

パンゲノムリファレンス収集のパイオニア

ロングリードシーケンスのリーダー



DuPont Pioneerの研究開発部では、トウモロコシ、ダイズ、モロコシ、ヒマワリ、ムラサキウマゴヤシ、キャノーラ、コムギ、イネなどの商業作物における形質の遺伝基礎を明らかにするためにDNAシーケンスを利用しています。しかし、以前のように高品質リファレンスゲノムの構築に数年を費やしたり、遺伝資源における遺伝的多様性を1つのリファレンスゲノムのみで表したりすることに限界を感じていました。

そこで、高速かつ高スループットのシーケンスとアセンブリを基盤とするパンゲノム(種間ゲノム比較により得られる遺伝子レパートリー)のリファレンス収集を目的とした意欲的なプロジェクトが始まりました。このプロジェクトでは、育種選抜された独自のトウモロコシの遺伝資源、そして興味のある形質を有するその他のトウモロコシ品種を対象としています。

トウモロコシは二倍体ゲノムで、10の染色体から構成される2.3 Gbから2.7 Gbのゲノムサイズであると考えられています。転移因子によるリピートやリボゾームDNA (rDNA)、高いコピー数で存在する短い縦列反復(主にテロメア、セントロメア、そしてヘテロクロマチンに位置)が存在し、1遺伝子あるいは少数の遺伝子が集まる独自のDNAアイランドで区切られています。

トウモロコシは非常に多様性のある生物種です。2005年にDuPont Pioneerの研究者はトウモロコシ近交系B73とMo17におけるアレルのゲノムセグメント解析結果を発表し、広範囲におよぶ非相同性のDNA配列を示しました。その結果、比較した遺伝子座間において、平均すると配列の5割、そして遺伝子の1/3が異なっていることがわかりました。参考までに、ヒトとチンパンジーは異なる生物種ですが、98%以上の配列相同性を示しています。こうしたことから、野生品種と疾患耐性などの形質をもつ栽培化された品種を解析対象としている研究者にとって、1つのトウモロコシリファレンスゲノムを使用することは十分ではないことが明らかとなったのです。

農作物植物の多くのゲノムアセンブリは、過去10年にわたり次世代シーケンサー(NGS)により産出されました。しかしこれらのアセンブリは断片化されており、そのため情報の価値も限られたものでした。ショートリードのシーケンスにのみ頼っていると、せつかくの努力も、表現型に重要な役割を果たす1塩基多型や構造多型を含むゲノムの複雑性を適切に再構築することができませんでした。



Kevin FenglerはDuPont Pioneerで数百にも及ぶ所内バイオインフォマティクスシステムと処理を基盤とする高品質リファレンスゲノムの開発に取り組んでいます。

最初の高品質なトウモロコシリファレンスゲノム(B73)は1分子リアルタイム(SMRT®)シーケンスを使い、Cold Spring Harbor LaboratoryのDoreen WareのチームとUSDA Agricultural Research Serviceにより実施され、トウモロコシコミュニティに多くの利便をもたらしました。しかしPioneerグループにとって、商業的および野生株の両品種におけるリファレンスゲノムが必要であることが明らかになりました。

Pioneerではまた、アセンブリ処理の最適化にかかる時間を改善したいと考えていました。彼らの取り組みは、シーケンスに2週間、アセンブリからポリッシング、染色体レベルのスケヤホールディングに10日間、と一連の処理を1か月以内にまで短縮する新たな高スループット手法の構築を生み出しました。

「何度も試行錯誤し、その度に処理を改善することでこのパワーが生まれたのです」とDuPont PioneerのData Science and Informatics部門で研究サイエンティストとして働くKevin Fenglerは言います。「これまでにすでに数回実施しているので、高品質のリファレンスゲノムを得るには何をすれば良いかわかるようになりました。しかし、パンゲノムの視点を可能にするには、これらを高スループットで行うワークフローを開発しなくてはならなかったのです」。



DuPont Pioneerの科学者たちは、Pioneerが独自に有する選抜品種やその他重要なトウモロコシ品種を対象に、SMRTシーケンスと*de novo*アセンブリを組み合わせてトウモロコシにおけるパンゲノムリファレンスを産出し、収集しています。

成功へのコツ

こうしたワークフローの構築には何が必要なのでしょう？それは、迅速なロングリードシーケンスです。Dupont Pioneerでは複数のSequel®システムを所有し、これを達成しています。

