

特集：カイコから害虫ゲノムへの展開

カイコゲノム情報統合データベース KAIKObase

—完全長 cDNA 配列に基づくゲノムアノテーション—

(独)農業生物資源研究所 ^{すえ}末 ^{つぐ}次 ^{よし}克 ^{なか}行

はじめに

コナガ (*Plutella xylostella*) やハスモンヨトウ (*Spodoptera litura*) 等が属する鱗翅目昆虫は農業害虫の中で最も重要なグループである。カイコ (*Bombyx mori*) は鱗翅目昆虫のモデル生物であり、鱗翅目昆虫を標的とした殺虫剤などの新規害虫防除技術開発にはゲノム情報が必要であることからカイコのゲノム解析がこれまで進められてきた。カイコゲノム解析の近年の大きな成果の一つとしては、2008 年に日本 (Mita et al., 2004) と中国 (Xia et al., 2004) の WGS シークエンスの統合によるゲノムシークエン스가公開されたことが挙げられる (The International Silkworm Genome Consortium, 2008)。アセンブリデータは 432 Mb であり、ゲノムサイズ (475 Mb) のおよそ 91% に相当し、アセンブリの評価指標の一つであるスキファールド N50 サイズ*は 3.7 Mb であり、カイコにおいても高精度なゲノムシークエン스가利用可能となった。今後は、カイコのゲノムシークエン스에何を書き込まれているかを明らかにしていくことが課題となり、カイコにおいてもポストゲノム研究の時代となったと言える。我々の研究グループでは、ポストゲノム研究において重要なタスクとなるゲノムのアノテーション (genome annotation: ゲノムの注釈付け) に取り組んでいる。現在、高品質なアノテーション情報を得るために必要な完全長 cDNA ライブラリの作成と解読、効率的なマニュアルアノテーションを実現するソフトウェア基盤構築を進めている。それに伴い、我々のグループで開発・公開しているカイコゲノム情報統合データベース (KAIKObase (<http://sgp.dna.affrc.go.jp/KAIKObase/>)) の改良も行っているところである。

そこで本稿では、今後進められるカイコゲノムのアノテーションの概要と、カイコゲノム統合データベースである KAIKObase の開発の方向性について述べていきたい。

I ゲノムアノテーションとは

ゲノムアノテーションとは、ゲノムシークエン스에生物学的な注釈を付与することである。ゲノムアノテーションは、構造アノテーション (structural annotation) と、機能アノテーション (functional annotation) に大別することができる。構造アノテーションでは、ゲノムシークエン스中のどの部分が遺伝子をコードしているのかなどを記述する。一方、機能アノテーションにおいては、その遺伝子が新規の遺伝子なのか、または既知の遺伝子とどのような類似性を有するのか、どのような性質の遺伝子なのか等といったことを記述していく。多くの場合、構造、機能アノテーションともコンピュータによる自動アノテーションが行われる。構造アノテーションを行う際には、cDNA といったトランスクリプトとゲノムシークエン스とをアラインさせて行う方法や、*ab initio* 法 (統計情報に基づく予測)、すなわち、既知の遺伝子構造から“遺伝子らしさ”のモデルを作成し、そのモデルを基に予測する方法が採られる。一方、機能アノテーションは BLAST のような相同性検索や InterProScan のようなドメイン検索といった解析を基に行われる。コンピュータによるゲノムアノテーションは高速であり、ゲノムの全体像を捉えるのには適しているが精度はそれほど高いものではない。より精度の高いゲノムアノテーション情報とするには、専門家によるマニュアルアノテーションが必要となってくる。さらに、より完全な形の cDNA である完全長 cDNA (full-length cDNA) も重要なリソースとなる。完全長 cDNA は、RNA の全長を含んだ cDNA のことである。EST (expressed sequence tag) のような断片的な cDNA とは異なり、完全長 cDNA はタンパク質を合成するうえでの全情報を有しているために、転写開始点やプロモーター領域の解析等にとって非常に重要な情報源となる。

KAIKObase : an Integrated Silkworm Genome Database and Data Mining Tool. By Yoshitaka SUETSUGU

(キーワード: ゲノムアノテーション, *Bombyx mori*, データベース)

* 50% 以上の塩基が N50 値以上の長さのスキファールドに含まれることを意味している。

II カイコゲノムアノテーション

1 カイコゲノムアノテーションの進め方

これまでに多くの生物種においてゲノムアノテーションが行われてきたが、マニュアルゲノムアノテーションの進め方として、①少数の専門家（キュレーター）によりアノテーションを行う **museum model**、②アノテーションワークショップを開催し、集まった主要な研究者らが共同作業を行い、アノテーションを終えてしまうやり方である **jamboree model**、③世界中に分散する各研究グループから募られたアノテーターより構成する **cottage industry model** 等のモデルに分類することができる（STEIN, 2001）。

