

GUIDE

PCR-EZ

Long-Read PCR Sequencing

クイックスタートガイド



目次

1. はじめに	21
プロジェクトレポートの閲覧	
サンプルレポートの閲覧	
<hr/>	
2. PCR配列の閲覧	23
Fastaファイルの確認	
Fastqファイルの確認	
<hr/>	
3. サンプル品質の評価	24
リード長分布の確認	
品質（クオリティ）分布の確認	
各塩基位置のカバレッジスコアの評価	
<hr/>	
4. バリエーションコール	28
<hr/>	
5. アセンブリ成功のポイント	29
濃度とクローンDNA	
アセンブリの失敗	
<hr/>	

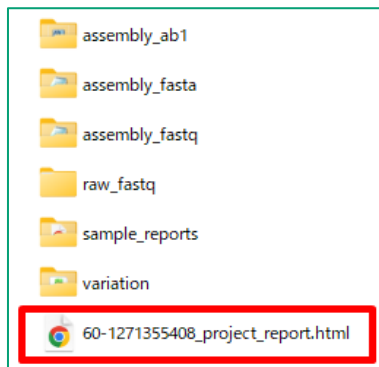
PCR-EZ クイックスタートガイド

1. はじめに

60-xxxxxxx.QC.html ファイルをクリックしてください。
このファイルには、全体的な QC レポートが含まれています。

プロジェクトレポートの閲覧

1.1 Results summaryには、アセンブリの簡易的な合否判定とシーケンス長が表示されます。下図の例では、3つのサンプルがコンティグとして正常にアセンブルされており、それぞれのシーケンス長は 1181、1554、1611 となっています。



1 ANALYSIS RESULTS

1.1 Results summary

Table 1.1 shows the assembled consensus sequence summary for all samples in this order.

Raw reads were cleaned, downsampled, assembled, polished and trimmed (remove low coverage regions), resulted in the final consensus sequence.

Note that the consensus sequence will miss up to 25 bp at the beginning and/or end of the linear/PCR product.

Table 1.1: Summary of the analysis results

Project	Sample	Assembly Status	Consensus sequence length
30-1140375569	BI1	Passed	1181
30-1140375569	BI10	Passed	1554
30-1140375569	BI11	Passed	1611

1.2 Raw data summaryは、RAW リードに関する主要な統計値が示されています。

この表には、総リード数、総塩基数、最小および最大リード長が一覧で表示されます。多くの場合、RAW リードの総数は、サンプルの濃度とサイズに相関します。

1.2 Raw data summary

Table 1.2 shows the summary of the raw data.

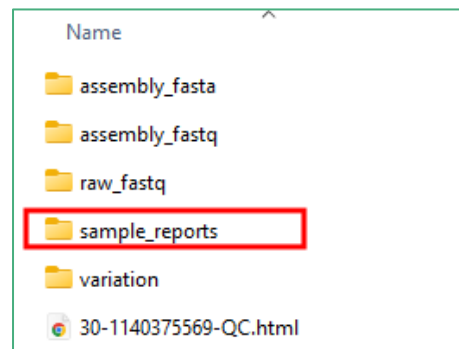
Table 1.2: Summary of the raw reads statistics

Project	Sample	Number of raw reads	Total Bases	Min read length	Max read length
30-1140375569	BI1	3941	2695339	115	2501
30-1140375569	BI10	1488	1114701	106	4042
30-1140375569	BI11	1394	1021351	101	3792

PCR-EZ クイックスタートガイド

サンプルレポートの閲覧

Sample_reportsフォルダを開き、60-xxx_report.htmlを参照します。



サンプルレポートには、サンプル情報、解析結果、データ品質管理（QC）指標が含まれています。解析結果のセクションでは、アセンブルされた配列と RAW リードの概要がまとめられています。一方、データ品質管理のセクションでは、RAW リードの長さおよび品質ごとの分布が表示されます。Assembly QC のドットプロットは、アセンブリに使用されたリードを視覚的に表しています。

これらの表に関する追加の詳細は、3章「サンプル品質の評価」に記載されています。

1 SAMPLE INFORMATION

Quote 30-1140375569

Sample BI1

2 ANALYSIS RESULTS

2.1 Results summary

Table 2.1 shows the raw data and the assembled consensus sequence summary. Raw reads were downsampled, assembled, polished and trimmed (remove low coverage regions) resulted in the final consensus sequence.

