

## Adaptation of *Plasmodium falciparum* to its transmission environment

Martin K Rono<sup>1,2,3</sup>, Mary A Nyonda<sup>4</sup>, Joan J Simam, Joyce M Ngoi<sup>1</sup>, Sachel Mok<sup>5</sup>, Moses M Kortok<sup>1</sup>, Abdullah S Abdullah<sup>6</sup>, Mohammed M Elfaki<sup>7</sup>, John N Waitumbi<sup>8</sup>, Ibrahim M El-Hassan<sup>9</sup>, Kevin Marsh<sup>1,3</sup>, Zbynek Bozdech<sup>10</sup> and Margaret J Mackinnon\*

<sup>1</sup>KEMRI-Wellcome Research Programme, Kilifi, 80108, Kenya. <sup>2</sup>Pwani University Bioscience Research Centre, Pwani University, Kilifi, 80108, Kenya. <sup>3</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, OX3 7FZ, UK. <sup>4</sup>Department of Microbiology and Molecular Medicine, Medical Faculty, University of Geneva, Geneva, 1211, Switzerland. <sup>5</sup>Columbia University Medical Center, New York, NY, 10032, USA. <sup>6</sup>Rochester Regional Health - Unity Hospital, Rochester, NY, 14626, USA. <sup>7</sup>Department of Microbiology and Parasitology, Faculty of Medicine, Jazan University, Gizan, 45142, Jazan, Kingdom of Saudi Arabia. <sup>8</sup>Walter Reed Army Institute of Research/Kenya Medical Research Institute, Kisumu, 40100, Kenya. <sup>9</sup>Faculty of Public Health and Tropical Medicine, Jazan University, Gizan 45142, Jazan, Kingdom of Saudi Arabia. <sup>10</sup>School of Biological Sciences, Nanyang Technological University, Singapore, 637551, Singapore. \*email: [mackinnon.mackinnon@gmail.com](mailto:mackinnon.mackinnon@gmail.com)

**Success in eliminating malaria will depend on whether parasite evolution outpaces control efforts. Here, we show that *Plasmodium falciparum* parasites (the deadliest of the species causing human malaria) found in low transmission intensity areas have evolved to invest more in transmission to new hosts (reproduction) and less in within-host replication (growth) than parasites found in high transmission areas. At the cellular level, this adaptation manifests as increased production of reproductive forms (gametocytes) early in the infection at the expense of processes associated with multiplication inside red blood cells, especially membrane transport and protein trafficking. At the molecular level, this manifests as changes in expression levels of genes encoding epigenetic and translational machinery. Specifically, expression levels of the gene encoding AP2-G - the transcription factor that initiates reproduction – increase as transmission intensity decreases. This is accompanied by down-regulation and up-regulation of genes encoding HDAC1 and HDA1, two histone deacetylases that epigenetically regulate the parasite's replicative and reproductive life stage programmes, respectively. Parasites in reproductive mode show increased reliance on the prokaryotic translation machinery found inside the plastid-derived organelles. Thus our dissection of the parasite's adaptive regulatory architecture has identified new potential molecular targets for malaria control.**

**Abbreviations:** H, high transmission; L, low transmission; HR, host responsiveness; IDC, intraerythrocytic cycle.

During their life cycle, malaria parasites alternate between asexual replication in the bloodstream of their vertebrate host and reproduction in their mosquito vector. Asexual replication leads to disease, reproduction leads to transmission to new hosts, and these trade against each other. Evolutionary theory predicts that parasites will adapt their replication-reproduction schedule according to the transmission opportunities and constraints imposed by the prevailing environment<sup>1-3</sup>. This has consequences to population disease burden as well as risk to individuals, potentially undermining the efficacy of control programmes<sup>4</sup>. In the light of recent widespread success in reducing malaria transmission<sup>5</sup>,

here we tested whether parasites are naturally pre-adapted to their transmission environment and therefore might evolve in response to human intervention. We compared transcriptomes of *P. falciparum* parasites *ex vivo* from non-immune children in East Africa living in areas with long-term differences in malaria transmission intensity (Fig. 1, Supplementary Fig. 1). Three independent experiments were performed: two compared geographically isolated populations (high vs. medium and medium vs. low transmission) and a third compared populations from before and after a major decline in malaria within a medium transmission area. Data on 4615 genes measured in 96 parasite isolates at all stages of the parasite's 48h replication cycle in the bloodstream (intraerythrocytic development cycle, IDC, 580 samples in total, Dataset 1, Supplementary Table 1), revealed more than double the expected number of genes with significantly different expression levels between 'High' vs. 'Low' populations ('H-L genes', Supplementary Tables 2 and 3). Many genes, especially those H-L down-regulated, showed consistent effects across the three experiments and steady declines in expression levels across the full range of transmission intensities (Fig. 2c, Supplementary Table 3, Supplementary Figs. 2 and 3). These results were robust to substantial biases in the data caused by host, gene and parasite factors (Supplementary Tables 4 and 5, Supplementary Figs. 2, 4, 5, 6 and 7). Our data support the hypothesis that natural populations of *P. falciparum* are locally adapted to the level of transmission intensity in their environment.

*Fig. 1 here*

Because parasites in H and L populations were sourced from non-immune children (three or fewer previous clinical episodes) at the acute stage of their infections (Supplementary Table 1, Supplementary Fig. 6), we attribute these H-L differences to general properties of the parasite population rather than to facultative responses by individual parasite isolates to the immune or physiological status of their current host. To prosecute this claim, we augmented the data from one of the populations with parasite isolates from semi-immune children (Dataset 2, Supplementary Table 1) and then analysed within this population for genes showing significant responsiveness in expression levels to host immunity-related clinical factors ('host responsive genes, HR', Supplementary Table 4). H-L and HR genes overlapped by only 13% (Supplementary Fig. 2). Furthermore, the set of genes significant for H-L differences did not change when adjustments for host factors were incorporated into the analyses (Supplementary Fig. 2). Thus we conclude that H-L differences were not due to parasite plasticity in response to cues from their immediate host, but instead reflected population-level adaptation to their wider environment.

### **H-L genes cluster by function**

By mapping H-L genes onto a whole-genome co-expression network annotated for putative function, we revealed that H-L up-regulated genes localised to regions associated with asexual stages in the blood and liver (where parasites spend ~10 days following a mosquito bite before emerging into the bloodstream), while H-L down-regulated genes associated with sexual stages in the blood (gametocytes) and mosquito (Fig. 2, Supplementary Figs. 8 and 9, Supplementary Tables 2 and 6). H-L regulated genes tended to fall in the periphery of the network (Fig. 2b), implying their role in non-core processes. Gene set enrichment analyses confirmed H-L associations with asexual vs. sexual function: while H-L up-regulated genes were significantly concentrated in functional pathways involved in biosynthesis and development of asexual stages inside the red blood cell, especially membrane transport,

protein trafficking, haemoglobin digestion, glycolysis and post-translational modifications, H-L down-regulated genes were enriched in pathways involved in transmission and related functions (gametocyte and ookinete production, motility and lipid metabolism) (Fig. 2d, Supplementary Fig. 9, Supplementary Tables 2 and 7). Thus parasites from high transmission areas appear to invest more in asexual replication inside their host and less in transmission between hosts than their counterparts from low transmission areas.

*Fig. 2 here*

### **Individual H-L genes and their functions**

The metabolic functions most associated with H-L up-regulation were transport and protein trafficking (Fig. 2d, Supplementary Figs. 10 and 11). Both these processes are fundamental to supporting the high biosynthesis demand of ~10-fold replication every 48h in the human bloodstream<sup>6</sup>. Most facets of membrane transport (nutrient import, maintenance of ion concentrations, removal of toxins) were H-L up-regulated, with the exception of transport across the membranes of the mitochondrion and apicoplast, the two organelles of plastid origin found in the parasite's cytosol. Examples of H-L up-regulated membrane transport genes were nucleoside transporter 1, NT1, which imports purines, the precursors for DNA which the parasite cannot make for itself<sup>7</sup>, and components of V-type ATPase proton pumps which remove from the cytosol excess H<sup>+</sup> generated by glycolysis (the main source of energy during the IDC) and maintain high H<sup>+</sup> in the digestive compartment to enable breakdown of haemoglobin for supply of amino acids for protein synthesis<sup>8</sup>. Likewise, most protein trafficking pathways were H-L up-regulated, particularly for transfer of haemoglobin to the digestive compartment, nucleus-cytosol traffic, and ER-Golgi traffic via vesicles. Localisation of most transport and trafficking genes to the liver or early bloodstream regions of the network (lime green, pink and maroon network modules, Supplementary Figs. 9, 10 and 11) indicate their specialised roles for life inside human cells. Genes involved in housekeeping functions such as DNA replication, cell division and supply of tRNAs for protein synthesis (magenta module), and those regulating clonal antigenic variation (see below, pale green module) were generally not H-L regulated (Fig. 2d, Supplementary Table 6 and Table 7).

