Received: 14 December 2017 | Revised: 21 August 2018 | Accepted: 28 August 2018

DOI: 10.1111/1755-0998.12941

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

NEOGEN: A tool to predict genetic effective population size (N_e) for species with generational overlap and to assist empirical N_e study design

Dean C. Blower[1,2] | Cynthia Riginos[1] | Jennifer R. Ovenden[2,3]

[1]School of Biological Sciences, The University of Queensland, St. Lucia, Queensland, Australia

[2]Molecular Fisheries Laboratory, The University of Queensland, St. Lucia, Queensland, Australia

[3]School of Biomedical Sciences, The University of Queensland, St. Lucia, Queensland, Australia

Correspondence
Dean Blower, The University of Queensland, St. Lucia, QLD, Australia.
Email: dean.blower@uqconnect.edu.au

Funding information
Fisheries Research and Development Corporation, Grant/Award Number: 2010/062

Abstract

Molecular genetic estimates of population effective size (N_e) lose accuracy and precision when insufficient numbers of samples or loci are used. Ideally, researchers would like to forecast the necessary power when *planning* their project. NEOGEN (genetic N_e for Overlapping Generations) enables estimates of precision and accuracy in advance of empirical investigation and allows exploration of the power available in different user-specified age-structured sampling schemes. NEOGEN provides a population simulation and genetic power analysis framework that simulates the demographics, genetic composition, and N_e, from species-specific life history, mortality, population size, and genetic priors. NEOGEN guides the user to establish a tractable sampling regime and to determine the numbers of samples and microsatellite or SNP loci required for accurate and precise genetic N_e estimates when sampling a natural population. NEOGEN is useful at multiple stages of a study's life cycle: when budgeting, as sampling and locus development progresses, and for corroboration when empirical N_e estimates are available. The underlying model is applicable to a wide variety of iteroparous species with overlapping generations (e.g., mammals, birds, reptiles, long-lived fishes). In this paper, we describe the NEOGEN model, detail the workflow for the point-and-click software, and explain the graphical results. We demonstrate the use of NEOGEN with empirical Australian east coast zebra shark (*Stegostoma fasciatum*) data. For researchers wishing to make accurate and precise genetic N_e estimates for overlapping generations species, NEOGEN facilitates planning for sample and locus acquisition, and with existing empirical genetic N_e estimates NEOGEN can corroborate population demographic and life history properties.

KEYWORDS
conservation genetics, genetic modelling, individual-based simulation, iteroparous, power analysis, sampling planning

1 | INTRODUCTION

In this era of rapid environmental change, evaluating the extinction vulnerability of populations is a high priority. To ensure effective population assessments for conservation management or sustainable exploitation, a clear understanding of a species' life history attributes, population demographics, and genetics is often necessary (Lacy, 1993; Lindenmayer, Lacy, Thomas, & Clark, 1993). Population assessment methods such as Capture–Mark–Recapture (CMR) tagging, species surveys, and biomass estimates often require extensive research, complicated modelling and expensive monitoring

(McAllister et al., 2009; Musick & Bonfil, 2005). Genetic population assessments incorporating effective population size (N_e) estimates can facilitate rapid management decisions regarding population viability with a relatively small investment in sampling and genetic marker development (Ovenden et al., 2016; Shafer et al., 2015; Waples & Do, 2010).

For the Wright-Fisher population model in which all individuals die after a single reproductive cycle (i.e., semelparous species with nonoverlapping generations), N_e equals the size of a population that constitutes a given quantity of genetic drift (Kimura & Crow, 1963; Wright, 1931). In such a population where individuals cannot breed with previous generations, N_e represents the number of successfully reproducing individuals and may equal the total population size (N). In contrast, for species that live long enough to mate multiple times (iteroparous species), and breed with conspecifics from older or younger age cohorts, overlapping generations occur. Overlapping generations contravene the assumptions of the Wright-Fisher model but typify a great number of species for which genetic estimates of population size would be highly useful, such as iteroparous species of conservation concern. For such species, the number of individuals successfully reproducing is a subset of total population size (N), and N_e is influenced by the number of reproductively mature individuals (N_c) and the extent of cross-cohort mating (Waples, Luikart, Faulkner, & Tallmon, 2013). N_e estimates also capture the future evolutionary resilience of a population (Frankham, 1995; Waples & Do, 2010). For reproductively isolated populations, reduced N_e may indicate loss of genetic adaptive potential and increased extinction risk (Frankham, 2005). At low N_e, deleterious mutations can accumulate and reproductive fitness may be lost (Frankham, 2005). These risks are inflated for species that have low to moderate fecundity, are long-lived or are slow to mature. Such characteristics can slow population growth and genetic adaptation, thereby increasing extinction vulnerability. Thus, as an index of population size and an indicator of evolutionary resilience, N_e is of significant utility to conservation managers (Frankham, 2005; Hare et al., 2011).

Despite great promise, multiple impediments to practical application of genetic population assessments remain and must be overcome, as recommended by Bernatchez et al. (2017). Practical estimation of empirical N_e faces some well-known hurdles. For accurate N_e estimation, the focal population needs to be assessed for continuity of population size, the extent of selection, mutation and gene flow, and for characteristics such as sex ratio and mating patterns. The effects of these conditions on N_e have been extensively investigated (Barker, 2011; Charlesworth, 2009; Luikart, Ryman, Tallmon, Schwartz, & Allendorf, 2010; Palstra & Ruzzante, 2008). Furthermore, there is uncertainty surrounding the interpretation of N_e estimates made for iteroparous species. Recent breakthroughs in this area have revealed that some iteroparous life history characteristics and population demographic factors exert substantial influence on N_e (Waples et al., 2013). Deterministic modelling (AgeNe; Waples, Do, & Chopelet 2011) demonstrated that both N_e and N_b (the effective number of breeders in one reproductive cycle), were strongly influenced by sexual maturation age, longevity (Waples et al., 2013), and mortality (Waples, 2016a). Waples, Antao, and Luikart (2014) further improved the practicality of effective size estimation by evaluating the relationship between demographically derived N_e and N_b (as enumerated by AgeNe), and estimation of genetic N_e using the linkage disequilibrium (LD) method (inbreeding effective size $N_{e.LD}$ estimated with the LDNe algorithm; Waples, 2006; Waples & Do, 2008, 2010). It was revealed that $N_{e.LD}$ estimates exhibit bias relative to the true N_e and N_b depending on the cohorts being sampled. $N_{e.LD}$ estimates from samples of a single age cohort were influenced equally by N_e and N_b, providing a general formula for predicting $N_{e.LD}$. However, combining samples from multiple cohorts (often necessary when opportunistically sampling rare or elusive species) exhibits a less predictable bias on $N_{e.LD}$ relative to the true N_e and N_b, highlighting the necessity for species-specific evaluations.

Integration of these findings into practical genetic population assessments remains constrained by two further obstacles. Firstly, most population genetics software currently lacks the complexity and flexibility to model iteroparous species. The previously cited breakthroughs linking life history to N_e were based on simulations involving small populations and a limited (albeit broad) range of key life history characteristics. However, to provide information that is suitably accurate for management of a specific population, customized and nuanced analyses are required to investigate how N_e is influenced by each species' complex life history properties, and each population's unique demographic and genetic circumstances (Waples & Do, 2010; Waples et al., 2013). The underlying design of most computational models, particularly genetic simulators for iteroparous species, significantly constrains life history refinement, genetic complexity, and the size of populations that may be simulated. Furthermore, user-friendly environments that cohesively link powerful simulations to key demographic and genetic analyses are essential to facilitate the practical use of genetic population assessment techniques. The second obstacle is obtaining reliable empirical genetic N_e estimates with limited resources. How to best divide resources between samples and genetic markers (loci) is a perennial issue. Locus availability is now significantly more tractable with the recent advent of rapid throughput genome sequencing, but sample acquisition remains resource intensive. The widely used temporal genetic N_e (variance effective size) estimation method requires multiple sets of samples from separate cohorts (Jorde, 2012; Waples & Yokota, 2007) which is often impractical for rare or elusive species that can only be sampled opportunistically. Avoiding this constraint, the $N_{e.LD}$ method requires just a single sample set and may require fewer samples and loci than the temporal method (Waples & Do, 2010). However, as with other methods of estimating N_e, the statistical validity of $N_{e.LD}$ is linked to three crucial factors that are often opaque when designing a study: (a) population size, (b) required sample number, and (c) the number of genetic markers (Waples, Larson, & Waples, 2016) and their allelic diversity (Hare et al., 2011; Palstra & Ruzzante, 2008; Waples & Do, 2010). When too few samples and genetic markers are assessed, $N_{e.LD}$ estimates suffer from low

accuracy and low precision, often having large or indeterminate upper confidence intervals (Waples & Do, 2010).

