

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 26 日現在

機関番号：11301

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24657011

研究課題名(和文) 遺伝子コピー数変異の野外集団解析による新たな疾病生態学の開拓

研究課題名(英文) The analysis of copy number variations in the field populations for disease ecology

研究代表者

河田 雅圭 (Kawata, Masakado)

東北大学・生命科学研究科・教授

研究者番号：90204734

交付決定額(研究期間全体)：(直接経費) 3,200,000円、(間接経費) 960,000円

研究成果の概要(和文)：ヤマアカガエルの抗菌ペプチドのコピー数変異のゲノムでの重複の状態を調べるため、フォスミドライブラリーを作成し、抗菌ペプチドを含むクローンの配列を十世代シーケンサーでの決定を試みた。次世代シーケンサーで解読したリードを複数のアセンブラでアセンブルした結果、約40kbpの配列のうち33kbはtemporinを含まない非コード領域で、残りの配列の中に、temporin-10jと10cを含む少なくとも3コピーの配列があることが明らかになった。今回得られたコンティグの情報では、temporin-10遺伝子の詳細のゲノム上の位置を断定することはできなかった。

研究成果の概要(英文)：Our previous study indicated that anti peptide, Temporin-1 in *Rana ornativentris* shows copy-number variations. In order to determine DNA sequences in the regions including temporin-1, we created fosmid library and chose the clone libraries including temporin-1. Then we tried to determine these DNA sequences using the next generation sequencer (Ion PGM). The results showed that among about 40kbp DNA region, those in 33kbp were non-coding regions. The rest of the sequences contained at least three copies of temporin-1, including temporin-10j and temporin-10c. This suggests that these copies of temporin might be tenderly repeated. At the present study, we could not determine complete sequences of these regions.

研究分野：生物学

科研費の分科・細目：基礎生物学、生態・環境

キーワード：分子生態 疾病生態学 保全生態学

1. 研究開始当初の背景

生態学や進化生物学において、近年の生態ゲノミクスの分野の進展により、生態的に重要な性質に関わる遺伝的変異が検出されるようになってきた。これらの研究で扱われている遺伝的変異のほとんどは、DNA の配列の違いや欠失・挿入で生じる遺伝子座内の対立遺伝子多型である。しかし、近年、ヒトの研究において、遺伝子数の個体間での変異がゲノム全体にわたって大きな割合で存在し、それらが病気やその他の機能的な違いに関係していることが明らかになってきた。しかし、生態学や分子生態学においては、野外の生物における遺伝子コピー数変異は全く考慮されてこなかった。それは、生態学的性質と関係しているコピー数多型がみつかっていないことと、多くの個体を調べてコピー数多型を検出するのは、特にゲノム配列の解読されていない生物では技術的困難であることが起因していると思われる。しかし、コピー数変異は、野外の生物にとって、集団の維持などに直結する遺伝的変異である可能性がある。

抗菌ペプチドは、幅広い生物において感染抵抗性を担う自然免疫系で機能するタンパク質である。抗菌ペプチドはアミノ酸配列によって抗菌性に違いがあり、種間・種内の抗菌ペプチドの多様性は、病原体に対する抵抗性の多様性と関係すると考えられている。ヤマアカガエル (*Rana ornativentris*) では、これまでに 16 種類の抗菌ペプチドの発現が確認され、そのうち 7 種類は Temporin1 ファミリーに分類されている。しかし、抗菌ペプチドをコードするゲノム DNA の多様性については不明であった。申請者らは、1 個体に存在する Temporin1 の多様性を調べ、その結果、ヤマアカガエルの Temporin1 ファミリーは多様な種類をコードする複数の遺伝子座から構成されており、個体間に遺伝子コピー数の多型が存在することが示唆された。これら抗菌ペプチドのコピー数変異を多数の個体を調べて集団内あるいは集団間の変異を調べることが、技術的・コスト的に容易になれば、コピー数変異と個体および集団の病気への抵抗性との関係、病気や有害遺伝子と集団の脆弱性や頑健性を明らかにする疾病生態学 (Disease Ecology) への貢献へ繋がる。

