The human parasite *Plasmodium malariae* has relatives infecting African apes (*Plasmodium rodhaini*) and New World monkeys (*Plasmodium brasilianum*), but its origins remain unknown. Using a novel approach to characterise *P. malariae*-related sequences in wild and captive African apes, we found that this group comprises three distinct lineages, one of which represents a previously unknown, highly divergent species infecting chimpanzees, bonobos and gorillas across central Africa. A second ape-derived lineage is much more closely related to the third, human-infective lineage *P. malariae*, but exhibits little evidence of genetic exchange with it, and so likely represents a separate species. Moreover, the levels and nature of genetic polymorphisms in *P. malariae* indicate that it resulted from the zoonotic transmission of an African ape parasite, reminiscent of the origin of *P. falciparum*. In contrast, *P. brasilianum* falls within the radiation of human *P. malariae*, and thus reflects a recent anthroponosis.
Plasmodium malariae is one of six Plasmodium species that commonly cause malaria in humans. Although geographically widespread across tropical and subtropical malaria-endemic regions, P. malariae is the least well characterised human parasite, in part because it reaches only very low levels of parasitaemia, is generally associated with mild or no disease, and frequently coexists with other malaria parasites in multi-species infections. Unlike other human parasites, P. malariae infects only mature red blood cells, with an intraerythrocytic cycle of 72 h, which causes the characteristic quartan fever. However, the parasite can also persist chronically and recrudesce years or decades after the initial infection. Because of a lack of systematic studies due to the perceived low disease burden, the prevalence of P. malariae is likely underestimated.

While the origins of the major human malaria parasites P. falciparum and P. vivax have recently been traced to parasites infecting African apes, the evolutionary history of P. malariae remains unknown, despite the long-known existence of close relatives in non-human primates. A quartan parasite morphologically indistinguishable from P. malariae was observed in the early 20th century infecting wild chimpanzees, and given the name P. rodhaini. However, in subsequent transmission studies P. rodhaini was successfully transferred from chimpanzees to humans by blood inoculation, while P. malariae was transmitted from humans to chimpanzees by blood and infective mosquitoes. This apparent lack of host specificity suggested that ape and human parasites represented members of the same species, but this could not be verified because strains of P. rodhaini were not maintained and thus were never genetically characterised.

More recently, genetic studies have been applied to ape-infecting relatives of P. malariae. First, analyses of mitochondrial DNA (mtDNA) sequences derived from faecal samples from wild-living chimpanzees, gorillas and bonobos indicated the existence of two distinct P. malariae-related lineages infecting apes: one that is closely related to the human parasite and another that is substantially more divergent. Echoing this, mtDNA sequences obtained from samples of mosquitoes (Anopheles vinckeii) that feed on apes also seemed to fall into two divergent lineages. Recently, genome sequences, one near-complete (PmlGA01) and one partial (PmlGA02), were recovered from the blood of two chimpanzees living at a wildlife reserve in Gabon. Although genome-wide analyses revealed a close relationship to P. malariae, these “P. malariae-like” parasites were considered a separate species, that might have diverged from P. malariae more than 3 million years ago. However, PmlGA01 and PmlGA02 were as divergent from one another as they were from P. malariae, and the study provided no information as to the relationship of these two parasites to the two divergent lineages of ape parasites revealed by mtDNA.

Finally, another P. malariae-related lineage, P. brasiliannum, infects New World monkeys throughout Central and South America. Molecular characterisation of P. brasiliannum, albeit at only a limited number of genetic loci, found it to be near-identical to P. malariae, again suggesting that the two lineages may in fact be a single species. In addition, P. brasiliannum was transmissible from New World monkeys to humans by mosquitoes and transferred back to monkeys via infected blood, and New World monkeys could be experimentally infected with P. malariae from a human. These findings suggested that the New World monkey parasite arose as an anthropoponosis from P. malariae-infected humans. However, comparisons of genetic diversities led some authors to conclude that the opposite scenario was also possible, i.e., that P. malariae originated from P. brasiliannum.

To resolve the number of P. malariae-related species and their host ranges, we used limiting dilution PCR to generate new sequences from faecal and blood samples of wild-living and captive chimpanzees and gorillas. We also developed a new approach to identify P. malariae-related sequences in published genome databases generated from apes with mixed Plasmodium infections. Here, we show that the clade of P. malariae-related parasites comprises three distinct lineages, one of which represents a previously unknown, distantly related species that infects wild-living chimpanzees, bonobos and gorillas across central and west Africa. We also show that “P. malariae-like” parasites represent a second ape-infective species, which is much more closely related to the human-infective lineage, but lacks evidence of genetic exchange with the human parasite. Moreover, comparison of genetic polymorphisms between the two lineages indicates that the human parasite has undergone a severe genetic bottleneck, followed by rapid population expansion. Finally, we show that P. brasiliannum does not represent a distinct lineage but falls within the radiation of human P. malariae strains. These results thus place the origin of the P. malariae-related clade in Africa and indicate that the human parasite arose following a host switch of an ape parasite, while P. brasiliannum arose after the transmission of the human parasite to New World monkeys.

Results

Three distinct lineages within the P. malariae-related group. Although wild apes have long been known to harbour P. malariae-related parasites (defined as a group of parasites that have much greater sequence similarity to P. malariae than to any other Plasmodium species, and which includes P. brasiliannum and the human parasite itself, see Supplementary Note 1), only a few infections have been characterised. Previous field studies of wild apes and their mosquitoes demonstrated the presence of P. malariae-related parasites, but recovered a very limited number of sequences, derived almost entirely from a single mitochondrial gene (cytB). To increase the number of parasite sequences for phylogenetic studies, we thus selected ape blood and faecal samples previously shown by diagnostic PCR to be P. malariae-positive, and subjected these to single genome amplification (SGA) targeting different genomic regions. SGA amplifies single templates and thus avoids generating parasite sequences with Taq polymerase artefacts, including chimeras produced by recombination, which is important for samples with multi-species infections. Using this approach we were able to amplify additional mtDNA and nuclear gene (ldh and asl) sequences. While most of these were derived from blood samples of sanctuary chimpanzees, P. malariae-related sequences were also recovered from one chimpanzee and two gorillas sampled in the wild (Supplementary Tables 1 and 2). We also derived newly assembled mtDNA genomes for the two “P. malariae-like” chimpanzee parasites GA01 and GA02.

