

【話 題】

International Plant & Animal Genome XXIV の参加報告

三嶋 賢太郎^{*,1}・平尾 知士¹・高橋 誠¹

大会の概要

2016年1月9～13日、アメリカ合衆国カリフォルニア州サンディエゴ市の Town and Country Hotel において、Plant & Animal Genome Conference XXIV (植物及び動物ゲノムに関する学会) が開催された。本大会では、植物及び動物を対象としたゲノム研究における最新の研究成果と育種への利用などが報告された。参加人数は約4,000人であり、173件のワークショップ(うち、企業のワークショップ26件)、1,293件のポスター発表、120社の企業による研究機材や試薬などの展示が行われた。著者らは Forest Tree (林木) ワークショップでポスター発表を行った。大会のすべては網羅できないが、本報告ではいくつかのテーマについて報告させて頂く。

Forest Tree ワークショップ

Forest Tree ワークショップでは、18件の口頭発表が行われ、ヒロハヤマナラシ (*Populus deltoides*)、ユーカリ (*Eucalyptus globulus* と *E. glandis*) をはじめとした広葉樹種及びダグラスファー (*Pseudotsuga menziesii*) やシュガーパイン (*Pinus lambertiana*) などの針葉樹種に関するゲノム研究の報告が行われた。一連の研究発表を通し、次世代シーケンサーを利用した大規模な遺伝子発現解析や遺伝子型の決定が行われており、それらの大規模な情報からいかに有益な情報を引き出すのかといった情報処理 (バイオインフォマティクス) に関する研究発表が行われた。以下に本ワークショップで行われた発表のいくつかを報告する。

E. grandis では、RNA の分解やエピジェネティックな改変を通し、遺伝子の発現を制御する低分子非コード RNA (small noncoding RNA ; sRNAs) の研究が行われていた。microRNA (miRNA) は sRNA の一つであ

り、保存性が高く維管束組織分化で重要な役割を担うことが知られている。一方で、低分子干渉 RNA (small interfering RNA ; siRNAs) は複雑かつ多様化しており、非常に多面的な sRNA のタイプであり、種特異的であるとされている。*E. grandis* では木部形成とリグノセルロース系生体高分子合成経路を制御する転写ネットワークと相互作用を解明するために、葉、二次師部、未成熟木部における sRNAs の包括的カタログ化と定量化を行っていた。これまでに135個の miRNA と 11,888個の siRNA を単離し、これらの sRNAs の発現解析により、維管束や葉の組織において組織特異的にクラスターを形成することを報告していた。

ダグラスファーでは針葉樹における季節刺激と季節反応を理解するために、昼間及び概日スケールで次世代シーケンサーを使った針葉の遺伝子発現解析を行っていた。異なる気候帯の3ヶ所に植栽された19個体から、4時間おきに2日間のサンプリングを行い、次世代シーケンサーによる mRNA のシーケンスが行われ、162,326個の転写産物が得られていた。これをリファレンスにして、さらに日周性、季節性、年周性の変異を検出するために166サンプルについて RNA-seq を行っていた。結果として、15,487個の転写産物は日周性を刻み、中でも4,912個の転写産物は高い振幅を示し、モデル植物で単離されているコアな時計遺伝子群と相同性を示した。また24,688個の転写産物は年周性を刻み、そのうち11,963個の転写産物は高い振幅を示すとのことだった。

セイヨウトネリコ (*Fraxinus excelsior*) における灰枝枯れ病 (病原体 *Hymenoscyphus fraxineus*) に関する研究では、Associative Transcriptomics のテクノロジーを利用して耐性遺伝子の特定が試みられていた。この方法は、遺伝子上の変異と遺伝子発現変化を統合したアプローチであり、174,470の SNP と 32,441の発現マーカーを利用して、パネル集団186個体の分析を行った結果、SNP

* E-mail: mishimak@affrc.go.jp

¹みしまけんたろう、ひらおともりのり、たかはしまこと 森林総合研究所林木育種センター

マーカーと発現マーカーを検出し、特に MADS 遺伝子が耐性形質と関係するとのことであった。

シュガーパイナップルにおいては、サビ病抵抗性のメカニズムを解明するために、様々な組織やストレス環境下における発現遺伝子の情報収集(アセンブル・アノテーション)を行っていた。遺伝子情報の収集については、次世代シーケンサー (Illumina HiSeq と Miseq) から得られた 25 億リードと第三世代シーケンサー (PacBio) の Iso-seq テクノロジーにより 16 億リードを取得し、33,720 個の単一遺伝子を単離していた。Iso-seq テクノロジーは、ショートリードの次世代シーケンサーではカバーできないスプライシングバリエーションを網羅的に取得できる。33,720 個の単一遺伝子のうち、10,026 個の単一遺伝子はどのライブラリーにも共通する遺伝子であり、17,505 個の単一遺伝子は、各組織に特異的な遺伝子であることが報告されていた。

