INTRODUCTION

Effective population size ($N_e$) is one of the key parameters in evolutionary biology. It is defined as the size of an idealized population (i.e., Wright-Fisher population; Fisher, 1930; Wright, 1931) that would have the same rate of genetic drift and inbreeding as the population in focus. $N_e$ determines the amount and distribution of genetic variation in a population in interaction with several evolutionary forces like mutation, recombination, selection and migration (Crow & Kimura, 1970). As a consequence, it is a good indicator of evolutionary potential and fitness (Lynch et al., 1995). It is also necessary to predict fixation probabilities of deleterious and beneficial alleles (Robertson, 1961), and is essential to infer demographic history (Hsieh et al., 2016; Juric et al., 2016; Ostrander et al., 2017). Several ways to estimate $N_e$ based on genetic data have been developed (reviewed in Luikart et al., 2010; Wang, 2005; Wang et al., 2016). These approaches differ in the mathematical framework used and range from approaches based on classic population genetic theory (Crow &
Methods that infer historical $N_e$ have been applied to understand demographic history in a wide range of species (Charlesworth, 2009). The advent of next-generation sequencing (NGS) techniques and whole-genome resequencing protocols has allowed attempts to infer complex demographic scenarios (Beichman et al., 2018). The advances of demographic history inference include, for example, estimation of $N_e$ of multiple nonequilibrium populations with different levels of connectivity (Steinrücken et al., 2019) and inference of historical $N_e$ fluctuations over relatively short or long periods of time (Barbato et al., 2015; Browning & Browning, 2015; Li & Durbin, 2011; Santiago et al., 2020; Terhorst et al., 2017). In contrast, contemporary $N_e$ remains difficult to estimate in nature, especially for large populations (Gilbert & Whitlock, 2015; Marandel et al., 2019; Serbezov et al., 2012).

In principle, by having information on demographic and life history parameters (such as census size, sex ratio, variance in reproductive success, mating system and/or pedigree) it is possible to estimate current $N_e$ (Caballero, 1994; Wang, 2005; Wang & Caballero, 1999; Waples & England, 2011). However, detailed data on these parameters are rarely available and generally difficult to collect. In practice, genetic methods therefore have to be widely used (Palstra & Fraser, 2010). However, in theory, data from a large number of loci should be required to make precise estimates of $N_e$. DNA is small and when the signal of drift can be seen, even with small sample sizes and limited number of loci/alleles. In contrast, estimating current $N_e$ in large populations ($N_e > 1000$) has proven to be challenging, at least when the number of markers is low, and often results in $N_e$ estimates indistinguishable from infinity (Marandel et al., 2019; Waples & Do, 2010). However, in theory, data from a large number of loci should harbor sufficient information to provide information on $N_e$ of large populations (Luikart et al., 2010; Wang, 2016; Waples & Do, 2010). How much data is needed to precisely estimate $N_e$ in such cases remains unclear and conclusions are often based on simulation studies.

Here, we use large-scale genomic data to estimate contemporary $N_e$ of an island population of collared flycatchers (Ficedula albicollis). We sampled 85 individuals at two time points 22 years (about nine generations) apart on Gotland, a Baltic Sea island that is thought to have been colonized by collared flycatchers and with approximately 4500 current breeding pairs (L. Gustafsson, personal observation). Based on high coverage, whole-genome resequencing data, we used both temporal and LD methods to estimate contemporary $N_e$ with data at a scale that rarely has been applied to natural populations before.

2 | MATERIALS AND METHODS

2.1 | Study population

We sampled 85 collared flycatcher (Ficedula albicollis) individuals from the Baltic island of Gotland. Forty-five adult birds were sampled in 1993 and another 40 in 2015 (22 years or approximately nine generations apart). The collared flycatcher is a small passerine bird that breeds mainly in southeast Europe and southwest Asia but isolated populations are also found at two Swedish islands in the Baltic Sea (Gotland and Öland). Gotland has been inhabited for at least 150 years but the detailed colonization history remains unknown (Lundberg & Alatalo, 1992).

