

Plasmid-EZ スタートガイド

納品データの解説およびよくあるご質問と回答

お問い合わせ先：



アゼンタ株式会社
次世代シーケンシング 受託サービス
〒142-0043 東京都品川区二葉二丁目9番15号 NFパークビルディング 4F
電話：03-6628-2950 メール：NGS.japan@azenta.com

最終更新日：11-22-2023 R1.1

目次

1. はじめに
 - プロジェクトレポート
 - サンプルレポート
2. プラスミド配列の見方
 - コンティグ配列のアノテーション
 - Snap Gene Viewerでの確認
 - シークエンス、アミノ酸配列、アノテーションの見方
 - エクセルのアノテーションの見方
 - FASTAファイルの開き方
3. クオリティ値の確認
 - リードの長さおよびクオリティ値の分布
 - ベース毎のクオリティ値
4. バリエントコール
 - ベース毎のバリエント
 - 複数コンティグによるアノテーション
5. アセンブリ不良の要因について
6. よくあるご質問と回答

1. はじめに

初めに、xx-xxxxxxx-QC.htmlというファイルを開いてください。
このファイルはQC結果のまとめと各サンプルへのリンクが含まれています。

pGEM	2023/10/30 9:54	ファイル フォルダー	
01-00000011-QC.html	2023/10/30 9:54	Microsoft Edge HTML D...	2 KB
Plasmid-EZ Bioinformatics FAQ.pdf	2023/10/30 9:54	Microsoft Edge PDF Do...	218 KB

QCレポートには各サンプルごとのQC結果が記載されております。
サンプルの名前をクリックすると、詳細な結果を確認することができます。

1. Quality assessment of sequencing:

Sample	Min_length	Mean_length	Max_length	Q1_length	Median_length	Q3_length	N50_length	Min_qscore	Mean_qscore	Ma
pGEM	124	2923.69	12119	3230	3277	3291	3281	9.05	14.59	22.

2. Quality assessment of assembly:

The following table reports statistics on mapping of raw sequencing reads to the *de novo* assembled contig.

Sample	Total_Reads	Mapped_reads	Unmapped_reads	Supplementary_mappings*
pGEM	7689	7673 (99.79%)	16 (0.2%)	6610 (85.96%)

* "Supplementary" reads are those reads that span the 3-prime - 5-prime boundary of the linearized contig sequence.
Calculation of percentages are based on the "Total_Reads" value reported.

また、個々のサンプルのデータは各サンプルのフォルダに格納されております。

pGEM	2023/10/30 9:54	ファイル フォルダー	
01-00000011-QC.html	2023/10/30 9:54	Microsoft Edge HTML D...	2 KB
Plasmid-EZ Bioinformatics FAQ.pdf	2023/10/30 9:54	Microsoft Edge PDF Do...	218 KB

個々のサンプルの詳細な結果については、xxx-AssemblyReport.htmlをご確認ください。

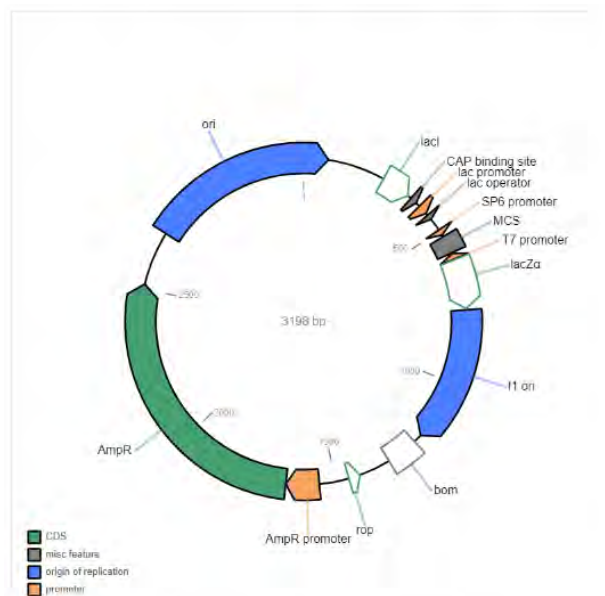
pGEM.fastq.gz	2023/10/30 9:54	GZ ファイル	18,458 KB
pGEM_contig.fasta	2023/10/30 9:54	FASTA ファイル	4 KB
pGEM_contig.fastq	2023/10/30 9:54	FASTQ ファイル	7 KB
pGEM_contig_annot.csv	2023/10/30 9:54	Comma Separated Value...	5 KB
pGEM_contig_annot.gbk	2023/10/30 9:54	GenBank DNA	9 KB
pGEM_contig_annot.html	2023/10/30 9:54	Microsoft Edge HTML D...	75 KB
pGEM_contig-readCounts-variation.csv	2023/10/30 9:54	Comma Separated Value...	119 KB
pGEM_contig-readCounts-variation.xls	2023/10/30 9:54	Microsoft Excel 97-2003...	606 KB
pGEM-AssemblyReport.htm	2023/10/30 9:54	Microsoft Edge HTML D...	84 KB
summary_pGEM.html	2023/10/30 9:54	Microsoft Edge HTML D...	4,678 KB

