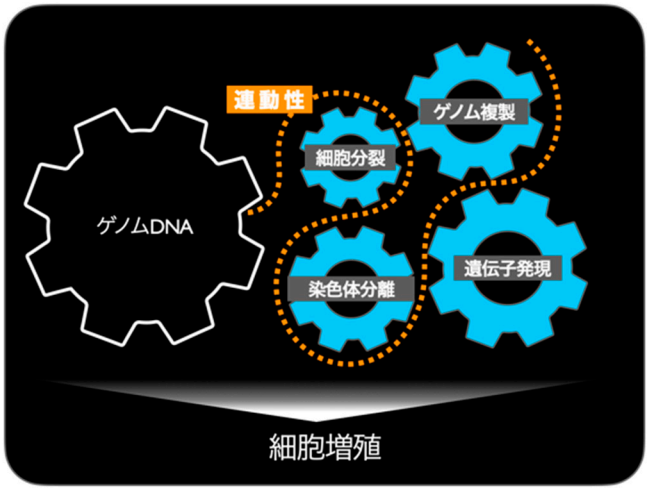
202　細胞増殖を動的に制御するゲノム構造体の理解	河野　暢明

【目的】細胞増殖のカスケードは細胞周期として理解され、真核生物ではチェックポイントなどと共に良く研究されている。しかしながら原核生物における細胞周期はほとんど解明されていないのが現状である。特にゲノム DNA それ自体はこれまで単なる情報として扱われることが多く、それ自体が与える影響は矮小化されてきた。情報密度の高い原核生物ゲノムでは一対の複製開始・終結点が対称に配置され、その軸から線対称に遺伝子の配向性、モチーフ配列の分布、そして塩基組成の偏りなどが存在している。ところがシミュレーションでは遺伝子発現の ON／OFF、バイオインフォマティクスでも遺伝子の有無だけが扱われ、こうしたゲノム情報構造と生命機能の関係性は未だ明らかにされていない。ところが実際はゲノムの情報構造を起点としたゲノム複製、染色体分離、核酸挙動、細胞分裂、そして各生化学反応といった複数レイヤーをまたぐシステム同士の連動性を支える高次な構造体としてシームレスな動態を制御している。ゲノム DNA が如何にゲノム複製、遺伝子発現、染色体分離、そして細胞分裂を連動させながら細胞増殖を機能させているのかを、合成生物学・バイオインフォマティクスを組み合わせたアプローチで解明し、細胞増殖という生命機能の発現・制御メカニズムの普遍的な説明を目指す必要がある。

【方法】ゲノム上の情報構造がより高次元の現象にどのような影響を与えているのかを定量的に理解するためには、バクテリアを対象としたマーカーレスのゲノム構造多型検出アルゴリズムを開発し、ゲノムシャッフリングの影響探索を可能にする必要がある。そこで本研究ではロングリードシーケンサーによるゲノム構造多型検出手法の開発を行った。さらに高効率に同時多発的な変異をゲノムスケールで引き起こすことのできる MAGE 法の新たな可能性について検証した。

【結果】　ロングリードシーケンサーを用いたゲノム構造多型検出プログラムの構築に成功した。このプログラムは既存ツールとカスタムスクリプトを組み合わせた解析パイプラインとなっており、低カバレッジのシーケンスデータにも対応可能となっている。MAGE 法の改良版となる TM-CspRecT-MAGE を検証したところ、条件付きでの変異導入効率の改善を観察することに成功した。しかし全ての環境で変異導入効率が改善されたわけではなく、今後の課題を明らかにした。