2.2 | Sequencing and data filtering

DNA was extracted from blood following established protocols (described in e.g., Burri et al., 2015). All individuals were sequenced with a paired-end approach on an Illumina HiSeqX instrument for a read length of 150 bp and an insert size of 350 bp. Reads were mapped to a repeat-masked collared flycatcher reference genome assembly, version FicAlb1.5 (GenBank Accession GCA_000247815.2), using BWA mem 0.7.13 (Li & Durbin, 2009) and further processed with Samtools 1.3 (Li et al., 2009). Reads were deduplicated with PICARD 2.0.1 (http://broadinstitute.github.io/picard/), and realigned and recalibrated with GATK3.6 (DePristo et al., 2011). Variants were called independently for each time sample with GATK's HaploTypeCaller and GenotypeGVCFs 3. The mean genome-wide coverage varied from 30 to 50 among all sequenced individuals. After variant calling the data consisted of 19.8 million single nucleotide polymorphisms (SNPs) in the 1993 sample and 19.3 million in the 2015 sample. Aiming for a very high quality data set, we applied a strict filtering using VCFtools (Danecek et al., 2011). Specifically, we conservatively removed all SNPs where any of the individuals had a coverage lower than 10 or higher than 100, or a mapping quality below 20. Additionally, we removed all sites within 1 kb from scaffold ends. We considered variants that were segregating in both time data sets as well as variants that were only segregating in one of the two cohorts.

To obtain a large number of independent markers in a computationally efficient way, we proceeded in two steps. We first sampled 1,000,000 SNPs (from 29 chromosomes and the two largest unassigned linkage groups) before further removing all sites with neighbouring SNPs within a distance of 2 kb. This distance is known to be the approximate distance at which linkage gets back to background levels in the collared flycatcher (Ellegren et al., 2012; Figure S1; total map length has been estimated at 3132 cM [mean recombination rate for the whole genome equals
3.1 but ranges from 2.0–11.1 cM/Mb; large chromosomes have smaller recombination rate; Kawakami et al., 2014). We used VCFtools to investigate genetic structure within and between time cohorts using a principal component analysis (PCA). \( F_{ST} \) (Weir & Cockerham, 1984) was also obtained between time cohorts, confidence intervals were obtained by resampling the data set 500 times before computing \( F_{ST} \).

### 2.3 | Contemporary \( N_e \) estimation

We used the temporal and LD-based methods to estimate contemporary \( N_e \). Temporal methods use changes in allele frequencies over several generations to estimate recent \( N_e \). This approach relies on the idea that the variance in allele frequency change between generations is a function of \( N_e \) (Krimbas & Tsakas, 1974; Waples, 1989). The larger changes in allele frequencies over time, the smaller the inferred \( N_e \). Linkage disequilibrium methods are based on the fact that random genetic drift in a finite population creates associations between linked and unlinked alleles and is therefore informative about \( N_e \) (Waples, 1989; Waples & Do, 2010). For both types of analyses, we filtered the data by excluding annotated genes, conserved elements (Craig et al., 2018), and regions with estimated recombination rate of zero (Kawakami et al., 2014).

