

# **TITLE**

Examination of the efficacy of small genetic panels in genomic conservation of companion animal populations

# **RUNNING TITLE:**

Small genetic panel diversity simulations

# **AUTHORS**

Aaron J. Sams<sup>\*1</sup>, Brett Ford<sup>1</sup>, Adam Gardner<sup>1</sup>, Adam R. Boyko<sup>1,2</sup>

# **CONTACT INFORMATION**

1. Embark Veterinary, Incorporated, Boston, MA 02111, USA
2. Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

\* Corresponding Author: [asams@embarkvet.com](mailto:asams@embarkvet.com)

# **ABSTRACT**

In many ways dogs are an ideal model for the study of genetic erosion and population recovery, problems of major concern in the field of conservation genetics. Genetic diversity in many dog breeds has been declining systematically since the beginning of the 1800's, when modern breeding practices came into fashion. As such, inbreeding in domestic dog breeds is substantial and widespread and has led to an increase in recessive deleterious mutations of high effect as well as general inbreeding depression. Pedigrees can in theory be used to guide breeding decisions, though are often incomplete and do not reflect the full history of inbreeding. Small microsatellite panels are also used in some cases to choose mating pairs to produce litters with low levels of inbreeding. However, the long-term impact of such practices have not been

thoroughly evaluated. Here, we use forward simulation on a model of the dog genome to examine the impact of using limited markers panels to guide pairwise mating decisions on genome-wide population level genetic diversity. Our results suggest that in unsupervised mating schemes, where breeding decisions are made at the pairwise- rather than population-level, such panels can lead to accelerated loss of genetic diversity compared to random mating at regions of the genome unlinked to panel markers and demonstrate the importance of genome-wide genetic panels for managing and conserving genetic diversity in dogs and other companion animals.

## KEYWORDS

dogs, conservation genetics, simulation, companion animal, microsatellites

## INTRODUCTION

Loss of genetic diversity in small populations is a major concern of the global conservation community because it can reduce fitness and adaptability, and can ultimately lead to breed, population, or species extinction (O'Grady et al., 2006). For example, Target 13 of the Convention on Biological Diversity's Strategic Plan (<https://www.cbd.int/sp/>) is aimed at "minimizing genetic erosion" of both socio-economically and culturally valuable species and "safeguarding their genetic diversity."

It has been repeatedly demonstrated that the optimal way to reduce genetic erosion in management programs is to minimize the mean kinship within a population (Ballou & Lacy, 1995; Fernández, Toro, & Caballero, 2001; Sonesson & Meuwissen, 2001). Kinship can be estimated using pedigrees, however, pedigrees are frequently inaccurate, incomplete, or missing (Cassell, Adamec, & Pearson, 2003). In lieu of deep, high quality pedigrees, molecular data is used. While there has been a push in recent years to move towards monitoring and managing populations of conservation concern with genomic data in the form of whole genome

genotyping or whole genome sequencing (Flanagan, Forester, Latch, Aitken, & Hoban, 2017; Ivy, Putnam, Navarro, Gurr, & Ryder, 2016; Leroy et al., 2017; Shafer et al., 2015), many populations are still monitored with small sets of neutral markers, typically tens to hundreds of microsatellites (Abdul-Muneer, 2014; Attard et al., 2016; Kaczmarczyk, 2016; Kirk & Freeland, 2011; Pedersen, Pooch, & Liu, 2016; Song et al., 2018; Toro, Fernández, & Caballero, 2009).

While many researchers have cautioned against using small marker panels alone to guide captive breeding (see for example (Miguel Toro et al., 1998; Wang & Hill, 2000)) at least some previous research has suggested that small microsatellite panels can be used to maintain genetic diversity in captive breeding programs (Kaczmarczyk, 2016). Importantly though, as Nicholas and colleagues (\*Nicholas, Mellersh, & Lewis, 2018) recently pointed out, such small microsatellite panels do not effectively survey genome-wide genetic diversity. Rather, they survey genetic diversity at and near (depending on the extent of linkage disequilibrium) the assayed microsatellites. Direct comparison of microsatellite panel and genome-wide (e.g. single nucleotide polymorphisms (SNPs)) panel estimates of genetic diversity are rare, but the argument of Nicholas and colleagues is supported by a recent direct comparison in *Arabidopsis halleri*, which found that microsatellite-based estimates of genetic diversity and population differentiation differ substantially from unbiased estimates from SNPs (Fischer et al., 2017).

In many ways dogs are an ideal model for the study of genetic erosion and population recovery. Genetic diversity in many common domestic dog breeds has been declining systematically since the beginning of the 1800's, when modern breeding practices came into fashion (Jansson & Laikre, 2018). As such, inbreeding in domestic dog breeds is substantial and widespread (Freedman et al., 2014; Kettunen, Daverdin, Helfjord, & Berg, 2017; Pedersen, Pooch, & Liu, 2016; Sams & Boyko, 2018) and has led to an increase in recessive deleterious mutations of high effect (Jagannathan et al., 2019; Marsden et al., 2016) as well as general inbreeding depression (Chu et al., 2019).

Dog breeders and breed clubs are increasingly aware of the serious consequences of diversity loss, and with robust panels of both microsatellite markers and genome-wide SNP arrays widely available commercially, there is great potential for breeders to use genetic testing in ways that ultimately improve (or worsen) genetic diversity. However, a key challenge, at least in the United States, is a lack of population-level management. Rather, individual dog breeders or groups of breeders typically manage small subsets of a breed, often relying on pedigrees or commercially available molecular tests to minimize known genetic health risks and sometimes overall inbreeding in individual litters of dogs. For breeds still managed in such a way, it is critical to long-term breed health and survival to understand the long-term impacts on genome-wide genetic diversity of chosen mating strategies and the molecular tools used to guide those strategies.