The most H-L down-regulated functional pathway was that for early stage gametocytes (Stages I and II, turquoise module, Fig. 2b), with many genes encoding hallmark proteins of gametocytes and gametes<sup>9</sup> showing significant H-L down-regulation in all experiments (Fig. 2c, Fig. 3, Supplementary Table 2). Of particular note amongst these was the gene encoding AP2-G (Fig. 2d), the DNA-binding transcription factor that is the 'master key' to unlock the developmental programme of gametocytes<sup>10</sup> and a member of a family of DNA-binding transcription factors in *Plasmodium* credited with regulating life stage-specific transcriptional programmes<sup>11</sup>. During the IDC, *AP2-g* is epigenetically silenced by heterochromatin: this is mediated by the histone modifier Class II histone deacetylase 2 (HDA2)<sup>12</sup>. HDA2 also causes general silencing of *var* genes<sup>12</sup>, the large gene family that encode proteins (PfEMP1) exported by the parasite to the red blood cell surface where they bind other host cells, causing life-threatening disease<sup>13</sup>. By virtue of PfEMP1's high antigenicity and variability, coupled with sequential and mutually exclusive expression afforded by HDA2-mediated background silencing, the *var* gene family is believed responsible for prolonging infections in the face of the adaptive immune system<sup>14</sup>. HDA2, by

silencing both *AP2-g* and *var* genes during the IDC, has thus been proposed as the molecular nexus governing the replication-reproduction divide<sup>12</sup>.

### Epigenetic control of H-L adaptation

In our data, the gene encoding HDA2 (PF10\_0078) was not significantly H-L up-regulated ( $P = 0.73$ ). Instead, two other members of the five genes encoding Class II histone deacetylases possessed by *P. falciparum* were - one up-regulated (*HDAC1*, PF11260,  $P = 0.04$ ) and one down-regulated (*HDA1*, PF 14\_0690,  $P = 0.004$ ). To further understand the role of Class II histone deacetylases and other histone modifiers in the asexual-sexual divide, we built a sub-network consisting of the core epigenetic machinery and H-L genes. We further included central components of the translational machinery (Supplementary Table 8) which is known to act as a major post-transcriptional regulator of life stage differentiation in *Plasmodium*<sup>15</sup>. This sub-network structured into distinct regulatory modules which broadly corresponded to life stages (Fig. 3). At the heart of each was a single deacetylase or demethylase surrounded by a unique set of acetylases, methylases and AP2 transcription factors. This structure is consistent with a model in which the former act as general silencers that 'wipe' the marks from the previous life stage ready for the latter to 'write' new sets of marks that coordinate specific within-stage transcriptional programmes. Based on their locations in the network, we deduce that HDAC1 and HDA1 provide the background silencing that underpins the general IDC and gametocyte transcriptional programmes, respectively, and we confirm the role of HDA2 in underpinning *var* gene regulation<sup>12</sup> (Fig. 3c, Supplementary Table 6). This, and their significant differential regulation, leads us to conclude that down-regulation of *HDAC1* and up-regulation of *HDA1* are the baseline events that entrain sexual stage development in *P. falciparum*.

Fig. 3 here

### Translational control of H-L adaptation

Overlaying the epigenetic control of asexual-sexual differentiation was a corresponding compartmentalisation of translational machinery into eukaryotic vs. prokaryotic types (Fig. 3d). The parasite's mitochondrion and apicoplast both derive from prokaryotes and perform essential metabolic functions on behalf of the parasite<sup>16</sup>. However, the genomes of these organelles house just a few genes that encode these functions, the rest having transferred to the nuclear genome over evolutionary time<sup>17</sup>. These organelle genes are translated *in situ* using prokaryotic translation machinery<sup>18</sup> which, as for the proteins that perform functions inside the organelles, is partly organelle-encoded and partly nucleus-encoded. We found that almost all prokaryotic translation machinery genes fell within the sexual stage regions of the network (Fig. 3, Fig. 4a) and that an unusually high number were H-L down-regulated (Fig. 4b). This was especially so for those encoded by the apicoplast genome (6 of 16 vs. 3 of 36 nuclear-encoded apicoplast-targeted translation machinery genes,  $P = 0.03$ , and vs. 15 of all 280 translation machinery genes,  $P < 0.001$  by hypergeometric tests), and for components not shared between the eukaryotic and prokaryotic machinery (organelle-specific ribosomal proteins, modifiers of the initial methionine, initiation and elongation factors, tRNA synthetases) but not for shared components such as the tRNAs (Fig. 3e). A clear example of this pattern is the gene encoding peptide deformylase (PDF, PF10380c,  $P = 0.002$ , Fig. 3e) which, in prokaryotes but not eukaryotes, removes formylated methionine from the start position of the newly synthesised protein. (In *Plasmodium*, this requirement

seems to apply to translation in the apicoplast but not the mitochondrion<sup>19</sup>). Overall, our data indicate that translation inside the organelles is at a premium during the sexual stages. While both organelles are active in asexual blood stages, and indeed are rendered indispensable during this phase by some of their metabolic functions<sup>20,21</sup>, this conclusion is supported by the following prior evidence: of the four metabolic functions performed by the apicoplast, three are essential during the sexual stages<sup>22-24</sup> while the fourth maintains the apicoplast and hence the other three<sup>25</sup>; most mutations in genes of mitochondrial pathways are lethal only in the sexual stages<sup>26</sup>; upon conversion from asexual blood stages to gametocytes, the parasite transfers its burden of energy production from glycolysis in the cytosol to aerobic respiration in the mitochondrion<sup>27,28</sup>; and drugs that target prokaryotic translation machinery show high efficacy against sexual as well as asexual stages<sup>29</sup>. This conclusion has vital practical significance: resistance mutations arising in blood-stage infections treated with drugs that inhibit prokaryotic translation, several of which are already in use, will not spread to the general population if these mutations are lethal during the sexual stages<sup>30</sup>.

*Fig. 4 here*

What controls the shift towards prokaryotic translation at the asexual-sexual transition? The switch is likely to fall high in the regulatory hierarchy because many organelle- and nucleus-encoded prokaryotic translation components need to be induced. A second shift in translation machinery upon switching to reproductive mode, namely, the change in species of eukaryotic ribosomal RNA (rRNA) - the RNA partner to ribosomal proteins that together form the ribosomes - when the parasite differentiates to mosquito stages, as reported previously<sup>31</sup>, would likewise require coordination of the parasite's many rRNA-encoding genes. Both these shifts may represent a mechanism to selectively translate subsets of mRNAs and thereby gain control over synthesis of the highly specialised sexual stage proteins<sup>31</sup>. Our network suggests that the forward switch (i.e., transfer from human to mosquito) involves epigenetic silencing by histone lysine demethylase, JmjC2, and the reverse switch upon transfer back involves activation of the liver stage AP2 transcription factor, AP2-L, accompanied by general silencing orchestrated by histone lysine demethylase, JmjC1 (Fig. 3c, Supplementary Table 6). The latter proposition is supported by evidence that dormant liver stages reactivate upon inhibition of histone lysine methylases<sup>32</sup>. We speculate that the remaining two of the five Class II histone deacetylases, Sir2A and Sir2B, might be additional partners in the eukaryotic-prokaryotic translation shift given that Sir2A activity represses production of rRNA<sup>33</sup> and that their encoding genes locate to network modules associated with general IDC function (maroon) and mosquito stages (khaki), respectively (Fig. 3).

In addition to the eukaryotic-prokaryotic shift, there was evidence of H-L associated transcriptional regulation within the eukaryotic arm of translation. Connectivity between translational and epigenetic machinery genes was higher than expected (Fig. 3e) with many that encode eukaryotic translation initiation and elongation factors among the most connected of all genes (Fig. 3e). These included eukaryotic elongation factor 2 (eEF2) and eukaryotic initiation factors (EIF) 2, 3, 4, 5 and 6. eEF2 and all subunits of the alpha and beta subunits of EIF2 were associated with IDC function and connected to H-L up-regulated genes more often than expected, though were not themselves significantly up-regulated. By contrast, genes for three of the nine subunits of EIF3 were significantly regulated (Fig. 3e), six of which belonged to liver stage modules, with subunit K (EIF3K), being particularly

pivotal between mosquito and liver stages (Fig. 3d). Different EIF3 subunits may thus play distinct roles in translational regulation at different life stages of *Plasmodium*. The alpha subunit of EIF2, EIF2 $\alpha$ , which prepares the ribosome for translation by loading methionine-bound tRNA to the start codon of mRNAs, is well-known as a mediator of global translational repression in *Plasmodium*. This manifests as temporary stalling of development (latency) during life-stage transitions while the parasite awaits the opportunity to transfer to a new host (i.e., in gametocytes<sup>34</sup> and sporozoites<sup>35</sup>), as well as within the IDC while the parasite undergoes the energy-costly process of DNA replication for production of daughter parasites<sup>34</sup>. Within the IDC, upon amino acid starvation leads to translational stalling via EIF2 $\alpha$  and this slows down the entire transcriptional programme<sup>36</sup>, implying feedback between the transcriptional and translational machineries. Tight co-regulation of the translational and transcriptional programmes is a feature of the IDC under normal growth and development conditions too<sup>37,38</sup>. Ribosome accumulation on 5' leaders of many mRNAs indicates there is widespread regulation of start codon recognition<sup>38</sup>. Combined, the evidence suggests that the translation initiation complex is central to coordination of the transcriptional and translation programmes throughout the parasite's life cycle. We have revealed potential upstream epigenetic regulators of this alliance (Supplementary Table 6) and propose that downstream regulators may be found among post-translational modifiers of EIFs, as in yeast<sup>39</sup>.

