Genomic adaptation underlies nutrient-driven reprogramming in the malaria-causing parasite

Qian Li[1,4], Leonie Vetter[1], Cagla Sahin[1], Ákos Végvári[2], Ulf Ribacke[1], Mats Wahlgren[1], Ola Larsson[3], Sherwin Chun-Leung Chan[3,4*]

[1] Department of Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet, Stockholm, Sweden
[2] Division of Chemistry I, Department of Medical Biochemistry and Biophysics (MBB), Karolinska Institutet, Stockholm, Sweden
[3] Department of Oncology-Pathology, Science for Life Laboratories, Karolinska Institutet, Stockholm, Sweden
[4] These authors contributed equally

*Correspondence: Sherwin Chun-Leung Chan

Abstract (Summary)

Plasmodium falciparum has an AT-rich genome with highly biased codon and amino acid usage that is not balanced by the tRNA genes repertoire. To determine whether individual tRNA genes can be differentially expressed to compensate for the tRNA-codon imbalance, we characterized the tRNA profiles of the parasite. While we found that the parasite stably maintained a biased expression of different tRNAs throughout the intraerythrocytic cycle (IDC), the tRNA-codon imbalance remains. Specifically, the requirements for a limited tRNA subset in the codons of a gene negatively correlated with its mRNA level. A detailed analysis revealed that these observations implicate host adaptation during *P. falciparum* genome evolution, whereby housekeeping genes evolved to minimize dependence on exogenous amino acids, allowing these genes to mitigate the negative effects during acute nutrient deprivation. We found that these host-adapted features can derive a decentralized nutrient-dependent regulatory programme via sensing isoleucine as a proxy for nutrient stress, which causes a reduction in aminoacylated Ile-tRNA level. This, in turn, rapidly reprogrammes the global protein output and transcriptome in a targeted fashion. Our findings provide strong evidence showing that the primary sequences of proteins underlie gene expression programmes, suggesting new perspectives on protein evolution.

Introduction

Plasmodium falciparum is a protozoan that causes malaria, which imposes a significant global health burden. The parasite genome is AT-rich (81%), which is reflected by a high frequency of AT-rich codons and biased in amino acid (AA) usage[1,2]. The number of tRNA gene copies is usually positively correlated with the codon usage in unicellular organisms[3]. In *P. falciparum, however,* tRNA genes are nonredundant. This raises the question of how the parasite compensates for its codon bias during protein synthesis. Due to extensive posttranscriptional modifications and redundancy in most organisms, profiling tRNAs requires tailored sequencing protocols which include ARM-seq, hydro-tRNAseq and mim-tRNAseq[4-8]. These protocols allow simultaneous quantifications of tRNA levels, posttranscriptional modifications and aminoacylation ("charge") status[9]. In contrast to the commonly held belief that tRNAs are stably expressed house-keeping molecules, the application of these methods has revealed dynamic tRNA expression across tissues and cell lines[10-12]. Dynamic alterations in the levels of individual tRNAs may modulate protein synthesis by affecting translation elongation. Although rarely explored, this regulatory mechanism has been implicated in cell-fate decisions and pathological conditions[11,13,14]. Intriguingly, regulation of tRNA expression has been reported with changes in growth conditions and stress levels[15-17]. In *P. falciparum*, nutrient stress induces Asp-tRNA cleavage, which may promote gametocytogenesis[18], while MAF1-dependent inhibition of tRNA transcription is essential for mounting an effective response to AA starvation[19]. In addition to these examples, stress sensing and downstream responses in *P. falciparum* remain poorly understood, in part due to the absence of the conserved mTOR pathway[20], which integrates a multitude of intracellular and extracellular cues in other organisms[21]. Moreover, although AA and glucose starvation can be sensed in *P. falciparum* via the conserved GCN2-dependent phosphorylation of eIF2α and a genus-specific serine/ threonine kinase, KIN, respectively[22-24], AA starvation is still effectively sensed in the absence of GCN2 or when only a nonphopshorylatable eIF2α variant is expressed[23]. This suggests additional stress sensing mechanisms in *P. falciparum.* We therefore sought to understand how translation is regulated in *P. falciparum* under nutrient stress.

Results

P. falciparum maintains an imbalanced tRNA-codon profile throughout the IDC

The lack of correlation between codon usage and tRNA gene copy number in *P. falciparum* suggests a disparity in tRNA availability and demand (highlighted by the demand for Asn-GTT~14.41% vs Pro-CGG~0.09%)[25] that is not seen in *Toxoplasma gondii*, a related apicomplexan (Extended Data Fig. 1). To assess whether tRNA dynamics and/or expression may compensate for this low correlation, we applied an improved TGIRT-seq protocol to profile mature tRNAs (the 45 nuclear tRNAs are referred to as the anti-codon pool or tRNA supply) throughout the IDC (Ring, 8-12 hours post invasion (h.p.i.); Trophozoite, 24-28 h.p.i.; Schizont, 40-44 h.p.i.) with a high mapping ratio (Extended Data Fig. 2a-b)[8]. Approximately 80% of the RNA sequencing reads mapped to tRNA genes retained full 5' to 3' coverage (Extended Data Fig. 2c), with only a small fraction mapped to host tRNAs (from 8.7% in Ring to 0.85% in Schizont) (Extended Data Fig. 3a). These profiles showed substantial tRNA abundance variation among tRNA species (Fig. 1a). In particular, all Ser-tRNAs, which all carry an A-box element that diverges from the consensus sequence (Extended Data Fig. 3b)[27], appeared to be highly abundant (28.7-35.4% of the anti-codon pools). However, in contrast to the cyclic pattern of the coding transcripts[28], minimal differences in the anticodon pools were found across stages (Fig. 1a). Furthermore, the discrepancy in charged tRNA ratio throughout the IDC was minimal, so that a relatively high uncharged-to-charged tRNA ratio was maintained (Fig. 1b, Extended Data Fig. 4). We also quantified fluctuations in mRNA expression-weighted tRNA usage (referred to as the codon pool or tRNA demand). The results

132 revealed that despite stage-dependent transcriptions (Extended Data Fig. 3c), the codon pools
133 exhibited limited fluctuation (Fig. 1c, Extended Data Fig. 3d). As pools of both anticodons and
134 codons were relatively constant, the codon composition of a transcript determines its
135 accessibility to the required anticodon pool throughout the IDC. Surprisingly, the anticodon
136 pools correlated poorly with the codon pools, which allowed us to identify a subset of tRNAs
137 that were either under- or overrepresented (Fig. 1d). Since isoacceptors for a specific AA are
138 often found concertedly to be under- or overrepresented, this observation suggested that AA
139 composition, in addition to codon usage, may influence translation efficiency of a transcript.
140
141 **tRNA modification profiles exhibit limited variations throughout the IDC**
142 tRNA posttranscriptional modifications can directly and indirectly affect translation[16,29]. We
143 assessed whether these modifications can be dynamically modulated throughout the IDC
144 despite the stable anticodon pool. Some modifications frequently cause position-specific
145 mutations[6], and these mutation signatures allow *a priori* inference about the modification status
146 at reported tRNA-modification sites[5,6,30]. Using this strategy, we generated a tRNA-
147 modification atlas including m^1A on A9 and A58, m^1G on G9 and G37, $m^{2,2}G$ on G26, m^3C on
148 C32, and inosine on A34 (I34) in *P. falciparum* (Extended Data Fig. 5a,c & 6a). As a validation
149 for the inferences, m^1A, m^1G and m^3C were found to be sensitive to *E. coli* AlkB demethylase
150 treatment, as previously reported (Extended Data Fig. 6b)[5,6,8]. Interestingly, we observed A59
151 mutations in four Leu-tRNAs and mutations at the discriminator U73 base in all Thr- and Lys-
152 tRNAs (Extended Data Fig. 5a,c). Although identities of these signatures cannot be inferred, as
153 no relevant annotations have been deposited in MODOMICS[31], the A59 signatures were
154 sensitive to demethylase treatment, suggesting that A59 is methylated (Extended Data Fig. 6b).
155 While the same types of modification are often found on isoacceptors for a specific AA, some
156 isoacceptors exhibit distinct modification patterns among themselves (Extended Data Fig. 5a;
157 e.g. m^1G37 in Ile-tRNAs). Modifications on the anticodon loop (C32 and G37), which can
158 directly influence translation dynamics, generally exhibit lower stoichiometry (Extended Data
159 Fig. 5a)[16,32,33].
160 Furthermore, we performed bisulfite (BS) conversion to estimate the abundance of m^5C
161 (Extended Data Fig. 5b). Positions with heavily deposited m^5C included the conserved C38 of
162 Asp-GUC and the variable loops of most tRNAs. Clusters of m^5C were observed at low
163 stoichiometry on both sides of the T-stem of many tRNAs (Extended Data Fig. 5b,c).
164 Meanwhile, m^1G37 is the only modification detected on apicoplast tRNAs (Extended Data
165 Fig.7a). However, a frequent truncation at the wobble position of one Met-tRNA strongly
166 suggests K^2C34 modification, which would convert the codon specificity from AUG (Met) to
167 AUA (Ile) (Extended Data Fig. 7b-d)[34,35].
168 Notably, we only observed a modest global reduction of modifications in trophozoites
169 (Extended Data Fig. 6c). To clarify this finding in more detail, we specifically examined Tyr-
170 GUA. Because Tyr-GUA is the only tRNA that splices an 11-nt intron from pre-tRNA, the ratio
171 of unspliced-to-spliced Tyr-GUA reads was used to estimate nascent tRNA transcription. As
172 expected from their higher transcriptional activity, trophozoites exhibited higher levels of
173 unspliced Tyr-GUA than other stages (Extended Data Fig. 6d). Furthermore, the unspliced
174 reads harbored less abundant modifications at all sites, with m^1G37 completely absent
175 (Extended Data Fig. 6d). Therefore, the global reduction in tRNA modification abundance is
176 likely a result of the relatively high abundance of pre-tRNAs in trophozoites, suggesting the
177 lack of systemic modulation of these modifications during stage progression. Together, these
178 findings support the notion that the codon composition of a transcript can primarily affect
179 translation efficiency throughout the IDC.
180

Highly expressed genes minimize their dependence on AAs with lower usage frequency in host haemoglobin

To further explore the relationship between the codon and the anticodon pools, we sought to identify sources of variation in tRNA demand across genes. To this end, we used principal component analysis with data describing the relative demand for each tRNA across all transcripts (i.e. after normalizing the tRNA demand per transcript). The resulting PCA suggested that nine tRNAs deviated in principal component 1 (PC1; which explains 80.4% total variance; Fig. 2a) and thus showed the greatest frequency variation across transcripts (among the highest in combined contribution in PC1 and PC2). Notably, their corresponding codons account for a major fraction of the codon pool, but the supply of these tRNAs was underrepresented in the anticodon pool (Fig. 2b; i.e. they belong to the "underrepresented" tRNA group identified above). As the supply of these tRNAs was relatively restricted, variation in the transcripts' demand of them may affect translation efficiency. Interestingly, genes with the highest mRNA levels was associated with a lower demand for six of these tRNAs (Asn-GUU, Ile-AAU, Ile-UAU, Lys-UUU, Phe-GAA and Tyr-GUA) compared to genes with the lowest mRNA levels (Fig. 2c,d). Moreover, transcripts encoding highly abundant ribosomal proteins show reduced demand for these tRNAs, except for Lys-UUU (Fig. 2c). Finally, the demand for these tRNAs was negatively correlated with the gene expression throughout the IDC (Extended Data Fig. 8a,b). Accordingly, to bolster their translation, highly expressed genes may have evolved to avoid competing for these decoding-limiting tRNAs. Perplexingly, five of these tRNAs exclusively decode asparagine, isoleucine, phenylalanine and tyrosine. Concomitantly, the AA composition of proteins translated from these highly abundant transcripts is low in these AAs. Consistent with this finding, the frequencies of these AAs are negatively correlated with mRNA levels (Extended Data Fig. 8c), suggesting the possibility that the AA composition influences translational efficiency.

In the IDC, *P. falciparum* digests host haemoglobin (HB) to acquire most of the AAs needed for growth[36]. Human HB, however, lacks isoleucine (Ile), which the parasite takes up from the external milieu[23]. Intriguingly, the usage frequency of the two Ile-tRNAs was negatively correlated with the mRNA level, and among all AAs, Ile AA frequency showed the strongest negative correlation with mRNA level (Fig. 2c,e, Extended Data Fig. 8a-c). We therefore hypothesized that highly expressed genes in *P. falciparum* may have evolved to reduce their dependence on exogenous AA as a feature of host-adaptation, which allows their expression to be unperturbed during external nutrient deprivation. To test this hypothesis, we compared the AA demand in the codon pools to the AA composition of HB (considered a stable endogenous AA supplier). This revealed that the AAs showing the strongest negative correlation between demand and mRNA level (Tyr [Y], Ile [I] and Asn [N] in Extended Data Fig. 8c) were inadequately supplied by HB (Fig. 2f; Y, I and N are indicated in red and fall below the identity line). Furthermore, the AA composition encoded by genes with the highest mRNA levels showed higher Spearman correlation to the AA composition of HB compared to genes with the lowest mRNA levels (Fig. 2g), a pattern that was generalizable to all the coding genes (Fig. 2h). These analyses support that highly expressed transcripts encode an AA composition which corresponds well with the HB AA composition and the proteins encoded by these transcripts are likely to be more subsistent under nutrient stress, during which AAs supply will increasingly depends on AAs released from HB digestion. Accordingly, this intricate interplay between AA availability and demand may facilitate translation of mRNA subsets in response to exogenous nutrient stress by the parasite.

Isoleucine availability serves as a proxy for external nutrient stress

We next sought to study whether AA availability under stress conditions may selectively control protein synthesis. To this end, we first assessed the tRNA profile after acute depletion

231 of AA in the medium, which forced the parasites to acquire AAs only from HB digestion. Six
232 hours of AA depletion in late trophozoites (32-36 hpi) resulted in a reduction in the level of
233 charged Ile-tRNAs (Ile-AAU -53% and Ile-UAU -19%) (Fig. 3a), but did not induce significant
234 changes in the composition or modification of the anticodon pool (Extended Data Fig. 9a,b).
235 This decrease was milder than might be expected by the complete lack of exogenous Ile,
236 suggesting that Ile was recycled from the degradation of endogenous proteins. However, the
237 level of charged Ile-tRNAs was not reduced if elongation was also blockaded by halofuginone,
238 which is a prolyl tRNA synthetase inhibitor that abolishes aminoacylation of all Pro-tRNAs
239 (>85%) (Fig. 3a)[37]. This is consistent with depletion of charged Ile-tRNAs upon an unbalanced
240 demand and supply of Ile during translation elongation. Interestingly, the reduction of charged
241 Ile-AAU, which is the preferred Ile-tRNA used by highly expressed genes (Extended Data Fig.
242 9c), was more pronounced than that of Ile-UAU (Fig. 3a). The higher depletion rate of Ile-
243 AAU may imply that highly expressed genes were still actively, and might have been
244 selectively, translated during starvation. Accordingly, our data consolidate isoleucine as an
245 essential exogenous AA for *P. falciparum* development and that isoleucine availability may be
246 key to nutrient-sensing in this parasite[23,38].
247
248 **Stage progression and parasite multiplication rates were sensitive to the isoleucine level**
249 Next, we sought to examine biological processes that may be affected by Ile withdrawal (such
250 as that during deprivation of the exogenous AA supply). Delay in stage progression is
251 commonly observed upon prolonged AA starvation; however, the parasite can reportedly
252 recover and resume normal progression upon nutrient resupply, reminiscing a "hibernatory
253 state"[23,38]. We reaffirmed the stage-delay phenotype even after a short duration of AA depletion
254 (6 hours, Extended Data Fig. 10). Notably, we found that genes with upregulated mRNA
255 expression during the ring-to-trophozoite and the trophozoite-to-schizont transitions showed
256 higher average Ile-encoding content than genes with downregulated mRNA expression, while
257 the opposite pattern was observed when the schizont-to-ring transition was analysed (Fig. 3b,
258 Extended Data Fig. 11a,b). Furthermore, genes with mRNA levels that were the most variable
259 throughout the IDC encoded a higher Ile content than genes with mRNA expression showing
260 the least variation between stages (Extended Data Fig. 11c). The usage discrepancy, however,
261 are not specific for isoleucine, since leucine, which exhibits biochemical properties similar to
262 those of Ile, displays similar trends. Yet overall, this analysis provides a rationale for the stage
263 delay phenotype and the higher sensitivity to Ile withdrawal reported in pre-S phase compared
264 to post-S phase parasites[38].
265 Meanwhile, stress parasites consistently multiplied less during the immediate invasion cycle
266 than the unstressed parasites. We observed the differences despite the stressed parasites had
267 fully recovered to normal progression. Furthermore, Ile supplementation during the acute stress
268 period could restore the multiplication rate in a dose-dependent manner (Fig. 3c). Specifically,
269 a reduction in multiplication rate was observable only under starvation conditions in the late
270 trophozoite stage, but not in the early stage (8-12 hours) (Fig. 3c). These combined to suggest
271 trophozoites-specific biological processes were affected by acute AA depletion, which led to
272 reduced proliferation and failure to achieve full recovery.
273 To explore the underlying reasons for these effects, we performed a quantitative proteomic
274 analysis and found that 56 out of the 78 downregulated proteins (56%) in the starved late
275 trophozoites were merozoite-related proteins, which mostly associate to membrane and
276 organellar compartments, particularly the inner membrane complex (Fig. 3d, Supplementary
277 table 1). Interestingly, gene ontology (GO) analysis of the 5% genes with the highest Ile content
278 (Ile-rich, "IR") revealed enrichment of lipid metabolic pathways (Fig 3e). In contrast, genes
279 with the highest leucine content, which is also decoded by AT-rich codons, show no pathway
280 enrichment. Since lipid metabolic pathways are fundamental to membrane biogenesis needed

