

ARTICLE

Received 11 Sep 2013 | Accepted 29 Jan 2014 | Published 21 Feb 2014

DOI: 10.1038/ncomms4346

# African origin of the malaria parasite *Plasmodium vivax*

Weimin Liu<sup>1</sup>, Yingying Li<sup>1</sup>, Katharina S. Shaw<sup>2</sup>, Gerald H. Learn<sup>1</sup>, Lindsey J. Plenderleith<sup>3</sup>, Jordan A. Malenke<sup>1</sup>, Sesh A. Sundararaman<sup>1,4</sup>, Miguel A. Ramirez<sup>1</sup>, Patricia A. Crystal<sup>1</sup>, Andrew G. Smith<sup>1</sup>, Frederic Bibollet-Ruche<sup>1</sup>, Ahidjo Ayouba<sup>5</sup>, Sabrina Locatelli<sup>5</sup>, Amandine Esteban<sup>5</sup>, Fatima Mouacha<sup>5</sup>, Emilande Guichet<sup>5</sup>, Christelle Butel<sup>5</sup>, Steve Ahuka-Mundeki<sup>5,6</sup>, Bila-Isia Inogwabini<sup>7</sup>, Jean-Bosco N. Ndjango<sup>8</sup>, Sheri Speede<sup>9</sup>, Crickette M. Sanz<sup>10,11</sup>, David B. Morgan<sup>11,12</sup>, Mary K. Gonder<sup>13</sup>, Philip J. Kranzusch<sup>14</sup>, Peter D. Walsh<sup>15</sup>, Alexander V. Georgiev<sup>16,†</sup>, Martin N. Muller<sup>17</sup>, Alex K. Piel<sup>15,18</sup>, Fiona A. Stewart<sup>15</sup>, Michael L. Wilson<sup>19</sup>, Anne E. Pusey<sup>20</sup>, Liwang Cui<sup>21</sup>, Zenglei Wang<sup>21</sup>, Anna Färnert<sup>22</sup>, Colin J. Sutherland<sup>23</sup>, Debbie Nolder<sup>23</sup>, John A. Hart<sup>24</sup>, Terese B. Hart<sup>24</sup>, Paco Bertolani<sup>25</sup>, Amethyst Gillis<sup>26</sup>, Matthew LeBreton<sup>26</sup>, Babila Tafon<sup>27</sup>, John Kiyang<sup>28</sup>, Cyrille F. Djoko<sup>26</sup>, Bradley S. Schneider<sup>26</sup>, Nathan D. Wolfe<sup>26</sup>, Eitel Mpoudi-Ngole<sup>29</sup>, Eric Delaporte<sup>5</sup>, Richard Carter<sup>30,31</sup>, Richard L. Culleton<sup>32</sup>, George M. Shaw<sup>1,4</sup>, Julian C. Rayner<sup>33</sup>, Martine Peeters<sup>5</sup>, Beatrice H. Hahn<sup>1,4</sup> & Paul M. Sharp<sup>3,31</sup>

*Plasmodium vivax* is the leading cause of human malaria in Asia and Latin America but is absent from most of central Africa due to the near fixation of a mutation that inhibits the expression of its receptor, the Duffy antigen, on human erythrocytes. The emergence of this protective allele is not understood because *P. vivax* is believed to have originated in Asia. Here we show, using a non-invasive approach, that wild chimpanzees and gorillas throughout central Africa are endemically infected with parasites that are closely related to human *P. vivax*. Sequence analyses reveal that ape parasites lack host specificity and are much more diverse than human parasites, which form a monophyletic lineage within the ape parasite radiation. These findings indicate that human *P. vivax* is of African origin and likely selected for the Duffy-negative mutation. All extant human *P. vivax* parasites are derived from a single ancestor that escaped out of Africa.

<sup>1</sup> Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup> Department of Biological Sciences, Columbia University, New York, New York 10027, USA. <sup>3</sup> Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. <sup>4</sup> Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>5</sup> Unité Mixte Internationale 233 Institut de Recherche pour le Développement and University of Montpellier 1, 34394 Montpellier, France. <sup>6</sup> Institut National de Recherche Biomedicale, BP 1197 Kinshasa, Democratic Republic of the Congo. <sup>7</sup> Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden. <sup>8</sup> Department of Ecology and Management of Plant and Animal Resources, Faculty of Sciences, University of Kisangani, BP 2012 Kisangani, Democratic Republic of the Congo. <sup>9</sup> Sanaga-Yong Chimpanzee Rescue Center, IDA-Africa, Portland, Oregon 97204, USA. <sup>10</sup> Department of Anthropology, Washington University, St Louis, Missouri 63130, USA. <sup>11</sup> Wildlife Conservation Society, Congo Program, BP 14537 Brazzaville, Republic of the Congo. <sup>12</sup> Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo, Chicago, Illinois 60614, USA. <sup>13</sup> Department of Biological Sciences, University at Albany, State University of New York, Albany, New York 12222, USA. <sup>14</sup> Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. <sup>15</sup> Division of Biological Anthropology, University of Cambridge, Cambridge CB2 1QH, UK. <sup>16</sup> Department of Human Evolutionary Biology, Harvard University, Boston, Massachusetts 02115, USA. <sup>17</sup> Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131, USA. <sup>18</sup> Department of Anthropology, University of California, San Diego, California 92093, USA. <sup>19</sup> Department of Anthropology, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>20</sup> Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA. <sup>21</sup> Department of Entomology, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>22</sup> Infectious Disease Unit, Karolinska Institute, Department of Medicine Solna, Karolinska University Hospital, Stockholm SE-17176, Sweden. <sup>23</sup> Public Health England Malaria Reference Laboratory, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. <sup>24</sup> Lukuru Wildlife Research Foundation, Tshuapa-Lomami-Lualaba Project, BP 2012 Kinshasa, Democratic Republic of the Congo. <sup>25</sup> Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge, CB2 1QH, UK. <sup>26</sup> Global Viral Forecasting Initiative, San Francisco, California 94104, USA. <sup>27</sup> Ape Action Africa, BP 20072 Yaounde, Cameroon. <sup>28</sup> Limbe Wildlife Centre, PO Box 878 Limbe, Cameroon. <sup>29</sup> Institut de Recherches Médicales et d'Études des Plantes Médicinales Prévention du Sida au Cameroun, Centre de Recherche Médicale, BP 906 Yaoundé, Republic of Cameroon. <sup>30</sup> Institute of Immunology and Infection Research, University of Edinburgh, Edinburgh EH9 3JT, UK. <sup>31</sup> Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3JT, UK. <sup>32</sup> Malaria Unit, Institute of Tropical Medicine, Nagasaki University, Nagasaki 852-8523, Japan. <sup>33</sup> Malaria Programme, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. † Present address: Institute for Mind and Biology, The University of Chicago, Chicago, Illinois 60637, USA. Correspondence and requests for materials should be addressed to P.M.S. (email: paul.sharp@ed.ac.uk).

Of the five *Plasmodium* species known to cause malaria in humans, *P. vivax* is the most widespread<sup>1</sup>. Although highly prevalent in Asia and Latin America, *P. vivax* is thought to be absent from west and central Africa due to the near fixation of a mutation that causes the Duffy-negative phenotype in indigenous African people<sup>1,2</sup>. The Duffy antigen receptor for chemokines (DARC) is used by *P. vivax* merozoites to invade red blood cells<sup>3</sup>. Since the absence of DARC on the surface of erythrocytes confers protection against *P. vivax* malaria, this parasite has long been suspected to be the agent that selected for this mutation<sup>2,4,5</sup>. However, this hypothesis has been difficult to reconcile with the proposed evolutionary origin of *P. vivax*<sup>6–8</sup>. The closest relative of *P. vivax* is believed to be *P. cynomolgi*<sup>9</sup>, which infects macaques. These two parasites form a lineage within a clade comprised of at least seven other *Plasmodium* species, all of which infect primates found only in Asia. The consensus view has thus been that *P. vivax* emerged in Southeast Asia following the cross-species transmission of a macaque parasite<sup>6–9</sup>. Under this scenario, the Duffy-negative mutation prevalent in west central African people was selected by another unidentified pathogen<sup>4</sup>, and its high frequency prevented *P. vivax* from entering central Africa.

Recently, *P. vivax*-like parasites have been identified in a limited number of African apes<sup>10–13</sup> and some mosquitoes (*Anopheles* species) trapped in their vicinity<sup>13,14</sup>. Molecular characterization of these parasites showed that they are very similar to, but apparently distinct from, human *P. vivax*<sup>13</sup>. These findings raised the possibility that a sylvatic *P. vivax* reservoir exists in wild-living apes or other African primate species. However, since only very few of these sylvatic parasites have been identified, information concerning their geographic distribution, host species association, prevalence and relationship to human *P. vivax* is lacking. In fact, most evidence of this natural *P. vivax* reservoir has come from pets and apes in wildlife rescue centres. Since captive apes can become infected with *Plasmodium* species that do not normally infect them in their natural habitat<sup>10,15</sup>, studies of wild-living populations are essential.

