21. J. L. Pruneda-Paz, G. Breton, A. Para, S. A. Kay, *Science* **323**, 1481 (2009).
22. Materials and methods are available as supporting material on *Science* Online.
23. M. F. Covington, J. N. Maloof, M. Straume, S. A. Kay, S. L. Harmer, *Genome Biol.* **9**, R130 (2008).
24. S. P. Hazen *et al.*, *Genome Biol.* **10**, R17 (2009).
25. T. P. Michael *et al.*, *PLoS Genet.* **4**, e14 (2008).
26. A. M. Lloyd, M. Schena, V. Walbot, R. W. Davis, *Science* **266**, 436 (1994).
27. A. Baudry *et al.*, *Plant Cell* **22**, 606 (2010).
28. N. Nakamichi *et al.*, *Plant Cell* **22**, 594 (2010).

Acknowledgments: We thank E. Meyerowitz for comments on the manuscript, E. Centeno for help with ChIP-Seq analysis, O. Casagran for help with the microfluidic arrays, M. Perales for help with the ChIP, L. Rico for help with the inducible constructs, and L. Schaeffer and D. Trout (Jacobs Genetics and Genomics Laboratory at Caltech) for the primary sequence data processing. This work was supported by grants to P.M. from the Ramón Areces Foundation, the Spanish Ministry of Science and Innovation (MICINN), the European Molecular Biology Organization (EMBO) Young Investigators Program, and from the European Heads of Research Councils and the European Science Foundation through the European Young Investigator Award; to J.L.R. from the European Comission (EC) Marie Curie program and MICINN; and to A.J.M. and others from the EC FP7 Collaborative Project TiMet. The Centre for Systems Biology at Edinburgh is a Centre for Integrative and Systems Biology supported by the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council award D019621. W.H. is supported by a Juan de la Cierva contract (MICINN) and P.P.-G. by a Formación de Personal Investigador fellowship (MICINN). Sequencing data have been deposited with the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE35952.

Supporting Online Material
www.sciencemag.org/cgi/content/full/science.1219075/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S7
References (*29*–*45*)

12 January 2012; accepted 27 February 2012
Published online 8 March 2012;
10.1126/science.1219075

A Major Genome Region Underlying Artemisinin Resistance in Malaria

Ian H. Cheeseman,[1] Becky A. Miller,[2] Shalini Nair,[1] Standwell Nkhoma,[1] Asako Tan,[2] John C. Tan,[2] Salma Al Saai,[1] Aung Pyae Phyo,[3] Carit Ler Moo,[3] Khin Maung Lwin,[3] Rose McGready,[3,4,5] Elizabeth Ashley,[3,4,5] Mallika Imwong,[4] Kasia Stepniewska,[4,5,7] Poravuth Yi,[8] Arjen M. Dondorp,[4,5] Mayfong Mayxay,[6] Paul N. Newton,[5,6] Nicholas J. White,[4,5] François Nosten,[3,4,5] Michael T. Ferdig,[2] Timothy J. C. Anderson[1]*

Evolving resistance to artemisinin-based compounds threatens to derail attempts to control malaria. Resistance has been confirmed in western Cambodia and has recently emerged in western Thailand, but is absent from neighboring Laos. Artemisinin resistance results in reduced parasite clearance rates (CRs) after treatment. We used a two-phase strategy to identify genome region(s) underlying this ongoing selective event. Geographical differentiation and haplotype structure at 6969 polymorphic single-nucleotide polymorphisms (SNPs) in 91 parasites from Cambodia, Thailand, and Laos identified 33 genome regions under strong selection. We screened SNPs and microsatellites within these regions in 715 parasites from Thailand, identifying a selective sweep on chromosome 13 that shows strong association ($P = 10^{-6}$ to 10^{-12}) with slow CRs, illustrating the efficacy of targeted association for identifying the genetic basis of adaptive traits.

