



**HAL**  
open science

# Estimating recent divergence history: making the most of microsatellite data and Approximate Bayesian Computation approaches

Takeshi Kawakami, Concetta Burgarella

► **To cite this version:**

Takeshi Kawakami, Concetta Burgarella. Estimating recent divergence history: making the most of microsatellite data and Approximate Bayesian Computation approaches. 2020, pp.100091. 10.24072/pci.evolbiol.100091 . hal-02551743

**HAL Id: hal-02551743**

**<https://hal.science/hal-02551743>**

Submitted on 23 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Estimating recent divergence history: making the most of microsatellite data and Approximate Bayesian Computation approaches

Open Access

[Takeshi Kawakami](#) and [Concetta Burgarella](#) based on reviews by Michael D Greenfield and 2 anonymous reviewers

Published: 20 January 2020

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

A recommendation of:

Marie-Pierre Chapuis, Louis Raynal, Christophe Plantamp, Christine N. Meynard, Laurence Blondin, Jean-Michel Marin, Arnaud Estoup.

**A young age of subspecific divergence in the desert locust *Schistocerca gregaria*, inferred by ABC Random Forest (2020)**, *bioRxiv*, 671867, ver. 4 peer-reviewed by Peer Community in Evolutionary Biology. [10.1101/671867](https://doi.org/10.1101/671867)

*Submitted: 20 June 2019, Recommended: 17 January 2020*

**Cite this recommendation as:**

Takeshi Kawakami and Concetta Burgarella (2020) Estimating recent divergence history: making the most of microsatellite data and Approximate Bayesian Computation approaches. *Peer Community in Evolutionary Biology*, 100091. [10.24072/pci.evolbiol.100091](https://doi.org/10.24072/pci.evolbiol.100091)

The present-day distribution of extant species is the result of the interplay between their past population demography (e.g., expansion, contraction, isolation, and migration) and adaptation to

the environment. Shedding light on the timing and magnitude of key demographic events helps identify potential drivers of such events and interaction of those drivers, such as life history traits and past episodes of environmental shifts. The understanding of the key factors driving species evolution gives important insights into how the species may respond to changing conditions, which can be particularly relevant for the management of harmful species, such as agricultural pests (e.g. [1]). Meaningful demographic inferences present major challenges. These include formulating evolutionary scenarios fitting species biology and the eco-geographical context and choosing informative molecular markers and accurate quantitative approaches to statistically compare multiple demographic scenarios and estimate the parameters of interest. A further issue comes with result interpretation. Accurately dating the inferred events is far from straightforward since reliable calibration points are necessary to translate the molecular estimates of the evolutionary time into absolute time units (i.e. years). This can be attempted in different ways, such as by using fossil and archaeological records, heterochronous samples (e.g. ancient DNA), and/or mutation rate estimated from independent data (e.g. [2], [3] for review). Nonetheless, most experimental systems rarely meet these conditions, hindering the comprehensive interpretation of results. The contribution of Chapuis et al. [4] addresses these issues to investigate the recent history of the African insect pest *Schistocerca gregaria* (desert locust). They apply Approximate Bayesian Computation-Random Forest (ABC-RF) approaches to microsatellite markers. Owing to their fast mutation rate microsatellite markers offer at least two advantages: i) suitability for analyzing recently diverged populations, and ii) direct estimate of the germline mutation rate in pedigree samples. The work of Chapuis et al. [4] benefits of both these advantages, since they have estimates of mutation rate and allele size constraints derived from germline mutations in the species [5]. The main aim of the study is to infer the history of divergence of the two subspecies of the desert locust, which have spatially disjoint distribution corresponding to the dry regions of North and West-South Africa. They first use paleo-vegetation maps to formulate hypotheses about changes in species range since the last glacial maximum. Based on them, they generate 12 divergence models. For the selection of the demographic model and parameter estimation, they apply the recently

