

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 9 日現在

機関番号：32508

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25440195

研究課題名(和文) 昆虫類核ゲノムに転移した細菌由来遺伝子群の探索とその進化的役割

研究課題名(英文) Lateral gene transfer from Bacteria to Insect genome and its evolutionary impact.

研究代表者

二河 成男 (Nikoh, Naruo)

放送大学・教養学部・教授

研究者番号：70364916

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：種の障壁を越えた水平転移による新規遺伝子の獲得は、遺伝情報の進化において重要な役割を担う。昆虫類の全ゲノムの塩基配列情報から、水平転移により細菌から獲得した遺伝子を同定し、昆虫がこれらの転移遺伝子群をどのように獲得したか。そして、それらの転移遺伝子が、どのような制約を受けて進化しているかを解明することを目指して研究を行った。細菌由来の転移遺伝子を持つ可能性が高いと考えられる、必須共生生物を内部共生させている生物種やその近縁種を対象として、水平転移遺伝子の検索を進めた。合計9種について解析を行い、数の違いはあれ、いずれの昆虫種からも、細菌由来の転移遺伝子の候補となるゲノム領域の同定に成功した。

研究成果の概要(英文)：Acquisition of genetic material through lateral gene transfer is important for the genome evolution of organisms. Our aim for this project is to reveal the evolutionary history of transferred genes and their functions in the recipients. Using assembled genome data of various insects, we tried to find transferred genes, their potential donors and their evolutionary patterns. Initially, we used genomes from 9 insects, which harbor obligate symbionts or are closely related with species establishing a symbiotic relationship with bacteria. All of the genomes investigated contained the candidates of bacterial-derived sequences.

研究分野：分子進化学

キーワード：遺伝子水平転移 共生 ゲノム

1. 研究開始当初の背景

原核生物のみならず、真核生物でも様々な生物種でゲノム解析が行われた結果、これまでの安定的、固定的なものであった生物の遺伝情報の進化に対するイメージは、変更を余儀なくされた。それは、塩基置換や重複といった自身のゲノムが持つ、既存の遺伝子情報の改変に加えて、種の障壁を越えた水平転移による新規遺伝子の獲得によって、生物の持つ遺伝情報は進化しているということである。生物の遺伝情報の進化は、突然変異とその集団への浸透、そして固定という過程を経る。遺伝子の水平転移による進化も、当然この道筋が当てはまる。異なる種の遺伝情報がある個体の生殖細胞の DNA に取り込まれ、それが遺伝的浮動や自然選択によって集団中に広まり、そして固定した結果を、私たちは観察している。

生殖細胞と体細胞が発生初期に分離する動物では、外来の DNA が生殖細胞と接する機会が少なく、水平転移の頻度は低いのではないかと考えられていた。しかし、共生細菌が細胞内によく観察される昆虫類や線虫類では、生殖細胞でも細菌と接触する機会が多く、事実、水平転移により獲得した遺伝子とその核ゲノム内に保持する例が多数報告されている。このような水平転移による遺伝子の獲得が、宿主昆虫の生存や繁殖にどのような利点をもたらしているのかは興味を持たれる。

2. 研究の目的

ゲノム解析技術の進展により、比較的ゲノムサイズが大きい生物群である動物においても、様々な生物種でゲノムプロジェクトが進行している。特に昆虫類では、様々な種での全ゲノム配列の決定が進行中である。本研究では、昆虫類を中心として、データベースに登録されている全ゲノム決定が進められているアセンブルされた全ゲノムの塩基配列情報から、細菌由来の水平転移により獲得

した遺伝子を同定し、昆虫がこれらの転移遺伝子群を、どのように獲得し、そしてそれらの転移遺伝子が、どのような制約を受けて進化しているかを解明する。そして、生物進化における遺伝子水平転移の役割を明らかにすることを目的とする。

3. 研究の方法

(1) 相同性検索用データベースの準備 極めて大量のデータについて相同性検索を行うため、既存のあらゆる配列情報が登録されているデータベースでは計算時間が膨大になる。よって、国際 DNA データベースから主要な昆虫のタンパク質アミノ酸配列と細菌のタンパク質アミノ酸配列をダウンロードし、それぞれについてデータベースを作成する。さらに、それらのデータベースに対して、Blast によって相同性検索可能なシステムをワークステーション上に構築する。

(2) 昆虫ゲノムの塩基配列情報の準備 次に、昆虫のゲノム配列情報を準備する。多くの昆虫のデータがあるが、水平転移遺伝子を保持する可能性が高いと考えられる、必須共生細菌を持つ昆虫を中心に行う。