Note that the consensus sequence will miss up to 25 bp at the beginning and/or end of the linear/PCR product.

Table 2.1: Summary of the analysis results

Project	Sample	Consensus sequence length	Number of raw reads	Total Bases	Min read length	Max read length
30-1140375569	BI1	1181	3941	2695339	115	2501

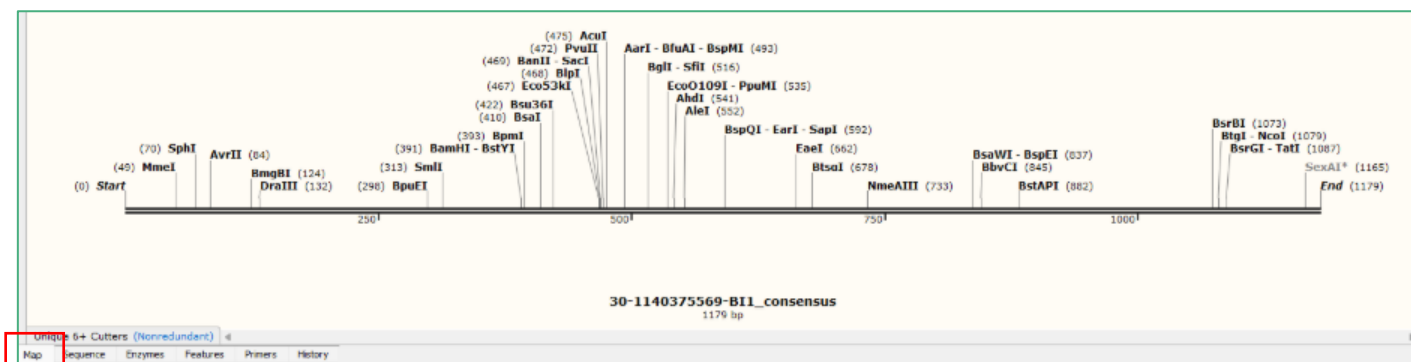
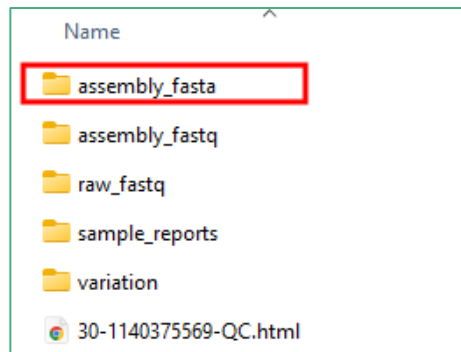
PCR-EZ クイックスタートガイド

