

## **Chapter 3: Detection, surveillance, and evaluation of molecular markers of piperazine resistance: *plasmepsin 2-3* copy number assays and functional studies**

---

### **3.1 Declaration of work**

The following chapter includes excerpts of data from a manuscript in which I am co-first author submitted for pre-print publication on bioRxiv on 4<sup>th</sup> July 2019 (Jacob et al., 2019). This chapter also contains copy number data that was performed for a manuscript in preparation in which I am a contributing author (Amaratunga et al., 2019). The content included in this chapter is my own work and does not include work done in collaboration with others unless directly stated.

#### **3.1.1 Significance and purpose of study**

Antimalarial drug resistance is an unrelenting obstacle to malaria control programs. As countries push toward malaria elimination, drug resistance and the potential for resistance to develop pose constant threats to treatment efforts which rely solely on chemotherapy. In Southeast Asia (SEA), parasites have developed some degree of resistance to nearly every malaria drug currently available, with the most recent emergence to artemisinin combination therapies (ACTs) (World Health Organization, 2018). ACTs are the recommended front-line treatments for *Plasmodium falciparum* malaria worldwide (World Health Organization, 2018) and decreased susceptibility of parasites to both artemisinin and one of the widely used partner drugs, piperazine (PPQ), have been reported in multiple locations throughout SEA (Duru et al., 2015, Leang et al., 2015, Spring et al., 2015, Saunders et al., 2014, Amaratunga et al., 2016, Chaorattanakawee et al., 2015, Imwong et al., 2017). It is therefore necessary to have reliable methods for detecting and evaluating resistant phenotypes. The purpose of this study was to combine clinical data from Cambodia with genomic studies to develop assays for detecting one of the molecular markers of piperazine (PPQ) resistance, a copy number variation (CNV) in the *plasmepsin 2* and *plasmepsin 3* (*PM2-PM3*) genes. In addition to assay development, this study designed and performed *in vitro* functional studies to examine the role of the *PM2-3* CNV in PPQ resistance.

### 3.1.2 Introduction

In 2008, dihydroartemisinin-piperaquine (DHA-PPQ) was deployed as frontline treatment for uncomplicated *P. falciparum* malaria in Cambodia and implemented nationwide in 2010 (Amaratunga et al., 2016). Since that time, drug efficacy studies have monitored for treatment failures, which are defined as greater than 10% recrudescence (persistence of the infection). As soon as 2010 in Western Cambodia, approximately 25% and 11% failure rates to DHA-PPQ treatment were observed in Pailin and Pursat provinces, respectively (Leang et al., 2013). By 2013—less than 3 years after nationwide drug implementation—widespread failure rates were observed throughout Cambodia, reaching as high as 50% (Duru et al., 2015, Leang et al., 2015, Spring et al., 2015, Saunders et al., 2014, Amaratunga et al., 2016, Chaorattanakawee et al., 2015). Genomic studies (Amato et al., 2017, Witkowski et al., 2017) on parasite samples collected from patients in three Cambodian provinces from 2012-2013 (Amaratunga et al., 2016) found several regions of the parasite genome that associated with treatment failures in patients and reduced PPQ susceptibility *in vitro*. Specifically, genome-wide association studies (GWAS) revealed copy number variations (CNVs) in the *plasmepsin 2* (*PM2*) and *plasmepsin 3* (*PM3*) genes on chromosome 14 as well as nonsynonymous single-nucleotide polymorphisms (SNPs) on chromosome 13 in genes encoding a putative exonuclease, PF3D7\_1362500 (*exo-E415G*), and a putative mitochondrial carrier protein, PF3D7\_1368700 (*mcp-N252D*) (Amato et al., 2017, Witkowski et al., 2017). Further work also reported mutations in the *chloroquine resistance transporter* gene (*pfcr1*) that confer varying levels of resistance to PPQ in both field isolates and the lab-adapted strain, Dd2 (Ross et al., 2018a, Agrawal et al., 2017, Duru et al., 2015). These molecular markers are on a mutant kelch13 (K13) propeller domain background, the genetic marker for artemisinin resistance, so it is hypothesized that the rapid decline of piperaquine efficacy was due to widespread artemisinin resistance in Western Cambodia, which put increased pressure on the partner drug, allowing parasites to quickly develop resistance to piperaquine.

Copy number variations or polymorphisms describe an amplification (increase) or decrease in the number or copies of a gene in the genome. CNVs are common in many eukaryotes, including humans, mice, and malaria parasites and are often associated with drug resistant phenotypes (Freeman et al., 2006, Anderson et al., 2009). In *P. falciparum*, the significant role CNVs could play in drug resistance was seen in the *multidrug resistant-1* gene (*pfmdr1*) (Anderson et al., 2009). An increase in the copy number of the *pfmdr1* gene was found to confer resistance to the ACT partner drug, mefloquine (Wilson et al., 1989, Wilson et al.,

1993, Price et al., 2004). Mefloquine was used in Cambodia from 2001 as the ACT, artesunate-mefloquine (AS-MQ) until 2008, when the partner drug failed with widespread amplification of *pfmdr1*. Since the introduction of DHA-PPQ in 2008, the *pfmdr1* amplification has declined with parasite populations predominantly reverting back to single copies of *pfmdr1* and increased copies of *PM2-3* (Amaratunga et al., 2016, Amato et al., 2017, Witkowski et al., 2017).

The *PM2-3* CNV observed in Cambodian parasite isolates exposed to DHA-PPQ treatment consists of an amplification of both the *PM2* and *PM3* genes. Whole genome sequencing (WGS) data shows that the breakpoint of parasites with at least two copies of *PM2-3* lies in the ends of the *plasmepsin 1* and *plasmepsin 3* genes, creating a duplication of the full *plasmepsin 2* gene and a chimeric *plasmepsin 3* in which the 3' end of *PM3* is replaced with the 3' end of *plasmepsin 1* (depicted in **Figure 3.5**) (Amato et al., 2017). Plasmepsins 1-4 (PM1-PM4) are enzymes located in the digestive vacuole of the parasite that function as proteases during the intraerythrocytic stages of the parasite lifecycle by degrading hemoglobin (Banerjee et al., 2002). Hemoglobin degradation also generates heme, which is a toxic by-product that causes oxidative stress to the parasite, so it is converted to inert hemozoin crystals (Wunderlich et al., 2012). Though the mechanism of piperazine action is unknown, one hypothesis is that it interferes with the hemoglobin degradation pathway, similar to chloroquine, and prevents hemoglobin degradation or heme detoxification. Thus a potential role for the *PM2-3* gene amplification could be to overcome any inhibitory effects that piperazine may have on the process. It should also be noted that the plasmepsins are redundant enzymes, and additional proteases such as falcilysins and falcipains also facilitate hemoglobin degradation (Liu et al., 2006), so it is possible that piperazine is affecting earlier stages of the degradation pathway in which the plasmepsins are involved. However, these are speculations that have yet to be tested in molecular studies.

The emergence of DHA-PPQ resistance greatly threatens the efficacy of the remaining ACTs worldwide. With the availability of molecular markers of piperazine resistance, it is imperative to have robust assays that can be used to monitor the presence and frequency of these markers in contemporary parasite isolates across endemic regions. Functional studies are also necessary for providing insight into the biological mechanisms of resistance. This study has developed quantitative PCR (qPCR) assays to measure the copy number of *PM2* and *PM3* using genomic DNA. I have also developed a PCR-based breakpoint assay to detect the presence of the chimeric *PM3-PM1* hybrid created by the duplication. The two copy

number assays are sensitive, reliable, and cost-efficient methods for detecting the breakpoint quickly. In addition to assays for monitoring the *PM2-3* CNV, I performed functional studies to test for the effects of increased copies of *PM2-3*.

