

植物および植物関連微生物のゲノム情報データベース統合と高度化

(ゲノムアノテーションの高度化とそれを支援する技術、ツール、手法の開発)

かずさ DNA 研究所

1. 委託事業の3月末時点の判断基準になる目標

- ・ Gene Indexing 型、Complex Indexing 型アノテーションを合計して 19 万件以上の蓄積。
- ・ かずさナビゲーション（研究者 SNS）利用者 120 名。
- ・ Biomart 植物関連微生物 5 種類、主要植物 4 種類追加。データ追加と維持管理支援のためのプログラム群を整備する。

2. 3月末時点の達成状況

- ・ Gene Indexing 型アノテーション 186954 件（2/15 時点）達成。
- ・ ゲノムアノテーションジャンボリーでのかずさナビゲーション（研究者 SNS）の利用。利用者 120 名達成。
- ・ Biomart 植物関連微生物 18 種類、主要植物 8 種類追加。統合データベース講習会や学会での広報活動を行った。

3. プロジェクト終了後の維持、活用策

- ・ プロジェクト終了後も KazusaAnnotation ツール群と蓄積された情報ならびに CyanoBase, RhizoBase は、引き続き維持される予定である。また、現在提案中の関連プロジェクトが採択された場合、ツールの開発の継続と、主要なモデル微生物全般を対象とした形での展開を継続していく予定である。

4. 成果の概要

1) 高度情報集積データベースの運用と改良

KazusaAnnotation (<http://a.kazusa.or.jp>) は、分子データ上の位置情報を統合基盤情報とし、アノテーション・キュレーションの統合と高度化を可能にする新しいタイプのデータベースである。平成 21 年度までに、当初計画していた基本機能の開発と改善をほぼ完了している。各サービスのアクセス状況は、表 1 にまとめた。KazusaAnnotation の情報から CyanoBase/RhizoBase で遺伝子単位の文献情報を一覧できる表示に、MeSH ターム、関連文献、文献中に言及されている他の遺伝子を表示する機能を追加更新した（図 1）。

表 1 かずさにおける各サービスのアクセス状況

サービス名	URL	訪問者数	ユニーク訪問者数	ページビュー
KazusaAnnotation	http://a.kazusa.or.jp	13660	3260	128605
KazusaNavigation	http://navi.kazusa.or.jp	35753	22893	62016
KazusaWiki	http://wiki.kazusa.or.jp	10689	5333	29442
CyanoBase	http://genome.kazusa.or.jp/cyanobase	70801	18628	533348
RhizoBase	http://genome.kazusa.or.jp/rhizobase	15076	4465	112705

