Direct long read visualization reveals metabolic interplay between two antimalarial drug targets

Short title: metabolic connection drives adaptive evolution

Shiwei Liu[1, 2], Emily R. Ebel[3], Aleksander Luniewski[1], Julia Zulawinska[1], Mary Lewis Simpson[1], Jane Kim[1], Nnenna Ene[1], Thomas Werner Anthony Braukmann[4], Molly Congdon[5], Webster Santos[5], Ellen Yeh[4], Jennifer L. Guler[1*]

[1]University of Virginia, Department of Biology, Charlottesville, VA, USA

[2]Current affiliation: Indiana University School of Medicine, Indianapolis, IN, USA

[3]Stanford, Departments of Pediatrics and Microbiology & Immunology, Stanford, CA, USA

[4]Stanford University, Departments of Pathology and Microbiology & Immunology, Stanford, CA, USA

[5]Virginia Tech, Department of Chemistry, Blacksburg, VA, USA

*Corresponding Author

Email: jlg5fw@virginia.edu

Abstract

Increases in the copy number of large genomic regions, termed genome amplification, are an important adaptive strategy for malaria parasites. Numerous amplifications across the *Plasmodium falciparum* genome contribute directly to drug resistance or impact the fitness of this protozoan parasite. During the characterization of parasite lines with amplifications of the *dihydroorotate dehydrogenase* (*DHODH*) gene, we detected increased copies of an additional genomic region that encompassed 3 genes (~5 kb) including *GTP cyclohydrolase I* (*GCH1* amplicon). While this gene is reported to increase the fitness of antifolate resistant parasites, *GCH1* amplicons had not previously been implicated in any other antimalarial resistance context. Here, we further explored the association between *GCH1* and *DHODH* copy number. Using long read sequencing and single read visualization, we directly observed a higher number of tandem *GCH1* amplicons in parasites with increased *DHODH* copies (up to 9 amplicons) compared to parental parasites (3 amplicons). While all *GCH1* amplicons shared a consistent structure, expansions arose in 2-unit steps (from 3 to 5 to 7, etc copies). Adaptive evolution of *DHODH* and *GCH1* loci was further bolstered when we evaluated prior selection experiments; *DHODH* amplification was only successful in parasite lines with pre-existing *GCH1* amplicons. These observations, combined with the direct connection between metabolic pathways that contain these enzymes, lead us to propose that the *GCH1* locus is beneficial for the fitness of parasites exposed to *DHODH* inhibitors. This finding highlights the importance of studying variation within individual parasite genomes as well as biochemical connections of drug targets as novel antimalarials move towards clinical approval.

Author Summary

Malaria is caused by a protozoan parasite that readily evolves resistance to drugs that are used to treat this deadly disease. Changes that arise in the parasite genome, including extra copies of important genes, directly contribute to this resistance or improve how well the resistant parasite competes. In this study, we identified that extra copies of one gene (*GTP cyclohydrolase* or *GCH1*) were more likely to be found in parasites with extra copies of another gene on a different chromosome (*dihydroorotate dehydrogenase* or *DHODH*). A method that allows us to view long pieces of DNA from individual genomes was especially important for this study; we were able to assess gene number, arrangement, and boundary sequences, which provided clues into how extra copies evolved. Additionally, by analyzing previous experiments, we identified that extra *GCH1* copies improved resistance to drugs that target DHODH. The relationship between these two loci is supported by a direct connection between the folate and pyrimidine biosynthesis pathways that the parasite uses to make DNA. Since *GCH1* amplicons are common in clinical parasites worldwide, this finding highlights the need to study metabolic connections to avoid resistance evolution.

Introduction

Malaria is a disease caused by the protozoan *Plasmodium* parasite and *P. falciparum* is the leading cause of human malaria deaths (1). Due to the lack of effective vaccines against malaria infection, antimalarial drugs are the primary approach for malaria treatment (2). However, drug efficacy is mitigated by the frequent emergence of antimalarial resistant parasites (3).

63 Changes in the copy number of large genomic regions, termed copy number variations or CNVs,
64 are an important adaptive strategy for malaria parasites (4–12). Amplification, one type of CNV
65 with increased gene copy number, plays an essential role in the evolution of resistance to various
66 antimalarials (6,7,9,11,13–16). As one example, amplification of the *dihydroorotate*
67 *dehydrogenase* (*DHODH*) gene directly confers resistance to DHODH inhibitors (i.e. DSM1) in
68 parasites propagated in vitro (11). DHODH is an important enzyme in the *P. falciparum*
69 pyrimidine biosynthesis pathway that contributes resources for nucleic acid synthesis (17,18).
70 *DHODH* amplicons increase transcription and presumably translation of the drug target to directly
71 impact drug sensitivity (11).

72 In another example, amplification of the *GTP cyclohydrolase 1* (*GCH1*) gene increases the fitness
73 of clinical parasite populations that are resistant to antifolates (i.e. pyrimethamine and sulfadoxine)
74 (4,7,8,16). GCH1 is the first enzyme in the folate biosynthesis pathway and increased flux through
75 this pathway compensates for fitness costs of resistance-conferring mutations in *dihydropteroate*
76 *synthase* (*DHPS*) and *dihydrofolate synthase* (*DHFS*) genes (4,8,15). Although the contribution of
77 *GCH1* amplicons to antifolate resistance is well studied, this locus has not been reported to
78 contribute to resistance to other antimalarials (15,19,20).

79 Typically, gene copy number is studied using widely accessible high coverage short read
80 sequencing (11,21–24). However, this approach has limitations including non-unique mapping of
81 repetitive sequences, the inability to resolve complex genomic rearrangements, and the overall
82 difficulty of detecting structural variations (25–27). These challenges are exacerbated by the high
83 AT-content of the *P. falciparum* genome (28,29). Long read technologies such as Oxford
84 Nanopore sequencing have the potential to span low complexity and repetitive regions and more
85 accurately represent structural variation (Cretu Stancu et al., 2017; Sedlazeck et al., 2018a, 2018b;

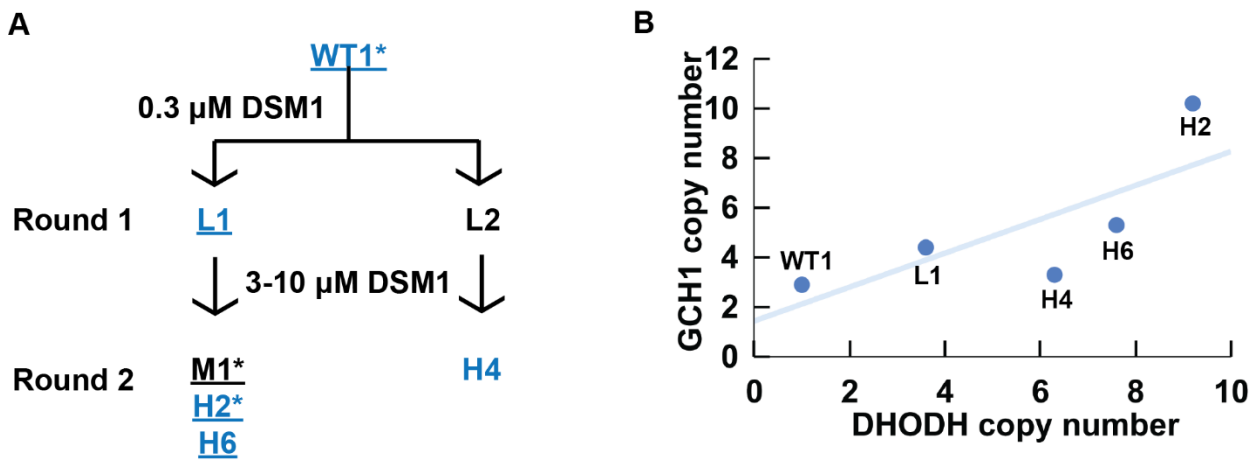
86 Ho et al., 2020). Moreover, this single molecule sequencing approach allows the examination of
87 clonal heterogeneity within cellular populations (34,35).

88 In this study, we identified an unprecedented association between *GCH1* and *DHODH* copy
89 numbers from a previously acquired short read data set. To explore this relationship further, we
90 performed long read sequencing and directly observed the step-wise expansion of the *GCH1*
91 amplicon in parasites with elevated *DHODH* copies. Using single long read visualization, we
92 determined that the structure and orientation of the amplicon were preserved during expansion and
93 amplicons increased in a step-wise manner. The positive correlation between *GCH1* and *DHODH*
94 copy number presented here, combined with the recognition that resistance to DHODH inhibitors
95 has only been selected in parasite lines with *GCH1* expansions, suggest the adaptive evolution of
96 these two genomic loci. In addition, the direct connection between folate and pyrimidine
97 biosynthesis supports metabolic interplay between the two pathways. Further study of this
98 relationship is important considering the potential use of DHODH inhibitors to treat clinical
99 malaria with preexisting *GCH1* amplicons.

100

101 **Results**

102 Through expanded analysis of short read data from a family of parasites selected with DSM1,
103 originally presented in (11) (**Figure 1A**), we observed a positive association between *GCH1* and
104 *DHODH* copy number (**Figure 1B**). Using droplet digital PCR on the same parasite lines that had
105 recently been propagated in our laboratory, we confirmed that *GCH1* copy number trended higher
106 as *DHODH* copy number increased (**Table 1**).



107

108 **Fig 1. *GCH1* copy number increase is positively correlated with *DHODH* copy number in**
109 **one family of DSM1 selected parasites. (A)** DSM1 selection schematic, as presented previously
110 (11). Blue text: Illumina short read sequenced lines. Underline: modern lines confirmed by ddPCR
111 analysis (Table 1). Asterisk (*): lines subjected to long read sequencing in this study. Wild-type
112 (WT1, *Dd2*) *P. falciparum* was selected with DSM1 in two steps; the first step selected for low-
113 level (L) resistance and the second step selected for moderate- (M) or high-level (H) resistance.
114 DSM1 EC50 values are as follows: L1 (1 µM), L2 (0.9 µM), M1 (7.2 µM), H2 (85 µM), H6 (56
115 µM), H4 (49 µM). All values were previously reported and clone names were adapted as
116 previously (11,36). **(B)** Relationship between *GCH1* and *DHODH* copy number in DSM1 selected
117 parasites as quantified using short read data from Guler et al. 2013. A trend line was added to show
118 the relationship between *GCH1* and *DHODH* copy numbers but a correlation coefficient could not
119 be calculated due to the small sample size (n=5) and dependence among the lines.

120

121

122 **Table 1. Positive *GCH1*:*DHODH* association is validated using Droplet Digital PCR on**
123 **modern parasite lines.**

Line	Sample type	Average copy number (SD)*			*GCH1* CN relative to parent	*DHODH* CN relative to parent
		GCH1	*DHODH*	*ATP6*		
WT1	Parent	2.5 (0.2)	0.8 (0.1)	1.1 (0.0)	1.0	1.0
L1	Round 1 selection	3.9 (0.3)	3.2 (0.1)	1.0 (0.1)	1.6	4
M1	Round 1 selection	3.9 (0.1)	5.2 (0.3)	1.0 (0.1)	1.6	6.5
H2	Round 2 selection	4.6 (0.2)	6.0 (0.3)	1.1 (0.1)	1.8	7.5
H6	Round 2 selection	4.3 (0.3)	7.0 (0.4)	1.0 (0.1)	1.7	8.8

124 *The average copy number is calculated by comparing the ddPCR signal to a single copy gene signal (*Seryl tRNA*
125 *synthetase*, PF3D7_0717700). N=4. *ATP6* copy number is a single copy gene expected to be 1 in all parasite lines.

