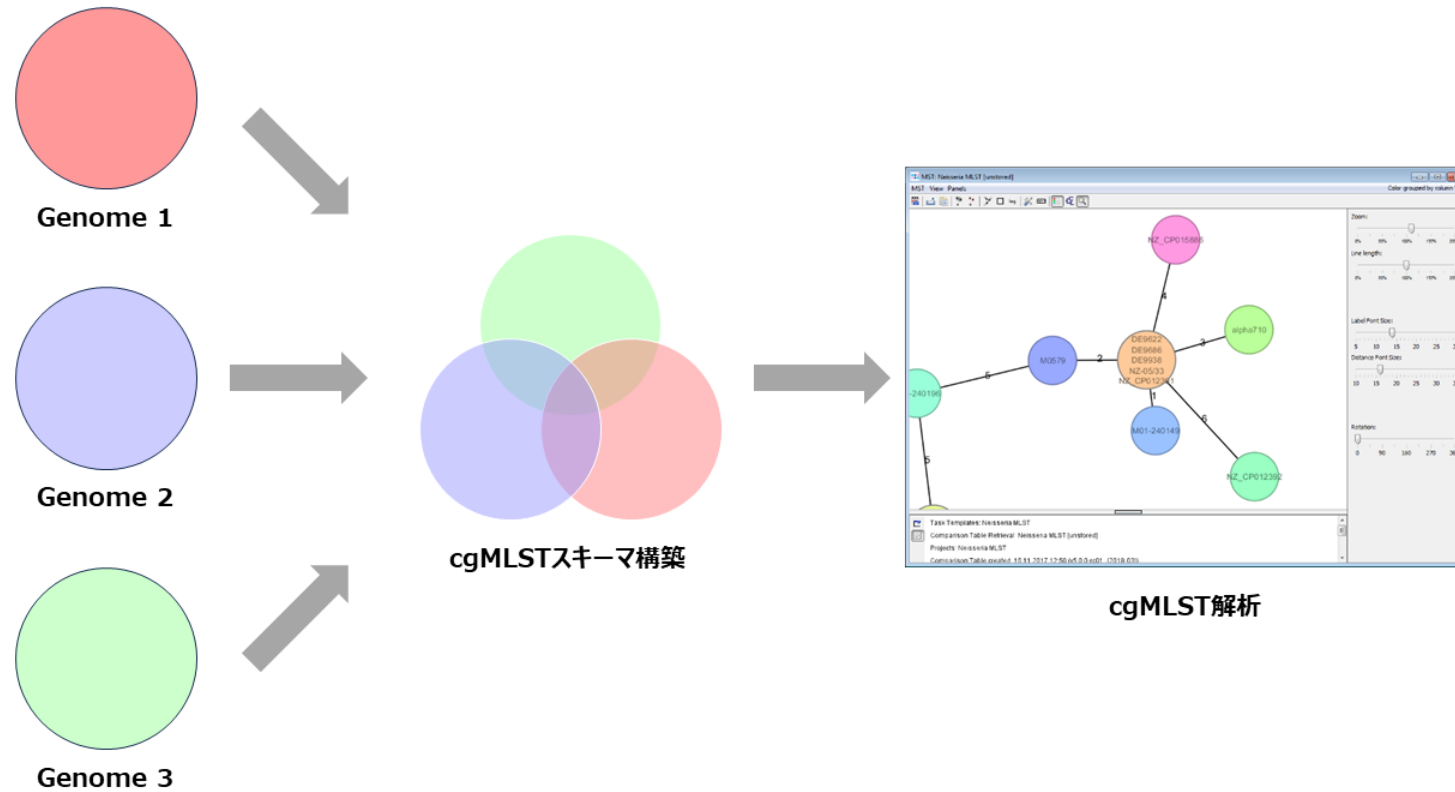
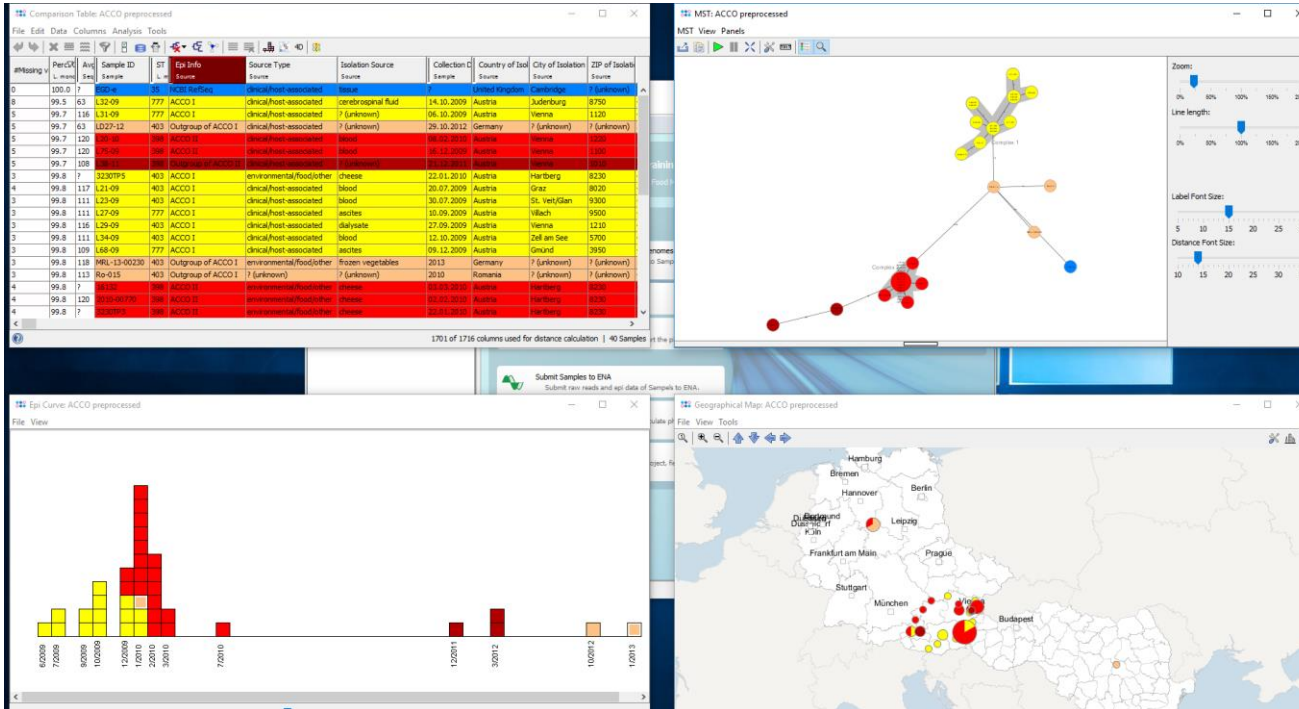


