

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 affects interactions with pathogens.

- Parasite genomics
- Pathogen biology
- Virus genomics
- Bacterial pathogenesis
- Pathogen genomics
- Experimental genetics of rodent malaria
- Malaria genomics
- Malaria programme

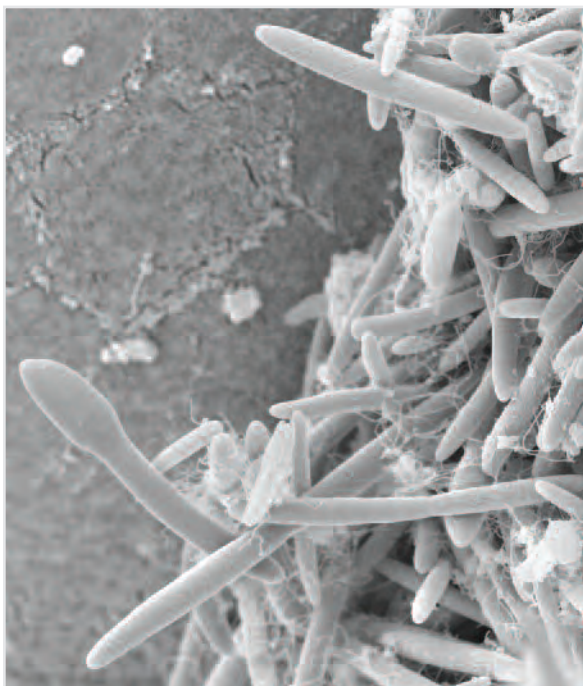
Detailed genetic analysis is enabling us to define pathogen populations and generate high-resolution phylogenies to track their evolution over timescales ranging from weeks to decades, and through environments ranging from individuals to continents. Moreover, in combination with phenotypic data, detailed genetic characterisation can identify links between genotype and phenotype, for important traits such as virulence or drug resistance. In terms of the host response, a combination of improved *in vitro* and *in vivo* models and a better understanding of host genetics has enabled us to identify mammalian genes contributing to infection control. All these developments are made possible by the core pipelines available at the Sanger Institute.

Bacteria

We are using high-throughput sequencing to study the biology and genetics of bacteria in unprecedented detail. Our first proof-of-principle study was to look at 65 isolates of a single type of MRSA, ST239, which has spread through hospitals worldwide over the past 30 years. A second, deeper study looked at the global spread of a pandemic strain of *Streptococcus pneumoniae* (Spanish 23F), revealing regional variation driven by differing selective pressures of antibiotics and vaccines. We are now expanding this approach to many other microorganisms, including *Mycobacterium tuberculosis*, *Bordetella pertussis*, *Salmonella Typhi*, *Shigella sonnei*, *Chlamydia* and *Vibrio cholerae*.

Alongside these population studies, we are using high-throughput phenotyping to characterise mechanisms of drug resistance and growth requirements for up to 100 strains simultaneously. We can also couple sequencing techniques to saturation mutagenesis to map millions of individual knockout mutants in a single pool, monitoring the function of genes in growth, infection and transmission.

Healthcare-associated infections present a major challenge to clinical practice and public health control – the modern hospital environment selecting for super-fit variants of commensal, environmental or opportunistic pathogens. We are investigating the genetic traits underlying the pathogenesis, persistence and transmission of *Clostridium difficile*, a major cause of diarrhoea associated with antibiotic use. We are currently developing a novel therapy that restores a healthy microbial community to the intestine of affected individuals. We are also working with the UK National Health



Dave Goulding, Genome Research Limited

Electron microscopy picture of *Clostridium difficile*. We are working with the UK National Health Service to develop systems to track the transmission and evolution of this bacteria.

Faculty members

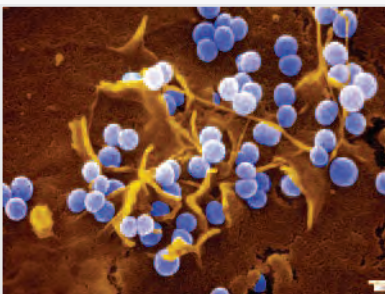
Gordon Dougan, Head
Matt Berriman
Paul Kellam
Trevor Lawley
Julian Parkhill
Oliver Billker
Dominic Kwiatkowski
Julian Rayner

Service, Health Protection Agency and local hospitals to develop a phylogenetic-based typing system for tracking transmission and evolution of *C. difficile* at the hospital and global level.

Other studies are examining bacterial populations in the lungs of cystic fibrosis patients, in the gut of people following different diet regimes, and in those with inflammatory bowel diseases.

Parasite genomes and biology

Control of malaria is particularly difficult as parasites are continually acquiring genetic changes that enable them to overcome control measures; conversely, human populations in malaria-endemic regions have acquired genetic changes offering protection against the disease. In the past year we have made progress in identifying genes undergoing evolution in both human and parasite populations.



Sharon Peacock, University of Cambridge

Staphylococcus aureus cells. Next-generation sequencing is allowing the transmission and evolution of this bacteria to be elucidated at the global and local level.



Crossing continents

Genomic techniques have been used to track MRSA strains across decades and continents, and between individual hospital patients.

MRSA (methicillin-resistant *Staphylococcus aureus*) is a significant problem in many healthcare settings. Traditional methods of categorisation have limited power to distinguish strains, but new genomic tools provide a way to track its long-term evolution as well as its short-term spread within disease outbreaks.

In a suitably international collaboration, Institute researchers examined a strain of MRSA that is dominant across much of Asia and also present in South America and eastern Europe and causes occasional outbreaks in western Europe. Some 43 isolates collected globally between 1982 and 2003 were analysed, alongside 20 isolates obtained from patients at a Thai hospital over a seven-month period.

The phylogenetic tree of the global isolates sheds light on the likely evolution and dispersal of the strain, which probably arose in Europe in the 1960s. Patterns of genetic diversity strongly suggest that its colonisation of new geographic areas – such as South America – is linked to individual cases of infection.

