



Perspective

On the feasibility of estimating contemporary effective population size (N_e) for genetic conservation and monitoring of forest trees

Luis Santos-del-Blanco^{a,*}, Sanna Olsson^a, Katharina B. Budde^b, Delphine Grivet^a,
Santiago C. González-Martínez^c, Ricardo Alía^a, Juan J. Robledo-Arnuncio^{a,*}

^a Institute of Forest Sciences (ICIFOR-INIA), CSIC, Ctra. De la Coruña km 7.5, 28040 Madrid, Spain

^b Faculty of Forest Sciences and Forest Ecology, Forest Genetics and Forest Tree Breeding, University of Göttingen, Büsgenweg 2, 37077, Göttingen, Germany

^c INRAE, Univ. Bordeaux, BIOGECO, F-33610 Cestas, France



ARTICLE INFO

Keywords:

Effective size
Linkage disequilibrium
Genetic drift
Isolation by distance
Estimation bias
Maritime pine

ABSTRACT

Estimates of contemporary effective population size (N_e) can provide valuable information for genetic conservation and monitoring, pinpointing populations at higher risk of genetic erosion, decreased fitness, maladaptation and, ultimately, demographic decline. There are however potential limitations in the application of commonly employed genetic estimators of contemporary N_e to widespread forest tree populations. Genetic isolation by distance within populations, small and spatially restricted samples, among-population gene flow and overlapping generations are factors that can potentially affect the accuracy of marker-based estimates of contemporary N_e , depending on the demographic scenario. In particular, we illustrate the uncertainty faced by forest researchers and managers when interpreting contemporary N_e estimates obtained for continuously distributed tree populations with large census size N . To that end, we use previously published genotypic data of 21 *Pinus pinaster* populations, including distinct sampling schemes, together with a widely used method based on linkage disequilibrium patterns observed in a single (one-time) population sample. We hypothesize that spatially restricted sampling might be the main putative factor behind the apparently low N_e/N estimates obtained for the large and continuously-distributed populations studied here. Because of its statistically-inherent assessment difficulty, we call for caution when interpreting marker-based estimates of contemporary N_e for monitoring widely-distributed tree populations or small genetic conservation units embedded within large continuous tree populations.

1. Introduction

Genetic monitoring, defined as the quantification of temporal changes in population genetic or demographic parameters using molecular markers (Schwartz et al., 2007), is considered a key tool in biological conservation, especially in the conservation of within-species genetic diversity (Hoban et al., 2020). The rationale is that, in order to guide conservation management decisions, genetic diversity or its determinants have first to be assessed, and then continuously monitored. This way, early actions can be taken in case of population genetic decline signals, even before they produce detectable phenotypic or demographic effects. Contributions from marine and riverine conservation biology have greatly contributed to genetic monitoring theory and practice during the last years, including software solutions for sampling design and data analysis, as well as real implementation reports (Ovenden

et al., 2006, 2016; Portnoy et al., 2009; Hare et al., 2011; Brauer et al., 2016; Waples, 2016; Pita et al., 2017; Yates et al., 2017; Bernos et al., 2018; Marandel et al., 2019; Blower et al., 2019). The interest in genetic monitoring of marine species has been largely driven by the need to assess the consequences of harvesting and management, particularly restocking, in order to ensure the long-term conservation of species (Ovenden et al., 2016). Similar management concerns apply to the case of European forest trees, for which genetic monitoring of in situ dynamic genetic conservation units (GCUs) is recognized as a valuable tool for both conservation of forest genetic resources and sustainable forest management (Aravanopoulos et al., 2015). These units represent the core of the pan-European strategy for genetic conservation of forest trees, being designated forest stands where natural evolutionary processes are to be maintained in order to allow the long-term conservation of populations and species (Koskela et al., 2013; de Vries et al., 2015;

* Corresponding authors.

E-mail addresses: santos.luis@inia.csic.es (L. Santos-del-Blanco), juan jose.robledo@csic.es (J.J. Robledo-Arnuncio).

<https://doi.org/10.1016/j.biocon.2022.109704>

Received 11 March 2022; Received in revised form 11 August 2022; Accepted 13 August 2022

Available online 19 August 2022

0006-3207/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

EUFORGEN, 2021).

Several genetic parameters have been proposed for genetic monitoring of GCUs of trees and other plant and animal species, such as allelic richness and diversity, heterozygosity, effective population size, inbreeding coefficients, outcrossing rates, and frequency of putatively adaptive or deleterious alleles (Schwartz et al., 2007; Hansen et al., 2012; Aravanopoulos et al., 2015; Fussi et al., 2016). Among them, effective population size (N_e) appears as a particularly relevant synthetic parameter, being defined as the size of an idealized Wright-Fisher population (a random-mating, constant-size population with non-overlapping generations; Fisher, 1930; Wright, 1931) that shows the same value of some specified genetic property as the real population (Crow and Kimura, 1970). Each considered genetic property corresponds to a different N_e concept, such as the inbreeding effective size, variance (of allelic frequencies) effective size, eigenvalue effective size, coalescent effective size, linkage disequilibrium effective size or additive variance effective size (see recent reviews in Wang et al., 2016; Ryman et al., 2019). These effective sizes are the same in the simplest case of an isolated population of constant size, but they can diverge in more realistic demographic settings (Ryman et al., 2019). In general, populations with smaller N_e tend to suffer more inbreeding and larger drift-driven stochastic allele frequency changes, which may result in inbreeding depression, genetic diversity loss, and less efficient natural selection (Frankham, 1995). If accurate, N_e estimates might thus provide very valuable information for conservation management, pinpointing populations at higher risk of genetic erosion, decreased fitness, maladaptation and, ultimately, demographic decline (Frankham, 1996; Hoban et al., 2020).

