Edinburgh Research Explorer

Multigenomic Delineation of Plasmodium Species of the Laverania Subgenus Infecting Wild-living Chimpanzees and Gorillas

Citation for published version:

Digital Object Identifier (DOI):
10.1093/gbe/evw128

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genome Biology and Evolution

Publisher Rights Statement:
© The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Multigenomic Delineation of *Plasmodium* Species of the *Laverania* Subgenus Infecting Wild-Living Chimpanzees and Gorillas

Weimin Liu1, Sesh A. Sundararaman1,2, Dorothy E. Loy1,2, Gerald H. Learn1, Yingying Li1, Lindsey J. Plenderleith3, Jean-Bosco N. Ndjango4, Sheri Speede5, Rebeca Atencia6, Debby Cox6,7, George M. Shaw1,2, Ahidjo Ayouba8, Martine Peeters8, Julian C. Rayner9, Beatrice H. Hahn1,2, and Paul M. Sharp3,*

1Department of Medicine, Perelman School of Medicine, University of Pennsylvania
2Department of Microbiology, Perelman School of Medicine, University of Pennsylvania
3Institute of Evolutionary Biology, and Centre for Immunity, Infection and Evolution, University of Edinburgh, United Kingdom
4Faculty of Sciences, University of Kisangani, Democratic Republic of the Congo
5Sanaga-Yong Chimpanzee Rescue Center, IDA-Africa, Portland, Oregon
6Tchimpounga Chimpanzee Rehabilitation Center, Pointe-Noire, Republic of the Congo
7Africa Programmes, Jane Goodall Institute, Vienna, Virginia
8UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175, and University of Montpellier, France
9Malaria Programme, Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

*Corresponding author: E-mail: paul.sharp@ed.ac.uk.

Accepted: May 24, 2016

Data deposition: This project has been deposited at NCBI GenBank under the accession numbers listed in supplementary table S5, Supplementary Material online.

Abstract

*Plasmodium falciparum*, the major cause of malaria morbidity and mortality worldwide, is only distantly related to other human malaria parasites and has thus been placed in a separate subgenus, termed *Laverania*. Parasites morphologically similar to *P. falciparum* have been identified in African apes, but only one other *Laverania* species, *Plasmodium reichenowi* from chimpanzees, has been formally described. Although recent studies have pointed to the existence of additional *Laverania* species, their precise number and host associations remain uncertain, primarily because of limited sampling and a paucity of parasite sequences other than from mitochondrial DNA. To address this, we used limiting dilution polymerase chain reaction to amplify additional parasite sequences from a large number of chimpanzee and gorilla blood and fecal samples collected at two sanctuaries and 30 field sites across equatorial Africa. Phylogenetic analyses of more than 2,000 new sequences derived from the mitochondrial, nuclear, and apicoplast genomes revealed six divergent and well-supported clades within the *Laverania* parasite group. Although two of these clades exhibited deep subdivisions in phylogenies estimated from organelle gene sequences, these sublineages were geographically defined and not present in trees from four unlinked nuclear loci. This greatly expanded sequence data set thus confirms six, and not seven or more, ape *Laverania* species, of which *P. reichenowi*, *Plasmodium gaboni*, and *Plasmodium billcollinsi* only infect chimpanzees, whereas *Plasmodium praefalciparum*, *Plasmodium adleri*, and *Plasmodium blacklocki* only infect gorillas. The new sequence data also confirm the *P. praefalciparum* origin of human *P. falciparum*.

