nature portfolio

Peer Review File

Single-cell analyses of polyclonal Plasmodium vivax infections and their consequences on parasite transmission

Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript provides interesting insights into the establishment of polyclonal infections using an animal model coupled with short read single cell RNA sequencing to profile the parasite transcriptome and genotypes. First, they explore monoinfections of two different strains, finding that among the cells where they have power (e.g. confident strain assignment, and good number of cells) strains show some DE but this is largely in PIRs or in genes that they observe to be deleted in one of the strains. Then they carried out sequential infections with a two week gap but these failed to establish polyclonal infections and did not show compelling DE compared to monoclonal infections. Then they explored inoculations with two strains simultaneously – again they did not find much DE here between strains. Then they fed mosquitoes on monkeys carrying two strains and explored the sporozoites by scRNAseq as well as inoculating two more monkeys with these sporozoites. The sporozoite data showed all three expected types (pure strains and crosses) but the monkey infections did not have pure Chesson strain present. They compared expression between NIH parasites in an infection resulting from blood stage inoculation vs sporozoite inoculation, finding an intriguing rhoptry signature. Lastly, they used the genotypes to explore the haplotypes present in the monkeys inoculated with sporozoites from a two strain cross finding only 4 haplotypes from one monkey and 2 from the other.

The overall study design is original and addresses important questions in the field of P.vivax malaria transmission, including beginning to better understand multi-strain infections. The limited sample number, and lack of certain controls prevents a more confident interpretation of the observations but the authors recognize this in the manuscript. They do overstate findings in the conclusions and need to tone this down (eg. line 318 regarding “inability”) when all sorts of other factors might have been at play as the authors themselves mention earlier.

There are areas where minor improvements will substantially improve the paper. These are listed below.

All figures need improvement on font sizes, clarity, and standard colour schemes, and too much white space.

The manuscript will benefit from a schematic of all the animal infections and workflows, giving IDs to the sampling points and the monkeys used for sampling, as provided in the supplementary file. This will help the reader follow the results and the experiments. Also, please indicate the ID of the monkeys in the individual figures. Please also name the species and strain of mosquito used in the main text and be much clearer in the methods as to how the two species of mosquitoes were used in the work (and why).

The authors need to discuss why only female gametocytes are observed in the scRNAseq data. Fertilisation was possible for the generation of sporozoites upon feeding mosquitoes with parasites from co-infected monkeys indicating male gametocytes were in circulation. Could the parasite

isolation and scRNAseq protocol have failed to capture the males as a limitation?

It was not clear why genotyping of the monoclonal infections was necessary – each parasite's genotype should have already been known based on which monkey it came from.

DE was compared between 3 NIH-1993-F3 mono-infection samples VS 3 Chesson mono-infection samples. It would be good to perform DE between NIH-1993-F3 VS Chesson within each of the 2 co-infected samples because these are not affected by batch and confirm these results are similar to those presented comparing gene expression between the strains in the monoclonal infections.

In 4B, there's already a bias towards NIH with respect to the number of sporozoites that were subsequently used for infection. Could you comment on this? Could this be a reason for having no Chesson selfed parasites in the new infections?

The conclusion that only very few hepatic schizonts seed infections due to strong "liver bottleneck" can be strengthened by correlating with the number of distinct haplotypes in the sporozoites from the same set as those used in the inoculum. To do this, the haplotypes from the sporozoites should be reconstructed as the authors have done for the blood stages, similar to supplemental figure 6. It should then also be possible to understand approximately how many oocysts contributed to the sporozoites that were in the scRNAseq sample set – were sporozoites observed in each of the 20 mosquitoes or were the mosquitoes mashed and pooled and some could have been uninfected? A deeper exploration of the haplotypes using the SNP data is warranted here.

The figures and tables need improvements as mentioned above, as well as
Figure 1 and Supp Figure 3: Color by stage labels and/or the pseudobulks (Group A,B,C,S) used to perform differential gene expression. Also, In Fig 1, A & B, images of PCA/UMAP plot colored by days post-infection, sampling points, and donors would be informative.

Figure 2 C: indicate the number of samples per group. There's a black dot (in addition to the red/blue dots) that might be misleading in the box plots. Label the dots in the boxplot by timepoint, to check progression of sexual proportion over time. Similarly for the other box plots representing sexual proportions

Figure 4E: x-axis label is cropped.

Supplemental Figure 4: No letters in the caption for the figure

Table 1 is not very clear. Is it all the cells from the study? The number of blood stage cells don't add up to 117,350. Information also seems to be conflicting with line 378 - "we collected 1 ml of blood from each animal to prepare 10X 3' end scRNA-seq libraries (for a total of 3-4 samples per animal)"

Other minor comments:
Line 87: do you mean 2 million parasites, or 2 million red blood cells, some small percentage of which have parasites? Please be clear about the parasitemia of the inoculation throughout (e.g. line 91 says it is 2 million parasitized RBCs, so presumably many more RBCs in total injected).

Line 98: Which co-infected monkeys are these? “Simultaneous infections”

Line 175: Following re-infection with the second strain, sampling was done on day 4,7 and 9?days. As rightly mentioned, these parasites might not have had enough time to establish infection and to show up in the blood stream. Is there any deeper analysis of the existing data or a more sensitive alternative method (PCR of a locus) that can be used to detect the presence of the second strain?

In line 173-175, “In a first attempt to establish P. vivax polyclonal infections, we re-inoculated, two weeks after the first inoculation, animals initially infected with one of the two strains with 1 million fresh RBCs parasitized with the other strain” is contradicted in Line 179-183 : “One explanation for the patterns observed is that once one infection is established, with hundreds of thousands (or millions) of parasites circulating, subsequent inoculation (or superinfection36) of a few thousand parasites from a different strain is unlikely to lead to a significant presence in the blood unless they have a significant growth advantage”

It would help if there is an estimate of the number of parasites of the first strain in circulation prior to the consecutive infection?

Line 242: what genotypes are included in the 13k sporozoite inoculation – is this from the exact same set of parasites just explored above using scRNAseq?

Line 264: “Third, we do not know if all salivary gland sporozoites are fully mature”. Can you check markers in the scRNA-seq data for the maturation state of the sporozoites? Could you explore maturity with RNA velocity? Could you perform pseudotime analysis on the sporozoites to explore if one strain seems younger than another?

Line 353 - 2 cryopreserved and 8 fresh RBC seems to conflict with information in lines 87-90 where it is indicated that 2+2 monkeys were infected with cryopreserved RBCs.

