Molecular identification of livestock breeds: a tool for modern conservation biology

Mohammed Yaro\textsuperscript{1}, Kylie A. Munyard\textsuperscript{1}, Michael J. Stear\textsuperscript{2} and David M. Groth\textsuperscript{1*}

\textsuperscript{1}School of Biomedical Sciences, CHIRI Biosciences Research Precinct, Faculty of Health Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

\textsuperscript{2}Department of Animal Production and Public Health, Faculty of Veterinary Medicine, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK

\textsuperscript{*}Author for correspondence (Tel.: +618 92667475; fax: +618 92662342; E-mail: D.Groth@curtin.edu.au).

ABSTRACT

Global livestock genetic diversity includes all of the species, breeds and strains of domestic animals, and their variations. Although a recent census indicated that there were 40 species and over 8,000 breeds of domestic animals; for the purpose of conservation biology the diversity between and within breeds rather than species is regarded to be of crucial importance. This domestic animal genetic diversity has developed through three main evolutionary events, from speciation (about 3 million years ago) through domestication (about 12,000 years ago) to specialised breeding (starting about 200 years ago). These events and their impacts on global animal genetic resources have been well documented in the literature. The key importance of global domestic animal resources in terms of economic, scientific and cultural heritage has also been addressed. In spite of their importance, there is a
growing number of reports on the alarming erosion of domestic animal genetic resources. This erosion of is happening in spite of several global conservation initiatives designed to mitigate it. Herein we discuss these conservation interventions and highlight their strengths and weaknesses. However, pivotal to the success of these conservation initiatives is the reliability of the genetic assignment of individual members to a target breed. Finally, we discuss the prospect of using improved breed identification methodologies to develop a reliable breed-specific molecular identification tool that is easily applicable to populations of livestock breeds in various ecosystems. These identification tools, when developed, will not only facilitate the regular monitoring of threatened or endangered breed populations, but also enhance the development of more efficient and sustainable livestock production systems.

Key words: conservation, diversity, genetic resources, global livestock, FAO, molecular techniques, threats, breeds.

CONTENTS
I. Introduction ...........................................................................................................3
II. Key scientific, cultural and economic importance of global livestock genetic resources...7
III. Threats to global livestock genetic resources.....................................................8
IV. Assessment of livestock genetic diversity and conservation status....................12
   (1) Variable number of tandem repeats (VNTRs)...............................................16
   (2) Single nucleotide polymorphism (SNP) markers...........................................18
V. Genomic methods for breed prediction.............................................................19
VI. Conservation methods for domestic animal genetic resources..........................25
VII. Modelling: the way forward...........................................................................29
VIII. Conclusions.....................................................................................................32
I. INTRODUCTION

A large amount of the genetic variation present in wild animal lineages prior to domestication has been conserved during the domestication process, and persisted within the respective domesticates (Dobney & Larson, 2006). Currently, most of these wild lineages are either extinct or critically endangered (Taberlet et al., 2011). Over the 12,000 years since farm animals were first domesticated, their genetic make-up has undergone subtle adaptation due to both natural (speciation) and artificial (domestication/breeding) selection pressures exerted by their specific environments and human activity, respectively (Banik, Pankaj & Naskar, 2015; Hoffmann & Scherf, 2005; Jensen, 2006; Mignon-Grasteau et al., 2005; Morris, 2006; Naskar, Gowane & Chopra, 2015; Price, 1999; Vigne, 2011; Zeder et al., 2006). These selection pressures have culminated in the development of a rich global domestic animal diversity with thousands of breeds (Ajmone-Marsan & The Globaldiv Consortium, 2010; Groeneveld et al., 2010). Each of these breeds is characterised by their unique morphology and productivity related to specific environmental and applied farming conditions (Lopes et al., 2015; Shand, 1997). A livestock breed can be generally defined as either a homogenous group with unique and identifiable phenotypic features that distinguish it from other subgroups within the same species, or a homogenous group for which geographical isolation from other groups of the same species has resulted in their acceptance as unique entities (FAO, 2000; Rege, 2003). Recently, a more refined definition of a breed concept to encompass the history of the livestock was proposed by Felius, Theunissen and Lenstra (2014) and Tixier-Boichard (2014). The scope of this new definition conforms to current practical reality, as not all breeds by definition actually represent unique genetic resources.
Breeds can therefore be regarded as the unit of management for livestock instead of the unit of conservation so as to make it a more useful instrument for conservation purposes (Felius et al., 2014; Groeneveld et al., 2010).

A recent report on livestock breed diversity stated that there were 7,202 local breeds (breeds found in only one country), 509 regional trans-boundary breeds (breeds found in different countries within one region) and 551 international trans-boundary breeds (breeds found in different countries in different continents) (FAO, 2013). These breed classifications cover the seven main mammalian livestock species (sheep, goats, cattle, pigs, buffalo, horses, and asses/donkeys), four main avian livestock species (chicken, turkeys, ducks, and geese) and eight minor livestock species (alpacas, yaks, llamas, camels, elephants, musk oxen, and guinea pigs). However, since the concept of selective breeding only emerged in the last 200 years, and subsequently through more intensive selection in the last few decades, domestic animal diversity has been under sustained threat of significant erosion (Ajmone-Marsan & The Globaldiv Consortium, 2010; Köhler-Rollefson, 1997). In 2012, an analysis of data from 182 countries by the Global Databank for farm animal genetic resources revealed that approximately 8% of all farm animal (local, regional trans-boundary and international trans-boundary) breeds could already be considered extinct, 22% were at varying degrees of extinction risk, and the risk status of 34% was unknown (FAO, 2013). The report, which was an update of the previous 2010 edition, brings the total farm animal extinction to a staggering 12% since 1999. This is quite significant given the fact that a total of only 16% extinctions was recorded in the preceding century (1900–1999) (FAO, 2013). The report also indicated that only approximately 36% of global farm animal genetic resources were not at any risk of immediate extinction.

This growing threat to the world’s animal genetic resources was recognised by the Food and Agriculture Organisation (FAO) of the United Nations (UN) as an emerging global challenge,
and this recognition has led to the ratification by 109 countries, in 2007, of the Interlaken Declaration on world animal genetic resources (Rischkowsky, Pilling & Commission on Genetic Resources for Food Agriculture, 2007). The Interlaken Declaration was the first global action plan specifically aimed at conserving our current animal genetic resources. The declaration called for urgent and prompt measures to be undertaken to mitigate the risk of large-scale loss of defined breeds in the face of challenges such as increasing human population, climate change and emerging diseases. It was also envisaged that such intervention, when successful, would also make a significant contribution to Millennium Development Goals 1 and 7: eradication of extreme poverty and hunger, and ensuring environmental sustainability, respectively. The Millennium Development Goals (or agenda) are a blueprint of eight goals referred to as the UN Millennium Declaration, which was commissioned by the UN general assembly in September, 2000 (United Nations, 2000). The objective of the declaration is to galvanize unprecedented efforts from all member countries to reverse the poverty, hunger and disease affecting billions of people around the world within a 15-year time frame. Despite the historic breakthrough at the Interlaken Summit, little progress has been made so far, especially in developing countries, due to several factors, the most prominent being a general lack of technical capacity and financial resources (FAO, 2007).

The Domestic Animal Diversity Information System (DAD-IS) is an information and communication tool that was set up to coordinate management strategies developed for domestic animal diversity at global, regional and national levels. This system has challenges, especially regarding the quality of entries from developing countries (Tixier-Boichard, 2014). Most of the data submitted, especially from Africa, require regular updating to make them relevant to the current situation. For example 48% and 53% of mammalian and avian breeds recorded in DAD-IS were found to lack sufficient demographic information necessary for the
assessment of their precise risk status (Groeneveld et al., 2010). Furthermore, 87% of entries regarding breed demographics were found to be based on a survey or census, thus presenting a significant limitation, and might be unreliable (Groeneveld et al., 2010).

