Population Genomics of *Plasmodium vivax* in Panama to Assess the Risk of Case Importation on Malaria Elimination

Lucas E. Buyon[1,3], Ana Maria Santamaria [2], Angela M. Early[1,3], Mario Quijada [2], Itza Barahona[4], Jose Lasso[4], Mario Avila[4], Sarah K. Volkman [1,3,5] Matthias Marti[6], Daniel E. Neafsey[1,3], Nicanor Obaldia III[1,2,6]

[1] Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Boston, MA. USA, [2]Instituto Conmemorativo Gorgas de Estudios de la Salud, Panama City, Panama, [3]The Broad Institute, Cambridge, MA., USA [4]Ministerio de Salud, Panama, Panama, [5]Simmons University, College of Natural, Behavioral and Health Sciences, Boston, MA, USA [6]University of Glasgow, Glasgow, Scotland, UK.

To whom correspondence should be addressed: nobaldia@gorgas.gob.pa and neafsey@hsph.harvard.edu

Abstract

Malaria incidence in Panama has plateaued in recent years in spite of elimination efforts, with almost all cases caused by *Plasmodium vivax*. Notwithstanding, overall malaria prevalence remains low (fewer than 1 case per 1000 persons). We used selective whole genome amplification to sequence 96 *P. vivax* samples from Panama collected between 2007 and 2019 to study the population structure and transmission dynamics of the parasite. Imported cases resulting from increased levels of human migration could threaten malaria elimination prospects, and four of the samples evaluated came from individuals with travel history. We explored patterns of recent common ancestry among the samples and observed that a single highly genetically related lineage was dominant among the samples (47 out of 59 samples with good sequencing coverage), spanning the entire period of the collection (2007-2019) and all regions of the country. We also found a second, smaller clonal lineage of four parasites collected between 2017 and 2019. To explore the regional context of Panamanian *P. vivax* we conducted principal components analysis and constructed a neighbor-joining tree using these samples and samples collected worldwide from a previous study. Three of the four samples with travel history clustered with samples collected from their suspected country of origin (consistent with importation), while one appears to have been a result of local transmission. The small number of genetically unique Panamanian *P. viva*x samples clustered with samples collected from Colombia, suggesting they represent the genetically similar ancestral *P. vivax* population in Panama or were recently imported from Colombia. The low diversity we observe in Panama indicates that this parasite population has been previously subject to a severe bottleneck and may be eligible for elimination. Additionally, while we confirmed that *P. vivax* is imported to Panama from diverse geographic locations, the lack of impact from imported cases on the overall parasite

population genomic profile suggests that onward transmission from such cases is limited and that imported cases may not presently pose a major barrier to elimination.

Author Summary

Panama has greatly reduced *P. vivax* incidence, however, this progress has plateaued. Understanding parasite transmission patterns and identifying imported cases is critical to help Panama and other countries succeed in their elimination efforts. Genomic epidemiology and population genomics can help provide information needed to inform malaria control policy. In this study, we collected 100 Panamanian *P. viva*x samples from two collection periods (2007-2009 and 2017-2019), of which 59 yielded usable sequencing data. 4 samples had patient travel history data associated with them. We found that the majority of samples belong to a single highly related lineage, termed CL1. This lineage has persisted since at least 2007. We also highlight how genomic epidemiology can be used to spotlight parasites that may be imported as a result of human migration, as well as corroborate or refute the country of origin as suggested by patient travel history. We observe no evidence of outcrossing between these potentially imported parasites and the local Panamanian parasite population, suggesting that imported parasites are not driving ongoing malaria transmission in Panama. The low diversity we observe in Panama indicates that this parasite population has been previously subject to a severe bottleneck and may be eligible for elimination.

Introduction

Malaria is a parasitic disease transmitted by the bite of female *Anopheles* mosquitoes. Malaria parasites cause approximately 219 million cases and 435,000 deaths each year, the vast majority

in sub-Saharan Africa (1). *Plasmodium falciparum,* the most virulent of the five *Plasmodium* species that infect humans (*P. falciparum*, *P. vivax, P. malariae, P. ovale, and P. knowlesi*) causes the majority of these cases (1). Though billions of dollars have been devoted to the control and eradication of malaria caused by *P. falciparum,* comparatively little attention is given to *P. vivax,* the most prevalent malaria parasite outside Africa (2). The impact of *P. vivax* on human health was once considered minimal relative to the more virulent *P. falciparum* (1,2). However, recent studies suggest this parasite species causes a significant global health burden (1,2). The *P .vivax* cycle includes a dormant liver "hypnozoite" stage, unique among human malaria parasites. (2) The hypnozoite stage can cause a relapse of malaria years after initial infection, thus beginning the cycle of infection and complicating control efforts. (2)

Sixty percent of the population of Central and South America live in areas with ongoing malaria transmission, predominantly caused by *P. vivax* (3). The region experiences about 700,000 *P. vivax* cases each year (1). Between 2000 and 2015 the incidence rate of malaria fell 37% globally and 42% in Africa (1). In the Americas during the same period, malaria mortality decreased by 72% (1). Unfortunately, recent evidence suggests that this trend has stalled, and in some countries, malaria incidence has even increased (1). Panama eliminated the autochthonous transmission of *P. falciparum* in 2010*,* outside of a small outbreak on the Colombian border in 2015 (4). Since 2010, *P. vivax* has caused almost all the malaria cases in Panama (1,5,6). *P. vivax* cases in Panama have declined precipitously since 2005, from close to 1 case per 1000 persons, to under 0.25 cases per 1000 persons in 2017 (7). However, malaria incidence in Panama has plateaued since 2008. This plateau in incidence could be due to low levels of transmission and/or imported cases that are re-seeding infections.

93 Human movement leading to parasite migration is a potentially significant
94 epidemiological threat to malaria control in Panama. Parasite importation stemming from human
95 migration is a challenge to elimination programs in other countries around the world (8–10).
96 The unique geographic position of Panama makes it a crossroads for human migration to the
97 United States from South America (5,11). Migrants enter Panama through two paths: through
98 the Darien jungle region on the border with Colombia, and through the Kuna Yala Amerindian
99 reserve ('Comarca') on the Caribbean coast (5,6,11). Previous studies implicate these regions as
100 focal points of ongoing malaria transmission in Panama and suggest this is partly due to
101 imported parasites (5,11). It is estimated that approximately 60,000 continental and extra-
102 continental migrants crossed the southern border of Panama through the Darien jungle region in
103 2015 and 2016 (12).
104 To inform effective malaria elimination strategies in Panama, it is critical to
105 differentiate between imported infections or sustained local infection as the underlying cause of
106 persistent *P. vivax* incidence in Panama (4,5). Whole-genome sequencing can help paint a
107 detailed picture of parasite movement and transmission within and between countries (10,12,13).
108 However, *P. vivax* cannot be grown *in vitro*, and the difficulty of sequencing *P. vivax* from
109 clinical samples dominated by host DNA has hindered parasite population studies(14). Recent
110 advances such as hybrid capture (15) and selective whole genome amplification (SWGA)
111 mitigate this problem by allowing for parasite DNA to be selectively enriched before
112 sequencing(14). Both methods have allowed for population genomic studies of *P. vivax* using
113 samples directly from patients.
114 In this study, we describe the population genomics of *P. vivax in* Panama over a 12-year
115 time span, with the aim of understanding patterns of genetic variation and recent shared ancestry