We used three different temporal \( N_e \) estimators: a likelihood based \( N_e \) by Hui and Burt (2015), and two different \( F \)-statistics: \( F_s \) by Jorde and Ryman (2007) and \( F_c \) by Nei and Tajima (1981). The likelihood-based estimator uses a computationally efficient hidden Markov algorithm and continuous approximation of allele frequencies. This approach makes the method well suited for estimation of larger \( N_e \). The method is implemented in \texttt{HAP} package (NB) and we used a slightly modified version where we allow for a noninteger number of generations. We used a generation time of 2.5 years, \( N_e \) prior ranging from 50 to 100,000. The two additional \( F \)-statistics are moment-based estimators and can be calculated by obtaining standardized variance of allele frequencies changes. Both \( F \)-statistics were calculated in \texttt{NeESTIMATOR v2.1} (Do et al., 2014). We used the plan I sampling procedure (sampling adults after the reproduction or before reproduction but returning them to population; Waples, 1989). To estimate LD-based \( N_e \) we used an approach developed by Waples and Do (2008), LDNe, implemented in \texttt{NeESTIMATOR v2}. LDNe is based on the mean of squared interlocus correlations of allele frequencies obtained from the Burrows method (Waples, 2006; Weir, 1996). We ran LDNe using two variations of the method, first including comparisons between all the SNPs in the data set and then omitting comparisons of loci on the same chromosomes while still comparing each SNP to all the SNPs on different chromosomes. The latter removed any physical linkage but greatly reduced the number of comparisons. We performed analysis for each time cohort independently. We performed jackknife to estimate 95% confidence intervals for \( F \)-statistics and LD-based estimate. In all methods we ignored SNPs with a frequency lower than 5%. The number of SNPs after filtering was 78,636 and the median distance between neighbouring SNPs was >3 kb.

Additionally, in order to investigate the power of all methods, we created several smaller data sets for both time cohorts by varying the number of SNPs from 1000 to 78,636 in increments of 500 sites. To further investigate the influence of physical linkage, we varied the minimum distance between SNPs in the analysed data sets from 1 to 40 kb with increments of 200 bp, thus creating 196 additional data sets per time cohort. Importantly, those data sets dramatically vary in number of SNPs as number of SNPs and average distance between SNPs are inescapably correlated.

### 2.4 | Recent \( N_e \) changes over time

We used the GONE method to infer recent changes of \( N_e \) over time from linkage disequilibrium and SNP data (Santiago et al., 2020). We performed the analysis for each time cohort separately and applied no frequency filters to the SNP data sets. We used recombination rates from a collared flycatcher linkage map (Kawakami et al., 2014) and set maximum recombination rate between pairs of analysed SNPs to 0.01 (hc = 0.01).

### 3 | RESULTS

#### 3.1 | Summary of data

We performed whole-genome resequencing of 85 adults of collared flycatcher from the Baltic Sea island Gotland at a mean coverage of 38.6× (range: 29.5×–50.2×). Forty-five birds were sampled in 1993 and 40 in 2015. We stringently filtered genotypes based on the coverage and mapping quality before removing any SNPs with missing data. Additional filtering of annotated genes, conserved elements, regions with estimated recombination rate of zero and non-independent SNPs based on physical proximity reduced our data to 131,902 SNPs.

This data set was used in all downstream analysis and filtered to 78,636 SNPs according to software settings. The average LD between the pairs of SNPs was \( r^2 = .0228 \) and \( r^2 = .0257 \) for 1993 and 2015 time cohorts, respectively. Additional data sets were created to test LD-based method performance as described above.

#### 3.2 | Genetic structure

A PCA of genetic variation (Figure 1) suggested that there was no clear structure either between or within the time samples. This was corroborated by the observation of very low genetic differentiation between the cohorts (\( F_{ST} = 2.6 \times 10^{-4}; \) 95% CI: \( 1.73 \times 10^{-4} - 3.34 \times 10^{-4} \)). The genomic \( F_{ST} \) landscape was flat and no \( F_{ST} \) peaks were visible on any of the chromosomes (Figure S1).
Contemporary $N_e$ estimates are summarized in Table 1. In general, all temporal methods gave relatively similar results of contemporary $N_e$ ranging from approximately 4000 to 7000 individuals. The likelihood-based estimator provided the highest estimate of $\hat{N}_B = 6921$ (95% CI: 5015–11,079; Figure 2). Both $F$-statistics provided lower point estimates, $F_c = 5804$ (95% CI: 4837–10,341) and $F_s = 3921$ (95% CI: 3148–5198). In concordance with the results from temporal methods, GONE analysis showed that $N_e$ for the most recent past equalled approximately few thousands (Figure 3). Additionally, the analysis indicated that the collared flycatcher population used to be higher (approximately 10,000) and slightly declined over the last 100 generations.