細胞増殖を導く連動性



202. 細胞増殖を動的に制御するゲノム構造体の理解

河野　暢明

慶應義塾大学 大学院政策・メディア研究科

Key words：ゲノム情報構造，合成生物学，構造多型

緒　言

微生物による「ものづくり」は古くは発酵食品やバイオ医薬品、近年ではバイオプラスチックや生分解性素材など、食糧、健康、医療、環境、エネルギー、そして材料などに通じる社会課題の解決や付加価値の増加に大きく寄与している。そしてこの「ものづくり」の鍵を握っているのが最適な宿主微生物の選定である。目的産物の毒性を抑え、如何に生産効率をあげられるかが最重要課題であり、それゆえに最適な宿主の不足が微生物ものづくりの発展を大きく妨げてしまっている。どんなに有用なマテリアルが自然界から発見されても、その生産効率は限りある選択肢の中でしか見込めない。そこでもし、目的産物に合わせて宿主微生物を創出する事ができれば、微生物ものづくりの可能性を飛躍的に高められるのではないか。ここで生まれる課題は、宿主微生物をどうデザインすればよいかという知見を我々人類は持ち合わせていないという点にある。既存の宿主微生物に付加価値を付ける程度であれば、生物機能を除去、改変、または移植するという遺伝学・構成生物学的なアプローチで「編集」することで済んできた。しかし目的産物に合わせてゲノムをフルスクラッチで「記述する」ためにはDNA合成やゲノム編集技術に加えて、ゲノムをデザインする際の設計ルールを徹底的に整備しなければならない。ところが現状、ゲノムスケールで遺伝子やタンパク結合サイトをどう配置するか、染色体を細胞内にどうパッキングするか、DNA コピー数をどの程度管理すべきかなど、ゲノムの構造はほとんど理解されていない。 例えば真正細菌の環状染色体で対称に位置する複製開始点と終結点のバランスを変えるだけで複製挙動や増殖効率が著しく低下し[1]、また遺伝子のもともと置かれていた鎖が逆になって方向が変わっただけで、周辺遺伝子を含む発現量が大きく抑制され、DNA 損傷や変異率が上昇したりする現象は個別には報告されている[2]。あるいは複製・修復関連タンパクのターゲットサイトとなるモチーフ配列もまたゲノム上で規則的に配置され、ゲノム情報構造を形成している。例えば複製終結に関わるTer配列は秩序だって向かい合わせに配置されており、それぞれがTusタンパクと結合することでフォークトラップを形成し、複製フォークの進行を阻害することで複製を特定の領域に終結させているが、報告者のこれまでの研究で Ter 配列の配置が複製挙動の安定維持に寄与していることが証明されて、染色体分離とは別のネットワークの存在が示されている［3，4］。つまり、複製、遺伝子発現、染色体分離、そして細胞分裂がゲノム高次構造体にどう制御されているかを構成的に記述、物理空間で検証を繰り返しながら、細胞増殖という生命機能の発現・制御メカニズムの普遍的な説明を目指す必要がある。

方　法

1.　構造多型検出手法の開発

配列処理にはseqkit v0.15.0［5］、マッピングにはBWA-MEM v0.7.15［6］、そしてマップデータ処理にはSAMtools v 1.3［7］を用いた。カスタムスクリプトは全てPerlで実装した。

2.　TM-MAGE の検討

TM-MAGEにはプラスミドpMA7-SacB（addgene#79968）および、CspRecTをクローニングしたpNK1255を用いた。pNK1255 は pORTMAGE-Ec1（Addgene #138474）から CspRecT 領域を PCR 増幅し、pMA7-SacB のλ-red Beta 領域と置き換える事で作製した。TM-MAGE は先行研究のプロトコルにしたがって実施した［8］。ここで

用いた ssDNA は全て *lacZ* を対象としており、効率は青白スクリーニングで計測した。

結果および考察

1. 構造多型検出手法の開発

ゲノム情報構造の定量的な意義を整備するため、本研究では多様なゲノム情報構造を網羅的に創り出し（ゲノムシャッフリング）、各構造変異株の表現型をオミクス解析で定量化することを目指した。その全てに先立ち、ゲノム情報構造の改変株を網羅的に獲得するためにはマーカーなどによるセレクションというボトルネックを排除する必要がある。そこで報告者はまず、ロングリードシーケンサーでゲノムシャッフリングを経たバクテリアゲノムの構造多型情報を得るためのプロトコル整備を行った。ロングリードから構造多型を検出するためには、リファレンスゲノム配列と比較してキメラリードに注目する必要がある。そこでまず、ロングリードの分割、リファレンスゲノムへのマッピング、キメラリードの検出、そしてキメラリードそれぞれがリファレンスゲノムのどの位置由来かの探索までを一気に行う構造多型解析ワークフローを開発した（図 1）。また同時に構造多型解析ワークフローが適用可能なロングリードシーケンスのシーケンスカバレッジの算出を行った。その結果 50×以上のカバレッジがあれば高い再現性を持って構造多型を検出できることが明らかとなった。1 回のナノポアシーケンスで 4 Gb のデータ量を出力できると想定した際に、20 サンプル以上のマルチプレックスが可能であることを示す。