In summary, despite the benefits of rapid and inexpensive genetic population monitoring and development of the $N_{e.LD}$ method that is ideal for opportunistic sampling, considerable hurdles still deter researchers from routinely evaluating $N_{e.LD}$. These include the lack of a powerful, unified, and user-friendly analysis framework that predicts $N_{e.LD}$ given population-specific life history, demographics, and genetics. Additionally, there is limited ability to forecast the numbers of samples and loci required for accurate and precise $N_{e.LD}$ estimates. Quantities of informative genetic markers are expected to increase but cannot wholly replace low numbers of samples. Evaluating which proportions of samples and loci have sufficient analytical power to make accurate and precise $N_{e.LD}$ estimates is critical for efficient and informative genetic population assessments. These barriers to practical $N_{e.LD}$ estimation are addressed by our user-friendly point-and-click software package, NEOGEN (genetic N_e for Overlapping Generations; pronounced Neo-gen). NEOGEN software provides a straightforward demographic and genetic simulation framework for populations of species with overlapping generations. Integrating the key analyses, the straightforward workflow of NEOGEN gives a clear understanding of the sample numbers and genetic marker quantities that supply sufficient power to estimate $N_{e.LD}$ for a focal population. In this capacity, NEOGEN assists with experiment budgeting and resource efficiency as sampling and locus development progresses. In this paper, we describe the practical use of NEOGEN, the analysis workflow and interpretation of the results. We demonstrate NEOGEN functionality, especially in conjunction with pre-existing pilot data, using simulated and empirical data from zebra sharks (*Stegostoma fasciatum*).

2 | METHODS

A forward-time, individual-based population simulator SIMUPOP (Peng & Kimmel, 2005) supplies the underlying functions of NEOGEN's overlapping generations model. In NEOGEN's model, each individual is assigned an age, sex, reproductive status, and a compact genome. Simulated individuals are born, mature, and can mate multiple times over their lifespan, passing their genes to offspring before dying. Opposite sex individuals of any mature age cohort may mate, thereby simulating mating between overlapping generations. Mating is performed annually, whereby mature adults are paired at random, with each pair producing a litter size with a mean and variance defined by the researcher. Natural mortality proportions specified for each age (by sex) modify the size of each cohort prior to mating. The N_e estimation methods employed by NEOGEN assume population size is constant, and NEOGEN maintains constant population size by increasing mating to provide sufficient offspring to compensate for mortality following Waples et al. (2011) and Waples et al. (2014). Furthermore, Waples (2006) suggests that changing population size modestly affects $N_{e.LD}$ estimates and only for a few generations (Russell & Fewster, 2009). Each modelled individual carries a multilocus genotype from which alleles are passed to offspring gametes at mating according to Mendelian inheritance. Loci are unlinked; however, in a population of finite size that is closed to migration, offspring genotypes acquire statistical linkage disequilibrium (LD; alleles that associate nonrandomly) proportional to the breeding population size. This LD is quantified by $N_{e.LD}$ estimates using the LDNe algorithm of NEESTIMATOR v2.01 (Do et al., 2014).

NEOGEN population simulations are run for an initial burn-in period of two lifespans (which is alterable by the user), whilst the demographics equilibrate in response to mortality. During burn-in, the genetic linkage disequilibrium also stabilizes; LD can reach equilibrium in as few as two generations (Waples et al., 2014). Post burn-in, continued annual matings and mortality produce the demographics and population genetic data of the focal population. After the final mating, the simulated population is subsampled as designated by the user for a range of sample sizes taken from specified cohorts and a range of genetic marker quantities. Then, for each combination of sample size and locus quantity, the $N_{e.LD}$ is estimated with confidence intervals.

SIMUPOP, the underlying simulation engine for NEOGEN, can simulate millions of individuals each with thousands of loci, assuming sufficient memory (RAM) to hold large populations and processing speed (CPU) to perform operations on individuals. Default maximums of population size (N = 10,000), number of offspring (100 per litter), maximum age (200 years), and number of loci (6,000 loci per individual), can easily be modified in the NEOGEN settings file. The lower limits of minimum and maximum age of reproduction are 1 and 2, respectively, to ensure generational overlap. The maximum mating age may be specified separately to the maximum age to simulate a period of reproductive senescence for species that cease reproduction prior to dying from old age.

NEOGEN's life history flexibility has been tested with iteroparous species such as zebra shark (*Stegastoma fasciatum*), Atlantic cod (*Gadhus morhua*), and bottlenose dolphin (*Tursiops truncates*). The NEOGEN model was validated against theoretical expectations for demographic and genetic compositions. The AgeNe algorithm of Waples et al. (2011) mathematically evaluates demographic properties such as population size, age-cohort sizes, demographic N_e, and variance in reproductive success, given species-specific life history, reproduction, and mortality metrics for each cohort. NEOGEN's simulated demographic results were compared with expectations from the AgeNe algorithm, using similar starting parameters. Genotypes simulated by NEOGEN were validated against accepted population genetic expectations that arise from the Wright-Fisher theoretical population, such as Hardy–Weinberg equilibrium (HWE), inbreeding (F_{is}), and post-mating changes to population allele frequencies and heterozygosity (H_e). Linkage disequilibrium and $N_{e.LD}$ were validated with a nonoverlapping generations variant of the NEOGEN model under two sampling strategies (S = 50 and S = 200) for a range of minimum allele frequency thresholds (P_{crit} values 0.1 – 0.01). Concluding NEOGEN validation, NEOGEN's results for bottlenose dolphin and Atlantic cod were compared to Waples et al. (2014) population modelling of the same species. Both investigations generated genetic $N_{e.LD}$ (LDNe method) estimates and demographic N_e and N_b (AgeNe

method) estimates from simulated populations. We expected concordance between like metrics.

When planning a study, NEOGEN's graphical user interface (GUI) guides the researcher with parameter entry and running the underlying NEOGEN model. Initially, NEOGEN predicts the demographics and multilocus genotypes for a focal population of a particular species. Then NEOGEN can perform a power analysis to determine optimal quantities of samples and microsatellite or SNP loci that provide accurate and precise $N_{e.LD}$ estimates of the focal population. The key results of this process are visually presented in three plots and detailed results are available in spreadsheets. The NEOGEN analytical framework incorporates the key analyses into two major phases (Figure 1). Phase one involves defining a population-specific "Scenario" that directs the simulated evolution of an individual-based population (independently replicated multiple times), resulting in a population "Demographic Profile." The researcher initially defines a Scenario based on fundamental life history properties of the target species, such as maturation age, longevity, age-specific mortality (Table 1), and population genetic parameters such as the number of loci and a range of allele frequency distributions (simulated or user-supplied). Next, the simulated individuals of the population are subjected to repeated cycles (e.g., years) in which they age annually, successively experiencing sexual maturation, mating, and senescence, whilst subject to age- and sex-specific mortality. With each annual mating event, offspring inherit a simulated multilocus genotype from randomly selected parents, and linkage disequilibrium develops as a consequence of the finite population size. Phase one concludes with multiple independently replicated populations, each comprising individuals of all ages. Scenario population replicates are summarized by the population Demographic Profile (Figure 1, phase I and Figure 2), which plots the mortality, births, and individuals remaining in each age cohort (averaging values for all population replicates). The Demographic Profile gives the researcher the opportunity to inspect the demographics for plausibility, and to identify which cohorts to subsample for the subsequent Sampling Strategy. In phase two (Figure 1, phase II), the in-silico individuals of the previously simulated

