The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.

The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.

The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.

The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.

The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.

The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC

Robert Verity, Ozkan Aydemir, Nicholas F. Brazeau, Oliver J. Watson, Nicholas J. Hathaway, Melchior Kashamuwa Mwandagalirwa, Patrick W. Marsh, Kyaw Thwai, Travis Fulton, Madeline Denton, Andrew P. Morgan, Jonathan B. Parr, Patrick K. Tumwebaze, Melissa Conrad, Philip J. Rosenthal, Deus S. Ishengoma, Jeremiah Ngondi, Julie Gutman, Modest Mulenga, Douglas E. Norris, William J. Moss, Benedicta A. Mensah, James L. Myers-Hansen, Anita Ghansah, Antoinette K. Tshefu, Azra C. Ghani, Steven R. Meshnick, Jeffrey A. Bailey & Jonathan J. Juliano

The Democratic Republic of the Congo (DRC) harbors 11% of global malaria cases, yet little is known about the spatial and genetic structure of the parasite population in that country. We sequence 2537 *Plasmodium falciparum* infections, including a nationally representative population sample from DRC and samples from surrounding countries, using molecular inversion probes - a high-throughput genotyping tool. We identify an east-west divide in haplotypes known to confer resistance to chloroquine and sulfadoxine-pyrimethamine. Furthermore, we identify highly related parasites over large geographic distances, indicative of gene flow and migration. Our results are consistent with a background of isolation by distance combined with the effects of selection for antimalarial drug resistance. This study provides a high-resolution view of parasite genetic structure across a large country in Africa and provides a baseline to study how implementation programs may impact parasite populations.
Malaria remains one of the largest global public health challenges, with an estimated 219 million cases world-wide in 2017. Despite decades of scale-up in control, there has been a recent resurgence, particularly in high transmission countries in sub-Saharan Africa. In addition, the emergence of antimalarial resistance poses a major threat to current control and elimination efforts worldwide, and new tools are needed to quantify the changing landscape of drug resistance on timescales relevant to malaria control programmes. Genomics has emerged as an useful method for better understanding parasite populations that can be leveraged to support the design of effective interventions against a continually evolving parasite.

Data from genomic studies provides information that is complementary to epidemiological data, and can help to answer several key questions, including how parasites are transmitted, how drug resistance spreads, and how malaria control efforts impact the diversity of the parasite population. However, to date, efforts to use genomics to inform malaria control efforts have suffered from three major limitations. First, much of the work has been conducted in low-transmission regions, such as Asian transmission fringe regions of Africa, leaving it unclear how useful information can be gathered in the highest transmission settings. Some of these high burden regions have experienced increasing malaria prevalence in recent years and are now the center of strategic plans for control efforts. Second, most genomic studies in Africa have relied upon convenience sampling from a few sites usually collected for other purposes, rather than population-representative samples. Lastly, studies have either relied on relatively few genetic markers, providing limited insight into the complete genome, or on expensive whole-genome sequencing, limiting the number of samples studied. Overcoming these limitations is essential for genomics to have broader impacts on malaria control.

Within Africa, parasite populations have been shown to vary significantly between East and West, as demonstrated by their distinct antimalarial drug susceptibilities and population genetics. However, few genomic studies have incorporated samples from central Africa, limiting our understanding of the connectivity of parasite populations across the continent. The Democratic Republic of the Congo (DRC) is the largest malaria-endemic country in Africa, borders nine countries, and harbors ~11% of global *P. falciparum* malaria cases. The DRC harbors a large, understudied parasite population that likely serves as a bridge between African parasite populations. Limited previous work has shown that the DRC represents a watershed between the East and West African clusters.

### Results

#### Sample quality and filtering

We obtained 2537 samples collected in 2013–2015 from the DRC and surrounding countries (DRC = 2039, Ghana = 194, Tanzania = 120, Uganda = 63, Zambia = 121). All samples were sequenced using two separate MIP panels: a genome-wide panel designed to capture overall levels of differentiation and relatedness, and a drug resistance panel designed to target polymorphic sites known to be associated with antimalarial resistance. The genome-wide panel included 739 ostensibly geographically informative SNPs, chosen on the basis of high differentiation ($F_{ST}$) between surrounding African countries in publicly available genomic sequences made available by the PDB project (see Supplementary Note 1 and Supplementary Data 1), and 1151 putatively neutral SNPs distributed throughout the genome, with an overlap of 56 SNPs that were both neutral and geographically informative. The drug resistance panel included SNPs in known and putative drug resistance genes and has been described elsewhere. The median number of unique molecular identifiers (UMIs) per MIP was 31 (range: 1–8490) for the genome-wide panel, and 10 (range: 1–32,511) for the drug resistance panel. Complete UMI depth distributions are shown in Supplementary Fig. 1. After filtering for samples and loci with sufficient UMI coverage, we were left with 1382 samples and 1079 loci from the genome-wide panel, and 674 samples and 1000 loci from the drug resistance panel, with an overlap of 452 samples between both panels. In addition to these samples, 114 controls consisting of known mixtures were sequenced and used to assess the accuracy of allele calls and frequencies. Expected versus measured allele frequencies for each SNP, calculated from these controls, are shown in Supplementary Fig. 2.

