OPEN ACCESS Freely available online

PLoS one

Gene-Specific Signatures of Elevated Non-Synonymous Substitution Rates Correlate Poorly across the *Plasmodium* Genus

Gareth D. Weedall[1]*, Spencer D. Polley[1], David J. Conway[1,2]

1 Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Medical Research Council Laboratories, Fajara, Banjul, The Gambia

Abstract

Background: Comparative genome analyses of parasites allow large scale investigation of selective pressures shaping their evolution. An acute limitation to such analysis of *Plasmodium falciparum* is that there is only very partial low-coverage genome sequence of the most closely related species, the chimpanzee parasite *P. reichenowi*. However, if orthologous genes have been under similar selective pressures throughout the *Plasmodium* genus then positive selection on the *P. falciparum* lineage might be predicted to some extent by analysis of other lineages.

Principal Findings: Here, three independent pairs of closely related species in different sub-generic clades (*P. falciparum* and *P. reichenowi*; *P. vivax* and *P. knowlesi*; *P. yoelii* and *P. berghei*) were compared for a set of 43 candidate ligand genes considered likely to be under positive directional selection and a set of 102 control genes for which there was no selective hypothesis. The ratios of non-synonymous to synonymous substitutions (dN/dS) were significantly elevated in the candidate ligand genes compared to control genes in each of the three clades. However, the rank order correlation of dN/dS ratios for individual candidate genes was very low, less than the correlation for the control genes.

Significance: The inability to predict positive selection on a gene in one lineage by identifying elevated dN/dS ratios in the orthologue within another lineage needs to be noted, as it reflects that adaptive mutations are generally rare events that lead to fixation in individual lineages. Thus it is essential to complete the genome sequences of particular species of phylogenetic importance, such as *P. reichenowi*.

Citation: Weedall GD, Polley SD, Conway DJ (2008) Gene-Specific Signatures of Elevated Non-Synonymous Substitution Rates Correlate Poorly across the *Plasmodium* Genus. PLoS ONE 3(5): e2281. doi:10.1371/journal.pone.0002281

Editor: Cecile Fairhead, Pasteur Institute, France

Received March 14, 2008; **Accepted** April 13, 2008; **Published** May 28, 2008

Funding: This work was supported in part by a UK Medical Research Council PhD studentship for GDW, and Wellcome Trust programme grant 074695/Z/04/B to DJC.

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

* E-mail: gweedall@hotmail.com

Introduction

Identifying genes under positive directional selection can help understand how parasites adapt to new survival or reproductive challenges. The dN/dS ratio (non-synonymous substitutions per non-synonymous site divided by synonymous substitutions per synonymous site) is commonly applied to scan for evidence of positive selection in comparative genomic analysis [1,2]. Analyses of polymorphism among genome sequences of the human malaria parasite *P. falciparum* [3–5], and divergence between *P. falciparum* and the partially available genome sequence of the chimpanzee parasite *P. reichenowi* [3] show elevated dN/dS ratios in genes encoding membrane and exported proteins (considered to be under positive selection), as well as genes that are expressed at low abundance or at only one stage of the life cycle (considered to be under relaxed negative selection). However, the incompleteness of the *P. reichenowi* genome sequence (available sequence reads aligned to only ~ 42% of the *P. falciparum* 3D7 genome sequence) means that most loci could not be effectively analysed for inter-specific divergence [3], so most signatures of positive directional selection have not yet been discriminated.

Pairwise analyses with other malaria parasite species may also identify loci under positive selection. However, given the great evolutionary distance between many of the species, such as between *P. falciparum* and the rodent parasite *P. yoelii* [6], studies of pairwise dN/dS suffer from too high a sequence divergence, causing synonymous substitutions to be saturated and making estimates of dN/dS rate ratios unreliable. Analyses of closely related species are preferable, and pairwise dN/dS analysis among the genomes of the rodent malaria parasites, *P. yoelii*, *P. berghei* and *P. chabaudi* [7], showed a similar overall trend to the *falciparum-reichenowi* analysis, with putative membrane proteins displaying higher dN/dS values than other genes. Could the results of that analysis (or analysis of other closely related species pairs such as *P. vivax* and *P. knowlesi*) be extrapolated to *P. falciparum* genes for which *P. reichenowi* orthologous sequences are not available? This study tests whether signatures from one clade of the *Plasmodium* genus can be used to predict those in other clades. The

distributions of dN/dS values are compared for sets of orthologous loci in three phylogenetically independent species pairs, investigating a set of 43 candidate genes that are considered likely to be under positive selection and a set of 102 control genes for which there is no selective hypothesis.

Results and Discussion

For each of the 43 candidate ligand genes analysed, inter-specific dN/dS ratios are shown for each of the three closely related species pairs, *P. falciparum* / *P. reichenowi*, *P. vivax* / *P. knowlesi*, and *P. yoelii* / *P. berghei* (Table 1, further details in table S1). To test whether this candidate ligand gene dataset is enriched in genes under positive selection, dN/dS values were compared with the control gene dataset (table S2) for each species pair (Fig. 1A) using Wilcoxon's rank sum test. For all three species pairs the median dN/dS ratio was significantly greater in the candidate ligand gene set than in the control set (*falciparum-reichenowi*, $P = 0.0084$; *vivax-knowlesi*, $P = 0.0175$; *yoelii-berghei*, $P = 0.0003$) (Fig. 1B). This was also seen for dN values (Fig. 1C), though not for dS (Fig. 1D), indicating a signature of positive selection on non-synonymous mutations leading to elevated dN/dS values in a proportion of the candidate ligand genes. Relaxed selective constraint could also result in elevated dN/dS, although there is no reason to expect that the candidate ligand genes should be under any less selective constraint than the control set of genes to maintain protein structure and function.