N_e can be estimated using genetic marker data or based on pedigrees or relevant demographic and reproductive parameters (Wang et al., 2016). Both the choice of N_e estimation method and the sampling design determine the temporal and spatial scale of N_e , which may be adjusted depending on management goals (Hare et al., 2011), and should be clearly stated to facilitate interpretation of estimated values and subsequent decision making. Historical N_e estimators pertain to studies interested in ancient-past demographic and population genetic processes, typically before recent human intervention, while contemporary N_e estimators reflect recent or ongoing demographic and reproductive processes, which may eventually trigger future population genetic and demographic changes (Hare et al., 2011; Nadachowska-Brzyska et al., 2022). Contemporary N_e estimators are therefore especially relevant for real-time genetic monitoring and, among them, those based on genetic marker data have become the most widely used for natural populations (Wang, 2005). Marker-based estimators of contemporary N_e are either based on a single (e.g. the linkage-disequilibrium or LD method; Hill, 1981) or several temporally spaced (e.g. Nei and Tajima, 1981) contemporary samples. Temporal methods are based on allele frequency changes observed among population samples collected at two or more points in time, while the LD method is based on the stochastic association between alleles at different loci in a single (one-time) population sample, caused by drift at a rate inversely proportional to N_e (Hill, 1981). The LD approach has become a genetic method of choice to estimate contemporary N_e in applied conservation biology, due to its sampling ease relative to temporal methods (especially for long-living species such as trees) and its estimation accuracy relative to other one-sample approaches, such as those based on heterozygote excess or sibship frequency (Gilbert and Whitlock, 2015; Wang, 2016; Waples, 2021).

There is an ongoing debate on whether the implementation of N_e -based indicators is generally feasible for monitoring forest trees, with opposing views on the practical limitations associated with the challenge of defining populations in widely and continuously distributed tree species with long-distance dispersal, and of setting generally relevant management thresholds for N_e (Fady and Bozzano, 2021; Hoban et al., 2021). A working group of the European Forest Genetic Resources Programme (EUFORGEN, www.euforgen.org) did recently propose to

use marker-based estimators of N_e as verifiers for genetic monitoring of GCUs of European forest trees (Aravanopoulos et al., 2015). More recently, another EUFORGEN working group did not include N_e as an indicator for assessing the conservation status of European forest genetic resources at the national level (Lefèvre et al., 2020). Our goal here is to participate in this important debate by elaborating on potential limitations in the application of commonly employed contemporary marker-based N_e estimators to widespread forest tree populations. These potential limitations have been long acknowledged and quantified in theoretical studies and animal (especially fish) conservation management literature. However, we believe that a cautionary note is necessary to rise comparable awareness in applied forest conservation genetics, especially because some of the demographic and life-history characteristics that have been associated with potentially inaccurate estimation of N_e in fish populations are shared to some extent by forest tree species.

2. Demographic and life-history characteristics of forest trees potentially complicating marker-based estimation of contemporary N_e

First, both marine fish and tree species are often widely distributed, forming a network of local populations connected by gene flow (a metapopulation), especially in the case of wind-pollinated, wind-dispersed trees or of marine species with larval dispersal (Kinlan et al., 2005; Kremer et al., 2012). As a result, the definition and estimation of N_e over a relevant spatial scale becomes complicated. In particular, genetic monitoring targets such as tree GCUs typically consist of a single local population, or rather a delimited stand within a local population. Even if adequate random sampling is conducted over the entire local target area, the corresponding contemporary N_e estimates obtained with commonly employed methods (which assume population isolation) may be biased, unless incoming gene flow is small (<5–10 %; Waples and England, 2011). Specifically, the bias expected from violating the isolation assumption is positive in the case of the LD method (which will tend to estimate the contemporary N_e of the entire metapopulation for increasingly larger gene flow rates) and negative for the temporal method (Waples and England, 2011). The expected biases can increase in non-equilibrium metapopulations, if for instance gene flow occurs episodically or if it originates from strongly genetically divergent sources (Waples and England, 2011), e.g. a recently introduced allochthonous population. More worryingly, although contemporary N_e estimators tend to reflect the actual rates of inbreeding and loss of additive genetic variation (most relevant for conservation) of isolated populations, they are expected to underestimate them consistently for populations under gene flow (Ryman et al., 2019).

Second, neither marine fish nor tree distributions conform well to the random-mating population scenario assumed by common contemporary N_e estimation methods; instead, each local population is typically continuously distributed across a large area, within which genetic isolation by distance (IBD) builds up from the balance between local genetic drift and spatially restricted mating and dispersal. Such balance can be characterized by the neighbourhood size (NS), the inverse of the probability of identity by descent of two uniting gametes (Wright, 1943), which depends on effective population density and both the scale and shape of historical effective dispersal (Rousset, 2000). NS determines the decrease in the probability of gene identity with spatial distance within a continuous population (Rousset, 2000), an internal spatial genetic structure property that, unlike N_e , does not provide direct information on the rate of inbreeding or loss of genetic variation in the population as a whole (Nunney, 2016). Under IBD, contemporary N_e estimates obtained in a continuously distributed local population using the LD or other commonly employed methods are not expected to reflect its true N_e , but rather NS if sampling scale is small relative to the unknown effective dispersal scale, or a quantity between NS and N_e otherwise (Neel et al., 2013; Nunney, 2016). N_e underestimation induced by IBD has been suggested as a potential factor behind the

exceptionally low N_e/N values ($<10^{-3}$) estimated for widely distributed marine fish populations with large census size N (Neel et al., 2013; Nunney, 2016). The fact that the scale and shape of historical effective dispersal is generally unknown, and may substantially depart from those of contemporary dispersal in heterogeneous environments (Robledo-Arnuncio and Rousset, 2010), further complicates the interpretation of contemporary N_e estimates obtained in continuously distributed populations.

Third, as is the case for many marine organisms, trees are iteroparous and grow in age-structured populations, which violates the assumption of discrete generations on which most N_e estimators are based. The expectation of one-sample contemporary N_e estimates obtained for age-structured populations of iteroparous species is complex and not yet fully resolved, but it has been found to depend among other factors on the distribution of individual samples across age classes and on life-history traits such as adult longevity and age at maturity (Waples et al., 2014). Regarding sampling distribution, estimates based on random adult samples have been shown to produce estimates 25–30 % lower than the true N_e in simulated age-structured populations (Waples et al., 2014). Estimates based on single-cohort offspring samples, on the other hand, will be influenced by both the effective number of breeders that gave rise to the sampled cohort (N_b) and the per-generation effective population size (N_e), being thus unreliable estimators of the latter unless adjusted with independent (and accurate) information on life-history traits and vital rates (Waples et al., 2014). Note that N_b is not necessarily lower than N_e ; empirical estimates of the ratio N_b/N_e actually exceed one in many taxa including some plants, being greater for species with larger adult longevity and lower age at maturity (Waples et al., 2013).

