CHAPTER TWO

Literature overview

2.1 Definition of indigenous, native and local fowl

This study refers to the characterization of fowl, which is found primarily in rural areas of South Africa, associated with backyard farming and or extensive small-scale agriculture. The terminology used for these birds is often confusing, as they are referred to as "indigenous", "native", "local" or even a "village chicken". Therefore, it was necessary to decide which term would be most appropriate to describe these fowl populations. The Concise Oxford Dictionary (1990) provides appropriate definitions for the following terms:

- Indigenous : occurring or living naturally in an area; not introduced; native
- Native : belonging by birth or origin to a specified country or place; inborn; indigenous as opposed to foreign or exotic; unaffected by artificial influences
- Local : a native inhabitant; of or pertaining to a place

The word indigenous, originating from the Latin: *indigena*, implies that the individual and its ancestry originated from a specific country or continent. As the common ancestor of the fowl is dated back to the Indus Valley 3250 BC (Moiseyeva, 1988) and was introduced to South Africa by settlers and traders (Ramsey *et al.*, 2000), the fowls could hardly be labelled as "indigenous" fowl. It could be argued that "local" fowl might be the most correct term for South African birds, but it could then mean that the locally bred Ross and other broiler lines (originating from imported stock) are now also "local" birds. The definition for native refers to place of birth and distinguishes the inhabitant from exotic or foreign but it also includes indigenous (*vida supra*).

Most of the fowls associated with rural and small-scale agriculture are found in the developing world and terminology ranges from indigenous fowl in Asia (Mukherjee, 1990) to local Malawi fowl in Africa (Safalaoh *et al.*, 1996). All breeds of fowl in this study originated from birds introduced to South Africa at least 350 years ago, left to

scavenge around the household and were not subjected to artificial selection over the years. Therefore, it was decided to use "native" in the characterization of the fowls in this study.

2.2 Conservation of native fowl populations

It is inevitable that selection, inbreeding and various crossbreeding systems may lead to the loss of genetic variation within breeds and that the breed itself may become extinct. For this reason, the scientific community identified the need for conservation of livestock resources, including poultry, during the late 1980's. During 1992 the Food and Agricultural Organization (FAO) launched a program for Global Management of Farm Animal Genetic Resources, with the main objective being to stimulate conservation activities and create an awareness of possible losses of genetic resources on an international basis (Scherf, 1995; Gandini & Oldenbroek, 1999). An information system, namely the Domestic Animal Diversity Information System (DAD-IS), forms one of the main components of the program and provides a list and description of all breeds in existence. During 1999 there was 332 cattle, 407 sheep, 123 goat, 156 pig and 213 horse breeds listed in DAD-IS from 37 European countries (Gandini & Oldenbroek, 1999). The biodiversity of 52 chicken breeds from Europe were assessed in a European Union project and a chicken DNA bank and poultry biodiversity database have been established at INRA Jouy-en-Josas (Weigend, 2000). In South Africa the Farm Animal Conservation Trust (FACT) was established in 1994 to facilitate and promote conservation of native animal genetic resources. Three South African native fowl breeds and one locally developed breed are listed by FACT (Ramsey *et al.*, 2000).

In order to make informed decisions on conservation of poultry breeds, Crawford & Christman (1992) emphasized that one should take into account the diverse groups of poultry and the reasons why they are kept. These authors identified the following categories:

- Food producers: indigenous types, middle-level and industrial.
- Fancy or exhibition poultry: kept by the hobbyists.

- Fighting stock: although illegal in many countries, these still exist.
- Wild ancestors of poultry: mostly endangered and the concern of aviculturalists.
- Inbred and specialised lines: developed for research purposes.

In this study only the food producers are of interest, as native fowls are mainly applied for household food production. A few companies in the world are responsible for breeding commercial stock that is bred from a relatively narrow genetic base. As these broilers and layers are selected for maximum production, they are replacing most of the middle-level poultry, the dualpurpose predecessor breeds, as well as the indigenous birds. Although the indigenous or native stock has a poor performance, relative to highly selected commercial lines, they do have the ability to survive in challenging environments. Very little is known about their potential to disease resistance and adaptation mechanisms (Crawford & Christman, 1992).