Key words: *Laverania*, *Plasmodium* parasites infecting chimpanzees and gorillas, cryptic *Plasmodium* species, single genome sequencing, *P. falciparum*.
Introduction

Of the five Plasmodium species known to commonly infect humans, Plasmodium falciparum is by far the most pathogenic, causing over 200 million clinical cases of malaria and over half a million malaria-related deaths annually (White et al. 2014). Given this public health impact, there is an urgent need to elucidate new strategies to combat this pathogen. One avenue for new discovery is to study P. falciparum in the context of its closest simian parasite relatives. For nearly a century, Plasmodium parasites, such that we did not consider these to represent two separate species (Liu et al. 2010a; Rayner et al. 2011). Nevertheless, others continue to recognize both (Prugnolle et al. 2011; Snounou et al. 2011; Pacheco et al. 2013; Boundenga et al. 2015; Herbert et al. 2015; Makanga et al. 2016), implying a total of seven Laverania species infecting apes, and some have split these lineages even further, suggesting at least ten new Laverania species (Zilversmit and Awadalla 2011; Zilversmit et al. 2013).

Perspectives on the number of Laverania species could have been influenced by limited geographic sampling and by the fact that most of the existing genetic information derives from the mitochondrial genome, a small, unparentally inherited, nonrecombining molecule that may not be ideal for delineating species (Galtier et al. 2009). In an attempt to resolve these issues, we have amplified parasite sequences from a large number of additional chimpanzee and gorilla samples collected in geographic regions that have previously been underrepresented. Testing nearly 250 new Laverania positive specimens, we almost tripled the number of existing Laverania sequences. Importantly, we have analyzed multiple unlinked loci from the parasite nuclear genome. Our results indicate that there are six, and not seven (or more), clearly defined ape Laverania species, and that mitochondrial sequences alone are not sufficient to delineate cryptic Plasmodium species reliably.

Expanding the Laverania Sequence Database

To increase the geographic representation of Laverania sequences, we selected 248 samples from more than 100 sanctuary and wild-living apes initially surveyed to characterize the molecular epidemiology of primate lentiviruses (Keefe et al. 2006; Neel et al. 2010; Li et al. 2012; D’arc et al. 2015). Samples from eastern chimpanzees (Pan t. schweinfurthii) and western lowland gorillas (G. g. gorilla) were specifically targeted, because these were underrepresented in previous studies of Laverania infections. All specimens were Laverania positive as determined by either diagnostic or limiting dilution polymerase chain reaction (PCR) amplification of mtDNA sequences. Fecal samples (n = 216) were derived from nonhabitatuated ape populations at 30 field sites across central Africa, including six locations where chimpanzees and gorillas have not previously been screened for Plasmodium infections (fig. 1), while blood samples (n = 32) were obtained from 25 chimpanzees at two sanctuaries (Sanaga Yong Wildlife Rescue Center, SY, and Tchimpounga Chimpanzee Rehabilitation Center, TC), and from one gorilla bushmeat sample (SA) of unknown geographic origin (supplementary tables S1 and S2, P. adleri (G2), and Plasmodium blacklocki (G3) (Liu et al. 2010a; Rayner et al. 2011). Sequences from the fourth proposed chimpanzee parasite species, P. billbrayi (Krief et al. 2010), were very closely related to sequences from P. gaboni parasites, such that we did not consider these to represent two separate species (Liu et al. 2010a; Rayner et al. 2011). Nevertheless, others continue to recognize both (Prugnolle et al. 2011; Snounou et al. 2011; Pacheco et al. 2013; Boundenga et al. 2015; Herbert et al. 2015; Makanga et al. 2016), implying a total of seven Laverania species infecting apes, and some have split these lineages even further, suggesting at least ten new Laverania species (Zilversmit and Awadalla 2011; Zilversmit et al. 2013).
Although the sanctuary chimpanzees were almost all healthy at the time of sampling, their blood samples had higher parasite loads than fecal samples and usually contained one predominant \textit{Laverania} species. In contrast, fecal samples, which contain DNA from both liver and blood stage parasites, contained multiple \textit{Laverania} species (Liu et al. 2010a; Abkallo et al. 2014).

