Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Potential of Short Tandem Repeat (STR) Polymorphisms to Fill the Missing Inheritance Gap of Artemisinin Partial Resistance of Human Malaria Parasites

Zbynek Bozdech

zbozdech@ntu.edu.sg

Nanynag Technological University https://orcid.org/0000-0002-9830-8446

Sourav Nayak

Nanynag Technological University https://orcid.org/0000-0002-7659-5712

Michal Kucharski

Nanynag Technological University https://orcid.org/0000-0002-1505-5608

Tien Quang Huy Duong

Nanynag Technological University https://orcid.org/0009-0004-6051-3200

Xavier Roca

Nanyang Technological University https://orcid.org/0000-0003-3853-5273

Thomas Peto

Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University
https://orcid.org/0000-0003-3197-9891

Rupam Tripura

Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University

James Callery

Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University
https://orcid.org/0000-0002-3218-2166

Lek Dysoley

National Center for Parasitology, Entomology and Malaria Control

Nghia Ho

6Oxford University Clinical Research Unit, Hospital for Tropical Diseases,

Rob van der Pluijm

Mahidol Oxford Tropical Medicine Research Unit

Dong Nguyen

10Khanh Hoa Hospital for Tropical diseases, Khanh Hoa province

Ranitha Vongpromek

Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University

Huy Rekol
4Centre for Parasitology, Entomology and Malaria Control,
Hoang Chau Nguyen
6Oxford University Clinical Research Unit, Hospital for Tropical Diseases https://orcid.org/0000-0002-7826-3850
Olivo Miotto
Mahidol-Oxford Tropical Medicine Research Unit https://orcid.org/0000-0001-8060-6771
Mavuto Mukaka
Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University
Lorenz von Seidlein
Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University
https://orcid.org/0000-0002-0282-6469
Mallika Imwong
Faculty of Tropical Medicine, Mahidol University
Nicholas P. Day
University of Oxford
Nicholas White
University of Oxford https://orcid.org/0000-0002-1897-1978
Arjen Dondorp
Mahidol-Oxford Tropical Medicine Research Unit https://orcid.org/0000-0001-5190-2395

Article

Keywords:

Posted Date: August 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10598830/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

The Potential of Short Tandem Repeat (STR) Polymorphisms to Fill the Missing Inheritance Gap of Artemisinin Partial Resistance of Human Malaria Parasites.

Sourav Nayak[1#], Michal Kucharski[1,14#], Duong Tien Quang Huy[1], Thomas J Peto[2,3], Rupam Tripura[2,3], James J Callery[2,3], Dysoley Lek[4,5], Ho Dang Trung Nghia[6,7], Rob W van der Pluijm[2,3,8,9], Nguyen Dong[10], Le Thanh Long[11], Ranitha Vongpromek[2,12], Huy Rekol[4], Nguyen Hoang Chau[6], Olivo Miotto[2,3], Mavuto Mukaka[2,3], Lorenz von Seidlein[2,3], Mallika Imwong[2,13], Nicholas PJ Day[2,3], Nicholas J White[2,3], Xavier Roca[1], Arjen M Dondorp[2,3,8], Zbynek Bozdech[1,2,3*]

# authors contributed equally

* Corresponding author (ZB) zbozdech@ntu.edu.sg

[1] School of Biological Sciences, Nanyang Technological University, Singapore.

[2] Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.

[3] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[4] Centre for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia

[5] National Institute for Public Health, Phnom Penh, Cambodia

[6] Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

[7] Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam

[8] Center of Tropical Medicine and Travel Medicine, Department of Infectious Diseases, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

[9] Institut Pasteur, Université Paris Cité, G5 Infectious Disease Epidemiology and Analytics, Paris, France.

[10] Khanh Hoa Hospital for Tropical diseases, Khanh Hoa province, Vietnam

[11] Phuoc Long Hospital, Binh Phuoc province, Vietnam

[12] WorldWide Antimalarial Resistance Network - Asia-Pacific Regional Centre, Bangkok, Thailand.

[13] Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

[14] Amsterdam UMC, University of Amsterdam, Department of Global Health, Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands.

Abstract

It is becoming clear that the Artemisinin Partial Resistance (ART-R) phenotype of *Plasmodium falciparum* is a complex genetic trait underlying a multifaceted molecular mechanism. Here, we investigated short tandem repeat (STR) polymorphisms to identify sequence polymorphisms in putative DNA regulatory elements that could drive transcriptional variation and contribute to *P. falciparum* phenotypic plasticity, including ART-R. We identified a broad spectrum of polymorphic STR (pSTR) genotypes in the whole-genome sequence dataset from a cohort of *P. falciparum* natural-infection parasite samples. These included microsatellite-like sequence-length variations in A/T-rich, low-complexity sequences, located in the *P. falciparum* gene's upstream regions, untranslated regions UTRs, introns, and coding sequences. Using expression quantitative trait loci, we identified 220 of such pSTRs that associate with transcriptional levels of their linked genes (e.g., *cis*-eSTR). Interestingly, we observed a highly significant overlap between the *cis*-eSTRs and SNP-based *cis*-eQTLs, indicating their ancestral origin. Crucially, genetic polymorphisms at 85 of the identified *cis*-eSTRs showed a significant association with partial artemisinin resistance, as determined in the TACT-CV parasite cohort by parasite clearance half-life (PC1/2) > 5hr (AR-eSTR). 18 of these cis-eSTR-linked genes also exhibited differential transcript levels that were associated with artemisinin resistance. These included a hemoglobin digestion-mediating cysteine protease, *falcipain 2b,* and a ribosomal assembly factor, MAK16, suggesting their validity as novel putative markers of ART-R. Our results also indicate that, like in higher eukaryotes, pSTRs could represent the "missing heritability" of key genetic traits of *P. falciparum* parasites, presumably via their activity in regulating gene expression.

Introduction

The declining efficacy of artemisinin combination therapies (ACTs) to treat uncomplicated infections of the deadliest malaria parasite, *Plasmodium falciparum,* poses a major public health threat[1]. First reported in the Western Cambodia province of Pailin in 2009, the compromised efficacy of ACT manifested as slower clearance of parasitemia in the early stages of treatment, as reflected by an increased parasite clearance half-life ($PC^{1/2}$ > 5 hours)[2]. The prolonged persistence of *P. falciparum* parasites in the patient's blood, resulting from increased $PC^{1/2,}$ became a well-defined phenotype that subsequently spread throughout the Greater Mekong Subregion (GMS) over the following half-decade[3] and has since completely dominated *P. falciparum* parasite populations within the region through a series of selective sweeps[4]. The phenotype manifests as delayed clearance without complete treatment failure, reflecting stage-specific ring survival rather than loss of drug efficacy across the life cycle — and was therefore termed partial resistance ART-R by the WHO. It is defined by significant increases in $PC^{1/2}$ of *P. falciparum* infections after administration of ACTs, leading to the persistence of detectable parasitemia past 3 days of the treatment course[5]. ART-R is believed to promote the emergence of resistant alleles to the non-artemisinin components of the ACT (partner drugs), which ultimately leads to their failures[6]. Indeed, ACT treatment failures resulting from the spread of ART-R in the GMS are now well documented[7,8]. Following Southeast Asia, ART-R has emerged in South America[9], Papua New Guinea[10], and, most importantly, Sub-Saharan Africa[11,12], which is the region with the highest malaria burden. This is a highly worrisome development, stressing the need for constant improvement of all malaria control program strategies, which is, amongst many things, based on a comprehensive understanding of *P. falciparum* parasite pathophysiological aspects of ART-R.

Through genome-wide association studies (GWAS) and *in vitro* experiments, nonsynonymous mutations in Pfkelch13 (PfK13) have been firmly established as molecular markers of artemisinin resistance[3,13,14]. In particular, non-synonymous Single Nucleotide Polymorphisms (SNPs) in the propeller KELCH domain of PfK13 have been shown to associate with ART-R in a mutually exclusive manner with high statistical confidence and are now the most informative markers of its emergence and

spread[5]. *In vitro*, amino acid changes within the PfK13 propeller domain cause resistance in the *P. falciparum* parasite in the very early stages of the intraerythrocytic developmental cycle (IDC, 0-3hr)[14,15], presumably by reducing endocytosis of hemoglobin and thus drug activation[16]. PfK13 SNPs, however, also cause broad gene expression rewiring with profound changes in the parasite's cellular physiology that could have a profound effect on drug sensitivities of other IDC stages that are currently uncharacterized[17]. Although a number of PfK13nsSNPs have been highly informative in studies of ART-R, since their discovery, it has been clear that their penetrance for both *in vitro* (ring stage survival) and *in vivo* (PC1/2 >5hr) phenotypes never reaches 100%[18]. Measured by the Ring Survival Assay (RSA)[19], the PfK13-mediated *in vitro* resistance of *P. falciparum* varied markedly among parasite strains[20]. This has been largely attributed to the need for a genetic background comprising genetic variants that either directly contribute to the resistance mechanism (in a PfK13-dependent or -independent manner) or alleviate the fitness costs associated with PfK13 SNPs and other ART-R-mediating polymorphisms[18]. Indeed, there is a growing list of PfK13 co-variants identified by population genetic studies of ART-R over the last two decades, ranging from the ~5 genetic regions of ART-R architectures identified by Miotto, O. et al (2015) in the early stages of ART-R emergence in the GMS[13] to SNP variants on chromosome 11 (PfWD11) and chromosome 13, in the vicinity of PfK13, (PfRad5 and PfVPS13) persisting with ART-R in GMS throughout the last decade[21]. Moreover, the ART-R *P. falciparum* parasites exhibit a specific transcriptional reprogramming while in their natural infections, involving factors of proteostasis, oxidative stress response, unfolded protein response signalling, and cell-cycle modulation[17,22-24]. Later termed the Artemisinin resistance-associated transcriptional profile (ARTP), this reprogramming contributes to the resistance mechanism by altering the levels of specific genes that act epistatically with PfK13 SNPs, thereby constituting *the* physiological state of *P. falciparum* resistant to artemisinin[25]. Although the exact molecular mechanism mediating this transcriptional reprogramming is currently unknown, there is mounting evidence that, in at least some cases, it is driven by genetic polymorphisms at gene expression DNA regulatory elements. This was demonstrated by Kucharski, M. et al (2023), detecting a sequence length polymorphism at a short tandem repeat (STR) within the promoter region of *P. falciparum* cyclophilin 19B (PfCYP19B), which led to tolerance to artemisinin (and piperaquine) by increased *pfcyp19b* transcript levels[26]. These observations raised the

possibility that, similar to other eukaryotic systems, including human[27], polymorphic STRs (pSTRs) modulate gene expression in *P. falciparum* parasites, conferring substantial phenotypic plasticity and, as such, contributing to the ART-R genetic background.