(3) 相同性検索 (2)で得られた塩基配列は、アセンブルされたものであり、配列長が長いものが多い。目的としている細菌に由来する転移遺伝子が部分的にしか含まれていない場合、何も処理をせずに相同性検索の問い合わせ配列として用いると、検出できない可能性が高い。したがって、1000 塩基以上の配列は、500 塩基の重なりを許して、1000 塩基以下の断片に分割する。このようにして作成した昆虫の問い合わせ配列を用いて、(1)で作成したデータベースに対して Blast による相同性検索を行う。

(4) 細菌のタンパク質と類似した塩基配列の抽出 相同性検索で得られた相同性の評価スコアを比較して、他の昆虫よりも細菌のタンパク質とより高い相同性を示す塩基配

列を抽出する。これらの候補を問い合わせ配列として、今度は国際 DNA データベースが作成している、既知のあらゆるタンパク質の情報を含む non-redundant データベースに対して、Blast による相同性検索を行う。これによって、既知のタンパク質の中で最も相同性が高いものを同定する。昆虫ゲノムにコードされているにも関わらず、細菌と最も相同性が高い遺伝子を、水平転移により獲得した遺伝子の候補とする。

(5) アライメントや分子系統樹を用いた確認

(4) の過程で得られた他の相同性を示すタンパク質も加えて、アライメントを行い、統計的にも有意な相同性があるかどうかを確認する。また、アライメントより分子系統樹を推定し、その樹形が生物種の系統関係と統計的にも有意に矛盾するかどうかを確認する。このような過程を経て、明らかに水平転移により獲得された遺伝子を検出する。

(6) 分子系統進化学的な解析

水平転移由来の遺伝子と判断したものについては、その由来と本来の機能を推定する。方法は、(5) で作成した分子系統樹を元に、その由来となった細菌種を同定する。また、その情報から供与側の生物におけるタンパク質の機能を検索し、昆虫での機能や役割を推定する。また、転移遺伝子に機能する上で必要な読み枠が存在するか、同義座位と非同義座位での置換率を比較し、機能的な制約が働いているかといった点から、現在機能しているかどうかを確認する。

また、その遺伝子の供与側となった生物において、転移した遺伝子の近傍にある他の遺伝子が水平転移しているかどうか、あるいは、転移した遺伝子が酵素なら、それが関わる代謝経路を構成する他の酵素遺伝子が水平転移しているかどうかを調べる。さらに、転移の時期を明らかにするには、近縁な生物群において、転移した形跡がないことを示す必要がある。RNA-Seq や全ゲノムシーケンスの

データがあればその解析も可能である。これらの結果を総合して、遺伝子水平転移と機能獲得のダイナミクスを明らかにする。

4. 研究成果

(1) 相同性検索用の昆虫、細菌タンパク質データベースの準備

昆虫のタンパク質データベースとして、以下の7種の昆虫、エンドウヒゲナガアブラムシ、ネッタイシマカ、セイヨウミツバチ、カイコ、キイロショウジョウバエ、ヒトジラミ、ゴミムシダマシを用いた。合計、124,206 の CDS に相当する。既存のデータの中で、アノテーションがしっかりしており、系統的に異なる種類を用いた。細菌ゲノムについては、枯草菌、バクテロイデス腸内細菌、ブラジリゾビウム根粒菌、パーコルデアリア土壌菌、大腸菌 K12、大腸菌 O157、ネンジュモ、緑膿菌、ロゼオバクター紅色細菌、ストレプトミセス放線菌、ボルバキア昆虫寄生菌、ボルバキア線虫共生菌の、合計 12 種類を用いた。合計、62,687 の CDS に相当する。代表的な細菌の中で、比較的ゲノムが大きいものや、真核生物との生態的に密接な関係があるもの、異なる系統に由来するもの、以上の基準を基に選択した。

かなり厳選したが、現状の処理速度は高速の CPU2 つ分(32 スレッド)で1分間に 400kbp のゲノムが解析できる程度であり、1G のゲノム DNA では、最短 3.5 日かかる。よって、相同性検索用のデータベースについては、相同性の高い配列は除くなどの工夫も課題である。

(2) ゲノム配列の準備

転移遺伝子を探る昆虫ゲノムとして、必須共生生物を体内に持つ生物種やその近縁種をまずは対象とした。そのような生物は、その共生細菌自体や、他の細菌からの水平転移遺伝子を保持する例が複数で知られているためである。現在は 9 種、ツエツェバ工類

