repeat array (MS20) showed low levels of variation in the South Korean population. Also in the same population, MS8 and MS12 showed low levels of variation, although the loci were not short repeat arrays. The genetic diversities of the loci from Sri Lankan and Brazilian populations (Table 2) were higher than those from the South Korean population. Therefore, the uniqueness of the diversity would not be solely dependent on the characteristics of the loci. In fact, mutations of microsatellite loci are generally considered to be neutral. However, if the loci are in a certain gene or close to a certain gene, the mutation may not be strictly neutral. Indeed, 6 of the 10 loci examined in this study (MS4, MS5, MS8, MS9, MS15, MS20) were in a gene coding a hypothetical protein or a known protein (Table 2). One of the 10 loci (MS7) was between a gene coding a hypothetical protein and a gene coding a merozoite surface protein-7 which is expected to be under strong selective pressure. Therefore, the mutation of those 7 loci may not be strictly neutral. The allelic data suggested that the frequencies of strand-slippage events of the microsatellite loci during mitotic replication in the South Korean *P. vivax* population were very low because identical alleles in the known loci have been found for 10 years or longer in this population.

Multiple genotype infection (MGI) is one of the important indexes of population genetics and epidemiology of malaria parasites because MGI is the first step in recombination of the parasite genome between different clones. In the case of *P. falciparum*, the rate of MGIs per population is basically associated with the endemicity [4, 5]. That is, the MGI rate of *P. falciparum* population is higher in high transmission areas and lower in low transmission areas. However, this is not the case with *P. vivax* populations because high MGI rates were observed among the *P. vivax* populations in low transmission areas [11, 12]. This feature could be attributed to relapse owing to hypnozoites in the liver of a vivax malaria patient. Although MGI is an important index, the methods or criteria of determining MGI is problematic. When any locus of the 10 loci showed more than 1 allele, we regarded the isolate as an example of MGI. Using this method, 85 (97.7%) out of the 87 isolates were MGIs. Focusing on each locus, the MGI rate per locus varied from 0.0% to 83.9% (average 29.1%) (Table 1). Focusing on the number of MGI loci per isolate, we found an interesting distribution pattern, similar to an F-distribution (Fig. 2). In the present study, the highest frequency of MGI loci per isolate was 2 (found in 25 isolates). The frequency decreased gradually, that is, 3 MGI loci; 19 isolates, 4 MGI loci; 12 isolates, 5 MGI loci; 6 isolates, and so on. We suspect that this distribution pattern may vary in each endemic area with different endemicity.

In the case of *P. falciparum* populations, the levels of genetic diversity are normally associated with the levels of malaria endemicity. That is, the levels of genetic diversity of the parasite populations are higher in high transmission areas and lower in low transmission areas [4,

5], although some exceptions have been reported [36].

We suspect that there will also be some association between the levels of genetic diversity and the levels of malaria endemicity in *P. vivax* populations, even though the correlation between these factors is not clearly understood at the time of writing. In *P. vivax* populations, the levels of genetic diversity tend to be high even in low transmission areas [11-13, 28]. This tendency can likely be attributed to unique biological features of *P. vivax*, such as early gametocytogenesis and relapse. Early gametocytogenesis may enhance the efficiency of transmission to *Anopheles* mosquitoes, allowing transmission to occur before symptoms appear – or, more importantly, before antimalarial drugs are administered. Relapses may also enhance the transmission and increase the genetic diversity of *P. vivax* populations, because the relapse will increase the probability of the coexistence of multiple genotype clones in a single patient, which are subsequently sucked up by an *Anopheles* mosquito in a single meal. Thus, the levels of genetic diversity of *P. vivax* populations could be higher than those observed in *P. falciparum* populations even in low transmission areas.

In the present study, the levels of genetic diversity of the South Korean population between 1994 and 2000 (when the number of malaria cases increased) were relatively lower than the levels of genetic diversity between 2001 and 2008 (when the number decreased). On the contrary, the levels of multilocus LD in the population between 1994 and 2000 were relatively higher than those between 2001 and 2008. These results suggested that the latter population was more genetically diverse and had less inbreeding. Furthermore, we divided the population into 3 groups each covering 5-year periods (1994-1998, 1999-2003 and 2004-2008) and reexamined the levels of genetic diversity and multilocus LD. Then, we again observed that the levels of genetic diversity in the populations had gradually increased, whereas the levels of multilocus LD had gradually decreased even though there was still strong multilocus LD in the most recent population (2004-2008). This result was surprising to us because we expected that the effective population size of the latter population would have decreased due to the reduction in the number of alleles in the population. However, this was not the case. In the South Korean populations, the association between the diversity and the endemicity of the *P. vivax* population is elusive.

There are, however, at least two possible explanations for this result. One is that the levels of genetic diversity of the *P. vivax* population increased in North Korea from 2001 to 2002, while the number of vivax malaria cases was very high (296,540 cases in 2001, 241,190 cases in 2002) [1]. Some of the isolates might have then been introduced to South Korea from North Korea by *Anopheles* mosquitoes. The other possible explanation is that the genetic diversity began accumulating in the South Korean population after the re-emergence in 1993. If the latter hypothesis is correct

then the malaria control program conducted by the South Korean government might not have affected the parasite population structure.

One of the clear differences between the *P. vivax* population in South Korea and populations in tropical and subtropical areas is the pattern of transmission: in South Korea, vivax malaria is seasonally prevalent with a peak during July and August and no transmission in the winter season [37] and very long incubation periods with 8 to 13 months [14], suggesting that the chance for the recombination of the genome is limited to specific time periods within the year, possibly once or at most twice a year. In fact, we found strong LD in the South Korean population, suggesting that the frequency of recombination in this population would be very limited. However, these results might be associated with the location of the examined MS DNA loci: the 6 loci are in a gene coding a protein and another locus is between 2 genes coding respective proteins. In tropical and subtropical areas, on the other hand, vivax malaria is prevalent throughout the year, and thus recombination may occur throughout the year; this would lead to an increase in the levels of genetic diversity in tropical and subtropical areas. Indeed, in the populations from the Brazilian Amazon, identical haplotypes were rarely observed two years in a row, even in the same endemic area [12]. This would suggest that frequent recombinations occurred between the clones in the population.

The present study showed evidence of a low recombination rate and low frequencies of strand-slippage events of the microsatellite loci during mitotic replication in the *P. vivax* population of South Korea in comparison to populations in tropical and subtropical areas [12], and demonstrated that the 2 dominant haplotypes (H16 and H25) had been transmitting for several years (H16; 1996, 1998-2001, 2005, H25; 1995-2001, 2003) (Table 5). This continuous existence of the same haplotypes for several years is definitive evidence of a low recombination rate in the South Korean *P. vivax* population.

This continuous existence of the same haplotypes could be explained by a local adaptation to vector species. According to Joy et al. [38], for example, *P. vivax* in southern Mexico was genetically differentiated into 3 populations. They suggested that this differentiation would be the result of adaptation to different *Anopheles* species. On the other hand, in South Korea, *Anopheles sinensis* is a main vector of *P. vivax* and the other *Anopheles* species are very minor. Therefore, continuous existence of the predominant haplotypes could not be explained by a local adaptation to certain vector species in this country. There might be some other advantages of these haplotypes, or simply, the variation in the *P. vivax* population on the Korean peninsula had been very small owing to an effective national eradication program conducted by the National Malaria Eradication Service under the operation of the South Korean government with the support of the WHO in the 1970s [14-16].