126

127 We conducted long read sequencing to more precisely define *GCH1* copy number and amplicon
128 structure in the parental line versus two DSM1-selected lines (WT1, M1, and H2, **Table S1**). We
129 directly visualized single reads using an application that represents gene segments of individual
130 long reads (*Materials and Methods*). Small amplicons, like those including *GCH1*, are especially
131 conducive to this approach because long reads span multiple copies of the amplicon as well as
132 both up and downstream regions. Read visualization showed that the general structure of two 3-
133 gene amplicons, separated by an inversion of the same 3-genes, was conserved between the
134 parental and selected lines (**Figure 2A-C**). This amplicon structure was reported previously in the
135 WT1 (*Dd2*) parental line (**Figure 2D**) (4).

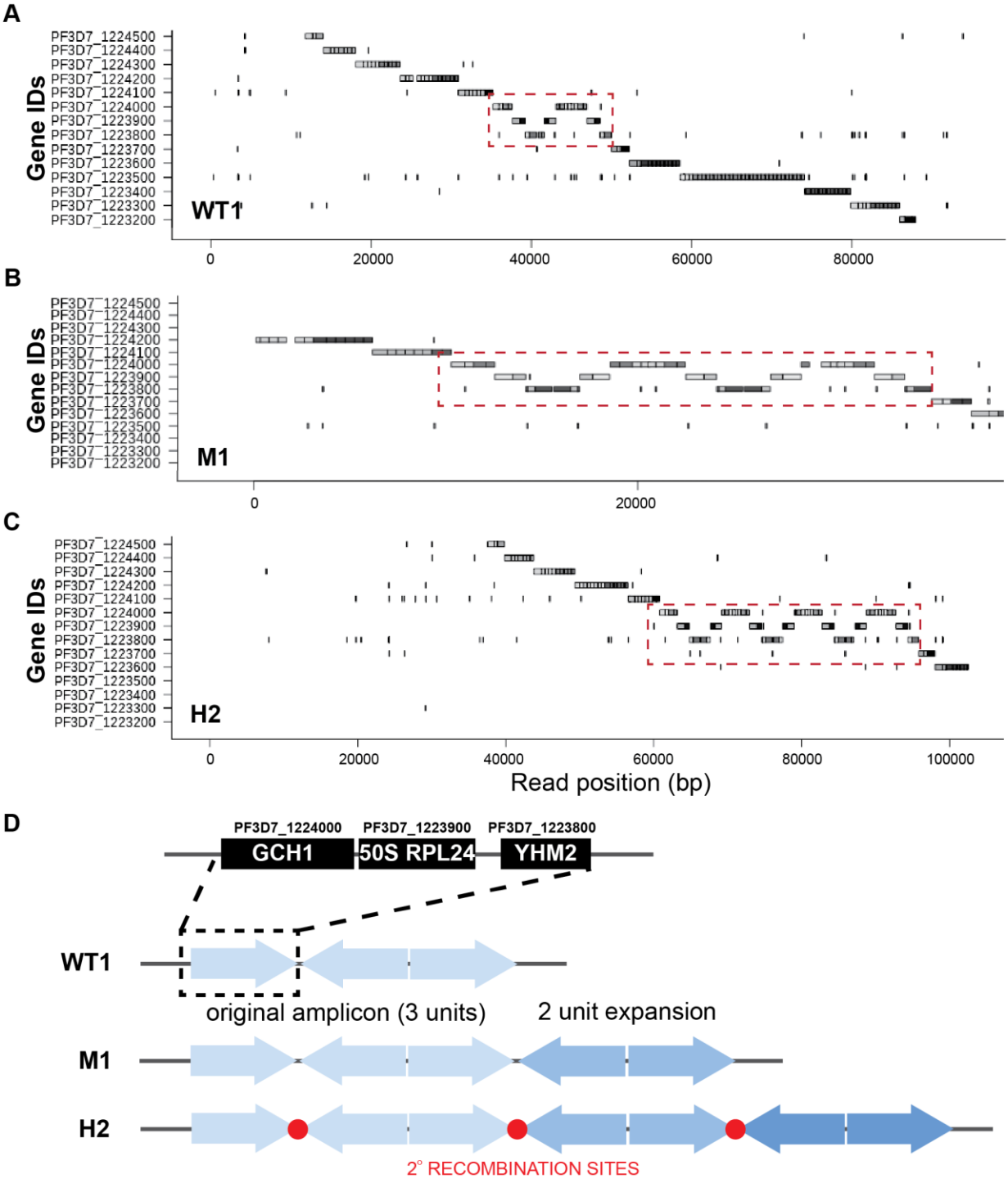


Fig 2. Long-read visualization shows conserved boundaries and structure of *GCH1* amplicon from parental and resistance parasites. (A-C) Representative images from the Shiny application comparing the *GCH1* amplicon in WT1 (**A**), M1 (**B**), and H2 (**C**) spanning reads. Red dashed square: copies of *GCH1* amplicon covering 3 genes. Each gene sequence from the *3D7* reference genome (no *GCH1* amplicon represented) was split into <=500 bp fragments and blasted against individual Nanopore reads (dark gray: genic regions, light grey: intergenic regions, more details

143 in *Materials and Methods*). **(D)** Orientation of *GCH1* amplicons from each parasite line. The three
144 genes within the *GCH1* amplicon unit include PF3D7_1224000 (*GTP cyclohydrolase I*, *GCH1*),
145 PF3D7_1223900 (*50S ribosomal protein L24*, putative, *50S RPL24)*, and PF3D7_1223800
146 (*citrate/oxoglutarate carrier protein*, putative, *YHM2*). Note: The gene order (*GCH1-50S RPL24-*
147 *YHM2*) is reversed compared to the 3D7 reference genome (*YHM2-50S RPL24-GCH1*) in order to
148 facilitate comparison with the read images in panels A-C. Red circles: amplicon junction
149 sequences that act as potential secondary recombination sites.

150

151 Using direct read visualization, we also recorded the number of *GCH1* amplicon units (**Figure 2D**)
152 from both spanning and non-spanning reads (**Figure 3A, Table S2**) Spanning reads cover genes
153 both upstream and downstream of the amplicon while non-spanning reads end within the amplicon.
154 We consistently detected up to 3 amplicon units on reads from the parental line. Although the
155 overall mean copy number was similar for all parasite lines (**Figure 3A**, WT1, 2.5 copies; M1, 3.6
156 copies; H2; 3.2 copies, $F(2,95)=[2.61]$, $p = 0.07$), we observed more reads that encompassed a
157 higher number of *GCH1* amplicon units in the selected lines (up to 9 units, **Figure 3A**). When we
158 restricted our analysis to spanning reads, or those that displayed the entire amplified region, we
159 detected a significant difference in mean copy number between the parasite lines (WT1 vs M1,
160 $p=0.0003$ and WT1 vs H2, $p=0.01$; **Figure S1**). The presence of longer reads was not contributing
161 to this observation; there was not a significant difference in the length of reads used for *GCH1*
162 analysis from each parasite line ($F(2,96)=[2.51]$, $p = 0.09$, **Figure 3B**, median read lengths: WT1,
163 51144bp; M1, 49018bp; H2, 26429bp). The larger variation observed for the M1 sample (**Figure**
164 **3B**) was due to an improvement in ONT technology and methods leading to an increase in the
165 number of longer reads (i.e. WT1 N50 of 15.6kb on R9.4.1 flow cell and M1 N50 of 99.7kb on
166 R10.4.1 flow cell, **Table S1,** see *Materials and Methods*).

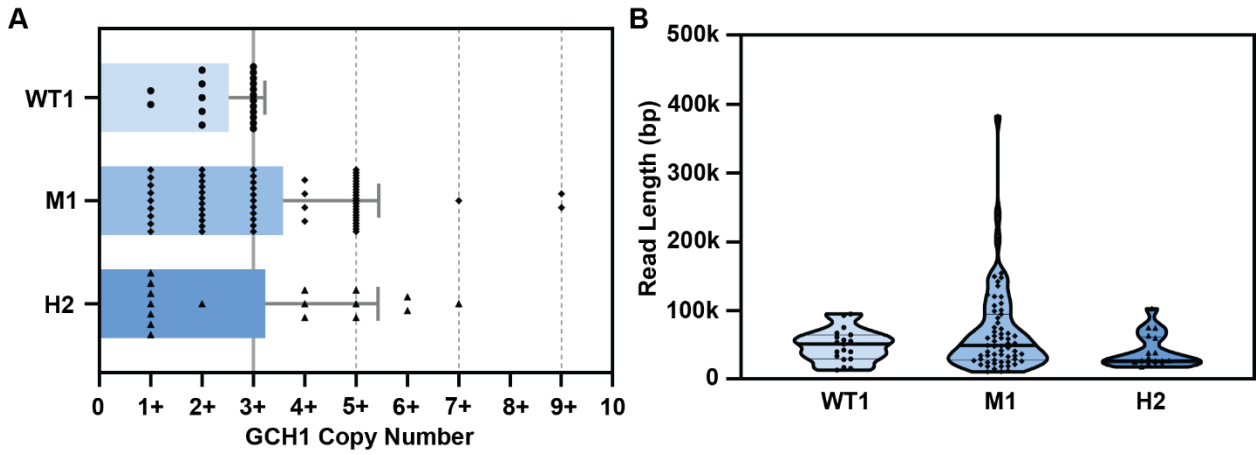


Fig 3. Quantification of long reads displays an increase in *GCH1* amplicon number in DSM1 selected parasites that is not dependent on overall read length. (**A**) *GCH1* copy number from WT1 and selected (M1 and H2) parasite lines. Spanning and non-spanning reads are included; spanning reads were grouped with their corresponding non-spanning count (e.g. a read showing 2 copies of the *GCH1* amplicon is counted as a 2+ read). Bars represent mean, error bars represent standard deviation (dotted lines, values that represent detected spanning reads, see **Figure S1**). (**B**) Read length distribution from long reads (>=10kb) covering the *GCH1* amplicon**.** Thick line represents median, thin line represents quartiles.

During this analysis, we identified that >50% of reads from M1 and H2 lines depicted amplicon copy numbers greater than the expected WT copy number (**Figure 3A**, **Table S2**). To assess whether variation in the copy number of the *GCH1* locus was common in other laboratory-adapted parasite lines, we evaluated several additional long read datasets (**Table S2**, *Supporting Information,* Vembar et al., 2016). In general, parasite lines exhibited the expected *GCH1* amplicon size (i.e. *Dd2* versus *3D7*, as reported previously (4) and visualized in **Figures 2D** and **S2**) and the amplicon copy number for each line was relatively stable across diverse datasets (limited to <25% of reads above the expected copy number).