Phylogenetic analyses of both newly derived (N = 21) and previously published (N = 44) parasite sequences revealed that P. malariae and primate parasites related to P. malariae fall into two highly divergent lineages in trees derived from both mtDNA (Fig. 1a, Supplementary Fig. 1) and nuclear (Fig. 1c, d) gene sequences. One of these (initially termed M1) included all P. malariae parasites from humans, together with P. brasiliannum and some samples from apes. A second lineage (termed M2, see Supplementary Note 1) was quite distantly related to P. malariae, although clearly closer to P. malariae than to any other previously described Plasmodium species (see below). The M2 lineage included sequences obtained from samples from gorillas, chimpanzees and bonobos, as well as all 10 P. malariae-related sequences previously obtained from mosquitoes that feed on apes (Fig. 1a).

Closer examination of M1 sequences indicated that they could be divided into two lineages. In a tree derived from our longest
We attempted selective whole genome amplification (SWGA) of restricted regions of whole blood DNA, but failed to obtain a sufficient number of whole blood DNA samples. As an alternative approach, we asked whether we could discern one clade comprising all human P. malariae sequences, P. brasilianum and a few sequences from captive apes, within which sequences differed by no more than 10 nucleotides from the consensus P. malariae sequence; we retained the term M1 for this clade (see Supplementary Note 1). The other non-M2 sequences showed greater diversity, and included samples derived from both wild and captive apes, including the two chimpanzee-derived strains (“P. malariae-like” GA01 and GA02) for which genome sequences had been derived. Here, we term these M1-like, to avoid confusion with the term “P. malariae-like” in the literature (see Supplementary Note 1). While M1 and M1-like strains were not separated in trees derived from partial nuclear gene sequences (Fig. 1c, d), subsequent comparisons of nuclear genome sequences confirmed that these two lineages are distinct.

**M2 represents a new P. malariae-related species that infects apes across Africa.** To further characterise the M2 lineage, we first examined one blood sample from a sanctuary chimpanzee (SYptt92, sequences 2 and 55 in Fig. 1) that had yielded partial mitochondrial and nuclear sequences that clustered within M2. To improve readability, one contig with a nucleotide identity of 70% has been omitted from the plot. Contigs between 83% and 92% nucleotide identity to PmGA01 (dashed red lines) were assumed to originate from M2 and used for further assembly; contigs between 92% and 97% identity or below 83% identity were evaluated manually and included as M2 contigs if appropriate. Boxplot of nucleotide divergences between M1 and M1-like, and between M1 and M2; n = 285 genes, totalling 350 kb of aligned sequence. A vertical centre line indicates the median; the box indicates the interquartile range; whiskers indicate 1.5x the interquartile range; black circles represent outliers.

![Fig. 1 Phylogenetic relationships among M1 strains (including P. malariae) and the two ape-infective lineages M1-like and M2. a]({#}) Maximum likelihood tree derived from a 641 bp mitochondrial DNA cytB fragment; this tree includes all available M2 cytB sequences, but for clarity includes only a few, representative M1 and M1-like sequences. b]({#}) Maximum likelihood tree derived from a 2,038 bp mitochondrial cytB/cst/1 fragment; this tree includes all available sequences of sufficient length from these lineages, and was generated from a longer alignment that encompasses the region in part a (see Supplementary Fig. 1). c]({#}) Maximum likelihood tree derived from a 682 bp ad/g fragment. d]({#}) Maximum likelihood tree derived from a 724 bp idh fragment. Boxes represent sequences and are coloured according to the host from which the sample was derived, as indicated in the key. A black bar at the left of a box indicates a sample from a captive or sanctuary ape; asterisks denote newly derived sequences; numbers in the boxes give the sequence code (Supplementary Table 2). Bootstrap support from 1,000 replicates was 100% for the nodes separating M2 from M1 and M1-like, and for the node separating M1 from M1-like in part b. Scale bars indicate the number of nucleotide changes.

![Fig. 2 P. malariae-related contigs from chimpanzee sample PGABG03. a]({#}) Identity to M1-like genome assembly PmGA01 of de novo contigs assembled from PGABG03 sequencing reads mapped to P. malariae. To improve readability, one contig with a nucleotide identity of 70% has been omitted from the plot. Contigs between 83% and 92% nucleotide identity to PmGA01 (dashed red lines) were assumed to originate from M2 and used for further assembly; contigs between 92% and 97% identity or below 83% identity were evaluated manually and included as M2 contigs if appropriate. b]({#}) Boxplot of nucleotide divergences between M1 and M1-like, and between M1 and M2; n = 285 genes, totalling 350 kb of aligned sequence. A vertical centre line indicates the median; the box indicates the interquartile range; whiskers indicate 1.5x the interquartile range; black circles represent outliers.

The available mtDNA alignment (~2 kb) for which we had numerous samples (Fig. 1b, Supplementary Fig. 1, Supplementary Tables 2 and 3), we could discern one clade comprising all human P. malariae sequences, P. brasilianum and a few sequences from captive apes, within which sequences differed by no more than two nucleotides from the consensus P. malariae sequence; we retained the term M1 for this clade (see Supplementary Note 1). The other non-M2 sequences showed greater diversity, and included samples derived from both wild and captive apes, including the two chimpanzee-derived strains (“P. malariae-like” GA01 and GA02) for which genome sequences had been derived, here, we term these M1-like, to avoid confusion with the term “P. malariae-like” in the literature (see Supplementary Note 1). While M1 and M1-like strains were not separated in trees derived from partial nuclear gene sequences (Fig. 1c, d), subsequent comparisons of nuclear genome sequences confirmed that these two lineages are distinct.

