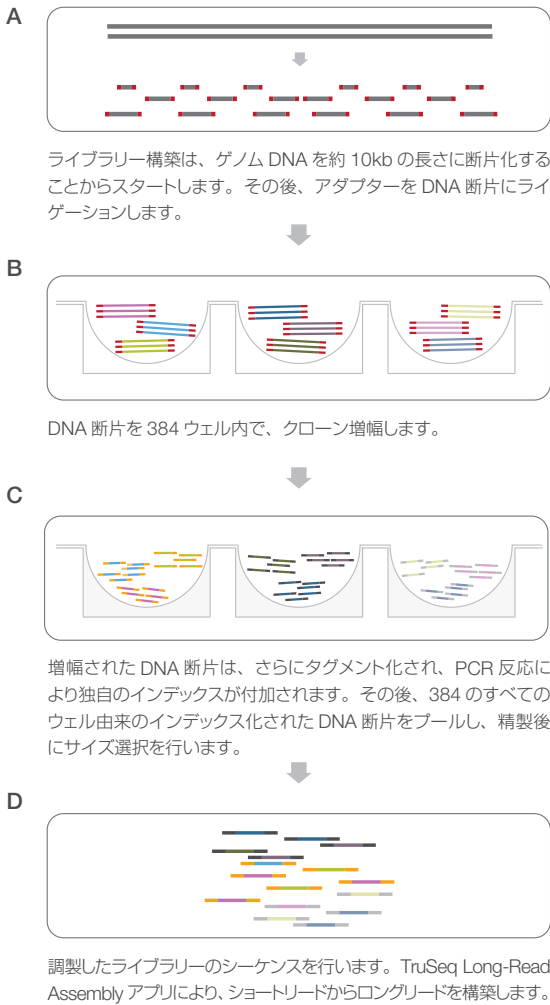


図 2: TruSeq Synthetic Long-Read による DNA ライブラリー調製ワークフロー



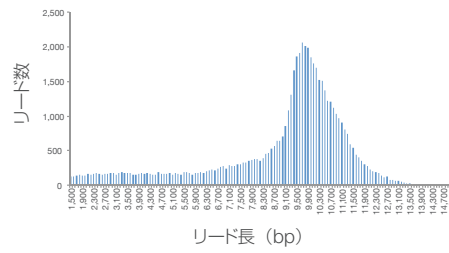
TruSeq Synthetic Long-Read DNA ライブラリー調製キットは、シーケンス用に DNA ライブラリーを調製します。TruSeq Long-Read Assembly アプリにより、ショートリードをロングリードにアセンブルします。

表 1: *C. elegans* (線虫) のゲノムアセンブル

パラメーター	結果
ロングリードの GC 含量	35.6%
ロングリードのリファレンスゲノムへのアライメント	99.9%
コンティグのミスアセンブル率	1.0%
N50 サイズ	9,125bp
ロングリードに完全にアセンブルされた塩基	1,182Mb

GC 含量は QUAST³ を使用して算出しました。リファレンスゲノムにアラインしたロングリードの割合は、MUMmer⁴ を使用して決定しました。コンティグのミスアセンブル率は、QUAST により定義されたミスアセンブルブレイクポイントにアラインするロングリードの合計の割合を算出することによって決定しました。

図 3: *C. elegans* (線虫) 由来の、エンドマークできたロングリードのサイズ分布



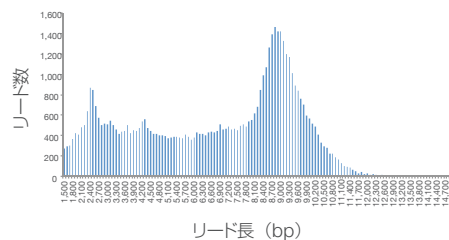
C. elegans 由来の DNA ライブラリーは TruSeq Synthetic Long-Read DNA ライブラリー調製キットを使用して調製し、ロングリードは TruSeq Long-Read Assembly アプリを使用してアセンブルしました。その結果をリード長 (X 軸) ごとに完全にアセンブルされたロングリードの数 (Y 軸) で示しています。この図に使ったロングリードは、両末端がマーカーでラベル化され、かつ末端からもう一方の末端へアセンブルできたリードのサブセットを表しています。リード長の中央値は約 8~10kb です。

表 2: *O. sativa* (イネ) のゲノムアセンブル

パラメーター	結果
ロングリードの GC 含量	40.5%
ロングリードのリファレンスゲノムへのアライメント	99.8%
コンティグのミスアセンブル率	2.5%
N50 サイズ	7,095bp
ロングリードに完全にアセンブルされた塩基	1,088Mb

GC 含量は QUAST を使用して算出しました。リファレンスゲノムにアラインしたロングリードの割合は、MUMmer を使用して決定しました。コンティグのミスアセンブル率は、QUAST により定義されたミスアセンブルブレイクポイントにアラインするロングリードの合計の割合を算出することによって決定しました。