Sequence comparisons also shed intriguing light on the likely 'natural history' of certain outbreaks. A group of Portuguese isolates, for example, appear to have been imported from South America – suggesting that the strain has traversed the Atlantic in both directions during its 40-year life. An 'Asian' variant in Denmark was probably brought in by a Thai patient, while an anomalous outbreak in a London hospital was probably seeded by a variant brought in from south-east Asia.

Remarkably, as well as being able to track the strain's evolution over decades, genomic comparisons could also shed light on transmission within the Thai hospital over a period of months. The pattern of transfer reflected the location of patients within the hospital, highlighting the potential use of such methods for infection control.

Strikingly, a small number of genetic changes do not fit the general inheritance patterns, and probably result from convergent evolution. In many cases, the affected genetic site is involved in antibiotic resistance, suggesting that the changes reflect the selective pressures of clinical practice. Hence analysis of similarly anomalous sites could shed light on the strain's response to control measures.

Harris SR et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010; 327(5964):469–74.

Working with partners in Africa through the MalariaGEN consortium, we have identified regions of the human genome that show evidence of recent selection. A genome-wide survey of *Plasmodium falciparum* populations in Africa, Asia and Oceania, using next-generation sequencing technology, has revealed many genes showing biologically fascinating differences between populations, notably in candidate genes for drug resistance and vector specificity. An important aspect of this work is a user-friendly web application to make data accessible to the wider scientific community.

We have also played a key role in a study of malaria parasites infecting wild gorillas and chimpanzees in central Africa, which revealed that *P. falciparum* probably originated from a parasite infecting western lowland gorillas. Further studies are exploring whether such cross-species transmissions still occur. We are using molecular and cellular tools to study host–parasite interactions during the blood stages of the *P. falciparum* life-cycle, particularly erythrocyte invasion. A new high-throughput fluorescence-activated cell-sorting platform is being used to quantify invasion and investigate the influence of natural genetic variation initially in the parasite, and eventually also in the host.

High-throughput experimental approaches developed by teams in the Institute are being used to identify new interactions between *P. falciparum* and erythrocyte surface proteins. We have identified several novel receptor–ligand pairs, which are currently being functionally validated. Additionally, this research has generated novel methods for producing usable amounts of recombinant *P. falciparum* merozoite surface proteins, of potential importance in malaria vaccine development.

We have continued to produce and curate high-quality reference genomes of different species of human and animal malaria parasites. The complete life-cycle of the rodent malaria parasite *P. berghei* can be studied in the laboratory, providing opportunities to explore their basic biology and interactions with hosts and vectors. Importantly, *P. berghei* is readily tractable to genetic manipulation. Scaling up experimental approaches has been a major aim this year. In a pilot study we systematically targeted each of 66 protein kinase genes, in order to prioritise targets for drug development. We are working on new technologies to achieve further scale-up.

We use two main approaches to study other major parasites. First, comparisons between genomes from related species highlight key similarities and differences. Comparisons between the clinically important yeast *Candida dubliniensis* and the more pathogenic *C. albicans*, the cause of thrush, revealed key gene families expanded in *C. albicans* as well as gene loss in *C. dubliniensis*, changes that could account for their differing pathogenicity.

Our second approach is to use large-scale RNA transcript sequencing to identify changes in gene expression. For example, during the parasite's life-cycle we can use RNA sequencing to define new genes and gene complexity.

Our helminth studies have been boosted by new bioinformatics tools that dramatically reduce the time and effort required to build a genome. This has enabled us to produce improved genome drafts of *Schistosoma mansoni* (blood fluke), *Echinococcus multilocularis* (tapeworm), *Strongyloides ratti* (threadworm) and *Trichuris muris* (whipworm).



Wellcome Library, London

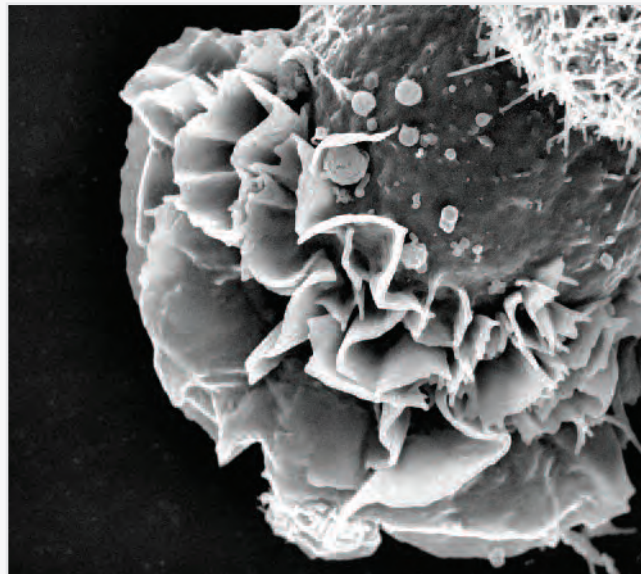
Mosquito dissection. We are able to study the complete life cycle of the rodent malaria parasite *P. berghei*. This year we systematically targeted each of 66 kinase genes to prioritise targets for drug development.

Virus epidemiology

Virus gene sequencing and phylogenetics can be used to study the epidemiological dynamics of rapidly evolving viruses. With complete genome data, individual transmission chains of viruses can be traced during the course of an epidemic. We have established methods for next-generation sequencing and assembly of whole virus genomes and techniques for deep sequencing of virus populations within an individual.

For pandemic influenza A virus H1N1 (swine flu), whole-genome analysis has been used to track individual influenza lineages in fine detail during an epidemic. Monitoring of such lineages could help to identify emerging antigenic variants.

We now plan to combine analysis of virus and host genome variation, using both defined gene knockouts in mouse models of virus infection and human genetic variation in diverse populations and in extreme disease phenotypes.



Dave Goulding, Genome Research Limited

Salmonella. The genome sequence of ST313 provides compelling evidence that it is becoming a specialized pathogen of people.