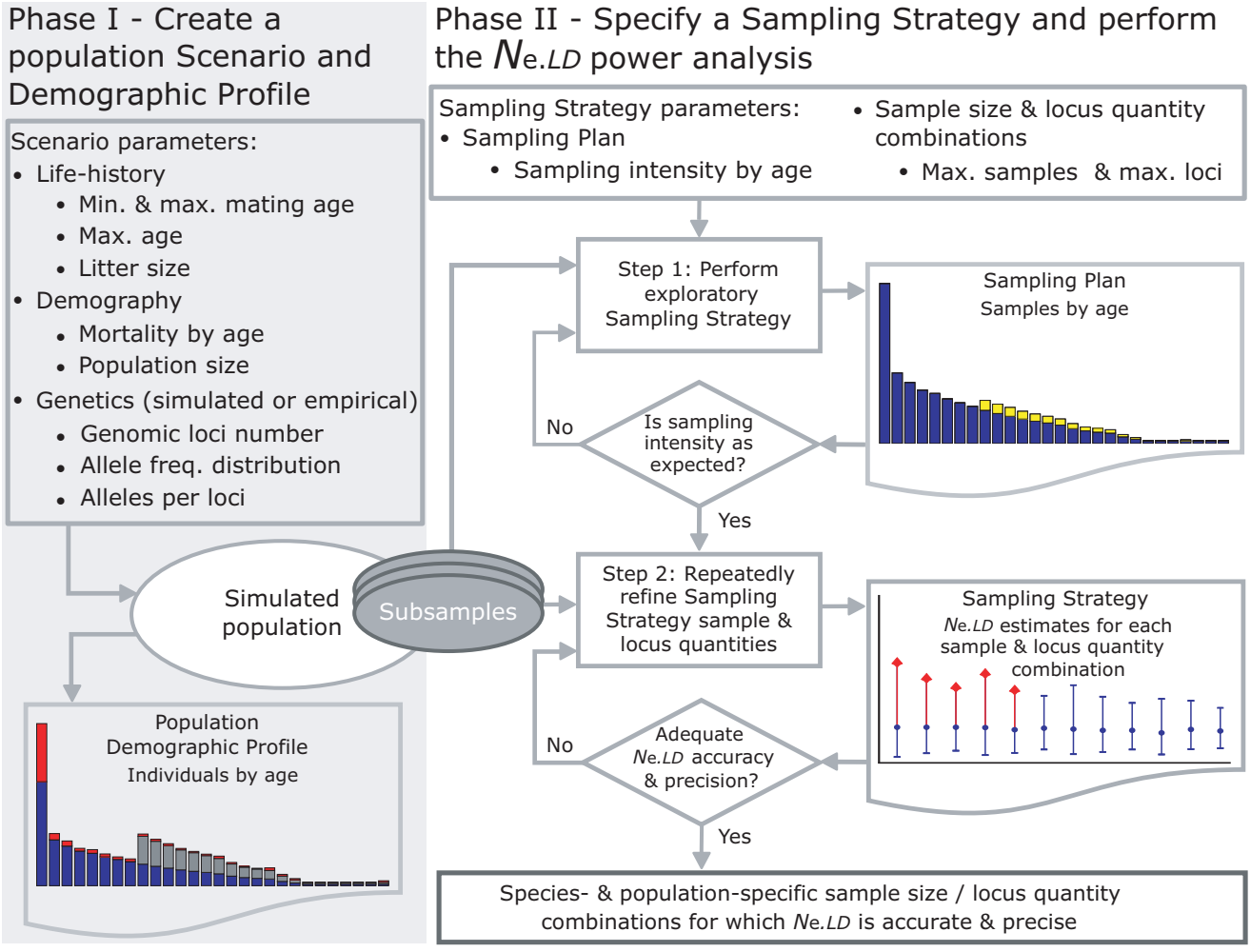

FIGURE 1 The two phases of the NEOGEN workflow: (I) create a species- and population-specific "Scenario" population; (II) subsample that population with a "Sampling Strategy" to identify sample and locus quantities that produce accurate and precise $N_{e.LD}$ estimates. See text for explanation, and Figures 2 and 3 for explanation of the plots [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Parameters required to run the Scenario and Sampling Strategy phases in NEOGEN

Category	Prior name	Description
Phase I —Scenario creation		
Life history	Maximum age	Longevity; the maximum age at which animals die
	Maximum mating age	The age of reproductive senescence; the age at which animals can no longer reproduce
	Minimum mating age	Maturation age; the age at which animals become reproductively capable
	Offspring per mating	Litter size per mating pair; the offspring number per parent (or overall offspring distribution) produced at each mating event
Demographic	Population size	An approximate total population size estimate (*N*), that is, estimate of all individuals comprising the population
	Natural mortality rates	Estimates of the probability of individual mortality by age and sex
Genetic	Number of loci per individual[a]	Number of genetic loci per individual available for interrogation; size of each individual's simulated genome
	Alleles per loci[a]	Locus allelic diversity across the simulated genome
	Allele frequencies[a]	The probability of encountering rare or common alleles per locus
Simulation length	Simulation burn-in length	Annual matings required to equilibrate demography and genetics
	Simulation temporal evolution length	Annual matings required for data gathering
	Number of replicate simulations	Number of populations independently generated with identical Scenario parameters.
Phase II —Sampling strategy and sampling plan creation		
Sampling strategy	Maximum samples	Maximum number of samples that are feasible to obtain.
	Sample increment	Factor that determines the sample sizes to evaluate up to the maximum
	Maximum loci	Maximum number of loci that are feasible to obtain.
	Loci increment	Factor that determines the loci numbers to evaluate up to the maximum
Sampling plan	Sampling Plan	Relative proportions to sample from each age cohort
	LDNe P_{crit}	LDNe method parameter; minimum allele frequency below which alleles are excluded
Simulation length	Number of $N_{e.LD}$ replicates	The number of times a sample/locus combination is independently subsampled and evaluated for $N_{e.LD}$

Notes. NEOGEN phase I Scenario parameters for simulated populations that reflect the life history, demography, and genetics of the target species. Phase II Sampling Strategy parameters that specify the Sampling Plan, sample sizes, and loci quantities required to estimate $N_{e.LD}$ from the simulated populations in order to evaluate $N_{e.LD}$ accuracy and precision for each sample size and locus quantity combination.
[a]May also be supplied by importing a file containing empirical genetic data.

population replicates are repeatedly sub-sampled for combinations of user-specified sample sizes and genetic marker quantities. This sampling is defined by a "Sampling Strategy" (Figure 1, phase II, steps 1 & 2), which incorporates a "Sampling Plan" (Figure 1, phase II, step 1) specifying the proportion of individuals to be sampled from each cohort. For each sample size and locus quantity combination defined by the Sampling Strategy, average $N_{e.LD}$ estimates (Do et al., 2014; Waples & Do, 2008) are generated and plotted with confidence intervals alongside an accuracy guideline (Figure 3). To provide a statistically robust indication of the power of any particular sample/locus quantity combination, $N_{e.LD}$ is repeatedly estimated with that combination across all replicate populations and averaged (using the harmonic mean) to a single $N_{e.LD}$ estimate. The estimation power of sample/locus quantity combinations is judged by the accuracy and precision of the $N_{e.LD}$ estimates they generate. The accuracy of each $N_{e.LD}$ estimate is gauged by comparison to the $N_{e.LD}$ obtained from exhaustive sampling of simulated genotypes (see "accuracy guideline," Figure 3), and precision is judged by the width of confidence intervals. Thus, the plotted Sampling Strategy results highlight the sample size and locus quantity combinations that are expected to produce accurate and precise empirical $N_{e.LD}$ estimates for the target population.

