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 26 年度採択)
課題名「植物ゲノム情報活用のための統合研究基盤の構築」(研究代表者　田畑　哲之)

植物ゲノム情報の活用基盤の構築

田畑　哲之
(かずさ DNA 研究所)

わが国では、モデル実験植物や実用植物を含む多種多様な植物種を対象に遺伝学研究やゲノム研究が精力的に行われている。これらの研究成果であるゲノム・遺伝子塩基配列、遺伝子発現、代謝産物などの情報は論文とデータベース (DB) を通して公開されているが、個々の研究グループが独自に DB を立ち上げた結果、異なるプラットフォームによる DB が散在し利便性を欠くことが指摘されている。ユーザが必要とする情報を迅速に入手するには、散在する様々な植物ゲノム関連情報が有機的に格納された情報基盤を構築することが必要である。我々は、植物ゲノム関連情報を統合する Plant Genome DataBase Japan (PGDBj、http://pgdbj.jp) を開発し、その拡充を進めている。

PGDBj のコアともいえる遺伝子オルソログ DB では、これまでに NCBI RefSeq DB から収集した緑色植物 20 種の 50 万配列とラン藻 111 種の 50 万配列を用いてオルソログ情報を生成し、アミノ酸配列の類似性や種間系統関係とオルソログ情報を一体化した DB を構築した。更に、オルソログ DB 内の遺伝子と外部の植物ゲノム DB 内の遺伝子をアミノ酸配列レベルでリンクすることにより、オルソログ DB をハブとする外部 DB への検索が可能となった。

実用性の高いゲノム情報である DNA マーカーについても情報の拡充を図っている。これ

までに、30 種の植物について外部 DB と論文から約 9 万件の DNA マーカー、約 1800 件の QTL 情報を収集、公開した。今後は植物種を追加し、病害抵抗性や環境ストレス耐性など農業上重要な形質に関連する DNA マーカーをゲノム配列上に位置付け、配列情報に基づく種間比較などを行うシステムを構築する予定である。

植物リソース DB では、理研 BRC の SABRE システムを拡充し、BRC と NBRP が保有する 14 種 150 万件のバイオリソース情報を PGDBj 横断検索システムから検索することができる。カンキツ等新たな植物種リソースの追加や、配列情報との関連付けによるオルソログ DB との連携を計画している。

PGDBj 全般については、今後セマンティックウェブ（RDF）化を進め、PGDBj の内部 DB ならびに統合化推進プログラムの他課題との一層の連携を図る。そのために必要な植物関連オントロジーの整備や、ゲノム情報アノテーションの高度化も進めてゆきたい。