SECTION 5

Developing and using gene bank collections

Developing and using gene bank collections

Coralie Danchin-Burge, Institut de l'Elevage, Paris, France
Christian Reimer, Institute of Farm Animal Genetics, Friedrich-Loeffler-Insitut, Neustadt, Germany
Jack Windig, Centre for Genetic Resources, the Netherlands (CGN) of Wageningen University and Research, Wageningen, Netherlands

5.1 INTRODUCTION

As noted in Section 1, gene banks have value not only as a backup to recreate a breed in case of disasters, but also to serve breeding programmes in existing *in situ* populations, to develop new populations, and to support research. In contrast to live populations, a gene bank collection does not "evolve". Once a collection is established, genetic drift and associated loss of alleles does not occur, nor does adaptation to the environment. Consequently, a collection can play different roles in the conservation of a breed (see Figure 5.1). An advantage of gene banks relative to live populations is that the gene bank represents genetic variation in the population at the time of sampling, which can be continual. As a result, alleles can be present in a gene bank that had been lost from the live population, and gene banks thus can help to restore genetic diversity in the live population (Dechow *et al.*, 2020).

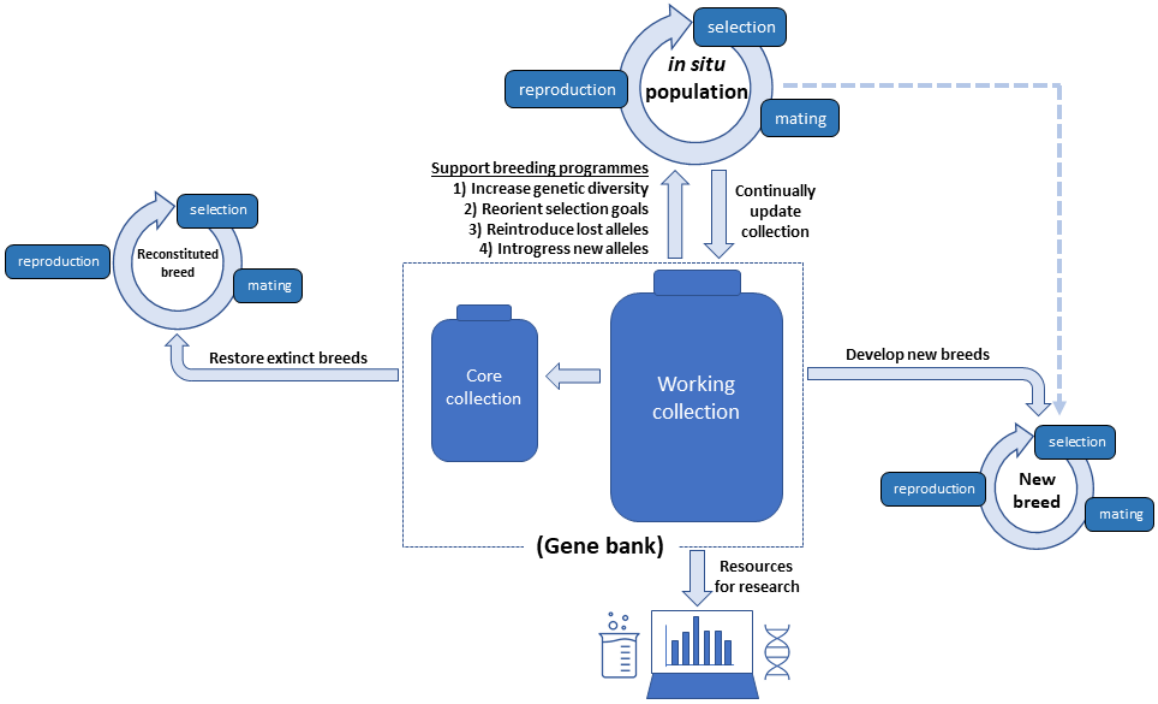

FIGURE 5.1
Uses of gene bank collection (*ex situ in vitro* cryocollection)
Note: Adapted from Berg and Windig, 2017; Tixier-Boichard and Crooijmans, 2019.

Figure 5.1 and the previous FAO guidelines for *Cryoconservation of animal genetic resources* (FAO, 2012) refer to the maintenance of "core" and "working" collections, for which the core collection is material kept for reconstitution of breeds in the case of a disaster, while the working collection is used for all other purposes such as research or supporting breeding programmes. The previous FAO guidelines (FAO, 2012) contain detailed instructions on the amounts of genetic material needed for the core collection. As time has passed and gene banks have become more and more utilized for purposes

other than breed reconstitution, the distinction between these two collections has become blurred. In many cases, scarcity of specimens available for storage may even prevent the creation and maintenance of separate core and working collections. Therefore, many gene banks no longer make such a distinction, but rather choose to develop a single sufficiently large collection and to manage it strategically to serve multiple purposes. With this in mind, gene bank managers and stakeholders for animal genetic resources (AnGR) are advised to consider their national *in vitro* collections to be dynamic resources and to routinely identify and develop opportunities for continual exploitation and replenishing of those collections.

With regard to supporting *in situ* populations, gene bank collections can be used to introduce valuable traits into live populations (see Box 5.1), serve as an archive for research, and improve breeding to better meet changing breeding goals. The latter may be because of changes in the market, due to environmental conditions or because of emerging diseases (Gandini and Oldenbroek, 2007).

BOX 5.1

A practical application of a marker-assisted introgression of a specific trait

A genetic resource population conserved in a gene bank can be used to take a specific trait of this population and to transfer it into a second population. The use of genetic markers can accelerate this process. Within the Horizon 2020 European Union project Innovative Management of Animal Genetic Resources (IMAGE, 2020), the marker-assisted transfer of a specific trait, the blue eggshell colour, from a donor population, the Araucana breed of chickens, to a commercial breeding line was demonstrated. Researchers of the University of Göttingen and the Friedrich-Loeffler-Institut in Germany integrated this causal mutation into a commercial White Leghorn breeding line to induce production of blue shelled eggs.

An initial F1 generation, two marker-assisted backcross generations (BC1 and BC2) and a final intercross generation (IC) were generated, aiming at developing a high performing White Leghorn like line that is homozygous for blue eggshell colour. To achieve this, all birds of the study were genotyped with a custom-made 52 000 single nucleotide polymorphism (SNP) array and 24 newly developed breed/line specific SNPs at the introgression locus. Genotyping results were analysed for the detection of haplotypes at the introgression locus. Each recombinant animal contains a different combination of alleles from either of its parents. For the BC2, only recombinant BC1 cocks were used to decrease the Araucana genome content flanking the introgressed locus.

Selection criteria were heterozygosity or homozygosity for blue eggshell colour locus, high similarity to the WL and high genetic diversity. On average, marker-assisted selection increased the proportion of the White Leghorn genome in the BC2 generation by 4.4 percent relative to the expectation (87.5 percent), thus decreasing the donor genome of Araucana accordingly. In 2019, the IC population hatched, of which 188 animals were homozygous carriers for the blue eggshell colour allele. Preliminary results of performance tests for the IC population yielded promising results. The laying rate was quite similar between the blue layer IC and commercial White Leghorn hens, while the mean egg weight was only slightly lower. The eggshell strength increased from generation to generation, but was still lower in the IC compared to the White Leghorn. Homozygous IC hens and cocks are the basis for a high performing blue egg layer line that is highly similar to the White Leghorn.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

In the first case, conserved genetic material can be used to enlarge the effective population size of the living population. Because populations of breeds at risk of extinction are small by definition, the probability that a long-term viable effective population size of at least 50 animals can be maintained *in situ* is small. Breeds with a low effective population size may consequently suffer from inbreeding depression and show a reduced selection response because of limited additive genetic variance.

