

3. 研究内容(別紙)

研究課題

内在性ウイルス配列の探索による節足動物ウイルスの多様性および進化の解析

堀江 真行¹

大阪府立大学大学院生命環境科学研究科

研究目的

研究の背景

これまでのヒトや家畜等の新興感染症の多くは、ヒト以外の動物を自然宿主とするウイルスによって引き起こされてきた。このような新興感染症を引き起こすウイルスは、顕著な被害が出たときにはじめて同定される。しかし、このような対応では感染源となる動物の特定などにおいて遅れが生じ、今後の新興感染症の発生時に被害が拡大するおそれがある。そのため、ウイルス感染症の制御には、どの動物に、どのようなウイルスが感染しているのかを事前に知っておくことが有用である。すなわち、ウイルスの多様性を徹底的に解明する必要がある。実際に、関連した種々のプロジェクト(Global virome プロジェクトなど; <http://www.globalviromeproject.org>)が世界中で展開されている。しかし、未だウイルスの探索は十分ではなく、地球上のウイルスの99.99%以上は未同定であると考えられている(Geoghegan et al., Open Biol. 2017)。

節足動物はジカウイルスやデングウイルスのような新興感染症から、日本脳炎や黄熱といった様々なウイルス感染症を媒介する。節足動物においても上記の通り、世界中においてウイルスの探索が行われている。しかしその探索は十分ではなく、さらなるウイルス探索が必要であると考えられている(Geoghegan et al., Open Biol. 2017)。

現在のウイルス探索法の主流はウイルス分離およびハイスループットシーケンス解析である。両者ともに様々なメリット・デメリットが存在するが、どちらの手法においても、採材した個体に感染しているウイルスのみしか検出ができないという欠点がある。ウイルスの多様性の解明には、過去に感染してきたウイルスについて知ることも必

須であるが、これらの手法では過去のウイルス感染について知ることはできない。

生物のゲノムにはウイルス由来の遺伝子配列が多数存在し、それらは内在性ウイルス様配列(EVE: endogenous viral elements)と呼ばれる。EVE は祖先生物におけるウイルス感染を示す「ウイルスの分子化石」であり、EVE を解析することによって過去のウイルス感染に関する貴重な情報(年代、宿主、遺伝子配列)が得られる(図 1)。申請者らはこれまでに、ボルナウイルスという RNA ウイルスをモデルとして、ゲノムデータベースに存在する EVE を網羅的に探索・解析し、脊椎動物における約 1 億年間のボルナウイルス感染履歴を明らかにした(Horie et al., Nature 2010; Kawasaki et al., PNAS 2021 など)。これらの研究では、これまでに魚やヘビにしか感染しないと思われていた系統のウイルスが、過去には霊長類やコウモリを含む様々な哺乳動物に感染していたことがわかり、現代においても哺乳動物に未知のウイルスが存在することが示唆された。このように EVE を解析することによって、現存するウイルスの研究のみからは得られない貴重な知見が得られる。

目的

本研究では節足動物、特にヒトや哺乳動物に感染症を媒介し得る「蚊」やその近縁動物におけるウイルスの解明を目的とし、EVE を利用して過去のウイルス感染の解明さらには分子系統解析等により有用な情報を得る。本研究では特に、EVE 検出の基礎情報を得るための最初のステップとなる EVE 検出法の改良について研究を行った。

研究内容と研究成果

1. EVE 検出方法の改良

はじめに、EVE 検出方法の改良とその検討を行った。これまでの EVE の検出方法の主流は、現代のウイルスのタンパク質のアミノ酸配列をクエリー、生物のゲノムの塩基配列をデータベースとして用いる tBLAST 検索であった(Katzourakis et al., PLoS Genet 2010)。tBLASTn は最も標準的な方法であるものの、フレームシフトが考慮されないため、変異によってフレームシフトが頻繁に起こっている EVE については十分に検出できない可能性がある。そのため、存在する EVE を見逃す可

能性がある。

FASTA は BLAST と同様に古くから使用されている配列類似性検索ソフトである(Pearson Curr Protoc Bioinformatics 2016)。FASTA パッケージでは tBLASTn と同様にアミノ酸配列をクエリーとした塩基配列データベース検索を行う tfastx および tfasty プログラムが存在する。特筆すべきは、tfastx および tfasty ではフレームシフトを考慮した検索を行うことができるということである(Pearson et al., Genomics 1997)。そのため、上記の tBLASTn の欠点を補い、今まで見逃していた EVE を検出し、より網羅的に過去のウイルス探索ができる可能性がある。そこで本研究では tfastx および tfasty を用い、EVE 検出法の改良とその検討を行った。ここでは検出法の検討材料として、染色体レベルでゲノムが利用可能であり、これまでに tBLASTn を用いた EVE 探索が何度も行われたことのあるヒトゲノムを用いた。また、ヒトゲノムに存在する唯一のレトロウイルス以外の EVE が存在する、ボルナウイルスの配列(RefSeq データベースに存在する *Bornaviridae* のアミノ酸配列)をクエリーとして用いた。

上記のクエリーとデータベースを用いて、tBLASTn、tfastx、tfastyにより検索を行った。その結果、tBLASTn、tfastx、tfastyそれぞれにおいて、24、26、27 遺伝子座位の配列がボルナウイルス由来 EVE の候補としてヒットした。それぞれの手法においてヒットした遺伝子座位の関係をベン図で表す(図 1)。なお、tBLASTn による結果はこれまでの過去の報告と一致した(Kawasaki et al., PNAS 2021)。

tfastx あるいは tfasty による検索でヒットした遺伝子座位は、tBLASTn によって検出した 24 遺伝子座位のすべてを含んでいた(図 1)。tfastx と tfasty はそれぞれ tBLASTn では検出できなかった遠い類似性を持つ遺伝子座位についても検出することが可能であった。

