16. A. F. Cowman, S. Karcz, D. Galatis, J. G. Culvenor, *J. Cell Biol.* **113**, 1033 (1991).
17. M. B. Reed, K. J. Saliba, S. R. Caruana, K. Kirk, A. F. Cowman, *Nature* **403**, 906 (2000).
18. M. T. Duraisingh, C. Roper, D. Walliker, D. C. Warhurst, *Mol. Microbiol.* **36**, 955 (2000).
19. S. J. Foote *et al.*, *Nature* **345**, 255 (1990).
20. L. K. Basco, J. Le Bras, Z. Rhoades, C. M. Wilson, *Mol. Biochem. Parasitol.* **74**, 157 (1995).
21. S. C. Chaiyaroj, A. Buranakiti, P. Angkasekwinai, S. Looressuwan, A. F. Cowman, *Am. J. Trop. Med. Hyg.* **61**, 780 (1999).
22. S. K. Martin, A. M. Oduola, W. K. Milhous, *Science* **235**, 899 (1987).
23. R. K. Mehlotra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12689 (2001).
24. D. J. Krogstad, I. Y. Gluzman, B. L. Herwaldt, P. H. Schlesinger, T. E. Wellems, *Biochem. Pharmacol.* **43**, 57 (1992).
25. P. G. Bray, M. Mungthin, R. G. Ridley, S. A. Ward, *Mol. Pharmacol.* **54**, 170 (1998).
26. G. E. Childs *et al.*, *Am. J. Trop. Med. Hyg.* **40**, 7 (1989).
27. L. K. Basco, J. Le Bras, *Am. J. Trop. Med. Hyg.* **48**, 120 (1993).
28. P. G. Bray, S. R. Hawley, M. Mungthin, S. A. Ward, *Mol. Pharmacol.* **50**, 1559 (1996).
29. P. Olliaro *et al.*, *Lancet* **348**, 1196 (1996).
30. M. Adjuik *et al.*, *Lancet* **359**, 1365 (2002).
31. D. De, F. M. Krogstad, F. B. Cogswell, D. J. Krogstad, *Am. J. Trop. Med. Hyg.* **55**, 579 (1996).
32. R. G. Ridley *et al.*, *Antimicrob. Agents Chemother.* **40**, 1846 (1996).
33. R. N. Price *et al.*, *Antimicrob. Agents Chemother.* **43**, 2943 (1999).
34. N. J. White, *Parassitologia* **41**, 301 (1999).
35. M. Foley, L. Tilley, *Pharmacol. Ther.* **79**, 55 (1998).
36. P. M. O'Neill, P. G. Bray, S. R. Hawley, S. A. Ward, B. K. Park, *Pharmacol. Ther.* **77**, 29 (1998).
37. D. J. Krogstad *et al.*, *Science* **238**, 1283 (1987).
38. L. M. B. Ursos, S. Dzekunov, P. D. Roepe, *Mol. Biochem. Parasitol.* **110**, 125 (2000).
39. J. C. Wootton *et al.*, *Nature* **418**, 320 (2002).
40. D. A. Fidock, T. E. Wellems, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10931 (1997).
41. C. B. Mamoun, I. Y. Gluzman, S. Goyard, S. M. Beverley, D. E. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8716 (1999).

42. We thank T. Wellems, P. Roepe, and S. Krishna for many helpful discussions. We also thank A. Talley, R. Cooper, D. Jacobus, S. Ward, P. Bray, W. Ellis, P. Ringwald, and N. Shulman. Supported by National Institute of Allergy and Infectious Diseases grant R01 AI50234, the Speaker's Fund for Biomedical Research, a New Scholar Award in Global Infectious Disease from the Ellison Medical Foundation, and the Howard Hughes Medical Institute Research Resources Program for Medical Schools.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5591/210/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 and S2