## Discussion

We interpret the increase in replicative vs. reproductive effort in *P. falciparum* parasites from high vs. low transmission intensity areas as the outcome of differential selection on parasite life history trade-offs under two different evolutionary models (Fig. 5). Both these models predict higher early-life investment in replication in environments where immunity and within-host competition are common, as in high transmission areas, but they differ fundamentally in the nature of the trade-off assumed to drive parasite populations to different optima. Under Model 1, increased replication is assumed to cost early life reproduction via decreased gametocyte conversion rates pre-peak. Under Model 2, the cost is higher risk of host death: selection is thus not necessarily on conversion rates *per se* since higher replication may be achieved through other means. Since our blood samples contained mixed populations of asexuals and gametocytes, and because microarrays measure relative, not absolute amounts, we cannot invoke one model over the other. We note, nonetheless, that the H-L up-regulation of within-IDC functions but not of between-IDC functions expected to affect replication rates *per se* (cell division, DNA replication), render our results more parsimonious with Model 1, although do not preclude Model 2.

*Fig. 5 here*

Since patients were recruited upon clinical presentation and excluded if they had progressed to severe malaria, our measurements took place during infection period most relevant to Models 1 and 2. We did not assay parasites in semi-immune hosts, the environment in which the selection is proposed to have occurred: nor did we measure lifetime fitness in both non-immune and semi-immune hosts. We cannot definitively conclude, therefore, that these models explain the H-L differences observed here. We note that our findings are consistent with those from our earlier experimental evolution study in laboratory mouse malaria in which we demonstrated that pre-immunity is a highly potent selection force for increased rates of pre-peak asexual replication, leading to greater lifetime reproductive output in both non-

immune and semi-immune hosts<sup>40</sup>, as predicted by Models 1 and 2. Nonetheless, alternative models – albeit less validated by theory, although well-substantiated by lifetime fitness data on *P. falciparum* in its natural setting – also predict H-L differences in response to selection later in the infection and from transmission-related factors other than immunity and in-host competition such as mosquito population density and seasonality<sup>41</sup>.

We considered whether H-L differences were due to ascertainment bias. Since gametocytes increase as the infection progresses (even if conversion rates do not<sup>42</sup>), lower expression of gametocyte genes in H populations may have resulted from capturing the infection earlier than in L populations. We think this unlikely since any prior immunity in H hosts would delay time to clinical presentation. On the other hand, the constant arrival of new infections in H populations would be expected to reduce average age of the infection at presentation and thus dilute the gametocyte signal. The gene encoding PUF1, which is dominantly expressed in sporozoites, was found at higher levels than the reference transcriptome in Kisumu and Kilifi-pre parasites but was almost undetectable in Kilifi-post and Sudan parasites (Fig. 3e), perhaps supporting such an explanation. However, this argument is countered by the fact that H-L differences were robust to adjustment for multiplicity of infection (Supplementary Fig. 2).

Our conclusion that epigenetic and translationally-mediated ‘soft-wired’ regulation of the transcriptional programme underlies the ‘hard-wired’ H-L differences discovered here does not sit well with the traditional roles assigned to these machineries, namely, to enable the parasite to adjust to the environment of its current host which varies unpredictably with each transmission cycle. Indeed, a plastic approach to reproductive investment is expected from theory<sup>2,43</sup>, and is empirically supported by observations that gametocyte conversion rates respond to environmental stressors such as drugs<sup>44</sup> and anaemia<sup>45</sup>. However, the fact that we found H-L differences in the ‘common garden’ of non-immune hosts leads us to propose that the soft-wired control of H-L differentiation described above simply reflects downstream effectors in a cascade controlled by a higher level switch that is subject to fixed-level variation in propensity to activate. Prior evidence supports the existence of hard-wired differences in rates of conversion to reproduction, namely: *in vivo*-maintained lines of the rodent malaria species, *P. chabaudi*, display stable differences in gametocyte production in a uniform host environment<sup>46</sup>; *P. falciparum* lines maintained in the same *in vitro* environment can vary in requirements for mitochondrial function<sup>47</sup>; *in vivo* transcriptional programmes of parasites infecting children in west Africa show distinct patterns that are gametocyte-related but, as in the study here, are not able to be explained by host factors<sup>48</sup>; and two west African populations of *P. falciparum* exhibit strong differentiation in a genomic region housing a gene affecting development of early gametocytes<sup>49</sup>. Only the latter specifically implicates genetic rather than epigenetic control of these differences. Based on our evidence of robustness of H-L differences to a wide range of environmental factors, we have interpreted the H-L differences as fixed, hard-wired, and therefore likely to be genetic rather than epigenetic in origin, but without deep exploration of the DNA variability associated with these differences, this conclusion cannot be definitive.

The practical implications of our findings are that when malaria control programmes lead to a reduction in transmission intensity, gametocyte production early in life will increase and asexual replication rates will decrease. Such counter-adaptation is unlikely to threaten malaria elimination, however. This is, first, because there exist other constraints that maintain low absolute levels of reproductive activity<sup>3,43,50</sup>, second, because efficacy of

transmission-reducing devices (e.g., bednets) far exceeds the capacity of the parasite to compensate via up-regulation of gametocyte production, and third, because early-life gametocyte production is small relative to total lifetime production. If death and infection length also decrease concomitantly with replication rates, as they appear to do in human and mouse malaria<sup>46</sup>, evolution of lower early-life replication in response to transmission control programmes is expected to generate even greater reductions in overall disease burden. Such a win-win situation reinforces our earlier conclusion that malaria control programmes which directly target transmission are likely to be more sustainable in the long-term than those which target asexual replication<sup>4</sup>. With this principle in mind, and armed with knowledge of the crucial molecules that dictate fitness of *P. falciparum* in its natural environment, as revealed here, new and sustainable control strategies to achieve malaria elimination can be devised.

## References

- 1 Stearns, S. C. *The evolution of life histories*. (Oxford University Press, 1992).
- 2 Reece, S. E., Ramiro, R. S. & Nussey, D. H. Plastic parasites: sophisticated strategies for survival and reproduction? *Evol Appl* **2**, 11-23 (2009).
- 3 Mideo, N. & Day, T. On the evolution of reproductive restraint in malaria. *Proc Biol Sci* **275**, 1217-1224 (2008).
- 4 Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F. Imperfect vaccines and the evolution of parasite virulence. *Nature* **414**, 751-755 (2001).
- 5 Noor, A. M. *et al.* The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-10: a spatial and temporal analysis of transmission intensity. *Lancet* (2014).
- 6 Kirk, K. & Lehane, A. M. Membrane transport in the malaria parasite and its host erythrocyte. *Biochem J* **457**, 1-18 (2014).
- 7 El Bissati, K. *et al.* The plasma membrane permease PfNT1 is essential for purine salvage in the human malaria parasite Plasmodium falciparum. *Proc Natl Acad Sci U S A* **103**, 9286-9291 (2006).
- 8 Saliba, K. J. & Kirk, K. pH regulation in the intracellular malaria parasite, Plasmodium falciparum. H(+) extrusion via a V-type H(+)-ATPase. *J Biol Chem* **274**, 33213-33219 (1999).
- 9 Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics* **9**, 1437-1448 (2010).
- 10 Kafsack, B. F. *et al.* A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 248-252 (2014).
- 11 Painter, H. J., Campbell, T. L. & Llinas, M. The Apicomplexan AP2 family: integral factors regulating Plasmodium development. *Mol Biochem Parasitol* **176**, 1-7 (2011).
- 12 Coleman, B. I. *et al.* A Plasmodium falciparum histone deacetylase regulates antigenic variation and gametocyte conversion. *Cell Host Microbe* **16**, 177-186 (2014).
- 13 Smith, J. D. *et al.* Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101-110 (1995).
- 14 Recker, M. *et al.* Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature* **429**, 555-558 (2004).
- 15 Zhang, M., Joyce, B. R., Sullivan, W. J., Jr. & Nussenzweig, V. Translational control in Plasmodium and toxoplasma parasites. *Eukaryot Cell* **12**, 161-167 (2013).



366 16 Sheiner, L., Vaidya, A. B. & McFadden, G. I. The metabolic roles of the  
367 endosymbiotic organelles of Toxoplasma and Plasmodium spp. *Curr Opin Microbiol*  
368 **16**, 452-458 (2013).

