Distribution of clones among hosts for the lizard malaria parasite *Plasmodium mexicanum*

Allison T Neal [Corresp. 1]

1 Department of Biology, Norwich University, Northfield, Vermont, United States

Corresponding Author: Allison T Neal
Email address: aneal1@norwich.edu

Background. Malaria parasites reproduce asexually, leading to the production of large numbers of genetically identical parasites, here termed a clonal line or clone. Infected hosts may harbor one or more clones, and the number of clones in a host is termed multiplicity of infection (MOI). Understanding the distribution of parasite clones among hosts can shed light on the processes shaping this distribution and is important for modeling MOI. Here, I determine whether the distribution of clones of the lizard malaria parasite *Plasmodium mexicanum* differ significantly from statistical distributions commonly used to model MOI and logical extensions of these models.

Methods. The number of clones per infection was assessed using four microsatellite loci with the maximum number of alleles at any one locus used as a simple estimate of MOI for each infection. I fit statistical models (Poisson, negative binomial, zero-inflated models) to data from four individual sites to determine a best fit model. I also simulated the number of alleles per locus using an unbiased estimate of MOI to determine whether the simple (but potentially biased) method I used to estimate MOI influenced model fit.

Results. The distribution of clones among hosts at individual sites differed significantly from traditional Poisson and negative binomial distributions, but not from zero-inflated modifications of these distributions. A consistent excess of two-clone infections and shortage of one-clone infections relative to all fit distributions was also observed. Any bias introduced by the simple method for estimating of MOI did not appear to qualitatively alter the results.

Conclusions. The statistical distributions used to model MOI are typically zero-truncated; truncating the Poisson or zero-inflated Poisson yield the same distribution, so the reasonable fit of the zero-inflated Poisson to the data suggests that the use of the zero-truncated Poisson in modeling is adequate. The improved fit of zero-inflated distributions relative to standard distributions may suggest that only a portion of the host population is located in areas suitable for transmission even at small sites (<1ha). Collective transmission of clones and premunition may also contribute to deviations from standard distributions.

1 # Distribution of clones among hosts for the lizard
2 # malaria parasite *Plasmodium mexicanum*

4 Allison Theresa Neal[1]

6 [1] Department of Biology, Norwich University, Northfield, Vermont, USA

8 Corresponding Author:
9 Allison Neal[1]
10 158 Harmon Drive, Northfield, VT, 05663, USA
11 Email address: aneal1@norwich.edu

Abstract

14 **Background.** Malaria parasites reproduce asexually, leading to the production of large numbers
15 of genetically identical parasites, here termed a clonal line or clone. Infected hosts may harbor
16 one or more clones, and the number of clones in a host is termed multiplicity of infection (MOI).
17 Understanding the distribution of parasite clones among hosts can shed light on the processes
18 shaping this distribution and is important for modeling MOI. Here, I determine whether the
19 distribution of clones of the lizard malaria parasite *Plasmodium mexicanum* differ significantly
20 from statistical distributions commonly used to model MOI and logical extensions of these
21 models.
22 **Methods.** The number of clones per infection was assessed using four microsatellite loci with
23 the maximum number of alleles at any one locus used as a simple estimate of MOI for each
24 infection. I fit statistical models (Poisson, negative binomial, zero-inflated models) to data from
25 four individual sites to determine a best fit model. I also simulated the number of alleles per
26 locus using an unbiased estimate of MOI to determine whether the simple (but potentially
27 biased) method I used to estimate MOI influenced model fit.
28 **Results**. The distribution of clones among hosts at individual sites differed significantly from
29 traditional Poisson and negative binomial distributions, but not from zero-inflated modifications
30 of these distributions. A consistent excess of two-clone infections and shortage of one-clone
31 infections relative to all fit distributions was also observed. Any bias introduced by the simple
32 method for estimating MOI did not appear to qualitatively alter the results.
33 **Conclusions**. The statistical distributions used to model MOI are typically zero-truncated;
34 truncating the Poisson or zero-inflated Poisson yield the same distribution, so the reasonable fit
35 of the zero-inflated Poisson to the data suggests that the use of the zero-truncated Poisson in
36 modeling is adequate. The improved fit of zero-inflated distributions relative to standard
37 distributions may suggest that only a portion of the host population is located in areas suitable for
38 transmission even at small sites (<1ha). Collective transmission of clones and premunition may
39 also contribute to deviations from standard distributions.

40

Introduction

41

42

43 It is not uncommon for hosts to harbor multiple parasites of the same species. For malaria
44 parasites (*Plasmodium* and related genera), detectable infections always consist of numerous
45 individual parasites. Many of these parasites are produced via asexual reproduction from a
46 single precursor cell in the insect vector (zygote/ookinete) and continue asexual replication in the
47 vertebrate host; all of these parasites are genetically identical (or nearly so) and are here referred
48 to collectively as a parasite clone. Vertebrate hosts may also harbor multiple parasite clones,
49 which are thought to be acquired primarily via bites from multiple infected insect vectors. The
50 distribution of malaria parasite clones among vertebrate hosts is thus shaped in part by the
51 transmission biology of these parasites. The distribution of clones is of practical interest because
52 interactions among co-infecting clones may influence health-related outcomes such as virulence,
53 transmissibility and resistance to superinfection (Smith et al. 1999, Read and Taylor 2001,
54 Vardo-Zalik and Schall 2009).

55

56   Modeling plays an important role in our understanding of clone distributions, including
57 for estimating two related parameters: multiplicity of infection (MOI; e.g. Hill and Babiker 1995,
58 Schneider and Escalante 2014) and complexity of infection (COI; e.g. Galinsky et al. 2015,
59 Chang et al. 2017). Though there is some inconsistency in the use of these terms, MOI often
60 describes the number of parasite clones entering a host, whether or not those clones are distinct
61 (i.e. a host may be bitten by multiple vectors carrying the same parasite clone), while COI
62 describes the number of genetically distinct clones in a host. The distribution of clones among
63 hosts is often modeled using either the Poisson or negative binomial (NB) distribution
64 (Schneider 2021); it is therefore important to know how often these assumed distributions are
65 observed in nature. Additionally, clone distributions may provide insight into the underlying
66 biological processes than generate and maintain these distributions. Below, I summarize the
67 Poison and NB distributions in the context of *Plasmodium* biology and provide three examples of
68 how complexities in the biology of *Plasmodium* biology that are supported by empirical data
69 might lead to alterations to these distributions. This list is not meant to be exhaustive, but merely
70 to demonstrate how the biology of *Plasmodium* can impact the distribution of clones among
71 hosts. Table 1 provides a summary of these points.

