university of groningen

umcg

| 1

Use of Next Generation Sequencing for clinical testing and epidemiology in medical microbiology - Troubleshooting NGS workflow

John WA Rossen

Senior Researcher Genomics for Infection Prevention
Head of the Molecular and Sequence Unit Clinical Bacteriology
Medical Molecular Microbiologist
Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Disclosure
Ion Torrent PGM was borrowed from LifeTechnologies through a collaboration with Pathogenica till 31 January 2014

No personal benefits

university of groningen

umcg

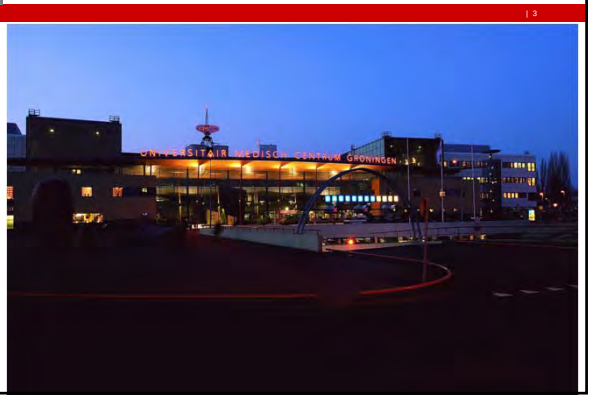
| 2

Groningen, The Netherlands

university of groningen

umcg

| 3



university of groningen umcg

| 4

university of groningen umcg

| 5

DNA isolation

› Manual which kit?
- Depending on application
- Deplete human DNA ?
- Enrich bacterial DNA ?

› What if you have 2000 strains?

› Automation which robot?
- Depending on the person you speak with
- Where to place a robot

university of groningen umcg

| 6

Library preparation

› Measuring DNA/RNA concentration/quality
- Qubit, nanodrop, bioanalyser or equivalent, agarose gel

› Method of Library prep
- Mechanical or enzymatic
- Costs vs hands-on time vs reproducibility
- Automation

› Size selection
- Beads (ratio beads vs DNA)
- Bioanalyser or equivalent
- Gel?

university of groningen

umcg

| 7

Sequencing

› Loading the flow cell or chip

› Coverage vs number of samples

› Speed vs read length vs output

university of groningen

| 8

Ion PGM™ System Performance Specifications

		Ion 314™ Chip v2	Ion 316™ Chip v2	Ion 318™ Chip v2
Output*	200 base 400 base	30–50 Mb 60–100 Mb	300–400 Mb 600 Mb–1 Gb	2 Gb 1.2 Gb
Reads		400–550 thousand	2–3 million	5.5 million
Run time	200 base 400 base	2.3 hr 3.7 hr	3.0 hr 4.7 hr	4.4 hr 7.3 hr
Research areas		Cancer research Genetic disease research Microbiology	Stem cell research Agriculture Epigenomics	Metagenomics Forensic science Ancient DNA genomics
Key applications		Targeted DNA sequencing Copy number analysis Targeted RNA sequencing Small RNA sequencing	*de novo* microbial sequencing Bacterial typing Viral typing Metagenomics	CNV sequencing Methylation analysis SNP validation Sequence-based genotyping
Target selection solutions		AmpliSeq™ Technology	Ion TargetSeq™ Technology	
Library solutions		Ion AmpliSeq™ Library Kit Ion Xpress™ Plus Fragment Library Kit	Ion Total RNA-Seq Kit v2 Ion Library Equalizer™ Kit	384 barcodes supported by Torrent Suite™ Software

university of groningen

| 9

MiSeq System Performance Parameters

MiSeq Reagent Kit v2

Read Length	Total Time*	Output
1 × 36 bp	~4 hours	540–610 Mb
2 × 25 bp	~5.5 hours	750–850 Mb
2 × 150 bp	~24 hours	4.5–5.1 Gb
2 × 250 bp	~39 hours	7.5–8.5 Gb

Reads Passing Filter†

Single Reads	12–15 M
Paired-End Reads	24–30 M

Quality Scores††

> 90% bases higher than Q30 at 1 × 36 bp
> 90% bases higher than Q30 at 2 × 25 bp
> 80% bases higher than Q30 at 2 × 150 bp
> 75% bases higher than Q30 at 2 × 250 bp

MiSeq Reagent Kit v3

Read Length	Total Time*	Output
2 × 75 bp	~24 hours	3.3–3.8 Gb
2 × 300 bp	~65 hours	13.2–15 Gb

Reads Passing Filter†

Single Reads	22–25 M
Paired-End Reads	44–50 M

Quality Scores††

> 85% bases higher than Q30 at 2 × 75 bp
> 70% bases higher than Q30 at 2 × 300 bp

university of groningen

| 10

Table 1 Insertion/deletion and substitution errors on read level for benchtop NGS platforms

