Bundesinstitut für Risikobewertung

DTU Food National Food Institute

MINISTRY OF FOOD AND DRUG SAFETY
National Institute of Food and Drug Safety Evaluation

anses
French agency for food, environmental and occupational health & safety
Investigate, evaluate, protect

Rapid Detection and Identification of Food-borne Pathogens using Single Nucleotide Polymerphism (SNP) Profiling of Their Whole Genome Sequences (WGS)

2019. 3. 26.

Ju-Hoon Lee, Ph.D.

Dept of Food Science and Biotechnology, Kyung Hee University
National Institute of Food and Drug Safety Evaluation, South Korea

Importance of Foodborne Pathogen Study

Foodborne Outbreaks in South Korea

<Foodborne outbreaks>

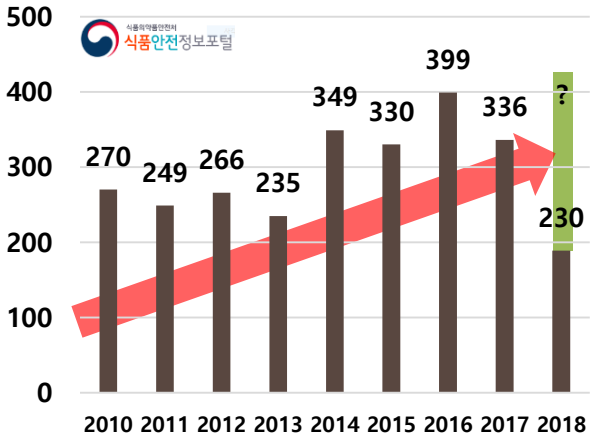
(Source: KBS)

(Source: KTV)

(Source: YTN, 2017)

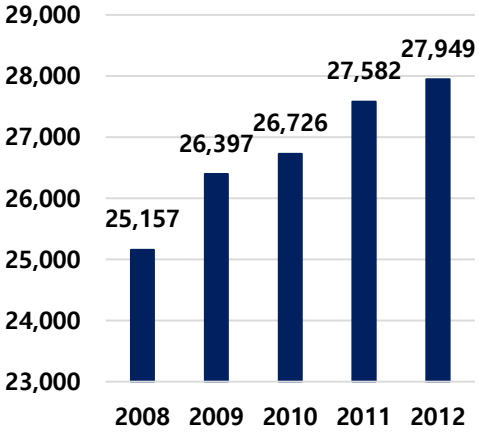
- Happened every year
- Outbreak-causing foods should be investigated for prevention

<# of outbreaks>



- # of outbreaks increase last 5 years (Avg: Food poisoning 332 cases, Patients 6,257)
- 26% of outbreaks caused by foodborne pathogens and 45% by unknown (Source: 2018 Korean food safety information portal)

<Economic loss>

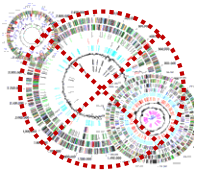


- >2.5 Billion US$ and increase every year

(Source: Korean NIFDS 2018)

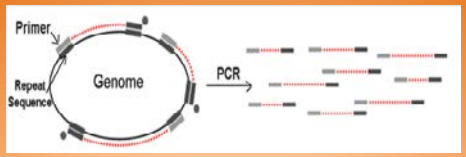
Systematic management for foodborne pathogens is required in omics level

Current Detection Limitation of Foodborne Pathogens (FP)

Current detection methods	Problems	Omics-based identification
1 Culture-based biochemical tests 	➢ Inefficiency by long cultivation time ➢ No information about FP genomes ➢ No information about Food-specific virulence factor gene expression ➢ No detection of unculturable FP  	➢ Acquisition of FP genome information and rapid identification ➢ Transcriptome-based identification of food-specific VFs ➢ Metagenome-based identification of unculturable FP
2 Rep-PCR, PFGE DNA-based tests 	➢ Low identification fidelity even due to short DNA sequence modulation or point mutation ➢ Low accuracy due to short DNA sequences 	➢ DB construction of FP genome/transcriptome/metagenome ➢ Development of rapid FP identification pipeline program using specific SNP patterns

