

COVGAS

---

Corona virus Genome Assembly and subtyping

# Administrative User Guide

Version r.20210115

ADMINISTRATIVE USER GUIDE

## **Corona virus Genome Assembly and subtyping**

---

© world fusion  
1-38-12 kakigara-cho nihonbashi  
chuoku, Tokyo 103-0014  
Phone 03.3662.0521 • Fax 03.3662.0522

Copyright (C) 2002 World Fusion Co.,LTD.

All Rights Reserved. under World Fusion Inc. No part of this documentation may be reproduced, stored, or transmitted in any form without the prior permission of World Fusion Inc.

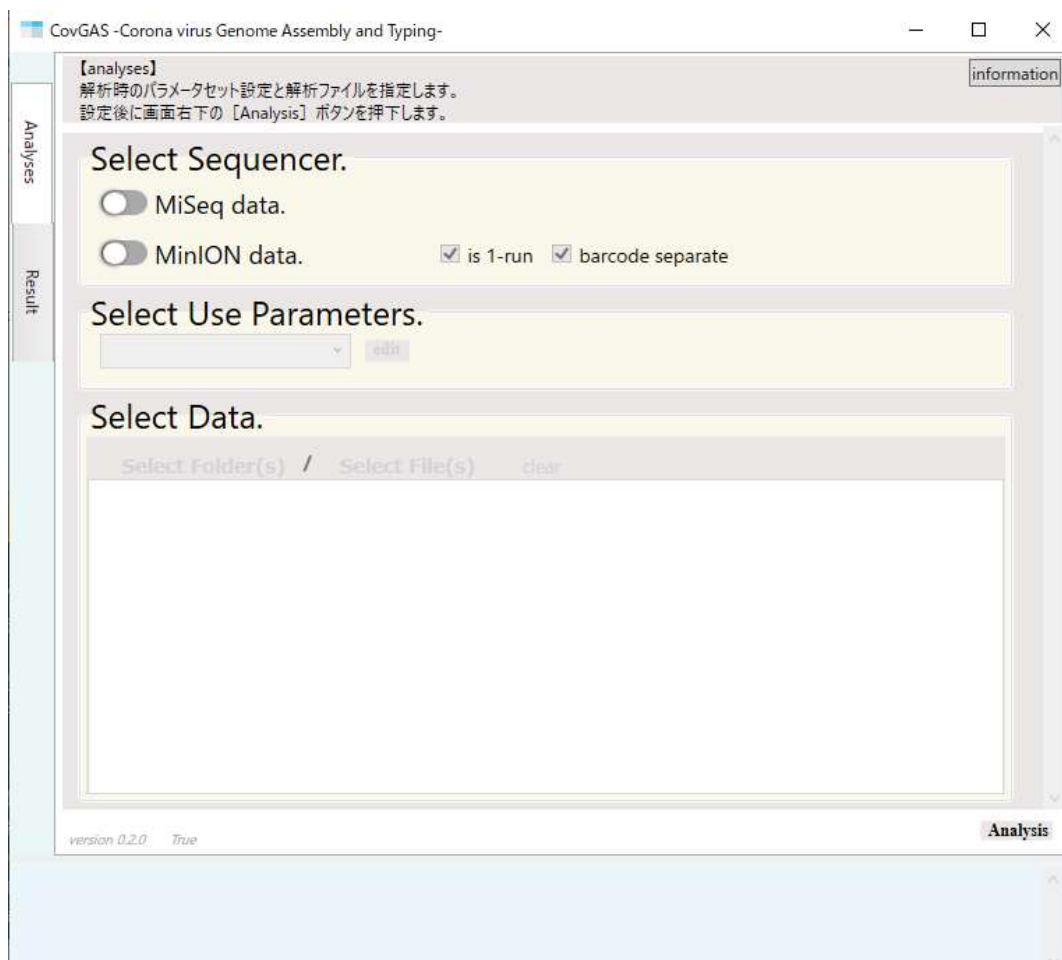
---



## マウス ウィンドウ項目 【画面構成】

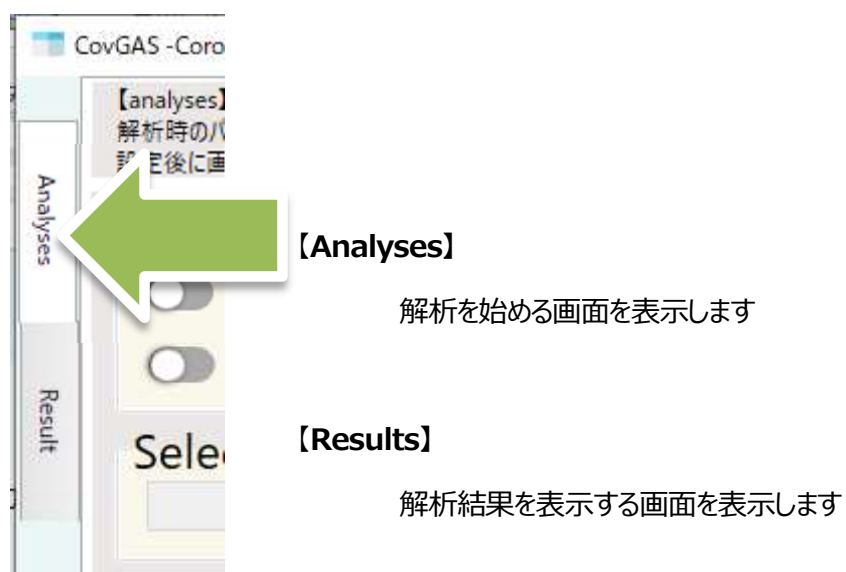
### ・初期画面

プログラムの実行は、CovGAS.exe プログラム本体、もしくはショートカット をダブルクリックして開始します。



※初回起動時はライセンスと Guppy プログラムの設定を行うダイアログが表示されます。

・画面項目（左側メニュー）



・画面項目（下部ログメッセージ）



アプリケーションで行われている内容を表示します。実行中のタスクやエラーログなどが表示されます。

- ・解析画面

【analyses】  
解析時のパラメータセット設定と解析ファイルを指定します。  
設定後に画面右下の【Analysis】ボタンを押下します。

Select Sequencer.

☐ MiSeq data.

☐ MinION data. ☒ is 1-run ☒ barcode separate

Select Use Parameters.

Select Data.

Select Folder(s) / Select File(s) clear

**【Select Sequencer】**

解析に利用するデータを出力したシーケンサを選択します

**【Select Use Parameters】**

解析時に利用するパラメータを選択します

初期状態は【default】が選択されています。右にある【edit】ボタンをクリックすると選択されているシーケンサに対して設定できる画面へ遷移します

**【Select Data】**

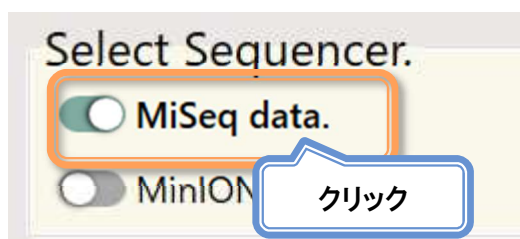
解析を行うデータを格納している【フォルダ】、もしくは【ファイル】を、枠内へドラッグ・アンド・ドロップします  
【Select Folder(s)】や【Select File(s)】をクリックすることでも選択が可能です