コアゲノムMLSTスキーマの構築

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



- 細菌のコアゲノムMLST (cgMLST) 解析を行うには、菌種ごとのコアゲノム情報のスキーマが必要
- Ridom SeqSphere+では、外部データベースやシーケンスにより得られたゲノムデータを用いて、コアゲノムMLSTスキーマの構築が可能



■ 菌種ごとの遺伝子配列やアレル情報データベースのダウンロード

- MLST / コアゲノムMLST (cgMLST) スキーマ
- 薬剤耐性因子 / 病原因子
- Serotype / Spatype

■ NGSデータ解析パイプラインによる遺伝子型タイピング

- クオリティチェック
- リードアセンブル / マッピング

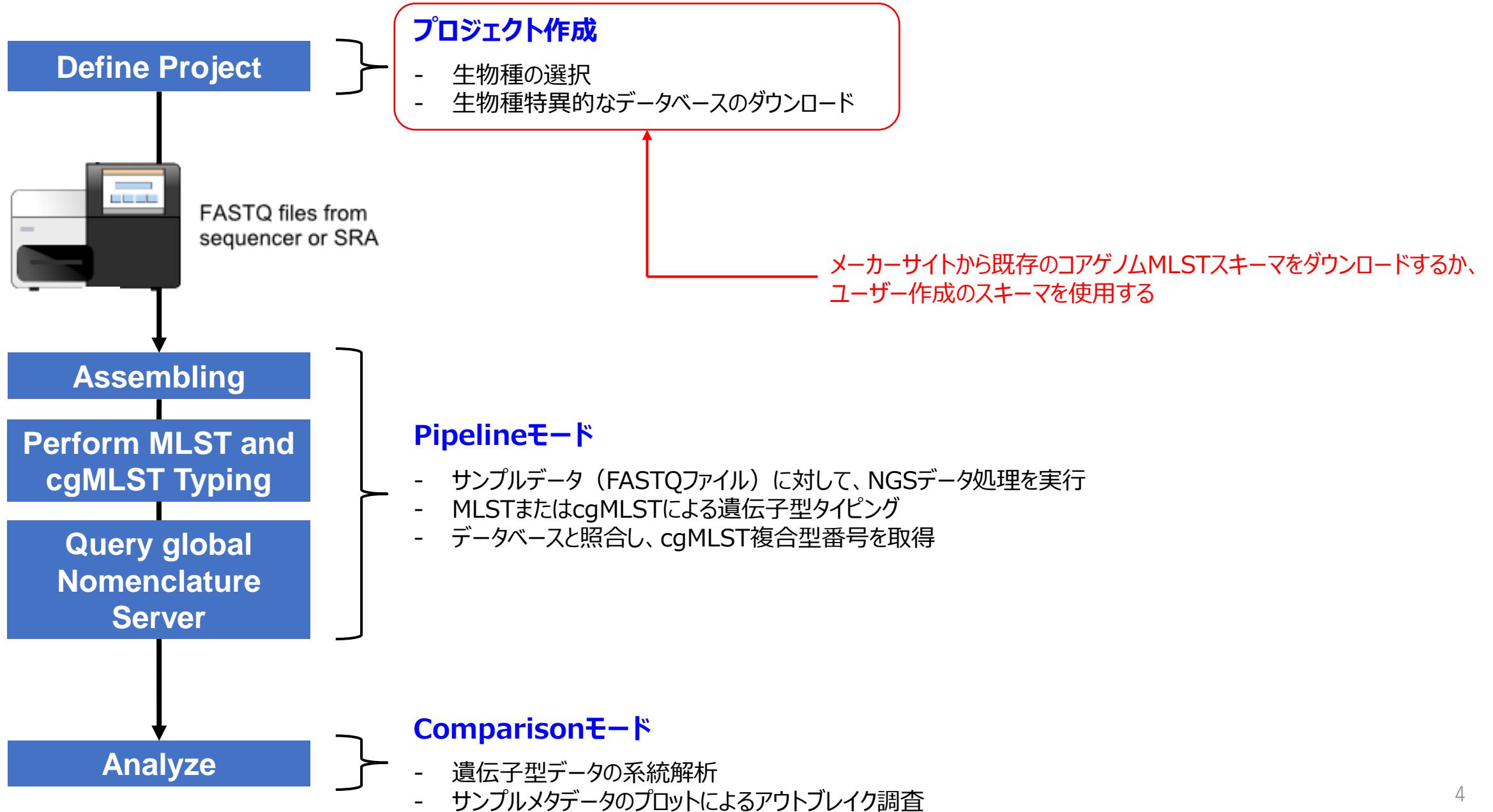