As a first step in understanding the impact of such population management, we conducted individual-based forward-time population genetic simulations of linked genetic diversity on a model of the dog genome using SLiM 3 (Haller & Messer, 2019). We apply a range of human-directed mate choice models to ask how well different mate choice schemes applied to a restricted panel of 33 “microsatellite” locations in our model dog genome (referred to throughout as MS33) affect genetic diversity genome-wide. More specifically, we evaluate several combinations of metrics calculated on this set of 33 multi-allelic markers to guide diversity-based mate choice including heterozygosity, internal relatedness (IR) (Amos et al., 2001) and average genetic relatedness (AGR) (Wang, 2002). Importantly, we do not model the generally recommended strategy of selecting parents to minimize population-level kinship, but rather simulate individual mating decisions aimed at optimizing genetic measurements for single offspring. We ask how mate choice using these methods impacts genome-wide genetic variation as reflected by heterozygosity and allelic richness (average number of alleles per locus) compared to random mate choice, as well as mate choice guided by two generation

pedigree awareness, genome-wide heterozygosity, and relatedness calculated from genome-wide identity by descent (see Materials and Methods).

## **MATERIALS AND METHODS**

### ***Genome Model***

We implemented all simulations in SLiM (v3.3) (Haller & Messer, 2019) and all simulations were run in parallel using Amazon Web Services EC2, SQS, and auto-scaling services. The genome model in our simulations is a rough approximation of the canine genome. For computational efficiency, we model genetic variation as non-recombining 0.5 megabase multi allelic haplotype blocks. These haplotype blocks are represented by mutation type 1 (m1) in the simulation template (Appendix S1). Chromosomes are created by dividing the genome into 38 sets of 120 haplotype blocks, approximating a genome size of 2.28 gigabases. Recombination rate between chromosomes is 0.5 and within chromosomes is 0.005, based on an overall mutation rate per base-pair of  $10^{-8}$ . Additionally, we modeled 33 microsatellite loci (m2 in the simulation template) spaced across the first 25 chromosomes as such: 3 on chromosome 1, 2 each on chrs 2 — 7, 1 each on chrs 8 — 25. This distribution of markers across chromosomes is similar to the 33 STR panel used in (Pedersen, Pooch, & Liu, 2016). Given a number of unique haplotype blocks and microsatellite alleles at the start of the simulation burn-in, we evenly distributed those alleles across individuals in the founding population (see Appendix S2 — Figure S2 for a graphical example of this model).

### ***Demographic Model***

We created a relatively simple demographic model in which a single ancestral population evolves for a burn in and drift period to allow founding genetic diversity to recombine sufficiently, and experience sufficient genetic drift. This is followed by a short immediate bottleneck. Finally, the population expands and goes through a number of mate-choice generations. Population genetic data is collected at the beginning of the first generation of mate choice, and again at the

very end of the simulations (see Appendix S2 — Figure S3 for a graphical example of this model).

### ***Life Cycle Model***

Generations in the initial burn in/drift period and bottleneck follow the standard Wright-Fisher model in SLiM. However, during mate-choice, we induce a slightly different model. Creation of offspring in the mate-choice schema includes (see Appendix S2 — Figure S4 for a graphical example of this model):

1. Random sampling of a parent from the full population (sample individual that has not been mated > MAXIMUM\_NUMBER\_OF\_MATINGS times).
2. Choose second parent:
  - a. Sub-sampling a fraction of the remaining population randomly (according to MATING\_POOL\_SIZE) from which to choose potential mates. This step is intended to model the fact that within an effective population of purebred dogs, individual dogs only have access to a limited number of other dogs as mates. This sampling is repeated if no eligible (has been mated < MAXIMUM\_NUMBER\_OF\_MATINGS times) dogs are sampled.
  - b. A mate for the first individual is chosen according to a specific mate choice model (see Mate Choice Models) from this pool.
3. In the event of a layered mate-choice model in which two statistics are used to choose mates, a PROPORTION\_OF\_MATES\_FOR\_LAYERED\_MATE\_CHOICE parameter is used to sub-sample the potential mates based on each statistic used.

### ***Mate Choice Models***

Mate choice models that we implemented in this study primarily differ in how a second parent is selected. First parents are selected randomly from the entire population as described above.

Random: Second parent is randomly selected from the sampled mating pool.

Pedigree: Second parent is the individual with the lowest relatedness as calculated from three generation pedigrees, randomly sampled in ties. In other words, using the keepPedigrees=True option in SLiM 3, pedigree relatedness between individuals in the current generation can be calculated from pedigrees. This option maintains pedigrees including all current individual's parents and grandparents.

Heterozygosity models [Microsatellite (MS33-HET) and Genome-wide (GW-HET)]: For these models we calculate the expected heterozygosity for offspring between the first parent and all individuals in the mating pool as the average pairwise observed homozygosity across all four pairwise combinations of parental genomes (assuming no recombination). We choose as the best mate the individual that would produce offspring with the highest heterozygosity and select randomly amongst ties.

Internal Relatedness (MS33-IR): For this model, as above, we calculate mean Internal Relatedness (Amos et al., 2001) amongst four possible gametic pairs at each microsatellite locus and then average across all loci and choose the individual that produces the lowest IR value, selecting randomly from ties.

Average Genetic Relatedness (MS33-AGR): Here we apply a method designed to calculate relatedness between individuals based on small panels of SNPs or microsatellites (Wang, 2002). We based the code in our SLiM template on the implementation of this calculation found in the R package Demerelate (<https://github.com/cran/Demerelate>) and ran several tests to ensure that our SLiM implementation and the R version produced identical results. We chose as the second parent the individual that produced the lowest AGR value, selecting randomly from ties.

Whole Genome Relatedness (GW-REL): We calculate whole genome relatedness as in (Hedrick & Lacy, 2015) and as above, calculated relatedness across all four possible gametic pairs between two individuals. We chose as the second parent the individual least related to the first randomly selected parent and chose amongst ties randomly.

Layered Mate Choice (MS33 IR AGR): We additionally investigated a single microsatellite mate choice model combining the IR and AGR statistics. First, a fraction of individuals in the mating pool are chosen based on the IR statistic, and then an individual is chosen from that sample based on AGR (see Life Cycle Model above).