16 May 2002; accepted 5 August 2002

Genetic Loci Affecting Resistance to Human Malaria Parasites in a West African Mosquito Vector Population

Oumou Niaré,[1,2]* Kyriacos Markianos,[3]* Jennifer Volz,[5]* Frederick Oduol,[1] Abdoulaye Touré,[2] Magaran Bagayoko,[2] Djibril Sangaré,[2] Sekou F. Traoré,[2] Rui Wang,[5] Claudia Blass,[5] Guimogo Dolo,[2] Madama Bouaré,[2] Fotis C. Kafatos,[5] Leonid Kruglyak,[3,4] Yeya T. Touré,[2] Kenneth D. Vernick[1]†

Successful propagation of the malaria parasite *Plasmodium falciparum* within a susceptible mosquito vector is a prerequisite for the transmission of malaria. A field-based genetic analysis of the major human malaria vector, *Anopheles gambiae*, has revealed natural factors that reduce the transmission of *P. falciparum*. Differences in *P. falciparum* oocyst numbers between mosquito isofemale families fed on the same infected blood indicated a large genetic component affecting resistance to the parasite, and genome-wide scanning in pedigrees of wild mosquitoes detected segregating resistance alleles. The apparently high natural frequency of resistance alleles suggests that malaria parasites (or a similar pathogen) exert a significant selective pressure on vector populations.

Mosquitoes infected with malaria incur quantifiable costs in reproductive fitness for measures such as longevity, fecundity, and flight distance (*1–3*). In addition, mosquito immune genes respond transcriptionally to malaria parasite infection (*4–6*), indicating that malaria parasites are detected by mosquito immune surveillance. Inbred lines selected from mosquito colonies can inhibit parasite development through at least two distinct mechanisms: melanotic encapsulation and intracellular lysis (*7, 8*). Therefore, selective pressure by parasite infection could drive the natural selection of resistance mechanisms in mosquitoes. However, previous studies of resistance used laboratory strains of mosquitoes and malaria parasites, which are known to display biological aberrations in comparison to natural populations (*9–11*), and thus the relevance of these studies to natural malaria transmission may be questioned. Malaria resistance mechanisms and their significance for the biology of malaria transmission have not been examined previously in natural field populations of vectors.

For study of the vector-parasite interaction using field populations, standard genetic techniques that require the creation of inbred or isogenic lines were not applicable. However, the large number of progeny produced by single wild females permitted a novel study design based on isofemale pedigrees, which is quite distinct from designs previously used in human and animal genetics (fig. S1). In nature, *A. gambiae* females mate only once (*12*), and the stored sperm is used to fertilize successive egg batches that mature when the female feeds on blood. We captured pre-mated blood-fed females resting in village dwellings in Bancoumana, Mali (West Africa); allowed them to oviposit in an environmentally controlled chamber; and then raised their respective progeny in the environmental chamber (*13*). We used the isofemale families at either the F_1 or the F_2 stage (the latter produced by mass mating of F_1 mosquitoes). All females of a family were challenged with malaria parasites by being fed on blood from the same *P. falciparum* gametocyte carrier (*14*). Unfed mosquitoes were removed (*15*). At 8 days after infection, the progeny were dissected, and the surviving oocyst-stage parasites were counted in each midgut. The number of oocysts per mosquito constituted the quantitative infection phenotype.

We investigated the genetic dependence of mosquito resistance to malaria infection in two stages. First, in a familiality study, we compared parasite infection levels between mosquito families that were each generated by a different founding pair of wild mosquitoes. The goal was to measure interfamily differences in infection distributions, and therefore only phenotype measurements were used. Groups of three F_1 families were reared together in an environmental chamber and were fed simultaneously on blood from a

[1]Department of Medical and Molecular Parasitology, New York University School of Medicine, 341 East 25th Street, New York, NY 10010, USA. [2]Département d'Epidémiologie des Affectations Parasitaires, Faculté de Médecine, de Pharmacie et d'Odonto-Stomatologie, Boite Postale 1805, Bamako, Mali. [3]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA. [4]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA. [5]European Molecular Biology Laboratories, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

