2nd Conference

"Rapid Microbial NGS and Bioinformatics: Translation Into Practice"

Hamburg/Germany, June 9-11, 2016

Base pairs 2 Bedside

Bacterial Genome Detectives

A conference organized by the EU funded PathoNGenTrace project (Grant Agreement No 278864) and the Robert Koch Institute.

RaMi-NGS

More information:
www.RaMi-NGS.org

Abstracts

Forschungszentrum Borstel

GenoScreen

APPLIED MATHS

ridom BIOINFORMATICS

Westfälische Wilhelms-Universität Münster

PiEXT

University of Oxford

Robert Koch Institut

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No 278864.

RaMi-NGS

www.RaMi-NGS.org

The 2nd Rapid Microbial NGS and Bioinformatics: Translation Into Practice Conference

The consortium of the EU-funded project PathoNGenTrace is delighted to welcome you to the 2nd **"Rapid Microbial NGS and Bioinformatics: Translation Into Practice"** conference which takes place in Hamburg/Germany from June 9th to 11th, 2016. The bi-annual meeting gathers experts from all over the world active in applying Next Generation Sequencing (NGS) techniques to discover the epidemiology, anti-microbial resistance, ecology and evolution of microorganisms.

The 2016 conference is targeted at scientists from the academic as well as the public health, food, veterinary, clinical and microbial forensics sectors. In addition, the conference particularly addresses the interests of NGS data software developers intending to build a bridge between software developers and end-users.

Program chairs:
Stefan Niemann (Research Center Borstel, Germany), Dag Harmsen (University of Muenster, Germany), Lothar H. Wieler (Robert Koch Institute Berlin, Germany)

Scientific program advisors:
Thomas Dobner (Heinrich Pette Institute, Germany), Rolf Horstmann (Bernhard Nocht Institute for Tropical Medicine, Germany), Martin Maiden (University of Oxford, United Kingdom), Bruno Pot (Applied Maths NV, Belgium), Philip Supply (GenoScreen, France)

Venue:
Erika-Haus (University Medical Center Hamburg-Eppendorf, Germany)
Martinistraße 52, 20246 Hamburg

Date:
June 9th - 11th, 2016

Website:
http://rami-ngs.org/

Twitter:
#RaMi-NGS

Program Management:
Birte Kretschmer, Ina Filla (Eurice GmbH, Germany)

Table of Content

Oral presentations (listed in order of the scientific program)

Session: NGS for Public Health surveillance

PLENARY LECTURE 1

Genome based surveillance and implication for public health

Lothar H. Wieler, Torsten Semmler, Tim Eckmanns, Andreas Nitsche, Guido Werner, Bernhard Renard

Robert Koch Institute, Berlin, Germany

Diagnostics and surveillance are key pillars Public Health actions are based on. To implement specific and rapid interventions the Robert Koch Institute has to utilize cutting edge technology. Next generation sequencing (NGS) has paved the way for developing surveillance platforms which - particularly in the field of Infectious diseases - enable timely exchange of information for rapid and specific risk assessment. This digitalised information exchange (Public Health 4.0) also is a key technology implemented in the Robert Koch Institute's surveillance of bacterial as well as viral pathogens. While analyses of bacterial pathogens focuses on identifying clonal relationships, virulence and antimicrobial resistance features for risk assessment, diagnostics of viral pathogens concentrates on metagenomic approaches. NGS and bioinformatics core facilities guarantee coordinated development of an integrated surveillance enabling rapid outbreak detection and implementation of interventions. This paper will give some insights into the current use of NGS in the above mentioned areas. In addition we will give insight into further research projects which do not focus on obvious Public Health needs, but highlight further applications of NGS.

PLENARY LECTURE 2

ECDC roadmap for integration of molecular and genomic typing into European level surveillance 2016-19

Marc Struelens, Barbara Albiger, Joana Revez and Katrin Leitmeyer

European Centre for Disease Prevention and Control, Stockholm, Sweden

In 2015 ECDC strategy sated that, in five years' time, the agency will have contributed to the establishment of standards and systems enabling the EU wide use of Whole Genome Sequencing (WGS) as the method of choice for typing of microbial pathogens, replacing other methods. This will improve the accuracy and effectiveness of disease surveillance, outbreak investigation and evaluation of prevention policies through enhanced assessment of disease and drug resistance transmission dynamics. The rapid shift of practice of molecular typing to WGS for public health surveillance and outbreak studies, as well as the rise of new public health needs led ECDC to revise its roadmap priority by pathogens/diseases. The roadmap - Version 2.1, 2016-19 presents the recommended list of pathogens/diseases and technical options for the medium term integration of molecular/genomic typing into EU level surveillance and epidemic preparedness. This is a fast-moving area, where there are marked variations in the pace of change between pathogens/diseases and between countries. ECDC has prioritised its implementation plans according:

- Public health priority and potential added-value
- Feasibility and capacity
- Level of resource required
- Potential synergies with activities of third parties.

Information related to these criteria was provided by the EU Disease Networks and a survey by National Focal Points for Microbiology (NMFP) to map the EU capability to use WGS for national surveillance. A Molecular typing for Surveillance Task Force composed of both (NMFPs) and National Focal Points for Surveillance ranked 13 proposals for integration of disease specific molecular and genomic typing data into EU level surveillance.

Based on the above, ECDC categorised pathogens/diseases by time priority for implementation:

- EU wide WGS-based surveillance systems in the near term: start implementation of WGS-based surveillance for L. monocytogenes, Neisseria meningitidis, carbapenemase-producing Enterobacteriaceae and antibiotic-resistant N. gonorrhoeae ;
- Operationalisation of WGS-based surveillance systems deferred until the required technical capacity across the EU/EEA is met: for surveillance of human influenza virus, Salmonella enterica, Shiga-Toxin producing E. coli (STEC) and multidrug-resistant Mycobacterium tuberculosis;
- Exploring opportunities and challenges: PCR-ribotyping for Clostridium difficile surveillance, and sequence-based surveillance of anti-viral drug resistance in human immunodeficiency virus (HIV) and Hepatitis C virus (HCV);

- Postponed until next roadmap revision in 2018: West Nile virus (WNV) and meticilin-resistant Staphylococcus aureus (MRSA).

To foster the implementation of its WGS strategy, ECDC together with public health partners in the EU member states is currently developing pilot implementation studies of WGS-based surveillance for the top priority diseases in collaboration with relevant EU and global initiatives. It will further evaluate the public health benefits of these activities. ECDC supports the EU/EEA countries for the transition to the genome-based typing methods by providing technical guidance, contributing to set up international surveillance standards and nomenclature as well as providing multidisciplinary training.

Session: Tracing foodborne pathogens

KEY NOTE LECTURE 1

Implementing NGS and WGS strategies for surveillance, attribution, and outbreak investigation of Campylobacter

Martin C. J. Maiden

University of Oxford, Oxford, United Kingdom

Campylobacter remains one of the most important sources of food-borne bacterial gastrointestinal disease worldwide. In high income countries campylobacteriosis has a high reported morbidity but this is perhaps ony 20-% of the total disease burden. While most cases are relatively mild and self-limiting, the economic burden is large because of the volume of cases and some cases are severe, with occasional severe neuropathological sequelae. In Europe about 90% of cases are *Campylobacter coli*, with most of the remaining cases *Campylobacter jejuni*.

In the early 2000s, our understanding of the epidemiology of *Campylobacter* was transformed by the implementation of high-throughput sequence-based typing approaches. For the first time, these approaches enabled the precise identification of different genotypes and their relationship to isolation source, improving epidemiology at all levels from the attribution of genotypes to source to the identification of disease outbreaks. In recent years, whole genome sequencing (WGS) has been use for these functions. *Campylobacter* genomes can be readily assembled from next generation sequencing data to produce high quality draft genomes of very high accuracy which can be annotated and deposited on public access databases such as PubMLST.org/neisseria. These data represent more than 95% of the genome and provide a large volume of detailed genetic data in addition to the loci used for conventional typing. The Oxfordshire *Campylobacter* surveillance study has generated data on all *Campylobacter* isolates obtained in the Oxfordshire area, which represents approximately 1% of the UK population, from 2003 with WGS data available from 2010. This has enabled the validation of these approaches as a routine tool and shows the feasibly of these approaches.

Practical application of Whole Genome Sequencing in the food industry

Katia Rouzeau-Szynalski[1], Deborah Moine[2], Caroline Barretto[1], Catherine Ngom-Bru[1], Coralie Fournier[2], Mohamed Kassam[2], Johan Gimonet[1], Pieter Breeuwer[3] and Leen Baert[1]

[1]Nestec Ltd - Nestlé Research Center, Lausanne, Switzerland; [2]Nestlé Institute of Health Sciences SA, Lausanne, Switzerland; [3]Nestlé Nutrition, La Tour-de-Peilz, Switzerland

Due to its high discriminatory power, authorities are currently using whole genome sequencing (WGS) in parallel to classical typing methods, such as PFGE, to establish links between food bacterial isolates and clinical isolates to support foodborne outbreak investigations. The potential value and challenges of WGS for the food industry was assessed by investigating a historical environmental factory contamination. Classical serotyping of the isolates indicated two serotypes Salmonella Kentucky and Salmonella enterica subsp. enterica 8, 20: i:- (monophasic strain).

The WGS analysis was carried out as follows: DNA was extracted from 108 Salmonella factory isolates and sequenced with 250 bp paired-end reads at a coverage of approximately 100x on Illumina MiSeq or HiSeq platforms. In parallel, selected isolates from our internal collection were sequenced with PacBio long read sequencing technology to create a reference genome. Afterwards the bacterial genomes were de novo assembled using hierarchical genome assembly process (HGAP), followed by genome closure using Circlator.

Bioinformatics analysis consisted in clustering the isolates and then performing the in depth SNP (Single Nucleotide Polymorphism) analysis with the adapted CFSAN-FDA pipeline including the closest public genomes available to put the results into a biological context. The bioinformatics pipeline was set up as following: (I) mapping of Illumina reads on the PacBio reference genome, (II) SNP identification, (III) phylogenic tree and SNP distance matrix construction.

WGS analysis showed that the isolates serotyped as Salmonella enterica subsp. enterica serovar Kentucky and Salmonella enterica subsp. enterica 8, 20: i:- (monophasic strain) are in fact belonging to the same serotype. WGS confirmed that the isolates were nearly identical (< 3 SNPs differences) and thus that these isolates were originating from the same contamination source. WGS was able to provide a more accurate picture than serotyping to support source tracking in a factory environment.

This study also highlighted challenges for a food company to implement the WGS approach. WGS analysis requires several competences (e.g. in sequencing, bioinformatics and biological interpretation). Especially bioinformatics analysis which is a crucial step. A user-friendly interface which could be used and understood by microbiologists and quality managers is one element which is needed to facilitate the use of WGS in food companies.

Source identification of a Campylobacter fetus outbreak through next generation sequencing

Birgitta Duim, Aldert Zomer, Hetty Koppenaal, Fleur Groenendijk, Marcel van den Berge, Erwin Verkade, Aloys Tijsma, Marcel Spierenburg, Kees Verduin, Margreet te Wierik and Jaap A. Wagenaar

Utrecht University, Department of Infectious Diseases and Immunology, The Netherlands

Background: Campylobacter fetus subsp. fetus (Cff) is primarily a veterinary pathogen and occasionally associated with severe infections in humans. The high level of genome similarity in Cff requires high-resolution methods for studying the epidemiology. From May 2015 to October 2015 an unexpectedly high number (6) of invasive Cff infections in human patients were reported in the province of Zeeland (n=5) and the neighbouring province of Brabant (n=1) in the Netherlands. Based on epidemiological information and patient questionnaires, it was concluded that the patient from Brabant most probably acquired the infection abroad, whereas all patients in Zeeland consumed unripened sheep cheese from unpasteurized milk. For 4 of the patients, the product could be traced to one sheep farm, while for the 5th patient a second farm was indicated. Microbiological investigation of sheep faeces revealed the presence of Cff in sheep from the first farm but not of the second. Methods: To support the source investigation, MiSeq genome sequencing was used to investigate isolates from patients, from the suspected sheep flock, and a set of Dutch Cff reference genomes from human and ruminant origin. The Harvest suite was used for core-genome alignment and reconstruction single-nucleotide polymorphism (SNP) based phylogeny. Results: The SNP profiles irrevocably showed that 5 of the outbreak isolates belonged to a distinct clone while the 6th isolate from the patient from Breda belonged to a separate clone. The 5 isolates were different from epidemiological unrelated human isolates obtained during the same period in the Netherlands. The outbreak isolates generally differed between 0-7 SNPs in their core genome, while reference isolates from the same MLST sequence type differed by at least 26 SNPs. The outbreak isolates harboured the determinants for the serum resistance UDP-galactopyranose mutase (glf) and the surface array protein (Sap), both associated with invasive infections. NGS of the sheep isolates identified isolates with identical genomes to the patient isolates, confirming the epidemiological source identification. Conclusions: NGS combined with epidemiological data identified the most probable scenario, that faecal contamination of milk from a single sheep farm resulted in Cff contaminated cheese, leading to an outbreak of invasive Cff infections after cheese consumption. The link between the 5th patient and the first farm remains to be resolved.

Session: Metagenomics

KEY NOTE LECTURE 2

Computational analysis of microbiome sequencing data

Daniel Huson,

University of Tuebingen, Center for Bioinformatics, Tuebingen, Germany

Taxonomic and functional analysis of microbiome sequencing data is challenging due to the large number and size of samples that usually considered (hundreds of samples, billions of reads). In this talk, we first discuss the computational aspects of the challenge and describe a number of methods and programs that we have developed to address it in an efficient manner.