### 3.1.3 Objectives

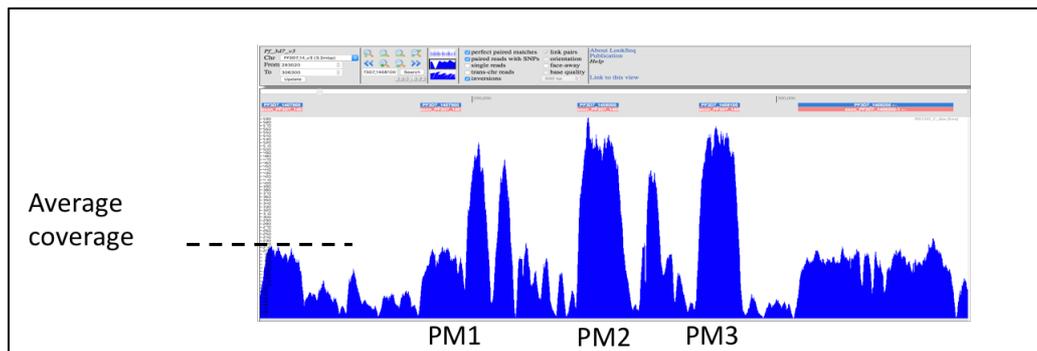
Examine the role of the amplification of hemoglobin degrading proteases, *PM2* and *PM3* (*PM2-PM3*), in piperazine resistance

- (1) Development of qPCR and breakpoint PCR-based assays to detect *PM2-3* resistance-associated copy number variation
  - a. Breakpoint PCR validation and comparison with SYBR-based qPCR assay using samples from DHA-PPQ study (Amaratunga et al., 2016)
  - b. Analysis of *PM2-3* copy numbers in samples from AS-MQ study (Amaratunga et al., 2019)
- (2) Study the functional role of the *PM2-3* amplification by genetic modulation of *PM2-3* copy numbers *in vitro*
  - a. Overexpression of *PM2* and both *PM2-3* in a piperazine-sensitive Cambodian field isolate with a single copy of the *PM2-3* genes

## 3.2 Materials and methods

### 3.2.1 Parasite samples

Field isolates used in breakpoint PCR and qPCR assays were obtained from clinical trials conducted in 3 sites (**Ch. 2, Figure 2.2**) in Cambodia: Pursat, Preah Vihear, and Ratanakiri between 2012-2015 (clinicaltrials.gov ID: NCT01736319) (Amaratunga et al., 2016). In the clinical trial from 2012-2013, patients presenting with uncomplicated *P. falciparum* malaria were treated with DHA-PPQ at all sites (Amaratunga et al., 2016). In the clinical trial from 2014-2015, patients were treated with artesunate-mefloquine (AS-MQ) at all sites (Amaratunga et al., 2019). Genomic DNA (gDNA) was obtained from 200  $\mu$ L of whole-blood taken at the time of diagnosis. Control samples used for validation and assay optimization were obtained from gDNA extracted from the lab strains, 3D7 and Dd2, and from field isolates (**Ch. 2, Table 2.1**) growing in continuous culture for which WGS data was available (**Figure 3.1**).



**Figure 3. 1. Whole genome sequencing data displaying average coverage on chromosome 14 for isolate 163-KH1-004 (PH1345\_C) as viewed using the LookSeq browser.** (Manske and Kwiatkowski, 2009). The dashed line represents the average coverage and the blue represents the number of reads at the approximate positions on chromosome 14: 283100-300550, encompassing the genes of *plasmepsins 1, 2, and 3*, PF3D7\_1407900-PF3D7\_1408100. All field isolates used for copy number controls were validated for increased *PM2-3* copy number using the LookSeq browser.

### 3.2.2 *Plasmepsin 2-3* duplication breakpoint PCR assay

Whole genome sequencing data from piperazine-resistant *P. falciparum* isolates obtained from clinical trials in Cambodia were used to identify the breakpoint of the *PM2-3* amplification. PCR primers were manually designed to amplify the region surrounding the breakpoint (Amaratunga et al., 2016, Amato et al., 2017, Witkowski et al., 2017). Primers AF (forward 5'-CCACGATTTATATTGGCAAGTTGATTTAG-3') and AR (reverse 5'-CATTCTACTAAAATAGCTTTAGCATCATTACAG-3') amplify a 623 base pair (bp) product surrounding the breakpoint located at the 3' end of *plasmepsin 1* (*PM1*). Primers BF (forward - 5'-CGTAGAATCTGCAAGTGT TTTCAAAG-3') and BR (reverse 5'-AATGTTATAAATGCAATATAATCAAACGACATCAC-3') amplify a 484 bp product surrounding the breakpoint located at the 3' end of *PM3*. BF + AR amplify the junction between the breakpoint and produce a 497 bp product in isolates with *PM2-3* amplifications. These primers face opposite directions in samples without duplications and are not expected to amplify a product in single copy isolates. Both control (AF + AR; BF+ BR) and duplication (BF + AR) primer sets were used with all samples and single copy isolates were only noted if the control primer sets amplified a product and duplication PCR was negative. Two or more copies were annotated as >1 copy of *PM2-3* only if both the control and duplication primer sets produced a product. PCR reactions contained 10 µl SapphireAmp Fast PCR Master Mix (Takara Bio Inc), 0.3 µl of each primer (Integrated DNA Technologies), 1 µl of genomic DNA up to 20 µl final volume with water. PCR conditions were: 92°C for 2 minutes, followed by 30 cycles of 92°C for 30 seconds, 59°C for 30 seconds, 66°C for 1.5 minutes, followed by a 1 minute extension at 66°C.

### 3.2.3 SYBR green validation of *plasmepsin 2-3* copy number using quantitative PCR

PCR primers for estimating *PM2* copy number amplification were designed manually inside the *PM2* gene (forward - 5'-CTTATACTGCTTCAACATTTGATGGTATCCTTGG-3'; reverse - 5'-GGTTAAGAATCCTGTATGTTTATCATGTACAGGTAAG-3'). Previously described primers for *P. falciparum* lactate dehydrogenase (*ldh*) (Lim et al., 2013), were used as the control for a single copy gene (forward - 5'-AGGACAATATGGACTCCGAT-3'; reverse - 5'-TTTCAGCTATGGCTTCATCAA-3'). Similar to the assays performed by Jacob *et al.* (Jacob et al., 2019), assays for amplifying *PM2* and *PM3* individually were compared to determine if each gene could be used as a marker of duplication and no major differences were observed (data not shown). *PM2* primers only were used for higher throughput in SYBR green assays, since each gene, *PM2* and *ldh* must be analyzed in separate reactions. Real-time PCR reactions were carried out in 20  $\mu$ l volumes in a 96-well plate (Bio-Rad) containing 10  $\mu$ l SensiFAST SYBR No-ROX mix (2x) (Bioline Inc.), 300 nM of each primer, and 2  $\mu$ l genomic DNA. Reactions were performed using a CFX Connect Real-Time PCR Detection System (Bio-Rad) using the following conditions: 5 minutes at 95°C, followed by 40 cycles of 10 seconds at 95°C, 20 seconds at 58°C, and 20 seconds at 60°C. Relative copy number was calculated on the basis of the  $2^{-\Delta\Delta C_t}$  method for relative quantification.  $\Delta\Delta C_t$  was calculated as  $(C_{t_{ldh}} - C_{t_{pfplasmepsin2}}) - (C_{t_{ldh\ cal}} - C_{t_{pfplasmepsin2\ cal}})$ , where cal is the calibration control of genomic 3D7 DNA with one copy of both *ldh* and *PM2*. DNA from an isolate with two copies of *PM2-3* (PH1387-C) (Amaratunga et al., 2016) was used as an internal plate control. All samples were analyzed in triplicate and each plate was replicated in triplicate. A relative copy number of 1.5 or greater was recorded as 2 copies.