Software/Packages used for different analyses are not listed in methods eg pseudotime

Supplementary tables need better legends/keys

Line 250: “This observation was unexpected given that the Chesson strain, in monoclonal infections, is able to be transmitted from mosquitoes to Saimiri monkeys.”. Citation/data needed

Line 409 - “We then use custom scripts to count the number of Chesson alleles and NIH alleles present in each cell, considering, for blood-stage parasites, only positions sequenced by at least five reads, with 75% of the reads carrying the same allele”. Does samtools mpileup consider PCR duplicates? Please clarify

Reviewer #2 (Remarks to the Author):

In this manuscript by Hazzard and colleagues, single cell RNAseq (scRNAseq) analyses of the Plasmodium vivax malaria parasite are described following a unique set of infection experiments with non-human primates (Saimiri monkeys) and mosquitoes. The findings are generally well-described, and will be of interest to the malaria community given the novelty and rarity of such controlled infection experiments in non-human primates. Some of the observations remain anecdotal owing to the practical constraints of doing such experiments, however the results remain interesting. The most important findings in this manuscript include: 1) evidence of superinfection resistance, at least when mediated by infected red blood cell injection, 2) simultaneous co-infection by two parasite strains does not impact sexual commitment or gene expression of each strain, 3) very severe bottlenecks in liver stage infection following injection of sporozoites.

The manuscript could be improved through further interpretation/clarification of some of the findings. Specific comments:

Line 180: The authors speculate that in the consecutive infection experiments, existing parasitemia could be too high to enable detection of the strain introduced in the second inoculation 4, 7, or 9 days later. Was the parasitemia of the animals assessed at 4, 7, and 9 days? It could be helpful to more quantitatively predict the power to detect the secondarily introduced strain as a function of growth rate under the null hypothesis of no competition, and the number of infected RBCs profiled at 4/7/9 days.

Line 254: Again, it would be useful to more quantitatively assess the hypothesis that the Chesson strain was not established in monkeys after sporozoite injection due to chance. If 13K sporozoites were injected, and Chesson selfed sporozoites were x% of that inoculum, how tight would the bottleneck need to be to make it likely that Chesson would not be observed in the bloodstage?

Line 278: The authors note that the proportion of gametocytes is lower in infections initiated by sporozoites rather than by infected RBCs. One of the frequently cited biological differences between P vivax and P falciparum is that the former undergoes sexual commitment immediately (or very soon) upon egress from the liver, however the data supporting this observation are thin. Can the authors infer the timing of sexual commitment upon emergence from the liver from their data?

Figure 5 A and B: On some chromosomes these two outcrossed parasites show evidence of very many recombination events, representing crossovers and/or gene conversion events. Does this observation accord with previous estimates of Plasmodium recombination rates from other primate and mouse crossing experiments? Could genotyping error contribute to some of the surprising intermingling of red and blue alleles on some chromosomes?

Small comments:

Multiple sentences begin with a numeral, negatively impacting readability.

Line 240: . . . some [sporozoites, rather than gametocytes] carried a mixture of NIH-1993-F3. . .

Figure 4D: What is signified by the differently sized points on this plot?

Reviewer #3 (Remarks to the Author):

This paper represents a major experimental undertaking and one that has great potential benefit to the scientific community. The authors have used an experimental primate model to assess single cell gene expression profiles of Plasmodium vivax blood stage and sporozoite stage parasites. Since these parasites cannot be cultured in in vitro systems, these data provide the first detailed analysis of gene expression and the resulting data is an incredible resource for the field. Of particular note is the differential expression of the PvAP2-G gene in the Chesson and NIH-1993-F3 parasites and the correlated difference in the proportion of female gametocytes in blood stage infections. This gene is critical in the developmental switch from asexual replication to the sexual stage developmental pathway essential for transmission by the mosquito.
Overall, the authors are able to delineate the expression profile for parasites throughout the asexual replication cycle, the stage IV-V gametocytes and salivary gland sporozoites. This is a valuable resource that can underpin many future studies.

The remainder of the manuscript explore the potential impact of co-infection on parasite gene expression, exploring to potential routes to co-infection - sequential blood infection or simultaneous blood infection. While these experiments represent a major technical feat, the conclusions are not well supported. The authors conclude the sequential infection is less successful, but there could be many reasons for this and the experiments presented do not offer insights into potential biological mechanisms. These infections were initiated by injecting blood stage parasites - not the natural route of infection. The authors imply that this could be due to immune mechanisms or lack of adequate reticulocytes but provide no additional data. It is not clear what the authors intended - the simultaneous injection model seems to have resulted in the two parasites being present and from these data the authors conclude that co-infection does not impact gene expression in the co-infecting strains. While interesting, this result is not particularly surprising and has been explored in the other major human malaria, P. falciparum.

The final part of the manuscript is again a major technical achievement and provides some insight into both sporozoite development and subsequent transmission into primate hosts. Mosquitos were fed on the co-infected primates. Sporozoites were analayzed after 17-20 days. Both the NIH-1993-F3 and Chesson geneotypes were present (presumably via selfing) as well as hybrid genotypes

reflecting sexual exchange during mosquito transmission. It would be useful to look more carefully at the hybrid genotypes, both in terms of proportions and inherited regions.

Subsequent sporozoite transmission to new animals demonstrated that only the NIH-1993-F3 and hybrid genotype parasites were present in the subsequent blood stage infection. Parasites with the Chesson only genotype were not present. This is a novel result and is perhaps consistent with observation that the NIH-1993-F3 genotype has a higher proportion of female gametocytes. The quantitative analysis of the individual sporozoite genotypes demonstrates a higher proportion were the NIH-1993-F3 genotype. It would be interesting to map the hybrid genotypes in the succesful blood stage infections to determine specific regions were inherited from the NIH-1993-F3 parent.

We would like to thank the reviewers for their positive and constructive assessment of our work and our manuscript. We have addressed below their specific comments and suggestions, with the line numbers referring to the changes made in the manuscript with track changes.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript provides interesting insights into the establishment of polyclonal infections using an animal model coupled with short read single cell RNA sequencing to profile the parasite transcriptome and genotypes. First, they explore monoinfections of two different strains, finding that among the cells where they have power (e.g. confident strain assignment, and good number of cells) strains show some DE but this is largely in PIRs or in genes that they observe to be deleted in one of the strains. Then they carried out sequential infections with a two week gap but these failed to establish polyclonal infections and did not show compelling DE compared to monoclonal infections. Then they explored inoculations with two strains simultaneously – again they did not find much DE here between strains. Then they fed mosquitoes on monkeys carrying two strains and explored the sporozoites by scRNAseq as well as inoculating two more monkeys with these sporozoites. The sporozoite data showed all three expected types (pure strains and crosses) but the monkey infections did not have pure Chesson strain present. They compared expression between NIH parasites in an infection resulting from blood stage inoculation vs sporozoite inoculation, finding an intriguing rhoptry signature. Lastly, they used the genotypes to explore the haplotypes present in the monkeys inoculated with sporozoites from a two strain cross finding only 4 haplotypes from one monkey and 2 from the other.

The overall study design is original and addresses important questions in the field of P.vivax malaria transmission, including beginning to better understand multi-strain infections. The limited sample number, and lack of certain controls prevents a more confident interpretation of the observations but the authors recognize this in the manuscript. They do overstate findings in the conclusions and need to tone this down (eg. line 318 regarding “inability”) when all sorts of other factors might have been at play as the authors themselves mention earlier.
We would like to thank the reviewer for the encouraging assessment of our work and its relevance. We have rephrased the “inability” sentence (now on line 346), as well as the rest of the discussion, to moderate our conclusions.

