

[研究課題名]

de novo RNA-seq 解析によるモレレットワニ (*Crocodylus moreletii*) の遺伝子配列と機能の同定

[共同研究者名 (所属を含む)]

岩田久人(愛媛大学・沿岸環境科学研究センター)

[研究目的]

モレレットワニはユカタン半島の淡水生態系における最上位捕食者の爬虫類である。本種は、個体数の減少からワシントン条約の絶滅のおそれのある種に指定されている。個体数減少の原因として、生息地域における環境汚染の可能性が考えられているが、化学物質曝露と生体影響の関係については明確な結論が出ていない。これらの関係の科学的裏付けをするにはゲノムや遺伝子配列情報などの生体情報が必要であるが、本種のゲノム配列のドラフトは存在せず、遺伝子配列情報は3遺伝子のみしか公開されていない。そこで、本研究では、モレレットワニの鱗甲軟組織から取得した試料を用いて de novo RNA-seq 解析を行い、モレレットワニの遺伝子配列情報とそれらの機能を明らかにすることを目的とした。

[研究方法]

・試料

本研究で用いた試料は DRA004128-41 に収録されているミシシッピワニ (*Alligator mississippiensis*) のサンプルである。具体的には、DRX043892 に収録されている、DRR048609_1.fastq.gz, DRR048609_2.fastq.gz を使用した。

試料採取の詳細については、Yatsu et al. (2015) に記載されているのでここでは、簡単に試料について記載する。このサンプルは、フロリダ州魚類野生生物保護委員会(Florida Fish and Wildlife Conservation Commission) と米国魚類野生生物局(the U.S. Fish and Wildlife Service) の許可 (Permit #: SPGS-10-44) のもと、2010年6月

に米国フロリダ州ボルシア郡にあるウッドラフ湖国立野生生物保護区で 5 つのクラッチから収集されたミシシッピワニの卵である。

ワニ卵は発生ステージ 19 になるまで、室温 30°C条件下で培養された。その後、ワニ胚から生殖腺-副腎中膜複合体 (GAM) を採取し、RNAlater (Ambion/Thermo Fisher Scientific, Waltham, MA, USA) に-20°Cで保存し、さらに解剖顕微鏡下で GAM 複合体から生殖腺組織を分離した。

・試料からの total mRNA 抽出

試料からの total mRNA 抽出の詳細については、Yatsu et al. (2015)に記載されている。Total mRNA は ISOGEN 試薬 (ニッポンジーン、日本、富山) を用いて個々の生殖腺から全 RNA を抽出し Promega SV Total RNA Isolation system、(Promega, Madison, WI, USA) を用いて精製された。

・シーケンス用ライブラリ調整とシーケンス条件

詳細については、Yatsu et al. (2015)に記載されている。生殖腺サンプルから得られた 500 ng の total RNA を、Illumina TruSeq RNA sample preparation v2 kit (Illumina, San Diego, CA, USA) を用いて、ライブラリ調製している。Illumina HiSeq2500 instrument (Illumina) を用いて塩基配列を決定した。シーケンスは 101 bp, ペアエンドリードとして 3 レーンに形成された。本研究では、このうちの 1 レーン分を用いた。

・使用したマシン

本研究で使用したマシンのスペックは、MacOS Monterey バージョン 12.3、プロセッサ 2.9GHz クアッドコア Intel Core i7、メモリ 16G である。アッセンブラーである Trinity を動かすため、本マシンに Parallels Desktop 17 を用いて、Ubuntu 20.04.2 LTS を仮想マシンとして作成し、使用した。

・使用したソフトウェア

シーケンスリードのクオリティチェックと低品質シーケンスのトリミング機能を持つ fastp (0.20.0)、アッセンブラーの trinity(v2.15.1)を取得した。下記にイン

ストール用コマンドを示す。アセンブリ以降の解析は、OmixBox を用いて行った。

```
sudo apt-get update

sudo apt-get install fastp

sudo apt-get install trinity
```

Apple が独自開発したチップ (M1 または M2) を搭載する Mac で仮想マシンを使用しない場合、Docker による Trinity のインストールと使用が有用である。この場合、fastp は Rosseta2 を起動した状態のターミナルから conda もしくは mamba によりインストールおよび実行が可能である。Docker を run する際は、通常のターミナルに戻って実行する。

[研究成果]