When genetic material from the gene bank is available, the potential breeding population consists not only of the individuals from the current population *in vivo*, but also of the individuals with material in the gene bank. The impact of gene bank material on the effective population size can be estimated depending on the proportion of the population whose parents have material stored in the gene bank (Sonesson, Goddard and Meuwissen, 2002). The inbreeding level of offspring from animals in the gene bank sampled *k* years ago will reduce to the level of inbreeding in the population at that point in time, and the effective generation interval will consequently increase to a maximum of *k* years.

Gene bank collections play an important role in the long-term conservation of AnGR. The previous guidelines (FAO, 2012) contain recommendations on how to develop gene bank collections. These recommendations start with (i) assessing the status of AnGR populations; and (ii) determining which populations should provide material to be conserved in the gene bank. The latter step considers aspects such as cultural, societal and historical importance, genetic uniqueness, and economic importance. There is no single "correct" approach that is appropriate for all countries. Some countries will aim to conserve all breeds, whereas others will target breeds at the greatest risk of extinction. Other countries may target strategies to support the most commercially important breeds, even if they are not at risk of extinction.

Once the decision is made to conserve a population, one must decide which type of genetic material (such as semen or embryos – see Section 3) should be conserved, and from which animals and how much. The guidelines (FAO, 2012) provide detailed explanations on how to determine the amounts of material to be stored per breed in the gene bank. In most instances, these calculations were based on quantities of material required to reconstitute an extinct breed from material in the gene bank. These calculations remain valid using the most commonly available and reproductive technologies, and especially for many local breeds. There are instances, however, where the estimates may be conservative, such as for countries with access to the most advanced reproductive technologies and for breeds for which such techniques have been developed and refined. The size needed for the collection depends on how intensively the stored material is used. A good practice is that, whenever breeders use samples from the gene bank, an effort is made to obtain samples from the resulting offspring. This ensures not only that the size of the collection remains sufficiently large, but also that it keeps pace with changes in the *in situ* population.

As explained in Section 1, nowadays more attention is given to uses of gene bank collections other than the opportunity to reconstitute an extinct breed. Development and use of gene bank collections have received increased attention since the release of the previous guidelines. The possibilities of using genomics to characterize populations and gene bank collections have also increased, and software has been developed to better predict the impacts and guide the use of genetic material from gene bank collections in live populations. This section provides updates on these methods, and describes how gene bank collections can be used in management of *in situ* populations.

5.2 ANALYZING GENETIC VARIABILITY CHANGES IN A GENE BANK

5.2.1 Pedigree analysis

Analysis of the genetic variability based on pedigrees is a widely acknowledged method that can be used effectively to assess the variability of past and potential gene bank donors versus the *in situ* population. Numerous scientific publications (e.g. Maignel, Boichard and Verrier, 1996; Gutiérrez and Goyache, 2005) are available on indicators calculated for a wide array of breeds and species, with different methods described in the previous cryoconservation guidelines (FAO, 2012). The main advantage of pedigree analysis is that assessment methods are simple, low cost and robust. Pedigree analysis does not involve any additional cost other than labor, assuming that pedigree data are available. A drawback of pedigree analysis is that estimates of inbreeding and other paramters are based on theoretical expectations, and deviation from these expectation (i.e. due to random assortment of chromosomes) can only be determined with genomic data (see Subsection 5.2.2 below).

Various indicators of genetic diversity can be calculated. One of the most relevant indicators is the kinship of gene bank donors versus the live population (where kinship, also referred to as "coancestry", defined as the the probability that single alleles drawn from the same locus of each of two individuals are identical by descent from a common ancestor). For instance, one could calculate the kinship of all males with semen in the gene bank versus males in the living population. The males in the live population with low average kinship with gene bank donors may then be targeted for collection. Another approach is to estimate the kinship of the donors with the *in situ* population's main ancestors (see, for instance, Boichard, 2002), and then collect materials from the descendants of ancestors that are not yet well-represented in the gene bank.

Pedigrees can also be used to guide collection development by calculating the genetic relationships among animals in a breed's *in situ* population and gene bank donors. Cluster analysis can then be applied to the resulting data (i.e. a genetic relationship matrix) to group closely related animals, by using statistical methods like Ward's Minimum-Variance. Once the two populations are clustered, the gene bank collections and *in situ* populations can be compared to assess the gene bank for completeness and identify potential *in situ* donors that would contribute the greatest amount of genetic variation to the gene bank collection (Blackburn, 2009).

5.2.2 Molecular analysis

Molecular analysis generates more accurate estimation than pedigree analysis, and is particularly recommended for any breed with unavailable and/or unreliable pedigree data. The disadvantage is the frequent scarcity of available DNA samples from animals in the living population and almost complete absence of samples from previously living individuals. Also, when the collection of genetic material was not done in conjunction with DNA sampling, a dose of material from the gene bank must be utilized to obtain DNA and perform the analysis. Finally, there is an associated cost for DNA extraction and genotyping. In the case of breeds with no routine DNA analysis, a new field sample of the *in situ* population will also be required.

Similar to pedigrees, the most common indicators that can be calculated are kinships between gene bank donors and *in situ* breeding males or the main breeding males at the time of the study (see Eynard *et al.*, 2015 for an overview of different methods). In addition to the effective population size, changes in allele frequencies and inbreeding over time are also parameters that may be used for managing collection. At present, there is no consensus on how to calculate the above-mentioned indicators from the molecular data, but most approaches provide similar results when applied to the same data. Box 5.2 briefly describes how molecular data is currently being used for management of the collection in the French gene bank.

BOX 5.2
Utility of genomic analysis for an existing collection

The French National Cryobank was set up in 1999 and currently (2022) preserves reproductive biological material from 241 breeds of 12 different species of livestock (in addition to material from nine aquaculture species). For rare pig and goat breeds, a large collection has been established, which was set up mostly for long-term conservation. The samples in the collection were gathered at the beginning of the conservation programmes, and very little information about the male animals was recorded at the time. The recent development of genomic tools offers a unique opportunity to answer some very essential questions for the conservation programmes, such as whether the founder animals were related, if they represent the history of the breed as individuals, and how well they are represented in the *in situ* population.

Genetic variability indicators such as molecular kinship have been calculated from genotyped data to answer these questions and to aid the breeder associations. The results allowed the breeders to optimize the breeding of the next generation of males to be sampled for the gene bank. The new generation of male donors was chosen by targeting breed origins that had not been previously sampled for the gene bank, based on molecular analysis of both gene bank material and animals from the living population.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

5.3 NEW DEVELOPMENTS IN GENOMICS

Since the turn of the millennium, sweeping advancements have been made in the ability to study the genomes of organisms. Box 5.3 includes a glossary of commonly used methods in genomic analysis. Prior to this, molecular analyses were based on single loci or small sets of genetic markers on relatively coarse genetic maps. The advent of high-density SNP arrays has fundamentally changed genome analysis and insights into genetic diversity. With these arrays, genetic analysis is no longer restricted to individual loci or imprecise maps, but can be extended to the entire genome (Mei *et al*., 2000). Among the most important applications that have emerged from this in the field of animal science are the estimation of the genetic value of an individual by using thousands of typed markers and the subsequent use of these estimated breeding values in efficient selection programmes, better known as “genomic breeding value estimation” and “genomic selection” (Meuwissen, Hayes and Goddard, 2001). It has also become possible to study the genetic architecture of important traits in more detail through so-called genome-wide association studies (GWAS). Last but not least, the realized genetic relationships derived from the marker data (VanRaden, 2008) can be used for more sustainable management of genetic diversity (Sonnesson *et al.*, 2012). The FAO *Draft practical guide on genomic characterization of animal genetic resources* (FAO, 2021) provides more detailed information about the utilization of genomic data in the assessment of the diversity of animal genetic resources. Box 5.4 provides two examples of the use of pedigrees and/or genetic markers in evaluation of gene bank collections.

