Joint Cold Spring Harbor Laboratory/Wellcome Trust Conference

INFECTIOUS DISEASE GENOMICS & GLOBAL HEALTH

September 12–September 15, 2010

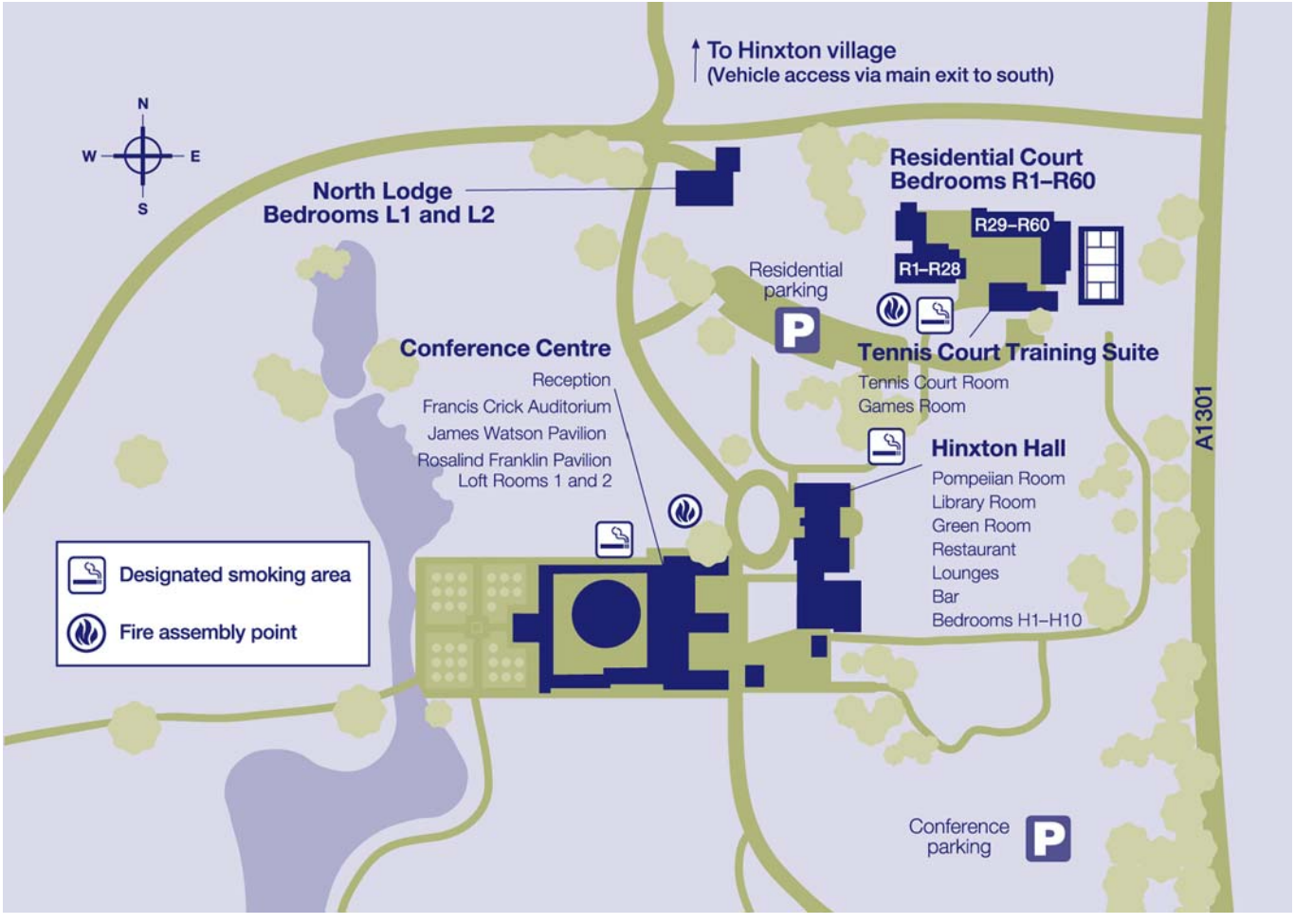
Joint Cold Spring Harbor Laboratory/Wellcome Trust Conference

INFECTIOUS DISEASE GENOMICS & GLOBAL HEALTH

September 12–September 15, 2010

Arranged by

Matthew Berriman, *Wellcome Trust Sanger Institute, UK*
Jane Carlton, *New York University, USA*
Julian Parkhill, *Wellcome Trust Sanger Institute, UK*
George Weinstock, *Washington University School of Medicine, USA*


To Hinxton village
(Vehicle access via main exit to south)
N
W
E
S
North Lodge
Bedrooms L1 and L2
Residential Court
Bedrooms R1–R60
R29–R60
R1–R28
Residential
parking
P
Conference Centre
Reception
Francis Crick Auditorium
James Watson Pavilion
Rosalind Franklin Pavilion
Loft Rooms 1 and 2
Tennis Court Training Suite
Tennis Court Room
Games Room
Hinxton Hall
Pompeiian Room
Library Room
Green Room
Restaurant
Lounges
Bar
Bedrooms H1–H10
A1301
Designated smoking area
Fire assembly point
Conference
parking
P

SCHEDULE AT A GLANCE

Sunday 12th September 2010

15.00-16.10	Session 1: Epidemiology and Public Health
16.10-16.30	Break
16.30-17.40	Session 1, continued
17.40-18.30	Wine and Cheese Party
18.30-19.30	Dinner
19.30-20.40	Session 2: Viruses I
20.40-21.00	Break
21.00-21.50	Session 2, continued

Monday 13th September 2010

07.30-09.00	Breakfast
09.00-10.40	Session 3: Viruses II
10.30-10:50	Morning Coffee
10:50-12:00	Session 3, continued
12.00-14.00	Lunch
14.00-15.30	Session 4: Population Genomics
15.30-15.50	Break
15.50-17.00	Session 4, continued
17.00-19.30	Poster Session I
19.30-21.00	Dinner

Tuesday 14th September 2010

07.30-09.00	Breakfast
09.00-10.30	Session 5: Bacteria
10.30-10.50	Morning Coffee
10.50-12.20	Session 5, continued
12.20-14.00	Lunch
14.00-15.30	Session 6: Parasites and Vectors
15.30-15.50	Break
15.50-17.00	Session 6, continued
17.30-18.30	Keynote Speaker: Dominic Kwiatkowski
18.30-19.30	Drinks Reception
19.30	Conference Banquet

General Information

Conference Badges
Please wear your name badge at all times to promote networking and to assist staff in identifying you.

If you have advised us of any dietary requirements, you will find a blue dot on your badge. Please make yourself known to the catering team and they will assist you with your meal request.

Scientific Session Protocol
Photography, audio or video recording of the scientific sessions is not permitted.

Internet Access
Wireless internet access is available throughout the campus. Please inquire at reception for a Wireless Connection token.

Conference Meals
Lunch and dinner will be served in the Hall Restaurant. Please refer to the conference programme in this book as times will vary based on the daily scientific presentations.

Presentations
If you are an invited speaker or your abstract has been selected for an oral presentation, please provide an electronic copy of your talk to a member of the AV team who will be based in the auditorium.

Poster Sessions
Posters will be displayed throughout the conference. Please post your materials in the Cloisters on arrival or during the afternoon break. The abstract page number indicates the assigned poster board number.

For Wellcome Trust Conference Centre Guests
Check in
If you are staying on site at the Wellcome Trust you may check into your room from 2.00pm. If you plan to arrive late at night you can check into your room as the conference centre reception is open 24 hours. Please note there will be no lunch or dinner facilities available outside of the conference timetable; however, there is a local public house (The Red Lion), serving both lunch and evening meals, located just 2 minutes walk from the campus in the village of Hinxton.

Breakfast
Your breakfast will be served in the Hall restaurant from 07.30 – 09.00

Telephone
If you are staying on-site and would like to use the telephone in your room, you will need to contact the Reception desk (Ext. 5000) to have your phone line activated - we will require your credit card number and expiry date to do so.

Departures
You must vacate your room by 10.00 on the day of your departure. Please ask at reception for assistance with luggage storage in the Conference Centre.

Taxis
Please find a list of local taxi numbers should you require transfers outside of the scheduled times:
Panther – 01223 715715
Mid Anglia - Tel: 01223 836000
Phil's Taxi Services - Tel: 01223 521918
A&M Carriages (Airport Specialist) - Tel: 01223 513703

Return Ground Transportation
Complimentary return transportation to Heathrow, Stansted Airport and the Cambridge Train Station and City Centre have been arranged for 9.00am on Wednesday, 15 September. Please note: a sign up sheet will be available at the registration desk. Places are limited so you are advised to book early.

Messages and Miscellaneous
All messages will be posted on the registration desk in the Conference Centre Foyer.

A number of toiletry and stationery items are available for purchase at the conference centre reception. Cards for our self-service laundry are also available.

If you have any queries or comments, please do not hesitate to contact a member of staff who will be pleased to help you.

Joint Cold Spring Harbor Laboratory/Wellcome Trust conferences at Hinxton are supported in part with funding courtesy of The Wellcome Trust.

Additional funding for this conference was provided by **Burroughs Wellcome Fund** and **New England Biolabs.**

These abstracts should not be cited in bibliographies. Material contained herein should be treated as personal communication and should be cited as such only with consent of the author.

Printed on 100% recycled paper.

PROGRAM

SUNDAY, September 12—3:00 PM

SESSION 1 EPIDEMIOLOGY AND PUBLIC HEALTH

Chairperson: **C. Arias,** University of Texas Medical School, Houston, USA
P. Bejon, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

Use of whole genome sequencing to track the in vivo evolution of antimicrobial resistance in enterococci
Cesar A. Arias.
Presenter affiliation: University of Texas Medical School at Houston, Houston, Texas. 1

Diversity and evolution of *Staphylococcus aureus* clonal complexes across space and time
Ruth Miller, Tanya Golubchik, Daniel Wilson, Helen Farr, Martin Llewelyn, Rory Bowden, Hanna Larner, John Paul, Nick Price, Nick Day, Sarah Walker, Tim Peto, Derrick Crook.
Presenter affiliation: University of Oxford, Oxford, United Kingdom. 2

High-resolution genomic typing of *Clostridium difficile* for hospital-based transmission studies
Rosalind M. Harding, Madeleine L. Cule, Tanya Golubchik, Ann S. Walker, Daniel Wilson, Camilla Ip, Hanna Larner, Teresa Street, Ruth Miller, Tim E. Peto, Kate E. Dingle, David Griffiths, Derrick W. Crook, Rory Bowden.
Presenter affiliation: Oxford University, Oxford, United Kingdom. 3

Philip Bejon.
Presenter affiliation: KEMRI Wellcome Trust Research Program, Kilifi, Kenya.

Mutations underlying the recent clonal expansion of a dominant subclone of the *Mycobacterium tuberculosis* Beijing genotype
Kristin Kremer, Anita Schuerch, Rob M. Warren, Roland J. Siezen, Noel H. Smith, Dick van Soolingen.
Presenter affiliation: RIVM, Bilthoven, the Netherlands. 4

***Escherichia coli* O157:H7 isolates related to the 2006 spinach-associated outbreak exhibit distinct transcriptional profiles and phenotypes—The spinach bag as a selective source**
Craig T. Parker, Jennifer L. Kyle, Michelle Q. Carter, Maria T. Brandl, Steven Huynh, Robert E. Mandrell.
Presenter affiliation: USDA, Agricultural Research Service, Albany, California. 5

SUNDAY, September 12—5:45 PM

Wine and Cheese Party

SUNDAY, September 12—7:30 PM

SESSION 2 VIRUSES I

Chairperson: **R. Heyderman,** Malawi-Liverpool Wellcome Trust, Blantyre, Malawi
P. Kellam, Wellcome Trust Sanger Institute, Hinxton, United Kingdom

Pneumococcal pathogenesis in the context of HIV—The biology and the genome
Robert S. Heyderman.
Presenter affiliation: Malawi-Liverpool/Wellcome Trust Clinical Research Program, Chichiri, Malawi. 6

Volant viruses—Metaviromic study of a West African bat population
Kate S. Baker, Mark Alston, Gordon Daly, Nick Bexfield, Paul Kellam, Jonathan L. Heeney, David T. Hayman, Andrew A. Cunningham, James L. Wood, Mario Caccamo, Pablo R. Murcia.
Presenter affiliation: University of Cambridge, Cambridge, United Kingdom. 7

Bioinformatics analysis over recently sequenced adenoviruses strains 53, 54 and intermediate 15-29-9 causing epidemic kerato-conjunctivitis outbreaks in Japan
Gabriel Gonzalez, Hidemi Watanabe, Kanako Koyanagi, Koki Aoki, Hisatoshi Kaneko.
Presenter affiliation: Hokkaido University Graduate School of Information Science and Technology, Sapporo Hokkaido, Japan. 8

MONDAY, September 13—9:00 AM

SESSION 3 VIRUSES II

Chairperson: **C. Simmons,** Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
L. Kramer, Wadsworth Center, SUNY Albany, New York, USA

MONDAY, September 13—2:00 PM

SESSION 4 POPULATION GENOMICS

Chairperson: **D. Conway,** Medical Research Council, The Gambia
P. Keim, Northern Arizona University, Flagstaff, USA

MONDAY, September 13—5:00 PM

POSTER SESSION

TUESDAY, September 14—9:00 AM

SESSION 5 BACTERIA

Chairperson: **N. Day,** Mahidol Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand
J. Vinetz, University of California, San Diego, USA

TUESDAY, September 14—2:00 PM

SESSION 6 PARASITES AND VECTORS

Chairperson: **T. Anderson,** Southwest Foundation for Biomedical Research, San Antonio, Texas, USA
A. Das, / National Institute of Malaria Research (ICMR), New Delhi, India

TUESDAY, September 14—5:30 PM

KEYNOTE SPEAKER

Dominic Kwiatkowski
Oxford University, Sanger Institute

TUESDAY, September 14

CONFERENCE BANQUET

Drinks 6:30 PM Dinner 7:30 PM

USE OF WHOLE GENOME SEQUENCING TO TRACK THE IN VIVO EVOLUTION OF ANTIMICROBIAL RESISTANCE IN ENTEROCOCCI

Cesar A Arias

University of Texas Medical School at Houston, Infectious Diseases, Houston, TX, 77030

Enterococci are microorganisms with a particular ability to adapt to the environment and acquire antibiotic resistance determinants. Of particular concern is the worldwide dissemination of vancomycin-resistant enterococci (VRE), which has resulted in a major decrease in therapeutic options to treat enterococcal infections. Apart from vancomycin, VRE are usually resistant to ampicillin and exhibit high-level of resistance to aminoglycosides, three of the traditionally most useful anti-enterococcal antibiotics. Currently, the only bactericidal antibiotic with in vitro and in vivo activity against VRE available in clinical practice is daptomycin (DAP). Although DAP does not have an FDA-approved indication for the treatment of VRE, it is often used to treat VRE infections due to the lack of other options. However, the development of resistance during therapy of VRE is a major drawback for this antibiotic and can severely jeopardize the only bactericidal compound currently available with potential activity against enterococci, including VRE. There is an important vacuum in the knowledge related to the factors and mechanisms involved in the development of DAP-R in VRE. Thus, using whole-genome sequencing, we aimed to understand the genetic changes associated with the development of DAP-resistance using isolates recovered from the same patient during DAP therapy. In May 2004, a woman was admitted to a hospital in Indiana with bacteremia with a vancomycin-resistant E. faecalis. After the failure of various antibiotics, she was re-admitted with fever and bacteremia with the same organism which was susceptible to DAP with a minimal inhibitory concentration (MIC) of 1 μg/ml. Daptomycin was started but after ca. two weeks, the patient returned bacteremic. The organism recovered from the blood had the same pulsed field gel electrophoresis (PFGE) pattern as the previous isolate and identical susceptibilities to antibiotics, except that the daptomycin MIC was now 16 μg/ml (isolate R712). We acquired the S613 and R712 isolates and performed comparative whole genome analysis of these isolates. A total of 53 non-synonymous changes in predicted proteins of R712 were found. None of the mutations in genes previously associated with non-susceptibility to daptomycin in other organisms (i.e. S. aureus) were identified in the VRE orthologs, suggesting that VRE utilizes a novel genetic strategy to develop resistance to DAP. These findings will be of paramount importance to understand the mechanism of in vivo DAP-resistance in VRE and, thus, design future strategies to prevent its emergence and “protect” this useful antibiotic for clinical use

DIVERSITY AND EVOLUTION OF *STAPHYLOCOCCUS AUREUS* CLONAL COMPLEXES ACROSS SPACE AND TIME

Ruth Miller[1], Tanya Golubchik[2], Daniel Wilson[1], Helen Farr[2], Martin Llewelyn[3], Rory Bowden[1,2], Hanna Larner[1], John Paul[3], Nick Price[3], Nick Day[1], Sarah Walker[1], Tim Peto[1], Derrick Crook[1]

[1]University of Oxford, Nuffield Department of Medicine, Oxford, , United Kingdom, [2]University of Oxford, Department of Statistics, Oxford, OX1 3TG, United Kingdom, [3]Brighton and Sussex University Hospital, , Brighton, , United Kingdom

Much of UK S. aureus, especially methicillin-resistant *S. aureus* (MRSA), occurs in two clonal complexes; CC22 and CC30. Evidence is accumulating that next-generation sequencing can be used to investigate population structure, even at fine scales of time and space, where isolates tend to be indistinguishable using lower-resolution techniques such as multi-locus sequence typing and spa typing. We have used the Illumina GAII platform to generate whole-genome data (>50× mean coverage) for CC22 and CC30 population samples comprising a total of >150 isolates separated in space (Oxford and Brighton) and time (1997-2007) to develop a snapshot of the population and molecular evolution of *S.aureus* in the UK.
We used a mapping-based approach (Maq software), with a high false positive rate for variant detection, coupled with a new goodness-of-fit filtering algorithm with high specificity and sensitivity to detect variants in unique regions of the genome.
In both CCs, methicillin-sensitive *S. aureus* (MSSA) was more diverse, and evidently ancestral to MRSA lineages. There is only evidence for one emergence of MRSA in each CC, in the sampled populations. CC30 is more diverse than CC22, and the CCs are each at different stages of their history, which is confirmed by epidemiological data. Clustering consistent with epidemiological reports implied a recent common origin of all isolates in recent MRSA outbreak(s) in Brighton. A lack of sharing of recently diverged clusters across space suggested that whereas there has been general spread of CCs across the UK, mixing is not fast enough to prevent the appearance of geographically specific types.
Whole genome sequencing captures the diversity of closely related strains in a CC, providing a deeper history than investigation of a single 'sequence type'. This allows us to build a picture of *S. aureus* evolution, including rates of spread and mixing between locations, as well as identifying emerging lineages that would not be distinguishable using lower resolution techniques.

