

【本件リリース先】

文部科学記者会、科学記者会、広島大学関係  
報道機関、府中市政記者クラブ

広島大学広報室

〒739-8511 東広島市鏡山 1-3-2

TEL : 082-424-3749 FAX : 082-424-6040

E-mail: koho@office.hiroshima-u.ac.jp



広島大学



理化学研究所



国立大学法人

東京農工大学

NEWS RELEASE

令和 4 年 7 月 7 日

## 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～

### 論文掲載

#### 【本研究成果のポイント】

- 公共データベースを利用した、昆虫の遺伝子機能アノテーション (1) のためのコンピュータプログラムによる一連のデータ解析の流れ(ワークフロー(2))を開発しました。
- 開発したワークフロー「Fanflow4Insects」は、従来のタンパク質配列からの機能アノテーションだけでなく、遺伝子発現情報も利用することによって、より情報量の多い機能アノテーションが可能になりました。
- 「Fanflow4Insects」によって得られる機能アノテーション情報は、ゲノム編集を用いた昆虫を利用した研究の可能性を大きく広げることが期待されます。

#### 【概要】

広島大学大学院統合生命科学研究科の坊農秀雅特任教授は、理化学研究所生命医科学研究センターの粕川雄也チームリーダー、東京農工大学グローバルイノベーション研究院の坂本卓磨特任助教、天竺桂弘子教授と共同で、昆虫の遺伝子機能アノテーションのワークフロー（コンピュータによる一連のデータ解析の流れ）「Fanflow4Insects」を開発しました。

原核生物ゲノムアノテーション（米国 NIH の NCBI Prokaryotic Genome Annotation Pipeline）やヒトリシーケンスデータ解析（国立遺伝学研究所日本 DNA データバンク (DDBJ) など生命科学塩基配列データ解析において、それぞれのデータ解析に応じたワークフローが広く使われています。

食品ロスからコオロギなどを育て新たなタンパク質を生み出す昆虫食が昨今話題になるなど、世界的な食糧難や有用物質生産のためのプラットフォームとして昆虫が注目されています。しかしながら、昆虫の持つ遺伝子の機能は研究の進んでいるヒトやマウスと比べて未知のものが多く、今後の研究が待たれているところです。

昆虫遺伝子研究を加速するため、本研究グループは 2010 年ごろからカイコガの研究のためにその機能アノテーションパイプライン (3) を構築し、全ての遺伝子に対してその機能を予測してアノテーションする技術を開発してきました。今回、カイコガ以外の多くの昆虫種でも利用可能となるように、タンパク質をコードした遺伝子だけでなく、非コード遺伝子も対象とし、その遺伝子発現情報をも機能アノテーションに利用する方法を取り込むことによって、昆虫に特化した機能アノテーションワークフローとして「Fanflow4Insects」を開発しました。

今回開発した機能アノテーションワークフローによって、昆虫での有用物質生産をデザインするためのゲノム編集研究が加速することが期待されます。

本研究は、スイスの出版社 Multidisciplinary Digital Publishing Institute (MDPI) の *Insects* 誌に 2022 年 6 月 27 日に掲載されました。

< 発表論文 >

著者

Hidemasa Bono\*, Takuma Sakamoto, Takeya Kasukawa, Hiroko Tabunoki

\* Corresponding author (責任著者)

論文題目

Systematic functional annotation workflow for insects

掲載雑誌

*Insects*

DOI: <https://doi.org/10.3390/insects13070586>

## 【背景】

ヒトやマウスではそのゲノム解読や理化学研究所が中心となって進めてきた FANTOM 国際共同研究などによって遺伝子機能の研究が進んだ結果、それらのデータベースが整備され、それらを統合したデータ解析が可能となっています。しかし、非モデル生物である昆虫のゲノムやトランスクリプトームの塩基配列解読することが可能になったものの、この技術の利用はしばしば標的遺伝子や関連遺伝子の塩基配列の取得に限定され、他の利用可能な配列が十分に注釈されていないために、取得した配列の多くが未利用のままとなっています。

ゲノム編集技術によって、遺伝子の機能解析のための昆虫研究を加速させることが可能ですが、そのためにはゲノムにコードされたすべての遺伝子の機能を研究する必要があります。生物のゲノム解読が完了しても、その生物が持つすべての遺伝子の機能情報についての知識はまだ不完全なままなのが現状で、これまでの生物学的研究の成果から得られた遺伝子に対してのみ既存のデータベースに記録されているにすぎません。しかしながら、これらの情報はゲノム編集の標的遺伝子を決定する上で重要となってきています。これらの情報は、今後大量情報処理のためにコンピュータプログラムによって解析されることになるので、人間が読んだらわかる形式ではなく、コンピュータによって読み取り可能な形式であることが非常に重要となっています。

