

NGSデータ解析入門Webセミナー： De Novoシーケンス解析編

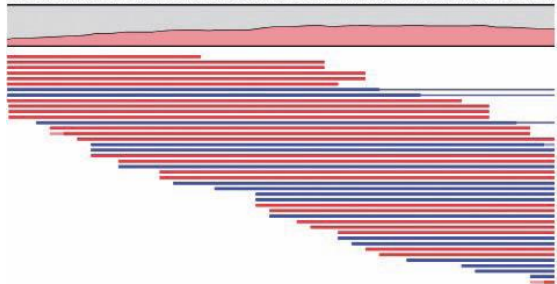
NGS新規ゲノム配列解析の手順



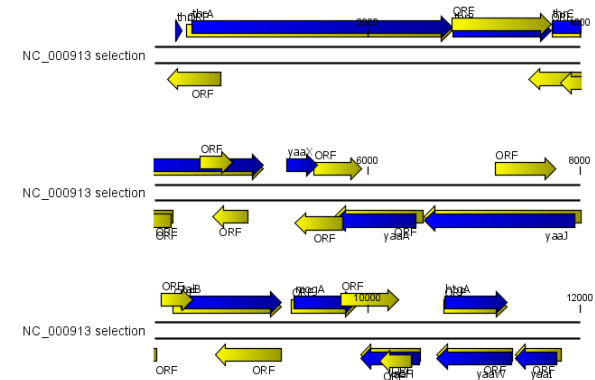
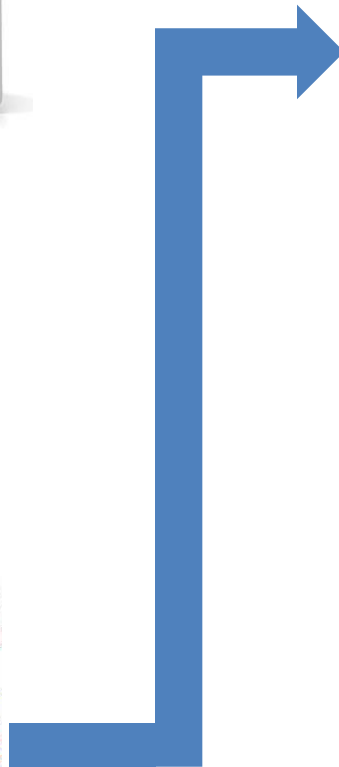
✓ シークエンス



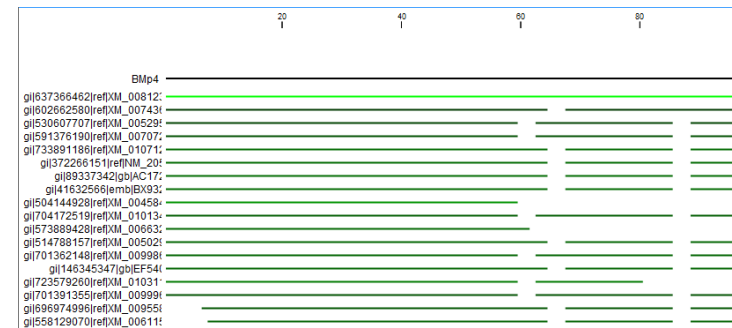
GCCGACAATAATGACGGGAATGGTGTAAAGCCATCACGCCA



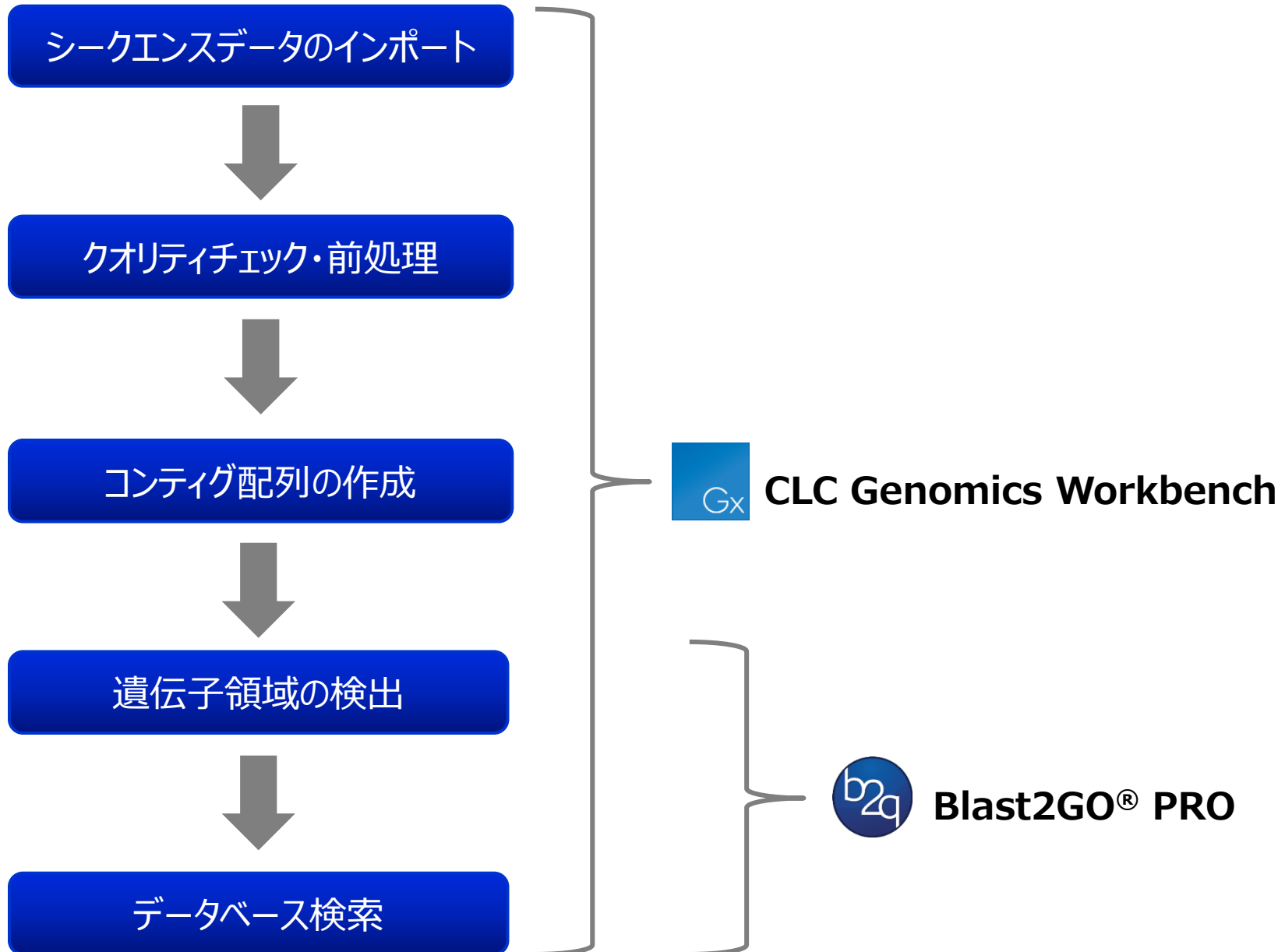
✓ アセンブル



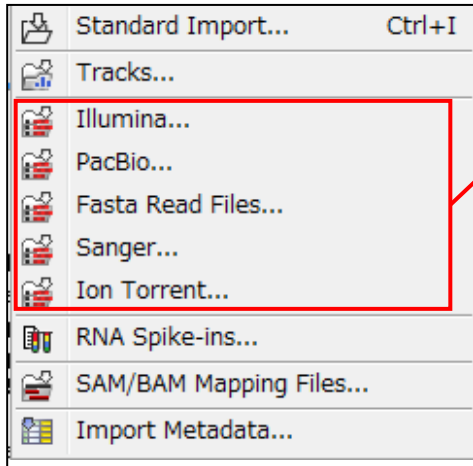
✓ 遺伝子領域の検出



✓ データベース検索



CLC Genomics Workbench

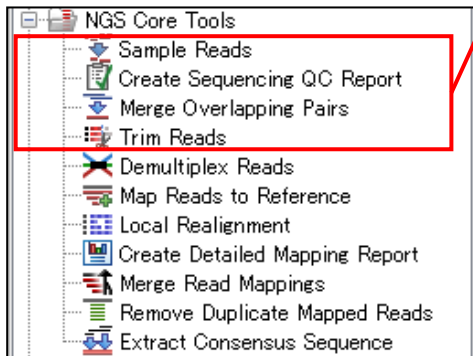


シークエンスデータのインポート

- NGS data import

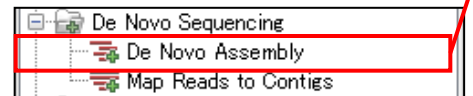
クオリティチェック・前処理

- Sample Reads
- Create Sequencing QC Report
- Merge Overlapping Pairs
- Trim Reads