Although the predominant haplotypes (H16 and H25) and their relatives had been transmitting in the DMZ for a long time, their transmission ended in 2005. We speculate that these predominant haplotypes were probably eliminated by the malaria control programs conducted by the North Korean government. In fact, according to the WHO World Malaria Report 2010, the number of vivax malaria cases in North Korea decreased substantially (2001: 296540 cases, 2005: 11507 cases). The reason for this reduction was not mentioned in detail, however this is probably due to the effect of mass drug administration by the North Korean government supported by South Korea. We suspect that the population structure of *P. vivax* in North Korea was changed dramatically and that these predominant "old" haplotypes were eliminated completely or became very minor in both the North Korean and South Korean populations. We suspect that the South Korean *P. vivax* population is a subpopulation of the North.

Our previous genetic epidemiological analyses of the South Korean *P. vivax* population using antigenic molecules [21-25] and the mitochondrial genome [39] showed that there were 2 types (or groups) of parasite populations in the endemic area. In these previous studies we examined groups of isolates collected from vivax malaria patients in the DMZ in 1997 [21], 1998 [22-24], 1999 [39]. In the present study, we examined isolates collected from patients in the DMZ between 1994 and 2008 using 10 highly polymorphic microsatellite loci. Once again, we observed two types of parasite populations (Fig. 4). However, some other haplotypes (clones) have been observed in the endemic area since 1998. The new haplotypes were genetically different to the 2 major groups that have been transmitted since the beginning of the re-emergence (Fig. 4). This finding was consistent with the results of analyses by Choi et al. using the DNA sequences of 2 antigenic molecules (circumsporozoite protein, merozoite surface protein-1) of isolates collected in the DMZ from 1996 to 2007 [40]. They also reported that new genotypes have been observed since 2000 and that the new genotypes had been rapidly disseminated in the endemic area.

The genetic differences between the 2 major groups and the new haplotypes in our data suggested two possibilities: the new haplotypes could have arisen in the DMZ in South Korea through recombination between existing clones in the population; or their emergence could be attributed to a continual introduction of *P. vivax* from other population sources, probably from North Korea. The present study suggested a low recombination rate in the South Korean population and would seem to indicate that the latter possibility is more likely.

A less likely possibility is that all of the isolates examined in this study were continually introduced from North Korea because all the isolates were collected from South Korean soldiers who served in the DMZ. These patients were normally

treated by chloroquine within 4 days of the onset of the symptoms, and then treated by primaquine as a radical cure. The recurrence rate (both new infection and relapse may be included) of vivax malaria among them is 1.6% (62 cases of 3881 cases) and the definitive relapse rate is only 0.2% (8 cases of 3881 cases) [41]. In addition, the incubation period of *P. vivax* on the Korean peninsula is very long (8 months to 13 months) [14] and the transmission is mainly in summer [38]. Moreover, the period of conscription is about 2 years. Therefore, it might have been very difficult to transmit continuously among the South Korean soldiers in the DMZ, leading to the high recombination rate of the genome within the parasite population of the study area.

There are a number of sampling limitations to the present study. The number of isolates per year was relatively small (2 to 14 isolates, average: 5.8 isolates/year), and the sample size during the years 2004-2008 was particularly small (2 to 7 isolates, average: 3.6 isolates/year). Moreover, all of the isolates used in this study were collected only from South Korean soldiers or veterans and not from civilians, whose proportion among vivax malaria patients in South Korea has been gradually increasing [18]. In order to overcome these limitations and more accurately estimate the current status of the parasite population in South Korea, it will be necessary to include new isolates collected from civilians in the endemic areas and to increase the sample size of recent years.

Although travel between South and North Korea is basically restricted and the malaria control programs in the two countries may not be the same, we suspect that the South Korean *P. vivax* population is a subpopulation of the North Korean population because the majority of malaria patients live near the border [42]. *Anopheles* mosquitoes can fly over the DMZ, and South Korean travelers are allowed to visit some parts of North Korea, such as Kaesong and Kumgang-san, which are very famous for sightseeing. Furthermore, from 2001 to 2009 the number of vivax malaria cases in North Korea ranged from twice as high as the number in South Korea to many times higher, indicating that the size of the parasite population in North Korea is probably larger. Thus, the inclusion of North Korean isolates in the analyses would greatly enhance the accuracy of the estimation of the parasite population structure and the transmission dynamics and provide a more complete picture of the *P. vivax* population in the Korean peninsula; unfortunately the feasibility of doing this is low.

In conclusion, molecular epidemiology using highly polymorphic DNA markers of the *P. vivax* population is a very useful tool for assessing the population structure and transmission dynamics of the parasites, the knowledge of which may lead to the effective control of vivax malaria in the respective endemic areas.

Funding: This work was financially supported by a Grant-in-Aid for Scientific Research (B) (19406013) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, the Grant for National Center for Global Health and Medicine (21A107), and a Grant for Global Health Promotion (H22-KUNII-SHITEI-004) from the Ministry of Health, Labour and Welfare of Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions

Conceived and designed the experiments: MI MF WGK SK. Performed the experiments: MI MF SYH SHK. Analyzed the data: MI MF. Contributed regents, materials, and analysis tools: WGK. Wrote the paper: MI MF WGK SK. All authors gave their final approval of this version to be published for *PLoS Neglected Tropical Diseases*.

References

1. World Health Organization (2010) World Malaria Report 2010. Geneva: WHO.
2. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl Trop Dis 4: e774.
3. Enserink M (2010) As challenges change, so does science. Science 328: 843.
4. Iwagami M, Rivera PT, Villacorte EA, Escueta AD, Hatabu T, et al. (2009) Genetic diversity and population structure of *Plasmodium falciparum* in the Philippines. Malar J 8: 96.
5. 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.
6. Machado RL, Povoa MM, Calvosa VS, Ferreira MU, Rossit AR, et al. (2004) Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. J Infect Dis 190: 1547-1555.
7. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, et al. (2005) Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. J Infect Dis 191: 1558-1564.
8. Mu J, Joy DA, Duan J, Huang Y, Carlton J, et al. (2005) Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Mol Biol Evol 22: 1686-1693.
9. Jongwutiwes S, Putaporntip C, Iwasaki T, Ferreira MU, Kanbara H, et al. (2005) Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. Mol Biol Evol 22: 1733-1739.
10. Cornejo OE, Escalante AA (2006) The origin and age of *Plasmodium vivax*. Trends Parasitol 22: 558-563.
11. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, et al. (2007) Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. Int J Parasitol 37: 1013-1022.
12. Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, et al. (2007) Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. J Infect Dis 195: 1218-1226.
13. Gunawardena S, Karunaweera ND, Ferreira MU, Phone-Kyaw M, Pollack RJ, et al. (2010) Geographic structure of *Plasmodium vivax*: microsatellite analysis of parasite populations from Sri Lanka, Myanmar, and Ethiopia. Am J Trop Med Hyg 82: 235-242.
14. Chai JY (1999) Re-emerging *Plasmodium vivax* malaria in the Republic of Korea. Korean J Parasitol 37: 129-143.
15. Ree HI (2000) Unstable vivax malaria in Korea. Korean J Parasitol 38: 119-138.
16. Shin EH, Guk SM, Kim HJ, Lee SH, Chai JY

(2008) Trends in parasitic diseases in the Republic of Korea. Trends Parasitol 24: 143-150.