Last, small sample bias can be an obstacle for practical estimation and monitoring of contemporary N_e in many tree populations. Except for endangered species or marginal populations, forest trees, as many marine species, are expected to have large effective population sizes, as a result of their typically enormous census population size and lifetime fecundity, high outcrossing, long-distance dispersal, and largely pre-reproductive density-dependent mortality (Petit and Hampe, 2006). The estimation of contemporary N_e from genetic data is problematic when its value is large, because the drift or inbreeding signal to be detected is inversely proportional to N_e and may thus be rapidly dwarfed by sampling noise as N_e increases (Wang et al., 2016; Waples, 2016). Already Hill (1981) warned that, since obtaining reliable estimates of N_e with the LD method required sample sizes of the same order as N_e , the estimation would not be feasible in natural populations of effective size in the tens of thousands or more. Recent analysis have estimated that, in practice, sample sizes of at least 1 % of the census population size are necessary to reduce the signal-to-sampling-noise ratio to minimally acceptable levels when estimating contemporary N_e with the LD method (Marandel et al., 2019). Both in the case of marine animal and forest tree populations, this sampling requirement could often translate into hundreds (if not thousands) of individuals to be sampled and genotyped per population. Unfortunately, these sample sizes are rarely achievable in practice, an additional factor that might also help explain the exceptionally low contemporary N_e/N estimates obtained for some large populations of marine species using genetic methods (Waples, 2016; Marandel et al., 2019).

The combination of large N_e and relatively small samples is indeed expected to produce downwardly biased contemporary N_e estimates from genetic methods (England et al., 2006; Wang, 2016; Marandel et al., 2019). For instance, the LD method has an expected bimodal distribution of N_e estimates when applied to ideal isolated populations of large constant size, with a substantial proportion of negative estimates (implying that the best possible estimate is infinitely large; Hill, 1981) and most of the remainder being orders of magnitude below the true value (Waples, 2016). Independently of potential biases, statistically detecting actual temporal changes in N_e for genetic monitoring may become virtually impossible in large populations, because of the low

precision and the wide (frequently including infinity) confidence intervals of N_e estimates. In addition, as the drift signal becomes smaller with increasing N_e , the confounding effect of factors such as migration gets amplified, so that the performance of contemporary N_e estimators deteriorates further for large and non-isolated populations (Gilbert and Whitlock, 2015), a most frequent scenario for forest trees.

In the paragraphs above, we have highlighted potential practical problems associated to marker-based estimation of contemporary N_e in natural tree populations with different demographic characteristics. Published studies have identified and quantified these inference problems mostly via simulation analyses, in which estimation errors were calculated by comparing assumed N_e values against estimates obtained from simulated data sets. We do not intend to replicate or extend such numerical analyses here, but rather present three empirical case studies that may help illustrate the uncertainty faced by forest researchers and managers when interpreting contemporary N_e estimates obtained for large continuous tree populations (of unknown N_e), given potential estimation errors identified in the literature. Focusing on maritime pine (*Pinus pinaster* L.), an ecologically and commercially important species with large continuous populations scattered across south-western Europe, we attempt to estimate N_e using the widely-used LD method in: (i) 19 large and continuously-distributed populations where a relatively small number of individuals were sampled from small sampling areas, (ii) a large and continuously-distributed population for which a comparatively larger, but still rather spatially restricted, sample was available, and (iii) an intensively sampled small isolated population.

3. Case studies of contemporary N_e estimation in tree populations

3.1. Small and spatially restricted samples from large continuous tree populations

Pinus pinaster occurs in the Iberian Peninsula, Southern France (including Corsica), Italy and Northern Africa, exhibiting a fragmented distribution across its range, with many large populations each comprising tens of thousands of individuals growing continuously over large areas, but also a few small marginal populations (Fig. A1 in Online Appendix 1). It is a wind-pollinated, wind-dispersed and highly outcrossing species, with early reproductive maturity, annual reproduction, and a maximum lifespan of around two centuries (Alía and Martín, 2003). The set of local populations found across the species range might be considered a metapopulation, as genetic connectivity has been demonstrated among distant wind-pollinated pine populations (Kremer et al., 2012). Available estimates of historical migration rates among southern Iberian populations of the species are however of a magnitude (mean pairwise $\widehat{mN} = 1.6$; González-Martínez et al., 2007) that is expected to have a minor impact on contemporary estimation of local N_e (Waples and England, 2011). Common garden experiments and molecular surveys have revealed substantial genetic divergence among maritime pine populations (Jaramillo-Correa et al., 2015; Fig. A1 in Online Appendix 1), and there are currently 62 proposed genetic conservation units across its distribution (<http://www.euforgen.org/species/pinus-pinaster>), most of them consisting of an artificially delimited stand of up to a few tens of hectares containing a few thousand individual trees, embedded within a much larger and continuously distributed local population.

In our first case study, we consider 19 large populations (Table 1 and Online Appendix 1), each of them growing continuously over hundreds to thousands of hectares, with census sizes in the order of 10^4 – 10^5 trees per population. We consider each entire population as the target population for N_e estimation, as it would be challenging to delimit a biologically meaningful discrete sub-population within their continuous distribution. Sample sizes per population ranged from 10 to 30 (mean 21.5) adult trees (Table 1). Sampled trees were spaced 30–50 m apart

from each other, typically (and similarly to GCUs) covering only a few hectares of each local target population. These small and spatially restricted samples were intended for phylogeographic and population genomic analyses not having the same sampling requirements as N_e estimation does. But, precisely, reporting N_e estimates based on data sets that were not originally intended for this purpose may represent a frequent source of negative bias in the literature (Marandel et al., 2019), which we intentionally wanted to highlight. Individual samples were genotyped using nine nuclear microsatellites (SSRs), an Illumina GoldenGate assay (177 single nucleotide polymorphisms, SNPs) and an Illumina Infinium assay (3514 SNPs) (see Online Appendix 1 for details). SNP sets included only one randomly selected locus per contig to minimize physical linkage among loci, so that N_e estimates using the LD method reflect contemporary rather than past demographic history (Wang, 2005; Saura et al., 2015). Most of the SNPs from the Illumina GoldenGate assay were also included in the Infinium assay. We estimated contemporary N_e using the bias-corrected LD method of Waples and Do (2008), as implemented in NeEstimator v2.0 software (Do et al., 2014). We obtained separate N_e estimates for the nine SSRs, for the 177 SNPs (lower SNP genotyping effort), and for the 3514 SNPs (higher SNP genotyping effort).

