

プラスミド全長シーケンスサービス 結果解釈ガイド

1. プラスミド全長シーケンスサービスの精度について

プラスミド全長シーケンスサービスで使用するケミストリーおよびフローセルに関して、オクスフォードナノポア(ONT)社の仕様によると、読み取り精度は99%を超えているとされています。一般に、カバレッジが高いほど、つまりコンセンサス構築に利用できるリード数が多いほど、結果の精度が向上する傾向があります。

一方、レポート内で、最終プラスミド全長シーケンスサービス内のより信頼性の低い位置を検出するバリエーションコールも提供しています。

当社のコンセンサス配列とお客様のリファレンス配列との間に矛盾が見られる場合は、プラスミド配列のうち、一部の欠落、突然変異、またはその他の要因により、提供されたコンストラクトがリファレンス配列と異なっている可能性があります。

● 信頼性の低い塩基について

当社では、ディープシーケンシングを利用して、個々の塩基レベルで高レベルの精度を実現しています。ただし、ONTのロングリードデータは、ある特定の配列に対しては不得意な一面を持っています。その課題を解決するため、シーケンス結果はポリッシュされます。

より信頼性の低い塩基を特定するために、高品質のコンセンサスアセンブリに対してリードをマッピングしています。このプロセスでは、特定のポジションでの各ヌクレオチドの頻度が決定されます。信頼性の高い塩基を持つ領域では、リードの大部分がアセンブルされた塩基と同じ塩基を含んでいます。ただし、Dcmメチル化部位（CC[A/T]GG）などのモチーフやホモポリマー塩基の長い鎖長が含まれる領域では、アセンブルされた塩基が正しい可能性があるにもかかわらず、リードの同じ位置で異なるヌクレオチドが同定される可能性があります。

アセンブリが期待と異なる場合は、これらの要素を考慮することが重要です。

ホモポリマー領域またはDcmメチル化部位のエラーについて

ホモポリマーストレッチ内の欠失と、Dcmメチル化部位（CCTGGおよびCCAGG）の中間位置でのエラーは、本解析で最も生じる可能性の高いエラーとなっています。

● プラスミドのシーケンシングカバレッジ

生成されるリード数はサンプルの品質によって大きく異なる可能性があるため、特定のレベルのカバレッジ保証を提供することはできません。通常、推奨濃度で送付されたサンプルが成功すると、数十から潜在的には数百、さらには数千に及ぶリードが得られます。レポートには平均カバレッジが示されており、約20倍以上のカバレッジは、精度の高いコンセンサスであることを示唆しています。

2. データの解釈について

● リード長ヒストグラム

プラスミドの配列を決定する前に、ライブラリー調製によって環状DNAが線形化され、主に全長シーケンスリードが取得されます。結果レポートでは、サンプルに存在する（シーケンス可能な）すべてのDNA分子からのリード長を示すリード長ヒストグラムをプロットします。

ヒストグラム内の1つの主要なピークは、通常、十分な濃度のクリーンなプラスミド調製物を示します（下記優れたプラスミド調製例を参照）。

● 非加重ヒストグラムと加重ヒストグラム

シーケンスデータからのリード長の分布はヒストグラムに示されています。リード長ヒストグラムはシーケンスデータの品質を評価するために使用され、リード長の分布は抽出品質や断片化、夾雑物の存在、またはシーケンスプロセスのバイアスを示す可能性があります。

また、サンプルの品質に応じて、短いプラスミドのサイズを決定するために使用することもできます。

● 非加重ヒストグラム(Non weighted histogram of read lengths)

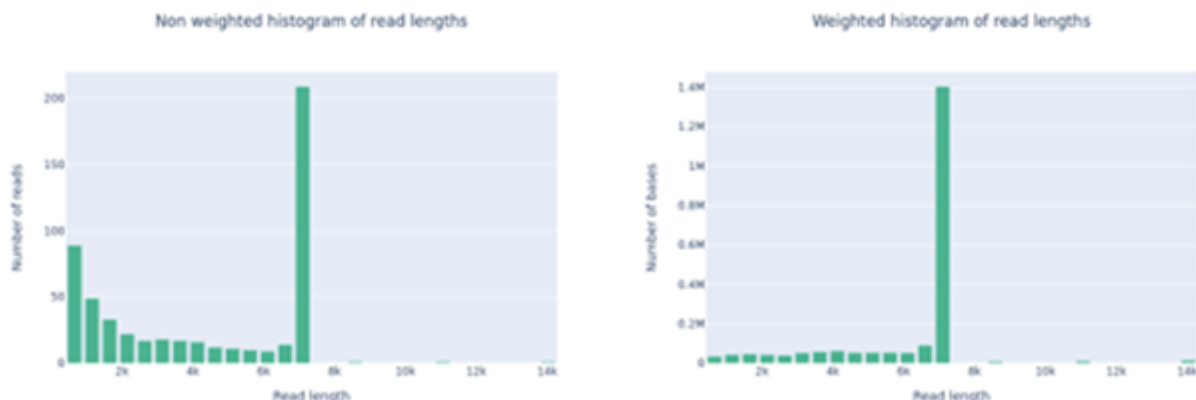
最初のヒストグラムは、Y軸にリード数、X軸にリード長を表示します。ヒストグラムの各バーはリード長の範囲を表し、バーの高さはその範囲内にあるリードの総数を示します。

