



～ゲノムアノテーション解析～

フィルジェン株式会社 バイオインフォマティクス部(biosupport@filgen.jp)

FASTAファイル 塩基、アミノ酸配列の情報を含むファイルフォーマット



De-Novo Assemblyで作成した
コンティグデータ



データベースで公開されているもの
情報が不足しているデータ

```
1      10      20      30
AAACTCCCACCCGGCATTCTTGTTCCTGAT
71     80     90     100
GATACACATATGATCTCAATGACAAGAAAGA
```

```
1      10      20      30
MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFG
71     80     90     100
IKRSDARTAPHGHVMVELVAELEGIQYGRSGE
```

塩基、アミノ酸配列のみの情報のみ・・・

その配列が持つ機能についてはこのファイルだけでは不十分

アノテーション解析



様々なデータベースを利用して類似した遺伝子名や
遺伝子機能情報を付与する。



塩基（アミノ酸）配列だけの情報からどのような類似する配列
がありどのような機能を持つものがあるか
または、どのような遺伝子機能が豊富に含まれるかどうかを調べ
ることが可能。

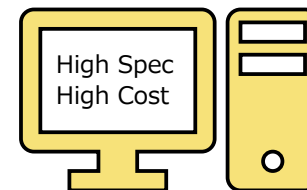
解析の課題...



複数のデータベースを使用
するので解析が煩雑化



計算量が多いため解析時間がかかる、
さらに
解析のため高度なスペックのPCが必要



OmicBoxはこれらの問題を解決



✓非モデル生物に対応

リファレンスゲノムがないデータでも解析が実行可能。
農学系のユーザーに適した解析ツールを多数搭載。

✓初めてでも簡単に使用することができます。

実績は高いがコマンドライン型であったりOSに制限がある
オープンソースソフトウェアを多数組み込みマウス操作で簡単に解析できる。



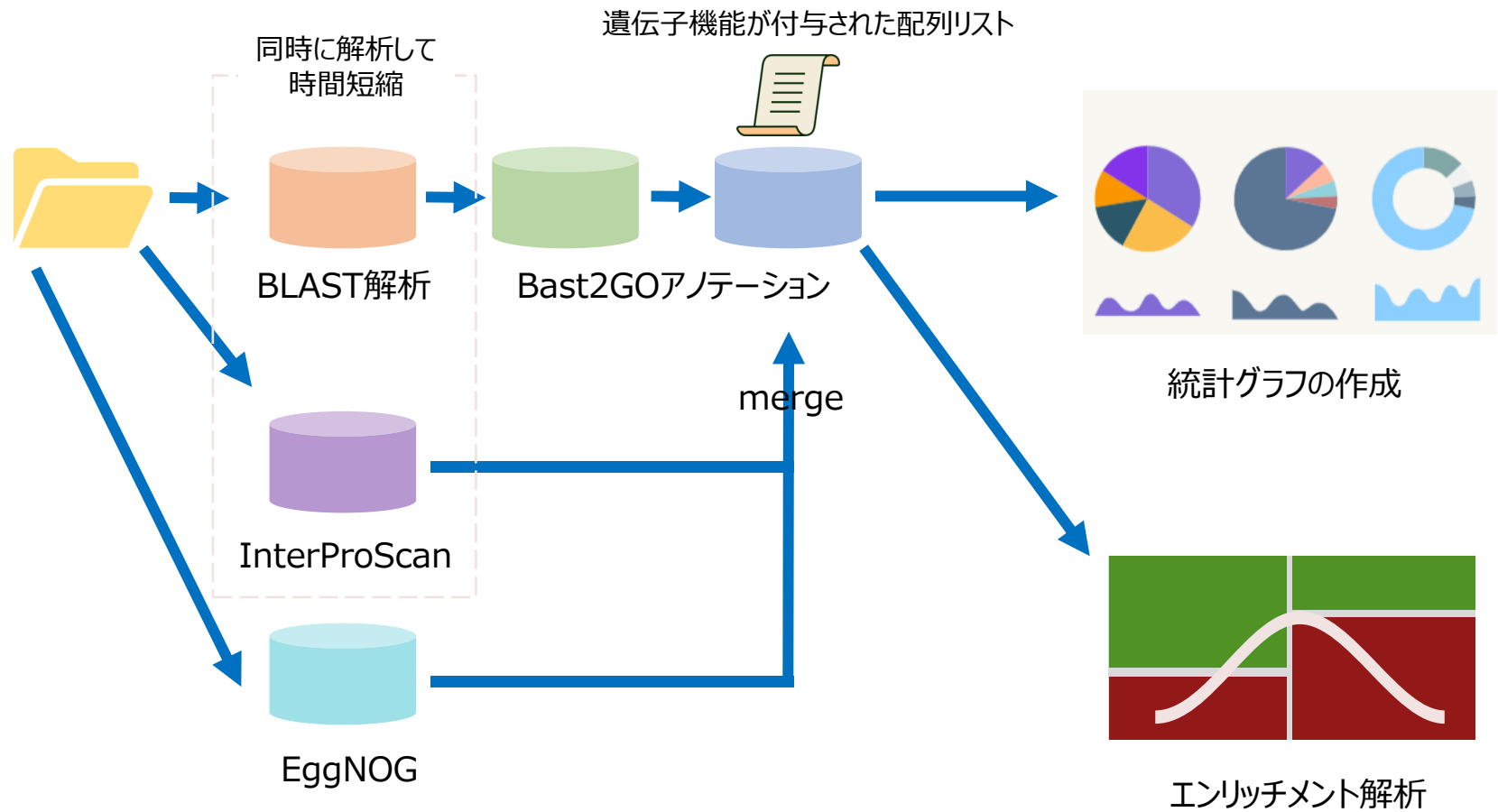
✓高価で高スペックなPCは不要

解析や計算は、統合させたウェブサイトや
BioBam社のクラウドを通して行われるため、
安定したインターネット接続があれば解析が可能。

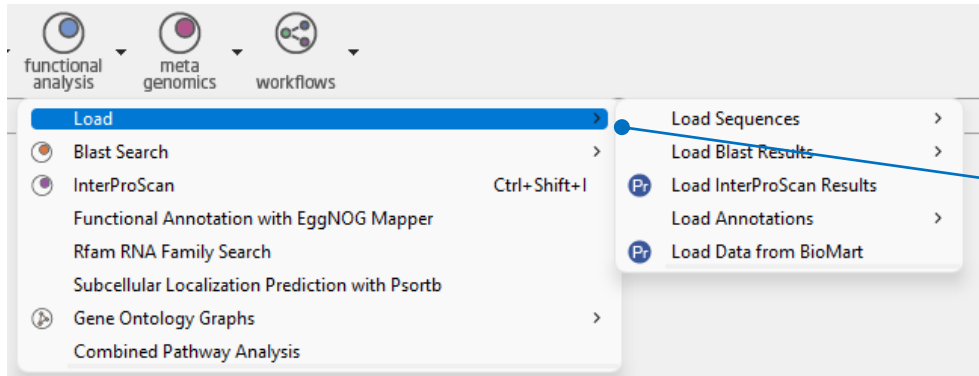
データインポート

遺伝子機能アノテーション付与

統計解析



OmicsBoxソフトウェア上ですべて解析可能



POINT 簡単にファイルをロード

Functional Analysis → **File** → **Load**
からファイルをロードできます。

<input checked="" type="checkbox"/>	Nr	Tags	SeqName	Description	Length	#Hits	e-Value	sim mean	#GO
<input checked="" type="checkbox"/>	1		YP_0097243...		7096				
<input checked="" type="checkbox"/>	2		YP_0097243...		1273				
<input checked="" type="checkbox"/>	3		YP_0097243...		275				
<input checked="" type="checkbox"/>	4		YP_0097243...		75				
<input checked="" type="checkbox"/>	5		YP_0097243...		222				
<input checked="" type="checkbox"/>	6		YP_0097243...		61				
<input checked="" type="checkbox"/>	7		YP_0097243...		121				
<input checked="" type="checkbox"/>	8		YP_0097243...		121				
<input checked="" type="checkbox"/>	9		YP_0097243...		419				
<input checked="" type="checkbox"/>	10		YP_0097252...		38				

Hide Side Panel

Blast

Blast Pr

Remove Blast Pr

Charts Ch

Blast Description Annotator Pr

Retrieve Blast Top-Hit Pr

Retrieve Blat Top-Hit Pr

画面中央にロードしたデータが表示
サイドパネルから各分析を実行できます。


↑ 次の解析項目の
BLAST解析に関するサイドパネル

POINT 類似する配列情報を付与できるBlast解析を実行できます。


Run Blast (ExampleSequences)

Blast Options


CloudBlast

 CloudBlast is a cloud-based OmicsBox Community Resource for massive sequence alignment tasks. It allows you to execute standard NCBI Blast+ searches directly from within OmicsBox in our dedicated computing cloud. This is a high-performance, secure and cost-optimized solution for your analysis. Check your available CloudUnits under Menu => View => Cloud Usage.