It should be noted that analysis of any single one of these genes in isolation would not lead to a strong conclusion of positive selection, since none showed a dN/dS value >1. Inter-specific dN/dS values for whole genes are hardly ever >1 even when positive selection occurs, due to the effect of negative background selection on many sites within most genes [1,2], so comparison of relative dN/dS values across sets of genes is a more sensitive way of scanning for evidence of positive selection than searching for individual values above 1 or any other arbitrary cut off.

To assess the predictive power of dN/dS across the *Plasmodium* genus, rank correlations (Spearman's $\rho = rSp$) were applied to test whether similar relative selective forces operate on orthologous genes in different species. Table 2 shows the correlation of dN/dS, dN and dS indices for all genes among the three different species pairs. Pairwise scatterplots of dN/dS values are shown in Fig. 2. The predictive power is quantified by r^2Sp which represents the amount of variability in one axis which can be explained by variability in the other. dN/dS was significantly, though poorly, positively correlated between independent species pairs for both candidate ligand genes and control genes. Correlations were greater for dN than for dS, supporting the idea that selection affects the correlations while synonymous substitutions are mostly stochastic. However, the predictive power of dN/dS for one species pair on another is lower for candidate ligand genes (25 %, 21 % and 31 % for *Pf/Pr* versus *Pv/Pk*, *Pf/Pr* versus *Py/Pb*, and *Pv/Pk* versus *Py/Pb* respectively) than for control genes (55 %, 35 % and 44 % for the respective three comparisons). This indicates that the correlation is not improved by positive selection but is actually made worse. Discrete processes of positive selection will have occurred in different species lineages, against a background of selective constraint that varies among genes in a manner that is apparently more homogeneous between different lineages.

Thus, although broadly similar signatures indicating positive selection on distinct classes of genes may be seen in different parts of the *Plasmodium* phylogeny, predictions about positive selection on individual genes for which sequence data are currently missing in particular species cannot be reliably extrapolated from orthologues in other parts of the phylogeny. To detect loci that have undergone positive directional selection in the lineage of a particular species, sequences must be directly compared with orthologues of a closely related species. As *P. falciparum* is currently the most important human parasite, completion of the closely related *P. reichenowi* genome sequence should now have particularly high priority [3].

Materials and Methods

Sets of candidate genes and controls

A set of 55 single-locus genes encoding surface proteins that are putatively ligands at various life cycle stages was first defined. These genes are candidates to display signatures of positive selection due to their likely role in host-parasite interaction, and of these, 43 could be included in comparative dN/dS analyses as noted in the following section. Loci in this candidate gene dataset were compared with loci from a control dataset chosen to represent an unbiased sample of genes not hypothesised to be under positive selection. The control set was of loci on *P. falciparum* chromosome 3 that contained one or more nucleotide difference among the sequences of five isolates as published [8] with data searchable on PlasmoDB (www.plasmodb.org) [9]. Of the 104 such loci identified, two (PFC0210c and PFC0420w) were already included in the candidate ligand gene dataset and were thus excluded from the control dataset, which therefore consisted of 102 genes.

Defining orthologous genes for analysis of sequence divergence between species

Pairwise nucleotide divergence was estimated for 3 pairs of closely related species: *P. falciparum* and *P. reichenowi*; *P. vivax* and *P. knowlesi*; and *P. yoelii* and *P. berghei* (Fig. 1A). Two other species for which genome sequence data are available, *P. gallinaceum* and *P. chabaudi*, were not included in the present analysis as the former is not very closely related to any other species [10–13], and the latter would add little extra information to the *yoelii-berghei* pair [7]. Protein-coding gene sequences in the *P. falciparum* 3D7 genome sequence (release date 11/02/2005), produced by a consortium of the Wellcome Trust Sanger Institute (WTSI), the Institute for Genomic Research (TIGR) and Stanford University [14], were downloaded from the PlasmoDB website (http://www.plasmodb.org/common/downloads/) [9]; sequences from *P. vivax* (release date 03/11/2005) and *P. yoelii* (23/07/2004) [15], produced by the Institute for Genomic Research (TIGR), were downloaded from the TIGR website (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/); shotgun sequences from *P. reichenowi* (11/03/2004) [3], and gene sequences from *P. knowlesi* (06/01/2006) and *P. berghei* (08/06/2004) [7] were produced by the Wellcome Trust Sanger Institute and were downloaded from the WTSI website (ftp://ftp.sanger.ac.uk/pub/pathogens/).

Orthologues to *P. falciparum* predicted protein sequences were defined by BLASTp (protein vs. protein) searches against databases of *P. yoelii*, *P. berghei*, *P. vivax* and *P. knowlesi* predicted proteins, and required a reciprocal best match against the *P. falciparum* predicted protein database. For added stringency, each pair of putative orthologues (*yoelii-berghei*, *vivax-knowlesi*, *falciparum-reichenowi*) were BLASTed against the database of the other species of the pair to ensure that the best matches to the *P. falciparum* sequences in each species were also reciprocal best matches to each other. Where this was not the case the pair was not analysed (detailed results of BLAST searches are shown in tables S1 and S2).