Artemisinin-based combination therapies (ACTs) are the first-line treatment in nearly all malaria-endemic countries (*1*) and are central to the current success of global efforts to control and eliminate *Plasmodium falciparum* malaria (*2*). Resistance to artemisinin (ART) in *P. falciparum* has been confirmed in Southeast Asia (*3*), raising concerns that it will spread to sub-Saharan Africa, following the path of chloroquine and anti-folate resistance (*4*). ART resistance results in reduced parasite clearance rates (CRs) after treatment (Fig. 1A) and is principally due to parasite genetics, which determines 58 and 64% of the variance in parasite CRs in western Cambodia and western Thailand, respectively (*5*, *6*). The resistance mechanism is unknown, but reduced killing of "ring"-stage parasites (*7*) and quiescence have been implicated (*8*). The genetic basis is likely to be simple. A single mutation in the *ubp1* gene confers ART resistance in the *P. chabaudi* mouse malaria model (*9*). Similarly, resistance to other antimalarials in *P. falciparum* involves single major-gene effects or is oligogenic (*10*).

Cross-population genomic comparisons offer a means to identify putative targets of natural selection (*11*–*14*). Targeted association analyses of genome regions under selection can then be used to directly identify the genetic basis of adaptive traits (*13*, *14*), minimizing the multiple-testing penalties constraining standard genome-wide association studies. We compared three neighboring Southeast Asian *P. falciparum* populations (in Laos, Thailand, and Cambodia) with low levels of genetic differentiation (Fig. 1B), but differences in CRs after ART treatment (Fig. 1C), to detect recent selective sweeps that may underlie resistance. Parasites from Laos are cleared rapidly (*15*) [median –log(CR) = 1; half-life of parasite decline = 2 hours], parasites from western Cambodia clear slowly [median –log(CR) 2.15, half-life 6 hours], and parasites from western Thailand show a wide range of clearance [median –log(CR) 1.4, half-life 3 hours]. CR distributions are significantly different between all locations (Thai-Cambodia, $D = 0.58$, $P < 0.001$; Thai-Laos, $D = 0.68$, $P < 0.001$; Cambodia-Laos, $D = 0.93$, $P < 0.001$; two-sided Kolmogorov-Smirnov test).

We genotyped 91 genetically unique single-clone parasites (27 from Laos, 30 from Cambodia, and 34 from Thailand) by hybridization to a custom Nimblegen genotyping array that scores single-nucleotide polymorphisms (SNPs) [1 SNP per 500 base pairs (bp)] and copy number variation (CNV) (*16*, *17*). Principal-components analysis (PCA) and global fixation indices (F_{ST}) confirmed low but significant differentiation between the three populations (Fig. 1, B and D). We characterized CNV across the three populations, identifying 78 common CNVs [minor allele frequency (MAF) > 5%] containing 209 genes (*17*).

For each SNP (MAF > 5%, $n = 6969$), we calculated two statistics to identify genome regions under strong selection for our three populations, measuring differentiation in haplotype structure [XP-EHH (*18*)] and allele frequency (F_{ST}) and classifying selected regions using a 10-kb sliding window approach (*17*) (Fig. 2 and fig. S1). Thirty-three nonoverlapping regions, comprising 2.4% of the 23-Mb genome, showed evidence for selection (top 1% of genome-wide values in both tests) in one or more populations and were ranked by the proportion of significant tests (table S1).

Known antimalarial resistance genes accounted for 10 out of 33 (10/33) genome regions. Three genes (*pfcrt, dhps,* and *dhfr*) were identified directly, whereas *GTP-cyclohydrolase I* showed evidence of selection within 5 kb (Fig. 2B). *pfmdr1* was not identified, most likely because the amplicon containing *pfmdr1* has multiple origins (*19*). An additional 23 genome regions showed strong signatures of selection (table S1 and fig. S1). Three regions, on chromosomes (chrs) 6, 13, and 14, were significant at multiple adjacent windows and ranked eighth, first, and second, respectively (table S1 and Fig. 2C). We did not observe evidence for selection at two proposed candidate loci, *atpase6* (Fig. 2A) (*20*) or *Part (21)*. Two

[1]Texas Biomedical Research Institute, San Antonio, TX 78245, USA. [2]The Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. [3]Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand. [4]Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. [5]Centre for Tropical Medicine, Churchill Hospital, Oxford, UK. [6]Wellcome Trust–Mahosot Hospital–Oxford Tropical Medicine Research Collaboration, Mahosot Hospital, Vientiane, Lao People's Democratic Republic. [7]Worldwide Antimalarial Resistance Network, Oxford, UK. [8]The National Center for Parasitology, Entomology, and Malaria Control, Phnom Penh, Cambodia.

