

The pathogen programme uses a combination of genotypic and phenotypic approaches to characterise bacteria, eukaryotic parasites (including the malaria parasite and helminths) and viruses that affect human health. Key aims are to identify the full extent of genetic variation in disease-causing organisms and to identify variants that influence their spread or interaction with human and veterinary hosts. Complementary studies investigate variation in the host that affect interactions with pathogens.

- Parasite genomics
- Experimental genetics of rodent malaria
- Pathogen biology
- Virus genomics
- Malaria genomics
- Understanding pathogen diversity
- Malaria programme



Dave Gaufking

Clostridium difficile attaching to the mucosal lining of hamster caecum.

Bacteria

Our main focus is on understanding the population structure of a variety of pathogens, in order to identify the genetic basis of traits such as virulence or patterns of transmission. We are using next-generation technologies to determine whole-genome sequences of pools of microorganisms representative of the global population or of a particular subset circulating in a local area. Phylogenetic analysis can then provide insight into the natural history and evolution of microbial populations, and reveal the genetic basis of key phenotypic traits such as antibiotic resistance or virulence. As examples, we work on *Salmonella enterica* (particularly serovars Typhi, Paratyphi A and Typhimurium), *Vibrio cholerae*, *Shigella sonnei*, *Staphylococcus aureus* (hospital clades), *Clostridium difficile* and *Bordetella pertussis*.



Strangles at birth

The streptococcus causing the horse disease strangles is exchanging genes with both human and horse pathogens.

Strangles is a common and potentially deadly respiratory tract infection in horses. It is caused by *Streptococcus equi*, which is thought to have evolved from another horse pathogen, *S. zooepidemicus*. Both are closely related to the human pathogen *Streptococcus pyogenes*, responsible for a range of conditions from sore throats to scarlet fever and necrotising fasciitis.

A comparison of these three genome sequences, carried out by Matt Holden, Julian Parkhill and colleagues, has shed light on the genetic changes that have enabled *S. equi* to occupy its new environmental niche. As seen in other pathogens that have taken on a more restricted niche, *S. equi* has lost a number of genes. It has also gained, via bacteriophage, at least three sets of genes that support its new lifestyle and promote more severe disease.

The comparative analysis also suggests that the three streptococci are still exchanging phage, raising the possibility that novel genes could be transferred to the human pathogen.

Working closely with the Sanger Institute's *S. equi* genome analysis team, colleagues at the Animal Health Trust in Newmarket have developed a new diagnostic test for strangles and a vaccine which is currently undergoing trials.

Holden MT et al. Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS Pathog.* 2009 Mar;5(3):e1000346. PMID: 19325880

Faculty members

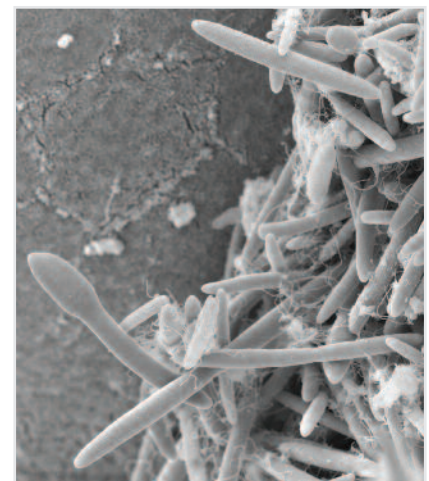
Gordon Dougan, Head
Matthew Berriman
Oliver Billker
Paul Kellam
Dominic Kwiatkowski
Julian Parkhill
Julian Rayner

An early indicator of the power of this approach was our work on the evolution of virulence and antibiotic resistance in *S. Typhi* (a cause of human typhoid). Here we were able to show that previously unrecognised lineages of *S. Typhi* present potential pathotypes. Further antibiotic-resistant isolates fell into a particular expanding haplotype/genotype.

Phylogenetic information can also be used to discriminate between isolates circulating in a particular geographical region. Working with external collaborators, we have been able to map such 'haplotyped' isolates back onto geographical areas within cities such as Kathmandu and Kolkata. This has allowed us to identify potential routes responsible for the spread of infection, and suggest vaccination strategies to interrupt transmission.

Genetic approaches are increasingly being complemented by transcriptomic studies. Gene expression has traditionally been assessed by microarray analysis, but new sequencing approaches now enable entire RNA populations to be sequenced directly. We have developed a novel approach (single strand RNA sequencing) that allows the transcriptome to be investigated in a strand-specific manner. Applied to *S. Typhi*, and other pathogens the approach generated an abundance of new information, revealing new transcribed sequences, confirming predicted coding sequences and identifying a novel population of small RNA molecules.

As well as transcriptomics, the Sanger Institute's mass spectrometry facility can extend analysis from the RNA to the protein level. A novel high-throughput phenotyping platform, Biolog, has been used to identify novel metabolic traits in wild-type or mutant bacteria. For example, we have been able to characterise the full proteome of spores produced by the healthcare-acquired pathogen *Clostridium difficile*. We have also used Illumina-based sequencing to carry out whole-genome screens for genes essential for life or for particular virulence traits. We sequence high-density pools of randomly mutated bacteria and produce detailed chromosomal maps of mutations affecting survival or virulence. This approach has so far been applied to *Salmonella* and *Clostridium* species.



C. difficile spore forming (bottom left) amid a raft of vegetative cells and the mucosal epithelial cells (hexagonal) of the caecum.