BOX 5.3

A glossary of the most commonly used methods in genomics

Whole genome shotgun sequencing: Whole genome shotgun sequencing (WGS) or short-reads/next-generation sequencing (NGS) is a technology that decomposes the whole DNA sequence of an individual into short fragments. Those fragments are subsequently sequenced to a defined length (often 50-150 bp), either from one side (single-reads) or from both sides, generating a connected pair of reads (paired-end sequencing). With such reads, the assembly of the whole genome becomes possible by merging reads via overlapping patterns into larger segments, called contigs (shorter) and scaffolds (longer), and eventually chromosomes, to generate a reference genome.

Long-read sequencing: One of the latest developments is long-read sequencing, using single-molecule real-time sequencing (SMRT) or nanopore technology. The output of both technologies consists of very long sequence reads, with a length up to several kilobases. Those technologies hold great promise in the successful detection of large structural variants, such as huge deletions and assembly of complex chromosomal regions (Miga *et al*., 2020).

Resequencing: This method aims at sequencing of an individual's genome in order to detect differences between the individual and the standard reference genome of the species. Sequence alignment can detect many sites of variation in genes and intergenic regions for studying functional genomics or genetics differentiation.

SNP array: A single nucleotide polymorphism (SNP) is a DNA sequence variation, occurring when a single nucleotide in the genome differs among individuals, and can therefore be used as a marker for the underlying linked haplotype. These markers are normally bi-allelic SNP markers. With a SNP array, a high number of such markers can be typed from the extracted DNA, usually from several animals (typically 48 or 96) in parallel. Common SNP arrays contain marker numbers ranging from a few thousand to more than a million, depending on the species. The chips currently (i.e. 2022) used in genomic selection belong to the middle category (around 50 000). The typing of an animal costs about USD 50 with decreasing tendency. Naturally, these arrays contain a predefined set of variations, which are selected according to certain criteria from a reference population (Kranis *et al*., 2013). These criteria may include, for example, a certain distribution of the minor allele frequencies or an equidistant coverage of the genome.

Note that due to the technical characteristics of the methods, the results obtained may differ significantly between full sequences and selected panels of SNP on arrays. This is because genotyping arrays contain sets of pre-ascertained SNPs, which may result in bias.

BOX 5.4

Complementary use of pedigrees and molecular information to evaluate the genetic variability in Holstein cattle gene banks

A. Analysis of Holstein gene banks

France, the Netherlands, and the United States of America (USA) all maintain Holstein-Friesian (HF) gene bank collections. Genetic variability of the collections within and between countries was assessed and compared with active male populations in each country by using pedigree data (Danchin-Burge, Hiemstra and Blackburn, 2011). Measures of genetic diversity such as probability of gene origin, inbreeding and kinship were calculated. The three gene banks have captured significant amounts of genetic diversity for the HF compared with the current populations. Although a substantial part of the USA, French and Dutch collections seems to be genetically similar, the USA collection in particular represents an interesting reservoir of HF genes of the past which is not present in the current *in situ* population.

B. Change in genetic diversity in the Dutch Holstein population determined by pedigree and genomic data analysis of gene bank samples

A recent study in the Netherlands used pedigree and genotype data of more than 6 000 bulls to assess trends in genome-wide inbreeding and kinship (Doekes *et al*., 2018a,b). Gene bank samples contributed to the study. The study estimated inbreeding trends in specific chromosomal regions by detecting runs of homozygosity (ROH) and changes in allele frequency over time. Two major points of inflection were observed in the estimated trend of genetic diversity. Around the year 2000, inbreeding and kinship both temporarily decreased. Then, from 2010 onwards, they began to steeply increase, with estimates of inbreeding rates up to 2.8 percent per generation, depending on the method used. The amount of inbreeding varied according to the genomic region. A large proportion of the marker alleles had changes in frequency that could not be explained by random genetic drift. Although cause and effect could not be proven, the decreases in inbreeding observed after 2000 corresponded to the introduction of optimal contribution selection and a shift in the breeding goal. The increases in rates of inbreeding and kinship occurring after 2010, on the other hand, corresponded closely with the adoption of genomic selection. The observed trends in genetic diversity reflect major changes in the Dutch-Flemish HF breeding programmes over the past 30-plus years.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

A drawback of SNP arrays is their limitation to a non-random set of SNP markers which were selected (ascertained) during the design process. This leads to under-representation of globally rare variants on SNP arrays compared to whole genome resequencing data, a phenomenon known as SNP "ascertainment bias" (Nielsen, 2004; Lachance and Tiskoff, 2013). Some array design schemes further increase this bias by intentional over-representation of common SNPs with regard to the natural background distribution. To a degree, this bias can be addressed by constructing sub-panels composed of minor alleles (Blackburn *et al*., 2014). A direct implication for monitoring genetic diversity is that this bias towards common alleles increases heterozygosity estimates, and populations therefore show an upwardly biased amount of genomic variation, particularly for the populations from which the SNP were ascertained.

Within this design process, such target breeds are screened for variant sites that are especially informative in the target breeds. Such sites may not necessarily be as informative in other, non-target breeds, and it is therefore inherent that important variation in such alternative breeds is not represented in the panel. This might be, for instance, be the case for variants that were lost in highly selected commercial breeds. The limitation to variants of a small number of populations additionally introduces the problem that populations that are distantly related to the populations for which the array was originally developed show reduced estimates of genomic variability compared to reality, which is not the case with whole genome sequencing (WGS) data (Malomane *et al.*, 2018). This population-specific ascertainment bias also inflates genomic estimates of genetic distinction between populations. Ascertainment bias may therefore have a consequence for gene banking, if breeds are prioritized for collection based on genomic estimates of genetic variation. Gene bank managers should be aware of this possible bias. Whole genome sequencing offers a mechanism to alleviate biases that are introduced with smaller SNP panels (Eynard *et al.*, 2015).

Although SNP technology is a widespread standard approach and has generated extensive knowledge, it also has disadvantages relative to WGS besides the ascertainment bias. For example, the detectable variants are usually limited to SNPs. Unfortunately, this class of variants can only explain part of the genetic variance of traits. Variants such as insertions and deletions of nucleotides and inversions of genomic regions may not be detected. Therefore, sequencing (either WGS or specific sequencing of target genomic regions) may need to be applied if selection of donors is to be based on these other types of genomic variants.

Today, many breeding programmes in most major livestock species in industrialized countries incorporate the use of genomic data, and genotyping with an SNP array is by far the most routinely used strategy to generate that data. This option also merits consideration for gene banking and may present an opportunity to use genotypic data that has already been generated.

