



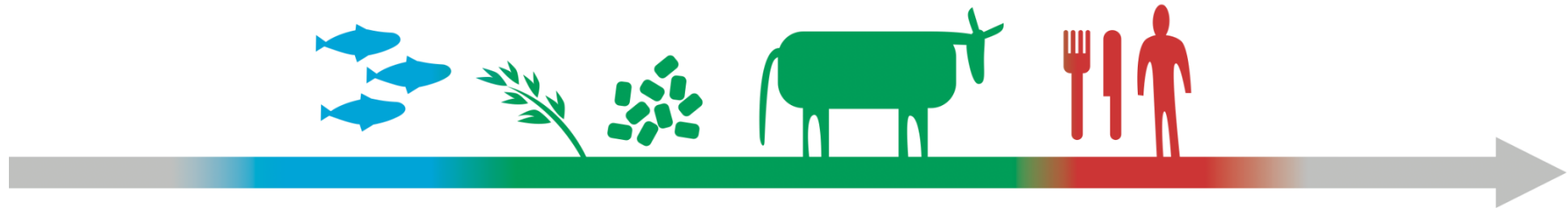
Veterinærinstituttet  
Norwegian Veterinary Institute

# SIBYL

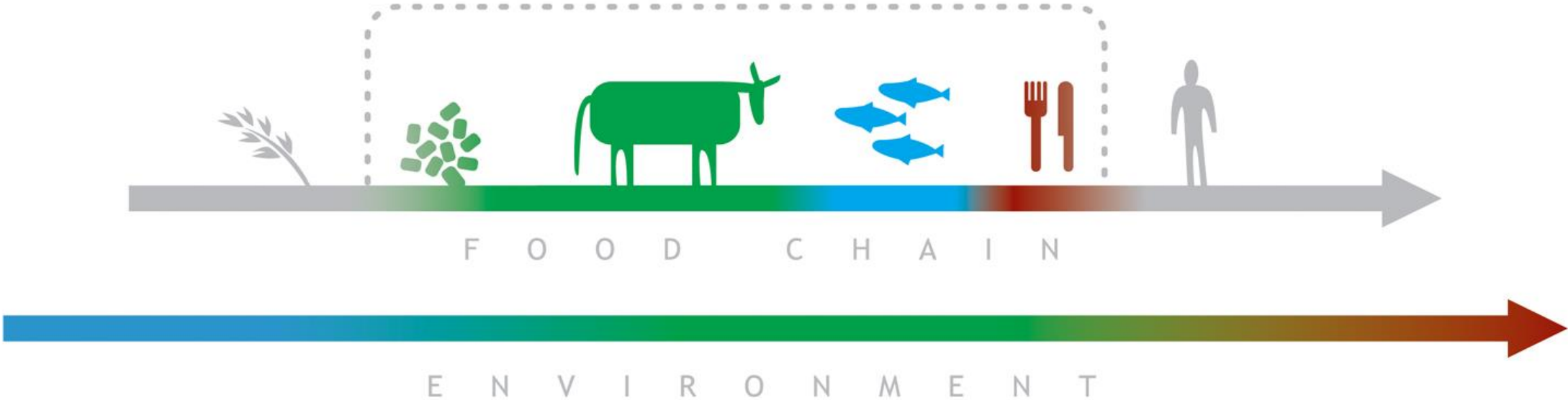
high throughput Sequencing and Bioinformatics analysis pLatform

Karin Lagesen

Parma 2023-09-05



# Areas of focus NVI



# A bit of history

- 2018/19: got two large projects focused on implementation of WGS for the institute
- ORION: One health suRveillance Initiative on harmOnization of data collection and interpretation
- SEQ-TECH: Implementation of NGS and e-infrastructure at the NVI
- Currently: NGS adopted into the digital strategy at the institute



# What analyses are we thinking of?

- Analysis of bacterial genomes (for now)
- Standard analysis types - diagnostics
  - Characterization – AMR, virulence, serotype, MLST ++
  - Clustering – place into context, have we seen it before
- Surveillance – results comparable over time
  - Apples to apples, not to pears