KEY NOTE LECTURE 3

SLIMM: Species level identification of microorganisms from metagenomes

Temesgen H. Dadi[1,2,3], Martin S. Lindner[4], Torsten Semmler[2,4], Bernhard Y. Renard[4], Lothar H. Wieler[2,4] and Knut Reinert[1]

[1]Free University Berlin, Institute for Bioinformatics, Berlin, Germany; [2]Free University Berlin, Institute for Microbiology and Epizootics, Berlin, Germany; [3]International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany; [4]Robert Koch Institute, Berlin, Germany

Identification and quantification of microorganisms is an important step in studding the alpha and beta diversities within and between microbial communities respectively. Both identification and quantification of a given microbial community can be carried out using whole genome shotgun sequences with less bias than using 16S-rRNA sequences. Shared regions of DNA among reference genomes and taxonomic units pose significant challenge to assign reads correctly to their true origins. The existing microbial community profiling tools deal with this problem by either preparing signature based unique references or assigning an ambiguous read to its least common ancestor up in a taxonomic tree. The former method is limited to making use of the reads which can be mapped to the curated regions while the later suffer from the lack of unique reads at lower (more specific) taxonomic ranks. Moreover, even if the tools exhibited generally good performance in calling the organisms present in a sample there is a room for improvement in calling the correct relative abundance of the called organisms. We present a new method Species Level Identification of Microorganisms from Metagenomes (SLIMM) which addresses the above issues by using coverage information of reference genomes to knockout unlikely genomes from the analysis and subsequently gain more unique reads to assign at lower ranks of a taxonomic tree.

Session: NGS implementation in Public Health

Recent advances of genome-based surveillance for tracing emergence and spread of MDR hospital pathogens – the perspective of the National Reference Centre for Staphylococci and Enterococci

Birgit Strommenger, Franziska Layer, Christiane Cuny, Ingo Klare, Jennifer Bender, Torsten Semmler, Bernhard Renard, Stephan Fuchs and Guido Werner

Robert Koch Institute, Berlin, Germany

The National Reference Centre (NRC) for Staphylococci and Enterococci collaborates with more than 250 microbiological, clinical laboratories in Germany and receives about 4000 staphylococcal and 1000 enterococcal strains per year. Most frequent requests address validation of resistance diagnostics and analyses regarding strain relatedness with respect to outbreak elucidation. NGS-based strain characterization offers maximum resolution and is, in principle, capable to provide all relevant diagnostic (virulence, resistance) and typing information (cgMLST, SNP and variant detection). Here we demonstrate applicability and suitability of NGS-based approaches for answering classical NRC requests and elaborate benefits but also limitations of the currently available techniques and software solutions.

We will shortly introduce the following studies on Staphylococcus:

- using NGS-based approaches for studying 280 S.aureus isolates received during one month of routine NRC work
- outbreak investigations based on NGS-data (when time/space allows)
- using pooled NGS data to extrapolate plasmid sequences of coagulase-negative staphylococci

We will shortly introduce the following studies on Enterococcus:

- using NGS data to disclose the emergence of a prominent ST192 VRE lineage and the acquisition of a chromosomally located vanB-resistance transposons
- using NGS-based approaches for studying strain transmission within hospital ICUs based on a set of 130 VRE isolates

Conclusions. NGS data provide important information relevant to most of the requests addressed to a microbial reference laboratory. Whereas the benefit of NGS-based strain typing over classical sequence- or fragment-based typing in case of outbreak analyses is beyond doubt, its general applicability for all (other) NRC-relevant typing and confirmation requests has limitations (resistance prediction, resolution of accessory genome content/plasmids, time-to-result, etc.). Although costs and running/analysis times have dropped dramatically in recent years, it is nowadays still impossible to implement NGS as a frontline diagnostic method for all incoming routine requests; establishing a guideline for considering distinct aspects for a valid strain pre-selection is inevitable.

Next-generation sequencing of a large MRSA outbreak shows that transmission likely occurred between Dutch healthcare centers

Thijs Bosch, Karen Heemstra, Sandra Witteveen, Masja Leenderste, Martin Becker, Annet Troelstra, Christine Vandenbroucke-Grauls, Jan Kluytmans and Leo Schouls

National Institute for Public Health and the Environment, Bilthoven, The Netherlands

In 2013, the first eight MRSA isolates with MLVA-type MT4239 emerged in the Netherlands. The prevalence of MT4239 MRSA rapidly increased to 121 isolates in 2015. Until now, 171 unique MT4239 isolates have been submitted to the Dutch MRSA reference center (RIVM), but only 151 carried the mecA gene. Fifteen isolates were mecA negative, and both mecA positive and negative isolates were cultured from five persons. Virtually all MT4239 isolates from 2013 originated from a single healthcare center (HCC), whereas the isolates from 2014 onwards were mainly submitted from another HCC closely located to the first HCC.

In order to elucidate the origin and transmission route of this S. aureus variant, NGS was performed on a subset of MT4239 isolates. The subset included both mecA positive and negative isolates from multiple sampling years and from both HCCs. NGS data were used for core genome MLST (cgMLST) and characterization of the SCCmec region.

cgMLST based on 1,210 genes in the context of 119 reference genomes showed that all MT4239 isolates clustered closely together in a minimum spanning tree (MST). A MST based on 1,816 genes present in all MT4239 isolates, showed that two isolates from 2013 differed in 75 genes from the isolates originating from 2014, 2015 and 2016 that all clustered very closely together. The latter isolates were all obtained from the second HCC. The distance between the two 2013 isolates each obtained from a different HCC, was 14 genes. Two mecA negative isolates from 2014 and 2015 clustered in the group of mecA positive isolates from 2014, 2015 and 2016 with limited distance between the isolates. Two other mecA negative isolates, cultured from the same patient on the same day, differed in 10 genes, whereas the distance to the closest neighbor was 65 genes. All isolates carried SCCmec type V (5C2&5), but only a single isolate from 2016 had the complete cassette. Both isolates from 2013 had a 4.6 kb deletion in the distal part of the SCCmec. All other isolates had a 7.6 kb deletion, also in the distal part of the cassette. The four mecA negative isolates still carried parts of the SCCmec cassette and these remaining SCCmec sequences were identical in the four isolates.

At this moment, the outbreak of MT4239 is restrained, but the source of the outbreak has not yet been identified. NGS showed that transmission between the HCCs has likely occurred. To identify the source NGS analysis of MRSA from patients and staff will be essential.

NGS surveillance and diagnostics, from development to routine

Erik Alm

Public Health Agency of Sweden, Stockholm, Sweden

The Public Health Agency of Sweden has traditionally used a wide plethora of different molecular and serological techniques for typing, surveillance and diagnosis of pathogenic agents. We are currently in the process of replacing all traditional methods with next generation sequencing. With an institute-wide IonTorrent based NGS platform we are now routinely sequencing 100-200 samples every week with an approximately 30-hour response time, and have replaced traditional cumbersome methods for a number of agents, and improved the resolution in outbreak investigations.

We have set up automatic analysis pipelines for bacterial/viral typing, detection of antibiotic resistance genes, phylogenetic outbreak analysis, allowing microbiologists to respond without additional input from bioinformaticians. Additionally, metagenomics methods for hypothesis-free detection of pathogens are being employed, using both a 16S approach for bacteria and a general metagenomics approach for pathogens of unknown taxonomic origin.

I will present the steps we have taken to get where we are in using NGS in the routine workflow, with special emphasis on bioinformatics, and detail the plans ahead to move most remaining analyses to NGS.

Session: Antimicrobial resistance – genome to phenome

KEY NOTE LECTURE 4

The comprehensive antibiotic resistance database – a platform for antimicrobial resistance surveillance

Andrew McArthur[1], Baofeng Jia[1], Amogelang R. Raphenya[1], Peiyao Guo[2], Kara Tsang[1], Biren Dave[1], Brian Alcock[1], Briony Lago[1], Nicholas Waglechner[1] and Gerard D. Wright[1]

[1]McMaster University, Hamilton, Canada; [2]University of Waterloo, Canada

Antimicrobial resistance (AMR) is among the most pressing public health crises of the 21st Century. Despite the importance of resistance to health, this field has been slow to take advantage of genome scale tools. Phenotype based criteria dominate the epidemiology of antibiotic action and effectiveness. There is a poor understanding of which antibiotic resistance genes are in circulation, which a threat, and how clinicians and public health workers can manage the crisis of resistance. However, DNA sequencing is rapidly decreasing in cost and as such we are on the cusp of an age of high-throughput molecular epidemiology. What are needed are tools for rapid, accurate analysis of DNA sequence data for the genetic underpinnings of antibiotic resistance. In an effort to address this problem, we have created the Comprehensive Antibiotic Resistance Database (card.mcmaster.ca). This database is a rigorously curated collection of known antibiotics, targets, and resistance determinants. It integrates disparate molecular and sequence data, provides a unique organizing principle in the form of the Antibiotic Resistance Ontology (ARO), and can quickly identify putative antibiotic resistance genes in raw genome sequences using the novel Resistance Gene Identifier (RGI). Here we review the current state of the CARD, particularly recent advances in the curation of resistance determinants and the structure of the ARO. We will also present our plans for development of semi- and fully-automated text mining algorithms for curation of broader AMR data, construction of meta-models for improved AMR phenotype prediction, and release of portable command-line genome analysis tools.

KEY NOTE LECTURE 5

Deeplex-MycTB, an all-in-one NGS-based diagnostics of M. tuberculosis

Cyril Gaudin[1], Stéphanie Duthoy[1], Mathilde Mairey[1], Mihanta Ramaroson[1], Katrien De Bruyne[2], Davy Hollevoet[2], Bram de Greve[2], Yannick Laurent[1], Guy Delcroix[3], Nadine Lemaître[3], Christelle Vluggen[4], Karine Soutaert[4], Vanessa Mathys[4], Leen Rigouts[5], Bouke de Jong[5], Hannes Pousseele[2], Caroline Allix-Béguec[1] and Philip Supply[1,6]

[1]GenoScreen, Lille, France; [2]Applied Maths, Sint-Martens-Latem, Belgium; [3]University Hospital, Lille, France; [4]Scientific Institute for Public Health, Brussels, Belgium; [5]Institute of Tropical Medicine, Antwerp, Belgium; [6]Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019-UMR 8204-Center for Infection and Immunity of Lille, France

Standard drug susceptibility testing by mycobacterial culturing is slow. Albeit faster, current molecular tests only interrogate a limited fraction of mutations commonly associated with drug resistance. Application of whole genome sequencing on clinical isolates for prediction of drug resistance requires costly selective enrichment methods, and yields relatively low genome coverage depths. In order to by-pass these limitations, a new assay directly applicable on clinical samples and called Deeplex®-MycTB has been developed. This assay uses ultra-deep sequencing of a single 24-plexed amplicon mix for simultaneous mycobacterial species identification, genotyping and prediction of drug resistance of Mycobacterium tuberculosis complex strains. Forty-eight samples can be analysed in one MiSeq run. The test includes the main gene targets associated with first- and second-line drug resistance. The achieved high coverage depths enable detection of low frequency variants that can cause clinical resistance. The kit design includes the use of a dedicated cloud-based application, for easy analysis and interpretation of the sequence data to identify drug resistance-associated mutations, using integrated databases compiling reference results e.g. from systematic whole genome sequencing-based studies (1). First evaluation results will be presented.

References:

- Walker TM et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* 15:1193-202 (2015).

Using whole genome sequence data to predict the antimicrobial resistance profiles of human campylobacteriosis isolates

Alison J. Cody[1], Jasna Kovač[1,2], Noel D. McCarthy[1,3,4,5], James E. Bray[1], Kate E. Dingle[6] and Martin C. J. Maiden[1,4]

[1]University of Oxford, Department of Zoology, Oxford, United Kingdom; [2]University of Ljubljana, Department of Food Science and Technology, Ljubljana, Slovenia; [3]Health Protection Agency, London, United Kingdom; [4]University of Warwick, Warwick Medical School, Coventry, United Kingdom; [5]University of Oxford, NIHR Health Protections Research Unit in Gastrointestinal Infections, Oxford, United Kingdom; [6]University of Oxford, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford, United Kingdom

The determination of antimicrobial susceptibility and resistance in Campylobacter currently relies on phenotypic approaches that have variable reliability and reproducibility. Analysis of whole genome sequence (WGS) provides an alternative paradigm for investigating and reporting such data. Antibiotic resistance and susceptibility genotypes, together with the inheritance patterns of resistance-conferring alleles, were determined for 2,556 human campylobacteriosis isolates collected over a three year period, from Oxfordshire, United Kingdom, using BIGSdb analysis tools. Multi-locus sequence typing (MLST), core-genome MLST (cgMLST), and antibiotic resistance-associated coding sequences were identified in these publicly available WGS data, available in the PubMLST.org/campylobacter database. Ciprofloxacin resistance–associated gyrA mutations were identified in 40.2% of isolates, and were significantly associated with ST-353, ST-354, ST-464, ST-607 clonal complexes and unassigned C. jejuni isolates (P>0.05). The tetO locus, identified in 40.5% of isolates, was found located on either chromosomal or plasmid contigs, and was significantly associated with ST-206, ST-354, ST-443, ST-464, ST-574 complexes and unassigned C. jejuni (P>0.05). Genotype data for gyrA and tetO accurately predicted resistance phenotypes, regardless of other resistance mechanisms. Phylogenetic analyses using cgMLST data demonstrated likely clonal expansion of isolates belonging to ST-354 and ST-464 complexes, between which the position of the chromosomally encoded tetO alleles differed. Genes conferring aminoglycoside resistance were identified in less than 1.0% of isolates. This methodology successfully identified ciprofloxacin and tetracycline resistance-associated genotypes among Campylobacter isolates and confirmed high levels of aminoglycoside susceptibility. Identification of clonal expansion events in multiple clonal complexes provided evidence of continued convergent evolution in antibiotic resistant clinical campylobacteriosis isolates.

