

Reply to Fountain *et al.* “Human facilitated metapopulation dynamics in an emerging pest species, *Cimex lectularius*”

10 November 2013

1 Summary

This article studies the genetic composition on different levels of a human-associated pest species, the common bed bug *Cimex lectularius*. One level is that of infestations in the same or different cities; two infestations from Birmingham and London each, and one from New South Wales, Australia. Each of these infestations is composed of several so-called refugia, which are treated as the atomic unit of population structure here. The second level is that of 13 different refugia within a particular infestation in London.

In the first part, the authors do a standard descriptive analysis of extant genetic diversity and differentiation on both these levels. They find significant differentiation based on F_{ST} between multi-refuge infestations. Only one infestation, located in London, revealed significant sub-structure among refugia. Some infestations show an excess, others a deficiency of heterozygotes. On the second level, the authors find significant differentiation between refugia within the focal infestation in London. There is no isolation by distance and no significant deviation from Hardy-Weinberg proportions.

In the second part, the authors focus exclusively on the within-infestation level and try to link observed genetic structure to metapopulation dynamics. Using ABC, two alternative scenarios are compared, the ‘propagule pool’ and the ‘migrant pool’ model of colonisation. In the first one, all founder individuals of a given refuge are assumed to come from a single source refuge that is also part of the metapopulation. In the second one, founders of a given refuge represent a random sample taken across all extant refugia in the metapopulation. The authors seem to find strong support for the propagule model. They then estimate parameters belonging to this model, focussing in particular on the number of founder individuals of a typical refuge, the typical age of a refuge until detection (which is equivalent to the time of sampling) and the typical age of a refuge before it acts as a source of founders. The authors conclude that refugia are commonly founded by small numbers of individuals and do not exist for a long time (because they are usually detected early on). They interpret their results as strong evidence in favour of the propagule model.

The implications of this study for pest management remain somewhat unclear, although it seemed to be one of the goals to provide answers here. I am concerned about the simplified versions of the migrant and propagule model employed here. Moreover, in the apparent absence of dispersal other than in the form of colonisation, the migrant pool model loses its justification. The conclusion that ABC can easily distinguish between alternative scenarios in a metapopulation context is an overstatement. It may do so in this case, but this cannot be generalised. Also, the scenarios modeled here have essentially been reduced to classical models of structured populations, without any features typical of metapopulations being modeled *explicitly*.

Regarding the application of ABC, I missed more details about its performance, and parameter estimates inferred under the migrant pool model. This is because I wonder if the weak support for the migrant pool model is a consequence of the choice of the priors. In particular, I wonder if allowing

for much smaller values of the parameter N_s would give more support to scenario 1 (migrant pool model). A respectable contribution of this study is the characterisation of 21 new microsatellites.

It would have been nice to see if individuals of a given refuge can be assigned to (a combination of) other refugia. Such an analysis might corroborate the authors' preference for the propagule model. The reservation against using software that does assignment seems to stem from uncertainty about Hardy-Weinberg equilibrium. However, in the absence of strong deviations, I think the authors could use a software like STRUCTURE or alike.