In this study, we systematically explored the genome-wide distribution of pSTRs in a *P. falciparum* population during natural infections and subsequently examined their associations with gene expression and ART-R. By integrating high-resolution STR genotyping, expression Short Tandem Repeat (eSTR) mapping, expression Quantitative Trait Locus (eQTL) mapping, and pSTR-PC$^{1/2}$ association, we identify eSTRs associated with *in vivo* transcriptional modulation and STR markers for ART-R. These findings expand the current framework of resistance genetics beyond canonical coding mutations and highlight repetitive variation as a potential contributor to the adaptive landscape of malaria parasites, helping explain components of phenotypic variance not captured by single-nucleotide variants alone.

Results

Polymorphic STRs in *Plasmodium falciparum* isolates from the TACT-CV cohort.

In the first step of the study, we explored pSTR within the TACT-CV *P. falciparum* cohort, collected across two regions of Cambodia and Vietnam between 2018 and 2020, which represented a peak in the spread of artemisinin resistance in the Greater Mekong Subregion (GMS)[28]. For this purpose, we adapted the bioinformatics pipeline previously described by Han et al. (2022), with modifications suited to the nature of the available TACT-CV genomic data{Han, 2022 #17}. In the first step, we defined 561,342 loci with STRs of any length present within the 14 chromosomes of the reference 3D7 *P. falciparum genome* **[Supp. Figure S1].** 388,442 of these STR loci were represented by 1 to 6 bp repeats, with at least 3 repeats identified by an alignment score of 10 or higher and a maximum of 10% indels (**Materials and Methods**). In this intermediate dataset, mononucleotide STRs were the highest in numbers (n = 181,560), contributing to ~47% of total STR loci, followed by dinucleotide repeats with ~ 34% (n = 131,495) **[Supp. Figure S2]**. This collection of 388,442 STRs subsequently served as a template dataset for mapping of SRT regions across the whole-genome sequence (WGS) of the TACT-CV cohort to classify their

polymorphisms. Namely, we utilized WGS for 239 isolates that exhibited a minimum genome coverage of 10X and a mean genome coverage of 151X, which met the minimum criteria for confident STR detection and polymorphism classification **[Supp. Figure S3]**. Overall, we identified a total of 86,030 loci (~22% of the investigated STRs) with polymorphic STR lengths (pSTRs) within the core genome[30] of the TACT-CV parasite cohort. Initially, an STR locus was considered polymorphic if at least one isolate carried a different STR length compared to reference 3D7, and/or differed from other isolates. Next, we considered only those pSTRs whose polymorphism could be confidently inferred in at least 70% of the 239 isolate samples. As a result, we could define a final set of 22,072 pSTRs that could be called in as many as 222 isolates, which we used as the foundation for the subsequent analysis. In contrast to the genomic representation of all STRs, where mononucleotide repeats dominated, the highest proportion of pSTRs is represented by dinucleotides (~64%, n=14193), followed by trinucleotides (~14%, n=3171), with mononucleotide repeats being in the minority (~11%, n=2361) [**Figure 1a**]. This distribution could reflect the origin of sequence polymorphisms, generated by DNA replication "slippage" caused by stem-loop structures resulting from DNA strand unwinding, which is unlikely to form with mononucleotide repeats (see below).

There was a significant enrichment of the pSTRs near protein-coding genes, with a distinct peak at transcriptional start sites (TSS) [**Figure 1b**]. Specifically, 4,709 (21%) and 12,012 (54%) of the pSTRs were located within 500 bp and 2 kb upstream/downstream of the TSS,respectively, with an even distribution within +/-500bp around TSS, dominated by dinucleotide repeats [**Figure 1b,** *inset*]. From the within-transcript positioning classifications, 56% (n=12,361) of total pSTRs were found in 5'UTRs or upstream intergenic STRs [**Figure 1c**]. A somewhat smaller number of pSTRs were also found within 3'UTRs, introns, and coding exons. pSTRs located outside of the protein-coding genic regions are enriched by dinucleotide repeats, whereas protein-coding pSTRs are predominantly represented by trinucleotide repeats (83%, n=2761 out of 3321 coding pSTRs). The 22,072 pSTR loci in the entire cohort of TACT-CV isolates (239), ultimately defined ~1.5 million deletions and 1.1 million insertions compared to the 3D7 reference genome [**Figure 1d**]. Even though this represents a slight skew, there was no bias in the size distribution towards either deletion or insertion. Notwithstanding the distinct (African) origin of the 3D7 *P.*

falciparum strain, placing this reference genome in the middle of the distribution could reflect its prolonged *in vitro* culturing that eliminated STR sequence polymorphism variations due to the lack of external adaptation pressures (*author's note*). Investigating the allelic diversity at each pSTR locus revealed their polyallelic character [**Figure 1e**]. The highest diversity was found among dinucleotide pSTRs, with a median of 4 alleles per locus. This was followed by the tri-, tetra-, and penta-nucleotide pSTRs, with mono-nucleotide and hexa-nucleotide pSTRs, which are the lowest in allelic diversity. This indicates that di-nucleotides represent the most variable pSTR class, being highly abundant in essentially all noncoding parts of the *P. falciparum* genome.

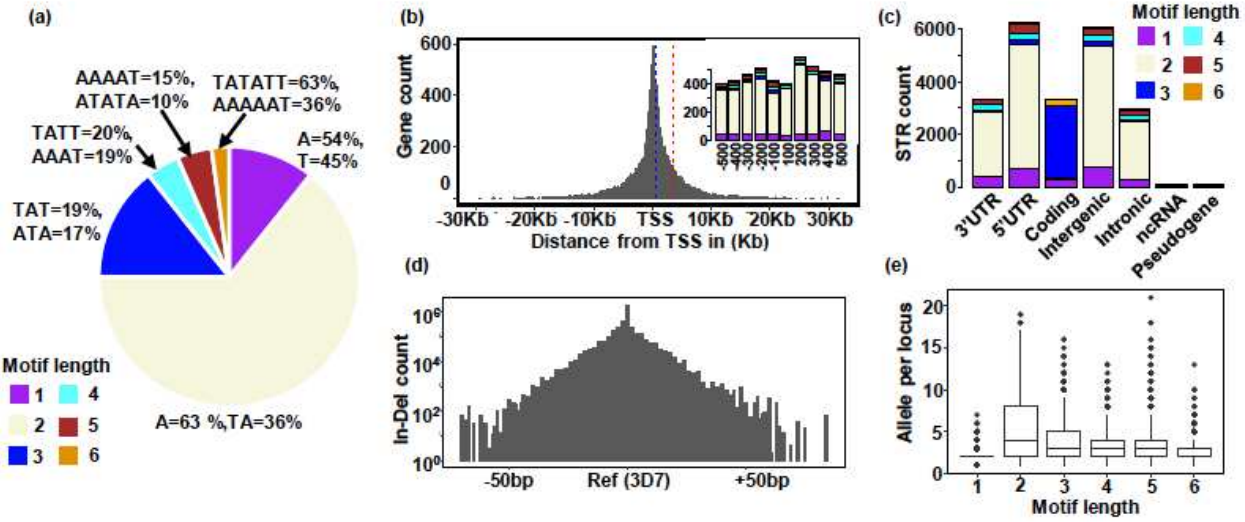


Figure 1: Polymorphic STRs from TACT-CV parasites: **(a)** Distribution of polymorphic STRs (pSTRs) according to their motif length detected from the TACT-CV parasite cohort. For each group, the highest two representative sequences have been shown along with their within-group percentage. Dinucleotide motifs contribute the highest proportion of pSTRs, followed by trinucleotide and mononucleotide motifs. **(b)** The distance between the STRs from the nearest ' Transcription Start Site (TSS) is shown in the form of a barplot. Each line represents 100 bp bin. The blue vertical line represents the mean UTR length, whereas the red vertical line mean gene length. The A zoom view, along with the motif length distribution for each binned are are shown for +/-500bp of TSS. **(c)** Count of pSTRs for each genomic region along with motif length distribution. Clearly, pSTRs detected within the non-coding regions are predominantly represented by di-nucleotide motifs, and pSTRs from coding regions are represented by tri-nucleotide motifs. **(d)** The pSTRs can alternatively be defined insertion-deletion events compared to the reference genome. This barplot shows the length distribution inserted or deleted sequence length compared to the 3D7 reference. **(e)** This boxplot shows the multi-allelic nature of pSTRs for each group of STR length. The X-axis represents groups of pSTRs based on their motif length, and the y-axis represents the number of alleles for each pSTR.

eSTR analysis found a subset of pSTRs associated with gene expression

Given the enrichment around the TSS that we identified in the TACT-CV cohort **[Figure 1b]**, it is reasonable to speculate that, similar to higher eukaryotes[27], pSTRs

have the capacity to modulate gene expression in *P. falciparum*. To investigate this, we applied the AssociaTR algorithm, which detects associations between sequence length variations in pSTRs and phenotypes (e.g., transcript levels) while mitigating the effects of population structure and transcriptomic variation by incorporating principal components into the model[31]. Using a minor allele frequency >0.05 as an inclusion criterion, we carried out association studies between 10,538 pSTRs and mRNA abundance levels of 2,740 *P. falciparum* genes that were previously established in our RNA-Seq study[21]. Out of the 8,828 associations (p-values < 1e-05) [**Supp. Table 1**], we selected 1,351 eSTRs associated with mRNA abundance levels of 249 genes, using a genome-wide significant p-value threshold of 4e-08, determined as 1/(number of LD-pruned STRs * number of genes from expression data) (materials and methods) [**Figure 2a**]. At this threshold, we identified 220 (~16%) eSTRs that could be classified as participating in *cis*-interactions, comprising eSTRs within the “gene boundary” regions of 98 unique genes [**Figure 2a & b**]. There were an additional 1,131 (84%) eSTRs with interactions involving 220 unique genes found across all 14 chromosomes, with their respective *trans*-eSRTs lying outside the gene boundaries, which included 58 (~60%) of the unique *cis*-STR **[Supp. Table S1]**. Crucially, the within-gene boundary distribution of the *cis*-eSTR is signified by the distribution peak around the gene’s TSS **[Figure 2b],** which is also reflective of the overall pSTR distribution identified in the representation of *P. falciparum* WGS within the TACT-CV cohort **[Figure 1b]**.

