



# Uncover bacterial pathogen identity using Nanopore MinION

A collaboration between WorldFish, The University of Queensland, WilderLab, Centex and GeneSEQ

Online virtual workshop 9th of August 2021



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB



# Inspire Challenge project



Platform for  
Big Data  
in Agriculture

[ABOUT](#) ▾

[INSPIRE CHALLENGE](#) ▾

[SHARED SERVICES](#)

[COMMUNITIES OF PRACTICE](#) ▾

[NEWS & EVENTS](#) ▾

[RESOURCES](#) ▾



## 2019 WINNER

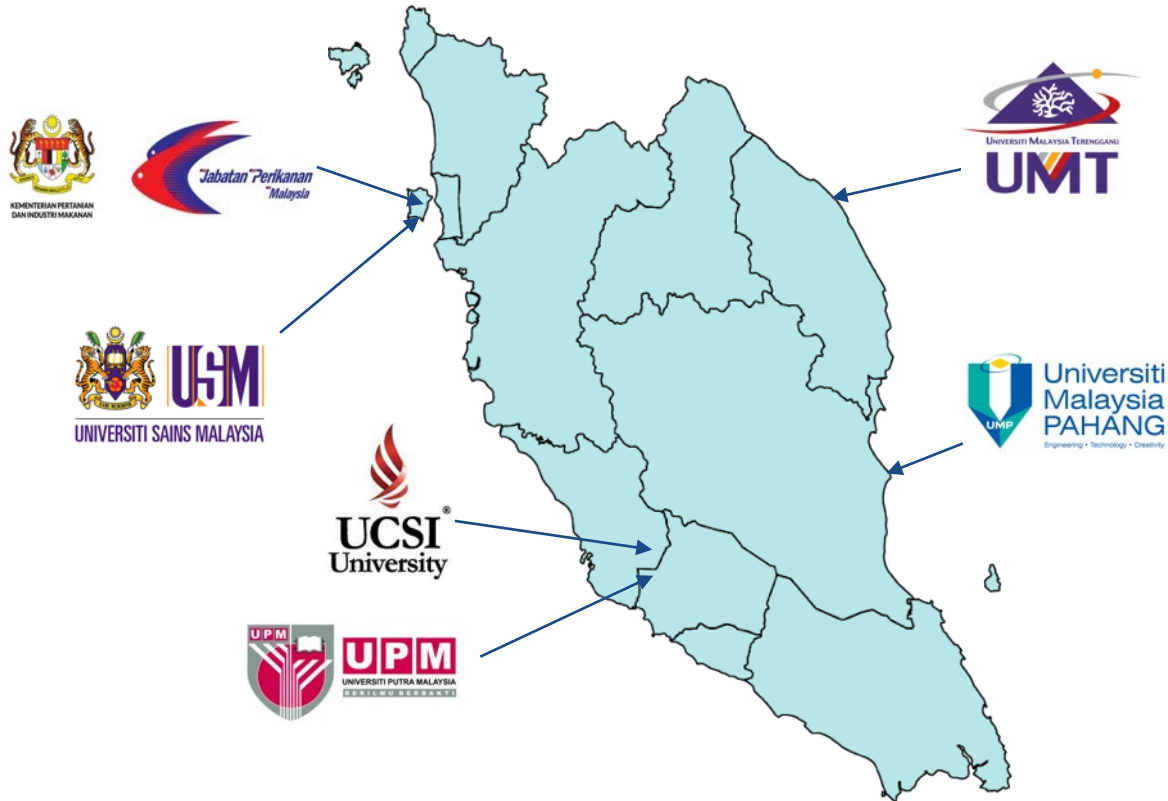
Rapid genomic detection  
of aquaculture  
pathogens



Malaysia, Bangladesh

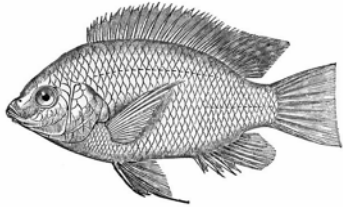


# Malaysian institutions and participants



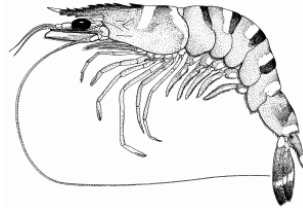
# Hosts & pathogens from Malaysia

***Oreochromis* sp.**



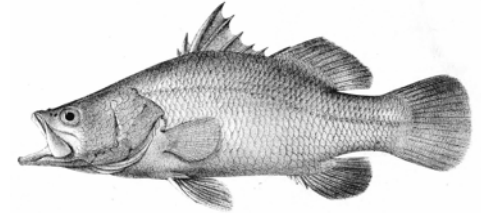
*Streptococcus agalactiae*

***Penaeus monodon***



*Vibrio parahaemolyticus* / *V. owensii* / *V. harveyi*

***Lates calcarifer***



*Vibrio parahaemolyticus*

**Seawater**



*Photobacterium* sp. / *Bacillus* sp.  
*Vibrio sagamiensis*

**Bioflocs**



*Enterobacter* sp.

***Holothuria leucospilota***



unknown

**Sediment**

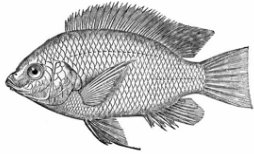


*Bacillus cereus*

# Hosts & pathogens from other parts of the world

## Tilapia

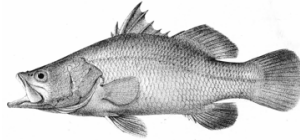
*Oreochromis* sp.



*Streptococcus* sp.; *Edwardsiella* sp.;  
*Aeromonas* sp.; *Vibrio* sp.; Infectious Spleen  
and Kidney Necrosis Virus (ISKNV); Tilapia  
lake virus (TiLV)

## Barramundi

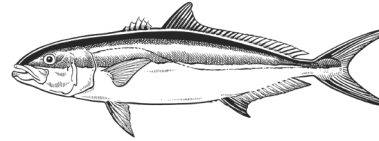
*Lates calcarifer*



*Vibrio* sp.; *Streptococcus iniae*

## Kingfish

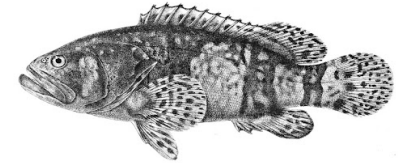
*Seriola lalandi*



*Photobacterium damsela*

## Qld giant grouper

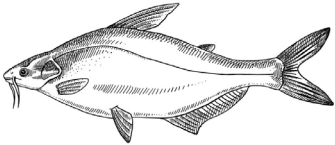
*Epinephelus lanceolatus*



*Streptococcus agalactiae*

## Pangasius

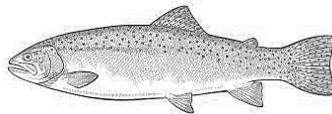
*Pangasius* sp.



*Aeromonas hydrophila*

## Rainbow trout

*Oncorhynchus mykiss*



*S. iniae*; *Yersinia ruckeri*

## Atlantic salmon

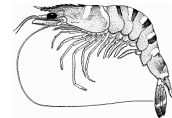
*Salmo salar*



*Y. ruckeri*; *Tenacibaculum maritimum*

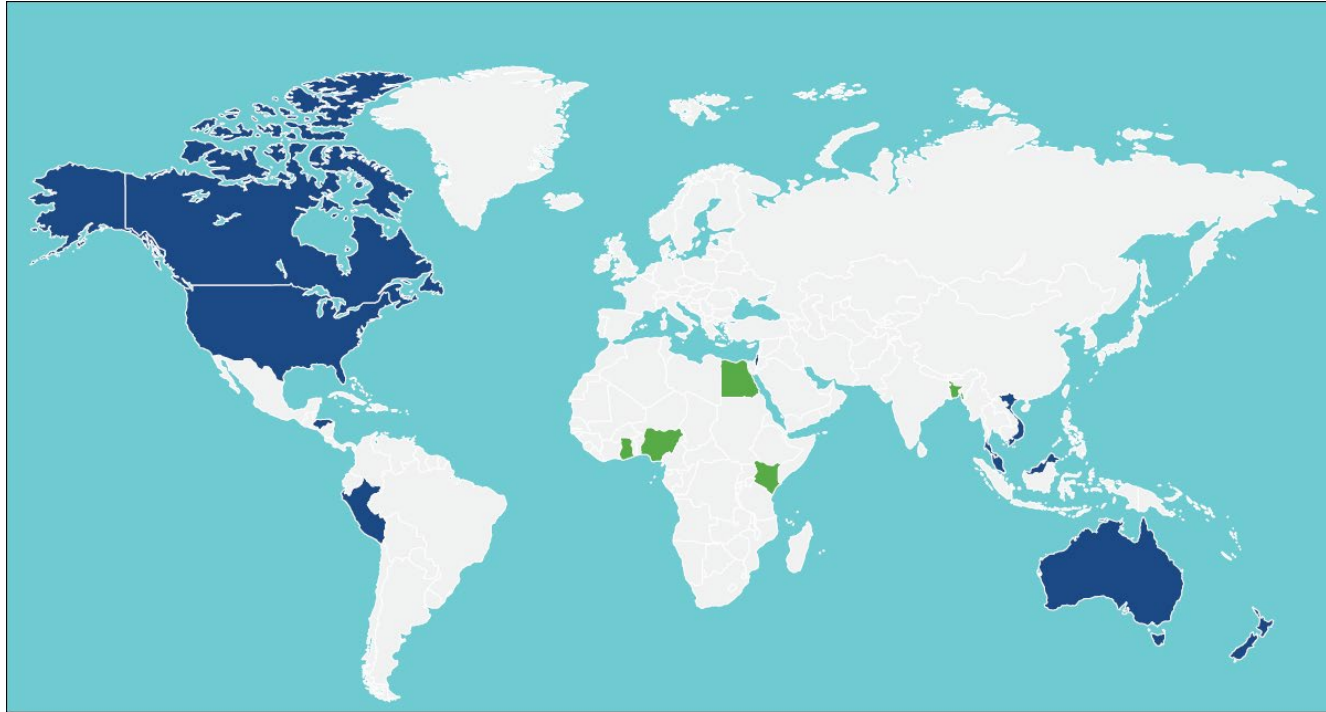
## Shrimps

*Penaeus monodon/vannamei*



*Vibrio* sp.; *Aeromonas* sp.

# Origins of the pathogens



Aquaculture pathogens sequenced  
in this project

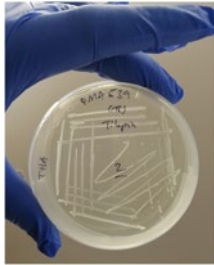
Honduras, Ecuador, Peru, USA,  
Canada, Israel, Australia, New-  
Zealand, Vietnam, Thailand, Malaysia



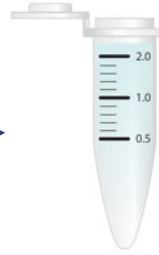
Future works using Nanopore

Bangladesh, Egypt, Ghana, Nigeria  
and Kenya

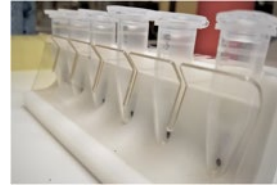
# Bacterial samples DNA extraction and Illumina sequencing



Participants  
bacterial isolates



bacterial suspension  
inactivated in 100%  
ethanol

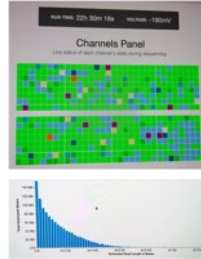
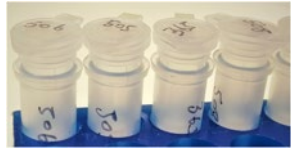


DNA extraction/  
quantification



Illumina sequencing  
(short reads)

# DNA samples for Nanopore sequencing



Raw Nanopore and Illumina  
data combined for analysis

# Today's agenda

---

## Part 1

- Sequencing technologies from Sanger to Nanopore
- Applications

## Part 2

- Sample collection & DNA extraction

## Part 3

- DNA library prep for Nanopore sequencing

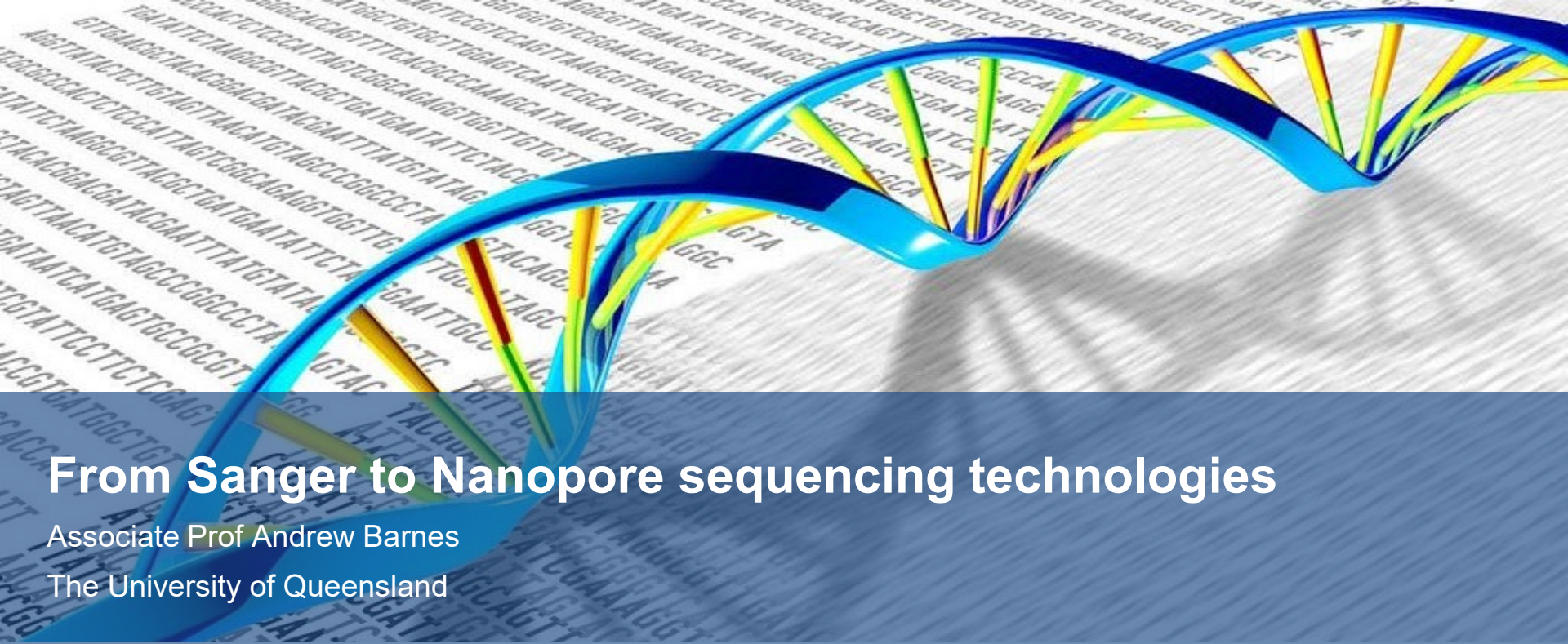
## Part 4

- Bioinformatics analyses

## Part 5

- Participant Data Overview



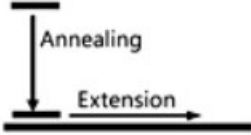

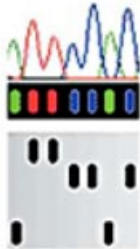


# From Sanger to Nanopore sequencing technologies

Associate Prof Andrew Barnes  
The University of Queensland

# First generation sequencing: Sanger technology



		Reagents	Reaction	Products	Capillary electrophoresis	Performance
First-generation sequencing	Sanger technology	Primers Template + dNTPs Enzyme + Dye-labeled terminators <b>A C G T</b>				Applied Biosystems® 3500 Series Genetic Analyzers  Read length up to 850 bp  Maximum throughput 138,000–414,000 bp/day  Run time at least 30 min