116 (relatedness) at different geographical and temporal scales. We found the vast majority of *P.*
117 *vivax* cases in Panama belong to a single highly related lineage that has persisted for at least a
118 decade. Furthermore, we observed a second smaller clonal lineage concentrated near the
119 Panamanian-Colombian border. We also found several samples that shared no relatedness with
120 any other sample, which may represent either localized pockets of outbred *P. vivax* transmission
121 or imported cases. These findings suggest the Panamanian *P. vivax* population has extremely low
122 genetic diversity and is on the cusp of elimination. Revealing these patterns of relatedness among
123 parasite infections can help inform best strategies for targeting interventions or case investigation
124 methods to increase the likelihood of successful elimination. The results obtained from this study
125 will help inform future elimination efforts in Panama and the rest of Meso-America.
126

127 ## Results

128 **Sample Extraction, Sequencing, and Data Filtering**

129 We successfully generated usable sequencing data from 35/60 (58%) Panamanian *P. vivax*
130 samples collected between 2007-2009 and 24/40 (60%) collected between 2017-2019, for a total
131 of 59 samples (Fig 1A). All Panamanian samples had an F_{ws} statistic greater than 0.95, indicating
132 that they were all monoclonal (Fig 1B).

133 We next analyzed the 59 Panamanian genomes in the context of 109 previously
134 published *P. vivax* genomes, generating a filtered VCF consisting of 168 samples and 62,211
135 high-quality biallelic SNPs.

136 **Recent Common Ancestry Analysis Reveals Single Highly Related Lineage of Parasites**

137 We used hmmIBD (23) to estimate the proportion of the genome that is IBD among
138 Panamanian sample pairs to understand patterns of recent common ancestry. We subsetted the

139 VCF to contain only samples collected in Panama and Colombia to estimate pairwise IBD. We
140 strictly filtered sites on the basis of minimum and maximum read depth (five and thirty
141 respectively), resulting in a VCF with 89 samples and 15,788 sites for input into hmmIBD.
142 We observed a bimodal distribution of pairwise IBD in Panamanian samples, with peaks
143 near zero and 0.95 (Fig 2A). Forty-seven of the 59 Panamanian samples shared high IBD
144 (>0.875) with each other, indicating very recent common ancestry. Four other Panamanian
145 samples, all collected in the Kuna Yala Province, shared 100% IBD with each other, and 0-10%
146 IBD with any other sample, Panamanian or Colombian (Fig 2A and S1 Fig). Another four
147 Panamanian samples exhibited no IBD with each other nor any of the other Panamanian samples.
148 All four of these samples were collected in the Darien jungle region or Kuna Yala, which are the
149 two main points of entry for migrants traveling through Panama. These four samples drive the
150 modal peak of pairwise IBD at zero.
151 The variable degree of relatedness among the 47 samples sharing > 0.85 IBD suggested
152 that data quality potentially impacted the estimation of IBD. We plotted the relationship between
153 IBD and sample quality, measured by the average proportion of high coverage sites in each
154 sample pair, to determine if pairwise sample data quality affected the estimation of IBD (Fig
155 2B). We defined high coverage sites as sites with greater than 5x coverage (the cutoff for site
156 filtering). We observed that as the average proportion of high coverage sites for sample pairs
157 increased, pairwise IBD estimates correspondingly increased as well (Fig 2B). This relationship
158 suggests that poor data quality can lead to underestimation of IBD. It is possible that the majority
159 of the pairwise IBD estimations would be closer to one had the overall sample sequence quality
160 been higher. The prevalent highly genetically related lineage is referred to henceforth as cluster
161 one (CL1). CL1 samples share an IBD fraction of at least 0.875 with other samples in this

162 cluster. We also concluded that the four samples that shared 100% IBD with each other
163 constituted a second completely clonal lineage, henceforth referred to as cluster two (CL2).
164 Next, we examined how these two clusters and the other Panamanian samples not
165 belonging to either lineage were geographically distributed in Panama (Fig 3). Samples from
166 CL1 were found across Panama. Notably, samples collected from both 2007-2009 and 2017-
167 2019 were found in this lineage. The inclusion of samples from both collection periods
168 demonstrates that this lineage has persisted throughout Panama for at least a decade. We did not
169 find any evidence of structure in the *P. vivax* population by region or relative to the Panama
170 Canal, as was previously observed for *P. falciparum* (4).
171 We only observed samples belonging to CL2 in a specific locality, Puerto Obaldia, in the
172 Kuna Yala Amerindian territory (Comarca) along the Atlantic Coast. We lacked geographic
173 information for one of the four samples in CL2. Collection location data for the other three
174 samples show they were collected in roughly the same location, Puerto Obaldía in Kuna Yala.
175 Additionally, of four samples that shared no recent common ancestry with any sample in the
176 dataset, three were collected in Darien province, along the Colombian border and one was
177 collected in Kuna Yala.
178 After identifying two highly related lineages in Panama, we explored an approach for
179 determining whether the samples excluded from analysis due to low coverage could belong to
180 one of these lineages. We identified a set of 264 genotyped SNPs that were called in at least 80%
181 of samples across both Panama sample collection periods. We then calculated Nei's standard
182 genetic distance on all pairwise sample comparisons. The majority of excluded samples in both
183 collection periods (17/21 and 4/16 for the 2007-2009 and 2017-2019 collection periods
184 respectively) exhibited very low levels (0-1%) of genetic distance with CL1 samples (S2 Fig).

185 Seven of the 2017-2019 samples exhibited 0-1% genetic distance with CL1 also showed 0-1%
186 genetic distance with CL2, making the identity of these samples uncertain. Three excluded
187 samples from 2007-2009 collected in the Darien Jungle Region had relatively high genetic
188 distance from all other samples in the dataset.

189 **Exploring the regional context of Panamanian *P. vivax***

190 We built a neighbor-joining tree using the Panamanian samples plus previously
191 sequenced samples (20) (Fig 4A) to understand the Panamanian *P. vivax* population in a global
192 context. As noted in previous studies (20,30), we observed clusters of samples corresponding to
193 different geographic regions, with a large cluster of Central and South American samples. CL1
194 and CL2 formed distinct clusters within the Central and South American cluster with 100%
195 bootstrap support. CL2 is situated in a cluster containing samples from Colombia, with 100%
196 bootstrap support at deep nodes. While these four samples are clustered together with 100%
197 support and short branch length, a long branch connects them to the rest of the Colombian
198 cluster. The Panamanian samples that shared zero IBD with either cluster also grouped with the
199 Colombian samples. These samples seem to share distant ancestry with each other and the rest of
200 the Colombian samples. The samples also formed their own sub-cluster within the Colombian
201 cluster.

