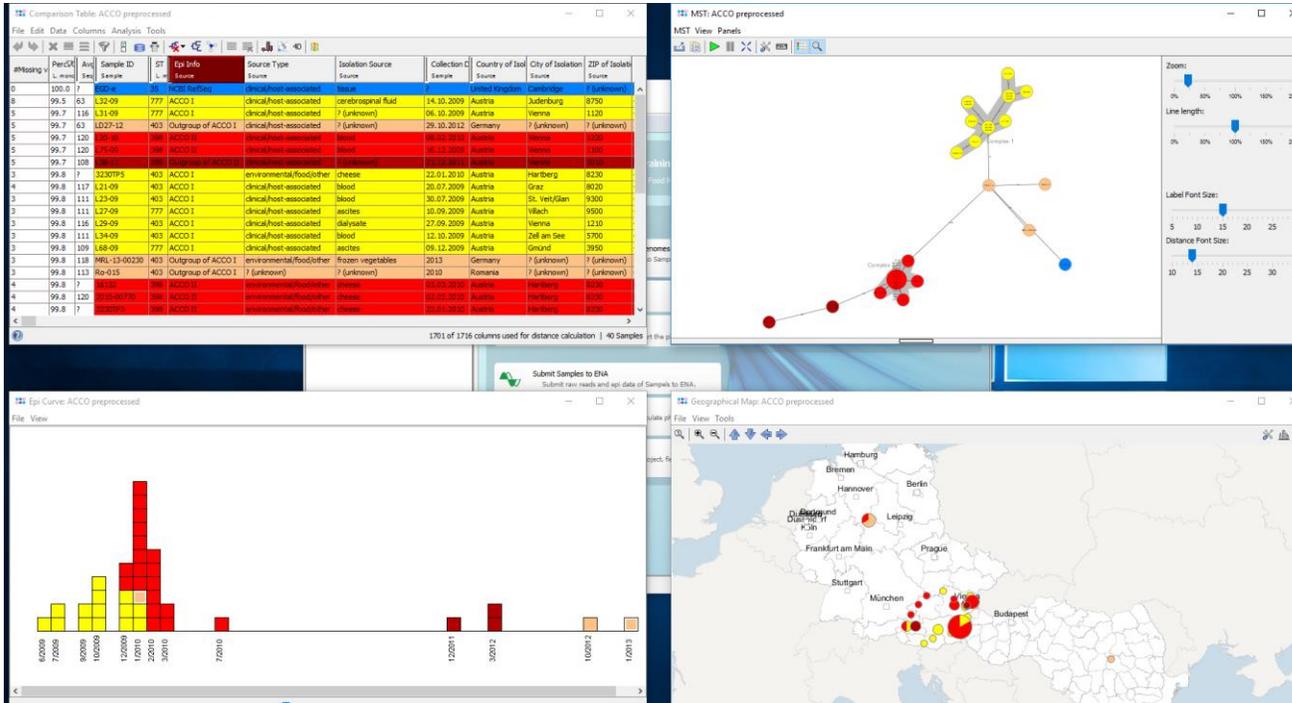


全ゲノムシーケンスによる 病原微生物の分子疫学解析

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)



