MALARIA: THE PAINFUL FACES ASK, CAN WE NOT CURE?

B. MISHRA

JANUARY 19, 2012

[1]

[1] (Professor of Computer Science, Mathematics, Cell Biology, Courant Institute and NYU School of Medicine; Visiting Scholar, Cold Spring Harbor Laboratory; Adjunct Professor, Tata Institute of Fundamental Research; Professor of Human Genetics, Mt Sinai School of Medicine)

Curse Against Curse:

Suppose a choice is presented, that by sacrificing one out of every hundred children from one's future generation, a cruel and capricious force can be kept at bay; it might be possible to stop it from fully unleashing its devastating and uncontrollable fury. Is it ever possible to acquiesce to it?

What if that force is malaria, which continues to cause more than half a billion cases of fever and several millions of deaths annually? The answer appears to have been "yes." Tens of thousands of years ago, humans seem to have consented to a cruel covenant of this kind. Often, in textbooks, such an agreement is presented as a quirky example of how strange population genetics can be — a phenomenon that goes by the name of "*heterozygotes advantage*."

Sickle-cell anemia (SCA) [2], [3] is a recessive genetic disorder and is caused by the presence of a mutant genomic allele, variant a; the normal version of the genomic variant is the major allele, variant A. Since an individual in a population will have two copies of a gene (one from each parent), there are *three* choices: the individual may have two mutant alleles (homozygous in minor allele, aa), one of each allele (heterozygous, aA) or two good (wild-type) alleles (homozygous in major allele, AA). Since sickle cell anemia (SCA) is a recessive disease, the sickle cell anemia trait is fully expressed in the individuals with two copies of the SCA (sickle cell anemia) mutant alleles (aa). The patients' red blood cells, when exposed to low-oxygen conditions, lose their normal spherical shape, become sickle-shaped and get stuck to the capillaries. The patients are deprived of oxygen and their internal organs fail under repeated assaults of the disease. They suffer from anemia with an increased risk of stroke. For the patients, these painful bouts of oxygen deprivation end with premature deaths.

Why is this tolerated? Why hasn't Darwinian selection purified the population of such a scourge? A back-of-the-envelope calculation is in order (at least to see that no "intelligent design" is at play here). If the frequency of the minor allele is 10% (1 a-allele individual for every 9 A-allele individuals in the population), then

[2] Serjeant, GR and Serjeant, BE. *Sickle Cell Disease*. Oxford University Press, 2001.

[3] Allison, AC. The sickle-cell and Haemoglobin C genes in some African populations. *Ann. Human Genet.*, 21:67–89, 1956.

one child in hundred will be homozygous SCA mutant (1% of the population) and will almost surely be lost to full-blown sickle cell anemia. However, some 18% of the population will be heterozygotes (one of each allele, aA) and will not exhibit the sickle cell anemia trait severely enough to be harmful. Fortunately, they will also be immune to malaria. However they are the carriers of the SCA traits, and if two such SCA-carriers have a child there is a 25% chance that the child will have SCA (and 50% chance that the child will be an unaffected SCA-carrier). The remaining 81% of the population will be normal and will have the usual risk of dying after contacting malaria.

Without SCA, if for instance 10% of the population is expected to have died of malaria, then with SCA, the population has traded off this risk for a portfolio of 1% sure death from SCA and 8.1% expected death from malaria — a saving of 0.9% from the malarial death's jaws. Thalassamia is yet another example of heteryzotes advantage, almost as tightly bound to malaria as sickle cell anemia — except thalassmia affects its patients by failing to synthesize one of the globin chains that make up hemoglobin.

In India, Sickle Cell Anemia occurs in some populations with a rate of about 0.75%, mainly in tribal regions. For instance, in the tribal population of Gujarat (numbering about 6.5 millions) about 48,700 individuals are suspected of suffering from sickle cell disease; the number of SCA-carriers is about 647,000 (about 10%).

Million Murders

A question arises: Are there no other ways of confronting this relentless force? Malaria: "the unseen, small, but million-murdering cause," as Sir Ronald Ross wrote in 1895 ("In this, O Nature, yield I pray to me").

Seven years later, Sir Ronald Ross [4], a dedicated and diligent scientist, was to receive the Nobel Prize (for physiology or medicine) in 1902 — the second one to receive this honor. Ross's citation for the Nobel Prize in Physiology or Medicine (1902) read: Ronald Ross received the Prize "*for his work on malaria, by which he has shown how it enters the organism and thereby has laid the foundation for successful research on this disease and methods of combating it*."