<https://doi.org/10.1159/000477808>

# 2G NGS platforms: Life Technologies Roche

NGS: Next Generation Sequencing



<http://www.marbigen.org/content/ffx-454-sequencer>

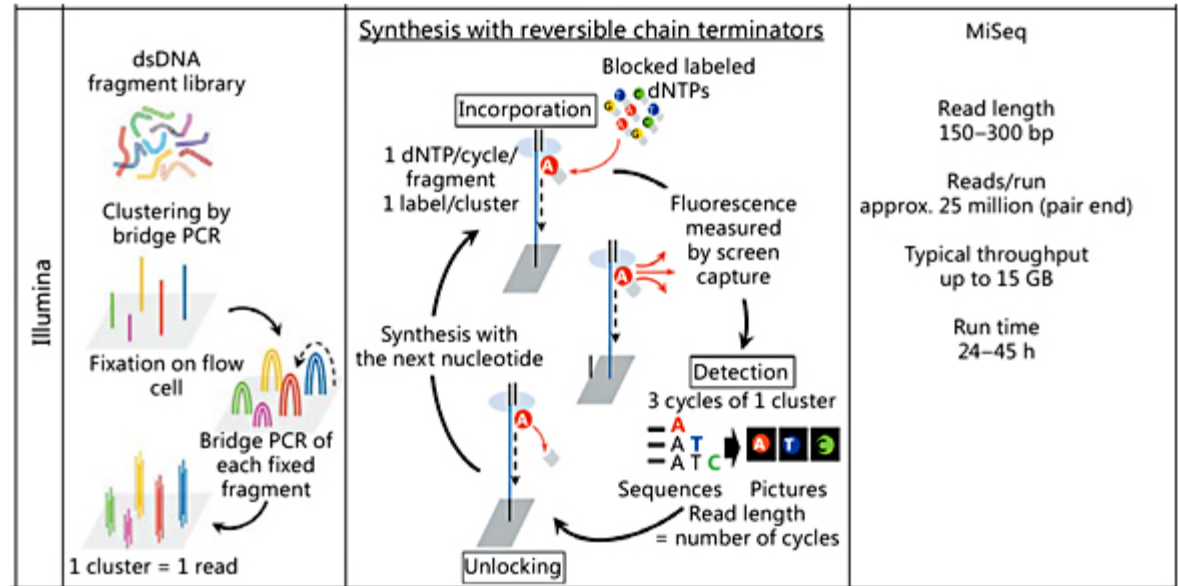


<https://tools.thermofisher.com/content/sfs/brochures/PGM-Specification-Sheet.pdf>

	Library and clonal amplification	Reaction	Performance
Roche	<p>ssDNA fragment library</p> <p>Fixation on beads</p> <p>Emulsion PCR</p> <p>1 fragment = 1 bead = 1 read</p>	<p><b>Pyrosequencing</b></p> <p>Sulfurylase</p> <p>PPi + APS → ATP</p> <p>Luciferase</p> <p>ATP → Light</p> <p>Degradation with apyrase</p> <p>Light measurement</p>	<p>454 Roche GS FLX+ systems</p> <p>Read length up to 600 bp</p> <p>Reads/run 700,000–1,000,000</p> <p>Typical throughput 450–700 MB</p> <p>Run time 10–23 h</p>
Life Technologies	<p>1 fragment = 1 bead = 1 read</p>	<p><b>pH-based sequencing</b></p> <p>H<sup>+</sup></p> <p>ΔpH</p> <p>ΔQ</p> <p>Oxidation of metal sensing layer</p> <p>Generation and recording of electric signal</p> <p>H<sup>+</sup>: protons Q: charge V: voltage</p>	<p>Ion Personal Genome Machine® (PGM™) System</p> <p>Read length 35–400 bases (average: 200 bases)</p> <p>Reads/run approx. 400,000–5,000,000</p> <p>Typical throughput up to 2 GB</p> <p>Run time approx. 2 h</p>

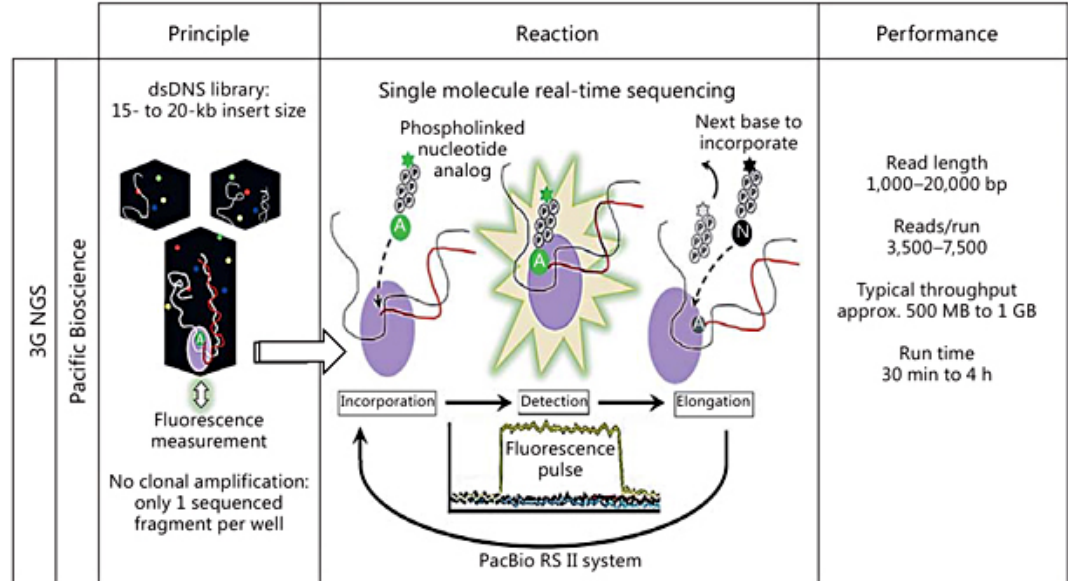
<https://doi.org/10.1159/000477808>

# 2G NGS platforms: Illumina



<https://doi.org/10.1159/000477808>


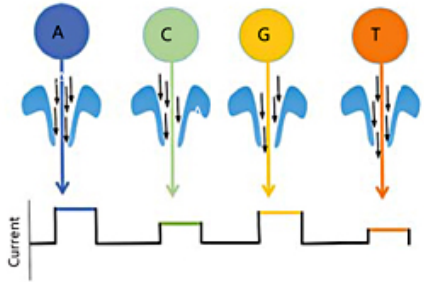
# 3G NGS Pacific Bioscience



<https://doi.org/10.1159/000477808>

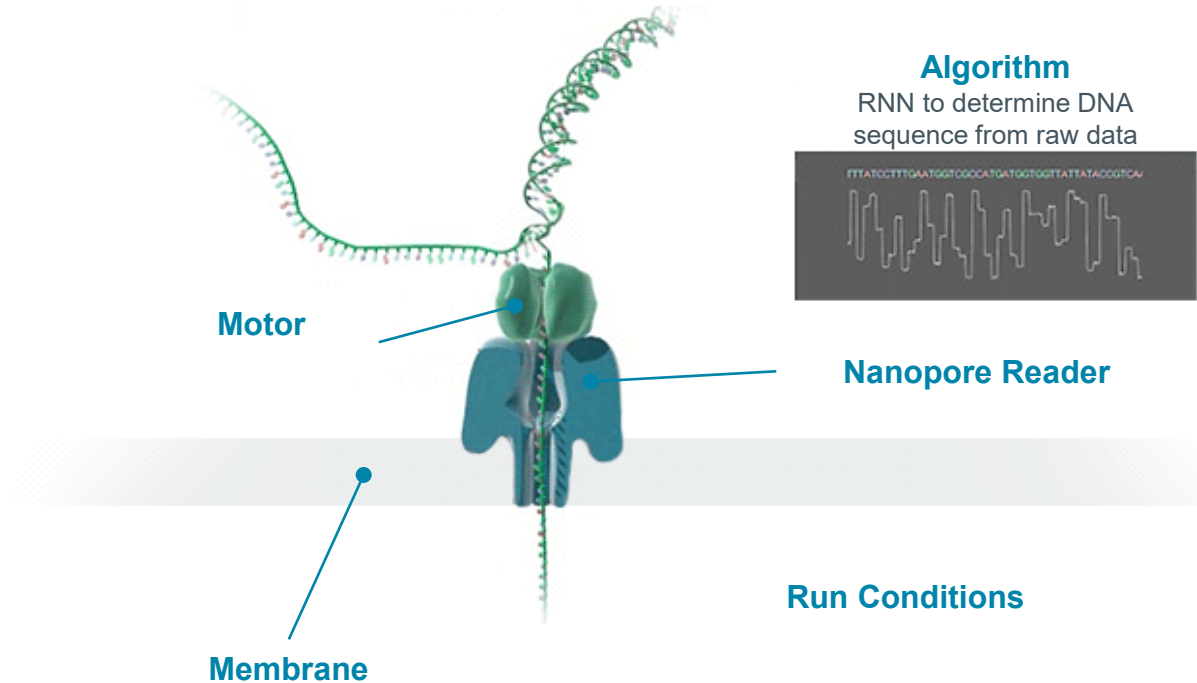
# 3G NGS Oxford Nanopore



		Principle	Reaction	Performance
4G NGS	Oxford Nanopore	<p>dsDNA library: 15- to 20-kb insert size</p>  <p>No clonal amplification A single molecule per nanopore per well</p>	<p>Single molecule real-time sequencing</p>  <p>Nucleotides pass through the pores and modify the ionic current</p> <p>Base calling → Measurement of the modulation</p>	<p>MinION</p> <p>Read length 5,000–200,000 bp</p> <p>Reads/run &gt;1,000,000 reads</p> <p>Typical throughput &gt;1 GB</p> <p>Run time depends on the input speed: 1 bp/ns</p> <p>Weight 90 g</p>

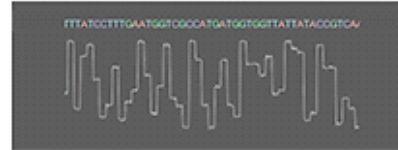
<https://doi.org/10.1159/000477808>

# Nanopore DNA/RNA sequencing - how does it work?



## Algorithm

RNN to determine DNA sequence from raw data



Multiple nanopore sensors arrayed in one device sequencing in parallel

Operate independently but at the same time

A top-down view of a large number of fish, likely koi or goldfish, swimming in a body of water. The fish are densely packed, moving in various directions. The water is a deep blue color, and there are many small, light-colored bubbles visible throughout the scene. The overall image has a strong blue tint.

# **Application of NGS in aquaculture**

# Fish farmers' questions during disease outbreak

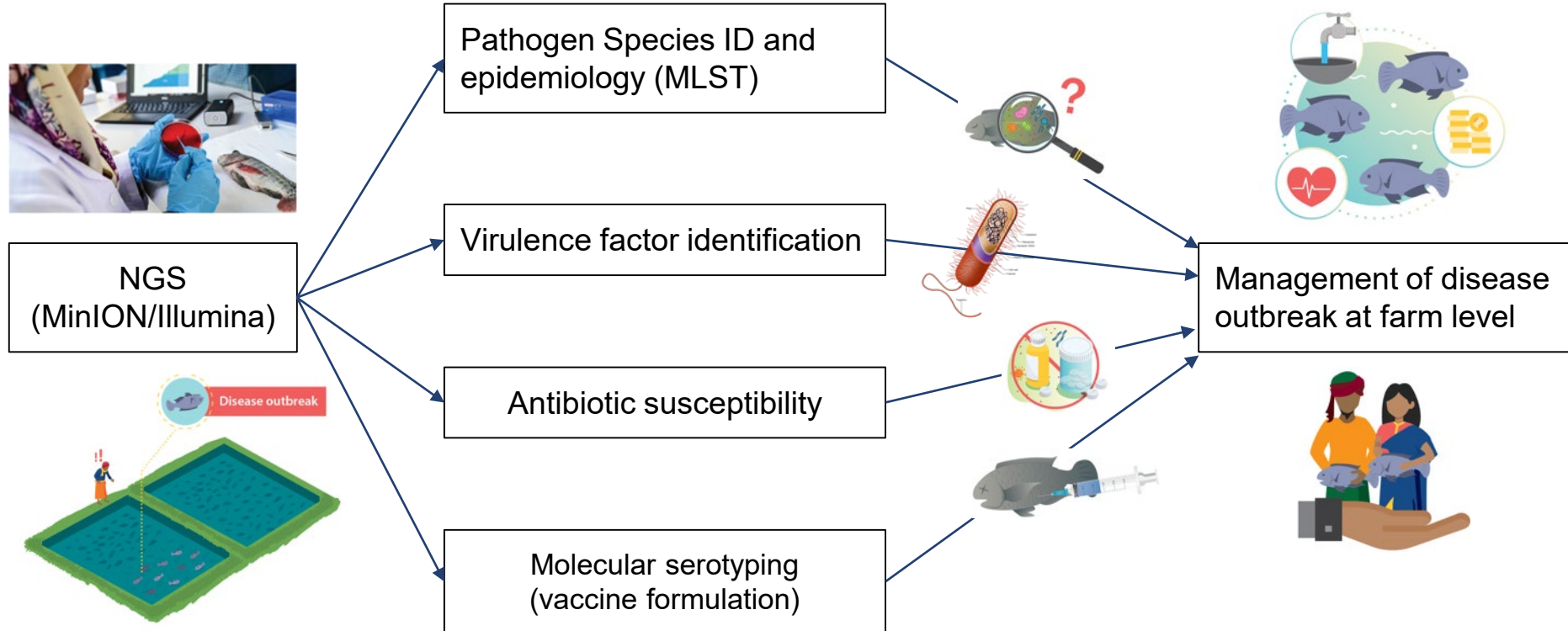
---

What is  
killing my  
fish?

How can I  
treat them?

How can I  
prevent  
disease in  
the next  
crop?

# Genome-based diagnosis of pathogens in aquaculture





# Sample collection & DNA extraction

Dr Jerome Delamare-Deboutteville

WorldFish



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB



# Sampling materials for fish bacteriology



Anesthetic (e.g. MS-222, benzocaine, clove oil).



Spray or squirt bottle filled with suitable disinfectant (e.g. normal grade 70% ethanol).



Sterile single-use loops, plain cotton swabs or reusable metal loops.



Plastic beaker (50–100 mL) to disinfect instruments in 70% ethanol.



Post-mortem sheet, clean tissue paper or aluminium fold to create a clean surface to sample the fish.



Sterile dissection kit (scissors, scalpel, forceps, tweezers etc.).



Trypticase Soy Agar (TSA) plates (or any other bacteriological agar as per study requirement).



Box, transport container, strong resealable plastic bags to pack and transport samples.



Paper towels (enough for each fish and to clean surfaces and instruments).



Portable Bunsen burner to create a sterile environment around the sampling area and to sterilize dissection instruments between fish.

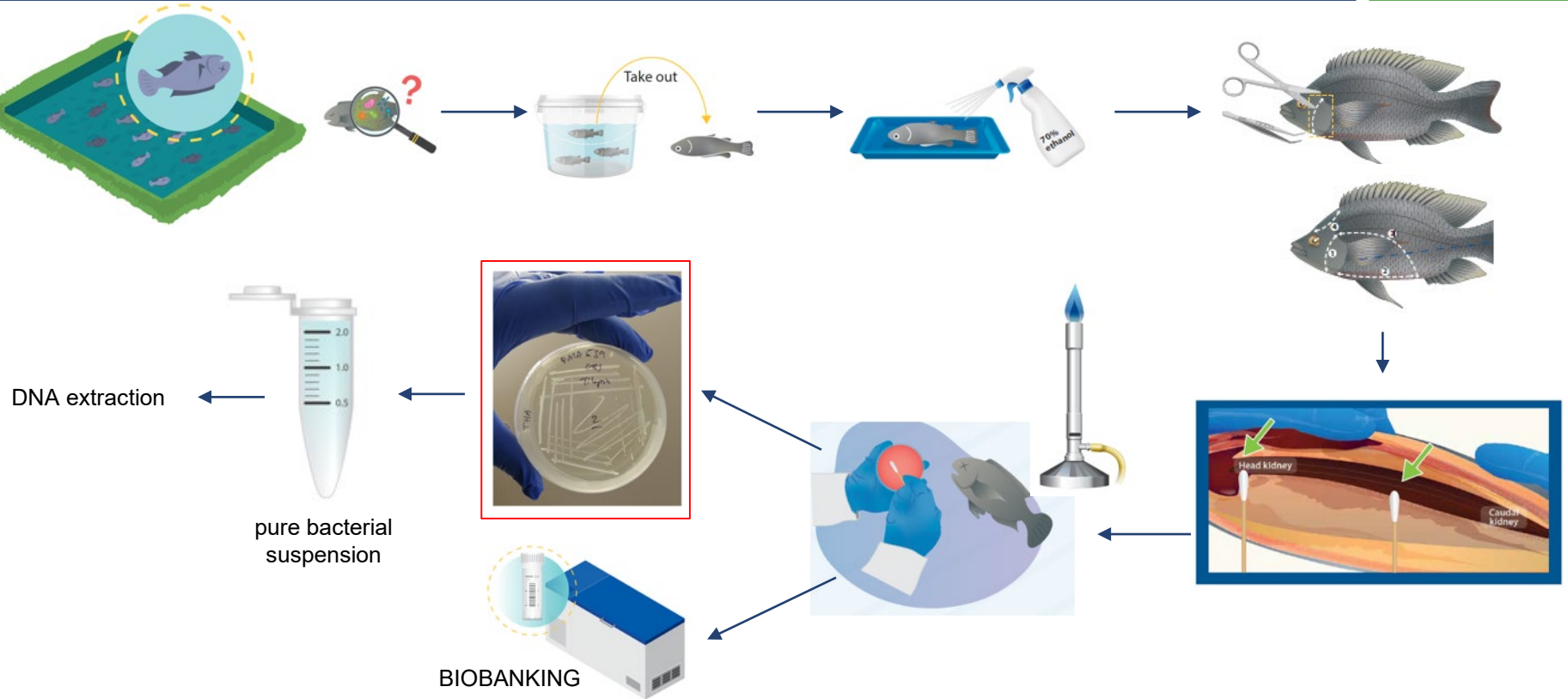


Bacteriological transport swabs (if no TSA plates to be used for field sampling).