**M2 represents a new P. malariae-related species that infects apes across Africa.** To further characterise the M2 lineage, we first examined one blood sample from a sanctuary chimpanzee (SYptt92, sequences 2 and 55 in Fig. 1) that had yielded partial mitochondrial and nuclear sequences that clustered within M2. We attempted selective whole genome amplification (SWGA) of whole blood DNA, but failed to obtain a sufficient number of sequences. As an alternative approach, we asked whether sequencing projects targeting any other ape Plasmodium species may have amplified sequences from M2 parasites as a by-product; we reasoned this was possible because M1 and M2 have previously been observed as coinfections with other Plasmodium species. We thus screened parasite sequence databases from published Laverania (13 samples) and P. vivax-like (seven samples) infections of chimpanzees and gorillas for the presence of P. malariae-related sequences. We mapped all reads to a combined Plasmodium reference sequence including the P. malariae reference genome PmUG01, followed by de novo contig assembly of read pairs that mapped to P. malariae and
alignment of contigs with the M1-like genome assembly PmlGA01\textsuperscript{24} ("P. malariae-like"). This approach identified one sample (PGABG03\textsuperscript{13}) from a chimpanzee in a Gabonese wildlife reserve that yielded contigs with a bimodal distribution of nucleotide identities to the M1-like reference (Fig. 2a). One peak comprised contigs exhibiting more than 97\% nucleotide identity to PmlGA01, while a second, broader peak was centred around 88\% identity, suggesting coinfection with both M1-like and M2 parasites. We aligned high-quality contigs from this second peak (see Methods for details of how these were selected) with the P. malariae reference genome and annotated genes by automatic transfer followed by manual curation. Because the preparation of read libraries from the PGABG03 blood sample involved a whole genome amplification step\textsuperscript{13}, DNA fragments may have been generated that were mosaics of coinfecting M1-like and M2 parasite genomes. We therefore performed several quality control steps to identify and remove mosaic contigs and other possible errors (see Methods for details). These assembly steps generated a partial (497 kb) genome sequence, with orthologues of 290 P. malariae genes annotated; however, most gene sequences were incomplete, and there were only 26 full-length genes. Although neither of the two nuclear genes amplified by SGA was covered entirely in this partial genome sequence, we identified 11 reads mapping to one of them (asl), yielding a consensus sequence with only two differences from one of the M2 amplicons (Fig. 1b, Supplementary Fig. 2). Although this similarity indicates that M2 repre

| Species 1 | Species 2 | % Amino acid divergence\textsuperscript{a} | Relative divergence\textsuperscript{b} |
|-----------|-----------|----------------------------------|----------------------------------|
| M1 (P. malariae) | M1-like | 0.44 | 1 |
| M1 (P. malariae) | M2 | 8.49 | 19.2 |
| M1-like | M2 | 8.47 | 19.2 |
| P. falciparum | P. praefalciparum | 0.28 | 0.6 |
| P. falciparum | P. reichenowi | 1.58 | 3.6 |
| P. falciparum | P. gaboni | 5.37 | 12.2 |
| P. ovale-curtisi | P. ovale-wallikeri | 2.80 | 6.3 |
| human P. vivax | ape P. vivax\textsuperscript{c} | 0.81 | 1.8 |
| P. vivax | P. knowlesi | 9.00 | 20.4 |
| P. berghei | P. chabaudi | 6.50 | 14.7 |

\textsuperscript{a}From 186 protein sequence alignments. 
\textsuperscript{b}Divergence of each species pair relative to the divergence between M1 and M1-like. 
\textsuperscript{c}As yet it is unclear whether human P. vivax and ape P. vivax are separate species.
humans (M1), and, on the other hand, Plasmodium malariae-related sequences obtained from wild apes (M1-like), although sequences from captive apes fell into both M1 and M1-like clades. This distinction was not apparent in comparisons of nuclear gene sequences (Fig. 1c, d), which showed less variability than the mtDNA sequences. Understanding the relationship between these two lineages is key to elucidating the zoonotic potential of parasites from wild apes, and so we attempted to increase the sampling of M1-like genomes. We performed SWGA\textsuperscript{11,33} on two P. malariae-positive samples, one from a Cameroonian sanctuary chimpanzee (MOpte51017) and the other from a wild-living chimpanzee (Ptv_Leo) from Cote d’Ivoire who had died from an infectious disease. We specifically selected these samples because they lacked other Plasmodium species and could thus be amplified without the risk of generating inter-species recombinants. Each sample was amplified on multiple occasions using different SWGA primer sets, with individual SWGA reactions pooled prior to MiSeq sequencing. Reads from MOpte51017 and Ptv_Leo were mapped to the P. malariae reference genome, covering 23.6 Mb and 0.3 Mb respectively (Supplementary Table 4). For comparison, the two published M1-like genomes covered 20.4 Mb and 6.2 Mb of the same reference (Supplementary Table 4).

SGA sequences of\textit{ asl} and\textit{ ldh} from MOpte51017 were identical to those from\textit{ P. malariae} (sequence 52 in Fig. 1c, and sequence 60 in Fig. 1d), suggesting that this captive chimpanzee was infected with a human parasite, although an M1-like strain particularly closely related to M1 could not be excluded on the basis of these two short sequences. While the Ptv_Leo sample was obtained from a wild ape, the previously sequenced\textit{ cytB} amplicon was insufficient to determine whether this parasite represented an M1 or an M1-like infection\textsuperscript{22,32}. To better characterise these samples, we combined the newly derived genome sequences with data from published human-derived M1 (i.e.,\textit{ P. malariae}) parasites\textsuperscript{24,34,35} (N = 22) and chimpanzee-derived M1-like parasites\textsuperscript{24} (N = 2), identifying SNPs by comparison with the\textit{ P. malariae} reference genome, which we also included in the analysis. The genome sequences varied in quality, and in their coverage of the\textit{ P. malariae} reference (Supplementary Table 4), with very little of the reference genome covered by all strains.