A change for the worse

A new and deadly strain of *Salmonella* Typhimurium has emerged in Africa.

Salmonella are a large and complex group of microorganisms. Some cause only mild disease, while others are potentially deadly. The roots of their pathogenicity lie in their genetic make up – a point well illustrated by the discovery that a novel invasive form of *Salmonella* Typhimurium in sub-Saharan Africa has now emerged on the coat-tails of the HIV pandemic.

Salmonella Typhimurium usually causes gastroenteritis – unpleasant, but not a major threat to health. Over the past several decades, however, it has been associated with severe disease in a growing number of people, killing up to a quarter of those affected.

Most adults with the invasive disease were also infected with HIV, although children without HIV are also infected, and it is likely that severity was a consequence of compromised immune systems. However, an analysis of disease isolates, carried out by the Sanger Institute and colleagues in Malawi, Kenya, revealed that invasive disease was being caused by a genetically distinct strain of Typhimurium, ST313.

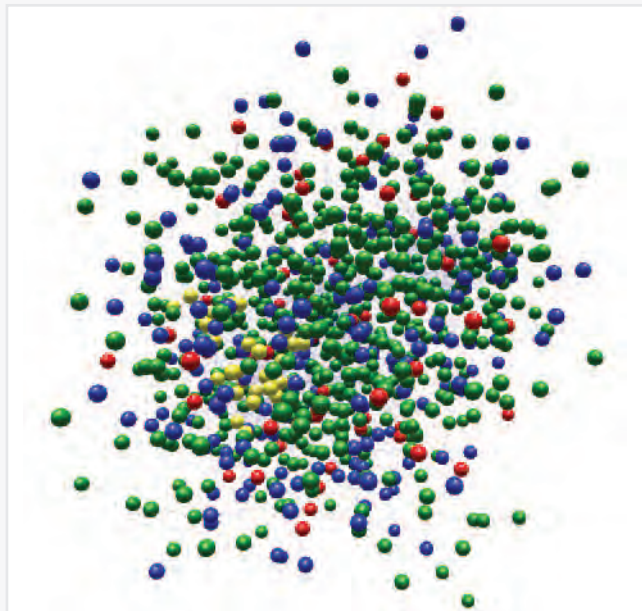
The genome sequence of ST313 provides compelling evidence that, unlike conventional strains, which can survive in a wide range of animal hosts, it is adapting to life exclusively in susceptible humans. Substantial chunks of its genome are degraded, which is likely to hinder its survival in other species. Moreover, some of these genomic regions are also disrupted in *Salmonella* Typhi, a relative that has also become a specialised pathogen of people. The strain's resistance to commonly used drugs stems from its acquisition of a mobile genetic element carrying multiple antibiotic resistance genes.

Multidrug-resistant ST313 has swept through human populations at remarkable speed. Within 18 months of being detected in 2002, it had come to represent 95 per cent of *S. Typhimurium* isolates identified in Blantyre, Malawi. This appears to have occurred also in Kenya, and clues that this may have occurred across the whole of the sub-Saharan African continent are beginning to emerge.

The results emphasise how entirely new infectious diseases can arise and spread with great rapidity. Genomic comparisons provide an important tool by which such emerging infections can be identified and tracked, and may also provide leads for the development of new drugs or vaccines.

Kingsley RA et al. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* 2009; 19(12):2279-87.

We are studying the genomes of parasites that cause neglected tropical diseases and using comparative genomics to probe their biology.



Jackson AP et al. *PLoS Negl Trop Dis* 2010; 4:e658
doi: 10.1371/journal.pntd.000658

Variant surface glycoprotein (VSG) repertoire of *T. b. brucei* (927) represented as a three-dimensional network graph. Nodes are shaded by type: orthologous sequence in *T. b. gambiense* (blue), orthologous sequence in *T. b. gambiense*, but closest relative in *T. b. brucei* (green), no corresponding sequence in *T. b. gambiense* (red), metacyclic-stage VSG (purple) and VSG-related (VR) proteins.

By sequencing and analysing the genomes of parasites, we aim to uncover the genomic features that underlie their unusual biology. We produce high-quality reference genomes and compare them with the genomes of related strains or other species to provide evidence of drug selection, identify genes involved in host-pathogen interactions, or uncover other biologically interesting features.

Parasitic worms (helminths) cause almost immeasurable morbidity worldwide. Helminth genomes are typically five to ten times the size of the malaria parasite genome and much more complex. Over the past year we have continued high-throughput data production on helminth genomes and also tackled the informatics challenges associated with piecing these data together. We have developed software to visualise the outputs from next-generation sequencing in the context of genes, as well as algorithms to reduce the amount of time required to correct errors or bridge sequence gaps.

As a follow-up to the African trypanosome (*Trypanosoma brucei*), we have sequenced the genome of the subspecies responsible for most human deaths in Africa. The highly conserved core genome indicates that findings made with the commonly used laboratory 'model' *T. brucei* are applicable to

the 'wild' organism. Outside the core, the variant surface antigen repertoire is extremely diverse. Nevertheless, its structure pre-dates the split of the subspecies.

In comparative genomics, we generated a complete genome sequence of the fungal pathogen *Candida dubliniensis* and compared it to available sequence of its more pathogenic relative *Candida albicans*, uncovering striking features that may account for differences in virulence. Many pathogenesis-associated genes have degenerated in *C. dubliniensis*, while other virulence genes have expanded into larger families in *C. albicans*.

We have continued to explore the use of next-generation sequencing technology to analyse gene function in parasites. In the first study of its kind in parasites, we sequenced the transcriptome of the malaria parasite *Plasmodium falciparum*, and are now applying high-throughput transcriptome sequencing to other parasites, including helminths.



Wellcome Library, London

A comparison of *Candida* species reveals a family of transcription factors and transmembrane proteins as putative virulence factors.

