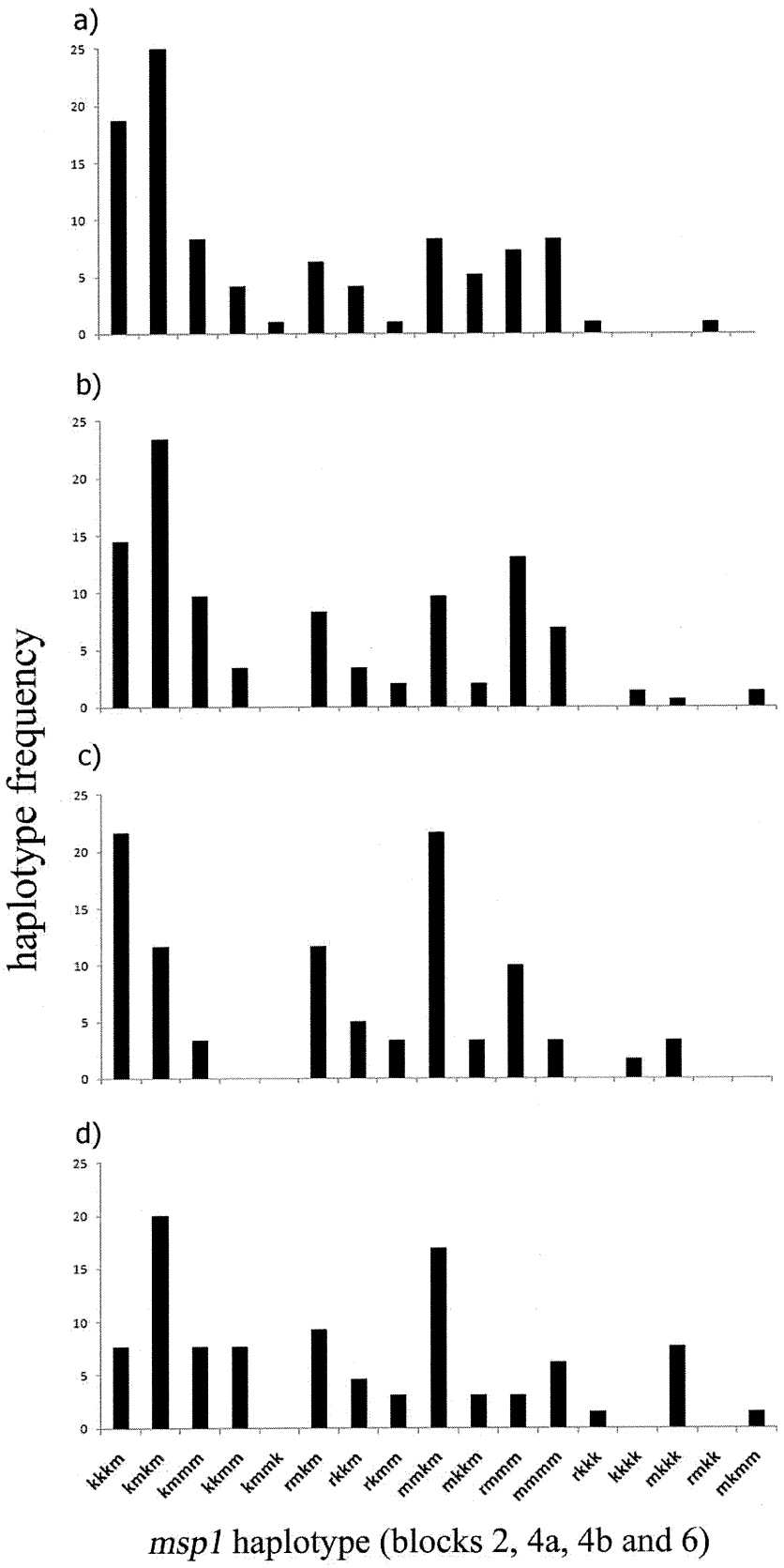
a)
b)
c)
d)
haplotype frequency
0
5
10
15
20
25
kkkm
kmkm
kmmm
kkmm
kmmk
rmkm
rkkm
rkmm
mmkm
mkkm
rmmm
mmmm
rkkk
kkkk
mkkk
rmkk
mkmm
msp1 haplotype (blocks 2, 4a, 4b and 6)

Figure 3. *msp1* haplotype frequencies of *Plasmodium falciparum* parasites in the urban (a) and peri-urban (b) areas of Brazzaville, and in Gamboma (d) and Pointe-Noire (c).
doi:10.1371/journal.pone.0023430.g003

between the urban and peri-urban regions, the percentage similarity was determined for both Point-Noire and Gamboma against a combined "Brazzaville" population. The haplotype pattern was less similar between Brazzaville and Pointe-Noire (Ps = 0.71, 95% range of permutation, 0.73–0.88) and Brazzaville and Gamboma (Ps = 0.77, 95% range of permutation, 0.70–0.86), than it was between the urban and peri-urban regions of Brazzaville.

Population structure, parasite population genetic diversity and linkage disequilibrium analysis based on microsatellite markers

In order to determine whether the parasites in circulation in the peri-urban and urban areas were discrete populations or in panmixis, an F_{ST} value was determined based on the microsatellite markers analyzed using FSTAT v.2.9.4. F_{ST} for the combined population was 0.004, (SE 0.005), a very low value indicative of panmixia. There was no significant population differentiation between areas (G-test, 10,000 randomizations, $P = 0.1$). h was calculated for the two populations based on the mean genetic diversity of all microsatellite loci using LIAN v3.5. For the urban area, $h = 0.81$, SE = 0.04, and for the peri-urban area, $h = 0.83$, SE = 0.04. The effective population sizes (N_e) based on these h values were 6703 (95% CI, 2880–15,269) for the urban area, and 7676 (95% CI, 3299–17,486) for the peri-urban area. Individual values of h were also determined for each microsatellite loci as above, and are shown in Table 1. There were no significant differences in the h values of any loci between the two areas.

The degree of linkage disequilibrium was determined for both populations using LIAN v3.5. For the urban area, $I^S_A = -0.0034$ ($P = 0.585$), and for the peri-urban area $I^S_A = -0.0119$ ($P = 0.786$) indicating the complete absence of linkage disequilibrium in both areas. Pair-wise tests of linkage disequilibrium between pairs of microsatellite markers were performed using FSTAT v.2.9.4. There was no significant linkage between the majority of markers for either the urban or peri-urban areas, with the exceptions of TA53 (Chromosome (Chr) 9) and Poly-α (Chr 4) in the peri-urban area and in the combined population (considering both the peri-urban and the urban areas as one population) ($P<0.01$); TA60 (Chr 13) and Poly-α, also in the peri-urban area ($P = 0.05$); TA60 and TA53 in the urban area ($P = 0.01$); and TA43 (Chr14) and TA81 (Chr5) in the combined population ($P = 0.03$). There were no significant differences in percentage similarities (Ps) between samples from the urban and the peri-urban regions for each individual marker (Table 1).

Allele frequencies of genes associated with resistance to chloroquine and pyrimethamine

The prevalence of drug-resistance associated mutations in the *crt* and *dhfr* genes, linked to resistance to pyrimethamine and chloroquine respectively, were assessed by PCR amplification and direct sequencing of products. For *dhfr*, we found a high proportion of parasites from both regions carried mutations at amino acid positions 51, 59 and 108, but that the 164 (I → L), mutation, known currently to be rare in Africa, was not observed. For *crt*, a high prevalence of parasites carrying the amino acid position 76 (K → T) was observed in both areas. Allele frequencies were determined using a maximum likelihood algorithm carried out using *MalHaploFreq* [45], and are shown in Table 4. We found that there was a significant difference in the allele frequencies at the *crt* locus between the urban and peri-urban areas, with significantly fewer wild-type parasites in circulation in the peri-urban area (log likelihood ratio test $\chi^2 = 5.19$, $P = 0.023$, d.f. = 1). There was also a significant difference in the allele frequencies of *dhfr* mutations between the two areas, with the highly resistant triple mutant more common in the peri-urban area (Log Likelihood ratio test $\chi^2 = 19.4$, $P<<0.01$, d.f. = 4

Discussion

Our results indicate that the *Plasmodium falciparum* parasite population in contiguous areas of urban and peri-urban Brazzaville, Republic of Congo circulates freely between both areas, and forms one population. Within this setting, we found no differences in the multiplicities of infections (MOI) of patients, parasite genetic diversity or linkage disequilibrium between the two areas, despite the large variation in the intensity of transmission between them (2–12 infective mosquito bites per person per year (ib/p/y) for the urban area, and ≈ 50 ib/p/y in the peri-urban area). It seems very likely that the relatively low transmission rates in the urban area result in a lower degree of immunity in the population residing there in comparison to those residing in the peri-urban area. This is supported by our data in two ways; firstly, by the larger proportion of patients displaying sub-microscopic parasite infections in the peri-urban area (20.1% of all PCR positive patients, compared with 11% in the urban area), and secondly by the fact that mean age of the sub-microscopic carriers was higher in the urban than in the peri-urban area (21.0 vs 15.4, respectively, although this was not significant, probably due to the low numbers of sub-microscopic carriers in the urban area). Increased vectorial capacity is thought to lead to higher strain diversity [10,55], with the result that it is expected that greater parasite genetic diversity will be found in areas with higher transmission rates. However, our results show that parasite genetic diversity is uniform throughout an area with varying transmission rates, presumably due to gene flow within a population in panmixis. As immunity to malaria is thought to be acquired in a strain-specific manner [6,7,8], then, given homogeneity of parasite strain diversity within a given area, transmission intensity becomes the major factor for determining the rate at which immunity is achieved against the parasite population circulating within it. One possible consequence of increasing urbanisation is that decreasing transmission intensity may lead to a decrease in parasite genetic diversity, and so increase the rate at which immunity may be achieved by the population. However, our results indicate that parasite genetic diversity may not alter significantly, when the parasite population extends to the surrounding peri-urban areas, which seems the likely situation when considering the process of urbanisation, in which highly urbanized centers extend out into the less urbanized peripheries. This suggests that the acquisition of immunity in urban areas will take longer than it will in peri-urban and rural areas, and will lead to increased risk of symptomatic malaria for a longer period of childhood for those living there. Our results are in agreement with other surveys of the relationship between transmission rates and parasite genetic diversity [15,16,17], but in disagreement with others [12,13]. In most of these studies, regions separated by large distances were compared; whereas our study is the first to compare contiguous regions where the potential of genetic transfer between high and low transmission areas exists.

Table 4. Allele frequencies at the *dhfr* and *crt* loci for the urban and peri-urban areas, estimated using MalHaploFreq based on prevalence data combined with the number of clones per sample (determined by haplotyping *msp1*).

Allele	Frequency (95% CI)		*P*-value
dhfr	Peri-urban	Urban	
NCSI (wild type)	0.02 (0.00–0.05)	0.02 (0.00–0.07)	0.001
ICNI (double mutant)	0.29 (0.21–0.39)	0.36 (0.24–0.49)	
NRNI (double mutant)	0	0.13 (0.06–0.23)	
IRNI (triple mutant)	0.68 (0.59–0.77)	0.50 (0.37–0.63)	
crt			
CVMNK (wild type)	0.03 (0.01–0.07)	0.12 (0.06–0.20)	0.023
CVIET (mutant)	0.97 (0.92–0.99)	0.88 (0.79–0.94)	

P-values were determined using log-liklelihood tests.
doi:10.1371/journal.pone.0023430.t004

We found significantly greater selection of parasites carrying drug-resistance associated alleles of *dhfr* and *crt*, linked to SP and CQ resistance, respectively, in the peri-urban area compared to the urban area. This result fits with the predictions of some theoretical models that drug resistance will be selected predominantly in high transmission areas [19], and with empirical evidence of the same trend from studies conducted in Western Uganda [34]. In these latter studies, the prevalence of *dhfr* mutants increased directly with transmission intensity, while the prevalence of *crt* mutations, whilst highest at high transmission areas, and lower at intermediate levels, increased again at the low transmission extreme. It seems likely that the "low" transmission setting (the urban area) in this study is analogous to the intermediate setting of the Ugandan study in terms of transmission intensity, and so the two data sets concur.

Our data does not appear to support the hypothesis that increased host immunity associated with high transmission areas can act as a "refuge" for wild type parasites [20], at least not in regions with a parasite pool shared between areas of low and high transmission. This hypothesis is based on a number of very reasonable assumptions; i) that high host immunity may reduce the selection advantage of resistant parasites in the presence of drugs [27], ii) that increased parasite genetic diversity leads to more clones per infection, and a greater opportunity for intra-host competition to select for sensitive parasites in the absence of drugs due to fitness costs associated with resistance [24] iii) that at high levels of population immunity, less antimalarial drugs are used per infection due to host immunity limiting the severity of the symptoms of malaria [20,56]and iv) that in high transmission areas, higher recombination rates lead to faster breakdown of drug resistance caused by multi-loci mutations. In our study, MOI and parasite genetic diversity was not significantly different between the high and low transmission regions, presumably due to the sharing of the panmitic parasite pool between them. Thus, it is not likely that there is differential selection pressure mediated by intra-host competition between these areas. Similarly, we found no differences in linkage disequilibrium between areas, indicative of comparable rates of recombination, again probably due to the shared parasite population, thus there should be equal opportunities for the break-up of multi-locus resistance genotypes in both areas. We did not directly measure drug usage in both areas, so we do not know if there is less drug use in the peri-urban area compared to the urban area (as hypothesized in point iii), above), although this seems unlikely given that almost equal proportions of patients from both areas presenting to the health centers with fevers had taken some form of anti-malarial treatment themselves at home prior to presentation (Table 3). Given this, we consider that the most parsimonious explanation for the differences in the frequencies of drug resistance associated alleles of *dhfr* and *crt* between the urban and peri-urban areas is due primarily to greater drug pressure in the peri-urban area due to greater parasite prevalences. Thus our data suggests that urbanisation may lead to a decrease in the selection pressure for drug resistance by driving down transmission rates while not affecting parasite genetic diversity.

An alternative explanation for the observation of greater selection for drug resistance in the high transmission areas is that although moderately high levels of transmission may indeed provide a refuge for wild type parasites [20,56], at very high levels of transmission, a threshold may be crossed at which point the resistant parasites again gain a selection advantage. This might happen when the duration of infections increases to the point that any fitness incurred by the drug resistant parasite is counter-balanced by the disadvantage of being sensitive to the drugs still applied to the population [23]. It may well be that the field setting considered here is more of a moderate versus very high transmission, than of a low versus high transmission.