As highlighted above, the practical use of NEOGEN starts with gathering parameters specific to the species and population of interest (Figure 1, phase I; Table 1). Some information may be known or inferred by the researcher through previous research, and some may be gathered from pilot sampling. The primary assumption of the NEOGEN analysis is that the focal population is a discrete genetic deme. NEOGEN also requires a coarse estimate of the total population size (*N*). This estimate may be imprecise as the researcher can use NEOGEN to test a range of probable population sizes and compare the plausibility of the resulting demographics and $N_{e.LD}$ with independent estimates. Ideally, population size estimates provided to NEOGEN will be based upon independent research, such as capture–mark–recapture (CMR) studies. Researchers must also consider the cohorts upon which a population estimate is based. For example, CMR estimates

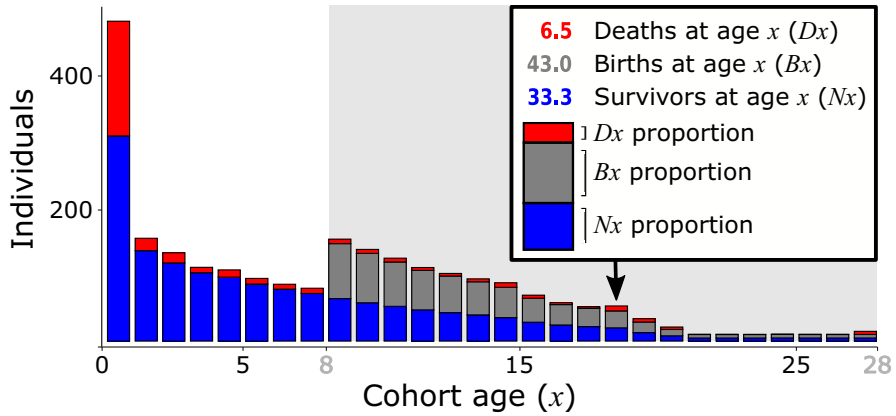

FIGURE 2 A NEOGEN population Demographic Profile resulting from many rounds of mating, directed by priors of life-history, mortality, and population size, and averaged over multiple independently replicated populations. This plot shows the population demographics for a NEOGEN *Stegostoma fasciatum* population (N = 1,400) simulated over four lifespans (112 years). The resulting adult population size (N_c = 459) approximates the empirical CMR N_c estimate of 458 (Dudgeon et al., 2008). The light grey shaded plot area highlights the sexually mature age cohorts, delineated by the minimum mating age (8 years) and the maximum mating age (28 years), shown with grey numbers. Column totals (only shown for age 18) indicate the number of deaths (Dx), births (Bx), and surviving individuals (Nx) in red (grey), grey (white), and blue (black) respectively for each age cohort (x), averaged across 30 population replicates [Colour figure can be viewed at wileyonlinelibrary.com]

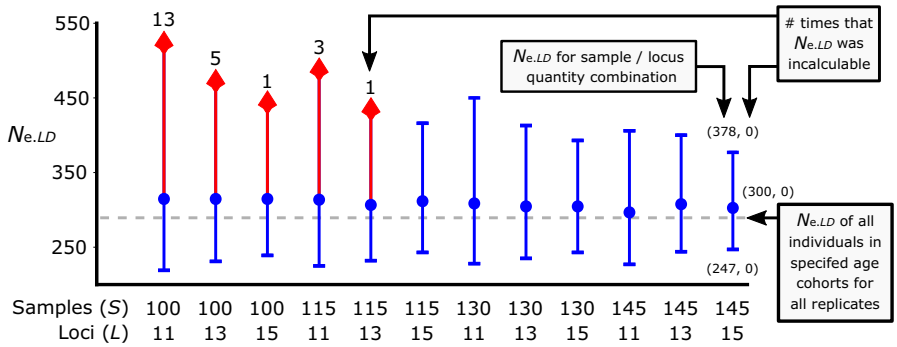

FIGURE 3 This NEOGEN Sampling Strategy plot allows the user to evaluate the power of a sampling strategy against an accuracy guideline and confidence intervals. The x-axis presents user-specified sample/locus quantity combinations. The y-axis shows the corresponding $N_{e.LD}$ (blue [black] circles: harmonic mean averaged over the total replicates; total replicates = population replicates × $N_{e.LD}$ replicates) with lower 5% and upper 95% confidence intervals (CI) for each combination. All $N_{e.LD}$ estimates are represented by two values enclosed by parentheses (shown here just for the 145/15 sample/loci combination). The first value indicates the relevant estimate, and the second, the number of times the estimate was incalculable (i.e., was negative or approached infinity) over all replicates. Incalculable CI's are displayed with a red (grey) arrow cap with the number of incalculable estimates, and CI's with adequate power are in blue (black) with a flat cap. Each combination's $N_{e.LD}$ precision is assessed by the width of the CI's. The accuracy of $N_{e.LD}$ point estimates can be judged by similarity to the dashed shaded "accuracy guideline" which equals the $N_{e.LD}$ estimated from all loci and all individuals from the same age cohorts as sampled for the sample/locus combinations. This plot displays the $N_{e.LD}$ power of 100–145 samples and 11–15 simulated loci sampled from a NEOGEN simulated *Stegostoma fasciatum* population (N = 1,400, N_c = 458, 30 population replicates, 50 $N_{e.LD}$ replicates per population replicate, P_{crit} = 0.02). Here, NEOGEN recommends a minimum of 115 samples with 15 loci or 130 samples with 11 or more loci for adequate $N_{e.LD}$ accuracy and precision. Inadequate power to estimate $N_{e.LD}$ with precision occurred with 115 or fewer samples combined with 13 or less loci. Estimate accuracy increases relatively consistently with more samples and loci. For example, the sample/loci combination 100/11 is 93% accurate relative to the "accuracy guideline" ($N_{e.LD}$ = 290, 95% CI: 286–293) compared with 96% accuracy for the 130/13 combination and 97% accuracy for the 145/15 combination [Colour figure can be viewed at wileyonlinelibrary.com]

resulting from tagging adults may only estimate the N_c, which excludes juveniles, rather than N that includes all individuals. NEOGEN produces simultaneous estimates of N_c and N, allowing the researcher to extrapolate N from an empirical estimate of N_c.

After population simulation, the researcher creates a Sampling Strategy (Figure 1, phase II) to evaluate the accuracy and precision of N_e estimates, determining the number of individuals that could be feasibly sampled, and a realistic quantity of loci that could be acquired given constraints (time, money, effort, etc.). In executing a Sampling Strategy, the replicate Scenario populations are subsampled with the proportions specified by the Sampling Plan (Figure 1, phase II, step 1). NEOGEN allows any proportion or combination of cohorts to be sampled, however when defining the Sampling Plan, the researcher should ideally consider existing field research (e.g., pilot

sampling) and specify the proportions of each cohort likely to be sampled in the target population. For instance, if trial sampling indicates that particular age cohorts or a life stage (e.g., adults) are most easily sampled, the researcher should specify a Sampling Plan that samples those particular cohorts.

The researcher then enacts a power analysis (Figure 1, phase II, step 2) to identify which sample size and locus quantity combinations are sufficiently economical and produce adequate $N_{e.LD}$ accuracy and precision to meet the researcher's aims. The Sampling Strategy plot (Figure 3) shows average $N_{e.LD}$ estimates for specified sample size and locus quantity combinations along with indications of their statistical properties. Inaccurate or imprecise sample/locus combinations can then be identified. The researcher will either eliminate those combinations or increase the sample or locus quantities for the subsequent Sampling Strategy. Combinations which produce acceptable $N_{e.LD}$ estimates but could be more economical in sample and locus quantities can be refined. This power analysis cycle is repeated until consecutive sample/locus combinations give $N_{e.LD}$ estimates of acceptable accuracy, precision, and economy. These NEOGEN recommendations can then be used as a guide when budgeting, planning, and executing empirical sampling and locus development. If after repeated cycles it is apparent that the desired number of samples and loci will not produce adequately accurate and precise $N_{e.LD}$ estimates for the population in question, the researcher may need to commit more resources to sample or locus acquisition.

