WEB WATCH

Malaria on the web

- PlasmoDB: http://PlasmoDB.org
- MR4: http://www.malaria.mr4.org

This October, a new version of the PlasmoDB database — PlasmoDB 4.0 — was launched to mark the publication of the complete genome sequence of the malaria parasite *Plasmodium falciparum*.

PlasmoDB 4.0 provides a single point of access to a wide range of genomics data on *P. falciparum*. Curated annotations of the parasite's genome sequence are available here, together with raw sequence data and automated predictions of gene and protein structures. Also available here are population-specific SNP data, genetic and optical maps, and microarray- and SAGE-expression, as well as proteomic, data. PlasmoDB also provides a suite of analysis tools, such as tools for finding protein motifs or for analysing expression data. The secret behind the complexity of this database is that it is not really a database but a single Web interface that sits atop a relational database, which houses and integrates all of the above-mentioned data, and more.

MR4 (the Malaria Research and Reference Reagent Resource Center) is a different kind of community database altogether. Developed since 1997, it provides quality-controlled malaria-related reagents and information to the international malaria research community — such as: various *Plasmodium* and *Anopheles* species; human reagents, such as sera and cells; molecular biology reagents, such as libraries and probes; and *Plasmodium* recombinant proteins and antigens. These resources are provided free of charge to malaria researchers on registration. For a more genomic view of the *Anopheles gambiae* genome, go to EBI or NCBI, where annotated versions of this genome are available.

Jane Alfred

GENOMICS

Malaria – the enemy's battle plan

Every day, at least 2,700 children die from an infection by the human malarial parasite *Plasmodium falciparum*. Almost all of these cases are in sub-Saharan Africa, where, if the children live through their earliest infections, they might go on to fight up to 40 separate malarial infections in their lifetime. Malaria is the third leading killer of humans, and increasing resistance of the parasite to current drugs is alarming. Many scientists have devoted their lifetimes to combating this scourge, and they've just been handed their enemy's battle plan: its genome.

After six years of difficult sequencing and assembly, the Malaria Genome Project published the mostly complete sequences of two species of *Plasmodium* — *P. falciparum*, the human parasite, and *P. y. yoelii*, a rodent parasite — in the October 3 issue of *Nature*. The extremely high A+T content, more than 80%, of both these 23-Mb genomes was often arranged in runs of 50 As or Ts, embodying a genome assembler's worst nightmare. A partial solution, also used for other (A+T)-rich genomes, was to separate most of the 14 chromosomes and sequence each separately.

Of course, of most interest are potential drug or vaccine targets. About 5,300 genes were identified in *P. falciparum*, and ~60% of those showed no significant similarity to previously known proteins. Of particular therapeutic interest are the 10% of all proteins targeted to the apicoplast, a parasite-specific subcellular compartment that is essential for its survival but of unknown function. Gardner *et al.* suggest that several apicoplast metabolic pathways, such as isopentyl diphosphate biosynthesis, should be investigated further for antimalarials. The need for developing new therapeutics is urgent, as all current treatments were developed between 50 and 4,000 years ago.

Florens *et al.* and Lasonder *et al.* looked at proteins expressed at different stages of the parasite's life cycle, a task made more intriguing by the fact that over half of the proteins were hypothetical. They focused on novel cell-surface proteins and secreted proteins, which, if parasite specific, would make excellent drug or vaccine targets.

Of course, the bonus to malaria researchers is doubled with the sequence of the insect vector *Anopheles gambiae* (see Highlight below). Combined with the human genome, we now have the

GENOMICS

Deadly vector unveiled

Malaria has been with us for thousands of years — Alexander the Great and Oliver Cromwell are thought to have been among its victims. Mosquito resistance to insecticides has thwarted efforts to control the disease. But the limited understanding of mosquito biology that made us lose this battle is set to change with the publication in *Science* of the genome sequence of *Anopheles gambiae*, the principle vector of the malaria parasite *Plasmodium falciparum*. This sequence and its accompanying analysis, together with the publication of the *P. falciparum* genome (see above Highlight) provide vitally important information in the fight to control malaria transmission.

The combined efforts of Celera, GenoScope and TIGR produced a tenfold shotgun coverage of the *A. gambiae* genome. At 278 Mb, it is more than twice the size of the fruitfly genome, mainly because of the expansion of non-coding DNA. The *A. gambiae* genome is also highly variable. Most of the variable regions in the sequenced strain fall into two haplotypes, suggesting that it derives from two genetically distinct populations. This variation might be of use in the future — for example, SNP markers should allow malaria researchers to follow the evolution and the spread of insecticide resistance genes.

The analysis of the genome revealed several key features of the *A. gambiae* genome. Gene-prediction algorithms indicate that *A. gambiae* has ~15,000 genes. A comparison between mosquito and fly genomes and proteomes revealed that, despite considerable similarities, there were significant differences — a sign of adaptation to different life strategies and a target for altering the mosquito's life cycle and its vector properties.

Other reports in this issue of *Science* include an analysis of immunity-related genes, and of those that

blueprints for all members of this deadly play. Thus, in one historic week, we've entered the post-genomic period in malaria research, full of challenges but with incredible promise for future preventative measures.