African apes are highly endangered and live in remote forest regions, rendering invasive screening for infectious agents both impractical and unethical. As an alternative, we have developed methods that permit the detection and amplification of pathogen-specific nucleic acids from ape faecal DNA<sup>16–19</sup>. This approach enabled us to trace the origins of human immunodeficiency virus type 1 (HIV-1) to chimpanzees (*Pan troglodytes*) in west central Africa<sup>18</sup> and to identify the precursor of human *P. falciparum* in western gorillas (*Gorilla gorilla*)<sup>10</sup>. Here, we used a similar approach to investigate the molecular epidemiology of *P. vivax* infection in wild-living apes. Screening more than 5,000 faecal samples from 78 remote forest sites, we tested ape communities throughout central Africa. Since conventional PCR analysis is error prone and has the potential to confound phylogenetic analyses<sup>10</sup>, parasite sequences were generated using single-genome amplification (SGA), which eliminates *Taq* polymerase-induced recombination and nucleotide substitutions from finished sequences<sup>20,21</sup>.

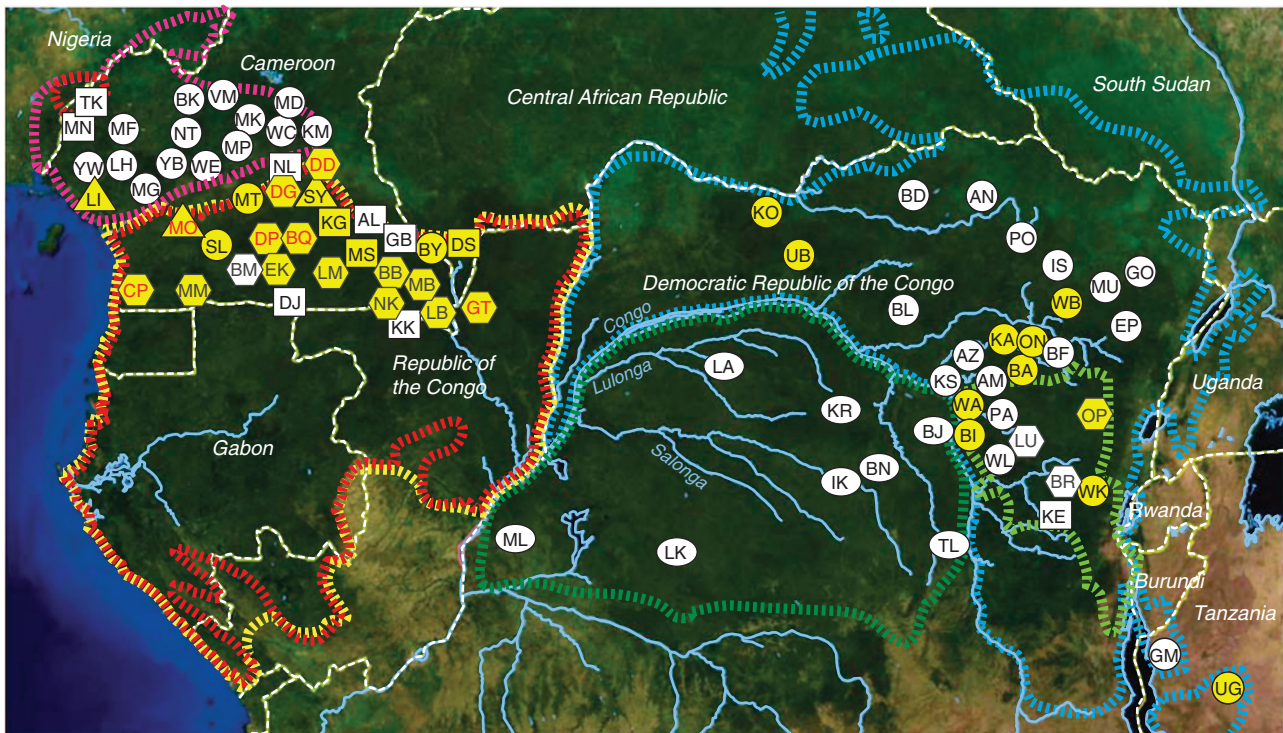
In this study, we show that western (*G. gorilla*) and eastern gorillas (*G. beringei*) and chimpanzees (*P. troglodytes*), but not bonobos (*P. paniscus*), are endemically infected with *P. vivax*-like parasites, and that infection rates in wild ape communities are similar to those in human populations with stable parasite transmission. Analysing over 2,600 SGA-derived mitochondrial, apicoplast and nuclear sequences, we also show that ape parasites are considerably more diverse than human parasites and do not cluster according to their host species. In contrast, human parasites form a monophyletic lineage that falls within the radiation of the ape parasite sequences. These results indicate that

human *P. vivax* arose from within a *Plasmodium* species that infect wild-living chimpanzees and gorillas, and that all extant human *P. vivax* parasites evolved from a single ancestor that spread out of Africa.

## Results

**Host species association and distribution of ape *P. vivax*.** Using PCR primers designed to amplify *P. vivax* mitochondrial (mt) DNA, we screened 5,469 faecal specimens from ape communities sampled at 78 forest sites throughout sub-Saharan Africa (Fig. 1). Except for 196 samples from habituated apes, all other specimens were derived from non-habituated communities (Supplementary Table 1). Ape species and subspecies were identified by faecal mtDNA analysis<sup>10,18,19</sup>. A subset of specimens was also subjected to microsatellite analyses to estimate the number of sampled individuals (Supplementary Table 1). Targeting a 297-base-pair (bp) mtDNA fragment (Supplementary Fig. 1a), we found *P. vivax*-like sequences in faecal DNA from western gorillas (*G. gorilla*), eastern gorillas (*G. beringei*) and chimpanzees (*P. troglodytes*), but not from bonobos (*P. paniscus*) (Table 1). Infections were most common in central chimpanzees (*P. t. troglodytes*) and western lowland gorillas (*G. g. gorilla*), with infected individuals identified at 76% of field sites, including six locations where *P. vivax* was found in both of these species (Fig. 1). Ape *P. vivax* was also endemic in eastern chimpanzees (*P. t. schweinfurthii*) and eastern lowland gorillas (*G. b. graueri*), with infected apes documented at 38% of field sites. Despite this wide geographic distribution (Fig. 1), the proportion of ape faecal samples that contained *P. vivax*-like sequences at any given field site was low: among 2,871 chimpanzee and 1,844 gorilla samples that were analysed, only 45 and 32 were found to be PCR positive, respectively (Table 1). Correcting for specimen degradation and redundant sampling, and taking into account the sensitivity of the non-invasive diagnostic test, we estimated the proportion of *P. vivax*-sequence-positive individuals for each field site (Supplementary Table 1). The resulting values of 4–8% (Table 1) were lower than prevalence rates previously determined for *P. falciparum*-like (*Laverania*) parasites in wild apes<sup>10</sup>, but they were very similar to *P. vivax* parasite rates reported for endemically infected human populations<sup>1</sup>. In humans, point estimates of patent blood infection rarely exceed 7%, even in hyperendemic areas, and a parasite rate of greater than 1% indicates stable transmission<sup>1</sup>.

Since human *P. vivax* can induce dormant liver infections, we considered the possibility that ape parasite DNA might be excreted into faeces in the absence of a productive blood-stage infection and thus inflate our infection rate estimates. To examine this, we compared the sensitivity of PCR-based parasite detection in blood and faecal samples from captive chimpanzees housed at a wildlife rescue centre (SY). Importantly, these chimpanzees were kept in outside enclosures immediately adjacent to the habitat of wild apes and were thus exposed to the same mosquito populations. Although blood and faecal samples were not matched, 11 of 48 chimpanzees (23%) were found to be *P. vivax* positive by blood analysis, as compared with 1 of 68 chimpanzees (1.5%) by faecal analysis (Supplementary Table 2). Thus, faecal *P. vivax* detection is considerably less sensitive than blood detection, most probably because of lower parasite loads, and may underestimate the number of infected apes by an order of magnitude. This likely explains our failure to detect ape *P. vivax* in wild-living Nigeria–Cameroonian chimpanzees (*P. t. ellioti*) and Cross River gorillas (*G. g. diehli*), for which only very few faecal samples were available (Table 1). Indeed, we subsequently confirmed *P. vivax* infection by blood analysis in five Nigeria–Cameroonian chimpanzees that were sampled in captivity



**Figure 1 | Geographic distribution of *P. vivax* in wild-living apes.** Field sites are shown in relation to the ranges of three subspecies of the common chimpanzee (*P. t. ellioti*, magenta; *P. t. troglodytes*, red; and *P. t. schweinfurthii*, blue), western (*G. gorilla*, yellow) and eastern (*G. beringei*, light green) gorillas, as well as bonobos (*P. paniscus*, green). Circles, squares and hexagons identify field sites where wild-living chimpanzees, gorillas or both species were sampled, respectively. Ovals indicate bonobo sampling sites. Triangles denote the location of wildlife rescue centres (see Supplementary Table 1 for a list of all field sites and their two-letter codes). Forested areas are shown in dark green, while arid and semiarid areas are depicted in yellow and brown, respectively. Major lakes and major rivers are shown in blue. Dashed white lines indicate national boundaries. Sites where ape *P. vivax* was detected are highlighted in yellow, with red lettering indicating that both chimpanzees and gorillas were infected.