2. 研究の目的

本研究では、ヤマアカガエルの抗菌ペプチド (Temporin1) の遺伝子コピー数変異を、次世代シーケンサーを使って低コストで調べる技術を確認する。そのために、Temporin 配列がゲノム上で、どのように存在しているかを明らかにするために、Temporin 配列を含むゲノム領域の配列を解読する。配列が解読されれば、sureselect などの方法で遺伝子コピー数を同定することができる。それを応用して、個体間の遺伝子コピー数の違いが

細菌への抵抗の違いに関係しているかどうか、遺伝子コピー数変異に集団間変異があるか、集団のサイズや隔離の程度は遺伝子コピー数の変異の大きさと関係しているかどうか、などを明らかにする。

3. 研究の方法

16 種類の抗菌ペプチドのうち Temporin-10 を対象とした (図 1)。これまでに、ヤマアカガエルの Temporin1 ファミリーは多様な種類をコードする複数の遺伝子座から構成されており、個体間に遺伝子コピー数の多型が存在することが示唆されている (図 2)。temporin-10 遺伝子がゲノム中にどのように存在しているのかを明らかにするために、ヤマアカガエル (*Rana ornativentris*) 1 個体を用いて DNA を抽出し、フォスミドライブラリの作製および temporin-10 遺伝子を含むフォスミドクローンのスクリーニングをタカラバイオ株式会社に委託した。スクリーニングのプロンプトには複数タイプ存在する temporin-10 遺伝子に広く保存された領域の塩基配列を用いた。



Temporin-1 の特徴

- 10 - 14 アミノ酸残基からなる
- C末端がアミド化している



図 1. *Rana ornativentris* の分泌物から確認された 16 種類の抗菌ペプチドと Temporin-1 の構造 (Ohnuma et al. 2010).

スクリーニングの結果、temporin-10 遺伝子を含むと思われるフォスミドクローンが合計 3 つ得られ、それぞれをフォスミド A、B、C とした。得られた陽性フォスミドクローンは、断片化後、次世代シーケンサー Ion PGM によってシーケンスを行い、De-novo アセンブリによる配列決定を行った。フォスミド B

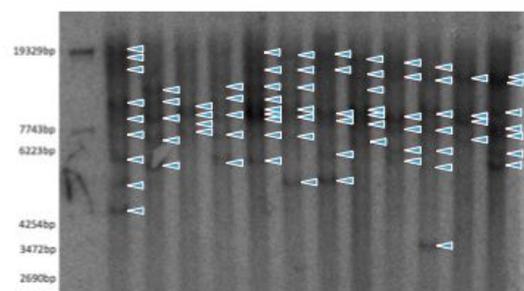


図 2. 異なる *Rana ornativentris* の個体を用いた temporin-10 のザンハイブリダイゼーションのバンド。個体によって、4-9 本のバンドが確認され、コピー数多型である可能性が示された。

については納品サンプルの DNA 濃度が低く、Ion PGM によるシーケンスを実施しなかった。フォスミド A、C の両陽性フォスミドクローンについてライブラリサイズ 200bp のシーケンスを複数回行った。しかしシーケンスにより得られた塩基配列を用いて De-novo アセンブラ SOAPdenovo によるアセンブリを行った結果、temporin-10 遺伝子を含むと思われるコンティグは数十～数百塩基と非常に短く、長いコンティグはライブラリ作製の際に混入する宿主大腸菌のゲノムの塩基配列ばかりであった。このことは目的とする temporin-10 遺伝子ならびにフォスミドベクターの配列を含む断片のカバレッジが低い、または宿主大腸菌ゲノムが予想以上に多く混入していることが原因として考えられた。しかし改善のため宿主大腸菌ゲノムの除去ならびにシーケンスの追加を行ったところ、アセンブリの結果に大きな変化は無かった。これらの状況を受け根本的なアセンブリの向上をはかり、ライブラリサイズ 400bp へ移行し、異なる複数のアセンブラを使用することにした。また陽性フォスミドクローンの適切な De-novo アセンブリ手法の確立のため、先立ってフォスミド C のシーケンスを集中的に行うことにした。

