

令和 5 年 6 月 5 日現在

機関番号：15401

研究種目：挑戦的研究（萌芽）

研究期間：2021～2022

課題番号：21K19118

研究課題名（和文）データ駆動型ゲノム育種を実現するデータ解析基盤技術の開発

研究課題名（英文）Development of data analysis platform technology for data-driven genomic breeding

研究代表者

坊農 秀雅（Bono, Hidemasa）

広島大学・統合生命科学研究科（理）・特任教授

研究者番号：20364789

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：これまでの機能アノテーションは、塩基配列とそれをコードする予測アミノ酸配列の特徴から機能推定を行う手法であり、今回さらにこの配列情報ベースの機能アノテーションに加えて遺伝子発現情報から機能を推定する手法の開発を試みた。開発した手法はFanflow4Insectsと名づけられ、利用可能なRNA配列解読(RNA-Seq)データの種別はそれぞれの生物種ごとにまちまちであるものの、RNA-Seq発現定量解析から得られる発現データから組織特異的発現情報として使うことで配列情報ベースの機能アノテーションでは手がかりが全く得られなかった転写配列に関して機能アノテーションを付与することが可能となった。

研究成果の学術的意義や社会的意義

開発した機能アノテーションワークフローFanflow4Insectsは、昆虫のみならず陸上植物のアカシソにおける機能アノテーション手法として利用され、そのゲノム配列解読後にFanflow4Plantsとして得られた遺伝子配列の機能アノテーションに用いた。二次代謝の例としてアントシアニンの合成経路の再構築に応用し、ゲノム中に多コピー存在する酵素遺伝子の同定を行った。このことは開発した手法が今後、非モデル生物の機能アノテーションのみならず、ゲノム編集ターゲット遺伝子選定に寄与し、データ駆動型ゲノム育種に大きく貢献すると考えられる。

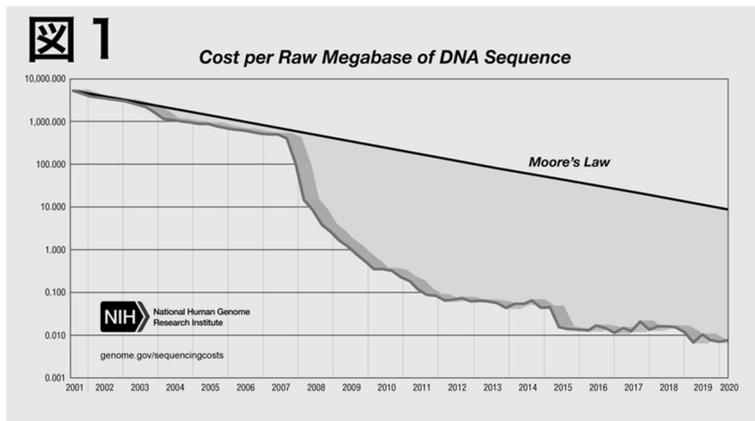
研究成果の概要（英文）：In addition to the sequence-based functional annotation, we have attempted to develop a method to infer function from gene expression information. Although available RNA sequencing (RNA-Seq) data varies among species, the developed method, named Fanflow4Insects, can be used as tissue-specific expression information from expression data obtained from RNA-Seq expression quantification analysis, thereby providing a method to infer function from sequence information-based functional annotation. The use of tissue-specific expression information from RNA-Seq expression quantitative analysis has made it possible to provide functional annotation for transcriptional sequences for which no clue could be obtained from sequence-based functional annotation.

研究分野：ゲノム情報科学

キーワード：ゲノム育種 機能アノテーション データ駆動型 バイオDX 新規モデル生物 塩基配列解析 遺伝子発現 ゲノム配列

1. 研究開始当初の背景

米国 NIH のレポートによると、塩基配列解読コストは 2007 年以降急激に安くなり、ここ 5 年ほどは百万塩基(1MB)を解読するコストはわずか\$0.01 ほどとなっている(図 1)。この大価格破壊は、百塩基長程度のリードが大量に読める技術によるものであったが、ここ数年でさらに数万から数十万塩基長の配列解読(長読み)ができる塩基配列解読技術が成熟してきた。



その結果、さまざまな生物種においてゲノム配列解読の機運が高まっている。2020年10月現在、ゲノム配列が解読されて公表されたデータが登録されているデータベース(DB)である NCBI Genome に登録されている生物の数は五千あまりで、名前が付けられているものだけでも 27 万 7 千種ある植物のうち 690 種のみ、百万種以上いるという昆虫も 544 種のみにはすぎない。特に育種のターゲットとなりうる有用物質生産生物のゲノム配列解読が近々の課題となっている。