### Complexity of infection in the DRC

Initial analyses focused on the genome-wide MIP panel only. Complexity of infection (COI) for each sample was estimated using THE REAL McCooL (Supplementary Fig. 3). The mean COI was estimated at 2.2 (range 1–8) for the study as a whole. We observed significant differences in COI between countries (Ghana: 1.55 (non-parametric bootstrap 95% CI: 1.39–1.73), DRC: 2.23 (2.15–2.31), Tanzania: 2.17 (1.83–2.51), Zambia: 2.68 (2.39–3.00), Uganda 2.18 (1.87–2.51), and within the DRC we observed a statistically significant relationship between COI and *P. falciparum* prevalence by microscopy at both the province and cluster levels (Supplementary Fig. 4), with higher COIs observed at higher prevalences.

### Population structure in the DRC

We explored population structure through principal component analysis (PCA) evaluated on within-sample allele frequencies at all 1079 genome-wide loci. We found the same separation between East and West Africa described in previous studies (Fig. 1) as well as finer structure between regions within East Africa. DRC samples comprised a continuum between the East and West African clusters.
The relative contribution of each locus to each principal component was quantified through normalized loading values. Relative contributions to the first four principal components are shown in Fig. 2. After the fourth principal component the percent variance explained by subsequent components plateaued (Supplementary Fig. 5). For principal component 1 (PC1) large contributions came from loci distributed throughout the genome, and a relatively larger contribution (65.2%) came from putatively geographically informative SNPs (non-parametric bootstrap, p < 0.001). In contrast, contributions to PC2 were concentrated in a region on chromosome seven in close proximity to *P. falciparum* chloroquine resistance transporter (*pfcrt*), a known drug resistance locus, suggesting that resistance to chloroquine or amodiaquine may be driving differentiation along this secondary resistance locus.

Fig. 1 Principal component analysis. The first two (a) and three (b) principal components calculated from within-sample allele frequencies using the genome-wide MIP panel. Colors indicate country of origin of each sample.

Fig. 2 Per-locus contributions to principal components. The relative contribution (%) of each locus to the first four principal components. Chromosomes are plotted in order, separated by vertical white gridlines. Point colors indicate sites that were chosen in the design based on *F*_ST values to be geographically informative (blue) or not (red).
axis. For PC3, locus contributions were concentrated in three genic regions: PF3D7_0215300 (8.5%), PF3D7_0220300 (5.0%), and PF3D7_1127000 (4.3%). The first and largest of these encodes an acyl-CoA synthetase and is part of a diverse gene family known to undergo extensive gene conversion and recombination14. For PC4 we observed a region of high locus contribution on chromosome eight in close proximity to the known antifolate drug resistance gene dihydropteroate synthase (dhps). Combined, these results suggest that geography and drug resistance are both contributors to the observed population structure.

The relationship between the PCA results and the spatial distribution of parasites was explored by plotting raw principal component values against the geographic location of samples (Fig. 3a–d). For PC1 this revealed a complex pattern of spatial variation, containing both north–south and east–west clines. For PC2 and PC4 the maps essentially recapitulate the known geographic distribution of pfcrt and dhps resistance mutations, respectively (Fig. 3e, f). For PC3 the map indicates some east–west spatial structuring that is not explained by known markers of antimalarial resistance and warrants further investigation.

**Between sample relatedness of parasites.** The relatedness of all pairs of samples was explored through pairwise identity by descent (IBD), estimated using a maximum likelihood approach. IBD describes the relatedness of samples in terms of their shared evolutionary history, and consequently is not influenced by a particular allele frequency distribution. This makes it a better measure than simple identity by state (IBS) when comparing between studies, as values can be compared directly15. We first carried out a simulation-based analysis to explore the accuracy of our maximum likelihood estimator (see Supplementary Note 2 and Supplementary Fig. 6), finding that we were conservatively biased in cases of high polyclonality. Hence, we expect to underestimate true IBD by this method. This result did depend on the number of genotyped positions, with estimates becoming increasingly unreliable for smaller datasets of 100 or 20 SNP loci. In the real data, the overall distribution of pairwise IBD was found to be heavy-tailed, consisting of a large body of weakly related samples and a tail of very highly related samples (Fig. 4).

Mean IBD was significantly higher within clusters compared to between clusters (0.06 vs. 0.02, two-sample t-test, p < 0.001). When plotted against geographic separation there was a clear fall-off of IBD with distance (Fig. 5a), consistent with the classical pattern expected under isolation-by-distance16,17. Focussing on the tail of highly related samples, which includes the major strain in complex infections, there were 12 sample pairs with a

![Fig. 3 Spatial patterns in principal components.](image)

**Fig. 3 Spatial patterns in principal components.** a–d Show the mean principal component value per DHS cluster. e, f Show estimated distributions of the prevalence of molecular markers of resistance for pfcrt and pfdhps.
relatedness greater than IBD = 0.9. Comparison of raw allele frequency distributions confirmed that these were likely clones (Supplementary Fig. 7). These highly related pairs were found more often within the same cluster than in different clusters (7 vs. 5, respectively, chi-squared test, $p < 0.001$), suggesting the presence of local clonal transmission chains. The five between-cluster highly related pairs (Fig. 5b) were spread over large geographic distances (281–1331 km), far beyond the normal expected scale of the breakdown in genetic relatedness (Fig. 5a), suggesting recent long distance migration.