EUCAST sub-committee report - can whole genome sequencing replace antimicrobial susceptibility testing?

Matthew J. Ellington[1], Frank M. Aarestrup[2], Rafael Canton[3], Michel Doumith[1], Oskar Ekelund[4], Christian Giske[5], Hajo Grundman[6], Henrik Hasman[7], Matthew T. G. Holden[8], Katie L. Hopkins[1], Jon Iredell[9], Gunnar Kahlmeter[4], Claudio U. Köser[10], Alasdair MacGowan[11], Dik Mevius[12], Mike Mulvey[13], Thierry Naas[14], Tim Peto[15], Jean-Marc Rolain[16], Ørjan Samuelsen[17] and Neil Woodford[1]

[1]Public Health England, London, United Kingdom; [2]Technical University of Denmark, Lyngby, Denmark; [3]Hospital Ramón y Cajal and Institute Ramón y Cajal for Health Research, Madrid, Spain; [4]Central Hospital, Växjö, Sweden; [5]Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden; [6]University of Freiburg, Freiburg, Germany; [7]Statens Serum Institute, Copenhagen, Denmark; [8]University of St Andrews, St Andrews, United Kingdom; [9]University of Sydney and Marie Bashir Institute, Sydney, Australia; [10]University of Cambridge, Cambridge, United Kingdom; [11]North Bristol NHS Trust, Bristol, United Kingdom; [12]Utrecht University, Utrecht, The Netherlands; [13]National Microbiology Laboratory, Winnipeg, Canada; [14]University Paris Sud, Le Kremlin-Bicêtre, France; [15]University of Oxford, Oxford, United Kingdom; [16]The Méditerranée Infection Foundation, Marseille, France; [17]University Hospital of North Norway, Tromsø, Norway

Background: WGS offers the potential to predict antimicrobial susceptibility from a single assay. The European Committee on Antimicrobial Susceptibility Testing (EUCAST) established a sub-committee to review and report on the current development status of WGS for bacterial AST. Findings and recommendations: Review of the literature indicated that the available evidence for using WGS as a tool to infer antimicrobial susceptibility accurately is currently either poor or non-existent and that the evidence and knowledge base require significant expansion. The primary comparators for assessing genotypic-phenotypic concordance from WGS data should be changed to epidemiological cut-off values (ECOFFs) in order to better differentiate wild-type from non-wild-type isolates (harbouring an acquired resistance). As a secondary comparator, the subcommittee recommend assessment against clinical breakpoints. This assessment will reveal if genetic predictions could also be used to guide clinical decision making. Bioinformatics tools for predicting antimicrobial susceptibility should perform to minimum standards and the results shown to be equivalent. Internationally agreed principles and quality control (QC) metrics will facilitate early harmonization of analytical approaches and interpretative criteria for WGS-based predictive AST. Only datasets that pass agreed QC metrics should be used in AST predictions, since resistance genes/mutations might be missed in sequences of poor quality. To facilitate measurement of comparative accuracies across different WGS processes a single public database of all known resistance genes/mutations should be established. Such a database should be regularly updated and strictly curated with minimum standards for the inclusion of new resistance loci, and control of resistance gene nomenclature. For most bacterial species and in most countries the current high cost and limited speed of inferring antibiotic susceptibility from WGS data remain prohibitive to wide adoption in clinical laboratories. Another limitation is the dependency of prior culture since analysis directly on specimens remains too challenging. Conclusions: Currently there is insufficient evidence to support the use of WGS-inferred AST to guide clinical decision making. WGS-AST should be a funding priority if it is to become a rival to phenotypic AST. We anticipate that our report will be updated regularly as WGS becomes more widely applied and the available evidence increases.

Session: Pathogen discovery

KEY NOTE LECTURE 6

Unbiased NGS for clinical infectious disease diagnostics

Adam Grundhoff

Heinrich Pette Institute, Hamburg, Germany

Conventional diagnostic methods such as PCR, serology, or microbial culture are generally cost-effective, robust and sensitive, but have a limited detection scope and usually require an accurate initial hypothesis as to the type of pathogen(s) that may be present in a clinical sample. With the widespread availability of NGS technologies, it is now possible to perform metagenomic shotgun sequencing of nucleic acids isolated from primary diagnostic material. However, while these techniques have the potential to permit broad-range detection of known, distantly related or even potentially novel infectious agents, there is still a profound lack of harmonization and validated standards for NGS-based microbial diagnostics. There also remains a fundamental need for analysis software that is streamlined towards the requirements of diagnostic laboratories, especially with regard to processing speed and ease of use for non-bioinformatically trained personnel. In this talk, I will present our work on the development of a clinically oriented pathogen detection platform and its application to the analysis of diagnostic samples of human and animal origin.

TAXONOMER, an integrated ultra-fast tool for metagenomic sequence analysis

Robert Schlaberg[1], Stephen Flygare[2],Hajime Matsuzaki[2], Heng Xie[2], Qing Li[2], Yuying Mei[2], Wan Xie[2], Keith Simmon[3], Chase Miller[1], Yi Qiao[1], EJ Osborne[1], Gabor Marth[1], Karen Eilbeck[1], Guochun Liao[2] and Mark Yandell[1]

[1]University of Utah, Salt Lake City, United States of America; [2]IDbyDNA Inc, Salt Lake City, United States of America

Metagenomics enables profiling of microbial communities in the environment and the human body at unprecedented depth and breadth. Enrichment independent metagenomics holds great promise for unbiased, hypothesis-free detection and molecular typing of pathogens directly from patient samples. However, the complexity of laboratory protocols, quality control requirements, and the need for rapid, accurate, and user-friendly data analysis solutions limit wider adoption in diagnostic laboratories.

We have developed Taxonomer, an integrated, ultra-fast tool for metagenomics sequence analysis. Taxonomer enables novel analysis modalities of unmatched complexity in an easy-to-use format including: (1) comprehensive panmicrobial detection and discovery, (2) host-response profiling, (3) interactive result visualization, and (4) access through an IOBIO-based web user interface, which eliminates the need for specialized hardware or expertise. Taxonomer computes at speeds comparable to the fastest existing tools, but supports both nucleotide and protein-based classification using a single integrated algorithmic framework, k-mer weighting-based approach, and classification algorithm. The result is greater tolerance for sequencing errors, greater sensitivity, more accurate microbial abundance estimates, and execution times exceeding the fastest published protein search tools.

A diagnostic version of Taxonomer, called TaxonomerDx, further reduces the barrier for diagnostic laboratories to adopt metagenomics based testing. TaxonomerDx is accessed through a web portal generating lists of detected pathogens based on raw RNA-seq and/or DNA-seq data, incorporates user-defined confidence thresholds for pathogen detection, provides interactive visualization of classification results, and creates diagnostic reports.

Here we present analyses of patient samples harboring viral, bacterial, and fungal pathogens, including pathogens of great public health concern, demonstrating that Taxonomer provides effective means for rapid pathogen detection for patient care and discovery in public health emergencies. As costs and turn-around times for high-throughput sequencing continue to fall, Taxonomer will enable a rapidly growing number of laboratories with access to sequencing instruments to analyze data in a meaningful timeframe without having to invest in computational infrastructure or bioinformatics expertise.

6380-year-old oral pathogens from hunter-gatherer remains found in Kiel bay: A metagenomic approach to ancient DNA

Sabin-Christin Kornell[1,2], Marcel Nutsua[1,2], Julian Susat[1], Femke Heinsen[1], Sönke Hartz[3], Julia Goldhammer[3], Johannes Müller[2,4] and Ben Krause-Kyora[1,2,5]

[1]Kiel University, Institute of Clinical Molecular Biology, Kiel, Germany; [2]Kiel University, Graduate School Human Development in Landscapes, Kiel, Germany; [3]Stiftung Schleswig-Holsteinische Landesmuseen, Schleswig, Germany; [4]Kiel University, Institute for Pre- and Protohistory, Kiel, Germany; [5]Max Planck Institute for the Science of Human History, Jena, Germany

Dental calculus is an excellent source for well-preserved ancient DNA – even in several thousand years old remains. Dental plaque is fossilized during lifetime into dental calculus, which can be defined as a calcified bacterial biofilm. This dental calculus preserves bacterial DNA, and allows investigation of the ancient oral microbiome. Furthermore, it is possible to understand the evolutionary history of the human diet as well as causes and evolution of specific oral diseases.

The Mesolithic site called Strande LA 163 was discovered in Kiel Bay in October 2011. At a depth of 6 metres, divers unexpectedly came upon the trunks of fallen oak trees. On closer inspection they found bone fragments and teeth with dental calculus, which were identified as dating to the late Mesolithic period (6380±35 BP).

DNA was extracted from dental calculus, next generation libraries were built and shotgun sequencing applied. Subsequent, bioinformatic analyses were carried out using MALT (MEGAN alignment tool) and MEGAN (metagenome analyzer) to search for pathogens. Interesting findings were further analysed using bowtie2, SAMtools, bedtools, DeDup (part of the EAGER pipeline) and statistical methods.

We were able to identify several oral pathogens, including different strains involved in the pathogenesis of periodontal disease. Amongst them, Tanerella forsynthia, Porphyromonas gingivalis and Treponema denticola, which are all members of the so called “red complex bacteria”. This is a group of bacteria associated with a severe form of periodontal disease.

These findings are very rare due to the fact that Mesolithic remains infrequently possess dental calculus. The formation of dental calculus depends of the dietary content and can be found regularly in early populations after the Neolithic revolution. The presence of dental calculus in our case could be seen as a first hint for an imbalance in the oral bacterial community, which has been verified with our metagenome analyses.

Session: Tracing hospital acquired pathogens

KEY NOTE LECTURE 7

Hospital real time prospective surveillance

Alexander Mellmann

University Hospital Muenster, Muenster, Germany

Driven by the rapid development of next generation sequencing (NGS) technologies, in the near future shotgun whole genome sequencing (WGS) of bacterial pathogens will be applied in clinical microbiology and infection control to unravel both the molecular epidemiology and further information such as the pathogenicity make-up and antibiotic resistance traits. Whereas the laboratory workflow to generate WGS data is nowadays already quite convenient and suitable for integration into a routine laboratory environment, data analysis and interpretation is still the major obstacle for broad usage of WGS. This presentation will demonstrate the applicability of WGS in a routine clinical laboratory. Over a 6-month period, all multidrug-resistant bacteria (MDRO) that were isolated from patients at the University Hospital Muenster, Germany, were subjected to WGS in a prospective manner to monitor their molecular epidemiology and to determine potential transmission events. In a second 6-month period, we investigated in addition transmission rates of MDRO after changing isolation procedures. Here, not only the molecular data will be shown but also the technical feasibility, the potential impact of patient care, and the cost-effectiveness of prospective WGS in a routine environment will be discussed.

Real-time whole-genome sequencing to investigate a vancomycin-resistant Enterococcus faecium outbreak at a tertiary care hospital

Mohamed M. H. Abdelbary[1], Claire Bertelli[2], Gilbert Greub[2], Delphine Hequet[1], Giorgio Zanetti[1], Laurence Senn[1] and Dominique Blanc[1]

[1]Lausanne University Hospital, Service of Hospital Preventive Medicine, Lausanne, Switzerland; [2]Lausanne University Hospital, Institute of Microbiology, Lausanne, Switzerland

Background: Between January and March 2015, an outbreak of vancomycin-resistant Enterococcus faecium (VREfm) occurred at the University Hospital of Lausanne. We performed whole-genome sequencing (WGS) to determine characteristic genetic signatures in the outbreak isolates for developing a specific multiplex PCR that could be implemented in the outbreak control measures. Furthermore, we investigated the relatedness, molecular features and the transmission events of the outbreak isolates.

Materials and Methods: In total, 39 VREfm isolates were retrieved from clinical samples and rectal swabs of 30 patients during the outbreak. WGS of 15 VREfm isolates was performed at the beginning of the outbreak (end of January 2015), while the remaining 24 isolates were sequenced retrospectively (July 2015). All isolates were sequenced using the Illumina MiSeq platform. Genome sequences were subjected to in-silico multi-locus sequence typing (MLST) and the identification of genetic determinants of antibiotic resistance. Comparative genome analysis of the assembled ordered draft genomes was performed using Mauve software. To infer the isolates' phylogeny, all paired-end reads were mapped to the most related reference genome and, subsequently, a maximum likelihood tree was constructed using single nucleotide polymorphisms (SNPs) from a core genome alignment. Estimations of the transmission tree, dates of infections, and imported cases were performed using the R-package outbreaker.

Results: The in-silico MLST analysis of the initial run revealed that fourteen of the sequenced isolates represented multilocus sequence type (ST) 17, while the remained isolate belonged to ST80. Using the comparative genomic approach, we have identified genetic loci responsible for polysaccharide biosynthesis that were specific to all ST17 outbreak isolates. These loci were used to develop a multiplex PCR to differentiate between outbreak and non-outbreak isolates, which preceded and followed the outbreak period. This specific PCR revealed that 15 further patients were part of the outbreak, and we were able to exclude ten patients from the outbreak.

The phylogenetic analysis demonstrated that outbreak isolates were very closely related with only 0-4 SNPs differences. The transmission tree revealed that three main clusters of cases that are likely linked (patient-to-patient transmission) and patient 01 (from November 2014) is likely to be the index case.

Gene-by-gene diversity exploration in large genome data sets

Sofia Hauck

University of Oxford, Oxford, United Kingdom

Background: With the rise of high throughput whole genome sequencing, thousands of genomes for bacterial species of public health interest have become available. While these genomes are often individually of poor quality, in combination they are a useful resource for exploring the diversity and selective pressures acting on a species at a gene-by-gene level.