しかしながら、ゲノム配列が解読できてもレゴに例えるとその「ブロック」を揃えたにすぎず、そのブロックをどう組み合わせるその生物が利用しているかといったことはわからない。そのゲノム配列にどのような遺伝子がコードされているかをまず知る必要がある。それに関しては転写された RNA の配列情報を解読するトランスクリプトーム解読によって遺伝子をコードする領域を測定することができる。しかしながらそれが分かっても、遺伝情報がどう生物によって使われているかはわからない。その遺伝子に対する機能を注釈付ける機能アノテーションが必要不可欠である。

しかしながら、ゲノム配列が解読できてもレゴに例えるとその「ブロック」を揃えたにすぎず、そのブロックをどう組み合わせるその生物が利用しているかといったことはわからない。そのゲノム配列にどのような遺伝子がコードされているかをまず知る必要がある。それに関しては転写された RNA の配列情報を解読するトランスクリプトーム解読によって遺伝子をコードする領域を測定することができる。しかしながらそれが分かっても、遺伝情報がどう生物によって使われているかはわからない。その遺伝子に対する機能を注釈付ける機能アノテーションが必要不可欠である。

2. 研究の目的

これまで機能アノテーションに関しては誰かがやらないといけないにもかかわらず、その解析手法は驚くほど開発されてきていない。基本的には、DB に登録されたこれまで知られている遺伝子配列との配列類似性に頼ってその遺伝子の機能アノテーションはなされているが、それ以上の手法開発はなされてこなかった。

そこで、本研究ではデータ駆動型ゲノム育種を実現するデータ解析基盤技術に必要なこれまでの配列類似性のみならず、遺伝子発現データを利用した機能アノテーション手法の開発を試みる。開発した手法はいずれ応用され、ゲノム配列解読がなされた後の遺伝子機能アノテーションの標準となるよう、技術移転されやすい形での実装を試みる。

3. 研究の方法

I. 発現データから機能アノテーションする手法の開発

まず、これまで FANTOM 共同研究において 20 年かけて確立してきたヒトやマウスなどの哺乳類の機能アノテーション手法を、ゲノム育種のターゲットとなる植物や昆虫で利用可能となるように汎用化を行う。この機能アノテーションは、塩基配列とそれのコードするアミノ酸配列の特徴から機能推定を行うもので、これまで長らく使われてきた手法である。

さらに、配列情報ベースの機能アノテーションに加えて遺伝子発現情報から機能を推定する手法の開発を試みる。RNA 配列解読(RNA-Seq)から得られる遺伝子発現データも情報として機能アノテーションに利用する手法を開発する。かつての EST(Expressed Sequence Tag)解析においてその配列をクローニングした cDNA ライブラリ情報として保持し、遺伝子機能の推定に利用していた。例えば、複数の組織で発現解析したデータを見て 'brain specific gene' と結論づけるなどがそれである。RNA-Seq データ解析から得られるそれらの大量の情報から機能情報としてアノテーションする手法の開発を試みる。

II. 開発手法を実データに適用

開発した新規機能アノテーション手法を用いて実際に機能アノテーションを行う。これまでに遺伝子配列解読のなされていない生物種、特に本研究課題においては昆虫をターゲットにする。公共 DB 中のデータは遺伝子発現データがないことも多いので、そのために独自に RNA 配列解読を行う。従来の Illumina 社シーケンサーによるショートリードを de novo アセンブルする手法(短い配列断片をつなぎ合わせる方法)に加えて、ロングリード(長読み)シーケンサーを用いた RNA 配列解読で配列解読を行う手法を検討する。

4. 研究成果

理化学研究所での FANTOM 共同研究において 20 年以上かけて確立してきた哺乳類(ヒトやマウス)の機能アノテーション手法を、ゲノム育種のターゲットとなる植物や昆虫において利用可能となるよう、汎用化を試みた。そのために必要な改良に関して研究代表者と研究分担者でオンラインで議論を重ねた。

これまでの機能アノテーションは、塩基配列とそれをコードする予測アミノ酸配列の特徴から機能推定を行う手法であり、今回さらにこの配列情報ベースの機能アノテーションに加えて遺伝子発現情報から機能を推定する手法の開発を試みた。RNA 配列解読(RNA-Seq)から得られる遺伝子発現データも情報として機能アノテーションに利用する手法の開発を行った。

並行して、ロングリード(長読み)シーケンサーを用いた RNA 配列解読で配列解読を行う手法の検討を行ったが、用いた Oxford Nanopore Technologies 社のシーケンサーでは当時の技術では十分な品質とリード数が得られず、それ以上の利用行わなかった。