# Function

Sample submission	Lab tracking	Sequencing Demult.plx	Compute	Workflow engine	Workflow run starter/monitor	Metadata db/raw data storage	Results database	Interpret.n. platform
-------------------	--------------	-----------------------	---------	-----------------	------------------------------	------------------------------	------------------	-----------------------

---

# The lab end - Clarity

- Illumina owned NGS LIMS developed specifically to support Illumina sequencing
- Hosted AWS Germany
- Tracks process data - what happens from sample to sequence
- Default workflows, but can be modified
- Configurable with python and groovy

The screenshot displays the Illumina NGS LIMS interface, divided into several sections:

- WORK IN PROGRESS:** Shows a workflow for 'Illumina DNA Prep v3.3' at Step 4 (Normalization - EpMotion) with 89 days ago and a 'Continue' button.
- AVAILABLE WORK:** A table listing various workflow steps and their total counts:

Workflow	Total
Prepared Libraries - Validation v1.0	8
DNA Extraction v3.0	6
Step 1 » Sort Incoming Samples [DNA Extraction v3.0]	1
Step 2 » In Freezer [DNA Extraction v3.0]	0
Step 3 » On Agar/Broth for Extraction [DNA Extraction v3.0]	0
Step 4 » QIAcube [DNA Extraction v3.0]	0
Step 5 » QIAamp [DNA Extraction v3.0]	5
DNA Tubes for Placement v1.0	0
- RECENT ACTIVITIES:** Lists recent user activities, such as 'Wenche Stoldal' performing 'Fluorescence - Quantity QC' on 'Illumina DNA Prep v3.3'.
- Step Details:** A detailed view of the 'Illumina DNA Prep v3.3' workflow, showing 'Assign Next Steps (24 samples)'. It includes fields for 'Minimum Pipetting Volume (µl)' (2), 'Maximum Pipetting Volume (µl)' (30), and 'Total Volume per Sample (µl)' (30). It also has checkboxes for 'Enforce Minimum Pipetting Volume' and 'Enforce Maximum Pipetting Volume', both set to 'Yes'. The 'Experiment Name' is 'LP230322-RF'. Instructions and EPmotion Notes are provided below.
- Files:** A list of files associated with the workflow, including 'Driver file EpMotion - sample', 'Driver file EpMotion - buffer', 'Deck setup pdf', and 'Manual pipetting pdf', each with a corresponding file link.
- Samples that Passed Normalization:** A table showing a list of samples, including 'BioRad HSP9601: LP230322-RF-Norm (24 samples)'.

NVI Clarity LIMS sample submission list. For Sample Type "WGS colonies"

See instructions in headers below (steps 1-2)

**IMPORTANT:**  
\* This template is for up to 96 samples. If you are submitting > 96 samples, move </SAMPLE ENTRIES> further down to add more  
Questions? Contact the NVI LIMS team: lims\_clarity\_lab@vetinst.no

start here

1. Insert sample name & Freezer Location			2. Make selection in drop-down menus	
Sample Name manually insert	Freezer Location manually insert (optional)	Optional: Ekstra kontaktperson	Are your samples located in the freezer or on a dish? Drop-down menu	Select your genus/VIGAS project Drop-down menu
Name is copied to Sample/Name (column G) while removing any blank space and ÆØÅ		Project account vil ha en standard kontaktperson. Her legges til eventuelt ekstra kontaktpersoner (EPOST!)	<u>WGS colonies - freezer</u>	<u>Escherichia spp</u>