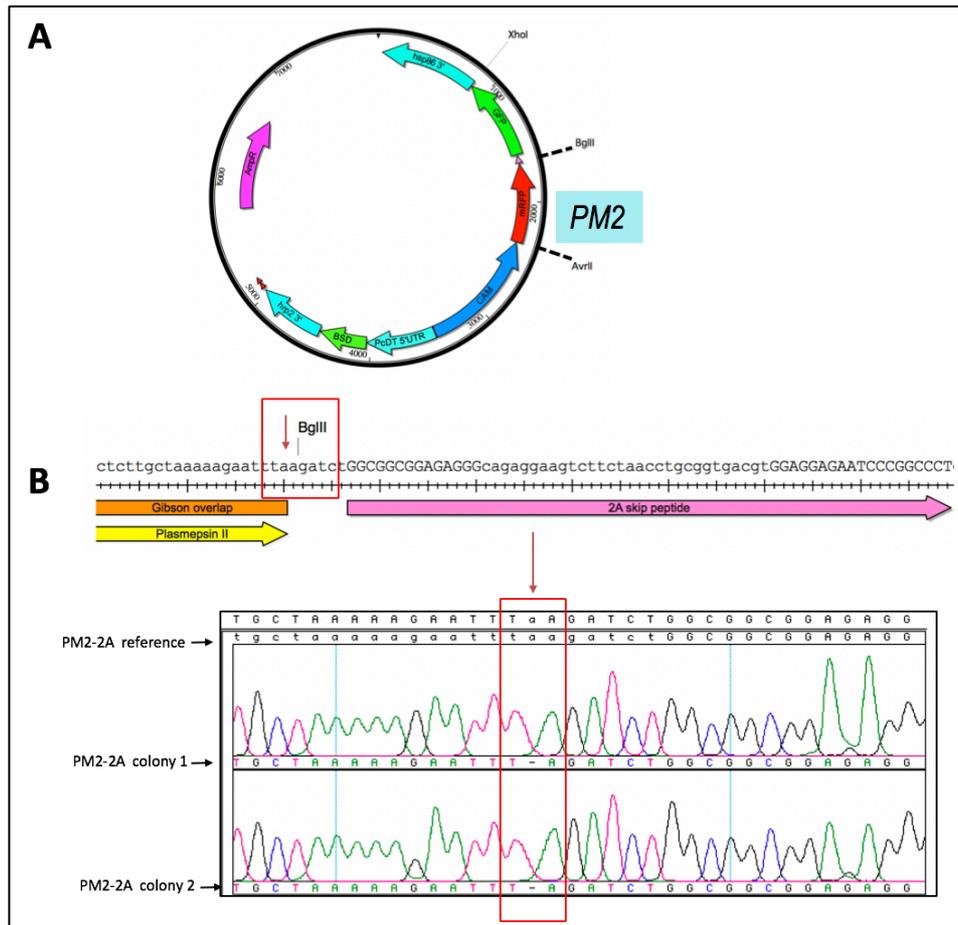
### 3.2.4 Overexpression of *PM2* and *PM3*

For overexpression of both the *PM2* and *PM3* genes, I used the pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid (Straimer et al., 2012). This plasmid was selected to enable stoichiometric amounts of both *PM2* and *PM3* to be expressed by utilizing the same promoter and a 2A-“skip” peptide to separate the two genes. Gibson assembly for all plasmids was performed using the HiFi DNA Assembly reagent and accompanying protocol from New England Biolabs (NEB).

### 3.2.4.1 Gibson assembly of a *PM2*-only overexpression plasmid: pDC2-*cam-PM2-2A-gfp-bsd-attP*

Before assembling a *PM2* and *PM3* overexpression plasmid, a single *PM2*-only expression plasmid was made by inserting the *PM2* gene into the pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid. The *mrfp* segment was excised from the pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid using the enzymes, *AvrII* and *BglIII* (NEB) (**Figure 3.2A**). The *PM2* gene was PCR-amplified from genomic 3D7 DNA using the primers, *PM2\_OE\_F* and *PM2\_OE\_R* (**Table 3.1**), which included overlap (lowercase letters) with the plasmid for Gibson assembly. No stop codon was included after the *PM2* gene to ensure translation would carry on through the 2A-skip peptide to the 2A-*gfp* segments (**Figure 3.2A**).

The reverse primer (*PM2\_OE\_R*) used to amplify the *PM2* gene contains a *BglIII*-2A overlap region (**Figure 3.2A**) by including the *BglIII* restriction site (AGATCT), but uses the “T” from the end of *PM2* and this T is highlighted in red in primer, *PM2\_OE\_R* (**Table 3.1**). Because of this, the plasmid generated by Gibson assembly was missing this nucleotide (read as an “A” in the forward sequence) (**Figure 3.2B**). This does not affect *PM2* expression, but changes the reading frame downstream of *PM2*, so *gfp* is not translated. The shift in the reading frame causes the nearest stop codon to be located in the 2A-skip peptide and no longer at the end of *gfp*, as originally designed. Since the reading frame was shifted after the full length *PM2* gene, this study was able to use this plasmid to examine the role of *PM2*-only gene overexpression (since it would be translated) but needed to use a different approach to generate a *PM2-2A-PM3* overexpression plasmid, described below.



**Figure 3.2. Plasmid design and assembly for overexpression of *plasmepsin 2* (*PM2*).** A.) The pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid was digested with AvrII and BglII to obtain the plasmid backbone and the *PM2* gene (light blue) was inserted via Gibson assembly. B.) Sequence chromatograms of pDC2-*cam-PM2-2A-gfp* colonies (colonies 1 and 2) aligned with the reference sequence of the inserted *PM2* through the 2A peptide sequence. A red box highlights the region missing the “A” nucleotide (indicated with a red arrow) in the forward sequence. This missing nucleotide does not affect *PM2* expression, but shifts the reading frame downstream of *PM2* so that the nearest stop codon is located in the 2A-skip peptide and *gfp* is not translated.

### 3.2.4.2 Gibson assembly of the *PM2-2A-PM3* overexpression plasmid

Briefly, the *mrfp-2A-gfp* segment was excised from the pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid using the enzymes, AvrII and XhoI (NEB) (Figure 3.3A). The *PM2* gene was PCR-amplified from the *PM2-2A-gfp* plasmid made by this study and described above (3.2.4.1, Figure 3.2) using the primers, p1632 (forward - 5'-ACCTAATAGAAATATATCACCTAGG - 3') and p1633 (forward - 5'-GCCAGATCTTAAATTCTTTTGTAGCAAGAGC - 3') (Table 3.1). The reverse primer, p1633 includes the “T” (highlighted in red) that was not included in the previous plasmid to correct the shift in the reading frame (Table 3.1). A *2A-PM3* fragment was synthesized (Life Technologies) including regions for overlap with the pDC2 plasmid backbone and the *PM2* fragment. The three fragments (Figure 3.3A): *PM2*, *2A-PM3*, and digested pDC2 plasmid

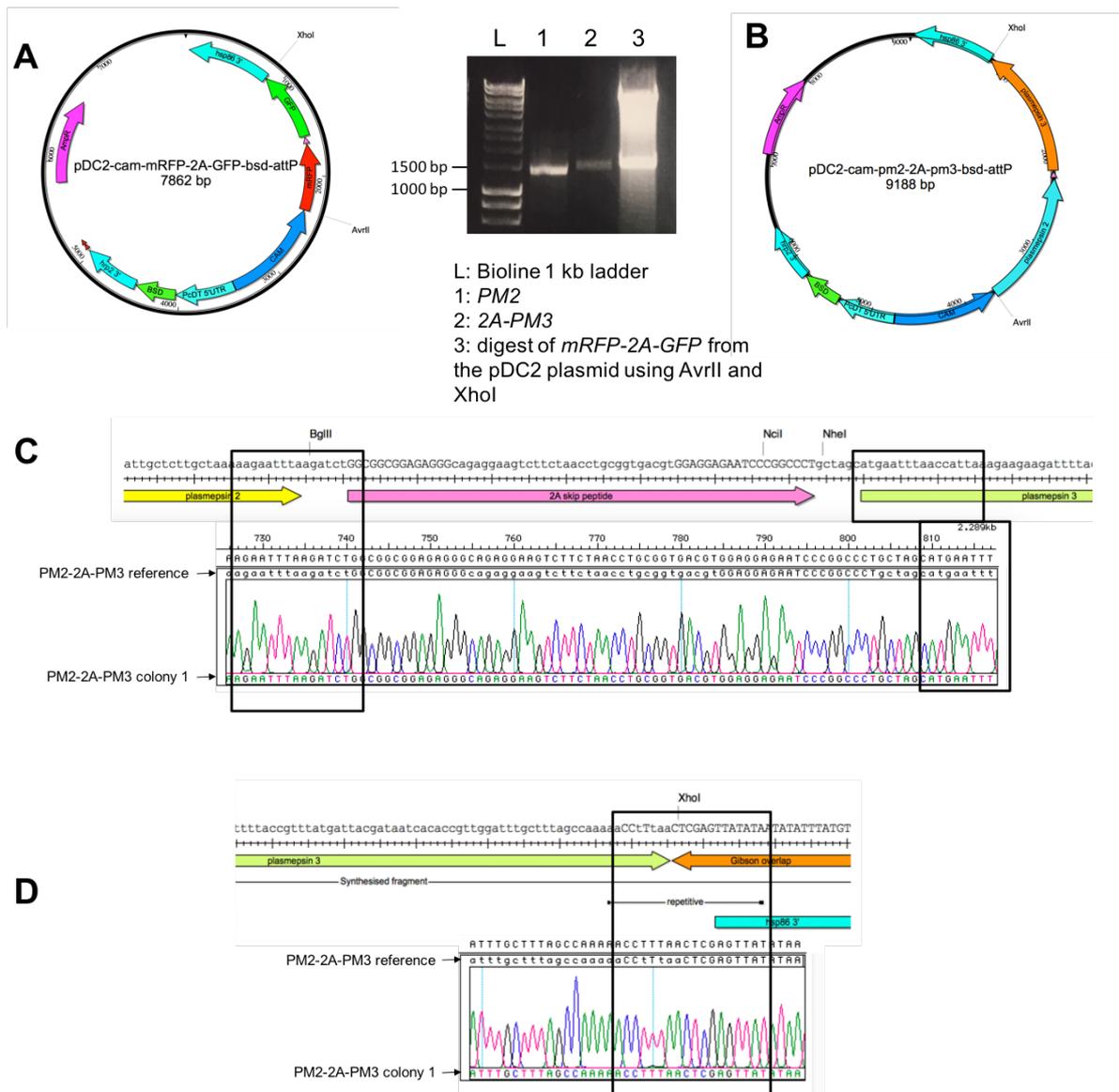
(100 ng) were assembled using a ratio of 2:1 = insert:vector and incubated at 50°C for 20 minutes. A no insert control with digested plasmid only was tested using an equivalent amount of plasmid and no insert. After assembly, 5 µl of the reactions were transformed into competent *E. coli* (Ch. 2, Section 2.8.1) and LB + ampicillin plates were screened for colonies the next day.