17. Chai IH, Lim GI, Yoon SN, Oh WI, Kim SJ, et al. (1994) [Occurrence of tertian malaria in a male patient who has never been abroad]. Korean J Parasitol 32: 195-200 [Article in Korean, English abstract available].
18. Park JW, Jun G, Yeom JS (2009) *Plasmodium vivax* malaria: status in the Republic of Korea following reemergence. Korean J Parasitol 47 (Suppl): S39-50.
19. Han ET, Lee DH, Park KD, Seok WS, Kim YS, et al. (2006) Reemerging vivax malaria: changing patterns of annual incidence and control programs in the Republic of Korea. Korean J Parasitol 44: 285-294.
20. Feighner BH, Pak SI, Novakoski WL, Kelsey LL, Strickman D (1998) Reemergence of *Plasmodium vivax* malaria in the Republic of Korea. Emerg Infect Dis 4: 295-297.
21. Kho WG, Park YH, Chung JY, Kim JP, Hong ST, et al. (1999) Two new genotypes of *Plasmodium vivax* circumsporozoite protein found in the Republic of Korea. Korean J Parasitol 37: 265-270.
22. Kho WG, Chung JY, Sim EJ, Kim DW, Chung WC (2001) Analysis of polymorphic regions of *Plasmodium vivax* Duffy binding protein of Korean isolates. Korean J Parasitol 39: 143-150.
23. Chung JY, Chun EH, Chun JH, Kho WG (2003) Analysis of the *Plasmodium vivax* apical membrane antigen-1 gene from re-emerging Korean isolates. Parasitol Res 90: 325-329.

24 Kim SH, Hwang SY, Shin JH, Moon CS, Kim DW et al. (2009) Molecular genetic characterization of the merozoite surface protein 1 Gene of *Plasmodium vivax* from reemerging Korean isolates. Clin Vaccine Immunol 16: 733-738.

25. Hwang SY, Kim SH, Kho WG (2009) Genetic characteristics of polymorphic antigenic markers among Korean isolates of *Plasmodium vivax*. Korean J Parasitol 47 (Suppl): S51-58.
26. Hartl DL (2004) The origin of malaria: mixed messages from genetic diversity. Nat Rev Microbiol 2: 15-22.
27. Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual. 3rd edition. New York: Cold Spring Harbor Laboratory Press. pp. 6.4-6.12.
28. Karunaweera ND, Ferreira MU, Hartl DL, Wirth DF (2007) Fourteen polymorphic microsatellite DNA markers for the human malaria parasite *Plasmodium vivax*. Mol Ecol Notes 7: 172-175.
29. Maynard-Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci USA 90: 4384-4388.
30. Hudson RR (1994) Analytical results concerning linkage disequilibrium in models with genetic transformation and recombination. J Evol Biol 7: 535-548.
31. Haubold B, Hudson RR (2000) LIAN 3.0:

detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847-848.

32. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186: 1518-30.
33. Orjuela-Sánchez P, da Silva NS, da Silva-Nunes M, Ferreira MU (2009) Recurrent parasitemias and population dynamics of *Plasmodium vivax* polymorphisms in rural Amazonia. Am J Trop Med Hyg 81: 961-968.
34. Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Day NPJ, et al. (2006) Microsatellite variation, repeat array length and population history of *Plasmodium vivax*. Mol Biol Evol 23: 1016-1018.
35. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435-445.
36. Branch OH, Sutton PL, Castro JC, Barnes C, Hussin J, et al. (2011) *Plasmodium falciparum* genetic diversity maintained and amplified over 5 years of a low transmission endemic in the Peruvian Amazon. Mol Biol Evol 28: 1973-1986.
37. Park JW, Klein TA, Lee HC, Pacha LA, Ryu SH, et al. (2003) Vivax malaria: a continuing health threat to the Republic of Korea. Am J Trop Med Hyg 69: 159-167.
38. Joy DA, Gonzalez-Ceron L, Carlton JM, Gueye A, Fay M, et al. (2008) Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. Mol Biol Evol 25: 1245-1252.
39. Iwagami M, Hwang SY, Fukumoto M, Hayakawa T, Tanabe K, et al. (2010) Geographical origin of *Plasmodium vivax* in the Republic of Korea: haplotype network analysis based on the parasite's mitochondrial genome. Malar J 9: 184.
40. Choi YK, Choi KM, Park MH, Lee EG, Kim YJ, et al. (2010) Rapid dissemination of newly introduced *Plasmodium vivax* genotypes in South Korea. Am J Trop Med Hyg 82: 426-432.
41. Moon KT, Kim YK, Ko DH, Park I, Shin DC, et al. (2009) Recurrence rate of vivax malaria in the Republic of Korea. Trans R Soc Trop Med Hyg 103: 1245-1249.
42. Roll Back Malaria, World Health Organization, UNICEF (2005) Democratic People's Republic of Korea, Country Profile, World Malaria Report 2005. Geneva: WHO and UNICEF. Available: http://www.rollbackmalaria.org/wmr2005/profiles/dprkorea.pdf Accessed: September 14, 2011.

Figure legends

Figure 1. Genetic diversity of the *P. vivax* population in South Korea

A: Average number of the alleles ± SE, H_E: Average expected heterozygosity ± SE.

Numbers and numbers in parentheses represent number of haplotypes and isolates observed each year, respectively. The graph was made based on numbers of reported vivax malaria cases in South Korea. n: represents the number of isolates. The data were obtained from the *World Malaria Report 2010* (WHO) [1].

Figure 2. Frequency of MGI loci per isolate

Zero in the x-axis indicates that no MGI loci were observed in a particular isolate, that is, it represents a single clone infection isolate.

Figure 3. Genetic diversity of the *P. vivax* population in South Korea

A: Average number of the alleles ± SE, H_E: Average expected heterozygosity ± SE.

Numbers and numbers in parentheses represent the number of haplotypes and isolates observed each year, respectively. The graph was made based on numbers of reported vivax malaria cases in South Korea. n: represents the number of isolates. The data were obtained from the *World Malaria Report 2010* (WHO) [1].

Figure 4. Relationships among the 40 haplotypes of *P. vivax* (n=87) in South Korea estimated by eBURST analysis

Relationships among 40 microsatellite haplotypes in the 87 isolates collected in South Korea as defined by eBURST analysis [32]. H1, H2, ... H40 represent the microsatellite haplotype. Red numbers represent the number of isolates that showed the haplotype.

When the haplotype was found in only one isolate, the red number was omitted. Numbers in parentheses represent the year the haplotype was observed.

H6 and H7 were included in the analysis but were classified into a different group because they were considerably different from others and were omitted from Figure 4.

Table 1. Multiple Genotype Infection rate per locus of the *P. vivax* population in South Korea.

Locus	No. of MGI isolates	MGI rate (%)
MS1	9	10.3
MS4	52	59.8
MS5	6	6.9
MS6	3	3.4
MS7	73	83.9
MS8	35	40.2
MS9	10	11.5
MS12	38	43.7
MS15	0	0.0
MS20	27	31.0
Average	25.3	29.1

MGI: Multiple genotype infection.
Sample size: 87 isolates.

Table 2. Genetic diversity of *P. vivax* populations in South Korea, Sri Lanka and the Brazilian Amazon.