NCBI Blast

 Use the public NCBI Blast service (QBlast) to blast your sequences against public databases. Performance and results depend on the NCBI Blast web service.

Local Blast

 Use NCBI blast+ software to perform Blast searches locally on your PC against a local database. Use an own, formatted database or download a pre-formatted sequences database from the NCBI (<ftp.ncbi.nlm.nih.gov/blast/db>). Simply select the database you want to blast against and run your blast searches locally.

Custom Database CloudBlast

 Run BLAST against a database made of your custom selected fasta files. Custom Database CloudBlast is a cloud-based OmicsBox Community Resource for sequence alignment tasks. It allows you to execute standard NCBI Blast+ searches directly from within OmicsBox in our dedicated computing cloud. This is a high-performance, secure and cost-optimized solution for your analysis. Check your available CloudUnits under Menu => View => Cloud Usage.

Default < Back **Next >** Cancel Run

Cloud Blast

メーカー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ことができます。

NCBI Blast

NCBIサーバー上で、BLAST解析を実行します。

Local Blast

ローカル/独自のデータベースに対し解析を実行します。あらかじめ検索したいデータベースの情報をダウンロードし設定しておく必要があります。

Custom Database CloudBlast

メーカーのクラウドサーバー上でローカル/独自のデータベースに対し解析を実行します。

POINT 複雑な設定は不要です。

Cloud Blastの設定画面

● 使用したいアルゴリズムを選択

例)

blastp	クエリ配列：アミノ酸配列	DB：タンパク質配列
blastx	クエリ配列：塩基配列	DB：タンパク質配列

● 検索するデータベースを選択

nr・swissprotなど

● サブセットの選択

分類を選択することによってより高速のBlast結果が得られ、Cloud Computation Unit*の消費量が少なくなり、機能アノテーションの具体性が高まります。

*Computation Unitsは、CloudBlast解析を行うごとに消費されます。(InterProScanのCloud解析でも消費) すべてのUnitsを使い果たった場合は、追加で 6 million Computation Units を購入することができます。

POINT 視覚的に結果を理解できます。

BLASTED 正常なBLAST結果を有する配列はオレンジ色

NO-BLAST 与えられたシーケンスに対して検索結果が得られない場合、暗赤色

SeqName	Description	Length
Solyc00g005000.3.1	Eukaryotic aspartyl protease fami...	1377
Solyc00g005040.3.1	Potassium channel (AHRD V3.3 *-...	357
Solyc00g005050.3.1	UPF0664 stress-induced protein ...	588
Solyc00g005060.1.1	LOW QUALITY:Cyclin/Brf1-like TB...	273
Solyc00g005080.2.1	LOW QUALITY:Protein Ycf2 (AHR...	288

SeqName	Description	Length	#Hits	e-Value	sim mean
Solyc00g005000.3.1	protein ASPARTIC PROTEASE IN GUARD CELL 2-like	1377	20	0E0	82.76%
Solyc00g005040.3.1	potassium channel KAT3-like isoform X2	357	20	3.09E-81	97.91%
Solyc00g005050.3.1	UPF0664 stress-induced protein C29812.11c	588	20	1.19E-143	92.53%

✓類似する配列の機能が配列に付与されます。

BLAST解析

Query Name: C02006A04
Database: nr
Length: 571
Program: BLASTX 2.2.31+

E-value cut-off: 0.001
Filters: L;

Alignments

> 200
80-200 50-80
40-50 < 40

#	Sequences Producing Significant Alignments	Scientific Taxonomy	E-Value	Hit length	Align length	Pos	Sim	Hsp/Hit	Hsp/Query	Hsps	Frame	Gene identifier
1	1. hypothetical protein CICLE_v10022212mg [Citrus clementina] gi 567903570 ref XP_006444273.1 gi 557546535 gb E5R57513.1 2. PREDICTED: glutathione S-transferase F6 [Citrus sinensis] gi 56883489 ref XP_006479908.1 3. hypothetical protein CISIN_1g046920mg [Citrus sinensis] gi 641868631 gb KDO87315.1	Citrus clementina Citrus sinensis	4.92276e-123	213	173	173	100.0%	81.2%	90.9%	1	1	557546535 567903570 568852489 641868631
2	1. phi class glutathione S-transferase protein [Bruguiera gymnorhiza] gi 329130894 gb AEB77871.1	Bruguiera gymnorhiza	4.14973e-93	214	173	148	85.5%	80.8%	90.9%	1	1	329130894
3	1. phi class glutathione S-transferase protein [Bruguiera gymnorhiza] gi 329130898 gb AEB77873.1	Bruguiera gymnorhiza	2.29516e-92	214	173	148	85.5%	80.8%	90.9%	1	1	329130898
4	1. hypothetical protein CICLE_v10024356mg.partial [Citrus clementina] gi 567903568 ref XP_006444272.1 gi 557546534 gb E5R57512.1	Citrus clementina	1.01671e-88	212	173	144	83.2%	81.6%	90.9%	1	1	557546534 567903568
5	1. PREDICTED: glutathione S-transferase PARB [Ricinus communis] gi 225582148 ref XP_002531868.1 2. glutathione s-transferase theta, gst, putative [Ricinus communis] gi 223528476 gb EEF30505.1	Ricinus communis	6.98634e-88	213	173	143	82.7%	81.2%	90.9%	1	1	223528476 255582148
6	1. hypothetical protein CISIN_1g038052mg [Citrus sinensis]	Citrus sinensis	8.70789e-88	225	173	143	82.7%	76.9%	90.9%	1	1	641868633

Table mode

> 200
80-200 50-80
40-50 < 40



<input checked="" type="checkbox"/>	Nr	Tags	SeqName	Description	Length	#Hts	e-value	sim mean	#GO
<input checked="" type="checkbox"/>	1	BLASTED	C02006A02	organic cation/carnitine tr...	602	20	8.61E-138	92.46%	
<input checked="" type="checkbox"/>	2	BLASTED	C02006A04	glutathione S-transferase	571	20	9.87E-125	84.27%	
<input checked="" type="checkbox"/>	3	BLASTED	C02006A08	two-component response	620	20	6.16E-69	83.58%	
<input checked="" type="checkbox"/>	4	BLASTED	C02006B02	GPI transamidase compon...	533	20	3.56E-123	88.92%	
<input checked="" type="checkbox"/>	5	BLASTED	C02006B04	40S ribosomal protein S11	558	20	9.52E-102	98.93%	
<input checked="" type="checkbox"/>	6	BLASTED	C02006B10						
<input checked="" type="checkbox"/>	7	BLASTED	C02006C02						
<input checked="" type="checkbox"/>	8	BLASTED	C02006C04						
<input checked="" type="checkbox"/>	9	BLASTED	C02006C06						
<input checked="" type="checkbox"/>	10	BLASTED	C02006C08						
<input checked="" type="checkbox"/>	11	BLASTED	C02006C10						
<input checked="" type="checkbox"/>	12	BLASTED	C02006C12						
<input checked="" type="checkbox"/>	13	BLASTED	C02006D02						
<input checked="" type="checkbox"/>	14	BLASTED	C02006D04						
<input checked="" type="checkbox"/>	15	BLASTED	C02006D06						
<input checked="" type="checkbox"/>	16	BLASTED	C02006D08						

Show BLAST Results

Show InterProScan Result

Show Mapping Result

Show GO Descriptions

Copy Sequence Names to Clipboard

Change Annotation and Description

Annotate Sequence

Load Pathway-Maps from KEGG (Online)

Make Graph of GO-Mapping-Results with Annotations Score

Extract Selection to New Tab

Copy Selection to Clipboard (tabular format)

Copy Content of Column: SeqName to Clipboard

Create ID List of Column: SeqName

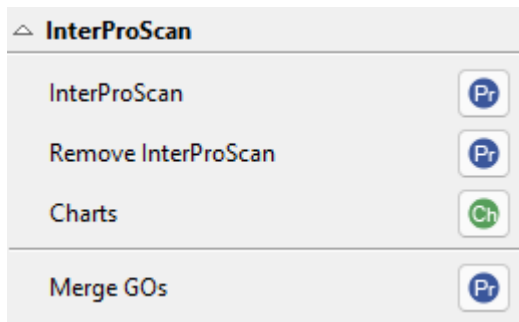
Welcome Message

配列を指定後マウスで右クリック
Show BLAST Resultsは、
選択された配列の類似性検索の
結果に関する情報を含む結果を表示します。

✓表として個々のBLAST結果や
アラインメントビューで個々のBLAST結果を見ることができます。

POINT

タンパク質のドメイン構造やモチーフなどの特徴を推定できます。
BLAST解析と同時に実行できるため解析時間の短縮ができます。



画面中央にロードしたデータが表示した状態で
サイドパネルから解析を実行できます。

InterPro Options



CloudIPS

CloudIPS is a cloud-based OmicsBox PRO community resource for fast and reliable InterPro analysis for everything from small to big data-sets. It allows to execute the original InterPro algorithms against up-to-date databases in our dedicated computing cloud. This is a high-performance, secure and cost-optimized solution for your analysis.