2. PCR配列の確認

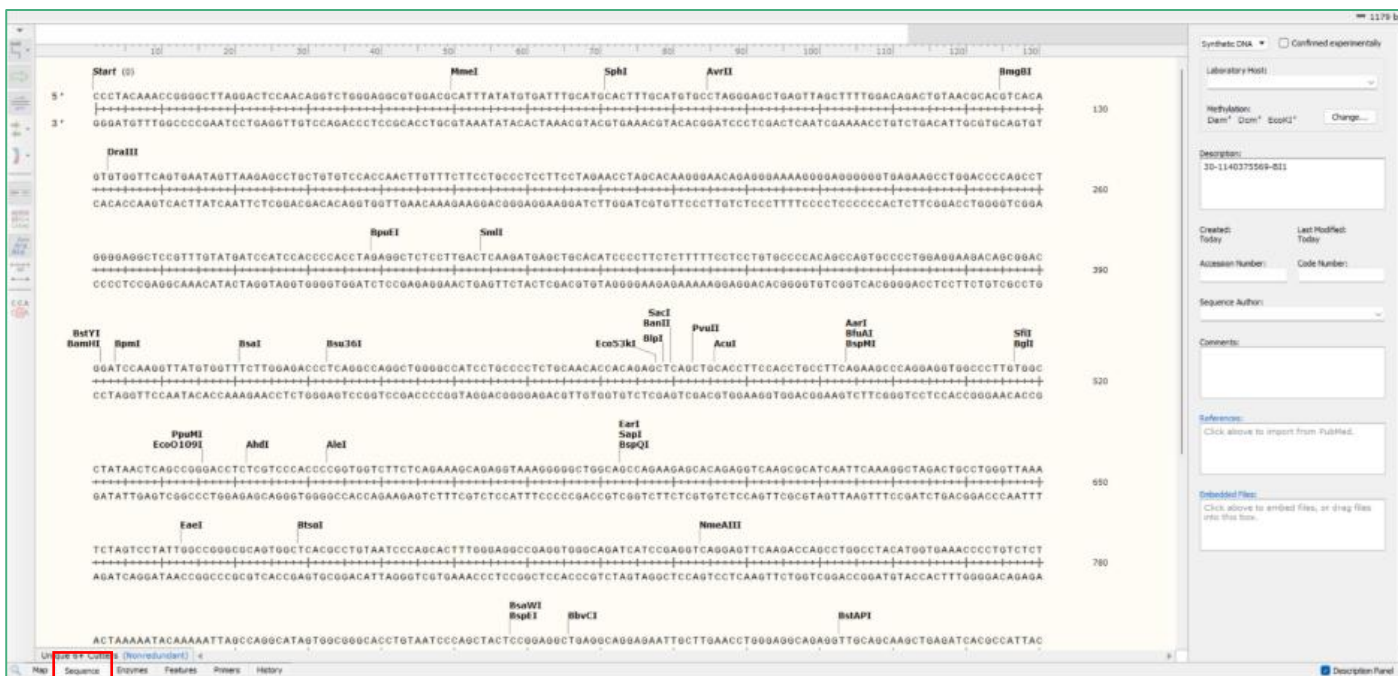
Fastaファイルの確認

アセンブルされた PCR 配列を確認するには、右側に表示されている assembly_fasta フォルダを開きます。

配列の閲覧には SnapGene Viewer、または任意の FASTA ビューアを使用できます。アセンブリ結果は、下図のように「Map」タブ（赤枠）で確認でき、アセンブルされた配列全体の長さや制限酵素部位をチェックできます。



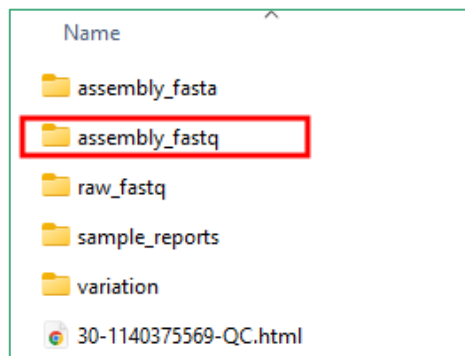
また、「Sequence」タブ（下の赤枠）をクリックすると、ヌクレオチド配列が表示されます。