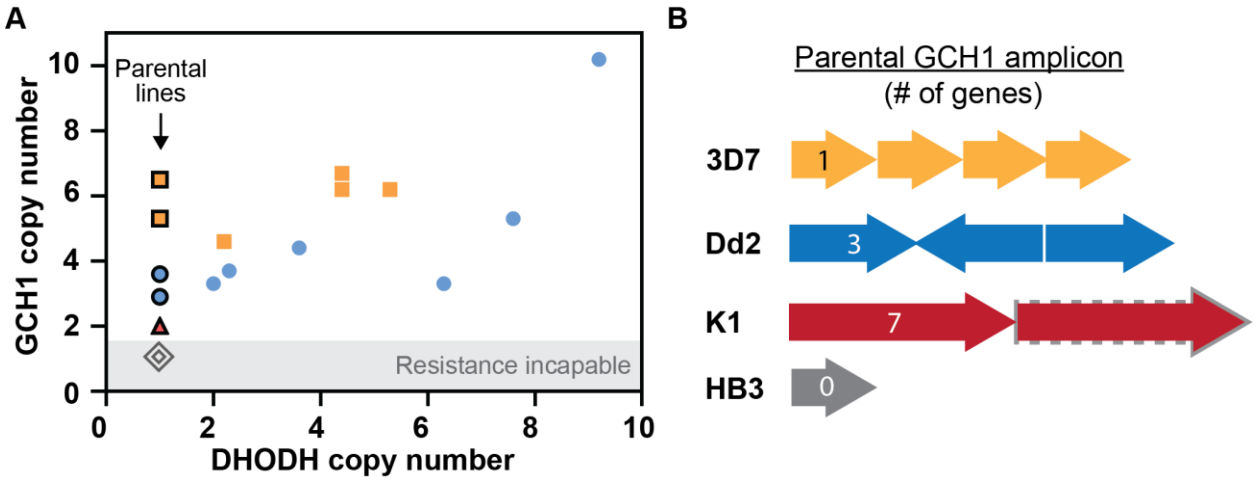
When we evaluated the amplicon structure, all M1 and H2 reads covering the *GCH1* region began with a set of 3 amplicon units, as observed in the parental line (**Figure 2A**). This basic structure was followed by groups of 2 units, where one copy was inverted and the other was in normal orientation (**Figure 2D**). This suggests that increases in copy number were arising as two unit

189 steps. We identified further evidence of the two unit amplicon expansion when we limited our
190 analysis to only spanning reads, which allow direct quantification of amplicon copy number. Here,
191 we did not detect any spanning reads that carried *even GCH1* copy numbers (only 3, 5, 7, or 9
192 copies, **Figure S1**).

193 To investigate whether an increase in *GCH1* amplicons was contributing generally to resistance to
194 DHODH inhibitors, we evaluated the copy number of *GCH1* amplicons using short read data
195 collected from parasites selected with other DSM compounds in two studies (18,37). Contrary to
196 DSM1-selected parasites (**Figure 1B**), we did not detect increases in *GCH1* amplicon number in
197 parasites selected with DSM265, DSM267, or DSM705 compared to parental *Dd2* or *3D7* parasite
198 lines (**Figure S3**, **Table S3**). For this analysis, we only included parasite lines that acquired
199 *DHODH* amplicons to confer resistance.

200 The absence of a correlation between *DHODH* and *GCH1* copy number in these other datasets
201 suggests that *GCH1* may not play a direct role in resistance. However, parental lines used in these
202 two studies harbored pre-existing *GCH1* amplicons (~3-6 copies, **Figure 4A**), indicating that this
203 locus may support the evolution of resistance involving *DHODH* amplification. Therefore, we
204 compiled information on all previous resistance selections conducted in parasite lines that naturally
205 vary in *GCH1* copy number (7 total studies, **Table 2**). *3D7* and *Dd2* parasite lines, which carry the
206 highest *GCH1* copy numbers (3-6 copies) with the smallest overall amplicon sizes of common
207 laboratory parasites (1 and 3 genes, respectively, **Figure 4B**), were most commonly used in
208 successful resistance selections (6 and 4 studies, respectively, **Table 2**) and were capable of
209 acquiring *DHODH* amplicons (**Figure 4A**). *K1* parasites harbor a larger low copy *GCH1* amplicon
210 (anticipated as 2 copies of 7 genes, **Figure 4B**) and although resistance could be selected in one
211 study, it was 1000-fold less efficient than *Dd2* selections run in parallel (38). The *DHODH* copy

212 number status of this K1-derived selected line was not investigated. Finally, *Hb3*, which harbors
213 one *GCH1* copy (**Table 2, Figure 4B**), was incapable of developing resistance in two independent
214 studies (11,38) (**Figure 4A**).



215
216 **Fig 4. Preexisting *GCH1* amplicons confer resistance competence.** (**A**) Relationship between
217 *GCH1* and *DHODH* copy number in parental (black outline) and DHODH inhibitor-selected
218 parasite lines as quantified using short read data (see **Table 2** for source data). Each data point
219 represents a copy number from an individual parasite clone or line from a distinct selection (see
220 **Tables S3** and **S4** for copy number assessment). *3D7*, yellow squares (2 selections); *Dd2*, blue
221 circles (2 selections); *K1*, red triangle (1 selection, unknown *DHODH* copy number post-
222 selection); *HB3*, grey diamonds (2 selections). (**B**) *GCH1* amplicon size and gene number vary
223 between parasite lines used for DHODH inhibitor selections (see **Table 2** for selection details).
224 *3D7*, 4 copy, single gene (PF3D7_1224000, 2kb, this study); *Dd2*, 3 copy, three genes
225 (PF3D7_1223800- PF3D7_1224000, 5kb, this study); *K1*, unknown copy number (grey dash,
226 anticipated 2 copy), 7 genes (PF3D7_1223500 - PF3D7_1224100, 19kb) (39); *HB3*, 1 copy of
227 region (40).

228
229

230 **Table 2. Parental line *GCH1* copy number determines resistance competence**

Study (author, year)	Selection Compounds	Success of Resistance Selection				*DHODH* amplicon
		Dd2 (*GCH1*=3)	*3D7* (*GCH1*=4-6)	*Hb3* (*GCH1*=1*)	Other lines	
Guler, 2013 (11)	DSM1	**Yes**	N/A	**No**	N/A	Yes
Palmer, 2021 (37)	DSM705	**Yes**	**Yes**	N/A	N/A	Yes
Phillips, 2015 (38)	DSM265	**Yes**	N/A	**No**	K1-Yes**	Yes
Mandt, 2019 (18)	DSM265, DSM267	**Yes**	**Yes**	N/A	N/A	^Yes
White, 2019 (41)	DSM265	**Yes**	N/A	N/A	N/A	No
Ross, 2014 (42)	DSM74	**Yes**	**Yes**	N/A	N/A	No
Lukens, 2014 (43)	Genz-666136, DSM74	N/A	**Yes**	N/A	N/A	No

231 *The *GCH1* copy number for the *HB3* parasite line is single copy based on whole genome sequencing (40). Prior
232 studies detected two *GCH1* copies using quantitative PCR but did not detect an increase in expression from this region
233 (8,15).
234 **The *K1* parasite line has a *GCH1* amplicon that is ~19kb (961,455–980,630, (39)) but the exact copy number has
235 not been reported. Phillips et al. reported that it was 100x more difficult to select for DSM265 resistance using *K1*
236 than *Dd2* (MIR- 2 x 10^8 versus 2 x 10^6, respectively, (38). Resistance mechanisms (i.e. mutation or amplification)
237 were not evaluated for the resistant *K1* line in this study.
238 ^Although both *Dd2* and *3D7* were used in selections, only two *3D7* clones acquired *DHODH* amplifications in this
239 study (18) and were evaluated for *GCH1* copy number in our study (**Tables S3 and S4**).
240

241 **Discussion**

242 Using both short-read sequencing as well as an accurate PCR-based method, ddPCR, we
243 uncovered that *GCH1* copy number is positively associated with *DHODH* copy number in malaria
244 parasites selected with DSM1 (**Figures 1B and 4, Table 1**). However, limitations in each of these
245 methods contribute to inaccuracies in copy number results. Most importantly, both of these
246 methods provide an average value for the parasite population and do not reflect variation from
247 individual genomes. Long-read sequencing combined with a custom visualization tool (44)
248 allowed us to directly quantify the copy number of the *GCH1* locus and assess the structure of all
249 amplicons (**Figures 2** and **3**). In the context of this haploid, single cellular parasite, each read
250 represents a distinct DNA strand from a single genome; this approach is an accurate way to
251 visualize copy number and junction heterogeneity across a population (**Figure 2**, **Figure S2**). We
252 showed that the detection of more *GCH1* amplicons on reads from DSM1 selected parasite lines

253 (**Figure 3A**) was not caused by sample preparation (**Figure 3B**, **Table S2**) or natural variation
254 during in vitro culture (**Table S2**). The visualization of tandem amplicons on single reads (**Figure**
255 **2A-C**) and the step-wise increase by 2 amplicon units (from 3 to 5 to 7 copies, **Figures 2D** and
256 **S1**) provides direct evidence for *GCH1* amplicon expansion in the DSM1 resistance context.

257 The two-unit steps suggest that there is a secondary recombination site that leads to facile re-
258 amplification of the region (**Figure 2D**). In the high coverage M1 sample, we identified long AT-
259 rich sequences at the end of the amplicon unit (**Figure S4**) that may participate in further DNA
260 breakage (AT dinucleotides at the *GCH1* end, **Figure S4C**) and recombination after the initial
261 event (homopolymeric A/T tracks at the *YHM2* end, **Figure S4A**)(24). Multiple junctions at these
262 positions provide evidence of the initial event that formed the 3 unit amplicon as well as a
263 secondary event resulting in a two-unit step-wise expansion.

264 As recently recognized in organisms from bacteria to humans (34,35,45–47), long read sequencing
265 is an exceptional tool to investigate genomic heterogeneity within populations. This method is not
266 without its limitations; reads that end within the amplicon region lead to an underestimation in the
267 total copy number; in our study, 65/98 (66%) of reads were non-spanning, which contributed to a
268 similar overall mean copy number between parasite lines (**Figure 3A**). However, the visualization
269 of spanning reads overcame this limitation, clearly showing the two-unit expansion of *GCH1*
270 amplicons (**Figures 2A-C** and **S1**). The quantification of spanning reads also confirms that
271 variation in *GCH1* amplicon numbers likely arises stochastically in the parasite population; while
272 most spanning reads from the high coverage M1 line showed 5 amplicon copies (15/21 reads), we
273 observed higher amplicon copies on a minority of reads (7 and 9 copies, 3/21 reads). In line with
274 our previous model for malaria genome evolution (11,24), parasites with higher copy numbers are
275 present within the population, poised to be selected under specific conditions.