In recognition of these and other shortcomings in attempts at addressing global animal genetic resource erosion issues, the European Union has recently commissioned a three-year global programme named ‘The GLOBALDIV Project’ (Ajmone-Marsan & The Globaldiv Consortium, 2010). The GLOBALDIV project also known as ‘Global View of Livestock Biodiversity and Conservation’ had representations from the FAO of the UN, the International Livestock Research Institute (ILRI), the International Atomic Energy Agency (IAEA), and 34 individual international researchers from key institutions that are working in areas related to the characterisation of farm animal genetic resources (Globaldiv consortium, 2010). The main aim of this project is to integrate and disseminate the experience of past, large-scale, biodiversity projects and to review the main drivers of biodiversity loss, and then to implement strategies for the conservation of farm animal genetic diversity. Notable among the recommendations of the GLOBADIV project is the need for amalgamation of the disciplines of genetics, socioeconomics and geographic information science for efficient valuation of domestic animal genetic resources.

Currently, improved geo-referencing methods, for example global positioning systems (GPS), are being used as part of a range of measures to provide better production-environment descriptors (Groeneveld et al., 2010). However, because of the dynamic nature of domestic livestock diversity, it is now obvious that more innovative interventions are required to provide precise information on breed structure and status and effectively halt the rapid loss of global livestock genetic diversity. For any livestock breed considered to be at risk, it is recommended that the monitoring of population trends in terms of population size and structure must be carried out at least once per generation (Groeneveld et al., 2010). More
recently, it was indicated that currently available data are inadequate for the ascertainment of the real extent of domestic animal genetic erosion (Bruford et al., 2015). The development of breed-specific identification tools for each characterised livestock breed will not only facilitate the process of regular monitoring of population trends and demographics, but also promote conservation.

This review summarises our knowledge of (i) the key importance of domestic animal genetic resources, (ii) the threats to this resource diversity, (iii) the current status of domestic animal genetic resources, and (iv) conservation methods, with specific emphasis on a molecular genetics approach. We conclude with an assessment of the potential development and use of reliable breed identification tools for livestock breeds for enhancing modern conservation biology studies and preservation of livestock breed diversity.

II. KEY SCIENTIFIC, CULTURAL AND ECONOMIC IMPORTANCE OF GLOBAL LIVESTOCK GENETIC RESOURCES

Domestic livestock are known directly to provide food and livelihoods to more than 90% of the 1.97 billion people who live on less than one US$ a day (Anderson, 2003; International Livestock Research Institute, 2009). With a total global asset value of US$ 1.4 trillion, domestic livestock is reported to contribute 33% and 17% to global protein and kilocalorie consumption, respectively (Herrero et al., 2009). In many developing countries, apart from the provision of food and income, livestock transactions also develop and foster meaningful and emotional social relationships between and among communities (McCorkle & James, 1996). The so-called minor livestock species, although fewer in population number and distribution, are typically of critical importance in terms of cultural heritage and for the livelihood of their owners (McCorkle & James, 1996; York & Mancus, 2013). For instance draught-animal power plays an essential role in the livelihoods of marginal communities in
many developing countries in Asia, sub-Saharan Africa, and Latin America (Barrett, 1992; Lawrence & Pearson, 2002; Teweldmehidin & Conroy, 2010). In addition to these traditional important uses, several species of animals are now used as models in toxicology studies to ascertain the hazard level to humans of prospective drugs (Olson et al., 2000). For example, the miniature pig was identified as an ideal non-primate model for chromosomal abnormalities, skin cell therapy and neural stem cell studies (Vodička et al., 2005). Also a strain of rabbit referred to as ‘Watanabe heritable hyperlipidemic’ was found to be a good model for the study of human myocardial infarction (Shiomi et al., 2003). It has been recommended that comparative medicine, which entails disease studies across animals and human species, holds the key to efficient prevention and control strategies for many zoonotic diseases (Kahn, 2006). Livestock diversity should not only be considered on the basis of global food security, but also as having critical cultural, economic and scientific importance, both currently and into the future.

III. THREATS TO GLOBAL LIVESTOCK GENETIC RESOURCES

The global domestic animal or livestock genetic resources (AnGR) are defined as the sum total of animal species, breeds and strains that currently are, or may be, of future economic, scientific and cultural heritage importance to humans. For the purpose of conservation it is usually breed diversity rather than species diversity that is of greater importance (Philipsson et al., 2011). According to the latest report by the commission on animal genetic resources the percentage of local livestock breeds considered to be at risk of extinction increased by two percentage points between 2010 and 2012 (FAO, 2013). This outlook on the prevailing extinction rate of livestock, although alarming, is likely to be an under-estimation of the actual situation, especially in relation to estimates for developing regions of the world such as sub-Saharan Africa (Rege & Gibson, 2003).
The loss of livestock genetic diversity reduces the range of opportunities available to confront the challenges of unpredictable future events, such as climate change, social change, disease epidemics, selection errors, and many others (Anderson, 2003; Anderson & Centonze, 2007). Some less-common or rare breeds of livestock may also contain rare gene variations that provide greater resistance/resilience to disease and parasites. For many years, the Djallonke sheep and Ndama cattle breeds of West Africa were regarded as less-desirable livestock because of their generally lower productivity, until these breeds were found to possess an inherent ability to resist the debilitating African animal trypanosomiasis disease (Dolan, 1987; Geerts et al., 2009; Goossens et al., 1999; Mwai et al., 2015; Tano et al., 2003). These breeds have since gained popularity in the region, particularly in the trypanosomiasis endemic areas, prompting their inclusion in strategies for mitigating the effects of the disease (Murray et al., 1984; Naessens, Teale & Sileghem, 2002). Several quantitative trait loci studies for trypanotolerance in these breeds have been undertaken to facilitate this process (Dayo et al., 2011; Gautier et al., 2009; Hanotte et al., 2003; Kemp & Teale, 1998). In another example, an approximate 50% increase in weaning rate was attained when the Borroola FecB gene of the lower productivity small Garole sheep breed was introgressed into the highly productive but low fecundity Deccani sheep breed in India (Marshall et al., 2011). Furthermore, behavioural traits such as ability to use a greater range of food sources (which may not normally be efficiently digested in the more common commercial breeds), tolerate heat and/or cold, and even display differences in mothering abilities are all important heritable traits that should be preserved. As many of these uniquely talented breeds are in developing countries and are not currently adequately characterised, it is therefore important to conserve as much as possible of this existing genetic diversity, because we do not know its genetic potential (Mwai et al., 2015). While extinction is a natural process due to the presence of inferior traits (for example, the Djallonke sheep breed not being commercially
desirable), until these breeds are fully genetically characterised, it is not known what genetic potential we are losing for future generations that face different challenges. Many of the presently uncharacterised breeds might serve as important genetic reservoirs for future selection options (Ciani et al., 2013). A more developed strategy of conservation such as has been employed in the preservation of plant germplasm is probably critical for future sustained food security (Sachs, 2009).

There is a wide spectrum of interrelated man-made and natural factors that pose varying levels of threats to global AnGR (Philipsson et al., 2011; Rege & Gibson, 2003). The factors that are responsible for the erosion of genetic diversity are often a function of the size of the population under consideration (Barbato et al., 2015). Generally, the smaller a livestock population, the greater is its vulnerability to extinction (Biscarini et al., 2015; Henson, 1992; Ramstad et al., 2004). Human factors offer the greatest threat to global livestock diversity (Biscarini et al., 2015; Frankham, 1995). The human factors include, but are not limited to; intensive selective breeding, overexploitation, political instability and wars (Goe & Stranzinger, 2002), indiscriminate crossbreeding (Alvarez et al., 2009; Wollny, 2003) and general neglect or lack of breeding programmes (Rege et al., 2011; Wollny, 2003).