2. プラスミド配列の見方

xxx-AssemblyReport.htmlというファイルには、生成されたコンティグ配列についてのプラスミドマップが表示されています。

マウスを図に移動させるとその領域の説明が表示されます。

また、アノテーションされている領域は下部に表で記載されています。



Data extracted from pGEM_contig_annot.csv

Feature	Type	percent identity	percent match length	Description
f1 ori	rep_origin	100.0	100.0	f1 bacteriophage origin of replication; arrow indicates direction of (+) strand synthesis
AmpR promoter	promoter	100.0	100.0	bla
AmpR	CDS	99.76	100.0	β-lactamase; bla; confers resistance to ampicillin
ori	rep_origin	99.83	100.0	high-copy-number ColE1/pMB1/pBR322/pUC origin of replication
MCS	misc_feature	100.0	100.0	pUC18/19 multiple cloning site
lac promoter	promoter	100.0	100.0	promoter for the E. coli lac operon
CAP binding site	protein_bind	100.0	100.0	CAP binding activates transcription in the presence of cAMP. E. coli catabolite activator protein
T7 promoter	promoter	100.0	100.0	promoter for bacteriophage T7 RNA polymerase
SP6 promoter	promoter	100.0	100.0	promoter for bacteriophage SP6 RNA polymerase

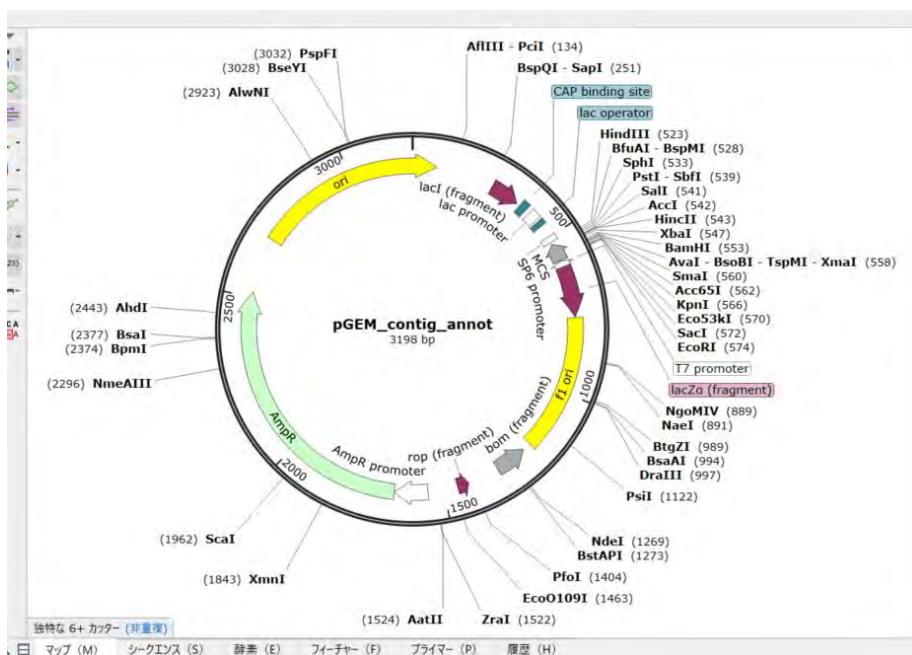
Gen Bankフォーマットでのアノテーションもxxx_contig_annot.gbkという名前で納品されています。このファイルはSnap Gene Viewerなどで開いてください。