276 The *GCH1* locus is a proven hotspot for genetic change (8,39). *GCH1* amplicons have previously
277 been associated with antifolate resistance; an increase in flux through the folate biosynthesis
278 pathway via *GCH1* alleviates fitness effects of mutations that confer pyrimethamine and
279 sulfadoxine resistance (4,8,16,19,20). A similar contribution to resistance to DHODH inhibitors
280 would not be surprising, given the close connection of the folate and pyrimidine biosynthesis
281 pathways (**Figure 5**); they both contribute to nucleotide biosynthesis and converting dUMP to
282 dTMP in pyrimidine biosynthesis requires reciprocal conversion of N^5, N^{10}-methylene-THF to
283 DHF by the folate pathway. Strong evidence for this metabolic connection comes from the
284 observation that parasite lines without *GCH1* amplicons are not able to develop resistance to
285 DHODH inhibitors (**Table 2**) (11,38). Assessment of resistance competence, combined with our
286 observations in the DSM1 context and folate-pyrimidine metabolic interplay, lead us to propose
287 that increases in *GCH1* copy number are also beneficial for the fitness of parasites carrying
288 *DHODH* amplicons. During resistance evolution, increased DHODH protein levels interact with
289 more of the inhibitor to cause resistance. When the inhibitor is not present, higher DHODH levels
290 increase flux through pyrimidine biosynthesis, which in turn requires more THF for the completion
291 of the pathway (**Figure 5**).

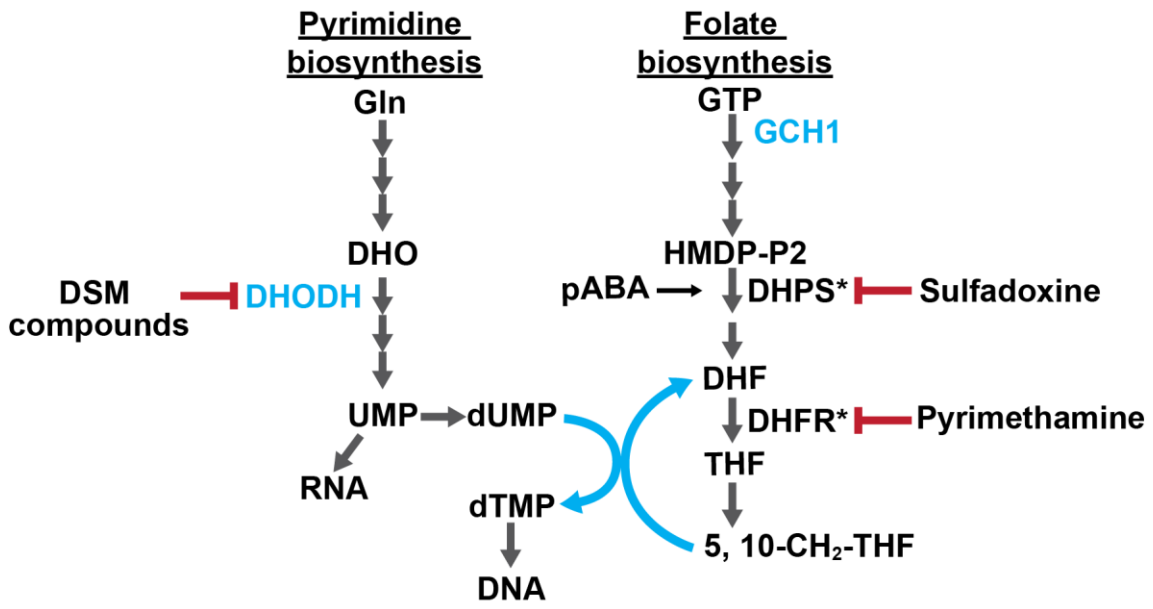


Fig 5. The connection between pyrimidine and folate biosynthesis pathways. Enzymes with gene copy number variations are indicated in blue text (DHODH: dihydroorotate dehydrogenase; GCH1: GTP cyclohydrolase 1) and connections between the two pathways in blue arrows. Gln: Glutamine; DHO: Dihydroorotate; UMP: Uridine monophosphate; dUMP: Deoxyuridine monophosphate; dTMP: deoxythymidine monophosphate; GTP: Guanosine-5'-triphosphate; DHPS: Dihydropteroate synthase; DHFR: Dihydrofolate reductase; DHF: Dihydrofolate; THF: Tetrahydrofolate; HMDP-P2: 6-hydroxymethyl-7, 8-dihydropterin diphosphate; pABA: para-amino-benzoic acid; 5, 10-CH2-THF: 5,10-Methylenetetrahydrofolate. **Dd2* carries 3 mutations in both DHPS and DHFR (6 total) and HB3 and 3D7 have a single mutation each in DHFR and DHPS, respectively (15,48).

There are a few factors that may impact parasite reliance on folate biosynthesis including parasite genetic background and the presence of folate precursors in the host environment. Antifolate resistant parasites from different regions of the world that harbor specific mutations rely more heavily on increased GCH1 flux to alleviate fitness effects (i.e. DHFR and DHPS mutants, (4,8,15,19), **Figure 5**). While *Dd2* and *3D7* have distinct origins and mutational patterns (49,50) (**Figure 5**), the fact that they both maintain the ability to develop DHODH inhibitor resistance excludes the possibility that DHFR and DHPS mutations matter in this context. Additionally, levels of folate precursors, such as para-aminobenzoic acid (pABA), vary widely in host versus in

312 vitro environments (51–55) and exert different selective pressures on parasites during drug
313 selection (56). Higher pABA levels in human serum, as were used during DSM1 selections (11),
314 may drive the accumulation of higher *GCH1* copy numbers and hold extra benefits by bolstering
315 both folate and pyrimidine biosynthesis (**Figure 5**). Our observation of some level of variation in
316 *GCH1* amplicon number in standard parasite lines (**Table S2**) suggests that differing in vitro media
317 formulations may drive copy number changes at this locus.

318 Importantly, DHODH inhibitors have entered trials for the treatment of clinical malaria (18) and
319 resistance-conferring *DHODH* amplicons can evolve in vivo (57). Meanwhile, antifolate
320 resistance and *GCH1* amplifications are widespread in clinical *P. falciparum* isolates, especially
321 in Southeast Asia (4,7,8,39) and now Africa (16,58). Additionally, a zoonotic malaria species, *P.*
322 *knowlesi,* also acquires extra *GCH1* copies after adaptation to human blood (59). The potential use
323 of DHODH inhibitors on clinical parasites carrying *GCH1* amplicons necessitates further
324 evaluation of the interplay between these two important metabolic pathways.

325

326 **Materials and Methods**

327 *DSM1 and parasite clones*

328 DSM1 inhibits dihydroorotate dehydrogenase enzyme of *Plasmodium* pyrimidine biosynthesis
329 (60). DSM1-resistant parasites were selected during a previous study and cloned by limiting
330 dilution at the time of isolation (11) (**Figure 1A**). The naming scheme representing low (L),
331 moderate (M), and high (H) levels of DSM1 resistance follows that used previously by our group
332 (36).

333 *Parasite culture*

334 We thawed erythrocytic stages of *P. falciparum* (*Dd2*, MRA-150; *3D7*, MRA-102, Malaria
335 Research and Reference Reagent Resource Center, BEI Resources and DSM1 selected clones as
336 highlighted in **Figure 1A**, a generous gift from Pradipsinh Rathod, University of Washington)
337 from frozen stocks and maintained them as previously described (61). Briefly, we grew parasites
338 at 37°C in vitro at 3% hematocrit (serotype A positive human erythrocytes, Valley Biomedical,
339 VA or BioIVT, NY) in RPMI 1640 medium (Invitrogen, USA) containing 24 mM NaHCO3 and
340 25 mM HEPES, and supplemented with 20% human type A positive heat inactivated plasma
341 (Valley Biomedical, VA or BioIVT, NY) in sterile, sealed flasks flushed with 5% O_2, 5% CO_2,
342 and 90% N_2 (11). We maintained the cultures with media changes every other day and sub-cultured
343 them as necessary to keep parasitemia below 5%. We determined all parasitemia measurements
344 and staging using SYBR green-based flow cytometry (62). We routinely tested cultures using the
345 LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich, USA) to confirm negative
346 Mycoplasma status.

347 *Short read sequencing analysis and CNV detection*

348 We analyzed amplifications in Illumina short read datasets of *P. falciparum* parasites selected by
349 different DSM antimalarial drugs (DSM1, DSM265, DSM267, and DSM705, **Table S3**)
350 (11,18,37). We first processed and mapped the reads to the reference genome as previously
351 described (24,63). Briefly, we trimmed Illumina adapters from reads with the BBDuk tool in
352 BBMap (version 38.57) (64). We aligned each fastq file to the *3D7* reference genome with
353 Speedseq (version 0.1.0) through BWA-MEM alignment (65). We discarded the reads with low-
354 mapping quality scores (below 10) and removed duplicated reads using Samtools (version 1.10)

355 (66). We analyzed split and discordant reads from the mapped reads using LUMPY in Speedseq
356 to determine the location and length of the previously reported *GCH1*, *DHODH*, and multidrug
357 resistance protein 1 (*MDR1*) amplicons (**Table S4**) (67). For read-depth analysis, we further
358 filtered the mapped reads using a mapping quality score of 30. To determine the copy number of
359 the *GCH1*, *DHODH*, and *MDR1* amplicons, we used CNVnator (version 0.4.1) with a bin size of
360 100 bp; the optimal bin size was chosen to detect GCH1 amplicons in all analyzed samples (68).

361 *Droplet Digital PCR*

362 Prior to Droplet Digital (dd) PCR, we digested DNA with restriction enzyme RsaI (Cut Site:
363 GT/AC) following the manufacturer's instructions (New England Biolabs, Ipswich, MA, USA) in
364 37°C incubation for one hour. We selected the restriction enzyme RsaI to cut outside of the ddPCR
365 amplified regions of desired genes and separate amplicon copies to be distributed into droplets.
366 We diluted the digested DNA for ddPCR reactions. We performed ddPCR using ddPCR Supermix
367 for Probes (no dUTP, Bio-Rad Laboratories, California, USA) with DNA input 0.1 ng (in duplicate
368 per sample), 0.025 ng (in duplicate per sample) as previously described (36). The primers and
369 probes used in reactions are included in **Table S5**. The PCR protocol for probe-based assay was
370 95°C for 10 min, followed by 40 rounds of 95°C for 30 sec and 60°C for 1 min. *Seryl-tRNA*
371 *synthetase* and *calcium-transporting ATPase* (*ATP6*) served as single copy reference genes on
372 chromosome 7 and chromosome 1 respectively; *dihydroorotate dehydrogenase* (*DHODH*) and
373 *GTP cyclohydrolase 1* (*GCH1*) are multi-copy genes (**Table S5**). We performed droplet generation
374 and fluorescence readings per the manufacturer's instructions. For each reaction, we required a
375 minimum number of 10,000 droplets to proceed with analysis. We calculated the ratio of positive
376 droplets containing a single- (*ATP6*) or multi-copy gene (*GCH1*, *DHODH*) versus a single-copy

377 gene (*seryl-tRNA synthetase*) using the Quantasoft analysis software (QuantaSoft Version 1.7,
378 BioRad Laboratories) and averaged between independent replicates.