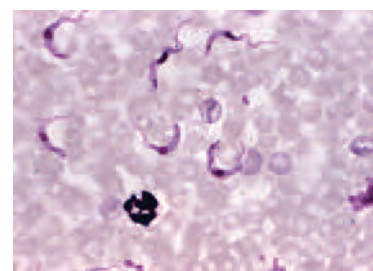
Jackson AP et al. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res* 2009; 19:2231-44.

Deep sequencing of RNA sheds light on gene composition and extent of alternative splicing in the malaria parasite.

Otto TD et al. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol* 2010; 76:12-24.

The genome sequence of the African trypanosome responsible for most human disease.

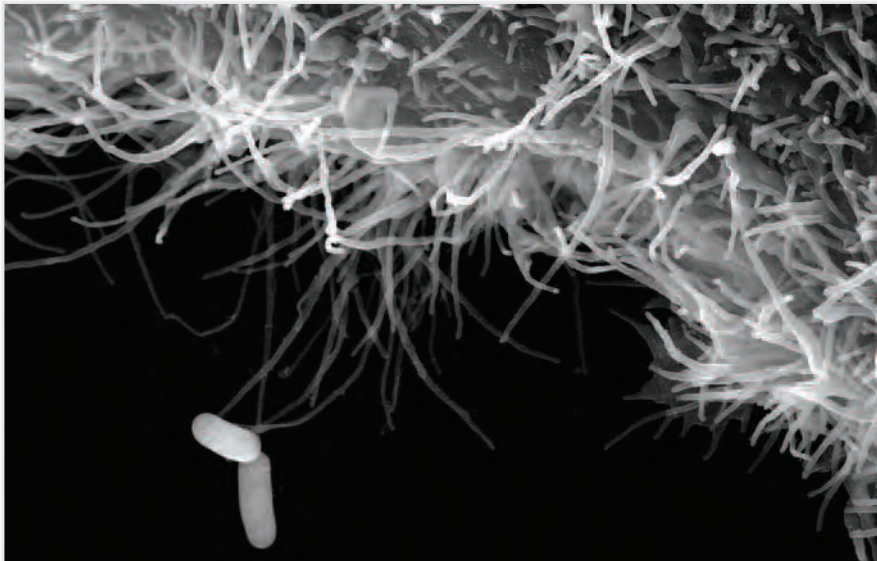
Jackson AP et al. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. *PLoS Negl Trop Dis* 2010; 4:e658.



Trypanosoma brucei gambiense: some 60 million people are at risk of trypanosomiasis in sub-Saharan Africa.

CDC/Dr. Mae Melvin

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



Dave Goulding, Genome Research Limited

Electron micrograph showing *Salmonella* attaching to a human cell. A new type of invasive *Salmonella* has emerged in Africa, possibly linked to the spread of HIV.

We have constructed a high-resolution phylogenetic tree of *Salmonella* Typhi, the cause of human typhoid, based on shared inheritance of single nucleotide polymorphisms (SNPs), insertions and deletions, and plasmids. All *S. Typhi* isolates map onto the tree in an unequivocal manner. Clinical, phenotypic and geographical data can be mapped back to the tree, linking genotype to phenotype. We are currently using this genotype data to map the geographical distribution of *S. Typhi* within cities such as Kathmandu, Nairobi and Kolkata. The aim is to identify transmission routes and support vaccination-based control programmes.

We have extended our phylogenetic analysis to other pathogens. For example, we have identified a novel *Salmonella* Typhimurium genotype associated with invasive salmonellosis in Africa, in which multiple antibiotic resistance genes are encoded on a mobile DNA cassette and virulence-associated plasmid. We have mapped the spread of this clade, known as ST313, across sub-Saharan Africa and believe the epidemic is linked to the spread of HIV in the same region. ST313 isolates show signs of genome

degradation, suggesting that they may be adapting to the human host. Similar studies are being undertaken in other pathogens including *Vibrio cholerae*, *Shigella sonnei* and *Clostridium difficile*.

We have continued to play a key role in the Mouse Genetics Programme by operating an infection challenge pipeline based on *Citrobacter rodentium* and *S. Typhimurium*. We have identified more than ten novel gene knockout mice with altered susceptibility to infection. We are currently screening approximately five novel mouse lines a week and are considering introducing viral and *Plasmodium* (malaria parasite) challenges. We are collaborating with both internal and external groups to follow up these findings through secondary phenotyping efforts.



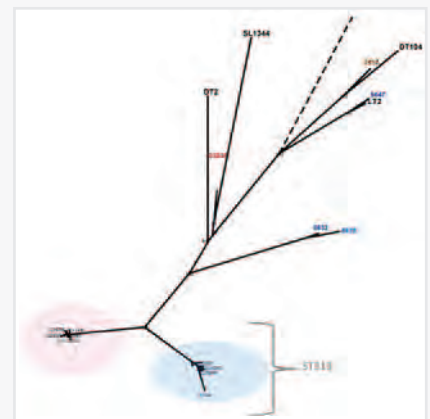
Wellcome Library, London

➤ **Genomic analysis identifies a new type of multidrug-resistant *Salmonella* Typhimurium causing invasive disease in Africa.**

Kingsley RA et al. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* 2009; 19(12):2279-87.

➤ **SNP typing study in Kathmandu identifies variants responsible for disease in children.**

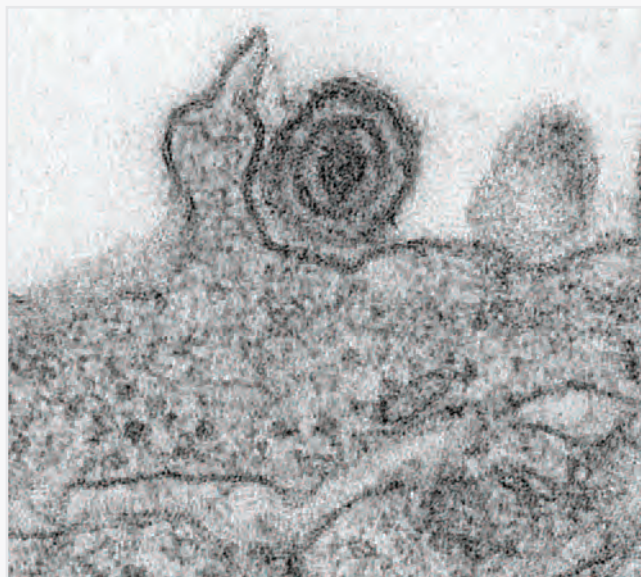
Holt KE et al. High-throughput bacterial SNP typing identifies distinct clusters of *Salmonella* Typhi causing typhoid in Nepalese children. *BMC Infect Dis* 2010; 10:144.