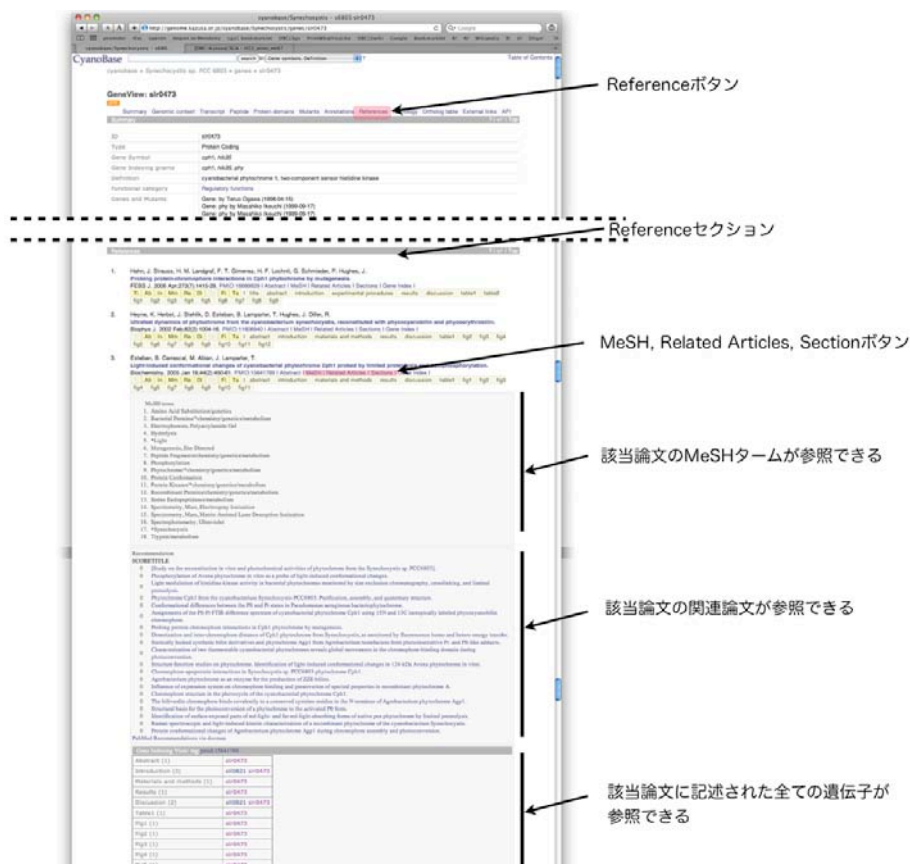


図1 CyanoBase, RhizoBase の遺伝子情報ページにある文献情報「Reference」に関連した情報を閲覧する機能追加した。この情報は、KazusaAnnotation に蓄積された Gene Indexing データから自動更新されている。

平成 22 年度には、KazusaAnnotation を利用して、植物関連土壌細菌のゲノムアノテーションジャンボリーを行った。公開前のゲノム情報をジャンボリー参加者のみで共有する仕組みを KazusaAnnotation に追加した。また、複数人数でゲノムアノテーションを行うための標準作業手順など協働作業手法を開発し KazusaWiki などを使い情報共有した。これらのシステム組み合わせることによりゲノムアノテーションを成功させた。

さらに、新しいアノテーションとして、シアノバクテリアの転写制御に関する情報を文献全文から抽出している。これは、KazusaAnnotation を用いて複雑な構造をもったデータを、タグ入力の形式で取り扱うための試みである。転写調節に関する生物学的な視座をタグとして設計し入力支援するツールを KazusaAnnotation システムの中で実現した (図 2)。また、標準作業手順を設計し、Wiki システムで共有することにより、アノテーション内容の一貫性と品質を担保している。この試みをとおして、遠隔地からの分散アノテーション作業により、専門性の高いデータベースを、構築することが可能なことが確かめられた。