Interestingly, these human factors vary across both developed and developing regions of the world. In the developed regions, the threat to livestock diversity is mostly associated with overexploitation such as specialised breeding in response to dynamic socioeconomic pressures (Groeneveld et al., 2010). This trend is also expedited partially by often misguided or inappropriate application of advanced molecular biology technologies (Tisdell, 2003). Conversely in developing countries, the main factors are a general neglect of livestock and or poorly structured breeding programmes driven in part by lack of technical knowledge and financial resources (Alvarez et al., 2009; Biscarini et al., 2015; Philipsson et al., 2011). In the face of this clear dimorphism, it is of utmost importance to take measures necessary to
minimise the ‘Swanson dominance effect’ (Tisdell, 2003). The Swanson dominance effect refers to a phenomenon in which the choices made by the earliest developing societies influence the later pattern of development in later societies. There have been reports of livestock keepers in parts of sub-Saharan Africa abandoning their locally adapted breeds in favour of specialised potentially highly productive, but non-adapted exotic breeds, thereby leading to a decline in diversity (Groeneveld et al., 2010; Wollny, 2003). Nonetheless, regardless of the region of the world, general increases in human population tend to impact negatively on livestock diversity.

Natural events that have commonly been cited as major causes of erosion of livestock genetic resources include tsunamis, earthquakes, hurricanes, droughts, disease epidemics, famine and floods (Prentice & Anzar, 2011). In the past two or more decades, climate change has emerged as a higher-level driving force for reduction in AnGR (Nardone et al., 2010; Thornton et al., 2009). Many reports have described the expected impact of climate change on livestock production systems and diversity (Banik et al., 2015; Herrero et al., 2009; Hoffmann, 2010; Kantanen et al., 2015; McMichael et al., 2007; Naskar et al., 2015). This is mainly because of the direct and indirect implications of climate change on both the frequencies and intensities of most of the causative factors for genetic erosion mentioned previously (Naskar et al., 2015). The irony, however, is that a few livestock species contribute significantly to climate change, as they contribute about a fifth of global greenhouse gas emissions (Garnett, 2009; Gavrilova et al., 2010; McMichael et al., 2007; Shields & Orme-Evans, 2015).

Natural and human-made evolutionary forces either directly or indirectly can cause a reduction in the effective population size ($N_e$) of a livestock breeding population. Therefore, the genetic variability of subsequent populations is drastically reduced because it is derived from the genetic constitution of the few survivors remaining from the original population
In population genetic studies these reductions in population size are referred to as bottlenecks. A population that passes through a bottleneck loses alleles and usually shows reduced average heterozygosity (Allendorf, 1986; Nei, Maruyama & Chakraborty, 1975), but could also lead temporarily, to an increase in heterozygosity if more rare alleles are lost in the process (Hundertmark & Van Daele, 2010; Luikart & Cornuet, 1998). This temporary increase in heterozygosity occurs only if the loss of the rare alleles due to the bottleneck event (mutation-drift equilibrium) has more effect on the expected heterozygosity of a given set of alleles than what is to be expected for a set of alleles under Hardy-Weinberg equilibrium. However, it is the overall decrease in genetic variation of the population post-bottleneck events that is of major relevance. Regardless of the cause of a bottleneck, it may take many generations to restore the original level of heterozygosity through new mutations (Chakraborty & Nei, 1977). Generally, the impact of a bottleneck is logically more profound on small breeding populations because of the larger correlative effect of the resultant diminished genetic variability on population fitness compared to large breeding populations. In population genetic studies, a bottleneck effect is referred to as a founder effect if it is associated with the founding of a new population (Dlugosch & Parker, 2008; Ramstad et al., 2004; Templeton, 1980). Random events such as founder and bottleneck effects that imperfectly eliminate genes and reduce variability within a population are also described as genetic drift (Newman & Pilson, 1997; Ramstad et al., 2004). Reduction in heterozygosity in a livestock population can be associated with decline in fitness of individual members, as is often the case in wild populations (Worley et al., 2010). This is because within small populations, the rate of inbreeding is much higher and consequently there is higher likelihood of the expression of deleterious recessives in a homozygous state. The expression of deleterious alleles has adverse effects on the livestock population, often presenting as reduced production, reproduction and survival (Dlugosch & Parker, 2008; Lacy,
1997). Frankham (1995) and Lacy (1997) have described the positive correlation between inbreeding and risk of extinction. The effective population size model takes into account important population variables such as age and structure, inbreeding rates, genetic drift, genetic diversity and sex ratio. For example, a population of four males and four females constitutes the same effective population size as that of 100 females and only two males (Henson, 1992). Therefore, the effective population size is the preferred indicator of livestock conservation risk status (Dlugosch & Parker, 2008; Nei et al., 1975). In a breed regeneration programme, the effective population size can be enhanced by equalising the male to female ratio, and standardising litter size and longevity within the breeding population, so as to ensure that each animal contributes equally to the next generation. However, it is apparent that the estimation of the effective population size and subsequent determination of its conservation status for a given breed is limited by the lack of availability of a reliable breed identification tool for any specific breed.

IV. ASSESSMENT OF LIVESTOCK GENETIC DIVERSITY AND CONSERVATION STATUS

In order to manage livestock genetic resources sustainably a comprehensive knowledge of diversity within and between breed populations is required (Groeneveld et al., 2010). A major step towards standardising the assessment criteria for livestock breed conservation status was the establishment of a universal classification framework by the FAO for categorising risk status. The current classification of livestock conservation risk status contains seven categories: extinct, critical, critical-maintained, endangered, endangered-maintained, not at risk, and unknown (FAO, 2013). Regular assessment of genetic conservation status of livestock is of fundamental importance to prevent genetic erosion and to preserve diversity.
Key to achieving an effective assessment of livestock conservation status is a reliable mode of identification of members of a target breed. There are two broad methods for identifying individual members of a livestock breed, and their merits and demerits have been discussed thoroughly elsewhere (Agaviezor et al., 2012; Ashley & Dow, 1994; Birteeb et al., 2012). These methods comprise phenotypic and molecular identification techniques. Traditionally phenotypic identification has been used to identify the breed of an individual in livestock genetic diversity studies. The phenotypic variables usually used comprise physical features (e.g. shape of horn, ears, body measurements, colour, etc.), production traits (e.g. growth parameters), reproductive traits (e.g. fecundity) and survival traits (e.g. disease resistance, drought resistance) (Brinks et al., 1964; Gwakisa, Kemp & Teale, 1994; Reverter et al., 2003). These methods are used extensively not only because they are inexpensive and often do not require the use of sophisticated equipment, but also may be useful criteria to some breed societies. However, the major disadvantage is that the genetic diversity is observed only at the phenotypic level and this does not always correspond to actual diversity at the DNA level (Felius et al., 2014).

It is possible to find different phenotypes with similar genotypes, typically due to genotype–environment interactions, for example as observed in Brazilian sheep breeds (Paiva et al., 2005) and Egyptian sheep breeds (Ali, 2003). Similar phenotypes with different genotypes also occur, as observed between the West African Djallonke sheep and F1 Djallonke–Sahelian crossbreeds (Alvarez et al., 2012; 2009; Wafula et al., 2005). As a result, the use of molecular tools in many assessment studies of genetic diversity in different regions of the world revealed varying degrees of unexpected introgression and admixture in livestock populations. These studies include the Djallonke sheep breed of sub-Saharan Africa (Alvarez et al., 2009; Wafula et al., 2005), Herdwick sheep of the United Kingdom (Bowles, Carson & Isaac, 2014) and alpaca and llama of Latin America (Kadwell et al., 2001). This obvious
shortcoming has rendered the use of phenotypic methods in isolation as unreliable for
determination of livestock breeds for the purpose of genetic diversity studies.

