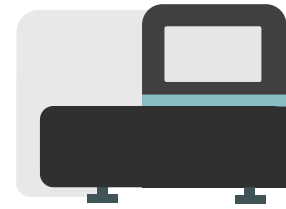


# Genome Analysis

## 生データからDNA配列を組み立てる

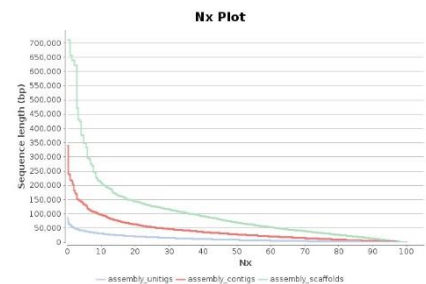
NGSより出力されるデータからクオリティチェックを実行し、リードをフィルタリングし、低品質の塩基を削除します。リファレンスゲノムの無い生物種でもDe-Novo Assemblyによって配列を組み立てることができます。現在、ショートリード・ロングリード・ショートリードとロングリードを組み合わせたハイブリッドアセンブリ（バクテリア）の解析が可能です。



**Illumina・IonTorrent  
PacBio・Nanopore**

## 統計

さまざまな統計チャートとレポートにより、ゲノムの組み立てと特性評価ができ、生物学的解釈の結果を支援することができます。

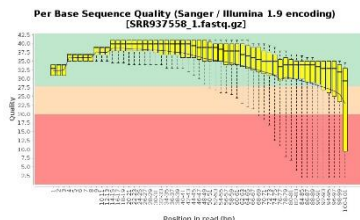


## 探索的分析

豊富なユーザーインターフェイスにより、データ量が多いゲノムアノテーションでも簡単に処理できます。遺伝子アノテーションは、フィルタリング、ソート、および他の結果セットと組み合わせることができます。表から直接表示する遺伝子を選択します。

SeqID	Source	Type	Start	End	Score	Strand	Frame	Acc ID	Gene Name	db Name
21	Prokaryote	gene	21114638	21114729	1	+	1	GeneID:100000000	gene1	Prokaryote
21	Prokaryote	mRNA	21114638	21114729	1	+	1	GeneID:100000000	gene1	Prokaryote
21	Prokaryote	5'UTR	21114638	21114638	1	+	1	GeneID:100000000	gene1	Prokaryote
21	Prokaryote	CDS	21114638	21114729	1	+	1	GeneID:100000000	gene1	Prokaryote
21	Prokaryote	3'UTR	21114729	21114729	1	+	1	GeneID:100000000	gene1	Prokaryote

また得られた配列情報はFunctional Analysisモジュールを使用してさらに機能情報を付与することが可能です。



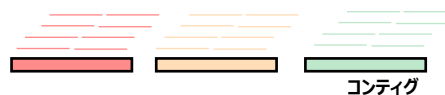
## Quality Control

FastQCとTrimmomaticを使用して品質管理を行います。低品質の塩基やアダプター除去、クオリティチェックができます。



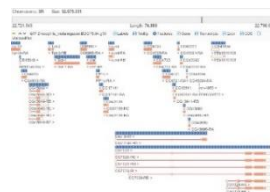
## Gene Finding

原核生物(Glimmer)および真核生物(Agustus)のORF遺伝子予測を行います。真核生物はリピート配列と低複雑度領域のDNA配列をマスクできます。



## DNA-seq De-Novo Assembly

ABYSS、SPAdesを使用したショートリードまたはハイブリッドアセンブリ、Flyeを使用したロングリードの配列を組み立てます。リファレンスを必要とせずに全ゲノム配列を構築することができます。



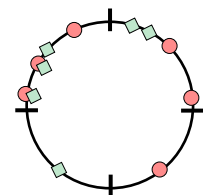
## Genome Browser

アノテーションをトラックの形で視覚化して、ゲノム配列(.fasta)とアライメント(.bam)、イントロン - エクソン構造(.gff)、などを組み合わせます。



## Alignment & Polishing

BWAを使用して短いリードをリファレンスゲノムにマッピングできます。さらにPilonを使用して作成したマッピングデータを修正します。



## MLST

7つのハウスキーピング遺伝子の内部フラグメントのシーケンスを使用して、細菌種の分離株を特徴付けることができます。



# Genome Analysisモジュールで何ができる？

ゲノムデータに合わせたアルゴリズムを選択できます！！

DNAシーケンスは、DNA内の核酸配列を決定するプロセスであり、ゲノムを特徴付けることができる技術です。現在のDNAシーケンシングテクノロジーはゲノム全体を一度に読み取ることにはできませんが、テクノロジーに応じて20~30,000塩基の小さな断片を読み取ります。元のシーケンスを再構築するには、シーケンスアセンブリ技術が必要です。