Genome Research Limited

A phylogenetic tree of *S. Typhimurium* showing the outlying ST313 clade as independent branches. Our *Salmonella* Typhi family tree is being used to map routes of transmission and plan vaccination campaigns against typhoid fever.

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



Dave Goulding, Genome Research Limited

Electron micrograph of a Herpes Simplex Virus on the surface of a murine ES cell. We are sequencing the complete genomes of herpesviruses and influenza viruses to understand the genomic diversity of DNA and RNA viruses.

To assess genome sequence diversity in RNA and DNA viruses, we are using next-generation technology to sequence the complete genomes of collections of herpesviruses, influenza viruses and viruses important to animal health.

We have used these methods to track human influenza A H1N1(2009) ('swine flu') infections within the UK during the recent pandemic. Using whole-genome sequencing methods we analysed 154 influenza genomes from UK isolates from the first (128 isolates) and second (26 isolates) waves of the 2009 pandemic. Using these sequences and their dates of isolation we inferred the genetic epidemiology of the UK epidemic. We determined that the UK epidemic is composed of many co-circulating lineages, of which at least 12 were exclusively or predominantly UK clusters.

We are using similar methods to understand influenza and HIV diversity within infected individuals. Next-generation sequencing technology is also allowing us to examine the viral 'metagenome' in human and animal reservoirs, to identify the presence of new and known virus genomes.

Computational and data resources are essential for our work and we have developed methods for virus genome data processing and assembly of complete virus

genomes. In collaboration with Oliver Pybus (University of Oxford) and Andrew Rambaut (University of Edinburgh), we are able to use phylogenetic methods to understand the impact of virus genetic variation and how this can uncover virus evolutionary histories. These methods use the date of virus sampling, the whole genome sequence of each virus and the virus evolutionary rate to accurately document the most recent common ancestor of a group of related viruses. These groups define chains of transmission between people.

In collaboration with colleagues at University College London, we are investigating the impact of host and virus genetic diversity on disease. We have identified some of the mechanisms by which Kaposi's sarcoma-associated herpesvirus (KSHV) responds to the intracellular environment. We have showed that in B-cell tumours infected with both KSHV and Epstein Barr Virus (EBV), the B-cell transcription factor X-box binding protein-1 (XBP-1) specifically induces the replication on KSHV and not EBV. Additionally, we have demonstrated how KSHV proteins interfere with cellular proteins counteracting the host cell's antiviral defences.

Model of H1N1 influenza virus. The UK pandemic was the result of multiple, distinct Influenza A H1N1 (2009) introductions.



Wellcome Library, London

➤ **Multiple genetically distinct subpopulations of equine influenza virus are transmitted between animals.**

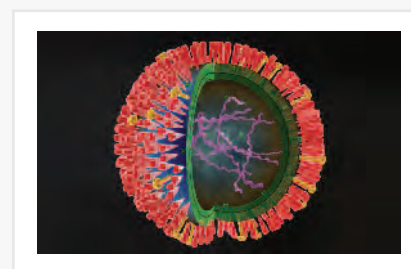
Murcia PR et al. Intra- and interhost evolutionary dynamics of equine influenza virus. *J Virol* 2010; 84(14):6943-54.

➤ **The herpesvirus that causes Kaposi's sarcoma produces microRNAs that reprogramme lymphatic endothelial cells.**

Hansen A et al. KSHV-encoded miRNAs target MAF to induce endothelial cell reprogramming. *Genes Dev* 2010; 24(2):195-20.

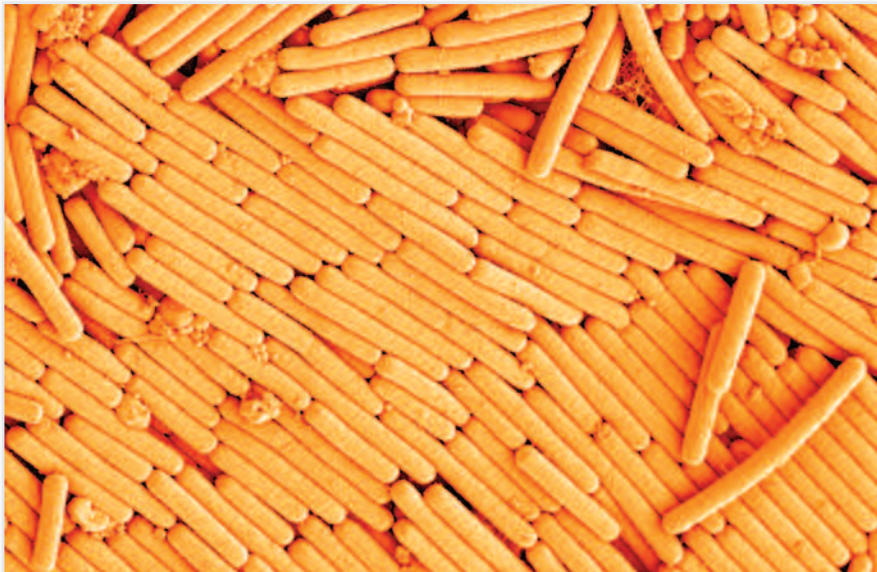
➤ **Kaposi's sarcoma-associated herpesvirus disables a host cell defence protein, tetherin, that inhibits the release of viral particles from the cell surface.**

Pardieu C et al. The RING-CH ligase K5 antagonizes restriction of KSHV and HIV-1 particle release by mediating ubiquitin-dependent endosomal degradation of tetherin. *PLoS Pathog* 2010; 6(4):e1000843.