A phylogenetic network generated from SNPs in genome regions covered by all four chimpanzee-derived parasite genomes and by the six highest coverage human-derived parasite genomes revealed a monophyletic M1 lineage that included MOpte51017 (Fig. 4a), confirming that the captive chimpanzee was indeed infected by a parasite of human origin. In contrast, the Ptv_Leo sequence was similar to the two previously characterised M1-like genomes (GA01 and GA02) in being quite distinct from the M1 clade, consistent with it being a wild-acquired chimpanzee parasite. These three M1-like genomes, plus the M1 clade, were all approximately equidistant (Fig. 4a), suggesting that the ancestor of the M1 clade was a member of the M1-like lineage.

Reticulation at the centre of the network (Fig. 4a) was consistent with gene flow among the various M1-like lineages, but there was little to indicate that there had been any exchange between M1 and any M1-like strain subsequent to the radiation within the M1 clade. Principal component analysis (PCA) of polymorphisms across 1.5 Mb of sequence across high-coverage M1 (N = 16) and M1-like (N = 2) strains was consistent with this conclusion, showing no M1 genomes intermediate between M1 and M1-like (Supplementary Fig. 4). We also searched for any genomic regions exhibiting unusually high diversity within M1, where some strains might have low divergence from M1-like, potentially reflecting recent introgression from an M1-like strain. Comparing each sample to the\textit{ P. malariae} reference genome, we plotted divergent SNPs per callable site for each strain in sliding windows along each chromosome; unlike the phylogenetic network and PCA, this approach allowed inclusion of genomic regions with coverage from only one or two strains.

Across almost all of the genome (>99% of 20 kb windows in 22 Mb examined), each M1 strain differed from the reference at less than 0.001 SNPs per site, whereas the three M1-like strains (GA01, GA02 and Ptv_Leo) differed at ~0.005–0.015 SNPs per site (Supplementary Fig. 5). We found only one candidate region, extending over about 45 kb on chromosome 10, where some M1 strains were as divergent from the reference as GA01 was, while GA02 exhibited regions where divergence from the reference was unusually low (Fig. 5a, Supplementary Fig. 6).

Across this region, no single strain was consistently divergent from the reference or consistently similar to GA02 (Supplementary Fig. 6). Thus, if this region does reflect introgression from an M1-like strain, it appears to have subsequently undergone recombination with other M1 strains. Consistent with this, phylogenetic networks derived from this region showed almost complete reticulation, whether based on only the small number of strains that included most of the region (Fig. 5b) or on more strains covering fewer SNPs (Fig. 5c, Supplementary Fig. 7). The region of high diversity extends across 12 genes, of which four appear to be house-keeping (the likely products are a translation initiation factor, ADP-ribosylation factor, methionine-tRNA ligase, and U2 snRNA/5S RNA pseudouridine synthase), while two encode proteins of unknown function conserved in\textit{ P. falciparum} and\textit{ P. vivax}, and the other six are unidentified and have no clear orthologues in other Plasmodium species. It is therefore unclear how or why the novel alleles might have been selectively retained following introgression. These findings suggest that, in this region of chromosome 10, variation among human-derived M1 strains reflects hybridisation with an M1-like strain from apes, but since this region covers only about 0.2% of the genome sequence examined, introgression appears to have been very rare.

Reduced diversity and increased nonsynonymous polymorphism in M1. Previous studies comparing human\textit{ Plasmodium} species (i.e.,\textit{ P. falciparum} and\textit{ P. vivax}) with their close relatives infecting apes have revealed different levels and patterns of nucleotide polymorphism in parasites from the two hosts\textsuperscript{11,12,36}. For\textit{ P. malariae}, the phylogenetic network (Fig. 4a) indicated that the M1-like lineage had considerably higher diversity than M1, consistent with a previous study of a smaller dataset\textsuperscript{24}. Calculating mean pairwise nucleotide diversity (\(\pi\)), we found that genetic variation among the three M1-like parasites from chimpanzees was 1.04%, nearly 40-fold higher than the value (0.026%) seen for M1 strains (Table 2). To investigate the possibility that

### Table 2 Nucleotide diversity and polymorphisms in M1 and M1-like parasites.

| Statistic\textsuperscript{a} | M1 (\textit{P. malariae})\textsuperscript{b} | M1-like\textsuperscript{b} |
|-----------------------------|---------------------------------|-------------------|
| \(\pi\)                     | 0.00026                         | 0.01043           |
| \(\theta_0\)                 | 0.00024                         | 0.00761           |
| \(\theta_s\)                 | 0.00040                         | 0.02241           |
| \(\theta_s/\theta_0\)        | 0.594                           | 0.340             |
| NS                           | 1981                            | 13,073            |
| S                             | 1031                            | 10,594            |
| NS/S                         | 1.921                           | 1.234             |
| Neutrality index             | 1.551                           | 0.996             |

\textsuperscript{a}\textit{ \(\pi\)} means pairwise nucleotide diversity; \(\theta_0\) means pairwise diversity at zerofold degenerate sites; \(\theta_s\) means pairwise diversity at fourfold degenerate sites; NS, number of nonsynonymous polymorphisms; S, number of synonymous polymorphisms.

\textsuperscript{b}Divergence between M1 and M1-like is 0.00961 substitutions/site.
the higher diversity among M1-like strains was caused by an elevated rate of sequencing errors, we compared the numbers of changes that were due to transition or transversion mutations, restricting the analysis to sites within coding sequences where all possible changes are synonymous (fourfold degenerate sites). Sequencing errors are much more often transitions than transversions, and so a data set with more errors is expected to have a higher transition:transversion ratio than one with fewer errors. However, the transition:transversion ratio for SNPs in M1-like (1.25) was slightly lower than that in M1 (1.31), indicating that increased sequencing error is unlikely to explain the much greater diversity in M1-like.

