Supplementary Information for

The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*

Daniel E. Neafsey[1], Kevin Galinsky[1], Rays H. Y. Jiang[1], Lauren Young[1], Sean M. Sykes[1], Sakina Saif[1], Sharvari Gujja[1], Jonathan M. Goldberg[1], Sarah Young[1], Qiandong Zeng[1], Sinéad B. Chapman[1], Aditya P. Dash[2], Anupkumar R. Anvikar[2], Patrick L. Sutton[3], Bruce W. Birren[1], Ananias A. Escalante[4], John W. Barnwell[5], and Jane M. Carlton[3]

[1]Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA

[2]National Institute of Malaria Research, Indian Council of Medical Research, Sector 8, Dwarka, New Delhi 110 077, India

[3]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA

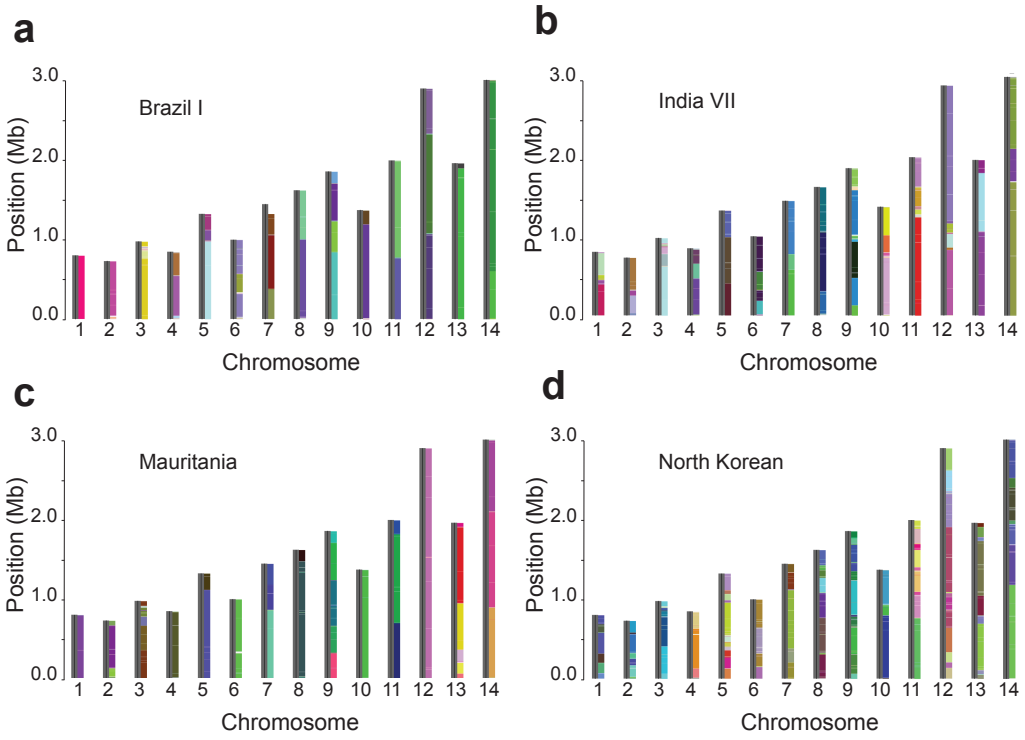
[4]Center for Evolutionary Medicine & Informatics, The Biodesign Institute, School of Life Sciences
Arizona State University, PO Box 874501, Tempe AZ 85287-450, USA

