A year in clinical bioinformatics

Helena Seth-Smith, PhD

October 2018
"...the application of next generation sequencing to clinical samples in order to recover information of clinical relevance"
Meeting report

Messages from the first International Conference on Clinical Metagenomics (ICCMg)

Etienne Ruppé a,*, Gilbert Greub b, Jacques Schrenzel a,c

Fig. 2. Pictures from the conference. C: Aerial view of the attendance.
ICCMg

Quality

Speed

Applications

Communication

- contamination
- ISO accreditation: ISO/IEC 17025
- gold standards
- bottleneck of DNA extraction
- sequencing platform
- analysis, universal pipelines
- resistance determination
- metagenomic diagnosis, novel pathogens
- therapeutics
- transmission detection and outbreak control
- reporting
- data storage/sharing
- interactive analysis

Theme section editors: Jacob Moran-Gilad and Panayotis T. Tassios — Next generation sequencing (NGS) for clinical microbiologists and infectious disease specialists dealing with bacterial infections

Helena Seth-Smith
Clinical microbiology and NGS: Quality

- ISO accreditation
  - wet and dry lab SOPs
  - quality management system
- Validation
  - improvements: maintaining validity
- Backwards compatibility
- Reference material
- External Quality Assessment
- Pipelines:
  - commercial vs open source
  - variant calling

Gargis J Clin Microbiol 2016
Rossen Clin Microbiol Infect 2018
Balloux Trends Microbiol 2018
NGS workflow at University Hospital Basel

1. Sample
2. Isolate
3. MALDI-TOF MS
4. DNA Extraction
5. DNA Quantification
6. Library Preparation
7. Sequencing
Sequencing platforms

4-24 samples per run

24-96 samples per run

long reads
QC pipeline at University Hospital Basel

Raw reads ➔ Trimming ➔ High quality reads ➔ De novo assembly ➔ Genome scaffolds ➔ Annotation ➔ Annotated genome assembly

Remapping ➔ Read quality ➔ Coverage, Variants ➔ Genome size, N50 ➔ 16S rRNA gene ➔ rMLST ➔ AMR genes ➔ Virulence genes

QC
ID
Phenotype prediction

Meta-genomic analysis

Helena Seth-Smith
18.10.2018
Clinical microbiology and NGS: Applications

- Species identification
- Factor identification
  - resistance determinants
  - virulence determinants
- Outbreak and transmission analysis
  - typing: highest discrimination
  - infection control and epidemiology
Analysis software

Ridom SeqSphere: cgMLST typing

**aroE**

**cgMLST loci**

**gdh**

**pdhC**

**pgm, adk**

**abcZ**

CLC Genomics Workbench: SNP analysis

Seth-Smith *EID* submitted
Public health: *Shigella sonnei* as an STI

- First cases in Switzerland
- Case histories underpinned by genomic analysis
- Higher antimicrobial resistance (azithromycin, quinoline)
- Related to known MSM associated clades
- Carry pKSR100 with macrolide resistance determinants
Public health: *Campylobacter* in Basel

- Identifying foodborne sources by hierarchical genomic analysis

Isolates
n= 685

- Stool isolates 2015
- Stool isolates 2016
- Stool isolates 2017
- Stool isolates 2018
- Blood isolates 2015
- Blood isolates 2016
- Blood isolates 2017
- Blood isolates 2018
- Chicken meat isolates 2015
- Chicken meat isolates 2016
- Chicken meat isolates 2017
- Chicken meat isolates 2018

Seth-Smith *et al* in preparation
Public health: *Campylobacter* in Basel

- Transmission cluster n= 20
- Patient only cluster n= 26
- Chicken only cluster n= 47

Seth-Smith *et al* in preparation
Outbreaks: Vancomycin Resistant Enterococci (VRE)

- Leading multidrug resistant hospital pathogen
- >400 VREs sequenced
- Ridom SeqSphere: MLST (n=7) and cgMLST (n=1423) loci
- Within clusters, CLC Genomics Workbench: SNP trees

- Highly virulent ST796, Australia (2011)
- Bern outbreak Dec-Apr (n=76)
  - Wassilew, Eurosurveillance, 2018
- Dec-Oct, n=241 patients

Buultjens PeerJ 2017
Outbreaks: VRE

MLST Sequence Type:
- 1: 203
- 17: 375
- 18: 555
- 32: 721
- 78: 761
- 80: 787
- 117: 796
- 133: 820
- 192: no group assigned
Outbreaks: VRE
Outbreaks: VRE

Coloured by source hospital
Outbreaks: VRE
Outbreaks: VRE

Coloured by patient with repeat samples
Reporting

Carriço *Clin Microbiol Infect* 2018  
Gómez-López *Brief Bioinf* 2017
Molekulare Typisierung mittels Genom-Sequenzierung (WGS)

**Zusammenfassung**


**Interpretation**

Die Isolate von XX ist ein STXXX und trägt ein JXX Gen. Die core genome MLST (cgMLST) Grafik zeigt die Verwandtschaftsverhältnisse der Isolate auf.

---

**Diagramme**

1. Minimal Spanning Tree von der Core Genome Multi Locus Sequencing Typing Analysis (cgMLST)
2. SNP Tree

---

**Referenz**

Crisan Peer J 2017

Helena Seth-Smith

18.10.2018
Data sharing

Atul Butte
@atulbutte

I really believe that in 5 years, any academic medical center that ISN’T using their clinical data to improve the practice of medicine will be seen as irresponsible!

9:51 AM - 4 Oct 2018 from San Francisco, CA

Fig. 1 Number of genome assemblies available in the National Center for Biotechnology Information (NCBI) database per year

Tagini *Eur J Clin Microbiol Infect Dis* 2017
Data coordination

- Swiss Pathogen Surveillance Platform

- **nextstrain** for interpretation and modelling

*Mycobacterium tuberculosis* sublineage 4.6.2

Emma Hodcroft, Walker *Lancet ID* 2018, Gagneux

http://spsp.ch
Conclusions

- Continuous improvement necessary
  - speed
  - resistance, mobile elements, plasmids, recombination…
  - data mining
  - machine learning

- Coordination of sequence and metadata
  - surveillance platforms
  - online tools

- Standardised workflows as goal
  - pathogens are plastic: can one size fit all?
Acknowledgements

Dr Adrian Egli PD
Dr Daniel Wüthrich
Josiane Reist
Dr Ferdinando Bonfiglio
NGS Team
Division of Clinical Microbiology

helena.seth-smith@usb.ch
https://appliedmicrobiologyresearch.net/
https://www.unispital-basel.ch/
Projects: outbreaks: VRE: dynamics

Swiss samples 2018

Australia 2014-16

Buultjens PeerJ 2017
Leong Sci Rep 2018

Helena Seth-Smith

18.10.2018
Outbreaks: VRE

FOKUS: Der Kampf gegen einen gefährlichen Keim
Aus 10vor10 vom 07.09.2018, 21:50 Uhr
Also, if you want to highlight the importance of “bioinformaticians” or “clinical bioinformaticians”, I really like the figure 1 of this paper: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-8-125; Sboner Genome Biol 2011

and also this paper: https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbx144/4565521; Gómez-López Briefings in Bioinformatics 2017
Projects: novel

- *Candida albicans*:
  - typing and Echinocandin resistance detection
- *Aspergillus fumigatus*
  - typing and Azole resistance detection
- Description of species / cases:
  - *Blastobotrys*
  - *Auritidibacter*
  - *Lawsonella*
  - *Mycobacterium basiliense*
Outbreaks: VRE