369 17 Wilson, R. J. M., Gardner, M. J., Feagin, J. E. & Williamson, D. H. Have malaria  
370 parasites three genomes? *Parasitology Today* **7**, 134-136 (1991).

371 18 Chaubey, S., Kumar, A., Singh, D. & Habib, S. The apicoplast of Plasmodium  
372 falciparum is translationally active. *Mol Microbiol* **56**, 81-89 (2005).

373 19 Pino, P. *et al.* Mitochondrial translation in absence of local tRNA aminoacylation and  
374 methionyl tRNA Met formylation in Apicomplexa. *Mol Microbiol* **76**, 706-718 (2010).

375 20 Yeh, E. & DeRisi, J. L. Chemical rescue of malaria parasites lacking an apicoplast  
376 defines organelle function in blood-stage Plasmodium falciparum. *PLoS Biol* **9**,  
377 e1001138 (2011).

378 21 Painter, H. J., Morrissey, J. M., Mather, M. W. & Vaidya, A. B. Specific role of  
379 mitochondrial electron transport in blood-stage Plasmodium falciparum. *Nature* **446**,  
380 88-91 (2007).

381 22 Ke, H. *et al.* The heme biosynthesis pathway is essential for Plasmodium falciparum  
382 development in mosquito stage but not in blood stages. *J Biol Chem* **289**, 34827-  
383 34837 (2014).

384 23 van Schaijk, B. C. *et al.* Type II fatty acid biosynthesis is essential for Plasmodium  
385 falciparum sporozoite development in the midgut of Anopheles mosquitoes. *Eukaryot*  
386 *Cell* **13**, 550-559 (2014).

387 24 Wiley, J. D. *et al.* Isoprenoid precursor biosynthesis is the essential metabolic role of  
388 the apicoplast during gametocytogenesis in Plasmodium falciparum. *Eukaryot Cell*  
389 **14**, 128-139 (2015).

390 25 Gisselberg, J. E., Dellibovi-Ragheb, T. A., Matthews, K. A., Bosch, G. & Prigge, S. T.  
391 The suf iron-sulfur cluster synthesis pathway is required for apicoplast maintenance  
392 in malaria parasites. *PLoS Pathog* **9**, e1003655 (2013).

393 26 Jacot, D., Waller, R. F., Soldati-Favre, D., MacPherson, D. A. & MacRae, J. I.  
394 Apicomplexan Energy Metabolism: Carbon Source Promiscuity and the Quiescence  
395 Hyperbole. *Trends Parasitol* **32**, 56-70 (2016).

396 27 Lang-Unnasch, N. & Murphy, A. D. Metabolic changes of the malaria parasite during  
397 the transition from the human to the mosquito host. *Annu Rev Microbiol* **52**, 561-590  
398 (1998).

399 28 MacRae, J. I. *et al.* Mitochondrial metabolism of sexual and asexual blood stages of  
400 the malaria parasite Plasmodium falciparum. *BMC Biol* **11**, 67 (2013).

401 29 Delves, M. *et al.* The activities of current antimalarial drugs on the life cycle stages of  
402 Plasmodium: a comparative study with human and rodent parasites. *PLoS Med* **9**,  
403 e1001169 (2012).

404 30 Goodman, C. D. *et al.* Parasites resistant to the antimalarial atovaquone fail to  
405 transmit by mosquitoes. *Science* **352**, 349-353 (2016).

406 31 Gunderson, J. H. *et al.* Structurally distinct, stage-specific ribosomes occur in  
407 Plasmodium. *Science* **238**, 933-937 (1987).

408 32 D  mb  le, L. *et al.* Persistence and activation of malaria hypnozoites in long-term  
409 primary hepatocyte cultures. *Nat Med* **20**, 307-312 (2014).

410 33 Mancio-Silva, L., Lopez-Rubio, J. J., Claes, A. & Scherf, A. Sir2a regulates rDNA  
411 transcription and multiplication rate in the human malaria parasite Plasmodium  
412 falciparum. *Nat Commun* **4**, 1530 (2013).

413 34 Zhang, M. *et al.* PK4, a eukaryotic initiation factor 2alpha(eIF2alpha) kinase, is  
414 essential for the development of the erythrocytic cycle of Plasmodium. *Proc Natl*  
415 *Acad Sci U S A* **109**, 3956-3961 (2012).

416 35 Zhang, M. *et al.* The Plasmodium eukaryotic initiation factor-2alpha kinase IK2  
417 controls the latency of sporozoites in the mosquito salivary glands. *J Exp Med* **207**,  
418 1465-1474 (2010).

419 36 Babbitt, S. E. *et al.* Plasmodium falciparum responds to amino acid starvation by  
420 entering into a hibernatory state. *Proc Natl Acad Sci U S A* **109**, E3278-3287 (2012).

- 37 Bunnik, E. M. *et al.* Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome Biol* **14**, R128 (2013).
- 38 Caro, F., Ahyong, V., Betegon, M. & DeRisi, J. L. Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *Elife* **3** (2014).
- 39 Beilsten-Edmands, V. *et al.* eIF2 interactions with initiator tRNA and eIF2B are regulated by post-translational modifications and conformational dynamics. *Cell Discov* **1**, 15020 (2015).
- 40 Mackinnon, M. J. & Read, A. F. Immunity promotes virulence evolution in a malaria model. *PLoS Biol* **2**, E230 (2004).
- 41 Mackinnon, M. J. & Marsh, K. The selection landscape of malaria parasites. *Science* **328**, 866-871 (2010).
- 42 Greischar, M. A., Mideo, N., Read, A. F. & Bjornstad, O. N. Quantifying Transmission Investment in Malaria Parasites. *PLoS Comput Biol* **12**, e1004718 (2016).
- 43 Greischar, M. A., Mideo, N., Read, A. F. & Bjornstad, O. N. Predicting optimal transmission investment in malaria parasites. *Evolution* **70**, 1542-1558 (2016).
- 44 Buckling, A. G. L., Crooks, L. & Read, A. F. *Plasmodium chabaudi* : effect of antimalarial drugs on gametocytogenesis. *Experimental Parasitology* **93**, 45-54 (1999).
- 45 Reece, S. E., Duncan, A. B., West, S. A. & Read, A. F. Host cell preference and variable transmission strategies in malaria parasites. *Proc Biol Sci* **272**, 511-517 (2005).
- 46 Mackinnon, M. J. & Read, A. F. Virulence in malaria: an evolutionary viewpoint. *Philos Trans R Soc Lond B Biol Sci* **359**, 965-986 (2004).
- 47 Ke, H. *et al.* Variation among *Plasmodium falciparum* strains in their reliance on mitochondrial electron transport chain function. *Eukaryot Cell* **10**, 1053-1061 (2011).
- 48 Daily, J. P. *et al.* In vivo transcriptional profiling of *Plasmodium falciparum*. *Malaria Journal* **3:30** (2004).
- 49 Mobegi, V. A. *et al.* Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol* **31**, 1490-1499 (2014).
- 50 Taylor, L. H. & Read, A. F. Why so few transmission stages? Reproductive restraint by malaria parasites. *Parasitology Today* **13**, 135-140 (1997).

## Methods

### Experimental design

With permission from the National Ethical Review Committees of Kenya (protocol SSC1292) and Sudan, parasite isolates were obtained from children below 13 years of age with three or fewer previous clinical episodes of malaria who were diagnosed with uncomplicated *P. falciparum* malaria at hospitals and dispensaries in western Kenya (2008, “Kisumu”), coastal Kenya (“Kilifi”) and eastern Sudan (2007, “Sudan”). These populations were confirmed to have high, medium and low malaria transmission intensities at the time of sampling based on malaria surveillance data from the Malaria Atlas Project<sup>51</sup>. Within the Kilifi population, parasites were sampled prior to and after a sustained reduction in transmission intensity (1994-1996, “Kilifi-pre”; 2010-2012, “Kilifi-post”). Parasite transcriptomes from ‘High’ vs. ‘Low’ transmission populations, as measured against a reference genome, were compared in three separate experiments, namely, Kisumu vs. Kilifi-post (Expt. A), Kilifi-post vs. Sudan (Expt. B) and Kilifi-pre vs. Kilifi-post (Expt. C). This yielded two geographic and one temporal High vs. Low population comparisons, each involving a common reference population (Kilifi-post) (Dataset 1, Supplementary Table 1). In a fourth experiment, data from parasites taken

from children in the medium-low transmission population with a history of many previous clinical episodes were introduced into the Kilifi 2010-2012 data (Dataset 2). These data were then analysed for within-population responsiveness to host clinical traits ('host responsiveness, HR').