[5]Divison of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA

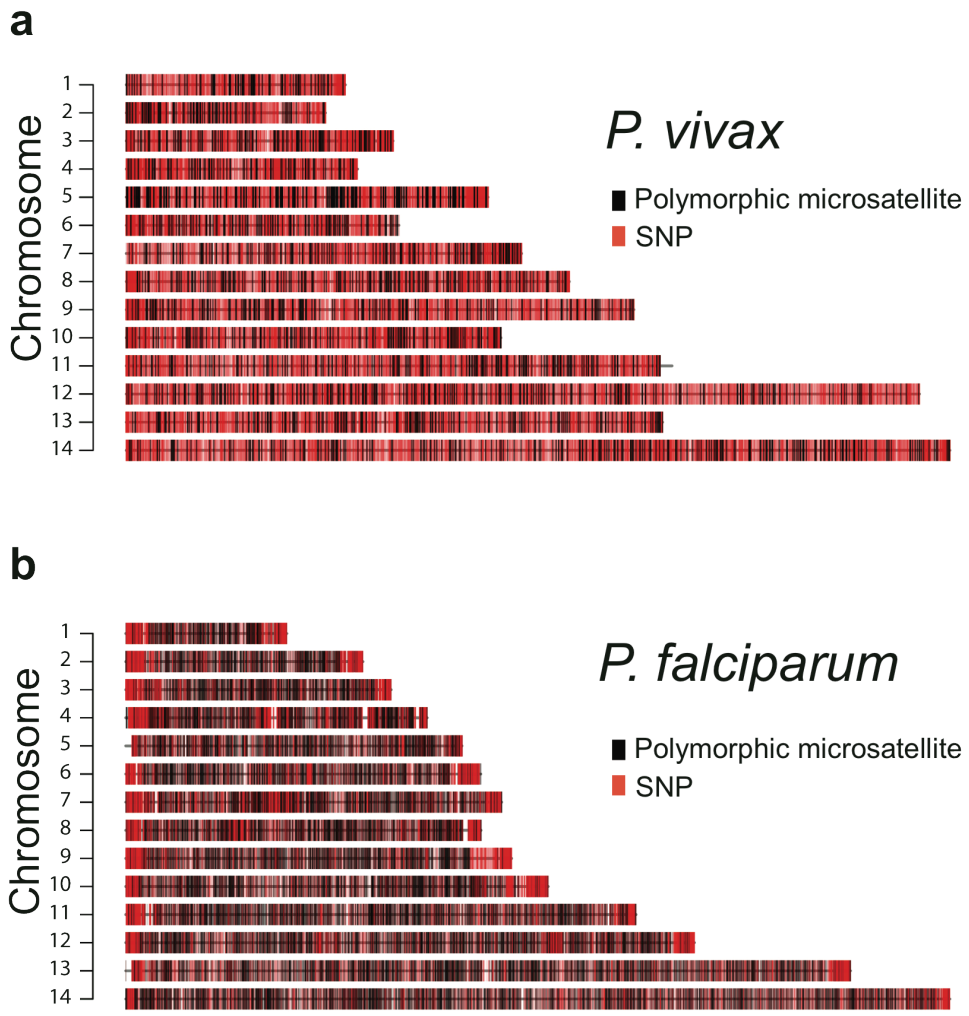
Correspondence should be addressed to J.C. (jane.carlton@nyu.edu).

List of Supplementary Information:

