**Title: Allelic bias when performing in-solution enrichment of ancient human DNA**

**Running title: Allelic bias in ancient DNA enrichment**

**Authors:** Roberta Davidson[1#], Matthew P. Williams[1,2], Xavier Roca-Rada[1], Kalina Kassadjikova[3], Raymond Tobler[1,4,5], Lars Fehren-Schmitz[3,6#^], Bastien Llamas[1,5,7,8#^]

# Corresponding authors: roberta.davidson@adelaide.edu.au, bastien.llamas@adelaide.edu.au, lfehrens@ucsc.edu

^ Equal contributor

**Affiliations**

1. The Australian Centre for Ancient DNA and the Environment Institute, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
2. Biology Department, The Pennsylvania State University, University Park, Pennsylvania, USA
3. UCSC Paleogenomics, Department of Anthropology, University of California, Santa Cruz, California, USA
4. Evolution of Cultural Diversity Initiative, Australian National University, Canberra, ACT 2601, Australia
5. Centre of Excellence for Australian Biodiversity and Heritage, The University of Adelaide, Adelaide, SA 5005, Australia
6. UCSC Genomics Institute, University of California, Santa Cruz, California, USA
7. National Centre for Indigenous Genomics, Australian National University, Canberra, ACT 0200, Australia

8. Indigenous Genomics, Telethon Kids Institute, Adelaide, SA 5000, Australia

## Abstract

In-solution hybridisation enrichment of genetic variation is a valuable methodology in human paleogenomics. It allows enrichment of endogenous DNA by targeting genetic markers that are comparable between sequencing libraries. Many studies have used the 1240k reagent—which enriches 1,237,207 genome-wide SNPs—since 2015, though access was restricted. In 2021, Twist Biosciences and Daicel Arbor Biosciences independently released commercial kits that enabled all researchers to perform enrichments for the same 1240k SNPs. We used the Daicel Arbor Biosciences Prime Plus kit to enrich 132 ancient samples from three continents. We identified a systematic assay bias that increases genetic similarity between enriched samples and that cannot be explained by batch effects. We present the impact of the bias on population genetics inferences (e.g., Principal Components Analysis, $f$-statistics) and genetic relatedness (READ). We compare the Prime Plus bias to that previously reported of the legacy 1240k enrichment assay. In $f$-statistics, we find that all Prime-Plus-generated data exhibit artefactual excess shared drift, such that within-continent relationships cannot be correctly determined. The bias is more subtle in READ, though interpretation of the results can still be misleading in specific contexts. We expect the bias may affect analyses we have not yet tested. Our observations support previously reported concerns for the integration of different data types in paleogenomics. We also caution that technological solutions to generate 1240k data necessitate a thorough validation process before their adoption in the paleogenomic community.

**Keywords**

Bias, enrichment, 1240k, ancient DNA

## INTRODUCTION

One of the major challenges faced when working with ancient DNA is the high proportion of exogenous DNA contamination present in the DNA extract. This contamination is mostly from microbes that invade the organism post-mortem or are present in the soil where the specimen was buried. It can also be introduced during laboratory experiments. To address this challenge, one method that has become popular in paleogenomic studies is the in-solution enrichment of target genomic regions using pre-designed oligonucleotides as molecular “probes” or “baits”. This technique increases the proportion of endogenous DNA of interest in a sequencing library, reducing sequencing cost and leading to larger quantities of comparable data across individual samples compared to using shotgun sequencing. Ultimately, the increase in analytical power resulting from target enrichment has enabled large-scale ancient human population genetic (i.e., paleogenomic) studies (Fumagalli, 2013).

In human paleogenomic studies, in-solution enrichment assays can be designed to target complete mitochondrial genomes (Briggs et al., 2009; Brotherton et al., 2013; Llamas et al., 2016; Maricic et al., 2010), the non-recombining portion of chromosome Y (Cruz-Dávalos

et al., 2018; Rohrlach et al., 2021), genome-wide SNPs (Fu et al., 2015; Haak et al., 2015; Mathieson et al., 2015), exomes (Castellano et al., 2014), autosomes (Fu et al., 2013), or complete genomes (Ávila-Arcos et al., 2015; Carpenter et al., 2013). In 2012, Patterson and colleagues proposed a molecular bait design that used a specific ascertainment strategy to enable unbiased population genetics studies of global human populations through time (Patterson et al., 2012). The first baits using this design were applied in a handful of landmark paleogenomic studies of European populations (Fu et al., 2015; Haak et al., 2015; Mathieson et al., 2015). These bait sets were subsequently restricted to 1240k genome-wide SNPs, which paved the way for the generation of thousands of individual datasets and launched the paleogenomic era (Marciniak & Perry, 2017; Olalde & Posth, 2020; Skoglund & Mathieson, 2018). Hereafter, we will refer to this set of baits as the 'legacy 1240k reagent'.

Since the initial publication of the molecular baits sequences in 2015, the legacy 1240k reagent was only available to a handful of laboratories through a commercial agreement. Subsequently, in 2021 two biotechnology companies—Daicel Arbor Biosciences and Twist Bioscience—released commercial in-solution enrichment kits that target the same set of 1240k SNPs (Rohland et al., 2022). In particular, the release of the suite of myBaits Expert Human Affinities kits by Daicel Arbor Biosciences aimed to make the 1240k enrichment assay accessible to all research groups worldwide, including those with limited human resources and funding. Importantly, the molecular design of both the new bait sets differs

from the legacy 1240k design, with the differences specific to the Twist reagents described in a recently published study (Rohland et al., 2022). While the bait design used in the myBaits Expert Human Affinities kits is also known to differ from both the legacy 1240k and Twist assays (Daicel Arbor Biosciences, personal communication), it is a proprietary product and the specific details are unknown to the present authors. The bait design and associated laboratory protocol steps of any enrichment reagent are key to its efficacy to enrich endogenous DNA and to achieve an unbiased enrichment that respects the true proportions of reference and alternate alleles at heterozygous sites.

Here, we report genomic data from 132 ancient human individuals from western Eurasia, South and Central America (from periods preceding the colonial contact period between these populations), which were generated using myBaits Expert Human Affinities Kit "Prime Plus" (Daicel Arbor Biosciences) for various independent studies. In addition to the standard set of 1240k SNPs, the Prime Plus (PP) probes target 46,218 Y Chromosome SNPs, as well as human, Neanderthal and Denisovan whole mitochondrial genomes. Notably, when performing routine population genetic analyses on the PP-generated dataset using $f$-statistics (Patterson et al., 2012), we observed a significant excess of shared genetic drift between any two populations enriched with the PP bait set, compared with published data generated with the legacy 1240k reagent or shotgun sequencing. These unexpected affinities could not be explained by batch effects or previously reported demographic events.

In the following sections, we outline our investigation and characterisation of the bias observed in PP-enriched ancient DNA data via principal components analyses, genetic relatedness analyses and $f$-statistics. In total, we attempted 20 strategies to reduce or negate the observed bias, which included two wet-lab protocol adjustments and 18 bioinformatic techniques, but we were unable to correct the $f$-statistics scores to a satisfactory standard. We conclude by recommending a series of principles for researchers to apply when co-analysing PP-enriched data with 1240k legacy and shotgun (SG) data, which mitigate the assay bias by adjusting the construction of $f$-statistics. These guidelines may also be useful for reviewers and general readers when assessing the robustness of studies that have used data generated with the myBaits Expert Human Affinities Kit "Prime Plus".

## MATERIALS AND METHODS

The data used in this study were initially generated for various studies intended by the authors. However, we collated all independent datasets into one overarching dataset after identifying some problems with Prime Plus data, resulting in the present study.

### Samples

We generated PP-enriched data for a total of 132 ancient samples, sourced from Spain, Mesoamerica, Peru, Bolivia and Chile. All samples from the Americas pre-date 16th-century European contact (see Table S2 for sample metadata). In our analyses, we use a

comparative dataset of 115 shotgun data samples and 428 legacy 1240k samples (Bergström et al., 2020; Bongers et al., 2020; de la Fuente et al., 2018; González-Fortes et al., 2019; Günther et al., 2015; Lindo et al., 2018; Lipson et al., 2017; Mallick et al., 2016; Martiniano et al., 2017; Mathieson et al., 2015; Moreno-Mayar et al., 2018; Nakatsuka, Lazaridis, et al., 2020; Nakatsuka, Luisi, et al., 2020; Olalde et al., 2018, 2019; Popović et al., 2021; Posth et al., 2018; Prüfer et al., 2014; Raghavan et al., 2015; Salazar et al., 2023; Valdiosera et al., 2018; Villalba-Mouco et al., 2019). A complete list of the comparative samples used in this study and relevant citations are provided in Table S3.

## Data generation

### *DNA extraction*

For the samples from Spain, DNA was extracted from tooth root powder using a method optimised to retrieve highly degraded ancient DNA fragments (Dabney et al., 2013) at the Autonomous University of Barcelona (UAB). For some Mesoamerican samples, the same methods were undertaken at the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide. For the rest of the samples from the Americas, DNA was extracted similarly, but with a modification to the protocol that includes a bleach pre-wash step (Boessenkool et al., 2017; Dabney et al., 2013), at the University of California, Santa Cruz (UCSC) Human Paleogenomics Lab. Table S2 provides a full summary of which lab processed which samples at each step.

151 *Library preparation*

152 For the Spanish and some Mesoamerican samples, partially UDG-treated double-stranded
153 DNA libraries (Rohland et al., 2015) were generated at ACAD. For the Bolivian, Peruvian,
154 Chilean and the rest of the Mesoamerican samples, partially UDG-treated single-stranded
155 DNA libraries (Kapp et al., 2021) were prepared at UCSC. Note that the difference in
156 library structures should not impact the enrichment bias as libraries are first denatured
157 and then hybridised to the baits.

158

159 *Library enrichment*

160 Before enrichment, libraries were over-amplified in order to reach the minimum required
161 DNA input of 500 ng per sample. Library enrichment was then performed according to
162 the suggested myBaits protocol of two enrichment rounds with hybridisation temperature
163 at 70°C and 20 amplification cycles in the first enrichment followed by 10 cycles in the
164 second round. Samples were grouped in pools of 2–4 libraries according to the endogenous
165 DNA proportions obtained from screening of the initial shotgun sequences. Samples were
166 processed at UCSC or at ACAD (see Table S2 for a breakdown of which lab processed
167 which samples at each step).

