REVIEWS

THE ORIGIN OF MALARIA: MIXED MESSAGES FROM GENETIC DIVERSITY

Daniel L. Hartl

Over the past 35 years, the incidence of malaria has increased 2–3-fold. At present, it affects 300–500 million people and causes about 1 million deaths, primarily in Africa. The continuing upsurge has come from a coincidence of drug-resistant parasites, insecticide-resistant mosquitoes, global climate change and continuing poverty and political instability. An analogous rapid increase in malaria might have taken place about 10,000 years ago. Patterns of genetic variation in mitochondrial DNA support this model, but variation in nuclear genes gives an ambiguous message. Resolving these discrepancies has implications for the evolution of drug resistance and vaccine evasion.

TROPICAL INFECTIOUS DISEASES

Malaria remains an urgent problem in global public health. The annual death toll is 0.7–2.7 million, with more than 75% of the victims being African children — the unfortunate among the 400–900 million children under five years of age who suffer acute fever[1]. An upsurge in the prevalence of malaria has occurred over the past 35 years due to a 'perfect storm' of mutually reinforcing factors. In 1955, the World Health Organization began an ambitious programme to eradicate malaria through clinical treatment using chloroquine and by control of the mosquito population using DDT (dichlorodiphenyl-trichloroethane). Phased out in the late 1960s, the programme nevertheless resulted in an important and sustained reduction in the burden of the disease in many countries throughout the world[2] (FIG. 1a); however, in many other countries there has been a resurgence in malaria (FIG. 1b). This has resulted from the emergence and spread of drug-resistant parasites[3–5]; the evolution of pesticide-resistant mosquitoes[6–8]; increased population density (the world population has doubled since 1963); global warming[9] (which has allowed the spread of vectors into areas that were previously outside their range); and continuing poverty, political instability[10]

and loss of productivity due to infectious diseases[11], all of which undermine the maintenance of a stable public-health infrastructure for the treatment and control of the disease.

The recent upsurge of malaria owing to a coincidence of mutually reinforcing factors might echo the first expansion of malaria, which is thought to have taken place about 10,000 years ago[12,13]. The events that could have conjoined to create this earlier expansion include climate change in Africa after the last glaciation (which led to optimal warm and humid conditions in the equatorial regions about 10,000 years ago[14]), the increase in human population density owing to the spread of swidden ('slash-and-burn') agriculture into Africa at about the same time[15] and the proliferation and rapid diversification of the highly anthropophilic *Anopheles* mosquito vectors[16,17], which efficiently transmit the parasite *Plasmodium falciparum* that causes the most lethal form of the disease. Determining whether a major expansion of malaria took place about 10,000 years ago is of interest because understanding the causes of this expansion might provide clues for controlling malaria today.

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
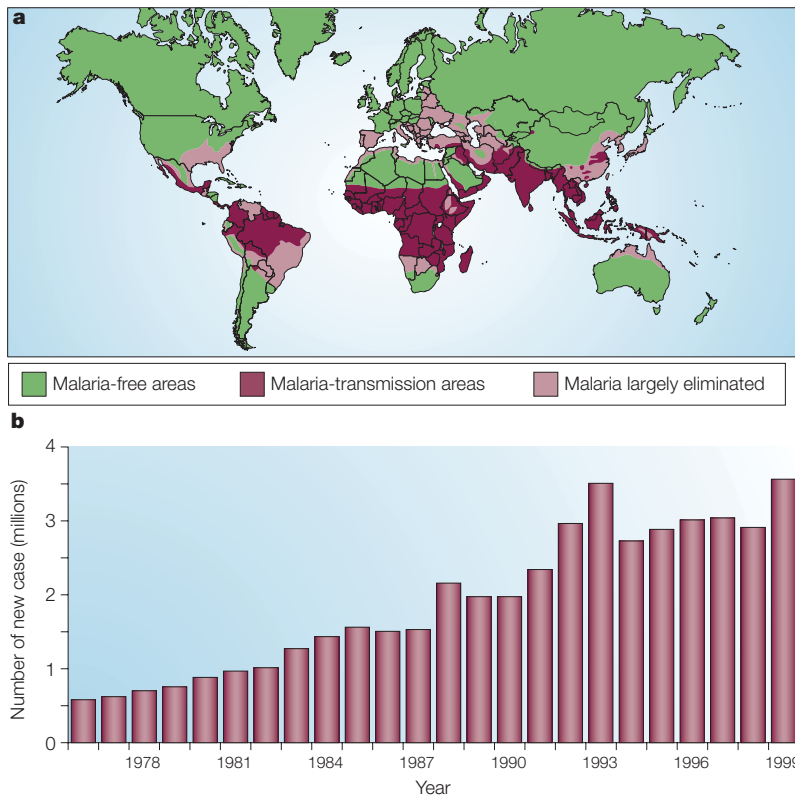*e-mail: dhartl@oeb.harvard.edu*

Figure 1 | **Distribution of malaria, and the increase in its incidence in recent years.**
a | The worldwide distribution of *Plasmodium falciparum* malaria in 2003. Modified with permission from the World Health Organization © (2003). **b** | The incidence of malaria in Zambia since 1975. The histogram illustrates the upsurge in malaria cases in recent years in some countries.

VIRULENCE
A measure of the harm that an infectious disease inflicts on infected individuals, estimated variously as the case-fatality rate, the reduction in expected lifespan or the lethal dose of the infectious agent.

CASE-FATALITY RATE
The number of deaths due to a disease expressed as a percentage of total cases.