The LD-based method (with no linkage information included) provided higher estimates of contemporary $N_e$ and varied from approximately 20,000 for the 2015 cohort ($N_e = 20,094$; 95% CI: 8430–Infinity) to 33,000 individuals for the 1993 cohort ($N_e = 32,534$; 95% CI: 10,670–Infinity). We applied two corrections presented by Waples et al. (2016, Equations 1a and 1b) to account for linkage between SNPs in the data set. The corrected $N_e$ estimates equaled approximately 38,000 and 23,000 for 1993 and 2015 time cohorts, respectively, for both corrections. The LD-based method that is restricted to interchromosomal comparisons, effectively taking into account physical linkage information by comparing each SNP to all SNPs on the other chromosomes, returned negative $N_e$ with infinite credible estimates suggesting large $N_e$.

We varied the number of SNPs using both types of methods (temporal and LD) to explore the effect of sampling on $N_e$ estimation (Figure 4 and Figure S2). Data sets with less than 30,000–40,000 SNPs provided $N_e$ estimates with high variance, ranging from approximately 1000 to infinity. Data sets with a higher number of SNPs provided similar results to that obtained in the full analyses.

We varied the minimal distance between SNPs (1–40 kb; LD method) to explore the influence of physical linkage on $N_e$ estimation (Figures S3 and S4). The median distance was always higher than the minimal distance and ranged from 1.8 to 63 kb. We observed a large variance in the obtained $N_e$ estimates ranging from 9921 (2015 time cohort) and 12,642 (1993 time cohort) to infinity. Approximately half of the data sets with a distance of >10 kb between SNPs gave infinite $N_e$ (44% and 53% for 1993 and 2015 time cohorts). The variance among data sets with SNPs with larger distance was also higher.

### DISCUSSION

Effective population size is a key concept in population genetics and evolutionary biology. Paradoxically, it is at the same time one of the most difficult parameters to estimate, especially for contemporary $N_e$ and when $N_e$ is large (1000 or larger). In this study, we estimated contemporary $N_e$ in a wild bird population using methods based on linkage disequilibrium and temporal comparisons of allele frequencies. Our study is one of very few applying whole-genome resequencing data to estimate current $N_e$ and using temporal and LD methods to estimate effective population size in a large natural population.
4.1 Effective population size of Gotland collared flycatchers

We have previously estimated the historical, long-term $N_e$ of collared flycatchers using Approximate Bayesian Computation modeling (Nadachowska-Brzyska et al., 2013; Nater et al., 2015) and the pairwise sequentially Markovian coalescent (PSMC; Nadachowska-Brzyska et al., 2016). About 200,000 years ago, $N_e$ was large, ≈500,000–600,000. Populations of collared flycatchers decreased towards the middle of the Last Glacial Period (50,000 years ago), and then showed signs of steady increase towards 10,000 years ago. In contrast to mainland collared flycatcher populations, the Baltic population declined to a level below 200,000 individuals around Last Glacial Maximum. This difference may reflect different ancestry of current collared flycatcher populations. It is important to note that PSMC analysis did not provide information on recent $N_e$, reflects historical $N_e$ of an ancestral population and is sensitive to ancestral population structure and admixture (Li & Durbin, 2011).

The colonization history of collared flycatchers on Gotland is unknown, including lack of knowledge about when (after the Last Glacial Period) colonization took place and if it was associated with a severe bottleneck. The distance to the central European mainland is about 400 km, and the distance to breeding areas further south where collared flycatchers are abundant is larger than that. Collared flycatchers were registered on Gotland some 150 years ago (Lundberg & Alatalo, 1992), while Carl von Linné (Carl Linnaeus) did not make notes of the species when visiting Gotland in the summer of 1741 (Linnaeus, 1745). However, Linnaeus arrived on Gotland on one of the first days of July, when flycatchers no longer sing and are less conspicuous (unless seen feeding nestlings).