Plastic biohazard clinical bags for waste disposal.

# Bacteriology sampling from diseased fish

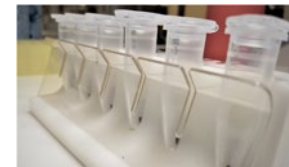
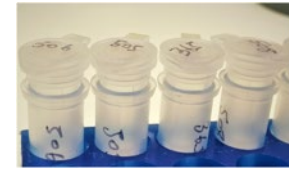
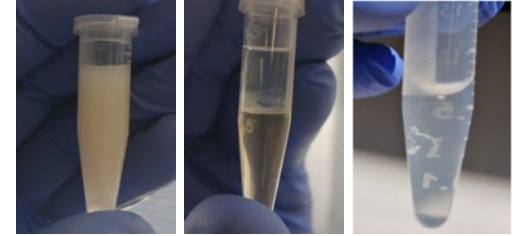


# Cells lysis methods

	Advantages	Disadvantages
SDS+Proteinase K	Overall good performer for DNA extraction	May lead to co-precipitation of carbohydrate
CTAB	Sample with high polysaccharide (e.g. capsulated/mucoid microbe)	CTAB is detrimental to environmental
Lysozyme pretreatment +SDS	Suitable for gram positive microbe	May not work across all gram positive microbe
Mechanical disruption (Bead beating)	Samples with tough/thick cell wall	Can lead to fragmentation of DNA (higher smearing)

# DNA purification methods

	Advantages	Disadvantages
Phase separation (chloroform extraction)	High DNA yield and integrity Cheap	Requires equipment, generates toxic chemicals (chloroform and phenol), relatively time consuming
Column-based separation	Fast and convenient.	Requires equipment, expensive, often low-yield and molecular weight and require multiple centrifuge steps. Not scalable
Magnetic silica/ carboxylated beads	Scalable, fast and convenient, only magnet needed, high DNA yield and integrity	High cost of commercially produced beads



# DNA concentration



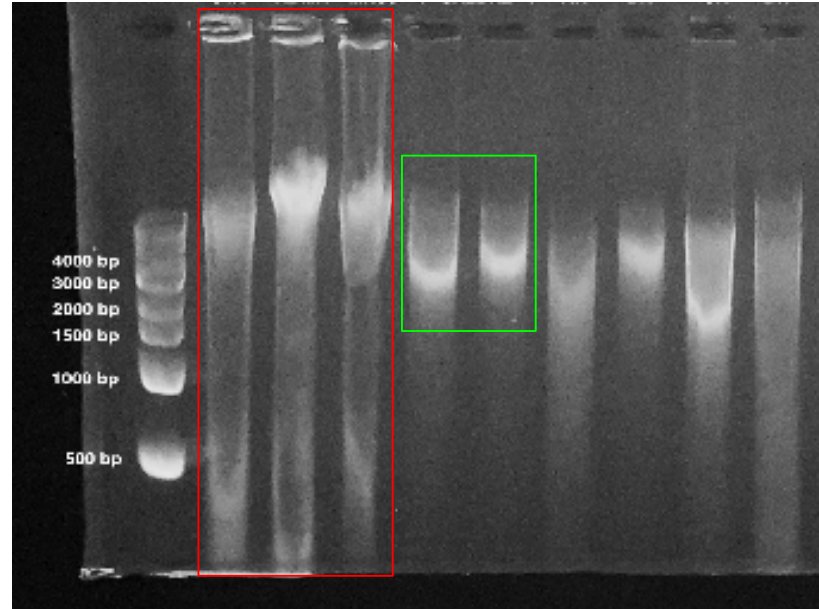
Qubit Fluorometer. Measure the concentration of dsDNA based on the fluorescence emitted by proprietary dsDNA-specific binding dye.



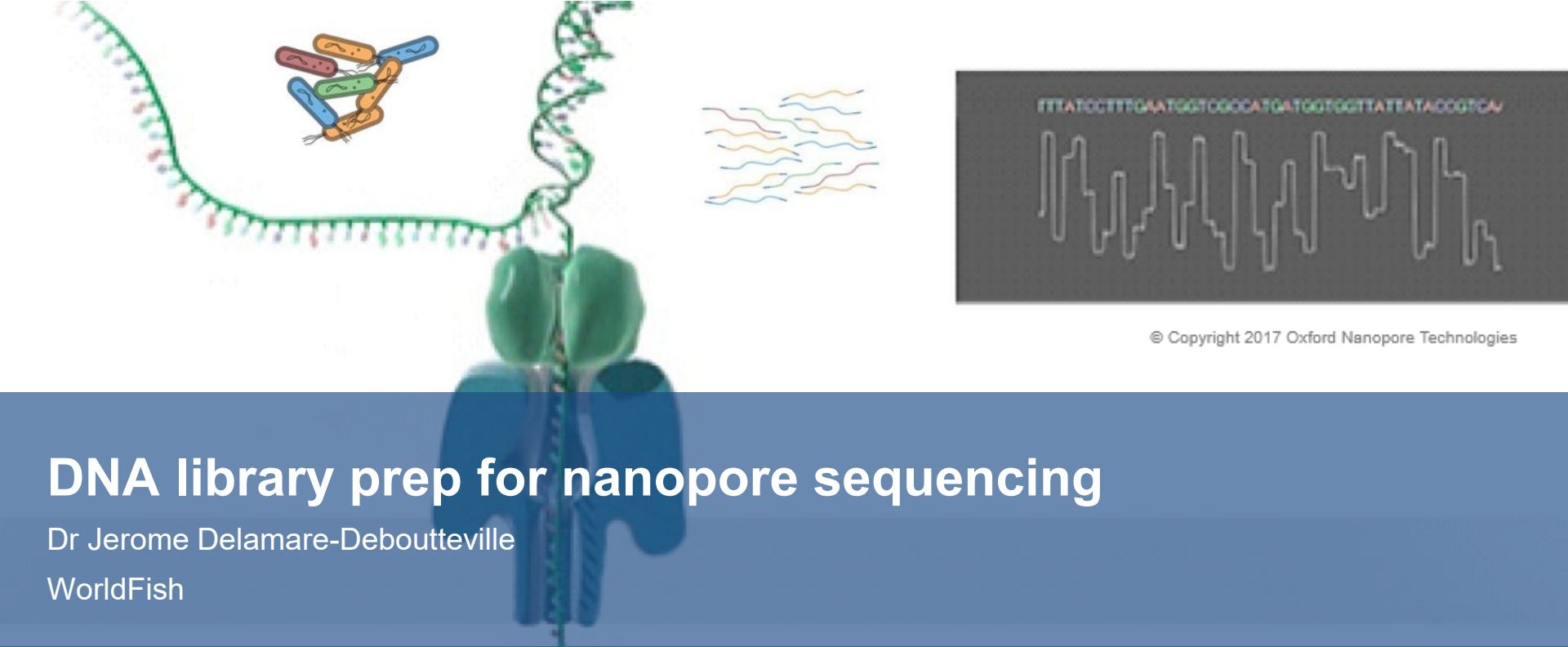
# DNA quality control



Nanodrop spectrophotometer to estimate DNA purity based on absorbance measurement.



DNA integrity assessed by agarose gel electrophoresis.



© Copyright 2017 Oxford Nanopore Technologies

# DNA library prep for nanopore sequencing

Dr Jerome Delamare-Deboutteville

WorldFish



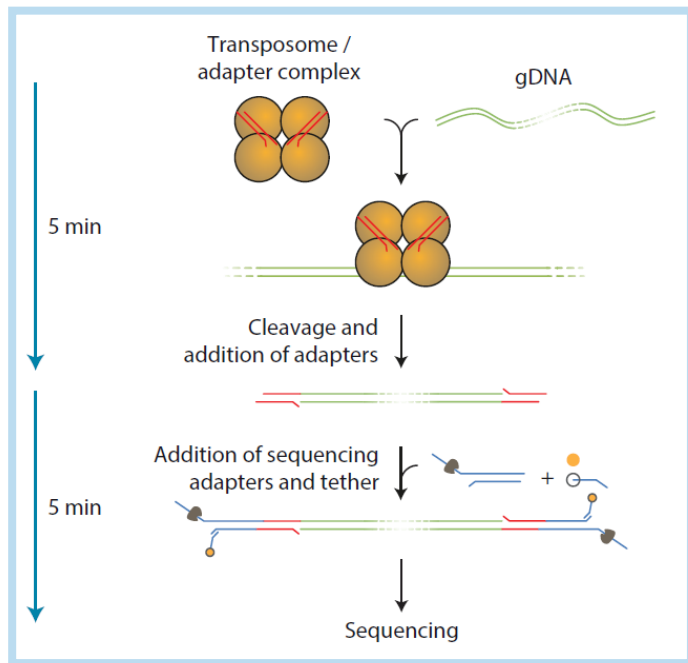
THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB



# Library Preparation: Transposase Approach (Rapid Kit)



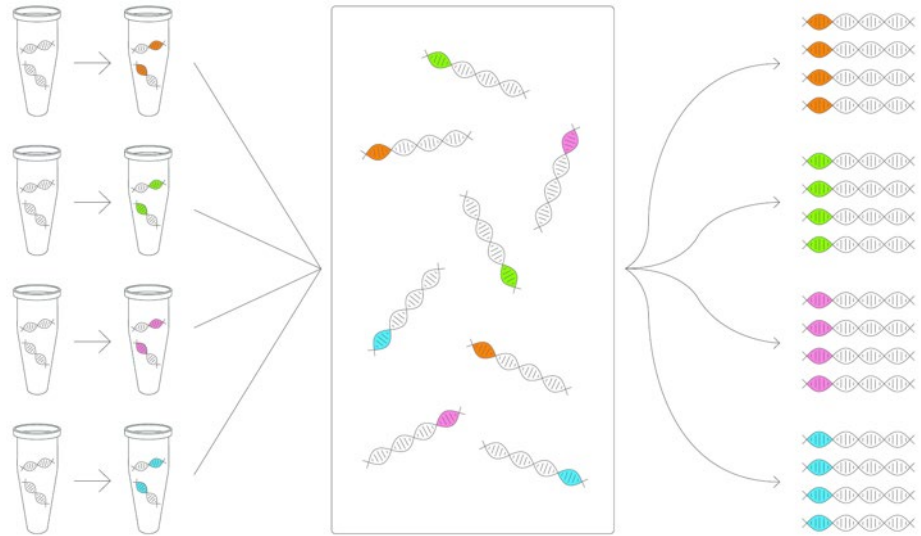
© Copyright 2017 Oxford Nanopore Technologies

- Very rapid library generation with minimum laboratory requirement
- Fragmentation: transposase based
- Input amount: 400 ng HMW gDNA
- Multiplexing options
- Can generate several Gbases in a run (~3-5Gb)

# Barcoding - Multiple Sample in One Nanopore Run

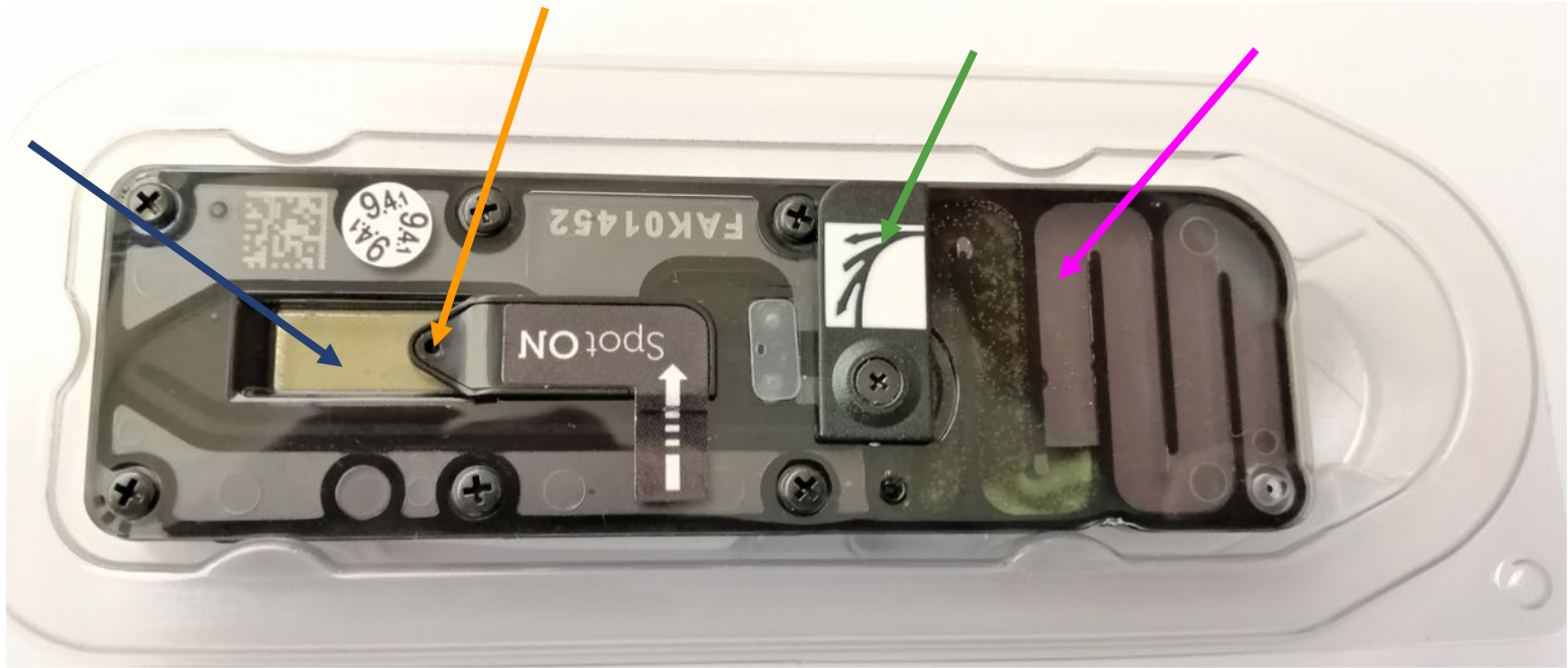
Kits are available to multiplex several samples, to maximise flow cell usage.

Up to 96 barcodes are supplied in Oxford Nanopore kits.