Table 1. A set of 43 candidate ligand gene loci with dN/dS ratios for three phylogenetically independent *Plasmodium* species pairs (*Pf/Pr*, *Pv/Pk* and *Py/Pb*)

Pf locus ID	Gene product	Evidence for ligand role	References	*Pf*/*Pr*	*Pv*/*Pk*	*Py*/*Pb*
PFB0310c	MSP4	D_a, SE_m	[20,21]	0.84	0.48	-[n]
PFB0305c	MSP5	SE_m	[21,22]	0.80	0.59	-[n]
PF13_0201	TRAP	B, IIA, D_b, AE_s, SE_s	[23–25]	0.70	0.35	0.72
MAL13P1.60	EBA140	B, IIA, AE_m	[26–28]	0.57	0.42	-[n]
PF10_0352	MSP11	SE_m	[29]	0.49	-[n]	-[n]
PF11_0486	MAEBL	B, IIA, D_b, SE_s, AE_m	[30–32]	0.48	0.37	0.30
PFA0125c	EBA181	B, AE_m	[33]	0.47	-[n]	-[n]
PFE0080c	RAP2	AE_m	[34,35]	0.44	-[n]	-[n]
PF10_0302	P28	IIA, D_b, SE_o	[36,37]	0.43	-[n]	0.57
PF13_0248	P47	SE_g	[38]	0.42	0.12	0.55
PFD1150c	RH4	AE_m	[39–41]	0.42	-[n]	-[n]
PFD0210c	P36	D_b	[42]	0.42	0.18	0.32
PF10_0303	P25	IIA, D_b, SE_o	[36,37]	0.38	0.45	0.81
PF14_0102	RAP1	IIA, AE_m	[34,35,43]	0.37	0.19	0.85
PFF0615c	Pf12	SE_m	[44]	0.37	0.14	0.12
PFF0995c	MSP10	D_a, AE_m	[21,45]	0.30	0.38	-[P]
PF11_0344	AMA1	B, IIA, D_a, AE_m/SE_m, AE_s/SE_s	[46–48]	0.30	0.14	0.30
PFL0800c	celTOS	D_b, AE_o	[49]	0.26	0.63	-[n]
PF10_0344	GLURP	SE_m	[50]	0.25	-[n]	-[n]
PFE0395c	Pf38	SE_m	[44]	0.24	0.11	0.25
PFI1730w	CLAG9	D_c, AE_m	[51,52]	0.24	-[n]	-[n]
PFC0640w	CTRP	D_b, AE_o	[53,54]	0.24	0.25	0.34
PFC0210c	CSP	B, SE_s	[55–57]	0.20	0.56	0.41
PFI1445w	RhopH2	AE_m	[58]	0.18	0.07	0.10
PF13_0247	P48/45	D_d, SE_g	[59]	0.14	0.15	0.21
PFI0265c	RhopH3	D_a, AE_m	[60,61]	0.11	-[n]	0.32
PFL2510w	CHT1	D_b, $AE_{o(secreted)}$	[62,63]	0.10	0.14	0.13
PFE0075c	RAP3	AE_m	[34]	0.09	-[n]	0.89
PFL0870w	PTRAMP	AE_m	[64]	0.08	0.18	0.10
MAL7P1.208	RAMA	B, D_a, AE_m	[21,65]	0.08	0.23	0.47
PFB0405w	P230	B, IIA, SE_g	[66,67]	0.08	0.11	0.23
PFC0120w	CLAG3.2	AE_m	[68]	0.07	-[n]	-[n]
PFB0570w	SPATR	B, IIA, SE_s	[69]	0.001	0.12	0.09
PF08_0003	TryThrA	IES	[70]	-[n]	-[P]	0.10
PF14_0040	SOAP	D_b, AE_o	[71]	-[n]	0.56	-[n]
PF13_0338	Pf92	D_a, SE_m	[21,44]	-[n]	0.27	-[n]
PF08_0136b	WARP	AE_o	[72]	-[n]	0.26	0.43
PFD0215c	P36p	D_b, SE_s	[42]	-[n]	0.25	0.30
PFE0120c	MSP8	$SE_{m(in\ P.\ yoelii)}$	[73]	-[n]	0.18	0.12
PFI1145w	PLP3/MAOP	D_b, AE_o	[74]	-[n]	0.17	0.13
PFL1385c	MSP9	B, SE_m	[75,76]	-[n]	0.17	0.07
PFD0240c	Pf41	SE_m	[44]	-[n]	0.16	0.30
PFC0420w	CDPK3	D_b, AE_o	[77,78]	-[n]	0.08	0.18

B = binding assay. IIA = invasion inhibition assay. D = gene disruption experiment which either ($_a$) could not produce viable parasites in asexual culture; ($_b$) reduced or abolished the traversal of cell membranes or tissue layers; ($_c$) abolished receptor binding; or ($_d$) reduced fertilization. SE/AE/IES = surface/apical/infected erythrocyte surface expression at ($_s$) sporozoite, ($_m$) merozoite, ($_g$) gametocyte, or ($_o$) ookinete stage. -[n] unambiguous orthologues could not be identified in one or both species; -[P] orthology could not be resolved among alternative possible orthologues. For each gene a maximum of three references are given. Twelve other candidate ligand loci could not be analysed due to complex sequence evolution (see Methods).