There are areas where minor improvements will substantially improve the paper. These are listed below.

All figures need improvement on font sizes, clarity, and standard colour schemes, and too much white space.
We have standardized the font size and the colors across all figures (blue always indicates Chesson, red always NIH) and added extra information to reduce white space.

The manuscript will benefit from a schematic of all the animal infections and workflows, giving IDs to the sampling points and the monkeys used for sampling, as provided in the supplementary file. This will help the reader follow the results and the experiments. Also, please indicate the ID of the monkeys in the individual figures. Please also name the species and strain of mosquito used in the main text and be much clearer in the methods as to how the two species of mosquitoes were used in the work (and why).

Thank you for these suggestions. We have now added a figure (Fig. 1A) summarizing the overall experimental design and indicating the IDs of all monkeys (those have also been added to all figures). The species and strains of the mosquitoes used are now clearly indicated on lines 99, 242 and 248. We chose to use *An. freeborni* and *An. stephensi* (Nijmegen strain) because both *P. vivax* strains are successfully transmitted by these species and are readily accessible in our insectary colonies.

The authors need to discuss why only female gametocytes are observed in the scRNAseq data. Fertilisation was possible for the generation of sporozoites upon feeding mosquitoes with parasites from co-infected monkeys indicating male gametocytes were in circulation. Could the parasite isolation and scRNAseq protocol have failed to capture the males as a limitation?
The reviewer is correct: male gametocytes were indeed present for both strains in the blood (since fertilization occurs in both directions) but these parasites were lost during the scRNA-seq experiments. The most likely explanation for this loss is that male gametocytes exflagellated during the enrichment with MACS columns and/or the loading on the 10X instrument (that were both performed at room temperature). This issue is now specifically discussed on lines 147-151.

It was not clear why genotyping of the monoclonal infections was necessary – each parasite's genotype should have already been known based on which monkey it came from.
We chose to include the genotyping of monoclonal infections as a proof-of-principle, to demonstrate that we could accurately genotype single cells and therefore rigorously analyze genotypes in polyclonal infections. This is indicated on line 133-134.

DE was compared between 3 NIH-1993-F3 mono-infection samples VS 3 Chesson mono-infection samples. It would be good to perform DE between NIH-1993-F3 VS Chesson within each of the 2 co-infected samples because these are not affected by batch and confirm these results are similar to those presented comparing gene expression between the strains in the monoclonal infections.
We conducted the analyses suggested by the reviewer. The results were overall highly similar, suggesting that animals and batch effects did not dramatically affect the differential gene expression analysis. For example, 96 genes were differentially expressed between NIH-1993-F3 and Chesson among early asexuals (group A) from the coinfections and 83 of those (86%) were also identified as differentially expressed between strains (and in the same direction) using the monoinfections. Similarly, 50 genes were differentially expressed in sexual parasites in coinfections and 46 of those (92%) were also deemed differentially expressed (and in the same direction) in the monoinfections. It is however worth noting that since much fewer cells are characterized in the coinfections than in monoinfections (14,005 vs 2,390 in Group A and 7,926 vs 1,975 in sexual parasites), the power is lower in these analyses (and could explain why some differences are not recapitulated). These analyses are now discussed briefly on line 219-223 and the results included in Supplemental Table S3.

In 4B, there's already a bias towards NIH with respect to the number of sporozoites that were subsequently used for infection. Could you comment on this? Could this be a reason for having no Chesson selfed parasites in the new infections?
The reviewer is correct about the bias towards NIH in sporozoites. One likely explanation for this distortion is that the NIH strain produces more female gametocytes than Chesson and therefore contributes to a larger amount of sporozoites (although the lack of information on male gametocytes precludes drawing definitive conclusions). This point is now discussed on line 259-262-.
This bias is relatively moderate (2- to 5- fold) but it might be sufficient to explain the disappearance of the Chesson genotype in a strong bottleneck (this is discussed on line 278-279, see also table below). Further studies, with more animals and better characterization of the sporozoites (using smart-seq?) will

be needed to rigorously test this scenario and examine the alternative, non-exclusive, hypotheses discussed in the manuscript (i.e., differences in sporozoites maturation rate and/or pre-erythrocytic development).

The conclusion that only very few hepatic schizonts seed infections due to strong "liver bottleneck" can be strengthened by correlating with the number of distinct haplotypes in the sporozoites from the same set as those used in the inoculum. To do this, the haplotypes from the sporozoites should be reconstructed as the authors have done for the blood stages, similar to supplemental figure 6. It should then also be possible to understand approximately how many oocysts contributed to the sporozoites that were in the scRNAseq sample set – were sporozoites observed in each of the 20 mosquitoes or were the mosquitoes mashed and pooled and some could have been uninfected? A deeper exploration of the haplotypes using the SNP data is warranted here.
These are great suggestions that we would love to test. However, these suggestions are difficult to implement with our current data. Most of the reads (>90%) generated by scRNA-seq from the dissected salivary glands derive from mosquito RNA contamination (and not *Plasmodium* sporozoite RNA). As a consequence, each sporozoite is only characterized by 250 reads on average (compared to 17,594 reads for an average blood stage parasite) greatly limiting the number of positions that can be genotyped: while we are able to accurately genotype hundreds of segregating positions for blood stage parasites (Fig 4C) and therefore reconstruct haplotypes across the entire genome, less than 20 positions are "genotypable" in most sporozoites (Fig 4B) preventing such analyses (we only have ~1 genotype per chromosome). This massive contamination with ambient RNA is an issue that we, and others in the field, are trying to solve but have not successfully addressed yet.
The comment of the reviewer on the number of oocysts contributing to the sporozoites is interesting and would be fascinating to examine further. Again, the data quality is limiting and, in addition, the sporozoites analyzed in our study derived from pooled dissected salivary glands from ~20 mosquitoes, preventing determination of the sporozoite infection rate (clarified on line 244). We did check the proportion of mosquitoes carrying oocysts and most mosquitoes were infected (see below).

Blood source	5163 (35 dpi)		5164 (35 dpi)		5164 (36 dpi)	
Mosquito	*An. freeborni* (n=16)	*An. stephensi* (n=20)	*An. freeborni* (n=16)	*An. stephensi* (n=19)	*An. freeborni* (n=11)	*An. stephensi* (n=18)
%infected	75%	40%	56%	84%	91%	83%
#oocysts (per infected gut)	15	2	45	31	45	41

The figures and tables need improvements as mentioned above, as well as
Figure 1 and Supp Figure 3: Color by stage labels and/or the pseudobulks (Group A,B,C,S) used to perform differential gene expression. Also, In Fig 1, A & B, images of PCA/UMAP plot colored by days post-infection, sampling points, and donors would be informative.
We improved the figures and tables and have now included the figures requested by the reviewer's comments in Supplemental information (the PCA and UMAP colored by the pseudobulk groups in Figure S6 and separated cells based on days post infection in Figure S3).