In conclusion, we have shown, using a field setting in which contiguous urban and peri-urban areas sharing a parasite population but with very different levels of malaria transmission intensity, that urbanisation is likely to lower transmission rates without affecting parasite genetic diversity, and that this may lead to a reduced pressure for the selection of drug resistance. Targeting peri-urban areas with appropriate treatment strategies would, therefore, be important in minimizing the risk of the selection of drug resistance.

Acknowledgments

We thank all the patients who contributed samples for this study, and to the health workers at Tenrikyo and Madibou health centers.

Author Contributions

Conceived and designed the experiments: RC. Performed the experiments: YT RC NH RI HU MI. Analyzed the data: YT RC TS KT. Contributed reagents/materials/analysis tools: MN PC OK. Wrote the paper: RC YT. Drafted the manuscript: SN.

References

1. Nations U (2002) World Urbanization Prospects: The 2001 Revision. In: Population Division DoEaSA, United Nations Secretariat, ed. New York: United Nations.
2. Donnelly MJ, McCall PJ, Lengeler C, Bates I, D'Alessandro U, et al. (2005) Malaria and urbanization in sub-Saharan Africa. Malar J 4: 12.
3. Coene J (1993) Malaria in urban and rural Kinshasa: the entomological input. Med Vet Entomol 7: 127–137.
4. Omumbo JA, Guerra CA, Hay SI, Snow RW (2005) The influence of urbanisation on measures of Plasmodium falciparum infection prevalence in East Africa. Acta Trop 93: 11–21.
5. Contamin H, Fandeur T, Rogier C, Bonnefoy S, Konate L, et al. (1996) Different genetic characteristics of Plasmodium falciparum isolates collected during successive clinical malaria episodes in Senegalese children. Am J Trop Med Hyg 54: 632–643.
6. Forsyth KP, Philip G, Smith T, Kum E, Southwell B, et al. (1989) Diversity of antigens expressed on the surface of erythrocytes infected with mature Plasmodium falciparum parasites in Papua New Guinea. Am J Trop Med Hyg 41: 259–265.
7. Marsh K, Otoo L, Hayes RJ, Carson DC, Greenwood BM (1989) Antibodies to blood stage antigens of Plasmodium falciparum in rural Gambians and their

relation to protection against infection. Trans R Soc Trop Med Hyg 83: 293–303.
8. Newbold CI, Pinches R, Roberts DJ, Marsh K (1992) Plasmodium falciparum: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. Exp Parasitol 75: 281–292.
9. Arnot D (1998) Unstable malaria in Sudan: the influence of the dry season. Clone multiplicity of Plasmodium falciparum infections in individuals exposed to variable levels of disease transmission. Trans R Soc Trop Med Hyg 92: 580–585.
10. Gupta S, Trenholme K, Anderson RM, Day KP (1994) Antigenic diversity and the transmission dynamics of Plasmodium falciparum. Science 263: 961–963.
11. Paganotti GM, Babiker HA, Modiano D, Sirima BS, Verra F, et al. (2004) Genetic complexity of Plasmodium falciparum in two ethnic groups of Burkina Faso with marked differences in susceptibility to malaria. Am J Trop Med Hyg 71: 173–178.
12. Konate L, Zwetyenga J, Rogier C, Bischoff E, Fontenille D, et al. (1999) Variation of Plasmodium falciparum msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. Trans R Soc Trop Med Hyg 93 Suppl 1: 21–28.
13. Babiker HA, Lines J, Hill WG, Walliker D (1997) Population structure of Plasmodium falciparum in villages with different malaria endemicity in east Africa. Am J Trop Med Hyg 56: 141–147.
14. Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, et al. (2009) Comparison of Plasmodium falciparum allelic frequency distribution in different endemic settings by high-resolution genotyping. Malar J 8: 250.
15. Bendixen M, Msangeni HA, Pedersen BV, Shayo D, Bodker R (2001) Diversity of Plasmodium falciparum populations and complexity of infections in relation to transmission intensity and host age: a study from the Usambara Mountains, Tanzania. Trans R Soc Trop Med Hyg 95: 143–148.
16. Peyerl-Hoffmann G, Jelinek T, Kilian A, Kabagambe G, Metzger WG, et al. (2001) Genetic diversity of Plasmodium falciparum and its relationship to parasite density in an area with different malaria endemicities in West Uganda. Trop Med Int Health 6: 607–613.
17. Soulama I, Nebie I, Ouedraogo A, Gansane A, Diarra A, et al. (2009) Plasmodium falciparum genotypes diversity in symptomatic malaria of children living in an urban and a rural setting in Burkina Faso. Malar J 8: 135.
18. Schultz L, Wapling J, Mueller I, Ntsuke PO, Senn N, et al. (2010) Multilocus haplotypes reveal variable levels of diversity and population structure of Plasmodium falciparum in Papua New Guinea, a region of intense perennial transmission. Malar J 9: 336.
19. Mackinnon MJ, Hastings IM (1998) The evolution of multiple drug resistance in malaria parasites. Trans R Soc Trop Med Hyg 92: 188–195.
20. Klein EY, Smith DL, Boni MF, Laxminarayan R (2008) Clinically immune hosts as a refuge for drug-sensitive malaria parasites. Malar J 7: 67.
21. White N (1999) Antimalarial drug resistance and combination chemotherapy. Philos Trans R Soc Lond B Biol Sci 354: 739–749.
22. White NJ (1999) Delaying antimalarial drug resistance with combination chemotherapy. Parassitologia 41: 301–308.
23. Artzy-Randrup Y, Alonso D, Pascual M (2010) Transmission intensity and drug resistance in malaria population dynamics: implications for climate change. PLoS ONE 5: e13588.
24. Hastings IM (1997) A model for the origins and spread of drug-resistant malaria. Parasitology 115(Pt 2): 133–141.
25. Hastings IM, D'Alessandro U (2000) Modelling a predictable disaster: the rise and spread of drug-resistantmalaria. Parasitol Today 16: 340–347.
26. Plowe CV, Kublin JG, Doumbo OK (1998) P. falciparum dihydrofolate reductase and dihydropteroate synthase mutations: epidemiology and role in clinical resistance to antifolates. Drug Resist Updat 1: 389–396.
27. Cravo P, Culleton R, Hunt P, Walliker D, Mackinnon M (2001) Antimalarial drugs clear resistant parasites from partially immune hosts. Antimicrob Agents Chemother 45: 2897–2901.
28. White NJ (1998) Preventing antimalarial drug resistance through combinations. Drug Resist Updat 1: 3–9.
29. Dye C, Williams BG (1997) Multigenic drug resistance among inbred malaria parasites. Proc Biol Sci 264: 61–67.
30. Paul RE, Packer MJ, Walmsley M, Lagog M, Ranford-Cartwright LC, et al. (1995) Mating patterns in malaria parasite populations of Papua New Guinea. Science 269: 1709–1711.
31. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, et al. (2003) A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol Biol Evol 20: 1526–1536.
32. Wellems TE, Plowe CV (2001) Chloroquine-resistant malaria. J Infect Dis 184: 770–776.
33. Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, et al. (2002) Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature 418: 320–323.
34. Talisuna AO, Okello PE, Erhart A, Coosemans M, D'Alessandro U (2007) Intensity of malaria transmission and the spread of Plasmodium falciparum resistant malaria: a review of epidemiologic field evidence. Am J Trop Med Hyg 77: 170–180.
35. Trape JF (1987) Malaria and urbanization in central Africa: the example of Brazzaville. Part I: Description of the town and review of previous surveys. Trans R Soc Trop Med Hyg 81 Suppl 2: 1–9.
36. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW (2005) Urbanization, malaria transmission and disease burden in Africa. Nat Rev Microbiol 3: 81–90.
37. Trape JF (1987) Malaria and urbanization in central Africa: the example of Brazzaville. Part IV. Parasitological and serological surveys in urban and surrounding rural areas. Trans R Soc Trop Med Hyg 81 Suppl 2: 26–33.
38. Trape JF, Quinet MC, Nzingoula S, Senga P, Tchichelle F, et al. (1987) Malaria and urbanization in central Africa: the example of Brazzaville. Part V: Pernicious attacks and mortality. Trans R Soc Trop Med Hyg 81 Suppl 2: 34–42.
39. Trape JF, Zoulani A (1987) Malaria and urbanization in central Africa: the example of Brazzaville. Part III: Relationships between urbanization and the intensity of malaria transmission. Trans R Soc Trop Med Hyg 81 Suppl 2: 19–25.
40. Trape JF, Zoulani A (1987) Malaria and urbanization in central Africa: the example of Brazzaville. Part II: Results of entomological surveys and epidemiological analysis. Trans R Soc Trop Med Hyg 81 Suppl 2: 10–18.
41. Ndounga M, Casimiro PN, Miakassissa-Mpassi V, Loumouamou D, Ntoumi F, et al. (2008) [Malaria in health centres in the southern districts of Brazzaville, Congo]. Bull Soc Pathol Exot 101: 329–335.
42. Culleton R, Mita T, Ndounga M, Unger H, Cravo P, et al. (2008) Failure to detect Plasmodium vivax in West and Central Africa by PCR species typing. Malar J 7: 174.
43. Sakihama N, Ohmae H, Bakote'e B, Kawabata M, Hirayama K, et al. (2006) Limited allelic diversity of Plasmodium falciparum merozoite surface protein 1 gene from populations in the Solomon Islands. Am J Trop Med Hyg 74: 31–40.
44. Isozumi R, Uemura H, Le DD, Truong VH, Nguyen DG, et al. (2010) Longitudinal survey of Plasmodium falciparum infection in Vietnam: characteristics of antimalarial resistance and their associated factors. J Clin Microbiol 48: 70–77.
45. Hastings IM, Smith TA (2008) MalHaploFreq: a computer programme for estimating malaria haplotype frequencies from blood samples. Malar J 7: 130.
46. Greenhouse B, Dokomajilar C, Hubbard A, Rosenthal PJ, Dorsey G (2007) Impact of transmission intensity on the accuracy of genotyping to distinguish recrudescence from new infection in antimalarial clinical trials. Antimicrob Agents Chemother 51: 3096–3103.
47. Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP (1999) Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology 119(Pt 2): 113–125.
48. Haubold B, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847–848.
49. Hudson RR (1994) Analytical results concerning linkage disequilibrium in models with genetic transformation and conjugation. Journal of evolutionary biology 7: 535–548.
50. Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. Journal of Heredity 86: 485–486.
51. Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.
52. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, et al. (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol Biol Evol 17: 1467–1482.
53. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, et al. (1999) A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. Science 286: 1351–1353.
54. Cohen j (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20: 37–46.
55. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, et al. (2007) Population genomics of the immune evasion (var) genes of Plasmodium falciparum. PLoS Pathog 3: e34.
56. Talisuna AO, Langi P, Bakyaita N, Egwang T, Mutabingwa TK, et al. (2002) Intensity of malaria transmission, antimalarial-drug use and resistance in Uganda: what is the relationship between these three factors? Trans R Soc Trop Med Hyg 96: 310–317.

Linkage maps from multiple genetic crosses and loci linked to growth-related virulent phenotype in *Plasmodium yoelii*

Jian Li[a,b,1], Sittiporn Pattaradilokrat[b,1], Feng Zhu[a,1], Hongying Jiang[b], Shengfa Liu[a], Lingxian Hong[a], Yong Fu[c], Lily Koo[d], Wenyue Xu[c], Weiqing Pan[e], Jane M. Carlton[f], Osamu Kaneko[g], Richard Carter[h], John C. Wootton[i], and Xin-zhuan Su[b,2]

[a]State Key Laboratory of Stress Cell Biology, School of Life Sciences, Xiamen University, Xiamen, Fujian 361005, People's Republic of China; [b]Laboratory of Malaria and Vector Research, and [d]Research Technologies Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; [c]Department of Pathogenic Biology, Third Military Medical University, Chongqing 400038, People's Republic of China; [e]Department of Pathogen Biology, Second Military Medical University, Shanghai 200433, People's Republic of China; [f]Department of Medical Parasitology, Langone Medical Center, New York University, New York, NY 10010; [g]Department of Protozoology, Institute of Tropical Medicine and the Global Center of Excellence Program, Nagasaki University, Nagasaki 852-8523, Japan; [h]Division of Biological Sciences, Institute of Cell, Animal and Population Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom; and [i]Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Edited by Thomas E. Wellems, National Institutes of Health, Bethesda, MD, and approved May 27, 2011 (received for review February 9, 2011)

***Plasmodium yoelii* is an excellent model for studying malaria pathogenesis that is often intractable to investigate using human parasites; however, genetic studies of the parasite have been hindered by lack of genome-wide linkage resources. Here, we performed 14 genetic crosses between three pairs of *P. yoelii* clones/subspecies, isolated 75 independent recombinant progeny from the crosses, and constructed a high-resolution linkage map for this parasite. Microsatellite genotypes from the progeny formed 14 linkage groups belonging to the 14 parasite chromosomes, allowing assignment of sequence contigs to chromosomes. Growth-related virulent phenotypes from 25 progeny of one of the crosses were significantly associated with a major locus on chromosome 13 and with two secondary loci on chromosomes 7 and 10. The chromosome 10 and 13 loci are both linked to day 5 parasitemia, and their effects on parasite growth rate are independent but additive. The locus on chromosome 7 is associated with day 10 parasitemia. The chromosome 13 locus spans ~220 kb of DNA containing 51 predicted genes, including the *P. yoelii* erythrocyte binding ligand, in which a C741Y substitution in the R6 domain is implicated in the change of growth rate. Similarly, the chromosome 10 locus spans ~234 kb with 71 candidate genes, containing a member of the 235-kDa rhoptry proteins (Py235) that can bind to the erythrocyte surface membrane. Atypical virulent phenotypes among the progeny were also observed. This study provides critical tools and information for genetic investigations of virulence and biology of *P. yoelii*.**

genetic mapping | inheritance | crossover | rodent

The rodent malaria parasite *Plasmodium yoelii* is an important model for studying malaria biology and pathogenesis. Because a malaria disease phenotype represents the outcome of the host-parasite interaction, the use of inbred mice to control host genetic background variation is critical for studying the influence of parasite virulent factors on a disease phenotype. Many genetically distinct (or similar) strains of *P. yoelii* and subspecies exhibiting a wide range of variations in growth rate and pathogenicity in their rodent hosts are available, which can be explored for studying disease and/or growth phenotypes. Compared with *Plasmodium falciparum* and *Plasmodium chabaudi chabaudi* (1–3), however, genetic studies in *P. yoelii* have been limited (2, 4, 5), partly because of the lack of genetic markers and well-characterized phenotypes. Recently, hundreds of polymorphic microsatellite (MS) markers have been developed from the *P. yoelii* genome (6), setting the stage for development of genome-wide genetic maps for this parasite. Additionally, a strategy called linkage group selection (LGS) was developed to map the determinants affecting selectable rodent malaria traits (2, 7). Indeed, a C713R substitution in the gene encoding the *P. y. yoelii* erythrocyte binding ligand (PyEBL) was recently linked to parasite growth rate and virulence using the LGS technique, although other determinants are likely to play a role (8, 9). [Note: There are three subspecies of *P. yoelii* (*P. yoelii yoelii, P. yoelii nigeriensis*, and *P. yoelii killicki*) and two subspecies of *P. chabaudi* (*P. chabaudi chabaudi* and *P. chabaudi adami*). *P. yoelii* is used here to refer generally to *P. yoelii* lines and subspecies; subspecies and lines will be specified in the text when necessary.] For mapping genes affecting complex traits or phenotypes that cannot be selected, however, evaluation of phenotypes from individual progeny of genetic crosses is necessary. Development of a genetic map and collection of genetic cross progeny with differences in disease phenotypes will provide important tools for studying such malaria disease phenotypes in detail.