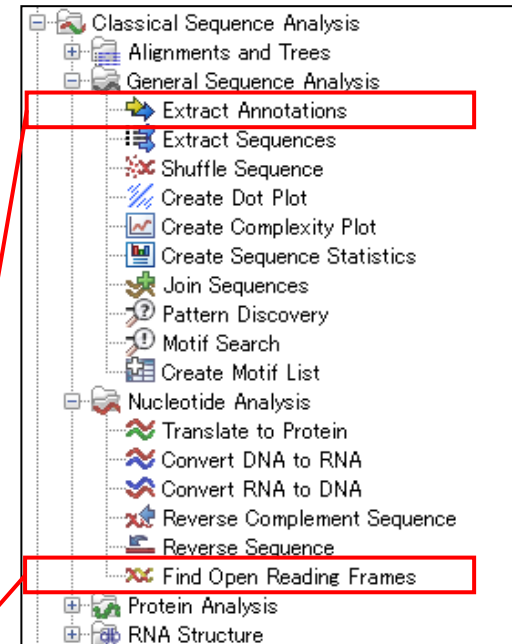
コンティグ配列の作成

- De Novo Assembly



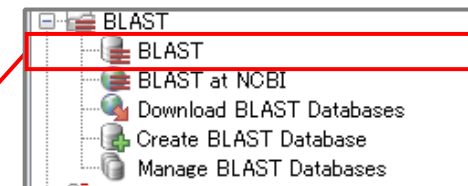
遺伝子領域の検出

- Find Open Reading Frames
- Extract Annotations

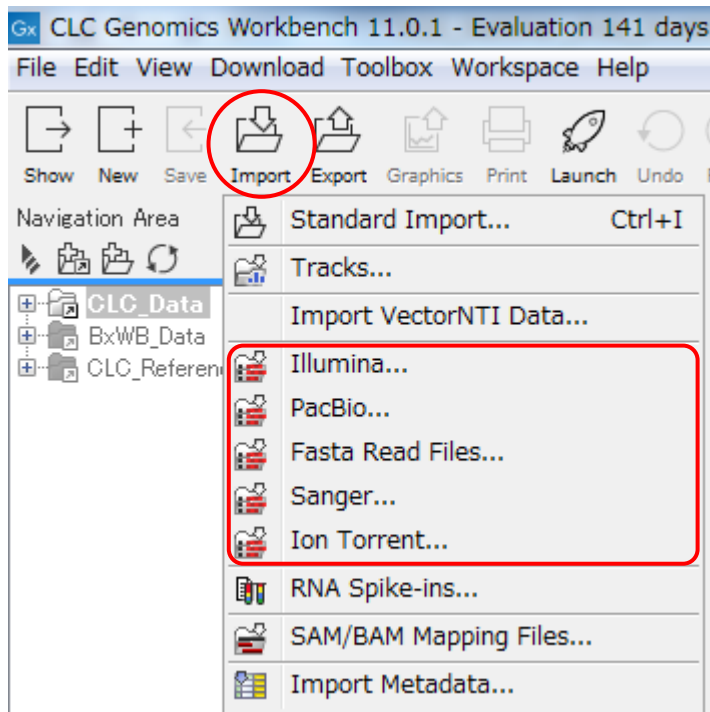


データベース検索

- BLAST



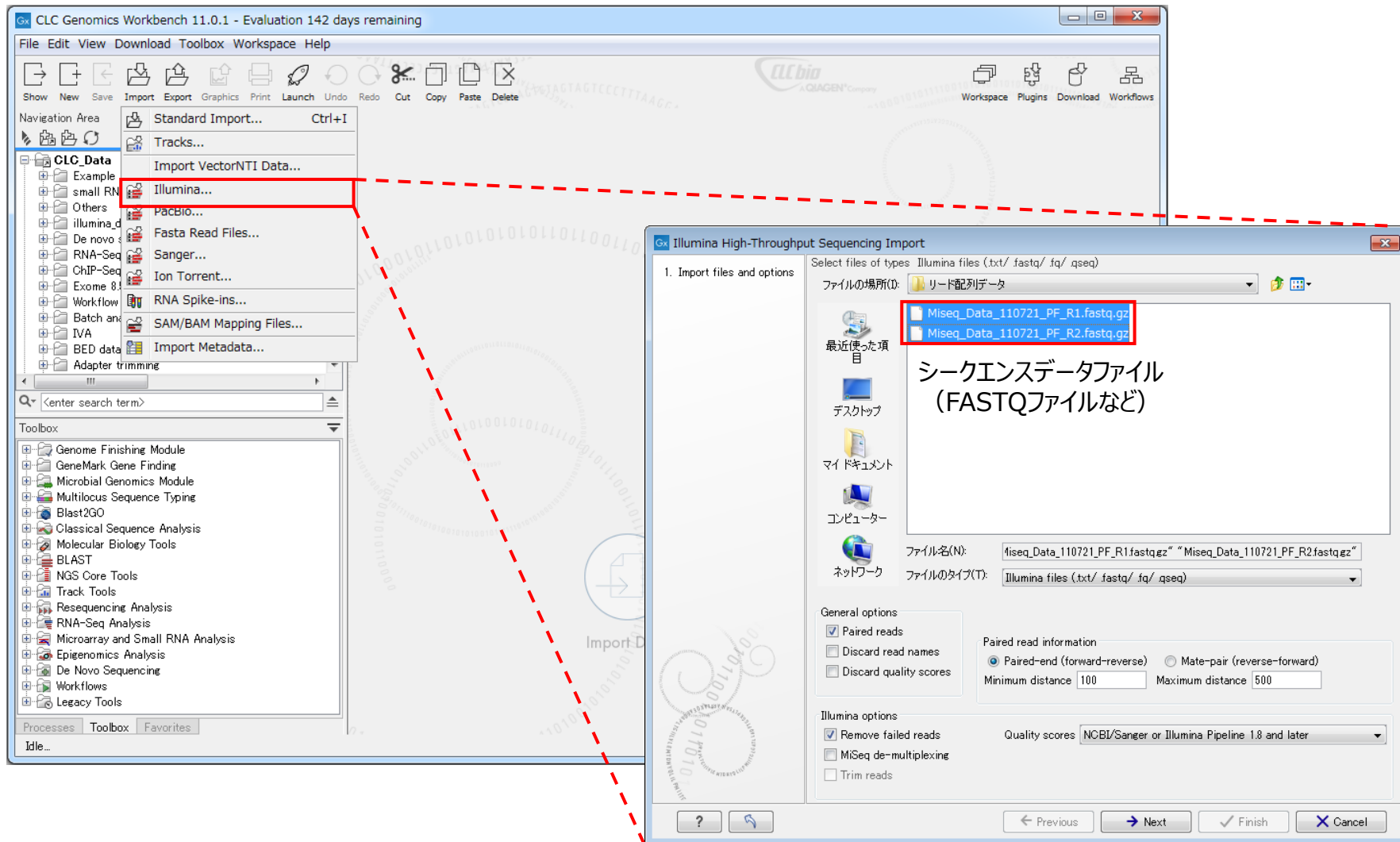
- ✓ CLC Genomics Workbench, Biomedical Genomics Workbenchともに、シーケンサー機種やファイルフォーマットに合せたインポートメニューを利用可能
- ✓ ToolbarのImportアイコンから表示されるインポーターから選択して、インポートを実行



プラットフォーム	ファイル形式
Illumina	.txt .fastq .fq .qseq
PacBio	.bas.h5/ .bax.h5 .fastq .fq .fasta .fa .fna
Ion Torrent[※]	.sff .fastq .fq

※Ion TorrentのUnmapped BAMファイルは、Standard Importよりインポートを行う

High-Throughput Sequencing Import

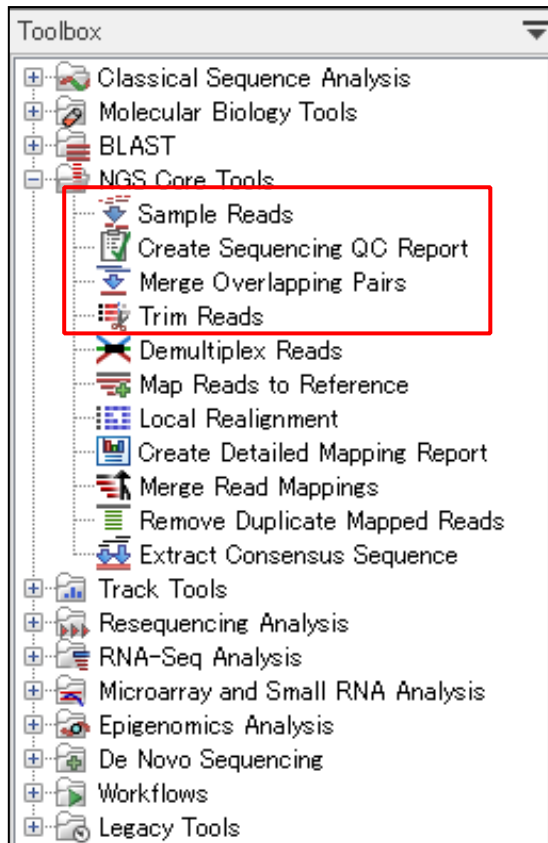


- ✓ シークエンサー機種などに合わせてメニューを選択し、シークエンスデータファイルを選択
- ✓ ペアエンドシークエンスデータのインポートにも対応

The screenshot displays the CLC Genomics Workbench 11.0.1 interface. The main window shows a list of sequence reads with their corresponding quality scores. The reads are displayed in a grid format, with the sequence and quality scores for each read. The quality scores are represented by a bar chart below the sequence. The reads are labeled with coordinates and strand information, such as 12001:1533 1:N:0:0 and 12001:1533 2:N:0:0. The interface includes a navigation area on the left, a toolbox at the bottom left, and a settings panel on the right. The settings panel is titled 'Sequence List Settings' and includes options for sequence layout, double stranded, numbers on sequences, relative to, numbers on plus strand, hide labels, lock labels, sequence label, annotation layout, show annotations, position, offset, label, show arrows, use gradients, restriction sites, motifs, residue coloring, nucleotide info, find, and text format.

- ✓ シーケンスデータがインポートされ、各種解析に使用できるようになる
- ✓ 各リードの塩基配列やクオリティスコアを確認できる

- ✓ クオリティチェックでは、インポートしたシーケンスデータに対して、クオリティチェックレポートの作成や、低クオリティリードの除去などを行う
- ✓ 必要に応じて、ペアリードの結合や、シーケンスデータのサンプリングなどの、コンティグ作成前の前処理を行う



クオリティチェック用ツール

Create Sequencing QC Report

- インポートしたシーケンスデータのクオリティやPCR Duplicate の状況などを確認するためのレポートを作成

Trim Reads

- アダプターの除去、クオリティスコアによる除去、長さを指定した除去などを選択・組み合わせて、リードのトリミングを実行

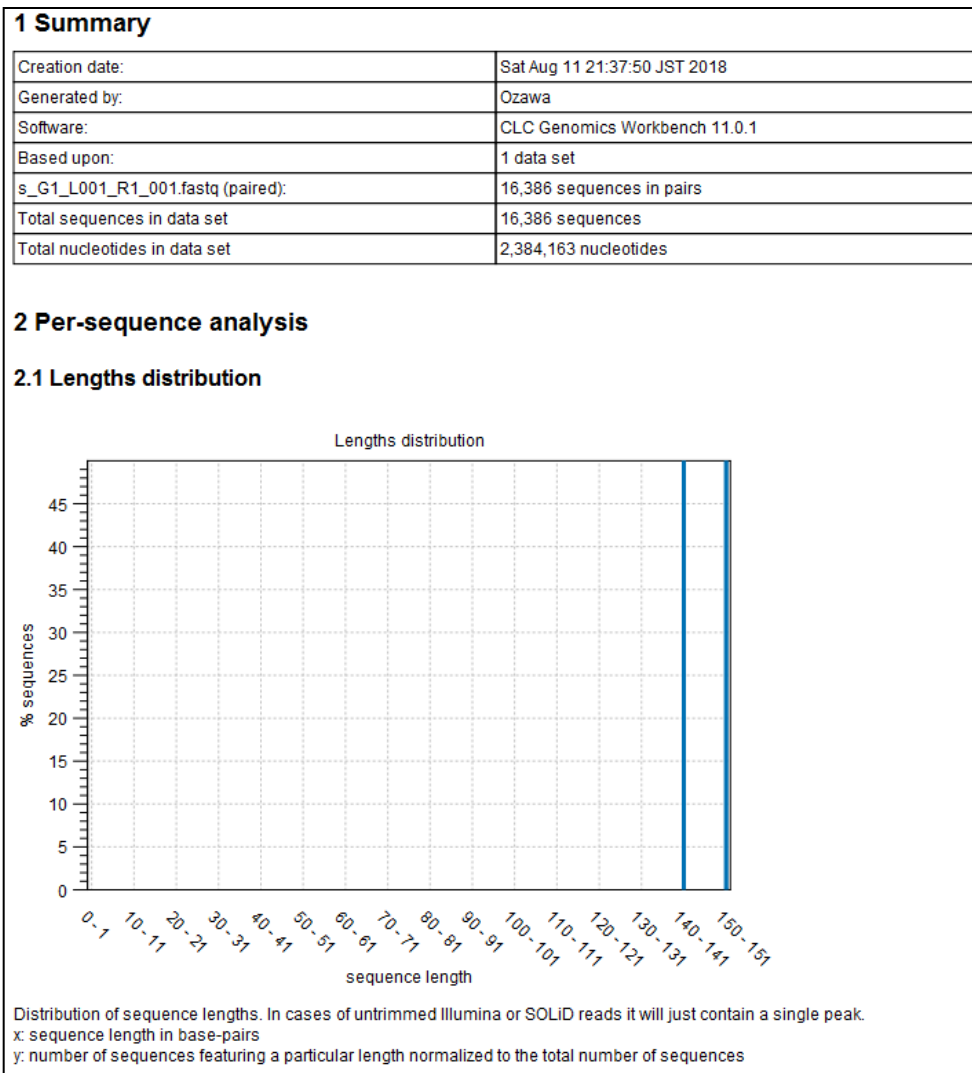
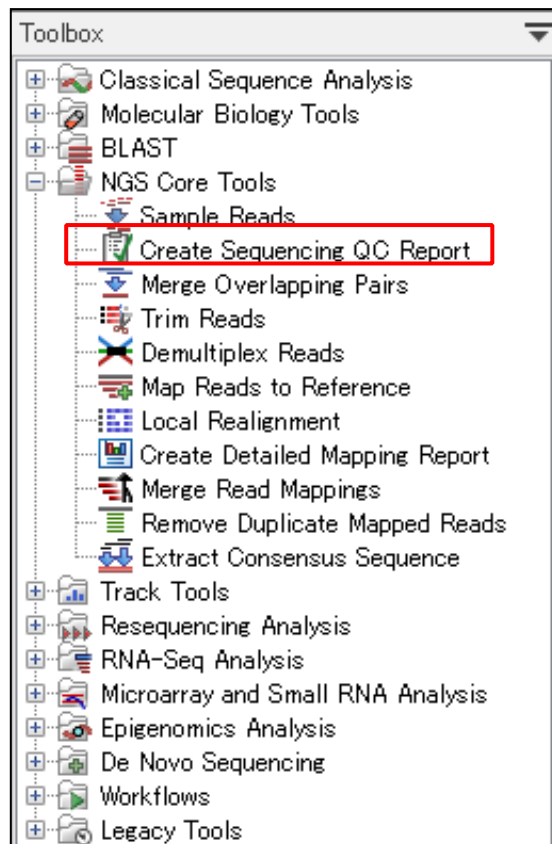
前処理用ツール

Sample Reads

- シーケンスデータのサンプリングを行う

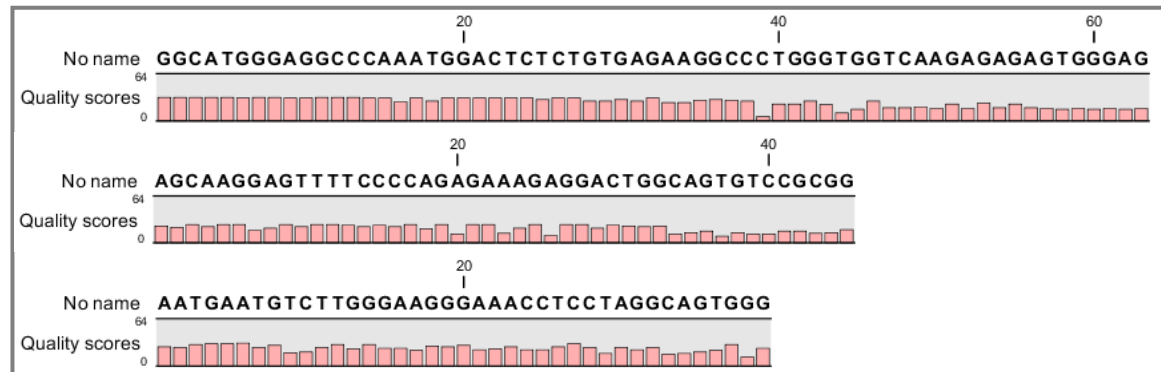
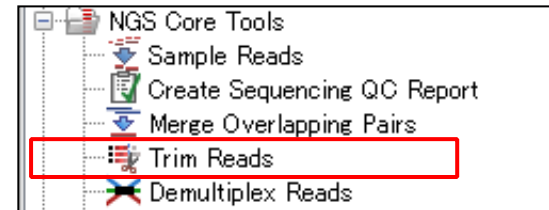
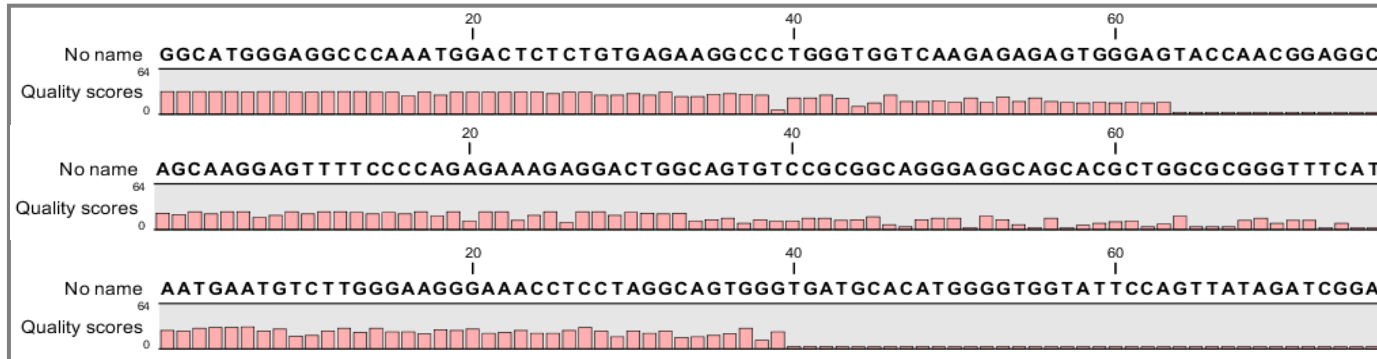
Merge Overlapping Pairs