## 【研究成果の内容】

本研究では、昆虫の転写配列情報およびリファレンスゲノム配列・タンパク質配列データベースを利用した、基づく遺伝子機能アノテーションのためのワークフロー（コンピュータプログラムによる一連のデータ解析の流れ）を開発し、従来のタンパク質配列に翻訳できた遺伝子のみを対象とする方法を大幅に改善しました。また、RNA 配列解読によって得られる配列情報を発現量データとしても利用することにより、得られた配列の頻度情報から発現値データの特徴から、機能アノテーションする手法も開発しました。

開発したワークフローでは、トランスクリプトーム解析で得られたタンパク質配列だけでなく、同時に得られた非コード RNA (4) 配列もアノテーションします。さらに、トランスクリプトーム解析で得られた発現レベルの情報を統合し、機能アノテーション情報として利用します。このワークフローを用いて、アマミナナフシとカイコガのトランスクリプトームから得られた配列に対して機能アノテーションを行ったところ、従来の研究よりも豊富な機能アノテーション情報を得ることができました。

開発したプログラム群はオープンソースソフトウェア (5) として GitHub 上で公開されています (<https://github.com/bonohu/SAQE>)。

## 【今後の展開】

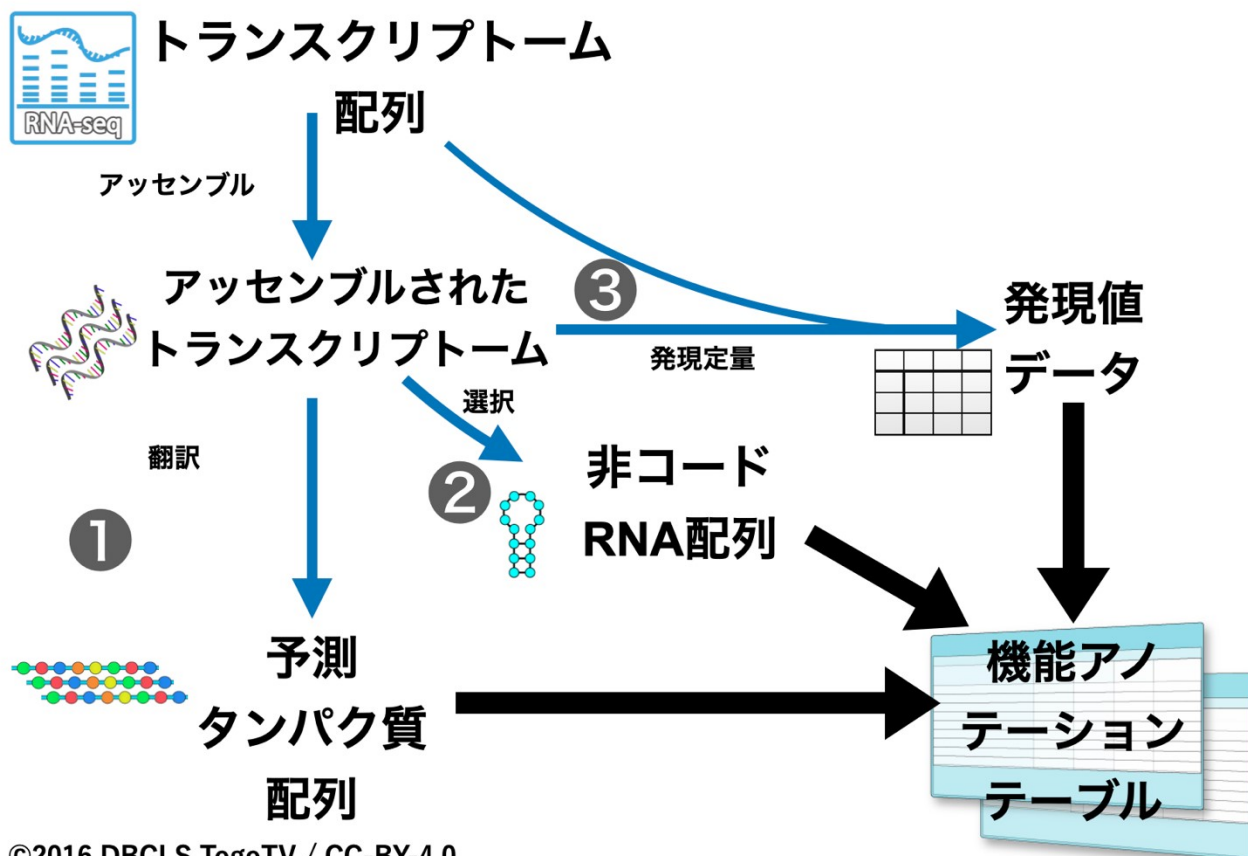
本ワークフローで得られる機能アノテーション情報はゲノム編集を用いた昆虫学研究の可能性を大きく広げるものです。また、作成したワークフローの植物や菌類への