Methods: Whole genomes from public archives such as the Sequence Read Archive (SRA) were added to the PubMLST database, following assembly via a Velvet optimiser pipeline. This database runs the BIGSdb (Bacterial Isolate Genome Sequence database) software, which allows for genome annotation based on a set of reference sequences that defines genes, and an iterative scanning and definition process for new alleles of those genes. Measures of diversity are then calculated from the length, count of unique alleles, and ratio of variable sites in each defined locus, using the mean for the genome as the point of comparison. An R Shiny interface allows for interactive plots to be made from the data in real-time on a browser for an easy-to-use exploration of the results.

Results: This method was first applied to a data set of 7681 Mycobacterium tuberculosis genomes, using the nearly four thousand genes in the H37Rv annotation as the reference. This identified both conserved and diverse genes, which were respectively often in the information pathway and with surface-bound products. Genes in the ESAT-6 family appeared to be genetically diverse but functionally conserved, suggesting perhaps two stages of evolution with different pressures. The distribution of synonymous mutations indicated that the vast majority of the M. tuberculosis genes are under purifying selection, while genes in Neisseria meningitidis showed an evenly distributed level of selection, and N. gonorrhoeae showed many positively selected genes in a background on mild purifying selection.

Conclusion: The availability of thousands of genomes and fully annotated reference genomes, together with population annotation software, allows for genes to be explored as "gene pools" that can be compared in order to reveal the evolutionary forces which act of them. This method allows to simple and rapid evaluation of genes using a large number of isolates to avoid sampling bias and increase power. This can be applied to choosing markers, identifying genes under strong evolutionary pressure or observing genomic trends in a population.

Genomic epidemiology of multiple Acinetobacter baumannii outbreaks in a veterinary intensive care unit

Aldert Zomer[1], Birgitta Duim[1], Els Broens[1], John W. Rossen[2], Joris Robben, Masja Leendertse[3], Jaap Wagenaar[1,4] and Joost Hordijk[1]

[1]Utrecht University, Department of Infectious Diseases and Immunology, Utrecht, The Netherlands; [2]University of Groningen, University Medical Center Groningen, Department of Medical Microbiology, Groningen, The Netherlands; [3]Alrijne Ziekenhuis, Department of Medical Microbiology, Leiderdorp, The Netherlands; [4]Central Veterinary Institute of Wageningen UR, Lelystad, The Netherlands

Introduction: Acinetobacter baumannii is an emerging pathogen primarily found in health-care associated settings. Due to its potential to acquire and control a large reservoir of antimicrobial resistance genes and its ability to survive for extended periods in the environment, outbreaks are a threat and challenge for health care settings. Molecular epidemiology of A. baumannii is difficult because existing typing methods often lack the resolution to identify transmission chains in repeated outbreaks in the same setting. Phenotypic tests on antibiotic susceptibility are less informative because of the ease A. baumanni acquires and regulates an antimicrobial resistance repertoire. As an alternative, whole genome sequencing (WGS) forms an attractive method to investigate whether repeated outbreaks are independent events or are linked. Here, we investigated two outbreaks of A. baumannii in the companion animal intensive care unit (caICU) at Utrecht University which infected 4 dogs in 2012 and 3 dogs in 2014.

Methods: Multi Locus Sequence Typing (MLST) was performed for both outbreaks. Additionally, all outbreak strains, three European Control strains (EC-I, -II and –III) and four environmental strains were sequenced on an Illumina Miseq platform. Genomes were assembled using SPAdes, horizontal gene transfer regions were detected using ClonalFrame, time-measured phylogeny was reconstructed using BEAST and results were visualized using iTOL.

Results:Multi Locus Sequence Typing (MLST) of the isolates revealed that all animal isolates from both outbreaks in the caICU belong to sequence type 2. Isolates from the outbreaks had similar resistance phenotypes, suggesting that the outbreak events are linked. Time-measured phylogenetic reconstruction on WGS data of the clinical isolates, reference isolates and public genome sequences revealed, however, that the outbreaks are two independent events and that isolates from both outbreaks diverged at least 20 years before.

Conclusion: Our study shows the benefit of WGS in outbreak management in a clinical setting for pathogens for which traditional molecular typing methods lack resolution.

Session: Comparative microbial genomics tools

Hierarchical clustering of core genome MLST data of S. aureus for rapid assessment of the genetic relatedness

Bruno Pichon, Michel Doumtih, Neil Woodford and Angel Kearns

Public Health England, London, United Kingdom

SNP analysis based on whole genome sequence data has been proposed as an alternative method for inferring the phylogenetic relatedness of staphylococci associated with hospital and community outbreaks. However, the results are dependent on the reference genome used, the bioinformatics pipeline and user expertise in phylogenetic interpretation.We have explored an alternative approach based on hierarchical clustering of core genome (CG) MLST data.

Core genes (n=1334) in 54 complete genomes (53 S. aureus and 1 S. argenteus), available from Genbank, were identified by a BLAST search. Allelic profiles were assigned using a unique identifier - a CG type (CGT). CG clustering of allelic profiles was performed based on hierarchical clustering of pairwise matrix of locus differences (LD) to reconcile clusters on 10 successive thresholds of LD (3, 10, 25, 50, 75, 100, 150, 200 and 300). CGTs and associated clusters were used to populate a reference database.

Newly sequenced genomes (n=317) were selected from 37 MLST lineages to assess the typeability of the method. Among them, 151 originated from 62 independent reports of suspected clusters of cases of EMRSA-15 (ST22 MRSA-IV) occurring in the UK over the period 2010-2014.

A total of 275 distinct CGTs were identified among the test genomes. CG clustering separated them into 27 major clusters, each composed of genomes clustering at the 300 LD threshold, and corroborated with MLST lineages. High genomic variability was observed among the EMRSA-15 isolates with 130 CGTs identified. CG clustering detected 41 outbreaks (previously confirmed by standard typing and epidemiological investigations) composed of samples having <3 LD (34 outbreaks) or <10 LD (7 outbreaks). Ten suspected outbreaks were classified as probable wherein isolates were clustered at 25 LD threshold. Eleven incidents were judged to be pseudo-outbreaks based on CG clustering which corroborated with standard strain characterisation (spa typing and PFGE) and epidemiological data. Interestingly, CG clustering led to the identification of previously undetected possible transmission on the same hospital or between geographically adjacent hospitals.

CG clustering is a promising approach for micro-epidemiological investigations. The method is rapid, stable and independent of the need for a reference genome. Moreover, it affords a more standardised approach for nomenclature and exchanging comparative genomic data between centres and countries.

Common languages in genomic epidemiology: from ontologies to algorithms

João André Carriço and Mário Ramirez

University of Lisboa, Lisboa, Portugal

The advent and recent evolution of High-Throughput Sequencing (HTS) has enabled Genomic Epidemiology studies to be performed by multiple research groups all over the world. The analysis approach used to compare draft genomes produced by HTS is maturing to the point that they can be easily used by non-specialists and without the need of expensive sequencing or computational infrastructures. However, the ability to effectively compare results from disparate groups and to do incremental studies without having to repeat the analysis pipelines every time a new sample is added requires standards. These common languages can be defined in three tiers.

The first tier captures the fact that each data analysis pipeline involves a plethora of algorithms/software, each with different parameterizations, and versions, which can impact /alter results.

The second tier are the ontologies required to produce machine actionable and exchangeable genomic analysis results and of auxiliary metadata. This metadata can range from the details of the sequencing procedure or analysis but also on the epidemiological metadata required to investigate outbreaks or to define microbial lineages. Several Ontologies are currently being developed such as GenEpiO, TypOn, NGSOnto to deal with each of this needs and their integration will cover all these steps.

The third tier arises from the amount of data being handled and is related to the creation of application programming interfaces (APIs) allowing operations between the existing databases, in order to enable effective communication between analysis software and distributed databases using the developed ontologies or controlled vocabularies.

In the present talk we will present the state-of-the-art in each of these tiers and discuss what the current needs for harmonization of ongoing efforts are in order to have effective multi-level surveillance of microbial infections and outbreak detection.

The bacterial isolate genome sequence database (BIGSdb) genomics platform

Keith Jolley

University of Oxford, Oxford, United Kingdom

Bacterial Isolate Genome Sequence Database (BIGSdb) is a web-based genomics platform designed to store and analyse sequence data for bacterial isolates [1]. BIGSdb extends the principle of MLST to genomic data, and can group any number of loci in to schemes for typing or for population annotation. The software powers the PubMLST databases that host conventional MLST and core genome MLST (cgMLST) schemes for a wide range of bacteria, as well as ribosomal MLST (rMLST) for Bacterial domain-wide universal characterization [2]. The system is highly scalable with the rMLST database alone now hosting over 150,000 bacterial genomes. Identification and tagging of alleles can be performed via the web interface, or more routinely, using command line tools that can utilize the power of multiple core servers (>1000 genomes can be tagged at 2000+ cgMLST loci in 1 hour on a single server).

Comparative genomics can be performed using the built-in Genome Comparator tool. This generates a distance matrix and graphical outputs based on pairwise allelic differences of defined cgMLST schemes or using loci defined in a reference genome, with further outputs including details of locus variability.

BIGSdb includes a comprehensive RESTful API allowing all hosted data to be made available to third party tools. The API additionally facilitates data submission of alleles, MLST profiles and genome assemblies, automatically notifying appropriate curators as required. Access to protected resources can be made available using OAuth authentication from registered third party clients.

The software is published under an open source license and is freely available (https://github.com/kjolley).

References:

[1] Jolley KA & Maiden MC BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595 (2010).

[2] Jolley KA et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158:1005-15 (2012).

SeqSphere+ software for prospective bacterial genomic surveillance and resistome analysis

Jörg Rothgänger

Ridom GmbH, Muenster, Germany

SeqSphere+ was introduced in the year 2013 (Nat Biotechnol. 31: 294, 2013) and supports genome-wide allele and single nucleotide variant (SNV) calling from whole genome sequence (WGS) data either on core genome and/or accessory genome level. However, the recommended (initial) analysis is a core genome MLST (cgMLST) allele typing as a global and uniform nomenclature service is maintained to ensure for a 'molecular typing Esperanto'. A number of cgMLST schemes using the software have been published recently; e.g., for M. tuberculosis (JCM 52: 2479, 2014) or L. monocytogenes (JCM 53: 2869, 2015). These schemes are available for download within the software. In addition, users can define with the included cgMLST Target Definer on the fly own 'ad hoc' schemes. Furthermore, the software supports setup of resistome and virulome schemes. Place, time, 'person', and type dimensions can be visualized with built-in geographic information system (GIS), epi-curve, epi-data coloring, and phylogenetic tree (among others the minimum spanning tree algorithm is supported) functionality. The software can generate sample-reports for senders with a summary of the analytical results, QC data, and extensive documentation of the analytical procedure.

SeqSphere+ is designed for distributed work-groups and requires no scripting skills. It allows automatic processing and analyzing of next generation sequence (NGS) and Sanger data for prospective bacterial genomic surveillance. De novo assembly or reference mapping of NGS read data is achieved with the incorporated Velvet or BWA tools, respectively. Defining and starting a pipeline to down-sample, assemble, and analyze data processes NGS data fully automated, e.g., by fetching the raw reads from a benchtop-sequencer as soon as data are generated. For speeding- and scaling-up the analysis simply an additional computer can be added for processing data in parallel. Experiment and epidemiologic meta-data are stored in an integrated searchable SQL database together with the DNA data. New sequence entries can be compared against stored data and automatic cluster alerts of possible outbreaks can be triggered. Meta- and sequence data can be (semi)-automated submitted to the EBI ENA archive. An audit trail of user actions is maintained. SeqSphere+ is commercially available from Ridom GmbH (Muenster, Germany) for Windows and Linux operation systems. A fully functional trial version can be requested at http://www.ridom.de/seqsphere/.

Whole genome sequence analysis using BioNumerics

Katrien de Bruyne

Applied Maths, Ghent, Belgium

With the increasing availability of next generation sequencing technology, whole genome sequencing (WGS) methods are increasingly being used for bacterial typing. Key challenge is the ability to rapidly extract the relevant information from large sequence data files.

BioNumerics provides two easy-to-use pipelines for high resolution WGS-based molecular typing: pan-genomic whole genome multilocus sequence typing (wgMLST) and whole genome single-nucleotide polymorphism analysis (wgSNP).

Whole genome multilocus sequence typing (wgMLST) is applied to WGS data with the purpose to detect clusters of highly related strains. wgMLST schemes have been created for multiple organisms, each time based on a limited set of reference sequences, from which all coding regions are extracted and used to create a set of discernible loci. Two allele calling approaches, an assembly-free and a BLAST-based allele calling algorithm are used in parallel to independently determine locus presence and allelic variants in a quality-controlled manner.

A cluster defined by wgMLST can then be further characterized by whole genome single-nucleotide polymorphism analysis (wgSNP). The wgSNP pipeline, tuned to reduce false positives while maximizing resolution, detects SNP variants by mapping the WGS reads to a reference sequence internal or external to the data set.

BioNumerics 7.6 offers a powerful platform where both wgMLST and wgSNP analysis can be performed at high-throughput rates. For both methods, all calculation-intensive data processing steps are performed on the BioNumerics Calculation Engine, deployed locally or in the cloud. Using BioNumerics, WGS analysis results can be validated against traditional data such as MLST or PFGE, rapidly providing a robust, portable and high resolution picture of molecular typing data. It has been demonstrated that wgMLST is suitable for the analysis of very large datasets, making it a suitable technique for outbreak surveillance. The added resolution of wgSNP against an internal reference sequence increases the confidence in the detected clusters.

Recently, functional genotyping tools have been developed that predict e.g. virulence, resistance, O/H type, pathotype, prophages and plasmids from E.coli WGS data. The functional genotyping results for each assembled genomic sequence are recorded as database information fields and as a detailed report, including a summary of results and detailed results for O/H serotype, virulence genotype, resistance genotype, shiga toxin subtype, pathotype, and plasmids present.