281 for merozoite production during schizogony, AA depletion, which abrogates the Ile supply, can
282 hamper the efficient translation of IR transcripts and eventually caused a reduction in
283 merozoites number and multiplication rate. Strikingly, >80% (43/53) of conserved IR genes
284 (defined as having >100 orthologues) encode an Ile content that is >3 standard deviations higher
285 than the mean Ile content in the respective orthologue groups (Fig. 3f)[39], suggesting that the
286 high Ile content encoded by IR genes is unlikely due to functional conservation in their proteins,
287 because if this is the cause, high Ile content would have been universally retained in their
288 orthologues, as is the case for genes encoding leucine-rich proteins (Fig. 3f; only 3/55 genes
289 have a leucine content >3 S.D. than the mean). The analysis provides a strong basis that the
290 high Ile content encoded by IR genes may have evolved uniquely in *P. falciparum* as an
291 adaptation.
292
293 **Stress-mediated reprogramming is defined by isoleucine usage**
294 To assess the impact of AA starvation on translation, we performed polysome-profiling, which
295 involves fractionation of mRNA on the basis of its ribosome-association. AA starvation led to
296 a reduction in mRNA associated with multiple ribosomes (polysomes) with a concomitant
297 increase in the 80S monosome peak (Fig. 4a). This profile is commonly observed in eukaryotes
298 under stress conditions and reflects the global suppression of translation initiation due to eIF2α
299 phosphorylation[40,41]. To determine how individual mRNAs are regulated upon AA-starvation,
300 we first performed RNA-sequencing to quantify polysome-associated mRNA. As 6h of AA
301 starvation led to a reduction in charged Ile-tRNAs, we compared Ile usage between the 247 and
302 152 transcripts that showed increased and decreased polysome-association, respectively,
303 following AA starvation (Fig. 4b; FDR<0.05 and $|\log_2$ fold-change$| > 1$). Indicative of increased
304 frequency of stalled Ile codon translation during elongation, transcripts with increased
305 polysome association showed a higher Ile demand than those with decreased polysome
306 association (Fig. 4c). Consistently, whereas changes in polysome-associated mRNA and
307 protein levels showed a substantial correlation for transcripts with decreased polysome
308 association (the larger subset with a $\log_2$ fold-change <0 for polysome-associated mRNA), the
309 corresponding correlation was lower for transcripts with increased polysome association (Fig.
310 4d). Therefore, increased ribosome stallings, but not translation efficiency, likely explains the
311 increase in polysome association.
312 The Kozak context of a start codon is a major determinant of the translation rate. Surprisingly,
313 the 247 transcripts with increased polysome association showed a Kozak contexts than the 152
314 with decreased polysome association (Fig. 4e, Extended Data Fig. 12a)[42]. Ribosome collisions
315 driven by excessive ribosome stalling can trigger the conserved No-Go-Decay (NGD) pathway,
316 leading to degradation of transcripts with collided ribosomes[43]. Therefore, transcripts with a
317 higher demand for Ile-tRNAs may have evolved to prevent ribosome collision by acquiring
318 lower translation initiation potential and thus escape NGD surveillance[44]. Hypothetically, this
319 would permit Ile-dependent stabilization of transcripts, which may then contribute to global
320 transcriptome reprogramming. To assess this possibility, we quantified total mRNA in samples
321 collected in parallel with the polysome-associated mRNA samples described above and
322 identified 555 and 353 genes with increased and decreased mRNA levels, respectively, after
323 AA starvation (Extended Data Fig. 12b). Because AA starvation delays stage progression,
324 starved parasites were expected to be less abundant in transcripts that are typically upregulated
325 in schizonts compared to trophozoites and vice versa. We thus categorized differentially
326 expressed genes as stage dependent and stage independent (Extended Data Fig. 12c). Changes
327 in expression of stage dependent genes are expected due to delayed stage progression, whereas
328 stage-independent changes would represent transcripts associated with the stress-
329 reprogramming. Among the stage-independent changes, higher Ile usage and a weaker initiation
330 context were identified in parasites with increased mRNA levels compared to those with

331 decreased mRNA levels. In contrast, no difference in Ile usage and only a minor difference in
332 initiation context were found when stage-dependent changes were compared among these
333 groups (Fig. 4f,g, Extended Data Fig. 12d). The shared features associated with polysome-
334 associated mRNA and stage-independent total mRNA changes suggest crosstalk between
335 polysome association and mRNA levels. Indeed, a stronger correlation was found between
336 changes in polysome-associated and total mRNA among mRNAs with increased polysome
337 association compared to those with decreased polysome association (Fig. 4h, Extended Data
338 Fig. 12e). This finding supports the idea that increased polysome association stabilizes
339 transcripts. Furthermore, many transcripts were not regulated as expected due to delayed stage
340 progression (Extended Data Fig. 12c), suggesting that these transcripts may have been
341 stabilized (i.e., they were expected to be downregulated but were unchanged) or destabilized
342 (i.e., they were expected to be upregulated but were unchanged) by reprogramming.
343 Accordingly, the stabilized transcripts also showed higher Ile usage and weaker initiation
344 context than the destabilized transcripts (Extended Data Fig. 12f,g), These features are
345 associated with increased polysome association and point to a stabilizing effect due to increased
346 polysome association. Interestingly, many transcripts encoding ribosomal and proteasomal
347 proteins were found in the subset of stage-independent downregulated genes (Extended Data
348 Fig. 12h). These changes may lead to reduced proteome turnover and potentially linked to the
349 "hibernatory state" reported during prolonged AA starvation[23]. In contrast, upregulated stage-
350 independent genes were enriched in several metabolic pathways (Extended Data Fig. 12i). As
351 upregulation may depend on transcript stability but not necessarily increased protein output,
352 these changes may prime a cell for rapid metabolic reactivation after nutrient stimulation.
353 Collectively, our data suggest that transcripts can be stabilized following ribosome stalling in
354 combination with averted ribosome collision, which would otherwise trigger NGD surveillance.
355 We used anota2seq to further explore the relationship between polysome-associated and total
356 mRNA. Anota2seq identifies changes in polysome-associated mRNA occurring independent of
357 changes in total mRNA (denoted "translation"); congruent changes in total and polysome-
358 associated mRNA ("abundance"); and changes in total mRNA not reflected by corresponding
359 changes in polysome-association ("buffering")[45]. Notably, anota2seq analysis revealed many
360 buffered transcripts (Extended Data Fig. 13a). Technically, buffered transcripts with increased
361 ("buffered, mRNA up") or decreased ("buffered, mRNA down") total transcript levels
362 correspond to transcripts with relatively decreased or increased ribosome association,
363 respectively. Buffered transcripts therefore exhibit a discordance between ribosome association
364 and total mRNA levels and therefore would be expected to show differed RNA features
365 compared to those of the aforementioned transcripts with altered polysome association (Fig.
366 4b-d). Indeed, genes in the "buffered, mRNA down" subset have higher Ile usage than genes in
367 the "buffered, mRNA up" subset (Extended Data Fig. 13b), as well as having stronger Kozak
368 initiation context (Extended Data Fig. 13c). Tentatively, we reason that despite the enrichment
369 of ribosomes on Ile-codons, mRNAs that are reduced but buffered are destabilized due to the
370 use of strong initiation context, which lead to ribosome collision and NGD degradation.
371 As stalling at Ile codon is instrumental to the observed transcriptome reprogramming, we
372 believe that the expansion of homo-Ile repeats, which are potent stalling motifs, would have
373 been under purifying selection. Without functional constraint, AA repeats often appear at a
374 higher frequency than expected by chance due to replication slippage[46]. In contrast, di-Ile and
375 all successive expansions of Ile unit in *P. falciparum* genome occur at a frequency lower than
376 expected by chance, which is not the case in other hydrophobic AAs (Extended Data Fig. 14).
377 Therefore, occurrence of Ile repeats are likely under non-neutral selection, further asserting a
378 central role for Ile in stress-mediated reprogramming.
379 In summary, we proposed a mechanism in which the parasite senses nutrient stress through a
380 change in the charged-to-uncharged ratio of Ile-tRNAs. The stress-induced accumulation of

uncharged Ile-tRNAs not only induces eIF2α phosphorylation and repressed translation initiation, but also allows targeted proteome and transcriptome reprogramming due to its uniquely evolved genome (Fig. 4i).

Similar cases of adaptive evolution may be conserved in other organisms

Our analyses with *P. falciparum* reveal an evolutionary selection that influenced the protein sequences on a genome-wide scale. This selection strongly implicates a feature of metabolic adaptation to its HB dependence. We may therefore expect other *Plasmodium* spp., which also digest HB, to show a similar evolutionary selection. However, the GC-rich genome of *P.vivax*, for example, does not favor Ile codons and would apparently contradict the idea of a selection that is primarily reliant on Ile. To investigate the generalizability of the adaptation, we compared AA compositions of the components of glycolysis and the pentose phosphate pathway (PPP) to the AA composition of HB in several organisms with different genome GC-contents. Both pathways are highly conserved and use glucose-6-phosphate as input; however, glycolysis generates ATP, while PPP supplies ribose-5-phosphates for nucleic acid synthesis[47]. Therefore, although these pathways compete for a common input, glycolysis is essential for energy homeostasis, whereas PPP favours proliferation. Importantly, our results indicate that glycolysis pathway components of all the investigated *Plasmodium* and *Babesia* spp., which are RBC-resident parasites[36], show a higher correlation to host HB than the correlation shown by the components of the PPP (Fig. 5). This finding supports an adaptation that leads to the stronger harmonization of the AA usage in housekeeping glycolytic enzymes to the stable nutrient source. It implies that synthesis of the glycolytic pathway maybe prioritized under exogenous nutrient stress. Interestingly, despite being RBC-resident, *Theileria* spp. do not exhibit this discrepancy (Fig. 5). Schizongy in *Theileria* exclusively takes place in nucleated blood cells[48], but not in RBCs, which may have reduced the selection force needed to drive similar evolution. Indeed, *Babesia* and *Theileria* share a high lineage relatedness and the opposite observations between the two genera suggest that the selection is dynamic and depends on the parasitism mode.

Discussion

P. falciparum genome presents a peculiar feature in which the extreme codon and AA usage bias is not balanced by an expansion of tRNA genes. We studied this feature by simultaneously followed the dynamics of codon and anticodon pools using tailored tRNA-seq. While detection sensitivities may vary among tRNAs due to sequence-specific biases, such as differences in posttranscriptional modifications, we found that the supply of anticodon correlates poorly to the demand of the codon pool, reducing the availability of cognate tRNAs to certain codons. Although our method enabled only an assessment of the relative compositions, but not the absolute abundance, between the codon and the anticodon pools, our study concludes that *P. falciparum* has evolved to exploit the translation elongation process as an important regulatory mechanism. This is a sound strategy because the parasite encodes few regulators for transcriptional control and, compared to other eukaryotes, has longer average coding sequences to accentuate the relative importance of translation elongations in determining protein output rate[1,49]. Indeed, our previous study also concluded that the excess usage of GU wobble codons in the parasite genome serves to fine-tune but not to maximize global protein output[25].

In this study, we identified sweeping selection that shaped the evolution of *P. falciparum* proteome, characterizing it as an adaptation to the lifestyle of parasitizing the RBCs. Proteins with housekeeping function evolved to rely mainly on a stable nutrient supply provided by host haemoglobins and minimize dependency on external nutrients. Therefore, translation of these proteins can be buffered during acute nutrient deprivation. Phosphorylation of eIF2α is the only characterized response pathway to AA starvation, however, the parasite can still mount an

431 effective response and recovered from AA starvation in the absence of phosphorylatable eIF2α
432 [23,24]. Therefore, an alternative stress response mechanism has long been expected. Our findings
433 suggest that the proteome architecture itself derives such an alternative mechanism through
434 which the parasite can make a decentralized, rapid, targeted and graded response by simply
435 sensing the isoleucine level, which is a proxy for AA starvation, resulting in a dose-dependent
436 autoregulation of its multiplication rate. These features bear remarkable resemblance to a
437 market economy in human society, where efficient resource allocation is based solely on
438 demand and supply.
439 Notably, isoleucine is the third-most encoded AA in *P. falciparum* genome; given that
440 isoleucine must be acquired exogenously, why has the genome not evolved to reduce its usage
441 altogether so as to minimize the exposure to nutrient deprivation? We believe this evolutionary
442 outcome confers two additional advantages. First, branched chain AAs, including isoleucine,
443 are potent activators of mTOR, which positively regulate haemoglobin synthesis in
444 erythroblasts[50,51]. Therefore, host nutrient status indirectly affects the internal nutrient supply
445 available for the parasite with the RBCs. By sensing the same cue that regulates haemoglobin
446 synthesis, the parasite can rapidly synchronize with the host nutrient status. Supporting this
447 hypothesis, isoleucine has been shown to be the only metabolite that synchronizes parasite
448 growth with the host feed-fast cycle *in vivo*[52]. Second, when the parasite was subjected to acute
449 AA depletion, we observed only a minor reduction in the charged Ile-tRNA level, indicating
450 that a continuous supply of isoleucine was maintained through recycling of endogenous
451 proteins. Since isoleucine usage is efficiently reallocated under nutrient stress condition, a
452 proteome with high isoleucine content can thus serve as a reservoir of isoleucine and enable the
453 parasite to endure prolonged starvation.
454 The high AT content of *P. falciparum* genome is maintained by a significant excess of G:C to
455 A:T transition, prompting the speculation that such mutational bias contributes to adaptive
456 evolution[53]. Coincidentally, AAs that are negatively associated with gene expression are
457 invariably decoded by AT-rich codons. A hypothesis suggesting that the substitution and
458 insertion of these AAs may be continuously selected is compelling, because these mutations
459 can successively increase the efficiency of resource reallocation during nutrient stress. This
460 evolutionary adaptation can account for the AT-biased mutational rate, at least within the
461 coding region of the genome. However, despite similar mode of parasitism, *P. vivax* harbors a
462 GC-rich genome and the emergence of similar isoleucine-sensing mechanism, although not
463 analyzed, is less favored. *P. vivax* shows a clear preference to invade reticulocytes, which
464 retains a substantial amount of host ribosomes; therefore, *P. vivax* can likely digest these
465 proteins to acquire isoleucine and is not a strict isoleucine auxotroph. Infections with *P. vivax*
466 also often result in lower parasitemia, as parasites compete for fewer nutrients in the host and
467 are less likely to cause severe nutrient stress. Together, these factors may weaken the selection
468 force that drives the adaptive evolution seen in *P. falciparum*.
469 A significant implication of our finding is the demonstration of a global selection sweep that
470 affected the primary sequences of many proteins but may not necessarily impart a functional
471 advantage. Although functional selection and neutral drift are believed to be the main drivers
472 of protein evolution[54,55], this study suggests that many non-functional missense mutations can
473 be under positive selection for their regulatory effect, which are otherwise manifested as neural
474 drift[55]. This adds new perspectives to the understanding of genome and protein evolution, such
475 as on the thesis of adaptive mutation bias and on the frequent occurrence of low complexity
476 regions in *P. falciparum* genome[2,46,54]. This finding also draws attention to the limitation of
477 studying translation dynamics by considering only codon usage bias, since in addition to codon-
478 identity, AA-identity can also be a prominent factor evolved to influence translation.
479 While NGD pathway has not been experimentally validated in *P. falciparum* and the genome
480 does not encode most of the homologues involved in the pathway[56], our findings implicate a

481 mechanism similar to NGD in *P. falciparum*. At present, only a highly divergent non-sense
482 mediated decay pathway has been experimentally validated in *P. falciparum*[57]. Similarly, NGD
483 may also be highly divergent, since the mechanism is poorly conserved outside metazoans.
484 Further investigation on this will be needed.
485 Artemisinin resistance is closely related to an enhanced stress response mediated through
486 pathways involved in proteostasis[58,59]. Furthermore, reduced haemoglobin digestion has been
487 shown to be a resistance-associated trait[60], and parasites with this trait exhibits enhanced
488 dependency on external nutrients to compensate for fitness loss[61,62]. Accordingly, the switch in
489 nutrient acquisition from a balanced nutrient source to predominantly external source may
490 influence global translation dynamics and contribute to the resistance mechanism. It warrants
491 further study on this hypothesis.
492
493
494 # References
495

496 1 Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium
497 falciparum. *Nature* **419**, 498-511 (2002). https://doi.org:10.1038/nature01097
498 2 Singh, G. P. *et al.* Hyper-expansion of asparagines correlates with an abundance of
499 proteins with prion-like domains in Plasmodium falciparum. *Mol Biochem Parasitol*
500 **137**, 307-319 (2004). https://doi.org:10.1016/j.molbiopara.2004.05.016
501 3 Percudani, R., Pavesi, A. & Ottonello, S. Transfer RNA gene redundancy and
502 translational selection in Saccharomyces cerevisiae. *J Mol Biol* **268**, 322-330 (1997).
503 https://doi.org:10.1006/jmbi.1997.0942
504 4 Schwartz, S. & Motorin, Y. Next-generation sequencing technologies for detection of
505 modified nucleotides in RNAs. *RNA Biol* **14**, 1124-1137 (2017).
506 https://doi.org:10.1080/15476286.2016.1251543
507 5 Cozen, A. E. *et al.* ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a
508 complex landscape of modified tRNA fragments. *Nat Methods* **12**, 879-884 (2015).
509 https://doi.org:10.1038/nmeth.3508
510 6 Zheng, G. *et al.* Efficient and quantitative high-throughput tRNA sequencing. *Nat*
511 *Methods* **12**, 835-837 (2015). https://doi.org:10.1038/nmeth.3478
512 7 Gogakos, T. *et al.* Characterizing Expression and Processing of Precursor and Mature
513 Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep* **20**, 1463-1475 (2017).
514 https://doi.org:10.1016/j.celrep.2017.07.029
515 8 Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling
516 of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol Cell*
517 **81**, 1802-1815.e1807 (2021). https://doi.org:10.1016/j.molcel.2021.01.028
518 9 Evans, M. E., Clark, W. C., Zheng, G. & Pan, T. Determination of tRNA aminoacylation
519 levels by high-throughput sequencing. *Nucleic Acids Res* **45**, e133 (2017).
520 https://doi.org:10.1093/nar/gkx514
521 10 Pinkard, O., McFarland, S., Sweet, T. & Coller, J. Quantitative tRNA-sequencing
522 uncovers metazoan tissue-specific tRNA regulation. *Nat Commun* **11**, 4104 (2020).
523 https://doi.org:10.1038/s41467-020-17879-x
524 11 Gingold, H. *et al.* A dual program for translation regulation in cellular proliferation and
525 differentiation. *Cell* **158**, 1281-1292 (2014). https://doi.org:10.1016/j.cell.2014.08.011
526 12 Yu, P. *et al.* Dynamic landscapes of tRNA transcriptomes and translatomes in diverse
527 mouse tissues. *Genomics Proteomics Bioinformatics* (2022).
528 https://doi.org:10.1016/j.gpb.2022.07.006