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

single *P. falciparum* gametocyte carrier. The three families constituted a comparison group; within such a group, environmental differences could be discounted. We collected six such three-family comparison groups. If there were no segregating mosquito genes that influence the infection phenotype, then parasite numbers in the different families of the same group should appear to be drawn from the same distribution. Conversely, the action of genes that affect resistance should cause family-specific changes in the distribution of parasite numbers. We performed pairwise comparisons (a:b, a:c, and b:c) of infection phenotype for the three families of each comparison group. Because the phenotypic distribution of oocyst infection in mosquitoes tends to be non-normal (*16–18*), we used the nonparametric Kolmogorov-Smirnov (KS) test to evaluate significance [see methods (*13*)]. Of 18 pairwise comparisons, 9 were significantly different at $P < 0.05$, including 5 that were significant at $P < 0.01$ and 3 that were significant at $P < 0.001$ (table S1). There was at least one significant pairwise difference in all but one of the comparison groups.

Thus, the familiality study provided strong evidence that alleles segregating in a natural population of *A. gambiae* have a significant effect on susceptibility to *P. falciparum*, and that such alleles are collectively frequent. This information has not been available for natural populations of malaria vectors and was further validated in the second stage of this study. In the second stage, we performed a genome-wide scan for loci that influence resistance or susceptibility. Two independent wild isofemale F_2 families were infected on the blood of different gametocyte carriers, and then we determined phenotypes and genotypes of each mosquito at a broadly distributed set of 24 microsatellite loci (fig. S2).

Because the genetic size of the *A. gambiae* genome is 215 centimorgans (cM) (*19*), the resolution of the scan was approximately 9 cM. At this marker density, an unknown locus of interest is, on average, 4.5% recombination distance (~6 Mb) away from a marker locus. For the genome scan, we compared parasite count distributions for mosquitoes that shared a single allele or allelic combination (genotype) versus all other mosquitoes in the same family. We used F_2 families because they offered a substantial increase in the number of mosquitoes per family, as compared to F_1s. It should be noted that although either generation (F_1 or F_2) has a maximum of four alleles sampled from the natural population, the maximum number of genotypes per locus is four for the F_1 and 10 for the F_2 generation.

Table 1 presents the results from the two F_2 families. At every marker, we split the members of the F_2 generation into two samples, using a genotype or allele observed at that marker as the selection criterion. For example, if six genotypes were observed in the F_2 sample, we performed six genotype tests for the marker: We compared the phenotype distribution for each genotype against the pooled phenotype distribution of all other genotypes. For every test, we computed the probability of encountering the observed difference under the null hypothesis of no linkage between this marker and malaria resistance (*20*). As was the case in the familiality study, we used a nonparametric test to compare phenotype distributions. For each family, we performed approximately 60 allele tests and 100 genotype tests. To adjust for multiple testing, we used an empirical permutation test (*21*, *22*). Statistical test results were interpreted according to accepted criteria (*23*).

Both of the F_2 families displayed significant linkage between parasite count and DNA markers (Table 1). In family 98BF214 (n = 83 mosquitoes), one of the two observed marker alleles at the chromosome 2L marker *H603* had a strong correlation with infection phenotype (genome-wide P value = 4×10^{-7}; phenotype distributions are shown in Fig. 1A). The effect was seen in both genotypes that included allele 1, which suggests that this marker is located close to a locus where the resistance allele is semidominant. The putative resistance allele linked to marker allele 1 can explain almost all of the parasite-free mosquitoes (24 out of 27) in the family. The linked resistance locus is provisionally named *Pfin1*, for *P. falciparum infection intensity 1*. In family 97F2B4A5 (n = 82), a different marker (*H290* on chromosome 2R) showed significant linkage between genotype and parasite count (Fig. 1B). Here, one of the eight genotypes at *H290* (genotype 2–4) showed significant resistance (genome-wide P value = 0.025), suggesting recessive inheritance of resistance (*24*). This locus is provisionally named *Pfin2*.