- オーバーラップしているペアリード同士の間を結合を行う

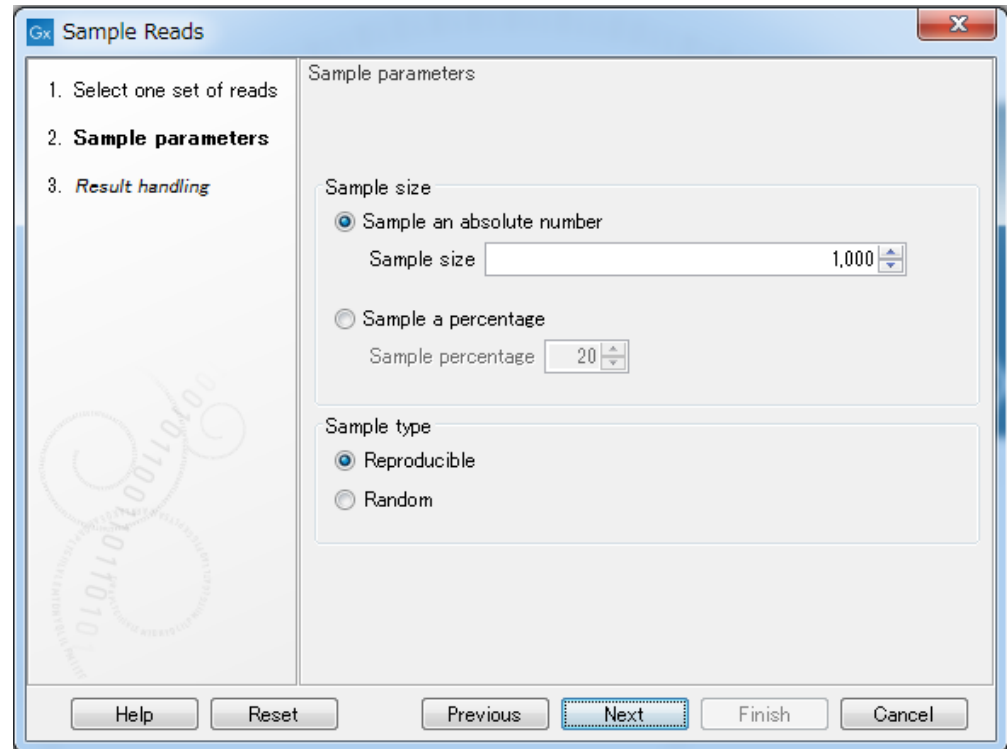
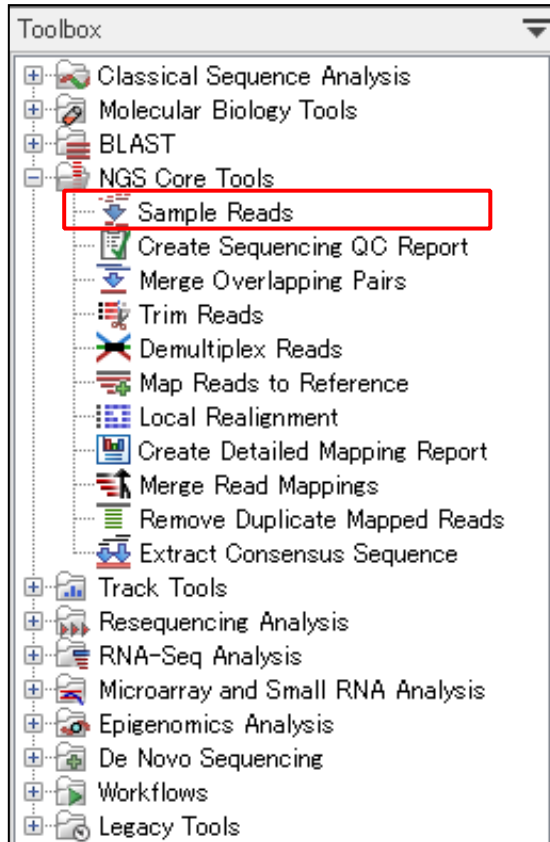


- ✓ Create Sequencing QC Reportでは、シーケンスデータのクオリティ情報をまとめたレポートが作成される
- ✓ GC含量やクオリティスコア分布などのグラフデータや数値データを確認が可能

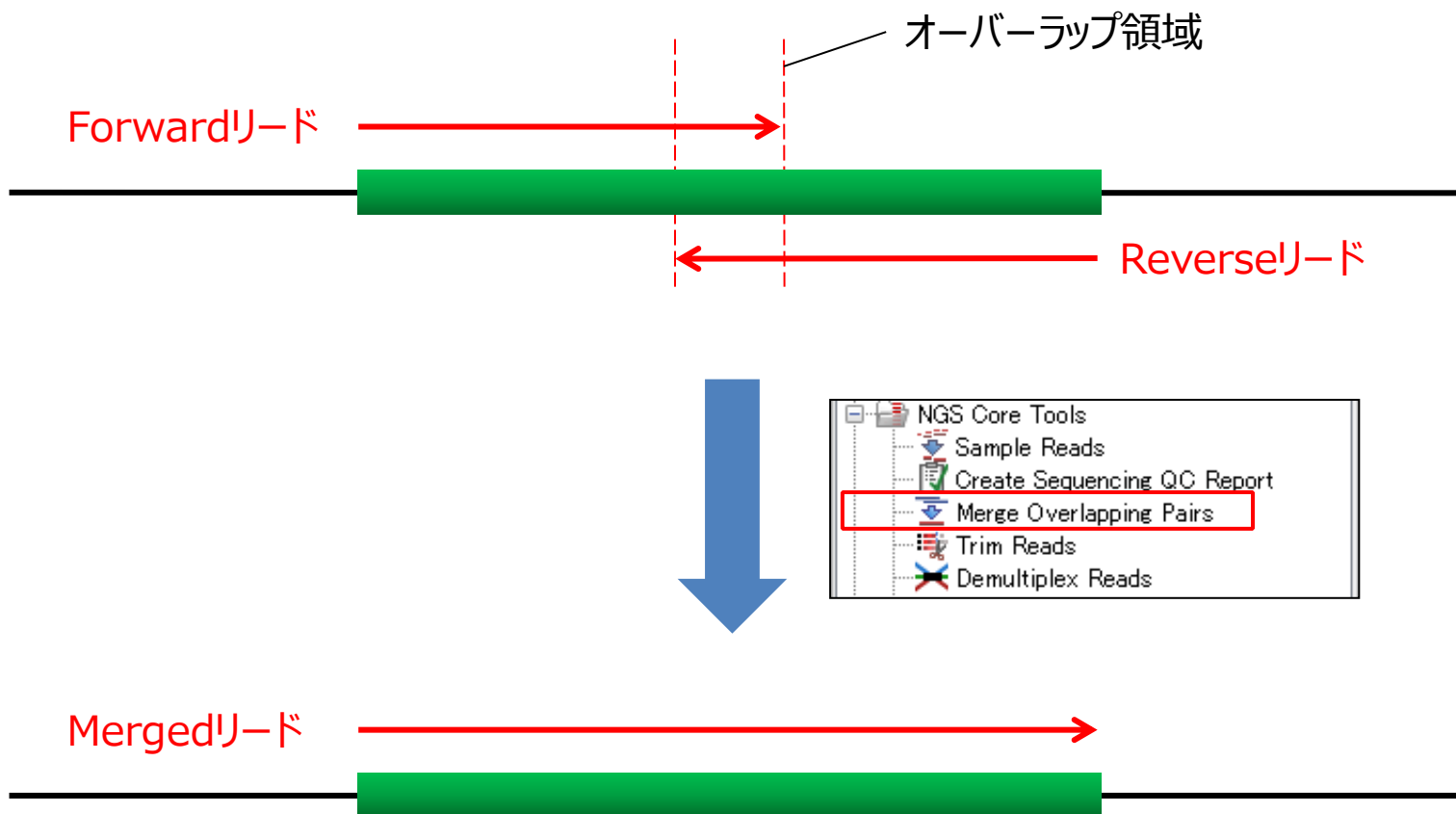
Trim Reads



- ✓ Trim Readsの使用により、各リードの低クオリティ部分がカットされる
- ✓ その他、アダプター配列の除去なども可能

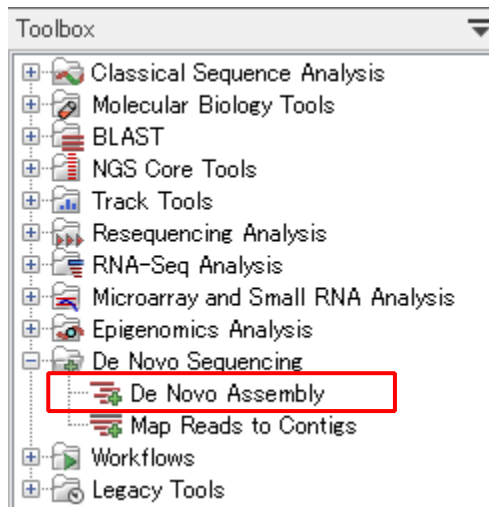


- ✓ シークエンスデータのリード配列数を減らす設定が可能
- ✓ シークエンスデータのデータ量が、サンプルのゲノムサイズより極端に大きい場合は、De Novo Assembly実行前に、本ツールでデータ量を最適な値（ゲノムサイズの30x）に揃える必要がある



- ✓ ペア同士のForward / Reverseリード配列間のオーバーラップ領域をマージし、一つのリード配列に統合する
- ✓ ペアエンドシーケンスを行っており、かつペアリード配列同士がオーバーラップする場合のみ、この処理が必要となる
- ✓ 先にTrim Readsを実行している場合、各リード配列の末端部分がトリミングされ、マージが上手く実行できない場合もあるので注意する

- ✓ リード配列データのクオリティチェック・前処理が終了したら、De Novo Assemblyによってコンティグの作成を行う
- ✓ コンティグのカバレッジや、リード配列のバリエーションを確認したい場合は、作成コンティグに対してリード配列データのマッピングを実行する