Figure 2 C: indicate the number of samples per group. There's a black dot (in addition to the red/blue dots) that might be misleading in the box plots. Label the dots in the boxplot by timepoint, to check progression of sexual proportion over time. Similarly for the other box plots representing sexual

proportions
We have indicated the number of samples in the figure legend of Fig 2C and removed the black dot. There are no obvious changes in gametocyte proportion over time (this is now clearly depicted on Fig S6C).

Figure 4E: x-axis label is cropped.
Fixed

Supplemental Figure 4: No letters in the caption for the figure
Fixed

Table 1 is not very clear. Is it all the cells from the study? The number of blood stage cells don't add up to 117,350. Information also seems to be conflicting with line 378 - "we collected 1 ml of blood from each animal to prepare 10X 3' end scRNA-seq libraries (for a total of 3-4 samples per animal)".
Thank you for noticing this mistake. The cell counts are now consistent between the text, tables and figure legends.

Other minor comments:
Line 87: do you mean 2 million parasites, or 2 million red blood cells, some small percentage of which have parasites? Please be clear about the parasitemia of the inoculation throughout (e.g. line 91 says it is 2 million parasitized RBCs, so presumably many more RBCs in total injected).
2 million parasitized RBCs. We have clarified this point in the text (line 87).

Line 98: Which co-infected monkeys are these? "Simultaneous infections"
Yes. We have harmonized the terminology throughout the manuscript.

Line 175: Following re-infection with the second strain, sampling was done on day 4,7 and 9?days. As rightly mentioned, these parasites might not have had enough time to establish infection and to show up in the blood stream. Is there any deeper analysis of the existing data or a more sensitive alternative method (PCR of a locus) that can be used to detect the presence of the second strain?
We did detect parasites from the second strain, albeit at very low proportion (<1%, see Fig. 2F). We could try to validate this result by PCR but we believe that this approach is likely to be less sensitive than the scRNA-seq data: e.g., 5 in 4,027 parasites derives from the second strain in one sample and we would need very high sequence coverage of amplicons to detect them after PCR (probably >10,000 X) and it would still be challenging to differentiate true genotypes from PCR or sequencing errors.

In line 173-175, "In a first attempt to establish P. vivax polyclonal infections, we re-inoculated, two weeks after the first inoculation, animals initially infected with one of the two strains with 1 million fresh RBCs parasitized with the other strain" is contradicted in Line 179-183 : "One explanation for the patterns observed is that once one infection is established, with hundreds of thousands (or millions) of parasites circulating, subsequent inoculation (or superinfection36) of a few thousand parasites from a different strain is unlikely to lead to a significant presence in the blood unless they have a significant growth advantage"
Our apologies for the confusion. We have rephrased the second sentence to reflect more accurately the experiment conducted.

It would help if there is an estimate of the number of parasites of the first strain in circulation prior to

the consecutive infection?
We have added this estimate on line 190-193. Thank you for the suggestion.

Line 242: what genotypes are included in the 13k sporozoite inoculation – is this from the exact same set of parasites just explored above using scRNAseq?
The sporozoites used for the inoculation derived from a different batch than those used for scRNA-seq but were collected from one of the same strain of mosquitoes and feed on the same animals (we have switched one of the figure to show sporozoite data from both mosquito species). We have no reason to believe that the genotypes would be different but have included this limitation on line 293-294.

Line 264: "Third, we do not know if all salivary gland sporozoites are fully mature". Can you check markers in the scRNA-seq data for the maturation state of the sporozoites? Could you explore maturity with RNA velocity? Could you perform pseudotime analysis on the sporozoites to explore if one strain seems younger than another?
Again, we are concerned about overanalyzing the sporozoites given the overall scRNA-seq data quality for these samples. We examined the expression of genes specifically expressed early and late during sporozoite development (based on markers of the development of *P. falciparum* and *P. berghei* sporozoites) and found no statistical differences in the maturation of Chesson and NIH sporozoites.

		NIH	Chesson
	Total Cells	5416	3368
Based on *P. falciparum* expression (Real et al., 2021)	Early spz development (CRPM1, CRMP2, CRMP3, P41)	62 (1.45%)	42 (1.24%)
	Late spz development (LSAP1, microneme associated antigen, TLP)	1208 (22.30%)	678 (20.13%)
Based on *P. berghei* expression (Bogale et al., 2021)	Early spz development (MAEBL, SPELD, TRAP)	1169 (21.58%)	620 (18.41%)
	Late spz development (UIS3)	188 (3.47%)	130 (3.86%)

Line 353 - 2 cryopreserved and 8 fresh RBC seems to conflict with information in lines 87-90 where it is indicated that 2+2 monkeys were infected with cryopreserved RBCs.
We apologize for the confusion. We have corrected the numbers.

Software/Packages used for different analyses are not listed in methods eg pseudotime
We have clarified the methods used for different analyses in the Materials and Method section.

Supplementary tables need better legends/keys
We have edited the legends of the supplementary tables.

Line 250: "This observation was unexpected given that the Chesson strain, in monoclonal infections, is able to be transmitted from mosquitoes to Saimiri monkeys.". Citation/data needed
We have now added the relevant citations on line 275.

Line 409 - “We then use custom scripts to count the number of Chesson alleles and NIH alleles present in each cell, considering, for blood-stage parasites, only positions sequenced by at least five reads, with 75% of the reads carrying the same allele”. Does samtools mpileup consider PCR duplicates? Please clarify

PCR duplicates were removed by the custom scripts. This is now indicated on line 449.

Reviewer #2 (Remarks to the Author):

In this manuscript by Hazzard and colleagues, single cell RNAseq (scRNAseq) analyses of the Plasmodium vivax malaria parasite are described following a unique set of infection experiments with non-human primates (Saimiri monkeys) and mosquitoes. The findings are generally well-described, and will be of interest to the malaria community given the novelty and rarity of such controlled infection experiments in non-human primates. Some of the observations remain anecdotal owing to the practical constraints of doing such experiments, however the results remain interesting. The most important findings in this manuscript include: 1) evidence of superinfection resistance, at least when mediated by infected red blood cell injection, 2) simultaneous co-infection by two parasite strains does not impact sexual commitment or gene expression of each strain, 3) very severe bottlenecks in liver stage infection following injection of sporozoites.

We thank the reviewer for the assessment of our work.

The manuscript could be improved through further interpretation/clarification of some of the findings. Specific comments:

Line 180: The authors speculate that in the consecutive infection experiments, existing parasitemia could be too high to enable detection of the strain introduced in the second inoculation 4, 7, or 9 days later. Was the parasitemia of the animals assessed at 4, 7, and 9 days? It could be helpful to more quantitatively predict the power to detect the secondarily introduced strain as a function of growth rate under the null hypothesis of no competition, and the number of infected RBCs profiled at 4/7/9 days.