■ 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

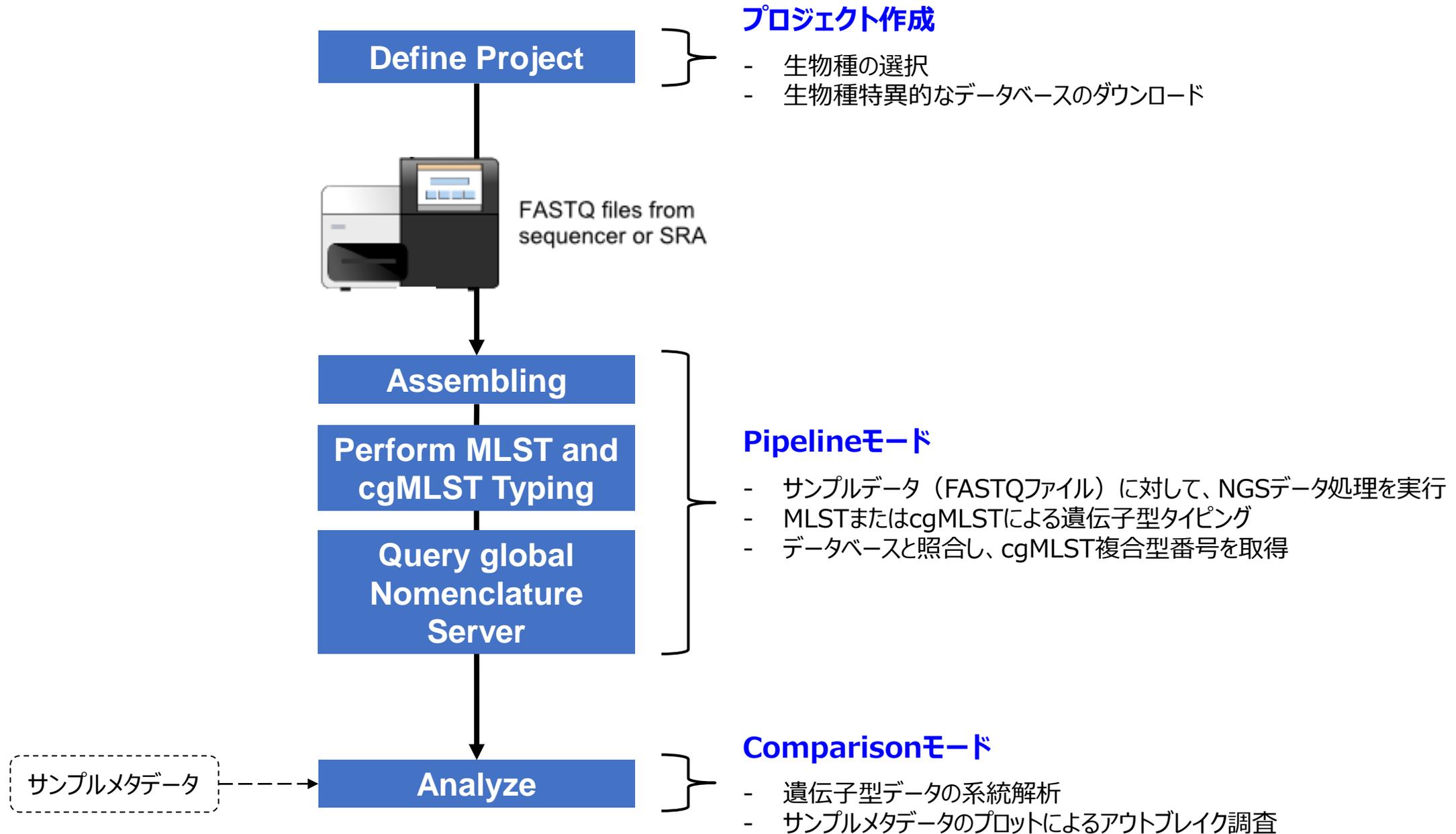
■ サンプルごとのcgMLST複合型の照合

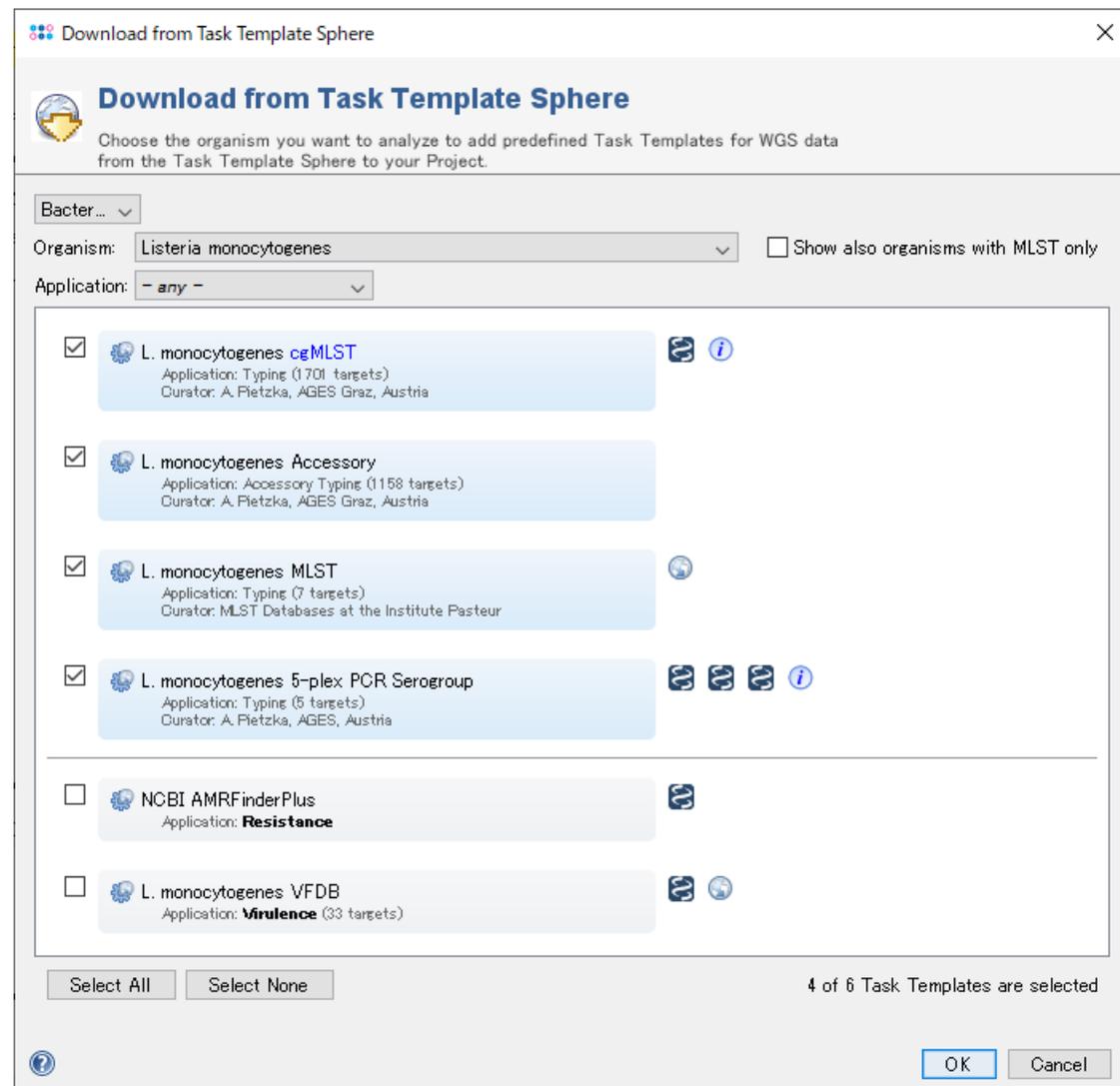
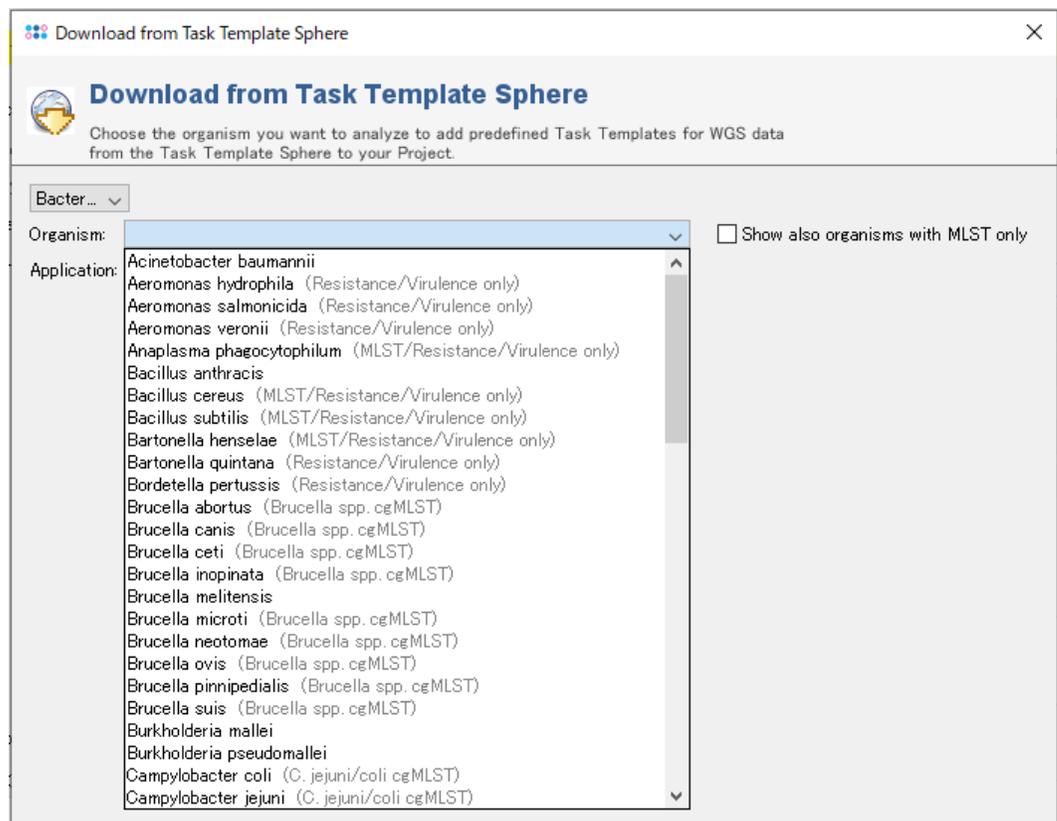
■ 系統解析