かつての EST(Expressed Sequence Tag) 解析においてその配列をクローニングした cDNA ライブラリの由来組織などのメタ情報を遺伝子機能の推定に利用していた(例えば、複数の組織において発現解析したデータセットから脳だけに発現が見られる場合に 'brain specific gene' と結論づける)が、それを RNA-Seq データ解析によって得られる大量の情報からコンピュータによって人手を介さず機能アノテーションする手法の開発を試みた。

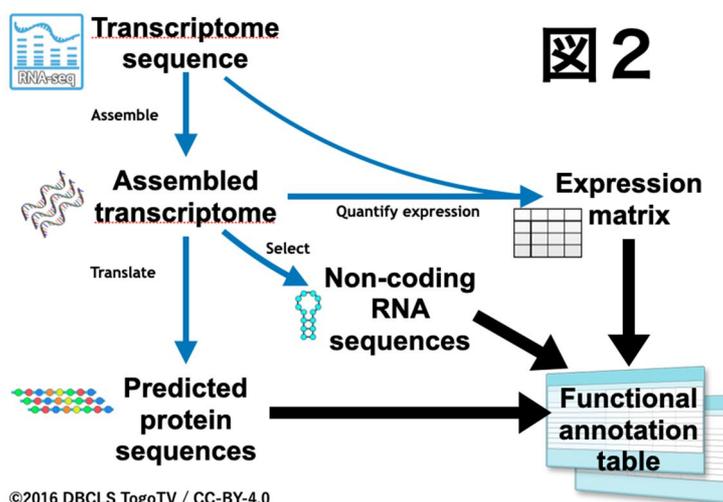
利用可能な RNA-Seq データの種類はそれぞれの生物種ごとにまちまちであるものの、RNA-Seq 発現定量解析から得られる発現データから組織特異的発現情報として使うことで配列情報ベースの機能アノテーションでは手がかりが全く得られなかった転写配列に関して機能アノテーションを付与することが可能となった。本研究課題で開発してきた機能アノテーション手法を 'Systematic Functional Annotation Workflow for Insects' と題した論文としてまとめ、査読を経て公開した

(<https://doi.org/10.3390/insects13070586>)。開発した手法は Fanflow4Insects と名づけられ、カイコとアマミナナフシに関して機能アノテーションした実例を発表した。

また、Fanflow4Insects のうち、汎用性の高い配列情報(予測アミノ酸配列と非コード RNA 配列)からの機能アノテーションに関して、実行可能なコードを GitHub より公開した(<http://github.com/bonohu/SAQE/>)。

リファレンスとする生物種を検討することにより、セイヨウミツバチとアワヨトウに対して Fanflow4Insects を用いた機能アノテーションを行った。セイヨウミツバチにおいては公共データベースで利用可能な全ての RNA シーケンスデータからリファレンス遺伝子セットを作成する研究(<https://doi.org/10.3390/insects13100931>)、アワヨトウに関してはそのゲノム配列を決定する研究(<https://doi.org/10.3390/insects13121172>)において Fanflow4Insects の結果が活用された。

さらに昆虫のみならず、陸上植物のアカシソにおける機能アノテーション手法として検討し、アカシソのゲノム配列解読後に Fanflow4Plants として得られた遺伝子配列の機能アノテーションに用いた。二次代謝の例としてアントシアニンの合成経路の再構築に応用し、ゲノム中に多コピー存在する酵素遺伝子の同定を行った。この情報は、ゲノム編集ターゲット遺伝子選定の際に重要な参考情報となり、データ駆動型ゲノム育種に寄与するものと考えられる。



©2016 DBCLS TogoTV / CC-BY-4.0

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 7件/うち国際共著 0件/うちオープンアクセス 8件）