Dave Goulding

Parasites

Parasite work is now focused on 'trityps' (trypanosomes and *Leishmania*) together with several helminth genomes. A recent landmark has been the publication of a schistosome genome (see box: A fluke of nature). *Schistosoma mansoni* is responsible for the neglected tropical disease schistosomiasis, which affects 210 million people worldwide. Sequence analysis revealed a number of peculiarities in the parasite genome. As well as identifying potential new drug targets, the sequence also offers insights into early events in the evolution of animals, including the development of a bilateral body plan and organs.

Studies on the human genetics of malaria encompass one of the world's largest malaria projects, involving the genetic analysis of around 50 000 human samples collected globally, with a focus on Africa. The power of the study should facilitate the identification of human genes influencing susceptibility to malaria disease. It is being organised through the Malaria Genomic Epidemiology Network, an international consortium led by Dominic Kwiatkowski.

Genetic work is supported by phenotypic studies, particularly on the mechanisms by which the parasite invades human erythrocytes and the impact that natural genetic variation in the host has on invasion. In collaboration with Gavin Wright's group, protein–protein interactions between large numbers of *Plasmodium* and erythrocyte proteins are being studied to identify key interactions that could be exploited to block infection.

Studies on diversity within the malaria parasite complement those on the human host. Whole-genome sequencing of multiple *P. falciparum* isolates is shedding light on parasite population structure and the impact of recombination on parasite populations. The results are helping to map the spatial distribution of parasite variants, laying the foundation of a 'geobiological' approach to malaria. To this end, software and informatics tools are being developed to link genotype, phenotype, environmental and clinical data.



A research assistant at Papua New Guinea Institute for Medical Research, extracting DNA from samples taken from children with severe malaria.



A fluke of nature

The genome sequence of the blood fluke *Schistosoma mansoni* is of evolutionary as well as medical significance.

Schistosoma mansoni affects more than 200 million people in 76 countries, killing 280 000 people in sub-Saharan Africa alone. The genome sequence of *S. mansoni*, completed by Matt Berriman and colleagues, not only reveals a wealth of potential leads for drug development but also sheds light on a key stage of animal evolution.

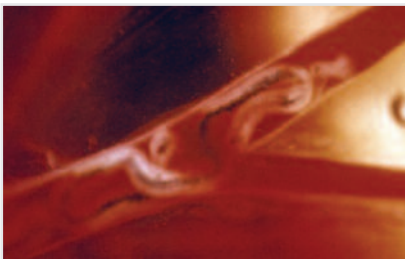
Genomically, *S. mansoni* shows some peculiarities. Its 12 000 genes show an unusual intron size distribution (small 5' introns, much larger 3' introns) and new families of micro-exon genes that undergo frequent alternative splicing, probably contributing to protein variability.

The genome encodes a wide range of proteins typically targeted by drugs, such as ion channels and membrane-bound receptors. The sequence also reveals the full complement of kinases and proteins involved in degradation, also popular drug targets. A family of novel neuropeptides not seen in vertebrates may provide further drug development options. Bioinformatic analysis positioned putative protein products in metabolic pathways, identifying potential 'chokepoints' – targets for pathway-blocking drugs.

A promising therapeutic strategy is to identify existing drugs that might have activity on pathogens. A database search identified several schistosome relatives of 'druggable' human or pathogen proteins, while a number of parasite proteins resemble the targets of drugs already being used in humans.

Comparisons with the sea anemone *Nematostella vectensis* provide clues to the genetic innovations responsible for features such as the three-layered body plan, parasitism and the formation of organs – *S. mansoni*'s tissues resemble simple organs. Similarly, comparisons with more complex organisms provide clues to the development of more complex features.

Berriman M et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009 Jul 16;460(7253):352-8. PMID: 19606141



Wellcome Library, London

Schistosoma mansoni adult worms in blood vessel.

The mouse malaria project is focusing on mouse-adapted parasites such as *P. berghei*. The parasite is being studied in both the insect and mammalian host, using genetic approaches to identify key infection pathways. The intention is to set up a large-scale gene knockout programme in the parasite and complement this with studies using knockout mice from the Mouse Genetics Programme. Small-molecule screening approaches are being used (with collaborators at Imperial College London) to test potentially druggable targets identified in the parasite.

Viruses

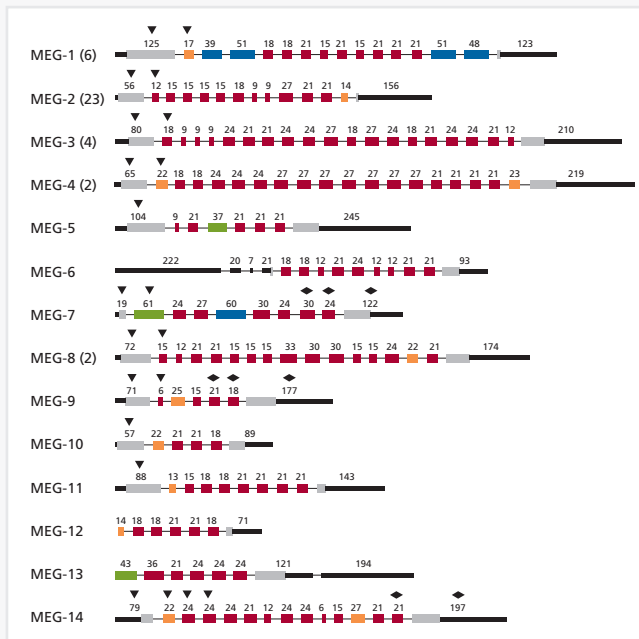
With the recruitment of Paul Kellam, we are now well-positioned to investigate the population structure of several different viruses using high-throughput sequencing. We have established an influenza virus sequencing pipeline that can deal with hundreds of viruses at a whole genome level and could be used to study influenza virus genomes as the virus moves through human or zoonotic populations. We plan to establish such pipelines for a number of viruses.

