University of Groningen

Inferring past demographic changes from contemporary genetic data

Cabrera, Andrea A; Palsbøll, Per J

Published in:
Molecular Ecology Resources

DOI:
[10.1111/1755-0998.12696](https://doi.org/10.1111/1755-0998.12696)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Cabrera, A. A., & Palsbøll, P. J. (2017). Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIYABC. *Molecular Ecology Resources*, 17(6), e94-e110. <https://doi.org/10.1111/1755-0998.12696>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIYABC

Andrea A. Cabrera | Per J. Palsbøll

Marine Evolution and Conservation,
Groningen Institute of Evolutionary Life
Sciences, University of Groningen,
Groningen, The Netherlands

Correspondence

Andrea A. Cabrera and Per J. Palsbøll,
Marine Evolution and Conservation,
Groningen Institute of Evolutionary Life
Sciences, University of Groningen,
Groningen, The Netherlands.
Emails: andrea_ca_gt@yahoo.com;
palsboll@gmail.com

Funding information

University of Groningen

Abstract

Inferring the demographic history of species and their populations is crucial to understand their contemporary distribution, abundance and adaptations. The high computational overhead of likelihood-based inference approaches severely restricts their applicability to large data sets or complex models. In response to these restrictions, approximate Bayesian computation (ABC) methods have been developed to infer the demographic past of populations and species. Here, we present the results of an evaluation of the ABC-based approach implemented in the popular software package DIYABC using simulated data sets (mitochondrial DNA sequences, microsatellite genotypes and single nucleotide polymorphisms). We simulated population genetic data under five different simple, single-population models to assess the model recovery rates as well as the bias and error of the parameter estimates. The ability of DIYABC to recover the correct model was relatively low (0.49): 0.6 for the simplest models and 0.3 for the more complex models. The recovery rate improved significantly when reducing the number of candidate models from five to three (from 0.57 to 0.71). Among the parameters of interest, the effective population size was estimated at a higher accuracy compared to the timing of events. Increased amounts of genetic data did not significantly improve the accuracy of the parameter estimates. Some gains in accuracy and decreases in error were observed for scaled parameters (e.g., $N_e\mu$) compared to unscaled parameters (e.g., N_e and μ). We concluded that DIYABC-based assessments are not suited to capture a detailed demographic history, but might be efficient at capturing simple, major demographic changes.

KEYWORDS

approximate Bayesian computation, demographic inference, model selection, population genetics

1 | INTRODUCTION

Inferring the demographic history of conspecific populations is fundamental to our understanding of the impacts of past environmental and anthropogenic changes upon the distribution, abundance and

adaptations of contemporary populations (Girod, Vitalis, Leblois, & Freville, 2011; Lorenzen et al., 2011). The demographic history can be inferred from the fossil record (e.g., Newbrey & Ashworth, 2004; Parducci, Suyama, Lascoux, & Bennett, 2005). However, the collection of fossils (e.g., animal remains, pollen, environmental DNA) from

historic populations is often challenging or impossible. Alternatively, the demographic history can be inferred from contemporary genetic data (Aise, 2004; Lawton-Rauh, 2008; Marjoram & Tavaré, 2006).

Most population genetic approaches, aimed at inferring demographic history, utilize likelihoods (Griffiths & Tavaré, 1994) and typically rely upon classical population genetic or coalescent theory in combination with Markov chain Monte Carlo sampling (Beaumont & Rannala, 2004; Csillery, Blum, Gaggiotti, & Francois, 2010). Likelihood-based methods are implemented in a variety of software, such as *BATWING* (Wilson & Balding, 1998; Wilson, Weale, & Balding, 2003), *MSVAR* (Beaumont, 1999), *IM* and *IMA* (Hey & Nielsen, 2004, 2007), *LAMARC* (Kuhner, 2006), *VAREFF* (Nikolic & Chevalet, 2014) and *BEAST* (Drummond & Rambaut, 2007). The aforementioned approaches differ mainly in the assumptions they make regarding the underlying demographic model and the mutational processes generating genetic variation (Girod et al., 2011). However, the high computational overhead of likelihood-based approaches imposes severe constraints upon the amount of data and the complexity of the underlying models in the analysis. Although these computational issues might appear trivial, most current implementations of likelihood-based approaches are incapable of accommodating the large data sets emerging from the massively parallel sequencing methods (Marjoram & Tavaré, 2006; Sunnaker et al., 2013). Consequently, the above issue has spurred the development of approximate, as opposed to full-likelihood, methods capable of accommodating large data sets and complex models (Marjoram & Tavaré, 2006). One popular implementation of such approximate methods is approximate Bayesian computation (ABC, Beaumont, Zhang, & Balding, 2002).

Approximate Bayesian computation is a Monte Carlo-based approach of approximate Bayesian statistical inference, which employs summary statistics to estimate the posterior probability of a specific hypothesis, that is, a specific combination of parameter values (Beaumont et al., 2002). ABC approach was first introduced to population genetics by Beaumont et al. (2002), and has gained in popularity, in part due to the lower computational overhead enabling the analysis of larger data sets and more complex models (Aeschbacher, Beaumont, & Futschik, 2012; Beaumont et al., 2002).

A “generic” ABC-based estimation is comprised of the following steps (Csillery et al., 2010; Sunnaker et al., 2013): (i) an input parameter value is sampled at random from a prespecified prior probability distribution and used to simulate data sets; (ii) Selected summary statistics are computed from the simulated data; (iii) The summary statistic values computed from the simulated data are compared to the summary statistic values computed from the empirical data; (iv) The input parameter value is retained if the simulated summary statistics are deemed close to the summary statistics of the empirical data; and (v) Once a prespecified number of simulated data sets have been retained, a posterior probability distribution is approximated from the distribution of accepted input parameter values.

During recent years, the use of ABC-based approaches has increased in population genetics to infer demographic history (e.g., Fontaine et al., 2012; Hoffman, Grant, Forcada, & Phillips, 2011; Lombaert et al., 2011; Nadachowska-Brzyska et al., 2013; Phillips

et al., 2013; Rovito, 2010; Row et al., 2011). Several software packages implement ABC approaches, such as *DIYABC* (Cornuet et al., 2008), *ONESAMP* (Tallmon, Koyuk, Luikart, & Beaumont, 2008), *POPABC* (Lopes, Balding, & Beaumont, 2009), *ABC-SYSBIO* (Liese et al., 2010), *ABCTOOLBOX* (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010) and several *R-CRAN* packages (R Development Core Team 2013), such as “*ABC*” (Csillery, Francois, & Blum, 2012) or “*EASYABC*” (Jabot, Faure, & Dumoulin, 2013). However, the assessment of the performance of these implementations of ABC to infer demographic history has so far been relatively limited (e.g., Cornuet, Ravigne, & Estoup, 2010; Cornuet et al., 2008; Pelletier & Carstens, 2014; Robert, Cornuet, Marin, & Pillai, 2011). The choice of which ABC implementation to employ depends upon the available computational resources, the nature of the molecular data and the complexity of the estimation. *DIYABC* is one of the most commonly employed ABC-based software, has a user-friendly graphical interface and executables for all major operating systems, such as Microsoft Windows™, Linux and OS X™. *DIYABC* enables the comparison of multiple complex historical models and accommodates (in a single estimation) commonly employed molecular data types (Cornuet et al., 2008, 2014). Cornuet and coworkers conducted an evaluation of *DIYABC* regarding model selection and parameter estimation of the effective population sizes, population divergence time and admixture rates from microsatellite genotypes. Employing a single replicate per demographic model, the authors concluded that *DIYABC* performed well; that is, the “true” model was recovered at high rates and the estimates of the effective population size and admixture rates were accurate. However, the assessment also revealed some degree of bias in the estimates of temporal parameters, particularly the timing of the most recent event. In a subsequent assessment, Cornuet et al. (2010) looked at the effects of combining nuclear microsatellite genotypes, mitochondrial DNA sequences and/or nuclear autosomal DNA sequences. The assessment revealed substantial differences in model recovery rates among different molecular data types. In line with expectations, data sets including data from different molecular data types performed comparatively better (Cornuet et al., 2010). The assessments undertaken by Cornuet et al. (2008, 2010) revealed some bias in the estimation of key parameters, which was inferred by the authors to stem from differences in the degree of “informativeness” among different molecular data types.

