High proportion of genome-wide homology and increased pre-treatment *pvcrt* levels in *Plasmodium vivax* late recurrences: a chloroquine therapeutic efficacy study

Eduard Rovira-Vallbona[1]#, Nguyen Van Hong[2], Johanna H. Kattenberg[1], Ro Mah Huan[3], Nguyen Thi Huong Binh[2], Nguyen Thi Hong Ngọc[2] , Pieter Guetens[1], Nguyen Luong Hieu[2], Nguyen Thị Thu Hien[2], Vu Thi Sang[2], Nguyen Duc Long[2], Erin Sauve[1], Tran Thanh Duong[2], Nguyen Xuan Xa[2], Annette Erhart[1,4], Anna Rosanas-Urgell[1]#

[1] Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium.

[2] National Institute of Malariology, Parasitology and Entomology, Hanoi, Vietnam.

[3] Centre for Disease Control and Prevention, Gia Lai, Vietnam.

[4] Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, The Gambia.

E.R.V. and N.V.H. contributed equally to this work.

Supplemental Material

Table S1. PCR reaction mixes and thermocycling conditions.

A) PCR reaction mixes and thermocycling conditions for microsatellite genotyping.

Gene	Primer sequences	qPCR mix (µl)		Thermocycling conditions	
pvmsp1F3	msp1F3-fwdp:	Hotstart 10x Buffer	2	1x	95°C, 10min
	5'-GGAGAACATAAGCTACCTGTCC-3'	dNTP mix (20mM)	2		95°C, 15sec
	msp1F3-revp:	MgCl2 (25 mM)	1.6	25x	59°C, 30sec
	5'-GTTGTTACTTGGTCTTCCTCCC-3'	Forward Primer (10µM)	0.5		72°C, 30sec
		Reverse Primer (10µM)	0.5	1x	72°C, 10min
		Hotstart Taq Polymerase (Qiagen)	0.2		
		H_2O Molecular Biology grade	8.2		
		DNA	5		
	msp1F3-fwdn:	Hotstart 10x Buffer	2	1x	95°C, 10min
	VIC-5'-CAAGCCTACCAAGAATTGATCCCCAA-3'	dNTP mix (20mM)	2		95°C, 15sec
	msp1F3-revn:	MgCl2 (25 mM)	1.6	25x	59°C, 30sec
	5'-ATTACTTTGTCGTAGTCCTCGGCGTAGTCC-3' + tail	Forward Primer (10µM)	0.5		72°C, 30sec
		Reverse Primer (10µM)	0.5	1x	72°C, 10min
		Hotstart Taq Polymerase (Qiagen)	0.2		
		H_2O Molecular Biology grade	8.2		
		Primary PCR product	5		
MS4	MS4-fwd:	Hotstart 10x Buffer	2.5	1x	95°C, 10min
	PET-5'- CGATTTACTGTTGACGCTGAA-3'	dNTP mix (20mM)	0.5		95°C, 30sec
	MS4-rev:	Forward Primer (10µM)	2	40x	60ºC, 40sec
	5'-CTGTCTTCAAAGGAACATGCTCGATGA-3'	Reverse Primer (10µM)	2		72°C, 40sec
		Acetylated BSA (0,1µg/µl)	0.25	1x	72°C, 10min
		Hotstart Taq Polymerase (Qiagen)	0.2		
		H_2O Molecular Biology grade	12.55		
		DNA	5		
MS10	MS10fwd:	Hotstart 10x Buffer	2.5	1x	95°C, 10min
	FAM-5'-TTATCCCTGCTGGATGTGAA-3'	dNTP mix (20mM)	0.5		95°C, 30sec
	MS10rev:	Forward Primer (10µM)	2	40x	58°C, 40sec
	5'-CTGTCTTTCCTTCAGGTGGGACTTGTT-3'	Reverse Primer (10µM)	2		72°C, 40sec
		Acetylated BSA (0,1µg/µl)	0.25	1x	72°C, 10min
		Hotstart Taq Polymerase (Qiagen)	0.2		
		H_2O Molecular Biology grade	12.55		
		DNA	5		
PvSal1814	PvSal1814-fwd:	Hotstart 10x Buffer	5	1x	95°C, 5min
	FAM-5'-AAACAGGCATTAGGTTTAAGAGTG-3'	dNTP mix (20mM)	0.5		95°C, 30sec
	PvSal1814-rev:	Forward Primer (10µM)	1	40x	54ºC, 40sec
	CAGTGGCTTCTTCTTTAGTGG	Reverse Primer (10µM)	1		72°C, 40sec
		acetylated BSA (0,1µg/µl)	0	1x	72°C, 8min
		Hotstart Taq Polymerase (Qiagen)	0.2		
		H_2O Molecular Biology grade	32.3		
		DNA	10		

B) PCR reaction mixes and thermocycling conditions for *pvmdr1* genotyping.