202 PCA conducted with all samples worldwide showed tight clustering of all Central and
203 South American samples, with only one Panamanian sample falling outside this Central and
204 South American cluster (S3 Fig). PCA restricted to the samples collected from Central and South
205 America is heavily influenced by covariation among samples within the two clusters (S4 Fig).
206 PCA performed with a single consensus sequence representing each cluster revealed CL1
207 clusters with samples from Peru and Brazil and CL2 clusters with the Colombian samples (Fig

208 4B). All four outbred Panamanian samples that shared no recent ancestry with the other samples
209 also clustered with the Colombian samples. Principal component one differentiated CL1 and
210 samples from Brazil and Peru from CL2 and the rest of the Central and South American samples.
211
212 **Genomic data both support and refute travel history data**

213 Four of the 59 samples had travel history data associated with them. All of these samples
214 with travel history data were collected during the 2007-2009 period. Travel history information
215 suggested that two samples were originally from Brazil, one sample from India, and one sample
216 from China. The two samples with Brazilian travel history fell within the Brazilian cluster on the
217 NJ tree and clustered with Brazilian samples on PCA (Fig 2A and 2B). The one sample with
218 Indian travel history grouped with the other Indian samples on the NJ tree, and clustered with the
219 other Indian samples via PCA as well (Fig 2A and S4 Fig). This sample was the only one
220 collected in Panama to fall outside of the Central and South American cluster in the PCA with
221 the worldwide sample set. For the two samples with Brazilian travel history and one sample with
222 Indian travel history, genomic data supported the same country of origin as the travel history
223 information.

224 The sample with Chinese travel history had a discrepancy between the region of origin
225 suggested by its travel history information and its genomic data. This sample clusters with the
226 Central and South American samples on the worldwide PCA instead of with the samples from
227 China. This sample clustered with the Colombian samples in the PCA conducted with only the
228 Central and South American samples (Fig 4B). Similarly, on the NJ tree, this sample fell within
229 the Colombian cluster with 100% bootstrap support along with the four Panamanian samples that
230 shared zero IBD with other Panamanian samples in the dataset.
231

232 **Discussion**

233 Panama is on the cusp of eliminating malaria after several decades of intervention (5).
234 We found extremely low clonal diversity in the Panamanian *P. vivax* population, observing that
235 the majority of the successfully sequenced samples (47/59) belonged to a single highly related
236 lineage, CL1. CL1 has persisted throughout Panama for at least a decade, in spite of ongoing
237 elimination efforts. Sample contamination could not explain this pattern as samples were
238 collected in two collection periods 10 years apart and extracted, amplified, and sequenced
239 separately. Our study suggests that the Panamanian *P. vivax* population has been through a
240 strong bottleneck due to reduced transmission, resulting in the majority of the population
241 belonging to a single highly related lineage. Several studies of the relationship between *P.*
242 *falciparum* genetic diversity and transmission intensity from Senegal (30), Thailand (31, 32),
243 Colombia (32), have documented that when there is a decline in parasite transmission, there is
244 an increase in the clonal composition of the population. Furthermore, there is evidence of
245 persistence and transmission of *P. falciparum* clonal lineages in Colombia (33), Ecuador (34),
246 and Haiti (35). This study demonstrates a similar relationship in Panama between low
247 transmission and extremely low genetic diversity of the *P. vivax* population. Almost all CL1
248 samples (46/47) share IBD > 0.95 with at least one other CL1 sample, suggesting a substantial
249 fraction of this population is clonal. Interestingly, the CL1 consensus sequence clusters with
250 samples from Brazil and Peru on PCA, suggesting a greater degree of *P. vivax* population
251 panmixia across Central and South America than previously understood, and that the population
252 has become more fractured over time (20). Alternatively, CL1 could have been imported into
253 Panama at an earlier point in time and has supplanted the ancestral *P. vivax* population in the
254 country, becoming the predominant lineage. Both scenarios point to the need for further

255 longitudinal genomic studies of *Plasmodium* parasites to better understand population dynamics
256 over space and time.

257 Previous studies have indicated that Panama has focal transmission in indigenous regions
258 (Comarcas) (5,6,11) Malaria transmission in Panama is increasingly concentrated in the
259 Comarcas, with the proportion of total malaria cases in Panama reported from Comarcas rising
260 from 41.8% in 2005 to 90% in 2019 (36). Prior work shows that low transmission can lead to
261 population structure in part due to variable malaria transmission (31). The finding that CL1 is
262 distributed ubiquitously throughout Panama is unexpected given the concentration of the malaria
263 epidemic within spatially separated regions of the country. The geographic distribution of CL1
264 suggests that parasites have historically moved throughout the country, founding new
265 populations or supplanting small existing ones. Case investigations and understanding human
266 movement patterns throughout Panama will be critical to achieving elimination.

267 This study had some limitations. We were unable to generate high-quality sequencing
268 data from ~40% of the samples. Fortunately, we were able to extract a small subset of 264
269 genotypable SNPs from the excluded samples. We used this SNP set to calculate Nei's standard
270 genetic distance to determine if the excluded samples were genetically distant from CL1 or CL2.
271 We found that ~57% of the excluded samples were genetically similar to CL1 (S2 Fig). This
272 result indicates that our assessment of relatedness within the Panamanian *P. vivax* population is
273 not biased by parasitemia or other factors that could have affected sequencing success. This
274 finding also suggests that we did not miss additional genetically distinct circulating Panamanian
275 *P. vivax* strains by excluding these samples. Seven samples from 2017-2019 exhibited low
276 genetic distance from both CL1 and CL2. However, these seven samples were missing ~60-80%
277 of calls even after stringent filtration. The high missing rate in these seven samples potentially

278 explains the low genetic distance to most other samples in the dataset. Additionally, all samples
279 that did yield usable sequence data represented almost all localities across Panama. The
280 exception to this was a group of samples collected near the Panamanian-Costa-Rican border that
281 we were unable to generate usable sequence data from. However, most ongoing malaria
282 transmission in Panama occurs East of the Panama Canal, where most of the samples that
283 generated usable sequencing data were collected (6,11). Due to the geographic sampling
284 coverage of regions with ongoing malaria transmission, we believe these data are reflective of
285 the current state of the Panamanian *P. vivax* population. We also lacked geographic collection
286 data for two of the successfully sequenced samples, and they were excluded from the geographic
287 analysis. The lack of geographical data is unlikely to bias our conclusions since these samples
288 came from both different collection periods and regions. The two samples also constitute a small
289 proportion of samples in the final dataset.

290 Genomic epidemiology can help to support malaria elimination efforts in Panama in
291 multiple ways. First, genomic data can help identify genetically distinct cases that may be
292 imported. Panama sits at the crux of migration paths to the United States, and it is possible that
293 genetically distinct samples collected in Panama represent imported cases. Integrating travel
294 history information with genomic data can help solidify the identification of imported cases.
295 Four samples had patient travel history information, and genomic data supported the presumed
296 country of origin for three of them. The fourth sample was collected from a subject with travel
297 history from China. However, it clustered with Colombian samples on the NJ tree and the PCA,
298 suggesting the infection was likely acquired somewhere in Central or South America, rather than
299 China.