A common use of ABC approaches (*DIYABC* in particular) is to assess past population size changes in a single isolated population. Inferences regarding past population size changes from contemporary population genetic data have been widely employed to assess the effects of past environmental changes, such as the most recent glaciations (e.g., Bisconti, Canestrelli, Colangelo, & Nascetti, 2011; González-Wevar, Saucedo, Morley, Chown, & Poulin, 2013; O’Corry-Crowe, 2008) and impacts of human exploitation (e.g., Fontaine et al., 2012; Hoffman et al., 2011). Here, we presented the results of an evaluation of *DIYABC*’s ability to recover past population demographic events as well as the accuracy and error of the targeted parameter estimates (i.e., effective population size (N_e), time of population size changes (T) and mutation rates (μ)). We aimed specifically

at population size changes in the relatively distant past, that is, during the Last Glacial Maximum (LGM) some 26,500 to 15,000 years ago (Clark et al., 2009), which has been the target of many studies. The climatic and ecological changes during and after the LGM have exerted significant and lasting effects upon the distribution and levels of genetic diversity we observe in contemporary populations across a wide range of habitats (Cahill et al., 2013; Fontaine et al., 2012; Hewitt, 1996, 2000; Hofreiter & Stewart, 2009; Nadachowska-Brzyska et al., 2013; Rovito, 2010). We simulated data representing the three most commonly employed molecular data types in DIYABC-based estimations (mitochondrial DNA sequences, microsatellite genotypes and single nucleotide polymorphism (SNP), e.g., Rovito, 2010; Lombaert et al., 2011; Fontaine et al., 2012; Lippens et al., 2017) under five different and increasingly complex models of past population size changes.

2 | METHODS

We simulated data sets of population genetic data under five different demographic models. These population genetic data sets were subsequently employed to evaluate the ability of DIYABC to recover past demographic events and estimate the parameter values of interest.

2.1 | Generation of pseudo-empirical data sets

Simulated data were generated under five different demographic models including growing, declining and stable populations. The changes in population size were modelled to have taken place during the LGM in the Northern Hemisphere (Clark et al., 2009). The five models were (CON) constant population size, (DEC) a single instantaneous decrease in population size, (INC) a single instantaneous increase in population size, (INCDEC) a single instantaneous increase followed by a single instantaneous decrease in population size and finally (DECINC) a single instantaneous decrease followed by a single instantaneous increase in population size (Figure 1, Table 1). In the case of the models DEC and INC, the change in population size occurred after the LGM (i.e., approximately 11,250 years ago), and in the case of the models INCDEC and DECINC, the changes in

population size occurred just before and after the LGM (i.e., approximately 22,500 and 11,250 years ago, respectively). We employed a generation time of 15 years, that is, in the range of generation times estimated for large, long-lived mammals, such as marine mammals and primates (e.g., Pacifici et al., 2013).

Each simulated data set was comprised of a sample of 100 individuals from one single panmictic population using DIYABC v. 2.0.4 (Cornuet et al., 2008, 2014). The simulated data sets were comprised of either (i) mitochondrial (haploid) DNA sequences of 450 nucleotides; (ii) 20 diploid microsatellite genotypes; (iii) a combination of 20 diploid microsatellite genotypes and mitochondrial DNA sequences of 450 nucleotides; or (iv) 1,000 SNP genotypes. The mutation model and rates used to simulate each molecular data type are listed in Table 2. The values are similar to those estimated for large mammals (Carapelli, Comandi, Convey, Nardi, & Frati, 2008; Dallas, 1992; Estoup, Jarne, & Cornuet, 2002; Excoffier & Yang, 1999; Hasegawa, Kishino, & Yano, 1985; Pesole, Gissi, De Chirico, & Saccone, 1999; Reyes, Gissi, Pesole, & Saccone, 1998; Santos et al., 2008; Schlötterer, Ritter, Harr, & Brem, 1998; Tamura & Nei, 1993; Weber & Wong, 1993). Nuclear genotypes were unlinked and selectively neutral.

In order to evaluate the effect of effective population size and different proportional changes in effective population size, we simulated two or three different combinations of N_e for each demographic model (two for the models INCDEC and DECINC and three for the models CON, DEC and INC, denoted as submodels A to M in Table 1). Ten replicated simulated data sets were generated for each combination of molecular data types (four combinations, i to iv above, Figure 2) and N_e (13 in total, submodels A to M above, Figure 2). In total, this amounted to 520 simulated data sets. Below, we refer to these simulated data sets as the “pseudo-empirical” data sets.

2.2 | Defining hypothesized models, prior information and summary statistics

The model selection and parameter estimations for each pseudo-empirical data set were conducted employing DIYABC v.2.0.4 (Cornuet

FIGURE 1 Schematic representation of the demographic models evaluated. The areas of the figures represent changes in population size through time. Effective population size (N_e) is represented by N_i (i.e., $N_{0(1-5)}$, N_{112} , N_{153} , N_{114} , N_{254} , N_{155} and N_{215}). Most recent events are at the top of the figure. The time, T (in number of generations), when a demographic change occurred is as T_i ($T_{T(2-3)}$, $T_{1(4-5)}$ and $T_{2(4-5)}$)

TABLE 1 Demographic models and model parameterization for the generation of the “pseudo-empirical data sets”

Parameters	Model 1: CON			Model 2: DEC			Model 3: INC			Model 4: INCDEC		Model 5: DECINC	
	A	B	C	D	E	F	G	H	I	J	K	L	M
$N_{0(1-5)}$	100	1,000	10,000	100	1,000	100	10,000	10,000	1,000	100	1,000	10,000	10,000
N_{112}				10,000	10,000	1,000							
N_{153}							100	1,000	100				
N_{114}										10,000	10,000		
N_{254}										100	1,000		
N_{155}												100	1,000
N_{215}												10,000	10,000
$T_{T(2-3)}$				750	750	750	750	750	750				
$T_{1(4-5)}$										750	750	750	750
$T_{2(4-5)}$										1,500	1,500	1,500	1,500

The demographic models refer to the five demographic models depicted in Figure 1. $N_{0(1-5)}$ denotes the most recent effective population size for all demographic models. N_{112} , N_{153} , N_{114} , N_{254} , N_{155} and N_{215} denote the large and small historical effective population sizes. Time is in number of generations assuming a generation time of 15 years; $T_{T(2-3)}$ denotes a unique event after the LGM, $T_{1(4-5)}$ and $T_{2(4-5)}$ denote an event after and before the LGM. The submodels A to M represent different combinations of N_e for each demographic model evaluated.

TABLE 2 Mutation model parameterization for the generation of the “pseudo-empirical data sets”

Autosomal microsatellite genotypes		Mitochondrial DNA sequences		SNP genotypes
Mutation model	GSM	Mutation model	HKY	No mutation model parameterization was required for SNPs. Polymorphism was assumed to result from a single mutation during the whole population gene tree and that SNPs were biallelic (Cornuet et al., 2014).
MEAN- μ_{MIC}	1.00E-04	MEAN- μ_{MIT}	5.00E-06	
INDL- μ_{MIC}	GA[1E-5, 1E-3, 2]	INDL- μ_{MIT}	GA[1E-7, 1E-5, 2]	
MEAN-P	0.22	MEAN-K1	10	
INDL-P	GA[0.01, 0.9, 2]	INDL-K1	GA[0.05, 20, 2]	
MEAN-SNI	1.00E-08	% <i>invariable sites</i>	80	
INDL-SNI	GA[1E-9, 1E-04, 2]	<i>Shape</i>	2	
		Nucleotide composition	30% A, 20% C, 20% G, 30% T	