doi:10.1371/journal.pone.0002281.t001

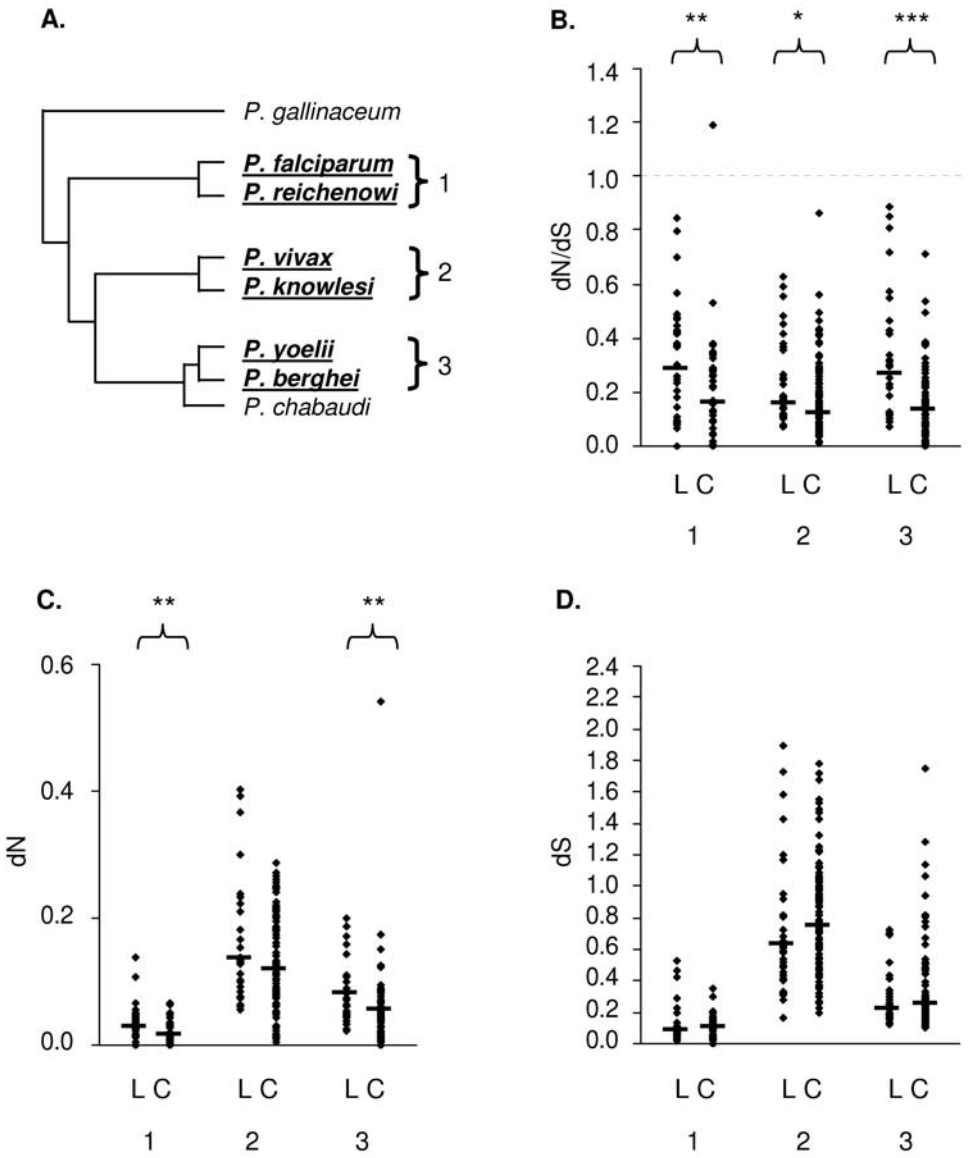

Figure 1. Genetic divergence among three pairs of *Plasmodium* species. A. Schematic representation of the phylogenetic relationship between sequenced *Plasmodium* genomes. Three pairs of closely related species (*falciparum-reichenowi*, *vivax-knowlesi* and *yoelii-berghei*) used for analysis are labelled clade 1, 2 and 3, respectively. (The phylogenetic position of *P. gallinaceum* in relation to the other species is not yet confirmed and awaits full genomic analysis, but is either an outgroup as illustrated here [10,11] or more closely related to the *falciparum-reichenowi* clade). **B.** The distribution of dN/dS for candidate ligand genes and control genes (labelled 'L' and 'C') between species of each clade defined in panel A. Sample sizes were: clade 1, L = 33, C = 37; clade 2, L = 32, C = 92; clade 3, L = 29, C = 70. Asterisks indicate a significant difference between gene datasets by Wilcoxon's rank sum test (*$<$0.05, **$<$0.01, ***$<$0.001). **C.** The distribution of dN for the same loci. One extreme value (PY05686 vs. PB000528.03.0, dN = 8.06) is not shown. **D.** The distribution of dS for the same loci. Two extreme values (PY05686 vs. PB000528.03.0, dS = 45.69; PY02848 vs. PB100183.00.0, dS = 108.79) are not shown.
doi:10.1371/journal.pone.0002281.g001

No database of predicted proteins existed for *P. reichenowi*, so *P. falciparum* predicted protein sequences were used to search the *P. reichenowi* genomic contig database using tBLASTn (protein versus DNA translated in all 6 possible reading frames). In a number of cases where *P. reichenowi* orthologues could not be identified in the contig data, published *P. reichenowi* sequences were obtained from GenBank or sequences built from shotgun sequencing reads were used (table S3). For each gene, the *P. falciparum* coding sequence (introns excluded) was aligned to the best matching *P. reichenowi* contig using the SeqMan II program (DNASTAR, Madison, WI) to define the start and end of the coding sequence and the intron-exon boundaries. *P. reichenowi* contig sequences contained some regions of single-read coverage, so nucleotide mismatches in regions of single-read coverage were edited to match the *P. falciparum* sequence, and only well supported nucleotide mismatches in regions of multiple-read coverage were used for analysis. If a *P. reichenowi* gene sequence contained apparent frameshifts supported by multiple-read coverage, it was considered to be a pseudogene and not analysed.