In 1882, Ross started to work on malaria at the Presidency General Hospital, Calcutta (now Kolkata), West Bengal, India. To study mosquitoes and what role they play in spreading malaria, he built his own laboratory at Mahanad village, which he visited frequently focusing on a wide variety of mosquitoes in Mahanad as well as in adjoining villages. Next year, in 1883, he moved to Bangalore (now Bengaluru), Karnataka, India, where he was

[4] Aronson, SM. The man who understood mosquitoes. *Med Health R I.*, 79:351–352, 1996.

posted as the Acting Garrison Surgeon. His attention there focused primarily on strategies for managing mosquito populations by controlling their access to water, while it still remained a mystery how to connect the relation between malaria, mosquitoes and marshes. Ross, who would be called a century later "*the man who understood mosquitoes*," proposed also an epidemiological (mathematical) model of malaria transmission with human hosts and mosquito vectors. This model would be incorporated into the classical Ross-Macdonald Model — still the basis of understanding how to control malaria.

Much later, around 1897, when Ross was working at Osmania University's medical school in Secunderabad, Andhra Pradesh, India, his attention moved to malaria pathogen. By artificially feeding mosquitoes on a malaria patient (a Mr. Hussain Khan), he discovered the presence of the malarial parasite within the Anopheles species of mosquito. In 1899, Ross went back to Britain and joined the Liverpool School of Tropical Medicine as a professor of tropical medicine. In 1911 he was elevated to the rank of Knight Commander of the Order of Bath. In 1932, Ross died at the Ross Institute, London.

Postgenomic Plasmodium

Because of Ross's research, finally, Malaria was demystified, however, not defeated. Malaria appears to have existed for many millennia. The *Charaka Samhita* and *Sushruta Samhita* [5], dating back to the 6th century BCE, document Malaria, list the main symptoms as fever and enlarged spleens, and implicate insect bites. In the meantime in western medical canons, it remained much less understood – it had gone by the names Roman Fever, Marsh Fever, Ague (Acute Fever), etc. and had been associated with marshland and swamps. Until Ross, not much else seems to have been described about its mode of propagation, as the name malaria itself indicated: *mala aria* — "bad air" in medieval Italian.

Ross's work had brought a focus on the malaria parasite, *Plasmodium*, and how it is transmitted by the mosquitoes. This human pathogen, *Plasmodium*, is thought to have coexisted along with humans, for almost the entirety of the species' history. Close relatives of human malaria parasites occur also in other primate species, e.g., chimpanzees and gorillas.

Plasmodium is a genus of parasitic protists (eukaryotic microorganisms with nucleus containing genomic DNA) and was first described in 1885 by Ettore Marchiafava and Angelo Celli. There are about 200 species of *Plasmodium*, of which about 11 species are responsible for human malaria [6]. Other species infect other animals, including monkeys, rodents, birds, and reptiles. *Plasmodium*

[5] Rudolf Hoernle, AF. *Studies in the Medicine of Ancient India*. Clarendon Press, 1907.

[6] Roy, SW and Irimia, M. Origins of human malaria: rare genomic changes and full mitochondrial genomes confirm the relationship of *Plasmodium falciparum* to other mammalian parasites but complicate the origins of *Plasmodium vivax*. *Mol Biol Evol.*, 25:1192–1198, 2008.

vivax causes a form of malaria in humans, which is characterized by periodic fevers, and usually treatable by quinine. However, a more deadly species in this genus, *Plasmodium falciparum* is more difficult to treat and frequently results in rapid deterioration and death of the patients. It infects hundreds of millions of humans a year (accounting for 90% of malaria related death), with about 120 million in India.

Responding to the growing global risks due to *P. falciparum* and with the goal of developing new means to treat or prevent malaria, in 1993, the *Malaria Genome Consortium* was formed. The consortium consisted of laboratories distributed throughout the world and established the *Wellcome Trust Malaria Genome Mapping Project*. Using a novel mapping technology, six years later, a multidisciplinary team of biologists, chemists, computer scientists and mathematicians at the University of Wisconsin-Madison and New York University produced the first, large-scale map of the *Plasmodium falciparum* genome. The Wisconsin-New York team used an unusual physical mapping technique, now referred to as "shotgun optical mapping" method [7], [8], [9], [10], [11], [12]. To perform optical mapping, large chunks of *P. falciparum*'s genomic DNA were immobilized on plates and cut up by site-specific enzymes in situ. A laser technique, image processing software and sophisticated statistical inference algorithms then picked out important markers in every imaged fragment, and assembled millions of pieces of information into a map of markers, as they would be distributed along the *P. falciparum*'s whole genome. Using these markers, one could tell how one species of *Plasmodium* differed from another, how they might have evolved, and whether there are regions common to multiple species that can be targeted by one (or small number of) universal malaria vaccine(s).

But to understand exactly how the malaria parasite functions, one would need more. To know and understand the function of all its genes, it helps to have the base-by-base (e.g., A, T, C, G) sequence content of its DNA. Fortunately, once a physical map (like an optical map) exists, one can scaffold to the map short DNA sequence reads from malaria's genome, and obtain in a relatively straightforward manner its entire genomic DNA sequence, which can be stored in a computer as a long string of four bases of A, C, G, and T. In October 2002, in a special issue of *Nature*, Hall et al. [13], Hyman et al. [14] and Gardner et al. [15], finally published the genome sequence of all 14 chromosomes of *Plasmodium*, using a shotgun sequencing approach, but validating the results with the optical maps. Optical maps played similar critical roles in the sequencing of many other organisms (*E. coli*, *D. radiodurans*, etc.), and are likely to play important roles in the near future to create correct haplotypic human sequence assembly, a

[7] Mishra, B. *Encyclopedia of the Human Genome*, chapter Optical Mapping, pages 448–453. Nature Publishing Group, Macmillan Publishers Limited, London, UK, 2003.

