

平成 23 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業「統合化推進プログラム」
平成 23 年度採択 研究代表者

田畑哲之
財団法人かずさディー・エヌ・エー研究所 副所長

ゲノム情報に基づく植物データベースの統合

§1. 研究実施体制

(1) かずさグループ

- ① 研究代表者: 田畑 哲之 (かずさ DNA 研究所、副所長)
- ② 研究項目
 - ・植物 DNA マーカー、連鎖地図、リソースおよび関連情報の統合

(2) 新潟大グループ

- ① 主たる共同研究者: 中谷 明弘 (新潟大学研究推進機構超域学術院、准教授)
- ② 研究項目
 - ・遺伝子オルソログ DB の構築と植物ゲノム DB の統合

§2. 研究実施内容

本研究では、国内に散在する植物データベース(以下、DB)の統合を行い利便性の向上を図ることを目的として、以下の4項目の研究開発を実施した。

(1) 遺伝子オルソログDBの構築とそれに基づく植物ゲノムDBの統合

現在モデル植物や作物についてのゲノム塩基配列の解読が精力的に進められており、それに伴い遺伝子配列も蓄積しているが、植物を主な対象としたオルソログ関係の整理は十分になされていない状況にある。そこで、植物についてのゲノム関連 DB(以下、植物ゲノム DB)のを遺伝子レベルで統合するために、生物種間の進化系統関係を反映させた階層的な遺伝子オルソログ DBを作成し、遺伝子間のオルソログ関係によって植物ゲノムDB群のエントリを相互リンクする。これによって、項目の細分化が進んでいるDBの全体像を、遺伝子IDをタグとして統一かつ俯瞰的に探索可能にする。当初の統合対象DBは、整備が進んでいるモデル植物・作物のゲノムDB(研究代表者グループ保有)や、シロイヌナズナ、イネ、ヒメツリガネゴケ、ゼニゴケ等のゲノムDB(各関連機関保有)とし、状況に応じて順次拡大する。平成23年度は、階層的なオルソログDBの仕様(内部的なデータ形式)を決定し、遺伝子のアミノ酸配列間の総当たりのBLAST検索とクラスタリングによって階層的なオルソログ情報の生成を行い、開発用のウェブサイトを作成した(<http://gene.bri.niigata-u.ac.jp/OD/>)。一連の生成処理はアミノ酸配列と生物種間の系統関係の情報を与えると自動的に実行できる。統合対象DBのエントリは、遺伝子名や遺伝子ID(例えばGenBank ID)の同一性、アミノ酸配列の類似性を用いてオルソログDBに連結することによって、生物種あるいは系統群を跨がって関連している遺伝子群(オルソログ)を検索することが可能になる。今後は、開発した階層的なオルソログDBをハブとした植物ゲノムDBの統合部分の処理を実装する。検索GUI/APIを実装して課題(4)で開発中のポータルサイト等との連携を図り、セマンティックウェブ(RDF)化にも対応する。また、オルソログDBとオントロジー情報との対応付けを検討する。

(2) DNA マーカーおよび連鎖地図情報に基づく植物ゲノムDBの統合

本研究項目では、ゲノム塩基配列の解読や国内でDNAマーカーが開発されている植物を対象として、DNAマーカーやQTL、連鎖地図関連のDBを調査し、DBに記載されていないこれらの情報を文献や総説から収集する。さらには、収集したDNAマーカーやQTLの位置をゲノム塩基配列や遺伝的地図上において特定するシステムを構築する。平成23年度はNCBIやGOLD等のゲノムデータベースを調査することにより、55の植物種を本研究課題の対象として選定した。これらのうちトマトについて、主にSSLPとSNPマーカーのプライマー配列やQTL情報を文献から抽出した。さらに、ジャガイモやナス等のナス科植物やイネ科植物に関するマーカー情報を収集した。また、DNAマーカー付随配列をゲノム塩基配列上にマッピングすることで、ゲノム塩基配列と遺伝的地図間におけるマーカーの対応関係を特定するシステムを構築した。なお、イネ科やマメ科等の主要作物に関しては、文献の量が膨大であり国際的標準のDBが既に存在しているため、リンクに留めることにした。今後は、植物種をさらに拡大し情報の収集を加速させ、DNAマーカーの位置を特定するシステムの充実を図る。