Forty three of the 55 candidate ligand gene loci examined could be analysed for pairwise divergence between orthologues. Twelve loci (*msp1*, *msp2*, *msp3*, *msp6*, *msp7*, *eba175*, *eba165*, *ebl1*, *rh1*, *rh2a*,

Table 2. Spearman's rank correlation (rSp) of pairwise sequence divergence estimates for orthologous loci among different species pairs

Species pairs compared	Gene dataset	N	Index	rSp	r^2Sp	
falciparum-reichenowi versus *vivax-knowlesi*	Ligand	23	dN/dS	0.50	0.25	*
			dN	0.56	0.32	**
			dS	0.04	0.002	
	Control	35	dN/dS	0.74	0.55	***
			dN	0.76	0.58	***
			dS	0.49	0.24	**
falciparum-reichenowi versus *yoelii-berghei*	Ligand	21	dN/dS	0.46	0.21	*
			dN	0.41	0.17	
			dS	−0.01	0.0002	
	Control	26	dN/dS	0.59	0.35	**
			dN	0.39	0.15	*
			dS	−0.19	0.03	
vivax-knowlesi versus *yoelii-berghei*	Ligand	25	dN/dS	0.56	0.31	**
			dN	0.54	0.30	**
			dS	0.25	0.06	
	Control	67	dN/dS	0.66	0.44	***
			dN	0.54	0.29	***
			dS	0.38	0.15	**

N = number of gene loci analysed for the pairwise correlations between each independent species pair. * $P<0.05$, ** $P<0.01$, *** $P<0.001$

doi:10.1371/journal.pone.0002281.t002

rh2b, *rh3*) were not analysed, because (i) unambiguous orthologues could not be defined, or (ii) molecular evolution appeared complex such that dN and dS may not represent the accumulation of substitutions between species (some genes had dimorphic alleles that were more divergent than the paired species sequence, and others showed evidence of gene conversion with paralogues), or (iii) an orthologue appeared to be a pseudogene. Of the 102 control gene loci, all orthologous pairs identified were analysed unless they contained a pseudogene. The relatively low number of *falciparum-reichenowi* gene pairs analysed (37 of 102 control genes, compared to 92 and 70 for the other species pairs) reflects the low sequence coverage of the *P. reichenowi* genome to date.

Analysis of pairwise between-species dN and dS values for individual genes

Orthologous protein sequence pairs were aligned using clustalW [16] and the protein alignments imposed upon the nucleotide sequences using the program pal2nal [17]. For each sequence pair, pairwise dN, dS and dN/dS indices were estimated by maximum likelihood using the codeml program [18]. Maximum likelihood estimates of dN/dS were used since they are more accurate than approximate methods such as the Nei-Gojobori method when transition/transversion rate biases and nucleotide composition or codon frequency biases exist [19]. Three independent runs of codeml were made with different initial estimates for the transition/transversion rate ratio (k) and the dN/dS ratio (run 1: $k = 1$, dN/dS = 1; run 2: $k = 0.1$, dN/dS = 10; run 3: $k = 10$, dN/dS = 0.1) so that each run began at a different point in the likelihood space. If different final results were obtained between runs, those with the highest log likelihood value (lnL) were used, lower lnL values being assumed to represent local likelihood optima. Non-parametric statistical tests, Wilcoxon's rank sum test and Spearman's rank correlation, were carried out using STATA 9 (StatCorp LP, Texas, USA) as the indices of divergence were not assumed to be normally distributed.

Supporting Information

Table S1 Results of BLAST sequence similarity searches to determine orthologuous pairs of loci in six Plasmodium genomes for 43 candidate ligand genes

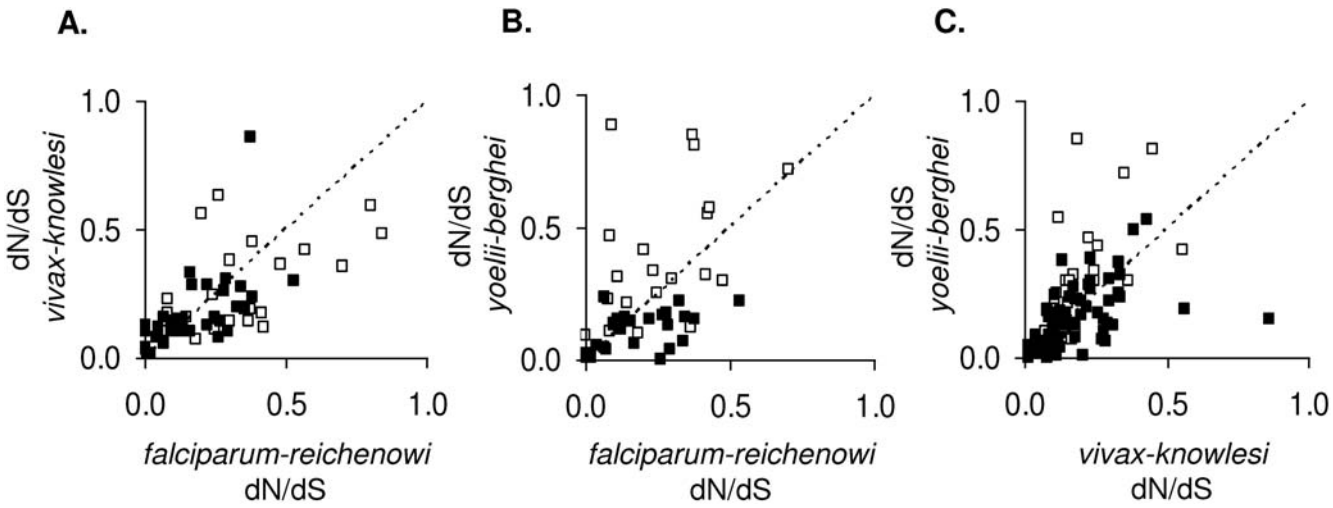

Figure 2. Scatterplots of dN/dS estimates for orthologous loci in independent *Plasmodium* species pairs. A. *vivax-knowlesi vs. falciparum-reichenowi*, **B.** *yoelii-berghei vs. falciparum-reichenowi* and **C.** *yoelii-berghei vs. vivax-knowlesi*. A line of identity representing equal selective constraint and/or positive selection in orthologous genes in different species is shown on each plot (dotted line). Filled squares represent gene pairs from the control gene dataset, open squares gene pairs from the set of candidate ligand genes. Sample sizes and results of Spearman's rank correlation analysis are shown in Table 2.