## 解析を行う 【Illumina シーケンサ】

Illumina Illumina シーケンサから出力された Fastq（または gz 圧縮された Fastq）ファイルを解析します。

- ・ [Select Sequencer]設定の [Illumina data] チェックボタンをクリックします



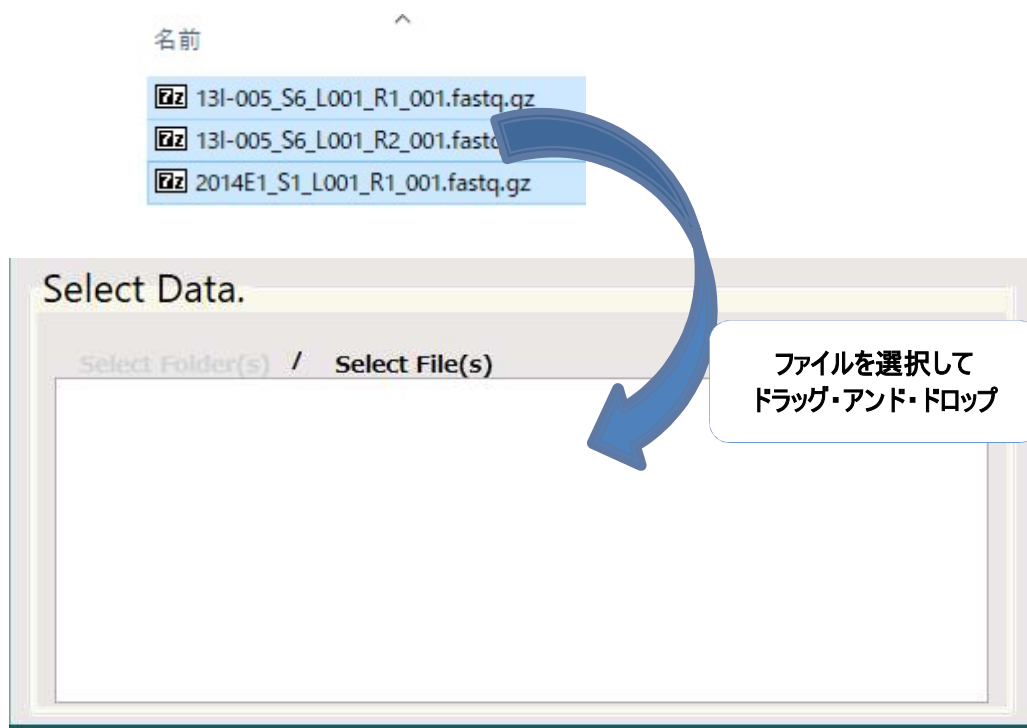
※ [Illumina data]と[Minion data]は排他です。混合サンプルの解析には対応していません

- ・ [Select Use Parameters]設定で使用するパラメータセットを選択します



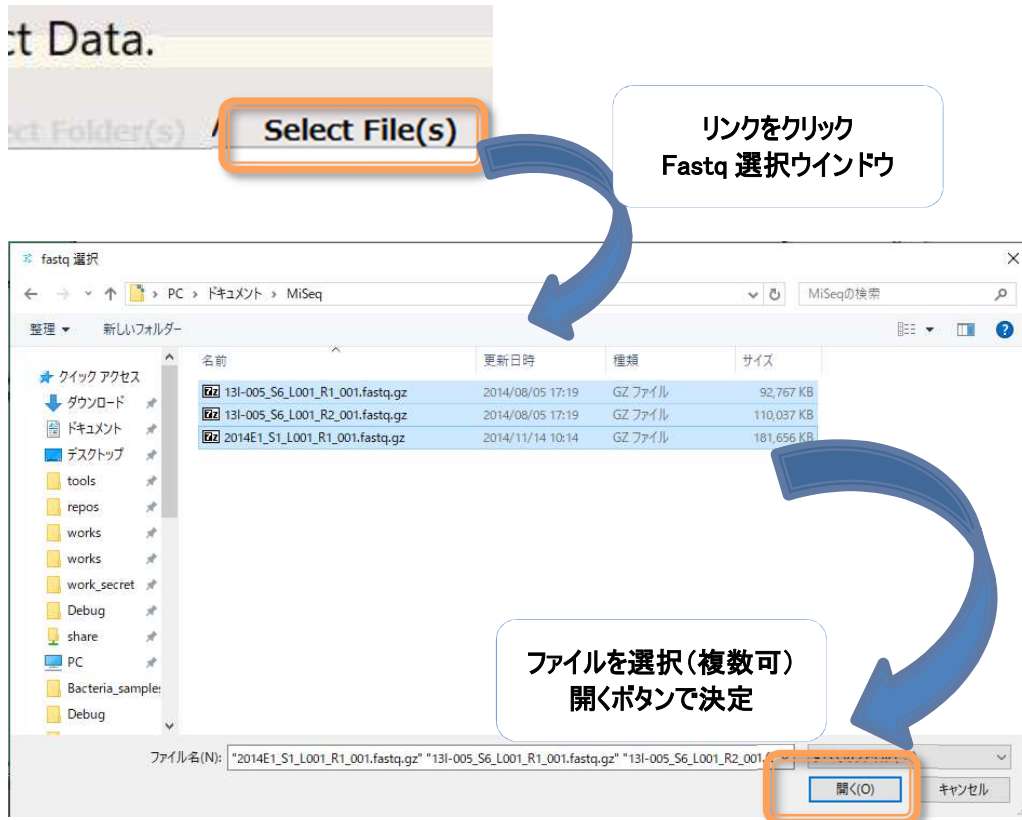
※ 解析時のパラメータを選択します。CovGAS は解析実行前にパラメータを全て設定する必要があります。解析時に使用する詳細なパラメータを変更するには横の【edit】ボタンをクリックしパラメータ画面にて編集を行います。

- ・ [Select Data]設定のテキストボックス内へ Fastq ファイルをドラッグ&ドロップします



※ Illumina のデータは、拡張子が「fastq」または「fastq.gz」の拡張子ファイルが指定可能です。上記以外のファイルを指定するとエラーとなります。

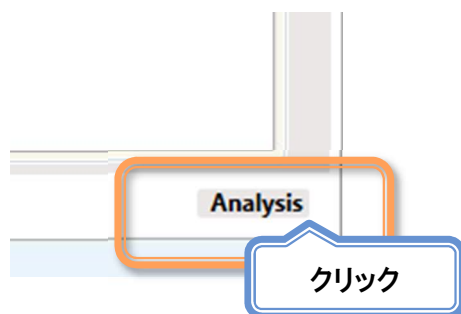
または [Select File(s)] ボタンをクリックすることにより、ファイル選択ウィンドウから、解析を行う[Fastq ファイル]を選択（複数サンプル可）もできます。



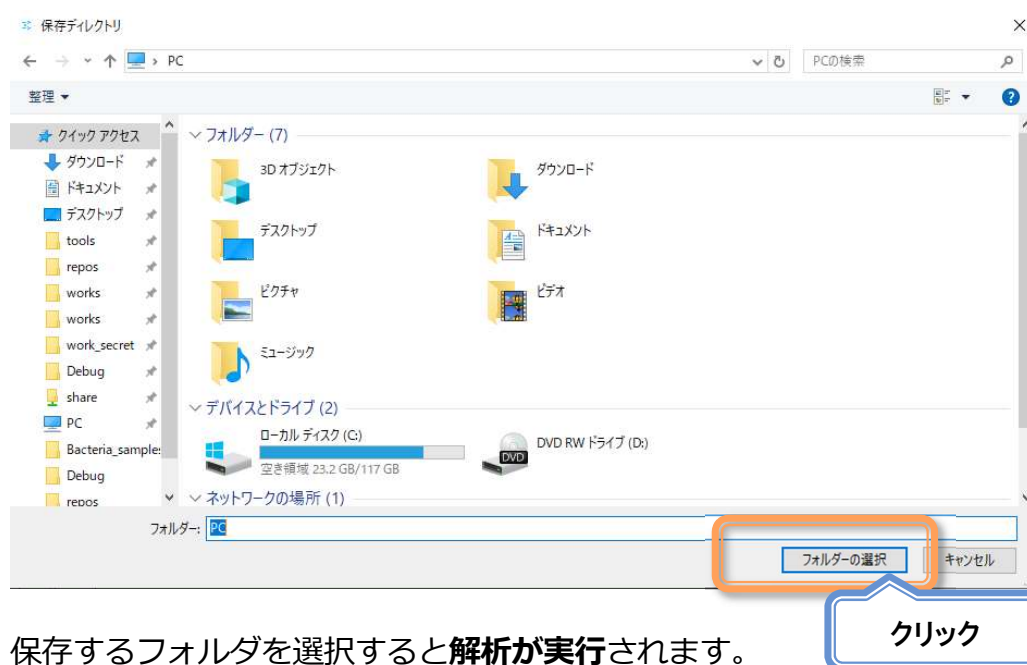
※ Illumina シーケンサのデータを解析する場合、フォルダの選択はできません。



- ・ [Analysis]ボタンをクリックし保存場所を指定します。



解析データの指定を行ったのちに[Next]ボタンをクリックすると、解析ファイルの保存場所を指定するウインドウが表示されます。



保存するフォルダを選択すると**解析が実行**されます。

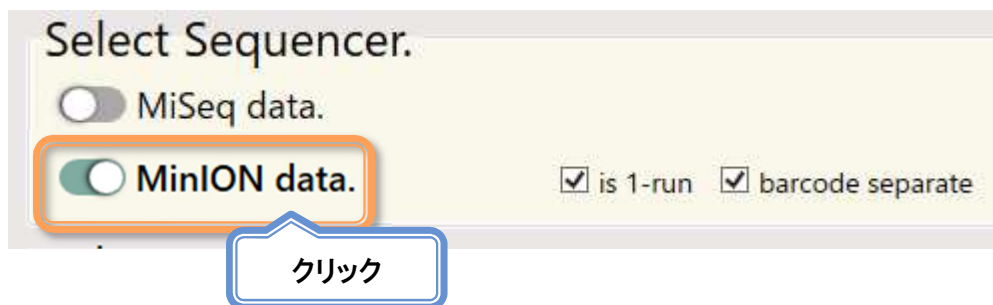
※ 日本語やスペースを含むフォルダを指定しないでください。外部ツールでエラーとなる場合があります。