Platform	Sequencing kit	Library	Strain	Date of sequencing	Indels per 100 bp	Indels per read	Substitutions per 100 bp	Substitutions per read
GSJ	GSJ Titanium	Nebulization / AMPure XP	Sakai	June 2012	0.4011	1.8351	0.0543	0.2484
MiSeq	2 × 150-bp PE	Nextera	Sakai	June 2012	0.0009	0.0013	0.0921	0.1318
MiSeq	2 × 250-bp PE	Nextera	Sakai	September 2012	0.0009	0.0018	0.0940	0.2033
PGM	100 bp	Bioruptor / Ion Fragment Library	Sakai	July 2011	0.3520	0.3878	0.0929	0.1024
PGM	200 bp	Ion Xpress Plus Fragment	Sakai	July 2012	0.3985	0.6811	0.0303	0.0521
PGM	300 bp	Ion Xpress Plus Fragment	Sakai	August 2012	0.7054	1.4457	0.0861	0.1765
PGM	400 bp†	Ion Xpress Plus Fragment	Sakai	November 2012	0.6722	1.8726	0.0790	0.2202

Error rates were calculated by counting indels and substitutions in the mapping against the E.HLO Sakai reference sequence for each uniquely mapped read.
†Kit was not officially available during time of study

Jünemann, nature biotechnology volume 31 number 4 APRIL 2013

university of groningen

| 11

conclusion ?

- no single platform comes out on top
- 454 GSJ wins when read length
- The Illumina MiSeq is best with respect to throughput per run and least number of consensus errors.
- Both the Ion Torrent PGM and the GSJ are well suited for sequencing amplicons

Jünemann, nature biotechnology volume 31 number 4 APRIL 2013

university of groningen

| 12

Analysis

- Commercial software
- In house pipelines
- Settings

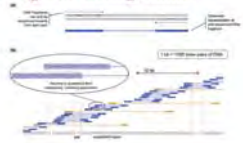
university of groningen

| 13

Data-analysis

Mapping to a reference genome:

Assembling a de novo genome:

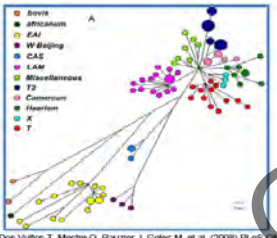


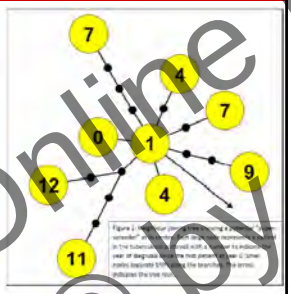
Courtesy of Jonathan Green, HPA Colindale

university of groningen

SNV-analysis

| 14





Dos Vultos T, Mestre O, Rauzier J, Golec M, et al. (2008) PLoS ONE 3(2): e1538. doi:10.1371/journal.pone.0001538
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0001538

Example applications :

- (Recent) evolution of closely related strains.
- Transmission studies
- Minority species/anti-viral resistance studies

university of groningen

| 15

Results vary per pipeline

	Streptococcus pyogenes MGAS315 (1,900,921 bp)			Chlamydia trachomatis D/UW-3/CX (1,042,519 bp)			Escherichia coli 0157:H7 EDL933 (5,526,445 bp)		
	GATK - PHRED 1 (Illumina)								
	GQ cut off value	False -VE	False +VE	False -VE	False +VE		False -VE	False +VE	
Novoalign	29	16	4	3	0		16	4	
	30	29	4	3	0		29	4	
BWA	29	4	0	0	0		4	0	
	40	11	0	0	0		11	0	
Bowtie 2	29	28	76	0	0		28	76	
	40	25	29	5	0		25	31	
Stampy .hsq	20	16	4	4	0		16	4	
	30	23	4	4	0		23	4	
Stampy	26	59	271	16	3		59	271	
	30	59	209	19	3		59	209	
	SAMtools (Illumina)								
Novoalign	29	89	5	62	23		91	39	
	30	92	0	70	10		111	18	
BWA	29	107	0	62	0		95	16	
	30	107	0	73	0		138	10	
Bowtie 2	29	92	35	62	38		94	102	
	30	96	28	192	4		272	33	
Stampy.hsq	29	87	39	62	27		96	82	
	30	92	7	69	23		116	52	
Stampy	29	94	21	69	30		126	173	
	30	96	21	73	30		126	101	

university of groningen

| 16

Core and Pan Genome

› A core genome is the set of genes found across all strains of a pathogen species.