Most of the research efforts towards the conservation of genetic animal resources in Africa have been on large (cattle) and small (sheep and goats) livestock. Native chickens were mostly disregarded and no data on breeds or lines have been reported (Setshwaelo & Adebambo, 1992; Hofmeyr *et al.*, 1998). One reason for this lack of data is that the need for the conservation of livestock resources is still an issue of debate among scientists. Two different approaches to conservation are recognised: the utilizationist versus the preservationist (Mason & Crawford, 1993). The utilizationist is more concerned with the usefulness of the genetic resource and the loss of breed identity is of lesser importance, while the preservationist views the conservation of the breed on the long term as the main objective. In the developing countries, where food security is a problem, the approach for conserving the native fowl should be more towards utilisation, rather than preservation of breed characteristics.

The project launched by the Poultry Supply Unit of the Animal Improvement Institute of the Agricultural Research Centre at Irene, was not only to conserve the native fowl populations found in South Africa, but to promote their role in rural agriculture (Joubert, 1996). As described in chapter one the native fowl populations included in the project

was originally collected from rural areas of most parts of South Africa and the fowls referred to as Ovambo's were collected in Ovamboland in Namibia. FAO guidelines are followed in terms of population size to control inbreeding. The "Fowls for Africa" include populations, such as the Lebowa-Venda, Ovambo, Naked Neck and Koekoek, as well as middle-level breeds for example the New Hampshire, Rhode Island Red, Black Australorp and Cornish (Honeyborne [personal communication], 1999). Except for the Koekoek, New Hampshire and Australorp, the other populations have not been described according to phenotype or included in the breed standards of the South African Poultry Society. The Lebowa-Venda is in the process of being described for inclusion in the breed standards (Honeyborne [personal communication], 2001).

2.3 Origin of the native fowl in South Africa

The Potchefstroom Koekoek has for many years widely been used by South African farmers for egg production for household purposes. The breed originated from a cross between Black Australorp and the White Leghorn during the 1950's, at the former Potchefstroom Agricultural College. The Plymouth Rock was later included into the breeding program and eventually the Koekoek was registered as a South African breed with the South African Poultry Association in 1976 (Viljoen, 1986). These birds have a characteristic black and white speckled colour pattern, which is present in as many as nine different poultry breeds. The males inherit the bar gene, a sex-linked gene and they are easily distinguished, having light grey bars on the feathers, while the females are darker (Figure 2.1a & b). Koekoek is classified as a heavy breed, with the average adult body weight varying between 3-4 kg for cocks and 2.5 - 3.5 kg for hens (Viljoen, 1986). The average egg weight is 55.7 g and the colour of the eggs are brown (Ramsey *et al.*, 2000).

The New Hampshire originated in the United States, bred from Rhode Island Reds. The first importation to South Africa was in 1947. These birds are also classified as a heavy breed with adult body weight varying between 3.9 kg for cocks and 3 kg for hens. Plumage colour is a chestnut red with a light salmon colour on the breast area (Figure

2.2a & b). Egg colour is light brown. Although the New Hampshire is not native to South Africa, it was included as it is very often kept as a dualpurpose line in rural areas.

Figure 2.1a: A Koekoek cock and two hens.

Figure 2.1b: Koekoek fowls.

Figure 2.2a: A New Hampshire hen.

Figure 2.2b: New Hampshire fowls.

Traders from Malaysia and settlers that came with the Dutch East Indian Company most probably introduced the Naked Neck chicken to South Africa (Ramsey *et al.*, 2000). These chickens carry the major gene *Na-* for Naked Neck. The gene is inherited autosomal with incomplete dominance and was mapped on chromosome 1 of the chicken

genome (Pitel *et al.*, 2000). Chickens that are homozygous for the *Na*-gene have no feathers on the neck area, while the heterozygous have a little tuft of feathers on the lower portion of the neck. The *Na*-gene is associated with significantly less plumage cover than chickens not carrying the Naked Neck gene. Naked Neck chickens were found to have better heat resistance (Pech-Waffenschmidt *et al.*, 1995). Naked Necks of South Africa are very colourful - white, red and black feather combinations are found (Joubert, 1996) (Figure 2.3a & b). Egg colour varies from cream to light brown.

Figure 2.3a: A heterozygous and homozygous Naked Neck fowl.

Figure 2.3b: Naked Neck fowls.

The first recording of the Lebowa-Venda chickens was made during 1979 in Venda in the Northern Province. They are quite common in this area and the name Lebowa-Venda is associated with the two largest ethnic groups residing in the Northern Province. These chickens have white and black or white and brown plumage with shades of dark green on the feather tips (Joubert, 1996) (Figure 2.4a & b). The colour of the eggs is cream and also sometimes tinted (Ramsey *et al.*, 2000).