Genetic control of the previously described encapsulation-mediated resistance mechanism in *A. gambiae* (*7*) was composed of three loci, where the major locus, *Pen1*, was closely linked to marker *H175* and near marker *H290* (*19*, *25*). Despite this genetic proximity, it is unlikely that *Pfin2* is the same as *Pen1* for two reasons. First, the phenotypes are distinct. The influence of *Pen1* was expressed as melanotic encapsulation of malaria ookinetes, with no effect on infection intensity (*25*), whereas the influence of *Pfin2* was exclusively expressed as variation in infection intensity, without melanotic encapsulation. Second, the encapsulation trait largely controlled by *Pen1* was strongest against allopatric (non-African) strains of *P. falciparum* but had a much reduced effect on African parasite strains (*7*), whereas *Pfin2* was identified specifically because of its action against sympatric parasites. It is likely that *Pfin2* and *Pen1* are distinct loci, perhaps in a resistance gene cluster, although the possibility cannot currently be excluded that they might represent disparate alleles at the same locus.

The *Pfin1* marker *H603* lies within the large chromosome inversion 2La (*19*). Karyotype analysis of polytene ovarian nurse cell chromosomes showed that family 98BF214 was fixed for the homokaryotypic inverted 2La/a chromosome arrangement

Table 1. Statistical tests for genetic loci controlling midgut oocyst number. F_2 families of *A. gambiae* from wild single-pair crosses were fed on natural *P. falciparum*–infected blood, and individual microsatellite genotypes were determined at a resolution of ~9 cM. Nominal (P^n) and genome-wide (corrected for multiple testing) P values are shown for the KS test. Genome-wide P values were generated by randomizing the correlation between marker genotype and phenotype distributions. We repeated the genome scan in 10^5 (or 10^7 for 98BF214 allele 1) randomized data sets.

Family 98BF214, allele and genotype test. Female family members were n = 83. At chromosome 2L, marker *H603*, three genotypes were observed: 1-2 (n = 44), 2-2 (n = 33), and 1-1 (n = 6). Fifty-eight allele tests and 100 genotype tests were performed.

	Nominal	Genome-wide
Presence of allele 1 (n = 50) versus all other genotypes (n = 33)	$P^n = 6.2 \times 10^{-8}$	$P = 4 \times 10^{-7}$
Genotype 1-2 (n = 44) versus all other genotypes (n = 39)	$P^n = 1.2 \times 10^{-5}$	$P = 1.9 \times 10^{-3}$

Family 97F2B4A5, genotype test. Female family members were n = 82. At chromosome 2R, marker *H290*, eight genotypes were observed: 3-4 (n = 6), 2-3 (n = 18), 1-2 (n = 19), 2-2 (n = 4), 1-3 (n = 14), 1-1 (n = 4), 1-4 (n = 9), and 2-4 (n = 8). One hundred genotype tests were performed. Allele tests for recessive traits are not informative.

	Nominal	Genome-wide
Genotype 2-4 (n = 8) versus all other genotypes (n = 74)	$P^n = 0.00097$	$P = 0.023$

(*26*). Thus, the linkage between *H603* and resistance was due to linkage between two freely segregating loci and not due to disequilibrium by suppression of recombination within a polymorphic inversion.

Outbred animal populations have been used previously for fine mapping of traits (*27*, *28*), but rarely for initial gene discovery by genome scan in field populations (*29*).