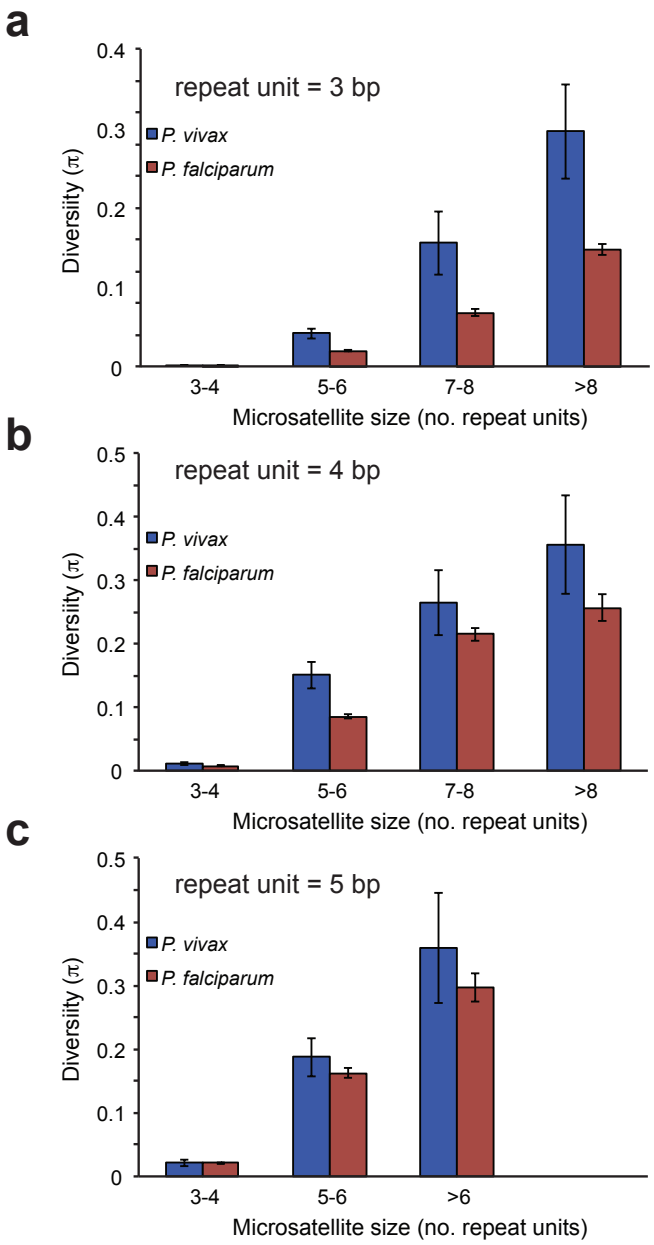
Supplementary Figure 1. Synteny conservation in *P. vivax*. Panels represent scaffold alignments to the Salvador I assembly for the Brazil I (**a**), India VII (**b**), Mauritania (**c**), and North Korean (**d**) isolates. The vast majority of scaffolds in the newly sequenced isolates map contiguously to single chromosome in the Salvador I assembly. Colors were assigned to scaffolds arbitrarily. An interactive synteny browser for these genome assemblies is located at http://www.broadinstitute.org/annotation/genome/plasmodium_vivax/ToolsIndex.html.



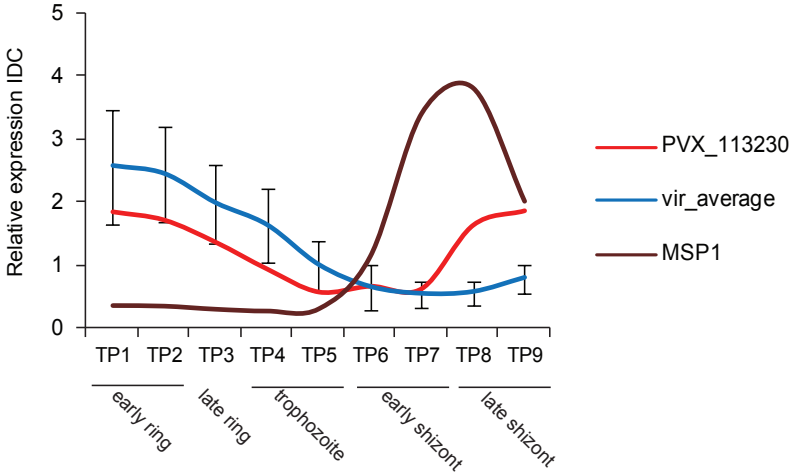
Supplementary Figure 2. Genomic distribution of SNPs and microsatellite polymorphisms for *P. vivax* (a) and *P. falciparum* (b). In both panels, SNP positions are indicated by red bars and polymorphic microsatellite positions are indiciated by black bars.



Supplementary Figure 3. Microsatellite diversity comparison for repeat unit motif sizes of 3 (a), 4 (b), and 5 bp (c). Error bars indicate standard errors. Comparisons were undertaken only for classes of loci (repeat unit size, number) with at least 25 instances in both *P. vivax* and *P. falciparum*.