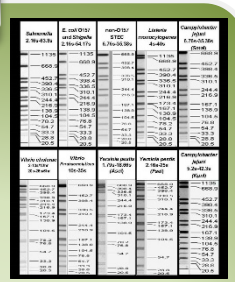
Molecular Identification Techniques for Rapid Detection of FP

Rep-PCR



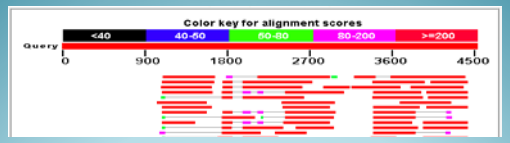
- Rep-PCR is performed using PCR with repeat sequence-targeting primer
- According to the PCR band patterns, the strain is identified
- Advantage: Rapid bacterial identification is possible, even though its genome sequence is unknown
- Disadvantage: There is a limitation and low accuracy for bacterial identification with very short PCR band patterns, according to the locations of repeat sequences

PFGE



- PFGE analysis is based on the locations of specific restriction enzyme (RE) recognition sites
- According to the DNA band patterns after specific RE digestion, the strain is identified
- Advantage: PulseNet DB is well-developed and organized for rapid identification with RE band patterns, even though its genome sequence is unknown
- Disadvantage: Only a point mutation in RE sites can change PFGE band patterns, indicating low accuracy

16S rRNA analysis



- 16S rRNA sequence analysis is based on sequence homology for bacterial identification
- PCR and sequencing of 16S rRNA gene can be done quickly
- Advantage: Massive amount of bacteria 16S rRNA sequences are accumulated in many DBs.
- Advantage: Accurate detection and identification are possible in genus and even species level
- Disadvantage: Relatively low resolution and accuracy comparing to ANI analysis with whole genome sequences

Based on Foodborne pathogen whole genome sequences, rapid and accurate identification is possible for advanced food safety

Next-Generation Sequencing (NGS)