**Prevalence of markers of antimalarial resistance.** Based on previous findings of an east–west divide in molecular markers of antimalarial resistance in the DRC$^8,9$, all samples in the DRC were divided by geographically weighted K-means clustering into two populations (Supplementary Fig. 8). The prevalence of every mutation identified by the drug resistance MIP panel was calculated in eastern and western DRC, as well as at the country level. Table 1 gives a summary of all mutations that reached a prevalence $>5\%$ in any geographic unit, and a complete list of all identified mutations along with their prevalence is given in Supplementary Data 2. Note that in the *dhps* mutation $G437A$ the reference is resistant, hence this is re-coded as $A437G$ and prevalence values indicate the prevalence of the reference allele. Estimated prevalences of these alleles in the DRC as a whole were broadly similar to previously published estimates$^{10}$. However, we did identify several polymorphisms in known and putative resistance genes not previously reported in the DRC, including *kelch* K189T and *pfatp6* N569K, both of which have been described at appreciable frequencies elsewhere in Africa$^{18–20}$.

**Geographic distribution of antimalarial resistant haplotypes.** Previous studies have demonstrated that mutations associated with antimalarial resistance are clustered into east–west groupings within DRC$^8,10$. Focusing on the 107 samples from DRC that were identified as monoclonal from the REAL McCOIL analysis, we explored the joint distribution of all combinations of mutant haplotypes in both the *dhps* and *crt* genes. Raw combinations of mutations were visualized using the UpSetR package in R$^{21}$, and the spatial distribution of haplotypes in the DRC was explored by plotting these same mutant combinations against their corresponding DHS cluster locations (Fig. 6). Our results for *dhps* recapitulate those found previously, showing a clear east–west divide with the K540E and A581G mutants concentrated in the east, and S436A and A437G concentrated in the west. For *crt* we also find evidence of an east–west divide, with haplotypes containing N326S and F325C concentrated in the east and those containing I356T concentrated in the west.

**Selective sweep and haplotype analysis of antimalarial resistance.** Using the antimalarial resistance MIPs and genome-wide SNP MIPs combined, the extended haplotypes of the monoclonal infections were determined for 200 kb upstream and downstream of each putative drug resistance allele that had at least 5\% overall prevalence in the DRC. The CVIET haplotype within the *crt* gene showed a signal of positive selection, with longer haplotype blocks in western DRC as compared to eastern DRC (Fig. 7; $p$-XP-EHH $D < 0.05$). In the east, patterns of haplotype homozygosity are consistent with positive selection for the derived I356T haplotype (Supplementary Fig. 9), although a XP-EHH statistic could not be calculated for this locus because the derived haplotype was absent in western DRC, supporting the geographic localization of the I356T mutation in the east (Fig. 6).

Mutations in *dhps* were more difficult to interpret. This gene has undergone multiple selective sweeps associated with increasing drug resistance. The most recently introduced mutation into the DRC, *dhps* A581G, showed relatively conserved local haplotypes around the mutation in both eastern and western DRC (Supplementary Fig. 10). Extended haplotypes around the other mutations (Supplementary Figs. 11 and 12) are inconsistent with a classical hard sweep, perhaps due to selection on multiple independent haplotypes or to interference between A581G and other linked alleles. Finally, we did not detect any strong signals of differing patterns of recent positive selection between the eastern and western DRC among the *dhfr* and *mdr2* genes (Supplementary Table 1, Supplementary Fig. 13).

**Discussion**

Here we provide the first large-scale, robustly sampled study of *falciparum* malaria in central Africa using MIP capture and sequencing, a high-throughput genotyping approach that is appropriate for large population based surveys. Using a panel of probes designed to detect genome-wide SNPs, combined with a second panel targeting drug resistance genes, we were able to show that the parasite population in the DRC contains a signal of differentiation by geographic separation, consistent with the
Table 1 Prevalence (%) of mutations identified by the drug resistance MIP panel.

