

Peer review on manuscript

"Determining causes of genetic differentiation in [REDACTED] to identify the need for conservation actions"

by Peer 578

**ADDED INFO
ABOUT
FEATURED
PEER REVIEW**

This peer review is written by Dr. Sean Hoban,
Post-doctoral Research Fellow at
National Institute for Mathematical and Biological Synthesis
University of Tennessee, USA

PEQ = 4.8 / 5
Peer reviewed by 3 Peers

Introduction

This manuscript concerns conservation genetics of [REDACTED] in [REDACTED]. The authors use population-genetic simulations combined with empirical genetic data to infer past demographic processes (population size, connectivity) of a population that is potentially isolated from the main range by urbanization and agriculture. Using simulations for inference is a powerful approach that is increasingly common in population and theoretical genetics (1,2) but has rarely been employed by applied conservation biologists and conservation geneticists. Therefore, in my opinion, the manuscript contributes substantial insight to this system and the topic of connectivity, and is a valuable framework for objective, quantitative analyses and conclusions in future studies.

Revision Recommendations

Question:	Accept
Data:	Minor
Methods:	Major
Inference:	Minor
Writing:	Minor

Knowledge of population history is important because management interventions and population protection status are most effective when tailored to particular processes, e.g. historic isolation, anthropogenic barriers, etc. (3,4). Simulations can help determine what past processes might have influenced the genetic patterns currently observed in natural populations (1,5). The idea is to produce simulated genetic data under several plausible, competing demographic and genetic scenarios (e.g. fragmentation, bottlenecks, colonization), record statistics summarizing the simulated data under these scenarios, and compare the simulated distribution of summary statistics to statistics calculated on real data (6,7). If the real data "match" the simulated data (i.e. observed genetic signatures are not significantly different from data under a particular simulated scenario), that scenario is a likely candidate for the true population history.

After inferring bottlenecks and connectivity reductions, the authors then use additional simulations to forecast future population responses to management intervention (sensu (8,9)), to determine for maintaining genetic diversity.

Merits

Conservation and population genetics have traditionally relied on ad hoc interpretations of genetic summary statistics, e.g. inferring a bottleneck if diversity is low relative to another populations/species, often leading to "story-telling," confirmation bias, and overinterpretation of data (2,10,11). As the authors astutely indicate, such subjective conclusions are problematic because low diversity, for example, can stem from numerous processes (12,13).

The simulation approach has facilitated stronger inference over the past ten years in molecular ecology and population genetics. Importantly, I have seen few papers in conservation biology employing simulations for inference; this manuscript is certainly at the forefront of its field.

The introduction provides clear, instructive background on simulations and their utility and connects their study to broader issues in conservation biology (e.g. lines 72-76, 77- 88, 91-92, 132-138, 626-632). The paper is well-structured (logical, step-by-step) and well-written (cohesive paragraphs, articulate statements).

The authors also employ sensitivity analysis (14,15) to explore a range of values for key parameters in the simulation models- migration rates, population sizes and time-scale.

Critique

Unfortunately, there are substantial flaws in the simulations and in reporting of simulated data which prevent me from recommending publication, at present.

A crucial flaw is that the authors "hide" the simulation data (on which the major conclusions are based). Appendix-1, comparing simulated and observed data, simply states whether the comparison was "high," "low," or "overlap/non-significant difference". It is necessary for the authors to report (numerically) the simulated data range, to show whether the observed data falls within, and "how close" the real and simulated data are, which is key information for readers. This is often shown graphically (see (16) Fig 3, (17) Fig 3, (18) Fig 6), but the authors could simply replace "high", "low" etc., with the numerical range from the simulations.

Additionally more simulation replicates are vitally needed- only 10 were performed per scenario, while similar investigations typically involve >500-1000 replicates (13,16,19- 24). Fifty is a bare minimum, to obtain 90-95% intervals to compare to the observed data. The authors

(unconventionally) use standard error (line 241) instead of 95% intervals (standard error will likely be much narrower), but a sample of 10 is a tenuous basis for concluding "no significant difference." This is not merely a picky point about "the authors need more data"; unlike in ecology field studies, computer simulations entail no real cost for sufficient sampling (see additional comments for suggestions).

Another principal flaw is absence of discussion of several central caveats/assumptions. The Discussion must admit that the simulations do not employ realistic life history (for example, generational overlap, which affects effective size and genetic drift (25-27)). The authors should also remark that more complex simulations (e.g. Vortex (28,29)) might be a better model for genetic diversity loss and may show different results. Additionally the assumptions of N_e analysis, which are violated (30-32), must be noted (no migration, no generation overlap, no recent demographic change).