Illumina HiSeq 2500

Illumina MiSeq

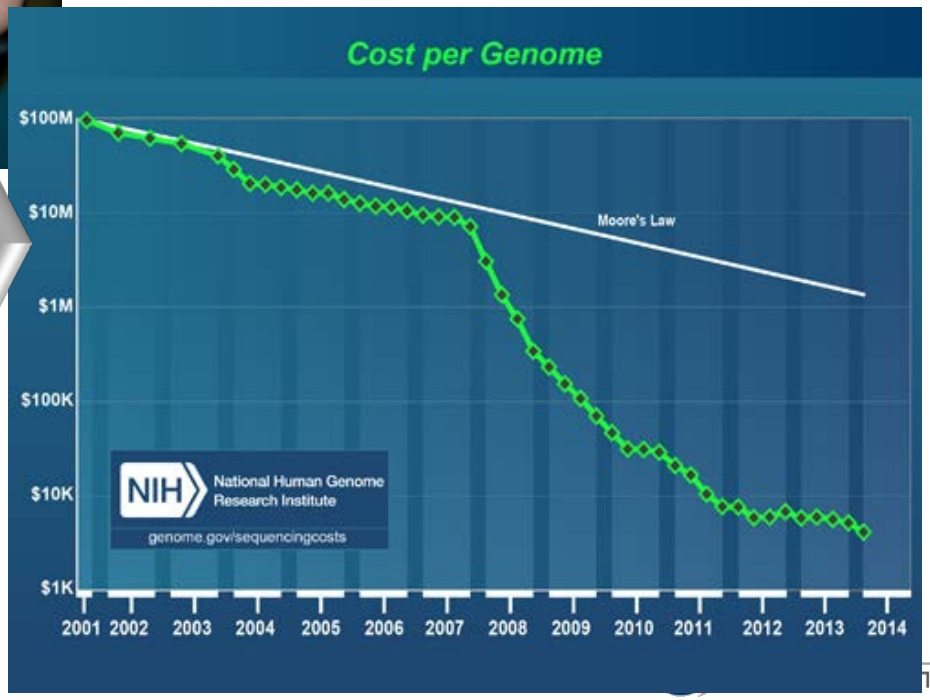
NGS

PacBio RS II

NanoPore

Platform Features

Feature	HiSeq2500 - Highoutput	HiSeq2500 – Rapid mode	MiSeq	PacBio RSII
Number of reads	150-180M/lane	100-150M/lane	12-15M (v2) 20-25M (v3)	50-80K/SMRT cell
Read length	2 x 100 bp	2 x 150 bp	2 x 300 bp (v3)	~ 10-20 kb
Yield per lane (PF data)	up to 35 Gb	up to 45Gb	up to 15 Gb	up to 0.4 Gb
Instrument Time	~12-14 days	~2 days	~2 days	~2 hours
Pricing per Gb	$59 (PE100)	$53 (PE150)	$108 (PE300)	$697



Lowering sequencing cost by NGS

- → High accessibility of genome study, especially foodborne pathogen genome study
- → Transcriptome study for VFs
- → Metagenome study for FP composition
- → Rapid detection and accurate identification

International Trend of Omics Study for Foodborne and Clinical Pathogens

U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

100K GENOME PROJECT

UC DAVIS
UNIVERSITY OF CALIFORNIA

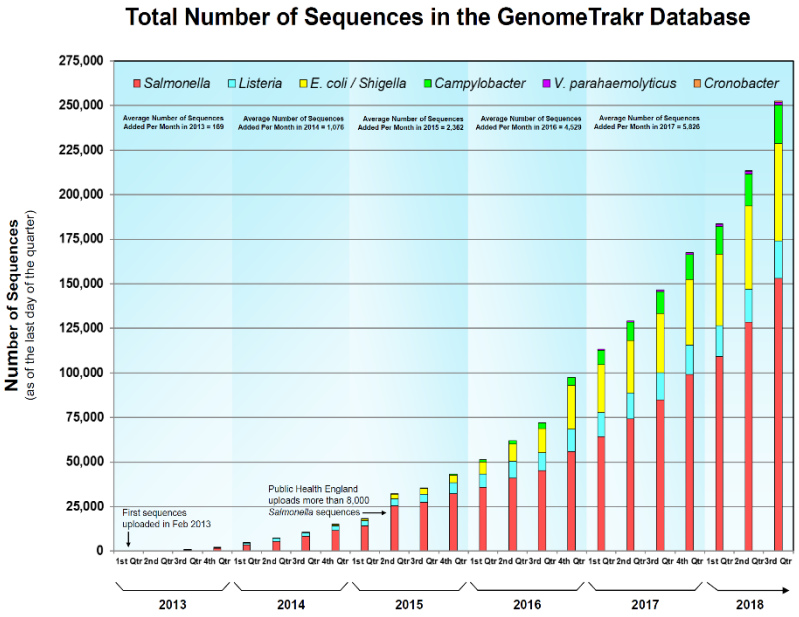
- Launched in 2012 by UC Davis (Dr. Bart Weimer)
- FDA, Agilent Technologies, BGI supported

[Collaborators]

- China FDA for 10K (100K Genome Project China)
- NIFDS/FORC for 1K (100K Genome Project Korea)
- Health Canada for 10K *Salmonella* (100K Genome Project Canada)