*To whom correspondence should be addressed. E-mail: tanderso@txbiomedgenetics.org

CNVs showed strong differentiation (F_{ST} > 0.4). A chr-12 CNV containing *GTP-cyclohydrolase I* (F_{ST} = 0.6; Thailand versus Laos) is driven by anti-folate treatment (*22*); whereas a deletion of *surfin4.1* (F_{ST} = 0.51; Cambodia versus Laos) is not a strong candidate, because this is present in >20% of Lao infections and in African samples. The 78 CNVs did not overlap with the 33 regions under selection and were not considered further.

We examined the association of each of these 33 regions with parasite CRs in an independent parasite population. Targeted association allowed us to exploit an extensive archive of blood-spot DNA samples from Thai patients with detailed (6-hourly) CR data. Between 2001 and 2010, 3202 patients with uncomplicated malaria were treated with ART in four clinics on the Thai-Burmese border (*5*). The proportion of parasites with slow clearance [–log(CR) > 1.89] rose from <5% in 2001 to >50% in 2010, with resistance spreading more rapidly north of Mae Sot as compared to the south (Fig. 3A). In 1689 patients with available blood spots, we identified (fig. S4) all genetically unique infections containing a single parasite clone [n = 715; 417 from the north (96 from 2001–2004 and 321 from 2007–2010) and 298 from the south (121 from 2001–2004 and 177 from 2007–2010)]. We genotyped 90 SNPs targeting the 33 selected regions and 4 SNPs in *atpase6* and *pfmdr1* (table S2), in addition to the 93 genome-wide control SNPs genotyped previously (*5*, *17*).

There were no associations in 2001–2004 (Fig. 3B) or at either *pfmdr1* or *atpase6*. Two adjacent SNPs (14 kb apart) on chr 13 showed strong association ($P = 5.0 \times 10^{-6}$ to 6.5×10^{-7}, Fishers' exact test) in the north (2007–2010, Fig. 3D). The strongest signals of association in the 2007–2010 southern samples lay adjacent to these SNPs (Fig. 3C). Quantile-quantile plots indicate moderate inflation of association P values (inflation factor = 1.24 to 1.27, fig. S2), but the two SNPs remain significantly associated after adjustment for this inflation.

We fine-mapped this region using 19 microsatellite markers spanning 550 kb surrounding the strongest association signal in 417 northern Thai, 88 Lao, and 83 Cambodian parasites. Four microsatellites spanning ~35 kb showed strong association with CR ($P = 2.8 \times 10^{-9}$ to 6.6×10^{-12}; χ^2 test) in 2007–2010 but no association in 2001–2004 (Fig. 4A). This analysis uses a threshold to define "resistant" parasites. Reanalysis of quantitative CR data using a general linear model (*23*) confirms the results of the categorical χ^2 test ($P = 1.1 \times 10^{-5}$ to 1.6×10^{-10}), and use of a more stringent threshold [–log(CR) > 2.19] for defining resistance boosted maximal significance to $P = 9 \times 10^{-16}$. Alleles at the marker showing maximal association show a threefold difference in CR (Fig. 4D).

Assuming a recent selective sweep, we expected reduced allelic variation in Cambodia and Thailand relative to Laos. We observed maximal diversity reduction in Cambodia [expected heterozygosity (H_e) = 0.24 ± 0.07 (1 SD)] at eight markers spanning 105 kb (Fig. 4B). In contrast, diversity is high across this region in Laos (H_e = 0.79 ± 0.17 SD), whereas Thai parasites showed intermediate levels of diversity (H_e = 0.63 ± 0.09 SD). There was no separation between highly resistant and sensitive parasites, perhaps due to multiple origins of resistance alleles or evolution from standing variation (*24*). Analysis of haplotype structure provides strong evidence for recent selection in Thailand (Fig. 4C). Extended haplotype homozygosity (*EHH*) decays to background levels (0.05) over 99 kb (range = 24 to 190 kb) in sensitive [–log(CR) < 1.3] parasites, whereas in resistant [–log(CR) > 1.89] parasites, decay to background levels is over 375 kb (range = 140 to >550 kb). Permutation testing reveals this difference to be highly significant (P = 0.0003, (*17*) and fig. S5).