## **Population Genetic Statistics**

Observed Heterozygosity: We calculated per individual as the fraction of all genotypes in an individual that are heterozygous. In some cases we averaged observed heterozygosity across all individuals in a population.

Allelic Richness: We calculated allelic richness as the total number of unique alleles at each position. In some outputs (results not presented here but data available) we also calculated richness for alleles  $\geq 0.05$  frequency.

Coefficient of inbreeding: We calculated the coefficient of inbreeding (COI) for individual dogs as the fraction of all mutations of type 'm1' that are identical within an individual divided by the total genomic length.

## **Statistical Analyses**

For nearly all statistical comparisons, including calculations of 95% confidence intervals used in Figure 1 we have utilized estimation graphics and statistics (Ho, Tumkaya, Aryal, Choi, & Claridge-Chang, 2019) as implemented in the python package dabest (v0.2.5 - <https://acclab.github.io/DABEST-python-docs/index.html>). See python script used to generate results for specific uses (github link will be added upon acceptance).

## **RESULTS**

### ***Genetic variation in simulated populations compared to present-day dogs***



For each of eight mate choice models (7 described in Materials and Methods and an additional compound mate choice model of MS33-IR and MS33-AGR), we generated 100 replicate simulations for each of 21 different demographic parameter sets. Three baseline parameter sets vary in the effective population size used during the burn in period. Additional simulations are variants of these three baseline parameters, varying one other parameter (Appendix S3 — Table S1).

The simulated populations show realistic genetic variation compared to present-day domestic dog breeds. For example, across all models, prior to mate selection the mean coefficient of inbreeding (COI) across all 800 replicates (100 replicates x 8 model types) for parameter set 1 (PS-1) is approximately 0.17 — 0.21, and across all simulations is ~0.12 — 0.38, well within the range of several common dog breeds today (Sams & Boyko, 2018). Similarly, mean internal relatedness (IR) varies in the range of -0.013 — 0.034 across PS-1 and -0.015 — 0.1 across all parameter sets (See Appendix S3 — Tables S2 and S3 for COI and IR means by model and parameter set). Pedersen and colleagues (Pedersen, Pooch, & Liu, 2016) observed a mean IR value of 0.007 in 102 Bulldogs.

### ***MS33-based mate-choice models lead to more diversity loss than random mate-choice or short pedigrees over time***

Performance of all models was measured as the loss of genetic diversity relative to the loss of genetic diversity observed across the random mating model. Of the MS33-based models tested, Internal Relatedness-based mate choice (MS33-IR) performed best overall after 40 generations of mate choice. This model limited loss of heterozygosity comparably or slightly better than random mating across all parameter sets tested (Figure 1, Appendix S3 — Table S4). However, this model performed worse than random mating at preserving allelic richness across all parameter sets but one (PS-19) at which it was indistinguishable from random mating (Appendix S3 — Table S5).

With a few exceptions, all other MS33-based models performed worse than random mating using both heterozygosity and allelic richness as a metric of diversity loss. In these cases, diversity loss is as great or in some cases substantially greater than random mate choice and recent pedigree-based mate choice. Importantly, even in cases where MS models preserve heterozygosity more than random mating, the trajectory of diversity loss over time in these simulations suggests that given enough time these models would also perform worse overall than random mating (Figure 1, Appendix S2 — Figure S1).

Among the MS33-based models, MS33-IR mate choice lost the least amount of genetic diversity, and MS33-AGR mate choice lost the greatest amount of genetic diversity. In all cases, the accelerated loss of genetic diversity compared to random mating is due to preserving diversity at a small number of loci at the expense of the remainder of the genome. In other words, by avoiding inbreeding with individuals more closely related at a small number of loci scattered throughout the genome, the effective population size at unlinked loci, which are evolving under drift, is further reduced.

As direct evidence of this reduction of effective size at unlinked loci, we observe that genetic diversity at MS loci is preserved well, but is lost more than random mating away from these MS positions (Figure 2A). This pattern is consistent regardless of the MS-based mate-selection model we examine and is not observed in genome-wide models described below (Figure 2B).

### ***Number of repeated matings can moderate “popular parent” effects***

We examined whether adjusting the total number of individuals that any single individual can access as potential mates within the population, termed here “mating pool size,” as well as the maximum number of times any single individual can contribute to the next generation, termed here “number of repeated matings,” has a substantial impact on the magnitude of the loss of diversity in MS-based mate selection models.

We reduced the maximum number of matings per individual per generation from 200 (the population size) to 50 (PS10-12) and 5 (PS13-15). Reducing from 200 to 50 had no significant impact on diversity metrics (Figure 3A, Appendix S3 — Tables S6-8). However, reducing from 200 to 5 led to a substantial reduction in loss of heterozygosity and allelic richness, and lower increase in COI (Figure 3B, Appendix S3 — Tables S6-8).

Additionally, we found that halving mating pool size from 50 to 25 (0.25 to 0.125 of the total population; PS16-18) and from 50 to 10 (PS-19-21), unexpectedly led to a notable reduction in the amount of genome-wide diversity loss (heterozygosity, richness, or coefficient of inbreeding) in all MS-based mate choice schema (Figure 3C,3D, Appendix S3 — Tables S6-8). Upon further inspection, we realized this result is consistent with the mating pool size acting to moderate the number of times the same individual is chosen as a parent in a single generation and that the two parameters discussed here are not as independent as we originally envisioned when designing our simulations. This may be because the “mating pool” is randomly chosen each time a first parent is sampled, such that if the same individual is chosen twice, it will have different mating pools each time. In other words, reducing the number of individuals available as mates for a given individual reduces the number of times that individuals with low mean genetic similarity *to the rest of the population* will be over represented in the next generation. However, while limiting the number of possible mates for each individual decreased the loss of heterozygosity for MS-based mate choice models, it has no impact on the random and pedigree-based mate choice models and increases the loss of heterozygosity slightly in the two genome-wide genetic models (Figure 3C, 3D), supporting the idea above that higher variance in heterozygosity when marker sample size is low drives the unexpected relationship between preservation of heterozygosity and an individual’s mating pool size.