[8] Lai Z, Jing J, Aston C, Clarke V, Apodaca J, Dimalanta ET, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Paxia S, Hoffman SL, Craig Venter J, Huff EJ, and Schwartz DC. A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nat Genet.*, 23:309–313, 1999.

[9] Jing J, Lai Z, Aston C, Lin J, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Tettelin H, Cummings LM, Hoffman SL, Venter JC, and Schwartz DC. Optical mapping of *Plasmodium falciparum* chromosome 2. *Genome Res.*, 9:175–181, 1999.

[10] Aston, C, Mishra, B and Schwartz, DC. Optical mapping and its potential for large-scale sequencing projects. *Trends Biotechnol.*, 17:297–302, 1999.

[11] Anantharaman, TS, Mishra, B, and Schwartz, DC. Genomics via optical mapping. III: Contiging Genomic DNA and Variations. *Proceedings 7th Intl. Cnf. on Intelligent Systems for Molecular Biology: ISMB '99*, 7:18–27, 1999.

[12] Anantharaman, TS, Mishra, B, and Schwartz, DC. Genomics via optical mapping. II: Ordered restriction maps. *J Comput Biol.*, 4: 91–118, 1997.

[13] Hall N, et al. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature*, 419:527–531, 2002.

[14] Hyman RW, Fung E, Conway A, Kurdi O, Mao J, Miranda M, Nakao B, Rowley D, Tamaki T, Wang F, and Davis RW. Sequence of *Plasmodium falciparum* chromosome 12. *Nature*, 419:534–537, 2002.

[15] Gardner MJ, et al. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature*, 419:531–34, 2002.

crucial but unmet need.

In the intervening nine years, since Plasmodium was sequenced, the field of genomics has moved forward with an exponential growth, in its own biotech's version of Moore's law. It is now possible to buy optical mapping machines (e.g., OpGen's Argus or a BioNanoMatrix machine), very-high throughput next-generation sequencing machines (e.g., Illumina's HySeq, or Life Technology's IonTorrent) and fast multi-core processors for genome mapping and sequencing, thus making it routine for a small laboratory to reproduce the work, which, only about a decade ago, took many scientists many months in a well-funded genome institute (e.g., BGI, Sanger or Broad Institute). New algorithms, software and platforms now exist enabling whole-genome sequence assembly, while self-validating the assembly using long-range information (e.g., optical maps) [16], [17], [18], since these new algorithms use sophisticated branch-and-bound approaches to constrained global optimization. Grass-root genomics appears to be on the verge of sprouting up in gazillion garages across the globe.

Based on the sophisticated genomic analyses, we now know that *Plasmodium falciparum*'s genome consists of 14 chromosomes in its nuclear DNA with sizes ranging between 25-30 Mb (strain 3D7 has a size of 23.3 Mb) with an A+T content of 81%. Additional genetic materials exist in its mitochondria (linear repeat of 6 Kb) and apicoplast (circular plastid-like structure of 35 Kb). The two organelles (mitochondria and apicoplast) are thought to have resulted from ancient endosymbiotic events, thus making the unique structure of apicoplasts an attractive target for antimalarial rational drug design.

Each of *P. falciparum*'s chromosomes is made up of a very long stretch of DNA, the molecule that carries the genetic code and orchestrates how the parasite metamorphoses in mosquitoes and humans. The chromosomes range in size between 0.75 to 3.5 Mb in size. The genomic DNA is long double-stranded polymer composed of four bases A (adenine), C (cytosine), T (thymine) and G (guanine). Small substrings of this long string of letters construct the genes, which are transcribed into RNA and then finally translated to proteins. Controlled by regulatory and metabolic processes, the proteins carry out the biological functions that the parasite needs at different stages — some during its life in mosquitoes and some in humans. These genes can undergo mutations, subtly changing the function of the gene (its genotype), and giving the parasite ability to change its overall functional abilities (its phenotype). Even though most random mutations could only have either deleterious or no effect, using persistently many Darwinian trial-and-error steps, the parasite has the ability to evolve and become drug-resistant, or adapt to different human (or non-

[16] Menges F, Narzisi G, and Mishra B. TOTALRECALLER: Improved Accuracy and Performance via Integrated Alignment & Base-Calling. *Bioinformatics*.

[17] Narzisi G, and Mishra B. Comparing de novo genome assembly: the long and short of it. *PLoS One*.

