

1 **bistro: An R package for vector bloodmeal identification by short tandem repeat overlap**

2 Zena Lapp¹, Lucy Abel², Judith Mangeni³, Andrew A. Obala⁴, Wendy O'Meara^{1,5}, Steve M.
3 Taylor^{1,5}, Christine F. Markwalter¹

4
5 ¹Duke Global Health Institute, Duke University, Durham, NC, USA

6 ²Academic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya

7 ³Department of Epidemiology and Medical Statistics, School of Public Health, Moi University,
8 Eldoret, Kenya

9 ⁴School of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya

10 ⁵Division of Infectious Diseases, School of Medicine, Duke University, Durham, NC, USA

11

12 **Corresponding author:** Christine Markwalter

13 310 Trent Dr., Durham NC, 27710; christine.markwalter@duke.edu

14

15 **Keywords:** bloodmeal, bloodmeal matching, DNA matching, malaria, match, short tandem
16 repeat, str, vector-borne disease

17

18 **Abstract**

19 1. Measuring vector-human contact in a natural setting can inform precise targeting of
20 interventions to interrupt transmission of vector-borne diseases. One approach is to
21 directly match human DNA in vector bloodmeals to the individuals who were bitten using
22 genotype panels of discriminative short tandem repeats (STRs). Existing methods for
23 matching STR profiles in bloodmeals to the people bitten preclude the ability to match
24 most incomplete profiles and multi-source bloodmeals to bitten individuals.

25

26 2. We developed bistro, an R package that implements 3 preexisting STR matching
27 methods as well as the package's namesake, bistro, a new algorithm described here.
28 bistro employs forensic analysis methods to calculate likelihood ratios and match human
29 STR profiles in bloodmeals to people using a dynamic threshold. We evaluated the
30 algorithm's accuracy and compared it to existing matching approaches using a publicly-
31 available panel of 188 single-source and 100 multi-source samples containing DNA from
32 50 known human sources. Then we applied it to match 777 newly field-collected
33 mosquito bloodmeals to a database of 645 people.

34

35 3. The R package implements four STR matching algorithms in user-friendly functions with
36 clear documentation. bistro correctly matched 99% (184/185) of profiles in single-source
37 samples, and 63% (225/359) of profiles from multi-source samples, resulting in a
38 sensitivity of 0.75 (vs < 0.51 for other algorithms). The specificity of bistro was 0.9998
39 (vs. 1 for other algorithms). Furthermore, bistro identified 80% (729/909) of all possible
40 matches for field-derived mosquitoes, yielding 1.4x more matches than existing
41 algorithms.

42

43 4. bistro identifies more correct bloodmeal-human matches than existing approaches,
44 enabling more accurate and robust analyses of vector-human contact in natural settings.
45 The bistro R package and corresponding documentation allow for straightforward uptake
46 of this algorithm by others.

47

48

49 **Introduction**

50 Vector-borne diseases cause over 700,000 deaths each year (WHO, 2020). Understanding
51 vector biting behavior in a natural setting can inform precise targeting of interventions to
52 efficiently interrupt transmission. One approach is to identify human factors associated with
53 increased vector biting by matching the human DNA in vector bloodmeals to the individuals who
54 were bitten. Analogous to forensic DNA analysis, this can be done by genotyping panels of
55 short tandem repeats (STRs) in a human population and in vector bloodmeals.

56

57 STR genotyping has been used to profile human bloodmeals in sandflies (Inbar et al., 2016),
58 lice (Mumcuoglu et al., 2004), and mosquitoes (Soremekun et al., 2004; Gonçalves et al., 2017;
59 Keven et al., 2021; Mbewe et al., 2023). Existing methods for matching STR profiles in
60 bloodmeals to the people bitten require exact, or near-exact, matching of all alleles at all loci
61 between a bloodmeal and the person bitten. This precludes the ability to match most incomplete
62 locus profiles, and any multi-source bloodmeals with more than two alleles at a locus, to bitten
63 individuals. While there will inevitably be incomplete and multi-source bloodmeals that cannot
64 be confidently matched to any person, for forensic applications, there are robust and well-
65 established methods for matching evidence from incomplete and multi-source samples to
66 persons of interest. These methods provide a likelihood ratio that the DNA of a person of
67 interest is in the evidence sample (vs. a random contributor), which for legal proceedings is
68 typically interpreted with extremely high stringency. Less stringency is needed for matching
69 vector bloodmeals to bitten individuals; in this research context, a small sacrifice in specificity,
70 which would be unacceptable for forensic analysis, can provide valuable gains in sensitivity that
71 lead to useful insights into disease transmission epidemiology.

72

73 We adapted forensic analytic approaches to develop bistro (bloodmeal identification by STR
74 overlap), a user-friendly R package that enables matching of bloodmeal STR profiles to people
75 based on log₁₀ likelihood ratios (log₁₀LRs). The underlying algorithm attempts to match not
76 only complete STR profiles from single-source bloodmeals, but also incomplete STR profiles
77 and multi-source bloodmeals. We evaluated the algorithm's efficacy using samples containing
78 human DNA from known sources, and applied it to mosquito bloodmeals from unknown
79 sources.