Deep sequencing approaches are also being used to look at the impact of natural variation within hosts or in the face of selective pressures such as vaccine-induced immunity. A programme of this type on equine influenza infections has recently been completed and we were able to monitor key mutational changes in the viral haemagglutinin protein as the virus grew within or moved between hosts.

Forward look

Sequencing is being used in multiple ways to support functional studies on bacterial pathogens. Our emphasis continues to shift from draft sequencing and annotation to more functional analysis. Bioinformatic support is still essential and new tools are under development or consideration including novel methods to link geobiology to genotype and, through collaboration with the European Bioinformatics Institute, to curate microbial genomes through Ensembl.

We are studying a range of eukaryotic parasites, particularly the malaria parasite and parasitic worms, and their interactions with mammalian hosts.



Schistosoma mansoni micro-exon genes. The micro-exons (red) have lengths that are divisible by three bases and are therefore skipped extensively during transcription.

Berriman M et al. Nature 2009 460: 352
doi:10.1038/nature08160

By genomic approaches, we aim to gain fundamental insights into parasite biology, identify possible targets for anti-infective agents and develop community resources for neglected tropical diseases. High-quality reference genomes underpin comparative studies, to identify either candidate genes or pathways that could account for major biological differences, or conserved features, such as common regulatory sequences. Deeper comparisons of isolates or strains may reveal loci related to key phenotypes, such as drug resistance.

Most of our *de novo* sequencing efforts are currently focused on parasitic helminths. Over the last year, we have begun large-scale sequence production, allowing us to evaluate multiple sequencing platforms. We are also using Illumina technology to sequence transcriptomes directly, to identify every expressed gene in the genome and to provide quantitative data on gene expression throughout a parasite's life cycle. In addition, we have developed software to automate techniques needed to improve sequence accuracy.

We have published reference genome sequences for *Plasmodium knowlesi* and *Schistosoma mansoni* (the first from our helminth genomes initiative). We have also sequenced the full repertoire of subtelomeric regions from a single strain of *Trypanosoma brucei*, which are under-represented in genome sequences but important in immune evasion.

Supporting studies in *Plasmodium* diversity, we have released new versions of the genome annotation of *P. falciparum* and markedly improved annotated genome sequences for two rodent malaria species. These data are made available through the GeneDB database, which has now been adapted to allow remote curation and tighter integration with other parasite resources globally.



Schistosoma mansoni trematodes under a low magnification of 78X and stained using an indirect fluorescent antibody.



Wellcome Library, London

The draft genome of the blood fluke *Schistosoma mansoni* reveals oddities such as genes with tiny 5' introns or multiple tiny exons, every one of which is alternatively spliced.

Berriman M et al. The genome of the blood fluke *Schistosoma mansoni*. Nature 460: 352 (2009)

Systematic sequencing of subtelomeric regions in the Lister 427 strain of *Trypanosoma brucei* sheds light on genomic sites where variant surface antigens are expressed.

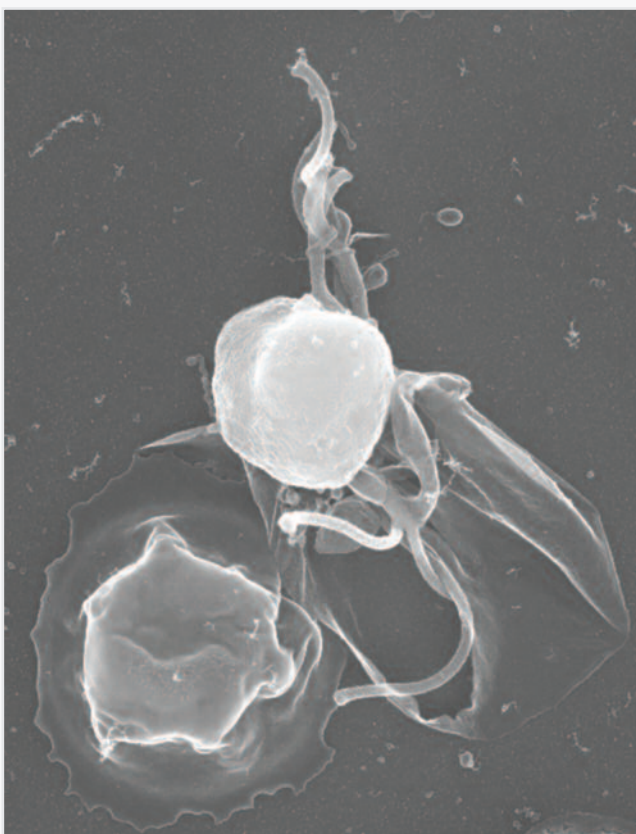
Hertz-Fowler C et al. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. PLoS ONE 3: e3527 (2008)

Genome sequence of the fifth human malaria parasite *Plasmodium knowlesi* reveals an unusual organisation of genes encoding host-interacting factors and a novel type of molecular mimicry.

Pain A et al. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. Nature 455: 799 (2008)

We aim to identify critical molecular mechanisms in malaria, by studying the effects in rodent models of natural and engineered variation in the parasite and host genome.

A male transmission stage of *Plasmodium berghei*, which has just emerged from a red blood cell and is now releasing sperm-like gametes. Billker O et al. Cell 2004 117(4): 503-14 PMID: 15137943



Copyright © 2004 by Cell Press

A cornerstone of our work is our ability to study the entire life cycle of rodent malaria species in mice and mosquitoes. To exploit this capability, we are collaborating with other Sanger Institute teams to generate gene targeting vectors. We are particularly interested in developing robust technologies that will allow us to generate new resources for the malaria research community.