Contemporary N_e estimates based on nine SSRs were infinite for five (26 %) of the 19 large maritime pine populations considered, while estimates for the other 14 populations ranged from $\hat{N}_e = 19$ to 2084 (mean 431) with the upper bound of 95 % confidence intervals (CIs) being infinite for all but one populations (Table 1). Estimates based on SNP markers were finite for all populations, and lower on average than those obtained with SSRs, with values ranging across populations from 19 to 2772 (mean 219) for the small subset of 177 loci, and from 39 to 1608 (mean 374) for the larger set of 3514 loci. Besides producing a larger average N_e estimate, increasing SNP genotyping effort also resulted in a lower proportion of infinite CI upper bounds (58 % vs 84 % for the larger and smaller SNP sets, respectively). There was no correlation between N_e estimates obtained for the 19 populations with each of the three considered marker sets (Spearman rank correlation: $r^2 < 0.06$ with $p > 0.3$ for all three possible pairwise comparisons). This result, along with the wide or infinite CIs frequently obtained, and the bimodal distribution of estimates for SSRs, indicates large estimation uncertainty because of a low signal-to-noise ratio.

Overall, estimates (when not infinite) correspond to an approximate average \hat{N}_e/N ratio in the order of 10^{-3} to 10^{-2} , or lower, resembling the

suspected underestimates reported for large and widely distributed fish populations (Waples, 2016; Marandel et al., 2019). Obviously, the true N_e of sampled populations is unknown, and thus we cannot draw conclusions about estimation bias. We believe, however, that estimates as low as the ones we have obtained should be regarded with caution when target populations are continuously distributed and as large as the studied *P. pinaster* ones, because they might be biased downwards. Two of the potentially problematic factors discussed in the previous section would particularly justify a cautionary negative bias expectation for \hat{N}_e across the studied populations: (i) relatively small spatial coverage of continuously distributed populations (and thus estimates tending to reflect a quantity between NS and N_e , rather than N_e ; Neel et al., 2013), and (ii) small adult samples taken from large populations. The substantial overlap of the range of contemporary N_e estimates that we obtained across the *P. pinaster* studied populations with independent estimates of NS based on patterns of genetic isolation by distance in some Iberian populations of the species ($\widehat{NS} = 38\text{--}154$; De-Lucas et al., 2009) would be consistent with the hypothesis of spatially limited sampling of continuous populations as source of \hat{N}_e bias. This hypothesis has been similarly formulated for widely distributed fish populations (Neel et al., 2013; Nunney, 2016), though it cannot be formally tested given that the true N_e of the studied natural populations remains unknown.

3.2. Larger yet spatially restricted sample from a large continuous tree population

Our second case study focuses on another widely distributed *P. pinaster* population, extending over >5000 ha in Sierra de Espadán Natural Park (Castellón, Spain), with a census size of many thousand individuals, and presumably large N_e . The sample size reached in this case 150 sampled adults but, as in the previous case study, collected from a fairly small area as compared to the total continuous extent of the population (25 trees sampled at each of six plots of 0.6–1.0 ha, totalling around five hectares; see Online Appendix 1 and Budde and González-Martínez, 2022). Our target population for N_e estimation is the entire continuously distributed population. Individuals were genotyped using 3133 polymorphic SNPs from the same Illumina Infinium assay used in the previous case study, including again only one randomly selected locus per contig (Online Appendix 1). We first estimated contemporary N_e using the LD method and the full sample of 150 individuals. The

Table 1

Contemporary N_e estimates for *Pinus pinaster* populations based on different molecular markers sets (SSRs or SNPs). Point estimates and 95 % adjusted parametric confidence intervals (95 % CI) were obtained using the linkage disequilibrium method of Waples and Do (2008) as implemented in NeEstimator v2.0 software (Do et al., 2014), based on a random sample of n adult genotypes after removing loci with minor allele frequency below 0.05.

Population	9 SSRs			177 SNPs			3514 SNPs		
	n	N_e	95 % CI	n	N_e	95 % CI	n	N_e	95 % CI
Arenas de San Pedro	27	64.9	(16.4–Infinite)	27	50.6	(15.0–Infinite)	17	364.1	(132.6–Infinite)
Bayubas de Abajo	27	Infinite	(39.7–Infinite)	27	88.2	(19.0–Infinite)	18	39.1	(11.4–Infinite)
Cadavedo	10	Infinite	(8.4–Infinite)	10	32.6	(4.8–Infinite)	10	156.4	(61.2–Infinite)
Castropol	10	Infinite	(74.0–Infinite)	10	30.4	(13.9–533.5)	10	729.1	(212.1–Infinite)
Coca	19	102.4	(13.3–Infinite)	18	61.5	(33.4–239.8)	18	299.6	(146.4–Infinite)
Cuellar	28	Infinite	(87.7–Infinite)	28	44.7	(17.0–Infinite)	28	799.7	(610.8–1154.7)
Le Verdon sur Mer	30	54.0	(18.5–Infinite)	29	42.9	(16.8–Infinite)	18	194.0	(128.0–387.1)
Leiria	24	418.9	(22.8–Infinite)	24	2772.2	(226.0–Infinite)	27	456.1	(203.4–Infinite)
Mimizan	19	29.2	(5.3–Infinite)	19	49.6	(18.9–Infinite)	22	166.6	(84.1–2294.7)
Olba	22	117.7	(17.0–Infinite)	21	473.3	(99.9–Infinite)	24	154.7	(62.3–Infinite)
Olonne sur Mer	27	144.6	(17.1–Infinite)	25	120.3	(16.4–Infinite)	26	300.1	(165.4–1404.9)
Oria	29	904.2	(32.1–Infinite)	29	27.3	(8.9–Infinite)	24	411.5	(227.9–1908.5)
Pineta	22	22.0	(3.0–Infinite)	23	19.1	(7.5–134.5)	16	402.0	(115.8–Infinite)
Pinia	14	Infinite	(15.6–Infinite)	14	72.7	(23.6–Infinite)	20	517.0	(327.2–1214.1)
S.C. de Ribaterme	12	32.1	(9.4–Infinite)	11	64.7	(20.5–Infinite)	14	122.0	(52.0–Infinite)
Sergude	21	19.2	(8.7–80.3)	21	32.3	(10.7–Infinite)	21	82.2	(48.4–227.6)
St-Jean des Monts	28	2025.3	(46.1–Infinite)	26	31.0	(10.6–Infinite)	28	74.3	(26.9–Infinite)
Tamrabta	24	2083.8	(20.5–Infinite)	24	126.3	(29.8–Infinite)	19	239.7	(129.8–1318.2)
Valdemaqueda	16	22.7	(6.7–Infinite)	16	31.4	(6.7–Infinite)	12	1608.1	(537.1–Infinite)