ベイスギ (*Thuja plicata*) では、シカの忌避物質や心材腐朽抵抗性を特定するために、候補遺伝子の探索及び生化学的生成経路の探索が RNA-seq を使って行われていた。シカの忌避物質の特定では、食害を受けない個体ではモノテルペンの一種の濃度が高いことが明らかにされており、食害を受けない個体と野生型のタイプについて、葉のサンプルから RNA-seq を行い、得られた遺伝子情報を比較し、テルペノイド生成経路に係わる約 500 個の遺伝子を候補遺伝子として特定していた。

ヒロハヤマナラシでは、生物燃料産業に適合した栽培品種を育種するため、産業上有用な形質に関連する遺伝的変異の検出を GWAS により行っていた。血縁関係のない 500 個体から構成される集団に対し、18,153 個の遺伝子についてターゲットシーケンスを行い、PLINK、SKAT、BURDEN といった様々な解析手法で関連解析を行っていた。リグニン含量や葉のバイオマスといった形質で遺伝的変異との関連性が検出され、113 遺伝子は 8 つの形質で共通して関連性があるとのことだった。また、レアな変異がリグニン含量に関与する可能性が示唆されており、解析においてはレアな変異も含めることも重要であるとの報告を行っていた。

E. globulus と *E. glandis* では、量的形質や複雑な表現形質を遺伝的に理解するために大規模な DNA マーカーを利用してゲノム全体の組み換え (連鎖不平衡) パターンを明らかにする研究が進められていた。2014 年に全ゲノム塩基配列が決定された *E. glandis* (691 Mb) について、36 個体を Illumina Infinium で 21,517 個の SNP マーカーと 1300 万個の SNP を用いてジェノタイピングし、連鎖不

平衡は 4 ~ 6kbp であることを報告した。現在、60,000 個の SNP マーカーが搭載されたチップ (EuCHIP60K) が作成されており、ゲノム中で約 20kbp 間隔 (1cM あたり 20 マーカー) でゲノムワイドにカバーされたチップであり、GWAS や Genomic Selection など実際の育種集団に高い精度で選抜できるチップであると報告があった。

ゲノム解析技術

今回の PAGXXIV において、遺伝子情報を用いた研究を行うにあたって、次世代シーケンサーから出力される大量のゲノムデータを使いこなすことは、もはや常道となっているとの認識を強くした。ここでは、ゲノム解析のための新技術や分析機器の新機種の発表があったものについて報告したい。

これまで多くのゲノムプロジェクトにおいて対象となる生物種のゲノム情報については、次世代シーケンサーや第 3 世代シーケンサーを駆使して対象となる生物種のゲノム情報が取得されてきている。既に、様々なスペックの次世代・第 3 世代シーケンサーが利用可能となっており、それぞれの研究目的に応じて、それに合った機種を選択し利用することが普通になっているとの感があった。ユーザー側の多くが選択しているのは、Illumina 社の Hi-seq や Mi-seq、Life Technologies 社の Ion Torrent, Ion Proton など数百 bp のショートリード型のシーケンサーと一分子シーケンスシステムである PacificBioscience 社の RS II や BioNano 社の Irys system のようなロングリード型の 2 つである。Irys system では、高分子 DNA の全体をラベリングした後、チップ上で 1 本に引き伸ばした DNA をスキャンし、イメージ画像を使ってアセンブルするという手法をとっている。今回発表あった多くのメガクラスのゲノムをもつ生物種の研究発表では、この 2 種類のタイプの機種を併用して、効率的かつ高速にゲノムプロジェクトが進められていた。PacificBioscience 社については、今回新たな機種の発表等はなかったが、現在のケミストリーである P6-C4 ケミストリーのさらなる改良により、最長約 30kbp まで取得リード長が長くなり、精度についても改良されるとのことであった。PacificBioscience 社では、既に汎用機種となっている RS II の上位機種である Sequel system があるが、今回の発表ではコスト (per SMRT Cell) は、現在の \$7,000 から次年度 \$3,000、次々年度 \$1,000 まで下がることが報告されていた。Illumina 社では、新た