| Gene  | Chromosome | Position | Mutation Name | Overall | DRC | DRC West | East | Ghana | Uganda | Zambia |
|-------|------------|----------|---------------|---------|-----|---------|------|-------|--------|--------|
| atp6  | chr1       | 267007   | I723V         | 1.1     | 0.3 | 0.7     | 0.0  | 4.2   | 7.3    | 0.0    |
| atp6  | chr1       | 267257   | G639D         | 2.0     | 1.8 | 2.9     | 1.0  | 0.0   | 7.3    | 0.0    |
| atp6  | chr1       | 267467   | N569K         | 24.1    | 21.9| 18.8    | 24.0 | 16.7  | 41.5   | 28.9   |
| atp6  | chr1       | 267882   | E431K         | 15.3    | 17.0| 18.8    | 15.7 | 16.7  | 9.8    | 6.7    |
| atp6  | chr1       | 267970   | L402V         | 7.1     | 8.2 | 10.1    | 6.9  | 12.5  | 0.0    | 2.2    |
| dhfr-ts | chr4    | 748239   | N51I          | 83.0    | 79.5| 81.2    | 78.4 | 75.0  | 100.0  | 97.8   |
| dhfr-ts | chr4    | 748262   | C59R          | 71.2    | 63.2| 63.0    | 63.2 | 95.8  | 95.1   | 97.8   |
| dhfr-ts | chr4    | 748410   | S108N         | 97.8    | 97.1| 97.1    | 97.1 | 100.0 | 100.0  | 100.0  |
| dhfr-ts | chr4    | 748577   | I64L          | 3.1     | 0.6 | 0.0     | 1.0  | 0.0   | 29.3   | 0.0    |
| mdr1  | chr5       | 958145   | N86Y          | 12.4    | 14.3| 18.8    | 11.3 | 16.7  | 7.3    | 0.0    |
| mdr1  | chr5       | 958440   | Y184F         | 37.4    | 36.5| 39.9    | 34.3 | 58.3  | 31.7   | 37.8   |
| mdr1  | chr5       | 958484   | T199S         | 1.3     | 0.0 | 0.0     | 0.0  | 0.0   | 14.6   | 0.0    |
| mdr1  | chr5       | 958584   | S232Y         | 2.7     | 3.5 | 5.1     | 2.5  | 0.0   | 0.0    | 0.0    |
| mdr1  | chr5       | 961625   | D1246Y        | 4.4     | 2.9 | 3.6     | 2.5  | 0.0   | 24.4   | 0.0    |
| crt   | chr7       | 403620   | M74I          | 30.3    | 28.7| 37.7    | 22.5 | 16.7  | 85.4   | 0.0    |
| crt   | chr7       | 403621   | N75E          | 30.3    | 28.7| 37.7    | 22.5 | 16.7  | 85.4   | 0.0    |
| crt   | chr7       | 403625   | K76T          | 30.3    | 28.7| 37.7    | 22.5 | 16.7  | 85.4   | 0.0    |
| crt   | chr7       | 404407   | A220S         | 28.1    | 24.6| 31.9    | 19.6 | 8.3   | 100.0  | 0.0    |
| crt   | chr7       | 405600   | I356T         | 7.1     | 9.4 | 21.0    | 1.5  | 0.0   | 0.0    | 0.0    |
| dhps  | chr8       | 549681   | S436A         | 15.0    | 17.3| 28.3    | 9.8  | 37.5  | 0.0    | 0.0    |
| dhps  | chr8       | 549865   | G437A         | 26.8    | 32.7| 27.5    | 36.3 | 4.2   | 0.0    | 17.8   |
| dhps  | chr8       | 549993   | K540E         | 25.4    | 17.0| 9.4     | 22.1 | 0.0   | 85.4   | 48.9   |
| dhps  | chr8       | 550117   | A581G         | 8.2     | 6.1 | 2.2     | 8.8  | 0.0   | 34.1   | 4.4    |
| k13   | chr13      | 1726431  | K189T         | 14.8    | 14.9| 18.8    | 12.3 | 54.2  | 0.0    | 6.7    |
| mdr2  | chr14      | 1956202  | I431V         | 23.2    | 21.3| 22.5    | 20.6 | 20.8  | 31.7   | 31.1   |
| mdr2  | chr14      | 1956408  | F423Y         | 31.4    | 30.1| 28.3    | 31.4 | 29.2  | 36.6   | 37.8   |

Includes all mutations that reached a prevalence >5% in any given geographic unit.
panel can be leveraged for conducting similar analyses. For example, the differences in CV between the West and East suggests that the CV in the West has potentially been more recently introduced, has experienced less breakdown through recombination, or has undergone stronger recent positive selection as compared to the East. Redesign of the selected targets with denser sampling around known drug resistance genes will allow for more robust assessment of these selected regions.

DRC’s location in central Africa and the enormous number of malaria cases in the country means that malaria control in Africa likely depends on improving our understanding on Congolese malaria. This represents the largest study of falciparum population genetics in the DRC and, unlike other large population genetic studies of malaria in Africa, leverages a nationally representative sampling approach. Thus, this study provides the first data on fine-scale genetic structure of parasites at a national scale in Africa, and provides a baseline that can be used to study how implementation programs impact parasite populations in the region. The MIP platform represents a highly scalable and cost-effective means of providing genome-wide genetic data, relative to whole-genome sequencing. The highly flexible nature of the platform allows it to be rapidly scaled in terms of targets and samples leading it to be applicable across malaria-endemic countries.

**Methods**

**Study populations.** Chelex-extracted DNA samples from dried blood spots, collected as part of the 2013–2014 DRC Demographic Health Survey (DHS), were tested using quantitative real-time PCR to detect *Plasmodium falciparum* lactate dehydrogenase (*dhb*) previously used for sequencing, which was applied to the remaining DRC samples (*n* = 1450), resulting in a total of 2039 DRC samples sent for sequencing. These samples were selected as part of the overall 539 DHS clusters. In addition, dried blood spot samples from four further counties were used: Ghana (*n* = 194), Tanzania (*n* = 120), Uganda (*n* = 63), and Zambia (*n* = 121). Samples from Ghana were collected in 2014 from symptomatic RDT-positive patients of all ages at Kharumwa Health Center in Northwest Tanzania.

