ARTICLE

doi:10.1038/nature09442

Origin of the human malaria parasite *Plasmodium falciparum* in gorillas

Weimin Liu[1], Yingying Li[1], Gerald H. Learn[1], Rebecca S. Rudicell[2], Joel D. Robertson[1], Brandon F. Keele[1]†, Jean-Bosco N. Ndjango[3], Crickette M. Sanz[4,5], David B. Morgan[5,6], Sabrina Locatelli[7], Mary K. Gonder[7], Philip J. Kranzusch[8], Peter D. Walsh[9], Eric Delaporte[10], Eitel Mpoudi-Ngole[11], Alexander V. Georgiev[12], Martin N. Muller[13], George M. Shaw[1,2], Martine Peeters[10], Paul M. Sharp[14], Julian C. Rayner[1,15] & Beatrice H. Hahn[1,2]

Plasmodium falciparum **is the most prevalent and lethal of the malaria parasites infecting humans, yet the origin and evolutionary history of this important pathogen remain controversial. Here we develop a single-genome amplification strategy to identify and characterize *Plasmodium* spp. DNA sequences in faecal samples from wild-living apes. Among nearly 3,000 specimens collected from field sites throughout central Africa, we found *Plasmodium* infection in chimpanzees (*Pan troglodytes*) and western gorillas (*Gorilla gorilla*), but not in eastern gorillas (*Gorilla beringei*) or bonobos (*Pan paniscus*). Ape plasmodial infections were highly prevalent, widely distributed and almost always made up of mixed parasite species. Analysis of more than 1,100 mitochondrial, apicoplast and nuclear gene sequences from chimpanzees and gorillas revealed that 99% grouped within one of six host-specific lineages representing distinct *Plasmodium* species within the subgenus *Laverania*. One of these from western gorillas comprised parasites that were nearly identical to *P. falciparum*. In phylogenetic analyses of full-length mitochondrial sequences, human *P. falciparum* formed a monophyletic lineage within the gorilla parasite radiation. These findings indicate that *P. falciparum* is of gorilla origin and not of chimpanzee, bonobo or ancient human origin.**

Malaria is a blood infection caused by mosquito (*Anopheles* spp.)-borne apicomplexan parasites of the genus *Plasmodium*[1–3]. Of five *Plasmodium* species known to infect humans, *P. falciparum* causes by far the greatest morbidity and mortality, with several hundred million cases of clinical malaria and more than one million deaths occurring annually[1,2]. Although much progress has been made in the treatment and prevention of *P. falciparum*[4], the origin and natural reservoirs of this and related plasmodial pathogens remain controversial. Until recently, the closest known relative of *P. falciparum* was a chimpanzee parasite, *Plasmodium reichenowi*, which was assumed to have diverged from its human counterpart at the same time as the ancestors of chimpanzees and humans, more than 5,000,000 years ago[5–8]. Within the past year, other closely related *Plasmodium* strains have been detected in chimpanzees, western gorillas and bonobos, raising the possibility that *P. falciparum* in humans could have arisen as a consequence of cross-species transmission from one or more of these apes[9–12]. However, all of these studies were limited by an analysis of only few apes, many of which were captive and living in close proximity to humans. In addition, all prior studies used non-limiting dilution polymerase chain reaction (PCR) amplification methods that are prone to generating artefactual mosaic sequences by recombination between genetically distinct templates. Here we used conventional and single-template PCR amplification methods to screen and analyse wild-living chimpanzee, bonobo and gorilla populations across sub-Saharan Africa for parasites related to *P. falciparum*.

Prevalence of ape *Plasmodium* infections

To determine the geographic distribution, species association and prevalence of ape *Plasmodium* spp. infections, we adapted a previously described PCR-based diagnostic method[10] to amplify a 956-base-pair (bp) fragment of *Plasmodium* cytochrome *b* (*cytb*) sequences from faecal DNA (Supplementary Fig. 1a). Ape faecal samples were selected from existing specimen banks that we had collected earlier for molecular epidemiological studies of simian retrovirus infections[13–16]. Except for 28 samples from one habituated gorilla community at field site DS (Fig. 1), all other specimens were derived from non-habituated apes living in remote forest areas (Supplementary Table 1). Chimpanzee ($n = 1{,}827$), gorilla ($n = 805$) and bonobo ($n = 107$) samples were subjected to diagnostic PCR, and all amplification products were sequenced to confirm *Plasmodium* infection. In addition, we subjected a subset of samples ($n = 1{,}027$), including all specimens from eastern gorillas and bonobos, to microsatellite analysis of host genomic DNA[14–16] to determine the number of individuals tested at particular field sites (Supplementary Table 1). Microsatellite analysis also provided quantitative estimates of specimen integrity (Supplementary Table 2) and redundant sampling (Supplementary Table 3), thereby allowing us to determine the sensitivity of the non-invasive diagnostic test by identifying the proportion of PCR-positive specimens from infected apes that were sampled more than once. From 32 such individuals, we estimated the test sensitivity to be 57% (Supplementary Table 4) and calculated the prevalence of ape infection at each field

[1]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [2]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [3]Department of Ecology and Management of Plant and Animal Resources, Faculty of Sciences, University of Kisangani, Kisangani, BP 2012, Democratic Republic of the Congo. [4]Department of Anthropology, Washington University, Saint Louis, Missouri 63130, USA. [5]Congo Program, Wildlife Conservation Society, Brazzaville, BP 14537, Republic of the Congo. [6]Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo, Chicago, Illinois 60614, USA. [7]Department of Biological Sciences, University at Albany, State University of New York, Albany, New York 12222, USA. [8]Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [9]VaccinApe, Bethesda, Maryland 200816, USA. [10]Institut de Recherche pour le Développement and University of Montpellier 1, 34394 Montpellier, France. [11]Institut de Recherches Médicales et d'Etudes des Plantes Médicinales Prévention du Sida au Cameroun, Centre de Recherche Médicale, BP 906, Yaoundé, République du Cameroun. [12]Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [13]Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131, USA. [14]Institute of Evolutionary Biology and Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3JT, UK. [15]Sanger Institute Malaria Programme, The Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. †Present address: The AIDS and Cancer Virus Program, Science Applications International Corporation-Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland 21702, USA.

