

委託業務の題目  
植物オミックス情報および蛋白質構造情報  
受託機関名  
独立行政法人理化学研究所

### 1. 委託事業の9月末時点の判断基準になる目標

(サブテーマ①) 文献情報から収集したシロイヌナズナ変異体の遺伝子および表現型の変異情報に基づき、表現型をオントロジーを用いて整理すると同時にそれぞれの表現型変異に対応する遺伝子 (TAIR の AGI コード) を列挙する。9月末までに 400 種類の表現型を登録し、330 の遺伝子コードへの対応付けを行い、これらすべてを理研サイネスを通じて提供する。

(サブテーマ②) 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タンパク質構造データのうち、累計 3 万 1 千件の回折実験データ (画像データ枚数) を公開し、それに関わる実験情報についても関連づけを完了させ、9月末までに可能な約 50 万件の結晶観察データについても提供する。

(サブテーマ③) Bacpedia のオリジナル DB (実験データを有するタンパク質約 8700 種類) を公開し、一般の研究者にとって使いやすい環境を整える。

(サブテーマ④) SciNeS 上の検索システムの改良作業を進め、9月末までに植物関係の 35 のデータベースについて横断的な関連情報の検索機能を提供する。

### ・ 2. 9月末時点の達成状況

(サブテーマ①) シロイヌナズナミュータントの表現型情報を扱った論文 266 件を対象に、マニュアルキュレーションを実施し、表現型変異・その変異を観察した実験条件・関与が考えられる遺伝子などの情報を抽出した。これに基づき、438 種類の表現型をオントロジーを用いた標準化し、370 種類の遺伝子 (AGI コード) との関連付けを行った。これらの結果をデータベース化し、サイネスを通じて公開した。

(サブテーマ②) 上記 1. の目標を達成した。

- ・ 2009 年 5 月：回折実験データ 2 万件と付随する実験データを理研サイネス上に公開した。
- ・ 2010 年 8 月：回折画像データ 1 万 1 千件と付随する実験データを理研サイネス上に公開した。
- ・ 2010 年 9 月：約 50 万件の結晶観察データと付随する実験データを提供した。

(サブテーマ③) Bacpedia のオリジナル DB (実験データを有するタンパク質約 8700 種類) を公開し、一般の研究者にとって使いやすい環境を整えた。

(サブテーマ④)

- ・ 植物統合データベース (<http://scinets.org/db/plant>) を開設し、この下に植物関係のデータベースを統合化した。
- ・ SciNetS 上でセマンティックリンクにより関連付けられた、遺伝子や表現型などのバイオデータアイテムを実用的速度で横断検索する Web 準拠のサービスを実現させた。

### 3. 上記達成状況を踏まえたプロジェクト終了までの目標

(サブテーマ①)

- ・ 未処理の文献のキュレーションを進める。年度末までに 270 件ほどの文献を処理し、環境応答性や種子生産性に関する表現型情報を中心に標準化とデータベース化を進める。

・シロイヌナズナ変異体データベースなどのDBと上記マニュアルキュレーションの成果との統合化をすすめる。植物オミックス研究者の意見を聞きながら、実験データ・文献データを問わずフェノーム情報から各種オミックス情報の検索が容易にできるようにシステムを整備する。

(サブテーマ②) 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タンパク質構造データ (X線による結晶構造解析) のうち、3万1千件の回折実験データ (画像データ枚数) を公開し、それに関わる実験情報についても関連づけを完了させ、可能なものについては公開する。

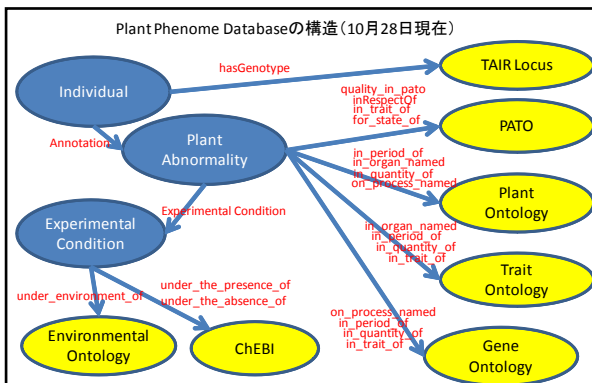
(サブテーマ③) 理研播磨研究所における微生物由来の蛋白質構造データに付随する実験データに関し、統合データベースの一部として生命科学分野の幅広い研究者に役立つ体制を確立する。