Anna Tanczos, Wellcome Images

Our research focuses on identifying the mechanisms that *Clostridium difficile* uses to cause intestinal disease and transmit between hosts, to inform infection-control measures and development of therapeutics.



Dave Goulding, Genome Research Limited

Electron microscope image of *C. difficile* colony. *C. difficile* has existed for millions of years but human infections were first detected only in 1978.

We use phylogenetic, proteomic, phenotypic and epidemiological approaches to study *Clostridium difficile* disease and transmission in experimental mouse infections and within healthcare settings.

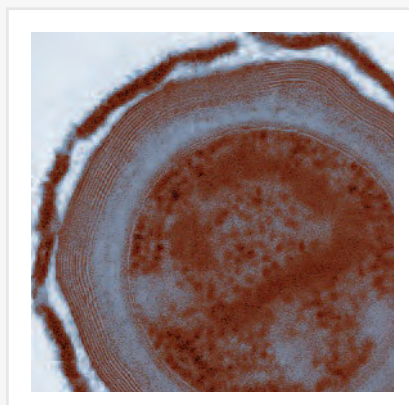
C. difficile has rapidly emerged as a significant cause of healthcare infection and is the leading cause of diarrhoea associated with antibiotic use. Antibiotics used to treat other infections can trigger *C. difficile* disease and spore-mediated transmission, whereas antibiotics used to treat *C. difficile* infections are becoming less effective. Thus, *C. difficile* is challenging standard clinical practices and infection-control measures.

In collaboration with the Pathogen Genomics group we have begun to define the *C. difficile* phylogeny. Genomic analyses suggest that *C. difficile* has existed for millions of years, and that antibiotic-resistant forms have arisen independently in different branches of its family tree on multiple occasions.

Building on this work, we are now using high-throughput whole-genome sequencing to track *C. difficile* transmission and genome evolution within hospitals. We expect this

approach to identify patterns of transmission and persistence that can be targeted for intervention. Ultimately we believe that genome sequencing technologies can be used for *C. difficile* surveillance and epidemiological studies at the local, national and international level.

We have developed a novel mouse infection model that mimics colonisation, disease and transmission in humans. We are now using this model to study host disease susceptibility and to develop novel therapies including a protective vaccine and bacteriotherapy – ‘probiotic’ inoculation with harmless bacteria to inhibit colonisation by *C. difficile*.



Dave Goulding, Genome Research Limited

C. difficile spore. Spores are the main route of transmission and are completely resistant to most healthcare disinfectants.



Genome Research Limited

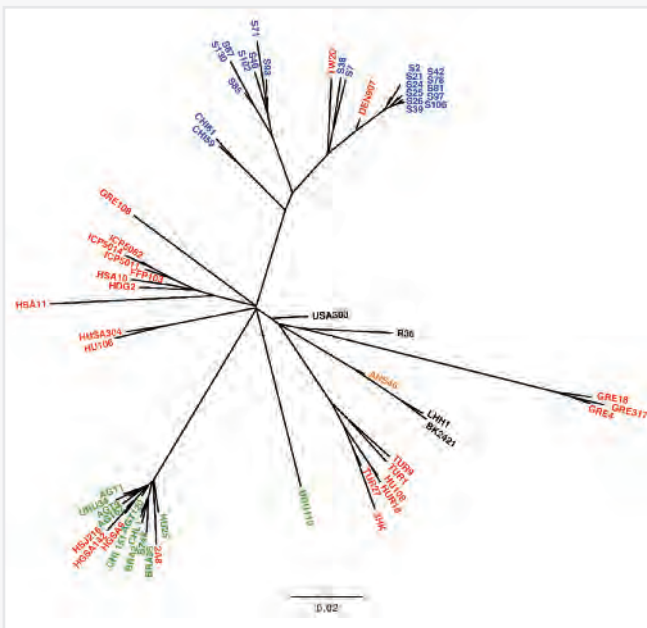
➤ ***C. difficile* is a genetically diverse species with multiple, independent lineages causing serious disease in humans and animals.**

He M et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci USA* 2010; 107;16;7527-32.

➤ **Oxidising disinfectants inactivate spores and block environmental spore transmission.**

Lawley TD et al. Use of purified *Clostridium difficile* spores facilitate the evaluation of healthcare disinfection regimens. *Appl Environ Microbiol.* 2010; 76(20): 6895-900.

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



Simon Harris, Genome Research Limited

Whole-genome studies have shown how an MRSA strain (ST239) was transmitted from Europe (red) to South America (green) and back again.



Wellcome Library, London

Whole-genome sequencing of 65 closely related isolates of MRSA identifies transmission events between continents and within a hospital.

Harris SR et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010; 327: 469-74.

Combining high-throughput sequencing and mutagenesis allows the analysis of knockouts of every gene in a bacterium in a single experiment.

Langridge GC et al. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* 2009; 19: 2308-16.

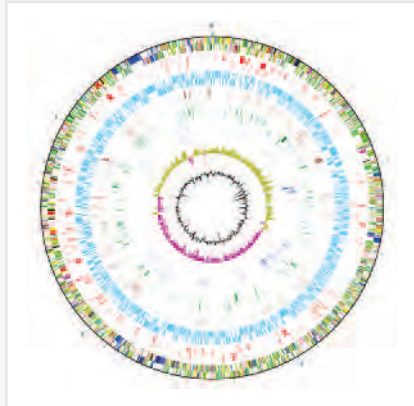
High-throughput sequencing allows us to study the biology and genetics of bacteria in unprecedented detail. Bacterial genomes are small, and new technologies can be used to sequence hundreds of bacterial genomes to very high accuracy. At the same time, we can use sequencing as an assay to follow gene regulation and identify gene function genome-wide.