379 *DNA extraction for long read sequencing*

380 We lysed asynchronous *P. falciparum*-infected erythrocytes with 0.15% saponin (Sigma-Aldrich,
381 USA) for 5 min at room temperature and washed them three times with 1× PBS (diluted from 10×
382 PBS Liquid Concentrate, Gibco, USA). We then lysed parasites with 0.1% Sarkosyl Solution
383 (Bioworld, bioPLUS, USA) in the presence of 1 mg/ml proteinase K (from *Tritirachium album*,
384 Sigma-Aldrich, USA) overnight at 37°C. We first extracted nucleic acids with
385 phenol/chloroform/isoamyl alcohol (25:24:1) pH 8.0 (Sigma-Aldrich, USA) three times using
386 1.5ml light Phase lock Gels (5Prime, USA), then further extracted nucleic acids with chloroform
387 twice using 1.5ml light Phase lock Gels (5Prime, USA). Lastly, we precipitated the DNA with
388 ethanol using the standard Maniatis method (69). To obtain high molecular weight genomic DNA,
389 we avoided any pipetting during the extraction, transferred solutions by direct pouring from one
390 tube to another, and mixed solutions by gently inverting the tubes.

391 *Oxford Nanopore long read sequencing and analysis*

392 For WT1 and H2 samples, we subjected ~1µg of high molecular weight genomic DNA to library
393 preparation for Oxford Nanopore sequencing following the Nanopore Native barcoding genomic
394 DNA protocol (version: NBE_9065_v109_revAB_14Aug2019) with 1x Ligation Sequencing kit
395 (SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK). We performed DNA repair and
396 end preparation using NEBNext FFPE DNA Repair Mix (New England Biolabs, Ipswich, MA,
397 USA) and NEBNext End Repair/dA-Tailing Module (New England Biolabs, Ipswich, MA, USA).
398 We cleaned the A-tailed fragments using 0.9X AMPure XP beads (Beckman Coulter, High

399 Wycombe, UK). We then ligated barcodes to the end-prepped DNA using the Native Barcoding
400 Expansion 1–12 kit (EXP-NBD104, Oxford Nanopore Technologies, Oxford, UK) and Blunt/TA
401 Ligase Master Mix (New England Biolabs, Ipswich, MA, USA). We cleaned the barcoded samples
402 using 0.9X AMPure XP beads. Then we pooled barcoded samples in equimolar ratios and
403 subjected them to an adaptor ligation step, using the Adapter Mix II from the Native Barcoding
404 Expansion 1–12 kit and NEBNext Quick Ligation Reaction Buffer (New England Biolabs,
405 Ipswich, MA, USA) as well as Quick T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA).
406 After adaptor ligation, we cleaned the WT1 and H2 libraries using AMPure XP beads.

407 For M1 samples, we isolated high molecular weight genomic DNA from 10 flasks of parasites (1-
408 4% parasitemia, ~30-40% late stage parasites) and prepared libraries as above except that we used
409 the cleanup and precipitation reagents provided in the Ultra-long DNA Sequencing kit (Oxford
410 Nanopore Technologies, Oxford, UK, V14). We quantified the adapter-ligated and barcoded DNA
411 using a Qubit fluorimeter (Qubit 1X dsDNA High Sensitivity Assay Kit, Life Technologies,
412 Carlsbad, CA).

413 We sequenced the WT1 and initial H2 libraries using the R9.4.1 flow cell (FLO-MIN106D),
414 another H2 library using the R10 flow cell (FLO-MIN111), and M1 libraries using the R10.4.1
415 flow cell (FLO-MIN114) on MinION (Oxford Nanopore Technologies, Oxford, UK). For WT and
416 H2 samples, we ran flow cells for 48 hours (controlled and monitored using the MinKNOW
417 software 3.6.5). For the M1 samples, we ran one-third of the library sequentially on the flow cell
418 for 24 hours each (up to 72 hours in total).

419 For base calling and demultiplexing of the Nanopore sequencing reads, we used Guppy (version
420 3.4.5+fb1fbfb) with the parameter settings “-c dna_r9.4.1_450bps_hac.cfg --barcode_kits "EXP-

421 NBD104" -x auto" for samples sequenced with R9.4.1 flow cell and "-c dna_r10_450bps_hac.cfg
422 -x auto" for the sample sequenced with R10 flow cell. We checked the read length and read quality
423 using "Nanoplot" (version 1.0.0) (see **Table S1**). We trimmed the adapters with "qcat" (version
424 1.1.0) (70) and filtered the reads with a cutoff "length ≥ 500 and Phred value ≥ 10" using the
425 program " filtlong version 0.2.0" (https://github.com/rrwick/Filtlong). To estimate the coverage of
426 sequencing reads in each sample, we aligned the filtered reads to *3D7* reference genome
427 (PlasmoDB) using "minimap2" (version 2.17) (71). "QualiMap" (version 2.2.1) (72) was used to
428 calculate the coverage of the aligned reads (see **Table S1**).

429 For junction analysis, we aligned Q10-filtered M1 reads (.bam) to the *3D7* reference genome
430 (PlasmdDB) and visualized read sequence at junctions of the GCH1 amplicon using Integrated
431 Genome Viewer (73,74).

432 *Direct visualization of long reads*

433 To visualize structural variants in the parasite genome, we used a custom R Shiny script to plot the
434 arrangement of reference gene segments on individual Nanopore reads (44). Briefly, we defined a
435 target region in the *3D7* reference genome (chromosome 12: 932916bp – 999275bp) that contained
436 3 genes in the *GCH1* amplicon and 11 flanking genes. We extracted the reference sequences of
437 these genes and subsequent intergenic regions, then split these sequences into fragments of 500-
438 1000 base pairs. We compared these fragments to individual Nanopore reads (>10kb) using
439 BLAST (75). We used the BLAST output as input for a custom R script, which drew blocks
440 representing homology between the defined genes (y-axis) and each individual read (x-axis). If
441 genes on the read were arranged in the same order as the reference genome then the blocks would
442 create a diagonal line on the plot. If there was a CNV that altered the gene order or direction then

443 the diagonal line is disrupted. We allowed the percent identity required to draw a homologous
444 rectangle to vary between reads, which varied in quality, using a slider in the Shiny application.
445 To filter out long reads with potentially spurious hits to gene fragments, we also used BLAST to
446 compare Nanopore reads to the reference genome and removed reads with <90% identity to
447 chromosome 12. To compare the mean copy number of GCH1 amplicon and read lengths between
448 WT1, M1, and H2 reads, we performed a one-way ANOVA in Graphpad Prism with an alpha
449 value of 0.05. If significant, differences were further explored using Tukey's multiple comparisons
450 testing.

451

452 **Acknowledgments**

453 Our thanks to Michelle Warthan, Noah Brown, and Ali Guler (University of Virginia) for
454 experimental, analysis, and statistical support, Martin Wu and Mercedes Campos-Lopez
455 (University of Virginia) for Oxford Nanopore sequencing support, and the laboratory of Dr.
456 William Petri Jr (University of Virginia) for their helpful discussions and insight. We are grateful
457 to David Fidock (Columbia University) and Dyann F. Wirth (Harvard University) for sharing
458 sequencing data and information on DSM-selected cell lines.

459

460 **References**

461 1. Rich SM, Leendertz FH, Xu G, LeBreton M, Djoko CF, Aminake MN, et al. The origin of
462 malignant malaria. Proc Natl Acad Sci U S A. 2009 Sep 1;106(35):14902–7.

463 2. Casares S, Brumeanu TD, Richie TL. The RTS,S malaria vaccine. Vaccine. 2010 Jul
464 12;28(31):4880–94.

465 3. Blasco B, Leroy D, Fidock DA. Antimalarial drug resistance: linking Plasmodium
466 falciparum parasite biology to the clinic. Nature medicine. 2017 Aug 4;23(8):917–28.

467 4. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, Zhou Y, et al. A Systematic Map
468 of Genetic Variation in Plasmodium falciparum. PLOS Pathogens. 2006;2(6):e57.

469 5. Conway DJ. Molecular epidemiology of malaria. Clin Microbiol Rev. 2007 Jan;20(1):188–
470 204.

471 6. Hyde JE. Drug-resistant malaria - an insight. FEBS J. 2007 Sep;274(18):4688–98.

472 7. Ribacke U, Mok BW, Wirta V, Normark J, Lundeberg J, Kironde F, et al. Genome wide
473 gene amplifications and deletions in Plasmodium falciparum. Molecular and Biochemical
474 Parasitology. 2007 Sep 1;155(1):33–44.

475 8. Nair S, Miller B, Barends M, Jaidee A, Patel J, Mayxay M, et al. Adaptive Copy Number
476 Evolution in Malaria Parasites. PLOS Genetics. 2008;4(10):e1000243.

477 9. Cheeseman IH, Gomez-Escobar N, Carret CK, Ivens A, Stewart LB, Tetteh KKA, et al.
478 Gene copy number variation throughout the Plasmodium falciparum genome. BMC
479 Genomics. 2009 Aug 4;10:353–353.

480 10. Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, et al. Mitotic
481 Evolution of Plasmodium falciparum Shows a Stable Core Genome but Recombination in
482 Antigen Families. PLOS Genetics. 2013 Feb 7;9(2):e1003293.

483 11. Guler JL, Freeman DL, Ahyong V, Patrapuvich R, White J, Gujjar R, et al. Asexual
484 Populations of the Human Malaria Parasite, Plasmodium falciparum, Use a Two-Step
485 Genomic Strategy to Acquire Accurate, Beneficial DNA Amplifications. PLOS Pathogens.
486 2013 May 23;9(5):e1003375.

487 12. Menard D, Dondorp A. Antimalarial Drug Resistance: A Threat to Malaria Elimination.
488 Cold Spring Harb Perspect Med. 2017;7(7):a025619.

489 13. Foote SJ, Thompson JK, Cowman AF, Kemp DJ. Amplification of the multidrug resistance
490 gene in some chloroquine-resistant isolates of P. falciparum. Cell. 1989 Jun 16;57(6):921–
491 30.

492 14. Wilson CM, Volkman SK, Thaithong S, Martin RK, Kyle DE, Milhous WK, et al.
493 Amplification of pfmdr1 associated with mefloquine and halofantrine resistance in
494 Plasmodium falciparum from Thailand. Molecular and Biochemical Parasitology. 1993 Jan
495 1;57(1):151–60.

496 15. Heinberg A, Siu E, Stern C, Lawrence EA, Ferdig MT, Deitsch KW, et al. Direct evidence
497 for the adaptive role of copy number variation on antifolate susceptibility in Plasmodium
498 falciparum. Molecular microbiology. 2013 May;88(4):702–12.

499 16. Osei M, Ansah F, Matrevi SA, Asante KP, Awandare GA, Quashie NB, et al. Amplification
500 of GTP-cyclohydrolase 1 gene in Plasmodium falciparum isolates with the quadruple mutant
501 of dihydrofolate reductase and dihydropteroate synthase genes in Ghana. PLOS ONE.
502 2018;13(9):e0204871.

503 17. Phillips MA, Rathod PK. Plasmodium dihydroorotate dehydrogenase: a promising target for
504 novel anti-malarial chemotherapy. Infect Disord Drug Targets. 2010;10(3):226–39.

505 18. Mandt REK, Lafuente-Monasterio MJ, Sakata-Kato T, Luth MR, Segura D, Pablos-Tanarro
506 A, et al. In vitro selection predicts malaria parasite resistance to dihydroorotate
507 dehydrogenase inhibitors in a mouse infection model. Science Translational Medicine.
508 2019;11(521):eaav1636.