The difference in pairwise nucleotide diversity between M1 and M1-like strains was about 32-fold at sites where all changes result in a change to the amino acid sequence (zerofold degenerate sites), but about 56-fold at fourfold degenerate sites (Table 2), indicating that M1 has a relative excess of nonsynonymous diversity compared with M1-like. Calculating the ratio of nonsynonymous to synonymous polymorphisms for each lineage, we found that while both lineages had more nonsynonymous than synonymous polymorphisms, the ratio (NS/S) was considerably higher for M1 (1.92) than for M1-like (1.23). Among fixed differences between M1 and M1-like strains, the NS/S ratio was 1.24, yielding a neutrality index of 0.996, very close to the neutral expectation of 1. In contrast, the NI for M1 was 1.551, again indicating an excess of nonsynonymous polymorphisms, and suggesting that it is the M1 lineage that is unusual in this regard.

We wanted to examine the distribution of this pattern of nonsynonymous polymorphism across genes. Many genes in M1 had no synonymous polymorphisms, and thus the NI value could not be defined. Instead we used the Direction of Selection (DoS) statistic, which contrasts the fraction of nonsynonymous changes among fixed differences with that among polymorphisms, and can be calculated for genes with no synonymous polymorphisms. DoS values for M1-like strains were centred around the neutral expectation of 0, whereas for M1 strains the entire distribution was shifted towards negative values (Fig. 6a), indicating that a higher fraction of polymorphisms than of fixed differences were nonsynonymous. Thus, the excess of nonsynonymous polymorphisms in M1 is not due to a small number of unusual genes but is common across the genome.

The site frequency spectrum (SFS) of polymorphisms is expected to vary dependent on the level of selection, with more strongly deleterious mutations expected to be rarer. Synonymous changes are usually assumed to be neutral, and so the SFS of polymorphisms at fourfold degenerate sites can be used as an approximation of the neutral expectation. The unfolded SFS for polymorphisms at zerofold degenerate sites in M1 strains did not
show the expected excess of these nonsynonymous changes in the singleton class, compared to polymorphisms at fourfold degenerate sites (Fig. 6b). This result suggests that these nonsynonymous polymorphisms in the M1 lineage have been subject to little or no purifying selection.

Origin of *P. brasilianum*. In our analyses of mtDNA and nuclear loci, the *P. brasilianum* sequence was identical to that of *P. malariae* from humans (Fig. 1). Previous studies considering the origin of *P. brasilianum* have similarly used data from only a small number of loci4,24–26,29–31. To investigate this question in more detail, we compared the *P. brasilianum* genome40 to numerous *P. malariae* sequences, generating a phylogenetic network from 2,050 SNPs covered in all M1 strains (Fig. 4b). This network divided human M1 (i.e., *P. malariae*) into two clusters, corresponding to strains from Africa and Asia, as reported recently35. Unsurprisingly, the M1 sample from a captive Cameroonian chimpanzee (MOpte51017) clustered with African strains. Addition of the M1-like strain GA01 to this network suggested that the M1 root lay near the base of the African cluster (Supplementary Fig. 8), although resolution in this region is poor because the network is dominated by differences between M1 and M1-like. *P. brasilianum* clustered with the African strains of *P. malariae* (Fig. 4b), suggesting that its origin was a human strain imported to South America from Africa.

Discussion
It has been known for about a century that chimpanzees are infected by parasites morphologically indistinguishable from *P. malariae* in humans41,42. The chimpanzee parasite was named *P. rodhaini*17, but the question of whether the ape and human parasites represent two distinct species was debated, with the eventual consensus being that they do not16. Here we show that *P. malariae*-related sequences from ape samples in fact belong to two quite divergent species, and that both infect gorillas as well as chimpanzees. We expect that these two species are...
two may have split several million years ago. However, with the characterisation of a third Plasmodium-like genome, our analyses indicate a different scenario: P. malariae falls within the radiation of the Plasmodium-like parasites, being no more divergent from the ape parasites than they are from each other (Fig. 4a). Moreover, the fact that the human parasites exhibit substantially less genetic variation than the strains from apes (Fig. 4; Table 2) indicates that P. malariae arose from a recent transmission of an ape parasite to humans. P. malariae has clearly experienced a severe genetic bottleneck, which is most parsimoniously attributed to the cross-species transmission event. The increased fraction of polymorphisms within P. malariae that are non-synonymous, compared to polymorphisms within P. malarielike, is consistent with rapid population expansion subsequent to that bottleneck. Our scan across the core genome found only a single region, on chromosome 10, where diversity among P. malariae strains is substantially higher than elsewhere in the genome. This region amounts to less than 0.21% of the core genome, suggesting that there has been very little introgression from ape parasites into the human parasite population since it started to diversely following its zoonotic origin.

This relationship between the human and ape P. malariae-related parasites is remarkably similar to that between human P. falciparum and ape parasites from the subgenus Laverania. P. falciparum appears to have been derived from one species of gorilla parasite (P. praefalciparum) quite recently, exhibits much reduced genetic diversity relative to the gorilla parasites, and there is little sign of introgression from the gorilla parasites subsequent to the origin of the human parasite. No evidence has been found of P. praefalciparum, or other Laverania, infecting humans living in regions close to infected apes, just as P. malariae-related infections of ape origin have not been found in humans.

