nature portfolio

Peer Review File

High-Throughput Genotyping of Plasmodium vivax in the Peruvian Amazon via Molecular Inversion Probes

Corresponding Author: Dr Zachary Popkin-Hall

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Popkin-Hall and colleagues have standardized four genotyping panels for the neglected human malaria parasite Plasmodium vivax, using molecular inversion probes (MIP), and applied this novel genotyping tool to 866 field-collected isolates collected in the Amazon Basin of Peru between 2011 and 2018.

The development of a new genotyping tool and its application to a large number of samples are both major accomplishments. However, in its current format, the manuscript provides neither a detailed account on the development and use of the MIP panels nor an appropriate population genomics analysis.

Methods:

1. Global variation study. The authors state in the Supplemental Material (lines 28-30) that “918 published P. vivax sequences representing all known NGS data were downloaded from the European Nucleotide Archive (accession numbers listed in Supplemental Table 2) using a custom bash script”. However, in the main text they cite a recent paper (reference 33) that describes 1,895 high-quality P. vivax genome datasets. From Supplemental Table 2 it becomes clear that (a) several genome sequences included in the global variation study were not derived from reference 33 and (b) many (most likely, most) genome sequences from reference 33 were not considered in the global variation study. Why? How “global” can this analysis be?

2. “Drug resistance mutations”.The authors correctly state that “there are no experimentally validated antimalarial resistance markers in P. vivax” (line 62, Supplemental Material), but have selected for inclusion in their MIP panel some “mutations associated with reduced antimalarial efficacy in field settings” in the following genes: chloroquine resistance transporter (Pvcrt), bifunctional dihydrofolate reductase-thymidylate synthase (Pvdhfr-ts), multidrug resistance protein 1 (Pvmdr1), and hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase (Pvpppk-dhps). Unfortunately, however, the reported association between mutations in the Pvcrt and Pvmdr1 genes and “chloroquine resistance” is much looser than originally thought.

Regarding Pvmdr1, a recent review (doi: 10.1016/j.ijpddr.2021.04.002) concludes that:

“Sequencing of pvmdr1 across several regions of the world has revealed more than fifty polymorphisms in this gene, as well as copy number variants (CNVs). These single nucleotide polymorphisms (SNPs) correspond to amino acid changes throughout the protein sequence, including in the ATP binding domains and multiple transmembrane regions (Sá et al., 2005). No single SNP, or set of SNPs, have emerged as definitive drug resistance markers. However, six SNPs have been reported at high frequency in multiple studies in regions with reported drug resistance; (relative to the SAL-1 reference) S513R, G698S, M908L, T958M, Y976F and F1076L (Brega et al., 2005; Huang et al., 2014; Joy et al., 2018; Orjuela-Sánchez et al., 2009; Vargas-Rodŕiguez et al., 2012). Amino acid changes at Y976F and F1076L in particular have been cited as possible markers of drug resistance (Brega et al., 2005; Spotin et al., 2020; Suwanarusk et al., 2008). However, these SNPs are also found in regions without reported CQR, making their association with drug resistance uncertain (Spotin et al., 2020). The T958M, Y976F, F1076L variants (and possibly others) in PvMDR1, have been shown to arise independently, even within the same population (Schousboe et al., 2015). This finding suggests that pvmdr1 mutations can arise on different genetic backgrounds and that they have not been subject to a selective sweep, possibly due to a lack of

direct pressure from drug (Schousboe et al., 2015). If, and how, these SNPs play a role in mediating P. vivax drug resistance remains to be directly demonstrated in functional studies."

Data are even less conclusive for Pvcrt (doi: 10.1016/j.ijpddr.2021.04.002):

"(…) very few SNPs (~10) in pvcrt have been reported, however, most occur at very low frequency (Barnadas et al., 2008; Joy et al., 2018; Noisang et al., 2020). The most common PvCRT polymorphism is a lysine insertion at position 10 (K10) (Barnadas et al., 2008; Joy et al., 2018; Noisang et al., 2020; Silva et al., 2018a). [Is this the "K9 duplication" mentioned in Popkin-Hall's Table 1?] The K10 insertion has been observed in both Southeast Asian and South American parasites (Noisang et al., 2020; Silva et al., 2018b.; Zhao et al., 2020). However, no association between the K10 insertion and in vitro P. vivax drug resistance has been found (Silva et al., 2018a; Suwanarusk et al., 2007). It is currently unknown if pvcrt variants that may be associated with CQR have a deleterious impact on parasite fitness, which is the case for P. falciparum and could explain the low frequency of pvcrt mutations (Ord et al., 2007; Petersen et al., 2015)."

Mutations at the Pvmdr1 and Pvcrt loci might be worth studying even if their impact on drug resistance remains very unlikely. However, labelling them as "putative drug resistance mutations" (as in Table 1) is misleading given available evidence.

3. Parasite samples. In a genomic epidemiology study, the study area and population are expected to be described in detail. However, parasite samples used in the genotyping study were succinctly described in the main text as follows: "Samples were collected by NAMRU SOUTH between 2011 and 2018, primarily in the vicinity of Iquitos, the major city in the Peruvian Amazon, within Loreto Department. The sampling scheme has been described in detail elsewhere [62,63]" (lines 350-352). Only this! Importantly, references 62 and 63 do not provide additional information regarding the sampling scheme for P. vivax samples included in the current study: reference 62 describes an analysis of a convenience sample of P. falciparum isolates for HRP2/HRP3 mutations and reference 63 describes a study of P. vivax diagnosis carried out on venous blood samples from febrile patients between 2019 and 2020!

There is no information on levels of malaria transmission (e.g., API) in the study sites at the time samples were collected. Are APIs comparable among sites? How patients were selected and enrolled is not mentioned. Is this a convenience sample or there was a population-based study in these communities? Were all samples collected from symptomatic patients? Range of parasite densities? Were all infections detected by microscopy or were there submicroscopic infections? Were the authors able to fully genotype at least some submicroscopic infections?

4. DNA extraction. The authors succinctly state that "DBS [not spelled out] were extracted using the Chelex method, as previously described [32,64] (line 354)." If the newly described genotyping method is to be widely used, please indicate the starting blood volume (and minimum parasitemia) for extraction, the concentration of genomic DNA after extraction etc.

5. Sample processing. Were samples tested with quantitative PCR prior to sequencing? Can the authors correlate cycle threshold (Ct) values with MIP coverage to make recommendations for readers? For instance, according to Aydemir et al. 2018 (https://doi.org/10.1093/infdis/jiy223): "MIP coverage began to drop off at qPCR cycle threshold (Ct) values of 31, which equates to a parasite level of approximately 100 parasites/uL" (…) "The median Ct value of the samples that yielded no MIP coverage was 34".

5. Genotyping. What is the minimum parasite density required for reliable genotyping results? In the Discussion (lines 217-218), the authors state that their method works "reliably on P. vivax
isolates with parasitemias ≥100 parasites/μL" They mention in the Supplemental Material that "(…) sample missingness was very low (<12%) for PvG even at low parasitemia (≤100 parasites/μL)", but no data are provided to support this statement. Moreover, the authors do not explain how they dealt with missing data in PCA, IBD and other analyses.

6. Filtering. Sequence filtering was based on allele frequency and UMI counts. Were any other quality metrics used in filtering? How many samples were covered with anUMI depth of at least 3? How polyclonal samples were handled for IBD analysis? Does the MIPanalyzer accommodate polyclonal samples? Please indicate the functions used for IBD analysis, as they are not listed in https://mrc-ide.github.io/MIPanalyzer/reference/index.html.

Results:

1. Clonal outbreak in Belén. From a genomic epidemiology perspective, this is the most interesting finding, but it remains largely unexplored. The reader is not provided with basic information to interpret the finding of a clonal outbreak spanning several months in one of the study sites. How were the 269 samples collected in Belén? How many (sampled and unsampled) infections were diagnosed in the community during the study period? Were all sample donors symptomatic? Can we take the 269 samples as representative of P. vivax lineages causing clinical malaria in the site? How connected is Belén with other study sites? What is the maximum percent IBD detected in pairs of samples from different communities/sites?