168

169 Additionally, two experiments that examined the utility of adjustments to standard
170 enrichment protocols were performed on a subset of libraries at UCSC (Table 1). In
171 experiment A, a single round of hybridisation at 70°C was performed with 12 cycles of

172 amplification. A single round of hybridisation was also used in Experiment B but at 62.5°C
173 and with 13 cycles of amplification.
174
175 *Sequencing*
176 Libraries enriched at UCSC were paired-end sequenced on a HiSeq 4000 at Fulgent
177 Genomics (Los Angeles, USA). Paired-end sequencing using an Illumina NovaSeq 6000
178 (200 cycles) was performed by the Kinghorn Centre for Clinical Genomics (Sydney,
179 Australia) for libraries enriched at ACAD (Table S2).
180
181 *Data processing and genotype calling*
182 Data were processed either using nf-core/eager (Fellows Yates et al., 2021), or UCSC's in-
183 house processing pipeline, Batpipe (https://github.com/mjobin/batpipe). The terminal 2bp
184 at either end of each read were trimmed to remove ancient DNA damage and then
185 mapped to the GRCh37d5 reference genome using bwa-aln with parameters -l 1024 -n
186 0.01 -o 2 (Oliva et al., 2021). Other mapping parameters vary slightly between different
187 datasets used in this study, such as mapping quality threshold 20-30 and minimum read
188 length 20-30 (see Table S2 for details of processing parameters for each population).
189 Pseudo-haploid genotypes were called at the enrichment target sites using the random
190 haploid mode within pileupCaller (https://github.com/stschiff/sequenceTools).
191
192 **Population genomics assessment of capture bias**

*PCA*

We computed principal components analyses using both EMU (Meisner et al., 2021) and smartPCA version 18140 in EIGENSOFT 8.0.0 (Patterson et al., 2006; Price et al., 2006) with and without projection to test how the data bias affected each type of PCA.

*f-statistics*

Outgroup-$f_3$ and $f_4$ statistics were computed using the ADMIXTOOLS package (Patterson et al., 2012).

*READ*

We tested the impact of the enrichment bias on genetic relatedness using READ, i.e. Relatedness Estimation for Ancient DNA (Monroy Kuhn et al., 2018), which is a widely used PMR-based estimator for genetic relatedness for ancient DNA data.

**Filters and strategies to resolve bias**

In order to estimate the effect of various bioinformatic approaches to mitigate the bias observed in $f_4$ statistics, we calculated $f$-statistics in the following four forms:

1) $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.Shotgun)
2) $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.1240k)
3) $f_4$(Mbuti, Spain.1240k; Target, Peru.Shotgun)
4) $f_4$(Mbuti, Spain.1240k; Target, Peru.1240k)

214 In each, the target population was the Peru.PP population after application of one of the
215 20 strategies to reduce bias detailed below. These strategies are also summarised in Table
216 S5.
217
218 *SNP Missingness*
219 Amongst the set of 1240k SNPs assayed by the PP probes, each SNP was missing at least
220 once in our dataset. Hence, we used the full data matrix with missing genotypes and
221 reported the number of overlapping SNPs included in each $f$-statistic comparison where
222 relevant.
223 To the target population (Peru.PP), we applied two data missingness filters using the --
224 *geno* flag in PLINK 1.9 (Chang et al., 2015)—namely, 0.0 and 0.5, which correspond to
225 removing SNPs with >0% or >50% missingness, respectively.
226
227 *Minor allele frequency*
228 We tested some simple filters for minor allele frequency (i.e., 0.3, 0.1 and 0.01) using the -
229 -*maf* flag in PLINK 1.9 (Chang et al., 2015).
230
231 *Linkage disequilibrium*
232 We filtered for linkage disequilibrium (LD), with the intention to remove potential
233 artefactual linkage introduced by one bait targeting multiple SNP sites. We used the
234 PLINK 1.9 LD pruning parameter --*indep-pairwise* with a window size of 3 variants,

235 sliding in 1 variant steps and tested correlation threshold values of $r^2$ = 0.1, 0.3, 0.5, 0.7
236 and 0.9 (Chang et al., 2015). These filters resulted in datasets of 300,838, 346,428, 379,345,
237 388,038, and 388,678 SNPs, respectively.
238
239 Given that each $r^2$ threshold produced approximately 300,000 SNPs, to assess if the
240 improvement in $f_4$-value was driven by the reduction in bias attributable to artefactual
241 linkage, rather than changes to other statistical properties of the data (such as decreased
242 power due to a reduced number of SNPs), we repeated our analyses with random sub-
243 samples of 300,000 SNPs, replicated five times, using the PLINK 1.9 flag *--thin-count*
244 *(Chang et al., 2015)*.
245
246 *Transversions*
247 Sequencing libraries were partially treated with UDG (Kapp et al., 2021) and the terminal
248 two nucleotides at either end of all mapped reads were trimmed when processing the data
249 with nf-core/eager (Fellows Yates et al., 2021). Therefore, we did not expect to observe a
250 strong impact of ancient DNA damage in the analyses. Nevertheless, we filtered out
251 transitions to minimise the potential impact of ancient DNA damage, which is typically
252 characterised by C-to-T substitutions at the 5' end of sequencing reads and C-to-T or G-
253 to-A substitutions at the 3' end of sequencing reads when using single-stranded or double-
254 stranded libraries, respectively (Krause et al., 2010; Meyer et al., 2012).
255

256 *Application of Rohland et al. 2022 filter*

257 We applied the filter suggested by Rohland and colleagues (Rohland et al., 2022) to extract
258 SNPs that will allow co-analysis between 1240K data generated with shotgun sequencing,
259 or the legacy 1240k, Prime Plus or Twist enrichments.

260

261 *Read length and mapping quality threshold*

262 We filtered out reads that were shorter than 50 bp and/or with mapping quality less than
263 20, as suggested by Günther & Nettelblad to reduce reference bias when mapping
264 paleogenomic data (Günther & Nettelblad, 2019).

265

266 *Off-target genotypes*

267 Given the hypothesis that the bias is caused by bait-target interactions and that the
268 enriched library efficacy is not 100%, we called off-target (non-specific) genotypes on the
269 assumption that they would not be impacted by any bait-mediated mechanism driving the
270 bias. The genotyping of off-target sites was performed using a standard pseudo-haploid
271 pileupCaller call with the dbSNP set 138 (Sherry et al., 2001) after filtering out all PP
272 target sites. Additionally, we called genotypes that were at least 100 bp away from target
273 sites to eliminate SNPs that may have been covered by a bait but were not the target
274 locus.

275

276 *Majority call pseudo-haploid and random diploid genotypes*

In addition to standard pseudo-haploid calls based on the random selection of a single allele at prespecified SNPs, majority allele pseudo-haploid genotypes were also called at sites with greater than 5X depth-of-coverage using the majority call mode in pileupCaller. Finally, diploid genotypes were called using the random diploid call mode in pileupCaller, where one read is chosen at random, followed by a second chosen at random without replacement, and sites covered by only one read are considered missing.

### Investigation of allele call bias

We investigated the rate of each allele call at heterozygous SNP sites for the most common substitutions: (A/C, T/G, A/G, T/C), where the inverse substitution was also included such that i.e., A/C and C/A SNPs are collated together. The relative allele call frequencies were calculated per SNP (ignoring missing SNPs) in aggregate across all population datasets for shotgun-, legacy-1240k- and PP-enriched genotypes. SNP missingness was quantified across 100 bins with equal numbers of SNPs, and the mean allele frequency per bin was plotted against SNP missingness.

## RESULTS

### Global bias assessment

To quantify the extent of the PP-associated bias in outgroup-$f_3$ statistics, we calculated statistics of the form $f_3$(Mbuti; Target, X), where Target refers to samples that were enriched using PP and originate either from ancient Spain (Spain.PP.2rnd;

298 unpublished) or Peru (Peru.PP.2rnd; unpublished), and X denotes a set of South American
299 and European ancient reference populations either generated from shotgun sequencing or
300 enriched with the PP or legacy 1240k kits.
301
302 For $f_3$ statistics that use Spain.PP.2rnd as the Target population (Fig 1A), all non-PP
303 European populations (shotgun or legacy 1240k data) exhibit more shared drift with the
304 Spain.PP.2rnd population than do the South American populations of either enrichment
305 type (Fig 1A). However, the three South American PP populations share more genetic
306 drift with the Spain.PP.2rnd target population relative to other non-PP South American
307 populations. Similarly, the inter-continental relationship is preserved when Peru.PP.2rnd
308 is used as the target population, as all South American populations have substantially more
309 shared drift with the Peru.PP.2rnd target population than with European populations, yet
310 the intra-continental relationships appear to be violated by the biased shared drift
311 between PP populations (Fig 1B). Further, Spain.PP.2rnd exhibits a strong attraction
312 toward Peru.PP.2rnd samples relative to the other Spanish non-PP populations.
313
314 We also observe variance between 1240k and shotgun datasets, which is likely due to a
315 combination of different ancestries and 1240k bias (Rohland et al., 2022), however in most
316 cases the shotgun and 1240k error bars overlap, whereas the Prime Plus bias is more
317 pronounced. Clearly, the bias observed in PP-generated data is not unique to South

318 American populations but is ubiquitous across the highly diverged populations of Europe
319 and South America, strongly suggesting that the bias affects all global populations.
320
321 **Principal Components Analysis**
322 We tested the impact of PP bias on multiple PCA approaches, including the widely used
323 smartPCA (Patterson et al., 2006; Price et al., 2006) and EMU, which is reported to handle
324 data missingness well (Meisner et al., 2021). For each approach, we computed PCAs at
325 continental and global scales to investigate how missingness and the PP bias manifested at
326 different geographical scales.
327
328 *Continental PCA*
329 Initially, we performed a PCA using EMU on 362 ancient genome-wide datasets from
330 South America including shotgun (n = 38), legacy 1240k (n = 223) and PP-enriched
331 (n = 101) data (see Table S1 for a detailed list of samples). Sample clustering in the first
332 three principal components of the PCA is driven by population structure (Fig 2A and
333 there is no apparent bias exhibited by the data source type (Fig 2A). When we used
334 smartPCA on the same 362 ancient genome-wide datasets (without projection onto
335 modern diversity, due to lack of sufficient modern data), the resulting PCA produced
336 sample clustering along PC2 that was mildly separated by the assay type for Chilean
337 samples, though this may be driven by local ancestries (Fig 2B). Note that no individual
338 samples were assayed with all three methods investigated, but many samples of different

