

別紙

「植物オミックス情報および蛋白質構造情報」

19年度 研究成果報告書

平成20年5月

独立行政法人理化学研究所

豊田 哲郎

委託業務成果報告書の無断複製等禁止の標記について

委託業務に係る成果報告書の無断複製等の禁止の標記については、次によるものとする。

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19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。

従って、本報告書の著作権は、文部科学省に帰属しており、本報告書の全文又は一部の無断複製等の行為は、法律で認められたときを除き、著作権の侵害にあたるので、これらの利用行為を行うときは、文部科学省の承認手続きが必要です。

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 19 年度の実施内容

2.1 実施計画

理研の公開データは膨大かつ多岐にわたる為、今回の統合化事業で一度にすべてを統合化することはできないが、今回、シロイヌナズナのオミックスデータとタンパク 3000 プロジェクトの成果データを統合化対象として代表的に選び、本事業への統合化を成功させることで、今後、理研の他のデータベース群を統合化していくためのモデルケースとしての流れを作る必要がある。よって本プロジェクトの意義は、オミックス情報のデータベース統合研究の最初のステップとして情報統合化のプロセスを確立し、植物を含む他分野の研究者が本プロジェクトの結果を参考にしながらオミックス情報を扱えるように基盤指針を提供し、データベース統合化への参加と貢献をより広範囲に広げていくことにある。本プロジェクト自体は、シロイヌナズナオミックス情報注釈、高等動植物等由来蛋白質構造のアノテーション、微生物由来蛋白質構造のキュレーションとアノテーション、アノテーションシステムの開発運用とデータ変換、の 4 つの大きなサブテーマに分割される。下記にサブテーマ毎の実施計画を示す。

シロイヌナズナオミックス情報注釈

平成 19 年度は、アノテーション作業の実施に必要な計算機を理化学研究所の横浜研究所に設置し、シロイヌナズナのトランスクリプトーム、メタボローム、フェノーム、リソースのアノテーション作業を共同で進められるような計算環境を構築する。また、平成 20 年度から本格的にアノテーション作業を開始する準備として、情報・システム研究機構と連携して方針を策定するとともに、アノテーション作業に携わる人材を確保する。

高等動植物等由来蛋白質構造のアノテーション

平成19年度は、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の分子・生理機能情報等のアノテーションの準備、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の準備を行う。

微生物由来蛋白質構造のキュレーションとアノテーション

平成19年度は、蛋白質の試料調製・結晶化・回折実験に関するデータ、変異導入蛋白質の結晶構造に関するデータ、重原子導入蛋白質の結晶構造に関するデータ、の各々について、理化学研究所の播磨研究所に存在する実験データのキュレーションを行い、アノテーション作業に着手する。

アノテーションシステムの開発運用とデータ変換

平成19年度は、アノテーション等のデータを安全に保持できるデータストレージの準備を行い、システム全体の運用体制を整える。データストレージは4年間のデータ編纂作業に十分な容量を準備するとともに、災害対策として理化学研究所の横浜研究所および和光研究所で同じ内容のストレージを設置し同期をとるようにシステムを構築する。

2.2 実施内容(成果)

平成19年度の実施期間は平成19年11月から平成20年3月までの5カ月間であった。下記に「植物オミックス情報および蛋白質構造情報」のサブテーマ毎の成果実績を説明する。

業務の実績

.シロイヌナズナオミックス情報注釈

アノテーションの実作業を行う環境整備には、1) 計算機環境の整備と2) アノテーション作業用の編集ソフトウェアの整備、におおまかに分ける事が出来る。このうち、1) 計算機環境の整備については後述の項目[4]にて詳細を記述する。オミックス情報注釈の実績として、シロイヌナズナのトランスクリプトーム、メタボローム、フェノーム、リソースのアノテーション作業を共同で進めるためのソフトウェアを本事業の目的で使用できる環境を構築した。また、平成20年度から本格的に植物のアノテーション作業を開始する準備として、中核機関である大学共同利用機関法人情報・システム研究機構と連携して方針を策定するとともに、アノテーション作業に携わる人材を確保した。

