

#### USING WGS DATA FOR OUTBREAK DETECTION

#### cgMLST vs SNPs

### Senior scientist HENRIK HASMAN Statens Serum Institut (SSI) Denmark



: MLST

- cgMLST / SNP (Core/Whole Genome Comparison)
- Presence/absence of genes and mobile elements
- .....often a combination of the above is used to study outbreaks.





- Both SNP and cgMLST can be used to study clonal relationships.
  Sometimes a combination gives a "second view" when in doubt
- They both have advantages and disadvantages...and potential pitfalls
- Genetic distances (SNPs or alleles) to define an outbreak will vary depending on 1) the organism, 2) the outbreak (length ect) and 3) how the method is applied (references, inclusion criteria, pruning ect)
- Only perform a SNP analysis on isolates, which are relatively closely related (less than 500-1000 SNPs as a rule of thumbs)
- Using raw data takes longer time, but allows to adjust for ambiguous calls

# Clone theory 101

Textbook – A clone is:

"a group of genotypic identical isolates descending from a common ancestor as part of a direct chain of replication"

A more realistic definition:

"the word clone will be used to denote bacterial cultures isolated independently from different sources, in different locations, and perhaps at different times, but showing so many identical phenotypic and genotypic traits that the most likely explanation for this identity is a common origin" (Ørskov & Ørskov, 1983)





- Randomly generated across the chromosome over time ("The mutation rate")
- ...but influenced by external factors...

# Horizontal gene transfer

The Chromosomographs evil nemesis



- Horizontal gene transfer circumvents the linearity of the evolutionary clock
- ...and needs to be addressed in any whole genome analysis such as SNPs...

## Advanced clone theory Clonal drift



- The more discriminatory a typing method is, the more difficult it will be for it to accommodate *biological variation* caused by clonal drift over time (stability issues).
- On top of this, all typing methods will add *methodological variation* (repeatability and reproducibility issues) thus blurring the picture even more.

### SINGLE SOURCE OUTBREAKS





Single source Short time span "Contaminated dish" "Single infected patient"

Single source – local spread Long time span "Hospital or regional outbreak"

0-2

0-2

0-2

0-2

### **COMPLICATED OUTBREAKS**





Single source Long time span "Contaminated processing plant / industry"

"Long-term colonized patient / healthcare worker"

International sciences Long time al clones "Importernations source" "Traventernation source"

### PO = POSSIBLE OUTBREAKS (E. COLI)





#### **Tentative definition of possible outbreak (PO)**

If two isolates have a SNP distance  $\leq 10$  (termed PO<sub>10</sub>), they are considered to be so genetically related that they may be part of the same outbreak.

# Phylogenetic analysis

Core genome MLST (cgMLST) vs Single Nucleotide Polymorphism (SNP)





- Reference based gene-by-gene comparison
- Super MLST"
- Increased number of genes  $\rightarrow$  Increased discriminatory power
- Requires curated and validated schemes
- Requires software to remove gene homologues if you want to build your own scheme.





#### cgMLST.org Nomenclature Server

This server controls the allelic nomenclature of core genome MLST (cgMLST) bacterial gene schemes. Currently submission of new alleles and optional metadata is only possible by use of the SeqSphere+ software. A cgMLST scheme is a fixed and agreed upon number of genes for each species or group of closely related species that is ideally suited to standardize whole genome sequencing (WGS) based bacterial genotyping. By cgMLST very closely related genomes are 'lumped' together in a **Complex Type** (CT). In addition, this server controls the allelic nomenclature of the **accessory genes** of the species seed genomes.

We care about your privacy. Read our privacy policy.

Scheme	Target Count	Strain Count
Acinetobacter baumannii cgMLST	2,390	8,258
Bacillus anthracis cgMLST	3,803	209
Brucella melitensis cgMLST	2,704	89
Brucella spp. cgMLST	1,764	1
Burkholderia mallei (FLI) cgMLST	2,838	1
Burkholderia mallei (RKI) cgMLST	3,328	13
Burkholderia pseudomallei cgMLST	4,221	21
Campylobacter jejuni/coli cgMLST	637	4,643
Clostridioides difficile cgMLST	2,147	1,621
Clostridium perfringens cgMLST	1,431	99
Enterococcus faecalis cgMLST	1,972	3,743
Enterococcus faecium cgMLST	1,423	17,491
Escherichia coli cgMLST	2,513	13,983