2. Monoclonal infections. The finding of a vast majority (91.4%) of monoclonal P. vivax infections in Loreto is rather surprising, given the relatively high levels of genetic diversity of malaria parasites and the relatively intense transmission during the study period (2011-2018). Several studies using a range of genotyping methods (e.g., microsatellites, AmpliSeq etc.) have shown much higher proportions of multiple-clone infections in P. vivax populations from sites around Iquitos (see, e.g., https://doi.org/10.1038/s41598-024-66925-x). Any explanation for this unexpected finding? Any change MIP genotyping

failed to detect a large proportion of minor clones?

3. Copy number variation. For reproducibility, please indicate cut-offs values for UMI depth prior to CNV calculation.

Multicopy genes have a faster rate of evolution than single-copy genes, which reflects in an increased nucleotide diversity (π) (e.g.: https://doi.org/10.1186/1471-2164-10-353). The comparison of π between duplicated genes and single-copy genes can give greater strength to these findings.

(Remarks on code availability)
As noted above, the functions used for IBD analysis are not listed in https://mrc-ide.github.io/MIPanalyzer/reference/index.html. Otherwise, information appears to be complete.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Overview

The study by Popkin-Hall and colleagues presents the first set of MIP panels for P. vivax molecular surveillance. This is an important advance, and the panels have potential to be informative for both malaria control program staff and researchers. However, the major flaw of the manuscript as it currently stands is that there is far too little information on the laboratory methods. Critical details such as the probe sequences are missing, rendering it impossible for potential users to implement the assays. Given that the assays are the main novelty of the study, as the biological and epidemiological insights from the Peruvian data are not likely to be of huge interest to others than those with very closely related interests in this geographic region, it is essential for the authors to provide more information on these methods to allow the assays to be reproduced in other laboratories.
In addition to the lack of detail on the laboratory methods, there was very little information on the evaluation and optimization of the assays. Critical details that one would want to know about before committing to setting up the assays were missing or difficult to derive, such as assay specificity, ability to detect polyclonal infections, and expected read depth for more than just the PvG panel.
In principle, the marker selection across the four panels sounds impressive and we were excited to see its application, but the results were unfortunately very sparse and didn't quite deliver.
Specific suggestions on areas for improvement are provided below.

Major comments
1. Critical information to allow reproducibility of the assays needs to be provided, including:
a. List of targets in all of the panels (not just the PvG panel)
b. Primer/probe sequences
c. Methods for the library preparation steps

2. Given that the authors have gotten the assays to work on dried blood spot samples (not an easy task!), I think there will be many keen users. However, more information on the evaluation and optimization of the laboratory assays would be very helpful for potential users to gauge the suitability of the assays for their own use. Specifically:
a. As MIPs are quite new to the malaria field (this is indeed why these assays are fairly novel), it might help to give more description on how UMIs correlate with e.g. amplicon sequencing read depth and coverage stats. This is key to really appreciating the amount of data one can expect from these assays.
b. Also, regarding data quantity, it wasn't clear why there were only heatmaps for the PvG panel – what about the performance in other panels?
c. More information within the legend of the heatmaps would help interpretation. What are the grey bands – zero coverage or depth?
d. There was no description on the specificity of the assays. Is there any cross-species amplification of P. falciparum or other common human-infecting Plasmodium spp? Or human cross-amplification?
e. Whilst there appear to be a number of WGS samples from Peru that helped to inform marker selection, it wasn't clear whether any of these samples were used for validation of the MIP data?
f. The authors repeatedly refer to the cost-efficacy of the assays. How much would it cost to run one or all 4 panels on a set of e.g. 96 samples?

3. More information on the panel design and analytical evaluation would also improve the manuscript. Specifically:
a. Why were Plasmodium spp. that don't traditionally infect humans, and the ancient cadaver DNA (N = 1), included?
b. Supplementary Table 3 does not include Malaysian isolates even though other samples from the Nat Comms study are

described. Were they excluded? It might also help to add the references directly to the Table for easier cross-referencing. I assume the design happened before MalariaGEN Pv4 was released?
c. It appears that the assays were designed using MIP tools - https://github.com/bailey-lab/MIPTools. Perhaps more details on the QC from running MIP tools could be included in supplementary methods?
d. Why was a 2 x 75bp design used for the SNPs rather than 2 x 150bp? Cost?

4. As mentioned above, the great breadth of markers is very exciting, but the results fell somewhat short in providing detailed evaluation of different use cases.
a. While the authors reported on a few monoclonal infections, it wasn't clear how well the assays can detect these - were any experiments conducted to assess this directly?
b. Data on MIP regions extracted from the WGS panel of isolates would be really beneficial for people interested in using these assays to demonstrate the versatility of the panels in showing various spatiotemporal trends. For example, a global PCA would be excellent. Or localized IBD networks.
c. The authors present several IBD plots but the ability of the panel to capture IBD is not evaluated. As above, could extracted MIPs from the WGS data be used for comparison against WGS data? Or Paneljudge software?
d. Were any experiments conducted on P. vivax to validate CNV detection with these panels? It seemed a little strange that one sample explained so many copy number variants – do the authors have any ideas why this might be?

5. Data analysis considerations
a. Although we commend the authors on being moderate in their interpretation of the pvmdr1 loci as candidates (i.e. not validated) markers of P. vivax resistance to various antimalarials, it appears they have made some errors in their interpretation of the results. It appears to be a classic error of using the PvP01 reference, which has several pvmdr1 variants associated with CQ resistance – perhaps it was incorrectly assumed that the reference was wildtype? It looks like the Peruvian population actually has near-fixation of the F976Y and L1076F variants, which are deemed 'wildtype'.
b. The IBD networks could be improved in several ways. Firstly, the IBD thresholds of >99% are very strict and might occlude parts of the story. Siblings, for example, which reflect very recent relatives, would not be detected (IBD >45%). It was also not possible to see how each of the clonal clusters connected with one another. Lastly, it would also be helpful to see all infections in the plots (not just the connected ones), for assessment against the baseline.
Minor comments
1. The introduction is very Pf-focused, which is a little detracting from the Pv story. It is also somewhat contradictory as the authors are trying to emphasize the relative neglect of Pv genetic/genomic epidemiology work but have not given much light to the Pv research that has been undertaken in this field. Many lines and references refer to Pf studies when there are some good Pv examples that could instead be described.

2. The Pv panels in Table 1 are not interpreted in the full context of their design features. For example, some were not designed to look at drug resistance but rather to focus on IBD or importation. Perhaps the authors could add a column indicating what their key use cases were.

3. Please provide more information on the sequencing filtering thresholds. Specifically:
a. What was the read depth threshold (DP = X?) Only QD < 4, FS > 10, MQ < 50, MQRankSum < -2.5, or ReadPosRankSum < -5
b. Missingness threshold?

4. Why is the PCA supplemental? This might be better placed in the main text and could also be improved as follows:
a. For clarity, it would be helpful to directly state the number of samples and SNPs used in this analysis in the legend; this is true for other plots in the manuscript.
b. As described in major comment 4b, it would be great to provide a PCA panel showing how Peru fits within the global context using WGS data with extracted MIP loci.
c. The authors mention separation by city and time within the PCA. Can they also plot by year to remove temporal vs geographical trends?

5. Minor comments on the IBD plots.
a. Are these isolates restricted to monoclonal infections?
b. Figure 2 would benefit from labels (e.g. A, B, C...)