[18] Narzisi G, and Mishra B. Scoring-and-unfolding trimmed tree assembler: concepts, constructs and comparisons. *Bioinformatics*, 2: 153–160, 2011.

human) or mosquito populations. Perennially, running against a Red Queen race, *Plasmodium* evolves rapidly using its highly mutable AT-rich genome. This evolvability of *Plasmodium* makes it an elusively obscure but obstinate adversary.

The malaria parasite *Plasmodium falciparum* exhibits a complex metamorphic life cycle involving a mosquito vector and a human host, since it must adapt to two drastically different environments during its life. Once the infection is initiated via sporozoites injected with the saliva of a feeding mosquito, *P. falciparum*'s major life cycle phases in humans commence. These phases are: liver stage, blood stage, sexual stage, and sporogony. Young female Anopheles mosquitoes ingest the malaria parasite by taking a blood meal from an infected human malaria-parasite-carrier (either symptomatic or asymptomatic) and carry *Plasmodium* sporozoites in their salivary glands for rest of its life (about 29 days). The parasite gametocytes in mosquitoes further differentiate into male or female gametes and fuse in the mosquito's gut. This phase produces an ookinete that penetrates the gut lining and produces an oocyst in the gut wall. When the oocyst ruptures, it releases sporozoites that migrate through the mosquito's body to the salivary glands, ready to infect a new human host. Mosquitoes inject sporozoites into humans, where they start a new life by migrating to the liver and developing via asexual reproduction. Transformed into Merozoites, they invade human patients' blood cells, bursting them.

The blood stage is characterized by a number of distinct and carefully programmed substages, which include the ring, trophozoite and schizont; these are referred to collectively as the intraerythrocytic developmental cycle (IDC). Several systems biological studies have carefully analyzed the IDC of P. falciparum [19], [20], [21], [22]. The new systems biology models along with *P. falciparum*'s genome sequence allows one the opportunity to gain further insight into the role of *P. falciparum*'s approximately 5,400 genes and the proteins they encode, the majority of whose biological functions remain unknown. It has been shown that a large percentage of the genome is active during the intraerythrocytic developmental cycle (IDC) and that the regulation pattern is such that as one set of genes is deactivated, another is being turned on, causing the so-called "continuous cascade" of activity, in which transcriptional regulation (genes encoding proteins that activate other genes) is controlled in a tightly timed choreography. Traditional approaches to understand the structure of the temporal relations among these key processes have been difficult, and usually required tedious manual intervention. In recent works of a systems biology team from New York University, it was demonstrated how model checking techniques and statistical inference could recon-

[19] Mitrofanova A, Kleinberg S, Carlton J, Kasif S, and Mishra B. Predicting malaria interactome classifications from time-course transcriptomic data along the intraerythrocytic developmental cycle. *Artif Intell Med*, 49: 167–176, 2010.

[20] Kleinberg S, Casey K, and Mishra B. Systems biology via redescription and ontologies (I): finding phase changes with applications to malaria temporal data. *Syst Synth Biol*, 1:197–205, 2007.

[21] Llinás M, Bozdech Z, Wong ED, Adai AT and DeRisi JL. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res*, 34: 1166–1173, 2006.

[22] Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J and DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*, 1, 2003.

struct the main features of the system, including the cascade of gene expression, as well as the stages of the intraerythrocytic developmental cycle (IDC) and their associated processes. Powerful formal methods that combined information theoretic techniques developed by engineers with the redescription theoretic techniques of philosophers finally revealed how these cascades of gene expressions regulates complex host-pathogen interactions [23], [24], [25], [26], [27]. Similar approach has also been tried to understand other host-pathogen interactions with considerable success (e.g., anthrax infection analysis by GOALIE).

The problem of protein function prediction for P. falciparum, though crucial for developing vaccines, is even more demanding, where genetic and biochemical investigations are much more challenging, especially for parasites. For example, it is problematic to isolate a malaria parasite at various stages of its development (e.g., the life-cycle of P. falciparum is very rapid, ookinetes are difficult to isolate in large numbers, the liver stage of a parasite's development is hard to study because of technical difficulties, etc.). Such obstacles manifest themselves in a paucity of information on the protein properties, interactions, localization and motifs of *Plasmodium* species. Thus it is not surprising that more than half of *Plasmodium* proteins have still remained uncharacterized and therefore unusable for clinical trials. The task is further complicated by a rapidly morphing life cycle of the parasite, which allows for rapid evolutionary changes and diversity among related strains, thus making precise targeting of the appropriate proteins for vaccination a technical challenge. However, by capitalizing on the importance of the intraerythrocytic developmental cycle data and gene expression changes during its five phases, it has been possible to make small but significant progress in this direction (see Mitrofanova et al. [28]).

Extending the method of temporal analysis of gene expression profiles with protein-protein interaction data, sequence similarity scores, and metabolic pathway information it was possible to produce a set of predicted protein functions that can be used as targets for vaccine development. The researchers used a Bayesian statistical inference approach, which assigns a probability of having (or not having) a particular function to each protein, given the various sources of evidence. Their algorithm was able to assign meaningful functions to 628 out of 1439 previously unannotated proteins, which are first-choice candidates for experimental vaccine research.