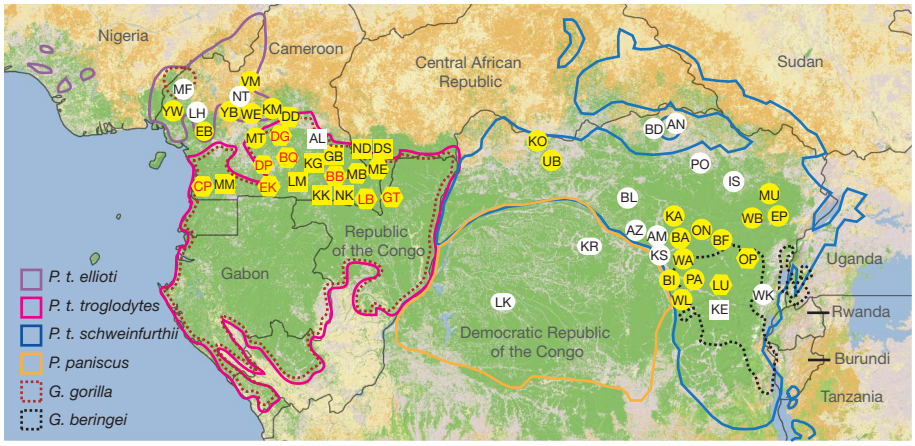


Figure 1 | Location of ape study sites. Field sites are shown in relation to the ranges of three subspecies of the common chimpanzee (*P. t. ellioti*, *P. t. troglodytes* and *P. t. schweinfurthii*), the western (*G. gorilla*) and eastern (*G. beringei*) gorillas, and the bonobo (*P. paniscus*) in sub-Saharan Africa (coloured as indicated). Forested areas are indicated in green, and arid areas are shown in yellow (map courtesy L. Pintea, The Jane Goodall Institute). Circles, squares and hexagons identify field sites where chimpanzees, gorillas or both were sampled, respectively; ovals indicate bonobo collection sites. Sites where ape malaria was detected are highlighted in yellow, with red lettering indicating that both chimpanzees and gorillas were infected.

site (Supplementary Table 1). The results revealed widespread *Plasmodium* infection in chimpanzees and western gorillas, but not in eastern gorillas or bonobos.

Ape malaria parasites were detected at 32 of 45 chimpanzee collection sites, and at 17 of 20 western gorilla collection sites (Fig. 1), including every site where at least ten individuals were estimated to have been sampled. *Plasmodium* infection was endemic in Nigeria–Cameroon chimpanzees (*Pan troglodytes ellioti*), central (*Pan troglodytes troglodytes*) and eastern (*Pan troglodytes schweinfurthii*) chimpanzees, and in western lowland gorillas (*Gorilla gorilla gorilla*), with estimated prevalence rates ranging from 32% to 48% (Table 1). The true infection rates are likely to be even higher, because *Plasmodium* detection in faecal samples can be expected to be less sensitive than detection in blood, as is the case for urine and saliva samples[17]. Although wild-living western chimpanzees (*Pan troglodytes verus*) and Cross River gorillas (*Gorilla gorilla diehli*) were not tested in this study, these two subspecies have previously been shown to harbour *Plasmodium* parasites in the wild[9,10]. On the basis of these data, it is clear that chimpanzees and western gorillas represent substantial *Plasmodium* reservoirs. Unexpectedly, we did not find this to be true for eastern gorillas or bonobos. Screening 71 and 58 members of these respective species at multiple field sites, we failed to detect *Plasmodium* infection in any of them (Supplementary Table 1). These findings suggest that malaria parasites are rare or absent in some wild-living ape communities, possibly reflecting regional, ecological or seasonal differences in the distribution and/or host specificities of the transmitting mosquito vectors. Additional field studies are needed to determine whether eastern gorillas and bonobos are infected by *Plasmodium* parasites at other locations or if they harbour divergent parasites not detected by current diagnostic assays.

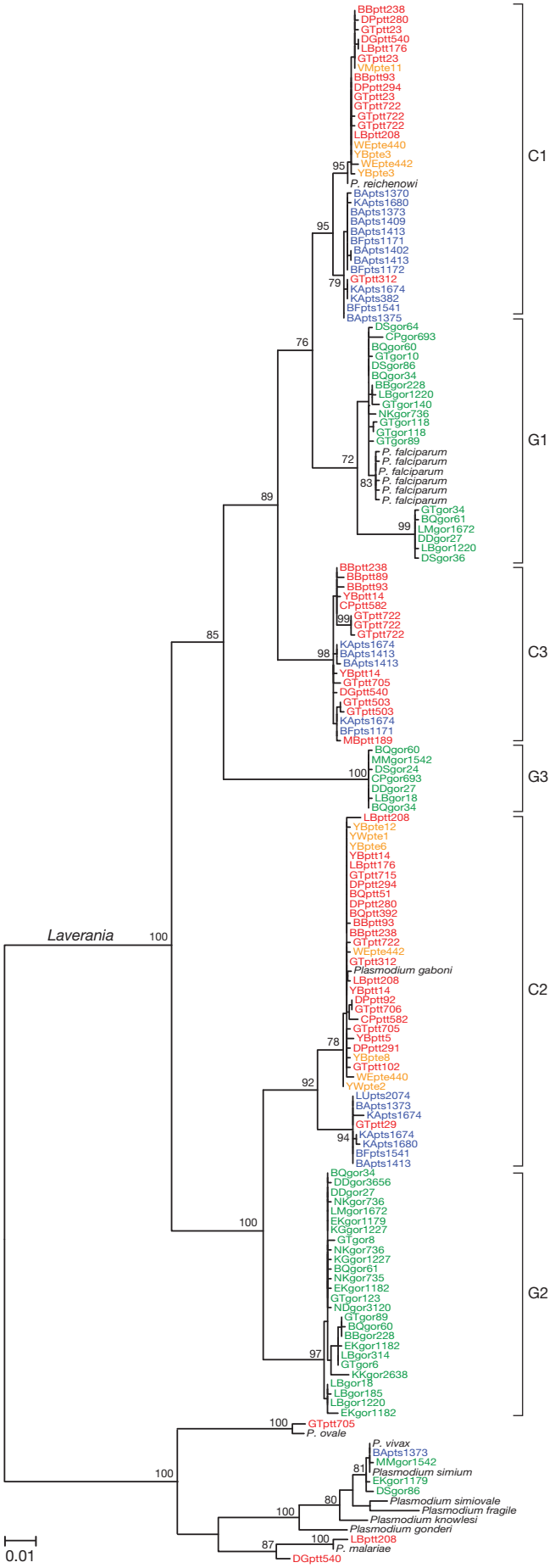
Analysis of ape *Plasmodium* species by single-genome amplification