The combination of these complementary approaches, wgMLST, wgSNP and genotyping, on a virtually unlimited number of samples, managed by a single software platform that also stores historical data as well as metadata, opens many perspectives for food safety and public health monitoring programs.

Canada's IRIDA platform for genomic epidemiology

Gary Van Domselaar

University of Manitoba, Winnipeg, Canada

Public health, food safety, and clinical microbiology labs around the world are embracing whole genome sequencing and genomic epidemiological approaches to modernize their infectious disease research, surveillance, diagnostics, and outbreak investigation programs. This new technology requires powerful, yet easy-to-use, intuitive software in order to transition these new “big data” methods out of the lab and into the front lines of public health response. The Integrated Rapid Infectious Disease Analysis (IRIDA) project is a Canadian-led initiative to build an open source, end-to-end platform for infectious disease genomic epidemiology. In this talk, Dr. Van Domselaar will present an overview the storage, management, analysis, QA/QC, and reporting capabilities of the IRIDA platform.

Phylogenies from large samples of bacterial genomes

Fabian Klötzl and Bernhard Haubold

Max-Planck-Institute for Evolutionary Biology, Ploen, Germany

Sequencing has become so cheap that it is feasible to follow bacterial epidemics by sequencing the complete genome of each clinical isolate. For example, in 2014 a sample of 3085 Streptococcus pneumoniae genomes were collected during an outbreak. A typical first step in the analysis of such data is to summarize them as a phylogeny. Phylogenies are usually based on multiple sequence alignments. However, aligning thousands of bacterial genomes is challenging even with the best tools available. To address this problem, we have developed a method for estimating evolutionary distances between genomes based on approximate pairwise local alignment rather than full multiple alignment. The resulting program, andi, computes the pairwise distances between all 3085 S. pneumoniae genomes in 4:37 h on a 24-core computer using 9.2 GB RAM. I explain the algorithm underlying andi and demonstrate its application to simulated and real data. I also describe how we have recently added support values to the trees computed from andi-distances.

References:

- Haubold B, Klötzl F & Pfaffelhuber P. Fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* 31:1169-1175 (2015);
- Klötzl F & Haubold B. Support values for genome phylogenies. *Life* 6:11 (2016).

Session: Tracing community acquired pathogens

KEY NOTE LECTURE 8

MDR tuberculosis evolutionary history and transmission

Matthias Merker

Research Center Borstel, Borstel, Germany

Transmission of multidrug resistant (MDR, resistance to isoniazid and rifampicin) Mycobacterium tuberculosis complex (MTBC) strains represents a major risk factor for global tuberculosis (TB) infection control and jeopardize the successful implementation of new TB drugs and drug regimens especially in high MDR-TB incidence settings in Russia, Eastern Europe and Central Asia.

In the past two decades strains belonging to the MTBC Beijing lineage have been often associated with local MDR-TB outbreaks in different settings worldwide and, besides socio-economic factors, its specific genetic background is still under debate to contribute to an enhanced transmissibility, high rates of drug resistances and/or increased virulence. Thus the global extend of the MDR-TB epidemic with regard to pathogenomic aspects, resistance levels and the temporal and spatial distribution of resistant strains is key to understand the evolution of tuberculosis drug resistance and its impact on recent tuberculosis treatment programs.

In that regards we analyzed different datasets using classical genotyping and whole genome sequencing approaches. These include a global collection of nearly 5,000 clinical Beijing isolates, MDR-TB cohorts from different hotspot regions in Russia and Eastern Europe, and genome data from a Beijing outbreak clone comprising over 700 clinical isolates.

From an evolutionary point of view, we could define different ancestral and modern sub-groups among strains of the MTBC Beijing lineage that show global phylogeographical expansion signatures. Originated some 6,000 years ago in Eastern Asia important human interventions such as Industrialization, World War I and the introduction of antibiotics shaped the population structure of this pathogen. In particular highly clonal modern strain types that dominated in Eastern Europe acquired MDR phenotypes already in Soviet Union times and are continuously isolated from MDR-TB patients in different settings across Eurasia with increasing drug resistance and possibly also enhanced bacterial fitness levels.

The recent evolutionary history of these outbreak clones is associated with inappropriate and delayed TB diagnosis, absence of comprehensive drug susceptibility testing, and interrupted drug supply which is common to other settings globally and has to be considered for prospective implementation of new TB drug regimens and treatment programs.

KEY NOTE LECTURE 9

Temporal and spatial analysis of the Ebola virus outbreak

Stephan Günther

Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

The European Mobile Laboratory (EMLab) was among the first laboratories deployed to West Africa to perform diagnostics in the Ebola virus (EBOV) outbreak. Retrospective deep sequencing of patient samples processed by the EMLab unit deployed to the epicentre of the outbreak in Guinea reveals an epidemiological and evolutionary history of the epidemic from March 2014 to January 2015. Our results confirm that the EBOV from Guinea moved into Sierra Leone, most likely in April or early May. The viruses of the Guinea / Sierra Leone lineage mixed around June / July 2014. Viral sequences covering August, September and October 2014 indicate that this lineage evolved independently within Guinea. In 2015, we introduced the MinION technology in Guinea to generate genomic sequence data in real-time to assist the tracing of transmission chains. The technology is still in operation in Guinea. The use of sequence data in conjunction with epidemiological information has become an integral part of the outbreak response.

Trace and control the regional expansion of extended-spectrum B-Lactamase producing ST15 Klebsiella pneumoniae

John W. A. Rossen, Mariette Lokate, Ruud H. Deurenberg, Marga Tepper, Jan P. Arends, Erwin G. C. Raangs, Jerome Lo-Ten-Foe, Hajo Grundmann, Kai Zhou and Alexander W. Friedrich

University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Background: Klebsiella pneumoniae (KP) has emerged as an important nosocomial pathogen. We present our experience in using whole-genome sequencing (WGS) and further developing an outbreak clone-specific PCR for tracing and controlling a regional and inter-institutional outbreak caused by a ST15 CTX-M-15-KP clone between May 2012 and September 2013 in the north of Netherlands.

Materials & Methods: In total, 19 K. pneumoniae isolates obtained from patients and environment were included. Acquired antimicrobial resistance genes were identified by uploading assembled genomes to the Resfinder server. Other genes relating to resistance and virulence were detected using the mapping unit of CLC Genomics Workbench to map and/or by blasting assembled genomes to a pseudomolecule generated by concatenating a set of target genes. The core genome of outbreak isolates was blasted against our local K. pneumoniae genome database. Unique fragments were extracted and blasted against GenBank. Non-match fragments not related to mobile genetic elements were considered as the DNA signatures for the outbreak clone.

Results: A suspected epidemiological link between clinical K. pneumoniae isolates was supported by patient contact tracing and genomic phylogenetic analysis from May to November 2012. By May 2013, a patient treated in three institutions in two cities was involved in an expanding cluster caused by this high-risk clone (HiRiC) (local expansion, CTX-M-15 producing, and containing hypervirulence factors). A clone-specific multiplex PCR was developed for patient screening by which another patient was identified in September 2013. Genomic phylogenetic analysis resolved a close homology with isolates previously found in the USA, and split ST15 strains into two clades. Two serotypes (K24 and K60) were carried by the two clades, respectively. Environmental contamination and lack of consistent patient screening were identified as being responsible for the clone dissemination.

Conclusion: The investigation addresses the advantages of WGS in the early detection of HiRiC with a high propensity of nosocomial transmission and prolonged circulation in the regional patient population. It shows the added value of designing a clone-specific multiplex PCR for effective and adequate infection control management. Our study suggests the necessity for inter-institutional/regional collaboration for infection/outbreak management of K. pneumoniae HiRiCs.

Session: Future microbial genomics developments

Whole genome analysis of Campylobacter fetus indicates recent diversification and reveals inconsistency of phenotypic and genomic characteristics

Linda van der Graaf-van Bloois, Birgitta Duim, William G. Miller, Ken J. Forbes, Jaap A. Wagenaar and Aldert L. Zomer

Utrecht University, Utrecht, The Netherlands

Introduction: The pathogen Campylobacter fetus (C. fetus) can cause disease in both animals and humans. The species C. fetus is divided into three subspecies; C. fetus subsp. venerealis (Cfv), C. fetus subsp. fetus (Cff) and C. fetus subsp. testudinum. Subspecies identification of mammal-associated C. fetus strains is crucial in the control of Bovine Genital Campylobacteriosis (BGC), a syndrome associated with Cfv. The prescribed methods for subspecies identification of the Cff and Cfv isolates are: tolerance to 1% glycine and H2S production.

Methods: In this study, we performed a genome-wide SNP-based phylogenetic analysis of 41 C. fetus genomes and a BEAST analysis to estimate the divergence dates of Cff and Cfv strains. Additionally, we investigated whether the genomes contain specific SNPs or genes that could be associated with the phenotypes and different clinical features of the C. fetus strains.

Results: Phylogenetic reconstruction of the genome-wide single nucleotide polymorphisms (SNPs) within Cff and Cfv strains divided these strains into five different clades. BEAST analysis showed that the Cfv clade and a Cff clade evolved recently from a Cff ancestor under diversifying selection. We observed the deletion of a putative cysteine transporter in the Cfv strains, which are not able to produce H2S from L-cysteine in the H2S production test.

Conclusions: The partial deletion of a putative cysteine transporter is observed in all H2S production negative Cfv strains. Genome-wide SNP analysis divided the mammal-associated C. fetus strains into 5 different clades, which were not consistent with the phenotypic differentiation of the strains. This suggests the need for a closer evaluation of the current C. fetus subspecies differentiation, considering that the phenotypic differentiation is still applied in BGC control programs.

Mycobacterium tuberculosis resistance prediction from genome sequencing: comparison of automated software tools

Viola Schleusener, Silke Feuerriegel and Stefan Niemann

Research Center Borstel, Borstel, Germany

Multi-drug resistant or even extensively drug resistant Mycobacterium tuberculosis complex (MTBC) strains are a threat for global health and tuberculosis (TB) control. Next-generation-sequencing (NGS) approaches enable cost effective, rapid and comprehensive resistance prediction from genome data that can ensure timely diagnosis and individualized treatment regimens. Thus, NGS is providing effective diagnostics for successful treatment that also prevents further resistance development and MDR/XDR transmission. However, data analysis and interpretation of NGS data is complex and usually not possible by non-specialized users. Recently developed automated tools promise to overcome these limitations, however, their performance has not been systematically evaluated.

To address this question, we compared five different tools, CASTB, KvarQ, MykrobePredictor TB, PhyResSE and TBProfiler, with regard to accuracy in resistance prediction and phylogenetic classification. Raw sequencing data from a well characterized strain collection from Sierra Leone (n=92) was individually processed for each tool. Results were compared to Sanger sequencing results and a standardized marker-based TB-genotype classification.

While lineage classification of clinical MTBC strains was mainly concordant between all tools, major discrepancies occurred in the resistance prediction to all first-line TB drugs even for isoniazid and rifampicin. Both, technical and database related limitations were found to restrict the output of individual resistance prediction pipelines.

For clinical use a shared consensus catalogue of high confident resistance markers will be crucial to take advantage of sequencing-based tuberculosis resistance diagnostics in the treatment of drug resistant tuberculosis patients.

Poster presentations

Topic 1: Comparative microbial genomics tools

Whole Genome Sequencing of Leishmania donovani revealed presence of an endosymbiotic bacterium Parvibaculum lavamentivorans post kala azar dermal leishmaniasis strain

Anil Kumar Gupta, Saumya Srivastava, Amit Singh and Sarman Singh

All India Institute of Medical Sciences, New Delhi, India

Background: Post-kala-azar dermal leishmaniasis (PKDL) is a well-known dermal sequel of visceral leishmaniasis (VL) or kala-azar. PKDL manifests as maculopapular skin lesions several months or years after the cure of VL, but only in 5 to 10% of patients. This was not known whether this is due to changes in the parasite genome or some host factors. Our group has been of the opinion that PKDL is a result of in vivo generation of quasi-species of Leishmania donovani either as in vivo hybridization of various endemically circulating species within the host cells or due to superinfection with other organisms.

Methodology: Transmission Electron Microscopy (TEM) was used to search out the presence of micro-organism present inside the parasite. We found few structures resembling to bacteria. For confirmation weather it is bacteria or other species, Whole Genome Sequencing was performed using Hiseq 2500 illumina platform.

Results: In TEM, we observed some suspicious and unsure structure inside the parasite, we found that structure resembling to bacteria. This indicates the possibility of endosymbiotic infection of the Leishmania. Whole genome sequencing produced 36.3 million reads, corresponding to more than 110-fold sequencing depth. At least 138 tRNAs were identified in the genome which is significantly higher than other Leishmania isolates sequenced so far. It showed the clear difference between the VL and PKDL species. The BLAST analysis suggested that the 893 contigs of a heterotrophic bacterium Parvibaculum lavamentivorans DS-1 in the sequencing data, which covers 53.8% of the Parvibaculum lavamentivorans DS-1 genome. This indicates the confirmation of endosymbiotic infection of the Leishmania leading to PKDL manifestations.

Conclusion: Presence of Parvibaculum lavamentivorans inside the L. donovani, indicates a possibility of endosymbiotic infection/superinfection of the Leishmania leading to PKDL manifestations.

Time-critical NGS analyses of a cluster of CTX-M-15 Klebsiella pneumoniae isolates

Laura Becker, Stephan Fuchs, Jennifer Bender, Sebastian Haller, Tim Eckmanns, Torsten Semmler and Guido Werner

Robert Koch Institute, Berlin, Germany

These days, the spectacular progress in whole genome sequencing technologies enables molecular typing and epidemiological studies in maximum resolution.