Having assembled a geographically diverse sample set, we next aimed to amplify regions of the mitochondrial, apicoplast, and nuclear parasite genomes. Apes can be simultaneously infected with multiple \textit{Plasmodium} species, and so conventional (bulk) PCR is not appropriate to generate \textit{Plasmodium} sequences for phylogenetic analyses, because Taq polymerase has the propensity to switch templates and to generate in vitro recombinants (Liu et al. 2010a). In contrast, limiting dilution PCR, also termed single-genome amplification (SGA), precludes such Taq-induced artifacts and provides a proportional representation of the \textit{Laverania} variants present in the sample (Liu et al. 2010b). Using this approach, we amplified a 956-bp fragment spanning most of the mitochondrial cytochrome B (\textit{cytB}) gene. From 451 new \textit{cytB} sequences, exclusion of identical sequences from the same sample yielded 259 new \textit{cytB} haplotypes, bringing the total number to 709 (table 1).

Phylogenetic analyses of the newly derived \textit{cytB} sequences showed that they clustered in six divergent and well-supported clades as reported previously (Liu et al. 2010a). Three of these clades (C1–C3) included only samples from chimpanzees, whereas the other three (G1–G3) included only samples from gorillas (fig. 2). Two of the chimpanzee parasite clades (C1 and C2) contained deep subdivisions, correlating with the geographical origin of the samples: in both cases, one sublineage contained only sequences from Nigerian–Cameroonian (\textit{Pan t. ellioti}) and central (\textit{Pan t. troglodytes}) chimpanzees from west central Africa (fig. 2; orange and red, respectively), while the other was almost exclusively comprised of parasite sequences from eastern chimpanzees (\textit{P. t. schweinfurthii}) from the Democratic Republic of the Congo (fig. 2; blue). These subdivisions were apparent in our previous survey (Liu et al. 2010a), but at the time we had only relatively few samples from \textit{P. t. schweinfurthii}, especially within the C2 clade. The many additional \textit{P. t. schweinfurthii} \textit{cytB} haplotypes obtained here (supplementary fig. S1a and e, Supplementary Material online) strongly reinforce this phylogeographic
### Table 1

<table>
<thead>
<tr>
<th>Species/Subspecies</th>
<th>Fecal</th>
<th>Mitochondrial</th>
<th>Nuclear</th>
<th>Apicoplast</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New (hapl) / Pub (hapl)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria-Gammaxx</td>
<td>22(6)</td>
<td>296(114)</td>
<td>105(37)</td>
<td>26(20)</td>
</tr>
<tr>
<td>Chimpanzee (P.t. ellioti)</td>
<td>76(17)</td>
<td>483(177)</td>
<td>43(13)</td>
<td>13(9)</td>
</tr>
<tr>
<td>Central chimpanzee</td>
<td>118</td>
<td>214(144)</td>
<td>15(7)</td>
<td>6(6)</td>
</tr>
<tr>
<td>Eastern chimpanzee</td>
<td>126</td>
<td>168(94)</td>
<td>15(7)</td>
<td>6(6)</td>
</tr>
<tr>
<td>Western lowland gorilla</td>
<td>342</td>
<td>451(259)</td>
<td>67(27)</td>
<td>21(16)</td>
</tr>
<tr>
<td>Total</td>
<td>362</td>
<td>843(514)</td>
<td>121(47)</td>
<td>45(38)</td>
</tr>
</tbody>
</table>

**Note:** NA = not available. Single-template-amplified regions of Laverania mitochondrial (cytB), nuclear (eba165, eba175, p47, and ldh), and apicoplast (clpM) genes have previously been described (Liu et al. 2010a; Sundararaman et al. 2016), with bracket indicating distinguishable haplotypes (hap).