80

81 **Matching algorithms**

82 While matching people to bloodmeals is not a new concept, to our knowledge no software
83 packages exist that provide user-friendly implementations of bloodmeal-human matching
84 algorithms. Thus, the bistro R package implements not only our newly proposed bistro
85 algorithm, but also three pre-existing methods for matching STR profiles: exact (Soremekun et
86 al., 2004), similarity (Keven et al., 2021; Mbewe et al., 2023), and static threshold (Bleka et al.,
87 2019) (**Table 1**). Exact and similarity matching have been used to match mosquito bloodmeals
88 to people bitten, while static threshold matching is commonly used in forensics. Below, we
89 describe each of the algorithms before explaining their implementation.

90

91 **Table 1: Comparison of matching algorithms**

Method	Match samples with incomplete profiles	Match samples predicted to be multi-source	Threshold	Use case	Reference
Exact	No	No	Complete identity	Bloodmeal matching	(Soremekun et al., 2004)
Similarity	Limited	No	Highest human similarity	Bloodmeal matching	(Keven et al., 2021; Mbewe et al., 2023)
Static threshold	Yes	Yes	Static log ₁₀ LR	Forensics	(Bleka et al., 2017, 2019)
Bistro	Yes	Yes	Dynamic log ₁₀ LR	Bloodmeal matching	This paper

92

93 **Exact matching**

94 The complete sample STR profile must match the complete human profile.

95

96 **Similarity matching**

97 The proportion of exact locus matches between each sample and human is identified, and any

98 similarities above maximum pairwise similarity values of the human database are matches.

99

100 **Static threshold matching**

101 Matches are bloodmeal-human pairs with a $\log_{10}LR$ above a user-defined threshold.

102

103 **bistro matching**

104 To identify human contributors to a bloodmeal, bistro uses pairwise bloodmeal-human $\log_{10}LR$ s

105 (**Figure 1**). The $\log_{10}LR$ s compare the likelihood that the DNA of a person is in the bloodmeal

106 sample (vs. a random contributor). To be considered a potential match, the likelihood ratio must

107 be above 1.5 (i.e. 30x greater likelihood). For each bloodmeal with at least 1 $\log_{10}LR \geq 1.5$, we

108 initialize a set of matches using a minimum $\log_{10}LR$ threshold of $\lfloor \max(\log_{10}LR * 2) \rfloor / 2$. We then

109 decrement the threshold by 0.5 until the number of matches is \geq the estimated number of

110 contributors (NOC) and the matches identified by the current threshold are identical to the

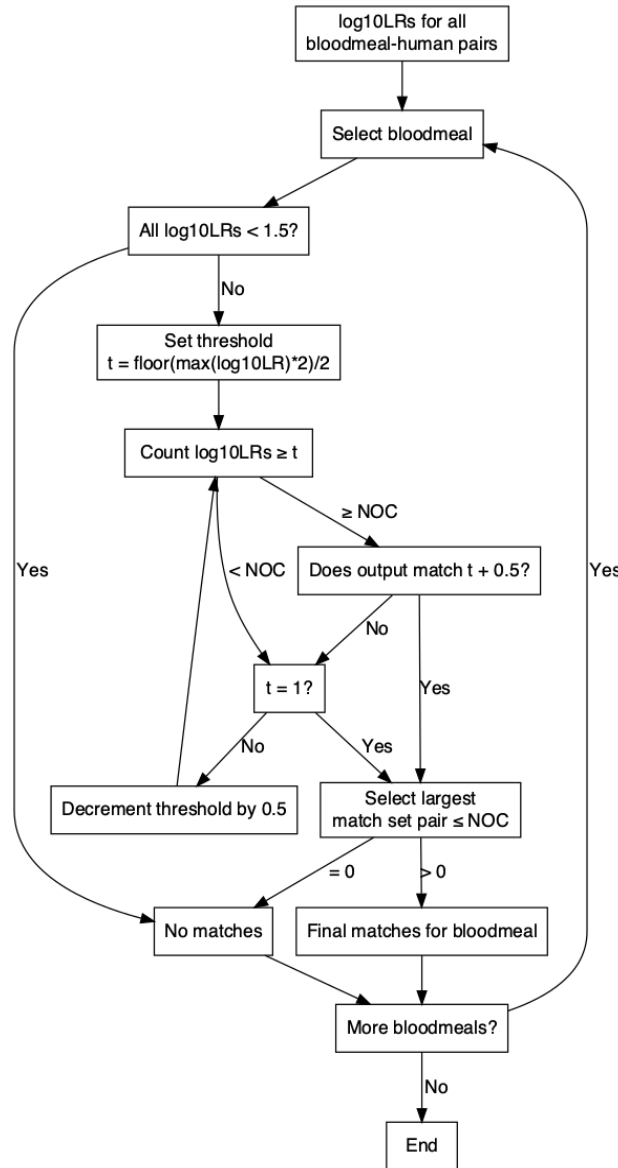
111 matches identified by the previous threshold. The most confident match set is considered to be

112 the one that was identified by at least two thresholds and where the number of bloodmeal-

113 human pairs is maximized under the constraint that there are \leq NOC pairs. The algorithm returns

114 no bloodmeal-human pairs if no confident match sets are identified by a $\log_{10}LR$ threshold = 1,

115 or if all match sets have $>$ NOC pairs.