6. Figure 3 may fit better as a supplementary figure, and the PCA might make more sense in the main body of text.

(Remarks on code availability)
The bioinformatic processing pipelines were largely acceptable. Rather, the key challenges, as per the review above, were with the limited laboratory processing information.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have carefully revised the manuscript and addressed all questions raised by this reviewer.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
1. We previously commented on the need for more information on the assays for transparency and reproducibility. The authors response to rectify this was to make some edits to the Readme on their GitHub repository, which unfortunately still leaves keen readers with a large amount of work to do to find the information needed. We were hoping for a more genuine effort to rectify the shortfalls, which might consider some of the items listed below.
a) The SNPs themselves may not be of much biological interest but it would nonetheless be helpful for readers to understand the selection process for these targets in case there were need to redesign or modify the panels. The MIPtools software explanation should help but it would make things much easier for the reader to provide a brief description of the steps in the main text.
b) The GitHub link provided offers detailed explanation on the software, but it is still very difficult to find the primer sequences as they are deep within the project repository. This could prove difficult for researchers not familiar with GitHub and how to navigate within it. The GitHub provides this description: “This repository also contains sequences for the MIPs, which can be found within the vivax_project_resources directory and the respective "mip_ids" subdirectory for each probe set. The easiest way to find them without cloning the entire repository is to navigate to "probe_info.csv" in the "Go to file" box. However, if you are interested in actually using the MIPs, the more useful files can be found by searching for "probe_order.xlsx" within vivax_project_resources and the associated probe set subdirectory. Should you intend to use MIPTools with these panels, you will need to clone the full repository, as the project resources are necessary for MIPTools to run effectively.” It would surely be easier just to include the information in these files within the supplementary material of this paper.
c) Although the library preparation steps may not differ substantially from those for P. falciparum, a bit more explanation would be helpful for those readers who are not familiar with the P. falciparum work. For example, is it possible to follow all the same cycling conditions as the previously described methodology?

2. We previously requested more information on the evaluation and optimization of the laboratory assays for potential users to gauge the suitability of the assays for their own use. The authors generally made a good effort to clarify many of the queries raised and to ensure that other readers would not face the same confusion. However, there are a few minor queries that are still unclear as outlined below.
a. The description of MIP panels versus amplicon panels is appreciated and provides some useful context for assay sensitivity. However, it is still unclear about the direct comparison between UMIs and read depth. Is the minimum read depth for population studies being met (i.e. DP > 5)?
b. Is it possible for the authors to provide average fail rates of probes per run? It might be useful to know how much data is generally achieved from a run even with failures. Are specific SNPs excluded from analysis if there are a large number missing in the dataset?
c. It is an unfortunate short-fall that the authors did not do any tests of non-specific (off-target), capture. The explanation given that P. vivax is the main species in Peru is unfortunately not very helpful as this methodology is being sold as a global tool… Off-target amplification could greatly limit the usability of this assay in the many malaria-endemic areas where P. vivax co-circulates with other Plasmodium Spp. Bioinformatic processing is generally not sufficient to remove cross-species contamination, particularly in low-quality, low parasitemia, infections. In the absence of any data I would advise that the authors at least describe this transparently as a potential limitation in the discussion section. Also, is this statement in the abstract still accurate: “Our MIP panels… are poised for successful deployment in other regions of P. vivax transmission”?
d. A good and detailed cost breakdown was provided by the author, this would be useful to add to the manuscript itself.

3. The authors have addressed our major queries concerning panel design and analytical evaluation (former Q3).

4. We previously asked the authors to provide more details on a selection of use cases that were presented for the panel (former Q4). Most of our queries have been addressed. One area of potential concern that remains unresolved is the accuracy of copy number variation detection. This is not just a challenge for MIPs for many assays! We appreciate that the authors may not have access to any samples with WGS or other data that could be used for validation. In the absence of samples with WGS data, an alternative might be to assess even just a few of the detected CNVs with qPCR; if this is not feasible (we appreciate expense and DNA availability limitations), perhaps even just a statement on this potential limitation would be good for transparency in the discussion section.

5. The authors have largely addressed our queries on data analysis considerations (former Q5). Amongst these considerations, we thank the authors for double-checking and correcting the pvmdr1 976 error. After this correction, it seems that the drug resistance candidate findings from the study are not quite as exciting as originally posed. In this light, the authors might consider cropping out the reference to “mutations in antimalarial resistance orthologs” in this sentence of the abstract as it is not a major highlight of the work: “We deployed these MIP panels on 866 infections from the Peruvian Amazon and identified transmission networks with clonality (IBD>0.99), copy number variation in Pvdbp and multiple 45 Pvrbps, mutations in antimalarial resistance orthologs, and balancing selection in 13 vaccine 46 candidate genes”. However, this is a very minor suggestion that can be decided by the authors.

(Remarks on code availability)
Aside from some minor analytical queries mentioned in the report above, the GitHub details were largely sufficient for data analysis.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Popkin-Hall and colleagues have standardized four genotyping panels for the neglected human malaria parasite Plasmodium vivax, using molecular inversion probes (MIP), and applied this novel genotyping tool to 866 field-collected isolates collected in the Amazon Basin of Peru between 2011 and 2018.

The development of a new genotyping tool and its application to a large number of samples are both major accomplishments. However, in its current format, the manuscript provides neither a detailed account on the development and use of the MIP panels nor an appropriate population genomics analysis.

Methods:

1. Global variation study. The authors state in the Supplemental Material (lines 28-30) that "918 published P. vivax sequences representing all known NGS data were downloaded from the European Nucleotide Archive (accession numbers listed in Supplemental Table 2) using a custom bash script". However, in the main text they cite a recent paper (reference 33) that describes 1,895 high-quality P. vivax genome datasets. From Supplemental Table 2 it becomes clear that (a) several genome sequences included in the global variation study were not derived from reference 33 and (b) many (most likely, most) genome sequences from reference 33 were not considered in the global variation study. Why? How "global" can this analysis be?

> **Response:** We regret the confusion. The malariaGEN manuscript was released after the MIP panels had already been designed. As such, these genomes were not available for variant calling. We have modified the text accordingly. It now reads:
>
> "918 published *P. vivax* sequences representing all known NGS data available at time of design (November 2021) were downloaded from the European Nucleotide Archive (accession numbers listed in **Supplemental Table 3**) using a custom bash script."

2. "Drug resistance mutations".The authors correctly state that "there are no experimentally validated antimalarial resistance markers in P. vivax" (line 62, Supplemental Material), but have selected for inclusion in their MIP panel some "mutations associated with reduced antimalarial efficacy in field settings" in the following genes: chloroquine resistance transporter (Pvcrt), bifunctional dihydrofolate reductase-thymidylate synthase (Pvdhfr-ts), multidrug resistance protein 1 (Pvmdr1), and hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase (Pvpppk-dhps). Unfortunately, however, the reported association between mutations in the Pvcrt and Pvmdr1 genes and "chloroquine resistance" is much looser than originally thought.

Regarding Pvmdr1, a recent review (doi: 10.1016/j.ijpddr.2021.04.002) concludes that:

“Sequencing of pvmdr1 across several regions of the world has revealed more than fifty polymorphisms in this gene, as well as copy number variants (CNVs). These single nucleotide polymorphisms (SNPs) correspond to amino acid changes throughout the protein sequence, including in the ATP binding domains and multiple transmembrane regions (Sá et al., 2005). No single SNP, or set of SNPs, have emerged as definitive drug resistance markers. However, six SNPs have been reported at high frequency in multiple studies in regions with reported drug resistance; (relative to the SAL-1 reference) S513R, G698S, M908L, T958M, Y976F and F1076L (Brega et al., 2005; Huang et al., 2014; Joy et al., 2018; Orjuela-Sánchez et al., 2009; Vargas-Rodríguez et al., 2012). Amino acid changes at Y976F and F1076L in particular have been cited as possible markers of drug resistance (Brega et al., 2005; Spotin et al., 2020; Suwanarusk et al., 2008). However, these SNPs are also found in regions without reported CQR, making their association with drug resistance uncertain (Spotin et al., 2020). The T958M, Y976F, F1076L variants (and possibly others) in PvMDR1, have been shown to arise independently, even within the same population (Schousboe et al., 2015). This finding suggests that pvmdr1 mutations can arise on different genetic backgrounds and that they have not been subject to a selective sweep, possibly due to a lack of direct pressure from drug (Schousboe et al., 2015). If, and how, these SNPs play a role in mediating P. vivax drug resistance remains to be directly demonstrated in functional studies.”