2 General issues

1. To my understanding, the differentiation between the 'migrant pool' and the 'propagule pool' model of colonization goes back to a question that WADE and MCCAULEY (1988) phrased as "Is colonization a behavior distinct from migration?". Clearly, the migrant pool model assumes that colonization follows the same rules as migration, where in this context migration is assumed to be according to the island model. The propagule model is to say that colonization is different from migration. Here, the authors on the one hand seem to imply that migration among extant refugia is almost absent. I conclude this from their scenarios in Figure 1, where migration is not explicitly modeled between the present and times t_2 and t_3 . On the other hand, going back in time beyond t_2 (t_3), the whole metapopulation is modeled as one panmictic Wright-Fischer population. This assumes at least some exchange between refugia, either via migration or colonization. If there were no migration at all (but only colonization), then I am not sure if the long-term behavior of the metapopulation can be summarized by just one effective size N_s .
2. I am worried that scenario 1 (migrant pool model) might be misspecified (Figure S5). It received substantially less support in the ABC-type model comparison. However, it was not obvious to me if this is a consequence of prior choice. In particular, could it be possible that scenario 1 is much better compatible with the data if the prior for N_s allowed for much smaller values?
3. The view of this study proving that ABC can easily distinguish between alternative metapopulation scenarios is problematic. First, the scenarios shown in Figure 1 are highly simplified and do not explicitly model any feature specific to a metapopulation (namely extinction and colonization dynamics). Rather, the scenarios are those of static structured populations. ABC has been (supposedly successfully) applied in much more complicated models of structured populations. Second, a point that this study *could* aim at making, but currently does not, is that the perhaps more complicated metapopulation dynamics can be reduced to the simpler versions analyzed so far. However, this would imply implementing the more complicated scenarios explicitly and compare them to the simpler ones. I do not think this would allow for a statement about applicability of ABC, but it might allow for a (much more interesting?) statement about identifiability and reducibility of metapopulation models to simpler models. One way this could be achieved is by comparing the current setting to a scenario in which the panmictic phase (where the metapopulation is represented by $N_s * u$) is omitted. Instead, the colonization and extinction process of individual refugia would be modeled explicitly over longer periods of time. This could shed light on whether or not the more complicated dynamics on the metapopulation level can be represented in a single value of N_s or not.
4. Both scenarios assume that colonists come from within the metapopulation (i.e. infestation). What about introductions from other infestations, e.g. from the London area? Could available data be used to randomly draw colonists from the other London refugia?
5. Given that several approaches for choosing summary statistics in ABC have recently been proposed, I think one should make use of them. It is reasonably well established that no single method will perform best in all scenarios (BLUM *et al.*, 2013), but this cannot serve as an argument against using any of them at all (see comments to 1.274–279).

3 Specific comments

3.1 Abbreviations used

Q Question

C Comment

S Suggestion

R Re-formulation or change needed (usually followed by a suggestion)

→ Suggested change/correction

3.2 Title

1.1 **S:** ‘Human facilitated...’ → Human-facilitated...

3.3 Abstract

1.29 **S:** Insert comma after ‘metapopulation framework’

1.37 **R:** ‘...within-infestations...’ → ...within infestations,...

1.38–39 **S:** ‘...origin of founders and that...’ → ...origin of all founders of a given sub-population, and that...

1.39 **S:** Split the sentence between ‘were low;’ and ‘implying that’: ...were long. This implies that...

1.42 **C:** Again, it should be clear that this means ‘all founders of a given sub-population’, as opposed to all founders of any sub-population. Compare e.g. to 1.68–69, where this is made clear.

1.43–44 **C:** I missed the conclusions/suggestions for targeted control that are promised here. If this is seen as a future step, I would not necessarily mention it in the Abstract, but only in the Discussion.

3.4 Introduction

1.67–79 **C:** I wonder to what extent these predictions – derived under the assumption of the infinite-alleles model and applicable to the infinite-sites model – carry over to the (generalized) stepwise model of mutation assumed in this study. Also, if any refuge would be founded by essentially one fertilized female, isn’t it that the propagule and migrant pool models coincide? In that case, I would expect a problem of identifiability if founder numbers are very small. **Q:** Related to this, I missed a mentioning of traumatic insemination being typical for bed bugs, and potential implications for this study. Is it known whether females have control over sperm-egg allocation after multiple matings? Could it be that, even though mated by multiple males, females select sperm essentially from one male only? This would lead to a difference between ecological and effective genetic colonization dynamics and should at least be mentioned.

1.105–106 **Q:** Why is this expected? What exactly is meant by ‘limited’?

3.5 Materials and Methods

1.147–149 **C:** I did not understand this sentence. There should be a concise and explicit definition of terms like property, refuge, infestation, and sub-population at some point.

1.174 **S:** Insert comma after ‘For each dataset’ **C:** Please be consistent and either use ‘dataset’ or ‘data set’ throughout the whole manuscript.

1.181–182 **R:** ‘...to allow for multiple tests...’ → ...to account for multiple testing... **C:** What would less stringent methods of correction suggest, e.g., based on false discovery rates as sug-

gested by BENJAMINI and YEKUTIELI (2001) or BENJAMINI and HOCHBERG (1995). Later in the paper, such alternatives are used.

1.188 **S**: ‘...errors and were bootstrapped ...’ → ...errors, and bootstrapped...