***Chris Gunter, Associate Editor,* Nature**

References and links

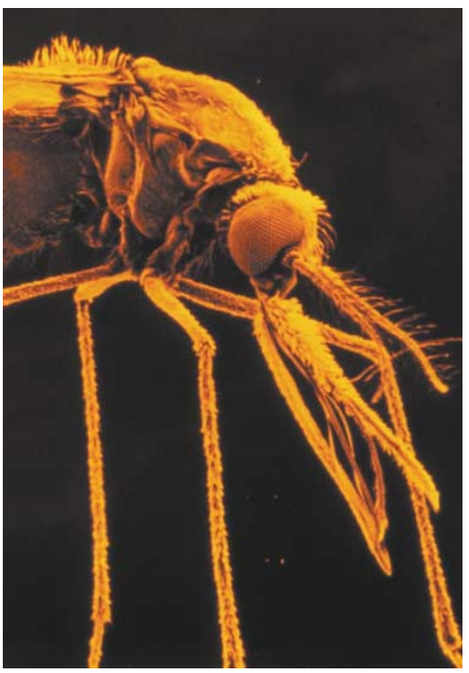
ORIGINAL RESEARCH PAPERS Gardner, M. J. *et al*. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002) | Florens, L. *et al*. A proteomics view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002) | Lasonder, E. *et al*. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002)

FURTHER READING Carlton, J. M. *et al*. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002) | Hall, N. *et al*. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9, and 13. *Nature* **419**, 527–531 (2002) | Gardner, M. J. *et al*. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11, and 14. *Nature* **419**, 531–534 (2002) | Hyman, R. W. *et al*. Sequence of *Plasmodium falciparum* chromosome 12. *Nature* **419**, 534–537 (2002)

WEB SITES

TIGR microbial database: http://www.tigr.org/tdb/e2k1/pfa1

***Nature* malaria web focus:** http://www.nature.com/nature/malaria



such as its response to the malaria parasite, taste for the human host and its reproduction. Which of them should be targeted to break the malaria life cycle remains to be seen.

A wealth of information on genome structure, evolution and function, and on insect biology can be found in this collection of papers. The direct application of these, and future, findings to the control of malaria is the next pressing goal.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPERS Holt, R. A. *et al.* The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002) | Zdobnov, E. M. *et al.* Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002)

FURTHER READING Christophides, G. K. *et al.* Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**, 159–165 (2002) | Riehle, M. A. *et al.* Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* **298**, 172–175 (2002) | Hill, C. A. *et al.* G protein-coupled receptors in *Anopheles gambiae*. *Science* **298**, 176–178 (2002) | Ranson, H. *et al.* Evolution of supergene families associated with insecticide resistance. *Science* **298**, 179–181 (2002)

WEB SITE

Mosquito Genome Browser: http://www.ensembl.org/Anopheles_gambiae

encode neuropeptides, peptide hormones and G-protein-coupled receptors. These gene products are involved in different aspects of the mosquito's biology and behaviour,

IN BRIEF

HUMAN GENETICS

Paternal inheritance of mitochondrial DNA.

Schwartz, M. & Vissing, J. *N. Engl. J. Med.* **348**, 576–580 (2002)

Mammalian mitochondrial DNA (mtDNA) is strictly maternally inherited, or so it was thought.The authors report a novel, sporadic mtDNA mutation that is present on a paternal mtDNA haplotype in a man with mitochondrial myopathy. The patient's muscle mtDNA carries this haplotype but his blood mtDNA is maternally derived. This sporadic mutation might have conferred an early and selective proliferative advantage on paternal mitochondria, which are normally eliminated during early embryogenesis.

CIRCADIAN RHYTHMS

Mutations in *Rab3a* alter circadian period and homeostatic response to sleep loss in the mouse.

Kapfhamer, D. *et al. Nature Genet.* **32**, 290–295 (2002)

Kapfhamer *et al.* identified the mouse mutant *earlybird* in an ENU screen for circadian mutants. These mice have a short circadian period that is probably caused by a point mutation in *Rab3a*, which reduces Rab3a levels in homozygous mutant mice by 73%. In support of this, Kapfhamer *et al.* also identified circadian and sleep abnormalities in *Rab3a*-null mice. Rab3a is an abundant brain protein that functions in synaptic vesicle transport, so these findings indicate that Rab3a-mediated synaptic plasticity might underlie circadian behaviour and sleep homeostasis.

FUNCTIONAL GENOMICS

Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags.

Chen, J. *et al. Proc. Natl Acad. Sci. USA* **19**, 12257–12262 (2002)

Identifying the number of genes in the human genome is an important challenge. Many short cDNA tags that have been identified by serial analysis of gene expression (SAGE) have no known match among the genes identified by conventional means. Experimental and SAGE database analysis revealed that most unmatched SAGE tags are truly novel, probably corresponding to new transcripts or previously unknown genes.

NEUROGENETICS

Genetic and physiological data implicating the new human gene *G27* and the gene for D-amino acid oxidase in schizophrenia.

Chumakov, I. *et al. Proc. Natl Acad. Sci. USA* **99**, 13675–13680 (2002)

Chumakov *et al.* genotyped 191 SNPs in a 5-Mb region of 13q22–q34 in a Canadian sample, identifying markers in two regions that showed association with schizophrenia. Two markers in one of these regions also showed association in a Russian sample. Sequencing in this region identified two novel genes: one of them (*G72*) interacted physically with D-amino acid oxidase (DAAO), which is linked to the activation of the NMDA-type glutamate receptor. Four SNPs from DAAO also showed association with schizophrenia in the Canadian sample.