Data are even less conclusive for Pvcrt (doi: 10.1016/j.ijpddr.2021.04.002):

“(…) very few SNPs (~10) in pvcrt have been reported, however, most occur at very low frequency (Barnadas et al., 2008; Joy et al., 2018; Noisang et al., 2020). The most common PvCRT polymorphism is a lysine insertion at position 10 (K10) (Barnadas et al., 2008; Joy et al., 2018; Noisang et al., 2020; Silva et al., 2018a). [Is this the “K9 duplication” mentioned in Popkin-Hall’s Table 1?] The K10 insertion has been observed in both Southeast Asian and South American parasites (Noisang et al., 2020; Silva et al., 2018b.; Zhao et al., 2020). However, no association between the K10 insertion and in vitro P. vivax drug resistance has been found (Silva et al., 2018a; Suwanarusk et al., 2007). It is currently unknown if pvcrt variants that may be associated with CQR have a deleterious impact on parasite fitness, which is the case for P. falciparum and could explain the low frequency of pvcrt mutations (Ord et al., 2007; Petersen et al., 2015).”

Mutations at the Pvmdr1 and Pvcrt loci might be worth studying even if their impact on drug resistance remains very unlikely. However, labelling them as “putative drug resistance mutations” (as in Table 1) is misleading given available evidence.

> **Response:** We thank the reviewer for this comment as it is a fair point. Putative was too ambiguous given the multiple meanings, although we were using it in the sense of

"suggested not proven " to the reviewer's point rather than "commonly accepted as true". We have modified the text accordingly to refer to these loci only as polymorphisms in antimalarial resistance orthologs of *P. falciparum* genes, including in the labels for **Tables 1 and 2**.

3. Parasite samples. In a genomic epidemiology study, the study area and population are expected to be described in detail. However, parasite samples used in the genotyping study were succinctly described in the main text as follows: "Samples were collected by NAMRU SOUTH between 2011 and 2018, primarily in the vicinity of Iquitos, the major city in the Peruvian Amazon, within Loreto Department. The sampling scheme has been described in detail elsewhere [62,63]" (lines 350-352). Only this! Importantly, references 62 and 63 do not provide additional information regarding the sampling scheme for P. vivax samples included in the current study: reference 62 describes an analysis of a convenience sample of P. falciparum isolates for HRP2/HRP3 mutations and reference 63 describes a study of P. vivax diagnosis carried out on venous blood samples from febrile patients between 2019 and 2020!

There is no information on levels of malaria transmission (e.g., API) in the study sites at the time samples were collected. Are APIs comparable among sites? How patients were selected and enrolled is not mentioned. Is this a convenience sample or there was a population-based study in these communities? Were all samples collected from symptomatic patients? Range of parasite densities? Were all infections detected by microscopy or were there submicroscopic infections? Were the authors able to fully genotype at least some submicroscopic infections?

Response: We thank this reviewer for this comment. We have increased the description of the collection of the samples to now read (page 17-18, lines 349-365):

"Samples were collected between 2011 and 2018 as part of a malaria passive surveillance study conducted in Iquitos, the main city in the Peruvian Amazon Basin, and its surrounding areas in the Loreto region. Loreto consistently contributed 90-95% of all malaria cases reported in Peru during the study period (2011-2018) with an approximate distribution of *P. vivax* and *P. falciparum* of 80% and 20%, respectively. The mean API in Loreto region for vivax malaria during the study period was 33.0 cases/1000 population with a range from 9.2 (2011) to 48.9 (2014) and malaria incidence was mainly homogeneous in all the study sites.

We enrolled individuals older than 1 year old, with fever or history of fever during the previous 72 hours in the absence of another cause of fever. Enrolled participants provided a written and/or assent form. We collected clinical, demographic and epidemiological information and up to 8 milliliters of venous whole blood. Two thin and thick smears were prepared and stained

with 10% Giemsa for each participant. Slides were read independently by two dedicated microscopists; discordant results were solved by a third microscopist. All participants were microscopy positive to *Plasmodium vivax* with a parasitemia range of 12 to 141,373 par/ul during the study period."

4. DNA extraction. The authors succinctly state that "DBS [not spelled out] were extracted using the Chelex method, as previously described [32,64] (line 354)." If the newly described genotyping method is to be widely used, please indicate the starting blood volume (and minimum parasitemia) for extraction, the concentration of genomic DNA after extraction etc.

Response: We have clarified that "DBS" refers to dried blood spots. The Chelex extraction method is well known and widely used, with specifics given in the listed references. In addition, the protocols used for this study do not differ substantially from those used for *P. falciparum*, merely the target organism. We have added additional detail on the extraction and MIP process to the manuscript, which now reads (page 18, lines 367-369):

"Briefly, DNA was extracted from three 6-mm punches, which represent approximately 25 µL of blood. Extracted DNA was used for molecular inversion probe (MIP) captures, which were sequenced as previously described for *P. falciparum*[20,44]. Briefly, all MIPs in a given panel were pooled, then samples were captured in individual reactions, followed by exonuclease treatment, amplification, and finally sequencing. "

5. Sample processing. Were samples tested with quantitative PCR prior to sequencing? Can the authors correlate cycle threshold (Ct) values with MIP coverage to make recommendations for readers? For instance, according to Aydemir et al. 2018 (https://doi.org/10.1093/infdis/jiy223): "MIP coverage began to drop off at qPCR cycle threshold (Ct) values of 31, which equates to a parasite level of approximately 100 parasites/uL" (…) "The median Ct value of the samples that yielded no MIP coverage was 34".

Response: Parasitemia was determined for this study via microscopy. We appreciate the comment on providing a guideline for readers, but would note that C_T values depend on the target and cycling conditions, so parasitemia, which can be determined via either microscopy or qPCR with a standard curve is more broadly applicable as well as less variable.

5. Genotyping. What is the minimum parasite density required for reliable genotyping results? In the Discussion (lines 217-218), the authors state that their method works "reliably on P. vivax

isolates with parasitemias ≥100 parasites/μL" They mention in the Supplemental Material that "(…) sample missingness was very low (<12%) for PvG even at low parasitemia (≤100 parasites/μL)", but no data are provided to support this statement. Moreover, the authors do not explain how they dealt with missing data in PCA, IBD and other analyses.

> **Response**: These data are presented in **Supplemental Figure 2**. The sample missingness scale for the inset plot of PvG (0 - 0.12) is appreciably lower than for the SNP panels, which range from 0 - 0.4 and 0 - 0.5. There is no parasitemia value at which MIP capture categorically does not work, but 100 p/μL or higher had consistent captures.
>
> Missingness was filtered both on a per-sample and a per-variant basis, as written in the Methods (page 18, lines 393-396):
>
> > "Samples with >50% missingness on any panel were excluded from the VCFs, which were then sorted and concatenated with *bcftools*. Finally, *bcftools* was used to subset the resulting VCF to include only biallelic SNPs and exclude SNPs with >50% missingness across remaining samples."
>
> These filters were applied prior to the other analyses.

6. Filtering. Sequence filtering was based on allele frequency and UMI counts. Were any other quality metrics used in filtering? How many samples were covered with anUMI depth of at least 3? How polyclonal samples were handled for IBD analysis? Does the MIPanalyzer accommodate polyclonal samples? Please indicate the functions used for IBD analysis, as they are not listed in https://mrc-ide.github.io/MIPanalyzer/reference/index.html.

> **Response**: The quality metrics used in filtering are indicated in the "MIP Variant Calling and Filtering" section of the methods, in addition to the quality filtering included in the MIPTools pipeline. Sample numbers that passed each filtering step are indicated in the results, e.g.:
>
> > "Of **697** isolates (all samples that passed quality filters, regardless of metadata) analyzed with THEREALMcCOIL".
>
> These numbers are also indicated in **Supplemental Figure 3**. We have also added additional details to this section on the number of samples that passed MIP filtering cutoffs upstream of the other analyses. It now reads:
>
> > "For the SNP panels, we specified a minimum UMI depth of 3, a within-sample allele frequency threshold of 0.5, and a minimum alternate read count of 2 to obtain 616 variant SNPs across 760 samples (PvFST), 437 variant

SNPs across 763 samples (PvMAF5), and 377 variant SNPs across 750 samples (PvMAF40).“

All analysis steps and associated functions are demarcated here: https://github.com/IDEELResearch/PvMIPs/blob/main/Analysis/Pv_MIP_analyses.Rmd

IBD analysis begins on line 741. As indicated at line 1,021, the data frame was filtered to include only monoclonal samples to produce the IBD networks seen in the manuscript. We have also clarified this point in the Methods.