72

73 *Distributions*
74 Two distributions are commonly used to model count data like the number of infectious bites per
75 host: the Poisson distribution and the NB distribution (Hilbe 2014). These distributions model
76 the number of times an event occurs in a given interval of time or space. When applied to the
77 distribution of malaria clones, the 'events' are the clones themselves (or, equivalently, bites by
78 an infected insect vector) and the units of space are the hosts. In practice, the Poisson and NB
79 distributions used are usually conditional (they are zero-truncated, meaning that they do not

80   include not infected hosts, Schneider 2021), in part to account for the fact that many data sets do
81   not include robust sampling of not infected hosts (Hill and Babiker 1995).
82
83   The simpler of the two is the Poisson distribution. The Poisson distribution is described by a
84   single parameter, μ, which is both the mean and variance of the distribution. The Poisson
85   distribution has a number of assumptions, including the independence of clones (i.e. having a
86   certain clone does not affect the chances of having another) and the equivalence of the mean and
87   variance (i.e. as the average number of clones per infection increases, so does the variance in
88   clones per infection).
89
90   In reality, most parasite distributions are overdispersed relative to the Poisson distribution,
91   meaning that the variance-to-mean ratio is greater than 1. For microparasites like malaria,
92   overdispersion often means that parasite clones show significant clumping or aggregation, with a
93   smaller number of hosts containing a larger number of clones than would be expected if clones
94   were distributed at random. This clumping can occur for a variety of reasons, including if
95   individual hosts vary in their susceptibility or exposure or if parasite propagules are distributed
96   unequally in the environment (e.g. due to variation in vector abundance or vectors carrying
97   multiple clones; Boulinier et al. 1996). The negative binomial (NB) distribution is often a better
98   fit in such cases (Anderson and May 1992). The NB is described by two parameters: the mean,
99   μ, and the dispersion parameter, α. When α is 0, the NB is equivalent to the Poisson. Higher
100 values of α indicate greater aggregation (i.e. parasite clones are concentrated in a smaller
101 proportion of the host population). Note that the statistical software R, which is used in this
102 paper, uses θ (or 'size') in place of α. θ is inversely related to α, thus small values of θ indicate
103 substantial clumping of clones and large values indicate similarity to a Poisson distribution
104 (Hilbe 2014). Hilbe (2014) provides useful coverage of both the Poisson and NB distributions.
105
106 Deviation from these basic distributions could occur for a number of reasons. For example,
107 clones may not be independent of one another, such as if multiple malaria clones are ingested by
108 a vector and transmitted simultaneously to a new host. Hill and Babiker (1995) dismissed this as
109 unlikely, but experimental data suggest not only that multiple clones can transfer to the vector
110 and develop there (e.g. Huber et al. 1998, Vardo-Zalik 2009), but also that clones that were
111 underrepresented in the vertebrate may actually tend to increase their relative representation in
112 the vector (Nwakanna et al. 2008, Taylor et al. 1997, Vardo-Zalik 2009). Hill and Babiker
113 (1995) outline an approach to model the distribution of clones in such a case, but note that we do
114 not have sufficient information about the transmission biology of *Plasmodium* to build a
115 meaningful model at this time. Nonetheless, it seems that this situation would lead to an
116 overabundance of multi-clone infections relative to the Poisson or negative binomial distribution.
117
118 Transmission may also be extremely localized such that some portion of an apparently
119 continuous population has little to no chance of contracting the parasite, while another portion is

120   at much higher risk than the overall prevalence of the parasite would otherwise suggest.
121   Although some variation in exposure or susceptibility can generate the clumping and
122   overdispersion characteristic of the negative binomial distribution (Boulinier et al. 1996), major
123   differences would cause deviations. Data suggest that such heterogeneity in exposure is
124   common. Transmission of human malaria is known to show hotspots, where transmission in
125   small areas is much higher than the surrounding region (Gaudart et al. 2006; reviewed in
126   Bousema et al. 2012). Similar patterns are seen in wildlife parasites. For example, Eisen and
127   Wright (2001) measured multiple landscape features at different spatial scales and found that
128   ground cover within 10m of a host's point of capture was the best predictor of infection status
129   with the lizard malaria parasite *Plasmodium mexicanum*.
130
131   An extreme case of heterogeneity in parasite exposure in which a portion of the population is not
132   exposed at all and a portion has ordinary variation in exposure can be modeled by a zero-inflated
133   distribution. This distribution assumes some proportion of the population has zero probability of
134   becoming infected while the remainder of the population has clones distributed in accordance
135   with the Poisson or NB distribution. Zero-inflated models are mixed models that include a
136   parameter (the 'zero-probability') describing the proportion of observed 0s (uninfected hosts)
137   that are not a result of the processes that generate the other distribution (Poisson or NB; Hilbe
138   2014); in other words, some portion of uninfected hosts are not infected because they are not
139   exposed to infectious vectors at all and other uninfected hosts just happen to have not been bitten
140   (but had the same chance of being bitten as the infected hosts). The number of not exposed hosts
141   can be estimated as the zero probability by fitting this zero-inflated distribution. Note that fitting
142   a zero-inflated model should predict a very similar distribution of clones among infected hosts to
143   fitting a zero-truncated (conditional) model, but the former may provide greater insight on
144   whether or not infected hosts are subject to the same underlying forces (e.g. exposure to infected
145   vectors) as infected hosts.
146
147   One final example of why standard distributions may not accurately predict clonal distributions
148   relates to intraspecific dynamics within infected hosts; negative interactions among *Plasmodium*
149   clones could lead to a paucity of multi-clone infections. For example, there is ample evidence
150   that established infections can prevent the entry of additional clones, a phenomenon known as
151   premunition (e.g. Vardo et al. 2007, de Roode et al. 2005, reviewed for *P. falciparum* in Smith et
152   al. 1999). Blocking the entrance of additional clones would reduce the number of multi-clone
153   infections because a host might harbor only a single parasite clone even if it had received
154   multiple infectious bites. This might alter the distribution of clones in various ways depending
155   on the transmission intensity and strength of premunition, but it seems likely that it would lead to
156   data that are underdispersed relative to the Poisson. At the extreme, with very high transmission
157   and perfect premunition (and ignoring the possibility that vectors may transmit multiple clones at
158   once), most hosts would be infected by a single clone; the mean number of clones per infection
159   would be low (near 1 clone per host), but the variance would be even lower.

160

161 *Research Question*

162   Obtaining reliable data for testing the distribution of clones among hosts can be
163 challenging. One reason is that some parasites (e.g. *Plasmodium falciparum*) are known to
164 sequester in capillaries, making it necessary to take multiple samples from a patient to ensure
165 that sequestered clones are not missed (Tusting et al. 2014). Additionally, control efforts like
166 anti-malarial prophylactics (Tusting et al 2014) and vaccine candidates (reviewed in Arnot 2002)
167 can alter MOI. Third, reliable counts of uninfected hosts are not always available due to study
168 design (e.g. uninfected hosts may not be sampled) or risk of false negatives (Hill and Babiker
169 1995).