The mutation model parameters of the autosomal microsatellite genotypes were the mutation rate (μ_{MIC}) per locus per generation, the parameter determining the shape of the gamma distribution of individual genotypes mutation rate (P) and the single insertion nucleotide rate (SNI). The mutation rate of the microsatellite genotypes was based on published mutation rate estimates for mammals, which are in the range of 10^{-3} to 10^{-5} per locus per generation (Dallas, 1992; Weber & Wong, 1993). The length and nucleotide composition of the mitochondrial DNA sequences were simulated based on commonly used values for the mitochondrial DNA control region or D-loop (Carapelli et al., 2008; Reyes et al., 1998). The mutation model parameters for the mitochondrial DNA sequences were the mutation rate (μ_{MIT}) in substitution per site per generation, the transition/transversion ratio ($K1$) parameter, the fraction of constant sites (% *invariable sites*) and the shape of the gamma distribution of mutations among sites (*Shape*). HKY is the Hasegawa–Kishino–Yano model (Hasegawa et al., 1985). GSM is the generalized stepwise-mutation model (Estoup et al., 2002). The mutation rate value of the mitochondrial DNA sequences is within the range of mutation rates found in the literature for the control region of the mitochondrial DNA in mammals (Excoffier & Yang, 1999; Pesole et al., 1999; Santos et al., 2008; Tamura & Nei, 1993). MEAN denotes mean value of the parameter, and INDL denotes the individual locus value with a gamma distribution (GA) around mean. Inside brackets are the minimum, maximum and shape of the parameter. Sex ratio for all simulated populations was 0.5.

et al., 2014) following the procedure common to most published studies based upon *DIYABC* (e.g., Fontaine et al., 2012; Lombaert et al., 2011; Louis et al., 2014). The five models, which guided the generation of the pseudo-empirical data sets (i.e., CON, DEC, INC, INCDEC and DECINC in Figure 1), were employed as candidate models during the *DIYABC* estimations. The prior distributions for the demographic and genetic parameter estimations are listed in Tables 3 and 4, respectively. Employing the same simulation software to both generate and analyse the pseudo-empirical data implies that the pseudo-empirical data sets represent the best-case scenario. In other words, empirical data sets collected from natural populations that have undergone demographic changes similar to those

simulated in our assessment, are likely subject to greater uncertainty and bias compared to the “pseudo-empirical” data employed in this study.

The summary statistics employed in ABC estimation are tabulated in Table 4 by molecular marker type. The selected summary statistics were those typically used in previous empirical studies (e.g., Fontaine et al., 2012; Lombaert et al., 2011; Louis et al., 2014).

2.3 | ABC

As the basis for the ABC estimation, we generated 10^6 simulated data sets for each demographic model and combination of parameter

FIGURE 2 Schematic representation of the pseudo-empirical data sets generated per demographic model and genetic data set combination. The first column represents the demographic model evaluated. The CON model is used as an example. The second column represents the different N_e evaluated. We referred to these combinations of N_e as submodels in the text. The third column represents the different genetic data sets evaluated: mitochondrial DNA sequences (MIT), microsatellite genotypes (MIC), a combination of mitochondrial DNA sequences and microsatellite genotypes (MITMIC) and SNP genotypes (SNP). For each combination of genetic data set per demographic model, ten simulated data sets were generated. The models CON, DEC and INC have the same schematic representation. The models INCDEC and DECINC have only two submodels. N_{01} is the effective population size

values from prior distributions as described above. The 10^4 prior parameter value combinations that yielded summary statistics closest to the values estimated from the pseudo-empirical data set were retained and the posterior probability distributions were estimated (Beaumont et al., 2002; Cornuet et al., 2008). These 10^4 retained simulations represented the default and fixed threshold value of 0.01 in *DIYABC*. The posterior probability distributions were employed for the model selection and parameter estimation. In our assessment, we treated the model selection and the parameter estimation as two separate assessments.

2.4 | Model selection and assessment of the confidence of the recovered model

The posterior probability for each demographic model was estimated from the polytomous weighted logistic regression (Beaumont, 2008) from which the point estimate and the 95% confidence interval were

computed for each candidate model as described by Cornuet et al. (2008). The most probable demographic model was deemed successfully recovered if the maximum posterior probability of the model was excluded from the 95% confidence intervals estimated for all other candidate models (Cornuet et al., 2008). The model recovery was classified as “undefined” in those cases where the above criterion failed. The model selection was conducted twice for each pseudo-empirical data set, that is, with three (CON, DEC and INC) or five (CON, DEC, INC, INCDEC and DECINC) candidate models, resulting in a total number of assessments at 1,040.

We estimated the classification error (referred to as the type I error in *DIYABC*) and the misclassification error (referred to as the type II error in *DIYABC*) from 500 pseudo-observed data sets (hereafter named pods) for each candidate demographic model (Cornuet et al., 2010). The pods (i.e., simulated data sets for which we knew the true values of the parameters) were simulated from the initial prior distributions and compared against the original 10^6 ABC simulated

TABLE 3 Demographic models and model parameterization in the ABC analysis

Parameters	Model 1: CON	Model 2: DEC	Model 3: INC	Model 4: INCDEC	Model 5: DECINC
$N_{0(1-5)}$	UN[100, 200,000]	UN[100, 200,000]	UN[100, 200,000]	UN[100, 200,000]	UN[100, 200,000]
N_{112}		UN[100, 200,000]			
N_{153}			UN[100, 200,000]		
N_{114}				UN[100, 200,000]	
N_{254}				UN[100, 200,000]	
N_{155}					UN[100, 200,000]
N_{215}					UN[100, 200,000]
$T_{T(2-3)}$		UN[100, 2,000]	UN[100, 2,000]		
$T_{1(4-5)}$				UN[100, 1,000]	UN[100, 1,000]
$T_{2(4-5)}$				UN[1,000, 2,000]	UN[1,000, 2,000]

According to models from Figure 1: N_0 , N_{112} , N_{153} , N_{114} , N_{254} , N_{155} and N_{215} most recent and historical effective population sizes. $T_{T(2-3)}$, $T_{1(4-5)}$ and $T_{2(4-5)}$, time in generations. UN: uniform distribution with a minimum and maximum parameter inside brackets.

TABLE 4 Mutation model parameterization and summary (test) statistics considered in the ABC analysis

Mutation model parameterization and summary statistics				
Microsatellite genotypes		Mitochondrial DNA sequence		SNP genotypes
Mutation model	GSM	Mutation model	HKY	No mutation model was required
MEAN- μ_{MIC}	UN[1.0 E-5, 1.0 E-3]	MEAN- μ_{MIT}	UN[1.00E-7,1.00E-5]	
INDL- μ_{MIC}	GA[1.00E-5, 1.00E-2, 2]	INDL- μ_{MIT}	GA[1.00E-7,1.00E-5,2]	
MEAN-P	UN[0.1, 0.3]	MEAN-K1	UN[0.050,20]	
INDL-P	GA[0.01, 0.9, 2]	INDL-K1	GA[0.050,20,2]	
MEAN-SNI	LU[1.00 E-9, 1.00E-4]	% invariable sites	80	
INDL-SNI	GA[1.00 E-9, 1.00 E-4, 2]	Shape	2	
Summary (test) statistics: (i) mean genetic diversity across loci (Nei, 1987); (ii) mean number of alleles across loci; (iii) mean of Garza-Williamson's (M) index across loci (Garza & Williamson, 2001); and (iv) mean allele size variance across loci		Summary (test) statistics: (i) number of distinct haplotypes; (ii) number of segregating sites; (iii) mean pairwise difference; (iv) variance of the number of pairwise differences; (v) Tajima's D statistics (Tajima, 1989); (vi) private segregating sites; (vii) mean of numbers of the rarest nucleotide at segregating site; and (viii) variance of numbers of the rarest nucleotide at segregating sites.		Summary (test) statistics: (i) proportion of zero values; (ii) mean of nonzero values; (iii) variance of nonzero values; and (iv) mean of complete distribution.

The mutation model parameters of the autosomal microsatellite genotypes were the mutation rate (μ_{MIC}) per locus per generation, the parameter determining the shape of the gamma distribution of individual loci mutation rate (P) and the single insertion nucleotide rate (SNI). The mutation model parameters for the mitochondrial DNA sequences were the mutation rate (μ_{MIT}) in substitution per site per generation, the transition/transversion ratio (K1) parameter, the fraction of constant sites (% invariable sites) and the shape of the gamma distribution of mutations among sites (Shape). HKY is the Hasegawa–Kishino–Yano model (Hasegawa et al., 1985). GSM is the generalized stepwise-mutation model (Estoup et al., 2002). MEAN denotes mean value of the parameter, and INDL denotes the individual locus value. GA: gamma distribution around mean. UN: uniform distribution. LU: log-uniform distribution. Inside brackets are the minimum and maximum in uniform distributions and minimum, maximum and shape of the parameter in gamma distribution.

data sets to calculate the posterior parameter distributions. The classification error was estimated as the proportion of pods among which the true model was incorrectly excluded. The misclassification error was estimated as the proportion of pods among which an incorrect model was recovered.