### Projects

Projects

Display Projects with Status: Pending Open Closed

PROJECT NAME	PROJECT STATUS	SAMPLES	LAST PROJECT NOTE	DATE SUBMITTED	LAST UPDATE
230825-23011-animalia-AGL	Open	1	haster	2023-08-25	2023-08-28 08:53 AM
230825-42015-NORM-VET-AGL	Open	11	--	2023-08-25	2023-08-28 08:53 AM
230822-21124-APEC-NorskKylling-GB	Open	2	--	2023-08-22	2023-08-22 02:36 PM
230822-21121-136-GB	Open	1	--	2023-08-22	2023-08-22 12:21 PM
230822-12160-136-NRL-GB	Open	3	--	2023-08-22	2023-08-22 12:21 PM
Diginostics - 31205	Open	5	--	2023-08-21	2023-08-21 02:20 PM
230816-13100-130A-FLAVO	Open	3	--	2023-08-16	2023-08-16 03:32 PM
230815-UG-DAA	Open	95	--	2023-08-15	2023-08-15 12:29 PM
Matbakt_21121-136	Open	3	--	2023-08-15	2023-08-15 11:30 AM
230605-WGS-AI-CAB	Open	28	--	2023-06-06	2023-08-15 10:40 AM

< 1 2 3 4 5 6 7 8 9 10 > Last

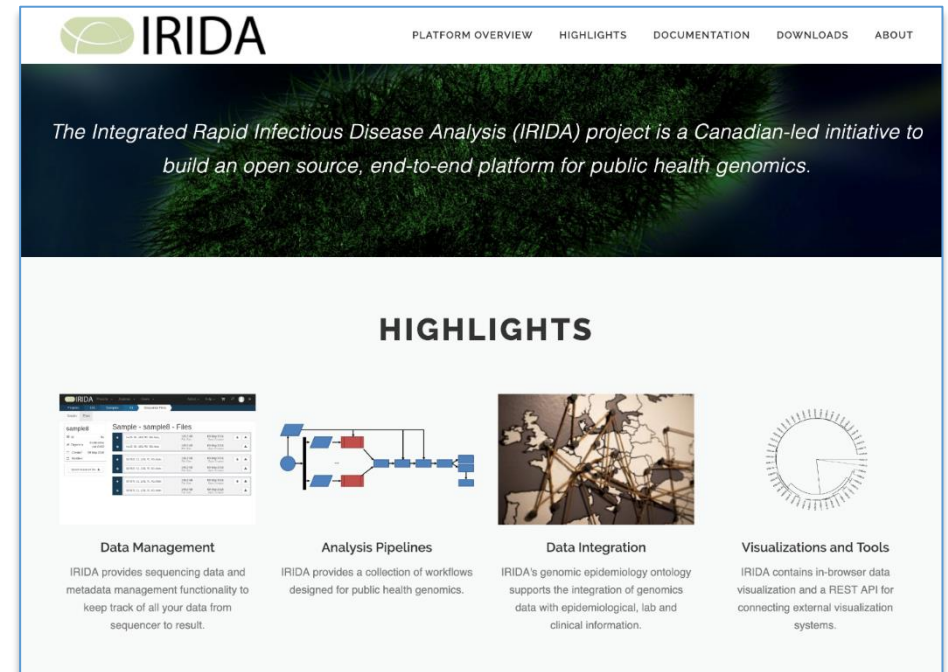
UDF/Freezer Location	UDF/Sample Type	UDF/Genus/VIGAS Project	UDF/Species	UDF/Project Account	UDF/Requested Sequencer	Sample/Name	Contai
----------------------	-----------------	-------------------------	-------------	---------------------	-------------------------	-------------	--------

Sample 1	Location 1	testperson@vetinst.no	WGS colonies - freezer	Escherichia spp	Escherichia coli	1124-Helgenomsekvenserir	MiSeq or NextSeq	Sample1	T
Sample 2	Location 2	testperson@vetinst.no	WGS colonies - freezer	Escherichia spp	Escherichia coli	1124-Helgenomsekvenserir	MiSeq or NextSeq	Sample2	T



# The bioinformatics end - IRIDA

- Created by Canadian Health authorities and other academic/research institutions
- Local install
- Web based analyses
- Group based access to data
- Integrates with HPC/slurm
- Galaxy - pipelines shareable
- Extendable