Our first proof-of-principle was to look at 65 isolates of a single type of methicillin-resistant *Staphylococcus aureus* (MRSA), ST239, which has spread world-wide during the last 30 years. These isolates are almost indistinguishable by current typing techniques, but can all be resolved by whole-genome sequencing. This enabled us to track transmission routes across the world, for example from Europe to South America and back, and to differentiate strains taken from patients in the same hospital just weeks apart. These sequences also reveal the changes in the genome that have enabled strains to become resistant to antibiotics.

We are now expanding this approach to other organisms, including *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, *Bordetella pertussis*, *Salmonella*, *Shigella*, *Chlamydia* and many others. Alongside these population studies, we are using high-throughput phenotyping to look at drug resistance and growth requirements for up to 100 strains simultaneously, correlating phenotypic characteristics with whole-genome sequences.

We can also couple sequencing techniques to saturation mutagenesis to follow a million individual knock-out mutants in a single pool of cells, monitoring the function of genes in growth, infection and transmission in pathogens such as *Salmonella* Typhi.

As well as answering basic questions about biology, these techniques can be used to track infections in real time, to identify the cause of outbreaks and follow their spread. We plan to work with colleagues in the Health Protection Agency and in local hospitals to pilot the use of these technologies in clinical investigations.



A high-throughput study using a million mobile genetic elements knocked out every single *Salmonella* Typhi gene. The map shows the genes that are essential (red) and dispensable (blue) for growth.

Langridge GC et al. *Genome Res* 2009; 19: 2308-16. doi: 10.1101/gr.097097.109

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.



CDC/Im Gathany

Malaria kills more than a million people a year, mainly young children.

Malaria parasites of rodents, such as *Plasmodium berghei*, can be studied in mice and in mosquitoes. Robust tools to modify the parasite's genome allow us to study their basic biology and to find out how the parasites interact with their host and their vector. They also give us a unique opportunity to ask which biochemical pathways the parasites require *in vivo* at different stages of their transmission cycle. On the basis of this knowledge it will be easier to decide which molecular targets new drugs should aim to disrupt.

To test this idea we studied a family of 66 *P. berghei* protein kinase genes, all but one of which are also found in human malaria parasites. Human protein kinases are important targets for cancer drugs, and parasite kinases may be good targets for antimalarial drugs. However, identifying the right candidates within this large gene family is a major challenge.

Through a gene knockout screen, we showed that more than one-third of *P. berghei* kinases are not required for the asexually reproducing blood stages, which in humans cause all the pathology of malaria. By studying the transmission of 23 mutants through the parasite life cycle, we identified key regulators of mosquito transmission. We can also now predict that the best primary kinase targets to treat blood-stage infections will be found among the group of genes that we failed to knock out, most probably because they are essential for parasite development in the blood.

We believe it will be possible to use genetic modification on a genome-wide scale to study parasite biology and identify drug targets. During the next year we will aim to improve the efficiency of experimental genetic technologies further and to provide the research community with new tools and resources for the genetic manipulation of malaria parasites.



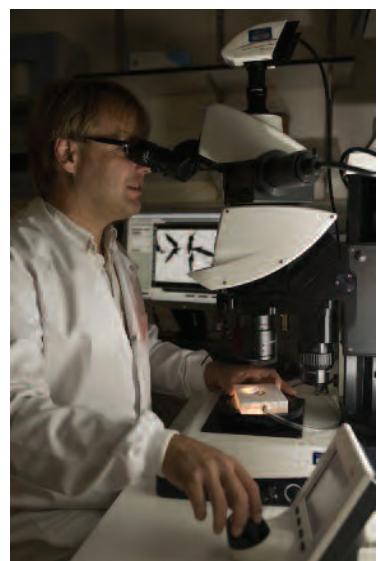
Wellcome Library, London

A large gene knockout study identifies functions of parasite protein kinases in malaria transmission.

Tewari R et al. The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* 2010; 8(4):377-87.

A set of genes controlling the development of malaria parasites through an intracellular messenger.

Moon RW et al. A cyclic GMP signalling module that regulates gliding motility in a malaria parasite. *PLoS Pathog* 2009; 5(9): e1000599.



Wellcome Library, London

Using rodent malaria allows us to study the entire parasite life cycle under laboratory conditions.

We aim to understand the clinical and biological consequences of genome diversity in *Plasmodium*, *Anopheles* and human populations, and to translate this knowledge into practical applications for malaria control.

MalariaGEN

MalariaGEN website. African populations are the world's genetically most diverse, creating significant challenges for genome-wide association studies.

A major obstacle to the prevention of malaria is the continual evolution of *Plasmodium* populations in response to selective pressures from the human immune system and antimalarial drugs. To help establish a global surveillance system to provide early warning of clinically or biologically important changes in parasite populations, we have used next-generation sequencing to identify 86 000 single nucleotide polymorphisms (SNPs) across the *P. falciparum* genome, and have mapped variation in SNP frequencies in hundreds of samples from Africa, Asia and Oceania.

These geographical comparisons offer important insights into biological determinants of evolutionary selection in the parasite. For example, the genes that determine parasite fertility when it undergoes sexual reproduction within mosquitoes differ markedly between populations. We are also using deep sequencing data to assess genetic diversity

of parasites within individual infections and to estimate levels of inbreeding within local parasite populations, which is important for understanding the spread of drug resistance.

We are working to develop ways in which clinicians and researchers in malaria-endemic countries can analyse patterns of genome variation in the samples that they have collected. We have developed a web application, MapSeq, that includes user-friendly tools for visualisation and analysis of *Plasmodium* genome variation data. We hope it will encourage researchers in malaria-endemic countries to collect samples for sequencing, and take the lead in applying these data to problems of local clinical importance.