(サブテーマ④) シロイヌナズナで実施した統合化をモデルケースとして、表現型情報の標準化・上位オントロジーの整備を進め、シロイヌナズナ以外の生物種 (マウス等) をも含めたデータベース統合化を実現する。

#### 4. 成果の概要

(サブテーマ①)

- ・植物表現型統合データベースをサイネス上に構築し、<http://scinets.org/item/cria224uli> から共有化した。
- ・表現型の標準化は6種類の公開オントロジーへの意味リンクとして表現した。このDB構造は他のDBでも共通に使えるよう考慮して設計した。

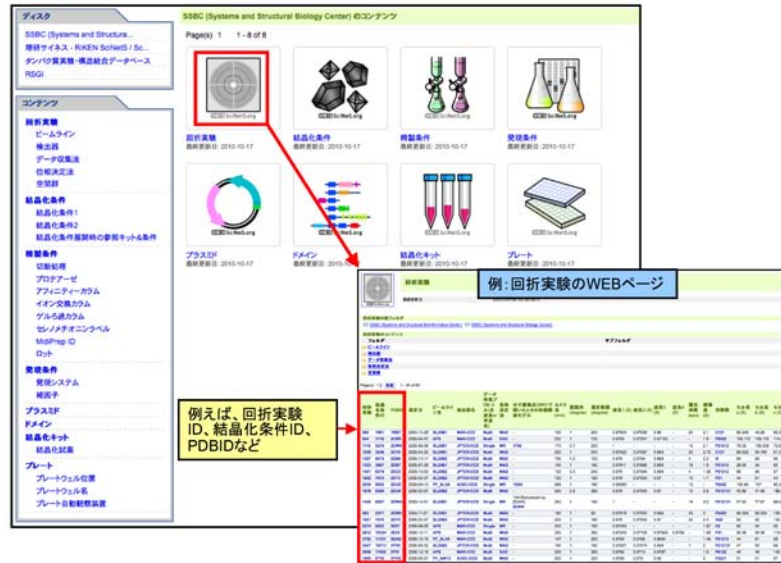


統合データベースプロジェクト作業部会分科会資料

(サブテーマ②) 理研サイネスの下記 URL に回折実験データ累計約 3 万 1 千件と付随する実験データを公開した。(結晶観察データ約 50 万件をテストサーバで編集中。)

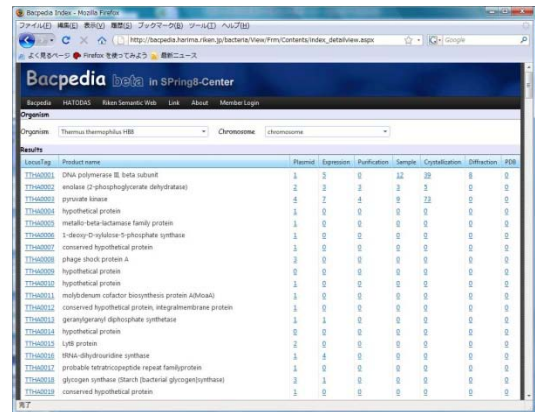
URL:

<https://database.riken.jp/sw/links/ja/ria46i/>



(サブテーマ③) Bacpedia オリジナル DB の公開画面 (2010 年 7 月 6 日公開, 現在まで 4280 ページビュー)

<http://bacpedia.harima.riken.jp>



(サブテーマ④) 各種オミックス間の関係を定義すべく新規に開発された「上位オントロジー」に準拠して、対象データベースを分類し直した。トップページに示すメニューにはこの分類体系を反映させている。

・トップページから遺伝子を検索する場合、遺伝子データにキーワードが含まれない場合であっても遺伝子に関連付けられているキーワードを含む表現型データを見出し、当該遺伝子を推論し検索することができるようになった。また検索速度については、データアクセス権のチェックも含めたいての場合高々1秒で結果が表示されるまで高速化が実現できた。さらには、JSONP を介したキーワードを含むデータアイテム数を返すインターフェイスを構築し、ライフサイエンス統合データベースの横断検索機能との連携が実現された。