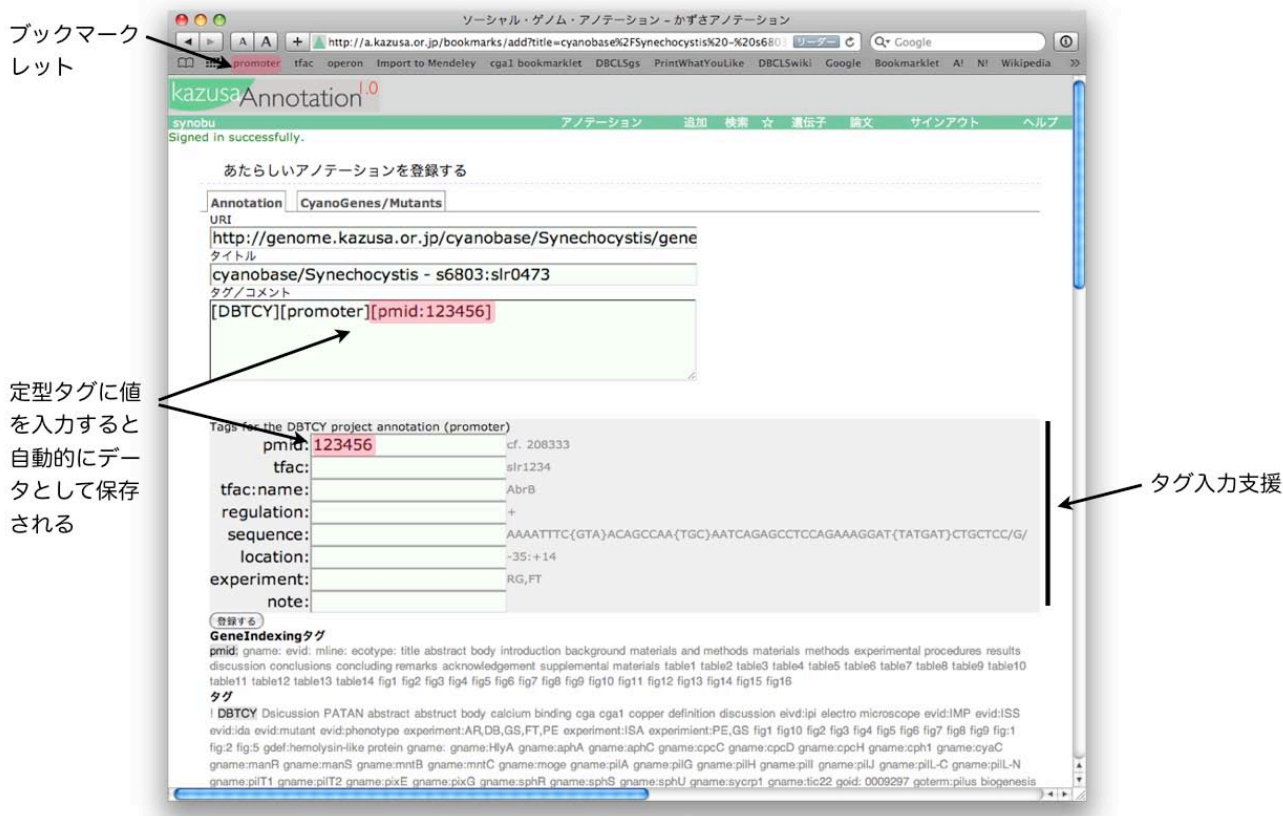


図2 KazusaAnnotationによる複雑なデータの入力支援ツール

2) ゲノムアノテーション情報の蓄積と高度化

平成22年度も論文中出现する遺伝子・タンパク質記載のインデクス情報(Gene Indexing型アノテーション)をKazusaAnnotationにより蓄積した。具体的な生物としては、*Synechocystis* sp. PCC 6803をはじめ13種類の光合成細菌に関する合計約4600報の論文について、また、*Mesorhizobium loti* MAFF303099をはじめ5種類の窒素固定根粒菌に関する合計約2600報の論文について遺伝子の位置情報の登録を完了している(詳細は表)。その他、高等植物としてマメ科モデル植物ミヤコグサ(*Lotus japonicus*)に関する約300報の論文も登録した。合計19生物種の約7500報の文献から遺伝子とタンパク質の情報を18万6千件抽出蓄積しゲノムの位置情報と論文記載情報の統合を図った(表2)。

KazusaAnnotationシステムを用いたジャンボリー形式のゲノムアノテーションでは、全国から11の大学や研究機関に所属する31名が参加した。3日間の合宿と、その後2週間の作業で6948遺伝子に対して2068の遺伝子名の改善を行った。

KazusaAnnotationシステムによる、専門分野データベース構築の試みであるシアノバクテリアの転写制御にかんする情報収集は、52報の論文から、プロモーター、転写因子、オペロンに関する情報を211件蓄積した。現在も進行中である。

これらの情報は上記のデータベースから即時公開し、広く生物学者の利用に供している。また、活用方法を一般の生物学者に周知するため、統合データベース講習会や植物研究関係のワークショップで紹介し利用を促した(平成22年度4月より10月末までの実績。講習会4回、

ワークショップ 3 回、招待講演 1 回、アノテーションジャンボリー 1 回)。

表 2 Gene Indexing 型アノテーションの対象生物種とアノテーション済み論文数

生物種	論文数
<i>Lotus japonicus</i>	296
<i>Mesorhizobium loti</i> MAFF303099	184
<i>Sinorhizobium melliloti</i> st.1021	176
<i>Bradyrhizobium japonicum</i> USDA110	947
<i>Rhizobium</i> sp. NGR234	128
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	1147
<i>Anabaena</i> sp. PCC 7120	722
<i>Anabaena variabilis</i> ATCC 29413	98
<i>Arthrospira platensis</i> NIES-39	8
<i>Chlorobium tepidum</i> TLS	177
<i>Gloeobacter violaceus</i>	27
<i>Nostoc punctiforme</i> ATCC 29133	95
<i>Prochlorococcus marinus</i> MED4	36
<i>Prochlorococcus marinus</i> MIT 9313	23
<i>Prochlorococcus marinus</i> SS120	20
<i>Synechocystis</i> sp. PCC 6803	2100
<i>Synechococcus elongatus</i> PCC 7942	744
<i>Synechococcus</i> sp. PCC 7002	269
<i>Thermosynechococcus elongatus</i> BP-1	287
Total	7484

3) 植物と関連微生物ゲノム情報の統合

生物種をまたいだ横断的なゲノム情報の取得や、理化学研究所など他機関の植物オーミクス情報との連携のため、KazusaMart (<http://mart.kazusa.or.jp>) を構築した。KazusaMart は、BioMart システム (<http://www.biomart.org/>) を利用している。このシステムは、グラフィカルインターフェイスを備え生物学者がデータベースに複雑な問い合わせをしてデータ取得することができる。植物関連データエントリに付与した種々の属性により高度なデータの絞り込みが出来るように各種データ調査、注釈、整形し KazusaMart に格納した。22 年 10 月以降に維持管理支援プログラム群を整備することで植物関連微生物 5 種類ならびに主要植物 4 種類を追加し、合計 64 生物種 (主要植物 12 種、植物関連微生物 52 種) のゲノム情報を公開している。また、統合データベース講習会にて広報活動を行なった。