### Phylogenetic Analyses of Laverania Nuclear and Apicoplast Gene Sequences

To ascertain whether the phylogenies obtained from mtDNA sequences were an accurate reflection of parasite relationships, we amplified fragments from four unlinked nuclear loci (table 1). These included portions of genes encoding lactate dehydrogenase (ldh; 772 bp), the erythrocyte binding antigens 165 (eba165; 790 bp) and 175 (eba175; 397 bp), and the gametocyte surface protein P47 (p47; 800 bp). A limited number of ldh, eba165, and eba175 sequences have previously been described (Liu et al. 2010a; Wanaguru et al. 2013; Sundararaman et al. 2016), but the p47 locus, encoding a determinant of mosquito host specificity (Molina-Cruz et al. 2015), has not yet been characterized for ape Laverania species (table 1). Overall, more than 1,000 new nuclear gene sequences were generated from either blood or fecal samples (supplementary table S3, Supplementary Material online). Phylogenetic analyses of these sequences again pointed to six major clades within the Laverania, corresponding to the three chimpanzee and three gorilla specific clades found in the mtDNA phylogeny (fig. 3). Although the eba175 sequences were quite short, the overall topologies of all four nuclear gene phylogenies were identical to the cytB tree: the C1 and G1 clades clustered together, as did the C2 and G2 clades, with C3 and G3 branching between these two pairs. Although all of these phylogenies were midpoint rooted, the placement of the root of the Laverania cytB tree corresponds to that found when more distant Plasmodium species are included as outgroups (Liu et al. 2010a). With midpoint rooting, the C2-G2 pair consistently grouped on one side of the root (figs. 2 and 3). The six major clades were each monophyletic in all four nuclear gene trees, with the single exception of the C1 clade in the ldh tree (fig. 3A). This lower resolution for ldh probably results from the lower level of divergence among the ldh sequences (as is apparent from the scale bar). For all four nuclear loci, geographically diverse representatives of human P. falciparum clustered within the radiation of G1 parasites (fig. 2).
The geographically based subdivisions within the C1 and C2 clades, which were apparent in the mtDNA phylogeny, were not seen in the nuclear gene trees (fig. 3). Although the C1 clades contained only few sequences from eastern chimpanzees (blue), there was no evidence that these clustered apart from those from central chimpanzees (red). The C2 clades comprised the largest number of parasite sequences and exhibited considerable substructure; however, again sequences from samples from the different chimpanzee sub-species were interspersed. Interestingly, as in the mtDNA tree, one P. t. schweinfurthii sample (PAPS368) yielded a C1-related eba165 sequence that was highly divergent from all other C1 sequences (fig. 3); however, repeated attempts to amplify other nuclear regions from this sample were unsuccessful.

Finally, we amplified a 390-bp fragment of the clpM gene from the apicoplast genome (note that this gene has previously been referred to as clpC; Liu et al. 2010a). Additional SGA-derived clpM sequences were obtained from 116 chimpanzee samples (19 from blood and 97 from feces) and 28 gorilla fecal samples. The phylogenetic relationships among these sequences were generally less well supported than for the mtDNA and nuclear loci, which were generally longer and/or faster evolving, and so exhibited more diversity. However, all but one of the sequences fell into the same six major clades (fig. 4). The single exception was a sequence from an eastern chimpanzee sample that fell between the C2 and G2 clades, but the bootstrap support for the branch separating this sequence from C2 was only around 50%. For clpM, the relationship among the six clades differed from that found for other genes in two respects: C3 was not more closely related to C1/G1 (but, again, the bootstrap support for this was low), and midpoint rooting would place C1 plus G1 to one side of the root. Because of the relative lack of resolution of the clpM tree, it is as yet unclear whether these differences reflect a truly different evolutionary history for the apicoplast genome within Laverania. While the short length of the clpM fragment, and its relative conservation, resulted in only few nucleotide differences among sequences within the major clades, there