1. 著者名 Suzuki Takayuki, Ono Yoko, Bono Hidemasa	4. 巻 9
2. 論文標題 Comparison of Oxidative and Hypoxic Stress Responsive Genes from Meta-Analysis of Public Transcriptomes	5. 発行年 2021年
3. 雑誌名 Biomedicines	6. 最初と最後の頁 1830
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/biomedicines9121830	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Tamura Keita, Sakamoto Mika, Tanizawa Yasuhiro, Mochizuki Takako, Matsushita Shuji, Kato Yoshihiro, Ishikawa Takeshi, Okuhara Keisuke, Nakamura Yasukazu, Bono Hidemasa	4. 巻 30
2. 論文標題 A highly contiguous genome assembly of red perilla (<i>Perilla frutescens</i>) domesticated in Japan	5. 発行年 2022年
3. 雑誌名 DNA Research	6. 最初と最後の頁 dsac044
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/dnares/dsac044	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Ono Yoko, Bono Hidemasa	4. 巻 6
2. 論文標題 Exploratory meta-analysis of hypoxic transcriptomes using a precise transcript reference sequence set	5. 発行年 2022年
3. 雑誌名 Life Science Alliance	6. 最初と最後の頁 e202201518
掲載論文のDOI (デジタルオブジェクト識別子) 10.26508/lisa.202201518	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Yokoi Kakeru, Wakamiya Takeshi, Bono Hidemasa	4. 巻 13
2. 論文標題 Meta-Analysis of the Public RNA-Seq Data of the Western Honeybee <i>Apis mellifera</i> to Construct Reference Transcriptome Data	5. 発行年 2022年
3. 雑誌名 Insects	6. 最初と最後の頁 931
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/insects13100931	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yokoi Kakeru, Furukawa Seiichi, Zhou Rui, Jouraku Akiya, Bono Hidemasa	4. 巻 13
2. 論文標題 Reference Genome Sequences of the Oriental Armyworm, <i>Mythimna separata</i> (Lepidoptera: Noctuidae)	5. 発行年 2022年
3. 雑誌名 Insects	6. 最初と最後の頁 1172
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/insects13121172	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Suzuki Takayuki, Bono Hidemasa	4. 巻 5
2. 論文標題 GEM: Genome Editing Meta-database, a dataset of genome editing related metadata systematically extracted from PubMed literatures	5. 発行年 2023年
3. 雑誌名 Gene and Genome Editing	6. 最初と最後の頁 100024
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ggedit.2022.100024	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Bono Hidemasa, Sakamoto Takuma, Kasukawa Takeya, Tabunoki Hiroko	4. 巻 13
2. 論文標題 Systematic Functional Annotation Workflow for Insects	5. 発行年 2022年
3. 雑誌名 Insects	6. 最初と最後の頁 586
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/insects13070586	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計7件 (うち招待講演 1件 / うち国際学会 2件)

1. 発表者名 鈴木貴之, 坊農秀雅
2. 発表標題 ゲノム編集ターゲット遺伝子のデータベースの構築とその解析
3. 学会等名 日本ゲノム編集学会第7回大会
4. 発表年 2022年

1. 発表者名 鈴木貴之, 坊農秀雅
2. 発表標題 ゲノム編集メタデータベースの開発 -文献からの情報抽出-
3. 学会等名 トーゴの日シンポジウム2022
4. 発表年 2022年

1. 発表者名 鈴木貴之, 坊農秀雅
2. 発表標題 ゲノム編集メタデータベースの構築 -文献情報からの知識抽出-
3. 学会等名 第45回日本分子生物学会年会
4. 発表年 2022年

1. 発表者名 坊農秀雅, 坂本卓磨, 粕川雄也, 天竺桂弘子
2. 発表標題 ゲノム編集のための昆虫遺伝子機能アノテーションワークフロー
3. 学会等名 第45回日本分子生物学会年会
4. 発表年 2022年

1. 発表者名 Takayuki Suzuki, Hidemasa Bono
2. 発表標題 Genome Editing Meta-Database By Computationally Automated Knowledge Extraction from Literatures
3. 学会等名 Plant & Animal Genome Conference (PAG30) (国際学会)
4. 発表年 2023年

1. 発表者名 Hidemasa Bono, Kakeru Yokoi
2. 発表標題 Meta-Analysis of Public Apis Genomes and Transcriptomes
3. 学会等名 Plant & Animal Genome Conference (PAG30) (国際学会)
4. 発表年 2023年

1. 発表者名 坊農秀雅
2. 発表標題 バイオDX産学共創拠点におけるデータ解析基盤技術開発
3. 学会等名 第67回日本応用動物昆虫学会大会 (招待講演)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>【研究成果】メタ解析によって酸化ストレスと低酸素刺激の共通因子を同定 https://www.hiroshima-u.ac.jp/news/68057 広島大学の最先端研究「ゲノム編集」と「バイオインフォマティクス」 https://tgo.hiroshima-u.ac.jp/tgo-seminar-20220326/ 最新研究紹介 https://doi.org/10.7875/togotv.2022.026 ゲノム編集のための昆虫遺伝子機能アノテーションワークフローを開発-バイオDXによる昆虫機能利用に道- https://www.hiroshima-u.ac.jp/news/71959 精密な転写産物参照配列セットを用いた低酸素応答性の評価-公共遺伝子発現データのDX- https://www.hiroshima-u.ac.jp/news/73471 公共データベースからのミツバチ参照遺伝子セットの構築-農畜産物のゲノム編集に向けたBioDX https://www.hiroshima-u.ac.jp/news/74153 赤シソの高精度なゲノム配列情報を決定-デジタル育種に向けた基盤情報を取得 https://www.hiroshima-u.ac.jp/news/74543</p>

6. 研究組織			
	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	粕川 雄也 (Kasukawa Takeya) (10304031)	国立研究開発法人理化学研究所・生命医学研究センター・ チームリーダー (82401)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------