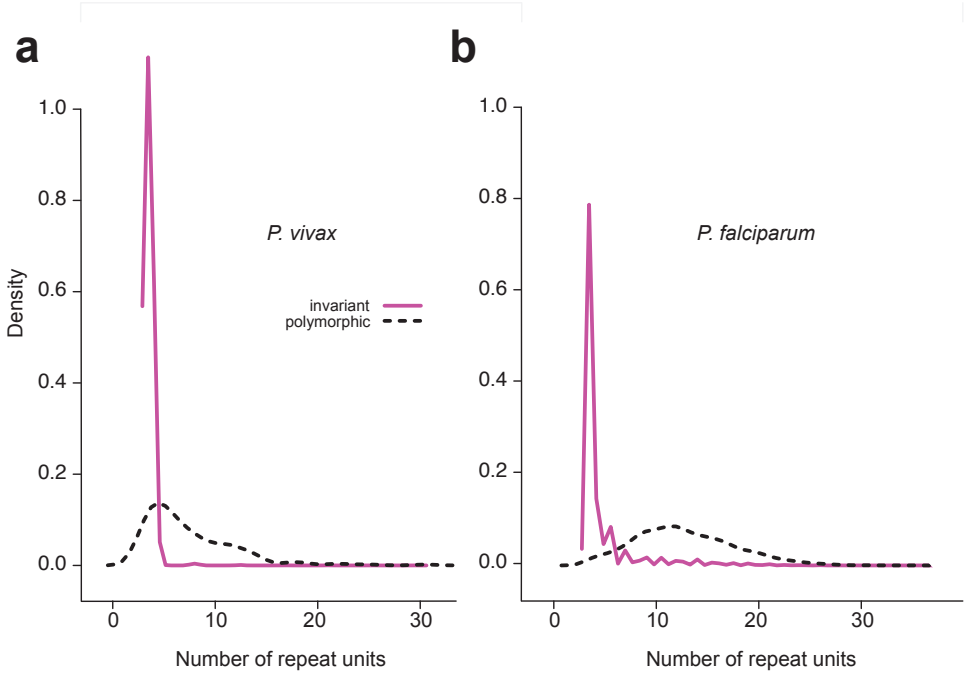


Supplementary Figure 4. Expression profile of ultra-conserved PVX_113230 compared to 50 other *vir* genes and MSP1. Data from Bozdech et al.[1] Error bars represent standard error. The 50 *vir* genes used as comparator loci were the most highly expressed members of the family.



1. Bozdech, Z. *et al.* The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.*

Supplementary Figure 5. Density profile of polymorphic vs. invariant microsatellites in *P. vivax (a) and P. falciparum (b).* Longer microsatellites are more likely to be polymorphic in both species, presumably due to a greater opportunity for replication slippage.



Supplementary Table 1. *P. falciparum* assembly quality relative to *P. vivax* assembly quality. Assemblies for both species were produced using similar levels of sequencing coverage and identical methodologies.

Strain	Assembly size (Mb)	Contig N50 (kb)	No. contigs	Scaffold N50 (kb)	No. scaffolds	% coverage of reference
***P. vivax* (mean)**	29.1	27.7	2,342	685.7	394	98.2
***P. falciparum* 3D7**	20.5	4.2	6,819	384.5	810	89.3

Supplementary Table 2. SNP counts by sequence class.

		SNP Counts					
Species	**Isolate**	**Intergenic**	**Intronic**	**Syn**	**Nonsyn**	**Total SNPs**	**Total Called Bases**
P. falciparum	87_239	917	871	3,368	8,682	19,356	21,587,453
P. falciparum	Dd2	1,968	1,775	4,789	12,182	34,044	21,921,621
P. falciparum	HB3	135	596	2,295	5,597	12,208	20,763,965
P. falciparum	ML-14	475	527	2,864	7,296	14,184	21,493,537
P. falciparum	SenTh231.08	908	1,153	4,127	10,413	22,950	22,099,238
P. vivax	Brazil I	708	2,293	4,494	8,005	31,982	22,161,108
P. vivax	India VII	626	2,562	5,374	8,911	36,315	22,198,012
P. vivax	Mauritania I	689	2,886	5,920	10,368	41,205	22,275,676
P. vivax	North Korean	825	2,973	5,833	10,276	41,027	22,270,333
P. vivax	IQ07	264	1,692	3,414	5,360	19,552	20,363,534

Supplementary Table 3. Gene Ontology (GO)-based functional enrichment analysis of non-synonymous SNP diversity. Positive Z scores indicate a significant (P< 0.05 after Bonferroni correction, Mann Whitney U test) excess of nonsynonymous SNPs, as measured by the π statistic for nonsynonymous SNPs. Negative Z scores indicate a significant deficit. d_N/d_S values indicate mean evolutionary rate of proteins in each group between *P. vivax* and *P. knowlesi*.