HIGH-RESOLUTION GENOMIC TYPING OF CLOSTRIDIUM DIFFICILE FOR HOSPITAL-BASED TRANSMISSION STUDIES

Rosalind M Harding[1], Madeleine L Cule[2], Tanya Golubchik[2], Ann S Walker[3], Daniel Wilson[2], Camilla Ip[2], Hanna Larner[2], Teresa Street[2], Ruth Miller[3], Tim E Peto[3], Kate E Dingle[3], David Griffiths[3], Derrick W Crook[3], Rory Bowden[2]

[1]Oxford University, Zoology and Statistics, Oxford, OX1 3TG, United Kingdom, [2]Oxford University, Statistics, Oxford, OX1 3TG, United Kingdom, [3]Oxford University, NIHR Oxford Biomedical Research Centre, Oxford, OX3 9DU, United Kingdom

Ongoing advances in whole genome pathogen sequencing and bioinformatics have now reached a stage where we can envisage their application to analysis of hospital-based outbreaks. We apply these methodologies to study transmission of Clostridium difficile infection (CDI), the most important cause of hospital-acquired diarrhoea in the UK.

C. difficile is a 4.3Mb anaerobic bacterium that is present in the gut of up to 7% of healthy adults and 66% of infants, where it is kept in check by the normal intestinal flora. Some antibiotics can lead to the rapid multiplication of C. difficile, raising toxin levels that lead to diarrhoea and other gastrointestinal symptoms, which can be fatal. Increases in the rate and severity of nosocomial CDI between 2000 and 2008 have been attributed to the epidemic spread of a hypervirulent clone (ribotype 027).

The aim of the current work is to use deep sequencing of whole C. difficile genomes to distinguish ward-based transmission clusters within strains that have been defined by multiple locus sequence typing (MLST). In the current study 190 genomes have been sequenced, representing bacterial isolates sampled from symptomatic patients in the Oxford Radcliffe Hospital Trust (2007-2009). Illumina (Solexa) 51 bp paired-end reads have been mapped to reference genomes. The genomes include technical replicates at two levels (same DNA preparation or same isolate with different DNA library preparations), and re-sampling from the same patient. Our results confirm the potential for reference-based mapping to provide high resolution sub-typing. Furthermore, they provide evidence for patient-to-patient transmission events that can be distinguished from new infections of the same MLST strain within the same ward and month. We hope to use these results to apportion hospital cases of CDI to either, (i) local transmission chains or, (ii) independent introductions of CDI, which may be circulating in the community.

MUTATIONS UNDERLYING THE RECENT CLONAL EXPANSION OF A DOMINANT SUBCLONE OF THE *MYCOBACTERIUM TUBERCULOSIS* BEIJING GENOTYPE

Kristin Kremer[1], Anita Schuerch[1,3], Rob M Warren[2], Roland J Siezen[3], Noel H Smith[4], Dick van Soolingen[1]

[1]RIVM, Tuberculosis Reference Laboratory, Bilthoven, 3721 MA, the Netherlands, [2]Stellenbosch University, MRC Centre for Molecular and Cellular Biology, Cape Town, 7505, South Africa, [3]Radboud University Nijmegen Medical Centre, NCMLS, Centre for Molecular and Biomolecular Informatics, Nijmegen, 6500 HB , the Netherlands, [4]Veterinary Laboratories Agency, Bovine Tuberculosis, Surrey, KT15 3NB, United Kingdom

The Beijing genotype family is an epidemiologically important sub-group of *Mycobacterium tuberculosis*. It has been suggested that the high frequency of the Beijing isolates in some areas could be explained by a selective advantage. Beijing strains however do not show uniform phenotypes, probably due to different sublineages within the Beijing clade. The emerging "Typical Beijing" lineage may have the ability to circumvent BCG-induced immunity.
To investigate the phylogeny of the Beijing genotype of *M. tuberculosis*, the genome of six Beijing strains from three different countries was sequenced with next-generation sequencing. The phylogeny of these strains was established using single nucleotide polymorphisms (SNPs). The three Typical Beijing strains clustered very tightly in the phylogeny within other Beijing strains suggesting that Typical Beijing are a monophyletic lineage and resulted from recent diversification. Typing of 150 *M. tuberculosis* strains with a subset of the SNPs and comparison to a database of Beijing IS*6110* restriction fragment length polymorphism types revealed that 79.6% of all Beijing strains belong to the Typical Beijing subclone, which indicates clonal expansion. To reconstruct genomic changes that are characteristic for all Typical Beijing strains and were derived from their most recent common ancestor the presence of SNPs were assayed in other Beijing strains.
Typical Beijing strains are the result of a recent clonal expansion and we identified 53 SNPs that define the minimal set of polymorphisms for all Typical Beijing strains. Nonsynonymous polymorphisms in genes coding for the regulatory network were over-represented in this set of mutations. The only mutation in a 3R gene (DNA repair, recombination and replication) that was identified was a silent mutation. The possible consequences of these mutations which might have enabled Typical Beijing strains to circumvent the BCG-induced immunity and become a dominant clonal complex of *M. tuberculosis*, are discussed.

ESCHERICHIA COLI O157:H7 ISOLATES RELATED TO THE 2006 SPINACH-ASSOCIATED OUTBREAK EXHIBIT DISTINCT TRANSCRIPTIONAL PROFILES AND PHENOTYPES: THE SPINACH BAG AS A SELECTIVE SOURCE

Craig T Parker, Jennifer L Kyle, Michelle Q Carter, Maria T Brandl, Steven Huynh, Robert E Mandrell

USDA, Agricultural Research Service, Produce Safety & Microbiology, Albany, CA, 94710

Food-borne outbreaks of illness due to *Escherichia coli* O157:H7 linked to the consumption of ready-to-eat leafy vegetables are a mounting concern. The likely sources of pre-harvest contamination are soil and water that become contaminated via cattle or feral pigs in the proximity of the vegetable fields. In this study, we compared the transcriptional profiles of twelve isolates that possess the same PFGE pattern and are related temporally or geographically to the 2006 *E. coli* O157:H7 outbreak attributed to fresh spinach in the United States. These *E. coli* O157:H7 isolates included three clinical isolates, five isolates from separate bags of spinach, and single isolates from soil, water, a cow and a feral pig. From stationary phase cultures, we observed the decreased expression of many RpoS-regulated genes including *gadA*, *osmE*, *osmY*, and *katE*, in all three clinical isolates and two spinach isolates as compared to the soil, water, cow, feral pig, and the other three spinach bag isolates. The decreased expression of several RpoS-regulated genes correlated with decreased resistance to acid stress (pH 2.5), oxidative stress (15 mM H2O2), and osmotic stress (2.5 M NaCl) in these isolates. The transcriptional and phenotypic differences of the spinach bag isolates of *E. coli* O157:H7 suggest that certain variants within the bag retained characteristics of the pre-harvest isolates while other variants with altered expression and phenotype infected the human host.

PNEUMOCOCCAL PATHOGENESIS IN THE CONTEXT OF HIV - THE BIOLOGY AND THE GENOME

Robert S Heyderman

Malawi-Liverpool-Wellcome Trust Clinical Research Program, ., Chichiri, Blantyre 3, Malawi

HIV seroprevelence in many sub-Saharan African countries, exceeds 10% of the overall population rising to over 30% in pregnant women. Increased carriage and susceptibility to disease by mucosal bacteria such as Streptococcus pneumoniae is one of the earliest manifestations of HIV infection. Although anti-retroviral therapy (ART) has been aggressively rolled out in many countries in the region, the risk of invasive infection by mucosal bacteria even in those well established on ART remains higher than in those that are uninfected. We have demonstrated numerous antigen specific T cell and B cell defects in anti-pneumococcal immune memory in asymptomatic HIV infected adults. Here we will explore ways in which changes in strain diversity driven by this early immune dysregulation impacts on pneumococcal pathogenesis and therefore disease.

VOLANT VIRUSES: METAVIROMIC STUDY OF A WEST AFRICAN BAT POPULATION

Kate S Baker[1], Mark Alston[2], Gordon Daly[1], Nick Bexfield[1], Paul Kellam[3], Jonathan L Heeney[1], David T Hayman[1], Andrew A Cunningham[4], James L Wood[1], Mario Caccamo[2], Pablo R Murcia[1]

[1]University of Cambridge, Department of Veterinary Medicine, Cambridge, CB3 0ES, United Kingdom, [2]BBSRC, Genome Analysis Centre, Norwich, NR4 7UH, United Kingdom, [3]Wellcome Trust Sanger Institute, Virus Genomics Team, Hinxton, CB10 1SA, United Kingdom, [4]Zoological Society of London, Institute of Zoology, London, NW1 4RY, United Kingdom

Bats constitute a significant reservoir of pathogens that can spill-over and/or emerge in both animal and human populations. This is illustrated by past outbreaks of SARS and Ebola haemorrhagic fever in humans, as well as Hendra and Nipah virus diseases in humans, horses and pigs. Active surveillance in areas with a high risk of spill over to characterise the viral richness within bat populations is thus critical for determining the potential for the emergence of new and known viruses from these animals.
Multiple geographical regions have been predicted as 'hot-spots' for infectious disease emergence from wildlife, including Ghana (West Africa). Here, we study a population of fruit bats that roosts in large numbers in close association with people in Accra, the capital city of Ghana, where they are also a source of bush meat. We have previously shown that this fruit bat population has neutralising antibodies against both henipaviruses and a Type II Lyssavirus.
Next generation sequencing (NGS) technology has greatly enhanced sequencing capabilities and has proved a unique tool for detecting previously-uncharacterized viruses. We applied a NGS approach to detect viruses in the aforementioned Ghanaian fruit bat population and identified numerous novel viral sequences, including members of the Herpesviridae, Papillomaviridae, Picornaviridae and Poxviridae. We discuss the importance of these findings for our understanding of the process of viral emergence .

BIOINFORMATICS ANALYSIS OVER RECENTLY SEQUENCED ADENOVIRUSES STRAINS 53, 54 AND INTERMEDIATE 15-29-9 CAUSING EPIDEMIC KERATO-CONJUNCTIVITIS OUTBREAKS IN JAPAN

Gabriel Gonzalez[1], Hidemi Watanabe[1], Kanako Koyanagi[1], Koki Aoki[2], Hisatoshi Kaneko[3]

[1]Hokkaido University Graduate School of Information Science and Technology, Laboratory of Genome Sciences, Research groups of Bioinformatics,, Sapporo Hokkaido, 060-0814, Japan, [2]Hokkaido University Graduate School of Medicine, Department of Ophthalmology and Visual Sciences, Sapporo Hokkaido, 060-0814, Japan, [3]Fukushima Medical University School of Medicine, Department of Microbiology, Fukushima, 060-0814, Japan

Every year, there is an outbreak of a new EKC strain in Japan and several other cases worldwide. EKC stands for Epidemic Kerato Conjunctivitis, and it is produced by adenovirus affecting the conjunctive tissue in eyes resulting in an inflammation of the cornea and conjunctive. EKC is highly contagious and it has the tendency to occur in epidemics. From 54 known human adenovirus strains, only 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 14, 15, 19a, 22, 37, 53 and 54 are related with EKC, but how they evolve and how they are related is still unknown. This study's pursuit is obtaining more insights about how adenoviruses could evolve into new strains. With this purpose on mind, we analyzed recently sequenced strains of 53, 54 and an intermediate product between 15, 29 and 9; all of these isolated in Japan in the last six years. Our analyses included recombinant analyses looking for recombinant events signals, and next, using these signals, we created phylogenetic trees over different sections of aligned genomes based on suspected recombinant regions; this process allowed us to look for relationships between them and confirm or reject the possibility of recombination. As outcome, we found "hot spots" for recombinant events in genes coding penton, hexon and fiber, which are the main components of the virion's capsid, and also in the E3 region which mostly encode the proteins for disguising the infection from the immune system. Besides, we obtained evidence which supports an origin for these new strains as result of recombinant processes from previously identified strains, like 53 which evidence supports an origin from a recombinant process between 8, 22 and 37; also, 54 which seems to be originated from 8, 9, 19 and 22. All of these related with previous EKC outbreaks. For example, 8 was first reported and identified during the 50's. This work can be the base for stating new questions regarding results of these recombinant events in the infection mechanisms or contagion symptoms. All this information could help us to identify necessary conditions behind recombinant events in adenoviruses which are reported to obtain some kind of variability through such method.

WHOLE VIRUS GENOME SEQUENCING AND THE TRACKING OF H1N1PDM INTRODUCTIONS AND VIRUS PERSISTENCE IN THE UK

Gregory J Baillie[1], Monica Galiano[2], Steve Platt[2], Richard Myers[2], Rachael Chiam[1], Astrid Gall[1], Anne Palser[1], Simon Watson[1], Paul-Michael Agapow[2], Anthony Underwood[2], Andrew Rambaut[3], Jonathan Green[2], Rod Daniels[4], Oliver G Pybus[5], Maria Zambon[2], Paul Kellam[1]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, United Kingdom, [2]Health Protection Agency, Centre for Infections, London, NW9 5EQ, United Kingdom, [3]Ashworth Laboratories, Institute of Evolutionary Biology, Edinburgh, EH3 9JT, United Kingdom, [4]National Institute for Medical Research, Division of Virology, London, NW7 1AA, United Kingdom, [5]University of Oxford, Dept. of Zoology, Oxford, OX1 3PS, United Kingdom

Virus gene sequencing and phylogenetics can be used to study the epidemiological dynamics of rapidly-evolving viruses. With complete genome data, it becomes possible to identify and trace individual transmission chains of viruses such as influenza during the course of an epidemic. Using both capillary and next generation sequencing methods we sequenced 154 influenza H1N1v genomes from UK isolates from the first (128 isolates) and second (26 isolates) waves of the 2009 pandemic, and have used the sequences and their dates of isolation to infer the genetic epidemiology of the UK epidemic. We demonstrate that the UK epidemic is composed of many co-circulating lineages, of which at least 12 were exclusively or predominantly UK clusters. The estimated divergence times of two of the clusters pre-date the detection of H1N1pdm in the UK, suggesting H1N1pdm was already circulating in the UK before the first clinical case. Crucially, 3 clusters contain isolates from the second wave of infections in the UK, two of which represent chains of transmission that appear to have persisted within the UK between the first and second waves, demonstrating that whole genome analysis can track in fine detail the behaviour of individual influenza lineages during the course of a single epidemic or pandemic. Forensic approaches to evolution of circulating influenza lineages will assist the early detection of emerging antigenic variants

SENSITIVE POPULATION PROFILING AND GENOME ASSEMBLY OF HIV AND FLAVIVIRUSES USING ULTRA-DEEP SEQUENCING TECHNOLOGIES

Matthew R Henn[1], Niall J Lennon[1], Ruchi Newman[1], Patrick Charlebois[1], Aaron Berlin[1], Elizabeth Ryan[1], Christine Malboeuf[1], Monica Casali[1], Lisa Green[1], Sarah Young[1], Joshua Levin[1], Michael S Diamond[4], Laura D Kramer[5], Gregory D Ebel[6], Eva Harris[3], Todd M Allen[2], Bruce W Birren[1]

[1]Broad Institute, Genome Biology Program, Cambridge, MA, 02142, [2]Ragon Institute, Virology , Charlestown, MA, 02129, [3]University of California, School of Public Health, Berkeley, CA, 94720, [4]Washington University, School of Medicine, St. Louis, MO, 63110, [5]Wadsworth Center, Dept. of Infectious Disease, Albany, NY, 12201, [6]University of New Mexico, Health Sciences Center, Albuquerque, AZ, 87131

Viral diseases such as HIV/AIDS and Dengue have an enormous impact on human health worldwide. Despite this, application of new sequencing technologies to viral genomics has lagged. We are using genome sequence data to study how populations of single stranded RNA viruses, including HIV, DENV, WNV and HCV, evolve within infected individuals in response to host immune, therapeutic, and vaccine pressures. We have developed high-throughput sequencing, assembly and population profiling pipelines based on 454 and Illumina technology that are tuned to the specific needs viruses. These strategies can assemble full genome sequences and can quantify sequence diversity at each residue in the genome with unprecedented sensitivity.

We can detect rare variants to below 1% frequency, revolutionizing our ability to accurately assess the earliest events in viral evolution. We have demonstrated effective assessment of genome-wide diversity during acute HIV infection, enabling rapid and affordable identification of the earliest cellular immune responses to HIV. This has allowed us to detect earlier evolutionary events, demonstrating, for example, that HIV cytotoxic T-lymphocyte (CTL) escape can occur much faster than previously known. In addition, we have shown that the extent of intra-host diversity in Flaviviruses such as DENV and WNV is different between these closely related viruses with the latter exhibiting greater genetic diversity. The results presented here demonstrate the power of scalable, next generation sequencing-based methodologies as a genome-wide and unbiased global approach to profiling genomic diversity in intra-host populations of single stranded RNA viruses. This project was funded by the NIAID NIH, Dept. HHS (Contract No. HHSN272200900006C) and the Gates Foundation.

DENGUE IN VIETNAM : GENOMICS OF THE VIRUS - HOST INTERACTION

Cameron Simmons

Oxford University Clinical Research Unit , Hospital for Tropical Diseases, Ho Chi Minh City, District 5,, Vietnam

Dengue is a major public health problem in Asia and Latin America. The ecology of dengue viruses is constantly changing, with pronounced oscillations in both serotype and genotype prevalence in endemic settings. To explore this in detail, genome sequences of ~1000 DENV sampled between 2001 and 2008 coupled with phylogeographic techniques has allowed us to define the spatial and temporal dynamics of virus trafficking within and between communities, and between countries. Strikingly, a genotype replacement event within DENV-2 began in 2003, with an Asian 1 lineage virus replacing the resident American/Asian lineage and precipitating a major increase in disease incidence that was associated with higher viraemia levels in patients. The implications of this for vector control activities and vaccine implementation will be discussed.

In children, dengue shock syndrome (DSS) is the most common life-threatening complication. Predicting patients who may develop DSS may improve triage and treatment. Strikingly, the only differences in the transcriptional signatures of early DSS and uncomplicated dengue cases was the greater abundance of several neutrophil-associated transcripts in patients who progressed to DSS, a finding supported by higher plasma concentrations of several canonical proteins associated with neutrophil-degranulation (BPI, ELA2, DEF1A). Collectively, these data suggest an hitherto unrecognised association between neutrophil activation, pathogenesis and the development of DSS and point to future strategies for guiding prognosis.

