

# 大規模生物情報を利用した 糸状菌の有用遺伝子探索

SATテクノロジー・ショーケース2015

## ■ はじめに

糸状菌や植物、放線菌が生体内で合成する二次代謝産物は極めて多様な構造を持つ低分子化合物である。そして、その多くが生理活性を持つため、創薬や化粧品、機能食品開発等の分野で有用であると考えられている。例えば、抗生物質であるペニシリンや高コレステロール血症の治療薬であるロバスタチンなどは糸状菌から発見された二次代謝産物として広く知られている。二次代謝産物の生合成に関わる遺伝子は生物実験のみで同定することが難しく、明らかになっていないものが数多く存在する。一方で近年、ハイスループットな次世代型のDNAシーケンサーの登場により一生物種の持つ全DNA配列(ゲノム)の決定が高速・低コスト化されつつあり、公共データベース中には微生物を中心に様々な生物種のゲノム情報が急速に蓄積されている。これを受けて生物種ごとの網羅的な遺伝子の情報に対し情報技術を利用して、二次代謝産物の生合成に関する遺伝子を予測し、遺伝子機能同定を効率的に行うための研究が進められてきた。

## ■ 活動内容

### 1. 次世代シーケンサーを利用した全ゲノム決定

糸状菌の多くはすべての染色体を合わせて、 $3 \sim 4 \times 10^7$ 塩基程度のゲノムサイズを持つことが知られている。これに対して次世代シーケンサーは、一本当たり数十～数百塩基の短いDNA配列を $10^9$ 本オーダーで読むことのできる装置である。そのため大量の短い配列間の重なりを利用してゲノム配列の再構成(アセンブル)を行う必要がある(図1)。ゲノムアセンブルを行うソフトウェアは様々なものが知られているが、多くの場合再構成されるゲノム配列断片の数や長さはパラメータ設定に大きく影響を受ける。また次世代シーケンサーがDNA配列を読む際に個々の塩基に対して出力する信頼度の情報をもとに、アセンブルに利用するデータのフィルタリングを行うことも結果に影響を与えることが知られている。我々の研究室では麹菌のシーケンスデータを利用して効果的なゲノムアセンブルの方法について検討を行った(1)。

(1) Umemura M *et al.*, 2013. PLoS One 8:e63673

### 2. ゲノム間の比較解析による二次代謝系遺伝子の予測

生物において多くの場合、遺伝子は単体ではなく他の遺伝子と協調的に働くことで一つの機能を成している。

例えば二次代謝系では、原料となる化合物から複数の化学反応を経て目的の化合物が合成されることが多く、それぞれの反応において触媒の機能を果たす酵素の情報をコードしている遺伝子や酵素遺伝子の働きを調整することで反応を制御するタンパク質をコードしている遺伝子などが協調的に働く。糸状菌の二次代謝系においては多くの場合、一つの機能に関連し協調的に働く一連の遺伝子群がゲノム配列上で近接した位置にコードされていることが知られている。さらに、異なる生物種において異なる二次代謝産物の生合成に関する遺伝子群の間であっても類似の配列を持った遺伝子から構成されている例が多数見つかっている。我々の研究室ではこのような二次代謝系遺伝子のゲノム配列上の特徴を利用し、文字列比較のアルゴリズムを応用することで、複数の生物種のゲノム上の遺伝子の並びを比較し、二次代謝系遺伝子を予測するMIPS-CG法を開発した(図2)(2)。

(2) Takeda I *et al.*, 2014. DNA Res. 21(4):447-57

## ■ 関連情報等(特許関係、施設)

MIPS-CG法 web service

<http://www.fung-metb.net/>



図1. 次世代シーケンサーを用いた全ゲノム決定

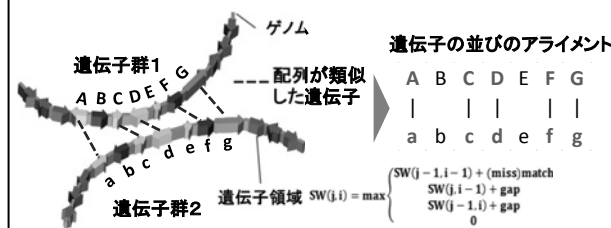


図2. 比較ゲノム解析を用いた二次代謝系遺伝子予測

代表発表者 竹田 至 (たけだ いたる)  
所属 東京農工大学 工学府 生命工学専攻  
連携 町田・中村研究室 博士後期課程  
問合せ先 〒184-8588 小金井市中町 2-24-16  
TEL: 042-388-7081 FAX: 042-388-7736

■ キーワード: (1) バイオインフォマティクス  
(2) ゲノム  
(3) 二次代謝