300 Second, genomic data will be critical to determine if imported parasites are contributing
301 to local transmission and/or admixing with the local parasite population. For example, we did not
302 observe evidence of admixture between the imported samples from India and Brazil and the
303 samples that comprise CL1, or evidence of onward clonal transmission of the imported samples.
304 Our data cannot distinguish whether CL2 is a native Panamanian parasite lineage or if it has been
305 imported from Colombia. However, the four CL2 samples displayed a genomic and
306 epidemiological pattern consistent with recent local transmission, as all CL2 samples are
307 virtually identical and were collected from the same municipality in 2019.
308 Overall, the existence of one main parasite genetic lineage exhibiting no recent evidence
309 of outcrossing with imported infections suggests that Panama is ripe for the elimination of *P.*
310 *vivax*. While case importation remains a threat, the lack of evidence of outcrossing suggests it
311 may not be sufficient to prevent elimination under present circumstances. Ongoing genomic
312 surveillance and case containment efforts will be needed, however, to mitigate the risk of
313 outbreaks resulting from imported cases and reversal of the impressive progress that has been
314 recently made towards malaria elimination in Panama.
315

316 ## Materials and Methods

317 **Sample collection**

318 We collected 96 *P. vivax* samples from infected consenting volunteers identified through passive
319 or active surveillance by technicians from the National Department of Vector Control from the
320 Ministry of Health (MINSA) of Panama. The Gorgas Memorial Institute of Health Studies
321 Bioethics Committee gave the study ethical approval (Permits: DIDETEC/06/ET-PA/02 and
322 154/CBI/ICGES/17). Two groups of DNA samples from infected patients were used in this

323 study: 1) 56 DNA samples collected during 2007-2009 and 2) 40 DNA samples collected during
324 2017-2019. The Gorgas Memorial Institutional Animal Use and Care Committee (CIUCAL)
325 (Permit: 002/CIUCAL-ICCES-2012) approved the use of Aotus *P. vivax* AMRU-1 and SAL-1
326 infected blood samples as a source of control DNA. Patient blood samples were collected by
327 finger-prick with a lancet and spotted into EBF 903 Five Spot Blood Cards (Eastern Business
328 Forms, INC, SC, USA). The samples were then transported at ambient temperature to the
329 laboratory and stored at –20℃ until processing. We also created thin and thick blood smears
330 from patient samples. The blood smears were stained with Giemsa for percent parasite density
331 determination, species identification, and stage differential counts. Each volunteer donated ~150
332 µL of blood.

333 **Information survey**

334 We collected demographic, geographic, socioeconomic, and epidemiological info from each
335 study subject using an epidemiological form developed for the Survey123 for ArcGIS online
336 survey program (Esri, Redlands, CA).

337 **Malaria microscopy**

338 Giemsa stained thick and thin blood smears were examined by light microscopy for parasite
339 density determinations, P*lasmodium* species confirmation, and parasite lifecycle stage
340 count. Parasite densities were calculated by quantifying the number of malaria-infected red
341 blood cells (iRBCs) among 500 – 2000 RBCs on a thin blood smear and expressing the result as
342 % parasitemia (% parasitemia = parasitized RBCs /total RBCs) x 100).

343 **DNA extraction**:

344 We extracted DNA from the filter paper blood spots using the Chelex method as described (16)
345 for samples obtained during 2007-2009 and with the Qiagen® DNA mini kit for samples
346 obtained during 2017-2019.

347 **Molecular confirmation of *P. vivax* infection**

348 We confirmed *P. viva*x infection for all samples collected during 2017-2019 by amplification of
349 the *P. vivax* PVX_18SrRNA gene using a qRT-PCR assay as described (17).

350 **Selective whole genome amplification and sequencing**

351 We carried out DNA pre-amplification as described (14). Briefly, the thermocycler was
352 preheated to 35℃. We dispensed aliquots of 37µl of Power SYBR Green Master Mix, plus 3µl
353 phi39 into each PCR tube, next adding DNA and water to achieve a final volume of 47µl.
354 Thermocycler settings were as follows: 35℃ x 10 min; 34℃ x 10 min; 33℃ x 10 min; 32℃ x 10
355 min; 31℃ x 10 min 30℃ x 16 hours, 65℃ x 10 min, 4 ℃ for infinity. SWGA reaction products
356 were diluted with 50 µl of water. We purified 50 µl of the diluted product using 50 µl AmPURE
357 beads according to the instructions of the manufacturer. We then eluted beads in 30 µl of water.
358 Approximately 60-120 ng/µl of DNA was obtained after bead purification of the SWGA
359 reaction. We measured DNA concentration using Nanodrop quantitation.

360 **Whole-genome sequencing**

361 We performed whole-genome sequencing (WGS) on all 96 *P. vivax* samples using Nextera
362 libraries and an Illumina HiSeq X platform. Sample reads were aligned to the P01 reference
363 genome assembly using BWA-MEM, version 0.7 (18). Illumina sequencing reads are available
364 through the NCBI Sequence Read Archive with BioProject accession numbers
365 SAMN15722613–SAMN15722671.

366 **SNP Discovery and Quality Filtering**

367 We marked duplicate reads using the MarkDuplicates tool from Picard tools. We next
368 performed local realignment around indels using the Genome Analysis Toolkit (GATK)
369 RealignerTargetCreator and GATK IndelRealigner (GATK Version 3.5.0). We called variants
370 using GATK HaplotypeCaller using best practices to call and filter SNPs and generate gVCFs
371 for each sample. We called samples in two batches, one containing samples collected in 2007-
372 2009 and one containing samples collected in 2017-2019. We performed joint variant calling on
373 the sets separately using GATK GenotypeGVCFs tool with GATK hard filters, including calls in
374 subtelomeric regions. The resulting VCFs consisted of 56 samples and 407,554 sites for the
375 2007-2009 samples, and 40 samples and 171,433 variants for the 2017-2019 samples. We
376 retained samples for analysis if they exhibited a minimum mean read depth of five and had calls
377 at more than 80% of variant sites in the VCF corresponding to their collection period, including
378 those in subtelomeric regions. We calculated and evaluated data quality measures using the
379 VCFtools package and custom R scripts (19). Thirty-five samples from 2007-2009 and 24
380 samples from 2017-2019 passed these filters and were kept for further analysis.
381 We next used GenotypeGVCFs tool to construct a joint VCF with the 59 Panamanian
382 samples plus a collection of previously collected global samples (Bioproject numbers
383 PRJNA240356-PRJNA240533 (20)). The joint VCF contained 168 samples and 2,4250,245
384 variants. We filtered sites on the basis of quality (GQ > 40), passing VQSR truth sensitivity level
385 of 0.99 or greater, missing rate (having a call at that site in > 85% of samples). We also
386 excluded any sites that were not bi-allelic and indels. The joint VCF generated after filtering
387 contained 168 samples and 62,211 sites.
388 Lastly, we generated a VCF containing SNPs found jointly in 80% of both the 2007-2009
389 and 2017-2019 samples. We also filtered sites in this VCF by excluding non-biallelic SNPs and

390 on the basis of quality (GQ > 30), and passing GATK filters. The resultant VCFs contained 56
391 samples and 2335 SNPs for the 2007-2009 samples, and 40 samples and 1301 SNPs for the
392 2017-2019 samples. For these samples, we generated a highly filtered variant set containing
393 biallelic SNPs that passed the GATK filters (GQ > 30, truth sensitivity level > 0.99, Mean DP >
394 five) and were called in at least 80% of the samples from both time periods (2007-2009; 2017-
395 2019). Calls from the two sample sets were merged to create a unified VCF of 96 samples and
396 264 genotyped SNPs.