Locus	Chr (ID)	Core repeat sequence in the Salvador-I strain	Type of region	South Korea (n=87)		Sri Lanka (n=25)		Brazil (n=99)	
				A	H_E	A	H_E	A	H_E
MS1	3 (CM000444)	(GAA)11	Repeat region	5	0.61	7	0.77	5	0.69
MS4	6 (CM000447)	(AGT)18	Gene coding a hypothetical protein	3	0.55	6	0.75	6	0.69
MS5	6 (CM000447)	CCTCTT(CCT)11	Gene coding a hypothetical protein	6	0.62	6	0.81	16	0.87
MS6	11 (CM000452)	$(TCC)_2(TCT)_3(CCT)_2(TCC)_2$ $GCTTCT(TCC)_{10}$	Repeat region	7	0.66	6	0.83	8	0.66
MS7	12 (CM000453)	(GAA)9	Between two genes*	6	0.58	6	0.65	3	0.48
MS8	12 (CM000453)	$(CAG)_2(CAA)_{11}$	Gene coding a 3'-5' exonuclease domein containing protein	3	0.05	11	0.84	16	0.84
MS9	8 (CM000449)	(GGA)18	Gene coding a hypothetical protein	3	0.57	6	0.80	7	0.78
MS12	5 (CM000446)	(TTC)10(TGC)4	Repeat region	3	0.07	6	0.81	8	0.74
MS15	5 (CM000446)	$(TCT)_{10}$	Gene coding a CW-type zinc finger domain containing protein	5	0.54	8	0.86	16	0.74
MS20	10 (CM000451)	$(GAA)_{11}GAG(GAA)_{13}(CAA)_4$ $GAA(CAA)_5$	Gene coding a tryptophan-rich antigen	2	0.12	13	0.91	11	0.82
Average				4.3	0.43	7.5	0.80	9.6	0.73

Chr: Chromosome No, ID represents Chromosome ID of *P. vivax* Salvador 1 in the GenBank Database, A: Number of alleles, H_E: Expected heterozygosity, Data of Sri Lankan and Brazilian populations were obtained from Karunaweera, et al. (2007) [28] and Orjuela-Sánchez, et al. (2009) [33], respectively.
n: Number of isolates. * Two genes are encoding a hypothetical protein and a merozite surface protein-7.

Table 3. Multilocus linkage disequilibrium in the two *P. vivax* populations.

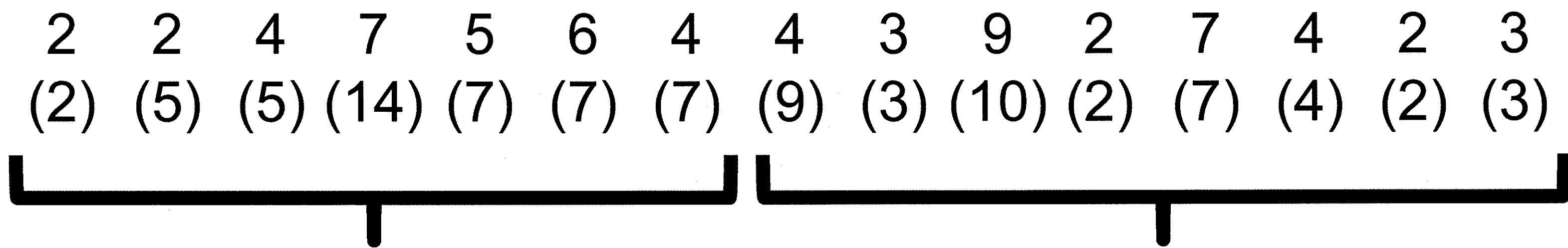
Population	Single Clones		Unique Haplotypes Only	
	No.	I_A^S	No.	I_A^S
1994-2000	47	0.529***	19	0.346***
2001-2008	40	0.218***	27	0.173***

Single clones show all haplotypes in single-clone infections. Unique haplotypes show haplotypes excluding duplicates of any multiply represented infection. No. indicates the number of isolates for each measure. *** $P < 0.001$

Table 4. Multilocus linkage disequilibrium in the three *P. vivax* populations.

Population	Single Clones		Unique Haplotypes Only	
	No.	I_A^S	No.	I_A^S
1994-1998	33	0.584***	13	0.408***
1999-2003	36	0.315***	17	0.231***
2004-2008	18	0.140***	18	0.153***

Single clones show all haplotypes in single-clone infections. Unique haplotypes show haplotypes excluding duplicates of any multiply represented infection. No. indicates the number of isolates for each measure. *** $P < 0.001$



$A = 2.70 \pm 0.26$
$H_E = 0.36 \pm 0.06$
1994-2000 (n=47)

$A = 3.80 \pm 0.57$
$H_E = 0.50 \pm 0.10$
2001-2008 (n=40)