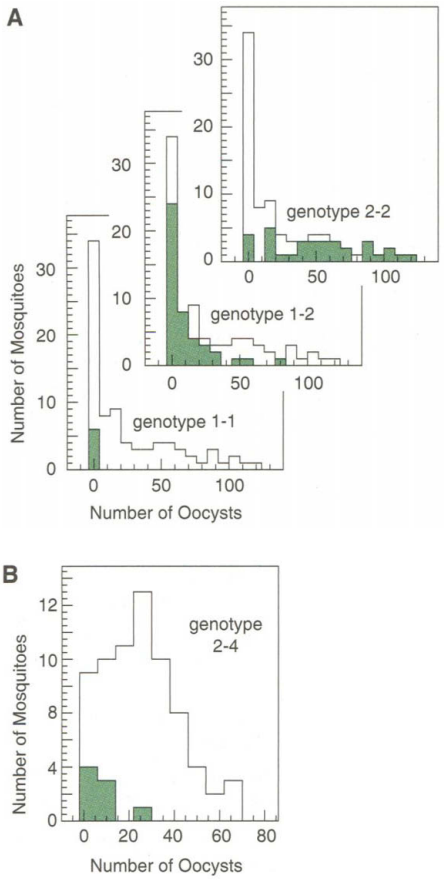

Fig. 1. Infection phenotype of wild *A. gambiae* pedigrees with natural *P. falciparum*. Histograms show the infection phenotype of F_2 isofemale families by microsatellite marker genotype. Green bars show phenotypic distributions for the genotype indicated in the panel. Open bars show the overall phenotypic distribution for all phenotypes of the family combined. **(A)** Family 98BF214. Green bars show the phenotype for the indicated genotype at marker *H603*. Mean oocyst number per mosquito for genotype 1-1 (n = 6) is 0.17 oocyst; for genotype 1-2 (n = 44), 9.9 oocysts; for genotype 2-2 (n = 33), 50.6 oocysts. **(B)** Family 97F2B4A5. Green bars show the phenotype for genotype 2-4 (n = 8) at marker *H290*. This family has seven other genotypes (Table 1). Mean oocyst number per mosquito for genotype 2-4 (n = 8) is 8.0 oocysts; for all other genotypes (n = 74), it is 28.6 oocysts. In all panels, the width of the bars represents a range of eight counts per bin. The first bar is centered on 0 and has an effective width of 0 to 3 oocysts. The second bar includes the bin from 4 to 11 oocysts, and subsequent bins follow the same pattern.

The novelty of the current study design lies in the mass mating step used to produce the F_2 generation. We chose the F_2 study design because it offers a large sample size. Although mass mating among F_1 mosquitoes leads to loss of some inheritance information (such as paternal relationships), phenotypic effect and marker informativeness were sufficiently large to help us identify genomic regions with a strong influence on vector-parasite interaction. We demonstrate that field populations of *A. gambiae* exhibit significant variation in permissiveness for parasite development, and we point to two genomic regions that affect this trait.

It is likely that there are other such resistance loci, and the strategy we present can be used to more extensively screen natural mosquito populations. By implementation of a modified protocol to maintain pedigrees beyond the initial genome scan of the F_2 generation, segregating resistance alleles could be mapped at high resolution and positional candidates could be identified. Extant pedigrees would also permit characterization of the mechanisms of resistance. In the current work, we detected resistance as a reduction in oocyst number 8 days after a blood meal. This is an aggregate phenotype that summarizes all preceding events in parasite development, and it is likely that different resistance alleles cause developmental blockades at different critical points. It is also possible that some resistance mechanisms may interact or synergize with human host factors such as transmission-blocking antibodies or cytokines in the infecting blood meal.

It would be of interest to identify the molecular nature of these resistance loci, which are probably rapidly evolving components of the genome that define the points of greatest adaptive friction between parasite virulence factors and the mosquito host. In plants, most genetically identified resistance genes are pattern-recognition receptors for pathogen virulence factors (*30*, *31*), but comparable information is lacking in mosquitoes. Continued parasite transmission by a vector with a high frequency of segregating resistance factors suggests that the parasite has made adaptive responses to mosquito resistance, perhaps by evolving multigenic and/or polymorphic virulence factors for the insect stages of the life cycle analogous to those found in asexual-stage parasites (*32*). Virulence in the mosquito host is under parasite genetic control (*33*), but specific virulence factors are not known.