実行中は下部ログメッセージに実行中のコマンドが表示されます。

## 🖱️ 解析を行う 【MinION シーケンサ】

Nanopore MinION シーケンサから出力された Fast5 または Fastq ファイルを解析します

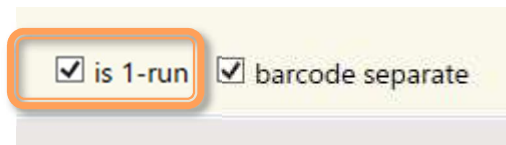
- ・ [Select Sequencer]設定の [MinION data] ボタンをクリックします



※ [Illumina data]と[Minion data]は排他です。混合サンプルの解析には対応していません

MinION Fast5 から解析する場合、フォルダの指定が可能です。

選択するフォルダ・ファイルが 1 ランのデータの場合 [is 1-sample]をチェックします。

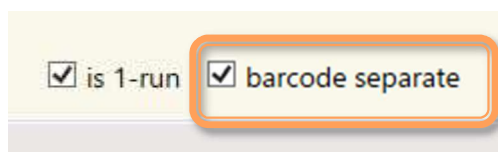


チェックが入っている場合、[Select Data]に指定したフォルダもしくはファイルを全てが 1 ランのサンプルとして、1 解析分のデータとして利用され、Barcode に基づくサンプル分割が行われます

チェックを外すと、指定された 1 ファイルもしくは 1 フォルダを別々のサンプルとして解析します。

MinION Fast5 から解析する場合、バーコード分割を行います。バーコード分割は PC にインストールされた Guppy のプログラムを使用します。

対象をバーコード分割する場合は、[barcode separate]をチェックします。

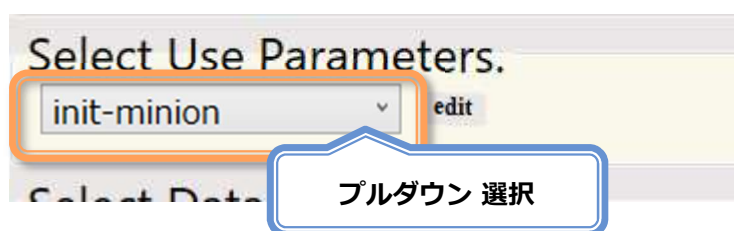


チェックが入っている場合、[Select Data]に指定した対象を Basecall したのちにバーコード分割を行います。  
チェックを外すと、指定された 1 ファイルもしくは 1 フォルダを別々のサンプルとして解析します。

[barcode separate]にチェックが入っている場合、次画面で各バーコードに対応したサンプル名が指定可能です。

※詳細はバーコード分割画面を参照してください。

- ・ [Select Use Parameters]設定で使用するパラメータセットを選択します



※ 解析時のパラメータを選択します。CovGAS は解析実行前にパラメータを全て設定する必要があります。解析時に使用する詳細なパラメータを変更するには横の【edit】ボタンをクリックしパラメータ画面にて編集を行います。

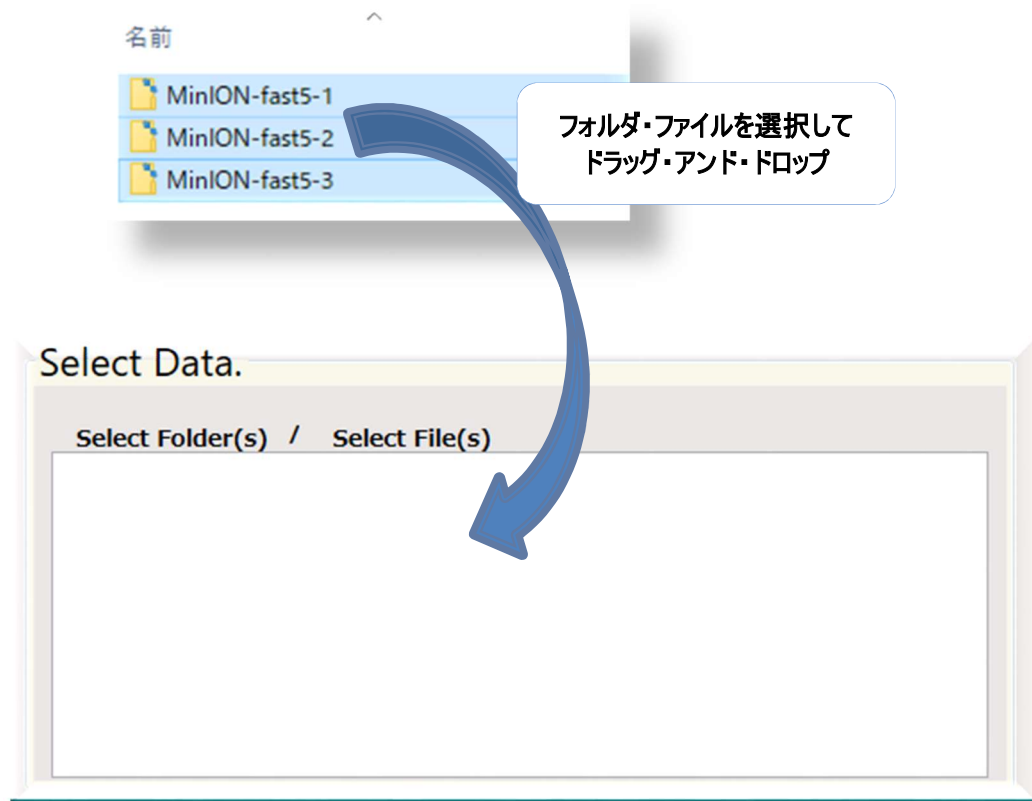
Fast5 から Fastq へ変換するコンバートツール、マッピングツールや、ホモロジー検索（Blast）実行時のパラメータを設定できます。（詳細は Parameters の項目を参照してください）



※ [edit] ボタンをクリックすることにより、マッピングツール（minimap2）や、ホモロジー検索（Blast）実行時のパラメータを設定できます。

詳細は [パラメータ設定] の項目を参照してください。

- ・ [Select Data]設定のテキストボックス内へ、[Fast5 フォルダ] [Fastq ファイル] をドラッグ・アンド・ドロップします

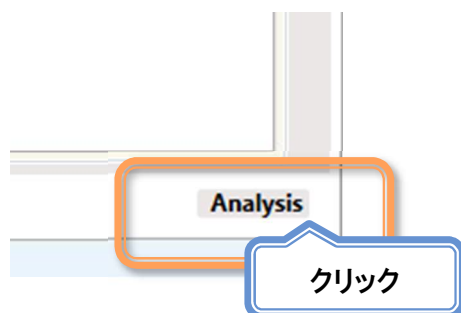


MinION シーケンサで指定できるデータは以下の内容です

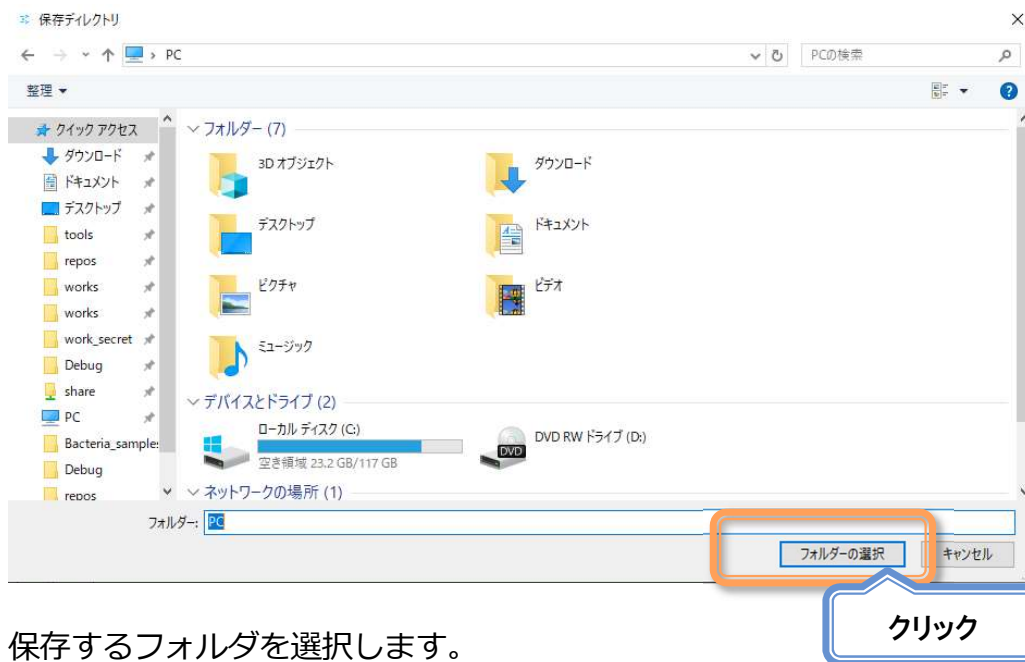
- ※ Fast5 ファイル (※ 1)
- ※ Fast5 が保存されているフォルダ
- ※ 付属のコンバートツールで Fastq へ変換後の Fastq ファイル