Snap Gene Viewerのダウンロード先：<https://www.snapgene.com/snapgene-viewer>

pGEM.fastq.gz	2023/10/30 9:54	GZ ファイル	18,458 KB
pGEM_contig.fasta	2023/10/30 9:54	FASTA ファイル	4 KB
pGEM_contig.fastq	2023/10/30 9:54	FASTQ ファイル	7 KB
pGEM_contig_annot.csv	2023/10/30 9:54	Comma Separated Value...	5 KB
pGEM_contig_annot.gbk	2023/10/30 9:54	GenBank DNA	9 KB
pGEM_contig_annot.html	2023/10/30 9:54	Microsoft Edge HTML D...	75 KB
pGEM_contig-readCounts-variation.csv	2023/10/30 9:54	Comma Separated Value...	119 KB
pGEM_contig-readCounts-variation.xls	2023/10/30 9:54	Microsoft Excel 97-2003...	606 KB
pGEM-AssemblyReport.html	2023/10/30 9:54	Microsoft Edge HTML D...	84 KB
summary_pGEM.html	2023/10/30 9:54	Microsoft Edge HTML D...	4,678 KB

2. プラスミド配列の見方 (続き)

Gen Bankフォーマットのファイル (xxx_contig_annot.gbk) をSnap Gene Viewerで開くと、下記のようなプラスミドマップと制限酵素サイトの情報が表示されます。



画面下部にある、“シーケンス”タブをクリックすると、アノテーション情報と共に配列を確認することができます。



2. プラスミド配列の見方 (続き)

アノテーションの情報だけ確認する際には、xxx_contig_annot.csvファイルをご確認ください。

pGEM.fastq.gz	2023/10/30 9:54	GZ ファイル	18,458 KB
pGEM_contig.fasta	2023/10/30 9:54	FASTA ファイル	4 KB
pGEM_contig.fastq	2023/10/30 9:54	FASTQ ファイル	7 KB
pGEM_contig_annot.csv	2023/10/30 9:54	Comma Separated Value...	5 KB
pGEM_contig_annot.gbk	2023/10/30 9:54	GenBank DNA	9 KB
pGEM_contig_annot.html	2023/10/30 9:54	Microsoft Edge HTML D...	75 KB
pGEM_contig-readCounts-variation.csv	2023/10/30 9:54	Comma Separated Value...	119 KB
pGEM_contig-readCounts-variation.xls	2023/10/30 9:54	Microsoft Excel 97-2003...	606 KB
pGEM-AssemblyReport.html	2023/10/30 9:54	Microsoft Edge HTML D...	84 KB
summary_pGEM.html	2023/10/30 9:54	Microsoft Edge HTML D...	4,678 KB

エクセルで開くと下記のようにアノテーション情報が表形式で得られます。

sseqid	start	locat	end	locati	strand	percent	id full	length	length of	percent m	fragment	database	Feature	Type	Descriptio	sequence
f1_ori	763		1192		1	100		429	429	100	FALSE	snappgene	f1 ori	rep_origin	f1 bacteri	ACGCGCCC
AmpR_pro	1550		1655		1	100		105	105	100	FALSE	snappgene	AmpR pro	promoter	bla	CGCGGAAC
AmpR_(2)	1655		2516		1	99.768		861	861	100	FALSE	snappgene	AmpR	CDS	*i-lactam	ATGAGTAT
ori	2687		78		1	99.83		589	589	100	FALSE	snappgene	ori	rep_origin	high-copy	TTGAGATCC
MCS_(8)	522		579		-1	100		57	57	100	FALSE	snappgene	MCS	misc_feat	pUC18/19	GAATTCGA
lac_promo	401		432		1	100		31	31	100	FALSE	snappgene	lac promo	promoter	promoter	TTTACACTT
CAP_bind	365		387		1	100		22	22	100	FALSE	snappgene	CAP bind	protein_b	CAP bind	TAATGTGA
T7_promo	581		600		-1	100		19	19	100	FALSE	snappgene	T7 promot	promoter	promoter	TAATACGA
SP6_prom	497		516		1	100		19	19	100	FALSE	snappgene	SP6 prom	promoter	promoter	ATTTAGGTC
lac_oper	439		456		1	100		17	17	100	FALSE	snappgene	lac operat	protein_b	The lac re	TTGTGAGCC
lacZ_alpha	602		760		1	100		174	158	90.8046	TRUE	snappgene	lacZ#7	CDS	LacZ#7 fr	ATTCACTGC
bom	1219		1319		-1	100		141	100	70.92199	TRUE	snappgene	bom	misc_feat	basis of m	CTGGCTTA
lacI	260		353		1	100		1083	93	8.587258	TRUE	snappgene	lacI	CDS	lac repres	GCGCCCAA
rop	1420		1456		-1	100		192	36	18.75	TRUE	snappgene	rop	CDS	Rop prote	CTGCGCGG