Figure 2.4a: A Lebowa-Venda hen.

Figure 2.4b: Lebowa-Venda fowls.

The Ovambo was originally found in the rural areas of Namibia and Ovamboland and the name refers to the geographical area where they were first distinguished by the local people as a group of chickens native to the area. These chickens have predominantly dark feathers. They tend to be aggressive if their young is in danger and they often roost in trees to avoid predators (Joubert, 1996) (Figure 2.5a & b).

Figure 2.5a: Ovambo hens.

Figure 2.5b: Ovambo fowls.

The Black Australorp was imported to South Africa from Australia during 1925, where they have been developed from the Black Orpington, which has been selected for egg production. Over the years different strains of Australorp were developed in South Africa with feather colours varying from white, golden "wheaten laced" and black. All the strains were primarily used as dualpurpose breeds, with an average mature body weight varying from 3.6 – 4.5 kg for cocks and 2.7 – 3.9 kg for hens (Viljoen, 1986). The Black Australorps in the "Fowls for Africa" program are a pure line, which originates from the former Poultry Research Unit at Potchefstroom.

Figure 2.6 The Black Australorp fowl

2.4 Markers for studying genetic variation in farm animals

Biochemical systems have been extensively applied over the years in genetic and physiological research of farm animals. Discoveries made by Landsteiner in the early

1900's on human blood group variability and Erlich & Morgenroth and Todd & White on blood groups in farm animals, set the scene for research in immuno genetics and genetic differences among animals (Hines, 1999). Stormont did further work during 1950 on the blood group systems in cattle. Biochemical polymorphisms, historically, have often been used in detection of genetic differences in farm animals. As early as 1940, Irwin and co-workers at the University of Wisconsin used blood group antigens for parentage verifications in the Holstein Friesians (Hines, 1999). The applications of blood groups proved to be a powerful tool in detection of incorrect parentage. Biochemical systems consist of a variety of proteins found in blood plasma, serum and milk such as haemoglobin, transferrin, albumin casein kappa and erythrocyte GSH, K and Na-concentrations. More recently biochemical polymorphisms were also applied in genetic characterization of cattle and goats (Baker & Manwell, 1980; Tucker, *et al.*, 1983; Arranz *et al*, 1996a). However, the development of molecular biology techniques during the late eighties has, opened up new ways for the study of genetics. DNA has become the alternative for the research of various genetic, breeding and physiological questions in farm animals.

2.4.1 Genetic markers

PCR (Polymerase Chain Reaction) technology

PCR technology has become an essential tool in molecular biology. It is a relatively easy *in vitro* method for amplification of a specific DNA-sequence from genomic DNA or a cDNA population (Strachen & Read, 1996). A PCR consists of a number of cycles of denaturation, annealing and extension. A basic PCR cycle will include the following steps:

i. DNA template is placed in a tube with specific primers (synthetic oligo-nucleotides), deoxynucleotides (dNTP's) and a heat stable polymerase enzyme, a buffer and $MgCl_2$.

ii. The mixture is heated (± 95°C) to denature or separate the two DNA strands.

iii. Temperature is decreased to ± 55°C to allow primers to bind/anneal to form a complementary strand. The annealing temperature will depend on the primer length and sequence.

iv. Temperature is increased to ± 72°C for DNA polymerase to act and add nucleotides to the 3' end of each primer, allowing for extension of new strands.

v. Cycle two will then commence by heating the reaction again.

A typical cycle takes approximately 5 minutes and amplification is repeated for 30-35 cycles on most PCR-machines. The product synthesized in one cycle serves as a template in the next, so that the original DNA is doubled in every cycle. The amplicon or PCR-product can then be visualized on an agarose or polyacrylamide gel (Nicholas, 1996; Turner *et al.*, 1998; Erlich, 1991). Often PCR requires optimization to ensure a good amplicon, distinguished as a defined band on the gel. Usually the annealing temperature and or $MgCl_2$ concentration may be adjusted towards obtaining optimal amplification.