次に、tBLASTn では検出できなかった 4 遺伝子座位のボルナウイルス由来 EVE 候補配列について詳細な解析を行った。配列類似性検索および比較ゲノム解析を行った結果、tfastx および tfasty の両方で検出された遺伝子座位については過去の挿入配列、すなわち EVE であることが示唆された。一方、tfasty でのみ検出された遺伝子座位のうち1つについては単純な繰り返し配列であり、その他の 2 遺伝子座位については周辺に反復配列が多く存在するため判断が困難であった。これらのことから、tfastx および tfasty を用いた検索は tBLASTn よりもより高感度に EVE

を検出できることが示唆された。なお計算時間を考慮し、以降の解析では tfastx のみを用いた。

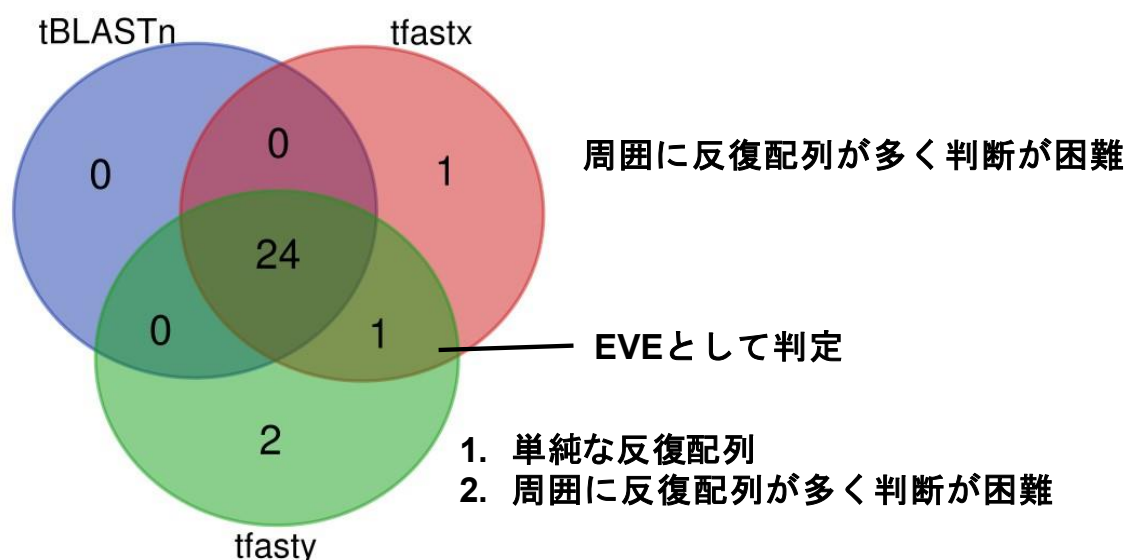


図 1. tBLASTn、fastx、fasty の比較

それぞれの手法によってヒットした遺伝子座位についてベン図で示した。クエリーおよびデータベース配列は本文を参照。作図には Draw Venn Diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) を用いた。

2. ヒトスジシマカおよびネッタイシマカゲノムにおける EVE の探索

次にヒトスジシマカ (*Aedes aegypti*) およびネッタイシマカ (*Aedes albopictus*) のゲノムにおける EVE の探索を行った。これらの生物のゲノムにはフラビウイルス由来の EVE が多数存在することが報告されている。本研究では、クエリー配列としては NCBI RefSeq データベースに存在するフラビウイルス (*Flaviviridae*) のアミノ酸配列を用いた。なおヒトスジシマカのゲノムは5つ、ネッタイシマカのゲノムは6つアセンブリが公開されているため、すべてを用いて検索を行った。tBLASTn と tfastx による検索を行った結果、ヒトゲノムにおける結果と同様に、tfastx では検出できるが、tBLASTn では検出できないフラビウイルス由来 EVE 候補が存在した。

現在、これらの配列について EVE であるかどうかを詳細に検証している。また、過去の報告とも照合し、検出した EVE 候補配列が新規 EVE であるかどうかも調査中である。さらに、フラビウイルス以外のウイルスについても tfastx および tfasty を用いた大規模検索を行っている。

3. まとめ

本研究において、従来の tBLASTn 検索では検出できなかった EVE の存在が示唆された。本研究によって得られた知見は今後の EVE 探索の礎となるとともに、2種の生物ゲノムから得られたデータは今後の節足動物における過去のウイルス感染についての知見を得るための重要なデータとなるであろう。

成果発表

なし

今後の問題点

本研究では EVE 探索法を改良することによって、今までに見過ごしていた EVE を検出することが可能となったと考えられる。一方で、これらの配列は極めてスコアが低く、統計的な信頼をあらわす E-value が比較的高かった。そのため、検出した配列については偽陽性である可能性も考えられるため、特に E-value が高いヒット配列については、ヒトゲノムでの検証例のような詳細な解析が必要であると考えられる。

今後はこれらの研究成果を基礎として、挿入年代の推定を行うとともに、その他の系統のウイルスや生物種へと拡張し、過去のウイルスの多様性に関する情報を得る。