

## サンドイッチ反復配列：CRISPR

真核生物のゲノムのかなりの割合を反復配列が占めていることは周知の事実である。原核生物でも、真核生物ほどではないが、反復配列は存在する。例えば、トランスポゾンやグループ I、グループ II イントロンなどは複数のコピーがゲノム中に存在している。これらは独自の機構によって増殖したり、転移したりしている。一方、数塩基から数十塩基の短い反復配列となると、複製エラーによって偶然生まれたものか、あるいは、独自の機構によって能動的に増殖するものなのか区別がつきにくい。短い反復配列のいくつかはレトロンの合成する msDNA がゲノム中に挿入されたものらしい。今回紹介する CRISPR は能動的に合成される反復配列と考えられている。

CRISPR は clustered regularly interspaced short palindromic repeat（群生性等間隔短回文反復配列）の略称であり、それまで SRSR（short regularly spaced repeat）などさまざまに呼ばれていた反復配列に Jansen らが付けた統一的呼称である（Jansen et al. 2002）。CRISPR は真正細菌と古細菌の両方の原核生物に分布するが、真核生物では報告されていない。CRISPR はサンドイッチかハンバーガーのような特殊な構造を持っている。特定の配列（サンドイッチのパンの部分）が等間隔で並び、その間隔（具の部分）には、反復しない配列（スペーサー）が挿入されている。反復する配列は同じクラスター内ではほぼ同じ配列で同じ長さのパリンドロームに似た配列だが、異なるクラスターでは異なる場合が多い。通常、20 塩基から 40 塩基程度である。CRISPR の反復配列が似ていると、片側の端には数百塩基の似た配列が存在する。この配列は leader 配列と呼ばれている。leader 配列には配列上の類似性は低く、唯一共通する特徴は AT リッチで同じ塩基のストレッチが観察されることくらいである。leader 配列の反対側の CRISPR の反復配列は途中で途切れていることが多い。

単純に同じ配列がタンデムに並んでいるだけならば DNA 複製の際のスリップによって誕生した反復配列に思えるが、CRISPR の最大の特徴は途中に挿入されている非反復配列の存在である。近年、CRISPR に含まれるスペーサーがバクテリオファージや conjugative plasmid の配列とよく似ていることが発見された

（Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005）。Mojica らはさまざまな生物から CRISPR を探索し、そのスペーサーのうち 88 個が既知の別の配列と相同性が高いことを発見した。その内、47 個はバクテリオファージに、10 個はプラスミドに、残り 31 個はそれ以外の染色体上に相同配列が見つかった。Bolotin らも同様の解析を行ない、相同配列を発見した 44 個のスペーサーの内、29 個がファージ由来であることを報告している。Pourcel らは *Yersinia pestis* の多数の株で CRISPR 中のスペーサーの配列を調べることで、新しく付加されたスペーサ

一を同定した。*Yersinia pestis* には3座位のCRISPRが存在している。同一個体内でのleader配列と反復配列は非常に似ているが、スペーサーは全く共通していなかった。株間では、スペーサーの配列と反復配列の数に多型が観察された。反復配列はleader配列から遠い側に付加される形での多型が認められ、そこに新たなスペーサーが挿入されていた。一方、反復数が減る場合もランダムであった。特徴的なことに、観察された31種類の新しいスペーサーの内、24個は実際にラムダ様プロファージの配列に由来していた。以上のことからCRISPRのスペーサーが失われる場合には、反復配列同士の組換えにより切り出される機構が想定され、一方でスペーサーが新しく付加される場合には、反復配列の新しい単位がleader配列から遠い側に作られ、同時にファージ、プラスミドなど寄生性遺伝因子の配列をコピーして間隔にスペーサーが挿入されるという機構が想定される。Bolotin とも *Streptococcus* 属の2種のたくさんの株間でCRISPRのスペーサーを比較し、新規のスペーサーの多くがファージの配列と類似していることを発見している。また、Bolotin らはスペーサーとしてコピーされるファージの配列の4塩基下流側にA/G-T/C-A-A-Aをコンセンサスとする5塩基が存在することを示している。この配列がスペーサー配列の選択とコピーに働いているのかもしれない。

CRISPRの合成機構に示唆を与えてくれるのはCRISPRとリンクする遺伝子の存在である。全てのCRISPRの付近に必ず存在する遺伝子はcas1と命名されており、配列の類似性からcas1Aとcas1Bに大別される(Jansen et al. 2002; Bolotin et al. 2005)。cas1Aの付近にはcas2、cas3、cas4の3種類の遺伝子が存在し、多くの場合、cas3-cas4-cas1A-cas2-CRISPRの順に並ぶが、並び順が異なる場合も多い。cas4とcas1Aはしばしば融合した蛋白質になっている。一方、cas1Bの付近には、cas2、cas3、cas4の代わりにcas5、cas6の2つの遺伝子が存在し、多くはcas5-cas1B-cas6の順に並んでいる。CRISPRの位置はさまざまだが、cas5の上流の場合が多い。このcas5、cas6は後述するHaftら(2005)のcas5、cas6とは別物であるので、ここでは区別のため、Bolotinらのcas5、cas6をcas5(Bolotin)、cas6(Bolotin)、Haftらのcas5、cas6をcas5(Haft)、cas6(Haft)と表記することにする。ところで、cas6(Bolotin)は原論文では他に近縁な遺伝子は存在しないと記されているが、Haftらの論文では、cas2に内包されている。そこで、試みに*Campylobacter jejuni*と*Wolinella succinogenes*のcas6(Bolotin)としてBolotinらの論文で報告されているCj1521Cとws1443の2つをクエリーにBLASTP検索を行なったところ、cas2やDUF196と名づけられたものがヒットしてきた。しかし、これらは実際には本来のcas2とは近縁ではなかった。Haftらの論文では、同一のHMMで検出していると表記されているのでもしかすると近縁なのかも