Results:

1. Clonal outbreak in Belén. From a genomic epidemiology perspective, this is the most interesting finding, but it remains largely unexplored. The reader is not provided with basic information to interpret the finding of a clonal outbreak spanning several months in one of the study sites. How were the 269 samples collected in Belén? How many (sampled and unsampled) infections were diagnosed in the community during the study period? Were all sample donors symptomatic? Can we take the 269 samples as representative of P. vivax lineages causing clinical malaria in the site? How connected is Belén with other study sites? What is the maximum percent IBD detected in pairs of samples from different communities/sites?

Response: Additional details about sample collection has been provided above in response to query 1. We do not claim to have captured all infections during this period within each site or across sites. However, the near identical nature of these isolates support that they are clonal in nature and thus strongly supports the transmission of this clone across years. We have replaced all instances of the word “outbreak” with “clonal transmission” to clarify what we actually detected.

The IBD distribution when comparing Belén to other sites closely resembles the overall IBD distribution, with most sample pairs not being closely related but a notable minority having very high IBD > 0.9. This is also true when comparing isolates within Belén, though the proportion of samples with very high IBD is larger. This difference is highly significant, with a t-test comparing the two IBD distributions finding a p-value <0.0001. The maximum IBD detected in sample pairs from different communities is 0.99.

2. Monoclonal infections. The finding of a vast majority (91.4%) of monoclonal P. vivax infections in Loreto is rather surprising, given the relatively high levels of genetic diversity of malaria parasites and the relatively intense transmission during the study period (2011-

2018). Several studies using a range of genotyping methods (e.g., microsatellites, AmpliSeq etc.) have shown much higher proportions of multiple-clone infections in P. vivax populations from sites around Iquitos (see, e.g., https://doi.org/10.1038/s41598-024-66925-x). Any explanation for this unexpected finding? Any change MIP genotyping failed to detect a large proportion of minor clones?

> **Response:** COI estimates vary widely between studies and collection years, including one study that found a higher monoclonal ratio than our study (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6032790/) and multiple others that find similar COI ratios (https://www.nature.com/articles/s41598-022-21028-3, https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-8) in addition to some that find higher rates of polyclonality as the reviewer points out. In this context, we do not find our high rate of monoclonality to be unexpected. In addition, COI can be estimated using a variety of computational methods. These methods are susceptible to various biases, but on the whole tend to overestimate rather than underestimate COI. While this is the first study where we have estimated COI using MIP data from *P. vivax*, we have done so numerous times in *P. falciparum* and recovered many polyclonal infections. For these reasons, we think it is unlikely that the high rate of monoclonality in our study reflects anything other than the biology of the samples included in this study.

3. Copy number variation. For reproducibility, please indicate cut-offs values for UMI depth prior to CNV calculation.

Multicopy genes have a faster rate of evolution than single-copy genes, which reflects in an increased nucleotide diversity (π) (e.g.: https://doi.org/10.1186/1471-2164-10-353). The comparison of π between duplicated genes and single-copy genes can give greater strength to these findings.

> **Response:** These cutoffs are indicated in the Methods, i.e. “Prior to CNV calculation, samples and MIPs were filtered for missingness as follows: samples with <5 unique molecular identifier (UMI) counts for more than 50% of MIPs were excluded, as were MIPs with <5 UMI counts for more than 50% of samples sequenced.”
>
> We appreciate the comment on looking at π in these genes. The gene duplications are a relatively minor finding in this study, but we will bear this in mind going forward as we capture more *P. vivax* samples.

Reviewer #1 (Remarks on code availability):

As noted above, the functions used for IBD analysis are not listed in https://mrc-ide.github.io/MIPanalyzer/reference/index.html. Otherwise, information appears to be complete.

> **Response:** We did not write the aforementioned code. As indicated in the code availability statement, all of our code is available from https://github.com/IDEELResearch/PvMIPs/ and the IBD analysis is detailed starting on line 741 of https://github.com/IDEELResearch/PvMIPs/blob/main/Analysis/Pv_MIP_analyses.Rmd

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3 (Remarks to the Author):

Overview

The study by Popkin-Hall and colleagues presents the first set of MIP panels for P. vivax molecular surveillance. This is an important advance, and the panels have potential to be informative for both malaria control program staff and researchers. However, the major flaw of the manuscript as it currently stands is that there is far too little information on the laboratory methods. Critical details such as the probe sequences are missing, rendering it impossible for potential users to implement the assays. Given that the assays are the main novelty of the study, as the biological and epidemiological insights from the Peruvian data are not likely to be of huge interest to others than those with very closely related interests in this geographic region, it is essential for the authors to provide more information on these methods to allow the assays to be reproduced in other laboratories.
In addition to the lack of detail on the laboratory methods, there was very little information on the evaluation and optimization of the assays. Critical details that one would want to know about before committing to setting up the assays were missing or difficult to derive, such as assay specificity, ability to detect polyclonal infections, and expected read depth for more than just the PvG panel.
In principle, the marker selection across the four panels sounds impressive and we were excited to see its application, but the results were unfortunately very sparse and didn't quite deliver.
Specific suggestions on areas for improvement are provided below.

Major comments

1. Critical information to allow reproducibility of the assays needs to be provided, including:

a. List of targets in all of the panels (not just the PvG panel)

b. Primer/probe sequences

c. Methods for the library preparation steps

> **Response:** The targets within the SNP panels are largely not of any biological interest and these panels were not designed for any purpose other than to detect population structure, geographic origin, and complexity of infection. Nonetheless, their genomic locations and all of the sequences in question are available on the GitHub repository linked in the "code availability" section. However, thanks to the reviewer's comment, we realized they were very difficult to find. As such, we have specified that they can be found in the GitHub repository and have edited the Readme on the GitHub repository to point to their locations, which can be accessed here: https://github.com/IDEELResearch/PvMIPs/blob/main/README.md.
>
> With respect to the library preparation steps, the protocols used for this study do not differ substantially from those used for *P. falciparum*, merely the target organism. Details on MIP capture for *P. falciparum* are already published. As per reviewer 1, we have expanded these methods in the paper.

2. Given that the authors have gotten the assays to work on dried blood spot samples (not an easy task!), I think there will be many keen users. However, more information on the evaluation and optimization of the laboratory assays would be very helpful for potential users to gauge the suitability of the assays for their own use. Specifically:

a. As MIPs are quite new to the malaria field (this is indeed why these assays are fairly novel), it might help to give more description on how UMIs correlate with e.g. amplicon sequencing read depth and coverage stats. This is key to really appreciating the amount of data one can expect from these assays.

b. Also, regarding data quantity, it wasn't clear why there were only heatmaps for the PvG panel – what about the performance in other panels?

c. More information within the legend of the heatmaps would help interpretation. What are the grey bands – zero coverage or depth?

d. There was no description on the specificity of the assays. Is there any cross-species amplification of P. falciparum or other common human-infecting Plasmodium spp? Or human cross-amplification?

e. Whilst there appear to be a number of WGS samples from Peru that helped to inform marker selection, it wasn't clear whether any of these samples were used for validation of the MIP data?

f. The authors repeatedly refer to the cost-efficacy of the assays. How much would it cost to run one or all 4 panels on a set of e.g. 96 samples?