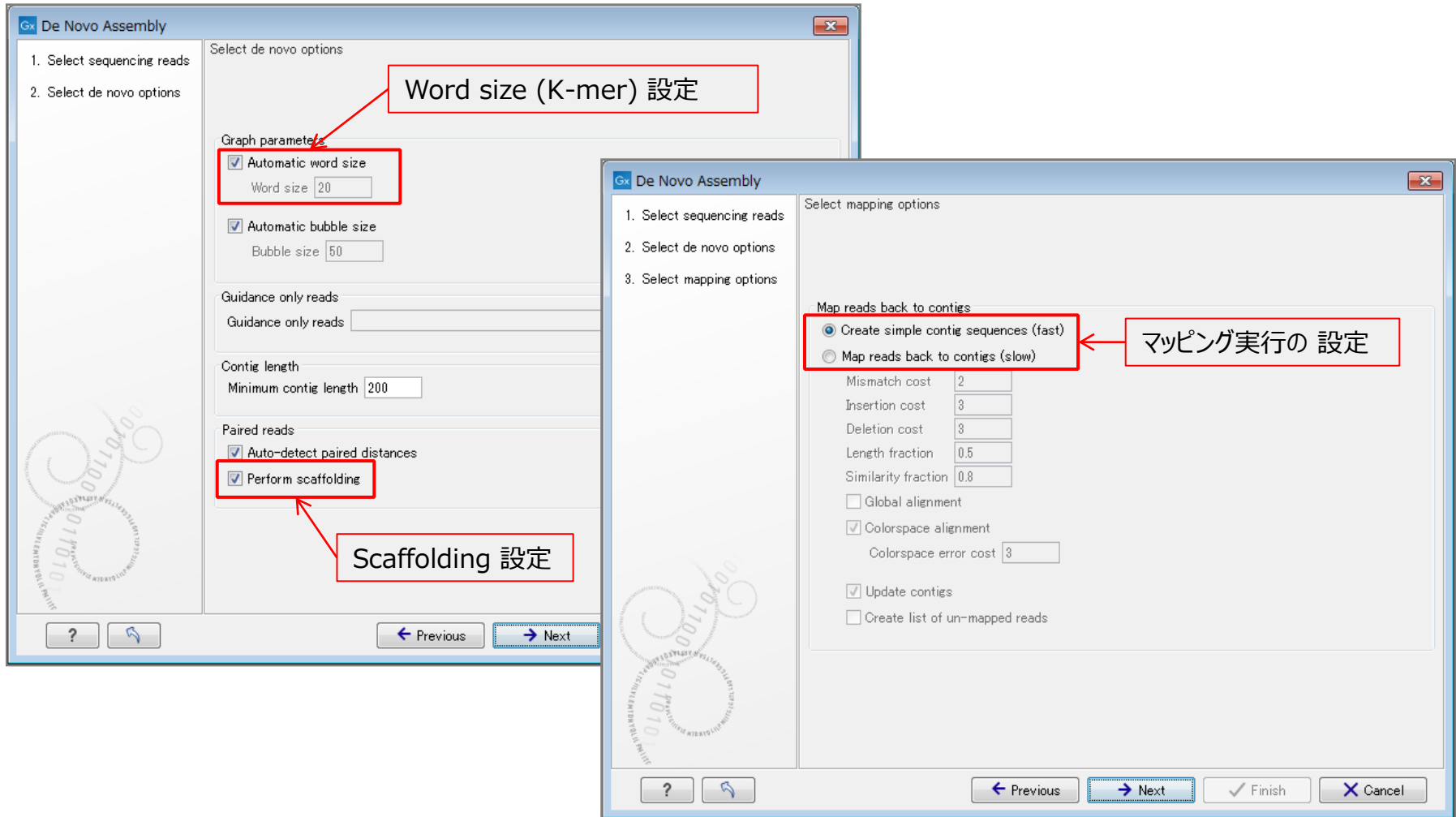


De Novo Assembly

- ショートリードデータを用いて、ゲノム配列のアセンブルを実行する
(ロングリード、およびトランスクリプトーム配列のアセンブルは非対応)

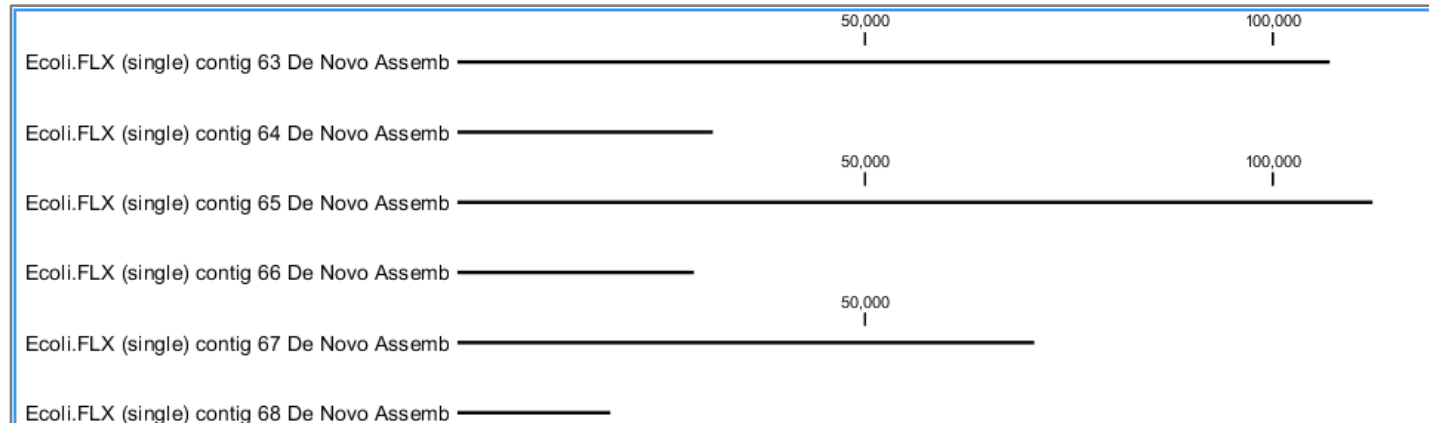
Map Reads to Contigs

- コンティグ配列を参照ゲノム配列として使用し、リード配列のマッピングを行う

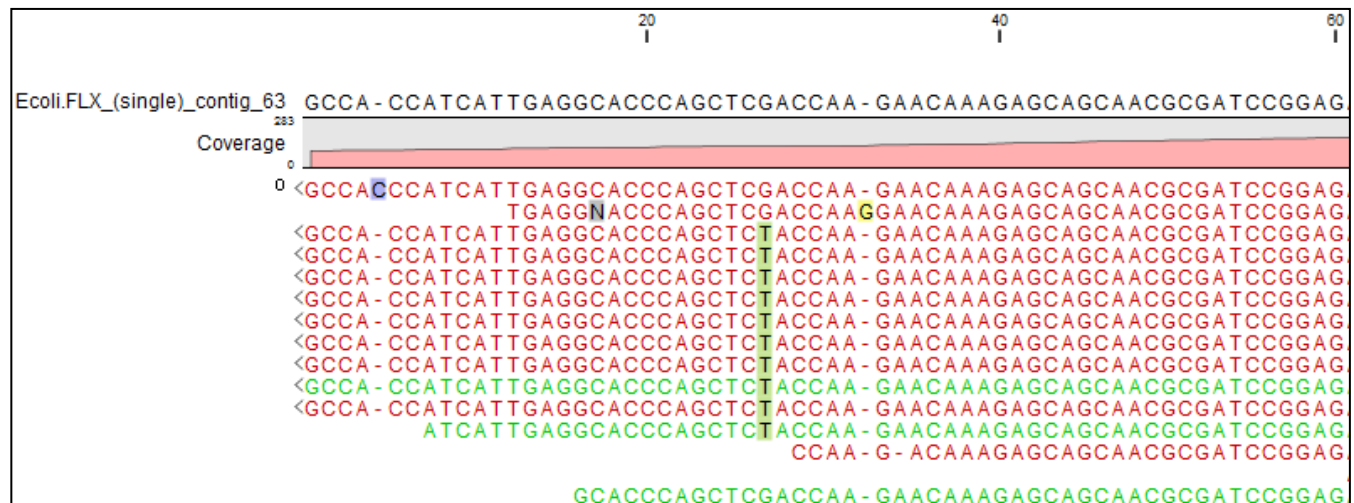


- ✓ パラメータ設定画面では、アセンブル実行時のWord size (K-mer)の値や、ペアリード使用時のScaffolding、さらにアセンブルと同時にマッピングも行うかどうかなどの設定を行うことができる

Simple Contigs

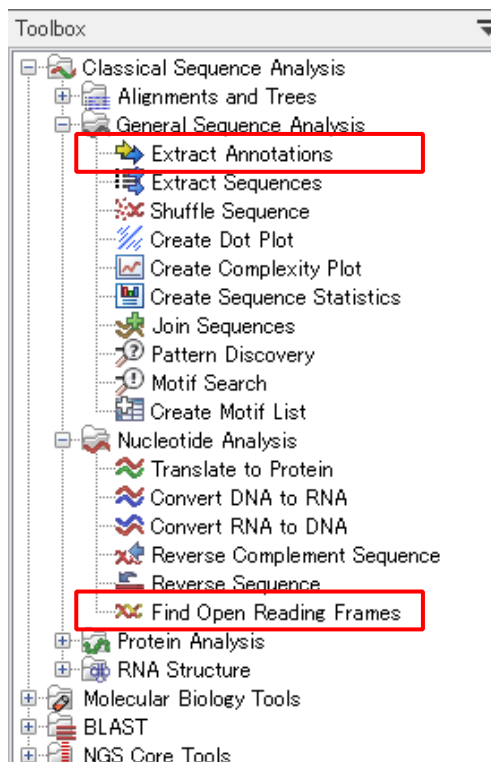


Mapping data



- ✓ Simple Contigsを選択した場合、各コンティグが一覧になって表示され、そのまま配列解析用ツールに使用することが可能
- ✓ アセンブル結果の統計値（コンティグ配列の最長塩基数やN50など）をまとめたレポートも出力される

- ✓ ゲノムシーケンズデータより作成したコンティグ配列には、非遺伝子領域の配列も含まれており、そのままではデータベース検索に使用できないため、コンティグ上の遺伝子領域を検出し、その領域のみの配列データを作成する必要がある
- ✓ CLC Genomics Workbenchでは、本機能については、基本的に原核生物ゲノム配列の解析しか実行できないため、真核生物ゲノムの解析には、Biobam社Blast2GO PROソフトウェアを使用する



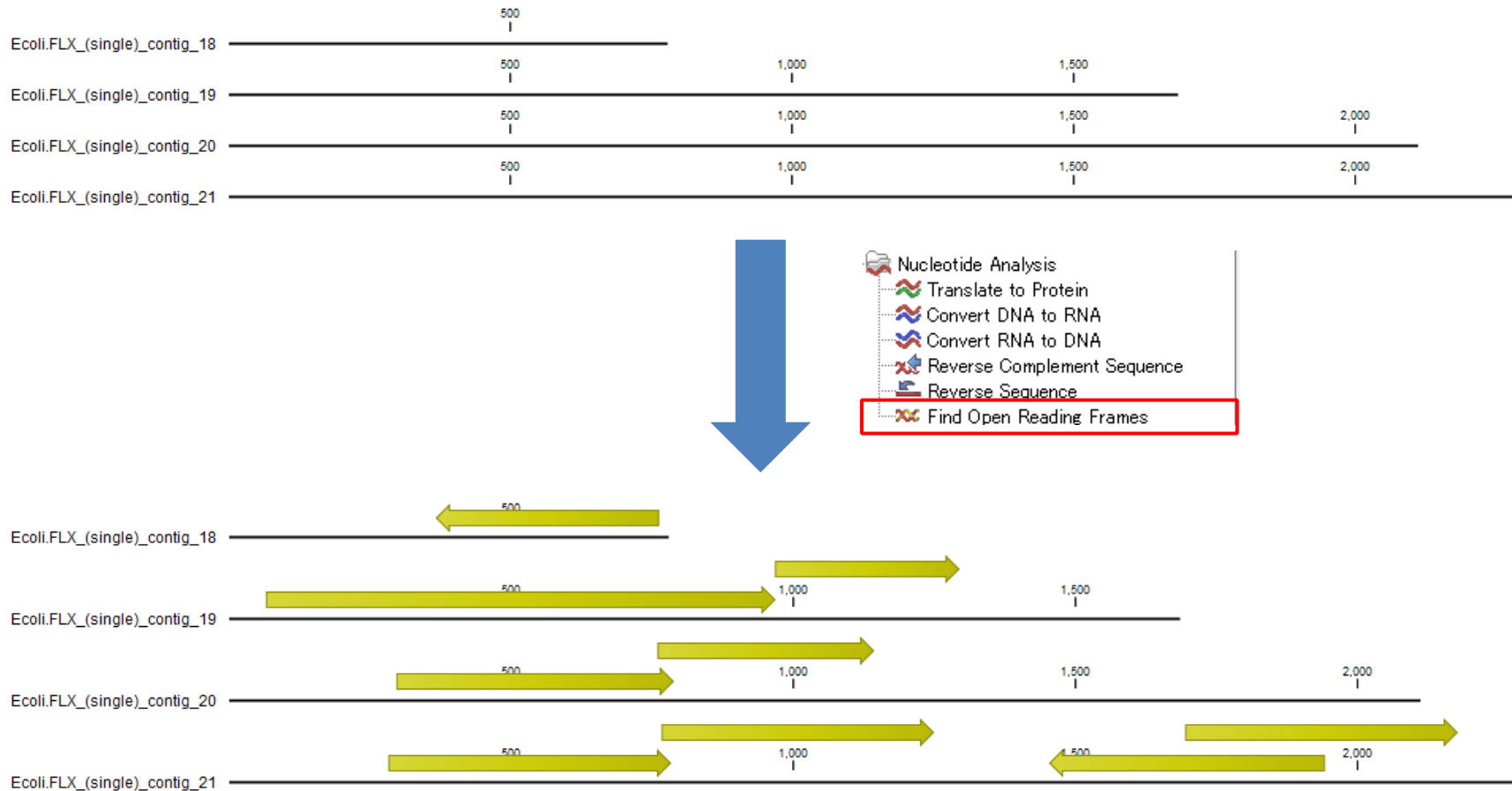
Find Open Reading Frames

- 配列データ上のORF (Open Reading Frame)領域を検出して、アノテーション付加を行う
(真核生物ゲノムのような、エクソン - イントロン構造をもつ配列データには非対応)

