

# バイオインフォマティクス技術を駆使した 糸状菌トランスクリプトーム解析

兒島 孝明\*・岡 大椰・中野 秀雄

## はじめに

高速DNAシーケンシングは、従来のDNA配列解析法であるサンガー法に比べて圧倒的なスループットでのデータ取得を可能とする。たとえば、イルミナ社が提供するシーケンシングプラットフォームで最高スペックのNovaSeq™ 6000システムは、1ランあたり最大6 Tbの出力量を誇る(2018年12月現在)。これを用いた場合、1回の解析で得られる情報量はサンガー法の最高機種に比べて $10^7$ 倍以上も大きく、同時に解析できるサンプル数を考慮すればコストパフォーマンスの差は歴然である。このように、最近の技術革新と低コスト化の実現により、高速DNAシーケンシングは、生物工学分野を含むさまざまな分野で研究者に身近な存在になってきた。その一方で、高速DNAシーケンシングの解析から出力される膨大なデータは、多くの研究者にとって新たな悩みの種となっている。取得データが膨大であるがゆえに、その解析作業には多大な時間と労力を要するためである。そしてこの骨の折れる作業には、バイオインフォマティクス技術が必要不可欠である。

筆者らは、かねてからDNA結合型転写因子(以下、転写因子)とDNA間相互作用の解析に関する研究を行ってきたが、この過程で高速DNAシーケンシング解析の必要性に直面し、かつその威力を肌で実感した。本稿では、この高速DNAシーケンシングとバイオインフォマティクス技術を用いた糸状菌における転写因子を介した転写制御機構の網羅的解析法について紹介する。

## 高速DNAシーケンシングを用いた 転写因子の標的遺伝子のゲノムワイド解析システム

転写因子は、標的DNA配列に特異的に結合し遺伝子の発現を制御するタンパク質の総称である。それらは生体内のさまざまな機能に深く関与しており、多様な生命現象を理解するうえで、どの転写因子がどの遺伝子の発現量の調節に直接的/間接的に関与しているかを、網羅的に同定することは非常に重要である。

現在、転写因子の標的遺伝子の網羅的解析を目的として、さまざまなアプローチが試みられている。目的転写

因子を発現する野生株およびその欠損株よりmRNAを調製し、これを高速DNAシーケンシングに供するRNA-Seq、もしくは標的生物のcDNAを固定化したマイクロアレイに供するDNAマイクロアレイ解析は、転写因子依存的に発現レベルが大きく変化する遺伝子(発現変動遺伝子, differentially expressed gene; DEG)を網羅的に同定できる。しかし、これらの手法によって検出される発現変動は、目的転写因子の間接的な影響をも含むため、検出DEGが目的転写因子によって実際に直接的に制御されているかどうかを、この情報だけで判別することは難しい。この判別のためには、ゲノム上に点在する目的転写因子の結合部位の特定が必要である。

一方、gSELEX-Seq (genomic systematic evolution of ligands by exponential enrichment) は、標的生物由来のゲノムDNAより作製したライブラリーから目的転写因子に *in vitro* で結合するDNA領域を選択濃縮し、選択DNAプールの配列を、高速シーケンシングを用いて解析する手法である。gSELEX法の手順は、大まかに分けて、1) 目的転写因子とDNAの *in vitro* 結合反応、2) 転写因子-DNA複合体の分離、3) PCRによる選択的に濃縮された結合DNAの回収、の3ステップで構成される(図1)。このgSELEXは、*in vitro* で転写因子-DNA間の結合反応を行うため、クロマチン免疫沈降法(ChIP)と比べ、一連の操作を迅速かつ簡便に遂行できる。さらに、結合条件を自在にコントロールできることから、任意の選択圧条件下での結合DNAの選択的濃縮を再現性