本研究では、モレレットワニのサンプルから得られた total mRNA をもとに、モレレットワニの遺伝子の同定と機能解析を行う予定であった。しかし、メキシコからサンプルが到着する時期が予定より遅れた。またサンプルから質の高い total mRNA が抽出できなかった。このため、対象種を変更しドライ解析のワークフローと環境構築を行うこととした。

ミシシッピワニ (*Alligator mississippiensis*) はモレレットワニと生息域が近く、ゲノムが解読されており (assembly ASM28112v4)、タンパク質数(74,857)、遺伝子数(30,217)、ヌクレオチド数(71,873)が公開されている。登録されているヌクレオチド数のうち、mRNA は 47,962 である。これらのことから、モレレットワニではなくミシシッピワニを用いて、ワニの de novo RNA-seq のワークフローを作成することにした。

1. フィルタリング

クオリティーチェックを行ったのち、低クオリティーだったリード及び、欠

損 (N) が多すぎるリードをフィルタリングした。Fastp(デフォルト設定)によるフィルタリングの結果、サンプルのリード数は 20,513,146 (10,256,573)リードから 19,860,928(993,464)リードとなった。

Trimomaticによるフィルタリングでは以下*の設定を使用した。この結果、19500350(9,750,175)リードとなった。一般に Trimomaticによるフィルタリングの方が、以降で得られるリードのクオリティが高い。このことから、本研究では Trimomatic によって得られたリードを使用することにした。

2.アセンブル

得られたリードを用いて、Trinityによりアセンブルをおこなった。本研究では、k_mer =1, k_mer=3 の2条件において、得られるリード数とそのクオリティを比較した(表1)。この結果、Trinityによって得られるリード数は k_mer =1 の時、k_mer =3 の時よりも多かった。一方で、CD-HITによって、配列をクラスタリングした結果、冗長性のないリードのクラスター数はほぼ同数であった。このことから、Trinityの Min.Kmer Coverageの変更は、アセンブリ以降の結果にあまり寄与しないことが考えられた。

表1 Trinityの設定の違いによる再構築リード数の変化

	Min. Kmer Coverage = 1	Min. Kmer Coverage = 3
Trinity Results		
Total transcripts	106,265	56,316
Total genes	85,732	45,475
Percent GC	46.84	46.81
Total assembled bases (all transcripts)	108,642,273	60,148,351
Total assembled bases (longest isoform per gene)	70,764,846	43,088,599
N50	2,268	2,095
BUSCO Results		
Complete Single-copy	1,982	1,848
Complete Duplicated	675	337
Fragmented	318	459
Missing	379	710
CD-HIT Results		
Cluster size =1	94262	94261

3. コーディング領域の予測と同定

Trinity (k_mer =3)から得られたアセンブル配列から、TransDecoder (v.5.5.0)を用いて、コーディング領域を予測した。助長性のないリード 94,261 配列を用いた予測の結果、38,185 の配列がコーディング領域であることが予測された。

これらの配列を用いて、blast search を実施した。この結果、29,966 配列が blast で Hit した。このうち、ミシシッピワニの配列は 21,873 配列であった。

本研究課題では、de novo RNA-seq を解析するための環境を構築した。ミシシッピワニの試料を用いて Trinity により mRNA の再構築を行った。この結果、既知の mRNA 配列の約 40%に当たる約 20,000 個の配列が再構築された。

[今後の課題]

今回の研究では、アセンブラーとして Trinity のみを用いた。一方で、de novo RNAseq で mRNA を再構築する場合、さまざまな種類のアセンブラーによって再構築されたリードを使用して、遺伝子を同定することが望ましい。こそこで、今後は Trinity 以外のアセンブラーである Velvet/Oases、SoapDenovo-Trans、でも同様にアセンブルを行い、得られた配列を統合して、mRNA の再構築を行う。さらに、本研究で構築したパイプラインを用いて、モレレットワニ (*Crocodylus moreletii*) の遺伝子配列と機能の同定を行う。モレレットワニの配列が得られない場合、より近縁種であると考えられるイリエワニ (*Crocodylus porosus*)のデータを用いて、de novo RNAseq 解析を行い、より精度の高いパイプラインの構築を目指す。

```
*ILLUMINACLIP:<fastaWithAdaptersEtc>:<seed mismatches>:<palindrome clip threshold>:<simple clip threshold>:<minAdapterLength>:<keepBothReads> = ILLUMINACLIP:TruSeq3.fasta:2:30:15:8:true, SLIDINGWINDOW:<>windowSize>:<requiredQuality> = SLIDINGWINDOW: 4:15, MINLEN:36
```