Current CloudUnit balance: 60,000



EMBL-EBI InterPro

Use the public EMBL-EBI InterPro web-service to scan your sequences against InterPro's signatures. Performance and results depend on the EBI web-server.



CloudIPS

メーカー専用のクラウドサーバー上で、
InterProアルゴリズムを実行することができます。
Cloud IPSは、Computation Unitsを使用し、解析を
行うごとに消費されます。

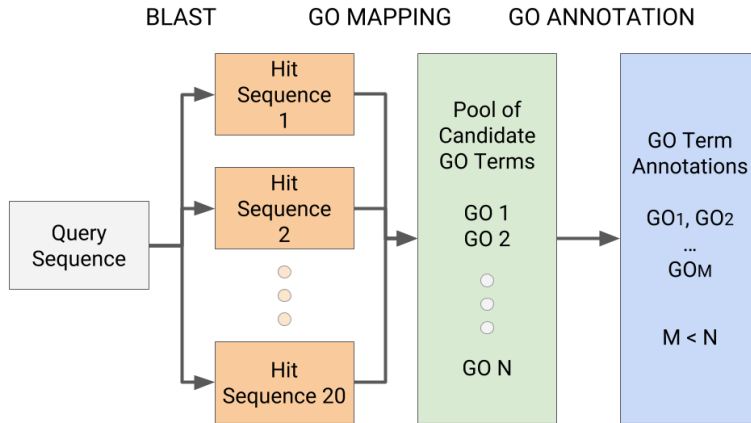
EMBL-EBI interPro

EBIのパブリックWebサービスを介して、
InterProScanを実行します。

Table: blast2go_1000sequences_subset ブラスト								Table: blast2go_1000sequences_subset interpro		
<input checked="" type="checkbox"/>	Nr	Tags	SeqName	Description	Length	#Hits	e-Value	InterPro IDs	InterPro GO IDs	InterPro GO Names
<input checked="" type="checkbox"/>	1	INTERPRO	C02006A02		602			IPR021109 (G3DSA:2.40.70.GENE3D); IPR032799 (PFAM); IPR032861 (PFAM); IPR021109 (G3DSA:2.40.70.GENE3D); IPR001461 (PANTHER); IPR001461 (PANTHER); PTHR13683:SF265 (PANTHER); PTHR13683:SF265 (PANTHER); IPR033121 (PROSITE_PROFILES); IPR021109 (SUPERFAMILY)	F:GO:0004190; P:GO:0006508	F:aspartic-type endopeptidase activity; P:proteolysis
<input checked="" type="checkbox"/>	2	INTERPRO	C02006A04		571			IPR000595 (PFAM); IPR014710 (G3DSA:2.60.120.GENE3D); PTHR10217:SF551 (PANTHER); PTHR10217 (PANTHER); IPR000595 (PROSITE_PROFILES); IPR000595 (CDD); IPR018490 (SUPERFAMILY)	no GO terms	no GO terms
<input checked="" type="checkbox"/>	3	INTERPRO	C02006A08		620			PTHR31606:SF12 (PANTHER); PTHR31606 (PANTHER); cd13214 (CDD); SSF50729 (SUPERFAMILY)	no GO terms	no GO terms
<input checked="" type="checkbox"/>	4	INTERPRO	C02006B02		533					

✓ InterPro ID・GO IDとその名前が配列に付与されます。

POINT 7000以上の研究引用実績のある高品質な遺伝子機能アノテーション



メーカー独自の有名なBlast2GO Methodologyを利用して高品質な遺伝子機能アノテーションを得ることができます。7000以上の研究引用があり、科学コミュニティによって国際的に認められている、新規ゲノムの分析のための主要なバイオインフォマティクスプラットフォームです。

△ **GO Mapping**

- Run GO Mapping Pr
- Remove GO Mapping Pr
- Charts Ch

△ **GO Annotation**

- Run GO Annotation Pr
- Remove GO Annotation Pr
- Charts Ch

GO Mapping

BLAST検索によって得られた情報に関連するGO Termを検索するプロセスです。Blast解析が完了したデータを表示させサイドパネルから解析を実行できます。

GO Annotation

GO Mapping後のデータを使って、さまざま基準で適切なGO Termを抽出します。ヒットシーケンスとクエリシーケンスの類似性、各GOtermのエビデンスコード、およびGOterm候補の存在が考慮されます。

2ステップで簡単に解析できます。

Blast2GO Annotation

#GO	GO IDs	GO Names
7	F:GO:0004190; P:GO:0006508; F:GO:0008233; C:GO:0016020; C:GO:0016021; F:GO:0016787; P:GO:0030163	F:aspartic-type endopeptidase activity; P:proteolysis; F:peptidase activity; C:membrane; C:integral component of membrane; F:hydrolase activity; P:protein catabolic process
9	F:GO:0005216; F:GO:0005249; P:GO:0006811; P:GO:0006813; C:GO:0016020; C:GO:0016021; P:GO:0034220; P:GO:0055085; P:GO:0071805	F:ion channel activity; F:voltage-gated potassium channel activity; P:ion transport; P:potassium ion transport; C:membrane; C:integral component of membrane; P:ion transmembrane transport; P:transmembrane transport; P:potassium ion transmembrane transport

GO Mappingの解析結果画面

#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Names
4	F:GO:0004190; P:GO:0006508; C:GO:0016020; P:GO:0030163	F:aspartic-type endopeptidase activity; P:proteolysis; C:membrane; P:protein catabolic process	EC:3.4.23	Acting on peptide bonds (peptidases)
3	F:GO:0005249; C:GO:0016021; P:GO:0071805	F:voltage-gated potassium channel activity; C:integral component of membrane; P:potassium ion transmembrane transport		

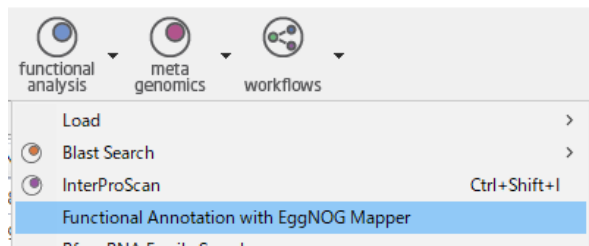
GO Annotationの解析結果画面

Annotation付きGOだけがGO ID列に残ります。

✓アノテーションプロセスによって最適な機能アノテーション情報を付与できます。

POINT

EggNOGベースのオルソロジー割り当てを使用した、新規配列（遺伝子またはタンパク質）の高速機能アノテーションのためのツールです。



Functional Analysis → EggNOG Annotation → EggNOG Mapper
より解析を実行できます。

設定画面ではアノテーション付与の対象となるFastaファイルを選択します。













Type	Query ID	Gene Name	EggNOG Description	E-Value	Bit-Score	#GO	GOs	GO Names	KEGG	Matching OGs
COG ENOG	k141_22_ORF	HDA	Mediates the interaction of DNA replication initiator protein DnaA with ...	1.00E-18	97.1	2	C:GO:0044425; P:GO:0032297	P:negative regulation of DNA-depen...	K10763	08KZ1@bactNOG,0...
COG ENOG	k141_34_ORF		ABC transporter	1.20E-28	130.6	10	C:GO:0016021; F:GO:0042626; P:...	P:purine ribonucleoside triphosphate...	K03529...	0KSKC@eurNOG,0...
COG ENOG	k141_43_ORF		two component, sigma54 specific, transcriptional regulator, Fis family	1.10E-8	64.3	0				05C1W@bactNOG,...
COG ENOG	k141_52_ORF	GOR	Glutathione reductase	3.70E-24	115.5	5	P:GO:0055114; C:GO:0016020; P:...	Figlutathione-disulfide reductase acti...	K00383	05DC8@bactNOG,0...
COG ENOG	k141_53_ORF	DPO2	DNA polymerase I	6.30E-18	94.7	0			K02319	0KRQ8@eurNOG,0...
COG ENOG	k141_55_ORF	NRDA	Provides the precursors necessary for DNA synthesis. Catalyzes the biosy...	5.40E-37	158.3	7	P:GO:0055114; F:GO:0005524; C:...	P:2'-deoxyribonucleotide biosyntheti...	K00525	05BZH@bactNOG,0...
COG ENOG	k141_59_ORF	BOP	bacteriorhodopsin	5.60E-25	118.6	0			K04641	0KZUS@eurNOG,0...
COG ENOG	k141_65_ORF	MUTS	that it carries out the mismatch recognition step. This protein has a weak ...	6.80E-23	111.7	11	C:GO:0032300; P:GO:0006200; F:...	C:mismatch repair complex; P:purine ...	K03555	0KRPX@eurNOG,0...
ENOG	k141_75_ORF			8.10E-20	101.3	0				05MIV@bactNOG,0...
ENOG	k141_80_ORF			2.00E-12	76.6	0				069RU@bactNOG,0...
COG ENOG	k141_85_ORF		amino acid	5.30E-45	185.3	0			K03294	07THB@bactNOG,0...
COG ENOG	k141_92_ORF	DESR	regulator, LuxR family	4.90E-43	178.7	1	C:GO:0005737	C:cytoplasm	K07693	08S01@bactNOG,0...
COG ENOG	k141_95_ORF	LIGA	DNA ligase that catalyzes the formation of phosphodiester linkages betw...	3.70E-62	242.7	10	F:GO:0000287; P:GO:0006266; P:...	P:DNA ligation; F:magnesium ion bin...	K01972	05C77@bactNOG,0...