A lot more remains to be done before reliable vaccine targets can be identified. Nonetheless, there is hope and excitement for unprecedented progress, primarily because of the availability of sophisticated biotechnology and information technology tools.

[23] Ramakrishnan N, Tadepalli S, Watson LT, Helm RF, Antoniotti M and Mishra B. Reverse engineering dynamic temporal models of biological processes and their relationships. *Proc Natl Acad Sci U S A*, 107:12511–12516, 2010.

[24] Mishra, B. Intelligently deciphering unintelligible designs: algorithmic algebraic model checking in systems biology. *J R Soc Interface*, 6:575–597, 2009.

[25] Tadepalli S, Ramakrishnan N, Watson LT, Mishra B, and Helm RF. Simultaneously segmenting multiple gene expression time courses by analyzing cluster dynamics. *J Bioinform Comput Biol*, 7:339–356, 2009.

[26] Mishra B, et al. A sense of life: computational and experimental investigations with models of biochemical and evolutionary processes. *OMICS*, 7:253–268, 2003.

[27] Antoniotti M, Policriti A, Ugel N, and Mishra B. Model building and model checking for biochemical processes. *Cell Biochem Biophys*, 38:271–286, 2003.

[28] Mitrofanova A, Kleinberg S, Carlton J, Kasif S, and Mishra B. Predicting malaria interactome classifications from time-course transcriptomic data along the intraerythrocytic developmental cycle. *Artif Intell Med*, 49: 167–176, 2010.

malaria:
the painful faces ask,
can we not cure? 8

Biotechnology can now provide tools for single molecule optical mapping, next-generation sequencing, high-resolution time-course gene-expression profiling, and proteomics and protein-protein interaction data. Such high-throughput data can be fed to information technology algorithms ranging from map and sequence assembly, rapid resequencing and alignment and temporal transcriptional regulation models to model checking, statistical inference and many other tools of systems biology. The anticipated deluge of data and models requires planning for a careful interpretation, especially, at population levels (involving humans, mosquitoes and malaria parasites).

Malaria: M.I.A.

By the late 19th century and subsequent to the industrialization and economic progress in the United States and Europe, malaria was nearly eradicated in many developed countries. As a result, malaria became relegated to the category of third-world diseases, receiving scant or no attention from pharmaceutical industries of the west. Nonetheless, as indicated earlier, malaria has still remained eerie, enigmatic and endemic to the most of the human population living in South-east Asia and Africa. India is one of 11 countries in South-East Asia with nearly 980 million people at risk — approximately one-third of the global population at risk. At present, India has the largest number of malaria cases occurring outside of Africa.

As India has prospered in recent years, it now possesses the ability to launch a "*third-world (but first-class) genomics*" effort against the third-world diseases (many mosquito-born and other infectious diseases). Starting with the work of Sir Ronald Ross, India's long history and leadership in attacking malaria are highly commendable, though less-than-adequate. Organized control programs, started in the 1940s, used DDT to control mosquitoes — nearly eliminating the disease by 1961. But that success proved to be short-lived, as malaria soon re-established itself in India. In 1998, WHO (World Health Organization), the World Bank and several charity organizations launched the Roll Back Malaria Partnership (RBM), a global initiative that coordinates actions against malaria. Roll Back Malaria, RBM, identified the following four major goals: (1) Reduce global malaria cases from 2000 levels by 50% in 2010 and by 75% in 2015; (2) Reduce global malaria deaths from 2000 levels by 50% in 2010, and to near zero by 2015; (3) Eliminate malaria in 8-10 countries by 2015; and eventually (4) Achieve eradication of malaria world-wide.

Classical epidemiological models for malaria can be created using the so-called SIR models based on three (ordinary differential)

equations (ODEs) [29]. SIR models are rooted in the earliest mathematical models of Daniel Bernoulli that estimate how an infectious disease, such as smallpox, spreads. Daniel Bernoulli, son of Johann Bernoulli and nephew of Jacob Bernoulli of the illustrious mathematical family of Bernoullis, was trained as a physician, and used his mathematical model to defend and advocate the practice of inoculating against smallpox. Daniel Bernoulli predicted that universal inoculation against smallpox could add 2 years and 2 months to the average life expectancy.