To facilitate genotyping of animals in gene banks, a multi-species SNP array has been developed (see Annex 5.1). By using this array (or any comparable assay) donor animals can be selected on the basis of the real diversity they carry instead of just expectations made from pedigree evaluations. For example, full sibs are expected to have 50 percent of additive genetic variation in common; in reality this will vary around 50 percent. When selecting donors from among a group of full sibs, results from an SNP array can be used to determine the optimal animals to sample to capture the maximum amount of genetic variation. Also, a library of variation in the gene bank will be known, and can be of immense importance in identifying the genetic originality of individuals when samples from the gene bank are subsequently used for breeding in the *in situ* populations. Especially for the latter aspect, complete genotyping of all samples is valuable. Examples of the use of genotype information are presented below in Subsection 5.6 about software programs.

It must be noted, however, that in addition to financial aspects, genomic analysis has costs in terms of time, and requires personnel that are capable of performing the required analyses.

5.4 SOFTWARE FOR MANAGEMENT OF GENE BANK COLLECTIONS

Today, interested animal breeders or conservation geneticists can, for many species, obtain high-throughput genotyping data at a reasonable cost, and exploit these data in breeding and conservation decisions. Apart from the general tasks of data and herd book management and breeding value

estimation, breeding programme design and strategies for managing diversity are major concerns of a more general interest. In utilization of gene bank collections, decisions on the use of collected samples can be supported by such approaches. The following section presents two examples of software for management of animal genetic diversity. The first one is the Modular Breeding Program Simulator (MoBPS), a flexible simulation framework to simulate breeding programs and thus evaluate the impact of breeding decisions on a population. The second example demonstrates how optimal contribution software can be used to manage genetic selection programmes while efficiently limiting inbreeding.

5.4.1 Modular Breeding Program Simulator (MoBPS)

Breeding programmes aim at improving the genetic properties of livestock populations with respect to a given goal, such as increased productivity, fitness and adaptation or some combination of traits. Progress towards the target is limited by the available resources, but also by negative effects of selection such as inbreeding depression, decreased fitness or loss of genetic diversity. These effects should thus be minimized, as much as possible. Additionally, population history, such as fluctuating population size and selection pressure, has an impact on the current genomic architecture and thus the potential for future improvement. Hence, the allocation of resources to gene banks and design of a breeding programme are complex optimization problems.

MoBPS is an *R*-package to perform stochastic simulation of breeding programmes, and thus assist breeders in evaluating and optimizing their breeding programmes (Pook, Schlather and Simianer, 2020). Similar to the gene-flow concept introduced by Hill (1974), MoBPS allows grouping of individuals into cohorts that have similar characteristics such as age, sex and genetic origin. Thus, MoBPS provides a highly flexible tool to allow detailed modelling of today's complex breeding programmes, which may include cohorts of animals with material stored in gene banks. MoBPS includes a variety of pre-implemented functions for common breeding practices such as optimum genetic contribution selection and estimation of genomic breeding values. Although breeds subject to conservation are often small, MoBPS allows for the simulation of breeding programmes with millions of animals, or population genetic studies with thousands of generations.

To make the MoBPS simulation framework accessible to a wider audience like breeders and scientists with limited experience in programming or as a tool for teaching, it also includes a front-end user interface that can be accessed via a web-browser at www.mobps.de (Pook *et al*., 2020) that includes most of the functionality of the *R*-package.

One example for the use of MoBPS is the simulation of a backcross breeding programme to reconstitute an extinct breed from gene bank material, as was described in the previous guidelines (FAO, 2012). In such a breeding programme, only semen has been stored in the gene bank. Females from another breed must thus be inseminated with cryoconserved semen of the breed to be recreated. The offspring are inseminated again with semen from the gene bank, and this is repeated until the offspring is "pure" enough (4 generations) to be considered as the original breed. This scheme has been successfully simulated with the MoBPS program (see Figure 5.2). Results showed that the purity of the recreated breed closely followed the expectation for each generation (e.g. 50 percent for BC1, 75 percent for BC2 and 87.5 percent for BC3, etc.) with hardly any variance.

The input modules in MoBPS allow users to adjust parameters, such as pregnancy and survival rates, to account for differences among species, breeds and production systems. By doing so, gene bank managers can more precisely estimate the quantities of material to be collected for the prevailing

circumstances in their country, and even explore the impacts of changing these parameters on the probability of successful breed reconstitution.

Such an exercise may be particularly helpful for the development and utilization of collections consisting of older material and/or material from local breeds, for which knowledge about pregnancy and survival rates with cryopreserved material is typically limited. Although more research is required on such populations to both improve pregnancy and survival rates and to obtain accurate estimates of their values, in the meantime using simulation to perform sensitivity analyses can help a gene bank manager optimize collection procedures to obtain satisfactory rates of breed reconstruction success at a reasonable cost.

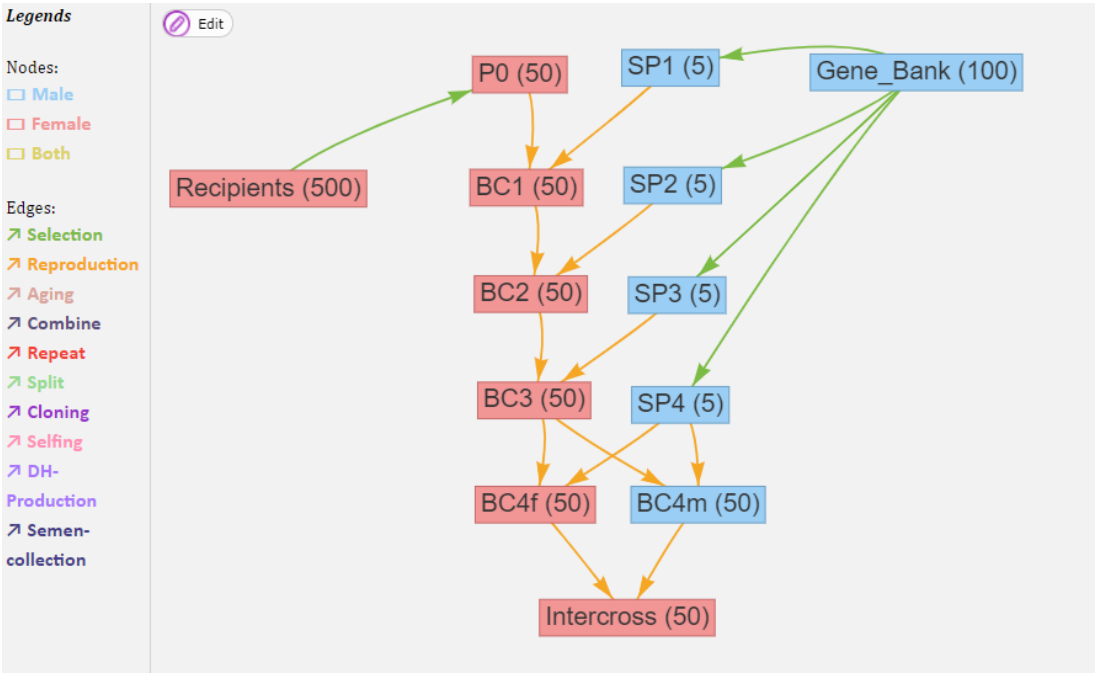

FIGURE 5.2

Set up of the MoBPS program to simulate reconstitution of a breed by backcrossing with semen from a gene bank

Note: Females are selected from another breed (recipients) to be inseminated by sperm (SP) stored in the gene bank. Each resulting backcross (BC) is inseminated again with sperm from the gene bank until the 4th backcross generation from which males and females are mated to produce the final generation (intercross) that is considered as the recreated breed. In brackets, the number of individuals used in the breeding programme is shown.