From April to December 2015, 39 multidrug-resistant Klebsiella pneumoniae were isolated from colonized or infected patients who were admitted to a single neurorehabilitation centre in Germany. Despite intensified hygiene measures and ward closure, transmission to different wards could not be prevented. NGS-based analyses of the isolates revealed a cluster of 28 Klebsiella pneumoniae isolates (sequence type 15) carrying CTX-M-15. Interestingly, the development of colistin-resistance was observed in two isolates consecutively isolated from one patient under therapy. Comparative genome analyses revealed a causative SNP in mgrB which leads to a premature stop codon and, thus, a truncated gene product (Poirel et al. 2014).

Here, we provide insights into our bioinformatics pipelines integrating tools for reference-based, reference-free, and alignment-free approaches for genome reconstruction and/or phylogenetic analyses. For mapping-based strategies we developed RefFinder to assist reference selection. For this, random read subsets are built from original datasets and numbers of mapped reads are evaluated for comparison. After final mapping and variant calling, SNPs are filtered based on ambiguities, gaps, and distance to other SNPs considering replicon circularity. Our reference-free strategy is based on assemblies and the maximum common genome approach (von Mentzer et al. 2014). The above mentioned isolates were analyzed using both reference-based and -free analyses revealing concordant results. Based on the reconstructed genome and plasmid sequences multiplex PCR based on three different primer pairs have been designed to facilitate a fast classification of new isolates.

References:

- Poirel L et al. Emergence of the 16S rRNA methylase RmtG in an extended-spectrum-beta-lactamase-producing and colistin-resistant Klebsiella pneumoniae isolate in Chile. *Antimicrob Agents Chemother* 58:618-9 (2014);
- von Mentzer A et al. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. *Nat Genet* 46(12):1321-6 (2014).

Analysis of New Delhi metallo-β-lactamase 1 carrying Enterobacteriaceae isolates from Southern Germany

André Frühauf[1], Yvonne Pfeifer[1], Stephan Fuchs[1], Maria Martin[2], Dirk Luft[2], Niels Pfennigwerth[3], Miriam Korte-Berwanger[3], Martin Kaase[3] and Guido Werner[1]

[1]Robert Koch Institute, Wernigerode, German; [2]Regionale Gesundheitsholding Heilbronn-Franken GmbH, Bad Friedrichshall, Germany; [3]Ruhr University Bochum, National Reference Laboratory for multidrug-resistant gram-negative bacteria, Bochum, Germany

Antibiotic resistance mediating genes can be spread among different species of Enterobacteriaceae by various, possible routes; detailed procedures of which few are so far poorly understood. Recent papers challenge the current model of “plasmid hospitalism” and rather favor mobile transposable elements as possible vehicles of dissemination. Of special concern are carbapenemase-producers, e.g. bacteria carrying the blaNDM-1 gene that encodes the New Delhi metallo-β-lactamase 1 (NDM-1), because these isolates are resistant to all β-lactams, and most often also to aminoglycosides and fluoroquinolones leaving only few treatment options. Here we report an ongoing prevalence of NDM-1 producing isolates of different species in a region in Southern Germany.

Since July 2015 NDM-1 positive, carbapenem-resistant isolates of the species Escherichia coli (11), Klebsiella pneumoniae (1), Citrobacter freundii (3), Morganella morganii (2) and Providencia stuartii (1) were identified from rectal swabs. Antimicrobial susceptibilities were determined by broth microdilution according to EUCAST. Genomic DNA from all 18 isolates was extracted and subjected to whole-genome sequencing (WGS) by using Illumina MiSeq V3 chemistry. De novo assembly of sequence reads was performed using A5.

All 18 isolates carried carbapenemase gene blaNDM-1, and the majority was additionally positive for extended-spectrum β-lactamase genes blaCTX-M-1or-15 (13/18) and β-lactamase gene blaTEM-1 (14/18). The isolates were resistant to ampicillin, cefotaxime, ceftazidime, meropenem, imipenem, ciprofloxacin, sulfamethoxazole/trimethoprim, gentamicin (15/18) and amikacin (13/18) but remained susceptible to colistin. Three E. coli sequence types (ST744, ST410, ST405) were obtained from several patients. The plasmid content varied among the isolates and involved different replicon types (IncN, IncA/C2, IncR, IncQ1, IncX1 and IncI1). The blaNDM-1-carrying plasmid contigs appear to be highly modular and differed in size and gene arrangement.

The presence of several NDM-1-producing Enterobacteriaceae suggests a horizontal gene transfer between the different species. First WGS data did not imply the spread by a single blaNDM-1-carrying plasmid but rather postulate transfer of a NDM-1 carrying genetic element. Our aim is to further monitor the emergence of NDM-1-producers and the spread of blaNDM-1 by comparing available and establishing new strategies to deduce plasmid-encoded information from WGS data.

Elaboration of a core genome MLST scheme for high-resolution typing of Enterococcus faecalis

Bernd Neumann[1], Jennifer Bender[1], Torsten Semmler[2], Lothar H. Wieler[2], Dag Harmsen[3] and Guido Werner[1]

[1]Robert Koch Institute, Wernigerode, Germany; [2]Robert Koch Institute, Berlin, Germany; [3]University of Muenster, Department for Periodontology, Muenster, Germany

Enterococci can act as opportunistic pathogens and are known as the third most common cause of nosocomial infections e.g. urinary tract infections, bloodstream infections and endocarditis. A high genetic variability enables enterococci widespread exchanges of genetic determinants involved in virulence traits and acquired antibiotic resistances. Several plasmids and mobile genetic elements such as transposons are jointly responsible for this genetic plasticity.

As a consequence it is important to monitor the prevalence and spread of the pathogens of interest. Whole genome sequencing (WGS) by utilizing state of the art next generation sequencing (NGS) approaches using the Illumina® technology will be used to analyse and characterise enterococci on a molecular level. The gathered global data from WGS approach shall be connected with the multi-locus sequence typing (MLST) approach. MLST has been established to survey the molecular epidemiology of various pathogen bacteria like E. faecalis and allows to register and to catalogue the genetic diversity for these bacterial pathogens. One major issue of MLST is that it bases on just seven genetic loci, a fact that may result in an average typing resolution [1]. Our approach intends to establish a core genome MLST (cgMLST) typing scheme for E. faecalis using the Ridom SeqSphere+ software, according to an existing cgMLST scheme for E. faecium [2]. This approach should combine the discriminatory power of MLST with the extensive genetic and genealogical data from WGS to create a high-resolution tool for genotyping.

For this approach we will use a collection of disposable E. faecalis sequences from open sources to elaborate a cgMLST scheme. Further we create a collection of various E. faecalis strains isolated while hospital outbreak events, from food, animals and colonization of healthy persons. The sequences of this collection will be used to evaluate and where appropriate to interpolate the new cgMLST scheme.

In addition, we intend to identify by genome-wide association studies aspects of host adaption and host association, recombination between and among clonal lineages and linked to this the extent of transmission of mobile genetic elements including resistance determinants.

References:

[1] Jolley KA & Maiden MC Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. *Future Microbiol* 9(5):623-30 (2014);

[2] De Been M et al. Core Genome Multilocus Sequence Typing Scheme for High- Resolution Typing of Enterococcus faecium. *J. Clin. Microbiol* 53(12):3788-97 (2015).

Whole genome based genotyping of Mycobacterium tuberculosis complex isolates using a standardised and easily expandable genome-wide MLST approach

Thomas A. Kohl[1], Roland Diel[2], Dag Harmsen[3], Jörg Rothgänger[4], Karen Meywald Walter[5] and Stefan Niemann[1]

[1]Research Center Borstel, Borstel, Germany; [2]Schleswig-Holstein University Hospital, Institute for Epidemiology, Kiel, Germany; [3]University Hospital Muenster, Department of Periodontology, Muenster, Germany; [4]Ridom GmbH, Muenster, Germany, [5]Public Health Department Hamburg-Central, Hamburg, Germany

In recent years, the success of dropping total case numbers of tuberculosis has become challenged by the increasing incidence of multiple (MDR) or extensively resistant (XDR) TB. As there is virtually no environmental reservoir of the MTBC bacteria, the disease can be controlled by public health interventions. This necessitates reliable genotyping of bacterial isolates for monitoring treatment success, local outbreak detection, and regional surveillance. The results of traditional typing methods such as spoligotyping or MIRU-VNTR can be easily expressed in a number format similar to a barcode, automatically grouping the isolates in a meaningful way by simply collecting those isolates with an identical barcode pattern in a clonal complex. Thereby, the results of newly analysed isolates can easily be combined with existing data and shared between laboratories.

Several studies have already shown that the use of whole genome sequencing (WGS) allows for a much higher resolution and simultaneous resistance phenotype prediction. However, at present, the key limiting factors for widespread use of WGS genotyping are non-standardised analysis pipelines and the problems inherent in the commonly used SNP based analysis for data exchange between laboratories in an easily extendable classification scheme. In fact, with SNP based pipelines, the addition of new isolates into an existing study usually requires extensive recalculation.

One solution is the use of a multi locus sequence type (MLST) scheme encompassing the entire core genome set of genes (cgMLST). This strategy has been successfully employed for several bacterial species. Previously, we demonstrated the usefulness of a core genome MLST scheme for M. tuberculosis, with sufficient resolution to resolve individual outbreaks. In this study, we introduce and evaluate the performance of a cgMLST scheme for the whole MTBC using a reference collection reflecting the known diversity. From our results, the suggested MLST scheme consisting of 2891 genes was able to reliably classify isolates, with at least 97% of the genes reliably identified in all MTBC groups, and allowing for in depth resolution of individual outbreaks.

Application of NGS as method for typing of Salmonella enterica serovar Typhi

Eva Trost[1], Sandra Simon[1], Jennifer Bender[2], Stephan Fuchs[2], Wolfgang Rabsch[1], Erhard Tietze[1], Angelika Fruth[1] and Antje Flieger[1]

[1]National Reference Centre for Salmonella and other Bacterial Enterics and unit for Enteropathogenic Bacteria and Legionella, Berlin, Germany; [2]Division for Nosocomial Pathogens and Antibiotic Resistances, Robert Koch Institute, Wernigerode, Germany

Since 1938, isolates of Salmonella enterica serovar Typhi (S. Typhi) are sub classified by phage typing. This method is based on lysis of bacterial strains by specific determined bacteriophages. The set from Public Health London comprises 110 Vi-adapted phages to distinguish just as many lysotype. Nevertheless evaluation of results requires a high degree of expert knowledge to grant reproducibility.

In recent year's introduction of benchtop sequencers has made it possible to add next generation genome sequencing (NGS) as typing tool in clinical and public health laboratories. Here we describe the routine sequencing and analysis of S. Typhi isolates sent to the NRC in 2015/2016. Our aim is to establish NGS as a routine typing method. Therefore we evaluate SNP- and core genome (cgMLST)-based approaches and compare the outcome with classical typing methods.

DNA-Libraries were sequenced with the Illumina MiSeq sequencer (2x300 bp). Our in house pipeline was used for quality control and trimming. Mapping to the reference genome S. Typhi strain CT18 and variant calling was done using bowtie2, samtools and varscan. SNP-based phylogenetic analyses were done with Geneious (Biomatters Ltd., Auckland). SeqSphere+ (Ridom® GmbH) was applied for cgMLST analyses.

In case of S. Typhi isolates classified as ubII are lysed by S. Typhi specific Vi-phages, but not by ViII-typing phages, preventing a reliable statement about epidemiological contexts. In these cases NGS based methods showed a high degree of differentiation between of isolates of phage type ubII. The minimum-spanning tree (MST) resulting from cgMLST analysis is in most cases supported by the tree based on concatenated SNP positions. NGS data were also used to identify isolates that group in the multidrug-resistant lineage H58. The lineage is characterized by 44 SNPs and multiple antibiotic resistances including nalidixic acid (NAL) resistance. We were able to identify 43 of these SNPs in 13 isolates. Phenotypic antibiotic resistance tests showed that four of these isolates have resistance to multiple antibiotics including NAL.

From our analysis we conclude that NGS is a highly suitable technique to differentiate S. Typhi isolates. The discriminatory power of NGS exceeds that of classical typing methods. Both, SNP- and cgMLST-based analyses show congruent results and seem appropriate as routine methods for pathogen surveillance and outbreak detection.

Genomic characterization of Staphylococcus petrasii

Veronika Vrbovská[1], Petr Petráš[2], Pavel Švec[1], Kamila Kýrová[1], Ondrej Šedo[1], Vojtěch Kovařovic[1], Jiří Doškař[1] and Roman Pantůček[1]

[1]Masaryk University, Brno, Czech Republic; [2]National Institute of Public Health, Prague, Czech Republic

Coagulase-negative staphylococci are common human commensals that were increasingly recognized as potentially opportunistic and nosocomial pathogens. Staphylococcus petrasii is recently described coagulase-negative, oxidase-negative and novobiocin-susceptible species. Phylogenetically, it belongs to the Staphylococcus haemolyticus cluster group. It is considered to be a possible opportunistic pathogen, recovered mostly from young children and older patients. Four subspecies of S. petrasii have been described so far, S. petrasii subsp. petrasii, S. petrasii subsp. croceilyticus, S. petrasii subsp. jettensis and S. petrasii subs. pragenis.

Total of 66 S. petrasii strains were isolated from clinical material in the Czech Republic, USA and Belgium.The strains were identified using (GTG)5 repetitive PCR fingerprinting to the subspecies level. The genomes of type strains of all S. petrasii subspecies and strain S. petrasii subsp. jettensis 5404 were sequenced by Ion Torrent PGM next generation sequencing technology and analysed by comparative genomics tools. All strains were analysed by typing methods, such as macrorestriction profiling performed by pulse field gel electrophoresis (PFGE), MALDI-TOF mass spectrometry, plasmid DNA analysis and PCR detection of methicillin resistant genes.