509 19. Kümpornsin K, Modchang C, Heinberg A, Ekland EH, Jirawatcharadech P, Chobson P, et al.
510 Origin of Robustness in Generating Drug-Resistant Malaria Parasites. Mol Biol Evol. 2014
511 Jul;31(7):1649–60.

512 20. Heinberg A, Kirkman L. The molecular basis of antifolate resistance in Plasmodium
513 falciparum: looking beyond point mutations. Ann N Y Acad Sci. 2015/02/18 ed.
514 2015;1342(1):10–8.

515 21. Herman JD, Rice DP, Ribacke U, Silterra J, Deik AA, Moss EL, et al. A genomic and
516 evolutionary approach reveals non-genetic drug resistance in malaria. Genome Biol.
517 2014;15(11):511–511.

518 22. Manary MJ, Singhakul SS, Flannery EL, Bopp SER, Corey VC, Bright AT, et al.
519 Identification of pathogen genomic variants through an integrated pipeline. BMC
520 Bioinformatics. 2014 Mar 3;15(1):63.

521 23. Cowell AN, Istvan ES, Lukens AK, Gomez-Lorenzo MG, Vanaerschot M, Sakata-Kato T, et
522 al. Mapping the malaria parasite druggable genome by using in vitro evolution and
523 chemogenomics. Science. 2018;359(6372):191–9.

524 24. Huckaby AC, Granum CS, Carey MA, Szlachta K, Al-Barghouthi B, Wang YH, et al.
525 Complex DNA structures trigger copy number variation across the Plasmodium falciparum
526 genome. Nucleic Acids Research. 2018 Dec 21;47(4):1615–27.

527 25. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nature
528 Reviews Genetics. 2011 May 1;12(5):363–76.

529 26. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational
530 challenges and solutions. Nature Reviews Genetics. 2012 Jan 1;13(1):36–46.

531 27. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation
532 of structural variation detection algorithms for whole genome sequencing. Genome Biology.
533 2019 Jun 3;20(1):117.

534 28. Beghain J, Langlois AC, Legrand E, Grange L, Khim N, Witkowski B, et al. Plasmodium
535 copy number variation scan: gene copy numbers evaluation in haploid genomes. Malaria
536 Journal. 2016 Apr 12;15(1):206.

537 29. Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural
538 variation, and recombination drive genomic diversity in Plasmodium falciparum. Genome
539 Res. 2016/08/16 ed. 2016;26(9):1288–99.

540 30. Cretu Stancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middelkamp S, de Ligt J, et
541 al. Mapping and phasing of structural variation in patient genomes using nanopore
542 sequencing. Nature Communications. 2017 Nov 6;8(1):1326.

543 31. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al.
544 Accurate detection of complex structural variations using single-molecule sequencing. Nat
545 Methods. 2018/04/30 ed. 2018;15(6):461–8.

546 32. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of
547 long-range sequencing and mapping. Nature Reviews Genetics. 2018 Jun 1;19(6):329–46.

548 33. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nature Reviews
549 Genetics. 2020 Mar 1;21(3):171–89.

550 34. Belikova D, Jochim A, Power J, Holden MTG, Heilbronner S. “Gene accordions” cause
551 genotypic and phenotypic heterogeneity in clonal populations of Staphylococcus aureus.
552 Nature Communications. 2020 Jul 14;11(1):3526.

553 35. Rugbjerg P, Dyerberg ASB, Quainoo S, Munck C, Sommer MOA. Short and long-read ultra-
554 deep sequencing profiles emerging heterogeneity across five platform Escherichia coli
555 strains. Metabolic Engineering. 2021 May 1;65:197–206.

556 36. McDaniels JM, Huckaby AC, Carter SA, Lingeman S, Francis A, Congdon M, et al.
557 Extrachromosomal DNA amplicons in antimalarial-resistant Plasmodium falciparum. Mol
558 Microbiol. 2020/11/19 ed. 2021;115(4):574–90.

559 37. Palmer MJ, Deng X, Watts S, Krilov G, Gerasyuto A, Kokkonda S, et al. Potent
560 Antimalarials with Development Potential Identified by Structure-Guided Computational
561 Optimization of a Pyrrole-Based Dihydroorotate Dehydrogenase Inhibitor Series. Journal of
562 Medicinal Chemistry. 2021 May 13;64(9):6085–136.

563 38. Phillips MA, Lotharius J, Marsh K, White J, Dayan A, White KL, et al. A long-duration
564 dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria.
565 Sci Transl Med. 2015 Jul 15;7(296):296ra111.

566 39. Turkiewicz A, Manko E, Sutherland CJ, Benavente ED, Campino S, Clark TG. Genetic
567 diversity of the Plasmodium falciparum GTP-cyclohydrolase 1, dihydrofolate reductase and
568 dihydropteroate synthetase genes reveals new insights into sulfadoxine-pyrimethamine
569 antimalarial drug resistance. PLOS Genetics. 2020 Dec 31;16(12):e1009268.

570 40. Sepúlveda N, Campino SG, Assefa SA, Sutherland CJ, Pain5 A, Clark TG. A Poisson
571 hierarchical modelling approach to detecting copy number variation in sequence coverage
572 data. BMC Genomics. 2013;14:128.

573 41. White J, Dhingra SK, Deng X, El Mazouni F, Lee MCS, Afanador GA, et al. Identification
574 and Mechanistic Understanding of Dihydroorotate Dehydrogenase Point Mutations in
575 Plasmodium falciparum that Confer in Vitro Resistance to the Clinical Candidate DSM265.
576 ACS Infect Dis. 2019 Jan 11;5(1):90–101.

577 42. Ross LS, Gamo FJ, Lafuente-Monasterio MJ, Singh OMP, Rowland P, Wiegand RC, et al. In
578 vitro resistance selections for Plasmodium falciparum dihydroorotate dehydrogenase
579 inhibitors give mutants with multiple point mutations in the drug-binding site and altered
580 growth. J Biol Chem. 2014 Jun 27;289(26):17980–95.

581 43. Lukens AK, Ross LS, Heidebrecht R, Javier Gamo F, Lafuente-Monasterio MJ, Booker ML,
582 et al. Harnessing evolutionary fitness in Plasmodium falciparum for drug discovery and
583 suppressing resistance. Proc Natl Acad Sci U S A. 2014 Jan 14;111(2):799–804.

584 44. Ebel ER, Kim BY, McDew-White M, Egan ES, Anderson TJC, Petrov DA. Antigenic
585 diversity in malaria parasites is maintained on extrachromosomal DNA. bioRxiv. 2023 Jan
586 1;2023.02.02.526885.

587 45. Girgis HS, DuPai CD, Lund J, Reeder J, Guillory J, Durinck S, et al. Single-molecule
588 nanopore sequencing reveals extreme target copy number heterogeneity in arylomycin-
589 resistant mutants. Proc Natl Acad Sci U S A. 2021 Jan 5;118(1):e2021958118.

590 46. Grigorev K, Foox J, Bezdan D, Butler D, Luxton JJ, Reed J, et al. Haplotype diversity and
591 sequence heterogeneity of human telomeres. Genome Res. 2021 Jul;31(7):1269–79.

592 47. Dixon K, Shen Y, O'Neill K, Mungall KL, Chan S, Bilobram S, et al. Defining the
593 heterogeneity of unbalanced structural variation underlying breast cancer susceptibility by
594 nanopore genome sequencing. Eur J Hum Genet. 2023 May;31(5):602–6.

595 48. Pholwat S, Liu J, Stroup S, Jacob ST, Banura P, Moore CC, et al. The Malaria TaqMan
596 Array Card Includes 87 Assays for Plasmodium falciparum Drug Resistance, Identification
597 of Species, and Genotyping in a Single Reaction. Antimicrobial Agents and Chemotherapy
598 [Internet]. 2017 May 1 [cited 2020 Jul 29];61(5). Available from:
599 https://aac.asm.org/content/61/5/e00110-17

600 49. Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, et al.
601 Genetic analysis of the human malaria parasite Plasmodium falciparum. Science. 1987 Jun
602 26;236(4809):1661–6.

603 50. Wellems TE, Panton LJ, Gluzman IY, do Rosario VE, Gwadz RW, Walker-Jonah A, et al.
604 Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross.
605 Nature. 1990 May 17;345(6272):253–5.

606 51. Chulay JD, Watkins WM, Sixsmith DG. Synergistic Antimalarial Activity of Pyrimethamine
607 and Sulfadoxine against Plasmodium falciparum In Vitro. The American Journal of Tropical
608 Medicine and Hygiene. 1984 May 1;33(3):325–30.

609 52. Watkins WM, Sixsmith DG, Chulay JD, Spencer HC. Antagonism of sulfadoxine and
610 pyrimethamine antimalarial activity in vitro by p-aminobenzoic acid, p-
611 aminobenzoylglutamic acid and folic acid. Molecular and Biochemical Parasitology. 1985
612 Jan 1;14(1):55–61.

613 53. Wang P, Brobey RKB, Horii T, Sims P, Hyde JE, Schapira A, et al. Utilization of exogenous
614 folate in the human malaria parasite Plasmodium falciparum and its critical role in antifolate
615 drug synergy The Susceptibility of Plasmodium Falciparum to Sulfadoxine and
616 Pyrimethamine: Correlation of in Vivo and in Vitro Results Synergistic Antimalarial
617 Activity of Pyrimethamine and Sulfadoxine against Plasmodium falciparum In Vitro.
618 Molecular Microbiology. 1986 Mar 1;32(2):239–45.

619 54. Salcedo-Sora JE, Ochong E, Beveridge S, Johnson D, Nzila A, Biagini GA, et al. The
620 Molecular Basis of Folate Salvage in Plasmodium falciparum: CHARACTERIZATION OF
621 TWO FOLATE TRANSPORTERS*. Journal of Biological Chemistry. 2011 Dec
622 30;286(52):44659–68.

623 55. Valenciano AL, Fernández-Murga ML, Merino EF, Holderman NR, Butschek GJ, Shaffer
624 KJ, et al. Metabolic dependency of chorismate in Plasmodium falciparum suggests an
625 alternative source for the ubiquinone biosynthesis precursor. Scientific Reports. 2019 Sep
626 26;9(1):13936.

627 56. Kumar S, Li X, McDew-White M, Reyes A, Delgado E, Sayeed A, et al. A Malaria Parasite
628 Cross Reveals Genetic Determinants of Plasmodium falciparum Growth in Different Culture
629 Media. Front Cell Infect Microbiol. 2022 May 30;12:878496.

630 57. Llanos-Cuentas A, Casapia M, Chuquiyauri R, Hinojosa JC, Kerr N, Rosario M, et al.
631 Antimalarial activity of single-dose DSM265, a novel plasmodium dihydroorotate
632 dehydrogenase inhibitor, in patients with uncomplicated Plasmodium falciparum or
633 Plasmodium vivax malaria infection: a proof-of-concept, open-label, phase 2a study. Lancet
634 Infect Dis. 2018 Aug;18(8):874–83.

635 58. Ravenhall M, Benavente ED, Mipando M, Jensen ATR, Sutherland CJ, Roper C, et al.
636 Characterizing the impact of sustained sulfadoxine/pyrimethamine use upon the Plasmodium
637 falciparum population in Malawi. Malar J. 2016 Nov 29;15:575.