To examine the evolutionary relationships of the newly identified *Plasmodium* parasites, we constructed phylogenetic trees for a subset of the diagnostic *cytb* sequences. This analysis showed that all sequences, except for one *Plasmodium ovale*-like strain, fell into one large monophyletic clade that also included *P. reichenowi* and *P. falciparum* (Supplementary Fig. 2). Parasites related to *P. reichenowi* and *P. falciparum* have previously been classified into a subgenus, termed *Laverania*, to distinguish them from more-divergent *Plasmodium* species[18]. Our results thus indicated that parasites from this subgenus were common and widespread among wild ape populations. However, the topology of the *Laverania* clade was highly unusual, characterized by only few discrete clades and multiple sequences with very short branches attached to internal branches. Moreover, repeated PCR analysis of the same faecal samples yielded sequences that clustered variably in different parts of the tree (Supplementary Fig. 2). These findings indicated simultaneous infection with genetically diverse *Plasmodium* parasites and suggested that conventional (bulk) PCR amplification had generated *in vitro* recombinants. To examine this possibility, we reanalysed the same *Plasmodium*-positive faecal samples by single-genome amplification (SGA), a molecular strategy that has been used extensively to characterize the genetic identity and quasispecies complexity of human and simian immunodeficiency viruses[19–23]. Faecal DNA was diluted so that fewer than 30% of all PCR reactions yielded an amplification product, which ensured amplification of single *Plasmodium* templates in most reactions[19–23]. All amplicons were sequenced directly and sequences containing mixed bases indicative of more than one amplified template were discarded. Using this approach to characterize the genetic complexity of malaria parasites in faecal samples, we could

Table 1 | Prevalence of *Plasmodium* spp. infection in wild-living African apes

Species or subspecies	Field sites*	Faecal samples tested	Faecal samples positive†	*Plasmodium* spp. prevalence (CI)‡
Nigeria–Cameroon chimpanzee (*Pan troglodytes ellioti*)	9	119	19	32% (23–46%)
Central chimpanzee (*Pan troglodytes troglodytes*)	13	612	147	48% (44–53%)
Eastern chimpanzee (*Pan troglodytes schweinfurthii*)	24	1,096	187	34% (30–40%)
Western lowland gorilla (*Gorilla gorilla gorilla*)	20	659	120	37% (32–41%)
Eastern lowland gorilla (*Gorilla beringei graueri*)	3	146	0	0% (0–4%)
Bonobo (*Pan paniscus*)	2	107	0	0% (0–6%)

* Field sites are listed in Supplementary Table 1 and their location is shown in Fig. 1.
† Faecal samples were tested for *Plasmodium* mitochondrial (*cytb*) DNA by diagnostic PCR; all amplification products were sequenced to confirm *Plasmodium* infection.
‡ Percentage prevalence of *Plasmodium* infection with brackets indicating 95% confidence intervals (CI). Values were estimated on the basis of the proportion of PCR-positive faecal samples, correcting for specimen degradation (Supplementary Table 2), repeat sampling (Supplementary Table 3), and the sensitivity of the diagnostic test (Supplementary Table 4). We note that these prevalence rates represent minimum estimates, because the extent to which infected apes shed *Plasmodium* DNA into their faeces is not known.



eliminate both Taq-polymerase-induced recombination (template switching) and nucleotide misincorporations in finished sequences, thereby ensuring an accurate representation of plasmodial variants as they existed *in vivo*[21–23].

Figure 2 depicts the phylogenetic relationships of a subset of SGA-derived mitochondrial *cytb* sequences (the entire set of 697 sequences is analysed in Supplementary Fig. 3). As in the corresponding tree of bulk-PCR-derived sequences (Supplementary Fig. 2), all SGA-derived sequences, except for seven *P. ovale*, *Plasmodium vivax* and *Plasmodium malariae*-like strains, grouped within the *Laverania* radiation. However, unlike in the bulk PCR tree, *Laverania* sequences in the SGA tree clustered in a strictly host-species-specific manner, forming three chimpanzee-specific (C1–C3) and three gorilla-specific (G1–G3) clades, each supported by high bootstrap values. This host specificity did not extend to the subspecies level, because *P. t. ellioti*, *P. t. troglodytes* and *P. t. schweinfurthii*-derived sequences were interspersed; however, *cytb* sequences from *P. t. schweinfurthii* segregated into distinct subclades within two of the three chimpanzee lineages (C1 and C2), suggesting a phylogeographic distribution of certain *Plasmodium* variants (Supplementary Fig. 3a, b). None of 363 chimpanzee-derived *Plasmodium cytb* sequences was closely related to human *P. falciparum*. Instead, all human sequences grouped within a single clade of parasites (G1) that infected western gorillas at numerous sites in Cameroon (LB, BB, CP, NK, BQ, DD, MM and LM), the Central African Republic (DS and ND) and the Republic of the Congo (GT) (Fig. 2 and Supplementary Fig. 3d). A notable finding of the SGA analysis, which was obscured by bulk PCR analysis, was that most apes were co-infected with parasites representing multiple different plasmodial lineages, including variants from the same *Laverania* clade, from different *Laverania* clades or from *Laverania* and non-*Laverania* clades (Supplementary Table 5). Of 65 chimpanzee and 53 gorilla samples characterized, 48 (74%) and 37 (70%), respectively, harboured more than one genetically distinct parasite strain, and 36 (55%) and 23 (43%) contained members of two or more major *Plasmodium* clades (Supplementary Fig. 3). Given this high frequency of co-infection with divergent parasites, conventional recombination-prone PCR approaches are not appropriate for generating ape *Plasmodium* sequences for phylogenetic analysis. Moreover, previously reported ape *Plasmodium* sequences[9–12] must be interpreted with caution because they were subject to these same confounding factors.

To test the robustness of the phylogenetic relationships depicted in Fig. 2, we used SGA to amplify additional genomic regions from *cytb*-positive faecal samples, targeting loci in the mitochondrial, apicoplast and nuclear *Plasmodium* genomes. These regions included 390 bp of the caseinolytic protease C (*clpC*) gene ($n = 126$), 772 bp of the lactase dehydrogenase (*ldh*) gene ($n = 46$), and 3.4-kilobase (kb) ($n = 165$) and 3.3-kb ($n = 127$) fragments that together spanned the entire mitochondrial genome (Supplementary Fig. 1a). Phylogenetic analyses of each of these genomic loci revealed very similar topologies. In trees of *clpC* (Supplementary Fig. 4), *ldh* (Supplementary Fig. 5) and mitochondrial sequences (Supplementary Figs 6 and 7), *Laverania* sequences formed the same number of chimpanzee-specific (C1–C3) and