OCCULT HEPATITIS C VIRUS INFECTION: THE PAKISTAN EXPERIENCE

Muhammad M Idrees[1], Amreek A Lal[2], Sadia S Butt[1]

[1]University of the Punjab, Molecular Virology, Lahore, AL, 53700, [2]Sidu Medical College, Pathology, Mingora Swat, AL, 6600, [3]University of the Punjab, Molecular Virology, Lahore, AL, 53700

BACKGROUND: Occult hepatitis C infection should be considered when diagnosing patients with a liver disease of unknown etiology.
AIM: The aim of the present study was to determine the presence of HCV RNA in the liver biopsies of patients with abnormal liver tests but without detectable serum HCV RNA and anti-HCV antibodies in sera.
METHOD: Liver biopsies and whole blood of total 31 patients who were negative for anti-HCV antibodies with elevated liver function tests were received at Division of Molecular Virology National Centre of Excellence in Molecular Biology, University of the Punjab Pakistan from January 2002 to June 2009 for the detection of HCV RNA. HCV RNA status of the subjects was tested by reverse-transcription polymerase chain reaction (RT-PCR) and quantified using SmartCycler II real-time PCR, in their liver biopsies, sera and peripheral blood mononuclear cells (PBMCs). HCV genotyping was carried out in HCV RNA positive samples using type-specific PCR.
RESULTS: HCV RNA was detected in liver-biopsy specimens from 23 (74.2%) of the total 31 patients negative for anti-HCV antibodies and undetectable serum HCV RNA. HCV RNA of both negative and positive polarity was found in the livers of 8 (25.8%) patients. Genotyping analysis showed that 65% patients were infected with HCV 3a, 17% with 3b, 13% with 1a and 4% patients were found with untypable genotype. In a multivariate logistic regression model, patients having previous history of sharing needles, presence of steatosis, elevated liver enzymes, history of previous surgeries, male sex, age above 30 years and with liver fibrosis stage F1 were significantly associated with the presence of occult HCV infection.
CONCLUSIONS: Patients with elevated liver enzymes and negative HCV antibodies and negative serum RNA may have intra-hepatic HCV RNA. The chance of occult HCV infection increases if the patient has previous history of sharing needles, elevated liver enzymes, history of previous surgeries, male sex and above 30 years of age.

POSITIVE SELECTION OF HIV HOST FACTORS AND THE EVOLUTION OF LENTIVIRUS GENES.

Katarzyna Bozek, Thomas Lengauer

Max Planck Institute for Informatics, Computational Biology and Applied Algorithmics, Saarbrücken, 66123, Germany

Phylogenetic studies have shown that HIV emerged in humans through at least eleven cross-species transmission events of simian immunodeficiency virus (SIV) from non-human African primates. SIV infection of African non-human primate host species (including sooty mangabeys, African green monkeys, mandrills, and several others) is non-pathogenic despite high levels of viremia. Different levels of pathogenicity of immunodeficiency viruses in their host species as well as the lack of adaptation to their non-natural species show how interspecies differences can impact viral infectivity and drive virus adaptation. Positive selection of rapidly evolving immune system genes reflects the evolution of the host defense in response to various infections and can be a marker of factors relevant for infection.

Our comparative and evolutionary analysis of 1439 HIV-interacting genes of four primate species and 175 lentivirus genomes points to specific host factors of high genetic variability that could account for differences in susceptibility to disease and indicate specific mechanisms of host defense and pathogen adaptation. We find that the largest amount of genetic change occurs in genes coding for cellular membrane proteins of the host as well as in the viral envelope genes suggesting cell entry and immune evasion as the primary evolutionary interface between host and pathogen. We additionally detect the innate immune response as a gene functional group harboring large differences among primates that could potentially account for the different levels of immune activation in the HIV/SIV primate infection. We find a significant correlation between the evolutionary rates of interacting host and viral proteins pointing to processes of the host-pathogen biology that are relatively conserved among species and to those undergoing accelerated genetic evolution.

Many of the proteins identified in our study are not SIV/HIV-specific but interact with a variety of pathogens (dengue, hepatitis, rhinovirus, tuberculosis, ebola, etc). Thus these results not only further the understanding of the interaction of primate hosts with lentiviruses but potentially also offer insights into other host-pathogen systems. Individual host factors pointed to by our analysis might merit further study as potential targets of antiretroviral therapies.

GENOMIC ANALYSIS OF MUTANT SPECTRUM, FITNESS, AND PHENOTYPE IN RNA VIRUS POPULATIONS

Adam S Lauring[1], Ashley Acevedo[2,3], Raul Andino[3]

[1]University of California, San Francisco, Medicine, San Francisco, CA, 94143-2280, [2]University of California, San Francisco, Tetrad Graduate Program, San Francisco, CA, 94143-2280, [3]University of California, San Francisco, Microbiology and Immunology, San Francisco, CA, 94143-2280

Studies of microbial pathogenesis and drug resistance have relied heavily on analyses of specific clones and their genetic determinants of virulence. It is often difficult to apply this deterministic approach to the study of RNA viruses, which by virtue of their very high mutation rates, exist as a complex mixture of mutants. Ensembles of mutants rather than individual clones maintain viral genetic information, and a growing body of evidence indicates that the structure and composition of viral populations determine their phenotype. Theoretical models further suggest that the location of populations within genetic sequence space can profoundly influence the evolvability of viral populations. We have directly tested this hypothesis using three different strains of poliovirus, a prototypical RNA virus with evolutionary dynamics similar to those of many emerging viruses. Compared to the wild type reference strain, the two variant polioviruses, Max and SD, encode 566 and 934 synonymous mutations, respectively, within the viral capsid. While all three viruses have identical GC content, codon usage, and amino acid consensus, they occupy distinct regions of sequence space and will give rise to genetically divergent populations. We show that wild type, Max, and SD replicate with similar kinetics and express nearly identical levels of protein – recognized correlates of viral fitness. Despite these similarities, direct competition assays indicate that SD is much less fit than wild type, and Max is slightly superior. Analyses of fitness variance and mutagen sensitivity suggest that SD is less fit because it gives rise to a population of mutants with lower average fitness than wild type. We are using next generation sequencing to define the mutant spectrum of all three viral populations and determine how sequence space influences viral phenotype and adaptability. We will also present pathogenesis data from a transgenic mouse model. Because the evolutionary dynamics of poliovirus closely resemble those of other RNA viruses, our results have profound implications for understanding the behavior of many emerging infectious agents.

UNDERSTANDING SELECTIVE PRESSURES ON ARBOVIRUSES USING EXPERIMENTAL IN VIVO PASSAGE OF FLAVIVIRUSES

Alex T Ciota[1], Dylan J Ehrbar[1], Yongqing Jia[1], Graham J Willsey[1], Greta V Jerzak[1], David J Young[1], Laura D Kramer[1,2]

[1]Wadsworth Center, New York State Dept Health, Zoonotic Diseases, Albany, NY, 12159, [2]School of Public Health, State University of New York, Biomedical Sciences, Albany, NY, 12222

West Nile virus (WNV) and St. Louis encephalitis virus (SLEV) are flaviviruses (Flaviviridae), which are predominantly maintained in an enzootic transmission cycle between mosquitoes and birds. SLEV was the major cause of epidemic flaviviral encephalitis in the U.S. prior to the 1999 introduction of WNV. Recently, the level of activity and geographic dispersal of SLEV have been significantly more limited than WNV, the former sporadic and focal, while WNV remains highly active and is now widespread throughout the Americas. The need for arboviruses to replicate in disparate hosts is thought to result in constraints on both evolution and host-specific adaptation. This predicts that arboviruses are generalists, with a cost of suboptimal adaptation to each individual host. If cycling is the cause of genetic stability observed in nature and arboviruses lack host specialization, than sequential passage should result in both the accumulation of mutations and specialized viruses better suited for replication in that host. Our previous in vitro studies suggest that WNV and SLEV may differ in their capacity for both genetic change and host specialization, and in the costs each accrues from specializing. In an attempt to clarify how these selective pressures contribute to epidemiological patterns of WNV and SLEV, we evaluated both genetic and phenotypic changes for WNV and SLEV following sequential passages in natural hosts, either Culex pipiens mosquitoes or chickens. Analysis of intrahost genetic diversity of SLEV during passage demonstrates that, similar to WNV, the size of the mutant spectrum is host-dependent and the capacity for genetic change is large. Despite this, a general lack of fixed consensus changes in either virus resulted from passage in either host, a result that contrasts with the idea that significant constraints on evolution exist. Passage of WNV in mosquitoes resulted in adaptation to mosquitoes in terms of both infectivity and replication kinetics, yet did not result in any apparent fitness cost in chickens. A lack of both genetic and phenotypic change was measured in WNV following 20 passages in chickens, suggesting further host specialization was not attainable. Similarly, no additional specialization to mosquitoes was attainable for SLEV following 20 passages in mosquitoes and, although a strain more infectious to chicks did emerge from chick passage of SLEV, no cost was measured for this strain in terms of mosquito infectivity. These results suggest that these viruses may be capable of residing at or near maximum fitness peaks in both hosts and that ultimately, coadaptation may be attainable for arboviruses with minimal cost in either host.

NOVEL HANTAVIRUS DETECTED IN YUNNAN RED-BACKED VOLE,EOTHENOMYS MILETUS

Yun-zhi Zhang, Zheng-li Shi

Chinese Academy of Sciences, Wuhan Institute of Virology, Wuhan, 430071, China

Hantaviruses cause two main human zoonoses, hemorrhagic fever with renal syndrome (HFRS) in Asia and Europe, and hantavirus pulmonary syndrome (HPS) in North and South America. Rodents are natural reservoirs of hantavirus. Long-term surveillance of HFRS is necessary for preventing the diseases transmission from animal to human. In this study, 330 rodents covering 12 species were captured in hantavirus endemic areas (Luxi county), Yunnan province, China, during autumn of 2009 to the spring of 2010, and detected for the infection of hantavirus. Among the samples, 23 out of 133 Eothenomys miletus, 3 out of 96 Rattus flavipectus and 2 out of 25 Rattus nitidus showed positive by immunofluorescence assay (IFA). However, the virus was detected only in 23 Eothenomys miletus. The sequences of small fragment (S), encoding the nucleocapsid protein, was obtained from virus-infected animal tissue. The sequence comparison of S fragment with other hantavirus indicated that the hantaviruses from Eothenomys miletus are more closer to Tula virus (88% in amino acid sequence). Moreover, the hantavirus sequences detected from positive samples displayed genetic diversity. To our knowledge, this is the first discovery of Tula virus in Eothenomys miletus.

HUMAN GENETIC FACTORS THAT MODULATE HIV TYPE 1 INFECTION IN RURAL POPULATIONS IN CAMEROON

Judith N Torimiro[1], Nathan D Wolfe[2], Donald S Burke[3], Jean K Carr[4], Michael Dean[5], Joseph Sodroski[6]

[1] University of Yaounde I , Faculty of Medicine and Biomedical Sciences, Yaounde, 03077, Cameroon, [2]Global Virus Forecasting Initiative, Epidemiology, San Francisco, CA, 94104, [3]Graduate School of Public Health, International Health, Bradford, PA, 16701, [4]Institute of Human Virology, Epidemiology, Baltimore, MD, 21201, [5]National Cancer Institute, Frederick, MD, 21702, [6]Dana-Farber Cancer Institute, Boston, MA, 02115

HIV infection in Cameroon is driven by several genetic variants. However, this old infection represents less than 3% of the rural populations in the southern rainforest region of Cameroon. HIV-1 gains entry into a competent cell through receptors (CD4 and co-receptor CCR5) to continue its replication process. This process can be controlled by both virus and host genetic factors that may lead to induction of disease or virus clearance. We investigated the anti-viral activity of TRIM5 alpha which is primate species-specific and which shows different degrees of potency to block the uncoating step in retrovirus replication, CCR5 Δ32 frequency as well as HIV Type 1 genetic variants that circulate in these communities. METHODS: Full length HIV-1 genotyping was performed by sequencing and phylogenetic analysis using the MEGA 3 software, CCR5 genotyping by Restriction Fragment Length Polymorphism assay and TRIM5 alpha genotyping by sequencing and TRIM5 alpha protein analysis RESULTS: The most prevalent HIV-1 variants identified were CRF02_AG (66.5%) and unique recombinant forms (10.4%) and less so Subtype A2, D, F2, G, CRF01_AE, CRF11_cpx, CRF13_cpx, CRF18_cpx,CRF22_01A1, CRF25_cpx, and CRF37_cpx. No individual with CCR5 Δ32 allele was identified. A rare allele of TRIM5 alpha with negative-dominant effect on its anti-viral activity was identified among 4.2% of pygmies in this study. CONCLUSION: Data on these human genes may be useful in explaining zoonotic retroviral infections and possible targets for anti-viral drug development and provide clues to our understanding of population risks to emerging viral infections.

A TIME FOR CHANGE: THE MUTATION RATE OF *MYCOBACTERIUM TUBERCULOSIS* DURING LATENT INFECTION.

Christopher B Ford *[1], Philana L Lin *[2], Michael R Chase[1], Rupal R Shah[1], Oleg Iartchouk[3], James Galagan[4], Marc Lipsitch[5,1], JoAnne L Flynn[6], Sarah M Fortune[1]

[1]Harvard School of Public Health, Immunology and Infectious Diseases, Boston, MA, 02115, [2]Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center, Department of Pediatrics, Pittsburgh, PA, 15224, [3]Harvard Medical School, Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA, 02115, [4]Boston University, Department of Biomedical Engineering and Department of Microbiology, Boston, MA, 02115, [5]Harvard School of Public Health, Center for Communicable Disease Dynamics, Department of Epidemiology, Boston, MA, 02115, [6]University of Pittsburgh School of Medicine, Department of Microbiology and Molecular Genetics, Pittsburgh, PA, 15261

* These authors contributed equally to this work

Mycobacterium tuberculosis (Mtb) poses a global health catastrophe that has been compounded by the emergence of highly drug resistant Mtb strains. In Mtb, all drug resistances are the result of chromosomal mutations and depend on the bacterium's capacity for mutation during the course of infection. We used whole genome sequencing (WGS) to directly compare the accumulation of mutations in Mtb isolated from cynomolgus macaques with active, latent and early reactivation disease. Based on the distribution of single nucleotide polymorphisms (SNPs) observed, we calculated the mutation rates for these disease states. Our data suggest that during latency, Mtb acquires a similar number of chromosomal mutations as would be expected to emerge in a logarithmically growing culture over the same period of time despite reduced bacterial replication during latent infection. The pattern of polymorphisms indicates that the mutational burden in vivo is due to oxidative DNA damage that occurs in the host environment. Thus, we demonstrate that during clinical latency, Mtb continues to acquire chromosomal mutations and provide an explanation for the observation that isoniazid (INH) monotherapy for latent tuberculosis is a risk factor for the emergence of INH resistance.

HYPER-RECOMBINATION IN *STREPTOCOCCUS PNEUMONIAE*: QUANTITATION AND POTENTIAL IMPLICATIONS

Christophe Fraser[1], Nicholas J Croucher[2], Stephen D Bentley[2], William P Hanage[3]

[1]Imperial College London, Dept of Infectious Disease Epidemiology, London, W2 1PG, United Kingdom, [2]The Wellcome Trust Sanger Institute, Pathogen Genomics, Cambridge, CB10 1SA, United Kingdom, [3]Harvard School of Public Health, Department of Epidemiology, Boston, MA, 02115

The importance of homologous recombination in *Streptococcus pneumoniae* in the evolution of the species is appreciated.

In a first step, an analysis of 1,930 pneumococcal genotypes from six housekeeping genes highlighted however that homologous recombination does not seem to affect the species homogeneously. A subset of strains showed evidence of enhanced admixture with other populations, and was significantly enriched in lineages resistant to a range of antibiotics. It is hypothesized that these strains were the result of a history of hyper-recombination, which is expected to introduce both divergent alleles and resistance determinants at an elevated rate.

Subsequently, we analyzed 215 complete full-length genomes from a lineage suspected of having experienced a history of hyper-recombination. The Pneumococcal Molecular Epidemiology Network clone 1 (PMEN1) lineage is typically identified as being penicillin resistant and serotype 23F. The sample was mostly composed of a single lineage, ST 81, and close variants as defined by commonly applied molecular typing methods (see also related abstract by Croucher *et al*).

Recombination events were identified with a moving window χ^2 test; as well as being surprisingly frequent, affecting over half the genome within this sample alone, and with some identified hotspots, recombination events were not distributed evenly over a genealogy constructed from non-recombinant sequences. After characterizing the size and frequency distribution of recombination events, our model identified at least eight braches within the phylogeny affected by hyper-recombination events, within which recombinations were not only more frequent but also uncharacteristically large.

These results shed new light on the nature and dynamics of evolution within *Streptococcus pneumoniae*. One might expect selection to rapidly purge or fix selectively significant genetic changes, and so it appears surprising that clinically and selectively important traits such as antibiotic resistance should maintain a long-term association with hyper-recombination within whole subpopulations co-exisiting with other strains. We hypothesise that these subpopulations may be adapted to specific niches where co-infection and thus recombination is more common.

GENOME-WIDE ANALYSIS OF MICROEVOLUTION IN *S. SONNEI* ASSOCIATED WITH INCREASING DISEASE SEVERITY AND DRUG RESISTANCE IN VIET NAM

Kathryn E Holt[1], Nicholas R Thomson[2], Stephen Baker[3]

[1]University of Melbourne, Microbiology & Immunology, Melbourne, 3010, Australia, [2]Wellcome Trust Sanger Institute, Pathogen Genomics, Cambridge, CB101SA, United Kingdom, [3]Hospital for Tropical Diseases, Oxford University Clinical Research Unit, Ho Chi Minh City, district 5, Vietnam

Shigella sonnei is the leading cause of bacterial dysentery (Shigellosis) in industrialised countries, while *Shigella* spp. *flexneri, dysenteriae* and *boydii* occur frequently in developing countries. In Ho Chi Minh City, Viet Nam, the proportion of Shigellosis caused by *S. sonnei* has shifted from 30% in 1995-1996 to 70% in 2007-2008 (Vinh et al, 2009, *BMC Infect Dis* 9:204). This shift coincided with increases in the severity and drug resistance of *S. sonnei* infections.

We generated whole genome sequence data for 91 *S. sonnei* isolated from patients Ho Chi Minh City during three study periods, 1995-1996 (N=21), 2000-2002 (N=38) and 2006-2008 (N=39), using multiplex paired-end sequencing on the Illumina platform. We used bwa to identify single nucleotide polymorphisms (SNPs) between the Vietnamese isolates and two available reference sequences.

We identified 2,800 SNPs in total, which were used to construct a phylogenetic tree. Eighty-three (91%) of the Vietnamese isolates formed a single clonal complex in which each isolate differed from all others by no more than 100 SNPs. The clonal complex consisted of three clonal groups (A, B, C), which arose successively (A->B->C) and were closely associated with year of isolation. The successive clonal groups account for the increasing trends in drug resistance observed between study periods, with increasing rates of extended spectrum beta-lactamase (ESBL) production (A-0%, B-5%, C-30%) and resistance to nalidixic acid (A-0%, B-20%, C-95%). The succession of clones also accounted for the increased disease severity, with average length of hospital stay rising from 2 to 4 to 6 days and average total duration of illness rising from 4 to 5 to 6 days. Distance from the root of the phylogenetic tree was linearly correlated with year of isolation, with an average accumulation of 3.2 SNPs per genome per year. The clonal groups were each defined by multiple SNPs and other variants, which may be associated with the differences in disease severity or drug resistance observed between the groups.