Whole-genome resequencing data indicated that the degree of genetic diversity in this island population ($4.5 \times 10^{-3}$) is comparable to mainland populations (Burri et al., 2015). Both temporal and LD-based methods as well as GONE indicated a contemporary $N_e$ of at least few thousands. This is in line with the results from a detailed analysis of identity by descent (IBD) segments of collared flycatchers from the nearby island Öland (Kardos et al., 2017), an island that was probably colonized by flycatchers from the Gotland population about 50 years ago. That analysis indicated that ancestral $N_e$, which probably reflects the $N_e$ of the source (Gotland) population, was at least 5000. All these results suggest that the Baltic Sea population has been large for a relatively long time, and there was no strong bottleneck associated with the colonization event on Gotland island. Similarly, although GONE analysis indicated some decline of Gotland population over the last 100 generations, no drastic bottleneck was detected.

4.2 Large $N_e$ - performance of genetic methods to estimate contemporary $N_e$

Genetic methods to estimate $N_e$ rely on the genetic drift signal present in the data. It follows that the larger the population the more difficult it is to estimate its effective population size. In practice, the estimates for large populations may be indistinguishable from infinity, especially when the number of individuals and loci analyzed are small (Waples & Do, 2010). Simulation studies have been used to evaluate the performance of methods based on temporal approaches and linkage disequilibrium, and usually considered populations of small to moderate size (Wang, 2002; Waples, 2006; Waples & Do, 2010; Waples & England, 2011). The amount of information used by temporal methods increases linearly with the number of loci. For LD-based methods, it increases with the square of the number of loci as it uses information on LD across all loci. Wang (2016) explored the possibility of estimating contemporary $N_e$ of
large populations and showed that LD-based methods can provide reasonably good estimates of $N_e$ even for populations of a size as high as 30,000. These results were obtained from simulations with a sample size of 100 individuals and only 20 microsatellite loci. Another simulation-based study (Waples & Do, 2010) indicated that LD methods work well for small populations (100–200 or less), and temporal methods are more precise in contemporary $N_e$ estimation. The study indicated that it is challenging to obtain estimates for large ($N_e > 1000$) populations due to very weak signal of drift. Similarly, results from a recent study by Waples et al. (2020) suggested that precision of LD based methods does not increase much with data sets larger than a few thousand loci. All studies mentioned above were based on simulations.

We investigated the possibility of estimating relatively large $N_e$ in a natural population and tested a similar number of individuals but many more loci (78,636 SNPs after filtering) than in the simulations of Wang (2016). Using temporal methods, we were able to estimate contemporary effective population size with quite narrow credible intervals. All three temporal methods gave similar results suggesting that $N_e$ of an order of a few thousand can be estimated when temporal genomic data are available. On the other hand, LD-based results were higher and were associated with much higher uncertainty. Several estimates obtained using LD-based methods on different number of SNPs and variable physical distance between neighbouring SNPs resulted in infinitive credible intervals. This result is in line with simulation studies indicating low power of LD-based methods.
for contemporary $N_e$ estimation when applied to large populations and with limited number of loci. Nonindependence of loci used in LD-based approaches may not only bias results but significantly also limit the power of LD-based methods due to pseudoreplication (Waples et al., 2020).

### 4.3 Confounding factors

Genetic drift is not the only evolutionary force that changes allele frequencies over time and thereby affects patterns of LD along the genome. In the case of contemporary $N_e$ estimation, migration is a force that can drastically influence the distribution of alleles in the population. Importantly, migration may bias $N_e$ estimation in different ways. We expect overestimation of $N_e$ when there is immigration from populations with limited differentiation to the focal population; we expect to see less drift due to the influx of alleles that are already present in the population. In this case the estimated $N_e$ reflects the $N_e$ of a metapopulation. Alternatively, we expect underestimation of LD-based $N_e$ when the population exchanges migrants with substantially differentiated populations. In this case many foreign alleles enter the population, giving the impression of stronger drift.