**Table 1 | Magnitude of the sylvatic *P. vivax* reservoir.**

Species/subspecies	Field sites tested*	Field sites positive†	Faecal samples tested	Faecal samples positive‡	Infection rate (CI)§
Nigeria-Cameroon chimpanzee ( <i>P. t. ellioti</i> )	14	0	126	0	0% (0-5%)
Central chimpanzee ( <i>P. t. troglodytes</i> )	25	11	1,130	25	8% (6-10%)
Eastern chimpanzee ( <i>P. t. schweinfurthii</i> )	28	10	1,615	20	4% (3-7%)
Cross River gorilla ( <i>G. g. diehli</i> )	2	0	80	0	0% (0-8%)
Western lowland gorilla ( <i>G. g. gorilla</i> )	22	14	1,575	30	7% (5-9%)
Eastern lowland gorilla ( <i>G. b. graueri</i> )	4	1	189	2	4% (1-9%)
Bonobo ( <i>P. paniscus</i> )	8	0	754	0	0% (0-1%)

CI, confidence intervals.

\*Field sites are listed in Supplementary Table 1 and their locations are shown in Fig. 1.

†Field sites where sylvatic *P. vivax* was found are highlighted in Fig. 1.

‡Faecal samples were tested for *P. vivax* mitochondrial DNA by diagnostic PCR; all amplicons were sequence confirmed.

§Ape *P. vivax* infection rates were estimated based on the proportion of PCR-positive samples, correcting for specimen degradation, redundant sampling and the sensitivity of the diagnostic test. Brackets indicate 95% CI. Since faecal *P. vivax* detection is less sensitive than blood detection, the values represent minimum estimates.

(Supplementary Table 3), and although not tested in this study, western chimpanzees (*P. t. verus*) have previously been shown to carry this parasite in the wild<sup>12</sup>. Thus, all four subspecies of the common chimpanzee as well as western and eastern gorillas are infected with *P. vivax*, indicating the existence of a substantial sylvatic reservoir.

Given the widespread infection of both chimpanzee and gorilla populations, the fact that over 700 bonobo faecal samples from eight different collection sites were *P. vivax* negative came as a surprise (Fig. 1). Since wild-living bonobos also lack *P. falciparum*-related parasites<sup>10</sup>, yet are susceptible to infection with human *P. falciparum* in captivity<sup>21</sup>, this finding may reflect a

paucity of transmitting mosquito vectors. In humans, a mutation (T to C, at position -33) in the GATA-1 transcription-factor-binding site within the promoter region of the DARC gene<sup>22</sup> yields resistance to *P. vivax* infection, but sequence analysis of the same region in 134 ape samples, including 28 from bonobos, indicated that none had this substitution (Supplementary Fig. 2). In addition, all ape DARC genes analysed encoded a blood-group antigen with aspartic acid at amino acid 42, rather than the glycine found in the protective Fy<sup>a</sup> allele in humans (Supplementary Fig. 2)<sup>23</sup>.

Finally, we asked whether other primates in central Africa might harbour *P. vivax*-like parasites. Using the same

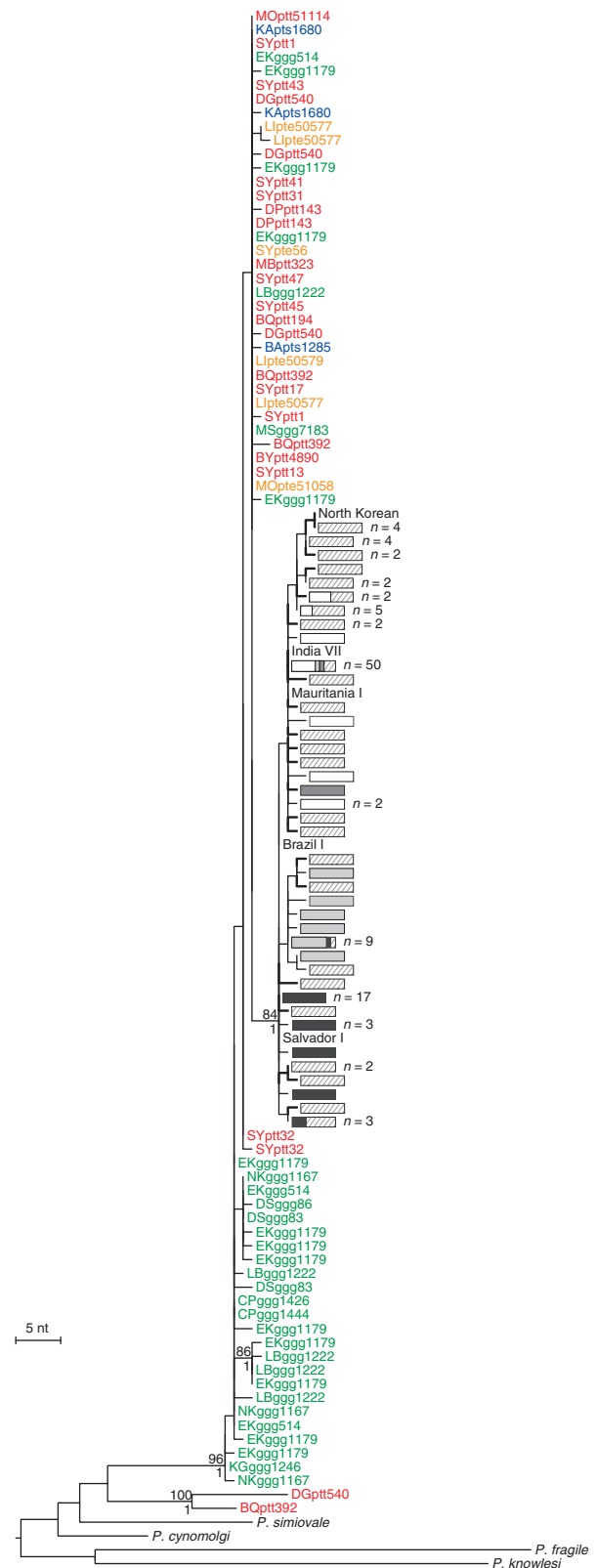


*P. vivax*-specific PCR primers, we screened 998 blood samples from 16 Old World monkey species that had previously been collected for molecular epidemiological studies of simian immunodeficiency viruses<sup>24,25</sup>. Testing samples from 11 different locations in southern Cameroon and the western parts of the Democratic Republic of the Congo (DRC), we failed to detect *P. vivax* infection in any of the animals tested (Supplementary Fig. 3). Although 501 of the 998 blood samples (50.2%) yielded a PCR amplicon, all of these represented *Hepaticystis* spp. infections as determined by sequence analysis (Supplementary Table 4). Thus, we found no evidence for a *P. vivax* reservoir in these African monkey species.

**SGA of *P. vivax* sequences.** To examine the evolutionary relationships of ape and human parasites, we amplified the complete *P. vivax* mitochondrial genome in three partially overlapping fragments (Supplementary Fig. 1a). This was done using SGA followed by direct amplicon sequencing, which eliminates *Taq* polymerase-induced recombination and nucleotide substitution errors, and provides a proportional representation of the parasite sequences that are present *in vivo*<sup>10,20</sup>. Alignment of these sequences revealed two single-nucleotide variants (SNVs) that distinguished all ape from all human parasites (Supplementary Fig. 1b; a third previously proposed SNV<sup>13</sup> was polymorphic among the ape samples in our dataset). We thus designed PCR primers to amplify a fragment (fragment D) that included both SNVs on the same SGA amplicon (Supplementary Fig. 1). Although only a subset of *P. vivax*-positive faecal samples yielded this larger mtDNA fragment, we were able to generate fragment D sequences from 22 chimpanzees and 9 gorillas, 17 of which were sampled in the wild (Supplementary Table 3). Since most database sequences are derived by conventional PCR approaches, we also used SGA to amplify fragment D sequences from the blood of *P. vivax*-infected humans to produce *Taq* polymerase error-free sequences<sup>20</sup>. These samples included 94 international travellers, who had acquired *P. vivax* while visiting malaria endemic areas, as well as 25 *P. vivax*-infected individuals from China, Thailand, Myanmar and India, who sought treatment for clinical malaria (Supplementary Table 5), and thus provide a globally representative sample of human *P. vivax* infections.