When planning a project with NEOGEN, the researcher would gather life history, demographic, and genetic parameters (Table 1) from a variety of empirical sources and follow the preceding workflow description. To illustrate this process with a real-life example, we present results for zebra shark (*Stegostoma fasciatum*) from south-eastern Queensland. Empirical investigation of this reproductively isolated *S. fasciatum* population provided a CMR N_c estimate of the adult population size (CMR N_c = 458, 95% CI: 298–618; Dudgeon, Noad & Lanyon, 2008). Additionally, field collecting experience indicated the numbers and age classes that could be captured for genetic tissue sampling. Using these key pieces of information, we used NEOGEN to determine a hypothetical future Sampling Plan to maximize accuracy and precision. Empirical *S. fasciatum* life history, demographic, and genetic parameters were entered to generate a NEOGEN Scenario population (comprising 30 independent replicates) for which the NEOGEN N_c approximated the CMR N_c. The resulting Demographic Profile (Figure 2) enabled specification of a Sampling Strategy and Sampling Plan. Following the empirical genetic study where only adult *S. fasciatum* were sampled, the Sampling Plan sampled all adult cohorts, specifically excluding subadults. Executing this Sampling Strategy against the simulated populations allowed evaluation of the numbers of samples and loci that yielded precise $N_{e.LD}$ estimates that were accurate relative to the simulated "accuracy guideline" (i.e., $N_{e.LD}$ obtained from sampling of all simulated individuals in the specified cohorts).

In addition to using *S. fasciatum* as a demonstration of NEOGEN's functionality, we verified the consistency of NEOGEN's outcomes to *S. fasciatum* empirical results. When project planning with NEOGEN, a precise empirical $N_{e.LD}$ estimate is typically unavailable but is the ultimate goal of NEOGEN's Sampling Strategy. However, for this *S. fasciatum* population, Dudgeon and Ovenden (2015) empirically estimated $N_{e.LD}$ from 113 adults using 13 microsatellite loci. Having empirical $N_{e.LD}$ and CMR N_c estimates for the same cohorts of this zebra shark population provided three comparisons to the NEOGEN simulated *S. fasciatum* populations: (a) to see if NEOGEN would predict a similar $N_{e.LD}$ when the simulated N_c matched the empirical CMR N_c, (b) to see if NEOGEN would predict an N_c that approximated the CMR N_c when the simulated $N_{e.LD}$ matched the empirical $N_{e.LD}$, and (c) for both cases, to verify if sample sizes and locus quantities predicted to give accurate and precise $N_{e.LD}$ estimates by NEOGEN would be similar to those used empirically. For each of these situations, two sources of genetic information were considered: (a) simulated genetic data with an alleles per locus distribution approximating *S. fasciatum* empirical genetic data (binomial distribution with mean = 15.0 and standard deviation = 10.0; Dudgeon & Ovenden, 2015), and (b) empirical population allele frequencies imported into the NEOGEN simulations. For practical purposes, the *S. fasciatum* empirical $N_{e.LD}$ was taken to be accurate and the associated sample and locus sizes adequate on several criteria: (a) closeness of the $N_{e.LD}$ to the CMR estimate of N_c (Dudgeon & Ovenden, 2015), a trend supported in sharks by several other studies (Andreotti et al., 2016; Blower, Pandolfi, Bruce, Gomez-Cabrera, & Ovenden, 2012; Portnoy, McDowell, McCandless, Musick, & Graves, 2009), (b) the stability of the $N_{e.LD}$ when re-estimated with samples or loci removed (Dudgeon & Ovenden, 2015), and c) the tight confidence intervals suggesting sufficient power for accurate $N_{e.LD}$ estimates. These evaluations complement the model validation and illustrate the consistency of NEOGEN projections against empirical evaluation of $N_{e.LD}$ for a wild species population.

3 | RESULTS

The NEOGEN model validation process produced simulated populations that conformed to population genetics theory, AgeNe demographic results, and $N_{e.LD}$ expectations. The demographic proportions of the NEOGEN individual-based simulations matched the AgeNe demographic results. Only minor discrepancies were observed due to simulation stochasticity and rounding error. NEOGEN genetic results met expectations for HWE, F_{is}, change of population allele frequencies, and H_e after mating. Evaluation of $N_{e.LD}$ for a discrete generations variant of the NEOGEN model (Figure 4) differed little from Waples and Do (2008). Additionally, for simulations based on bottlenose dolphin and Atlantic cod (Figure 5), a very high average similarity was observed between NEOGEN and the simulations of Waples et al. (2014). NEOGEN $N_{e.LD}$ concordance with Waples $N_{e.LD}$ estimates was higher for bottlenose dolphin at 98.1% than Atlantic cod at 95.4%. This difference was attributed to the greater suitability of the NEOGEN model to the life history properties of the bottlenose dolphin.

The *S. fasciatum* example illustrates how pilot information can be used to guide sampling design using NEOGEN. Empirical estimates of *S. fasciatum* population size enabled three points of comparison with

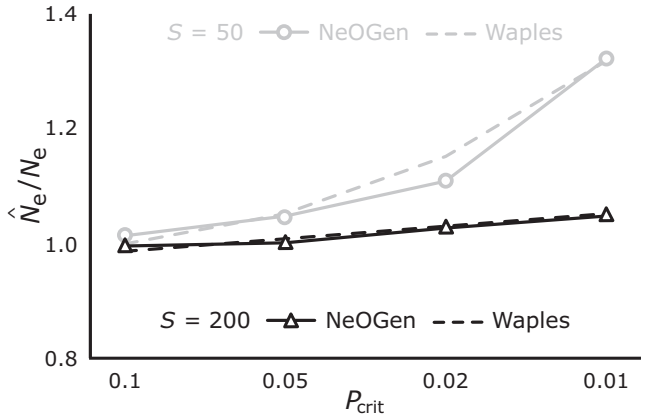

FIGURE 4 NEOGEN recovers similar results to Waples and Do (2008) using a similar discrete generations model. For both models, the ratio of $\hat{N}_e$ (N_e estimated with LDNe method, $N_{e.LD}$) to true N_e (500) are presented for two sample sizes, S = 50 and S = 200, and P_{crit} values 0.1–0.01. Allele frequencies below the P_{crit} are excluded as rare alleles can bias LDNe estimates. Waples and Do (2008) data was extracted from their Figure 1 top panel

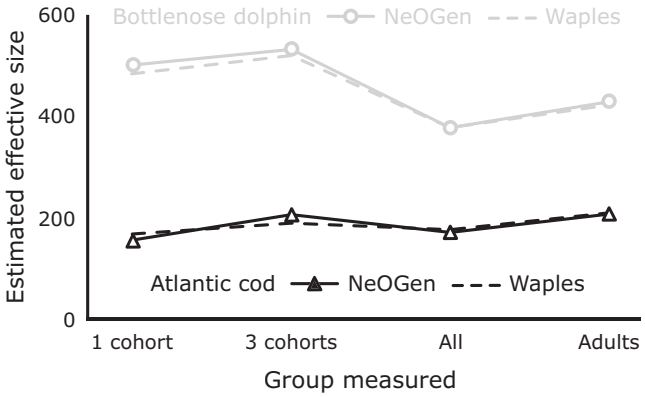

FIGURE 5 NEOGEN's overlapping generations model recovers similar results to Waples et al. (2014), in which simulated genetic data provided estimates of effective size (N_e estimated with LDNe method, $N_{e.LD}$) for populations created with AgeNe life-history parameters. Results from two test species are shown: bottlenose dolphin (*Tursiops truncatus*) and Atlantic cod (*Gadhus morhua*). Each estimate comprised one hundred samples (S = 100) and $N_{e.LD}$ results are for P_{crit} = 0.05. Waples et al. (2014) data was extracted from their supporting information Figure S5 top panel, Atlantic cod, and Figure S6 bottom panel, bottlenose dolphin