■ サンプルごとのcgMLST複合型の照合

■ 系統解析

- 系統樹 (Neighbor Joining Tree)
- ネットワークグラフ (Minimum Spanning Tree)

■ アウトブレイク調査

- Epi Curve
- Geographical Map

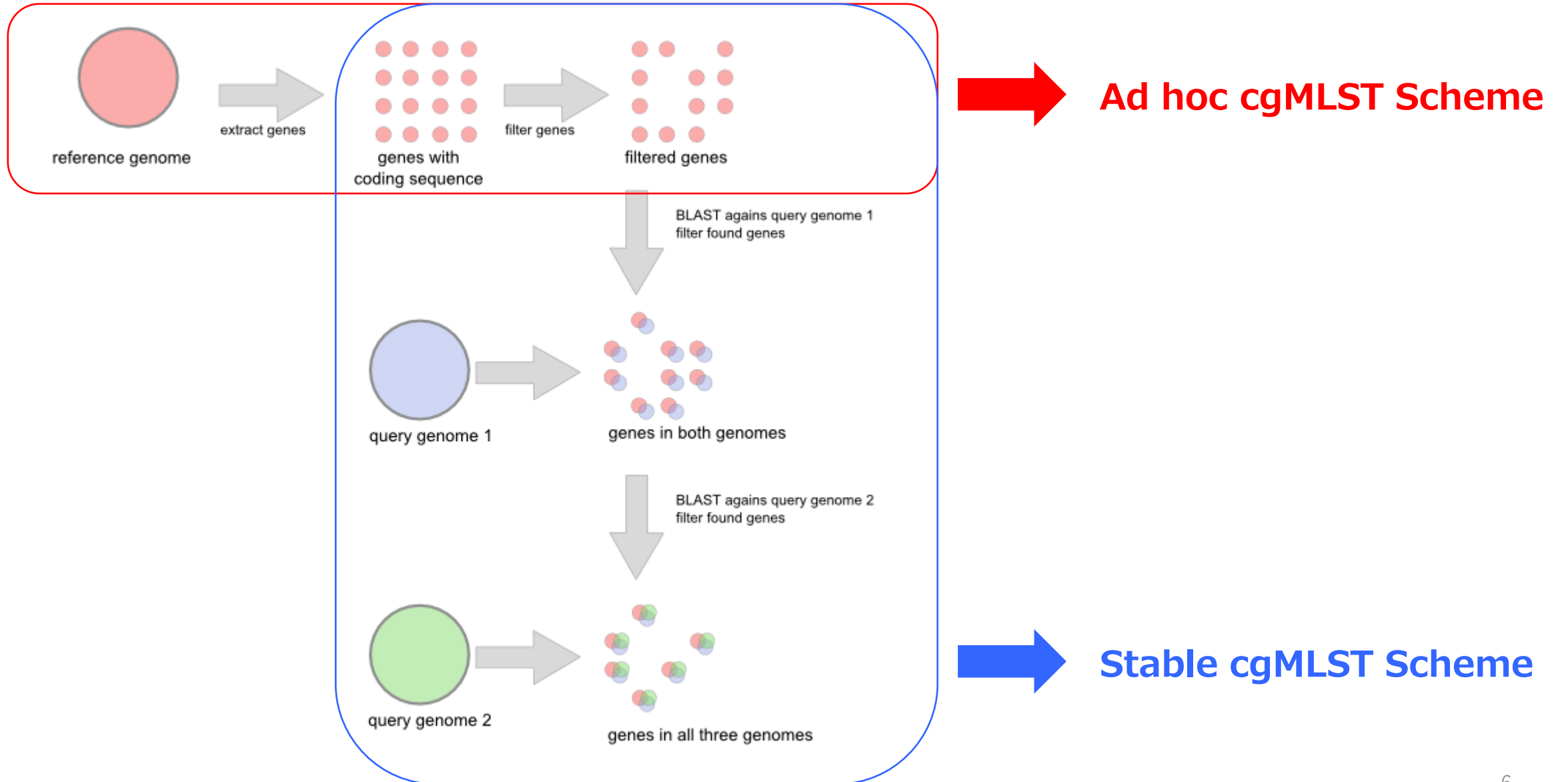


Ad hoc cgMLST Scheme

- ✓ 1種類のシードゲノム (Seed Genome) のみを用いて、コアゲノムMLSTスキーマを作成
- ✓ シードゲノムは、配列データと遺伝子アノテーションデータの両方が必要
- ✓ 相同遺伝子や配列長の短い遺伝子の除去、スタート/ストップコドンの有無の判別などの基礎的なフィルターのみを利用して、コアゲノムの決定を行う