✓EggNOG Mapperで転送できたアノテーションが表にまとめられます。

Annotation details for C02006C04

EggNOG Description:	Pentatricopeptide repeat-containing protein
EggNOG Protein:	XP_006426850.1 (Citrus clementina)
E-Value:	2.9E-9
Bit Score:	68.9
Best Taxonomic Level:	Streptophyta
Taxonomic Scope:	Viridiplantae[1]
KEGG KO:	K17710
BRITE:	ko00000, ko03016, ko03029
Matching Orthologous Groups:	37SKF@Viridiplantae, 3G7B8@Streptophyta, KOG4197@root, KOG4197@Eukaryota
COG Categories:	S

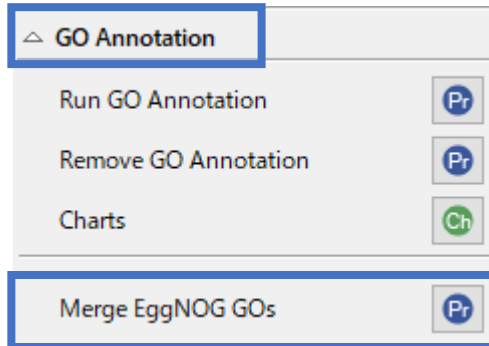
Related GOs

 GO:0090305 	 GO:0008380 	 GO:0003723 	 GO:0009737 
Name nucleic acid phosphodiester bond hydrolysis 	Name RNA splicing 	Name RNA binding 	Name response to abscisic acid 
Definition The nucleic acid metabolic process in which the phosphodiester bonds between nucleotides are cleaved by hydrolysis.	Definition The process of removing sections of the primary RNA transcript to remove sequences not present in the mature form of the RNA and joining the remaining sections to form the mature form of the RNA.	Definition Interacting selectively and non-covalently with an RNA molecule or a portion thereof.	Definition Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an abscisic acid stimulus.

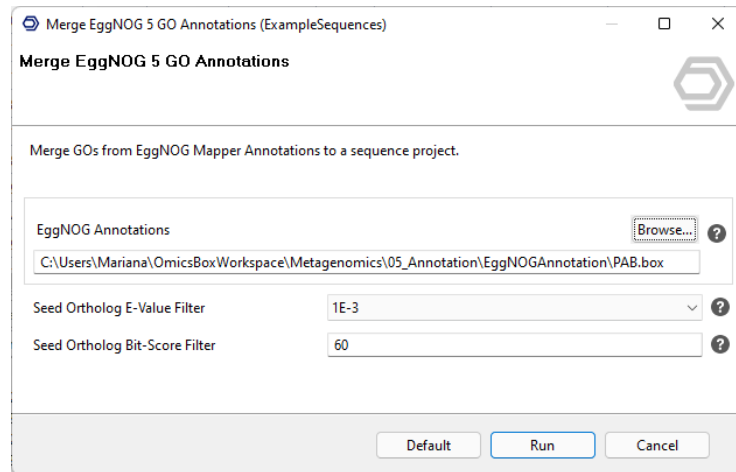
✓レポートを作成するとGOの総数、COGカテゴリ、およびオーソロググループの分布に関する情報を閲覧できます。

Merge EggNOG to Annotation

EggNOG to Annotationの結果を機能アノテーション結果にマージする。



Blast2GO Annotation解析済み（青色のテーブルが）表示された状態でサイドパネルから解析を実行できます。

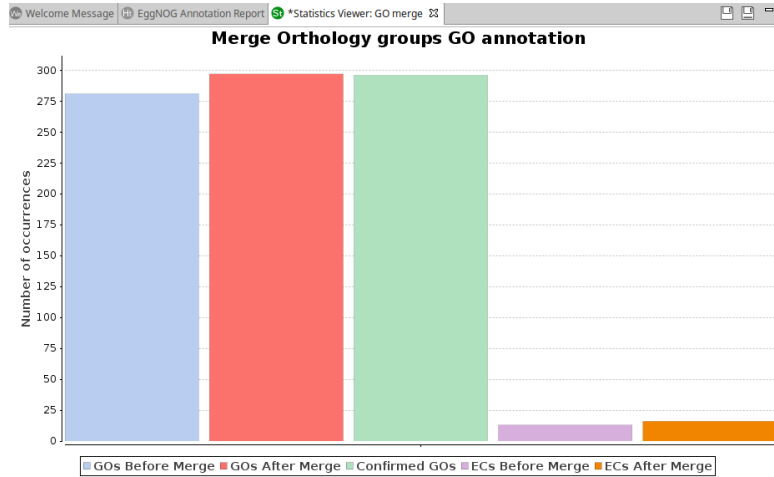


EggNOG to Annotationデータをこちらで選択します。



EggNOG to Annotationデータは事前に保存する必要があります。結果画面テーブルの右上のSave asより.box形式で保存することが可能です。

Merge EggNOG to Annotation

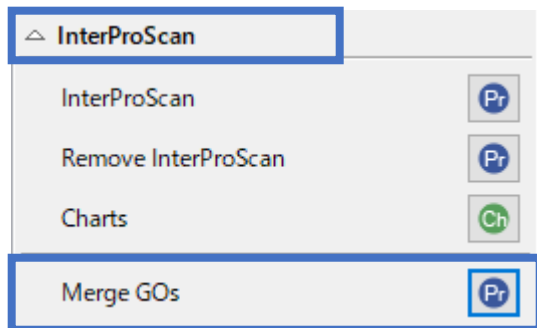


マージが完了すると、現在のアノテーション結果に追加（または確認）されたGOの数を示す分布チャートが結果メニューに表示されます。

<input checked="" type="checkbox"/>	3	INTERPRO BLASTED	Solyc00g005050.3.1	UPF0664 stress-induced...	588	20	1.19E-143	92.53%		
<input checked="" type="checkbox"/>	3	INTERPRO BLASTED ANNOTATED	Solyc00g005050.3.1	UPF0664 stress-induced...	588	20	1.19E-143	92.53%	1	C:GO:0005886 C:plasma membrane