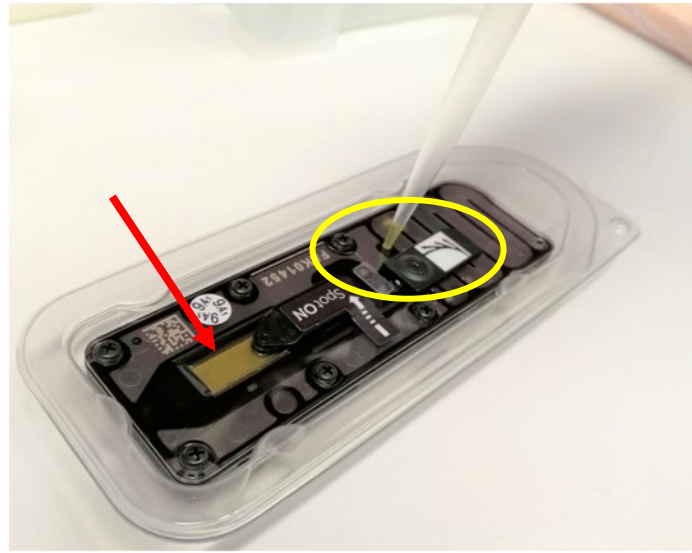


Barcode multiple samples → Pool and sequence → Separate and analyse

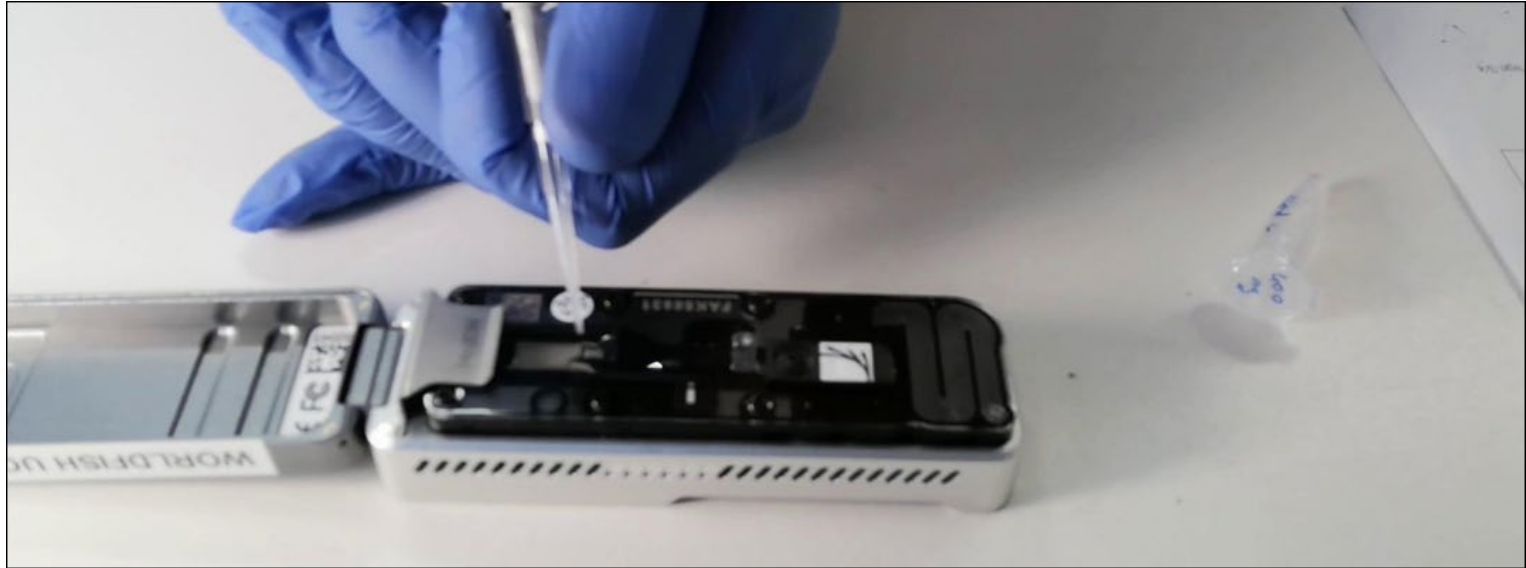
# A functional MinION Flowcell for Sequencing



# Flow cell Priming



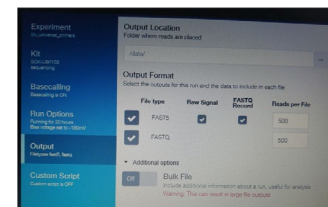
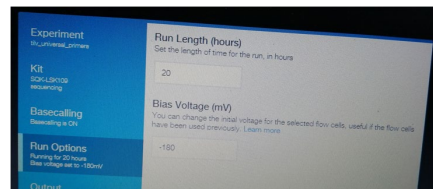
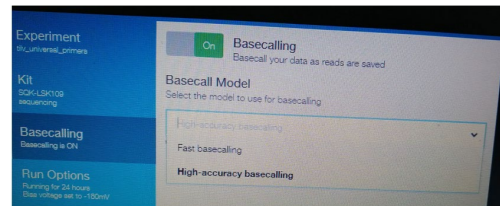
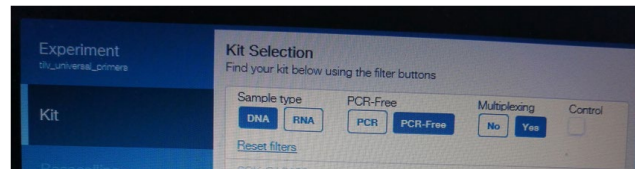
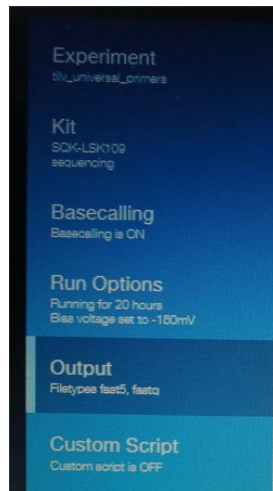
# Library loading on Minion flow cell



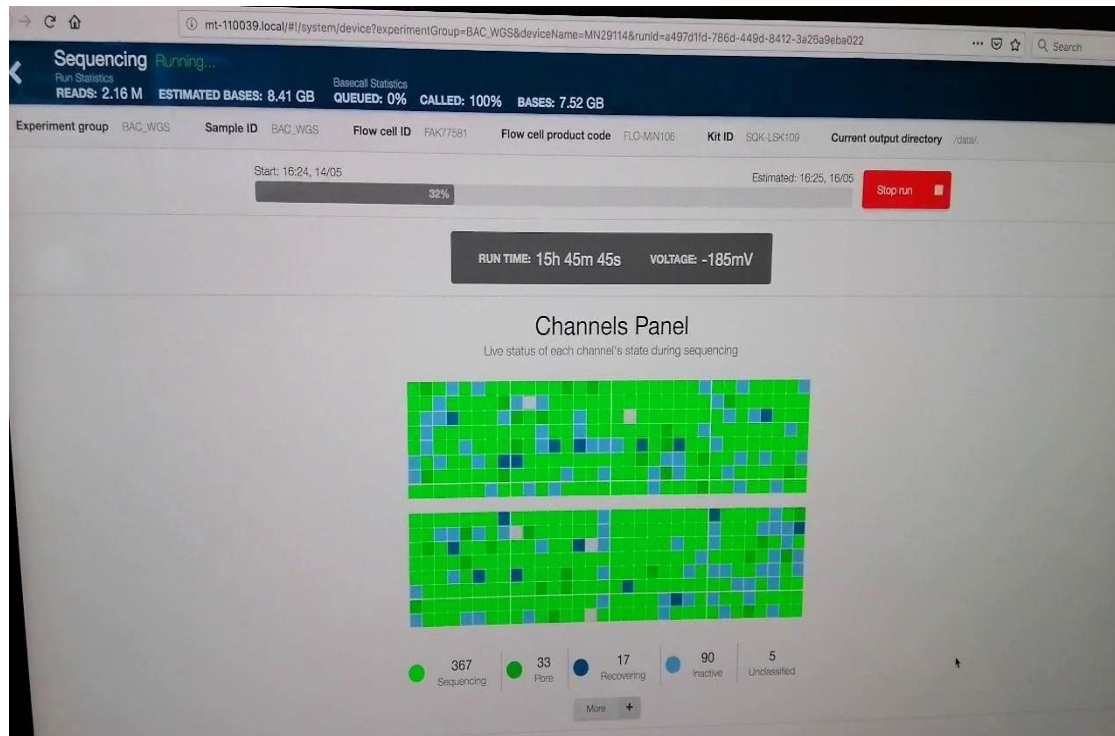
# Library loaded



# Starting a sequencing run



# Sequencing run



```

ptions:
-i, --in1          read1 input file name (string [=])
-o, --out1         read1 output file name (string [=])
-I, --in2          read2 input file name (string [=])
-O, --out2         read2 output file name (string [=])
--unpaired1       for PE input, if read1 passed QC but read2 not, it will be written to unpaired1. Default is to discard it
--unpaired2       for PE input, if read2 passed QC but read1 not, it will be written to unpaired2. If --unpaired2 is same c
unpaired1 (default mode), both unpaired reads will be written to this same file. (string [=])
--failed_out      specify the file to store reads that cannot pass the filters. (string [=])
-m, --merge       for paired-end input, merge each pair of reads into a single read if they are overlapped. The merged read
s disabled by default.
--merged_out      in the merging mode, specify the file name to store merged output, or specify --stdout to stream the merg
output (string [=])
--include_unmerged in the merging mode, write the unmerged or unpaired reads to the file specified by --merge. Disabled by c
lt.
-6, --phred64     indicate the input is using phred64 scoring (it'll be converted to phred33, so the output will still be p
33)
-z, --compression compression level for gzip output (1 ~ 9). 1 is fastest, 9 is smallest, default is 4. (int [=4])
--stdin          input from STDIN. If the STDIN is interleaved paired-end FASTQ, please also add --interleaved_in.
--stdout        stream passing-filters reads to STDOUT. This option will result in interleaved FASTQ output for paired-e
nt. Disabled by default.

```

# Bioinformatics: Simplifying Big Data

Dr Gan Han Ming

GeneSEQ



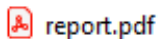
THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB



# Nanopore Raw Data: Sequencing report



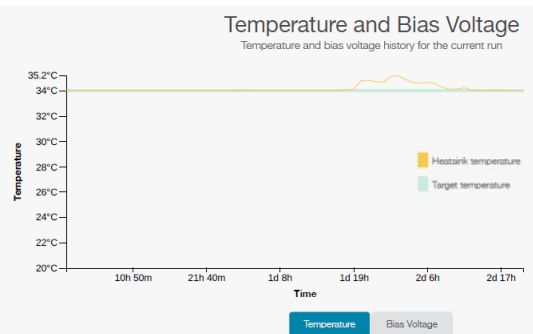
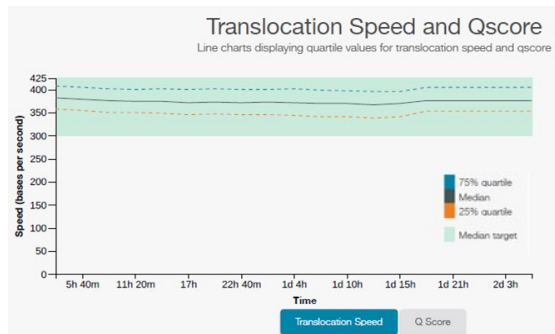
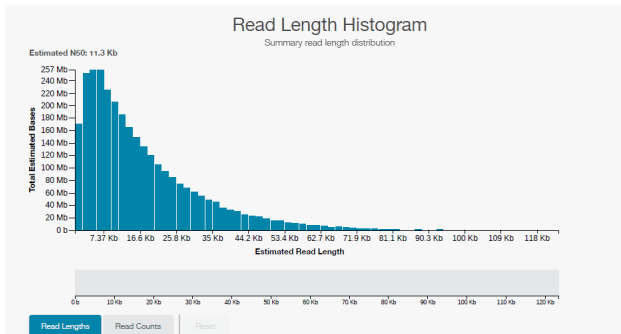
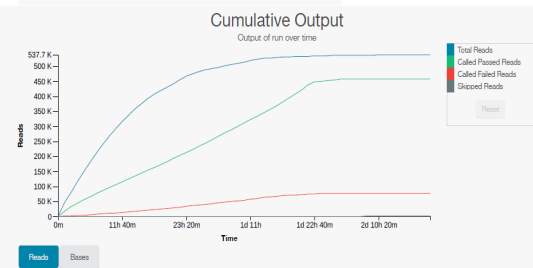
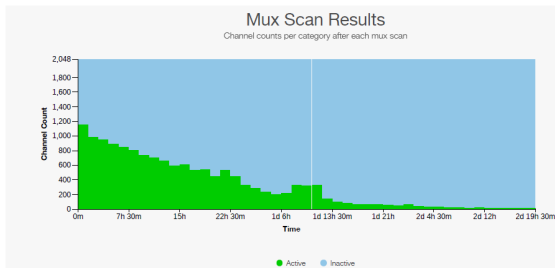
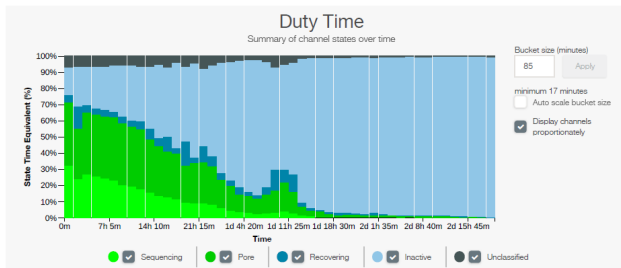
**Sequencing** Stopped by user

Run Statistics  
READS: 537.7 K ESTIMATED BASES: 3.06 Gb

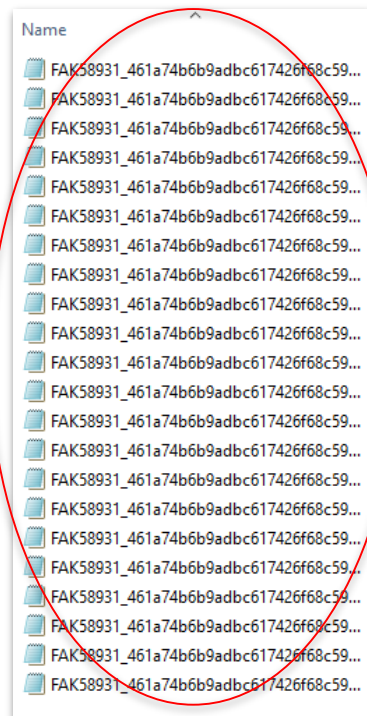
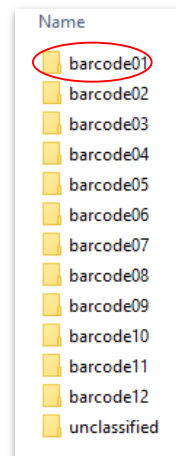
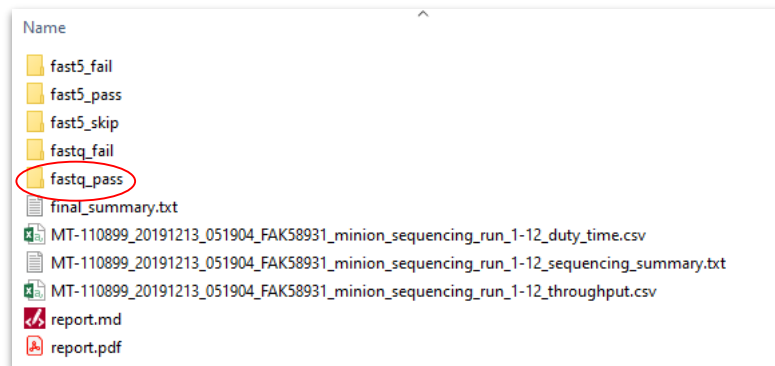
Basecall Statistics  
QUEUED: 1% CALLED: 99% BASES: 2.95 Gb SKIPPED READS: 2.26 K

Position MN32043 Experiment group inspire\_run2 Sample ID 1-12 Flow cell ID FAK38931 Current output directory /data Basecall model High-accuracy basecalling

TOTAL RUN TIME 2d 19h 37m 9s



# Nanopore Raw Data: File structure



# Complexity to Simplicity

---



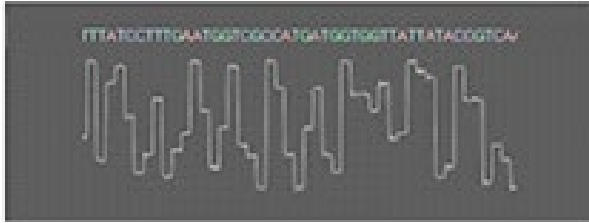
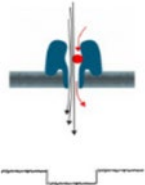
Raw Data



Usable (or “edible”) Information

## Conversion of Raw Signal Data (fast5) into FastQ file

## Biological nanopore



- Basecaller - **Machine Learning**  
Algorithm to determine DNA sequence  
from raw data signal

Source: Nanopore

[illegible]

# Different Basecaller = Different Sequence Accuracy

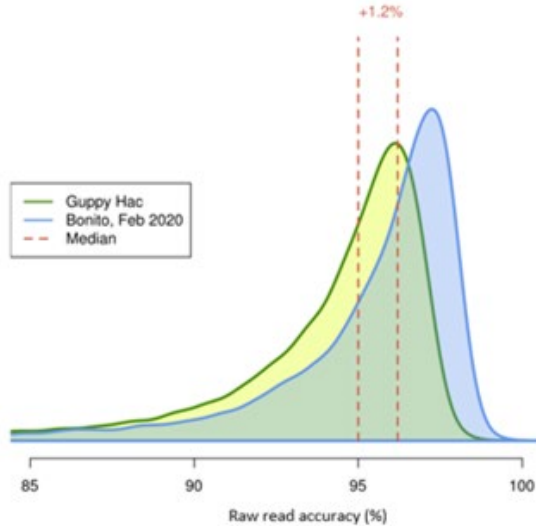


Figure 1: Raw read accuracy of Bonito basecaller on the human reference genome NA12878 against high-accuracy Guppy, currently integrated into MinKNOW onboard nanopore devices.