SEVERE MALARIA
Cases of malaria that are marked by one or more of the following symptoms: coma, convulsions, a severe drop in blood pressure, difficulty in breathing, jaundice, blood in the urine, extreme weakness or prostration, kidney impairment, severe anaemia or hypoglycaemia.

MICROSATELLITE POLYMORPHISM
A polymorphism in a population that results in a difference in the number of tandem repeats of a short (1–8 bp) DNA sequence.

Traditional arguments for a dramatic expansion of malaria about 10,000 years ago often cite the VIRULENCE of the disease[18], on the grounds that virulence should decrease as the parasite evolves. Although *P. falciparum* malaria does have a high CASE-FATALITY RATE — estimated as 9.4% in one multicentre study of 1,230 children under the age of five with SEVERE MALARIA (REF. 19) — the intuitively appealing and comforting notion that evolution inevitably leads to an attenuation of virulence does not stand up to rigorous theoretical analysis (BOX 1). An expansion of malaria that was coincident with the origin of agriculture is supported by an analysis of the geographical distribution of the sickle-cell haemoglobin mutation[12,13] and by estimates of the age of the glucose-6-phosphate dehydrogenase mutation[20]. Both of these mutations confer partial resistance to malaria (probably by facilitating the clearance of infected red blood cells), but inferring the history of a parasite from this type of mutation in the host is indirect. Evidence that has a more direct bearing on the origin of epidemic malaria has come from an unexpected source — the level of genetic variation in the parasite itself. The evidence is not altogether consistent, however. Some studies reveal a low level of genetic variation in *P. falciparum,* indicating a relatively recent expansion, whereas other studies reveal a high level of genetic variation, indicating a large population size that has been maintained over hundreds of thousands of years.

Evidence from protein-coding genes

Rapid population expansion can leave a characteristic signature on genetic polymorphisms in the form of a reduced overall level of variation, as well as a frequency distribution of variants that is skewed towards rare alleles. Neither of these signatures is observed in *P. falciparum* for genes that encode antigens, or for MICROSATELLITE POLYMORPHISMS. However, neither antigen-encoding genes nor microsatellite polymorphisms are ideal for estimating the level of genetic variation in a population (TABLE 1). Genes that encode antigens are subject to strong selection for diversity to evade the host immune response, and microsatellites have an extraordinarily high mutation rate, owing to REPLICATION SLIPPAGE[21,22].

The possibility of a rapid expansion of *P. falciparum* about 10,000 years ago received renewed interest after an analysis of protein-coding sequences that were available in GenBank[23]. The data comprised 2–32 alleles of each of 10 genes, which are scattered among 6 of the 14 chromosomes in *P. falciparum*, and include highly polymorphic genes encoding antigens, such as circumsporozoite antigen protein 1, as well as general metabolic 'housekeeping' genes, such as triose phosphate isomerase, which functions in glycolysis. Among 16,449 nucleotide sites in these coding regions, there were 51 (0.3%) NONSYNONYMOUS POLYMORPHISMS and no SYNONYMOUS POLYMORPHISMS[23]. In view of the high level of polymorphism in genes encoding antigens and in microsatellites, the complete absence of synonymous polymorphism in the *P. falciparum* data was completely unexpected. It can be compared with data from enteric bacteria[24], for which 30,203 nucleotide sites in 28 coding sequences yielded 3.2% nonsynonymous polymorphisms and 10.2% synonymous polymorphisms.

How can the lack of synonymous polymorphism be explained? The favoured explanation for this was that the alleles of each gene were all very closely related. The level of synonymous polymorphism is consistent with a scenario in which all extant parasites derive from a single individual that existed no more than 24,500–57,500 years ago[23]. This range corresponds to the upper limits of the 95% confidence intervals that have been estimated using different values for nucleotide substitution rates derived from either the chimpanzee parasite *Plasmodium reichenowi* or the rodent parasite *Plasmodium berghei*. As the range 24,500–57,500 years is an upper limit, the authors suggested that the actual common ancestry was probably more recent, and consistent with the expansion of the parasite population 10,000 years ago[23]. An alternative explanation for the low level of synonymous polymorphism is based on CODON-USAGE BIAS. In many organisms, the use of codons that specify the same amino acid is nonrandom, and in highly expressed genes, codon-usage bias often favours codons that correspond to the most abudant transfer RNA molecules in the cell. This explanation was originally regarded as implausible[23], and it seems even less likely now that the complete genomic sequence of *P. falciparum* is available[25]. In contrast to many other organisms, in which there are

Box 1 | The evolution of parasite virulence

The idea that parasite virulence should decrease as a result of natural selection has intuitive appeal. The implicit logic is that highly virulent parasites, by killing their hosts, will impair their own transmission, and hence less virulent genotypes should have a selective advantage over more virulent ones. In one of the simplest theoretical models of parasite transmission[54], this logic is embodied in equation 1:

$$R_0 = \frac{\beta N}{\alpha + b + \nu} \qquad\qquad (1)$$

in which R_0 is the number of secondary infections resulting from a single infected host (a measure of parasite fitness), β is the transmissibility, N is the density of susceptible hosts, and α, b, and ν are the rate constants for parasite-induced mortality, natural host mortality and recovery, respectively. All else being equal, it follows that a decrease in α will increase R_0, and hence a lower level of virulence will be favoured.

One problem with this lies in the phrase 'all else being equal', because transmissibility itself can be a function of virulence. In such models, there is an interdependence between parasite transmissibility (β) and parasite-induced host mortality (α), with trade-offs resulting in a maximum for R_0 at some optimal level of virulence[55]. Whether the optimum virulence is high or low depends on the biological specifics of the trade-offs.