Stable cgMLST Scheme

- ✓ シードゲノムと複数のクエリーゲノム (Query Genome) を用いて、コアゲノムMLSTスキーマを作成
- ✓ クエリーゲノムは、配列データのみの場合も利用可能
- ✓ シードゲノムとすべてのクエリーゲノムで共通に存在する遺伝子のみを抽出することで、菌種内で保存されている遺伝子のみをコアゲノムとして使用することができる



- Ridom SeqSphere+ で使用できるコアゲノムMLSTスキーマ作成用ツール
- シードゲノムやクエリーゲノム、プラスミドなどの除外する遺伝子を指定すると、コアゲノムのターゲット遺伝子を自動で決定し、その情報に基づいてコアゲノムMLSTスキーマの作成を行う
- 各ゲノムや遺伝子データは、ファイル（GenBank, FASTAなど）を選択するか、NCBI Accession番号による指定が可能
- NCBIに登録されているゲノムデータの検索機能もあり、菌種名を入力することで、すべての登録されている株のゲノムを確認し、スキーマ作成に利用することが可能

The screenshot shows the 'cgMLST Target Definer' application window. It features a 'Settings' tab with three sub-tabs: 'Filter for seed genome', 'Filter for penetration query genomes', and 'Manually exclude targets'. The 'Seed Genome' field is empty, with 'Open File' and 'Get from NCBI' buttons. Below, the 'Genomes used for testing of 100% gene penetration ('penetration query genomes')' section has an empty list box and 'Add from File', 'Add from NCBI', and 'Remove' buttons. The 'Exclude Genes with BLAST matches > 90% and > 100bp length in the following sequences:' section also has an empty list box and similar buttons. The 'Thresholds' section is active, showing 'Blast options' with 'Required gene identity found in penetration query genome ≥ 90 %' and 'Required gene overlap found in penetration query genome ≥ 100 %'. At the bottom, there is a 'Find taxonomic and quality outliers' button and 'Start' and 'Cancel' buttons.

cgMLST Target Definer

Settings Filter for seed genome Filter for penetration query genomes Manually exclude targets

Seed Genome: **NC_008767.1 (12-MAR-2023), 2194961 bases, 2111 genes (Neisseria meningitidis)** Open File Get from NCBI

Genomes used for testing of 100% gene penetration ('penetration query genomes'):

- NC_008112.2 (04-DEC-2023), 2272860 bases, 2178 genes with CDS (Neisseria meningitidis MC58, com
- NC_008116.1 (24-DEC-2022), 2184406 bases, 2121 genes with CDS (Neisseria meningitidis Z2491, cor
- NC_010120.1 (04-DEC-2023), 2153416 bases, 2073 genes with CDS (Neisseria meningitidis 053442, co
- NC_013016.1 (30-DEC-2022), 2145295 bases, 2049 genes with CDS (Neisseria meningitidis alpha14, co
- NC_017501.1 (24-DEC-2022), 2277550 bases, 2195 genes with CDS (Neisseria meningitidis 8013, comp
- NC_017505.1 (22-FEB-2023), 2242947 bases, 2173 genes with CDS (Neisseria meningitidis alpha710, c
- NC_017512.1 (30-DEC-2022), 2227255 bases, 2171 genes with CDS (Neisseria meningitidis WUE 2594,
- NC_017513.1 (22-FEB-2023), 2184862 bases, 2098 genes with CDS (Neisseria meningitidis G2136, com
- NC_017514.1 (12-FEB-2023), 2223518 bases, 2150 genes with CDS (Neisseria meningitidis M01-24014
- NC_017515.1 (22-FEB-2023), 2250449 bases, 2158 genes with CDS (Neisseria meningitidis M04-24019
- NC_017516.1 (06-SEP-2023), 2240883 bases, 2151 genes with CDS (Neisseria meningitidis H44/76, co
- NC_017517.1 (22-FEB-2023), 2287777 bases, 2206 genes with CDS (Neisseria meningitidis M01-24035
- NC_017518.1 (22-FEB-2023), 2248966 bases, 2190 genes with CDS (Neisseria meningitidis NZ-05/33,

Exclude Genes with BLAST matches > 90% and > 100bp length in the following sequences:

- NZ_CM000956.1 (12-FEB-2023), 7242 bases, 11 genes with CDS (Neisseria meningitidis K1207 plasmid
- NZ_CM000958.1 (12-FEB-2023), 7242 bases, 11 genes with CDS (Neisseria meningitidis S0108 plasmid