529 13 Orioli, A. tRNA biology in the omics era: Stress signalling dynamics and cancer
530 progression. *Bioessays* **39** (2017). https://doi.org:10.1002/bies.201600158
531 14 Knight, J. R. P. *et al.* Control of translation elongation in health and disease. *Dis Model*
532 *Mech* **13** (2020). https://doi.org:10.1242/dmm.043208
533 15 Pang, Y. L., Abo, R., Levine, S. S. & Dedon, P. C. Diverse cell stresses induce unique
534 patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA
535 copy number. *Nucleic Acids Res* **42**, e170 (2014). https://doi.org:10.1093/nar/gku945
536 16 Chan, C. T. *et al.* Highly Predictive Reprogramming of tRNA Modifications Is Linked to
537 Selective Expression of Codon-Biased Genes. *Chem Res Toxicol* **28**, 978-988 (2015).
538 https://doi.org:10.1021/acs.chemrestox.5b00004
539 17 Sørensen, M. A., Fehler, A. O. & Lo Svenningsen, S. Transfer RNA instability as a stress
540 response in Escherichia coli: Rapid dynamics of the tRNA pool as a function of demand.
541 *RNA Biol* **15**, 586-593 (2018). https://doi.org:10.1080/15476286.2017.1391440
542 18 Hammam, E. *et al.* Malaria Parasite Stress Tolerance Is Regulated by DNMT2-Mediated
543 tRNA Cytosine Methylation. *mBio* **12**, e0255821 (2021).
544 https://doi.org:10.1128/mBio.02558-21
545 19 McLean, K. J. & Jacobs-Lorena, M. Plasmodium falciparum Maf1 Confers Survival upon
546 Amino Acid Starvation. *mBio* **8** (2017). https://doi.org:10.1128/mBio.02317-16
547 20 van Dam, T. J., Zwartkruis, F. J., Bos, J. L. & Snel, B. Evolution of the TOR pathway. *J Mol*
548 *Evol* **73**, 209-220 (2011). https://doi.org:10.1007/s00239-011-9469-9
549 21 Sengupta, S., Peterson, T. R. & Sabatini, D. M. Regulation of the mTOR complex 1
550 pathway by nutrients, growth factors, and stress. *Mol Cell* **40**, 310-322 (2010).
551 https://doi.org:10.1016/j.molcel.2010.09.026
552 22 Mancio-Silva, L. *et al.* Nutrient sensing modulates malaria parasite virulence. *Nature*
553 **547**, 213-216 (2017). https://doi.org:10.1038/nature23009
554 23 Babbitt, S. E. *et al.* Plasmodium falciparum responds to amino acid starvation by
555 entering into a hibernatory state. *Proc Natl Acad Sci U S A* **109**, E3278-3287 (2012).
556 https://doi.org:10.1073/pnas.1209823109
557 24 Zhang, M. *et al.* PK4, a eukaryotic initiation factor 2α(eIF2α) kinase, is essential for the
558 development of the erythrocytic cycle of Plasmodium. *Proc Natl Acad Sci U S A* **109**,
559 3956-3961 (2012). https://doi.org:10.1073/pnas.1121567109
560 25 Chan, S., Ch'ng, J. H., Wahlgren, M. & Thutkawkorapin, J. Frequent GU wobble pairings
561 reduce translation efficiency in Plasmodium falciparum. *Sci Rep* **7**, 723 (2017).
562 https://doi.org:10.1038/s41598-017-00801-9
563 26 Fuchs, R. T., Sun, Z., Zhuang, F. & Robb, G. B. Bias in ligation-based small RNA
564 sequencing library construction is determined by adaptor and RNA structure. *PLoS One*
565 **10**, e0126049 (2015). https://doi.org:10.1371/journal.pone.0126049
566 27 Marck, C. *et al.* The RNA polymerase III-dependent family of genes in
567 hemiascomycetes: comparative RNomics, decoding strategies, transcription and
568 evolutionary implications. *Nucleic Acids Res* **34**, 1816-1835 (2006).
569 https://doi.org:10.1093/nar/gkl085
570 28 Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of
571 Plasmodium falciparum. *PLoS Biol* **1**, E5 (2003).
572 https://doi.org:10.1371/journal.pbio.0000005
573 29 Zaborske, J. M. *et al.* A nutrient-driven tRNA modification alters translational fidelity
574 and genome-wide protein coding across an animal genus. *PLoS Biol* **12**, e1002015
575 (2014). https://doi.org:10.1371/journal.pbio.1002015

30 Ng, C. S. *et al.* tRNA epitranscriptomics and biased codon are linked to proteome expression in. *Mol Syst Biol* **14**, e8009 (2018). https://doi.org:10.15252/msb.20178009
31 Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**, D303-D307 (2018). https://doi.org:10.1093/nar/gkx1030
32 Masuda, I. *et al.* Loss of N^1-methylation of G37 in tRNA induces ribosome stalling and reprograms gene expression. *Elife* **10** (2021). https://doi.org:10.7554/eLife.70619
33 Hou, Y. M., Masuda, I. & Gamper, H. Codon-Specific Translation by m^1G37 Methylation of tRNA. *Front Genet* **9**, 713 (2018). https://doi.org:10.3389/fgene.2018.00713
34 Muramatsu, T. *et al.* Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**, 179-181 (1988). https://doi.org:10.1038/336179a0
35 Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189-1232 (2002). https://doi.org:10.1017/s1355838202022021
36 Sojka, D., Jalovecká, M. & Perner, J. Babesia, Theileria, Plasmodium and Hemoglobin. *Microorganisms* **10** (2022). https://doi.org:10.3390/microorganisms10081651
37 Herman, J. D. *et al.* The cytoplasmic prolyl-tRNA synthetase of the malaria parasite is a dual-stage target of febrifugine and its analogs. *Sci Transl Med* **7**, 288ra277 (2015). https://doi.org:10.1126/scitranslmed.aaa3575
38 McLean, K. J. & Jacobs-Lorena, M. The response of Plasmodium falciparum to isoleucine withdrawal is dependent on the stage of progression through the intraerythrocytic cell cycle. *Malar J* **19**, 147 (2020). https://doi.org:10.1186/s12936-020-03220-w
39 Altenhoff, A. M. *et al.* OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* **49**, D373-D379 (2021). https://doi.org:10.1093/nar/gkaa1007
40 Nagarajan, S. & Grewal, S. S. An investigation of nutrient-dependent mRNA translation in Drosophila larvae. *Biol Open* **3**, 1020-1031 (2014). https://doi.org:10.1242/bio.20149407
41 Gobet, C. *et al.* Robust landscapes of ribosome dwell times and aminoacyl-tRNAs in response to nutrient stress in liver. *Proc Natl Acad Sci U S A* **117**, 9630-9641 (2020). https://doi.org:10.1073/pnas.1918145117
42 Patakottu, B. R., Singh, P. K., Malhotra, P., Chauhan, V. S. & Patankar, S. In vivo analysis of translation initiation sites in Plasmodium falciparum. *Mol Biol Rep* **39**, 2225-2232 (2012). https://doi.org:10.1007/s11033-011-0971-3
43 Ikeuchi, K., Izawa, T. & Inada, T. Recent Progress on the Molecular Mechanism of Quality Controls Induced by Ribosome Stalling. *Front Genet* **9**, 743 (2018). https://doi.org:10.3389/fgene.2018.00743
44 Simms, C. L., Yan, L. L. & Zaher, H. S. Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Mol Cell* **68**, 361-373.e365 (2017). https://doi.org:10.1016/j.molcel.2017.08.019
45 Oertlin, C. *et al.* Generally applicable transcriptome-wide analysis of translation using anota2seq. *Nucleic Acids Res* **47**, e70 (2019). https://doi.org:10.1093/nar/gkz223
46 Radó-Trilla, N. & Albà, M. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins. *BMC Evol Biol* **12**, 155 (2012). https://doi.org:10.1186/1471-2148-12-155

623 47 Bertels, L. K., Fernández Murillo, L. & Heinisch, J. J. The Pentose Phosphate Pathway in
624 Yeasts-More Than a Poor Cousin of Glycolysis. *Biomolecules* **11** (2021).
625 https://doi.org:10.3390/biom11050725
626 48 Mans, B. J., Pienaar, R. & Latif, A. A. A review of Theileria diagnostics and epidemiology.
627 *Int J Parasitol Parasites Wildl* **4**, 104-118 (2015).
628 https://doi.org:10.1016/j.ijppaw.2014.12.006
629 49 Tiessen, A., Pérez-Rodríguez, P. & Delaye-Arredondo, L. J. Mathematical modeling and
630 comparison of protein size distribution in different plant, animal, fungal and microbial
631 species reveals a negative correlation between protein size and protein number, thus
632 providing insight into the evolution of proteomes. *BMC Res Notes* **5**, 85 (2012).
633 https://doi.org:10.1186/1756-0500-5-85
634 50 Chung, J. *et al.* The mTORC1/4E-BP pathway coordinates hemoglobin production with
635 L-leucine availability. *Sci Signal* **8**, ra34 (2015).
636 https://doi.org:10.1126/scisignal.aaa5903
637 51 Knight, Z. A., Schmidt, S. F., Birsoy, K., Tan, K. & Friedman, J. M. A critical role for
638 mTORC1 in erythropoiesis and anemia. *Elife* **3**, e01913 (2014).
639 https://doi.org:10.7554/eLife.01913
640 52 Kimberley, F. P. *et al.* Synchrony between daily rhythms of malaria parasites and hosts
641 is driven by an essential amino acid. *bioRxiv* (2021).
642 https://doi.org:10.1101/2020.08.24.264689
643 53 Hamilton, W. L. *et al.* Extreme mutation bias and high AT content in Plasmodium
644 falciparum. *Nucleic Acids Res* **45**, 1889-1901 (2017).
645 https://doi.org:10.1093/nar/gkw1259
646 54 Jayaraman, V., Toledo-Patiño, S., Noda-García, L. & Laurino, P. Mechanisms of protein
647 evolution. *Protein Sci* **31**, e4362 (2022). https://doi.org:10.1002/pro.4362
648 55 Svensson, E. I. & Berger, D. The Role of Mutation Bias in Adaptive Evolution. *Trends*
649 *Ecol Evol* **34**, 422-434 (2019). https://doi.org:10.1016/j.tree.2019.01.015
650 56 Pavlovic Djuranovic, S. *et al.* translational machinery condones polyadenosine repeats.
651 *Elife* **9** (2020). https://doi.org:10.7554/eLife.57799
652 57 McHugh , E., Bulloch , M. S., Batinovic , S., Sarna , D. K. & Ralph, S. A. A divergent
653 nonsense-mediated decay machinery in Plasmodium falciparum is inefficient and non-
654 essential. *BioRxiv* (2021). https://doi.org:10.1101/2021.04.14.439394
655 58 Zhang, M. *et al.* Inhibiting the Plasmodium eIF2α Kinase PK4 Prevents Artemisinin-
656 Induced Latency. *Cell Host Microbe* **22**, 766-776.e764 (2017).
657 https://doi.org:10.1016/j.chom.2017.11.005
658 59 Mok, S. *et al.* Drug resistance. Population transcriptomics of human malaria parasites
659 reveals the mechanism of artemisinin resistance. *Science* **347**, 431-435 (2015).
660 https://doi.org:10.1126/science.1260403
661 60 Birnbaum, J. *et al.* A Kelch13-defined endocytosis pathway mediates artemisinin
662 resistance in malaria parasites. *Science* **367**, 51-59 (2020).
663 https://doi.org:10.1126/science.aax4735
664 61 Bunditvorapoom, D. *et al.* Fitness Loss under Amino Acid Starvation in Artemisinin-
665 Resistant Plasmodium falciparum Isolates from Cambodia. *Sci Rep* **8**, 12622 (2018).
666 https://doi.org:10.1038/s41598-018-30593-5
667 62 Mesén-Ramírez, P. *et al.* The parasitophorous vacuole nutrient channel is critical for
668 drug access in malaria parasites and modulates the artemisinin resistance fitness cost.

Cell Host Microbe **29**, 1774-1787.e1779 (2021). https://doi.org:10.1016/j.chom.2021.11.002

Acknowledgement

We thank all members of the Larsson laboratory and the Ribacke laboratory for their constructive comments and feedback. We thank Pan Tao and Addgene for providing us the demethylase plasmids and E. Strandback, H.A. Korsah and T. Nyman in the Protein Science facility of Karolinska Institutet for assistance in synthesizing the recombinant proteins. Protein identification and quantification were carried out by the Proteomics Biomedicum core facility, Karolinska Institutet (https://ki.se/en/mbb/proteomics-biomedicum). The authors acknowledge the support of the Freiburg Galaxy Team: Person *X* and Björn Grüning, Bioinformatics, University of Freiburg (Germany) funded by the Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC. This work was supported by a grant from Knut and Alice Wallenberg Foundation (2017.0055.) to M.W. and U.R, a grant from the Swedish Research Council (2021-03141) to S.CL.C. and M.W, a grant from Tore Nilsons Stiftelse För Medicinsk Forskning (2017-00532) to S.CL.C., S.CL.C. was also supported by a postdoctoral fellowship grant from Svenska Sällskapet för Medicinsk Forskning, Q.L. was supported by a scholarship from Chinese Scholarship Council and a doctoral (KID #2018-01037) grant from Karolinska Institutet.

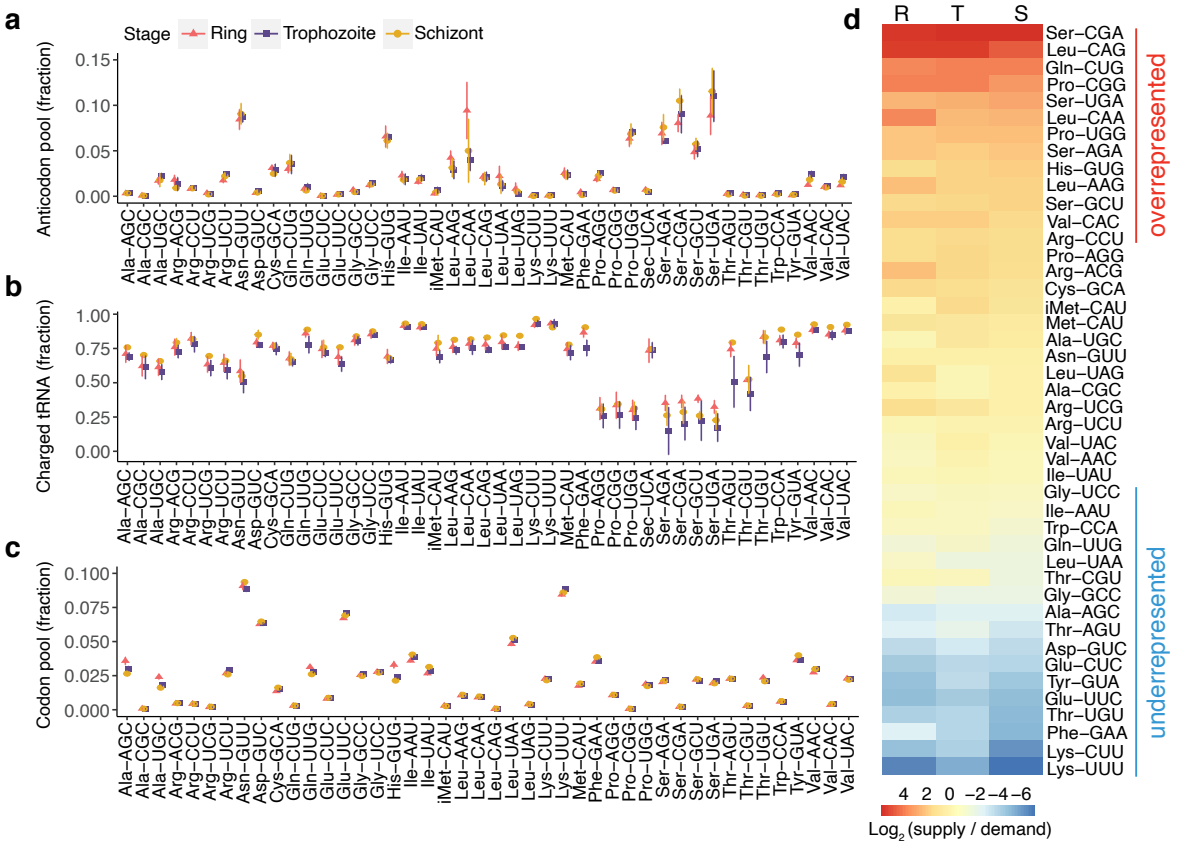


Figure 1. Maintenance of an imbalanced codon and anticodon pools throughout the IDC. **a,** The dynamics of the anticodon pool during the IDC. The fraction of each tRNA to the total aligned tRNA reads in each stage is shown. **b,** The dynamics of tRNA charging throughout the IDC. Charged tRNAs are represented by reads with 3' -CCA ends after periodate oxidation and beta-elimination reactions. **c,** The dynamics of the codon pools throughout the IDC. The codon usage frequency of the transcriptome in each stage was estimated by the codon usage of each expressed gene weighted by the expression level (transcript per million), and the fraction of codons that needed to be decoded by each tRNA is shown. **d,** Heatmap showing $\log_2$-transformed averaged ratio of supply to demand (anti-codon/codon) for all tRNAs across the IDC. R: Ring; T: Trophozoite; S: Schizont. Error bars represent the standard deviation (SD) of biological replicates. For a & c, N=3 each for the Ring, Trophozoite and Schizont stages. For b, Ring, Trophozoite: N=3, Schizont: N=2. All apicoplast tRNAs are excluded.

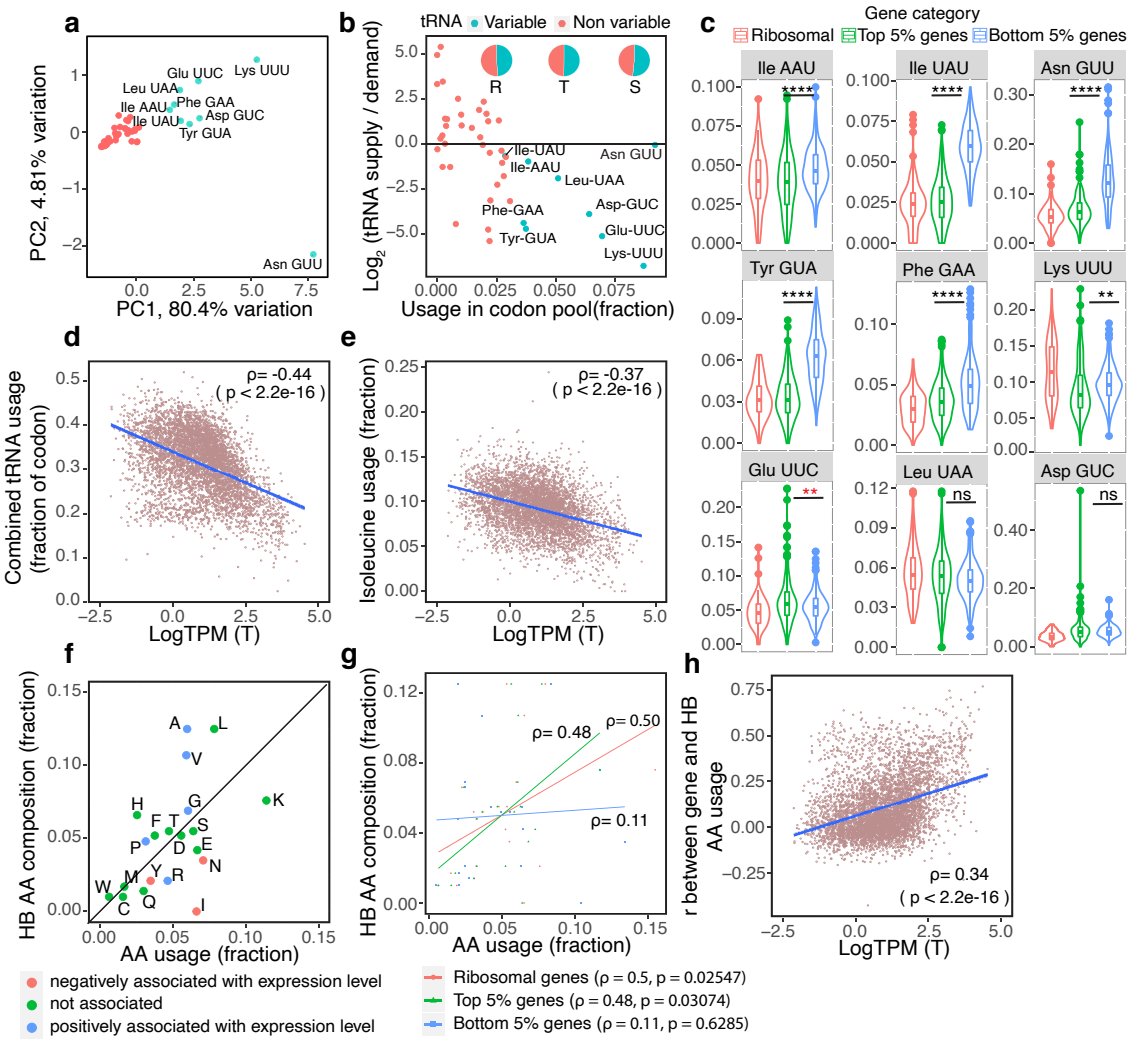


Figure 2. tRNA usage delineates gene expression level and haemoglobin dependency. **a,** Principal component analysis of tRNA usage among all the coding genes in the genome. The most variably used tRNAs among genes (cyan) were named and used for subsequent analyses. **b,** Plot showing the relationship of the usage of each tRNA in the codon pool (x-axis) and the $\log_2$-transformed ratio of supply to demand per tRNA (y-axis) in the trophozoite stage. Variable tRNAs (cyan) represent tRNAs named in (a). The line demarcates regions of underrepresentation (below the line) and overrepresentation (above the line). Pie charts show the combined usage of these tRNAs in the codon pool. **c,** tRNA usage in different gene categories. Genes were sorted by the averaged TPM values in the trophozoite stage and classified into three categories; Ribosomal: genes encoding ribosomal proteins (n=84), Top 5% genes: the 5% of the most highly expressed coding genes (n=258), Bottom 5% genes: the 5% of the least expressed coding genes (n=258). P-values of Mann-Whitney U-test between the top and bottom 5% genes are denoted by **** ($p<0.0001$); ** ($p<0.01$); ns (not significant, $p>0.05$). Glu-UUC shows an opposite trend with higher mean usage in the top 5% genes (red asterisk). **d,** Correlation of the gene expression levels in the trophozoite stage and the combined usage of Asn-GUU, Ile-AAU, Ile-UAU, Phe-GAA and Tyr-GUA tRNAs of genes. **e,** Correlation of the gene expression levels in the trophozoite stage and isoleucine usage of genes. **f,** Plot showing the relationship of AA usage frequency in the codon pool (x-axis) and in host haemoglobin (y-axis). AAs are defined as negatively associated with expression level if correlation between usage frequency in genes and gene expression level was < -0.2, not associated if $-0.2 < \rho < 0.2$, positively associated if $\rho > 0.2$. **g,** Correlation between the AA usage frequency in different gene categories and in host haemoglobin. **h,** The relationship