### SEQSPHERE+ SOFTWARE





#### Available schemes:

S. aureus – E. coli – E. faecium – A. baumannii – K. pneumoniae ...and more

### CORE GENOME MLST (cgMLST)

STATENS	_HHH
SERUM	
INSTITUT	

888	582								Con	npari	ison Ta	able:	C_fr	euno	dii_S	T18 [	unsto	ored]																_ 0	×			
File Edit	Data C	olumns Analysis Tools																																				
8	44	🗙 🎟 🛱   😵 🔋 🔮 🙀 ≰	- E	€0			<b>≣</b>   .	<b>L</b>	🧧 4D	GTG GRG																												
#Missing v	Perc 🕅	A Sample ID Sample	ST c.	E C		C Ci	IZIIL SoS	a La	C Pri	ofile freundii MLS	т	asp c.t	dp)f	fad mdl	arıdı c.c	h lysf		AF AF	AF A c.c	FAF 4	AF AF	AF AF	AF AF c.c.	AF AF	AF AL	FAFA	FAF	AF AF c.c.	AF AF	AF A c. c	FAF/	AF AF	AF AF	AF /	F AR322_RS23	AF c.	AR322_RS027 C. freundii Acce	AF A
84	100.0	AMA332	18	??	? ?	?	??	?	? 10	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	? (not found)	1 1
75	100.0	CPO20160068_CLCc	18	??	? 7	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
136	98.2	GCF_001880795_1_ASM188079v1_g	18	??	? 7	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
135	98.2	GCF_001880955_1_ASM188095v1_g	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	. 1	1 1	1 1	1	1	1	1 1
126	98.4	CPO20170015_contigs	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1 :	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	2	1 1
119	99.0	CPO20150022	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	. 1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
105	99.2	AMA393	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	55	14	1 1	. 1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
98	99.3	AMA653	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	. 1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	. 1	1 1	1 1	1	1	1	1 1
96	99.5	AMA528	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	? (not found)	1 1
91	99.5	AMA533	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
93	99.5	AMA535	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	? (not found)	1 1
91	99.6	AMA426	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	? (not found)	1	1	1 1
88	99.7	AMA463	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 2
82	99.7	AMA570	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	. 1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	. 1	1 1	1 1	1	1	1	1 1
85	99.7	CPO20150067_CLC	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
80	99.8	CPO20150038	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	2 1	1	1 1	1 1	1	1	1	1 1
81	99.8	CPO20160021	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	55	14	1 1	. 1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 2
79	99.9	AMA1576_CLCc	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
78	99.9	CPO20150079_CLC	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
75	99.9	CPO20160010	18	??	??	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1	1 1	1 1	1	1	1	1 1
75	99.9	CPO20160011	18	??	? ?	?	??	?	? 10,	14, 17, 1	2, 5, 5, 14	10	14 1	7 12 5	5 5	14	1 1	1	1 1	1 1	1 1	. 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	. 1	1 1	1 1	1	1	1	1 1

- All isolates are assigned to specific Complex Types (CTs)
- Different cgMLST schemes use different cut-off values for new CTs



Ridom – SeqSphere+

#### CPO IN DENMARK – E. COLI ST410





### CORE GENOME MLST (cgMLST)









Center for Genomic Epidemiology											
	Home	Services	Publications	Contact							
<b>cgMLS</b> Service	TFinder 1.2	2 Output Article abstract Citations									
Software Database	version: 1.0.1 (202 e: Available here	1-08-29)									
Select sp Campylo	<b>becies</b> bbacter (PubMLST)	~									

#### Input file(s)

fastq and fasta formats are supported both as plain text and gzipped files. Data from several isolates can be uploaded together.

Choose File(s)



# But what if my favorite organism is not in the cgMLST.org list?

# ....or if I want to make a lineage-specific scheme?

#### Core Genome MLST (cgMLST) А В С D Е F G Н 100% 81% 74% 66% 78% 85% 59% 97%







- Gene homologues (in reference) removed.
- Intergenic regions are not included in analysis.
- Each gene variant is assigned to an allele number.
- Robustness is ensured by including a diverse dataset covering the genetic variation of the whole species.
- Validation using defined outbreak isolates with known epidemiology is required to determine CT-groups.