This year we completed a pilot for large-scale, systematic gene targeting in *Plasmodium berghei*, focusing on the parasite's complement of protein kinase genes – work begun in our Imperial College laboratory, and continued this year in collaboration with Rita Tewari (now University of Nottingham). Phenotyping of more than 20 mutants has led to the discovery of many new regulators of parasite maturation and life cycle transitions. This work has also highlighted apparently essential protein kinases, possible targets for rational drug development.

Of particular interest is a family of closely related protein kinases that have different essential functions, and are therefore possible targets for a multi-kinase inhibitor. In collaboration with Jane Endicott in Oxford and the Drug Development Institute at Imperial College London, we have begun structural biology and drug discovery programmes with selected *P. falciparum* members of this kinase family.

With Dominique Soldati in Geneva and Manuel Llinas at Princeton University, we have produced a prototype conditional gene expression system, which could provide a much-needed tool for studies of essential genes in malaria parasites. We are evaluating its use in *P. berghei*.

We are also developing simple, robust and scalable assays for routinely challenging knockout mice with rodent malaria. We will expand these activities, to identify host genes that modulate susceptibility to malaria resistance and pathology.



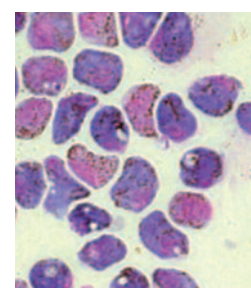
Wellcome Library, London

An overview of the current state of protein kinase research in malaria.

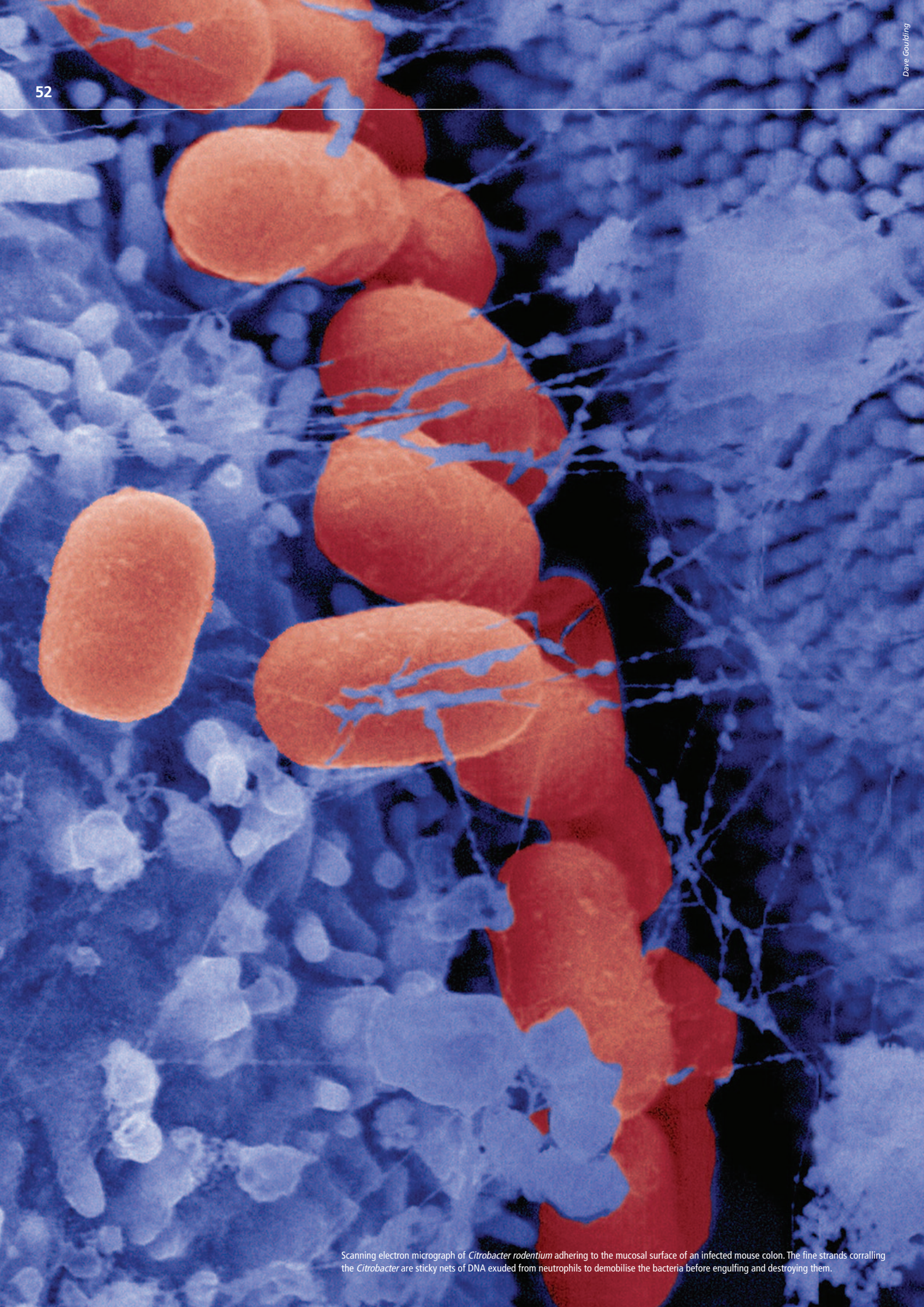
Doerig C et al. Protein kinases of malaria parasites: an update. Trends Parasitol 24: 570 (2008)

A review of calcium-dependent signalling in the malaria parasite and its relatives.