図 4: *O. sativa* (イネ) 由来の、エンドマークできたロングリードのサイズ分布



O. sativa Nipponbare 由来の DNA ライブラリーは TruSeq Synthetic Long-Read DNA ライブラリー調製キットを使用して調製し、ロングリードは TruSeq Long-Read Assembly アプリを使用して作成しました。その結果をリード長 (X 軸) ごとに完全にアセンブルされたロングリードの数 (Y 軸) で示しています。この図に使ったロングリードは、両末端がマーカーでラベル化され、かつ末端からもう一方の末端へアセンブルできたリードのサブセットを表しています。リード長の中央値は約 8~10kb です。

高精度なゲノムアセンブル

TruSeq Synthetic Long-Read DNA ライブラリー調製キットおよびバーコードキットは、実績のある TruSeq テクノロジー¹ を使用して、ロングリード用のライブラリーを調製します。また、TruSeq Long-Read Assembly アプリでは、得られたショートリードからロングリードを構築します。TruSeq Synthetic Long-Read テクノロジーを使うことで、追加のシーケンサーや更なるカバレッジを必要とすることなく、高精度な結果をもたらします。同様の精度を達成するために、より高いシーケンス深度および特別な設備を必要とする従来の方法への依存が回避されます。

TruSeq 法では、約 6~10kb の長さの隣接フラグメント（コンティグ）が構築されるため、相同性の高い反復領域間の識別には有効です。これらの長いコンティグは、ゲノムカバレッジを向上させ、*de novo* ゲノムアセンブルをサポートします（表 1~2）²。サイズ選択された長い断片には両末端に特異的タグまたはエンドマーカが含まれているため、TruSeq Long-Read Assembly アプリを使用して完全にアセンブルされたロングリードの数を可視化することができます（図 3~4）。TruSeq Synthetic Long-Read テクノロジーは、精度を犠牲にすることなく、解析が困難な領域のカバレッジを向上させるため、得られるアセンブルの信頼性が高まります。

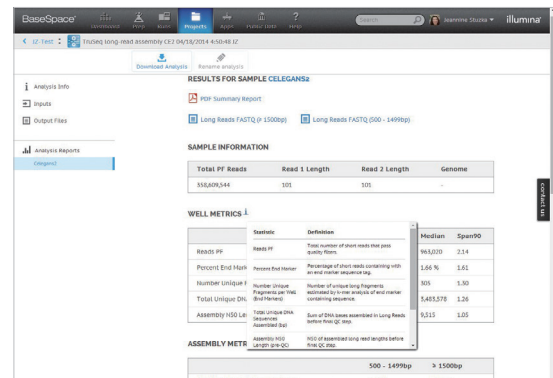
装置のフレキシビリティ

TruSeq Synthetic Long-Read DNA ライブラリーは、イルミナの HiSeq[®] 2500 または HiSeq 2000 システム上でシーケンスすることが可能です。TruSeq 法では、複数のアプリケーションに対して 1 台のシーケンス装置で貴重なロングリード情報へのアクセスが可能のため、特別な装置を追加する必要はありません。この方法により、従来のアプローチにかかるコストのほんの一部で、ゲノムのより詳細な情報が提供されます。

シンプル解析およびアセンブル

BaseSpace 環境でのプッシュボタン解析により、ロングリードのアセンブルを簡素化します。データはイルミナのシーケンサーから BaseSpace クラウドに即座に転送されます。TruSeq Synthetic Long-Read DNA ライブラリー調製キットおよびバーコードキットとともに使用するようデザインされた TruSeq Long-Read Assembly アプリ⁵ は、ショートリードからロングリードを構築します。直観的なユーザーインターフェース（図 5）は、データ解析を簡素化し、プロジェクトおよびファイルの保存先を選択するだけでデータ解析が行えます。アプリではショートリードを処理し、オーバーラップベースの方法を使用してイニシャルコンティグをアセンブルし、最後に合成的にロングリードを構築するためにコンティグのスカフォールドを作成します（図 6）。アセンブルされたリードはスタンダードの FASTQ フォーマットとして出力され、さらなる解析のためにダウンストリームのアセンブルツールに直接取り込むことが可能です。TruSeq Long-Read Assembly アプリにより、初心者ユーザーでも、高度な専門知識または基盤を必要とせずに、ワンクリックのインフォマティクスが可能となります。

図 5：簡素化された解析



直観的なユーザーインターフェースを特長とする TruSeq Long-Read Assembly アプリは、バイオインフォマティクスに関する専門的知見に関わらず、あらゆる生物学研究者のデータ解析を簡素化します。結果は分かりやすい表およびグラフにより表示されます。

図 6：TruSeq Long-Read Assembly アプリのワークフロー



TruSeq Long-Read Assembly アプリはシーケンスデータを使用してロングリードを構築し、結果は FASTQ ファイルで提供されます。

まとめ

ショートリードと TruSeq Synthetic Long-Read テクノロジーを組み合わせることで、1 台のシーケンサーから、従来のアプローチと比較して高精度なロングリードを入手することが可能となります。生物学者による利用を考慮してデザインされた TruSeq Long-Read Assembly アプリは、バイオインフォマティクスを簡素化することにより、研究者がデータ解析に費やす時間を短縮し、さらに研究に集中することを可能とします。TruSeq Synthetic Long-Read テクノロジーは、ゲノムアセンブルおよびゲノムフィンッシングのための包括的ソリューションを提供します。