Extract Annotations

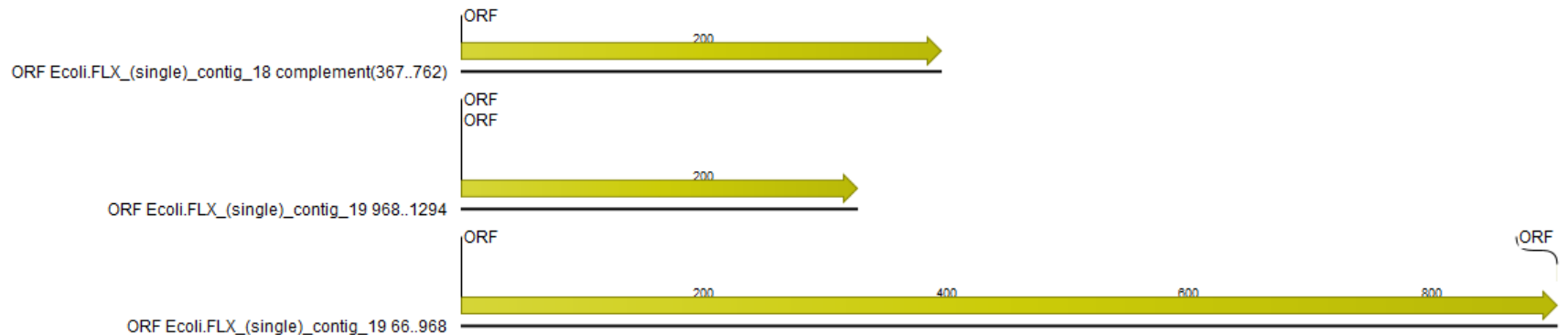
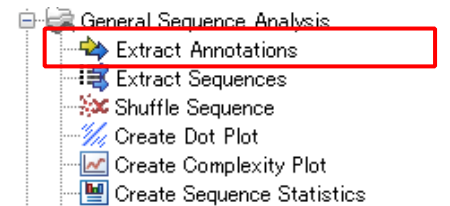
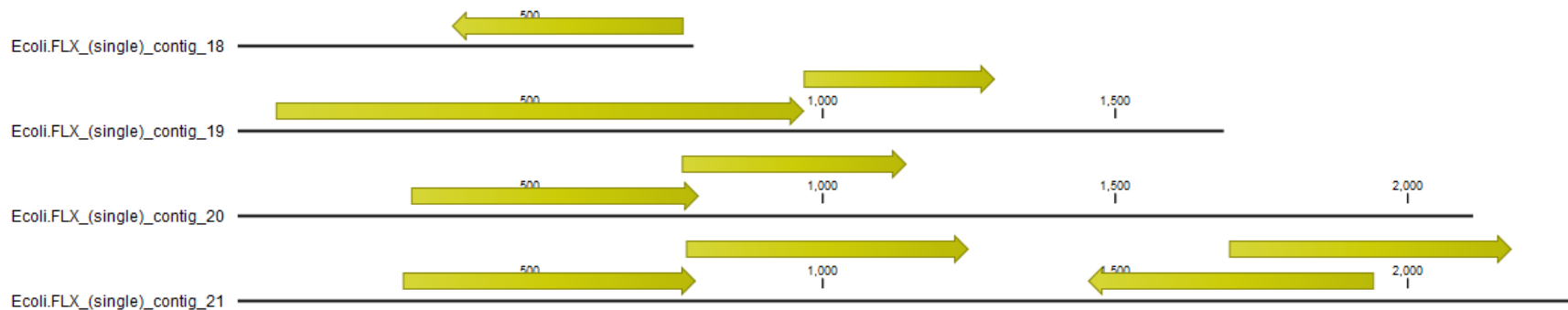
- 配列データ上の、任意のアノテーション領域の配列の抽出を行う

Find Open Reading Frames



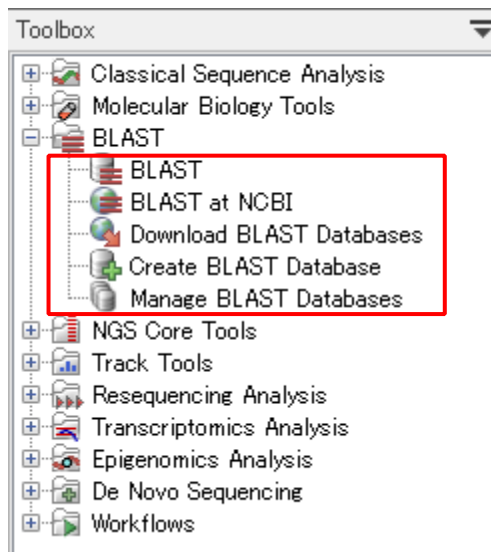
- ✓ Find Open Reading Framesを実行すると、もとの配列データ上に、ORF領域を示すアノテーションが追加される
- ✓ 用いるスタートコドンやストランドなどの各種パラメータ設定を調整することが可能

Extract Annotations



- ✓ Extract Annotationsでは、配列データ上の任意のアノテーション領域の配列を抽出することができる

- ✓ 抽出された各遺伝子配列に対し、BLASTで相同性検索を実行することで、各遺伝子の機能情報などを取得する
- ✓ CLC Genomics WorkbenchではBLASTツールの他、検索に用いるデータベースのダウンロードやカスタム作成用のツールが搭載されている



BLAST

- ローカルコンピュータを使用してBLAST検索を行う

BLAST at NCBI

- NCBIサーバーでBLAST検索を行う

Download BLAST Database

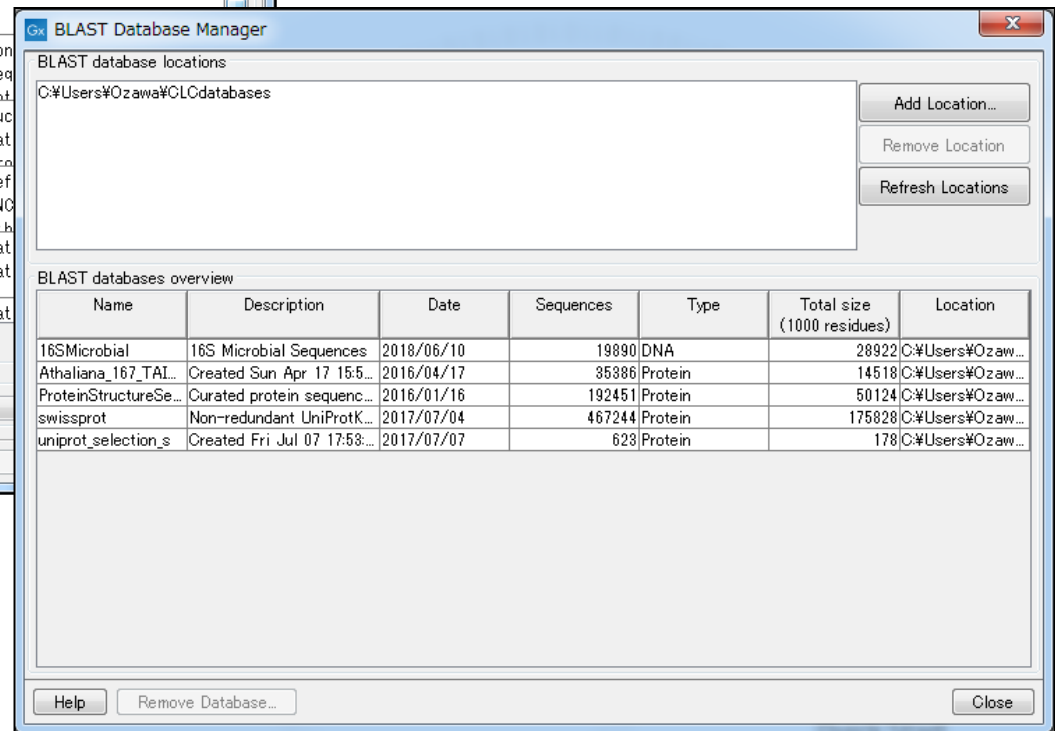
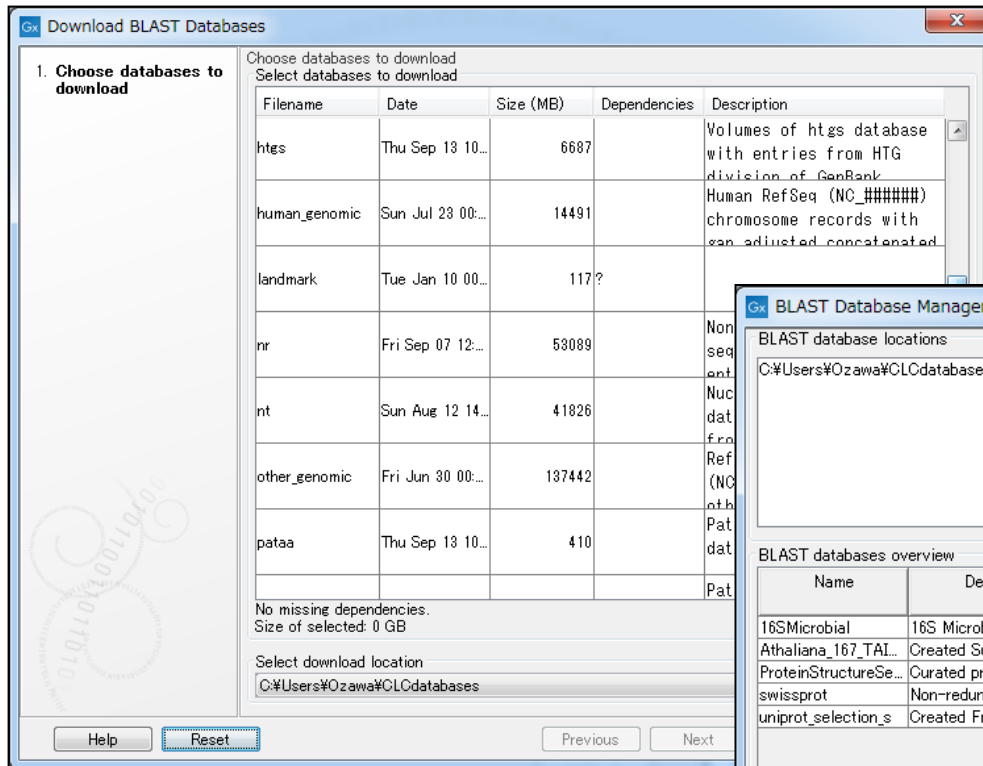
- NCBIなどで公開されている、様々なカテゴリごとに分類されたターゲット配列をダウンロードする

Create BLAST Database

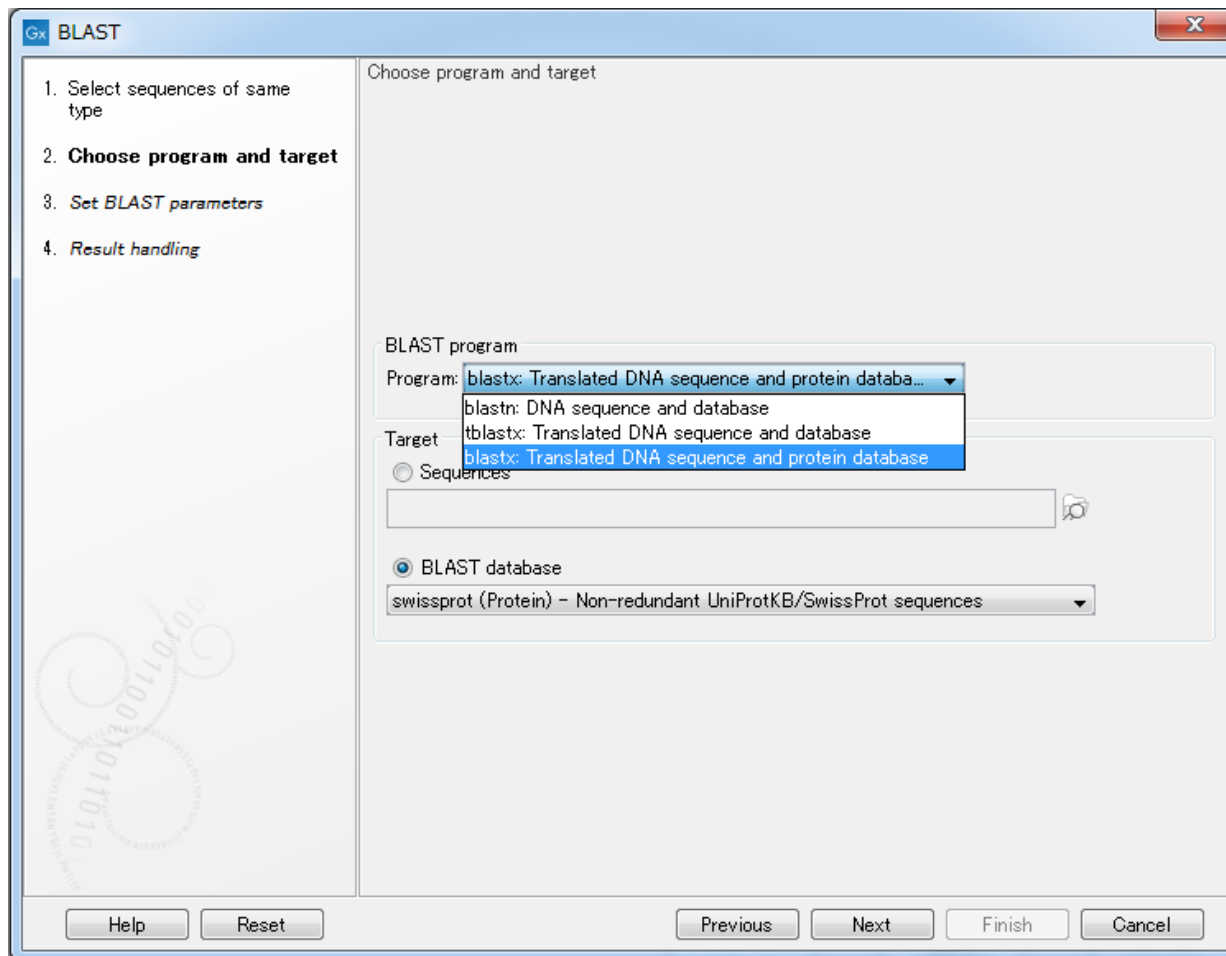
- 配列データをBLAST解析用のDatabaseデータとして登録する

Manage BLAST Database

- 登録したDatabaseデータの管理を行う



- ✓ ローカルBLASTを使用する場合、Download BLAST Databasesでnr/ntなどのデータベースをダウンロードしたり、カスタム作成の配列データをCreate BLAST Databasesでデータベース登録が可能