Figure 2 | Phylogeny of *Plasmodium* parasites from wild-living chimpanzees and western gorillas. A representative subset of 146 SGA-derived *Plasmodium* mitochondrial cytochrome *b* sequences (956 bp) is shown in relation to human and simian *Plasmodium* reference sequences (for accession numbers, see Supplementary Tables 6 and 7). The full set of 697 SGA-derived ape *Plasmodium* sequences is analysed in Supplementary Fig. 3. Sequences are colour-coded, with capital letters indicating the field site (Fig. 1) and lower-case letters denoting species and subspecies origin (ptt: *P. t. troglodytes*, red; pte: *P. t. ellioti*, orange; pts: *P. t. schweinfurthii*, blue; gor: *G. g. gorilla*, green). C1–C3 and G1–G3 represent chimpanzee- and gorilla-specific *Plasmodium* species, respectively, all of which are included within the subgenus *Laverania*[18]. The tree was inferred using maximum-likelihood methods[30]. Bootstrap values (>70%) are indicated for major nodes only (the scale bar represents 0.01 substitutions per site).

gorilla-specific (G1–G3) clades, albeit with some variations in the relationships among these lineages. Importantly, there was no evidence of recombination between chimpanzee- and gorilla-specific parasites, although many of them infected apes at the same field sites. This suggested that *Laverania* parasites are largely host specific (recombination between parasites infecting the same host species could not be assessed because of mixed *Plasmodium* infections). These findings, together with the extent of genetic diversity that distinguishes the various clades, argue strongly for the existence of six distinct *Plasmodium* species within the *Laverania* subgenus (Supplementary Figs 3–8). Formal classification of these lineages must await additional taxonomic evaluation.

Origin of human *P. falciparum*

The new SGA-derived ape *Plasmodium* sequences call for a reassessment of the origin of human *P. falciparum*. Among over 600 sequences derived from ape samples spanning most of central Africa, we failed to find a single chimpanzee parasite that was sufficiently closely related to *P. falciparum* to represent a progenitor (Fig. 2 and Supplementary Figs 3–8). Thus, *P. reichenowi*, as well as other chimpanzee *Plasmodium* species, can be excluded as precursors of *P. falciparum*. Instead, all new phylogenetic evidence points to a western gorilla origin of human *P. falciparum* (Fig. 2 and Supplementary Figs 3–8). To investigate how often gorilla parasites might have colonized humans, we constructed phylogenetic trees from complete mitochondrial genome equivalents of the closest *Plasmodium* relatives of human *P. falciparum* (Fig. 3). In a tree of concatenated CYTB, COXI and COXIII protein sequences (980 amino acids), all available human *P. falciparum* sequences ($n = 105$) coalesced to a single common ancestor nested within the G1 clade of gorilla parasites (Fig. 3a). Nucleotide sequences from the remaining (non-coding) portions of the mitochondrial genome yielded a very similar topology, again showing that human *P. falciparum* formed a monophyletic lineage within the gorilla *P. falciparum* radiation (Fig. 3b). These findings, together with the observation that human parasites exhibit substantially less sequence diversity than the various ape *Plasmodium* species, including the closest gorilla relatives (Table 2), provide compelling evidence for a gorilla origin of human *P. falciparum*. Moreover, the monophyly of the human parasite sequences (Fig. 3) may indicate that all extant human strains evolved from a single gorilla-to-human cross-species transmission event. Notably, four recently reported *Plasmodium* sequences from captive bonobos[11] also clustered closely with *P. falciparum*. However, unlike the gorilla sequences, the bonobo sequences were interspersed with the human sequences (Fig. 3). This finding, together with the fact that the bonobo parasites encoded dihydrofolate reductase/thymidylate synthase (*dhfr-ts*) drug-resistance mutations prevalent in the local human population[11], suggests that the bonobos became infected with human parasites while housed in an urban sanctuary. In fact, the topologies in Fig. 3 are consistent with more than one human-to-bonobo transmission, although some (or all) of the substitutions that distinguish bonobo and human sequences could represent PCR misincorporations because they were not generated by SGA methods[11].

Discussion

Using single-template amplification strategies and a much larger collection of ape specimens than previously analysed, we show here that wild-living chimpanzees and western gorillas are naturally infected with at least nine *Plasmodium* species. Among more than 1,100 SGA-derived mitochondrial, apicoplast and nuclear gene sequences from 80 chimpanzee and 55 gorilla samples, we found a total of nine

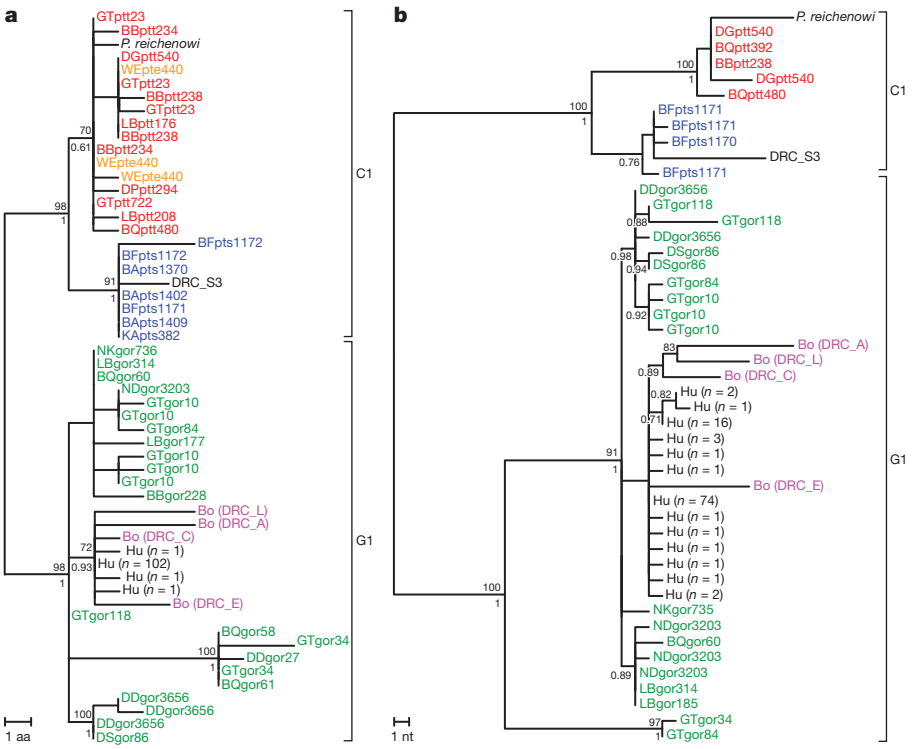


Figure 3 | Evolutionary relationships of ape and human *Plasmodium* parasites in mitochondrial coding and non-coding regions.
a, **b**, Phylogenetic trees of SGA-derived mitochondrial CYTB, COXI and COXIII proteins (980 amino acids; **a**) and non-overlapping nucleotide sequences (2,501 bp; **b**) of the closest chimpanzee (C1) and gorilla (G1) relatives of human *P. falciparum* (see Supplementary Fig. 1a for the position of the corresponding 3.4-kb and 3.3-kb SGA amplicons). *Plasmodium* sequences are labelled and colour-coded as in Fig. 2, except for sequences derived from captive bonobos[11] (Bo), which are shown in magenta. Human *P. falciparum* (Hu) and chimpanzee reference sequences are depicted in black, with the number of human sequences representing the same haplotype shown in parentheses (for accession numbers, see Supplementary Table 7). The trees were inferred using maximum-likelihood methods[30]; numbers at nodes indicate bootstrap (above) and posterior probability (below) values, respectively (only bootstrap values >70 and posterior probabilities >0.7 are shown). The scale bars represent one amino acid (aa; **a**) and one nucleotide (nt) substitution per site (**b**).