Billker O et al. D Calcium-Dependent Signaling and Kinases in Apicomplexan Parasites. Cell Host Microbe 5: 612 (2009)

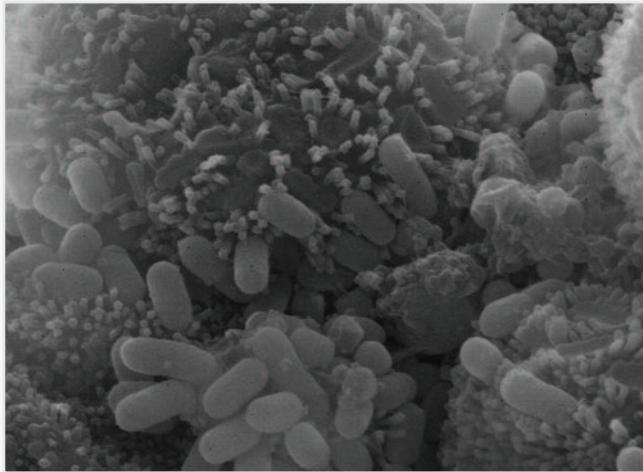


Male (pink) and female (purple) transmission stages of a rodent malaria parasite purified from blood.



Scanning electron micrograph of *Citrobacter rodentium* adhering to the mucosal surface of an infected mouse colon. The fine strands corralling the *Citrobacter* are sticky nets of DNA exuded from neutrophils to demobilise the bacteria before engulfing and destroying them.

We focus on three areas: (a) the phylogenetic and phenotypic analysis of *Salmonella*; (b) linking genotype to phenotype in other bacterial pathogens; and (c) screening mice for infection susceptibility phenotypes.



Simon Clare and Dave Goulding

Scanning electron micrograph of a *Citrobacter rodentium* invading a host enterocyte.

Following on from our *Salmonella* Typhi (typhoid) project, phylogenetic analysis has been extended to other pathogens, including *Salmonella* Typhimurium, *Clostridium difficile*, *Shigella sonnei* and *Vibrio cholerae*. Phylogeny is mainly built on analysis of SNPs, and other types of genome variation (such as plasmid content).

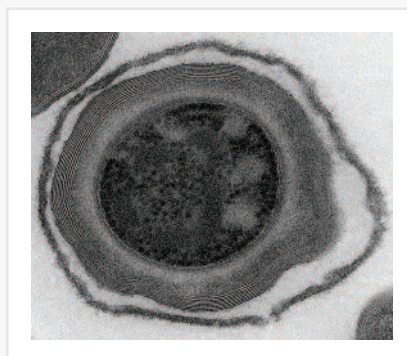
For typhoid, we are using genotype data to map the geographical distribution of strains within cities, including Kathmandu and Kolkata. The eventual aim is to identify transmission routes and to bring in vaccination-based control programmes.

We have identified a novel *S. Typhimurium* genotype associated with invasive salmonellosis in Africa, in which antibiotic resistance is associated with a novel locus on the virulence plasmid. The strains show signs of genome degradation, suggesting that they may be adapting to the human host.

Our *V. cholerae* project is focusing on novel hybrid strains which are replacing the previously dominant El Tor strain. In the host-adapted pathogens *S. Typhi* and *S. Gallinarum*, we have identified core gene sets involved in survival both within and outside the host.

We have also established a murine model of *C. difficile* infection that shares many of the features of human infection. We have identified the spore as the key vehicle of transmission and are using proteomic approaches to define its composition. We are establishing collaborations with hospitals to test practical applications of our model.

Over the last year we have challenged over 90 novel lines of mutant mice generated by the Mouse Genetics Programme with *Citrobacter rodentium* and *S. Typhimurium*, and are currently screening five novel lines a week. As of August 2009 we had identified approximately 10 lines with altered susceptibility to infection (three to *C. rodentium* and five to *S. Typhimurium*). We are collaborating with both internal and external groups to follow up these findings.



Trevor Lawley and Dave Goulding

Transmission electron micrograph of a *Clostridium difficile* spore.



Wellcome Library, London



Whole-genome sequencing of 19 *Salmonella* Typhi isolates finds little evidence of immune-driven selection, emphasising the likely role of asymptomatic carriers in transmission.

Holt KE et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat Genet* 40: 987 (2008)



A new technique for exploring the transcriptome, strand-specific RNA-seq, reveals hidden features in the *Salmonella* Typhi genome.

Perkins T et al. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella* Typhi. *PLoS Genetics* PLoS Genet 5: e1000569 (2009)

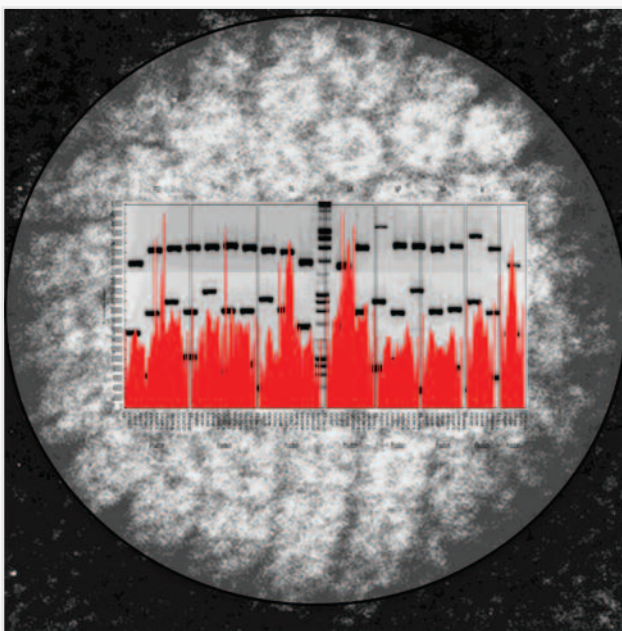


Antibiotic treatment of mice infected with *Clostridium difficile* turns them into 'supershedders' of highly infectious spores.