The 220 *cis*-eSTRs participated in as many as 304 interactions, including 120 and 78, positioned in up- and downstream intergenic regions, respectively **[Figure 2c]**. In addition, we identified 44, 16, 14, and 10 *cis*-eSTRs located in 5’UTRs, introns, coding regions, and 3’UTRs, respectively. Note: due to incopled annotations of the P. falciparum genomes, some of the *cis*-eSTRs participate in more than one interaction, being positioned in individual regions of at least two genes equally correlated in their expression **[Supp. Table S2].** Although the numbers of cis-eSTRs within the gene bodies were comparatively lower, their presence also indicates their potential to affect their respective steady-state mRNA levels. According to the expectation, eSTRs in the coding regions were dominated by tri-nucleotide repeats, while all other *cis*-acting eSTRs involved mainly di-nucleotides with only a minority representation of other types of pSTRs to up to 6-mer repeats [**Figure 2c]**. **Figure 2d** presents several notable

examples of the eSTRs depicting the linear regression model by which eSTRs were characterized. These include cysteine proteinase *falcipain2a* (PF3D7_115700), whose mRNA abundance correlated with a mononucleotide bi-allelic pSTR variation (8/9T) located 131 bp upstream of its TSS. On the other hand, di-nucleotide tri-allelic (-8/0/+8/AT) pSTR in the 5'UTR of ubiquitin carboxyl-terminal hydrolase 1 (UBP1, PF3D7_0104300); tri-nucleotide tri-allelic (-39/-15/0/TCA) in the first exon of a nucleolar preribosomal assembly protein (MAK16, PF3D7_0203700); mono-nucleotide bi-allelic (-1/0/T) in the intron of a flippase (ATP2, PF3D7_1219600) and di-nucleotide bi-allelic (0/=1/AT) in the 3'UTR of the FHA domain protein (PF3D7_1133700) all correlated with the abundance of their respective mRNA. Finally, the di-nucleotide tri-allelic (0/2/4/TA) represented a significant eSTR located in the downstream intergenic region of the erythrocyte binding antigen (EBA175, PF3D7_0731500). Beyond these examples, a functional analysis of the differentially expressed genes linked to the *cis*-eSTR revealed a narrow set of putative biological roles for these interactions in parasite adaptations during natural infections [**Figure 2e**]. These included factors involved in protein metabolism, such as ribosomal assembly and ribosomal structure, as well as protein phosphorylation, folding, degradation, and transport. In addition, protein kinases and acidocalcisome cofactors were represented among the *cis*-eSTR-interacting genes. Notwithstanding the need for further validation, it is tempting to speculate that these pathways play key adaptive roles for *P. falciparum* in its natural host, mediated by eSTR-causal associations that modulate altered gene expression levels.

Local eSTR-eQTL overlap revealed a complex expression regulation landscape.

In our earlier study, we explored SNP-based expression quantitative trait loci (eQTL) in a cohort of *P. falciparum* infections at the early onset of artemisinin resistance in the Greater Mekong Subregion (GMS), collected during the Tracking Artemisinin Resistance Collaboration (TRAC) between 2009 and 2011[32]**.** We demonstrated that putative cis-acting eQTLs define distinct regions of the GMS, suggesting their role in recent genetic diversification, possibly through gene expression regulation. Here, we sought to explore this concept by comparing the identified local eSTR with SNP-based local eQTLs (further eQTLs) in the TACT-CV cohort. Considering SNPs in the WGS of the TACT-CV cohort filtered based on mapping quality (MQ=60), detection rate (70% or higher), and minor allele frequency

(MAF=0.05) we carried out an eQTL study with 2,740 genes for which mRNA abundance was characterized by RNASeq analysis in the same sample set[21]. Overall, we identified 51790 SNP-based eQTL interactions, of which 51,119 and 671 were *trans*- and *cis*-acting, respectively **[Supp. Figure S4]**. The latter fell within the gene boundaries and were detected with the genome-wide significance threshold of 1.7e-05, involving 644 SNPs linked to 374 genes, respectively. Next, we wished to compare the set of 671 *cis*-eQTLs with the 304 *cis*-eSTRs in terms of their relative positional

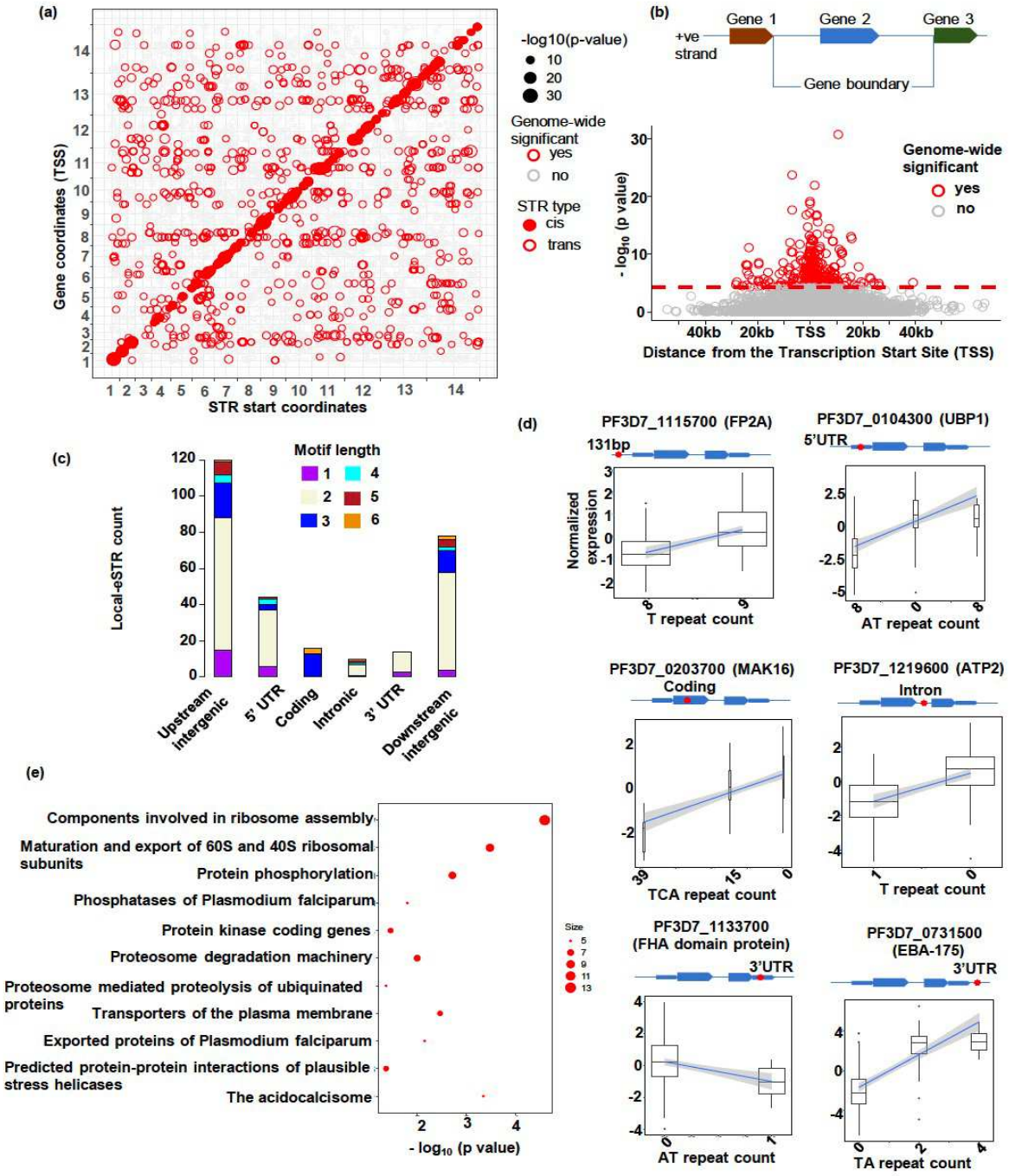


Figure 2: Expression-associated eSTRs: **(a)** Cis-trans plot showing pSTRs associated with gene expression. In this plot, the X-axis represents the genomic coordinates of STRs, and the y-axis represents the transcription start site of the genes investigated for association. Here, the red colored circles denote the genome-wide significant association, and the grey-colored circles denote association with p-value < 1e-05. Filled circles are STRs associated through local interactions, while hollow circles are trans-eSTRs. The size of the circles increases with higher association. **(b)** The top panel shows a schematic of the gene boundary or local interactions definition. The panel below shows all the local eSTRs. Here, the x-axis denotes the distance between the local e-STRs and the transcription start site, and the y-axis denotes the p-value of association. Clearly, the most significant local eSTRs are clustered around TSS. **(c)** This barplot shows genomic locations of local-eSTRs and their corresponding counts. Further, each bar is stacked with different groups of STRs based on motif size **(d).** Representative examples of local-eSTRs located at different genomic regions. Examples are shown based on STR locations from left to right according to genomic context, i.e., upstream intergenic, 5'U, coding, intron, 3'U, and downstream intergenic. In each boxplot, the y-axis shows the count of repeat gain or loss, and y-axis represents expressions of genes normalised for age. The blue line drawn for each plot shows the slope, and the grey area shows confidence intervals. Top of each barplot a schematic of the gene, along with the corresponding position of the STR is marked by red star (not to be scaled). **(e)** Gene found involved in local eSTRs (n=156 unique genes) subjected to pathways enrichment analysis. This bubble plot shows any pathway enriched with at least 5 genes from the local-eSTR-associated genes. In this plot, the x-axis shows p-values derived from the hypergeometric test, and the y-axis shows the name of the pathway defined by Malaria Parasite Metabolic Pathways.

