

「植物オミックス情報および蛋白質構造情報」

21年度 研究成果報告書

平成22年4月

独立行政法人理化学研究所

豊田 哲郎

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、独立行政法人理化学研究所が実施した平成21年度「植物オミックス情報および蛋白質構造情報」の成果を取りまとめたものです。

1. 委託業務の目的	2
2. 平成21年度（報告年度）の実施内容	2
2.1 実施計画	2
2.2 実施内容(成果)	2
2.3 成果の外部への発表	6
2.4 活動（運営委員会等の活動等）	6
2.5 実施体制	7
「学会等発表実績」	7
「整備実績一覧」	15

1. 委託業務の目的

シロイヌナズナにおけるオミックス情報、並びに理化学研究所がタンパク 3000 プロジェクト等で解明した蛋白質構造データ（高等動植物由来及び微生物由来）及び付随する実験データについて、キュレーションやアノテーション作業を行い、それら利用価値を高めたデータを中核機関である大学共同利用機関法人情報・システム研究機構の方針に従って統合データベース事業に提供する。さらに上記アノテーション作業に必要なシステムの開発と運用を行い、提供する際に必要となるデータ変換を実施する。平成 22 年度には統合データベースの一部として生命科学分野の幅広い研究者に役立つ体制を確立し、将来的に理化学研究所の他のデータベースも統合化していくためのモデルケースをつくる。このため、情報・システム研究機構と共同で業務を行なう。

2. 平成 21 年度（報告年度）の実施内容

2.1 実施計画

①シロイヌナズナオミックス情報注釈

平成 20 年度に引き続き、これまでに公開したデータについて、アノテーション作業を行い、アノテーションデータを公開する。また、それ以外のデータについても国際的な合意を得つつ、データ公開を進める。

②高等動植物等由来蛋白質構造のアノテーション

平成 20 年度に引き続き、タンパク 3000 プロジェクトで解明された高等動植物等由来の蛋白質構造データのアノテーション、およびそれに付随する実験データのキュレーションとアノテーションを行う。

③微生物由来蛋白質構造のアノテーション

平成 20 年度に引き続き、これまでに公開したデータについてアノテーション作業を行い、アノテーションデータを公開する。またそれ以外のデータについても国際的な合意を得つつデータ公開を進める。

④アノテーションシステムの開発運用とデータ変換

平成 20 年度に引き続き、これまでに公開したデータ、および、理研の他のデータベースについて、アノテーション作業を行うためのシステムを運用・改善し、作成されたアノテーションを国際的な合意をえた基準で公開用データに変換する。

アノテーションの交換基準は国際的な合意を得たもので実施する必要性があり、国際的な場でこちらのデータベースを開示して国際的な合意を形成するための交流活動を国内外で行う。

2.2 実施内容(成果)

①シロイヌナズナオミックス情報注釈

理研サイネクス上で公開されているシロイヌナズナ変異体に関する 3 データベース (Fox-hunting, DS transposon, Activation tagging line) に含まれる表現型情報を対象に、新しい統一的なアノテーションを考えた。これは、従来それぞれの DB 開発者がそれぞれ独自の流儀で表していた表現型情報を「該当器官名+形質状態」として捉え、既存の公開オントロジー (Plant Ontology および PATO) に定義された用語の組み合わせによって統一的に再定義しようとするものである。具体的な形質を一つ例にとると、「開花が遅い」という形質は P0:0007016+PATO:000303 という組み合わせで表現できる (図 1 参照)。この手法によって前述の 3 データベース中に記述されている表現型情報を整理すると全部で 115 種類となった。この 115 種類を新しいデータベース「シロイヌナズナ変異体情報」に収納し、3 データベースそれぞれに既存の観察情報に対して、13142 件, 221 件, 1268 件のリンクを設定して公開した。ここまでの、明確な表現型情報が公開されているレコード全てについて新しいアノテーション情報が関連付けられた。この新データベースを公開したことにより、アノテーション情報の公開という当初目標は達成された。

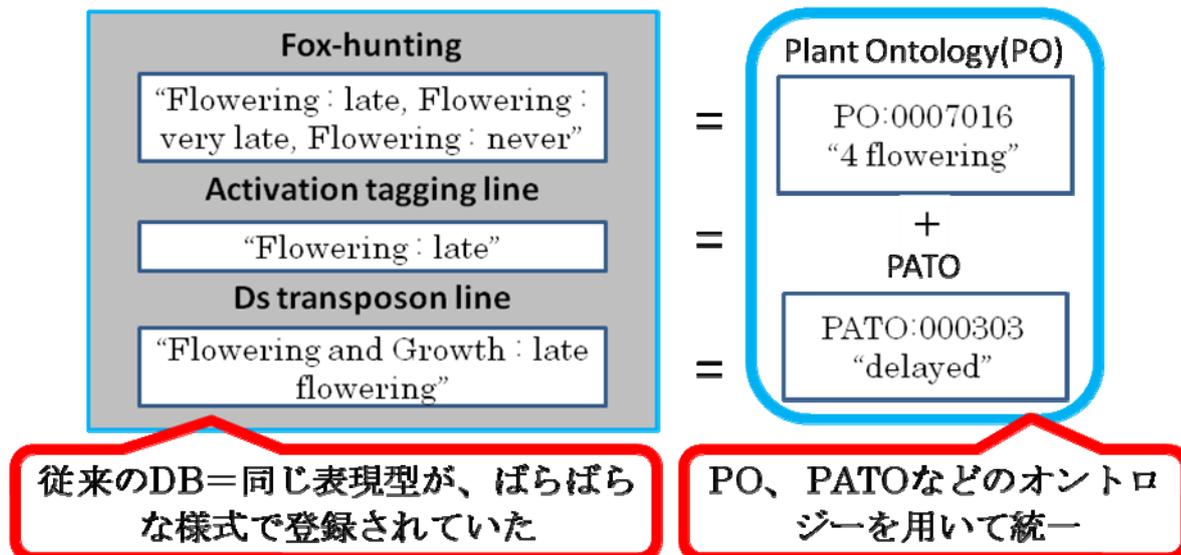


図1： 公開オントロジーを用いた表現型情報のアノテーションの例

②高等動植物等由来蛋白質構造のアノテーション

高等動植物タンパク質の試料調製に関わる発現検討実験（無細胞タンパク質発現用発現系構築実験4万8千件、スモールスケール発現実験8万件）に基づいて X 線結晶構造解析のための大量調製実験（大量合成および大量発現実験4千件、精製実験5千件）を行い、結晶化実験（結晶化条件138万件、結晶観察1517万件）へと進め、タンパク3000プロジェクトで解明されたタンパク質構造データのうち平成21年度分としては1万1千件の回折実験データ（画像データ枚数）を公開可能とした。