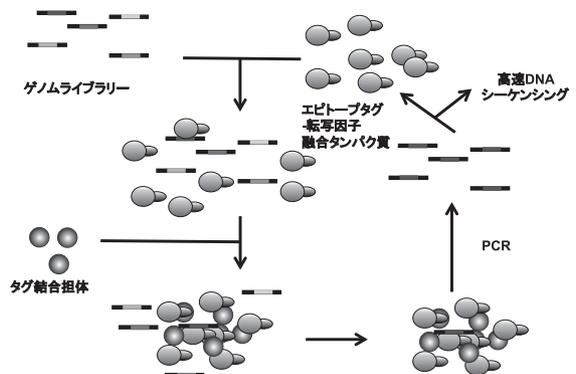


図1. gSELEX-Seqの概要(文献1のFig. 2を一部改訂)。

\*著者紹介 名古屋大学大学院生命農学研究科応用生命科学専攻(講師) E-mail: kojimat@agr.nagoya-u.ac.jp

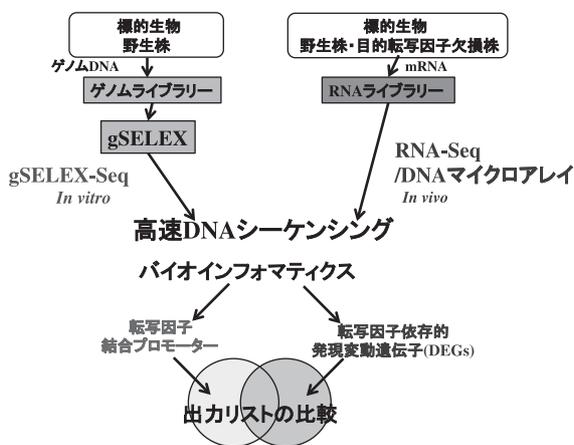


図2. 本研究の解析システムの流れ

良く行えることも大きな魅力である。

筆者らは近年、このgSELEX-Seqと上記発現変動遺伝子解析を相補的に用いることで、目的転写因子によって直接的に発現制御される遺伝子を、網羅的かつ高精度に検出するシステムの確立に成功した<sup>1)</sup>(図2)。gSELEX-Seqにより *in vitro* で濃縮選択したゲノム上の目的転写因子の結合領域の情報が得られ、一方、RNA-SeqもしくはDNAマイクロアレイ解析により、生細胞での各遺伝子の発現量の情報が得られる。この両手法によって出力された情報を比較することで、目的転写因子によって直接的に発現制御される遺伝子の検出が可能となる。

以下、この手法を用いた糸状菌由来の転写因子を介した転写制御機構の解析について2例紹介する。

### アミラーゼ代謝関連転写因子 AmyR の制御遺伝子の網羅的同定

AmyRは *Aspergillus* 属のアミラーゼ遺伝子群のプロモーター領域に結合し、その下流遺伝子の発現を促進する糸状菌由来のZn<sub>2</sub>Cys<sub>6</sub>型の転写因子である。筆者らは、かねてから *A. nidulans* 由来のAmyRの結合配列の解析に取り組んできた。このAmyRは、CGGN<sub>8</sub>CGG配列などに特異的に結合することが以前より知られていたが<sup>2)</sup>、筆者らは、ランダムなDNAライブラリーをマイクロビーズ上に作製し、セルソーターを用いてAmyRに結合するDNAを濃縮・解析した。その結果、このCGG tripletに挟まれたN<sub>8</sub>領域がまったくのランダムでなく、AmyRが結合を示す配列には、このN<sub>8</sub>領域にTが優先的に存在することを明らかにした<sup>3,4)</sup>。