enrichments within their corresponding loci/gene boundaries. Indeed, there was a near-perfect overlap, with 97 out of the 98 cis-eSTRs-linked genes also containing at least one eQTL **[Figure 3a, Supp. Table S3]**. This overlap signifies similar origins and/or functional relevance of the two mutation types, and their putative linkage to the characteristic gene expression within each gene locus. This model is further supported by the fact that most of these genes contained more than one eSTR or eQTL, with the "lead eSTR" exhibiting the strongest correlation with the corresponding mRNA abundance **[Figure 3b]**. Out of the 97 eSTR-linked gene loci, 30 contained more than three eSTRs, and of these, 19 also contained more than three eQTLs [**Figure 3c]**. On the other hand, 24 loci contained at least three eQTLs, but fewer than three eSTRs **[Figure 3d]**, and finally, 43 loci carried fewer than three of both. This represents highly statistically significant results (p-value 0.000026), suggesting that cis-acting genetic polymorphisms at both STRs and SNPs are under selection pressures in *P. falciparum* during natural infections by regulating the corresponding gene expression.

Focusing on the group of 19 genes with >3eSTRs and eQTLs **[Figure 3c]**, the top example of the genes with> 3eSTRs is ATP-dependent RNA helicase (DDX60 PF3D7_0903400), which contained 10 *cis*-eSTRs and 34 *cis*-eQTLs **[Figure 3e]**. Although the lead eSTR (marked by a grey box) is in the downstream intergenic region, most of the significant SNPs are in the upstream intergenic region. All the significant

associated eSTRs/eQTLs are in mid (R^2>0.2) to high (R^2>0.5) linkage disequilibrium (LD) with the lead STRs, signifying their co-segregation within this parasite population. The contrasting example is PHISTb (PF3D7_0424600), a factor of host cytoplasm remodeling, with 6 *cis*-eSTRs and 25 *cis*-eQTLs all within the gene boundary, with the majority of eQTLs clustered with the lead sSTR at the 3' intergenic region **[Figure 3f]**. Here, the LD values calculated between lead eSTRs and the rest of the eSTRs and SNPs are within mid to high levels. Within the 19 highly linked genes, we observed highly significant LD for eQTL-associated SNPs associated with eSTR/eQTLs, which exhibit significantly higher-level LD with the lead eSTR compared to the rest of the

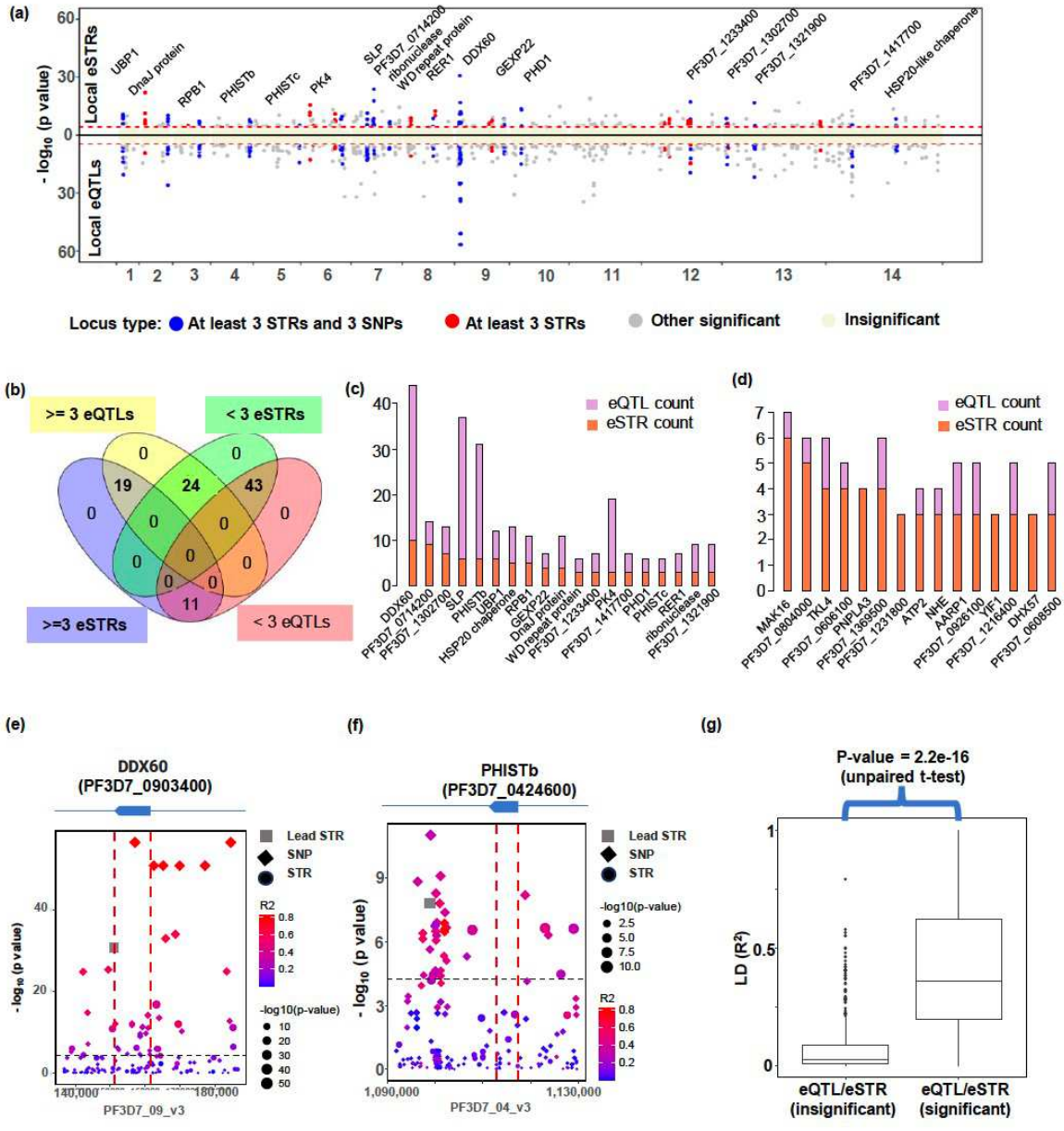


Figure 3: *Cis*-eSTR-eQTL overlap: **(a)** Miami plot shows local-eSTR and eQTLs. The upper part of the plot shows local-eSTRs associated, whereas the bottom part represents local eQTLs. For local-eSTRs, circles above the red line signify, and for local-eQTLs, circles below the the redline denotes genome-wide significant association. Blue colored and red colored circles denote individual gene-based loci, where blue color denotes the genes found to be associated with at least 3 STRs and 3 SNPs, and red color denotes genes found to be associated with at least 3 STRs and no or fewer than 3 SNPs. Genes associated with at least 3 STRs and 3 SNPs are also labeled either by their short names or by their IDs if unannotated. **(b)** 97 Genes associated with both local eSTR and eQTLs are categorized into different classes according to the number of associations found with SNPs and STRs. The primary aim of this figure is to highlight strong expression-associated loci, either defined by STRs alone or SNPs alone or together with STRs and SNPs. **(c)** Bar plot showing the short names of the genes or their IDs (if unannotated) in the x-axis (n=19), the number of STRs (at least 3 STRs), and the number of SNPs (at least 3 SNPs) associated with their expression. Essentially, these are the strongest gene-based expression-associated loci, as these are supported by both STRs and SNPs. **(d)** Similar barplot showing genes found to be associated with at least 3 STRs and < 3 SNPs, thus primarily supported by STRs. In this plot, the x-axis shows (n=14) short names of the genes or their IDs (if unannotated), and the y-axis shows the count of eSTRs and eQTLs. **(e)** A detailed view of an example locus involving the gene ATP-dependent RNA helicase (PF3D7_0903400). The loci were found in both the local-eSTR (n=10) and local-eQTL (n=34) analyses. This figure shows all SNP and STR variations detected within the gene boundary and the p-value of association either by eSTR or eQTL analysis. The X-axis of this plot shows genomic coordinates of the STRs or SNPs, and the y-axis denotes p-values of association. Here, the vertical line shows the gene area, and the horizontal line denotes genome-wide significance. The shape of the points shows the type of variants, where the grey box denotes lead-STRs, circles denote STRs and diamonds denote SNPs. The color of the points is based on their LD R2 value with lead STR. **(f)** The second example involves another locus involving gene PHISTb (PF3D7_0424600). The locus is defined by local-eSTR (n=6) and local-eQTL (n=25) This figure shows all SNP and STR variations detected within the gene boundary and the p-value of association either by eSTR or eQTL analysis. The X-axis of this plot shows genomic coordinates of the STRs or SNPs, and the y-axis denotes p-values of association. Here, the vertical line shows the gene area, and the horizontal line denotes genome-wide significance. The shape of the points shows the type of variants, where the grey box denotes lead-STRs, circles denote STRs, and diamonds denote SNPs. The color of the points is based on their LD R2 value with lead STR. **(g)** This box plot is drawn to compare the LD R2 values with lead STR between two groups of variants, significant eQTLs/eSTRs and insignificant eQTLs/eSTRs. Essentially, the groups differ based on an unpaired t-test, significantly showing expression associated-eQTL/eSTRs have higher values of LD compared to eQTL/eSTRs with no/low association with expression.