Using a general linear model (*23*), we estimated that 22.5% of the variation in CRs was determined by this region in 2007–2010. Given the heritability of CR in this population [64% (*5*)], we conservatively estimate that this locus explains at least 35.2% (22.5/64) of the heritable component of CR, suggesting that it is a major determinant of ART resistance.

Within this 35-kb region there are multiple candidate ART-resistance genes (table S4). These include the genes encoding lipoate synthase (lipoic acid salvage/biosynthesis), aminomethyltransferase (glycine cleavage pathway), and heat shock protein 70 (stress response/molecular chaperone). We sequenced the coding regions of six of the seven genes (18,747 bp) in 8 Lao, 8 Cambodian, and 24 to 31 Thai parasites (fig. S3). Six nonsynonymous, derived mutations in 3/6 genes are at high frequency in ART-resistant populations including SNPs 1 and 2 (Fig. 3F and table S4). However, these were present in parasites from Southeast Asia before the origin of ART resistance, suggesting that they are associated with but do not directly underlie CR and that noncoding regulatory mutations may be involved. Analysis of transcriptional changes in resistant lines offers a means to further prioritize genes for follow-up. Three genes within the region show significant changes in transcript levels during at least one life-cycle stage in ART-resistant lines profiled in a recent transcriptomic study (*25*) (table S4).

The spread of ART-resistant parasites would be catastrophic for malaria control. We have shown that a strongly selected region on chr 13, containing several candidate genes, explains a large proportion of variation in CR. Future functional dissection of loci within this region will be dependent on the

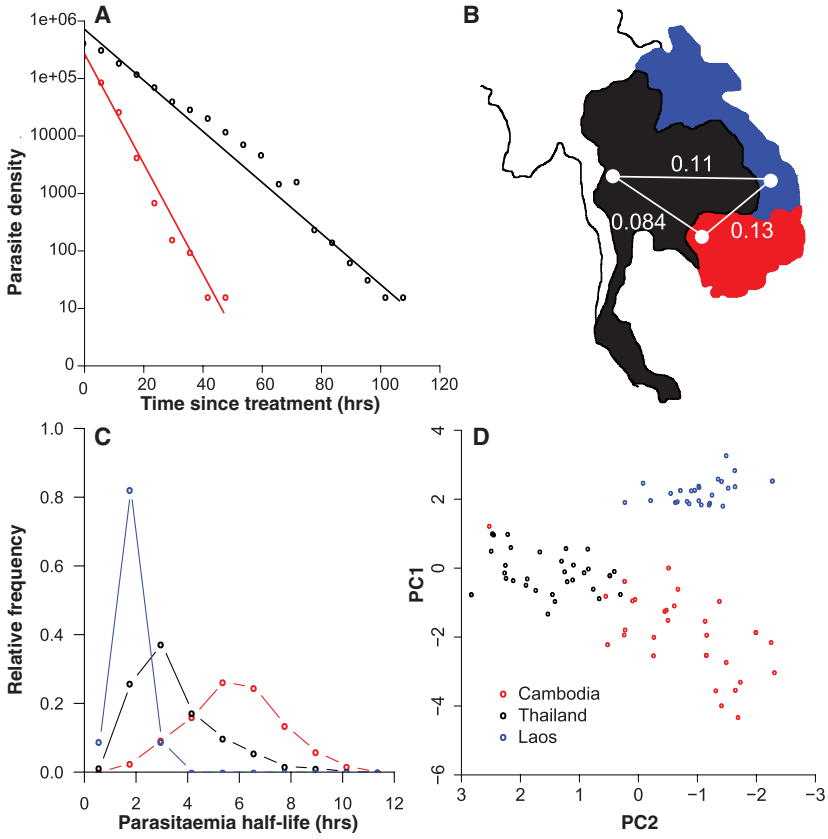

Fig. 1. Phenotypic and genetic differentiation between Southeast Asian parasites. (**A**) Patterns of parasite clearance from two Thai patients (black, slow CR; red, fast CR) after treatment. (**B**) F_{ST} between locations for 5662 SNPs with MAF > 5%. (**C**) Clearance half-lives from Laos (blue, n = 44 patients), Cambodia (red, n = 64), and Thailand (black, n = 3202). (**D**) PCA of parasites; 1770 SNPs were used, with MAF >5% and no missing data. PC1 and -2 are the first two principal components of the data.