このアプローチでは濃縮されたDNA断片をクローニングしたうえで、DNA配列を一つずつ地道に確認する

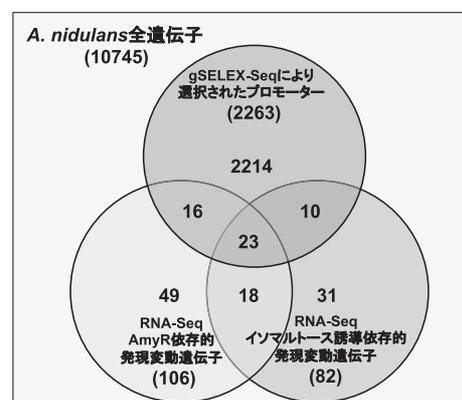


図3. gSELEX-SeqおよびRNA-Seqから得られたAmyR関連遺伝子数(文献1のFig. 5を一部改訂)。

手法を採用していたため、この一連のDNAシーケンス解析作業がボトルネックとなっていた。一方、結合部位周辺のDNA構造が転写因子の結合に影響を与えることが知られており<sup>5)</sup>、標的転写因子のゲノム中の結合部位をより正確に同定するためには、ランダム合成DNAライブラリーではなく、ゲノム由来のライブラリーを用いることが望ましいと考えられた。そこで、前述したgSELEX-SeqをAmyRの結合配列の解析に適用し、*A. nidulans*ゲノム中に点在するAmyRの結合部位の網羅的同定を試みた。

まず、*A. nidulans*ゲノムを剪断したDNAライブラリーを調製し、このライブラリーとMalE-AmyR融合タンパク質を用いたgSELEXを3ラウンド実施した。各ラウンドより選択されたDNAプールを高速DNAシーケンサーに供し、得られた配列データを用いてバイオインフォマティクス解析を行った。その結果、AmyRの既知の結合モチーフであるCGGN<sub>8</sub>CGGが抽出された<sup>1)</sup>。また、このCGG tripletに挟まれたN<sub>8</sub>領域には5番目の塩基をはじめTが比較的高い頻度で検出され、上述の筆者らのアプローチの結果と完全に一致した。さらに、AmyRに結合するDNA配列をプロモーター領域に有する2000種以上の遺伝子が同定された(図3)。これらの遺伝子は、AmyRによって発現制御される候補遺伝子ではあるものの、あくまでも *in vitro* での結合が確認されただけで、これらの遺伝子すべてが実際に生きた細胞内でAmyRによって発現制御されているとは限らない。

そこで、生細胞中でのAmyRを介した発現変動遺伝子の情報を取得するため、*A. nidulans* AmyR欠損株を用いたRNA-Seqを実施した。*A. nidulans*野生株およびAmyR欠損株をアミロース代謝関連酵素群の誘導剤であるイソマルトース誘導/非誘導の条件下で培養し、これ

らの菌株より調製したmRNAライブラリーを高速DNAシーケンサーによって解析した。これより得られた配列データをもとにバイオインフォマティクス解析を実施し、AmyR依存性およびイソマルトース誘導依存的な41種類の発現変動遺伝子 (DEGs) を検出した (図3)。

このRNA-Seqより得られた41種のDEGsとgSELEX-Seqより得られた候補プロモーター下流に位置する遺伝子リストを照合したところ、AmyRによる発現制御を受けることが既知であった8遺伝子のうちの7遺伝子 (*agdA*, *agdB*, *agdE*, *agdF*, *amyA*, *amyB* および *amyF*) を含む23種類の遺伝子が検出された。これらの遺伝子は、AmyR結合領域を上流に保持し、かつAmyRを介した発現変動を示すことから、AmyRによって直接的に発現制御を受ける遺伝子であると推定された。

以上の結果から、筆者らは、高速DNAシーケンシングおよびバイオインフォマティクス手法を駆使した本解析パイプラインが、目的転写因子によって直接的に発現制御される遺伝子を高精度かつゲノムワイドに検出できる転写制御機構の解析法であることを示した。