Supplementary Material online). Sequences are color coded to indicate host: chimpanzees (Pan t. elliottae orange, Pan t. troglodytes in red, and Pan t. schweinfurthii in blue), or gorillas (G. g. gorilla in green), with a two-letter code for the field site of origin (fig. 1). Reference sequences for human P. falciparum and chimpanzee P. reichenowi are in black. C1, C2, and C3 represent the chimpanzee parasite species P. reichenowi, P. gaboni, and P. billcollinsi; G1, G2, and G3 represent the gorilla parasite species P. praefalciparum, P. aden, and P. blacklocki. Bootstrap values are shown for major nodes only (the scale bar represents 0.01 substitutions per site).
was a clear tendency for clpM sequences from chimpanzees to cluster based on geographic or host subspecies origin in both C1 and C2, but not C3, clades (fig. 4; supplementary fig. S6, Supplementary Material online). Thus, the apicoplast genome sequences behave similarly to those from the more rapidly evolving mitochondrial genome, while the four nuclear genes do not support any systematic subdivision within the C1 and C2 clades.

**Discussion**

Classifying *Laverania* species solely on the basis of genetic information has been controversial (Valkiunas et al. 2011). Although a full taxonomic description would be preferable, it is unrealistic to expect life history and morphological data of the various parasite species to be forthcoming. First, the endangered status of chimpanzees and gorillas precludes invasive studies of any kind, including malaria transmission studies. Second, the high frequency of mixed *Laverania* species infections renders a correlation of parasite morphology with genetic information from the same sample difficult. Although blood samples can be obtained from sanctuary apes for health reasons, these animals are usually subpatently infected and exhibit exceedingly low parasitemia levels (Sundararaman et al. 2016). Thus, laser microdissection of single infected erythrocytes on blood films as proposed (Valkiunas et al. 2011) is not feasible. Given the need to provide a standardized nomenclature of these parasites for future investigations, it seems desirable to arrive at a definition of species based on a comprehensive genetic analysis.

Phylogenetic analyses of more than 3,000 mitochondrial, apicoplast, and nuclear *Laverania* sequences, contributed by this and previous studies, consistently point to the existence of six distinct *Laverania* clades. This, together with the strict host specificity of these clades, even at field sites such as GT where all six *Laverania* clades are cocirculating in sympatric chimpanzee and gorilla populations, indicates that there are strong isolating mechanisms preventing interspecific hybridization. While it has been argued that detection of parasite sequences in either feces or blood is not sufficient to prove productive infection (Valkiunas et al. 2011), the high prevalence of *Laverania* infections in wild ape populations and their widespread geographic distribution across central Africa provide compelling evidence for ongoing transmission of these parasites, even if gametocytes on blood smears have not been formally described. Thus, the overwhelming amount of evidence points to the existence of six distinct *Laverania* species infecting apes.

![Figure 3](https://academic.oup.com/gbe/article-abstract/8/6/1929/2574109/fig3) **Fig. 3.**—Maximum likelihood phylogenies of *Laverania* nuclear gene sequences. (A) *ldh* (772 bp), (B) *eba175* (397 bp), (C) *eba165* (790 bp), and (D) *p47* (800 bp) gene sequences. Only distinct haplotypes per field site are shown (the full sets of sequences are shown in supplementary figs. S2–S5, Supplementary Material online). Sequences and clades are colored and labeled as in figure 2. Numbers on branches indicate percent bootstrap support. The scale bars indicate 0.01 substitutions per site.