Gene	Primer sequences	qPCR mix (µl)		Thermocycling conditions	
pvmdr1	GolPvmdr-1 OF 5'-CGCCATTATAGCCCTGAGCA-3' GolPvmdr-1 OR 5'-TCTCACGTCGATGAGGGACT-3'	5X Buffer MgCl2 (25 mM) dNTP mix (20mM) Forward Primer (10µM) Reverse Primer (10µM) GoTaq Polymerase (Promega) H_2O Molecular Biology grade DNA	10 5 0.5 1 1 0.25 27.25 5	1x 33x 1x	95°C, 5min 95°C, 15sec 55°C, 30sec 72°C, 1min 72°C, 7min
	Lin-mdr1_Fw 5'-ATAGTCATGCCCCAGGATTG-3' Lin-mdr1_Rev 5'-CCTTTCGAAGGACAGCTTTG-3'	Buffer MgCl2 (25 mM) dNTP mix (20mM) Forward Primer (10µM) Reverse Primer (10µM) GoTaq Polymerase (Promega) H_2O Molecular Biology grade Primary Product	10 5 0.5 1 1 0.25 31.25 1	1x 33x 1x	95°C, 5min 95°C, 15sec 55°C, 30sec 72°C, 1min 72°C, 7min

Data S1. (Excel [Data_S1_GRC.xlsx]). Genetic Report Card (SpotMalaria). The dataset contains drug-resistance genotypes as well as the SNP barcode (Sheet A). An explanatory table with detail of sequenced positions and reference/mutant codon is provided in Sheet B.

Table S2. Whole-genome sequencing coverage.

Patient code	Day 0			Day of recurrence		
	parasites/µl	coverage (mean X ± SD)	% missing positions	parasites/µl	coverage (mean X ± SD)	% missing positions
007	41506	41.2 ± 52	0.8%	769	21.4 ± 33	4.4%
008	6068	27.6 ± 51	13.8%	3428	45.0 ± 43	0.4%
010	5431	6.8 ± 11	0.9%	212	1.6 ± 15	88.7%
018	2408	25.5 ± 37	2.9%	103	3.3 ± 22	81.9%
022	13430	38.9 ± 54	1.6%	444	44.0 ± 46	0.4%
027	5517	37.1 ± 57	4.6%	3268	42.2 ± 42	0.4%
034	4010	42.1 ± 56	2.1%	1536	45.5 ± 47	0.4%
037	11401	38.5 ± 55	2.5%	310	37.5 ± 41	0.6%
043	55892	42.7 ± 39	0.3%	7477	34.2 ± 40	0.8%
044	274	46.4 ± 51	0.6%	68	2.5 ± 9	48.6%
047	4927	25.0 ± 55	27.8%	91	35.0 ± 39	0.6%
048	13965	50.1 ± 47	0.3%	856	16.9 ± 41	35.5%
053	11262	8.6 ± 40	81.0%	2469	21.8 ± 52	34.4%
054	14096	32.3 ± 50	4.0%	440	35.7 ± 55	3.9%
All		**33.1 ± 47**	**10.2%**		**27.6 ± 38**	**21.5%**