CONTRASTING CLONAL AND NON CLONAL BACTERIAL PATHOGENS' POPULATION GENETICS: *BACILLUS ANTHRACIS, F. TULARENSIS, Y. PESTIS* AND *BURKHOLDERIA PSEUDOMALLEI*

Paul S Keim

NAU/TGen, Pathogen Genomics, Flagstaff, AZ, 86011-4073

Many infectious diseases have emerged and circulated around the world with the development of human civilizations and global commerce. Anthrax, plague and tularemia are three such zoonotic diseases that have been intensely studied through genome characterization and phylogeographic analyses. A few highly fit genotypes within each of the causative species represent the vast majority of observed disease cases. Mutational and selective forces working together create highly adapted pathogens, but this has to be coupled with ecological opportunities for global expansion. In contrast, *Burkholderia pseudomallei* is an opportunistic pathogen (melioidosis) with a large environmental reservoir. Infecting humans, and perhaps all higher animals, is relatively unimportant in the evolution and ecology of this highly non clonal bacterium. But, global dissemination of the pathogen has occurred and population genetic patterns are still discernable using statistic, rather than cladistic, analyses. Population genetic analysis provides the underlying basis for both molecular epidemiology and microbial forensics. A basic understanding of the inheritance mechanisms for each pathogen is essential for validating model assumptions.

GENOMIC INSIGHTS INTO THE CONVERGENCE OF CAMPYLOBACTER SPECIES

Samuel K Sheppard[1], Xavier Didelot[1], Aaron Darling[2], Michael Egholm[3], Steven D Bentley[4], Julian Parkhill[4], Daniel Falush[5], Martin C Maiden[1]

[1]University of Oxford, Zoology, Oxford, OX13PS, United Kingdom, [2]Univeristy of California, Davis, Biology, Davis, CA, 95616, [3]454 Life Sciences, Research and Development, Branford, CT, 06405, [4]Wellcome Trust Sanger Institute, Pathogen Genomics, Cambridge, CB10 1SA, United Kingdom, [5]Univeristy College, Cork, Environmental Research Institute, Cork, N/A, United Kingdom

Current observations of the population genetics and evolution of Campylobacter jejuni and Campylobacter coli have proved instructive in modeling divergence in bacteria. In a recent paper we described a change in the patterns of inter-species gene flow which are consistent with despeciation of these bacteria. This has generated considerable interest and positioned the genus at the centre of research into fundamental biological processes that generate and maintain bacterial species. Here we describe a genomics approach to investigate the patterns of gene flow among Campylobacter populations, by identifying homologous and non-homologous recombination in 30 sequenced genomes. A whole genome neighbor-joining genealogy confirmed the 3-clade structure within C. coli and analysis using a progressive Mauve backbone output to quantify the "core" and "pan" genomes demonstrated that while the average genome size (approximately 1.6 Mb) varies little between groups, C. coli clade 1 has a smaller core genome and larger pan genome per strain than the other two C. coli clades or C. jejuni. Probabilistic assignment of nucleotide polymorphisms (239543) to species, using STRUCTURE, demonstrated that, while C. coli clades 2 and 3 and C. jejuni recombined mostly within species, as much as 20% of the C. coli clade 1 genomes is of C. jejuni origin. Further analysis identified 2237 homologous and 590 non-homologous recombination events (>50 bp) within C. coli clade 1, using CLONALFRAME and GENOPLAST respectively. The results supported the hypothesis that this is largely due to recent recombination between C. jejuni and C. coli clade 1. This is consistent with our previous findings concerning species convergence.

UNDERSTANDING EVOLUTION THROUGH GENOME SEQUENCING OF METHICILLIN RESISTANCE *STAPHYLOCOCCUS AUREUS* USA300 CLINICAL ISOLATES

Ryan Tewhey[1,2], Christopher R Cannavino[3,4,5], John A Leake[3,4], Vikas Bansal[1], Eric Topol[1,6], Ali Torkamani[1], John S Bradley[4,5], Nicholas J Schork[1,6]

[1]Scripps Translational Science Institute, Scripps Genomic Medicine, La Jolla, CA, 92037, [2]University of California, San Diego, Division of Biological Sciences, La Jolla, CA, 92093, [3]Rady Children's Hospital San Diego, Department of Pediatrics, Division of Infectious Diseases, San Diego, CA, 92123, [4]University of California, San Diego, Department of Pediatrics, Division of Infectious Diseases, San Diego, CA, 92123, [5]Rady Children's Hospital San Diego, Department of Pediatrics, Division of Hospital Medicine, San Diego, CA, 92123, [6]The Scripps Research Institute, Department of Molecular and Experimental Medicine, La Jolla, CA, 92037

With the rapid advances in sequencing technologies the ability to produce many complete bacterial genomes cost effectively allows for an unprecedented view into the evolution of bacterial populations. We have applied whole genome sequencing to 36 clinical isolates of community acquired methicillin resistant *staphylococcus aureus* (MRSA). The isolates were selected from infected individuals with a broad array of clinical presentations allowing us to look at genomic differences that may play a role in pathogenesis and drug resistance. We sampled individuals without co-infection, compromised immune systems, and other phenomena that may confound our conclusions. We ultimately find evidence for selection within the MRSA genome over time as well as for the potential role of particular pathways and processes that mediate pathogenesis. The results of this study lay a foundation for large-scale whole genome association studies of infectious diseases that will be able to highlight pathways and single genes directly influencing virulence

PUBLIC HEALTH GENOMICS TO IMPROVE GLOBAL HEALTH

Elena Ambrosino, Angela Brand

Maastricht University, Institute for Public Health Genomics, Maastricht, 6211RE, the Netherlands

In today's world, more then ever, Public Health is a global issue. While on one side globalization is seen as a tool to overcome poverty and diseases, on the other it is felt as a threat to be counteracted. Whatever the view on globalization is, the consensus is that it affects Public Health, thus a more comprehensive approach should consider the global perspective (Global Public Health), beyond concerns of nations.
Global Public Health-related problems consist of both health matters, as global issues indirectly related to, but affecting Public Health. In the former group fall the HIV and malaria epidemics; while in the latter rural development and climate change.
Within Public Health, the field of Public Health Genomics (PHG) combines knowledge from genetic science with insights from population and social sciences, and is defined as the effective translation of genome-based knowledge for the benefit of population health.
This study aims at evaluating the potential of PHG to face existing global issues, by exploring the integration of genome-based knowledge and technologies into Global Public Health.
Existing evidence supports the successful implementation of PHG in some Public Health areas, as molecular epidemiology, diagnostics and vaccinology. In addition, advances in genomics could be used in an even wider range of fields to tackle Global Health. This is the case of nutritionally enhanced food, environment bioremediation and disease resistant-crops. Additionally, PHG can improve Global Health by means of genome-based information of basic mechanism of diseases, thus guiding the development of future health interventions. Such advances can influence the prevention, diagnosis and treatments of some of the major Global Health treats, as malaria, tuberculosis and HIV.
Furthermore, a key reason to invest in genomics and integrate its knowledge into Global Public Health is to prevent the increase in the existing gap in genome-based research and use of genome-based information around the globe. Such "genomic divide", if unchecked, could lead to even greater disparities in Global Health.
Finally, implementing PHG should be seen as a mean of capacity and infrastructure building and scientific development in countries around the world.
In conclusion, Global PHG has the potential to meet the Public Health needs of both developed and developing countries. In the implementation of Global PHG countries' health needs and known benefits of genome-based information and technologies should be carefully assessed. Finally integration of PHG into Global Health will allow a fairer use of scientific advances around the globe and the reduction of the genomic divide.

DRAFT GENOME SEQUENCE OF A *TRYPANOSOMA CRUZI* I (TCI) STRAIN.

Oscar Franzén[1,2], Stephen Ochaya[1,2], Ellen Sherwood[1,2], Martin Llewellyn[3], Michael Lewis[3], Michael A Miles[3], Björn Andersson[1,2]

[1]Karolinska Institutet, Department of Cell and Molecular Biology, Stockholm, 17177, Sweden, [2]Science for Life Laboratory, KISP, Solna, 17121, Sweden, [3]London School of Tropical Medicine and Hygiene, Department of Infectious and Tropical Diseases, London, WC1E 7HT, United Kingdom

Trypanosoma cruzi is the causative agent of Chagas disease. The genome sequence of the *T. cruzi* reference strain CL Brener was published in 2005, and has greatly facilitated research into this disease. However, there are many *T. cruzi* evolutionarily diverged lineages with different geographical distributions, hosts and possibly also different clinical manifestations. We have therefore performed draft genome sequencing of the *T. cruzi* I isolate Sylvio X10/1 and performed comparative analysis to CL Brener (TcVI). No lineage specific genes could be identified for either strain. A large amount of allelic polymorphism was present, including nucleotide polymorphisms, small insertion-deletion events and microsatellites. Many genes show evidence of positive selection. The overall repeat content of Sylvio X10/1 is lower than that of CL Brener, including repeated genes, and an analysis of sequence reads revealed a 9-30% divergence of the major surface antigen family genes. Estimation of the number of genes present in the large *T. cruzi* gene families indicates that the mucin multigene family is the most diverged and expanded in these two genomes.

SEEKING NEW GENETIC TARGET OF CHLOROQUINE RESISTANCE IN INDIAN *PLASMODIUM FALCIPARUM* WITH EVOLUTIONARY GENOMIC APPROACH

Gauri Awasthi[1], Aparup Das[1]

[1]National Institute of Malaria Research, Evolutionary genomics and Bioinformatics, New Delhi, 110077, India, [2]National Institute of Malaria Research, Evolutionary genomics and Bioinformatics, New Delhi, 110077, India

Emergence and spread of chloroquine-resistant malaria parasites is one of the major problems in the treatment and prophylaxis of malaria. Evidences so far have suggested the role of the *Pfcrt* gene, located on chromosome 7, in conferring resistance against chloroquine (CQ). In particular, an amino acid mutation in the 76th position of the *Pfcrt* protein has been directly implicated. However, several bodies of evidences have directly or indirectly suggested that *Pfcrt* might not be the sole gene governing chloroquine resistance (CQR) in *P. falciparum*. Among all this, observed linkage disequilibrium between the *Pfcrt* and *Var S4* region suggest that genes other than the *Pfcrt* present in this region (alone or in association with *Pfcrt*) might also govern CQR in *P. falciparum*. To test this hypothesis we followed evolutionary genomic approaches to scan the 530 kb DNA fragment of Chromosome 7 encompassing the *Pfcrt* and the *var* genes including seven other transporter genes tandemly placed as cluster (*CG3*, *CG4*, *CG8*, *CG1*, *CG6*, *CG2* and *CG7*). We designed primers to amplify and sequence 10 non-coding DNA fragments (one marker each in 10-15Kb) present in intergenic regions in 15 CQ sensitive (CQS) and an equal number of *P. falciparum* CQR isolates from India. Results on DNA polymorphism and summary statistics were different for CQS and CQR types. Most significantly, the patterns of nucleotide diversity across the 10 fragments were quite dissimilar (and opposite) for CQR and CQS populations. While no definite pattern could be observed in CQS isolates, in the CQR isolates, nucleotide diversity was found to dip suddenly and reached almost zero (a characteristic feature of a recent "selective sweep") while approaching the *CG2* gene, but again took a leap at the next fragment (located at the end of the *CG2* gene). Thus to verify if *CG2* gene is also included under the "selective sweep" region, we sequenced four different fragments of *CG2* gene in both the CQS and CQR types and find very meager nucleotide diversity in CQR types in comparison to the CQS types, signifying that the *CG2* gene might be under the influence of positive natural selection in CQR but not in the CQS isolates. Since the *CG2* gene has earlier been implicated in chloroquine resistance in *P. falciparum*, the role of *CG2* gene in chloroquine resistance might not be ruled out.

GENOMICS WITHOUT GENETICS: THE *SIMULIUM* GENOMICS PROJECT.

Charles L Brockhouse[1], Soochin Cho[1], Guishan (Gary) Xiao[2], Eric J Haas[3], John K Colbourne[4]

[1]Creighton University, Biology, Omaha, NE, 68178, [2]Creighton University, Medical Center, Omaha, NE, 68178, [3]Creighton University, Chemistry, Omaha, NE, 68178, [4]Indiana University, The Center for Genomics and Bioinformatics, Bloomington, IN, 47405

The family Simuliidae (Black Flies) is generally regarded as the second most significant insect disease and pest taxon inflicting humankind. The biting adult females transmit a wide variety diseases, including onchercercisis, leishmaniasis, and many arboviruses. Much of the health and economic impact of the family is also caused by sheer mass biting, causing severe health problems in the tropics, and heavy economic losses to agriculture world-wide. In spite of this medical and economic impact, molecular studies of this family have languished since the early DNA-DNA reassociation kinetics analysis, largely because of the near impossibility of establishing laboratory colonies. Instead, cytogenomic analysis of natural populations have yielded a high resolution picture of the family-wide population structure unmatched in any other taxon. In the absence of (molecular) genetic maps, we have initiated a genomics project focusing on the only extant black fly colony (*Simulium vittatum*) with a transcriptome analysis (https://wiki.cgb.indiana.edu/display/grp/Simulium). To supplement the standard characterization of the transcriptome, we have used proteomic approaches such as 2D protein electrophoresis, post-translation modification studies and mass spectrometry. Using *de novo* sequencing ms-ms data we have isolated ESTs corresponding to novel proteins of interest, focusing on novel silk, storage and reproductive proteins. We anticipate that genomic/proteomic approaches for non-laboratory organisms will find wide-spread use as genomic studies of arthropod disease vectors reaches past dependence on laboratory colonies into natural populations.

INCREASED RESOLUTION OF LARGE MULTI-HOST *CAMPYLOBACTER JEJUNI* AND *C. COLI* CLONAL COMPLEXES BY EXTENDED MULTI-LOCUS SEQUENCE TYPING.

A J Cody[1], S K Sheppard[1], K E Dingle[2], M C Maiden[1]

[1]University of Oxford, Department of Zoology, Oxford, OX1 3PS, United Kingdom, [2]University of Oxford, National Institute for Health Research, Oxford, OX3 9DU, United Kingdom

Introduction:
The population structure of *Campylobacter jejuni* and *C. coli* has been well characterised by use of multilocus sequence typing (MLST) with 43 clonal complexes described to date. Association between some clonal complexes and isolation source has been exploited to identify the reservoir of origin of particular strains. However, several multi-host clonal complexes, notably sequence type (ST) 21 complex, which are large and diverse cannot be resolved by 7-locus typing. Our aim was to determine whether data from additional loci would provide further insight into *Campylobacter* population genetics, host association and evolution.

Methods:
Fragments of 14 housekeeping genes, either previously used in alternative *Campylobacter* typing schemes or employed in MLST schemes for other organisms, were sequenced from; (i) *C. jejuni* isolates from the predominant ST of the ST-21 complex (including ST-19, ST-21, ST-50, ST-51 and ST-104), (ii) *C. coli* isolates from clades 1, 2 and 3, (iii) representative(s) of the central genotype of the other 42 clonal complexes. Genealogical analyses of the data included the use of ClonalFrame.

Results:
Allelic information from 14 loci increased the resolution of large multi-host complexes and subspecies *C. coli* clades. Isolates from 5 predominant STs of the multi-host ST-21 complex were sub-divided to create a total of 30 21-locus STs on differences identified at between 10 and 2 of the 14 additional loci. ClonalFrame analysis confirmed the association between ST-48, ST-206 and ST-21 complexes.

Conclusions:
Analyses of additional loci is informative in increasing resolution between isolates belonging to large multi-host clonal complexes. Further analyses of a larger data set is required to determine whether sequencing from more alleles is informative with respect to host-association.

COMPARATIVE SOLID™ SEQUENCING REVEALS GENOMIC PATTERNS OF VIRULENCE EVOLUTION IN THE BACTERIAL GENUS *LISTERIA*

Craig A Cummings*[1], Henk C den Bakker*[2], Vania Ferreira[2], Paolo Vatta[1], Renato H Orsi[2], Lovorka Degoricija[1], Melissa Barker[1], Olga Petrauskene[1], Manohar R Furtado[1], Martin Wiedmann[2]

[1]Applied Biosystems, a part of Life Technologies Corporation, Foster City, CA, 94044, [2]Cornell University, Department of Food Science, Ithaca, NY, 14853

Although most species in the genus *Listeria* are non-pathogenic, saprophytic (decomposing organic material) bacteria, two species—the dangerous foodborne microbe *L. monocytogenes*, and *L. ivanovii*—are highly virulent and able to cause severe disease in infected hosts. Both of these species carry homologous virulence genes, suggesting shared mechanisms of pathogenicity, but the avirulent species of *Listeria* also share some of the putative virulence genes, thus making the specific identification of pathogenic *Listeria* challenging. To better understand evolution of *Listeria* virulence at the genome level, and catalog the set of genes that may be critical for virulence, we used SOLiD™ system sequencing to generate draft genomes for seven strains representing *Listeria* species or subtypes for which genome sequences were not available. Comparative analyses of these draft genomes and six publicly available genomes suggested a relatively closed *Listeria* pan-genome with 2,032 core and 2,918 accessory genes, limited gene loss and acquisition with conserved genome size (between 2.8 and 3.2 Mb), and a highly syntenic chromosome. Loss of multiple virulence-associated genes is a common feature of all non-pathogenic *Listeria* strains, although a consistent pattern of specific gene loss is not apparent. The data support a model in which modern *Listeria* species diverged, beginning in the Eocene, from a pathogenic common ancestor that contained key virulence genes, with emergence of saprophytic *Listeria* lineages occurring multiple times during the evolution of this genus. The genus *Listeria* thus provides an example of a group of bacteria that evolve through a loss of virulence rather than acquisition of virulence characteristics, as observed for many other infectious bacteria. While *Listeria* includes a number of species-like clades, many of these include subtypes or strains with atypical virulence-associated characteristics. The information obtained in this study will allow for the development of genetic and genomic criteria for pathogenic strain identification.

* These two authors contributed equally

HIGH-THROUGHPUT SEQUENCING OF 16 SALMONELLA GENOMES ON THE SOLID™ 3 PLUS SYSTEM

Lovorka Degoricija[1], Craig Cummings[1], Gregory R Govoni[1], Elena Bolchakova[1], Matthew L Ranieri[2], Andrea Switt[2], Henk C den Bakker[2], Martin Wiedmann[2], Manohar Furtado[1]

[1]Life Technologies Corporation, Applied Molecular Testing, Foster City, CA, 94404, [2]Cornell University, Food Science, Ithaca, NY, 14853

Salmonellosis is an important global public health issue and rapid detection of pathogenic bacteria in food supplies has become an increasing priority. The SOLiD™ System was used for whole genome sequencing of 16 less common human disease associated *Salmonella* serotypes to (i) expand our knowledge of the *rfb* locus and to support DNA-based approaches for serotyping and (ii) to characterize mobile elements and plasmids.

In this study, the SOLiD™ 3 Plus System was used to sequence whole genomes of 16 different serovars of *Salmonella*. Mate paired libraries with approximately 1.5 kb inserts were constructed for each genome and beads were generated by emulsion PCR prior to deposition on an eighth of a slide. Sequencing was carried out on a SOLiD™ System at 25 base pairs for the R3 and F3 tags. After correcting errors in colorspace reads, *de novo* assembly was performed using the SOLiD *de novo* pipeline.