Correct insertion of *PM2* and *PM3* into the pDC2 plasmid was confirmed by sequencing (Figure 3.3C-D), using the primers: p364 (5'- GGGCCCCGCATGCTTAGCTAATTCG - 3') and p648 (5'- GATTCTTCTTGAGACAAAGGC - 3'), which produce a 649 bp product. Following sequence confirmation, the plasmid was midi-prepped (Ch.2, Section 2.8.2) and parasites were transfected (Ch.2, Section 2.3.2).

**Table 3. 1 Primers for the development of *PM2-PM3* overexpression plasmids**

Primer name	Primer (5' - 3') sequence	Description
PM2_OE_F	acctaataagaaatatatcacctaggATGGATATTACAGTAA GAGAACATG	PM2 forward primer with Gibson overlap
PM2_OE_R	ctctgccctctccgccgagatcTAAATTCTTTTAGCA AGAGCAATAC	PM2 reverse primer with Gibson overlap
p1632	ACCTAATAGAAATATATCACCTAGG	PM2-2A forward primer, binds CAM promoter upstream of <i>PM2</i> insertion site in the pDC2-cam-PM2-2A-GFP-bsd-attP plasmid
p1633	GCCAGATcTAAATTCTTTTAGCAAGAGC	PM2-2A reverse primer, binds 2A region downstream of <i>PM2</i> insertion site in the pDC2-cam-PM2-2A-GFP-bsd-attP plasmid

The “T” highlighted in red in p1633 is the nucleotide that was not included in the *PM2*-only plasmid reverse primer (PM2\_OE\_R). This corrected the shift in the reading frame in the *PM2-2A-PM3* plasmid

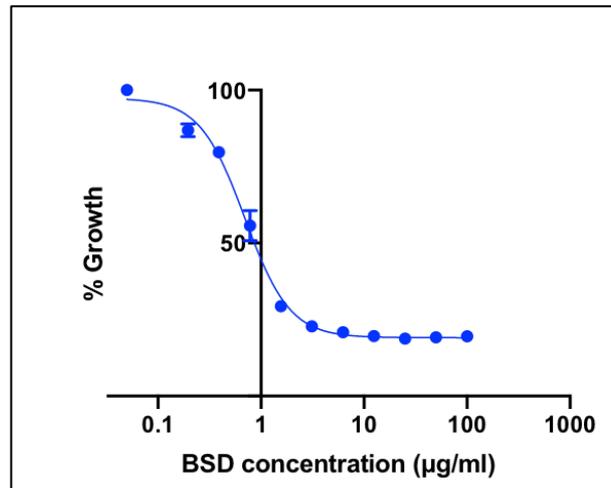


**Figure 3.3. Plasmid design and assembly for overexpression of *PM2* and *PM3*.** **A.)** The pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid was digested with *AvrII* and *XhoI* to obtain the plasmid backbone. The fragments for Gibson assembly: *PM2* (1384 bp), *2A-PM3* (1487 bp), and digested plasmid (6380 bp) with an excised *mrfp-2A-gfp* fragment (1482 bp) were run on a 1% agarose gel to confirm sizes. **B.)** The pDC2-*cam-PM2-2A-PM3-bsd-attP* plasmid after Gibson assembly. **C.)** Sequence of *PM2-2A-PM3* colony 1 aligned with the reference sequence of the inserted *PM2* through the 2A peptide to *PM3*. Black boxes indicate regions of overlap with the sequence alignment and plasmid map. **D.)** Sequence chromatogram of *PM2-2A-PM3* colony 1 aligned with the reference sequence through the end of the inserted *PM3* gene. Black boxes indicate regions of overlap with the reference sequence alignment and plasmid map. The top sequence above the reference is the consensus sequence for all reads.

### 3.2.4.3 Determination of drug selection concentration by $IC_{50}$ assays

Successful transfection of the episomal plasmid was verified by drug selection with blasticidin (BSD). To determine the concentration of BSD to use on the transfected field isolate, 163-KH3-005 (005), (a “PPQ-sensitive” parasite, **Ch. 2, Table 2.1**) an  $IC_{50}$  assay was performed (**Figure 3.4**). Based on the  $IC_{50}$  results, it was determined that 2  $\mu$ g/ml BSD,

which is the same concentration used for lab strains, could be used for the 005 isolate. The transfected parasites were kept under constant drug selection in order to maintain the episomal plasmid.



**Figure 3. 4. Dose-response curve for field isolate 163-KH3-005 after 72-hour exposure to blasticidin (BSD).** The curve represents 4 technical replicates with an  $IC_{50}$  value of 0.697 Error bars represent SD.

#### **3.2.4.4 Use of flow cytometry to screen for pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid in transfected parasites**