A notable difference between the two Plasmodium clades concerns their host specificity. In the wild, ape Laverania appear to be strongly host-specific: the three species infecting gorillas have not been found in wild chimpanzees, while the four species infecting chimpanzees and/or bonobos (two extremely closely related hosts) have never been detected in samples from wild gorillas. The basis of these barriers to cross-species transmission of Laverania are, to some extent, understood. In contrast, the two P. malariae-related ape parasite species have been found in samples from both chimpanzees and gorillas, and there seems no reason a priori to expect that they could not infect humans. Indeed, some 70–80 years ago, it was reported that P. malariae could be transmitted to chimpanzees, and that P. rodhaini (the P. malariae-related parasite seen in chimpanzee blood) could be transmitted to humans. However, most of these experiments involved inoculation of parasitised blood, where the number of parasites transmitted would greatly exceed those from a mosquito bite. Later, Bray was able to infect Anopheles gambiae mosquitoes with P. malariae from humans, which were then able to transmit these parasites to chimpanzees. These observations are consistent with our finding of P. malariae of human origin infecting captive apes. In contrast, Bray was unable to infect A. gambiae with P. malariae-related parasites from chimpanzees, and he concluded that the risk of a zoonotic transmission of P. malariae-related parasites from chimpanzees to humans was much lower than the reverse anthroponosis. Indeed, our genomic analyses indicating a scarcity of introgression from ape parasites into P. malariae provide direct evidence that zoonotic transmission is extremely rare, and that P. malariae has effectively become a new species, isolated from the related ape parasites. If recognised as such, P. malariae-like might retain the name P. rodhaini, or, given that the original “P. rodhaini” was likely a mix of two divergent species, it could be renamed as P. praemalariae, recognising its role in the origin of the human parasites. Importantly, until we know the basis of the apparent isolation between the ape and human parasites, we cannot exclude the possibility that P. praemalariae could emerge in humans again.

It has long been known that there is a very close relationship between human P. malariae and P. brasilianum, a parasite infecting New World monkeys, but the direction of transfer between hosts has been debated, even quite recently. Now that P. malariae has been shown to have a relatively recent African origin, it is clear that P. brasilianum arose subsequent to the
export of the human parasite to South America, most likely within the last 500 years. As yet, only a single *P. brasilianum* genome has been characterised20,29, and so it remains unknown whether these parasites, which have been found to infect around 30 different monkey species in several countries across South and Central America30, are the result of a single introduction, or multiple transmissions, from humans.

The existence of the second *P. malariae*-related ape parasite, termed M2 above, was first suggested by an analysis of mtDNA sequences (Fig. 1a; see also ref. 30). We have assembled a partial M2 genome sequence, from reads that had gone unnoticed in a multiply-infected chimpanzee sample used to generate a *P. gaboni* genome sequence31. These data indicate that the extent of divergence between M2 and *P. malariae* is similar to that between *P. vivax* and its relatives infecting Asian monkeys (e.g., *P. knowlesi*), and greater than that among the most divergent *Laverania* species, such as *P. falciparum* and *P. gaboni* (Fig. 3; Table 1). Thus, M2 parasites clearly represent a previously undescribed species, for which we suggest the name *Plasmodium celatum* (in reference to its being hitherto hidden). Sequence data from this parasite have been found in samples from chimpanzees, bonobos and western gorillas, as well as in mosquitoes, with the latter species, *A. vinckeii*, found to be infected with a variety of other chimpanzee and gorilla *Plasmodium* parasites23. Moreover, the M2-positive samples were collected at several sites in Cameroon, Gabon and the Republic of the Congo, as well as one site in the eastern Democratic Republic of the Congo. Thus, this parasite is widely distributed and appears to exhibit no host specificity in apes, although it has never been found in humans. Since two of the three species within the *P. malariae*-related parasite lineage have only been found infecting African apes, while the third reflects a recent zoonosis, it seems clear that this lineage evolved in Africa, and indeed it seems highly likely that the entire radiation of mammalian *Plasmodium* parasites originated there36.

The timescale of the origins of the *Plasmodium* species infecting humans has been the subject of much speculation. Rutledge and colleagues estimated that the divergence of *P. malariae* and *P. malariae*-like occurred at the same time as the divergence, within the *Laverania*, of the ancestors of *P. falciparum* and the chimpanzee parasite *P. reichenowi*24. They suggested that these two events in the two distinct lineages each involved a move between non-human primates and humans, and that their similar timing may indicate that they both resulted from a common underlying historical event34. However, it is now clear that neither of these divergence events involved ape-to-human transmissions. Indeed, it was already known that this divergence within the *Laverania* involved the split between the ancestors of *P. reichenowi* and *P. praefalciparum*, a gorilla parasite51, and that the gorilla-to-human transmission that gave rise to *P. falciparum* occurred much later11,13. Similarly, our analyses now indicate that the common ancestor of *P. malariae* and any one genome from *P. malariae*-like, as analysed by Rutledge et al., in fact reflects the coalescence of alleles within *P. malariae*-like, and again the ape-to-human transfer giving rise to *P. malariae* occurred much later. One estimate suggests that *P. falciparum* originated about 50,000 years ago13,43, although we have suggested a more recent timeline, of 5000–10,000 years ago36. Whatever the age of *P. falciparum*, the origin of *P. malariae* is similarly recent: levels of genetic diversity in *P. malariae* are even lower than in *P. falciparum*36, although this may be due in part to a slower mutation rate in the quartan parasite24. Thus, there may indeed have been a common historical event that promoted the emergence of both *P. falciparum* and *P. malariae* in humans, such as the beginnings of agriculture in Africa.

**Methods**

**Ape DNA samples.** Ape samples were selected from existing specimen banks previously collected from captive and wild-living chimpanzees and gorillas in west and central Africa (Supplementary Table 1). Whole blood and dried blood spots were obtained from central (Pan troglodytes troglodytes) and Nigeria-Cameroon (P. t. ellioti) chimpanzees housed at the Sanaga Yong (SY) Chimpanzee Rescue Center and the Mblo (MO) National Park Wildlife Rescue Center in Cameroon12,13. Additionally, a small amount of blood (Swg565) was obtained from a western lowland gorilla (*Gorilla gorilla gorilla*), which was killed by hunters and confiscated by the anti-poaching programme of the Cameroononian Ministry of Environment and Forestry32. Spleen and lung samples were collected from a wild-living habituated western chimpanzee (*P. t. verus*) (Leo), who died in 2002 in the forest of the Cote d’Ivoire33. Faecal samples were collected non-invasively from non-habituated apes at numerous sites across equatorial Africa8,10,21,32. DNA was extracted from whole blood and dried blood spots using the QiAamp Blood DNA Mini Kit or the Puregene Core Blood Kit (Qiagen). Faecal DNA was extracted using the QiAamp Stool DNA mini kit (Qiagen). All specimens were subjected to host mtDNA and to confirm their species and subspecies origin. Sample collection at the time was approved by the Ministry of Environment and Forestry of the respective countries, and all 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.