Although the *P. y. yoelii* genome was the first rodent malaria parasite genome sequenced, the assembly is still fragmented because of the currently low coverage and lack of a genetic map to guide the assembly (10). Recently, a rodent malaria syntenic map was constructed based on genomic sequences from three rodent malaria parasites (*P. y. yoelii*, *P. c. chabaudi*, and *Plasmodium berghei*) and sequence synteny to the genomic sequence of *P. falciparum* (11); however, thousands of sequence gaps still exist, and many contigs are yet to be assigned to their proper chromosomes. Increasing sequence coverage may close additional gaps, but development of physical and genetic maps will be necessary for assigning all the contigs to chromosomal positions and for assembling the chromosomes completely.

Here, we have performed 14 individual genetic crosses using six parasite lines/subspecies, cloned 75 independent recombinant progeny from the crosses, genotyped 82 recombinant progeny from genetic crosses of four parental pairs (including 7 progeny from a previous *P. y. yoelii* YM × *P. y. yoelii* A/C cross) (12) with hundreds of MS markers, and developed a high-resolution linkage map. We also identified three genetic loci, including the gene encoding PyEBL, linked to quantitative growth-related virulent pheno-

Author contributions: X.-z.S. designed research; J.L., S.P., F.Z., Y.F., and L.K. performed research; S.L., L.H., W.X., W.P., and O.K. contributed new reagents/analytic tools; J.L., S.P., H.J., L.K., J.M.C., R.C., J.C.W., and X.-z.S. analyzed data; and J.L., S.P., J.M.C., O.K., R.C., J.C.W., and X.-z.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]J.L., S.P., and F.Z. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: xsu@niaid.nih.gov.

See Author Summary on page 12575.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102261108/-/DCSupplemental.

types (GRVPs) using trait data from individual nonselected progeny clones.

Results

Frequencies of Clonal Infection and Independent Recombinant Progeny. We performed 14 independent genetic crosses using the six *P. yoelii* lines or subspecies (*P. y. yoelii* 17XNL × *P. y. nigeriensis* N67, *P. y. yoelii* BY265 × *P. y. nigeriensis* NSM, and *P. y. yoelii* YM × *P. y. yoelii* 33X). For simplicity, these parasite strains are referred to as 17XNL, N67, BY265, NSM, YM, 33X, and A/C (for *P. y. yoelii* A/C), respectively. Mice (n = 2,326) were injected with 0.6–1.5 infected red blood cells (iRBCs) obtained from mice that were infected with sporozoites derived from mosquitoes fed on blood samples containing both parental parasites. Of the 2,326 mice, 889 (38.2%) had parasites in their blood 7–9 d postinjection (Table 1). DNA samples from the infected mice were genotyped with 45 MS markers, 18 for the BY265 × NSM cross, 20 for the YM × 33X cross, and 28 for the 17XNL × N67 cross, with some markers typed in more than one cross. For example, Py2699 was used to type progeny from the 17XNL × N67, YM × 33X, and BY265 × NSM crosses to verify clonal infections and to identify recombinant progeny (Datasets S1 and S2). Because the parasite has a haploid genome in the mouse host, the progeny that carry only one of the parental alleles at all MS loci typed for each cross are considered clonal. Among the 889 infected mice, 488 (54.9%) were found to have clonal infections, including 75 (8.4%) independent recombinant progeny with unique genotypes and 248 (27.9%) having parasites with parental genotypes (Table 1). We also genotyped 7 progeny from the YM × A/C cross performed at the University of Edinburgh, bringing the total number of progeny typed with MSs to 82.

MS Polymorphism and Genetic Variations Between Parasite Strains. The 82 independent recombinant progeny identified from the initial MS typing were further analyzed with additional genome-wide MS markers. To determine which MSs are polymorphic between a particular parental pair of the crosses, we first genotyped DNA samples from the parental parasites [N67, BY265, and 17XNL were genotyped previously (6)] with 591 MS markers (Dataset S1). A panel of polymorphic MS markers for each cross was selected for typing progeny from the crosses after comparing PCR product sizes between the parents. Eventually, 485, 499, 339, and 182 MS markers were shown to be polymorphic between the parental pairs of BY265 × NSM, 17XNL × N67, YM × 33X, and YM × A/C crosses, respectively, and were used to type the progeny of the crosses. Four hundred seventy-seven MS markers produced genotypes from the 32 progeny of the BY265 × NSM cross, 486 MS markers produced genotypes from the 25 progeny of the 17XNL × N67 cross, 330 MS markers had genotypes from the 18 progeny of the YM × 33X cross, and 178 MS markers had genotypes from the 7 progeny of the YM × A/C cross, generating 33,939 MS genotypes with a genotype-calling rate of ~96.0% (Dataset S2 and Table S1).

More than 82% of the 591 MSs were polymorphic between the parental pairs of BY265 × NSM and 17XNL × N67 (Table S1), indicating highly diverse genomes of the *P. yoelii* strains or subspecies from different geographic origins. In contrast, ~70% and 43% of the MS markers that were shown to be polymorphic among seven isolates previously (6) were monomorphic between the parents of the YM × A/C and YM × 33X crosses, respectively. Parasites YM, 17XL, and 17XNL emerged as very similar, with fewer than 25 polymorphic MS markers, which reflects their common origin from 17X and is consistent with a previous study based on amplified fragment length polymorphism (AFLP) (13) (Figs. S1 and S2). These results showed close relationships and potentially shared chromosomes or chromosomal segments between the 17X, YM, A/C, and 33X. Because A/C is a progeny clone from a genetic cross between *P. y. yoelii* 17XA (17XA) that was derived from the isolates 17X and 33X, which originated from the same locality in the Central African Republic as parasite 17X (Fig. S2), it is not surprising to see a close genetic relationship or shared chromosome segments between17X/YM and 33X.

Comparative Genetic Linkage Maps and Estimates of Genetic Distances. We first constructed three individual genetic maps from the three crosses (excluding YM × A/C, which has only 7 progeny) and estimated the genetic distances from the crosses using 37 markers that were physically mapped to specific chromosomes previously (11, 14–16) (Fig. S3 and Table S2). These markers therefore anchored specific linkage groups to their respective chromosomes and were useful in resolving some linkage groups into independent chromosomes. Although the genome-wide distances from the 17XNL × N67 and BY265 × NSM crosses were similar [431.2 and 473.7 centimorgan (cM), respectively], the total genetic distance from the YM × 33X cross was approximately double those from the other two crosses (807.0 cM) (Table 2). Despite these differences in estimates of genetic distance, all three crosses gave relatively even marker intervals (in cM) on all linkage groups, and the marker orders on each linkage group of the three crosses were essentially the same. We therefore combined the genotypes from the 82 progeny of the four crosses to construct a composite linkage map (Fig. 1 and Fig. S3). The resulting map totaled 579.2 cM, with markers quite evenly distributed across each of 14 linkage groups (no interval was >10 cM) (Fig. 1 and Table 2). Using the estimated genome size of 23 Mb (10), we obtained an average genome-wide unit recombination rate of 39.7 kb/cM (25.2 cM/Mb) for *P. yoelii*.

Shared Chromosomal Segments and Lack of Significant Segregation Bias. Analysis of the marker inheritance patterns also revealed that some entire chromosomes and long chromosomal segments were shared among the YM, A/C, and 33X parasites, explaining the smaller numbers of polymorphic MS markers in the crosses of these parasites. Chromosomes 2, 3, 5, 7, and 10 [our linkage groups are numbered to match the chromosome numbers in the syntenic map of Kooij et al. (11)] and parts of many other chromosomes were monomorphic between YM and A/C (Table 2 and

Table 1. Genetic crosses of *P. yoelii* performed and numbers of recombinant progeny obtained during this study

Crosses	Crosses performed	Mice injected	Mice infected	% mice infected	Clonal progeny	Parental clones	Recom progeny	% recom progeny	No. IRP	% IRP
17XNL × N67	4	501	145	28.9	122	84	38	26.2	25	17.2
YM × 33X	3	209	77	36.8	58	10	48	62.3	18	23.4
BY265 × NSM	7	1,616	667	41.3	308	154	154	23.1	32	4.8
Total	14	2,326	889	38.6	488	248	240	27	75	8.4

Clonal progeny, numbers of progeny that are clonal after being typed with MS markers; Mice infected, numbers of mice infected with parasites; % Mice infected, percentage of mice infected with parasites; Mice injected, numbers of mice injected with diluted blood/parasites; No. IRP, numbers of independent recombinant progeny; % IRP, percentage of independent recombinant progeny from total infected mice; Recom progeny, numbers of recombinant progeny; % Recom progeny, percentage of recombinant progeny from total infected mice.

Table 2. Informative MS markers, crossover counts, and genetic distances from crosses of different parental combinations

	YM × A/C			BY265 × NSM			17XNL × N67			YM × 33X			All crosses		
	(7 progeny)			(32 progeny)			(25 progeny)			(18 progeny)			(82 progeny)		
Chr	MS	CO	G. Dis, cM	MS	CO	G. Dis, cM	MS	CO	G. Dis, cM	MS	CO	G. Dis, cM	MS	CO	G. Dis, cM
1	6	1	—	31	6	20.9	31	5	21.8	27	3	18.0	34	15	21.1
2	0	0	—	11	7	23.0	12	1	4.2	1	0	—	12	8	14.5
3	0	0	—	15	7	23.2	17	3	13.1	4	1	5.9	17	11	19.2
4	12	6	—	24	8	27.0	26	9	38.7	20	9	58.8	27	32	42.6
5	1	0	—	28	5	17.4	25	1	5.0	10	2	12.0	30	8	15.7
6	11	4	—	20	7	26.8	23	3	13.2	16	13	88.0	28	27	38.3
7	1	0	—	24	3	10.4	24	9	38.0	16	11	70.1	25	23	33.4
8	30	4	—	38	12	40.2	36	13	56.6	35	20	123.6	41	49	64.9
9	23	2	—	39	15	50.5	39	11	47.5	38	9	55.1	47	37	54.8
10	1	0	—	31	10	33.4	30	7	30.6	1	0	—	33	17	31.7
11	8	4	—	48	17	61.1	49	12	60.1	41	17	106.0	52	50	70.0
12	12	0	—	39	12	43.8	40	5	22.3	11	0	—	44	17	40.7
13	50	4	—	56	16	54.9	59	12	52.7	56	27	180.7	67	59	77.3
14	12	6	—	49	12	41.1	51	6	27.4	37	13	88.8	53	37	55.0
UA	11			24			24			17			29		
Total	178	31	—	477	137	473.7	486	97	431.2	330	125	807.0	539	390	579.2

Chr, chromosome; CO, crossover counts; G. Dis, genetic distances in cM; MS, numbers of polymorphic microsatellites; UA, unassigned.

Dataset S2); similarly, markers from chromosomes 2 and 10 and parts of chromosomes 3, 5, 6, 7, and 12 were monomorphic in the YM × 33X cross. Because YM, which derives from 17X, and 33X were from two independent wild isolates of the same location (Fig. S2*B*), the results suggested that either 17X or 33X was a progeny of a cross involving an ancestor of the other parent that occurred in the wild before they were isolated.

We also examined the possibility of segregation bias by plotting the ratios of parental genotypes along each of the 14 chromosomes (Fig. 2*A*). No statistically significant distortion was

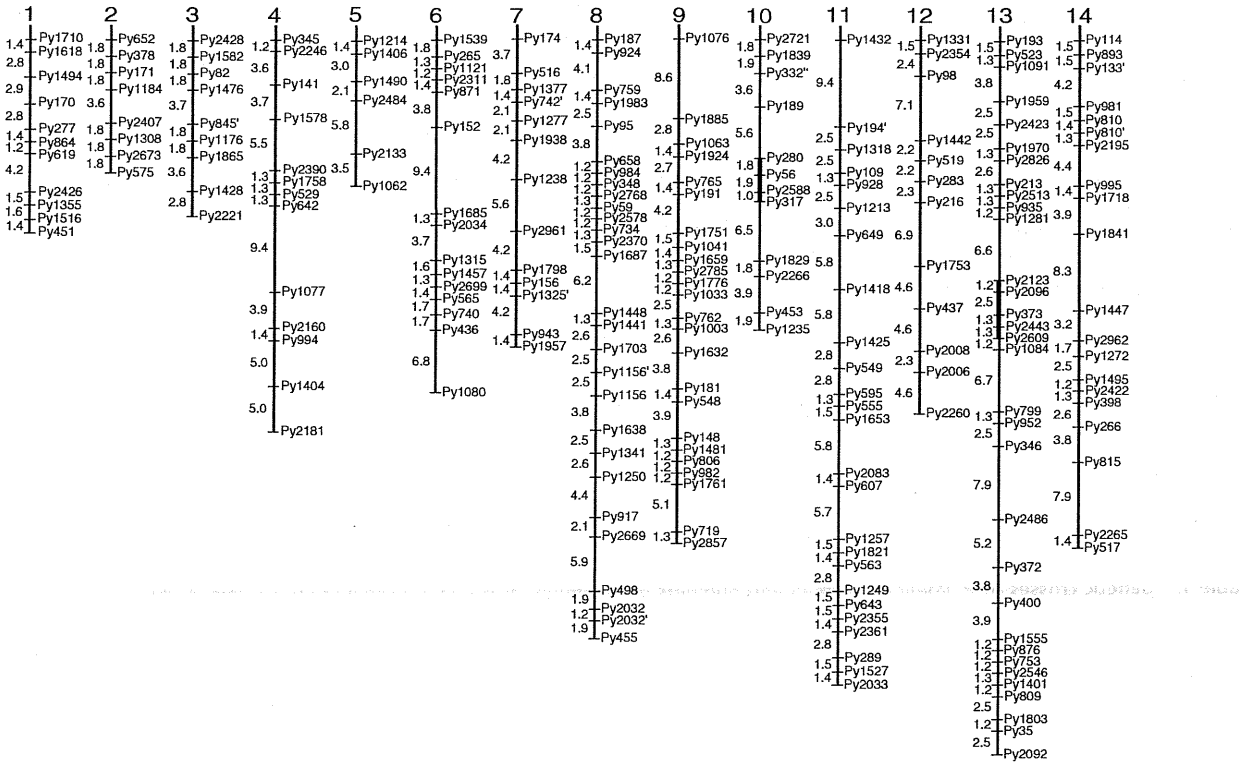

Fig. 1. A composite *P. yoelii* genetic map constructed from genetic crosses of four parental combinations. Genetic distances between MS markers on the 14 chromosomes of 82 progeny were calculated using Mapmaker/Exp3.0. The linkage groups corresponding to the 14 chromosomes in the syntenic map (11) are marked from 1 to 14. The numbers between two ticks on the left side of the vertical lines are genetic distances in centimorgans, and the names of the MS markers are on the right side of vertical lines. Additional names and clusters of the MS markers can be found in Fig. S3. Note that chromosome numbering in *P. yoelii* and other rodent *Plasmodium* sp. is different from that of *P. falciparum*. The thick bars on chromosomes 10 and 13 mark the two loci linked to a GRVP.