Another flaw is the reliance on only diversity-based summary statistics; differentiation-based statistics are not deployed at all, despite "differentiation" appearing in the title and throughout the manuscript. Both diversity- and differentiation-based summary statistics (usually F_{ST}) are nearly always used in similar investigation of multiple population simulations. Using STRUCTURE to examine population subdivision is interesting and useful, but not sufficient for comparing simulated and observed data- it is somewhat qualitative, and was used on only three datasets per scenario. The hallmark of this manuscript is the objective, quantitative nature of simulation work, but the interpretation of STRUCTURE results (lines 364-370, 376-378, 389-395, etc.) is entirely subjective (e.g. different cluster membership thresholds per scenario, discussing results in "cherry-picking" fashion, STRUCTURE results not used for scenario 5). Using F_{ST} (and/or R_{ST}) solves these problems, allowing the same quantitative comparison used on diversity statistics- e.g. does the observed value fall within the range of simulated data or not. Additionally, F_{IS} and M -ratio are valuable summary statistics (33-35) for comparing scenarios, due to the unique information they capture. Testing F_{ST} , F_{IS} , and M -ratio (and placing results as columns in Appendix-2) is highly advised.

Discussion

The authors use simulation methods to infer processes influencing genetic data, which represents a paradigm shift in conservation genetics methodology. However, the authors do not use sufficient simulation replicates to establish convincing conclusions, do not report summaries of simulation data, and do not calculate sufficient summary statistics. It is therefore not possible to judge, as yet, the reliability of the conclusions. Fortunately, these deficiencies are easily rectified.

References

1. Hoban S, Bertorelle G, Gaggiotti OE. Computer simulations: tools for population and evolutionary genetics. *Nat. Rev. Genet.* [Internet]. Nature Publishing Group; 2012 Feb [cited 2012 Mar 2];13:110-22. Available from: <http://dx.doi.org/10.1038/nrg3130>
2. Marjoram P, Tavaré S. Modern computational approaches for analysing molecular genetic variation data. *Nat. Rev. Genet.* [Internet]. Nature Publishing Group; 2006 Oct [cited 2013 May 21];7(10):759-70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16983372>
3. DeYoung RW, Honeycutt RL. The molecular toolbox: genetic techniques in wildlife ecology and management. *J. Wildl. Manage.* [Internet]. The Wildlife Society; 2005;69(4):1362-84. Available from: [http://dx.doi.org/10.2193/0022-541X\(2005\)69\[1362:TMTGTI\]2.0.CO;2](http://dx.doi.org/10.2193/0022-541X(2005)69[1362:TMTGTI]2.0.CO;2)
4. Santamaria L, Mèndez PF. Evolution in biodiversity policy-current gaps and future needs. *Evol. Appl.* 2012;5:202-18.
5. Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, et al. Utility of computer simulations in landscape genetics. *Mol. Ecol.* [Internet]. 2010;19(17):3549-64. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2010.04678.x>
6. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian Computation in Population Genetics. *Genetics* [Internet]. 2002;162(4):2025-35. Available from: <http://www.genetics.org/cgi/content/abstract/162/4/2025>
7. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* [Internet]. 1999;16(12):1791-8. Available from: <http://mbe.oxfordjournals.org/content/16/12/1791.abstract>
8. Bruford MW, Ancrenaz M, Chikhi L, Lackman-Ancrenaz I, Andau M, Ambu L, et al. Projecting genetic diversity and population viability for the fragmented orang-utan population in the Kinabatangan floodplain, Sabah, Malaysia. *Endanger. Species Res.* 2010;12:249-61.
9. Hedrick PW. Gene Flow and Genetic Restoration: The Florida Panther as a Case Study. *Conserv. Biol.* [Internet]. Blackwell Publishing for Society for Conservation Biology; 1995;9(5):996-1007. Available from: <http://www.jstor.org/stable/2387039>
10. Osmond DL, Nadkarni NM, Driscoll CT, Andrews E, Gold AJ, Allred SRB, et al. The role of interface organizations in science communication and understanding. *Front. Ecol. Environ.* [Internet]. 2010 Aug [cited 2012 Jul 28];8(6):306-13. Available from: <http://www.esajournals.org/doi/abs/10.1890/090145>
11. Sousa V, Hey J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat. Rev. Genet.* [Internet]. Nature Publishing Group; 2013; Available from: <http://dx.doi.org/10.1038/nrg3446>
12. Hundertmark KJ, Daele LJ Van. Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conserv. Genet.* 2010;11:139-47.
13. Brekke P, Bennett PM, Santure AW, Ewen JG. High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. *Mol. Ecol.* [Internet]. Blackwell Publishing Ltd; 2011;20(1):29-45. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2010.04923.x>