SNPs also identified in each locus but not associated with the lead eSTR or mRNA levels; tested by unpaired t-test (p-value < 2.2e-16) **[Figure 3g]**. Indeed, complex structures in eSTR/eQTL positioning can be observed in 97 identified loci, each, however, characterized by statistically significant LD for eQTL-SNPs exclusively **[Supp. Figure S5].** These results suggest that the identified *cis*-eSTR polymorphisms are stable within the TACT-CV, reflecting a selection process presumably directed by the altered gene expression they mediate.

Associating STR genotypes with artemisinin resistance (strGWAS)

As mentioned above, the TACT-CV epidemiological survey was designed to study the efficacy of new formulations of triple artemisinin combination therapies

(TACT)[28]. As such, the efficacy of the artemisinin component of the TACT was characterized for each cohort sample by measuring the parasite clearance half-life ($PC^{1/2}$), a key parameter for defining ARP ($PC^{1/2}$ >5 hours)[5]. In our earlier study, we explored the association of SNPs with $PC^{1/2,}$ confirming the role of PfK13, as well as several additional genetic SNPs, in the spread and maintenance of partial artemisinin resistance in the GMS[21]. Having established pSTR and a new set of polymorphisms, we here evaluated their associations with partial artemisinin resistance. For that, we again used AssociaTR with parasite $PC^{1/2}$, as a continuous variable incorporating 10 principal components of the derived pSTR for population stratification (materials and methods). A total of 10,538 pSTRs (maf > 0.05) from 220 isolates were correlated with $PC^{1/2,}$ which led to identifying 85 pSTRs with significant associations (p-value < 9.5e-05) **[Figure 4a, Supp. Table S3]**. These 85 artemisinin resistance-associated pSTRs (AR-pSTRs) appear to be distributed across 12 of the 14 *P. falciparum* chromosomes (excluding chromosomes 3 and 7), suggesting a broad spectrum of genetic determinants contributing to ART-R. Again, there was a peak distribution of the AR-pSTRs around gene TSS, with enrichment within the first 1kb bin upstream and downstream (n=48, 57%) **[Figure 4b]**. Specifically, 56 (~65%) and 31 (~36%) of AR-pSTRs are located within 10kb up and downstream of TSS. Based on their proximity to the TSS, the 85 identified AR-pSTRs can be linked to 72 genes **[Figure 4c]**. In these, ~90% (n=79) are in the non-coding regions of the linked genes, either in upstream promoter regions or within UTRs and introns. This suggests a putative role in transcription or mRNA stability of these genes, which can be altered by their polymorphisms.

Figure 4a Manhattan plot depicts at least five AR-pSTR clusters at 2, 11, 12, and 13 (2), the latter two of which are linked but not directly overlapping the *pfk13* gene. Indeed, the chromosome 13 cluster was the target of the largest AR-STR aggregation, further indicating that selection at this genetic locus was driven largely, though possibly not entirely, by the functional relevance of the PfK13 SNPs. Although absent in pfk13, at least two AR-pSTR were observed in neighboring PfRAD5 (PF3D7_1343400) and PfVPS13 (PF3D7_1343800), being most significantly associated with $PC^{1/2}$ within the locus **[Figure 4d]**. The AR-pSTR of PfRad5 involved a dinucleotide repeat in its promoter regions, which could be consistent with its role in the gene's transcription. The PfVPS13-linked ARpSTR involved a tri-nucleotide expansion within the coding

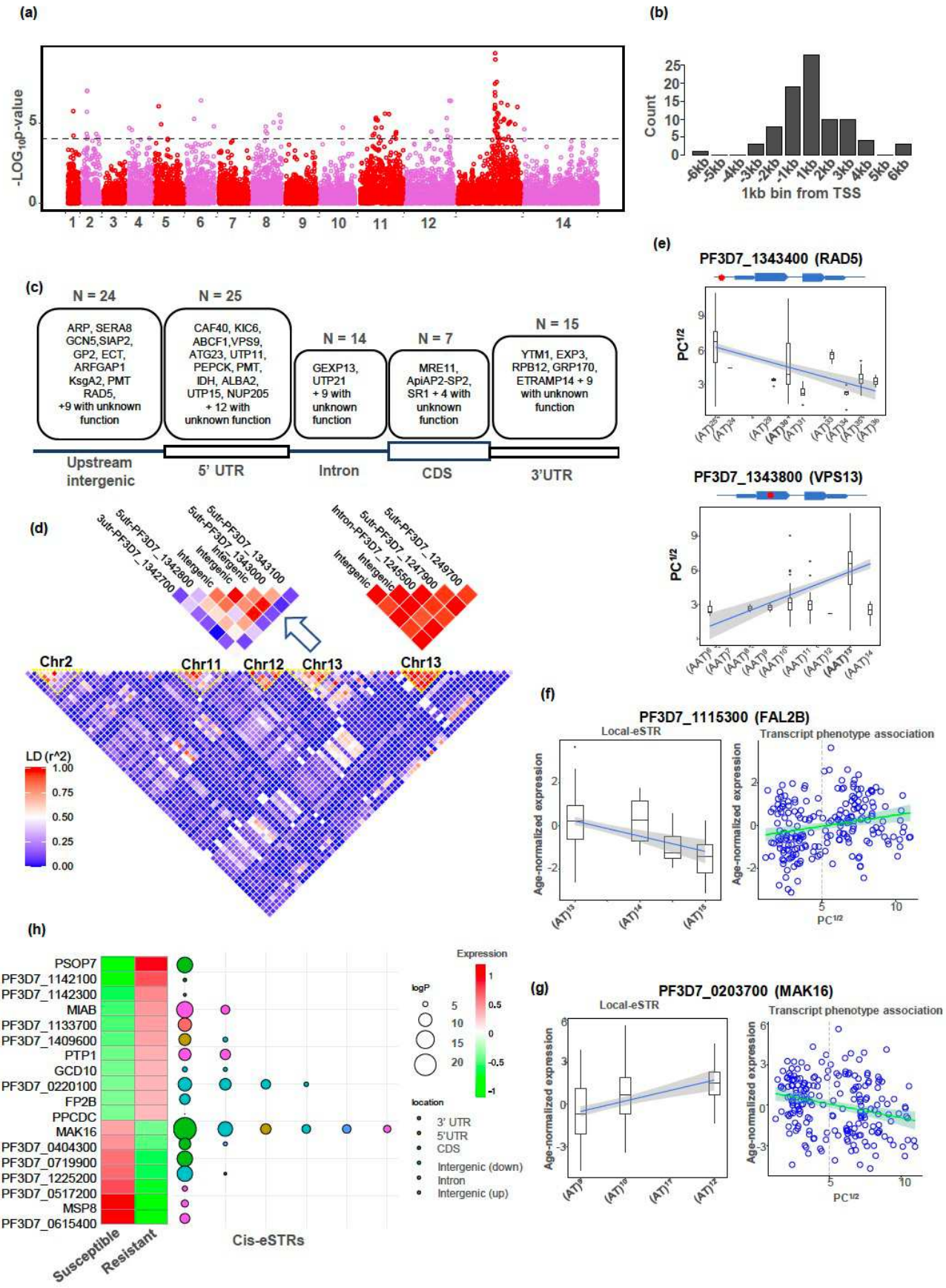

(a)
-LOG10p-value
1 2 3 4 5 6 7 8 9 10 11 12 14
(b)
Count
1kb bin from TSS
(c)
N = 24
N = 25
N = 14
N = 7
N = 15
ARP, SERA8 GCN5,SIAP2, GP2, ECT, ARFGAP1 KsgA2, PMT RAD5, +9 with unknown function
CAF40, KIC6, ABCF1,VPS9, ATG23, UTP11, PEPCK, PMT, IDH, ALBA2, UTP15, NUP205 + 12 with unknown function
GEXP13, UTP21 + 9 with unknown function
MRE11, ApiAP2-SP2, SR1 + 4 with unknown function
YTM1, EXP3, RPB12, GRP170, ETRAMP14 + 9 with unknown function
Upstream intergenic
5' UTR
Intron
CDS
3'UTR
(d)
Chr2
Chr11
Chr12
Chr13
Chr13
LD (r^2)
(e)
PF3D7_1343400 (RAD5)
PF3D7_1343800 (VPS13)
PC1/2
(f)
PF3D7_1115300 (FAL2B)
Local-eSTR
Transcript phenotype association
Age-normalized expression
(g)
PF3D7_0203700 (MAK16)
(h)
PSOP7
PF3D7_1142100
PF3D7_1142300
MIAB
PF3D7_1133700
PF3D7_1409600
PTP1
GCD10
PF3D7_0220100
FP2B
PPCDC
MAK16
PF3D7_0404300
PF3D7_0719900
PF3D7_1225200
PF3D7_0517200
MSP8
PF3D7_0615400
Susceptible
Resistant
Cis-eSTRs
Expression
logP
location
3' UTR
5'UTR
CDS
Intergenic (down)
Intron
Intergenic (up)