The parasitemia was indeed measured throughout the infection and is displayed on Supp Fig 2. We have now estimated the number of parasites that could result from the initial infection (assuming no competition, no parasite clearance and no limitations on the number of available reticulocytes) and the consequences on our ability to detect the second strain (see line 190 and our response to Reviewer 1).

Line 254: Again, it would be useful to more quantitatively assess the hypothesis that the Chesson strain was not established in monkeys after sporozoite injection due to chance. If 13K sporozoites were injected, and Chesson selfed sporozoites were x% of that inoculum, how tight would the bottleneck need to be to make it likely that Chesson would not be observed in the bloodstage?

If p is the proportion of Chesson sporozoites observed in the salivary glands, the probability of not seeing any Chesson blood stage parasites (x=0) if n sporozoites contribute to the blood stage infection is: $P(x=0) = \binom{n}{x} p^x (1-p)^{n-x}$

Or based on the number of Chesson sporozoites observed in the scRNA-seq data:

Proportion of Chesson in SG (based on scRNA-seq data)	0.16 (406/(406+1289+852))	0.027 (3/(3+50+59))
Number of hepatic schizonts	Probability of not seeing Chesson blood stage parasites	
1	0.840	0.973

5	0.418	0.872
10	0.175	0.761
15	0.073	0.663
20	0.031	0.578
25	0.013	0.504
50	1.6E-04	0.254
100	2.6E-08	0.065

These crude statistics are discussed on lines 280-283.

Line 278: The authors note that the proportion of gametocytes is lower in infections initiated by sporozoites rather than by infected RBCs. One of the frequently cited biological differences between P vivax and P falciparum is that the former undergoes sexual commitment immediately (or very soon) upon egress from the liver, however the data supporting this observation are thin. Can the authors infer the timing of sexual commitment upon emergence from the liver from their data?
We are not sure how to conduct the analyses suggested by the reviewer.

Figure 5 A and B: On some chromosomes these two outcrossed parasites show evidence of very many recombination events, representing crossovers and/or gene conversion events. Does this observation accord with previous estimates of Plasmodium recombination rates from other primate and mouse crossing experiments? Could genotyping error contribute to some of the surprising intermingling of red and blue alleles on some chromosomes?
It is possible that some of the multiple crossovers spuriously derive from genotyping errors (especially, on Fig 5A and 5B where individual genotypes are represented). However, the profiles depicted on Figure 5C, that also occasionally show two crossovers per chromosome, are very consistent with the profiles from mouse crossing experiments (e.g., Fig 2 in Vaughan et al., 2015).

Small comments:

Multiple sentences begin with a numeral, negatively impacting readability.
We have modified these sentences to start with a word.

Line 240: . . . some [sporozoites, rather than gametocytes] carried a mixture of NIH-1993-F3. . .
Thank you for noticing this mistake!

Figure 4D: What is signified by the differently sized points on this plot?
The size of the point is proportional to the number of cells characterized in that infection. This is now clearly indicated in the figure legend.

Reviewer #3 (Remarks to the Author):

This paper represents a major experimental undertaking and one that has great potential benefit to the scientific community. The authors have used an experimental primate model to assess single cell gene expression profiles of Plasmodium vivax blood stage and sporozoite stage parasites. Since these parasites cannot be cultured in in vitro systems, these data provide the first detailed analysis of gene expression and the resulting data is an incredible resource for the field. Of particular note is the differential expression of the PvAP2-G gene in the Chesson and NIH-1993-F3 parasites and the correlated difference in the proportion of female gametocytes in blood stage infections. This gene is critical in the developmental switch from asexual replication to the sexual stage developmental pathway

essential for transmission by the mosquito.
Overall, the authors are able to delineate the expression profile for parasites throughout the asexual replication cycle, the stage IV-V gametocytes and salivary gland sporozoites. This is a valuable resource that can underpin many future studies.
Thank you for the kind assessment

The remainder of the manuscript explore the potential impact of co-infection on parasite gene expression, exploring to potential routes to co-infection - sequential blood infection or simultaneous blood infection. While these experiments represent a major technical feat, the conclusions are not well supported. The authors conclude the sequential infection is less successful, but there could be many reasons for this and the experiments presented do not offer insights into potential biological mechanisms. These infections were initiated by injecting blood stage parasites - not the natural route of infection. The authors imply that this could be due to immune mechanisms or lack of adequate reticulocytes but provide no additional data. It is not clear what the authors intended - the simultaneous injection model seems to have resulted in the two parasites being present and from these data the authors conclude that co-infection does not impact gene expression in the co-infecting strains. While interesting, this result is not particularly surprising and has been explored in the other major human malaria, P. falciparum.

The final part of the manuscript is again a major technical achievement and provides some insight into both sporozoite development and subsequent transmission into primate hosts. Mosquitos were fed on the co-infected primates. Sporozoites were analayzed after 17-20 days. Both the NIH-1993-F3 and Chesson geneotypes were present (presumably via selfing) as well as hybrid genotypes reflecting sexual exchange during mosquito transmission. It would be useful to look more carefully at the hybrid genotypes, both in terms of proportions and inherited regions.
While the analyses suggested by the reviewer would be fascinating, we feel that the very small number of hybrid genotypes obtained (n=4!) would preclude any robust conclusions. We hope to soon expand this work and generate more complex offspring pools to enable the type of analyses suggested.

Subsequent sporozoite transmission to new animals demonstrated that only the NIH-1993-F3 and hybrid genotype parasites were present in the subsequent blood stage infection. Parasites with the Chesson only genotype were not present. This is a novel result and is perhaps consistent with observation that the NIH-1993-F3 genotype has a higher proportion of female gametocytes. The quantitative analysis of the individual sporozoite genotypes demonstrates a higher proportion were the NIH-1993-F3 genotype. It would be interesting to map the hybrid genotypes in the succesful blood stage infections to determine specific regions were inherited from the NIH-1993-F3 parent.
We agree but, again, believe that the small number of hybrid progenies prevents such analyses (for now). This is discussed on line 297-300.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors responded to all our comments and made the text and figures clearer to interpret.
The manuscript significantly improved with the edits following the reviewers' suggestion.
The additional analysis done provides more clarity to the interpretations and the language has been toned down where requested.

Reviewer #2 (Remarks to the Author):

I am satisfied with the authors' response to my comments and those of the other reviewers.

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 #4 (Remarks on code availability):

The code at the GitHub repository is described with enough detail to utilize and reproduce the results from the paper if needed.

Reviewer #5 (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.