**Figure 2 | Evolutionary relationships of ape and human *P. vivax* parasites in mitochondrial gene regions.** The phylogenetic positions of mitochondrial fragment D (2,524 bp; Supplementary Fig. 1a) sequences from ape and human *P. vivax* strains are shown in relation to human and macaque parasite reference sequences. All sequences were generated by SGA<sup>20</sup>, except for human (Salvador I, India VII, Mauritania I, North Korean and Brazil I) and simian reference strains from the database (see Supplementary Tables 6–8 for GenBank accession numbers). Ape 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; ggg: *G. g. gorilla*, green). Human sequences are depicted by haplotype (rectangles) and labelled according to their geographic origin in Oceania (light grey), Africa (white), South and Central America (black), South and Southeast Asia (striped) and the Middle East (dark grey). Haplotypes that include more than one sequence are indicated, with the numbers of sequences listed to the right. A second lineage of related parasite sequences from chimpanzee samples DGpnt540 and BQpnt392 likely represents a new *Plasmodium* species. The tree was inferred using maximum likelihood methods<sup>56</sup>. Numbers above and below nodes indicate bootstrap values ( $\geq 70\%$ ) and Bayesian posterior probabilities ( $\geq 0.95$ ), respectively (the scale bar represents five nucleotide substitutions).

Phylogenetic analysis of all SGA-derived *P. vivax* mtDNA sequences showed that the ape parasites formed two distinct clades (Fig. 2). One clade, represented by sequences from just two chimpanzee samples (termed BQpnt392 and DGpnt540; see legend of Fig. 2 for an explanation of sample nomenclature), was almost



as divergent from the remaining ape and human parasites as were other *Plasmodium* species, and thus likely represents a previously unidentified species. All other ape parasite sequences were closely related to each other and to human *P. vivax* sequences, and thus appear to represent a single species (Fig. 2). Within this *P. vivax* clade, chimpanzee- and gorilla-derived sequences were interspersed, but all human-derived sequences formed a single well-supported lineage that fell within the radiation of the ape parasites. Inclusion of previously published non-SGA sequences confirmed this topology, although many of the database sequences exhibited long branches suggestive of PCR errors (Supplementary Fig. 4a). Interestingly, the one *P. vivax* sequence recently identified in a European traveller who became infected after working in a central African forest<sup>13</sup> did not fall within the human *P. vivax* lineage, but clustered with parasites obtained from wild-living chimpanzees and gorillas (Supplementary Fig. 4a). This confirms the suspicion that this traveller acquired his infection by cross-species transmission from a wild ape.

To examine the robustness of these phylogenetic relationships, we selected additional genomic regions that had previously been used for evolutionary studies of *P. vivax*<sup>7,26</sup>. These included portions of the apicoplast caseinolytic protease C (*clpC*) gene as well as the nuclear genes encoding lactate dehydrogenase (*ldh*), adenylosuccinate lyase (*asl*), cell division cycle 2-related kinase (*crk2*) and  $\beta$ -tubulin ( *$\beta$ -tub*). Although fewer ape samples yielded amplification products, this was most likely due to lower copy numbers of apicoplast and nuclear genomes compared with mtDNA. Nonetheless, many samples yielded more than one *P. vivax* haplotype, indicating coinfection with multiple locally circulating variants (Supplementary Table 3). To increase the number of suitable reference sequences, we also amplified these same fragments from *P. vivax*-positive human samples, and for some gene regions from related macaque parasites. The resulting phylogenies yielded very similar topologies, with human parasites always forming a monophyletic lineage. Moreover, this lineage fell within the radiation of the *P. vivax*-like ape sequences for five of the six loci tested (Fig. 3; Supplementary Figs 5–7). In contrast, chimpanzee and gorilla sequences were again interspersed, suggesting that *P. vivax* is often transmitted between the two ape hosts.

The apicoplast and nuclear gene sequences also confirmed the existence of the second, closely related, ape *Plasmodium* species. Sequences obtained for the *clpC*, *ldh* and *crk2* genes (Fig. 3; Supplementary Fig. 6) from one or the other of the same two chimpanzee samples that yielded divergent mtDNA sequences (DGptt540 and BQptt392) were again clearly distinct from *P. vivax*. In the mtDNA, *ldh* and 5' *crk2* phylogenies, this new species was the closest relative of *P. vivax* (although strong support for this was only found in the latter two trees), while in the other trees relationships were not well resolved. While the new species appears to be rare, it was found in samples from two locations about 110 km apart. Moreover, its detection depended on the cross-reactivity of *P. vivax*-specific PCR primers, indicating that its prevalence and host association remain to be determined. Nonetheless, existing parasite sequences indicate that African apes harbour not only *P. vivax*, but also its closest relative.

**Relative diversity of ape and human *P. vivax*.** Since the phylogenetic analyses indicated that human *P. vivax* strains were derived from within the radiation of ape parasites, we expected the human strains to exhibit lower genetic diversity. To test this directly, we calculated the relative nucleotide diversity of SGA-derived ape and human *P. vivax* sequences. Indeed, values for the average number of nucleotide differences per site ( $\pi$ ) were higher

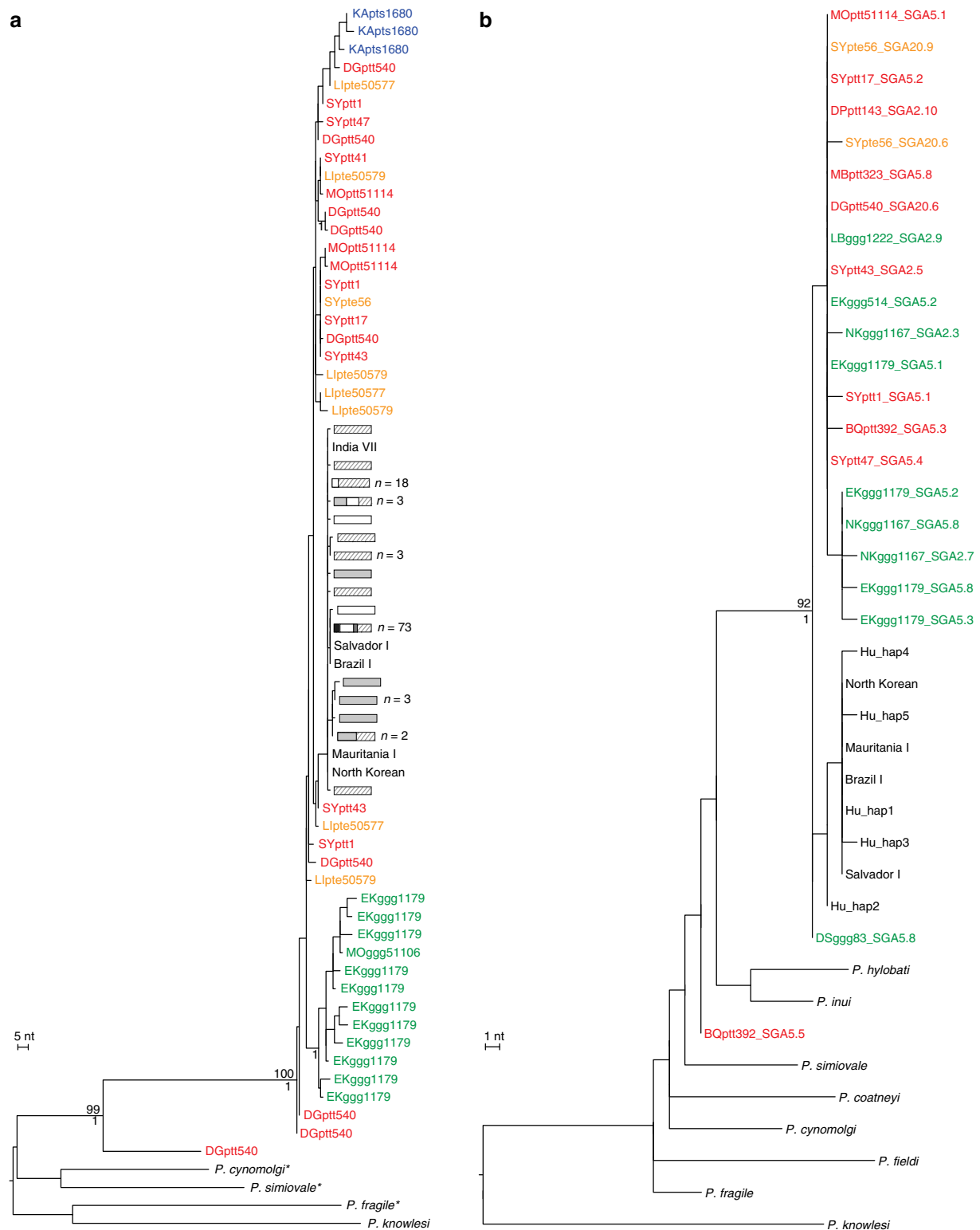
for ape parasites than for human parasites at all loci tested. However, the extent of this increased diversity varied among genes (Table 2). For example, the diversity of mtDNA sequences was only 1.4 times higher among the ape parasites than among the human parasites, while for the apicoplast sequence this value was 6 times higher. For nuclear genes the ape sequences were 9 (*asl*, *ldh*) to 50 (*crk2*) times more diverse (Table 2). If the levels of diversity were in fact similar in the ape and human parasites, it would be most unlikely to observe this difference consistently across this number of loci (for example, Mann–Whitney *U*-test applied to the four nuclear loci,  $P=0.014$ ).