エクセル内の各項目については以下のFAQファイル、または本資料のFAQのセクションをご参照ください。

pGEM	2023/10/30 9:54	ファイル フォルダー	
01-00000011-QC.html	2023/10/30 9:54	Microsoft Edge HTML D...	2 KB
Plasmid-EZ Bioinformatics FAQ.pdf	2023/10/30 9:54	Microsoft Edge PDF Do...	218 KB

また、コンティグなどの配列情報については、Fastaフォーマットで納品されております。メモ帳やMicrosoft Wordなどのテキストエディタでご確認ください。

```
pGEM_contig.fasta - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
>pGEM_contig1 len=3195
AGTCCTGTGCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGGCGI
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTCCAAACTGGAACAACA
GCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACAT
CGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGCGGGACAGGTATCCGGTAAGCGGI
```

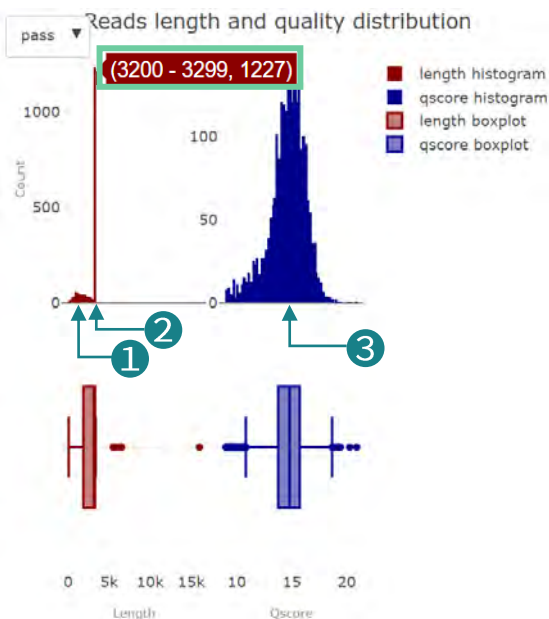
3. データ品質の確認

Assembly Report内の中段にはリードの長さや品質に関する統計値が記載されています。

2. Read Lengths

Summary statistics of read lengths.

read_type	min_length	mean_length	max_length	q1_length	median_length	q3_length	n50_length
all	98	2633.55	18209	1849.75	3266	3288	3277
pass	98	2633.37	16001	1858	3268	3289	3278



プロット左の赤いバーはリード長（x軸：鎖長 y軸：リード数）、プロット右の青いバーはクオリティ値（x軸：クオリティ値 y軸：リード数）を表しております。

プロットの任意の場所にマウスカーソルを合わせると統計値が表示されます。

プロット左、緑枠の例では、鎖長3200-3299 bpの配列が1227リード検出されたことを意味しています。

また、下部にはコンティグ配列へのマッピング率を含むアセンブリの統計値も記載されています。

3. Mapping of sequencing reads to the assembly

The following table reports statistics on mapping of raw sequencing reads to the *de novo* assembled contig.

Sample	Total_Reads	Mapped_reads	Unmapped_reads	Supplementary_mappings*
pGEM-10	2385	2370 (99.37%)	15 (0.62%)	1880 (78.82%)

* "Supplementary" reads are those reads that span the 3-prime - 5-prime boundary of the linearized contig sequence. Calculation of percentages are based on the "Total_Reads" value reported.

下記の結果を満たしていない場合、
ご提出のプラスミドDNAの濃度・品質をご確認ください。

- 鎖長の短いリードが比較的少ないこと。
- 最も豊富なリードの鎖長が実際のプラスミドDNAの長さとおおよそ一致していること。
- 大部分のリードのクオリティ値がQ15程度あるいはそれ以上であること。
- 生成されたコンティグに対し、多くのリードがマップされていること。

3. データ品質の確認（続き）

納品物の中には配列情報にクオリティ値が付随するFASTQフォーマットのファイルも含まれています。このファイルはテキストエディタもしくは、Snap Gene Viewerなどで開いてください。

pGEM.fastq.gz	2023/10/30 9:54	GZ ファイル	18,458 KB
pGEM_contig.fasta	2023/10/30 9:54	FASTA ファイル	4 KB
pGEM_contig.fastq	2023/10/30 9:54	FASTQ ファイル	7 KB
pGEM_contig_annot.csv	2023/10/30 9:54	Comma Separated Value...	5 KB
pGEM_contig_annot.gbk	2023/10/30 9:54	GenBank DNA	9 KB
pGEM_contig_annot.html	2023/10/30 9:54	Microsoft Edge HTML D...	75 KB
pGEM_contig-readCounts-variation.csv	2023/10/30 9:54	Comma Separated Value...	119 KB
pGEM_contig-readCounts-variation.xls	2023/10/30 9:54	Microsoft Excel 97-2003...	606 KB
pGEM-AssemblyReport.html	2023/10/30 9:54	Microsoft Edge HTML D...	84 KB
summary_pGEM.html	2023/10/30 9:54	Microsoft Edge HTML D...	4,678 KB