1.192–194 **C**: I did not understand this sentence. Where the intervals chosen such that each interval is comprised of the same number of samples (sample locations)?

1.200–203 **C**: From the results section (e.g. Table 3), I got the impression that there was no significant deviation from Hardy-Weinberg equilibrium. Therefore, I wonder why the authors did not complement their analysis by running a clustering and assignment method like STRUCTURE. In particular, an assignment test could give important hints regarding the pattern of colonization. I might corroborate the support for the propagule model.

1.204 **S**: Delete ‘subsequent’

1.208 **Q**: Is ‘proportion of successful observed discriminations’ really what is meant here?

1.211 **Q**: Does ‘group’ mean ‘infestation’ in the context of dataset one, and ‘refuge’ in the context of dataset two? Does the issue of small numbers of samples really apply to both levels? Table S1 suggests that at least the infestations in London and Australia were sampled at considerable depth.

1.212 **S**: ‘... principal components (PCs) but in both cases the...’ → ... principal components (PCs), but for both datasets, the...

1.218 **R**: ‘ABC allows likelihood-free inference in complex scenarios with many parameters, by comparing...’ → ABC allows inference without explicit calculation of likelihoods in complex scenarios with many parameters. It compares... **C**: Likelihoods are present implicitly in ABC, which is why the term ‘likelihood-free’ is misleading.

1.221–222 **S**: ‘...of historical population models,...’ → of models of population history, ...

1.222 **C**: ABC is not restricted to coalescent simulations. **R**: ‘...simulation of many data sets using a coalescent approach and ...’ → ...simulation of many data sets (e.g. using a coalescent approach) and ... **C**: See earlier comment about using ‘dataset’ or ‘data set’ throughout.

1.224 **C**: I have some reservation against ‘rigorous’ here, as there is a lot of non-automated tuning and uncertainty involved in ABC.

1.234 **R**: ‘...such that colonists all come...’ → ...such that colonists founding a particular infestation all come...

1.242–257 **C**: I think that t_2 and t_3 should be allowed to vary among refugia, unless it is possible to show that variation in t_2 and t_3 is sufficiently captured in N_e and N_b . But, on top of that, refugia might also vary in N_e and N_b . This could be modeled using a hierarchical approach, with hyper parameters describing the distribution of, say t_2 , among refugia, and each individual t_2 being drawn from this distribution. Check recent ABC literature for examples.

1.259 **C**: Please avoid ‘uninformative priors’, as there is strictly speaking no uninformative prior. In this particular case, for instance, it makes a huge difference whether one chooses a uniform or a log-uniform prior, without there being any plausible confidence in one or the other being more uninformative.

1.274–276 **C**: This is no argument against applying at least one of the methods proposed over the last five years. For some of these methods (e.g. WEGMANN *et al.*, 2010; AESCHBACHER *et al.*, 2012), there exists software and it should not be too hard to apply them.

1.276–279 **C**: I don’t think this is a valid argument neither. The above-mentioned methods do not only pick summary statistics from a set of candidates, but may also help reducing the dimensionality of the problem, and hence improve the robustness of the inference.

1.291 **S**: Insert comma after ‘scenarios’. **Q**: Is the polychromatic weighted regression here the same as a weighted logistic regression?

- 1.292 Q:** How robust are the results to alternative choices of the rejection tolerance?
- 1.293–295 S:** Without going too much into detail, could the authors summarize describe the idea of this approach?
- 1.298–299 C:** Importance of assessing goodness of fit is not specific to ABC, but to any model-based approach to inference.
- 1.302–304 R:** In spite of the risk of over fitting, I would like to see the fit with the original summary statistics, just for comparison.
- 1.314** §I suggest adding a sentence saying that ABC was run on each POD.

