Supporting Information

Elguero et al. 10.1073/pnas.1505665112

SI Materials and Methods

Blood Sample Collection. Blood samples were usually collected in the village health care centers into 7-mL vacutener tubes containing EDTA (VWR International). The tubes were transported daily to the field laboratory for centrifugation (10 min, 2,000 × *g*). Plasma, buffy coat, and red blood cells were stored separately. Samples were preserved at −20 °C until the end of the field mission and then transferred on dry ice to the Centre International de Recherches Médicales de Franceville (CIRMF) and kept at −80 °C until analysis. Red blood cell samples were then processed for *Plasmodium* detection.

DNA Extraction and *Plasmodium* cytb Amplification

For each sample, DNA was extracted from 200 μL of red blood cell clot by using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's recommendations and eluted in 100 μL of elution buffer.

To screen for *Plasmodium* infections, we tested the human samples for *Plasmodium* cytb mitochondrial sequences by a nested PCR already described in Ollomo et al. (1).

Sample Preparation for Multiplex 454 Sequencing. To identify the *Plasmodium* species present in each positive sample and detect mixed infections even when one species constitutes only a small fraction of the parasite load, we used a deep-sequencing approach. More particularly, we used the 454 GS-FLX Titanium technology (Roche) on multiplexed-tagged amplicons of several hundreds of samples amplified for a fragment of 201 bp of the cytb gene, containing SNVs (single nucleotide variants) and allowing to discriminate all human *Plasmodium* species (*P. falciparum, P. malariae, P. ovale*, and *P. vivax*).

The target mitochondrial fragment was amplified by using the amplicons of the first round of PCR obtained during *Plasmodium* diagnostics. Based on these amplicons, a second round of amplification was performed by using the following primers designed for this study (forward: 5′-WAATTAYCCATGYCCATTRAA-3′ and reverse: 5′-CCWGTWGCYTGCATYTATCT-3′). Forward and reverse primers were designed by adding a GS FLX Titanium Primer sequence 7 bp multiplex identifier (MID) tags published by Galan et al. (2). Because these MID tags differ from the others by at least three substitutions, they are tolerant to several errors to avoid misassignment of reads. Every 25-μL reaction mix was composed of 12.5 μL of Multiplex PCR mix 2× (Qiagen), 2.5 μL of Q solution (Qiagen), 0.5 μL of each tagged primer (10 pM each) and 1 μL of the first PCR product. Amplifications were carried out in a thermal cycler by using the following reaction conditions: 95 °C for 15 min, followed by 35 cycles of 94 °C for 30 s, 51 °C for 30 s, 72 °C for 1 min, and a final extension step of 72 °C for 10 min. These tagged primers were validated for amplification of sequences of the appropriate length by using *P. falciparum* 3D7, *P. malariae*, *P. ovale,* and *P. vivax* genomic DNA. Each isolate was amplified by using a unique combination of forward and reverse MID tags. PCR amplification was confirmed visually by nucleic acid staining (EZ VISION DNA Dye, Ambresco), followed by gel electrophoresis (2 g agarose in 100 ml of TBE buffer 0.5×), demonstrating a band of the appropriate size (∼317 bp comprising adaptors). Several positive and negative controls (no template) were used for quality assurance (see Table S1 for the list of the positive controls used in the study).

Amplicons Library Preparation. The PCR products were first purified by using the SPRI method (solid-phase reversible immobilization) (Agencourt, AMPure XP). Then, PCR amplicon concentrations were measured by using the Quant-iT PicoGreen dsDNA kit per manufacturer's instructions (Invitrogen). Known concentrations of control DNA were prepared as directed by the Roche Technical Bulletin (454 Sequencing Technical Bulletin No. 005–2009). We assayed fluorescence intensity by using a Perkin-Elmer VICTOR ×3 multilabel plate reader, with fluorescein excitation wavelength of ∼480 nm and emission of ∼520 nm wavelength.

454 GS-FLX Titanium Sequencing Approach. We prepared four PCR amplicons library pools, each containing equimolar amounts of up to 700 PCR amplicons with unique MID tag combinations. These four pools were sequenced in forward and reverse directions on segregated regions of one full 454 plate by using GS FLX Titanium chemistry (Roche). Sequencing was performed by Beckmann Coulter Genomics.

Sequence Analysis Pipeline. A custom pipeline was developed to demultiplex, denoise and remove PCR and sequencing artifacts from the cytb reads. Sff files obtained from each region on the 454 plate were divided into smaller isolate specific sff files by identification of reads with exact matching MID sequences in both ends. Ambiguous primer sites were then identified (exact match) and trimmed off the flowgrams, and reverse reads were reverse complemented. Forward and reverse reads were then combined to take advantage of bidirectional amplicon sequencing, because the forward read has the highest quality in the 5′-end of the target sequence, and the reverse read improves the 3′-end quality. For each individual, each read was aligned to a set of reference sequences (Table S2) by the Muscle program (3) and then BLAST against these references for species assignment by using the BLASTn program. The maximal *e* value was retained for each read. Reads were discarded if they had a similar BLAST score (similar maximal *e* value) for two reference sequences belonging to two different *Plasmodium* species. For the others, each read was assigned to the species for which it obtained the highest score.

Because several biases are associated with this kind of experimental design (454 sequencing, nested PCR, tagged individuals), which can lead to low-level cross contamination and misassignment of read to individuals, it is advised to remove low-frequency reads within each individual. We here took a conservative 5% cutoff value. This cutoff value represents the lowest proportion of reads one species must have within an individual to be considered as present in this individual.

1. Ollomo B, et al. (2009) A new malaria agent in African hominids. *PLoS Pathog* 5(5): e1000446.
2. Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* 11:296.
3. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Table S1. Positive controls used in the study

Positive control name	Geographic origin	Host species	*Plasmodium* species
Pfalciparum	French Guiana	Human	*P. falciparum*
Pmalariae	Ivory Coast	Human	*P. malariae*
Povale	Central African Republic	Human	*P. ovale*
Pvivax	Soudan	Human	*P. vivax*

Table S2. Reference sequences used to determine the origin of the 454 reads

Reference name	Species	Accession no.
PfalciparumACBS01001974	*P. falciparum*	ACBS01001974
PfalciparumAJ298775	*P. falciparum*	AJ298775
PfalciparumAY282924	*P. falciparum*	AY282924
PfalciparumAY282947	*P. falciparum*	AY282947
PfalciparumAY282957	*P. falciparum*	AY282957
Pfalciparum3D7AY282930	*P. falciparum*	AY282930
PmalariaeAB489194	*P. malariae*	AB489194
PmalariaeAB354570	*P. malariae*	AB354570
PovalecurtisiHQ712052	*P. ovale*	HQ712052
PovalecurtisiAB354571	*P. ovale*	AB354571
PovalewallikeriHQ712053	*P. ovale*	HQ712053
PovalewallikeriKC175307	*P. ovale*	KC175307
PvivaxAY598060	*P. vivax*	AY598060
PvivaxAY598119	*P. vivax*	AY598119
PvivaxY17721	*P. vivax*	Y17721
PvivaxSal-1NC007243	*P. vivax*	NC007243

Other Supporting Information Files

Dataset S1 (XLSX)