Starting in 1927, Kermack and McKendrick [30], [31], [32] developed a more general model in which they considered a fixed population consisting of three compartments, susceptible: S, infected: I, and recovered: R, the total adding up to a constant population size of $N = S + I + R$. Transitions can occur from the state S to state I or from state I to state R with certain probabilities. The expected number of susceptible individuals then reduces by a rate that depended on the product of S and I ($\dot{S} = -\beta SI$), as the infected individuals can infect other susceptible (but not yet infected) individuals. The expected number of recovered individuals would increase at a rate that depended on just I ($\dot{R} = \gamma I$). Two parameters (β and γ) determined these two rates, respectively. To keep everything consistent, expected number of infected individuals would then increase at a rate that is just the difference in the rate of susceptible individuals getting infected and the rate at which the infected individuals recover ($\dot{I} = \beta SI - \gamma I$). Each of these rates lead to a single ordinary differential equation. From these, one could calculate a reproduction number R_0, which predicts whether the infection leads to epidemics, $R_0 > 1$, or it dies out, $R_0 < 1$. In the bare-bone SIR model, R_0 turns out to be $\frac{\beta N}{\gamma}$. To control, the disease, either one has to reduce how fast an infected individual can infect others (e.g., through quarantine or vaccines) or one has to improve how quickly an infected individual can recover (e.g., through treatments). But the key number to watch is just one, R_0, often hovering around 1 — like a sword of Damocles.

To extend this model, as it would apply to malaria, one needs to add other terms to account for the mosquito (female Anopheles) population, and track the expected number of infected mosquitoes (Y), which would increase at the rate depending on the difference of two terms: one depending on the "bite rate" (r) multiplied by uninfected mosquito population and infected humans, and another term modeling the rate at which infected mosquitoes die out (d), i.e., $\dot{Y} = r(M - Y)I - dY$, where M is the constant population size of the female Anopheles mosquitoes. Sir Ronald Ross was the first one to propose such a model, which was later extended to Ross-Macdonald Model.

Yet another simple structural change to this epidemiological

[29] May, RL and Anderson RM. *Infectious diseases of humans: dynamics and control*. Oxford, Oxford University Press, 1991.

[30] Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. *Proc. R. Soc. Lond. A*, 115:700–721, 1927.

[31] Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity. *Proc. R. Soc. Lond. A*, 138:55–83, 1932.

[32] Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity. *Proc. R. Soc. Lond. A*, 141: 94–122, 1933.

model is critical, but, often swept under the rugs: a term for "*asymptomatic malaria-parasite-carrier population*" (*C*), an empirically disfavored state, as it is difficult to parameterize — it sits between the susceptible state (*S*) and the infected state (I) in the SIR. Varying from regions to regions and also by genetics and age (no report of gender bias), there seem to exist individual groups of different sizes, who, without exhibiting any clinical symptoms, are able to carry *P. falciparum* up to six months and multiple times. Nonetheless, they can infect other susceptible individuals, as the parasites can be transmitted from asymptomatic malaria-parasite-carriers to uninfected individuals by mosquito bites.

In some recent models, more suitable for the Indian context, additional modifications have been made to account for a growing human population of India, to permit recovered individuals to be repeatedly reinfected, to allow for an improving rate of treatment-recovery rate (in India), etc. As before, one can recalculate a more realistic reproduction number R_0, especially, for India.

Martcheva and Hoppensteadt from University of Florida and New York University proposed one such model recently [33] and fitted the model to the data for *P. falciparum* cases in India over the period 1983–2009. They discovered that the disease reproduction number has been reduced from $R_0 = 1.00732$ to 0.999457 as a consequence of the introduction of RBM (Roll Back Malaria) measures. The model suggested a reduction of 36% in P. falciparum malaria cases over a period from 2000 (11,40,000 cases) to 2010 (7,34,000 cases), which, though impressive, is far short of the desired reduction of 50%.

[33] Martcheva M and Hoppensteadt F. India's Approach to Elliminating emphPlasmodium falciparum Malaria: A Modeling Perspective. *J. Biol. Systems*, 18:867–891, 2010.

As suggested before, there is one problem with estimating the model parameters, since apparently healthy asymptomatic carriers of *Plasmodium* do not report their conditions — estimations, when they exist, are obtained only by pure chance (routine blood tests given to pregnant women). Since it would be impractical to test everyone, irrespective of any clinical symptom, for the *P. falciparum* density in blood, it is important to design subsampling schemes, which would be based on age and genetic markers associated with asymptomatic-malaria-parasite-carrier status. For this purpose, it would be highly desired that India invest in genome-wide associations studies (GWAS) to find related genetic variants (i.e., polymorphisms) and to develop simple genetic-test kits to check for the implicated biomarkers. So far, the genetic studies for this purpose have been unsatisfactory: (1) they have involved only low-resolution scans using few candidate genes, (2) no causal connection to specific genes have been inferred, and (3) no widely accepted standard exists for phenotyping (classifying an individual as an asymptomatic malaria-parasite-carrier), characterizing population background (with admixture), or accounting for en-

vironmental covariates. A locus on the q-arm of chromosome 5 (5q31-q33) has been implicated by multiple studies, and additional evidence for two other loci, p-arm of chromosome 5 (5p15-p13) and q-arm of chromosome 13 (13q13-q22), have also been presented, but these variants are likely to be telling less than the full story, since these studies have limited power, as reflected by modest LOD scores [34]. Better (higher resolution) analysis points to the need for large-scale genome-wide association studies (GWAS) with cases and controls.