③微生物由来蛋白質構造のアノテーション

平成20年度に閲覧可能な状態で公開したデータについて国際基準に沿った形式への変換を行い、アノテーションシステムからダウンロード可能な状態にした。このアノテーションデータ公開に合わせてプレス発表を行った（平成21年7月23日）（添付資料1）。また、国際的な合意を得るため、台湾の3カ所で海外デモンストレーションを実施した（国家同步輻射研究中心（新竹，平成21年9月14日）・国立台湾大学（台北，同15日）・国立成功大学（台南，同16日））（添付資料2）。さらに、追加データ（試料調製データ・結晶化実験データ・回折実験データ・重原子導入蛋白質データ）のキュレーション作業を行った。

④アノテーションシステムの開発運用とデータ変換

所定の表現型について、関連する遺伝子情報を対象生物種や実験手法等の違いに関わらず検索可能となるようにシステムの改良を行った。まず、生物種を問わず適用可能で表現型・遺伝子型・実験条件・対応する文献などの関連情報を網羅するテンプレート（セマンティックウェブ構築のための枠組み。図2参照）を作成した。前述した「シロイヌナズナ変異体情報」データベースもこのテンプレートに沿って作成した。これにより、表現型をキーとしたデータベース横断型のオミックス情報検索が容易に可能となった。オントロジー体系とデータ公開に関する国際的な合意を形成するための交流活動としては、平成21年7月12、13の両日に滋賀県で「国際マウス表現型情報統合シンポジウム（International Phenome Integration Meeting）」を開催した。この場で、EU・米国・カナダの研究機関との間で、国際的な研究の分担やデータ共有に向けて共同で取り組むこと、理研としてはセマンティックウェブを基盤としたマウス表現型データの標準化作業を行うことなどが合意された（添付資料3）。また、マウス表現型データの標準化に関しては4th AMMRA Annual Meeting、The 20th International Conference on Genome Informatics、第32回日本分子生物学会年会等で発表を行ない、関係する研究者と議論を進めた（添付資料4・5・6）。また、項目①の全データをRDF形式に変換して統合データベースセンター（かずさDNA研究所のグループ）に提供した。

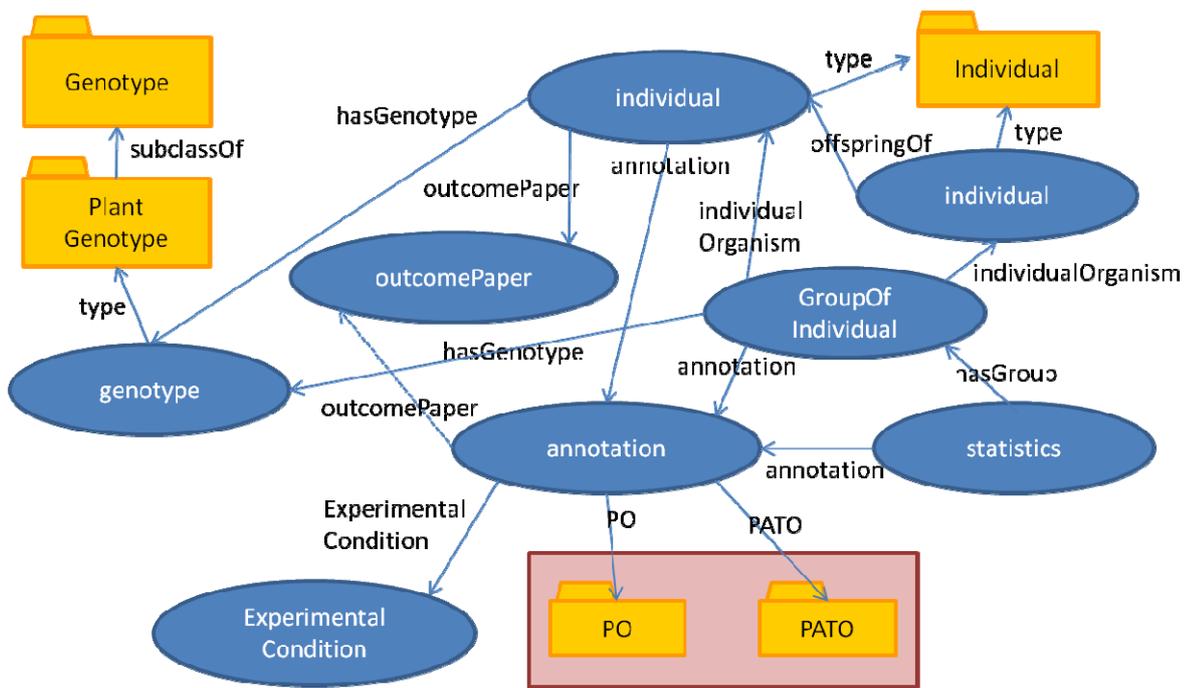


図 2：生物種を問わず適用できるフェノーム情報表現のためのテンプレート

理研が公開するデータベースの統合状況

理研のトップページ

理研サイネス DBレジストリ



データベース一覧
をクリック



105件の理研DBサイトが
収集されている
うち38サイトは実データを
理研サイネスにも収録済

理研サイネス登録データの統計値 (2010年3月現在)

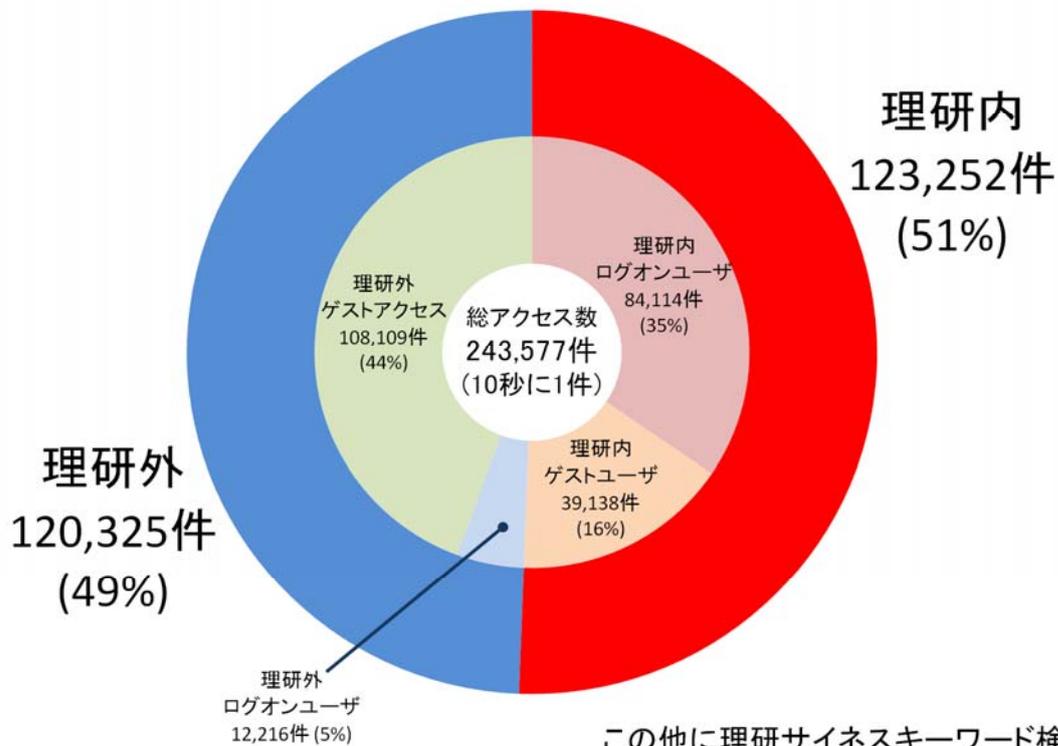
<http://database.riken.jp> 理研内外からもアクセス可能



