

PC クラスタによるゲノム解析

¹ 国立遺伝学研究所 生命情報・DDBJ センター, ² 東京理科大学薬学部
阿部貴志¹、宮崎智²、菅原秀明¹

1995年に *H. influenzae* の全ゲノム配列決定が行われて以来、そこにさらに大きな展開をもたらした。1995年から2004年8月までに、国際塩基配列データベースINSD (International Nucleotide Sequence Databases: DDBJ/EMBL database/GenBank) に登録公開されたゲノム配列件数(菌株数)は、180株を超えている (Genome Information Broker : <http://gib.genes.nig.ac.jp/> [1])。原核生物に限っても、ここ数年で1,000種類以上の全ゲノム配列が決定・公開されるといわれている。現在、ゲノム配列情報は、生物に関わるあらゆる研究にとって必須のデータ資源となっている。しかしながら、1995年以来さまざまなグループによって解析されたデータとアノテーションがINSDに蓄積されてきたため、それらを一様に取り扱うことは躊躇される。一方、多数のゲノム配列データとその生物学的意味(アノテーション)をINSDからオンラインで入手できるようになったこと、また、同じ種に属する複数の菌株のゲノム配列のセットもいくつか公開されているため、比較ゲノム (comparative genomics) として、複数の生物種のゲノム配列情報を比較しながら data mining (データ・マイニング) することによって、生物の多様性、遺伝子の機能、ゲノム構造の進化などについて新たな知見を獲得する試みが行われるようになった。

しかしながら、大量、かつ複雑なゲノム配列情報から比較ゲノムの観点により、広範囲な生物の多様性の実体などを統合的に理解するためには、情報資源の統一化、大規模データからの特徴抽出、PCクラスタのような大規模な計算機の利用などが必須となりつつある。

本公演では、我々が、

- 1、情報資源の信頼性の観点から、微生物ゲノムの遺伝子位置の同定、かつ統一的なアノテーションといったゲノム配列情報を比較するために必要な基礎的なデータをゲノム配列情報から共通の手法で整理しようという試みを行った例。
- 2、比較ゲノムとして、特に生物多様性の解明の観点から、自己組織化マップ (Self-Organizing Map) を用いて [2]、ゲノム配列に潜む多様な生物種固有の特徴抽出を行った例。

を中心にゲノム解析におけるPCクラスタの必要性とゲノム解析の実例を報告する。

参考文献

- 1, Fumoto, M., Miyazaki, S. and Sugawara, H. (2002). *Genome Information Broker (GIB): data retrieval and comparative analysis system for completed microbial genomes and more*, Nucleic Acids Res. **30**: 66-68
- 2, Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. and Ikemura, T. (2003). *Informatics for unveiling hidden genome signatures*. Genome Res. **13**: 693-702

付録 : プレゼンテーション資料

<http://www.ssken.gr.jp/nl/2004/sci/1/ppt4.pdf>