On average, 28.5 million reads from each strain were generated for the F3 and R3 tags. The deep read coverage ranged from 143 to 200 fold per nucleotide. *De novo* assembly of the Salmonella short reads resulted in 2,345 contigs per genome. The sum of the length of the contigs is between 4.7 Mb, which is comparable to genome sizes of previously sequenced *Salmonella* genomes. Widely distributed *rfb* genes *wzx* and *wzy*, encoding for an O-antigen export unit and O-unit polymerase are present in the *rfb* clusters associated with most *Salmonella* serogroups (except serogroups O:2(A), O:4(B), or O:9(D1), and represent possible targets for molecular serotyping. Mobile elements and plasmids were identified by finding scaffolds that do not match any *Salmonella* chromosome. Two different plasmids were identified in *S*. Montevideo: (i) a conjugative plasmid of 53 kb encoding antimicrobial resistance genes and (ii) a 299 kb conjugative plasmid of the IncH12 incompatibility group. One IncI1 plasmid of approximately 120 kb and one element of approximately 91 kb were found in serotype Inverness; the 91 kb element shows characteristics of a mobile element, which was also identified in serotypes Rubislaw and Urbana. Putative IncI1 plasmids harboring virulence genes were identified in Mississippi and Urbana isolates. A serotype Montevideo isolate was found to contain two plasmids, one IncW plasmid carrying an integron class 1 element that encodes sulfonamide and aminoglycoside resistance, and one IncH12 plasmid encoding resistance to several heavy metals, disinfectants, and tetracycline.

SALMONELLA ENTERICA GENOMICS: CRISPR ELEMENTS AND PHENOTYPE EVOLUTION

W. Florian Fricke[1], Mark Mammel[2], J. Eugene LeClerc[2], Jacques Ravel[1], Thomas A Cebula[1]

[1]Institute for Genome Sciences, Microbiology & Immunology, Baltimore, MD, 21201, [2]Food and Drug Administration, Center for Veterinary Medicine, College Park, MD, 20740

Salmonella is one of the leading causes of bacterial foodborne gastroenteritis worldwide. The over 2,500 *Salmonella* serotypes that exist can differ substantially in their geographic distribution, virulence, host adaptation and antimicrobial resistance. To date, few genome sequencing projects have studied the genetic determinants responsible for evolution of the *Salmonella* genus or the phenotypic features that distinguish different serotypes. We present the results of a genome sequencing project of 17 non-typhoidal *S. enterica* isolates from human, animal and food sources, including serotypes Heidelberg, Newport, Kentucky, St. Paul, Schwarzengrund, Agona, Dublin, Javiana, Weltevreden, Hadar, Virchow, and I4,[5],12:i:-.

Strains were selected to represent a variety of common serotypes and include isolate pairs from the same serotype that showed different antimicrobial resistance phenotypes (Newport, Schwarzengrund, Heidelberg, Kentucky) or that had previously been shown to belong to different sub-lineages within the same serotype (St. Paul, Newport). Using comparative genomics approaches, we identified large variations in the genome content of prophage regions and plasmid elements between and among the different serotypes - mobilome components, which are predicted to substantially impact the overall phenotypes of each serotype.

To investigate potential mechanisms responsible for different phage and plasmid contents, we analyzed Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) elements found in each of the genomes. Our results show that inter- and intra-serotype phenotype variations are reflected by different sets of CRISPR elements. Furthermore, *S.* Newport and *S.* St. Paul genomes from different lineages within the same serotype are characterized by different sets of CRISPR spacer elements, possibly reflecting adaptation to varying plasmid and/or phage contents.

Our results suggest that CRISPR elements play an important role in serotype evolution within the *Salmonella* genus with impact on virulence, antimicrobial resistance and host specificity and should be further examined in the future.

COMPARISON OF FOUR COMPLETE GENOMES OF JAPANESE *HELICOBACTER PYLORI*: GENOME REARRANGEMENTS AND GENOMIC ISLANDS

Yoshikazu Furuta[1], Mikihiko Kawai[2], Koji Yahara[3], Takeshi Tsuru[1], Noriko Takahashi[1], Naofumi Handa[1], Kenshiro Oshima[4], Masahira Hattori[4], Masaru Yoshida[5], Takeshi Azuma[5], Ikuo Uchiyama[2], Ichizo Kobayashi[1,6]

[1]Graduate School of Frontier Sciences, University of Tokyo, Department of Medical Genome Sciences, Minato-ku, Tokyo, 108-8639, Japan, [2]National Institute for Basic Biology, Laboratory of Genome Informatics, Okazaki, Aichi, 444-8585, Japan, [3]Kurume University, Graduate School of Medicine, Kurume, Fukuoka, 830-0011, Japan, [4]Graduate School of Frontier Sciences, University of Tokyo, Department of Computational Biology, Kashiwa, Chiba, 277-8561, Japan, [5]Kobe University, Graduate School of Medicine, Chuou-ku, Kobe, Hyogo, 650-0017, Japan, [6]University of Tokyo, Institute of Medical Science, Minato-ku, Tokyo, 108-8639, Japan

A clinical study in Japan strongly suggested that *Helicobacter pylori* is responsible for recurrence of gastric cancer. There have been hypotheses linking the genotype of East Asian *H. pylori* to the high incidence of gastric cancer in East Asia. In order to fully understand the East Asian strains, our group determined entire genome sequences of 4 *H. pylori* strains from Japanese patients by Sanger sequencing and compared them with the 6 complete genome sequences. We analyzed ortholog relationships, phylogenies, horizontal transfers, gene gains and losses, and genome rearrangements.

Potential macro-regional genome rearrangements were analyzed and rearrangement pathways were reconstructed by combination of regional inversion events. Even within the closely-related Japanese isolates, many genome inversions have occurred at various positions. Their recombination points were frequently linked to restriction-modification (RM) genes, which suggests their involvement. We also detected RM systems and genomic islands specific to the East Asian strains. In a Japanese strain, we found a genomic island with features of a prophage. This may be the first report of prophage-like region in *H. pylori*. Analysis of the other plastic regions and genes will also be reported.

DETECTION OF DNA METHYLATION IN *SCHISTOSOMA MANSONI*

Kathrin K Geyer, Carlos M Rodriguez Lopez, Michael J Wilkinson, Karl F Hoffmann

Aberystwyth University, IBERS, Aberystwyth, SY23 3DA, United Kingdom

With over 200 million people suffering from schistosomiasis, it is the second most significant tropical parasitic disease after malaria. Due to the absence of a vaccine, patients rely on the single anthelmintic agent praziquantel for treatment, which has worryingly led to fears of drug resistance.

As expected for a complex metazoan parasite, schistosomes exhibit a high degree of gene regulation. Large-scale transcriptome studies have shown stage-specific, as well as gender-specific expression of multiple genes. DNA methylation-mediated gene silencing as a means of transcriptional control has evolved in higher eukaryotic organisms. While recent studies have established that schistosomes post-translationally modify histones and express small regulatory RNAs, no compelling evidence exists that demonstrates this parasite's genome is methylated.

However, the sequencing of the *Schistosoma mansoni* genome revealed the presence of key proteins involved in the DNA methylation machinery - a putative Dnmt2-like protein (SmDNMT2) as well as a Methyl-Binding-Domain Protein (SmMBD). Expression analysis has shown that SmMBD and SmDNMT2 are developmentally co-regulated - suggesting an important role of DNA-methylation during lifecycle progression, intra-snail maturation and female biology.

To provide preliminary evidence that the genome of this blood fluke is indeed methylated, methylation sensitive amplification polymorphism (MSAP) analyses were performed on various schistosome lifecycle stages. Here, cytosine methylation was detected in both free-living and parasitic life stages as well as in both sexes. Strikingly, culturing *S. mansoni* adult pairs in the presence of the DNA-methyltransferase inhibitor 5-Azacytidine (5-AzaC) resulted in a general decline of egg production with increasing drug concentration. Furthermore, MSAP analysis showed a change in methylation pattern after 5-AzaC treatment. Therefore, as the eggs are the primary cause of the pathology associated with schistosomiasis, targeting the DNA methylation system of the parasite could be an exciting treatment opportunity.

Collectively these observations will contribute to a greater understanding of Platyhelminthes developmental biology, assist in the evolutionary reinterpretation of DNA methylation across the Metazoa and perhaps, provide a cautionary note for transgene expression studies within the parasite.

WHOLE GENOME SEQUENCING REVEALS FINE-SCALE GENETIC HETEROGENEITY AMONG *S. AUREUS* ISOLATES FROM A SINGLE INDIVIDUAL

Tanya Golubchik[1], Ruth Miller[2], Helen Farr[2], Daniel Wilson[1,2], Hanna Larner[2,3], Rowena Fung[2], Heather Godwin[2], Derrick Crook[2], Rory Bowden[1]

[1]University of Oxford, Department of Statistics, Oxford, OX1 3TG, United Kingdom, [2]NIHR, Oxford Biomedical Research Centre, Oxford, OX3 9DU, United Kingdom, [3]Wellcome Trust Centre for Human Genetics, -, Oxford, OX3 7BN, United Kingdom

Staphylococcus aureus is a gram-positive bacterium that has gained notoriety in recent years as a major cause of serious hospital-acquired infections, including soft tissue infections and bacteraemia. *S. aureus* is carried asymptomatically by around one third of healthy adults, and this reservoir of carried strains is thought to play an important role in the generation of genetic diversity among *S. aureus*. As part of a carriage surveillance study in Oxford, we have collected 1991 *S. aureus* isolates, including multiple longitudinal samples from healthy participants taken at two-month intervals over a period of up to 24 months. We are interested in the genetic diversity among carried and disease isolates, with the aim of using this information to inform studies into transmission patterns of *S. aureus*. Samples were selected such that each individual carried only a single *S. aureus* sequence type, as determined by *spa* typing. For each sample, 12 isolates per timepoint were cultured from nasal swabs and sequenced on the Illumina platform. Illumina reads were mapped to the closest available reference genome, using the Maq mapping software. We applied a custom, highly specific filtering algorithm to identify high-quality variable sites for which there was good support in the data. Even within these apparently homogeneous samples, whole genome sequencing clearly revealed multiple distinguishable genotypes carried by a single individual at a given timepoint, as well as evidence of continuing evolution within each of those genotypes. It is evident that conventional typing techniques underestimate the diversity of *S. aureus* genotypes in carriage, and that this level of diversity within a single individual has to be taken into account when examining possible routes of transmission between epidemiologically linked cases.

ETIOLOGY OF DIARRHEA IN BOLIVIA.

Lucia Gonzales[1], Enrique Joffre[2], Rosario Rivera[2], Samanta Sanchez[2], Ann-Mari Svennerholm[1], Asa Sjoling[1], Volga Iñiguez[1]

[1]University of Gothenburg, Institute of Microbiology and Immunology, Gothenburg, 40530, Sweden, [2]Universidad Mayor de San Andres, Instituto de Biologia Molecular y Biotecnologia, La Paz, NA, Bolivia

Diarrhea is one of the most common causes of mortality and morbidity among infants and children in developing countries. In Bolivia, diarrhea is the second cause of mortality in children under 5 years. However, little is known concerning the principal pathogens involved with the disease. The prevalence of Diarrheagenic *E. coli* (DEC) and rotavirus was determined using routine bacteriology and molecular assay techniques in 1673 stool samples collected from children with diarrhea less than five years of age from hospitals of La Paz and Cochabamba, Bolivia during 2007 and 2008. The most frequently isolated enteric pathogens were rotavirus (46.6%), *enteroaggregative E. coli* (EAEC, 12.6%), *enterotoxigenic E. coli* (ETEC, 7.7%) and *enteropathogenic E. coli* (EPEC, 6.6%). Multiple pathogens were recovered from 12.5 of the diarrheagenic children with the virus-bacteria combination as the most frequent (10.3%). The seasonal distribution showed a high prevalence of rotavirus during the dry and cold season (June) and the presence of biannual distribution for ETEC with peaks during the transition time (May and September). Rotavirus infection is mainly present in children under 12 months, while infection with bacteria is present between 6 and 18 months for EAEC and EPEC and between 8 and 24 months for ETEC. High frequency of resistance to ampicilin (75.5%), trimetoprim/sulfamethoxazole (59%) and tetracycline (44.3%) was displayed among *Escherichia coli* strains. Etiologic data on diarrheal disease and susceptibility patterns of diarrheal pathogens are important tools for clinical management and control strategies planning.

POPULATION GENOMICS OF INDIAN *PLASMODIUM VIVAX* USING NON-CODING SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)

Bhavna Gupta, Aparup Das

National Institute of Malaria Research, Genomics and Bioinformatics, New Delhi, 110077, India

India is one of the very few countries where two major human malaria parasites, *Plasmodium falciparum* and *P. vivax* coexist. Since malaria due to *P. vivax* is comparatively rarely lethal than *P. falciparum*, research on *P. vivax* has attracted less attention. However, malaria morbidity caused by *P. vivax* infections contributes significantly to high socioeconomic burden in India. Population genetic studies can contribute significantly to understand evolutionary history of species populations in general, and of gene and genome evolution, in particular. However, very little is known about the genetic structure and demographic history of *P. vivax* in India. Thus, in order to gain insights into population structure and demography of *P. vivax*, we developed 11 putatively neutral DNA markers located in non-coding regions of a highly conserved ~200Kb region of *P. vivax* 13th chromosome. Sixty one *P. vivax* isolates were collected from six geographically distant regions of India. Altogether thirty five single nucleotide polymorphisms (SNPs) were identified in these 11 non-coding regions spanning about 5 Kb region in each *P. vivax* isolate. We utilized the data on the distribution of the SNPs to estimate genetic diversity and calculate different other population genetic parameters. The net genetic diversity was found to be variable across populations, and all the populations were found to follow the model of constant population size. Genetic differentiation between pairs of populations was not remarkable, suggesting that Indian populations of *P. vivax* are less genetically structured. Several other population parameters were estimated and will be discussed in detail.

ANALYSIS OF BACTERIA FROM INTESTINAL TRACT OF FAP PATIENTS FOR THE PRESENCE OF APC-LIKE SEQUENCES

Vladimir Holec[1,2], Sonia Ciernikova[3], Lenka Wachsmannova[3], Zuzana Adamcikova[3], Michal Mego[4], Viola Stevurkova[3], Vladimir Zajac[2,3]

[1]Comenius University, Medical faculty, Pathological Anatomy, Bratislava, 81108, Slovakia, [2]University Children's Hospital, Clinic of Pediatric Oncology and Hematology, Banska Bystrica, 97409, Slovakia, [3]Cancer Research Institute, Cancer Genetics, Bratislava, 83391, Slovakia, [4]National Cancer Hospital, Internal, Bratislava, 83101, Slovakia

Familial adenomatous polyposis (FAP) is an autosomal dominant hereditary predisposition presented by hundreds or thousands of polyps in the intestine with nearly 100% penetrance. This disease is induced by germ mutations in the tumorsupressor APC gene, which is in the process of carcinogenesis a critical event. We have found that the presence of APC-like sequences in bacteria of FAP patients could be beneficial in analyzing the causes of the existence of a disproportionate number of germinal mutations in the APC gene. We could hypothetically consider about horizontal transfer of genetic information between eukaryotic and prokaryotic cells. In the event of confirmation of this fact, this original model could bring new opportunities in research, diagnosis and therapy of this disease. Our present results support this hypothesis, because most of the APC-like sequences in bacteria are located in the MCR (mutation cluster region), where the majority of germ mutations is located. This fact will be proven by sufficiently large group of patients.
Thus, if one of the mutation sources were bacteria containing mutations in APC-like sequences, their limitation would be possible by means of probiotic bacteria. Pathological microflora could be competitively expelled by application of probiotics and thereby is created a basic form of treatment. Beneficial effect in preventing colon cancer with probiotic bacteria have been already demonstrated in some studies.

THE ROLE OF CONTINGENCY GENES IN THE ADAPTATION OF *CAMPYLOBACTER JEJUNI* TO A NOVEL HOST

John P Jerome[1], Julia A Bell[1], Anne E Plovanich-Jones[1], Jeff E Barrick[1], Charles T Brown[1,2], Linda S Mansfield[1]

[1]Michigan State University, Microbiology and Molecular Genetics, East Lansing, MI, 48824, [2]Michigan State University, Computer Science and Engineering, East Lansing, MI, 48824

Background: The food-borne pathogen *Campylobacter jejuni* is a leading cause of bacterial gastroenteritis in humans. Multiple highly mutable loci, or contingency genes, have been detected in *C. jejuni* and are coincident in the genome with genes responsible for synthesis of surface exposed molecules. These regions consist of mononucleotide tracts that are prone to slipped-strand mispairing during replication that leads to an insertion or deletion (indel) mutation. It has been suggested that variability at these loci is a mechanism for adaptation to a dynamic intestinal environment and/or immune avoidance. We hypothesized that specific phases of contingency genes would be selected due to serial passage in a mouse model.
Methods: *Campylobacter jejuni* NCTC 11168 was serially passaged four times in mice to adapt the bacteria to this host. Phenotype tests, microarray analysis, PFGE, and re-sequencing of the adapted *C. jejuni* has been performed.
Results: Serial passage resulted in decreased time to develop pathology, and increased severity of gross pathology and histopathology. These measures indicate a trend toward increased *C. jejuni* virulence after adaptation to the mouse gastrointestinal tract. Interestingly, standard *in vitro* assessments of virulence would suggest the adapted bacteria was less virulent. To understand the genetic mechanism of this adaptation, microarray analysis was first performed. Overexpression of a region in the flagellar glycosylation locus was detected in bacteria isolated after passage. This region contains mononucleotide tracts, and sequence analysis revealed indels in these loci after passage. Illumina re-sequencing to ~250-400X coverage allowed us to define all contingency genes in the genome before and after *in vivo* adaptation. Single nucleotide polymorphisms and small indels were discovered in our lab's *C. jejuni* relative to the published reference genome, but the only detected change in the adapted *C. jejuni* was in contingency genes.
Conclusions: Specific phases of contingency genes appear to be selected during *in vivo* passage of *C. jejuni*. This finding, coupled with the apparent lack of other genetic changes, is the first experimental evidence that these loci play a significant role in *C. jejuni* adaptability to a novel environment. Because the natural reservoir of *C. jejuni* is in avian species, these regions are likely to be significant for the ability of the bacteria to colonize and cause inflammation in another novel host, humans.

NEISSERIA POPULATION GENOMICS: INTEGRATING WHOLE GENOME DATA WITH MULTI LOCUS APPROACHES TO EPIDEMIOLOGY AND POPULATION BIOLOGY.

Keith A Jolley, Martin C Maiden

University of Oxford, Department of Zoology, Oxford, OX1 3PS, United Kingdom

Within the genus *Neisseria* pathogenesis has emerged from harmless progenitors on at least several occasions. It is yet to be fully understood as it is a complex polygenic trait in both gonococci and hyperinvasive meningococci. Comparative whole genome studies of multiple bacterial isolates with defined phenotypes provide a potentially powerful approach to studying meningococcal and gonococcal disease, but such studies present appreciable challenges in integrating and analysing large, complex datasets.