・登録ユーザ数	172 ユーザ	
・登録ディスク(データベース)数	273 件	(うちBacpediaデータ 1 件)
・登録フォルダ(テーブル)数	16,478 件	(うちBacpediaデータ 29 件)
・登録ページ数	20,164,571 件	(うちBacpediaデータ 15,481,267 件)
・データ容量	12.1 TB	(うちBacpediaデータ 6.56 TB)
・ファイル数	3億2300万件	(うちBacpediaデータ 1億8500万件)

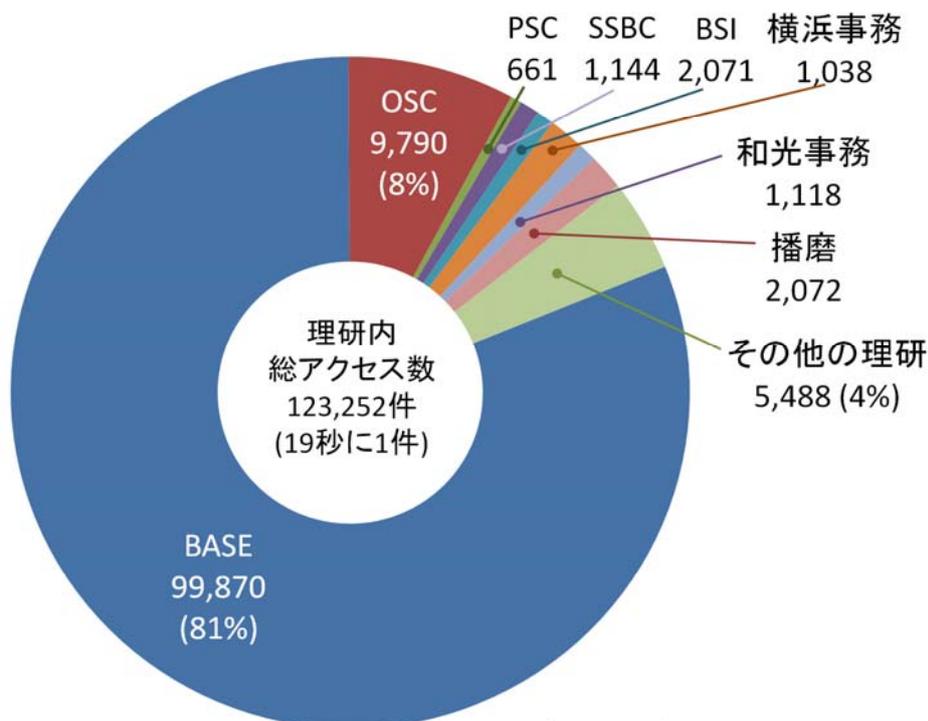
2

理研サイネスユーザアクセス数 2010年2月



この他に理研サイネスキーワード検索ページへのアクセスが32,303件あった。

理研サイネス 理研内組織別アクセス数 2010年2月



BASEはデータ作成を行っていたため、書き込みアクセス件数が多くなった。

2. 3 成果の外部への発表

後掲「学会等発表実績」(7ページ以下)を参照。

2. 4 活動 (運営委員会等の活動等)

概要	開催日	活動内容
運営委員会	2009/5/29	統合DB作業部会分科会 (LS統合DBセンター、文京区)
運営委員会	2009/6/25	第7回統合DBプロジェクト研究運営委員会 (ライフサイエンス統合データベースセンター大会議室)
シンポジウム	2009/10/5	シンポジウム「科学における情報の上手な権利化と共有化」(東京大学農学部構内 弥生講堂(一条ホール))
運営委員会	2010/1/25	「統合データベースプロジェクト」研究運営委員会(第9回) (情報・システム研究機構事務局会議室)

運営委員会	2010/2/2	作業部会分科会 立体構造・植物、人材育成機関
-------	----------	------------------------

2.5 実施体制

アノテーション作業と中核機関へのデータ提供を円滑に行うために、理研生命情報基盤研究部門の豊田が責任者として全体のとりまとめを行い、下記アノテーション作業のインフラシステムの運用とデータ管理、および生成されたアノテーションデータの変換等を主に担当した。トランスクリプトームについては理研植物科学研究センターの関・櫻井・花田がデータの解析に基づくアノテーションの作成を担当した。フェノームについては理研植物科学研究センターの松井と黒森が中心となってアノテーション作業を担当した。リソースについては理研バイオリソースセンターの深海がシロイヌナズナの保有リソース情報のデータ変換と提供を担当した。タンパク3000のデータについては理研生命分子システム基盤研究領域の横山が責任者となり、中山、寺田が中心となって高等動植物等由来の蛋白質構造のアノテーションとそれに付随する実験データのキュレーションを担当した。また微生物由来の蛋白質構造については理研放射光科学総合研究センターの国島が責任者となり、キュレーションは浅田が、アノテーションは国島・菅原が主に担当した。

平成 21 年度における実施体制

業 務 項 目	担当機関等	研究担当者
サブテーマ① シロイヌナズナオミックス情報注釈	理研生命情報基盤研究部門 部門長	◎ 豊田 哲郎
	理研植物科学研究センター チームリーダー	関 原明
	理研植物科学研究センター ユニットリーダー	櫻井 哲也
	理研植物科学研究センター 研究員	花田 耕介
	理研植物科学研究センター 副グループディレクター	有田 正規
	理研植物科学研究センター グループディレクター	松井 南
	理研植物科学研究センター 研究員	黒森 崇
	理研バイオリソースセンター 室長	深海 薫
	理研バイオリソースセンター ユニットリーダー	榎屋 啓志
	理研生命情報基盤研究部門 リサーチアソシエイト	浅野 (松下) さとみ
	理研生命情報基盤研究部門 研究員	蒔田 由布子
	理研生命情報基盤研究部門 研究員	土井 考爾
	理研生命情報基盤研究部門 リサーチアソシエイト	原田 えりみ