339 data types derive from the same localised ancestry groups. This result concurs with a
340 recent study (Rohland et al., 2022), which also reported that PCA is robust to
341 combinations of these three data types.
342
343 *Global PCA*
344 Subsequently, we performed another PCA testing both EMU and smartPCA without
345 projection on a merged dataset of 399 ancient South American and European genome-
346 wide datasets including shotgun (n = 38), legacy 1240k (n = 282) and PP-enriched (n = 79)
347 data. In both the EMU and smartPCA outputs, PC1 is driven by the ancestral difference
348 between South American and European populations, PC2 is driven by the diversity within
349 South America and PC3 by the diversity represented in Iberia (Fig 2C and 2D).
350
351 *Global PCA with smartPCA projection*
352 To assess the effect of the enrichment bias on a global scale, we applied both EMU and
353 smartPCA with projection methods to 436 ancient genome-wide datasets from South
354 America, Mesoamerica and Europe, either generated by shotgun sequencing or enriched
355 using the PP or legacy 1240k kits, onto a PCA of modern diversity that used 1,071
356 individuals from the HGDP dataset (Bergström et al., 2020) (Fig 3). Both PCAs look very
357 similar, the ancient individuals are clustering with similar modern ancestries.
358
359 **Genetic relatedness**

360 We ran READ for pairwise combinations of four individuals from Peru for which we had
361 generated shotgun data and single round PP-enriched data as a part of Experiment B (see
362 Methods). These individuals represent the only samples for which both shotgun data and
363 PP-enriched data were available. Given that only a single round of enrichment was used
364 in these PP assays, technical bias is expected to be reduced relative to double-round
365 enrichments.
366 The four artefactual "true twins" pairs (i.e., the same individual assayed with different
367 methods; e.g., IndA.PP.1rnd.B and IndA.Shotgun) were the most related of all
368 combinations with all average P0 values <0.4 (Fig S1). The remaining "non-twin" pairs all
369 had a P0 value of close to 0.5, with small increments between them. Notably, PP-enriched
370 pairs (e.g., Ind1.PP.1rnd.B and Ind2.PP.1rnd.B) were always more closely related (lower
371 P0) than the respective shotgun pairing (e.g., Ind1.Shotgun and Ind2.Shotgun) (Fig S1).
372 This suggests that the PP bias effects READ inferences, although it is subtle in this
373 instance.
374

375 **Comparison of strategies to reduce bias**

376 We conducted two wet-lab experiments with altered enrichment protocols and applied a
377 total of 18 bioinformatic strategies, aiming to find a strategy that best reduced the Prime
378 Plus bias. All strategies are detailed in Table S5. Bioinformatic strategies are categorised
379 according to whether they 1) reduce mapping reference bias, 2) reduce the impact of
380 ancient DNA damage, 3) remove specific "biased" SNPs, 4) employ statistical filters based

on allele frequencies or 5) call genotypes with alternate methods. Our test population comprised the ancient Peruvian samples, the only population for which we generated both Prime Plus and shotgun data (Bongers et al., 2020). This allowed us to compare the effect of bias-reduction strategies in an $f_4$ test using statistics of the form $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.Shotgun), where the target rotates between each of the different bias reduction strategies applied to the Peruvian population. We chose Spain.PP.2rnd for this assessment for the following reasons: i) the Spanish dataset was generated in a different lab, by different individuals, and was sequenced on a different sequencing machine, than the Peruvian dataset (Table S2), thereby ruling out batch effects and isolating the Prime Plus enrichment kit as the underlying cause of any bias; and ii) the Spanish and Peruvian populations are sufficiently divergent and lack evidence of recent admixture (the Peruvian data date to pre-colonial times) such that the impact of Prime Plus bias is readily observable, but is not stronger than the population affinities, as would be the case if we were comparing two more closely related groups (such as populations from the same continent). To quantify the impact of Prime Plus bias relative to population affinity between the Spanish and Peruvian samples, we included Spain.1240k and Spain.Shotgun from publicly available data as target populations.

In total, we calculated $f_4$ statistics in the following four configurations to robustly estimate the Prime Plus bias and determine the best strategies to reduce it (Fig 4):

1) $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.Shotgun)

2) $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.1240k)

3) $f_4$(Mbuti, Spain.1240k; Target, Peru.Shotgun)

4) $f_4$(Mbuti, Spain.1240k; Target, Peru.1240k)

The reason for using the first two configurations is that by comparing the target populations with Peruvian shotgun data, we are measuring our attempts to resolve the Prime Plus bias against the least biased data currently available. However, this may be an unreasonable standard to meet, given that it is known that legacy 1240k data also has some biases (Rohland et al., 2022), and the bulk of published ancient DNA data is legacy 1240k data. Therefore, while this comparison may not be a perfect estimation of total bias reduction, it will provide more useful insights for researchers given that most projects will leverage large proportions of 1240k reference data. Following the same logic, in the third and fourth configurations, we also computed $f_4$ statistics with Spanish legacy 1240k data (Fig 4), to gain a deeper understanding of the merits of alternate strategies for reducing bias in different combinations of PP, legacy 1240k and shotgun data. The full list of $f_4$ statistics computed in this study are detailed in Table S6.

As a positive control, we included a different set of ancient shotgun genomes from Peru (Peru.other.Shotgun) as a Target population, which resulted in $f_4$ results that were not significantly different from zero ($Z < 3$), as expected. The main aim of this process was to identify a strategy that would behave like the positive control, i.e., that would reduce the enrichment bias (Peru.PP.2rnd; dashed vertical line, Fig 4) so the $f_4$ values are not significantly different from 0. In the first configuration $f_4$(Mbuti,

Spain.PP.2rnd; Target, Peru.Shotgun), a result not statistically different from 0 would indicate the bias was alleviated, and in the second configuration $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.1240k), the same result would suggest that the PP data is at least co-analysable with 1240k data under specific configurations. Importantly, the third configuration $f_4$(Mbuti, Spain.1240k; Peru.PP.2rnd, Peru.Shotgun) (Fig 4) resulted in an $f_4$ result that was not significantly different from zero, indicating that there was no significant bias between these PP and legacy 1240k datasets. Finally, for the fourth configuration $f_4$(Mbuti, Spain.1240k; Peru.PP.2rnd, Peru.1240k) (Fig 4) the $f_4$ result is significantly positive, providing further evidence for the previously reported bias in the legacy 1240k enrichment (Rohland et al., 2022). Therefore, the purpose of including Spain.1240k among our tested configurations is to find a bias-reduction strategy that does not significantly impact these $f_4$ statistics, and that allows the confident application of bias reduction strategies to different combinations of the Prime Plus and other data types and under alternate $f_4$ statistic configurations.

*Ancient DNA damage or reference bias*

The filter for minimum read length (Peru.PP.2rnd.>50bp) suggested by (Günther & Nettelblad, 2019) to reduce mapping reference bias had minimal effect on the $f_4$ statistics, suggesting that the observed assay bias is not strongly impacted by very short reads.

Filtering out transition SNPs (A/G and C/T), thus removing any possible ancient DNA damage (Peru.PP.2rnd.transversion) had a negligible effect on the $f_4$ statistics scores. Indeed, because all libraries were half-UDG treated and the ends of sequencing reads trimmed, we did not anticipate that ancient DNA damage would have a large impact.

*SNP-set based filters*

Strategies to reduce bias by removing certain SNPs included filtering to include only the Human Origins SNP set (Peru.PP.2rnd.HO; 593,120 SNPs retained) and a filter previously reported (Rohland et al., 2022) to enable the co-analysis of legacy 1240k and Prime Plus data (Peru.PP.2rnd.Rohland; 803,662 SNPs retained). Retaining only the Human Origins SNPs had minimal effect on the $f_4$ results. Notably, the application of the Rohland filter did not alleviate bias in any of the configurations of the $f_4$ statistics that we tested. In some cases application of the Rohland filter resulted in the $f_4$ values shifting slightly toward zero, but in other cases slightly away, which is likely due to an increase in uncertainty from the approximately 2-fold reduction in the number of SNPs. Overall, this indicates the Rohland filter does not work as reported on the present dataset, suggesting that it is not a universally applicable solution for capture bias.

*Allele frequency-based filters.*

We further examined the bias reduction efficacy of a series of filters based on allele frequencies, including filters for SNP missingness, minor allele frequency and linkage disequilibrium. Of the SNP missingness filters, a strategy that excluded sites missing

in >50% of individuals (Peru.PP.2rnd.miss0.5) did not notably shift the $f_4$ value towards zero. However, excluding all SNPs exhibiting any missingness (Peru.PP.2rnd.miss0) did move the value toward zero, but it also reduced the number of SNPs so drastically that this strategy would probably not be suitable for most paleogenomic studies. For SNP filters based on minor allele frequency (Peru.PP.2rnd.MAF0.3, Peru.PP.2rnd.MAF0.1 and Peru.PP.2rnd.MAF0.01), $f_4$ results were shifted toward zero as the stringency of the filter increased, but this also led to increasingly large reductions in the number of analysable SNPs. Interestingly, in the $f_4$ configurations with Spain.1240k, the minor allele frequency filters produced positive $f_4$ values indicating a bias towards the Peruvian data regardless of whether shotgun or legacy 1240k are used.

When applying LD-based filters, we identified that the shared drift with Spain.PP.2rnd reduces as the stringency of the filter increases, though this once again came at the cost of increasingly fewer retained SNPs. In the configuration $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.Shotgun), the $f_4$ result for $r^2 = 0.1$ was very close to 0 (Peru.PP.2rnd.LDr0.1). Filters for less stringent $r^2$ values filtered out fewer SNPs but did not improve $f_4$ results to the same degree (Fig 4). In the configuration $f_4$(Mbuti, Spain.PP; Target, Peru.1240k), the less stringent $r^2 = 0.3$ produces an $f_4$ value not significantly different from 0, with the retention of more SNPs than the $r^2 = 0.1$ filter. The use of this filter to alleviate PP bias could be considered by users in specific cases with high sequencing depth to accommodate the expected SNP loss. Randomly downsampling the Peru.PP data to an equivalent

number of SNPs (300,000) yielded $f_4$ values with additional variance around the value of the unfiltered Peru.PP data (dashed lines, Fig 4) and higher absolute $Z$-scores (Fig 4). This suggests that the bias may in fact be creating artificial "linkage" among neighbouring SNPs, but that sufficiently stringent LD pruning to "correct" the value to zero requires discarding the majority of targeted SNPs. Accordingly, the value of linkage disequilibrium pruning may be considered as a potential solution to the PP bias on a case-by-case basis, with the caveat that improved bias correction will be offset to some degree by the loss of precision from reduced statistical power, and consideration will need to be given to the effect of LD pruning on the results of various analyses.