Table 2 | Genetic diversity within different *Laverania* species

Species	*cytb**				mtDNA-3.4 kb*				mtDNA-3.3 kb*			
	n	Mean	Median	Max.	*n*	Mean	Median	Max.	*n*	Mean	Median	Max.
C1	58	0.0055	0.0031	0.0136	25	0.0069	0.0024	0.0148	9	0.0035	0.0052	0.0060
G1(gor/all)†	51	0.0085	0.0021	0.0251	31	0.0058	0.0015	0.0193	19	0.0029	0.0016	0.0108
G1(gor/main)†	40	0.0018	0.0011	0.0073	26	0.0012	0.0012	0.0027	17	0.0015	0.0012	0.0040
G1(hum)†	105	0.0005	0	0.0021	105	0.0003	0.0003	0.0015	105	0.0002	0	0.0012
C2	109	0.0077	0.0021	0.0262	39	0.0078	0.0024	0.0202	16	0.0044	0.0020	0.0112
G2	92	0.0031	0.0031	0.0115	25	0.0032	0.0036	0.0071	36	0.0015	0.0016	0.0036
C3	48	0.0029	0.0031	0.0094	13	0.0035	0.0030	0.0071	8	0.0026	0.0028	0.0044
G3	23	0.0017	0.0021	0.0042	14	0.0010	0.0009	0.0033	4	0	0	0

* Pairwise distances were determined for SGA-derived *cytb*, mtDNA-3.4 kb and mtDNA-3.3 kb sequences for each of the six *Laverania* species as shown in Supplementary Figs 3, 6 and 7, respectively. The number of sequences (*n*) analysed for each clade is indicated. Sequences were aligned and gap-stripped to remove ambiguous sites. Alignments were then used to determine the mean, median and maximum sequence distances. Identical sequences from different samples were included (identical sequences from the same sample were excluded). The sequence overlap between the mtDNA-3.3 kb and mtDNA-3.4 kb fragments was removed from the mtDNA-3.3 kb comparisons.

† The *P. falciparum* clade (G1) was subdivided into three separate groups: G1(gor/all) included all gorilla *P. falciparum* sequences, G1(gor/main) included only the closest gorilla relatives of human *P. falciparum* and G1(hum) included only human *P. falciparum* sequences. The comparisons show that human *P. falciparum* strains have much lower diversity than the various clades of ape-derived strains (except for G3, where only very few sequences were available for the 3.3-kb fragment); in particular, the G1(gor/main) clade has substantially higher diversity than the G1(hum) clade.

sequences that were related to *P. malariae*, *P. ovale* or *P. vivax* (Supplementary Table 5). All others grouped within one of six chimpanzee- or gorilla-specific lineages representing distinct *Plasmodium* species, three of which had not previously been described. Significantly, all currently available human *P. falciparum* sequences constitute a single lineage nested within the G1 clade of gorilla parasites. This indicates that human *P. falciparum* is of gorilla origin, and not of chimpanzee[9,10,12], bonobo[11] or ancient human[5] origin, and that all known human strains may have resulted from a single cross-species transmission event. What is still unclear is when gorilla *P. falciparum* entered the human population and whether present-day ape populations represent a source for recurring human infection. It has been suggested that the limited levels of genetic diversity seen at many loci in human *P. falciparum* reflect a relatively recent selective sweep[8]. Our data suggest that this bottleneck or 'Eve event' was instead the consequence of cross-species transmission of a gorilla parasite. It is difficult to date this event without having reliable dates with which to calibrate the *Plasmodium* phylogenetic trees. Previous estimates of dates in the evolution of *Plasmodium* have relied largely on the belief that *P. falciparum* and *P. reichenowi* diverged at the same time as the ancestors of humans and chimpanzees[6–8,24], an assumption that is now groundless. Others have proposed a much shorter timescale coincident with the emergence of agricultural societies in sub-Saharan Africa, the incomplete penetration of protective human gene polymorphisms (for example haemoglobin C) that are selected by *P. falciparum* infection, or the speciation of African mosquito vectors[3,25]. Complete sequence analysis of members of the ape *Plasmodium* species identified here may help to resolve this conundrum. The second question, of whether additional cross-species transmissions of *Laverania* parasites have given rise to human infections, is more immediately addressable. An alignment of over 100 ape *Plasmodium* mitochondrial genome sequences reveals ape-specific single nucleotide polymorphisms (Supplementary Fig. 1b), which can now be used to screen plasmodial sequences from humans living in close proximity to wild gorillas and chimpanzees. Such studies can inform malaria eradication efforts about potential zoonotic *Plasmodium* reservoirs and provide insights into adaptive changes that might be required for ape *Plasmodium* infection of humans[26].

METHODS SUMMARY

Ape samples. We selected faecal samples from wild-living chimpanzees, gorillas and bonobos from existing specimens banks[13–16] on the basis of their geographic location, available host genetic information, and species and subspecies origin (Supplementary Table 1).

Conventional PCR. *Plasmodium* mitochondrial, apicoplast and nuclear gene sequences were amplified as previously described[9,10,27,28], but using modified primers and PCR conditions suitable for faecal DNA. Bulk-PCR-positive faecal samples were subsequently subjected to SGA analysis.

Single-genome amplification. We performed SGA analysis of *Plasmodium* sequences from faecal DNA as previously described[21,22]. All amplicons were sequenced directly, and sequences containing double peaks were discarded.

Prevalence estimations. The prevalence rates of *Plasmodium* infection were estimated on the basis of the proportion of PCR-positive faecal samples, correcting for specimen degradation, repeat sampling and the sensitivity of the diagnostic test.