(3) 植物リソースDBの統合

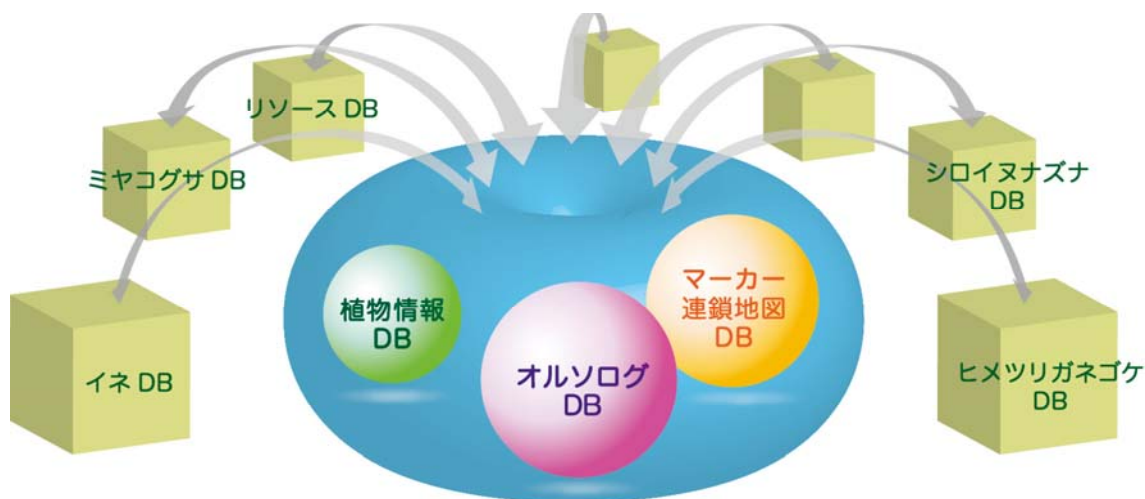
本研究項目では、国内に散在する植物バイオリソース情報を統合的に検索可能にするこ

て、利便性を高めることをめざす。平成 23 年度は、文部科学省の植物バイオリソースが集約されている二大センターとして、ナショナルバイオリソースプロジェクトDB(以下 NBRP)ならびに理研バイオリソースセンター(以下 BRC)が提供する情報の調査と横断検索のための開発を実施した。両者に於いては既に NBRP サイトにて BRC リソースの検索が可能となっているが、調査の結果、リファレンスゲノムに遺伝子 ID のバージョン情報が明示されていない場合があること、リファレンスゲノムの指定が適切でなく消失している場合がある等の問題点があった。そこで、これらの問題を解消し、リソース情報とゲノム情報の対応状況を改善した。また BRC から NBRP 由来情報を統合的に検索できるよう情報整備の支援を実施した。さらに、BRC 提供の SABRE データベース上への NBRP 情報の追加と REST API による統合機能を追加した。今後は、引き続き主要リソース DB のゲノム情報との連携の改善をすすめるとともに、それ以外の国内の主要なバイオリソースの調査を行い、集約されていない植物リソースの情報を収集し検索可能にする。

(4) 植物研究に関連する情報基盤の構築

本課題では、ゲノム塩基配列の解読や国内で DNA マーカーが開発されている植物について、植物学や農学、園芸学、育種学上の特徴、有用性、ゲノム塩基配列の決定方法、解析手法等の学術情報を文献(論文や辞典、総説)から収集する。また、ゲノムやトランスクリプトーム、メタボローム、インタラクトーム等のオミックス研究に関する国内外の DB を調査し情報を収集する。平成 23 年度は、課題(2)で選定した 55 の植物種を対象とし、学術的諸情報の収集を完了した。また、国内外における 1,093 件の植物のオミックス関連の DB を調査しリンク先情報を収集した。さらに、DB のリンク切れを自動的に検知するシステムを組み込んだ。今後は、収集した 1,093 件のオミックス関連の DB を整理・分類しポータルサイトに掲載する。また、DB の調査は継続しリンク情報を更新する。

上記課題(1)、(2)、(4)については、オルソログに関する計算結果や収集した諸情報を格納するための DB を構築した。また、ポータルサイトを立ち上げ、課題(2)と(4)により得られたマーカー情報やマーカー表示システム、植物の学術的情報といったコンテンツを掲載した。今後は、コンテンツがある程度揃った時点(平成 24 年 8 月を予定)でポータルサイトを一般に公開し、ユーザーからの意見を取り入れることでコンテンツの充実を図る。最終的には、本研究開発の 4 課題をポータルサイトに掲載することで情報の連携を図り、横断的な検索機能を付与させる(図参照)。



§3. 成果発表等

(3-1) データベースおよびウェブツールの構築と公開

① 公開中のデータベース・ウェブツール等

なし

(3-2) 原著論文発表

なし