Genome sequencing of the strains provided approximately 135 to 344 contigs in each strain with total length of genomes 2,5 – 2,6 Mbp. PFGE showed differences in macrorestriction patterns of individual strains. MALDI-TOF analysis clearly distinguished two subspecies. Plasmid analysis detected numerous different plasmids in most of the strains. Methicillin resistance was confirmed by the presence of mecA gene in 9 from 21 S. petrasii subsp. jettensis strains. New types of prophages and SCCmec cassettes were detected.

Occurrence of S. petrasii in human clinical material, the resistance to methicillin and possible opportunistic pathogenicity indicate importance of the study. The knowledge of genome sequences of the type strains is important for understanding the evolution, subspecies differentiation, as well as for genotypic and phenotypic characterization and identification of virulence factors of S. petrasii. Differences in genotypic characteristics between four subspecies can be used for designing new diagnostics tools.

Acknowledgement: This work was supported by Student Project Grant at Masaryk University (specific research, rector's programme) (MUNI/A/0967/2015).

Defining and implementing a core genome multi-locus sequence typing (cgMLST) scheme for Campylobacter jejuni and C. coli.

Alison J. Cody

University of Oxford, Oxford, United Kingdom

The increasing availability of large Campylobacter collections (thousands of isolates) that have been characterised by whole-genome sequencing (WGS) has created a need for structured approaches that can effectively summarise and analyse these extensive data. The hierarchical gene-by-gene comparison of genomes enables the study of isolates from 'domain to strain' and a widely accepted set of core loci common to Campylobacter isolates (cgMLST) is one means of facilitating such analyses. The focus of this sudy was to define a core genome (v1.0) for comparative genomic analyses of Campylobacter jejuni and C. coli, based on human clinical disease isolates, from the 1,643 loci defined by the re-annotation of the NCTC11168 genome. Loci appearing in 95% or more of 2,742 Oxfordshire human surveillance draft genomes were identified using the Genome Comparator (GC) function of BIGSdb. Potential paralogues were identified and excluded from the set. Sequence similarities among cgMLST loci were identified using GeneDB. Analyses of 2,742 genomes from 2,449 (89.3%) C. jejuni and 293 (10.7%) C. coli isolates identified 1,365 shared loci present in 95% of isolates. A total of 22 potential paralogous loci were identified and removed from the core list to provide a core genome multi-locus sequence typing (cgMLST) scheme of 1,343 loci.

The cgMLST scheme (v1.0) has been has been defined for comparative analyses of human Campylobacter disease isolates, by public health and research communities worldwide and is available at http://pubmlst.org/campylobacter/.

Topic 2: Pathogen discovery

The significance and health implications of antimicrobial susceptibility pattern of bacteria isolated from locally made snacks in Nigeria

Grace J. Ajayi, Adewale A. Ifesanwo, Olayinka O. Abass and Michael A. Ajayi

Ogun State Institute of Technology, Igbesa, Nigeria

Seven different traditional snacks samples were purchased from different locations in parts of Nigeria. Microbiological analysis was carried out on the snacks sample so as to identify bacteria present and to determine their antimicrobial susceptibility pattern. Bacteria isolated from snack samples include Stapphyloccucus aereus, Streptococcus faecalis, Baccillus spp, Micrococcus spp, Eschericha coli, Klebsiella pneumonia, Pseudomonas aeruginosa and Enterrobacter spp. The bacterial count on snacks sample shows that "Aadun and "Kulikuli" had higest bacterial load of (5.6 x 10.7 and 6.6x10.7 cfu/ml). Antimicrobial susceptibility pattern shows that Pseudomonas aeruginosa resisted almost all antimicrobial agents tested. This study reveals the presence of antimicrobial resistance of bacteria in traditionally prepared snacks and its public health significance.

NGS characterization of bovine mastitis Escherichia coli: new insights into diversity and putative virulence mechanisms

Florent Kempf[1], Cindy Slugocki[1], Shlomo E. Blum[2], Gabriel Leitner[2] and Pierre Germon[1]

[1]University of Tours, Nouzilly, France; [2]Kimron Veterinary Institute, Bet Dagan, Israël

Escherichia coli is one of the main causative agents of bovine mastitis, responsible for significant losses in dairy farms. Fine characterization of strains isolated from mastitis cases may help to better understand the E. coli pathogenicity mechanisms involved in this context. In the present study we have considered 5 currently available mastitis E. coli genome sequences. This included 2 novel strains isolated from cases of severe, per-acute mastitis and sequenced on an Illumina sequencing system. Data from 13 additional E. coli strains were also retrieved from public databases.

Genomic properties of mastitis E. coli were first investigated by phylogenetic analyses and whole genome comparison. This approach revealed a strong bias in their distribution among E. coli historical phylogenetic groups. Most of mastitis E. coli are indeed related to A and B1 strains.

Genetic relationships were next inferred using core or accessory genomic data. Interestingly, both approaches gave very similar clustering patterns. This may suggest that each phylogroup carry specific accessory genes, and may thus exhibit different phenotypic traits. In particular, the ability to trigger bovine mastitis may rely on various group-specific gene sets and mechanisms.

In line with this, comparative genomics revealed that very few genes were shared by all mastitis E. coli. This result argues for a reduced set of common virulence genes. This point was also investigated by a top-down search based on factors already described in other pathogenic E. coli strains. Again, the analysis did not uncover specific properties of mastitis isolates. Altogether, these results indicate that mastitis E. coli isolates are rather characterized by a lack of bona fide currently described virulence genes.

Sequencing of avian pathogenic Escherichia coli from a colibacillosis outbreak

Tarja Pohjanvirta, Troels Ronco,Sinikka Pelkonen and Karl Pedersen

Finnish Food Safety Authority Evira, Helsinki, Finnland

Avian pathogenic Escherichia coli (APEC) are responsible for colibacillosis in poultry. The most frequent form of avian colibacillosis is a systematic infection that starts in the respiratory tract and is later characterized by fibrinous inflammation in internal organs and septicaemia. APEC pathotype is poorly defined, and usually strains isolated from birds suffering from colibacillosis can be very heterogenous. Recently colibacillosis has caused high mortality in Finnish broiler production although predisposing factors like viral and mycoplasma infections do not occur in commercially raised broilers. We isolated E. coli strains during one month from 40 broiler farms from bone marrow of broilers that had died because of colibacillosis. Altogether 295 isolates from 40 farms were screened by PCR. Two main groups were detected. Group one strains occurred both in parent and production frams, group two isolates were seen only in production farms. Five isolates from different farms from each group were sequenced using Illumina MiSeq platform.The raw reads were de novo assembled using CLCbio's Genomics Workbench 7.5. The de novo assembled contigs were in silico MLST typed, serotyped, and virulence and antibiotic resistance genes were identified using Center for Genomic Epidemiology (DTU) programs. The relationship between the 10 isolates was inferred by investigating the genetic diversity in the core genome of the isolates against E. coli CFT073 using Burrows-Wheeler aligner and maximum parsimony tree was made using MEGA 6.0.6. All 10 isolates belonged to phylogroup D, MLST 117. The group one isolates were found to be of serotype O78:H4, carried the vat, iucD and iss genes and had no antibiotic resistance genes. The group two isolates were serotype O53:H4, carried the cva, iucD, irp2 and iss genes and 4/5 had the aminoglycoside encoding resistance genes. The phylogenetic analysis based on 58 710 SNPs showed that all 10 Finnish isolates cluster together compared to CFT073. Two subclades were seen according to seropathotype with an approximately 2900 SNPs difference between the subclades. In conclusion, phylogroup D, ST117 APEC caused severe colibacillosis problems in Finnish broiler farms. Serogroup O78 was isolated from birds in parent as well as production farms, and was likely transferred via eggs from parents to next generation. Serogroup O53 strains were isolated only from production farms. Antibiotic resistance was not common in Finnish APEC.

Topic 3: Tracing hospital acquired pathogens

Elucidation of transmission events of vancomycin resistant Enterococcus faecium isolates in intensive care units of a German tertiary-care hospital

Jennifer Bender, Carola Fleige, Torsten Semmler, Sandra Schneider, Petra Gastmeier and Guido Werner

Robert Koch Institute, Berlin, Germany

Background: According to the recent EARS-Net report, Germany is one of the few European countries where numbers of vancomycin resistant enterococci (VRE) isolated from bloodstream infections have constantly increased since 2002. Also, data from the German National Nosocomial Surveillance System report a significant increase of VRE frequency in intensive care units (ICUs) (Gastmeier et al., doi: 10.1093/jac/dku035). Further molecular typing of isolates received by the National Reference Centre for Staphylococci and Enterococci at the Robert Koch Institute revealed distinct sequence types highly prevalent in German hospitals. However, little is known about clonal relatedness and success of dominant lineages in ICUs with a persistent burden of VRE. Thus, analyzed a set of 176 VRE derived from a study on Surveillance of Antibiotic Usage and Bacterial Resistance (SARI-IQ) from 6 ICUs with respect to population structure and dynamics.

Methods: Enterococcus faecium isolates were collected by the Charité hospital between 2014 and 2015 and sent to the NRC for detailed molecular analyses. The entire set of strains was analyzed by the rep-PCR-based method DiversiLab® and whole genome sequencing (WGS) utilizing Illumina technology. WGS data were used to infer phylogenies by means of a mapping and SNP-based approach and compared to analyses based on core genome MLST data extracted by the commercially available software SeqSphere+ (Ridom, Muenster, Germany).

Results: Due to limited discriminatory power, initial MLST analyses proved to be insufficient to allocate specific isolates to entry or transmission events across ICUs. Conflicting results were obtained by analyzing the population structure derived from DiversiLab® data. As an example, several smaller (n<10) or larger (n>25) DL clusters combined obviously different VRE isolates demonstrating vanA and vanB genotypes, varying presence of virulence determinants and different antibiotic susceptibility profiles. In addition, some DL clusters contained isolates with up to 4 different MLST types. In contrast, WGS improved resolution in a way to enable to trace VRE transmission between wards.

Conclusion: WGS provides maximal discriminatory power to monitor entry and/or transmission events between hospital wards which is of utmost importance in order to reduce the possibility of VRE infections in high risk patients.

Topic 4: Metagenomics

Metagenomics analysis of archaeological human remains samples from Medieval Latvia as a primary screening tool for the identification of pathogen genomes

Alisa Kazarina[1], Guntis Gerhards[2], Elīna Pētersone-Gordina[2], Viktorija Leonova[2], Ilva Pole[1], Inta Jansone[1] and Renate Ranka[1]

[1]Latvian Biomedical Research and Study Centre, Riga, Latvia; [2]University of Latvia, Riga, Latvia

The study of ancient DNA (aDNA), recovered from archaeological human remains samples, has enormous potential in the study of origin and evolution of tuberculosis and its pathogens (Mycobacterium tuberculosis complex).

Usually M. tuberculosis infects human lungs, but under particular conditions it can travel via blood and lymphatic system and replicate near or within bones. The replication results in characteristic lesions, which make it possible to suggest the presence of M. tuberculosis in archaeologically found human bone remains.

Metagenomics is an approach that allows to retrieve bacterial genetic information directly from each environmental sample. Together with Next-Generation Sequencing (NGS) it allows to obtain a broad spectrum of data about the microbial communities inhabiting target archaeological samples.

The aim of this study was to evaluate the use of metagenomics sequencing as a primary screening tool for the identification of pathogen genomes in the study of archaeological human remains samples.

Throughout the study, archaeological bone samples were collected from Medieval Latvian cemeteries and were carefully examined in order to detect TB characteristic lesions. Those, which possessed the lesions, were selected for DNA isolation. Isolated DNA was used for the 16S amplicon library preparation. Full microbiomes of the samples were gathered by shotgun sequencing method using Ion Personal Genome Machine (PGM) following manufacturer's instructions.

A combination of computational methods was used in order to analyse the resulting data. Quality control was done using Galaxy public server. Alignment (via Silva reference database) and taxonomic representation was done using Mothur program. Further processing and representation was done using R programming language.

Eventually, the average number of 279 bacteria genus per archaeological sample was detected with Mycobacterium genus being present among others, providing us the basis for further investigation in the field of whole-genome sequencing in order to identify the bacteria species.

Our results show that metagenomics method powered by NGS is a useful tool for primary identification of pathogen genomes. Obtained data also reveals a need to consider burial environment microbiome as a contamination factor, in further investigations concerning ancient microbiomes and ancient pathogens.

Acknowledgments: The research was supported by the Latvian National Research Programme „BIOMEDICINE”.

Pathogen Search Pipeline – identification of pathogens in NGS data from clinical samples

Andreas Andrusch, Claudia Kohl, Jeanette Klenner, Piotr Wojciech Dabrowski and Andreas Nitsche

Robert Koch Institute, Berlin, Germany

Traditionally, most pathogens in clinical samples are diagnosed by PCR-based approaches which are fast, cost-effective and precise, but limited to the specific detection of organisms. In order to achieve an open view into the sample, the focus in diagnostics of clinical samples is increasingly turning to approaches utilizing Next Generation Sequencing (NGS) techniques. These techniques allow the generation of vast amounts of data which can contain genomic sequences from a variety of sources. From among this plethora of information, sequences originating from the host and organisms known not to be pathogenic to humans, ubiquitous or regular contaminants in NGS samples need to be filtered in order to focus on the identification of pathogens. The remaining sequences are then assigned to their organism of origin. Here, we present the Pathogen Search Pipeline (PAIP-line). The PAIP-line offers a streamlined workflow for a quick and sensitive comparison of reads to a large database of known sequences and their subsequent sensitive and specific taxonomic assignment. Our approach of combining Bowtie2, BLAST and our own algorithms enables a reliable identification of pathogenic organisms which is magnitudes faster than pure BLASTing of the same data while achieving a much higher level of certainty than that provided by any approach based on mapping alone. The analysis results are presented in a highly accessible manner which allows both a quick overview and deep analysis. This enables the researcher to glean optimal information from the data provided by an NGS experiment.