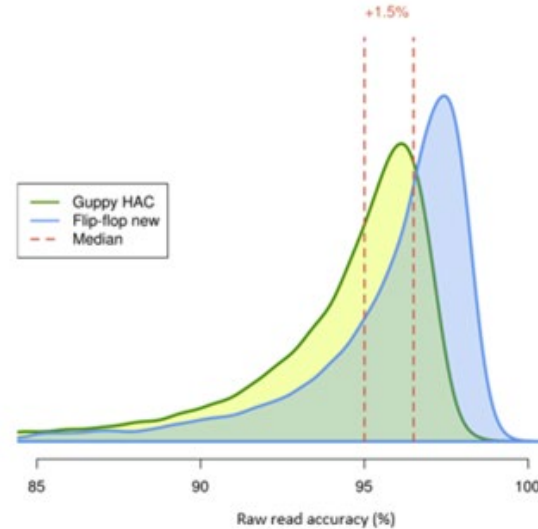


Figure 2: Raw read accuracy of new flip-flop basecaller on the human reference genome NA12878 against high-accuracy Guppy, currently integrated into MinKNOW onboard nanopore devices.

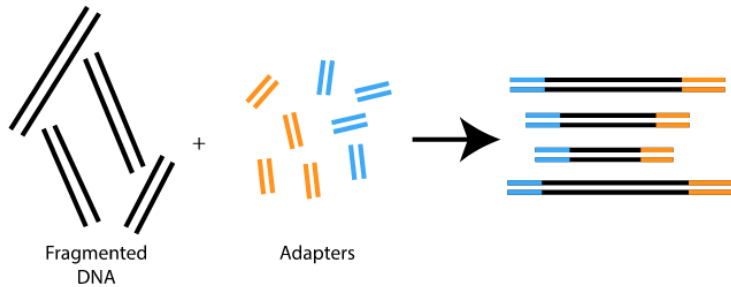
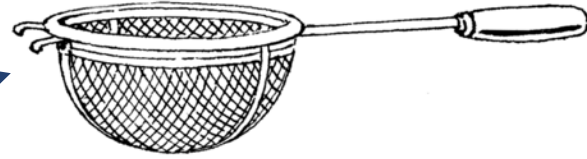
<https://nanoporetech.com/about-us/news/new-research-algorithms-yield-accuracy-gains-nanopore-sequencing>

Same Raw Fast5 file > Different Basecaller > Different Accuracy  
Keep your fast5 file

[illegible]

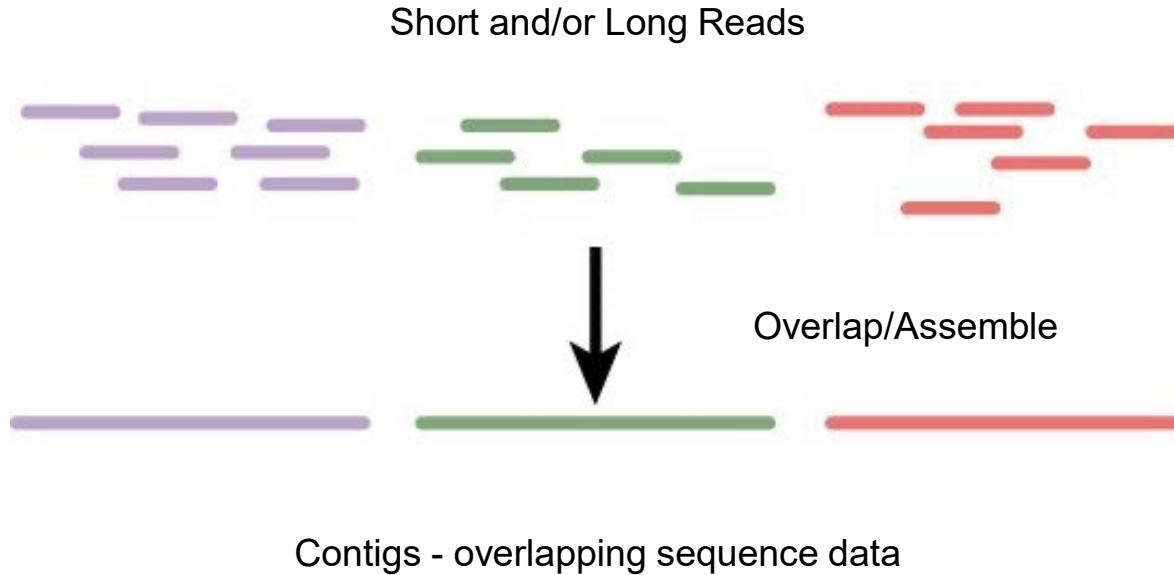
# Filtering and trimming reads

- Adapter read through
- Low-quality reads (low accuracy)
- Filter = to remove completely
- Trimming = to remove certain part of the DNA reads



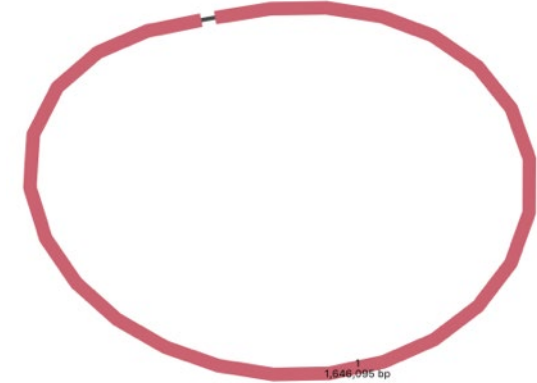
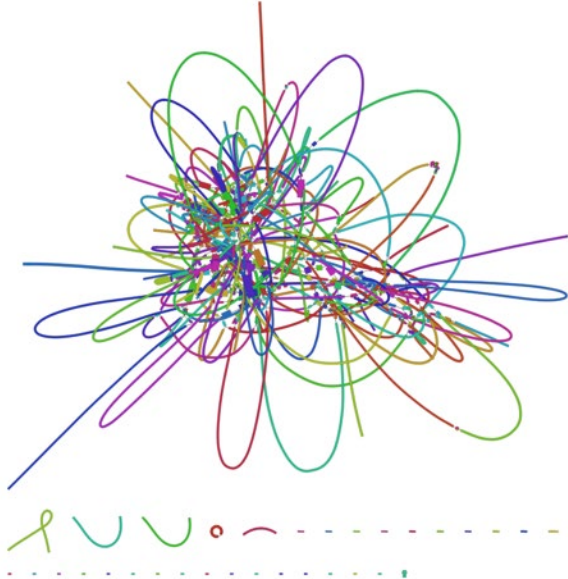
# Piecing the “clean” puzzles together

---



# Genome assembly statistics outcome depends on read length

Short Reads Assembly Graph:  
Multiple Contigs and unresolved knots



Long Reads - Resolve repeats - Simpler Graph

# Major currencies in NGS

---

## Fast5

Raw basecall signal (but specific to Nanopore)  
Large but can be re-basecalled to give improved quality

### Approximate disk storage size

800 Mb

## FastQ

Raw data (can be submitted to NCBI SRA)

100 Mb (8x smaller than fast5)

## Fasta

Assembly file (draft or complete)  
usually similar size as the genome)

2.1 Mb (A typical GBS genome)

# Key Metrics of Genome Assembly (Fasta file)

---

Number of contigs

N50

Total Assembled Length

GC content

# Key Metrics of Genome Assembly

---

- Number of contigs - Ideal Assembly = 1 contig --- 1 chromosome
- N50 = Shortest contig length at 50% of total assembled length
- Total Assembled Length = Should be similar to expected size
- GC content = No major deviation from other closely related strains

# Web-based tools

- Upload **Fasta (or FastQ)** files
- Open-source = Free to use
- Peer-reviewed
- Doesn't rely on your hardware
- Internet Connection (cloud)



# Quality Assessment of Genome Assemblies

## QUAST web server

### Quast

Quality Assessment Tool for Genome Assemblies by [CAB](#)

**QUAST evaluates genome assemblies by computing various metrics, including**

N50, length for which the collection of all contigs of that length or longer covers at least 50% of assembly length,

NG50, where length of the reference genome is being covered,

NA50 and NGA50, where aligned blocks instead of contigs are taken,

misassemblies, misassembled and unaligned contigs or contigs bases,

genes and operons covered.

**Builds convenient plots for different metrics**

cumulative contigs length,

all kinds of N-metrics,

genes and operons covered,

GC content.

[Report example](#)

More details are on [the project page](#) and in [Gurevich et al \(2013\)](#), [Bioinformatics](#).  
[Supplementary material](#) for the paper.

[Download console tool](#)

For installation details and usage instructions, please read [the manual](#).

*We will be thankful if you help us make QUAST better by sending your comments, bug reports, and suggestions to [quast.support@cab.spbu.ru](mailto:quast.support@cab.spbu.ru).*

### Quality Assessment

Assemblies

Select files

File size limit is 100Mb

or drop files here

Skip contigs shorter than  bp

☐ Scaffolds (adds assemblies splitted by fragments of N's  $\geq 10$  bp)

Email

[Get personal page](#)

We will email you a link to your page with your quality assessment reports.

We will also notify you when your report is finished, and contact you if any problems arise.

If you leave the email address blank, your reports are going to be visible for your browsing session.

Upload your fasta here!

Source: <http://cab.cc.spbu.ru/quast/>

# QUAST Output

## Quast

Quality Assessment Tool for Genome Assemblies by [CAB](#)

### NF\_3

01 August 2021, Sunday, 04:19:03

[View in Icarus contig browser](#)

[Download report](#)

Text, TSV and Latex versions  
of the table, plots in PDF.  
Additionally, detailed contigs  
and genome statistics.

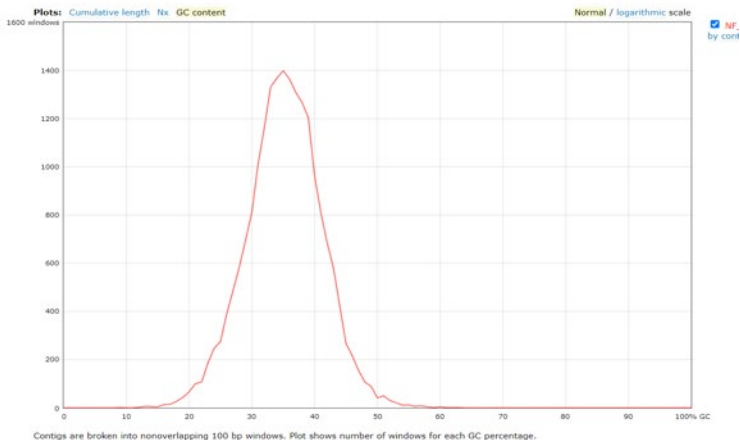
All statistics are based on contigs of size  $\geq 500$  bp, unless otherwise noted (e.g., "# contigs ( $\geq 0$  bp)" and "Total length ( $\geq 0$  bp)" include all contigs).

#### Statistics without reference ☰ NF\_3.contigs

# contigs	19
# contigs ( $\geq 0$ bp)	19
# contigs ( $\geq 1000$ bp)	17
# contigs ( $\geq 5000$ bp)	16
# contigs ( $\geq 10000$ bp)	15
# contigs ( $\geq 25000$ bp)	13
# contigs ( $\geq 50000$ bp)	10
Largest contig	491 086
Total length	1 996 902
Total length ( $\geq 0$ bp)	1 996 902
Total length ( $\geq 1000$ bp)	1 995 577
Total length ( $\geq 5000$ bp)	1 990 605
Total length ( $\geq 10000$ bp)	1 981 044
Total length ( $\geq 25000$ bp)	1 936 495
Total length ( $\geq 50000$ bp)	1 830 621
N50	261 619
N75	140 661
L50	3
L75	5
GC (%)	35.18

#### Mismatches

# N's	0
# N's per 100 kbp	0



Source: <http://cab.cc.spbu.ru/quast/>

# Assessment of Genome Completeness

## BUSCO 5

Assembled Genome -> Identification of Conserved Gene (Bacteria ODB10) -> Summarize

100% complete and single-copy gene = target

High Duplicated rate and Low Single-copy rate (S:0%,D:100%) = possible contamination

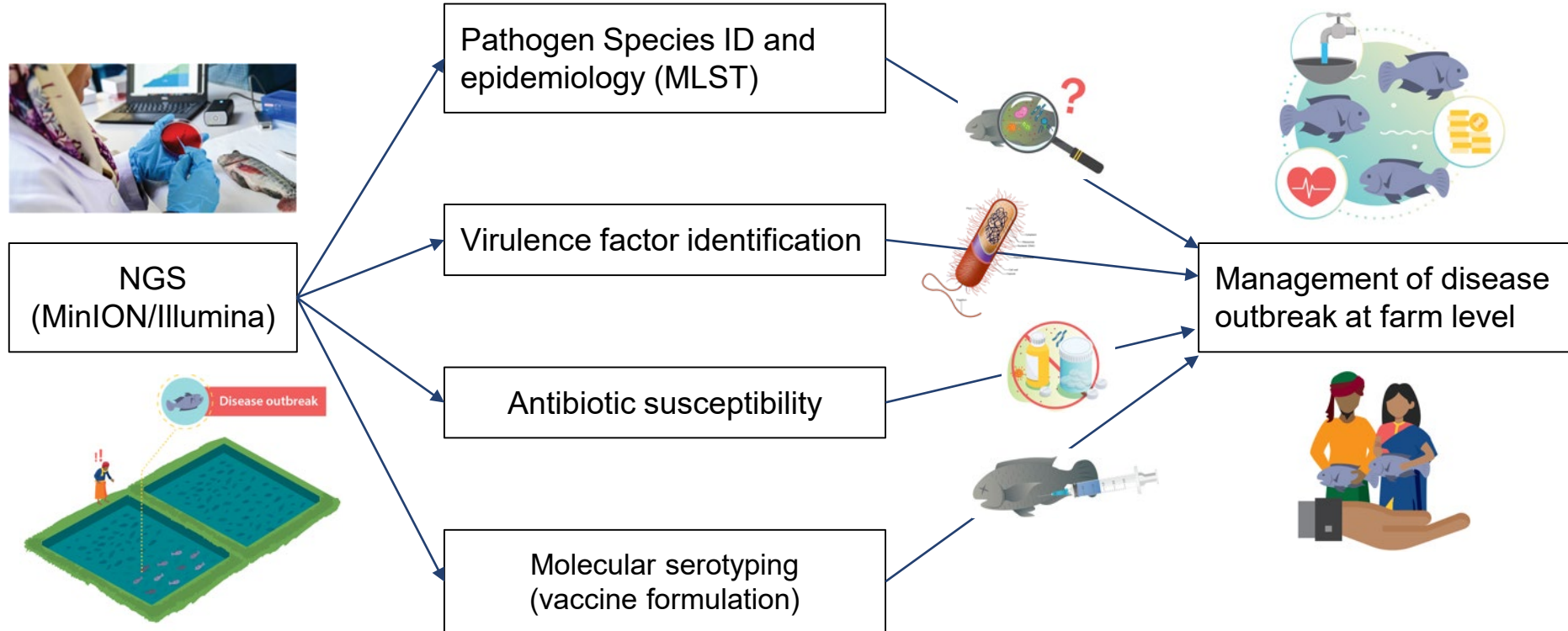
High fragmented rate indicates fragmented assembly

```
# BUSCO version is: 5.0.0
# The lineage dataset is: bacteria_odb10 (Creation date: 2020-03-06, number of species: 4085, number of BUSCOs: 124)
# Summarized benchmarking in BUSCO notation for file /media/gs/3TB/Jerome_WGS/Final/keep/Presentation_Rename/USM_3.contigs.fasta
# BUSCO was run in mode: genome
# Gene predictor used: prodigal

***** Results: *****

C:99.2%[S:99.2%,D:0.0%],F:0.8%,M:0.0%,n:124
123   Complete BUSCOs (C)
123   Complete and single-copy BUSCOs (S)
0     Complete and duplicated BUSCOs (D)
1     Fragmented BUSCOs (F)
0     Missing BUSCOs (M)
124   Total BUSCO groups searched
```

# Genome-based diagnosis of pathogens in aquaculture



# Web-based Species ID, MLST, Serotype, AMR identification

## Center for Genomic Epidemiology (CGE)

---

Center for Genomic Epidemiology

Home

Services

Publications

Contact

Source: <http://www.genomicepidemiology.org/services/>

# CGE: Taxonomic Classification

## (A sanity check)

A Fast K-mer based approach (Kmer finder)

**Center for Genomic Epidemiology**

Username:   
Password:

HomeServicesInstructionsOutputArticle abstract

### KmerFinder 3.2

Software version: [3.0.2 \(2020-10-30\)](#)  
Database: [Available here](#)  
View the [version history](#) of this server.