Samples from Uganda were collected in 2013 from RDT-positive individuals from a community survey of all ages in Nchelenge District in northeast Zambia on the border with the DRC. All non-DRC samples were Chelex extracted, except for the Ghanaian samples which were extracted using QiaQuick per protocol (Qiagen, Hilden, Germany). This study was approved by the Internal Review Board at UNC and the Ethics Committee of the Kinshasa School of Public Health.

**Design of MIP panels.** We used two distinct MIP panels—a genome-wide panel designed to capture overall levels of differentiation and relatedness, and a drug resistance panel designed to target polymorphic sites known to be associated with antimalarial resistance (Supplementary Note 1). When selecting targets for the genome-wide panel, we used the publicly available *P. falciparum* whole-genome sequences provided by the Pf3k and *P. falciparum* Community projects from the MalariaGen Consortium. This consisted of sample sets from Cameroon (*n* = 134), Uganda (*n* = 1450), and Zambia (*n* = 121). Samples from Ghana were collected in 2014 from symptomatic RDT-positive individuals presenting at health care facilities in Begoro (*n* = 94) and Cape Coast (*n* = 98). Samples from Tanzania were collected in 2015 from symptomatic RDT-positive patients of all ages at Kharumwa Health Center in Northwest Tanzania. Samples from Uganda were collected in 2013 from RDT-positive symptomatic patients at Kanungu in Southwest Uganda. Finally, samples from Zambia were collected in 2013 from RDT-positive individuals from a community survey of all ages in Nchelenge District in northeast Zambia on the border with the DRC. All non-DRC samples were Chelex extracted, except for the Ghanaian samples which were extracted using QiaQuick per protocol (Qiagen, Hilden, Germany). This study was approved by the Internal Review Board at UNC and the Ethics Committee of the Kinshasa School of Public Health.
mitigated in the east.

In the eastern Democratic Republic of the Congo (DRC) and western DRC, monoclonal samples with no missing genotype data 200 kb upstream and downstream from the K76T core single-nucleotide polymorphism in centimorgans among the samples from the eastern DRC and western DRC. For each haplotype, there is strong evidence for recent positive selection of the core allele ancestry and the eastern DRC and western DRC for a subsetted region. Position is considered in kilobases, and segregating sites unless otherwise indicated. Best Practices with minor modifications. The Pf3k strategy consistent with the Genome Analysis Toolkit (GATK, version 3.6) was carried out as described in the Supplementary Note 1. Samples from Nigeria and Uganda were dropped after variant filtering due to small sample sizes, and the filtered sequences were used to calculate Weir and Cochran’s $F_{ST}$ with respect to country for each biallelic locus. The 1000 loci with the highest $F_{ST}$ values were considered for MIP design as phylogeographically informative loci. Of these 1000 potential loci, 739 were identified as regions that were suitable for MIP-probe design. Separately, from the combined SNP file, we identified 1595 loci that had a minor-allele frequency >5%, had an $F_{ST}$ value between 0.005 and 0.2, and were annotated by SnpEff (version 4.3) as functionally silent mutations. These loci were identified as putatively neutral SNPs, and in the case of the genome-wide panel these were filtered to the pre-designed biallelic target SNP sites. Any variant that was represented by a single UMI in a sample, or that was invariant across the entire dataset after filtering, was eliminated. Any site that was invariant across the entire dataset after this procedure was dropped. Samples were assessed for quality in terms of the proportion of low-coverage sites, where low-coverage was defined as fewer than 10 supporting UMIs. Samples with >50% low-coverage loci were dropped. Variant sites were then assessed by the same means in terms of the proportion of low-coverage sites, and sites with >50% low-coverage were dropped. Samples were then combined with metadata, including geographic information, and were only retained if there were at least 10 samples in a given country. This resulted in dropping Tanzanian samples from the drug resistance dataset, but no other countries were dropped. Post-filtering, genome-wide data consisted of 1382 samples (DRC = 1111, Ghana = 114, Tanzania = 30, Uganda = 45, Zambia = 832 and 1079 loci, and drug resistance data consisted of 674 samples (DRC = 557, Ghana = 29, Uganda = 43, Zambia = 45) and 1000 loci. The complete bioinformatic pipeline is shown in Supplementary Fig. 15.

**MIP variant calling and filtering.** MIP variant calling is summarized in the Supplementary Note 1. Within each sample, variants were dropped if they had a Phred-scaled quality score of <20. Across samples, variant sites were dropped if they were observed only in one sample, or if they had a total UMI count of <5 across all samples. This data set was considered the final raw data used for additional filtering.