※1 Nanopore 社が提供している変換・マッピングツール「Guppy」が必須です  
2020.01 時点 Guppy 3.4.4 で検証を行っています

- ・ [Analysis]ボタンをクリックし保存場所を指定します。



解析対象が Fastq ファイルの場合、解析ファイルの保存場所を指定します。



指定されたデータが MinION (fast5) データの場合、ベースコールとバーコード分割を行います。対象のデータに対応するコンフィグの選択と、サンプル名を指定します。

ポップアップウィンドウが表示されます

Select Config

dna\_r9.4.1\_450bps\_fast\_prom.cfg

Barcode -> sample name

TemplateExcel

Barcode	Name
barcode01	barcode01
barcode02	barcode02
barcode03	barcode03
barcode04	barcode04
barcode05	barcode05
barcode06	barcode06
barcode07	barcode07
barcode08	barcode08
barcode09	barcode09
barcode10	barcode10
barcode11	barcode11
barcode12	barcode12
barcode13	barcode13
barcode14	barcode14

Select Config

dna\_r9.4.1\_450bps\_fast\_prom.cfg

MinION (Fast5 形式) データを指定し[barcode separate]にチェックが入っていた場合、guppy\_barcode プログラムによりバーコード分割を行います。バーコードに対応するサンプル名[name]を指定します。

Barcode -> sample name

Barcode	Name
barcode01	barcode01
barcode02	barcode02
barcode03	barcode03
barcode04	barcode04
barcode05	barcode05
barcode06	barcode06
barcode07	barcode07
barcode08	barcode08
barcode09	barcode09
barcode10	barcode10
barcode11	barcode11

Barcode に対応するサンプル名を入力します

バーコードの名前はエクセルからも指定することができます。

Barcode -> sample name

Barcode	Name
barcode01	barcode01
barcode02	barcode02
barcode03	barcode03
barcode04	barcode04
barcode05	barcode05
barcode06	barcode06

TemplateExcel

クリック

ウインドウ右上にある、[TemplateExcel]ボタンを押下し、エクセルファイルの保存場所を指定します。テーブル表示されている内容と同じ表が取得できます。

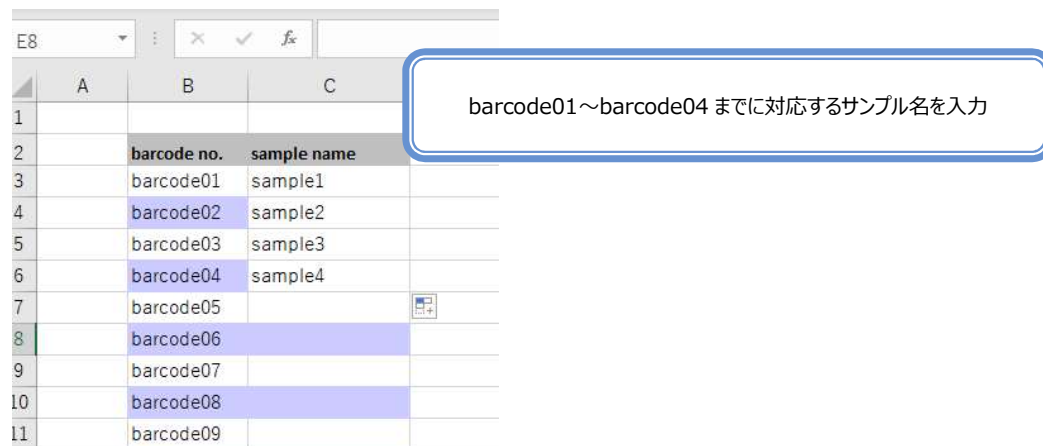
	A	B	C
1			
2		barcode no sample name	
3		barcode01	
4		barcode02	
5		barcode03	
6		barcode04	
7		barcode05	
8		barcode06	
9		barcode07	
10		barcode08	
11		barcode09	
12		barcode10	

ダウンロードしたエクセルファイルを編集します。C 列にサンプル名を入力して保存します。バーコード数はデフォルトで 24 件です。

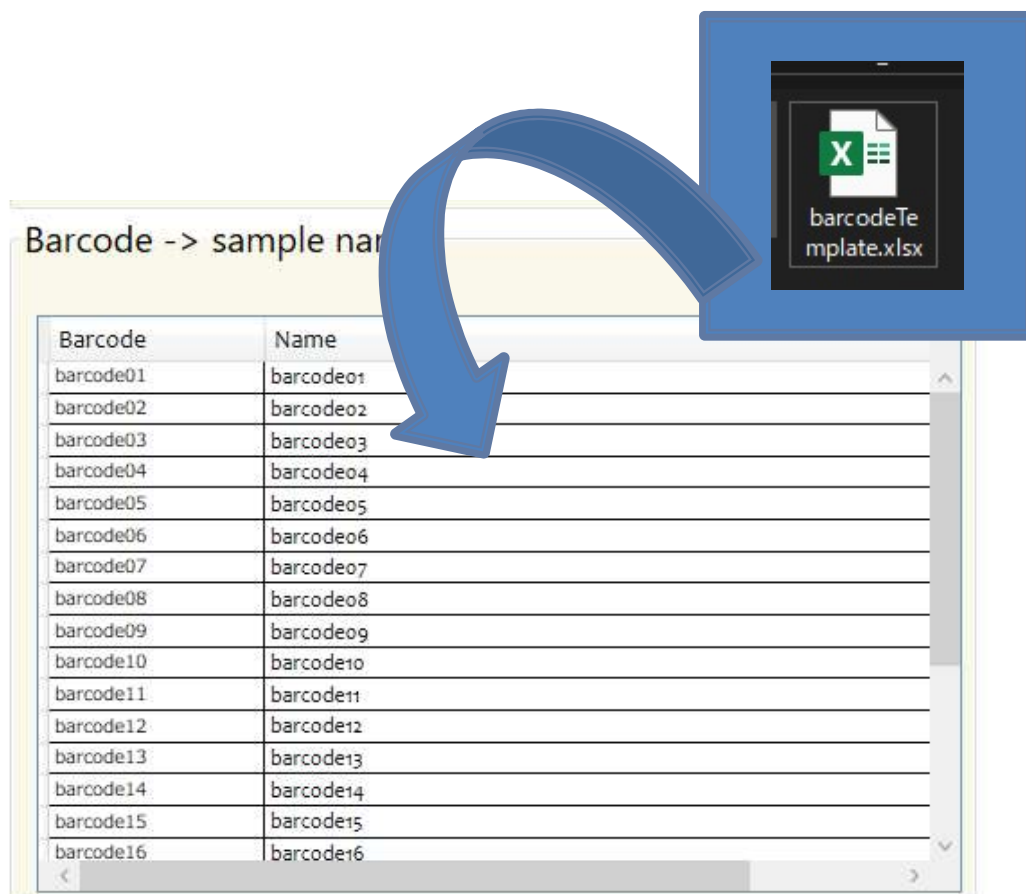
※B 列にはバーコードの一意名となっているので編集しないでください。



- ・エクセルファイルを、テーブル表示内にドラッグ&ドロップすると、内容が反映されます。



	A	B	C
1			
2		barcode no.	sample name
3		barcode01	sample1
4		barcode02	sample2
5		barcode03	sample3
6		barcode04	sample4
7		barcode05	
8		barcode06	
9		barcode07	
10		barcode08	
11		barcode09	



Barcode -> sample name

Barcode	Name
barcode01	barcode01
barcode02	barcode02
barcode03	barcode03
barcode04	barcode04
barcode05	barcode05
barcode06	barcode06
barcode07	barcode07
barcode08	barcode08
barcode09	barcode09
barcode10	barcode10
barcode11	barcode11
barcode12	barcode12
barcode13	barcode13
barcode14	barcode14
barcode15	barcode15
barcode16	barcode16