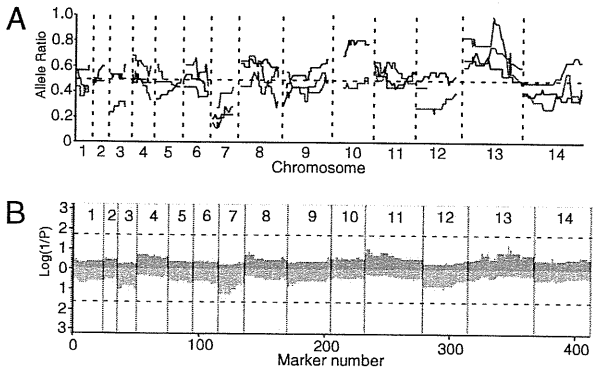

Fig. 2. Plots of parental genotype inheritance among the progeny of the three genetic crosses. (*A*) Ratios of alleles from one parent of each of the crosses were calculated and plotted. Green, ratios of YM alleles over the total alleles from the progeny of the YM × 33X cross; red, ratios of 17XNL alleles over the total alleles from the progeny of the 17XNL × N67 cross; and black, ratios of BY265 alleles over the total alleles from the progeny of the BY265 × NSM cross. (*B*) Log(1/*P*) values are plotted upward (gray) for distortion toward 17XNL and downward (blue) for distortion toward N67. Dotted lines represent the $P < 0.01$ genome-wide significance threshold estimated using a Bonferroni correction based on the number of genetic intervals in the genome map. The largest (nonsignificant) deviation from the expected 1:1 ratio occurred at regions of chromosomes 7 and 13 close to the loci associated with the GRVP (Fig. 3).

present (Fig. 2*B*); however, the largest (nonsignificant) deviation from the expected 1:1 ratio occurred at regions of chromosomes 7 and 13 close to loci associated with growth phenotypes (see below), which may reflect contributions of these loci to parasite fitness. The relative lack of segregation distortion is an advantage for genetic mapping in *P. yoelii* and contrasts with the significantly skewed inheritance reported for a few chromosomal regions in genetic crosses of *P. falciparum* (1, 3) and *Toxoplasma gondii* (17).

Assignment of Orphan Contigs to Chromosomes. The linkage maps were compared with the composite synteny chromosome maps of three rodent malaria parasites (11). Except for four contigs, the linkage maps placed the contigs with MS markers in the same order of those in the synteny map (Dataset S2). A contig with MS Py2653 was assigned to chromosome 7 in the synteny map, but the inheritance pattern of the MS in the progeny of the BY265 × NSM and 17XNL × N67 crosses matched those of Py1080 on chromosome 6. Similarly, the contig with Py1484 initially placed on chromosome 7 of the synteny map was assigned to chromosome 12, and the contig with Py517′ initially assigned to chromosome 13 matched MS Py353′ on chromosome 14 (Dataset S2). Finally, the positions of the contigs containing Py1803 and Py35 on chromosome 13 were reversed in the synteny map.

Our linkage maps also assigned 28 orphan contigs (~159 kb) that were not assigned to the synteny maps previously to 13 of the 14 linkage groups (those in gray background in Dataset S2). These genetic maps and the placement of the contigs in linkage groups will greatly facilitate the assembly and completion of the genome sequence.

Mapping Genes Contributing to GRVPs. We next applied the linkage map to analyze the GRVP in the 17XNL × N67 cross. Between the two parasites used as parents in the genetic crosses, N67 grows faster than 17XNL, showing a strong early burst of rapid growth associated with greater virulence. The difference in parasitemia between N67 and 17XNL was the largest at day 5 postinjection of 1×10^5 iRBCs, when N67 parasitemia reached 20–60% but the 17XNL parasitemia was 1–10% (Fig. 3 *A–C*). Accordingly, we counted day 5 parasitemia in replicate mice injected with the 25 progeny from the 17XNL × N67 cross (Fig. 3*D* and Dataset S3). The 25 progeny from the 17XNL × N67 cross produced relatively consistent growth phenotypes in replicate mice and could be largely classified into either a fast- or slow-growth phenotype, representing the phenotypes of N67 or 17XNL, respectively; however, some progeny clones had unstable phenotypes (e.g., inconsistently producing either slow or fast growth in replicate infections of isogenic mice) (Dataset S3). For example, three of the eight mice infected with progeny G007#3 had a fast-growth phenotype at day 5, whereas the remaining five had a slow-growth phenotype (a third mouse reached 32% parasitemia at day 7).

Quantitative trait loci (QTL) linkage analysis was performed using the parasitemia data from the 25 progeny of the 17XNL × N67 cross. We first used two distinct conservative statistical phenotype-genotype association strategies (nonparametric Wilcoxon rank statistics and mutual information) that make no assumptions about pedigree, given that we had high marker density but relatively few progeny and that the parents can also be included to

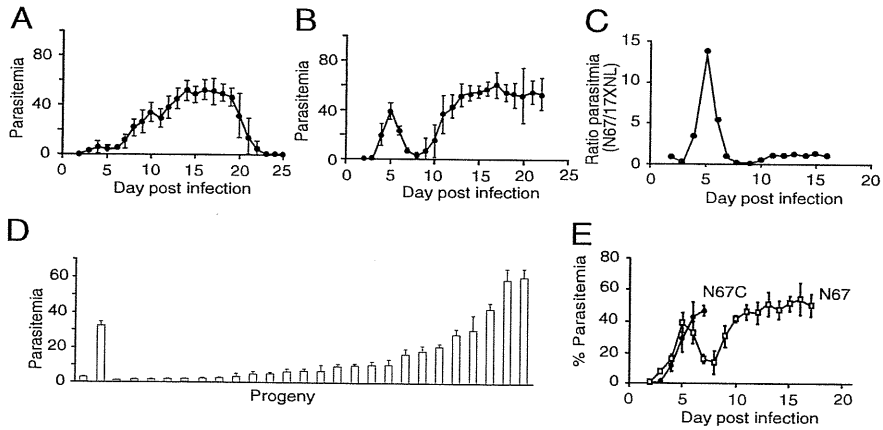

Fig. 3. Measurements of growth rate (parasitemia) of the parents of the 17XNL × N67 cross. Mean parasitemia and SEs of the 17XNL parasite (*A*), mean parasitemia and SEs of the N67 parasite (*B*), ratios of parasitemia N67/17XNL at days 2–16 postinfection (*C*), and day 5 parasitemia from the parents and progeny of the 17XNL × N67 cross (*D*) are shown. In *D*, the first two bars from the left are the parents 17XNL and N67, respectively. SEs were from at least four mice. (*E*) Mean parasitemia with SEs from mice infected with N67 and N67C parasites. A total of 15 and 5 C57BL/6 mice were injected with 1×10^5 N67C and N67 parasites, respectively. All the mice infected with N67C died at day 7, whereas the majority of the mice infected with N67 were still alive at day 17 when experiments were stopped.

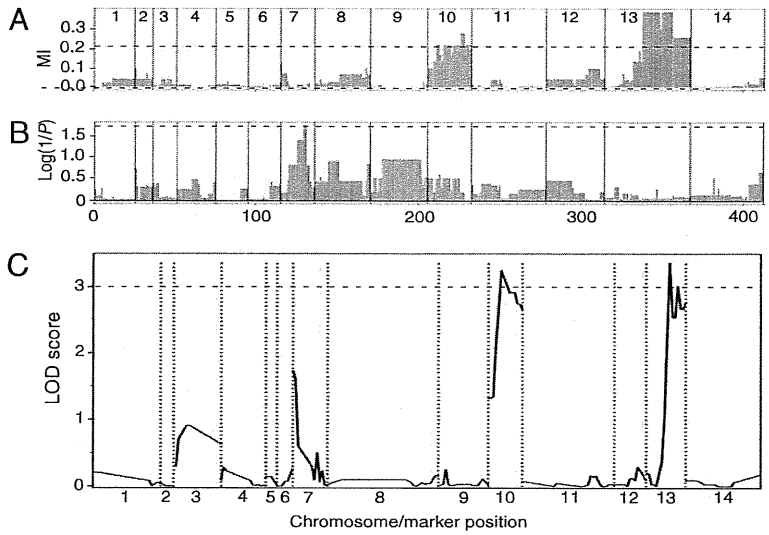

Fig. 4. Genetic loci linked to the GRVP. (*A*) Plots of mutual information scores and MS markers across the 14 *P. yoelii* chromosomes using day 5 parasitemia from the 25 progeny and the parents of the 17XNL × N67 cross. MI, mutual information. (*B*) Plots of log(1/*P*) and MS markers using day 10 parasitemia. (*C*) Plot of logarithm of odds scores using R/qtl (*Materials and Methods*) after natural logarithm transformation of the day 5 parasitemia data so that the dataset is normally distributed. LOD, logarithm of odds. Multiple calculation methods, including maximum likelihood and extended Haley–Knott regression, were evaluated; all produced the essentially the same results. The horizontal dashed lines indicate significant levels at $P = 0.001$ (*A*), $P = 0.05$ (*B*), and $P = 0.05$ (*C*).

increase the sample size using these methods (*Materials and Methods*). The result identified a major peak on chromosome 13 containing the PyEBL gene (genome-wide $P < 1 \times 10^{-5}$) and a minor peak on chromosome 10 ($P < 1 \times 10^{-2}$) (Fig. 4*A*). As a distinct phenotype, day 10 parasitemia was analyzed and found to be weakly associated with a locus on chromosome 7 ($P = 0.05$ after 1,000 permutations) (Fig. 4*B*). We also used R/qtl (18) to identify loci linked to the GRVP; again, the chromosome 13 and chromosome 10 loci were the two major peaks with logarithm of odds scores of 3.4 and 3.2, respectively (Fig. 4*C*). There was also a peak on chromosome 7, but it was not statistically significant. The chromosome 13 locus spans a DNA segment of ~220 kb (between MS markers Py2123 and Py2609) and contains 28 *P. yoelii* contigs and 51 predicted genes, including the gene encoding PyEBL (Table S3), and the chromosome 10 locus covers ~234 kb (between MS markers Py280 and Py1597) containing 28 *P. yoelii* contigs and 71 predicted genes (Table S4), including a gene encoding a member of the 235-kDa rhoptry proteins that can bind to the erythrocyte surface membrane (19).

We also investigated the possibility of interactions among genome-wide markers, particularly between the two loci linked to day 5 parasitemia in the 17XNL × N67 cross. Using standard interval mapping (the expectation and maximization method) in R/qtl, we found no statistically significant epistasis among genome-wide markers (Fig. 5*A*), including no interaction between the loci on chromosome 10 and chromosome 13 ($P = 0.31$); instead, the three notable two-locus effects were all additive, namely, chromosome 10/chromosome 13 ($P = 0.005$), chromosome 7/chromosome 10 ($P = 0.012$), and chromosome 13/chromosome 13 ($P = 0.011$) (Fig. 5*A*). The two interacting loci on chromosome 13 are 18 cM apart and are inversely additive, with one containing *pyebl* and one at the beginning of the chromosome that has a minimum QTL peak (Fig. 4*C*). The chromosome 10 and chromosome 13 loci are estimated to explain ~23% and ~25% of the phenotype variance (totaling ~70% if combined; $P < 0.001$), respectively. From the mutual information method, the percentages explained by the chromosome 10 and chromosome 13 loci to the association with the day 5 parasitemia are 23% and 37%, respectively. Given the relatively broad bands of uncertainty on estimates of relative contributions and the small sample size of 25 progeny, these values from R/qtl and mutual information can be considered to be in good agreement. Indeed, the progeny carrying N67 alleles at both chromosome 13 and chromosome 10 loci had the highest level of parasitemia at day 5 compared with those carrying one or both of the 17XNL alleles (Fig. 5*B*).