The primers in a primer pair are designed to be of similar lengths (usually 18-30nt in length) with a similar G + C content, in order to anneal to the complementary DNA at approximately the same temperature. The primers will extend towards each other by the addition of the enzyme (Taq polymerase) and nucleotides (DNTP's) to the 3' end. A suitable heat stable DNA-polymerase enzyme is an essential component in PCR, as the enzyme initiates the synthesis of the new strands and must survive the first heating step at 95°C. *Taq* polymerase, derived from the thermophilic bacteria *Thermus aquaticus*, is most often used. Incorrect copying may occur using *Taq*, because it has no proof reading function. This refers to the presence of a 3' - 5' exonuclease activity, which reduces the chance of incorrect copying of nucleotides during the synthesis of the new strand. Other polymerase enzymes are available that can be used if a higher accuracy is required (Strachan & Read, 1996; Turner, 1998).

The source of DNA used for PCR can be genomic DNA (whole blood or tissue), or forensic specimens and ancient biological samples (Turner *et al.*, 1998). The aim of PCR is to multiply the given small segment of template DNA, to provide a fragment large

enough for cloning, sequencing or detection on a polyacrylamide gel. Prior knowledge is required of the sequence of at least a segment of the DNA to be amplified. This information is used to synthesize an oligonucleotide, or referred to as a primer, which is used to prime the synthesis of the new DNA strands (Nicholas 1996; Turner *et al.*, 1998).

PCR technology is used to amplify known sequences of a sample of DNA or for arbitrary priming of variable regions of the genome. It is therefore possible to use PCR for Variable Number Tandem repeats (VNTR's), Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) and Restricted Fragment Length Polymorphism (RFLP) (Erlich, 1991). All these markers can be used for estimation of genetic variability. PCR has the advantage of being a relatively fast, sensitive and reliable method. It cannot only amplify very small amounts of DNA, but also amplify degraded or poor sources of DNA (Strachen & Read, 1996; Erlich, 1991).

2.4.2 DNA-based markers

Repetitive DNA

The structure and composition of DNA provides the basis for the understanding of the genetic information stored, coded and transmitted for production of the proteins necessary for all metabolic functions. Genomic DNA consists of non-coding and coding DNA. Coding DNA encode for the synthesis of proteins, a process, which involves the translation of RNA, which is transcribed from the DNA template. Non-coding DNA makes up a large proportion of the genomes of the eukaryotes. These non-coding regions contain regulatory elements such as promoters and enhancers, but can in many cases also contain repetitive elements, eg. multiple repeats (Turner *et al.*, 1998). These repetitions include satellite DNA, which are thousands of tandem repeats in one site, as well as mini- and microsatellite DNA depending on the number of repeats. Many satellites usually range in size from 10-100 bases and microsatellites from 1-6 bases (Nicholas, 1996; Turner *et al.*, 1998). Although both minisatellites and microsatellites occur throughout the genome, the minisatellites tend to be concentrated in the telomere regions and sites associated with a high frequency of recombination (Nicholas, 1996). The number of

repeats in the satellite DNA is highly hypervariable and differs among individuals of a species. For this reason these repeats may be applied as DNA markers in the study of genetic differences.

Classification

Genetic markers associated with DNA are commonly grouped into Type I and Type II markers (O'Brien, 1991). Type I markers are usually associated with a gene with a known function, while Type II markers refer to anonymous gene segments. Alternatively, DNA-based markers are also grouped as clone/sequence-based markers (CSB) such as microsatellites and fingerprint markers (Dodgson *et al.*, 1997).

CSB markers require isolation of a DNA fragment and the determination of the sequence of the fragment. For Fingerprint markers the sequence of the DNA region is unknown and the markers include randomly amplified polymorphic DNA (RAPD) (Dodgson *et al.*, 1997). Table 2.1 provides a summary of the properties of DNA-based markers, which are often used in the studies of genetic variability in farm animals.

Table 2.1 Properties of DNA-based markers (Dodgson *et al.*, 1997)

	Clone Sequence Based (CSB)		**Fingerprint (FP)**		**Single Nucleotide Polymorphism (SNP)**
Variable	RFLP	Microsatellites	RAPD	Mini-satellite	
Genome distribution	Ubiquitous	Ubiquitous	Ubiquitous	Heterochromatien	Ubiquitous
Genome surveyed	sc & mr	Sc	Sc & mr	mr-hr	Sc
Typical PIC	Low	High	moderate	High	Low
Typical allele number	2	2-10	2	2	2
Inheritance mode	Co-dominant	Co-dominant	Dominant	Dominant	Co-dominant
Type of loci (O'Brien)	I and II	II > I	II	II	I and II
Reliability	High	High	Low	High	High
Speed of assay	Low	High	High	Low	High
Initial investment	Moderate	High	High	Low-moderate	High

Sc = single copy, mr = moderately repetitive, hr = highly repetitive and
PIC = polymorphic information content.