Lawley TD et al. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect Immun* advanced online publication (29 June 2009)

Our aim is to understand how genetic changes in viruses and their hosts influence viral pathogenesis, persistence and host susceptibility to infection.

Paul Kellam joined the Sanger Institute in January 2009.



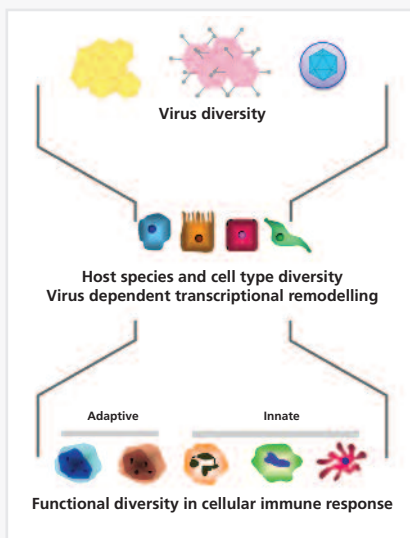
Next generation sequencing of virus genomes.

To assess genome sequence diversity in RNA and DNA viruses, we are using next-generation technology to sequence the complete genomes of collections of herpes viruses, HIV and viruses important to animal health. We have also established an influenza virus genome sequencing pipeline, which was used to sequence some of the initial UK isolates of H1N1 swine flu. This pipeline remains prepared for any further developments of this unfolding pandemic.

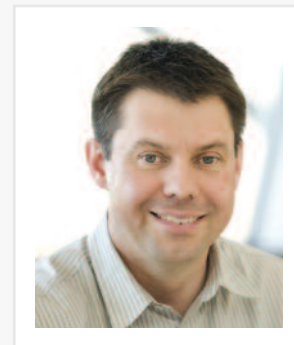
We are examining the viral 'metagenome' in human and animal reservoirs, to identify the presence of new and known virus genomes. Using computational methods and data resources, we aim to get a better understanding of the full extent of viral genetic variation.

How viruses change the intracellular environment has been a focus of our past research. This will continue at the Sanger Institute, where we will use RNAseq to

investigate the patterns of gene expression and transcript-processing dynamics in cells infected with human pathogenic viruses, in collaboration with the University College London/Medical Research Council Centre for Medical Molecular Virology. Using the mouse genetic resources of the Sanger Institute we will also develop methods for investigating the molecular detail of virus pathogenesis.



Levels of genetic and functional diversity that contribute to virus pathogenesis.



Wellcome Library, London

➤ **Hypoxia-triggered activation of Kaposi's sarcoma-associated herpesvirus from its latent to lytic state depends on the transcription factor X-box binding protein-1.**

Dalton-Griffin L et al. X-box binding protein-1 contributes to the induction of the KSHV lytic cycle under hypoxia. *J Virol* 83: 7202 (2009)

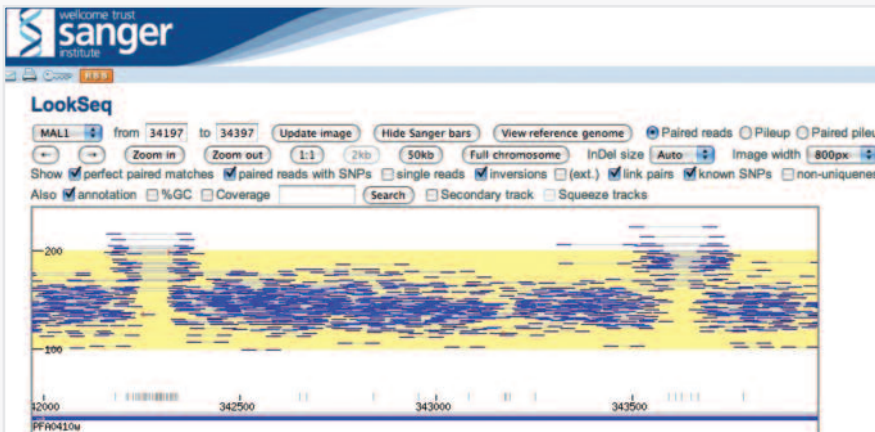
➤ **Genome-wide transcriptional profiling in HIV-infected macrophages suggests that, even though the key NF-kB pathway is significantly perturbed, gene expression is only mildly affected.**

Noursadeghi M et al. Genome wide innate immune responses in HIV 1 infected macrophages are preserved despite attenuation of the NF-kB activation pathway. *J Immunol* 182: 319 (2009)

➤ **A comparison of gene expression in cells infected by oncogenic and non-oncogenic adenoviruses reveals a complex web of pathways altered in virus-infected cells.**

Strath J et al. Identification of genes differentially expressed as result of adenovirus type 5- and adenovirus type 12-transformation. *BMC Genomics* 10: 67 (2009)

By analysing genome variation in humans and *Plasmodium*, we aim to understand the evolution and biology of host–parasite interactions that are critical for disease control.