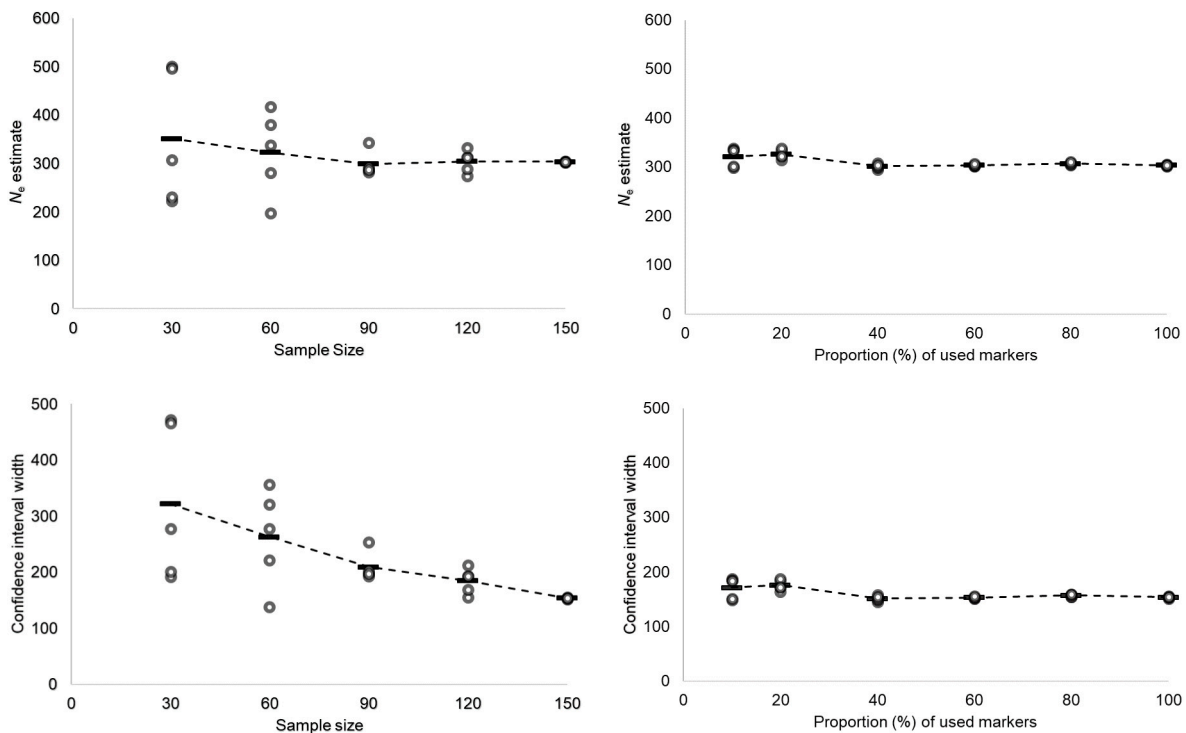


Fig. 1. Point estimates (top panels) and adjusted 95 % confidence interval width (bottom panels) of contemporary N_e from linkage disequilibrium in a large continuous *Pinus pinaster* population (Sierra de Espadán, Spain), based on random subsamples of increasing size up to $n = 150$ individuals (left panels) and on increasing proportions of loci randomly sampled from the total array of 3133 SNP loci (right panels). In all cases, each independent replicate initially involved randomly sampling one SNP per contig to avoid physically close loci. Dotted lines join the means (horizontal bars) of estimates at each subsample size.

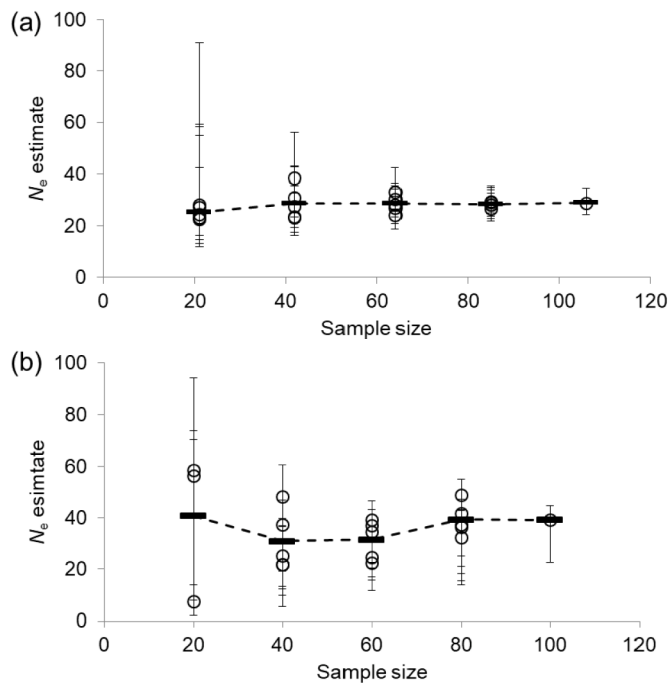


Fig. 2. Estimates of contemporary N_e from linkage disequilibrium at (a) 334 SNPs and (b) nine SSRs in a small isolated *Pinus pinaster* population (Fuencaliente, Spain), based on random subsamples of increasing size (five independent replicates per size) up to the full field sample size ($n = 106$ and 100 for the SNPs and SSRs, respectively). Vertical bars represent 95 % confidence intervals. Dotted lines join the means (horizontal bars) of estimates at each subsample size.