397 **Determination of Sample Clonality**

398 We estimated sample clonality using the F_{ws} statistic. F_{ws} measures the within-sample
399 genetic diversity (measured by heterozygosity H_w) relative to the overall population genetic
400 diversity (H_s) (21). The underlying theory assumes that a monoclonal (single strain) infection has
401 extremely low genetic diversity relative to overall population genetic diversity. By contrast, a
402 polyclonal (multiple strain) infection has high diversity relative to overall population diversity
403 (compared to a monoclonal infection). By estimating the ratio between within-host diversity and
404 population diversity, we can distinguish between monoclonal and polyclonal infections (21). A
405 sample with an F_{ws} statistic of 0.95 or greater (>=0.95) is considered monoclonal. We calculated
406 F_{ws} using the R package moimix (22).

407 **Analysis of Recent Common Ancestry**

408 We used hmmIBD (23) to estimate the proportion of sites identical by descent (IBD) between
409 sample pairs to ascertain recent common ancestry among Panamanian and Colombian samples
410 collected previously from the global *P. vivax* population study (20). We estimated minor allele
411 frequency (MAF) for IBD inference using the genetically distinct Panamanian samples, a
412 representative sample from each of the two highly related Panamanian clusters, and the

413 Colombian samples. We included Colombian samples to improve MAF estimation given the
414 greater clonal diversity of the Colombian parasite population and presumed historical gene flow
415 with Panama. We subsetted the master VCF file to keep only samples collected in Panama and
416 Colombia. Sites were excluded sites on the basis of minimum and maximum read depth (five and
417 thirty respectively) to ensure that we were using only high-quality SNPs. The input VCF for
418 hmmIBD contained 89 samples (59 Panamanian samples and 30 Colombian samples) and 15,788
419 variant sites. We then re-formatted the data using a custom perl script for input into hmmIBD
420 along with the MAF estimates. We conducted analysis and visualization of the hmmIBD output
421 using custom R scripts.

422 **Analysis of Population Structure**

423 We employed principal components analysis (PCA) and a neighbor-joining tree to study
424 the population structure of Panamanian samples in the context of the worldwide *P. vivax*
425 population (20). We used a strictly filtered SNP set for PCA, keeping only variants with calls in
426 at least 95% of samples. This input VCF consisted of 168 samples and 2,428 variants. We used
427 the R package SNPRelate to conduct PCA (24). Covariation within the two clusters heavily
428 influenced the PCA of all samples, so we also performed PCA using a single consensus sequence
429 for each cluster.

430 We used the R packages ape, StAMPP, pegas, and adegenet, (25–28) to generate the
431 neighbor-joining tree and genetic distance statistics. First, we calculated Nei's distance for all
432 pairwise sample combinations using the master VCF consisting of 168 samples and 62,211 sites
433 to generate a distance matrix. The distance matrix was used to generate a tree. We used the
434 bootphylo function in the ape package to bootstrap the dataset 100 times to estimate nodal
435 support. We then visualized the final tree with support values using the FigTree program

(29). We used R software (R version 3.6.1) to carry out statistical analysis and data visualization.

Acknowledgments

The authors thank Selina Bopp and Becky Kuzma at HSPH for conducting SWGA of the *P. vivax* samples. We also thank at ICGES, Néstor Sosa, Director, José Calzada, Azael Saldaña and Gladys Tuñón, for their administrative support; also at ICGES Marlon Núñez, Ariel Magallón, and José C. Marín for their technical and logistic support; at the University of California-San Diego, Elizabeth Winzeler and Annie Cowell for their technical advice on establishing SWGA for *P. vivax*. Department of Research and Development (I+D), Secretary of Science, Technology, and Innovation (SENACYT) of the Government of the Republic of Panamá, Iriela Aguilar, and Milagros Mainieri for their administrative and technical support with project management.

References

1. World Health Organization. World Malaria Report 2019. Vol. 1. WHO; 2019.
2. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. Am J Trop Med Hyg. 2007 Dec;77(6 Suppl):79–87.
3. Arevalo-Herrera M, Quiñones ML, Guerra C, Céspedes N, Giron S, Ahumada M, et al. Malaria in selected non-Amazonian countries of Latin America. Acta Trop. 2012 Mar;121(3):303–14.
4. Obaldia N, Baro NK, Calzada JE, Santamaria AM, Daniels R, Wong W, et al. Clonal outbreak of Plasmodium falciparum infection in eastern Panama. J Infect Dis. 2015 Apr 1;211(7):1087–96.
5. Carrera LC, Victoria C, Ramirez JL, Jackman C, Calzada JE, Torres R. Study of the epidemiological behavior of malaria in the Darien Region, Panama. 2015–2017. PLOS ONE. 2019 Nov 15;14(11):e0224508.
6. Calzada JE, Marquez R, Rigg C, Victoria C, De La Cruz M, Chaves LF, et al. Characterization of a recent malaria outbreak in the autonomous indigenous region of Guna Yala, Panama. Malar J. 2015 Nov 17;14:459.
7. World Health Organization. Panama Malaria Country Profile 2017. WHO Malaria Country Profiles [Internet]. 2017;1(1). Available from: https://www.who.int/malaria/publications/country-profiles/profile_pan_en.pdf?ua=1
8. Ruktanonchai NW, DeLeenheer P, Tatem AJ, Alegana VA, Caughlin TT, Zu Erbach-Schoenberg E, et al. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. PLoS Comput Biol. 2016 Apr;12(4):e1004846.
9. Marshall JM, Touré M, Ouédraogo AL, Ndhlovu M, Kiware SS, Rezai A, et al. Key traveller groups of relevance to spatial malaria transmission: a survey of movement patterns in four sub-Saharan African countries. Malar J. 2016 Apr 12;15:200.
10. Wesolowski A, Taylor AR, Chang H-H, Verity R, Tessema S, Bailey JA, et al. Mapping malaria by combining parasite genomic and epidemiologic data. BMC Medicine. 2018 Oct 18;16(1):190.
11. Lainhart W, Dutari LC, Rovira JR, Sucupira IMC, Póvoa MM, Conn JE, et al. Epidemic and Non-Epidemic Hot Spots of Malaria Transmission Occur in Indigenous Comarcas of Panama. PLoS Negl Trop Dis . 2016 May 16;10(5). doi: 10.1371/journal.pntd.0004718
12. José Arcia. Darién, convertido en un río imparable de migrantes irregulares [Internet]. La Estrella de Panamá. 2019 [cited 2020 Apr 29]. Available from: https://www.laestrella.com.pa/nacional/190709/rio-darien-imparable-migrantes-convertido
13. Neafsey DE, Volkman SK. Malaria Genomics in the Era of Eradication. Cold Spring Harb Perspect Med. 2017 Aug;7(8):a025544.
14. Cowell AN, Loy DE, Sundararaman SA, Valdivia H, Fisch K, Lescano AG, et al. Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. mBio. 2017 07;8(1).
15. Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU, et al. Whole genome (1)sequencing analysis of Plasmodium vivax using whole genome capture. BMC Genomics. 2012 Jun 21;13(1):262.
16. Wooden J, Kyes S, Sibley CH. PCR and strain identification in Plasmodium falciparum. Parasitol Today (Regul Ed). 1993 Aug;9(8):303–5.
17. Obaldia N, Meibalan E, Sa JM, Ma S, Clark MA, Mejia P, et al. Bone Marrow Is a Major Parasite Reservoir in Plasmodium vivax Infection. mBio. 2018 08;9(3).