## Core Genome MLST (cgMLST)



## Core Genome MLST (cgMLST)



## Core Genome MLST (cgMLST)

#### Main advantages

- Common nomenclature (Cluster types)
- Fixed set of reference genes
- Recombination has been filtered out
- Curated database
- Fast, as it runs on draft assemblies

#### Main disadvantages

- Requires a validated cgMLST scheme
- May be sensitive to assembly method
- Requires a curator to manage the database
- The discriminatory power may be a bit lower than for SNP analysis
- Have a tendency to drift over time especially in long-lasting outbreaks

## **SNP** analysis

### Core, auxiliary and pan genomes

I		L J
	γ	
	Core	utiliat.
	genor den	Joenn
	ine ome	me

# SNP analysis practical considerations

- Choosing the best reference
- Global SNP vs HQ SNP analysis
- Detecting contamination
- Recombination events

## Choosing the best reference



Most of the reference covered by isolate genomes

 $S1 \rightarrow REF1 = 1 SNP$ 

S2  $\rightarrow$  REF1 = 3 SNP

 $S1 \rightarrow S2 = 2 SNP$ 

## Choosing the best reference



### Choosing only closely related isolates



## Choosing the best reference



- In general, a closely related reference is desired
- A best match in NCBI RefSeq can be searched using KmerFinder
- Complete genomes can also be searched at NCBI (but is not easy to use..)
- A draft genome of the index isolate can be considered to be used
- Or you can make your own complete genome by using MinION or PacBIO

### SNP analysis practical considerations

- Finding the best reference
- Global SNP vs HQ SNP analysis
- Detecting intraspecies contamination
- Recombination events



- "Draft assembly" to give a "Draft genome"
- Different assemblers may give different results  $\rightarrow$  cgMLST/SNP variation



Assembled genomes will not pick up contamination issues



And any QC method should be able to pick up an increased level of hetogeneous positions and warn of possible contamination

#### cgMLST vs SNP А В С D Е F G A В С Е F G D 100% 100% 100% 100% 100% 99% 100%

cgMLST  $\rightarrow$  1 allele difference SNP  $\rightarrow$  2 allele difference



# SNP analysis practical considerations

- Finding the best reference
- Global SNP vs HQ SNP analysis
- Detecting contamination
- Recombination events

## **Recombination events**

- Horizontal gene transfer
- Repetitive elements (IS-elements, AMR genes ect..)
- Gene duplication and diversification

Can to some extend be removed by using bioinformatic tools such as GUBBINS or by ignoring SNPs that at "close" to each other (called *pruning*).



352

A.C. Schürch et al. / Clinical Microbiology and Infection 24 (2018) 350-354

#### Table 1

Examples of relatedness criteria for wg/cgMLST and SNP typing schemes of representative clinically relevant bacteria

Organism	Relatedness threshold <sup>a</sup>		References
	wg/cgMLST (allele) SNPs		
Acinetobacter baumannii	≤8	≤3	[25,26]
Brucella spp.	Epidemiologic validation in progress <sup>b</sup>		http://www.applied-maths.com/applications/wgmlst
Campylobacter coli, C. jejuni	≤14	≤15	[27,28]
Cronobacter spp.	Epidemiologic validation in progress <sup>b</sup>		http://www.applied-maths.com/applications/wgmlst
Clostridium difficile	Epidemiologic validation in progress <sup>b</sup>	≤4	[29], http://www.cgmlst.org/ncs, http://www.applied-
			maths.com/applications/wgmlst
Enterococcus faecium	≤20	≤16	[30]
Enterococcus raffinosus	Epidemiologic validation in progress <sup>b</sup>		http://www.applied-maths.com/applications/wgmlst
Escherichia coli	≤10	≤10	[31,32], https://enterobase.warwick.ac.uk/
Francisella tularensis	<u>si</u>	≤2	[33,34]
Klebsiella oxytoca	Epidemiologic validation in progress <sup>b</sup>		http://www.applied-maths.com/applications/wgmlst
Klebsiella pneumonia	≤10	≤18	[35,36]
Legionella pneumophila	≤4	≤15	[37]
Listeria monocytogenes	≤10	≤3	[38,39]
Mycobacterium abscessus		≤30	[40]
Mycobacterium tuberculosis	≤12	≤12	[41]
Neisseria gonorrhoeae	Epidemiologic validation in progress <sup>b</sup>	≤14	[42], http://www.applied-maths.com/applications/wgmlst
Neisseria meningitidis	Epidemiologic validation in progress <sup>b</sup>		http://www.cgmlst.org/ncs
Pseudomonas aeruginosa	≤14	≤37	[31,43]
Salmonella dublin	Epidemiologic validation in progress <sup>b</sup>	≤13	[44], https://enterobase.warwick.ac.uk/
Salmonella enterica	Epidemiologic validation in progress <sup>b</sup>	≤4	[45], http://www.cgmlst.org/ncs, http://www.applied-
			maths.com/applications/wgmlst, https://enterobase.warwick.ac.uk/
Salmonella typhimurium	Epidemiologic validation in progress <sup>b</sup>	≤2	[46], https://enterobase.warwick.ac.uk/
Staphylococcus aureus	≤24	≤15	[47,48]
Streptococcus suis		≤21	[49]
Vibrio parahaemolyticus	≤10		[50]
Yersinia spp.	0		[51]

