

武漢コロナウイルスのホモログマップ作成法の紹介

天野晃 *

* 無所属

*amano.au1@gmail.com

概要 2020 年初頭、新型コロナウイルスの感染が拡大し、各国で緊急事態宣言が発せられるまでに至った。当該ウイルス（だけではないが）の感染検査には、主に PCR 法が用いられるが、プライマー設計はその精度を左右する大きな要因の一つである。特に False-positive を忌避する場合は他のウイルス/生物のゲノム（断片）のコンタミネーションに対してロバストである必要があるが、ホモログマップによるゲノム特徴の可視化は、その判断の参考となる。本報告では、ホモログマップの作成法について紹介する。また、医療系、生物系が専門でない参加者の方々のために、テクニカルタームの説明を付録として用意する。ポスター閲覧の際の参考にされたい。

Wuhan corona virus homologue mapping

Kou AMANO*

*independent

1 はじめに

ウイルス等の感染検査の一つに、PCR 検査がある。この検査は、文字通り検出対象となる DNA(RNA)断片が存在するかを、PCR 増幅により直接的に検出・確認する方法である。PCR の際には、ターゲットとなるゲノム断片の一部と相同性を持つ、さらに短い DNA(RNA)断片をプライマーとすることによりターゲットを特異的に増幅させるが、当然、ターゲット以外にもプライマーと相同な配列を持つゲノム（断片）は存在し、これらがコンタミネーションを起こしている場合は、False-Positive を導く。そのような場合も配列解析を行うことによりより正確な検出が可能となるが、コストは大きくなる。

2 目的

PCR プライマーの設計において、ロバストネスの判断の参考となり得る、簡易かつ低コストなホ

モログマップの作成方法を紹介する。

具体的には、(1) 宿主側ゲノムに対するマップ、(2) ウイルスゲノムに対するマップ、(3) 自身のゲノムの特徴化、について述べる。

3 マッピング方法

3.1 宿主側

[DB 側ゲノム] 宿主のゲノムとして、turkey、rock pigeon、pig、rabbit、Mouse、human、ferret、dog、cat、camel、beluga、bat を用いた。配列情報の取得は、NCBI のサイト [1] より 2020 年 3 月に行った。完全ゲノムを用いた種と、全ゲノムショットガンシーケンシングの結果を用いた種を含む。DB 作成は、makeblastdb コマンドにより、デフォルトで行った。

[クエリー側ゲノム] 武漢コロナウイルス完全ゲノム、MN908947.3。megablast を利用。クエリー条

件は、10 塩基以上のマッチ。その他はデフォルト。

3.2 ウイルス側

[DB 側ゲノム] 配列情報の取得は、NCBI のサイト [1] より 2020 年 3 月に行った。NCBI のサイトには、ウイルスおよびファージのゲノム配列がまとめられたセクションがあり、これを一括ダウンロードした。

[クエリー側ゲノム] 3.1 に同じ。

3.3 自身のゲノム

自身のゲノムにおいて、どの領域がどの程度のコピー数を持つか、を示すものである。

[DB 側ゲノム] ゲノムを 25 塩基ごとに分割したデータを作成し、これをもとに DB を作成した。

[クエリー側ゲノム] DB 作成時と同じ配列情報をクエリーとして、megablast にて 10 塩基以上のマッチを行った。その他はデフォルト。

[Window-fourier] 次のような条件にて、フーリエ変換を行った。

1. 塩基情報を次のように数値列に変換する : A - i
1、G - i I、C - i -I、T - i -1

4 マッピング結果の可視化

4.1 バーマップ

4.2 頻度マップ

4.3 フーリエマップ

4.4 領域マップ

附録：テクニカルターム解説

- PCR : Polymerase Chain Reaction の略。DNA ポリメラーゼ（合成酵素）を利用して DNA を複製する系。DNA 鋳型、プライマー、合成酵素、DNA の構成要素である、アデニン、グアニン、シトシン、チミン等をバッファに投入し、温度サイクルを作成することにより DNA 複製が可能となる。産物を検出する際は電気泳動を行う。またはリアルタイム定量的（逆転写）PCR を用い同時に検出するのが一般的である。
- 電気泳動 : DNA やタンパクなど、電荷を持つ分子の分離を行う系。蛍光マーキング等により視覚的に産物の確認を行う。
- リアルタイム定量的（逆転写）PCR : DNA の定量を目的とする PCR。蛍光マーカー等を用い、これを測定することにより産物の量を計測（推測）する。増幅中にリアルタイムに計量を行う。RNA 量を計測する際には逆転写を行うので、このように呼ばれる。
- プライマー : PCR の際、鋳型 DNA に結合し合成開始のプライマーとなる短い DNA 断片。
- 全ゲノムショットガンシーケンシング : 配列決定を行う際、chromosome 全体を読み取することは困難なため、ゲノムをある程度の大きさに切断してシーケンシングを行い、後に計算機により可能性の高い配列を（接合）推

測することが一般的である。完全に接合されていない状態での配列情報をこう呼ぶ。

- 相同性：特に遺伝子およびアミノ酸の相同性を指す。基本は文字列の相同性を基にしているが、置換を受けやすい／受けにくいペアが判明しており、マッチングには遺伝学の知識が反映されている。
- blast： ”Basic Local Alignment Search Tools”の略。DNA（RNA）およびアミノ酸配列の相同部分を検索するシステム。複数のコマンドからなり、主にデータベース作成コマンド、データベース検索コマンドに分かれる。
- ホモログ： 検索による相同部分、あるいは相同な遺伝子をこう呼ぶ。

注・文献

- [1] <<https://ftp.ncbi.nlm.nih.gov/>>, (参照:2020-03)
- [2] 情報メディア学会編集委員会. 情報メディア学会誌論文等執筆要綱, 2003
- [3] 杉本重雄. メタデータについて: Dublin Core を中心として. 情報の科学と技術. Vol. 49, No. 1, 1999, p.3-10.
- [4] Fidel, R. Qualitative Methods in Information Retrieval Research. Library & Information Science Research. Vo.15, No.3, 1993, p.219-247.
- [5] 天野晃. 自己組織化による DNA シークエンスの分類. 情報メディア学会第 5 回研究会発表資料. 2003, p. 5-8.