高等動植物等由来蛋白質構造のアノテーション

高等動植物由来蛋白質の分子・生理機能情報等のアノテーションは各研究者の情報や多種多様なデータベースを参照することにより利用価値を高めることができるが、それらの情報の内容や形式等はさまざまである。また、それに付随する実験データは複数の既存のデータベースや各研究者のローカル環境にあり、試料調製から構造解析の各ステ

ップを関連付けることにより利用価値を高めることができる。さらにタンパク 3000 プロジェクトの実施によって集積されたデータ散逸を防ぐ必要がある。そこで平成19年度は分子・生理機能情報等のアノテーションを効率的に行うために関連データベース調査や論文情報の収集等の予備的な作業を進めた。また実験データについてもデータの集約化を進め、本格的なアキュレーションとアノテーションの作業の開始が可能な体制を整えた。

微生物由来蛋白質構造のキュレーションとアノテーション

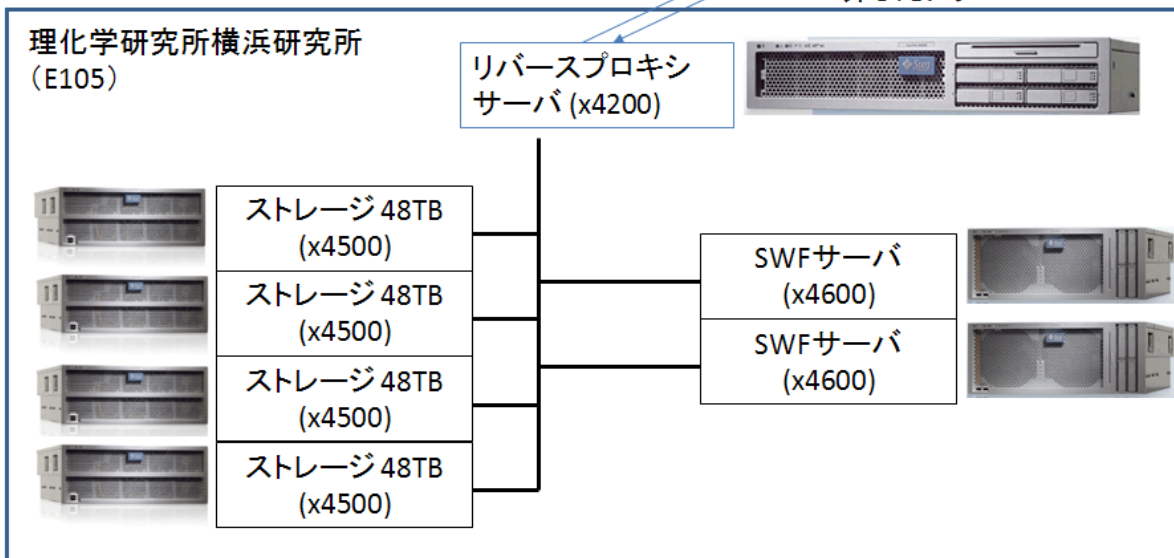
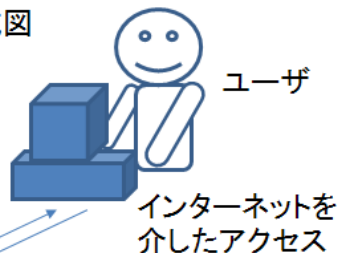
理研播磨研究所において、微生物由来蛋白質の実験データは複数の既存のデータベースや各研究者のローカル環境にあり、さらに情報の形式や保存メディア等も多種多様である。そのため、データを提供するにあたり統一的に扱いにくい。また、個々の既存のデータベース内でも実験の前後の段階（例えば試料調製と結晶化）の間で関連付けができていないものも多く、情報利用の観点から品質が比較的低い。さらにデータの生産元であるタンパク 3000 プロジェクトが既に終了していることから、緊急の課題としてデータ散逸を防ぐ必要がある。そこで平成19年度は膨大なデータを効率的に扱うためのキュレーションシステムを開発し、同システムを用いることでデータの集約化を行った。また変異導入蛋白質の結晶構造に関するデータ、重原子導入蛋白質の結晶構造に関するデータについても回収を行った。アノテーション作業は関連論文の収集等の予備的な作業を進め、本格的なアノテーション作業の開始がいつでも可能な体制を整えた。