LookSeq. A web viewer for next-generation deep sequencing data: <http://lookseq.sf.net>. This view shows two small deletions of sequence in a clinical isolate of *Plasmodium falciparum*.

Our work on human genetics is underpinned by a global partnership of malaria researchers – the Malaria Genomic Epidemiology Network – who have collected clinical data and DNA from more than 50 000 individuals for genetic studies aimed at accelerating malaria vaccine development by improving understanding of the molecular mechanisms of protective immunity.

Because of the genetic diversity of African populations, genome-wide association studies in Africa pose considerable methodological challenges – for example, signals tend to be much weaker than in European populations. We can boost the strength of signals by imputation, inferring SNPs on the basis of known haplotypes, as long as reference genome sequence data are available for the population in which the study was carried out. Once a signal has been discovered, however, it may be easier to identify the causal variant in African than in European populations. Our work in this area ties in closely with that of the 1000 Genomes Project.

Our work on *Plasmodium* genome variation is based on a partnership with researchers in over 10 malaria-endemic countries, who are conducting epidemiological studies of antimalarial drug resistance or clinical trials of new drugs or vaccines. Using next-generation technologies, we have carried out whole-genome sequencing of 200 isolates of *P. falciparum* from Africa, Asia, Oceania and

South America, using informatic and statistical methods to overcome the problems associated with highly repetitive and variable sequences in the *Plasmodium* genome.

Using this information, we can begin to identify the approximate geographical origin of a parasite. We can also see unusually long haplotypes, signifying recent evolutionary change due to antimalarial drug use and other factors. These advances point the way to genome-based methods of monitoring global patterns of parasite migration and of drug resistance, which could provide early warning of drug resistance and assist public health policy decision-making.



Working with communities affected by malaria. A health care worker collects a blood sample from a child near Kilifi, Kenya.



➤ **An introduction to the work of the Malaria Genomic Epidemiology Network, uniting researchers from 21 countries.**

Malaria Genomic Epidemiology Network. A global network for investigating the genomic epidemiology of malaria. *Nature* 456: 732 (2008)

➤ **A study of genetic resistance to malaria in 2500 Gambian children provides new methodological insights for genome-wide association studies in Africa.**

Jallow M et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet* 41: 657 (2009)

➤ **A guide to the LookSeq browser for viewing large quantities of sequence data at multiple levels.**

Manske HM and Kwiatkowski DP. LookSeq, a browser-based viewer for deep sequencing data. *Genome Res advanced online publication* (13 August 2009)

We use whole-genome sequencing, comparative sequencing and analysis of variation to explore the diversity of pathogens and how they cause disease in humans and other animals.



Gemma Langridge

tuberculosis and *Bordetella pertussis*, where we expect to find very little variation. We can use the same approach with subtypes of more diverse pathogens, including *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Salmonella* Typhimurium and *Clostridium difficile*. Our experimental work has expanded to include, for example, capsular biosynthesis and virulence gene expression in *Streptococcus pneumoniae*, mechanisms of lateral gene transfer in *Salmonella bongori*, and analysis of DNA from single cells of unculturable bacteria.

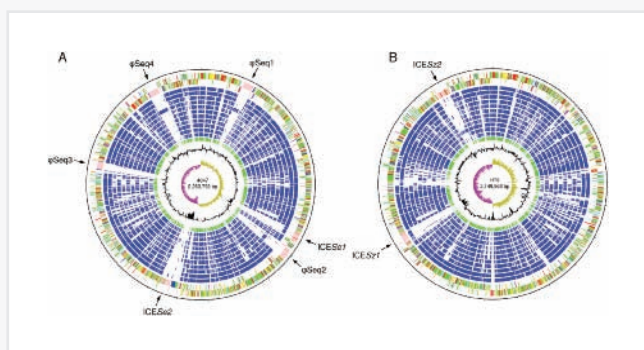
This figure shows all the genes of *Salmonella* Typhi around the outside, with essential and non-essential genes for various conditions, identified by TraDIS, in the central rings.

We use a 'broad and deep' approach, looking at a wide range of pathogens to understand the diversity of mechanisms underlying host interaction and pathogenicity, while also dissecting key differences among strains of important organisms such as *Salmonella* and *Clostridium* spp. We also collaborate internally and externally on more applied projects, for example using high-throughput sequence data to identify SNPs for bacterial epidemiology and bacterial population sampling to examine the health impact of host-associated microbiota.

Capitalising on next-generation sequencing technologies, we are sequencing tens to hundreds of genomes of *Salmonella* Typhi, *Salmonella* Paratyphi A, *Mycobacterium*

As well as *Streptococcus zooepidemicus* and *equi*, we have published genome sequences of the cystic fibrosis pathogens *Pseudomonas aeruginosa* and *Burkholderia cenocepacia*, a pandemic lineage of *Streptococcus pneumoniae*, enteropathogenic *Escherichia coli*, and the food poisoning organism *Salmonella* Enteritidis, along with its chicken-restricted relative *Salmonella* Gallinarum. We have also analysed two genomes and six plasmids from *Chlamydia trachomatis*, including a new Swedish variant that cannot be detected by conventional diagnostic PCR because of a specific deletion. Through detailed analysis we were able to suggest new diagnostic targets.

In addition, we have developed two innovative sequence-based technologies, to generate strand-specific information on transcription (ss-RNAseq) and to detect large numbers of transposon mutants in a bacterial population (TraDIS).