*Alternative genotyping*

We considered several genotyping approaches alternative to the standard random pseudo-haploidisation, including calling random diploid genotypes, using majority call pseudo-haploid genotypes at sites with > 5X coverage, genotyping off-target SNPs, and genotyping 'out-of-range' SNPs that were > 100 bp from the target sites. Calling off-target genotypes (Peru.PP.2rnd.OffTarget) and out-of-range genotypes (Peru.PP.2rnd.>100bpOffTarget) both resulted in a small shift of the $f_4$ values toward zero. Majority call genotypes (Peru.PP.2rnd.>5X.maj) behaved very differently between configurations. In configurations with Spain.1240k, the filter had a minimal effect, but with Spain.PP.2rnd, the majority call genotyping drastically shifted the $f_4$ results towards negative values, indicating that this approach increases the artefactual "shared drift" between PP data from Spain and Peru. Random diploid genotyping

produced an $f_4$ value that was even more biased (Peru.PP.2rnd.diploid), mildly so when compared with Spain.1240k, but much more strongly for Spain.PP.2rnd. In fact, the latter value is close to the $f_4$ result where the target dataset is Spain.Shotgun. This means that using random diploid genotype calls exacerbates the artefactual shared drift driven by the PP enrichment bias to approximately the same magnitude as true shared drift between Spanish populations (Fig 4).

*Laboratory experiments*

In our final set of bias reduction strategies, two non-standard lab experiment protocols that used a single round of enrichment (Peru.PP.1rnd.A and Peru.PP.1rnd.B) both led to minor bias reductions. Protocol B, which lowered the hybridisation temperature performed slightly better than Protocol A. While the overall improvement exhibited by both methods was minor, the bias remained considerably greater than the bias observed in the legacy 1240k data (Peru.1240k).

**Allele calls**

We observe strong GC bias amongst the data generated with the legacy 1240k baits, whereas the PP data display comparatively lower levels of bias that are broadly comparable with shotgun datasets (Fig 5). It is important to note that the bias apparent in the legacy 1240k data was not a concern prior to the release of alternate enrichment kits, simply because all the data generated using the legacy 1240k baits are comparable. We observe that across all three methods, the GC bias is stronger for transitions—i.e., allele

pairs A/G and C/T where both alleles are either purines or pyrimidines, respectively—which represent the majority of target sites (80%). Though our understanding of the biochemical mechanisms underlying the PP bias is limited by a lack of knowledge of the bait design, our data clearly demonstrate a systematic difference in the pattern of allelic biases between the legacy 1240k and PP assays.

## DISCUSSION

Results presented in this study support the previous report of a bias when enriching SNPs at ~1240k sites using the Daicel Arbor Biosciences Expert Human Affinities Prime Plus enrichment kit, and extend the previous findings to ancient Indigenous American (i.e., non-Eurasian) populations. In accordance with (Rohland et al., 2022) we find that the biased data strongly affect $f$-statistic calculations, where the shared genetic drift (i.e., the evolutionary history) between groups is explicitly estimated. Note that we consider the effect strong, not because of the magnitude of any quantitative measurements, but due to the corruption of expected demographic relationships between populations that is driven by this bias. In accordance with results reported by (Rohland et al., 2022), we find that the assay bias does not strongly affect PCA at least in the first three prinicipal components. Additionally, for the first time, we observe a (in this case mild) bias effect on the predicted relatedness between individuals using the ancient DNA relatedness test READ.

### Batch effects

555 The initial observation that PP data had an unexpected excess of shared drift with other
556 PP-generated data in $f$-statistics prompted us to investigate if this artefact was due to a
557 batch effect arising from certain samples, specific labs having done the work, DNA
558 sequencing machines, or resulted from the distinct demographic history of populations in
559 question. We can confidently rule out each of these hypotheses, as the lab work was
560 performed on numerous sample sets, in different labs, by multiple individuals, sequenced
561 in different facilities using different sequencers (Table S2), with the only common factor
562 to all biased samples being the PP enrichment. Our outgroup-$f_3$ statistics calculated
563 between European and South American populations (Fig 1) confirmed that the bias
564 affected these two highly diverged populations suggesting that it is most likely a
565 systematic bias that affects all human ancestries. However, we did not have access to the
566 appropriate shotgun data to determine if the enrichment bias has differential effects by
567 ancestry source, and this hypothesis requires additional testing.
568
569 ***Describing the Prime Plus bias***
570 *PCA*
571 We performed principal components analysis in several ways with the aim of addressing
572 multiple analytical approaches that researchers might take. To accommodate differing
573 geographical scales, we computed separate PCAs at continental and global scales. While
574 the approach of using smartPCA with projection has long been considered the gold
575 standard, it should be noted that this method works best with high-quality ancient

samples that have low missingness and a well-developed reference database of modern diversity that captures the genetic variation present in the ancient populations, therefore enabling the most accurate projection. Hence, the PCA published by Rohland et al. 2022 could be expected to show good results as they projected 15 high-quality ancient genomes onto the modern diversity of western Eurasia, the most studied region of the world. However, this is generally less achievable in understudied regions where modern reference data is lacking and/or when working with low-quality samples. This is what motivated the simultaneous use of EMU (designed to handle data missingness and does not use projection) (Meisner et al., 2021) and smartPCA both with and without projection onto modern diversity.

At both the the continental and global scale, both EMU and smartPCA without projection produced PCAs that reflects South American population structure in accordance with previous results (Posth et al. 2018; Nakatsuka et al. 2020; Nakatsuka et al. 2020; Martiniano et al. 2017; Olalde et al. 2019; Valdiosera et al. 2018) and patterns suggesting assay bias are not evident within the first three principal components (Fig 2). However, we caution that the results of smartPCA are variable by the version used and the data inputted, therefore these results do not preclude the emergence of assay bias in a lower principal component or within a different dataset.

At a global scale, both smartPCA with projection onto the modern diversity of the HGDP data (Bergström et al., 2020) and EMU performed similarly. Note that EMU does not have

projection functionality and instead includes all individuals in the calculation of eigenvectors while imputing sites with missing data (Meisner et al., 2021). There does not seem to be any obvious assay bias in either PCA output, although this might still be present in a lower-level PC. Furthermore, discrepancies between the PCA methods, and the effect of assay bias may become more impactful on results and interpretations at finer-scales.

Overall, we suggest constructing PCAs using smartPCA, projecting ancient genomes onto modern diversity where possible. However, in the case that modern reference data is lacking, PCA plots from smartPCA and EMU should be produced and compared, especially in cases where PP data, or data with high rates of missingness are included.

*Genetic relatedness*

We assessed the effect of PP bias on results from the genetic relatedness estimation software READ (Monroy Kuhn et al., 2018). We found that pairwise estimations between PP libraries were always more closely related than shotgun library pairs of the same four samples (Fig S1). Given that these PP libraries were only enriched for a single round, we would expect this bias to be even stronger in PP libraries that follow the standard two-round enrichment protocol (reflecting patterns observed in the $f_4$ statistics; Fig 4). In the present case, the bias is subtle and unlikely to lead to, for example, a false-positive first-degree relationship from an unrelated pair. While our ability to make more general

conclusions is clearly limited by the sample set, our results tentatively suggest that READ inferences should be reasonably robust if run on a dataset of only PP data, but users should exhibit caution when combining different data types.

### $f$-statistics

Our analyses of the PP bias in both outgroup-$f_3$ and $f_4$ statistics showed that the bias is strong enough to corrupt expected relationships between populations from the same continent (Fig 1). We reproduced the legacy 1240k bias that has been previously reported (Rohland et al., 2022) and showed that it is relatively subtle compared to the PP bias, indicating that it is less impactful on population genetic inferences.

Importantly, our analyses suggest that it should be possible to use Prime Plus data in a comparative outgroup-$f_3$ if the target population is the only one enriched with Prime Plus, that is statistics of the form $f_3$(Outgroup.notPP; target.PP, X.notPP) where both the outgroup and population X are not enriched with Prime Plus. This is suggested by our $f_3$ statistics (Fig 1A) where Peru.PP.2rnd is the Target population, as the other non.PP South American populations reproduce the expected relative relationships previously reported (Nakatsuka, Lazaridis, et al., 2020; Nakatsuka, Luisi, et al., 2020; Posth et al., 2018) despite none having been enriched with the Prime Plus baits. However, the limitation is that a population enriched with Prime Plus could not be split into subpopulations, such as Peru_A.PP.2rnd and Peru_B.PP.2rnd, and rotated in the X

population position, or in scenarios where one is X and the other Test, but then compared to $f_3$ statistics of a different form. Presumably, an $f$-statistic would also be viable if all X groups were PP-enriched because the enrichment bias would be shared by all data and should not corrupt demographic interpretations of the results. See Table S4 for a comprehensive list of acceptable outgroup-$f_3$ and $f_4$ statistic forms and, though not directly investigated here, using the same logic we also list qpWave and qpAdm configurations and their relative reliability under PP bias.

***Strategies to reduce bias***

The bias observed in $f$-statistics is both persistent and systematic, which allowed us to test a total of 20 experimental and analytical protocols to characterise the bias and alleviate its disruptive impact on inferences from common population genetic analyses. Many of these analytical protocols produced marginal improvements, but no protocol was sufficient to remove all biases ($Z$-score remained such that $|Z| > 3$, and/or resulted in low SNP retention).

*Ancient DNA damage or reference bias*

Neither ancient DNA damage nor reference bias appeared to contribute to the bias, as removing transition sites (all SNPs that could potentially be ancient DNA damage) and removing reads shorter than 50 bp had little impact on bias in the $f_4$ statistics (Fig 4).