The magnitude of the bias depends on the amount of migration between populations. Several simulation studies have evaluated the influence of migration on contemporary $N_e$ estimation (Gilbert & Whitlock, 2015; Ryman et al., 2014) and concluded that migration rates below 1% ($m = 0.01$) do not have a strong effect on estimates (in some cases even migration of the order of 5%–10% did not led to substantial bias; Waples & England, 2011). In the case of collared flycatchers on Gotland, a 1% migration rate would correspond to a large number of new birds coming to the island every year. The closest flycatcher population that could potentially serve as a migration source is located at Öland island. A substantial part of the Öland population is ringed and ringed individuals are extremely rare on Gotland island. While this observation excludes extensive migration between islands one cannot exclude immigration from other mainland populations.

Other factors that can potentially bias $N_e$ estimation include selection and overlapping generations in the studied population. Selection creates linkage disequilibrium and changes in allele frequencies at genomic region under selection. We sought to minimize this effect by conservatively filtering functional regions potentially under strong selection pressure (genes, conserved elements) and regions of very low recombination rate where linked selection might be prevalent.). When applying temporal methods, a bias might arise from using age-structured populations (overlapping generations). This can be minimized by taking temporal samples several generations apart, at least 3–5 generations apart (Waples & Yokota, 2007). With about nine generations between the two time samples analysed in this study, results comparable to that obtained with nonoverlapping generations can be expected.

It was suggested that whole-genome sequencing and/or genotyping of several thousands of loci can potentially overcome problems associated with estimation of high contemporary effective population size (Luikart et al., 2010; Wang, 2016; Waples & Do, 2010). Our results indicated that several classic ($F$-statistics) as well as new (likelihood-based) temporal methods can provide reliable estimates of high contemporary effective population size. Nevertheless, the contemporary $N_e$ estimation remains challenging. In particular, temporal methods need population samples taken several generations apart and such data may be unavailable for many populations of interest. Single sample estimators that rely on LD information do not have this limitation, but the results may provide very wide credible intervals and be difficult to interpret. The LD-based results for contemporary $N_e$ estimation should always be interpreted with caution and preferably be augmented with another method for estimating contemporary $N_e$ or recent population dynamics.

### ACKNOWLEDGEMENTS

We would like to thank Reija Dufva for help with DNA preparation. We thank Robin Waples, Tin-Yu Hui, Chi Do and Nils Ryman for fruitful discussions on contemporary effective population size estimation and/or software code sharing. This study was funded by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Sequencing was performed by the SNP&SEQ Technology Platform at Uppsala University, as part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

### CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

### AUTHOR CONTRIBUTION

Krystyna Nadachowska-Brzyska, Ludovic Dutoit and Hans Ellegren conceived of the study and wrote the manuscript. Ludovic Dutoit performed all main analyses. Linnea Smeds provided scripts and analysed raw data. Martin Kardos assisted with interpretation of the results and writing the manuscript. Lars Gustafsson provided samples.

### DATA AVAILABILITY STATEMENT

The genome resequencing data have been made freely available in EMBL-EBI European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession number PRJEB22864.

