

生体分子情報データベースの開発
Development of Databases for Biomolecular Information

情報・システム研究機構 データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター 五斗 進

研究成果概要

本研究では、京都大学化学研究所スーパーコンピュータシステムを利用し、生体分子情報データベースおよびバイオインフォマティクス技術の開発に取り組み、その成果をゲノムネット (<http://www.genome.jp/>) から広く公開している。平成 29 年度は、ウイルスゲノム系統解析のためのツールを開発し、ViPTree として公開した (<http://www.genome.jp/viptree/>) [2]。ViPTree では、Virus-Host Database に登録されているリファレンスウイルスゲノムの Proteomic tree が予め計算され、それを利用者の好みに応じて表示する。また、利用者がアップロードしたウイルスゲノムが Proteomic tree 上のどこに位置するかを計算し表示する。メタゲノムなどから新規ウイルスゲノムを同定した後にそれがどのようなウイルスかを調べるのに役に立つツールである。

また、引き続き開発を進めている MAPLE システムを応用し、溶菌酵素生産菌 *Lysobacter enzymogenes* M497-1 の遺伝子機能予測を試みた。MAPLE で代謝ポテンシャルを解析し、充足率が 100% に満たないモジュールをリストアップしたところ、充足率が 80% 以上かつ *Q* 値が 0.5 以下のものが 14 個あった。そのうち、近縁 *Lysobacter* 属で充足率が 100% のものは Ornithine biosynthesis (M00028)のみであった。そこで、このモジュールの抜けの部分 (アミノ酸 N-アセチル転移酵素) に対応する遺伝子を、ゲノムネットの遺伝子機能相関解析ツール GENIES で探索したところ、LEN_3144 が候補に含まれていた。アミノ酸配列やゲノム上の位置などをさらに調べたところ、この酵素遺伝子の候補として適当であることが判明した。逆に、近縁 *Lysobacter* でこの酵素に割り当てられている遺伝子が、実際には LEN_3144 のオーソログではなく、近縁 *Lysobacter* の機能割り当ての見直しが必要であることも判明した。

その他、ゲノムネットのメタゲノム・ゲノムデータを活用した解析と予測ツールの開発を行った [1,3]。

発表論文(謝辞あり)

1. Nishiyama, H., Nagai, T., Kudo, M., Okazaki, Y., Azuma, Y., Watanabe, T., Goto, S., Ogata, H. and Sakurai, T.; Supplementation of pancreatic digestive enzymes alters the composition of intestinal microbiota in mice. *Biochem. Biophys. Res. Commun.*, **495**:273-279 (2017).
2. Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H. and Goto, S.; ViPTree: the viral proteomic tree server. *Bioinformatics*, **33**:2379-2380 (2017).
3. Shimizu, Y., Ogata, H. and Goto, S.; Discriminating the reaction types of plant type III polyketide synthases. *Bioinformatics*, **33**:1937-1943 (2017).