746 between the abundance of transcripts (x-axis) and the correlation between the AA usage
747 frequency of the transcripts and the AA composition of haemoglobin (y-axis). For all panels,
748 ρ: Spearman's rank correlation coefficients. Coding genes, n=5267.
749
750

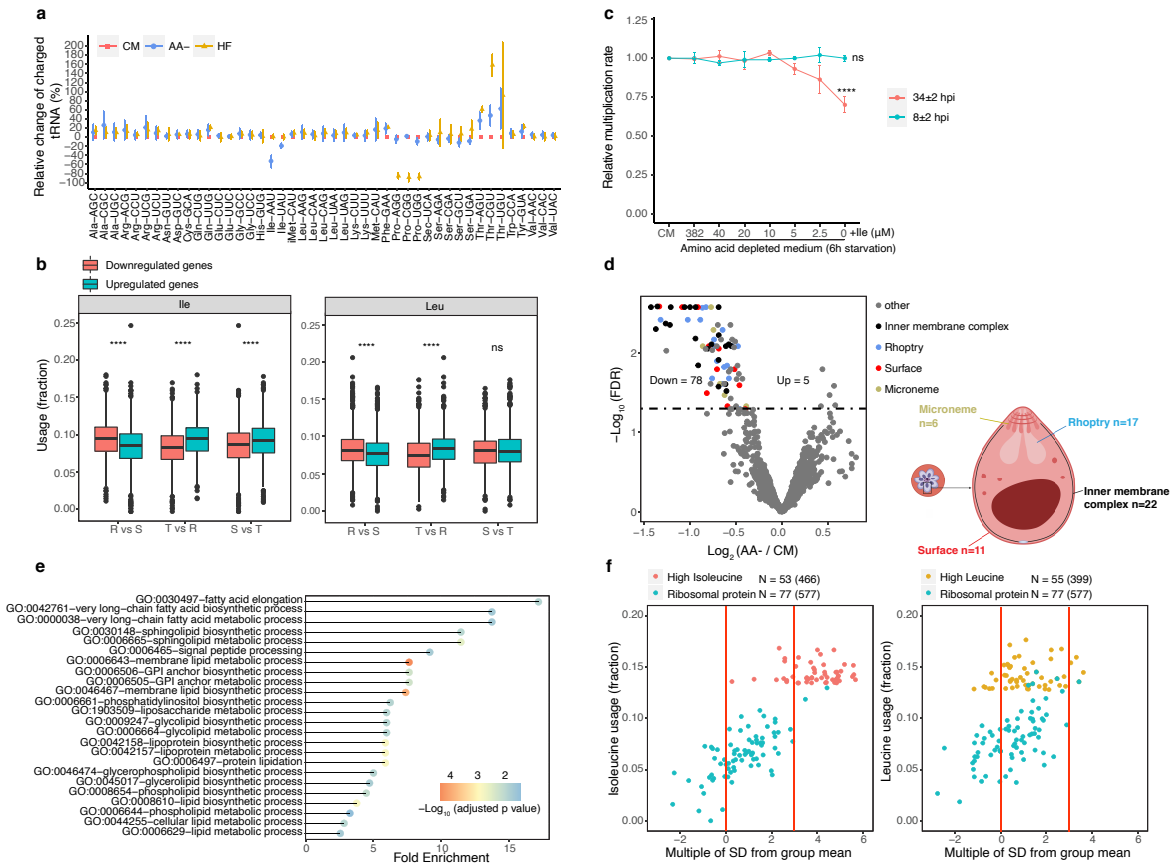


Figure 3. Isoleucine availability is sensed as a proxy of nutrient status. a, Changes in charged tRNA levels after 6 hours of treatment in late trophozoite cultures (CM: complete medium, AA-: AA depletion, HF: halofuginone (70 nM) and AA depletion). All changes are represented relative to the CM condition. Error bars represent the SD of N=3 biological replicates. **b,** Box plot showing isoleucine (left) and leucine (right) usage in up- and downregulated genes during stage progression. P-values of Mann-Whitney U-test are denoted by **** ($p<0.0001$); ns (not significant, $p>0.05$). **c,** The multiplication of parasites after 6 hours of AA depletion in the presence of varying concentrations of isoleucine. The parasites were starved from 34±2 hpi (red) or 8±2 hpi (cyan) for 6 hours and subsequently recovered by culturing in complete medium until all parasites had completed reinvasion. The multiplication rate was normalized to the complete medium (CM) group. Error bars represent the SD of N=3 biological replicates. P-values of the two-tailed unpaired t test are denoted by **** ($p<0.0001$); ns (not significant, $p>0.05$). **d,** (Left) Volcano plot showing 1146 proteins identified in all samples by LC-MS/MS after 6 hours of AA depletion. The dotted line indicates the cut-off of the FDR-adjusted p-value=0.05. 56 of the 78 downregulated proteins are associated with merozoite, and their reported localizations are indicated (right). **e,** Gene Ontology (GO) analysis of the top 5% isoleucine-rich genes (n=265). **f,** The number of SDs for the usage frequency in *P. falciparum* orthologue from the mean usage frequency of all orthologous proteins for isoleucine (left) and leucine (right). High isoleucine: top 5% of isoleucine-rich genes, high leucine: top 5% of leucine-rich genes, ribosomal protein: ribosomal protein encoding genes. Only genes with >100 orthologues were considered as functionally conserved and analyzed. Bracketed values = average orthologue counts. Red lines indicate 0 and 3 SD.

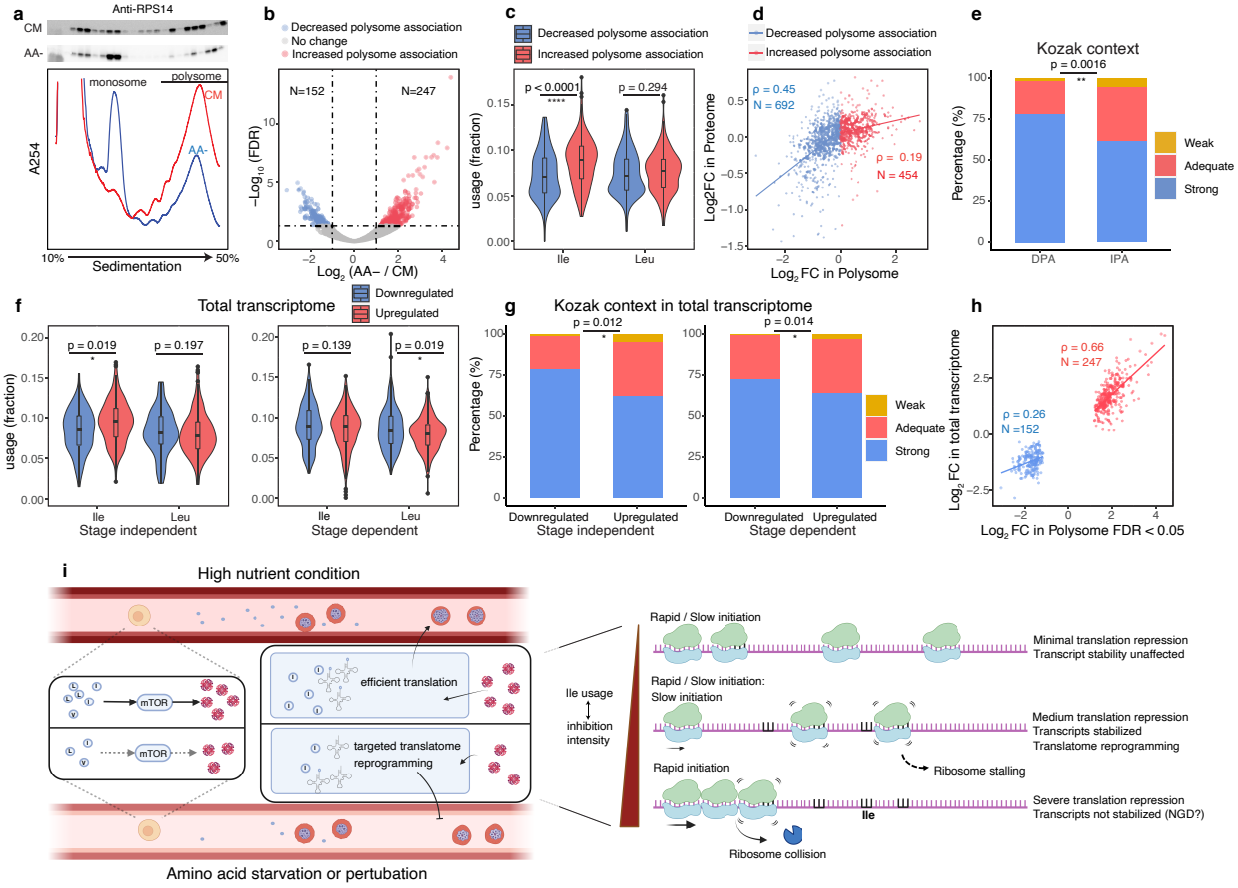


Figure 4. Polysome profiling reveals targeted reprogramming after AA starvation. **a,** (Below) Representative polysome traces of parasites growing in complete medium (CM) and AA-depleted medium (AA-) on a 10-50% sucrose gradient. Fractions containing monosomes and polysomes are indicated. (Upper) Immunoblotting of each fraction against RPS14. **b,** Differential expression analysis of the genes in the polysome fractions after AA starvation. The horizontal and vertical dotted lines show the cut-off of the FDR-adjusted p-value=0.05 and cut-off of twofold changes, respectively. **c,** The differences in isoleucine and leucine usages between genes with increased or decreased polysome association **d,** The Spearman correlation of the changes in transcript abundance in the polysome fractions and the changes in protein abundance. **e,** Kozak context of genes with decreased (DPA) or increased polysome association (IPA) **f, g,** The differences in **(f)** isoleucine and leucine usage; and **(g)** the Kozak context between genes up- and downregulated within (left) the stage independent and (right) the stage dependent transcriptome changes. Kozak context - strong: RnnATGR; adequate: RnnATGY/YnnATGR; and weak: YnnATGY (ATG= start codon. Mann-Whitney U test for the AA usage analysis and Fisher's exact test for the Kozak context analysis, * ($p<0.05$); ns (not significant, $p>0.05$). **h,** The Spearman correlation of the changes in transcript abundance in the polysome fractions and the total transcriptomes. Only transcripts with an FDR<0.05 and $|\log_2$ fold-change$| > 1$ in polysome association are shown. **i,** Proposed model of a decentralized nutrient-sensing mechanism. (Left) Nutrient deprivation due to host fasting reduces haemoglobin synthesis in erythroblasts via a dampened mTOR activation by branched chain AAs. Reduction in serum Ile levels below a threshold level causes an accumulation of uncharged Ile-tRNAs and the selective decrease in translational output of the translatome. Lipid metabolism genes are highly targeted due to generally high Ile usage, leading to lower merozoite production and potentially reduced endocytosis of host haemoglobins. (Right) The changes in ribosome loadings also reprogramme the transcriptome. Differential stabilization of transcripts depends on the interplay between translation initiation and Ile-codon-induced

802 stalling. Reprogramming in conjunction with a global reduction in translation initiation by
803 eIF2α phosphorylation drives the cell into a “hibernation state” signified by reduced proteome
804 turnover but a transcriptome primed for rapid metabolic reactivation upon the cessation of
805 nutrient deprivation.
806

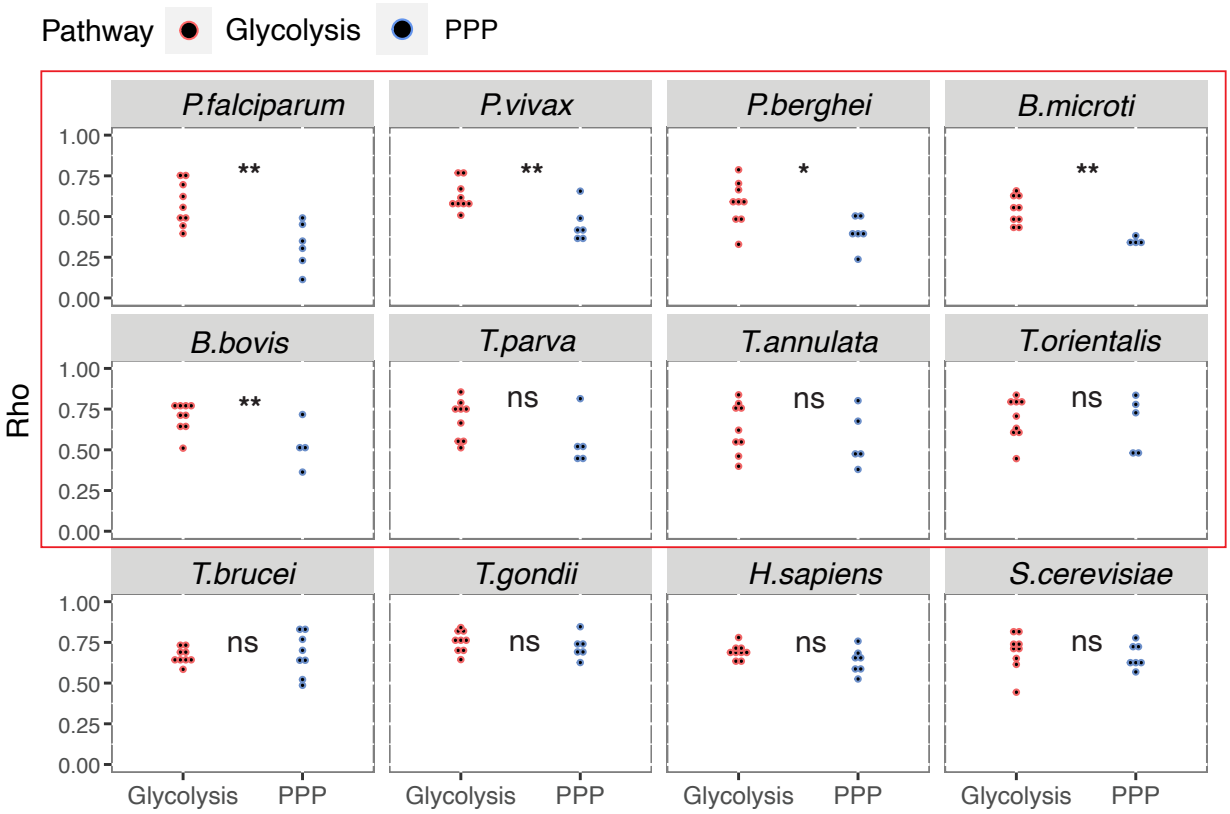


Figure 5. Reduced dependency of housekeeping genes on external nutrient sources is conserved. Spearman correlation (Rho) between the AA compositions of the host haemoglobin and that of the enzymes involved in glycolysis or the pentose phosphate pathway (PPP) in different organisms. Correlations were calculated using *Homo sapiens* haemoglobin for *P. falciparum*, *P. vivax*, *B. microti*, *T. brucei*, *T. gondii*. *H. sapiens* and *S. cerevisiae*; using *Bos Taurus* haemoglobin for *B. bovis*, *T. parva*, *T.annulata* and *T. orientalis*; and using *Mus musculus* haemoglobin for *P. berghei*. Non-parasitic multicellular (*H. sapiens*) and unicellular (*S. cerevisiae*) organisms are shown for comparison. The red box represents organisms with developmental stage within host RBCs. P-values of Mann-Whitney U-test are denoted by ** ($p<0.01$); * ($p<0.05$); ns (not significant, $p>0.05$).

Material Extension for

Genomic adaptation underlies nutrient-driven reprogramming in the malaria-causing parasite

Qian Li[1,4], Leonie Vetter[1], Cagla Sahin[1], Akos Vegvari[2], Ulf Ribacke[1], Mats Wahlgren[1], Ola Larsson[3], Sherwin Chun-Leung Chan[3,4*]

This file includes:
Materials and Methods
Extended Data Fig. 1 to 14
References

Materials and methods

Parasite Culturing

Plasmodium falciparum strain NF54 were cultured with human O+ erythrocytes in RPMI 1640 medium (Thermofisher Scientific, #42402-016) supplemented with 10% A^+ human serum, 50 μg/L gentamicin (Gibco) and 2mM L-glutamine (Hyclone). Cultures were maintained at 37°C in 5% O_2, 5% CO_2 gas mixture. Parasites were routinely synchronized by treating with 5% sorbitol solution (Sigma) for ring stage and by using percoll gradient (Cytiva) for late stage.

RNA isolation

For time series experiment, cultures were expanded to 30 ml at 3% hematocrit and ~8% parasitemia, 10 ml culture were collected for RNA extraction at each time point (8-12, 24-28 and 40-44 hpi).