### WHEN NATURE IS F.... WITH YOU. HYPERMUTATORS





### WHEN NATURE IS F.... WITH YOU. HYPERMUTATORS



#### Targets of Distance Columns (CPO C. freundii ST18)

tight-click on the allele type columns to jump to the according contig position in the Sample

3 🔟

arget	Begin	End	GenBank gene	GenBank product	GenBank note	GenBank protein_id	200117_A19 /	AMA003417	AMA003565	CPO20190159	AMA00338
.322_RS02285	465,622	467,868		phosphoenolpyruvateprotein phosphotransferase PtsP	member of a	WP_003033984.1	? (failed)	1	1	1	1
.322_RS03425	716,674	721,020	8. 17	autotransporter domain-containing protein	Derived by a	WP_071684359.1	? (failed)	1	1	1	? (not found
322_RS03765	802,371	803,735		PTS sugar transporter subunit IIC	Derived by a	WP_054528657.1	? (failed)	1	? (not found)	1	1
.322_RS04195	912,678	914,693		tRNA(Met) cytidine acetyltransferase TmcA	cetylates the	WP_054528641.1	? (failed)	1	1	1	1
.322_RS05765	1,252,228	1,254,714		fimbrial assembly protein	Derived by a	WP_054528576.1	? (failed)	1	1	1	1
.322_RS06635	1,433,991	1,435,370	55. 17	cobyrinic acid a,c-diamide synthase	Derived by a	WP_044701540.1	? (failed)	1	1	1	1
.322_RS06975	1,490,987	1,491,643		DNA-binding response regulator	Derived by a	WP_003030486.1	? (failed)	1	1	1	1
.322_RS09275	1,966,895	1,967,473		TetR family transcriptional regulator	Derived by a	WP_046670695.1	? (failed)	1	1	1	1
.322_RS09820	2,083,766	2,084,500		DNA-binding response regulator	Derived by a	WP_003836390.1	? (failed)	1	1	1	1
.322_RS11920	2,514,178	2,514,801		DSBA oxidoreductase	Derived by a	WP_003035975.1	? (failed)	1	1	1	1
.322_RS12760	2,702,258	2,703,679	0	2-oxoglutarate/malate translocator	Derived by a	WP_003837022.1	? (failed)	1	1	1	1
.322_RS13805	2,920,181	2,921,314		LPS O-antigen length regulator	Derived by a	WP_054528176.1	? (failed)	1	1	1	1
.322_RS14935	3,176,896	3,177,909	0	4-hydroxy-2-oxovalerate aldolase	Derived by a	WP_003021379.1	? (failed)	1	1	1	1
.322_RS15475	3,301,194	3,301,901		flagellar basal body L-ring protein	Derived by a	WP_042270212.1	? (failed)	1	1	1	1
.322_RS15765	3,357,582	3,359,564		type IV secretion protein Rhs	Derived by a	WP_072143931.1	? (failed)	1	1	1	1
.322_RS19670	4,225,537	4,226,988		potassium transporter	Derived by a	WP_003017848.1	? (failed)	1	1	1	1
.322_RS21515	4,624,296	4,625,513	0	MFS transporter	Derived by a	WP_054528867.1	? (failed)	1	1	1	1
.322_RS07405	1,575,702	1,576,793		enterohemolysin	Derived by a	WP_054528497.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS07605	1,603,957	1,604,262		hypothetical protein	Derived by a	WP_057101149.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS08700	1,830,199	1,830,762		hypothetical protein	Derived by a	WP_003843940.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS17560	3,773,522	3,773,764		transcriptional regulator	Qin prophag	WP_003839576.1	? (not found)	1	? (not found)	? (not found)	1
.322_RS22180	4,766,146	4,767,330		elongation factor Tu	Derived by a	WP_003031109.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
322_RS06380	1,382,478	1,383,593		amino acid oxidase	Derived by a	WP_054528547.1	? (not found)	1	? (not found)	? (not found)	? (not found
322_RS17175	3,686,035	3,686,850		AraC family transcriptional regulator	Derived by a	WP_054528023.1	? (not found)	1	1	1	1
322_RS17930	3,844,410	3,846,275		DNA mismatch repair protein MutL	Derived by a	WP_054527983.1	? (not found)	1	1	1	1
.322_RS20890	4,482,716	4,483,960	σ.	O-antigen polymerase	Derived by a	WP_046671022.1	? (not found	1	1	? (not found)	? (not found
.322_RS22035	4,737,170	4,739,713	0	nitrite reductase large subunit	Derived by a	WP_003023592.1	? (not found)	1	1	1	1
.322_RS23080	211,671	211,868		hypothetical protein	Derived by a	WP_072143936.1	1	1	1	1	1
.322_RS23130	545,890	546,069		hypothetical protein	Derived by a	WP_071524456.1	1	1	1	? (not found)	1
.322_RS02845	588,933	589,343		formate hydrogenlyase maturation protein HycH	required for	WP_016150885.1	1	2	1	1	2
.322_RS03030	623,789	624,124		L-valine transporter subunit YgaH	Derived by a	WP_054528723.1	1	1	? (not found)	? (not found)	1
.322_RS03050	627,414	628,478		proline/betaine ABC transporter permease ProW	Derived by a	WP_003846040.1	1	1	? (not found)	? (not found)	1
.322_RS03070	633,115	633,525	nrdI	ribonucleotide reductase assembly protein NrdI	in Salmonella	WP_003037273.1	1	1	? (not found)	? (not found)	1
.322 RS03235	667,222	668,508	01	capsular polysaccharide biosynthesis protein	Derived by a	WP 003839728.1	1	1	1	1	1