### ***Genome-wide metrics improve diversity preservation***

In addition to the MS-based mate selection models, we also included several models meant to capture the viability of using genome-wide metrics in general to preserve genetic diversity. These include tracking two generation pedigrees (pedigree)- to avoid matings between very close relatives, genome-wide heterozygosity (GW-HET)- to select mates which maximize heterozygosity in the offspring, and genome-wide relatedness (GW-REL)- which prefers the most distantly related individuals as mates. We find that these three models all lead to greater preservation of genetic diversity than random mating. Perhaps more importantly, these models reduce the rate of genetic drift over time (Random model), compared to the MS models which accelerate the rate of genetic drift (Figure 1).

## DISCUSSION

In this study, we used forward population genetic simulations of a model of the canine genome to investigate the efficacy of using a small genetic marker panel (e.g. a microsatellite panel) to guide mating aimed at preserving existing genetic variation in a population, when mating decisions are made at the pairwise, rather than population, level. We ran these simulations across a range of mate choice models and demographic parameterizations using a genomic model that included both genome-wide genetic markers and a set of 33 markers distributed across the genome (herein referred to as the GW and MS33 marker sets, see methods).

Most previous work on conservation management with molecular data has focused on cases where a population can be managed by selecting the entire configuration of parents for the next generation with marker assisted selection (MAS). For example, Fernandez and colleagues (\*Fernández, Toro, & Caballero, 2004) demonstrated that in a single population, management programs where parental contributions are chosen to maximize either heterozygosity or allelic richness at a set of multi-allelic markers are optimal at maintaining each of those statistics, but that heterozygosity can be better at maintaining allelic richness than vice-versa. However, no prior work, to our knowledge, has addressed whether similar strategies that

optimize at the level of individual mating pairs, rather than the entire population of parents, can similarly act to preserve diversity. Consistent with this prior work, our results suggest that selecting optimal mates for individuals from an entire population using heterozygosity and other kinship metrics can act to preserve genetic diversity at markers used to calculate the test statistic (for example see Figure 2).

Importantly, however, we also found that given enough generations using small panels of markers in such a mating scheme does not preserve diversity genome-wide. In fact, mate-choice models using the MS33 marker set over time led to greater loss of genetic variation compared to random mating in the form of reduced heterozygosity and allelic richness measured using the GW marker set. Lopez-Cortegano and colleagues (\*López-Cortegano et al., 2019) simulated management of subdivided populations and found that using a restricted number of markers was less effective than whole genome data but still more effective than random mating. However, the density of markers in their simulations is greater than typical microsatellite panels and they acknowledge that less dense panels would likely be less effective. Nonetheless, our results are partially consistent with this result, in that the MS33-IR model does preserve genome-wide genetic diversity better than random mating (but not allelic richness) during the course of our simulations.

Our results suggest that reducing the number of times that any given individual can contribute offspring to the next generation, either explicitly in the form of the “maximum number of matings” parameter or implicitly by reducing the “mating pool size” parameter, can act to moderate the severity of diversity loss compared to the random mating model. This finding is generally consistent with theory and prior simulation work which has demonstrated that optimal management schemes to preserve genetic diversity include limiting variance in family size, in other words, ensuring that no single individual contributes disproportionately to the next generation (Miguel Toro, Silió, Rodrigáñez, & Rodriguez, 1998).

Toro and colleagues (\*Miguel Toro, Silió, Rodrigañez, Rodriguez, & Fernández, 1999) demonstrated that irrespective of variance in family size MAS should lead to better preservation of diversity than using no genetic information at all. In contrast, our results suggest that in an unsupervised pairwise parental selection scheme, limited marker panels lead to substantially more diversity loss than using no genetic information at all. A general consensus in MAS of parental populations is that pedigrees should be the primary source of kinship calculations and that small microsatellite panels are generally only useful to supplement pedigrees (M. A. Toro, Fernández, & Caballero, 2009). Our results from pairwise parental selection are consistent with this, as we have shown that using only shallow pedigrees to minimize loss of genetic diversity is preferable to using a small panel of genetic markers alone.

Genetic drift comes from two primary sources in a diploid population: variation in genetic contribution between individuals in a population and variation in genetic diversity at a given locus within an individual (J Wang & Hill, 2000). Here, we have shown (Figure 2) that the added loss of genetic variation in our simulations relative to random mating is due to accelerated loss of diversity throughout the majority of the genome that is untagged by MS33 markers. While we did not specifically explore the causes of this difference, we suspect that even in our simulations which reduce the contributions of any given individual to the next generation, that groups of individuals which happen to be most distantly related to all other individuals in a given generation across the MS33 marker set are disproportionately chosen as mates for the next generation. This variance in contribution of families (as measured by the MS33 set) across generations will act to consistently reduce effective size at markers unlinked to the MS33 marker set.

Finally, for comparison we simulated several GW mate-choice models and found, by and large, that using genome-wide genetic data to monitor genetic diversity and make mate selection decisions is far superior to small marker panels, and typically, random mating. This result has important implications for the preservation of domestic dog breeds. Most academic

effort in the field of genetic diversity management over the past few decades has primarily been focused on optimal management for small populations of conservation concern where mating in the entire population can be controlled (Ballou & Lacy, 1995; Fernández, Toro, & Caballero, 2001, 2004; Kettunen, Daverdin, Helfjord, & Berg, 2017; López-Cortegano et al., 2019; Sonesson & Meuwissen, 2001). Similarly, in livestock, conservation of genomic diversity in combination with genomic selection can occur at the level of entire herds or regional populations, though also suffers from geographic partitioning and localization of conservation efforts (Bosse et al., 2015; Bruford et al., 2015; Herrero-Medrano et al., 2014; Ramljak et al., 2018; Zhao et al., 2019). In contrast, dog breeds, as well as breeds in other companion species such as cats and horses, have populations which are typically maintained by networks of individual breeders. Therefore, it is very important to understand the long term impact of different types of breeding practices in these systems.