Select database  
Bacteria organisms

**Upload file(s)**  
To input the sequences, upload a single FASTA file, or one/two FASTQ file(s), or one interleaved FASTQ file on your local disk by using the applet below. Both assembled genome (in FASTA format) and raw reads single end or paired end (in FASTQ format) are supported. Gzipped FASTA/FASTQ files are also supported.

If you get an "Access forbidden. Error 403". Make sure the start of the web address is https and not just http. Fix it by clicking [here](#)

Isolate File

Name	Size	Progress	Status
UPM_3.contigs.fasta	1.91 MB	<div></div>	

- User-Friendly
- Select Database
- Select Your Genome fasta/fastq file (isolate)
- Upload
- Species ID

# CGE: Taxonomic Classification

## (A sanity check)

### Center for Genomic Epidemiology

[Home](#)[Services](#)[Instructions](#)[Output](#)

#### KmerFinder-3.2 Server - Results

NF\_1

KmerFinder 3.2 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth	q_value	p_value
NZ_UHIL01000001.1 <i>Vibrio parahaemolyticus</i> strain NCTC10903, whole genome shotgun sequence	200463	84406	8	110506	47.29	78.01	0.76	47.29	78.01	0.76	84379.93	1.0e-26
NZ_CP014047.2 <i>Vibrio parahaemolyticus</i> strain ATCC 17802 chromosome 2, complete sequence	409450	46084	8	64881	25.82	71.11	0.71	25.97	71.41	0.71	46059.81	1.0e-26
NC_013456.1 <i>Vibrio antiquarius</i> chromosome 1, complete sequence	85991	1503	19	110687	0.84	1.37	0.01	9.16	14.81	0.15	1446.53	1.0e-26

### Center for Genomic Epidemiology

[Home](#)[Services](#)[Instructions](#)[Output](#)

#### KmerFinder-3.2 Server - Results

USM\_3

KmerFinder 3.2 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth
NZ_JMIB01000004.1 <i>Photobacterium galathea</i> strain S2753 contig0005, whole genome shotgun sequence	315626	1860	1	12265	1.03	15.19	0.15	1.03	15.19	0.15
NZ_JMIB01000006.1 <i>Photobacterium galathea</i> strain S2753 contig0007, whole genome shotgun sequence	315628	1851	1	9514	1.02	19.35	0.19	1.02	19.35	0.19

# Average Nucleotide Identity Calculation

## JSpecies Webserver

### In-silico genome-genome hybridization

JSpeciesWS

Home

Analyse

GenomesDB

Help ▾

News

🗑️ Genome Cart

📊 ANIb Result

📊 ANIm Result

📊 Tetra Result

📊 TCS Result

🔑 875A5E76C15E77E81EAE

This is the Genome Cart of your current session. You can add genomes to this cart either by uploading own genomes or by choosing a genome from the reference database GenomesDB.

A session is automatically deleted after a period of 14 days in which it has not been visited again (except projects from registered users).

Upload own genome (min 0.02MB - max. 15MB)

Genome as (multi)-FASTA.

📎 Select file

📁 Upload ZIP archive (New!)

Choose genome from GenomesDB

Genomes included (0/15)

Pairwise comparisons ?

📊 Start ANIb

📊 Start ANIm

📊 Start Tetra

Genome cart is empty. You can add max. 15 genomes ...

Source: <http://jspecies.ribohost.com/jspeciesws/>

# Average Nucleotide Identity Calculation

## JSpecies Webserver

Compare with TYPE STRAIN (Descendants of the original isolates used in species and subspecies descriptions). Not all strains are type strains!

ANIm Result Matrix    ANIm Result by Genome

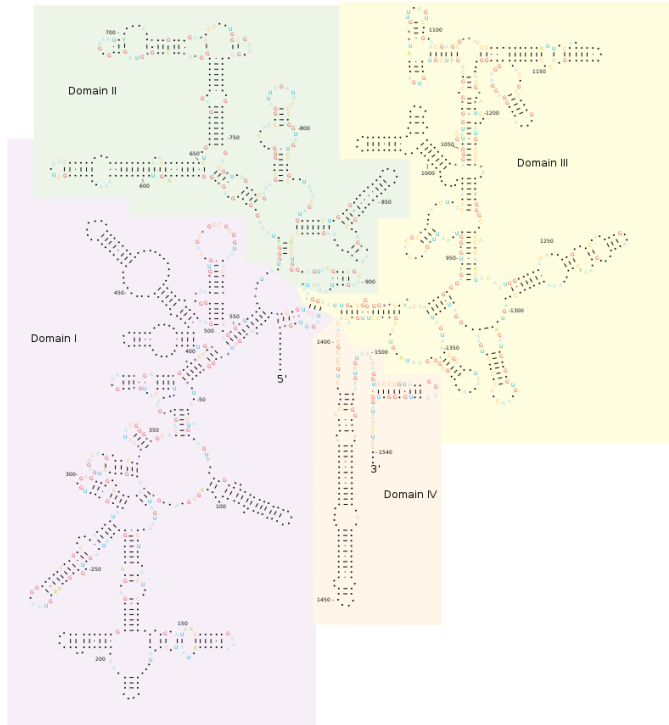
Show ANIm and [aligned nucleotides] [%] ▾    Download as .csv

Legend: Above cutoff (> 95%)    Below cutoff (< 95%)    Suspicious alignment!

	USM 3.contigs.fasta	Vibrio parahaemolyticus ATCC 17802 [T]	Vibrio alginolyticus FDAARGOS 97 [T]	Vibrio antiquarius EX25 [T]	Vibrio diabolus CNCM I-1629 [T]	Vibrio harveyi FDAARGOS 109 [T]	NF 1.contigs.fasta
USM 3.contigs.fasta	*	83.98 (1.39)	84.54 (1.37)	85.17 (0.92)	85.12 (1.22)	83.98 (1.53)	84.45 (1.41)
Vibrio parahaemolyticus ATCC 17802 [T]	84.03 (1.48)	*	86.79 (58.71)	87.57 (59.01)	87.39 (63.21)	85.30 (46.04)	98.58 (93.29)
Vibrio alginolyticus FDAARGOS 97 [T]	84.57 (2.28)	86.78 (57.65)	*	92.14 (81.04)	92.19 (86.04)	86.38 (44.02)	86.81 (59.47)
Vibrio antiquarius EX25 [T]	84.97 (1.67)	87.56 (61.52)	92.14 (85.28)	*	98.11 (92.71)	85.56 (41.26)	87.61 (62.70)
Vibrio diabolus CNCM I-1629 [T]	84.97 (1.32)	87.38 (62.35)	92.19 (85.15)	98.11 (87.07)	*	85.53 (41.87)	87.41 (62.80)
Vibrio harveyi FDAARGOS 109 [T]	84.01 (2.25)	85.31 (41.29)	86.39 (40.15)	85.57 (35.78)	85.54 (38.82)	*	85.35 (41.85)
NF 1.contigs.fasta	84.48 (1.42)	98.58 (90.68)	86.81 (57.53)	87.61 (57.81)	87.41 (61.88)	85.35 (44.39)	*

# Species Identification based on 16S rRNA similarity

## NCBI-BLAST Webserver



https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\_TYPE=BlastSearch&LINK\_LOC=blasthome

U.S. National Library of Medicine  
National Center for Biotechnology Information

**COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

BLAST® » blastn suite

Standard Nucleotide BLAST

blastn | blastp | blastx | tblastn | tblastx

BLASTN programs search nucleotide databases using a nucleotide query

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

paste 16S rRNA fasta sequence

Or, upload file

Job Title

Choose Search Set

Database

Standard databases (nr etc.): ☒ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

16S ribosomal RNA sequences (Bacteria and Archaea)

Organism

Enter organism name or id-completions will be suggested

Exclude

Limit to

Entrez Query

Program Selection

Optimize for

Highly similar sequences (megablast)

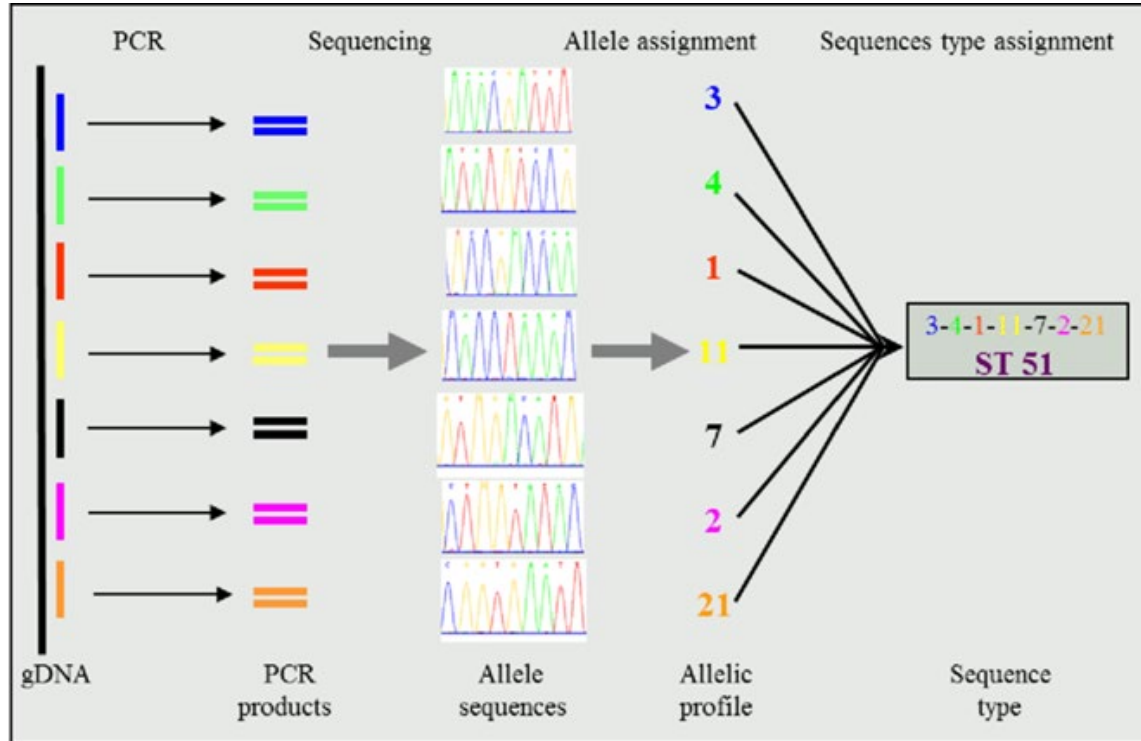
More dissimilar sequences (discontinuous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Choose the NCB rRNA/ITS databases

# Multi Locus Sequence Typing (MLST)



# Public databases for molecular typing and microbial genome diversity

**PubMLST** Public databases for molecular typing and microbial genome diversity MY ACCOUNT

[HOME](#) [ORGANISMS](#) [SPECIES ID](#) [ABOUT US](#) [UPDATES](#)

A collection of open-access, curated databases that integrate population sequence data with provenance and phenotype information for over 100 different microbial species and genera.

25,708,057  
ALLELES

869,401  
ISOLATES

620,309  
GENOMES

Organisms search

APPLY



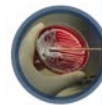
## Organisms

Choose your organism from a list of over 100 species and genera-specific databases. Access molecular typing and isolate records.



## Species ID

Use ribosomal MLST to accurately identify bacterial species from a genome assembly.



## Submit data

We welcome submissions to the databases we host. Submissions may consist of new allele sequences, MLST profiles, or isolate records. Isolates may be accompanied by a genome assembly.

Source: <https://pubmlst.org/>

# CGE: Genome-Based MLST

## MLST 2.0 (Multi-Locus Sequence Typing)

Center for Genomic Epidemiology

Username

Password

New Reset Login

Home

Services

Instructions

Output

Article abstract

### MLST 2.0 (Multi-Locus Sequence Typing)

Software version: [2.0.4 \(2019-05-08\)](#)  
Database version: [2.0.0 \(2021-07-26\)](#)

Momentanously, the species *Lactococcus Lactis* is unavailable.

Select MLST configuration  

Streptococcus agalactiae

MLST allele sequence and profile data are obtained from [PubMLST.org](#).

Please note that for four organisms, two or three different MLST schemes are available:

- *Acinetobacter baumannii* (*Acinetobacter baumannii* #1 [\[1\]](#), *Acinetobacter baumannii* #2 [\[link\]](#)).
- *Escherichia coli* (*Escherichia coli* #1 [\[4\]](#), *Escherichia coli* #2 [\[5\]](#)).
- *Pasteurella multocida* (*Pasteurella multocida* #1 (RIRD-C), *Pasteurella multocida* #2 (multihost)).
- *Leptospira* (*Leptospira* #1, *Leptospira* #2, *Leptospira* #3).

Note: *Campylobacter coli* and *Campylobacter jejuni* are considered together.

Select type of data input  
Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.  

Assembled or Draft Genome/Contigs\* (fasta)

Please note that "Assembled Genomes/Contigs" should be selected, if you have already assembled your short sequencing reads into one continuous genome or into several contigs. It is indifferent which type of short sequence reads were used to produce the genome/contigs.

Isolate File

Name	Size	Progress	Status

- Alternative to conventional MLST based on PCR / Sanger
- Fast and Efficient
- Digitalize
- Less Error-prone

Source: <https://cge.cbs.dtu.dk/services/MLST/>

# CGE: MLST from assembled contigs or raw data

## MLST 2.0 (Multi-Locus Sequence Typing)

### Center for Genomic Epidemiology

[Home](#)[Services](#)[Instructions](#)[Output](#)

#### MLST-2.0 Server - Results

mlst Profile: *sagalactiae*

Organism: *Streptococcus agalactiae*

Sequence Type: 283!

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
adhP	100	100	498	498	0	adhP_9
atr	100	100	501	501	0	atr_7
glcK	100	100	459	459	0	glcK_3
glnA	100	100	498	498	0	glnA_1
pheS	100	100	501	501	0	pheS_5
sdhA	100	100	519	519	0	sdhA_3
tkr	100	100	480	480	0	tkr_134!
tkr	100	100	480	480	0	tkr_2!

# CGE: AMR identification from assembled contigs or raw data

## ResFinder 4.1



**Center for Genomic Epidemiology**

Username  Password   
[New](#) [Reset](#) [Login](#)

[Home](#) [Services](#) [Instructions](#) [Output](#) [Overview of genes](#) [Article abstract](#)

### ResFinder 4.1

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.