Thresholds Blast options

Required gene identity found in penetration query genome ≥ 90 %

Required gene overlap found in penetration query genome ≥ 100 %

Find taxonomic and quality outliers

Start Cancel

cgMLST Target Definer Results

ターゲット遺伝子リスト

1295 of 2111 targets were defined for cgMLST

Export Results Create Task Templates

Target	Source	Query Result NC_008112.2 (04-DEC-2023)	Query Result NC_008116.1 (24-DEC-2022)	Query Result NC
NMC_RS00010	NC_008767.1 (12-MAR-2023)	NMB_RS00095	NMA_RS01380	NMCC_RS00105
NMC_RS00025	NC_008767.1 (12-MAR-2023)	NMB_RS00150	NMA_RS01435	NMCC_RS00140
NMC_RS00035	NC_008767.1 (12-MAR-2023)	NMB_RS00160	NMA_RS01450	NMCC_RS00150
NMC_RS00040	NC_008767.1 (12-MAR-2023)	NMB_RS00165	NMA_RS01455	NMCC_RS00155
NMC_RS00045	NC_008767.1 (12-MAR-2023)	NMB_RS00170	NMA_RS01460	NMCC_RS00160
NMC_RS00055	NC_008767.1 (12-MAR-2023)	NMB_RS00175	NMA_RS01470	NMCC_RS00170
NMC_RS00060	NC_008767.1 (12-MAR-2023)	NMB_RS00185	NMA_RS01475	NMCC_RS00175
NMC_RS00065	NC_008767.1 (12-MAR-2023)	NMB_RS00190	NMA_RS01480	NMCC_RS00180
NMC_RS00070	NC_008767.1 (12-MAR-2023)	NMB_RS00195	NMA_RS01485	NMCC_RS00185

1,295 rows

cgMLST Target Definer Results

Date: 2023/12/26 16:36
cgMLST Target Definer version: 1.5 (win)
Ridom SeqSpher+ version: 9.0.10
User: Analyzer
Server: Localhost (SeqSpher+ Server on OZAWA)

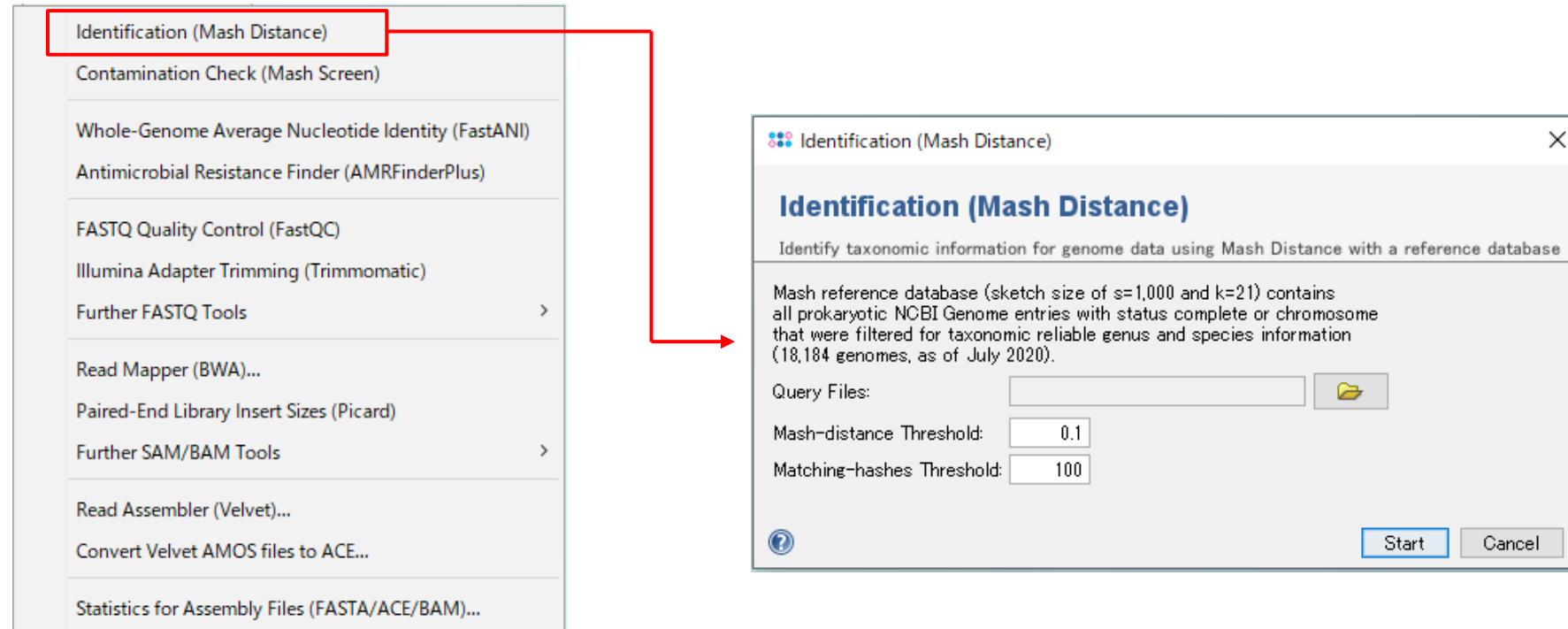
Resulting Targets:
1295 targets were defined for cgMLST (1221828 bases)
570 targets were used as Accessory targets (413877 bases)
246 targets were discarded