Here, we have shown that using small panels of molecular markers is no substitute for quality pedigree information, or more importantly whole genome characterization of genetic diversity using dense genetic markers or whole genome sequence data. Our results suggest that optimal management of unsupervised companion animal populations should 1) include strictly limiting individual and family contributions to the next generation and 2) the selection of mating pairs to minimize inbreeding in offspring using deep pedigree information or, more optimally, using dense genotype data to maximize heterozygosity/minimize inbreeding in offspring, as pedigrees are often incomplete and do not incorporate variance in inheritance of IBD segments amongst related individuals (Cassell, Adamec, & Pearson, 2003; Hill & Weir, 2011; Keller, Visscher, & Goddard, 2011).

We note that we did not directly compare our results to whole-population management schemes, as such management strategies are not currently feasible for most companion animals. We suspect that such schemes will be generally superior to the unsupervised mating methods examined here, as they are better able to optimize contributions from individuals and

choose the optimal (or near optimal) configuration of pairwise matings to preserve existing genetic diversity.

Most companion animal species remain relatively unmanaged with respect to genetic diversity at the breed level. As such, genetic diversity has rapidly decayed in many breeds over the past century (Jansson & Laikre, 2018). While we have not focused on optimizing the use of whole genome molecular data to preserve genetic diversity in this study, future species-specific analyses should aim to develop specific recommendations to individual breeders. For example, more realistic (non-Wright-Fisher) models would better reflect the breeding practices used in companion animal breeding. Further, in our whole genome mate choice methods we have focused primarily on maximizing heterozygosity, but preservation of allelic diversity is also an important metric to optimize, as the number of unique alleles creates a limit on the maximum heterozygosity attainable (Fernández, Toro, & Caballero, 2004). Finally, we have not considered here the needs of diversity management schemes to also consider balancing the goals of preserving genetic diversity with simultaneously eliminating deleterious variation from a population. In particular, due to the lack of past management, many companion animal breeds carry high effect deleterious mutations, and care must be taken to purge such variation without reducing linked neutral variation (Fernández et al., 2004; Hedrick & Garcia-Dorado, 2016).

Eventually, companion animal breeding may benefit from large-scale participation in databases and services aimed at tracking breed-wide whole genome genetic diversity, including awareness of adaptive and deleterious variation, to limit variance in family contributions, maximize the inclusion of genetic variation in subsequent generations, and purge deleterious variation over time. Experimentation and optimization of such a system applied to breeds in a large and diverse species such as domestic dogs would provide critical case studies to the conservation genetics community (Shafer et al., 2015), help breeders and breed organizations understand the limits of truly closed breeding, and better conserve some of the world's most precious animal resources.



## DATA ARCHIVING STATEMENT

Data and code for this study are available at: to be completed after manuscript is accepted for publication.

## ACKNOWLEDGEMENTS

We thank the customers of Embark, who's participation and curiosity made this work possible.

We also thank our colleagues at Embark (in particular Alison Ruhe and Tiago Antao for several helpful discussions), the Embark scientific advisory board, Cornell University, and the Kevin M. McGovern Family Center for Venture Development in the Life Sciences, for their guidance and encouragement. This study was funded by Embark.

## LITERATURE CITED

Abdul-Muneer, P. (2014). Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies. *Genetics Research International*, 2014(6), 1–11.

<https://doi.org/10.1155/2014/691759>

Amos, W., Wilmer, J., Fullard, K., Burg, T., Croxall, J., Bloch, D., & COULSON, T. (2001). The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society B: Biological Sciences*, 268(1480), 2021–2027. <https://doi.org/10.1098/rspb.2001.1751>

Attard, C., Möller, L., saki, Hammer, M., Bice, C., Brauer, C., ... Beheregaray, L. (2016). A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conservation Biology*, 30(5), 1060–1069. <https://doi.org/10.1111/cobi.12699>

Ballou, J., & Lacy, R. C. (1995). Identifying genetically important individuals for management of genetic variation in pedigreed populations. *Population Management for Survival & Recovery*, 76–111.

Bosse, M., Megens, H.-J., Madsen, O., Crooijmans, R. P., Ryder, O. A., Austerlitz, F., ... de Cara, A. M. (2015). Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Research*, 25(7), 970–981. <https://doi.org/10.1101/gr.187039.114>

Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F. J., ... Zhan, X. (2015). Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Frontiers in Genetics*, 6(109), 337–11. <https://doi.org/10.3389/fgene.2015.00314>

Cassell, B., Adamec, V., & Pearson, R. (2003). Effect of Incomplete Pedigrees on Estimates of Inbreeding and Inbreeding Depression for Days to First Service and Summit Milk Yield in Holsteins and Jerseys. *Journal of Dairy Science*, 86(9), 2967–2976. [https://doi.org/10.3168/jds.s0022-0302\(03\)73894-6](https://doi.org/10.3168/jds.s0022-0302(03)73894-6)

Chu, E. T., Simpson, M. J., Diehl, K., Page, R. L., Sams, A. J., & Boyko, A. R. (2019). Inbreeding depression causes reduced fecundity in Golden Retrievers. *Mammalian Genome*, 1–7. <https://doi.org/10.1007/s00335-019-09805-4>

Fernández, J., Toro, M., & Caballero, A. (2001). Practical implementation of optimal management strategies in conservation programmes: a mate selection method. *Animal Biodiversity and Conservation*, 24(2), 17–24.

Fernández, J., Toro, M., & Caballero, A. (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Wiley Online Library*, 18(5), 1358–1367.