FASTQファイルをSnap Gene Viewerで開くと、下記のようにクオリティ値を各ベース毎に確認することができます。

棒グラフはクオリティを表しており、高いほどクオリティが高いことを示しております。



4. バリエントコール

各ベースにおけるバリエントの結果はエクセルファイルで納品されております。

pGEM.fastq.gz	2023/10/30 9:54	GZ ファイル	18,458 KB
pGEM_contig.fasta	2023/10/30 9:54	FASTA ファイル	4 KB
pGEM_contig.fastq	2023/10/30 9:54	FASTQ ファイル	7 KB
pGEM_contig_annot.csv	2023/10/30 9:54	Comma Separated Value...	5 KB
pGEM_contig_annot.gbk	2023/10/30 9:54	GenBank DNA	9 KB
pGEM_contig_annot.html	2023/10/30 9:54	Microsoft Edge HTML D...	75 KB
pGEM_contig-readCounts-variation.csv	2023/10/30 9:54	Comma Separated Value...	119 KB
pGEM_contig-readCounts-variation.xls	2023/10/30 9:54	Microsoft Excel 97-2003...	606 KB
pGEM-AssemblyReport.html	2023/10/30 9:54	Microsoft Edge HTML D...	84 KB
summary_pGEM.html	2023/10/30 9:54	Microsoft Edge HTML D...	4,678 KB

各ベースのバリエントが、各塩基のリード数として表示されています。

10%以上のバリエントが存在する場合、黄色で強調されます。

また、バリエント一覧の表には塩基の挿入(Insertion)や欠損(Deletion)も確認できます。

	A	B	C	D	E	F	G	H	I	J	K
1	AZENTA LIFE SCIENCES - Plasmid-EZ										
2											
3	Color Legend:										
4		>= 10% deviation									
5	These data relate to *raw* sequencing reads										
6	Position	Reference	Coverage	A	T	G	C	N	Insertions	Top Insertion	Deletions
7	1	T	8000	0	8000	0	0	0	0	-	0
8	2	C	8001	0	0	0	8001	0	0	-	0
9	3	A	8002	8002	0	0	0	0	0	-	0
10	4	C	8003	0	0	0	7911	0	3	G (1)	0
11	5	C	8004	0	0	0	8001	0	2	CA (1)	0
12	6	C	8005	0	0	0	7983	0	6	AG (1)	0
13	7	A	8006	7987	0	0	0	0	4	C (1)	3
14	8	G	8003	0	0	7977	0	0	9	AA (1)	1
15	9	A	8007	7916	0	0	0	0	1	G (1)	2
16	10	A	8004	7999	0	0	0	0	1	GG (1)	0
17	11	A	8010	8004	0	0	0	0	3	CT (1)	2
18	12	C	8007	0	0	0	7982	0	3	T (1)	2
19	13	G	8008	0	0	7977	0	0	2	C (1)	2
20	14	C	8009	0	0	0	7941	0	3	A (1)	1
21	15	T	8010	0	7986	0	0	0	8	GCG (1)	3
22	16	G	8008	0	0	7958	0	0	4	C (1)	4
755	749	A	7974	7022	0	0	0	0	5	AC (1)	5
756	750	G	7933	0	0	7648	0	0	10	AT (1)	3
757	751	A	7966	7797	0	0	0	0	9	C (1)	2
758	752	C	7949	0	0	0	7800	0	15	GGG (1)	2
759	753	A	7964	7668	0	0	0	0	4	AG (1)	2
760	754	G	7947	0	0	7055	0	0	11	A (1)	2
761	755	A	7967	7948	0	0	0	0	6	AT (1)	2
762	756	T	7966	0	7955	0	0	0	11	TC (1)	2
763	757	C	7964	0	818	0	6962	0	6	GCT (1)	3
764	758	G	7973	0	0	7928	0	0	5	A (1)	3
765	759	C	7977	0	0	0	7933	0	4	T (1)	2
766	760	T	7980	0	7779	0	0	0	20	TG (1)	5
767	761	G	7969	0	0	7554	0	0	11	T (1)	4
768	762	A	7800	7546	0	0	0	0	3	T (1)	2
769	763	G	7956	0	0	7615	0	0	6	A (1)	5