決定されたコアゲノムとアクセサリーゲノムのターゲット遺伝子数

Close

■ スキーマ作成に用いるゲノムを指定

- 任意のシードゲノムやクエリーゲノムを指定してツールを実行すると、コアゲノムとアクセサリーゲノムが自動で決定される
- 決定されたコアゲノムとアクセサリーゲノムのターゲット遺伝子リストと、コアゲノムMLSTスキーマが出力される



- MinHash アルゴリズムを使用して、ゲノム配列間の距離を高速に推定するツール
- NCBIに登録されている原核生物の完全長ゲノムからつくられたMash Referenceデータベースを内蔵しており、サンプルゲノムの配列データを指定するだけで、データベース登録されているゲノムとの類似度を容易に計算可能
- サンプルゲノムのデータには、FASTAファイルやFASTQファイルが使用可能

- ツールを実行すると、サンプルゲノムの入力ファイルごとに、近縁ゲノムのリストが出力される
- Mash-distanceフィールドによるゲノム配列間の変異率や、Matching-hashesフィールドによる一致したMinHash数などに基づき、サンプルと距離が近いゲノムを決定することができる
- シードゲノムとしてどの株のゲノムを用いればよいか分からない場合に、本ツールの結果を参考にシードゲノムの選択を行う

Mash Identification Results

Mash Identification Results

Taxonomic Info	Reference Sketch-ID	Mash-distance	Matching-hashes	P-value
Neisseria meningitidis M22293	NZ_CP031334	0.039	283/1000	0.0
Neisseria meningitidis FDAARGOS_214	NZ_CP020401	0.04	278/1000	0.0
Neisseria meningitidis NCTC12163	NZ_LR134528	0.04	278/1000	0.0
Neisseria meningitidis DE8669	NZ_CP012391	0.04	275/1000	0.0
Neisseria meningitidis NZ-05/33	NC_017518	0.04	274/1000	0.0
Neisseria meningitidis 53442	NC_010120	0.04	272/1000	0.0
Neisseria meningitidis 14-563	NZ_CP021516	0.04	272/1000	0.0
Neisseria meningitidis M01-240149	NC_017514	0.041	271/1000	0.0
Neisseria meningitidis B6116/77	NZ_CP007667	0.041	271/1000	0.0
Neisseria meningitidis NM3686	NZ_CP009418	0.041	271/1000	0.0
Neisseria meningitidis 12-221	NZ_CP021517	0.041	271/1000	0.0
Neisseria meningitidis 12-176	NZ_CP021518	0.041	271/1000	0.0
Neisseria meningitidis 11-7	NZ_CP021520	0.041	271/1000	0.0
Neisseria meningitidis 06-178	NZ_CP021522	0.041	271/1000	0.0
Neisseria meningitidis 98-182	NZ_CP021523	0.041	271/1000	0.0
Neisseria meningitidis 13-600	NZ_CP021723	0.041	271/1000	0.0
Neisseria meningitidis 12-330	NZ_CP021724	0.041	271/1000	0.0
Neisseria meningitidis M21374	NZ_CP031326	0.041	271/1000	0.0
Neisseria meningitidis AUSMDU00005726	NZ_CP045960	0.041	271/1000	0.0
Neisseria meningitidis FAM18	NC_008767	0.041	270/1000	0.0

180 rows

Close

コアゲノムMLSTスキーマ作成例

菌種

- *Neisseria meningitidis*

シードゲノム

- NC_008767 (*Neisseria meningitidis* FAM18)

クエリーゲノム

- NC_003112 (*Neisseria meningitidis* MC58)
- NC_003116 (*Neisseria meningitidis* Z2491)
- NC_010120 (*Neisseria meningitidis* 053442)
- NC_013016 (*Neisseria meningitidis* alpha14)
- NC_017501 (*Neisseria meningitidis* 8013)
- NC_017505 (*Neisseria meningitidis* alpha710)
- NC_017512 (*Neisseria meningitidis* WUE 2594)
- NC_017513 (*Neisseria meningitidis* G2136)
- NC_017514 (*Neisseria meningitidis* M01-240149)
- NC_017515 (*Neisseria meningitidis* M04-240196)
- NC_017516 (*Neisseria meningitidis* H44/76)
- NC_017517 (*Neisseria meningitidis* M01-240355)
- NC_017518 (*Neisseria meningitidis* NZ-05/33)

除外する遺伝子

- NZ_CM000956 (*Neisseria meningitidis* K1207 plasmid)
- NZ_CM000958 (*Neisseria meningitidis* S0108 plasmid)