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., ... Widmer, A. (2017). Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics*, 1–15. <https://doi.org/10.1186/s12864-016-3459-7>

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2017). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, 11(7), 1035–1052. <https://doi.org/10.1111/eva.12569>

Freedman, A. H., Gronau, I., hweizer, R., Vecchyo, D., Han, E., Iva, P., ... Novembre, J. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genetics*, 10(1), e1004016. <https://doi.org/10.1371/journal.pgen.1004016>

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, 36(3), 632–637. <https://doi.org/10.1093/molbev/msy228>

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding Inbreeding Depression, Purging, and Genetic Rescue. *Trends in Ecology & Evolution*, 31(12), 940–952. <https://doi.org/10.1016/j.tree.2016.09.005>

Hedrick, P. W., & Lacy, R. C. (2015). Measuring relatedness between inbred individuals. *The*

*Journal of Heredity*, 106(1), 20–25. <https://doi.org/10.1093/jhered/esu072>

Herrero-Medrano, J., Megens, H.-J., Groenen, M. A., Bosse, M., Perez-Enciso, M., & Crooijmans, R. (2014). Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics*, 15(1), 601–612. <https://doi.org/10.1186/1471-2164-15-601>

Hill, W., & Weir, B. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, 93(1), 47–64. <https://doi.org/10.1017/s0016672310000480>

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: data analysis with estimation graphics. *Nature Chemical Biology*, 1–4. <https://doi.org/10.1038/s41592-019-0470-3>

Ivy, J. A., Putnam, A. S., Navarro, A. Y., Gurr, J., & Ryder, O. A. (2016). Applying SNP-Derived Molecular Coancestry Estimates to Captive Breeding Programs. *Journal of Heredity*, 107(5), 403–412. <https://doi.org/10.1093/jhered/esw029>

Jagannathan, V., Drogemuller, C., Leeb, T., (DBVDC), D., Aguirre, G., André, C., ... Wade, C. (2019). A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Animal Genetics*, 99, 6693–10. <https://doi.org/10.1111/age.12834>

Jansson, M., & Laikre, L. (2018). Pedigree data indicate rapid inbreeding and loss of genetic diversity within populations of native, traditional dog breeds of conservation concern. *PloS One*, 13(9), e0202849-17. <https://doi.org/10.1371/journal.pone.0202849>

Kaczmarczyk, D. (2016). Selection of optimal spawning pairs to maintain genetic variation among captive populations of Acipenseridae based on the polymorphism of microsatellite loci. *Archives of Polish Fisheries*, 24(2), 77–84. <https://doi.org/10.1515/aopf-2016-0009>

Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics*, 189(1), 237–249. <https://doi.org/10.1534/genetics.111.130922>

Kettunen, A., Daverdin, M., Helfjord, T., & Berg, P. (2017). Cross-Breeding Is Inevitable to Conserve the Highly Inbred Population of Puffin Hunter: The Norwegian Lundehund. *PloS One*, 12(1), e0170039-16. <https://doi.org/10.1371/journal.pone.0170039>

Kirk, H., & Freeland, J. R. (2011). Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *International Journal of Molecular Sciences*, 12(6), 3966–3988. <https://doi.org/10.3390/ijms12063966>

Leroy, G., Carroll, E. L., Bruford, M. W., DeWoody, A. J., Strand, A., Waits, L., & Wang, J.

(2017). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, 11(7), 1066–1083.  
<https://doi.org/10.1111/eva.12564>

López-Cortegano, E., Pouso, R., Labrador, A., Pérez-Figueroa, A., Fernández, J., & Caballero, A. (2019). Optimal Management of Genetic Diversity in Subdivided Populations. *Frontiers in Genetics*, 10, 76–10. <https://doi.org/10.3389/fgene.2019.00843>

Marsden, C. D., Vecchyo, D., O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., ... Lohmueller, K. E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America*, 113(1), 152–157. <https://doi.org/10.1073/pnas.1512501113>

Nicholas, F. W., Mellersh, C., & Lewis, T. (2018). Letter to the editor regarding an autosomal recessive mutation in *SCL24A4* causing enamel hypoplasia in Samoyed and its relationship to breed-wide genetic diversity. 1–4. <https://doi.org/10.1186/s40575-018-0059-7>

O'Grady, J. J., Brook, B. W., Reed, D. H., Ballou, J. D., Tonkyn, D. W., & Frankham, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, 133(1), 42–51. <https://doi.org/10.1016/j.biocon.2006.05.016>

Pedersen, N. C., Pooch, A. S., & Liu, H. (2016). A genetic assessment of the English bulldog. *Canine Genetics and Epidemiology*, 1–16. <https://doi.org/10.1186/s40575-016-0036-y>

Ramljak, J., Bunevski, G., Bytyqi, H., Marković, B., Brka, M., Ivanković, A., ... Medugorac, I. (2018). Conservation of a domestic metapopulation structured into related and partly admixed strains. *Molecular Ecology*, 27(7), 1633–1650. <https://doi.org/10.1111/mec.14555>

Sams, A. J., & Boyko, A. R. (2018). Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. *G3: Genes/Genomes/Genetics*, g3.200836.2018-7.  
<https://doi.org/10.1534/g3.118.200836>

Shafer, A., Wolf, J. B., 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(2), 78–87. <https://doi.org/10.1016/j.tree.2014.11.009>

Sonesson, A., & Meuwissen, T. (2001). Minimization of rate of inbreeding for small populations with overlapping generations. *Genetics Research*, 77(3), 285–292.  
<https://doi.org/10.1017/s0016672301005079>

Song, Z., Liu, D., Zhou, Y., Zhang, X., Chen, Y., Li, C., ... Song, Z. (2018). Low genetic diversity in broodstocks of endangered Chinese sucker, *Myxocyprinus asiaticus*: implications for artificial propagation and conservation. *ZooKeys*, 792(2), 117–132.

<https://doi.org/10.3897/zookeys.792.23785>

Toro, MA, Fernández, J., & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120, 174–195.

<https://doi.org/10.1016/j.livsci.2008.07.003>

Toro, Miguel, Silió, L., Rodrigáñez, J., & Rodriguez, C. (1998). The use of molecular markers in conservation programmes of live animals. *Genetics Selection Evolution*, 30(6), 585–17.