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

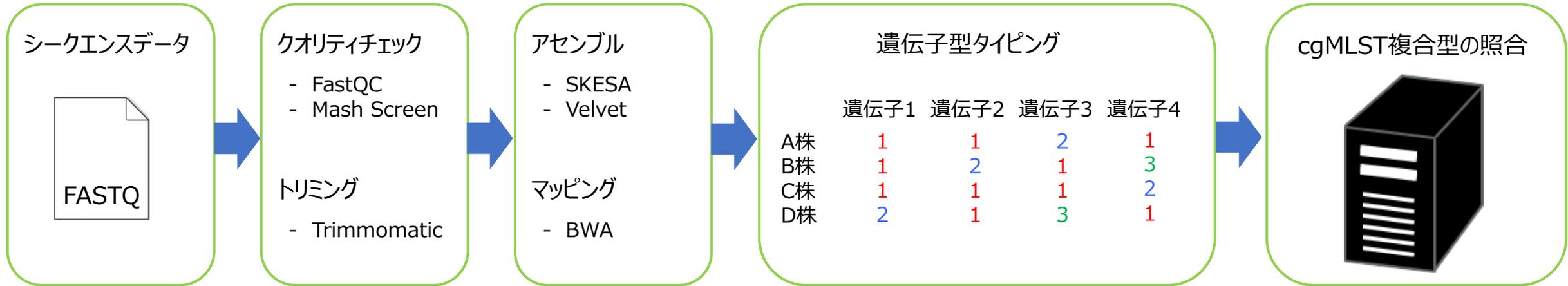
■ アウトブレイク調査

- Epi Curve
- Geographical Map





- プロジェクト作成時には、180種類以上の菌種が選択可能
- 菌種を選択すると、データ解析に利用可能なデータベースの一覧が表示される



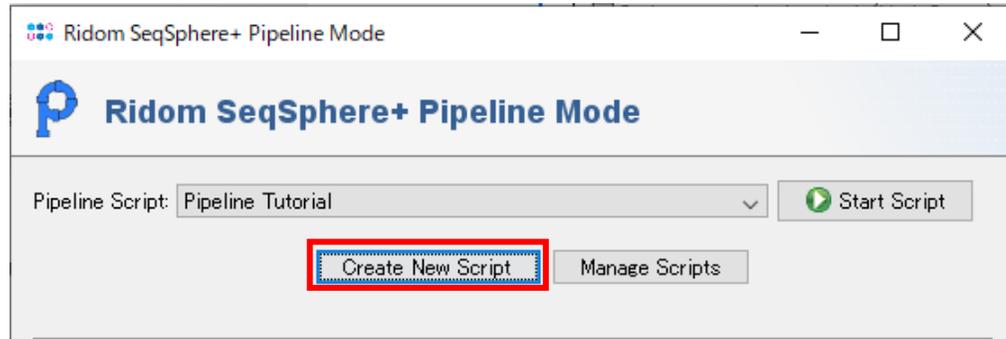
```

SeqSphere+ Pipeline: Lm ACCO demo
Whole Pipeline: Sample L14-10 (1/2)
Current Step: Assembling/Mapping reads...

07.06.16 17:10:08 Ridom SeqSphere+ 3.2.2-rc14_h_(2016-06) (workshoplivedemo)
07.06.16 17:10:08 Start Pipeline Script 'Lm ACCO demo'
07.06.16 17:10:08 Connecting server localhost
07.06.16 17:10:08 Login user: livedemo (host: LenovoZ70)
07.06.16 17:10:08 Found files for 2 Samples

07.06.16 17:10:08 Initialize Sample L14-10
07.06.16 17:10:08 Start processing files L14-10_1.fastq.gz, L14-10_2.fastq.gz
07.06.16 17:10:08 No Project acronym found
07.06.16 17:10:08 Using default Project for import: ACCO
07.06.16 17:10:09 Found existing Sample L14-10 in database
07.06.16 17:10:09 Start assembling with Velvet...
07.06.16 17:10:09 Velvet params: downsample to 120x coverage, trim-qual: 30, trim-wsize: 20 bp
07.06.16 17:10:09 Trimming at 5' and 3' end until average quality is 30 in a window of 20 bases.
07.06.16 17:10:10 Trimmed 105277 bases.
07.06.16 17:10:10 Merging to interleaved read file.
07.06.16 17:10:10 Merging 2 files.
07.06.16 17:10:10 After processing, read files contain 2,220,847 bases in 9,900 reads (avg. read
    
```

- サンプルごとのFASTQファイルより、データの各種QCとアセンブル・マッピングといったNGSの基本的なデータ処理を実行
- データ処理が完了すると、ダウンロードしたスキーマ (MLST、cgMLST) に基づき、各遺伝子の遺伝子型が決定される
- cgMLST遺伝子型データを外部サーバーに送信し、複合型の番号を取得



- Pipelineモードでは、NGSデータ処理パイプラインのスクリプトを事前に作成してデータ解析を実行する
- 処理内容のパラメータ設定はソフトウェア上で対話形式で行い、項目を選択するだけで設定可能なため、簡単な操作で行うことができる
- 作成したスクリプトは、別サンプルの解析時に再利用したり、設定内容の編集も可能

■ クオリティチェック・トリミング設定

- Perform read data quality and adapter control (FastQC)
- Perform Illumina adapter trimming (Trimmomatic) if adapters found by FastQ... ▾
- Perform contamination check (Mash Screen)
- Store read data and/or base qualities for targets with QC errors/warnings if available
- Continuous Mode: Run pipeline as continuous process that monitors the input sources for new files

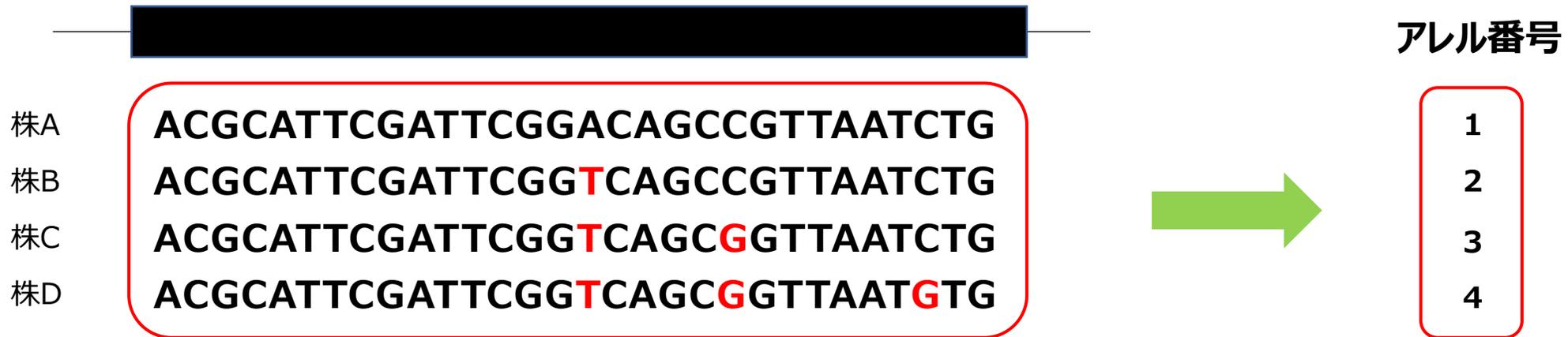
■ アセンブル・マッピング設定

- Perform Assembling/Mapping for read files
- For 'Illumina': SKESA (de novo assembling) ▾ Advanced Settings... 📄
- Seed Genome: NC_003210.1 Alternative Seed Genome...
- Expected Genome Size: 2.9 Mbp