Figure 4: Artemisinin resistance and STRs: **(a)** Manhattan plot shows the association between length variation of STRs with parasite clearance half-life (PC1/2). The X-axis shows genomic locations of the STRs, and the y-axis denotes the p-value of association. Clearly, locus from chromosome 13 shows the most significant association, along with many other candidates across the genome with resistance-associated STRs. **(b)** Resistant-associated STRs and their distance from the transcription start site (TSS) of the nearest gene. Here, the X-axis denotes 1kb bins across upstream and downstream of the TSS, and the y-axis denotes the count of resistance-associated STRs. Approximately 57% of resistance-associated STRs (n=48 out of 85) are within the 1kb upstream or downstream of their nearest genes. **(c)** This schematic shows genomic locations of resistance-associated STRs with respect to the nearest genes. For each genomic region, short names of the genes are mentioned at the top of the schematic. **(d)** LD heatmap showing R2 values involving all the pSTRs found significantly associated with PC1/2. Although there were no strong linkages observed between any of these candidates and PfKelch13 C580Y, several chromosomal regions showed stronger LD patterns and are marked by the triangles with chromosome identifiers written above each region. Two of such regions were zoomed and shown as a panel on top. This boxplot shows an example of a resistance-associated STR with a negative slope. **(e)** (Upper panel) In this example, a pSTR was found 274 upstream of RAD5 (PF3D7_1343400) associated with PC1/2 level. The X-axis of this plot shows variations in repeat numbers, and the y-axis denotes PC1/2. The blue line drawn for this plot and the other two example plots shows the sign of the slope, and the grey area shows confidence intervals. The point to be mentioned here is that RAD5 is a neighbor of PfKelch13. (Bottom panel) Here is another example of a resistance-associated STR with a positive slope. The boxplot gene shows another resistance-associated STRs found within the Pfkelch13 locus, VPS13 (PF3D7_1343400), with a coding pSTRs associated with PC1/2. **(f)** Boxplot shows the first example of the alternative (indirect) way to link STRs with artemisinin resistance. A STR with an AT motif located ~9kb upstream of cysteine proteinase falcipain 2b (PF3D7_1115700) linked with the expression changes and found to be differentially expressed in resistant parasites compared to susceptible parasites. **(g)** Another example involves a STR with an AT motif located 272kb upstream of MAK16 (PF3D7_0203700) linked with the expression changes and found to be differentially expressed in resistant parasites compared to susceptible parasites. (h) This figure shows an alternative ways to link pSTRs with resistance phenotype. Left heatmap shows 18 genes and their differential expression pattern in resistant and susceptible parasites. And the bubble plot on the left shows number of cis-STRs associated with these 18 genes and their location related to the genes. Size of the bubble denotes the p-value of association.

nonsynonymous SNPs in both genes were also associated with ART-R in our previous study, making them the most genetically altered in resistant parasites in GMS identified so far. Interestingly, all four AR-pSTR clusters also overlapped with or were closely linked to regions of linkage disequilibrium detected exclusively in the resistant parasite WGS **[Figure 4e]**. These included two clusters on chromosome 13, both linked to pfk13. The first of these involved AR-pSTR in, a heat shock-regulating ApiAP2 transcription factor (AP2-HS, PF3D7_1342900). The role of this transcription factor was previously shown to be important in the *P. falciparum* response to heat[33], which is also highly analogous to an unfolded protein response, a key factor in artemisinin resistance[23]. Besides this there are additional four LD clusters at chromosome 2, including DNA-directed RNA polymerase III subunit, RPC10 (PF3D7_0206600)

spanning across the serine repeat antigen (SERA) locus to a putative sugar transporter (PF3D7_0216600); chromosome 11 including six genes (between PF3D7_1119400 and PF3D7_1126700); chromosome 12 cluster around (PF3D7_1245500) and chromosome 13 downstream of the pfk13 locus, involving putative zinc finger proteins (PF3D7_1344300 & PF3D7_1345000) and isocitrate dehydrogenase (IDH, PF3D7_1345700). Although the putative roles of these genes in partial artemisinin resistance remain unknown, the genetic association revealed by this study warrants further investigations.

18 of the 76 AR-pSTR-linked genes were also found to be either up-regulated or down-regulated in resistant parasites compared to susceptible parasites, and this deregulation is linked to at least one pSTR (e.g. *cis*-eSTR) **[Figure 4f, Supp. Table S4]**. Interestingly, expression of 8 of these 18 genes was linked to more than one pSTR, including all possible positions within and outside the transcript regions. One notable example is cysteine proteinase falcipain-2b, which has been found to be up-regulated in resistant parasites **[Figure 4g].** Its differential mRNA levels appeared to be linked to a pSTR 9Kb downstream intergenic region correlated with its expression change, with a negative slope (i.e., expression decreases with the gain of repeats). The opposite trend was found for gene MAK16, whose expression decreases with the gain of AT-repeats in the 5'UTR that characterize the resistant parasites **[Figure 4h]**. Moreover, MAK16 appears to be linked with at least six *cis*-eSTRs, including polymorphisms in essentially all parts within its gene boundary **[Figure 4f & 2d]**. This abundant linkage could indicate strong selection pressure on these genes, possibly due to drug selection, further strengthening their role in partial artemisinin resistance. Indeed, MAK16 (PF3D7_0203700) is a *Plasmodium* homolog of the highly conserved MAK16 protein family, which is essential for ribosomal assembly in all eukaryotic cellular systems, including human cells. Protein metabolism was also strongly linked to artemisinin resistance in our earlier population transcriptomics analysis, further supporting its relevance.

5.6: Discussion

Genotyping of pSTR in this study is synonymous with microsatellite-based genetic studies widely used in *Plasmodium* research in the past[34]. Applicable to all *Plasmodium* species, microsatellites were particularly useful in *P. falciparum,* whose genome is abundant in STRs, comprising ~11% of its genome, reflecting its AT-richness[35]. Due to their high resolution and relatively simple methodological applications, microsatellite genotyping was used to study various biological aspects of *Plasmodium* parasites *in vivo*, such as population structure, vector-parasite dynamics, transmission patterns, and, crucially, the emergence and spread of drug resistance[34,36-38], including the PfK13-linked ART-R[3,14]. A recent example was the use of microsatellites to monitor selective sweeps of several *pfk13* alleles selected for by different ACT regimens in Southeast Asia over the last decade[39-42]. Microsatellite-assisted genotyping is now also being implemented to monitor the simultaneous spread of *the pfk13* SNP and *hrp2* deletion alleles, which mediate partial artemisinin resistance and rapid diagnostic failures, respectively, and are currently spreading in Sub-Saharan Africa[43]. Recognizing the impact of microsatellites (a.k.a. STRs), Han, J. et al, (2022) carried out the first genome-wide exploration of SRTs using the public database of *P. falciparum* and *P. vivax* WGS[29]. Classifying 6,768 pSTRs in the WGS dataset from 3000 *P. falciparum* specimens, Han, J., and colleagues were able to reconstruct the population structure and several genomic signatures of selection in the *P. falciparum* cohorts that were comparable to those obtained with the use of SNPs. Here, we identified 22,072 pSTRs in a WGS form from 239 parasite samples that provided a high statistical confidence for association analyses, including both eSTR and AR-pSTRs. This likely reflects the nature of the input WGS data, which in Hans' study included a broad and heterogeneous sample set from publicly available databases, whereas here we used in-house-generated high-coverage, high-quality sequencing data from the TACT-CV cohort, with well-defined drug-resistance phenotypes. This suggests that future improvements in sequencing technologies can significantly increase the information yield of *P. falciparum* population genetics, going beyond exonic SNPs.

The key appeal of STRs as genetic markers is their potential to have tangible functional implications for gene function and expression. Over the last decade, studies across diverse eukaryotic systems have demonstrated the role of short tandem repeat (STR) sequence length polymorphism in the quantitative modulation of gene

expression, rapid adaptation, and phenotypic plasticity[27,44,45]. STR variations were found to alter amino acid tracks, affecting protein sequence and structure, or to modulate transcriptional levels when located in promoters, untranslated regions (UTRs), or introns[46-48]. STR polymorphisms in regulatory regions can alter promoter architecture by increasing or decreasing the spacing between transcription factor binding sites and the downstream gene, affecting transcriptional efficiency[49,50]. Studies have also found that polymorphic STRs (pSTRs) influence promoter methylation and nucleosome occupancy, and ultimately local chromatin structure[51-53]. pSTRs within UTRs, on the other hand, have the capacity to alter mRNA secondary structure, hence altering the transcript stability, whereas intronic pSTRs can regulate splicing efficiency[54,55]. Due to higher mutation rates, pSTRs represent an active source of genetic variation with the potential to contribute to phenotypic plasticity, thereby enhancing adaptability and evolution[55,56]. Recognized as neutral genetic variants, these pSTRs were commonly used in population genetics, forensic genetics, paternity analysis, et[29,57-60]. Currently, however, an increasing number of human disease susceptibilities and trait modulations are being associated with pSTRs, thereby marking them as an important source of genetic variation[61-64]. Sometimes referred to as the “missing heritability,” sequence length polymorphisms are now recognized as equally important genetic factors as SNPs, often filling the gaps in understanding complex genetic traits[65]. In this study, we provide evidence for the potential of pSTR to fill inheritance gaps in complex genetic traits in malaria parasites, particularly *P. falciparum,* through their direct effects on mRNA levels. Given the genetic streamlining that characterizes the haploid *P. falciparum* genomes, it is tempting to suggest that gene expression variation may be a more efficient way for the parasite to adapt to varying external conditions than amino acid changes in robustly evolved protein sequences.