116

117 **Figure 1: bistro matching algorithm for identifying matches between bloodmeals and**
118 **people.** log10LRs = log10 likelihood ratios; t = threshold; NOC = estimated number of human
119 contributors.

120

121 **bistro R package**

122 The open-source bistro R package hosted on GitHub ([https://github.com/duke-malaria-](https://github.com/duke-malaria-collaboratory/bistro)
123 [collaboratory/bistro](https://github.com/duke-malaria-collaboratory/bistro)), implements the matching algorithms described above and provides
124 examples using a built-in dataset. Documentation on the package website ([https://duke-malaria-](https://duke-malaria-collaboratory.github.io/bistro/index.html)
125 [collaboratory.github.io/bistro/index.html](https://duke-malaria-collaboratory.github.io/bistro/index.html)) describes step-by-step instructions for installing the

126 package, running the algorithms, and understanding the output. Below, we briefly describe the
127 implementation of the bistro algorithm.

128

129 The bistro function takes as input bloodmeal STR profiles, human STR reference profiles, a
130 bloodmeal STR allele peak detection threshold, the STR genotyping kit name, and optionally
131 human population allele frequencies. Then, it performs the following steps:

- 132 1. If no human population allele frequencies are supplied, compute them from the provided
133 human STR profiles.
- 134 2. Remove bloodmeal STR alleles below the user-provided peak threshold.
- 135 3. Estimate the NOC for each bloodmeal using the equation: $\lceil \max(a)/2 \rceil$, where a is the
136 number of alleles at a locus.
- 137 4. Calculate log10LRs for each bloodmeal-human pair using `euroformix::contLikSearch()`
138 (`Oyvind Bleka, 2023`) with the parameter defaults in **Table 2**.
- 139 5. Use the bistro matching algorithm described above to identify, from the provided human
140 reference database, contributors to each bloodmeal.

141

142 **Table 2: bistro parameters for `euroformix::contLikeSearch()`**

contLikSearch parameter	Description	bistro default value
threshT	peak detection threshold in RFU	user-provided
NOC	number of contributors	min(NOC, 3)
Difftol	tolerance of difference in log likelihood across 2 iterations	1
modelDegrad	whether to model degradation	TRUE
modelBWstutt	whether to model backward stutter	FALSE
modelFWstutt	whether to model forward stutter	FALSE
prC	allele drop-in probability	0.05
Lambda	parameter in modeled peak height shifted exponential model	0.01
nDone	number of optimizations required providing the same log-likelihood	2
Steptol	minimum allowable relative step length	0.001

143

144 The output of the bistro function is a dataset with matches for each bloodmeal, including why
145 there was or was not a match for each.

146

147 The current version of bistro (v0.1.1) requires R ($\geq 4.0.0$) (R Core Team, 2021), depends on
148 euroformix ($\geq 4.0.7$) (Oyvind Bleka, 2023); and imports codetools ($\geq 0.2.19$) (Tierney, 2023),
149 R.utils ($\geq v2.12.2$) (Bengtsson, 2022), and dplyr ($\geq 1.1.3$), stringr ($\geq 1.5.0$), tibble ($\geq 3.2.1$), and
150 tidyr ($\geq 1.3.0$) (Wickham et al., 2019).

151

152 **Snakemake workflow**

153 We also provide a template Snakemake (Köster & Rahmann, 2012) workflow that enables
154 straightforward parallelization of match identification using bistro ([https://github.com/duke-](https://github.com/duke-malaria-collaboratory/bistro_pipeline)
155 [malaria-collaboratory/bistro_pipeline](https://github.com/duke-malaria-collaboratory/bistro_pipeline)).

156

157 **Analysis methods**

158 **Ethics**

159 The human and mosquito data in this study were collected under protocols approved by the
160 IRBs of Duke (Pro00082000) and Moi (IREC\2017\36) Universities.

161

162 **Publicly available STR profiles**

163 To evaluate bistro, we used data from the public PROVEDIt database, which provides single-
164 and multi-source, complete and incomplete STR profiles derived from 50 reference individuals
165 (Alfonse et al., 2018). Ratios of sample material from each contributor to multi-source profiles
166 were pre-specified in the dataset. We used profiles originating from untreated samples
167 evaluated with the Identifiler Plus STR kit (28 cycles) on a 3130 Genetic Analyzer with a 10s
168 injection. We evaluated samples with ≥ 0.1 ng template mass and applied a peak height
169 threshold of 200 relative fluorescence units (RFUs), similar to the criteria applied to STR profiles

170 from mosquito bloodmeals below. Population allele frequencies from the revised NIST
171 Population dataset of Autosomal STRs (Hill et al., 2013; Steffen et al., 2017) were used.

172

173 **Bloodmeal and human STR profiles from a malaria-endemic region**

174 Indoor resting mosquitoes were collected by aspiration weekly from July 2020 to September
175 2021 in 75 households across 5 villages in Webuye, Kenya. Bloodfed abdomens (n=1065) were
176 separated from head/thoraces and pressed in filter paper to preserve bloodmeals for STR
177 analysis. Dried blood spot samples (Whatman Protein Saver 903) were collected from each
178 individual who slept in the study households during the study period.