FDA's GenomeTrakr

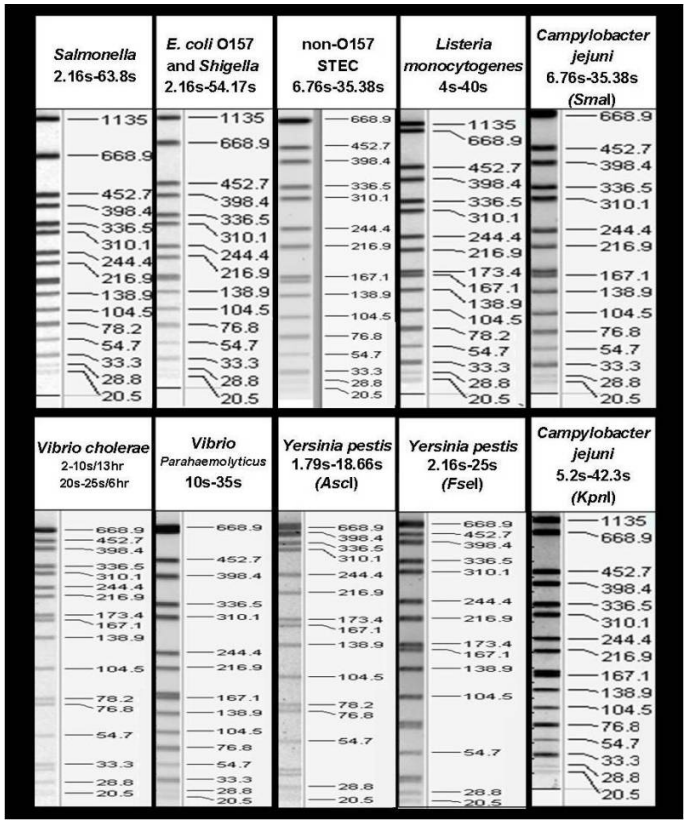
- Launched in 2012 by US FDA
- Collaborated with US CDC for *Listeria*
- Collaborated with MN/WA/NY/FDA for Real-time *Salmonella*
- In addition, >24 national labs joined this project for pathogen genome sequencing
- *E. coli, Campylobacter, Vibrio, Cronobacter,* etc.



식품의약품안전처
식품의약품안전평가원

Omics Research Trend in South Korea : FORC by NIFDS

PulseNet
The National Molecular Subtyping Network for Foodborne Disease Surveillance
CDC



NGS

NIFDS

(2014-2018)

FORC 식품의약품안전처 식중독균유전체연구사업단
Food-borne pathogen Omics Research Center

FORCDB

FORC SNPing

- FP complete genome
- Food metagenome
- VF transcriptome

- FP WGS
- SNP analysis for Rapid detection and identification

Research Results of Food-borne Pathogen Omics Research Center (FORC) in South Korea (2014-2018) Supported by NIFDS