496 18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
497 Bioinformatics. 2009 Jul 15;25(14):1754–60.
498 19. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call
499 format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.
500 20. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, et al. Population
501 genomics studies identify signatures of global dispersal and drug resistance in Plasmodium
502 vivax. Nat Genet. 2016;48(8):953–8.
503 21. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization
504 of Within-Host Plasmodium falciparum Diversity Using Next-Generation Sequence Data.
505 PLOS ONE. 2012 Feb 29;7(2):e32891.
506 22. Lee S, Harrison A, Tessier N, Tavul L, Miotto O, Siba P, Kwiatkowski D, Müller I, Barry
507 AE and Bahlo M. Assessing clonality in malaria parasites using massively parallel
508 sequencing data. in preparation. 2016;
509 23. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer
510 pairwise identity by descent between haploid genotypes. Malar J. 2018 May 15;17(1):196.
511 24. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance
512 computing toolset for relatedness and principal component analysis of SNP data.
513 Bioinformatics. 2012 Dec 1;28(24):3326–8.
514 25. Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic
515 differentiation and structure of mixed-ploidy level populations. Molecular Ecology
516 Resources. 2013;13(5):946–52.
517 26. Paradis E. pegas: an R package for population genetics with an integrated-modular
518 approach. Bioinformatics. 2010 Feb 1;26(3):419–20.
519 27. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
520 analyses in R. Bioinformatics. 2019 Feb 1;35(3):526–8.
521 28. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.
522 Bioinformatics. 2011 Nov 1;27(21):3070–1.
523 29. Rambaut A, Drummond AJ. FigTree version 1.4. 0. 2012.
524 30. Rougeron V, Elguero E, Arnathau C, Acuña Hidalgo B, Durand P, Houze S, et al. Human
525 Plasmodium vivax diversity, population structure and evolutionary origin. PLoS Negl Trop
526 Dis . 2020 Mar 9 ;14(3). doi: 10.1371/journal.pntd.0008072
527 31. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic
528 correlates of declining transmission in a human pathogen. Mol Ecol. 2013 Jan;22(2):273–
529 85.
530 32. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis
531 of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012
532 Jul 19;487(7407):375–9.
533 33. Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ. Long term persistence
534 of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region.
535 BMC Genetics. 2013 Jan 7;14(1):2.
536 34. Sáenz FE, Morton LC, Okoth SA, Valenzuela G, Vera-Arias CA, Vélez-Álvarez E, et al.
537 Clonal population expansion in an outbreak of Plasmodium falciparum on the northwest
538 coast of Ecuador. Malar J. 2015 Dec 10;14. doi: 10.1186/s12936-015-1019-2
539 35. Charles M, Das S, Daniels R, Kirkman L, Delva GG, Destine R, et al. Plasmodium
540 falciparum K76T pfcrt Gene Mutations and Parasite Population Structure, Haiti, 2006–
541 2009. Emerg Infect Dis. 2016 May;22(5):786–93.

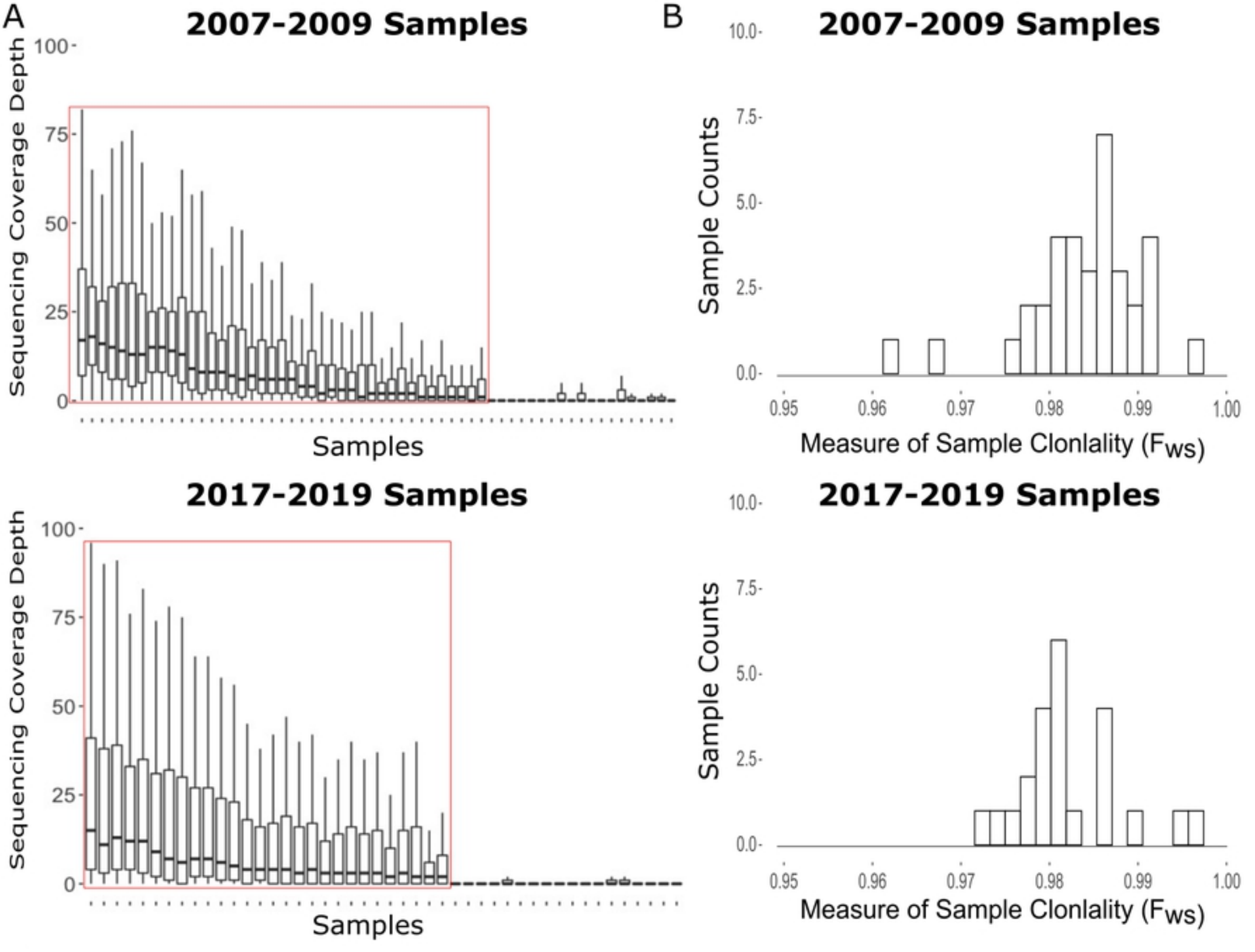
36. Hurtado L, Cumbrera A, Rigg C, Perea M, Santamaría AM, Chaves LF, et al. Long-term transmission patterns and public health policies leading to malaria elimination in Panamá. Malar J. 2020 Jul 23 ;19. doi: 10.1186/s12936-020-03329-y

Fig 1. Sequencing and Sample Assessment at Variant Sites. A) Distribution of variant site read coverage for each sample stratified by the collection period. Coverage values > 100 were censored for visualization purposes. Samples within the red boxes were kept for analysis. B) Distribution of F_{ws} values for all samples, stratified by the collection period. We interpreted F_{ws} values > 0.95 as evidence of sample monoclonality.