- ✓ ローカルBLASTのパラメータ設定画面では、BLASTプログラム（BLASTNやBLASTPなど）の選択と、ターゲットのBLAST databaseの種類などを指定する

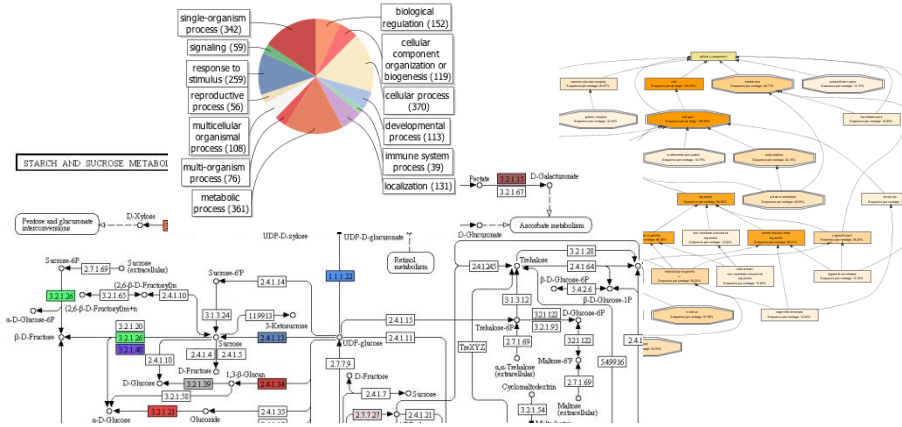
Query	Number of HSPs	Lowest E-value	Accession (E-value)	Description (E-value)
ORF Ecoli.FLX(single)_contig_1 3190.3906	10	2.72E-180	P69856	RecName: Full=Probable N-acetylneuraminic acid outer membrane channel protein NanC; AltName: Full=NanR-regul...
ORF Ecoli.FLX(single)_contig_1 9508.10254	228	1.28E-175	P99367	RecName: Full=Uncharacterized protein YjhP
ORF Ecoli.FLX(single)_contig_1 12448.13761	6	0.00	P99365	RecName: Full=Putative permease IIC component; AltName: Full=Putative PTS system EIIc component
ORF Ecoli.FLX(single)_contig_1 13774.14580	27	0.00	P99364	RecName: Full=Putative <i>sec</i> region protein SgcQ
ORF Ecoli.FLX(single)_contig_1 15154.15786	48	1.20E-145	P99362	RecName: Full=Protein SgcE
ORF Ecoli.FLX(single)_contig_1 16888.17676	32	0.00	P99360	RecName: Full=Uncharacterized HTH-type transcriptional regulator YjhI
ORF Ecoli.FLX(single)_contig_1 20671.22020	15	0.00	P99357	RecName: Full=Uncharacterized permease YjhF
ORF Ecoli.FLX(single)_contig_1 26014.28338	85	0.00	P13036	RecName: Full=Fe(3+) dicitrate transport protein FecA; AltName: Full=Iron(III) dicitrate transport protein FecA; Flag...
ORF Ecoli.FLX(single)_contig_1 30277.31233	69	1.25E-162	P15029	RecName: Full=Fe(3+) dicitrate transport system permease protein FecD; AltName: Full=Iron(III) dicitrate transport ...
ORF Ecoli.FLX(single)_contig_1 31234.32001	251	1.09E-178	P15031	RecName: Full=Fe(3+) dicitrate transport ATP-binding protein FecE; AltName: Full=Iron(III) dicitrate transport ATP-...
ORF Ecoli.FLX(single)_contig_1 3818.5032	285	0.00	P99371	RecName: Full=N-acetylneuraminic acid epimerase; AltName: Full=N-acetylneuraminic acid mutarotase; Short=Neu5Ac mut...
ORF Ecoli.FLX(single)_contig_1 14711.15142	24	1.08E-103	P99363	RecName: Full=Putative phosphotransferase IIA component SgcA; AltName: Full=Putative PTS system EIIa compon...
ORF Ecoli.FLX(single)_contig_1 15803.16586	40	0.00	P99361	RecName: Full=Putative <i>sec</i> region transcriptional regulator
ORF Ecoli.FLX(single)_contig_1 17627.18586	250	0.00	P99359	RecName: Full=Uncharacterized lyase YjhH
ORF Ecoli.FLX(single)_contig_1 22475.23353	20	0.00	P99356	RecName: Full=Uncharacterized transcriptional regulator YjhU
ORF Ecoli.FLX(single)_contig_1 24974.25927	4	0.00	P23485	RecName: Full=Protein FecR
ORF Ecoli.FLX(single)_contig_1 26978.27301	16	2.63E-71	P13036	RecName: Full=Fe(3+) dicitrate transport protein FecA; AltName: Full=Iron(III) dicitrate transport protein FecA; Flag...
ORF Ecoli.FLX(single)_contig_1 29282.30280	76	2.13E-179	P15030	RecName: Full=Fe(3+) dicitrate transport system permease protein FecC; AltName: Full=Iron(III) dicitrate transport ...
ORF Ecoli.FLX(single)_contig_1 5097.6077	9	0.00	P99370	RecName: Full=Probable 9-O-acetyl-N-acetylneuraminic acid deacetylase; Short=Neu5,9Ac2 deacetylase; AltName: ...
ORF Ecoli.FLX(single)_contig_1 6642.7007	10	1.89E-79	P99369	PUTATIVE PSEUDOGENE: RecName: Full=Putative uncharacterized protein YjhR
ORF Ecoli.FLX(single)_contig_1 8907.9452	45	6.38E-134	P99368	RecName: Full=Uncharacterized N-acetyltransferase YjhQ
ORF Ecoli.FLX(single)_contig_1 11040.12161	20	0.00	P99366	RecName: Full=Putative aminopeptidase SgcX
ORF Ecoli.FLX(single)_contig_1 18597.20564	250	0.00	P99358	RecName: Full=Uncharacterized protein YjhG
ORF Ecoli.FLX(single)_contig_1 24456.24977	38	2.83E-126	P23484	RecName: Full=Probable RNA polymerase sigma factor FecI
ORF Ecoli.FLX(single)_contig_1 28383.29285	30	0.00	P15028	RecName: Full=Fe(3+) dicitrate-binding periplasmic protein; AltName: Full=Iron(III) dicitrate-binding periplasmic prot...
ORF Ecoli.FLX(single)_contig_1 complement(30738.31055)	62	1.80E-53	P15029	RecName: Full=Fe(3+) dicitrate transport system permease protein FecD; AltName: Full=Iron(III) dicitrate transport ...
ORF Ecoli.FLX(single)_contig_1 complement(28287.28862)	20	1.67E-111	P15028	RecName: Full=Fe(3+) dicitrate-binding periplasmic protein; AltName: Full=Iron(III) dicitrate-binding periplasmic prot...
ORF Ecoli.FLX(single)_contig_1 complement(26241.26612)	34	4.12E-69	P13036	RecName: Full=Fe(3+) dicitrate transport protein FecA; AltName: Full=Iron(III) dicitrate transport protein FecA; Flag...
ORF Ecoli.FLX(single)_contig_1 complement(20940.21413)	16	9.88E-98	P99357	RecName: Full=Uncharacterized permease YjhF
ORF Ecoli.FLX(single)_contig_1 complement(20460.20933)	12	7.76E-14	P99358	RecName: Full=Uncharacterized protein YjhG
ORF Ecoli.FLX(single)_contig_1 complement(13563.13946)	9	4.72E-34	P99364	RecName: Full=Putative <i>sec</i> region protein SgcQ
ORF Ecoli.FLX(single)_contig_1 complement(13086.13520)	9	7.46E-95	P99365	RecName: Full=Putative permease IIC component; AltName: Full=Putative PTS system EIIc component
ORF Ecoli.FLX(single)_contig_1 complement(6660.7676)	60	0.00	P99369	PUTATIVE PSEUDOGENE: RecName: Full=Putative uncharacterized protein YjhR
ORF Ecoli.FLX(single)_contig_1 complement(33137.33733)	36	2.02E-148	Q47718	PUTATIVE PSEUDOGENE: RecName: Full=Putative transposase <i>InsO</i> for insertion sequence element IS911B
ORF Ecoli.FLX(single)_contig_1 complement(32558.32971)	18	1.29E-101	Q47719	PUTATIVE PSEUDOGENE: RecName: Full=Putative uncharacterized protein b4286
ORF Ecoli.FLX(single)_contig_1 complement(29645.29968)	31	3.17E-51	P15030	RecName: Full=Fe(3+) dicitrate transport system permease protein FecC; AltName: Full=Iron(III) dicitrate transport ...
ORF Ecoli.FLX(single)_contig_1 complement(24515.24904)	21	9.17E-92	P23484	RecName: Full=Probable RNA polymerase sigma factor FecI
ORF Ecoli.FLX(single)_contig_1 complement(22304.22642)	22	9.00E-61	P99356	RecName: Full=Uncharacterized transcriptional regulator YjhU
ORF Ecoli.FLX(single)_contig_1 complement(17525.17959)	250	1.86E-61	P99359	RecName: Full=Uncharacterized lyase YjhH

- ✓ BLASTに使用したクエリー配列ごとに、ヒットしたターゲット配列数と、その中で一番相同性が高い配列の詳細情報などがテーブル表示される
- ✓ テーブルの各行をダブルクリックすると、二番目以降にヒットしたターゲット配列も確認が可能

Blast2GO[®] PRO

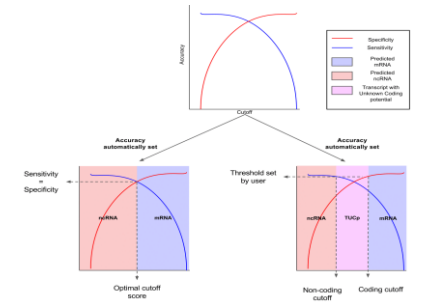
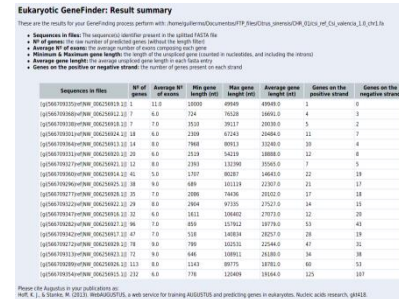
Gene Ontology Annotation

- BLASTとInterProScanによる、配列データへのアノテーション付け
- 取得したGene Ontology情報を使用した、遺伝子機能エンリッチメント解析やグラフ作成