PCR-EZ クイックスタートガイド

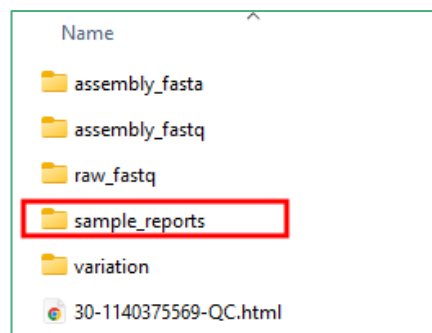
Fastqファイルの確認

各塩基ごとの信頼度（Qスコア）を含むFASTQファイルも参照可能です。このファイルはSnapGene Viewerなどの類似ツールで確認できます。バーが高いほど、その位置の塩基コールの信頼度が高いことを示します。逆にバーが低い場合は、低品質のシーケンスデータである可能性や、その部位に多型（ポリモーフィズム）が存在する可能性を示唆しています。



3. サンプル品質の評価

サンプルレポートには、各サンプルの品質を評価するための詳細な図が含まれています。



PCR-EZ クイックスタートガイド

リード長分布の確認

データ品質管理（QC）には、RAW リードの長さや品質ごとの分布が表示されます。下のグラフは、RAW リードの長さ分布を示したものです。理想的には、この分布は、予想される全長に達するまで、おおむね均一なリード数を示すことが望まれます。この例では、1,300 bp の PCR フラグメントに対する RAW リードの長さ分布が描かれています。

3 DATA QUALITY CONTROL

3.1 Raw data QC

Figure 3.1 shows the length distribution of the raw reads.

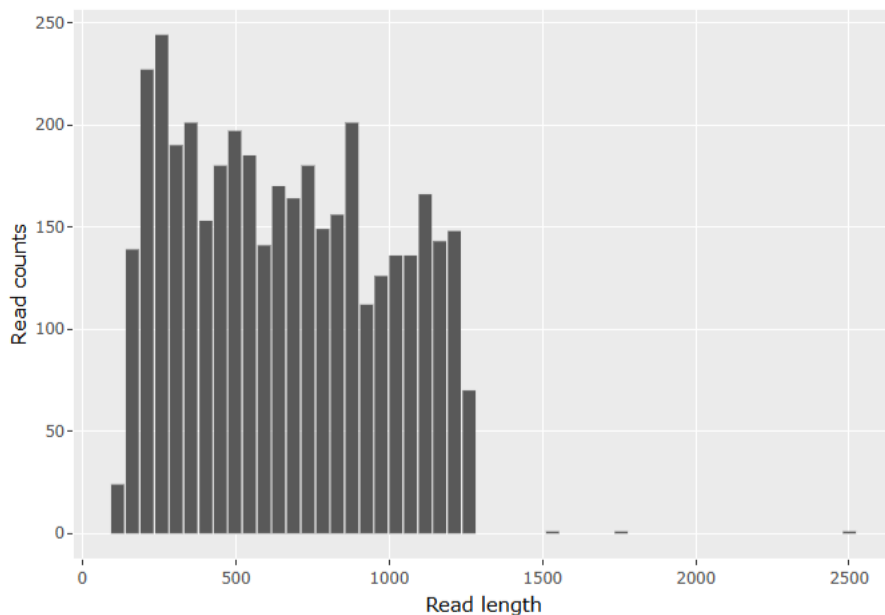


Figure 3.1: Length distribution of the raw reads

PCR-EZ クイックスタートガイド

品質（クオリティ）分布の確認

下図は、RAW リードの品質分布の一例です。理想的なグラフでは、右側に向かってピークが現れ、ほとんどのリードが高品質であることを示します。

Figure 3.2 shows the quality distribution of the raw reads.