4. 研究成果

フォスミド C を Ion PGM によりアセンブラを行った。アセンブラには、複数のアセンブラの方法をもちいた(表 1)。

表1. 異なるアセンブラによるクローンCのアセンブルの結果

アルゴリズム	De Bruijn graph				OLC	
アセンブラ	SOAPdenovo	ABYSS 1.3.7	Velvet 1.2.03	SPAdes 3.0.0	Newbler (454)	Mira 4.0.2
Number of contigs	176,006	190,583	21,328	5,655	93	5,890
Total bases	15,505,797	15,302,502	2,962,865	496,992	94,188	3,593,937
Average contig size	88	80	138	87	1,012	610
N50 contig size	80	75	132	80	1,450	638
Longest contig size	1,074	1,587	853	3,579	7,255	5,634

アセンブラにより結果が大きく変動することから、適切なアセンブラの選択が配列決定に重要となることがわかった。今回用いたアセンブラはそのアルゴリズムから大きく de Bruijn graph と Overlap Layout Consensus (OLC) に分けられるが、今回のアセンブルにおいては OLC アセンブラである Newbler、Mira の方がより優れていることがわかった。また de Bruijn graph の中で SPAdes、Platanus に関しては長いコンティグを作り出し、良い結果が得られた。

アセンブルの総合に評価した結果 temporin-10j 遺伝子全長を含むコンティグ (5111 bp) を最長とする、temporin-10 遺伝子を含むと考えられる複数のコンティグが得られた。これらのコンティグには temporin-10j 遺伝子の他に temporin-10c 遺

伝子を含むコンティグもあり、このことからフォスミド C のインサート中に複数の temporin-10 遺伝子が存在していることが示唆された(図 3)。また複数回行ったすべてのアセンブルのうち最長のコンティグは約 33kbp であり、temporin-10 遺伝子を含む抗菌ペプチド遺伝子を含まない非コーディング領域であることがわかった。フォスミドベクターのインサート配列の長さが約 40kbp であることを考えると、フォスミド C におけるインサートの多くの領域は非コーディング領域で構成されている可能性がある。

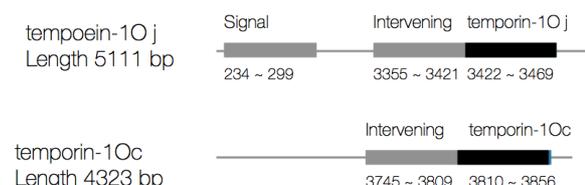


図 3. Temporin-10j, temporin-10c を含む最も長いコンティグ

今回得られたコンティグの情報では、temporin-10 遺伝子がゲノムにタンDEMに存在しているかどうかを断定することはできなかった。しかしフォスミド C におけるインサート配列の非コーディング領域を除いた、限られた領域中に temporin-10 遺伝子の異なるタイプが位置している可能性が高いことがわかった。また、temporin が存在している領域は、リピート配列が多いなど、次世代シーケンサーでの解読が困難な領域であったと推測された。

今後、異なる次世代シーケンサーを使って、ペアエンドで配列を読み、アセンブルをすることで、全配列を決定する予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 4 件)

1. Tamate, S., Kawata, M. and T. Makino (2014) Contribution of non-ohnologous duplicated genes to high habitat variability in mammals. *Molecular Biology and Evolution*, (Advanced Access Publication) 査読有 doi: 10.1093/molbev/mss133
2. Makino, T., McLysaght, A and Kawata, M. (2013). Genome-wide deserts for copy number variants in vertebrates. *Nature Commnuications* 4:2283. 査読有 doi:10.1038/ncomms3283
3. Iwasaki, W., Tsuda, M. and Kawata M.

- (2013) Genetic and environmental factors affecting cryptic variations in gene regulatory networks. *BMC Evolutionary Biology*, 13:19 査読有 doi:10.1186/1471-2148-13-91
4. Makino, T. and Kawata, M. (2012) Habitat variability correlates with duplicate content of *Drosophila* genomes. *Molecular Biology and Evolution*, 29, 3169-3179 査読有 doi:10.1093/molbev/mss133

〔学会発表〕(計 1 件)

1. Makino T. Genome-wide deserts for copy number variation in vertebrates. Plant and Animal Genomics XXII, 2014年1月10-15日, San Diego, USA

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計0件)

取得状況(計 0 件)

〔その他〕

ホームページ等

6. 研究組織

(1)研究代表者

河田 雅圭 (KAWATA, MASAKADO)

東北大学・大学院生命科学研究科・教授

研究者番号：90204734

(2)連携研究者

牧野 能士 (MAKINO, TAKASHI)

東北大学・大学院生命科学研究科・助教

研究者番号：20443442