な機種である Miniseq について報告があった。これは、Miseq の下位機種に位置付けられる最小クラスの機種であり、実験台に置くことのできる PCR ほど本体サイズである。これで近年発表された最上機種の Hi-seq X Ten, Mi-seq と Hi-seq2500 の中間機種である Nextseq 500 とあわせ、あらゆる分析スケールに対応したラインナップが出揃ったことになる。今回発表された miniseq の価格は \$49,500 であり、最大 2x150 bp・7.5 Gb を 24 時間で出力できる (High-output kit 使用時)。今回、多くの栽培作物において、GBS が使われていたが、この低価格機種の導入によって、ゲノム情報がないあらゆる生物種にわたって、さらにアプローチしやすくなるのではないかと考えられる。また、Dovetail-Genomics 社からは新しいロングリードライブラリーの作成技術が開発されていた。The Dovetail Chicago method と名付けられており、以下簡単にその原理を記すと、① 50 kbp 以上の配列から *in vitro* でクロマチン構造 (ヒストンにしっかり巻きついた状態の) を再構築し、固定液によって複雑な crosslink をとるように固定する。② crosslink されたクロマチンのまま制限酵素処理によって粘着末端化された後、ビオチン標識して平滑化しライゲーションを行う (このライゲーションは多くの場合、もともと近接していなかった断片同士で生じることになる)。③クロマチン構造を除去し、ビオチン標識されたキメラ状の配列を濃縮しライブラリーとする。シーケンスによって生じる配列は、少なくとも 2 つ以上の元の配列からなり、配列は HiRise という専用の Scaffolding Assembler によってアセンブルされる。このテクノロジーを用いることにより、植物においては、繰り返し構造を多く持つレタスのゲノムが効率よく解析されていた。この新しい技術によって、ロングリードでは、PacificBioscience 社、BioNano 社に続いて Dovetails 社も新たな選択肢になると考えられる。Dovetails 社のワーショップは大盛況で、今後のロングリードシーケンスの動向については注視する必要があると感じた。

ゲノム情報と表現型情報の育種への利用

今回の発表では、表現型測定、データベースに関する発表も多数みられた。ゲノム情報を活用した分子育種を進めようとする時、よりよい予測モデルを構築するためには遺伝子型データと表現型データのそれぞれを高精度にペアで収集し、それらを統合的に解析することが必要となる。遺伝解析の方は、技術が日進月歩の発展を続

けており、1 回の分析から得られるデータ量、データ精度ともに大きな進歩を遂げている。このような中であって、多くの研究者が近年課題と考えているのが、表現型データをいかに正確に高効率に収集するかということである。表現型測定の高精度化と高効率化を図るために、タブレット端末を利用した FiledNote 等のソフト開発が進められている。また、データをより大量に得るためのもう一つの方向性として共同化が進められている。農作物の分野では複数の国、あるいは米国内の複数の州にまたがったコンソーシアム形成により、複数機関が協力して表現型データを収集するという大規模なプロジェクトが多数進行している。こういった表現型データ収集の共同化を進める上で、データベースが重要な位置を占めている。iPlant (正式には iPlant Collaborative, アメリカ国立科学財団 (U.S. National Science Foundation ; NSF) が出資する資金により、アリゾナ大学 (University of Arizona)、テキサス大学の Texas Advanced Computing Center、コールド・スプリング・ハーバー (Cold Spring Harbor Laboratory)、ノースカロライナ州立大学 (University of North Carolina) の主導の下に 2008 年設立) や GOBII (Genomic and Open-source Breeding Informatics Initiative) がデータベースの構築を進めている。特徴としてはウェブ・ベースのデータベースで、メンバーが自由にデータを登録できる環境を提供している。集積されたデータにメンバーは自由にアクセスできるようになっており、解析ツールのパッケージも用意されていて、メンバーはウェブ上でデータ解析できる。ウェブサーバーにはスーパーコンピュータを配することによって各地にいるユーザーも高パフォーマンスのデータ解析能力の恩恵を享受できるようにしている。

また、複数機関が個別に収集するデータが効果的に用いられるようにするために、表現型形質の規格化の動きも進んでいる。一つの形質をいかに正確に定義するか、その形質の測定単位の統一、測定方法の統一化の作業が国際的レベルで進められており、それらの成果は、Plant Ontology や Trait Ontology という形でとりまとめられている。

終わりに

今回、私自身 (三嶋) は初めての本大会の参加であったが、内容はもとより、その規模や密度の高いスケジュールに圧倒された。毎朝 7:00 から 4,000 人の参加者全員がホールでビュッフェ形式の朝食を摂るところから 1 日

が始まり、夜 20:00 までプログラムが続く。また、特筆したいのは、後半の3日間、毎日 8:00 から 10:00 までの2時間に、2名のスピーカーがそれぞれ1時間ずつ行う Plenary lecture が設けられている点である。この時間は、参加者全員が大ホールで発表を聞く時間に充てられている。Plenary lecture のスピーカーには、一流誌に多数の論文を掲載している、文字通り“一流の”研究者が選ばれている。今回、私が特に印象に残っているのが、Erich Jarvis 博士の行った鳥類を対象とした「Molecular mechanisms underlying natural circuits for vocal learning」であった。内容はこの発表に限らず、どの発表も最新かつ大

量のデータを示した重層的なものであり、しかもプレゼンテーション自体が大変魅力的であった。最後になるが、本大会はゲノムを利用した（あるいはその周辺の）動植物育種の最新の研究成果とその手法の全体を俯瞰できる学会となっており、またゲノム情報を使った研究に携わる者であれば誰もが知っている企業のほとんどが展示を行っていて、最新の試薬及び機器の情報が得られる。私個人としては、PAG はゲノム情報を利用した世界中の研究の最新状況を把握するのにまたとない良い機会であると感じた。