4. バリエントコール (続き)

アセンブリの結果、複数のコンティグ配列が作られ場合はxxx_allContigs.fastaというファイルが納品されております。これは、すべてのコンティグ配列を含んでいるファイルです。

PGEM.fastq.gz	5/16/2023 6:42 PM	GZ File
PGEM_allContigs.fasta	5/16/2023 6:42 PM	FASTA DNA
PGEM_longestContig.fasta	5/16/2023 6:42 PM	FASTA DNA
PGEM_longestContig.fastq	5/16/2023 6:42 PM	FASTQ Sequence Trace
PGEM_longestContig_annot.csv	5/16/2023 6:42 PM	Microsoft Excel Comma Separat...
PGEM_longestContig_annot.gbk	5/16/2023 6:42 PM	GenBank DNA
PGEM_longestContig_annot.html	5/16/2023 6:42 PM	Chrome HTML Document
PGEM_longestContig-readCounts-variation.csv	5/16/2023 6:42 PM	Microsoft Excel Comma Separat...
PGEM_longestContig-readCounts-variation.xls	5/16/2023 6:42 PM	Microsoft Excel 97-2003 Worksh...
PGEM-AssemblyReport.html	5/16/2023 6:42 PM	Chrome HTML Document
summary_PGEM.html	5/16/2023 6:42 PM	Chrome HTML Document

Snap Gene Viewerで開くと各contigのリードの長さを確認することができます。

File Name
PGEM.tig00000003 len=3197 reads=125 class=contig suggestRepeat=yes suggestBubble=no suggestCircular=yes trim=0-3197
PGEM.tig00000004 len=3146 reads=1 class=contig suggestRepeat=yes suggestBubble=no suggestCircular=no trim=0-3146

上記の場合、二つのコンティグ配列が検出されています。

125リード得られたコンティグはsuggestCircularがyesとなっておりますが、もう一方はnoとなっております環状にはなっていません。

そのため、二番目のコンティグは、短いリードで生成された不十分なコンティグの可能性がります。

一方で、複数のコンティグが多くリードによって生成された場合は、バリエントの可能性がります。

複数のコンティグが生成された場合、最も多くのリードがマップされたコンティグを信頼性が高いものと判断し、アノテーションやバリエント解析を実施します。

その他のコンティグのアノテーション情報が知りたい場合は、Plannotate

(<http://plannotate.barricklab.org/>)にてご確認ください。このサイトではGenBankフォーマットのファイルとCSVファイルを出力します。

少量のリードによって複数のコンティグが生成された場合、DNAサンプルの品質が低い、あるいはDNA溶液の濃度が低い可能性があります。

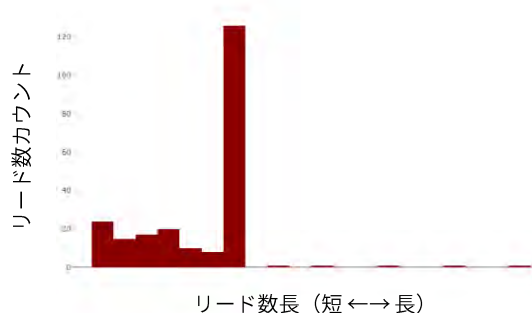
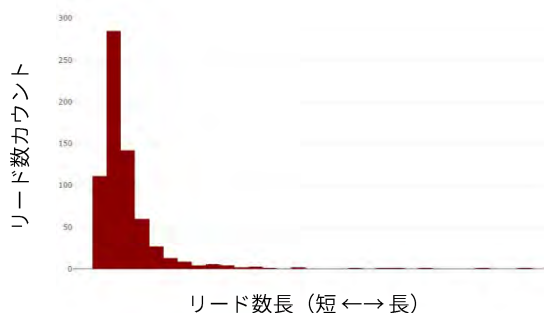
5. アセンブリ不良の要因について

アセンブリができなかった場合、フォルダ内には得られたリードのfastqとサマリーのレポートのみが含まれます。



Plasmid-EZのサービスでは、低価格で迅速にデータを納品させていただくため、サンプルのQCは行っておりません。そのため、アセンブリ不良の要因の究明について原則ご対応はいたしかねます。

一般に、サンプルの濃度が適正でない場合にアセンブリが上手くいかないことがあります。例えば、左下の図は得られたリードが短く、アセンブリすることができませんでした。理想的には右下の図のように長いリードが多く得られることが適切なアセンブリに必要とされます。