**Updates**

ResFinder and PointFinder software: [\(2021-05-27\)](#)  
ResFinder database: [\(2021-04-28\)](#)  
PointFinder database: [\(2021-02-01\)](#)

The database is curated by:  
**Frank Møller Aarestrup**  
([click to contact](#))

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☒

Select Antimicrobial configuration  
Select multiple items, with Ctrl-Click (or Cmd-Click on Mac) - as default all databases is selected

☐ Aminoglycoside  
☐ Beta-lactam  
☐ Colistin  
☐ Disinfectant  
☐ Fluoroquinolone  
☐ Fosfomycin

Select threshold for %ID

Select minimum length

Select species  
  
Chromosomal point mutation database exists

Select type of your reads

If you get an "Access forbidden. Error 403". Make sure the start of the web address is https and not just http. Fix it by clicking [here](#).

Name	Size	Progress	Status
------	------	----------	--------

- Homology-based AMR identification
- % ID vs % Length (Specific vs Sensitive)
- Select genome and Upload

# CGE: AMR identification from assembled contigs or raw data

## ResFinder 4.1



Macrolide									
Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
mre(A)	100.0	936/936	1..936	NODE_4_length_152740_cov_230.023915	31574..32509	erythromycin,azithromycin,spiramycin	9420045	<a href="#">U92073</a>	

Tetracycline									
Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.	Notes
tet(M)	100.0	1920/1920	1..1920	NODE_3_length_261619_cov_280.568306	233331..235250	doxycycline,tetracycline,minocycline	20546576	<a href="#">AM990992</a>	

Source: <https://cge.cbs.dtu.dk/services/ResFinder/>

# Genome Submission to NCBI



## COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)



## Login Update Warning

Important changes to how you log into NCBI coming in June. [Read more here.](#)



## Submission Portal

[Home](#)

[My submissions](#)

[Manage data](#)

[Templates](#)

[My profile](#)

## Genome

[New submission](#)



**Note:** To find submissions started before Feb. 3, 2014, go to the [previous version](#) of the WGS submission wizard.



**ATTN:** to fix or update a recent submission whose status is Queued, Processed-error or Processing, please use

- the FIX button on the existing submission
- or [email your request](#) to have the FIX button enabled for that submission.

Be sure to include the Submission ID and the reason that you need to send new files.

**Do not** create a new submission to fix or update an existing submission whose status is Queued, Processed-error or Processing!

Short description and brief instructions



Options to preload data:

<https://submit.ncbi.nlm.nih.gov/subs/genome/>

# A Typical Annotated Genome by NCBI

- .gbk format with GenBank Accession code for the assembly
- Draft or complete assembly will be accepted
- Unique Accession for each Gene
- Protein Translation
- Start and Stop
- Putative Function description

```
gene      5484..5560
          /locus_tag="HU005_00030"
tRNA      5484..5560
          /locus_tag="HU005_00030"
          /product="tRNA-Asp"
          /inference="COORDINATES: profile:tRNAscan-SE:2.0.4"
          /note="Derived by automated computational analysis using
          gene prediction method: tRNAscan-SE."
          /anticodon=(pos:5518..5520,aa:Asp,seq:gtc)

gene      5596..5672
          /locus_tag="HU005_00035"
tRNA      5596..5672
          /locus_tag="HU005_00035"
          /product="tRNA-Trp"
          /inference="COORDINATES: profile:tRNAscan-SE:2.0.4"
          /note="Derived by automated computational analysis using
          gene prediction method: tRNAscan-SE."
          /anticodon=(pos:5630..5632,aa:Trp,seq:cca)

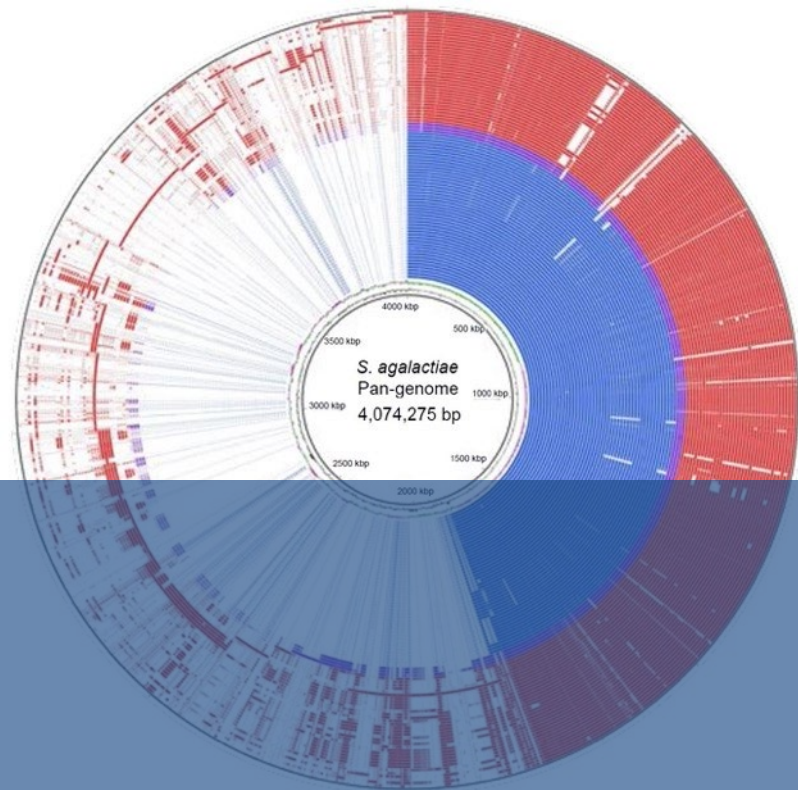
gene      6147..6479
          /locus_tag="HU005_00040"
CDS       6147..6479
          /locus_tag="HU005_00040"
          /inference="COORDINATES: similar to AA
          sequence:RefSeq:WP_017633933.1"
          /note="Derived by automated computational analysis using
          gene prediction method: Protein Homology."
          /codon_start=1
          /transl_table=11
          /product="DUF1904 domain-containing protein"
          /protein_id="OKS93699.1"
          /translation="MPHFRFRAVEPQAVQALSKPLTDELQPLMCPREDFTFEYIYTT
          FFNEGEVSAAYPFVEVLWFDGRGEVQDEVAKLITQVIRGIAGADIDVAVIFSALSPKA
          YDNGEHY"

gene      6492..7328
          /gene="mepA"
          /locus_tag="HU005_00045"
CDS       6492..7328
          /gene="mepA"
          /locus_tag="HU005_00045"
          /inference="COORDINATES: similar to AA
          sequence:RefSeq:WP_017633932.1"
          /note="Derived by automated computational analysis using
          gene prediction method: Protein Homology."
          /codon_start=1
          /transl_table=11
          /product="penicillin-insensitive murein endopeptidase"
          /protein_id="OKS93700.1"
          /translation="MRLSFIVTLGLCFASVCSTPMESAVNPTRNSSSSISGYANGCL
          DGLALPLPLDGVGYVLRKSKTKRYVGHGKTIETIENLAKAHQHLNTLLIGQLSLPRG
          GRFSSGHSWOTGLDIDWLRLADQPLSYNELQPKPMVSVDLKGYSILNHRWEERHF
          KLIYASKSKDVARIFVHPVIKEQLQLQENGKDRSLRKVRPMGHYHVFHRLSCPQ
          SSDYCVDSPPVPGGCGAELASWAPKAKPIDIPRVKTTSPKRRKVVPPQCLPLIN
          PN"
```

```

081 aatatcaaca ctatcaaatt ttgataa
141 cattactggt actactatth tcattat
201 ctttggttgg ccttatcatt tacagtga
261 ttatgtgaag ctttaccaaa ccccttt
321 ttagctgctg tagttgatgc aatgata
381 tcaattaatt tatagtaatc tttagta
441 cttggtggat caataacaat cacatca
501 tattcaaaaa catccataac gataaat

```



## Participant Data Overview

Dr Gan Han Ming

GeneSEQ



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB

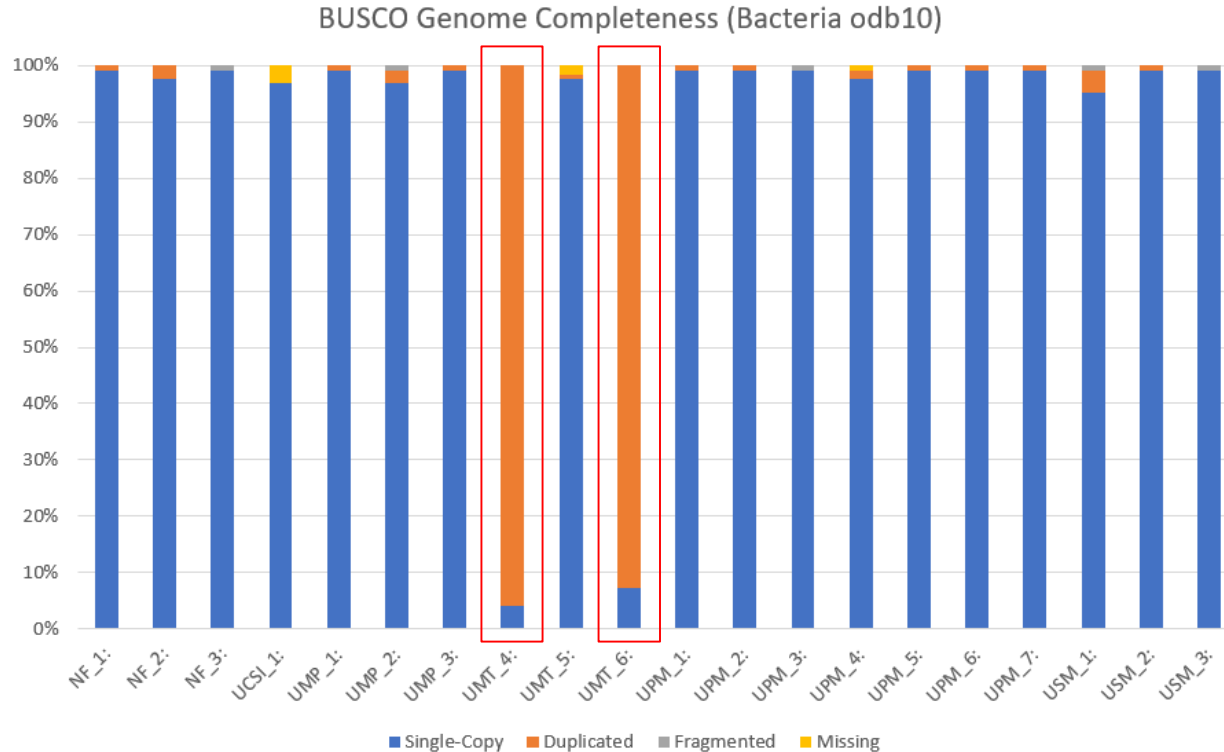


# Participants' Analyzed Data

## Putative Species Assignment Based on WGS

ID	family	genus	species
UMP_1.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
UMT_6.contigs.fasta	f__Bacillaceae_G	g__Bacillus_A	
USM_2.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio diabolus
UMP_2.contigs.fasta	f__Bacillaceae_H	g__Priestia	
UPM_7.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
UMP_3.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
UMT_4.contigs.fasta	f__DSM-18226	g__Cytobacillus	
UMT_5.contigs.fasta	f__Enterobacteriaceae	g__Enterobacter	
USM_3.contigs.fasta	f__Vibrionaceae	g__Photobacterium	s__Photobacterium galathea
UCSI_1.contigs.fasta	f__Enterobacteriaceae	g__Escherichia_C	s__Escherichia_C alba
USM_1.contigs.fasta	f__Bacillaceae_G	g__Bacillus_A	
NF_3.contigs.fasta	f__Streptococcaceae	g__Streptococcus	s__Streptococcus agalactiae
NF_2.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio harveyi
UPM_3.contigs.fasta	f__Streptococcaceae	g__Streptococcus	s__Streptococcus agalactiae
UPM_6.contigs.fasta	f__DSM-18226	g__Cytobacillus	s__Cytobacillus oceanisediminis_B
UPM_4.contigs.fasta	f__Aeromonadaceae	g__Aeromonas	
UPM_2.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
NF_1.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
UPM_5.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus
UPM_1.contigs.fasta	f__Vibrionaceae	g__Vibrio	s__Vibrio parahaemolyticus

# Assessment of Genome Completeness



# MLST on selected strains (*Vibrio parahaemolyticus*)

Sample ID	Species	Sequence Type (ST)	MLST Allele						
NF_1.contigs.fasta	vparahaemolyticus	799	dnaE(28)	gyrB(4)	recA(82)	dtdS(88)	pntA(63)	pyrC(187)	tnaA(1)
UMP_1.contigs.fasta	vparahaemolyticus	-	dnaE(71)	gyrB(13)	recA(97)	dtdS(~461)	pntA(~116)	pyrC(82)	tnaA(26)
UMP_3.contigs.fasta	vparahaemolyticus	-	dnaE(71)	gyrB(13)	recA(97)	dtdS(~461)	pntA(~116)	pyrC(82)	tnaA(26)
UPM_1.contigs.fasta	vparahaemolyticus	1913	dnaE(363)	gyrB(505)	recA(218)	dtdS(442)	pntA(30)	pyrC(303)	tnaA(26)
UPM_2.contigs.fasta	vparahaemolyticus	-	dnaE(~98)	gyrB(264)	recA(~105)	dtdS(~13)	pntA(2)	pyrC(~403)	tnaA(94)
UPM_7.contigs.fasta	vparahaemolyticus	392	dnaE(5)	gyrB(84)	recA(115)	dtdS(74)	pntA(63)	pyrC(159)	tnaA(84)

[https://pubmlst.org/bigsdb?page=profileInfo&db=pubmlst\\_vparahaemolyticus\\_seqdef&scheme\\_id=1&profile\\_id=1913](https://pubmlst.org/bigsdb?page=profileInfo&db=pubmlst_vparahaemolyticus_seqdef&scheme_id=1&profile_id=1913)

## Profile information for ST-1913 (MLST)

ST	dnaE	gyrB	recA	dtdS	pntA	pyrC	tnaA	clonal complex
1913	363	505	218	442	30	303	26	no

sender: Chrystine Yan, School of Science, Monash University Malaysia and 2. Genomics Facility, Tropical Medicine and Biology Platform, Monash University Malaysia

curator: Narjol Gonzalez-Escalona, Molecular Methods & Subtyping Branch, Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, FDA, MD, USA

update history: [1 update](#) [show details](#)

date entered: 2018-05-30

datestamp: 2018-05-30

# ST-1913 Background



Volume 366, Issue 17  
September 2019

< Previous Next >

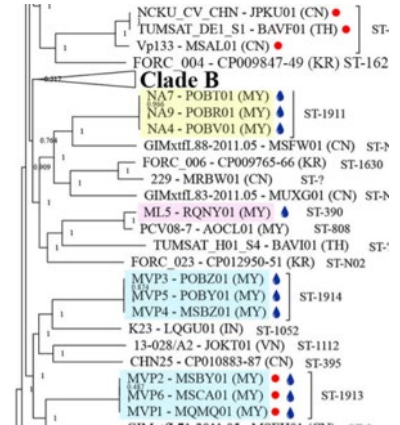
Genomic characterization of *Vibrio parahaemolyticus* from Pacific white shrimp and rearing water in Malaysia reveals novel sequence types and structural variation in genomic regions containing the *Photothabdus* insect-related (Pir) toxin-like genes

Christine Zou Yi Yan, Christopher M Austin, Qasim Ayub, Sadequr Rahman, Han Ming Gan ✉

FEMS Microbiology Letters, Volume 366, Issue 17, September 2019, fnz211,  
<https://doi.org/10.1093/femsle/fnz211>

Published: 07 October 2019 Article history ▼

- *pirA*<sup>+</sup>*B*<sup>+</sup>
- *pirA*<sup>+</sup>*B*<sup>-</sup>
- *pirA*<sup>-</sup>*B*<sup>-</sup> (pVa 'backbone')
- 💧 Rearing Water
- 🦐 Shrimp Individual 1
- 🦐 Shrimp Individual 2
- 🏠 Farm 1 (Northern Malaysia)
- 🏠 Farm 2 (Central Malaysia)
- 🏠 Farm 3 (Central Malaysia)



## Three major structural variants of the pVA plasmid among Asian *V. parahaemolyticus*

Of the 52 sequence types of *V. parahaemolyticus* identified in this current genomic sampling, *pirAB*<sup>Vp</sup> could only be detected among isolates from 8 validated (ST-2013, ST-114, ST-150, ST-392, ST-413, ST-970, ST-1166, and ST-1913) and 1 undetermined sequence types (Figure 1).

