

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）は存在する可能性があり、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、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。完全ゲノム情報を用いた種と、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

4.4 相同領域マップ

25 ベースごとにヒットした領域数を折れ線グラフで表す。武漢コロナウイルスの対象となるゲノム位置にあわせて表示する。プライマーに対応する配列が複数箇所が存在すれば、増幅されるコピー数が異なることになる。

4.5 高頻度ホモログ配列

全ゲノム対全ゲノムの megablast の結果より、ヒットが高頻度となる領域に対応する塩基配列パターンを抽出した。用いた脊椎動物ゲノム全て、およびウイルス・ファージゲノムに対してそれぞれ頻度を 5 回以上、21 回以上とした。さらに、それぞれ重なり合うホモログ群より、すべてに出現する塩基配列パターン部分を抽出した。

4.6 プライマー位置

マニュアル [5] に示されるプライマーの位置を武漢コロナウイルスのゲノム上にマップした。

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 の構成要素であるデオキシヌクレオチド（塩基が、アデニン、グアニン、シトシン、チミンの 4 種）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、増幅と同時に検出するのが一般的である。[5]
- 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。
- コーディング領域：DNA においてタンパク質に翻訳される領域。あるいは、そうであると予想される領域。
- ORF : "Open Reading Frame" の略。タンパク質への翻訳は、開始コドンであるメチオニンから始まり、終止コドンである 3 種のトリプレットで終了することが知られている。この領域をこう呼ぶ。
- コドン：DNA がタンパク質（アミノ酸配列）に翻訳される際、3 塩基が 1 組でひとつのアミ

ノ酸をコードする。この3塩基のコード（トリプレットコード）をコドンと呼ぶ。

注・文献

- [1] <<https://ftp.ncbi.nlm.nih.gov/>>, (参照:2020-03)
- [2] Sergey V. Petoukhov. The genetic code, 8-dimensional hypercomplex numbers and dyadic shifts. <http://symmetry.hu/isabm/petoukhov.html>.
- [3] GUY DODIN, PIERRE VANDERGHEYNST, PATRICK LEVOIR, CHRISTINE CORDIER, LAURENCE MARCOURT. Fourier and Wavelet Transform Analysis, a Tool for Visualizing Regular Patterns in DNA Sequences. J. theor. Biol. No. 206, 2000, p.323-326.
- [4] V. R. Chechetkina, V.V. Lobzinc aEngelhardt. Large-scale chromosome folding versus genomic DNA sequences: A discrete double Fourier transform technique. DOI:10.1016/j.jtbi.2017.05.0.
- [5] 病原体検出マニュアル 2019-nCoV Ver.2.6. <<https://www.niid.go.jp/niid/images/lab-manual/2019-nCoV20200217.pdf>>. (参照:2020-03)