Representations of the genomes of *Streptococcus equi* (A) and *Streptococcus zooepidemicus* (B) showing all genes on the outside, and genes conserved with various relatives on the inner circles. Holden MTG et al. *PLoS Pathogens* 2009 5(3): e1000346 doi:10.1371/journal.ppat.1000346



Wellcome Library, London

A comparative analysis of *Streptococcus equi* and *S. zooepidemicus* suggests that *S. equi* has acquired virulence genes from phage and is exchanging phage with the human pathogen *S. pyogenes*.

Holden MT et al. Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS Pathog* 5: e1000346 (2009)

Sequencing and experimental studies of an epidemic strain of *Pseudomonas aeruginosa* isolated from cystic fibrosis patients suggest that particular 'genomic islands' have promoted its spread.

Winstanley C et al. Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* 19: 12 (2009)

Whole-genome sequencing of 19 *Salmonella* Typhi isolates finds little evidence of immune-driven selection, emphasising the likely role of asymptomatic carriers in transmission.

Holt KE et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat Genet* 40: 987 (2008)

We are studying the impact of genetic variation on host-parasite interactions during the erythrocyte stages of the *Plasmodium falciparum* life cycle.



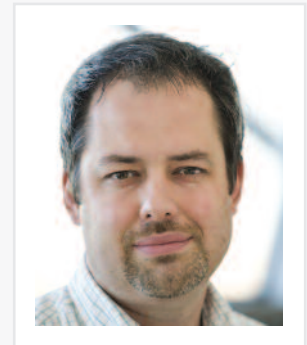
Through collaboration with Dr. OraLee Branch at NYU, the Rayner lab works with the Universidad Nacional de Amazonia Peruana at a research site near Iquitos, in the Peruvian Amazon, pictured above.

We are using genome-wide approaches to understand the molecular basis of the multiple host-parasite interactions that are central to *P. falciparum* erythrocyte invasion. Erythrocyte invasion is essential for parasite survival and pathogenesis, so understanding invasion will identify new targets for drug and vaccine development.

We are using flow cytometry to develop a high-throughput platform for quantifying erythrocyte invasion. This approach will allow us to assess how natural genetic variants (*P. falciparum* field strains and/or human erythrocytes from different genetic backgrounds) and genetically manipulated *P. falciparum* lines impact invasion efficiency and specificity. Measuring the impact of natural genetic variation on erythrocyte invasion will form a key interaction with Dominic Kwiatkowski's *P. falciparum* genetic diversity project. In pilot studies we have phenotyped the erythrocyte invasion pathways used by *P. falciparum* isolates from the Peruvian Amazon, while the Kwiatkowski group has carried out whole-genome sequencing of the same isolates.

With Gavin Wright's team, we are using AVEXIS (avidity-based extracellular interaction screen) technology to perform the first large-scale screen for receptor-ligand interactions between *P. falciparum* and the erythrocyte. Initial screening has identified candidate interactions that are undergoing *in vitro* validation. Once hits are confirmed, the same technology can be used to assess the impact of natural genetic variation on the affinity and/or specificity of specific receptor-ligand interactions, as well as to screen for inhibitors that could block these interactions, and hence potentially inhibit erythrocyte invasion.

Finally, in collaboration with the Sanger Institute proteomics core facility led by Jyoti Choudhary we are studying two post-translational modifications important in erythrocyte invasion – phosphorylation and palmitoylation. Initial screens have identified multiple new palmitoylated and phosphorylated proteins in invasive *P. falciparum* blood-stage parasites. In collaboration with Oliver Billker we are correlating these findings with changes in phosphorylation and palmitoylation in the model parasite *P. berghei*, in order to identify conserved mechanisms that regulate parasite motility and invasion across *Plasmodium* species and life-cycle stages.



Wellcome Library, London

↗ **A review of a multiprotein family, the reticulocyte-binding-like family, and its role in erythrocyte invasion.**

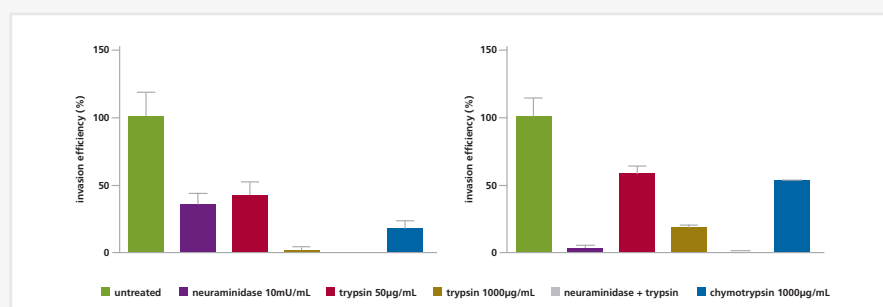
Rayner JC The merozoite has landed: RBLs and specificity of erythrocyte invasion. Trends Parasitol 25: 104 (2009)

↗ **An analysis of 630 *P. falciparum* infections over four years suggests that merozoite surface protein-3 is less variable over time than previously thought.**

Jordan SJ et al. Genetic diversity of the malaria vaccine candidate *Plasmodium falciparum* Merozoite Surface Protein-3 in a hypoendemic transmission environment. Am J Trop Med Hyg 80: 479 (2009)

↗ **Sequence analysis identifies *P. falciparum* equivalents of genes for all seven SNAREs, key proteins in vesicle transport and secretion.**

Parish LP and Rayner JC. *Plasmodium falciparum* secretory pathway: Characterization of PfStx1, a plasma membrane Qa-SNARE. Mol Biochem Parasit 164: 153 (2009)



Plasmodium falciparum strains can differ widely in their erythrocyte invasion profile. Here two strains from Peru have invasion profiles quite distinct from each other.