Response:

a. MIPs are much more easily multiplexed. Amplicon panels as well as MIPs can be deeply sequenced with thousands of reads but MIPs allow for oversequencing of the same molecule to be corrected for thus UMI counts tend to be much lower than amplicon panel read depth. MIPs do tend to be less sensitive given capture occurs initially whereas amplification of the template can occur in any of the initial rounds of PCR cycling. A direct comparison of MIPs to amplicons is difficult due to sample limitations and the lack of functional amplicon panels in our labs. We have added the following sentence to the discussion to address this question:

> "MIPs are more easily multiplexed than amplicon panels and incorporate unique molecular identifiers that enable correction of amplification bias, but are typically less sensitive than amplicons."

b. Heatmaps were initially only provided for PvG because PvG targets full genes that are of known biological interest. The other panels cover SNPs interspersed across the entire genome, which were not chosen based on their biological functions, but strictly for population genomics analyses. The performance of any individual probe is not a critical factor in data interpretation. Indeed, redundancy is built into these panels so that individual probes can fail to amplify without affecting data analysis and interpretation. With these caveats, we have provided heatmaps of performance in the other MIP panels as additional supplemental figures.

c. Grey bands represent zero coverage. We have updated the legend to make this explicit. It now reads:

> "Heatmaps showing MIP sequencing performance for all tiled genes included in PvG panel. Sequencing performance is shown as log(UMI count) for each MIP. Grey bands represent probes with zero coverage."

d. We did not test this. *P. vivax* is the major malaria parasite in Peru, but co-infections with *P. falciparum* are not uncommon. In general, there is non-specific and non-targeted capture that is filtered out routinely during bioinformatic processing and such off-target captures do not affect the results or analysis.

e. MIPs have been used extensively in *P. falciparum* in various geographic settings without reproducibility problems. In addition the UMIs protect from amplification errors. While validating performance on known controls would be preferred, we were unable to obtain any control DNA from MR4 or other sources for this unculturable parasite made this impossible.

f. Thank you for the question. We estimate that MIP capture costs $3.02 per sample, with sequencing costing $3.91 per sample per panel (i.e. a total of $10.84 for each sample including capture and sequencing for both the SNP panels and the tiled genes). Thus, capturing and sequencing 96 samples would cost $1,040.64. It is important to note that most of the cost is upfront, with one order being sufficient for 21,298 captures (in the case of the tiled gene panel) or 57,521 captures (in the case of the SNP panels). The total upfront cost for one order of plates, including reagents and plastics, is therefore about $27,000. Even following these captures, there will be leftover probes for the lower pooled volume probes, as only those of the highest pooling volume will have been completely used up.

3. More information on the panel design and analytical evaluation would also improve the manuscript. Specifically:
a. Why were Plasmodium spp. that don't traditionally infect humans, and the ancient cadaver DNA (N = 1), included?
b. Supplementary Table 3 does not include Malaysian isolates even though other samples from the Nat Comms study are described. Were they excluded? It might also help to add the references directly to the Table for easier cross-referencing. I assume the design happened before MalariaGEN Pv4 was released?
c. It appears that the assays were designed using MIP tools - https://github.com/bailey-lab/MIPTools. Perhaps more details on the QC from running MIP tools could be included in supplementary methods?
d. Why was a 2 x 75bp design used for the SNPs rather than 2 x 150bp? Cost?

Response:
a. These are a relic from a previous global study of *P. vivax* variation and are not expected to have meaningfully changed any results or interpretations.

b. Thank you. This is an oversight. 52 Malaysian samples were included and the table has been updated accordingly. The reviewer is correct that the design happened before the MalariaGEN Pv4 release (see response to reviewer #1).

c. We have added a section on panel optimization to the supplementary methods.

d. Shorter MIPs tend to perform better, with higher capture efficiency. In addition, 2 X 75bp sequencing is cheaper. Thus, this approach is valuable for SNP detection. We used longer sequencing with larger captures for genes we wanted more dense sequencing of through tiling. Here we are trading capture efficiency and sequencing costs to have fewer probes needed to span the gene.

4. As mentioned above, the great breadth of markers is very exciting, but the results fell somewhat short in providing detailed evaluation of different use cases.

a. While the authors reported on a few monoclonal infections, it wasn't clear how well the assays can detect these - were any experiments conducted to assess this directly?
b. Data on MIP regions extracted from the WGS panel of isolates would be really beneficial for people interested in using these assays to demonstrate the versatility of the panels in showing various spatiotemporal trends. For example, a global PCA would be excellent. Or localized IBD networks.
c. The authors present several IBD plots but the ability of the panel to capture IBD is not evaluated. As above, could extracted MIPs from the WGS data be used for comparison against WGS data? Or Paneljudge software?
d. Were any experiments conducted on P. vivax to validate CNV detection with these panels? It seemed a little strange that one sample explained so many copy number variants – do the authors have any ideas why this might be?

Response:

a. THE REAL McCOIL has been used in numerous previous studies of *P. falciparum* to estimate COI from MIP data. As such, and given that there is no universally agreed upon tool for COI estimation, we saw no need to conduct such an experiment. The WSAF and missingness filtering steps we performed serve to reduce the potential for the inflation of COI estimates, as erroneous polyclonal calls are much more likely than erroneous monoclonal calls. Mock samples can not be made as there is no renewable resource for parasite DNA as this organism is not culturable, as we have for Pf.

b. Thank you for the suggestion. We have included an additional supplemental PCA (**Supplemental Figure 11**) using the MIP regions extracted from the global VCF we used for panel design which replicates the WGS studies showing clear continental divisions. Although there is less structure when comparing Peruvian samples to other South American samples, this is consistent with other studies that have shown Peruvian and Brazilian *P. vivax* to cluster together due to substantial gene flow within Amazonia.

c. The use of MIP panels to estimate IBD in *Plasmodium* populations is well established and there is no reason to think that the method would fare differently in *P. vivax* than in *P. falciparum*. Here again, we do not have samples where both types of data are available generated in our lab.

We know that different laboratory methods will have different ability to detect IBD, with low density genotyping typically providing lower resolution than high density genotyping (like WGS). MIPs provide an intermediate step between multiplex amplicon and WGS for the number of SNPs. Simulation data for Pf (see https://doi.org/10.1038/s41467-020-15779-8 Supplement) suggests that this density of SNPs is more than sufficient and work by others suggest that our panel has more

than enough density for robust IBD calls in P vivax (see https://academic.oup.com/genetics/article/212/4/1337/5931242). We have added information on this to the second paragraph of the discussion:

> "The 1,200 SNPs genotyped across the three panels are more than sufficient to generate robust IBD estimates[34,35]."

d. The CNV detection method has been empirically validated in *P. falciparum*, as described in the methods: "The approach was previously validated in laboratory strains of *P. falciparum*, including 3D7, 7G8 and Dd2, in independent runs. Unlike 3D7 and 7G8, Dd2, the positive control with multiple copies of *mdr1*[71,72] consistently showed CNV > 2 for this gene."

P. vivax has no such strains for comparison, so there is no equivalent cross-check possible. There are examples from the literature of other field isolates with multiple duplications, so we did not find this result surprising. The reticulocyte binding proteins are also known to be highly diverse and under selective pressure, so there is presumably an advantage for this sample to have so many duplications.

5. Data analysis considerations

a. Although we commend the authors on being moderate in their interpretation of the pvmdr1 loci as candidates (i.e. not validated) markers of P. vivax resistance to various antimalarials, it appears they have made some errors in their interpretation of the results. It appears to be a classic error of using the PvP01 reference, which has several pvmdr1 variants associated with CQ resistance – perhaps it was incorrectly assumed that the reference was wildtype? It looks like the Peruvian population actually has near-fixation of the F976Y and L1076F variants, which are deemed 'wildtype'.

b. The IBD networks could be improved in several ways. Firstly, the IBD thresholds of >99% are very strict and might occlude parts of the story. Siblings, for example, which reflect very recent relatives, would not be detected (IBD >45%). It was also not possible to see how each of the clonal clusters connected with one another. Lastly, it would also be helpful to see all infections in the plots (not just the connected ones), for assessment against the baseline.

Response:

a. Thank you for catching this mistake. We have corrected it in the text and **Table 2**.

b. The very strict threshold of >99% was chosen so that the networks were interpretable and also to highlight the extremely high IBD in these communities. The aim of this analysis was to detect clonal transmission, not

siblings. With respect to seeing all infections on the plots, this is visually intractable. The overall IBD distribution however is visible in **Supplemental Figure 5**, where it is clear that while the clonal clusters are in the overall minority, they are by far the most abundant within isolate pairs with >50% IBD. We have added additional supplemental images at 0.25, 0.5, and 0.9 IBD).

Minor comments

1. The introduction is very Pf-focused, which is a little detracting from the Pv story. It is also somewhat contradictory as the authors are trying to emphasize the relative neglect of Pv genetic/genomic epidemiology work but have not given much light to the Pv research that has been undertaken in this field. Many lines and references refer to Pf studies when there are some good Pv examples that could instead be described.

Response: Thank you for pointing this out. The introduction focuses on Pf because the majority of genomic epidemiology studies and the entirety of MIP studies until now have been in Pf. However, we have added additional mentions and citations of the Pv genomic epidemiology studies as suggested.