NEOGEN's results. The first comparison required matching NEOGEN's simulated adult population size (N_c) to the *S. fasciatum* CMR N_c. Various estimates of N (total number of individuals including sub-adults) were simulated to ascertain that N of 1,400 (Figure 2) correlated with an adult number (N_c = 459), closely approximating the CMR adult population size estimate (CMR N_c = 458; Dudgeon et al., 2008). Initial Sampling Strategies using this scenario with simulated loci suggested exhaustive sampling of the target age cohorts gave an $N_{e.LD}$ estimate of 290 (95% CI: 286–293; see "accuracy guideline" Figure 3). A Sampling Strategy with samples sizes S = 100, 115, 130, and 145 and loci quantities L = 11, 13, and 15 produced $N_{e.LD}$ estimates with accuracy of 93% or greater (within 7% of the accuracy guideline) and reasonably precise CI's (upper CI's within 40% of the accuracy guideline) with 115 samples and 15 loci, or where a minimum of 130 samples were combined with 11 or more loci (Figure 3). Further Sampling Strategies favouring fewer samples indicated 120 samples with 13 loci, 110 samples with 14 loci, 100 samples with 15 loci, or 90 samples with 18 loci (Table 2, Test 1, Simulated loci) would be sufficient to estimate $N_{e.LD}$ with 93% average accuracy and relatively precise CI's. In summary, with simulated loci and when the NEOGEN $N_c \approx$ CMR N_c, NEOGEN simulated $N_{e.LD}$ was 23% smaller than the empirically derived *S. fasciatum* $N_{e.LD}$, but within the empirical confidence limits ($N_{e.LD}$ = 377, 95% CI: 274–584, P_{crit}: 0.02; Dudgeon & Ovenden, 2015). NEOGEN's recommendations for sample and locus quantities reasonably approximated the 113 samples and 13 loci used for the empirical study.

TABLE 2 NEOGEN results compared to the *Stegostoma fasciatum* empirical results obtained by Dudgeon and Ovenden (2015) with 113 samples and 13 loci

	NEOGEN recommended sample/locus quantity combinations			
Test situation	Samples	Simulated Loci	Samples	Empirical Loci
Test 1: NEOGEN $N_c \approx$ *S. fasciatum* empirical CMR N_c	90	18	130	13
	100	15	135	13
	110	14	145	12
	120	13	155	11
Test 2: NEOGEN $N_{e.LD} \approx$ *S. fasciatum* empirical $N_{e.LD}$	95	18	175	13
	125	15	185	12
	135	13	205	11
	160	11	220	11

Notes. For both simulated genetic loci and empirically derived genetic loci, NEOGEN sample and locus quantity recommendations for accurate and precise $N_{e.LD}$ estimates (P_{crit}: 0.02) are presented for two test situations: Test 1: NEOGEN simulated *S. fasciatum* population has an adult population size (N = 1,400; $N_c \approx$ 459) which approximates the *S. fasciatum* CMR N_c (458, 95% CI: 298–618; Dudgeon et al., 2008); Test 2: the simulated population (N = 1,800; $N_c \approx$ 591) has an $N_{e.LD}$ (370, 95% CI: 260–615) approximating the empirically derived *S. fasciatum* $N_{e.LD}$ (377, 95% CI: 274–584; Dudgeon and Ovenden 2015).

Second, continuing to use the NEOGEN $N_c \approx$ CMR N_c Scenario but incorporating the *S. fasciatum* empirical loci resulted in an "accuracy guideline" $N_{e.LD}$ estimate of 275 (95% CI: 241–317; Figure 6). Sampling Strategies limited to the 13 imported loci (L = 11, 12, 13) were evaluated. These analyses recommended 130 or more samples with 13 empirical loci, 145 or more samples with 12 empirical loci (Figure 6), or 155 samples with 11 empirical loci (Table 2, Test 1, Empirical loci), each combination providing an average 98% $N_{e.LD}$ estimate accuracy. Using empirical loci for the NEOGEN $N_c \approx$ CMR N_c scenario underestimated the empirical $N_{e.LD}$ by 27%, and with 13 loci slightly overestimated the required number of samples by 17 samples.

Third, the comparison of NEOGEN to empirical results required simulated populations for which the $N_{e.LD}$ approximated the empirically

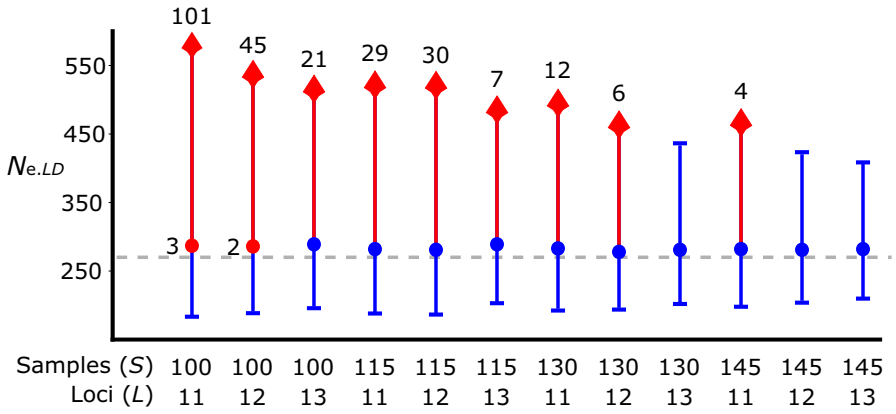

FIGURE 6 Incorporating *Stegostoma fasciatum* empirical genetic loci from Dudgeon and Ovenden (2015), this Sampling Strategy plot shows $N_{e.LD}$ accuracy and precision of sample/loci combinations sampled from a NEOGEN *S. fasciatum* population (N = 1,400, N_c = 458, 30 population replicates, 50 $N_{e.LD}$ replicates per population replicate, P_{crit} = 0.02). Mean $N_{e.LD}$ evaluated for 100–145 samples and 11–13 loci showed $N_{e.LD}$ estimates of adequate accuracy and precision for a minimum of 130 samples with 13 loci, or 145 samples with 12–13 loci. With 100 samples and 12 or less loci, combinations exhibit very low power with the $N_{e.LD}$ point estimate showing reduced ability to compute $N_{e.LD}$ (red [grey] circles with the number of incalculable estimates). Below 130 samples and 13 loci, estimates exhibit low precision with wide upper CI's and multiple incalculable estimates (red [grey] arrow capped CI's with a number of incalculable estimates). $N_{e.LD}$ point estimate accuracy can be gauged by percentage equivalence to the "accuracy guideline" (dashed shaded line) representing $N_{e.LD}$ ($N_{e.LD}$ = 275, 95% CI: 241–317) estimated with all loci and all individuals from the same cohorts as sampled for the sample/locus quantity combinations. Accuracy improved with increased samples and loci and was considered acceptable for all combinations. For example, the 100/11 combination had the fewest samples and loci despite which the $N_{e.LD}$ point estimate, when calculable, averaged 96% accuracy, and was incalculable for just 3 of 1,500 replicates [Colour figure can be viewed at wileyonlinelibrary.com]

derived *S. fasciatum* $N_{e.LD}$ estimate. For these populations, the NEOGEN N_c was expected to approximate the empirical CMR N_c. NEOGEN's simulated populations had N of 1,800 and N_c of 591 which was 133 individuals larger than the empirical *S. fasciatum* CMR N_c estimate of 458, but within the empirical confidence intervals (95% CI: 298–618). Sampling Strategy power analysis of these populations with simulated genetics recommended that 135 or more samples combined with 13 or more loci would result in 95% accurate $N_{e.LD}$ estimates (Table 2, Test 2, Simulated loci). With 13 loci, NEOGEN suggested 20% more samples than used by the empirical study. NEOGEN populations based on empirical genetics suggested sample sizes of 175 or more with 13 loci (Table 2, Test 2, Empirical loci). This overestimated the samples required for the empirical study by 55% but predicted 99% $N_{e.LD}$ estimate accuracy.

Overall, NEOGEN somewhat underestimated the $N_{e.LD}$ and N_c relative to the empirical $N_{e.LD}$ and empirical CMR N_c, and somewhat overestimated the samples required for accurate and precise $N_{e.LD}$ estimates. Despite this, the recommendations of NEOGEN from both simulated and empirical allele frequencies would have resulted in accurate and precise $N_{e.LD}$ estimates for this *S. fasciatum* population. To indicate the computing performance required for these examples, the running of a Scenario and a Sampling Strategy took ~2 hrs each on a desktop computer: MS Windows 10, Intel i7 CPU, 16 GB RAM.

