

Focus	Generation of high quality genome sequence of Gifu accession of <i>Lotus japonicus</i> to accelerate NBRP resource application
PI	Shusei Sato Graduate School of Life Sciences, Tohoku University
Period	FY2015
Overview	<p><i>Lotus japonicus</i> has been widely used as a model system to investigate the genetic background of legume-specific phenomena. The genome sequences of <i>L. japonicus</i> have been analyzed using the earliest flowering accession, MG-20, as a reference. In the previous program, we carried out re-sequencing of two types of RILs collected by NBRP program. By applying the information on high density SNP markers and high resolution genetic map obtained based on RILs re-sequencing data, improvement of MG-20 genome sequence information has been carried out.</p> <p>With the aim to establish the genome information on the accession Gifu (B-129), the other widely used experimental accession of , we are going to accumulate the Gifu genomic sequences by using PacBio RS II sequencer and the end sequences of BAC clones constructed from Gifu genomic DNA by applying 3D-pooling system. By taking the advantages of the long reads from PacBio platform and high density anchoring marker information based on RILs re-sequencing along with newly analyzed BAC end sequences, we will be able to generate high quality genome sequence information that will accelerate the application of information and material resources in NBRP.</p> <div data-bbox="572 835 1141 1395" data-label="Image"> </div> <p style="text-align: center;"><i>Lotus japonicus</i> accession Gifu</p>
Progress	<p>(written in Japanese)</p> <p>【概要】</p> <p>NBRP の第 1 期から収集対象とされてきたミヤコグサについて、実験系統として広く利用されている Gifu 系統のゲノム情報を整備し NBRP リソースの活用を図ることを目的として、第三世代の PacBio RSII シーケンサーを用いた配列情報収集を行うとともに、Gifu 系統 BAC ライブラリの末端配列を収集し、平成 26 年度の本事業で収集した高密度の SNP マーカー情報と合わせて高精度のゲノム情報の構築を行なった。デンマーク、オーストラリアのグループと共同でミヤコグサ Gifu 系統の PacBio シーケンサーを用い、最終的に 100 倍のゲノムカバーの配列情報を収集した。これらの配列情報を基に、大規模データに対応できるように調整した新規アセンブルプログラムを用いてアセンブルを行なった結果、最終的に N50 が 810 kb の 1685 コンティグにアセンブルすることができた。これらのコンティグを BAC ライブラリの末端配列情報や高密度の SNP マーカー情報を用いて整理化することにより、サイズベースで全コンティグの 99.4%に相当する 1561 コンティグで構成される pseudomolecules を構築することができた。これらの pseudomolecules にはテロメア配列や反復配列リッチなペリセントロメア領域の配列も含まれており、当初計画通りの高精度ゲノム配列情報を得ることができた。</p> <p>【詳細】</p> <p>(1) PacBio シーケンサーを用いた配列情報収集</p>

計画通りデンマーク、オーフス大のグループと共同でミヤコグサ Gifu 系統の PacBio シーケンサーを用いた配列情報収集を行った。アセンブル時の混乱を避けるため、SSR マーカーによる遺伝的背景を確認した Gifu 系統の同一個体から調整した DNA サンプルを用いて双方で配列解析を実施した。本プログラムでの PacBio シーケンサーによる配列解析は、20 kb ライブラリを用いて 10 SMART cells で実施し、平均リード長 10.7 kb, 総塩基数 12.4 Gb のデータが得られ、当初目標を超えるゲノムの 28 倍に相当する配列情報が収集できた。

(2) Gifu 系統 BAC ライブラリの末端配列解

NBRP ミヤコグサ・ダイズ中核機関からレプリカの提供を受けた全 96 プレーットの 3D プールを作製した。各プール DNA サンプルから、BAC クローンの末端配列領域のみを増幅して作製したシーケンスライブラリを MiSeq シーケンサーで解析し配列情報を収集した。各プールから得られた配列情報を、リファレンスゲノム配列にマップし、プール間で共通するマップ位置の情報からクローンを特定する作業を行った。情報を整理して得られた Gifu 系統 BAC ライブラリの末端配列情報は、アセンブルの精度確認作業に応用するとともに、リソースの利用促進に役立てるために NBRP ミヤコグサ・ダイズの web サイト、[LegumeBase](#) から公開する準備を進めている。

(3) Gifu 系統のゲノムアセンブリ

デンマーク、オーフス大のグループで最終的に 75 倍の PacBio シーケンサーを用いた Gifu 系統のゲノム配列情報が収集できたため、日本で取得した 28 倍の配列情報と合わせて目標の 100 倍を超える配列情報を収集することができた。これらの配列情報を基に、大規模データに対応できるように調整した新規アセンブルプログラムを用いてアセンブル作業を進めている。データ量が多いためプログラムの実施に試行錯誤している状況であるが、現時点で、最長コンティグのサイズが 4 Gb、N50 が 700 kb のアセンブル情報が得られている。アセンブルプログラムのパラメータと解析データ量の調整を行うことにより、コンティグ状況の向上を図るとともに、マーカー情報、末端配列情報との対応の情報解析を行い、高精度ゲノム情報の構成を完成させる計画である。