2.5 | Parameter estimation, bias and error

We estimated both unscaled (i.e., N_e , T and μ) and scaled (i.e., $N_e\mu$ and $T\mu$) parameters (Beaumont et al., 2002; Cornille et al., 2012;

Cornuet et al., 2008) for 200 pseudo-empirical data sets of the 540, that is, ten replicated data sets for five demographic models (submodels B, E, H, K and M, in Table 1) and four combinations of genetic data set. The five selected demographic models represented a single value of N_e per demographic model. In order to assess the accuracy and error of the parameter estimates, we compared the estimated values against the input parameter values used to generate the pseudo-empirical data sets (Table 1, Fig. S1a and Table S1a–e).

The accuracy and error of the parameter estimates were estimated for 20 pseudo-empirical data sets, that is, one data set for

each of the five demographic models (submodels B, E, H, K and M) and four combinations of genetic data set. The pseudo-empirical data sets were selected at random among the pseudo-empirical data sets in which the true model was correctly recovered. We estimated the accuracy as the average relative bias (ARB) and the error as the relative median of the absolute error (RMAE). We employed the same ARB and RMAE measures to assess the effect of the pseudo-empirical data upon the posterior probability distributions (Cornuet et al., 2008).

2.6 | Assessing the goodness-of-fit of the recovered demographic model

We assessed the goodness-of-fit of the recovered demographic model relative to the other four candidate demographic models for each of the 20 pseudo-empirical data sets described above (i.e., one data set for each of the five demographic models and the four kinds of genetic data sets).

The goodness-of-fit of each specific demographic model was estimated employing the model checking procedure implemented in DIYABC. The goodness-of-fit was based on the posterior distribution, which was used to estimate a posterior predictive distribution (Geisser & Eddy, 1979). The assessment was based upon 10^3 new simulated data sets per candidate model and a new set of summary statistics that served as test statistics. The posterior distributions estimated from the pseudo-empirical data were employed as prior distributions for these new simulated data sets (Gelman et al., 2014). The posterior predictive p -value was estimated for each test statistic of the pseudo-empirical data sets (Cornuet et al., 2010). If the model was a good fit, the posterior predictive p -value was expected to be close to .5 (Gelman, 2013). In order to evaluate the risk of overestimating the quality of the fit by using as test statistics the same statistics as in the model selection (Cornuet et al., 2010), we employed two sets of test statistics: the same statistics employed during the model selection, and a new, different set of statistics. During the first assessment, we employed the statistics listed in Table 4, which were the same test statistics employed as summary statistics in the model selection. The second assessment was based upon test statistics that were different from the summary statistics in the model selection. In the latter case, the odd numbered statistics in Table 4 were employed as summary statistics in the model selection and the even numbered statistics as test statistics.

3 | RESULTS

3.1 | Model choice: recovery rates

The overall ability of DIYABC to recover the correct model was low (255/520 assessments or 0.49, Figure 3). Among the five demographic models evaluated in this study (see Figure 1), three demographic models were "simple" (CON, INC and DEC) and two "complex" (INCDEC and DECINC). The correct model was recovered

at an estimated rate of 0.6 in the case of the simple models (Figure 3). In contrast, the correct model was recovered at a rate of 0.3 in the case of the complex models (Figure 3). The average classification error was substantially lower among the three simple models (between 0.44 and 0.47, Table 5) compared to the complex models (between 0.62 and 0.81, Table 5). In contrast, the average misclassification error was lower for the two complex models (between 0.07 and 0.09) compared to the simple models (between 0.15 and 0.20, Table 5). A chi-square test of goodness-of-fit was performed to assess whether the five models were recovered at equal rates. When the assessment included all possible outcomes (i.e., correct, incorrect, undefined, $n = 520$), we observed a significant difference in the recovery rates $\chi^2_{(df=8)} = 99.49$, $p < .01$, as was the case when the undefined category was excluded from the assessment ($n = 494$) $\chi^2_{(df=4)} = 92.20$, $p < .01$.

In total, 130 pseudo-empirical data sets were analysed for each of the four genetic data set combinations (see Figure 3). The recovery rate was higher for data sets comprising SNP genotypes (0.55) in comparison with mitochondrial DNA sequences (0.47), microsatellites genotypes or a combination of mitochondrial DNA sequences and microsatellite genotypes (0.47). The estimated recovery rates differed significantly among the four types of genetic data sets, $\chi^2_{(df=8)} = 17.01$, $p < .01$, when all outcomes were included in the assessment (i.e., correct, incorrect and undefined, $n = 520$). However, when we excluded undefined ($n = 26$, for a total of 494 assessments), we were unable to detect a statistically significant difference in the recovery rate among the four different types of genetic data sets ($\chi^2_{(df=4)} = 2.23$, $p = .53$).

The model recovery rates were highest for the DEC model (0.8, 0.9 and 0.7, Figure 3d–f) regardless of which combination of genetic data and demographic parameters was employed to generate the pseudo-empirical data. The lowest model recovery rates (0.2 and <0.1, Figure 3j and k) were observed for the DECINC model.

The model recovery rate among the simple models increased from 0.57 to 0.71 when the total number of candidate models was reduced from all five models to the three simplest models (Figures 3d–f and 4d–f). This increase was observed for all combinations of genetic data and parameter values (N_e and T) representing 360 pseudo-empirical data sets. The largest increase in model recovery rate, due to the reduction in the total number of candidate modes, was observed for the INC model (from 0.5 to 0.84).

When the true model was a complex model but the candidate models only contained the simple models, the most recent event of the true complex model was recovered with the highest rate. For example, the INC model when the true model was the DECINC model and the DEC model when the true model was the INCDEC model. In those cases, the recovery rates were between 0.6 and 0.9 (Figure 4j–m). For instance, among the 80 cases where the true model was the INCDEC demographic model, the DEC model was recovered as the most probable model 48 times, the CON model 30 times and the INC model once. One instance was undefined (Figure 4j and k). The outcomes were similar for the DECINC model (Figure 4l and m).

FIGURE 3 Best-supported model among five different competing demographic models in *DIYABC*. The columns represent the five demographic models evaluated. The rows correspond to the different combinations of demographic parameter values evaluated: three combinations for CON, DEC and INC models and two combinations for INCDEC and DECINC models. The simulated parameters for each demographic model (a-m) are shown on top of each figure and correspond to Table 2. The bars with a grey dot represent the expected demographic model based on the pseudo-empirical data set. The term “undefined” was employed when none of the candidate demographic models were recovered as the most probable demographic model. Each demographic model was tested for four genetic data sets: MIT, MIC, MITMIC and SNP

TABLE 5 Confidence in model selection for the different demographic models and molecular markers

Model	Classification error					Misclassification error				
	Molecular data					Molecular data				
	MIT	MIC	MITMIC	SNP	Mean	MIT	MIC	MITMIC	SNP	Mean
CON	0.653	0.557	0.547	0.104	0.466	0.199	0.256	0.191	0.166	0.203
DEC	0.493	0.511	0.441	0.324	0.442	0.211	0.18	0.173	0.148	0.178
INC	0.586	0.477	0.501	0.309	0.468	0.17	0.188	0.143	0.112	0.153
INCDEC	0.801	0.931	0.729	0.781	0.81	0.1	0.041	0.114	0.04	0.074
DECINC	0.633	0.7	0.597	0.551	0.621	0.112	0.13	0.084	0.051	0.094

Confidence in model selection was assessed employing the classification and misclassification error (referred as type I and II error in the software). Each estimate represents the mean value among three pseudo-empirical data sets selected at random per demographic model (CON, DEC, INC, INCDEC and DECINC) per genetic data set combination (MIT, MIC, MITMIC and SNP). All estimates were based on the logistic approach by Cornuet et al. (2010). Results are based on 500 simulated pseudo-observed data sets (pods) per model with parameter values drawn from the same distributions as the prior distributions given in Tables 3 and 4.