179

180 Anopheles bloodmeals and human DBS were extracted using Chelex-100 (Markwalter et al.,
181 2022). Briefly, blood spots were soaked overnight at 4°C in 0.5% saponin before two extractions
182 in 150µl of 6.7% chelex in a 95°C water bath. All extracts were cleaned with Zymo RNA Clean &
183 Concentrator-96 kit using Zymo XL-96 filter plates. Bloodmeal extracts were cleaned and
184 concentrated twice. A template volume of 1µl (human DBS) or 3.5µl (bloodmeal) was added to
185 10µl amplification reactions using the 10-locus Promega Geneprint10 kit (Promega Cat
186 #B9510). Products (1µl) were analyzed on an Applied Biosystems 3730xl DNA Analyzer with
187 0.5x injection with a 600 internal lane standard. Alleles were called using Osiris v2.15.1 (NCBI
188 et al., 2021) with a 200 RFU threshold, which eliminated artifacts from no-template control
189 samples.

190

191 Population allele frequencies were calculated among people from the cohort (n=645) to more
192 closely represent the allele frequencies of the population present in bloodmeals.

193

194

195

196 **Data analysis and visualization**

197 We compared the four matching algorithms described above for each dataset. For static
198 threshold matching, we used a threshold of 10. The AMEL marker was not used for methods
199 requiring log₁₀LRs. Data were analyzed and visualized using R v4.2.1 (R Core Team, 2021)
200 and RStudio v2022.12.0+353 (*RStudio*, 2023) with the following packages: tidyverse v2.0.0
201 (Wickham et al., 2019), ggpubr v0.6.0 (Kassambara, 2018), ggupset v0.3.0 (Ahlmann-Eltze,
202 2020). Data analysis scripts are on GitHub: [https://github.com/duke-malaria-](https://github.com/duke-malaria-collaboratory/bistro_validation)
203 [collaboratory/bistro_validation](https://github.com/duke-malaria-collaboratory/bistro_validation).