![Figure 4](https://academic.oup.com/gbe/article-abstract/8/6/1929/2574109/fig4) **Fig. 4.**—Maximum likelihood phylogeny of *Laverania* apicoplast DNA sequences. A subset of 80 SGA-derived caseinolytic protease M (clpM) sequences (390 bp) is shown; only distinct haplotypes per field site are shown (the full sets of 227 clpM sequences appear in supplementary fig. S6, Supplementary Material online). Sequences and clades are colored and labeled as in figure 2. Numbers on nodes indicate percent bootstrap support. The scale bar represents 0.01 substitutions per site.
Although genetic information can inform Plasmodium species taxonomy, new species should not be named based on limited sequence data sets. Plasmodium gaboni (Ollomo et al. 2009) and P. billbrayi (Krief et al. 2010) were each initially proposed based on a small number of mtDNA sequences, from samples of central and eastern chimpanzees, respectively. Subsequent comparisons revealed that these sequences are very closely related, falling within what we term the C2 clade, but the two names have continued to be used by some authors because mtDNA sequences form two distinct sublineages (Prugnolle et al. 2011; Snounou et al. 2011; Pacheco et al. 2013; Boundenga et al. 2015; Herbert et al. 2015; Makanga et al. 2016). By the same criterion, the C1 clade should also be split, because mtDNA again reveals two distinct subclades. However, the additional data reported here shed light on this in several ways. First, within both C1 and C2, it is clear that the mtDNA subdivisions reflect the geographical origins of the samples (fig. 2). Second, sequences from the other organelle genome (the apicoplast) suggest similar phylogeographic splits in both C1 and C2 (fig. 4). Third, in contrast, phylogenies from nuclear DNA sequences fail to show the same subclades within C1 or C2 (fig. 3). (For a quantitative assessment of the difference between organelle and nuclear sequences, with respect to phylogeographic segregation, see supplementary table S4, Supplementary Material online.) These findings are reminiscent of the genetic relationships among the chimpanzee hosts: the central (P. t. troglodytes) and eastern (P. t. schweinfurthii) chimpanzee subspecies seem phylogenetically distinct based on mtDNA (Gagneux et al. 1999), but less so with respect to nuclear genome variation (Gonder et al. 2011). However, we find no phylogenetic distinction between C1 and C2 parasites infecting the chimpanzees living to the north (P. t. ellioti) and south (P. t. troglodytes) of the Sanaga River in Cameroon, despite the stronger genetic differentiation between these two host subspecies (Gagneux et al. 1999; Gonder et al. 2011). These two subspecies carry distinct forms of simian foamy viruses (Liu et al. 2008), while simian immunodeficiency viruses infect chimpanzees to the south, but not to the north, of the Sanaga (Sharp and Hahn 2011). Transmission of these viruses requires intimate contact, whereas the mosquito vectors of Laverania parasites are evidently able to cross the Sanaga River. Thus, the geographic distance between central and eastern chimpanzees (fig. 1) seems to have led to recent population subdivision in their Laverania parasites, which has had a more pronounced effect on organelle genome sequences than on nuclear genes; this is to be expected since these organelle genomes have smaller effective population sizes and will tend to undergo lineage sorting more quickly. However, while mtDNA sequences may reveal incipient subspecies within C1 and C2, nuclear genes suggest that the subclades do not represent distinct species. Thus, there is currently clear evidence for only six Laverania species infecting apes, with C1, C2, and C3 clades representing the chimpanzee parasites P. reichenowi, P. gaboni, and P. billcollinsi, and G1, G2, and G3 representing the gorilla parasites P. praefalci.parum, P. adleri, and P. blacklocki, respectively.

The expanded Laverania sequence data set reported here also corroborates the origin of human P. falciparum. In all analyses, P. falciparum clusters closely with gorilla parasites within the G1 clade. Previous studies of mitochondrial gene sequences have shown that P. falciparum strains exhibit much less genetic diversity than each of the ape Laverania species and that all extant P. falciparum strains form a monophyletic lineage within the radiation of the G1 clade of gorilla parasites (Liu et al. 2010a; Sundararaman et al. 2013). These data indicate that parasites within the G1 clade ancestrally infected gorillas and that P. falciparum emerged from a more recent gorilla-to-human transmission event (Liu et al. 2010a). The same close relationships between human and gorilla parasites within the G1 clade are seen for nuclear (fig. 3) and apicoplast (fig. 4) gene sequences, although the relationships are generally less well resolved than with mtDNA sequences (fig. 2).