2. The Pv panels in Table 1 are not interpreted in the full context of their design features. For example, some were not designed to look at drug resistance but rather to focus on IBD or importation. Perhaps the authors could add a column indicating what their key use cases were.

Response: We have added the additional column.

3. Please provide more information on the sequencing filtering thresholds. Specifically:
a. What was the read depth threshold (DP = X?) Only QD < 4, FS > 10, MQ < 50, MQRankSum < -2.5, or ReadPosRankSum < -5
b. Missingness threshold?

Response:

a. Those filtering thresholds refer to the genomic variation study used to design the MIP panels. These are the metrics recommended in the GATK Best Practices pipeline (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants). The specific cutoffs were determined by examining the distribution of these metrics across all variants, as is described in the supplemental methods:

> “Variant hard filtering was performed using GATK to exclude variants with a QD < 4, FS > 10, MQ < 50, MQRankSum < -2.5, or

ReadPosRankSum < -5. These thresholds were identified by using the VariantsToTable tool within GATK to export metrics for all variants, the distributions of which were then plotted in R. “

b. Missingness was primarily a concern for COI and IBD. The threshold is noted in the Methods:

> “Samples with >50% missingness on any panel were excluded from the VCFs, which were then sorted and concatenated with *bcftools*. Finally, *bcftools* was used to subset the resulting VCF to include only biallelic SNPs and exclude SNPs with >50% missingness.”

With respect to the global variation study for MIP design, candidate SNPs were selected by calculating F_{ST} values and minor allele frequency across the entire dataset, so missingness was not a major concern.

4. Why is the PCA supplemental? This might be better placed in the main text and could also be improved as follows:
a. For clarity, it would be helpful to directly state the number of samples and SNPs used in this analysis in the legend; this is true for other plots in the manuscript.
b. As described in major comment 4b, it would be great to provide a PCA panel showing how Peru fits within the global context using WGS data with extracted MIP loci.
c. The authors mention separation by city and time within the PCA. Can they also plot by year to remove temporal vs geographical trends?

Response: The PCA is supplemental because it is difficult to interpret and not terribly informative due to the large number of sites and collection years incorporated within. The IBD figure is much clearer and more informative, but as noted above does not include all samples.

a. This information is available in the text and in **Supplemental Figure 3**. Nonetheless, we have modified the figure legends accordingly to minimize the need to reference back and forth.

b. Thank you for the suggestion. We have added this plot (**Supplemental Figure 11**).

c. This panel already exists (**Supplemental Figure 4c**). We have modified the main text to make this clearer:

> “The PCA revealed no broad geographical or temporal structure, but PC1 and PC2 included multiple clusters that were associated by both city and

year (plotted by city, collection year, and both in **Supplemental Figure 4**)."

5. Minor comments on the IBD plots.
a. Are these isolates restricted to monoclonal infections?
b. Figure 2 would benefit from labels (e.g. A, B, C...)

Response:

a. Yes, the IBD networks are restricted to monoclonal infections. All analysis steps and associated functions are demarcated here: https://github.com/IDEELResearch/PvMIPs/blob/main/Analysis/Pv_MIP_analyses.Rmd

   IBD analysis begins on line 741. As indicated at line 1,021, the data frame was filtered to include only monoclonal samples to produce the IBD networks seen in the manuscript.

   We have modified the figure legend to make this clearer.

b. Thank you. We have added the additional labels.

6. Figure 3 may fit better as a supplementary figure, and the PCA might make more sense in the main body of text.

Response: We appreciate the suggestion but we disagree, as discussed above.

Reviewer #3 (Remarks on code availability):

The bioinformatic processing pipelines were largely acceptable. Rather, the key challenges, as per the review above, were with the limited laboratory processing information.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have carefully revised the manuscript and addressed all questions raised by this reviewer.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3 (Remarks to the Author):

1. We previously commented on the need for more information on the assays for transparency and reproducibility. The authors response to rectify this was to make some edits to the Readme on their GitHub repository, which unfortunately still leaves keen readers with a large amount of work to do to find the information needed. We were hoping for a more genuine effort to rectify the shortfalls, which might consider some of the items listed below.
a) The SNPs themselves may not be of much biological interest but it would nonetheless be helpful for readers to understand the selection process for these targets in case there were need to redesign or modify the panels. The MIPtools software explanation should help but it would make things much easier for the reader to provide a brief description of the steps in the main text.
b) The GitHub link provided offers detailed explanation on the software, but it is still very difficult to find the primer sequences as they are deep within the project repository. This could prove difficult for researchers not familiar with GitHub and how to navigate within it. The GitHub provides this description: “This repository also contains sequences for the MIPs, which can be found within the vivax_project_resources directory and the respective "mip_ids" subdirectory for each probe set. The easiest way to find them without cloning the entire repository is to navigate to "probe_info.csv" in the "Go to file" box. However, if you are interested in actually using the MIPs, the more useful files can be found by searching for "probe_order.xlsx" within vivax_project_resources and the associated probe set subdirectory. Should you intend to use MIPTools with these panels, you will need to clone the full repository, as the project resources are necessary for MIPTools to run effectively.” It would surely be easier just to include the information in these files within the supplementary material of this paper.
c) Although the library preparation steps may not differ substantially from those for P. falciparum, a bit more explanation would be helpful for those readers who are not familiar with the P. falciparum work. For example, is it possible to follow all the same cycling conditions as the previously described methodology?

Response: Thank you for the comment and for your attention to detail. Responses to each of the points are below.

a. This information is already included in the text, in general terms in the main text (**lines 373-379**), and in detail in the supplement. Please see the below paragraph, copied from page 2 of the supplement:

   "Vcftools was used to calculate pairwise F_{ST} for all SNPs between all endemic regions (Southeast Asia/Papua New Guinea, South Asia, Africa, South America, and Central America) and the 400 SNPs with the highest F_{ST} (i.e. the best private alleles for each geographic region) were selected for MIP design along with previously identified SNPs of interest[5,27]. In addition to these differentiating SNPs, two sets of neutral (non-coding) SNPs were also generated. First, bedtools v2.30.0[28] was used to export only non-coding SNPs from the original filtered VCF. Then, bcftools was used to generate one set of rare neutral SNPs (MAF < 0.05) and one set of common neutral SNPs (MAF > 0.4). These SNP sets were then further filtered to minimize the number of missing genotypes for included sequences: all included variants had a GQ > 90 and were genotyped in > 67% of individuals (MAF < 0.05) or > 55% of individuals (MAF > 0.4)."

b. Thank you for the suggestion. The information in the GitHub repository is too extensive to be able to show in a simplified way in terms of numbers of files, etc. To try to simplify this, we are providing an additional Excel table with the primers listed. This ties to the next comment (c) where additional information on methods are provided.

c. We have added additional details to the methods, as excerpted below:
   "Sample Processing and Sequencing

   DBS Dried blood spots were extracted using the Chelex method, as previously described[44,47]. Briefly, DNA was extracted from three 6-mm punches, which represent approximately 25 µL of blood. Extracted DNA was used for molecular inversion probe (MIP) captures, which were sequenced as previously described for *P. falciparum*[20,44]. The human MIP protocol was used instead of the parasite MIP protocol due to the GC richness of the *P. vivax* genome. When comparing the parasite and human MIP protocols, the differing buffer and enzyme in the respective capture and PCR steps, in addition to the increased temperatures in the PCR cycling conditions allow for more efficient amplification. Due to the difficulty in capturing the *P. vivax* genome from human samples, the amount of DNA input into the capture reaction was increased from 3 uL to 5 uL and the volume of water in the capture master mix was decreased, in exchange. Four new MIP panels were designed for *P. vivax* based on the PvP01 genome assembly: 1) PvG, a tiled gene panel designed to densely genotype genes of interest, including putative drug antimalarial resistance ortholog markers and invasion ligands that are potential vaccine targets; 2) PvFST, a panel designed to capture geographically discriminating SNPs between and within continents; 3)

PvMAF5, a panel designed to capture rare neutral (non-coding) SNPs (minor allele frequency <0.05); and 4) PvMAF40, a panel designed to capture common neutral SNPs (MAF > 0.4). The details of the MIP design scheme are given in the **Supplemental Material**. The generated libraries were sequenced on an Illumina Nextseq 550 platform using 150 bp paired end sequencing with dual indexing using Nextseq 500/550 Mid-output Kit v2 in the case of the gene panels and 75 bp paired end sequencing in the case of the SNP panels."