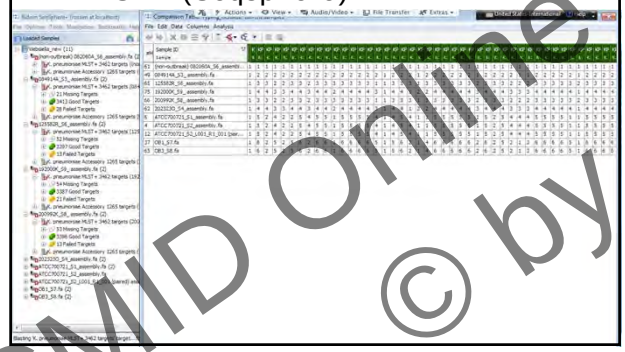
› A pan-genome is the entire gene pool for that pathogen species, and includes genes that are not shared by all strains.

university of groningen

umcg

| 17

MLST+ (Seqsphere)



Ridom SeqSphere+ - Analytical Tools Functionality

university of groningen | umcg

| 19

NGS – UMCG

Microbiology
* MiSeq
* PGM – Pathogenica kit

Pathology
* PGM - kits

Commercial
Own pipeline?

CLC-Bio
DNAstar
Seqsphere
Pathogenica
Open source

Clinical genetics
* MiSeq
* HiSeq

Own pipeline
Separating human/patho gen sequences

university of groningen | umcg

| 20

Typing

› A. Identification of both strain and drug-resistance information about a bacterial infection or colonization

› B. Epidemiology (Epi-Typing)
- find unrelated isolates -> de-compromising
- find related isolates
  - Place, Time, Person, Species, Subtype

› C. Patient management (Patho-Typing)
- Early detection of highly virulent/epidemic clones
- Focused infection control measures

university of groningen

|

Workflow:

› MiSeq/IonTorrent -> FastQ
› CLC Genomics Workbench –> assembly
› Mauve –> alignment:
- in-house script to determine core-genome
- look for unique marker (blast)

› CLC Genomics Workbench –> SNP calling
› ResFinder –> Resitome
› RAxML -> Phylogenetic Tree
› BRIG -> Circular Comparisons
› Seqsphere –> MLST+

Zhou, Kai

university of groningen

umcg

| 22

Outbreak description

› At the time of alert: 6 patients identified to be infected with ESBL producing *K. pneumoniae*

- Rehabilitation clinic (n=3)
- Thoracic chirurgical ward – UMCG (n=3)

university of groningen

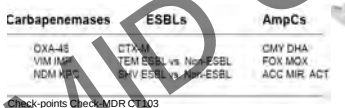
umcg

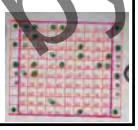
| 23

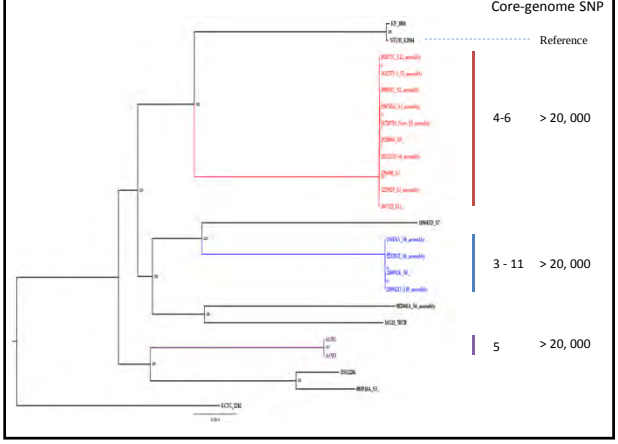
Microarray resistance genes

› Both strains at the rehabilitation center and the the university hospital carried identical (ESBL) genes:

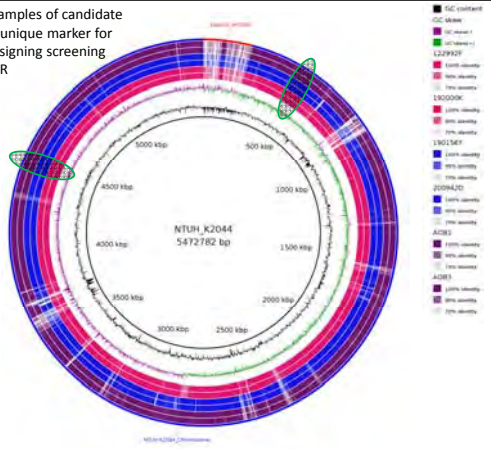
- TEM (wt)
- SHV (wt)
- CTX-M 1







Examples of candidate of unique marker for designing screening PCR



university of groningen

umcg

| 26

Unique region

- Used to develop a real-time PCR for screening (fast, no "background")
- Real-time PCR pick-up only strains in outbreak cluster…
- …and strains with the same MLST-ST
- Lack of database with WGS (FGS)