DB Construction

1

Complete genome	10
Transcriptome	4
Metagenome	4
VFs	13

Monitoring **93 cases**

DB update

2

Monitoring **412 cases**

Complete genome	20
Transcriptome	9
Metagenome	124
VFs	13

HGTree DB construction

Info integration

3

Complete genome	22
Transcriptome	7
Metagenome	174
VFs	10

Monitoring **320 cases**

FORC SNPing pipeline program

DB upgrade

4

Complete genome	19
WGS	583
Transcriptome	8
Metagenome	188
VFs	11

Monitoring **336 cases**

5

Complete genome	29
WGS	585
Transcriptome	8
Metagenome	272
VFs	10

Integrated Omics DB Service for Korean FP

Final Results

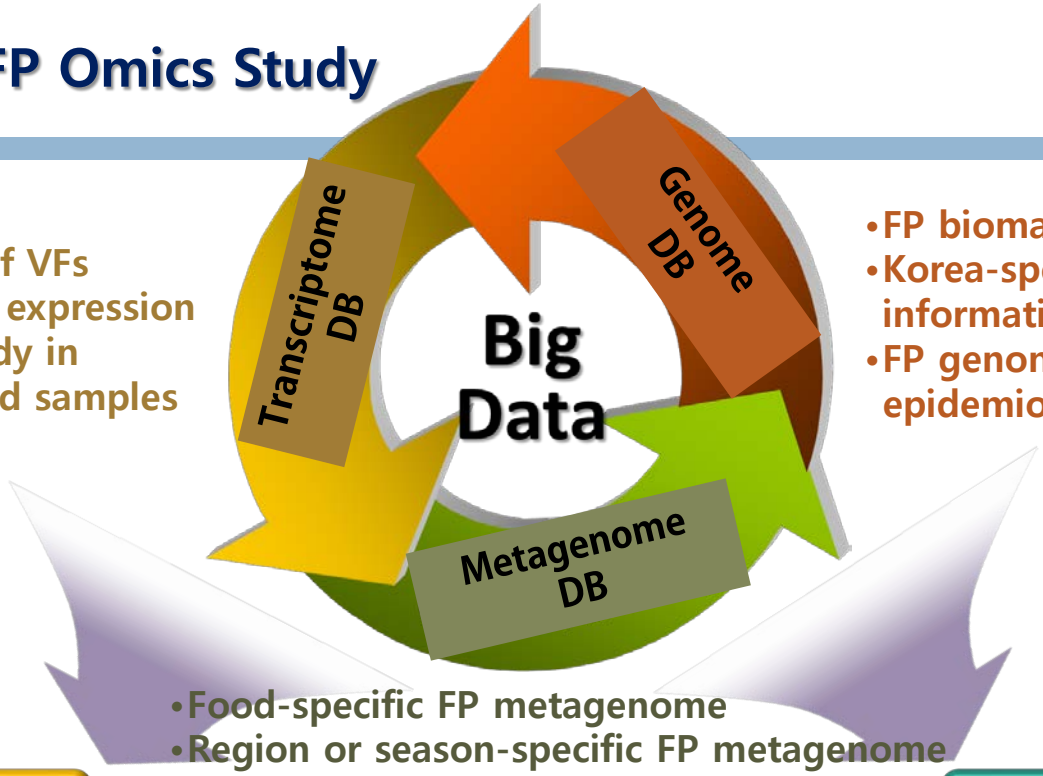
Comp. genome	100
WGS	1,168
Transcriptome	36
Metagenome	762
VFs	57

Copyright	1
Patents	2
Patent applications	20
SCI papers	39

식품의약품안전처
식품의약품안전평가원

Application of FP Omics Study

- Detection of VFs
- FP VF gene expression pattern study in specific food samples



- FP biomarker identification
- Korea-specific FP genome information
- FP genome evidence for epidemiological study

- Food-specific FP metagenome
- Region or season-specific FP metagenome

Domestic food safety

◎Outbreak Monitoring◎

- Food
- Region
- Season

} Real-time monitoring

FP outbreak surveillance

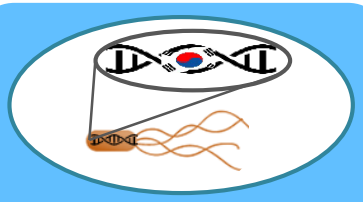
◎Epidemiological survey◎

- Patients
- Foods
- Equipments

} Genome analysis

Rapid ID of FP

Export/import food safety



◎FP genome barcode◎

Korean FP {
- Genome DB
- Genome features
- Genome evidence

International dispute solution

◎Ex/Import food safety◎

- Food safety inspection
- Food bacterial specification
- Enhanced ex/import food safety

National confidence

WGS-based GenomeTrakr SNP Analysis for Rapid ID of FP

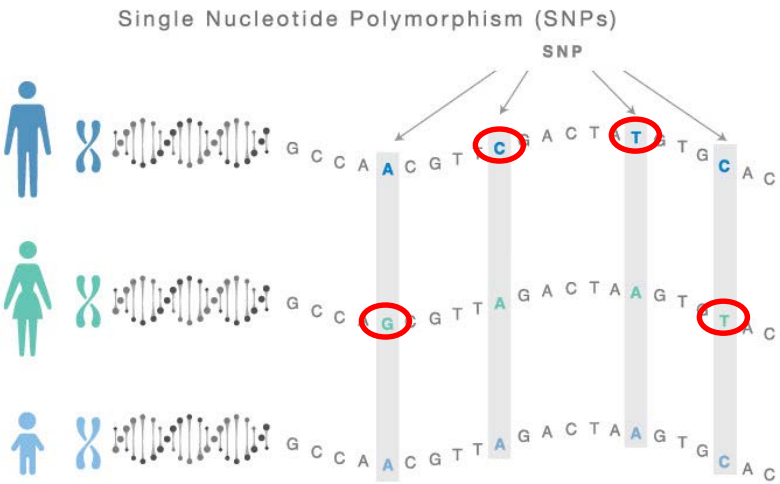
CFSAN SNP Pipeline (2014-15)