● 加重ヒストグラム(Weighted histogram of read lengths)

2番目のヒストグラムは、シーケンスされた塩基数（bp）をY軸に、リード長をX軸に表示します。リード総数の代わりに、バーの高さは、その範囲内にある塩基の総数（bp）を示します。

3. ヒストグラム例

A. 優れたプラスミド調製例



1つの主要なピークは、クリーンなモノクローナルプラスミド調製物を示しており、通常は良好なシーケンス結果が得られます（十分なカバレッジが達成されている場合）。

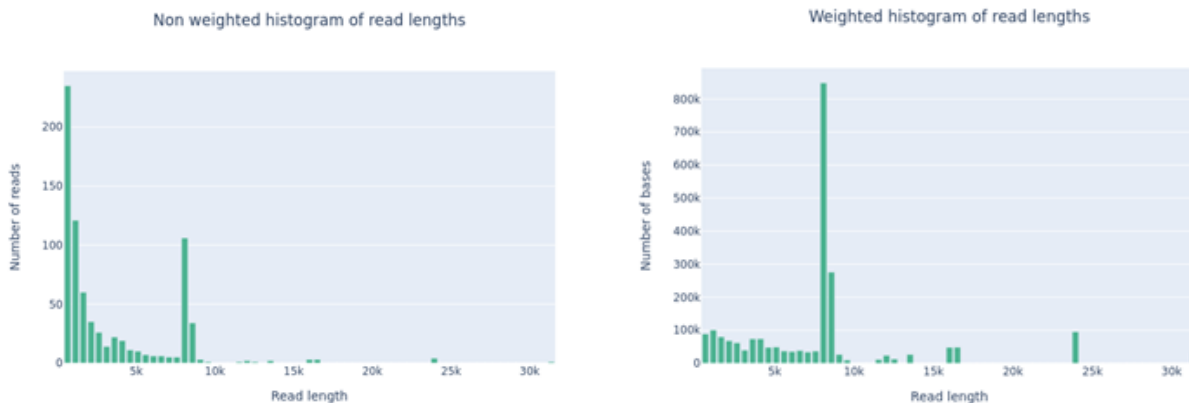
プラスミド混合物は、サンプル内のシーケンス相同性に応じて、誤ったアセンブリを引き起こす可能性があります。

ヒストグラム内の1つの見かけのピークは、同じサイズの複数のプラスミド、または同じビン内に含まれるさまざまな長さの複数のプラスミドを表す可能性があることに注意してください。

分析パイプラインでは、類似性の高い配列が単一種のバリエーションとして扱われ、その結果、単一のコンセンサス配列を生成しようとします（レポート内で信頼度が低いポジションがレポート内で報告されます）。

B. 必要な基準を満たすプラスミド調製例

ゲノム断片とプラスミド断片の両方を含む、大量の分解されたDNAとともに、顕著なピークを示すサンプルに遭遇することもあります。ほとんどの場合、リードカバレッジと精度が必要な閾値を満たしていれば、主要なピークは依然としてコンセンサスシーケンスを生成します。



リード長がヒストグラムの2つの異なるサイズのビンに分割されることがあります。これは次の可能性があります。

- 定義されたビン境界により、1つのクリーンなプラスミドの主要なリード長ピークが2つに分割されます。これは分析には影響しません
- プラスミドの小さな変化（例えば、異なるサイズのコモポリマー領域、小さな挿入および欠失（InDel））により、ビン境界付近のプラスミドサイズがわずかに異なる。
- 非常に類似したサイズの2つのプラスミドのプラスミド混合物