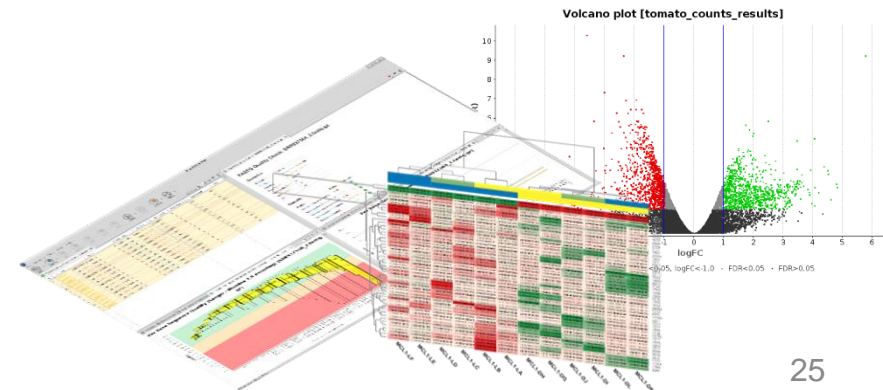
Genome Characterization

- 真核生物ゲノム用「Augustus」と原核生物用「Glimmer」プログラムを用いた、配列データ内の遺伝子領域予測

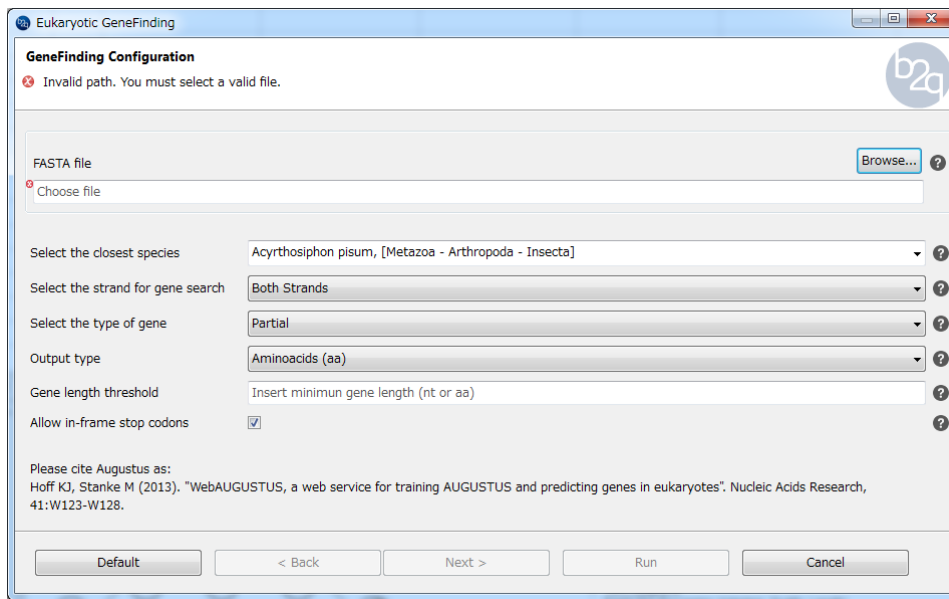
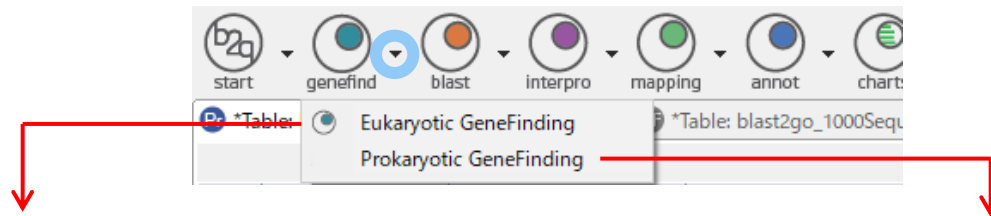


RNA-seq

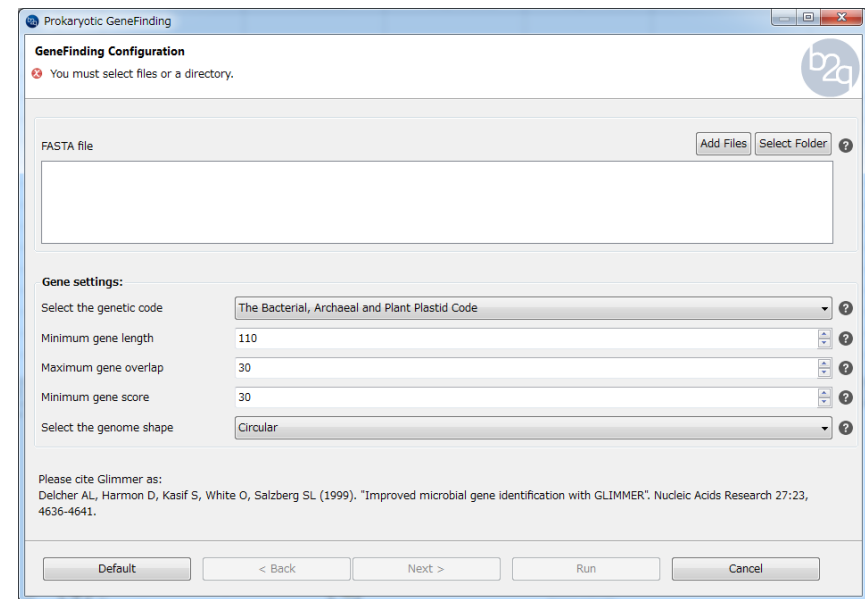
- FASTQファイルのクオリティチェック (FastQC) とトリミング
- RNA-Seq De Novo Assembly (Trinity)
- 新規転写物の発現量測定 (Fasta+Fastq)とサンプル間比較解析
- Volcano PlotやMA plot、Heatmapなどのグラフ作成



- ✓ De Novo Assemblyで作成したコンティグ配列などのDNA配列データを用いて、配列内の遺伝子やタンパク質コード領域を予測
- ✓ 原核生物と真核生物解析用の2種類のプログラムを搭載



真核生物用プログラム (Augustus)



原核生物用プログラム (Glimmer)

Select the closest species: **Caenorhabditis elegans [Metazoa - Nematoda - Chromadorea]**

Select the strand for gene search

Select the type of gene

Output type

Gene length threshold

Allow in-frame stop codons

Please cite Augustus as:
Hoff KJ, Stanke M (2013). "WebAUGUSTUS, a web service for training AUGUSTUS and predicting genes in eukaryotes." *Nucleic Acid Research*, 41:W123-W128.

Select the GeneFinding method

'Ab initio'

By using Hidden Markov Models we are able to build probabilistic models to predict the positions and composition of genes from DNA data.

With RNA-seq data

By adding introns as 'hints' from RNA-seq based alignments to the mathematical models, we can increase the accuracy of the gene prediction.

RNA-seq alignment file (.bam)

Qmap threshold: 30

Minimum read alignment: 11

Minimum intron length: 32

Minimum exon length: 300

Depth coverage: 20

- ✓ 真核生物ゲノム用プログラムでは、実行時に近縁生物種や、RNA-Seq マッピングデータを参照することが可能

Eukaryotic Gene Finding Results

Name: Cel_contig_list

Dataset Overview

- Number of sequences: 14,177
- Gene search based on *Caenorhabditis elegans* as closest related species
- GeneFinding method: 'Ab initio'

Results

Input Sequences		Found Genes				
Name	Length	Exons	Genes*	Genes per Strand +/-*	Length Min/Max**	Avg Length**
SRR7872897_trimmed_(paired)_contig_528	10,036	0	2	0 / 2	595 / 4,880	2,737.500
SRR7872897_trimmed_(paired)_contig_529	34,782	0	7	3 / 4	1,354 / 10,424	3,734.857
SRR7872897_trimmed_(paired)_contig_526	65,803	0	13	8 / 5	499 / 10,287	3,675.923
SRR7872897_trimmed_(paired)_contig_527	26,818	0	4	3 / 1	1,346 / 3,622	2,255.500
SRR7872897_trimmed_(paired)_contig_525	3,955	0	1	0 / 1	3,795 / 3,795	3,795.000

GO Version: Sep 17 2018

- ✓ 解析が終了すると、遺伝子領域の詳細をまとめたGFFファイルとCDS配列リスト、さらに解析結果をまとめたレポートが出力される

The screenshot displays the Blast2GO 5 PRO interface. The top menu includes File, Analysis, Tools, View, and Help. Below the menu is a toolbar with icons for start, genefind, blast, interpro, mapping, annot, charts, graphs, select, rna-seq, and workflows. The main window shows a table of annotations for the contig 'cel_contig_list_gff'. The table has columns for SeqID, Source, Type, Start, End, Score, Strand, Phase, Attr.ID, Attr.seqName, and Attr.Parent. Below the table, there is a progress log on the left and a genome browser on the right. The genome browser shows a chromosome with a length of 3,461 and a specific contig highlighted in orange. The contig is labeled 'SRR7872897 trimmed (paired) contig_3075.g117' and is surrounded by various annotations and coordinates.

SeqID	Source	Type	Start	End	Score	Strand	Phase	Attr.ID	Attr.seqName	Attr.Parent
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	gene	1	2885	0.2	-	.	SRR7872897_tri...	SRR7872897_tri...	
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	mRNA	1	2885	0.2	-	.	SRR7872897_tri...	SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	transcription_end_site	1	1	.	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	three_prime_utr	1	20	0.9	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	stop_codon	21	23	.	-	0		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	555	602	1	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	830	882	1	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	1156	1212	1	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	1310	1347	0.7	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	1416	1472	1	-	.		SRR7872897_tri...	SRR7872897_...
SRR7872897_trimmed_(paired)_contig_3075	AUGUSTUS	intron	1615	1659	1	-	.		SRR7872897_tri...	SRR7872897_...

✓ GFFファイルは専用のビューワーで、エクソン／イントロンなどの遺伝子構造を表示可能

The screenshot displays the Blast2GO 5 PRO software interface. The main window shows a table titled 'Table: cel_contig_list_seqs' with 11 rows of contig data. Below the table, the 'Sequence Viewer' is open for the first contig, showing its DNA sequence and amino acid translation.

Nr	Tags	SeqName	Description	Length	#Hits	e-Value	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme N...
1		SRR7872897_trimmed_(paired)_contig_1.g1		129							
2		SRR7872897_trimmed_(paired)_contig_1.g2		195							
3		SRR7872897_trimmed_(paired)_contig_1.g3		291							
4		SRR7872897_trimmed_(paired)_contig_1.g4		379							
5		SRR7872897_trimmed_(paired)_contig_1.g5		176							
6		SRR7872897_trimmed_(paired)_contig_1.g6		66							
7		SRR7872897_trimmed_(paired)_contig_1.g7		532							
8		SRR7872897_trimmed_(paired)_contig_1.g8		71							
9		SRR7872897_trimmed_(paired)_contig_3.g9		451							
10		SRR7872897_trimmed_(paired)_contig_3.g10		346							
11		SRR7872897_trimmed_(paired)_contig_4.g11		78							

The Sequence Viewer shows the following details for SRR7872897_trimmed_(paired)_contig_1.g1:

- Name: SRR7872897_trimmed_(paired)_contig_1.g1
- Description:
- Length: 129

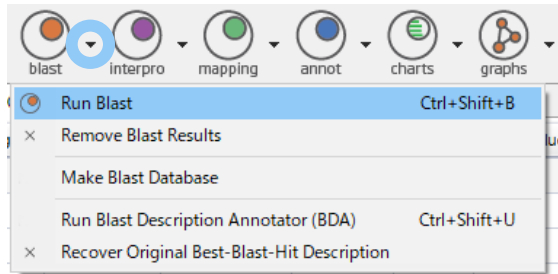
The DNA sequence is displayed in a color-coded format with positions 1 to 129 marked above. The amino acid translation is shown below the DNA sequence.

```

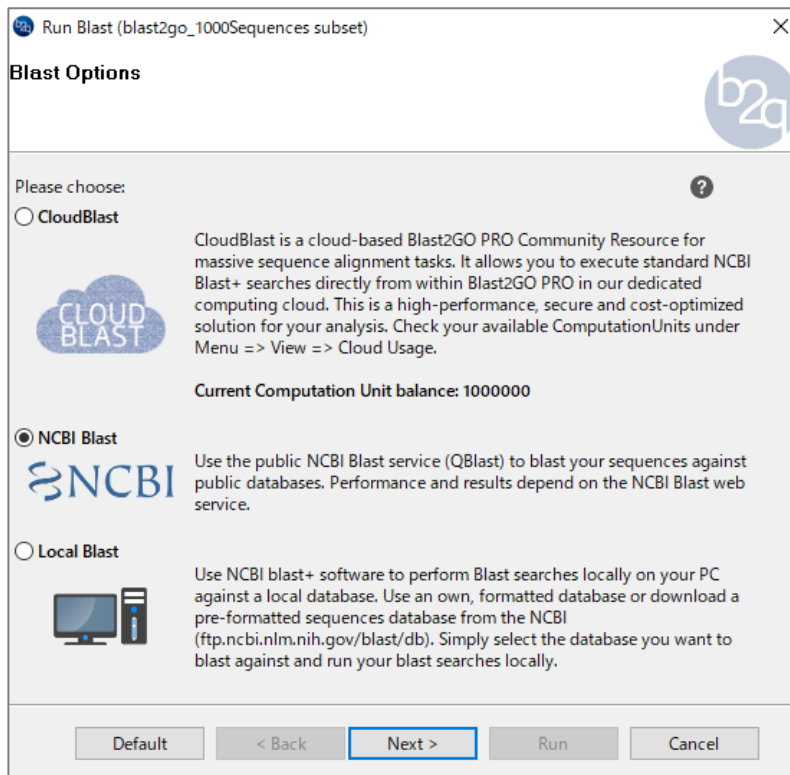
1      10     20     30     40     50     60
MNYNLLLSACIIFLIPEKSESICVTRRTDVGQFGSFFTDP L C D V W C R I R C G K G Q C R E D
61     70     80     90     100    110    120
P A T S N T A N C V C E K C Y R D D D G N V I F P D N D G F Q S R L N F D N S P T S S P W T M N Q R N E D D L Y P S
121
Q D R Y D T D R N
    
```

The bottom left corner of the interface shows the GO Version: Sep 17 2018.