Although P. falciparum has been detected in a small number of captive chimpanzees and bonobos living in close proximity of humans (Duval et al. 2010; Krief et al. 2010), none of the 216 ape fecal samples characterized in this study (supplementary table S1, Supplementary Material online) contained P. falciparum sequences, indicating that neither wild-living chimpanzees nor wild-living gorillas are naturally infected with the human parasite. Similarly, epidemiological data indicate that contemporary ape Laverania parasites do not normally infect humans (Sundararaman et al. 2013; Délicat-Loembet et al. 2015). Thus, the human parasites have become isolated from, and do not interbreed with, their progenitors infecting gorillas, such that the two are now distinct species and justifying a distinct name (P. praefalciparum) for the latter.

The classification established here can serve as a unifying framework for evolutionary and biological studies of members of the Laverania subgenus. For example, high-quality P. gaboni and P. reichenowi genome sequences were recently generated from subpatently infected chimpanzee blood samples using a select whole-genome amplification approach (Sundararaman et al. 2016). These samples were SGA typed at multiple loci to ensure amplification of members of only one, and not multiple, Laverania species. Comparison of these and one other full-length Laverania genome (Otto et al. 2014) showed that parasites classified as P. gaboni and P. reichenowi based on subgenomic regions indeed represented distinct species, with no evidence of cross-species mating. Given recent advances in selective amplification and nextgen sequencing approaches (Otto et al. 2014; Sundararaman et al. 2016), it is likely that whole genome sequences from members of all Laverania species will be generated long before species-specific morphological data or validated type specimens can be derived.
Materials and Methods

Ape Samples

Blood samples were obtained from sanctuary chimpanzees (P. t. ellioti and P. t. traglobydes) living in outside enclosures in close proximity to wild apes at the Sanaga Yong Wildlife Rescue Center in Cameroon (n = 30) and the Tchimpounga Chimpanzee Rehabilitation Center in the Republic of the Congo (n = 1) (supplementary table S1, Supplementary Material online). Blood samples were obtained for veterinary purposes only or represented leftover material from yearly health examinations. Most blood samples were preserved in RNAlater (1:1 vol/vol) or as dried blood spots on filter paper without further processing, except for eight samples, which were subjected to density gradient centrifugation in the field to enrich for red blood cells (supplementary table S2, Supplementary Material online). Small quantities of blood were also obtained from one western lowland gorilla (G. g. gorilla) of unknown geographic origin (SA), whose body was confiscated by the anti-poaching program of the Cameroonian Ministry of Environment and Forestry. DNA was extracted from whole blood and dried blood spots using the QiAamp Blood DNA mini Kit (Qiagen, Valencia, CA). Ape fecal samples (n = 216) were selected from an existing bank of chimpanzee and western gorilla specimens (Keefe et al. 2006; Neel et al. 2010; Li et al. 2012; D’arc et al. 2015). These specimens were collected from nonhabituated apes living in remote forest areas, with a two-letter code indicating their field site of origin (fig. 1). Fecal DNA was extracted using the QiAamp Stool DNA mini kit (Qiagen, Valencia, CA). Sample collection was approved by the Ministry of Environment and Forestry in Cameroon and by the Ministry of Forest Economy and Sustainable Development in the Republic of Congo. 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.