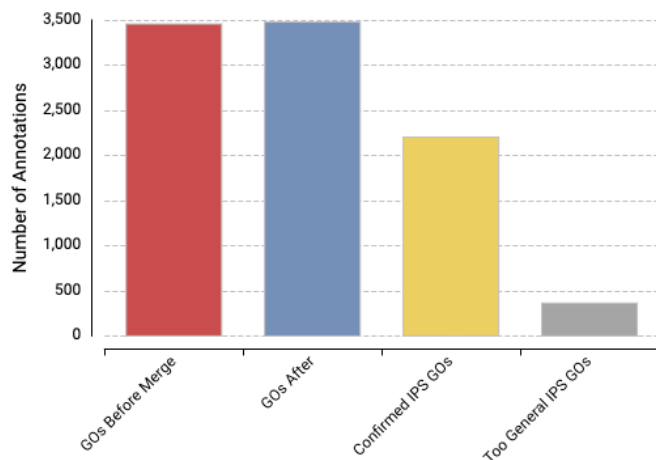
✓マージが完了すると、GO情報が新たに付与されたことが確認できます。

InterProScanの結果を機能アノテーション結果にマージする。




Blast2GO Annotation解析済み（青色のテーブルが）表示された状態でサイドパネルから解析を実行できます。

**Merge Interpro Annotation Results
[Example Sequences]**



マージが完了すると、現在のアノテーション結果に追加（または確認）されたGOの数を示す分布チャートがデータタブに表示されます。

Merge InterProScan to Annotation

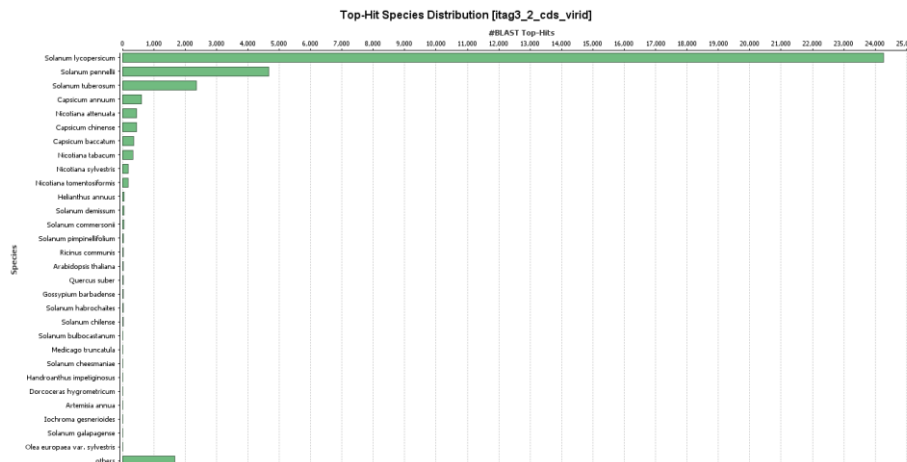
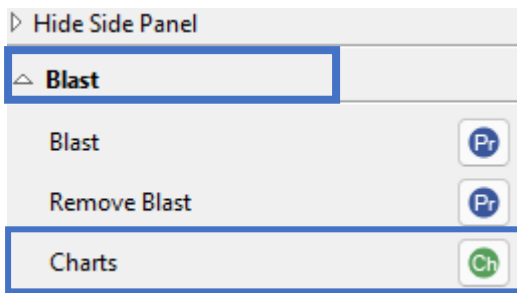
<input checked="" type="checkbox"/>	105		Solyc00g009030.2.1	sugar transport protein ...	1281	20	0E0	71.13%	3	F:GO:0005215; P:GO:0006810; C:GO:0016020	F:transporter activity; P:transport; C:membrane
-------------------------------------	-----	---	--------------------	-----------------------------	------	----	-----	--------	---	--	--



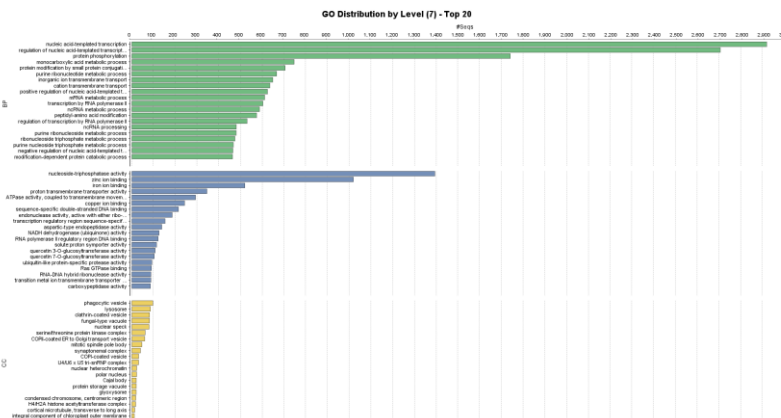
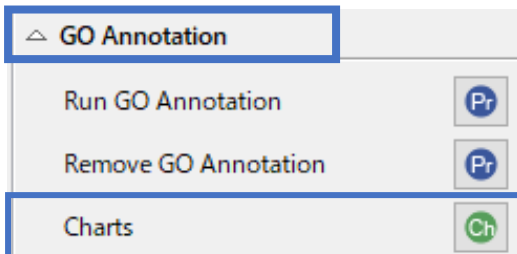
<input checked="" type="checkbox"/>	105		Solyc00g009030.2.1	sugar transport protein ...	1281	20	0E0	71.13%	3	C:GO:0016021; F:GO:0022857; P:GO:0055085	C:integral component of membrane; F:transmembrane transporter activity; P:transmembrane transport
-------------------------------------	-----	---	--------------------	-----------------------------	------	----	-----	--------	---	--	---

マージが完了すると、GOID、nameがInterProScan解析由来のものに置き換わります。
✓最も具体的なGotermにアノテーションを付けることができます。

POINT アノテーション解析結果から様々なグラフを簡単に作成できます。ここではその一部をご紹介します。

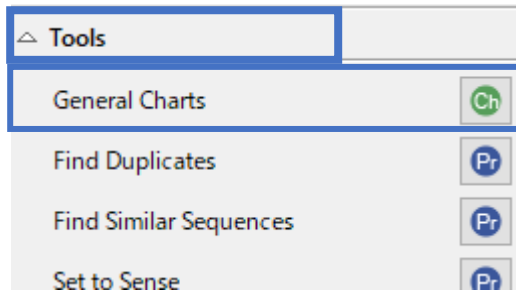


BLASTで検索された生物種のヒット数

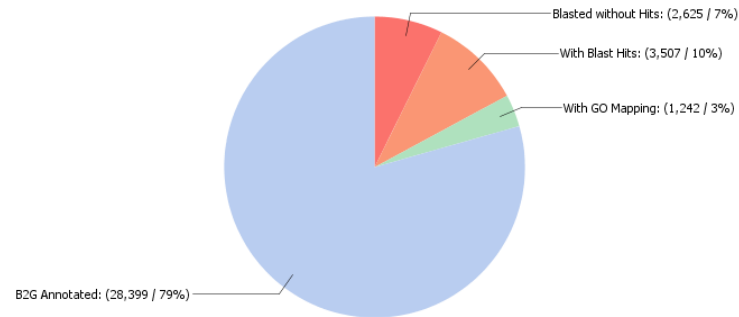


各アノテーションの結果項目より作図できます。

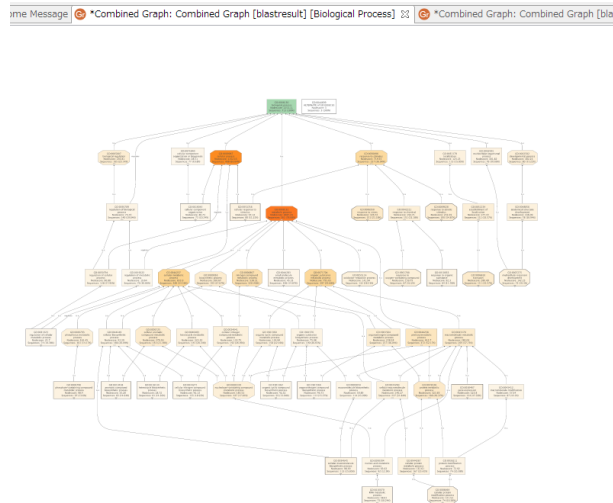
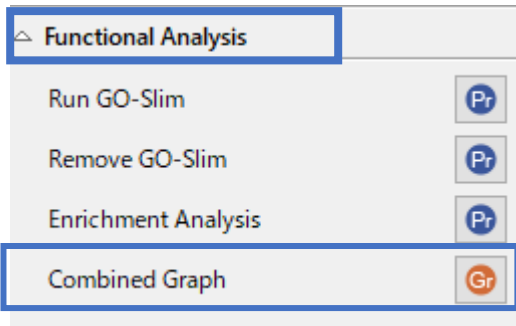
各GO結果のランク図



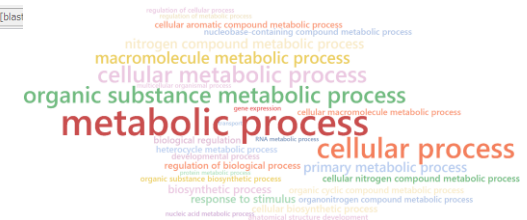
Data Distribution Pie Chart [ITSG3.2_CDS_6_EC]



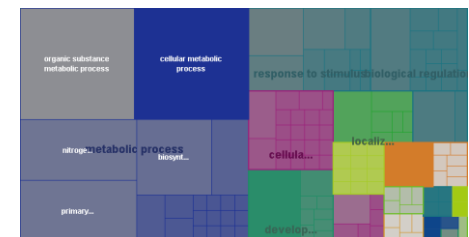
全ての解析結果の円グラフ



GOグラフ



ワードクラウド

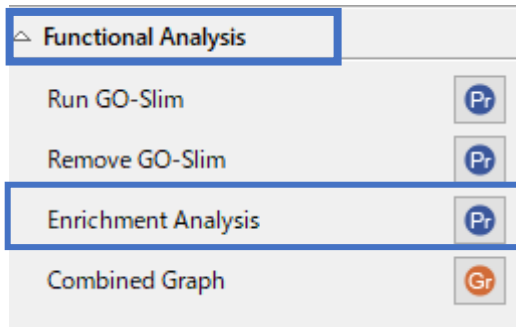


ツリービュー

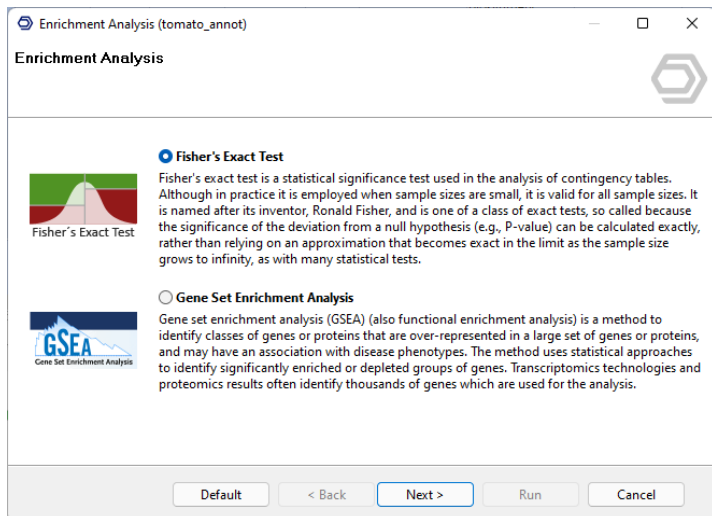
✓アノテーション結果をグラフィカルに要約するための多くの追加の視覚化ツールが搭載されています。