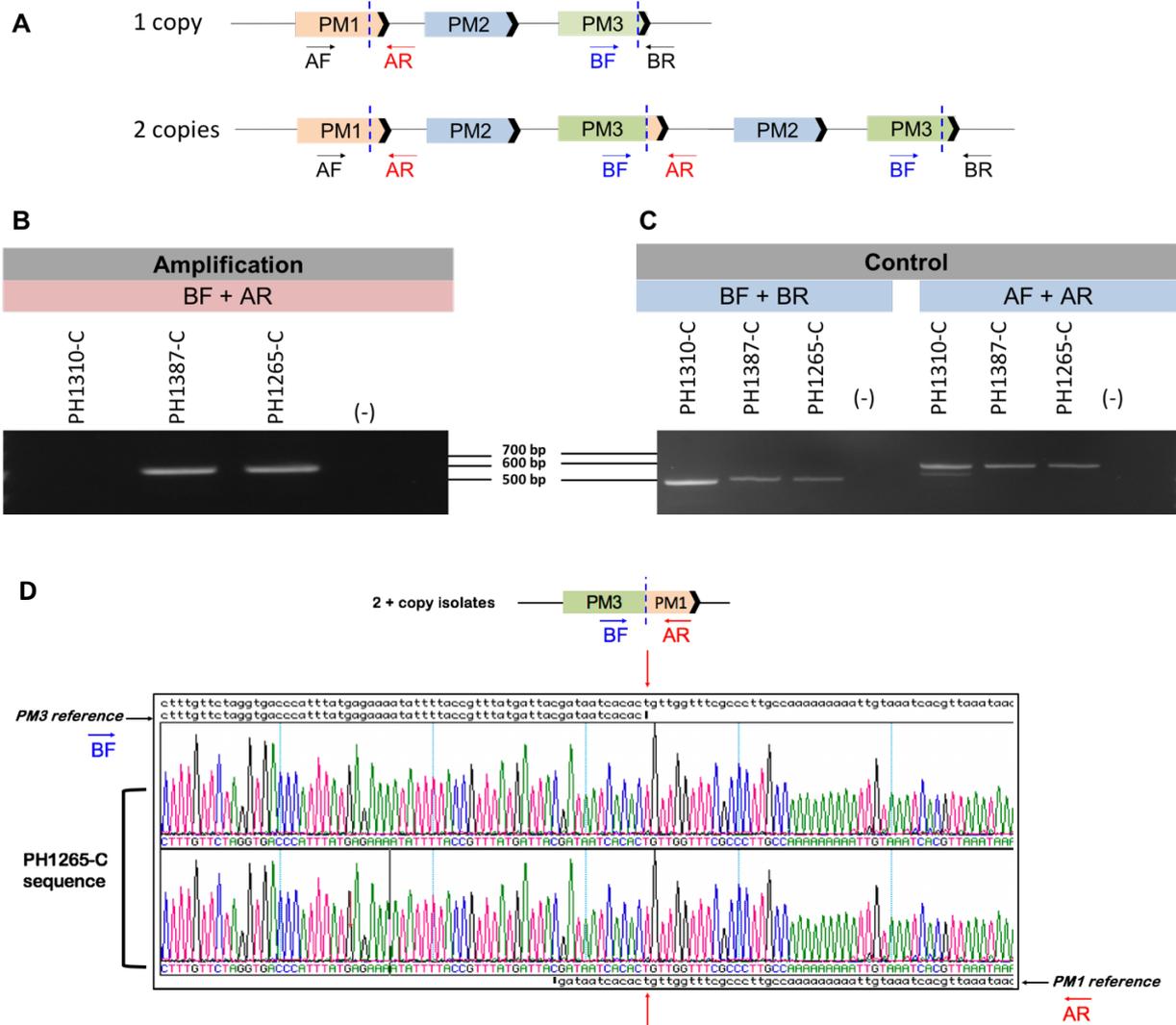
The transfected 005 and Dd2 parasites were screened for the presence of GFP using flow cytometry as described in Chapter 2 (**Ch. 2, section 2.7**). Approximately 50 µl of a 3% hematocrit culture transfected with the pDC2-*cam-mrfp-2A-gfp-bsd-attP* plasmid was added to 200 µl of PBS and washed twice, then resuspended in 200 µl of PBS and 10 µl of this dilution was added to a further 200 µl of PBS. No stain was added to any of the samples. A non-transfected 005 parent line was analyzed as a control along with the transfected 005 and Dd2 lines. Each sample was read on a CytoFLEX (Beckman Coulter) flow cytometer using the channels: FITC (to detect GFP) and PE (to detect RFP).

## 3.3 Results

### 3.3.1 *PM2-3* breakpoint assay validation

In order to develop a method for detecting the *PM2-3* copy number variations, I designed PCR primers that amplify the breakpoint of the *PM2-3* amplification observed in Cambodian isolates (**Figure 3.5A**). The breakpoint assay identifies copy number amplification in isolates that contain 2 or more copies of *PM2* and *PM3* (**Figure 3.5B**). As expected, no PCR products were observed in samples with a single copy of *PM2-3* (**Figure 3.5B**). Control primers confirmed that the regions surrounding the *PM2-3* amplification were present in all isolates (**Figure 3.5C**) and verified that absence of amplification using the duplication primer sets was not due to lack of DNA or low quality DNA.

Further verification of the breakpoint assay was performed via sequence analysis of the BF + AR PCR product. Sequence data revealed the same breakpoint location as observed by WGS (**Figure 3.5D**). All Cambodian samples that were positive for *PM2-3* amplifications as detected by the breakpoint assay (>1 copy) were in 100% concordance with qPCR and WGS data that called 2 or more copies of *PM2*. Sequencing chromatogram review of PCRs representing 2 and 3 copy samples showed that the breakpoints for representative 2 and 3 copy samples were identical. These sequencing results combined with the identical PCR sizes for all isolates indicates that the breakpoint is identical in all Cambodian isolates tested to date.



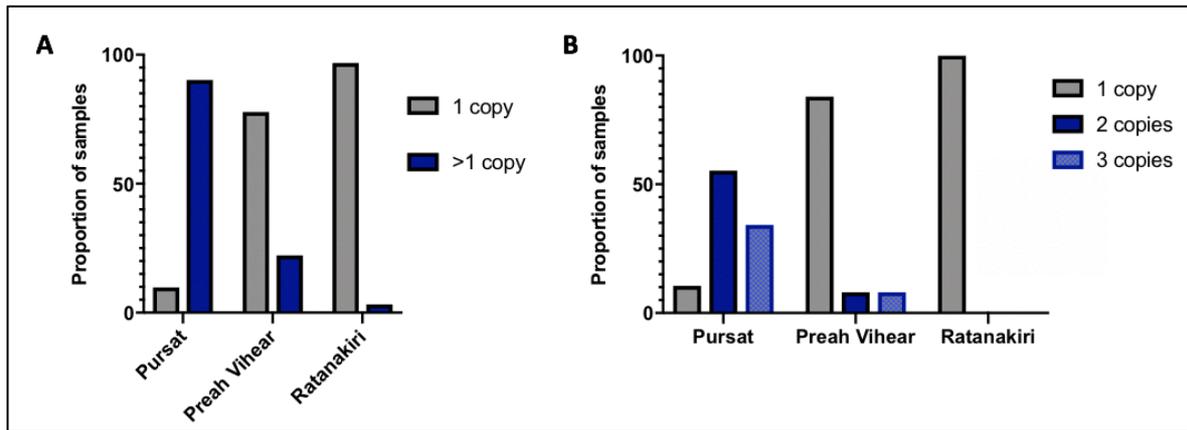
**Figure 3. 5. Schematic of *plasmepsin 2-3* gene duplication.** A.) Gene model depicting the *PM2-3* breakpoint (dashed blue lines) observed in Cambodian isolates. Primer positions are labeled in the single copy (top) and multi-copy (bottom) isolates. B.) Amplification primer set BF + AR amplifies a product in an isolate with two copies (PH1387-C) and three copies (PH1265-C) of *PM2-3*. No product is observed for the single copy (PH1310-C) isolate or in the DNA-negative control (-). C.) Control primers amplify a product in the single copy (PH1310-C) and multi-copy isolates (PH1387-C; PH1265-C). No product is observed in the DNA-negative control (-). D.) Sequence chromatograms of the breakpoint PCR product of primer set BF + AR using 3 copy isolate, PH1265-C. The primers sequence through the breakpoint between the 3' end of *PM3* and the 3' end of *PM1*. The sequence of isolate, PH1265-C is represented in color with the chromatograms. The reference sequences for *PM3* and *PM1* are shown in black and indicated with the black arrow. Viewed from the top, the *PM3* sequence stops (red arrows) and aligns with the *PM1* sequence due to the formation of the chimeric *PM3-PM1* gene. The sequence proceeds through the breakpoint (red arrows) to align with the sequence of *PM1* (bottom).

### 3.3.2 *PM2-3* breakpoint assay utility in comparison with SYBR green assay

To test the utility of using the breakpoint assay for rapid surveillance of large sample sizes for which WGS data is not yet completed or available, I performed breakpoint PCR on 99 samples from patients treated with DHA-PPQ in Cambodia from 2012-2013 (Amaratunga et al., 2016). I found that 90% of samples from Pursat had increased copies of *PM2-3*, 22% of samples from Preah Vihear had increased copies, and 3.2% of samples from Ratanakiri had increased *PM2-3* copies (**Figure 3.6A**).

The qPCR-based assay results for the same 99 samples from patients treated with DHA-PPQ showed approximately 90% of samples from Pursat had increased *PM2-3* copy numbers, while 16% of samples from Preah Vihear had increased copies and Ratanakiri samples were all single copy *PM2-3* by the qPCR method (**Figure 3.6B**).