	理研生命情報基盤研究部門 パートタイマー	西方 公郎
	理研生命情報基盤研究部門 パートタイマー	Deshpande Mrinalini
	理研生命情報基盤研究部門 パートタイマー	Bhatia Rinki
	理研生命情報基盤研究部門 パートタイマー	渡辺 敏明
	理研生命情報基盤研究部門 パートタイマー	荒井 ひろみ
サブテーマ② 高等 動植物等由来蛋白質 構造のアノテーション	理研生命分子システム基盤研究領域 領域長	○ 横山 茂之
	理研生命分子システム基盤研究領域 上級研究員	寺田 貴帆
	理研生命分子システム基盤研究領域 上級技師	中山 朋子
	理研生命分子システム基盤研究領域 特別研究員	吉良 聡
	理研生命分子システム基盤研究領域 特別研究員	高谷 大輔
サブテーマ③ 微生物 由来蛋白質構造のア ノテーション	播磨研究所放射光科学総合研究セン ター グループ副ディレクター	○ 国島 直樹
	播磨研究所放射光科学総合研究セン ター 研究員	菅原 道泰
	播磨研究所放射光科学総合研究セン ター リサーチアソシエイト	浅田 征彦
	播磨研究所放射光科学総合研究セン ター グループディレクター	倉光 成紀
	播磨研究所放射光科学総合研究セン ター チームリーダー	新海 暁男
サブテーマ④ アノ テーションシステム の開発運用とデータ 変換	理研生命情報基盤研究部門 部門長	○ 豊田 哲郎
	理研生命情報基盤研究部門 研究員	飯田 慶
	理研生命情報基盤研究部門 上級研究員	小林 紀郎
	理研生命情報基盤研究部門 技師	吉田 有子
	理研生命情報基盤研究部門 技師	望月 芳樹

理研生命情報基盤研究部門 テクニカルスタッフ	石井 学
理研生命情報基盤研究部門 リサーチアソシエイト	梅津 亮
理研生命情報基盤研究部門 特別研究員	川口 修治
理研生命情報基盤研究部門 テクニカルスタッフ	松嶋 明宏
理研生命情報基盤研究部門 特別研究員	神坂 紀久子

学 会 等 発 表 実 績

委託業務題

「植物オミックス情報および蛋白質構造情報」

目：

機 関 名：

独立行政法人理化学研究所

1. 学会等における口頭・ポスター発表

発表した成果（発表題目，口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
○ PosMed-plus: an information system making the utmost use of Arabidopsis knowledge to assist molecular breeding of crops	Makita Yuko, Kobayashi Norio, Iida Kei, Asano Satomi (Matsushita), Yoshida Yuko, Heida Naohiko, Deshpande Mrinalini, Bhatia Rinki, Matsushima Akihiro, Ishii Manabu, Mochizuki Yoshiki, Hanada Kousuke, Kawaguchi Shuji, Kuromori Takashi, Seki Motoaki, Shinozaki Kazuo, Toyoda Tetsuro	20th International Conference on Arabidopsis Research 2009	2009/6/30-7-4	国際会議
○ ARTADE2.0: A mathematical integration of tiling array, CAGE and sequence data to elucidate the transcriptional system's dynamics of Arabidopsis thaliana	Kawaguchi Shuji, Hanada Kousuke, Iida Kei, Mochizuki Yoshiki, Matsushima Akihiro, Okamoto Masanori, Seki Motoaki, Shinozaki Kazuo, Toyoda Tetsuro	20th International Conference on Arabidopsis Research 2009	2009/6/30-7-4	国際会議

○	Multiple alternative splicing events in individual transcripts; analyses using full-length cDNAs and tiling arrays	Iida Kei, Kawaguchi Shuji, Hanada Kousuke, Hukami-Kobayashi Kaoru, Toyoda Atsushi, Sakaki Yoshiyuki, Kobayashi Masatomo, Seki Motoaki, Shinozaki Kazuo, Toyoda Tetsuro	20th International Conference on Arabidopsis Research 2009	2009/6/30-7-4	国際会議
◎	シロイヌナズナオミックスに関する統合データベースの開発	土井 考爾, 原田 えりみ, 神坂 紀久子, 吉田 有子, 豊田 哲郎	文科省統合データベースプロジェクト「データベースが拓くこれからのライフサイエンス」	2009/6/12	国内会議
◎	RDF based data integration system, SciNeS	Toyoda Tetsuro	International Phenome Integration Meeting RIKEN/InterPhenome/CASIMIR	2009/7/12-13	国際会議
○	アラビドプシスを中心とした植物トランスクリプトーム DB と関連アプリケーションの紹介	蒔田 由布子	第 27 回日本植物細胞分子生物学会 (藤沢) 大会・シンポジウム	2009/7/30-31	国内会議
○	創緑研究に向けた高等植物におけるゲノム設計の自動化と微生物からのパスウェイ導入	豊田 哲郎	日本遺伝学会 第 81 回大会	2009/9/16-18	国内会議
○	創緑研究のためのゲノム設計とバイオインフォマティクスの基盤	豊田 哲郎	第 15 回理事長フェンドワークショップ「人類存続のための環境・エネルギー分野の研究」-バイオマスエンジニアリング研究プログラムの戦略-	2009/10/9-10	国内会議

○	Summary of international phenome integration meeting RIKEN/InterPhenome/CASIMIR, July 12-13, 2009, Kyoto Japan	Masuya Hiroshi, John Hancock, Paul Schofield, Joel Richardson, Nora Tsao, Hugh Morgan, Michael Grunberger, Kobayashi Norio, Takada Toyoyuki, Abe Kuniya, Ann Marie Mallon, Janan Eppig, Wakana Shigeharu, Shiroishi Toshihiko, Obata Yuichi, Toyoda Tetsuro	23rd International Mammalian Genome Conference	2009/11/1-4	国際会議
○	Toward the integration of mouse phenotype information, technology and international cooperation	Masuya Hiroshi, Tanaka Nobuhiko, Waki Kazunori, Mizoguchi Riichiro, Toyoda Tetsuro, Yoshiki Atsushi, Wakana Shigeharu	4th AMRA Annual Meeting	2009/12/17-18	国際会議
○	A trial to develop a integrative database of mouse phenotype related information	Masuya Hiroshi, Kobayashi Norio, Tanaka Nobuhiko, Waki Kazunori, Kozaki Kouji, Wakana Shigeharu, Mizoguchi Riichiro, Toyoda Tetsuro	The 20th International Conference on Genome Informatics	2009/12/14-16	国際会議
◎	セマンティックウェブを利用したシロイヌナズナの表現型および関連情報の統合の試み	土井 考爾, 小林 紀郎, 榎屋 啓志, 黒森 崇, 豊田 哲郎	第 32 回日本分子生物学会年会	2009/12/9-12	国内会議
○	タイリングアレイを用いた mRNA 中の cis 調節因子の発見	飯田 慶, 川口 修治, 原田 えりみ, 花田 耕介, 松井 章浩, 関 原明, 豊田 哲郎	第 32 回日本分子生物学会年会	2009/12/9-12	国内会議