GO category	Z score	GO category_description	Number of genes	*P.vivax vs. P.knowlesi* mean d_N/d_S
GO:0003774	5.21	motor activity	440	0.28
GO:0016459	4.55	myosin complex	268	0.23
GO:0030260	3.75	entry into host cell	69	0.19
GO:0022624	-3.68	proteasome accessory complex	76	0.112
GO:0030660	-3.81	Golgi-associated vesicle membrane	74	0.13
GO:0034641	-3.87	cellular nitrogen compound metabolic process	158	0.17
GO:0030119	-3.93	AP-type membrane coat adaptor complex	270	0.13
GO:0016796	-4.20	exonuclease activity	262	0.17
GO:0009408	-4.31	response to heat	224	0.21
GO:0065004	-4.33	protein-DNA complex assembly	26	0.17
GO:0044427	-4.33	chromosomal part	417	0.22
GO:0016896	-4.45	exoribonuclease activity, 5'phosphomonoesters	187	0.17
GO:0005794	-4.50	Golgi apparatus	109	0.11
GO:0005996	-4.72	monosaccharide metabolic process	42	0.075
GO:0017171	-5.16	serine hydrolase activity	196	0.16
GO:0005798	-5.36	Golgi-associated vesicle	144	0.16
GO:0006730	-5.48	one-carbon metabolic process	97	0.17
GO:0005838	-5.56	proteasome regulatory particle	127	0.17
GO:0005643	-5.63	nuclear pore	324	0.11
GO:0006333	-5.89	chromatin assembly or disassembly	243	0.18
GO:0051087	-5.91	chaperone binding	331	0.12
GO:0009262	-6.24	deoxyribonucleotide metabolic process	237	0.24
GO:0019843	-6.73	rRNA binding	94	0.05
GO:0005853	-6.74	eukaryotic translation elongation factor 1 complex	145	0.10
GO:0051246	-7.02	regulation of protein metabolic process	166	0.16
GO:0006839	-7.09	mitochondrial transport	284	0.13
GO:0016765	-7.42	transferase activity, transferring alkyl or aryl groups	430	0.14
GO:0031410	-7.49	cytoplasmic vesicle	212	0.15
GO:0022627	-8.11	cytosolic small ribosomal subunit	145	0.069
GO:0005832	-9.06	chaperonin-containing T-complex	259	0.13
GO:0003743	-9.15	translation initiation factor activity	347	0.13
GO:0016209	-9.17	antioxidant activity	380	0.16
GO:0006979	-9.19	response to oxidative stress	417	0.16

Supplementary Table 6. Pairwise genetic distances (per 10,000 sites) at fourfold degenerate sites.

P. falciparum	3D7	HB3	ML-14	87_239	Dd2
HB3	0.9031				
ML-14	0.7733	0.6047			
87_239	0.8862	0.8121	0.4597		
Dd2	0.9599	0.7766	0.5100	0.7429	
Th231.08	0.8757	0.6923	0.6375	0.8695	0.8338

P. vivax	Salvador I	Brazil I	IQ07	India VII	N. Korean
Brazil I	1.244				
IQ07	1.140	1.173			
India VII	1.515	1.572	1.555		
N. Korean	1.457	1.577	1.566	1.505	
Mauritania I	1.493	1.628	1.533	1.494	1.561

Supplementary Table 7. Empirical nucleotide substitution matrices at 4D synonymous sites.

P. falciparum	**Derived**			
Ancestral	**A**	**C**	**G**	**T**
A	50208	12	39	14
C	17	8391	5	21
G	34	7	8178	10
T	26	35	13	50212

P. vivax	**Derived**			
Ancestral	**A**	**C**	**G**	**T**
A	44645	36	95	16
C	42	51853	50	111
G	112	42	50954	45
T	26	100	38	44956