The relative diversity of the various genes also differed between ape and human parasites. For example, for ape parasites, nuclear gene sequences were 12–25 times more diverse than mtDNA, whereas this ratio was only 1–2 for human parasites (Table 2). In the absence of positive selection or demographic changes, relative diversities within species should be similar to those between species. Thus, for comparison, we calculated distances between orthologous sequences from *P. vivax* and the closely related *P. cynomolgi*<sup>9</sup>. For four of the five loci, the relative diversity (scaled to the mtDNA value) among ape *P. vivax* strains was remarkably similar to the relative interspecific divergence (Table 2); the single exception was *crk2*, which was unusually diverse among the ape parasites (compared with other nuclear loci), but also the most conserved nuclear gene between species. In contrast, the diversity values for nuclear and apicoplast genes (relative to mtDNA) among human *P. vivax* strains were strikingly low (Table 2). This reduced diversity among the human *P. vivax* strains most likely reflects a recent bottleneck that, depending on the composition of the founder population, could have affected the relative diversity of organelle and nuclear genomes differently.

## Discussion

Our finding that wild-living apes in central Africa show widespread infection with diverse strains of *P. vivax* and harbour a distinct but related *Plasmodium* species provides new insight into the evolutionary history of human *P. vivax*, and potentially solves the paradox that a mutation conferring resistance to *P. vivax* occurs at high frequency in the very region where this parasite is absent. These results indicate that human *P. vivax* arose from within a *Plasmodium* species that infects chimpanzees and gorillas and indicate an origin in Africa rather than, as previously assumed, in Asia<sup>6–8</sup>. One interpretation of the phylogenies is that a single host switch from apes gave rise to human *P. vivax*, analogous to the origin of human *P. falciparum*<sup>10</sup>. However, this seems unlikely in this case since ape *P. vivax* does not divide into gorilla- and chimpanzee-specific lineages and since humans are susceptible to both natural<sup>13</sup> and experimental<sup>27</sup> ape *P. vivax* infections. Thus, a more plausible interpretation is that an ancestral *P. vivax* stock was able to infect humans, gorillas and chimpanzees in Africa until the Duffy-negative mutation started to spread (perhaps around 30,000 years ago<sup>28</sup>) and eliminated *P. vivax* from humans there. Under this scenario, extant human-infecting *P. vivax* represents a bottlenecked lineage that survived after spreading out of Africa. Much more recently, concomitant with host migrations, human *P. vivax* has been reintroduced to Africa<sup>29</sup>.

Several alternative scenarios for the origins of ape and human *P. vivax* have recently been discussed<sup>13</sup>, but none of these seems plausible in light of the present data. All previous models assumed that *P. vivax* originated in humans in Asia, following the cross-species transmission of a monkey parasite, and that humans then brought the parasite to African apes. This assumption has been based on the fact that the closest known relatives



**Figure 3 | Evolutionary relationships of ape and human *P. vivax* parasites in nuclear and apicoplast gene regions.** The phylogenetic positions of (a) lactate dehydrogenase (*ldh*) gene (711 bp) and (b) caseinolytic protease C (*clpC*) gene (574 bp) sequences from ape and human *P. vivax* strains are shown in relation to human and macaque parasite reference sequences. All sequences were generated by SGA<sup>20</sup>, except for human (Salvador I, India VII, Mauritania I, North Korean and Brazil I) and simian reference strains from the database (asterisks indicate SGA-derived *ldh* sequences for *P. simiovale*, *P. cynomolgi* and *P. fragile*; see Supplementary Tables 6–8 for GenBank accession numbers). Newly derived ape *P. vivax* sequences are labelled and colour coded as in Fig. 2. Human and simian reference sequences are shown in black. Human *ldh* haplotypes are depicted as described in Fig. 2. Related parasite sequences from chimpanzee samples DGptt540 (*ldh*) and BQptt392 (*clpC*) likely represent a new *Plasmodium* species. Trees were inferred using maximum likelihood methods<sup>56</sup>. Numbers above and below nodes indicate bootstrap values ( $\geq 70\%$ ) and Bayesian posterior probabilities ( $\geq 0.95$ ), respectively (the scale bar represents five and one nucleotide (nt) substitutions, respectively).

**Table 2 | Nucleotide diversity in human and ape *P. vivax* lineages.**

<i>P. vivax</i> genome	Locus	Length (bp)	Human <i>P. vivax</i> *			Ape <i>P. vivax</i> *			Ape/human ratio <sup>†</sup>	Distance to <i>P. cynomolgi</i> <sup>‡</sup>	
			No.	$\pi (\times 10^{-3})$ <sup>§</sup>	Ratio <sup>  </sup>	No.	$\pi (\times 10^{-3})$ <sup>§</sup>	Ratio <sup>  </sup>		Distance	Ratio <sup>  </sup>
Mitochondrial	<i>cox1; cyt b</i>	2,443	138	0.75	1.0	62	1.02	1.0	1.4	0.012	1.0
Nuclear	<i>ldh</i>	679	114	1.58	2.1	42	14.07	13.8	8.9	0.161	13.5
Nuclear	<i>asl</i>	838	97	1.45	1.9	21	12.78	12.5	8.8	0.197	16.5
Nuclear	<i>crk2</i>	666	134	0.54	0.7	32	25.92	25.4	49.9	0.081	6.8
Nuclear	<i><math>\beta</math>-tub</i>	684	81	0.81	1.1	12	11.85	11.6	14.6	0.161	13.5
Apicoplast	<i>clpC</i>	574	70	0.34	0.5	21	1.99	2.0	5.9	0.024	2.0

\*All human parasite sequences were derived using SGA methods from a global sampling of *P. vivax* strains (Supplementary Table 5), except for five completely sequenced reference genomes from the database (Salvador I, Brazil I, Mauritania I, North Korea and India VII). Ape *P. vivax* sequences were also derived by SGA from faecal and blood samples of wild-living and sanctuary apes (Supplementary Table 3).

<sup>†</sup>Ratio of nucleotide diversity ( $\pi$ ) values in ape and human parasites.

<sup>‡</sup>Average distance between *P. vivax* and the orthologous sequences from *P. cynomolgi*.

<sup>§</sup>Nucleotide diversity ( $\pi$ ). Human and ape sequences were compared over the same length of sequence.

<sup>||</sup>Diversity (or distance) value expressed relative to that for mtDNA.

of *P. vivax* all seem to infect Asian primates<sup>6–9</sup>. However, we now show that chimpanzees harbour a *Plasmodium* species that is more closely related to *P. vivax* than are any of the Asian primate parasites. Thus, it is more parsimonious to assume that the common ancestor of these two species existed in Africa. How this lineage was introduced into African apes remains unknown; however, this appears to have occurred a long time before the origin of *P. vivax*.

To explain current levels of genetic diversity in ape and human *P. vivax* strains, previous models invoking an Asian origin either require human *P. vivax* in Asia to have gone extinct, prior to repopulation from Africa, or necessitate Asian *P. vivax* to have gone through a bottleneck (of unknown cause)<sup>13</sup>. In contrast, the African origin model does not require such an ancestral (now extinct) human *P. vivax* population in Asia, but explains the reduced diversity of human parasites, as resulting from an out-of-Africa bottleneck as seen in *P. falciparum*<sup>30</sup>, and in humans themselves<sup>31</sup>. It has also been suggested that *P. vivax* is more likely to have spread from Asia to Africa, because human *P. vivax* strains in Asia seem to be the most diverse<sup>13,32</sup> and because phylogeographic analyses indicated elevated migration rates from Asia (especially India) to Africa<sup>6</sup>. However, this is now explained more simply by the extinction of human *P. vivax* in Africa, which would have had high diversity, due to the spread of the Duffy-negative mutation. *P. vivax* strains currently infecting humans in Africa are indeed of Asian origin, but this reflects a reintroduction and occurred only recently, perhaps with the peopling of Madagascar from Asia within the last few thousand years<sup>33</sup>.

If the origin of *P. vivax* had been due to transmission from macaques in southeast Asia, this would imply a convoluted evolutionary history, given the timescales that have been invoked. Estimates of the time of the last common ancestor of human *P. vivax* are generally on the order of hundreds of thousands of years ago. For example, using mtDNA sequences this ancestor was estimated to have existed about 400,000 years ago<sup>32</sup>, and a more recent comparison of nuclear genome sequences suggested a similar date<sup>34</sup>. Modern humans are thought to have evolved in Africa and to first have entered Asia no more than 60,000 years ago<sup>35</sup>. Thus, if the estimates of the timescale of the coalescence of human *P. vivax* lineages are correct, the recipient of the transmission from macaques must have been some earlier hominin species, and *P. vivax* must have diversified for a long time in that host before numerous lineages were transmitted to modern humans after they emerged from Africa.