しれないが、判断がつきかねる。ここでは、別の遺伝子として扱うことにする。ちなみに cas5(Bolotin)は Haft らの論文の csn1 に対応する。

Cas1 は高い pI を持つ長さ約 350 アミノ酸残基の蛋白質である。機能は未知だが、保存されたアミノ酸からヌクレアーゼではないかと考えられている

(Makarova et al. 2002)。Cas2 は 100 アミノ酸程度の小さな蛋白質で他に類似した蛋白質は知られていない。600 から 800 アミノ酸の大きな蛋白質である Cas3 にはヘリカーゼスーパーファミリーII の 7 つのモチーフが見つかるため、ヘリカーゼであると予測される。Cas3 には HD ハイドロラーゼドメインを持つものもある。HD ヌクレアーゼドメインは独立した蛋白質としてコードされている場合もあるし、完全に失われている場合もある。Cas4 は 200 アミノ酸前後の蛋白質で、RecB エクソヌクレアーゼと近縁である。Cas5(Bolotin)は 1100 アミノ酸以上の巨大な蛋白質で中央部に HNH ヌクレアーゼのモチーフを持つので、ヌクレアーゼであろう。Cas6(Bolotin)は 100 アミノ酸程度の短い蛋白質で高い pI を持つが、既知の蛋白質との類似性は認められない。

これらを中心として、さまざまな遺伝子が CRISPR の周辺に分布することを Haft らは報告している (Haft et al. 2005)。cas5(Haft)は cas1A タイプの CRISPR と一緒に存在する遺伝子で 250 アミノ酸程度の蛋白質をコードする。N 末は保存性を示すが、それ以外の領域は近縁な CRISPR 同士でのみ類似する。cas6(Haft)も cas1A タイプの CRISPR と一緒に存在する遺伝子である。Cas6(Haft)は約 140 アミノ酸の蛋白質で C 末に GhGxxxxxGhG (h は疎水性残基) のモチーフを持っている。他には転写因子やポリメラーゼの機能が想定されている蛋白質や、機能未知の RAMP スーパーファミリーに属する蛋白質が複数の CRISPR に共通して見つかっている。Haft らは同時に、CRISPR をグループ分けし、cas2 が存在しないグループや cas4 が存在しないグループがあることも報告している。

共通する遺伝子として、ヌクレアーゼやヘリカーゼなど DNA に作用する遺伝子が見つかることから、これらは CRISPR の複製、転移などの機能を担っていると考えられる。しかし、同時にグループごとに遺伝子セットが大きく異なり、並び順もばらばらであるため、統一的な増殖機構は想定しづらい。あるいは普遍的に存在する cas1 だけで CRISPR の維持には十分なのかもしれない。

しかし、これらの蛋白質が CRISPR の複製にだけ関与しているわけではないのかもしれない。スペーサーの配列が寄生性遺伝因子であることから、これが原核生物の免疫系であると考えられる研究者もいる。興味深いことに、CRISPR が転写されていることが確認された (Tang et al. 2002)。CRISPR はリーダー配列から一方向にのみ転写されている。反復配列 1 つとスペーサー 1 つの長さを足した長さの倍数長の RNA が確認されることから、どうやら CRISPR はリーダー配列がプロモーターとなって全長が転写され、その後、反復配列の位置で切断さ

れるらしい。確認された RNA の 3' 末端のほぼ全てが反復配列の中央部であることは、切断が反復配列内部で起こることを示している。誰もがこの結果に真核生物の RNAi(RNA interference)/PTGS(post transcriptional gene silencing)の機構との類似性を認めるだろう。RNAi/PTGS の機構では、RNA は二本鎖の状態で切断され、一本鎖の相補鎖の形で mRNA の分解に関与する。そこでは、ヌクレアーゼ、ヘリカーゼ、ポリメラーゼなどの蛋白質が作用することが知られている。Cas1、Cas3 (HD モチーフを持つ場合)、Cas4、Cas5(Bolotin)はヌクレアーゼのモチーフを持っている。Cas3 はヘリカーゼであろう。ポリメラーゼも一部の CRISPR の近傍に存在する。これらから、CRISPR のスペーサー部分の配列が寄生性遺伝因子の配列と結合することによって選択的に分解する機構が想定できる。もしかしたら、Cas 遺伝子群は CRISPR の複製に関わるコアの部分と、生物学的機能を担うアクセサリーの部分との二重構造になっているのかもしれない。今後、蛋白質の機能解析を端緒に CRISPR の多くの謎が明らかになっていくことを期待したい。

Jansen R, Embden JD, Gaastra W, Schouls LM.

Identification of genes that are associated with DNA repeats in prokaryotes.

Mol Microbiol. 2002 Mar;43(6):1565-75.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.

Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.

Microbiology. 2005 Aug;151(Pt 8):2551-61.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E.

Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.

J Mol Evol. 2005 Feb;60(2):174-82.

Pourcel C, Salvignol G, Vergnaud G.

CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.

Microbiology. 2005 Mar;151(Pt 3):653-63.

Haft DH, Selengut J, Mongodin EF, Nelson KE.

A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes.

PLoS Comput Biol. 2005 Nov 11;1(6):e60 [Epub ahead of print]

Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV.

A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis.

Nucleic Acids Res. 2002 Jan 15;30(2):482-96.

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A.

Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*.

Proc Natl Acad Sci U S A. 2002 May 28;99(11):7536-41.

2006/01/20

小島 健司 著

禁 無断複写転載