我々は、このうちの3番目のモデル、すなわち、世界中に分散した研究グループから構成されるアノテーションコミュニティによりゲノムアノテーションを進めていくことにした。より具体的に言えば、世界中のカイコを中心とした昆虫領域の研究者が、各自の研究室よりネットワークを介してアノテーション作業を行っていくものである（図-1）。このような分散したコミュニティによるゲノムアノテーションの一般的な利点としては、①各分野の専門家の手による機能アノテーションが行われる

こと、②研究者が配列情報や発現を確かめるために実験を行うことが可能なこと ③多くの予測遺伝子モデルに対して視覚的にチェックを行い、仮の機能を割り当てることが比較的短期間にできること等が挙げられる。反対に、同じ対象物に対し複数のアノテーターからアノテーションが重複して行われたりして効率が悪くなる可能性がある。また、アノテーションの際の判断基準を曖昧にしてしまうと、アノテーターの主観に基づいたアノテーションが行われてしまう危険性も生じる。しかしながら、これらの危険性については、アノテーション対象に関する取り決めを事前に十分に行い、また、アノテーションに関する詳細なガイドラインを作成することによって回避することができると思われる。カイコゲノムアノテーションにおけるアノテーターには、日本、中国、フランス、アメリカ等10か国、118名（2010年12月10日現在）が参加表明をしており、国際コンソーシアム形式で進めていくことが予定されている。

first version となる今回のカイコゲノムアノテーションでは、**Locus** と呼ばれる単位毎にアノテーション情報を付加する。**Locus** とはこれまでに蓄積されてきた各種のカイコ配列情報（完全長cDNA、EST、公的データベースに登録済みの **complete CDS** 配列、予測遺伝子）を

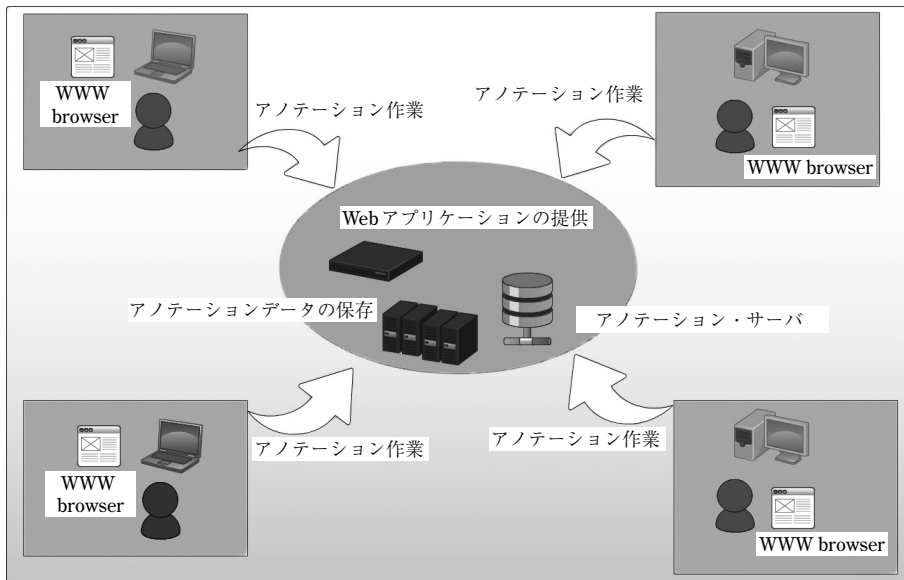


図-1 アノテーション・システムの概略

世界に分散しているアノテーション参加者は、アノテーション・サーバに接続してアノテーション作業を行う。アノテーション・システムはアノテーション情報の閲覧および編集用のWebアプリケーションを提供し、ユーザーはWebブラウザを介して利用する。各アノテーターによって編集された内容はリレーショナル・データベース管理システム（RDBMS）に保存される。

コンピュータで解析し、ゲノム上の位置を基にグループ化して作成されたものである。今回アノテーション対象となるのは少なくとも EST をメンバー配列としてもつ Locus であり、数としては 17,000 程度となる予定である。今回のカイコゲノムアノテーションでは、これらの Locus すべてに専用に開発されたソフトウェアを使って、予想される機能をアサインしていくことが主な作業となる予定である。

2 アノテーション支援ソフトウェアの開発

既に述べたが、カイコゲノムアノテーションは世界に分散したグループによりネットワークを介して進められる。分散型コミュニティアノテーションを成功させるためには、アノテーターのモチベーション維持のための仕組み、アノテーション作業を支援する優れたソフトウェアやマニュアル類、アノテーション作業に関するガイドライン作成等が重要となる (MAZUMDER, 2010)。アノテーターは昆虫学研究者より構成され、バイオインフォ