FIGURE 4 Best-supported model among three competing demographic models in DIYABC. The columns represent the five demographic models evaluated. The rows correspond to the different combinations of demographic parameters tested: three combinations for CON, DEC and INC models and two combinations for INCDEC and DECINC models. The simulated parameters for each demographic model (a–m) are shown on top of each figure and correspond to Table 2. The bars with a grey dot represent the expected model based on the pseudo-empirical data sets. The term “undefined” was employed when none of the candidate demographic models were recovered as the most probable demographic model. The demographic models j, k, l and m represent models in which the true model was not included in the comparison set. Each demographic model was tested for four genetic data sets: MIT, MIC, MITMIC and SNP

3.2 | Estimating effective population size, time and mutation parameter values

The effective population size (e.g., N_{02} , N_{03} , N_{112} and N_{153} , see Figure 5) was generally estimated with higher accuracy and lower error compared to T (e.g., T_{T2} and T_{T3} , Figure 5). The smallest N_e (recent N_e for DEC and historic N_e for INC, Figure 5) was estimated most accurately and precisely, irrespective of the nature of the genetic data.

Increasing the amount of genetic data (i.e., number of loci and/or type of genetic markers) improved the precision by reducing the credible interval of the parameter estimates (N_{02} , N_{03} , N_{112} , N_{153} , T_{T2} and T_{T3} , Figure 5, Fig. S1), but did not appear to improve the accuracy (Figure 5, Fig. S1 and Table S1a–f). The unscaled parameter values were estimated with higher error compared with the scaled parameter values (Figure 6, Fig. S1b–e and Table S1a–e).

We assessed the effect of the genetic data and demographic history upon the parameter estimates. We compared the difference

FIGURE 5 Inference of the effective population size and time parameters of pseudo-empirical data sets using DIYABC. Four genetic data set combinations are presented for two demographic models (top: population decline (DEC.E) and bottom: population increase (INC.H)) as described in Table 1. Each graphic presents the results of ten replicates per demographic model. The black circles represent the mode of the posterior probability distribution, and the grey vertical line the 95% credible interval for each parameter. The true parameter value is shown with a horizontal line. Parameter notations are the time of the demographic event ($T_{T(2-3)}$) in generations, the most recent effective population size ($N_{0(2-3)}$), the historic effective population size before the demographic decline (N_{112}) and before the demographic increase (N_{153}). See Table 3 and supplementary material S1 for all models

FIGURE 6 Marginal posterior distribution of unscaled and scaled parameters. Parameters were estimated from data generated under a demographic decline based upon data emulating mitochondrial DNA sequences. Unscaled parameters: (a) most recent effective population size N_{o2} , (b) historic effective population size N_{112} , (c) time of the population decline in generations T_{T2} , (d) mutation rate in substitutions per site per generation μ_{MIT} . Scaled parameters: (e) the product of most recent effective population size by mean mutation rate $N_{o2}\mu_{MIT}$, (f) the product of historic effective population size by mean mutation rate $N_{112}\mu_{MIT}$, (g) the product of time by mean mutation rate $T_{T2}\mu_{MIT}$. The demographic model corresponds to a population decline ($N_{o2} = 1,000$ $N_{112} = 10,000$ $T_{T2} = 750$) based on MIT (model DEC.E, Table 1). A dark grey line represents the prior distribution of the parameter values, and a light grey line the posterior distribution of the parameter values. The y-axis is the density of the distributions

between the posterior estimates obtained with and without the pseudo-empirical genetic data (Table S2a,b, Fig. S2a,b). The average ARB (bias) for estimates of N_e and T was improved from 1.49 to 0.64 (Table S2a). Excluding the CON model, the average improvement in ARB for N_e (from 2.12 to 0.74) was larger compared to the improvement observed in estimates of T (from 0.45 to 0.40). We observed a similar trend, in the case of error, where the average RMAE for estimates of N_e and T was improved from 0.34 to 0.29 (Table S2a). Excluding the CON, the average improvement in ARB was larger for N_e (from 0.37 to 0.27) compared to T (from 0.33 to 0.31).

Comparing the four different types of genetic data sets, we observed the largest improvement in ARB in analyses based upon only microsatellite genotypes (from 1.61 to 0.42) or a combination of microsatellite genotypes and mitochondrial DNA sequences (from 1.48 to 0.47). Despite much more data, the improvement in ARB was slightly less for the data sets comprising solely SNP genotypes (from 1.41 to 0.77) or mitochondrial DNA sequences (from 1.33 to 0.80, Table S2b). The improvement in RMAE was larger for data sets comprising solely SNP genotypes (from 0.35 to 0.25) or a combination of microsatellite genotypes and mitochondrial DNA sequences (from 0.30 to 0.25) compared to data sets comprising solely microsatellite genotypes (from 0.35 to 0.30) or mitochondrial DNA sequences (from 0.35 to 0.31, Table S2b).

3.3 | Assessing model fit

In general, the goodness-of-fit was higher for all test statistics for the recovered model (Table S3a–h). In cases when an incorrect model was recovered as the most probable model, the goodness-of-fit of the recovered model was usually only marginally higher relative to the goodness-of-fit of the true model. Among 100 assessments

(i.e., the goodness-of-fit of all five candidate models for each of the 20 selected pseudo-empirical data sets), 22 candidate models contained at least one test statistic estimate that was located below 0.05 or above 0.95. None of these 22 candidate models were recovered as the most probable model (Table S3a–d). We obtained comparable outcomes for both employing the same or different statistics to estimate the posterior distribution of the parameter values and the goodness-of-fit of the model, suggesting that the choice of test statistics employed to assess the goodness-of-fit might not be all crucial (Table S3a–h). However, this specific outcome might not apply to all demographic histories or data sets.

4 | DISCUSSION

We evaluated the performance of the ABC approach implemented in the popular software package *DIYABC* (Cornuet et al., 2014) both in terms of ability to recover the true demographic model and with regard to the error and accuracy of the parameter estimates. Our assessment focused on a few simple, single-population demographic models with one or two changes in population size in the relative distant past (i.e., the LGM), as the performance of *DIYABC* would likely decrease for more complex models (e.g., Briscoe & Feldman, 2011; Geman, Bienenstock, & Doursat, 1992). The graphical user interface (GUI), necessary for all analysis except the generation of simulated data sets, employed to estimate the summary statistics, is likely one of the main reasons why *DIYABC* is a popular choice. However, the mandatory use of GUI for the last steps in model selection and parameter estimation also constitutes a severe limitation for simulation-based assessments of *DIYABC*. This was why we were forced to limit the number of replicated pseudo-empirical data sets to ten for each combination of genetic data set and demographic model.

However, the total number of pseudo-empirical data sets (520) in our assessment is considerably larger to those employed in previous assessments of *DIYABC* (Cornuet et al., 2008, 2010). The earlier assessments of *DIYABC* were based upon a single replicate (either 10 autosomal microsatellite genotypes, 1,000 nucleotides of mitochondrial DNA sequences, five nuclear autosomal DNA sequences of 1,000 nucleotides each and all combinations of two and three types of markers) per demographic model (for a total of seven assessments). Here, we evaluated ten replicate data sets per combination of genetic data set and demographic model, resulting in a total of 520 data sets (Figure 2). Each pseudo-empirical data set was evaluated against five and three candidate models, representing a total of 1,040 assessments. In many instances, the differences among the aspects compared were such that ten replicates were sufficient. However, many comparisons were effectively made across multiple parameter combinations, and hence, the actual number of pseudo-empirical data sets employed in an assessment was much larger. Similar to most of such simulation-based assessments, the pseudo-empirical data were generated under assumptions that are identical, or very similar to the assumptions underlying the subsequent data analysis. In other words, the results presented here represent the best-case scenario. Consequently, it is safe to assume that the model selection and parameter value estimates obtained from empirical data sets likely will be subject to higher uncertainty and error compared to the assessment of the data analysed in this study.

In general, we found that *DIYABC* performs relatively poorly. In less than half of the assessments, *DIYABC* was able to recover the correct demographic model, and most demographic parameters were estimated with bias and with high error. Our findings are based upon sample sizes and data sets that closely resemble data sets in publications employing *DIYABC* (e.g., Besnard et al., 2014; Fontaine et al., 2012; Inoue, Monroe, Elderkin, & Berg, 2014; Kimura et al., 2014; Louis et al., 2014).

Our findings suggested that the ability to recover the correct demographic model, as well as the error and accuracy of the parameter estimates, was influenced by multiple factors, such as the number of candidate models; which parameters are targeted; the informativeness of each molecular marker type; and the specific demographic history. Below, we discuss some of these aspects in further detail and provide recommendations of how to remedy some of the potential caveats.