In livestock genetic diversity studies, the molecular method for determining breed identity
entails two main approaches based upon either protein markers or DNA markers (Ferguson et al., 1995; McMahon, Teeling & Höglund, 2014). Protein markers, also referred to as
allozymes, are based on the characteristic polymorphism of the blood group systems,
leucocyte antigens and enzymes (Dodgson, Cheng & Okimoto, 1997). This molecular method
employs these protein markers to estimate genetic variability in livestock populations as well
as phylogenetic relationships between breeds (Pepin & Nguyen, 1994; Witko-Sarsat et al.,
1996). Although better than the phenotypic method, the use of protein markers is too
expensive for a large number of loci, and lacks the power to resolve differences between
closely related breeds, because of limits of detection of genetic variation (Engel et al., 1996;
Ferguson et al., 1995; Toro, Fernández & Caballero, 2009). The use of DNA markers is the

Nuclear and mitochondrial DNA marker analyses have revealed detailed information on
many domestication events, such as their timing and location (Bruford, Bradley & Luikart,
2003; Zhao et al., 2013). DNA marker analyses provide an added opportunity for
investigating the genetic composition of both extinct and endangered breeds without
destructive sampling.

There are seven principal DNA marker techniques commonly used for livestock diversity
studies (Sunnucks, 2000). These seven DNA marker techniques have been discussed
thoroughly and their advantages and disadvantages are well documented. These techniques
are: restriction fragment length polymorphism (RFLP) (Beckmann & Soller, 1983; 1986;
Thurston et al., 2002), mitochondrial DNA barcoding (mtDNA) (Avise et al., 1987; Avise &
Ellis, 1986; Harrison, 1989; Kocher et al., 1989; Zhang & Hewitt, 1996), random amplified
polymorphic DNA (RAPD) (Ali et al., 2004; Dodgson et al., 1997; Koh et al., 1998; Levin, Crittenden & Dodgson, 1993), amplified fragment length polymorphism technique (AFLP) (Blears et al., 1998; Parsons & Shaw, 2001; Savelkoul et al., 1999), Y-chromosome technique (Bruford et al., 2003; Zeder et al., 2006), variable number of tandem repeats (VNTR) (minisatellite and microsatellite markers) (Chistiakov, Hellemans & Volckaert, 2006; Fan & Chu, 2007; Lopes et al., 2015; Zane, Bargelloni & Patarnello, 2002) and single nucleotide polymorphism (SNP) based techniques (Andersson & Georges, 2004; Liu & Cordes, 2004; McMahon et al., 2014; Morin, Luikart & Wayne, 2004; Tixier-Boichard, 2014; Vignal et al., 2002). The latter two DNA techniques are the most popular.

The advancement of DNA technologies during the past three decades, and particularly since 2007 when high-throughput next-generation sequencing became readily available, is revolutionising livestock population genetics studies (Helyar et al., 2011; Schlötterer et al., 2014a). This revolution is expedited by the concomitant advancement in bioinformatics tools and pipelines (Kofler, Nolte & Schlötterer, 2015). DNA markers have been used not only for diversity studies but also for molecular characterisation of numerous livestock breeds worldwide (Agaviezor et al., 2012; Al-Atiyat, Salameh & Tabbaa, 2014; Alvarez et al., 2012; Bowles et al., 2014; Chenyambuga et al., 2004; Mukesh et al., 2004). The dramatic reduction in the cost of use of DNA markers has facilitated their greater use by researchers. AFLP and RAPD markers are both bi-allelic and dominant in nature, and hence are less informative and also have low reproducibility compared to the other markers (Vignal et al., 2002). These characteristics have rendered them less popular for most animal-based molecular genetic studies. RFLP markers are bi-allelic and co-dominant, and were famously used in the first large-scale mapping of the human genome. However, RFLPs have now been superseded by the more informative microsatellite markers, a type of VNTR for both animal and human genome studies. In turn, microsatellite markers have been largely supplanted by single
nucleotide polymorphism (SNP) arrays. mtDNA along with microsatellite markers were once popular molecular genetic techniques of choice for evolutionary and ecological studies, however the molecular information provided by mtDNA markers is limited to only maternally inherited loci (Morin et al., 2004). The use of mtDNA techniques, in combination with archaeological data, has provided precise information on most of the important centres of domestication for the main livestock species around the world (Bruford et al., 2003; Globaldiv consortium, 2010; Guo et al., 2006; Zeder et al., 2006). Similarly limited, the use of Y-chromosomal haplotype markers elucidates specific molecular information only on paternally inherited traits (Luikart et al., 2006). The VNTR and the SNP techniques will be discussed in greater detail below because of their current wider application compared with the other molecular markers.

(1) Variable number of tandem repeats (VNTRs)

The application of VNTRs for assessment of genetic variation, sub-structuring and hybridisation in natural populations has been reviewed in great detail previously (Bruford & Wayne, 1993; Chistiakov et al., 2006; Fan & Chu, 2007; Sunnucks, 2000). The VNTR technique is based on the abundance of tandem repeats of simple sequences of nucleotides throughout the eukaryotic genome (Takezaki & Nei, 2008). These VNTRs have been categorised into minisatellites and microsatellites according to the number of nucleotides per motif of repeats. VNTRs of between 1 and 6 nucleotide base pair units are referred to as microsatellites (Ashley & Dow, 1994; Chistiakov et al., 2006; Fan & Chu, 2007), whereas a range of between 10 and 60 nucleotide base pair units is regarded as a minisatellite (Ashley & Dow, 1994; Wasko & Galetti, 2003). Whereas minisatellites are concentrated towards the telomere of chromosomes, microsatellites are randomly distributed in chromosomes. Microsatellite markers are highly polymorphic, co-dominant markers of relatively small size,
and hence are more amenable to polymerase chain reaction (PCR) typing than are minisatellites (Zane et al., 2002). Also, in comparison to the RFLP and RAPD techniques, the genetic basis of variability is readily apparent for microsatellites. Most microsatellites are located in non-coding regions of the genome (Chistiakov et al., 2006). Generally, microsatellite primers developed for one species of livestock are broadly applicable to other closely related species. For example, microsatellite markers developed for studies in bovine species are applicable to caprine and ovine species (Engel et al., 1996). This versatility has led to the popularity of microsatellite maps for economically important livestock species (Sunnucks, 2000).

Microsatellites have been used in linkage mapping in diverse organisms, for example in the bovine genome (Barendse et al., 1997), porcine genome (Rohrer et al., 1994), human genome (Dib et al., 1996), and ovine genome (Maddox et al., 2001). Microsatellites have also been employed for the identification of quantitative trait loci (QTL) in major livestock species, for example, carcass composition and growth rate in cattle (Casas et al., 2000), back fat thickness and intramuscular fat in pigs (Rohrer & Keele, 1998) and intestinal parasitic infection in sheep (Davies et al., 2006). Other population genetics studies accomplished with microsatellite markers include the determination of evolutionary relationships (Alvarez et al., 2012; Buchanan et al., 1994; Vanhala et al., 1998), estimation of pedigree errors (Visscher et al., 2002) and determination of genetic diversity among livestock populations (Alvarez et al., 2012; Alvarez et al., 2009; Curković et al., 2015; Marletta et al., 2006; Medugorac et al., 2011; Wafula et al., 2005). The genetic distance between individuals within a livestock population indicates the suitability of an individual for conservation purposes. Individuals within the same breed with the widest differences in genetic distances are deemed most suitable candidates for conservation programmes. The estimates of genetic distances are also relevant for the determination of divergence time and construction of phylogenies (Takezaki
Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology

& Nei, 1996). Prior to the use of SNP markers, microsatellites were the most popular and efficient technique for genetic-diversity investigation, not only in livestock but also in humans. Microsatellites continue to be seen as a method of choice for many researchers in breeding programs, particularly in third-world and developing countries, due to their low cost, relative ease of analysis and requirement for relatively unsophisticated scientific equipment (Rege et al., 2011). Whereas newer technologies offer better prospects, the enabling supporting infrastructure is often not available in developing regions of world. For example, the analyses of large-scale genomic data require ready internet access for web-based reference sequence information, which currently cannot be guaranteed in many sub-Saharan African countries (Gulati, 2008; Oyelaran-Oyeyinka & Lal, 2005). The same can also be said of the availability and reliability of electric power supplies necessary to support cryobanking of important genetic materials (Deichmann et al., 2011; Wolde-Rufael, 2006). Given the levels of existing infrastructure and human technical capacity in many developing countries, significant investment is required to implement some of the recent genomic technologies for sustainable livestock production and conservation (Rege et al., 2011).