Documentation: http://snp-pipeline.rtfd.org

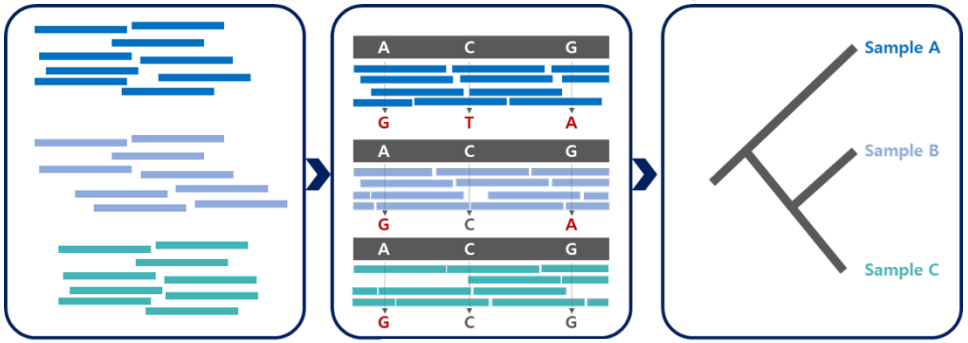
Source Code: https://github.com/CFSAN-Biostatistics/snp-pipeline

Pettengill JB, Luo Y, Davis S, Chen Y, Gonzalez-Escalona N, Ottesen A, Rand H, Allard MW, Strain E. (2014) An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. PeerJ 2:e620 http://dx.doi.org/10.7717/peerj.620

Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E. (2015) CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science 1:e20 https://dx.doi.org/10.7717/peerj-cs.20



Basic concept for SNP analysis



① Raw WGS by NGS ② Mapping with Ref. genome ③ SNP tree analysis

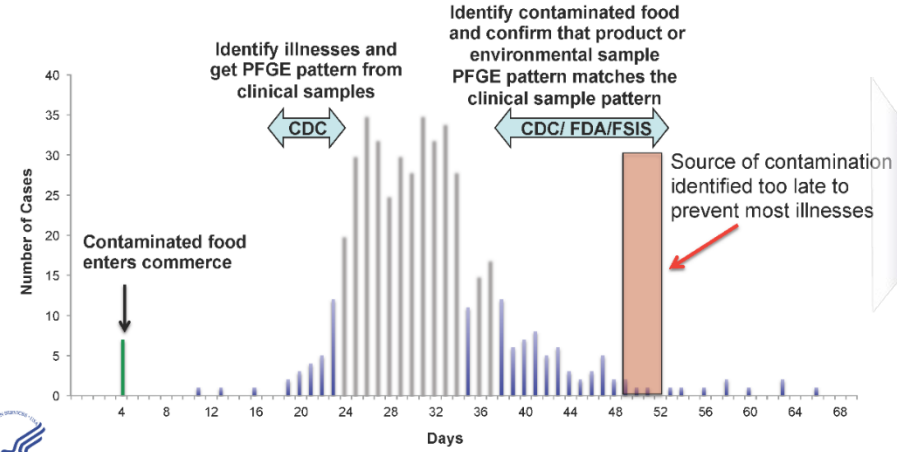
Comparative SNP tree analysis using WGS is a key method for rapid ID of FP