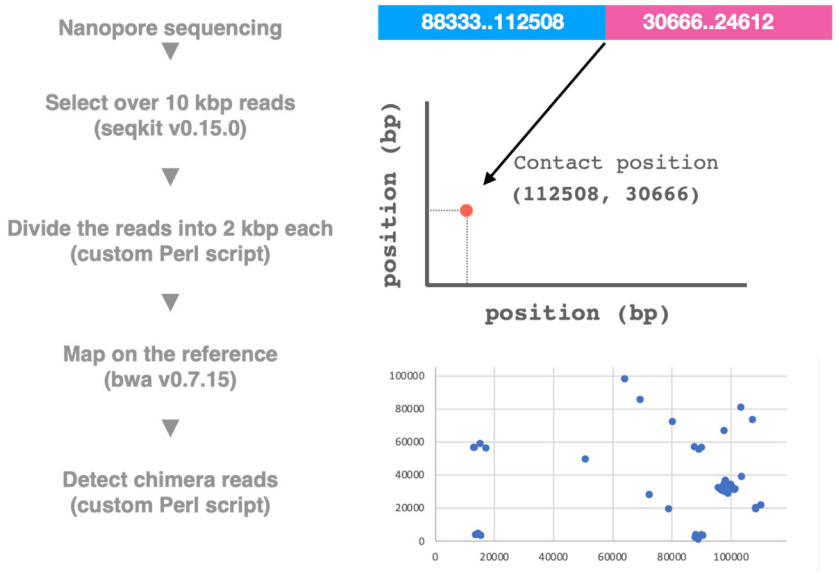


図 1. 構造多型検出アルゴリズムのフロー

ナノポアシーケンスで得られたシーケンスリードを 2 kbp 程度の短い断片に分割してリファレンスにマッピングする。その結果からキメラリードを検出し、それぞれの配列がリファレンス上どこ由来であったかを探索する。これで構造が改変されたスポットを発見することができる。

2. ゲノム情報構造改変ライブラリ構築手法の改良

ゲノムスケールで変異を及ぼすために我々は MAGE を用いることを想定した。MAGE は ssDNA をλ-red Beta タンパク質で組換える事で変異を起こす手法で、複数箇所をターゲットにする ssDNA を同時に組み込む事で同時多発的にゲノムスケールの変異を導入できるようになる。しかし MAGE は変異を安定維持するために *mutS* 変異を宿主ゲノムに導入する必要があった。*mutS* 遺伝子に変異があるとゲノムには自然発生的な変異が発生してしまうため、目的以外の改変が生まれてしまう危険性があるという問題点があった。そして最近になってλ-red Beta よりも効率的に

ssDNA 組換えを引き起こすことのできるタンパク質も発見された［9］。こうした背景を受けて、我々は従来の MAGE 法を改良する好機を迎えることができた。我々はメチル化によって変異箇所を保護することで、*mutS* 変異を不要とする TM-MAGE［8］を用い、λ-red Beta タンパク質とその代替案として考えられるリコンビナーゼで *Collinsella stercoris* phage 由来の CspRecT を比較解析した。その結果、2〜3 サイクル目までは CspRecT の方が効率的に組換えを引き起こしていたが、それ以降伸び悩み、最終的には変異導入効率は頭打ちになってしまった（図 2）。

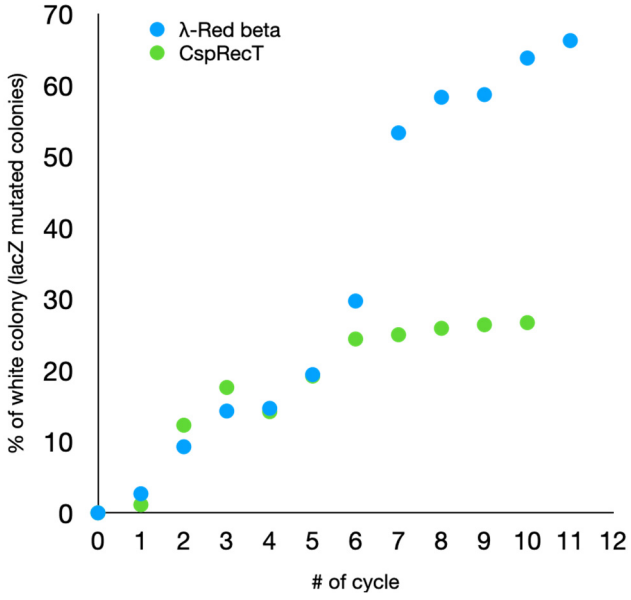


図 2. 組換えタンパクを変えた時の TM-MAGE 効率

TM-MAGE に用いる組換えタンパク質 2 種（λ-Red beta と CspRecT）の組換え効率を比較した図。TM-MAGE のサイクルを繰り返すごとにどちらも変異導入効率は上昇していったが、CspRecT の導入効率は一定以上には上がらなかった。