As such, pSTRs have a high potential to fill the missing inheritance of *P. falciparum*, the ART-R, that phenotype undermining global malaria control strategies[5]. Although firmly associated with nonsynonymous SNPs in the propeller domain of the *pfk13*, it is now clear that the molecular mechanism that mediates ARP is polygenic[18]. In earlier studies, it was shown that the ART-R is associated with a specific transcriptional reprogramming of the *P. falciparum* parasites in their natural infection settings[23,32] that likely arose from their initial transcriptional response *in vivo*[25]. Later,

the Artemisinin Resistance-associated Transcriptional Profile (ART-RP) was identified. This transcriptional reprogramming is closely related to the gene expression rewiring characterized in K13-Mutant *P. falciparum* lines *in vitro*[17]. The *in vivo* ARTP was also shown to have significant associations with SNP-based gene expression in an eQTL-like manner, suggesting it has lineage-based heritability throughout the emergence and spread of ART-R in Southeast Asia in the second decade of the 21st century[32]. However, the genetic factors mediating the heritability of the ART-RP remained unknown. The first indication of the role of pSRTs in transcriptional regulation came from the studies of a cyclophilin homolog, *PfCYP19B*, which was found transcriptionally upregulated in the ART-R parasites in Southeast Asia throughout the spread[26]. Here, a sequence-length polymorphism in the dinucleotide AT track in the promoter region of PfCYP19B was shown to directly regulate its mRNA levels, as demonstrated by a reporter gene assay *in vitro*. The resulting upregulation/overexpression of the *PfCYP19B* protein indeed mediated decreased susceptibility, thereby driving its upregulation and decreasing susceptibility of *P. falciparum* parasites to artemisinin derivatives (and piperaquine) *in vitro.* Taken together, the results presented in this study provide an important precedent for sequence-length polymorphism at STRs contributing to the genetic makeup of ART-R by altering gene expression, which can either directly contribute to the resistance mechanism or alleviate the fitness cost associated with it. With about a decade and a half delay from Southeast Asia, ART-R emerged in Sub-Saharan Africa with independent spontaneous rises of distinct PfK13 alleles[66,67]. In contrast to Southeast Asia, however, the relationship between the PfK13 genotypes and the ART-R phenotype, measured either in vivo by $PC^{1/2}$ or *in vitro* using the Ring Survival Assay (RSA), is more heterogeneous[68]. Moreover, the rise of the PfK13 mutants and accompanying ART-R phenotype in the African *P. falciparum* parasite is modulated by highly distinct genetic backgrounds, leading to a co-selection of distinct modifier gene polymorphisms compared to Southeast Asia, including Pfubp1[69], coronin[70], and recently discovered PfPX1[71]. Given the uniqueness but at the same time urgency of the inevitable rise of ART-R (likely followed by ACT failures) in this most important malaria endemic region, there is a need to “recalibrate our early detection signals” to prevent any “blind spots” in detecting its uncontrolled spread[68]. The results in this study provide a precedent for including polymorphic STR as an additional source of genetic

markers that, in addition to the currently used SNP-based genotyping, could considerably increase sensitivity as well as effectiveness of such detections.

Materials and Methods

Short tandem repeat detection in the *Plasmodium falciparum* genome

To detect polymorphic STRs in the dataset, first, STR regions within the genome were marked.We used Tandem Repeat Finder to mark the Tandem repeat region with parameter Match score =2, Mismatch penalty=15, Delta(Indel Penalty)=15, PM(Match probability)=80, PI(Indel probability)=10, Minscore=10, MaxPeriod=6. A total of 561,342 STRs were found in 14 chromosomes of *Plasmodium falciparum*. Further 2 criteria were imposed to narrow down the list: (1) the minimum repeat times need to be >= 3, and the minimum overall alignment score needs to be >= 10. According to this criterion, 388,442 STR regions were selected for genotyping.

2.19: Genotyping and filtering Short tandem repeat

All samples used in this study were collected from 310 patients involved in multi-site clinical trials at five hospitals or health centers in three locations (western Cambodia, eastern Cambodia, and southern Vietnam). The trial took place between March 18, 2018, and January 30, 2020. All details regarding sample collection, site locations, inclusion criteria, and parasitemia assessment are provided in the supplementary Table 1 [**Table S5**. The samples were chosen based on the x-coverage value such that if the mean genome coverage is >= 10, the samples were genotyped. We used HipSTR[72] for genotyping the defined STR regions out of 239 isolates. HipSTR is a specialized tool that takes care of each locus based on the sequence context. Further, it mines the population genomics data into each stutter model and minimizes the artifacts while aligning candidate alleles. More importantly, it uses SNP haplotypes to genotype and phase STRs. Further, these genotypes were filtered according to three conditions (a) minimum call quality is 0.9, ii) less than 85% of reads used for calling

the genotype do not have any flanking indel and iii) less than 85% of reads have any artifact due to stutter estimation.

2.20: Population structure analysis on Short tandem repeat

After achieving high-quality STR genotype data, the first goal was to look for any signature of population structure. STR polymorphisms are a form of short insertions and deletions. SNP-based models for population structure determination do not hold in these calculations. Bruvo's distance was used to generate the distance matrix among individuals. The Bruvo's distance follows stepwise mutation models, i.e. ATATAT allele is closer to the ATAT allele rather than the AT allele. The generated distance matrix is first used for multidimensional scaling to visualize how the samples are connected according to the STR genotypes. And then, using APE, a Neighbour-joining tree was generated with the distance matrix[73]. Both the Neighbour-joining tree and MDS values were compared with the SNP-based tree and PCA.

2.21: Detecting eSTR based on a linear regression model

To determine the expression associated with the polymorphic Short Tandem Repeat, data was cleaned as the first step. Since the Linear regression model was considered to carry out the association study, STRs with at least three alleles were considered, with each allele represented by 5 samples. Expression residual of each transcript was obtained from a previously carried out transcriptome-phenotype association study, and the expression outliers were marked for each of them. Expression values greater than or less than 3 times Median Absolute Deviation (3*MAD) were marked as NA values. This is a very important step because the presence of expression outliers artificially drives the association p-value towards higher significance, hence generating a false positive association between genotype and phenotype. Finally, the linear regression model was implemented using the R lm function and included the first three MDS values as a population and unknown covariates present in genotype data. It is worth noting that, for expression data, both age and detection level were considered while getting the expression residual values. All possible pairs between STR genotype and expression data were regressed with the above-mentioned model. The result was

initially filtered by a standard p-value cut-off 1*e05 and a Cis-Trans Plot was drawn, marking the associated genes that passed the FDR cut-off in red.

2.22: Defining local eSTRs

This study was focused on determining polymorphic STRs, being in the vicinity of genes, regulating their expression. For that, We defined local interaction by any STR presented (a) gene within the gene body or (b) upstream/ downstream intergenic regions of the gene on the same strand. In simple words, We defined the strand-specific local interaction between STR genotype and the expression phenotype. Now, once local interactions were marked, STRs were annotated using their positions corresponding to gene expression data. A variety of characteristic features were investigated e.g. slope of the association, motif sequence bias, etc. Lastly, the gene set enrichment analysis was carried out to understand the pathways affected by polymorphic STRs.

2.23: Associating STR genotypes with clearance half-life (strGWAS)

Finally, we wanted to associate these polymorphic STRs with the parasite clearance half-life of the isolates. For this purpose, we used 'TRTools: a toolkit for genome-wide tandem repeat analysis '[31]. The command line of ‘TRTools’ consists of a module named ‘AssociaTR’ to determine the length of STR with phenotype. A Manhattan plot was drawn based on the association analysis in search of a novel locus associated with artemisinin resistance. Further overlap with local eSTRs was mapped to understand the link among genotype, resistance phenotype, and gene expression.

References

1. Organisation, T.W.H. The World Malaria Report. (2024).
2. Dondorp, A.M. *et al.* Artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med* **361**, 455-67 (2009).

3. Ashley, E.A. *et al.* Spread of artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med* **371**, 411-23 (2014).
4. Balmer, A.J. *et al.* Understanding the global rise of artemisinin resistance: Insights from over 100,000 Plasmodium falciparum samples. *Elife* **14**(2025).
5. Organization, T.W.H. Malaria: Artemisinin partial resistance. (2025).
6. Bjorkman, A., Gil, P. & Alifrangis, M. Alarming Plasmodium falciparum resistance to artemisinin-based combination therapy in Africa: the critical role of the partner drug. *Lancet Infect Dis* **24**, e540-e541 (2024).
7. Na-Bangchang, K., Ruengweerayut, R., Mahamad, P., Ruengweerayut, K. & Chaijaroenkul, W. Declining in efficacy of a three-day combination regimen of mefloquine-artesunate in a multi-drug resistance area along the Thai-Myanmar border. *Malar J* **9**, 273 (2010).
8. van der Pluijm, R.W. *et al.* Determinants of dihydroartemisinin-piperaquine treatment failure in Plasmodium falciparum malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study. *Lancet Infect Dis* **19**, 952-961 (2019).
9. Mathieu, L.C. *et al.* Local emergence in Amazonia of Plasmodium falciparum k13 C580Y mutants associated with in vitro artemisinin resistance. *Elife* **9**(2020).
10. Miotto, O. *et al.* Emergence of artemisinin-resistant Plasmodium falciparum with kelch13 C580Y mutations on the island of New Guinea. *PLoS Pathog* **16**, e1009133 (2020).
11. Balikagala, B. *et al.* Evidence of Artemisinin-Resistant Malaria in Africa. *N Engl J Med* **385**, 1163-1171 (2021).
12. Stokes, B.H., Ward, K.E. & Fidock, D.A. Evidence of Artemisinin-Resistant Malaria in Africa. *N Engl J Med* **386**, 1385-1386 (2022).
13. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat Genet* **47**, 226-34 (2015).
14. Ariey, F. *et al.* A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* **505**, 50-5 (2014).
15. Davis, S.Z. *et al.* The extended recovery ring-stage survival assay provides a superior association with patient clearance half-life and increases throughput. *Malar J* **19**, 54 (2020).
16. Birnbaum, J. *et al.* A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. *Science* **367**, 51-59 (2020).
17. Mok, S. *et al.* Artemisinin-resistant K13 mutations rewire Plasmodium falciparum's intra-erythrocytic metabolic program to enhance survival. *Nat Commun* **12**, 530 (2021).
18. Kucharski, M., Nayak, S., Gendrot, M., Dondorp, A.M. & Bozdech, Z. Peeling the onion: how complex is the artemisinin resistance genetic trait of malaria parasites? *Trends Parasitol* **40**, 970-986 (2024).
19. Witkowski, B. *et al.* Novel phenotypic assays for the detection of artemisinin-resistant Plasmodium falciparum malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *Lancet Infect Dis* **13**, 1043-9 (2013).
20. Straimer, J. *et al.* Drug resistance. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. *Science* **347**, 428-31 (2015).
21. Nayak, S. *et al.* Population genomics and transcriptomics of Plasmodium falciparum in Cambodia and Vietnam uncover key components of the artemisinin resistance genetic background. *Nat Commun* **15**, 10625 (2024).