Figure 1

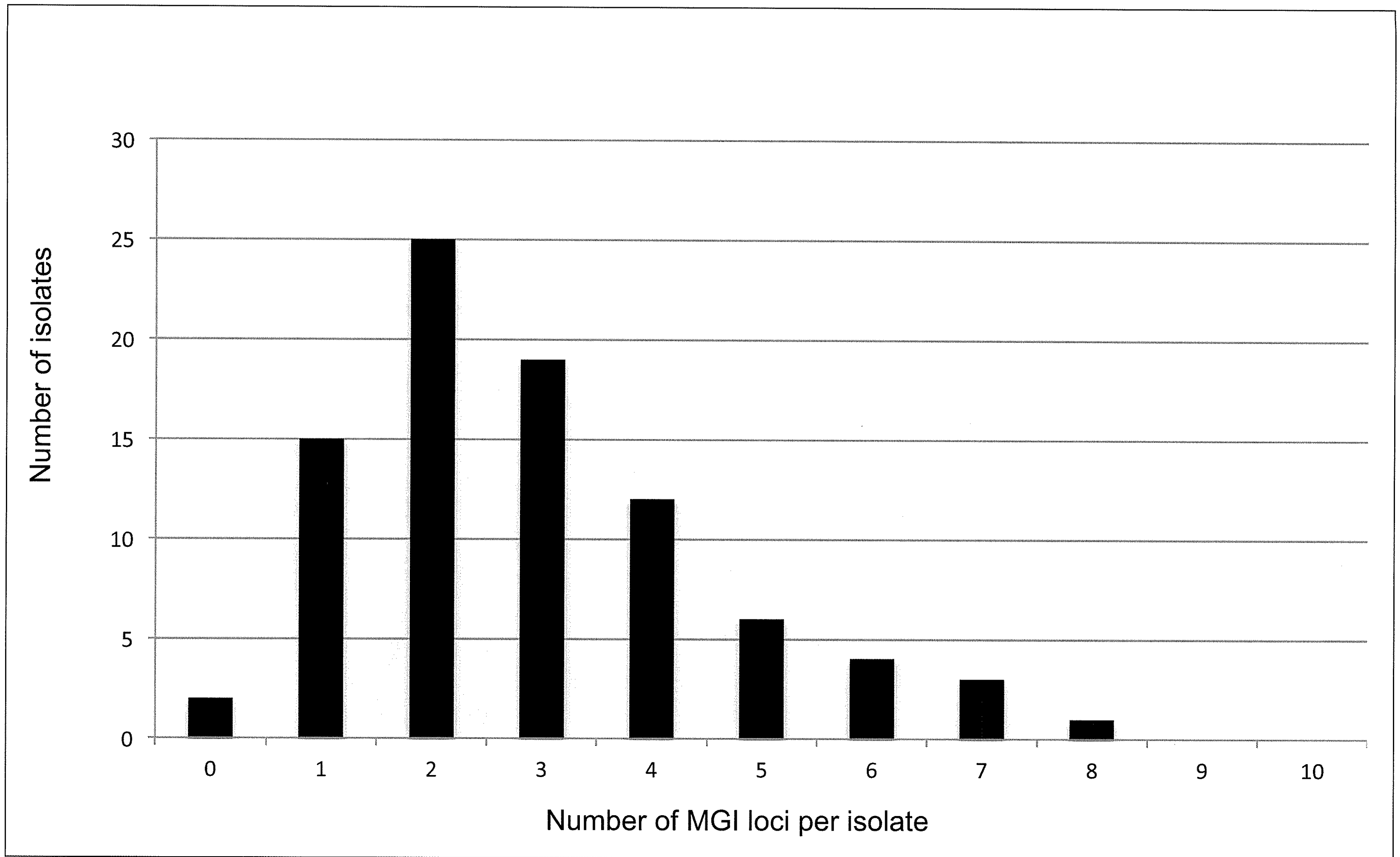
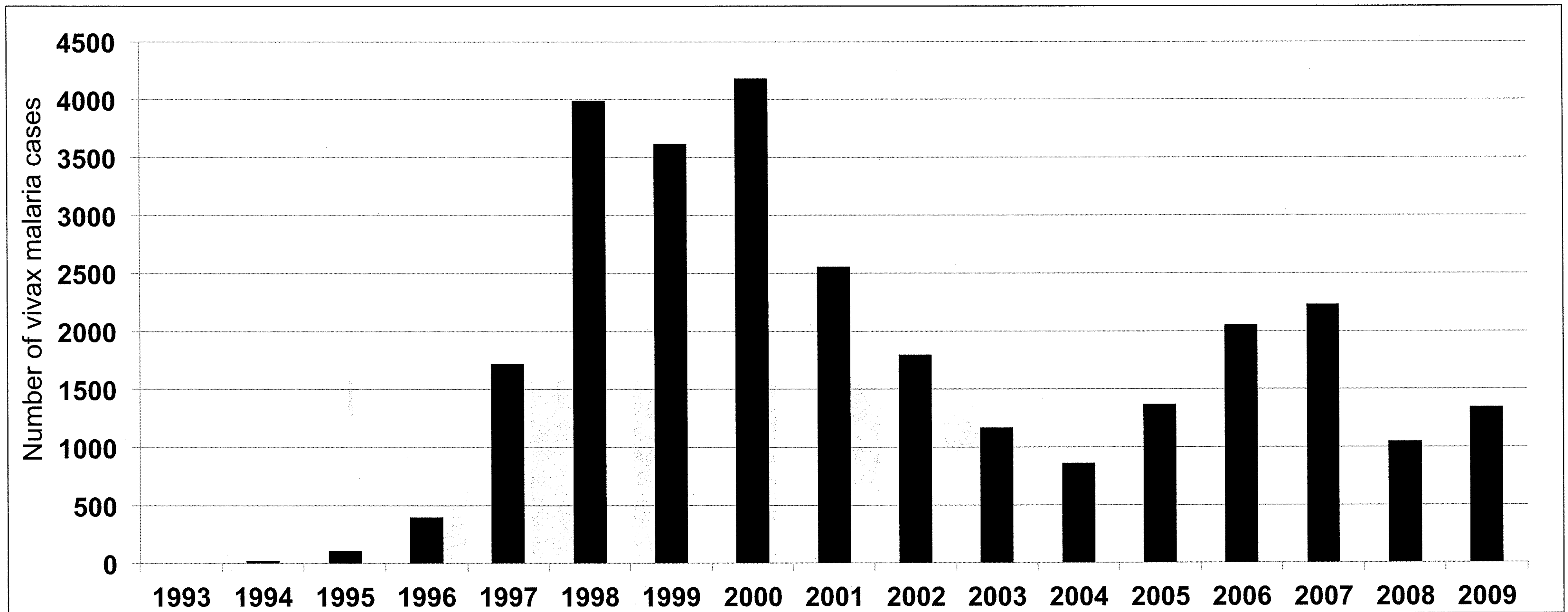


Figure 2



2 2 4 7 5 6 4 4 3 9 2 7 4 2 3

(2) (5) (5) (14) (7) (7) (7) (9) (3) (10) (2) (7) (4) (2) (3)

$A = 2.50 \pm 0.27$
$H_E = 0.31 \pm 0.05$
1994-1998 (n=33)

$A = 3.00 \pm 0.42$
$H_E = 0.42 \pm 0.09$
1999-2003 (n=36)

$A = 3.80 \pm 0.57$
$H_E = 0.56 \pm 0.10$
2004-2008 (n=18)