Phylogenetic analysis. We inferred the phylogenetic trees of newly derived ape *Plasmodium* mitochondrial, apicoplast and nuclear gene sequences by Bayesian[29] and maximum-likelihood methods[30] (GenBank accession numbers are listed in Supplementary Table 6).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 May; accepted 20 August 2010.

1. Greenwood, B. M., Bojang, K., Whitty, C. J. & Targett, G. A. Malaria. *Lancet* **365,** 1487–1498 (2005).
2. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434,** 214–217 (2005).
3. Carter, R. & Mendis, K. N. Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.* **15,** 564–594 (2002).
4. Kappe, S. H., Vaughan, A. M., Boddey, J. A. & Cowman, A. F. That was then but this is now; malaria research in the time of an eradication agenda. *Science* **328,** 862–866 (2010).
5. Escalante, A. A. & Ayala, F. J. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl Acad. Sci. USA* **91,** 11373–11377 (1994).
6. Escalante, A. A., Barrio, E. & Ayala, F. J. Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* **12,** 616–626 (1995).
7. Jeffares, D. C. *et al.* Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nature Genet.* **39,** 120–125 (2006).
8. Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95,** 4425–4430 (1998).
9. Rich, S. M. *et al.* The origin of malignant malaria. *Proc. Natl Acad. Sci. USA* **106,** 14902–14907 (2009).
10. Prugnolle, F. *et al.* African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **107,** 1458–1463 (2010).
11. Krief, S. *et al.* On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from bonobos. *PLoS Pathogens* **6,** e1000765 (2010).
12. Duval, L. *et al.* African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. *Proc. Natl Acad. Sci. USA* **107,** 10561–10566 (2010).
13. Liu, W. *et al.* Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. *PLoS Pathogens* **4,** e1000097 (2008).
14. Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313,** 523–526 (2006).
15. Neel, C. *et al.* Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *J. Virol.* **84,** 1464–1476 (2010).
16. Van Heuverswyn, F. *et al.* Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. *Virology* **368,** 155–171 (2007).
17. Nwakanma, D. C. *et al.* Quantitative detection of *Plasmodium falciparum* DNA in saliva, blood, and urine. *J. Infect. Dis.* **199,** 1567–1574 (2009).
18. Bray, R. S. Studies on malaria in chimpanzees. VI. *Laverania falciparum*. *Am. J. Trop. Med. Hyg.* **7,** 20–24 (1958).
19. Simmonds, P. *et al.* Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J. Virol.* **64,** 864–872 (1990).
20. Palmer, S. *et al.* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* **43,** 406–413 (2005).

21. Salazar-Gonzalez, J. F. *et al.* Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* **82,** 3952–3970 (2008).
22. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **105,** 7552–7557 (2008).
23. Keele, B. F. *et al.* Low dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. *J. Exp. Med.* **206,** 1117–1134 (2009).
24. Su, X.-Z., Mu, J. & Joy, D. A. The ''Malaria's Eve'' hypothesis and the debate concerning the origin of the human malaria parasite *Plasmodium falciparum*. *Microbes Infect.* **5,** 891–896 (2003).
25. 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).
26. Martin, M. J., Rayner, J. C., Gagneux, P., Barnwell, J. W. & Varki, A. Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. *Proc. Natl Acad. Sci. USA* **102,** 12819–12824 (2005).
27. Ollomo, B. *et al.* A new malaria agent in African hominids. *PLoS Pathogens* **5,** e1000446 (2009).
28. Duval, L. *et al.* Chimpanzee malaria parasite related to *Plasmodium ovale* in Africa. *PLoS ONE* **4,** e5520 (2009).
29. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
30. Guindon, S., Delsuc, F., Dufayard, J. F. & Gascuel, O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **537,** 113–137 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Neel, S. Loul, A. Mebanga, B. Yangda and F. Liegeois for field work in Cameroon; the Cameroonian Ministries of Health, Forestry and Wildlife, and Research for permission to collect samples in Cameroon; the Water and Forest Ministry for permission to collect samples in the Central African Republic; the Ministries of Science and Technology and Forest Economy for permission to collect samples in the Republic of the Congo; the Ministry of Scientific Research and Technology and the Department of Ecology and Management of Plant and Animal Resources of the University of Kisangani for permission to collect samples in the Democratic Republic of the Congo; M. Ndunda, S. Coxe, A. Lokasola, A. Todd and the staff of the World Wildlife Fund in the Central African Republic for logistical support; R. Carter for helpful discussions; M. Salazar, Y. Chen and B. Cochran for technical assistance; and J. White for artwork and manuscript preparation. This work was supported by grants from the National Institutes of Health (R01 AI50529, R01 AI58715, U19 AI 067854, R03 AI074778, T32 GM008111, T32 AI007245, P30 AI 27767), the Bill & Melinda Gates Foundation (37874), the National Science Foundation (0755823), the Agence Nationale de Recherche sur le Sida (12152/12182), the Great Ape Conservation Fund of the US Fish and Wildlife Service, the Arthur L. Greene Fund, the Wallace Global Fund, the Bristol Myers Freedom to Discover Program and the Wellcome Trust. R.S.R. was supported by a Howard Hughes Medical Institute Med-into-Grad Fellowship.

Author Contributions All authors contributed to the acquisition, analysis and interpretation of the data; W.L., M.P., J.C.R., P.M.S. and B.H.H. initiated and designed the study; W.L., Y.L. and J.D.R. performed non-invasive *Plasmodium* testing and SGA analyses; B.F.K, R.S.R and J.D.R. performed microsatellite analyses; P.M.S. calculated *Plasmodium* prevalence rates; G.H.L. and P.M.S performed phylogenetic analyses; J.-B.N.N., C.M.S., D.B.M., S.L., M.K.G., P.J.K., P.D.W., E.D., E.M.-N., A.V.G. and M.N.M. conducted and supervised all fieldwork; and W.L., G.M.S., M.P., P.M.S., J.C.R. and B.H.H. coordinated the contributions of all authors and wrote the paper.

Author Information SGA-derived *Plasmodium* nucleotide sequences have been deposited in GenBank under accession numbers HM234976–HM235117 and HM237301 (*cytb*), HM235118–HM235143 (*ldh*), HM235144–HM235170 (*clpC*), HM235171–HM235268 (mtDNA-3.3 kb) and HM235269–HM235404 (mtDNA-3.4 kb) (also see Supplementary Table 6). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.H.H. (bhahn@uab.edu).