**Diagnostic PCR for *P. malariae*-related parasites.** Ape blood and faecal samples were screened for *P. malariae*-related sequences by diagnostic PCR using both pan-*Plasmodium* and *P. malariae*-specific primer sets as previously described21.

Briefly, a 956 bp mitochondrial cytB fragment was amplified using pan-*Plasmodium* primers, and a 600 bp mitochondrial cytB fragment was amplified using *P. malariae*-specific primers (Supplementary Table 3). Amplicons were sequenced and used for cloning and phylogenetic analysis to determine the species of the amplified *Plasmodium* sequences. The chimpanzee spleen sample Ptv_Leo was previously reported to be positive for *P. malariae*32.

**Limiting dilution PCR.** Ape samples were subjected to SGA to derive *Plasmodium* sequences without PCR-induced artefacts30,12,21,25. Sample DNA was endpoint diluted in 96-well plates, and the dilution that yielded less than 30% positive wells was used to generate single template derived sequences. For *P. malariae*-positive samples, mitochondrial cytB and 2.5 kb cox1/cytB, and nuclear (ldh, ald) gene regions were amplified using *P. malariae*-specific primer sets (Supplementary Table 5, Supplementary Fig. 1) and previously reported amplification conditions36,8,11.

**Selective whole genome amplification.** Near full-length M1 and partial M1-like genomes (respectively) were amplified from chimpanzee blood (MCype31017) and spleen (Ptv_Leo) samples using SWGA as previously described11,22. Briefly, custom scripts were used to identify sequence motifs (6–12 bp) that occurred frequently in the *P. malariae* reference genome24, but only rarely in the human genome. These were filtered to remove primers that exhibited extreme melting temperatures, were predicted to form homodimers, or bound the *P. malariae* mitochondrial genome or its subtelomeric regions more than three times. The resulting primer sets, Pm_set37 and Pm_set31 (Supplementary Table 6), exhibited high frequency of occurrence in the *P. malariae* genome and low frequency of occurrence in the human genome: SWGA was performed by amplifying whole blood or tissue DNA (100–750 ng) in a 50 μl reaction with phi29 buffer, 1 mM dNTPs, 3.5 μM of SWGA primers (an equimolar mix of primers in the set), 1% BSA and 30 units of phi29 polymerase (New England Biolabs). SWGA conditions included a 1 h ramp-down step (35–30 °C), followed by an amplification step for 16 at 30 °C, followed by a phi29 denaturation step for 10 min at 65 °C. SWGA products were purified using AMPure beads (Beckman Coulter) and stored at 4 °C. To obtain sufficient quantities of parasite genomic DNA for sequencing, ape-derived DNA samples were subjected to multiple rounds (up to 4) of successive SWGA amplification, some of which were performed with alternating primer sets to improve genome coverage. Pooled SWGA products were MiSeq sequenced and the MiSeq reads trimmed using cutadapt15 to remove adaptor sequence.

**Data mining and assembly of an M2 genome.** Full details of the bioinformatics methods used in this section and the next are given in Supplementary Note 2.

Briefly, published read libraries13,35 were filtered to remove host reads by mapping to the host reference genome using bwa28 then mapped to a combined *Plasmodium* reference with bwa followed by a more stringent mapping with small (www.sanger.ac.uk/software/toolkit/small-0). α = 0.6, i.e., 60% of bases required to be identical to the reference). Read pairs with both mates mapped to the *P. malariae* reference genome PmUG01 were considered "*P. malariae*-related" reads and assembled de novo with SPades6. De novo contigs of at least 500 bp with k-mer coverage of at least 2X were aligned to the M1-like genome assembly PmgA0124 using MUMmer36, and contigs that were not mapped were aligned to PmgU01. Additionally, each *P. malariae*-related read set was aligned to M2 sequences for *asl* and *ldh* using small (α = 0.9, i.e., 90% of bases required to be identical to the reference), generating one *asl* sequence (number 42) used in Fig. 1c.
Sequence derived from sample PGAG03 (EBS33307313) was used to generate the M2 assembly. M2 contigs were identified by similarity to PmUG01 or PmlGA01, with 92% nucleotide identity to either reference. They were assumed to be M2, and contigs that had less than 83% or 92–97% nucleotide identity were inspected manually to evaluate if they were likely to be M2-derived. Pseudochromosome scaffolds were generated from the M2 contigs by ordering against PmUG01 using ABACAS35. To reduce assembly errors near contig ends, 100 bp were removed from the end of each contig, followed by re-extension with IMG/8E. Two checks were performed to identify misassemblies, including misincorporation of M1-like sequence: an alignment of P. malariae-related reads to the initial M2 assembly was inspected to identify contigs for which multiple different haplotypes were represented in the mapped reads; and blastin alignments of the assembly to PmUG01 and PmlGA01 were inspected to identify contigs with segments that had unusually high identity to PmUG01 or that mapped to different regions of the reference genomes.

Coding sequences in the final M2 assembly were annotated by transfer from PmUG01 using RATT31, followed by manual inspection and curation. Corrected genetic distances were calculated in R80 (TN93 model) from nucleotide alignments of each gene from M2 with its orthologues from PmUG01 and M1-like strain GA01. Amino acid alignments were generated for genes that had one-to-one orthologues in all Plasmodium species in the analysis (Fig. 3) and used to find the number of differences between pairs of species and thus calculate inter-species distances.