### **Sample preparation**

Blood samples from children (3-5ml) were washed and placed in *in vitro* cultures within 5h of sampling in Expts. A and B, or were revived from frozen ring-stage parasites in Expt. C using standard procedures. In order to obtain data from all stages of the parasite's 48h blood stage replication cycle (the intraerythrocytic cycle, IDC), every 10h, for up to 9 times, samples of the culture were washed, stored in Trizol and their RNA extracted using phenol-chloroform, as previously<sup>52</sup>. 500 ng of total RNA from each timepoint was linearly amplified and competitively hybridised against equivalent reference material (prepared from pooled RNA from six equally spaced timepoints in a synchronised culture of a laboratory adapted field isolate, P4<sup>52</sup>) to a DNA microarray containing ~10,000 70-mer oligonucleotide probes representing ~5500 genes<sup>53</sup>. Printing of microarray slides, cDNA preparation, reference isolate preparation and hybridisations were performed independently for each experiment. Within experiments, samples were randomised across population groups (High vs. Low transmission populations) and maturation time points prior to processing and then processed in batches of eight.

### **Stage of maturity estimation**

Stage of maturity of each sample at each time point was estimated to within 2h intervals by comparison with the reference time-course transcriptome from laboratory isolate Dd2 using previous methods<sup>52,54</sup>. Time point samples for which the Pearson correlation between the reference transcriptome and sample was < 0.2, and all samples from isolates for which the estimated stages of maturity did not increase across consecutive sampling time points, were excluded from the data for further analysis.

### **Data pre-processing**

Data from poor quality spots (regression ratio < 0.6 by GenePix Pro software), probes containing a SNP (based on data from 129 genomes in PlasmoDB), probes matching multiple genes in Version 3 of the 3D7 genome, genes annotated as pseudogenes, *var*, *rif* and *stevor* (i.e., highly variable) genes, and probes with 2-fold or greater differences between the samples and the reference isolate when averaged across all samples (an indication of a deletion or polymorphism in the reference isolate, 336 probes, 45 of which code for genes involved in invasion) were excluded from the analyses.  $\log_2$  expression ratios were normalized within and between arrays using the "normexp" and "quantile" methods, respectively, in the *limma* package<sup>55</sup> in R<sup>56</sup>.

### **Pre-adjustment for stage of maturity**

Loess curves were fitted to time-course data on each gene using all probes and data from all experiments combined. Curves were also fitted to data from each experiment separately in order to allow for potential systematic differences between experiments in maturation patterns. Residuals from these curves were used for tests of High vs. Low population differences and host responsiveness (see below). The Loess curves were used to

determine the time of maximum expression,  $t_{max}$ , and periodicity using the Lomb-Scargle method implemented in R<sup>57</sup>. Genes were classified as 'aperiodic' if their P-value for cyclic behaviour was  $> 10^{-15}$ .

### Testing for High vs. Low transmission intensity population differences

For each gene, the difference between High vs. Low transmission intensity populations (denoted 'H-L contrast'), its t-value and corresponding P-value, with and without Benjamini-Hochberg adjustment for multiple testing ('nominal' and 'false discovery rate, FDR' P-values, respectively), were estimated by fitting linear models in the *limma* package<sup>55</sup> to the expression data from Dataset 1. Analyses were performed on data with and without prior adjustment for stage of maturity by Loess curve-fitting. The first model fitted (the 'population-specific adaptation' model) included fixed effects for population type (High vs. Low) within experiment. Pairwise contrasts between all possible H vs. L population pairs, both within experiment and between different experiments were constructed (Supplementary Fig. 1). Where these involved different experiments, experiment-specific values for the reference (Kilifi-post) population were subtracted in order to remove systematic differences between experiments. Statistical tests of these pairwise population contrasts are denoted 'Test 1'. A second model (the 'generalised adaptation' model) was fitted with fixed main effects for population type (High vs. Low) and experiment: this model thus estimated the average H-L effect across experiments and assumes that these are similar across the range of transmission intensities represented in the three experiments. Statistical tests under this model are denoted 'Test 2'. A third model, likewise testing for generalised adaptation, fitted a fixed effect linear covariate for population transmission intensity thus estimating H-L effects as a function of gradient in transmission intensity while ignoring any systematic experiment effects. The regression slope was evaluated for significance ('Test 3'). In all models, a random 'block' effect for isolates that allowed for correlations between multiple time points on the same parasite isolate was included<sup>58</sup>.

### Testing for host responsiveness

Each gene was further analysed for its responsiveness in expression levels to host traits (denoted 'HR') by analysing within the Kilifi-post population after augmenting it with data from semi-immune hosts (Dataset 2). The model fitted a linear covariate for one of seven traits relating to host physiological or immune status (Supplementary Table 4) and a random effect for isolate. HR analyses were similarly performed in Dataset 1 by incorporating a covariate for host trait into the models testing for H-L effects described above (Supplementary Table 5).

### Post-adjustment for gene traits

H-L contrasts were post-adjusted for systematic bias from gene traits (sexual:asexual expression ratio, time of maximum expression, single nucleotide polymorphism (SNP) density and non-synonymous:synonymous SNP ratio) by fitting a linear model with a fixed effect for each of these traits using the 'lm' function in the *stats* package in R<sup>56</sup> and then taking the residuals for downstream analyses (Supplementary Table 4). Sexual:asexual expression ratios for each gene were calculated as the average expression values in Stage II gametocytes, Stage V gametocytes and ookinetes relative to the average expression in rings, early trophozoites, late trophozoites and schizonts based on RNASeq data<sup>59</sup>. SNP

variables were based on the values from 129 isolates using data downloaded from PlasmoDB<sup>60</sup>.

## **Analysis combinations**

To test for the influence on H-L effects of data pre-adjustment, data subsetting and biological factors (host and gene properties), analyses were performed for all combinations of the following variables: data without and with parasite densities < 25,000p/μl (Dataset 1.1 and 1.2, respectively); data from three alternative length of interval around the maximum (6h, 12h or 24h either side of  $t_{max}$ , i.e., representing quarter, half or the full 48h cycle); and three alternative methods of stage of maturity adjustment using Loess curves (across experiment, within experiment and none). For each of these 18 combinations, H-L effects were estimated without and with adjustment for one of seven host traits (by incorporation as a covariate in the linear model), and with and without post-adjustment for one or all four gene traits. For analyses of HR effects, the same model-dataset combinations as above were applied but to Dataset 2 and Dataset 1.2. This gave, for each gene, a total of 171 tests for H-L effects and 126 tests for HR effects (Supplementary Tables 4 and 5).

A 'reference analysis' for H-L effects was defined as that using Dataset 1.2, all time points (i.e., interval length of 48h), pre-adjusted by Loess curves fitted across experiments, with no covariates for host traits in the model and with no post-adjustment for gene traits.

## **Significance criteria**

For individual analysis combinations, H-L genes were declared significant if the 'H-L' nominal P-value was < 0.05, unless stated otherwise. For results from across analysis combinations, H-L genes were declared significant if the geometric mean nominal P-value across all 171 combinations was < 0.05, unless stated otherwise. Robustness was defined as the proportion of individual analysis combinations with nominal P-value < 0.05. These thresholds were applied to all H-L tests between pairs of experiments and across all experiments, i.e., for Tests 1-3 but, where reported, refer to Test 2 (H-L effects averaged over the three experiments) unless stated otherwise. For downstream analyses of the regulatory network and functional enrichment, significant genes were defined as those reaching the threshold for across-analysis combinations for Test 2.

## **Numbers of observed vs. expected significant genes**

Numbers of H-L significant genes expected under the null hypothesis of no H-L population differentiation were obtained by performing the reference analysis on data in which the H-L status of each isolate within experiment was randomly reassigned prior to analysis. One-tailed tests of excess of observed vs. expected numbers of significant genes were calculated from the empirical distribution generated from 105 such data permutations.

## **Tests of overlap**

Evidence of generalised adaptation was sought by testing for excess overlap between pairs of experiments in their sets of significant genes. Since the Kilifi-post population was shared by all three experiments, H-L contrasts from Expts. A, B and C were correlated (negatively between Expts. A and B, and between B and C, positively between A and C). This led to

higher than expected discordance or concordance between experiments in H-L up- and down-regulated genes than if the experiments were independent. To account for this, significant genes were defined as those outside of the 95% confidence limits of the bivariate distribution of H-L effects for each pair of experiments (computed from the covariance matrix between the two experiments using the *SIBER* package in R<sup>61</sup>). Genes were further re-classified as up- or down-regulated according to whether they fell above or below the regression lines ('adjusted axes') obtained by fitting H-L contrasts from each experiment on the other, i.e., by their residuals after accounting for the between-experiment relationships. The proportion of significant genes in common for each pair of experiments for each of the four possible combinations of up and down-regulation in the experiment pair was calculated relative to the total number of significant genes in the quadrants defined by the adjusted axes and which fell outside the 95% confidence limits. Values were compared to the empirical distribution obtained from 105 permuted datasets generated as described above.

As a further test for consistency in H-L effects across the range of transmission intensities, overlap tests were performed on individual experiment contrasts (i.e., by Test 1) formed not involving the same populations (e.g., Kisumu-Sudan contrast vs. Kilifi-pre-Kilifi-post contrast). Overlaps tests were also performed on results from an individual experiment and those from the other two experiments combined using Test. The above tests for overlap were applied to results from the reference analysis only, after excluding the five isolates in Expt. B that were shared with other experiments (Supplementary Table 1).