<https://doi.org/10.1186/1297-9686-30-6-585>

Toro, Miguel, Silió, L., Rodrigáñez, J., Rodriguez, C., & Fernández, J. (1999). Optimal use of genetic markers in conservation programmes. *Genetics Selection Evolution*, 31(3), 255.

<https://doi.org/10.1186/1297-9686-31-3-255>

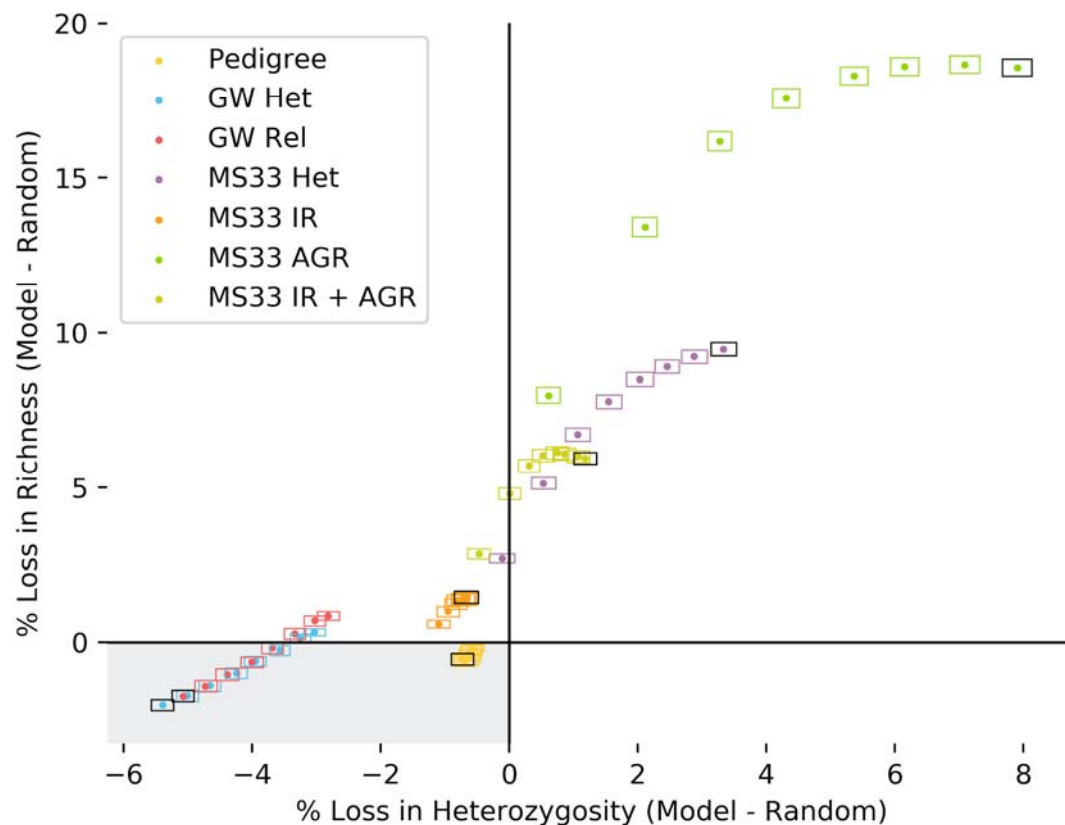
Wang, J., & Hill, W. (2000). Marker-assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics*, 154(1), 475–489.

Wang, Jinliang. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, 160(3), 1203–1215.

Zhao, Q.-B., Sun, H., Zhang, Z., Xu, Z., Olasege, B., Ma, P.-P., ... Pan, Y.-C. (2019). Exploring the Structure of Haplotype Blocks and Genetic Diversity in Chinese Indigenous Pig Populations for Conservation Purpose. *Evolutionary Bioinformatics Online*, 15, 1176934318825082.

<https://doi.org/10.1177/1176934318825082>

# FIGURES

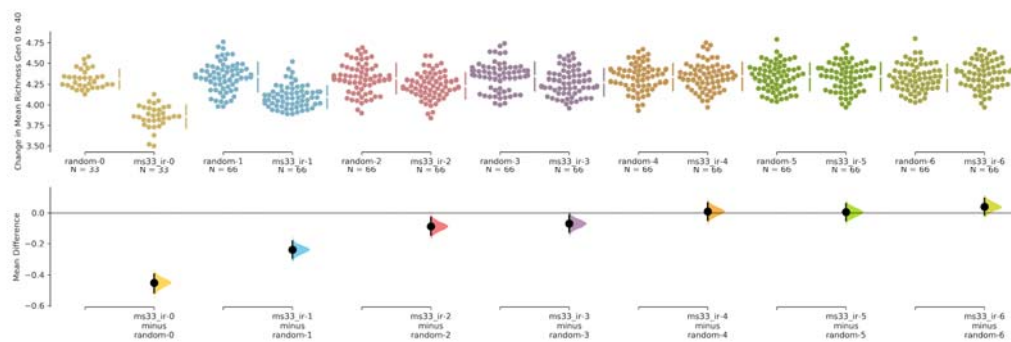


**Figure 1. Small marker panel mate selection performs worse and genome-wide marker panel mate selection improves over time relative to random mating**