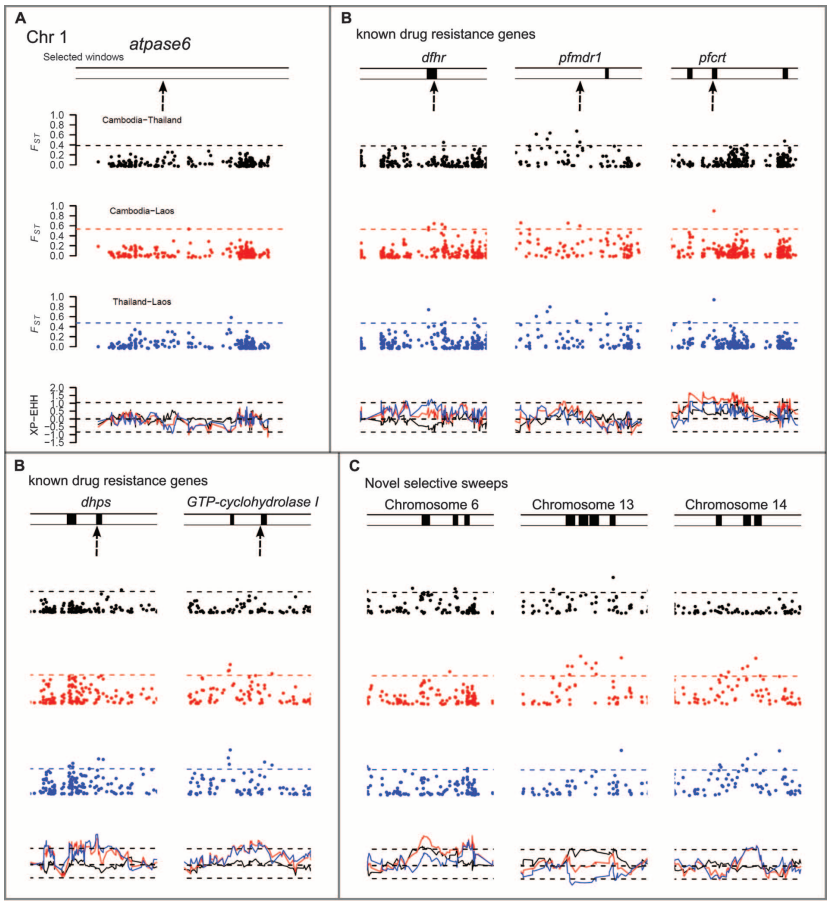

Fig. 2. Evidence for selection. Plots show F_{ST} and XP-EHH for each SNP (n = 6969) in a pairwise comparisons of countries. Dashed lines represent the 1% level from the genomic empirical distribution. Significant windows are shown by black bars above each plot. Red, Cambodia versus Laos; blue, Thailand versus Laos; black, Cambodia versus Thailand. (**A**) Chr 1. There is no evidence of selection surrounding *atpase6*. (**B**) Patterns of selection surrounding five known drug-resistance genes (marked by arrows). (**C**) Three loci putatively under selection.