Here we propose a scalable isolate-centric approach to whole genome analysis and have developed and implemented via the Internet software that realises this paradigm (Bacterial Isolate Genome Sequence Database, BIGSdb). For each isolate provenance and phenotypic properties are linked to the sequence data, which may range from a single gene fragment, through multi locus sequence data and partial genome assemblies, to a whole closed genome. These sequence datasets can be queried with standard bioinformatics algorithms using curated reference datasets for given genetic loci. This enables the rapid discovery and characterisation of genetic variation and its association with known phenotypes. These loci can be grouped into schemes with unlimited numbers of members to reveal higher order structure, as in MLST, but schemes for metabolic processes, antibiotic resistance, and antigenicity are also possible. BIGSdb is currently running the reference and isolate databases for *Neisseria* sequences (http://pubmlst.org/neisseria/) where it hosts provenance and genotypic data for over 17000 isolates, including 22 complete genomes. It is also currently being used as an in-house tool to analyse the genomes of 200 *N. meningitidis* isolates, but there are no practical limits to the number of bacterial genomes, loci, and schemes which can be accommodated. The platform is available as a resource for community annotation and investigation of phenotypic and genotypic variation.

GLOBAL EPIDEMIOLOGY OF ENTEROTOXIGENIC *ESCHERICHIA COLI* (ETEC) EXPRESSING COLI SURFACE ANTIGEN 6 (CS6).

Matilda Nicklasson[1], John Klena[2], Rania Abdel El-Khalek[2], Erik Nygren[1], Anders Janzon[1], Firdausi Qadri[3], Åsa Sjöling[1]

[1]University of Gothenburg, Dept. of Medical Microbiology and Immunology, Göteborg, 40530, Sweden, [2]NAMRU-3, Clinical Trials Program, Cairo, 11517, Egypt, [3]International Centre for Diarrhoeal Disease Research (ICDDR,B), Dept. of Immunology, Dhaka, 1000, Bangladesh

Enterotoxigenic *Escherichia coli* (ETEC) is one of the most common causes of acute watery diarrhoea in low-income countries, causing approximately 400 000 deaths in children under the age of five every year. ETEC produce two main plasmid-encoded virulence factors; colonization factors (CFs), which adhere to the small intestine, and a heat-stable (ST) and/or a heat-labile (LT) enterotoxin. More than 20 CFs have been identified so far. One of the most common CFs woldwide is coli surface antigen 6 (CS6). The aim of this study was to determine the genetic relationship between CS6 positive ETEC strains isolated worldwide. The chromosomal genetic relationship was determined by Multi Locus Sequence Typing (MLST) by a scheme based on sequencing of internal regions of *E. coli* housekeeping genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*. The ST type (STh or STp) was determined by multiplex PCR. The CS6-encoding operon was analyzed by sequencing of a 540 bp region previously shown by us to exhibit variability in LT-only strains. As previously reported by us, the MLST sequence type 398 was the most common among STp/CS6 isolates from children in a Guatemalan village. Sequence type 398 also infected adult travellers in Guatemala and persisted in the region for at least six years (1998-2003). The same sequence type has now also been found in an ST/CS6 isolate from Qatar. The second most common sequence type identified in Guatemala and Mexico, sequence type 182, has now been identified in Turkey and Egypt. In Guatemala and Mexico, all 23 strains positive for STp were identical in the sequenced part of the CS6 operon. The same CS6 sequence has now been found in 9 STp/CS6 isolates from Bangladeshi children aged 0-2 years. The CS6 sequence of an STh/CS6 isolate from a child in Guatemala differed greatly from the sequence in the STp/CS6 strains, and has now also been identified in an STh/CS6 isolate from a Bangladeshi child. The study is ongoing and is being expanded to include a larger number of clinical isolates from around the world.

INVESTIGATING THE GENOTYPE-PHENOTYPE LINK IN INVASIVE *SALMONELLA* TYPHIMURIUM.

Chinyere K Okoro, Robert A Kingsley, Simon R Harris, Julian Parkhill, Gordon Dougan

Wellcome Trust Sanger Institute, Pathogen Genomics/Microbial Pathogenesis, Cambridge, CB10 1SA, United Kingdom

A highly invasive form of non-typhoidal Salmonella (NTS) disease in humans, mostly associated with *Salmonella enterica* serovar Typhimurium isolates, has emerged in many parts of sub-Saharan Africa. Host factors associated with the diseases in clinical settings include HIV, young age, immunesuppression, malaria and anaemia. The invasive *S.* Typhimurium isolates are mainly of multilocus sequence types (ST) 313 and 19 and some harbour multi-drug resistance cassettes.

Relatively little is known about the biology or phylogeny of *S*. Typhimurium isolates and the evolutionary relationship between these and non-invasive gastroenteritis-associated *S*. Typhimurium from other parts of the world. Preliminary analyses of the genome sequence of three invasive isolate genome sequences have shown little sequence diversity. In our studies, the population structure of invasive *S*. Typhimurium isolates from sub-Saharan Africa has been analysed by determining whole-genome-SNP based phylogeny of invasive strains against the background of global gastroenteritis strains from other parts of the world. Three clusters have been identified within which there is strong evidence of on-going microevolution. Subsequent investigations will explore the underlying genetic variation that underpins evolutionary changes over time and geographical space.

The metabolic variation and differences in *in vivo* colonisation patterns between invasive and gastroenteritis *S*. Typhimurium isolates were examined using phenotypic microarray studies and the murine infection model, respectively. The invasive isolates formed distinct clusters based on metabolic profiles similar to the phylogenetic clustering, inferring a correlation between observable phenotypic characteristics to genetic variation. Results obtained from systemic and intestinal colonisation studies in mice reveal heterogeneity in the invasive potentials of different invasive *S*. Typhimurium.

FROM COMMENSAL TO LETHAL INFECTION: GENOMIC ANALYSIS OF *E. COLI* AND *SHIGELLA*

Jason W Sahl[1], Julia C Redman[1], Jennifer Emberger[1], Adam Phillippy[2], David A Rasko[1]

[1]Institute for Genome Sciences, Microbiology and Immunology, Baltimore, MD, 21201, [2]University of Maryland College Park, Computer Sciences, College Park, MD, 20742

Escherichia coli exists as the major facultative anaerobe in the human gastrointestinal tract, but also can cause serious disease especially in young children, travelers and those that do not have access to clean water. While *E. coli/Shigella* have been a model microbial system in use for decades the advent of the genomic era has allowed unprecedented insight into the variation of this important species. There are currently six readily identifiable and generally agreed upon pathogenic variants of diarrheagenic *E. coli/Shigella*. We have determined that there are ~2200 genes that all *E. coli/Shigella* share and each novel genome will provide approximately 300 new genes. These two factors together indicate that the pan-genome of *E. coli/Shigella* is open and ever expanding. We identified that there is relatively little conservation among many of the pathovars suggesting that there are multiple routes of evolution for each of these groups. We have focused initial efforts on sequencing multiple strains of *Shigella* from multiple geographic locations and representatives of each of the species. Gene independent genome analysis has identified core genomic content that highlights the diversity of this group of pathogens and provides a finer resolution view of chromosomal evolution of these pathogens. Gene based comparisons has identified genomic regions that are species-specific as well as clinical presentation-specific. Comparative genomics has reached a new era where the number of strains that can be rapidly sequenced will begin to address epidemiological questions and become a more integrated component of the global health care system.

METAGENOMIC ANALYSIS OF VIRUSES IN BAT INTESTINAL TRACT, IMPLICATION OF IMPORTANT ROLES PLAYED BY BATS IN ECOSYSTEM

Xing-yi Ge, Yan Li, Zheng-li Shi

Chinese Academy of Sciences, Wuhan Institute of Virology, Wuhan, 430071, China

Bat is the second most diverse mammals on the earth. Increasing data indicate that bats harbor diverse viruses and some of them cause severe human diseases. In this study, random-polymerase chain reaction (rPCR) and high throughput sequencing technology (Solexa) was applied for metagenomic analysis of viruses harbored in bat intestinal tract. Bat feces were collected from six different locations and used for virus concentration and purification. The nucleic acid of purified viruses was extracted and used for Solexa sequencing. A total of 8,746,417 reads with a length of 306,124,595bp, were obtained. Among them, 13541 reads have homologies to phages, 9170 reads to other viruses. A total of 105 assembled contigs (>75nt) have been constructed and compared with GenBank, 22 contigs showed identities of 68-99% to known virus genomic sequences, 83 showed identities of 20-95% to viral protein sequences. The most frequent reads and contigs are homologous to densoviruses, dicistroviruses, coronaviruses, paroviruses and tobamoviruses, covering invertebrate viruses, vertebrate viruses and plant viruses. This study provides a more comprehensive understanding of virome in bat intestinal tract which may guide the discovery.

SEROLOGICAL EVIDENCE OF SARS-COV LIKE VIRUS IN SOME SPECIES OF MAMMALS FROM HUBEI PROVINCE, CHINA

Hanzhong Wang

Chinese Academy of Sciences, Wuhan Institute of Virology, Wuhan, 430071, China

In a search for the wildlife reservoir of SARS coronavirus (SARS-CoV), we used an ELISA to screen individuals from two orders of wild and domestic mammals in Hubei Province, China, where bats have recently been found infected with bat SARS-like CoV, for SARS-like CoVs. We also screened samples by RT-PCR. We found serological evidence of exposure to SARS-like CoVs in 4/24 (16.67%) of Chinese ferret badgers (Melogale moschata); 1/7 (14.29%) of hog badgers (Actonyx collaris); 6/18 (33.33%) of Edward's rat (Leopoldamys edwardsi); and 2/37 (5.41%) of palm civets (Parguma larvata). We found no evidence of exposure in small numbers of wildcaught Eurasian badgers (Meles meles) or domestic dogs, pigs or cats. We also found no viral sequences in any of the samples by RT-PCR. This is the first demonstration of exposure to SARS-like CoVs in wild mammals other than bats outside wildlife markets in China, and it suggests that a large diversity of SARS-like CoVs exists in wild mammals in China, and/or that the viruses carried by bats also circulate within a range of wild mammal species.

PHYLOGENETIC ANALYSIS OF *ESCHERICHIA COLI* ISOLATES BASED ON INTERGENIC SEQUENCES REVEALS LIFESTYLE ADAPTATIONS BUT NOT HOST SPECIFICITY.

Aaron P White[1,2,3], Kristen McLeod[2], Ricci Schaefer[2], Chris D Sibley[3], Michael G Surette[3], Norm F Neumann[3,4]

[1]University of Saskatchewan, VIDO, Saskatoon, S7N 5E3, Canada, [2]Provincial Laboratory, Public Health, Calgary, T6G 2J2, Canada, [3]University of Calgary, Microbiology and Infectious Disease, Calgary, T2N 4N1, Canada, [4]University of Alberta, Public Health, Edmonton, T6G 2G3, Canada

We analyzed a collection of 284 diverse *E. coli* isolates to determine if there were molecular signatures associated with specific host types. This would lead to development of diagnostic tests to assess virulence potential and need for remediation when water sources are contaminated with *E. coli*. Phylogenetic analysis was performed using the sequences of three intergenic, promoter-containing regions of the *E. coli* genome, controlling production of curli fimbriae (*csgD/B*), flagella assembly and motility (*flhDC*), and nutrient import (*ompF*). These regions were chosen because the associated gene products are conserved, known to exhibit variation, and expressed on the cell surface. Isolates were also analyzed for the presence of key virulence genes and phenotypic and metabolic (BIOLOG) testing was performed. The *E. coli* isolates were clearly differentiated into four commonly known phylogenetic groups (A, B1, B2 and D), with virulence genes most prevalent in group B2 and D isolates, as expected. In general, isolates from the same host type did not cluster together, except that 50% of human isolates were in the B2 subgroup and 56% of non-human isolates were in the B1 subgroup. All B1 isolates were highly similar, whereas isolates in groups B2 and A displayed comparably greater genetic distances. This indicated that B1 isolates had few molecular adaptations to a particular host environment and were predicted to be “host-generalist”, whereas group A and B2 isolates were much more likely to be “host-adapted”. In agreement with this hypothesis, prevalence of the rdar morphotype, a phenotype associated with survivability in the environment and long-term persistence, was significantly higher in B1 isolates (84%) than either B2 (41%) or A isolates (34%). There were no clear trends for phylogenetic groups or host types with respect to metabolic capacity. Our overall conclusions were that isolates from humans are more likely to contain virulence genes and appear to be more host-adapted, while *E. coli* isolates from non-human sources are generally less virulent and more host-generalist. Given the high rates of recombination observed in many *E. coli* genomes, we hypothesize that host specificity is not reflected in genome-wide adaptations but rather is characterized by the presence or absence of specific genes.

CLUES TO TRANSMISSION: FROM SOURCE ATTRIBUTION TO REAL-TIME TRACKING USING PATHOGEN GENOMES

Daniel J Wilson

University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, OX4 1BN, United Kingdom

The field of Evolutionary Genetics has established itself as an indispensable tool for solving practical problems in Infectious Diseases, providing a framework for understanding the genetic diversity found in pathogen populations. Patterns of genetic variation are informative about transmission, allowing us to reconstruct the epidemiological dynamics of pathogen populations, identify dominant transmission routes, and even to infer direct transmission events between epidemiologically related individuals. As the genetic resolution at which we can feasibly type pathogens has steadily increased, the Evolutionary Genetics revolution has become increasingly ambitious in its goals.

The aim of the Modernising Medical Microbiology consortium – a collaboration between the University of Oxford, the Wellcome Trust Sanger Institute, the UK Health Protection Agency and the NHS – is to rapidly detect and track the spread of clinically relevant pathogens in near-to-real time in order to facilitate measures for control and prevention. Using a combination of high-throughput whole genome sequencing and novel evolutionary analyses, we can characterise the evolution and epidemiology of clinical and carriage isolates in order to understand transmission with a view to design control and prevention strategies.

Drawing on examples from species such as Campylobacter jejuni, Staphylococcus aureus, Clostridium difficile and norovirus, I will illustrate how the tools of Evolutionary Genetics illuminate our understanding of transmission, from attributing the population-of-origin of zoonotic pathogens, through quantifying the intrinsic rate of spread of outbreak strains, to reconstructing direct transmission pathways at the individual host level.

I will discuss the issues of reliably identifying variants and coping with the complexities of recombination, population structure and genome plasticity in order to fully exploit whole-genome sequences for clues they hold to disease transmission.

DETECTION OF PROTEINS HOMOLOGOUS WITH HIV-1 ANTIGENS IN BACTERIA OF HIV POSITIVE PATIENTS

Vladimir Zajac[1], Michal Mego[2], Lenka Wachsmannova[1], Zuzana Adamcikova[1], Vladimir Holec[3], Viola Stevurkova[1], Vladimir Krcmery[4]

[1]Cancer Research Institute, Cancer Genetics, Bratislava, 42183391, Slovakia, [2] National Cancer Hospital, Internal, Bratislava, 42183101, Slovakia, [3]Medical Faculty, Comenius University, Pathological Anatomy, Bratislava, 42181108, Slovakia, [4]Saint Elizabeth University , Health, Bratislava, 42181000, Slovakia

Bacterial DNA isolated from the intestinal tract of American and Slovak HIV/AIDS patients and bacteria and yeasts isolated from respiratory tract of Cambodian and Kenyan HIV positive children were positive in hybridization assay for HIV-1 specific sequences. PCR products using specific primers for gag, pol and env HIV-1 genes, with template of these DNA, were found to be for more than 90% homologous to the corresponding HIV-1 sequences. In Western blotting analysis were detected specific HIV-1 proteins using monoclonal antibodies against HIV-1 antigens p17, p24 and gp41. Molecular weight of detected proteins are mostly not in accordance with corresponding viral proteins. The greatest consent was found in WB using MAbs against gp 41 and p17 in bacterial samples of all aforementioned patients. Differences between profile of detected protein by using MAbs against HIV-1 p24 of American and Slovak on the one side and Kambodian and Kenyan patients on the other one, are expressive. These differences are very probably determined by bacterial family or are results of evolutionary process. Finally, the origin of bacterial HIV-like sequences and the role of their products in AIDS process is discussed.

MULTISTRAIN ANALYSIS OF *ANAPLASMA*

Anthony F Barbet[1], Basima Al-Khedery[1], Michael J Dark[1], Ulrike G Munderloh[2]

[1]University of Florida, Infectious Diseases & Pathology, Gainesville, FL, 32611, [2]University of Minnesota, Entomology, St. Paul, MN, 55108

Anaplasma sp. are organisms from the Order *Rickettsiales* transmitted by ticks. *Anaplasma marginale* is a major impediment to livestock production world-wide and *Anaplasma phagocytophilum*, known for many years to cause persistent disease in ruminants, has recently been recognized as causing an emerging disease of humans and dogs in the U.S. and Europe. The causative agents have small circular genomes, ~1.2 Mb for *A. marginale* and ~1.5 Mb for *A. phagocytophilum*. A prominent feature of these genomes is the presence of numerous "functional pseudogenes" and a genomic locus into which duplicated pseudogenes, or pseudogene segments, may recombine, be expressed and generate antigenically variant outer membrane proteins. These pseudogenes generally consist of a central hypervariable region and different 5' and 3' lengths of conserved sequence repeated many times in different genome locations. There are ~15 pseudogenes in *A. marginale* which generate variant MSP2 and MSP3 proteins and ~100 pseudogenes in *A. phagocytophilum* generating variant MSP2/P44 proteins. The structure of these pseudogenes in each strain determines the repertoire of variant molecules that can be expressed. In turn, these repertoire differences have been linked to the ability of a particular strain to overcome pre-existing immunity, superinfect an animal already persistently infected with a different strain, and cause disease spread. The repertoires may also be associated with invasion of new host species. Expressed outer membrane protein variants are similar in U.S. human and dog strains of *A. phagocytophilum* but markedly different from strains infecting European ruminants. Consequently, it is important to determine the structure of pseudogene repertoires in different strains and their propensity for evolutionary change. We have obtained 454 sequence (regular and 3kb paired-end) for strains of *Anaplasma* from different regions of the U.S. and evaluated different *de novo* and reference mapper assembly methods for their ability to resolve the numerous repeats and generate reliable information about the pseudogene repertoires. The data suggest that these repertoires may be extremely diverse even within the same geographic location, but that particular pseudogenes are conserved across numerous strains. High-throughput sequencing should allow us to better understand the molecular epidemiology of larger-genome infectious agents, as has been achieved with viruses.