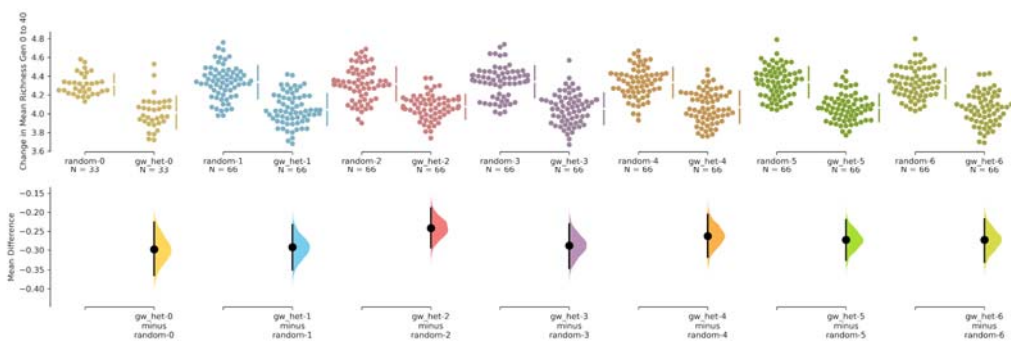
Illustration of the percent loss in heterozygosity and richness over the duration of simulations for parameter set 1, relative to loss during random mating simulations. Dots represent the mean difference between each mate choice model and random mating using a randomization method. Height and width of boxes represent 95% confidence intervals for richness and heterozygosity, respectively (see Materials and Methods). Black boxes represent the mean difference over the duration of simulations (40 generations of mate choice). Grey box represents the quadrant in which both loss of heterozygosity and richness is slowed relative to random mating. GW models

(including Pedigree) improve over the course of simulations, while MS models lose more diversity over time than random mating.

A



B

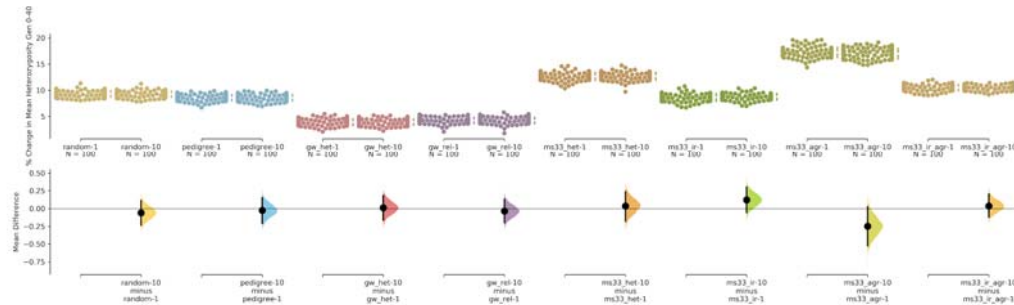


**Figure 2. Local preservation of allelic richness degrades with distance to MS markers in MS but not GW mate choice models**

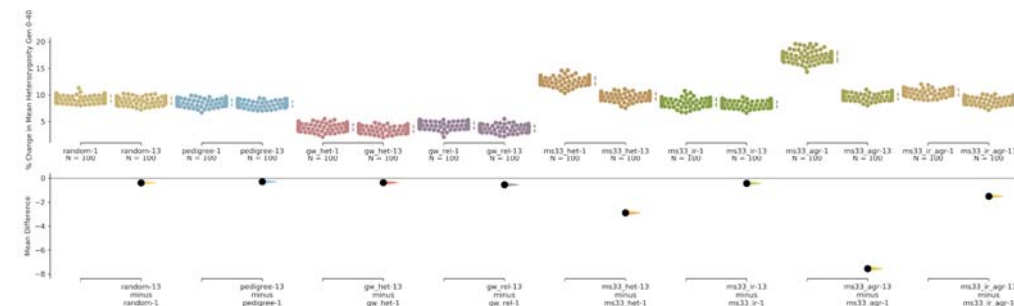
Mean difference in percent allelic richness lost during 40 generations of mate choice between Random and other models. A) In MS models such as MS33-IR, richness is better preserved near markers used during mate choice but degrades with distance. B) In GW models like GW-HET, richness is consistently preserved across all markers.



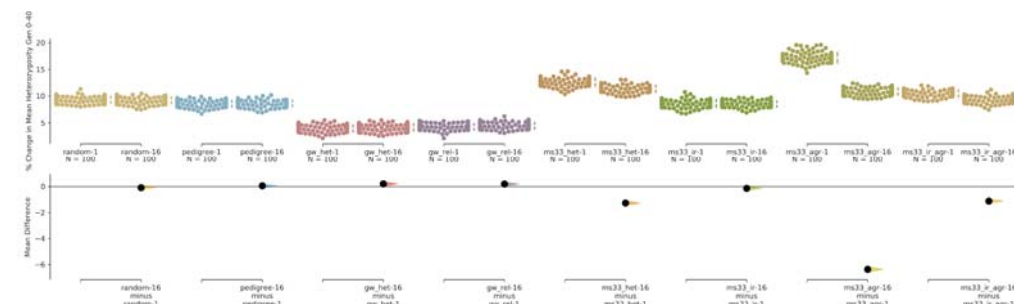
A



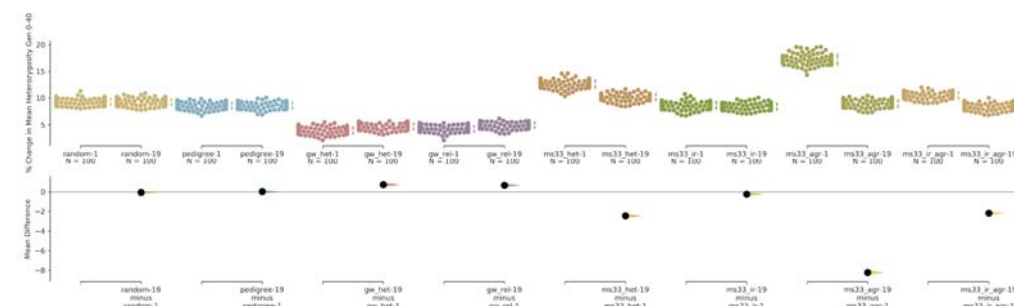
B



C



D



### **Figure 3. Limiting the number of matings per individual improves preservation of heterozygosity in MS models**

Change in loss of heterozygosity over 40 mate choice generations. Lower values correspond to better preservation of heterozygosity. A) PS1 compared to PS10

(maximum\_number\_of\_matings 200 to 50). B) PS1 compared to PS13

(maximum\_number\_of\_matings 200 to 5). C) PS1 compared to PS16 (mating\_pool\_size 50 to

25). D. PS1 compared to PS19 (mating\_pool\_size 50 to 10).