204

205 **Results**

206

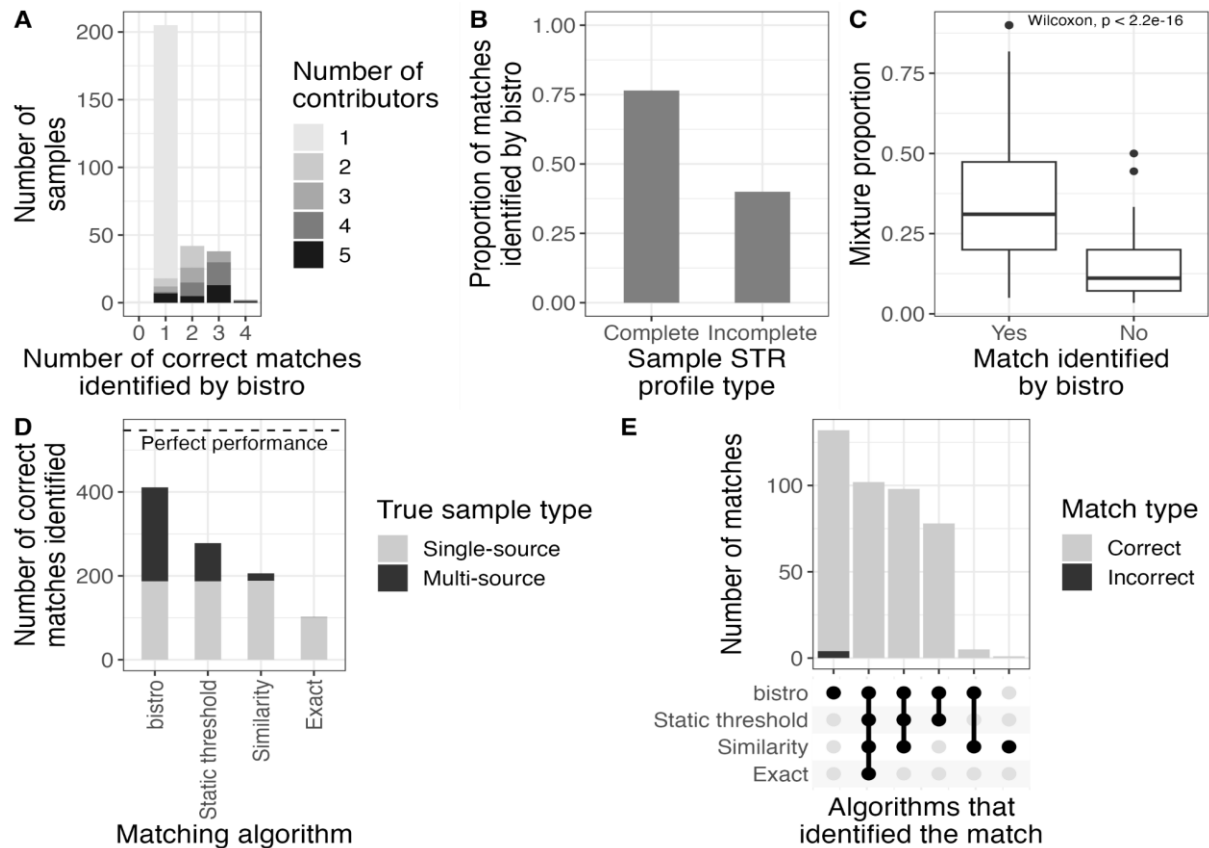
207 **bistro efficacy**

208 We first evaluated the efficacy of bistro by using it to match people to samples with known
209 sources (**Figure 2A, Table 3**): 188 single-source samples (n=188 sources) and 100 multi-
210 source samples (n=359 sources). For single-source samples, 99% (187/188) of matches were
211 correctly identified, including 3/3 samples with incomplete profiles. For multi-source samples we
212 identified at least one source in each sample and 62% (224/359) of all possible matches,
213 including 5/17 matches from incomplete profiles. Across all samples, those with complete
214 profiles had a higher matching rate (76%) than those with incomplete profiles (40%; **Figure 2B**).
215 Furthermore, within samples, DNA present at higher proportions matched more frequently than
216 DNA present in lower proportions (**Figure 2C**; Wilcoxon $p < 0.001$).

217

218 Next, we compared bistro to three other algorithms used for matching: exact, similarity
219 (similarity threshold = 0.375), and static threshold (**Figure 2D-E, Table 3**). bistro returned 4
220 false positives (specificity = 0.9998), while all other algorithms returned zero false positives
221 (specificity = 1). However, bistro (411 matches, sensitivity=0.75) identified 4x more matches

222 than exact (102 matches, sensitivity=0.19), 2x more matches than similarity (206 matches,
223 sensitivity=0.38), and 1.5x more matches than static threshold matching (278 matches,
224 sensitivity=0.51), yielding much higher sensitivity.
225



226

227 **Figure 2: Evaluating bistro using samples with known sources.** (A) Number of correct
228 matches for each sample colored by known number of contributors (NOC). (B) Proportion of
229 matches identified separated by whether the STR profile was complete (15 loci present) or not
230 (<15 loci present). The denominator is the total number of possible matches (m_{tot} , 547). (C) For
231 multi-source samples, the original DNA mixture proportion of people successfully matched vs.
232 not matched. (D) Total number of matches identified by each algorithm. The dotted line is m_{tot} .
233 Color indicates single- vs. multi-source sample based on the known NOC. (E) Number of
234 matches identified by each algorithm.
235

236

237

238

239

239 **Table 3: Comparison of matching algorithm performance on samples with known**
240 **sources**

Algorithm	Single-source	Multi-source		All		
	# correct sources identified	# samples with ≥ 1 correct source identified	# correct sources identified	# samples with ≥ 1 correct source identified	# correct sources identified	# incorrect matches
Bistro	187/188 (99%)	100/100 (100%)	224/359 (62%)	287/288 (99.7%)	411/547 (75%)	4
Static threshold	187/188 (99%)	68/100 (68%)	91/359 (25%)	255/288 (89%)	278/547 (51%)	0
Similarity	188/188 (100%)	18/100 (18%)	18/359 (5%)	206/288 (72%)	206/547 (38%)	0
Exact	101/188 (54%)	1/100 (1%)	1/359 (0.28%)	102/288 (35%)	102/547 (19%)	0