Simulations can be executed directly from the web-interface, as the corresponding *R*-package, which acts as the back-end simulator, is directly linked to it. After those simulations are done, a variety of functions to analyse and compare different breeding programmes with regard to breeding objectives can be performed (see Figure 5.3).

A key strength of a simulation approach lies in the fact that, in contrast to real-world experiments, far less time and money is needed. Also, potential harm to animals, such as adverse fitness effects, can be avoided. Furthermore, experiments can be repeated and modified, which leads to much greater statistical power when comparing scenarios. A huge advantage of stochastic simulations is that variation can be studied by repeating the process many times, which is very helpful for risk analysis. For example, by chance, low reproduction or unbalanced sex ratios may decrease numbers of offspring in a breed reconstitution programme. Simulation allows one to observe the likelihood of failure due to such

problems. Even if the absolute value for estimated effects, such as genomic gain, inbreeding rate, etc. might be slightly incorrect due to simplifications of reality, these effects should usually affect all scenarios considered and thereby still ensure comparability.

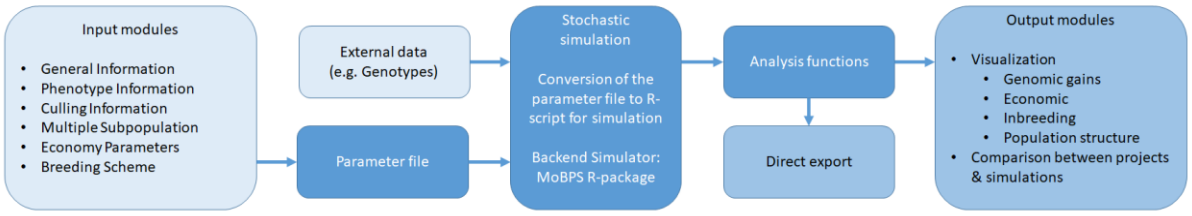

FIGURE 5.3
Setup of the MoBPS simulation framework

A key strength of a simulation approach lies in the fact that, in contrast to real-world experiments, far less time and money is needed. Also, potential harm to animals, such as adverse fitness effects, can be avoided. Furthermore, experiments can be repeated and modified, which leads to much greater statistical power when comparing scenarios. A huge advantage of stochastic simulations is that variation can be studied by repeating the process many times, which is very helpful for risk analysis. For example, by chance, low reproduction or unbalanced sex ratios may decrease numbers of offspring in a breed reconstitution programme. Simulation allows one to observe the likelihood of failure due to such problems. Even if the absolute value for estimated effects, such as genomic gain, inbreeding rate, etc. might be slightly incorrect due to simplifications of reality, these effects should usually affect all scenarios considered and thereby still ensure comparability.

5.4.2 Optimum contribution software

As discussed throughout these guidelines, the genetic material stored in the gene bank can serve different purposes. Regardless of the purpose, it is essential that genetic diversity in the gene bank is maximised, given restrictions such as the number of animals that can be sampled. Optimum contribution theory and associated software can assist in achieving this goal.

Optimum contribution selection was originally developed to determine the optimum number of offspring of each breeding individual in a selection programme subject to certain constraints (Meuwissen, 1997). The approach is equally applicable to management of gene banks (see Box 5.5). A common objective in a commercial breeding programme is to maximize the average breeding value while restricting the rate of increase in the average kinship and inbreeding. Another constraint in a breeding programme is that the total contribution of each sex must be 50 percent. For a gene bank, one logical objective may be to minimize the average kinship when sampling a fixed number of individuals from the live population.

With the development of genotyping, kinships can now be reliably estimated both by using genotypes and by using pedigrees. If pedigrees are used, care must be taken to ensure pedigree completeness because predictions of optimum contributions will otherwise be biased. If genomic kinships are used, missing genotypes can be problematic and methods that combine pedigree and genomic information may be needed. When using genotypes, the use of segment-based kinships, which can be computed from phased marker data, is recommended. This way, the segment-based kinship of two individuals equals the expected proportion of an offspring genome that is included in ROH. Different programs are available for optimum contribution selection, and include the free *R*-package optiSel (Wellmann,

2019), the free software EVA (Berg, Nielsen and Sørensen, 2006), the free software Gencont (Meuwissen, 2002) and the commercial software TGRM (Kinghorn, 2011).

Constraints due to sex and the rate of increase in kinship can influence the results. For example, conserved genetic material can be used for the recovery of "native" genomes of the local breeds that have undergone "upgrading". Many local breeds have been subject to generations of occasional or systematic crossing with other breeds. Continued crossing eventually leads to the genetic extinction of the local breed because the native alleles are lost. To prevent this, breeders (or other stakeholders) may wish to remove the introgressed "foreign" genetic material from the genomes of the local breed as much as possible. Older individuals of local breeds not only tend to have less foreign genetic material, but the foreign haplotypes also tend to be longer. Long haplotypes can more easily be removed than short haplotypes. The previously conserved genetic material from animals born years ago is therefore of high value for a breeding programme that aims at removing the foreign haplotypes from the population.

BOX 5.5

Using gene bank material to optimize selection in Creole cattle

Creole, or "Criollo" cattle are locally adapted cattle found throughout Latin America. Although they directly descend from breeds that originated in southwestern Europe during the latter half of the previous millennium, they are often considered "native" to the Americas due to their long-term presence there.

A recent study was undertaken to improve the management of the Creole cattle in Colombia by evaluating the *in situ* population, as well as the usability and usefulness of the samples in the local gene bank. The breed Blanco Orejinegro was studied in detail. The first step was to assess the genetic diversity of all available samples. In the follow-up steps, the gene-flow approach was used to model and optimize introgression and conservation schemes.

The optimum contribution method was used to develop a hypothetical breeding programme to maximize genetic gain for growth traits while constraining inbreeding in the Blanco Orejinegro breed, and while including the use of semen from the gene bank. Around 50 females and 50 males are registered in the Blanco Orejinegro herdbook each year. The gene bank collection consists of semen of 104 bulls, of which 28 have genotype data and are no longer part of the active population.

The value of the bulls in the gene bank was assessed in terms of genetic merit and diversity, by simulating an optimum contribution selection scheme with the 28 cryobank bulls as selection candidates, in addition to the current population. The MoBPS software was used. The simulation showed that the growth traits could be improved while controlling the level of inbreeding. In the simulated population, both the genetic gain and variability were improved by applying optimum contribution selection and using semen from the gene bank.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

The objective of such a selection programme could be to maximize the proportion of the local breed genome, while constraining the increase in the average kinship of native alleles below a certain threshold. An alternative is to reduce the frequencies of alleles associated with the foreign breed, while

maximizing diversity in the rest of the genome. For both procedures, the *R*-package *optiSel* can be used.

In many instances, local breeds that have been subject to crossbreeding will not have material from previous generations stored in a gene bank. Nevertheless, cryoconservation can still play an important role. In such cases, animals within the breed that have remained relatively free from crossbreeding could be targeted for collection for the current gene bank. This process would facilitate the use of the material by today's breeders for matings in the *in situ* population (now and in the future) to accelerate a breeding programme designed to purge foreign haploptes.

5.5 GENE BANKS AS A RESOURCE FOR RESEARCH

In addition to being used to directly support the maintenance and genetic improvement of animal populations, gene banks can be a vehicle to support research. Clearly, studies can be undertaken in gene banks to develop, test and adapt methods for collection, cryopreservation and utilization of germplasm. In addition, gene bank collections can serve as a source of material for genetic and genomic studies of population genetic variation and its changes over time.