の一種、トコジラミ、コナカイガラムシ類の一種、オオヨコバイ類の一種、トビイロウンカ、ナガカメムシ類の一種、キジラム類の一種、クサギカメムシ、サシガメについて、転移遺伝子の探索を行った。それぞれ、登録されているアセンブルデータのゲノムサイズは、370M、698M、285M、2.2G、1.1G、1.1G、708M、1.2G、707Mであった。細菌由来の領域を探索するために、1000塩基の断片になるように分割し、以下の相同性検索の問い合わせ配列とした。

(3) 相同性検索

細菌由来の領域を探索するため、上記の断片化した昆虫ゲノムを問い合わせ配列として、上記(1)で作製した細菌タンパク質データベースと昆虫タンパク質データベースに対して、Blastxによる相同性検索を行った。統計的な優位性の基準値であるE-valueにおいて、0.0001以下の場合を相同性があるとした。この中から相同性の度合いを示すスコアが、昆虫よりも細菌タンパク質に対して高い場合を、細菌由来のゲノム領域の候補とした。ただし、スコアが50未満のものは、多くの場合、単純反復配列が偶然類似性を示したものであるため、候補から除いた。

ツェツェバエ類の一種では、3,638の断片が細菌に対してより高いスコアを示した。ただし、そのうち3,364はある2つのscaffoldの断片であった。おそらく、この2つの断片は共生細菌か何らかの形で混入した細菌の可能性もある。また、これ以外にも、断片が細菌由来の可能性のある5のscaffoldでは、領域全体が細菌に類似しており、何らかの実験上混入したDNAである可能性が高い。これらを除くと、33領域に細菌由来の塩基配列が存在する可能性が残った。これらについて、細菌及び真核生物のアノテーションタンパク質からなるuniprotデータベースに対する検索から、昆虫ゲノム断片上に、細菌タンパ

ク質に対して高い相同性を示す塩基配列が存在する領域が、13あることがわかった。いずれも、2000塩基以下の断片的なものである。

コナカイガラムシ類の一種では、細菌にしか相同性を示さないcontigを除くと、176の領域に細菌由来の遺伝子である可能性が残った。ただし、ゲノムアセンブルの質が低いため(N50=9727)、この中にも実験上での混入配列が多く含まれていると予想される。

さらに解析と進め、昆虫ゲノム断片上に、細菌タンパク質に対して高い相同性を示す塩基配列が存在する領域が、74あることがわかった。

トビイロウンカは364の領域に細菌由来の塩基配列がある可能性が残った。より詳細な解析を行った結果、86の領域に昆虫ゲノム断片にも関わらず、細菌タンパク質に対して高い相同性を示す塩基配列が存在することがわかった。

その他も、同様な解析を行い、現在は6種、ツェツェバエ類の一種、コナカイガラムシ類の一種、トビイロウンカ、ナガカメムシ類の一種、キジラム類の一種、サシガメより、13、74、86、61、45、114の領域に昆虫ゲノム断片にも関わらず、細菌タンパク質に対して高い相同性を示す塩基配列が存在することがわかった。他の3種についても一通りの解析は終えている。

(5) ゲノムのアノテーション

以上のように、多数の細菌由来遺伝子の候補が得られた。これらが本当に細菌由来遺伝子であるのか、実験上の混入、あるいはミスアセンブルによる昆虫ゲノム配列への挿入などが起こった配列ではないかを確認する必要がある。そのためには、細菌由来遺伝子の候補を含むcontigやscaffoldに昆虫の遺伝子がコードされていることによって、まずはスクリーニングができる。そのためには、昆虫ゲノム自体の遺伝子予測とアノテーションが重要になる。しかし、現時点ではこれら

の配列の遺伝子予測は、一部の配列でなされているだけである。そして、論文上は遺伝子予測の結果が示されていても、データベースには配列のみの登録しかない場合が多く、自身で行う必要があることがわかった。

これまでの類似の解析ではゲノムの遺伝子予測済みのデータを使っていたので、この問題が生じなかった。しかし、contig や scaffold が昆虫由来であることを示すには、そのコードされた遺伝子がイントロンを含むことや、昆虫のタンパク質と相同性が高いこと、近傍の遺伝子構造などを確認することが必要である。よって、遺伝子予測の結果は重要である。