14. Naujokaitis-Lewis IR, Curtis JMR, Arcese P, Rosenfeld J. Sensitivity analyses of spatial population viability analysis models for species at risk and habitat conservation planning. *Conserv. Biol.* [Internet]. 2009 Feb [cited 2012 Mar 6];23(1):225-9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18798856>
15. Beissinger SR. Population Viability Analysis, past, present and future. In: Beissinger SR, McCullough DR, editors. *Popul. Viability Anal.* University of Chicago Press; 2002. p. 5-17.
16. Ficetola GF, Bonin A, Miaud C. Population genetics reveals origin and number of founders in a biological invasion. *Mol. Ecol.* [Internet]. 2008;17(3):773-82. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2007.03622.x>
17. Mardulyn P, Mikhailov YE, Pasteels JM. Testing phylogeographic hypotheses in a eurosiberian cold-adapted leaf beetle with coalescent simulations. *Evolution (N. Y.)*. [Internet]. Blackwell Publishing Inc; 2009;63(10):2717-29. Available from: <http://dx.doi.org/10.1111/j.1558-5646.2009.00755.x>
18. Mardulyn P, Milinkovitch MC. Inferring contemporary levels of gene flow and demographic history in a local population of the leaf beetle *Gonioctena olivacea* from mitochondrial DNA sequence variation. *Mol. Ecol.* [Internet]. Blackwell Science Ltd; 2005;14(6):1641-53. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2005.02537.x>
19. Alberto F, Raimondi PT, Reed DC, Coelho NC, Leblois R, Whitmer A, et al. Habitat continuity and geographic distance predict population genetic differentiation in giant kelp. *Ecology* [Internet]. 2010;91(1):49-56. Available from: <http://www.esajournals.org/doi/abs/10.1890/09-0050.1>
20. Estoup A, Guillemaud T. Reconstructing routes of invasion using genetic data: why, how and so what? *Mol. Ecol.* 2010;19:4113-30.
21. Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, Moritz C. Inferring Population History From Microsatellite and Enzyme Data in Serially Introduced Cane Toads, *Bufo marinus*. *Genetics* [Internet]. 2001;159(4):1671-87. Available from: <http://www.genetics.org/cgi/content/abstract/159/4/1671>
22. Puebla O, Bermingham E, McMillan WO. On the spatial scale of dispersal in coral reef fishes. *Mol. Ecol.* 2012;21:5675-88.
23. Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J-M. Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution (N. Y.)*. [Internet]. Blackwell Publishing Ltd; 2004;58(9):2021-36. Available from: <http://dx.doi.org/10.1111/j.0014-3820.2004.tb00487.x>
24. Johansson M, Primmer CR, Merilä J. History vs. current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). *Mol. Ecol.* [Internet]. 2006 Apr [cited 2012 Nov 30];15(4):975-83. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2006.02866.x>
25. Hoelzel AR. Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biol. J. Linn. Soc.* 1999;68:23-39.