Furthermore, the measure of virulence, α, as the parasite-induced instantaneous mortality rate is rather arbitrary. Alternative measures include the case-mortality rate, the expected host lifespan, and the lethal dose, any of which can have a predicted pattern of evolution that differs qualitatively from that of the parasite-induced instantaneous mortality rate[56]. Furthermore, counterintuitive patterns of virulence evolution can arise through co-evolution between parasite and host, particularly when effects on host fecundity are taken into account[57].

Comparisons among different types of pathogen indicate that vector-borne parasites are often more virulent that those that are transmitted directly[58]. Again, although this is often true, its theoretical underpinning depends on the details of the host-morbidity and parasite-fecundity schedules[59].

Overall, in regard to the evolution of parasite virulence, theoretical models have been both illuminating and frustrating. They are illuminating in that they clarify the quantitative relationships between the many parameters that affect parasite transmission and its biological effects on the host; however, they are frustrating in that generalities are few. In any particular case, the predicted evolution of virulence, regardless of how virulence is defined, depends intimately on the details of the biological interactions between parasite, vector (if vector-borne) and host[60].

REPLICATION SLIPPAGE
Aberrant replication across regions of DNA that contain short, tandemly repeated sequences that results in an increase or decrease in the number of repeats present in the daughter molecules.

NONSYNONYMOUS POLYMORPHISM
A single-nucleotide polymorphism in a coding region that results in an amino-acid replacement.

SYNONYMOUS POLYMORPHISM
A single-nucleotide polymorphism in a coding region that does not result in an amino-acid replacement.

CODON-USAGE BIAS
Nonrandom usage of synonymous codons, which specify the same amino acid in a polypeptide chain.

ALLOTYPES
Products of one or more alleles that result in inherited variants of a particular molecule, usually a protein.

GENE CONVERSION
A process in which the sequence of nucleotides in a gene or allele is changed by DNA-repair mechanisms, using the nucleotide sequence of a related gene or allele as the template for repair.

COALESCENT
The convergence, going backwards in time, of two or more gene lineages onto a single common ancestor.

MAXIMUM LIKELIHOOD
A method of statistical estimation that stipulates an underlying model of a process and that estimates any parameter as the value that maximizes the probability of the observed data given the correctness of the model.

duplicate copies of tRNA genes, which reflect the organism's overall codon usage[26], in *P. falciparum* there is a minimal amount of duplication of the tRNA genes.

Although a dramatic population expansion 10,000 years ago does not require that all extant *P. falciparum* parasites be descended from a single common ancestor, evidence for a single common ancestor at around this time provides strong support for this model of population expansion. However, the inference of a recent common ancestor was greeted with scepticism, and was contested almost immediately. First, the inference seems at odds with the high level of polymorphism that is found in many genes in *P. falciparum*, especially genes that encode antigens or membrane-bound proteins[27]. These antigens include MSP1 (merozoite surface protein 1), one of the most highly polymorphic proteins in eukaryotes, the ALLOTYPES of which can differ at more than 70% of the amino-acid residues in some regions[28,29]. For a protein to acquire such extreme divergence as MSP1 through the conventional processes of mutation and selection would require about 48 million years[30], which would make MSP1 the oldest eukaryotic polymorphism known. On the other hand, the genes that encode MSP1 and some other highly polymorphic proteins contain internally repeated regions that can generate high levels of allelic variation by means of replication slippage[29] or GENE CONVERSION[31], and there is no doubt that selection for diversity in MSP1 is strong, owing to its interaction with the host immune system.

More doubt was cast on the analysis of synonymous polymorphisms in GenBank by a report using essentially the same data but that reached opposite conclusions[32]. This analysis used COALESCENT theory to infer common ancestry of the alleles and, by means of MAXIMUM LIKELIHOOD, estimated that the population size of *P. falciparum* must have been of the order of at least 100,000 parasites for the past 300,000–400,000 years. How could such a large discrepancy have arisen from two analyses of what are essentially the same data?

Errors in older GenBank sequences

The great discrepancy between a minimum of 24,500–57,500 years[23] and a maximum of 300,000–400,000 years[32] results, in part, from which GenBank sequences were included or excluded from the analyses. Estimates of the age of the most recent common ancestor that include genes for antigens selected for diversity will be biased towards older values, and those including genes whose variation is constrained by selective forces will be biased toward more recent values. Another potential source of uncertainty results from errors, either in GenBank sequences or in sequence alignments, which might have arisen because of the difficulty that older sequencing technologies had in coping with the high A/T content of the *P. falciparum* genome (the average A/T content is 80.6%)[25]. To assess errors in the older GenBank sequences, we resequenced genes of interest from available isolates using modern methods[33]. We found, for example, that while aldolase

Table 1 | **Types of genetic variation used for tracing population history**

Type of genetic variation	Main advantages	Main disadvantages
Antigenic variation	Abundant	Strong selection for diversity to evade the host immune system
Microsatelllites	Abundant	High mutation rate; unknown pattern of mutation
Synonymous nucleotide sites	Low mutation rate; weak selective constraints	Possible selection against variation due to biased codon usage
Intron sites	Low mutation rate; weak sequence constraints	High A/T content; many microsatellites; possible selection against variation due to unrecognized selective forces
Upstream and downstream noncoding regions	Low mutation rate	High A/T content; many microsatellites; possible selection for diversity due to effects on gene expression
Mitochondrial DNA	Low mutation rate; no recombination	Possible selection on some mutations, and selective sweeps of the entire mitochondrial DNA molecule

is reported to have 14 polymorphic nucleotide sites, 13 of these are present in a single GenBank entry (accession number AF179421) and could not be confirmed by resequencing. As another example, falcipain 2 was reported to have eight segregating sites, all of which have subsequently been found to result from gene conversion between unrecognized paralogues[31].