図-2 Silkworm Genome Annotation Editor のスクリーンショット

(左上) トップページ。(右上) 検索結果の表示。指定した染色体番号上にマップされている Locus がテーブル上に表示されている。(左下) Locus メンバー配列の詳細情報。Locus を構成する各配列のマップ位置、機能アノテーション情報、配列情報等を閲覧できる。(右下) アノテーション情報編集画面。

マティストは極僅かである。そのため、コンピュータの操作が不慣れなユーザーが多いことが特徴として挙げられる。コミュニティの特徴にマッチしたアノテーション作業支援ソフトウェアの有無が重要であることから、我々は専用ソフトウェア (Silkworm Genome Annotation Editor) を開発している。開発において考慮したのは、① Web アプリケーションとすること、②直感的な UI (ユーザーインターフェース) になるようにすることである。Web アプリケーションとは、Web (ウェブ) 技術を使って作成されたアプリケーションソフトウェアの総称で、一般的にはクライアントソフトウェアである WWW ブラウザを使って利用する形態のものである。Web アプリケーションのメリットは、Web アプリケーションとすることで、アノテーション・システムを利用するうえで必要となるものが Web ブラウザのみとなることである。すなわち、コンピュータの知識があまりなくてもアノテーション作業を行うための環境を整備しやすい。また、Web アプリケーションとすることで、アノテーション・システムのバージョンアップ度に毎回インストーラーなどを作成してユーザーに配布する必要がないことも利点である。これにより、開発者・ユーザー側双方にとって負担が大幅に軽減される。また、ユーザーインターフェース (UI) をシンプルで直感的なものにすることにより、ユーザーは次にどの操作を行うべきかを毎回ドキュメント類なしにその場で判断することができるため、効率的にアノテーション作業、ひいてはアノテーターのモチベーション維持にも役立つ。我々は、既存の JavaScript ライブラリ (Yahoo! User Interface Library, jQuery) を積極的に利用することで、これら要求に見合ったソフトウェアを比較的短期間に作成できた。

図-1 は開発したカイコゲノムアノテーションシステムの概念図である。本システムでは、各アノテーターは研究室より Web ブラウザ経由でアノテーション・サーバに接続してアノテーション作業を進めることができる。アノテーターにより編集された内容はアノテーション・サーバ上のリレーショナル・データベース管理システム (RDBMS) に保存・管理されるようになっている。

3 KAIKObase との連携

既に述べたが、当研究グループではカイコゲノム情報統合データベース KAIKObase (SHIMOMURA et al., 2009) を開発・公開しており、年々機能拡充を進めてきている。今回行われるカイコゲノムアノテーションに合わせて、新たに KAIKObase version 3 をリリースする予定である。Silkworm Genome Annotation Editor は、アノテーションを行ううえで必要となる様々な情報へアクセス

できるように外部データベース (UniProtKB, NCBI など) へのリンクが張られているが、その一つとして KAIKObase 上のゲノムブラウザ (GBrowse) や GeneViewer へも直接アクセスできるように作成されている。また、KAIKObase version3 では、Silkworm Genome Annotation Editor の仕様に合わせて、完全長 cDNA や Locus 情報ページへアクセスできるようトラックの追加などが行われている。

また、カイコゲノムアノテーションにより蓄積された高品質なアノテーション情報を KAIKObase に反映するための仕組みも構築しており、カイコゲノムアノテーシ

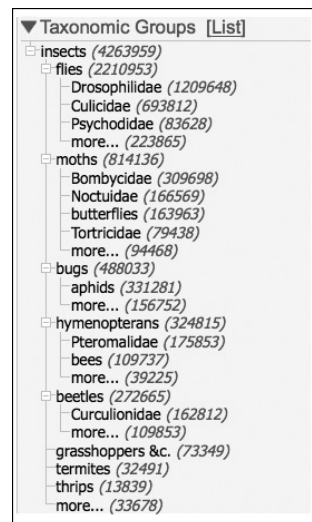


図-3 公的データベース (NCBI) に登録されている昆虫ゲノム情報 (EST) (2010/08/07 現在)
http://www.ncbi.nlm.nih.gov/nucest より引用。

ョンシステムと連携して最新のアノテーション情報を迅速に提供できるよう、KAIKObase 側の改良も進められている。

お わ り に

最後に、KAIKObase の今後の方向性について触れた。図-3 は、公的データベース上に登録されている昆虫の EST の数をまとめたものである。これまでに公的データベースに蓄積されてきた昆虫のゲノム情報は膨大な量となっている。今後次世代型 DNA シークエンサーのオミックス解析などへの利用拡大により、蓄積量は急速に増加すると思われる。KAIKObase を害虫防除技術開発にとって有効なツールとして発展させていくためには、高品質なアノテーション情報の付加されたカイコゲノム情報を基本としつつ、これら公的データベース上の