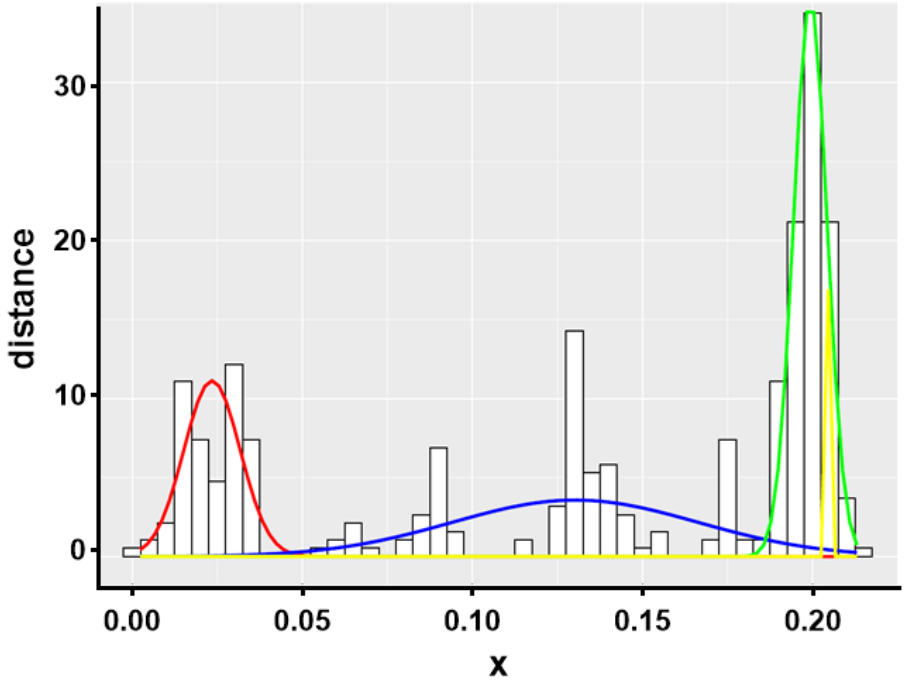

Fig S1. Distribution of discordant SNPs by whole-genome sequencing. The proportion of discordant SNPs was determined by calculating the Prevosti distance between all pair-wise comparisons (*i.e.* number of allelic differences/number of possible differences) using the R-package *poppr* (https://cran.r-project.org/web/packages/poppr/index.html). Due to the high likelihood of multiclonal infections in *P. vivax*, distances were fitted as the sum of four Gaussian distributions using the R-package *mixtools* (https://cran.r-project.org/web/packages/mixtools/index.html). The distribution with the lowest proportion of discordant SNPs (red line) was used to define the threshold (mean plus 3x the standard deviation for that distribution) of homologous recurrence for identify-by-state (IBS) analysis.

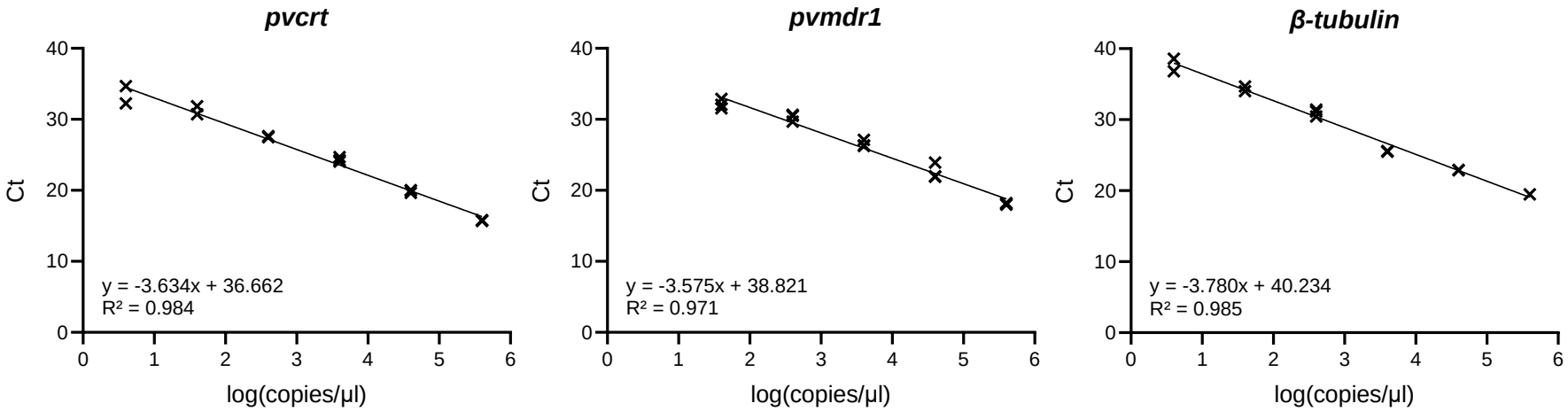

Fig S2. **qPCR standard curves for *pvcrt, pvmdr1* and *β-tubulin.*** Standard curves were prepared using 10-fold serial dilutions of dsDNA fragments covering the targeted region in each gene of interest (range 4000000-4 copies/µl). Amplification efficiencies calculated from cycle threshold (Ct) *versus* dsDNA copy number (log transformed) plots were 1.88, 1.90 and 1.84 for *pvcrt, pvmdr1* and *β-tubulin,* respectively.