In addition to this failure to confirm many of the reported polymorphisms by resequencing[33], the low level of synonymous polymorphism was also supported by data from a newly sequenced sample of genes. TABLE 2 shows the data for regions of 20 protein-coding genes from 11 chromosomes among 5–7 reference isolates. Among a total of 22,611 nucleotides that were sequenced for each strain, we found 21 nonsynonymous polymorphisms (D_n) and 1 synonymous polymorphism (D_s). Many of these genes were deliberately chosen to differ from those analysed previously to avoid potential bias due to the sample of genes. The low level of synonymous polymorphism revealed by these data also indicates a relatively recent common ancestor[23].

Introns and other noncoding sequences

While the data from coding sequences were being debated, the complete sequences of *P. falciparum* chromosomes 2 and 3 became available through the Malaria Genome Project[34,35]. This important information afforded a new approach to the study of genetic diversity in *P. falciparum* because it allowed noncoding sequences to be examined. The rationale is that if there are unrecognized constraints on synonymous sites, then noncoding sequences will be a more reliable indicator of the level of background polymorphism because of the relative lack of SELECTIVE CONSTRAINTS. This led to the examination of introns[36], which are among the most rapidly evolving DNA sequences, consistent with relatively weak selective constraints[37,38]. The initial study analysed 4,217 bp in 25 introns across chromosomes 2 and 3 in eight reference isolates. With an A/T content approaching 90%, the introns contain a large number of microsatellite sequences. These consist of tandem repeats of short sequences of 1–8 nucleotides in length, including many mononucleotide repeats (especially $(A)_n$ and $(T)_n$) and dinucleotide repeats (for example, $(AT)_n$). Approximately half of the microsatellite repeats in the

introns analysed were found to be polymorphic in the number of repeating units that were present in different isolates[36], attesting to the high rate at which these sequences undergo replication slippage to yield new microsatellite alleles[21,22]. Altogether, microsatellite repeats accounted for 19% (800 of 4,217 bp) of the intron sequences. However, five of eight SINGLE-NUCLEOTIDE POLYMORPHISMS (SNPs) were located in microsatellite repeats, a statistically significant excess. This finding suggested that microsatellite repeats might have a higher frequency of base-substitution mutation than more complex sequences, owing to nucleotide incorporation errors during MISMATCH REPAIR[39]. Excluding the five SNPs that are located in microsatellites, from this analysis, the estimated time to the most recent common ancestor of the reference strains was calculated to be 9,500–23,000 years. On the other hand, if the five SNPs in microsatellites are not excluded, then the most recent common ancestor estimate is 25,000–61,000 years. Which estimate is to be believed?

Additional uncertainty about the timing of the most recent common ancestor came from a study of 204 coding and noncoding sequences on chromosome 3 from five diverse isolates of *P. falciparum*[40]. On the whole, these data show less polymorphism than would be expected from an organism that had maintained a very large population size for several hundred thousand years. For example, across the central region of the chromosome, 150 of 204 genes contained no synonymous SNPs. However, the overall level of single-nucleotide polymorphism was sufficient to imply a most recent common ancestor for chromosome 3 that existed 100,000–180,000 years ago. This is much older than the previous estimates from coding sequences[23], and even older than the estimates obtained from introns[36]. Is it possible that, on average, genes in chromosome 2 share a more recent common ancestor than genes in chromosome 3?

Mitochondrial DNA may tell the tale

One problem with nuclear genes is that they can undergo recombination. Although inbreeding (mating between related genotypes) takes place in many local populations of *P. falciparum*[41–43], and this does reduce the effective rate of recombination[44], the level of

SELECTIVE CONSTRAINT
A restriction on the level or type of polymorphism that is likely to be found in a population owing to deleterious effects of the polymorphism on the ability of their carriers to survive and reproduce.

SINGLE-NUCLEOTIDE POLYMORPHISM
(SNP). A position in a genomic DNA sequence in which the particular nucleotide pair that is present can differ from one individual (or chromosome) to the next; it normally refers to genetic variation that is common in the population, and excludes rare mutational variants.

MISMATCH REPAIR
A process of DNA repair in which a mispaired region of a DNA duplex is excised and replaced by resynthesis using the remaining strand as a template.

Table 2 | **Single-nucleotide polymorphisms in coding sequences of *P. falciparum***

Gene product	Chromosome	Strains	Nucleotides per strain	D_n	D_s
GPSP	1	6	552	0	0
AcCoA 695	2	7	2,667	11	0
AcCoA 685	2	5	2,706	2	0
ATPase 3	5	6	546	0	0
GPAT	6	6	636	0	0
STARP	7	6	1,092	1	1
GPRK	7	6	1,395	1	0
GPCA	9	6	252	0	0
GRP78	9	6	426	0	0
LSA1	10	6	684	2	0
GBP130	10	6	96	0	0
Enolase	10	6	738	0	0
Falcipain 3	11	6	1,350	0	0
FuFo1	11	6	1,890	1	0
CPN60	12	6	42	0	0
GPCE	13	6	1,725	0	0
Actin 2	13	6	1,131	0	0
G6PD	14	6	1,884	3	0
Falcipain 1	14	5	1,710	0	0
Aldolase	14	7	1,089	0	0
Total			22,611	21	1

D_n is the total number of nonsynonymous polymorphisms; D_s is the total number of synonymous poplymorphisms.

inbreeding is usually low enough that recombination is effective in breaking up linked blocks of nucleotide sequence[44–46]. This means that regions from different chromosomes can have distinct ancestral histories, so the most recent common ancestor can differ from one chromosomal region to the next.