식품의약품안전처
식품의약품안전평가원

Advantage of GenomeTrakr WGS/SNP Pipeline over Traditional PFGE Analysis

PFGE

Timeline for Traditional Approach to Foodborne Illness Investigation

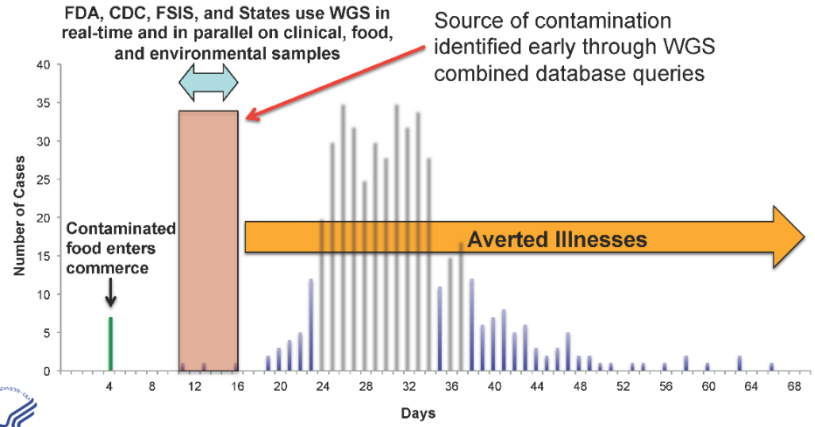


1. In the initial outbreak, CDC try to identify illnesses and isolates original FP strain from clinical sample for PFGE analysis
2. PFGE pattern is obtained from the strain in clinical sample
3. Contaminated food is identified by CDC/FDA/FSIS and FP strain is isolated from the food sample
4. PFGE pattern is obtained from the strain in food sample
5. The PFGE patterns between clinical and food isolates are compared for matching
6. Source of contamination is finally identified

→ It is too late to prevent the propagation of food outbreak

WGS/SNP

TARGET: Timeline for Foodborne Illness Investigation Using Whole Genome Sequencing



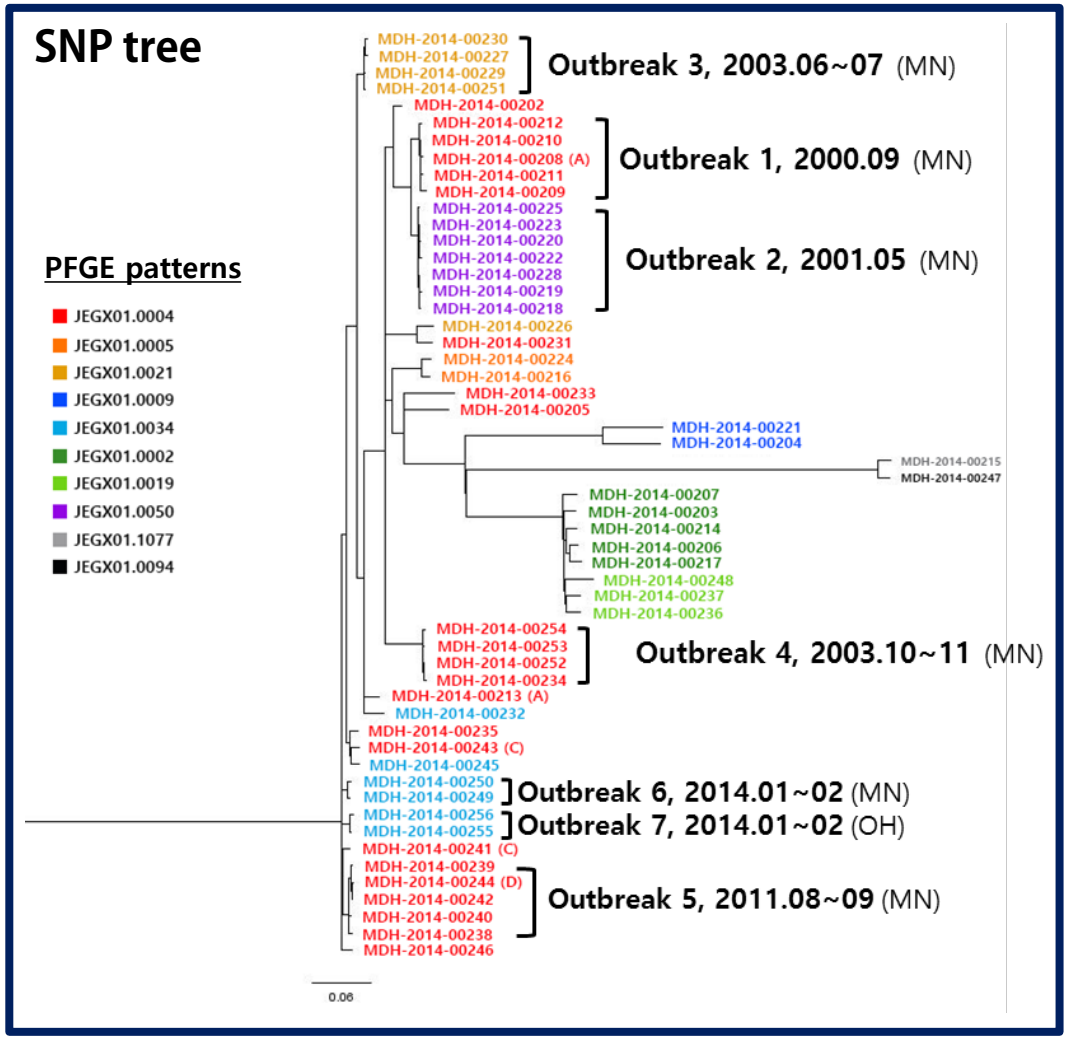
1. In the initial outbreak, CDC/FDA/FSIS isolate potential FP strains from clinical and food samples at the same time
2. WGS is performed using NGS and then original FP strain is identified with WGS DB, which is present in both samples
3. Source of contamination is finally identified
4. For further epidemiological study, SNP analysis is conducted with WGS data and then FP strain is confirmed in the SNP-based reference tree with its SNP pattern