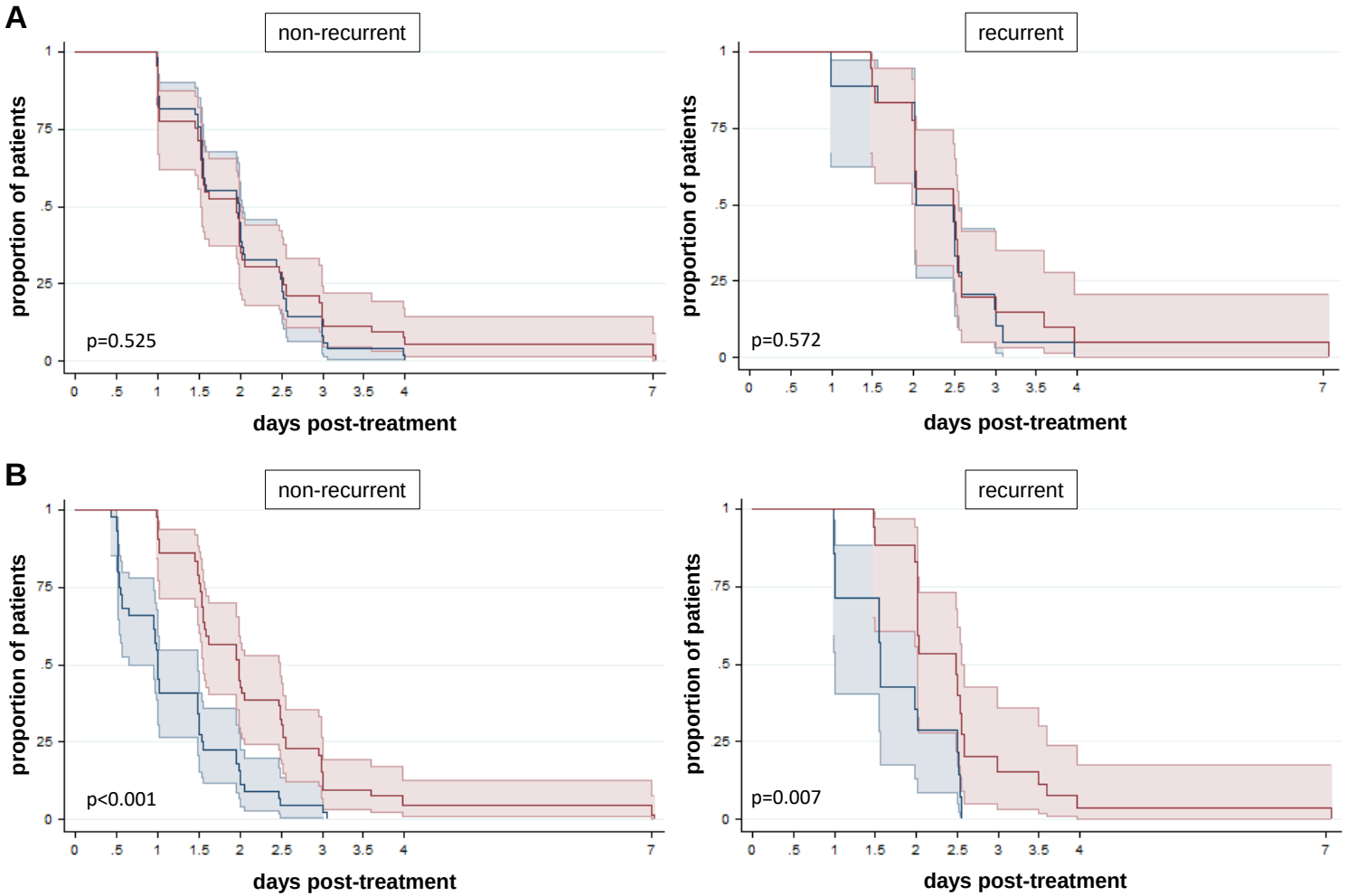

Fig S3. Time to parasite clearance by treatment outcome. Kaplan-Meier survival curves for total parasite clearance (A) and gametocyte clearance (B) separated by non-recurrent (n=49) or recurrent (n=18) courses of infection. Curves are stratified by light microscopy (blue line) and qPCR (A; red line) or RTqPCR (B; red line), respectively. Colored areas indicate 95% confidence intervals. P-values were calculated using log-rank test. One patient with no Day 7 sample by qPCR/RTqPCR was excluded from this analysis in the non-recurrent group.

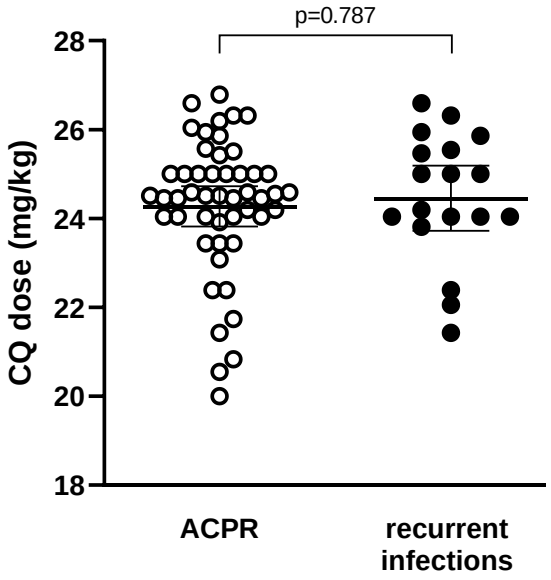

Fig S4. CQ dose received by patients. Graph shows the exact CQ dose administered in mg per kg of body weight, stratified by treatment outcome.

Table S3. **Microsatellites genotyping.** Results of allele calling using Genetools software. Shared alleles between sample pairs at Day 0 and Day of recurrence (DRec) are indicated in bold.

Patient code	Day	PvMSP1.F3					MS4		MS10		PvSal1814					
		Allele1	Allele2	Allele3	Allele4	Allele5	Allele1	Allele2	Allele1	Allele2	Allele1	Allele2	Allele3	Allele4	Allele5	Allele6
003	Day0	**264**	313				195	**198**	216		**528**	**543**	558			
	DRec	**264**						**198**			**528**	**543**				
007	Day0	229		264	278	**313**	198	**201**	**201**	207	**599**	**614**	**629**			
	DRec		259			**313**		**201**	**201**		**599**	**614**	**629**			
008	Day0	**276**					**195**		**201**		**563**	**578**				
	DRec	**276**	325				**195**		**201**		**563**	**578**				