As the 6-kb mitochondrial genome of *P. falciparum* is inherited through only one parent, it does not undergo recombination and is therefore potentially more informative for inferring population history than are nuclear genes. Evidence for a recent mitochondrial common ancestor emerged from an initial study of the complete sequence of the mitochondrial genome from four isolates, which revealed an extremely low level of synonymous nucleotide polymorphism[47]. Additional DNA typing of the polymorphisms among 104 diverse isolates showed a geographical pattern that indicated radiation out of Africa into southeast Asia and South America[47].

The inference of a rapid, recent expansion of *P. falciparum* has also been supported by sequences of the complete mitochondrial genome from 100 geographically diverse isolates[48]. Rapid population growth affects the distribution of polymorphisms because, in the genealogy of a gene, the extra terminal branches that are produced by the rapid expansion of the population provide more opportunity for mutation; mutations will therefore have a tendency to be of relatively recent origin. The result is that, when sequences are aligned in pairs, the distribution of the number of mismatches is smooth and unimodal,

with the position of the mode reflecting the time of the population growth (BOX 2). The mitochondrial data for African isolates show a mismatch distribution that is typical of rapid population growth, with a mode that is consistent with rapid expansion having occurred approximately 10,000 years ago[48]. The data also indicate that some lineages are more ancient (50,000–100,000 years), and that parasite populations migrated to southeast Asia and South America prior to the African expansion. The latter inference is surprising, as the pattern of microsatellite variation in southeast Asia and South America indicates a much more recent invasion[45], and is also at odds with the conventional wisdom that malaria was introduced into South America by Europeans and the slave trade[49].

However, for any particular nonrecombining DNA sequence, such as mitochondrial DNA, the pattern of sequence relationships has a large variation, which is due purely to the chance extinction of lineages that occurs over a period of time. Therefore, the mitochondrial genealogy might not necessarily coincide with that of any nuclear gene. Moreover, this analysis assumes that all nucleotide substitutions in mitochondrial DNA are selectively neutral. If some are selectively favoured, then the increased frequency of this type of substitution would make them seem substantially older than they actually are. The possibility of selection cannot be ruled out, since the mitochondrial gene encoding cytochrome b seems to underlie susceptibility to certain antibiotics[50].

Box 2 | Population expansion and the mismatch distribution

A sudden burst of population growth has characteristic effects on the distribution of mutations among homologous DNA sequences. Among the population statistics that are affected is the histogram of the number of nucleotide mismatches among all pairwise comparisons — known as the mismatch distribution[61,62].

For nonrecombining DNA sequences, such as mitochondrial DNA, after a burst of growth the mismatch distribution approximates a smooth curve with a single, well-defined peak, whereas in a steady-state population with a constant population size the histogram is ragged[63].

The ragged histogram in a steady-state population is due to the large variance in the time to the most recent common ancestor of any two randomly chosen lineages, which leads to a relatively large coefficient of variation (the ratio of the standard deviation to the mean) for the number of mismatches.

In contrast to a steady-state population, a sudden population expansion results in a rapid increase in the average number of mismatches and a precipitous drop in the coefficient of variation[61]. Both of these effects are due to the new mutations that occur in the extra lineages that are created by the population growth. If the population remains large, independent mutations continue to accumulate so that, with time, the mean of the mismatch distribution increases. Very gradually, as a new steady state is approached, random extinctions of lineages return the mismatch distribution to a ragged appearance, but this takes a long time. Consequently, if the mutation rate remains constant, a smooth mismatch distribution with a well-defined peak is indicative of population expansion.

The figure shows the mismatch distribution for mitochondrial DNA from African isolates of *Plasmodium falciparum*[48]. The histogram shows the distribution that is expected from a recent population expansion. Quantitatively, the position of the mean, τ, in the mismatch distribution equals 1.17 and corresponds theoretically to equation 2:
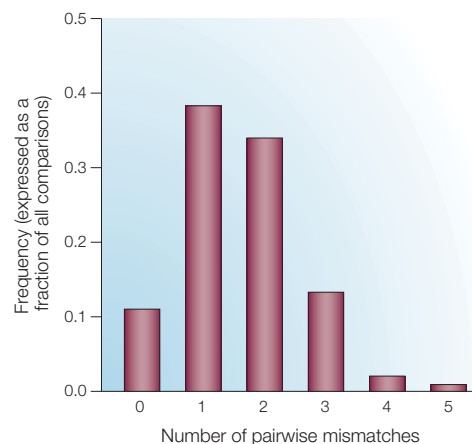
$$\tau = 2ut \qquad (2)$$

where u is the per-sequence mutation rate per generation and t is the time in generations[61]. Considering the divergence in mitochondrial DNA between *P. falciparum* and the chimpanzee parasite *Plasmodium reichenowi*, $2u$ can be estimated as ranging from 5.85×10^{-5} to 8.36×10^{-5} per molecule. This corresponds to a value of t in the range of ~14,000–20,000 generations. Assuming two generations per year, the time since the beginning of the rapid expansion is calculated as 7,000–10,000 years[48].