The screenshot shows the IRIDA website homepage. At the top left is the IRIDA logo, a green leaf-like shape next to the text "IRIDA". To the right of the logo is a navigation menu with links: "PLATFORM OVERVIEW", "HIGHLIGHTS", "DOCUMENTATION", "DOWNLOADS", and "ABOUT". Below the navigation is a dark blue banner with white text: "The Integrated Rapid Infectious Disease Analysis (IRIDA) project is a Canadian-led initiative to build an open source, end-to-end platform for public health genomics." Below the banner is a section titled "HIGHLIGHTS" in bold black text. This section contains four columns, each with a small image and a text block:



- Data Management:** The image shows a screenshot of a web interface with a table of data. The text below reads: "IRIDA provides sequencing data and metadata management functionality to keep track of all your data from sequencer to result."
- Analysis Pipelines:** The image shows a flowchart diagram with blue and red boxes connected by lines. The text below reads: "IRIDA provides a collection of workflows designed for public health genomics."
- Data Integration:** The image shows a map of the world with various locations marked. The text below reads: "IRIDA's genomic epidemiology ontology supports the integration of genomics data with epidemiological, lab and clinical information."
- Visualizations and Tools:** The image shows a circular diagram with a central point and lines radiating outwards. The text below reads: "IRIDA contains in-browser data visualization and a REST API for connecting external visualization systems."

# Data management

The screenshot displays the IRIDA web interface for managing Salmonella spp. data. The top navigation bar includes the IRIDA logo, 'Projects', 'Analyses', and a search bar. The breadcrumb trail shows 'Projects / Salmonella spp'. Below the breadcrumb, there are tabs for 'Samples', 'Line List', 'Analyses', 'NCBI Exports', 'Recent Activity', and 'Settings'. The 'Line List' tab is active. The main content area features an 'Export' dropdown, an 'Add to Cart' button, and a search bar. A table lists sample data with the following columns: Sample, Modified Date, Created Date, mlst/scheme, mlst/sequence\_type, and Reads\_QC/Coverage. The table contains 8 rows of data. At the bottom, it indicates 'No samples selected' and 'Displaying 478 of 478'.

<input type="checkbox"/>	Sample	Modified Date ↓	Created Date	mlst/scheme	mlst/sequence_type	Reads_QC/Coverage
<input type="checkbox"/>	<a href="#">2023-01-7117-1-1-1-1</a>	Aug 29, 2023, 3:52 PM	Aug 14, 2023, 1:00 PM	senterica	19	93
<input type="checkbox"/>	<a href="#">2023-01-6519-1-1-1-1</a>	Aug 29, 2023, 3:52 PM	Aug 14, 2023, 1:00 PM	senterica	40	62
<input type="checkbox"/>	<a href="#">2023-01-6941-1-1-1-1</a>	Aug 29, 2023, 3:52 PM	Aug 14, 2023, 1:00 PM	senterica	34	57
<input type="checkbox"/>	<a href="#">2023-04-21714-1-2-1-1</a>	Aug 29, 2023, 3:52 PM	Aug 28, 2023, 1:11 PM	senterica	128	50
<input type="checkbox"/>	<a href="#">2023-01-7560-1-1-1-1</a>	Aug 29, 2023, 3:25 PM	Aug 28, 2023, 1:11 PM	senterica	43	
<input type="checkbox"/>	<a href="#">2023-01-7715-1-1-1-1</a>	Aug 29, 2023, 2:01 AM	Aug 28, 2023, 1:11 PM	senterica	14	40
<input type="checkbox"/>	<a href="#">2023-01-6941-2-1-1-1</a>	Aug 29, 2023, 1:15 AM	Aug 28, 2023, 1:11 PM	senterica	4873	77