アノテーションシステムの開発運用とデータ変換

アノテーション等のデータを安全に保持できるデータストレージの準備を行い、システム全体の運用体制を整えた。データストレージは4年間のデータ編纂作業に十分な容量を準備した。またサービス品質を上げるための工夫として災害時の運用維持対策を考慮し、理研の横浜研究所および和光研究所で同じ内容のストレージを設置し同期をとるようにシステムを構築した。具体的な説明として、下図（追加資料A）に構築したデータベース処理の計算機環境の構成図を示す。データベースの統合処理を行うデータ処理用サーバーマシンと、テラバイトを超える大量データを保持するためのデポジトリ機能としてのディスクマシンを購入し導入を行った。また下図（追加資料B）に構築計算機環境の安全運用対策の結果を示す。メインのデータ管理を行うのは理化学研究所横浜研究所（横浜市鶴見区末広町）で、地震等の災害が起こった場合にはデータベースリソースの供給が途絶えてしまうために、常時サービスの提供が続けられる様にバックアップ機能用のマシン環境を埼玉県和光市にある理化学研究所和光研究所に導入した。

追加資料A-1:データベース処理の計算機環境の構成図

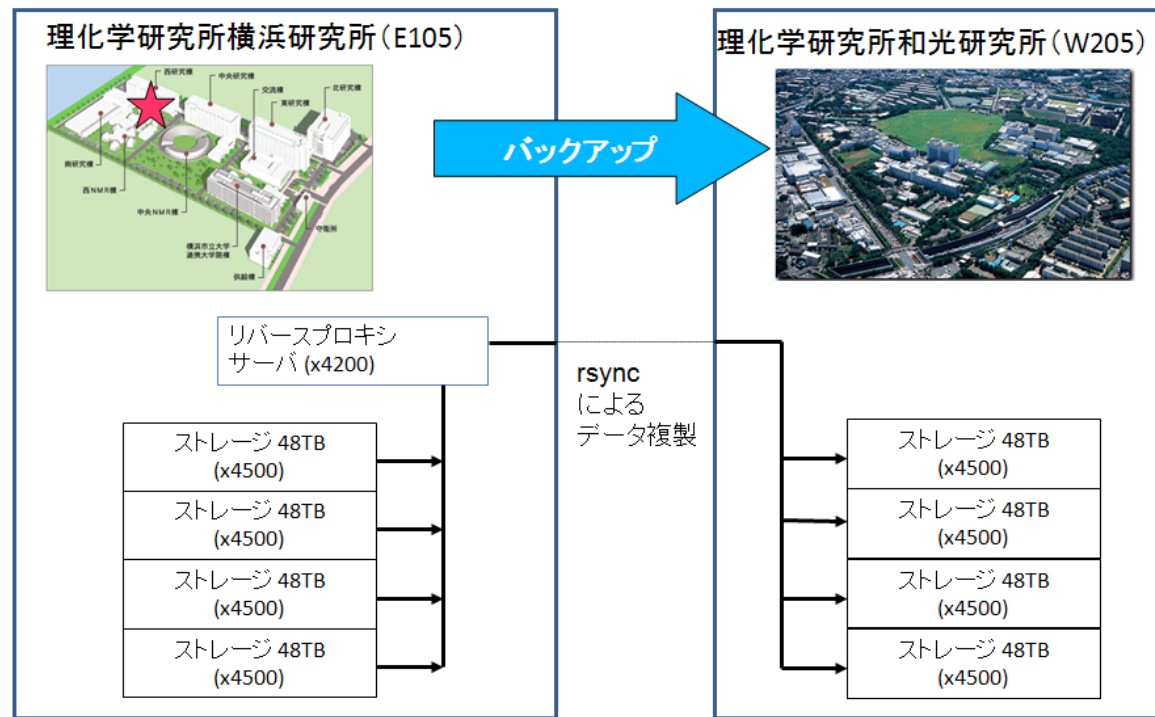
SWFサーバはストレージと連携し、データベースサービスをユーザに提供する。



(Reference: <http://jp.sun.com/products/servers/entry/x4700>)

追加資料B:構築計算機環境の安全運用対策(データバックアップ機能)

- *横浜研究所のデータを和光研究所に複製する
- *横浜研究所でストレージに障害が発生した時は、和光研究所のストレージに切り替えて運用が継続される



2.3 成果の外部への発表

論文寄稿

業務コード	実施年度	和誌 / 洋誌	論文タイトル	発表者名	発表誌名	巻号	ページ	掲載年月	メモ
	19	和誌	コンピュータの中の脳：情報基盤の進化論	豊田哲郎	生体の科学	4	1	20-32	2008年2月

講演

業務コード	実施年度	国内 / 国際	講演タイトル	発表者名	講演会名	発表年月日	メモ