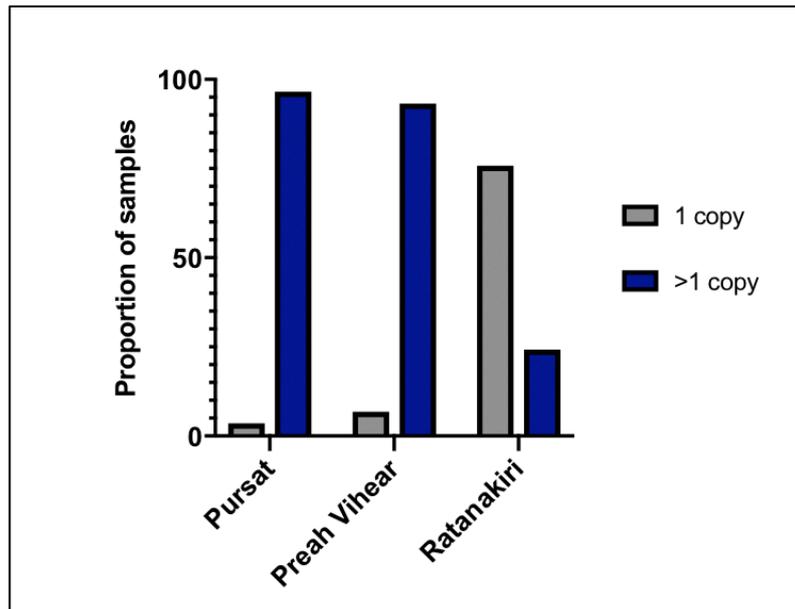
I compared the results of the breakpoint assay with the qPCR SYBR green assay and found that 93/99 samples tested with the breakpoint PCR assay matched qPCR data for the same samples, a 94% concordance. The six non-concordant samples were determined to be >1 copy by breakpoint PCR (**Figure 3.6A**), but qPCR read them as single copy isolates (**Figure 3.6B**). This included 1 sample from Ratanakiri, 2 samples from Preah Vihear, and 3 samples from Pursat that had >1 copy by breakpoint PCR but single copy by qPCR (**Figure 3.6**). The six non-concordant samples were repeated, always producing the same results. These results could suggest that the breakpoint PCR assay is more sensitive than the qPCR assay for detecting the breakpoint in samples that are polyclonal. Minor clones containing the duplication in field isolates would not be detected in the qPCR assay, however, they would be detected by the breakpoint assay. One limitation of the breakpoint assay is that there is no high-throughput way to determine if these samples are false positives. The best way to check would be to examine WGS data for each sample, however these samples were selected for breakpoint PCR because WGS data for the samples either failed or was not chosen for WGS analysis due to insufficient or low-quality DNA. Another way to assess if the six non-concordant breakpoint PCR samples were false positives would be to sequence the PCR product and see if the *PM3-PM1* chimeric sequence is present, which would provide evidence that the CNV is likely present.



**Figure 3. 6. Comparison of the breakpoint PCR and qPCR assays for measuring *PM2-3* copy numbers.** A.) Proportion of samples with 1 copy or >1 copy of *PM2-3* as determined by breakpoint PCR. B.) Proportion of samples with 1, 2, or 3 copies of *PM2* as determined by qPCR analysis. For both breakpoint PCR (A) and qPCR (B) assays a total of 99 isolates were analyzed from 3 Cambodian provinces during 2012-2013 after treatment with DHA-PPQ.

### 3.3.3 Analysis of *PM2-3* copy number in isolates from patients treated with artesunate-mefloquine

The breakpoint PCR assay was utilized to detect the *PM2-3* breakpoint in samples from a clinical trial conducted in 3 provinces in Cambodia from 2014-2015 to test the efficacy of artesunate-mefloquine (AS-MQ) (clinicaltrials.gov ID: NCT01736319) (Amaratunga et al., 2019). The prevalence of *PM2-3* amplifications in 295 total patient isolates from Pursat, Preah Vihear, and Ratanakiri were 96%, 93%, and 24%, respectively (**Figure 3.7**). As seen in Figures 3.6-3.7, the proportion of multi-copy parasites is continuing to increase in the 3 Cambodian provinces from 2012 to 2015 (**Figures 3.6-3.7**).



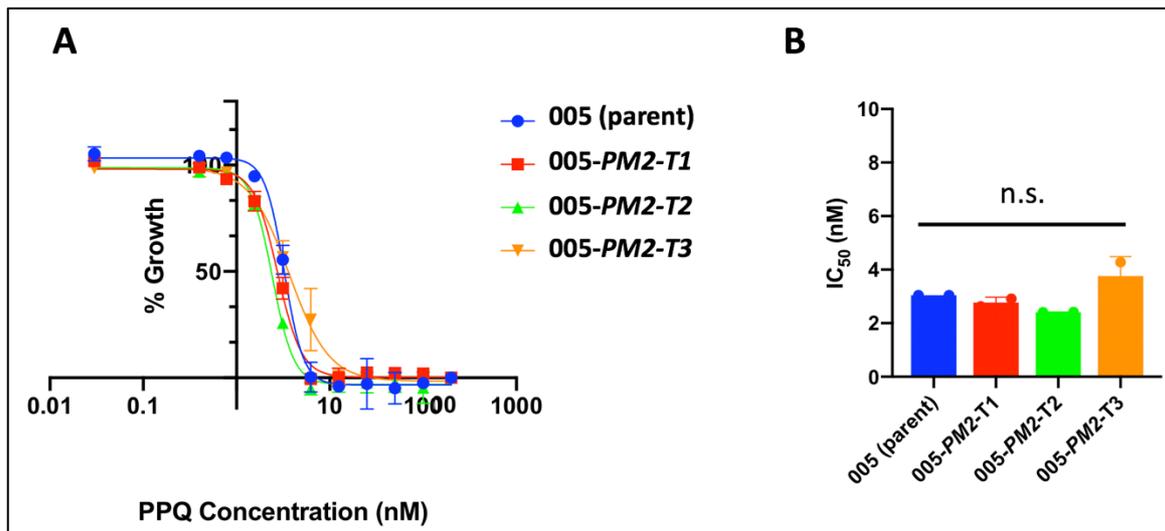
**Figure 3. 7. Proportion of samples with >1 copy or 1 copy of *PM2-3* as determined by breakpoint PCR analysis in 3 Cambodian provinces from 2014-2015.** A total of 295 patient isolates from 3 Cambodian provinces were analyzed after treatment with AS-MQ.

### 3.3.4 Overexpression of *PM2* and *PM3* in Cambodian isolates

It was hypothesized that the increased *PM2-3* copy numbers are not only associated with piperazine resistance, but could also serve a causal role in the resistant phenotype. To test this hypothesis directly, this study used transgenic approaches to generate isogenic parasite lines that overexpressed *PM2* and *PM3* in stoichiometric amounts. The best way to do this was to link expression of both genes to a single promoter, with the genes separated by a 2A viral skip peptide. The peptide results in translation that continues to the downstream gene with a break in the peptide bond, and has been used when 1:1 stoichiometric expression of two genes is desired, such as for immunoglobulin heavy and light chains or T-cell receptors (Szymczak et al., 2004). The plasmid used by this study was codon-optimized for *P. falciparum* and used successfully by Straimer *et al.* for dual protein expression of RFP and GFP (Straimer et al., 2012).

#### 3.3.4.1 Overexpression of *PM2* in Cambodian field isolate, 163-KH3-005

This study first assessed if overexpression of *PM2*-only had any effect on susceptibility to piperazine. I transfected the *PM2* overexpression plasmid into a PPQ-sensitive isolate from Cambodia, 005, and found little to no differences in parasite susceptibility to piperazine in the transgenic lines compared to the control (not transfected) parasites in IC<sub>50</sub> assays (**Figure 3.8**).



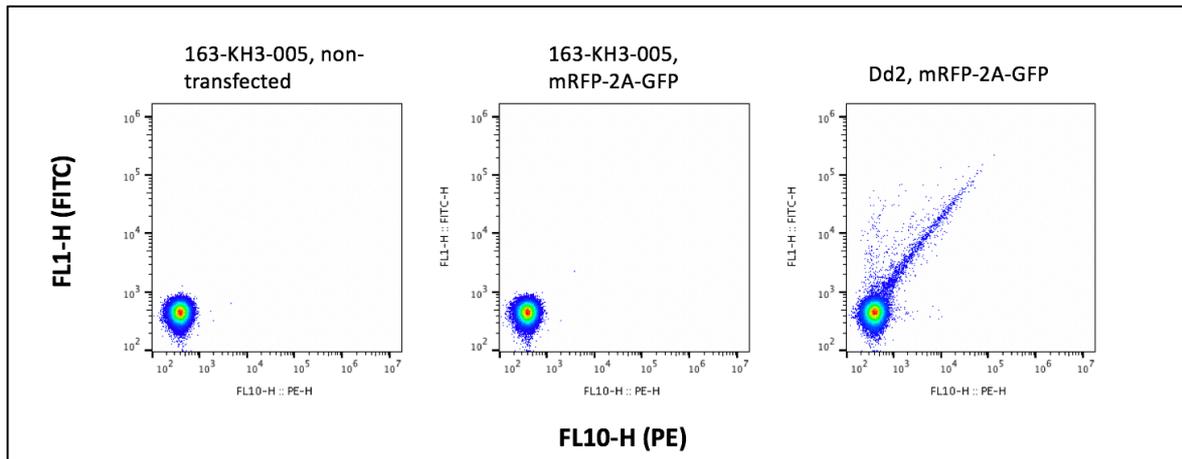
**Figure 3. 8. Transgenic piperazine-sensitive parasites (005) carrying the *PM2* overexpression plasmid showed no shift in PPQ  $IC_{50}$  values.** A.) Dose-response curves for three 005-transfectants with the *PM2* overexpression plasmid (T1, T2, T3) compared to the parental line (005). Error bars represent SD (n=2). B.) Significance was determined using Mann-Whitney U tests comparing the PPQ  $IC_{50}$  values of the three 005-*PM2* transfectants with the 005 parent line. n.s. indicates not significant ( $p>0.05$ ).