Cryopreservation has been widely used in cattle since the 1960s. Therefore, the implementation of gene bank collections has a long history in some countries. According to a survey carried out during the IMAGE project, (Passemard *et al*., 2018), the average onset of sampling for the collections in European cattle gene bank collections was 2002, which is quite recent. However, the oldest collection was sampled in 1963 for the Meuse-Rhine-Yssel cattle (MRY) breed in the Netherlands, and there are also samples for three other cattle breeds that were preserved in 1966 (Dutch Friesian, Polish Red and Original Austrian Brown cattle).

Some gene bank collections are highly interesting since they are a unique representation of breed evolution over time. Box 5.6 summarizes the results of a survey in which European gene banks reported on the inventory of their collections with respect to the species and breeds represented, and the history of collection activities. Box 5.7 reports on a study that examined retrospectively how the past use of semen from a gene bank bull subsequently affected the genetic variation of the *in situ* population. Box 5.8 refers to two studies (from the United States of America and Spain, respectively) where material in gene banks were used as raw material to compare historical changes in breeds, including change that occurred through genetic and geographic isolation and via natural and/or artificial selection. The studies demonstrated that gene bank collections often differ substantially from the live population. Similar results have been previously reported by Danchin-Burge, Hiemstra and Blackburn (2011).

BOX 5.6

European gene banks as gene archives

A survey was conducted during the Horizon 2020 European Union project Innovative Management of Animal Genetic Resources (IMAGE) study about gene bank collections. Some managers gave detailed answers about their collections, which revealed that there are at least 92 breeds from five distinct species that are sampled every year: cattle (62 breeds); goats (7 breeds); horses (2 breeds); pigs (8 breeds); sheep (13 breeds). These collections are kept in eight different European countries (Austria, France, Germany, Hungary, Iceland, Italy, Poland, and Spain) and 14 different organizations, and the time of first sampling varies between 1966 and 2014, and the average year of first sampling was 1997. These data show that some collections are likely to be distinctly different from the current *in situ* populations. The gene pool of a breed changes constantly over time due to genetic drift (which particularly affects rare breeds due to their small numbers), selection and inbreeding. These changes do not affect all breeds and species in the collections equally. For instance, due to their longer generation interval, the genetic diversity in cattle collections will not change as rapidly as collections for pigs, which have a short generation interval. On the other hand, the selection pressure in dairy cattle is stronger than in horses or sheep, but this may be compensated by a large census size such as in Holstein cattle. Therefore, the only way to accurately determine if a gene bank collection differs radically from a live population is to assess its genetic variability, by using either pedigrees, genotypes, or a combination of the two.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

BOX 5.7

Assessing the impact of the use of an old cryopreserved bull on the genetic variability of a breed

The aim of this case study was to assess the extent to which the use of old cryopreserved material can support reintroduction of genetic variability within a given breed. To do so, the impact of using cryopreserved material from a bull born in the 1970s to inseminate cows during the years 2004 to 2007 in a regional transboundary dairy cattle breed, the Abondance, was examined. Molecular data (50 000 SNP and high-density chips) as well as pedigree information were available for the cryopreserved bull as well as for recent reproducers (bulls, as well as a significant number of cows born over the last two years), including his descendants. Genealogical and molecular approaches were used in a complementary manner to evaluate the consequences of this reintroduction on neutral and genome-wide diversity of the breed. The results showed a favourable impact for the genetic variability of the breed, as well as some desireable genetic change for specific traits (mostly functional traits) which counterbalanced from an economic perspective the loss of genetic gain in production traits. However, one should bear in mind that this favourable result was obtained in a breed with annual genetic gains that are relatively small in comparison with large international breeds.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

BOX 5.8

Using gene banks samples to evaluate genetic changes of a breed

A. A genetic investigation of Island Jersey Cattle, the foundation of the Jersey breed

The genetic difference between the founding population of Jersey cattle (from Jersey Island) and non-Island Jersey cattle was analyzed by Huson *et al.* (2020). Samples from Jerseys raised in the USA (n = 49) and on Jersey Island (n = 34) were obtained from the gene bank of the United States of America. The cross-section of bulls were born from the 1960s to 2000s, and were lowly related to one another. The study revealed that the two Jersey populations had similar degrees of inbreeding, despite their vastly different census sizes. Signatures of past selection were revealed and demonstrated that the two different Jersey populations differed in terms of the genomic locations of key ROH regions. These data provided the first insights into the divergence of two subpopulations of the Jersey breed over decades of isolation between its place of origin and the United States of America.

Cattle samples by breed per birth decade

Population	Decade of Birth							
	1950	**1960**	**1970**	**1980**	**1990**	**2000**	**Unknown**	**Total**
Jersey_ISL		2	8	10	21	8	0	49
Jersey_USA	1	3	3	5	18	4	0	34
Total	1	5	11	15	45	17	1	95

B. Annotation of selection signatures in the bovine breed Asturiana de los Valles

This study was implemented to demonstrate the usefulness of gene banks for detection of recent selection and annotating signatures of historical selection. It was based on a Spanish autochthonous beef cattle breed, the Asturiana de los Valles, which is raised under semi-extensive breeding conditions. The gene bank collection enabled the analysis of evolution of genetic diversity across 35 years, from 1980 to 2015.

Generation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
First year	1980	1984	1988	1992	1996	2000	2004	2008	2012
Final sample size	0	4	8	13	17	28	29	9	9

The genome data analysis detected selection signatures at different sites which appeared over time. It revealed candidate genes for meat and milk production, immunity and olfaction. The study showed that time series material stored in gene banks served as rich information source in research on breed history and biology.

Source: Innovative Management of Animal Genetic Resources (IMAGE). www.imageh2020.eu

5.6 RECOMMENDATIONS FOR GENE BANK UPDATING

Gene banks are multi-functional. Because cryopreserved material retains its viability for many decades, some gene bank managers (i.e. stakeholder advisory boards) may be tempted to not update the conserved material once a quantity sufficient for reconstitution of an extinct breed has been cryopreserved. However, *in situ* populations are constantly evolving and therefore gene bank managers and breeding organizations should regularly monitor and update their gene bank collections, especially if the collection is being actively used to support breeding in the *in situ* population (see Figure 5.1). As a rule of thumb, the faster the breed develops and changes, the more often the gene bank collection should be evaluated and updated, if necessary. Estimates of how often this should be done have been made, and indicate roughly every 4 to 7 generations (see Box 5.9 and Blackburn, 2018).

In the case of endangered breeds, one goal is to store material from a sufficient number of animals needed to restore the breed's genetic variability. So far, a limited number of breeds have reached this goal (Leroy *et al*., 2019). Eventually, establishing breed reconstruction as the exclusive objective can become an obstacle in the utilization of gene bank material, as farmers could then be reluctant to use genetic material not corresponding to the breed's objectives. For example, an assessment was performed of dairy cattle breeds in French gene bank collections after 10 years of storage. Selected breeds are sampled on two principles: the first one is to gather a snapshot of the breed from a genetic variability point of view for a given year; and the second is to sample extreme bulls for various traits (Verrier *et al*., 2003). The assessment showed that for breeds intensively selected for a trait, such as milk yield or protein content, extreme bulls for these traits were quickly surpassed by contemporaries. Therefore, the gene bank sampling goals were changed for these breeds; animals were sampled according to their breeding values with extremes that differed according to the rate of change in traits subject to selection. For instance, for intensely selected traits such as milk production or stature, Holstein bulls are included in the cryobank if and only if their estimated breeding value is vastly superior to the average (based on the population of bulls in artificial insemination programmes) for the particular trait, on a given year. On the other hand, for traits such as fertility that have a lower genetic progress, the threshold is lowered.