doi:10.1371/journal.pone.0002281.g002

Found at: doi:10.1371/journal.pone.0002281.s001 (0.08 MB XLS)

Table S2 Results of BLAST sequence similarity searches to determine orthologuous pairs of loci in six Plasmodium genomes for 102 genes on P. falciparum chromosome 3.
Found at: doi:10.1371/journal.pone.0002281.s002 (0.12 MB XLS)

Table S3 Shotgun sequencing reads used to build 7 of the P. reichenowi gene sequences
Found at: doi:10.1371/journal.pone.0002281.s003 (0.03 MB DOC)

References

1. Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. Heredity 86: 641–647.
2. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203–218.
3. Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, et al. (2007) Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. Nat Genet 39: 120–125.
4. Mu J, Awadalla P, Duan J, McGee KM, Keebler J, et al. (2007) Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. Nat Genet 39: 126–130.
5. Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, et al. (2007) A genome-wide map of diversity in *Plasmodium falciparum*. Nat Genet 39: 113–119.
6. Hughes AL, Friedman R (2005) Amino acid sequence constraint and gene expression pattern across the life history in the malaria parasite *Plasmodium falciparum*. Mol Biochem Parasitol 142: 170–176.
7. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, et al. (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 307: 82–86.
8. Mu J, Duan J, Makova KD, Joy DA, Huynh CQ, et al. (2002) Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 418: 323–326.
9. Stoeckert CJ Jr, Fischer S, Kissinger JC, Heiges M, Aurrecoechea C, et al. (2006) PlasmoDB v5: new looks, new genomes. Trends Parasitol 22: 543–546.
10. Perkins SL, Schall JJ (2002) A molecular phylogeny of malarial parasites recovered from *cytochrome b* gene sequences. J Parasitol 88: 972–978.
11. Arisue N, Hirai M, Arai M, Matsuoka H, Horii T (2007) Phylogeny and evolution of the SERA multigene family in the genus *Plasmodium*. J Mol Evol 65: 82–91.
12. Rich SM, Ayala FJ (2003) Progress in malaria research: the case for phylogenetics. Adv Parasitol 54: 255–280.
13. Escalante AA, Ayala FJ (1994) Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proc Natl Acad Sci USA 91: 11373–11377.
14. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498–511.
15. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, et al. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature 419: 512–519.
16. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
17. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34: W609–612.
18. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556.
19. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17: 32–43.
20. Marshall VM, Silva A, Foley M, Cranmer S, Wang L, et al. (1997) A second merozoite surface protein (MSP-4) of *Plasmodium falciparum* that contains an epidermal growth factor-like domain. Infect Immun 65: 4460–4467.
21. Sanders PR, Kats LM, Drew DR, O'Donnell RA, O'Neill M, et al. (2006) A set of glycosylphosphatidyl inositol-anchored membrane proteins of *Plasmodium falciparum* is refractory to genetic deletion. Infect Immun 74: 4330–4338.
22. Marshall VM, Tieqiao W, Coppel RL (1998) Close linkage of three merozoite surface protein genes on chromosome 2 of *Plasmodium falciparum*. Mol Biochem Parasitol 94: 13–25.
23. Muller HM, Reckmann I, Hollingdale MR, Bujard H, Robson KJ, et al. (1993) Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. Embo J 12: 2881–2889.
24. Sultan AA, Thathy V, Frevert U, Robson KJ, Crisanti A, et al. (1997) TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. Cell 90: 511–522.
25. McCormick CJ, Tuckwell DS, Crisanti A, Humphries MJ, Hollingdale MR (1999) Identification of heparin as a ligand for the A-domain of *Plasmodium falciparum* thrombospondin-related adhesion protein. Mol Biochem Parasitol 100: 111–124.
26. Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, et al. (2003) *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nat Med 9: 87–92.
27. Narum DL, Fuhrmann SR, Luu T, Sim BK (2002) A novel *Plasmodium falciparum* erythrocyte binding protein-2 (EBP2/BAEBL) involved in erythrocyte receptor binding. Mol Biochem Parasitol 119: 159–168.
28. Thompson JK, Triglia T, Reed MB, Cowman AF (2001) A novel ligand from *Plasmodium falciparum* that binds to a sialic acid-containing receptor on the surface of human erythrocytes. Mol Microbiol 41: 47–58.
29. Pearce JA, Mills K, Triglia T, Cowman AF, Anders RF (2005) Characterisation of two novel proteins from the asexual stage of *Plasmodium falciparum*, H101 and H103. Mol Biochem Parasitol 139: 141–151.
30. Ghai M, Dutta S, Hall T, Freilich D, Ockenhouse CF (2002) Identification, expression, and functional characterization of MAEBL, a sporozoite and asexual blood stage chimeric erythrocyte-binding protein of *Plasmodium falciparum*. Mol Biochem Parasitol 123: 35–45.
31. Kariu T, Yuda M, Yano K, Chinzei Y (2002) MAEBL is essential for malarial sporozoite infection of the mosquito salivary gland. J Exp Med 195: 1317–1323.
32. Preiser P, Renia L, Singh N, Balu B, Jarra W, et al. (2004) Antibodies against MAEBL ligand domains M1 and M2 inhibit sporozoite development in vitro. Infect Immun 72: 3604–3608.
33. Gilberger T-M, Thompson JK, Triglia T, Good RT, Duraisingh MT, et al. (2003) A novel erythrocyte binding antigen-175 paralogue from *Plasmodium falciparum* defines a new trypsin-resistant receptor on human erythrocytes. J Biol Chem 278: 14480–14486.
34. Howard RF, Reese RT (1990) *Plasmodium falciparum*: hetero-oligomeric complexes of rhoptry polypeptides. Exp Parasitol 71: 330–342.
35. Ridley RG, Lahm H-W, Takacs B, Scaife JG (1991) Genetic and structural relationships between components of a protective rhoptry antigen complex from *Plasmodium falciparum*. Mol Biochem Parasitol 47: 245–247.
36. Hisaeda H, Stowers AW, Tsuboi T, Collins WE, Sattabongkot JS, et al. (2000) Antibodies to malaria vaccine candidates Pvs25 and Pvs28 completely block the ability of *Plasmodium vivax* to infect mosquitoes. Infect Immun 68: 6618–6623.
37. Tomas AM, Margos G, Dimopoulos G, van Lin LH, de Koning-Ward TF, et al. (2001) P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. Embo J 20: 3975–3983.
38. van Schaijk BC, van Dijk MR, van de Vegte-Bolmer M, van Gemert GJ, van Dooren MW, et al. (2006) Pfs47, paralog of the male fertility factor Pfs48/45, is a female specific surface protein in *Plasmodium falciparum*. Mol Biochem Parasitol 149: 216–222.
39. Gaur D, Singh S, Jiang L, Diouf A, Miller LH (2007) Recombinant *Plasmodium falciparum* reticulocyte homology protein 4 binds to erythrocytes and blocks invasion. Proc Natl Acad Sci U S A 104: 17789–17794.
40. Kaneko O, Mu J, Tsuboi T, Su X, Torii M (2002) Gene structure and expression of a *Plasmodium falciparum* 220-kDa protein homologous to the *Plasmodium vivax* reticulocyte binding proteins. Mol Biochem Parasitol 121: 275–278.
41. Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, et al. (2005) Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. Science 309: 1384–1387.
42. Ishino T, Chinzei Y, Yuda M (2005) Two proteins with 6-cys motifs are required for malarial parasites to commit to infection of the hepatocyte. Mol Microbiol 58: 1264–1275.
43. Sterkers Y, Scheidig C, da Rocha M, Lepolard C, Gysin J, et al. (2007) Members of the low-molecular-mass rhoptry protein complex of *Plasmodium falciparum* bind to the surface of normal erythrocytes. J Infect Dis 196: 617–621.