*SNP-set based filters*

Next, we investigated if a subset of the 1240k SNPs was driving the bias and could therefore be removed to allow safe co-analysis of PP enriched samples with other data types. This idea was initially suggested by Rohland and colleagues (Rohland et al., 2022), who published a list of SNPs that they suggested should be suitable for co-analyses of shotgun, legacy 1240k, Twist, and PP data. However, our tests of the Rohland et al. filter produced no notable improvement on the $f_4$ statistics when co-analysed with 1240k or PP data (Peru.PP.2rnd.Rohland, Fig 4). Note that to build this filter Rohland and colleagues looked for improved $Z$-scores, i.e., closer to zero, but did not report the $f_4$ statistics that are informative to estimate the effect on population relatedness. When using any SNP filter, the removal of SNPs will likely increase the standard error of the $f_4$ statistic, leading to decreased statistical power and precision. However, it is not possible to predict the deviation from the true $f_4$ statistic. Therefore $Z$-scores alone do not serve as a measure of the bias-reduction of a filtering strategy. This emphasises the importance of testing hypotheses in the context of the study and understanding the expected relationships of the populations in question.

Subsetting our data to include only the ~600k SNPs in the Human Origins set had a similar effect to the Rohland filter. This may be related to the overlap between the two sets, 61% of the Human Origins SNPs are also present in the Rohland filter set. (367,332 intersecting SNPs).

683

684 In summary, none of these filtering strategies corrected the PP bias satisfactorily.
685 Furthermore, while Rohland et al. acknowledge there is room for improvement in their
686 filtering approach, it is possible that their filter is well-calibrated for the Eurasian
687 ancestries present in their dataset but does not represent a universal filter to mitigate
688 Prime Plus capture bias in all ancestries. In any case, it is less than ideal to have
689 enrichment kits on the market that would require discarding a subset of target SNPs when
690 combining multiple data types. This will negatively affect not only users of the Daicel
691 Arbor Biosciences Human Affinities kits — and potentially other commercial kits — but
692 also impact subsequent analyses by other researchers that incorporate published PP
693 datasets.

694

695 *Alternative genotyping*

696 Genotype calling in paleogenomic studies typically involves randomly sampling one read
697 covering a targeted SNP site, known as 'pseudo-haploidisation', making it possible to
698 generate allele calls at sites with low coverage (Morris et al., 2011). Given that our
699 experiments with SNP filtering did not produce a universal solution and following the
700 assumption that every targeted SNP site may be affected by the bias, we investigated the
701 efficacy of alternate genotype calling for bias reduction. In our first assessment, we called
702 pseudo-haploid genotypes at off-target SNPs, which we hypothesised would be equivalent
703 to shotgun data and therefore unbiased if the bias was driven by the bait design. Similarly,

we also called random pseudo-haploid genotypes at sites more than 100 bp away from a target locus to mitigate the possibility of an "off-target" site being covered by a probe that had correctly hybridised to a nearby target locus. The $f_4$ statistic for both off-target and out-of-range genotypes did improve compared to the on-target results, producing values that were not significantly different from zero (Fig 4). Interestingly, the $f_4$ statistic for the out-of-range genotypes was not notably different from the off-target estimation. Given that all non-bound DNA fragments are theoretically washed away during the enrichment protocol, the off-target bait bias is likely due to an as-yet-unknown biochemical bias that systematically affects bait binding to DNA at non-target locations, but it is unlikely this is caused by on-target hybridised baits covering nearby SNPs within a 100 bp range.

Additionally, we used an alternate pseudo-haploid calling approach which requires target SNPs to have a higher minimum coverage (5X) and takes the allele represented by the majority of reads as the final call. This experiment (Peru.PP.2rnd.>5X.maj, Fig 4) yielded a dramatic shift of the $f_4$ results towards large negative values, further exacerbating the artefactual shared drift and bias. This is strong evidence that the bias is conferred by the bait design, and results in the persistent overrepresentation of one allele over the other at heterozygous sites. Concordantly, our random diploid genotype test produced even more biased $f_4$ statistics, such that the estimated levels of shared drift observed between the Spanish PP data and Peruvian PP data that used random diploid genotype calls is of a similar magnitude to the true shared

726 drift between Spanish PP generated and Spanish shotgun data. This is evidence of a
727 remarkably strong systematic bias that has the potential to corrupt all paleogenomic
728 studies combining PP with other data types if not handled carefully. These two results are
729 strongly suggestive of a systematic allelic bias that impacts all heterozygous SNPs and is
730 potentially caused by the molecular design of the PP baits.
731
732 *Laboratory protocol experiments*
733 Our two experiments with single-round PP enrichment yielded $f_4$ values closer to zero
734 than those produced by the standard two-round enrichment. This improvement is likely
735 because whatever bias occurs during the first round of bait–DNA binding becomes
736 amplified during the second hybridisation step because it is acting on an already biased set
737 of DNA fragments. This interpretation is supported by the sequencing statistics (Fig S3)
738 showing that repeated capture rounds result in an increased endogenous DNA percentage
739 but decreased library complexity (observed as fewer unique reads when sequenced).
740 Additionally, any biased sampling that has occurred during the initial binding will be
741 exponentially amplified during the PCR step, so we would expect a reduction in PCR
742 cycles to improve this. The reduced hybridisation temperature should also decrease the
743 specificity of hybridisation and thus reduce allelic bias, but simultaneously increase off-
744 target binding, thus reducing the percentage of endogenous DNA enriched.
745
746 ***Characterising the bias***

747 Given the evidence that the bias affects all targeted SNPs, we investigated whether this
748 effect was independent from the SNP transition, hoping that we might be able to apply
749 some relatively simple statistical scaling of the genotype calls that would provide co-
750 analysable genotypes. However, our investigations of allelic bias (Fig 5) clearly show that
751 the capture bias produces complex effects on allele calls. We can only confidently observe
752 that the pattern of GC bias in PP data is different to the 1240k and shotgun data. Further,
753 the patterns of missingness evident in the PCA analyses consistently differ between data
754 types (Fig S2), indicating that there may be several layers of biochemical bias at play
755 including that each assay preferentially captures certain SNPs, and/or favours the capture
756 of different alleles at heterozygous sites, with allelic preference being more exacerbated in
757 PP data.
758
759 ***Study limitations***
760 The main limitation of this study is that we did not have samples for which data was
761 generated in all of the three methods addressed. This is due to the opportunistic nature of
762 the study in that we collated this dataset when the PP bias was initially observed, after the
763 data generation process. However, we have made every effort to combat this limitation by
764 maximising the size of the dataset and explicitly stating assumptions where we've made
765 them.
766

767 In our READ analyses we were limited to a dataset of only four individuals for which we
768 had shotgun and PP data, and a further limitation is that this PP data was from one of our
769 single-round enrichment experiments and thus does not fully represent the PP bias under
770 investigation. To further understand the impact on READ, both shotgun and two-round
771 PP enriched data would need to be generated for a larger number of samples, from a more
772 diverse set of relationships to test the impact of bias on truly related individuals.
773
774 The main outstanding hypothesis to investigate is whether this bias has differential effects
775 on different ancestries. This would require a large dataset of PP and shotgun data from the
776 same individuals across more ancestries than in the present study. While an interesting
777 question, we believe this study presents sufficient evidence that the PP assay is biased.
778
779 *Implications*
780 For the past decade in-solution hybridisation enrichment has been a go-to method for
781 obtaining affordable, high-quality, and informative data for paleogenomic studies.
782 However, irrespective of the bait design, the biochemistry of DNA hybridisation means
783 that one bait will always have a higher binding affinity with its targeted allele than any
784 other. Greater differences in these affinities result in increased allelic biases in the
785 generated datasets. This is evidenced by our assessment of alternative enrichment
786 protocols, where the single-round enrichments have less endogenous DNA (Fig S3), but

also appear to be less biased (Fig 4). Therefore, we hypothesise that the enrichment efficiency of the PP reagent is in direct trade-off with the biochemical allelic bias.

It is worth reiterating here that all capture methods used in paleogenomic studies will lead to ascertainment bias, meaning that one must carefully consider both the selection of methods and the use of comparative published data. Previously, the monopoly on data generation by labs with access to the legacy 1240k enrichment protocol somewhat insulated the community from this compatibility issue. However, now that data can be generated using different enrichment kits, researchers face a new set of challenges whenever they integrate data generated with new capture assays with previously published data generated using the legacy 1240k—the most comprehensive dataset to date that underpins many of the findings in paleogenomics. Additionally, researchers utilising the Twist reagent should benchmark the data similarly to the present study to establish its co-analysability with other data types.

Here, we have tested some of the most used analyses in paleogenomic studies, though not an exhaustive list. Principal Components Analysis is a standard way to understand population structure, and outgroup-$f_3$ and $f_4$ statistics, along with their more complex derivatives, are integral to reconstructing population relationships and evolutionary history, while READ is a standard for genetic relatedness analysis. We have found that PP-enriched data bias the results of READ and $f$ statistics when used in combination with other non-PP datasets, and therefore we caution potential PP users that

809 analyses of such multi-assayed datasets may lead to erroneous conclusions if not
810 interpreted carefully.
811
812 Having comparative shotgun data for a subset of samples in a study, or from samples
813 expected to have a very similar ancestry should allow researchers to estimate and control
814 for the PP bias in the specific methods they wish to utilise. Further, while we explicitly
815 test PP data, we advise researchers to exhibit similar caution when using the Daicel Arbor
816 Biosciences Expert Human Affinities "Complete" and "Ancestral Plus" enrichment kits, as
817 the bait chemistry and enrichment protocol is the same as the PP kit, and only the
818 number of target loci differ.
819
820 *Guidelines*
821 - When calculating $f$-statistics only compare Prime-Plus-generated data with other
822 Prime-Plus-generated data; OR
823 - Only compare one Prime-Plus-generated target population/individual with other
824 non-Prime-Plus-generated populations/individuals (See Table S4 for a full
825 description).
826 - Beware that Prime-Plus-generated data attract each other and that this may
827 artificially confirm/reject your hypothesis depending on the study design.
828 - Explicitly assess the impact of Prime-Plus bias in your data when running analyses
829 that have not yet been investigated.

- Recommend to shotgun sequence a subset of enriched libraries as a control set if possible, ideally select libraries to represent a range of ancestries.
- Always annotate your data as Prime-Plus-generated when publishing, sharing with other researchers, or incorporating into any publicly accessible database.
- Always annotate the data type of both newly generated and published data used in your study so this can be factored into interpretations.
- For best results compute PCAs using smartPCA with projection where possible and otherwise with EMU to compare results; if clustering appears driven by capture assay, pursue alternative analyses.
- In general, pursue multiple lines of evidence to support your findings.