241

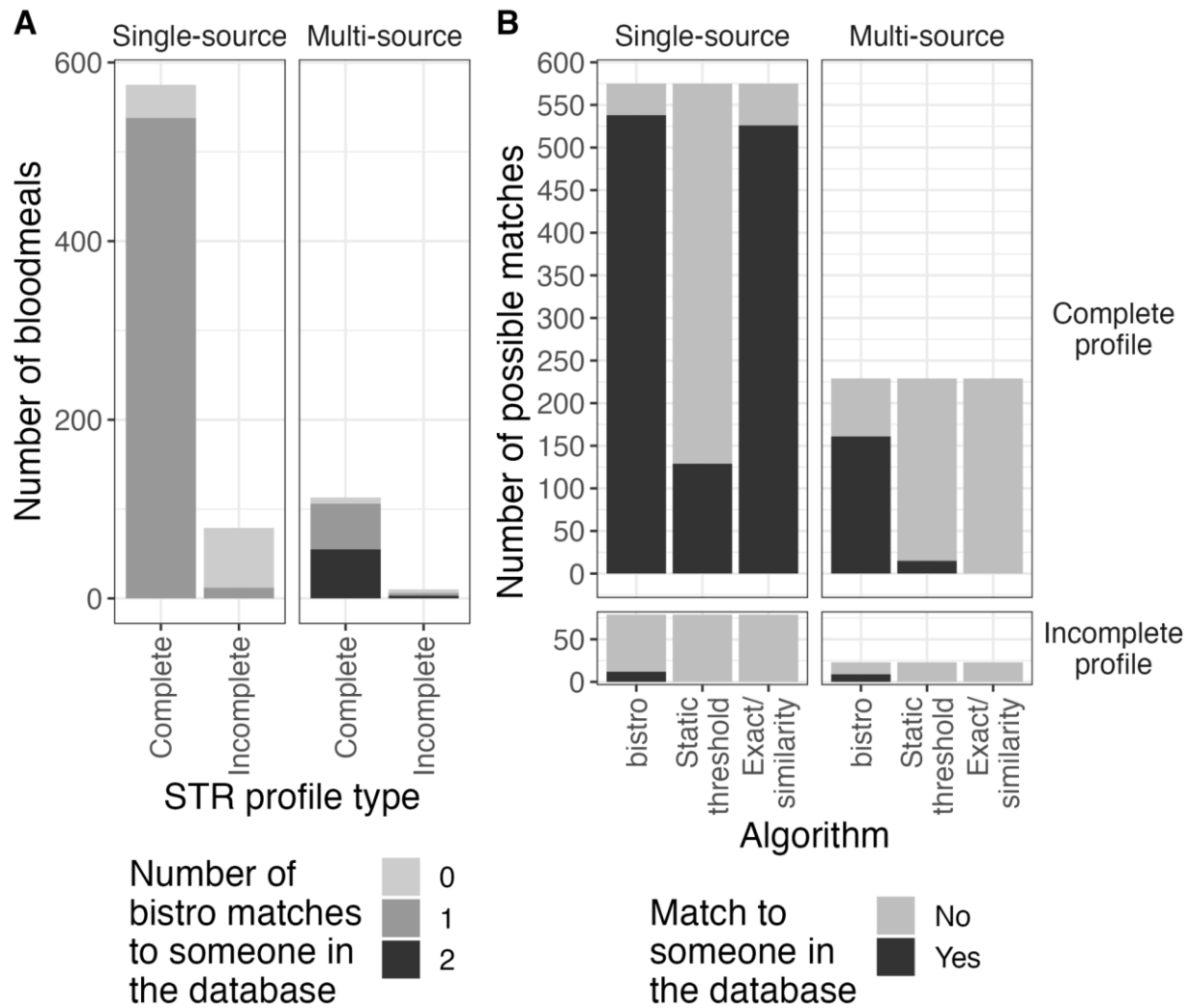
242 **bistro on mosquito bloodmeals**

243 We next applied bistro to identify matches between a freshly-collected database of human
244 samples (n=645) from Western Kenya and mosquito bloodmeals (n=1065) collected from their
245 households (**Figure 3A, Table 4**). About a quarter (288) of bloodmeals returned no usable
246 peaks from STR typing. Among the 777 bloodmeals that returned peaks, 654 (84%) were
247 estimated to be single-source (n=654 human sources), and 123 (16%) were estimated to be
248 multi-source (n=252 human sources), totalling 906 estimated human sources.

249

250 bistro identified a match to an individual in the database for 94% of complete (644/688) and
251 20% of incomplete (18/89) bloodmeals. Furthermore, bistro identified two matches for 47%
252 (58/123) of polyallelic profiles thought to represent multi-source bloodmeals. Therefore, we were
253 able to identify 79% (720/906) of all human sources of bloodmeals. In comparison, exact and
254 similarity matching, which are equivalent to each other for these data due to a high maximum
255 similarity value between cohort members (0.9), were only able to identify 58% (526/906) of
256 matches, and static threshold matching was able to identify 16% (144/906) (**Figure 3B, Table**

257 4). Therefore, bistro identified 1.39x more matches than exact/similarity and 5x more matches
258 than static threshold matching.
259



260

261 **Figure 3: Application of matching algorithms to mosquito bloodmeals.** (A) Number of
262 matches identified by bistro for each bloodmeal by single- vs. multi-source bloodmeal (based on
263 estimated NOC) and STR profile completeness. (B) Comparison of matches identified by
264 different matching algorithms by profile completeness and NOC. For this dataset, exact and
265 similarity matching return the same results because the highest similarity value in the cohort
266 was 9/10, requiring an exact match with all 10 STR-typed loci.

267

268

269

270 **Table 4: Comparison of matching algorithm performance on Anopheles bloodmeal**
271 **samples**

Algorithm	Predicted single-source	Predicted multi-source		All	
	# sources identified	# samples with ≥ 1 source identified	# sources identified	# samples with ≥ 1 source identified	# sources identified
bistro	550/654 (84%)	112/123 (91%)	170/252 (67%)	669/780 (86%)	720/906 (79%)
Static threshold	129/654 (20%)	15/123 (12%)	15/252 (6%)	530/780 (68%)	144/906 (16%)
Exact/similarity	526/654 (80%)	0/123 (0%)	0/252 (0%)	526/780 (67%)	526/906 (58%)

272

273

274 **Discussion: strengths and limitations of different matching algorithms**

275 We implemented one novel (bistro) and 3 preexisting (exact, similarity, static threshold)
276 matching algorithms in bistro, an easy-to-use R package. Each of the matching algorithms
277 tested have different capabilities and corresponding strengths and limitations. The benefit of
278 exact matching is that it provides extreme confidence in matches and is not computationally
279 intensive. However, many likely matches are missed with this method because it requires that
280 the estimated NOC equals one and that there are no missing or additional peaks in the
281 bloodmeal STR profile. Similarity matching improves upon this by allowing some non-matching
282 peaks, as long as the proportion of loci that match is greater than the maximum similarity
283 between two STR profiles in a human database. One limitation of this method is that it assumes
284 you have a comprehensive database of who could have been bitten, which may not always be
285 available. Furthermore, as in the bloodmeal data presented here, when cohort members have
286 highly related profiles or when fewer loci are typed, similarity matching becomes identical to
287 exact matching. Static threshold matching, as done in forensic analysis, still returns zero false
288 positives, but nevertheless results in many missing matches that have high log10LRs relative to