developed ABC-RF approach, a powerful inferential tool that allows optimizing the use of summary statistics information content, among other advantages [6]. Some methodological novelties are also introduced in this work, such as the computation of the error associated with the posterior parameter estimates under the best scenario. The accuracy of timing estimate is assured in two ways: i) by the use of microsatellite markers with known evolutionary dynamics, as underlined above, and ii) by assessing the divergence time threshold above which posterior estimates are likely to be biased by size homoplasy and limits in allele size range [7]. The best-supported model suggests a recent divergence event of the subspecies of *S. gregaria* (around 2.6 kya) and a reduction of populations size in one of the subspecies (*S. g. flaviventris*) that colonized the southern distribution area. As such, results did not support the hypothesis that the southward colonization was driven by the expansion of African dry environments associated with the last glacial maximum, as it has been postulated for other arid-adapted species with similar African disjoint distributions [8]. The estimated time of divergence points at a much more recent origin for the two subspecies, during the late Holocene, in a period corresponding to fairly stable arid conditions similar to current ones [9,10]. Although the authors cannot exclude that their microsatellite data bear limited information on older colonization events than the last one, they bring arguments in favour of alternative explanations. The hypothesis privileged does not involve climatic drivers, but the particularly efficient dispersal behaviour of the species, whose individuals are able to fly over long distances (up to thousands of kilometers) under favourable windy conditions. A single long-distance dispersal event by a few individuals would explain the genetic signature of the bottleneck. There is a growing number of studies in phylogeography in arid regions in the Southern hemisphere, but the impact of past climate changes on the species distribution in this region remains understudied relative to the Northern hemisphere [11,12]. The study presented by Chapuis et al. [4] offers several important insights into demographic changes and the evolutionary history of an agriculturally important pest species in Africa, which could also mirror the history of other organisms in the continent. As the authors point out, there are necessarily some uncertainties associated with the models of past ecosystems and climate, especially for Africa. Interestingly, the

authors argue that the information on paleo-vegetation turnover was more informative than climatic niche modeling for the purpose of their study since it made them consider a wider range of bio-geographical changes and in turn a wider range of evolutionary scenarios (see discussion in Supplementary Material).

Microsatellite markers have been offering a useful tool in population genetics and phylogeography for decades, but their popularity is perhaps being taken over by single nucleotide polymorphism (SNP) genotyping and whole-genome sequencing (WGS) (the peak year of the number of the publication with “microsatellite” is in 2012 according to PubMed). This study reaffirms the usefulness of these classic molecular markers to estimate past demographic events, especially when species- and locus-specific microsatellite mutation features are available and a powerful inferential approach is adopted. Nonetheless, there are still hurdles to overcome, such as the limitations in scenario choice associated with the simulation software used (e.g. not allowing for continuous gene flow in this particular case), which calls for further improvement of simulation tools allowing for more flexible modeling of demographic events and mutation patterns. In sum, this work not only contributes to our understanding of the makeup of the African biodiversity but also offers a useful statistical framework, which can be applied to a wide array of species and molecular markers (microsatellites, SNPs, and WGS).

## References

- [1] Lehmann, P. et al. (2018). Complex responses of global insect pests to climate change. *bioRxiv*, 425488. doi: [10.1101/425488](https://doi.org/10.1101/425488) [2] Donoghue, P. C., & Benton, M. J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution*, 22(8), 424-431. doi: [10.1016/j.tree.2007.05.005](https://doi.org/10.1016/j.tree.2007.05.005) [3] Ho, S. Y., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., & Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular ecology*, 20(15), 3087-3101. doi: [10.1111/j.1365-294X.2011.05178.x](https://doi.org/10.1111/j.1365-294X.2011.05178.x) [4] Chapuis, M.-P., Raynal, L., Plantamp, C., Meynard, C. N., Blondin, L., Marin, J.-M. and Estoup, A. (2020). A young age of subspecific divergence in the desert locust *Schistocerca gregaria*, inferred by ABC Random Forest. *bioRxiv*, 671867, ver. 4 peer-reviewed and recommended by PCI Evolutionary Biology. doi:

[10.1101/671867](https://doi.org/10.1101/671867) [5] Chapuis, M.-P., Plantamp, C., Streiff, R., Blondin, L., & Piou, C. (2015). Microsatellite evolutionary rate and pattern in *Schistocerca gregaria* inferred from direct observation of germline mutations. *Molecular ecology*, 24(24), 6107-6119. doi: [10.1111/mec.13465](https://doi.org/10.1111/mec.13465) [6] Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2018). ABC random forests for Bayesian parameter inference. *Bioinformatics*, 35(10), 1720-1728. doi: [10.1093/bioinformatics/bty867](https://doi.org/10.1093/bioinformatics/bty867) [7] Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology*, 11(9), 1591-1604. doi: [10.1046/j.1365-294X.2002.01576.x](https://doi.org/10.1046/j.1365-294X.2002.01576.x) [8] Moodley, Y. et al. (2018). Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros (*Ceratotherium simum*). *Proceedings of the Royal Society B*, 285(1890), 20181567. doi: [10.1098/rspb.2018.1567](https://doi.org/10.1098/rspb.2018.1567) [9] Kröpelin, S. et al. (2008). Climate-driven ecosystem succession in the Sahara: the past 6000 years. *science*, 320(5877), 765-768. doi: [10.1126/science.1154913](https://doi.org/10.1126/science.1154913) [10] Maley, J. et al. (2018). Late Holocene forest contraction and fragmentation in central Africa. *Quaternary Research*, 89(1), 43-59. doi: [10.1017/qua.2017.97](https://doi.org/10.1017/qua.2017.97) [11] Beheregaray, L. B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17(17), 3754-3774. doi: [10.1111/j.1365-294X.2008.03857.x](https://doi.org/10.1111/j.1365-294X.2008.03857.x) [12] Dubey, S., & Shine, R. (2012). Are reptile and amphibian species younger in the Northern Hemisphere than in the Southern Hemisphere?. *Journal of evolutionary biology*, 25(1), 220-226. doi: [10.1111/j.1420-9101.2011.02417.x](https://doi.org/10.1111/j.1420-9101.2011.02417.x)

## Revision round #2

2019-12-04

Dear Dr. Chapuis and co-authors,

Your revised manuscript ("A young age of subspecific divergence in the desert locust, *Schistocerca gregaria*"), which you submitted to PCI Evolutionary Biology, along with your response to the comments made by the three expert reviewers,

have now been evaluated by us. We agree that the issues raised by the reviewers were comprehensively addressed, and the quality of the manuscript improved significantly. In particular, the entire manuscript has been nicely streamlined while focusing on the major discoveries (pointed out by #rev 1 and #rev 2). We still notice some minor typos (see below), but those do not diminish the impact of the discovery that this study made, and, therefore, we have no hesitation to recommend this article in PCI Evolutionary Biology.

Since both of us are not capable of checking small errors and typos due to the time constraint, please do read the manuscript carefully to further improve the readability.

Typos: Line 29: “the lat[t]er providing” Figure 1 legend: “b)” appears twice. The latter should be “c)” Line 224: “dbf / Nbf” -> “dB / NB” to make it consistent with Table 3.

Sincerely,

Dr. Concetta Burgarella Dr. Takeshi Kawakami Recommenders, PCI Evolutionary Biology

### **PS: Additional requirements of the Managing Board**

**Mandatory modifications.** As indicated in the 'How does it work?' section and in the code of conduct, please make sure that: -Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data. -Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused. -Details on experimental procedures are available to readers in the text or as appendices. -Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing

this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders." In order to reach a better referencing and greater visibility of your recommended preprint, we suggest you to do the following modifications : (i) add the following sentence in the acknowledgements: "Version 3 of this preprint has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (<https://doi.org/10.24072/pci.evolbiol.100091>) » ⇒ If you use bioRxiv to post your preprint, add this sentence also in the footnote section (a specific section of bioRxiv). Note that this DOI is not the DOI of your article, but the DOI of the recommendation text. The DOI of your article remains unchanged. Doing so is very important because it would: -indicate to readers that, unlike many other preprint in this server, your pre-print has been peer-reviewed and recommended. -make visible this information in Google Scholar search (which is quite important). (ii) In addition, we suggest you to remove line numbering from the preprint.

**Optional modifications** ==> Third, (if you wish) we advise you to use templates (word docx template and a latex template) to format your preprint in a PCI style. This is optional. Here is the links of the templates: <https://peercommunityin.org/templates/> Please be careful to correctly update all text in these templates (doi, authors' names, address, title, date, recommender first name and family name ...). Please be careful to also choose the badge "Open Code" if appropriate (in addition to the "Open access", "Open data" and "Open Peer-Review" badges). Indicate in the "cite as" box the version of the article that you are currently formatting. This should be version 3. In the reviewer section, indicate "Michael D Greenfield and two anonymous reviewers".