Malaria-infected mosquitoes in nature typically carry fewer than 10 oocysts (*18*, *34*). Far higher parasite intensities can be achieved in laboratory infections of mosquitoes, particularly when genetically selected susceptible lines are used (*7*, *8*). Natural resistance alleles that limit parasite development in the vector, such as those described here, may prove to be an important factor underlying the small numbers of oocysts observed in wild infected mosquitoes. Genetic studies in the field as well as in the laboratory will be needed to elucidate the mechanisms that can limit the propagation of *P. falciparum* in *A. gambiae* and their respective importance for natural transmission of the disease.

References and Notes

1. R. A. Anderson, B. G. Knols, J. C. Koella, *Parasitology* **120**, 329 (2000).
2. J. A. Hopwood, A. M. Ahmed, A. Polwart, G. T. Williams, H. Hurd, *J. Exp. Biol.* **204**, 2773 (2001).
3. B. A. Schiefer, R. A. Ward, B. F. Eldridge, *Exp. Parasitol.* **41**, 397 (1977).
4. F. Oduol, J. Xu, O. Niare, R. Natarajan, K. D. Vernick, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11397 (2000).
5. A. M. Richman, G. Dimopoulos, D. Seeley, F. C. Kafatos, *EMBO J.* **16**, 6114 (1997).
6. S. Bonnet, G. Prevot, J. C. Jacques, C. Boudin, C. Bourgouin, *Cell Microbiol.* **3**, 449 (2001).
7. F. H. Collins *et al.*, *Science* **234**, 607 (1986).
8. K. D. Vernick *et al.*, *Exp. Parasitol.* **80**, 583 (1995).
9. K. P. Day *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8292 (1993).
10. D. E. Norris, A. C. Shurtleff, Y. T. Toure, G. C. Lanzaro, *J. Med. Entomol.* **38**, 336 (2001).
11. L. Pologe, *Mol. Biochem. Parasitol.* **68**, 35 (1994).
12. F. Tripet *et al.*, *Mol. Ecol.* **10**, 1725 (2001).
13. Materials and methods are available as supporting material on *Science* Online.
14. Experimental protocols involving human subjects were approved by the ethical review boards of the New York University School of Medicine (protocol H-7270), the Faculté de Médecine, de Pharmacie et d'Odonto-Stomatologie of Mali (protocol 2001-0007), and the U.S. National Institutes of Health (protocol S-16077).
15. Physical removal of unfed mosquitoes was verified at the time of dissection by determination that all mosquitoes displayed stages of ovarian development that require a blood meal. Thus, the absence of any oocysts resulted from failure of parasite development rather than lack of feeding.
16. G. Pichon, V. Robert, J. P. Verhave, *Parasite J. Soc. Fr. Parasitol.* **3**, 161 (1996).
17. G. F. Medley *et al.*, *Parasitology* **106**, 441 (1993).
18. P. F. Billingsley, G. F. Medley, D. Charlwood, R. E. Sinden, *Am. J. Trop. Med. Hyg.* **51**, 260 (1994).
19. L. Zheng, M. Q. Benedict, F. C. Kafatos, *Genetics* **143**, 941 (1996).
20. Linkage analysis was performed with software written specifically for this study (Outbred Lines, K.M. and L.K., unpublished). We did not use standard linkage analysis programs used for family studies because paternity and maternity information for the F_2 generation was lost because of mass mating of F_1 family members. Mass mating was necessary because *A. gambiae* breed very inefficiently as isolated pairs. The program reads phenotype and genotype information and automatically performs phenotype distribution comparisons between mosquitoes that carry a specific genotype or allele against the pooled phenotype distribution of all other family members. Under the null hypothesis of no correlation between genotype and phenotype, phenotype distributions for a genotype or allele would be random picks from the overall phenotype distribution. The nominal P value reported here is the probability that the two distributions were randomly drawn from the same parent distribution. Nominal P values were computed using the KS test, and similar values were obtained using the Wilcoxon-Mann-Whitney ranking test.
21. G. A. Churchill, R. W. Doerge, *Genetics* **138**, 963 (1994).
22. We performed a large number of tests per family and therefore had to correct for multiple testing. The standard Bonferroni correction (number of tests performed times nominal P value) is overly conservative because the tests are not independent: The markers