Unique C741Y Substitution in PyEBL Is Likely Associated with the GRVP in the N67 Background. The loci on chromosomes 7 and 10 have not been previously reported and require further investigation to identify the gene(s) playing a role in parasite growth; however, the locus on chromosome 13 contains the *pyebl* gene, which was previously linked to a GRVP both by LGS in the YM × 33X cross and by genetic manipulation implicating a single C713R substitution (position 713 in the YM sequence) in the PyEBL R6 domain as the crucial determinant (8, 9). Our mapping of the primary locus to chromosome 13 strongly suggests that PyEBL again plays a role in the GRVP in N67. We therefore sequenced the N67 PyEBL gene to determine whether this C713R substitution is also present in the N67 parasite. We found 39 amino acid substitutions and two indels between 17XNL and N67 (Fig. S4); however, the C713R substitution seen in YM does not exist in N67. Instead, a C741Y substitution (with the other changes) was present in the R6 domain (Fig. S4). Interestingly, a parasite submitted to MR4 (http://www.mr4.org/) under the name of *P. y. yoelii* 33X(Pr3) [for simplicity, 33X(Pr3) will be used] has an identical PyEBL sequence and a very similar genomic background to that of N67, except for the C741Y substitution in N67 (Fig. S4 and Table S5). Among 21 MS markers typed, only a single MS had different alleles between N67 and 33X(Pr3), whereas all the 21 MSs had different alleles among 33X and 33X (Pr3)/N67 (Table S5). The results suggest that N67 and 33X(Pr3) are closely related or isogenic and that the 33X(Pr3) in our hands was not the original parasite derived from 33X. We therefore designated this parasite N67C for having a 741C in PyEBL. Although N67C grows slightly slower than N67 in the early infection, it is more virulent than N67 because all the N67C-infected mice died at day 7, whereas mice infected with N67 died at approximately day 15 after a decline in parasitemia on day 7 (Fig. 3 *A* and *E*). The difference in the GRVP between N67 and N67C is

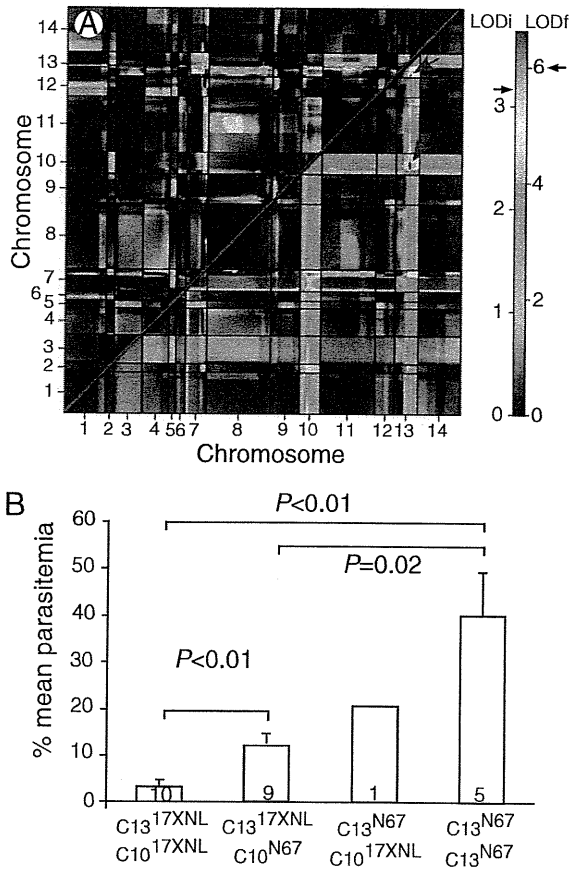

Fig. 5. Interaction and additive effect between the loci on the parasite chromosomes. (*A*) Loci with additive effects but no significant interactions genome-wide were detected. The upper left triangle displays interactions among pairwise loci. The lower right triangle displays additive effects. LODf, logarithm of odds additive effect; LODi, logarithm of odds interaction. Signals of additive effect between markers on chromosomes 10–13, chromosomes 7–10, and two loci on chromosome 13 (red arrows) were detected. The color scale bar shows LOD scores, with those on the left and right corresponding to LODi and LODf values, respectively. The two arrows indicate cutoff LOD scores from 1,000 permutations (LODi = 4.5; LODf = 5.9; $P < 0.05$). (*B*) Relationship of parasitemia and different combinations of PyEBL alleles. Progeny carrying N67 PyEBL alleles at the mapped loci also have significantly higher parasitemia than those carrying 17XNL alleles (unpaired *t* test). The numbers within each bar are the numbers of progeny in each group.

therefore likely caused by the C741Y substitution. Thus, two different single mutational substitutions associated with high day 5 parasitemia and with destabilizing effects focused on the R6 domain appear to have originated independently in the distinct PyEBL sequence backgrounds of 17XNL/YM and N67/N67C.

Discussion

Genetic crosses have been performed with several malaria parasite species, including *P. falciparum*, *P. c. chabaudi*, and *P. y. yoelii* (3, 8, 20–23). The process of cloning, genotyping, and identifying recombinant progeny has been the most time-consuming procedure in developing a linkage map and mapping malaria traits. The numbers of progeny isolated from *Plasmodium* crosses have been relatively small; for example, the *P. falciparum* Dd2 × HB3 and GB4 × 7G8 crosses have 35 and 33 progeny, respectively (3, 22, 24). Because cloning rodent malaria parasites requires injecting a single parasite into a large number of mice, only 7 progeny from the YM × A/C cross were previously obtained (8). In this study, we obtained 488 cloned parasites from 2,326 mice and identified 75 independent recombinant progeny from multiple genetic crosses after genotyping the parasites with hundreds of MS markers. This study describes the first *P. yoelii* linkage map, which represents a significant advance in malaria parasite genetics.

Performing multiple genetic crosses using different parents allowed comparison of inheritance bias and recombination frequency among crosses. Our study revealed a somewhat greater genome-wide genetic distance (centimorgans) in the YM × 33X cross than in the other crosses, possibly reflecting fewer progeny from this cross or some unknown genetic factors among *P. yoelii* strains in control of meiotic recombination frequencies. Nevertheless, in all these crosses, the recombination rates were substantially less than those previously reported for *P. falciparum* and *P. c. chabaudi*. We found relatively small inheritance bias, in contrast to previous findings with *P. falciparum* crosses (1, 3). Inheritance patterns of chromosomes 7 and 13 were marginally skewed, possibly reflecting alleles that may promote parasite survival (e.g., a chromosome 13 locus strongly linked to parasitemia at day 5, a chromosome 7 locus weakly linked to parasitemia at day 10).

The linkage map provides a solid framework for future improvement of the *P. yoelii* genome assembly. The *P. yoelii* genome is currently estimated at 23 Mb but is highly fragmented in 5,687 contigs (10). Although the syntenic map based on sequences from three rodent parasites and the genome sequence of *P. falciparum* had assigned ~2,400 *P. y. yoelii* contigs (total length of 15.2 Mb) to putative chromosomes, there are still many contigs that cannot be assigned (10, 11). We were able to assign 28 orphan contigs (~159 kb) to 13 chromosomes. Additional orphan contigs can now be assigned to chromosomes or linkage groups if polymorphic markers from the contigs can be typed on the progeny of the genetic crosses. Our genetic map has 14 linkage groups, in accordance with the 14 nuclear chromosomes of all *Plasmodium* species investigated, and the marker order in these linkage groups matched well with those in the chromosomal synteny maps. The linkage maps developed in this study provide essential tools for complete genome assembly of *P. yoelii*.

The same chromosome 13 locus containing the gene encoding PyEBL was identified by both the LGS (8) and our classic QTL mapping based on the phenotypes of individual nonselected progeny clones. Our study also identified a unique mutation and a potential different mechanism affecting parasite GRVPs. These mutations appear to have originated independently in the diverged PyEBL sequence backgrounds of *P. y. yoelii* 17X and *P. y. nigeriensis* N67, which differ by 39 amino acid substitutions and two indels. In theory, other genes in the chromosome 13 locus cannot be excluded as candidates contributing to the GRVP; however, the differences in the PyEBL are likely to be the primary determinant for the differences in the GRVP between N67 and 17XNL based on the two previous studies (8, 9). Furthermore, N67 and N67C have an almost identical genomic background and the same PyEBL sequence, except for the substitution at C741Y. The two different single mutational substitutions associated with higher early parasitemia, C713R in YM (8, 9) and C741Y in N67 identified here, are predicted to have similar disruptive effects on the R6 domain. Inspection of the homologous EBA-175 R6/KIX domain crystal structure (25) suggests that both the C713R and the C741Y substitutions break two different strongly conserved disulfide bonds (Fig. S4) and act indirectly, partly through solvation effects on the primary hydrophobic core and a hydrophobic groove in the dimer interface. The disruption of the disulfide bonds appears to correlate with higher early peak parasitemia before day 7. N67 (with a 741Y) grows slightly faster than N67C (741C) at days 2–5, and YM/17XL (with a 713R) also grows faster than 17XNL (713C). The C741Y substitution may partly explain the reduced virulence of N67 compared with N67C, but the drop in parasitemia after day 7 in N67 could be caused by other un-

known factors too, or it could represent a separate phenotype. The mechanism of how the change in amino acid in a merozoite protein affecting parasite early growth remains unknown.

The presence of loci other than the chromosome 13 locus was implicated in GRVPs in the previous studies (8, 9); however, the exact locations of the additional determinants were unknown. Our study now points to two loci on chromosomes 10 and 7 that are linked to day 5 and day 10 parasitemia, respectively. Interestingly, a copy of the Py235 rhoptry protein (PY06636) that has been implicated in RBC invasion is present in the chromosome 10 locus (19, 26). Members of the Py235 rhoptry proteins have been suggested to be potential factors that may be responsible for the difference in virulence (27). Moreover, our results show independent and additive effects for the QTL on chromosomes 13 and 10, with no evidence for significant epistasis genome-wide. Three additive effects were found, although the additive effect from the two loci on chromosome 13 could be artifact because of no obvious QTL peak at the locus on the chromosome end and the small sample size. Further studies are necessary to identify or confirm genes in the other two loci that contribute to the GRVP. Interestingly, genetic loci contributing to subtle but quantifiable strain-specific differences in proliferation rates of *P. falciparum* parasite Dd2 (faster growth) and HB3 during in vitro cultivation, which were attributed to a decreased cycle time and increased merozoite production and invasion rates, have been mapped recently, including a major genetic effect on chromosome 12 containing 165 candidate genes (28).

Recombinant progeny with atypical growth phenotypes were obtained after genetic recombination. Display of either fast or slow growth in a single progeny was observed, which was likely controlled through gene expression regulation or combinational effects of multiple genes, or possibly through subtle uncontrolled environmental effects influencing the mice. Progeny with an intermediate growth rate or growth rate higher than that of the parents were also obtained. These transgressive cases indicate that clones that are more virulent could be generated in the wild through genetic recombination in sexual crosses of parasites as well as by de novo mutations. Indeed, recombinant progeny more virulent than parental clones have been reported in *T. gondii*, suggesting that sexual recombination can be a powerful force driving the natural evolution of virulence (29). Alternatively, deleterious mutations can accumulate in clonally maintained organisms, and recombination may effectively remove these deleterious mutations from some progeny, restoring fitness and resulting in progeny with phenotypes greater than those observed in the parental lines.

Genetic mapping of rodent malaria traits has been largely based on LGS, which does not require cloning and typing individual progeny (7). In this procedure, a phenotype-specific selection pressure is applied to the uncloned progeny of a genetic cross between two parents with different relevant phenotypes. Selected and unselected progeny are analyzed using genome-wide quantitative genetic markers. Under selection, the sensitive allele of the target gene will be removed, leading to a reduced allele frequency or "selection valley" at the locus carrying the "resistant" gene (7). LGS has been successfully used to map loci affecting drug resistance, immunity, and difference in growth rate/virulence (2, 7, 8, 30), which has established LGS as a convenient approach for mapping selectable malaria traits. Cloning and evaluating individual progeny from genetic crosses now enables a powerful genetic mapping approach that can be applied to screen nonselectable parasite phenotypes or complex traits that are not amenable to LGS methods. The progeny obtained in this study not only provide powerful and durable resources for genetic studies but also allow investigation of inheritance bias and recombination parameters, such as hotspots. Moreover, the markers ordered on the 14 chromosomes now provide a chromosomal framework for improved assembly of the parasite genome sequence. As with studies that have been done using *P. falciparum* genetic crosses (22, 31–35), the progeny and genotypes we describe here can be used to map multiple segregating genetic determinants in this mouse model, which, in turn, will provide important information for studying human malaria.

Materials and Methods

Parasites, DNA Sequencing, and MS Typing. The origins of the *P. yoelii* lines and the relationships of the cloned lines are summarized in Fig. S2 and Table S6. Parasites 17XNL, N67, BY265, N67C [under *P. y. yoelii* 33X(Pr3)], and NSM were obtained from MR4 (http://www.mr4.org/) or were described previously (36). YM, 33X, A/C, and *P. y. yoelii* 17X(A) were obtained from frozen stocks of the University of Edinburgh (13). YM is a lethal parasite derived from the nonlethal uncloned isolate 17X following removal of a stabilate from liquid nitrogen storage in the early 1970s (12, 37); it also has the same genome and GRVP as 17XL (13). NSM is a mefloquine-resistant parasite selected from *P. yoelii* NS, a parasite line emerged from an isolate of *P. berghei* from Katanga, Belgian Congo, in the early 1970s; however, our genotyping showed that the N67, NSM, and *P. yoelii* NS in our hands had essentially the same genome (Dataset S1) (36). A/C is a progeny from a cross of 33X and 17XA, a pyrimethamine-resistant line that is genetically distinct from clones 17XNL, 17XL, and YM but originated from the same isolate 17X (12, 13) (Fig. S2). Mosquitoes were from a colony of *Anopheles stephensi* maintained at the Laboratory of Malaria and Vector Research, National Institutes of Health (NIH), and a colony of *A. stephensi* maintained at the Third Military Medical University of China. Female inbred strain C57BL/6 and outbred CD-1 and Kunming mice, aged 6–8 wk, were used in the experiments. All animal procedures were performed in accordance with animal study protocol LMVR 11E approved by the National Institute of Allergy and Infectious Diseases Animal Care and Use Committee (NIH) and with the approved protocols of the Third Military Medical University or Xiamen University in China.

DNA samples were extracted from 20 to 100 μL of heparinized tail blood from *P. yoelii*-infected mice using a High Pure PCR Template preparation kit (Roche). For DNA sequencing of *pyebl*, oligonucleotide primers 5-CCTCC-TGTTGCATAGTAGTATTGAT-3 and 5-TTTGATGAACCAAATGCATAGA-3, corresponding to positions 1,407–1,431 and 4,232–4,211 of the *P. y. yoelii* 17XNL contig MALPY01471 (GenBank accession no. AABL01001466) (10), were synthesized to amplify a full-length coding region. PCR reactions were performed in a 50-μL volume consisting of 100 ng of genomic DNA in 1× Ultra-high fidelity Accuzyme mix (Bioline). PCR conditions were as follows: one cycle at 94 °C for 5 min; followed by 35 cycles at 94 °C for 30 s, 55 °C for 1 min, and 68 °C for 4 min; and a final extension at 68 °C for 5 min. The products were treated with shrimp alkaline phosphatase and exonuclease I before sequencing (38). Sequencing reactions were performed in triplicate from different amplification tubes using an ABI Prism BigDye Terminator ready-to-use reaction kit (Applied Biosystems). Sequencing primers were described previously (8).

MSs used in this study have been described previously (6, 36). Briefly, PCR products without any labeling procedures were separated using capillary electrophoresis performed in a QIAxcel machine (QIAGEN) according to the manufacturer's instructions. Genotypes from genetic cross progeny were scored by matching the PCR product sizes with those from the parents. The genetic distances between the parasites were calculated using methods described (36).