○	PosMed-plus: 種間情報を推論的に統合する植物分子育種のためのインテリジェントサーチエンジン	蒔田 由布子, 小林 紀郎, 望月 芳樹, 吉田 有子, 松嶋 明宏, 石井 学, 飯田 慶, 花田 耕介, 黒森 崇, 関 原明, 篠崎 一雄, 豊田 哲郎	第 32 回日本分子生物学会年会	2009/12/9-12	国内会議
○	Genome-wide analysis of histone H4 acetylation status in hda6, Arabidopsis histone deacetylase6 mutant using ChIP-chip and ChIP-seq systems	金 鍾明, 藤 泰子, 遠藤 高帆, 田中 真帆, 大島 真澄, 倉田 哲也, 松井 章浩, 栗原 志夫, 石田 順子, 諸澤 妙子, 篠崎 一雄, 長谷部 光泰, 豊田 哲郎, 関 原明	第 32 回日本分子生物学会年会	2009/12/9-12	国内会議
○	Transcriptome analyses revealed diverse expression changes in agol and hyl1 Arabidopsis mutants	栗原 志夫, 神沼 英里, 松井 章浩, 川嶋 真貴子, 諸澤 妙子, 石田 順子, 田中 真帆, 望月 芳樹, 篠崎 一雄, 豊田 哲郎, 関 原明	第 32 回日本分子生物学会年会	2009/12/9-12	国内会議
○	Exploring post-transcriptional regulations using vast expression profile data.	飯田 慶	ADVANCED SCIENCE INSTITUTE & YOKOHAMA INSTITUTE FORUM(基幹研横浜研連携フォーラム)	2010/2/9	国内会議
○	Exploring post-transcriptional regulations using vast expression profile data.	飯田 慶, 川口 修治, 豊田 哲郎	ADVANCED SCIENCE INSTITUTE & YOKOHAMA INSTITUTE FORUM(基幹研横浜研連携フォーラム)	2010/2/9	国内会議
○	RIKEN SciNeS	豊田 哲郎	DBCLS BioHackathon 2010	2010/2/8	国際会議
○	Arabidopsis の新規アンチセンス RNA の網羅的解析に向けて	原田 えりみ, 松井 章浩, 川口 修治, 飯田 慶, 花田 耕介, 関 原明, 豊田 哲郎	第 51 回日本植物生理学会年会	2010/3/18-21	国内会議

○	Analysis of post-transcriptional gene regulation; RUST using tiling arrays and computational methods.	飯田 慶、川口 修治、原田 えりみ、花田 耕介、松井 章浩、岡本 昌憲、篠崎 一雄、関 原明、豊田 哲郎	第 51 回日本植物生理学会年会	2010/3/18-21	国内会議
○	乾燥, 低温, 塩ストレスおよび ABA 処理時におけるシロイヌナズナ small RNA の網羅的探索	松井章浩、神沼英里、諸沢妙子、中嶋舞子、豊田哲郎、篠崎一雄、関原明	第 51 回日本植物生理学会年会	2010/3/18-21	国内会議
○	理研総合データベースによる国内外への貢献	豊田 哲郎	第 18 回理事長ファンドワークショップ「理研総合データベースを核にした研究基盤連携網の構築」-課題解決型研究のための情報戦略と国内外統合 DB 事業との連携-	2010/3/12-13	国内会議
○	理研サイネスを活用したデータベース連携と統合	豊田 哲郎	第 18 回理事長ファンドワークショップ「理研総合データベースを核にした研究基盤連携網の構築」-課題解決型研究のための情報戦略と国内外統合 DB 事業との連携-	2010/3/12-13	国内会議
◎	「放射光科学総合研究センターにおけるタンパク質 X 線結晶構造解析実験データの公開」 (ポスター発表)	浅田征彦、水谷尚志、菅原道泰、新海暁男、倉光成紀、国島直樹	理研シンポジウム「高度好熱菌丸ごと一匹プロジェクト第 7 回連携研究会」, 理研播磨研究所	2009/8/21-23	国際会議
○	"RIKEN database for protein research", 口頭発表 (台湾の 3 カ所における海外デモ)	Yukuhiko Asada, Naoki Kunishima	国家同步輻射研究中心 (新竹) 国立台湾大学 (台北) 国立成功大学 (台南)	2009/9/14-16	国際会議
◎	「放射光科学総合研究センターにおけるタンパク質 X 線結晶構造解析実験データの公開」 (ポスター発表)	浅田征彦、水谷尚志、菅原道泰、新海暁男、倉光成紀、国島直樹	第 82 回日本生化学会大会, 神戸ポートアイランド	2009/10/21-24	国際会議

2. 学会誌・雑誌等における論文掲載

掲載した論文 (発表題目)	発表者氏名	発表した場所 (学会誌・雑誌等名)	発表した時期	国内・外の
---------------	-------	-------------------	--------	-------

				別	
○	OmicBrowse: a Flash-based high-performance graphics interface for genomic resources	Matsushima Akihiro, Kobayashi Norio, Mochizuki Yoshiki, Ishii Manabu, Kawaguchi Shuji, Endo Takaho, Umetsu Ryo, Makita Yuko, Toyoda Tetsuro	Nucleic Acids Research	2009年5月	国外
○	PosMed-plus: an Intelligent Search Engine that Inferentially Integrates Cross-species Information Resources for Molecular Breeding of Plants	Makita Yuko, Kobayashi Norio, Mochizuki Yoshiki, Yoshida Yuko, Asano (Matsushita) Satomi, Heida Naohiko, Deshpande Mrinalini, Bhatia Rinki, Matsushima Akihiro, Ishii Manabu, Kawaguchi Shuji, Iida Kei, Hanada Kousuke, Kuromori Takashi, Seki Motoaki, Shinozaki Kazuo, Toyoda Tetsuro	Plant and Cell Physiology	2009年6月	国外
○	Transcriptome analyses revealed diverse expression changes in ago1 and hyl1 Arabidopsis mutants	Kurihara Yukio, Kaminuma Eri, Matsui Akihiro, Kawashima Makiko, Tanaka Maho, Morosawa Taeko, Ishida Junko, Mochizuki Yoshiki, Shinozaki Kazuo, Toyoda Tetsuro, Seki Motoaki	Plant and Cell Physiology	2009年7月	国外

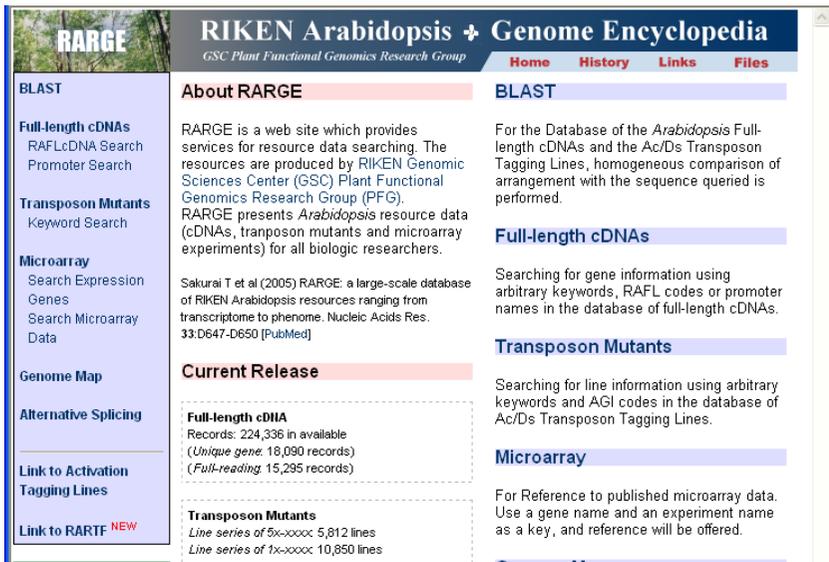
○	Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays	Okamoto Masanori, Matsui Akihiro, Tatematsu Kiyoshi, Morosawa Taeko, Ishida Junko, Tanaka Maho, Endo Takaho, Mochizuki Yoshiki, Toyoda Tetsuro, Kamiya Yuji, Shinozaki Kazuo, Nambara Eiji, Seki Motoaki	The Plant Journal	in press	国外
---	---	--	-------------------	----------	----