170

171 I therefore pursue a study of the distribution of clones among hosts using the lizard malaria
172 parasite *Plasmodium mexicanum* in its host the western fence lizard *Sceloporus occidentalis*.
173 This natural host-parasite system has been studied for decades (Schall and St. Denis 2013) and is
174 free from many of the challenges outlined above: the parasite is not known to sequester (Ford
175 and Schall 2011), the hosts are not exposed to chemotherapy or vaccination, and false-negatives
176 are known to be quite rare (Perkins et al. 1998). Additionally, despite differences in biology
177 between *P. mexicanum* and human parasites (e.g. *P. mexicanum* is transmitted by sand flies
178 rather than mosquitoes), many of the biological complexities potentially impacting the
179 distribution of clones among hosts have shown similar patterns between *P. mexicanum* and other
180 malaria parasites (see papers referenced under 'Distributions'). My aim was to determine
181 whether the distribution of clones among hosts differs significantly from the distributions
182 outlined in Table 1.

183

184 ## Materials & Methods

185 *Study sites and sampling*

186 The malaria parasite *Plasmodium mexicanum* infects *Sceloporus occidentalis* lizards in western
187 North America and has been studied since 1978 at the University of California Hopland
188 Research and Extension Center (HREC) near the town of Hopland in Mendocino County
189 California (Schall and St. Denis 2013, Vardo and Schall 2007). Lizards are captured using a
190 fishing pole with attached noose, and a small sample of blood is taken before the lizard is
191 released at its point of capture. Blood samples are used to make thin blood smears for processing
192 with Giemsa stain and a small amount of blood is stored frozen on filter paper (Whatman®
193 qualitative circles or similar) for genetic analysis.

194

195 Sampling at HREC is conducted at a number of well-established sites. The parasite shows
196 genetic differentiation among these sites, some of which are only a few hundred meters apart
197 (Fricke et al. 2010). Because combining data from multiple sites or years might cause deviations
198 from standard statistical distributions, I tested the fit of the selected distributions using data taken
199 from the four individual sites sampled in 2009 or 2010 that had the most infections: Water Tank

200 (WT, 2009), Parson's Creek (PC, 2009), Middle Lower Horse (MLH, 2010) and Gorge (GOR,
201 2010). The largest of these sites, MLH, is less than 4.5ha; Eisen and Wright (2001) describe this
202 site in detail, though the area sampled in 2009 – 2010 is a bit smaller than their study area. The
203 other sites are each closer to 1ha (Fig. 1).
204
205 *Plasmodium* infections in this system are known to be chronic; as a result, variation in the
206 duration of exposure may also cause deviations from standard distributions (i.e. older lizards
207 might harbor more parasite clones because they have been exposed to infectious bites over a
208 longer period). To limit the impact of variation in lizard age on clonal distributions, I retained
209 only lizards in a limited size range (40-64mm) in the final analysis. For lizards collected in early
210 summer (the vast majority of lizards in this data set), this range of sizes corresponds to lizards
211 that hatched in late summer of the previous year with a high level of confidence (Davis 1967).
212
213 *Genetic analysis*
214 All infections included in this study were genotyped previously; these data were published by
215 Neal and Schall (2014). Genotyping was performed for 4 microsatellite loci- Pmx306, Pmx732,
216 Pmx747 and Pmx839- using fluorescently labeled PCR primers and conditions described by
217 Schall and Vardo (2007). The PCR product was diluted and underwent fragment analysis at
218 Cornell University Biotechnology Resource Center.
219
220 The parasite is haploid in vertebrate blood, so each clone will have a single allele at each locus
221 and produce a single peak on an electropherogram. Clones that are at a low relative frequency in
222 the blood may be hard to detect in infections due to a weak signal on the electropherograms. No
223 peak-size cut-off was applied to correct for this, but a previous study demonstrated that even
224 alleles with 10-fold differences in concentration could usually be detected (Vardo-Zalik et al.
225 2009). Additionally, alleles were classified by relative electropherogram peak height (highest,
226 >1/3 highest, <1/3 highest [small]); these designations were not used in my analysis, but are
227 available in the data set at osf.io/6yvh5.
228
229 I estimated the number of clones per host as the maximum number of alleles seen at any one
230 locus (a common measure also used for human malaria [Zhong et al. 2018]). Loci with fewer
231 alleles may result if clones share alleles at that locus. This measure provides a minimum
232 estimate of the number of clones per infection because some clones may share alleles, and will
233 be referred to hereafter as the observed minimum number of clones, or observed MC. Anderson
234 et al. (2010) showed that it was possible to distinguish about 90% of *Plasmodium falciparum*
235 clones using only four microsatellites, and the loci in that paper had lower average
236 heterozygosity than the loci I used. This suggests that observed MC, while a minimum estimate
237 of MOI, may not greatly underestimate the number of distinct clones in most cases.
238
239 *Fitting distributions*

240   All analyses were performed in R (R Development Core Team, Vienna, Austria). The following
241   analyses were performed on observed MC data for each of the four single sites individually
242   (GOR [2010], MLH [2010], PC [2009] and WT [2009]). I fit the Poisson distribution using the
243   glm() function and the NB using the glm.nb() function. The zero-inflated models were fit using
244   the zeroinfl() from the R package "pscl". I assessed dispersion using the Pearson dispersion
245   statistic as suggested by Hilbe (2014) and compared each zero-inflated model to the
246   corresponding standard model with the Vuong test using the function vuong(). I performed a
247   chi-squared test comparing the observed vs. expected values of each distribution after grouping
248   multi-clone infections to reduce the number of categories with fewer than five observed
249   infections following (generally) the suggestion of Bliss and Fisher (1953; two sites unfortunately
250   had fewer than five single-clone infections: PC had four and WT had three). Due to the
251   relatively small number of infections at each site considered individually, I used Fisher's
252   combined probability test to obtain an overall assessment of whether the distribution of observed
253   MC at the four sites collectively differed significantly from each distribution.
254
255   As mentioned above, observed MC is known to provide a minimum (and therefore biased)
256   estimate of the true MOI. True MOI is almost certainly higher because some distinct clones will
257   share alleles and some hosts may have been infected by the same clone multiple times
258   (Schneider 2018). If the bias is slight, it may not greatly affect the fit of distributions, but if the
259   bias is greater, it could cause MOI estimates to deviate from expected distributions even if true
260   MOI does not. To account for this possibility, I used the likelihood approach of Schneider and
261   Escalante (2014) to obtain an unbiased (Schneider 2018) estimate of average MOI at each site,
262   simulated the observable number of alleles per locus using a Poisson distribution of clones per
263   infection and allele frequencies observed at each site, and compared this expected distribution to
264   the distribution of observed MC. True MOI and true allele frequencies were estimated from
265   genetic data for each site using the MLE() R function provided by Schneider (2018;
266   supplemental materials). A separate MOI estimate was obtained for each of the four
267   microsatellite loci, and these values were averaged to obtain a single overall estimate of MOI for
268   each site. I then simulated the distribution of clones among n_s hosts, where n_s is the number of
269   lizards collected for the observed data from each site s. Simulations were performed by drawing
270   n_s random numbers from a Poisson distribution with lambda equal to the estimated MOI. These
271   random numbers represent the number of infectious bites received by each of the n_s simulated
272   hosts. I then used the function mnom(), also from the R code in Schneider (2018, supplemental
273   materials), to randomly draw x_i alleles from a multinomial distribution with allele frequencies
274   estimated from the observed data (described above), where x_i is the random number of infectious
275   bites for each host drawn from the Poisson distribution in the previous step. Sampling of the
276   multinomial distribution was performed for each of the n_s hosts at each of the four microsatellite
277   loci. After allele sampling, a count of the number of distinct alleles per locus was tallied (note
278   that some alleles may be sampled more than once, but would only be counted once) and the
279   maximum number of distinct alleles at any locus was recorded as simulated MC. Taken