Genomics Technology部門で研究ディレクターを務めるGreg Mayが率いるシーケンス専門家チームが、まずPacBio®のデータを産出します。その後、データはFenglerらに渡り、アセンブリからコンティグを構築します。その際、非近交系やヘテロな株にはFALCONなどのツールを、近交系にはCanuを使ってアセンブリし、その後エラー補正を行います。

ロバストで大きなコンティグは、リファレンスゲノムのアセンブルプロジェクトに不可欠です。N50が1 Mb以上のコンティグとなるアセンブリは成果と見なされており、またバーチャルな“1 Mb Contig Club”の一員となり、オンラインのソーシャルネットワークにおいても認められます。近年、PacBioのSMRTシーケンスで得られるロングリードを使い、多くの植物ゲノムアセンブリがこのステータスにたどり着くことができました。

「トウモロコシにおける私たちの新しい基準は、2.8 MbのコンティグN50です。これからシーケンスする各トウモロコシゲノムにおいて、この基準を達成することを目標としています」とFenglerは話します。

Fenglerはさらに他の種類のデータを加えることで、迅速にコンティグのアセンブ

リを完全なリファレンスゲノムにまで高めています。ショートリードのクラウドを使ってコンティグをポリッシュし、光学マッピングを用いてコンティグをつなぐことでスカフォールド化し、さらにHi-Cシーケンスデータによりスカフォールドを染色体のpseudomolecule (仮想的な一続きの塩基配列)に変換しています。

「リファレンスゲノムプロジェクトは、よくコンティグから染色体に変換するところをつまづいてしまうことが多いのですが、所内で補足的なゲノム技術をアセンブリのコツとして重ねることで、こうした課題を克服しています」とFenglerは話します。

「コンティグアセンブリとスカフォールドの可能性と実用性は、リファレンスゲノムが完了した時点で最大化されます。私たちはアセンブリではなく、リファレンスゲノムを構築しています。リファレンスゲノムは、数百にも及ぶ所内のバイオインフォマティクスのシステムと処理を重ねることで得られます。すべてのアセンブリが同じような高いレベルで行われると、こうした作業も随分と楽になります。」

“高品質のPacBioゲノムにより、
遺伝子探索や製品開発のために
使用可能な配列情報を
研究者に提供できるように
なったのです。”

これまでのNGSを使ったアプローチでは、シーケンス情報が不完全だったり、検証が必要でした。「高品質のPacBioゲノムにより、遺伝子探索や製品開発のために使用可能な配列情報を研究者に提供できるようになったのです」とFenglerはさらに語ります。

それぞれの遺伝子型ごとに複数の高水準ゲノムアセンブリをもつことは、植物ゲノムにおける構造多型の影響をさらに理解する上でますます重要になりつつあります。そうでないと、ミスアセンブリやコンティグの誤配置を、変異(多型)として解釈してしまう可能性があるからです。

「真の構造多型にフォーカスし、これらのゲノムで見つける新しい探索に自信を持ちたいのです」とFenglerは話します。

合理化された全ゲノムシーケンス

SMRTシーケンスは多大な人的リソースや予算を必要としません。少数の研究者でも高品質のリファレンスゲノムをコスト効率良く構築できます。Fenglerは、ワークフローを簡素化することで処理に必要な資源を節約することができることも話しています。

高品質全ゲノムシーケンスがもたらすスピードおよびコスト効率性により、幾つかの既知手法が使われなくなりました。もしPioneerの研究者がある1遺伝子にのみ興味をもったとしても、全ゲノムをシーケンスしてアセンブルすることが最も簡単なアプローチとなることがあります。

「必要な配列情報を入手するために、私たちはこれまでBACシーケンスやターゲットキャプチャーなどといったターゲットゲノミクス手法も考慮しました。しかし、多くのケースで、最も迅速で簡単かつ経済的な方法は全ゲノムアプローチであったりするので、BACライブラリーを作ってスクリーニングを行い、ターゲット領域をシーケンスする時間があれば、その間に全ゲノムアセンブリを完了できてしまうのです」とFenglerは言います。

ロングリードシーケンスを使って全ゲノムアセンブリを実施すると、複雑な領域に集中する疾患耐性遺伝子の探索に特に威力を発揮します。こうした遺伝子を紐解くには、ロングリードが必要なのです。

DuPont Pioneerでは、トウモロコシゲノムをさらに深く探索することで多大なる利益を得ることができました。「これまでで初めて、存在すら知らなかったトウモロコシのパンゲノムの視点を捉えることができるようになったのです」とFenglerは語ります。

「これまでB73リファレンスにフォーカスしていたことで、私たちの視点は限られていました。今まさに、これまで見えていなかったことを、探索することができるようになったのです。」