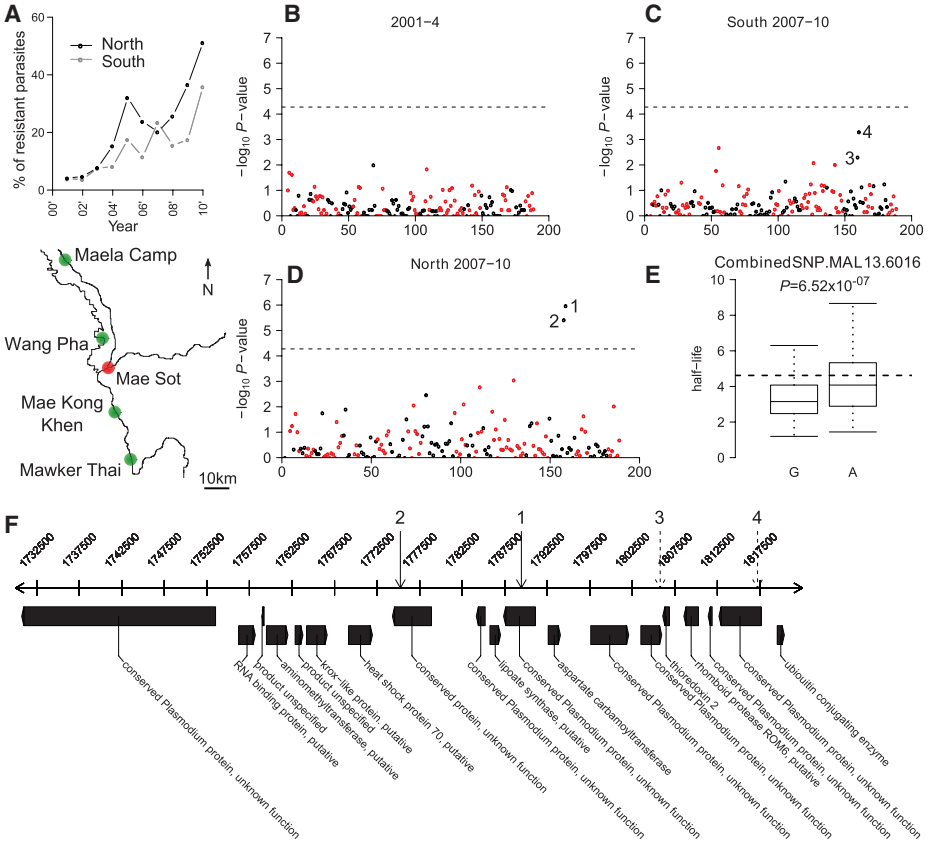

Fig. 3. Association testing. (**A**) Study sites on the Thai-Burma border. Northern populations were Maela Camp and Wang Pha, and southern populations were Mae Kong Khen and Mawker Thai. The clearance rate declined between 2001 and 2010 (shown above the map). The *y* axis shows the percentage of parasites with –log(clearance rate) > 1.89. (**B** to **D**) Association of 94 SNPs from selected regions (black) and 93 genome-wide SNPs (red). Two neighboring SNPs in the northern population from 2007 to 2010 showed strong association with CRs (denoted 1 and 2). In (C), 3 and 4 indicate the SNPs showing the strongest associations in regions under selection. (**E**) Box plot of the CR for each allele of SNP 1. (**F**) The chromosomal context surrounding these SNPs.