280   together, these steps were meant to simulate the expected underlying biology (Poisson
281   distribution of infectious bites) and known observational limitations (overlap in alleles among
282   clones, repeated infection with the same clones) of the observed data. The simulation was
283   repeated 1000 times to determine the mean, minimum and maximum number of infections
284   observed with each possible number of clones to determine if observed data fell in this range
285   (estimation of a 95% confidence interval did not seem necessary; see Results).
286

287   **Results**
288   Forty-five (WT) to 131 (MLH) lizards were sampled per site, with the number of infections per
289   sampled site ranging from 10 (PC) to 17 (MLH; Table 2). Despite these relatively small sample
290   sizes, there was still sufficient genetic variation to differentiate clones; the number of alleles per
291   locus at individual sites ranged from 5 to 10, with every site having at least one locus with 7 or
292   more alleles (See Table S1 for details).
293

294   The distribution of observed MC was overdispersed relative to the Poisson (dispersion parameter
295   range 1.55 [WT] – 2.12 [GOR]) and differed significantly from this distribution (combined P
296   <0.0001, X^2 = 112.5, df = 8). The distribution also differed significantly from the NB (combined
297   P = 0.0031, X^2 = 23.2, df = 8), showing significant underdispersion relative to the NB
298   (dispersion parameter range 0.75 [WT] – 0.80 [MLH]). The fit values of theta for the ZINB
299   were so large (all >20,000; Table S2) that the fit distributions were indistinguishable from the
300   ZIP, and the observed data did not differ from either of these fit distributions (combined P =
301   0.296, X^2 = 9.58, df = 8; values are for both ZIP and ZINB). Table 2 shows the fit values and
302   goodness-of-fit statistics for the NB and ZIP. Table S2 shows individual goodness-of-fit X^2 tests
303   for each site individually, and Table S3 lists the fit values for each of the four models for every
304   site. AIC and BIC are both lower for the traditional NB than the traditional Poisson at every site
305   (Table S4). The Vuong Test indicates that the ZIP models are preferable to the traditional
306   Poisson models (P<0.05 for all raw, AIC- and BIC-corrected tests; Table S5) and that the ZINB
307   models may be indistinguishable from or slightly better than the traditional NB (P-values mixed;
308   Table S5). A comparison of the observed and expected values for all distributions at all four sites
309   is shown in Fig. 2, and the amount each model deviated from observed values is shown in Fig.
310   3A-C.
311

312   The distribution of simulated MC differed considerably from observed MC. Most notably, the
313   simulation substantially overestimated the number of hosts that would be infected: estimates
314   ranged from 32 (WT) to 87 (MLH) infections per site, while observed numbers of infections
315   were much lower (10 [PC] to 17 [MLH]). For all four sites, the observed numbers of uninfected
316   lizards were well above the range observed in 1000 simulations and the observed numbers of
317   lizards with a single apparent clone (MC = 1) were below the range (Table S6). Looking just at
318   the distribution of clones among infected hosts, values of observed MC showed a consistent
319   deficit of single-clone infections and excess of two-clone infections compared with the relative

320 frequencies of these infection types in the simulation (Fig. 3D), much like the distributions fit to
321 observed MC (Fig. 3A-C).
322
Discussion
324 Many models of MOI for malaria parasites assume that *Plasmodium* clones are distributed
325 among hosts following the Poisson or NB distribution, but certain aspects of these parasites'
326 biology may cause substantial deviations from this prediction (e.g. Table 1). I compared the
327 observed numbers of clones per infection at four individual sites with those expected based on
328 four statistical distributions: the traditional Poisson, the traditional NB, the zero-inflated Poisson
329 (ZIP) and the zero-inflated negative binomial (ZINB). The observed data differed significantly
330 from both traditional distributions, but do not differ from either zero-inflated distribution. The fit
331 values of theta for the ZINB were so large (all >50,000) that this model essentially converges on
332 the ZIP distribution (recall from the introduction that large values of theta indicate similarity to
333 the Poisson). While the method used to measure MOI in this study is expected to show bias
334 (discussed above), incorporating an unbiased estimate of MOI (simulation) did not appear to
335 qualitatively alter the results: the Poison distribution still did a poor job of describing the
336 distribution of clones among hosts when not infected hosts were considered (Table S6), and still
337 appeared to overestimate the relative frequency of one-clone infections and underestimate the
338 relative frequency of two-clone infections among infected lizards (Fig. 3). Taken together, I
339 conclude that the best fitting distribution for the observed data is the ZIP.
340
341 The ZIP was selected for comparison with data because previous studies have shown that malaria
342 transmission can be highly localized. The fact that this distribution showed the best fit of those
343 tested may suggest that localized transmission is important in shaping the distribution of clones
344 among hosts. In fact, the fit value of one of the parameters, the zero probability, suggests that a
345 high percentage (60-82%) of lizards may have essentially no chance of contracting malaria even
346 at sites where *P. mexicanum* is relatively common. In other words, sites may contain small
347 patches where most transmission is focused surrounded by areas where transmission is much
348 lower or non-existent. The feasibility of such hotspots being established is facilitated by the fact
349 that *S. occidentalis* lizards show high site fidelity, often being found on the same rock, log or
350 fence post throughout the collecting season (pers. obs., Eisen and Wright 2001), and that the
351 insect vectors (*Lutzomyia* sand flies) are not strong fliers (de Queiroz Balbino et al. 2006).
352 Together, this sets up a situation in which transmission could be highly localized, and previous
353 research has shown that certain landscape features (e.g. leaf litter) are associated with higher
354 rates of infection with *P. mexicanum* on a scale of a few meters (Eisen and Wright 2001).
355
356 Even though the ZIP was the best fitting model, there were still consistent discrepancies between
357 the model's predictions and the observed data. Most notably, all four sites showed a consistent
358 shortage of single-clone infections and a consistent excess of two-clone infections relative to