4.1 | Selection of the demographic models: complexity, similarity and number of candidate demographic models

4.1.1 | Model complexity

The genetic make-up of contemporary populations is a product of the interaction of multiple, different underlying processes, which potentially may lead to similar outcomes. Acknowledging this underlying complexity encourages the inclusion of multiple and complex candidate models in the assessment of the demographic history.

However, it was evident from our results that multiple changes in N_e was challenging to identify compared to demographic histories with only a single change in N_e . This general observation was consistent across all combinations of genetic data sets. Consequently, it is probably overly optimistic to expect a *DIYABC*-based analysis to capture every detail of the demographic history; a more efficient strategy might instead be to conduct an estimation that focuses on a few simple candidate demographic models that aim to capture that major difference(s) with the large effect size(s).

4.1.2 | Similarity of the candidate models

Reassuringly our assessment revealed that in those cases when an incorrect model was recovered, the model deemed most probable was usually that most similar to the true model. For example, in the case of data generated under the INC model, the models recovered most frequently (when the INC model was recovered) were those with increases in N_e as well (i.e., DECINC and INCDEC). As the similarity of the candidate models increased so did the failure to recover the true model. These trends were also evident in increased ability to recover the true model when increasing the dissimilarity among the candidate models. Similar results were reported in other ABC assessments where model selection was employed to make phylogeographic inferences (Pelletier & Carstens, 2014). In this case, the authors concluded that increasing the similarity among the candidate models decreased the probability of recovering the true model and the expected posterior probability of true model reduced. Consequently, at least in the initial analysis, it is probably most efficient to avoid the inclusion of similar candidate models.

4.1.3 | Number of candidate models

In general, our results suggested that increasing the number of candidate models influences the rate of recovering the true demographic model. As noted by Pelletier and Carstens (2014), an increase in the number of candidate models typically increases the overall similarity among the candidate models, which, in turn, reduces the expected posterior probability of the true model. The obvious consequence is to recommend a reduction in the number of candidate models to the bare minimum. Although our study showed that reducing the number of candidate demographic models improved the probability of recovering the correct demographic model, such a reduction in the number of candidate demographic models harbours the inherent risk of excluding the true model from the analysis altogether.

Previous authors have proposed to circumvent this potential caveat by conducting the model selection in a hierarchical manner, making an initial model selection among relatively simple models and subsequently add complexity to a second set of the models deemed most probable (Chan, Schanzenbach, & Hickerson, 2014; Turner & Van Zandt, 2014). Although we did not evaluate such a hierarchical approach, we found that limiting the number of candidate models to the least similar models appeared to increase the

ability to recover the correct model, a finding that is consistent with the suggestion of conducting the model selection in a hierarchical manner. However, our simulations also revealed that models very similar to the true model are recovered sometimes, as the true model. Ideally, the posterior probabilities of similar candidate models that are close to the true model should yield similar posterior probabilities and hence result in an undefined outcome. However, this may not always be the case according to our results. Conducting the model selection in a hierarchical manner might also come with the risk of recovering a model, which is not, globally, the most probable model akin to some of the criticism directed towards nested clade analysis. Accordingly, we suggest that once a model has been recovered using or not a hierarchical approach, the assessment should be repeated with a different range of candidate models, including the selected model.

4.2 | Estimates of N_e and T

In general, we observed that estimates of T were more biased and subject to higher error than estimates of N_e . Biased and imprecise estimates of T appear to be a general feature of genetic estimates of T and hence not necessarily due to the ABC approach (e.g., Girod et al., 2011; Nikolic & Chevalet, 2014). Cornuet et al. (2008) suggested that the comparatively poor performance, in terms of estimating T , is caused by insufficient data. In other words, increasing the sequence lengths and/or the number of loci might improve the accuracy and error of T (e.g., Abdo, Crandall, & Joyce, 2004; Hoban, Gaggiotti, & Bertorelle, 2013). However, we did not observe a substantial improvement in data sets comprising 1,000 SNP genotypes or a combination of mitochondrial DNA sequences and microsatellite genotypes. Furthermore, while increasing the amount of genetic data might result in a higher precision, more data may not necessarily improve the accuracy (e.g., Peery et al., 2013).

4.3 | Unscaled versus scaled parameter estimation

Most studies aimed at elucidating the demographic history of natural populations are interested in obtaining estimates of the unscaled parameters, that is, N_e and T . However, the underlying coalescent-based simulations are based upon scaled parameter values (i.e., the product of N_e and μ or N_e and generation time) and converting the scaled estimates to their unscaled values comes at a cost. Our simulations did indeed confirm that unscaled parameters are estimated with a higher error and bias compared to scaled parameters. Additional uncertainty is introduced during the conversion of scaled parameter estimates to their unscaled components as their conversion requires estimates of additional nuisance parameters, such as μ or generation time, which are also subject to uncertainty which is often ignored (Girod et al., 2011). Bertorelle, Benazzo, and Mona (2010) suggested to employ scaled parameter estimates to the largest possible extent. Most aspects of interest to many studies may be informed from the relative differences in scaled parameter estimates (e.g., Fontaine et al., 2012).

4.4 | Molecular marker types and number of loci

Our findings suggested that the error of the parameter estimates and model selection was improved either with the increasing amounts of data or by including multiple genetic data types. For example, data sets comprised both microsatellite genotypes and mitochondrial DNA sequences or increasing the data to many hundreds of SNP genotypes, as opposed to relying only on mitochondrial DNA sequences or genotypes from a few microsatellite loci. While it may be that additional loci or longer DNA sequences will improve precision and bias, increasing the amount of data did not always result in a corresponding improvement in accuracy or precision. For instance, in the case of the DEC model (Figure 5), employing genotypes from 1,000 SNP genotypes greatly improved the error of N_{e2} , N_{112} and T_{T2} in comparison with data sets comprising only mitochondrial DNA sequences or genotypes from 20 microsatellite loci. However, the accuracy of same estimates did not improve in a similar manner. The mode of the posterior probability distributions of N_{e2} , N_{112} and T_{T2} based upon the 1,000 SNP data sets was upward bias in comparison with the posterior probability distributions obtained from data sets comprising solely mitochondrial DNA sequences or microsatellite genotypes.

This trade-off between error and bias was previously noted by Beaumont et al. (2002) who showed that an increase in the tolerance in the ABC estimation reduces error as the sample size from which the posterior distribution is estimated increases. However, the larger sample sizes may increase the bias due to uncorrected departures from additivity and linearity (Beaumont et al., 2002). Increasing data will also add variance and hence diminishing the return in terms of improving the error and bias (Robinson, Bunnefeld, Hearn, Stone, & Hickerson, 2014; Shafer, Gattepaille, Stewart, & Wolf, 2015). For example, Shafer et al. (2015) reported that 1,000 SNP genotypes performed equally well to 50,000 SNP genotypes in terms of parameter estimates such as time. An additional, unaccounted source of error, which increases with DNA sequence length, is recombination. Unfortunately, most implementations, including ABC-based, in population genetics ignore recombination. Simulations, as in this study, can be employed as an effective manner to identify the optimal balance between accuracy and error for each specific case.

During our simulations, we noted some consistently incorrect outcomes in some assessments based upon genetic data sets comprising SNP loci (e.g., submodels D, F and K), which suggested a potential computational bug in the simulations conducted by *DIYABC*.

The ability of all estimation approaches depends crucially upon the information content in the data that serves as the basis of estimation and, in the case of ABC-based approaches, the suitability of the specific summary statistics. Summary statistics differ in terms of the population genetic aspects they capture, the nature of the genetic data (e.g., nucleotide sequence or microsatellite loci) and the true population history. The primary aim of this study was not to evaluate the choice of individual summary statistics. Instead, we employed a suite of summary statistics that are common to past studies based on *DIYABC*. The very nature (and advantage in a ABC

context) of summary statistics is to reduce the overall information content, which in turn introduces additional bias, particularly in complex models (Aeschbacher et al., 2012; Robert et al., 2011; Sunnaker et al., 2013). Given this fundamental feature of ABC-based estimations may make it seem appealing to increase the number of summary statistics, thereby presumably improving the precision and bias of the parameter estimates. However, we observed bias across all four types of generic data sets, suggesting that the bias is not in order to improve the parameter estimates. However, we observed bias across all four kinds of genetic data sets, suggesting that the bias is not necessarily related to a particular set of summary statistics if one can assume that the same summary statistic results in different bias for different molecular data types. Additionally, the “curse of dimensionality” (Beaumont et al., 2002) also limits the benefits of increasing the number of summary statistics. Consequently, aiming for a few “optimal” summary statistics is likely a better strategy.