編集・保存したエクセルファイルを barcode テーブルへドラッグ&ドロップ

Barcode -> sample name

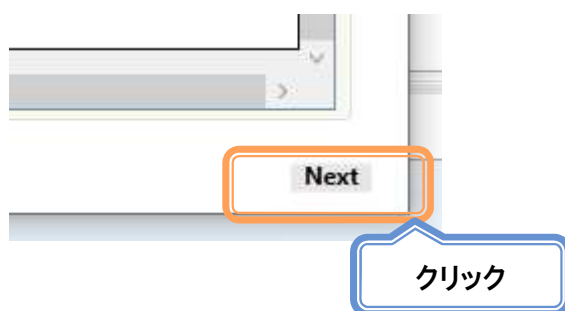
Barcode	Name
barcode01	sample1
barcode02	sample2
barcode03	sample3
barcode04	sample4
barcode05	barcode05
barcode06	barcode06
barcode07	barcode07
barcode08	barcode08
barcode09	barcode09

エクセルファイルに入力した「Name」だけ更新されます

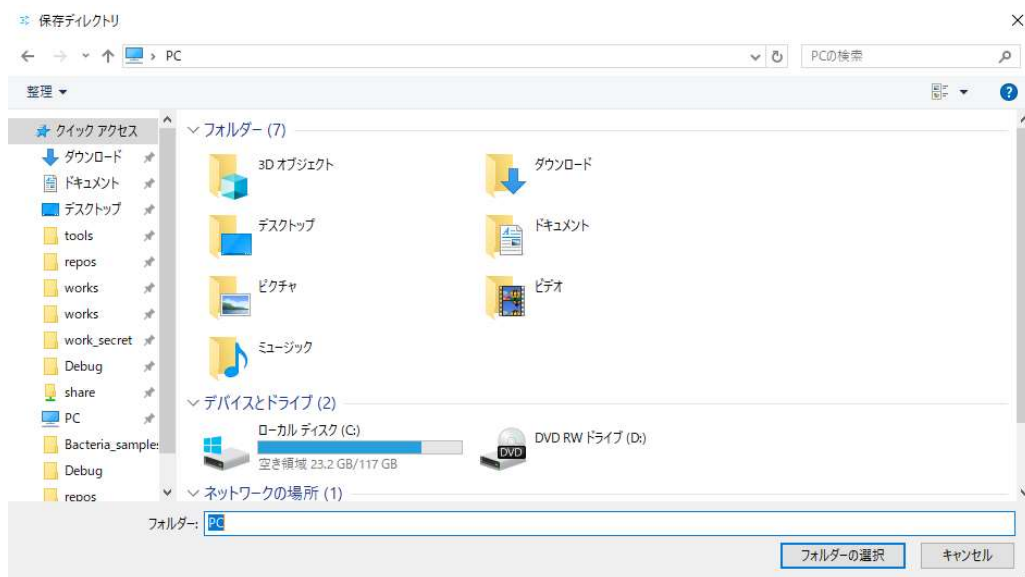
エクセルファイルに入力した[Name]がテーブル表示に反映されます。

※上記の例では「barcode01」～「barcode04」の[Name]を変更（sample1～sample4）して保存したエクセルファイルをドラッグ&ドロップした結果です。

barcode はデフォルトで 01～24 まで登録されています。25 以上のバーコードをご利用の場合は、エクセルファイル内で B 列の「barcode24」以下に、「barcode25」を増やし[Name]を指定してから保存したエクセルファイルをドラッグ&ドロップしてください。増加した分のバーコードが反映されます。



バーコードに対応するサンプル名を指定したのちに、[Next]ボタンを押下すると保存場所を指定するウィンドウが表示されます。



保存するフォルダを選択すると解析が実行されます。

実行中は下部ログメッセージに実行中のコマンドが表示されます。

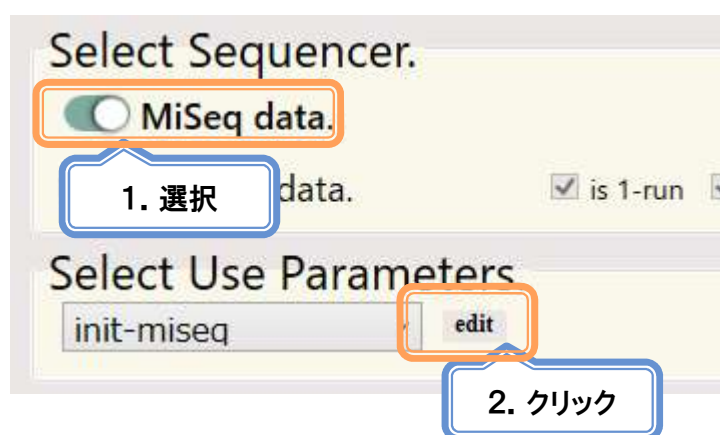
※ファイルの保存場所を指定する際に、日本語名（全角）・スペースを含むフォルダを指定しないでください。外部ツール実行時に正常動作しない場合があります。



## パラメータ設定を行う 【Illumina シーケンサ】

Illumina Illumina シーケンサから出力されたデータを解析する際に利用するツールの設定を行います。

[Select Sequencer] 項目で「Illumina data」を選択したのち、[Select Use Parameters] 項目で「edit」ボタンをクリックします。



## ・パラメータ設定 項目 【Illumina】

**Parameter name.**

MiSeq Parameter Name

**Commit**

☒ Analysis target is top3

**Mapping Parameter.**

☐ Bowtie2 Parameter edit?

mismatch penalties (--mp MX,MN)	3,2
penalty for positions (--np int)	4
read gap open (--rdg int1,int2)	4,1
reference gap open (--rfg int1,int2)	6,4
minimum alignment score (--score-min func)	L,10,1.3
disallow gaps within (--gbar int)	20
ambiguous characters (--n-ceil func)	L,0,0.3
consecutive seed extension (-D int)	15
re-seed reads (-R int)	2
mismatches to allowed (-N int)	0
seed substrings (-L int)	15
interval between seed (-i func)	L,15,10

### 【Parameter Name】（必須）

下部で設定したパラメータ項目にたいして名前を付けて保存します。

指定したパラメータを利用する場合は必須です。以前に同じパラメータ名がある場合は上書きされます。

### 【Analysis target is top3】

解析を行うさいにマッピング、Blast 結果の対象を上位 3 件にします（デフォルト）

チェックを外すとマッピング、Blast 結果の上位 1 件のみを対象とします

### 【Mapping Parameter】

Illumina データのマッピングを行う【bowtie2】実行時に指定するパラメータを設定します。

### 【Bowtie2 Parameter edit?】（非推奨）

Bowtie2 実行時に指定するパラメータを編集する場合にチェックボックスをオンにしてパラメータ値を変更します。（既存のパラメータセットは、A 型インフルエンザ・ウイルス Illumina シーケンスデータを前提としパラメータをセットしています）

詳細は以下のサイトにある内容を参照してください。

<http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>

### 【Sampling read count】

Fastq から 解析に利用するリード数を設定します。（Illumina シーケンサから出力されるリード数が過多の場合、解析に時間がかかるほか、利用 PC によってはメモリが足りない場合があります。このパラメータにより解析を行うリード数を制限することにより解析時間の短縮、PC のメモリ不足などに対応します）

【use FastQC 】

### 【FastQC】

Fastq から Phread score の高い塩基のみを取得する Quality control の設定を行います。

【use FastQC 】

Illumina データを利用する場合に、QC チェックを行う場合にチェックします。

【MinPhreadScore 】

QC に使用する Phread score のカットオフを設定します

【Window Size 】

QC 時に使用する Window 幅を塩基数で設定します

【Min Length】

解析に持ち込む Read の最小塩基長を指定します。QC での Read Trimming 後の Read 長がこの値以下の Read は破棄されます。

## ・パラメータ設定項目 【MinION】

**Parameter name.**

Minlon Parameter Name

**Commit**

☒ Analysis target is top3

**Convert tool Select.**

☒ Guppy

Program Path  **select**

**Mapping tool Select.**

☒ 1st-Mapping is selected reference

☐ GuppyAligner

Minimum coverage

☒ minimap2

<b>Indexing:</b>	k-mer size	<input type="text" value="15"/>
	minimizer window size	<input type="text" value="10"/>
<b>Alignment:</b>	matching score	<input type="text" value="2"/>
	mismatch penalty	<input type="text" value="4"/>
	gap open penalty	<input type="text" value="4,24"/>
	gap extension penalty	<input type="text" value="2,1"/>