昆虫ゲノム情報も活用し、種間解析結果を提供していくことが重要である。また、多様化・膨大化する生命科学情報を視覚化・データベース化してユーザーに提供するための技術を開発することも重要な課題となる。その実現には、シーケンシング技術の向上がもたらす情報量の爆発的な増加に対応できる、高速・低コストな計算機リソースを整備する必要がある、これまで以上に情報工学の重要性が増すと感じている。

引 用 文 献

- 1) MAZUMDER, R. et al. (2010): Biology Direct. 5: 12.
- 2) MITA, K. et al. (2004): DNA Res. 11: 27 ~ 35.
- 3) SHIMOMURA et al. (2009): BMC Genomics. 10: 486.
- 4) STEIN, L. (2001): Nat. Rev. Genet. 2: 493 ~ 503.
- 5) The International Silkworm Genome Consortium (2008): Insect Biochem. Mol. Biol. 38: 1036 ~ 1045.
- 6) XIA, Q. et al. (2004): Science 306: 1937 ~ 1940.

(登録が失効した農薬 4 ページからの続き)

「殺虫殺菌剤」

- イミダクロプリド・アジキシストロビン粒剤
20147: ゼネカ・アミスターアドマイヤー箱粒剤 (シンジェンタ ジャパン) 11/01/14
- BPMC・MEP・ジクロシメット粉剤
20383: 三共デラウススミバツサ粉剤 DL (ホクサン) 11/01/13
- ベンフラカルブ・ジクロシメット粒剤
20752: 大塚グランドデラウスオンコル粒剤 (大塚アグリテクノ) 11/01/22
- カルボスルファン・カルプロバミド粒剤
20787: FMC ウィンガゼット粒剤 (エフエムシー・ケミカルズ) 11/01/27
- MEP・ジクロシメット粉剤
21068: 三共デラウススミチオン (L) 粉剤 DL (ホクサン) 11/01/13
- クロチアニジン・ジクロシメット・フェリムゾン粉剤
21699: 三共ブラストップダントツ粉剤 DL (ホクサン) 11/01/13

「殺菌剤」

- 銅水和剤
18383: 三共コサイド DF (ホクサン) 11/01/13
22217: 三共コサイド DF (ホクサン) 11/01/13
- ジクロシメット粒剤
20361: 三共デラウス粒剤 (ホクサン) 11/01/13
- イプロジオン・テブコナゾール水和剤
21642: 三共ユキスター水和剤 (ホクサン) 11/01/13

「除草剤」

- MCPP 液剤
15019: 三共 MCPP 液剤 (ホクサン) 11/01/13
- MDBA 粒剤
16283: 三共バンベル-D 粒剤 (ホクサン) 11/01/13
- ピアラホス水溶剤
16647: 明治ハービエース水溶剤 (明治製菓) 11/01/20

●エンドタールニナトリウム塩液剤

- 18941: 三共エンドタール液剤 H (ホクサン) 11/01/13
- 18943: 三共エンドタール液剤 (ホクサン) 11/01/13
- ピリブチカルブ・ベンスルフロメチル水和剤
19136: [DIC] カルショット L フロアブル (日本曹達) 11/01/19
- ジメタメトリン・ピラゾレート・プレチラクロール・プロモブチド粒剤
20136: 三共スラッシャ 1 キロ粒剤 (三井化学アグロ) 11/01/14
- プレチラクロール・ベンゾフェナップ水和剤
20143: ホクコーチョップフロアブル (北興化学工業) 11/01/14
- フェントラザミド・ベンスルフロメチル粒剤
20524: 三共イノーバ 1 キロ粒剤 75 (ホクサン) 11/01/13
- フェントラザミド・ベンスルフロメチル水和剤
20540: 三共イノーバフロアブル (ホクサン) 11/01/13
- フェンメディファム水和剤
20571: 三共ベタナールフロアブル (ホクサン) 11/01/13
- オキサジクロメホン・クロメブロップ・ベンスルフロメチル粒剤
20745: JA ミスターホームラン L ジャンボ (全国農業協同組合連合会) 11/01/10
- オキサジアルギル水和剤
20747: フェナックスフロアブル (バイエルクロップサイエンス) 11/01/22
- フェントラザミド・プロモブチド・ベンスルフロメチル粒剤
21126: 三共イノーバ DX 1 キロ粒剤 75 (ホクサン) 11/01/13
- オキサジアルギル粒剤
22203: 三共キルクサ 1 キロ粒剤 (ホクサン) 11/01/13
- オキサジアルギル・プロモブチド・ベンゾフェナップ粒剤
22206: 三共パパール 1 キロ粒剤 (ホクサン) 11/01/13
- ピリブチカルブ・プレチラクロール乳剤
22326: 北海三共シング乳剤 (ホクサン) 11/01/13