resulting estimate was 303.6 (95 % CI: 300.5–306.8), a value that would again translate into a \hat{N}_e/N ratio in the order of 10^{-3} to 10^{-2} , which as in the previous case study we believe should be regarded with caution as a potential underestimate, although the true N_e and bias remain obviously unknown. The main reason why we would recommend caution before translating this estimate into conservation management decisions is, again, that the population is large and continuously distributed, and that IBD within such populations has been shown to negatively bias contemporary N_e estimates using the LD method, with estimated values approaching the neighbourhood size NS when sampling is spatially limited (Neel et al., 2013). A pattern of IBD does actually exist in the study population, and estimates of NS are in fact of the same order and somewhat lower than the LD-based contemporary N_e estimate ($\hat{N}_S = 133$ –174; see Online Appendix 2), which would be consistent with the hypothesis that N_e estimates for the Sierra de Espadán population are affected by the presence of IBD. If this hypothesis were true, we would not expect small sample size (independently of spatial coverage) to be an important source of additional N_e bias, because the restricted spatial distribution of samples, and not its number, would then be bounding N_e estimates. In other words, no matter how many more individuals are sampled, N_e estimates will still reflect the particular genealogy of individuals from a small deme within a spatially structured population, a relevant issue for all types of N_e estimates. A potential way to test sample size effects is to repeatedly subsample the population sample at increasingly larger sizes, up to the full sample, and examine if the resulting N_e estimates increase with subsample size (England et al., 2006). We would expect them to increase if the presence of IBD is not the predominant factor bounding N_e estimates and if sample size is actually well below N_e (which we ignore). To examine the relationship between \hat{N}_e and sample size, we generated random subsamples of increasing size (30, 60, 90 and 120) via sampling without replacement the population sample (five independent replicates per subsample size), estimating N_e for each subsample. Analogously, we also explored the potential effect

on \hat{N}_e of the number of used loci, by estimating N_e based on increasingly large random subsets of loci. Results did not show an increase of \hat{N}_e neither with sample size nor with number of loci (Fig. 1), even if larger sample size (but not locus number) reduced the width of confidence intervals around N_e estimates (Fig. 1). These results could be interpreted in two ways: (i) N_e estimates are biased because of spatially restricted sampling of a large population under IBD, so increasing sample size (or locus number) does not improve estimates; or (ii) neither small sample size nor the presence of IBD are biasing N_e estimates. Given the big difference between census size and \hat{N}_e , we favour the first interpretation, although we cannot discard the second because we ignore the true N_e .

3.3. Intensive and spatially exhaustive sampling of a small tree population

Our last case study involves the small *P. pinaster* population of Fuencaliente (Ciudad Real, Spain), with a census size of around 300 isolated individuals growing over approximately five hectares. The available sample is more adequate for N_e estimation in this case, involving both a larger proportion of individuals sampled ($n = 106$ adults, i.e. over 30 %) and more exhaustive spatial coverage (individuals randomly sampled over the entire extension of the stand; see Unger et al., 2014 for details). Using genotypic information at nine SSRs and 344 SNPs (Online Appendix 1), we estimated contemporary N_e using the LD method, first for the full data set with either marker type, and second for random individual subsamples of increasing size (with five independent replicates per subsample size). The estimated contemporary effective size of the small stand was $\hat{N}_e = 29.0$ (95 % CI: 24.5–34.6) for the SNPs, and somewhat higher and with wider uncertainty $\hat{N}_e = 39.2$ (95 % CI: 22.6–75.4) for the SSRs. These estimates translate into $\hat{N}_e/N \approx 10^{-1}$, a ratio one or two orders of magnitude larger than our estimates for the large populations. Since we again ignore the true N_e we cannot quantify actual bias, but the fact that the sample size reached over 30 % of the population census and was 3-fold larger than \hat{N}_e , along with the more “normal” (about 0.1 or higher, Waples, 2016) \hat{N}_e/N estimate, would suggest a much smaller potential negative bias than in the case of the large populations. The fact that N_e estimates do not increase with subsample size (Fig. 2) does also suggest the absence of small sample size bias, as isolation by distance cannot be expected to be bounding N_e estimates in the Fuencaliente population, which occupies a very small area (well within pollen dispersal range) that was exhaustively sampled (Neel et al., 2013; Nunney, 2016). Note that it would not be possible to estimate NS from spatial genetic structure patterns in this particular case study, because available methods strongly rely on the assumption of an infinite (or at least a very large) continuous population. All potential biases considered, we hypothesize that the Fuencaliente study represents a demographic and sampling scenario (the only one among our case studies) in which forest researchers and managers would not face strong uncertainty about the reliability of marker-based estimates of contemporary N_e .

4. Genetic estimation of contemporary N_e in forest tree genetic monitoring: Concluding remarks

Notwithstanding the theoretical appeal and general practical utility of N_e for genetic monitoring (Hoban et al., 2020, 2021, 2022), accurately estimating contemporary N_e in forest tree populations will frequently be difficult, requiring extensive spatial coverage, large sample sizes, and sufficient genotyping intensity. Sampling requirements and model assumptions could be easily met in small isolated stands, of greater conservation concern, but more hardly so in widespread tree populations. In particular, small sample size and spatially restricted sampling of large populations may both bias (downwards) contemporary N_e estimates based on genetic data, a bias that will not disappear with increasing numbers of loci. Importantly from the perspective of

genetic monitoring, power to detect actual ongoing changes in contemporary N_e of large populations may be limited, even with ample sample size, because of very large estimation variance. Overall, because of its statistically-inherent assessment difficulty, we call for caution when interpreting marker-based estimates of contemporary N_e for monitoring large continuous tree populations or small GCUs embedded within large continuous tree populations. In such demographic scenarios, it would be worth exploring whether monitoring demography (e.g. number of reproductive individuals), mating system (e.g. outcrossing and correlated paternity) and reproductive success (e.g. effective fecundity variance) could prove more rewarding for detecting temporal changes in contemporary N_e (e.g. via predictive equations; Wang et al., 2016) and in other genetic and demographic parameters relevant for conservation management. We anticipate that obtaining accurate demographic predictions of contemporary N_e of widespread tree populations will also be challenging, requiring extensive fieldwork to gauge spatial heterogeneity in demographic and reproductive parameters.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data available in the references indicated in Table A1 of Online Appendix 1

Acknowledgements

This work was supported by the Spanish Ministry of Agriculture, Fisheries and Food (MAPAMA) [grant number AEG 17-048] in the frame of the measure 15.2 “support to the conservation and use of forest genetic resources” and under Regulation (EU) No 1305/2013 of the European Parliament and of the Council of 17 December 2013 on support for rural development by the European Agricultural Fund for Rural Development (EAFRD) with 75 % co-financing. This paper was also supported by the European Union's Horizon 2020 research and innovation programme [grant agreement No 862221 (FORGENIUS)]. We thank two anonymous reviewers for their constructive criticism.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biocon.2022.109704>.