Additional filters were applied to both genome-wide and drug resistance datasets prior to carrying out analysis. Sites were restricted to SNPs, and in the case of the genome-wide panel these were filtered to the pre-designed biallelic target SNP sites. Any variant that was represented by a single UMI in a sample, or that had a within-sample allele frequency (WSAF), UMI count of allele/total UMI less than 1%, was eliminated. Any site that was invariant across the entire dataset after this procedure was dropped. Samples were assessed for quality in terms of the proportion of low-coverage sites, where low-coverage was defined as fewer than 10 supporting UMIs. Samples with >50% low-coverage loci were dropped. Variant sites were then assessed by the same means in terms of the proportion of low-coverage sites, and sites with >50% low-coverage were dropped. Samples were then combined with metadata, including geographic information, and were only retained if there were at least 10 samples in a given country. This resulted in dropping Tanzanian samples from the drug resistance dataset, but no other countries were dropped. Post-filtering, genome-wide data consisted of 1382 samples (DRC = 1111, Ghana = 114, Tanzania = 30, Uganda = 45, Zambia = 832 and 1079 loci, and drug resistance data consisted of 674 samples (DRC = 557, Ghana = 29, Uganda = 43, Zambia = 45) and 1000 loci. The complete bioinformatic pipeline is shown in Supplementary Fig. 15.

**Fig. 7 Extended haplotype homozygosity and bifurcation plots for pfcrt K76T.**

a. Display extended haplotype homozygosity (EHH) curves from the monoclonal samples with no missing genotype data 200 kb upstream and downstream from the K76T core single-nucleotide polymorphism in DRC-East.

b. Show haplotype bifurcation plots with respect to the core allele ancestry and the eastern DRC and western DRC for a subsetted region.

c. Position (cM) of the genome-wide panel these were filtered to the pre-designed biallelic target SNP sites.

d. Position (kbp) of the genome-wide panel these were filtered to the pre-designed biallelic target SNP sites.
Complexity of infection. We applied THE REAL McCoIL (v2) categorical method to the SNP genotyped samples to estimate the COI of each individual\(^1\). Details of the analysis are in the Supplementary Note 1.

Analysis of population structure. WSAFs were calculated for all genome-wide SNPs, with missing values imputed as the mean per locus. Principal component analysis (PCA) was carried out on WSAFs using the `prcomp` function in R version 3.5.1. The relative contribution of each locus was calculated from the loading values as \(l_i^2/\sum l_i^2\), where \(l_i\) is the absolute value of the loading at locus \(i\), and \(L\) is the total number of loci. PCA results were explored in a spatial context by taking the mean of the raw principal component values over all samples in a given DHS cluster, and plotting this against the geoposition of the cluster.

Identity by descent analysis. Pairwise IBD was calculated between all samples from the genome-wide SNPs. We used Malecot's\(^4\) definition of \(f\) as the probability of identity by descent, where \(f\) can be defined as the probability of a randomly chosen locus being IBD between samples \(u\) and \(v\). At locus \(i\), let \(A\) denote the reference allele, which occurs at population allele frequency \(p\), and let a denote the non-reference allele, which occurs at population allele frequency \(q = 1 - p\). Assuming that both samples \(u\) and \(v\) are monochonal, let \(X_a\) denote the observed allele at locus \(i\) in sample \(u\), and equivalently let \(X_v\) denote the observed allele in sample \(v\).

\[
\begin{align*}
Pr(X_u = A, X_v = a) &= f_{uv}p(1 - f_{uv})p^2 \\
Pr(X_u = A, X_v = A) &= f_{uv}(1 - f_{uv})p^2 \\
Pr(X_u = a, X_v = a) &= f_{uv}(1 - f_{uv})q^2 \\
Pr(X_u = a, X_v = A) &= f_{uv}(1 - f_{uv})q \\
Pr(X_u = A, X_v = a) &= f_{uv}(1 - f_{uv})p \\

L(X_u, X_v) &= \sum_i Pr(X_u = A, X_v = a). 
\end{align*}
\]

In practice, population allele frequencies \((p, q)\) were calculated using the mean WSAF for that locus over all samples. Samples were then coerced to monochonal by calling the dominant allele at every locus. The likelihood was evaluated using Eq. (2) in log-space for a range of values \(f_{uv}\) distributed between 0 and 1 in equal increments of 0.02. The maximum likelihood estimate \(f_{	ext{max}}\) was calculated for each group. Finally, sample pairs with IBD > 0.9 were identified, and explored in terms of their WSAFs and their spatial distribution.

Estimating mutation prevalence from drug resistance panel. Given previous
findings of an East–West divide in molecular markers of antimalarial resistance in the DRC\(^6\), all samples in the DRC were divided by geographically weighted \(K\)-means clustering into two populations. The prevalence of every mutation identified by the drug resistance MIP panel was then calculated in East and West DRC, as well as at the country level. Prevalences in each DHS cluster were used to produce smooth prevalence maps using the `prcomp` function in R version 1.4.2 in \(K\).

Analysis of monochonal haplotypes. Results of the previous COI analysis on the genome-wide SNPs with THE REAL McCoIL were used to identify samples that were monochonal with a high degree of confidence. Samples were defined as monochonal if the upper 95% credible interval did not include any COI greater than one. This resulted in 408 monochonal samples, of which 143 overlapping with the drug resistance MIP dataset and therefore could be used to explore the joint distribution of mutations in drug resistance genes. 107 of these were from DRC. Analysis focused on the \(d\)ips and \(c\)rt genes. Raw combinations of mutations were visualized using the `UpSet` package in R\(^6\), and the spatial distribution of haplotypes was explored by plotting these same mutant combinations against DHS cluster geoposition.