Topic 5: Tracing community acquired pathogens

Tracing multidrug-resistant tuberculosis in Africa – a genetic approach

Patrick Beckert[1,2], Andrea Rachow[3,4], Elisabeth Sanchez-Padilla[5], Solomon Ali[6], Ellen Bruske[7,8], Abraham Alabi[9], Bertrand Lell[9], Yassir A. Shuaib[1], Khalide Azam[10], Catia Bila[10], Saumu Pazia Juma[11,12], Daniel Mapamba[13], Thomas A. Kohl[1], Lynsey Isherwood[14], Ian Sanne[15], Nyanda Elias Ntinginya[13], Gibson Kibiki[11], Matthew Bates[16], Sofia Viegas[10], Matthias Frank[7,8], Martin P. Grobusch[7,17], Maryline Bonnet[5], Michael Hölscher [3,4] and Stefan Niemann[1,2]

[1]Research Center Borstel, Borstel, Germany; [2]German Center for Infection Research, Partner Site Hamburg-Borstel-Luebeck, Borstel, Germany; [3]University of Munich, Munich, Germany; [4]German Center for Infection Research, Partner Site Munich, Munich, Germany; [5]Epicentre, Paris, France; [6]Jimma University, Jimma, Ethiopia; [7]University of Tuebingen, Tuebingen, Germany; [8]German Center for Infection Research, Partner Site Tuebingen, Tuebingen, Germany; [9]Albert Schweitzer Hospital, Lambaréné, Gabon; [10]Ministry of Health*, Maputo, Mozambique; [11]Kilimanjaro Clinical Research Institute, Moshi, Tanzania; [12]Kilimanjaro Christian Medical Center, Moshi, Tanzania; [13]NIMR-Mbeya Medical Research Center, Mbeya, Tanzania; [14]National Health Laboratory Service, Johannesburg, South Africa; [15]University of the Witwatersrand, Johannesburg, South Africa; [16]University of Zambia – University College London Medical School, Lusaka, Zambia; [17]Center for Tropical Medicine and Travel Medicine, University of Amsterdam, Amsterdam, The Netherlands*

Tuberculosis (TB) challenges the health care systems worldwide; a special problem is the emergence and transmission of antibiotic-resistant Mycobacterium tuberculosis complex (MTBC) strains. In Eastern Europe multidrug-resistant (MDR) TB strains, strains resistant to at least isoniazid and rifampicin (RIF), are already spreading and causing outbreaks of MDR TB. In Africa with an estimated number of 2.7 million new TB cases annually, effective transmission of MDR TB is a frightening scenario; however detailed data on transmission dynamics are only sparsely available.

To address this question, we established the "Map-African TB" project that employs DNA fingerprinting (genotyping and whole genome sequencing, WGS) to analyse clinical MTBC strains across Africa. First, we focussed on Swaziland, a country with the highest TB and HIV incidence worldwide and increasing rates of MDR TB. Genotyping of 412 isolates indicated that the MDR TB epidemic is driven by two major outbreak strains. WGS of all drug-resistant (n=143) and a selection of fully susceptible isolates (n=130) confirmed the clonality of the outbreak strains and revealed that the strains of the largest outbreak (30% of all MDR strains) harbour a particular RIF resistance mediating mutation in rpoB (RpoB I491F), which is not detected by the commercial molecular resistance assays. To analyse cross-border transmission of such MDR strains, we established a database with genotyping patterns of more than 4500 isolates from 17 African countries. The analysis showed that 12% of the isolates are already MDR and MDR strains emerged in 13 countries. In addition to the outbreak in Swaziland, we could identify three larger MDR outbreaks in DR Congo, Sudan and Gabon.

In conclusion, our data demonstrate that MDR TB emerged several times in Africa and is spreading locally or to the neighbouring countries. In four countries larger MDT TB outbreaks have been detected. The MDR TB epidemic in Swaziland is mainly driven by transmission of highly resistant strains. High prevalence of RpoB I491F among circulating MDR strains reduces the sensitivity of Xpert

MTB/RIF for MDR detection, a finding with enormous public health implications for TB control. The emergence of several outbreaks of MDR TB harbour the potential risk for further spreading of those strains across the whole continent, highlighting a worrisome scenario of larger MDR TB epidemics in Africa comparable to the situation already present in Easte

Topic 6: Antimicrobial resistance genome to phenome

Tuberculosis in humans and animals in the eastern part of the Sudan

Yassir A. Shuaib[1,3,5], Stefan Niemann[1], Eltahir A. Khalil[2], Ulrich Schaible[1], Lothar H. Wieler [4,6], Mohammed A. Bakhiet[2], Abbashar O. Mohammed[4], Mohamed A. Abdalla[5] and Elvira Richter[1]

[1]Research Center Borstel, Borstel, German; [2]University of Khartoum, Khartoum, the Sudan; [3]Free University Berlin, Berlin, Germany; [4]Ministry of Health, Kassala, the Sudan; [5]Sudan University of Science and Technology, Khartoum North, the Sudan; [6]Robert Koch Institute, Berlin, Germany

Tuberculosis (TB) is a chronic bacterial disease of humans and animals and it is characterized by the progressive development of specific granulomatous tubercle lesions in affected tissues. In a six-month study, from June to November 2014, a total of 2,304 carcasses of cattle, camel, sheep, and goats slaughtered at East and West Gaash slaughterhouses, Kassala, were investigated during postmortem, in parallel, 101 sputum samples from TB suspected patients at Kassala and El-Gadarif Teaching Hospitals were collected in order to investigate tuberculosis in animals and humans. Only 0.1% carcasses were found with suspected TB lesions in the liver and lung and peritoneal cavity of two sheep and no tuberculous lesions were found in the carcasses of cattle, goats or camels. All samples, tissue lesions and sputum, were decontaminated by the NALC-NaOH method and cultured for mycobacterial growth at the NRZ for Mycobacteria, Research Center Borstel, Germany. Genotyping and molecular characterization of the grown strains were done by line probe assay (GenoType CM and MTBC) and 16S rDNA, rpoB gene, and ITS sequencing, spoligotyping, MIRU-VNTR typing and next generation sequencing (NGS). Culture of the specimens revealed growth of organisms from 81.6% of all samples. Mycobacterium tuberculosis (76.2%), M. intracellulare (14.2%), mixed infection with M. tuberculosis and M. intracellulare (6.0%) and mixed infection with M. tuberculosis and M. fortuitum and with M. intracellulare and unknown species (1.2%) were detected in the sputum samples and unknown species (1.2%) were detected in one of the animal tissue samples. From the 69 M. tuberculosis strains, 25 (36.2%) were showing either mono-drug-resistant or multi-drug-resistant or poly-drug-resistant but none was extensively drug-resistant. In conclusion, the prevalence of TB in animals was very low while in humans M. tuberculosis-Delhi/CAS lineage was responsible for most cases and there was an evidence of MDR transmission and acquisition.

Routine use of Whole Genome Sequencing approach for diagnosis of tuberculosis

Andrea Maurizio Cabibbe[1], Alberto Trovato[1], Emanuele Borroni[1], Enrico Tortoli[1], Maria Rosaria De Filippo[1], Tullia Simonetti[2], Anna Maria Bartolesi[2], Gian Maria Rossolini[2] and Daniela Maria Cirillo[1]

[1]IRCCS San Raffaele Scientific Institute, Milan, Italy; [2]Florence Careggi University Hospital, Florence, Italy

Rapid identification of M. tuberculosis complex (MTBC) and drug-resistance (DR) patterns are required in a routine laboratory to promptly start treatment of tuberculosis (TB) patients. The increasing use of Whole Genome Sequencing (WGS) approach potentially allows to overcome the limitations of time, feasibility and cost of the phenotypic standard testing, as well as the constrains of the WHO-endorsed molecular assays in terms of genomic targets analyzed, genotyping and information on specific nucleotide changes.

Aim of this study was to evaluate the feasibility to use WGS with a centralized approach to speed-up diagnosis of TB in a low-incidence Country. Time to diagnosis and accuracy of this technique were compared with standard testing routinely performed.

In this pilot project started on March 2016, we collected and processed by WGS all the early positive MGIT960 tubes from EBPU activity or referred by the TB Regional Reference Center of Tuscany. Two-mL aliquots of early positive MGIT cultures were processed at EBPU, after heat-inactivation performed on-site. DNA was isolated by using the Maxwell 16 MDx extraction system. Paired-end libraries were prepared using the Nextera XT DNA Sample Preparation kit, and sequenced on Illumina Miseq or HiSeq 2500 platforms (based on the first available run). Total variant calling was performed by using Phyresse web-tool.

DNA isolation step required 60'. Concentrations obtained were suitable for library preparation. Results were obtained in two working days, using benchtop sequencer. The percentage of reads mapped to reference genome was over 84% for 73% of samples, and between 20-75% for the rest. Mean reads coverage was 80x and GC content 64,8%. Main challenge was the presence of non-mycobacterial DNA contamination in variable amount. Lineage information was obtained for all cases, with advantage to provide early evidence for suspected TB outbreaks. Genomic variants associated to resistance to first-, second-line and new anti-TB drugs were examined, including antibiotics not routinely tested. Three strains were rapidly confirmed as Multi-DR, 3 isoniazid-R and 2 fluoroquinolone-mono R. Cost per strain was approximately €150 inclusive of staff, reagents and instruments, revealing its effectiveness.

WGS is a rapid, cost-effective technique that promises to integrate and replace the other tests in routine laboratories for an accurate diagnosis of DR-TB, although suitable nowadays for cultured samples only.

Topic 7: Tracing foodborne pathogens

Whole Genome Sequencing of isolates of Salmonella enterica subsp. enterica serovar 4,[5],12:i:– from a human outbreak reveals context-specific limitations of PFGE and MLVA

Erika Scaltriti, Marina Morganti, Luca Bolzoni, Norma Arrigoni, Gabriele Casadei and Stefano Pongolini

The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

More than one hundred human cases of salmonellosis by Salmonella enterica subsp. enterica serovar 4,[5],12:i:– were diagnosed in Piacenza Province (Northern Italy) from July to November 2013 and linked to pork-meat consumption. The epidemiological investigation indicated that processed pork meat from a local producer was the cause of the outbreak, while implication of the purchasing slaughterhouses and pig farms remained uncertain. PFGE typing revealed that all human and the majority of food isolates belonged to the STYMXB.0131 genotype, while isolates from a suspect farm belonged to a 96% similar genotype (STYMXB.0083). Multiple-Locus Variable number tandem repeat Analysis (MLVA) revealed that the majority of human and food isolates belonged to the outbreak MLVA type 3-13-9-NA-211, while other MLVA types were present among outbreak isolates showing that STTR5 and STTR6 loci rapidly changed, mainly in the late phase of the outbreak. As a matter of fact, PFGE and MLVA did not completely disclosed the epidemiological relationships within the outbreak, mostly because STYMXB.0131 is a very common type in the area, consequently PFGE was not sufficiently discriminatory to unequivocally identify outbreak isolates and MLVA showed many variant genotypes among isolates strongly suspected to belong to the outbreak. Therefore, to achieve accurate understanding of the infection chain and precise outbreak definition, a selection of human, food and farm isolates were whole genome sequenced (WGS) using an Illumina MiSeq sequencer with a 2x250 pair-end run after library preparation with Illumina Nextera XT. Reads were de novo assembled with MIRA 4.0, core single-nucleotide polymorphisms (SNPs) were extracted from genome assemblies by kSNP3 and used for Bayesian analysis with MRBAYES. Bayesian phylogenetic analysis confirmed the presence of an outbreak cluster and clearly excluded responsibility of an originally suspected farm and slaughterhouse (having outbreak-related PFGE and MLVA types). In-silico genomic MLVA performed on a selection of isolates with different MLVA types confirmed MLVA results obtained by classical capillary electrophoresis. In conclusion the WGS approach clarified the epidemiological links within the outbreak and solved responsibility issues left open by PFGE and MLVA. At the same time, WGS led to precise assignment of the relevant isolates to the outbreak clone overcoming the sensitivity and specificity limitations posed by PFGE and MLVA.

A high-throughput genome assembly pipeline to facilitate the development and validation of core genome MLST (cgMLST) schemes for Staphylococcus aureus, Campylobacter jejuni and Mycobacterium tuberculosis isolates

James E. Bray, Keith A. Jolley, Alison J. Cody, Sofia Hauck and Martin C. J. Maiden

University of Oxford, Oxford, United Kingdom

We describe the development of a high-throughput genome assembly pipeline that uses the VELVET assembly algorithm [1] coupled with a 64-core Dell PowerEdge R815 Server with 512 gigabytes of RAM. Whole genome DNA sequencing data is automatically downloaded from the EBI Sequence Read Archive (SRA) and assessed for the instrument platform, library layout, read lengths and file size. The number of processors, hard disk space and RAM requirements are automatically computed for each individual assembly. Jobs are submitted to a queuing system (SQS, sourceforge.net/projects/sqs) that allows multiple assembly jobs to run in parallel. We have implemented the VelvetOptimiser.pl wrapper script to sample all possible odd k-mer lengths at least 21 nucleotides long and to automatically calculate the expected sequence coverage.

All assemblies generated by the high-throughput pipeline are publicly available on the PubMLST website (pubmlst.org). To date, the pipeline has assembled draft genomes for 25,356 Staphylococcus aureus isolates (pubmlst.org/saureus), 7,291 Mycobacterium tuberculosis isolates (pubmlst.org/mycobacteria) and 4,590 Campylobacter jejuni/coli isolates (pubmlst.org/campylobacter). These draft genomes have been used for the development and validation of core genome MLST (cgMLST) schemes during the Patho-NGen-Trace project (www.patho-ngen-trace.eu).

References

- Zerbino DR & Birney E Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821-9. (2008).