For tRNA library

Small RNA (<200 nt) was isolated directly from the pellets of infected RBCs using the *mir*Vana miRNA isolation kit (Invitrogen) according to the manufacturer's instructions. 600 μl lysis buffer was added to 150 μl infected RBCs and the cell suspension was vortexed vigorously to completely lyse the cells. 1/10 volume of miRNA Homogenate Additive was added to the cell lysates and the samples were incubated 10min on ice. After addition of one v/v ratio of Acid-Phenol: Chloroform to the lysate, the samples were vortexed vigorously and centrifuged at 10,000 g for 5 min and the aqueous phase was transferred to a new tube. 1/3 volume of 100% ethanol was added to the aqueous phase and subsequently passed through the column. The flow through was mixed with 2/3 volume of 100% ethanol and passed through a second filter. Small RNAs were immobilized on the filter. Two consecutive washing were performed using miRNA wash solution 1 and wash solution 2/3, respectively. Small RNAs were eluted in 70μl pre-heated (95°C) nuclease-free water.

For mRNA library

Total RNA was extracted from 100 μl infected RBCs using Trizol reagent (Sigma). 1 ml Trizol was added to the cell pellet and pipetted up and down until complete homogenization of the lysate. The protocol was then carried out according to the manufacturer's instructions. Total RNAs and small RNAs were quantified by Qubit RNA high sensitivity kit (Invitrogen).

tRNA library construction

Periodate oxidation and β-elimination

RNA oxidation and β-elimination were performed as previously described[1]. 2 μg isolated small RNA was treated in 100 mM NaOAc/HOAc, pH 4.8 and freshly prepared 50 mM $NaIO_4$ at room temperature for 30 min. The reaction was quenched by incubating with 100 mM glucose for 5 min at room temperature. The treated RNA was purified first with RNA clean-up and concentration micro-elute kit (Norgen Biotek Corp.), and then by ethanol precipitation. For β-elimination, the purified RNA was treated in 60 mM sodium borate, pH 9.5 at 45°C for 90 min. Subsequently, the RNA was purified with RNA clean-up and concentration micro-elute kit followed by ethanol precipitation. Two short RNA standards with A-end and C-end were mixed in equimolar ratio and added to the purified RNA to a final concentration of 0.4 pmol/μg small RNA).

Two synthetic RNA standards:

A-ending: 5'-GCAGAUGGCUUCAAUUGCUAUUAAGGACCA

C-ending: 5'-GCAGAUGGCUUCAAUUGCUAUUAAGGACAC

Bisulfite conversion

Bisulfite conversion of RNA was performed to generate separate sequencing library. 1 μg of isolated small RNA was used for the reaction using EZ RNA methylation kit (Zymo Research).

tRNA library generation

This was carried out as previously described with minor modifications[2,3]. For cDNA synthesis, optimized R2R DNA (1:1 ratio of T-ending and G-ending oligos) was annealed to complementary R2 RNA in 100 mM Tris-HCl, 0.5 mM EDTA, pH 7.5 at 82°C for 2 min, then cooled down to 25°C with a ramp rate of 0.1°C /s[4]. A total of 100 ng tRNA (≈4 pmol) was mixed with 4 pmol primer mixture and pre-incubated at room temperature for 30 min with low salt buffer (50 mM Tris-HCl, pH7 .5, 75 mM KCl, 3 mM $MgCl_2$), 10 mM DDT and 500 nM TGIRT-III enzyme (InGex, Inc). dNTPs mix were added to a final concentration of 1 mM to initiate the reaction. The template switching reaction was performed at 42°C for 16 h.

The reactions were terminated by adding 1 μl of 5 M NaOH and incubated at 95°C for 3 min.The reactions were neutralized by addition of 1μl of 5M HCl. cDNAs were then purified by MinElute Reaction Cleanup Kit (QIAGEN) and eluted in 10 μl nuclease free water. 10 μM R1R DNA containing UMI sequences was adenylated with 2 μl of Mth RNA ligase (New England Biolab) at 65°C for 1 h and subsequently inactivated at 85°C for 5 min. 2 μM of the adenylated R1R DNA was ligated to the purified cDNAs at 65°C for 2 h with Thermostable 5' AppDNA/RNA ligase (New England Biolab). After inactivation at 90°C for 3 min, the ligated cDNAs were purified with MinElute Reaction Cleanup Kit (QIAGEN) and eluted in 23 μl nuclease free water. PCR amplification for Illumina sequencing was performed using 2X Phusion High-Fidelity PCR Master Mix (Thermofisher Scientific) for 12 PCR cycles (98°C 5s, 60°C 10s, 72°C 10s) using R1 multiplex PCR primer and barcoded primer. 1.3X Agencourt AMPure XP beads (Beckman Coulter) were used to clean up the libraries and get rid of primer dimers before Illumina sequencing. The quality and quantity of each library was analyzed on the Bioanalyzer with a High Sensitivity DNA analysis Kit (Agilent technologies) and Collibri Library Quantification Kit (Invitrogen).

R2 RNA:

5'-GAUCGGAAGAGCACACGUCUGAACUCCAGUCAC-3SpC3/

R2R DNA T-ending/ G-ending:

5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT/G

R1R DNA:

5'-/5Phos/GATANNNNNNNGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/

R1 multiplex PCR primer:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCT

Barcoded primer:

5'CAAGCAGAAGACGGCATACGAGAT*[BARCODE]*GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT

mRNA library construction

100 ng total RNA was used to construct mRNA library using TruSeq Stranded mRNA kit (Illumina). The protocol was carried out following manufacturer's instruction. The quality and quantity of each library was analyzed on the Bioanalyzer with a High Sensitivity DNA analysis Kit (Agilent technologies) and Collibri Library Quantification Kit (Invitrogen).

Recombinant synthesis of AlkB demethylases

144 pET30a-AlkB and pET30a-AlkB-D135S were gifts from Tao Pan (Addgene plasmid # 79050
145 and #79051). These plasmids contain wildtype *E.coli alkB* or D135S mutant sequences, both
146 with a deletion of 11 residues from the N-terminal[5]. Plasmids were transformed into *E.coli*
147 BL21(DE3) T1R pRARE2 and the cultures were grown in the presence of 50 μg/ml kanamycin
148 and 34 μg/ml chloramphenicol at 37°C until OD_{600} reached 2. Protein expression was then
149 induced with 0.5 mM IPTG at 18°C overnight. Cells were pelleted and resuspended in lysis
150 buffer (100 mM HEPES, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 0.5 mM TCEP,
151 protease inhibitor cocktail (Roche), pH 8.0). The Lysate was sonicated by pulsed sonication
152 (4s/4s on/off cycle for 4 min, 80% amplitude) and then centrifuged and filtered through 0.45
153 μm filters. The soluble proteins were purified first using HisTrap HP (GE Healthcare) and then
154 by gel filtration on a HiLoad 16/60 Superdex 200 column (GE Healthcare) on an ÄKTA Xpress
155 system. Thrombin was added (1 unit/mg protein) to the purified recombinant proteins and
156 incubated overnight at 4°C. The cleavage tag was then moved by HisTrap HP. The proteins
157 were then concentrated in storage buffer (20 mM HEPES, 300 mM NaCl, 10% glycerol, 2 mM
158 TCEP, pH 7.5).
159
160 **tRNA demethylation reaction**
161 Demethylation assay was performed to validation the identities of the detected mutation
162 signatures. 1 μg small RNA was treated with 80 pmol WT and 160 pmol D135S AlkB in
163 demethylase reaction buffer (300 mM KCl, 50 mM MES buffer (pH 5.0), 2 mM $MgCl_2$, 50 μM
164 of $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, 300 μM 2-ketoglutarate (2-KG), 2 mM L-ascorbic acid, 50 μg/ml
165 BSA). The reaction was incubated at room temperature for 2h and then quenched by adding 5
166 mM EDTA. The treated RNA was purified using RNA clean-up and concentration micro-elute
167 kit.
168
169 **Library sequencing and analysis**
170 *tRNA library read preprocessing*
171 All libraries were sequenced on an Illumina Nextseq 550 platform with a single-end read length
172 of 120 bp. Base calling and demultiplexing were performed with Generate FASTQ (v2.0.1).
173 Unique molecular identifiers (UMI) were extracted using UMI_tools (v1.1.2). Adapter
174 sequences were clipped from the reads using fastx_toolkit (v0.0.14) with a size filtering to
175 discard reads <15bp. These analysis was performed on the galaxy platform.
176 For our system, we expect that the low complexity tRNA transcriptome of the parasite would
177 generate the majority of reads sequenced in the libraries, with the few remaining reads
178 originated from the highly complex human tRNA transcriptome as well as from other non-
179 tRNA small RNA species. Therefore, a combined tRNA reference (including Plasmodium
180 nuclear and apicoplast tRNAs and all human tRNAs from GtRNAdb[6]) was first compiled and
181 curated using sequence data generated from a pilot sequencing trial. The curation includes re-
182 defining the exact 5' and 3' position of all tRNAs, addition of 3'-CCA to all tRNAs, changing
183 the reference from A to G for tRNAs that has inosine modification on position 34 and the
184 addition of a 5' G base on His-tRNA[7].
185 The sequence reads were then aligned to the curated tRNA reference using bowtie2 (v2.2.1)
186 with sensitive local mode[8]. The unaligned reads were then aligned to the 3D7 genome (v.
187 PlasmoDB-46) with sensitive local mode to identify mapping to other small RNA species. Two
188 addition steps were performed to assess the outcome of the mapping. First, the unaligned reads
189 from the genome mapping were again aligned to the tRNA reference with very sensitive local
190 mode, which should yield very few reads. Second, unaligned reads with 3'-CCA were
191 randomly assessed by blasting to 3D7 genome to ensure most reads with 3'-CCA were mapped
192 in prior operations. After alignment, reads with the same UMI and mapping coordinates were
193 de-duplicated using UMI_tools (v1.1.2).

194 tRNA composition was determined by the ratio of read counts per unique tRNA to the read
195 counts of total tRNAs found by htseq-count (v0.9.1) [9].
196 The determination for tRNA aminoacylation level is as follows: an aligned read is considered
197 CCA-ending(charged) if it aligns with no mismatches to the tRNA's 3' end. Consequently, if
198 the read aligns to the 3'end but only ends in -CC, it is considered uncharged. Aminoacylation
199 level were determined as the percentage of CCA-ending reads over the sum of the CCA-ending
200 and CC-ending reads, adjusted to the recovered ratio of the two spike-in RNA oligos.
201 Fractions of mutation signature were calculated as 1- (coverage of reference sequence at
202 position n/ total coverage at position n+1).
203
204 *mRNA library read preprocessing*
205 mRNA libraries were sequenced on an Illumina Nextseq 550 platform with a single-end read
206 length of 120 bp or 150bp (for the polysome profiles). Base calling was performed with
207 Generate FASTQ (v2.0.1). Sequence reads were first filtered with Trimmomatic (v0.38) using
208 the default setting[10]. The trimmed reads were then aligned to 3D7 genome (v. PlasmoDB-46)
209 with HISAT2 (v2.2.1) using default setting and read counts were obtained by htseq-count
210 (v0.9.1)[11].
211
212 **Amino acid starvation assay**
213 Parasite cultures were synchronized with two consecutive cycles of sorbitol and percoll
214 gradient treatment before the start of the assay. All starvation assays used parasites at 32-36
215 hpi maintained at 3% hematocrit and >5% parasitemia. The parasites were washed twice with
216 and resuspended in either AA-free RPMI (US biological, #R8999-04A) or complete RPMI (US
217 biological, #R8999-13) supplemented with 0.5% AlbumaxII and 20µg/ml hypoxanthine, pH
218 7.2. Parasites were incubated for 6 h and harvested immediately afterwards for the preparation
219 of tRNA library and quantitative proteomics analysis. 70 nM of halofuginone (Calbiochem) or
220 382 nM of L-isoleucine (Sigma) were also added where applicable.
221
222 **Quantitative proteomics analysis**
223 *Sample preparation*
224 Cell pellets were solubilized with 100 µL of 8M urea, 0.2% ProteaseMAX (Promega) and 5%
225 acetonitrile (ACN) in 100 mM Tris-HCl, pH 8.0 with 1 µL of 100x protease inhibitor cocktail
226 (Roche) during sonication in water bath for 10 min followed by probe sonicated with VibraCell
227 probe (Sonics & Materials, Inc.) for 20 s, with pulse 2/2, at 20% amplitude. Following
228 centrifugation at 10,000 g for 10 min at 4°C, the supernatant was transferred to a new tube and
229 the remaining pellet was supplemented with 50 µL of 1M urea in 100 mM Tris-HCl with 0.2%
230 ProteaseMAX and sonicated in water bath for 10 min. Protein concentration in the combine
231 supernatant was measured by BCA assay (Thermofisher Scientific). An aliquot of 25 µg
232 samples was transferred to a new tube and equalized with 50 mM Tris-HCl to a total volume
233 of 100 µL. Proteins were reduced with addition of 1 µL of 500 mM dithiothreitol (Sigma) at
234 25°C for 1 h and alkylated with 3 µL of 500 mM iodoacetamide for 1 h at room temperature in
235 the dark and quenched with 2 yL of dithiothreitol for 10 min. Proteolytic digestion was started
236 with 2 µL of 0.5 µg/µL Lys-C (Wako, Japan) at 37°C for 3 h with shake at 450 rmp and
237 supplemented with 300 µL of 100 mM Tris-HCl, pH 8.0 before adding 4 µL of 0.5 µg/µL
238 sequencing grade trypsin (Promega) and incubtaed over night at 37°C. The digestion was
239 stopped with 20 µL cc. formic acid, incubating the solutions at RT for 5 min. The sample was
240 cleaned on a C18 Hypersep plate with 40 µL bed volume (Thermofisher Scientific), dried using
241 a vacuum concentrator (Eppendorf). Biological samples were labeled with TMT-10plex
242 reagents in random order adding 200 µg TMT-reagent in 30 µL dry ACN to each digested
243 sample resolubilized in 70 µL of 50 mM triethylammonium bicarbonate (TEAB) and

incubating at room temperature (RT) for 2 h. The labeling reaction was stopped by adding 11 µL of 5% hydroxylamine and incubated at RT for 15 min before combining them in one vial.

Liquid Chromatography-Tandem Mass Spectrometry Data Acquisition

Peptides were reconstituted in solvent A and approximately, two µg samples injected on a 50 cm long EASY-Spray C18 column (Thermofisher Scientific) connected to an Ultimate 3000 nanoUPLC system (Thermofisher Scientific) using a 120 min long gradient: 4-26% of solvent B (98% acetonitrile, 0.1% FA) in 120 min, 26-95% in 5 min, and 95% of solvent B for 5 min at a flow rate of 300 nL/min. Mass spectra were acquired on a Q Exactive HF hybrid quadrupole orbitrap mass spectrometer (Thermofisher Scientific) ranging from m/z 375 to 1700 at a resolution of R=120,000 (at m/z 200) targeting 5x106 ions for maximum injection time of 80 ms, followed by data-dependent higher-energy collisional dissociation (HCD) fragmentations of precursor ions with a charge state 2+ to 8+, using 45 s dynamic exclusion. The tandem mass spectra of the top 18 precursor ions were acquired with a resolution of R=60,000, targeting 2x105 ions for maximum injection time of 54 ms, setting quadrupole isolation width to 1.4 Th and normalized collision energy to 33%.

Data analysis

Acquired raw data files were analyzed using Proteome Discoverer v2.5 (Thermofisher Scientific) with Sequest HT search engine against Plasmodium falciparum PF3D7 protein database (SwissProt). A maximum of two missed cleavage sites were allowed for full tryptic digestion, while setting the precursor and the fragment ion mass tolerance to 10 ppm and 0.02, respectively. Carbamidomethylation of cysteine was specified as a fixed modification, while TMT6plex on lysine and N-termini, oxidation on methionine as well as deamidation of asparagine and glutamine were set as dynamic modifications. Initial search results were filtered with 5% FDR using Percolator node in Proteome Discoverer. Quantification was based on the TMT-reporter ion intensities. Differential quantitation analysis was performed by DESeq2[12].

Multiplication rate assay

Parasite cultures were synchronized with two consecutive cycles of sorbitol and percoll gradient treatment before the start of the assay. Parasites at 8-12 hpi or 32-36 hpi were collected for the assay. The parasites were washed twice in AA-free RPMI and then resuspended at 2% hematocrit and 1% parasitemia in different conditions including in complete medium, AA-free medium and AA-free medium supplied with varying concentrations of isoleucine (382 µM, 40 µM, 20 µM, 10 µM, 5 µM, 2.5 µM) in a 96-well plates. After 6h of incubation, the starved cultures were re-supplemented with complete medium until all parasites had entered a new invasion cycle. Microscopy assessment was used to ensure all parasites were either in ring or trophozoite stage. Samples were then harvested for flow cytometry analysis.

Flow cytometry analysis

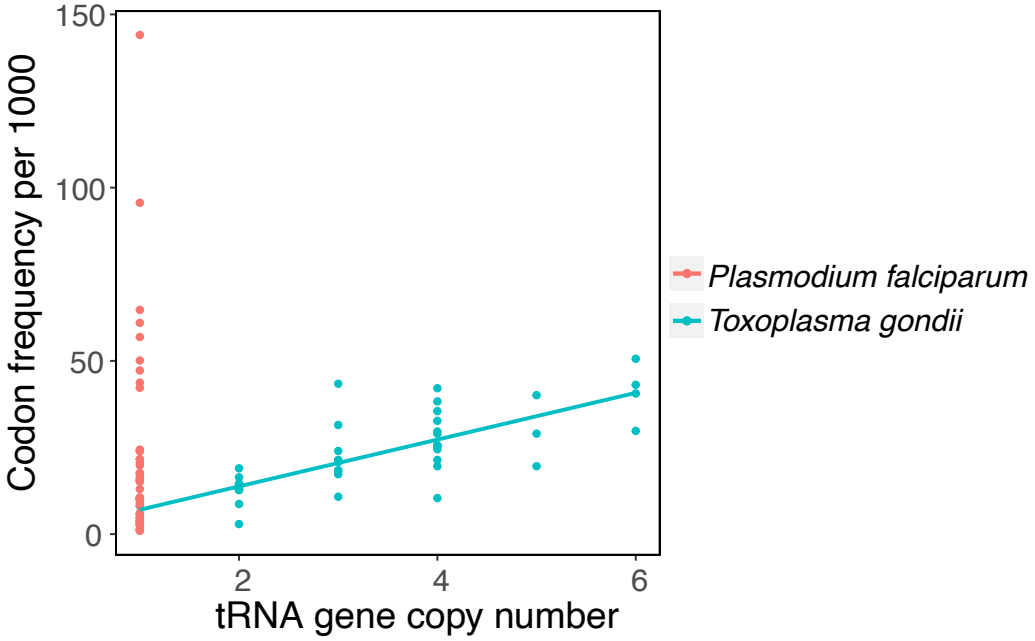
Samples were harvested for flow cytometry analysis at 0h, 6h, 24h and 48h after plate setup for 6-10 hpi parasites and at 0h, 6h, 12h, 18h and 32h for 32-36hpi parasites. 20µl of sample per timepoint were harvested and stained for 30 min at 37°C with Hoechst (Invitrogen) and DHE (Sigma) at a final concentration of 20 µg/ml and 10 µg/ml respectively. Stained cells were diluted in PBS to a final hematocrit of 0.2% prior to analysis. At least 3000 Hoechst+/DHE+ events/sample were acquired using a FACSVerse (BD Biosciences). Data was analyzed as described previously using FlowJo (Version 10.6.2)[13].