	19	国内	シロイヌナズナのオミックス統合データベース：オミックス進化論とデータベースの役割	豊田哲郎	第49回日本植物生理学会年会	2008年3月20日	
	19	国内	デジタル whole mount in situ hybridization に向けたシロイヌナズナ LucTag ラインの遺伝子発現時系列画像の定量化	神沼 英里、吉積 毅、栗山 朋子、越智子、武藤周、豊田哲郎、松井南、	第49回日本植物生理学会年会	2008年3月22日	
	19	国内	セマンティックウェブ技術による理研データベースの統合化	豊田哲郎	「生命をはかる」研究会第23回研究会	2008年2月18日	
	19	国内	シロイヌナズナ LucTag ラインを用いた遺伝子発現定量解析	神沼 英里、吉積 毅、栗山 朋子、越智子、武藤周、豊田哲郎、松井南、	人工知能学会第36回分子生物情報研究会 (SIG-MBI)	2008年1月11日	
	19	国内	シロイヌナズナのオミックス統合解析とデー	豊田哲郎	第30回日本分子生物学会年	2007年12月	

			データベース		会・第 80 回日本生化学会大会 (BMB2007)	11-14 日	
	19	国内	シロイヌナズナ miRNA の in silico ターゲット探索における特徴パラメータの決定木分析	神沼 英里 松井 章浩, 栗原 志夫, 諸澤 妙子, 関 原明, 豊田 哲郎	第 30 回日本分子生物学会年会・第 80 回日本生化学会大会 (BMB2007)	2007 年 12 月 11-14 日	
	19	国内	シロイヌナズナのオミックス統合解析とデータベース	豊田哲郎	理研シンポジウム「植物トランスクリプトーム解析の新展開:シロイヌナズナワークショップ 2007	2007 年 12 月 10 日	

プレス発表

業務コード	実施年度	発表タイトル	掲載新聞名	掲載日
	19	該当なし		

2.4 活動（運営委員会等の活動等）

本補完課題実施機関である理化学研究所が関係するシンポジウム/ワークショップ開催、運営委員会の開催状況として、下記に4件の活動実績を示す。運営委員会の活動実績として、理化学研究所と中核機関の間でオントロジーや辞書,API, ID, 作業フローの仕様を共通化する合意がなされた。