規定 (50 ng/μl) より低いDNA濃度でご提出された場合、シーケンシングライブラリの調製に使用するDNA量が相対的に減少します。その場合、ライブラリ調製の酵素反応の工程で、より高い頻度でDNAの断片化を生じることがあります。

より良好な結果を得るために、下記サンプル提出ガイドラインをご参考にプラスミドDNAサンプルのご用意をお願いします。

- サンプルタイプ：精製済み環状プラスミドDNA
- サイズ・塩基長：最大25 kb
- 濃度：50 ng/μl（弊社では濃度調製はいたしません）
- 濃度測定方法：QubitやPiroGreen等の蛍光強度に基づくものを強く推奨
- 液量：10 μl以上
- 純度 (OD260/280)：1.8-2.0
- 懸濁液：DNase/RNaseフリー水、Tris-HCl溶液（10 mM、pH 8.5前後）あるいはLow TE (<0.1M EDTA)

6. よくあるご質問と回答

1. 納品物にはどのようなファイルが含まれますか？

- htmlファイル
 - プラスミドのマッピングイメージ
 - アノテーションの表
 - リードの長さおよびクオリティ値
- プラスミドのアセンブリ結果(.fasta)
- クオリティ付きプラスミドのアセンブリ結果(.fastq)
- GenBankフォーマットのアセンブリ結果(.gbk)
- プラスミドアセンブリのアノテーション結果(.csv)
- バリエーションコールの結果(.csv と .xls)
- 得られたリード (.fastq.gz)

2. プラスミドの混合物を提出できますか？

このサービスでは保証しておりません。

3. コンティグが複数生成されることはありますか？

アセンブリした結果、コンセンサス配列が一つになった場合は、xx_conting.fastaに記載されます。

複数配列作られた場合には、xx_allContings.fastaにすべての配列が記載され、そのうち最も多くのリードがマップされた配列が、xx_selectedContig.fastaに記載されます。アノテーションおよびバリエーション一覧の結果は、xx_selectedContig.fastaの配列に対してのみ解析を実施、納品されます。

4. Variation.csvに記載の項目名はなんですか？

Position: 位置情報

Reference: 生成されたコンティグ配列の塩基

Coverage: 各ポジションに配置された塩基の個数

A/T/G/C/N: 塩基の種類

Insertions: そのポジションで検出された挿入の数

Top Insertion: 最も検出された挿入の数

Deletions: 欠失の数

6. よくあるご質問（続き）

5. Annotation.csvに記載の項目はなんですか？

sseqid: アノテーションID

Start location: コンティグ配列での各アノテーションの開始位置

End location: コンティグ配列での各アノテーションの終了位置

Strand: スtrandの方向

Percent Identity: 領域内の特徴と一致する塩基の割合

Full length of feature in db: データベースから得られたアノテーションの長さ

Length of found feature: アセンブリから得られたアノテーションの長さ

Percent match length: データベースから得られた長さに対してアセンブリ結果のアノテーションの長さの割合

Fragment: アノテーション情報がデータベースに登録された配列の一部に基づくかどうか (FALSEの場合、full length of feature in dbとlength of found featureが実質同じ、一方、TRUEの場合、full length of feature in dbの一部分のみがlength of found featureに一致していることを意味)

Database: データベース名

Feature: アノテーションの名前

Type: 配列の機能的な分類

Description: 説明

Sequence: 各アノテーションの生成されたコンティグでの配列

6. なぜ、アセンブルできなかつたのでしょうか？

資料の5. アセンブルできなかつた場合をご参照ください。

多くの場合、ご提出のプラスミドDNAの濃度が不足、フェノール等の不純物が含まれることによるライブラリ調製時の酵素反応不良、ニックや断片化が進んでいる、などの理由によります。

7. 生成されたコンティグと想定していた配列の長さが異なるのですがなぜですか？

kb単位の長いリピート配列を含むプラスミドDNAでは、アセンブルの不具合で、生成されたコンティグ配列から一部分が脱落していたり、2コピーがつながってコンティグが生じるなどの現象が、わずかですが確認されています。アセンブルのアルゴリズムの修正を進めています。