The existence of a sylvatic *P. vivax* reservoir has public health implications. First, it solves the mystery of *P. vivax* infections in travellers returning from regions where 99% of the human

population is Duffy negative<sup>36,37</sup>. Second, it raises the possibility that humans living in close proximity of chimpanzees and gorillas may become infected by ape *P. vivax*. A recent study of individuals attending a health clinic in the Republic of Congo revealed that 10% carried antibodies specific for preerythrocytic stages of *P. vivax*, suggesting continuing exposure to *P. vivax* sporozoites from an unidentified source<sup>38</sup>. Since ape *P. vivax* is highly prevalent, especially in west central Africa, wild-living chimpanzees and gorillas could serve as an infection reservoir, especially in areas where an influx of Duffy-positive humans through commerce and travel coincides with increasing forest encroachment and ape habitat destruction. Although Duffy-negative individuals are generally protected from blood-stage infections, recent studies in Madagascar<sup>39</sup> and Ethiopia<sup>40</sup> have shown that *P. vivax* is not absolutely dependent on the Duffy receptor. It will thus be important to assess the potential of ape parasites to acquire this phenotype, once the underlying genetic determinants have been identified in human strains. The possibility that ape *P. vivax* may spread via international travel to countries where human *P. vivax* is actively transmitted should also be considered. Since ape *P. vivax* is much more diverse than human *P. vivax* (Figs 2 and 3; Table 2), it is potentially more versatile to escape treatment and prevention measures, especially if human and ape parasites were able to recombine. Given the documented propensity of *P. vivax* for host switching<sup>13,27</sup>, it seems important to screen Duffy-positive and -negative humans in west central Africa, as well as transmitting mosquito vectors, for the presence of ape *P. vivax*. Such studies are now possible through the development of molecular tools that distinguish ape from human *P. vivax*, which also permit the screening of faecal samples for liver-derived parasite DNA in the absence of patent blood infection. These results are necessary to inform malaria control and eradication efforts and to assess future human zoonotic risk.

## Methods

**Ape samples.** A total of 5,469 faecal samples newly (TL site) or previously collected from wild-living chimpanzees (*P. troglodytes*), western gorillas (*G. gorilla*), eastern gorillas (*G. beringei*) and bonobos (*P. paniscus*) for molecular epidemiological studies of simian retroviruses<sup>18,19,41–44</sup> and *Laverania* parasites<sup>10</sup> were selected for *P. vivax* screening. All specimens were derived from non-habituated apes, except for 170 samples from chimpanzees in Gombe National Park (GM), Tanzania, and 26 samples from gorillas in the Dzanga-Sangha Reserve (DS), Central African Republic, who were habituated to the presence of human observers. Samples were collected (1:1 vol/vol) in RNAlater (Life Technologies), transported at ambient temperatures and stored at  $-80^{\circ}\text{C}$ . Faecal DNA was extracted using the QIAamp Stool DNA Mini Kit (Qiagen, Valencia, CA) and used to amplify portions of the host mitochondrial genome to confirm species and subspecies origin<sup>18,19,41–44</sup>. A subset was also subjected to microsatellite analysis at



4–8 polymorphic loci<sup>10,18,41–44</sup> to estimate the number of sampled individuals (Supplementary Table 1). In addition to faecal samples from wild populations, we also obtained stool and blood samples from sanctuary chimpanzees and gorillas that were kept in outside enclosures immediately adjacent to the habitat of wild apes. These included 113 faecal and 66 blood samples from chimpanzees housed at the Sanaga Yong Rescue Centre (SY), 2 faecal samples from gorillas and 14 blood samples from chimpanzees housed at the Limbe Wildlife Centre (LI) and 8 blood samples from 6 chimpanzees and 2 gorillas housed at the Mfou National Park Wildlife Rescue Centre (MO), all located in Cameroon. Faecal samples were collected from known individuals under direct observation. Blood samples were obtained by veni-puncture (dried blood spots, whole blood, buffy coats, red blood cells) and represented left-over material from routine health examinations or were collected for specific veterinary (diagnostic) purposes. Blood and faecal collections were approved by the Ministry of Forestry and Wildlife of Cameroon. Two chimpanzees at the SY sanctuary were suspected to suffer from malaria and sampled during severe febrile illnesses. One had positive blood smears on site and was subsequently identified to be PCR positive for *P. reichenowi*, while the other was PCR positive for ape *P. vivax*. Both responded to malaria treatment. A few additional chimpanzees who were PCR positive for either *Laverania* or ape *P. vivax* exhibited milder symptoms at or near the time of sampling, but the majority of captive apes, including several who were *Laverania* and/or non-*Laverania* sequence positive by blood analysis, were asymptomatic at the time of blood collection. These individuals were also blood smear negative. Samples were shipped in compliance with Convention on International Trade in Endangered Species of Wild Fauna and Flora regulations and country specific import and export permits. DNA was extracted from whole blood and dried blood spots using the QIAamp Blood DNA Mini Kit (Qiagen, Valencia, CA).

**Monkey samples.** To investigate the full host range of *P. vivax*, we screened blood samples from 998 non-human primates sampled in Cameroon and the Democratic Republic of the Congo (Supplementary Table 4). Collection and molecular characterization of these samples have been described<sup>24,25,45</sup>, except for samples from the BO, MO, MA and MS field sites (Supplementary Fig. 3). Briefly, samples were collected from primate bushmeat as dried blood spots or whole blood using strategies specifically designed not to increase demand<sup>24,25,45</sup>. DNA was extracted using the QIAamp blood kit (Qiagen, Courtaboeuf, France) and the species origin was determined by amplifying a 386 bp mtDNA fragment spanning the 12S rRNA gene using single round PCR<sup>24,25</sup>. The non-human primate samples were obtained with approval from the Ministries of Health and Environment and the National Ethics Committees of Cameroon and the Democratic Republic of the Congo.

**Human samples.** To increase the number of human *P. vivax* reference sequences and to generate sequences devoid of PCR error, we obtained a global sampling of human *P. vivax* infections. These included dried blood spots or DNA samples from (i) 61 international travellers diagnosed at the Malaria Reference Laboratory of the London School of Hygiene and Tropical Medicine, London, UK (designated MRL), (ii) 35 international travellers diagnosed at the Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden (designated SW), and (iii) 32 residents of malaria endemic areas in China (designated GX;  $n = 10$ ), Myanmar (VO;  $n = 10$ ), Thailand (PVAR;  $n = 2$ ) and India (IN;  $n = 10$ ), who sought treatment at local health clinics (Supplementary Table 5). In all cases, *P. vivax* infection was initially identified by microscopy and then confirmed by diagnostic PCR followed by direct amplicon sequencing. DNA was extracted from whole blood or dried blood spots using QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Anonymised DNA samples previously collected by the MRL were provided under its remit to undertake epidemiological surveillance relevant to imported malaria in the UK. All other subjects provided written informed consent for the collection and analysis of samples, which were sent without patient identifiers and other patient information, except for the country of (known or presumed) *P. vivax* acquisition. The study was approved by the Institutional Review Boards of the University of Pennsylvania, the Karolinska Institute and the Pennsylvania State University.

**Other Plasmodium species.** Genomic DNA from, or dried blood spots containing, *P. inui* (catalogue number MRA-486F), *P. simiovale* (MRA-488F), *P. cynomolgi* (MRA-350G) and *P. fragile* (MRA-352G) were obtained from the Malaria Research and Reference Reagent Resource Center (MR4) of the American Type Culture Collection (ATCC, Manassas, VA) to generate single-genome-derived reference sequences for phylogenetic analyses.

**Microsatellite analyses.** Ape faecal samples that have previously been subjected to microsatellite analyses have been reported<sup>10</sup>, except for bonobo samples obtained at the TL field site, which were obtained more recently. Briefly, faecal DNA was extracted and used to amplify eight polymorphic microsatellite loci. Amplification products were analysed on an automated sequencer (Applied Biosystems, CA) and sized using GeneMapper 4.0 (Applied Biosystems). For individual identification, samples were first grouped by mtDNA haplotype. Within each haplotype, samples were then grouped by microsatellite genotypes. Since samples were partially degraded due to prolonged storage at ambient temperatures, we allowed allelic mismatches at up to six loci to guard against allelic dropout, but

only if they represented a missing allele. This conservative genotyping approach likely underestimates the number of individuals screened and thus reflects a minimum estimate. Samples with evidence of DNA admixture (multiple peaks for the same locus) were discarded.

**P. vivax diagnostic PCR.** Faecal and blood samples were screened by conventional PCR for *P. vivax* mitochondrial (*cox1*) sequences as described<sup>146</sup>. Although originally designed to amplify *P. vivax* sequences from samples coinfecting with *Laverania* species, the primers were also found to amplify *Hepatozoon* spp. mtDNA (Supplementary Table 4). Primers Pv2768p (5'-GTATGGATCGAATC TTAATTATC-3') and Pv3287n (5'-AATACCAGATATAAAAGACCAACA ATGATA-3') were used in the first round of PCR, while Pv2870p (5'-TTGCA ATGATAAAAACCTTTAGTGC-3') and Pv3185n (5'-TCCTCCAAATTCTGCTGTCT GTAGATAAAATG-3') were used in the second round, yielding a 297 bp amplicon (Supplementary Fig. 1a). Amplicons were gel purified and sequenced directly to confirm *Plasmodium* infection.