010	Day0	259	306				204		213							
	DRec															
018	Day0	**259**	264	316			198	201	**204**	213	510	525	540			
	DRec	**259**							**204**							
021	Day0	253		300				204	195					573	588	603
	DRec		259				201			207	528	543	558			
022	Day0	219	**278**	327			**198**		**195**		**599**	614	629			
	DRec		**278**				**198**		**195**		**599**					
027	Day0			325	**374**		**201**		**198**		**555**	**584**	**599**			
	DRec	147	197		**374**		**201**		**198**		**555**	**584**	**599**			
028	Day0	147	197	325	374		201		198							
	DRec															
034	Day0	**325**					**201**		**198**			581		596		
	DRec	**325**					**201**		**198**		555		584			
037	Day0	259					201		201							
	DRec															
043	Day0	264	**272**						**204**		555	569	**584**			
	DRec		**272**				198		**204**				**584**	614		
044	Day0	**264**					204		**210**		**584**	**599**	**614**			
	DRec	**264**							**210**		**584**	**599**	**614**			
047	Day0	**253**	**264**				195	**201**	207		599					
	DRec	**253**	**264**					**201**	204							
048	Day0	**264**					**223**		**201**							
	DRec	**264**					**223**		**201**		599					
050	Day0	**264**					198		**213**		**620**	635				
	DRec	**264**							**213**		**620**					
053	Day0	**258**					**201**		**201**			**573**	**588**			
	DRec	**258**					**201**		**201**		558	**573**	**588**	603		
054	Day0	**250**					**213**		**204**		**620**	**635**				
	DRec	**250**					**213**		**204**		**620**	**635**				

Table S4. Genotyping and categorization of recurrent infections.