6. よくあるご質問（続き）

8. 想定していた配列と異なる配列が含まれていました。なぜでしょうか？
一般的な次世代シーケンシングでの手法と同様、ライブラリ調製時にバーコーディング（インデックス配列の付加）を行い、他のサンプルとプールしてランをします。ごく低頻度ですが、インデックスの読み取りエラーにより、一部のリードが誤ったサンプルとして認識されることがあります。
また、本アプリケーションで採用しているOxford Nanopore Technologiesのプラットフォームでは、その仕様として、フローセルを洗浄して複数回再利用することになっています。0.1%程度の低い割合ですが、前回までにランを行ったライブラリの配列が検出されることがあります。詳細はOxford Nanopore Technologies提供の下記の資料をご参照ください。
<https://store.nanoporetech.com/flow-cell-wash.html>
9. リニア状の配列は解析できますか？
本Plasmid-EZのアプリケーションの対象は、環状のプラスミドDNAとなります。
10. 25 kbを超えるサイズのプラスミドDNAは解析できますか？
本Plasmid-EZのアプリケーションでは25 kbまでの環状プラスミドDNAを対象にしています。これを超えるサイズにつきましては、弊社NGSカスタマーサポート（NGS.Japan@azenta.com）までご相談ください。
11. 複数種のプラスミドが混在したサンプルは解析できますか？
クローン化された配列を対象としています。複数のプラスミドが混在したサンプルもご提出は可能ですが、結果不良のリスクをご了承いただいた上での実施となります。
12. 濃度調製は必須ですか？
はい。あらかじめ50 ng/μlに調製した、精製済みのプラスミドDNAをご提出ください。原則弊社の方では希釈等の濃度調製操作はいたしません。
13. 濃度や品質確認はしてもらえますか？
本サービスは、費用対効果・短納期を重視したものになり、サンプル受領後の品質確認は実施しておりません。

6. よくあるご質問（続き）

14. 濃度測定にQubit/PicoGreen等の蛍光強度に基づいた手法を推奨とあります。NanoDropしか持っていないのですがどうすればよいでしょうか？
NanoDropでは、DNA/RNAの区別ができず、多くの場合、濃度を過剰に見積もる傾向があります（ゆえに、ライブラリ調製時のインプット量として相対的に低下することになります）。NanoDrop等の吸光度での測定法しか利用できない場合、80 ng/ μ lに調製をしてご提出ください。なお、結果に影響するリスクをご了承の上での実施となります。
15. 精製前のグリセロールストック、大腸菌プレートを提出することは可能ですか？
NGS.Japan@azenta.comまでお問合せください。
なお、遺伝子組み換え体の輸送となるため、国内カルタヘナ法に準じた事前の情報提供ならびに三重梱包等への対応が義務となります。
16. プラスミド調製は行ってもらえますか？
配列解析後、特定のプラスミドについてプラスミド調製サービスをご依頼いただくことが可能です。マイクログラムスケールからミリグラムスケールまで対応しており、アニマルフリー条件でのプラスミド調製、エンドトキシンフリー条件でのプラスミド調製等に対応しております。詳しくは、Project.Japan@azenta.comまでお問合せください。
17. アノテーションに使用しているデータベースは何ですか？また、それは外部のデータベースを参照していますか？
SnapGeneで採用されているデータベースを用いています。弊社内あるいは弊社管理のクラウド環境にスタンドアロンのデータベースを構築し、それに対し検索を行っております。

6. よくあるご質問（続き）

18. Plasmid-EZサービスの配列の正確性はどれくらいですか？

各リードの精度は、通常大部分がQ15前後で、これはエラー率3% = 100 bp読んだときにおおよそ3 bpの読み取りエラーを生じる = という品質になります。各リードの精度は一般的なショートリード型の次世代シーケンサーと比較すると高くはありませんが、数百から数千リードからなる、同じ領域の配列を重ねてコンセンサス配列を得ることで、大部分の領域についてQ50-60程度の精度を実現しています。

また、逆位反復配列など、エアピループ構造を取りやすい配列などに対しても高い信頼度で配列の取得が可能です。

一方、数十bpの連続した塩基（poly-A）では、正確な連続塩基数を取得することが困難であることが示されています。また、長鎖のリピートを含む配列では、アセンブルの工程でのエラーを生じる可能性があり、生成されたコンティグ配列が正確でないことがあります。

プラスミドDNAの濃度や純度が適切でない場合も、取得リード長やリード数により、アセンブルの精度が低下することもあります。

19. FASTQファイルは何でしょうか。どのように開くことが可能でしょうか？

FASTQファイルは、塩基配列情報と各塩基のクオリティスコアを含むフォーマットです。しばしば `xx.fastq.gz` という圧縮された形式で保存されています。ファイルを直接開くためには、Linux等のコマンドラインをご使用いただくか、解凍後にテキストエディタやSnapGene Viewerを使用する方法があります。

20. どのシーケンシングプラットフォームを使用されていますか？

Oxford Nanopore Technologiesのロングリードシーケンサーを使用しています。