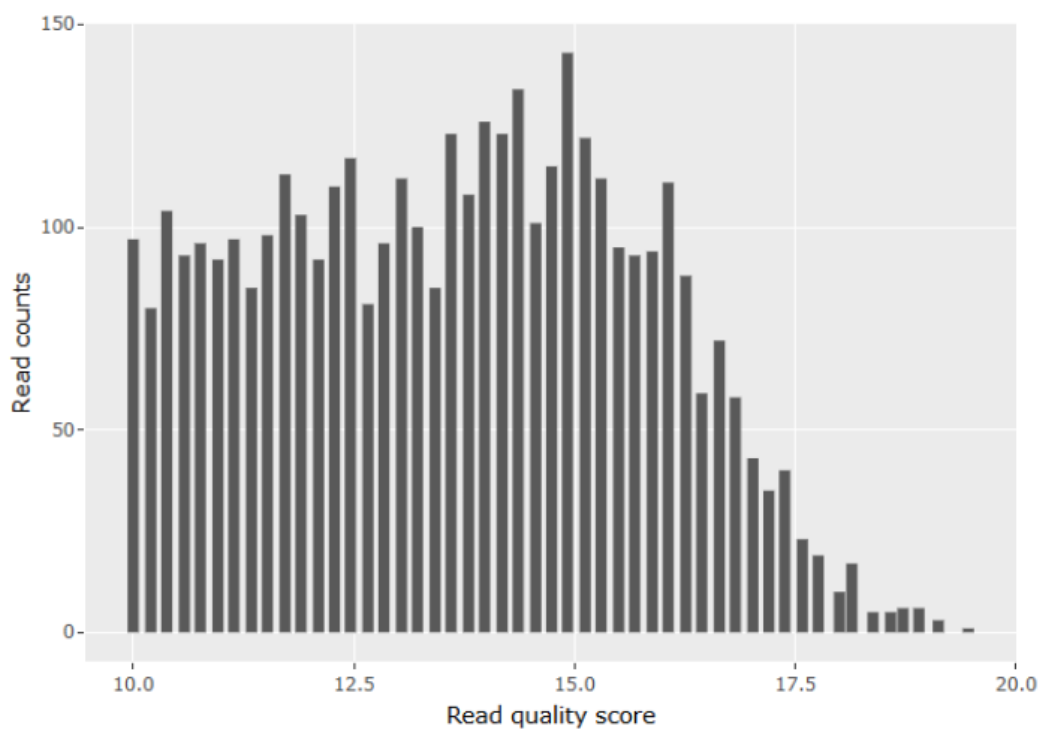


Figure 3.2: Quality distribution of the raw reads

PCR-EZ クイックスタートガイド

各塩基位置のカバレッジスコアの評価

Assembly QC グラフは、アセンブルされたコンセンサス配列全体にわたるカバレッジを表示します。通常、カバレッジはフラグメントの両端で低く、中央部分で高くなります。以下の例では、およそ 50–1,100 bp の位置でカバレッジが 400 を超えており、両端ではややカバレッジが低下していることがわかります。中央領域内に谷（dip）が見られる場合は、その位置でカバレッジが低下していることを意味します。

3.2 Assembly QC

Reads used for assembly and polishing were aligned to the final consensus sequence. Figure 3.3 shows the coverage along the polished consensus sequence.