Recent theoretical results indicate that tests for population expansion that are based on statistics from mismatch distribution are more conservative than tests that are based on either the distribution of mutation frequencies (the frequency spectrum) or the distribution of haplotypes (genetically linked polymorphisms)[64]. Hence, the conservative nature of tests based on mismatch distribution reinforces the inference of recent population growth from the data in the figure.

Although sudden population growth results in a unimodal mismatch distribution in a geographically homogeneous population, the pattern can be dramatically different when there is geographical subdivision. With a sufficiently small migration rate between subpopulations, the mismatch distribution can even become multimodal[65]. Geographical subdivision might explain some of the anomalies in the mismatch distributions that are observed for certain populations of *P. falciparum*, for example, in South America[48].

Malaria today

Ongoing studies are in progress to estimate the age of the most recent common ancestor for coding and non-coding sequences in various regions of the genome. However, it already seems to be clear that the level of polymorphism in *P. falciparum* is much lower than that found in many other microorganisms. Although this finding is not consistent with a continuously large population size, neither is it consistent with a single common ancestor. At present, the model that seems to be the best fit is consistent with a rapid expansion from a relatively small number of individuals (perhaps a few hundred) in Africa about 10,000 years ago. This conclusion is tentative and might need revision pending data from studies that are now underway.

It could be that each chromosome, or perhaps individual regions of chromosomes, have different ancestral histories due to a series of SELECTIVE SWEEPS, in which favourable mutations arise and rapidly become incorporated into the population.

Although the conflicting interpretations of the population history of *P. falciparum* might at first seem to be an empty academic dispute, the issue has important implications for the evolution of drug resistance and immune evasion, as well as for *P. falciparum* genetic-screening strategies (for example, identifying the optimal position on the spectrum between screening a small number of organisms at a large number of sites, or the other way around). For if *P. falciparum* has abundant genetic variation that is present throughout the world,

SELECTIVE SWEEP
The rapid increase in the frequency of a new favourable mutation to displace, or nearly displace, other alleles of the gene.
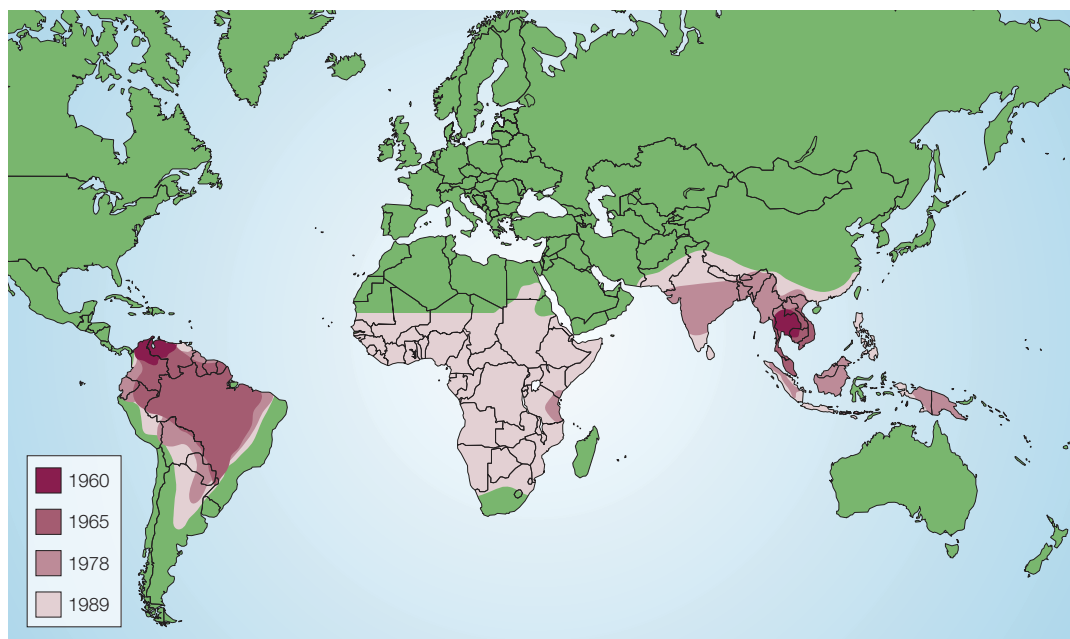
Figure 2 | **Origin and spread of chloroquine resistance.** Pockets of resistance were first noted almost simultaneously in southeast Asia and South America around 1960. These spread locally until 1978, when resistance was first noted in Africa. Since then, chloroquine resistance has spread through virtually the whole of Africa.

then resistance to new drugs could, in principle, evolve through recombination and selection of existing polymorphisms, perhaps almost simultaneously in several geographically distant locations. If, on the other hand, the level of genetic polymorphism in *P. falciparum* is very low, then drug resistance might require new favourable mutations to occur. In this case, they are likely to require a 'waiting time' until their first occurrence, and unless the rate of mutation is high, the resistance mutations are likely to originate at a single geographical location and spread from there. Present data indicate that the latter possibility is the more likely.

