

グループIIイントロンの網羅的同定および系統解析

慶應義塾大学大学院 政策・メディア研究科

先端生命科学 (BI) 修士課程2年

三浦昌浩

背景

グループIIイントロン (G2I) は微生物のゲノム中に見られ、増殖能を持つDNA領域である。同イントロンの増殖によって、生命の設計図たるゲノムDNAの一部に書き換えが生じる。微生物の中には、同イントロンが増殖している種と、全く増殖していない種とが存在しているが、なぜこういった違いが見られるのか、その理由は明らかとなっていない。本研究ではこの要因の解明を目的とする。1万種以上のゲノムDNAを含む公共データベースから網羅的に同イントロンを収集し、その増殖性を左右する特性として、同イントロンの配列的特徴、および同イントロンを持つ生物のゲノム配列の特徴を検討した。

結果

前年度までに、G2I内部に存在する IEP (Intron-Encoded Protein) の領域を探索することで、G2Iを収集するバイオインフォマティクスパイプラインを構築していた。14,506のバクテリアゲノムから12,153のIEPを抽出し、クラスタリングにより代表的なIEPの1,949配列を得た。収集したIEPの中には、終止コドンで中断された配列や、明らかに一部の機能ドメインが欠落している配列が存在した。これらの非正規型のIEPと正規型のIEPとを分類したところ、非正規型のIEPは全1,949配列の約30%を占めた。IEPの種やタイプに関係なく、これらの非正規型のIEPはバクテリアの系統樹に広く分布していた。一方、場合によっては、正規型IEPの一部と類似した塩基配列が停止コドンの後に生じるため、一部の非正規IEP配列はリードスルーまたはフレームシフト機構によって発現する完全なIEPをコードしている可能性がある。非正規型のIEPの有無とG2Iの個数とに明らかな関連性は見られなかった。

Toroらは4,338個のG2I IEPを含む大規模な逆転写酵素データセットを構築している (Toro et al., 2019)。その論文に記載された4,338個のG2I IEP配列を用い、非正規IEPを含まない我々の1,392個のIEPとの類似性解析を行い、Toroらのデータセットにおける各IEPタイプを特定した。結果、彼らのデータセットを我々の方法で解析しても、データセット内のバクテリア全体の各IEPタイプの割合は我々の割合と非常に類似しており、彼らのデータセットは我々の分類を支持した。これに加え、我々のデータには非正規型のIEPの情報も含まれており、有用だと考えられる。

さらに、G2Iの個数とゲノム上の特徴との関連性を明らかにするために、バクテリアゲノムにおける複製バイアスの影響を反映する generalized GC skew index (GCSI) (Arakawa et al., 2009) を用いて以下の解析を行った。ここで、GC skewとは、グアニンとシトシンの分布におけるストランドの非対称性を示す指標であり、GCSIは各DNA配列のGC skewの偏りが強い場合は1.0、偏りが弱い場合は0へ近づく。ほとんどの細菌ゲノムのGCSIは0.1程度であり、GCSIが0.1以下のゲノムは、GC skewに明確な偏りが見られない (Arakawa et al., 2009)。図 1は、バクテリア門における GCSI と G2I の数