359   model predictions (Fig. 3), regardless of the distribution examined. I have no definitive
360   explanation for this pattern, but I discuss a few possibilities below.
361
362   One possibility is that there is a consistent bias in the way that we estimate the number of clones
363   per infection. Estimates of MOI based on the maximum number of microsatellite alleles at a
364   locus do have known limitations. For one, microsatellite data itself is imperfect. PCR may
365   preferentially amplify more common or shorter fragments, and minor clones may be missed due
366   to the relative weakness of the electropherogram signal, especially if the minor clone carries an
367   allele that falls in the stutter (a common artifact in microsatellite analysis) of a major clone's
368   allele. While this remains a legitimate concern, previous research conducted on *P. mexicanum*
369   using the microsatellite loci reported on here suggest that, while not perfect, genotyping results
370   with these markers are fairly robust even with 10-fold differences in the relative proportions of
371   clones (Vardo-Zalik et al. 2009). A related concern is that *Plasmodium* clones may share alleles
372   at a locus or experience superinfection from the same parasite clone, which could true MOI
373   higher than observed MC. Although overlapping alleles would tend to underestimate the number
374   of clones found in an infection, possibly leading infections with three or more clones to be
375   misclassified as two-clone infections, it does not seem likely that this explains excess of two-
376   clone infections in the data because the same pattern was seen when comparing observed data to
377   the MLE simulation, which incorporates an unbiased estimate of MOI.
378
379   Another possibility is that the consistent excess of two-clone infections reflects something
380   biologically important. Two of the biological situations outlined in Table 1 were not connected
381   with specific distributions to be tested in this study, but I instead predicted that they would lead
382   to an overabundance (collective transmission) or underabundance (premunition) of multi-clone
383   infections relative to the fit distributions. What is observed in the data is a consistent excess of
384   multi-clone infections, but only those with two clones.   This does not lead to a clear support of
385   either scenario in isolation, but could conceivably result from the two forces combined. It is
386   possible that small numbers of clones (e.g. two) are transmitted collectively and, once
387   established, are resistant to superinfection (i.e. display premunition), perhaps more so than
388   single-clone infections. Data from *P. falciparum* do suggest that complex infections are more
389   likely to display premunition (Smith et al. 1999). It does not appear that is the case for *P.*
390   *mexicanum* (Vardo et al. 2007), though single and two-clone infections were grouped for
391   analysis in that experiment.
392

Conclusions
394   The data presented here do not provide any definitive answers about what shapes the distribution
395   of *Plasmodium* clones among hosts- that was never my intention. The observed distribution of
396   clones may follow or diverge from standard statistical distributions for any number of reasons.
397   What these data do provide is insight into whether commonly held assumptions about these
398   distributions are supported by data and guidance for shaping future hypotheses.

399

400 First, the data presented here provide support for the use of the conditional Poisson in modeling
401 MOI and the distribution of clones among hosts (as is used, for example, in Schneider and
402 Escalante 2014, Hill and Babiker 1995, Schneider 2021). Zero-truncated distributions are
403 conditional distributions that leave out uninfected individuals and have frequently been used in
404 these models. Their use was motivated primarily by limitations in data (e.g. uninfected
405 individuals are not sampled or may be misclassified due to low parasite density in the blood), but
406 conveniently, these zero-truncated distributions may also provide a good fit for data that would
407 reasonably be described by zero-inflated distributions, which could result from extremely patchy
408 transmission. My data suggest that zero-inflated models provide a better fit to MOI data for *P.*
409 *mexicanum* than traditional models, and zero-truncated models should therefore also do an
410 adequate job of modeling these data.

411

412 Second, this study suggests that transmission hotspots may be as important for the
413 transmission of *P. mexicanum* as they are for human malaria parasites. Highly localized
414 transmission may account for the improved fit of zero-inflated models relative to traditional
415 distributions. Additional research into fine-scale transmission patterns in both human and animal
416 parasites may prove fruitful.

417

418 Finally, the distribution of clones among hosts at the individual sites sampled here appear
419 to show a consistent excess of two-clone infections and shortage of single-clone infections
420 relative to the predictions of any of the fit distributions. I am at a loss to provide a clear
421 hypothesis regarding the source of this deviation, but it may be interesting to determine whether
422 such patterns are seen for other malaria parasites and an explanation for this pattern, when found,
423 may provide new insight into the biology of these organisms.

424

Acknowledgements

426 I am incredibly grateful to the six anonymous reviewers who provided invaluable suggestions at
427 various stages in the development of this manuscript. I also thank J. Schall and his students for
428 helping curate the data set analyzed here and the staff of the University of California Hopland
429 Research and Extension Center for their continuing support of this research.

430

References

432

433 Arnot DE. The influence of the genetic complexity of *Plasmodium falciparum* infections on the
434   epidemiology of malaria. Trans R Soc Trop Med Hyg. 2002;96(SUPPL. 1):S131.
435 Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford:
436   Oxford University Press; 1992.
437 Anderson TJC, Williams JT, Nair S, Sudimack D, Barends M, Jaidee A, et al. Inferred
438   relatedness and heritability in malaria parasites. Proc Biol Sci. 2010;277(1693):2531–40.