- ✓ CDS配列リストは、配列名をまとめたテーブルと、DNA配列またはアミノ酸配列のビューワーが表示される



- ✓ CLC Genomics Workbenchと同様に、NCBI BLASTと Local BLASTを使用できる他、高速なクラウドベースの BLASTツール「CloudBlast」も使用可能
- ✓ Local BLAST用のデータベース作成機能も搭載



CloudBlast

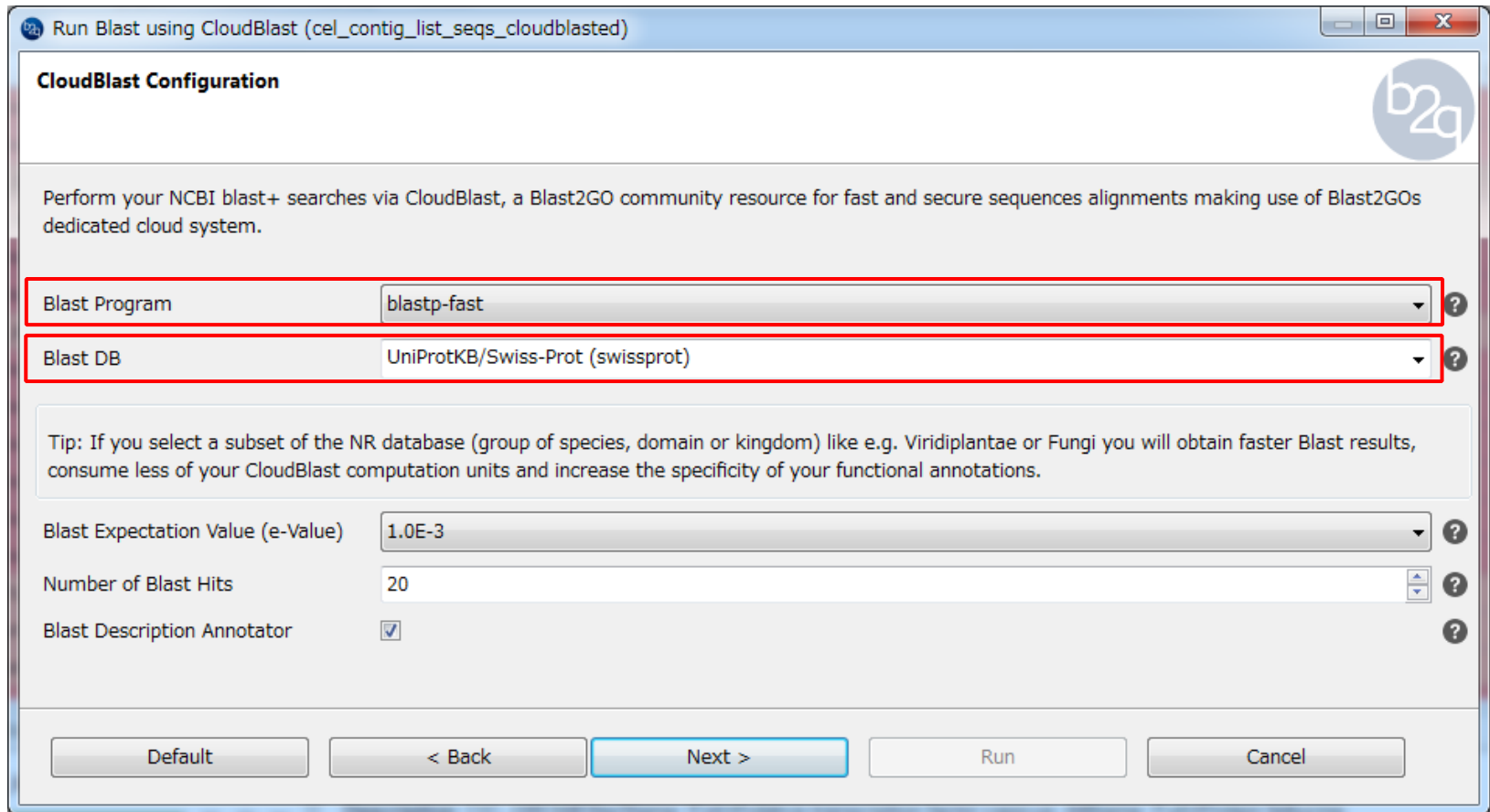
- Biobam社専用のクラウドサーバー上で、BLASTを安全にかつ高速に実行する
(Computation Unitsを消費)

NCBI Blast

- NCBIサーバーでBLAST検索を行う

Local Blast

- ローカルコンピュータを使用してBLAST検索を行う



Run Blast using CloudBlast (cel_contig_list_seqs_cloudblasted)

CloudBlast Configuration

Perform your NCBI blast+ searches via CloudBlast, a Blast2GO community resource for fast and secure sequences alignments making use of Blast2GOs dedicated cloud system.

Blast Program: blastp-fast

Blast DB: UniProtKB/Swiss-Prot (swissprot)

Tip: If you select a subset of the NR database (group of species, domain or kingdom) like e.g. Viridiplantae or Fungi you will obtain faster Blast results, consume less of your CloudBlast computation units and increase the specificity of your functional annotations.

Blast Expectation Value (e-Value): 1.0E-3

Number of Blast Hits: 20

Blast Description Annotator:

Default < Back Next > Run Cancel

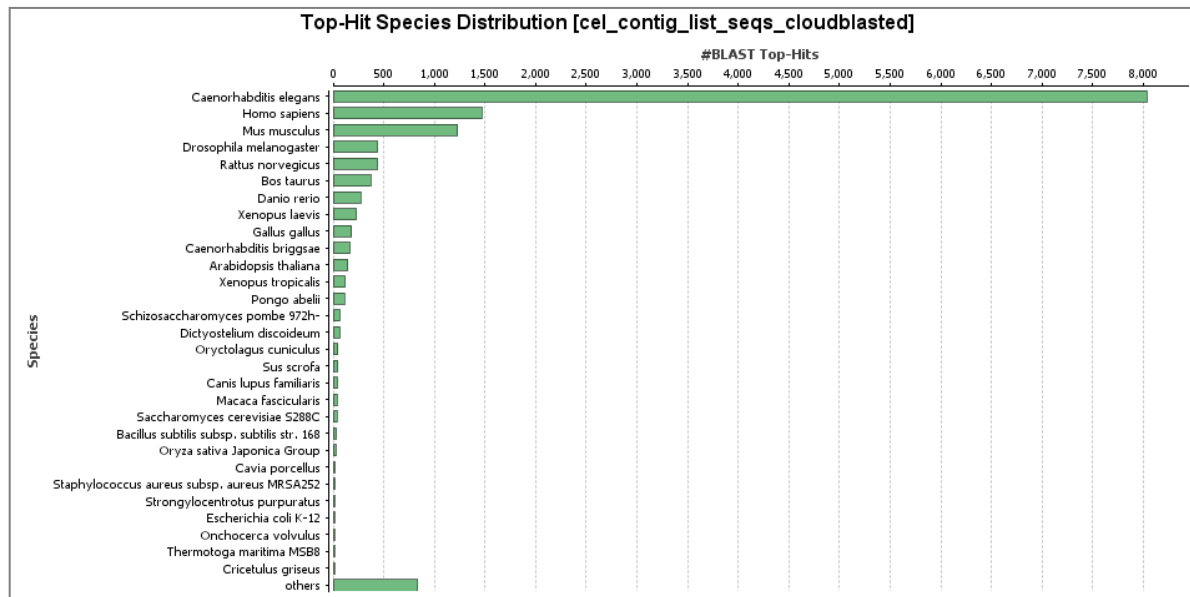
- ✓ どのBLASTを使用する場合でも、BLASTプログラムとBLASTデータベースを設定する
- ✓ CloudBlastとNCBI Blastでは、レディメイドのBLASTデータベースを使用するが、Local Blastではユーザー定義のデータベースを作成する必要がある

The screenshot displays the Blast2GO 5 PRO interface. The top window shows a table of sequence results with columns for Nr, Tags, SeqName, Description, Length, #Hits, e-Value, #GO, CO IDs, CO Names, Enzyme Codes, and Enzyme Names. The bottom window shows a detailed Blast Result for the query SRR7872897_trimmed_(paired)_contig_6.g23, including a color-coded alignment visualization and a table of sequences producing significant alignments.

Nr	Tags	SeqName	Description	Length	#Hits	e-Value	#GO	CO IDs	CO Names	Enzyme Codes	Enzyme Names
17	NO-BLAST	SRR7872897_trimmed_(paire...	---NA---	365							
18	BLASTED	SRR7872897_trimmed_(paire...	SRD19_CAEELRecName: Full=S...	316	11	0E0					
19	NO-BLAST	SRR7872897_trimmed_(paire...	---NA---	599							
20	NO-BLAST	SRR7872897_trimmed_(paire...	---NA---	166							
21	BLASTED	SRR7872897_trimmed_(paire...	NH167_CAEELRecName: Full=...	195	20	4.84E-52					
22	BLASTED	SRR7872897_trimmed_(paire...	NH167_CAEELRecName: Full=...	72	1	1.54E-6					
23	BLASTED	SRR7872897_trimmed_(paire...	NH167_CAEELRecName: Full=...	347	20	1.04E-136					
24	BLASTED	SRR7872897_trimmed_(paire...	VA5_LYCSIRecName: Full=Ven...	200	10	1.19E-10					
25	BLASTED	SRR7872897_trimmed_(paire...	NH167_CAEELRecName: Full=...	345	20	9.88E-142					
26	BLASTED	SRR7872897_trimmed_(paire...	NH167_CAEELRecName: Full=...	270	11	1.32E-88					
27	NO-BLAST	SRR7872897_trimmed_(paire...	---NA---	208							
28	BLASTED	SRR7872897_trimmed_(paire...	VGLU1_XENLARecName: Full=...	470	20	1.28E-28					

#	Sequences Producing Significant Alignments	Scientific Taxonomy	E-Value	Hit length	Align length	Pos	Sim	Hsp/Hit	Hsp/Query	Hsps	Frame	Gene identifier	Acc
1	1. RecName: Full=Nuclear hormone receptor family member nhr-167 gij 74957698 sp O17683.1 NH167_CAEEL	Caenorhabditis elegans	1.04459e-136	343	344	256	74.4%	100.3%	99.1%	1	0	74957698	O17683
2	1. RecName: Full=Nuclear hormone receptor family member nhr-9 gij 10720138 sp Q23489.1 NHR9_CAEEL	Caenorhabditis elegans	9.03817e-51	332	314	168	53.5%	94.6%	90.5%	1	0	10720138	Q23489
3	1. RecName: Full=Nuclear hormone receptor family member nhr-51 gij 74957904 sp O17927.3 NHR51_CAEEL	Caenorhabditis elegans	3.61706e-41	371	336	172	51.2%	90.6%	96.8%	1	0	74957904	O17927

- ✓ BLASTを実行すると、CDS配列リストに各配列のトップヒット配列の情報が表示される
- ✓ Blast Resultビューで、ヒットした配列の詳細情報を確認が可能



Nr	Tags	SeqName	Description	Length	#Hits	e-Value	#GO	GO IDs	GO Names
14	BLASTED MAPPED	SRR7872897.trimmed_(paired)_contig_4.g14	CYTX_ONCVORecName: Full=Onc...	93	6	4.14E-9	14	F.GO:0004869; C.GO:0005576; C.GO:0005615; C.GO:0005623; P.GO:0006952; P.GO:0006953; P.GO:0007204; P.GO:0010466; P.GO:0010951; P.GO:0030195; F.GO:0020414; P.GO:0042311; P.GO:0042742; P.GO:0050880	F:cysteine-type endopeptidase inhibitor activity; C:extracellular region; C:extracellular space; C:cell; P:defense response; P:acute-phase response; P:positive regulation of cytosolic calcium ion concentration; P:negative regulation of peptidase activity; P:negative regulation of endopeptidase activity; P:negative regulation of blood coagulation; F:peptidase inhibitor activity; P:vasodilation; P:defense response to bacterium; P:regulation of blood vessel size
15	BLASTED MAPPED	SRR7872897.trimmed_(paired)_contig_4.g15	PGAP3_DROMERecName: Full=Pos...	320	10	3.21E-37	15	C.GO:0000139; C.GO:0000329; F.GO:0003674; C.GO:0005773; C.GO:0005774; C.GO:0005783; C.GO:0005789; C.GO:0005794; P.GO:0006505; P.GO:0006506; C.GO:0016020; C.GO:0016021; F.GO:0016788; P.GO:0030026; C.GO:0031227	C:Golgi membrane; C:fungal-type vacuole membrane; F:molecular function; C:vacuole; C:vacuolar membrane; C:endoplasmic reticulum; C:endoplasmic reticulum membrane; C:Golgi apparatus; P:GPI anchor metabolic process; P:GPI anchor biosynthetic process; C:membrane; C:integral component of membrane; F:hydrolase activity, acting on ester bonds; P:cellular manganese ion homeostasis; C:intrinsic component of endoplasmic reticulum membrane
16	BLASTED MAPPED	SRR7872897.trimmed_(paired)_contig_4.g16	JKAMP_MOUSERecName: Full=JNK...	326	2	9.85E-46	8	C.GO:0005575; C.GO:0005783; C.GO:0005789; P.GO:0006986; C.GO:0016020; C.GO:0016021; P.GO:0030433; F.GO:0031625	C:cellular component; C:endoplasmic reticulum membrane; P:response to unfolded protein; C:membrane; C:integral component of membrane; P:ubiquitin-dependent ERAD pathway; F:ubiquitin protein ligase binding

- ✓ BLAST実行後は、ヒットした配列の詳細情報をグラフ化や、Gene Ontologyなどの遺伝子機能解析が可能

お問い合わせ先: フィルジェン株式会社

TEL 052-624-4388 (9:00~17:00)

FAX 052-624-4389

E-mail: biosupport@filgen.jp