In conclusion, the following recommendations can be used for establishing and maintaining a gene bank:

- Although some prioritization may be necessary, whenever feasible gene bank collections should be established as early as possible for all breeds, regardless of their census size. These collections can then be used to support breeding in the *in situ* population, establish new breeds or sub-populations and to provide a resource for research, in addition to "backing up" populations at risk of extinction.
- Maintenance of an *ex situ* collection will allow breeders and other stakeholders to select the most interesting males to be cryopreserved, while maintaining a sufficient degree of genetic variability in the combined *in situ* and *ex situ* "populations".
- Genomcs are a powerful tool for management of gene bank collections. Genomic data provide greater precision of estimates of genetic variation than do pedigrees. Costs of genomic analyses have decreased substantially in recent years, so assessment of all gene bank donors is recommended whenever feasible. Genomic analysis requires particular expertise, however, so specific capacity building is needed (see Section 10).

- Material from individuals in gene bank collections can be used to generate improved offspring which may better meet the recent breeding objectives and farmer expectations. When multiple offspring are available, genomics or evaluation of estimated breeding values can be to determine the best ones to be susbsequently sampled for the gene bank collection.
- The gene bank collection should be compared to the *in situ* population on a regular basis, and the gene bank collection should be updated with new material based on both genetic variability and genetic merit with respect to the traits in the breeding objectives. Collection activities should continue even if the collection has already met FAO standards for breed reconstitution. Box 5.9 reports on the results achieved by the gene bank of the United States of America through their procedures of continual updating the collection of material for the Guernsey breed of dairy cattle.
- Genetic simulation software such as MoBPS can be used to improve the management of gene bank collections.
- New knowledge and tools for measuring fertility rate in cryoconserved material are needed, since the current knowledge about pregnancy rates obtained with frozen semen is incomplete for many species and breeds, especially for avian species. This knowledge can be used as parameters in simulation studies to better establish collection goals and perform risk analysis.
- For intensively selected breeds, assessment of the collection may also be used to evaluate specific changes in genetics of the breed, for example, depending on: (i) the level of intensity of the selection for a given trait; (ii) the variability and extent of changes applied on selection goals over time; and (iii) the existence of similar collections of the breed in other countries.

BOX 5.9

Resampling populations to keep the collection current

A common criticism of gene banks is that the collections can become dated and thus lose utility over time, or make it more difficult to use due to its genetic differences with the existing *in situ* populations (Leroy *et al.*, 2011). This issue can be addressed by executing periodic sampling over time and collecting material from a wide range of animals, including genetically superior donors (Blackburn, 2018). Under a resampling protocol, gene banks would decide how frequently to resample a breed dependent upon the genetic change that is occurring.

For example, Guernsey is an "at risk" breed in the United States of America, but nevertheless maintains an active genetic improvement programme. The gene bank has samples from a time continuum from 1948. Figure 5.4 illustrates how resampling over time has enabled the collection to keep pace with the *in situ* population with respect to the predicted transmitting ability (PTA) for milk yield. The graph also illustrates the diversity collected by comparing individuals from the gene bank to the annualized *in situ* breed mean and ± two standard deviations for milk production. There are several years where animals in the collection exceed the *in situ* population by more than two standard deviations. Additionally, the graph shows that, while the breed is increasing for PTA for milk, the rate of increase has been moderate, such that the PTA for some bulls collected more 30 years ago have PTA greater than the current breed average. These data suggest that by selecting a semen from a diverse group of bulls, some animals collected years ago remain competitive for use within the current *in situ* population.

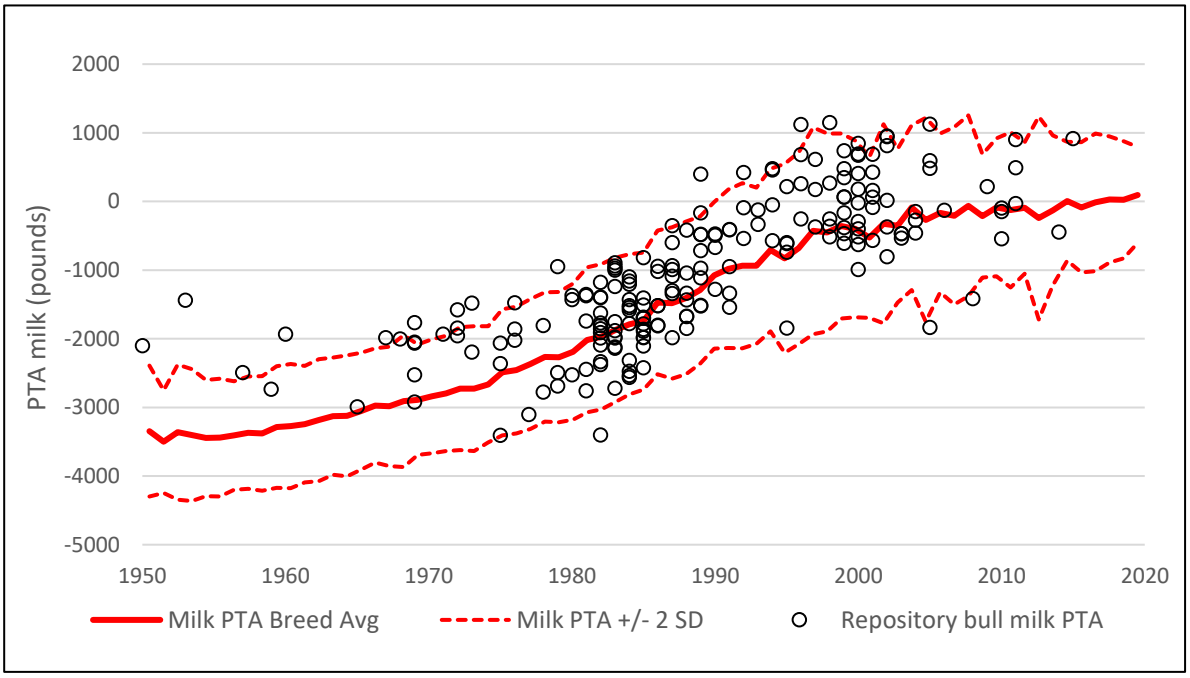

FIGURE 5.4

Distribution of predicted transmitting abilities (PTA) for milk yield of Guernsey bulls stored in the gene bank collection of the United States of America, relative to the distribution (mean and ±2SD) of the general population

Source: Harvey Blackburn

5.7 REFERENCES

Berg, P., Nielsen, J. & Sørensen, M.K. 2006. EVA: Realized and predicted optimal genetic contributions. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production.* Belo Horizonte, Brazil. 13-18 August 2006. pp. 27-09 ref.8

Berg P. & Windig J.J. 2017. Management of cryo-collections with genomic tools. In K. Oldenbroek, ed. *Genomic management of animal genetic diversity*, pp. 155-178. Wageningen, Netherlands, Wageningen Academic Publishers.