439   Bliss CI, Fisher RA. Fitting the negative binomial distribution to biological data. Biometrics.
440   1953;9(2):176–200.
441   Boulinier T, Ives AR, Danchin E. Measuring aggregation of parasites at different host population
442   levels. Parasitology. 1996;112:581–7.
443   Bousema T, Griffin J, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting
444   hotspots: spatial targeting of malaria for control and elimination. PLoS Medicine.
445   2012;9(1):1-7.
446   Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, et al. THE REAL
447   McCOIL: A method for the concurrent estimation o fthe complexity of infection and SNP
448   allele frequency for malaria parasites. PLoS Computational Biology. 2017;13(1):1-18.
449   Davis J. Growth and Size of the western fence lizard (*Sceloporus occidentalis*). Copeia.
450   1967;1967(4):721-31.
451   de Queiroz Balbino V, Coutinho-Abreu IV, Sonoda IV, Melo MA, de Andrade PP, de Castro
452   JAF, et al. Genetic structure of natural populations of the sand fly *Lutzomyia longipalpis*
453   (Diptera: Psychodidae) from the Brazilian northeastern region. Acta Trop. 2006;98(1):15–24.
454   de Roode JC, Helinski MEH, Anwar MA, Read AF. Dynamics of multiple infection and within-
455   host competition in genetically diverse malaria infections. Am Nat. 2005;166(5):531–42.
456   Eisen RJ, Wright NM. Landscape features associated with infection by a malaria parasite
457   (*Plasmodium mexicanum*) and the importance of multiple scale studies. Parasitology.
458   2001;122(Pt 5):507–13.
459   Ford AF, Schall JJ. Relative clonal proportions over time in mixed-genotype infections of the
460   lizard malaria parasite Plasmodium mexicanum. Int J Parasitol. 2011;41(7):731–8.
461   Fricke JM, Vardo-Zalik AM, Schall JJ. Geographic genetic differentiation of a malaria parasite
462   *Plasmodium mexicanum*, and its lizard host, *Sceloporus occidentalis*. Journal of Parasitology.
463   2010;96(2):308-313.
464   Galinsky K, Valim C, Salmier A, De Thoisy B, Musset L, Legrand E, et al. COIL: a
465   methodology for evaluating malarial complexity of infection using likelihood from single
466   nucleotide polymorphism data. Malaria Journal. 2015;14(1):1-9.
467   Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, et al. Space-time clustering of
468   childhood malaria at the household level: a dynamic cohort in a Mali village. BMC Public
469   Health. 2006;6:286.
470   Hilbe JM. Modeling Count Data. Cambridge University Press; 2014.
471   Hill WG, Babiker HA. Estimation of numbers of malaria clones in blood samples. Proc Biol Sci.
472   1995;262(1365):249–57.
473   Huber W, Haji H, Charlwood JD, Certa U, Walliker D, Tanner M. Genetic characterization of
474   the malaria parasite *Plasmodium falciparum* in the transmission from the host to the vector.
475   Parasitology. 1998;116:95–101.
476   Neal AT, Schall JJ. Testing sex ratio theory with the malaria parasite *Plasmodium mexicanum* in
477   natural and experimental infections. Evolution. 2014;68(4):1071-1081.
478   Nwakanma D, Kheir A, Sowa M, Dunyo S, Jawara M, Pinder M, et al. High gametocyte
479   complexity and mosquito infectivity of *Plasmodium falciparum* in the Gambia. Int J
480   Parasitol. 2008;38(2):219–27.

481 Perkins SL, Osgood SM, Schall JJ. Use of PCR for detection of subpatent infections of lizard
482   malaria: implications for epizootiology. Mol Ecol. 1998;7(11):1587–90.
483 Read AF, Taylor LH. The ecology of genetically diverse infections. Science.
484   2001;292(5519):1099–1102.
485 Schall JJ, St Denis KM. Microsatellite loci over a thirty-three year period for a malaria parasite
486   (*Plasmodium mexicanum*): bottleneck in effective population size and effect on allele
487   frequencies. Parasitology. 2013;140:21–8.
488 Schall JJ, Vardo AM. Identification of microsatellite markers in *Plasmodium mexicanum*, a
489   lizard malaria parasite that infects nucleated erythrocytes. Mol Ecol Notes. 2007;7(2):227–9.
490 Schneider KA. Charles Darwin meets Ronald Ross: a population-genetic framework for the
491   evolutionary dynamics of malaria. In: Infectious Diseases and Our Planet by MI Teboh-
492   Ewungkem and GA Ngwa (Eds.), Springer, 2021:149-191.
493 Schneider KA. Large and finite sample properties of a maximum-likelihood estimator for
494   multiplicity of infection. PLoS ONE. 2018;13(4):1-21.
495 Schneider KA, Escalante AA. A likelihood approach to estimate the number of co-infections.
496   PLoS One. 2014;9(7).
497 Smith T, Felger I, Tanner M, Beck HP. The epidemiology of multiple *Plasmodium falciparum*
498   infections 11. Premunition in *Plasmodium falciparum* infection: Insights from the
499   epidemiology of multiple infections. Trans R Soc Trop Med Hyg. 1999;93(SUPPL. 1):59–
500   64.
501 Taylor LH, Walliker D, Read AF. Mixed-genotype infections of malaria parasites: within-host
502   dynamics and transmission success of competing clones. Proc Biol Sci. 1997;264:927–935.
503 Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in *Plasmodium falciparum*
504   transmission: precision, accuracy and costs of metrics. Physiol. Behav. 2014;84:151–208.
505 Vardo AM, Kaufhold KD, Schall JJ. Experimental test for premunition in a lizard malaria
506   parasite (*Plasmodium mexicanum*). J Parasitol. 2007;93(2):280–2.
507 Vardo AM, Schall JJ. Clonal diversity of a lizard malaria parasite, *Plasmodium mexicanum*, in its
508   vertebrate host, the western fence lizard: role of variation in transmission intensity over time
509   and space. Mol Ecol. 2007;16(13):2712–20.
510 Vardo-Zalik AM. Clonal diversity of a malaria parasite, *Plasmodium mexicanum*, and its
511   transmission success from its vertebrate-to-insect host. Int J Parasitol. 2009;39(14):1573–9.
512 Vardo-Zalik AM, Ford AF, Schall JJ. Detecting number of clones, and their relative abundance,
513   of a malaria parasite (*Plasmodium mexicanum*) infecting its vertebrate host. Parasitol Res.
514   2009;105(1):209–15.
515 Vardo-Zalik AM, Schall JJ. Clonal diversity alters the infection dynamics of a malaria parasite
516   (*Plasmodium mexicanum*) in its vertebrate host. Ecology. 2009;90(2):529–36.
517 Zhong D, Koepfli C, Cui L, Yan G. Molecular approaches to determine the multiplicity of
518   Plasmodium infections. Malar J. 2018;17(1):1–9.

Overview of distributions of interest

1   **Table 1: Overview of distributions of interest**

Biological situation	Predicted distribution
Clones are independent; all hosts have equal susceptibility and likelihood of exposure	Poisson
Clones are independent; hosts show some variability in susceptibility or exposure that causes clones to clump together	Negative Binomial (NB)
Some portion of the population meets criteria above (Poisson or NB); some portion is not exposed	Zero-inflated model (Poisson or NB)
Clones are not independent; they are regularly transmitted collectively	Undefined, but likely with an excess of multi-clone infections relative to above distributions
Clones are not independent; premunition limits multi-clone infections	Undefined, but likely with a shortage of multi-clone infections relative to the above distributions

Summary and fit statistics for sites used to test the distribution of clones among hosts.

Site indicates where lizards were collected (Gorge [GOR], Middle Lower Horse [MLH], Parson's Creek [PC], and Water Tank [WT]). Year indicates the year samples were collected. For each site and year, I report: n_s- the number of lizards collected that were 40-64mm snout to vent; inf- the number of these lizards that were infected with *P. mexicanum* and have genetic data (only one lizard from GOR was infected but not genotyped); mu- mean clones per host from fit negative binomial or zero inflated Poisson distributions; theta- dispersion parameter of clones per host from fit negative binomial distribution; chisq- chi squared value from goodness of fit test (observed vs. expected from fit distribution); P- P-value from goodness of fit test (observed vs. expected from fit distribution); and zp- the zero probability from the fit zero inflated Poisson distribution. Two additional models were fit, with results reported in the supplemental materials. The model selected for presentation here are the best fitting traditional model (Negative binomial) and zero inflated model (ZI Poisson).