INTRACELLULAR PARASITISM, THE DRIVING FORCE OF EVOLUTION OF *LEGIONELLA PNEUMOPHILA* AND *LEGIONELLA LONGEBACHAE*

Carmen Buchrieser

Institut Pasteur, Genomes and Genetics, Paris, 75724, France

Legionella pneumophila and *Legionella longbeachae* are the etiological agents of Legionnaire's disease. They are Gram-negative bacteria present in fresh and artificial water environments but also in soil that replicate in protozoan hosts and is also found in biofilms. When aerosolized bacteria are inhaled, they are able colonize the respiratory tract, invade alveolar macrophages and replicate therein causing the disease. Replication within protozoa is essential for the survival of these bacteria. Analysis of the genome sequences of *L. pneumophila* and of *L. longbeachae* revealed the presence of an unexpected high number and variety of eukaryotic-like proteins, predicted to be involved in the exploitation of the host cellular cycle by mimicking specific eukaryotic functions. Among those are proteins containing F-Box domains, U-Box domains, Sel-1 domains, Ankyrin repeat proteins or a sphingosin-1-phospahte lyase. Eukaryotic-like proteins have also been identified in other bacterial pathogens, however a comparative analysis shows that *L. pneumophila* ranks as one of the pathogens that encodes the most and the widest variety of eukaryotic-like proteins or proteins with eukaryotic domains. Recent published results and ongoing work show that these different proteins are implicated in modulating host cell functions during the intracellular life of *L. pneumophila*. Phylogenetic analyses demonstrated that both lateral gene transfer from eukaryotic hosts and bacterial genes that became eukaryotic-like by convergent evolution, contributed to the evolution of these proteins within the species *L. pneumophila* and *L. longbeachae*.

HYPERVIRULENT *CHLAMYDIA TRACHOMATIS* CLINICAL STRAIN IS A RECOMBINANT BETWEEN LYMPHOGRANULOMA VENEREUM (L_2) AND D LINEAGES

Naraporn Somboonna[1], Raymond Wan[1], David M Ojcius[2], Matthew Pettengill[2], Sandeep Joseph[3], Alexander Chang[3], Ray Hsu[1], Timothy D Read[3], Deborah Dean[1,4]

[1]CHORI, Immunobiology and Vaccine Design, Oakland, CA, 94609, [2]Univ. of Calif. at Merced, Natural Sciences, Merced, CA, 95344, [3]Emory Univ., Infectious Diseases, Atlanta, GA, 30322, [4]Univ. of Calif. at Berkeley, Bioengineering, Berkeley, CA, 94720

Chlamydia trachomatis is an obligate intracellular bacterium that causes a diversity of severe and debilitating diseases in humans. It is the leading cause of preventable blindness in tropical developing countries and the leading cause of bacterial sexually transmitted diseases worldwide. Lymphogranuloma venereum (LGV) is caused by the LGV strains of *C. trachomatis* and represents a neglected tropical disease. Recently, sporadic and ongoing outbreaks of LGV among men who have sex with men have been documented in Europe and the US. These findings support the need for research on virulence factors associated with these organisms. Previous analyses of LGV included genome sequencing of the laboratory-adapted reference strain L2/434 and the outbreak strain L2b/UCH-1. We isolated an LGV strain, termed L2c, from a man with severe proctitis but no inguinal adenopathy. Morphologically, L2c developed non-fusing, grape-like inclusions in culture unlike other LGV strains described to date. Genome sequencing revealed L2c to consist of a community of recombinants with similar regions of genetic exchange. The consensus sequence contained a 74 kb region from *C. trachomatis* strain D and a partial, yet functional, cytotoxin gene. Indels (insertion/deletions) occurred in a promoter region for the *ftsK* gene, and in the *tarp* and *hctB* genes that encode key proteins involved in replication, intracellular inclusion formation, and histone H1-like protein activity, respectively. Analyses suggest that the indels affect protein function, supporting the clinical and *in vitro* findings for L2c. L2c represents the first genomic evidence for recombination between *C. trachomatis* strains. The finding of a heterogeneous community of recombinants suggests that L2 and D either produced a number of recombinants in the rectal mucosa of the patient or that there were a number of clonal populations of L2c recombinants circulating among the patient's core sexual group. The recombinants resemble the "quasi-species" of genetically diverse individuals frequently reported in viral populations. Given the lack of a genetic system for producing stable mutants of *C. trachomatis*, identifying naturally occurring clinical recombinants such as L2c can clarify the functions of specific genes and provide opportunities for discovering avenues for genomic manipulation.

INFERRING LEPTOSPIRAL PATHOGENICITY THROUGH WHOLE GENOME ANALYSIS

Joseph M Vinetz[1], Michael A Matthias[1], Jessica N Ricaldi[1,2]

[1]UCSD, Medicine, La Jolla, CA, 92093, [2]Univ Peruana Cayetano Heredia, AvHIMT, Lima, 15, Peru

Leptospirosis is a globally important zoonotic disease caused by diverse spirochetes in the genus *Leptospira* which is divided into pathogens, intermediates, and saprophytes. Genomic analyses of pathogens (*L. interrogans*, *L. borgpetersenii*) and a saprophyte (*L. biflexa*), have produced insights into genomic organization/reduction and genes selectively found in pathogens vs. saprophyte, but have not predicted pathogenicity mechanisms. To identify genes potentially associated with virulence, we employed two strategies: (1) comparative whole genome analysis of an intermediate *Leptospira*, the newly discovered *L. licerasiae*, from the Peruvian Amazon; and 2) identification of mutations in *L. interrogans* associated with *in vitro* attenuation by whole genome sequencing. 454 sequencing of *L. licerasiae* demonstrated a genome size (~4.2 Mbp) intermediate between that of pathogens (~4.8 Mbp) and saprophytes (~3.9 Mbp). *L. interrogans* and *L. licerasiae* share 2545 genes, of which 158 were not found in either the saprophyte, *L. biflexa*, or the strictly parasitic *L. borgpetersenii.* Pathogenic and intermediate *Leptospira* share more orthologs with each other than with saprophytes (378, 225 with predicted function, others being conserved hypothetical proteins), suggesting genes associated with virulence and/or mammalian host-pathogen interactions. Such genes include those implicated in environmental sensing and signal transduction, and cobalamin (cob operon, Vitamin B12) metabolism. Selective retention of the *cob* operon by infectious *Leptospira* suggests a role in virulence/pathogenicity, yet all sequenced leptospiral genomes encode two enzymes that require B12 for function: methylmelanoyl-CoA mutase and B12-dependent methionine synthetase, MetH. These results suggest that enhanced *in vivo* synthesis of vitamin B12 may contribute to survival within the host. To identify potential virulence genes, a highly pathogenic strain of *L. interrogans* Lai [Passage 1 (P1) LD50 < 100] was attenuated by serial passage. Comparison of P1 vs P19 Solexa-sequenced genomes demonstrated 15 non-synonymous SNPs, likely to represent pathogenicity-associated genes. These included 2 adenylate cyclases unique to pathogenic *Leptospira* with amino acid changes in essential functional domains. Despite recent microarray, transposon mutagenesis, and targeted gene deletion studies, translation of genomic information to the inference of specific leptospiral virulence mechanisms remains in its infancy. Genomic approaches to identify potential virulence-associated genes in *Leptospira* may yield new insights into pathogenetic mechanisms of leptospirosis.

TRANSCRIPTIONAL PROFILING OF *MYCOBACTERIUM TUBERCULOSIS* DRUG SENSITIVE SURVIVORS OF DRUG TREATMENT IDENTIFIES A REGULATOR OF PERSISTENCE

Brian Weinrick, William Jacobs, Jr.

Howard Hughes Medical Institute at Albert Einstein College of Medicine, Microbiology & Immunology, Bronx, NY, 10461

Persistent bacteria survive drug treatment despite lacking alleles that code for genotypic drug resistance. The phenomenon of persistence may be the basis of the need for lengthy treatment regimens to cure tuberculosis and could also contribute to the ability of the bacteria to establish a latent infection.
A basic experimental model of persistence involves treating *M. tuberculosis* cells with isoniazid, an antibiotic that provokes a lytic death of the cell. After four days of treatment, 99.9% of the population has been killed and the remaining cells are persisters; genotypically drug sensitive survivors of drug treatment. The transcriptional profile of these persister cells was compared to untreated cells. The most differentially expressed gene not affected by short-term isoniazid treatment was a putative transcription factor, *whiB6*. We created *whiB6* knock out and complemented strains and compared the transcriptional profile of the *whiB6* mutant and complemented strains with the wild type strain. Many of the differences in gene expression seen in the persister cells were absent in the *whiB6* mutant and these differences were restored with the complemented strain. We have investigated the behavior of the *whiB6* mutant in a mouse model of drug treatment, latency, and reactivation. The mutant was more effectively killed by drug treatment in the context of the mouse and may be less likely to enter a latent state.

ANALYSIS OF THE *NEISSERIA MENINGITIDIS* CAPSULE

Odile Harrison[1], Holly Bratcher[1], Keith Jolley[1], Heike Claus[2], Ulrich Vogel[2], Mike Quail[3], Stephen Bentley[3], Ian Feavers[4], Martin Maiden[1]

[1]University of Oxford, Zoology, Oxford, OX1 3PS, United Kingdom, [2]Universitat Wurzburg, Institut fur Hygiene und Mikrobiologie, Wurzburg, 97080, Germany, [3]Wellcome Trust Sanger Institute, Sequencing Unit, Cambridge, CB10 1SA , United Kingdom, [4]National Institute for Biological Standards and Control, Bacteriology, South Mimms, EN6 3QG, United Kingdom

Introduction

Capsular polysaccharide plays an important role in virulence and is the basis for immunological serogrouping. Thirteen serogroups have been described: A, B, C, D, 29E, H, I, K, L, W-135, X, Y and Z. This study explored the genetic relatedness of the capsule operon among all thirteen serogroups. It made use of the Bacterial Isolate Genome Sequence Database (BIGSdb) which is a scalable, open source, web-accessible database system.

Methods

The operon was sequenced in a representative *N. meningitidis* isolate from each serogroup. Serogroup A, B and C capsule operons were obtained from the genomes belonging to isolates FAM18 (C), 053442 (C), MC58 (B) and Z2491 (A). Sequences were deposited in the BIGSdb from which schemes for each capsule region were devised.

Results

Genes necessary for capsule expression are clustered within the capsule gene complex that comprises 5 regions (A–E). Genes in regions A, B and C are responsible for capsule synthesis, modification and transport while region D contains genes for LPS synthesis. Region E contains the *tex* gene. All of the serogroups examined had similar organisation of the capsule with regions occurring in the order D-A-C-E-D'-B. Serogroup H contained 4 genes in region A, two of which identical to *capZa* and *capZb* involved in serogroup Z capsule synthesis. Serogroups I and K had identical capsule operons with 5 genes in region A designated *capIKa-capIKe*. Three of these genes shared sequence identity with genes involved in capsule synthesis in *Pasteurella* suggesting horizontal genetic exchange. Serogroup D was found to be a serogroup C isolate.

Conclusion

Isolates with serogroup H, I, K and L capsules have been rarely reported since their description in the 1980s such that only 2 serogroup K and 8 serogroup H isolates have been deposited in the PubMLST database with no further descriptions of serogroups I and L. Sequencing of the capsule operon revealed that D was identical to C suggesting that this serogroup does not exist. It is proposed that *N. meningitidis* be classified into 12 serogroups A, B, C, 29E, H, I, K, L, W-135, X, Y and Z. All of the sequenced capsule operons have been deposited in BIGSdb and are being used as a reference set to probe for the capsule operon among *N. meningitidis* genomic DNA deposited in the database.

EMERGENCE OF A *STREPTOCOCCUS AGALACTIAE* LINEAGE HYPERVIRULENT IN FISH BY REDUCTIVE EVOLUTION

Philippe Glaser

Institut Pasteur, Evolution et Génomique Bactériennes, Paris, 75724, France

Streptococcus agalactiae is a leading cause of neonatal infections in humans. It is a commensal frequently isolated from the digestive and genitourinary tracts that could be transmitted from the mother to the newborn at birth. However, *S. agalactiae* was initially described as responsible for bovine mastitis and more recently for epidemic events of invasive diseases in fish farms. More specifically, a lineage was identified with specific phenotypic properties. One strain from this lineage, responsible for epidemic events in fish farms in Israel was initially described as a new streptococcus species named *S. difficile*.
The complete genome sequences of 8 strains of human origin are currently available. We have now sequenced the genomes of five strains responsible for infections in cold blood animals (fish and frogs) including the original *S. difficile* strain (2-22)
The three strains isolated from cold blood animals share a thermosensitive growth, a hydrophobic surface and the loss of the capacity to metabolize different sugars. They are genetically closely related with an overall polymorphism lower than 1 SNP per kb. The most remarkable feature of their genome is the massive reduction in genome size by more than 20% resulting from a large number of deletions and the loss of almost all genomic islands. Furthermore, more than 12% of the remaining genes are putative pseudogenes as a result of internal deletions, frame shift mutations and in frame stop codons. In agreement with the phenotypic properties, this gene loss affects mostly genes encoding metabolic functions, surface proteins, regulatory functions and almost all known virulence factors, the polysaccharide capsule remaining the only conserved one.
As described for human pathogens, like *Salmonella typhi* or *Bordetella pertussis*, the shift from a commensal to a professional pathogen is associated with genome reduction. The access to multiple strains from this lineage isolated in different parts of the world with different sets of deletions and pseudogenes allows to investigate its history and the molecular mechanisms underlying the genome reduction and provide a model to study increased virulence and host restriction associated to high population density of the host.

FROM GENOME SEQUENCE TO FORWARD GENETIC ANALYSIS OF SCHISTOSOMA MANSONI

Claudia Valentim[1,2], Phil LoVerde[1], Charles Criscione[3], Claudia Carvalho=Queiroz[1], Jason Tsai[4], Matt Berriman[4], Tim Anderson[2]

[1]University of Texas Health Science Center, Biochemistry & Pathology, San Antonio, TX, 78245, [2]Southwest Foundation for Biomedical Research, Genetics, San Antonio, TX, 78245, [3]Texas A&M University, Biology, , College Station, TX, 77843, [4]Wellcome Trust , Sanger Institute, Hinxton, CB10 1SA, United Kingdom

Schistosomiasis, the most important of the human helminth infections, is often described as a "neglected" disease. The publication of the Schistosoma mansoni genome sequence provides opportunities to reinvigorate research on this parasite. We have exploited the genome sequence to develop forward genetic methods (i.e. linkage mapping) to identify the genes that underlie phenotypic variation in this parasite. This approach is well suited to S. mansoni as the lifecycle can be maintained in the laboratory and clonal propagation of parasites within snails generates large numbers of genetically identical parasites. Furthermore schistosomes show heritable variation in many biomedically important traits such as drug resistance, host specificity, and virulence. To generate a genetic map and help assemble the fragmented genome sequence, we staged a genetic cross and genotyped parental, F1 and F2 parasites using microsatellite markers placed in the largest sequence scaffolds. In this way we generated a 5 cM genetic map and were able to assign and order ~70% of the scaffolds onto chromosomes. To provide proof-of-principal that linkage mapping is feasible and powerful we used the genetic map and a second cross to identify the genome region that underlies resistance to oxamniquine (OXA). Resistance to this drug has arisen in nature, has a simple recessive basis and results in ~500-fold reduction in drug sensitivity. We measured OXA-resistance by monitoring death of cultured worms following drug exposure and genotyped parents, F1 and F2 progeny using markers distributed across the genome. As expected trait segregation in the cross was consistent with recessive inheritance as F1s were sensitive and ~25% of F2 progeny were resistant. We found a strong QTL (LOD = 5.4) on chromosome 6 where microsatellite markers segregate closely with OXA resistance. The two parents of this cross have now been sequenced, simplifying fine mapping and identification of candidate genes. Successful identification of gene(s) that underlie OXA-resistance will provide insights into mode of drug action, allow development of modified compounds that kill resistant parasites, generate selectable markers for genetic manipulation, and set the stage for forward genetic analyses of a range of biomedically important traits including praziquantel resistance.

A GENOMIC APPROACH TO *TRYPANOSOMA CRUZI* LINEAGE-SPECIFIC SEROLOGY FOR CHAGAS DISEASE.

Tapan Bhattacharyya[1], Martin Llewellyn[1], Matthew Yeo[1], Michael Lewis[1], Andrew Falconar[1], Alejandro Luquetti[2], Michael A Miles[1]

[1]London School of Hygiene & Tropical Medicine, Dept. Infectious & Tropical Diseases, London, WC1E 7HT , United Kingdom, [2]Universidade Federal de Goias, Faculdade de Medicina, Goiania, CEP 74605-020, Brazil

Chagas disease, endemic to the Americas, is caused by life-long infection with the parasitic protozoan *Trypanosoma cruzi*. Genetically diverse, *T. cruzi* is classified into the intra-species lineages TcI-TcVI, with disparate ecologies and geographical distributions. Disease outcome may be linked to parasite lineage, complicated by mixed infections and divergent tissue tropism. Lineage-specific serology would be an important new tool in understanding Chagas disease epidemiology.

The *T. cruzi* trypomastigote small surface antigen (TSSA) has previously been described to serologically identify infection by TcII-VI, as distinct from TcI, on the basis of a biallelic *TSSA* resulting in two isoforms of the mature protein containing distinct epitopes. However, we have recently shown by analysis of *TSSA* nucleotide and predicted amino acid sequences across a panel of reference biological clones representing all lineages, that TSSA epitope polymorphism is greater than previously described. The epitope considered to be serologically characteristic of TcII-VI is restricted to TcII, V, and VI, while epitope forms found in TcIII and TcIV share key features with TcI. Furthermore, *TSSA* sequences inferred greater phylogenetic affinities of TcIII and TcIV to TcI than to TcII, V, or VI, and a high ratio of non-synonymous/synonymous nucleotide substitutions (ω=1.233) suggests that *TSSA* has been under positive selection pressure by the vertebrate host.

By using synthetic peptides based on these lineage-specific TSSA epitopes, we are currently developing an ELISA for serological studies of an individual's history of infection by different *T. cruzi* lineages. Early results using this method have identified evidence of TcII infection in patient sera from Brazil. Additionally, an *in silico* 'comparative genomics' approach to the recently sequenced genomes of *T. cruzi* lineage reference strains demonstrates a new means to identify novel lineage-specific antigens, which has broad applicability to neglected infectious diseases.

Supported by EC ChagasEpiNet.

MAINTENANCE AND EVOLUTION OF PLASMODIUM FALCIPARUM GENETIC DIVERSITY SUSTAINING A LOW TRANSMISSION MALARIA ENDEMIC IN THE PERUVIAN AMAZON

OraLee H Branch[1], Patrick L Sutton[1], Juan Carlos Castro-Gomez[2], Carmen R Barnes[3], Gisely Hijar-Guerra[3]

[1]New York University, Medical Parasitology, New York, NY, 10016, [2]Universidad Nacional Amazonia Peruana, Laboratorio Investigaciones Productos Naturales Antiparasitarios de la Amazonia, Iquitos, 065, Peru, [3]Instituto Nacional de Salud, Laboratorio de Biotecnologia y Biologia Molecular, Lima, 001, Peru

P. falciparum entered into the Peruvian Amazon in 1994, sparking an epidemic between 1995 and 1998. Since 2000, there has been sustained, low *P. falciparum* transmission. The Malaria Immunology and Genetics in the Amazon project has longitudinally followed members of the community of Zungarococha (N=1945, 4 villages) with active household and health center-based visits each year since 2003. We examined parasite population structure and traced the parasite genetic diversity throughout time and space. We typed infections occurring from 2003-2007 using 14 microsatellite (MS) markers scattered across 9 different chromosomes. Despite low transmission, there was considerable genetic diversity which we compared to other geographic regions. This study included 182 different haplotypes from 302 parasites in 217 discrete infections. Population structure was characterized by two different methods. The first method broadly grouped haplotypes into clusters of phylogenetically related clones, while the second method grouped haplotypes pairwise into haplotype families (hapfams) by considering haplotypes with three or less loci differences as members of the same family. In this population, five different clusters of related clones and 34 different hapfams were identified. Hapfams with the greatest amount of within-family diversity (polymorphic hapfams) were consistently redetected over time, relative to the monomorphic hapfams. Controlling for haplotype frequency and possible mutation, the monomorphic hapfams were shown to go extinct. Over time the diversity was maintained, within a village the hapfams persisted, but with one or two differences (polymorphisms). The population structure, genetic diversity, appearance/disappearance of the different haplotypes over time has provided a genome-wide, 'real-time' perspective into how *P. falciparum* parasites are maintained in endemic transmission. We relate this to the high level of asymptomatic malaria infections in this population, speculating that the MS markers might be linked to antigen genes and that the immunity in this population is selecting for variation across the genome.