4 | DISCUSSION

NEOGEN is a practical tool for guiding genetic linkage disequilibrium N_e ($N_{e.LD}$) studies of iteroparous species and is useful throughout a study's lifecycle. When planning experiments or as locus development and sampling progress, NEOGEN's genetic power analyses can use simulated or preliminary empirical genetic data to forecast the quantity of samples and loci required to make accurate and precise empirical $N_{e.LD}$ estimates. Additionally, when empirically derived $N_{e.LD}$ estimates are available, NEOGEN can be used for sensitivity analyses of $N_{e.LD}$ estimates, population size, life history characteristics, age-specific mortality and other properties. NEOGEN differs from currently available tools by providing a user-friendly and practical analysis framework, customizable to a focal population, which links powerful iteroparous species simulations to $N_{e.LD}$ estimates. The standalone software tools, LDNe and AgeNe, have had their outputs correlated with generalized formulae (Waples et al., 2014). However, NEOGEN avoids generalizing by quantifying the $N_{e.LD}$ arising from the explicit properties of the focal population. NEOGEN complements other genetic N_e estimation tools developed for iteroparous species, such as GONe (Coombs, Letcher, & Nislow, 2012) and *NEff* (Grimm, Gruber, Hoehn, Enders, & Henle, 2016). GONe links AgeNe-like demographic parameters to genetic N_e by the temporal (or allele frequency variance) N_e method. *NEff* is an individual-based iteroparous species simulator dependent on the R core environment, which uses a demographic formula to estimate inbreeding genetic N_e and as such may provide useful comparisons to NEOGEN's $N_{e.LD}$ estimates.

NEOGEN yielded results commensurate with Waples' AgeNe and $N_{e.LD}$ results for bottlenose dolphin and Atlantic cod (Figure 5). Bottlenose dolphin have few offspring, slow maturation, and long lifespan, life history properties ideal for NEOGEN simulations. Results for this species were highly concordant across methods. Atlantic cod

results were slightly less concordant; NEOGEN assumes constant fecundity across cohorts, yet the fecundity of cod can increase with age, resulting in higher reproductive success for older animals, thus influencing population N_e (Waples, 2016b). Fecundity changes with age are far less pronounced for dolphins and therefore better suited to NEOGEN's current capabilities. Empirical estimation of population size by CMR and $N_{e.LD}$ for an *S. fasciatum* shark population (Dudgeon & Ovenden, 2015) also demonstrated the effectiveness of the NEOGEN model. Using a minimum mating age of 8 years, NEOGEN simulated an adult population whose size (N_c) and $N_{e.LD}$ approximated empirical estimates from the wild population. Although NEOGEN somewhat overestimated the *S. fasciatum* sample quantities required, cautious but realistic recommendations that overstate the necessary samples and loci are desirable, as they increase the likelihood of obtaining valid empirical $N_{e.LD}$ estimates.

NEOGEN integrates two predictive models (demographic and genetic) to produce N_e predictions for the planning of projects and corroboration of existing studies. However, these two components make predictions that also have independent uses. Demographic changes can be assessed by cohort, identifying cohorts that are most resilient to depletion, allowing for strategies that maximize genetic diversity and population sizes under age-specific pressures, such as disease or poaching, or when developing culling or sustainable harvest strategies. Aside from sampling planning, the genetic N_e prediction component of NEOGEN allows comparison of empirical $N_{e.LD}$ estimates with a range of life history and demographic scenarios, allowing testing of hypotheses such as the sexual maturation age of a species, or the population size. Additionally, genetic diversity baselines may be predicted and compared with empirical estimates, facilitating genetic health monitoring of a population. NEOGEN also provides opportunities to generate N_e/N ratios for species, a frequently intractable measure that is a useful index of population productivity and vulnerability (Frankham, 1995). When estimating $N_{e.LD}$, NEOGEN allows complete freedom to specify a cohort sampling plan. This is important as it allows the estimation of N_b, the number of effective breeders per mating event (annually or otherwise), a parameter that is highly relevant to designing conservation strategies (Bernos & Fraser, 2016). NEOGEN N_b estimates can be generated by sampling a single age cohort. It is thus possible to explore similarity between single-cohort estimates of N_b and effective size estimated from sampling several cohorts. Furthermore, simulated life tables and AgeNe (Waples et al., 2014) algorithm output produced by NEOGEN can be used to explore the relationship between the genetic (NEOGEN) and demographic (AgeNe) estimates of effective size for individual and multiple cohorts.

NEOGEN analyses are subject to cautions and caveats. Predictions of population demographics rely upon accurate maturation age, maximum age, and mortality parameters. These factors determine cohort sizes and strongly influence N_e (Waples et al., 2013; Waples, 2016a). Inaccuracy in these factors may lead to erroneous population productivity and N_e estimates, which could invalidate population vulnerability assessment. Also, although the ratio of N_e to N varies across species types (Frankham, 1995), N is strongly linked to N_e. As such, the researcher should diligently ground-truth NEOGEN population size priors with other measures of N and ensure assumptions are defensible. Random mating is assumed in NEOGEN, but natural species often exhibit more complex mating behaviours. To reduce the complexity of analysis, modes of reproduction such as monogamy or polygamy have been avoided in the current version, although programmatic customization of the underlying NEOGEN model and SIMUPOP engine can achieve this. As mentioned, NEOGEN cannot accommodate fecundity that changes with age and size as observed for some species (e.g., bony fish), although this is less common for mammals and sharks. Also, natural variation in age of sexual maturation is not accommodated.

Genetic estimates of N_e are also subject to caveats. In practice, empirical sample sets yield just one estimate of $N_{e.LD}$. However, NEOGEN's $N_{e.LD}$ estimates are averaged from repeated $N_{e.LD}$ estimates sampled randomly from replicated populations. As such, these estimates may be statistically more accurate than a single empirical estimate and may thus underestimate the sample and locus power required for empirical sampling. Additionally, NEOGEN estimates of $N_{e.LD}$ for a population can be derailed by nonideal sample acquisition (e.g., nonrandom sampling; sampling a population receiving migrants), or relying upon empirical loci that have not been genotyped adequately or correctly (e.g., too few samples genotyped giving nonrepresentative allele frequencies; loci with null alleles). Managing these sources of error are standard practice for population genetics projects, though extra care must be taken when extrapolating results into management practice. Despite these constraints, NEOGEN can be heavily customized for species outside the software's standard parameters. Researchers may tailor the NEOGEN model and the immensely rich and capable underlying simulation engine, SIMUPOP, to better suit the complexities of their focal species. The NEOGEN GUI allows some customization of the options available; alternatively, the NEOGEN model can be run from the command line using easily modified parameter files.

5 | CONCLUSIONS

NEOGEN provides a unique, powerful, practical, and customizable connection between the life history, mortality, and genetics of a specific population and its consequent genetic N_e. NEOGEN is ideal for planning studies of iteroparous species with overlapping generations that are of conservation concern, or to add value to existing studies where N_e estimates are underutilized. Sample and locus acquisition requirements can be predicted by NEOGEN, thereby optimizing genetic population assessment efficiency, and increasing the probability of producing accurate and precise $N_{e.LD}$ estimates for natural populations. NEOGEN therefore facilitates collaboration between wildlife managers and wildlife geneticists by providing a platform for planning field and laboratory sampling requirements and comparing projected results with previous empirical research. The simulation model of NEOGEN is extensive, powerful and overlaid with an easy-to-use interface to encourage biologists and wildlife managers to explore this valuable new source of population information.

ACKNOWLEDGEMENTS

Funding was provided by Fisheries Research and Development Corporation (FRDC) grant (2010/062) to Fisheries New South Wales on behalf of the Australian Government. P. Butcher and V. Peddemors provided constructive support. Thanks to R. Waples for invaluable feedback on NEOGEN, and SIMUPOP guidance from B. Peng was crucial. We greatly appreciate the efforts of the many software testers and document reviewers. Thanks to the anonymous reviewers for their invaluable comments.