I hope this is clear. Do not hesitate to ask any help if you need. Once you have made these modifications, you should upload a new version of the article on the preprint server. Please tell us when you have done so. Thanks. Best, The MB of PCI Evol Biol.

*Preprint DOI:* <https://doi.org/10.1101/671867>



## Author's reply:

[Download author's reply \(PDF file\)](#)

## Revision round #1

2019-08-08

Dear Dr. Chapuis,

Three reviewers have assessed your manuscript. We both agree with reviewers that the study addresses several important issues (eg., the evolutionary history of agriculturally important species in the relatively understudied geographic region) by using carefully formulated analysis. From a technical point of view, it provides a useful statistical framework to discriminate possible demographic scenarios by using microsatellite markers.

At the same time, all three reviewers had numerous but constructive comments about the analyses and interpretation of the data. They pointed out that there are two or three different sections, such as the biological question of dating the divergence between the two subspecies and the methodological approach, but these are not optimally integrated. One possibility is to reduce the methodological details (by moving them to Supplementary Materials) and to focus more on the more biological themes in the main text. Other possible solutions are, of course, welcome.

A couple of minor comments: 1) Fig 1. It seems to me that the dark orange and light orange legend might be exchanged. Does the dark orange represent deserts (“extreme deserts” sensu Adams and Faure 1997) and the light orange represent xeric shrublands (“semi-deserts” sensu Adams and Faure 1997)?

2) Why distinguishing between untranscribed and transcribed microsatellites, if they have been previously shown to be independent and under neutrality (lines 544)? Have I missed the explanation?

We look forward to the revised manuscript.

Sincerely,

Dr. Concetta Burgarella Dr. Takeshi Kawakami Recommenders, PCI Evolutionary Biology

**Additional requirements of the managing board:** As indicated in the 'How does it work?' section and in the code of conduct, please make sure that: -Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused. -Details on experimental procedures are available to readers in the text or as appendices.

-Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders."

*Preprint DOI:* <https://doi.org/10.1101/671867>

Reviewed by anonymous reviewer, 2019-07-15 16:16

This preprint by Chapuis et al. estimates the divergence time of two subspecies of an African locust species, *Schistocerca gregaria*. The authors compared present day species distribution to the projected past distribution of associated habitats to formulate competing demographic models. They then used fast-evolving microsatellite markers and ABC-random forest inference for model selection and parameter estimation. The authors estimate a young subspecific divergence time with highest support for the demographic scenario with a bottleneck in the southern and ancestral population. I believe the authors did a thorough job in the analyses to infer the best demographic scenario and estimate parameters. The

authors provide a good empirical application of the abcrf method for demographic modeling which is an important contribution to make these types of inference more time-efficient and robust to correlated summary statistics.

My main concern with the manuscript at this state is that the main objective gets lost in the details of the ABC-RF methods. The manuscript initially reads as though the priority is to accurately date the divergence time of the two locus subspecies but then shifts priority to the utility of the ABC-RF method. As it is written, these two focuses are not connected cohesively in one story. If the focus is to lean towards the young divergence time of the subspecies (as the title suggests), there needs to be stronger background as to why estimating this parameter is particularly relevant. Why these two subspecies in particular? Why now? What new questions or avenues of research would open up? The divergence time of two subspecies would already be assumed to be quite young so the estimate should be made to be more broadly relevant outside of this species or at least expand on how this is particularly important. If the focus is to lean towards the application of ABCRF, then it would be important to actually include this in the title and have more of a focus in the introduction. It would then also be important to discuss what might be a novel contribution of applying the methods to this particular study system and question. Although it may already be discussed briefly, it would be important to further emphasize why this method provides a better or more time efficient alternative to traditional methods and how this particular study supports that.