### ORCID

Krystyna Nadachowska-Brzyska [https://orcid.org/0000-0002-8457-310X](https://orcid.org/0000-0002-8457-310X)
Ludovic Dutoit [https://orcid.org/0000-0002-0164-9878](https://orcid.org/0000-0002-0164-9878)
Linnea Smeds [https://orcid.org/0000-0002-8415-9259](https://orcid.org/0000-0002-8415-9259)
Lars Gustafsson [https://orcid.org/0000-0001-6566-2863](https://orcid.org/0000-0001-6566-2863)
Hans Ellegren [https://orcid.org/0000-0002-5035-1736](https://orcid.org/0000-0002-5035-1736)
Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, 25, 1058–1072. https://doi.org/10.1111/mec.13540

Nater, A., Burri, R., Kawakami, T., Smeds, L., & Ellegren, H. (2015). Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Systematic Biology*, 64, 1000–1017. https://doi.org/10.1093/sysbio/syv045

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

Ostrander, E. A., Wayne, R. K., Freedman, A. H., & Davis, B. W. (2017). *Wang, J. (2002). A pseudo-likelihood method for estimating effective population size. Genetics*, 191, 1358–1370. https://doi.org/10.1534/genetics.111.132233

Palstra, F. P., & Fraser, D. J. (2012). Effective / census population size ratio estimation: A compendium and appraisal. *Ecology and Evolution*, 2, 2357–2365. https://doi.org/10.1002/ece3.329

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104, 531–548. https://doi.org/10.1093/genetics/104.3.531

Palstra, F. P., & Fraser, D. J. (2012). Effective / census population size ratio estimation: A compendium and appraisal. *Ecology and Evolution*, 2, 2357–2365. https://doi.org/10.1002/ece3.329

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104, 531–548. https://doi.org/10.1093/genetics/104.3.531

Robertson, A. (1961). Inbreeding in artificial selection programs. *Genetical Research*, 2, 189–194.

Ryman, N., Allendorf, F. W., Jorde, P. E., Laikre, L., & Hössjer, O. (2014). Samples from subdivided populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. *Molecular Ecology Resources*, 14, 87–99. https://doi.org/10.1111/1755-0998.12154

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, 111.136580

Steinrücken, M., Kamm, J., Spence, J. P., & Song, Y. S. (2019). Inference of complex population histories using whole-genome sequences from multiple populations. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 17115–17120. https://doi.org/10.1073/pnas.1905060116

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49, 303–309. https://doi.org/10.1038/ng.3748

Wakely, J. (2008). Coalescent theory. Roberts and Company Publishers.

Wang, J. (2002). A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research*, 78, 243–257. https://doi.org/10.1017/S0016672301005286

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 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. *Molecular Ecology*, 25, 4692–4711. https://doi.org/10.1111/mec.13725

Wang, J., & Caballero, A. (1999). Developments in predicting the effective size of subdivided populations. *Hereditas*, 82, 212–226. https://doi.org/10.1038/sj.hdy.6884670

Wang, J., Santiago, E., & Caballero, A. (2016). Prediction and estimation of effective population size. *Hereditas*, 117, 193–206. https://doi.org/10.1038/hdy.2016.43

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121, 379–391. https://doi.org/10.1093/genetics/121.2.379

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, 7, 167–184. https://doi.org/10.1007/s10593-005-9100-y

Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 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. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Hereditas*, 117, 233–240. https://doi.org/10.1038/hdy.2016.60

Waples, R. S., Waples, R. K., & Ward, E. J. (2020). Pseudoreplication in genomics-scale datasets. *BioRxiv*. https://doi.org/10.1101/2020.11.12.380410

Waples, R. S., & Yokota, M. (2007). Temporal estimates of effective population size in species with overlapping generations. *Genetics*, 233, 219–233. https://doi.org/10.1534/genetics.106.065300

Weir, B. (1996). *Genetic data analysis II*. Sinauer Associates.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the population. *Evolutionary Ecology*, 3, 79–99. https://doi.org/10.1007/BF01234350

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–119. https://doi.org/10.1093/genetics/16.2.97

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Nadachowska-Brzyska, K., Dutoit, L., Smeds, L., Kardos, M., Gustafsson, L., & Ellegren, H. (2021). Genomic inference of contemporary effective population size in a large island population of collared flycatchers (*Ficedula albicollis*). *Molecular Ecology*, 00, 1–9. https://doi.org/10.1111/mec.16025