### WHEN NATURE IS F.... WITH YOU. HYPERMUTATORS



#### Targets of Distance Columns (CPO C. freundii ST18)

2. 111

Right-click on the allele type columns to jump to the according contig position in the Sample

	De site	<b>c</b>	Carperlanes		Conflored Conflored	C	1 2001	7 410 22552		****	60030100150	44440022002
arget	Begin	End	Genbank gene	VenBank product	GenBank note GenBani	protein_i	2001	1/_A19_22552	AMA003417	AMA003565	CPO20190159	AMA003382
1322_RS23015	4,734,020	4,734,190	-	DUF4223 domain-containing protein	Derived by a WP_0708	08248.1	1		1	1	? (not found)	1
1322_RS22140	4,760,051	4,760,869		peptidyl-prolyl cis-trans isomerase	otamase; D WP_0030	2364 .1	1		1	1	1	1
1322_RS06880	1,473,514	1,474,512		flagellar motor switch protein FliG	Derived by a WP_0030	3044.1	2		1	? (not found)	2	1
1322_RS07395	1,575,127	1,575,330		hypothetical protein	Derived by a WP_054	28499.1	2		2	? (not found)	2	2
1322_RS05555	1,207,439	1,208,275		S-formylglutathione hydrolase	perived by a WP_054	2859.1	2		? (not found)	? (not found)	? (not found)	? (not found)
1322_RS07680	1,614,513	1,617,746		host specificity protein	Derived by a WP_054	28464.1	2		1	1	1	1
1322_RS23505	2,199,438	2,200,406		hypothetical protein	Derived by a WP_0482	33636.1	2		2	2	? (not found)	? (not found)
₹322_RS13800	2,919,337	2,920,134		iron-enterobactin transporter ATP-binding protein	with FepBDE WP_0038	47439.1	2		2	2	2	2
1322_RS14230	3,021,614	3,022,951		putative heme utilization radical SAM enzyme HutW	Derived by a WP_0329	48719.1	2		2	2	2	2
1322_RS16445	3,516,539	3,518,305		peptidoglycan glycosyltransferase FtsI	enicillin-bind WP_0030	1878.1	2		1	1	1	1
1322_RS21860	4,701,589	4,702,308		DNA-binding response regulator	Derived by a WP_001:	5775.1	2		1	2	2	1
1322_RS07400	1,575,323	1,575,667		hypothetical protein	Derived by a WP_0545	28498.1	3		3	? (not found)	3	3
1322_RS03490	747,132	747,476		outer membrane protein assembly factor BamE	Derived by a WP_0038	2640.1	3		1	1	1	1
\322_RS03980	852,238	853,089		3-mercaptopyruvate sulfurtransferase	Derived by a WP_0030	3770.1	3		1	1	1	1
₹322_RS12515	2,648,226	2,649,140		LysR family transcriptional regulator	Derived by a WP_0038	3694.1	3		3	3	3	3
1322_RS19595	4,210,147	4,210,662		GTPase-activating protein	Perived by a WP_016:	5125.1	3		1	1	1	1
1322_RS23220	978,050	978,250		hypothetical protein	Perived by a WP_071	2430.1	4		1	? (not found)	1	1