# Custom pipelines

Pipelines		 Running 10	 Queued 10
<b>Assembly and Annotation Pipeline</b>	<b>Assembly_QC Pipeline</b>	<b>MLST Pipeline NVI</b>	
Shovill assembly, Prokka annotation and QUASt assembly assessment	Under Test: "DO NOT USE THIS" pipeline. Pipeline performs Shovill assembly and QUASt assembly quality assessment. And, uses checkm_analyze and checkm_qa for assessing the quality of genomes.	Scan contig files against PubMLST typing schemes	
<b>Phylo-CoreGenomeSNP Pipeline</b>	<b>Reads_QC Pipeline</b>	<b>ResPointFinder</b>	
Under Test: "DO NOT USE THIS" pipeline. Phylogenetic analyses of prokaryotic isolates. Suitable only for datasets where the samples are expected to be very closely related. Here, multiple sequence alignment is generated with ParSNP, which takes the whole genome into account. This pipeline is similar to ALPPACA on	Runs Kraken2/Bracken for species confirmation/contamination-check and fastqc for basic quality control parameters.	Uses ResFinder and PointFinder tools to map the Fastq reads to ResFinder and PointFinder databases and produces a combined formatted output as well.	

## Function

Sample submission	Lab tracking	Sequencing Demult.plx	Compute	Workflow engine	Workflow run starter/ monitor	Metadata db/raw data storage	Results database	Interpret.n. platform
-------------------	--------------	--------------------------	---------	-----------------	----------------------------------	------------------------------	------------------	-----------------------

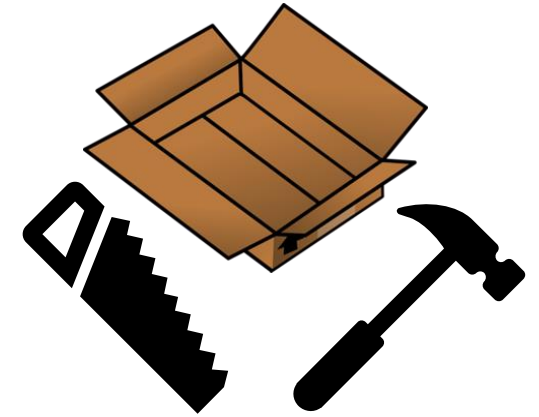
---

## Implementation

Lablink	Clarity	Illumina sequencers Local servers	NREC/ OpenStack	Galaxy	IRIDA Galaxy mariadb	Disk storage postgresql	IRIDA postgresql	IRIDA
---------	---------	--------------------------------------	--------------------	--------	----------------------------	----------------------------	---------------------	-------

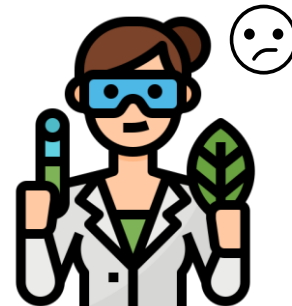
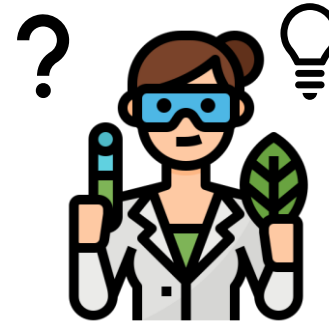
# Systems are IKEA flatpacks

- Modules premade
- But: nothing "works" out of the box
- Skilled lab techs/IT/bioinformaticians needed to:
  - Adapt systems to local instruments and infrastructure
  - Adapt analyses to local analysis requirements



# Two models

- Core facility model
  - Bioinformaticians analyze
  - Domain expert get static report
  - No option for DE to ask more without help
- Domain expert model
  - Bioinformaticians create analyses
  - Domain expert run analysis
  - Domain experts ask questions via analyses



