

論文の内容の要旨

論文題目 植物の完全長 cDNA 収集および DNA 多型探索によるゲノム研究

氏 名 櫻井 哲也

ゲノムという言葉は、遺伝子 (gene) と集団、総体を意味するオーム (-ome) を合わせた造語であり、1920 年にドイツのハンブルク大学の植物学者 Hans Winkler 博士が定義した。当時は、卵と精子に含まれる染色体の 1 組という意味で使われていたが、DNA という分子が遺伝情報の本体であることが分かった今日では、生物が持つ全遺伝子セットという意味に再定義されている。1990 年代に入り、配列決定の自動化技術が確立するとゲノム科学は急激に発展し、1990 年代半ばには、様々な生物のゲノム配列決定プロジェクトが進行した。ゲノム研究において、転写領域などの予測を行うゲノム注釈の作成は、その遺伝情報全体を理解するためのゲノム解析の基本であり、高品質であることが要求される。シロイヌナズナは、1970 年代から遺伝学実験の対象として盛んに使われ、植物研究において最も幅広く使用されている生物種の 1 つであるが、顕花植物として初めて全ゲノム塩基配列が解読されたことで、ゲノム研究においても、モデル植物としての地位を確立した。シロイヌナズナの包括的研究は盛んに展開され、cDNA および発現配列タグ(expressed sequence tag; EST)の収集により、遺伝子構造や遺伝子機能の解析が進んだ。しかし、収集された cDNA は必ずしも完全長ではなく、5'末端側のタンパク質コード領域などの情報が欠くため、転写領域の予測などのゲノム注釈情報に不正確な情報が含まれることが懸念されていた。また、生産された情報は、インターネットを通じて閲覧でき

るオンラインデータベースとして編纂されていたが、cDNA などの生物遺伝資源への関連付けなどに改善の余地がある。樹木や作物など他の有用植物のゲノム研究においても、ゲノム注釈情報は量、質ともに十分ではなく、機能ゲノム研究を推進するためには完全長 cDNA などの高品質な研究資源を整備し、モデル植物シロイヌナズナの研究成果の活用による効果的な比較ゲノム解析の実施が求められている。集積されたゲノム情報は、生物種内における比較ゲノム解析を可能にし、同定された DNA 多型は分子マーカーとして育種研究の推進に貢献するだけでなく、DNA 多型と遺伝子機能との関係性の理解にも有用な情報となる。

以上の背景をもとに、完全長 cDNA を活用したモデル植物シロイヌナズナ、モデル樹木ポプラとデンプン作物キャッサバのゲノム注釈改善による植物のゲノム研究基盤整備を推進し、各植物の遺伝子機能概観の把握および系統間 DNA 多型の同定、DNA 多型と遺伝子機能との関係性についての解析を行った。さらに、これらの研究成果についてのデータベースを構築し、情報基盤整備を行った。

第 1 章緒論に続き第 2 章では、植物の完全長 cDNA 解析として、(1)シロイヌナズナのゲノム注釈情報の改善とデータベース RARGE の構築、(2)ポプラ完全長 cDNA の収集とその配列解析、(3)キャッサバ完全長 cDNA の収集とその配列解析の 3 点について言及した。

(1)シロイヌナズナのゲノム注釈情報の改善とデータベース RARGE の構築においては、シロイヌナズナの完全長 cDNA に由来する約 30 万配列を精査することにより、シロイヌナズナの新規転写領域 503 ヶ所を同定、および 4,119 個の遺伝子についての新規遺伝子モデルを検出し、ゲノム注釈情報の改善を図った。同様に、理化学研究所の研究チームが作製したシロイヌナズナ *Ds* トランスポゾンタグ挿入変異体 17,671 系統についての *Ds* トランスポゾンタグの近傍ゲノム塩基配列を獲得し、この配列データをゲノム塩基配列へマップすることで、ゲノム上における *Ds* トランスポゾンタグ挿入位置を計算し、関連遺伝子とタグ挿入部位を同定した。本解析で得られたシロイヌナズナ完全長 cDNA と *Ds* トランスポゾンタグ挿入変異体の情報は、データベース RARGE(<http://rarge.psc.riken.jp/>)として編纂し、インターネット上へ公開した。

(2)ポプラ完全長 cDNA の収集とその配列解析においては、ポプラ(*Populus nigra*)の花芽、葉芽、根および各種ストレス処理を施した葉を出発材料とし、ビオチン化キャプトラッパー法を用いて完全長 cDNA ライブラリを作製した。作製した完全長 cDNA ライブラリから単離した各 cDNA の両末端配列を決定し、89,572 配列(5'末端読み:46,000、3'末端読み配列:43,572 配列、47,137cDNA クロー