Patient code	Parasite density by LM (p/µl)		Microsatellites				SNP barcode				Whole-genome sequencing					
			COI		Discordant MS	Type	COI		Disc. SNP	Type	COI		IBS		IBD	
	Day0	DRec	Day0	DRec			Day0	DRec			Day0	DRec	Disc. SNP	Type	% IBD	Type
003	3470	0	3	2	0/3	homologous	1	1	1/38	heterologous	-	-	-	-	-	-
007	41506	769	4	3	0/4	homologous	2	1	2/38	heterologous	>1	>1	14.3%	heterologous	35.2%	heterologous
008	6068	3428	2	2	0/4	homologous	1	1	0/37	homologous	1	1	2.2%	homologous	99.9%	homologous
010	5431	212	2	-	-	-	1	2	1/38	heterologous	-	-	-	-	-	-
018	2408	103	3	1	0/2	homologous	2	-	-	-	>1	-	-	-	-	-
021	3369	1225	3	3	4/4	heterologous	2	-	-	-	-	-	-	-	-	-
022	13430	444	3	1	0/4	homologous	1	2	0/37	homologous	1	>1	6.8%	heterologous	98.1%	homologous
027	5517	3268	3	3	0/4	homologous	1	-	-	-	1	1	2.3%	homologous	99.5%	homologous
028	6849	24	4	-	-	-	1	-	-	-	-	-	-	-	-	-
034	4010	1536	2	2	1/4	heterologous	1	2	0/38	homologous	1	1	1.6%	homologous	99.8%	homologous
037	11401	310	1	-	-	-	1	-	-	-	1	1	2.0%	homologous	99.7%	homologous
043	55892	7477	3	2	0/3	homologous	1	2	1/38	heterologous	>1	1	12.6%	heterologous	99.8%	homologous
044	274	68	3	3	0/3	homologous	-	2	-	-	>1	-	-	-	-	-
047	4927	91	2	2	1/3	heterologous	2	2	0/22	homologous	>1	>1	8.0%	heterologous	32.5%	heterologous
048	13965	856	1	1	0/3	homologous	1	2	0/20	homologous	1	1	1.7%	homologous	99.5%	homologous
050	5852	0	2	1	0/3	homologous	1	-	-	-	-	-	-	-	-	-
053	11262	2469	2	4	0/4	homologous	1	2	0/38	homologous	-	1	-	-	-	-
054	14096	440	2	2	0/4	homologous	1	2	5/38	heterologous	1	1	2.4%	homologous	100%	homologous
Homologous recurrence rate:			12/15 (80%)				6/11 (55%)				6/10 (60%)				8/10 (80%)	

LM, light microscopy; DRec, samples at day of recurrence; COI, complexity of infection; MS, microsatellites (Pvmsp1F3+MS4+MS10+PvSal1814); IBS, identity-by-state; IBD, identity-by-descent.