# Linux command line - An example with AMR identification



```
--threads [N] Use this many BLAST+ threads [1].
DATABASES
--setupdb      Format all the BLAST databases.
--list         List included databases.
--datadir [X]  Databases folder [/home/gan/miniconda3/envs/abricate/db].
--db [X]       Database to use [ncbi].
OUTPUT
--noheader     Suppress column header row.
--csv          Output CSV instead of TSV.
--nopath       Strip filename paths from FILE column.
FILTERING
--minid [n.n]  Minimum DNA %identity [75].
--mincov [n.n] Minimum DNA %coverage [0].
MODE
--summary      Summarize multiple reports into a table.
DOCUMENTATION
https://github.com/tseemann/abricate
(abricate) gan@gan: /mnt/c/Ubuntu_Shared/Jerome_WorkShop/keep/Presentation_Rename/Vpare$ abricate *.contigs.fasta
Using ncbi database ncbi: 5283 sequences - 2020-Feb-20
#FILE SEQUENCE START END STRAND GENE COVERAGE COVERAGE_MAP GAPS %COVERAGE %IDENTITY
DATABASE ACCESSION PRODUCT RESISTANCE
Processing: NF_1.contigs.fasta
Found 3 genes in NF_1.contigs.fasta
NF_1.contigs.fasta NODE_11_length_192793_cov_91.644690 64786 65250 - tet(34) 1-465/465 =====/=====
/2 99.78 83.48 ncbi NG_048129.1 oxytetracycline resistance phosphoribosyltransferase domain-containing protein Tet(34)
ETRACYCLINE
NF_1.contigs.fasta NODE_1_length_601117_cov_80.593360 329283 330134 - blaCARB-47 1-852/852 =====
/0 100.00 99.18 ncbi NG_050564.1 carbenicillin-hydrolyzing class A beta-lactamase CARB-47 BETA-LACTAM
NF_1.contigs.fasta NODE_4_length_470469_cov_78.618048 189617 191218 + tet(35) 1-1602/1602 =====
/0 100.00 99.19 ncbi NG_063830.1 tetracycline efflux Na+/H+ antiporter family transporter Tet(35) TETRACYCLINE
Processing: UPM_1.contigs.fasta
Found 3 genes in UPM_1.contigs.fasta
UPM_1.contigs.fasta NODE_1_length_862548_cov_114.102298 519864 520715 + blaCARB-33 1-852/852 =====
/0 100.00 99.53 ncbi NG_048737.1 carbenicillin-hydrolyzing class A beta-lactamase CARB-33 BETA-LACTAM
UPM_1.contigs.fasta NODE_4_length_471855_cov_114.677020 280941 282542 - tet(35) 1-1602/1602 =====
/0 100.00 98.94 ncbi NG_063830.1 tetracycline efflux Na+/H+ antiporter family transporter Tet(35) TETRACYCLINE
UPM_1.contigs.fasta NODE_9_length_179391_cov_118.384114 64784 65248 - tet(34) 1-465/465 =====/=====
/2 99.78 83.26 ncbi NG_048129.1 oxytetracycline resistance phosphoribosyltransferase domain-containing protein Tet(34)
ETRACYCLINE
Processing: UPM_3.contigs.fasta
Found 3 genes in UPM_3.contigs.fasta
UPM_3.contigs.fasta NODE_1_length_876017_cov_54.348261 685103 686704 - tet(35) 1-1602/1602 =====
/0 100.00 98.94 ncbi NG_063830.1 tetracycline efflux Na+/H+ antiporter family transporter Tet(35) TETRACYCLINE
UPM_3.contigs.fasta NODE_2_length_862548_cov_54.500130 341834 342685 - blaCARB-33 1-852/852 =====
/0 100.00 99.53 ncbi NG_048737.1 carbenicillin-hydrolyzing class A beta-lactamase CARB-33 BETA-LACTAM
UPM_3.contigs.fasta NODE_9_length_179145_cov_60.183578 64784 65248 - tet(34) 1-465/465 =====/=====
/2 99.78 83.26 ncbi NG_048129.1 oxytetracycline resistance phosphoribosyltransferase domain-containing protein Tet(34)
ETRACYCLINE
Processing: UPM_1.contigs.fasta
```

# Visualization of AMR profile among sequenced strains

#FILE	BclI	aph(3'')-Ib	aph(6)-Id	bla1	blaACT-70	blaCARB-26	blaCARB-33	blaCARB-38	blaCARB-42	blaCARB-45	blaCARB-47	blaLHK-6	blaOXA-724	blaP	blaTEM-1	blaVHH-1	catA10	catA2	cepH	dfrA14	dfrA17	fexA	fosB-38141535	fosB_gen	fosD	imiH	lsa(B)	mcr-7.1	mecA1	oqxA9	oqx89	qnrS2	rphC	sal(A)	satA_Ba	sul1	sul2	tet(34)	tet(35)	tet(45)	tet(A)	tet(K)	tet(M)
NF_1	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
NF_2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.77	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100	.	.	.	.	
NF_3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	
UCSI_1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
UMP_1	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.	
UMP_2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.88	.	.	.	.	99.78	100	.	.	.	.
UMP_3	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
UMT_4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	100	
UMT_5	.	100	100	.	100	.	.	.	.	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	100	.	100	98.95	.	.	.	.	100	.	.	.	.	.	
UMT_6	##	.	.	99	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	99.76	.	.	.	.	.	.	.	.	99.35	.	.	.	.	.	.	.	.	.	.
UPM_1	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
UPM_2	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	99.78	100	.	.	.	.	
UPM_3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
UPM_4	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.	100	.	.	.	.	.	100	.	86.48	.	.	.	100	.	.	.	100	.	.	.	100	.	.	
UPM_5	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
UPM_6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
UPM_7	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99.78	100	.	.	.	.
USM_1	##	.	100	.	.	.	.	.	.	.	.	.	80.84	.	.	.	.	.	.	.	.	.	31.18	100	.	.	98.99	.	.	.	.	99.5	.	98.38	.	.	.	100	99.38	.	.	.	.
USM_2	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	99.38	.	.	.	.
USM_3	.	.	.	.	.	.	.	.	.	.	11.08	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

Source: <https://github.com/tseemann/abricate>

# MLST on selected strains (GBS)

(1) Using **Linux** based approach (<https://github.com/tseemann/mlst>)

```
NF_3.contigs.fasta    sagalactiae    283    adhP(9) pheS(5) atr(7)  glnA(1) sdhA(3) glcK(3) tkt(2)
UPM_3.contigs.fasta  sagalactiae    283    adhP(9) pheS(5) atr(7)  glnA(1) sdhA(3) glcK(3) tkt(2)
```

SAME RESULTS

(2) Rapid and precise alignment  
draft genome/contigs (fasta) against  
redundant databases with KMA

## Center for Genomic Epidemiology

[Home](#)[Services](#)[Instructions](#)[Output](#)

### MLST-2.0 Server - Results

mlst Profile: *sagalactiae*

Organism: *Streptococcus agalactiae*

Sequence Type: **283!**

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
adhP	100	100	498	498	0	adhP_9
atr	100	100	501	501	0	atr_7
glcK	100	100	459	459	0	glcK_3
glnA	100	100	498	498	0	glnA_1
pheS	100	100	501	501	0	pheS_5
sdhA	100	100	519	519	0	sdhA_3
tkt	100	100	480	480	0	tkt_134l
tkt	100	100	480	480	0	tkt_2l

# CGE: ST-283 pathogenicity towards human hosts

## PathogenFinder 1.1

Center for Genomic Epidemiology

Home

Services

Instructions

Output

### PathogenFinder 1.1

View the [version history](#) of this server.

Choose the phylum or class of your organism:  
Choose 'All' if you want to use the model created using all bacteria.  
Automatic Model Selection

Sequencing Platform  
Select the sequencing platform used to generate the uploaded reads. (Note: Select 'Assembled Genome' if you are uploading preassembled reads)  
Proteome

Isolate File

Name	Size	Progress	Status
------	------	----------	--------

Upload Remove

**IMPORTANT NOTE:**  
To avoid problems caused by file names, we only allow a limited selection of ASCII characters (see below).  
A-Z  
a-z  
0-9  
\_ (underscore)  
- (hyphen)  
. (full stop)

**REFERENCES**

1. Camacho C, Coutinho G, Alvayn V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+ architecture and applications. *BMC Bioinformatics* 2009; 10:421.

**CITATIONS**

For publication of results, please cite:

- PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. Covert SD, Vothly Lensen M, Keller Aepresen F, Lind O (2013) *PLoS ONE* 8(10): e77352. PMID: 24204725 DOI: 10.1371/journal.pone.0077352

<https://cge.cbs.dtu.dk/services/PathogenFinder/>

Center for Genomic Epidemiology

Home

Services

Instructions

Output

### The input organism was predicted as human pathogen

Probability of being a human pathogen 0.887  
Input proteome coverage (%) 14.31  
Matched Pathogenic Families 282  
Matched Not Pathogenic Families 0

Sequences 1970  
Total bpp 596357  
Longest seq 1572  
Shortest seq 30  
Avg seq length 302.0

**Input Sequence**  
NODE\_4\_length\_152731\_cov\_249.280477\_37 # 36916 # 39543 # -1 # D=4\_37.partial=00.start\_type=ATG.rbs\_motif=GGAG/GAGG.rbs\_spacer=5-10bp; gc\_cont=0.326

**Matched Family**

PROTEIN ID	ACCESSION ID	ORGANISM	CLASS	PROTEIN FUNCTION	PROTEIN ID	SCORE (%)
326	CP000114	Streptococcus agalactiae A909, complete genome	Lactobacillales	ABC transporter, permease protein, putative	ABA45245	100.0

**Input Sequence**  
NODE\_5\_length\_91300\_cov\_280.705359\_76 # 73026 # 75344 # -1 # D=5\_76.partial=00.start\_type=ATG.rbs\_motif=AGGAG.rbs\_spacer=5-10bp; gc\_cont=0.359

**Matched Family**

PROTEIN ID	ACCESSION ID	ORGANISM	CLASS	PROTEIN FUNCTION	PROTEIN ID	SCORE (%)
326	CP000114	Streptococcus agalactiae A909, complete genome	Lactobacillales	prophage Sa05, membrane protein, putative	ABA45245	100.0

**Input Sequence**  
NODE\_2\_length\_472300\_cov\_263.810268\_103 # 118054 # 120315 # -1 # D=2\_103.partial=00.start\_type=ATG.rbs\_motif=GGAG/GAGG; rbs\_spacer=5-10bp; gc\_cont=0.313

**Matched Family**

PROTEIN ID	ACCESSION ID	ORGANISM	CLASS	PROTEIN FUNCTION	PROTEIN ID	SCORE (%)
326	CP000114	Streptococcus agalactiae A909, complete genome	Lactobacillales	fibrinogen-binding protein	ABA45245	100.0

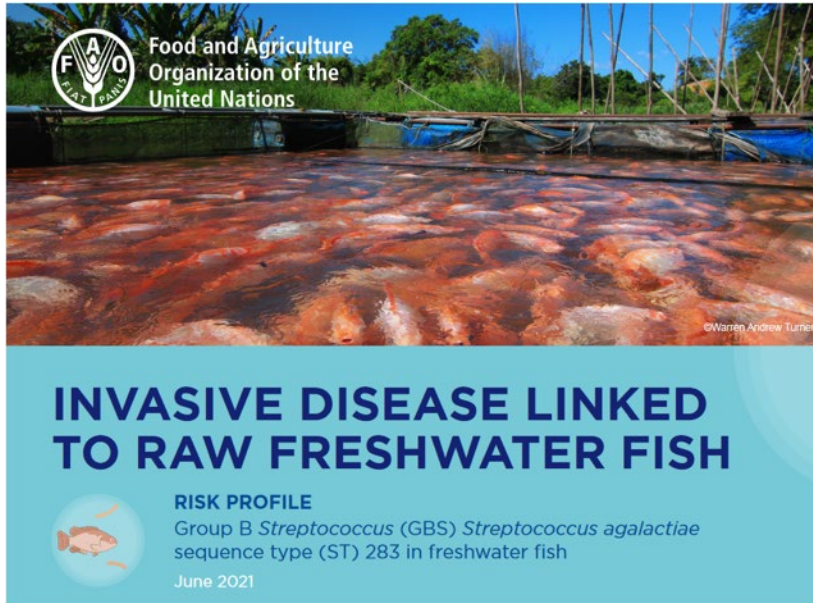
**Input Sequence**  
NODE\_2\_length\_472300\_cov\_263.810268\_261 # 291881 # 294076 # -1 # D=2\_261.partial=00.start\_type=ATG.rbs\_motif=GGAG/GAGG; rbs\_spacer=5-10bp; gc\_cont=0.362

**Matched Family**

PROTEIN ID	ACCESSION ID	ORGANISM	CLASS	PROTEIN FUNCTION	PROTEIN ID	SCORE (%)
326	CP000114	Streptococcus agalactiae A909, complete genome	Lactobacillales	cyt protein	ABA45245	100.0

<https://cge.cbs.dtu.dk/cgi-bin/webface.fcgi?jobid=6107946F000009F219F0D4A7>

# MLST: *Streptococcus agalactiae* sequence type (ST) 283



<http://www.fao.org/3/cb4901en/cb4901en.pdf>

Clinical Infectious Diseases

SUPPLEMENT ARTICLE



## 2015 Epidemic of Severe *Streptococcus agalactiae* Sequence Type 283 Infections in Singapore Associated With the Consumption of Raw Freshwater Fish: A Detailed Analysis of Clinical, Epidemiological, and Bacterial Sequencing Data

Shirin Kalimuddin,<sup>1,2</sup> Swaine L. Chen,<sup>2,3,4</sup> Cindy T. K. Lim,<sup>4,5</sup> Tse Hsien Koh,<sup>5</sup> Thean Yen Tan,<sup>6</sup> Michelle Kam,<sup>7</sup> Christopher W. Wong,<sup>2</sup> Kurosh S. Meherashahi,<sup>7</sup> Man Ling Chau,<sup>8</sup> Lee Ching Ng,<sup>4</sup> Wen Ying Tang,<sup>9</sup> Hishamuddin Badaruddin,<sup>10</sup> Jeanette Teo,<sup>11</sup> Anucha Apisarnthanarak,<sup>12</sup> Nuntra Suwantararat,<sup>12,13</sup> Margaret Ip,<sup>14</sup> Matthew T. G. Holden,<sup>15</sup> Li Yang Hsu,<sup>16</sup> and Timothy Barkham<sup>1</sup>

## PLOS NEGLECTED TROPICAL DISEASES

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## One hypervirulent clone, sequence type 283, accounts for a large proportion of invasive *Streptococcus agalactiae* isolated from humans and diseased tilapia in Southeast Asia

Timothy Barkham , Ruth N. Zadoks, Mohammad Noor Amal Azmai, Stephen Baker, Vu Thi Ngoc Bich, Victoria Chalker, Man Ling Chau, David Dance, Rama Narayana Deepak, H. Rogier van Doorn, Ramona A. Gutierrez, Mark A. Holmes, Lan Nguyen Phu Huong, [ ... ], Swaine L. Chen  [ view all ]

Published: June 27, 2019 • <https://doi.org/10.1371/journal.pntd.0007421>

# Molecular serotyping on selected strains (GBS)

BLAST-based approach  
(input = Fasta: assembled draft or complete genome)

<https://github.com/swainechen/GBS-SBG>

#	Name	Serotype
NF_3		GBS-SBG:III-4
#	Name	Serotype
UPM_3		GBS-SBG:III-4

SAME RESULTS

## AquaPath.

Rapid and accurate diagnosis for the control and prevention of diseases in aquatic animals

[Learn more >](#)

[ID a pathogen >](#)

Diagnostic k-mer-based approach  
(input = Fastq > 400 read sequences - Assembly-Free and Real-Time)

[In development by WorldFish, UQ and Wilderlab](#)

[Browse...](#) 1000.fastq

Upload a FASTQ file

Identify pathogen

```
{"data_filename":["1000.fastq"],"data_content_type":["application/octet-stream"],"matches":  
{"sa_sero":{"III":[5880],"Ib":[1071],"V":[781],"II":[572],"VI":[506],"Ia":[496],"IV":[95]}}
```

# Thank You



This work was undertaken as part of



RESEARCH  
PROGRAM ON  
Fish  
Led by WorldFish



Platform for  
Big Data  
in Agriculture



In partnership with



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA



WILDERLAB



GeneSEQ  
NEXT GENERATION SEQUENCING