(2) Single nucleotide polymorphism (SNP) markers

The growing importance of SNP marker applications in molecular genetics has been reviewed in detail by Barbato et al. (2015), Broxham (2015), Goddard and Hayes (2009), Vignal et al. (2002), Hamblin, Warburton and Buckler (2007) and Morin et al. (2004). SNPs represent a location within a DNA sequence for which more than one nucleotide type is present within a given population (Morin et al., 2004). In a strict molecular sense, SNPs are base substitutions within nucleotide sequences, and the very high density of their occurrence in the genomes of eukaryotes, including livestock, has been of great significance in population genomics studies (Goddard & Hayes, 2009; Vignal et al., 2002). Although SNPs
Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology

are bi-allelic (sometimes tri-allelic or quadri-allelic) co-dominant molecular markers, their high density permits, more than any other technique, very detailed information to be elucidated on genome dynamics within a study population (Hamblin et al., 2007; Morin et al., 2004). Furthermore, they provide deeper insight than microsatellites with respect to linkage disequilibrium and haplotype diversity, pedigree information and past demographic events, such as bottlenecks within a target population (Clarke et al., 2014; Gautier et al., 2007; Helyar et al., 2011; Morin et al., 2004; The Bovine HapMap Consortium, 2009; Vignal et al., 2002). SNP markers also allow for standardised data recording, and are stable over generations if selected from neutral genomic loci (Tixier-Boichard, 2014). These features of SNP markers are opening opportunities for wider applications of SNP markers in understanding of livestock genetic architecture, such as precise identification of genomic regions that control traits of economic and survival importance (Kohn et al., 2006; Matukumalli et al., 2009) and ultimately genomic selection (Choi et al., 2015; Clarke et al., 2014; Goddard & Hayes, 2009). These advances in genetic marker application for use in population genetic studies will not only enhance the development of improved livestock production systems, but most importantly will facilitate the development of efficient conservation strategies.

V. GENOMIC METHODS FOR BREED PREDICTION

The unique genetic structure of livestock breeds, shaped by their demographic history of natural and artificial selection, provides a basis for the assignment of an individual to a particular breed (Bertolini et al., 2015). The large numbers of SNPs identified in various domestic animal species have been used to develop species-specific standard technology products referred to as BeadChips or SNP chips (Wilkinson et al., 2011). These SNP chips are commercially available, and have been designed to amplify genome-wide SNP loci rapidly in an automated platform to generate large-scale genomic SNP data for analysis.
Examples are the Illumina ovine SNP 50 BeadChip and the Illumina bovine SNP50 BeadChip developed for sheep and cattle, respectively (Bertolini et al., 2015; Dodds et al., 2014). An analysis of the data generated using SNP chip technologies has shown that it is possible to assign an individual animal correctly to a specific breed (Table 1). Moderate- to high-density SNP genotyping assays are frequently used to capture common genomic variations within breed populations (Bertolini et al., 2015; Broxham, 2015; Dodds et al., 2014; Frkonja et al., 2012; Kijas et al., 2009; Lewis et al., 2011; Rolf et al., 2014; Sasazaki et al., 2007; Suekawa et al., 2010; Wilkinson et al., 2011). Bioinformatics and statistical tools such as STRUCTURE, principal component analysis (PCA) and discriminant analysis have been widely applied to these SNP data sets with varying levels of success (Gilbert et al., 2012; Herrero-Medrano et al., 2013; Hubisz et al., 2009; Schwartz & McKelvey, 2009).

Unlike STRUCTURE and PCA, discriminant analysis is not considered a multivariate statistical method for assignment of individuals to a population. Furthermore, the discriminant analysis does not permit the fractional or mixed prediction of individuals in a subject population (Dodds et al., 2014). This is a major limitation for studies that require multiple predictors, making it less suitable for breed predictions (Dodds et al., 2014).

However, with continuous advances in bioinformatics tools, many more tools are becoming available for this kind of analysis. Recently, two different analyses of Illumina OvineSNP50 genotyped data were used to assign four New Zealand sheep breeds correctly with high prediction accuracy (0.85–0.97) (Dodds et al., 2014). The two methods used were a regression analysis with a genomic selection algorithm that employed allele frequencies, and genomic Best Linear Unbiased Prediction (gBLUP) estimates respectively, derived from a pure-bred subset of each sheep breed population. These estimates were then used as the training data set for respective breed predictions in each of the four populations. The two methods produced different prediction accuracies that depended upon the breed structure of
the subject populations. It was concluded that the accuracy of breed prediction was enhanced if the composition of the training set is representative of the breed diversity within a subject population (Dodds et al., 2014). A recent study of the genomics of cattle in the USA beef industry has supported this conclusion, particularly for predictions in multi-breed beef cattle populations (Rolf et al., 2014). The accuracy of predictions obtained from both methods was similar to that recorded with STRUCTURE (Dodds et al., 2014). However, unlike these two methods, STRUCTURE does not provide a prediction equation for subsequent breed prediction in a subject population (Dodds et al., 2014). STRUCTURE analysis of data also has low reproducibility (Gilbert et al., 2012). In spite of these drawbacks, the STRUCTURE algorithm has been used extensively in clustering of genetic data (Falush, Stephens & Pritchard, 2007; Hubisz et al., 2009; Schwartz & McKelvey, 2009).

Principal component analysis (PCA) is also a powerful multivariate tool that facilitates the elicitation of unknown population clusters (Lewis et al., 2011). When applied to genomic data, it has been found to group individuals of the same breed together (Dodds et al., 2014). The use of a combination of ancestry-informative marker metrics and PCA using 30,501 SNPs on the Bovine HapMap accurately predicted 19 cattle breeds (Lewis et al., 2011). This result led to the conclusion that a carefully selected panel of 250–500 SNPs from the Bovine HapMap data set was sufficient for correct breed assignment. Justifiably, Lewis et al. (2011) also conceded that the sensitivity and the resolving power of their approach would be higher if applied to denser genomic data than the Bovine HapMap data set used. This view is also supported by Wilkinson et al. (2011). However, the PCA result does not readily translate to the actual breed proportion estimates in mixed breeds. Hence, it is more suitable for the verification of a pure-breed member (Dodds et al., 2014). Nonetheless, this combined tool approach is said to be suitable for the reliable tracing of breed-specific branded products in the meat industry (Lewis et al., 2011). Prior to this study, SNP-based markers derived from
the AFLP technique were used for distinguishing between Australian and Japanese beef (Sasazaki et al., 2007). Although the AFLP-derived markers were of low resolution, the power was sufficient to discriminate cattle breeds from the two countries. A more stringent breed-specific SNP marker panel was later developed from Bovine 50K SNP BeadChip data and this was able to discriminate between Japanese and America cattle products (Suekawa et al., 2010). The higher efficiency marker panel was developed in response to an outbreak of bovine spongiform encephalopathy (BSE) in the USA, and comprised only six highly informative breed-specific SNPs (Suekawa et al., 2010). More recently, the use of 48 and 96 highly informative SNP markers derived from a combined PCA in combination with a ranking algorithm (random forests) of Illumina Bovine SNP50 BeadChip genotyped data, correctly assigned four Italian cattle breeds (Bertolini et al., 2015). A few of the highly informative SNPs used in that study were found to be located in loci associated within important quantitative traits for some cattle breeds. A systematic assessment of the efficiency of four different methods for identifying population-informative SNPs from the SNP50 BeadChip data set showed no gain of further assignment power beyond the use of more than 200 SNPs in a panel, for all the approaches (Wilkinson et al., 2011). Wilkinson et al. (2011) also showed that a panel of 60 SNP markers was the minimum required for successful prediction of the cattle breeds investigated. However, more genetic markers (in excess of 200) will be required successfully to assign closely related breeds and far fewer for distantly related breeds. Hence, Wilkinson et al. (2011) provided evidence that the number of SNPs required for correct assignment of an individual to a breed is directly proportional to the genetic heterogeneity or homogeneity of the sampled population. In a more recent study, two separate panels of SNPs derived from 21 different sheep breeds from Italy and Slovenia were used to assign all the sheep correctly to the breeds (Dimauro et al., 2015). This study combined three different types of discriminant analyses on an Illumina Ovine SNP50
Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology

genotyped data set from all 21 breeds to produce a reduced panel of 108 and 110 SNP markers. Further advances in genomics have also shown that the use of SNP chip technology is prone to ascertainment biases because the discovery SNP panels are derived from small numbers of individuals from selected populations that are not representative of all the species populations (Albrechtsen, Nielsen & Nielsen, 2010; Foll, Beaumont & Gaggiotti, 2008). The bias in sheep was particularly evident for African and Asian breeds as the SNPs were originally identified from European breeds (Kijas et al., 2012). Such ascertainment biases are likely to skew inferences determined from genotyped data such as allele frequency spectra and genetic differentiation of subject populations.