# Training is paramount

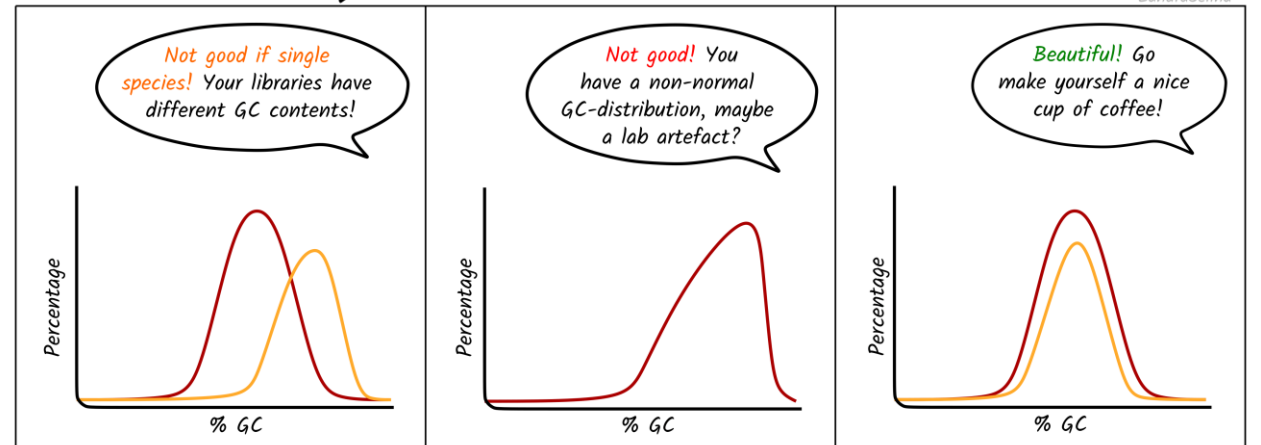
- *Listeria monocytogenes* - ST35



# Training is paramount

- *Listeria monocytogenes* - ST35
- Double GC content peak
- N50 – 15 000
- #Contigs – 453