- ゲノムのNCBI Accession番号を直接入力

OR

Chromosome Acces...	Organism/Name	Strain	Status
NC_003112.2	<i>Neisseria meningitidis</i> MC58	MC58	CompleteGenome
NC_003116.1	<i>Neisseria meningitidis</i> Z2491	Z2491	CompleteGenome
NC_008767.1	<i>Neisseria meningitidis</i> FAM18	FAM18	CompleteGenome
NC_010120.1	<i>Neisseria meningitidis</i> 053442	53442	CompleteGenome
NC_013016.1	<i>Neisseria meningitidis</i> alpha14	alpha14	CompleteGenome
NC_017501.1	<i>Neisseria meningitidis</i> 8013	8013	CompleteGenome
NC_017505.1	<i>Neisseria meningitidis</i> alpha710	alpha710	CompleteGenome
NC_017512.1	<i>Neisseria meningitidis</i> WUE 2594	WUE 2594	CompleteGenome
NC_017513.1	<i>Neisseria meningitidis</i> G2136	G2136	CompleteGenome
NC_017516.1	<i>Neisseria meningitidis</i> H44/76	H44/76	CompleteGenome
NC_017514.1	<i>Neisseria meningitidis</i> M01-240149	M01-240149	CompleteGenome
NC_017517.1	<i>Neisseria meningitidis</i> M01-240355	M01-240355	CompleteGenome
NC_017515.1	<i>Neisseria meningitidis</i> M04-240196	M04-240196	CompleteGenome
NC_017518.1	<i>Neisseria meningitidis</i> NZ-05/33	NZ-05/33	CompleteGenome
NZ_CP007524.1	<i>Neisseria meningitidis</i>	510612	CompleteGenome
NZ_CP009422.1	<i>Neisseria meningitidis</i>	M10208	CompleteGenome

- NCBI検索結果よりゲノムを選択

- クエリーゲノムを入力した場合は、シードゲノムと比較して共通に存在する遺伝子数をチェックし、少ないものがあつた場合はスキーマ作成前に除外する

Seed genome: NC_008767.1 (12-MAR-2023), 2194961 bases, 2111 genes with CDS (Neisseria meningitidis FAM18, complete sequence)
 Genes excluded by Multi Copy Filter: 185

Query genome	Genes with CDS	Genes not found	Genes with missing/wrong stop codons	Sum
NC_003112.2 (04-DEC-2023), 2272360 bases	2178 genes with CDS (Neisseria meningitidis MC58, complete sequence)	291	107	398
NC_003116.1 (24-DEC-2022), 2184406 bases	2121 genes with CDS (Neisseria meningitidis Z2491, complete sequence)	265	84	349
NC_010120.1 (04-DEC-2023), 2153416 bases	2073 genes with CDS (Neisseria meningitidis 053442, complete sequence)	255	100	355
NC_013016.1 (30-DEC-2022), 2145295 bases	2049 genes with CDS (Neisseria meningitidis alpha14, complete sequence)	276	82	358
NC_017501.1 (24-DEC-2022), 2277550 bases	2195 genes with CDS (Neisseria meningitidis 8013, complete sequence)	213	92	305
NC_017505.1 (22-FEB-2023), 2242947 bases	2173 genes with CDS (Neisseria meningitidis alpha710, complete sequence)	215	88	303
NC_017512.1 (30-DEC-2022), 2227255 bases	2171 genes with CDS (Neisseria meningitidis WJUE 2594, complete sequenc...	257	99	356
NC_017513.1 (22-FEB-2023), 2184862 bases	2098 genes with CDS (Neisseria meningitidis G2136, complete sequence)	101	78	179
NC_017514.1 (12-FEB-2023), 2223518 bases	2150 genes with CDS (Neisseria meningitidis M01-240149, complete seque...	207	88	295
NC_017515.1 (22-FEB-2023), 2250449 bases	2158 genes with CDS (Neisseria meningitidis M04-240196, complete seque...	227	96	323
NC_017516.1 (06-SEP-2023), 2240883 bases	2151 genes with CDS (Neisseria meningitidis H44/76, complete sequence)	263	97	360
NC_017517.1 (22-FEB-2023), 2287777 bases	2206 genes with CDS (Neisseria meningitidis M01-240355, complete seque...	221	91	312
NC_017518.1 (22-FEB-2023), 2248966 bases	2190 genes with CDS (Neisseria meningitidis NZ-05/33, complete sequenc...	206	93	299

Settings | Filter for seed genome | Filter for penetration query genomes | Manually exclude targets

- Minimum Length Filter
Discard all genes that are shorter than 50 bases.
- Start Codon Filter
Discard all genes that contain no start codon at the beginning of the gene.
- Stop Codon Filter
Discard all genes that contain no stop codon, more than 1 stop codon or if the stop codon is not at the end of the gene.
Does not consider any GenBank annotations that indicate non-continuous coding regions.
- Homologous Gene Filter
Discard all genes that have fragments that occur in multiple copies in seed genome (with identity >= 90% and more than 100 bases overlap).
- Gene Overlap Filter
If two genes overlap more than 4 bases, move the shorter gene to Accessory targets.

- ゲノム配列内の特定コドンの有無のチェックや相同遺伝子の除外といった、ターゲット遺伝子検索における基本的なフィルターも調整可能

Name	Organism	Owner	Application	Targets
N. meningitidis cgMLST 1780 targets FAM18 seed only	N. meningitidis	Analyzer	Typing	1,780
N. meningitidis Accessory 85 targets FAM18 seed only	N. meningitidis	Analyzer	Accessory Typing	85



N. meningitidis cgMLST 1780 targ...	
1524 Good Targets	
60 Failed Targets	
196 Not Found Targets	
Perc. Good Targets:	85.6

■ Ad hoc cgMLST Scheme作成例

■ Ad hoc cgMLST Schemeを用いた cgMLST解析結果

Name	Organism	Owner	Application	Targets
N. meningitidis cgMLST 1295 targets FAM18	N. meningitidis	Analyzer	Typing	1,295
N. meningitidis Accessory 570 targets FAM18	N. meningitidis	Analyzer	Accessory Typing	570