In our work on human genetic resistance to malaria, we have now completed our analysis of genome-wide association data from more than 12 000 individuals from three African populations. The next step is



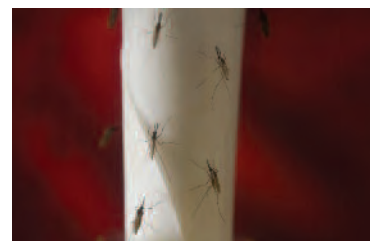
Genome Research Limited

➤ **A guide to the methodological approaches needed to carry out genome-wide association studies in Africa.**

Teo YY et al. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* 2010; 11:149–60.

➤ **A data release policy, developed by consultation with the MalariaGEN network, that balances the interests of researchers in rich and resource-poor countries.**

Parker M et al. Ethical data release in genome-wide association studies in developing countries. *PLoS Med* 2009; 11:e1000143.

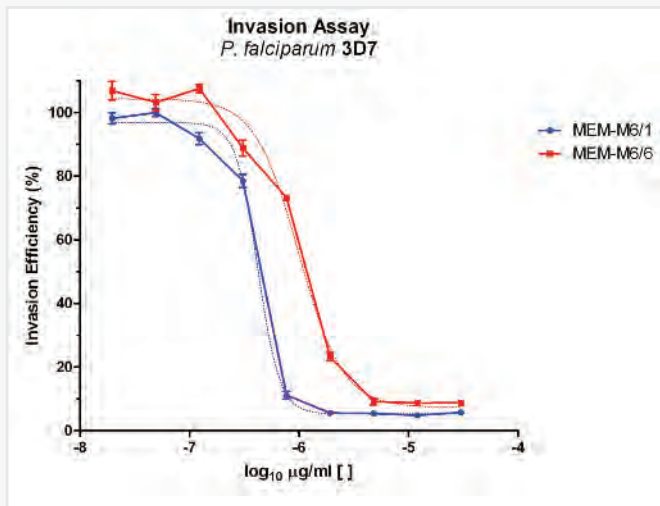


Wellcome library, London

Mosquitoes in Sanger Institute malaria laboratories. The *Plasmodium* genome shows signatures of selection in response to the human immune system.

to pool these data in a meta-analysis, which will require statistical imputation based on data being generated by the 1000 Genomes Project. This will be followed by large-scale replication studies across the 14 malaria-endemic countries involved in the MalariaGEN consortium.

We use molecular and cellular tools to study host-parasite interactions during the erythrocytic stages of the *Plasmodium falciparum* life cycle that could be targeted by new anti-malarials, with a particular focus on erythrocyte invasion.



Theron M et al. Cytometry A 2010; 77(11):1067-74.
doi: 10.1002/cyto.a.20972

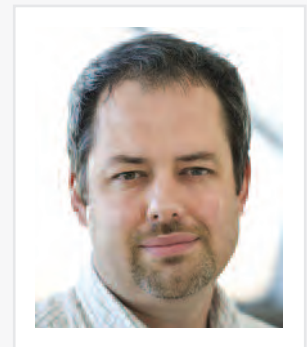
Plasmodium falciparum parasites can invade a human red blood cell in under a minute, and the process is driven entirely by the parasite.

Erythrocyte invasion by *Plasmodium falciparum* parasites is influenced by natural genetic variation in both humans and parasites. To measure the impact of these genetic differences, we have developed a new high-throughput approach that uses two-colour fluorescent labelling and flow cytometry to quantify erythrocyte invasion. We are now applying the method to multiple *P. falciparum* isolates that are also being sequenced using next-generation approaches as part of Dominic Kwiatkowski's *P. falciparum* genetic diversity project. Our goal is to combine the genotype and phenotype data to identify genes that influence the invasion process.

We are also exploring new experimental approaches to understand erythrocyte invasion at a molecular level. With Gavin Wright's team, we are using AVEIXIS (avidity-based extracellular interaction screen) technology to perform the first large-scale screen for receptor-ligand interactions between *P. falciparum* and the erythrocyte. In collaboration with Oliver Billker and the Sanger Institute proteomics core facility led by Jyoti Choudhary, we are using mass spectrometry to study two post-translational modifications that play key roles in regulating erythrocyte invasion – phosphorylation and palmitoylation.

Finally, in collaboration with researchers at the University of Alabama at Birmingham, we have played a leading role in a comprehensive genetic study of *P. falciparum*-related parasites infecting wild apes in central Africa. This study established that *Plasmodium* infection is widespread among gorillas and chimpanzees, with up to 50 per cent of individuals in some communities infected. At least six apparently independent *P. falciparum*-related species exist, and apes are often infected with multiple species simultaneously.

All human *P. falciparum* parasites fall within a single clade of parasites infecting western lowland gorillas, arguing that *P. falciparum* passed from gorillas to humans at some time during our evolutionary history. Further studies are aimed at understanding whether such cross-species transmissions still occur and what molecular characteristics control which host species a *Plasmodium* parasite can infect.



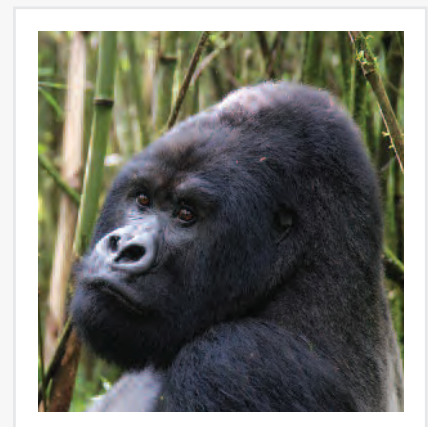
Wellcome Library, London

➤ **A study reveals wild apes are infected by multiple *Plasmodium* species, with human *P. falciparum* most closely related to a parasite of western lowland gorillas.**

Liu W et al. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 2010; 467: 420-5.

➤ **A new high-throughput method to quantitate the invasion of human erythrocytes by *P. falciparum* merozoites.**

Theron M et al. An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by *Plasmodium falciparum* parasites. Theron M et al. Cytometry A 2010; 77(11):1067-74.



stock.xchng

Plasmodium falciparum parasites, which are responsible for the majority of human deaths from malaria, originated in western gorillas.