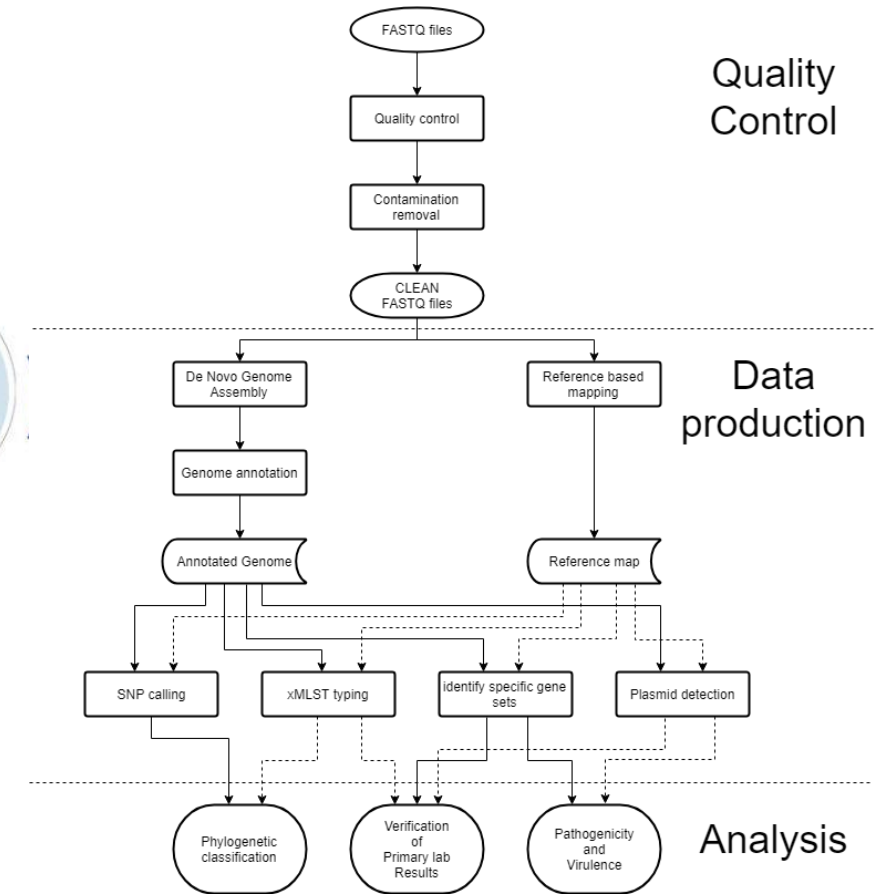
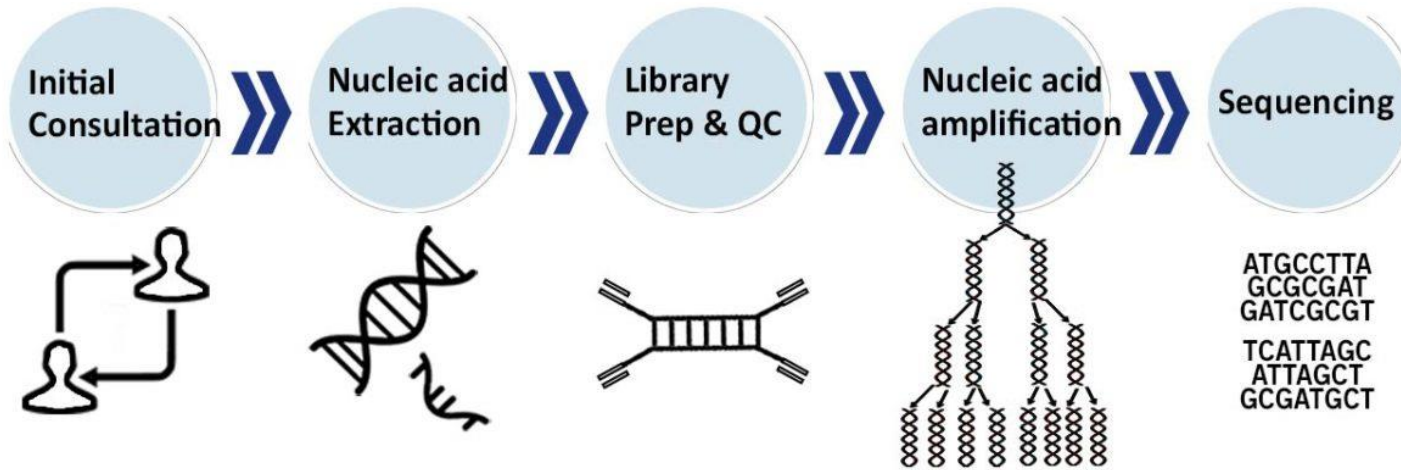
## FASTQC - Per sequence GC-content