289 other samples in the reference database. Additionally, the log₁₀LR magnitude depends on the
290 number of loci typed, so this approach will likely miss more correct matches when evaluating
291 STR panels with fewer loci. To circumvent the limitations of these preexisting algorithms for
292 bloodmeal matching applications, bistro builds upon methods originally developed for forensic
293 analysis to relax the stringency of matches to a degree suitable for research purposes. Since
294 matching bloodmeals in the context of scientific research does not require such high levels of
295 certainty as in forensics, after calculating likelihood ratios for each mosquito-human pair, bistro
296 determines an optimal log₁₀LR threshold for each individual mosquito. This avoids missing true
297 matches with lower log₁₀LRs. While this introduces some false positives, specificity remains
298 high while increasing sensitivity relative to the other algorithms described here. Thus, bistro
299 allows matching of more bloodmeals to people bitten than preexisting methods, providing more
300 matches for downstream analyses related to vector biting and disease transmission. A user-
301 friendly implementation of all four algorithms is provided as an R package with extensive
302 documentation.

303

304 **Acknowledgements**

305 This work was supported by the National Institute of Allergy and Infectious Diseases
306 [R01AI146849 to W.P.O. and S.M.T.]. C.F.M. was supported by F32AI149950 and
307 K01AI175527. We appreciate the study implementation skills of project managers and field
308 technicians in Webuye and Eldoret: J. Kipkoech Kirui, I. Khaoya, L. Marango, E. Mukeli, E.
309 Nalianya, J. Namae, L. Nukewa, E. Wamalwa, and A. Wekesa. We thank S. Langdon (DNA
310 Analysis Core, Department of Immunology), B. Freedman, J. Grassia, J. DeCurzio, and T.
311 Thane (each of Duke University) for their help with laboratory samples. Ultimately, we are
312 indebted to the people who participated in our study.

313

314

315 **Data availability**

316 PROVEDIt data are available at <https://doi.org/10.1016/j.fsigen.2017.10.006> (Alfonse et al.,
317 2018). Kenyan STR profiles are unshareable.

318

319 **Conflict of interest**

320 The authors declare no conflicts of interest.

321

322 **Author contributions**

323 CM, ZL, ST, and WPO conceptualized the project. CM and ZL performed methodology and
324 software development, formal analysis, investigation, data curation, original draft writing, and
325 visualization. LA managed the project. JM, AA, WPO, and ST supervised the project and
326 acquired funding. All authors reviewed and edited the manuscript. Our study includes scientists
327 based in the country where the example mosquito samples were collected. All authors
328 contributed intellectually to the research and study design. We cited literature published by
329 scientists from the region.

330

331

332

333 **References**

- 334 Ahlmann-Eltze, C. (2020). *ggupset: Combination Matrix Axis for “ggplot2” to Create “UpSet”*
335 *Plots* (0.3.0) [Computer software].
336 <https://cran.rstudio.com/web/packages/ggupset/index.html>
- 337 Alfonse, L. E., Garrett, A. D., Lun, D. S., Duffy, K. R., & Grgicak, C. M. (2018). A large-scale
338 dataset of single and mixed-source short tandem repeat profiles to inform human
339 identification strategies: PROVEDIt. *Forensic Science International: Genetics*, 32, 62–
340 70. <https://doi.org/10.1016/j.fsigen.2017.10.006>
- 341 Bengtsson, H. (2022). *R.utils: Various Programming Utilities* (2.12.2) [Computer software].
342 <https://cran.r-project.org/web/packages/R.utils/index.html>
- 343 Bleka, Ø., Bouzga, M., Fonnelløp, A., & Gill, P. (2017). dnamatch2: An open source software to
344 carry out large scale database searches of mixtures using qualitative and quantitative
345 models. *Forensic Science International: Genetics Supplement Series*, 6, e404–e406.
346 <https://doi.org/10.1016/j.fsigss.2017.09.160>
- 347 Bleka, Ø., Prieto, L., & Gill, P. (2019). CaseSolver: An investigative open source expert system
348 based on EuroForMix. *Forensic Science International: Genetics*, 41, 83–92.
349 <https://doi.org/10.1016/j.fsigen.2019.03.024>
- 350 Gonçalves, B. P., Kapulu, M. C., Sawa, P., Guelbéogo, W. M., Tiono, A. B., Grignard, L., Stone,
351 W., Hellewell, J., Lanke, K., Bastiaens, G. J. H., Bradley, J., Nébié, I., Ngoi, J. M.,
352 Oriango, R., Mkabili, D., Nyaurah, M., Midega, J., Wirth, D. F., Marsh, K., ... Bousema,
353 T. (2017). Examining the human infectious reservoir for *Plasmodium falciparum* malaria
354 in areas of differing transmission intensity. *Nature Communications*, 8, 1133.
355 <https://doi.org/10.1038/s41467-017-01270-4>
- 356 Hill, C. R., Duewer, D. L., Kline, M. C., Coble, M. D., & Butler, J. M. (2013). U.S. population data
357 for 29 autosomal STR loci. *Forensic Science International. Genetics*, 7(3), e82-83.
358 <https://doi.org/10.1016/j.fsigen.2012.12.004>