### 【Parameter Name】（必須）

下部で設定したパラメータ項目にたいして名前を付けて保存します。

指定したパラメータを利用する場合は必須です。以前に同じパラメータ名がある場合は上書きされます。

### 【Analysis target is top3】

解析を行うさいにマッピング、Blast 結果の対象を上位 3 件にします（デフォルト）

チェックを外すとマッピング、Blast 結果の上位 1 件のみを対象とします

## 【Convert tool Select】

MinION シーケンサから出力された Fast5 形式のファイルを Fasta 形式へ変換させるツールを選択します。入力ファイルが Fastq/Fasta 形式の場合、この項目は無視されます。

現バージョンは Guppy のみ利用可能です（以前のバージョンで選択可能だったツールは、廃止となりました）。

### 【Guppy】

Fast5 から Fasta へ変換する際にナノポア社が提供している【Guppy】を利用します。

※【Guppy】を利用する場合は、事前にナノポア社のダウンロードページより Windows 版インストールファイルを保存し実行します。

[select]ボタンから Guppy の「bin」フォルダを指定してください。デフォルトのインストール先は **C:¥Program Files¥OxfordNanopore¥ont-guppy-cpu¥bin** となります。

### ・パラメータ設定 項目 【共通】

Create consensus.

CNS Min Coverage	<input type="text" value="3"/>
Homo Frequency	<input type="text" value="0.7"/>
Variant Min Frequency	<input type="text" value="0.35"/>
Ins/Del Min Frequency	<input type="text" value="0.51"/>
1st mapping Less than Min Coverage	<input type="radio"/> None <input checked="" type="radio"/> N
final mapping Less than Min Coverage	<input checked="" type="radio"/> None <input type="radio"/> N

Consensus include 'N' cutoff.

Ratio of 'N' include(%)

## 【CNS Create consensus】

コンセンサス配列 作成時のパラメータを設定します。

### 【CNS Min Coverage】

指定されたカバレッジ以下の領域は下記のパラメータ【1st mapping Less than Min Coverage】、【final mapping Less than Min Coverage】に従います



【Homo Frequency】

指定されたアレル頻度を超えた塩基がコンセンサス配列に反映されます

【Variant Min Frequency】

指定されたアレル頻度以上の変異をコンセンサス配列へ反映されます

【Ins/Del Min Frequency】

指定されたアレル頻度以上の変異（Ins / Del）がコンセンサス配列に反映されます

【1st Mapping Less than Min Coverage】

1st Mapping 結果から作成される一時コンセンサス配列に対して、上記で指定された【CNS Min Coverage】の閾値を満たさなかったコンセンサス塩基に空白文字（None）、もしくは N を指定します

【1st Mapping Less than Min Coverage】

最終コンセンサス配列作成の Mapping 結果から、上記で指定された【CNS Min Coverage】の閾値を満たさなかったコンセンサス塩基に空白文字（None）、もしくは N を指定します

【Consensus include 'N' cutoff】

コンセンサス配列 として採用する閾値を設定します。

【Ratio of Include （%）】

作成されたコンセンサス配列に、指定された数値（%）以上の N が含まれていた場合、当該コンセンサス配列を破棄します

The screenshot shows a web interface for configuring BLAST parameters. It is divided into two main sections: 'Blast result cutoff.' and 'Select consensus Blast target.'.

**Blast result cutoff.**

- Blast E-value:** A text input field containing '0.001'.
- Blast Max-target:** A text input field containing '5'.

**Select consensus Blast target.**

- Blast Top result Select.** A dropdown menu with 'Length' selected.
- Blast cutoff value Select.** A group of four radio button options, each with a corresponding text input field:
  - E-value:** Selected with a radio button, input field contains '0.001'.
  - Identical (%):** Unselected, input field contains '90.9'.
  - Score:** Unselected, input field contains '120'.
  - Length:** Unselected, input field contains '120'.

## 【Blast result cutoff】

BLAST 実行時に指定するパラメータを設定します。

### 【Blast E-value 】

Blast 結果から Subtype 決定に使用する Subject を取得する際の E-value の下限を設定します

### 【Blast Max-target 】

Subtype 決定に使用する Blast 結果の Subject 数を設定します

指定された数の Subject がすべて同じ Subtype を示している場合検体の Subtype が決定されます

## 【Select consensus Blast target】

コンセンサス配列を作成するための 2nd-mapping リファレンスを決定する BLAST 結果の基準を設定します。

### 【Blast Top result Select. 】

Top Hit を選択する Blast 結果の項目を選択します。デフォルトは Length です。

※ Length の場合 BLAST の閾値を超えた結果の中から塩基配列数が最大のエントリーとなります

### 【Blast cutoff value Select.】

Blast 結果の閾値を指定します。

(上述の Blast Top result Select.項目と一緒に利用します)



## 【Cancel】 【Commit】 ボタン

パラメータ設定画面の値を保存するかを決定します。

### 【Cancel】

上記の設定項目・数値を破棄して、解析画面へ戻ります

### 【Commit】

上記の設定項目・数値を保存し、解析画面へ戻ります



## 解析結果を参照する 【Sample Result】

解析結果を参照します。

画面項目（左側メニュー） [Result] をクリックすると解析結果画面へ遷移します。



## ・解析結果画面 項目

**Analyses**

**Result**

**【Results】**  
解析結果を表示します。  
リスト選択してからDetailを表示します。

Date	Name
2021-01-19	rep-6042_S67_L001
2021-01-19	6042_S67_L001
2021-01-19	SRR1324315
2021-01-19	6042_S67_L001
2021-01-19	SRR1324315
2021-01-19	sample.fastq
2021-01-20	MI-MDHHS-SC23042_c
2021-01-20	rep-SRR12109250-4_na

**Sample Result.**

Name: **WF-sample-20210108** commit

ExecDate: 2021-01-20 09:27:15

**Result Detail.**

Top1

Reference: **MW277478**

"Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/AUS/VIC16104/2020, complete genome"

Spices: **Severe acute respiratory syndrome-related coronavirus**

Host: **Homo sapiens**

SequenceType: **GenBank**

GeoLocation: **Australia: Victoria**

Collection Date: **2020-08-02** Release Date: **2020-11-20T00:00:00Z**

GisaidClade: **GR:C241T,C3037T,A23403G,G28882A includes S-D614G + N-G204R**

NextstrainClade: **20I/501Y.V1:derived from 20B bearing S 501Y, S 570D, S 681H, ORF8 27\*, concentrated in the United Kingdom**

Created Consensus nucleotide sequence:

Reference Length: **29809** Ratio: **0.991** CoverAve: **5161.037**

ACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAATCTGTGTGGCTGCTCACTCGGCTGCATGCTTAGTGCACTCACGCAI  
ATATTCACTGGATGGTTATGTTACACCTTTAGTACCTTTCTGGATAACAATTGCTTATATCATTTGTATTTCCACAAAGCATTTCTATT  
TAGAGTGCTATCAACCTTAACCTTGCCTGGTTGTGATGGTGGCAGCTTTAGCTTGTGGGTTTACAAACAATTTGATACTTATAACCTC  
CCAAGGAAATTTGGGGACCAGGAACATAACAGACAAGGAAGTATTACAAACATTGGCCGCAATTCACAAATTTGCCCCAGCG

version 0.9.2 True

VCF-nucleotide VCF-aminoacid map view ExportResult Alignment Tree

## 【解析結果 リスト】

解析を行った日付（Date）と 解析実行時に指定したフォルダまたはファイルの代表名がリストされます。

リスト上部の [Date] [Name] をクリックすると、リスト部分がクリック対象の昇順・降順となります。デフォルトは日付（降順）です。



#### 【Hide/Delete】

解析結果リストに表示されるサンプルに対して一時的に非表示、または削除が可能です。解析結果リストの当該サンプル名を右クリックするメニューを表示させ、[Hide] / [Delete] を選択します。

- ・ [Hide] 一時的に解析結果リストより当該サンプルを非表示にします
- ・ [Delete] 管理しているサンプルの情報を削除します（削除されたデータは戻りません）

### 【Sample Result】

解析結果リスト行を選択（クリック）すると、選択された解析結果が表示されます。

#### 【Name】

解析実行時に入力したフォルダ/ファイル名をもとに解析結果の名前が付けられます。

[name] テキストボックスに任意の文字列を入力後、テキストボックス右にある [commit] ボタンをクリックすることによりデフォルトのフォルダ・ファイル名ではなく、ユーザが分かりやすい名前を付与する事ができます。