1322_RS22970	2,993,272	2,995,080		hypothetical protein	Derived by a WP_0634	56226.1	4		1	1	1	? (not found)
1322_RS15095	3,212,238	3,213,065		ABC transporter	perived by a WP_0466	69784.1	4		1	1	1	1
322_RS15610	3,324,822	3,325,604		flagellar biosynthetic protein FliR	Derived by a WP_0038	43839.1	5		1	1	1	1
1322_RS06570	1,422,859	1,423,350	1	microcompartment protein PduM	Perived by a WP_0038	3900 . 1	5		1	1	1	1
322_RS07180	1,531,973	1,532,617	5	protein phosphatase CheZ	perived by a WP_0030	34659.1	5		1	1	1	1
1322_RS10990	2,329,185	2,330,675		sensor domain-containing diguanylate cyclase	Derived by a WP_016:	50102.1	5		1	1	1	1
322_RS11990	2,527,371	2,528,186		histidinol-phosphatase	Perived by a WP_0482	3341.1	5		1	1	1	1
322_RS14630	3,109,037	3,109,486		NrdR family transcriptional regulator	perived by a WP_0030	2157 .1	5		1	1	1	1
322_RS16965	3,641,128	3,642,201		patatin family protein	perived by a WP_0038	3727.1	5		1	1	1	1
1322_RS17340	3,721,589	3,722,332		hypothetical protein	Derived by a WP_0549	28666.1	5		1	1	1	1
322_RS00135	16,098	16,670		L-threonylcarbamoyladenylate synthase type 1 TsaC	Derived by a WP_0038	42158.1	6		1	1	1	1
322 RS00920	179,070	181,364		formate acetyltransferase	Derived by a WP 0030	2487.1	6		1	1	1	1
322_RS01355	273,926	275,722		aryl-sulfate sulfotransferase	perived by a WP_0549	28788.1	6		1	1	1	1
1322 RS02250	457,972	460,131		bifunctional 2-acylolycerophosphoethanolamine acyltransferas	Derived by a WP_0030	3394.1	6		1	1	1	1
1322 RS02540	523,945	524,727		tRNA pseudouridine(65) synthase TruC	Derived by a WP_016:	5091.1	6		1	1	1	1
1322 RS02590	538,450	540,087		CTP synthetase	Derived by a WP 0030	3412 .1	6		1	1	1	1
1322 RS02715	563,064	564,491		phenolic acid decarboxylase	Derived by a WP 0545	2873.1	6		1	1	1	1
1322 RS03885	833,772	835,409		ribulokinase	Derived by a WP 0466	7015.1	6		1	1	1	1
1322 RS04140	902,354	903,817		hypothetical protein	Derived by a WP 0447	0178.1	6		1	1	1	1
322 RS04435	963,488	964,459		cysteine synthase A	Derived by a WP 0030	3809.1	6		1	1	1	1
1322 RS04845	1,046,871	1,047,764		epimerase	Derived by a WP 0030	28110.1	6		1	1	1	1
322 RS04915	1,060,772	1.061,989		aminotransferase AlaT	road specifi WP 0030	28082.1	6		1	1	1	1
322 RS05165	1,117,880	1,120,516		DNA gyrase subunit A	Derived by a WP 044	01169.1	6		1	1	1	1
1322 RS05325	1 153 837	1 155 597		hypothetical protein	Derived by a WP 0161	50671 1	6		1	1	1	1

#### THANKS FOR THE ATTENTION