***P. yoelii* Genetic Crosses.** Three combinations of genetically distinct clones of *P. yoelii* were chosen to produce genetic crosses. Two genetic crosses of (*i*) N67 (lethal) × 17XNL (nonlethal) and (*ii*) YM (lethal) × 33X (nonlethal) were performed at the NIH laboratory, whereas the genetic cross between BY265 and NSM was conducted at the Third Military Medical University and Xiamen University of China. The genetic cross between YM and A/C was performed in David Walliker's laboratory at the University of Edinburgh, and seven recombinant clones have been cryopreserved since 1976 (12). Because we could only handle a limited number of mice at a time, we cloned progeny from multiple individual crosses of each parental pair. The crosses were also performed with different phenotypes in mind. For example, we were interested in differences in GRVPs in the N67 × 17XNL and YM × 33X crosses and drug resistance in the BY265 × NSM cross (NSM is more resistant to mefloquine and chloroquine).

The experimental procedures for the production of genetic crosses in *P. yoelii* have been described previously (8, 39). Briefly, inbred C57BL/6 or Kunming outbred (BY265 × NSM cross) female mice were coinfected with two parasites (parents) to be crossed. Parasitemia (percent of parasitized erythrocytes from at least 1,000 cells) from the mice was monitored daily by microscopic examination of Giemsa-stained thin blood smears. On day 4 postinfection, each infected mouse with male and female gametocytes (av-

erage parasitemia ~10–30%) was anesthetized and fed to 50–100 female mosquitoes per infected mouse. The mouse was then euthanized while under anesthesia. Seventeen days after feeding, sporozoites were harvested from infected mosquitoes, and 1×10^5 to 5×10^5 sporozoites were injected i.p. into female CD-1 mice to obtain blood-stage parasites representing "uncloned progeny" of the genetic cross. When blood-stage parasites were microscopically detectable (~0.1–1% parasitemia), blood samples were drawn from the infected mice and diluted to 0.6–1.5 iRBCs per 100 μL (inoculum size) before injection i.v. into recipient mice. Five to nine days after injection, mice were monitored for the presence of blood-stage malaria parasites. A small volume (20–50 μL) of mouse tail blood was withdrawn and used for parasite genomic DNA extraction and MS genotyping. For screening of recombinant progeny, DNA from cross progeny was typed with a panel of MS markers distributed on chromosomes 1–14 (Dataset S1). Recombinant progeny were considered clonal when single MS alleles were found in the MS loci. To improve the recovery of independent progeny, the parasites were cloned as early as possible to prevent "overgrowth" of some fast-growing parasites. Moreover, the parental input ratio was adjusted appropriately for each particular cross based on the different growth rates of each pair of parents.

Development of Genetic Linkage Maps. We constructed a genetic linkage map from the MS genotypes and segregation patterns in the progeny of each individual cross using Mapmaker/Exp 3.0 (40) to order the markers and estimate genetic distances between them. We used 37 MS markers within the contigs that had been physically assigned to chromosomes previously (11, 14–16) as anchors for assigning linkage groups to chromosomes (Dataset S1 and Table S2).

GRVP and Linkage Analysis. To map the determinant(s) that affect the GRVP between N67 and 17XNL, we evaluated the growth rates of the progeny from the cross between 17XNL and N67 in female C57BL/6 mice (4–8 mice per single parasite clone). Each mouse was injected i.v. with an inoculum containing 1×10^5 iRBCs. Parasitemias were monitored daily, as measured by microscopic examination of Giemsa-stained thin tail blood smears, and recorded in Excel.

QTL Analysis and Statistics. For genome-wide scans of the day 5 parasitemia, the association between categorical trait value vectors and genotype vectors was evaluated using Shannon's mutual information (41) with the genome-wide $P < 0.001$ threshold (Fig. 4*A*, dashed horizontal line) estimated using a noncentral χ^2 test confirmed by simulation (41). The simulation estimated the null distribution of mutual information under nonassociation using 1,000 permutations and the method described by Nelson and O'Brien (42). The day 10 parasitemia was treated as a continuous trait, and associated QTLs were mapped using a *t* test statistic and 1,000 permutations. We also ran scans based on a nonparametric Wilcoxon test for all phenotypes to provide independent support for the associations, which gave essentially the same results.

QTL mapping was also conducted using the R/qtl library in R2.12.2 software (18). Day 5 parasitemia from the 25 progeny of the N67 and 17XNL cross was natural-log transformed to obtain a normally distributed dataset. One thousand permutations with a 5% threshold ($P = 0.05$) were performed using standard interval mapping (the expectation and maximization method). A single QTL genome scan was done first; a 2D scan and, finally, two QTL genome scans were then performed because of two similarly significant loci on chromosome 10 and chromosome 13. In principle, the 2D and two QTL scans could identify additional QTL and two-locus epistatic interactions if they are present.

Statistical Analysis of Inheritance Bias. We measured the statistical significance of deviations from the expected 1:1 segregation ratio. For each marker, we computed the Kendall τ-rank association statistic between the observed parents and progeny genotype vector and each of the two vectors, simulating complete bias (i.e., with all progeny assigned as one parental type or the other). The S-plus function cor.test was used with the method "kendall," where probabilities of dependence were based on the normal approximation as described by Prokhorov (43).

ACKNOWLEDGMENTS. We thank Drs. Karl W. Broman and Na Li for advice on QTL analysis and National Institute of Allergy and Infectious Diseases intramural editor Brenda Rae Marshall for assistance. This work was supported by grants from the National Basic Research Program of China, 973 Program (Grant 2007CB513103), the Science Planning Program of Fujian Province (Grant 2010J1008), and the 111 Project of Education of China (Grant B06016) as well as by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. J.C.W. was supported by the Intramural Program of the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health.

1. Su X, et al. (1999) A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science* 286:1351–1353.
2. Martinelli A, et al. (2005) A genetic approach to the de novo identification of targets of strain-specific immunity in malaria parasites. *Proc Natl Acad Sci USA* 102:814–819.
3. Hayton K, et al. (2008) Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of Plasmodium falciparum invasion. *Cell Host Microbe* 4: 40–51.
4. Su X, Hayton K, Wellems TE (2007) Genetic linkage and association analyses for trait mapping in Plasmodium falciparum. *Nat Rev Genet* 8:497–506.
5. Martinelli A, et al. (2005) An AFLP-based genetic linkage map of Plasmodium chabaudi chabaudi. *Malar J* 4:11.
6. Li J, et al. (2009) Hundreds of microsatellites for genotyping Plasmodium yoelii parasites. *Mol Biochem Parasitol* 166:153–158.
7. Culleton R, Martinelli A, Hunt P, Carter R (2005) Linkage group selection: Rapid gene discovery in malaria parasites. *Genome Res* 15:92–97.
8. Pattaradilokrat S, Culleton RL, Cheesman SJ, Carter R (2009) Gene encoding erythrocyte binding ligand linked to blood stage multiplication rate phenotype in Plasmodium yoelii yoelii. *Proc Natl Acad Sci USA* 106:7161–7166.
9. Otsuki H, et al. (2009) Single amino acid substitution in Plasmodium yoelii erythrocyte ligand determines its localization and controls parasite virulence. *Proc Natl Acad Sci USA* 106:7167–7172.
10. Carlton JM, et al. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature* 419:512–519.
11. Kooij TW, et al. (2005) A Plasmodium whole-genome synteny map: Indels and synteny breakpoints as foci for species-specific genes. *PLoS Pathog* 1:e44.
12. Walliker D, Sanderson A, Yoeli M, Hargreaves BJ (1976) A genetic investigation of virulence in a rodent malaria parasite. *Parasitology* 72:183–194.
13. Pattaradilokrat S, Cheesman SJ, Carter R (2008) Congenicity and genetic polymorphism in cloned lines derived from a single isolate of a rodent malaria parasite. *Mol Biochem Parasitol* 157:244–247.
14. Uzureau P, Barale JC, Janse CJ, Waters AP, Breton CB (2004) Gene targeting demonstrates that the Plasmodium berghei subtilisin PbSUB2 is essential for red cell invasion and reveals spontaneous genetic recombination events. *Cell Microbiol* 6:65–78.
15. Thompson J, et al. (2004) PTRAMP; a conserved Plasmodium thrombospondin-related apical merozoite protein. *Mol Biochem Parasitol* 134:225–232.
16. Janse CJ, Carlton JM, Walliker D, Waters AP (1994) Conserved location of genes on polymorphic chromosomes of four species of malaria parasites. *Mol Biochem Parasitol* 68:285–296.
17. Khan A, et al. (2005) Composite genome map and recombination parameters derived from three archetypal lineages of Toxoplasma gondii. *Nucleic Acids Res* 33:2980–2992.
18. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890.
19. Preiser PR, Jarra W, Capiod T, Snounou G (1999) A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* 398:618–622.
20. Walliker D, Carter R, Morgan S (1971) Genetic recombination in malaria parasites. *Nature* 232:561–562.
21. Walliker D, et al. (1987) Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science* 236:1661–1666.
22. Wellems TE, et al. (1990) Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature* 345:253–255.
23. Grech K, et al. (2002) Numerous, robust genetic markers for Plasmodium chabaudi by the method of amplified fragment length polymorphism. *Mol Biochem Parasitol* 123: 95–104.
24. Su X, Kirkman LA, Fujioka H, Wellems TE (1997) Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. *Cell* 91:593–603.
25. Withers-Martinez C, et al. (2008) Malarial EBA-175 region VI crystallographic structure reveals a KIX-like binding interface. *J Mol Biol* 375:773–781.
26. Khan SM, Jarra W, Bayele H, Preiser PR (2001) Distribution and characterisation of the 235 kDa rhoptry multigene family within the genomes of virulent and avirulent lines of Plasmodium yoelii. *Mol Biochem Parasitol* 114:197–208.
27. Culleton R, Kaneko O (2010) Erythrocyte binding ligands in malaria parasites: Intracellular trafficking and parasite virulence. *Acta Trop* 114:131–137.
28. Reilly Ayala HB, Wacker MA, Siwo G, Ferdig MT (2010) Quantitative trait loci mapping reveals candidate pathways regulating cell cycle duration in Plasmodium falciparum. *BMC Genomics* 11:577.
29. Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC (2001) Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. *Science* 294:161–165.
30. Pattaradilokrat S, Cheesman SJ, Carter R (2007) Linkage group selection: Towards identifying genes controlling strain specific protective immunity in malaria. *PLoS ONE* 2:e857.
31. Vaidya AB, et al. (1995) A genetic locus on Plasmodium falciparum chromosome 12 linked to a defect in mosquito-infectivity and male gametogenesis. *Mol Biochem Parasitol* 69:65–71.

32. Ferdig MT, et al. (2004) Dissecting the loci of low-level quinine resistance in malaria parasites. *Mol Microbiol* 52:985–997.
33. Hayton K, Su XZ (2008) Drug resistance and genetic mapping in Plasmodium falciparum. *Curr Genet* 54:223–239.
34. Gonzales JM, et al. (2008) Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol* 6:e238.
35. Yuan J, et al. (2009) Genetic mapping of targets mediating differential chemical phenotypes in Plasmodium falciparum. *Nat Chem Biol* 5:765–771.
36. Li J, et al. (2007) Typing Plasmodium yoelii microsatellites using a simple and affordable fluorescent labeling method. *Mol Biochem Parasitol* 155:94–102.
37. Yoeli M, Hargreaves B, Carter R, Walliker D (1975) Sudden increase in virulence in a strain of Plasmodium berghei yoelii. *Ann Trop Med Parasitol* 69:173–178.
38. Mu J, et al. (2007) Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nat Genet* 39:126–130.
39. Pattaradilokrat S, Li J, Su XZ (2011) Protocol for production of a genetic cross of the rodent malaria parasites. *J Vis Exp*, 10.3791/2365.
40. Lander E, et al. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
41. Dawy Z, et al. (2006) Gene mapping and marker clustering using Shannon's mutual information. *IEEE/ACM Trans Comput Biol Bioinformatics* 3:47–56.
42. Nelson GW, O'Brien SJ (2006) Using mutual information to measure the impact of multiple genetic factors on AIDS. *J Acquir Immune Defic Syndr* 42:347–354.
43. Prokhorov AV (2001) Kendall coefficient of rank correlation. *Online Encyclopedia of Mathematics*, ed Hazewinkel M (Springer, CWI, Amsterdam).

Tropical Medicine and Health Vol. 39 No. 3, 2011, pp. 77-82
doi:10.2149/tmh.2011-12

Original article

Positive selection on the *Plasmodium falciparum clag2* gene encoding a component of the erythrocyte-binding rhoptry protein complex

Jean SF Alexandre[1,2], Morakot Kaewthamasorn[1], Kazuhide Yahata[1], Shusuke Nakazawa[1] and Osamu Kaneko[1*]
Received 23 May, 2011 Accepted 24 May, 2011 Published online 6 August, 2011

Abstract: A protein complex of high-molecular-mass proteins (*Pf*RhopH) of the human malaria parasite *Plasmodium falciparum* induces host protective immunity and therefore is a candidate for vaccine development. Clarification of the level of polymorphism and the evolutionary processes is important both for vaccine design and for a better understanding of the evolution of cell invasion in this parasite. In a previous study on 5 genes encoding RhopH1/Clag proteins, positive diversifying selection was detected in *clag8* and *clag9* but not in the paralogous *clag2*, *clag3.1* and *clag3.2*. In this study, to extend the analysis of *clag* polymorphism, we obtained sequences surrounding the most polymorphic regions of *clag2*, *clag8*, and *clag9* from parasites collected in Thailand. Using sequence data obtained newly in this study and reported previously, we classified *clag2* sequences into 5 groups based on the similarity of the deduced amino acid sequences and number of insertions/deletions. By the sliding window method, an excess of nonsynonymous substitutions over synonymous substitutions was detected in the group 1 and group 2 *clag2* and *clag8* sequences. Population-based analyses also detected a significant departure from the neutral expectation for group 1 *clag2* and *clag8*. Thus, two independent approaches suggest that *clag2* is subject to a positive diversifying selection. The previously suggested positive selection on *clag8* was also supported by population-based analyses. However, the positive selection on *clag9*, which was detected by comparing the 5 sequences, was not detected using the additional 34 sequences obtained in this study.
Key words: malaria, rhoptry, polymorphism

1. INTRODUCTION

Plasmodium falciparum, the causative agent of malignant tertian malaria is, for part of its life cycle, an obligate intra-erythrocytic parasite. This stage of the life cycle is characterized by repeating cycles of erythrocyte invasion, followed by growth and schizogonic multiplication within the cell, egress, and re-invasion. The erythrocyte invasive unit, the merozoite, expresses a panel of proteins on its surface, all of which are exposed to the host immune system for the brief time that the parasite is free in the blood plasma. These proteins, exemplified by merozoite surface protein 1 (MSP1), tend to be more polymorphic (showing a higher degree of amino acid diversity between parasite strains) than other non-exposed parasite proteins. A possible explanation for this is that such proteins have undergone "positive diversifying selection", mediated by host immune pressure [1]. Proteins secreted by the parasite at the point of cell invasion are also exposed, however briefly, to the host immune system, and may be targeted by immune pressures, and polymorphism selected within them, in much the same way as merozoite surface proteins. Apical membrane antigen (AMA1), released from the merozoite micronemes during invasion, for example, is believed to have undergone positive diversifying selection in this manner [2].