Polysome fractionation

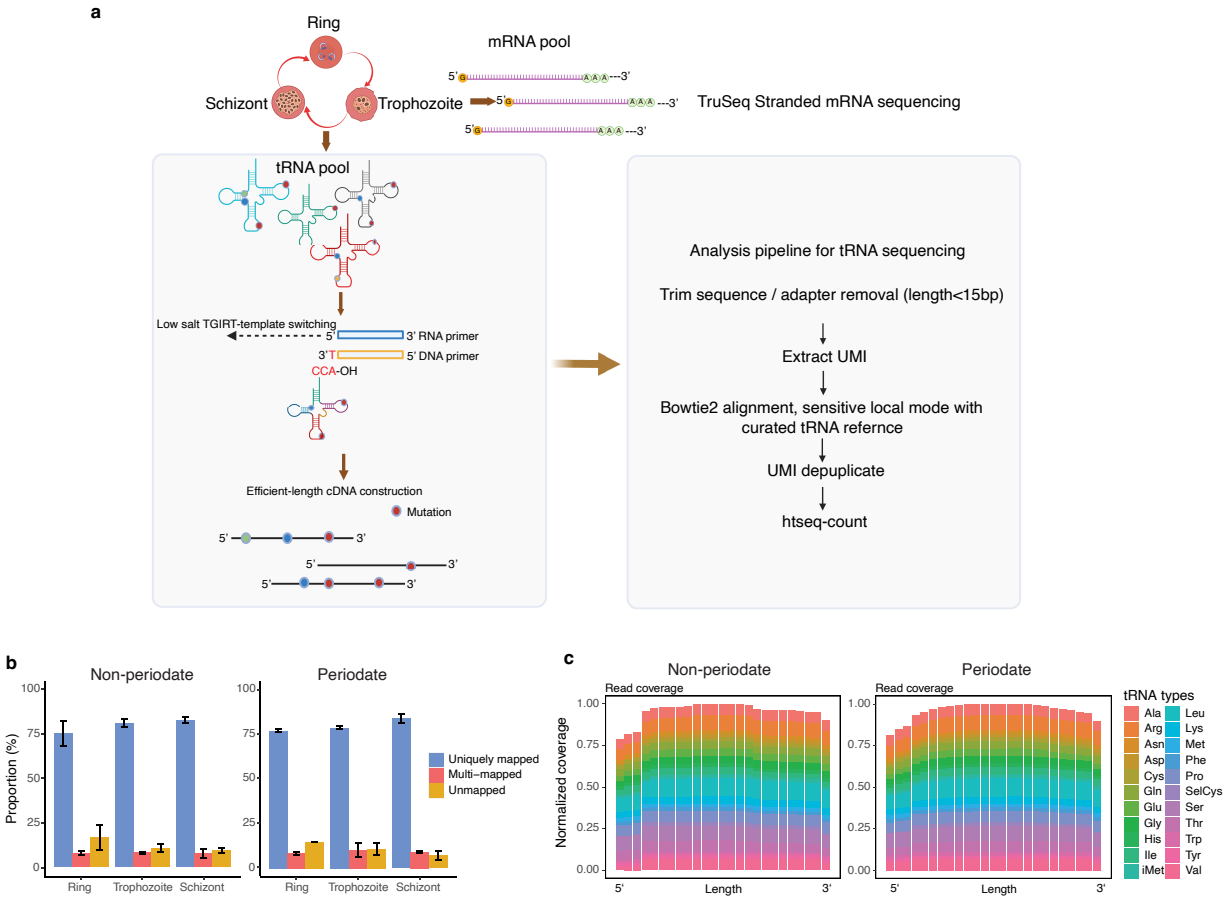
293 Polysome purification was adopted from a protocol previously reported but with major
294 modifications[14].
295 Cultures were expanded to 150 ml at 3% hematocrit and ~10% parasitemia and were
296 synchronized with two consecutive cycles of sorbitol and percoll gradient treatment before the
297 start of the assay. When the culture reached 32-36 hpi, infected RBCs were enriched with
298 magnetic separation using XS column (Miltenyi Biotec). After the enrichment, the cell pellet
299 was equally split and cultured either in AA-free RPMI or complete RPMI for 6h.
300 After the incubation, cycloheximide was added to a final concentration of 100 µg/ml to the
301 cultures. The cultures were briefly shaken and immediately incubated on ice for 10 min.
302 Afterwards, the cells were centrifuged and washed once with 10 ml ice-cold PBS supplemented
303 with 100 µg/ml cycloheximide. Equal volume of ice-cold 2x lysis buffer (800 mM potassium
304 acetate, 50 mM potassium Hepes, pH 7.2, 30 mM magnesium acetate, 10 µg/ml cycloheximide,
305 2 mM DTT, 2 mM AEBSF, 40 U/ml rRNasin (Promega), 1% sodium deoxycholate, 2% Igepal
306 CA-630) was added to the pellet and gently resuspended the pellet. The resuspended pellet was
307 incubated on ice for 5 min with a brief vortex every minute. The pellet was then centrifuged at
308 17000 g for 2 min to clear the lysate. 50 µl of the lysate was separated for total RNA extraction.
309 A 10% to 50% sucrose gradient solution in gradient buffer (20 mM HEPES pH 7.6, 100 mM
310 KCl, 5 mM $MgCl_2$) was prepared on a Gradient Station platform (Biocomp Instruments). 450
311 µl of the lysate was carefully layered on top of the sucrose gradient and centrifuged at 35000
312 rpm for 2h on a pre-chillled SW40 Ti rotor with the OptimaXL-100K ultracentrifuge system
313 (Beckman Coulter). After the centrifugation, the gradient was fractionated into 25 fractions on
314 a Gradient Station platform (Biocomp Instruments). RNA and proteins were extracted from the
315 fractions using Trizol reagent (Sigma) as per manufacturer's instruction and the RNA from the
316 polysomal fractions were pooled.
317
318 **Immunoblotting**
319 Western blot was performed with proteins purified from the polysome fractionation. Purified
320 proteins were solubilized in Laemmli buffer (Bio-Rad) with 2% β-mercaptoethanol and heated
321 at 95°C for 10 min. The proteins were resolved in a criterion TGX Stain-Free gel (Bio-Rad)
322 with Tris/Glycine/SDS running buffer (Bio-Rad). The resolved proteins were transferred to
323 nitrocellulose membrane in Tris/glycine buffer (Bio-Rad). The membrane was blocked with
324 5% skim milk, 2% BSA in PBS at 4°C overnight. The membrane was incubated with 1:2000
325 anti-RPS14 antibody (abcam, #ab174661) in 2% skim milk, 2% BSA in PBS for 2h at room
326 temperature. The membrane was washed thrice with 1x TBS, 0.1% Tween 20 and then
327 incubated for 30 min with 1:5000 HRP-conjugated anti-rabbit IgG (Amersham, #NA934).
328 Chemiluminescent signal was detected with a ChemiDoc imager (Bio-Rad) after developing
329 with WesternBright ECL kit (advansta).
330
331 **Orthologue analysis**
332 All orthologues to the isoleucine-rich genes of *P. falciparum* were obtained from OMA
333 orthology database using the OMA group fingerprint[15]. A cut-off of having at least 100
334 orthologues was applied to consider a *P. falciparum* protein as functionally conserved and used
335 for anaylsis.
336
337 **Kozak Similarity score (KSS)**
338 Calculation of KSS was inspired by a previous report[15]. Bit scores for each nucleotide from
339 position -3 to +4 of the annotation start codon was generated from all genes encoding ribosomal
340 protein and used as a reference. The KSS of a gene was calculated as the sum of all the bit
341 scores from these positions relative to the maximum possible score summed from these
342 positions.

Statistical analysis

Two tailed student's t-test, two tailed Mann-Whitney U test and Fisher's exact test were used for statistical analyses wherever specified. A p-value of <0.05 was used as the cut off for statistical significance. Multiple testings were corrected with Benjamini-Hochberg procedure, with a false discovery rate <0.05 used as the cut off for statistical significance.



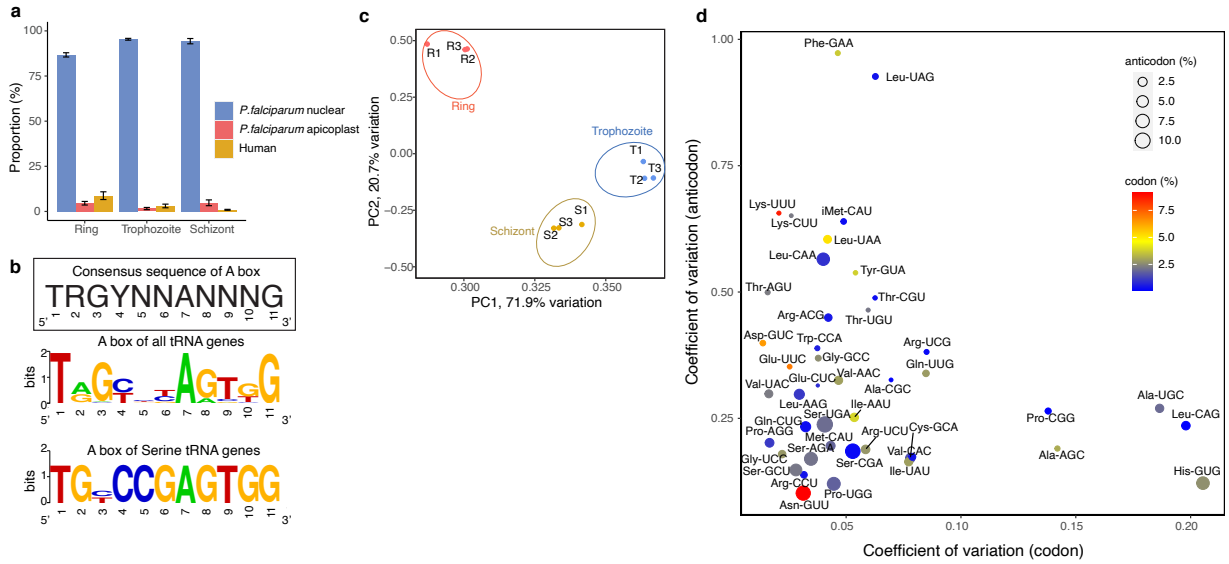
Extended Data Fig. 1
The codon usage bias in *P. falciparum* is not compensated by an expansion of tRNA genes. Correlations between tRNA gene copy numbers and codon usage in the genomes of *T. gondii* and *P. falciparum*



Extended Data Fig. 2

An optimized workflow showing full-length cDNA library construction and data analysis pipeline.

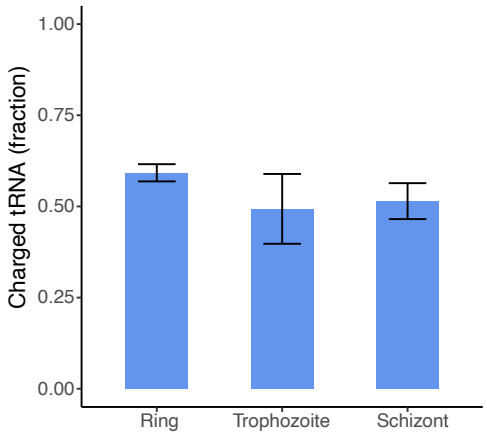
(a) Left panel: Schematic showing tRNA sequence library generation using TGIRT. A DNA/RNA duplex with a T overhang that is complementary to the terminal adenosine of mature tRNAs was used for the template-switching reaction. The 3' adapter was attached only after the cDNA was synthesized. Right panel: Strategy for read processing of the tRNA sequence libraries. **(b)** Bowtie2 alignment statistics for tRNA-sequencing data generated from non-periodate (untreated, N=3) and periodate-treated (N=3) libraries. **(c)** Representative scaled 5'-to-3' sequence coverage plots spanning all nuclear tRNAs. The full-length tRNA was divided into 25 bins with each bin representing 4% of the tRNA length, as shown on the x-axis. The Y-axis value for each tRNA was normalized to the bin with the maximum coverage for each respective tRNA. Left panel: samples without periodate treatment. Right panel: sample treated with periodate.



Extended Data Fig. 3

Characteristics of the anticodon and codon pools.

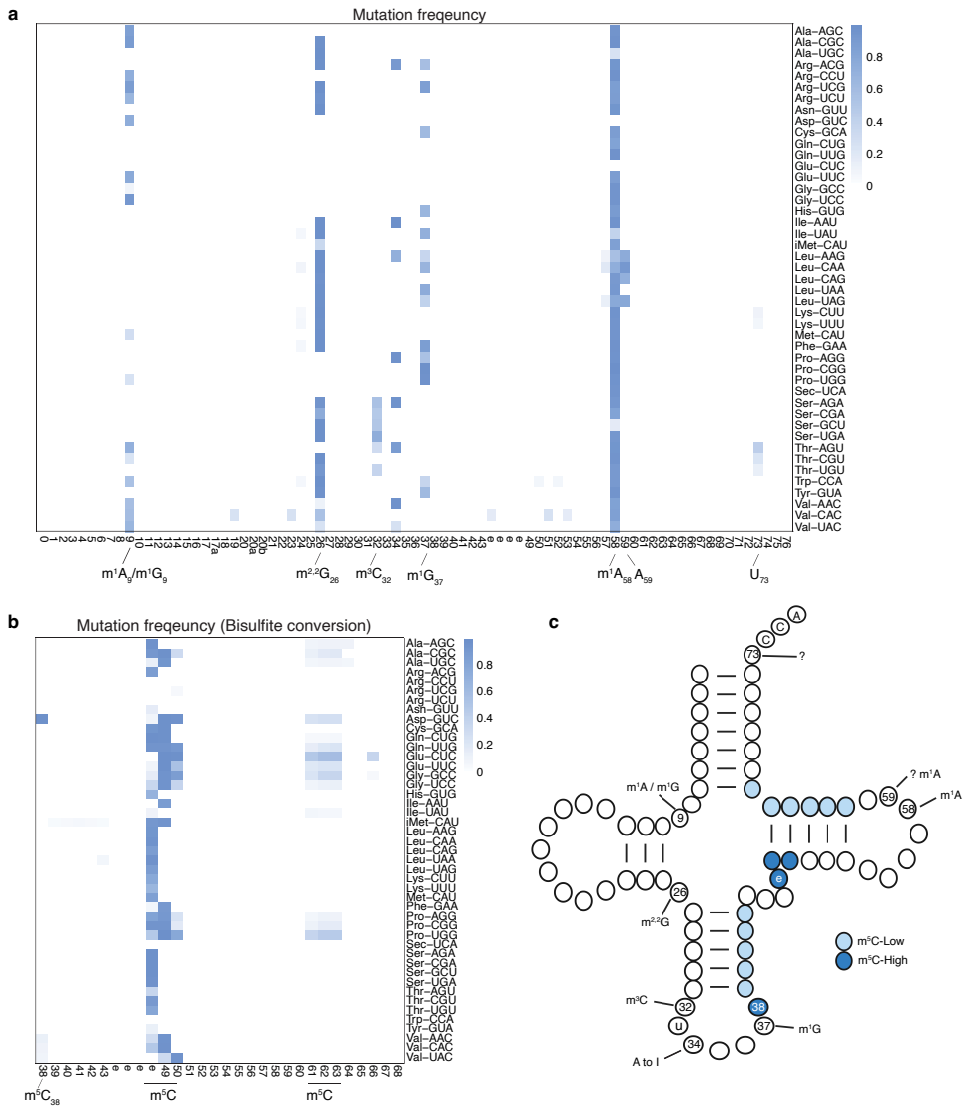
(a) Origin of all reads mapped to tRNAs throughout the IDC. Error bars represent the standard deviation of N=3 biological replicates. **(b)** Motif logo of the A-box in all *P. falciparum* tRNAs, except Ser-tRNAs (upper), and of all Ser-tRNAs (lower). With a G-to-Y substitution at the third position, the A-box of Ser-tRNAs diverges from the eukaryotic consensus (boxed). **(c)** Principal component analysis of the coding transcriptomes throughout the IDC. **(d)** Coefficient of variation of each anticodon (tRNA) and the respective codons it decodes. The abundance of each anticodon and its decoded codons as a percentage of the total anticodon (size) and codon pool (colour) are illustrated. Calculated from N=9, N=3 for the ring, the trophozoite and the schizont stages. All apicoplast tRNAs are excluded.



Extended Data Fig. 4

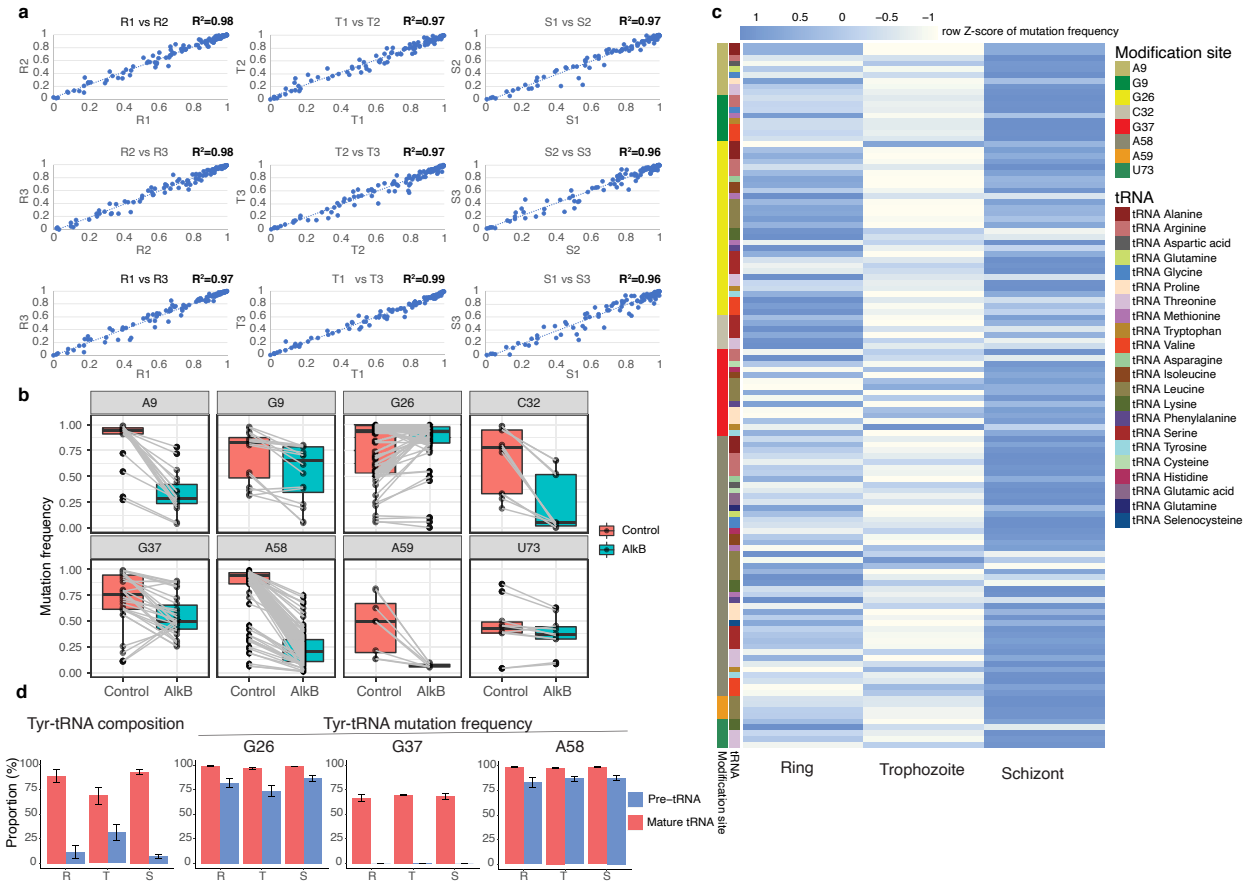
Charge level of total tRNA throughout the IDC.