プラスミド調製におけるこれらのバリエーションは、リード長ヒストグラムでのみ確認できますが、従来のサンガーシーケンスでは確認できません。

C. 複数のピークが認められる例

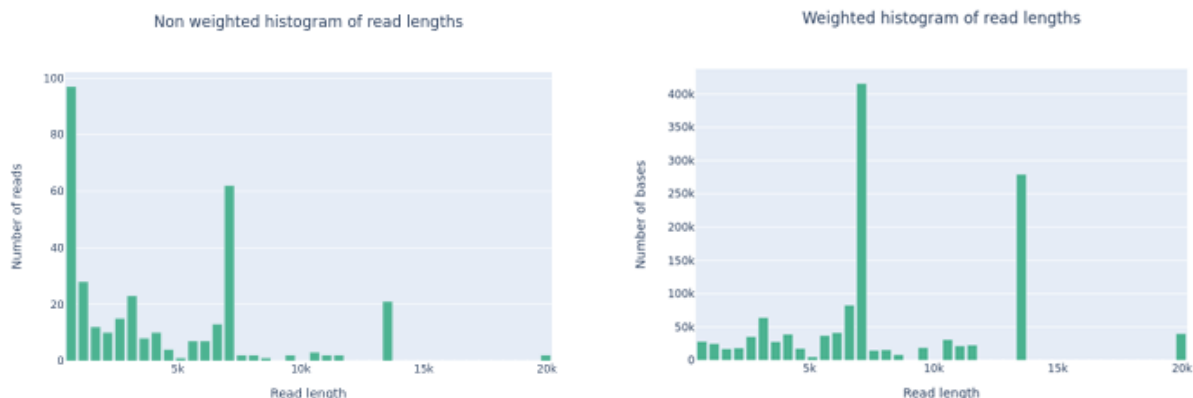
リード長ヒストグラムに複数のピークが観察される場合があります。これには3つの理由が考えられます。

- サイズの異なるプラスミドの混合物を含むサンプル
- 生物学的プラスミドのコンカテマー
- 欠失、組換え、または上記のコンカテマーの結果として生じる、クローニング株内でのプラスミドの増殖の予期しない副産物

私たちは、配列決定アーティファクトではないコンカテマーの存在を頻繁に観察します。コンカテマーはサンガーシーケンスでは検出できず、消化または線状化したプラスミドのゲルでは見ることはできません。その結果、それらに遭遇することに慣れていない人にとっては、それらは見慣れないものに見えるかもしれません。ただし、ゲル上でカットされていないサンプルを泳動すると、ダイマーなどのバンドが明らかになります。コンカテマーは、*recA*+株の増殖中に*in vivo*で形成されることがよくあります。この現象の詳細な説明については、下記をご参照ください。

<https://blog.addgene.org/plasmids-101-dimers-and-multimers>

当社のサービスは分子のプラスミドクローン集団に最適化されているため、ゲル上またはバイオアナライザー/フラグメントアナライザーを使用してサンプルの品質（すなわち、単一のゲルバンド）を管理することをお勧めします。



当社のパイプラインは、サンプルに含まれる主要なピークと対応するプラスミドのみを返すように設計されています（上記の加重と非加重を参照）。上記のサンプルがプラスミド混合物の場合、約7kbのプラスミドのみが報告されます。

D. プラスミド混合物の分離

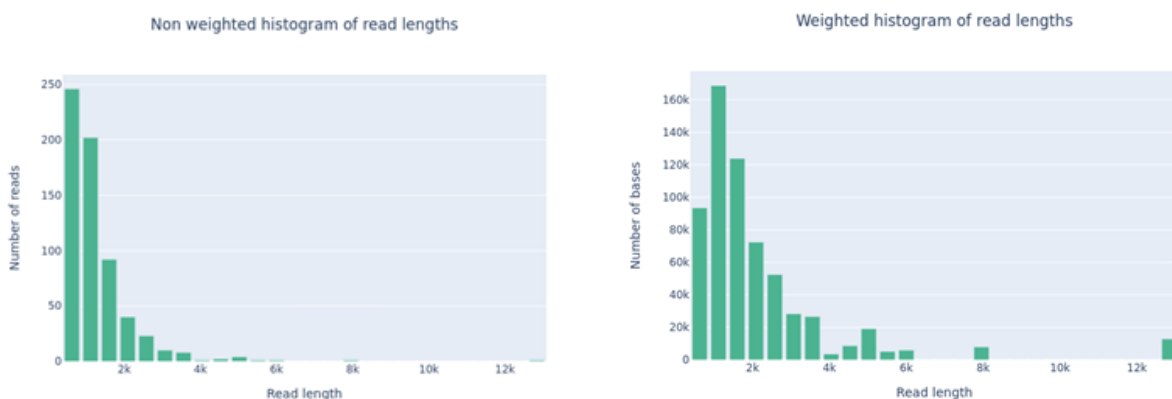
本サービスは、分子のクローン集団を分析するために設計されています。分子種の混合物を提出することは可能ですが、分析の結果を予測できないため、固有のリスクが伴います。

- プラスミド分子の長さや配列が非常に似ていて、ヌクレオチドがわずかに異なっている場合、分析パイプラインは通常、SNP/インデル位置に信頼性の低い位置を示す単一の.fastaコンセンサスファイルを作成します。これらの場所を特定するには、提供されたレポートを参照してください
- 種が十分な区別性を示している場合、プラスミドサイズが25kb未満であれば、パイプラインは最も豊富な種の.fastaコンセンサスファイルを作成します