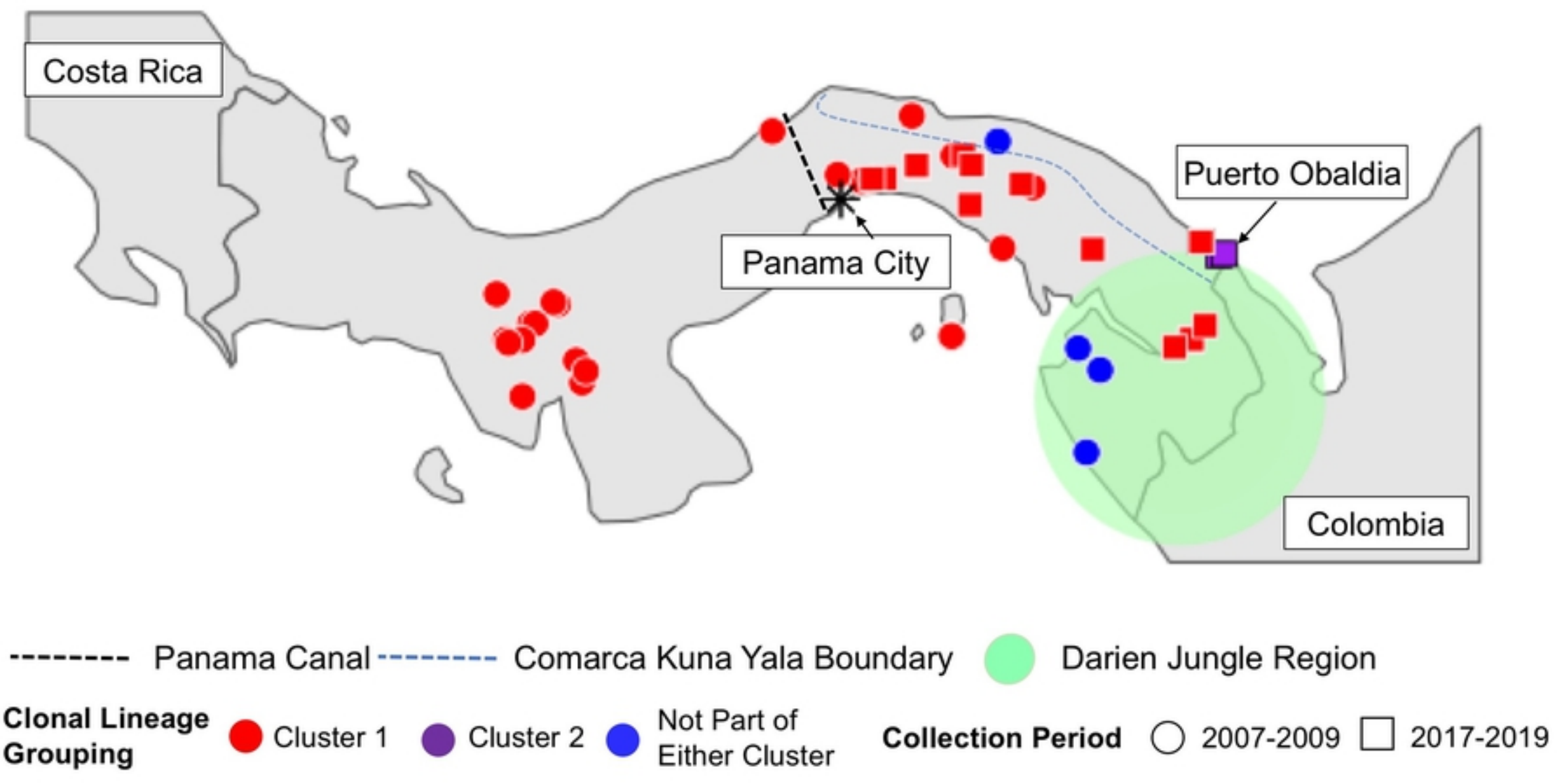
Fig 2. IBD analysis of the Panamanian samples. A) The distribution of pairwise IBD estimates among the Panamanian samples. IBD values near zero indicate no recent common ancestry. Values closer to one indicate that the sample pair are clonal or essentially clonal. B) Depicts pairwise IBD estimates for all Panamanian sample pairs with IBD > 0.875 plotted against the mean proportion of high coverage sites (sites with > 5x coverage) in each sample pair. The line indicates a linear regression, the box displays the Pearson correlation coefficient between the two axes variables.

Fig 3. Map of Panamanian sample collection sites. Sample colors show which cluster (or neither) each sample belongs too. Shape indicates the sample collection period. The dotted line shows the location of the Panama Canal. The Blue Line shows the border of the Comarca Kuna Yala. The Darien Jungle Region is indicated by the green shaded area.

