

課題名 (タイトル) :

ハブ (*Protobothrops flavoviridis*) のゲノムアセンブリ

利用者氏名 : ○三澤 計治*

所属 : *本所 情報基盤センター 技術開発ユニット

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

ハブ (*Protobothrops flavoviridis*) は、世界でも、鹿児島県の奄美群島と、沖縄県の沖縄諸島にのみ分布する日本固有種です。ハブは、毒の激しい作用で広く知られていて、平成 24 年度は鹿児島県内で 60 件以上、沖縄県内では 90 件以上の咬症が報告されています。残念ながら、毒が作用する仕組みはまだ明らかになっていないことも多く、毒液を構成するタンパク質成分についても未解明な点が多く残されています。

ハブゲノム配列が決定されれば、毒液作用機序の解明への手掛かりとなり、よりよい咬症治療薬の開発へと繋がるのが期待されます。さらに、ハブゲノム配列から見つかるタンパク質は、創薬シーズになる可能性があります。

2. 具体的な利用内容、計算方法

ゲノム配列決定に現在広く使われている次世代シーケンサは、100 文字程度の短い配列情報をたくさん出力します。このような配列情報から、ゲノム全体を復元するためにはアセンブルという作業をしなくてはなりません。このアセンブルには *de novo* アセンブルと *mapping* アセンブルという 2 つの方法があります。

de novo アセンブルとは、参考となるリファレンスゲノム配列が無い状態で、多数の短い配列を適切に繋ぎ合わせて、全ゲノムを構築する作業のことです。「*De novo*」とは「新規」という意味です。さまざまな *de novo* アセンブリのプログラムを試したところ、どれもメモリと計算時間を沢山使うことがわかりました。

他方、*mapping* アセンブルでは、近縁種のゲノム配列を利用して、高速にアセンブルを行うことができます。しかしながら、*reference* 配列として使われた配列が、近縁種のものではない時には、アセンブルエラーが生じることがわかっています。

そこで、私は、Barrows-Wheeler 変換による *index* を用いた一致配列探索と、BLAST の *extension* のアルゴリズムを組み合わせ、1 億年ほど昔に分岐した生物種でも、エラーが少なく、*mapping* アセンブルできるプログラムの開発に成功しました。このプログラムによって次世代シーケンサの配列を *mapping* アセンブルする時に、比較的遠い生物種のゲノムを *reference* として利用できます。この成果は投稿準備中です。

また *mapping* のエラーを引き起こす原因として繰り返し配列がありますが、それを取り除くプログラムも開発し、出版しました (Misawa 2013, *Genomics* **102**: 35-)。

3. まとめ

次世代シーケンサの配列を *mapping* アセンブルする時に、比較的遠い生物種のゲノムを *reference* として利用することができるプログラムの開発に成功しました。

4. 今後の計画・展望

今後は、今年度開発されたプログラムを用い、最近決定されたキングコブラのゲノム (Vonk et al. 2013, *PNAS* **110**: 20651-)、ならびにビルマニシキヘビ (Castoe et al. 2013, *PNAS* **110**: 20645-) を用い、ハブのゲノム解析を行って行きたいと考えています。

平成 25 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

Misawa, K., RF: A method for filtering short reads with tandem repeats for genome mapping.
Genomics **102**: 35-37 (2013)

【国際会議などの予稿集、proceeding】

【国際会議、学会などでの口頭発表】

三澤計治、ホモロジーが低い時でも高速にホモロジーを見つけるプログラムの開発、日本遺伝学会 85 回
大会、2013/9/19-21、神奈川県横浜市

三澤計治、京速コンピュータ「京」を用いた二型糖尿病のハプロタイプ関連解析、日本人類伝学会第 58
回大会、2013/11/20-23、宮城県仙台市

【その他】