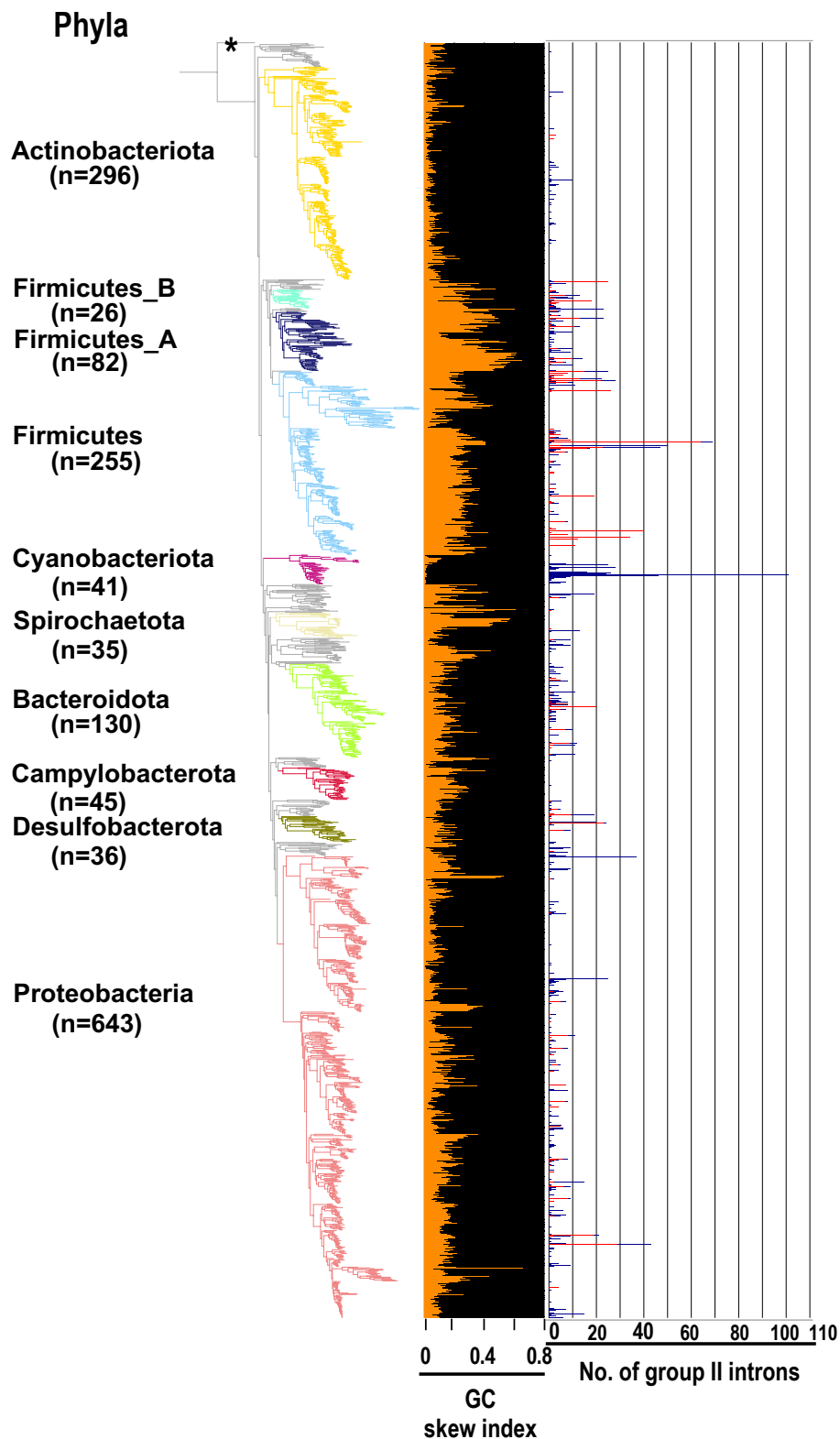


図1. バクテリアにおけるGCSIとG2Iの個数.

代表的な細菌全ゲノム (1,774種) のGC skew indexとG2Iの数を示す. 左側にバクテリア門 (Phyla) を示し, 系統樹の対応する各枝を色分けしている. 括弧内の数字は各門の種数を表す. アウトグループ (Candidatus Saccharibacteria oral taxon TM7x [RefSeq assembly accession: GCF_000803625.1]) の位置はアスタリスクで示されている. 中段のオレンジ色の線は, 各細菌種で最も長いゲノムのGC skew indexを示している. また, 右側にはG2Iの数を示した (赤線:bacterial-C型IEPを持つG2I, 青線:その他のIEPを持つG2I).

を示したものである。まず、細菌株とGCSI値は3つのグループでほぼ一致した。ActinobacteriotaとCyanobacteriotaでは、ほとんどのゲノムで GCSI は低く (それぞれ平均 0.06 と 0.03), 一方 , Firmicutesでは、ばらつきは大きいものの比較的高く (平均 0.28), また, GCSI は, G2I の数が多い (平均 0.28). さらに, 上述の通り, これら3分類群では, Actinobacteriota ではbacteiral-G [g1] 型, CyanobacteriotaではCL型, Firmicutesではbacterial-C型など, 主に特定のIEP型のG2Iが存在または増加していた (図1). しかし, 他の門では, GCSIのパターンは細菌の分類群との対応関係は不明瞭であり, 特定のIEP型の存在や増加は見られなかった.

総括

本研究では, G2Iを系統的に収集するバイオインフォマティクスパイプラインを開発し, 現在公開されている原核生物の全ゲノム (約15,000種) に適用することで, その全体像を明らかにした. 収集したG2Iを分類することで, IEPの配列に基づく新たなG2I群の存在の可能性や, 断片化した非正規型のIEPが原核生物に広がっていることを明らかにした. また, 収集したG2Iと宿主ゲノムの系統情報により, バクテリアとアーキアにおけるG2Iの詳細な分布を明らかにすることができた. G2Iは多くのバクテリア門に存在し (全バクテリアの25%), アーキアのHalobacteriota門に集中していた (全アーキアの9%). G2Iの数は概ね門レベルで規定されていたが, 近縁種間で大きく異なるケースも見られた. 原核生物におけるG2Iの劇的な増加は, 特定のIEPタイプと宿主の分類群の組み合わせで起こり, これらの増加は転写終結領域やGC skewといったゲノム上の特徴と関連している可能性がある.

参考文献

1. Arakawa K, Suzuki H, Tomita M. Quantitative analysis of replication-related mutation and selection pressures in bacterial chromosomes and plasmids using generalised GC skew index. *BMC Genomics*. 2009;10:640.
2. Toro N, Martinez-Abarca F, Mestre MR, Gonzalez-Delgado A. Multiple origins of reverse transcriptases linked to CRISPR-Cas systems. *RNA Biol*. 2019;16(10):1486-93.

主な業績

学会発表

1. **Masahiro C. Miura**, Shohei Nagata, Satoshi Tamaki, Masaru Tomita, Akio Kanai. Comprehensive Sequence Analysis Reveals Distinct Expansion Profiles of Group II Introns in Prokaryote, The 25th Annual Meeting of the RNA Society, Online, May 2021, Poster presentation.
2. **Masahiro C. Miura**, Shohei Nagata, Satoshi Tamaki, Masaru Tomita, Akio Kanai. Comprehensive Sequence Analysis Reveals Distinct Expansion Profiles of Group II Introns in Prokaryotes, The 22nd Annual Meeting of the RNA Society of Japan, Online, July 2021, Oral presentation.

査読付き論文

1. **Masahiro C. Miura**, Shohei Nagata, Satoshi Tamaki, Masaru Tomita, Akio Kanai. Distinct Expansion of Group II Introns during Evolution of Prokaryotes and Possible Factors Involved in its Regulation, *Frontiers in Microbiology*, in press.

謝辞

森泰吉郎記念研究振興基金はコンピュータおよび書籍の購入費用として使用させていただきました。ご支援に心より感謝申し上げます。ありがとうございました。