We would like to thank the reviewers for their positive and constructive assessment of our work and our manuscript. We have addressed below their specific comments and suggestions, with the line numbers referring to the changes made in the manuscript with track changes.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript provides interesting insights into the establishment of polyclonal infections using an animal model coupled with short read single cell RNA sequencing to profile the parasite transcriptome and genotypes. First, they explore monoinfections of two different strains, finding that among the cells where they have power (e.g. confident strain assignment, and good number of cells) strains show some DE but this is largely in PIRs or in genes that they observe to be deleted in one of the strains. Then they carried out sequential infections with a two week gap but these failed to establish polyclonal infections and did not show compelling DE compared to monoclonal infections. Then they explored inoculations with two strains simultaneously – again they did not find much DE here between strains. Then they fed mosquitoes on monkeys carrying two strains and explored the sporozoites by scRNAseq as well as inoculating two more monkeys with these sporozoites. The sporozoite data showed all three expected types (pure strains and crosses) but the monkey infections did not have pure Chesson strain present. They compared expression between NIH parasites in an infection resulting from blood stage inoculation vs sporozoite inoculation, finding an intriguing rhoptry signature. Lastly, they used the genotypes to explore the haplotypes present in the monkeys inoculated with sporozoites from a two strain cross finding only 4 haplotypes from one monkey and 2 from the other.

The overall study design is original and addresses important questions in the field of P.vivax malaria transmission, including beginning to better understand multi-strain infections. The limited sample number, and lack of certain controls prevents a more confident interpretation of the observations but the authors recognize this in the manuscript. They do overstate findings in the conclusions and need to tone this down (eg. line 318 regarding “inability”) when all sorts of other factors might have been at play as the authors themselves mention earlier.
We would like to thank the reviewer for the encouraging assessment of our work and its relevance. We have rephrased the “inability” sentence (now on line 346), as well as the rest of the discussion, to moderate our conclusions.

There are areas where minor improvements will substantially improve the paper. These are listed below.

All figures need improvement on font sizes, clarity, and standard colour schemes, and too much white space.
We have standardized the font size and the colors across all figures (blue always indicates Chesson, red always NIH) and added extra information to reduce white space.

The manuscript will benefit from a schematic of all the animal infections and workflows, giving IDs to the sampling points and the monkeys used for sampling, as provided in the supplementary file. This will help the reader follow the results and the experiments. Also, please indicate the ID of the monkeys in the individual figures. Please also name the species and strain of mosquito used in the main text and be much clearer in the methods as to how the two species of mosquitoes were used in the work (and why).

Thank you for these suggestions. We have now added a figure (Fig. 1A) summarizing the overall experimental design and indicating the IDs of all monkeys (those have also been added to all figures). The species and strains of the mosquitoes used are now clearly indicated on lines 99, 242 and 248. We chose to use *An. freeborni* and *An. stephensi* (Nijmegen strain) because both *P. vivax* strains are successfully transmitted by these species and are readily accessible in our insectary colonies.

The authors need to discuss why only female gametocytes are observed in the scRNAseq data. Fertilisation was possible for the generation of sporozoites upon feeding mosquitoes with parasites from co-infected monkeys indicating male gametocytes were in circulation. Could the parasite isolation and scRNAseq protocol have failed to capture the males as a limitation?
The reviewer is correct: male gametocytes were indeed present for both strains in the blood (since fertilization occurs in both directions) but these parasites were lost during the scRNA-seq experiments. The most likely explanation for this loss is that male gametocytes exflagellated during the enrichment with MACS columns and/or the loading on the 10X instrument (that were both performed at room temperature). This issue is now specifically discussed on lines 147-151.

It was not clear why genotyping of the monoclonal infections was necessary – each parasite's genotype should have already been known based on which monkey it came from.
We chose to include the genotyping of monoclonal infections as a proof-of-principle, to demonstrate that we could accurately genotype single cells and therefore rigorously analyze genotypes in polyclonal infections. This is indicated on line 133-134.

DE was compared between 3 NIH-1993-F3 mono-infection samples VS 3 Chesson mono-infection samples. It would be good to perform DE between NIH-1993-F3 VS Chesson within each of the 2 co-infected samples because these are not affected by batch and confirm these results are similar to those presented comparing gene expression between the strains in the monoclonal infections.
We conducted the analyses suggested by the reviewer. The results were overall highly similar, suggesting that animals and batch effects did not dramatically affect the differential gene expression analysis. For example, 96 genes were differentially expressed between NIH-1993-F3 and Chesson among early asexuals (group A) from the coinfections and 83 of those (86%) were also identified as differentially expressed between strains (and in the same direction) using the monoinfections. Similarly, 50 genes were differentially expressed in sexual parasites in coinfections and 46 of those (92%) were also deemed differentially expressed (and in the same direction) in the monoinfections. It is however worth noting that since much fewer cells are characterized in the coinfections than in monoinfections (14,005 vs 2,390 in Group A and 7,926 vs 1,975 in sexual parasites), the power is lower in these analyses (and could explain why some differences are not recapitulated). These analyses are now discussed briefly on line 219-223 and the results included in Supplemental Table S3.

In 4B, there's already a bias towards NIH with respect to the number of sporozoites that were subsequently used for infection. Could you comment on this? Could this be a reason for having no Chesson selfed parasites in the new infections?
The reviewer is correct about the bias towards NIH in sporozoites. One likely explanation for this distortion is that the NIH strain produces more female gametocytes than Chesson and therefore contributes to a larger amount of sporozoites (although the lack of information on male gametocytes precludes drawing definitive conclusions). This point is now discussed on line 259-262-.
This bias is relatively moderate (2- to 5- fold) but it might be sufficient to explain the disappearance of the Chesson genotype in a strong bottleneck (this is discussed on line 278-279, see also table below). Further studies, with more animals and better characterization of the sporozoites (using smart-seq?) will

be needed to rigorously test this scenario and examine the alternative, non-exclusive, hypotheses discussed in the manuscript (i.e., differences in sporozoites maturation rate and/or pre-erythrocytic development).

The conclusion that only very few hepatic schizonts seed infections due to strong "liver bottleneck" can be strengthened by correlating with the number of distinct haplotypes in the sporozoites from the same set as those used in the inoculum. To do this, the haplotypes from the sporozoites should be reconstructed as the authors have done for the blood stages, similar to supplemental figure 6. It should then also be possible to understand approximately how many oocysts contributed to the sporozoites that were in the scRNAseq sample set – were sporozoites observed in each of the 20 mosquitoes or were the mosquitoes mashed and pooled and some could have been uninfected? A deeper exploration of the haplotypes using the SNP data is warranted here.
These are great suggestions that we would love to test. However, these suggestions are difficult to implement with our current data. Most of the reads (>90%) generated by scRNA-seq from the dissected salivary glands derive from mosquito RNA contamination (and not *Plasmodium* sporozoite RNA). As a consequence, each sporozoite is only characterized by 250 reads on average (compared to 17,594 reads for an average blood stage parasite) greatly limiting the number of positions that can be genotyped: while we are able to accurately genotype hundreds of segregating positions for blood stage parasites (Fig 4C) and therefore reconstruct haplotypes across the entire genome, less than 20 positions are "genotypable" in most sporozoites (Fig 4B) preventing such analyses (we only have ~1 genotype per chromosome). This massive contamination with ambient RNA is an issue that we, and others in the field, are trying to solve but have not successfully addressed yet.
The comment of the reviewer on the number of oocysts contributing to the sporozoites is interesting and would be fascinating to examine further. Again, the data quality is limiting and, in addition, the sporozoites analyzed in our study derived from pooled dissected salivary glands from ~20 mosquitoes, preventing determination of the sporozoite infection rate (clarified on line 244). We did check the proportion of mosquitoes carrying oocysts and most mosquitoes were infected (see below).