METHODS

Ape faecal samples. To screen wild ape populations for *Plasmodium* infection, we selected 2,739 faecal samples from an existing bank of chimpanzee (*P. troglodytes*), western gorilla (*G. gorilla*), eastern gorilla (*G. beringei*) and bonobo (*P. paniscus*) specimens previously collected for molecular epidemiological studies of simian immunodeficiency virus (SIV_{CPZ} and SIV_{GOR})[14–16,31] and simian foamy virus (SFV_{CPZ})[13]. All of these specimens, except for 28 samples from a group of habituated western gorillas (Makumba group) at the DS field site, were collected from non-habituated apes living in remote forest areas. Faecal samples were first subjected to host mitochondrial DNA analysis to determine their species and subspecies origin[13–16,31]. A subset was then selected for host microsatellite analysis to determine the number of individuals at particular field sites (Supplementary Table 1). These included 198 chimpanzee samples from the GT field site, 189 eastern gorilla samples from the KE, LU and OP field sites, and 119 bonobo samples from the LK and KR field sites (Supplementary Table 2). For estimates of sample degradation (Supplementary Table 2) and oversampling (Supplementary Table 3), we also included microsatellite results that we had obtained earlier for specimens collected from central chimpanzees[14–16] and western gorillas[15] in Cameroon, as well as from eastern chimpanzees[31] in the Democratic Republic of the Congo.

Microsatellite analyses. Faecal DNA was extracted[14] and used to amplify four (GT) or eight polymorphic microsatellite loci (KE, LU, OP, LK and KR) as previously described[14,15]. Amplification products were analysed on an automated sequencer (Applied Biosystems) and sized using GENEMAPPER 4.0 (Applied Biosystems). For individual identification, samples were first grouped by field site and mitochondrial DNA haplotype. Within each haplotype, samples were then grouped by microsatellite genotypes, but allowing for one allelic mismatch between samples. Specimens were classified as degraded if they failed to amplify two or more (GT) or three or more (all other sites) microsatellite loci. Samples with evidence of DNA admixture (multiple peaks for the same locus) were discarded.

Amplification of ape *Plasmodium* sequences by conventional (bulk) PCR. Faecal samples were first screened for *Plasmodium cytb* sequences (956 bp) as previously described[10,27], but using modified PCR conditions as well as a different second-round reverse primer to generate an amplicon 166 bp longer. Nested PCR was performed using DW2 (5′-TAATGCCTAGACGTATTCCTGATTATCCAG-3′) and DW4 (5′-TGTTTGCTTGGGAGCTGTAATCATAATGTG-3′) in the first round of PCR and Pfcytb1 (5′-CTCTATTAATTTAGTTAAAGCACA-3′) and PLAS2a (5′-GTGGTAATTGACATCCWATCC-3′) in the second round. For the first round, 2.5 µl faecal DNA was used in a 25-µl reaction volume, containing 0.5 µl dNTPs (10 mM of each dNTP), 20 pmol of each primer (DW2 and DW4), 2.5 µl PCR buffer and 0.25 µl Expand Long Template enzyme mix (Roche). Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 15 cycles of denaturation (94 °C, 10 s), annealing (45 °C, 30 s) and elongation (68 °C, 2 min), followed by 35 cycles of denaturation (94 °C, 10 s), annealing (48 °C, 30 s) and elongation (68 °C, 2 min; with 15-s increments for each successive cycle), followed by a final elongation step of 10 min at 68 °C. For the second round of PCR, 1 µl of the first-round product was used in a 25-µl reaction volume containing 0.5 µl dNTPs (10 mM of each dNTP), 20 pmol of each primer (Pfcytb1 and PLAS2a), 2.5 µl PCR buffer and 0.25 µl Expand Long Template enzyme mix. Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 60 cycles of denaturation (94 °C, 10 s), annealing (52 °C, 30 s) and elongation (68 °C 1 min), followed by a final elongation step of 10 min at 68 °C. Amplified products were gel-purified and sequenced directly to confirm *Plasmodium* infection.

Samples positive for *Plasmodium cytb* sequences were then screened for apicoplast *clpC* (390 bp) and nuclear *ldh* (772 bp) sequences. Amplification of the *clpC* fragment was performed as previously described[9], but using modified PCR conditions and a different second-round reverse primer to generate an amplicon 117 bp longer. Nested PCR was performed using primers TFM1421+ (5′-AAAAC TGAATTAGCAAAAATATTA-3′) and TFM1423RC (5′-CGAGCTCCATATAA AGGAT-3′) in the first round of PCR and CLPCF1 (5′-TCTAAACAATTATTTG GTTCTG-3′) and CLPCR2 (5′-GTTAATCTATTTARTAATTCHGGTTTAA-3′) in the second round. Amplification of the *ldh* fragment was also performed as previously described[28,32,33], but using different PCR conditions. Nested PCR was performed using primers JNB272 (5′-ATGGCACCAAAAGCAAAAAT-3′) and JNB273 (5′-GCCTTCATTCTSYTAGTTTCAGC-3′) for the first round and LDH1 (5′-GGNTCDGGHATGATHGGAGG-3′) and Fv2n (5′-AACRASAGG WGTACCACC-3′) for the second round. PCR conditions were the same as described for the *cytb* fragment. Amplified products were gel-purified and sequenced directly to confirm *Plasmodium* infection.

Finally, *cytb*-positive samples were subjected to nested PCR with the aim of amplifying larger fragments (3.4 kb and 3.3 kb in length, respectively), which together spanned the entire *Plasmodium* mitochondrial genome (Supplementary Fig. 1a). The 3.4-kb fragment was amplified using Pf936p (5′-GAGAAAAA TGYAATCCWGTWACACAATA-3′) and DW4 in the first round of PCR and Pf1031p (5′-GATGCAAAACATTRWCCTAATAAGTA-3′) and PLAS2a in the second round. The 3.3-kb fragment was amplified using McytP (5′-TATCCAA ATCTATTAAGTCTTG-3′) and Pf1916n (5′-GCGTTCGTTCTTATAGTGTAG GC-3′) in the first round of PCR and Pf4450p (5′-CTGTTCCTATTATATGGTTT ATGTGTGC-3′) and Pf1880n (5′-CCTTTAATGTAGTTTCCTCACAGCTT-3′) in the second round. For the first round of amplification, 2.5 µl faecal DNA was used in a 25-µl reaction volume containing 0.5 µl dNTPs (10 mM of each dNTP), 20 pmol of each first-round primer, 2.5 µl PCR buffer and 0.25 µl Expand Long Template enzyme mix (Roche). Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 15 cycles of denaturation (94 °C, 10 s), annealing (45 °C, 30 s) and elongation (68 °C, 4 min), followed by 35 cycles of denaturation (94 °C, 10 s), annealing (48 °C, 30 s) and elongation (68 °C, 4 min; with 15-s increments for each successive cycle), followed by a final elongation step of 10 min at 68 °C. For the second round of amplification, 2 µl of the first-round PCR product was used in a 50 µl volume containing 1 µl dNTPs (10 mM of each dNTP), 40 pmol of each second-round primer, 5 µl PCR buffer and 0.5 µl Expand Long Template enzyme mix. Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 60 cycles of denaturation (94 °C, 10 s), annealing (52 °C, 30 s) and elongation (68 °C, 4 min), followed by a final elongation step of 10 min at 68 °C. Amplified products were gel-purified, but only a small fragment was sequenced to confirm *Plasmodium* infection.