### 転写因子AoXlnR制御遺伝子の網羅的同定とデータ・マイニング

糸状菌はセルロースやキシランなどのセルロース系バイオマス分解に関与するさまざまな酵素を分泌することから、その関連遺伝子の発現制御機構の解明を目的としたさまざまな研究が行われてきた。この発現制御機構解明の鍵を握るタンパク質の一つ、AnXlnRは*A. niger*におけるキシラン、セルロース分解性酵素の発現に関与するZn<sub>2</sub>Cys<sub>6</sub>型の転写因子として単離され、GGCTAAAなどを標的配列として結合することが知られている<sup>6)</sup>。また、このAnXlnRの*A. oryzae*におけるオルソログ、AoXlnRも同様の機能を保持し、その結合DNAモチーフがGGCT(A/G)Aであること、キシランやセルロース代謝関連の8種の遺伝子 (*xynF1*, *xynG1*, *xynG2*, *xyIA*, *celA*, *celB*, *celC* および *celD*) がAoXlnRの制御下にあることが知られている<sup>7,8)</sup>。これらに加えて2009年にはDNAマイクロアレイ解析によってペントース代謝関連遺伝子をはじめ、AoXlnRの過剰発現に応答する遺伝子が多数同定されている<sup>9)</sup>。

しかしながら、上述のようにDNAマイクロアレイ解析によって同定されたこれらDEGsが、実際にAoXlnRによって直接的に発現制御されているかどうか判別することは非常に難しい。このため、AoXlnRの発現制御機構の詳細な解析には、ゲノム上のどの部位にAoXlnRが結合し、どの遺伝子の発現制御に直接的に関与するかと

いう知見が必要不可欠であった。そこで筆者らは、上記DNAマイクロアレイ解析によって得られた既存の発現変動データに加えて、gSELEX-Seq解析を行い、バイオインフォマティクスを駆使することで、このAoXlnRによって直接的に発現制御される遺伝子の網羅的同定およびその制御機構を解明することを試みた。

まず、*A. oryzae*ゲノムライブラリーとMalE-AoXlnR融合タンパク質を調製した後、gSELEXを3ラウンド実施し、高速DNAシーケンシング解析を実施した。これによって得られた配列データを*A. oryzae*ゲノム配列にマッピングし、モチーフ解析を行ったところ、(c/a)GGcT(A/g)(A/t)(a/t)がAoXlnR結合DNAモチーフとして抽出された (図4)<sup>10)</sup>。これは、既知のAoXlnRの結合DNAモチーフであるGGCT(A/G)Aを含むことから、AoXlnRに結合するDNA断片が優先的に選択されたと考えられる。最近Ishikawaらは、AoXlnRはモノマーとしてCGGNTAA(A/T)、ダイマーとしてTTAG(G/C)CTAAにそれぞれ結合することを報告している<sup>11)</sup>。今回検出された結合DNAモチーフはこのモノマーに対する結合DNAモチーフに類似していることから、本gSELEX-SeqではAoXlnRモノマーに対する結合配列が優先的に選択濃縮されたと推定される。

一方、NoguchiらによるDNAマイクロアレイ解析により、AoXlnR発現に関連する75種類 (その後、データベースの更新により72種類) のDEGが同定されている<sup>9)</sup>。そこでまず、gSELEX-Seqによって出力されたデータをもとに、開始コドンより上流1000 bpにAoXlnR結合サイトを保持する約2000の遺伝子リストを作成した。次に、この遺伝子リストと上述のDEGsを照合したところ、AoXlnRの制御下にあることが既知の上記キシランやセルロース代謝遺伝子8種を含む51遺伝子において重複が確認された<sup>10)</sup>。すなわちプロモーター領域に結合部位を有し、かつ、発現変動を示すこれらの遺伝子は、AoXlnRによって直接的に発現制御される遺伝子であると考えられる。

さらに、AoXlnRによって調節される転写機構をより詳細に解析するために、AoXlnR依存性発現変動と制御

	Round 1	Round 2	Round 3
Motif			
E-value	3.2e <sup>-58</sup>	2.4e <sup>-78</sup>	2.2e <sup>-88</sup>

図4. AoXlnRの結合DNAモチーフ (文献10のFig. 2を一部改訂)