Acknowledgments

We thank Dr Richard Emes (University College London) and colleagues at the London School of Hygiene and Tropical Medicine for helpful discussions on the analyses. We also thank the Wellcome Trust Sanger Institute (WTSI) and the Institute for Genomic Research (TIGR) for making public the data from *Plasmodium* genome sequencing projects.

Author Contributions

Conceived and designed the experiments: DC GW. Performed the experiments: GW. Analyzed the data: SP DC GW. Wrote the paper: SP DC GW.

44. Sanders PR, Gilson PR, Cantin GT, Greenbaum DC, Nebl T, et al. (2005) Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. J Biol Chem 280: 40169–40176.
45. Black CG, Wang L, Wu T, Coppel RL (2003) Apical location of a novel EGF-like domain-containing protein of *Plasmodium falciparum*. Mol Biochem Parasitol 127: 59–68.
46. Silvie O, Franetich JF, Charrin S, Mueller MS, Siau A, et al. (2004) A role for apical membrane antigen 1 during invasion of hepatocytes by *Plasmodium falciparum* sporozoites. J Biol Chem 279: 9490–9496.
47. Triglia T, Healer J, Caruana SR, Hodder AN, Anders RF, et al. (2000) Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. Mol Microbiol 38: 706–718.
48. Kato K, Mayer DC, Singh S, Reid M, Miller LH (2005) Domain III of *Plasmodium falciparum* apical membrane antigen 1 binds to the erythrocyte membrane protein Kx. Proc Natl Acad Sci U S A 102: 5552–5557.
49. Kariu T, Ishino T, Yano K, Chinzei Y, Yuda M (2006) CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. Mol Microbiol 59: 1369–1379.
50. Borre MB, Dziegiel M, Hogh B, Petersen E, Rieneck K, et al. (1991) Primary structure and localization of a conserved immunogenic *Plasmodium falciparum* glutamate rich protein (GLURP) expressed in both the preerythrocytic and erythrocytic stages of the vertebrate life cycle. Mol Biochem Parasitol 49: 119–131.
51. Trenholme KR, Gardiner DL, Holt DC, Thomas EA, Cowman AF, et al. (2000) *clag9*: A cytoadherence gene in *Plasmodium falciparum* essential for binding of parasitized erythrocytes to CD36. Proc Natl Acad Sci U S A 97: 4029–4033.
52. Ling IT, Florens L, Dluzewski AR, Kaneko O, Grainger M, et al. (2004) The *Plasmodium falciparum clag9* gene encodes a rhoptry protein that is transferred to the host erythrocyte upon invasion. Mol Microbiol 52: 107–118.
53. Yuda M, Sakaida H, Chinzei Y (1999) Targeted disruption of the *Plasmodium berghei* CTRP gene reveals its essential role in malaria infection of the vector mosquito. J Exp Med 190: 1711–1716.
54. Dessens JT, Beetsma AL, Dimopoulos G, Wengelnik K, Crisanti A, et al. (1999) CTRP is essential for mosquito infection by malaria ookinetes. Embo J 18: 6221–6227.
55. Frevert U, Sinnis P, Cerami C, Shreffler W, Takacs B, et al. (1993) Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. J Exp Med 177: 1287–1298.
56. Cerami C, Frevert U, Sinnis P, Takacs B, Clavijo P, et al. (1992) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of *Plasmodium falciparum* sporozoites. Cell 70: 1021–1033.
57. Sidjanski SP, Vanderberg JP, Sinnis P (1997) *Anopheles stephensi* salivary glands bear receptors for region I of the circumsporozoite protein of *Plasmodium falciparum*. Mol Biochem Parasitol 90: 33–41.
58. Ling IT, Kaneko O, Narum DL, Tsuboi T, Howell S, et al. (2003) Characterisation of the *rhoph2* gene of *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 127: 47–57.
59. van Dijk MR, Janse CJ, Thompson J, Waters AP, Braks JA, et al. (2001) A central role for P48/45 in malaria parasite male gamete fertility. Cell 104: 153–164.
60. Cowman AF, Baldi DL, Healer J, Mills KE, O'Donnell RA, et al. (2000) Functional analysis of proteins involved in *Plasmodium falciparum* merozoite invasion of red blood cells. FEBS Lett 476: 84–88.