Figure 3

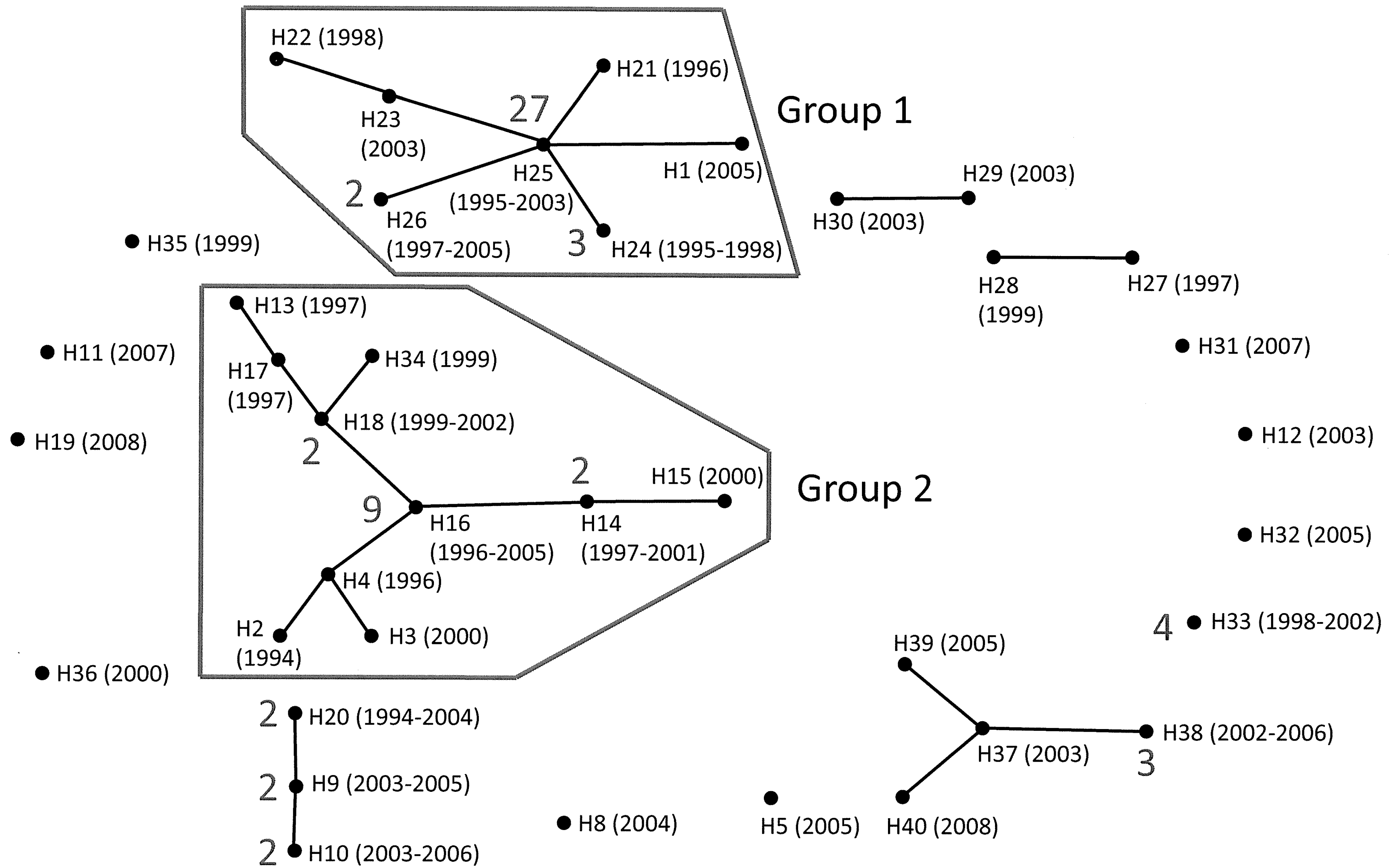


Figure 4

Rui *et al. Malaria Journal* 2011, **10**:192
http://www.malariajournal.com/content/10/1/192

MALARIA JOURNAL

METHODOLOGY Open Access

Plasmodium vivax: comparison of immunogenicity among proteins expressed in the cell-free systems of *Escherichia coli* and wheat germ by suspension array assays

Edmilson Rui[1], Carmen Fernandez-Becerra[1], Satoru Takeo[2], Sergi Sanz[1], Marcus VG Lacerda[3], Takafumi Tsuboi[2,4] and Hernando A del Portillo[1,5*]

Abstract

Background: In vitro cell-free systems for protein expression with extracts from prokaryotic (*Escherichia coli*) or eukaryotic (wheat germ) cells coupled to solid matrices have offered a valid approach for antigen discovery in malaria research. However, no comparative analysis of both systems is presently available nor the usage of suspension array technologies, which offer nearly solution phase kinetics.

Methods: Five *Plasmodium vivax* antigens representing leading vaccine candidates were expressed in the *E. coli* and wheat germ cell-free systems at a 50 μl scale. Products were affinity purified in a single-step and coupled to luminex beads to measure antibody reactivity of human immune sera.

Results: Both systems readily produced detectable proteins; proteins produced in wheat germ, however, were mostly soluble and intact as opposed to proteins produced in *E. coli*, which remained mostly insoluble and highly degraded. Noticeably, wheat germ proteins were recognized in significantly higher numbers by sera of *P. vivax* patients than identical proteins produced in *E. coli*.

Conclusions: The wheat germ cell-free system offers the possibility of expressing soluble *P. vivax* proteins in a small-scale for antigen discovery and immuno-epidemiological studies using suspension array technology.

Background

The recent call for malaria eradication has re-emphasized the importance of bringing *Plasmodium vivax* into the research agenda [1]. *Plasmodium vivax* remains the most widely distributed human malaria parasite with 2.85 billion people living at risk of infection [2]. Noticeably, the number of yearly clinical cases seems to be increasing from 70-80 million [3] to 300 million cases [4] and these include cases of severe disease and death exclusively associated with *P. vivax* [5,6]. Moreover, experts agree that present tools against *Plasmodium falciparum* will not be effective against *P. vivax*, reinforcing the development of control measurements for this species [7]. Among these tools, vaccines continue to represent the most cost-effective control measurement but unfortunately vaccine development in *P. vivax* lags well behind that of *P. falciparum* [8].

The genomes of human malaria parasites encode approximately 5,400 coding genes opening an avenue for antigen discovery in this species [9]. Unfortunately, cell-based expression systems have met limited success to obtain soluble proteins largely attributed to the high AT-content, the existence of long stretches of repeated amino acid sequences and much larger proteins than their homologues in other eukaryotes [10]. In contrast to cell-based systems, cell-free expression systems for protein synthesis with extracts from prokaryotic or eukaryotic cells has offered a valid alternative to express soluble proteins [11]. In the case of malaria, using the *Escherichia coli* cell-free system, Doolan and co-workers

* Correspondence: hernandoa.delportillo@cresib.cat
[1]Barcelona Centre for International Health Research (CRESIB), Hospital Clinic/IDIBAPS, Universitat de Barcelona, Roselló 153, 1a planta, 08036, Barcelona, Spain
Full list of author information is available at the end of the article

BioMed Central

© 2011 Rui et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

first reported on the expression of 250 *P. falciparum* proteins subsequently coupled to solid arrays and analysed with immune sera discovering putative new antigens [12]. Using this same approach, expression of 1,204 *P. falciparum* proteins later expanded these analysis and predicted new antigens [13]. Parallel efforts were reported on the use of cell-free extracts from wheat germ to similarly produce hundreds of *P. falciparum* proteins [14,15]. More recently, the wheat germ expression system has been used for antigen discovery in *P. vivax* [16]. Thus, 89 different soluble proteins were expressed and shown to be immunogenic on analyses of protein arrays and immune sera. Together, this data demonstrates that cell-free expression systems coupled to protein arrays offer a scalable platform for antigen discovery in malaria.

Suspension array technologies with high-throughput capacity to simultaneously analyse several proteins with minimal amount of immune sera have also been developed and used in analysis of multiple malaria vaccine candidates as well as in developing functional assays [17-20]. Suspension arrays offer several advantages as compared to flat protein arrays including nearly solution phase kinetics and total assay sensitivity [21]. The aim of this study was to develop a small-scale method for soluble expression of *P. vivax* proteins using the *E. coli* and wheat germ cell-free systems and to compare their usage by multiplexing assays.

Methods

Human samples

Human plasma samples were obtained from endemic areas of Brazil and from a non-endemic region. The first group comprised immune sera from adults living in the Brazilian Amazon [22]. The other group comprised sera from four healthy adult volunteers living in the city of Barcelona (Spain) that have never been exposed to malaria or visited malaria endemic regions. These studies received the ethical approval of Local Institutional Reviewing Boards.

Construction of plasmids

Plasmid pIVEX1.4d for expression in wheat germ and pIVEX2.4d for expression in *E. coli* were purchased from Roche and modified by inserting GST after the 6xHis tag sequence. Modified plasmids were termed pIVEXGST1.4d and pIVEXGST2.4d (Figure 1). Both vectors carry the same T7-DNA promoter elements, the ampicillin selectable marker and identical His-GST tags in the same positions. The following proteins were engineered into these vectors: PvMSP1-19 (1590-1699 aa, id PVX_099980) and PvMSP1-Nter (170-675 aa, id PVX_099980); PvDBP-RII (196-521 aa, id PVX_110810); PvCSP-S (51-319 aa, id PVX_119355); PvMSP5 (full length, id PVX_003770); PvMSP7 (full length, id PVX_082695 (Figure 2). Further information on these proteins and primers used for amplifications can be obtained as supplementary information (Additional file 1).

