

麹菌プロジェクト — 麹菌ゲノム解析への取り組み —

CBRCメンバー：浅井 潔、金 大真、熊谷俊高、寺井悟朗（以上、数理モデルチーム）、長崎英樹（アルゴリズムチーム）

麹菌(*Aspergillus oryzae*)は、我が国の伝統的発酵産業である清酒・味噌・醤油の製造に1000年以上の長期にわたって利用されてきた糸状菌であり、産業上もっとも重要な微生物の1つです。

1998年、我が国の国立研究所・大学・民間企業から構成される麹菌EST解析コンソーシアムが設立され、2000年までに17,000を超えるEST配列が解析されました。翌年には麹菌ゲノム解析コンソーシアムが立ち上げられ、麹菌の全ゲノム解読が始まりました。このプロジェクトでCBRCは、解読されたドラフトゲノムシーケンスに含まれる全ての遺伝子をいち早く解読するという、配列解析の部分を担当することになりました。

2003年第2四半期までに麹菌のゲノム配列の95%以上を確定し、我々はドラフトゲノム配列に含まれる全遺伝子のセットを予測しました。本稿ではCBRCが担当した、麹菌ゲノム解読の概要をご紹介します。

遺伝子発見の手法

麹菌ゲノムには、ゲノムがすでに解読された酵母やアカパンカビなどの生物が持つ既知のタンパク質配列と高い相同性を持つ遺伝子と、既知タンパク質データベースからの相同性検索ではうまく見い出せない麹菌特有の遺伝子の両方があると考えられます。一般的に、後者の予測は前者より難しく、そこでまずは既知のタンパク質と相同性がある遺伝子を全て麹菌ゲノムから抽出し、それ以外のゲノム上の領域から麹菌特有の遺伝子を探すという方針を取りました。

我々はまずBLASTXやALN^[1]などのソフトウェアを用いて、麹菌ゲノムに「埋まって」いる既知タンパク質配列と相同性の高い遺伝子群を同定しました。ALNは主にCBRCで開発されたアライメントソフトウェアで、真核生物ゲノムDNA配列とアミノ酸配列をアライメントする機能を持っています。この手法で遺伝子が見つからないゲノム上の領域については、CBRCで開発されたGeneDecoder^[2]とTIGRによって開発されたGlimmerM^[3]の両ソフトウェアを用いて麹菌特有の遺伝子予測を試みました。

上記の予測遺伝子群全てを矛盾のないように統合し、最終的に約14,000の遺伝子候補を同定したのです。

発見した遺伝子のアノテーション

予測された遺伝子について、その機能のある程度予測しカテゴリ分類を行うというアノテーション作業が必要となります。分類にはNCBIのCOG (Clusters of Orthologous Groups) を参照し、それぞれの予測遺伝

子にCOGの機能コードを割り振ることで機能予測を行いました。また、酵母や線虫などのすでに解析が進んでいるモデル生物のCOGに含まれる配列との比較をグラフィカルに行えるツールも作成しました。これらに基づいて2003年第4四半期に国内の糸状菌研究者による集中的なドラフトアノテーション作業が行われたのです。各予測遺伝子についての詳細な解析は現在進行中です。

また2003年に、近縁の糸状菌である*Aspergillus nidulans*、*Aspergillus fumigatus* の解析を進めているSanger Centre やTIGR、Whitehead Institute (MIT) といった欧米の複数の研究所と、配列情報や解析データなどを共有することが決定し、*Aspergillus*属のゲノム解析を全世界的に推進する体制が整いました。今後数ヶ月から数年のうちに、麹菌のゲノムレベルでの描像は飛躍的に明らかになっていくと期待されます。

(筆：熊谷俊高)

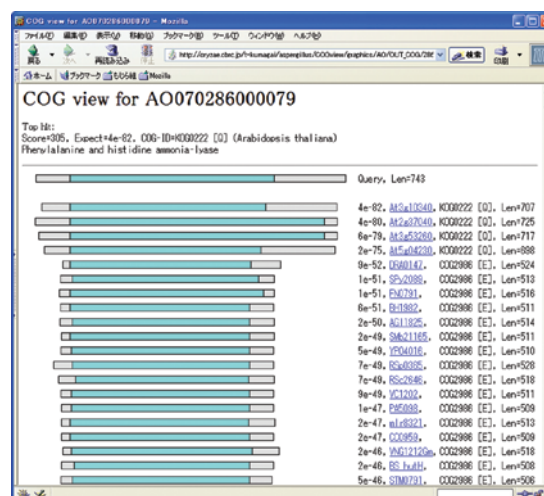


図 麹菌予測遺伝子と他のモデル生物の遺伝子との比較

[1] Gotoh O., Homology-based gene structure prediction: simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps. *Bioinformatics*. 16, 190-202 (2000).

[2] Asai K, Itou K, Ueno Y, Yada T., Recognition of human genes by stochastic parsing. *Pac Symp Biocomput*. 228-239 (1998).

[3] Majoros WH, Pertea M, Antonescu C, Salzberg SL., GlimmerM, Economy and Unveil: three ab initio eukaryotic genefinders. *Nucleic Acids Res*. 31, 3601-3604 (2003).