### 3.3.4.2 Confirmation of the pDC2-cam-*PM2*-2A-*PM3* plasmid and pDC2-cam-*mrfp*-2A-*gfp*-*bsd*-attP plasmid in transfected parasites by flow cytometry and PCR

To fully recapitulate the *PM2*-3 CNV observed in clinical isolates, I next aimed to overexpress both *PM2* and *PM3* in the 005 field isolate and Dd2.

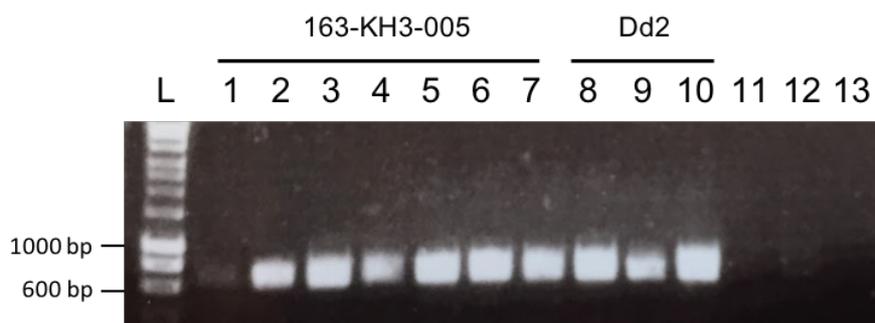
Field isolate, 005 and lab strain, Dd2 were transfected with the *PM2*-2A-*PM3* overexpression plasmid and the pDC2-cam-*mrfp*-2A-*gfp*-*bsd*-attP plasmid as a transfection control.

As soon as there was a detectable parasitemia (around days 14-15 post-transfection for both lines), the 005 isolate and Dd2 transfectants with the control pDC2-cam-*mrfp*-2A-*gfp*-*bsd*-attP plasmid were screened for the presence of GFP and RFP by flow cytometry (Figure 3.9). Fluorescence was detected in the Dd2 transfected parasite, as demonstrated by the presence of a population of cells positive in the FITC (GFP) and PE (RFP) channels (Figure 3.9, right panel). However, no fluorescence in the FITC or PE channels was observed for the 005 transfected parasite (Figure 3.9, middle panel), which showed a similar profile to the non-transfected 005 parent line (Figure 3.9, left panel). The experiment was repeated twice and detection of GFP and RFP was only observed in the Dd2 transfected parasites.



**Figure 3. 9. Flow cytometry assessment of transgenic parasites.** Flow cytometry assessment of GFP transgene expression (FITC channel) and RFP transgene expression (PE channel) in transfected parasite lines 005 and Dd2, compared with a non-transfected control line (left panel).

Although the *pDC2-cam-mrpf-2A-gfp-bsd-attP* plasmid was transfected as a control to monitor transfection success by measuring GFP/RFP fluorescence, it could have no relationship to whether or not the *PM2-2A-PM3* plasmid was successfully transfected. Therefore, to determine if the 005 and Dd2 parasite lines were successfully transfected with the both the *PM2-2A-PM3* plasmid and *pDC2-cam-mrpf-2A-gfp-bsd-attP*, the transfections were screened using plasmid-specific primers (**Figure 3.10**). As seen in Figure 3.10, both plasmids are present in the 005 and Dd2 transfected lines. Due to time constraints the PPQ response of the *PM2-3* transgenic lines was not yet tested (as well as RNA and protein expression) but will be prioritized in future work.



**Figure 3. 10. PCR confirmation of the *PM2-2A-PM3* plasmid in transfected 005 and Dd2 parasites.** PCR confirmation that the *pDC2-cam* plasmid backbone was present in transfections. Lane L is the 1 kb ladder (Bioline); lanes 1-2 and 5-7 are field isolate 005 transfected with the *pDC2-cam-PM2-2A-PM3* plasmid and lanes 3-4 are field isolate 005 transfected with the control *pDC2-cam-mrpf-2A-gfp-bsd-attP* plasmid. Lane 8 is Dd2 transfected with the control *pDC2-cam-mrpf-2A-gfp-bsd-attP* plasmid and lanes 9-10 are Dd2 transfected *pDC2-cam-PM2-2A-PM3* plasmid. Lanes 11 and 12 are the non-transfected parental lines 005 and Dd2, respectively. Lane 13 is the no template control.

### 3.4 Discussion and future work

The novel breakpoint and qPCR assays developed by this study are sensitive, reliable, and cost-efficient methods for detecting the *PM2-3* gene amplifications. They can be used in place of WGS or in areas where sequencing is not feasible, as they require considerably less time and facilities, allowing for copy number analysis within the same day of sample collection. Both assays can be performed with very little genomic DNA (2-5 ng of gDNA) which is useful in many endemic regions where only filter paper blood spots are available. The qPCR and breakpoint assays are also high-throughput, with the possibility of using 96-well plate analysis, making them ideal for surveillance throughout malaria endemic regions.

The qPCR-based *PM2-3* CNV assay measures discrete number of copies, which is especially useful in determining if the number of *PM2-PM3* copies is increasing in parasite populations throughout countries where DHA-PPQ was previously used or is still currently in use. One disadvantage of the qPCR method is the arbitrary relative assignment of 1 vs. 2 copies. For the purposes of this study, a value of “2 copies” was set as a relative copy number of 1.5 or greater. This leaves the potential that a sample read as 1.44 copies in the qPCR assay, for example, would be counted as single copy, when it actually had two copies of *PM2-3*. The same could be said for a 1.5 relative copy number actually being a single copy parasite, rather than multicopy. In spite of the inherent potential errors with relative copy number, because qPCR methods do not rely on the primer location in reference to an estimated breakpoint, they can be used broadly.

The breakpoint PCR assay (**Figure 3.4**) also effectively determines increased *PM2-3* copy number in field isolates and can be used in areas where qPCR is infeasible. Since the assay detects the presence of the breakpoint, it does not have the same problems as WGS and qPCR with calling average or relative copy numbers and is positive only if the breakpoint is present. The sensitivity of this assay also enables the detection of minor clones, which proves advantageous, given the potential for polyclonal infections. The basic laboratory equipment needed for breakpoint-PCR makes it possible for analysis to be conducted at field sites, rather than shipping the samples to another laboratory within the country or outside of the country to wait for results. However, a notable disadvantage of this method is that the breakpoint assay was designed with primers specific to the *PM2-3* breakpoint observed in Cambodian field isolates (Amato et al., 2017). The PCR primers used to detect the Cambodian breakpoint observed to date may not amplify a product in samples that contain different breakpoints. As previous studies have shown (Menard et al., 2013, Hostetler et al., 2016),

copy number variations in the same gene can arise independently on different genetic backgrounds and result in distinct molecular breakpoints of amplification. Though this study has observed over 300 samples from Cambodia with the same breakpoint (**Figures 3.6-3.7**), Witkowski *et al.* (Witkowski et al., 2017) have reported different breakpoints in Cambodian isolates. Therefore, both breakpoint and qPCR assays should be performed for the most accurate results, when possible.