Charged ratio of tRNAs are represented by reads with 3'-CCA/-CC ends after periodate oxidation and beta-elimination reactions. Error bars represent the standard deviation of biological replicates; Ring and Trophozoite: N=3; Schizont: N=2. All apicoplast tRNAs are excluded.



Extended Data Fig. 5

The mutation frequency measured by TGIRT-tRNAseq reflects modification stoichiometry. Global heatmap of the average proportion of reads with **(a)** mutation signatures and **(b)** m^5C methylation detected per position in all tRNAs at the ring stage. Numbers represent the canonical position, e denotes positions in the variable loop, inferred identities of the modifications that cause the mutations are shown (A_9, G_9, G_{26}, C_{32}, A_{34}, G_{37} and A_{58}), and identities of A_{59} and U_{73} cannot be inferred. **(c)** Schematic representation of the locations of the modifications in relation to the secondary structure of tRNA. Blue represents the presence of m^5C methylation with high (dark blue) or low (light blue) stoichiometry.



Extended Data Fig. 6

The dynamics of tRNA modifications throughout the IDC.

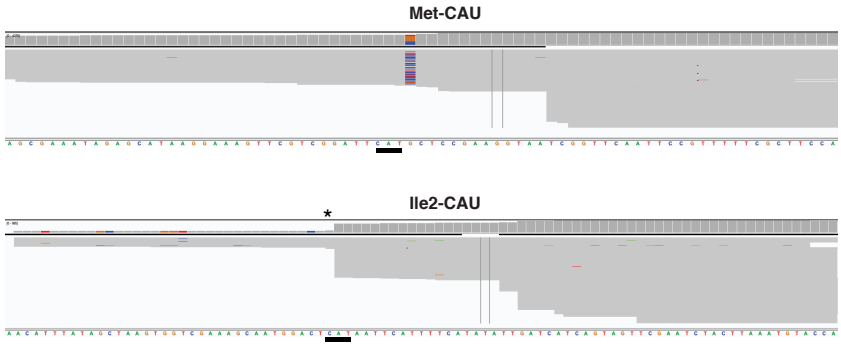
(a) Pairwise correlations of the mutation frequency at all detected sites between replicates. **(b)** Changes in the average mutation frequencies after AlkB demethylase treatment (N=2). Modifications of G26 and U73 were insensitive to treatment. **(c)** Heatmap showing changes in the mutation frequency at each site during IDC. **(d)** Left panel: proportion of pre- and mature Tyr-tRNA during IDC. Right panel: mutation frequencies of pre- and mature Tyr-tRNA at G26-, G37- and A58-modified sites during IDC. Error bars in all panels represent the SD of N=3 biological replicates.

a

Gene ID	Anticodon	pos37 mutation
PF3D7_API00200	His-GUG	+
PF3D7_API00300	Cys-GCA	
PF3D7_API00450	Leu-UAA	
PF3D7_API00600	Ile2-CAU	
PF3D7_API00700	Tyr-GUA	
PF3D7_API00800	Ser-GCU	
PF3D7_API00900	Asp-GUC	
PF3D7_API01000	Lys-UUU	
PF3D7_API01100	Glu-UUC	
PF3D7_API01200	Pro-UGG	+
PF3D7_API03100	Phe-GAA	+
PF3D7_API03200	Gln-UUG	
PF3D7_API03300	Gly-ACC	
PF3D7_API03400	Trp-CCA	
PF3D7_API03700	Gly-UCC	
PF3D7_API03900	Ser-UGA	+
PF3D7_API04800	Thr-UGU	
PF3D7_API05000	Met-CAU	+
PF3D7_API05100	Arg-ACG	
PF3D7_API05200	Val-UAC	
PF3D7_API05300	Arg-UCU	
PF3D7_API05400	Leu-UAG	
PF3D7_API05500	Asn-GUU	
PF3D7_API05600	Ala-UGC	
PF3D7_API05800	Ile-GAU	
PF3D7_API06000	Ala-UGC	Dup.
PF3D7_API06100	Asn-GUU	Dup.
PF3D7_API06200	Leu-UAG	Dup.
PF3D7_API06300	Arg-UCU	Dup.
PF3D7_API06400	Val-UAC	Dup.
PF3D7_API06500	Arg-ACG	Dup.
PF3D7_API06600	Met-CAU	Dup.
PF3D7_API06800	Thr-UGU	Dup.

A37
G37

b



c

aa	codon	frequency	tRNA	aa	codon	frequency	tRNA	aa	codon	frequency	tRNA	aa	codon	frequency	tRNA
F	TTT	457	0	L	CTT	4	0	I	ATT	520	0	V	GTT	49	0
F	TTC	5	1	L	CTC	0	0	I	ATC	8	1	V	GTC	0	0
L	TTA	889	1	L	CTA	10	2	I	ATA	759	0	V	GTA	79	2
L	TTG	19	0	L	CTG	0	0	M	ATG	86	3	V	GTG	1	0
S	TCT	131	0	P	CCT	76	0	T	ACT	106	0	A	GCT	0	0
S	TCC	5	0	P	CCC	1	0	T	ACC	1	0	A	GCC	16	0
S	TCA	97	1	P	CCA	27	1	T	ACA	101	2	A	GCA	41	2
S	TCG	3	0	P	CCG	1	0	T	ACG	2	0	A	GCG	1	0
Y	TAT	778	0	H	CAT	70	0	N	AAT	1117	0	D	GAT	128	0
Y	TAC	8	1	H	CAC	0	1	N	AAC	7	2	D	GAC	2	1
Stop	TAA	23		Q	CAA	117	1	K	AAA	917	1	E	GAA	145	1
Stop	TAG	0		Q	CAG	9	0	K	AAG	53	0	E	GAG	8	0
C	TGT	81	0	R	CGT	14	2	S	AGT	108	0	G	GGT	119	1
C	TGC	1	1	R	CGC	0	0	S	AGC	0	1	G	GGC	0	0
Stop	TGA	7		R	CGA	0	0	R	AGA	84	2	G	GGA	85	1
W	TGG	29	1	R	CGG	0	0	R	AGG	3	0	G	GGG	13	0

Minimum requirement:
24 full tRNA isoacceptors
+1 split tRNA
+ lysidine modification

Total:7391 sense
Splice sensitive:12.3% (908)
GU wobble: 44.1% (3259)
UG wobble: 1.3% (93)
Superwobbling: 5.3% (389)
2-thiolation sensitive: 16% (1179)
Lysidine sensitive: 10.3%(759)

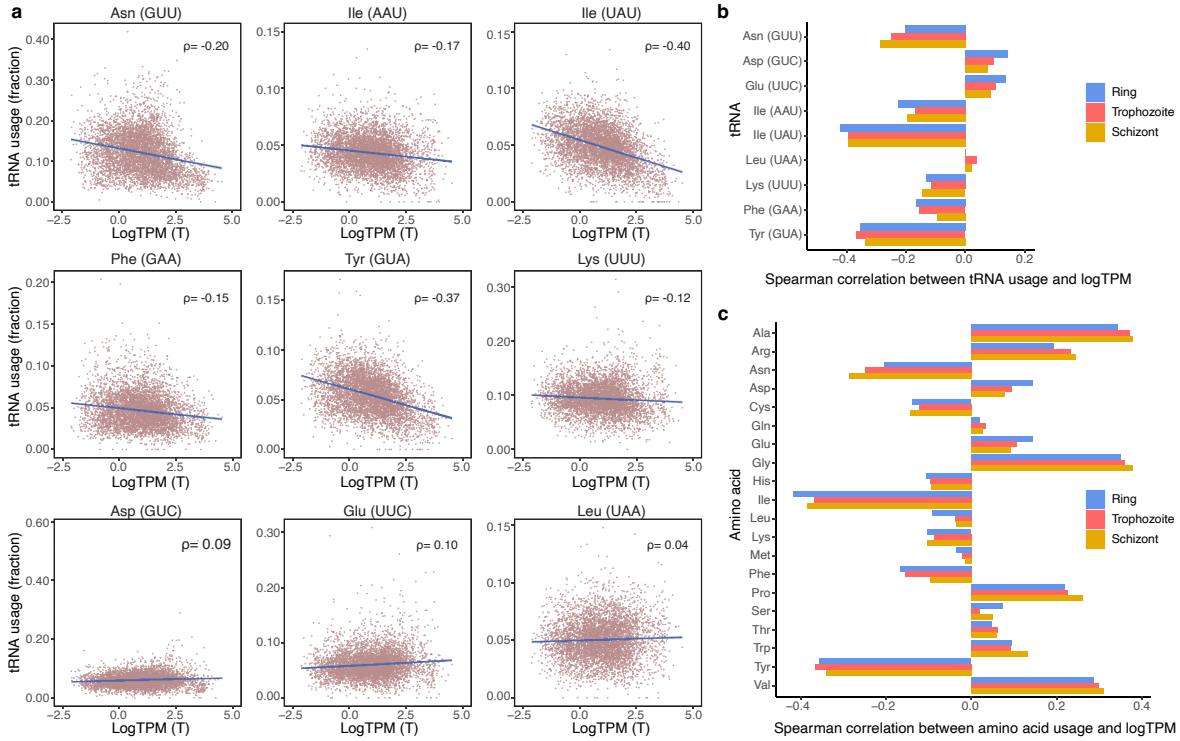
d

tRNAscan ID	Anticodon predicted isotype	Isotype Prediction (Anticodon v. isotype Model)	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Ile2	Leu	Lys	Met	Phe	Pro	SeC	Ser	Thr	Trp	Tyr	Val	tMet
Ile2-CAU.tRNA1	Met	Inconsistent	-5.8	32.4	10.9	No Hit	29.3	30.6	No Hit	26.1	13.9	26.9	41	No Hit	39.8	16.6	42.6	23.5	No Hit	26.7	32.8	2.3	18.9	26.9	No Hit
Met-CAU.tRNA1	Met	Consistent	29	26.8	No Hit	-1.4	33.6	11.8	No Hit	6.8	16.7	20.7	No Hit	16.1	23.3	37.6	28.8	10.5	No Hit	No Hit	31.8	11.1	No Hit	21.1	23.4
Ile-GAU.tRNA1	Ile	Consistent	2.9	26.7	No Hit	No Hit	13.1	2.7	18.0	23.0	-2.9	39.9	8.3	No Hit	No Hit	11.1	No Hit	No Hit	No Hit	No Hit	25.6	5.5	No Hit	15.3	No Hit

Extended Data Fig. 7

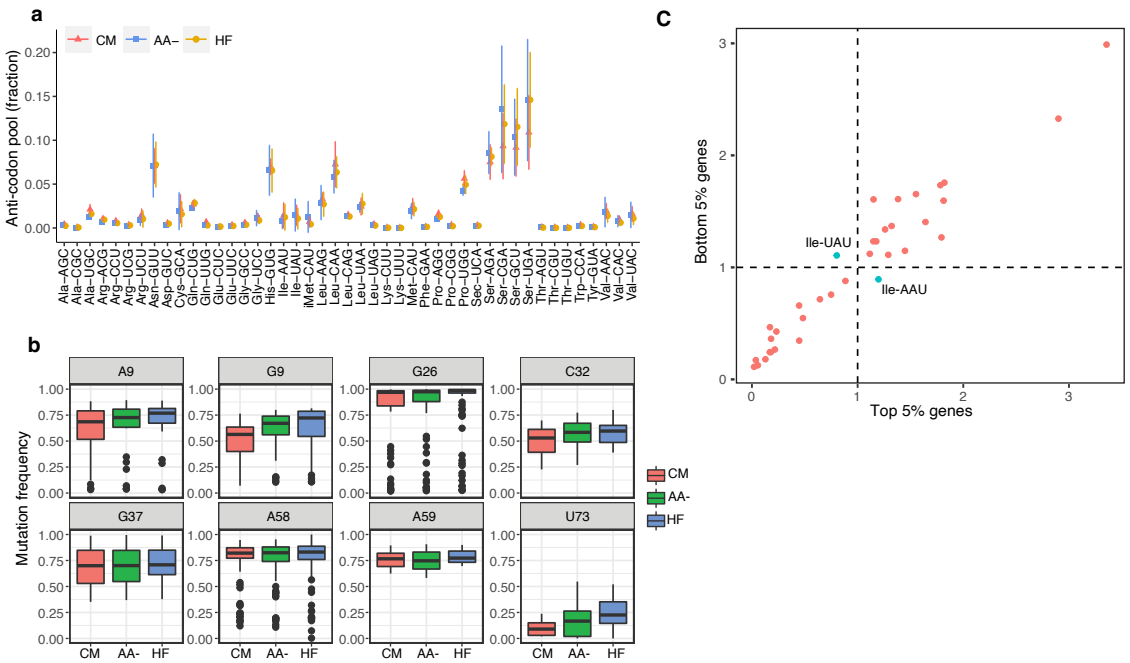
The characteristics of the apicoplast tRNA transcriptome.

(a) All G37 sites on apicoplast tRNAs are methylated. A37 (orange)/G37 (green) indicates the nucleotide identity at position 37. (+) denotes the detection of mutation signature. (Dup.) indicates duplicated genes. **(b)** Integrative Genomics Viewer (IGV) image showing the truncation at the wobble position * in (lower) a potential Ile2-tRNA (PF3D7_API00600), and no truncation was observed in (upper) the Met-tRNA (PF3D7_API05000). **(c)** An illustration of all possible codon anticodon pairings predicted by the wobble hypothesis. The cognate tRNA for the ATA codon was not found in the apicoplast genome. The number of tRNA genes per anticodon (tRNA) and the number of codons used (frequency) by the apicoplast genome are indicated. **(d)** tRNAscan-SE predicted an Ile2-tRNA.



Extended Data Fig. 8

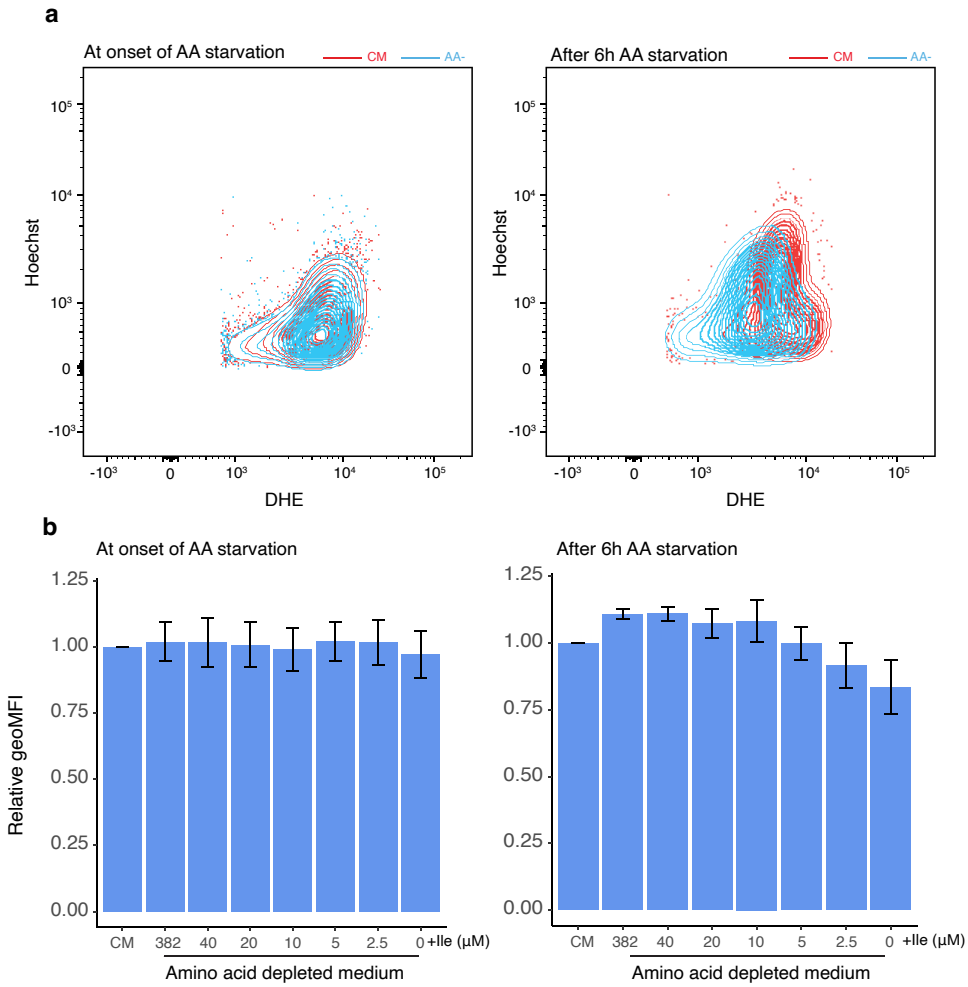
Highly expressed genes minimize dependency on certain types of amino acids. **(a)** Correlation between the usage frequency and gene expression level in the trophozoite stage (LogTPM) based on a fitted linear trendline of nine tRNAs. Spearman's rank correlation coefficients (ρ) are indicated. (n=5267)**.** **(b)** Spearman correlations between the demand of tRNA usage frequency and gene expression level (LogTPM) throughout the IDC. **(c)** Spearman correlation between amino acid usage frequency and gene expression level (LogTPM) throughout the IDC. The LogTPM value for each stage is the average of N=3.



Extended Data Fig. 9

Limited changes in the abundance and modification levels of the anticodon pool after acute amino acid depletion.