Unfortunately, however, genome-wide association studies have only had a dismal record so far, heaping on the field an accumulating disreputation (derided in American press as "recreational genomics"). It may be speculated that this failure has resulted from several factors: poor haplotyping, unreliable population stratification, ad hoc phenotyping, and rather non-rigorous statistical analysis (involving multiple hypotheses testing). Indian genomicists may wish to take up this challenge and turn their attention to developing a complete genomics toolbox ab initio: it would contain tools for human haplotypic whole-genome sequence assembly (using new sequencers and optical mapping, for instance), polymorphism-mapping, population stratification and finally, genome-wide association studies. And that would be time and energy very wisely spent!

[34] Sakuntabhai A, Ndiaye R, CasadŐmont I, Peerapittayamongkol C, Rogier C, Tortevoye P, Tall A, Paul R, Turbpaiboon C, Phimpraphi W, Trape JF, Spiegel A, Heath S, Mercereau-Puijalon O, Dieye A and Julier C. Genetic determination and linkage mapping of *Plasmodium falciparum* malaria related traits in Senegal. *PLoS One*, 3, 2008.

Planning against Plasmodium

The malaria epidemiology models suggest essentially three ways of attacking the disease at a population level: (1) introduction of (medicated) mosquito nets to reduce the bite rate affecting the number of infected female Anopheles, (2) more effective mosquito control affecting the death rate of infected mosquitoes (especially with a better tracking of mosquito habitats via satellite imaging), and finally (3) early detection and prompt treatment of malaria, especially as the new strains of *P. falciparum* emerge. As hinted, characterization of asymptomatic malaria-parasite-carriers via accurate and well-designed genome-wide association studies will also be critical.

Furthermore, one will also need to create much better spatio-temporal (multipatch-) models (with spatial migration), which could be tracked over India in real-time using high-speed broadband networks. It would be also interesting to cost-effectively control how different centralized intervention schemes translate locally to modulate the key parameters of the model (e.g., a portfolio of budgets for mosquito-nets, mosquito-eradication and treatment may percolate differently in different urban and rural areas of India as the money filters through various governmental

and non-governmental bureaucracies). Some form of transparencies can be aimed at through Internet and other publications. Ability to rapidly translate effective intervention in one village to another could be achievable. Most importantly, if there are many genomic laboratories (mini CDCs) spread through out the country and connected electronically, new strains can be tracked quickly. The parsites' lifecycles, infectiveness, and drug- or treatment-resistance can be better understood to suggest the most optimal way to mobilize resources to stop a new but avoidable epidemic.

It should be noted parenthetically that the structure proposed here could be easily translated into a more general bio-surveillance scheme for India as well as internationally. One hopes that the plan against *Plasmodium* would fit well with the goals of India's National Innovation Council, which aims to *re-define innovations beyond formal R&D*; *facilitate innovative solutions for inclusive growth for the people and by the people*; *foster an innovation eco-system to strengthen entrepreneurship*; *focus on key drivers of sustainability, durability and quality and expand the space for dialogue and discourse on innovation*.

In summary, one may wish to create:

1. Distributed genomics and bioinformatics centers with advanced platforms for mapping and sequencing with the aim of rapid detection of pathogen strains (e.g., malaria parasites).

2. A centralized genomics, systems biology and bio-medical-informatics institute, with powerful high-performance computing facilities and connected by high-speed broadband networks, to interpret data and disseminate results rapidly. This institute could also perform genome-wide association studies to characterize host biomarkers that make certain individuals malaria-asymptomatic.

3. Computational and mathematical epidemiological models to forecast and control disease reproduction numbers (in a multipatch model).

4. Grass-root level approaches to disease intervention strategies that replicate successes in one region in to others.

Finally, I would like to dedicate this essay to Sir Ronald Ross, whose poem below could be an inspiration to the new generation of biologists, be they bio-medical, molecular, mathematical or computational.

In this, O Nature, yield I pray to me.
I pace and pace, and think and think, and take

The fever'd hands, and note down all I see,
That some dim distant light may haply break.
The painful faces ask, can we not cure?
We answer, No, not yet; we seek the laws.
O God, reveal thro' all this thing obscure
The unseen, small, but million-murdering cause.

§§§

[35]

[35] *The paper has improved considerably following many insightful suggestions from several colleagues: most notably, F. Hoppensteadt, K.R. Sreenivasan, A. Witzel of NYU, R. Parikh of CUNY, and A. Nerode of Cornell.*

References

Allison, AC. The sickle-cell and Haemoglobin C genes in some African populations. *Ann. Human Genet.*, 21:67–89, 1956.

Anantharaman, TS, Mishra, B, and Schwartz, DC. Genomics via optical mapping. II: Ordered restriction maps. *J Comput Biol.*, 4: 91–118, 1997.

Anantharaman, TS, Mishra, B, and Schwartz, DC. Genomics via optical mapping. III: Contiging Genomic DNA and Variations. *Proceedings 7th Intl. Cnf. on Intelligent Systems for Molecular Biology: ISMB '99*, 7:18–27, 1999.