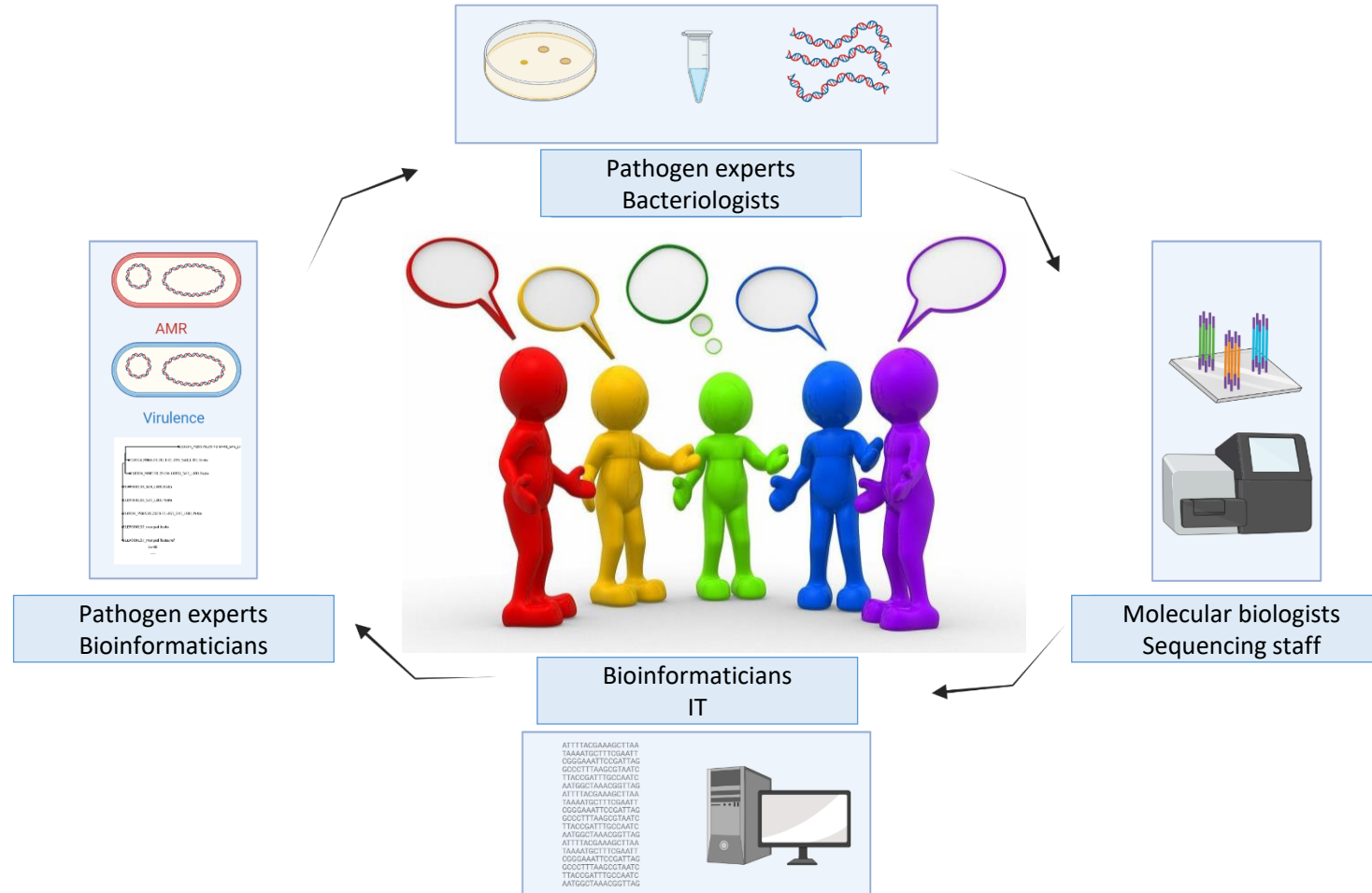


# The whole picture

## NGS Pipeline



# Sequencing is a circle



# Main working group



Magnus Leithaug  
Molecular biology  
lab



Camilla Sekse  
Bacteriology



Eve Fiskebeck  
Bioinformatics  
Phylogeny



Jeevan Karloss  
Antony-Samy  
Bioinformatics  
Programmer



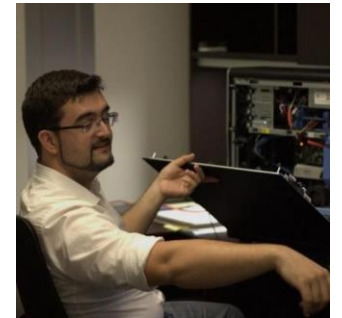
Jannice Schau  
Sletteemeås  
Antimicrobial  
resistance



Cathrine  
Arnason Bøe  
Molecular biology  
lab



Thomas Haverkamp  
Bioinformatics  
Metagenomics



George Marselis  
Sysadmin  
Programmer

*Faglig ambisiøs, fremtidsrettet og  
samspillende - for Én helse!*



**Veterinærinstituttet**  
— *Norwegian Veterinary Institute*

[www.vetinst.no](http://www.vetinst.no)