実際に遺伝子予測を行うには、mRNA の配列、EST 配列、近縁種のタンパク質配列といった遺伝子の情報が必須である。現在では RNA-Seq の配列からでも推測が可能である。しかし、これらもただ必要なデータを与えただけでは、実際にコードされている遺伝子の 1/4 程度しか予測されない。時間をかけて、予測パラメータを最適化することや、複数の方法を試す必要があり、現状では遺伝子予測のための PC 環境の整備は整ったが、発表できるレベルの予測ができていない。今後の課題である。

(6) まとめ

以上のように、昆虫ゲノムから細菌由来遺伝子を予測するシステムの構築はできた。そこから、細菌由来の転移遺伝子であることを確定し、さらにはその進化的な由来と宿主における機能の同定は現在進行中である(方法の(5),(6)の部分)。本来は RNA-Seq のデータを中心に進める予定であったが、研究計画時に予定していた配列を手に入れることができず、方針を変更した。そのため、より一般的で複数の生物種の解析ができたし、できる方法を構築できた。その分、個別の結果を深化させることができなかった。

(7) その他

本研究と合わせて、マツノマダラカミキリのゲノムに転移した、ボルバキア由来の遺伝子についての解析も進めた。また、リベリバクター植物病原菌が、その進化の過程で、キジラミの共生細菌の祖先から、水平転移によって、遺伝子を獲得したことを明らかにした。その他、データベース構築や共生に関する研究活動を行った。

5 . 主な発表論文等

[雑誌論文] (計 2 件)

Aikawa T, Nikoh N, Anbutsu H, Togashi H (2014) Prevalence of laterally transferred *Wolbachia* genes in Japanese pine sawyer, *Monochamus alternatus* (Coleoptera: Cerambycidae) *Applied Entomology and Zoology* **49**:337-346 doi:10.1007/s13355-014-0256-0 (査読有)

*Nakabachi A, *Nikoh N, Oshima K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima S, Hattori M, Fukatsu T (2013) Horizontal gene acquisition of *Liberibacter* plant pathogens from a bacteriome-confined endosymbiont of their psyllid vector. *Plos One* **8(12)**: e82612. (*Equal contributions) doi:10.1371/journal.pone.0082612 (査読有)

[学会発表] (計 5 件)

Nikoh N, Tsuchida T, Maeda T, Yamaguchi K, Shigenobu S, Takema Fukatsu (2015) Genome sequence of *Rickettsiella viridis*, a symbiotic bacterium modifying insect's body color. *Animal-Microbe Symbioses Identifying the Common Language of Host-Microbe Associations*, Gordon Research Conference (Waterville valley, USA) June 26 2015

Mochizuki T, Tanizawa Y, Fujisawa T,

Nikoh N, Toyoda A, Fujiyama A, Kurata N, Nagasaki H, Shimizu T, Kaminuma E, Nakamura Y (2015) DNA Polymorphism Database from New-Generation Sequence Read Archive and Analytical Workflow. Plant & Animal Genome XXIII 2015年1月14日

望月 孝子 , 谷澤 靖洋 , 藤澤 貴智 , 長崎 英樹 , 神沼 英里 , 清水 徳朗 , 豊田 敦 , 藤山 秋佐夫 , 倉田 のり , 二河成男 , 中村 保一 (2014) DNApod : NGS アーカイブ配列からの統合DNA多型注釈データベース 生物横断的解析への活用 第37回日本分子生物学会年会 2014年11月25日

望月孝子, 藤澤貴智, 谷沢靖洋, 長崎英樹, 神沼英里, 大柳一, 清水徳朗, 豊田敦, 藤山秋佐夫, 倉田のり, 二河成男, 中村保一 (2013) DNA多型統合データベースと解析ワークフローの構築 : 植物、微生物への取り組み 第36回日本分子生物学会年会 2013年12月5日

望月孝子 , 長崎英樹, 谷澤靖洋, 藤澤貴智, 神沼英里, 大柳一, 清水徳朗, 豊田敦, 藤山秋佐夫, 倉田のり, 二河成男, 中村保一 (2013) DNA多型統合データベースと解析ワークフローの構築 : 植物、微生物への取り組み NGS現場の会 第三回研究会 2013年9月4日

〔図書〕(計 1 件)

二河成男 (2014) 動植物と微生物との共生関係「現代生物科学」(松本忠夫、二河成男 編) 177-191 放送大学教育振興会

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

6. 研究組織
(1)研究代表者
二河 成男 (NIKOH, Naruo)
放送大学・教養学部・教授
研究者番号 : 70364916