POINT

遺伝子のセットの中で過大または過小な遺伝子機能を特定するエンリッチメント解析が搭載されています。



Blast2GO Annotation解析済み（青色のテーブルが）表示された状態でサイドパネルから解析を実行できます。



Fisher's Exact Test

2つの遺伝子リストの機能アノテーションを互いに比較するします。例えばGO Termについて、参照セットと比較して、より頻繁であるかどうかを検定します。

Gene Set Enrichment Analysis

ある遺伝子セット（同じGO Termでアノテーションされた遺伝子群）の遺伝子が、ランク付けされた遺伝子リストの上位または下位に集積しているかどうかを調べるツールです。遺伝子リストは、例えば発現値やメチル化レベルなど、生物学的な意味を持つ任意の指標でランク付けされたものを使用します。

Fisher's Exact Testの設定画面

テストセットに属するシーケンスを含むIDリスト

テストセットに該当するアノテーション結果を右クリックしIDリストを作成・保存したものをこちらで選択します。

参照セットに属するシーケンスを含むIDリスト

参照セットに該当するアノテーションの結果が表示されている場合は入力不要です。

しきい値の設定

FDRまたはP値のカットオフの値を指定します。

Two Tailed

両側検定の場合はこちらを選択し片側検定の場合はチェックをはずします。

Remove double IDs

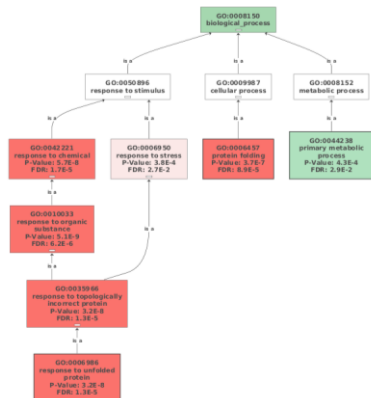
もし両方のリストに共通する配列があり、比較の方向に依存しない検定を行いたい場合は、こちらのオプション選択してください。

Enrichment Analysis (Fisher's Exact Test)

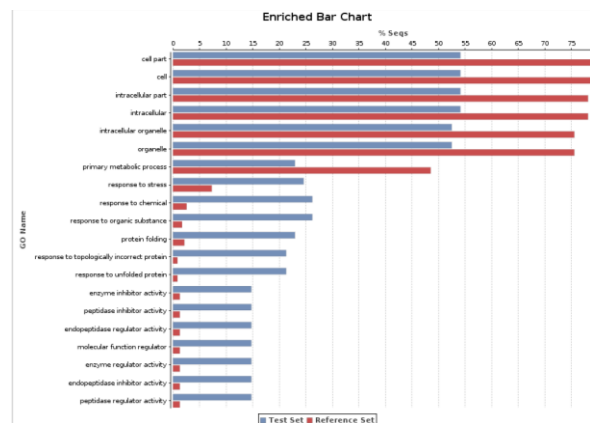
Nr	Tags	GO ID	GO Name	GO Cat...	FDR	P-Value	Nr Test	Nr Refe...	Non An...	Non An...
1	OVER	GO:0010...	response to organic substance	BIOLOGI...	6.24216...	5.14605...	16	4	45	229
2	OVER	GO:0035...	response to topologically incorrect protein	BIOLOGI...	1.31128...	3.24307...	13	2	48	231
3	OVER	GO:0006...	response to unfolded protein	BIOLOGI...	1.31128...	3.24307...	13	2	48	231
4	OVER	GO:0042...	response to chemical	BIOLOGI...	1.71910...	5.66892...	16	6	45	227
5	OVER	GO:0006...	protein folding	BIOLOGI...	8.93100...	3.68137...	14	5	47	228
6	OVER	GO:0098...	molecular function regulator	MOLECU...	5.63034...	5.57000...	9	3	52	230
7	OVER	GO:0030...	enzyme regulator activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
8	OVER	GO:0004...	enzyme inhibitor activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
9	OVER	GO:0004...	endopeptidase inhibitor activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
10	OVER	GO:0030...	peptidase inhibitor activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
11	OVER	GO:0061...	peptidase regulator activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
12	OVER	GO:0061...	endopeptidase regulator activity	MOLECU...	5.63034...	5.57000...	9	3	52	230
13	UNDER	GO:0044...	cell part	CELLULA...	2.13965...	2.46951...	33	184	28	49
14	UNDER	GO:0005...	cell	CELLULA...	2.13965...	2.46951...	33	184	28	49
15	UNDER	GO:0044...	intracellular part	CELLULA...	2.37036...	3.12661...	33	182	28	51
16	UNDER	GO:0005...	intracellular	CELLULA...	2.37036...	3.12661...	33	182	28	51
17	OVER	GO:0006...	response to stress	BIOLOGI...	2.72839...	3.8238E-4	15	17	46	216
18	UNDER	GO:0044...	primary metabolic process	BIOLOGI...	2.89128E-2	4.29044...	14	113	47	120
19	UNDER	GO:0043...	intracellular organelle	CELLULA...	4.75799...	7.84501E-4	32	176	29	57
20	UNDER	GO:0043...	organelle	CELLULA...	4.75799...	7.84501E-4	32	176	29	57