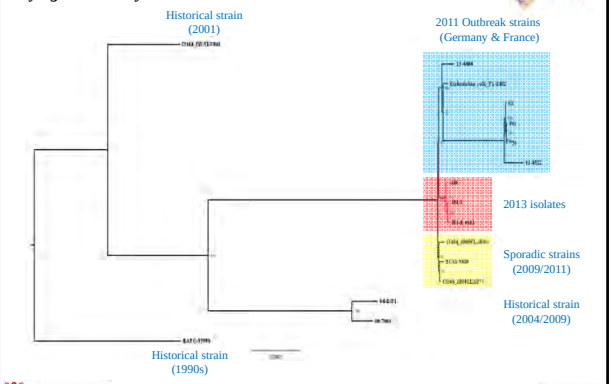
Using NGS in clinical diagnosis of STEC isolates from patients returning after holiday

Isolate ID	Date of isolate	Patient information	Clinical manifestations	MLST	ESBL	Source
338	2013	Travel back from Turkey	HUS			Netherlands (this study)
381-1	2013	Travel back from Turkey	Diarrhea			Netherlands (this study)
381-3	2013	Travel back from Turkey	Diarrhea			Netherlands (this study)
381-4	2013	Travel back from Turkey	Diarrhea			Netherlands (this study)
7N (LB233637-07N)	2011	Unknown	Unknown			Germany
8G (LB227687-08G)	2011	Unknown	Unknown			Germany
9Z (LB227103-09Z)	2011	Unknown	HUS			Germany
LB226692	2011	Unknown	HUS			Germany
TY-2482	2011	16-year-old girl	HUS	ST-678	+	Germany
HUSEC041	2001	Child	HUS	ST-678	-	Germany
55989	1990s	HIV patient	Diarrhea	ST-678	-	Central African Republic

university of groningen

umcg

Phylogenetic analysis



university of groningen

RAxML (1000x bootstrap)

umcg

university of groningen

umcg

| 29

Amplicon based sequencing (Pathogenica)

Acinetobacter baumannii	Enterobacter cloacae
Clostridium difficile	Klebsiella pneumoniae
Escherichia coli	Proteus mirabilis
Enterococcus faecalis	Pseudomonas aeruginosa
Enterococcus faecium	Coagulase-negative Staph (epidermidis, saprophyticus)
Enterobacter aerogenes	Staphylococcus aureus

CARB	PER	ermA
CMY	SHV	vanA
CTX-M	VEB	vanB
GES	VIM	mecA
IMP	NDM	mexA
KPC	OXA	TEM

to monitor ICUs and other high-risk areas for nosocomial infection

resistance gene presence

strain typing

equal to 144 real-time PCRs

university of groningen

umcg

| 30

cross-sectional survey in a nursing home on direct material

› Presented at ECCMID

In collaboration with Jacobien Veenemans and Jan Kluytmans, Amphia Hospital Breda

university of groningen umcg

| 31

Final thoughts

› high-throughput sequencing approaches can make a significant contribution to the investigation of outbreaks

› the integration of WGS **with epidemiological investigation**, diagnostic assays and antimicrobial susceptibility testing is radical changing clinical microbiology

› several challenges remain before WGS can be routinely used in outbreak investigation and clinical practice

university of groningen

| 32

Challenges

› Robotics: preparative and processing
› Genomics LIMS
› IT infrastructure (storage)
› Wet-lab staff
› Bioinformatics staff

university of groningen umcg

| 33

Challenges

› Moving towards culture-independent approaches (metagenomics) is key for diagnostics

› Standardization in genomics
- National and international initiatives
- Biological databases and maintenance
- Effective data and knowledge sharing

university of groningen

| 34

Workshop 'Rapid NGS for Clinical & Public Health Microbiology'
March 12th – 14th, 2014; Univ. Münster

First announcement

Workshop
'Rapid NGS for Identification and Typing of ESBL and CRE'
October 8th – 10th, 2014, UMCG Groningen, NL

Organizers

Molecular Typing Methods for Pathogens, ESCMID Postgraduate Education Course

30 June – 4 July 2014, Lyon, France

Organizers
- ESCMID Study Group for Epidemiological Markers (ESGEM)
- Lyon East Medical School

Supporters
- ESCMID Study Group for Legionella Infections (ESGLI)
- ESCMID Study Group for Molecular Diagnostics (ESGMD)
- ESCMID Study Group for Staphylococci and Staphylococcal Diseases (ESGS)

Course Coordinators
- Alex Friedrich, Groningen, The Netherlands
- Frédéric Laurent, Lyon, France

ESGEM

ESGMD

university of groningen

umcg

| 35