1 **Table 2: Summary and fit statistics for sites used to test the distribution of clones among hosts.**
2 Site indicates where lizards were collected (Gorge [GOR], Middle Lower Horse [MLH], Parson's
3 Creek [PC], and Water Tank [WT]).
4 Year indicates the year samples were collected. For each site and year, I report:
5 n_s- the number of lizards collected that were 40-64mm snout to vent
6 inf- the number of these lizards that were infected with *P. mexicanum* and have genetic data
7 (only one lizard from GOR was infected but not genotyped).
8 mu- mean clones per host from fit negative binomial or zero inflated Poisson distributions
9 theta- dispersion parameter of clones per host from fit negative binomial distribution
10 chisq- chi squared value from goodness of fit test (observed vs. expected from fit distribution)
11 P- P-value from goodness of fit test (observed vs. expected from fit distribution)
12 zp- the zero probability from the fit zero inflated Poisson distribution
13 Two additional models were fit, with results reported in the supplemental materials. The
14 model selected for presentation here are the best fitting traditional model (Negative binomial)
15 and zero inflated model (ZI Poisson).
16

				Negative binomial				ZI Poisson			
Site	Year	n_s	inf	*mu*	*theta*	chisq	P	*mu*	*zp*	chisq	P
GOR	2010	77	12	0.30	0.17	0.73	0.70	1.48	0.80	0.085	0.96
MLH	2010	131	17	0.22	0.17	3.48	0.18	1.18	0.81	0.833	0.66
PC	2009	78	10	0.23	0.15	11.91	**0.0026**	1.31	0.82	5.96	0.051
WT	2009	45	14	0.60	0.57	7.1	**0.029**	1.49	0.60	2.70	0.26

24
25

Figure 1

Map of the sites sampled for this study.

Sites are Gorge (GOR), Middle Lower Horse (MLH), Parson's Creek (PC) and Water Tank (WT).
Map data © 2021 Google.

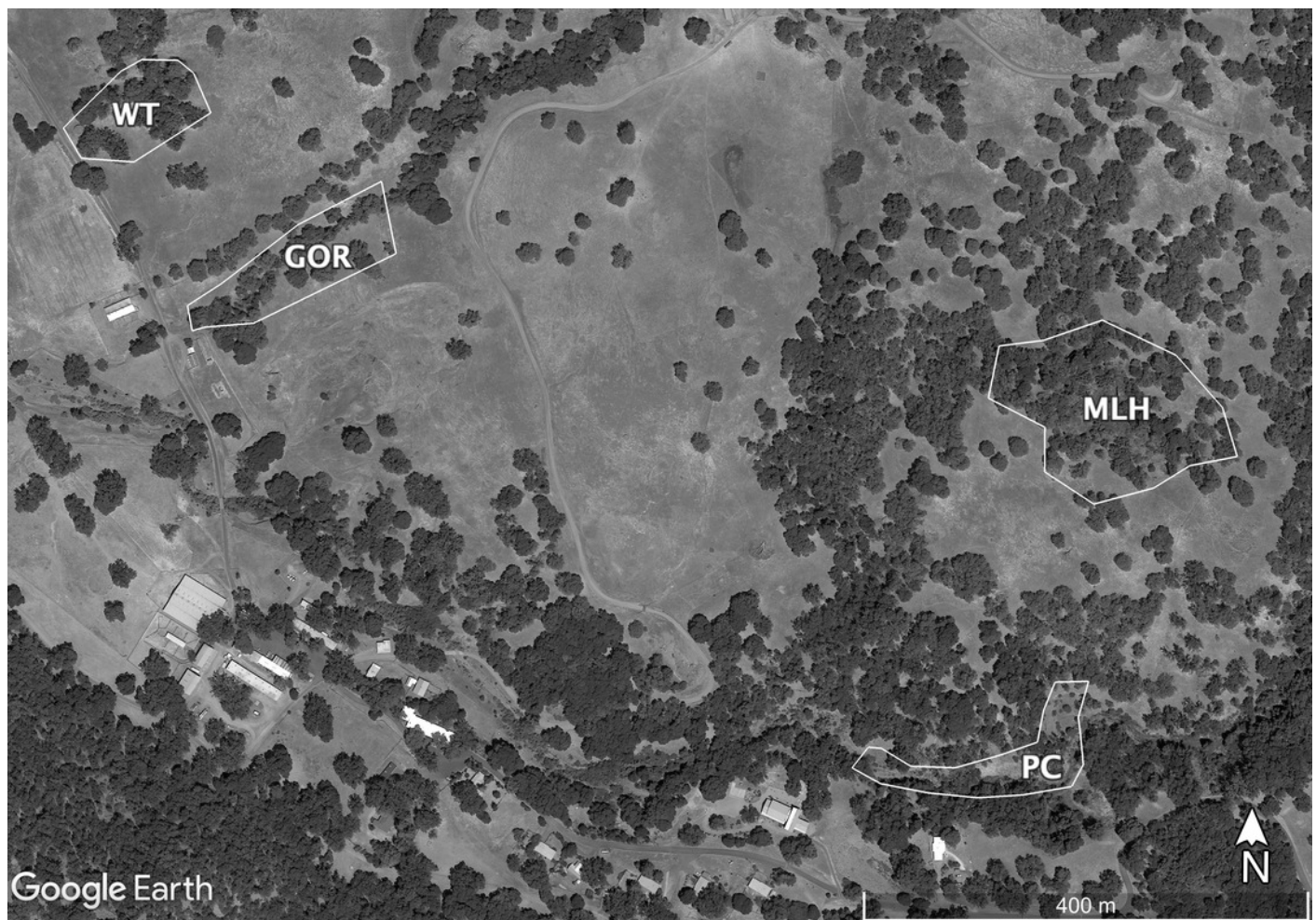



Figure 2

Observed (bars) vs. expected (lines) number of parasite clones per host at four individual sites.

Expected lines are based on traditional Poisson (solid black), traditional negative binomial (dotted black), and zero-inflated Poisson (solid green) models (the zero-inflated negative binomial fit is equivalent). The full height of the first bar (0 clones) represents the observed number of not infected lizards at the site; the dark gray portion of the bar represents the number of lizards estimated to be not exposed based on the fit zero-inflated distribution (the estimated zero probability). Collectively, the distribution of clones among infections differs significantly from both traditional models but not the zero-inflated negative models (see text for details).