61. Lustigman S, Anders RF, Brown GV, Coppel RL (1988) A component of an antigenic rhoptry complex of *Plasmodium falciparum* is modified after merozoite invasion. Mol Biochem Parasitol 30: 217–224.
62. Dessens JT, Mendoza J, Claudianos C, Vinetz JM, Khater E, et al. (2001) Knockout of the rodent malaria parasite chitinase pbCHT1 reduces infectivity to mosquitoes. Infect Immun 69: 4041–4047.
63. Tsai YL, Hayward RE, Langer RC, Fidock DA, Vinetz JM (2001) Disruption of *Plasmodium falciparum* chitinase markedly impairs parasite invasion of mosquito midgut. Infect Immun 69: 4048–4054.
64. Thompson J, Cooke RE, Moore S, Anderson LF, Janse CJ, et al. (2004) PTRAMP; a conserved *Plasmodium* thrombospondin-related apical merozoite protein. Mol Biochem Parasitol 134: 225–232.
65. Topolska AE, Lidgett A, Truman D, Fujioka H, Coppel RL (2004) Characterization of a membrane-associated rhoptry protein of *Plasmodium falciparum*. J Biol Chem 279: 4648–4656.
66. Eksi S, Czesny B, van Gemert GJ, Sauerwein RW, Eling W, et al. (2006) Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production. Mol Microbiol 61: 991–998.
67. Quakyi IA, Carter R, Rener J, Kumar N, Good MF, et al. (1987) The 230-kDa gamete surface protein of *Plasmodium falciparum* is also a target for transmission-blocking antibodies. J Immunol 139: 4213–4217.
68. Kaneko O, Yim Lim BY, Iriko H, Ling IT, Otsuki H, et al. (2005) Apical expression of three RhopH1/Clag proteins as components of the *Plasmodium falciparum* RhopH complex. Mol Biochem Parasitol 143: 20–28.
69. Chattopadhyay R, Rathore D, Fujioka H, Kumar S, de la Vega P, et al. (2003) PfSPATR, a *Plasmodium falciparum* protein containing an altered thrombospondin type I repeat domain is expressed at several stages of the parasite life cycle and is the target of inhibitory antibodies. J Biol Chem 278: 25977–25981.
70. Burns JM Jr, Adeeku EK, Dunn PD (1999) Protective immunization with a novel membrane protein of *Plasmodium yoelii*-infected erythrocytes. Infect Immun 67: 675–680.
71. Dessens JT, Siden-Kiamos I, Mendoza J, Mahairaki V, Khater E, et al. (2003) SOAP, a novel malaria ookinete protein involved in mosquito midgut invasion and oocyst development. Mol Microbiol 49: 319–329.
72. Yuda M, Yano K, Tsuboi T, Torii M, Chinzei Y (2001) von Willebrand Factor A domain-related protein, a novel microneme protein of the malaria ookinete highly conserved throughout *Plasmodium* parasites. Mol Biochem Parasitol 116: 65–72.
73. Shi Q, Cernetich-Ott A, Lynch MM, Burns JM Jr (2006) Expression, localization, and erythrocyte binding activity of *Plasmodium yoelii* merozoite surface protein-8. Mol Biochem Parasitol 149: 231–241.
74. Kadota K, Ishino T, Matsuyama T, Chinzei Y, Yuda M (2004) Essential role of membrane-attack protein in malarial transmission to mosquito host. Proc Natl Acad Sci U S A 101: 16310–16315.
75. Kushwaha A, Perween A, Mukund S, Majumdar S, Bhardwaj D, et al. (2002) Amino terminus of *Plasmodium falciparum* acidic basic repeat antigen interacts with the erythrocyte membrane through band 3 protein. Mol Biochem Parasitol 122: 45–54.
76. Li X, Chen H, Oo TH, Daly TM, Bergman LW, et al. (2004) A co-ligand complex anchors *Plasmodium falciparum* merozoites to the erythrocyte invasion receptor band 3. J Biol Chem 279: 5765–5771.
77. Ishino T, Orito Y, Chinzei Y, Yuda M (2006) A calcium-dependent protein kinase regulates *Plasmodium* ookinete access to the midgut epithelial cell. Mol Microbiol 59: 1175–1184.
78. Siden-Kiamos I, Ecker A, Nyback S, Louis C, Sinden RE, et al. (2006) *Plasmodium berghei* calcium-dependent protein kinase 3 is required for ookinete gliding motility and mosquito midgut invasion. Mol Microbiol 60: 1355–1363.