638 59. Moon RW, Sharaf H, Hastings CH, Ho YS, Nair MB, Rchiad Z, et al. Normocyte-binding
639 protein required for human erythrocyte invasion by the zoonotic malaria parasite
640 Plasmodium knowlesi. Proc Natl Acad Sci U S A. 2016 Jun 28;113(26):7231–6.

641 60. Phillips MA, Gujjar R, Malmquist NA, White J, El Mazouni F, Baldwin J, et al.
642 Triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors with potent and selective
643 activity against the malaria parasite Plasmodium falciparum. J Med Chem. 2008/06/04 ed.
644 2008;51(12):3649–53.

645 61. Haynes JD, Diggs CL, Hines FA, Desjardins RE. Culture of human malaria parasites
646 Plasmodium falciparum. Nature. 1976 Oct 1;263(5580):767–9.

647 62. Bei AK, Desimone TM, Badiane AS, Ahouidi AD, Dieye T, Ndiaye D, et al. A flow
648 cytometry-based assay for measuring invasion of red blood cells by Plasmodium falciparum.
649 Am J Hematol. 2010 Apr;85(4):234–7.

650 63. Liu S, Huckaby AC, Brown AC, Moore CC, Burbulis I, McConnell MJ, et al. Single-cell
651 sequencing of the small and AT-skewed genome of malaria parasites. Genome Medicine.
652 2021 May 4;13(1):75.

653 64. Bushnell B. BBMap [Internet]. [cited 2021 Dec 1]. Available from:
654 http://sourceforge.net/projects/bbmap/ (2019)

655 65. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq:
656 ultra-fast personal genome analysis and interpretation. Nat Methods. 2015 Oct;12(10):966–8.

657 66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
658 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

659 67. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for
660 structural variant discovery. Genome Biology. 2014 Jun 26;15(6):R84.

661 68. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype,
662 and characterize typical and atypical CNVs from family and population genome sequencing.
663 Genome Res. 2011/02/07 ed. 2011;21(6):974–84.

664 69. Maniatis T, Sambrook J, Fritsch EF. Molecular cloning: a laboratory manual. Cold Spring
665 Harbor, NY: Cold Spring Harbor Laboratory Press; 1989. xxxviii + 1546 pp.

666 70. Oxford Nanopore Technologies. qcat (version 1.1.0) [Internet]. [cited 2021 Jun 1]. Available
667 from: https://github.com/nanoporetech/qcat

668 71. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
669 2018;34(18):3094–100.

670 72. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al.
671 Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012 Aug
672 22;28(20):2678–9.

673 73. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
674 Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

675 74. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
676 performance genomics data visualization and exploration. Briefings in Bioinformatics. 2013
677 Mar 1;14(2):178–92.

678 75. Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. Nucleic
679 Acids Research. 2006;34(suppl_2):W6–9.

680

681

682 **Supporting Information Captions**

683 Figure S1: Spanning reads carry odd copy numbers of *GCH1* amplicons

684 Figure S2: Long-read visualization of *GCH1* amplicon from *3D7* parasite line

685 Figure S3: Parasites lines with *DHODH* amplicons are selected from parent lines with *GCH1*
686 amplicons

687 Figure S4: *GCH1* amplicon junctions show signs of primary and secondary expansion events

688 Supplemental Methods: Procedures used for the preparation of long read data by Ellen Yeh's
689 group, Stanford University, as presented in Table S2, including in vitro cultures, mutagenesis,
690 DNA extraction, and sequencing.

691 Table S1: Oxford Nanopore sequencing datasets new to this manuscript

692 Table S2: Quantification of long reads covering the *GCH1* amplicon displays greater variability in
693 amplicon number from DSM1 resistant parasites

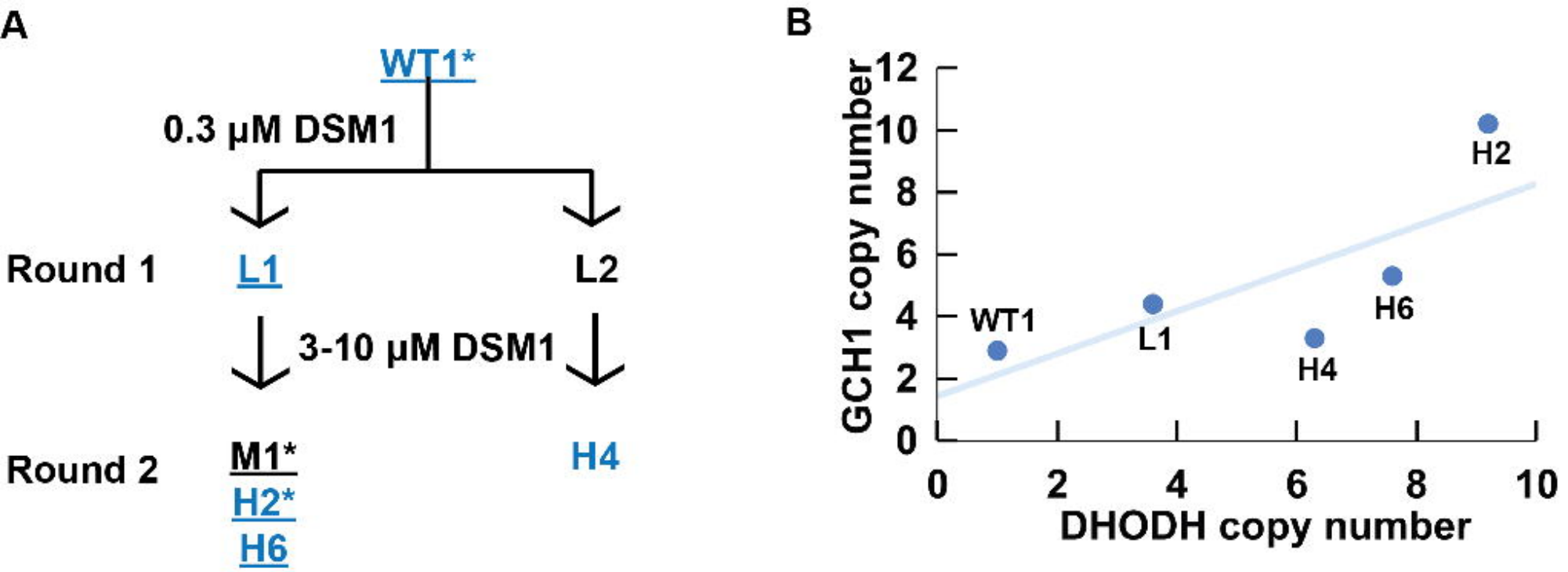
694 Table S3: *DHODH* and *GCH1* copy number in parasites resistant to DHODH inhibitors

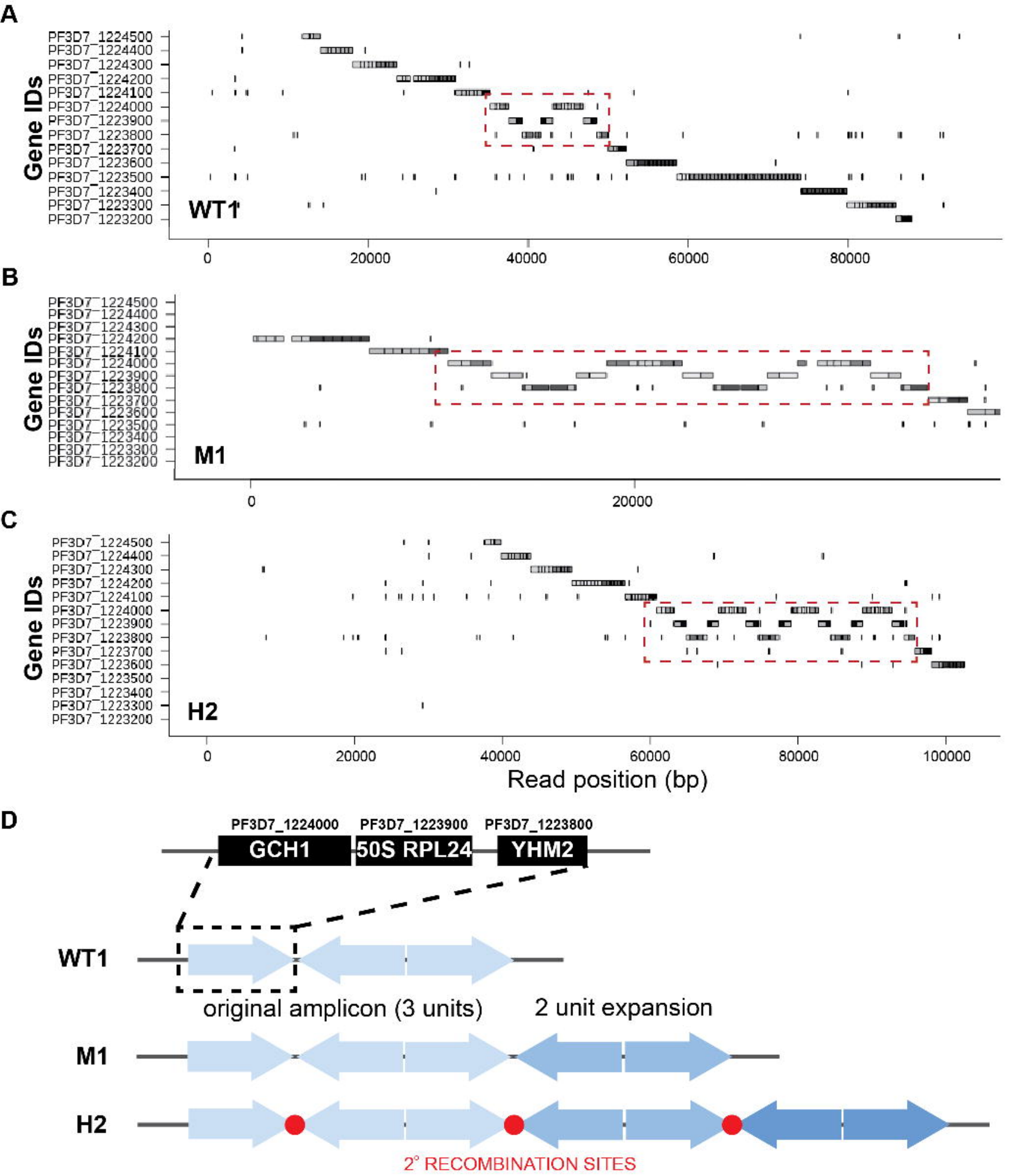
695 Table S4: Detection of known amplifications by Illumina short reads from parasites resistant to
696 DHODH inhibitors