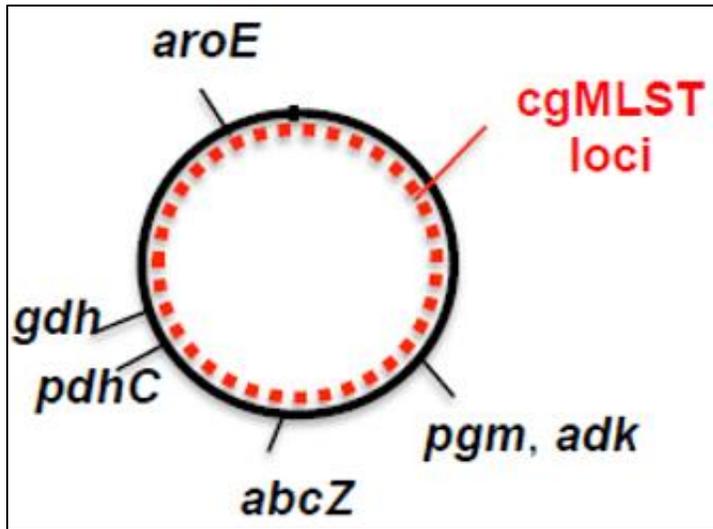
■ ファイル出力設定

- Assembly Result Files
- Do not keep aligned read files (ACE/BAM) in SeqSphere+
 - Copy aligned read files (ACE/BAM) files to folder: 📁
 - Create sub-folder for each Project
 - Export contig files (FASTA) to the folder

遺伝子



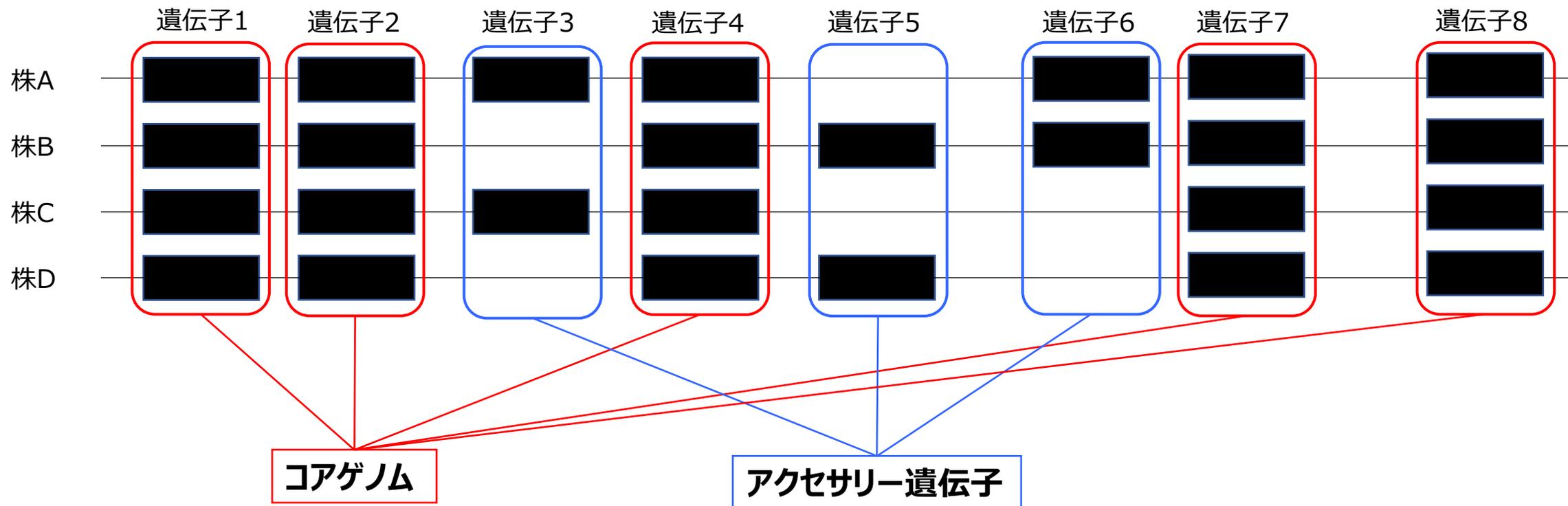
- MLSTによる遺伝子型タイピングでは、NGSデータ処理結果の遺伝子ごとの塩基配列より、スキーマ（MLST, cgMLST）に基づきアレル番号に変換して出力
- 配列が違えば異なるアレル番号がつく
- 塩基配列ではなく、遺伝子ごとのアレル番号の組み合わせ（プロファイル）データを用いて、系統解析を実行することが可能になる



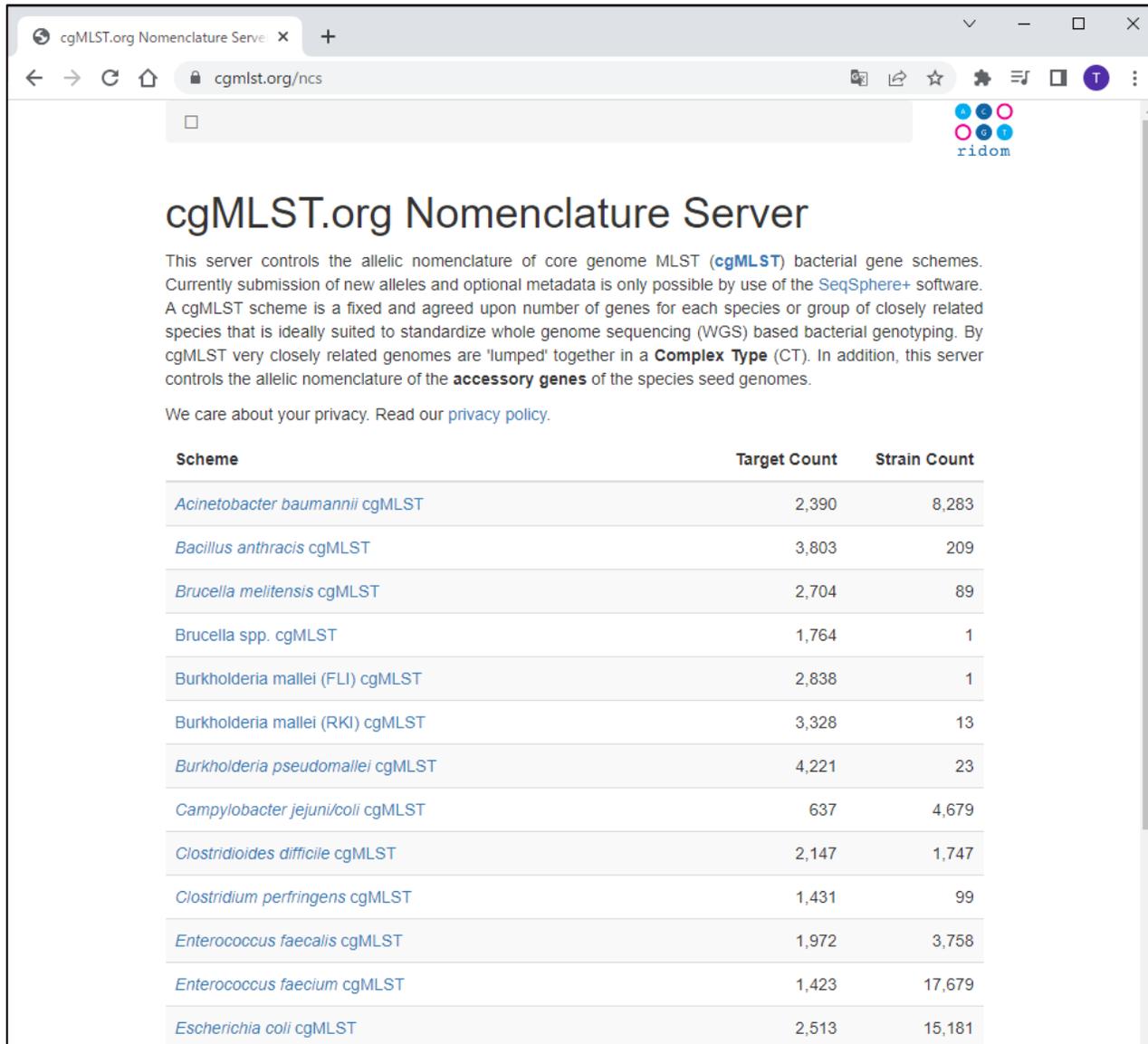
- MLSTでは少数の遺伝子のみタイピングを行うが、cgMLSTでは多数の遺伝子を用いるため株間の識別力が高い
- ゲノム上の対象領域が大きいため、次世代シーケンサーによる全ゲノムシーケンスを行うことが必要となる

■ MLST解析の場合： 6, 7遺伝子のみ
1-1-2-1-3-1-1 ➔ 株間の識別力が**低い**

■ cgMLST解析の場合： 1000 - 4000遺伝子
1-2-4-2...2-1-1-3 ➔ 株間の識別力が**高い**



- cgMLSTでは、菌種内で共通に存在する遺伝子（コアゲノム）を対象にしてタイピングを行う
- 一部の株のみが持つ遺伝子はアクセサリ-遺伝子とよばれ、cgMLSTによる遺伝子型タイピングには使用しない
- コアゲノムのみを用いることにより、全遺伝子を用いる場合と比べて研究室間での標準化が行いやすくなる



Scheme	Target Count	Strain Count
<i>Acinetobacter baumannii</i> cgMLST	2,390	8,283
<i>Bacillus anthracis</i> cgMLST	3,803	209
<i>Brucella melitensis</i> cgMLST	2,704	89
<i>Brucella</i> spp. cgMLST	1,764	1
<i>Burkholderia mallei</i> (FLI) cgMLST	2,838	1
<i>Burkholderia mallei</i> (RKI) cgMLST	3,328	13
<i>Burkholderia pseudomallei</i> cgMLST	4,221	23
<i>Campylobacter jejuni/coli</i> cgMLST	637	4,679
<i>Clostridioides difficile</i> cgMLST	2,147	1,747
<i>Clostridium perfringens</i> cgMLST	1,431	99
<i>Enterococcus faecalis</i> cgMLST	1,972	3,758
<i>Enterococcus faecium</i> cgMLST	1,423	17,679
<i>Escherichia coli</i> cgMLST	2,513	15,181

- cgMLSTでは、アレルプロファイルを外部サーバー（cgMLST.org）のデータと照合し、サンプルを識別するための複合型番号（Complex Type: CT）を取得する
- 類似したアレルプロファイルをもつサンプルは同一のCTが割り当てられるため、サンプルの系統分類を行うことができる

L. monocytogenes MLST (L10-10 20...)

abcZ: Target Contig

bgIA: Target Contig

cat: Target Contig

dapE: Target Contig

dat: Target Contig

ldh: Target Contig

lhkA: Target Contig

ST: 398

CC: CC19

Profile: 7, 13, 19, 6, 1, 7, 1

- MLSTスキームを用いた遺伝子型タイピングでは、6, 7個程度のハウスキーピング遺伝子が対象となる
- タイピング結果として、遺伝子ごとのアレル番号と、アレル番号の組み合わせにより決定された塩基配列型 (Sequence Type: ST) が出力される
- 菌種によっては、類似した塩基配列型をグループにまとめたClonal Complex (CC)も出力される
- 遺伝子ごとに、アセンブルされたコンティグ配列や標準配列との違いも確認可能

■ MLSTタイピング結果のサマリー

Contig: abcZ (L10-10 2014-06-20 15-34-18) (53...)

Editing Mode, locked for other users

Ref.-seq. |< >|

Consensus: |1| 5| 10| 15| 20| 25| 30| 35| 40| 45| 50| 55| 60| 65| 70| 75| 80| 85| 90| 95| 100| 105| 110| 115| 120| 125| 130| 135| 140| 145|

Ref.-seq.: AAAATCGACGAAACAGAATGCGTATAGGAGCTTTTCCGCAAGATGGAAAAACTATCGATTTCGTTTCTTCGATAGCCGCAATGATGGCGAAATGCTTAGCCGCTTCACTAGTGACTTAGATAAATTTTCCAATACACTAAACCAAGCATTT

Consensus AA: |Lys| |Ser| |Thr| |Asn| |Arg| |Met| |Arg| |Ile| |Gly| |Leu| |Phe| |Arg| |Lys| |Met| |Glu| |Lys| |Leu| |Ser| |Ile| |Arg| |Phe| |Phe| |Asp| |Ser| |Arg| |Asn| |Asp| |Gly| |Glu| |Met| |Leu| |Ser| |Arg| |Phe| |Thr| |Ser| |Asp| |Leu| |Asp| |Asn| |Ile| |Ser| |Asn| |Thr| |Leu| |Asn| |Gln| |Ala|

Ref.-seq. AA: Lys Ser Thr Asn Arg Met Arg Ile Gly Leu Phe Arg Lys Met Glu Lys Leu Ser Ile Arg Phe Phe Asp Ser Arg Asn Asp Gly Glu Met Leu Ser Arg Phe Thr Ser Asp Leu Asp Asn Ile Ser Asn Thr Leu Asn Gln Ala

Contig |< >|

Consensus: |5| Signature found |10| 15| 20| 25| 30| 35| 40| 45| 50| 55| 60| 65| 70| 75| 80| 85| 90| 95| 100| 105| 110| 115| 120| 125| 130| 135| 140| 145|

abcZ: AAAATCGACGAAACAGAATGCGTATAGGAGCTTTTCCGCAAGATGGAAAAACTATCGATTTCGTTTCTTCGATAGCCGCAATGATGGCGAAATGCTTAGCCGCTTCACTAGTGACTTAGATAAATTTTCCAATACACTAAACCAAGCATTT

■ コンティグ配列と標準配列の確認画面

L. monocytogenes cgMLST (L10-10 ...)

- 1697 Good Targets, 1 with warnings
- 2 Failed Targets
- 2 Not Found Targets

Complex Type: 39
Perc. Good Targets: 99.8

- cgMLSTスキームを用いた遺伝子型タイピングでは、菌種内共通に存在するコア遺伝子が対象となる
- タイピング結果として、遺伝子ごとのアレル番号と、アレル番号の組み合わせにより決定されたcgMLST複合型 (Complex Type: CT) が出力される
- 遺伝子ごと、あるいは全体のQC情報なども得ることが可能

■ cgMLSTタイピング結果のサマリー

cgMLST

Complex Type: 39

1701 targets: 1697 known alleles, 4 missing

Target	Allele Type	Genotyping Result	Target QC Procedure
lmo0001	3	known allele	OK
lmo0002	14	known allele	OK
lmo0003	14	known allele	OK
lmo0004	1	known allele	OK
lmo0005	15	known allele	OK
lmo0006	19	known allele	OK
lmo0007	21	known allele	OK
lmo0008	12	known allele	OK
lmo0009	5	known allele	OK

■ 遺伝子ごとのアレル番号

L. monocytogenes Accessory (L10-...

- 937 Good Targets, 1 with warnings
- 13 Failed Targets
- 208 Not Found Targets

■ アクセサリー遺伝子

L. monocytogenes 5-plex PCR Sero...

Imo0737: Target [TCR](#) [Contig](#)

Imo1118:

ORF2819:

ORF2110:

prs: Target [TCR](#) [Contig](#)

serogroup: IIa (serotype 1/2a and 3a)

■ Serotype (血清型)

NCBI AMRFinderPlus (L10-10 2014-...

Fosfomycin: fosX

Lincosamide: lin

■ 薬剤耐性因子

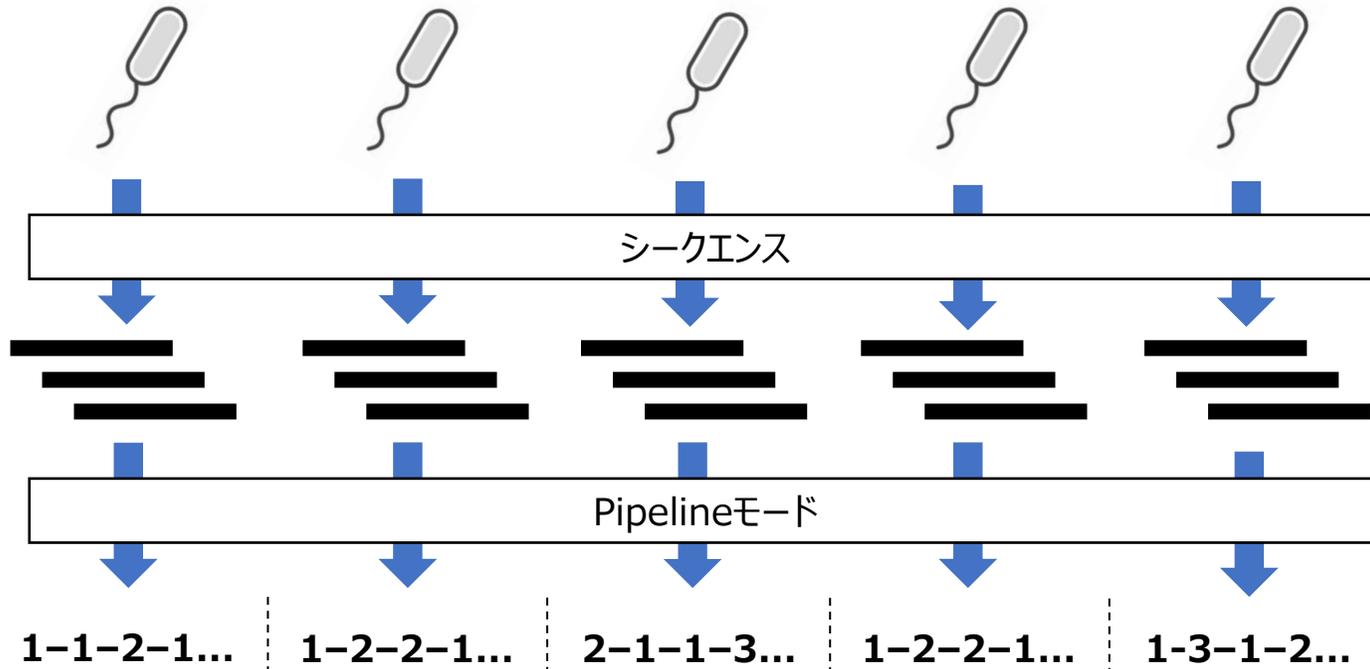
L. monocytogenes VFDB (L10-10 20...

- 32 Good Targets, 1 with warnings
- 1 Not Found Targets

Confident Targets: actA / aut / bsh / clpC ,

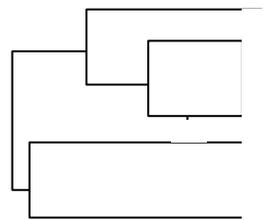
■ 病原因子

- MLST解析結果の遺伝子型タイピングデータに加え、他のデータベースを利用した解析結果も同時に出力される
- 各結果の詳細データ（遺伝子配列やデータテーブルなど）も表示可能

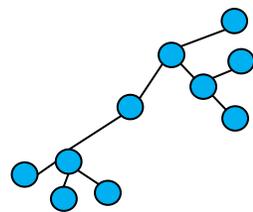


- Pipelineモードで全サンプルの遺伝子型タイピングが完了した後は、Comparisonモードでデータの俯瞰、サンプル間比較解析などを行う
- Comparisonモードでは、系統樹やネットワークグラフによる系統解析や、メタデータを用いたアウトブレイク調査などが可能

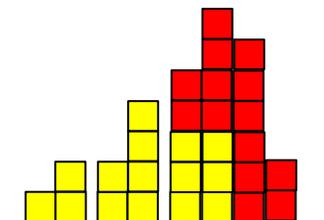
Comparisonモード



✓ 系統樹作成



✓ ネットワーク図作成



✓ アウトブレイク調査

Comparisonテーブルの作成

ST番号 CT番号 コアゲノムのアレル番号

ハウスキーピング遺伝子のアレル番号

Sample ID Sample	ST S. enterica MLST	Complex Type S. enterica cgMLST	SC0831 S. enterica	SEN0401 S. enterica cg	SNSL254 S. enterica c	SPAB_04503 S. enterica cgML	STM0834 S. enterica c	STM0942 S. enterica c
Senterica ERX248790	11	11677	3	5	4	1	4	5
Senterica ERX248791	11	384	3	5	4	1	4	5
Senterica ERX248792	11	11677	3	5	4	1	4	5
Senterica ERX248793	11	11681	3	5	4	1	4	5
Senterica ERX248794	11	11681	3	5	4	1	4	5
Senterica ERX248796	11	11846	3	5	4	1	4	5
Senterica ERX248797	11	11878	3	5	4	1	4	5
Senterica ERX248795	11	396	3	5	4	1	4	5

Sample ID Sample	ST S. enterica MLST	Complex Type S. enterica cgMLST	aroC S. enterica M	dnaN S. enterica M	hemD S. enterica M	hisD S. enterica M	purE S. enterica	sucA S. enterica M
Senterica ERX248790	11	11677	5	2	3	7	6	6
Senterica ERX248791	11	384	5	2	3	7	6	6
Senterica ERX248792	11	11677	5	2	3	7	6	6
Senterica ERX248793	11	11681	5	2	3	7	6	6
Senterica ERX248794	11	11681	5	2	3	7	6	6
Senterica ERX248796	11	11846	5	2	3	7	6	6
Senterica ERX248797	11	11878	5	2	3	7	6	6
Senterica ERX248795	11	396	5	2	3	7	6	6

■ cgMLST遺伝子型テーブル

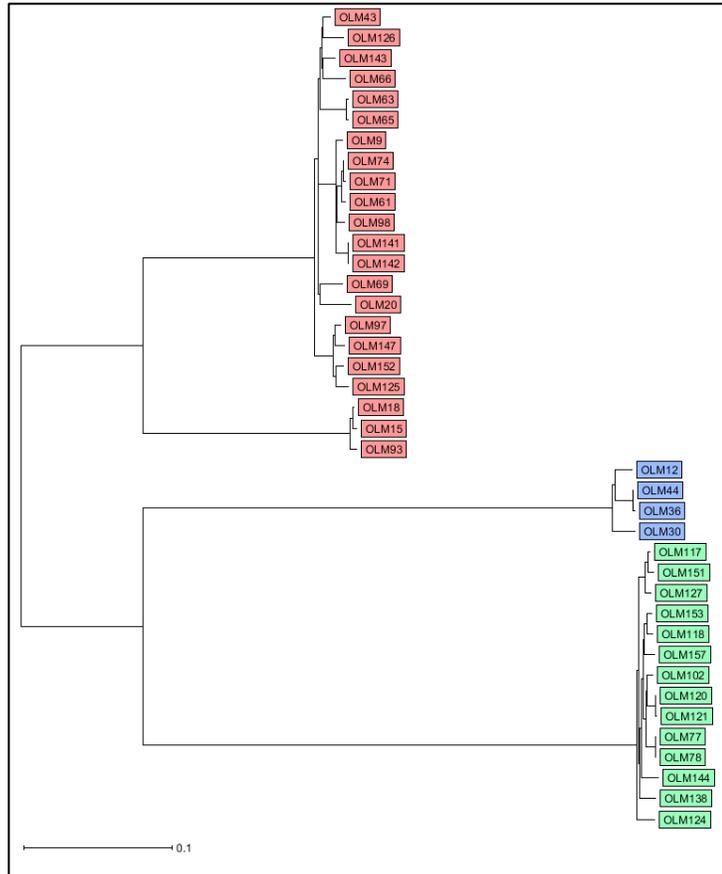
■ MLST遺伝子型テーブル

コアゲノムまたはハウスキーピング遺伝子のSNP

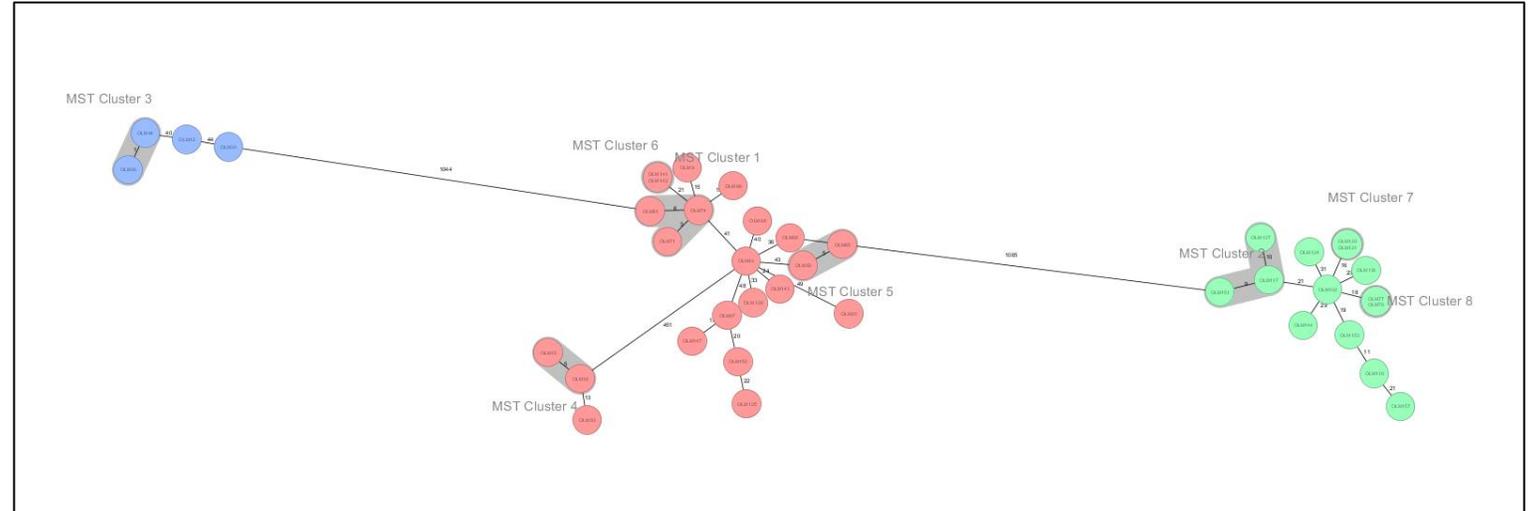
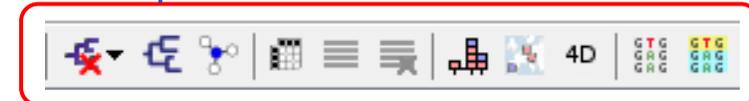
Sample ID Sample	ST S. enterica MLST	Complex Type S. enterica cgMLST	930566 SD0831	930573 SD0831	930590 SD0831	451452 SEN0401	451519 SEN0401	451555 SEN0401
Senterica ERX248790	11	11677	T	C	G	A	A	G
Senterica ERX248791	11	384	T	C	G	A	A	G
Senterica ERX248792	11	11677	T	C	G	A	A	G
Senterica ERX248793	11	11681	T	C	G	A	A	G
Senterica ERX248794	11	11681	T	C	G	A	A	G
Senterica ERX248796	11	11846	T	C	G	A	A	G
Senterica ERX248797	11	11878	T	C	G	A	A	G
Senterica ERX248795	11	396	T	C	G	A	A	G

■ SNPテーブル

- まず初めに、プロジェクト内全サンプルのタイピング結果をまとめたComparisonテーブルを作成
- テーブル表示の遺伝子型タイピングデータには、MLSTとcgMLSTの遺伝子別アレル番号、またはSNPを用いることが可能
- 必要に応じて、薬剤耐性因子などの補足データのフィールドも追加可能



■ Neighbor Joining Tree



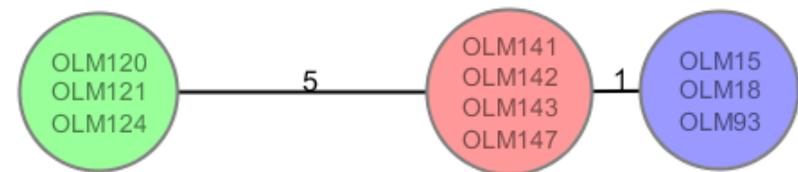
■ Minimum Spanning Tree

- Comparisonテーブルより、ボタンクリックだけで系統解析図を作成
- 系統解析は、テーブルに含まれている遺伝子型データ（MLST/cgMLSTのアレル番号、またはSNP）をもとに実行される

Sample ID	ST	abcZ	bgIA	cat	dapE	dat	ldh	lhkA
Sample	L. monocytogenes MLST	L. monoc						
OLM147	1	3	1	1	1	3	1	3
OLM142	1	3	1	1	1	3	1	3
OLM141	1	3	1	1	1	3	1	3
OLM143	1	3	1	1	1	3	1	3
OLM120	2	1	1	11	11	2	1	5
OLM121	2	1	1	11	11	2	1	5
OLM124	2	1	1	11	11	2	1	5
OLM15	308	3	1	1	1	18	1	3
OLM18	308	3	1	1	1	18	1	3
OLM93	308	3	1	1	1	18	1	3



少数遺伝子のみでの遺伝子型データでは、株間の識別が困難

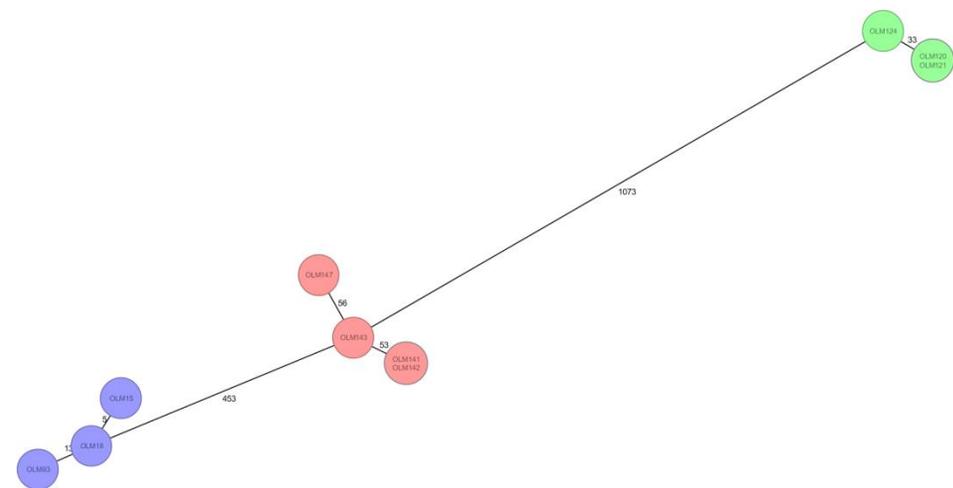


■ MLST遺伝子型テーブル

Sample ID	ST	lmo	lm	lmo0	lm	lmo0	lm										
Sample	L. monocytogenes MLST	L. m	L. r	L. mo	L. r	L. mo	L. r										
OLM147	1	2	2	2	4	4	4	7	2	6	2	6	4	2	4	4	4
OLM142	1	2	2	2	4	608	4	7	2	6	2	6	4	2	4	4	4
OLM141	1	2	2	2	4	608	4	7	2	6	2	6	4	2	4	4	4
OLM143	1	2	2	2	4	4	4	7	2	6	2	6	4	2	4	4	4
OLM120	2	2	2	116	2	2	2	2	2	2	2	2	2	2	2	2	2
OLM121	2	2	2	116	2	2	2	2	2	2	2	2	2	2	2	2	2
OLM124	2	2	2	866	2	2	2	2	2	2	2	2	2	2	2	2	2
OLM15	308	2	2	2	4	4	4	7	6	3	2	6	4	2	4	4	4
OLM18	308	537	2	2	4	4	4	7	6	3	2	6	4	2	4	4	4
OLM93	308	2	2	2	4	4	4	7	6	3	2	6	4	2	4	4	4



多数の遺伝子を用いれば、高解像度で識別が可能



■ cgMLST遺伝子型テーブル

特徴的なSNPの抽出

Abs. Position	Target	Rel. Position	Gene	Product	Protein_id	Ref.-Seq...	Target-Genomes (14)	Nontarget-Genomes (12)	Effect to Ref-...	Ref.-Seq. AA ...	Target-Genomes AA	Nontarget-Genomes AA
1214	lmo0001	897	dnaA	chromosome...	NP_463534.1	T	T (14)	C (12)	synonymous	V	V (14)	V (12)
2124	lmo0002	258	dnaN	DNA polyme...	NP_463535.1	C	C (14)	T (12)	synonymous	V	V (14)	V (12)
2505	lmo0002	639	dnaN	DNA polyme...	NP_463535.1	A	G (14)	A (12)	synonymous	K	K (14)	K (12)
3840	lmo0003	720		hypothetical ...	NP_463536.1	G	G (14)	A (12)	synonymous	G	G (14)	G (12)
5120	lmo0005	252	recF	recombinatio...	NP_463538.1	A	G (14)	A (12)	synonymous	E	E (14)	E (12)
5522	lmo0005	654	recF	recombinatio...	NP_463538.1	T	A (14)	T (12)	synonymous	L	L (14)	L (12)
5529	lmo0005	661	recF	recombinatio...	NP_463538.1	C	T (14)	C (12)	synonymous	L	L (14)	L (12)
5531	lmo0005	663	recF	recombinatio...	NP_463538.1	G	A (14)	G (12)	synonymous	L	L (14)	L (12)
6119	lmo0006	90	gyrB	DNA gyrase ...	NP_463539.1	A	G (14)	A (12)	synonymous	V	V (14)	V (12)
6248	lmo0006	219	gyrB	DNA gyrase ...	NP_463539.1	T	C (14)	T (12)	synonymous	I	I (14)	I (12)
6275	lmo0006	246	gyrB	DNA gyrase ...	NP_463539.1	C	T (14)	C (12)	synonymous	R	R (14)	R (12)
6290	lmo0006	261	gyrB	DNA gyrase ...	NP_463539.1	A	G (14)	A (12)	synonymous	G	G (14)	G (12)
6299	lmo0006	270	gyrB	DNA gyrase ...	NP_463539.1	A	T (14)					
6317	lmo0006	288	gyrB	DNA gyrase ...	NP_463539.1	C	T (14)					
6338	lmo0006	309	gyrB	DNA gyrase ...	NP_463539.1	A	T (14)					
9165	lmo0007	1101	gyrA	DNA gyrase ...	NP_463540.1	A	A (14)					

Sample ID	ST	Complex Typ	1214	2124	2505	3840	5120	5522	5529	5531
Sample	L. monocytogenes	L. monocytogenes	lmo0001	lmo0002	lmo0002	lmo0003	lmo0005	lmo0005	lmo0005	lmo0005
L10-10...	398	39	T	C	G	G	G	A	T	A
L42-10...	398	39	T	C	G	G	G	A	T	A
L14-10...	398	39	T	C	G	G	G	A	T	A
L16-10...	398	39	T	C	G	G	G	A	T	A
L17-10...	398	39	T	C	G	G	G	A	T	A
L18-10...	398	39	T	C	G	G	G	A	T	A
L19-10...	398	39	T	C	G	G	G	A	T	A
L20-10...	398	39	T	C	G	G	G	A	T	A
L30-10...	398	39	T	C	G	G	G	A	T	A
L32-10...	398	39	T	C	G	G	G	A	T	A
L33-10...	398	39	T	C	G	G	G	A	T	A
L38-11...	398	41	T	C	G	G	G	A	T	A
L4-10 ...	398	39	T	C	G	G	G	A	T	A
L75-09...	398	39	T	C	G	G	G	A	T	A
L20-09...	403	35	C	T	A	A	A	T	C	G
L21-09...	403	35	C	T	A	A	A	T	C	G
L23-09...	403	35	C	T	A	A	A	T	C	G
L29-09...	403	35	C	T	A	A	A	T	C	G
L33-09...	403	35	C	T	A	A	A	T	C	G
L34-09...	403	35	C	T	A	A	A	T	C	G
L71-09...	403	35	C	T	A	A	A	T	C	G
L9-10 ...	403	35	C	T	A	A	A	T	C	G
L27-09...	777	35	C	T	A	A	A	T	C	G
L31-09...	777	35	C	T	A	A	A	T	C	G
L35-09...	777	35	C	T	A	A	A	T	C	G
L68-09...	777	35	C	T	A	A	A	T	C	G

- 任意のサンプルグループ (ST分類、CT分類など) に対して、サンプルグループに特徴的なSNPを抽出

Find Group Specific SNVs in Distance Columns

Find Group Specific SNVs in Distance Columns

Preselect 'Target-Genomes' and 'Nontarget-Genomes' by Groups

	'Target-Genomes'	'Nontarget-Genomes'
<input checked="" type="checkbox"/> Group '398' (14 Samples):	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Group '403' (8 Samples):	<input type="radio"/>	<input checked="" type="radio"/>
<input checked="" type="checkbox"/> Group '777' (4 Samples):	<input type="radio"/>	<input checked="" type="radio"/>

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