are linked, multiple comparisons at the same locus are correlated, and not all DNA markers are equally informative. Instead of a Bonferroni correction, we used an empirical permutation test. We created replicates of the observed sample, in which genotype and phenotype distributions were preserved, but any linkage between genotype and phenotype was removed by randomly reassigning the observed phenotypic values. We then asked how often any replicates that conformed to the null hypothesis (the independence of genotype and phenotype) produced P values in excess of the P value observed for the real data. This test preserved both the observed marker informativeness and the overall phenotype distribution.
23. E. Lander, L. Kruglyak, *Nature Genet.* **11**, 241 (1995).
24. Linkage of the resistance allele to two distinct marker alleles, 2 and 4, could be explained by recombination between resistance and marker alleles or by multiple entry of the resistance allele into the family. An attempt to distinguish by reconstructing flanking markers, although not conclusive, appeared more consistent with the recombination hypothesis.
25. L. Zheng *et al.*, *Science* **276**, 425 (1997).
26. S. F. Traoré, O. Niare, K. D. Vernick, data not shown.
27. J. Flint, R. Mott, *Nature Rev. Genet.* **2**, 437 (2001).
28. C. J. Talbot *et al.*, *Nature Genet.* **21**, 305 (1999).
29. C. L. Peichel *et al.*, *Nature* **414**, 901 (2001).
30. E. B. Holub, *Nature Rev. Genet.* **2**, 516 (2001).
31. J. Bergelson, M. Kreitman, E. A. Stahl, D. Tian, *Science* **292**, 2281 (2001).
32. K. W. Deitsch, E. R. Moxon, T. E. Wellems, *Microbiol. Mol. Biol. Rev.* **61**, 281 (1997).
33. H. M. Ferguson, A. F. Read, *Proc. R. Soc. London Ser. B* **269**, 1217 (2002).
34. G. Pringle, *Trans. R. Soc. Trop. Med. Hyg.* **60**, 626 (1966).

35. We acknowledge the support of O. Doumbo and the Equipe Parasitologique (FMPOS, Mali) and the organizational assistance of S. Karambe, R. Sakai, and R. Gwadz. Supported by a grant from the National Institute of Allergy and Infectious Disease/NIH to K.D.V., a DOE/Sloan Foundation Fellowship in Computational Molecular Biology to K.M., DFG Sonderforscungbereich 544 to F.C.K., a Ph.D. Fellowship from the University of Heidelberg to J.V., and a James S. McDonnell Centennial Fellowship to L.K.

Supporting Online Material
www.sciencemag.org/cgi/content/full/298/5591/213/DC1
Materials and Methods
Figs. S1 and S2
Table S1
Data Sets

30 April 2002; accepted 9 September 2002

Excess Polymorphisms in Genes for Membrane Proteins in *Plasmodium falciparum*

Sarah K. Volkman,[1] Daniel L. Hartl,[2] Dyann F. Wirth,[1] Kaare M. Nielsen,[2,3] Mehee Choi,[2] Serge Batalov,[4] Yingyao Zhou,[4] David Plouffe,[4] Karine G. Le Roch,[4] Ruben Abagyan,[4] Elizabeth A. Winzeler[4,5]*