Extended haplotype homogeneity analysis. In order to improve our power to detect hard-sweeps and capture patterns of linkage-disequilibrium with EHH statistics among putative drug resistance SNPs, we combined the genome-wide and the drug resistance filtered biallelic SNPs into a single dataset (Supplementary Note 1). All associated EHH calculations were carried out using the R-package rehh (version 2.6.0), and were truncated when fewer than two haplotypes were present or the EHH statistic fell below 0.05\(^4\). In addition, we allowed EHH integration calculations to be made without respect to “neighbors,” which were frequent due to the MIP-probe design. Although this would result in an inflated integration statistic if the EHH statistic had not yet reached 0 within the region of investigation, this problem was mitigated by only comparing between subpopulations, and not between loci. EHH decay, bifurcation plots, and haplotype plots were adapted from the rehh package objects and modified using ggplot\(^3\).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. DHS data for the 2013 DRC DHS is available here: https://dhsprogram.com/what-we-do/survey/survey-display-421.cfm. This includes clinical and GPS information and is available upon request from the DHS program. All raw sequencing data is available at the NCBI SRA (Accession numbers: PRJNA454490, PRJNA545345, and PRJNA545347).

Code availability. Tools for MIP variant calling and filtering are available at https://github.com/bailey-lab/MIPTools (v.0.19.12.13) and https://github.com/Mrc-ide/mipanalyzer (v.1.0.0). Code and data are available for each figure at https://github.com/bobverity/antimalarial_resistance_DRC. Code access is unrestricted.

Received: 31 May 2019; Accepted: 28 March 2020; Published online: 30 April 2020

References

1. WHO. World Malaria Report 2017 (WHO, 2018).
2. Neafsey, D. E. & Volkman, S. K. Malaria genomics in the era of eradication. Cold Spring Harb. Perspect. Med. 7, https://doi.org/10.1101/cshperspect.a025544 (2017).
3. Bill & Melinda Gates Foundation. Malaria, https://gatesfoundation.org/What-We-Do/Global-Health/Malaria (2019).
4. World Health Organization. WHO: High Burden to High Impact. A Targeted Malaria Response (WHO, 2019).
5. Pearce, R. J. et al. Multiple origins and regional dispersal of resistant dhps in African Plasmodium falciparum malaria. PLoS Med. 6, e1000555 (2009).
6. Ocholla, H. et al. Whole-genome scans provide evidence of adaptive evolution in Malawian Plasmodium falciparum isolates. J. Infect. Dis. 210, 1991–2000 (2014).
7. Carrel, M. et al. The geography of malaria genetics in the Democratic Republic of Congo: a complex and fragmented landscape. Soc. Sci. Med. 133, 233–241 (2015).
8. Taylor, S. M. et al. Plasmodium falciparum sulfadoxine resistance is geographically and genetically clustered within the DR Congo. Sci. Rep. 3, 1165 (2013).
9. Antonia, A. L. et al. A cross-sectional survey of Plasmodium falciparum pfcrt mutant haplotypes in the Democratic Republic of Congo. Am. J. Trop. Med. Hyg. 90, 1094–1097 (2014).
10. Aydemir, O. et al. Drug resistance and population structure of Plasmodium falciparum across the Democratic Republic of Congo using high-throughput molecular inversion probes. J. Infect. Dis. https://doi.org/10.1093/infdis/iij228 (2018).
11. Verity, R. et al. Plasmodium falciparum genetic variation of var2csa in the Democratic Republic of the Congo. Malar. J. 17, 46 (2018).
12. O’Boak, B. J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338, 1619–1622 (2012).
13. Chang, H.-H. et al. THE REAL McCoIL: a method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. PLoS Comput. Biol. 13, e1005348 (2017).
14. Bethke, L. L. et al. Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of Plasmodium falciparum. Mol. Biochem. Parasitol. 150, 10–24 (2006).
15. Taylor, A. R., Jacob, P. E., Neafsey, D. E. & Buckee, C. O. Estimating relatedness between malaria parasites. Genetics https://doi.org/10.1534/genetics.119.302120 (2019).
16. Rousset, F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145, 1219–1228 (1997).
17. Crow, J. F. & Kimura, M. in An Introduction to Population Genetics Theory Ch. 9.9 (Harper & Row, New York, 1970).
18. Talundzic, E. et al. Molecular epidemiology of Plasmodium falciparum kelch13 mutations in senegal determined by using targeted amplicon deep

NATURE COMMUNICATIONS | (2020) 11:2107 | https://doi.org/10.1038/s41467-020-15779-8 | www.nature.com/naturecommunications

ARTICLE
 sequencing. Antimicrob. Agents Chemother. 61, https://doi.org/10.1128/AAC.02116-16 (2017).

19. Torrentino-Madate, M. et al. Limited polymorphisms in k13 gene in Plasmodium falciparum isolates from Dakar, Senegal in 2012–2013. Malar. J. 13, 472 (2014).

20. Dahliström, S. et al. Diversity of the sarco/endoplasmic reticulum Ca(2+)-ATPase orthologue of Plasmodium falciparum (PiATP6). Infect. Genet. Evol. 8, 340–345 (2008).

21. Lex, A., Gehlenborg, N., Strobl, H., Vuilleminot, R. & Pfister, H. UpSet: Visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. 20, 1983–1992 (2014).

22. Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat. Methods 6, 315–319 (2009).

23. Mvumbi, D. M. et al. Falciparum malaria molecular drug resistance in the Democratic Republic of Congo: a systematic review. Malar. J. 14, 355 (2015).

24. Leroy, D. et al. African isolates show a high proportion of multiple copies of Plasmodium falciparum plasmmeg-2 gene, a piperazine resistance marker. Malar. J. 18, 126 (2019).

25. Nikol Mandoko, P. et al. Prevalence of Plasmodium falciparum parasites resistant to sulfadoxine/pyrimethamine in the Democratic Republic of the Congo: emergence of highly resistant pfdhf/pfdhps alleles. J. Antimicrob. Chemother. 73, 2704–2715 (2018).

26. Baraka, V. et al. Impact of treatment and re-treatment with artmether-lumesfantrine and artemesunate-amodiaquine on selection of Plasmodium falciparum multidrug resistance gene-1 polymorphisms in the Democratic Republic of Congo and Uganda. PLoS One 13, e0191922 (2018).

27. Ruh, E., Batelo, J. P., Imir, T. & Taylan-Ozkan, A. Molecular identification of sulfadoxine-pyrimethamine resistance in malaria infected women who received intermittent preventive treatment in the Democratic Republic of Congo. Malar. J. 17, 17 (2018).

28. Mvumbi, D. M. et al. Molecular surveillance of Plasmodium falciparum resistance to artesinin-based combination therapies in the Democratic Republic of Congo. PLoS One 12, e0179142 (2017).

29. Taylor, S. M. et al. Absence of putative artesinin resistance mutations among Plasmodium falciparum in Sub-Saharan Africa: a molecular epidemiologic study. J. Infect. Dis. 211, 680–688 (2015).

30. World Health Organization. Status Report on Artemisinin and ACT Resistance (WHO, 2017).

31. Malaria GEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. eLife 8, e08714 (2016).

32. Ministère du Plan et Suivi de la Mise en œuvre de la Révolution de la Modernité (MPSMRM), Ministère de la Santé Publique (MSP) et ICF International. Enquête Démographique et de Santé en République Démocratique du Congo (MPSMRM, MSP & ICF, 2014).

33. Pickard, A. L. et al. Resistance to antimalarials in Southeast Asia and genetic polymorphisms in pfmdr1. Antimicrob. Agents Chemother. 47, 2418–2423 (2003).

34. Abouak, B. K. et al. Efficacy of Artemesunate/Amodiaquine in the treatment of uncomplicated malaria among children in Ghana. Am. J. Trop. Med. Hyg. 97, 690–695 (2017).

35. Ngondi, J. M. et al. Surveillance for sulfadoxine-pyrimethamine resistant malaria parasites in the Lake and Southern Zones, Tanzania, using pooling and next-generation sequencing. Malar. J. 16, 236 (2017).

36. Tumwebaze, P. et al. Changing antimalarial drug resistance patterns identified by surveillance at three sites in Uganda. J. Infect. Dis. 215, 631–635 (2017).

37. McKenna, A. et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

38. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinform. 43, 11.10.1–33 (2013).

39. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).

40. The PISK Project. PISK pilot data release 5, https://malariaigen.net/data/pisk-5 (2016).

41. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370 (1984).

42. Malecot, G. The Mathematics of Heredity (W.H. Freeman, 1970).

43. Giorgi, E. & Diggie, P. J. PrevMap: An R package for prevalence apping. J. Stat. Softw. 78, https://doi.org/10.18637/jss.v078.i08 (2017).

44. Gautier, M., Klassmann, A. & Vitalis, R. reh: 2.0: a reimplemention of the R package reh to detect positive selection from haplotype structure. Mol. Ecol. Resour. 17, 78–90 (2017).

45. Gautier, M. & Vitalis, R. reh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177 (2012).

46. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer, New York, 2009).

Acknowledgements

This work was supported by the National Institutes of Health (R01AI107949, R01AI1139520, K24AI134990, R21AI1121465, F30AI143172, U19AI089680). R.V is funded by a Skills Development Fellowship: this award is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP programme supported by the European Union. We would also like to thank everyone who participated in the studies and all members of the study teams in Ghana, Zambia, Uganda, and Tanzania. We would like to thank the DHS Program and USAID for the collection and access to samples from the DRC Demographic Health Survey.

Author contributions

R.V., O.A., N.F.B., J.A.B., and J.J.I contributed to data analysis, writing and experimental design. O.J.W contributed to data analysis and writing. N.I.H. and A.P.M contributed software design. M.K.M, J.B.P., P.K.T, M.C., P.J.R., D.S.I., J.N., J.G., M.M., D.E.N., W.I.M., B.A.M., J.L.M.H., A.G., and A.K.T contributed samples from studies conducted at their sites and reviewed the paper. A.C.G. contributed to analysis design and reviewed the manuscript. P.W.M., K.T., T.F., and M.D contributed laboratory analysis. S.R.M. contributed coordination with DRC investigators, experimental design, and writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-15779-8.

Correspondence and requests for materials should be addressed to R.V. or J.J.I.

Peer review information Nature Communications thanks Roberto Amato and the other anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2020