SCHISTOSOMA MANSONI DATA INTEGRATION AND MINING. CHARACTERIZATION OF NEW DRUG TARGETS.

Adhemar Zerlotini[2], Laila A Nahum[1], Luiza A Almeida[1], Larissa L Silva[1], Mariana C Simões[2], Fabiano S Pais[2], Raymond Pierce[3], Guilherme Oliveira[1,2]

[1]FIOCRUZ-Minas, Genomics and Computational Biology, Belo Horizonte, 30190-002, Brazil, [2]FIOCRUZ-Minas, CEBio, Belo Horizonte, 30190-110, Brazil, [3]Insitut Pasteur de Lille, Centre d'Infection et d'Immunité de Lille, Inserm U1019 – CNRS UMR 8204, Lille, BP 245 - 59019, France

SchistoDB (www.SchistoDB.net) is a genomic database for Schistosoma mansoni. The database uses the Genomics Unified Schema, GUS, and currently contains S. mansoni sequence data (Genomic, EST) and ~13 thousand automated gene predictions. SchistoDB offers a variety of tools including BLAST, protein motif searches, keyword searches of pre-computed BLAST results, GO assignments and protein family information that can be queried with Boolean logic. Data from a number of microarray, SAGE and ChipSeq were also uploaded. Integrated into SchistoDB are SchistoCyc and KEGG DRUG. SchistoCyc is a prediction of the metabolic pathways of the organism based on the genomic information. Genomic data were used to computationally predict 112 metabolic pathways. Reactions that might play a crucial role in parasite metabolism were identified. In order to enhance the annotation of genes we have reconstructed the evolutionary histories of all proteins encoded in the S. mansoni genome (phylome). The phylome was performed using an automatic pipeline as implemented in PhylomeDB. We analyzed 8.818 phylogenetic trees of 13.285 predicted proteins and homologs in 16 other organisms. We could transfer functional annotations from GO to 5.587 S. mansoni proteins (956 had no annotation). This approach provided us with a genome-wide view of S. mansoni, indicating gain, loss and gene duplication in this parasite. We have used the database to characterize two groups of proteins, Eukaryotic protein kinases (ePKs) and histone modifying enzymes (HMEs), as possible drug targets. ePKs play a central role in mediating signal transduction through complex networks. The analysis of the S. mansoni genomic data revealed 252 putative ePKs, which corresponds to 1.9% of the parasite predicted proteome. For this purpose, amino acid sequences corresponding to the conserved catalytic domain of ePKs were aligned by MAFFT and an HMM model generated to scan the parasite predicted proteome. Phylogenetic analysis based on a distance method was implemented in PHYLIP. Our results showed the functional classification of the S. mansoni ePKs into nine kinase groups (AGC, CaMK, CK1, CMGC, STE, RGC, TK, TKL, and Others). A similar approach was used for the identification of 67 HMEs (14 histone deacetylases, 12 histone acetyltransferases, 23 lysine methyltransferases, 7 arginine methyltransferases and 11 histone demethylases). Selected genes are now being confirmed by sequencing and coded proteins used for the identification of inhibitors by computational and experimental approaches.

COMPLEX MALARIA EPIDEMIOLOGY AND EVOLUTIONARY GENOMIC UNDERSTANDING FROM INDIAN MALARIA PARASITES, VECTORS AND HUMANS

Aparup Das

National Institute of Malaria Research, Evolutionary Genomics and Bioinformatics Laboratory, New Delhi, 110077, India

Malaria is one of the highly endemic infectious diseases in India with severe public health hazards, as it alone infects over 1.5 million Indians every year. Unlike other endemic countries, malaria in India presents much complex epidemiological patterns. These complexities chiefly been implicated to eco-climatic conditions that favors vector survival and growth, almost equal distribution of the two most-widespread species of human malaria parasites (*Plasmodium falciparum* and *P. vivax*), high proportion of mixed parasite species infections and malaria susceptible humans, among several others. Majority of the basic biological and epidemiological understandings related to malaria infection are still in dearth in India, which debars to formulate effective malaria intervention strategies. In this concern, recent advancements in genomic technologies offer excellent opportunities to unravel genetic mysteries of malaria infections. We have recently initiated evolutionary genomic research in Indian malaria parasites, vectors and human hosts. While results on Indian malaria vector phylogenomics and population genomics provide initial insights into evolutionary understanding, population genomics of both *P. falciparum* and *P. vivax* reflect genome variability and population structure in India. Further, important insights into evolutionary genetic outlines of antimalarial resistance in *P. falciparum* have also been gained. Additionally, human population genomics and genetic variations in drug metabolizing and malaria susceptibility genes have provided primary understanding on complexities of malaria infections in India. Results from all these studies will be presented.

ORIGINS AND EVOLUTION OF VARIABLE SURFACE ANTIGENS IN AFRICAN TRYPANOSOMES

Andrew Jackson[1], Andrew Berry[1], John Gamble[1], Nicola Corton[1], Heidi Hauser[1], Martin Aslett[1], Dave Barry[2], Christiane Hertz-Fowler[1], Matthew Berriman[1]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, United Kingdom, [2]Wellcome Trust Centre for Molecular Parasitology, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow, G12 8TA, United Kingdom

The expression of Variable Surface Glycoproteins (VSG) on the cell surfaces of African trypanosomes (*Trypanosoma brucei, T. congolense and T. vivax*) allows these parasites to evade host immune responses, and is a textbook example of antigenic variation. Our understanding of the genetic diversity of VSG, regulation of their expression, and placement within the genome comes mainly from *T. brucei*, where a single VSG is expressed from telomeric expression sites and periodically substituted through ectopic gene conversion. This results in high levels of mosaicism among *T. brucei* VSG that doubtless helps the parasite to stay ahead of host immunity. We have sequenced the genomes of *T. congolense* and *T. vivax*, two closely related species, to compare and contrast the molecular structure, genetic diversity, and genomic architecture of VSG in all three species. We identify a common structural signature to VSG from all species, which each species has modified in different ways. Most notably in *T. brucei*, the VSG has become substantially larger through the addition of W-based and C-based repetitive regions at 5' and 3' ends respectively. While all species have comparable numbers of genes, *T. brucei* and *T. vivax* VSG are more diverse than *T. congolense* VSG, and all three species display distinct phylogenetic structures that indicate a fundamentally different evolutionary dynamic. We find much greater evidence for recombination among *T. brucei* VSG than either other species, which probably accounts for the lack of phylogenetic structure in the *T. brucei* VSG phylogeny. From the small number of *T. congolense* and *T. vivax* telomeres assembly, we find no evidence for an expression site (as defined in *T. brucei*), and most gene components of the *T. brucei* expression site are specific to this species. Therefore, the canonical VSG expressed by *T. brucei* has been derived relative to the ancestral surface glycoprotein with respect to its structure, its genomic position and its evolutionary dynamic. *T. brucei* generates greater diversity among its surface antigens by recombination, perhaps due to a specialised telomeric expression site that is apparently absent in other species.

THE SRS SUPERFAMILY OF *TOXOPLASMA GONDII* IS LARGE, DIVERSE AND DYNAMIC

James Wasmuth[1], Martin Boulanger[2], Michael Grigg[3], John Parkinson[1]

[1]Hospital for Sick Children, Molecular Structure and Function, Toronto, M5T 1Z3, Canada, [2]University of Victoria, Biochemistry and Microbiology, Victoria, V8P 5C2, Canada, [3]National Institutes of Health, Laboratory of Parasitic Diseases, Bethesda, MA, 20892

Toxoplasma gondii is a highly successful parasite; it is able to infect almost any nucleated cell from any warm blooded animal. This ubiquity means that the parasite must be able to recognise a large range of host cell receptors and evade a variety of immune mechanisms, processes that may involve cell surface proteins. Among the most abundant is SAG1, now known to be part of a large superfamily of SAG1-related sequences (SRS), which typically consist of one or two SRS-domains. Previous analyses suggested *T. gondii* possess a complement of 160 SRS genes, however many have been lost in recent releases of the genome. Interest in the SRS superfamily continues; while SAG1 is immunogenic and SAG1 knockouts are less virulent, the detailed roles of other members are still unknown. The release of improved genome assemblies, together with the publication of four 3D structures provide an excellent opportunity to better understand into the function and evolution of the SRS proteins.

Using current models of the SRS domain, we searched the protein sets of three *T. gondii* strains. The resulting 427 domains (from 239 proteins) were clustered, allowing the definition of eight distinct domain families. Further searches were made with family-specific HMMs, producing a final set of ~100 SRS proteins and 35-70 pseudogenes per strain. Family pairings in multi-domain proteins are not random, but demonstrate strong associations between particular families. Mapping the SRS genes onto the genome, showed that the majority are present in loci of up to ten genes. These loci are not predominantly telomeric, as observed in other apicomplexan families of surface proteins. Given a phylogenetic reconstruction of 30 two-domain SRS genes, the hypothesis is that the whole locus duplication is followed by concerted evolution; where inter-locus recombination makes each locus appear monophyletic.

Conservation profiles of each domain family were mapped onto the available 3D structures. The majority of the family defining residues are found on two exposed β-sheets. Additionally, in two domain proteins, the surface of the domain distal to the membrane is hypervariable. These diverse regions may underlie different functions of each family.

Taken together, these results highlight key features of SRS evolution and how patterns of diversity may have a consequence on their structure and function.

Participant List

Dr. Olufemi Adedeji
University of Ibadan
oluadedeji2001@yahoo.com

Dr. Elena Ambrosino
Maastricht University
e.ambrosino@phg.unimaas.nl

Dr. Tim Anderson
Southwest Foundation for Biomedical Research
tanderso@sfbrgenetics.org

Dr. Bjorn Andersson
Karolinska Institutet
bjorn.andersson@ki.se

Dr. Cesar Arias
University of Texas Medical School
cesar.arias@uth.tmc.edu

Ms. Gauri Awasthi
National Institute of Malaria Research
gauri84@yahoo.com

Ms. Kate Baker
University of Cambridge
kate.s.baker@googlemail.com

Ms. Kate Baker
University of Cambridge
kate.s.baker@gmail.com

Dr. Anthony Barbet
University of Florida
barbet@ufl.edu

Dr. Philip Bejon
KEMRI Wellcome Collaborative Research Program
pbejon@kilifi.kemri-wellcome.org

Dr. Julia Bennett
University of Oxford
julia.bennett@zoo.ox.ac.uk

Dr. Stephen Bentley
Wellcome Trust Sanger Institute
sdb@sanger.ac.uk

Dr. Matthew Berriman
The Wellcome Trust Sanger Institute
mb4@sanger.ac.uk

Mr. Nicholas Bexfield
University of Cambridge
nb289@cam.ac.uk

Mr. Tapan Bhattacharyya
London School of Hygiene and Tropical Medicine
Tapan.bhattacharyya@lshtm.ac.uk

Dr. Rory Bowden
University of Oxford
bowden@stats.ox.ac.uk

Ms. Katarzyna Bozek
Max Planck Institute for Informatics
bozek@mpi-inf.mpg.de

Dr. OraLee Branch
NYU Medical School
oralee.branch@nyumc.org

Dr. Charles Brockhouse
Creighton University
charlesbrockhouse@creighton.edu

Dr. Angela Brueggemann
University of Oxford
angela.brueggemann@zoo.ox.ac.uk

Dr. Carmen Buchrieser
Institut Pasteur
cbuch@pasteur.fr

Prof. Jane Carlton
New York University School of Medicine
jane.carlton@med.nyu.edu

Dr. Alison Cody
University of Oxford
alison.cody@zoo.ox.ac.uk

Dr. David Conway
MRC Unit, The Gambia
dconway@mrc.gm

Dr. Madeleine Cule
University of Oxford
cule@stats.ox.ac.uk

Dr. Craig Cummings
Life Technologies
craig.cummings@lifetech.com

Dr. Aparup Das
National Institute of Malaria Research (ICMR)
aparup@mrcindia.org

Dr. Nick Day
Mahidol Oxford Tropical Medicine Research Unit (MORU)
nickd@tropmedres,ac

Prof. Deborah Dean
UC Berkeley and Children's Hospital Oakland
ddean@chori.org

Dr. Lovorka Degoricija
Life Technologies
lavorka.degoricija@lifetech.com

Dr. Elie Dolgin
Nature Medicine
e.dolgin@us.nature.com

Mr. Christopher Ford
Harvard School of Public Health
cford@hsph.harvard.edu

Prof. Christophe Fraser
Imperial College London
c.fraser@imperial.ac.uk

Dr. W. Florian Fricke
Institute for Genome Sciences
wffricke@som.umaryland.edu

Dr. Manohar Furtado
Applied Biosystems
manohar.furtado@lifetech.com

Dr. Yoshikazu Furuta
University of Tokyo
yfuruta@ims.u-tokyo.ac.jp

Ms. Kathrin Geyer
Aberystwyth University
kgg08@aber.ac.uk

Dr. Philippe Glaser
Institut Pasteur
pglaser@pasteur.fr

Dr. Tanya Golubchik
University of Oxford
golubchi@stats.ox.ac.uk

Ms. Lucia Gonzales
University of Gothenburg
lucia.gonzales@microbio.gu.se

Mr. Gabriel Gonzalez
Hokkaido University
gabrielsnake@gmail.com

Ms. Bhavna Gupta
National Institute of Malaria Research
bhavnaguptaster@gmail.com

Ms. Tran Ha
Hanoi University of Science
tranthuha400@yahoo.com

Dr. Rosalind Harding
Oxford University
rosalind.harding@zoo.ox.ac.uk

Dr. Odile Harrison
Oxford University
odile.harrison@zoo.ox.ac.uk

Dr. Matthew Henn
Broad Institute
mhenn@broadinstitute.org

Dr. Rob Heyderman
Malawi-Liverpool/Wellcome Trust
r.heyderman@liverpool.ac.uk

Dr. Vladimir Holec
Comenius University, Medical Faculty
mudr.holec@gmail.com

Dr. Kathryn Holt
University of Melbourne
kholt@unimelb.edu.au

Dr. Muhammad Idrees
Punjab University
idrees.khan96@yahoo.com

Dr. Michael Inouye
Walter and Eliza Hall Institute
inouye@wehi.edu.au

Ms. Camilla Ip
University of Oxford
camilla.ip@stats.ox.ac.uk

Dr. Andrew Jackson
Wellcome Trust Sanger Institute
aj4@sanger.ac.uk

Dr. John Jerome
Michigan State University
jeromejo@msu.edu

Dr. Aaron Jex
University of Melbourne
ajex@unimelb.edu.au

Dr. Keith Jolley
University of Oxford, UK
keith.jolley@zoo.ox.ac.uk

Dr. Paul Keim
NAU/TGEN
Paul.Keim@nau.edu

Prof. Paul Kellam
Wellcome Trust Sanger Institute
pk5@sanger.ac.uk

Dr. Laura Kramer
SUNY Albany, Wadsworth Center
kramer@wadsworth.org

Dr. Kristin Kremer
RIVM
kristin.kremer@rivm.nl

Prof. Dominic Kwiatkowski
Oxford University / Sanger Institute
dominic@well.ox.ac.uk

Dr. Adam Lauring
University of California, San Francisco
Adam.Lauring@ucsf.edu

Dr. Ngo Lien
Hanoi University of Science
ngogiangl@yahoo.com

Dr. Claudia Lupp
Nature
c.lupp@nature.com

Dr. Matilda Nicklasson
University of Gothenburg
matilda.nicklasson@microbio.gu.se

Ms. Chinyere Okoro
Wellcome Trust Sanger Institute
co1@sanger.ac.uk

Dr. Guilherme Oliveira
FIOCRUZ
oliveira@cpqrr.fiocruz.br

Prof. Subrat Panda
All India Institute of Medical Sciences
subrat@gmail.com

Dr. Craig Parker
USDA, Agricultural Research Service
craig.parker@ars.usda.gov

Dr. Julian Parkhill
The Wellcome Trust Sanger Institute
parkhill@sanger.ac.uk

Dr. David Rasko
Institute for Genome Sciences
drasko@som.umaryland.edu

Dr. Jason Sahl
Institute for Genome Sciences
jsahl@som.umaryland.edu

Dr. Samuel Sheppard
Univeristy of Oxford
samuel.sheppard@zoo.ox.ac.uk

Dr. Zhengli Shi
Chinese Academy of Sciences
zlshi@wh.iov.cn

Dr. Cameron Simmons
Hospital For Tropical Diseases
csimmons@oucru.org

Dr. Noel Smith
University of Sussex
noel@sussex.ac.uk

Dr. David Stewart
Cold Spring Harbor Laboratory
stewart@cshl.edu

Mr. Ryan Tewhey
The Scripps Translational Science Institute
rtewhey@scripps.edu

Dr. Judith Torimiro
Faculty of Medicine and Biomedical Sciences
jtorimiro@yahoo.co.uk

Dr. Ronald Van Rij
Radboud UMC
r.vanrij@ncmls.ru.nl

Dr. Joseph Vinetz
University of California San Diego
jvinetz@ucsd.edu

Dr. Richard Vipond
Health Protection Agency
richard.vipond@hpa.org.uk

Prof. Hanzhong Wang
Wuhan Institute of Virology, Chinese Academy of Sc
wanghz@wh.iov.cn

Dr. James Wasmuth
Hospital Sick Children
besnoitia@gmail.com

Dr. Brian Weinrick
HHMI at Albert Einstein College of Medicine
brian.weinrick@einstein.yu.edu

Dr. George Weinstock
Washington University School of Medicine
geowei@mac.com

Dr. Aaron White
University of Saskatchewan
Aaron.White@usask.ca

Dr. Daniel Wilson
University of Oxford
daniel.wilson@ndm.ox.ac.uk

Ms. Kelly Wyres
University of Oxford
kelly.wyres@zoo.ox.ac.uk

Dr. Neil Young
The University of Melbourne
nyoung@unimelb.edu.au

Prof. Vladimir Zajac
Cancer Research Institute
vladimir.zajac@savba.sk