26. Hill WG. A note on effective population size with overlapping generations. *Genetics*. 1979;92(317-322):317-22.
27. Nunney L. The influence of variation in female fecundity on effective population size. *Biol. J. Linn. Soc.* [Internet]. Blackwell Publishing Ltd; 1996;59(4):411-25. Available from: <http://dx.doi.org/10.1111/j.1095-8312.1996.tb01474.x>
28. Lacy RC. Vortex: a computer simulation model for population viability analysis. *Wildl. Res.* 1993;20:45-65.
29. Lacy RC. Structure of the VORTEX Simulation Model for Population Viability Analysis. *Ecol. Bull.* 2000;48:191-203.
30. Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. *Mol. Ecol.* [Internet]. Blackwell Publishing Ltd; 2007;16(5):953-66. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2006.03206.x>
31. Waples RS. Evaluating the effect of stage-specific survivorship on the $N(e)/N$ ratio. *Mol. Ecol.* [Internet]. 2002 Jun;11(6):1029-37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12090236>
32. Tallmon DA, Gregovich D, Waples RS, Baker CS, Jackson J, Taylor BL, et al. When are genetic methods useful for estimating contemporary abundance and detecting population trends? *Mol. Ecol. Resour.* [Internet]. 2010 Jul [cited 2012 Oct 5];10(4):684-92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21565073>
33. Peter BM, Wegmann D, Excoffier L. Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Mol. Ecol.* [Internet]. Blackwell Publishing Ltd; 2010;19(21):4648-60. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2010.04783.x>
34. Csilléry K, Blum MGB, Gaggiotti OE, François O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* [Internet]. 2010 Jul [cited 2012 Mar 11];25(7):410-8. Available from: <http://www.sciencedirect.com/science/article/B6VJ1-503WMFS-1/2/64bcd19f311b55ad28dc6bc889caad6d>
35. Ray N, Wegmann D, Fagundes NJR, Wang S, Excoffier L. A Statistical Evaluation of Models for the Initial Settlement of the American Continent Emphasizes the Importance of Gene Flow with Asia. *Mol. Biol. Evol.* 2010;27(2):337-45.
36. Bertorelle G, Benazzo A, Mona S. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* [Internet]. 2010 Jul [cited 2012 Mar 11];19(13):2609-25. Available from: <http://dx.doi.org/10.1111/j.1365-294X.2010.04690.x>
37. Beaumont MA. Approximate Bayesian Computation in Evolution and Ecology. *Annu. Rev. Ecol. Evol. Syst.* [Internet]. 2010;41(1):379-406. Available from: <http://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-102209-144621>
38. Pascual M, Chapuis MP, Mestres F, BalanyÀ J, Huey RB, Gilchrist GW, et al. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC

- methods. *Mol. Ecol.* [Internet]. Blackwell Publishing Ltd; 2007 Aug [cited 2012 Mar 19];16(15):3069-83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17651188>
39. Hardy OJ, Charbonnel N, Freville H, Heuertz M. Microsatellite Allele Sizes: A Simple Test to Assess Their Significance on Genetic Differentiation. *Genetics* [Internet]. Test to Assess Their Significance on Genetic Differentiation. *Genetics* [Internet]. 2003;163(4):1467-82. Available from: <http://www.genetics.org/cgi/content/abstract/163/4/1467>
40. Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C. An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. *Genetics* [Internet]. 1997;147(4):1943-57. Available from: <http://www.genetics.org/content/147/4/1943.abstract>
41. Paetkau D, Calvert W, Stirling I, Strobeck C. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* [Internet]. Blackwell Publishing Ltd; 1995;4(3):347-54. Available from: <http://dx.doi.org/10.1111/j.1365-294X.1995.tb00227.x>
42. Hoban SM, Gaggiotti OE, Bertorelle G. The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulationbased study. *Mol. Ecol.* 2013;22(13):3444-50.
43. Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont M a. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* [Internet]. 2010 Nov [cited 2012 Mar 12];186(3):983-95. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2975287&tool=pmcentrez&rendertype=abstract>
44. Paz-Vinas I, Quemere E, Chikhi L, Loot G, Blanchet S. The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Mol. Ecol.* 2013;22:3279-91.
45. Peng B, Chen H, Mechanic LE, Racine B, Clarke J, Clarke L, et al. Genetic Simulation Resources (GSR): A website for the registration and discovery of genetic data simulators. *Bioinformatics.* 2013;
46. Beerli P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Mol. Ecol.* [Internet]. Blackwell Science Ltd; 2004 Apr [cited 2013 May 23];13(4):827-36. Available from: <http://doi.wiley.com/10.1111/j.1365-294X.2004.02101.x>
47. Beerli P. *Migrate Documentation.* 2010;

Additional comments for authors

First some additional key comments (important!), and then line-by-line edits (less important)...

Regarding the additional replications needed for each scenario. I recognize that hundreds of simulations are computationally demanding, but they are within the capabilities of even small computer clusters or even on individual computers (over the time of days to weeks). If the authors

do not have access to their own or a government/university cluster, there are numerous institutions in Canada, the USA, and Europe which offer use of their computer cluster at small or no cost (<http://rna.urmc.rochester.edu/bioinfo.html>, <http://cse.ucsd.edu/node/251>, <http://www.biotech.uconn.edu/bf/>, <http://www.bioinformatics.ku.edu/resources/biocluster>, <http://cagt.bu.edu/page/Bcluster>, note: I am not advertising my own institution in this list in order to maintain neutrality). The time spent on additional simulations will be very much worth the small additional effort.