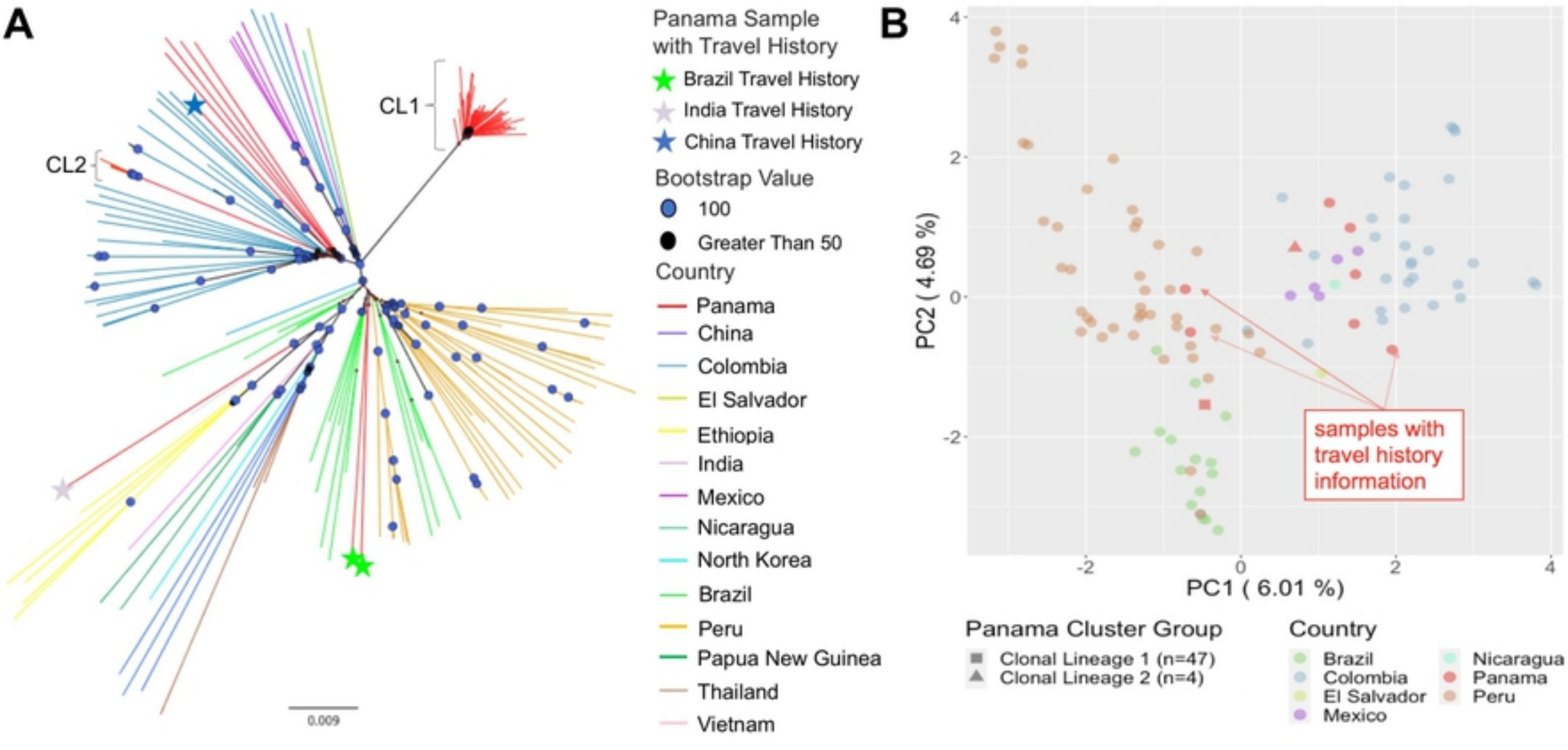
Fig 4. Population Structure. A) Neighbor-joining tree for all samples worldwide. Node symbols denote support values: circles indicate 100% support, triangles indicate > 50% support. Branch color indicates the country of collection for each sample. Panamanian samples with travel history are noted with the colored stars. B) PCA of Central and South American samples. Circle color indicates country of collection. Consensus sequences for cluster one and cluster two are noted as a square and triangle respectively. Panamanian samples with travel history are annotated.


A
2007-2009 Samples
Sequencing Coverage Depth
100
75
50
25
0
Samples
B
2007-2009 Samples
Sample Counts
10.0
7.5
5.0
2.5
0.0
0.95
0.96
0.97
0.98
0.99
1.00
Measure of Sample Clonlality (F_{WS})
2017-2019 Samples
Sequencing Coverage Depth
100
75
50
25
0
Samples
2017-2019 Samples
Sample Counts
10.0
7.5
5.0
2.5
0.0
0.95
0.96
0.97
0.98
0.99
1.00
Measure of Sample Clonlality (F_{WS})

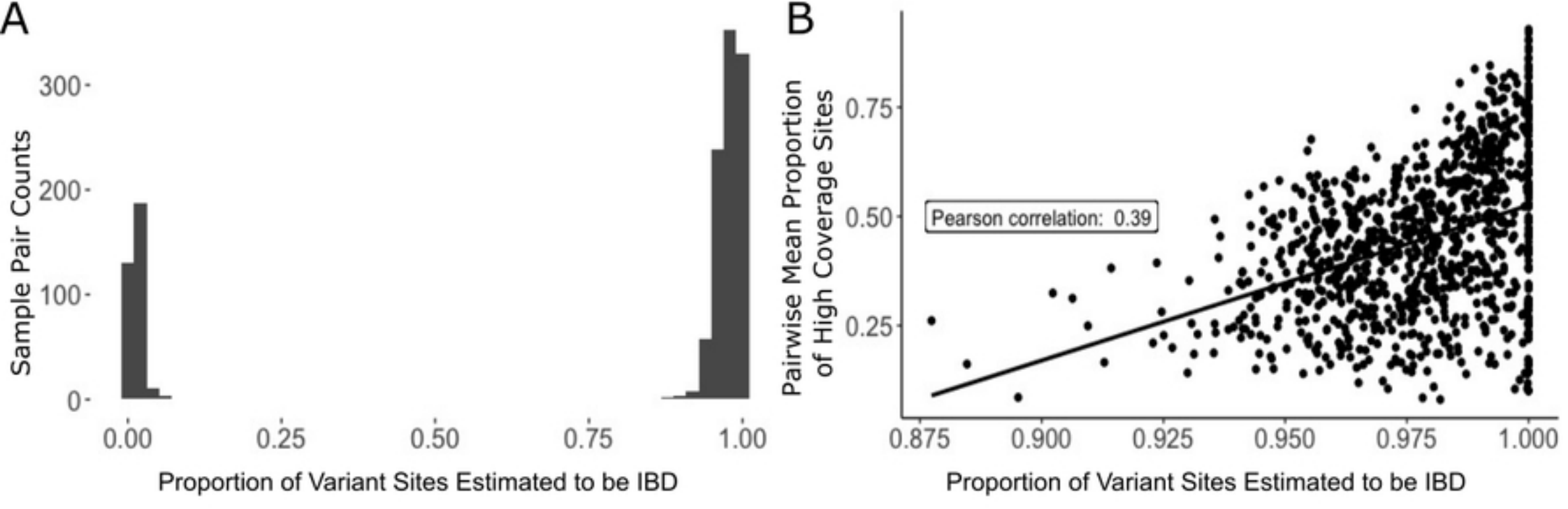
Figure


Costa Rica
Panama City
Puerto Obaldia
Colombia
Panama Canal
Comarca Kuna Yala Boundary
Darien Jungle Region
Clonal Lineage Grouping
Cluster 1
Cluster 2
Not Part of Either Cluster
Collection Period
2007-2009
2017-2019

Figure


A
CL1
CL2
0.009
Panama Sample with Travel History
Brazil Travel History
India Travel History
China Travel History
Bootstrap Value
100
Greater Than 50
Country
Panama
China
Colombia
El Salvador
Ethiopia
India
Mexico
Nicaragua
North Korea
Brazil
Peru
Papua New Guinea
Thailand
Vietnam
B
PC2 (4.69 %)
PC1 (6.01 %)
samples with travel history information
Panama Cluster Group
Clonal Lineage 1 (n=47)
Clonal Lineage 2 (n=4)
Country
Brazil
Colombia
El Salvador
Mexico
Nicaragua
Panama
Peru

Figure


A
300
200
100
0
Sample Pair Counts
0.00
0.25
0.50
0.75
1.00
Proportion of Variant Sites Estimated to be IBD
B
0.75
0.50
0.25
Pairwise Mean Proportion of High Coverage Sites
Pearson correlation: 0.39
0.875
0.900
0.925
0.950
0.975
1.000
Proportion of Variant Sites Estimated to be IBD

Figure