ン)を得た。獲得した cDNA の完全長率は 86%であり、通常の cDNA 収集よりも高品質であることを客観的に示した。獲得した cDNA 配列のアセンブル結果から、17,838 個の遺伝子を収集したことを推定した。cDNA の重複が 3 つ以下である遺伝子が 80%以上を占めることから、重複が少ない、効率的な cDNA 収集であることが確認された。本研究で収集したポプラ完全長 cDNA 配列を精査することにより、ポプラの新規転写領域 944 ヶ所を同定、および 1,087 個の遺伝子についての新規遺伝子モデルを検出し、ゲノム注釈情報の改善を図った。同様に、(3)キャッサバ完全長 cDNA の収集とその配列解析においても、キャッサバ(*Manihot esculenta* 系統 MTAI16)の各種ストレス処理を施した植物体の葉と根を出発材料とし、ビオチン化キャップトラッパー法を用いて完全長 cDNA ライブラリを作製した。作製した完全長 cDNA ライブラリから単離した各 cDNA の両末端配列を決定し、35,400 配列(5'末端読み:18,790、3'末端読み配列:16,610 配列、19,450cDNA クローン)を得た。獲得した cDNA の完全長率は 84%であり、通常の cDNA 収集よりも高品質であることを客観的に示した。獲得した cDNA 配列のアセンブル結果から、10,363 個の遺伝子を収集したことを推定した。ポプラの完全長 cDNA ライブラリ作製時と同様に、cDNA の重複が 3 つ以下である遺伝子が 80%以上を占めることから、重複が少ない、効率的な cDNA 収集であることが確認された。本研究で収集したキャッサバ完全長 cDNA 配列を精査することにより、キャッサバの新規転写領域 974 ヶ所を同定、および 751 個の遺伝子についての新規遺伝子モデルを検出し、ゲノム注釈情報の改善を図った。ポプラ、キャッサバいずれにおいても、ゲノム配列由来の予測遺伝子配列へ新たに同定した転写領域のスキヤフォールド配列を加えたものを全転写領域配列とみなし、シロイヌナズナとの比較解析を行い、各々の植物種の遺伝子機能概観を得た。

第 3 章では、ゲノム情報を活用した有用作物キャッサバの DNA 多型の探索と DNA 多型と遺伝子機能の関係性についての解析を行った。第 2 章で述べたキャッサバ完全長 cDNA の収集により、キャッサバの EST 数は増大し、キャッサバ系統間における比較ゲノム解析が可能になった。17 のキャッサバ系統間における DNA 多型探索を行った結果、10,546 ヶ所の SNP と 674 ヶ所の InDel を同定し、各 DNA 多型周辺のゲノム DNA 増幅用のプライマーペア配列を設計した。同定した SNP の内、タンパク質コード領域に位置するものについて、SNP により非同義塩基置換/同義塩基置換が生じるタンパク質の機能を解析した結果、ユビキチンファミリーや ATP 合成酵素などの遺伝子では、顕著に非同義塩基置換比が低く、対照的に、NB-ARC ドメインやロイシンリッチリピートを含む遺伝子は、

非同義塩基置換比が顕著に高かった。この解析から、SNP の遺伝子多様化への関与が示唆された。また、SNP による読み過ごし置換と遺伝子機能との関係性を解析した結果、「生物/非生物的刺激」「ストレス応答性遺伝子」「シグナル伝達」において、読み過ごし置換を生じる遺伝子群で有意に高い割合を示した。この結果は、環境応答性遺伝子で読み過ごし置換の保存性が高いことを示し、変異によるタンパク質長の延長がこれらの新規機能獲得へ関与することを示唆した。これらの解析で得られた結果をデータベース Cassava Online Archive(<http://cassava.psc.riken.jp/>)として構築し、インターネット上に公開した。

以上のように本研究では、モデル植物シロイヌナズナのゲノム注釈を改善し、その解析手法を活用して完全長 cDNA 収集による樹木ポプラとデンプン作物キャッサバのゲノム注釈改善を行った。さらに、キャッサバの系統間 DNA 多型の同定と分子マーカー開発、および SNP と遺伝子機能との関係性についての解析を行った。また、本研究で得られた成果をデータベースとして編纂し、インターネット上に公開した。本研究で得られた知見と構築されたデータベースは、今後の植物ゲノム研究や分子育種の推進に役立つと期待される。