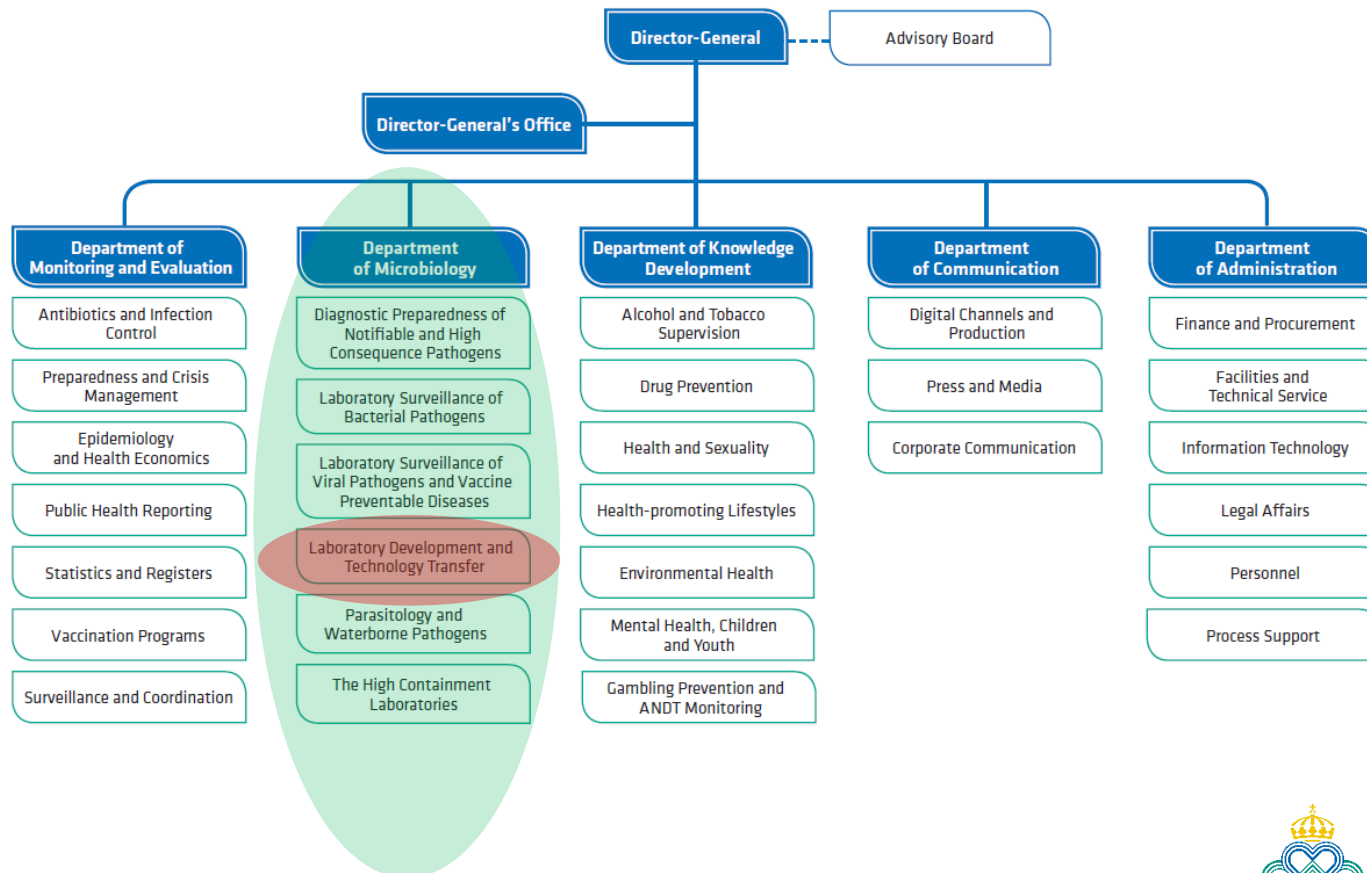


# NGS surveillance and diagnostics, from development to routine

Erik Alm, Bioinformatician, Head of NGS platform

Public Health Agency of Sweden  
2016-06-21

# Public Health Agency of Sweden



# PHA Sweden – Department of Microbiology

- Microbial typing - Surveillance
- Outbreak investigations
- BSL3 and BSL4 lab – Highly pathogenic bacteria and viruses, emerging viruses and bioterrorism.
- Diagnostics