→ It is possible to prevent the propagation of foodborne pathogen before outbreak

Evaluation of FORC SNPing Pipeline vs. PFGE/GenomeTrakr Pipeline

- FORC SNPing pipeline was evaluated with 55 *S.* Enteritidis strains from Minnesota and Ohio, USA
- WGS was obtained and SNP tree analysis was conducted
  - : PFGE pattern analysis is impossible to determine the original outbreak for specific FP
  - : SNPing pipeline analysis can determine the original outbreak for specific FP in SNP tree

	한국 식약처 FORC SNPing	US FDA CFSAN SNP pipeline
True Positive	49,483 / 50,000	49,358 / 50,000
True Negative	479,815,994 / 479866000	479,815,995 / 479866000
False Positive	6	5
False Negative	517	642
Sensitivity	99.0%	98.7%
Accuracy	99.9%	99.9%
Specificity	99.9%	99.9%



Prerequisites for enhanced accuracy of SNP tree analysis

- Various reference genome sequences with high accuracy and fidelity are required
- Massive WGS information and correct outbreak history are required
- **Highly accurate reference SNP tree should be constructed**
- Optimized NGS facility and most recently updated SNP pipeline program are required

Summary

1. Omics study for foodborne pathogen is required for advanced food safety

- Accumulation of **complete genome sequences** as reference genomes is important for accurate identification of foodborne pathogens
- **Transcriptomics** study is required to understand virulence factor gene expression in specific food environments for regulation of virulence and toxicity in foodborne pathogens
- **Metagenomics** study is required to elucidate composition and population of foodborne pathogens in specific foods for prevention of foodborne outbreaks by control of the food consumption

2. SNP analysis using WGS is required for practical application and further epidemiological survey

- **Accumulation of whole genome sequences** of various foodborne pathogens and their outbreak history are needed to overcome the limitation of PFGE analysis
- **WGS-based SNP analysis data** should be collected in database and the **reference SNP tree** should be constructed with the most updated SNP profiles
- **FORC SNPing pipeline program** is more sensitive and faster for identification of foodborne pathogens and their epidemiological survey than GenomeTrakr CFSAN pipeline program

3. Further SNP analysis study is important to improve efficiency and accuracy

- WGS data in all public databases and WGS of more than 6,000 foodborne pathogens will be collected for **update of SNP database in FORC DB**
- **FORC SNPing pipeline program** will be more optimized and upgraded for analysis service


Thank You for Attention