The RhopH complex is a high molecular mass

[1] Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global COE Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
[2] Centro Nacional de Control de Enfermedades Tropicales, Santo Domingo, República Dominicana
*Corresponding author:
Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global COE Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
Tel.: (+81) 95 819 7838
Fax: (+81) 95 819 7805
E-mail: okaneko@nagasaki-u.ac.jp
Sequence data from this article were deposited in the GenBank™/EMBL/DDBJ databases under accession numbers AB633214–AB633323, and AB634454.
Abbreviations: *Clag*, cytoadherence-linked asexual gene; DNA, deoxyribo–nucleotide acid; indels, insertions/deletions; nt, nucleotides; PCR, polymerase chain reaction

erythrocyte-binding protein complex secreted from the merozoite rhoptry during invasion, and antibodies raised against it have been shown to confer anti-parasite protection to the host [3–6]. The precise role of this complex in erythrocyte invasion remains unclear. Components of this complex have been detected on the erythrocyte cytosol side of the parasitophorous vacuole membrane and the parasite-infected erythrocyte membrane, suggesting a role in the formation of the parasite-restructured membranous architecture in the infected erythrocyte [7, 8]. The RhopH complex is itself comprised of three distinct proteins: RhopH1, RhopH2 and RhopH3, each encoded by separate genes [9–11]. RhopH1 is encoded by a multigene family termed the cytoadherence-linked asexual gene (*rhoph1/clag*) family, that consists of at least five paralogous genes (*clag2*, *clag3.1*, *clag3.2*, *clag8*, and *clag9*), and each RhopH complex contains one of the *rhoph1/clag* gene products [9, 12, 13].

In a previous study, the degree of inter-allelic polymorphism for seven RhopH complex-related genes, 5 *rhoph1/clag*, *rhoph2* and *rhoph3*, was evaluated by comparing nucleotide sequences from 5 culture-adapted parasite lines. It was found that *clag2*, *clag3.1*, *clag3.2* and *clag8* were highly polymorphic and that amino acid substitutions and insertions/deletions (indels) were found mainly in a region encompassing amino acid positions 1000–1200 of these gene products. The signature of positive selection was detected by an excess of nonsynonymous substitutions over synonymous substitutions for *clag8* and *clag9*. Further evidence for the involvement of positive selection in driving *clag8* polymorphism was independently gathered in an analysis that compared the *Plasmodium reichenowi* orthologous sequence for *clag8* to the *P. falciparum* gene [14].

In this study, we extend our analysis of *clag* polymorphism further by increasing the number of sequences and employing a population-based approach utilizing sequences obtained from parasites collected in Thailand.

2. MATERIALS AND METHODS

2.1. Parasite culture and DNA extraction

All 39 *P. falciparum* lines examined in this study (MS802, MS803, MS804, MS805, MS806, MS807, MS808, MS809, MS810, MS811, MS812, MS814, MS815, MS816, MS817, MS818, MS819, MS820, MS821, MS822, MS824, MS825, MS826, MS827, MS828, MS829, MS830, MS831, MS833, MS834, MS835, MS837, MS838, MS840, MS842, MS843, MS844, MS946, and MS947) were collected in Mae Sot, Thailand from November 21, 1988 to January 16, 1989 and were maintained in vitro essentially as previously described [15–17]. Since MS814 and MS822 showed unclear chromatograms for *clag* sequences, these lines were cloned by limiting dilution, yielding clones MS814K and MS814R, and MS822B6 and MS822G8, respectively. The human erythrocytes and plasma used for culture were obtained from the Nagasaki Red Cross Blood Center. Parasites were harvested when parasitemia reached about 2%, and parasite DNA was extracted using DNAzol BD (Invitrogen).

2.2. Polymerase chain reaction (PCR) amplification and sequencing

DNA fragments were PCR-amplified twice, independently, with oligonucleotide primers: TATATGGAAAAA GTAGTAATATACAGG and TACTAGTATGTGGTTGAT ATTCTTTTG for *clag2* (resulting PCR product with the size of 702 bp); GTTTATGGAAAAAGTGGTAAAATAGG and CTCTTTAAGTTTTCTTCTGAATAGTTC for *clag8* (750 bp); and ATAAACTTGATAGAATATATGGTAAAGC and ATTGAATAATCTTTTAATGTACATGCAC for *clag9* (764 bp) in a 20 μL reaction mixture using a high-fidelity KOD Plus DNA polymerase (TOYOBO, Japan). The PCR conditions were as follows: 94°C for 2 min; 40 cycles of 92°C for 15 sec, 54°C for 20 sec, 68°C for 1 min 10 sec; final extension step of 68°C for 5 min. PCR products were subjected to 1.5% agarose gel electrophoresis, and when a single band product with no background was observed, PCR-amplified DNA fragments were directly sequenced following treatment of PCR mixture with ExoSAP-IT (GE Healthcare, UK). Two independent PCR products were sequenced using a panel of primers described previously [14]; one in the forward direction and the other in the reverse direction, to ensure the accuracy of the obtained sequences. Sequencing reactions were performed using the BigDye® Terminator v1.1 Kit (Applied BioSystems, UK) with an ABI3730 DNA analyzer (Applied BioSystems). Sequences were manually corrected using BioEdit 7.0.0 software [18]. Sequences of *clag2* (AB250822) and *clag8* (AB250849) for MS838 were obtained from the database.

2.3. Statistical analyses of genetic diversity

Nucleotide diversity (π) and its standard error (SE) were computed by the Jukes and Cantor method using MEGA 4.0 software [19]. The mean number of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) and their standard errors were computed using the Nei and Gojobori method [20] with the Jukes and Cantor correction, implemented in MEGA 4.0. The statistical difference between d_S and d_N was tested using a one-tailed Z-test with 500 bootstrap pseudo-samples in MEGA 4.0. A value of d_N significantly higher than d_S at the 95% confidence level was taken as evidence for positive selection. The d_N/d_S ratio was evaluated using a sliding window method (90 bases with a step size of 3

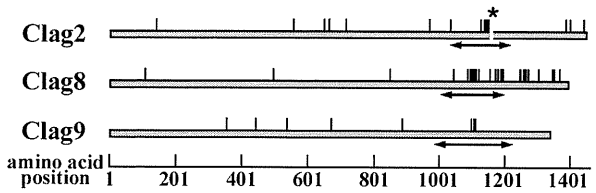

Fig. 1. Regions used for the analysis. Bars with arrows under each schematic for Clag protein indicate the regions used for the analysis in this study. Vertical bars and a gap with asterisk above each schematic indicate polymorphic sites among 5 *P. falciparum* laboratory isolates and an indel reported previously [14]. Amino acid position is according to the 3D7 line sequence.

bases) in DnaSP 4.0 [21]. Population genetics tests of neutrality were applied to the 3 *clag* gene sequences. Tajima's test was used to test for departure from neutrality as measured by the difference between π (observed average pairwise nucleotide diversity) and θ (expected nucleotide diversity under neutrality derived from the number of segregating sites, *S*). Under positive diversifying selection, rare alleles are selected and maintained at intermediate frequencies, elevating π above that expected under neutrality and making the value of the test statistic (*D*) positive [22]. Fu and Li's test was also used to evaluate positive diversifying selection by comparing estimates of θ based on the number of singletons and that derived from *S* (the *D** index) or π (the *F** index). Under positive diversifying selection, an excess of intermediate frequency polymorphisms and lower number of singletons make the value of *D** and *F** positive [23]. The regions analyzed in this study are shown in Fig. 1.

3. RESULTS AND DISCUSSION

3.1. Clag2 sequence is classified into 5 groups with the signature of positive diversifying selection detected on groups 1 and 2.

In this study, we obtained 35 sequences of the most polymorphic region of *clag2*, 40 sequences for *clag8* and 34 sequences for *clag9*. Iriko *et al.* (2008) did not detect the signature of positive diversifying selection on *clag2* using a set of sequences which excluded the most polymorphic regions with indels, due to difficulty in obtaining a reliable nucleotide sequence alignment [14]. Thus, we classified *clag2* into distinct groups using the deduced amino acid sequences from 35 *P. falciparum clag2* alleles generated during this work, 31 previously reported alleles and one *clag2* ortholog (*prclag2*) from the chimpanzee malaria parasite *Plasmodium reichenowi*, based on amino acid similarity and numbers of indels, and attempted to detect the signature of positive selection for each group separately. We classified *Pf*Clag2 sequences into five groups (Fig. 2). Among 66 *Pf*Clag2 sequences, 44 sequences (including the 3D7 line sequence) were classified in group 1, 11 sequences with a double amino acid insertion in group 2, five sequences with a seven amino acid insertion in group 3, four sequences with an eight amino acid insertion in group 4, and two sequences showing a high degree of similarity to the *P. reichenowi* Clag2 ortholog sequence in group 5. The observations that 1) amino acid sequences are clearly distinct between groups, 2) the group 5 sequence is similar to the *P. reichenowi* Clag2 ortholog sequence, and 3) the older origin of *clag2* polymorphism than the time to the most recent common ancestor of the extant *P. falciparum* population was previously proposed based on the analysis using sequences excluding indels [14], raise the possibility that these 5 distinct Clag2 groups may have been generated before the time to the most recent common ancestor of the extant *P. falciparum* population. This may be clarified by analyzing the sequences of Clag2 orthologs from the malaria parasite species recently discovered in gorillas, and currently thought to be the closest relative species [25].

We attempted to detect the signature of positive diversifying selection on *clag2* alleles belonging to groups 1 (n = 24) and 2 (n = 8) and *clag8* (n = 41) by comparing d_N and d_S for sequences obtained in Thailand during the 1988–1989 period ("Thai" in Fig. 3), or 44, 11, or 69 sequences, respectively, after combining sequences reported previously ("Thai + others" in Fig. 3) [14]. No significant difference was detected between d_N and d_S, when the nucleotide sequences corresponding to nucleotide position (nt) 3106–3642 for *clag2* or nt 3022–3591 for *clag8* (numbered according to the 3D7 line sequence) were analyzed. However, when d_N and d_S were compared for the region where a high d_N/d_S ratio was detected by the sliding window plot method, a significant excess of d_N over d_S was detected for group 1 *clag2* (mid point nt position of 3444 for "Thai + others", d_N/d_S = 3.13, $p<0.02$; mid point nt position of 3447–3453 for "Thai + others", d_N/d_S = 3.13–3.27, $p<0.05$), group 2 *clag2* (mid point nt position of 3396–3459 for "Thai" only, d_N/d_S = 5.90–7.65, $p<0.02$; mid point nt position of 3396–3459 for "Thai + others", d_N/d_S = 8.02–10.02, $p<0.02$), and *clag8* (mid point nt position of 3258–3309 for Thai only, d_N/d_S = 3.49–53.96, $p<0.02$; mid point nt position of 3258–3309 for Thai + others, d_N/d_S = 3.21–33.27, $p<0.02$; mid point nt position of 3531–3546 for Thai + others, d_N/d_S = 6.99–7.68, $p<0.05$), suggesting that positive diversifying selection had operated on these regions. Positive selection was not detected for *clag9* using sequences from Thai only (n = 34) and Thai + others (n = 39). Thus, in

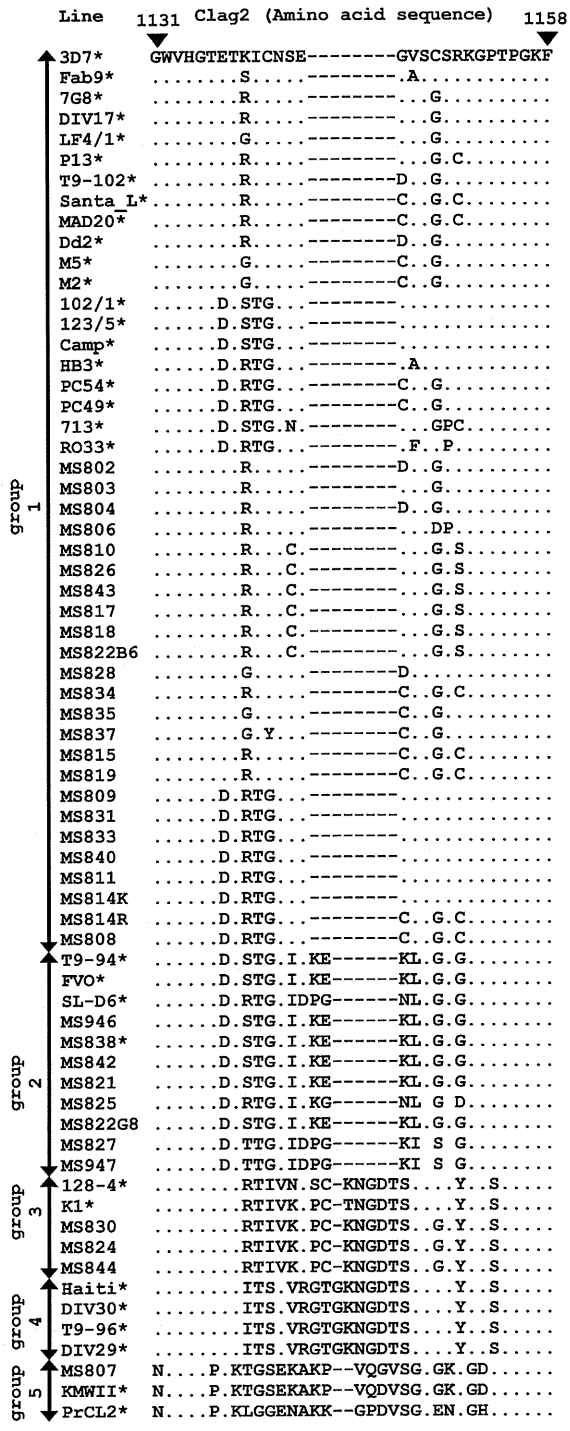

Fig. 2. Clag2 is classified into 5 groups based on the similarity in the variable region. Dots and bars indicate identical amino acid residues with the 3D7 line sequence and gaps. Asterisks indicate sequences reported previously [14]. PrCL2 indicate *Plasmodium reichenowi* Clag2. Amino acid positions shown above the sequences are according to the 3D7 line sequence.