Fisher's Exact tests were used to evaluate overlap between sets of H-L genes identified by different tests (Test 2 vs. Test 3), for results with vs. without host or gene trait adjustment, and for H-L vs. HR genes. These were performed on results pooled across all analysis combinations.

### **Gene set enrichment analyses**

Gene set enrichment tests were performed on sets of genes categorised into 255 'metabolic pathways' using information from the Metabolic Pathways for Malaria database<sup>62</sup> and other sources (Supplementary Table 7). These were applied to sets of H-L up-regulated and down-regulated genes separately using a categorical 'membership test' based on the hypergeometric distribution. These were applied to H-L genes based on results pooled across analysis combinations.

### **Full genome network analysis**

To identify sets of co-regulated genes and their response to H-L population status, gene correlation network analyses were conducted using WGCNA<sup>63</sup> in R following the method in ref.<sup>64</sup>. The input adjacency matrix was constructed from across-isolate gene-pair correlations computed from residuals after curve-fitting to data from Datasets 1 and 2 combined. After checking for outlier isolates by hierarchical clustering, the adjacency matrix was raised to the power of 6 to give a scale-free topology, and used to compute the topological overlap similarity matrix, from which gene clusters ('modules') were defined using hierarchical clustering. Pearson correlations between eigenvalues of these clusters and experimental factors (H-L status, host clinical variables) were calculated. For each gene, the number of connections to other genes among the strongest 5% of network links were computed within and between modules, and across the entire network. Networks were

visualised using the *igraph* package<sup>65</sup> and the Fruchterman-Reingold algorithm after pruning to the strongest 5% of linkages then excluding genes with no connections.

### Epigenetic and translation machinery sub-network

A sub-network was constructed that contained genes encoding central components of the epigenetic machinery (histone modifiers and deacetylases, histone acetylases, histone methylases, histone demethylases, AP2 transcription factors,  $N = 53$ ), translation machinery (ribosomal proteins, tRNAs, tRNA synthetases, translation initiation and elongation factors, and processors of the initial methionine involved in translation initiation,  $N = 280$ ), RNA-binding gene families implicated as regulators of gene subsets at life stage transitions (the DNA/RNA-binding family of ALBA proteins, believed to be master regulators of the life-stage specific translational programme in *Plasmodium*, and thus the translation equivalent of the ApiAP2 family ( $N = 6$ )<sup>66</sup>, and two other RNA-binding proteins from the Puf family, which have been shown to repress translation of subsets of mRNAs at the asexual-gametocyte and sporozoite-liver stage transitions<sup>67,68</sup>). In addition, H-L genes with  $P < 0.01$  ( $N = 136$ ) were included.

### Analysis of network hubs

To identify potential key transcriptional regulators ('hubs') and their relationship to metabolic function, the top 100 genes for number of connections to genes that were H-L significantly up-regulated, significantly down-regulated, significantly up- or down-regulated, and not significantly H-L regulated were identified. For each of these 400 hub genes, membership functional enrichment tests were performed on their connecting genes ('hub communities'). Each hub gene community was tested for enrichment of H-L genes, as above, and results were pooled across hubs belonging to the same network module.

Figures in circular format showing enrichment patterns and gene properties were constructed using the *circlize* package in R<sup>69</sup>.

### References for Methods

- 51 *Malaria Atlas Project*, <http://www.map.ox.ac.uk/>.
- 52 Mackinnon, M. J. *et al.* Comparative transcriptional and genomic analysis of *Plasmodium falciparum* field isolates. *PLoS Pathog* **5**, e1000644 (2009).
- 53 Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biology* **1**, 85-100 (2003).
- 54 Mok, S. *et al.* Artemisinin resistance in *Plasmodium falciparum* is associated with an altered temporal pattern of transcription. *BMC Genomics* **12**, 391 (2011).
- 55 Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds R. Gentleman *et al.*) 397-420 (Springer, 2005).
- 56 R: A Language and Environment for Statistical Computing (Vienna, Austria, 2015).
- 57 Glynn, E. F., Chen, J. & Mushegian, A. R. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics* **22**, 310-316 (2006).
- 58 Smyth, G. K., Michaud, J. & Scott, H. The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-2075 (2005).

- 683 59 Lopez-Barragan, M. J. *et al.* Directional gene expression and antisense transcripts in  
684 sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* **12**, 587  
685 (2011).  
686 60 *PlasmoDB* <http://plasmodb.org/plasmo/>.  
687 61 Jackson, A. L., Inger, R., Parnell, A. C. & Bearhop, S. Comparing isotopic niche  
688 widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in  
689 R. *J Anim Ecol* **80**, 595-602 (2011).  
690 62 *Malaria Parasite Metabolic Pathways*, <http://mpmp.huji.ac.il/home>.  
691 63 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network  
692 analysis. *BMC Bioinformatics* **9**, 559 (2008).  
693 64 *Tutorials for the WGCNA package*,  
694 [https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tut](https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html)  
695 [orials/index.html](https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html).  
696 65 Csardi, G. N., T. The igraph software package for complex network research.  
697 *InterJournal, Complex Systems* **1695** (2006).  
698 66 Vembar, S. S., Macpherson, C. R., Sismeiro, O., Coppee, J. Y. & Scherf, A. The  
699 PfAlba1 RNA-binding protein is an important regulator of translational timing in  
700 *Plasmodium falciparum* blood stages. *Genome Biol* **16**, 212 (2015).  
701 67 Muller, K., Matuschewski, K. & Silvie, O. The Puf-family RNA-binding protein Puf2  
702 controls sporozoite conversion to liver stages in the malaria parasite. *PLoS One* **6**,  
703 e19860 (2011).  
704 68 Miao, J. *et al.* The Puf-family RNA-binding protein PfPuf2 regulates sexual  
705 development and sex differentiation in the malaria parasite *Plasmodium falciparum*. *J*  
706 *Cell Sci* **123**, 1039-1049 (2010).  
707 69 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances  
708 circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).

709

## 710 **Acknowledgements**

711 The authors are grateful to the study participants and the parasite culture laboratory at the  
712 KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya, and for helpful comments from  
713 Geoff McFadden, Megan Greischar and Andrew Read, and assistance from Hagai Ginsburg  
714 with gene sets for enrichment tests. This work was supported by The Wellcome Trust (grant  
715 numbers 088634 to M.J.M., 092741 and 077176 to K.M.).

## 716 **Author contributions**

717 Data collection: MR, MAN, JJS, JNM, MME, ASA, MMK, MJM; provision of microarray  
718 materials: ZB, SM; organisation of field work: IME, JNW, KM, MJM; manuscript preparation:  
719 MJM, MR.

## 720 **Competing interests**

721 The authors declare they have no competing financial interests.

## 722 **Data availability**

723 The datasets generated during and/or analysed during the current study will be available in  
724 the Gene Expression Omnibus repository, <https://www.ncbi.nlm.nih.gov/geo/>.  
725

## 726 **Code availability**

727



Apart from that for data manipulation prior to analysis, custom code outside the R packages cited was not used.

## Permission

This paper is published with the permission of the Director of KEMRI.

## Figure legends

**Figure 1. Experimental design and analysis.** **a**, *P. falciparum* parasites were sampled from three pairs of high (H) vs. low (L) transmission populations in East Africa separated by space and time. **b**, Transcriptional profile ( $\log_2$  expression values relative to the reference isolate) for an example gene (PF11\_0513) during the 48h asexual stage maturation cycle. Dots represent observed values (orange, H: blue, L); the black line shows the Loess fitted curve. Analysis of the curve's residual values revealed highly significant up-regulation of this gene in H vs. L populations (right panel, mean and 95% confidence intervals,  $P < 0.001$  by significance Test 2, see Methods). **c**, Overview of the data analysis.