Comparison of M1 and M1-like strains. Chimpanzee-derived read libraries (Ptv_Leo, M0pte51017, GA01 and GA0024) were filtered to remove sequencing reads from the host and from coinfections with other Plasmodium species by mapping with bwa to a combined chimpanzee and Plasmodium reference. Read pairs aligned to P. malariae were used for variant calling, in combination with human-derived M1 read libraries from three reference genomes24,33. Variants from the reference genome (PmUG01) were called and filtered using GATK31 with Haplotype-Caller, excluding subtelomeres (Supplementary Table 7), low-complexity regions (identified using Dustmasker36) and sites where an individual sample had more than one alternative allele. Where a heterozygous SNP call was made, the sample genotype was taken to be the allele supported by most reads. Coverage of the assembly to PmUG01 and PmlGA01 were inspected to identify contigs with misincorporation of M1-like sequence: an alignment of

References

1. Sutherland, C. J. & Polley, S. D. In Genetics and Evolution of Infectious Diseases (ed. Tibayrenc, M.) 487–507 (Elsevier, 2017).
2. Sutherland, C. J. Persistent parasitism: the adaptive biology of Malariae and Ovale malaria. Trends Parasitol. 32, 808–819 (2016).
3. Mueller, I., Zimmerman, P. A. & Reeder, J. C. Plasmodium malariae and Plasmodium ovale – the ‘bashful’ malaria parasites. Trends Parasitol. 23, 278–283 (2007).
4. Collins, W. E. & Jeffery, G. M. Plasmodium malariae: parasite and disease. Clin. Microbiol. Rev. 20, 579–592 (2007).
5. McKenzie, F. E. & Bossert, W. H. Multispecies Plasmodium infections of humans. J. Parasitol. 85, 12–18 (1999).
6. Orioño, E. C., Amenga-Etego, L., Ichogomna, D. S. & Amambua-Ngwa, A. Plasmodium malariae, current knowledge and future research opportunities on a neglected malaria parasite species. Crit. Rev. Microbiol. 47, 44–56 (2021).
7. Scopel, K. K. G., Fontes, C. J. F., Nunes, Â. C., Horta, M. F. & Braga, É. M. High prevalence of Plasmodium malariae infections in a Brazilian Amazon endemic area (Aptacíus—Mato Grosso State) as detected by polymerase chain reaction. Acta Trop. 90, 61–64 (2004).
8. Maxyax, M., Pukrittayakamee, S., Newton, P. N. & White, N. J. Mixed-species malaria infections in humans. Trends Parasitol. 20, 233–240 (2004).
9. Liu, W. et al. Origin of the human malaria parasite Plasmodium falciparum in gorillas. Nature 467, 420–425 (2010).
10. Liu, W. et al. African origin of the malaria parasite Plasmodium vivax. Nat. Commun. 5, 3346 (2014).
11. Sundararaman, S. A. et al. Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria. Nat. Commun. 7, 11078 (2016).
12. Loy, D. E. et al. Evolutionary history of human Plasmodium vivax revealed by genome-wide analyses of related ape parasites. Proc. Natl Acad. Sci. USA 115, E8450–E8459 (2018).
13. Otto, T. D. et al. Genomes of all known members of a Plasmodium subgenus reveal path to virulent human malaria. Nat. Microbiol. 3, 687–697 (2018).
14. 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).
15. Gilabert, A. et al. Plasmodium vivax-like genome sequences shed new insights into Plasmodium vivax biology and evolution. PLoS Biol. 16, e2006035 (2018).
16. Coatney, G. R., Collins, W. E., Warren, M. & Contacos, P. G. The Primate Malariae (US Gov Print Off, 1971).
17. Brumpt, E. Les parasites du paludisme des chimpanzés. C. R. Soc. Biol. 130, 837–840 (1939).
18. Rodhain, J. & Dellaert, B. L’infestation à Plasmodium malariae du chimpanzé chez l’homme. Etude d’une premiere souche isolée de l’anthropoide Pan satyrsus versus. Ann. Soc. Belg. Med. Trop. 23, 19–46 (1943).
19. Rodhain, J. Les plasmodiums des anthropoïdes de l’Afrique centrale et leurs relations avec les plasmodiums humains. Bull. Acad. R. Med. Belgique 6, 21–60 (1941).
20. Bray, R. S. Studies on malaria in chimpanzees: VIII. The experimental transmission and pre-erythrocytic phase of Plasmodium malariae, with a note on the host-range of the parasite. Am. J. Trop. Med. Hyg. 9, 455–465 (1940).
21. Liu, W. et al. Wild bonobos host geographically restricted malaria parasites including a putative new Laverania species. Nat. Commun. 8, 1635 (2017).
22. Loy, D. E. et al. Investigating zoonotic infection barriers to ape Plasmodium parasites using faecal DNA analysis. Int. J. Parasitol. 48, 531–542 (2018).
23. Makanga, B. et al. Ape malaria transmission and potential for ape-to-human transfers in. Afr. Proc. Natl Acad. Sci. USA 113, 5329–5334 (2016).
24. Rutledge, G. G. et al. Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. Nature 542, 101–104 (2017).
25. Fandeur, T., Volney, B., Peneau, C. & Thiouy, B. D. Monkeys of the rainforest in French Guiana are natural reservoirs for P. brasilianum/P. malariae malaria. Parasitology 120, 11–21 (2000).

Data availability

The new sequence data reported in this paper have been deposited in the appropriate NCBI databases under accession numbers MN175636-MN175639, MZ555468-MZ555469, MZ927539 (GenBank, SQA sequences) and PRJNA767638 (Sequence Read Archive, SWGA sequences), as noted in Supplementary Table 1. Scaffolds from the new M2 assembly derived from sequence library EBS333073 are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK061131-BK061144. The study also analysed publicly available data from BioProjects PRJEB1439224 and PRJNA344798 (genome assemblies), and BioProjects PRJEB33434, PRJEB12860 and PRJEB7746 (read libraries), details in Supplementary Table 4.

Received: 21 October 2021; Accepted: 7 March 2022; Published online: 06 April 2022