[commit] ボタンにより更新が正常に行われると、[解析結果リスト] にも反映されます。

#### 【ExecDate】

解析を行った日時の詳細を表示します。

### 【Sample Detail】

解析結果 塩基配列に関する情報を表示します。

リスト行を選択（クリック）すると、選択された解析結果が表示されます。

表示される内容は以下の URL にて公開されている NCBI 登録情報です。

[https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType\\_s=Nucleotide&VirusLineage\\_ss=SARS-CoV-2,%20taxid:2697049](https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=SARS-CoV-2,%20taxid:2697049)

#### 【Top1 ～ Top3】タブ

解析結果のスコア上位の Top1～Top3 をタブ別に表示します。スコアはコンセンサス配列作成時のマッピング結果より算出されます。マッピング時のリファレンス配列に対するマッピング領域（Ratio）、マッピング平均アライメント（Cover average）の良い順になります。

コンセンサス配列作成時に設定しているパラメータにより、Top2 以降が閾値（各種 cutoff 値）を満たさない場合はタブが作成されません。

※Top1 が閾値を満たさない場合は、エラーとしてログに表示され Result リストに表示されません。

## 【Reference】

最終コンセンサス配列作成時に登録されている NSBI コロナウイルスから最も類似した NCBI アクセッションが表示されます。下部に当該配列のディスクリプションが表示されます。

※最終コンセンサス配列は当該配列のみに全てのリードがマッピングされています。

## 【Spices】

当該アクセッションに登録されている種が表示されます。（CovGAS では Homo Sapiens が主になります）

## 【Sequence Type】

当該アクセッションに登録されている Sequence Type（Genbank、Refseq など）が表示されます。

## 【Geo Location】

当該アクセッションに登録されているサンプルが取得された地域情報が表示されます。

## 【Collection Date】

当該アクセッションに登録されているサンプルが取得された年月日が表示されます。

## 【Release Date】

当該アクセッションに登録されているサンプルが登録された日時が表示されます。

## 【Gisaid Clade】

以下の GISAID サイトで定義されている Clade 情報を表示します。

当該箇所にある変異を満たす場合に情報が表示されます。

<https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/>

## 【Nextstrain Clade】

以下の Nextstrain サイトで定義されている Clade 情報を表示します。

当該箇所にある変異を満たす場合に情報が表示されます。

<https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming>

また情報は以下の URL で更新されています。

<https://raw.githubusercontent.com/nextstrain/ncov/master/defaults/clades.tsv>

## 【Reference Length】

当該アクセッションの塩基配列数が表示されます。

## 【Ratio】

最終コンセンサスと当該アクセッションのマッピング結果のリファレンスカバー率が表示されます。

## 【Cover Ave.】

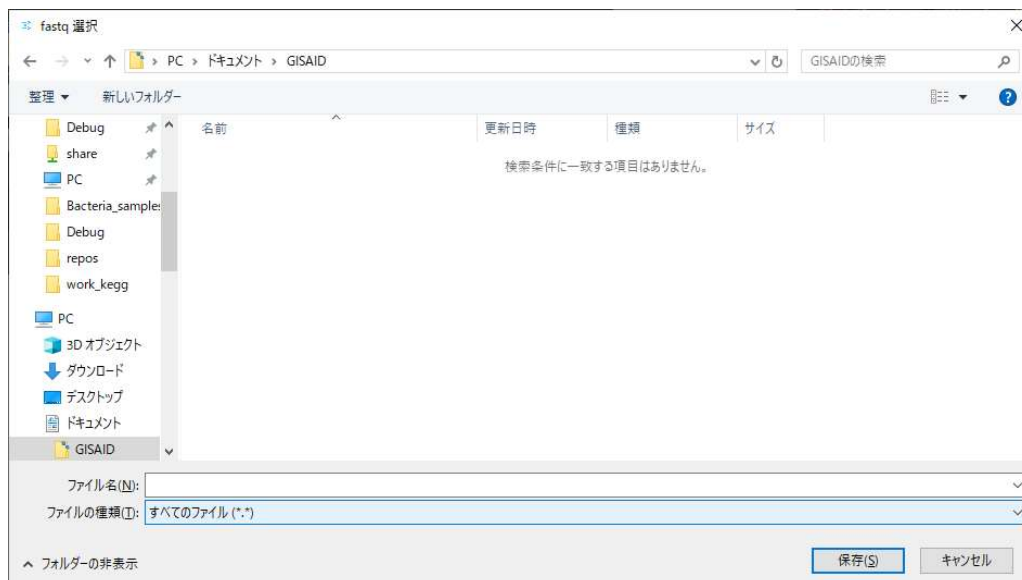
最終コンセンサスと当該アクセッションのマッピング結果の平均カバレッジが表示されます。

## 【塩基配列】

最終コンセンサスの塩基配列が表示されます。

## ・結果の外部出力

解析結果を表示している右下部にあるボタンをクリックすると、保存場所を指定するウィンドウが表示されます。それぞれ出力ファイルが作成され、保存されます。



## 【Export Result】

Result 画面に表示されている情報が hta 形式で保存されます。hta ファイルはダブルクリックするとブラウザベースの Windows 標準アプリケーションです。

## 【Alignment】

Result 画面に表示されているリファレンス塩基配列と最終コンセンサス配列のアライメントが .aln 形式で保存されます。.aln ファイルは AliView (<https://ormbunkar.se/aliview/>) が起動しリファレンス塩基配列の変異個所を確認することが可能です。

別途、解析結果のディレクトリが削除されていない場合、Result 画面で表示されているリファレンスに対するマッピング結果を Tablet (<https://ics.hutton.ac.uk/tablet/>) が起動し確認することが可能です。

## 【Tree】



Result 画面に表示されているリファレンスのアクセッションを含んだ系統樹作成を行います。系統樹作成に用いられるアクセッションは以下の通りです。

系統樹作成は Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) を用いて作成され、njplot (<http://doua.prabi.fr/software/njplot>) によって表示されます。

(系統樹作成は Clustal Omega が長時間に渡り実行されます)

※このアクセッションは初期マッピングに用いられる CovGAS で選択している標準株となります。

>NC\_004718.3 SARS coronavirus Tor2, complete genome

>NC\_006577.2 Human coronavirus HKU1, complete genome

>NC\_019843.3 Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome

>NC\_038294.1 Betacoronavirus England 1 isolate H123990006, complete genome

>NC\_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

## 【VCF-nucleotide】

Result 画面に表示されているリファレンスに対してマッピングした結果を基に変異検出結果を表示します。(実行されている PC にエクセルがインストールされている場合エクセルが起動します)

## 【VCF-aminoacid】

Result 画面に表示されているリファレンスに対してマッピングした結果から変異検出結果を基にアミノ酸変異を表示します。(実行されている PC にエクセルがインストールされている場合エクセルが起動します)

## 【VCF-aminoacid】

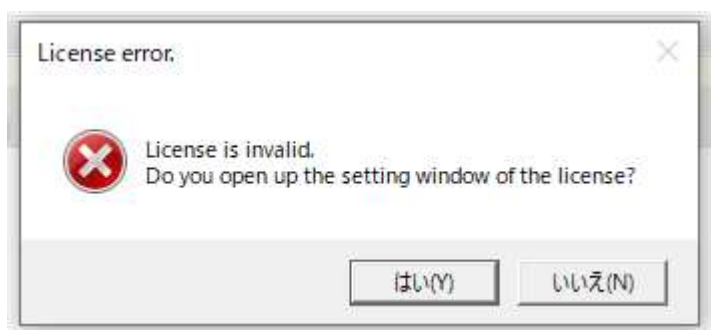
Result 画面に表示されているリファレンスに対してマッピングした結果のマッピング状況を表示します。マッピング結果は外部ツールの Tablet (<https://ics.hutton.ac.uk/tablet/>) が起動されます。(初回起動時に時間を要します。また左ペインにあるコンセンサス名をクリックする必要があります)