The circumsporozoite antigen of *P. vivax* is dimorphic based on the central repeat region and the two alleles, VK210- and VK247-type, share no immunological cross-reactivity [23]. Therefore, a recombinant chimeric PvCSP protein containing VK210-(PVX_119355) and VK247-type (GenBank#M69059, *P. vivax* PNG strain) amino acid repeat sequences (PvCSP-c) which may cover the vivax parasite population globally was developed (Figure 3). The PvCSP-c was constructed and expressed in a large-scale wheat germ cell-free system

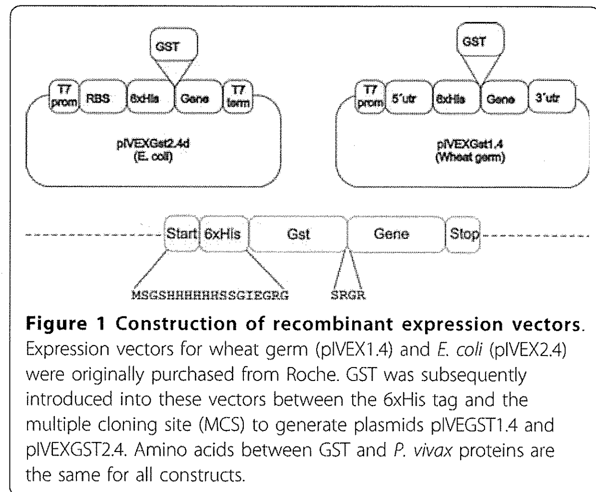


Figure 1 Construction of recombinant expression vectors. Expression vectors for wheat germ (pIVEX1.4) and *E. coli* (pIVEX2.4) were originally purchased from Roche. GST was subsequently introduced into these vectors between the 6xHis tag and the multiple cloning site (MCS) to generate plasmids pIVEGST1.4 and pIVEXGST2.4. Amino acids between GST and *P. vivax* proteins are the same for all constructs.

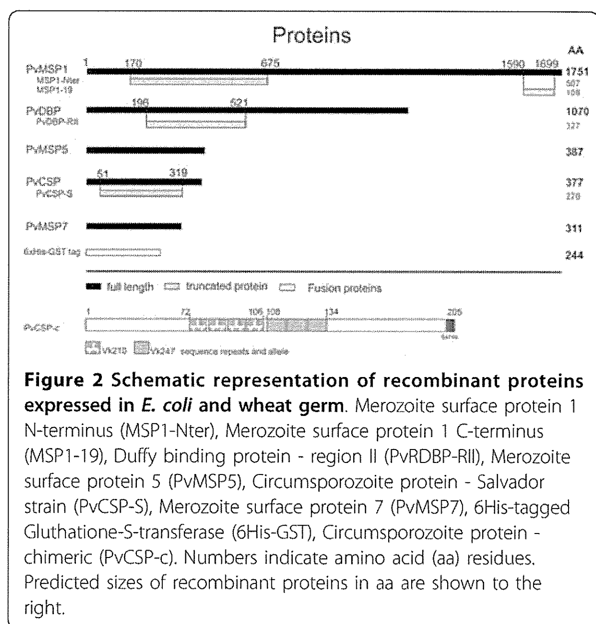


Figure 2 Schematic representation of recombinant proteins expressed in *E. coli* and wheat germ. Merozoite surface protein 1 N-terminus (MSP1-Nter), Merozoite surface protein 1 C-terminus (MSP1-19), Duffy binding protein - region II (PvRDBP-RII), Merozoite surface protein 5 (PvMSP5), Circumsporozoite protein - Salvador strain (PvCSP-S), Merozoite surface protein 7 (PvMSP7), 6His-tagged Gluthatione-S-transferase (6His-GST), Circumsporozoite protein - chimeric (PvCSP-c). Numbers indicate amino acid (aa) residues. Predicted sizes of recombinant proteins in aa are shown to the right.

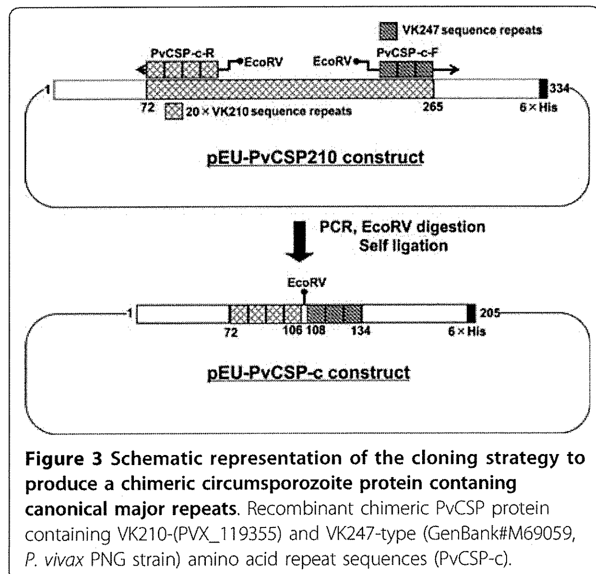


Figure 3 Schematic representation of the cloning strategy to produce a chimeric circumsporozoite protein contaning canonical major repeats. Recombinant chimeric PvCSP protein containing VK210-(PVX_119355) and VK247-type (GenBank#M69059, *P. vivax* PNG strain) amino acid repeat sequences (PvCSP-c).

(CellFree Sciences, Matsuyama, Japan). Briefly, the nucleotide sequences of PvCSP (SalI strain, VK210 type: PVX_119355) excluding the signal peptide and the GPI anchor signal, with addition of penta-His-tag sequence at the C-terminus, was amplified from SalI gDNA by PCR using VK210-F and VK210-R primers, and was cloned at the EcoRV site into the pEU-E01-MCS plasmid (CellFree Sciences) in the presence of both EcoRV restriction enzyme and T4 DNA ligase generating the pEU-PvCSP210 construct without original EcoRV site. The pEU-PvCSP210 was then inversely amplified by PCR using antisense-primer encoding the four times of the VK210-repeat amino acid sequence "GQPAGDRAD" at the 5' end with EcoRV site (PvCSP-c-R) (Additional file 1) and sense-primer encoding the three times of the VK247-repeat amino acid sequence "GANGAGNQP" at the 5' end with EcoRV site (PvCSP-c-F) (Additional file 1). Then the PCR product was digested with EcoRV, and self ligated after the gel-purification of the restricted DNA fragment. Finally, the presence of tetra-VK210-type sequence was confirmed followed by tri-VK247-type repeat amino acid sequences after the nucleotide sequencing of the final pEU-PvCSP-c plasmid (Figure 3). Deduced amino acid sequences, Gly_2 to Asp_{106} and Asn_{134} to Cys_{199} in PvCSP-c was identical to Gly_{23} to Asp_{127} and Asn_{286} to Cys_{347} based on the *SalI* sequence, PVX_ 119355, and Gly_{108} to Pro_{134} in PvCSP-c was identical to Gly_{248} to Pro_{274} based on the deduced amino acid sequence from *P. vivax* PNG strain, M69059.