## Acknowledgements

We thank Adam Rohrlach for insightful manuscript revisions and assistance with statistical interpretation. We thank members of the Australian Centre for Ancient DNA's Thesis Writing Group for their contributions to manuscript revisions, especially Olivia Johnson for assistance with R coding. We thank the members of the modern and ancient DNA research group at the Unitat d'Antropologia Biològica, Universitat Autònoma de Barcelona (Spain) for their contribution to data generation. We thank Linda R. Manzanilla and Grégory Pereira from the Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México (Mexico), and UMR 8096 Archéologie des Amériques, Centre National de la Recherche Scientifique (France), respectively, for their contribution to sample provision. LFS was supported by funding from the National Science Foundation (NSF - 1515138), RT by the Australian Research Council (DE190101069), BL by an Australian Research Council Future Fellowship (FT170100448).

## References

Ávila-Arcos, M. C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M. L., Malaspinas, A.-S., Wales, N., Peñaloza, F., Bustamante, C. D., & Gilbert, M. T. P. (2015). Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina

libraries. *Methods in Ecology and Evolution / British Ecological Society*, *6*(6), 725–734.

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J.-F., Cann, H., Mallick, S., Reich, D., Sandhu, M. S., Skoglund, P., Scally, A., Xue, Y., … Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* , *367*(6484), eaay5012.

Boessenkool, S., Hanghøj, K., Nistelberger, H. M., Der Sarkissian, C., Gondek, A. T., Orlando, L., Barrett, J. H., & Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*, *17*(4), 742–751.

Bongers, J. L., Nakatsuka, N., O'Shea, C., Harper, T. K., Tantaleán, H., Stanish, C., & Fehren-Schmitz, L. (2020). Integration of ancient DNA with transdisciplinary dataset finds strong support for Inca resettlement in the south Peruvian coast. *Proceedings of the National Academy of Sciences*, *117*(31), 18359–18368;

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V. B., Golovanova, L. V., de la Rasilla, M., Fortea, J., Rosas, A., & Pääbo, S. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, *325*(5938), 318–321.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S. M., Sarkissian, C. D., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M. B., Korlach, J., Luong, K., Ho, S. Y. W., Quintana-Murci, L., Behar, D. M., … Cooper, A. (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications*, *4*, 1764.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American Journal of Human Genetics*, *93*(5), 852–864.

Castellano, S., Parra, G., Sánchez-Quinto, F. A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., Dabney, J., Siebauer, M., White, L., Burbano, H. A., Renaud, G., Stenzel, U., Lalueza-Fox, C., Rasilla, M. de la, Rosas, A., … Pääbo, S. (2014). Patterns of coding variation in the complete exomes of three Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6666–6671.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.

Cruz-Dávalos, D. I., Nieves-Colón, M. A., Sockell, A., Poznik, G. D., Schroeder, H., Stone, A. C., Bustamante, C. D., Malaspinas, A.-S., & Ávila-Arcos, M. C. (2018). In-solution Y-chromosome capture-enrichment on ancient DNA libraries. *BMC Genomics*, *19*. https://doi.org/10.1186/s12864-018-4945-x

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763.

de la Fuente, C., Ávila-Arcos, M. C., Galimany, J., Carpenter, M. L., Homburger, J. R., Blanco, A., Contreras, P., Cruz Dávalos, D., Reyes, O., San Roman, M., Moreno-Estrada, A., Campos, P.

F., Eng, C., Huntsman, S., Burchard, E. G., Malaspinas, A.-S., Bustamante, C. D., Willerslev, E., Llop, E., … Moraga, M. (2018). Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), E4006–E4012.

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J., & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ*, *9*(e10947), e10947.

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PloS One*, *8*(11), e79667.

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prüfer, K., Meyer, M., Kelso, J., Reich, D., & Pääbo, S. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, *524*(7564), 216–219.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., & Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences*, *110*(6), 2223–2227.

González-Fortes, G., Tassi, F., Trucchi, E., Henneberger, K., Paijmans, J. L. A., Díez-Del-Molino, D., Schroeder, H., Susca, R. R., Barroso-Ruíz, C., Bermudez, F. J., Barroso-Medina, C., Bettencourt, A. M. S., Sampaio, H. A., Grandal-d'Anglade, A., Salas, A., de Lombera-Hermida, A., Fabregas Valcarce, R., Vaquero, M., Alonso, S., … Barbujani, G. (2019). A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proceedings. Biological Sciences / The Royal Society*, *286*(1895), 20182288.

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics*, *15*(7), e1008302.

Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, Ó. O., Daskalaki, E. A., Skoglund, P., Naidoo, T., Svensson, E. M., Bermúdez de Castro, J. M., Carbonell, E., Dunn, M., Storå, J., Iriarte, E., Arsuaga, J. L., Carretero, J.-M., Götherström, A., & Jakobsson, M. (2015). Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(38), 11917–11922.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., & Stewardson, K. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207.

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *The Journal of Heredity*, *112*(3), 241–249.

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., & Pääbo, S. (2010). A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology: CB*, *20*(3), 231–236.

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., Watson, J. T., Viviano Llave, C., Witonsky, D., Beall, C., Warinner, C., Novembre, J., Aldenderfer, M., & Di Rienzo, A. (2018). The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Science Advances*, *4*(11), eaau4921.

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., … Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European

farmers. *Nature*, *551*(7680), 368–372.
Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I. B., Espinoza, I. F., Cagigao, E. T., Jiménez, L. W., Makowski, K., Reyna, I. S. L., Lory, J. M., Torrez, J. A. B., Rivera, M. A., … Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, *2*(4), e1501385.
Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206.
Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews. Genetics*, *18*(11), 659–674.
Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS One*, *5*(11), e14004.
Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera, A. C., Porfirio, E., Boaventura, R., Silva, A. M., & Bradley, D. G. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics*, *13*(7), e1006852.
Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., & Novak, M. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503.
Meisner, J., Liu, S., Huang, M., & Albrechtsen, A. (2021). Large-scale Inference of Population Structure in Presence of Missingness using PCA. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btab027
Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Pruefer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, *338*(6104), 222–226.
Monroy Kuhn, J. M., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PloS One*, *13*(4), e0195491.
Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., Allentoft, M. E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Iraeta Orbegozo, M., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., … Willerslev, E. (2018). Early human dispersals within the Americas. *Science*, *362*(6419), eaav2621.
Morris, G. P., Grabowski, P. P., & Borevitz, J. O. (2011). Genomic diversity in switchgrass (Panicum virgatum): from the continental scale to a dune landscape. *Molecular Ecology*, *20*(23), 4938–4952.
Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins-Kinkaid, K., Ferry, M., Harney, É., Michel, M., Stewardson, K., Novak-Forst, J., Capriles, J. M., Alfonso Durruty, M., Aranda Álvarez, K., Beresford-Jones, D., Burger, R. L., Cadwallader, L., … Fehren-Schmitz, L. (2020). A paleogenomic reconstruction of the deep population history of the Andes. *Cell*, *181*(5), 1131–1145.e21.
Nakatsuka, N., Luisi, P., Motti, J. M. B., Salemme, M., Santiago, F., D'Angelo Del Campo, M. D., Vecchi, R. J., Espinosa-Parrilla, Y., Prieto, A., Adamski, N., Lawson, A. M., Harper, T. K.,

Culleton, B. J., Kennett, D. J., Lalueza-Fox, C., Mallick, S., Rohland, N., Guichón, R. A., Cabana, G. S., … Reich, D. (2020). Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. *Nature Communications*, *11*(1), 3868.

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., … Reich, D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, *555*(7695), 190–196.

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N., Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., … Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science*, *363*(6432), 1230–1234.

Olalde, I., & Posth, C. (2020). Latest trends in archaeogenetic research of west Eurasians. *Current Opinion in Genetics & Development*, *62*, 36–43.

Oliva, A., Tobler, R., Cooper, A., Llamas, B., & Souilmi, Y. (2021). Systematic benchmark of ancient DNA read mapping. *Briefings in Bioinformatics*, *22*(5), bbab076.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Popović, D., Molak, M., Ziółkowski, M., Vranich, A., Sobczyk, M., Vidaurre, D. U., Agresti, G., Skrzypczak, M., Ginalski, K., & Lamnidis, T. C. (2021). Ancient genomes reveal long-range influence of the pre-Columbian culture and site of Tiwanaku. *Science Advances*, *7*(39), eabg7261.

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., Rohland, N., Nägele, K., Adamski, N., & Bertolini, E. (2018). Reconstructing the deep population history of Central and South America. *Cell*, *175*(5), 1185–1197. e22.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., … Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Víctor Moreno-Mayar, J., Korneliussen, T. S., Pierre, T., … Willerslev, E. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. https://doi.org/10.1126/science.aab3884

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1660), 20130624.

Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., & Reich, D. (2022). Three assays for

in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Research*. https://doi.org/10.1101/gr.276728.122

Rohrlach, A. B., Papac, L., Childebayeva, A., Rivollat, M., Villalba-Mouco, V., Neumann, G. U., Penske, S., Skourtanioti, E., van de Loosdrecht, M., Akar, M., Boyadzhiev, K., Boyadzhiev, Y., Deguilloux, M.-F., Dobeš, M., Erdal, Y. S., Ernée, M., Frangipane, M., Furmanek, M., Friederich, S., … Haak, W. (2021). Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. *Scientific Reports*, *11*(1), 15005.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311.

Skoglund, P., & Mathieson, I. (2018). Ancient Genomics of Modern Humans: The First Decade. *Annual Review of Genomics and Human Genetics*, *19*(1), 381–404.

Valdiosera, C., Günther, T., Vera-Rodríguez, J. C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L. G., Martínez-Sánchez, R. M., Svensson, E. M., Malmström, H., Rodríguez, L., Bermúdez de Castro, J.-M., Carbonell, E., Alday, A., Hernández Vera, J. A., Götherström, A., Carretero, J.-M., Arsuaga, J. L., Smith, C. I., & Jakobsson, M. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3428–3433.

Villalba-Mouco, V., van de Loosdrecht, M. S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Salazar-García, D. C., Royo-Guillén, J. I., Kunst, M., Rougier, H., Crevecoeur, I., Arcusa-Magallón, H., Tejedor-Rodríguez, C., García-Martínez de Lagrán, I., Garrido-Pena, R., Alt, K. W., Jeong, C., Schiffels, S., Utrilla, P., … Haak, W. (2019). Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Current Biology: CB*, *29*(7), 1169–1177.e7.

## Author Contributions

R.D., B.L., L.F-S., M.P.W. and X.R-R. conceptualised the study. R.D., L.F-S., K.K., M.P.W. and X.R-R. performed laboratory work. R.D. and L.F-S. performed computational analyses. R.T., B.L., and L.F-S. funded the study. All authors contributed to manuscript writing and revisions.