References

- Alfà, R., Martin, S., 2003. EUFORGEN Technical Guidelines for Genetic Conservation and Use for Maritime Pine (*Pinus pinaster*). Rome, Italy. International Plant Genetic Resources Institute, Rome, Italy.
- Aravanopoulos, F.A., Tollefsrud, M.M., Graudal, L., 2015. Genetic Monitoring Methods for Genetic Conservation Units of Forest Trees in Europe. European Forest Genetic Resources Programme (EUFORGEN), Bioversity International, Rome, Italy.
- Bernos, T.A., Yates, M.C., Fraser, D.J., 2018. Fine-scale differences in genetic and census population size ratios between two stream fishes. *Conserv. Genet.* 19, 265–274. <https://doi.org/10.1007/s10592-017-0997-8>.
- Blower, D.C., Riginos, C., Ovenden, J.R., 2019. NEOGEN: a tool to predict genetic effective population size (N_e) for species with generational overlap and to assist empirical N_e study design. *Mol. Ecol. Resour.* 19, 260–271. <https://doi.org/10.1111/1755-0998.12941>.
- Brauer, C.J., Hammer, M.P., Beheregaray, L.B., 2016. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Mol. Ecol.* 25, 5093–5113. <https://doi.org/10.1111/mec.13830>.
- Budde, K.B., González-Martínez, S.C., 2022. *Pinus pinaster* genotypes from Sierra Espadán [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.6345964>.
- Crow, J.F., Kimura, M., 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York.