# History of NGS at Public Health Sweden

- 2007-2011: 454-pyrosequencer, used for research
- Early 2012: NGS-platform was established using Ion Torrent technology. 1 bioinformatician, 1 technician, 1 application specialist. Project to move Influenza surveillance to NGS
- 2014: Influenza surveillance moved to NGS. Listeria validated on NGS. Second bioinformatician. Second technician.
- 2015: Strategy to use NGS as main tool for typing. Third bioinformatician. Staff to run and maintain platform for routine analysis. Metagenomics tested for diagnostics.
- 2016: Most bacterial typings moved to NGS. Metagenomics in analysis catalogue.

# NGS-plattformen på Folkhälsomyndigheten



Reza Advani  
*PhD*



Maria Lind Karlberg  
*PhD*



Shahnaz Mahdavifar  
*Labtekniker*



Raija Ahmed  
*PhD*



Petra Holmström  
*PhD*



Erik Alm  
*bioinformatiker*  
*PhD,*  
*Plattformsansvarig*



Björn Hallström  
*bioinformatiker*  
*PhD*



Gabriel Östlund  
*bioinformatiker*  
*PhD*



Library Builder 3x



Ion Chef 2x

Capacity: 200  
genomes/week  
Max 3 days for  
results



Ion S5 XL 2x



Ion PGM 2x

# Roadmap for implementation of NGS

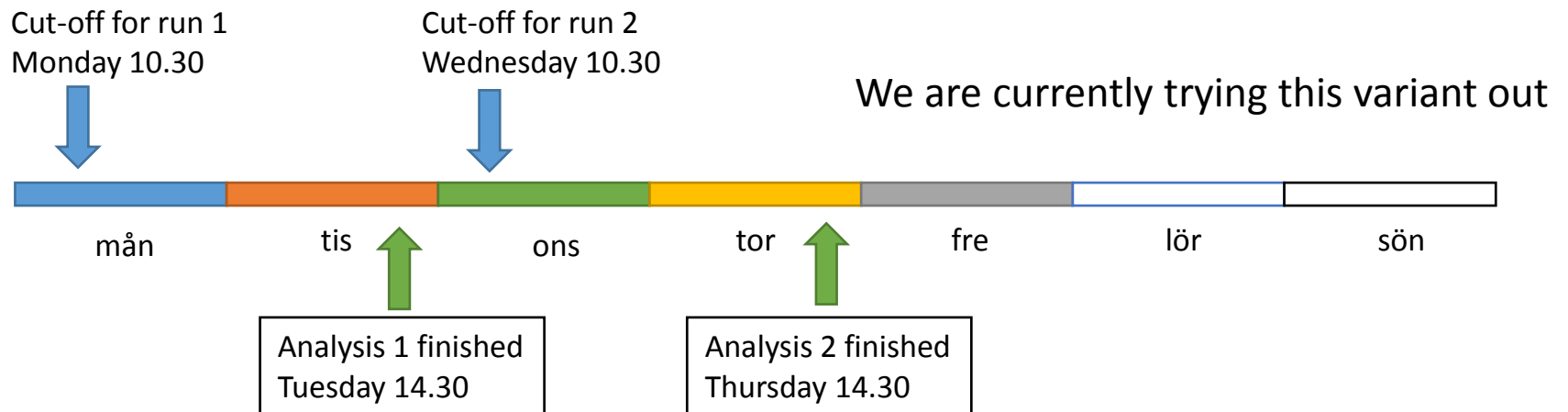
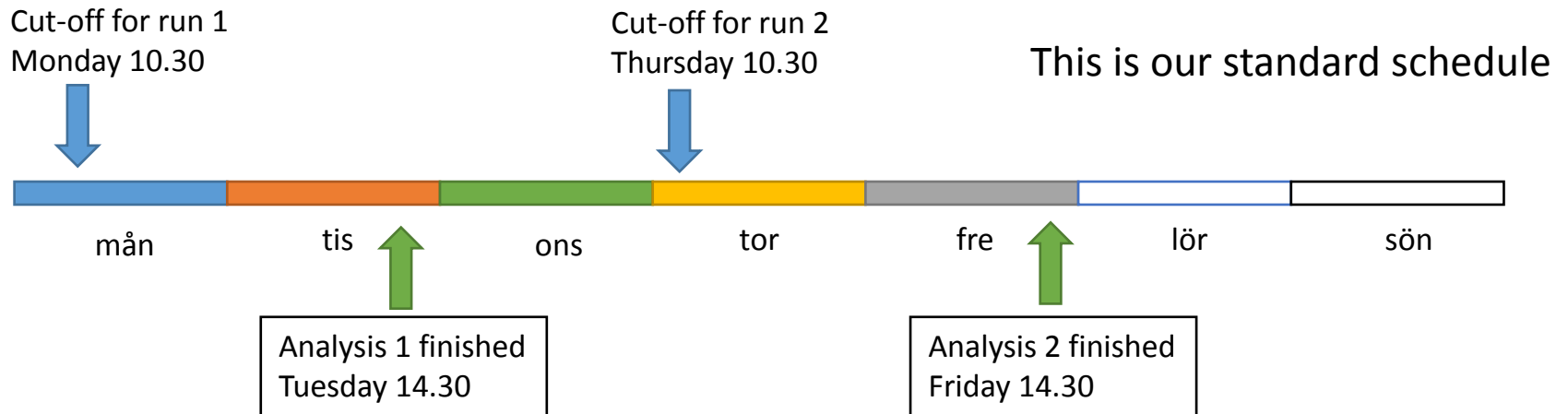
- Influenza A and B – Implemented for routine analysis (2014)
- Listeria – Implemented for routine- and outbreak analysis (2015)
- ESBL – Implemented for resistance gene characterization (2015)
- Streptococcus pneumoniae (MLST) – implemented (2015)
- Salmonella – Implemented for outbreak investigations only (2015)
- Vancomycin resistant enterococcus – Implemented (2015)
- Metagenomics – implemented (2016)
- STEC – Implemented for characterization and surveillance (2016)
- MRSA – Implemented for outbreaks (2016)
- Haemophilus influenzae – Implemented (2016)
- C. difficile – Implemented (2016)
- M. tuberculosis – Validation ongoing (Autumn 2016)
- Viruses – Whole genome amplification development (HBV, HAV, Rota)