## Data Accessibility and Benefit-Sharing Section

Genomic Data:

1078 The newly generated genotype data in this study are available with DOI:

1079 10.25909/24004665. We recommend against the inclusion of this data in other population

1080 genomic analyses due to the bias present.

1081

1082 Code Availability:

1083 Code used for analyses and plot generation is available at https://github.com/roberta-

1084 davidson/Davidson_etal_2023-Capture-bias-paper.

1085

1086 **Tables and Figures**

1087 **Table 1:** Relevant modified variables for each of the laboratory experiments.

| Variable | Standard Protocol | Experiment A | Experiment B |
|---|---|---|---|
| Enrichment rounds | 2 | 1 | 1 |
| Hybridisation temperature | 70°C | 70°C | 62.5°C |
| Amplification cycles | 20 (1st round) + 10 (2nd round) | 12 | 13 |

1088

1089

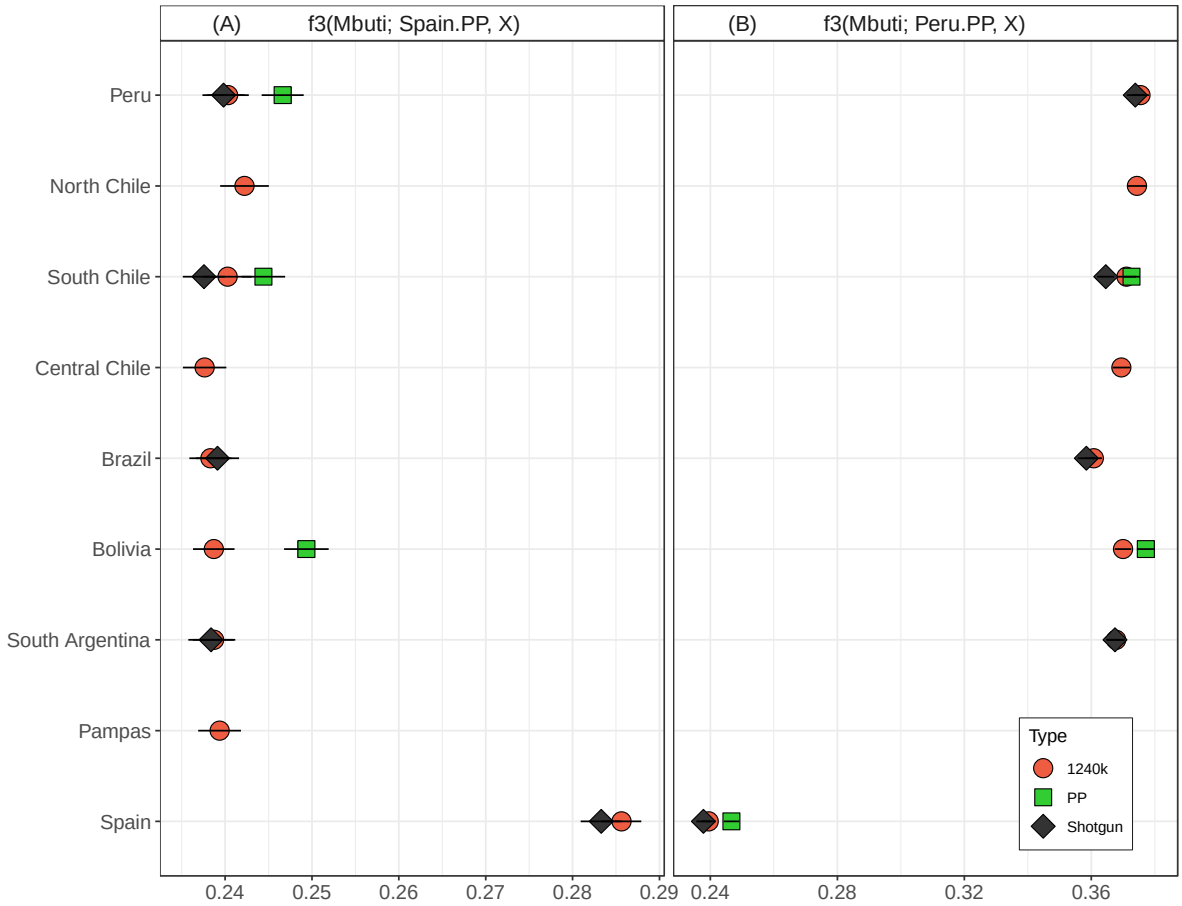


**Figure 1:** Outgroup $f_3$ statistics comparing ancient South American and Spanish populations by data type. Panels (A) and (B) show results for Spain.PP.2rnd and Peru.PP.2rnd as the target population, respectively. Note that not all regions have data of every type and different data types from the same region are not the same individuals but are expected to share broadly similar ancestries.

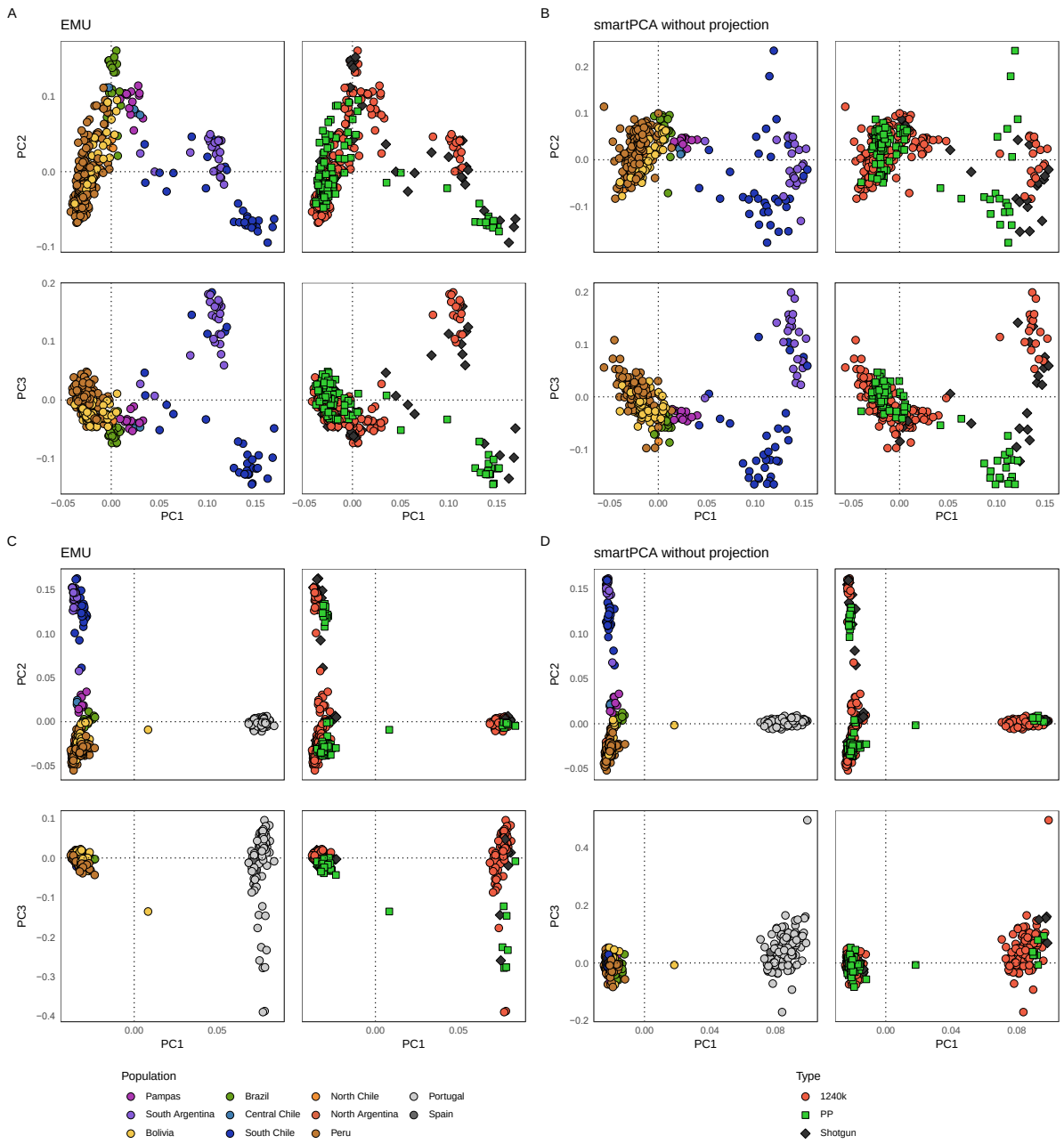


**Figure 2:** Comparing PCA methods at continental and global scales. Continental PCA (A and B) uses a collated dataset of 362 ancient genomes from South America. Global PCA (C and D) uses a collated dataset of 399 ancient genomes from South America and Iberia. Samples are coloured by region (left column of each panel) or data type (right column of each panel). (A) Continental PCA computed with EMU (Meisner et al., 2021). (B) Continental PCA computed with smartPCA, no projection (Patterson et al., 2006; Price et al., 2006). (C) Global PCA computed with EMU (Meisner et al., 2021). (D) Global PCA computed with smartPCA, no projection (Patterson et al., 2006; Price et al., 2006).