対応を今後検討する予定です。

### 【参考資料】

#### 図 1 作成した遺伝子機能アノテーションワークフロー Fanflow4Insects

解読したトランスクリプトーム配列をアッセンブル (6) したのちに、①タンパク質配列に翻訳、配列類似性検索とタンパク質ドメイン (7) 検索によって既知のタンパク質配列の機能情報から、②タンパク質配列に翻訳できない非コード RNA 配列に対して塩基配列レベルでの配列類似性検索と RNA ドメイン検索から、③得られた配列の頻度情報から発現値データの特徴から、機能アノテーションするワークフローを構築。



### 【研究資金】

本研究は、日本学術振興会が助成する科学研究費挑戦的研究（萌芽）「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」（21K19118）、科学技術振興機構（JST）が助成する共創の場形成支援プログラム（COI-NEXT）「Bio-Digital Transformation（バイオ DX）産学共創拠点」（JPMJPF2010）および情報・システム研究機構が助成する ROIS-DS-JOINT「昆虫由来化合物の新奇な構造を創る代謝酵素 CYP の同定」（004RP2017）によって実施されました。

### 【語句説明】

- (1) アノテーション：「注釈づけ」のこと。「機能アノテーション」とは、その遺伝子がどのような機能を持つか、注釈をつけることを意味する。
- (2) ワークフロー：一連のデータ解析の流れのこと。ワークフローのさまざまな箇所で行われるツールなどをまとめたものが「パイプライン」である。
- (3) パイプライン：元々は、石油・天然ガスなどを目的地まで輸送するために設置される管路の意味であるが、それと同様にデータを入力したら結果が出力されるプログラム群とその組み合わせのことを意味する。

- (4) 非コード RNA：ノンコーディング RNA と呼ばれ、タンパク質へと翻訳される伝令 RNA (messenger RNA, mRNA) を除く全ての RNA の総称。その機能についてはわかっていないものが多く、今後の研究が待たれる。
- (5) オープンソースソフトウェア：利用者の目的を問わずソースコードを使用、調査、再利用、修正、拡張、再配布が可能なソフトウェアの総称。無償で利用が可能となっている。
- (6) アッセンブル：一般に DNA や RNA は長く、一度に端から端まで配列解読することができない。そこでその配列の一部を多数解読し、それらの断片配列をコンピュータ上でつなげて対象塩基配列を復元する。これをアッセンブルと呼ぶ。
- (7) タンパク質ドメイン：タンパク質の配列、構造の一部で他の部分とは独立に進化し、機能を持ったパーツのこと。

#### 【お問い合わせ先】

<研究に関すること>

広島大学大学院統合生命科学研究科 特任教授 坊農秀雅

Tel：082-424-4013

E-mail：[bonohu@hiroshima-u.ac.jp](mailto:bonohu@hiroshima-u.ac.jp)

理化学研究所

生命医科学研究センター チームリーダー 粕川雄也

Tel：045-503-9604

E-mail：[takeya.kasukawa@riken.jp](mailto:takeya.kasukawa@riken.jp)

東京農工大学

農学研究院・生物生産科学部門 教授 天竺桂弘子

Tel：042-367-5613

E-mail：[h\\_tabuno@cc.tuat.ac.jp](mailto:h_tabuno@cc.tuat.ac.jp)

<広報・報道に関すること>

広島大学広報室

Tel：082-424-3749

E-mail：[koho@office.hiroshima-u.ac.jp](mailto:koho@office.hiroshima-u.ac.jp)

理化学研究所

E-mail：[ex-press@riken.jp](mailto:ex-press@riken.jp)

東京農工大学総務・経営企画部企画課広報係

Tel：042-367-5930

E-mail：[koho2@cc.tuat.ac.jp](mailto:koho2@cc.tuat.ac.jp)

発信枚数：A4版 4枚（本票含む）