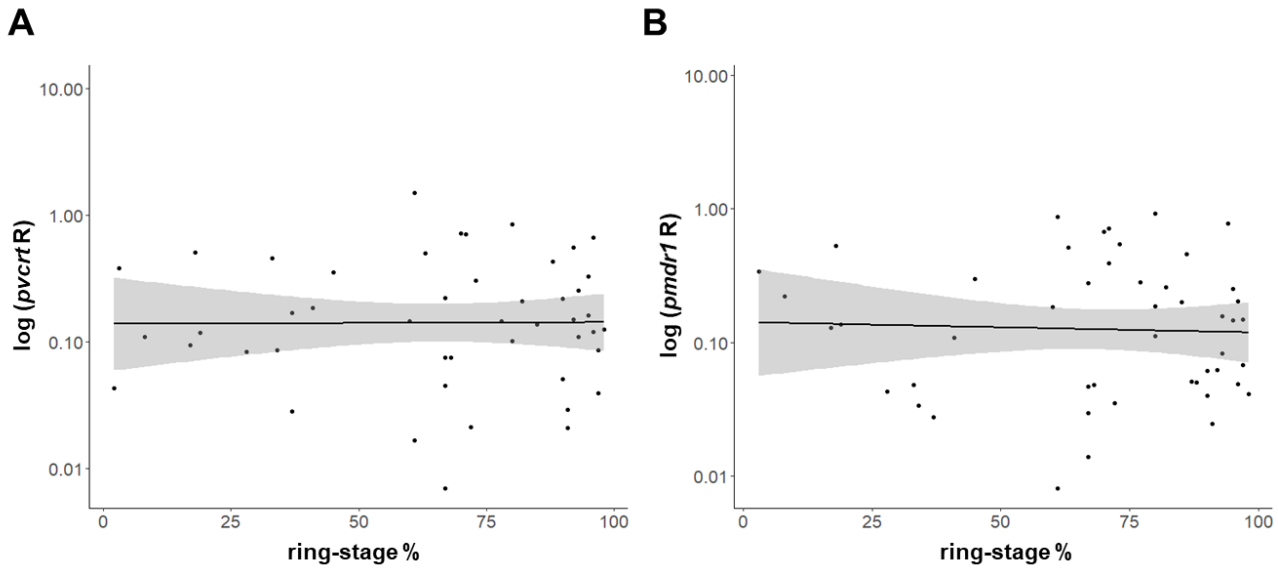

Fig S5. Gene expression of *pvcrt* and *pvmdr1* and proportion of ring-stages in infections at Day 0. Gene expression ratios (R) of *pvcrt* (A) and *pvmdr1* (B) relative to reference gene *β-tubulin* are displayed in log-scale (y axis). Linear regression was fitted to data (black line with standard error bounds in grey) did not show an association between R and ring-stage percentage (*pvcrt*: coef.=-0.0003, p=0.886; *pvmdr1*: coef.=-0.001, p=0.655).

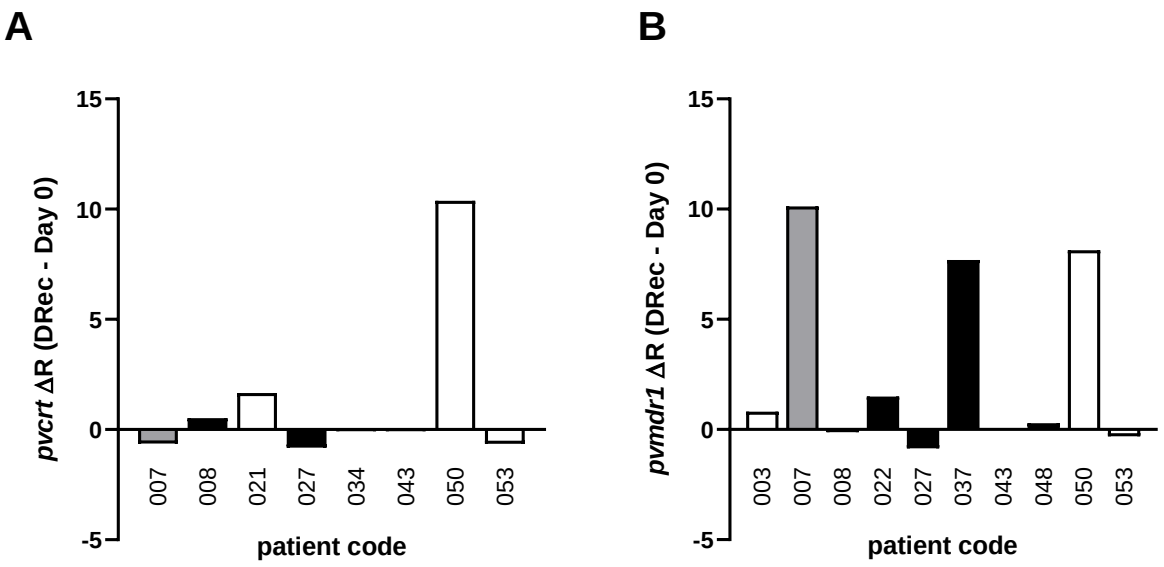

Fig S6. Gene expression of *pvcrt* and *pvmdr1* in Day 0/DRec paired samples. Differences in gene expression ratio R between infections at DRec and paired infections at Day 0 are shown for both *pvcrt* (A) and *pvmdr1* (B). Type of recurrences based on identity-by-descent (IBD) analysis is shown with coloured bars (black, IBD-homologous recurrence; grey, IBD-heterologous recurrence; white, undetermined by WGS).