AUTHOR CONTRIBUTIONS

D.B. wrote the software and the manuscript. D.B. and J.O. designed the simulation approach as implemented in the software. C.R. provided oversight of the software validation and manuscript production from the viewpoint of applied and theoretical population genetics.

DATA ACCESSIBILITY

The Microsoft Windows (64-bit ver. 7 or later) program, user manual and associated documentation (e.g. README.txt) are available from the Molecular Fisheries Laboratory web site (https://www.molecularfisherieslaboratory.com.au/neogen-v1-3-0-6-a1-software). A GNU General Public License (GPL) licence applies to the NEOGEN software, and the code (written in PYTHON v2.7 and SIMUPOP v1.1.3) is available on request. Example data for zebra shark, *Stegostoma fasciatum*, is included and accessed from within the NEOGEN program, and the documented analysis is available upon request.

ORCID

Dean C. Blower http://orcid.org/0000-0001-8363-6490
Cynthia Riginos http://orcid.org/0000-0002-5485-4197
Jennifer R. Ovenden https://orcid.org/0000-0001-7538-1504

REFERENCES

Andreotti, S., Rutzen, M., van der Walt, S., Von der Heyden, S., Henriques, R., Meÿer, M., … Matthee, C. A. (2016). An integrated mark-recapture and genetic approach to estimate the population size of white sharks in South Africa. *Marine Ecology Progress Series*, *552*, 241–253. https://doi.org/10.3354/meps11744

Barker, J. S. (2011). Effective population size of natural populations of *Drosophila buzzatii*, with a comparative evaluation of nine methods of estimation. *Molecular Ecology*, *20*, 4452–4471. https://doi.org/10.1111/j.1365-294X.2011.05324.x

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., … Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, *32*, 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bernos, T. A., & Fraser, D. J. (2016). Spatiotemporal relationship between adult census size and genetic population size across a wide population size gradient. *Molecular Ecology*, *25*, 4472–4487. https://doi.org/10.1111/mec.13790

Blower, D. C., Pandolfi, J. M., Bruce, B. D., Gomez-Cabrera, M. D., & Ovenden, J. R. (2012). Population genetics of Australian white sharks reveals fine-scale spatial structure, transoceanic dispersal events and low effective population sizes. *Marine Ecology-Progress Series*, *455*, 229–244. https://doi.org/10.3354/meps09659

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*, 195–205. https://doi.org/10.1038/nrg2526

Coombs, J. A., Letcher, B. H., & Nislow, K. H. (2012). GONe: Software for estimating effective population size in species with generational overlap. *Molecular Ecology Resources*, *12*, 160–163. https://doi.org/10.1111/j.1755-0998.2011.03057.x

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NEESTIMATOR V2: Re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, *14*, 209–214.

Dudgeon, C. L., Noad, M. J., & Lanyon, J. M. (2008). Abundance and demography of a seasonal aggregation of zebra sharks *Stegostoma fasciatum*. *Marine Ecology Progress Series*, *368*, 269–281. https://doi.org/10.3354/meps07581

Dudgeon, C. L., & Ovenden, J. R. (2015). The relationship between abundance and genetic effective population size in elasmobranchs: An example from the globally threatened zebra shark *Stegostoma fasciatum* within its protected range. *Conservation Genetics*, *16*, 1443–1454. https://doi.org/10.1007/s10592-015-0752-y

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. *Genetical Research*, *66*, 95–107. https://doi.org/10.1017/S0016672300034455

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*, 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Grimm, A., Gruber, B., Hoehn, M., Enders, K., & Henle, K. (2016). A model-derived short-term estimation method of effective size for small populations with overlapping generations. *Methods in Ecology and Evolution*, *7*, 734–743. https://doi.org/10.1111/2041-210X.12530

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., … Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, *25*, 438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x

Jorde, P. E. (2012). Allele frequency covariance among cohorts and its use in estimating effective size of age-structured populations. *Molecular Ecology Resources*, *12*, 476–480. https://doi.org/10.1111/j.1755-0998.2011.03111.x

Kimura, M., & Crow, J. F. (1963). Measurement of the effective population number. *Evolution*, *17*, 279–288.

Lacy, R. (1993). VORTEX: A computer simulation model for population viability analysis. *Wildlife Research*, *20*, 45–65. https://doi.org/10.1071/WR9930045

Lindenmayer, D., Lacy, R., Thomas, V., & Clark, T. (1993). Predictions of the impacts of changes in population size and environmental variability on Leadbeater's possum, *Gymnobelideus leadbeateri McCoy* (Marsupialia: Petauridae) using population viability analysis: An application of the computer program VORTEX. *Wildlife Research*, *20*, 67–85. https://doi.org/10.1071/WR9930067

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. *Conservation Genetics*, *11*, 355–373. https://doi.org/10.1007/s10592-010-0050-7

McAllister, M. K., Pikitch, E. K., & Babcock, E. A. (2009). Why are Bayesian methods useful for the stock assessment of sharks? In: M. D. Camhi, E. K. Pikitch, & E. A. Babcock (Eds.) *Sharks of the open ocean. Biology, fisheries and conservation* (pp. 351–368). Oxford, UK: Blackwell Publishing Ltd.

Musick, J. A., & Bonfil, R. (2005). FAO Fisheries Technical Paper 474: Management techniques for elasmobranch fisheries. In: FAO Fisheries Technical Paper, p. 261. Food and Agriculture Organization of the United Nations (FAO).

Ovenden, J. R., Leigh, G. M., Blower, D. C., Jones, A. T., Moore, A., Bustamante, C., … Dudgeon, C. L. (2016). Can estimates of genetic effective population size contribute to fisheries stock assessments? *Journal of Fish Biology*, *89*, 2505–2518. https://doi.org/10.1111/jfb.13129

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, *17*, 3428–3447. https://doi.org/10.1111/j.1365-294X.2008.03842.x

Peng, B., & Kimmel, M. (2005). SIMUPOP: A forward-time population genetics simulation environment. *Bioinformatics*, *21*, 3686–3687. https://doi.org/10.1093/bioinformatics/bti584

Portnoy, D. S., McDowell, J. R., McCandless, C. T., Musick, J. A., & Graves, J. E. (2009). Effective size closely approximates the census size in the heavily exploited western Atlantic population of the sandbar shark, *Carcharhinus plumbeus*. *Conservation Genetics*, *10*, 1697–1705. https://doi.org/10.1007/s10592-008-9771-2

Russell, J. C., & Fewster, R. M. (2009). Evaluation of the linkage disequilibrium method for estimating effective population size. In D. L. Thomson, E. G. Cooch, & M. J. Conroy (Eds.), *Modeling demographic processes in marked populations* (pp. 291–320). New York, NY: Springer Media.

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*, 78–87. https://doi.org/10.1016/j.tree.2014.11.009

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/s10592-005-9100-y

Waples, R. S. (2016a). Life-history traits and effective population size in species with overlapping generations revisited: The importance of adult mortality. *Heredity*, *117*, 241–250.

Waples, R. S. (2016b). Tiny estimates of the N_e/N ratio in marine fishes: Are they real? *Journal of Fish Biology*, *89*, 2479–2504.

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

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.

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, *3*, 244–262.

Waples, R. S., Do, C., & Chopelet, J. (2011). Calculating N_e and N_e/N in age-structured populations: A hybrid Felsenstein-Hill approach. *Ecology*, *92*, 1513–1522.

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. *Heredity*, *117*, 233. https://doi.org/10.1038/hdy.2016.60

Waples, R. S., Luikart, G., Faulkner, J. R., & Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1339–1433. https://doi.org/10.1098/rspb.2013.1339

Waples, R. S., & Yokota, M. (2007). Temporal estimates of effective population size in species with overlapping generations. *Genetics*, *175*, 219–233.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

How to cite this article: Blower DC, Riginos C, Ovenden JR. 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*. 2019;19:260–271. https://doi.org/10.1111/1755-0998.12941