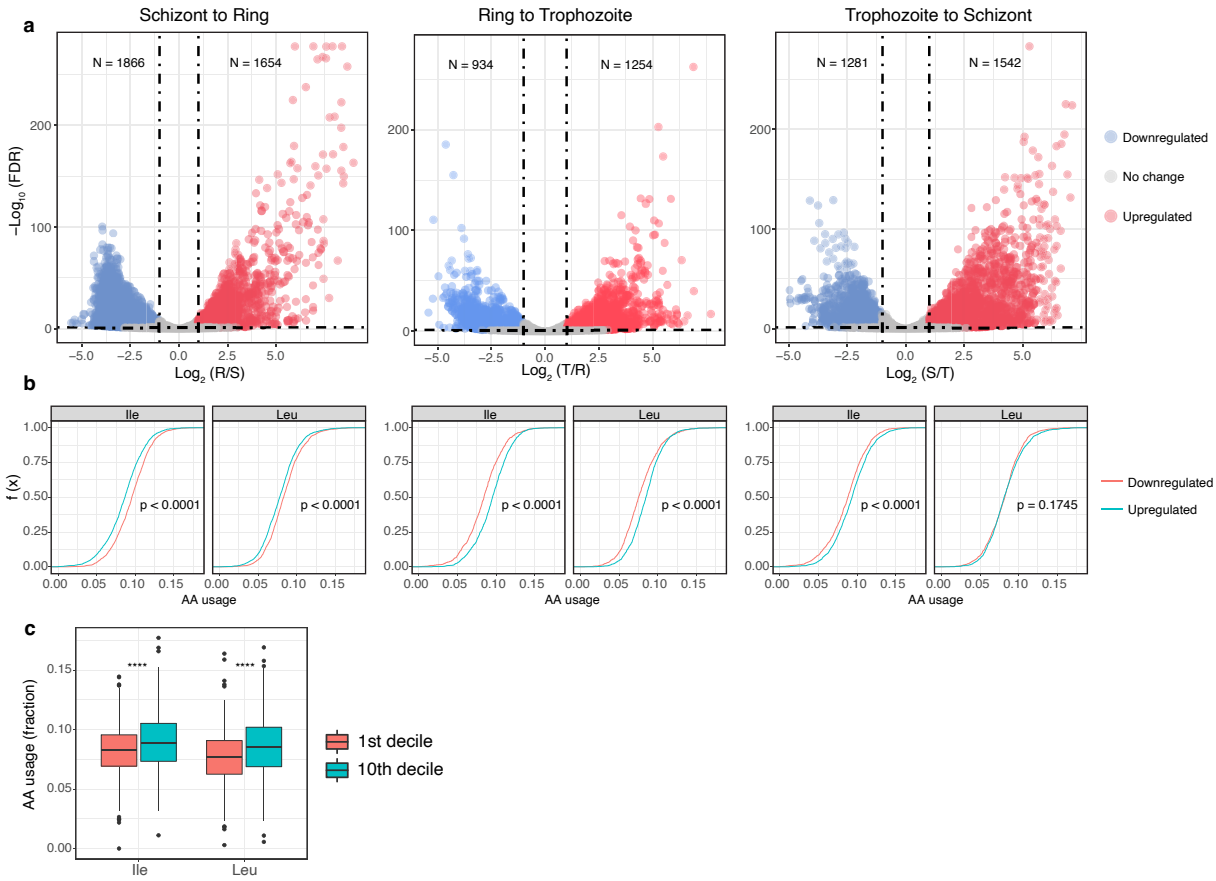
(a) Changes in tRNA composition and **(b)** mutation frequency at each position detected after 6 hours of treatment in trophozoites 32-36 h.p.i. (CM: complete medium; AA-: AA depletion; and HF: halofuginone (70 nM) and AA depletion). Error bars represent the SD of N=3 biological replicates. **(c)** The relative isoacceptor usage of the 5% most highly expressed coding genes (x-axis, n=258) and the 5% least expressed coding genes (y-axis, n=258). Relative isoacceptor usage for each tRNA was calculated as (the fraction of all codons encoding a specific amino acid that is decoded by an isoacceptor/the fraction of all isoacceptors for a specific amino acid that are equally used). The most highly expressed genes preferred Ile-AAU, but the least expressed genes preferred Ile-UAU.



Extended Data Fig. 10

Acute amino acid depletion causes delayed progression.

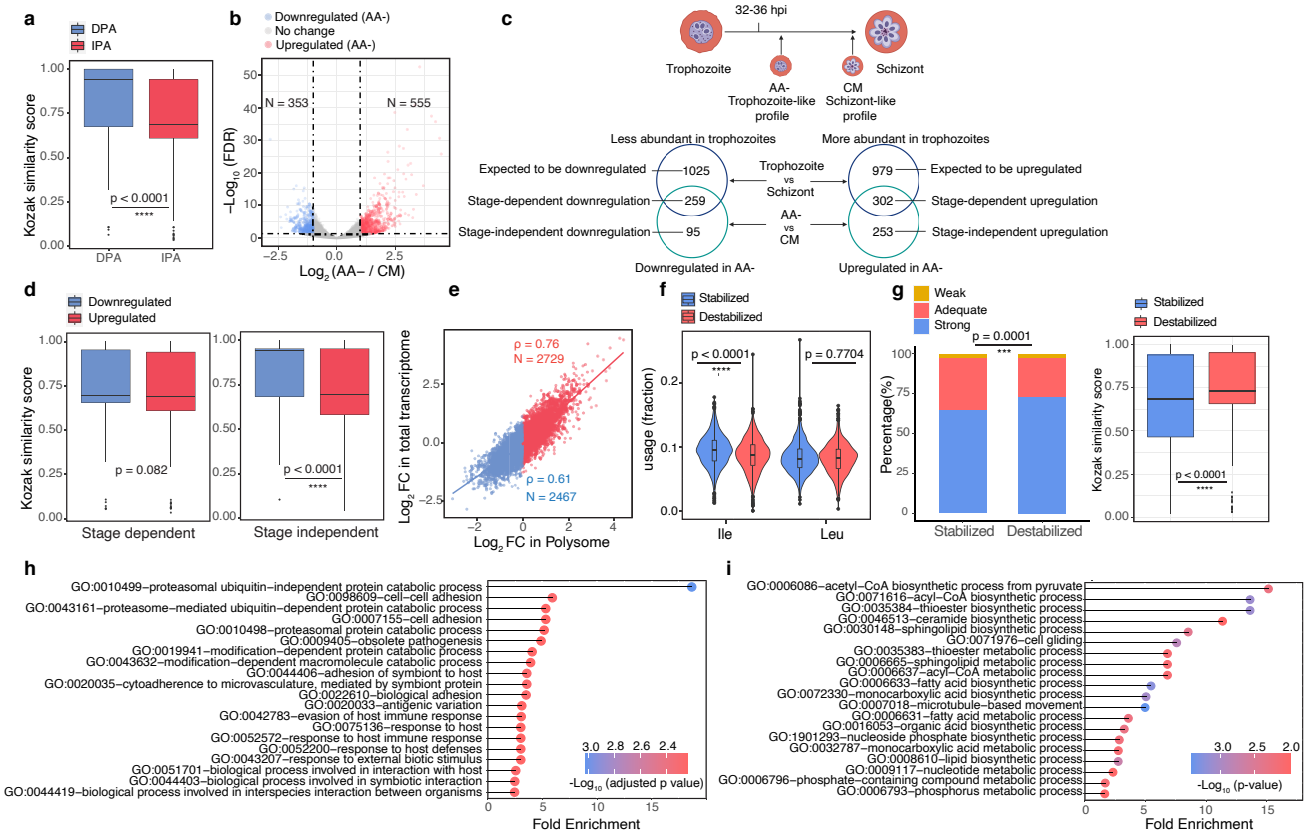
(a) FACS plots showing the fluorescence intensity of dihydroethidium (DHE) and Hoechst DNA staining in parasites (left) at the beginning and (right) after 6 hours of AA depletion (AA-) compared to parasites grown in complete medium (CM). Both DHE and Hoechst staining were reduced in AA- parasites after 6 hours. **(b)** The geometric mean of the Hoechst staining intensity in parasites (left) at the beginning and (right) after 6 hours of AA depletion relative to that in CM parasites. Supplementation of AA- parasites with isoleucine rescued the reduction in DNA content in a dose-dependent manner. Error bars represent the SD for N=3 biological replicates.



Extended Data Fig. 11

Stage progression is isoleucine sensitive.

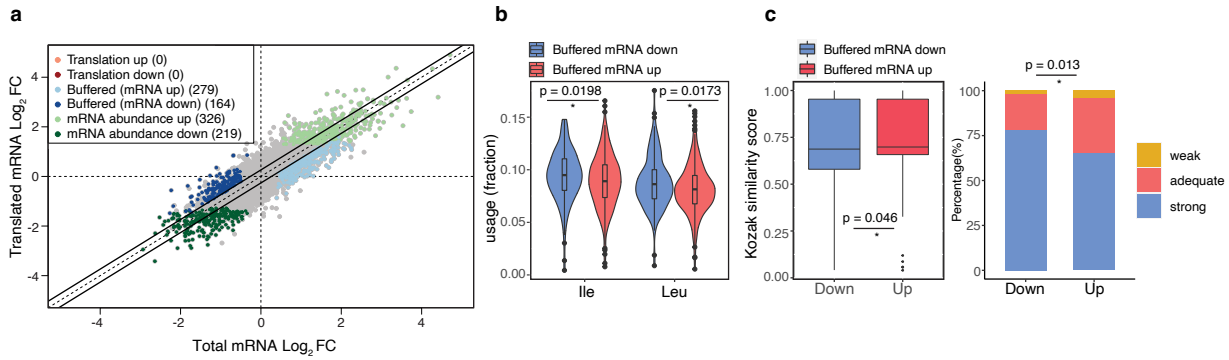
(a) DESeq2 differential gene expression analyses of the transcriptomes during stage progression from (left) the schizont-to-ring stage, (middle) the ring-to-trophozoite stage, and (right) the trophozoite-to-schizont stage. The horizontal and vertical dotted lines show the cut-off of the FDR-adjusted p-value=0.05 and cut-off of the twofold change, respectively. N=3 biological replicates. **(b)** Empirical CDF plot of isoleucine and leucine usage in up- (cyan) and downregulated (red) genes during stage progression. There was an increase in isoleucine sensitivity during the ring-to-schizont progression, which was reflected by higher isoleucine usage for the upregulated genes but reduced sensitivity during the schizont-to-ring transition. The p-values of Mann–Whitney U tests are indicated. **(c)** Isoleucine and leucine usage in the 10% of genes with the most variable expression (the 10th decile) and the 10% of genes with the least variable expression (1st decile) throughout the IDC. Variability was defined by the coefficient of variation in expression (CV of TPM) throughout the IDC.



Extended Data Fig. 12

Increased ribosome loading stabilizes transcripts and reprogrammes the transcriptome. **(a)** Kozak similarity scores (KSS, using ribosomal genes as reference) in transcripts with increased polysome association (IPA) or decreased polysome association (DPA) after amino acid depletion. Transcripts with increased polysome association exhibited lower scores, indicating an initiation context less similar to that of the highly translated ribosomal genes. The p-value of the Mann–Whitney U test is indicated. **(b)** DESeq2 differential gene expression analyses of the transcriptome after amino acid depletion. The horizontal and vertical dotted lines show the cut-off of the FDR-adjusted p-value=0.05 and cut-off of twofold change, respectively. N=3 biological replicates. **(c)** The strategy used to categorize stage-dependent and stage-independent changes in the transcriptome. The transcriptome of the AA-starved parasites (AA-) was expected to most closely resemble a trophozoite-like transcriptional profile due to delayed stage progression (upper). Transcripts that were more abundant in trophozoites were also expected to be upregulated in AA- parasites and vice versa, as illustrated in the Venn diagram. **(d)** The KSS of up- and downregulated genes in the stage-dependent changes (left panel) and stage-independent changes (right panel). Only stage-independent changes show a disparity in KSS between the up- and downregulated genes similar to that in the polysome fractions, supporting a positive feedback loop. The p-values of Mann–Whitney U tests are indicated. **(e)** The correlation of the changes in transcript abundance in the polysome fractions and those in the total transcriptome after amino acid depletion. The differences in **(f)** isoleucine and leucine usage and **(g)** the Kozak context in transcripts that remained stable after amino acid depletion but were expected to be downregulated (stabilized) or upregulated (destabilized) according to the categorization shown in Panel (c). Gene Ontology analysis of **(h)** the downregulated genes and **(i)** the upregulated genes in the stage-independent change group. Only the top 20 enrichment terms are shown. Downregulated genes were enriched in proteasomal pathways, suggesting a lower protein turnover. No pathway was significantly

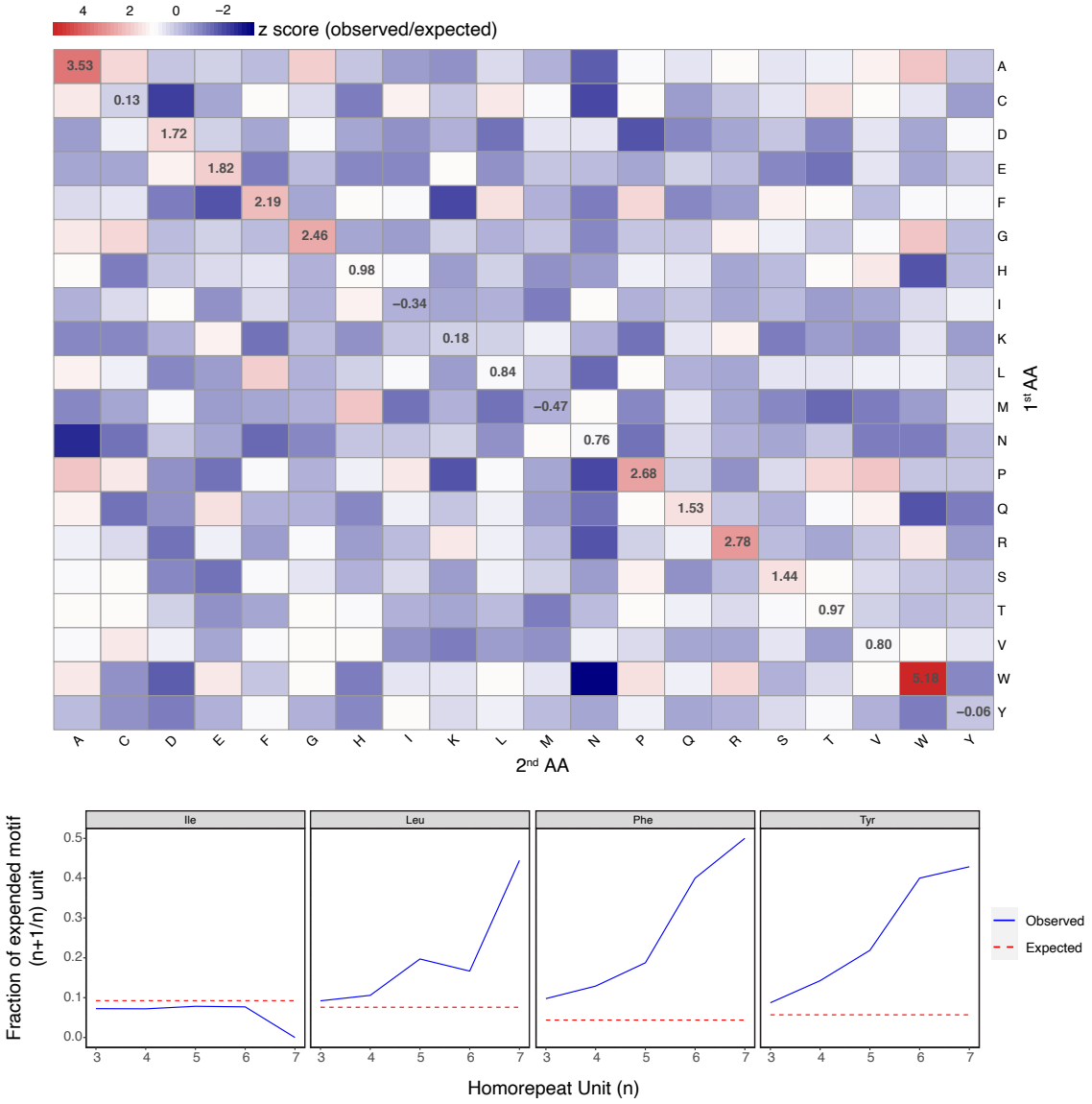
564 enriched (FDR<0.05) with upregulated genes, but many genes were found to be involved in
565 various metabolic pathways, potentially priming cells for rapid metabolic reactivation.
566
567
568



Extended Data Fig. 13

Transcript stabilization depends on the interplay between translation initiation and Ile codon-induced stalling.

(a) Anota2seq analysis of the polysome dataset using rlog normalization with default parameters. Anota2seq delineates different regulatory modes. Buffered (mRNA down) represents genes with a relative increase in ribosome loading but with no transcript-stabilizing effect. N=3 biological replicates. **(b)** Isoleucine and leucine usage in "buffered mRNA up" and "buffered mRNA down" transcripts. **(c)** (Left) KSS of "buffered mRNA up" and "buffered mRNA down" transcripts. The p-values of the Mann–Whitney U tests are indicated. (Right) The Kozak context of "buffered mRNA up" and "buffered mRNA down" transcripts. Kozak context: strong, RnnATGR; adequate, RnnATGY/YnnATGR; and weak, YnnATGY (ATG= start codon). The p-value of the Fisher's exact test is indicated.



583 Extended Data Fig. 14

584 **The expansion of Ile homorepeats is under purifying selection.**

585 (Upper) Heatmap showing the frequency of di-amino acid motifs relative to their expected
586 frequency. Values of (observed/expected frequency) were calculated for each di-amino acid
587 motif, and all values were Z-transformed to generate the heatmap. (Lower) The fraction of the
588 motifs retained in the successive additions of an extra residue. The dotted lines represent the
589 expected frequency based on the amino acid usage frequency in the genome. Four of the most
590 frequently used hydrophobic amino acids were analysed. Instead of being more likely to occur
591 due to replication slippage, both di-Ile and Ile homorepeat expansion were found to be less
592 frequent than expected.

593

594

595

596

597

598

Reference

1. Evans, M.E., Clark, W.C., Zheng, G., and Pan, T. (2017). Determination of tRNA aminoacylation levels by high-throughput sequencing. Nucleic Acids Res *45*, e133. 10.1093/nar/gkx514.
2. Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nat Methods *12*, 835-837. 10.1038/nmeth.3478.
3. Behrens, A., Rodschinka, G., and Nedialkova, D.D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Mol Cell *81*, 1802-1815.e1807. 10.1016/j.molcel.2021.01.028.
4. Xu, H., Yao, J., Wu, D.C., and Lambowitz, A.M. (2019). Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. Sci Rep *9*, 7953. 10.1038/s41598-019-44457-z.
5. Cozen, A.E., Quartley, E., Holmes, A.D., Hrabeta-Robinson, E., Phizicky, E.M., and Lowe, T.M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nat Methods *12*, 879-884. 10.1038/nmeth.3508.
6. Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res *44*, D184-189. 10.1093/nar/gkv1309.
7. Cooley, L., Appel, B., and Söll, D. (1982). Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA. Proc Natl Acad Sci U S A *79*, 6475-6479. 10.1073/pnas.79.21.6475.
8. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
9. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169. 10.1093/bioinformatics/btu638.
10. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120. 10.1093/bioinformatics/btu170.
11. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915. 10.1038/s41587-019-0201-4.
12. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.
13. Malleret, B., Claser, C., Ong, A.S., Suwanarusk, R., Sriprawat, K., Howland, S.W., Russell, B., Nosten, F., and Rénia, L. (2011). A rapid and robust tri-color flow cytometry assay for monitoring malaria parasite development. Sci Rep *1*, 118. 10.1038/srep00118.
14. Lacsina, J.R., LaMonte, G., Nicchitta, C.V., and Chi, J.T. (2011). Polysome profiling of the malaria parasite Plasmodium falciparum. Mol Biochem Parasitol *179*, 42-46. 10.1016/j.molbiopara.2011.05.003.
15. Altenhoff, A.M., Train, C.M., Gilbert, K.J., Mediratta, I., Mendes de Farias, T., Moi, D., Nevers, Y., Radoykova, H.S., Rossier, V., Warwick Vesztrocy, A., et al. (2021). OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. Nucleic Acids Res *49*, D373-D379. 10.1093/nar/gkaa1007.