Next-generation sequencing (NGS) of genomic DNA will provide denser SNP data than the Illumina OvineSNP50 BeadChip or any medium-density marker set that has previously been used in this type of study, and does not suffer from ascertainment bias. It is important to note that NGS data have inherent challenges arising from alignment and sequencing errors, but these are smaller in comparison to the biases of the SNP-chip genotyped data (Albrechtsen et al., 2010). For example, a whole genome of an animal from a Korean cattle breed that was sequenced on the Illumina HiSeq 2000 platform resulted in more than 10 million SNPs being identified, 54% of which were novel (Choi et al., 2014). Furthermore, another study showed a reduction in prediction accuracy when a SNP data set derived from Illumina Bovine SNP50 was replaced with that derived from an Illumina SNP3K genotyped data set (Kuehn et al., 2011). The higher resolving power of NGS has been shown to capture more rare breed-specific polymorphisms or more informative polymorphisms (with higher confidence) than bovine SNP50 BeadChip genotyping (Choi et al., 2015; 2014; Lee et al., 2013).

Collectively, these studies suggest that the use of high-density data will enable the real possibility of developing a smaller panel containing the most informative breed-specific SNPs having the highest sensitivity for resolving breed differences. Therefore, analyses and
use of such NGS data will lead to more accurate breed predictions, and the NGS of individual genomes at high coverage has been referred to as the ‘gold standard’ for generating quality data (Schlötterer et al., 2014b). In spite of the dramatic reduction in the cost of NGS, the cost of sequencing the large number of individuals required for population studies of this nature, at high coverage, is still economically prohibitive. However, it has also been shown that NGS of pools of individuals at a moderate coverage could provide a cost-effective and efficient alternative technique for generating very high density SNP data sets across a genome, compared with NGS of non-pooled individuals (Gautier et al., 2013; Kofler et al., 2015; Schlötterer et al., 2014b). Another promising cost-effective approach that was applied successfully to some plant species is referred to as genotyping by sequencing (GBS), and is based on the sequencing and analysing of more informative regions of the genome rather than the whole genome (Elshire et al., 2011). GBS is fast, highly specific and exceedingly reproducible, and could be used to complement the pool-sequencing approach through confirmatory testing where the need arises. For association-mapping studies, the analysis of pooled-sequenced data has more statistical power than SNP arrays (Futschik & Schlötterer, 2010; Gautier et al., 2013; Kofler, Pandey & Schlötterer, 2011). The high-density SNP data generated by this pool strategy was shown to facilitate the discovery of more accurate allelic frequency estimates across a genome (Futschik & Schlötterer, 2010). The advantages of a pooled-sequencing technique over individual sequencing have been reviewed previously (Schlötterer et al., 2014b). NGS sequencing of pools of unrelated individual purebreds from a subject population therefore will enable the identification the most informative breed-specific SNPs. The rationale for sampling unrelated animals instead of related ones is to enable the capturing of a wide spectrum of within-breed genetic diversity of the subject population. This strategy is necessary to minimise the introduction of ascertainment bias into the subsequent breed identification tool (McTavish & Hillis, 2015). A carefully selected panel of SNPs
Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology

derived from the identified breed-specific SNPs can serve as an efficient breed identification tool.

VI. CONSERVATION METHODS FOR DOMESTIC ANIMAL GENETIC RESOURCES

Conservation of AnGR comprises all the management practices carried out to preserve the pool of genetic diversity of livestock for the purposes of meeting current and future needs of humans (Rege & Gibson, 2003). The relevance of conservation of AnGR has been discussed from several different perspectives, including economic evaluation as a basis for AnGR conservation decisions (Drucker, Gomez & Anderson, 2001), the role of cryopreservation, reproductive technologies and genetic resource banks for AnGR conservation strategies (Hiemstra, van der Lende & Woelders, 2006b; Holt & Pickard, 1999; Mara et al., 2013), information on population kinships as a basis for AnGR conservation decisions (Eding & Meuwissen, 2001) and the challenge of conserving indigenous AnGR diversity (Mendelsohn, 2003). Each breed of livestock consists of unique sets of genes resulting from evolutionary events and diverse selection pressures imposed by the environment combined with the activities of humans over time. It is therefore difficult, if not impossible, to replace lost breeds of livestock, because those unique evolutionary processes cannot be re-created. There has been a general consensus on three critical approaches regarding the conservation of domestic livestock breeds: sustainable utilisation of available livestock breeds; appropriate diversity-based improvement strategies for livestock breeds; and development of appropriate assessment and preservation strategies (FAO, 2000; Hammond, 1999; Koehler-Rollefson & Meyer, 2014; Notter, Mariante & Sheng, 1994; Thornton et al., 2007). In addition to these approaches, the FAO has also recommended the regular monitoring of livestock breed conservation status (FAO, 2013). Currently the two main methods of AnGR conservation
applied are the *in situ* and the *ex situ* methods. The applicability of both conservation methods, and their respective merits and demerits has been reported extensively (Boettcher *et al.*, 2010; Hammond, 1994; Henson, 1992; Mara *et al.*, 2013; Rege & Gibson, 2003). *In situ* conservation can best be described as the sustainable breeding of an endangered livestock breed in their normal adaptive production environment, or as close to it as practically possible, to conserve genetic diversity over a long period (Andrabi & Maxwell, 2007; Henson, 1992). Notable features of *in situ* conservation therefore include selection and mating programmes that retain genetic variation within the target group, as well as management of the ecosystem to sustain their production. The basic requirements for *in situ* conservation programmes are generally readily available and affordable globally. There is a distinct difference between developed and developing countries regarding the minimum number of individuals required to commence an *in situ* programme. This is typically due to general differences in the efficiency of management of their respective livestock production systems. For example, whereas the minimum number for major livestock breeds (i.e. cattle, sheep, goats, pigs) required for *in situ* conservation is 100–1,000 breeding females in developed countries, no fewer than 5,000 breeding females is recommended for developing countries (Signorello & Pappalardo, 2003). Simon (1999) reported 500 breeding females for pigs and goats, 750 for cattle, and 1,500 for sheep for European breeds. It has been recommended that, ideally, for unrelated animals a minimum of 25 males and 50 females is sufficient to commence an *in situ* conservation programme, because the possible loss of genetic variability is estimated to be less than 1% per generation (Henson, 1992; Mara *et al.*, 2013). However, recent advances in the field of genomics have enabled the elucidation of abundant genomic information *via* high-throughput sequencing technologies and analysis, so a re-evaluation of these recommended numbers required for conservation programmes is overdue. This is because more accurate population genetic parameters such as allele
frequencies can be computed for target populations to allow for more precise determination of these numbers.

There are a number of flagship in situ conservation programmes in place to conserve and improve some disease-resistant breeds of livestock in some African countries, for example, Ndama cattle in the republic of Guinea (Yapi-Gnaoré, Dagnogo & Oya, 2003), Djallonke sheep in Ghana and Cote D’Ivoire (Kosgey & Okeyo, 2007), and Tswana sheep in Botswana (Henson, 1992). The unique advantage of the in situ conservation method is that the target livestock breed continues to be utilised in the process. However, the danger is that the target livestock breed remains susceptible to uncertain demographic threats such as natural disasters and disease epidemics.