22. Rosenthal, M.R., Vijayrajratnam, S., Firestone, T.M. & Ng, C.L. Enhanced cell stress response and protein degradation capacity underlie artemisinin resistance in Plasmodium falciparum. *mSphere* **9**, e0037124 (2024).
23. Mok, S. *et al.* Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**, 431-5 (2015).
24. Bridgford, J.L. *et al.* Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. *Nat Commun* **9**, 3801 (2018).
25. Zhu, L. *et al.* Artemisinin resistance in the malaria parasite, Plasmodium falciparum, originates from its initial transcriptional response. *Commun Biol* **5**, 274 (2022).
26. Kucharski, M. *et al.* Short tandem repeat polymorphism in the promoter region of cyclophilin 19B drives its transcriptional upregulation and contributes to drug resistance in the malaria parasite Plasmodium falciparum. *PLoS Pathog* **19**, e1011118 (2023).
27. Hannan, A.J. Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet* **19**, 286-298 (2018).
28. Peto, T.J. *et al.* Triple therapy with artemether-lumefantrine plus amodiaquine versus artemether-lumefantrine alone for artemisinin-resistant, uncomplicated falciparum malaria: an open-label, randomised, multicentre trial. *Lancet Infect Dis* **22**, 867-878 (2022).
29. Han, J., Munro, J.E., Kocoski, A., Barry, A.E. & Bahlo, M. Population-level genome-wide STR discovery and validation for population structure and genetic diversity assessment of Plasmodium species. *PLoS Genet* **18**, e1009604 (2022).
30. Miles, A. *et al.* Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Res* **26**, 1288-99 (2016).
31. Mousavi, N. *et al.* TRTools: a toolkit for genome-wide analysis of tandem repeats. *Bioinformatics* **37**, 731-733 (2021).
32. Zhu, L. *et al.* The origins of malaria artemisinin resistance defined by a genetic and transcriptomic background. *Nat Commun* **9**, 5158 (2018).
33. Tinto-Font, E. *et al.* A heat-shock response regulated by the PfAP2-HS transcription factor protects human malaria parasites from febrile temperatures. *Nat Microbiol* **6**, 1163-1174 (2021).
34. Verma, K., Razack, S.A., Schneider, K.A. & Bharti, P.K. Microsatellites markers fostering the understanding of malaria parasite biology, epidemiology and population genetics. *Diagn Microbiol Infect Dis* **111**, 116667 (2025).
35. McDew-White, M. *et al.* Mode and Tempo of Microsatellite Length Change in a Malaria Parasite Mutation Accumulation Experiment. *Genome Biol Evol* **11**, 1971-1985 (2019).
36. Mita, T. *et al.* Limited geographical origin and global spread of sulfadoxine-resistant dhps alleles in Plasmodium falciparum populations. *J Infect Dis* **204**, 1980-8 (2011).
37. Sa, J.M. *et al.* Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc Natl Acad Sci U S A* **106**, 18883-9 (2009).
38. Wootton, J.C. *et al.* Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature* **418**, 320-3 (2002).

39. Awor, P. *et al.* Indigenous emergence and spread of kelch13 C469Y artemisinin-resistant Plasmodium falciparum in Uganda. *Antimicrob Agents Chemother* **68**, e0165923 (2024).
40. Imwong, M. *et al.* The spread of artemisinin-resistant Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet Infect Dis* **17**, 491-497 (2017).
41. Tun, K.M. *et al.* Spread of artemisinin-resistant Plasmodium falciparum in Myanmar: a cross-sectional survey of the K13 molecular marker. *Lancet Infect Dis* **15**, 415-21 (2015).
42. Li, M. *et al.* Genetic characteristics of P. falciparum parasites collected from 2012 to 2016 and anti-malaria resistance along the China-Myanmar border. *PLoS One* **18**, e0293590 (2023).
43. Mihreteab, S. *et al.* The spread of molecular markers of artemisinin partial resistance and diagnostic evasion in Eritrea: a retrospective molecular epidemiology study. *Lancet Microbe* **6**, 100930 (2025).
44. Fotsing, S.F. *et al.* The impact of short tandem repeat variation on gene expression. *Nat Genet* **51**, 1652-1659 (2019).
45. Gymrek, M. *et al.* Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet* **48**, 22-9 (2016).
46. Chen, H.Y. *et al.* The mechanism of transactivation regulation due to polymorphic short tandem repeats (STRs) using IGF1 promoter as a model. *Sci Rep* **6**, 38225 (2016).
47. Wright, S.E. & Todd, P.K. Native functions of short tandem repeats. *Elife* **12**(2023).
48. Kim, J.H. *et al.* Short tandem repeat expansions in cortical layer-specific genes implicate in phenotypic severity and adaptability of autism spectrum disorder. *Psychiatry Clin Neurosci* **78**, 405-415 (2024).
49. Abe, H. & Gemmell, N.J. Evolutionary Footprints of Short Tandem Repeats in Avian Promoters. *Sci Rep* **6**, 19421 (2016).
50. Zhang, X. *et al.* Systematic evaluation of the impact of promoter proximal short tandem repeats on expression. *bioRxiv* (2025).
51. Sun, J.H. *et al.* Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **175**, 224-238 e15 (2018).
52. Martin-Trujillo, A., Garg, P., Patel, N., Jadhav, B. & Sharp, A.J. Genome-wide evaluation of the effect of short tandem repeat variation on local DNA methylation. *Genome Res* **33**, 184-196 (2023).
53. Quilez, J. *et al.* Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Res* **44**, 3750-62 (2016).
54. Kinney, N., Pathak, D., Evans, E. & Arias, P. Short tandem repeat variants are possibly associated with RNA secondary structure and gene expression. *PLoS One* **20**, e0326355 (2025).
55. Zhang, Z.Q. *et al.* Mutations of short tandem repeats explain abundant trait heritability in Arabidopsis. *Genome Biol* **26**, 242 (2025).
56. Verbiest, M. *et al.* Mutation and selection processes regulating short tandem repeats give rise to genetic and phenotypic diversity across species. *J Evol Biol* **36**, 321-336 (2023).
57. Alford, R.L., Hammond, H.A., Coto, I. & Caskey, C.T. Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am J Hum Genet* **55**, 190-5 (1994).

58. Mei, S. *et al.* The Polymorphism Analyses of Short Tandem Repeats as a Basis for Understanding the Genetic Characteristics of the Guanzhong Han Population. *Biomed Res Int* **2021**, 8887244 (2021).
59. Pajnic, I.Z. Autosomal STR Markers for Forensic Genetics: Applications, Challenges, and Future Directions. *Genes (Basel)* **17**(2026).
60. Wyner, N., Barash, M. & McNevin, D. Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype. *Front Genet* **11**, 884 (2020).
61. Margoliash, J. *et al.* Polymorphic short tandem repeats make widespread contributions to blood and serum traits. *Cell Genom* **3**, 100458 (2023).
62. Bustos, B.I. *et al.* Genome-wide contribution of common short-tandem repeats to Parkinson's disease genetic risk. *Brain* **146**, 65-74 (2023).
63. Zhang, G. & Andersen, E.C. Interplay Between Polymorphic Short Tandem Repeats and Gene Expression Variation in Caenorhabditis elegans. *Mol Biol Evol* **40**(2023).
64. Henden, L. *et al.* Short tandem repeat expansions in sporadic amyotrophic lateral sclerosis and frontotemporal dementia. *Sci Adv* **9**, eade2044 (2023).
65. Hannan, A.J. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet* **26**, 59-65 (2010).
66. Rosenthal, P.J. *et al.* The emergence of artemisinin partial resistance in Africa: how do we respond? *Lancet Infect Dis* **24**, e591-e600 (2024).
67. Rosenthal, P.J., Asua, V. & Conrad, M.D. Emergence, transmission dynamics and mechanisms of artemisinin partial resistance in malaria parasites in Africa. *Nat Rev Microbiol* **22**, 373-384 (2024).
68. Hosangadi, V. & Fidock, D.A. Artemisinin partial resistance at a crossroads: evidence for continental variation in Plasmodium falciparumclearance phenotypes. *Lancet Infect Dis* (2026).
69. Pratt, S. *et al.* Novel Pfk13 and Pfubp1 genotypes in African Plasmodium falciparum isolates exhibiting reduced susceptibility to the antimalarials artemisinin and lumefantrine. *mBio* **17**, e0367625 (2026).
70. Sharma, A.I. *et al.* Genetic background and PfKelch13 affect artemisinin susceptibility of PfCoronin mutants in Plasmodium falciparum. *PLoS Genet* **16**, e1009266 (2020).
71. Christopher Bower-Lepts , A.J., Sachie Kanatani, Abhai Tripathi, Godfree Mlambo, Alejandro McCotter-González, Linda Romano, Stephen Orena, Martin Okitwi, Karamoko Niaré, Philip J Rosenthal, Melissa Conrad, Photini Sinnis , Sachel Mok. A Plasmodium falciparum PX1 haplotype is associated with reduced susceptibility to artemisinin and lumefantrine. *bioRxiv preprint* (2026).
72. Amaratunga, C. *et al.* Dihydroartemisinin-piperaquine resistance in Plasmodium falciparum malaria in Cambodia: a multisite prospective cohort study. *Lancet Infect Dis* **16**, 357-65 (2016).
73. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Suppfig17.pdf
- Supplementarytable1.xlsx
- Supplementarytable2.xlsx
- Supplementarytable3.xlsx
- Supplementarytable4.xlsx
- Supplementarytable5.xlsx