The evolution of chloroquine resistance is a case in point. First synthesized in 1937 in Germany, and tested as an antimalarial by German troops in North Africa, chloroquine came into widespread civilian use after the Second World War. It was a magic bullet, cheap (the equivalent of US $0.10 for each course of treatment at current market prices) and effective. And it remained effective for 20 years, in spite of underdosing, noncompliance, and indiscriminate consumption, including its use as an additive in table salt in some countries[3]. The first signs of resistance appeared almost simultaneously in southeast Asia and South America in the late 1950s (FIG. 2), but it was not until the late 1970s that resistance appeared in Africa[51]. Although the mechanism of resistance is not known, the main resistance mutations are in *pfcrt*, a gene that encodes a protein containing 10 putative transmembrane domains that localizes to the digestive vacuole membrane[5,52]. Resistance is slow to appear because it evidently requires a sequential accumulation of mutations, with the early steps conferring increasing drug tolerance and the later steps progressing to full drug resistance[51]. In the absence of pre-existing

genetic variation, the waiting time for the occurrence and selection of a particular sequence of mutations is long. In spite of resistance in southeast Asia and South America, chloroquine remained effective in Africa because it took 20 years for the resistant parasites to spread. In the present era of global travel, one could not realistically expect such a long grace period.

Roll Back Malaria
The present need for malaria control is great. Without effective control, the worldwide burden of malaria is expected to double over the next 20 years[1]. (The doubling time for the incidence of malaria in FIG. 1b is 8.7 years.) In response to this need, in 1998 a consortium of the World Health Organization, the World Bank, the United Nations Development Programme and the United Nations Children's Fund initiated a new campaign entitled Roll Back Malaria. This intiative has been joined by others, such as the Multilateral Initiative on Malaria (MIM), the Medicines for Malaria Venture (MMV), the Malaria Vaccine Initiative (MVI) and the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM). Unlike the 1955 campaign, Roll Back Malaria does not have the unrealistic objective of eradicating malaria. Instead, its goal[53] is to 'calm the storm' by cutting the burden of malaria by half by the year 2010. The present level of funding for the initiative is only about 10% of that needed, and the time frame should probably be closer to 20 years rather than 7, but it is a beginning[2]. Our knowledge of the population genetics of *P. falciparum*, combined with appropriate screening strategies for detecting drug resistance, might prolong the effective lifetime of new drugs and vaccines beyond that of chloroquine.