ゲノムアセンブリとは、シーケンシング中に生成される多数のDNAフラグメントを取得し、元のゲノムを再構築するなどのバイオインフォマティクスプロセスを指します。同様の生物のゲノムが以前に配列決定されている場合、目的のゲノムは、この既知のゲノムと比較することにより組み立てられます。しかし、組み立てられるゲノムが、以前に配列決定された他の生物と類似していない場合、De-Novo Assembly戦略が一般的に行われます。

現在までに、多数のゲノムアセンブラが開発されています。アセンブラは、overlap-layout consensus (OLC) とハミルトン路、およびde Bruijn graphとオイラー路などに分類されます。さらに、配列技術向けに様々なアセンブラが設計されており、小さな（細菌）または大きな（真核）ゲノムのアセンブリなど、特定のニーズに合わせて調整されています。

OmicsBoxでは3つのアセンブリ戦略が利用可能です。

## Results Overview

Statistic	scaffolds	contigs
Number of Contigs (>= 0 bp)	13,051	13,302
Number of Contigs (>= 1000 bp)	1,270	1,449
Number of Contigs (>= 5000 bp)	961	1,098
Number of Contigs (>= 10000 bp)	807	916
Number of Contigs (>= 25000 bp)	594	643
Number of Contigs (>= 50000 bp)	389	388
Total Length (>= 0 bp)	58,906,803	58,713,232
Total Length (>= 1000 bp)	57,385,604	57,176,150
Total Length (>= 5000 bp)	56,626,754	56,309,997
Total Length (>= 10000 bp)	55,534,404	55,011,827
Total Length (>= 25000 bp)	52,027,676	50,528,481
Total Length (>= 50000 bp)	44,540,916	41,305,262
Number of Contigs	1,592	1,758
Largest Contig	769,910	646,233
Total Length	57,600,900	57,385,255
GC (%)	67.84	67.84
N50	100,779	86,721
N75	53,632	44,600
L50	166	194
L75	364	425
Number of N's per 100 kbp	314.12	0

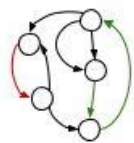
OmicsBoxのDe-Novo Assembly結果はFASTAファイルに加えて、シーケンスアセンブリの品質に関する統計レポートとチャートが生成されます。



ABxSS (ショートシーケンスによるアセンブリ) は、ショートリード用に設計されたde novoシーケンスアセンブラです。de Bruijn Graphを表すために、確率的データ構造であるブルームフィルターを使用するアルゴリズムを実装しています。ABxSSは大きなゲノムを組み立てることができます。



SPAdesは、de Bruijn Graphに基づくさまざまなアセンブリパイプラインを含むアセンブリツールです。SPAdesは小さなゲノム用に設計されており、細菌や単一細胞のゲノムを組み立てることができます。さらにSPAdesはIlluminaおよびIonTorrentのデータとPacBio、Oxford Nanopore、およびSangerのリードを組み合わせたハイブリッドアセンブリを実行することで信頼性の高いアセンブルデータを取得できます。



Flyeは、PacBioやOxford Nanopore Technologiesにデータなど、単一分子シーケンスリード用のde novoシーケンスアセンブラです。小さな細菌プロジェクトから哺乳類規模の大規模アセンブリまで、幅広いデータセット用に設計されています。

Genome Analysisモジュールではゲノムデータに最適なアルゴリズムで計算を行うことが可能です。

## 手間を省き正確に評価

## 高速計算

OmicsBoxは様々なオープンソースのソフトウェアを統合しているのでツール同士の行き来なく手間がかかりません。さらに負荷のかかる計算はクラウドにて高速計算されるため素早く結果を得ることができます。得られたデータは図や表を通して評価してから下流分析に臨めます。



De Novo Assemblyで得られたコンティグは配列情報のみの状態です。OmicsBoxでは配列情報だけを用いてゲノム上のORF遺伝子の位置を予測できます。さらにFunctional Analysisモジュールと組み合わせることで配列の機能情報などゲノム配列から様々な有用な情報を引き出すことが可能です。