This thesis (**Figures 3.6-3.7**) and other studies (Amato et al., 2017, Jacob et al., 2019, Witkowski et al., 2017) have shown that parasites with the *PM2-3* copy number expansion are increasing in prevalence in Cambodia. The breakpoint and qPCR assays have shown that the proportion of parasites with more than one copy of *PM2-3* has increased in the 3 Cambodian provinces where samples were collected from 2012 to 2015 (**Figures 3.6-3.7**). These parasite samples have been obtained from clinical studies carried out by Amaratunga *et al.* in 2012-2013 (Amaratunga et al., 2016) and 2014-2015 (Amaratunga et al., 2019). From these clinical trials, we know that the patients in all 3 provinces (Pursat, Preah Vihear, and Ratanakiri, **Ch. 2 Figure 2.2**) were treated with the same drug regimens, specifically DHA-PPQ in the 2012-2013 trial followed by AS-MQ in the 2014-2015 trial. As discussed in Chapter 2 (**Section 2.2.1**), field isolates from these three Cambodian provinces from 2012-2013 varied in their susceptibility to PPQ, with parasites from Ratanakiri (Eastern Cambodia), being largely susceptible to PPQ and also containing single copies of *PM2-3*. It will be important to determine if the parasites with increased *PM2-3* copy numbers from Ratanakiri in 2014-2015 (**Figure 3.7**) remain susceptible to PPQ, since the parasite samples from this study were successfully treated with AS-MQ, but not exposed to DHA-PPQ. The 2014-2015 parasite samples were the most recent samples this thesis had access to, but recent studies by van der Pluijm *et al.* reported that *PM2-3* amplifications were present in 72% of samples from Ratanakiri collected from 2015-2018 (van der Pluijm et al., 2019). Continued and future work is thus needed to monitor if and how the *PM2-3* copy number expansion changes in this region. In areas where WGS is not possible, the breakpoint and qPCR methods designed by this study can provide efficient, cost-effective methods for real-time analysis of *PM2-3* copy numbers.

In addition to assays for monitoring the *PM2-3* CNV, I developed methods to test for the effects of increased copies of *PM2-3* on piperazine response. To mimic the CNV observed in field isolates encompassing both *PM2* and *PM3*, I generated both a *PM2*-only overexpression vector (**Figure 3.2**) and a *PM2-2A-PM3* vector to co-express stoichiometric

amounts of both genes using the 2A skip peptide (**Figure 3.3**). Initial comparisons of *PM2*-only overexpression in PPQ-sensitive isolate, 005, showed no significant difference in parasite survival to piperazine (**Figure 3.7**). This could indicate that both genes, *PM2* and *PM3*, are needed to obtain a resistant phenotype. To test this hypothesis, I transfected the same PPQ sensitive isolate, 005, and the Dd2 lab strain with the *PM2-2A-PM3* vector. Confirmation of the presence of the plasmid in the PPQ-sensitive field isolate 005 and Dd2 lab strain was demonstrated via PCR (**Figure 3.10**). I also transfected the unmodified pDC2-*cam-mrffp-2A-gfp-bsd-attP* plasmid as a transfection control (**Figure 3.10**). Curiously, the study was only able to detect GFP and RFP fluorescence in the Dd2 transfectants (**Figure 3.9**) and not the 005 isolate, although both the Dd2 and 005 transfectants were PCR-positive for the plasmids and maintained on the appropriate BSD selection pressure to maintain the episomal plasmid (**Figure 3.10**). Since the pDC2-*cam-mrffp-2A-gfp-bsd-attP* plasmid was used as a transfection control to determine if the parasites could be successfully transfected, the lack of GFP and RFP fluorescence in the 005 transfectants will not be pursued further, since the PCR results and the growth of the parasites on BSD pressure show that both plasmids (*mrffp-2A-gfp* and *PM2-2A-PM3*) are present in the transfected parasites. However, plasmid-rescue of the episomal plasmid from the 005 parasites, and sequencing of the *mrffp-2A-gfp* cassette would be the first step in determining why expression of the reporter genes is not observed.

Future work will now prioritize further characterization of the *PM2-3* transfected lines. Before performing further parasite survival assays to assess the impact of *PM2-3* overexpression on PPQ sensitivity, it is necessary to examine both *PM2-3* RNA and protein levels to determine if the overexpression plasmids have resulted in increased RNA and protein in the transfected parasites compared to the non-transfected parent lines. Samples have been saved for qPCR analysis to compare *PM2-3* transcript expression levels and preliminary Western blots to compare protein expression in the parental non-transfected and transfected lines have been performed but must be optimized.

Recent functional studies have provided insight into the role of the *PM2-3* amplification in response to antimalarial drug pressure. Loesbanluechai *et al.* 2019 (Loesbanluechai *et al.*, 2019) demonstrated that overexpression of either *PM2* or *PM3* individually in the 3D7 parasite background had no effect on parasite susceptibility to piperazine, artemisinin, and chloroquine. Consistent with this result, I did not observe any survival difference in response to piperazine in initial studies on the PPQ-sensitive isolate, 005 overexpressing only *PM2*

(Figure 3.7). Interestingly, Bopp *et al.* 2018 found that increased *PM2-3* copy numbers in PPQ-resistant field isolates enhanced survival to PPQ (Bopp *et al.*, 2018). It should be noted that the isolates with increased *PM2-3* copy numbers and survival under PPQ pressure in the study by Bopp *et al.* were *kelch13* C580Y or Y493H mutants, and were not generated by transgenic methods but were naturally multicopy for *PM2-3* (Bopp *et al.*, 2018). Additional studies on *PM2* and *PM3* knockouts in 3D7 parasites showed decreased survival to PPQ, though the differences in PPQ IC<sub>50</sub> survival assays appeared minor (Mukherjee *et al.*, 2018). It is therefore possible that the *PM2-3* amplification needs to be present in the same genetic background of artemisinin resistance (K13 mutant or other potential genetic variants). Furthermore, it may be necessary for the *PM2-3* amplification to be in the presence of other molecular markers of PPQ resistance for the effects of the amplification to manifest. In subsequent chapters, I will explore other potential markers of piperazine resistance, including SNPs in the putative exonuclease protein, PF3D7\_1362500 (exo-E415G), putative mitochondrial carrier protein, PF3D7\_1368700 (MCP-N252D), and PfCRT. It is also necessary to further examine the role of knocking out or knocking down *PM2* and *PM3*, thus future functional work is critical.

I have made over four separate attempts (in duplicate) to transfect the *PM2-2A-PM3* plasmid into multiple field isolates with K13 C580Y mutations, one of the molecular markers of artemisinin resistance. Additional attempts were made with the generous help of my lab colleague, Dr. Sophie Adjalley. All attempts to date have been unsuccessful. Studies have shown that laboratory strains such as Dd2 and 3D7 are genetically different from more recently adapted field isolates (Mackinnon *et al.*, 2009), which could contribute to difficulties experienced in field isolate transfections. Due to the lack of additional field isolates transfected with the *PM2-2A-PM3* plasmid, my studies in field isolates with various genetic backgrounds (K13, exo-E415G, MCP-N252D, PfCRT) are ongoing. Transfection protocols have been optimized using different machines (Bio-Rad electroporator and Amaxa electroporator) and conditions (pre-loading RBCs versus direct transfection of ring-stage parasites and use of conditioned media) but these changes have yet to result in transfected field isolates. Transfections of schizonts rather than loaded RBCs or ring-stage parasites has also been widely used for transfections of *P. knowlesi* (Moon *et al.*, 2013) and could be attempted with the field isolates used in this study. Additional transfection attempts with field isolates could also vary the concentrations of BSD used for plasmid selection (increasing the BSD concentration) or perhaps different drug resistance genes could be used. Nevertheless,

progress has been made in transfecting the 005 isolate with the dual *PM2-2A-PM3* overexpression plasmids and future work on RNA and protein expression is forthcoming.

As antimalarial drug resistance continues to thwart treatment efforts, it is imperative to have assays and functional studies to assess molecular markers of drug resistance. This thesis has developed two novel assays that can be used to monitor for the *PM2-3* molecular marker of piperazine resistance. In a proof of concept of assay reliability and ease of use, I spent one month in Cambodia (April 2017) teaching our NIH-based team in Phnom Penh how to perform the assays, where the procedures were quickly introduced without any technical issues. In addition to the copy number assays, I have also developed a *PM2-3* overexpression plasmid that can be used to assess the effects of stoichiometric overexpression of both *PM2* and *PM3*. Though future work is necessary, the tools developed by this study have formed a fundamental basis for monitoring and determining the role of the *PM2-3* amplification in piperazine resistance.