The ex situ livestock conservation method is the preservation of endangered livestock outside their normal production systems (Henson, 1992; Hiemstra et al., 2006a). This method is normally applied to target groups that are faced with imminent extinction, and hence requires the use of high-level expertise and technology. The three main ex situ methods are cryopreservation, farm park conservation, and breed pools or composite preservation.

Cryopreservation, also referred to as in vitro ex situ is undoubtedly the most popular of the ex situ approaches to conservation of AnGR (Hiemstra et al., 2006a). This approach involves the cryopreservation of eggs, semen and or embryos of endangered or threatened animals in genome banks for use in managing diversity or regenerating the population decades, or even centuries, later (Chen, Zhang & Yu, 2008; Hanks, 2001; Russo et al., 2007; Xiao-Yong et al., 2008). Cryogenic storage of carefully evaluated genetic material from a target breed population is also seen as an insurance policy against future loss. The merits and demerits of using these approaches have been discussed previously (Boettcher et al., 2010; Munro & Adams, 1991; Philipsson, Rege & Okeyo, 2006; Pintado & Hourcade, 2011; Ruane & Sonnino, 2011). The application of cryopreservation formerly depended only on assisted
reproductive techniques such as artificial insemination and embryo transfer technologies. However, recent advances in reproductive biotechnologies including semen sexing, embryo micromanipulation and in vitro fertilisation have the potential to revolutionise the livestock cryopreservation approach (O’Brien, Steinman & Robeck, 2009; Prentice & Anzar, 2011). Cryopreserved genetic materials are shielded from the influence of unfavourable environmental conditions in existence in the normal production ecosystems. Regeneration of a breed through only preserved semen requires a number of back crosses (Andrabi & Maxwell, 2007). However, the exact genetic composition of an original breed after going through adaptive selection is not recoverable with only cryopreserved semen that was collected before the adaptation process. In practice, the in situ and ex situ conservation methods are not mutually exclusive because the cryopreservation approach can be used to complement the in situ method to achieve better regeneration of endangered populations. A range of combinations of in situ and ex situ conservation methods are being applied in a now-popular integrated conservation approach (de Souza et al., 2011; Hiemstra et al., 2006a). It has been recommended that a stock of cryopreserved semen from 25 unrelated sires is sufficient to provide a reasonable diversity for an endangered population (Bruns & Glodek, 1999; Mara et al., 2013).

The farm-park ex situ conservation approach is similar to the in situ conservation approach, except that the targeted breeds are preserved outside their normal production environment in a specialised institutional setting, also referred to as an Ark-farm (Simon, 1999). Farm-park animals are usually also subjected to more stringent management regimes to conserve natural levels of genetic variability within each species (Chesser, Smith & Brisbin, 1980). A notable feature of the farm-park approach is its popularity in attracting tourists, and hence creating awareness of the need to conserve endangered animals. The Cotswold farm park in the UK is an example where rare breeds of sheep, goats, cattle, pigs and horses are being conserved,
and it attracts more than 100,000 visitors yearly (Henson, 1992). The breed-pool preservation programme is unique in the sense that it involves the breeding together of a pool of two to four rare breeds with similar characteristics, and subsequently managing their offspring to conserve genetic variation (Henson, 1992). It is, however, recommended that the breed characteristics of each of the rare breeds is well ascertained prior to commencing a breed-pool programme (Santos et al., 2013). This method is particularly suitable for genes that control obvious morphological traits and extreme quantitative traits such as coat colour and prolificacy, respectively. Although this approach conserves useful genes from the pool, individual breeds are lost in the process. An example of the breed-pool approach is the conservation programme of four rare desert goat breeds in the north eastern part of Brazil (Henson, 1992).

Given that no single conservation method is capable of solving the myriad of challenges of domestic animal genetic resource erosion, an integrated conservation approach has been advocated to provide greater efficiency (Rege & Gibson, 2003).

VII. MODELLING: THE WAY FORWARD

The successful domestication of animals represents a pivotal historic event in the cultural and demographic development of humans. The importance of global domestic livestock diversity to human wellbeing is now well appreciated. This is evident from the globally coordinated efforts directed at halting the decline in AnGR as well as the sustainable utilisation of available livestock resources as discussed herein. These global initiatives have yielded several interventions which are being implemented at the international, regional and local level (Ajmone-Marsan & The Globaldiv Consortium, 2010; FAO, 2013). The main global body is the DAD-IS that coordinates regional bodies, for example the European Farm Animal Biodiversity Information System (EFABIS) and the Domestic Animal Genetic Resource
In Information System (DAGRIS) for European and African regions, respectively. The regional bodies in turn coordinate the local or national bodies, which essentially are the individual member states of the FAO of the UN. These efforts are being supplemented by the activities of other important organisations, prominent members including the GLOBALDIV, the International Society for Animal Genetics (ISAG), the SAVE foundation and several livestock breed societies worldwide (Broxham, 2015). Some notable progress has been made towards reducing the rate of erosion of global AnGR. For example, the status information for global mammalian and avian livestock breeds in the DAD-IS has increased from 43% and 39%, respectively in 2009, to 57% and 48%, respectively, in 2012 (FAO, 2013). The effective monitoring of breed conservation status of livestock requires at least one census per generation of that target breed (FAO, 2007; Groeneveld et al., 2010). A specific breed identification tool for each livestock breed will expedite this exercise. However, pivotal to the success of these conservation efforts is the reliability of genetic identification of individual members within a target breed. The advancement in molecular technology in the last two decades has significantly increased our understanding of the population genetics of domestic animals. It is apparent that the molecular characterisation of all domestic livestock breeds, particularly in developing countries, is a pre-requisite for their sustainable utilisation and conservation. This is because characterisation at the molecular level provides precise information for determination of the actual population characteristics such as genetic variation and effective population size (Luikart et al., 2003). Currently, many domestic livestock breeds, particularly those in the developing countries, have yet to be characterised due to myriad issues including the lack of financial and technological capacity. A recent report indicated that the risk status of 36% of all populations of the local livestock breeds still remains unknown (FAO, 2013).
The main molecular technique used for most livestock genetic characterisation was microsatellite markers [for example, in Spanish native cattle breeds (Martín-Burriel, García-Muro & Zaragoza, 1999), Aberdeen Angus cattle breeds (Vasconcellos et al., 2003), Austrian sheep breeds (Baumung et al., 2006) and indigenous goats in sub-Saharan Africa (Chenyambuga et al., 2004)]. Although highly informative, the current panels of microsatellites used for analyses are not capable of elucidating all the information required regarding breed variation in livestock (Toro et al., 2009). Recently, it is becoming more evident that SNP analysis is more suited for the high-throughput genotyping that is required to elucidate greater molecular insights such as historic signatures of selection (Qanbari et al., 2014), phenotypic variations within livestock breeds (Groenen et al., 2011) as well as linkage disequilibrium over short physical distances (Kijas et al., 2014). The availability and accessibility of comprehensive databases of genomic data for various uses has also facilitated population genetic studies globally, for example the National Centre for Biotechnology Information (NCBI) (Sayers et al., 2011), the Livestock Animal Quantitative Trait Loci database (Hu et al., 2013), and the University of California Santa Cruz (UCSC) genome browser (Dreszer et al., 2012).