3. プレス発表	発表タイトル	発表者氏名	発表した場所	掲載日
◎	「世界最大級のタンパク質結晶構造解析実験データベースを公開—90万件の結晶化条件などを整備した実験データ群が利用可能に—」	浅田征彦、国島直樹	理化学研究所	2009年7月

(1-3) データの管理状況

①更新頻度等の管理状況、体制	不定期にアノテーション情報を更新。
②その他、特記事項	

(1-4) データベース関係

(1)DB 管理者数	2 名
(2)キュレータ・アノテータ数	1 名
(3)データ構造	リレーショナルデータベース
(4)DB 管理ソフト	PostgreSQL
(5)サーバの OS	Red Hat Enterprise Linux ES
(6)サーバ規模	1 基
(7)DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8)独立 IP 数	1
(9)その他、特記事項	画面の様子 RARGE の画面 

--	--

[2 件目] トランスクリプトーム（タイリングアレイデータ）業務計画書に記載した業務項目①に対応

(1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリー等の種類、数	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19 種類（各々F および R アレイを用いた計 6 回のハイブリ実験を行う） GEO データベースに登録されている、シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データの内、3 回以上繰り返し実験を行ったものは、わずか 4 種類のみである。
(3) 測定方法	
(4) データの内容	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19 種類（各々F および R アレイを用いた計 6 回のハイブリ実験を行う） 実験内容：播種後 2 週間目の植物体を用いた乾燥、低温、塩ストレス、ABA 処理、再吸水処理による乾燥ストレスからの回復過程など
(5) その他、特記事項	

(1-2). データソース

(1) 現在のデータ量	シロイヌナズナ全ゲノムタイリングアレイを用いた発現プロファイル解析データ 19 種類
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他（下欄に詳細を記述）
(3) 将来の増加の見込み	さらに今後も新規のタイリングアレイデータを追加する一方で、同じサンプルをシーケンサー等の別の手段で計測したデータも生産される予定である。このため、メタデータでこれらに関連付けることで多様な計測手段による情報を組み合わせた統合解析が可能なデータとなり、バイオインフォマティクスの発展に大変有用なデー

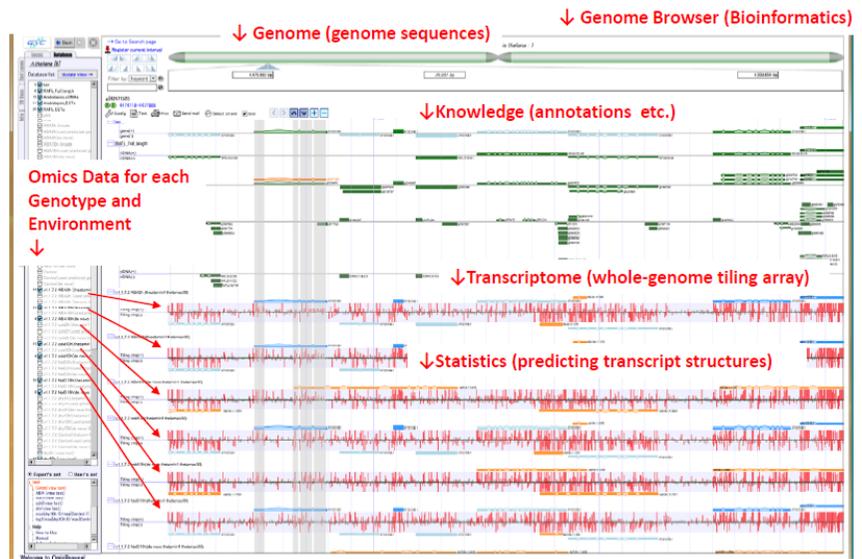
	データベースとなっていくことが期待される。
(4) 権利関係	所有者（理研PSC、植物ゲノム機能研究グループ） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他〔論文発表後データを公開する予定である。〕
(5) その他、特記事項	

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	実験データの取得に際して随時追加
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	2名
(2) キュレータ・アナデータ数	1名
(3) データ構造	ファイル
(4) DB 管理ソフト	ファイルシステム
(5) サーバの OS	CentOS
(6) サーバ規模	2ノード
(7) DB へのアクセス数	未公開
(8) 独立 IP 数	
(9) その他、特記事項	画面の様子 OmicBrowse の画面



--	--

[3 件目] 業務計画書に記載した業務項目①に対応

(1-1). データの種類

(1) 生物種	シロイヌナズナ
(2) 試料・ライブラリー等の種類、数	454 シーケンサーを用いた small RNA の大量解析データ 11 種類 454 シーケンサーを用いたシロイヌナズナの small RNA データは、これまでに 6 種類報告されている。
(3) 測定方法	
(4) データの内容	454 シーケンサーを用いた small RNA の大量解析データ 11 種類 用いた植物材料：播種後 2 週間目の植物体を用いた乾燥、低温、塩ストレス、ABA 処理、無処理の植物体など
(5) その他、特記事項	

(1-2). データソース

(1) 現在のデータ量	454 シーケンサーを用いた small RNA の大量解析データ 11 種類
(2) データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他（下欄に詳細を記述）
(3) 将来の増加の見込み	今後、454シーケンサーよりも大量の配列解析が可能なSolexaシーケンサーを用いてsmall RNAの大量解析を行う予定である。
(4) 権利関係	所有者（理研PSC、植物ゲノム機能研究グループ） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他〔論文発表後データを公開する予定である。〕
(5) その他、特記事項	