SGA of *Plasmodium* sequences from faecal DNA. To derive sequences suitable for phylogenetic analyses, a subset of bulk-PCR-positive chimpanzee ($n = 80$) and gorilla ($n = 55$) faecal samples was subjected to SGA analyses[21,22]. Following a Poisson distribution, the DNA dilution that yields PCR products in no more than 30% of wells contains one amplifiable template per positive PCR more than 80% of the time. Faecal DNA was thus end-point diluted in 96-well plates, and the dilution for which fewer than 30% of the wells were positive was used to generate between one and 40 different SGA sequences per sample (Supplementary Table 5). The same primers and PCR conditions used for bulk amplification of *cytb*, mtDNA-3.4 kb, mtDNA-3.3 kb, *clpC* and *ldh* fragments were also used for SGA analyses. Amplification products were gel-purified, and sequenced directly using SEQUENCHER 4.9 (Gene Codes Corporation). Sequences that contained double peaks as an indicator of more than one amplified template were discarded.

Sensitivity and specificity of *Plasmodium* nucleic-acid detection in faecal samples. To estimate the sensitivity of the diagnostic *cytb* PCR test, we determined the proportion of PCR-positive specimens from *Plasmodium*-infected apes that were sampled more than once on the same day. Other replicate samples were excluded because the duration of natural ape *Plasmodium* infections is unknown. The sensitivity of *Plasmodium* nucleic-acid detection was then calculated as the fraction of positive tests per total number of samples tested (Supplementary Table 4). Including data from 32 such apes, we estimated the sensitivity to be 57% (with confidence limits determined assuming binomial sampling). It should be noted that this approach led to a systematic (albeit small) overestimation of the assay sensitivity, because it did not account for infected apes that yielded only negative replicate samples. Moreover, *Plasmodium* detection in faecal samples is very probably less sensitive than in blood, as is the case in urine and saliva[17]. Thus, the prevalence rates in Table 1 and Supplementary Table 1 should be interpreted as minimum estimates of *Plasmodium* infection rates in wild apes. The specificity of faecal *Plasmodium* detection was 1.00, because all amplification products were sequence-confirmed.

Ape *Plasmodium* prevalence estimations. For sites where the number of sampled chimpanzees was known (Supplementary Table 1), *Plasmodium* prevalence rates were estimated on the basis of the proportion of infected individuals. For each ape, the probability that it would be detected as being infected, if it was truly infected, was calculated taking into consideration the sensitivity of the diagnostic PCR test and the number of samples analysed, with 95% confidence limits determined assuming binomial sampling. For the remaining field sites, where the number of sampled individuals was not known, prevalence rates were estimated on the basis of the number of faecal samples, but correcting for specimen degradation and oversampling. As shown in Supplementary Table 2, microsatellite analysis of 1,027 faecal samples indicated an average degradation factor of 13%. Microsatellite analyses also provided a quantitative estimate of oversampling. Because of regional differences in sample collection, oversampling values were calculated separately for the different ape species and subspecies. As shown in Supplementary Table 3, central chimpanzees, western gorillas, eastern chimpanzees, eastern gorillas and bonobos were each assumed to have been sampled on average 1.77, 1.84, 3.74, 2.01 and 1.84 times, respectively. Using these corrections, the proportion of *Plasmodium*-infected chimpanzees was estimated for each field site, again taking into account the sensitivity of the diagnostic test. From these determinations, prevalence rates and their confidence limits were calculated.

Phylogenetic analyses. Ape-derived *Plasmodium* sequences were aligned with human and simian reference sequences using CLUSTAL W[34]. Sites that could not be aligned unambiguously were excluded. Trees were constructed from mitochondrial *cytb* sequences (956 bp, Supplementary Figs 2 and 3; 240 bp, Supplementary Fig. 8), apicoplast *clpC* sequences (390 bp, Supplementary Fig. 4), nuclear *ldh* sequences (772 bp, Supplementary Fig. 5) and mitochondrial half-genomes (3,361 bp, Supplementary Fig. 6; 3,277 bp, Supplementary Fig. 7). In addition, trees were constructed from mitochondrial coding (Fig. 3a) and non-coding regions (Fig. 3b). Deduced COXI, COXIII, and CYTB protein sequences were concatenated into a single 980-amino-acid sequence. The non-protein-coding portion of the mtDNA-3.3 kb fragment comprised 2,447 nucleotides following the removal of ambiguous sites. Phylogenetic trees were inferred using PHYML[30]. The class of evolutionary model was chosen using MODELTEST[35], and parameters were iteratively estimated in PHYML, using the GTR+I+G model for nucleotide sequence trees and the LG+I+G model (ref. 36) for amino-acid sequence trees. Bootstrap values were calculated with 100 replicates[37]. Posterior probability values were calculated with MRBAYES[29], using an average standard deviation of partition frequencies <0.01 as a convergence diagnostic. A neighbour-joining phylogenetic tree (Supplementary Fig. 8) was calculated with CLUSTAL W, using the Kimura two-parameter model of evolution with bootstrap support based on 1,000 bootstrap replicates[34,37].

Nucleotide sequence accession numbers. All new SGA-derived ape *Plasmodium* sequences have been submitted to GenBank, with accession numbers listed in Supplementary Table 6.

31. Li, Y. *et al.* in *Proc. 17th Conf. Retroviruses Opportunistic Infections* abstr. 440, ⟨http://www.retroconference.org/2010/PDFs/440.pdf⟩ (2010).
32. Talman, A. M. *et al.* Evaluation of the intra- and inter-specific genetic variability of *Plasmodium* lactate dehydrogenase. *Malar. J.* **6,** 140 (2007).
33. Brown, W. M. *et al.* Comparative structural analysis and kinetic properties of *lactate dehydrogenases* from the four species of human malarial parasites. *Biochemistry* **43,** 6219–6229 (2004).
34. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
35. Posada, D. & Buckley, T. R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53,** 793–808 (2004).
36. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25,** 1307–1320 (2008).
37. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39,** 783–791 (1985).