

課題名 (タイトル) :

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

利用者氏名 : ○三澤 計治

所属 : 情報基盤センター計算工学応用開発ユニット

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 (*Protothrops 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できれば、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、すでに決められているゲノム配列に対し、読まれた配列情報がどの場所に対応するのかを推定する **mapping** という作業をする必要があります。以降、次世代シーケンサで決定した配列を **read**、**mapping** の対象となる配列を **reference** と呼びます。

生物が持つゲノム配列は、同種内の個体でも少しずつ違いますし、ましてや別種になると配列の違いは大きくなります。よく似ているが、少し異なる配列を対応させるためには、統計的な扱いが必要ですし、**reference** のサイズは数ギガバイト程度ありますので、高速計算も必要となります。

2013 年度に、私は、**read** の中に繰り返し配列があると、**mapping** の誤りが大きいことを示し、それをあらかじめ取り除く手法を開発しました (Misawa 2013, *Genomics* **102**:35-)。さらに、昨年、**reference** 側に繰り返し配列があるような場合で

も、**mapping** の成功率が下がることが示されました (Li et al. 2014, *BMC Bioinformatics* **15**:2-)。

そこで、私は、**reference** ゲノム配列中に、繰り返し配列がどのように分布しているのか、**hash table** を使い調べる方法を開発しました。

3. 結果

ハブのゲノム配列はまだ決定されていないので、比較的広く研究が進んでいるヒトゲノムを **reference** として、繰り返し配列がどのように分布しているか調べました。染色体の末端とセントロメアに繰り返し配列が多いことがわかりました。その他の部分では 8 割ほどは **unique** な配列であることがわかりました。この結果は論文で出版しました (Misawa 2015, *Austin J Comput Biol Bioinform.* **2**: 1010-)。

4. まとめ

繰り返し配列が多い所は、染色体末端やセントロメアなど、重要な遺伝子があまりない所が多いため、この領域を避けて **mapping** していけば、重要な遺伝子配列のかかなりの割合を決定できると推測されます。

5. 今後の計画・展望

2013 年に、ヘビの仲間のニシキヘビ (Castoe et al 2013, *PNAS* **110**: 20645-) ならびにコブラ (Vonk et al. 2013, *PNAS* **110**: 20651-) のゲノム配列が決定されました。今後は、これらを **reference** 配列として用い、昨年度および今年度に研究された繰り返し配列に注意しながら、ハブのゲノム配列を決定していきたいと考えています。

平成 26 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. (2015) Distribution of unique sequences in the human genome. *Austin J Comput Biol Bioinform* 2015, 2:1010.

【国際会議などの予稿集、proceeding】

Misawa, K. and Kamatani, N. (2014) ParaHaplo: A program package for haplotype-based whole-genome association study on k computer. The 5th AICS International Symposium.

【国際会議、学会などでの口頭発表】

三澤計治「MAFFT を応用した、エラー率が高い時でも使えるホモロジーサーチ法の開発」日本進化学会第 16 回大会 2014 年 8 月 21 日

三澤計治「MAFFT を応用した、エラー率が高い時でも使えるホモロジーサーチ法の開発」日本遺伝学会第 86 回大会 2014 年 9 月 18 日

三澤計治、田嶋文生「遺伝子の違いの地理的分布の理論的研究」第 59 回日本人類遺伝学会 2014 年 11 月 21 日

三澤計治、田嶋文生「遺伝子の違いの地理的分布の理論的研究」バイオスーパーコンピューティング 2015 ウィンタースクール 2015 年 1 月 30 日