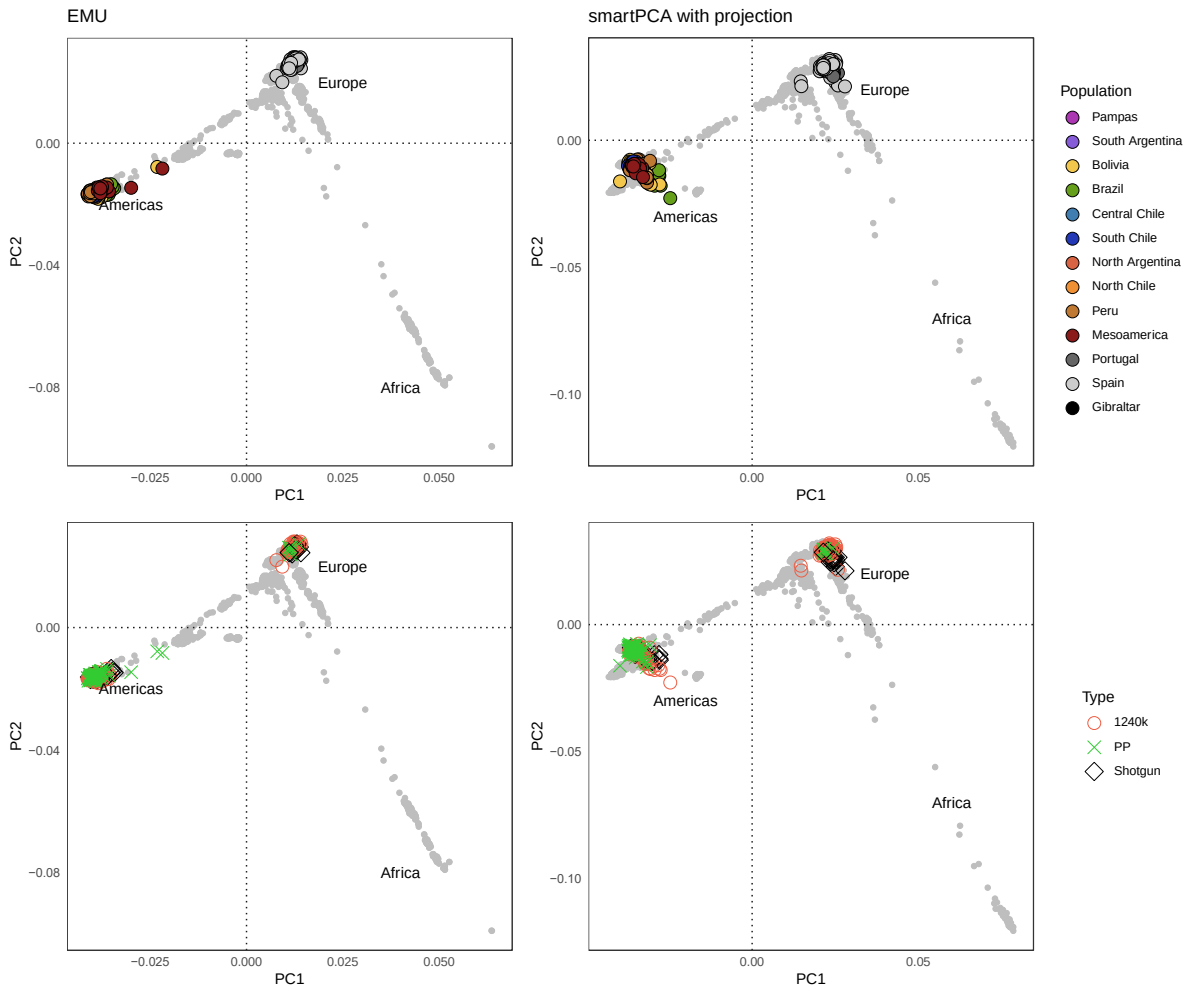


**Figure 3:** PCAs computed with EMU (left) and smartPCA (right) of ancient genomes from South America, Central America, and Iberia coloured by region (top row) or data type (bottom row). Ancient samples are projected by smartPCA onto the modern diversity of the HGDP shown in grey (Bergström et al., 2020).

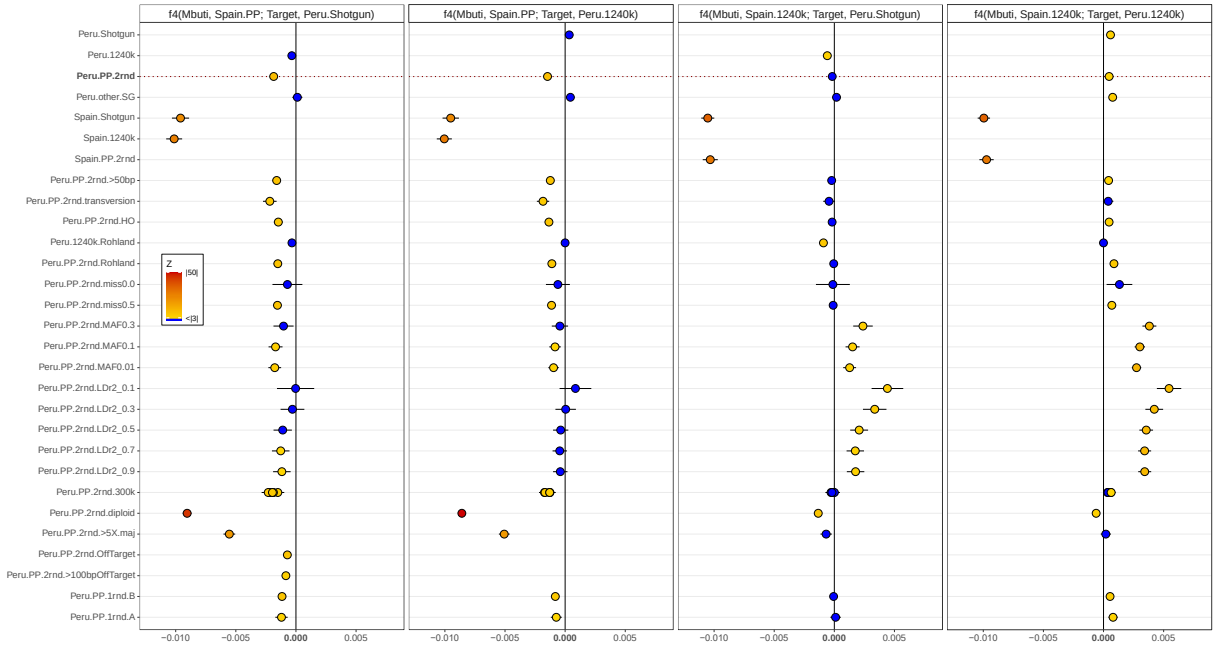


**Figure 4:** $f_4$ statistics showing the effectiveness of 20 bias-reduction strategies. Test configurations are, from left to right: $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.Shotgun), $f_4$(Mbuti, Spain.PP.2rnd; Target, Peru.1240k), $f_4$(Mbuti, Spain.1240k; Target, Peru.Shotgun), $f_4$(Mbuti, Spain.1240k; Target, Peru.1240k). "Target" groups are shown along the y-axis, being either one of the 29 strategies used to resolve the Prime Plus bias, or shotgun or 1240k data from Peru or Spain. Z-scores and the number of SNPs used in each test are recorded in Table S6. Non-significant $Z$-scores ($|Z| < 3$) are shown in blue, and significant $Z$-scores ($|Z| > 3$) are indicated with a yellow-red gradient. Error bars represent ± 2 se. The vertical dashed line represents the observed $f_4$ statistic in each configuration when Peru.PP.2rnd was the Target group. Note that bias-reduction strategies are only applied to the target population and the $f_4$ statistic is calculated from SNPs overlapping between the target and the other populations that remain unchanged between tests. Note that in the $f_4$ statistics where the target population is off-target or out-of-range genotypes (Peru.PP.2rnd.OffTarget, Peru.PP.2rnd.>100bpOffTarget), all four populations were genotyped in the respective alternative way to ensure overlapping SNPs, and therefore these are not exactly comparable.

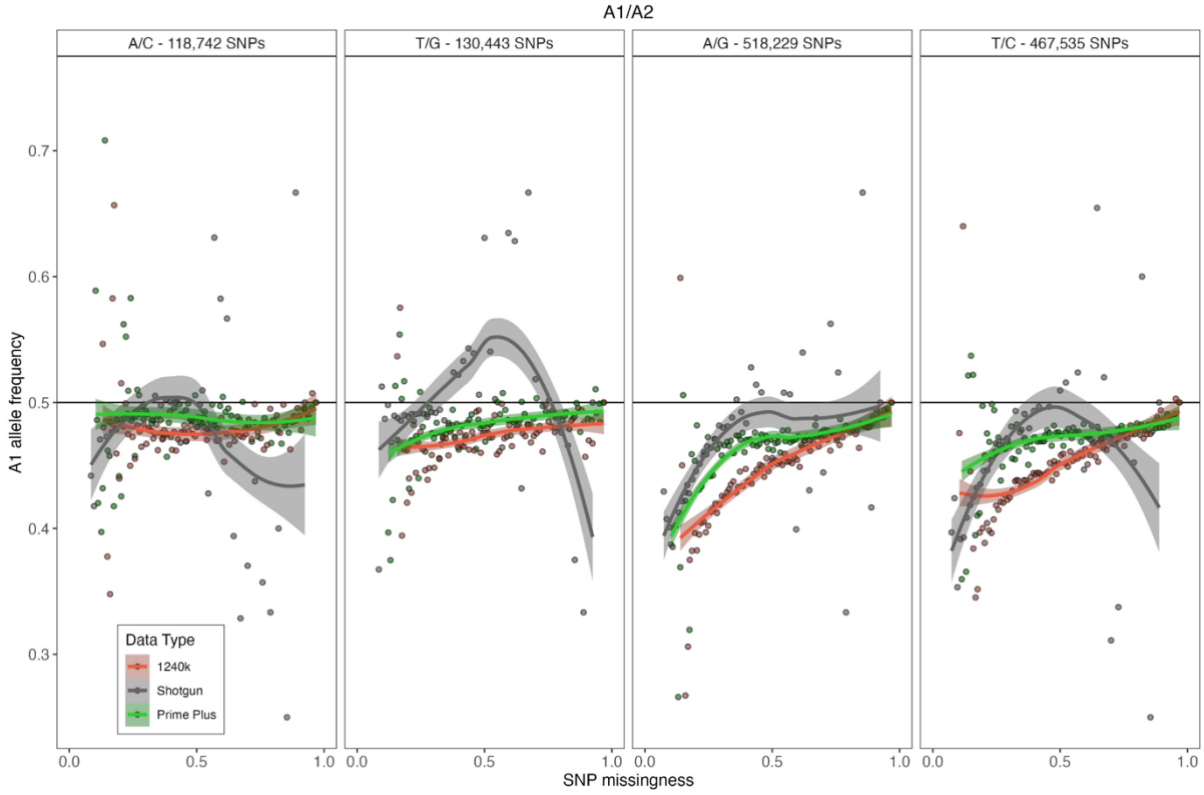


**Figure 5**: Mean frequency of A1 allele calls in different data types as a function of SNP missingness. Results for the four different A1 allele pairs are provided in separate panels, with inset data points showing the mean allele frequencies calculated across all individuals in successive SNP missingness bins. SNPs that are fixed within each dataset were removed to avoid skewing the mean A1 calls. The header of each panel shows the number of target SNPs of the corresponding A1/A2 pair observed in the Prime Plus assay.