- 359 Inbar, E., Lawyer, P., Sacks, D., & Podini, D. (2016). The Potential Use of Forensic DNA
360 Methods Applied to Sand Fly Blood Meal Analysis to Identify the Infection Reservoirs of
361 Anthroponotic Visceral Leishmaniasis. *PLOS Neglected Tropical Diseases*, 10(5),
362 e0004706. <https://doi.org/10.1371/journal.pntd.0004706>
- 363 Kassambara, A. (2018). *ggpubr: "ggplot2" Based Publication Ready Plots* (0.2) [Computer
364 software]. <https://CRAN.R-project.org/package=ggpubr>
- 365 Keven, J. B., Katusese, M., Vinit, R., Rodríguez-Rodríguez, D., Hetzel, M. W., Robinson, L. J.,
366 Laman, M., Karl, S., Foran, D. R., & Walker, E. D. (2021). Nonrandom Selection and
367 Multiple Blood Feeding of Human Hosts by Anopheles Vectors: Implications for Malaria
368 Transmission in Papua New Guinea. *The American Journal of Tropical Medicine and*
369 *Hygiene*, 105(6), 1747–1758. <https://doi.org/10.4269/ajtmh.21-0210>
- 370 Köster, J., & Rahmann, S. (2012). Snakemake—A scalable bioinformatics workflow engine.
371 *Bioinformatics*, 28(19), 2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>
- 372 Markwalter, C. F., Menya, D., Wesolowski, A., Esimit, D., Lokoel, G., Kipkoech, J., Freedman,
373 E., Sumner, K. M., Abel, L., Ambani, G., Meredith, H. R., Taylor, S. M., Obala, A. A., &
374 O'Meara, W. P. (2022). Plasmodium falciparum importation does not sustain malaria
375 transmission in a semi-arid region of Kenya. *PLOS Global Public Health*, 2(8),
376 e0000807. <https://doi.org/10.1371/journal.pgph.0000807>
- 377 Mbewe, R. B., Keven, J. B., Mangani, C., Wilson, M. L., Mzilahowa, T., Mathanga, D. P., Valim,
378 C., Laufer, M. K., Walker, E. D., & Cohee, L. M. (2023). Genotyping of Anopheles
379 mosquito blood meals reveals nonrandom human host selection: Implications for human-
380 to-mosquito Plasmodium falciparum transmission. *Malaria Journal*, 22, 115.
381 <https://doi.org/10.1186/s12936-023-04541-2>
- 382 Mumcuoglu, K. Y., Gallili, N., Reshef, A., Brauner, P., & Grant, H. (2004). Use of Human Lice in
383 Forensic Entomology. *Journal of Medical Entomology*, 41(4), 803–806.
384 <https://doi.org/10.1603/0022-2585-41.4.803>

385 NCBI, NLM, NIH, & US DHHS. (2021). *Osiris (Open Source Independent Review and*
386 *Interpretation System)* (2.15.1) [Computer software]. <https://www.ncbi.nlm.nih.gov/osiris/>

387 Oyvind Bleka. (2023). *euroformix: A Quantitative Model for Interpreting DNA Mixtures*
388 [Computer software]. <http://www.euroformix.com/>

389 R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation
390 for Statistical Computing. <https://www.R-project.org/>

391 *RStudio*. (2023). <https://www.rstudio.com/>

392 Soremekun, S., Maxwell, C., Zuwakuu, M., Chen, C., Michael, E., & Curtis, C. (2004).
393 Measuring the efficacy of insecticide treated bednets: The use of DNA fingerprinting to
394 increase the accuracy of personal protection estimates in Tanzania. *Tropical Medicine &*
395 *International Health*, 9(6), 664–672. <https://doi.org/10.1111/j.1365-3156.2004.01250.x>

396 Steffen, C. R., Coble, M. D., Gettings, K. B., & Vallone, P. M. (2017). Corrigendum to “U.S.
397 Population Data for 29 Autosomal STR Loci” [Forensic Sci. Int. Genet. 7 (2013) e82-
398 e83]. *Forensic Science International. Genetics*, 31, e36–e40.
399 <https://doi.org/10.1016/j.fsigen.2017.08.011>

400 Tierney, L. (2023). *codetools: Code Analysis Tools for R* (0.2-19) [Computer software].
401 <https://cran.r-project.org/web/packages/codetools/index.html>

402 WHO. (2020, March). *Vector-borne diseases*. [https://www.who.int/news-room/fact-](https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases)
403 [sheets/detail/vector-borne-diseases](https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases)

404 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemond, G.,
405 Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K.,
406 Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the
407 Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
408 <https://doi.org/10.21105/joss.01686>

409