A

GOR

Frequency

clones

B

MLH

Frequency

clones

C

PC

Frequency

clones

D

WT

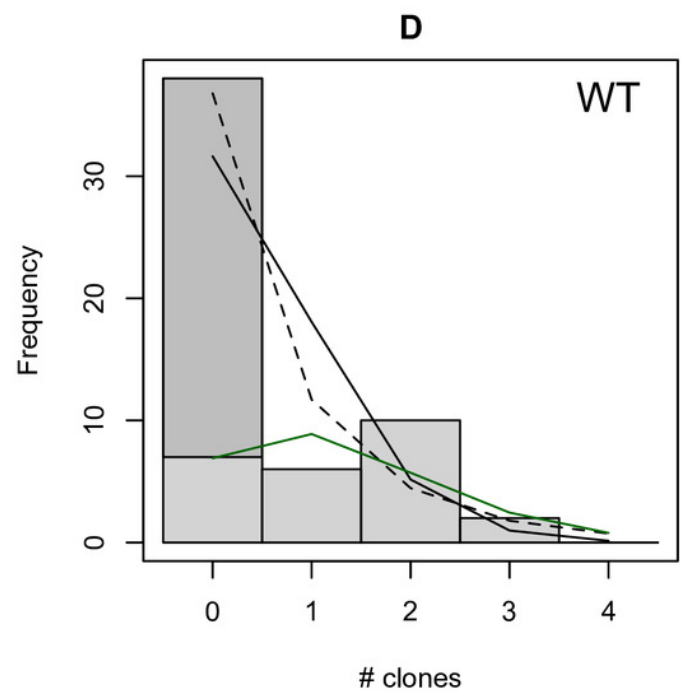
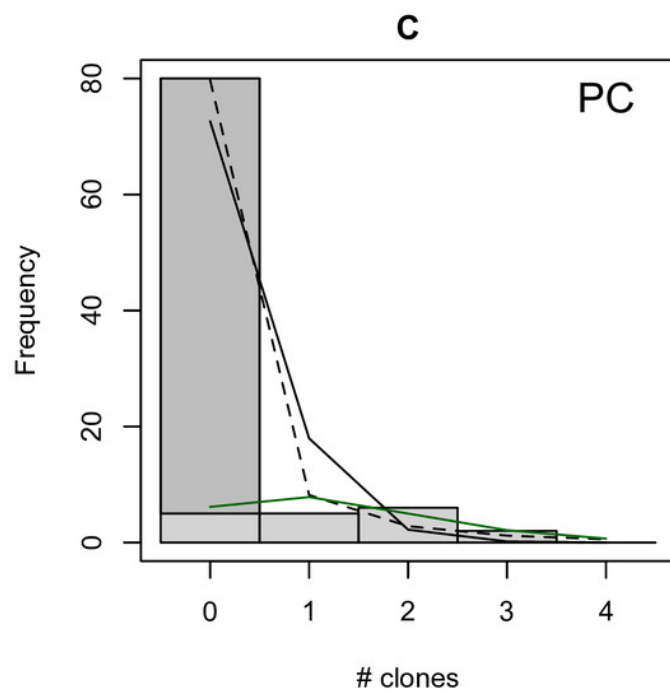
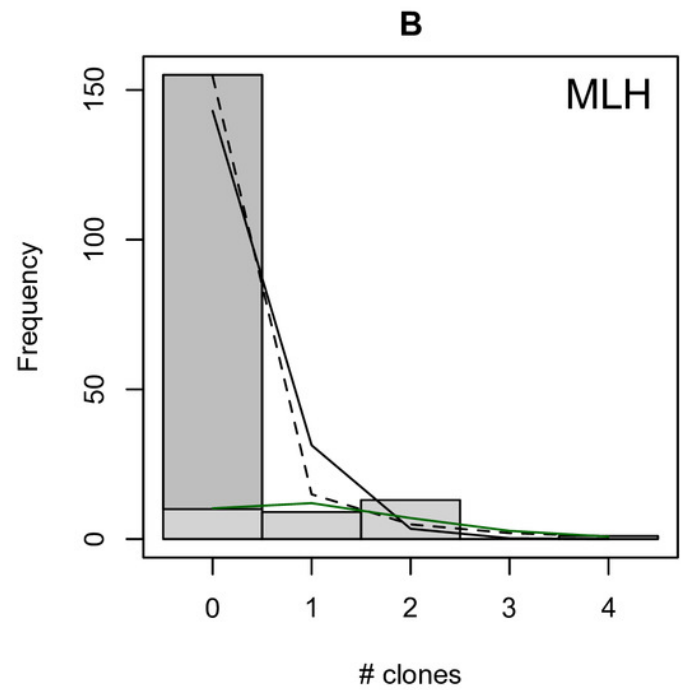
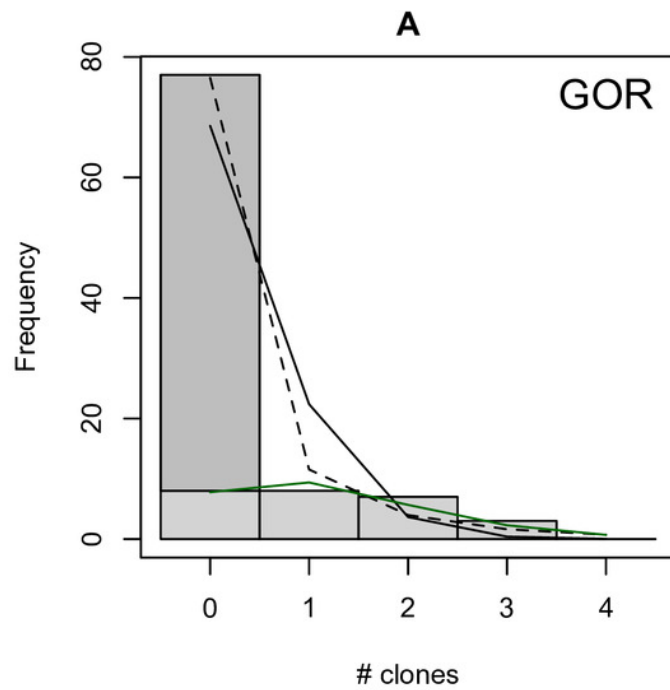
Frequency

clones

Figure 3

Deviation from fit distributions of number of *Plasmodium mexicanum* clones per infected host.

Fit distributions are A) Poisson, B) Negative binomial, C) Zero-Inflated Poisson (Zero-Inflated Negative Binomial is equivalent) and D) Simulated distribution based on MOI estimated using the method of Schneider and Escalante (2014). Data are shown for four individual sites: GOR (triangles), MLH (circles), PC ('X'), and WT (+). Because MOI was estimated from infected lizards only, simulated expectations were scaled down to equalize the total number of expected vs. observed infections; without this adjustment most simulated values were well above observed values.