- De-Lucas, A.I., González-Martínez, S.C., Vendramin, G.G., et al., 2009. Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* aiton. *Mol. Ecol.* 18, 4564–4576. <https://doi.org/10.1111/j.1365-294X.2009.04372.x>.
- Do, C., Waples, R.S., Peel, D., et al., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol. Ecol. Resour.* 14, 209–214. <https://doi.org/10.1111/1755-0998.12157>.
- England, P.R., Cornuet, J.M., Berthier, P., et al., 2006. Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conserv. Genet.* 7, 303–308. <https://doi.org/10.1007/s10592-005-9103-8>.
- EUFORGEN, 2021. Forest Genetic Resources Strategy for Europe. European Forest Institute.
- Fady, B., Bozzano, M., 2021. Effective population size does not make a practical indicator of genetic diversity in forest trees. *Biol. Conserv.* 253, 108904 <https://doi.org/10.1016/j.biocon.2020.108904>.
- Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. *Genet. Res. (Camb.)* 66, 95–107. <https://doi.org/10.1017/S0016672308009695>.
- Frankham, R., 1996. Relationship of genetic variation to population size in wildlife. *Conserv. Biol.* 10, 1500–1508. <https://doi.org/10.1046/j.1523-1739.1996.10061500.x>.
- Fussi, B., Westergren, M., Aravanopoulos, F.A., et al., 2016. Forest genetic monitoring: an overview of concepts and definitions. *Environ. Monit. Assess.* 188 <https://doi.org/10.1007/s10661-016-5489-7>.
- Gilbert, K.J., Whitlock, M.C., 2015. Evaluating methods for estimating local effective population size with and without migration. *Evolution (NY)* 69, 2154–2166. <https://doi.org/10.1111/evo.12713>.
- González-Martínez, S.C., Gómez, A., Carrión, J.S., et al., 2007. Spatial genetic structure of an explicit glacial refugium of maritime pine (*Pinus pinaster* Aiton) in southeastern Spain. In: Weiss, S., Ferrand, N. (Eds.), *Phylogeography of Southern European Refugia*. Springer, Dordrecht, pp. 257–269.
- Hansen, M.M., Olivieri, L., Waller, D.M., Nielsen, E.E., 2012. Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* 21, 1311–1329. <https://doi.org/10.1111/j.1365-294X.2011.05463.x>.
- Hare, M.P., Nunney, L., Schwartz, M.K., et al., 2011. Understanding and estimating effective population size for practical application in marine species management. *Conserv. Biol.* 25, 438–449. <https://doi.org/10.1111/j.1523-1739.2010.01637.x>.
- Hill, W.G., 1981. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* 38, 209–216. <https://doi.org/10.1017/S0016672300020553>.
- Hoban, S., Bruford, M., D'Urban Jackson, J., et al., 2020. Genetic diversity targets and indicators in the CBD post-2020 global biodiversity framework must be improved. *Biol. Conserv.* 248, 108654 <https://doi.org/10.1016/j.biocon.2020.108654>.
- Hoban, S., Paz-Vinas, I., Aitken, S., et al., 2021. Effective population size remains a suitable, pragmatic indicator of genetic diversity for all species, including forest trees. *Biol. Conserv.* 253, 108906 <https://doi.org/10.1016/j.biocon.2020.108906>.
- Hoban, S., Archer, F.I., Bertola, L.D., et al., 2022. Global genetic diversity status and trends: towards a suite of essential biodiversity variables (EBVs) for genetic composition. *Biol. Rev.* <https://doi.org/10.1111/brv.12852>.
- Jaramillo-Correa, J.-P.P., Rodríguez-Quilón, I., Grivet, D., et al., 2015. Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* 199, 793–807. <https://doi.org/10.1534/genetics.114.173252>.
- Kinlan, B.P., Gaines, S.D., Lester, S.E., 2005. Propagule dispersal and the scales of marine conservation process. *Divers. Distrib.* 11, 139–148. <https://doi.org/10.1111/j.1366-9516.2005.00158.x>.
- Koskela, J., Lefèvre, F., Schueler, S., et al., 2013. Translating conservation genetics into management: pan-european minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biol. Conserv.* 157, 39–49. <https://doi.org/10.1016/j.biocon.2012.07.023>.
- Kremer, A., Ronco, O., Robledo-Arnuncio, J.J., et al., 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol. Lett.* 15, 378–392. <https://doi.org/10.1111/j.1461-0248.2012.01746.x>.
- Lefèvre, F., Alía, R., Bakkebo Fjellstad, K., 2020. *Dynamic Conservation and Utilization of Forest Tree Genetic Resources: Indicators for In Situ and Ex Situ Genetic Conservation and Forest Reproductive Material*. European Forest Genetic Resources Programme (EUFORGEN), European Forest Institute.
- Marandel, F., Lorange, P., Berthel, O., et al., 2019. Estimating effective population size of large marine populations, is it feasible? *Fish Fish.* 20, 189–198. <https://doi.org/10.1111/faf.12338>.
- Nadachowska-Brzyska, K., Konczal, M., Babik, W., 2022. Navigating the temporal continuum of effective population size. *Methods Ecol. Evol.* 13, 22–41. <https://doi.org/10.1111/2041-210X.13740>.
- Neel, M.C., McKelvey, K., Ryman, N., et al., 2013. Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity (Edinb)* 111, 189–199. <https://doi.org/10.1038/hdy.2013.37>.
- Nei, M., Tajima, F., 1981. Genetic drift and estimation of effective population size. *Genetics* 98, 625–640. <https://doi.org/10.1093/genetics/98.3.625>.
- Nunney, L., 2016. The effect of neighborhood size on effective population size in theory and in practice. *Heredity (Edinb.)* 117, 224–232. <https://doi.org/10.1038/hdy.2016.76>.
- Ovenden, J.R., Peel, D., Street, R., et al., 2006. The genetic effective and adult census size of an australian population of tiger prawns (*Penaeus esculentus*). *Mol. Ecol.* 16, 127–138. <https://doi.org/10.1111/j.1365-294X.2006.03132.x>.
- Ovenden, J.R., Leigh, G.M., Blower, D.C., et al., 2016. Can estimates of genetic effective population size contribute to fisheries stock assessments? *J. Fish Biol.* 89, 2505–2518. <https://doi.org/10.1111/jfb.13129>.
- Petit, R.J., Hampe, A., 2006. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* 37, 187–214.
- Pita, A., Pérez, M., Velasco, F., Presa, P., 2017. Trends of the genetic effective population size in the southern stock of the european hake. *Fish. Res.* 191, 108–119. <https://doi.org/10.1016/j.fishres.2017.02.022>.
- Portnoy, D.S., McDowell, J.R., McCandless, C.T., et al., 2009. Effective size closely approximates the census size in the heavily exploited western Atlantic population of the sandbar shark, *Carcharhinus plumbeus*. *Conserv. Genet.* 10, 1697–1705. <https://doi.org/10.1007/s10592-008-9771-2>.
- Robledo-Arnuncio, J.J., Rousset, F., 2010. Isolation by distance in a continuous population under stochastic demographic fluctuations. *J. Evol. Biol.* 23, 53–71. <https://doi.org/10.1111/j.1420-9101.2009.01860.x>.
- Rousset, F., 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 13, 58–62. <https://doi.org/10.1046/j.1420-9101.2000.00137.x>.
- Ryman, N., Laikre, L., Hössjer, O., 2019. Do estimates of contemporary effective population size tell us what we want to know? *Mol. Ecol.* 28, 1904–1918. <https://doi.org/10.1111/mec.15027>.
- Saura, M., Tenesa, A., Woolliams, J.A., et al., 2015. Evaluation of the linkage-disequilibrium method for the estimation of effective population size when generations overlap: an empirical case. *BMC Genomics* 16, 922. <https://doi.org/10.1186/s12864-015-2167-z>.
- Schwartz, M.K., Luikart, G., Waples, R.S., 2007. Genetic monitoring as a promising tool for conservation and management. *Trends Ecol. Evol.* 22, 25–33. <https://doi.org/10.1016/j.tree.2006.08.009>.
- Unger, G.M., Vendramin, G.G., Robledo-Arnuncio, J.J., 2014. Estimating exotic gene flow into native pine stands: zygotic vs. gametic components. *Mol. Ecol.* 23, 5435–5447. <https://doi.org/10.1111/mec.12946>.
- de Vries, S.M.G., Alan, M., Bozzano, M., 2015. *Pan-European strategy for genetic conservation of forest trees and establishment of a core network of dynamic conservation units*. In: *European Forest Genetic Resources Programme (EUFORGEN)*. Bioversity International, Rome, Italy.
- Wang, J., 2005. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 360, 1395–1409. <https://doi.org/10.1098/rstb.2005.1682>.
- Wang, J., 2016. A comparison of single-sample estimators of effective population sizes from genetic marker data. *Mol. Ecol.* 25, 4692–4711. <https://doi.org/10.1111/mec.13725>.
- Wang, J., Santiago, E., Caballero, A., 2016. Prediction and estimation of effective population size. *Heredity (Edinb.)* 117, 193–206. <https://doi.org/10.1038/hdy.2016.43>.
- Waples, R.S., 2016. Tiny estimates of the N_e/N ratio in marine fishes: are they real? *J. Fish Biol.* 89, 2479–2504. <https://doi.org/10.1111/jfb.13143>.
- Waples, R.S., 2021. Relative precision of the sibship and LD methods for estimating effective population size with genomics-scale datasets. *J. Hered.* 112, 535–539. <https://doi.org/10.1093/jhered/esab042>.
- Waples, R.S., Do, C., 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol. Ecol. Resour.* 8, 753–756. <https://doi.org/10.1111/j.1755-0998.2007.02061.x>.
- Waples, R.S., England, P.R., 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 189, 633–644. <https://doi.org/10.1534/genetics.111.132233>.
- Waples, R.S., Luikart, G., Faulkner, J.R., Tallon, D.A., 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proc. R. Soc. B Biol. Sci.* 280, 20131339. <https://doi.org/10.1098/rspb.2013.1339>.
- Waples, R.S., Antao, T., Luikart, G., 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* 197, 769–780. <https://doi.org/10.1534/genetics.114.164822>.
- Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16, 97–159. <https://doi.org/10.1093/genetics/16.2.97>.
- Wright, S., 1943. Isolation by distance. *Genetics* 28, 114–138. <https://doi.org/10.1093/genetics/28.2.114>.
- Yates, M.C., Bernos, T.A., Fraser, D.J., 2017. A critical assessment of estimating census population size from genetic population size (or vice versa) in three fishes. *Evol. Appl.* 10, 935–945. <https://doi.org/10.1111/eva.12496>.