2. We previously requested more information on the evaluation and optimization of the laboratory assays for potential users to gauge the suitability of the assays for their own use. The authors generally made a good effort to clarify many of the queries raised and to ensure that other readers would not face the same confusion. However, there are a few minor queries that are still unclear as outlined below.

a. The description of MIP panels versus amplicon panels is appreciated and provides some useful context for assay sensitivity. However, it is still unclear about the direct comparison between UMIs and read depth. Is the minimum read depth for population studies being met (i.e. DP > 5)?

b. Is it possible for the authors to provide average fail rates of probes per run? It might be useful to know how much data is generally achieved from a run even with failures. Are specific SNPs excluded from analysis if there are a large number missing in the dataset?

c. It is an unfortunate short-fall that the authors did not do any tests of non-specific (off-target), capture. The explanation given that P. vivax is the main species in Peru is unfortunately not very helpful as this methodology is being sold as a global tool… Off-target amplification could greatly limit the usability of this assay in the many malaria-endemic areas where P. vivax co-circulates with other Plasmodium Spp. Bioinformatic processing is generally not sufficient to remove cross-species contamination, particularly in low-quality, low parasitemia, infections. In the absence of any data I would advise that the authors at least describe this transparently as a potential limitation in the discussion section. Also, is this statement in the abstract still accurate: "Our MIP panels… are poised for successful deployment in other regions of P. vivax transmission"?

d. A good and detailed cost breakdown was provided by the author, this would be useful to add to the manuscript itself.

Response: We have addressed these minor queries below.

a. We specified a minimum UMI depth of 3 based on previous MIP analysis which often incorporates a read depth of more than DP 5. The advantage and the ability to have a lower threshold with MIPs compared to ampicon reads is the error correction due to the incorporation of UMIs into the template capture, whereas sequencing of PCR amplicons allows for sequencing multiple, even thousands of reads that represent a single initial template molecule.

b. As we noted in the supplementary methods, we performed a probe rebalance to ensure even capture and sequencing to the extent possible. As we have noted

previously, the rationale for choosing 400 SNPs per panel is to reduce the impact of any failures. Since the panels were designed based on a global data analysis, it is expected that some SNPs may not be present in a given population. Because this is likely to vary by parasitemia and geography, an average failure rate is more likely to be misleading than useful. We have provided supplementary heat maps (**Supplemental Figure 1 and 2**) indicating capturing and sequencing success in our samples, which provide a general baseline for the inquisitive reader. These heat maps clearly show that failure rates are low in well-performing samples, but frequently occur in the same SNPs. This pattern is expected based on previous experience with MIP panels and is the reason for the high redundancy built into these panels.

c. The utility of a *P. vivax* genotyping tool is greatest in regions where *P. vivax* is highly prevalent, e.g. Latin America, Southeast Asia, the Horn of Africa. We do not propose using our panels to characterize known co-infections with low *P. vivax* parasitemia. Our panel was not designed to speciate malaria infections, so there is no reason to think that any investigators would attempt to use it to capture unspeciated *Plasmodium* infections in regions with high *P. falciparum* prevalence. It is furthermore unlikely that the panel would be used on isolates that had not undergone either microscopy or qPCR for speciation. As such, the likely impact of off-target amplification is minor. In addition, MIPs are substantially more precise than normal primers, which further reduces the likelihood of its occurrence. Finally, during individual variant calling, any off-target reads are removed, as they will not map properly to the vivax genome. Off-target sequences usually are extremely short and are removed in the MIPWrangler processing for haplotypes as aberrant lengths. Additionally, off-target MIPs will not map to the vivax genome to the regions we analyze--since they are off-target.

d. We have added this cost breakdown to the supplement, following the section on panel optimization:

   "*Cost Estimate:*

   We estimate that genotyping one sample with our panel costs $10.84, which includes $3.02 for MIP capture and $7.82 for sequencing ($3.91 each for the tiled genes and the SNPs). Thus, capturing and sequencing 96 samples would cost $1,040.64. However, this per-sample cost presumes an upfront cost of about $27,000, which is sufficient for 21,298 captures of PvG and 57,521 SNP captures. Even following these captures, there will be leftover probes for the lower pooled volume probes, as only those of the highest pooling volume will have been completely used up. As such, while ordering the reagents is a substantial initial investment, the panel is very cost-effective for a large number of samples."

3. The authors have addressed our major queries concerning panel design and analytical evaluation (former Q3).

4. We previously asked the authors to provide more details on a selection of use cases that were presented for the panel (former Q4). Most of our queries have been addressed. One area of potential concern that remains unresolved is the accuracy of copy number variation detection. This is not just a challenge for MIPs for many assays! We appreciate that the authors may not have access to any samples with WGS or other data that could be used for validation. In the absence of samples with WGS data, an alternative might be to assess even just a few of the detected CNVs with qPCR; if this is not feasible (we appreciate expense and DNA availability limitations), perhaps even just a statement on this potential limitation would be good for transparency in the discussion section.

Response: As noted in the manuscript, our CNV method was validated using the Dd2 lab strain, which is known to have duplication of the *mdr1* gene (positive control), and 3D7 and 7G8 carrying a single copy of this gene as negative controls. These lab strains were tested many times in different captures and sequencing runs to ensure that our CNV calculator algorithm was reproducible. The results we got were concordant with the values previously reported for these strains with qPCR methods (PMC1363351 , PMID: 33639924) . We have published very precise estimates with falciparum MIPs for MDR1 and plasmepsin 2 and 3 (see https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(21)00085-9/fulltext)

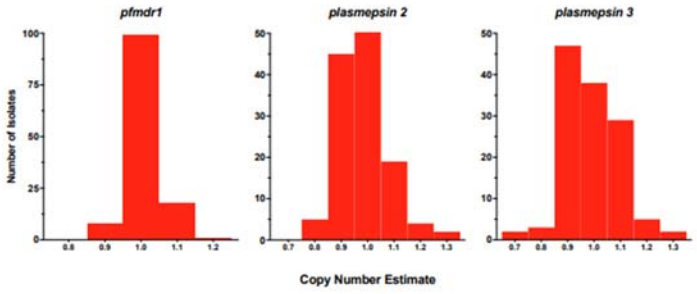

Figure S4. Distribution of gene copy number estimates based on MIP genotyping for the *pfmdr1* and *plasmepsin 2* and *3* genes. Calls are based on means from 31 unique probes for *pfmdr1* and 10 each for *plasmepsin 2* and *plasmepsin 3* (see appendix 2).

Essentially, MIPs performance is superior to qPCR given the ability to use tens, even hundreds of MIPs to target CNVs as well as the ability to control within sample capture based on non copy number variant MIPs. Together this provides the precision and accuracy needed for large scale studies with minimal false positives or negatives.

We have added a sentence with a reference to this previous work at the end of the methods:
"We have also previously used MIP data to generate precise copy number estimates[82]."

5. The authors have largely addressed our queries on data analysis considerations (former Q5). Amongst these considerations, we thank the authors for double-checking and correcting the pvmdr1 976 error. After this correction, it seems that the drug resistance candidate findings from the study are not quite as exciting as originally posed. In this light, the authors might consider cropping out the reference to "mutations in antimalarial resistance orthologs" in this sentence of the abstract as it is not a major highlight of the work: "We deployed these MIP panels on 866 infections from the Peruvian Amazon and identified transmission networks with clonality (IBD>0.99), copy number variation in Pvdbp and multiple 45 Pvrbps, mutations in antimalarial resistance orthologs, and balancing selection in 13 vaccine 46 candidate genes". However, this is a very minor suggestion that can be decided by the authors.

Response: Thank you for the suggestion. We elect to keep the sentence as is, given that it highlights the panel's ability to detect these mutations, even if we did not find anything of particular interest in our Peruvian samples.

Reviewer #3 (Remarks on code availability):

Aside from some minor analytical queries mentioned in the report above, the GitHub details were largely sufficient for data analysis.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.