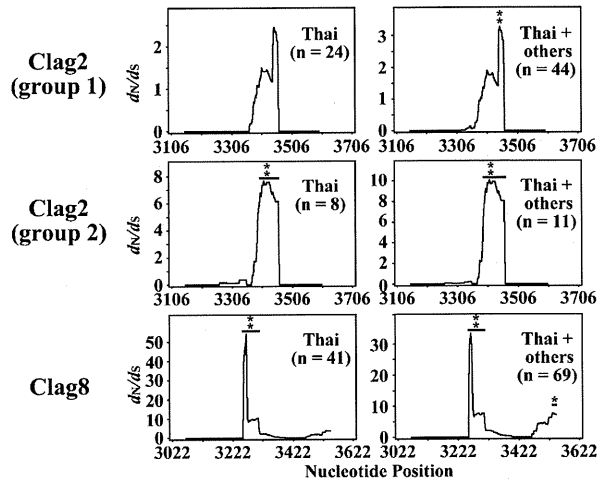

Fig. 3. Sliding window plot of d_N/d_S ratio for *Plasmodium falciparum clag2* (group 1 and 2) and *clag8* in Thai isolates and in a set combined with laboratory isolates. Nucleotide positions are according to the 3D7 line sequence. Window length is 90 bp, and step size is 3 bp. n indicates the number of samples analyzed. Asterisks indicate the region where a significant excess of non-synonymous substitutions (d_N) over synonymous substitutions (d_S) was observed (single for $p<0.05$, and double for $p<0.02$). The statistical difference between d_S and d_N was tested using a one-tailed Z-test with 500 bootstrap pseudosamples implemented in MEGA4.0. Others indicate sequence reported previously and shown in Fig. 2 [14].

addition to *clag8* and *clag9*, for which a positive selection was detected in a previous study [14], positive diversifying selection was detected here for at least two groups of *clag2* sequences (Fig. 3).

3.2. Population-based analyses also detected positive selection on group 1 *clag2* and *clag8*

Positive diversifying selection was evaluated for group 1 *clag2* (n = 24), *clag8* (n = 41), and *clag9* (n = 34) using sequences obtained from Thailand within two months by a population-based approach. Using Tajima's test and Fu and Li's test for nt sequences obtained in this study (nt 3106–3642 of *clag2*, 3022–3591 of *clag8*, and nt 2947–3639 of *clag9*; nucleotide positions are according to the 3D7 line sequence), no significant departure from the neutral expectation was detected. However, the sliding window plot method detected a significantly high Tajima's D value for group 1 *clag2* (mid point nt position of 3492, $D = 2.26$, $p<0.05$) and *clag8* (mid point nt position of 3363, $D = 2.07$, $p<0.05$), Fu and Li's D^* value for *clag8* (mid point nt position of 3258–3264, $D^* = 1.48$–1.56, $p<0.05$; mid point

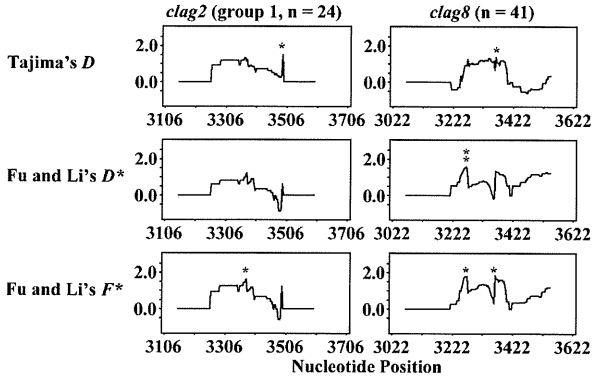

Fig. 4. Sliding window plot of Tajima's *D*, Fu & Li's *D** and *F** tests for *Plasmodium falciparum clag2* (group 1) and *clag8* genes in Thai isolates. Nucleotide positions are according to the 3D7 line sequence. Window length is 90 bp, and step size is 3 bp. n indicates the number of samples analyzed. Asterisks indicate the region where a significant departure from the neutrality was observed (single for $p<0.05$, and double for $p<0.02$).

nt position of 3267, $D^* = 1.61$, $p<0.02$), and Fu and Li's F^* value for group 1 *clag2* (mid point nt position of 3375, $F^* = 1.64$, $p<0.05$) and *clag8* (mid point nt position of 3261–3270, $F^* = 1.75$–1.81, $p<0.05$; mid point nt position of 3363, $F^* = 1.86$, $p<0.05$) (Fig. 4). Thus, in addition to the detection of positive selection on group 1 *clag2* and *clag8* detected by an excess of d_N over d_S, the results of the population-based method also support the finding that both group 1 *clag2* and *clag8* are subject to a positive diversifying selection.

In summary, two independent tests, one by comparing d_N and d_S and the other based on the population, suggest that the region of *clag2* is under positive diversifying selection. Previously suggested positive selection on *clag8* was also supported by population-based analyses. Such observations are consistent with the action of balancing selection maintaining allelic variation in the population. Contrary to a previous report [14], positive selection on *clag9* was not detected in this study using an additional 34 sequences. As only 5 *clag9* sequences were used to compare d_N and d_S in this previous study, a low sample number might have generated a false positive result. Further study is required to determine if *clag9* is under positive selection.

ACKNOWLEDGEMENTS

We thank R. Culleton for critical reading. We are grateful to I. Sekine, head of the Nagasaki Red Cross Blood Center for human erythrocyte and plasma. This work was supported in part by Grants-in-Aids for Scientific Research 19590428 (to OK) and the Global COE Program, Nagasaki University (to OK) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The nucleotide sequence data reported in this paper are available in the GenBank™/EMBL/DDBJ databases under the accession numbers: AB633214–AB633323, and AB634454. Part of this work was performed during Master of Tropical Medicine course in Nagasaki University supported by the Japanese International Cooperation Agency (JICA). J.A. is a recipient of MEXT PhD scholarship, Japan.

REFERENCES

1. Conway DJ. Natural selection on polymorphic malaria antigens and the search for a vaccine. Parasitol Today 1997; 13: 26–29.
2. Crewther PE, Matthew ML, Flegg RH, Anders RF. Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. Infect Immun 1996; 64: 3310–3317.
3. Siddiqui WA, Tam LQ, Kramer KJ, *et al.* Merozoite surface coat precursor protein completely protects Aotus monkeys against *Plasmodium falciparum* malaria. Proc Natl Acad Sci USA 1987; 84: 3014–3018.
4. Cooper JA, Ingram LT, Bushell GR, *et al.* The 140/130/105 kilodalton protein complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. Mol Biochem Parasitol 1988; 29: 251–260.
5. Doury JC, Bonnefoy S, Roger N, *et al.* Analysis of the high molecular weight rhoptry complex of *Plasmodium falciparum* using monoclonal antibodies. Parasitology 1994; 18: 269–280.
6. Rungruang T, Kaneko O, Murakami Y, *et al.* Erythrocyte surface glycosyl-phosphatidyl inositol anchored receptor for the malaria parasite. Mol Biochem Parasitol 2005; 140: 13–21.
7. Hiller NL, Akompong T, Morrow JS, Holder AA, Haldar K. Identification of a stomatin orthologue in vacuoles induced in human erythrocytes by malaria parasites. A role for microbial raft proteins in apicomplexan vacuole biogenesis. J Biol Chem 2003; 278: 48413–48421.
8. Goel S, Valiyaveettil M, Achur RN, *et al.* Dual stage synthesis and crucial role of cytoadherence-linked asexual gene 9 in the surface expression of malaria parasite *var* proteins. Proc Natl Acad Sci USA 2010; 107: 16643–16648.
9. Kaneko O, Tsuboi T, Ling IT, *et al.* The high molecular mass rhoptry protein, RhopH1, is encoded by members of the *clag* multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2001; 118: 223–231.
10. Ling IT, Kaneko O, Narum DL, *et al.* Characterization of the *rhoph2* gene of *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 2003; 127: 47–57.
11. Shirano M, Tsuboi T, Kaneko O, Tachibana M, Adams JH, Torii M. Conserved regions of the *Plasmodium yoelii*

rhoptry protein RhopH3 revealed by comparison with the *P. falciparum* homologue. Mol Biochem Parasitol 2001; 112: 297–299.
12. Holt DC, Gardiner DL, Thomas EA, *et al.* The cytoadherence linked asexual gene family of Plasmodium falciparum: are there roles other than cytoadherence? Int J Parasitol 1999; 29: 939–944.
13. Kaneko O, Yim Lim BY, Iriko H, *et al.* Apical expression of three RhopH1/Clag proteins as components of the *Plasmodium falciparum* RhopH complex. Mol Biochem Parasitol 2005; 143: 20–28.
14. Iriko H, Kaneko O, Otsuki H, *et al.* Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol 2008; 158: 11–21.
15. Trager W, Jensen JB. Human malaria parasites in continuous culture. Science 1976; 193: 673–675.
16. Jongwutiwes S, Tanabe K, Kanbara H. Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of *Plasmodium falciparum* from field isolates. Mol Biochem Parasitol 1993; 59: 95–100.
17. Nakazawa S, Culleton R, Maeno Y. In vivo and in vitro gametocyte production of *Plasmodium falciparum* isolates from Northern Thailand. Int J Parasitol 2011; 41: 317–323.
18. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999; 41: 95–98.
19. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24: 1596–1599.
20. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986; 3: 418–426.
21. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003; 19: 2496–2497.
22. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 1993; 135: 599–607.
23. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993; 133: 693–709.
24. Liu W, Li Y, Learn GH, *et al.* Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 2010; 467: 420–425.

Parasitology International 60 (2011) 313–320

Contents lists available at ScienceDirect

Parasitology International

journal homepage: www.elsevier.com/locate/parint

PEXEL-independent trafficking of *Plasmodium falciparum* SURFIN$_{4.2}$ to the parasite-infected red blood cell and Maurer's clefts

Jean Semé Fils Alexandre [a,b], Kazuhide Yahata [a], Satoru Kawai [c], Motomi Torii [d], Osamu Kaneko [a,*]

[a] Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global Center of Excellence Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
[b] Centro Nacional de Control de Enfermedades Tropicales, Santo Domingo, Dominican Republic
[c] Center for Tropical Medicine and Parasitology, Dokkyo Medical University, Tochigi 321-0293, Japan
[d] Department of Molecular Parasitology, Ehime University Graduate School of Medicine, Shitsukawa, Toon, Ehime 791-0925, Japan

ARTICLE INFO

Article history:
Received 22 April 2011
Received in revised form 9 May 2011
Accepted 10 May 2011
Available online 17 May 2011

Keywords:
Malaria
Maurer's clefts
Plasmodium falciparum
Protein trafficking
SURFIN

ABSTRACT

SURFIN$_{4.2}$ is a parasite-infected red blood cell (iRBC) surface associated protein of *Plasmodium falciparum*. To analyze the region responsible for the intracellular trafficking of SURFIN$_{4.2}$ to the iRBC and Maurer's clefts, a panel of transgenic parasite lines expressing recombinant SURFIN$_{4.2}$ fused with green fluorescent protein was generated and evaluated for their localization. We found that the cytoplasmic region containing a tryptophan rich (WR) domain is not necessary for trafficking, whereas the transmembrane (TM) region was. Two PEXEL-like sequences were shown not to be responsible for the trafficking of SURFIN$_{4.2}$, demonstrating that the protein is trafficked in a PEXEL-independent manner. N-terminal replacement, deletion of the cysteine-rich domain or the variable region also did not prevent the protein from localizing at the iRBC or Maurer's clefts. A recombinant SURFIN$_{4.2}$ protein possessing 50 amino acids upstream of the TM region, TM region itself and a part of the cytoplasmic region was shown to be trafficked into the iRBC and Maurer's clefts, suggesting that there are no essential trafficking motifs in the SURFIN$_{4.2}$ extracellular region. A mini-SURFIN$_{4.2}$ protein containing WR domain was shown by Western blotting to be more abundantly detected in a Triton X-100-insoluble fraction, compared to the one without WR domain. We suggest that the cytoplasmic region containing the WR may be responsible for their difference in solubility.

1. Introduction

During its asexual replication in the human host, *Plasmodium falciparum*, the apicomplexan parasite responsible for malaria, dramatically remodels the infected red blood cell (iRBC) [1]. This process involves the generation of a parasitophorous vacuole (PV) in which parasites reside and replicate, the transportation of parasite proteins into the iRBC across the PV membrane (PVM), the generation of parasite-derived membranous structures in the cytoplasm of the host RBC called Maurer's clefts that play a major role as protein-sorting points, and the formation of knobs on the iRBC surface [2–4]. Some of the most severe malaria pathologies caused by *P. falciparum*, such as cerebral and placental malaria, are specifically linked to the adherence of the iRBCs to capillary vessels (cytoadhesion) and to uninfected RBCs (rosseting). *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1), a parasite protein transported to the surface of the iRBC has previously been shown to mediate these phenomena [5–7]. Understanding the molecular mechanisms and pathways by which parasite-proteins such as *Pf*EMP1, are trafficked to the cytosol and thence to the surface of the iRBC is, therefore, critical for a clear insight into the pathogenesis of *P. falciparum* malaria.

RBCs lack a protein secretory apparatus so the parasite must establish de novo secretion machinery within the host cell cytoplasm in order to transport its own proteins into the iRBC across the PVM. The mechanisms that enable such trafficking are incompletely understood; however, many *P. falciparum* proteins destined for export into the iRBC contain both an N-terminal hydrophobic signal sequence and a short conserved pentameric host cell-targeting motif (RxLxE/Q/D) termed the *Plasmodium* export element (PEXEL) or the vacuolar transport signal (VTS) [8,9]. The N-terminal signal sequence is required for the proteins to enter the constitutive secretory pathway *via* the endoplasmic reticulum (ER) [10,11], where the PEXEL/VTS motif is cleaved by plasmepsin V, an ER residing aspartic protease [12,13], and the newly formed N-terminus (xE/Q/D) allows translocation into the iRBC cytosol by a PVM residing translocon called the "*Plasmodium* translocon of exported proteins" (PTEX) complex [14]. Our understanding of the mechanisms behind the transport of proteins within the iRBC remains vague, but some tentative explanations

Abbreviations: aa, amino acid(s); CRD, cysteine-rich domain; ER, endoplasmic reticulum; GFP, green fluorescence protein; IFA, indirect immunofluorescence assay; iRBC, infected red blood cell; PBS, phosphate buffered saline; PEXEL, *Plasmodium* export element; PNEP, PEXEL negative exported protein; PVM, parasitophorous vacuole membrane; TM, transmembrane; Tx, Triton-X 100; Var, variable region; WR, tryptophan-rich.

* Corresponding author. Tel.: +81 95 819 7838; fax: +81 95 819 7805.
E-mail address: okaneko@nagasaki-u.ac.jp (O. Kaneko).

1383-5769/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.parint.2011.05.003