Antoniotti M, Policriti A, Ugel N, and Mishra B. Model building and model checking for biochemical processes. *Cell Biochem Biophys*, 38:271–286, 2003.

Aronson, SM. The man who understood mosquitoes. *Med Health R I.*, 79:351–352, 1996.

Aston, C, Mishra, B and Schwartz, DC. Optical mapping and its potential for large-scale sequencing projects. *Trends Biotechnol.*, 17:297–302, 1999.

Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J and DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*, 1, 2003.

Gardner MJ, et al. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature*, 419:531–34, 2002.

Hall N, et al. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature*, 419:527–531, 2002.

Hyman RW, Fung E, Conway A, Kurdi O, Mao J, Miranda M, Nakao B, Rowley D, Tamaki T, Wang F, and Davis RW. Sequence of *Plasmodium falciparum* chromosome 12. *Nature*, 419: 534–537, 2002.

Jing J, Lai Z, Aston C, Lin J, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Tettelin H, Cummings LM, Hoffman SL, Venter JC, and Schwartz DC. Optical mapping of *Plasmodium falciparum* chromosome 2. *Genome Res.*, 9:175–181, 1999.

Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. *Proc. R. Soc. Lond. A*, 115:700–721, 1927.

Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity. *Proc. R. Soc. Lond. A*, 138:55–83, 1932.

Kermack WO, McKendrick AG . Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity. *Proc. R. Soc. Lond. A*, 141:94–122, 1933.

Kleinberg S, Casey K, and Mishra B. Systems biology via redescription and ontologies (I): finding phase changes with applications to malaria temporal data. *Syst Synth Biol*, 1:197–205, 2007.

Lai Z, Jing J, Aston C, Clarke V, Apodaca J, Dimalanta ET, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Paxia S, Hoffman SL, Craig Venter J, Huff EJ, and Schwartz DC. A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nat Genet.*, 23: 309–313, 1999.

Llinás M, Bozdech Z, Wong ED, Adai AT and DeRisi JL. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res*, 34:1166–1173, 2006.

Martcheva M and Hoppensteadt F. India's Approach to Elliminating emphPlasmodium falciparum Malaria: A Modeling Perspective. *J. Biol. Systems*, 18:867–891, 2010.

May, RL and Anderson RM. *Infectious diseases of humans: dynamics and control*. Oxford, Oxford University Press, 1991.

Menges F, Narzisi G, and Mishra B. TOTALRECALLER: Improved Accuracy and Performance via Integrated Alignment & Base-Calling. *Bioinformatics*.

Mishra, B. *Encyclopedia of the Human Genome*, chapter Optical Mapping, pages 448–453. Nature Publishing Group, Macmillan Publishers Limited, London, UK, 2003.

Mishra, B. Intelligently deciphering unintelligible designs: algorithmic algebraic model checking in systems biology. *J R Soc Interface*, 6:575–597, 2009.

Mishra B, et al. A sense of life: computational and experimental investigations with models of biochemical and evolutionary processes. *OMICS*, 7:253–268, 2003.

Mitrofanova A, Kleinberg S, Carlton J, Kasif S, and Mishra B. Predicting malaria interactome classifications from time-course transcriptomic data along the intraerythrocytic developmental cycle. *Artif Intell Med*, 49:167–176, 2010.

Narzisi G, and Mishra B. Scoring-and-unfolding trimmed tree assembler: concepts, constructs and comparisons. *Bioinformatics*, 2:153–160, 2011.

Narzisi G, and Mishra B. Comparing de novo genome assembly: the long and short of it. *PLoS One*.

Ramakrishnan N, Tadepalli S, Watson LT, Helm RF, Antoniotti M and Mishra B. Reverse engineering dynamic temporal models of biological processes and their relationships. *Proc Natl Acad Sci U S A*, 107:12511–12516, 2010.

Roy, SW and Irimia, M. Origins of human malaria: rare genomic changes and full mitochondrial genomes confirm the relationship of *Plasmodium falciparum* to other mammalian parasites but complicate the origins of *Plasmodium vivax*. *Mol Biol Evol.*, 25: 1192–1198, 2008.

Rudolf Hoernle, AF. *Studies in the Medicine of Ancient India*. Clarendon Press, 1907.

Sakuntabhai A, Ndiaye R, CasadŐmont I, Peerapittayamongkol C, Rogier C, Tortevoye P, Tall A, Paul R, Turbpaiboon C, Phimpraphi W, Trape JF, Spiegel A, Heath S, Mercereau-Puijalon O, Dieye A and Julier C. Genetic determination and linkage mapping of *Plasmodium falciparum* malaria related traits in Senegal. *PLoS One*, 3, 2008.

Serjeant, GR and Serjeant, BE. *Sickle Cell Disease*. Oxford University Press, 2001.

Tadepalli S, Ramakrishnan N, Watson LT, Mishra B, and Helm RF. Simultaneously segmenting multiple gene expression time courses by analyzing cluster dynamics. *J Bioinform Comput Biol*, 7:339–356, 2009.