N. meningitidis cgMLST 1295 targ...	
1269 Good Targets	
19 Failed Targets	
7 Not Found Targets	
Perc. Good Targets:	98.0

■ Stable cgMLST Scheme作成例

■ Stable cgMLST Schemeを用いた cgMLST解析結果

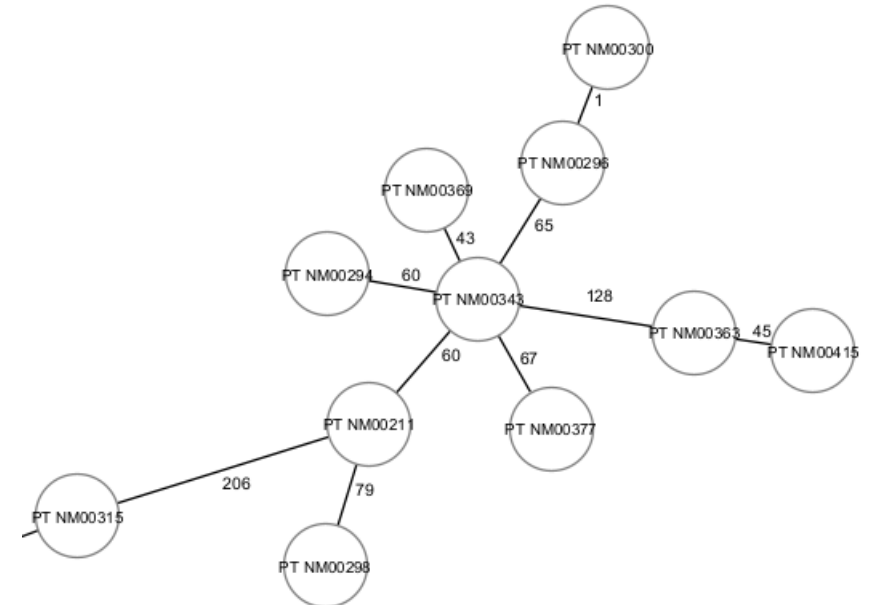
NMC_RS0	NMC_RS0	NMC_RS0037	NN	NN	NN	NN	NN	NN	NN	NN	NMC_RS1344	NN	NMC_RS0042	NMC_RS1416	NMC_RS0045
N. meningitidis	N. meningitidis	N. meningitidis	N.	N.	N.	N.	N.	N.	N.	N.	N. meningitidis	N.	N. meningitidis	N. meningitidis	N. meningitidis
? (failed)	1	? (not found)	1	1	1	1	1	1	1	1	? (failed)	1	? (not found)	? (not found)	? (not found)
? (failed)	1	? (not found)	1	1	1	1	1	1	1	1	? (failed)	1	? (not found)	? (not found)	? (not found)
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	2	2	
1	2	? (not found)	1	1	1	1	1	1	1	? (not found)	1	1	? (not found)	1	
1	2	? (not found)	1	1	1	1	1	1	1	? (failed)	1	1	1	1	
1	2	? (not found)	1	1	1	1	1	1	1	? (not found)	1	1	1	1	
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	2	2	
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	2	2	
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	2	2	
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	? (not found)	2	
1	? (failed)	1	1	2	2	2	1	2	2	? (not found)	2	2	2	2	
? (failed)	1	? (not found)	1	1	1	1	1	1	1	? (failed)	1	? (not found)	? (not found)	? (not found)	
? (failed)	1	? (not found)	1	1	1	1	1	1	1	? (failed)	1	1	? (not found)	? (not found)	



欠損データが多く、系統樹が作成できない

- Ad hoc cgMLST Scheme解析結果からつくられた Comparisonテーブル

NN	NN	NM	NM	NM	NN	NN	NN	NN	NN	NN	NM	NMC_RS0050	NMC_RS0051	NN	NN	NN	NMC_RS0053
N.	N.	N.	N.	N.	N.	N.	N.	N.	N.	N.	N.	N. meningitidis	N. meningitidis	N.	N.	N.	N. meningitidis
5	2	8	6	15	1	1	1	1	1	1	8	2	3	11	5	1	1
5	2	8	6	15	1	1	1	1	1	1	8	2	3	11	5	1	1
4	3	3	3	12	3	3	1	3	3	4	3	1	3	3	3	3	3
5	2	9	10	16	1	1	1	1	1	1	8	2	3	11	5	1	1
5	2	9	10	16	1	1	1	1	1	1	8	2	3	11	5	1	1
5	2	9	10	16	1	1	1	1	1	1	8	2	3	11	5	1	1
4	3	3	3	5	3	3	1	3	3	4	3	1	3	3	3	3	3
4	3	3	3	12	3	3	1	3	3	4	3	1	? (not found)	3	3	3	3
4	3	3	3	12	3	3	1	3	3	4	3	1	3	3	3	3	3
4	3	3	3	12	3	3	1	3	3	4	3	1	3	3	3	3	3
4	3	10	1	12	3	3	1	3	3	4	3	1	7	12	8	7	? (not found)



系統樹が作成可能

- Stable cgMLST Scheme解析結果からつくられた Comparisonテーブル

お問い合わせ先：フィルジエン株式会社

TEL: 052-624-4388 (9:00～18 : 00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp