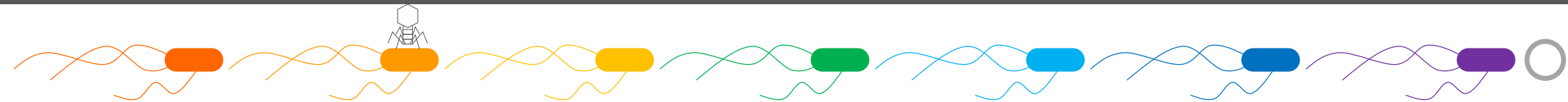


R3 希少感染症診断技術研修会 事例紹介 腸管出血性大腸菌(EHEC)関係

2022-02-18



国立感染症研究所・細菌第一部

李 謙一

※当日発表から一部改変しています。

※追加したスライドタイトルには[追加]を付けました。

Enterohemorrhagic *Escherichia coli* : EHEC

腸管出血性大腸菌(感染症法3類感染症)

≡Shiga toxin-producing *E. coli* :STEC

≡Vero toxin-producing *E. coli* :VTEC

志賀毒素(Shiga toxin: Stx)を産生する大腸菌

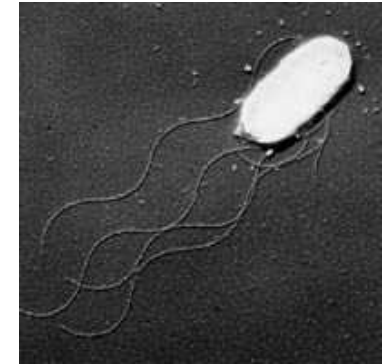
多数の血清型に分けられる: **O157:H7** など

病原性等

- 牛肉、野菜等の食品を介しての食中毒や、2次感染(ヒト-ヒト感染)を起こす
- 人での症状: 下痢、血便、溶血性尿毒症症候群 (HUS)
- 国内では年間約3,000名の感染者、~100名のHUS、~数名の死者

主な病原性機序

- **III型分泌装置**: 腸管上皮細胞への接着
 - **Stx**: タンパク質合成阻害による細胞傷害
- Stx**(学術的) = Vero toxin:**VT** (行政的) = Shiga-like toxin, Verocytotoxin etc.



EHECの全国サーベイランス

地方自治体
衛生研究所等

A県

B市

C県

菌株 (約3000株/年)

自治体間・内
比較情報

感染研
細菌第一部

MLVA

O157、O26、O111、O103、
O121、O145、O165、O91
→全体の**95%以上**

PFGE

その他O群

全ゲノム配列(whole-genome sequencing: WGS)解析による
精度の高いサーベイランスのために...

・解析パイプラインの構築・評価 ・ゲノム情報データベース化 ・従来法との比較

PFGEとWGSの比較
O121 広域散発事例
(Lee et al. Front. Microbiol. 2017)

MLVAとWGSの比較
類似MLVA型 O157
(Lee et al. Appl. Environ. Microbiol. 2019)

WGSを用いた
サーベイランス
(試行的)

1.WGSとは？

2.WGSとPFGEの比較 (0121)

3.WGSとMLVAの比較 (0157)

4.WGSによるサーベイランス (試行的)

1. WGSとは？

全ゲノム配列(whole-genome sequence: WGS)解析とは？

超並列型シーケンサー(次世代シーケンサー:NGS)を用いて、細胞内の全てのDNA配列(ゲノム)を解読する手法

	サンガー法	ショートリード型NGS	ロングリード型NGS
リード長	約600 bp	~300 bp	>10 kb
リード数	~数百	数兆以上	数億以上
リードのクオリティ	高	中	中~低
必要カバレッジ	2 (F & R)	数十以上	数十以上

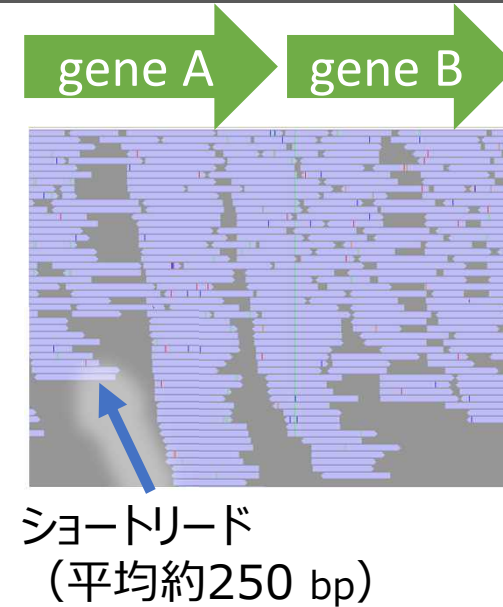
利点

- ・単一塩基多型(SNP)解析など
- ・ほぼ全ての遺伝子が解読可能

欠点

- ・費用が割高：1~2万円/株~
- ・解析が難しい

- 高解像度に菌株の類縁関係が分かる
- 網羅的な遺伝子型決定(病原因子、薬剤耐性遺伝子、MLST等)



1. WGSとは？

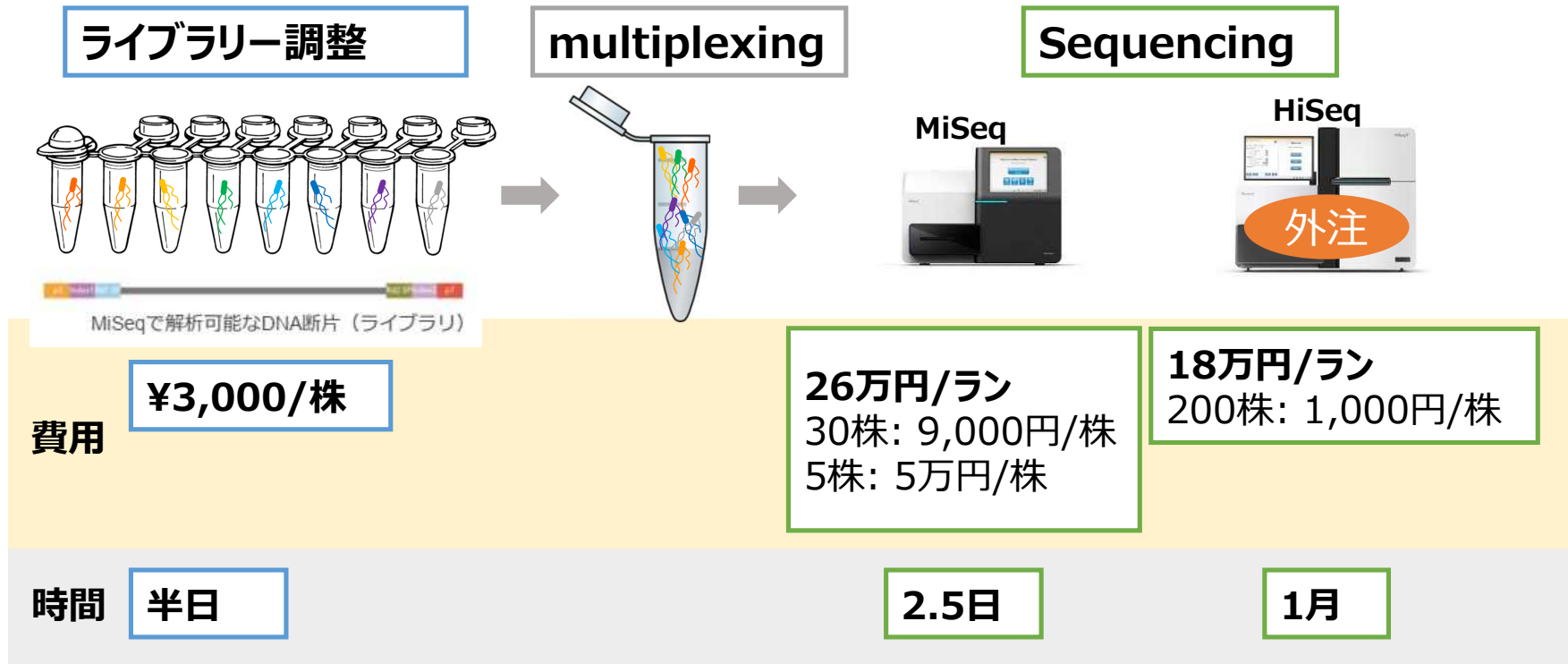
(感染研 細1で使用中的) 主なNGSシーケンサー

機器	リード長	精度	コスト (/株)	時間	目的
	-300bp	高	2万円～	60 h	SNP解析 ドラフトゲノム
 外注	-150bp	高	5,000円～	2か月 (外注のため)	RNA-seq メタゲノム
 外注	-40kb 平均: 10kb	高	10万円～	2か月 (外注のため)	完全長配列決定
	>100kb 平均: 10kb	中	2万円～	~48 h	完全長配列決定 SNP解析?

- ・**集団感染調査にはイルミナがスタンダード(今後はONTも?)**
- ・**外注やONTを用いることで小規模ラボでも実用可能**

1. WGSとは？

[追加]例. WGS取得の流れ(イルミナ)



- ・費用：1~2万円/株 (最低5,000円/株)
- ・時間：解析を含めて1週間~

1. WGSとは？

WGS取得後の解析



Linux OS+フリーソフト

おすすめ

- ゲノム研究者が使用
- コマンドライン操作が必要
- 仮想環境なら、Win/Mac PCで無料で使用可能

Galaxy

- ブラウザで操作可能な種々のツール

商用ソフト

- (BioNumerics, CLC, SeqSphere...)
- 高価...

解析サービスウェブサイト

おすすめ

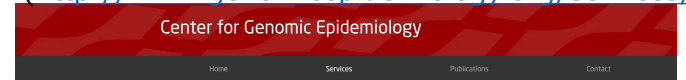
- (CGE server等)
- 基本的な解析なら十分

コマンドライン操作(慣れれば楽しいです)



CGE server

(<http://www.genomicepidemiology.org/services/>)



Overview of Services

Phenotyping	Phylogeny
SeqFinder Identification of acquired antibiotic resistance genes.	MINItree Identification of SNPs with automatic filtering, masking and site validation together with inferred phylogeny based on both long and short sequencing data.
SeqFinderEG Identification of functional metagenomic antibiotic resistance determinants.	CSIPhylogeny CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality SNPs.
LRE-Index Identification of genes and mutations leading to linezolid resistance.	NDtree NDtree constructs phylogenetic trees from Single-End or Pair-End FASTQ files.
KmerResistance Identification of acquired antibiotic resistance genes using Kmers.	Evergreen Evergreen generates a forest of constantly updated phylogenetic trees with publicly available whole-genome sequencing data from foodborne bacterial isolates that were deposited in the short sequencing read archives (NCBI SRA/ENA).
PathogenFinder Prediction of a bacteria's pathogenicity towards human hosts.	TreeViewer Phylogeny Tree Viewer.
VirulenceFinder Identification of acquired virulence genes.	
RestrictionModificationFinder Determination of Restriction-Modification sites (based on REBASE).	
SPIfinder SPIfinder identifies Salmonella Pathogenicity Islands.	

※発表者個人の見解です

1. WGSとは？

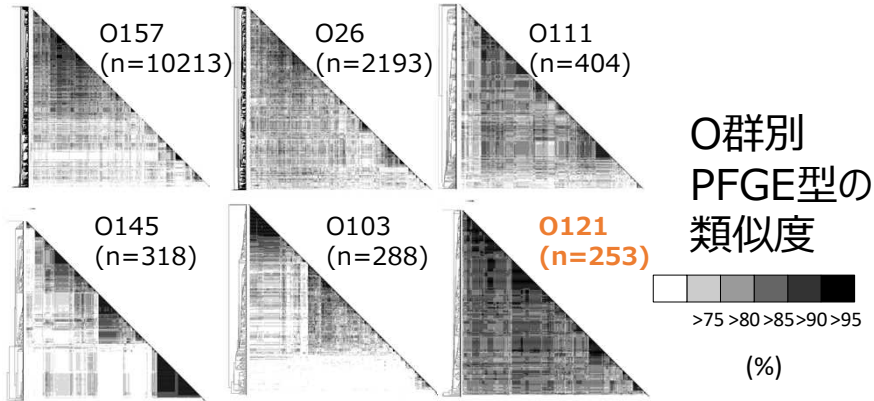
2. WGSとPFGEの比較 (0121)

3. WGSとMLVAの比較 (0157)

4. WGSによるサーベイランス (試行的)

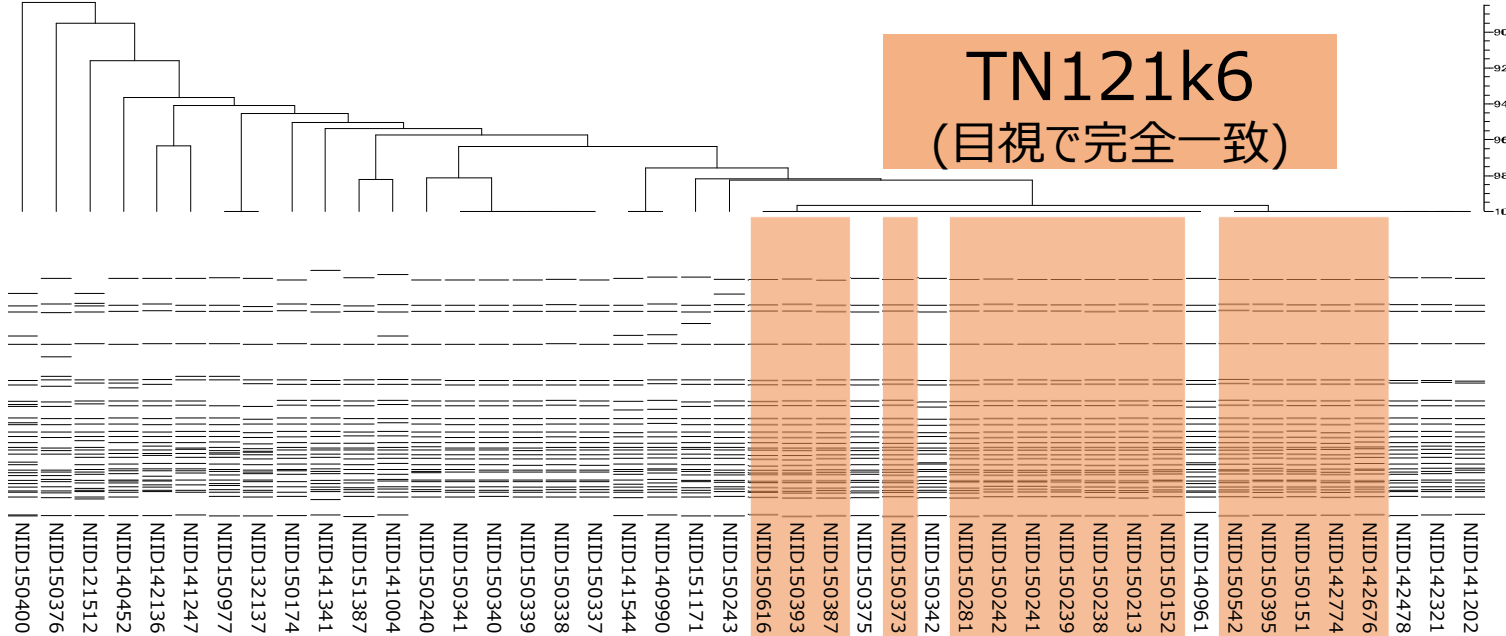
2. PFGEとWGSの比較 O121

背景・目的



- O121は、他のO群よりPFGEの多様性が低い
- 2014年12月-'15年3月に同一のPFGE型株が全国的に多数分離
- 一部で共通の食品喫食

しかし、類似株とPFGEで区別できない…



目的

より型別能が高い
散発集団感染の検出
= WGSを用いたSNP解析

方法

EHEC O121
44株のSNP解析

[追加]方法

菌株 (計44株)

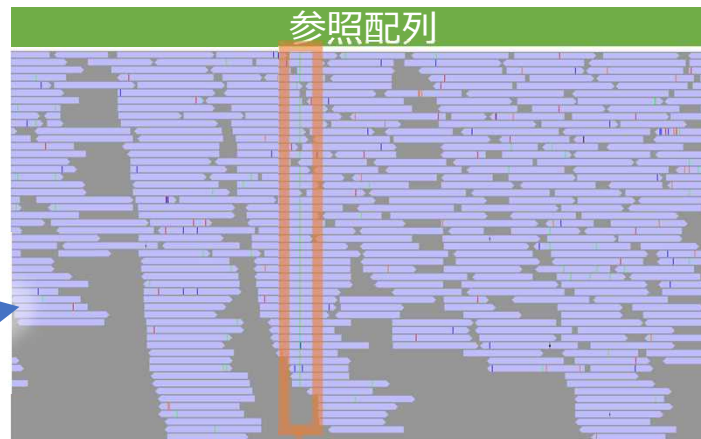
- TN121k6 : 16株
- 2012-2015年分離株: 28株

ゲノム配列取得と一塩基多型 (SNP) 抽出

- Illumina MiSeq (300 bp×2) による全ゲノム配列取得
- マッピングによるSNP抽出 (bwa, VarScanを使用)
- 参照配列 : O121 完全長配列解読株の染色体
- 抽出SNPからファージ、50 bp以上のリピート領域を除去

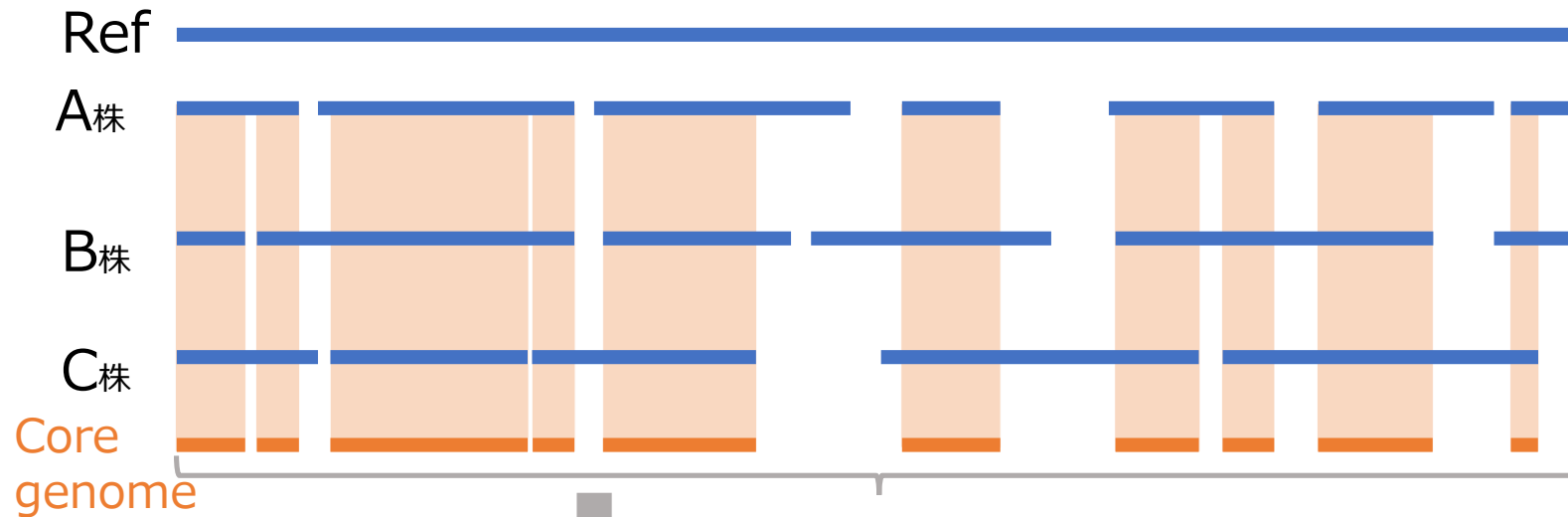
マッピングの例

ショートリード
(平均約250 bp)



[追加]方法：core genome SNP (cgSNP)抽出法

Core genome (解析する全ての株に共通する領域) の SNPをまとめる

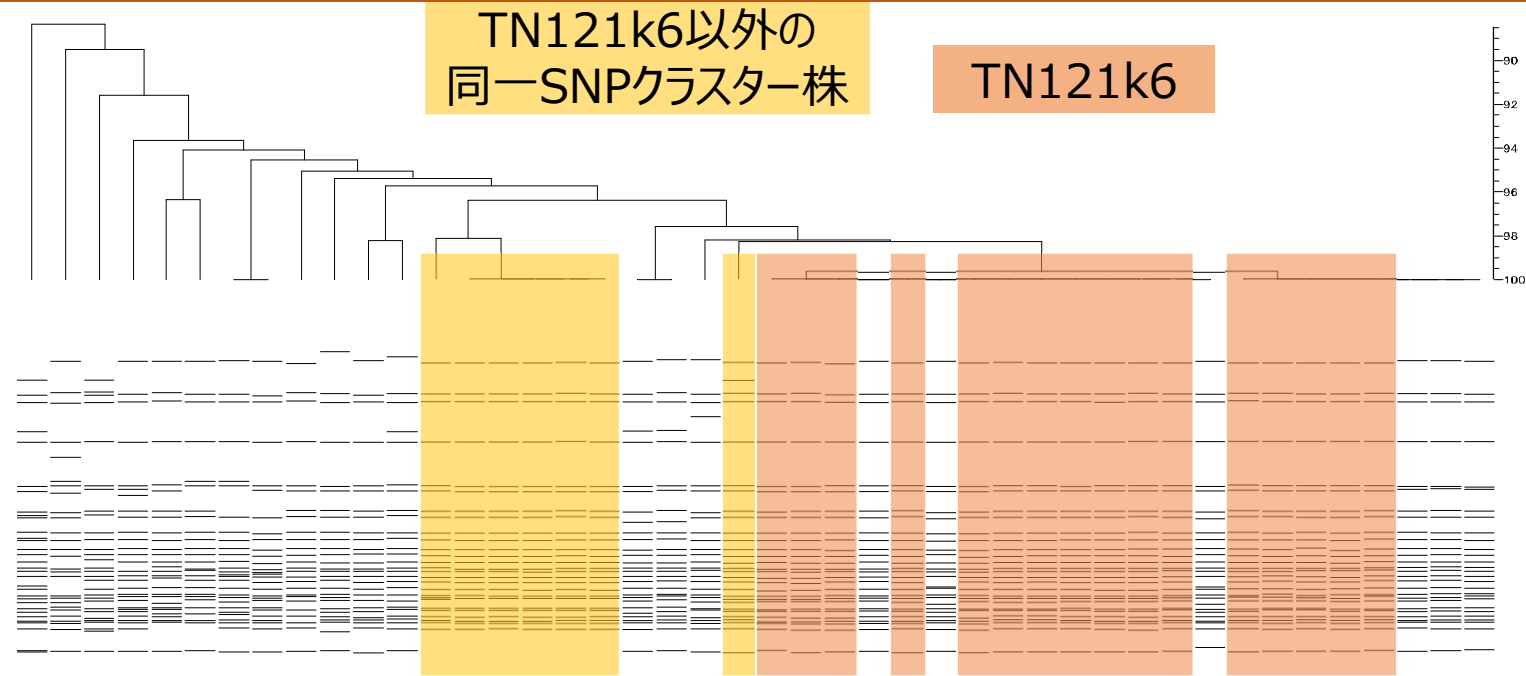
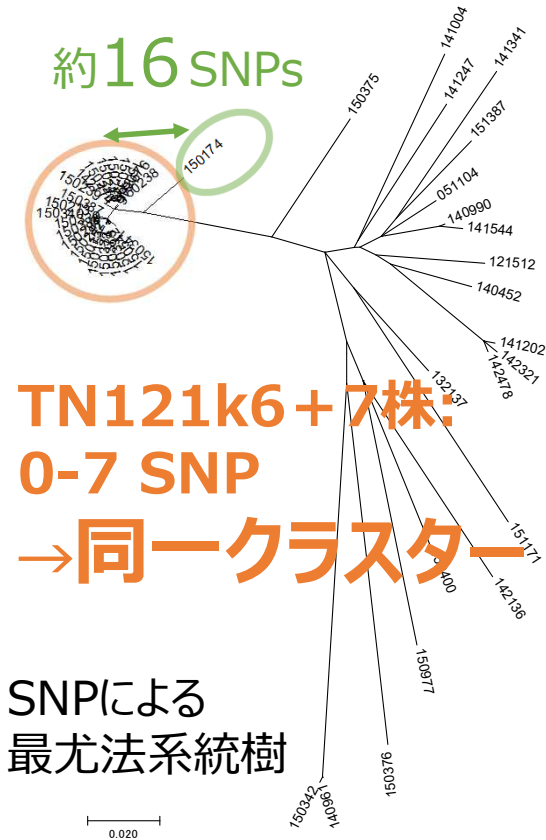


Position	Ref	A	B	C
5,514	C	C	G	G
6,065	C	T	C	C
6,456	A	C	A	A
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

- 系統解析
- pairwise SNPの算出

2. PFGEとWGSの比較 O121

PFGEとゲノム解析の比較



- TN121k6は全て同一SNPクラスターに含まれた
- PFGE型類似株から同一SNPクラスターとなる株が抽出された

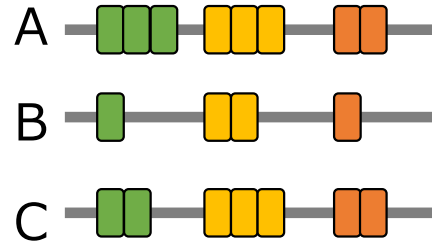
- PFGEで一致→ゲノムレベルでも一致
- PFGEで近縁→ゲノムレベルでは遠縁な株も含まれる

1. WGSとは？
2. WGSとPFGEの比較 (O121)
- 3. WGSとMLVAの比較 (O157)**
4. WGSによるサーベイランス (試行的)

O157におけるMLVA

17か所のリピート領域のパターン

...GTC **GTTGTTGTT**CGA...
...GTC **GTT**CGA...
...GTC **GTTGTT**CGA...



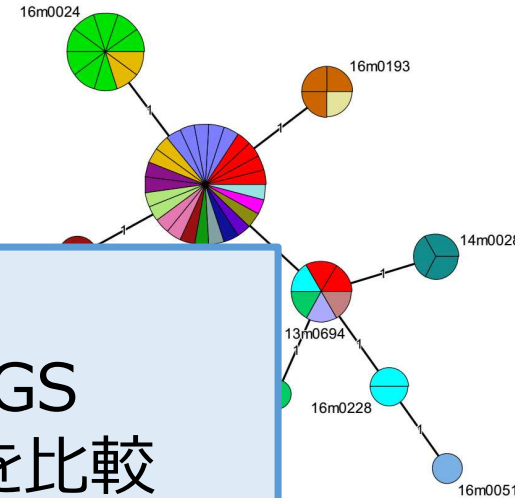
Isolate	No. repeat unit				
	Allele 1	2	3	...	17
A	11	3	9		4
B	5	3	9		2
C	5	3	7		4

現行のサーベイランス体制

- 疫学情報
- 1.5 年以内
- MLVAの差異：0-1カ所

菌株をグループ化

集団感染を検知



目的

1. WGSによるMLVAの評価
2. WGS解析手法の確立

方法

EHEC O157 369株のWGS
→SNP・MLVA・cgMLSTを比較

[追加]方法

使用菌株

EHEC O157, 369株

- ・分離年 2012-'17
- ・7種のMLVAグループ
= 主要MLVA型 + 近縁型

※SLV: single locus variant
DLV: double locus variant
TLV: triple locus variant

方法

WGS
(by MiSeq)



Analysis pipelines

1. **cgSNP解析**

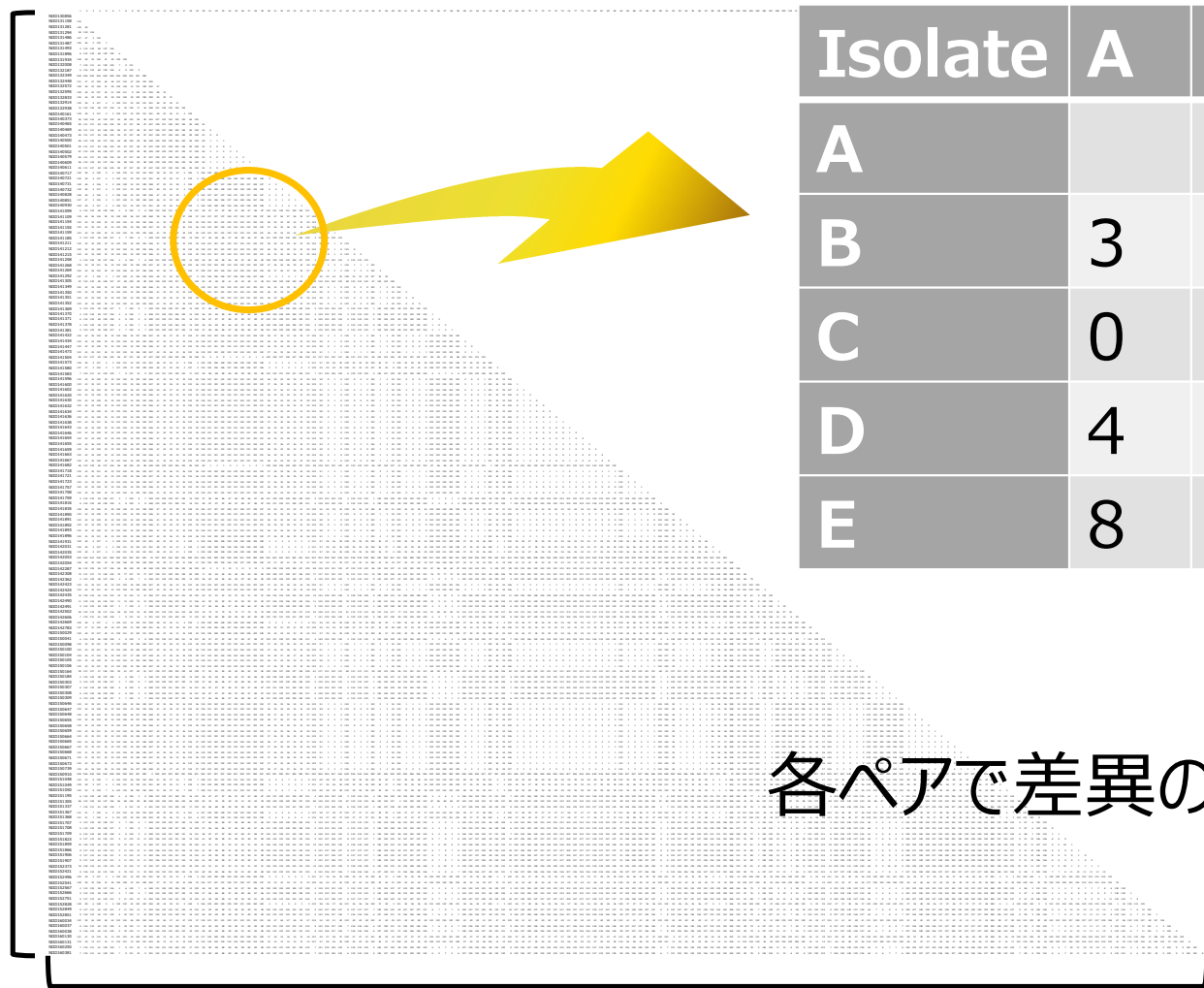
2. **MLVAとcgSNPの比較**

3. **cgMLSTとcgSNPの比較**

MLVA group	Index MLVA type	No. of isolates			
		Index	SLV	DLV	TLV
G1	13m0551	56	10	1	7
G2	15m0278	39	9	0	2
G3	16m0110	26	2	0	0
G4	17m0121	41	6	24	2
G5	13m0157	18	11	6	10
G6	13m0694	36	18	19	6
G7	14m0004	5	9	4	2

3. MLVAとWGSの比較 O157

[追加] Pairwise distanceの計算



Isolate	A	B	C	D
A				
B	3			
C	0	1		
D	4	7	8	
E	8	9	6	17

各ペアで差異の数を計算

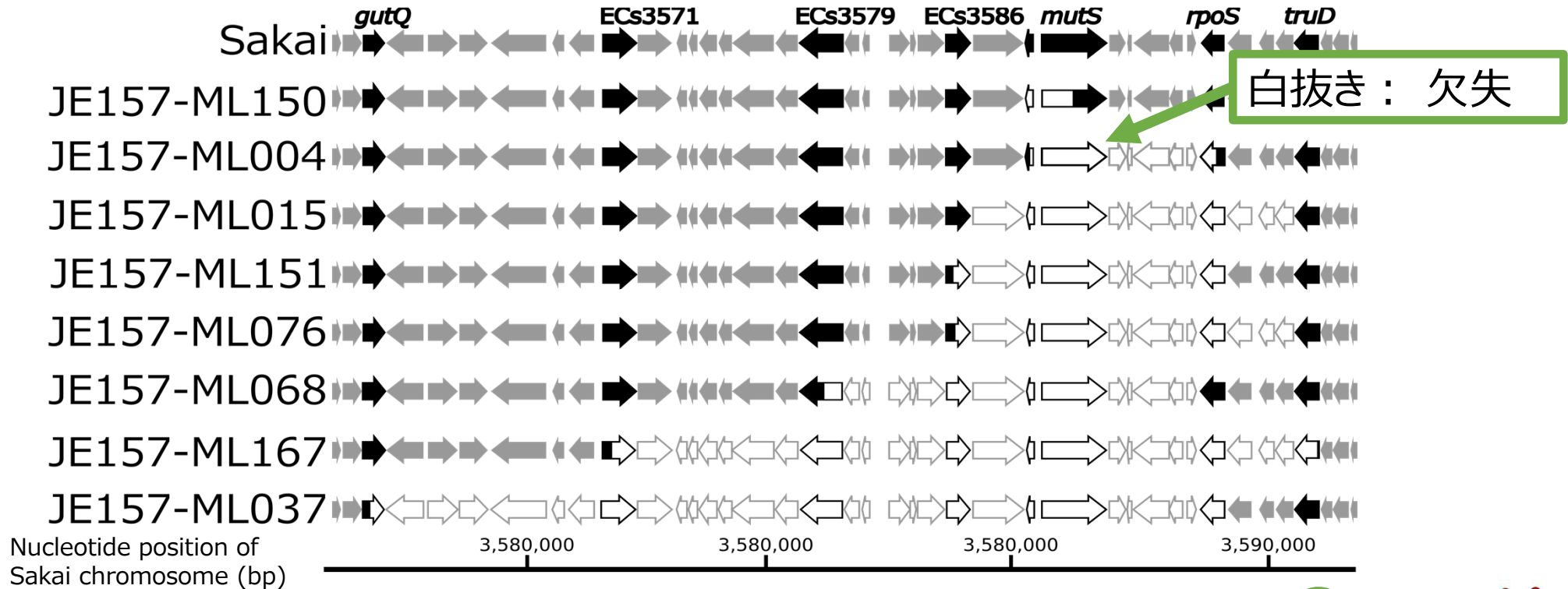
{ cgSNP
cgMLST
MLVA

369株

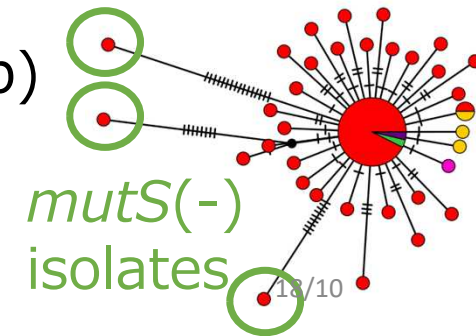
369株

3. MLVAとWGSの比較 O157

[追加] *mutS* (DNA修復タンパク質) 周辺の欠失

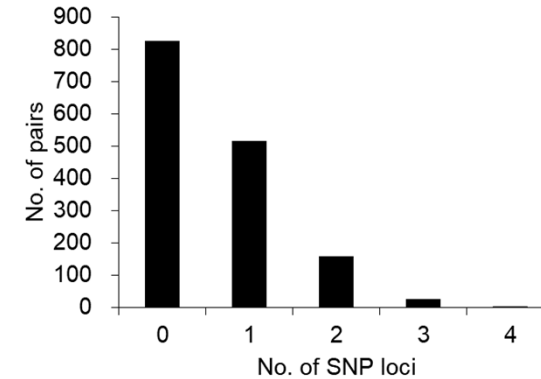


- 8/369 (2.2%) 株で*mutS*周辺の欠失 (55bp – 33.4kbp)
 - *mutS*(-) 株は疫学関連ある株間でも多数のSNP
- mutS* (&*mutL/H*) の確認が必要**



SNP解析（同一由来だと何カ所のSNP？）

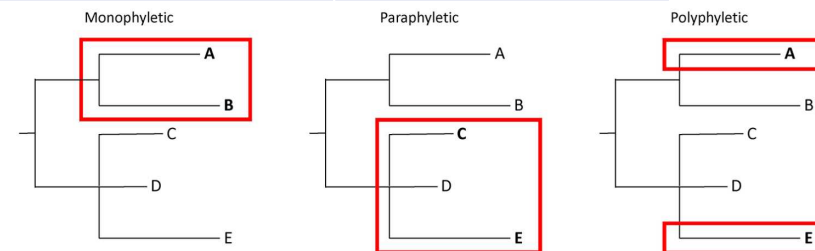
疫学リンクあり(同一感染源・二次感染)の株間SNP
 →4か所以内
 →先行研究でも大部分は**10か所以内のSNPs**
 ※大腸菌の場合



WGS解析時の菌株間関連性の基準(*E. coli*, *Salmonella*, *Listeria*)

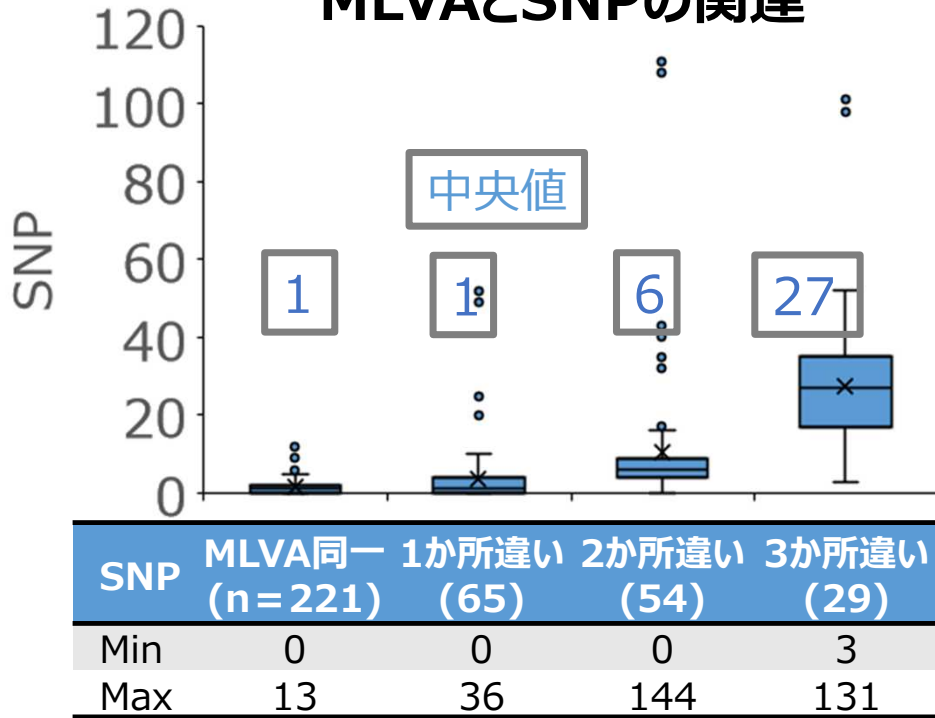
	Supports	Neutral	Does not support
SNP distance	< 21	21-100	> 100
Bootstrap support	> 0.89	0.80-0.89	< 0.80
Tree topology	Monophyletic	Paraphyletic	Polyphyletic

Pightling (FDA) et al. Front Microbiol. 2018など
 “WGSの適切な解釈には、
 SNP距離、ブートストラップ、系統樹、
 疫学情報、さかのぼり情報を
 総合的に考慮することが必要”



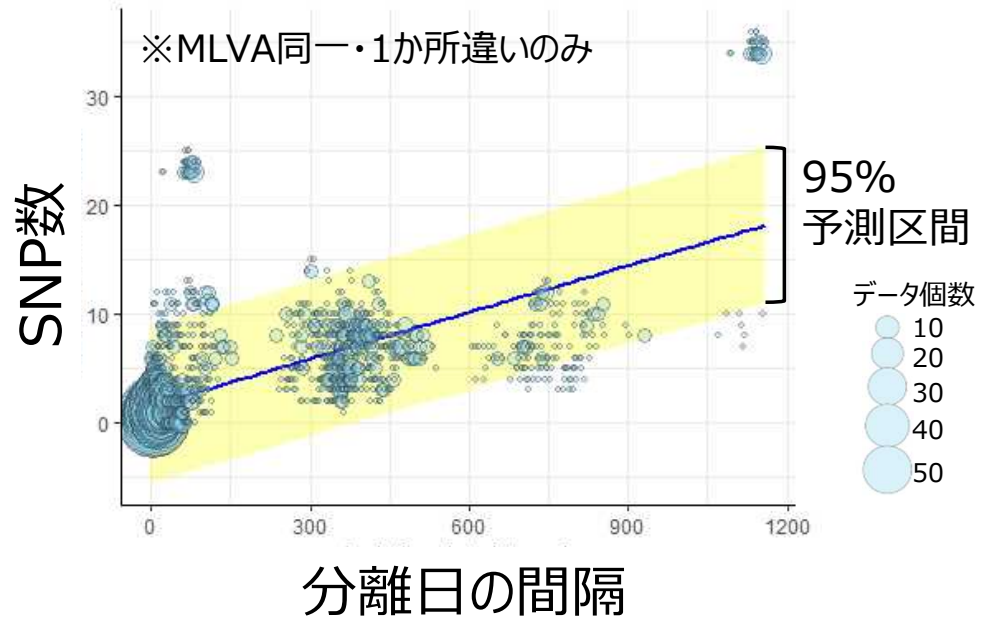
cgSNPとMLVAとの比較

MLVAとSNPの関連



- MLVAが同一/1か所違いでは遺伝的に近い
- 2か所以上異なる場合には、遺伝的に近い株と遠い株が含まれる

分離日が離れた株間のSNP



- MLVAが同一/1か所違いでも時間が経つとSNPが蓄積する

SNPとcgMLSTとの比較

SNP: (時間がかかる、コンピュータ負荷が高い)

cgMLST:

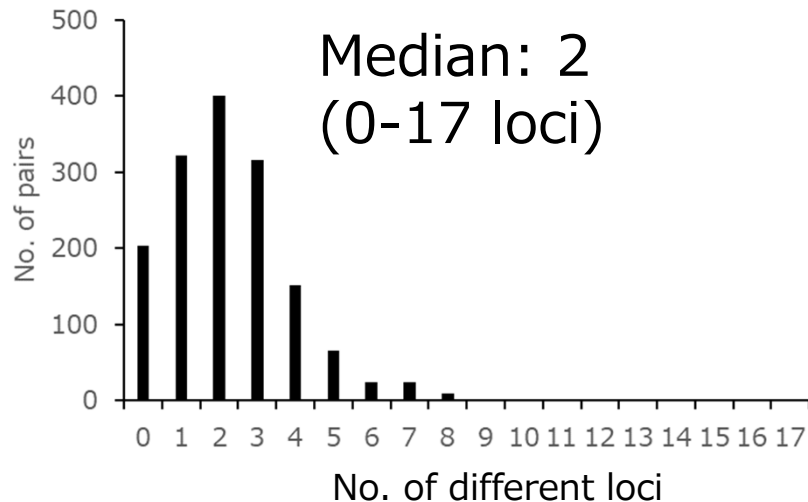
速い

コンピュータ負荷低い

DBとの比較容易

- ・保存性の高い2,513遺伝子のMLST(大腸菌)

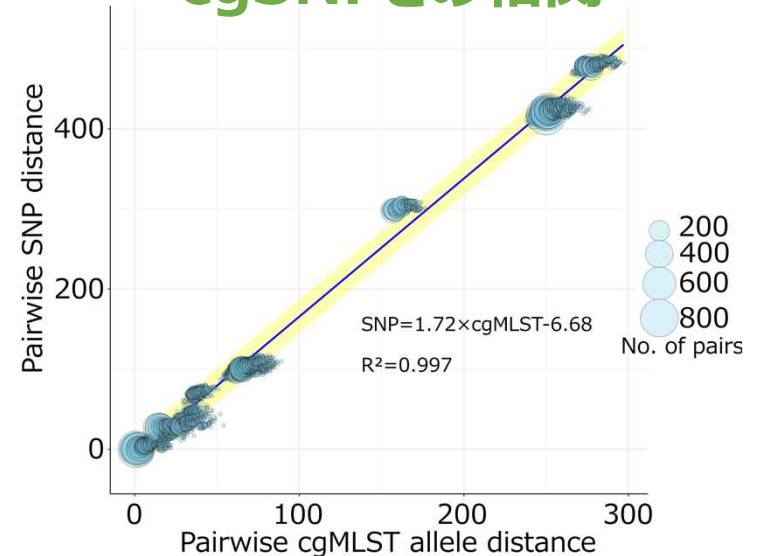
疫学関連ある株間での差異のヒストグラム



※擬陽性 (挿入・欠失など)の除去
→99.9%は差異無し

解像度はcgSNPより低い

cgSNPとの相関



高度に相関(R²=0.99)

3. MLVAとWGSの比較 O157

小括

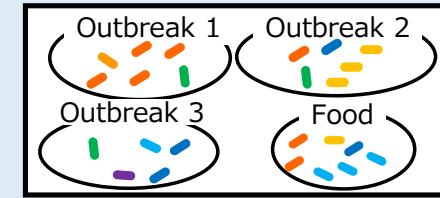
MLVAの違い	SNP distance	
	~1年	1年~
0-1か所	Clonal (< 10SNP)	Variable
2か所	Variable	Variable
3か所	Variable	Variable

MLVAは短期間の調査用には高い型別能
(迅速・安価・他検体処理などの利点)

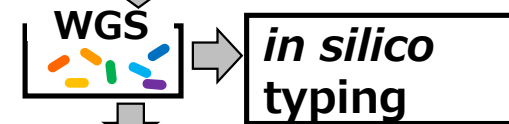
迅速・安価なサーベイランス :
Screening by MLVA
Detailed analyses by WGS

MLVA

- ・スクリーニング
- ・集団感染調査



代表的なMLVA型
→WGS



cgMLST

大規模比較

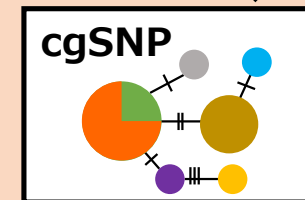


cgMLST comparison

近縁株のみで解析

SNP

詳細な
系統解析



Example of surveillance with MLVA & WGS

Lee et al.2019. *Appl Environ Microbiol*

1. WGSとは？
2. WGSとPFGEの比較 (O121)
3. WGSとMLVAの比較 (O157)
- 4. WGSによるサーベイランス (試行的)**

4. WGS用いたサーベイランス 対象菌株・目的

2012～'19年のMLVA対象EHEC菌株数

Year	All		O157		O26		O111		O121		O103		O145		O165		O91		
	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	No.	MLVA	
2012	22	18	21	17	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2013	2,057	958	1,475	695	508	216	74	47	0	0	0	0	0	0	0	0	0	0	0
2014	2,549	903	1,792	646	633	213	124	45	0	0	0	0	0	0	0	0	0	0	0
2015	2,317	860	1,548	573	683	234	81	51	0	0	5	2	0	0	0	0	0	0	0
2016	2,651	972	1,653	586	693	221	77	44	51	28	111	48	37	16	8	8	21	21	
2017	2,767	1,031	1,637	597	686	246	130	52	77	33	145	46	46	19	8	8	38	30	
2018	2,786	1,105	1,674	679	645	231	88	55	152	46	141	53	48	27	4	4	34	30	
2019	2,105	886	1,252	524	386	163	107	59	67	35	180	55	83	21	4	4	26	25	

2020年～

MLVA対象O群 : 全MLVA型

その他O群 (PFGE対象) : 全株

} 試験的にWGS解読

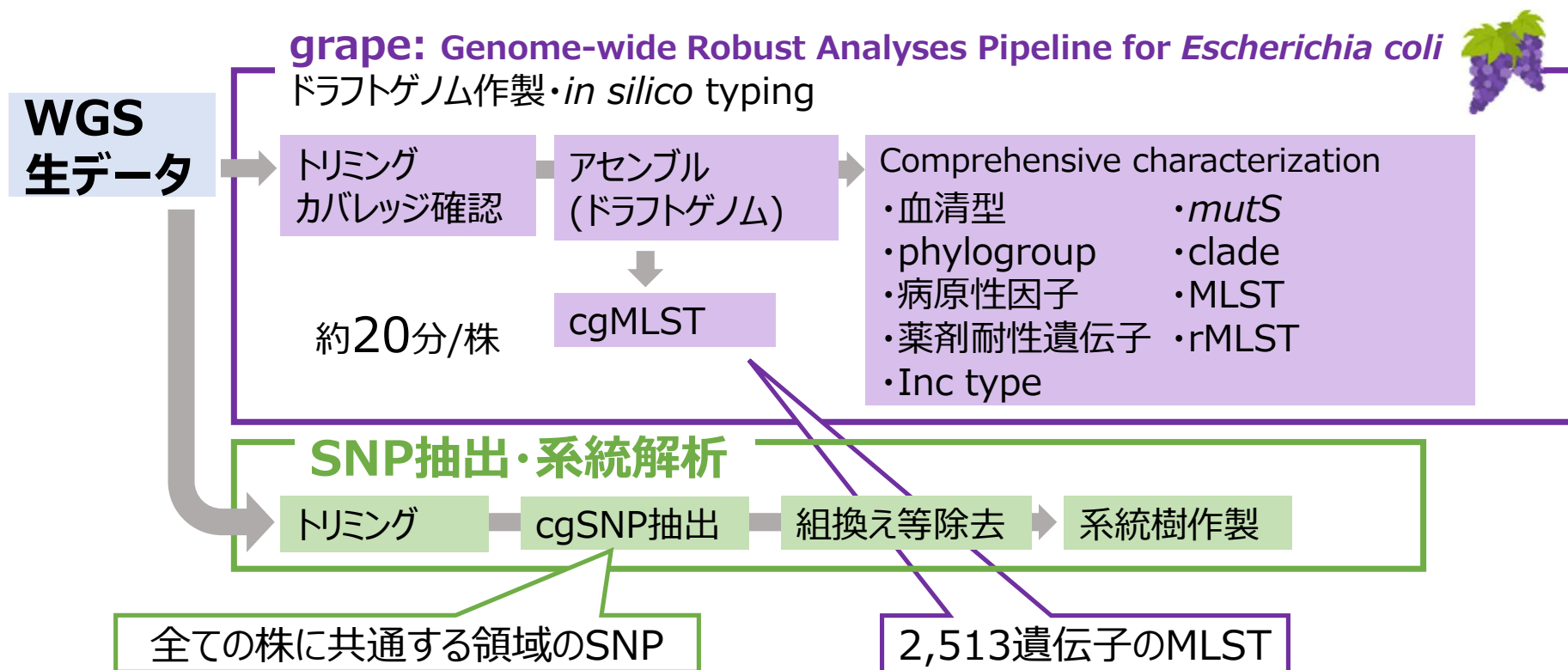
目的

- ・集団感染/散発株の関連性
- ・流行系統
- ・O157以外でのMLVAの評価

4. WGS用いたサーベイランス

WGS解析パイプラインの確立

ほぼ自動で基本的な解析（血清型、保有遺伝子、系統樹）ができる体制を構築



4. WGS用いたサーベイランス

WGSデータベース化と迅速解析法の確立



集団感染株のWGS

**cgMLST
データベース**
=大規模解析向き

国内EHEC
約5,000株

公共データベース
(主に海外株)
約19万件

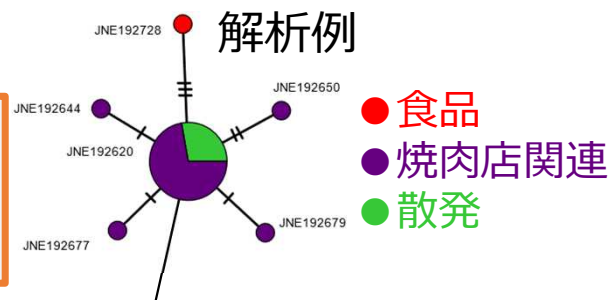
・国内EHECのcgMLST・
cgSNPのデータベース化

**cgSNP
データベース**
=詳細解析向き

国内EHEC
約5,000株

・新たなゲノム情報が得られた
際に、迅速にデータベースから
近縁株を抽出する手法を確立

国内・海外株データベースから
迅速に (4日~)
最も近縁な株を抽出



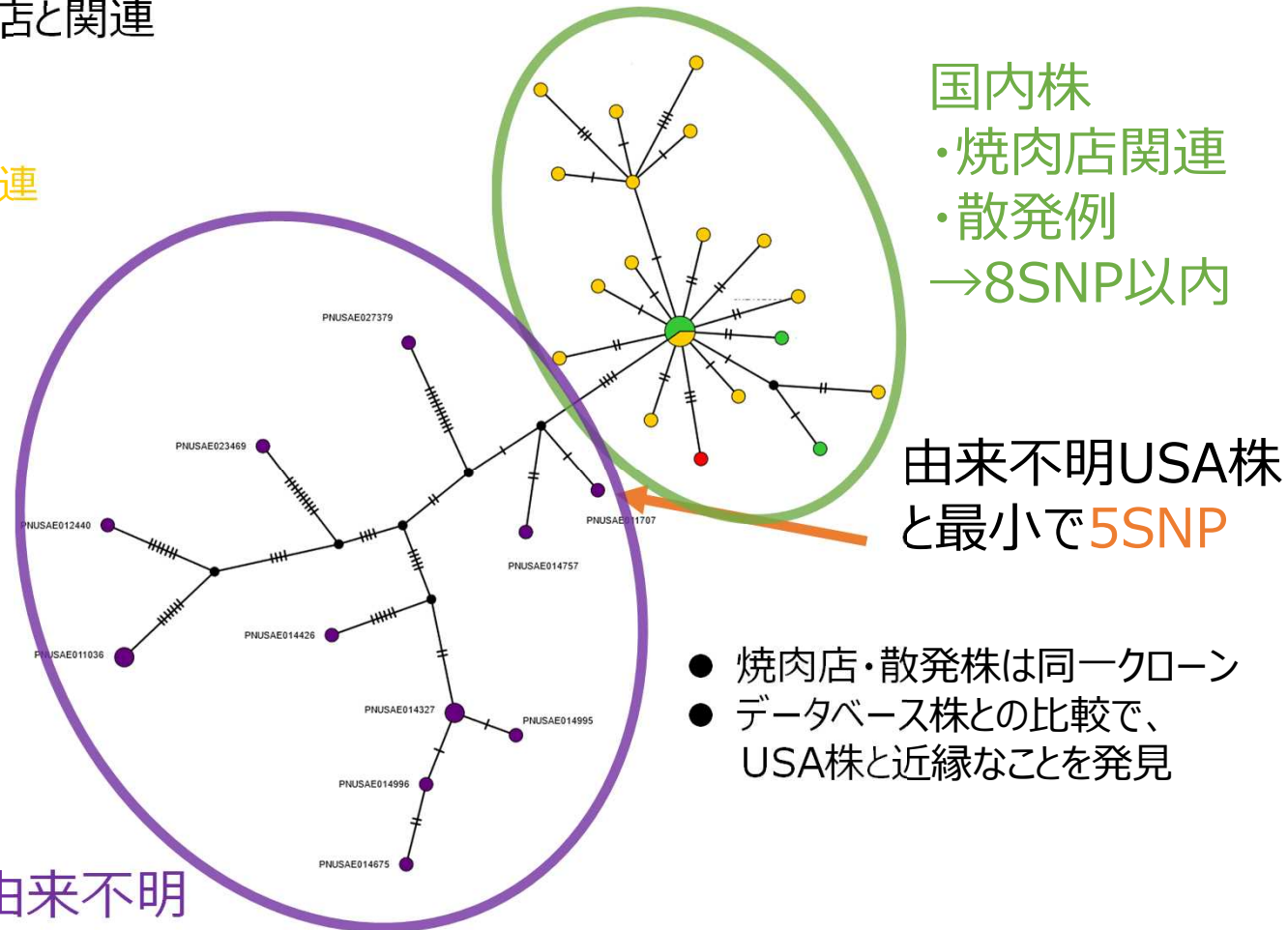
その他国内・海外株とは40SNP以上異なる

4. WGS用いたサーベイランス 実際の事例解析：2019年焼肉店食中毒と海外株

【背景】

- 2019年11月から全国的に類似したMLVAが分離一部は、焼肉チェーン店と関連

- 焼肉店関連
- 食品
- 散発
- アメリカ株



総括

PFGEとの比較 (O121)

- PFGE近縁株には、ゲノムレベルで離れた株も含まれる。

MLVAとの比較 (O157)

- 集団感染調査には十分高い型別能。
- MLVAで2カ所以上離れた株、分離日が離れた株では、遠縁な株も含まれる。

※いずれもO群によって結果は異なる。

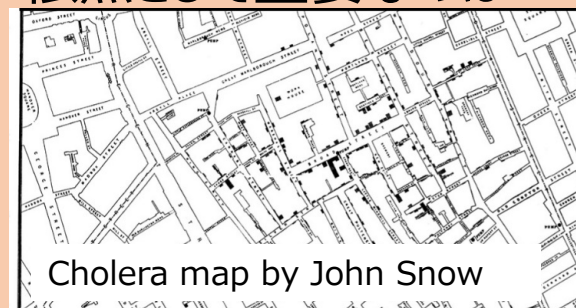
WGSを用いたサーベイランス

- cgMLST, cgSNPを用いることで国内外のデータとの迅速な比較が可能。

WGSサーベイランスの課題

- ・データ、解析法の共有
- ・コスト、解析労力(外注、フリーサービスで十分な場合も)
- ・遡り調査、宿主(動物等)調査の不足

依然として重要なのは…



疫学調査 / 病原体の分離

謝辞

感染研 細菌第一部

大西 真

明田幸宏

伊豫田 淳

泉谷秀昌

石原朋子

感染研 病原体ゲノム解析

研究センター

黒田 誠

関塚剛史

九州大学大学院 医学研究院

細菌学分野

林 哲也

小椋義俊

地方自治体 衛生研究所

保健所等の皆様

(敬称略)