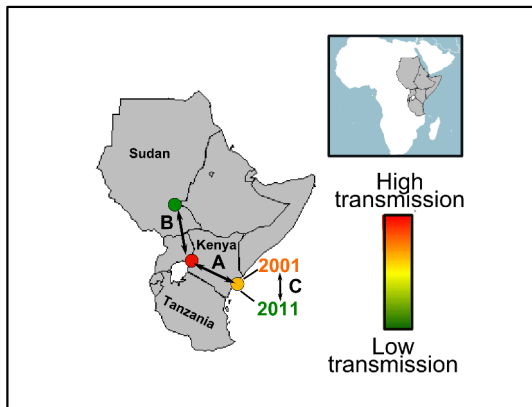
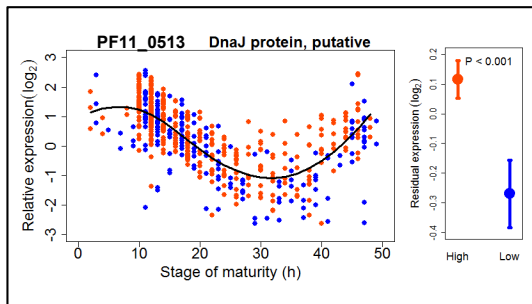
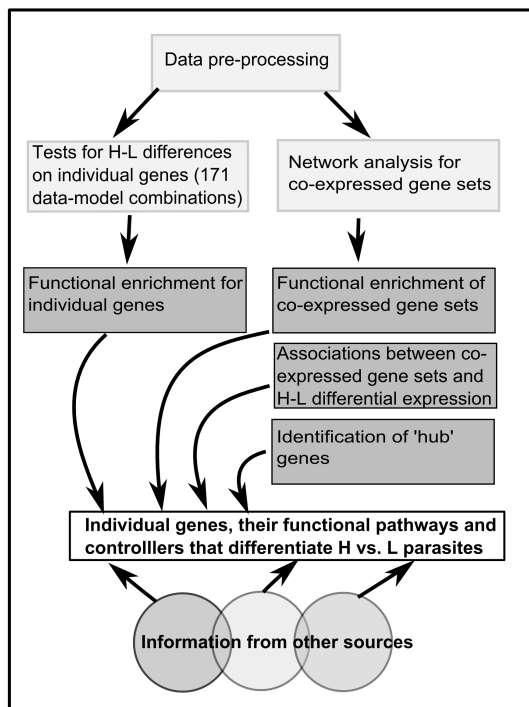
**Figure 2. Full-genome gene correlation network.** **a**, Network of co-regulated genes coloured by module. The network is built from the strongest 5% of linkages between genes and this visual representation includes only genes with at least 50 connections ( $N = 3401$  out of 4615 genes). Black curved lines approximately delineate the three life stages: asexual forms in the blood (maroon, pale green, magenta and forest green); mosquito stages (khaki, goldenrod); and liver stages (lime green, dark blue, cornflower blue, purple, pink) (Supplementary Fig. 8). Sexual stages in the blood and early in the mosquito phase occupy the periphery (turquoise, yellow, tomato). **b**, Proportion of network modules with significant H-L regulated genes ( $P < 0.05$  by significance Test 2), ordered from left to right by the module's average network connectivity, an index that negatively relates to occupancy of the network's periphery. **c**, Heatmap showing significance (orange-blue colour legend) for genes with consistent H-L differences in the three experiments (significance Test 1, first three columns) and across experiments (by significance Tests 2 and 3, last two columns). Left side colour bar indicates the network module in **a** to which the gene belongs. **d**, Functional enrichment of H-L significant genes ( $P < 0.05$  by Test 2). Bar heights on inner two tracks indicate proportion of 171 analysis combinations, scaled from 0 to 0.5, with significant enrichment ( $P < 0.05$  by membership enrichment test) of H-L genes (up-regulated, orange; down-regulated, blue) in each of 255 small functional groups (Supplementary Table 7) delineated by tick marks. Large and medium functional groups are indicated by the outer coloured track (coloured segments, adjacent labels), and inner grey-scale tracks ( $N = 32$ , black radiating labels), respectively. The third from outside track indicates the network module most strongly enriched for the small functional group. More details relating network module to functional group are given in Supplementary Fig. 9. **e**, Across experiment regression of expression ratio on transmission intensity (by significance Test 3) for the *AP2-g* gene (PFL1085w). Coloured dots and vertical bars indicate means and 95% confidence intervals for H (red colours) and L (blue colours) populations in each of the three experiments.

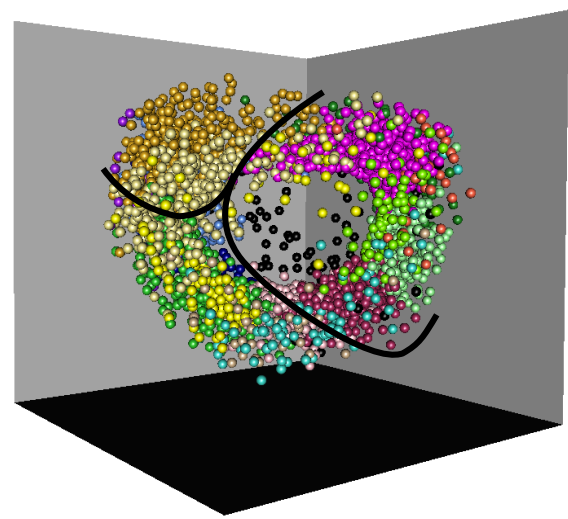
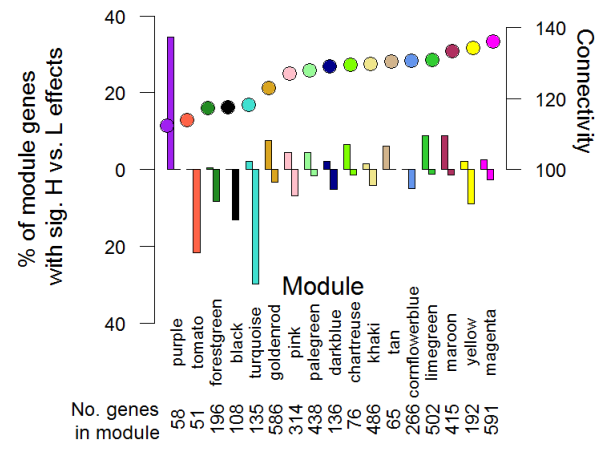
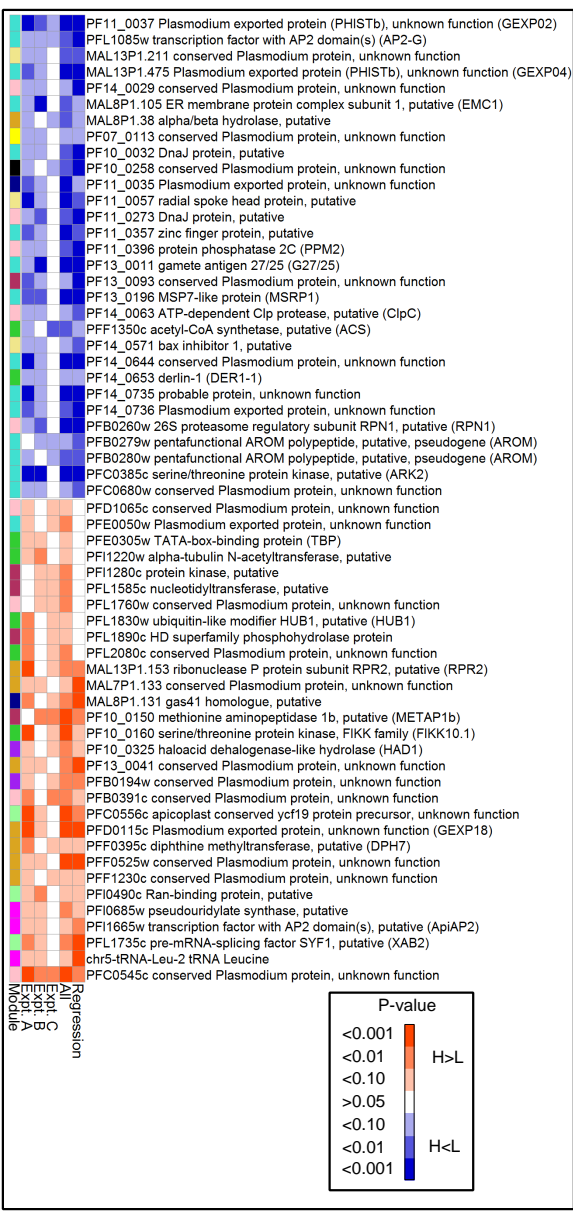
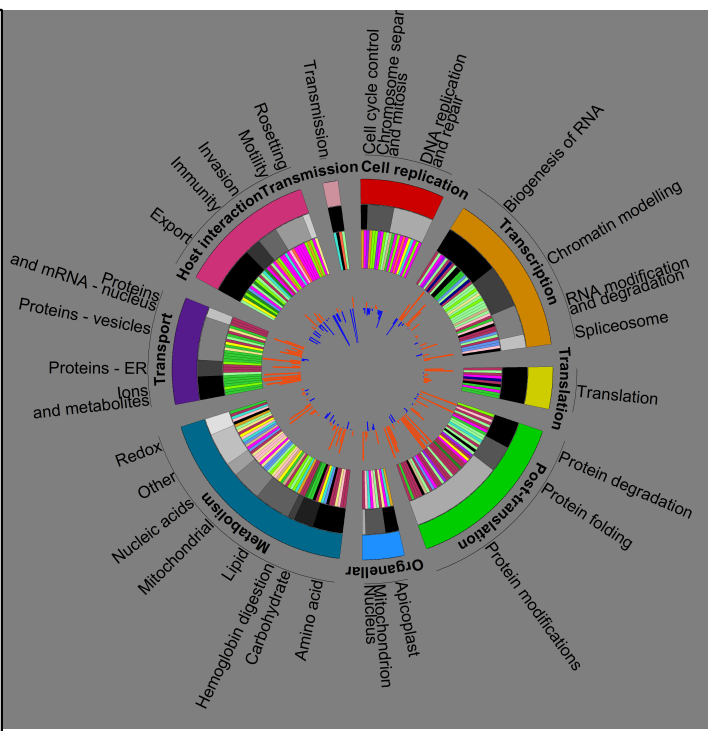
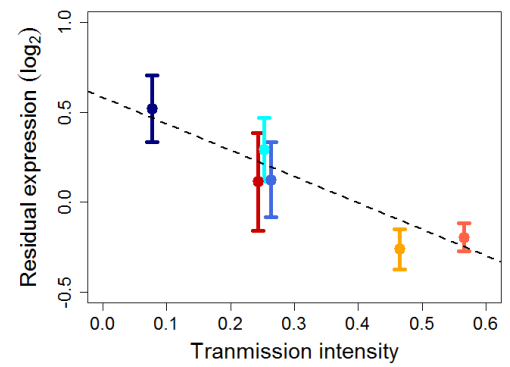
**Figure 3. Epigenetic and translation machinery networks.** Sub-network of genes coding for epigenetic and translational machinery alongside highly significant H-L regulated genes ( $P < 0.01$  by Test 2), coloured by **a**, network module; **b**, H-L significance (see legend); **c**,