Blackburn, H.D. 2009. Genebank development for the conservation of livestock genetic resources in the United States of America. *Livestock Production Science*, 120(3): 196–203. https://doi.org/10.1016/j.livsci.2008.07.004

Blackburn, H.D., Paiva, S.R., Sollero, B.P., Biegelmeyer, P., Caetano, A.R. & Cardoso, F.F. 2014. A dedicated SNP Panel for evaluating genetic diversity in a composite cattle breed. *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, 17-22 August 2014.

Blackburn, H.D. 2018. Biobanking genetic material for agricultural animal species. *Annual Review of Animal Biosciences*, 6: 69-82. https://doi.org/10.1146/annurev-animal-030117-014603

Boichard, D. 2002. Pedig: a Fortran package for pedigree analysis suited to large populations. *7th World Congress on Genetics Applied to Livestock Production*, Comm No. 28-13. Montpellier, France. August 19-23, 2002.

Danchin-Burge, C., Hiemstra, S.J. & Blackburn, H. 2011. *Ex situ* conservation of Holstein-Friesian cattle: Comparing the Dutch, French, and US germplasm collections. *Journal of Dairy Science*, 94(8): 4100–4108. https://doi.org/10.3168/jds.2010-3957

Dechow, C.D., Liu, W.S., Specht, L.W. & Blackburn, H. 2020. Reconstitution and modernization of lost Holstein male lineages using samples from a gene bank. *Journal of Dairy Science*, 103(5): 4510-4516. https://doi.org/10.3168/jds.2019-17753

Doekes, H.P., Veerkamp, R.F., Bijma, P., Hiemstra, S.J. & Windig J.J. 2018a. Value of the Dutch Holstein Friesian germplasm collection to increase genetic variability and improve genetic merit. *Journal of Dairy Science*, 101(11): 10022-10033. https://doi.org/10.3168/jds.2018-15217

Doekes, H.P., Veerkamp, R.F., Bijma, P., Hiemstra, S.J. & Windig, JJ. 2018b. Trends in genome-wide and region-specifc genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. *Genetics Selection Evolution*, 50(15). https://doi.org/10.1186/s12711-018-0385-y

Eynard, S.E., Windig, J.J., Leroy, G., van Binsbergen, R. & Calus, M.P.L. 2015. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genetics*, 16: 24. https://doi.org/10.1186/s12863-015-0185-0

FAO. 2012. *Cryoconservation of animal genetic resources.* FAO Animal Production and Health Guidelines No. 12. Rome. (also available at https://www.fao.org/3/i3017e/i3017e00.pdf).

FAO. 2021. *Draft practical guide on genomic characterization of animal genetic resources*. *CGRFA-18/21/10.2/Inf.2.* Rome. (also available at https://www.fao.org/3/ng883en/ng883en.pdf).

Gandini, G. & Oldenbroek, J.K. 2007. Strategies from moving from conservation to utilisation. In K. Oldenbroek, ed. *utilisation and conservation of farm animal genetic resources*, pp. 29-54. Wageningen, Netherlands, Wageningen Academic Publishers.

Gutiérrez, J.P. & Goyache, F. 2005 A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding and Genetics,* 122(3):172-176. https://doi.org/10.1111/j.1439-0388.2005.00512.x

Hill, W.G. 1974. Prediction and evaluation of response to selection with overlapping generations. *Animal Production,* 18(2): 117–139. https://doi.org/10.1017/S0003356100017372

Huson, H.J., Sonstegard, T.S., Godfrey, J., Hambrook, D., Wolfe, C., Wiggans, G, Blackburn, H. & VanTassell C.P. 2020. A Genetic investigation of Island Jersey Cattle, the foundation of the Jersey breed: Comparing population structure and selection to Guernsey, Holstein, and United States Jersey cattle. *Frontiers in Genetics,* 11:366. https://doi.org/10.3389/fgene.2020.00366

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P. Hocking, P., Fife, M., Salmon, N., Fulton, J., Strom, T., Haberer, G., Weigend, S., Preisinger, R., Gholami, M. Qanbari, S., Simianer, H., Watson, K., Woolliams, J. & Burt. D. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics,* 14(59). https://doi.org/10.1186/1471-2164-14-59

Kinghorn, B.P. 2011. An algorithm for efficient constrained mate selection. *Genetics Selection Evolution,* 43(4). https://doi.org/10.1186/1297-9686-43-4

Lachance, J. & Tishkoff, S.A. 2013. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays,* 35(9): 780-786. https://doi.org/10.1002/bies.201300014

Leroy, G., Danchin-Burge, C. &Verrier, E. 2011. Impact of the use of cryobank samples in a selected cattle breed: A simulation study. *Genetic Selection Evolution*, 43(36). https://doi.org/10.1186/1297-9686-43-36

Maignel, L., Boichard, D. & Verrier, E. 1996. Genetic variability of French dairy breeds estimated form pedigree information. *Proceedings of the open session of the Interbull annual meeting*, 23-24 June 1996, Veldhoven, The Netherlands. https://journal.interbull.org/index.php/ib/article/view/328

Malomane, D.K., Reimer, C., Weigend, S., Weigend, A., Sharifi, A.R. & Simianer, H. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics*, 19(22). https://doi.org/10.1186/s12864-017-4416-9

Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J. & Lockhart, D.J. 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Research,* 10(8): 1126–1137. https://genome.cshlp.org/content/10/8/1126

Meuwissen, T.H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science. 75(4): 934-40. https://doi.org/10.2527/1997.754934x.

Meuwissen, T.H.E. 2002. Gencont: an operational tool for controlling inbreeding in selection and conservation schemes. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, August 19-23, Montpellier, France.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4): 1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.M., Selvaraj, S., Dennis, M.Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R.A., Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E. & Phillippy, A.M. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature*. 585: 79–84. https://doi.org/10.1038/s41586-020-2547-7

Nielsen, R. 2004. Population genetic analysis of ascertained SNP data. *Human Genomics*, 1(3): 218-224. https://doi.org/10.1186/1479-7364-1-3-218

Passemard, A.S., Hiemstra, S.J., Tixier-Boichard, M., & Danchin-Burge, C. 2018. Mapping the diversity and characteristics of European farm animal genetic collections: banks or museums? *Proceedings of the World Congress on Genetics Applied to Livestock Production*, 11.

Pook, T., Schlather, M. & Simianer, H. 2020. MoBPS - Modular Breeding Program Simulator. *G3: Genes, Genomes, Genetics*, 10(6): 1915-1918. https://doi.org/10.1534/g3.120.401193

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11): 4414–4423. https://doi.org/10.3168/jds.2007-0980

Sonesson, A.K., Goddard, M.E. & Meuwissen, T.H.E. 2002. The use of frozen semen to minimize inbreeding in small populations. *Genetics Research*, 80(1): 27-30. https://doi.org/10.1017/S0016672302005712

Tixier-Boichard, M. & Crooijmans, R. 2019. A universal tool to share knowledge about animal gene banks. Press release 5 of IMAGE project. [online]. [Cited 22 |September 2020] www.imageh2020.eu/Publications/PR5.pdf

Verrier, E., C. Danchin-Burge, S. Moureaux, L. Ollivier, M. TixierBoichard, M. J. Maignel, J. P. Bidanel, & F. Clement. 2003. What should be preserved: Genetic goals and collection protocols for the French National Cryobank. In D. Planchenaults, ed. *Workshop on cryopreservation of AnGR in Europe*, pp. 79–89. Paris.

Wellmann, R. 2019. Optimum contribution selection for animal breeding and conservation: the R package optiSel. *BMC Bioinformatics*, 20(25). https://doi.org/10.1186/s12859-018-2450-5