**SGA of P. vivax gene sequences.** To derive *P. vivax* sequences devoid of PCR errors, including *Taq* polymerase-induced misincorporations and template switching, blood and faecal samples from *P. vivax*-infected apes and humans were subjected to SGA methods, essentially as described<sup>10,20</sup>. DNA was extracted and end point diluted such that fewer than 30% of PCR reactions yielded an amplification product. According to a Poisson distribution, a well yielding a PCR product at this dilution will contain only a single amplifiable template more than 83% of the time. Multiple different gene regions were amplified, including mitochondrial fragments S, A, B, C and D (Supplementary Fig. 1a), the apicoplast *clpC* gene (574 bp), and the nuclear genes *ldh* (713–724 bp), *asl* (838 bp), *crk2* (666 bp) and *β-tub* (684 bp) (Supplementary Table 3). Primers were designed in conserved regions of the *P. vivax* genome, except for a set of *ldh* primers intended to amplify more diverse macaque parasites, which were designed using both *P. vivax* and *P. knowlesi* consensus sequences (Supplementary Table 9). PCR cycling conditions were as described<sup>10</sup>, except for varying elongation times depending on the length of the amplicon (1 min per 1 kb amplicon length). Amplification products were sequenced directly and analysed using Sequencher (Gene Codes Corporation, Ann Arbor, MI). All sequences with double peaks in the chromatogram were discarded, except for sequences derived from ape faecal samples that contained a single double peak. Single double peaks in SGA sequences either indicate the presence of two *P. vivax* variants differing by a single nucleotide or a PCR misincorporation within the first or second round of PCR. We resolved these ambiguous sites by selecting the nucleotide that was identical to the ape *P. vivax* consensus. This conservative approach avoided the loss of otherwise valuable sequence information. GenBank accession codes of newly derived SGA sequences are listed in Supplementary Tables 6 and 7.

**DARC genotyping.** To examine whether *P. vivax*-infected and -uninfected apes exhibited sequence polymorphisms in their DARC promoter and/or adjacent coding sequences, we amplified a 1,286 bp region of the DARC gene in two partially overlapping PCR fragments from 134 ape (faecal and blood) samples (Supplementary Fig. 2a). Fragment A was amplified using DARCpF1 (5'-GCTG TCCATTGTCCCCTAG-3') and DARCpR8 (5'-GGCCCATACTACCCT GTGC-3') in the first round of PCR, and DARCpF3 (5'-GCACAATGATACA CAGCAAAC-3') and FYPdn (5'-CCATGGCACCGTTGGTTCAGG-3')<sup>47</sup> in the second round. Fragment B was amplified using DARC\_F5 (5'-AGGCAGTGGG CGTGGGGTAAG-3') and DARC\_R5 (5'-AGCCATACAGACACAGTAG CC-3') in the first round, and DARC\_NTF1 (5'-TTGGCTCTATCTTGG ACAC-3') and DARC\_NTR1 (5'-TGGTGAGGATGAAGAAGGGCAGT-3') in the second round. PCR conditions were the same as those used to amplify *P. vivax* mtDNA. Amplicons were directly sequenced and analysed using Sequencher.

**Ape P. vivax infection rates.** For sites where the number of sampled chimpanzees was known (Supplementary Table 1), *P. vivax* infection rates were estimated based on the proportion of positive individuals (but correcting for test sensitivity; see below), with 95% confidence limits determined assuming binomial sampling. For field sites where the number of sampled individuals was not known, infection rates were estimated based on the number of faecal samples, but correcting for specimen degradation, oversampling and the sensitivity of the diagnostic PCR test. The latter was estimated by determining the proportion of PCR-positive specimens from *P. vivax*-infected apes that were sampled more than once on the same day, as previously described<sup>10</sup>. Including data from 14 such apes (6 central chimpanzees, 5 eastern chimpanzees and 3 western gorillas), we estimated the sensitivity to be 32% (17 of 53 samples were positive; confidence limits were determined assuming binomial sampling). Using previously reported values for oversampling<sup>10</sup> (1.77, 1.84, 3.74, 2.06 and 1.84 for central chimpanzees, western gorillas, eastern chimpanzees, eastern gorillas and bonobos, respectively) and sample degradation<sup>10</sup> (0.13), the proportion of *P. vivax*-sequence-positive apes was estimated for each field site (Table 1 and Supplementary Table 1). However, since *P. vivax* detection in faecal samples is considerably less sensitive than in blood, the resulting prevalence rates should be interpreted as minimum estimates.



**Nucleotide diversity calculations.** Newly generated ape and human *P. vivax* sequences were aligned with sequences from the complete genomes of five *P. vivax* reference strains (Salvador I, India VII, Mauritania I, North Korean and Brazil I). Other than these five reference strains, only SGA-derived sequences were included in the analysis, so as to not inflate diversity values by including *Taq* polymerase-induced errors. Identical sequences from the same sample or individual were excluded; identical sequences from different samples or individuals were retained. For each gene, sequences were aligned using ClustalW<sup>48</sup>, and both ape and human *P. vivax* alignments were trimmed to the same length. The nucleotide diversity ( $\pi$ ) of human and ape parasite sequences was calculated for six *P. vivax* gene fragments (mtDNA fragment D, apicoplast *clpC*, nuclear *ldh*, *asl*, *crk2* and  $\beta$ -*tub*) using DnaSP version 5.10 (ref. 49).

**Distance calculation from *P. cynomolgi*.** For each *P. vivax* gene, the same set of sequences as used for the diversity calculation was aligned with one *P. cynomolgi* reference sequence<sup>9</sup>. The distance from each *P. vivax* sequence to the *P. cynomolgi* reference was calculated using the R package version 3.0–8 with the Tamura–Nei correction<sup>50,51</sup>, and the mean distance was calculated for each gene.

**Phylogenetic analyses.** Sequence alignments were constructed using ClustalW version 2.1 (ref. 48) and manually adjusted using MacClade<sup>52</sup>. Regions that could not be unambiguously aligned were omitted from subsequent phylogenetic analyses. Nuclear gene sequences were subjected to recombination analysis using GARD<sup>53</sup>. Evolutionary models for phylogenetic analyses were determined using the Akaike information criterion with Modeltest (version 3.7)<sup>54</sup> and PAUP\* (ref. 55). Maximum likelihood phylogenies with bootstrap support (100 replicates) were estimated jointly with model parameter values by means of PhyML (version 3)<sup>56</sup> using both nearest-neighbour interchange and subtree pruning and regrafting with Neighbor Joining and 10 random-addition starting trees<sup>57</sup>. Posterior probabilities for nodes in phylogenetic trees were calculated using MrBayes (version 3.2)<sup>58</sup>, using an average standard deviation of partition frequencies <0.01 as a convergence diagnostic. Trees were constructed from mitochondrial fragment D sequences (2,524 bp, Fig. 2 and Supplementary Fig. 4a), apicoplast *clpC* sequences (574 bp, Fig. 3b), and nuclear  $\beta$ -*tub* (664 bp, Supplementary Fig. 5), *ldh* (711 bp, Fig. 3a and Supplementary Fig. 4b), *crk2* (271 and 372 bp, Supplementary Fig. 6) and *asl* (838 bp, Supplementary Fig. 7) sequences.