The challenge now is to use these enhanced insights and understanding of molecular methods to develop breed-specific identification tools that are easily applicable to populations of livestock in different ecosystems. Such a breed-specific tool can be developed through identification and characterisation of unique phylogenomic SNPs in next-generation sequenced pooled genomic DNA from a selected representative small group of pure-bred individuals. The lessons derived from the ascertainment bias of genetic markers indicate that, for the purpose of conservation, it will be more suitable to develop a robust specific identification assay for each target breed rather than one assay for the identification of different breeds. This assertion does not discount the continued relevance of the use of
common marker sets across multiple breeds or even species in other molecular studies such as investigation of QTL for economically important production and disease-resistant traits. However, for a target breed population the selection of breed-specific SNPs from neutral regions of the genome would guard against loss of efficiency of the SNP assay developed over time through direct selection or hitchhiking effects. It is important to add that it is not always obvious which regions of the genome are under the influence of selection. In a recent study, an annotation of SNPs derived from the WGS Korean cattle breed using the bovine reference genome led to the suggestion that fixed, breed-specific SNPs might be useful for breed identification (Choi et al., 2015). That study described breed-specific fixation of many SNPs.

A growing number of software tools are being developed for the analysis of pooled-sequenced data (Kofler et al., 2015; Li & Durbin, 2010; Li et al., 2009). Read alignment of next-generation sequenced pooled genomic data to reference genomes has been achieved using the Burrows-Wheeler Tool (Kofler et al., 2011; Li & Durbin, 2010). The aligned reads are converted to a compatible pileup file format with SAMtools for subsequent analyses with the PoPoolation algorithm (Kofler et al., 2015; Li et al., 2009). Pooled-genomic sequenced data have been successfully analysed with PoPoolation accurately and efficiently to identify allele frequencies and population differentiation parameters of subject populations (Kofler et al., 2015; Kofler et al., 2011). Other analytical methods successfully applied include modified versions of popular genetic estimators such as the Watterson’s θ and Tajima’s π analyses (Futschik & Schlötterer, 2010; Gautier & Naves, 2011).

These tools enable the efficient analyses of whole-genome sequenced SNP data sets such as mapping to appropriate reference genomes, SNP calling and annotation. Additional benefits of such data sets include use for the investigation of traits of economic and adaptive importance for the breed (Choi et al., 2015; Qanbari et al., 2014). Information from
discovered phylogenomic SNPs can then be used to develop breed-specific SNP assays for the easy and precise identification of pure-bred members from mixed populations of breeds in various ecosystems. These tools will not only facilitate the timely diagnosis of the conservation status of livestock breeds, but will also permit the regular monitoring of endangered breed populations, particularly in developing countries where the lack of technical and financial capacity is reported to be a major impediment.

VIII. CONCLUSIONS

(1) Maintaining global domestic animal genetic diversity is important to human wellbeing.

(2) Breed-specific molecular identification tools are urgently needed to allow the reliable and expeditious identification of individual members of any given breed; this is a pre-requisite for sustainable utilisation and conservation of any breed.

(3) A growing number of studies have established that whole-genome sequencing of pools of individuals within a group or breed provides a great deal of information on genetic variation across the whole genome even when performed at relatively low coverage, but also at considerably lower cost (Clarke et al., 2014; Gautier et al., 2013; Kim et al., 2010; Kofler et al., 2015). This will be a cost-effective technique for the identification of breed-specific phylogenomic SNPs within a target breed for the purposes of developing breed-specific molecular identification tools.

(4) The knowledge and technological gap between the developed and developing worlds need to be addressed through the strengthening of collaborations of existing regional institutions to take up the mandate of developing molecular identification tools for regional breeds.

(5) It is important for countries that have livestock breeds in common such as international and regional trans-boundary breeds to work together for the purposes of standardisation and cost sharing.
(6) Special funds could be set aside for developing breed-specific molecular tools in disadvantaged regions of the world, to be managed by a global body such as the FAO.

IX. ACKNOWLEDGEMENTS

We are grateful to CHIRI Biosciences and Curtin University of Technology for the use of their facilities. We are particularly indebted to the anonymous reviewers for their invaluable comments. M.Y. was funded by a PhD grant from Curtin University International Strategic Research Scholarship (CIPRS).

X. REFERENCES


Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology


BRONXHAM, E. T. (2015). A case study on strains of Buša cattle structured into a meta-population to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties. *Frontiers in Genetics* 6, 73.


Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology


Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology


RISCHKOWSKY, B., PILLING, D. & COMMISSION ON GENETIC RESOURCES FOR FOOD AGRICULTURE. (2007). *The state of the world’s animal genetic resources for food and agriculture*. Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations, Rome, Italy.


Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology


Molecular Identification of Livestock Breeds: A Tool for Modern Conservation Biology


Table 1. Present status of efforts to identify livestock breeds using genomic methods.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genotyping platform</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian sheep breeds</td>
<td>Illumina OvineSNP50</td>
<td>Stepwise discriminant analysis</td>
<td>(Dimauro et al., 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canonical discriminant analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discriminant analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GENECLASS 2 software</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>POWERMARKER software</td>
<td></td>
</tr>
<tr>
<td>Italian cattle breeds</td>
<td>Illumina Bovine50</td>
<td>Principal components analysis, random forest</td>
<td>(Bertolini et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>BeadChip</td>
<td>regression</td>
<td></td>
</tr>
<tr>
<td>New Zealand sheep breeds</td>
<td>Illumina OvineSNP50</td>
<td>Regression and genomic BLUP</td>
<td>(Dodds et al., 2014)</td>
</tr>
<tr>
<td>Korean native and Holstein cattle breeds</td>
<td>Illumina HiSeq 2000</td>
<td>Mapping of Reads to Bovine Genome Assembly UMD 3.1</td>
<td>(Choi et al., 2014)</td>
</tr>
<tr>
<td>Yunnan (South China) chicken breeds</td>
<td>Microsatellite markers</td>
<td>Nei’s genetic distance, Hardy Weinberg analysis, GENALEX 6 Software, Weir &amp; Cockerham’s F&lt;sub&gt;ST&lt;/sub&gt;,</td>
<td>(Huo et al., 2014)</td>
</tr>
<tr>
<td>Swiss cattle breeds</td>
<td>Illumina Bovine50</td>
<td>STRUCTURE algorithm, BayesB, Partial least squares regression</td>
<td>(Frkonja et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>BeadChip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 worldwide cattle breeds</td>
<td>Bovine HapMap data set</td>
<td>Principal components analysis, Nearest neighbour classification Algorithm</td>
<td>(Lewis et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>SNP marker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European cattle breeds</td>
<td>Illumina Bovine50</td>
<td>Weir &amp; Cockerham’s F&lt;sub&gt;ST&lt;/sub&gt;, Delta, Wright’s F&lt;sub&gt;ST&lt;/sub&gt;</td>
<td>(Wilkinson et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>BeadChip</td>
<td>Principal component analysis methods</td>
<td></td>
</tr>
<tr>
<td>Japanese and USA cattle breeds</td>
<td>Illumina Bovine50</td>
<td>Allelic frequency method, BLAST programs, PCR restriction fragment length polymorphism</td>
<td>(Suekawa et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>BeadChip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian cattle breeds</td>
<td>Microsatellite markers</td>
<td>STRUCTURE algorithm, Wright’s F–statistics, MolKin V3.0 software</td>
<td>(Bozzi et al., 2009)</td>
</tr>
<tr>
<td>Local European and Asian chicken breeds</td>
<td>Microsatellite markers</td>
<td>Nei’s genetic distance, Hardy Weinberg analysis, GENECLASS 2 software, Multiple CO–inertia analysis</td>
<td>(Berthouly et al., 2008)</td>
</tr>
<tr>
<td>Breed Type</td>
<td>Markers Used</td>
<td>Methodology</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Japanese and Australian cattle</td>
<td>Microsatellite markers</td>
<td>Polymerase chain reaction–amplified fragment length polymorphism</td>
<td>(Sasazaki et al., 2007)</td>
</tr>
<tr>
<td>Iberian pig breeds</td>
<td>Microsatellite markers</td>
<td>PHYLIP software package</td>
<td>(Fabuel et al., 2004)</td>
</tr>
<tr>
<td>Asian and European pig breeds</td>
<td>Mitochondria DNA D-loop assay</td>
<td>Polymerase chain reaction–amplified fragment length polymorphism</td>
<td>(Kim et al., 2002)</td>
</tr>
</tbody>
</table>