Blood source	5163 (35 dpi)		5164 (35 dpi)		5164 (36 dpi)	
Mosquito	*An. freeborni* (n=16)	*An. stephensi* (n=20)	*An. freeborni* (n=16)	*An. stephensi* (n=19)	*An. freeborni* (n=11)	*An. stephensi* (n=18)
%infected	75%	40%	56%	84%	91%	83%
#oocysts (per infected gut)	15	2	45	31	45	41

The figures and tables need improvements as mentioned above, as well as
Figure 1 and Supp Figure 3: Color by stage labels and/or the pseudobulks (Group A,B,C,S) used to perform differential gene expression. Also, In Fig 1, A & B, images of PCA/UMAP plot colored by days post-infection, sampling points, and donors would be informative.
We improved the figures and tables and have now included the figures requested by the reviewer's comments in Supplemental information (the PCA and UMAP colored by the pseudobulk groups in Figure S6 and separated cells based on days post infection in Figure S3).

Figure 2 C: indicate the number of samples per group. There's a black dot (in addition to the red/blue dots) that might be misleading in the box plots. Label the dots in the boxplot by timepoint, to check progression of sexual proportion over time. Similarly for the other box plots representing sexual

proportions
We have indicated the number of samples in the figure legend of Fig 2C and removed the black dot. There are no obvious changes in gametocyte proportion over time (this is now clearly depicted on Fig S6C).

Figure 4E: x-axis label is cropped.
Fixed

Supplemental Figure 4: No letters in the caption for the figure
Fixed

Table 1 is not very clear. Is it all the cells from the study? The number of blood stage cells don't add up to 117,350. Information also seems to be conflicting with line 378 - "we collected 1 ml of blood from each animal to prepare 10X 3' end scRNA-seq libraries (for a total of 3-4 samples per animal)".
Thank you for noticing this mistake. The cell counts are now consistent between the text, tables and figure legends.

Other minor comments:
Line 87: do you mean 2 million parasites, or 2 million red blood cells, some small percentage of which have parasites? Please be clear about the parasitemia of the inoculation throughout (e.g. line 91 says it is 2 million parasitized RBCs, so presumably many more RBCs in total injected).
2 million parasitized RBCs. We have clarified this point in the text (line 87).

Line 98: Which co-infected monkeys are these? "Simultaneous infections"
Yes. We have harmonized the terminology throughout the manuscript.

Line 175: Following re-infection with the second strain, sampling was done on day 4,7 and 9?days. As rightly mentioned, these parasites might not have had enough time to establish infection and to show up in the blood stream. Is there any deeper analysis of the existing data or a more sensitive alternative method (PCR of a locus) that can be used to detect the presence of the second strain?
We did detect parasites from the second strain, albeit at very low proportion (<1%, see Fig. 2F). We could try to validate this result by PCR but we believe that this approach is likely to be less sensitive than the scRNA-seq data: e.g., 5 in 4,027 parasites derives from the second strain in one sample and we would need very high sequence coverage of amplicons to detect them after PCR (probably >10,000 X) and it would still be challenging to differentiate true genotypes from PCR or sequencing errors.

In line 173-175, "In a first attempt to establish P. vivax polyclonal infections, we re-inoculated, two weeks after the first inoculation, animals initially infected with one of the two strains with 1 million fresh RBCs parasitized with the other strain" is contradicted in Line 179-183 : "One explanation for the patterns observed is that once one infection is established, with hundreds of thousands (or millions) of parasites circulating, subsequent inoculation (or superinfection36) of a few thousand parasites from a different strain is unlikely to lead to a significant presence in the blood unless they have a significant growth advantage"
Our apologies for the confusion. We have rephrased the second sentence to reflect more accurately the experiment conducted.

It would help if there is an estimate of the number of parasites of the first strain in circulation prior to

the consecutive infection?
We have added this estimate on line 190-193. Thank you for the suggestion.

Line 242: what genotypes are included in the 13k sporozoite inoculation – is this from the exact same set of parasites just explored above using scRNAseq?
The sporozoites used for the inoculation derived from a different batch than those used for scRNA-seq but were collected from one of the same strain of mosquitoes and feed on the same animals (we have switched one of the figure to show sporozoite data from both mosquito species). We have no reason to believe that the genotypes would be different but have included this limitation on line 293-294.

Line 264: "Third, we do not know if all salivary gland sporozoites are fully mature". Can you check markers in the scRNA-seq data for the maturation state of the sporozoites? Could you explore maturity with RNA velocity? Could you perform pseudotime analysis on the sporozoites to explore if one strain seems younger than another?
Again, we are concerned about overanalyzing the sporozoites given the overall scRNA-seq data quality for these samples. We examined the expression of genes specifically expressed early and late during sporozoite development (based on markers of the development of *P. falciparum* and *P. berghei* sporozoites) and found no statistical differences in the maturation of Chesson and NIH sporozoites.

		NIH	Chesson
	Total Cells	5416	3368
Based on *P. falciparum* expression (Real et al., 2021)	Early spz development (CRPM1, CRMP2, CRMP3, P41)	62 (1.45%)	42 (1.24%)
	Late spz development (LSAP1, microneme associated antigen, TLP)	1208 (22.30%)	678 (20.13%)
Based on *P. berghei* expression (Bogale et al., 2021)	Early spz development (MAEBL, SPELD, TRAP)	1169 (21.58%)	620 (18.41%)
	Late spz development (UIS3)	188 (3.47%)	130 (3.86%)

Line 353 - 2 cryopreserved and 8 fresh RBC seems to conflict with information in lines 87-90 where it is indicated that 2+2 monkeys were infected with cryopreserved RBCs.
We apologize for the confusion. We have corrected the numbers.

Software/Packages used for different analyses are not listed in methods eg pseudotime
We have clarified the methods used for different analyses in the Materials and Method section.

Supplementary tables need better legends/keys
We have edited the legends of the supplementary tables.

Line 250: "This observation was unexpected given that the Chesson strain, in monoclonal infections, is able to be transmitted from mosquitoes to Saimiri monkeys.". Citation/data needed
We have now added the relevant citations on line 275.

Line 409 - “We then use custom scripts to count the number of Chesson alleles and NIH alleles present in each cell, considering, for blood-stage parasites, only positions sequenced by at least five reads, with 75% of the reads carrying the same allele”. Does samtools mpileup consider PCR duplicates? Please clarify

PCR duplicates were removed by the custom scripts. This is now indicated on line 449.