概要	開催日	活動内容	参加者（敬称略）
運営委員会	2007/11/7	補完課題と中核機関の連携体制の話し合い	研究代表者：豊田 研究分担者：横山, 国島 中核機関関係者：中村, 皿井, 高木, 西川, 永井, 藤枝; 文部科学省：田中
運営委員会	2007/11/16	理研（補完課題）とかずさ（中核の参画機関）の植物統合化に関わる連携について	研究代表者：豊田 研究分担者：篠崎, 関, 花田 中核機関関係者：中村, 中尾, 岡本, 高木, 大久保, 永井, 西川, 川本 文部科学省：田中
運営委員会	2007/2/18	理研補完課題「蛋白質構造情報」のための打合せ	研究代表者：豊田 研究分担者：浅田, 国島, 横山 中核機関関係者：中村, 西川, 大野

シンポジウム	2007/3/20	第 49 回日本植物生理学会（札幌コンベンションセンター）にて、「データベースの中に築く生物像-生命現象を読み解くためのデータベースとWEB リソース-」	研究代表者：豊田 研究分担者：深海、有田 中核機関関係者：中村
--------	-----------	---	---------------------------------------

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、理研ゲノム科学総合研究センター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の中間作業を主に担当する。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する。メタボロームについては理研植物科学研究センターの有田が中心となってアノテーション作業を担当する。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する。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する。タンパク 3000 のデータについては理研ゲノム科学総合研究センター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する。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、必要に応じて倉光・海老原が補助する。また、プロジェクト全期間を通じて可能な限り新たなデータを追加することで、データベースのさらなる充実を図る。また、中核機関と連携して、データ変換ツールやインターフェース等を作成する。

平成 19 年度に於ける実施体制

研究項目	担当機関等	研究担当者
サブテーマ シロイヌナズナオミックス 情報注釈	（独）理化学研究所ゲノム科学総合研究センター チームリーダー	豊田哲郎
	（独）理化学研究所植物科学研究センター チームリーダー	関原明
	（独）理化学研究所植物科学研究センター ユニットリーダー	櫻井哲也
	（独）理化学研究所植物科学研究センター 研究員	花田耕介
	（独）理化学研究所植物科学研究	有田正規

	センター 副グループディレクター	
	(独)理化学研究所植物科学研究 センター グループディレクター	松井南
	(独)理化学研究所植物科学研究 センター 研究員	黒森崇
	(独)理化学研究所ゲノム科学総 合研究センター 技師	平田直彦
	(独)理化学研究所ゲノム科学総 合研究センター テクニカルスタッフ	松下さとみ
サブテーマ 高等動植物等由来蛋白質構 造のアノテーション	(独)理化学研究所ゲノム科学総 合研究センター プロジェクトディレクター	横山茂之
	(独)理化学研究所ゲノム科学総 合研究センター シニアテクニカルサイエンティ スト	中山朋子
	(独)理化学研究所ゲノム科学総 合研究センター 上級研究員	寺田貴帆
	(独)理化学研究所ゲノム科学総 合研究センター リサーチアソシエイト	吉良 聡
	(独)理化学研究所ゲノム科学総 合研究センター リサーチアソシエイト	丹羽 英明
	(独)理化学研究所ゲノム科学総 合研究センター 技師	熊谷 紀和
サブテーマ 微生物由来蛋白質構造のキ ュレーションとアノテーシ ョン	(独)理化学研究所播磨研究所 放射光科学総合研究センター 上級研究員	国島 直樹
	(独)理化学研究所播磨研究所 放射光科学総合研究センター 研究員	菅原 道泰
	(独)理化学研究所播磨研究所 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
	(独)理化学研究所播磨研究所 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
	(独)理化学研究所播磨研究所 放射光科学総合研究センター チームリーダー	海老原 章郎
サブテーマ アノテーションシステムの 開発運用とデータ変換	(独)理化学研究所ゲノム科学総 合研究センター チームリーダー	豊田哲郎
	(独)理化学研究所ゲノム科学総 合研究センター	小西史一

	研究員	
--	-----	--