3.6 Results

- 1.324 Q:** Were 154 or 156 samples genotyped? Compare 1.324 to Table 2.
- 1.325–326 Q:** How was this test for deviation of HWE done? How many loci showed significant deviation from HWE, i.e. was the overall signal a consequence of a few outlier loci or was it a general trend?
- 1.330–334 Q:** What impact could the mating system have on low/variable F_{IS} , as opposed to the actual number of colonizing individuals? See also comment to 1.67–79 above.
- 1.335–337 Q:** Was there a correction for multiple testing here? Which one?
- 1.342 Q:** Are these now infestations with one refuge only?
- 1.350–351 R:** I would be interested in seeing results of an assignment method such as STRUCTURE.
- 1.354 C:** Remind the reader of what original data the PCA was based on.
- 1.355 C:** But Figure S4 also shows that scenario 1 was misspecified. It is then no surprise that only scenario 2 obtains support. However, Figure S4 also shows that there might be scope for models fitting the data even better than scenario 2. See general issues 1 to 3 above.
- 1.361 S:** Insert comma after ‘... low (5.2%)’
- 1.366–374 C:** I missed a comment on t_3 , for which information seems limited, too. Yet, this is *the* parameter making scenario 2 different from 1.

3.7 Discussion

- 1.379–381** Related to general issues 1 to 3, I am irritated about there being no explicit modeling of extinction and colonization. Hence, I think this study was not within a metapopulation framework. The models were stripped down to classical models of structured populations. Therefore, I think this is an overstatement here.
- 1.381–382 Q:** But did LON2 not show significant F_{ST} ? **S:** Insert comma after ‘within-infestation diversity’
- 1.385–387 C:** A single founding event can also occur under the migrant pool model (scenario 1).
- 1.386 C:** Another term (‘dwelling’) that was not properly defined and can lead to confusion.
- 1.388–392 C:** If passive dispersal is only via colonization (not migration between existing sub-populations), then modeling the migrant pool as being panmictic is wrong and the meaning of N_s questionable. See general issues 1 to 3.
- 1.397–398 C:** As mentioned in general issue 3, this is an overstatement. The models (scenarios) compared here are not necessarily equivalent to those analyzed by [WADE and McCAULEY \(1988\)](#) and [PANNELL and CHARLESWORTH \(1999\)](#).
- 1.400 S:** Insert comma after ‘problematic’
- 1.406–408 C:** As mentioned above, I think the choice of summary statistics should be done using one of the semi-automated approaches that have been proposed before.

- 1.408–410 **C:** As mentioned above (1.242–257), variation in ages and sizes among sub-populations has been ignored, although this represents an important feature of metapopulations.
- 1.412–415 **C:** This is an appreciable extension. However, I worry that the two bottlenecks following each other end up being confounded and lead to what is described as a loss of information here. It also raises the question of whether ABC is giving the right answer to the wrong model when it favors scenario 2.
- 1.425 **S:** Omit ‘the information loss caused by’
- 1.430 **S:** Insert comma after ‘that’ and ‘analysis’.
- 1.431–433 **C:** I don’t think this conclusion is valid in its generality. Conclusions from ABC studies should be made for the specific models and case studies, and for a set of parameters for which thorough tests of performance have been done. The RMAE values in Table 1 say something about the performance of ABC in this particular case, but should not be extrapolated to potential other scenarios.
- 1.438 **R:** Delete comma after ‘study’
- 1.445 **S:** Insert comma after ‘studies’
- 1.449–453 **Q:** In view of traumatic insemination, could it be that effective mating with one male is due to other reasons than all males being related? Can females select sperm from particular matings to fertilise all their eggs? See also previous comments (1.67–79).
- 1.453–454 **Q:** Did the mode have a higher or lower RMAE than the median? **R:** I would like to see RMAEs for all point estimators in Table 1.
- 1.458–461 **Q:** Could this also be partially due to small per-refuge sample sizes?
- 1.472–478 **C:** The difference between the migrant and propagule model is not in the number of founder individuals per se, but in their origin and diversity. This should be accounted for in the formulation of this paragraph.
- 1.478 **S:** New paragraph after ‘of multiple founders.’
- 1.478–479 **Q:** What exactly is meant by ‘stochastic distribution of genetic diversity’?
- 1.482 **S:** Insert commas after ‘distance’, ‘detect it’ and ‘and’.
- 1.491 **S:** Insert ‘temporal variation in the environment,’ after ‘discontinuous habitat’
- 1.505–506 **C:** I find this an overstatement, because the favored model did not include *repeated* bottlenecks explicitly, and because it is unclear whether a history of repeated earlier bottlenecks can be appropriately represented by a panmictic population of size N_s .
- 1.509 **S:** New paragraph after ‘levels of differentiation’.
- 1.510–511 **C:** This sentence did not seem well-embedded into a context to me.
- 1.517 **S:** Insert comma after ‘In combination’
- 1.528–530 **C:** I found this sentence too complicated. **Q:** Should it say ‘repeated’ instead of ‘repeat’? Same question applies to 1.532.
- 1.530–534 **C:** This sentence seemed very long and hence difficult to understand.
- 1.539–541 **S:** ‘These markers, ...’ → The markers developed here,.... **C:** But this requires LD between markers and selected genes. In other words: Is the map of markers available now dense enough for such an undertaking?