In vitro protein synthesis

In vitro protein synthesis followed the original manufacturers' instructions (Roche) and was done on a 50 μl scale, excepting for PvCSP-c (see below). Expressed proteins were purified on GST SpinTrap purification columns (GE Healthcare). Briefly, soluble fractions from cell-free system extracts were applied to a Glutathione Sepharose® 4B column that had been equilibrated with PBS. The column was washed with PBS and the bound GST-HBx fusion protein was eluted with 10 mM glutathione in 50 mM Tris-HCl, pH 8.0. Eluted proteins were extensively dialyzed in PBS to remove glutathione. Proteins were analysed by SDS-page and Western blot and quantified as described else where [24]

Larger scale wheat germ cell-free protein synthesis

The recombinant PvCSP-c protein was synthesized with the wheat germ cell-free protein expression system using the bilayer translation reaction method on a 30 ml scale as manufacturer's recommendation (CellFree Sciences) [14]. The PvCSP-c protein was affinity purified by Ni-affinity chromatography as described previously [25]. Briefly, add imidazole (pH 8.0) in the translation reaction mixture (final concentration, 20 mM) and then add Ni-NTA beads (QIAGEN, Valencia, CA). Incubate the tube for 16 h on a continuous rotator, at 4°C, for the binding of proteins on to the beads. Transfer the solution with the beads into a Poly-Prep chromatography column (Bio-Rad, Hercules, CA). Wash the beads by five bed-volumes of PBS containing 30 mM imidazole three times and then elute the recombinant protein with one bed-volume of PBS containing 500 mM imidazole.

Covalent coupling of recombinant proteins to beads

BioPlex carboxylated beads (Bio-Rad) were covalently coated with the different recombinant proteins following the manufacturer's instructions (BioPlex Amine Coupling Kit). Briefly, activated beads (1.25×10^6 beads) were resuspended in 100 μl of PBS and 1 μg of each recombinant protein used per coupling reaction. Incubation under rotation was done at 4°C overnight and coupled beads were washed with 500 μl of PBS pH 7.4. After re-suspending coupling beads in 250 μl of blocking buffer and further incubation under rotation at room temperature for 30 min, beads were washed with 500 μl of storage buffer and centrifuged for six minutes at 14,000 × g. Pellets were resuspended into 125 μl of the same buffer and stored at 4°C protected from light until use.

Analysis of coupled beads on the BioPlex system

Coupled beads were analysed in the Bioplex system as previously described [20] with slight modifications. Briefly, circa 3,000 coated beads were used for each assay. Frozen plasma samples were thawed at room temperature, diluted 1:50 in assay buffer and 50 μl aliquots

added to the beads (final plasma dilution 1:100). Aliquots of 50 μl of Biotinylated human IgG antibody (Sigma) diluted 1:10,000 and of phycoerythrin conjugated streptavidin diluted to 1 μg/ml were used in subsequent incubations. Beads were re-suspended in 125 μl of assay buffer (BioRad) and analysed on the BioPlex100 system and results were expressed as median fluorescent intensity (MFI).

Statistical analysis

T-test and chi-square or fisher exact test were used to compare mean levels for prevalence, respectively, between groups. Averages were expressed as geometric mean (GM) plus 95% confidence intervals (CI). To evaluate the statistical measure of agreement between two independent proteins the index Kappa was calculated.

Results

Cloning and expression of *Plasmodium vivax* proteins

Expression of genes encoding five *P. vivax* proteins: PvMsp1-19, PvMsp1-Nter, PvMsp5, PvMsp7, PvDBP-RII, PvCsp-S and GST as control was initially attempted in *E. coli* and wheat germ cell-free expression systems using commercially available vectors (Roche). Yields, however, were very low and highly degraded as detected by Western blot analysis. It was thus decided to incorporate a GST tag into these vectors (Figure 1) as GST increased the solubility and yields of different recombinant proteins [26]. Noticeably, when cloned into these vectors, both expression systems produced readily detectable proteins by Western blot analysis under reducing condition (Figure 4). Proteins expressed by the cell-free *E. coli* system, however, were mostly degraded and showed low amounts of intact proteins with predicted sizes (Figure 4A). In contrast, soluble proteins expressed in wheat germ cell-free system were of predicted sizes

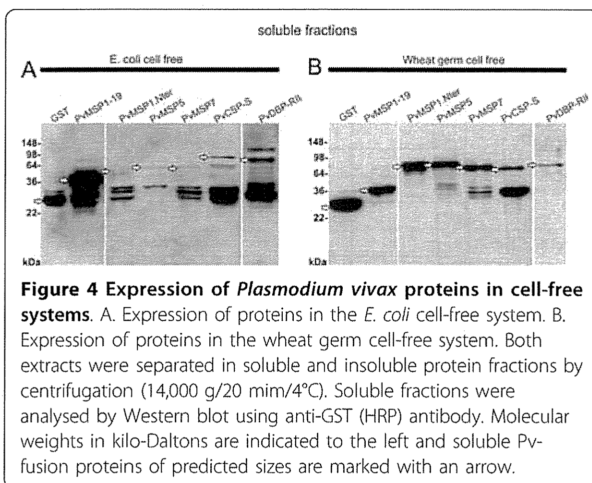


Figure 4 Expression of *Plasmodium vivax* proteins in cell-free systems. A. Expression of proteins in the *E. coli* cell-free system. B. Expression of proteins in the wheat germ cell-free system. Both extracts were separated in soluble and insoluble protein fractions by centrifugation (14,000 g/20 mim/4°C). Soluble fractions were analysed by Western blot using anti-GST (HRP) antibody. Molecular weights in kilo-Daltons are indicated to the left and soluble Pv-fusion proteins of predicted sizes are marked with an arrow.

and had much less degradation products (Figure 4B). All proteins produced by wheat germ system were affinity-purified to 60-85% and yielded 1-10ug/50ul (Additional file 2). Soluble purified proteins were coupled to individual bioplex beads and coupling efficiency was verified prior to multiplexing using an anti-GST or anti-his antibody (Additional file 3).

Proteins produced by wheat germ system are recognized by significantly higher number of immune sera than those produced by *E. coli*

Only three soluble proteins produced in the 50 μl scale in *E. coli* could be purified in a single-step and coupled to Bioplex beads using exactly the same methodology as those produced and purified by wheat germ system. A comparison of naturally acquired humoral IgG responses against these proteins was thus made using immune sera of 40 malaria patients from Brazil known to have large reactivity against PvMSP1 [22]. GST values were subtracted from MFI values obtained against individual recombinant proteins and the cut-off defined as the mean value of control sera +3 standard deviations. Noticeably, proteins produced in the wheat germ system were recognized in significantly higher numbers than those produced in the *E. coli* system (MSP1-19 wheat germ 37/40 (92.5%) vs MSP1-19 *E. coli* 19/40 (47.5%), p = 0.000; MSP1-N wheat germ 26/40 (65%) vs MSP1-N *E. coli* 8/40 (20%), p = 0.000; MSP5 wheat germ 34/40 (85%) vs MSP5 *E. coli* 23/40 (57.5%) (p = 0.001) (Figure 5). Moreover, values of geometric means of all proteins produced in wheat germ system were significantly higher than those produced in *E. coli* system and 95% confidence intervals reinforced such differences. This data demonstrates that identical soluble proteins expressed in wheat germ system and coupled to bioplex beads are better recognized by the same immune sera than those expressed by *E. coli* system.

Multiplex assays with proteins produced in wheat germ system as an alternative platform for antigen discovery

To illustrate the use of soluble proteins produced by wheat germ system in a 50 μl scale and multiplexing assays for immuno-epidemiological studies, the responses of other proteins also considered important targets for *P. vivax* vaccine development were determined. These include (besides MSP1-19, MSP1-N, and MSP5), MSP7 [27], PvDBP-RII [28], and CSP [29]. Moreover, a chimeric CSP protein produced in large-scale in wheat germ and containing the two major allele repeats of PvCSP was also included (Figure 3 and Additional file 1). Of note, for this analysis a different group of 40 sera pertaining to other individuals with no particular strong reactivity against PvMSP1 was used [22].