Reviewer #2 (Remarks to the Author):

In this manuscript by Hazzard and colleagues, single cell RNAseq (scRNAseq) analyses of the Plasmodium vivax malaria parasite are described following a unique set of infection experiments with non-human primates (Saimiri monkeys) and mosquitoes. The findings are generally well-described, and will be of interest to the malaria community given the novelty and rarity of such controlled infection experiments in non-human primates. Some of the observations remain anecdotal owing to the practical constraints of doing such experiments, however the results remain interesting. The most important findings in this manuscript include: 1) evidence of superinfection resistance, at least when mediated by infected red blood cell injection, 2) simultaneous co-infection by two parasite strains does not impact sexual commitment or gene expression of each strain, 3) very severe bottlenecks in liver stage infection following injection of sporozoites.

We thank the reviewer for the assessment of our work.

The manuscript could be improved through further interpretation/clarification of some of the findings. Specific comments:

Line 180: The authors speculate that in the consecutive infection experiments, existing parasitemia could be too high to enable detection of the strain introduced in the second inoculation 4, 7, or 9 days later. Was the parasitemia of the animals assessed at 4, 7, and 9 days? It could be helpful to more quantitatively predict the power to detect the secondarily introduced strain as a function of growth rate under the null hypothesis of no competition, and the number of infected RBCs profiled at 4/7/9 days.

The parasitemia was indeed measured throughout the infection and is displayed on Supp Fig 2. We have now estimated the number of parasites that could result from the initial infection (assuming no competition, no parasite clearance and no limitations on the number of available reticulocytes) and the consequences on our ability to detect the second strain (see line 190 and our response to Reviewer 1).

Line 254: Again, it would be useful to more quantitatively assess the hypothesis that the Chesson strain was not established in monkeys after sporozoite injection due to chance. If 13K sporozoites were injected, and Chesson selfed sporozoites were x% of that inoculum, how tight would the bottleneck need to be to make it likely that Chesson would not be observed in the bloodstage?

If p is the proportion of Chesson sporozoites observed in the salivary glands, the probability of not seeing any Chesson blood stage parasites (x=0) if n sporozoites contribute to the blood stage infection is: $P(x=0) = \binom{n}{x} p^x (1-p)^{n-x}$

Or based on the number of Chesson sporozoites observed in the scRNA-seq data:

Proportion of Chesson in SG (based on scRNA-seq data)	0.16 (406/(406+1289+852))	0.027 (3/(3+50+59))
Number of hepatic schizonts	Probability of not seeing Chesson blood stage parasites	
1	0.840	0.973

5	0.418	0.872
10	0.175	0.761
15	0.073	0.663
20	0.031	0.578
25	0.013	0.504
50	1.6E-04	0.254
100	2.6E-08	0.065

These crude statistics are discussed on lines 280-283.

Line 278: The authors note that the proportion of gametocytes is lower in infections initiated by sporozoites rather than by infected RBCs. One of the frequently cited biological differences between P vivax and P falciparum is that the former undergoes sexual commitment immediately (or very soon) upon egress from the liver, however the data supporting this observation are thin. Can the authors infer the timing of sexual commitment upon emergence from the liver from their data?
We are not sure how to conduct the analyses suggested by the reviewer.

Figure 5 A and B: On some chromosomes these two outcrossed parasites show evidence of very many recombination events, representing crossovers and/or gene conversion events. Does this observation accord with previous estimates of Plasmodium recombination rates from other primate and mouse crossing experiments? Could genotyping error contribute to some of the surprising intermingling of red and blue alleles on some chromosomes?
It is possible that some of the multiple crossovers spuriously derive from genotyping errors (especially, on Fig 5A and 5B where individual genotypes are represented). However, the profiles depicted on Figure 5C, that also occasionally show two crossovers per chromosome, are very consistent with the profiles from mouse crossing experiments (e.g., Fig 2 in Vaughan et al., 2015).

Small comments:

Multiple sentences begin with a numeral, negatively impacting readability.
We have modified these sentences to start with a word.

Line 240: . . . some [sporozoites, rather than gametocytes] carried a mixture of NIH-1993-F3. . .
Thank you for noticing this mistake!

Figure 4D: What is signified by the differently sized points on this plot?
The size of the point is proportional to the number of cells characterized in that infection. This is now clearly indicated in the figure legend.

Reviewer #3 (Remarks to the Author):

This paper represents a major experimental undertaking and one that has great potential benefit to the scientific community. The authors have used an experimental primate model to assess single cell gene expression profiles of Plasmodium vivax blood stage and sporozoite stage parasites. Since these parasites cannot be cultured in in vitro systems, these data provide the first detailed analysis of gene expression and the resulting data is an incredible resource for the field. Of particular note is the differential expression of the PvAP2-G gene in the Chesson and NIH-1993-F3 parasites and the correlated difference in the proportion of female gametocytes in blood stage infections. This gene is critical in the developmental switch from asexual replication to the sexual stage developmental pathway

essential for transmission by the mosquito.
Overall, the authors are able to delineate the expression profile for parasites throughout the asexual replication cycle, the stage IV-V gametocytes and salivary gland sporozoites. This is a valuable resource that can underpin many future studies.
Thank you for the kind assessment

The remainder of the manuscript explore the potential impact of co-infection on parasite gene expression, exploring to potential routes to co-infection - sequential blood infection or simultaneous blood infection. While these experiments represent a major technical feat, the conclusions are not well supported. The authors conclude the sequential infection is less successful, but there could be many reasons for this and the experiments presented do not offer insights into potential biological mechanisms. These infections were initiated by injecting blood stage parasites - not the natural route of infection. The authors imply that this could be due to immune mechanisms or lack of adequate reticulocytes but provide no additional data. It is not clear what the authors intended - the simultaneous injection model seems to have resulted in the two parasites being present and from these data the authors conclude that co-infection does not impact gene expression in the co-infecting strains. While interesting, this result is not particularly surprising and has been explored in the other major human malaria, P. falciparum.

The final part of the manuscript is again a major technical achievement and provides some insight into both sporozoite development and subsequent transmission into primate hosts. Mosquitos were fed on the co-infected primates. Sporozoites were analayzed after 17-20 days. Both the NIH-1993-F3 and Chesson geneotypes were present (presumably via selfing) as well as hybrid genotypes reflecting sexual exchange during mosquito transmission. It would be useful to look more carefully at the hybrid genotypes, both in terms of proportions and inherited regions.
While the analyses suggested by the reviewer would be fascinating, we feel that the very small number of hybrid genotypes obtained (n=4!) would preclude any robust conclusions. We hope to soon expand this work and generate more complex offspring pools to enable the type of analyses suggested.

Subsequent sporozoite transmission to new animals demonstrated that only the NIH-1993-F3 and hybrid genotype parasites were present in the subsequent blood stage infection. Parasites with the Chesson only genotype were not present. This is a novel result and is perhaps consistent with observation that the NIH-1993-F3 genotype has a higher proportion of female gametocytes. The quantitative analysis of the individual sporozoite genotypes demonstrates a higher proportion were the NIH-1993-F3 genotype. It would be interesting to map the hybrid genotypes in the succesful blood stage infections to determine specific regions were inherited from the NIH-1993-F3 parent.
We agree but, again, believe that the small number of hybrid progenies prevents such analyses (for now). This is discussed on line 297-300.