項	
---	--

(1-3). データの管理状況

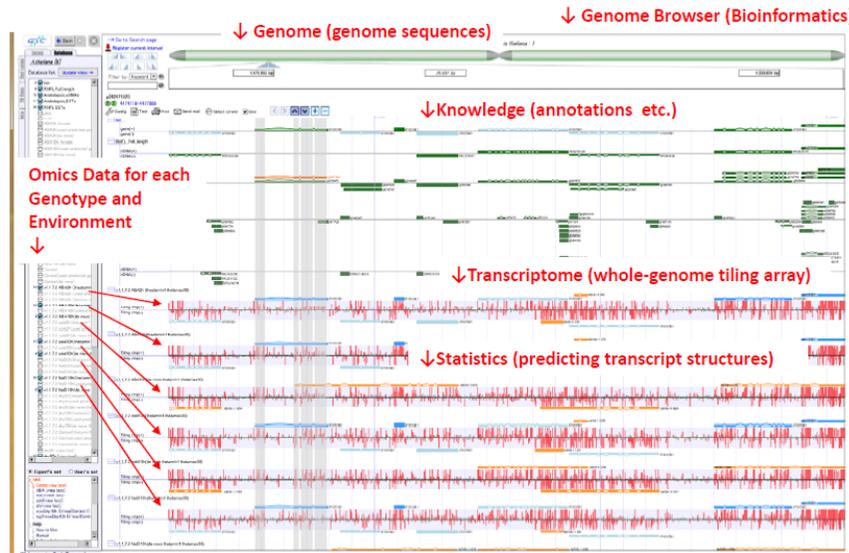
(1)更新頻度等の管理状況、体制	実験データの取得に際して随時追加
(2)その他、特記事項	

(1-4). データベース関係

(1)DB 管理者数	1名
(2)キュレータ・アナデータ数	1名
(3)データ構造	ファイル
(4)DB 管理ソフト	ファイルシステム
(5)サーバの OS	CentOS
(6)サーバ規模	1ノード
(7)DB へのアクセス数	未公開
(8)独立 IP 数	

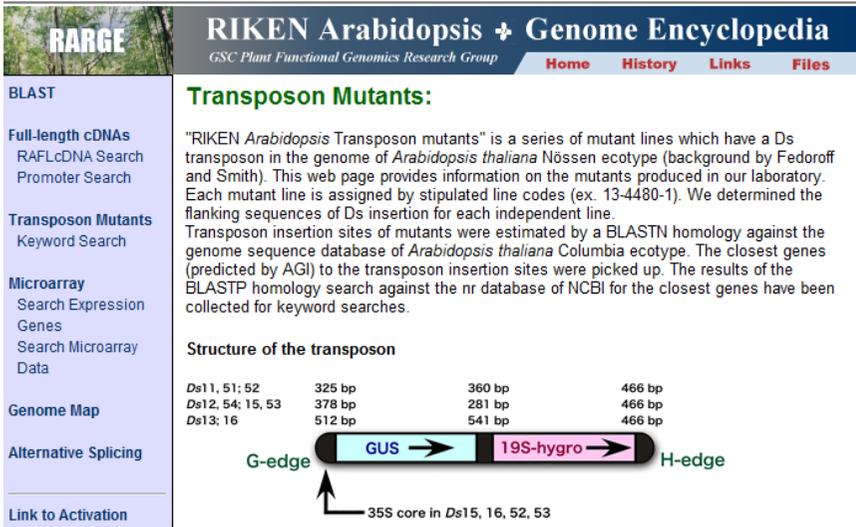
(9)その他、特記事項

画面の様子 OmicBrowse の画面

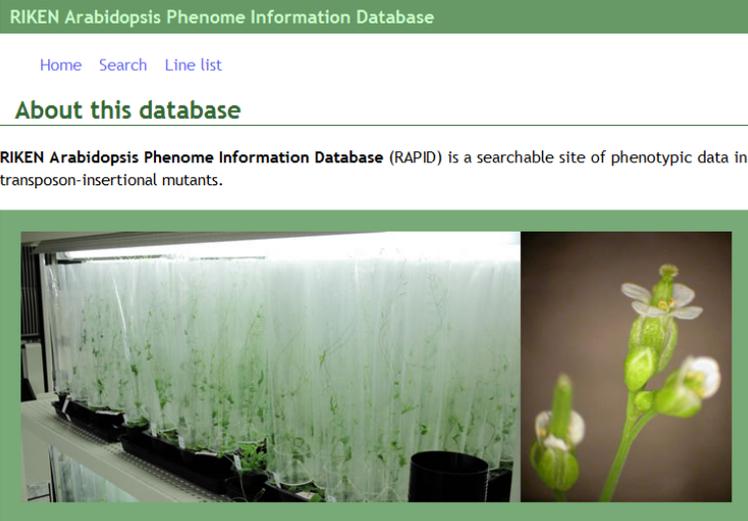


(1)更新頻度等の管理状況、体制	理化学研究所植物科学研究センターで維持されている他、生命情報基盤研究部門のサーバ（理研サイネス）上にも情報が統合されている。
(2)その他、特記事項	

(1-4). データベース関係

(1)DB 管理者数	2 名
(2)キュレータ・アナデータ数	1 名
(3)データ構造	リレーショナルデータベース
(4)DB 管理ソフト	PostgreSQL
(5)サーバの OS	Red Hat Enterprise Linux ES
(6)サーバ規模	1 基
(7)DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8)独立 IP 数	1
(9)その他、特記事項	画面の様子 

(1-4). データベース関係

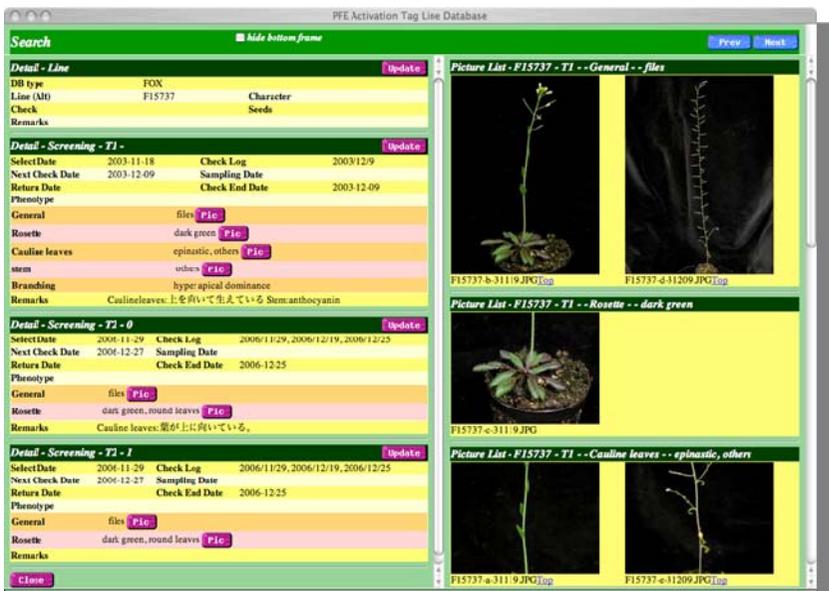
(1)DB 管理者数	2 名
(2)キュレータ・アナ データ数	1 名
(3)データ構造	リレーショナルデータベース、ファイル
(4)DB 管理ソフト	PostgreSQL、ファイルシステム
(5)サーバの OS	Red Hat Enterprise Linux ES
(6)サーバ規模	1 基
(7)DB へのアクセス数	データベース RARGE 総計 約 5000 ページ/月
(8)独立 IP 数	1
(9)その他、特記事項	画面の様子 