The detection of single-nucleotide polymorphisms in pathogenic microorganisms has normally been carried out by trial and error. Here we show that DNA hybridization with high-density oligonucleotide arrays provides rapid and convenient detection of single-nucleotide polymorphisms in *Plasmodium falciparum*, despite its exceptionally high adenine-thymine (AT) content (82%). A disproportionate number of polymorphisms are found in genes encoding proteins associated with the cell membrane. These genes are targets for only 22% of the oligonucleotide probes but account for 69% of the polymorphisms. Genetic variation is also enriched in subtelomeric regions, which account for 22% of the chromosome but 76% of the polymorphisms.

The complete genomic sequence of *P. falciparum* has been determined and is in the final stages of assembly and annotation (*1–3*). The great challenge now is how best to use the genome sequence for public health and clinical applications. One approach is to identify single-nucleotide polymorphisms (SNPs) in order to pinpoint the origin and map the spread of contagious diseases, to identify and track new mutations that confer resistance to drugs or vaccine-induced immunity, and potentially to identify candidate genes for novel therapeutic or immunological intervention. Although SNP detection on a genome-wide scale is technically difficult, we reasoned that it might be feasible to detect SNPs in *P. falciparum* by means of high-density oligonucleotide arrays, even though such arrays were originally developed for gene-expression studies.

Owing to the relatively short probe sequences used for oligonucleotide arrays, the strength of hybridization between a probe and its target sequence depends largely on perfect complementarity between the probe and the target. An SNP or a small deletion or insertion in the target will reduce the hybridization signal (*4*). Because the exact genomic position of each probe is known, the location of variant sequences can be found by comparing the intensity of oligonucleotide hybridization between genomic DNA from an unknown strain and that from the 3D7 reference strain of *P. falciparum*, from whose genomic sequence the oligonucleotides were designed (*5*). There are two major technical obstacles to the use of oligonucleotide arrays for *P. falciparum*. First, the AT content of the genome is unusually high even in coding regions, and in some genes the AT content of the third positions of codons is nearly as high as the genetic code allows. Second, preparations of DNA from *P. falciparum* may also contain quantities of human DNA.

The feasibility of the approach was tested with the complete sequence of chromosome 2, which contains 210 annotated genes (*6*). Using an oligonucleotide selector algorithm, we chose a unique 25-nucleotide (nt) sequence to match protein-coding sequences at intervals of ~200 nucleotides (yielding between 2 and 117 probes per gene, depending on size). The probes were designed to have similar melting temperatures and to avoid runs of As and Ts but otherwise to be as different from each other as possible. A total of 4167 single-stranded probes were designed (*7*), manufactured by means of photosensitive DNA synthesis technology (*8*), and positioned by Affymetrix onto a prototype array that also included 395,833 probes for ~80,000 different cDNAs from human tissues.

To evaluate the robustness and specificity of the *P. falciparum* probes and to rule out artifacts due to contaminating human DNA, we prepared genomic DNA from different parasite isolates that had been cultured in human erythrocytes, then labeled the DNA and hybridized it to the oligonucleotide array. After hybridization, the integrated intensity (an estimate of the copy number) was determined for each of the 210 *P. falciparum* genes probed on the array (Fig. 1). The mean of the integrated 3D7 intensity for probes at the chromosome ends was higher than for probes within the central region of the chromosome, presumably because of duplication of some of these probe sequences. Even though oligonucleotide probes are thought to differ substantially in their hybridization properties, the integrated signals were generally consistent, varying by not much more than a factor of 2 (fig. S1). In contrast, for the W2 isolate, the integrated intensity across a region on the left arm of the chromosome was 30- to 50-fold lower than the mean integrated intensity for genes in the central region of the

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA. [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Pharmacy, University of Tromso, Tromso N-9037, Norway. [4]Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA. [5]Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: winzeler@gnf.org