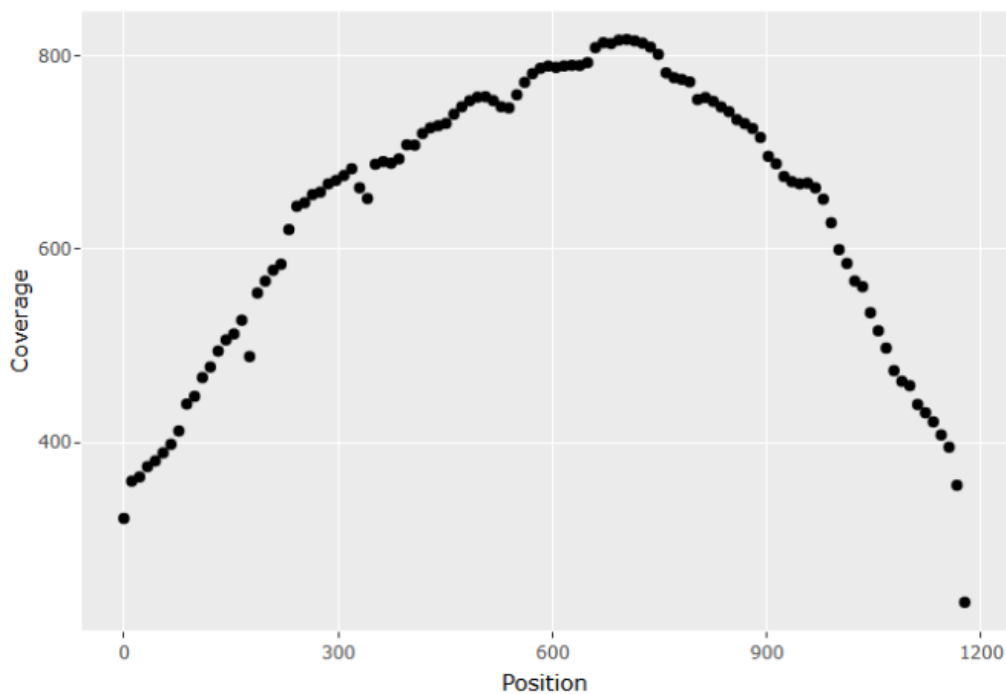
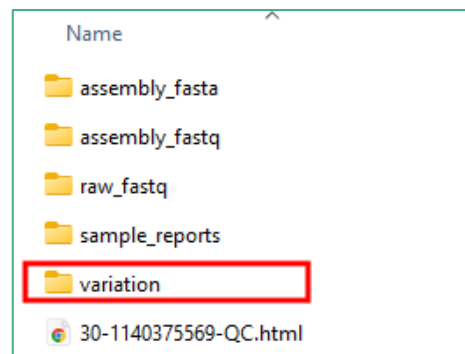


Figure 3.3: Coverage of the polished consensus sequence

4. バリエントコール

リードのうち 10% を超える割合で別の塩基（second nucleotide）が存在する位置は、vc.tsv ファイルに記載されています。下の例では、アセンブリ内の 4 つの位置で代替塩基コールが発生していることが示されています。AD 列には、その位置における「呼び出された塩基」と「代替塩基」のリード数の比率が示されています。AF 列は、その位置での混合塩基コールの割合（%）を示しています。



CHROM	POS	REF	ALT	GT	DP	AD	AF
30-1140375569-B11	218	A	G	0/1	580	439,136	0.2345
30-1140375569-B11	545	T	C	0/1	750	349,380	0.5067
30-1140375569-B11	551	C	T	0/1	758	596,160	0.2111
30-1140375569-B11	1009	C	T	0/1	594	299,282	0.4747

base_count.tsv ファイルには、各位置における各塩基のリード数が一覧で記載されています。また、挿入（insertion）や欠失（deletion）がある位置についても、それぞれの挿入・欠失に対するリード数が表示されます。