The authors should note in discussion that Approximate Bayesian Computation ((6,34,36,37)) is a more formal and advisable method for comparing simulated and observed data than the "grid" method used (which however was sufficient for the task).

I think the Methods is unbalanced regarding the main focus of the paper- the authors spend more than three pages describing the bottleneck and effective population size methods, which are a minor component of the study, and much of which could be placed in the appendix. More space should be devoted in the main manuscript to the simulations (parameter space, assumptions, sensitivity analysis), which only receive one paragraph currently. I understand that simulations are explained in appropriate detail in the Appendix 1, but unfortunately many readers do not utilize appendices and online material, so please attempt to place as much of this information as possible within the main manuscript Methods- several paragraphs at least.

You monitor heterozygosity as the response over the next hundreds of years with ██████, but why not monitor number of alleles or inbreeding coefficient (or level of relatedness), which might be more sensitive to low population size? These might also show "thresholds" or "vortexes" in which loss accelerates more than linearly, perhaps.

Again, I highly emphasize that using only within-population diversity measures for comparing simulated and observed data is unconvincing, and would set a poor precedent if the authors with this paper to be a model for future work (line 636)... As per this, note that other options for among-population diversity measures (other than FST/RST) are number of shared alleles, Goldstein's delta mu squared statistic, and Paetkau's mean assignment likelihood (38-41).

I think the Introduction is a bit lengthy, and is a bit heavy with citations- could be reduced some; same for the Discussion. Don't remove any points, just try to make the sentences tighter, use less repetition, etc..

A series of papers (24,42-44) have looked at the power and error rates of the M-ratio and heterozygote excess test to detect real bottlenecks, and some of their findings should be mentioned when interpreting the bottleneck results, in terms of possibility of false positives and also "missed" bottleneck signals.

It seems that (1) should be cited, as it is probably the most comprehensive recent review of the use of simulations in ecology and evolution and the most thorough resource of existing simulation software (see also (45)).

There is also insufficient/unclear explanation of how this paper and the data therein. There is also insufficient/unclear explanation of how this paper and the data therein overlaps with other recent papers on these populations... Some information on this can be garnered throughout the manuscript, but it is scattered. Please include several sentences that clearly state, in one paragraph, what data previous studies collected and what methods they used (did any use Structure or simulations?), and state clearly what the new study does provide.

If the authors continue to use STRUCTURE results, they should set some quantitative means to compare observed and simulated data, such as mean q values- the discussion of STRUCTURE results does not seem very quantitative. Q values are mentioned but it seems a different "threshold" was used each scenario rather than determining if the observed value fell within the standard deviation of the simulated. Of course, F_{ST} and the other statistics will help here. Importantly, the lengthy, somewhat qualitative presentation of STRUCTURE results for each simulated scenario is a substantial "bump" in the otherwise nice flow of the paper.

In some PVA studies of large mammals, it has been suggested that ■■■, as the authors allude to in the abstract. The authors only tested ■■■, and I think it is worth testing if ■■■. Line 338-339 seems to imply that the authors did test ■■■ but no results are reported regarding this (not included in Fig 4.3)... In any case please clarify.

Following are some small in-line changes

Line 68 change "and" to "or"

Line 86- suggest to use another term than "non-natural"... a very vague term... what is "natural?"
Same in line 456

Line 188 "small" is relative... in many parts of the ■■■, this size population would not be considered "small", so I suggest changing to "relatively small"

Line 227- should be "in order to"?

Line 296- "we can note that for microsatellite data, the SMM or TPM with $pg=90\%$ are usually more appropriate"... I think you meant 10%?

Line 492- what do you mean by "performing better"?

Line 605- you mean "phenotypic polymorphism" or "genetic"? Regarding source sink discussion on page 23, you could confirm this with the software MIGRATE (46,47) which should reveal asymmetries in migration rates.

Figure 4.1- please zoom in the figure 4, as the sampling zone only takes up maybe 20% of this figure, the rest is basically empty so the part we want to look at is tiny. It is hard to see the sampling symbols (and I have a young person's eyes). Also, the WMU boundaries (wildlife management units??) make it cluttered and hard to see the symbols. Maybe remove them? Lastly, numbers might be easier to read than the small triangles etc. ■ is hard to find, the symbols for ■ and ■ etc.

Figure 4.2- a bit confusing especially regarding the bottleneck- it doesn't look like a population reduction, it looks like a wavy line.