Furthermore, the direct benefit of the paleo-veg inference is not immediately obvious. It might be helpful to include examples of demographic scenarios that was ruled out by taking this first step. It is also not clear from the methods if the inference of the past distribution was done qualitatively or quantitatively. From the methods section, it seems that the present day distribution map was qualitatively matched with particular habitat and this habitat was used as a proxy to infer past distribution. If this were to be included, it would be important to run this in a more quantitative manner by conducting niche modeling (ex. MAXENT) to infer the present day distribution from occurrence points to then find the relevant bioclimatic variables and the project these to the distribution of the

variables in the past. Before conducting such analyses, it would first be important to pin down the main focus and then add why it is relevant to include paleo-veg information.

Lastly, the relevance of the ‘evolution of phase polyphenism’ in the discussion section escapes me. Although it is an interesting point, it seems out of place without being mentioned at all in the introduction. I think this should either be removed or brought it earlier as a way to emphasize the relevance of this system which goes beyond the estimate of the divergence time.

In the end, these concerns are mainly with the broad organization and relevance of the paper. This preprint has good potential contributions if the story is pinned down.

Figure 1. The colors need a legend and it would be beneficial to include a small title for each panel (A-F) in the figure itself.

Figure 2. The evolutionary events (c, b, sc) should be written out and the variable be in the parentheses so this would be more informative.

Reviewed by [Michael D Greenfield](#), 2019-07-23 22:19

Review of manuscript ‘A young age of subspecific divergence in the desert locust *Schistocerca gregaria*’ by Chapuis et al.

This is an impressive manuscript that focuses on several important subjects in evolutionary biology, and it certainly should be / will be publishable in a major journal(s) following some revision. The authors address three overlapping subjects, each of which represents a major theme : 1) biology of *Schistocerca gregaria*, a species of considerable economic importance during its swarm phase ; 2) the question of population divergence in animal species that disperse very effectively over long distances ; 3) using molecular and quantitative methods to estimate the ancestry of populations that have diverged recently and cannot be studied with standard / classical phylogenetic approaches. As currently written, the manuscript integrates all three themes and is quite long, even without the supplementary material sections at the end. Thus, the authors should carefully

weigh the positive and negative points of a single article of 'monograph' format versus 2 or 3 separate articles. And if they opt for a single monograph, the journal to which it would be submitted needs much consideration. Themes 2 and 3 are far too important to be 'buried' in a monograph devoted to *Schistocerca* biology, the worldwide pest status of this species notwithstanding. Specific points : Treatment of theme 2 would be improved by comparison with other species exhibiting similar ecologies and evolutionary histories. First, the origin of the New World *Schistocerca* species should be discussed, as it is argued (see papers by R.F. Chapman et al) that they are all descended from the Old World (African) *Schistocerca gregaria* : A single trans-Atlantic founder event to NE Brazil, followed by (adaptive) radiation. The difference between this case and that treated by the authors of the submitted manuscript is that the founder event in Brazil is a bit older (Pleistocene). Interestingly, none of the New World *Schistocerca* species exhibit a change from solitary to swarm phase. Second, the Monarch butterfly in North America exhibits a population structure that is roughly similar to *Schistocerca gregaria* in Africa : The major population is found in eastern North America, and a small population is found in on the West Coast. Both populations undergo an annual north-south migration each year, and admixture is believed to be minimal. Some of the points discussed in the supplementary material deserve integration in the main body of the manuscript. For example, the question raised in S4 on the possibility of Pleistocene colonization of southern Africa is too critical for relegation to an addendum, which is unlikely to be read. And regarding this possibility, one explanation is that such colonization had occurred but went extinct. This type of scenario has been proposed for the West Coast population of the Monarch butterfly in North America. Evaluation of the 8 different evolutionary scenarios for recent (late Holocene) divergence of *Schistocerca gregaria* is extremely difficult to follow in Figure 4 and the Tables. I recommend a much simpler presentation of the ABC-RF information in the Figure and Tables, with details placed in the supplementary materials. The writing is generally clear, particularly in the beginning of the manuscript, but there are placed where clarity could be / should be improved. I attach an annotated pdf with some suggestions.

Michael Greenfield



[Download the review \(PDF file\)](#)

Reviewed by anonymous reviewer, 2019-07-24 15:15

[Download the review \(PDF file\)](#)

**Author's reply:**

[Download author's reply \(PDF file\)](#)