# Implementation of NGS in the routine

- Started with Influenza A in the 2014/2015 season
- We passed critical mass for the platform to run routine once a week during early 2015
- We passed critical mass for two runs per week in the end of 2015
  
- We have now implemented most of the bacterial surveillance programs using WGS
  
- Four large volume exceptions remain:
  - Salmonella (various methods are currently used)
  - MRSA (spa-typing current method)
  - Pneumococci (only MLST amplicons on NGS, serotyping still done traditionally)
  - GAS (emm-typing currently used)



## Weekly schedule



# Dataanalys for routine: BactTyper

- In-house software for typing of bacteria
  - Species
  - MLST
  - Molecular serotype
  - Toxins, resistance genes, resistance mutations
- SNP-typing
- MST and phylogenetic trees
- Specific analyses for each bacteria
- Report
  - Detailed report per sample
  - Overview of many samples
  - Trees for molecular epidemiology
- External portal
  - External data can be imported into the system. Ion Torrent, Illumina, Pac Bio, Finished genomes

# BactTyper flow

Assembly



Species determination



MRSA

EHEC

VRE

ESBL

Listeria

C. Diff

TB

Utbrottsanalys



Automated emails with results

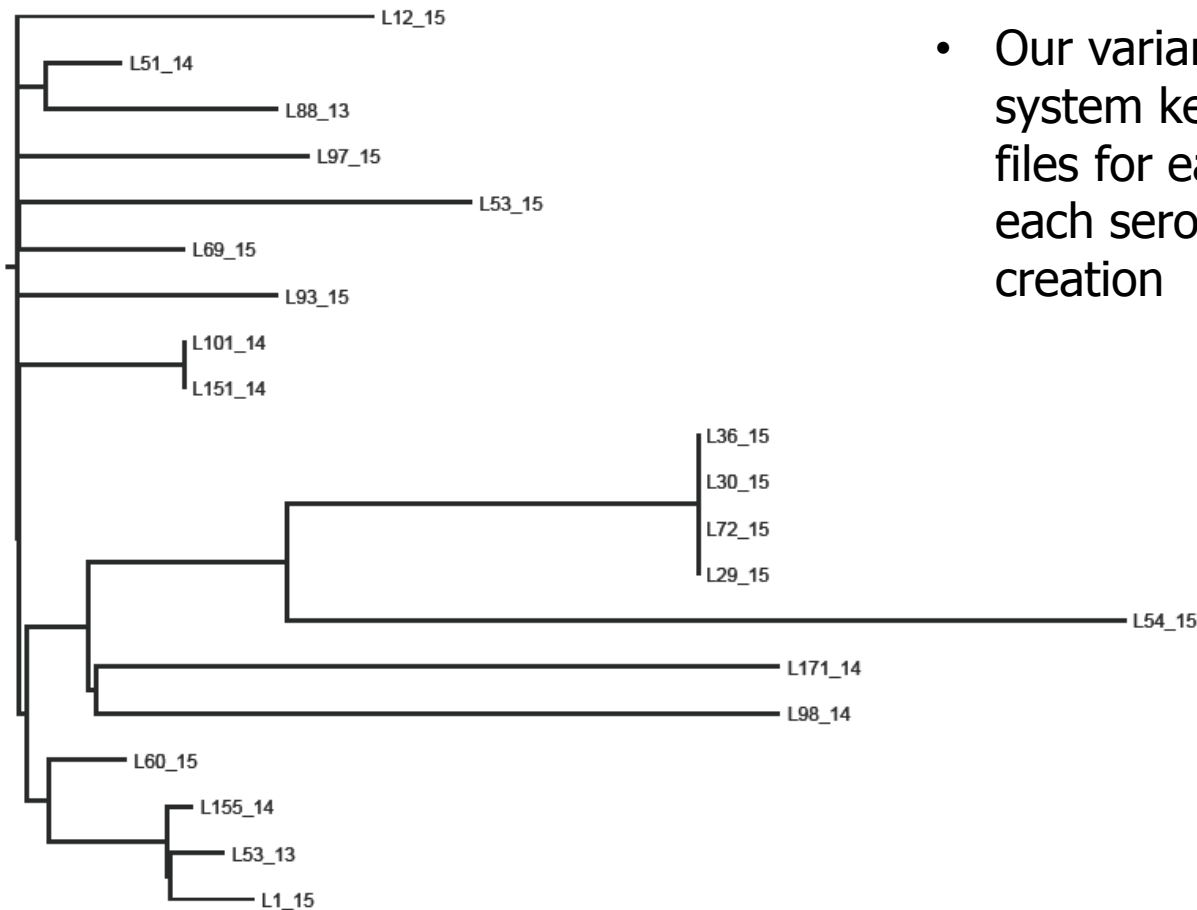
# Information extracted from the data

- In-silico PCR, replaces PCR-based methods
  - Serogroup determination for *Listeria monocytogenes*
  - Capsule type for HI
- Gene databases, finds specific genes
  - Serotyping for *E.coli*
  - Antibiotic resistance genes and resistance SNPs
  - Detection of toxins and other virulence factors for EHEC
- MLST