## References

1. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6**, e1814 (2012).
2. Mendis, K., Sina, B. J., Marchesini, P. & Carter, R. The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg.* **64**, 97–106 (2001).
3. Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *New Engl. J. Med.* **295**, 302–304 (1976).
4. Livingstone, F. B. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. *Hum. Biol.* **56**, 413–425 (1984).
5. Carter, R. & Mendis, K. N. Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.* **15**, 564–594 (2002).
6. Mu, J. *et al.* Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.* **22**, 1686–1693 (2005).
7. Escalante, A. A. *et al.* A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl Acad. Sci. USA* **102**, 1980–1985 (2005).
8. Carlton, J. M., Das, A. & Escalante, A. A. Genomics, population genetics and evolutionary history of *Plasmodium vivax*. *Adv. Parasitol.* **81**, 203–222 (2013).
9. Tachibana, S. *et al.* *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* **44**, 1051–1055 (2012).
10. Liu, W. *et al.* Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **467**, 420–425 (2010).
11. Krief, S. *et al.* On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from bonobos. *PLoS Pathog.* **6**, e1000765 (2010).
12. Kaiser, M. *et al.* Wild chimpanzees infected with five *Plasmodium* species. *Emerg. Infect. Dis.* **16**, 1956–1959 (2010).
13. Prugnolle, F. *et al.* Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. *Proc. Natl Acad. Sci. USA* **110**, 8123–8128 (2013).
14. Paupy, C. *et al.* *Anopheles moucheti* and *Anopheles vinckei* are candidate vectors of ape *Plasmodium* parasites, including *Plasmodium praefalciparum* in Gabon. *PLoS One* **8**, e57294 (2013).
15. 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).
16. Santiago, M. L. *et al.* SIVcpz in wild chimpanzees. *Science* **295**, 465 (2002).
17. Santiago, M. L. *et al.* Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J. Virol.* **77**, 7545–7562 (2003).
18. Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
19. Li, Y. *et al.* Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *J. Virol.* **86**, 10776–10791 (2012).
20. Liu, W. *et al.* Single genome amplification and direct amplicon sequencing of *Plasmodium* spp. DNA from ape fecal samples. *Protoc. Exch.* doi:10.1038/nprot.2010.156 (2010).
21. Rayner, J. C., Liu, W., Peeters, M., Sharp, P. M. & Hahn, B. H. A plethora of *Plasmodium* species in wild apes: a source of human infection? *Trends Parasitol.* **27**, 222–229 (2011).
22. Tournamille, C., Colin, Y., Cartron, J. P. & Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy negative individuals. *Nat. Genet.* **10**, 224–228 (1995).
23. King, C. L. *et al.* Fy<sup>A</sup>/Fy<sup>B</sup> antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to *Plasmodium vivax* malaria. *Proc. Natl Acad. Sci. USA* **108**, 20113–20118 (2011).
24. Aghokeng, A. F. *et al.* Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. *Infect. Genet. Evol.* **10**, 386–396 (2010).
25. Ahuka-Mundeye, S. *et al.* Novel multiplexed HIV/simian immunodeficiency virus antibody detection assay. *Emerg. Infect. Dis.* **17**, 2277–2286 (2011).
26. Kedzierski, L. *et al.* Phylogenetic analysis of the genus *Plasmodium* based on the gene encoding adenylosuccinate lyase. *Infect. Genet. Evol.* **1**, 297–301 (2002).
27. Contacos, P. G. *et al.* Transmission of *Plasmodium schwetzi* from the chimpanzee to man by mosquito bite. *Am. J. Trop. Med. Hyg.* **19**, 190–195 (1970).
28. Hamblin, M. T. & Di-Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66**, 1669–1679 (2000).
29. Culleton, R. & Carter, R. African *Plasmodium vivax*: distribution and origins. *Int. J. Parasitol.* **42**, 1091–1097 (2012).
30. Tanabe, K. *et al.* *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr. Biol.* **20**, 1283–1289 (2010).
31. Ramachandran, S. *et al.* Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947 (2005).
32. Cornejo, O. E. & Escalante, A. A. The origin and age of *Plasmodium vivax*. *Trends Parasitol.* **22**, 558–563 (2006).
33. Tofanelli, S. *et al.* On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol. Biol. Evol.* **26**, 2109–2124 (2009).
34. Neafsey, D. E. *et al.* The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* **44**, 1046–1050 (2012).
35. Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796–800 (2006).
36. Carter, R. Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol.* **13**, 214–219 (2003).
37. Guerra, C. A. *et al.* The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl. Trop. Dis.* **4**, e774 (2010).
38. Culleton, R. *et al.* Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *J. Infect. Dis.* **200**, 1465–1469 (2009).
39. Ménard, D. *et al.* *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl Acad. Sci. USA* **107**, 5967–5971 (2010).
40. Woldearegai, T. G. *et al.* *Plasmodium vivax* malaria in Duffy-negative individuals from Ethiopia. *Trans. R Soc. Trop. Med. Hyg.* **107**, 328–331 (2013).
41. Keele, B. F. *et al.* Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* **460**, 515–519 (2009).
42. Rudicell, R. S. *et al.* High prevalence of simian immunodeficiency virus infection in a community of savanna chimpanzees. *J. Virol.* **85**, 9918–9928 (2010).
43. Neel, C. *et al.* Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *J. Virol.* **84**, 1464–1476 (2009).
44. Etienne, L. *et al.* Non-invasive follow-up of simian immunodeficiency virus infection in wild-living nonhabituated western lowland gorillas in Cameroon. *J. Virol.* **86**, 9760–9772 (2012).
45. Aghokeng, A. F. *et al.* Widely varying SIV prevalence rates in naturally infected primate species from Cameroon. *Virology* **345**, 174–189 (2006).
46. Sundararaman, S. A. *et al.* *Plasmodium falciparum*-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. *Proc. Natl Acad. Sci. USA* **110**, 7020–7025 (2013).
47. Zimmerman, P. A. *et al.* Emergence of Fy<sup>A</sup> null in a *Plasmodium vivax*-endemic region of Papua New Guinea. *Proc. Natl Acad. Sci. USA* **96**, 13973–13977 (1999).

48. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
49. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
50. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
51. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
52. Maddison, W. P. & Maddison, D. R. *MacClade—Analysis of Phylogeny and Character Evolution—Version 4* (Sinauer Associates, Inc., 2001).
53. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901 (2006).
54. Posada, D. & Buckley, T. 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).
55. PAUP\* 4.0: *Phylogenetic Analysis Using Parsimony (\*and Other Methods) v. 4.0b2a* (Sinauer Associates, Inc., Sunderland, MA, 1999).
56. Guindon, S., Delsuc, F., Dufayard, J. F. & Gascuel, O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **537**, 113–137 (2009).
57. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
58. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

## Acknowledgements

We thank the staff of project PRESICA, the World Wildlife Fund for Nature (WWF/DRC), the Institut National de Recherches Biomédicales (INRB, Kinshasa, DRC), Global Viral Cameroon, the Bonobo Conservation Initiative and Vie Sauvage, as well as Didier Mazongo, Octavie Lunguya, Muriel Aloni and Valentin Mbenz for fieldwork in Cameroon and the DRC; field assistants in Gombe National Park and Ugalla for sample collection in Tanzania, the staff of the Sanaga Yong, Limbe and Ape Action Africa/Mfou National Park Wildlife Rescue Centres for collecting blood samples from captive apes; the Cameroonian Ministries of Health, Forestry and Wildlife, and Scientific Research and Innovation for permission to collect samples in Cameroon; the Water and Forest Ministry for permission to collect samples in the Central African Republic; the Ministry of Forest Economy and Sustainable Development for permission to collect samples in the Republic of Congo; the Ministry of Scientific Research and Technology, the Department of Ecology and Management of Plant and Animal Resources of the University of Kisangani, the Ministries of Health and Environment and the National Ethics Committee for permission to collect samples in the DRC; the Tanzania Commission for Science and

Technology and the Tanzania Wildlife Research Institute for permission to conduct research in Gombe National Park and Ugalla. This work was supported by grants from the National Institutes of Health (R01 AI091595, R37 AI050529, R01 AI58715, T32 AI007532, P30 AI045008), the Agence Nationale de Recherche sur le Sida (ANRS 12125/12182/12255), the Agence Nationale de Recherche (Programme Blanc, Sciences de la Vie, de la Santé et des Ecosystèmes and ANR 11 BSV3 021 01, Projet PRIMAL), Harvard University, the Arthur L. Greene Fund, the Jane Goodall Institute, the Wellcome Trust (098051), the Leakey Foundation, Google.org and the Skoll Foundation. This study was also made possible by the generous support of the American people through the United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

## Author contributions

All authors contributed to the acquisition, analysis and interpretation of the data; W.L., R.C., R.L.C., G.M.S., J.C.R., M.P., B.H.H. and P.M.S. initiated and conceived the study; W.L., Y.L., K.S.S., J.A.M., A.G. and C.F.D. performed non-invasive *P. vivax* testing and SGA analyses; J.A.M., M.A.R., P.A.C., A.G.S. and F.B.-R. performed genotyping; S.A.S. calculated ape *P. vivax* infection rates; G.H.L., L.J.P. and P.M.S. performed phylogenetic analyses; A.A., S.L., A.E., F.M., E.G., C.B. and S.A.-M. performed non-invasive *P. vivax* testing on apes and monkeys; S.L., S.A.-M., B.-I.I., J.-B.N.N., S.S., C.M.S., D.B.M., M.K.G., P.J.K., P.D.W., A.V.G., M.N.M., A.K.P., F.A.S., M.L.W., A.E.P., J.A.H., T.B.H., P.B., M.L., B.T., J.K., B.S.S., N.D.W., E.M.-N. and E.D. conducted or supervised fieldwork; L.C., Z.W., A.F., C.J.S. and D.N. characterized human *P. vivax* infections; and W.L., R.C., R.L.C., G.M.S., J.C.R., M.P., B.H.H. and P.M.S. coordinated the contributions of all authors and wrote the paper.

## Additional information

**Accession codes:** All newly derived ape and human *Plasmodium* sequences have been deposited in the GenBank Nucleotide Sequence Database under accession codes KF591752—KF591851 and KF618374—KF618618, respectively. DARC sequences have been deposited in the GenBank Nucleotide Sequence Database under accession codes KF618448—KF618495.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npublishing.nature.com/reprintsandpermissions/>

**How to cite this article:** Liu, W. *et al.* African origin of the malaria parasite *Plasmodium vivax*. *Nat. Commun.* **5**:3346 doi: 10.1038/ncomms4346 (2014).