[6 件目]フェノーム（アクティベーション・タグライン）業務計画書に記載した業務項目①に対応

(1)更新頻度等の管理状況、体制	1週間に1回アノテーション情報の更新を行っている。 理化学研究所植物科学研究センターで維持されている他、生命情報基盤研究部門のサーバ（理研サイネス）上にも情報が統合されている。
(2)その他、特記事項	

(1-4). データベース関係

(1)DB 管理者数	1名
(2)キュレータ・アノテータ数	1名
(3)データ構造	RDB
(4)DB 管理ソフト	PostgreSQL
(5)サーバの OS	Red Hat Enterprise Linux ES (外部公開用) Red Hat Enterprise Linux AS (内部用)
(6)サーバ規模	2CPU (Pentium Xeon) (外部公開用) 4CPU (Pentium Xeon) (内部用)
(7)DB へのアクセス数	登録外部ユーザ数 884 (2007年登録 82、2006年登録 142、2005年登録 141、2004年登録 327、2003年登録 192)
(8)独立 IP 数	1 (外部公開用)、1 (内部用)
(9)その他、特記事項	画面の様子



(5) その他、特記事項	http://www.brc.riken.jp/lab/epd/catalog/cdnac1one.html http://saber.epd.brc.riken.jp/sabre7/SABRE0101.cgi
--------------	--

(1-3). データの管理状況

(1) 更新頻度等の管理状況、体制	データの増加に応じて、年に数回までの更新あり 外部業者に運用を業務委託している
(2) その他、特記事項	

(1-4). データベース関係

(1) DB 管理者数	理研側担当者 2 名、業者側数名
(2) キュレータ・アナデータ数	なし
(3) データ構造	テキストファイル、リレーショナルデータベース
(4) DB 管理ソフト	PostgreSQL
(5) サーバの OS	Linux
(6) サーバ規模	パーソナルコンピュータ
(7) DB へのアクセス数	計測値なし
(8) 独立 IP 数	2
(9) その他、特記事項	画面の様子



--	--

[8 件目] シロイヌナズナ Fox-hunting 変異体データベース

(1-1) データの種類

①生物種	シロイヌナズナ
②試料・ライブラリ 一等の種類、数	FOX-Hunting 法によって cDNA を挿入することによって得られた変異体のデータベース。 当該実験のデータのうちデータ提供者が公開可能とした情報は全て公開した。
③測定方法	
④データの内容	挿入された完全長 cDNA の遺伝子 ID・挿入方向、観察された表現型情報のテキストベースの記載と植物体の画像データなど
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	登録した変異体 14125 系統 表現型情報 13292 件
②データ区分	<input checked="" type="checkbox"/> 自前 <input type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他（下欄に詳細を記述）
③将来の増加の見込み	なし
④権利関係	所有者（独立行政法人理化学研究所） 公開（ <input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否 <input type="checkbox"/> その他 []）
⑤その他、特記事項	http://nazunafox.psc.database.riken.jp にて公開

(1-3) データの管理状況

①更新頻度等の管理	理研サイネス上に構築されており、同システム上で統合的に管理されている。
-----------	-------------------------------------

状況、体制	
②その他、特記事項	

[9 件目] シロイヌナズナ変異体情報

(1-1) データの種類

①生物種	シロイヌナズナ
②試料・ライブラリ 一等の種類、数	理研サイネス上で公開されている 3 件のシロイヌナズナ変異体に関するデータベース (Fox-hunting, DS transposon, Activation tagging line) に含まれる表現型情報に関して、Plant Ontology (PO), PATO といった公開オントロジーを用いた新しい統一的なアノテーションを設定し、公開した。現在理研サイネス上で公開されているシロイヌナズナ変異体データベースの中で明確な記述のある表現型情報については全て新しいアノテーションで関連付けを実現し、統合的検索を可能とした。
③測定方法	
④データの内容	表現型アノテーション (PO・PATO データベースへのリンクを含む) シロイヌナズナ変異体データベースに含まれるフェノーム情報へのリンク
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	表現型アノテーション 115 件 シロイヌナズナ変異体データベース 3 件 (Fox-hunting, DS transposon, Activation tagging line) それぞれに含まれるフェノーム情報への表現型アノテーション情報からのリンク (それぞれ 13142 件, 221 件, 1268 件)
②データ区分	■自前 □第三者 □文献データ □計算結果等の二次データ □その他 (下欄に詳細を記述)
③将来の増加の見込み	
④権利関係	所有者 (独立行政法人理化学研究所) 公開 (■可 □否 □その他 [])
⑤その他、特記事項	http://scines.org/item/ria143i にて公開

	論文2 HATODAS II: heavy-atom database system with potentiality scoring. <i>J. Appl. Cryst.</i> , 42 , 540-544. (2009).
--	---

(2) データ (又はDB) の連結、統合化整備 (※試験的、限定的公開済みのものも含む。)

通番	データ (又はDB) の名称	公開 / 未公開	概要 (データの種類 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
	植物オミックス統合データベース		概要: シロイヌナズナオミックス情報の注釈づけと公開を進めた。 成果: FOX-hunting 変異体データベースを追加統合した。前年度までに統合済みであった DS transposon, Activation tagging line のデータベースを合わせた 3 データベースについて、そのフェノーム情報を統合すべく、PO/PATO などの公開オントロジーを活用した新たなアノテーションを試み、より高度な統合化を試みた。

(3) DB 基盤システム、ツール等開発成果物の整備 (※試験的、限定的公開済みのものも含む。)

通番	DB 基盤システム、ツール等の名称	公開 / 未公開	概要 (主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述)
	理研サイネス		「セマンティックウェブ形式」と呼ばれる国際標準規格に準拠したデータベース統合システムで、以下のような特徴を持つシステムとして開発した。 ①数万個以上の個別データベース構築活動を、大勢の研究者がインターネット経由で並行して実施できる。 ②大規模なデータを介した業務フローを柔軟に設定でき、人的連携や自動処理を容易化できる。 ③各活動群をセキュリティの高い状態で区切り、未公開の状態データベース構築ができる。 ④構築したデータベースをその基盤から直接公開できる。 ⑤公開後も、研究者がシステムの維持コストを負担することなく、その基盤でコンテンツを継続的に更新することができる。 ⑥複数の世界標準形式に準拠したデータ配信が容易。

			同システムは 2009 年 3 月にベータ版が公開され、植物統合 DB 計算機環境として利用されている。今年度は、タンパク質結晶構造解析実験データベースなどの大容量データも安定的に扱えるように、またデータ検索等がスムーズにできるようにシステムの改良が進められた。今後は、システムを運用・改善し、アノテーション等のデータ交換やプログラム交換を理研外部者でも実施できるオープンな参加型のシステムとして公開する。

(4) その他の成果物 ((2)、(3) に該当しないもの)

通 番	名称	公開／ 未公開	概要
	該当なし		