Single Template Amplification of mtDNA, Apicoplast, and Nuclear Gene Fragments

To derive Plasmodium parasite sequences devoid of PCR errors, including Taq polymerase-induced misincorporation and template switching, Laverania parasite-positive blood and fecal DNA were end point diluted such that fewer than 30% of PCR reactions yielded an amplification product (Liu et al. 2010b). 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 the mtDNA cytB (956bp), the apicoplast clpM gene (390bp), and the nuclear genes eba165 (790 bp), eba175 (390 bp), p47 (800 bp), and ldh (772 bp). Primers and PCR conditions for cytB, clpM, eba165, eba175, and ldh have previously been reported (Liu et al. 2010a; Wanaguru et al. 2013; Sundararaman et al. 2016). The p47 fragment was amplified using Pf47F449 (5’-GTAGATGATAATGAAAACCGG-3’) (Anthony et al. 2007) and Pf47R1 (5’-AATGTATTGGAAACACATTCCATATAC-3’) in the first round, and Pf47D2F1 (5’-TATCCCCAGGACAAGATAAAAATAT-3’) and Pf47R3 (5’-CAAGTTCCATTATGTYAAMATACAT-3’) in the second round of PCR. For the first round, 2.5 μl of end-point diluted sample DNA was used in a 25 μl reaction volume, containing 0.5 μl dNTPs (10 mM of each dNTP), 10 pmol of each first round primer, 2.5 μl PCR buffer, 0.1 μl BSA solution (50 μg/ml), and 0.25 μl expand long template enzyme mix (Expand Long Template PCR System, Roche). Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 15 cycles of denaturation (94 °C, 10 s), annealing (45 °C, 30 s), and elongation (68 °C, 1 min), followed by 35 cycles of denaturation (94 °C, 10 s), annealing (48 °C, 30 s), and elongation (68 °C, 1 min; with 10-s increments for each successive cycle), followed by a final elongation step of 10 min at 68 °C. For the second round PCR, 2 μl of the first round product was used in 25 μl reaction volume. Cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 60 cycles of denaturation (94 °C, 10 s), annealing (52 °C, 30 s), and elongation (68 °C, 1 min), followed by a final elongation step of 10 min at 68 °C. Amplification products were sequenced directly without interm cloning and analyzed using Sequencher (Gene Codes Corporation, Ann Arbor, MI). Sequences containing double peaks in the chromatogram, indicative of the presence of multiple templates or early PCR errors, were discarded. GenBank accession numbers of newly derived SGA sequences are listed in supplementary table S5, Supplementary Material online.

Phylogenetic Analyses

Sequence alignments were constructed using CLUSTAL W version 2.1 (Larkin et al. 2007) and manually adjusted using MacClade (Maddison and Maddison 1989). Regions that could not be unambiguously aligned were omitted from subsequent phylogenetic analyses. Nuclear gene sequences were subjected to recombination analysis using GARD (Kosakovsky Pond et al. 2006). Evolutionary models for phylogenetic analyses were determined using the Akaike information criterion with jModeltest (version 2.1.4) (Darriba et al. 2012) and PhyML (Guindon and Gascuel 2003). Maximum-likelihood phylogenies with bootstrap support (1,000 replicates) were estimated either jointly with model parameter values by means of PhyML using both nearest-neighbor interchange (NNI) and subtree pruning and regrafting (SPR) with Neighbor Joining and 10 random-addition starting trees (Guindon et al. 2010) or RAxML (Stamatakis 2014). Trees were constructed from partial sequences of the cytB gene from the mitochondrial genome (956bp), clpM gene from the apicoplast genome (390bp), and the ldh (772bp),
ebra175 (397 bp), ebra165 (790 bp), and p47 (800 bp) genes from the nuclear genome.

Data Availability
Limiting dilution PCR-derived sequences have been deposited in the GenBank nucleotide database under accession codes KU665647–KU665803.

Supplementary Material
Supplementary figures S1–S6 and tables S1–S5 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
We thank the staff of project PRESICA for fieldwork in Cameroon, the staff of the Sanaga Yong Rescue Center, and the Tchimpounga Chimpanzee Rehabilitation Center for providing left-over blood samples from captive chimpanzees; the Cameroonian Ministries of Health, Forestry and Wildlife, and Scientific Research and Innovation for permission to perform studies in Cameroon; 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. This work was supported by grants from the National Institutes of Health (R01 AI 091595, R01 AI 058715, R01 Al 120810, R37 Al 050529, T32 Al 007532, P30 Al 045008), the Agence Nationale de Recherche sur le Sida et des Écosystèmes and ANR 11 BSV3 021 01, Projet PRIMAL), and the Wellcome Trust (098051).

Literature Cited
Abikalho HM, et al. 2014. DNA from pre-erythrocystic stage malaria parasites is detectable by PCR in the faeces and blood of hosts. Int J Parasitol. 44:467–473.


Associate editor: Martin Embley