エンリッチメント解析結果テーブル

GO termが過剰か過少か示すタグとFDRとP値が記載されています。



GOグラフ
重要度の値に比例して色が強調表示されます。



棒グラフ
アノテーションごとに、シーケンスのパーセンテージの棒グラフを生成します。

その他ワードクラウドや Treemapを作図可能。

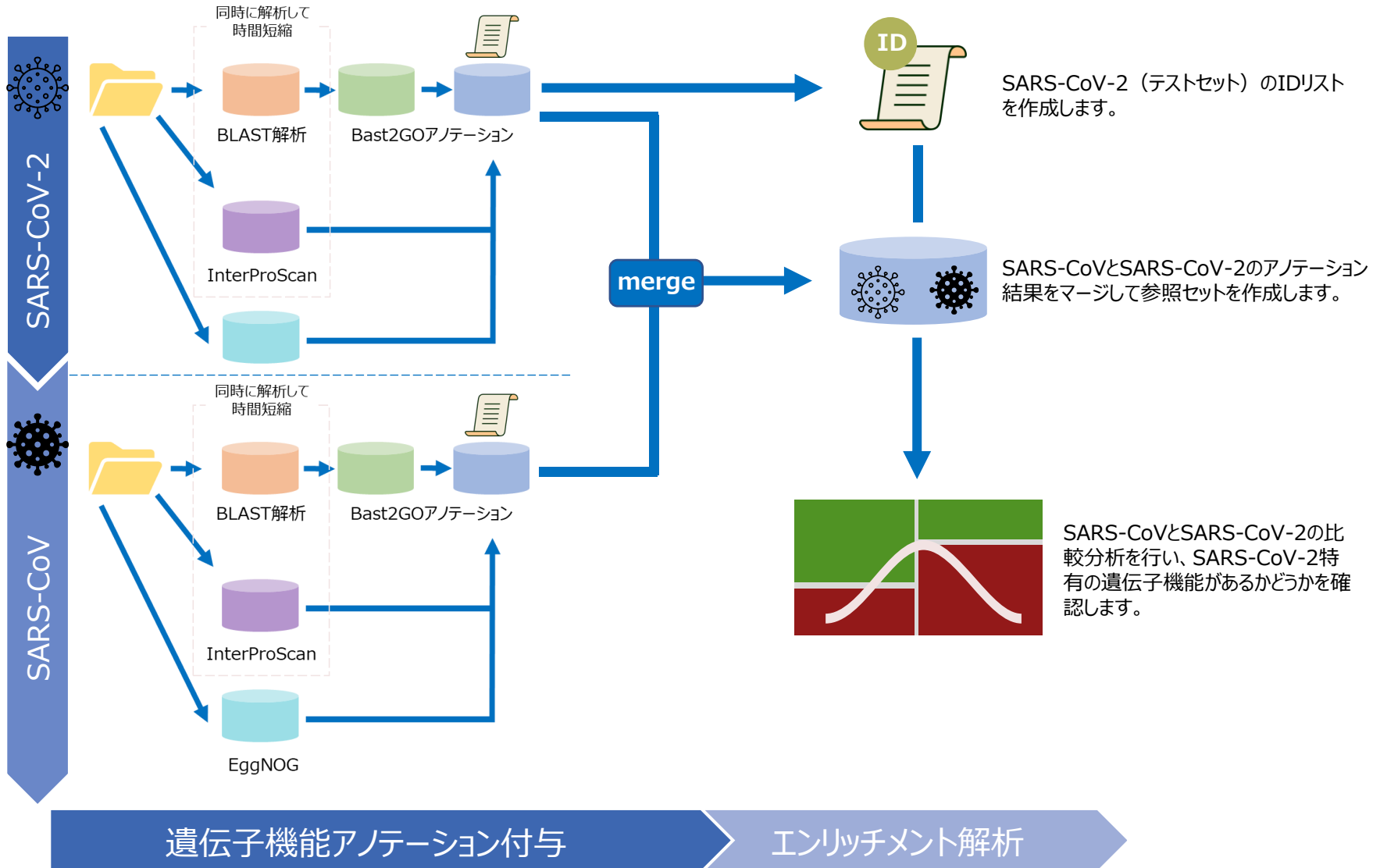
✓アノテーション結果を用いて統計検定を行い、さらにはグラフの作成まで可能です。

目的

- SARS-CoV-2タンパク質配列に遺伝子機能情報を付与します。
- SARS-CoVとSARS-CoV-2の比較分析を行い、SARS-CoV-2特有の遺伝子機能があるかどうかを確認します。



✓今回タンパク質配列について解析を行います。コンティグデータのような塩基情報に対しても同じような解析を行うことが可能です。

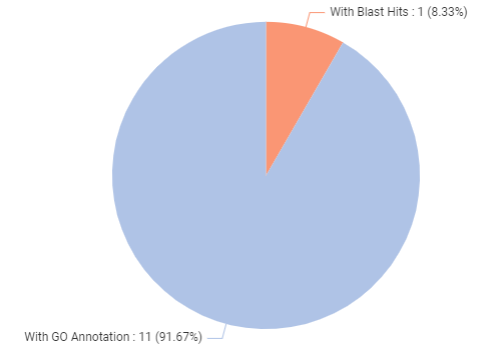


SARS-CoVとSARS-CoV-2それぞれ
遺伝子機能アノテーション解析を実行します。

チュートリアル

☑	Nr	Tags	SeqName	Description	Length	#Hits	e-Value	sim mean	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs	InterPro GO IDs	InterPro GO Names	
☑ 3		INTERPRO BLASTED MAPPED ANNOTATED	YP_009724391.1	ORF3a protein [S... 275	275	20	0E0	99.96%	28	P:GO:0019081; P:GO:004220; P:GO:0039707; F:GO:0005216; F:GO:0005515; C:GO:0005578; C:GO:0005788; C:GO:0005886; C:GO:016021; C:GO:0020002; C:GO:0041174; C:GO:0044178; C:GO:0044187; C:GO:0044218; C:GO:0055036	membrane or host cell; Fion channel activity; Fprotein binding; C:extracellular region; C:endoplasmic reticulum lumen; C:plasma membrane; C:integral component of membrane; C:host cell plasma membrane; C:host cell endosome; C:host cell Golgi membrane; C:host cell lysosome; C:other organism cell membrane; C:virion membrane	EC:7	Translocases	IPR024407 (PFAM); IPR024407 (CDD)	P:GO:0039707; F:GO:0005216	P:GO:0039707; F:GO:0005216	P:pore formation by virus in membrane of host cell; Fion channel activity
☑ 4		INTERPRO BLASTED MAPPED ANNOTATED	YP_009724392.1	envelope protein	75	20	2.7E-43	98.93%	28	P:GO:0019081; P:GO:0039707; P:GO:0044116; P:GO:0044662; P:GO:0046760; P:GO:0051704; P:GO:0051818; F:GO:0042802; C:GO:0005788; C:GO:0016021; C:GO:0019031; C:GO:0030666; C:GO:0042025; C:GO:0044165; C:GO:0044178; F:GO:0142317	P:oviral translation; P:pore formation by virus in membrane of host cell; P:induction by symbiont of host defense response; P:disruption by virus of host cell membrane; P:oviral budding from Golgi membrane; P:obsolete multi-organism process; P:obsolete disruption of cells of other organism involved in symbiotic interaction; P:identical protein binding; C:endoplasmic reticulum lumen; C:integral component of membrane; P:oviral			G3DSA6.10.250.1810 (GENE3D); IPR03873 (PFAM); IPR03973 (PROSITE_PROFILES); IPR043506 (HAMAP); IPR044377 (CDD)	P:GO:0019068; P:GO:0039707; P:GO:0046760; C:GO:0016021	P:oviral assembly; P:pore formation by virus in membrane of host cell; P:oviral budding from Golgi membrane; C:antigenic component of membrane	

Data Distribution Pie Chart [06_Merge_EggNOG]



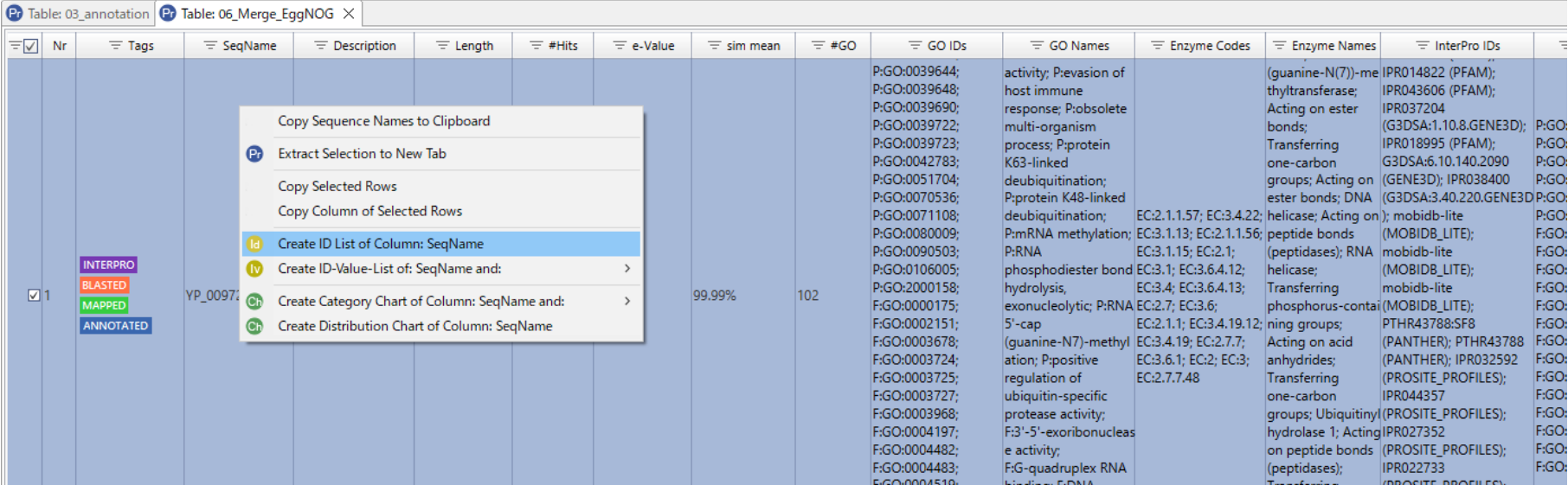
SARS-CoV-2のアノテーション解析結果 (エンリッチメント解析は除く)

高品質な
アノテーション

約32分
処理時間

CloudBlast : 19分/ 1174 ユニット
CloudInterProScan : 4分/ 194 ユニット
EggNOG-Mapper : 6分
GOマッピング : 2分
その他は 1 分未満