## ライセンスの登録【License】

CovGAS はインストール後 2 週間をデモ期間として利用いただけます。

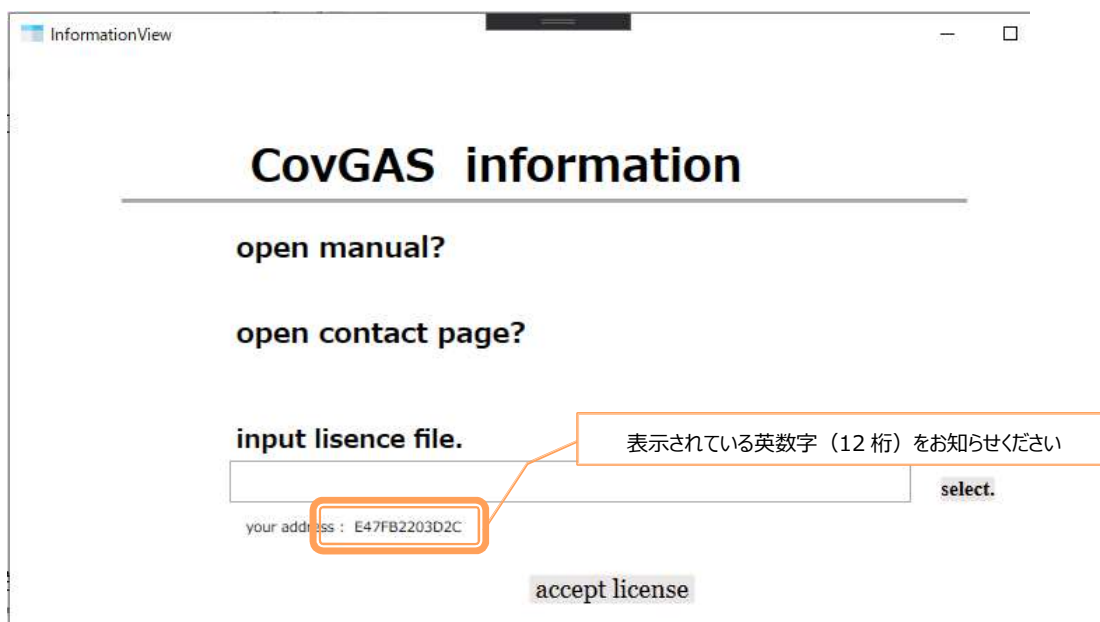
デモ期間を過ぎると、ライセンスエラーとなり起動時にダイアログが表示されます



〔ライセンス エラーダイアログ〕

「いいえ」を選択すると、アプリケーションが終了します。

「はい」を選択すると、インフォメーションウィンドウが表示されます。



「ライセンスファイル」を取得済みの場合は、[select] ボタンからライセンスファイルを選択し、[accept license] ボタンをクリックします。

「ライセンスファイル」を申請する際は、「your address 」部分に書いてある文字列を弊社にお送りください。  
[open contact page] から弊社のお問い合わせ画面、もしくは以下のメールアドレスよりご連絡ください。

URL : <https://www.w-fusion.co.jp/J/contactus.php>

メールアドレス : techsupport@w-fusion.co.jp



## 【Appendx】

## コンセンサス配列へ反映される SNP と挿入・欠失のルール

## コンセンサス配列へ反映される SNP と挿入・欠失のルール

CNS MIN COVERAGE で設定したカバレッジを越えた領域からコンセンサス配列が出力されます。この領域にある変異をコンセンサス配列に反映するルールは下記です。CNS MIN Variant Frequency で設定された頻度を越えた変異がコンセンサスに反映されます。

CNS MIN Variant Frequency 以上で CNS MIN Homozygous Frequency 以下の SNP が複数ある場合は IUPAC 表記のルールにのつった混合塩基がコンセンサスに反映されます。

SNP がひとつしかなく、頻度が CNS MIN Variant Frequency 以上で CNS MIN Homozygous Frequency 以下の場合はマッピングリード中の参照配列の塩基が CNS MIN Variant Frequency 以上で CNS MIN Homozygous Frequency 以下であれば参照配列塩基と SNP の IUPAC 表記となります。

複数の SNP のうちひとつでも CNS MIN Homozygous Frequency を越えた SNP があれば、その SNP の塩基がコンセンサスに反映されます。

複数の CNS MIN Homozygous Frequency を越えた SNP が存在した場合は頻度のもっとも高い変異が反映されます。

挿入・欠失がコンセンサス配列に反映されるには必ず CNS MIN Homozygous Frequency を超えていることが条件です。

CNS MIN Variant Frequency を超えただけでは挿入・欠失はコンセンサス配列に反映されません。

## バーコードとサンプル名

入力データの個数と、[is 1-run]、[barcode separate]のチェックボックスによって、バーコードに対応する表示名が変化します。

・[is 1-run] チェックボックスが **ON**、[barcode separate] チェックボックスが **ON** の場合

複数データ (fast5 ファイル、fast5 を含むフォルダ) を指定した際、ベースコール後の Fastq ファイルをマージしてバーコード分割を行うため入力されたファイルは 1fastq として扱われます。

このため、バーコード分割を行った場合でもサンプル数はバーコード分割後の数と同じになるため「barcode01」～ が一意名になります。

「is 1-run」 チェックボックスが **OFF**、[barcode separate] チェックが **ON** の場合

単一のデータを指定した場合、上記同様にベースコール後の入力データが 1fastq となるため、「barcode01」～ が一意名になります。

複数のデータが指定された場合、各データに対してバーコード分割が行われるため、各データのフルパス+Barcode 番号になります。

### name list 個数

uniq	1
uniqs	n
single	barcode24
multi	n x barcode24

is-1run	barcode	n=1	request
TRUE	TRUE	TRUE	single
TRUE	TRUE	FALSE	single
TRUE	FALSE	TRUE	uniq
TRUE	FALSE	FALSE	uniq
FALSE	TRUE	TRUE	single
FALSE	TRUE	FALSE	multi
FALSE	FALSE	TRUE	uniq
FALSE	FALSE	FALSE	unqs

## 塩基変異結果とアミノ酸変異結果に関して

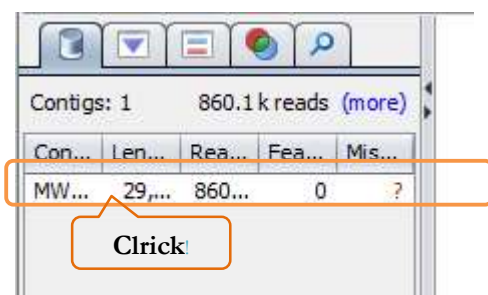
塩基変異結果は、1st-mapping 結果の BAM ファイルとリファレンスにより、bcftools を利用して vcf 形式にて出力されます。出力された塩基変異結果のリファレンスゲノム位置を基にアミノ酸変異を検出します。ゲノム位置とアミノ酸翻訳領域はマッピングリファレンスのアクセッション番号を基に NCBI Genbank 情報より取得しています。

## Alignment Viewer に関して

結果画面にあるボタンを押下することにより 2nd Blast 結果（もっとも近い NCBI アクセッション）とコンセンサス配列のアライメント情報を作成し保存します。保存が正常に行われると AliView (<https://ormbunkar.se/aliview/>) が起動します。AliView の操作は起動したツールのヘルプを参照してください。

## Mapping Viewer に関して

結果画面にあるボタンを押下することにより Tablet (<https://ics.hutton.ac.uk/tablet/>) が起動します。（起動時に解析時に結果出力したディレクトリから削除されている場合は起動しません）Tablet は最終解析結果のリファレンスとマッピング結果（BAM ファイル）を読み込んだ状態で起動します。起動後に左側にある配列名をクリックすることによりマッピング情報が表示されます。Tablet の操作は起動したツールのヘルプを参照してください。



## コンセンサス配列の作成に関して

CovGAS は以下の要項によってコンセンサス配列を作成します。

### Step1 : 1st-Mapping

Sample 配列 x リファレンス配列

### Step2 : mapping -> Pileup コマンド -> コンセンサス配列作成

マッピング結果 sam ファイルから、SNP/insert/delete を反映させた 1st コンセンサス配列作成

### Step3 : 1st-Blast

1st コンセンサス配列 x リファレンス配列

BLAST 結果から上位 3 件を 次項 2nd-Mapping リファレンス配列とする

### Step4 : 2ndMapping

Sample 配列 x 1st-Blast リファレンス配列

【mapping -> Pileup コマンド -> コンセンサス配列】

マッピング結果 sam ファイルから、SNP/insert/delete を反映させた 2nd コンセンサス配列作成

### Step5 : 2nd Blast

2nd コンセンサス配列 x リファレンス配列

解析結果コンセンサス配列に最も類似した NCBI 登録配列を策定

※ このワークフローは FluGAS（特許）と同様になります。

内容についてのお問い合わせ先：

株式会社ワールドフュージョン技術営業部

東京都中央区日本橋蛸殻町 1 丁目 38 番 1 2 号

油商会館ビル 2F

Tel : 03-3662-0521 Fax : 03-3662-0522

Mail : [techsupport@w-fusion.co.jp](mailto:techsupport@w-fusion.co.jp)

URL : <https://www.w-fusion.co.jp>