classes of epigenetic machinery and **d**, translational machinery (legend in panel **e**). Key genes mentioned in the main text are labelled with their abbreviated names (Supplementary Table 8). Symbol shapes indicate cellular location (square, mitochondrion; triangle, apicoplast; circle, cytosol). **e**, Connectivity of epigenetic machinery (right sector) and translational machinery genes (left sector). Colours in outer four tracks show H-L significance (legend in panel **b**), network module, and class of epigenetic machinery and translational machinery, respectively. The next track (grey bars) shows numbers of connections to all other genes as a proportion of all genes (range 0 to 0.25). Inner three tracks show the ratio of observed to expected number of connections to H-L up-regulated genes (orange), H-L down-regulated genes (blue), epigenetic machinery genes (gold) and translational machinery genes (dark gold) on a  $\log_2$  scale, ranging from 0 to 1. Ribosomal proteins in the bottom 80% for global connectivity were excluded in **e**.

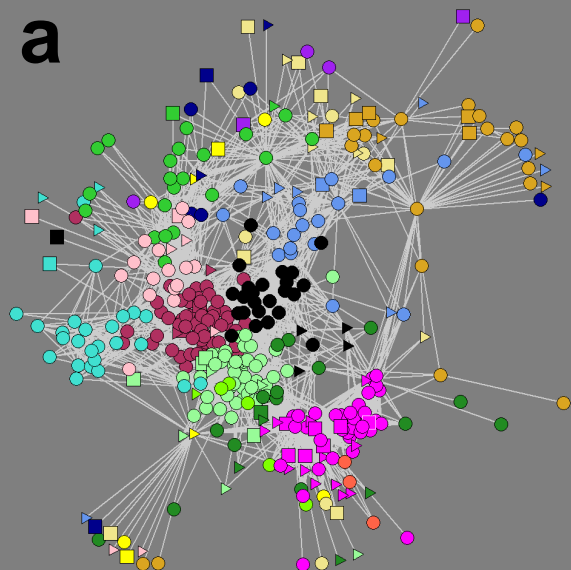
**Figure 4. Enrichment of organellar translation genes in down-regulated H-L genes and sexual stage network modules.** **a**, Distribution of classes of translation machinery genes (x-axis, Supplementary Table 8) across network modules (y-axis). Horizontal lines on the y-axis delineate asexual blood stages, sexual blood stages, mosquito stages and liver stages, from top to bottom (Supplementary Table 6). **b**, Proportions of H-L up-regulated, down-regulated and non-regulated genes by translation machinery classes. (Shared x-axis with **a**). Abbreviations: Api., apicoplast; Mit., mitochondrion; iMet, initiator methionine.

**Figure 5. Selection for environment-specific levels of replication and reproduction under two different trade-off models.** Differential selection (orange and blue vertical arrows) acting on life history trade-offs leads to different optimal levels of replication in high vs. low transmission environments (orange and blue lines and text). Under Model 1 (panel **a**), a direct trade-off between replication and reproduction, indicated by a single asterisk, arises from the fact that gametocytes, which are formed from asexual parasites, do not replicate. Theoretical models<sup>43</sup> and empirical data<sup>46</sup> have shown that low investment in reproduction (higher investment in replication) *early* in the infection leads to larger asexual populations and thus more gametocytes later in the infection, thus creating a trade-off between future and current reproduction<sup>2,43</sup>. When host immunity and within-host competition, as occur in high transmission areas, impose greater demands on replication ability, lower reproductive investment early in the infection is favoured<sup>43</sup>, thus creating different outcomes in high vs. low transmission environments (crossed blue and orange lines). **b**, Under Model 2, higher asexual replication rate is assumed to lead to more gametocytes throughout the infection (parallel orange and blue lines), and/or longer infections (persistence) but the trade-off (double asterisk) is higher risk of host death (virulence) which causes zero transmission. In high transmission areas where immunity protects hosts from dying, the cost of virulence is reduced thus allowing higher replication rates to evolve<sup>4</sup>. Model 2, like Model 1, predicts that in-host competition will further drive replication upwards in high transmission areas<sup>4</sup>. Note that the Model 2 trade-off may also operate within Model 1 (dotted arrows) and *vice versa* without altering their individual predictions of higher early asexual replication in high transmission environments. Traits that impact fitness through early-life decisions are shown in green. Fitness-trait relationships considered to be most influential under the model in question are shown with larger '+' or '-' symbols.

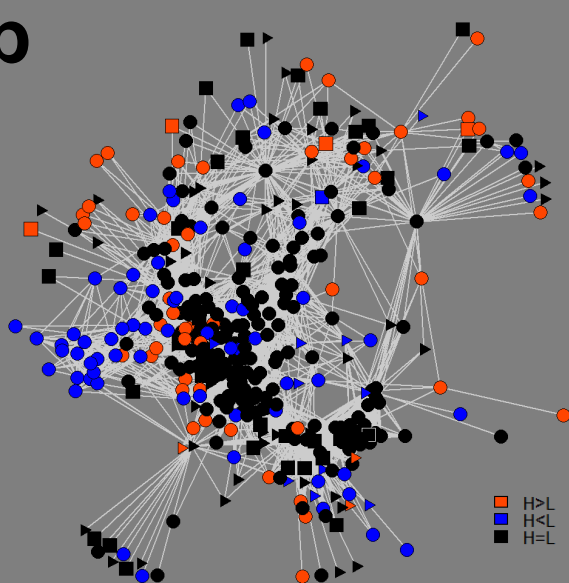
**a****b****c**

**a****b****c****d****e**

a

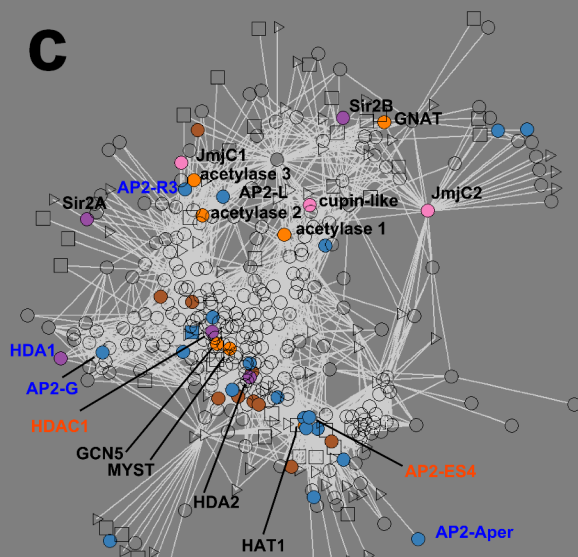


b

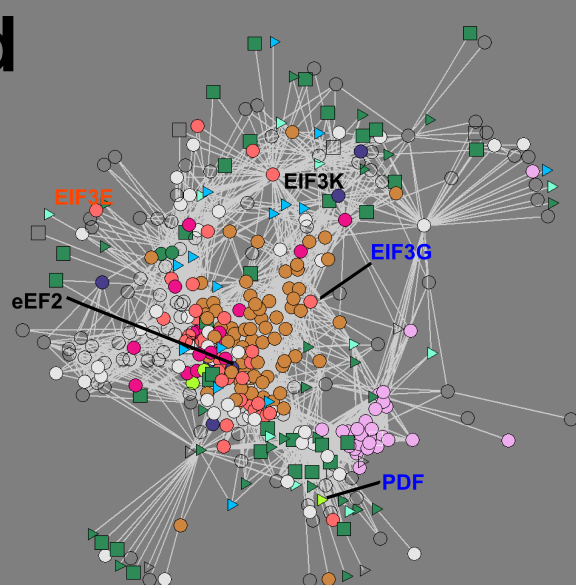


■ H>L  
■ H<L  
■ H=L

c



d



e