テストセットIDリストを作成する



Nr	Tags	SeqName	Description	Length	#Hits	e-value	sim mean	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Names	InterPro IDs
1	INTERPRO BLASTED MAPPED ANNOTATED	YP_0097...				99.99%		102	P:GO:0039644; P:GO:0039648; P:GO:0039690; P:GO:0039722; P:GO:0039723; P:GO:0042783; P:GO:0051704; P:GO:0070536; P:GO:0071108; P:GO:0080009; P:GO:0090503; P:GO:0106005; P:GO:2000158; F:GO:0000175; F:GO:0002151; F:GO:0003678; F:GO:0003724; F:GO:0003725; F:GO:0003727; F:GO:0003968; F:GO:0004197; F:GO:0004482; F:GO:0004483; F:GO:0004519;	activity; P:evasion of host immune response; P:obsolete multi-organism process; P:protein K63-linked deubiquitination; P:protein K48-linked deubiquitination; P:mRNA methylation; P:RNA phosphodiester bond hydrolysis, exonucleolytic; P:RNA 5'-cap (guanine-N7)-methylation; P:positive regulation of ubiquitin-specific protease activity; F:3'-5'-exoribonuclease activity; F:G-quadruplex RNA binding; F:DNA	EC:2.1.1.57; EC:3.4.22; EC:3.1.13; EC:2.1.1.56; EC:3.1.15; EC:2.1; EC:3.1; EC:3.6.4.12; EC:3.4; EC:3.6.4.13; EC:2.7; EC:3.6; EC:2.1.1; EC:3.4.19.12; EC:3.4.19; EC:2.7.7; EC:3.6.1; EC:2; EC:3; EC:2.7.7.48	helicase; Acting on one-carbon groups; Acting on ester bonds; DNA helicase; Acting on peptide bonds (peptidases); RNA helicase; Transferring phosphorus-containing groups; Acting on acid anhydrides; Transferring one-carbon groups; Ubiquitinyl hydrolase 1; Acting on peptide bonds (peptidases); Transferring	IPRO14822 (PFAM); IPR043606 (PFAM); IPR037204 (G3DSA:1.10.8.GENE3D); IPR018995 (PFAM); G3DSA:6.10.140.2090 (GENE3D); IPR038400 (G3DSA:3.40.220.GENE3D); mobidb-lite (MOBIDB_LITE); mobidb-lite (peptidases); RNA mobidb-lite (MOBIDB_LITE); mobidb-lite (MOBIDB_LITE); Transferring phosphorus-containing groups; (PANTHER); IPR032592 (MOBIDB_LITE); (PROSITE_PROFILES); IPR044357 (PROSITE_PROFILES); IPR027352 (PROSITE_PROFILES); IPR022733 (PROSITE_PROFILES); (PROSITE_PROFILES);

1. テストセットとなるアノテーション結果（今回はSARS-CoV-2データ）を表示させます。
2. Ctrl + A（WindowsおよびLinux） / Cmd + A（Mac）でリスト内を全選択します。
3. SeqNameの欄で右クリックし「Create ID List of Column: SeqName」を選択します。

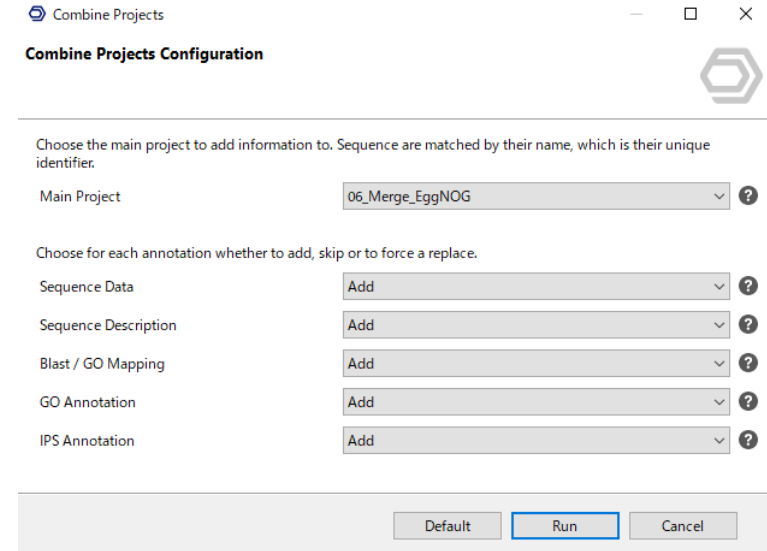
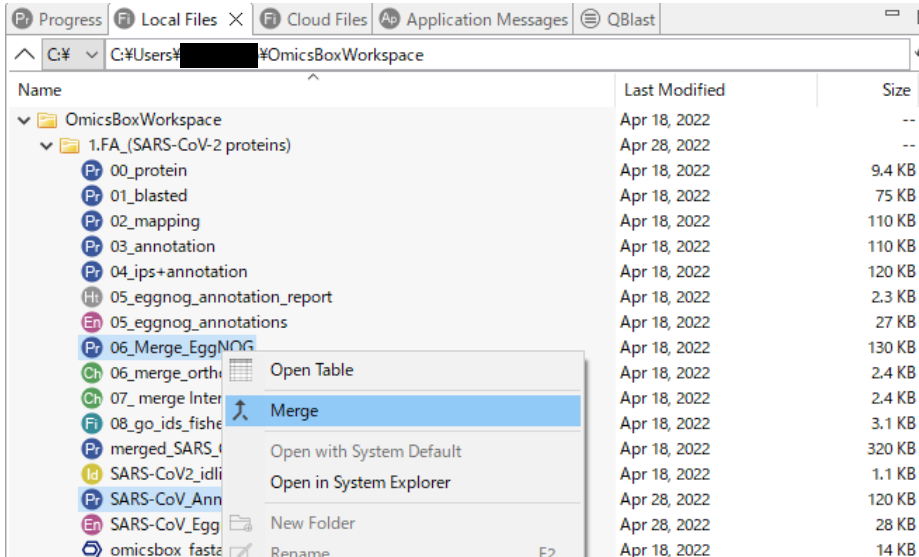
注） SeqNameの欄以外で右クリックするとその欄に対応したIDが作成されます。

例えばDescriptionの欄で右クリックするとCreate ID List of Column: Descriptionが表示されDescriptionのリストが作成されます。

4. 新しいタブにリストが作成されるので.box形式で保存します。



参照セット表示のため2つのアノテーション結果をマージして表示させる



1. 2つのアノテーション結果をOmicsBoxWork speceに表示させます。

注) OmicsBoxWork speceはLocal Filesのタブを選択するとファイルロケーションも表示されます。
この時Cloud Filesのタブは利用しないようにご注意ください。

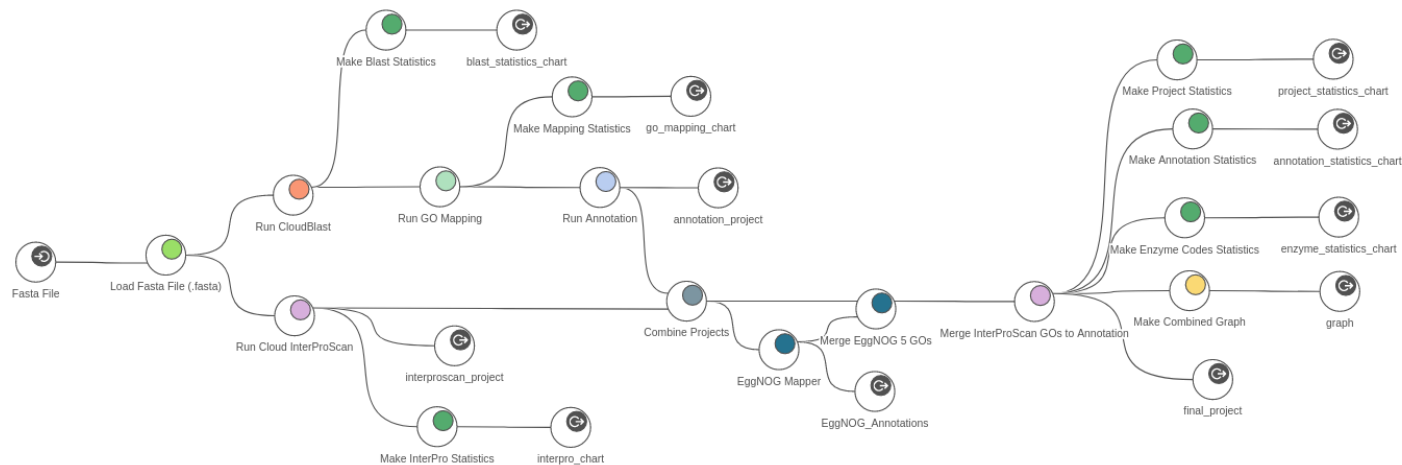
2. 2つのアノテーション結果選択し右クリックからMergeを選択します。

3. すべての結果を追加し、Runをクリックするとマージプロジェクトが開きます。

Nr	Tags	GO ID	GO Name	GO Category	FDR	P-Value
1	OVER	GO:0044174	host cell endosome	CELLULAR_COMPONENT	8.289486E-1	9.810479E-3

エンリッチメント解析結果

✓ SARS-CoV-2に特異的な1つのGene Ontology term
「host cell endosome」が抽出された。



各解析をワークフローにて1度に解析を実行することができます。
もちろん設定の変更なども柔軟に行うことができます。
すでに出来上がった既存のワークフローからも解析を行うことが可能です。

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