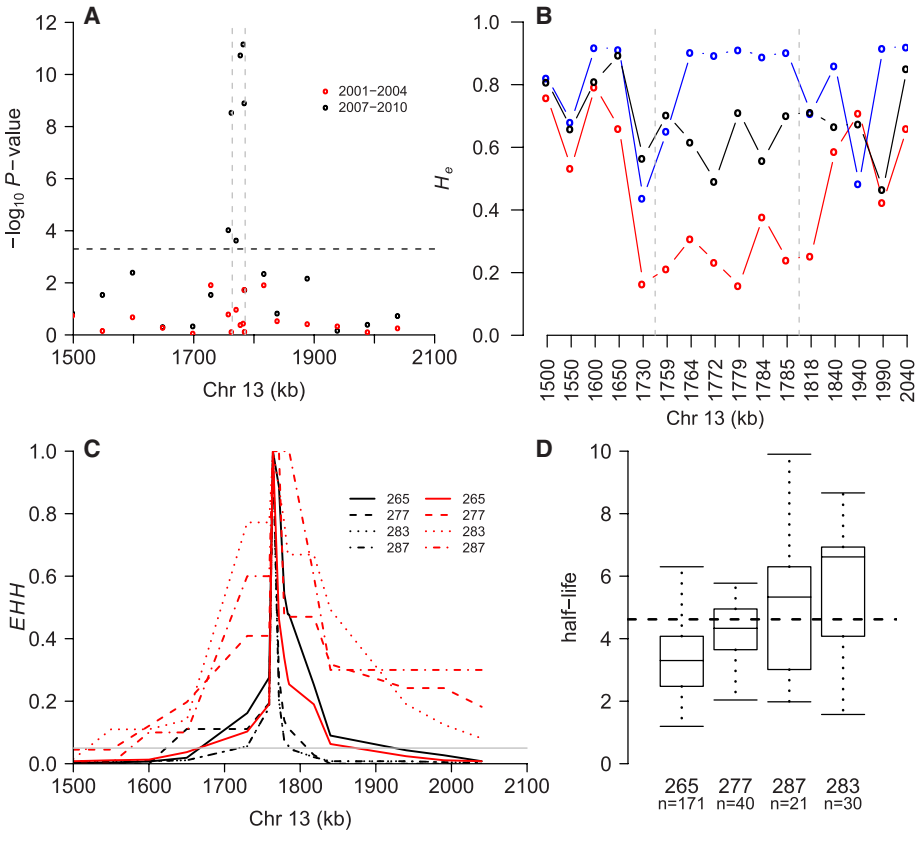

Fig. 4. Fine mapping using 19 microsatellites. (**A**) Association *P* values from the early (2001–2004, red dots) and late (2007–2010, black dots) periods in the northern Thai population. The Bonferroni correction threshold is shown by a horizontal dashed line. (**B**) Comparison of H_e in Thailand, Laos, and Cambodia. (**C**) *EHH* surrounding a microsatellite (position 1,763,950) in slow-clearing (half-life >4.6 hours) and fast-clearing (half-life <2.3 hours) parasites. (**D**) Phenotypic distribution at this locus in 2007–2010 ($P = 4 \times 10^{-12}$).

development of laboratory assays that replicate the CR phenotype. The mapping approach we have used—targeted association of selected genome regions—has broad utility for researchers wishing to map variants responsible for traits under strong recent positive selection.

References and Notes

1. N. J. White, *Science* **320**, 330 (2008).
2. R. T. Eastman, D. A. Fidock, *Nat. Rev. Microbiol.* **7**, 864 (2009).
3. A. M. Dondorp *et al.*, *N. Engl. J. Med.* **361**, 455 (2009).
4. C. Roper *et al.*, *Science* **305**, 1124 (2004).
5. A. P. Phyo *et al.*, *Lancet* **379**, 10.1016/S0140-6736(12)60484-X (2012).
6. T. J. Anderson *et al.*, *J. Infect. Dis.* **201**, 1326 (2010).
7. S. Saralamba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 397 (2011).
8. M. S. Tucker, T. Mutka, K. Sparks, J. Patel, D. E. Kyle, *Antimicrob. Agents Chemother.* **56**, 302 (2012).
9. P. Hunt *et al.*, *BMC Genomics* **11**, 499 (2010).
10. T. Anderson, S. Nkhoma, A. Ecker, D. Fidock, *Pharmacogenomics* **12**, 59 (2011).
11. T. J. Anderson *et al.*, *Mol. Biol. Evol.* **22**, 2362 (2005).
12. R. C. Lewontin, J. Krakauer, *Genetics* **74**, 175 (1973).
13. X. Yi *et al.*, *Science* **329**, 75 (2010).
14. T. S. Simonson *et al.*, *Science* **329**, 72 (2010).
15. M. Mayxay *et al.*, *Am. J. Trop. Med. Hyg.* **86**, 403 (2012).
16. J. C. Tan *et al.*, *Genome Biol.* **12**, R35 (2011).
17. Materials and methods are available as supplementary materials on *Science* Online.
18. P. C. Sabeti *et al.*; International HapMap Consortium, *Nature* **449**, 913 (2007).
19. S. Nair *et al.*, *Mol. Biol. Evol.* **24**, 562 (2007).
20. R. Jambou *et al.*, *Lancet* **366**, 1960 (2005).
21. G. Deplaine *et al.*, *Antimicrob. Agents Chemother.* **55**, 2576 (2011).
22. C. Kidgell *et al.*, *PLoS Pathog.* **2**, e57 (2006).
23. P. J. Bradbury *et al.*, *Bioinformatics* **23**, 2633 (2007).
24. M. Przeworski, G. Coop, J. D. Wall, *Evolution* **59**, 2312 (2005).
25. S. Mok *et al.*, *BMC Genomics* **12**, 391 (2011).