4 References

- 1.567–569 **R:** The article by [BLUM *et al.* \(2013\)](#) has been published by now.
- 1.576 **R:** Adjust style (format of volume, number and page range).
- 1.592 **R:** Check page range.

1.600 R: Add volume number and check page range.
1.640 R: Adjust style (format of volume, number and page range).
1.654 R: Adjust style (format of volume, number and page range).
1.666 R: Adjust style (format of volume, number and page range).
1.668 R: Adjust style (format of volume, number and page range).
1.674 R: Adjust style (format of volume, number and page range).
1.682 R: Adjust style (format of volume, number and page range).
1.686–687 R: Adjust style (format of volume, number and page range).
1.712 R: Adjust style (format of volume, number and page range).
1.749 R: Adjust style (format of volume, number and page range).

4.1 Figures and tables

1.781–791 S: I suggest using subscripts, e.g. N_s instead of Ns .
1.819 R: In Table 1, please give RMAE for all point estimators.
1.828–830 C: There is some confusion about refuge and locality being the same or not.
1.830–832 C: Only overall values of H_e and H_o are shown. Therefore, it is difficult to interpret F_{ST} .
1.838–839 Q: Just to be sure: Was the sampling scheme in the ABC simulations using the same low sample sizes as those given in Table 3?
Fig. 3 R: I would like to see plots of $t_3 - t_2$ to see if t_3 is high just because t_2 is high, or if t_2 and t_3 actually balance each other. I would also like to see posteriors for N_s and N_b under scenario 1 (migrant pool model).

5 Supporting Information

5.1 Supporting text

Microsatellite isolation Q: Where did the individuals used for isolation of micro satellites come from? What is known about global diversity and substructure in *C. lectularius*, and could this have implications for microsatellite isolation (ascertainment)?
Microsatellite characterisation C: I think Table 4 should be Table S2. Regarding heterozygosity: I thought high rather than low observed heterozygosity would be an indication of a recent bottleneck. This is actually correctly stated in the main text.

5.2 Supporting figures

Suppl. Fig. 1 R: ‘between infestation’ \rightarrow between-infestation. **S:** I would not start a sentence with abbreviations (h in this case). **R:** Please add a scale and a windrose to the map. Also, correctly refer to the source of the cartographic material.
Suppl. Fig. 2 Q: Should there be a full stop rather than a comma after ‘analysis’ in the last sentence of the caption?

5.3 Supporting table

Suppl. Tab. 2 R: Mention N_A explicitly in the caption (at least treat T_a and N_A consistently).
Suppl. Tab. 5 C: I wonder if a visual representation could be given in the SI. This might make it easier for the reader to extract the relevant information.

References

- AESCHBACHER, S., M. A. BEAUMONT, and A. FUTSCHIK, 2012 A novel approach for choosing summary statistics in approximate Bayesian computation. *Genetics* **192**: 1027–1047.
- BENJAMINI, Y. and Y. HOCHBERG, 1995 Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**: 289–300.
- BENJAMINI, Y. and D. YEKUTIELI, 2001 The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**: 1165–1188.
- BLUM, M. G. B., M. A. NUNES, D. PRANGLE, and S. A. SISSON, 2013 A comparative review of dimension reduction methods in approximate Bayesian computation. *Stat. Sci.* **189**: 513–525.
- PANNELL, J. R. and B. CHARLESWORTH, 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664–676.
- WADE, M. J. and D. E. MCCAULEY, 1988 Extinction and recolonization: Their effects on the genetic differentiation of local populations. *Evolution* **42**: 995–1005.
- WEGMANN, D., C. LEUENBERGER, S. NEUENSCHWANDER, and L. EXCOFFIER, 2010 ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**.