考　察

本研究で開発された構造多型の検出パイプラインは、マーカーレスな大規模ゲノムシャッフリングの網羅解析に用いることが可能である。通常ゲノムに変異を加えた際はその変異株を単離するために何かしらマーカーを用いる。しかしながらそのマーカーセレクションが増殖のバイアスを産み、構造多様性のボトルネックになってしまう。そのために多様なゲノム情報構造を得るためにはゲノムシャッフリングで生み出された多型をマーカーレスに単離・検出しなければならない。開発された構造多型の検出パイプラインはまだ変異株の単離には至っていないが、ヘテロな集団内における多型の割合や性質を網羅的に観察することを可能にした。今後はヘテロなゲノム集団を如何に単離していけるかが課題となっていくが、そのアイディアは既にあり、本成果を基に展開していく予定である。MAGE の効率をあげるべく、TM-MAGE の組換えタンパク質を CspRecT に置き換えた系を検証したが、予想に反して劇的な変異導入率の改善は見られなかった。この理由としては CspRecT の発現制御に起因している可能性が考えられる。また、MAGE サイクルを繰り返すごとに増殖効率が低下していったことから、宿主株への影響も考慮しなければならないことがわかった。

共同研究者・謝辞

本研究の遂行にあたり、シーケンス解析のサンプル処理をしてくださった高井幸および土門真綾に感謝申し上げます。最後に、ご支援を賜りました上原記念生命科学財団に深く御礼申し上げます。

文　献

1） Kuroki A, Toda T, Matsui K, Uotsu-Tomita R, Tomita M, Itaya M. Reshuffling of the Bacillus subtilis 168 genome by multifold inversion. J Biochem. 2008;143(1):97-105. Epub 2007/10/30. doi: 10.1093/jb/mvm197. PubMed PMID: 17965428.

2） Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, Pollock AJ, et al. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell. 2017;170(4):787-99 e18. Epub 2017/08/13. doi: 10.1016/j.cell.2017.07.044. PubMed PMID: 28802046; PubMed Central PMCID: PMCPMC5630229.

3） Kono N, Tomita M, Arakawa K. Accelerated Laboratory Evolution Reveals the Influence of Replication on the GC Skew in Escherichia coli. Genome Biol Evol. 2018;10(11):3110-7. Epub 2018/10/30. doi: 10.1093/gbe/evy237. PubMed PMID: 30371772; PubMed Central PMCID: PMCPMC6263442.

4） Kono N, Arakawa K, Sato M, Yoshikawa H, Tomita M, Itaya M. Undesigned selection for replication termination of bacterial chromosomes. J Mol Biol. 2014;426(16):2918-27. doi: 10.1016/j.jmb.2014.06.005. PubMed PMID: 24946150.

5） Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11(10):e0163962. Epub 2016/10/06. doi: 10.1371/journal.pone.0163962. PubMed PMID: 27706213; PubMed Central PMCID: PMCPMC5051824.

6） Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMCPMC2705234.

7） Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.

8） Lennen RM, Nilsson Wallin AI, Pedersen M, Bonde M, Luo H, Herrgard MJ, et al. Transient overexpression of DNA adenine methylase enables efficient and mobile genome engineering with reduced off-target effects. Nucleic Acids Res. 2016;44(4):e36. Epub 2015/10/27. doi: 10.1093/nar/gkv1090. PubMed PMID: 26496947; PubMed Central PMCID: PMCPMC4770203.

9） Wannier TM, Nyerges A, Kuchwara HM, Czikkely M, Balogh D, Filsinger GT, et al. Improved bacterial recombineering by parallelized protein discovery. Proc Natl Acad Sci U S A. 2020;117(24):13689-98. Epub 2020/05/30. doi: 10.1073/pnas.2001588117. PubMed PMID: 32467157; PubMed Central PMCID: PMCPMC7306799.