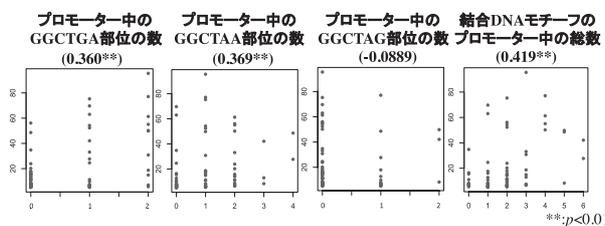


図5. AoXlnR 結合DNAモチーフとAoXlnR 依存的発現変動レベルの相関散布図（文献10のFig. 3を一部改訂）

プロモーター中のAoXlnR結合配列の頻度をパラメーターとしたデータ・マイニングを実施した。その結果、AoXlnRに高い親和性を示すことが以前より知られており、かつ、本結合DNAモチーフ解析において明瞭に検出されたGGCTAA配列の頻度と、AoXlnR依存的発現変動レベルと間に一定の相関が確認された。また、非常に興味深いことに、AoXlnRとの親和性がGGCTAA配列に比べ比較的低いとされていたGGCTGA配列の頻度とAoXlnRを介した下流の遺伝子発現との間にも、GGCTAA配列と同等の相関関係があることが示された(図5)<sup>10</sup>。

これらに加え、プロモーター中にGGCTAA、GGCTGA二種類の結合DNAモチーフを保持する場合、発現変動レベルが高い値を示すことを統計的に示した<sup>10</sup>。実際、これら二種類の結合DNAモチーフを保持する*xynF1*プロモーターを用いたレポーターアッセイでは、GGCTAA部位への変異導入により、下流レポーターの発現が劇的に低下する一方、GGCTGA部位への変異導入でも、レポーターの発現レベルが半減することが示されている<sup>7,8</sup>。本研究におけるこのデータ・マイニングの結果により、プロモーター中におけるGGCTAAとGGCTGA部位の共存がAoXlnRによる下流遺伝子の発

現活性化に寄与することが示唆された。

これらの一連の結果は、バイオインフォマティクスを駆使することでAoXlnRを介した転写制御機構の核心により迫ったことを示すものである。今回得られた知見は、*A. oryzae*におけるAoXlnRを介した転写制御ネットワークの解明に大きく寄与すると考えられる。

### おわりに

gSELEX-Seqによる標的転写因子のゲノム上の結合部位情報と、標的転写因子による遺伝子の発現変動情報から構成されるビッグ・データのバイオインフォマティクス解析により、遺伝子発現ネットワークのより詳細かつ包括的な理解が可能となる。本稿では糸状菌由来の転写因子を用いた例を述べるにとどめたが、本解析パイプラインは汎用性が高く、ここで紹介した微生物にとどまらず、生物系産業に用いられている、動植物の転写因子の機能解析にも実装可能である。

### 文 献

- 1) Kojima, T. *et al.*: *PLoS One*, **11**, e0159011 (2016).
- 2) Tani, S. *et al.*: *Biosci. Biotechnol. Biochem.*, **65**, 1568 (2001).
- 3) Kojima, T. *et al.*: *J. Biosci. Bioeng.*, **109**, 519 (2010).
- 4) Wang, P. *et al.*: *Biosci. Biotechnol. Biochem.*, **76**, 1128 (2012).
- 5) Dror, I. *et al.*: *Genome Res.*, **25**, 1268 (2015).
- 6) van Peij, N. N. *et al.*: *Mol. Microbiol.*, **27**, 131 (1998).
- 7) Marui, J. *et al.*: *Fungal Genet. Biol.*, **35**, 157 (2002).
- 8) Marui, J. *et al.*: *FEBS Lett.*, **528**, 279 (2002).
- 9) Noguchi, Y. *et al.*: *Appl. Microbiol. Biotechnol.*, **85**, 141 (2009).
- 10) Oka, H. *et al.*: *BMC Genomics*, **20**, 16 (2019).
- 11) Ishikawa, K. *et al.*: *Curr. Genet.*, **64**, 1245 (2018).