  - Replaces traditional seven gene MLST using WGS
  - Implementation of MLST+ (cgMLST/wgMLST) in progress
- wgSNP-analysis
  - High resolution cluster analysis
  - Variant files are stored in a repository
  - An in-house client software can create trees of selected strains in seconds

# Data storage and computing infrastructure

- Analysis done on instrument servers (Ion S5 XL)
- Data storage administered by IT department
  
- ~1-2 TB of data per year (gzip compressed fastq)

# SNP tree – *Listeria monocytogenes* ST451



- Our variant file repository system keeps one set of variant files for each ST and one set for each serogroup for fast tree creation

0.03

# The present collaboration between agencies on national level

- Listeria, Campylobacter, EHEC, Salmonella
- Public Health Agency supports the National Food Agency with WGS analysis
- National Veterinary Institute and Public Health share strains for WGS comparison
  
- A few local clinical labs have capabilities.
  - Örebro
  - Gothenburg
  - Linköping (starting up)

# Challenges on national and international level

- No standardized nomenclature
- Structures for systematic data sharing missing
- Interpretations of clusters: "are they alike or not?"
- How much meta data in addition to WGS data from different sources (human, food, animal) can be shared
- When and how can the results from joint cluster investigations be communicated?
- Hard to determine who does what on national level



# Wishlist

- Data sharing between sites in real time
  - National
  - International
- cgMLST services
- Common nomenclature for clusters
- Consensus on interpretation