E. DNAの分解または小さなフラグメントの混入

調製プロセス中にプラスミドDNAが分解された場合、結果として得られるシーケンシングリードは、リード数が高いにもかかわらず、主に主要なピークのない小さなフラグメントで構成されます。これにより、カバレッジが不十分になり、プラスミドコンセンサスシーケンスが生成されなくなる可能性があります。また、分解した宿主ゲノムDNAのコンタミも同様の結果を示す可能性があります。

当社のサービスは、分子のプラスミドクローン集団に最適化されているため、サンプル（すなわち単一ゲルバンド）の品質を、できれば線状プラスミドとして、ゲル上またはバイオアナライザー/フラグメントアナライザーを使用して確認することをお勧めします。



弊社による技術的な問題が特定されない限り、失敗したサンプルの再シーケンシングや払い戻しは行いません。必要なDNA濃度を満たしていないサンプル、複数のプラスミド、実質的な宿主汚染、または複数の大きな反復要素を持つプラスミドは、コンセンサス配列の生成に失敗する可能性が高くなります。サンプルのシーケンスに失敗した場合は、上記のサンプル要件を参照してください。サンプルの品質を向上させる方法に関する提案があります。

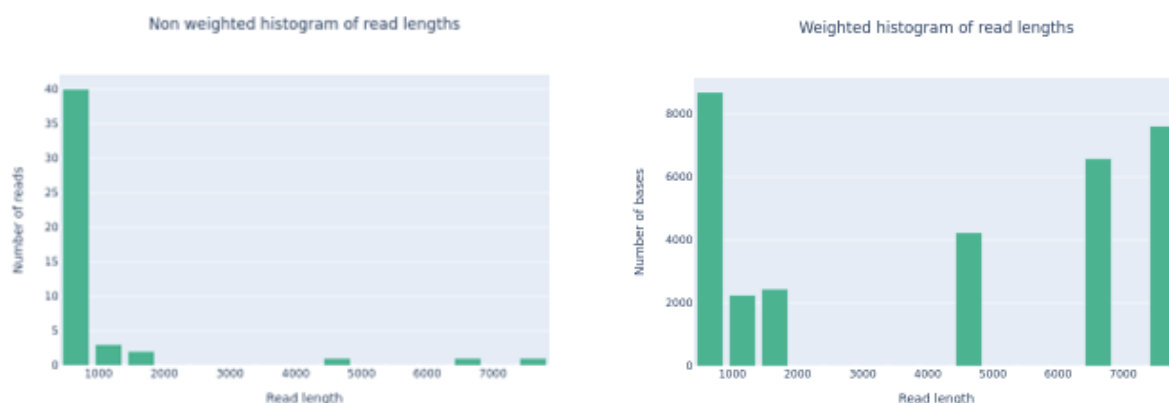
ユーロフィンジェノミクス株式会社

かずさ事業所 オミクス解析グループ NGS チーム

お問い合わせ先: Email: ngs-jp@gsjp.eurofinsasia.com TEL: 03-6631-0106

F. 不完全なデータ

個々のピークを区別したりコンセンサスを生成したりするには、リード数が不十分な場合がよく見られます。リード数が低すぎる場合は、通常、サンプルが50 ng/μlの仕様に従って、必要なDNA濃度で調製されていないことが原因です。吸光濃度測定（Nanodropなど）ではなく、二本鎖DNAの精度が非常に高い蛍光濃度測定（Qubitなど）を強くお勧めします。吸光測定では、サンプルのDNA濃度が実際より高く見積もられることがよくあります。当社は、顧客から依然として吸光測定で濃度測定しているプラスミドを受け取ることがよくありますが、これがプラスミド配列決定の失敗の最も一般的な理由です。



G. アセンブリ失敗/最終的なプラスミドサイズが合わない

考えられる理由は、通常、相同組換えを介して*recA*+株の増殖中に現れる生物学的プラスミドコンカテマー（上記参照）です。サンプル中のコンカテマーの割合が高い場合、それらが検出される可能性があります。コンカテマーは、消化/直鎖化されたプラスミドを含むゲルでは見ることはできませんが、スーパーコイル/未切断のプラスミドを泳動する場合に、スーパーコイルラダーとして見られます。

さらに、第3世代ONTサービスは、プラスミドのサイズに応じた長いリードでプラスミドのシーケンスを決定しますが、まれに、大きくて反復性の高いプラスミドがアセンブラーによって正しく解決されない場合があります。最後に、サンプルの品質が十分ではない可能性があります（クリーンな単一プラスミド調整には1つの主要なプラスミドピークが見られます）。複数のピークは、プラスミド混合物、または挿入/欠失や組換えなどの他の副産物を示します。