5 | CONCLUSIONS AND SUGGESTED GUIDELINES

Genetic diversity in natural populations is a product of multiple interacting evolutionary processes, which makes it a challenging exercise to discern among those processes and to obtain estimates of unbiased parameters with low error. However, the development of specific parameterized hypotheses as demographic models will aid a study to focus on the key parameters.

Our simulation-based evaluation of *DIYABC* demonstrated a low ability to recover the true demographic model and that some parameters were estimated with considerable error and bias. However, an ABC-based approach, such as that implemented in *DIYABC*, is an effective manner by which to obtain a *general* (as opposed to very specific) insight into the past of contemporary populations.

Our results suggested *not* to focus on capturing every detail and *not* to aim for an as complex and realistic model as possible. Instead, our findings suggest to focus on simple contrasting models that are likely to capture the key demographic events of large effect sizes. Our assessment also suggested that it might be counterproductive to include very similar candidate models in the estimation, which might increase the probability of recovering an incorrect model. Similarly, the number of candidate models in itself proved important in terms of recovering the true model. Accordingly, it is recommendable to reduce the number of candidate models (e.g., less than five candidate demographic models). Additional models could be included later in the assessment by conducting the model selection, for example, in a hierarchical manner. Finally, it is advisable to evaluate different combinations of candidate models in order to assure that the same candidate model is recovered consistently (akin to convergence in other Bayesian-based approaches).

The apparent difficulty in obtaining accurate and precise estimates of some parameter values, such as T , may be circumvented, to some extent, by focusing on scaled parameters or the confidence

interval of the estimate rather than a point estimate. Although most studies aim to estimate unscaled parameters, scaled parameters were estimated with comparatively less bias and lower error. Accordingly, it might prove beneficial to employ scaled parameter estimates when possible.

The development of massively parallel sequencing technologies has enabled the generation of very large data sets in most species. Although an increase in the amounts of genetic data seems a logical next step to improve the estimation, our simulations showed that “more is not always better.” Increasing the number of loci improved the error in the parameter estimates, but not necessarily the accuracy. The optimal number of samples and data is case-specific and depending upon aspects, such as the objectives and the population history. Accordingly, applying generic rules may prove somewhat counterproductive. A reasonably and readily available approach is to employ a simulation framework, as in this study, to estimate the optimal amount and kind of data given the logistic and monetary constraints in each particular case (e.g., Hoban, Bertorelle, & Gaggiotti, 2012).

The results reported here provide a general insight into key aspects of ABC-based approaches in terms of model recovery rates and the bias and precision of the parameter estimates. Our assessment showed that inferences drawn from ABC model comparisons may require additional simulations in order to assess whether the results are robust and likely to recover the true demographic history and parameter estimates. Shafer et al. (2015) recommended to apply both ABC and likelihood approaches. However, consistent results between two different analytical approaches, which fundamentally utilize the same aspects of the data, should not be considered as two independent tests. We considered that assessing the “sensitivity” of ABC-based (and likelihood) approaches in each specific case using simulations harbours substantial potential to gain insights into the specific estimation and identify which aspects are likely captured reliably. Unfortunately, the study of natural populations comes with limited prior knowledge and hence a considerable degree of uncertainty. Hence, it is most efficient to focus upon a few demographic events with large effects sizes, as it is almost impossible, at this time, to represent and resolve the enormous complexity of nature in a single model with contemporary genetic data.

ACKNOWLEDGEMENTS

We thank the members of the MarECon group for their helpful comments and discussions throughout this work, in particular Dr. Jungkoo Kang, Dr. Anna Kopps, Tom Oosting, Xênia Lopes and Vania Rivera. We are also indebted to Lionel Morgado and Bob Dröge for their support with the execution of scripts and jobs on the University of Groningen High Performance Computing cluster Peregrine. The University of Groningen provided the funding for Andrea Cabrera's PhD. We would also like to thank Brian Lavin, Hielko van der Hoorn, the five anonymous reviewers and the editor, Dr. Sean Schoville, for their constructive criticisms, which greatly improve our manuscript.

AUTHOR CONTRIBUTIONS

A.A.C. and P.J.P. conceived and designed the study. A.A.C. conducted the simulations, bioinformatics and analysed the data. A.A.C. and P.J.P. wrote the manuscript.

DATA ACCESSIBILITY

All relevant information to reproduce this study is included in this manuscript and supplementary material.

REFERENCES

- Abdo, Z., Crandall, K. A., & Joyce, P. (2004). Evaluating the performance of likelihood methods for detecting population structure and migration. *Molecular Ecology*, 13, 837–851.
- Aeschbacher, S., Beaumont, M. A., & Futschik, A. (2012). A novel approach for choosing summary statistics in approximate Bayesian computation. *Genetics*, 192, 1027–1047.
- Avise, J. C. (2004). *Molecular markers, natural history, and evolution*. Sunderland, MA: Sinauer Associates, Incorporated.
- Beaumont, M. A. (1999). Detecting population expansion and decline using microsatellites. *Genetics*, 153, 2013–2029.
- Beaumont, M. A. (2008). Joint determination of topology, divergence time, and immigration in population trees. In S. Matsumura, P. Forster, & C. Renfrew (Eds.), *Simulation, genetics and human prehistory* (pp. 135–154). Cambridge, UK: McDonald Institute Monographs.
- Beaumont, M. A., & Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics*, 5, 251–261.
- Beaumont, M. A., Zhang, W. Y., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162, 2025–2035.
- Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, 19, 2609–2625.
- Besnard, G., Dupuy, J., Larter, M., Cuneo, P., Cooke, D., & Chikhi, L. (2014). History of the invasive African olive tree in Australia and Hawaii: Evidence for sequential bottlenecks and hybridization with the Mediterranean olive. *Evolutionary Applications*, 7, 195–211.
- Bisconti, R., Canestrelli, D., Colangelo, P., & Nascetti, G. (2011). Multiple lines of evidence for demographic and range expansion of a temperate species (*Hyla sarda*) during the last glaciation. *Molecular Ecology*, 20, 5313–5327.
- Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118, 2–16.
- Cahill, J. A., Green, R. E., Fulton, T. L., Stiller, M., Jay, F., Ovsyanikov, N., ... Shapiro, B. (2013). Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *Plos Genetics*, 9, e1003345.
- Carapelli, A., Comandi, S., Convey, P., Nardi, F., & Frati, F. (2008). The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genomics*, 9, 315.
- Chan, Y. L., Schanzenbach, D., & Hickerson, M. J. (2014). Detecting concerted demographic response across community assemblages using hierarchical Approximate Bayesian Computation. *Molecular Biology and Evolution*, 31, 2501–2515.
- Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... McCabe, A. M. (2009). The last glacial maximum. *Science*, 325, 710–714.
- Cornille, A., Gladieux, P., Smulders, M. J. M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., ... Giraud, T. (2012). New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *Plos Genetics*, 8, e1002703.
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187–1189.
- Cornuet, J. M., Ravigne, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, 11, 401.
- Cornuet, J. M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J. M., Balding, D. J., ... Estoup, A. (2008). Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*, 24, 2713–2719.
- Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25, 410–418.
- Csillery, K., Francois, O., & Blum, M. G. B. (2012). abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479.
- Dallas, J. F. (1992). Estimation of microsatellite mutation-rates in recombinant inbred strains of mouse. *Mammalian Genome*, 3, 452–456.
- Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214.
- 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, 1591–1604.
- Excoffier, L., & Yang, Z. H. (1999). Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Molecular Biology and Evolution*, 16, 1357–1368.
- Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Öztürk, B., Öztürk, A., & Austerlitz, F. (2012). History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. *Proceedings of the National Academy of Sciences of the United States of America*, 109, E2569–E2576.
- Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, 10, 305–318.
- Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.
- Gelman, A. (2013). Two simple examples for understanding posterior p-values whose distributions are far from uniform. *Electronic Journal of Statistics*, 7, 2595–2602.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vejtaro, A., & Rubin, D. B. (2014). *Bayesian data analysis*. Florida: Taylor & Francis.
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1–58.
- Girod, C., Vitalis, R., Leblois, R., & Freville, H. (2011). Inferring population decline and expansion from microsatellite data: A simulation-based evaluation of the Msvar method. *Genetics*, 188, 165–179.
- González-Wevar, C. A., Saucedo, T., Morley, S. A., Chown, S. L., & Poulin, E. (2013). Extinction and recolonization of maritime Antarctica in the limpet *Nacella concinna* (Strebel, 1908) during the last glacial cycle: Toward a model of Quaternary biogeography in shallow Antarctic invertebrates. *Molecular Ecology*, 22, 5221–5236.
- Griffiths, R. C., & Tavaré, S. (1994). Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 344, 403–410.
- Hasegawa, M., Kishino, H., & Yano, T. A. (1985). Dating of the human ape splitting by a molecular clock of mitochondrial-DNA. *Journal of Molecular Evolution*, 22, 160–174.
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247–276.

- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913.
- Hey, J., & Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167, 747–760.
- Hey, J., & Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 2785–2790.
- Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: Tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13, 110–122.
- Hoban, S. M., Gaggiotti, O. E., & Bertorelle, G. (2013). The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: A simulation-based study. *Molecular Ecology*, 22, 3444–3450.
- Hoffman, J. I., Grant, S. M., Forcada, J., & Phillips, C. D. (2011). Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology*, 20, 3989–4008.
- Hofreiter, M., & Stewart, J. (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current Biology*, 19, R584–R594.
- Inoue, K., Monroe, E. M., Elderkin, C. L., & Berg, D. J. (2014). Phylogeographic and population genetic analyses reveal Pleistocene isolation followed by high gene flow in a wide ranging, but endangered, freshwater mussel. *Heredity*, 112, 282–290.
- Jabot, F., Faure, T., & Dumoulin, N. (2013). EasyABC: Performing efficient approximate Bayesian computation sampling schemes using R. *Methods in Ecology and Evolution*, 4, 684–687.
- Kimura, M. K., Uchiyama, K., Nakao, K., Moriguchi, Y., San Jose-Maldia, L., & Tsumura, Y. (2014). Evidence for cryptic northern refugia in the last glacial period in *Cryptomeria japonica*. *Annals of Botany*, 114, 1687–1700.
- Kuhner, M. K. (2006). LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, 22, 768–770.
- Lawton-Rauh, A. (2008). Demographic processes shaping genetic variation. *Current Opinion in Plant Biology*, 11, 103–109.
- Liepe, J., Barnes, C., Cule, E., Erguler, K., Kirk, P., Toni, T., & Stumpf, M. P. H. (2010). ABC-SysBio-approximate Bayesian computation in Python with GPU support. *Bioinformatics*, 26, 1797–1799.
- Lippens, C., Estoup, A., Hima, M. K., Loiseau, A., Tatar, C., Dalecky, A., ... Brouat, C. (2017). Genetic structure and invasion history of the house mouse (*Mus musculus domesticus*) in Senegal, West Africa: A legacy of colonial and contemporary times. *Heredity*, 1–12. <https://doi.org/10.1038/hdy.2017.18>
- Lombaert, E., Guillemaud, T., Thomas, C. E., Handley, L. J. L., Li, J., Wang, S., ... Estoup, A. (2011). Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: Case study of the invasive ladybird *Harmonia axyridis*. *Molecular Ecology*, 20, 4654–4670.
- Lopes, J. S., Balding, D., & Beaumont, M. A. (2009). PopABC: A program to infer historical demographic parameters. *Bioinformatics*, 25, 2747–2749.
- Lorenzen, E. D., Nogues-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., ... Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479, 359–364.
- Louis, M., Fontaine, M. C., Spitz, J., Schlund, E., Dabin, W., Deaville, R., ... Simon-Bouhet, B. (2014). Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20141558.
- Marjoram, P., & Tavare, S. (2006). Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics*, 7, 759–770.
- Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L., & Ellegren, H. (2013). Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome resequencing data. *Plos Genetics*, 9, e1003942.
- Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.
- Newbrey, M. G., & Ashworth, A. C. (2004). A fossil record of colonization and response of lacustrine fish populations to climate change. *Canadian Journal of Fisheries and Aquatic Sciences*, 61, 1807–1816.
- Nikolic, N., & Chevalet, C. (2014). Detecting past changes of effective population size. *Evolutionary Applications*, 7, 663–681.
- O'Corry-Crowe, G. (2008). Climate change and the molecular ecology of Arctic marine mammals. *Ecological Applications*, 18, S56–S76.
- Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Marasini, G. G., ... Rondinini, C. (2013). Generation length for mammals. *Nature Conservation*, 5, 89–94.
- Parducci, L., Suyama, Y., Lascoux, M., & Bennett, K. D. (2005). Ancient DNA from pollen: A genetic record of population history in Scots pine. *Molecular Ecology*, 14, 2873–2882.
- Peery, M. Z., Reid, B. N., Kirby, R., Stoelting, R., Doucet-Ber, E., Robinson, S., ... Palsbll, P. J. (2013). More precisely biased: Increasing the number of markers is not a silver bullet in genetic bottleneck testing. *Molecular Ecology*, 22, 3451–3457.
- Pelletier, T. A., & Carstens, B. C. (2014). Model choice for phylogeographic inference using a large set of models. *Molecular Ecology*, 23, 3028–3043.
- Pesole, G., Gissi, C., De Chirico, A., & Saccone, C. (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, 48, 427–434.
- Phillips, C. D., Hoffman, J. I., George, J. C., Suydam, R. S., Huebinger, R. M., Patton, J. C., & Bickham, J. W. (2013). Molecular insights into the historic demography of bowhead whales: Understanding the evolutionary basis of contemporary management practices. *Ecology and Evolution*, 3, 18–37.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reyes, A., Gissi, C., Pesole, G., & Saccone, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Molecular Biology and Evolution*, 15, 957–966.
- Robert, C. P., Cornuet, J. M., Marin, J. M., & Pillai, N. S. (2011). Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15112–15117.
- Robinson, J. D., Bunnefeld, L., Hearn, J., Stone, G. N., & Hickerson, M. J. (2014). ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology Notes*, 23, 4458–4471.
- Rovito, S. M. (2010). Lineage divergence and speciation in the Web-toed Salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada, California. *Molecular Ecology*, 19, 4554–4571.
- Row, J. R., Brooks, R. J., MacKinnon, C. A., Lawson, A., Crother, B. I., White, M., & Loughheed, S. C. (2011). Approximate Bayesian computation reveals the factors that influence genetic diversity and population structure of foxsnakes. *Journal of Evolutionary Biology*, 24, 2364–2377.
- Santos, C., Montiel, R., Arruda, A., Alvarez, L., Aluja, M. P., & Lima, M. (2008). Mutation patterns of mtDNA: Empirical inferences for the coding region. *BMC Evolutionary Biology*, 8, 167.
- Schltterer, C., Ritter, R., Harr, B., & Brem, G. (1998). High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides

- evidence for allele-specific mutation rates. *Molecular Biology and Evolution*, 15, 1269–1274.
- Shafer, A. B. A., Gattepaille, L. M., Stewart, R. E. A., & Wolf, J. B. W. (2015). Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: In silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology*, 24, 328–345.
- Sunnaker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology*, 9, e1002803.
- Tajima, F. (1989). Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.
- Tallmon, D. A., Koyuk, A., Luikart, G., & Beaumont, M. A. (2008). ONE-SAMP: A program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, 8, 299–301.
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10, 512–526.
- Turner, B. M., & Van Zandt, T. (2014). Hierarchical approximate Bayesian computation. *Psychometrika*, 79, 185–209.
- Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, 2, 1123–1128.
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11, 116.
- Wilson, I. J., & Balding, D. J. (1998). Genealogical inference from microsatellite data. *Genetics*, 150, 499–510.
- Wilson, I. J., Weale, M. E., & Balding, D. J. (2003). Inferences from DNA data: Population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society Series a-Statistics in Society*, 166, 155–188.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Cabrera AA, Palsbøll PJ. Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIYABC. *Mol Ecol Resour.* 2017;17:e94–e110. <https://doi.org/10.1111/1755-0998.12696>