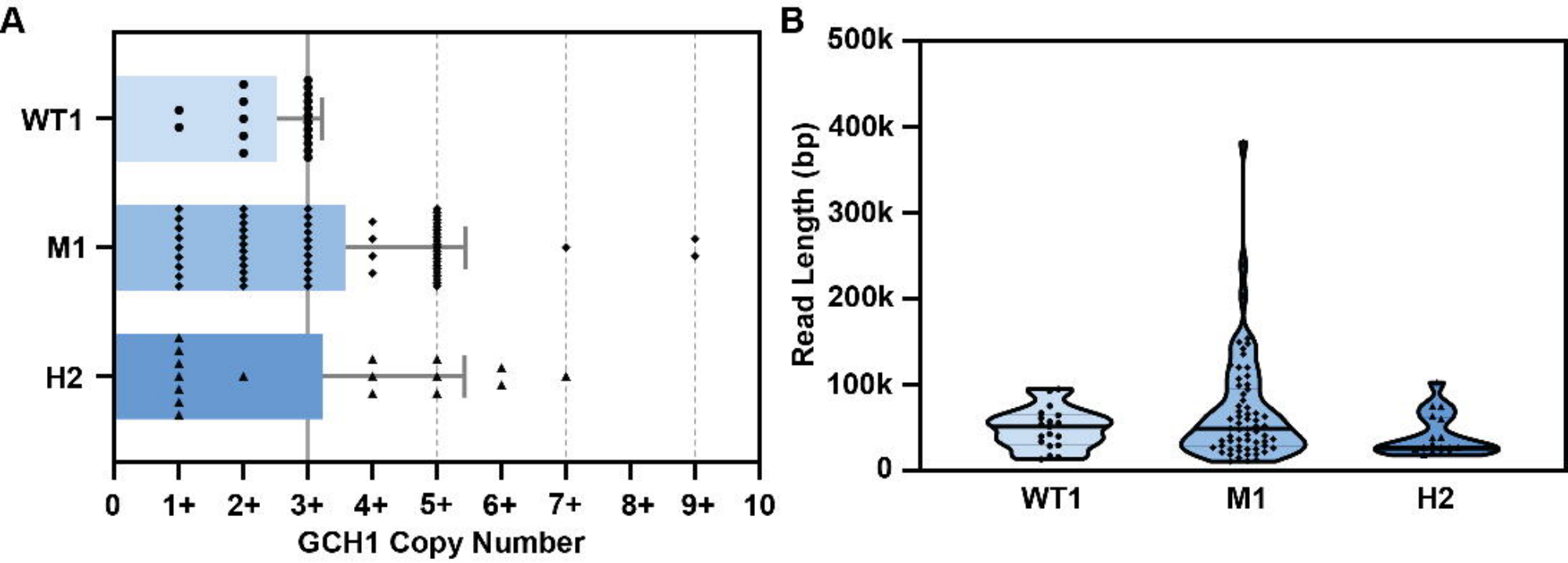
697 Table S5: ddPCR primer and probe sequences

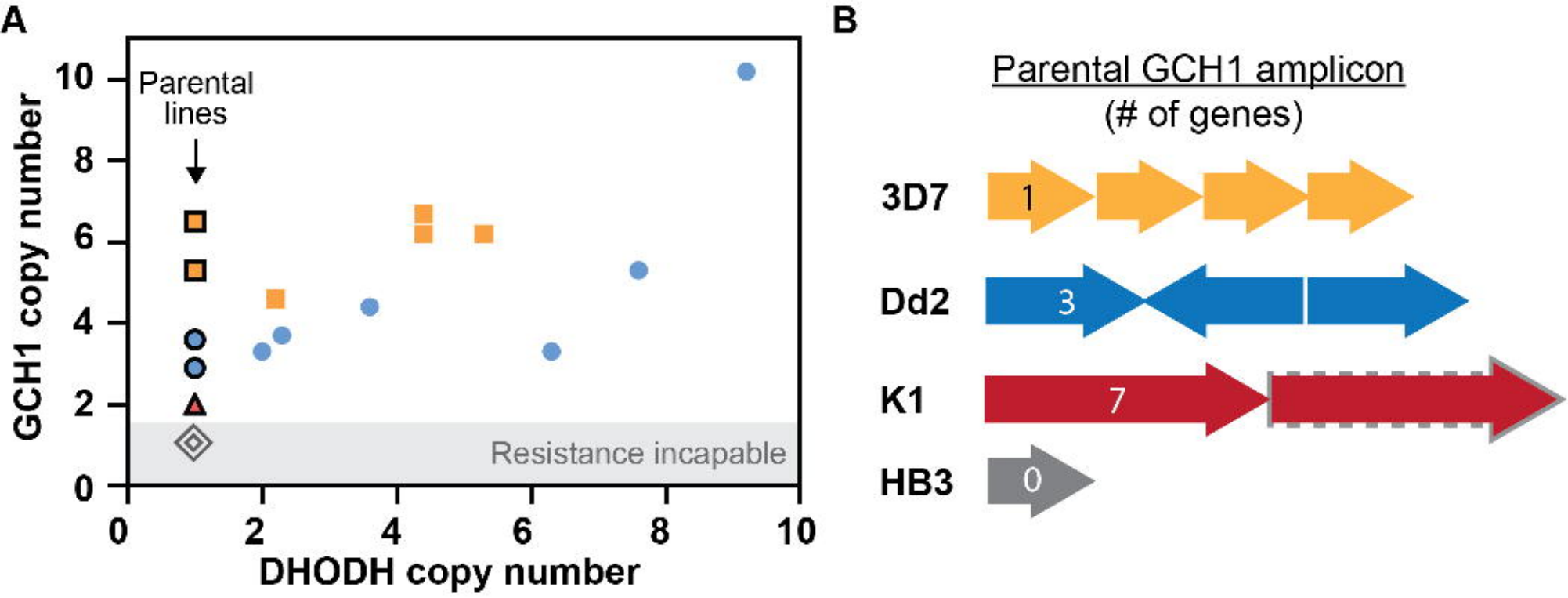
698

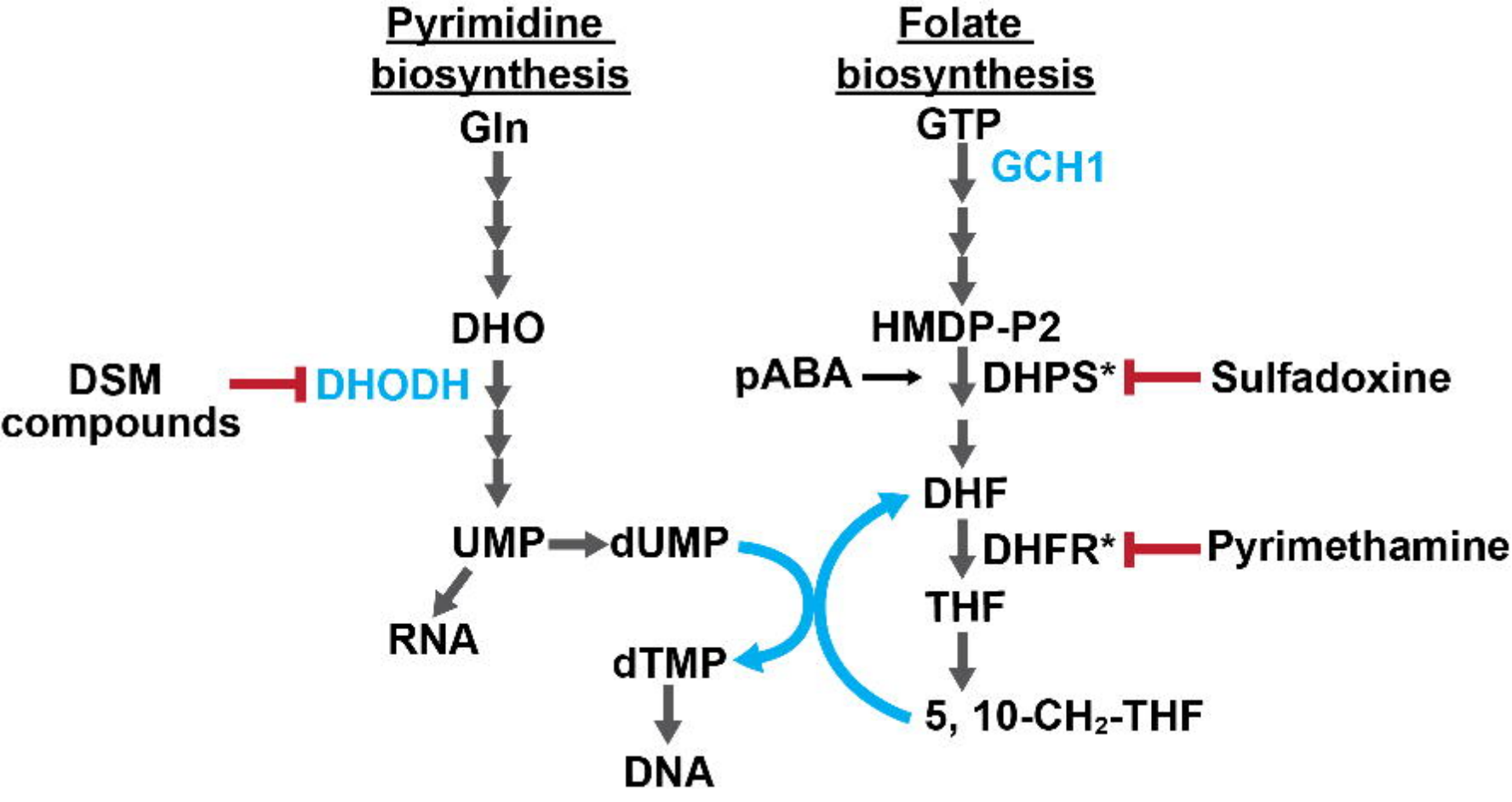
699


A
WT1*
0.3 μM DSM1
Round 1
L1
L2
3-10 μM DSM1
Round 2
M1*
H2*
H6
H4
B
GCH1 copy number
DHODH copy number
0
2
4
6
8
10
12
WT1
L1
H4
H6
H2


A
Gene IDs
PF3D7_1224500
PF3D7_1224400
PF3D7_1224300
PF3D7_1224200
PF3D7_1224100
PF3D7_1224000
PF3D7_1223900
PF3D7_1223800
PF3D7_1223700
PF3D7_1223600
PF3D7_1223500
PF3D7_1223400
PF3D7_1223300
PF3D7_1223200
WT1
0
20000
40000
60000
80000
B
Gene IDs
M1
0
20000
C
Gene IDs
H2
0
20000
40000
60000
80000
100000
Read position (bp)
D
PF3D7_1224000
PF3D7_1223900
PF3D7_1223800
GCH1
50S RPL24
YHM2
WT1
original amplicon (3 units)
2 unit expansion
M1
H2
2° RECOMBINATION SITES


A
WT1
M1
H2
0
1+
2+
3+
4+
5+
6+
7+
8+
9+
10
GCH1 Copy Number
B
Read Length (bp)
500k
400k
300k
200k
100k
0
WT1
M1
H2


A
Parental lines
Resistance incapable
GCH1 copy number
DHODH copy number
0
2
4
6
8
10
B
Parental GCH1 amplicon
(# of genes)
3D7
1
Dd2
3
K1
7
HB3
0


Pyrimidine biosynthesis
Gln
DHO
DSM compounds
DHODH
UMP
dUMP
RNA
dTMP
DNA
Folate biosynthesis
GTP
GCH1
HMDP-P2
pABA
DHPS*
Sulfadoxine
DHF
DHFR*
Pyrimethamine
THF
5, 10-CH2-THF