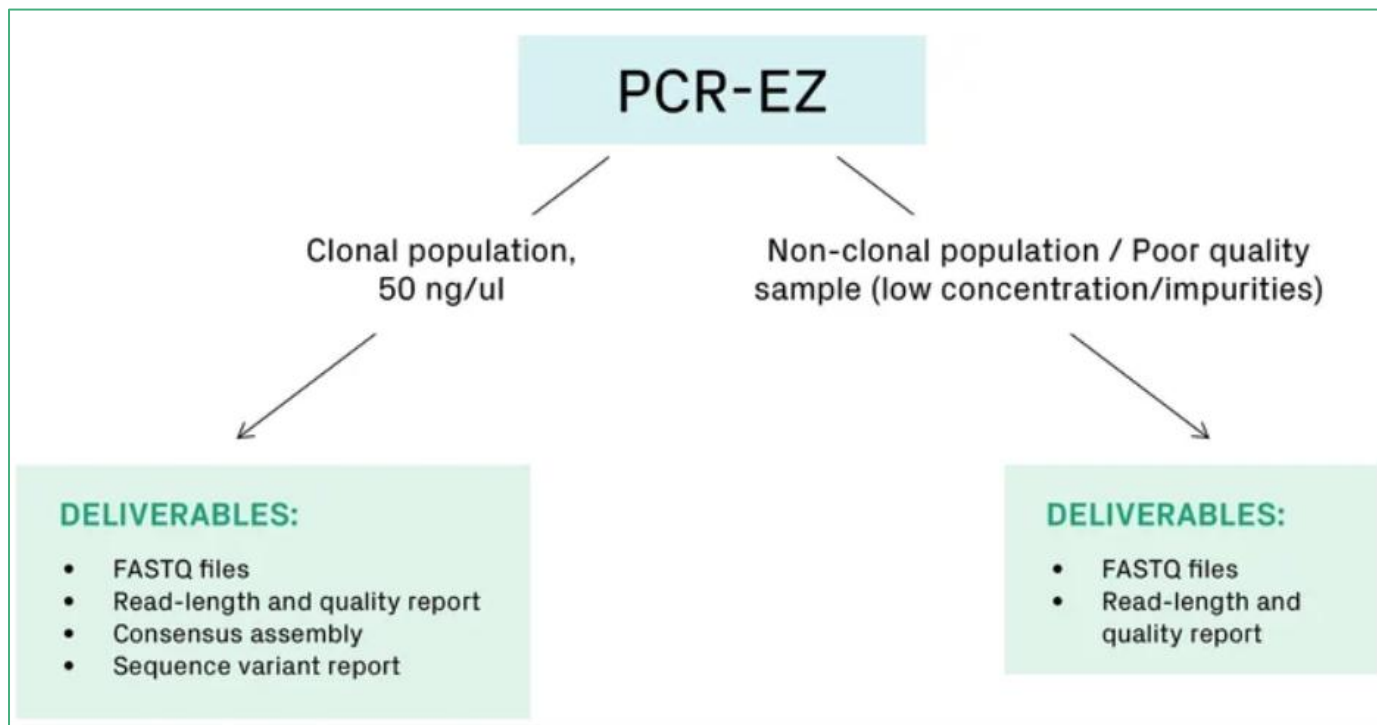
Position	Reference	Coverage	A	T	G	C	N	Insertions	Top Insert	Deletions
1	C	216	0	0	0	216	0	0 -	0	
2	C	275	0	0	0	275	0	0 -	0	
3	C	311	0	0	0	311	0	0 -	0	
4	T	317	1	311	2	3	0	3 G (3)	0	
5	A	337	337	0	0	0	0	0 -	0	
6	C	341	0	0	0	341	0	1 A (1)	0	
7	A	348	348	0	0	0	0	0 -	0	
8	A	348	348	0	0	0	0	0 -	0	
9	A	349	348	0	0	0	0	0 -	1	
10	C	349	1	0	0	345	0	0 -	3	

5. アセンブリ成功のポイント

濃度とクローンDNA

迅速な結果提供と低コストを実現するため、当社ではサンプルがアセンブリに失敗した理由を調べるための QC（品質管理）は実施していません。しかし、最も一般的な失敗理由は、サンプル濃度が必要条件である 50 ng/μl に達していないことです。濃度が低い場合、ライブラリー調製中に断片化が増える、あるいはサンプルに対して生成されるリード数が不足することがあります。そのため、サンプル濃度の確認と、失敗リスクを最小限に抑えるためのクローン DNA の使用を強く推奨します。

PCR-EZ でもクローン DNA が必須です。混合集団（mixed population）が含まれていると、アセンブリがうまくいかない原因となる可能性があります。アセンブリに失敗したサンプルについては、追加解析を行うよう、RAW の FASTQ ファイルを提供しています。



アセンブルの失敗

サンプルがアセンブリを生成できなかった場合、注文フォルダーにはRAW FASTQ リードと、RAW リード統計の概要を含む注文 QC レポートのみが保存されます。