1. Breman, J. G. The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **64**, S1–S11 (2001).
2. Sachs, J. D. A new global effort to control malaria. *Science* **298**, 122–124 (2002).
3. Hastings, I. M., Bray, P. G. & Ward, S. A. A requiem for chloroquine. *Science* **298**, 74–75 (2002).
4. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).
5. Sidhu, A. B. S., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfcrt* mutations. *Science* **298**, 210–213 (2002).
6. Bradley, D. J. The particular and the general: issues of specificity and verticality in the history of malaria control. *Parassitologia* **40**, 5–10 (1998).
7. Hemingway, J. & Ranson, H. Insecticide resistance in insect vectors of human disease. *Ann. Rev. Entomol.* **45**, 369–389 (2000).
8. Hemingway, J., Field, L. & Vontas, J. An overview of insecticide resistance. **298**, 96–97 (2002).
9. Report of the Intergovenmental Panel on Climate Change (IPCC). *Climate Change 2001: impacts, adaptation and vulnerability. Contribution of Working Group II to the Third Assessment Report of United Nations Environment Programme* (IPCC, Geneva, Switzerland, 2001).
10. World Health Organization. Press Release PR-2000-46. Third of African malaria deaths due to conflict or natural disaster. [online], (cited 11 Nov 2003), <http://www.who.int/inf-pr-2000/en/pr2000-46.html> (30 June 2000).
11. Gallup, J. L. & Sachs, J. D. The economic burden of malaria. *Am. J. Trop. Med. Hyg.* **64**, S85–S96 (2001).
12. Livingstone, F. B. Anthropological implications of sickle-cell gene distribution in West Africa. *Am. Anthropol.* **60**, 533–562 (1958).
13. Wiesenfeld, S. L. Sickle-cell trait in human biological and cultural evolution. *Science* **157**, 1134–1140 (1967).
14. Olago, D. O. Vegetation changes over palaeo-time scales in Africa. *Climate Res.* **17**, 105–121 (2001).
15. Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. & Martin, W. Genetics and geography of wild cereal domestication in the Near East. *Nature Rev. Genet.* **3**, 429–441 (2002).
16. Coluzzi, M., Petrarca, V. & Di Deco, M. A. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll. Zool.* **52**, 45–63 (1985).
17. Coluzzi, M. The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia* **41**, 277–283 (1999).
18. Boyd, M. (ed.) *Malariology: A Comprehensive Survey of All Aspects of This Group of Diseases from a Global Standpoint* (Saunders, Philadelphia, 1949).
19. Liaison Bulletin of the Malaria Programme WHO/AFRO. Severe malaria in the African region: results of a multicentre study. [online], (cited 11 Nov 2003), <http://www.afro.who.int/malaria/bulletins/2001-04_vol4-2.pdf> (April 2001).
20. Tishkoff, S. A. *et al.* Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
21. Su, X.-Z. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falcipirum*. *Science* **286**, 1351–1353 (1999).
22. Anderson, T. J. C., Su, X. Z., Roddam, A. & Day, K. P. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. *Mol. Ecol.* **9**, 1599–1608 (2000).
23. Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. Malaria's Eve: evidence of a recent population bottleneck throughout the world population of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 4425–4430 (1998).
   A key paper and important stimulus to research on the population genetics of *P. falciparum*, which concluded from an analysis of GenBank sequences that the present parasite population derives relatively recently from a relatively small number of common ancestors.
24. Hartl, D. L., Boyd, E. F., Bustamante, C. D. & Sawyer, S. A. in *Genomics and Proteomics* (ed. Suhai, S.) 37–49 (Plenum Press, New York, 2000).
25. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
26. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **203**, 1–13 (1985).
27. Volkman, S. *et al.* Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science* **298**, 216–218 (2002).
   This paper describes the unusual distribution of single-nucleotide polymorphisms in chromosome 2 of *P. falciparum*, with most polymorphisms concentrated in the subtelomeric regions and with a disproportionate number present in genes encoding proteins associated with the cell membrane.
28. Tanabe, K., Mackay, M., Goman, M. & Scaife, J. G. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **195**, 273–287 (1987).
29. Rich, S. M. & Ayala, F. J. Population structure and recent evolution of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **97**, 6994–7001 (2000).
30. Hughes, A. L. *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, New York, 1999).
31. Nielsen, K. M. *et al.* Gene conversion as a source of nucleotide diversity in *Plasmodium falciparum*. *Mol. Biol. Evol.* **20**, 726–734 (2003).
32. Hughes, A. L. & Verra, F. Very large long–term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. R. Soc. London B Biol. Sci.* **268**, 1855–1860 (2001).
33. Barry, A. E. *et al.* Artefacts in sequence data from the human malaria parasite, *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **130**, 143–147 (2003).
   A cautionary tale about the uncritical acceptance of some of the older GenBank sequences and about special problems encountered in the analysis of population genetic data from A/T-rich sequences.
34. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
35. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
36. Volkman, S. K. *et al.* Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293**, 482–484 (2001).
37. Li, W.-H. *Molecular Evolution* (Sinauer, Sunderland, Massachusetts, 1997).
38. Hartl, D. L. *A Primer of Population Genetics* (Sinauer, Sunderland, Massachusetts, 2000).
39. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive sequences in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
40. Mu, J. *et al.* Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* **418**, 323–326 (2002).
   Evidence that the average age of the most recent common ancestor for genes in chromosome 3 is substantially older than 10,000 years.
41. Babiker, H. A. *et al.* Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitol.* **109**, 413–421 (1994).
42. Paul, R. E. L. *et al.* Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**, 1709–1711 (1995).
43. Hill, W. G., Babiker, H. A., Ranfordcartwright, L. C. & Walliker, D. Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.* **65**, 53–61 (1995).
44. Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96**, 4506–4511 (1999).
45. Anderson, T. J. C. *et al.* Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
46. Hartl, D. L. *et al.* The paradoxical population genetics of *Plasmodium falciparum*. *Trends Parasit.* **18**, 266–272 (2002).
47. Conway, D. J. *et al.* Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* **111**, 163–171 (2000).
48. Joy, D. A. *et al.* Early origin and recent expansion of *Plasmodium falciparum*. *Science* **300**, 318–321 (2003).
   A study of mitochondrial DNA sequences indicating that rapid population growth of a relatively small number of lineages of *P. falciparum* started about 10,000 years ago.
49. Watts, S. *Epidemics and History: Disease, Power and Imperialism* (Yale Univ. Press, New Haven, 1997).
50. Vaidya, A. B., Lashgari, M. S., Pologe, L. G. & Morrisey, J. Structural features of *Plasmodium* cytochrome-b that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones. *Mol. Biochem. Parasitol.* **58**, 33–42 (1993).
51. Wellems, T. E. *Plasmodium* chloroquine resistance and the search for a replacement antimalarial drug. *Science* **198**, 124–126 (2002).
52. Wootten, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).
53. Remme, J. H. F., Binka, F. & Nabarro, D. Toward a framework and indicators for monitoring Roll Back Malaria. *Am. J. Trop. Med. Hyg.* **64**, 76–84 (2001).
54. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans* (Oxford Univ. Press, Oxford, UK, 1991).
55. Lenski, R. E. & May, R. M. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J. Theor. Biol.* **169**, 253–265 (1994).
56. Day, T. On the evolution of virulence and the relationship between various measures of mortality. *Proc. R. Soc. Lond. B Biol. Sci.* **269**, 1317–1323 (2002).
57. Day, T. & Burns, J. G. A consideration of patterns of virulence arising from host–parasite coevolution. *Evolution* **57**, 671–676 (2003).
58. Ewald, P. W. *Evolution of Infectious Diseases* (Oxford Univ. Press, Oxford, UK, 1994).
59. Day, T. Parasite transmission modes and the evolution of virulence. *Evolution* **55**, 2389–2400 (2001).
60. Ganusov, V. V. & Antia, R. Trade-offs and the evolution of virulence of microparasites: do details matter? *Theor. Popul. Biol.* **64**, 211–220 (2003).
61. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
62. Harpending, H. C. *et al.* Genetic traces of ancient demography. *Proc. Natl Acad. Sci. USA* **95**, 1961–1967 (1998).
63. Harpending, H. C. Signature of ancient population growth in a low–resolution mitochondrial DNA mismatch distribution. *Human Biol.* **66**, 591–600 (1994).
64. Ramos–Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100 (2002).
65. Ray, N., Currat, M. & Excoffier, L. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**, 76–86 (2003).

🌐 Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez/
Plasmodium falciparum

FURTHER INFORMATION
Daniel L. Hartl's laboratory:
http://www.oeb.harvard.edu/hartl/lab/
Malaria Foundation International: http://www.malaria.org/
Roll Back Malaria: http://www.rbm.who.int/
The *Plasmodium* genome resource: http://plasmodb.org/
What is malaria?:
http://www.malariasite.com/malaria/WhatIsMalaria.htm
WHO/TDR malaria database: www.wehi.edu.au/MalDB-www/who.html
Access to this interactive links box is free online.