Acknowledgments: Clinical work was funded by the Wellcome Trust. Molecular work was funded by the National Institutes of Health (grant R01 AI048071/AI075145) in facilities constructed with support from Research Facilities Improvement Program grant no. C06 RR013556 from the National Center for Research Resources. We thank patients and staff who contributed to data collection in Thailand, Cambodia (C. Nguon, C. Meng Chuor, and D. Socheat), and Laos (M. Khanthavong, O. Chanthongthip, B. Soonthornsata, T. Pongvongsa, S. Phompida, and B. Hongvanthong); K. Burgoine, P. Singhasivanon, P. Ringwald, M. Zlojutro Kos, and J. Currans' lab. Microarray data have been submitted to the National Center for Biotechnology Information's Gene Expression Omnibus under accession nos. GSM818073 to GSM818239. The authors declare no competing financial interests.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6077/79/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S4
References (*26*–*33*)

31 October 2011; accepted 2 March 2012
10.1126/science.1215966

Extrachromosomal MicroDNAs and Chromosomal Microdeletions in Normal Tissues

Yoshiyuki Shibata,[1]* Pankaj Kumar,[1]* Ryan Layer,[1] Smaranda Willcox,[2] Jeffrey R. Gagan,[1] Jack D. Griffith,[2] Anindya Dutta[1]†

We have identified tens of thousands of short extrachromosomal circular DNAs (microDNA) in mouse tissues as well as mouse and human cell lines. These microDNAs are 200 to 400 base pairs long, are derived from unique nonrepetitive sequence, and are enriched in the 5′-untranslated regions of genes, exons, and CpG islands. Chromosomal loci that are enriched sources of microDNA in the adult brain are somatically mosaic for microdeletions that appear to arise from the excision of microDNAs. Germline microdeletions identified by the "Thousand Genomes" project may also arise from the excision of microDNAs in the germline lineage. We have thus identified a previously unknown DNA entity in mammalian cells and provide evidence that their generation leaves behind deletions in different genomic loci.

Single-nucleotide polymorphisms and copy-number variations are known sources of genetic variation between individuals (*1*–*5*), but there is also great interest in variations that arise during generation of somatic tissues like the mammalian brain, leading to genetic mosaicism between somatic cells. To identify sites of intramolecular homologous recombination during brain development, we searched for extrachromosomal circular DNA (eccDNA) derived from excised chromosomal regions in normal mouse embryonic brains.

We purified eccDNA from nuclei of embryonic day 13.5 (ED13.5) mouse brain and removed linear DNA by digestion with an adenosine 5′-triphosphate (ATP)–dependent exonuclease (*6*) (fig. S1, table S1, and SOM methods). Multiple displacement amplification (MDA) with random primers (*7*, *8*) enriched circular DNA by rolling-circle amplification. The linear products of MDA were sheared to 500–base pair (bp) fragments

[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA. [2]Lineberger Cancer Center, University of North Carolina, Chapel Hill, NC, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: ad8q@virginia.edu

Science

A Major Genome Region Underlying Artemisinin Resistance in Malaria

Ian H. Cheeseman, Becky A. Miller, Shalini Nair, Standwell Nkhoma, Asako Tan, John C. Tan, Salma Al Saai, Aung Pyae Phyo, Carit Ler Moo, Khin Maung Lwin, Rose McGready, Elizabeth Ashley, Mallika Imwong, Kasia Stepniewska, Poravuth Yi, Arjen M. Dondorp, Mayfong Mayxay, Paul N. Newton, Nicholas J. White, François Nosten, Michael T. Ferdig and Timothy J. C. Anderson

Science **336** (6077), 79-82.
DOI: 10.1126/science.1215966

Narrowing Down Artemisinin Resistance

Knowing that antimalarial drug resistance is characterized by selective sweeps and reduced diversity around resistance mutations, **Cheeseman *et al.*** (p. 79) looked for signatures of selection in a modified genome-wide association study in parasite populations from Cambodia, Laos, and Thailand. Thirty-three regions showed evidence of selection and enrichment of known antimalarial resistance genes. Fine-mapping of parasite samples taken during the past decade narrowed the association down to a 35-kb region of seven genes on chromosome 13 that seemed to explain at least 35% of the observed reduction in parasite clearance rate. However, the absence of strong candidate mutations suggests the involvement of noncoding regulatory mutations.