2.4.2.1 Restricted Fragment Length Polymorphism (RFLP)

RFLP generally refers to the differences in banding patterns obtained, from DNA fragments, after digestion with restriction enzymes. Restriction enzymes (RE) bind to specific sequences and cut the DNA at a specific cleavage site. The DNA fragments of different lengths are then subjected to electrophoresis and fragments will migrate according to their weights, the smaller fragments faster and the larger fragments slower (Nicholas, 1996). This application of RFLP is the more conventional use for example for detection of diseases. Restriction enzymes are enzymes that are produced by bacteria, which protects the bacterial cell against foreign DNA by cutting it into fragments. Restriction enzymes are usually named after the bacteria where they were found, for example in BamH1, the **Bam** refers to the **B**acillus **amy**loliquefaciens where it was found and **H** refers to the strain and **1** indicates that it was the first restriction enzyme obtained from that bacteria. Various Restriction enzymes have already been shown (BamHI, EcoRI SacI and TaqI) to be useful in obtaining RFLP patterns for haplotype identification in individuals (Spike *et al.*, 1996; Smith *et al.*, 1996). It is often required, that many enzymes need to be tested in the initial phase to be able to identify the polymorphism, but even then it is still an easy and relatively cheap marker to use (Dodgson *et al.*, 1997). Potential disadvantages of the RFLP technique are the dimorphic nature, since a RFLP only indicates the presence or absence of a cleavage site, and therefore does not provide a great deal of genotypic information. Large amounts of DNA are also required for RFLP analyses and the technique is relatively time consuming.

RFLP techniques are also used in DNA fingerprinting which is based on the detection of Variable Number of Tandem Repeats (minisatellites) by firstly digestion of the DNA with a RE, followed by Southern analysis and hybridization techniques. The steps required for producing a DNA fingerprint for an individual are briefly as follows (Nicholas, 1996):

i. Isolation of the DNA template from for e.g. blood, tissue or hair.
ii. Digestion of the DNA with a Restriction enzyme.
iii. Separation of the fragments with gel electrophoresis.

iv. Transfer of the fragments onto a nylon membrane (blotting).

v. Membrane is " baked" to allow for DNA fragments to attach to membrane (hybridization).

vi. The solution with the probe (labeled DNA) is added. The probe is single stranded and will attach to any complementary DNA fragment on the membrane.

vii. Fragments can then be visualized by autoradiography.

A number of bands can be observed as a result and each individual has a unique set, referred to as the DNA fingerprint (DFP), which makes this technique very useful for parentage verifications.

2.4.2.2 Microsatellites

Microsatellites consist of tandem repeats between one and six bp, repeated up to 60 times and referred to as simple sequence loci. These domains were first demonstrated by Hamada and colleagues, during the early eighties (Tautz & Renz, 1984; Tautz, 1989; Smeets *et al.*, 1989). Repeat units may consist of $(A)_n$, $(TG)_n$, $(CA)_n$ or $(AAT)_n$ repeat. For example in most vertebrates the $(CA)_n$ repeat is the most common motif (Beuzen *et al.*, 2000). Microsatellites are highly polymorphic due to the variation in the number of repeats. It is not uncommon to find up to 10 alleles per locus and heterozygosity values of 60% in a relatively small number of samples (Goldstein & Polack, 1997). Microsatellites are well-distributed in animal genomes and are multi-allelic, co-dominant and can be detected by PCR-technology (Tautz, 1989).

The function of these simple regions of short sequence motifs is not clear. Tautz (1989) indicated that they could be involved in gene regulation or act as signals for recombination, as a certain amount of crossing over takes place within $(GT)_n$ repeat sequences. However, tandem repeat loci may also have a function in the packaging and condensing of the DNA in eukaryotic chromosomes (Stallings *et al.*, 1991). It was for example found that $(GT)_n$ repeat sequences are much more frequent in euchromatin than heterochromatin and the $(GT)_n$ repeat could therefore be an important determinant in

distinguishing between hetero- and euchromatin (Stallings *et al.*, 1991). Microsatellites are found in both coding and non-coding regions of eukaryotic genomes and are generated by a mechanism referred to as slippage (Tautz *et al.*, 1986). Slippage occurs when normal pairing of repeats is altered during replication. A backward slippage causes an insertion of a repeat in the new strand and a forward slippage a deletion (Strachan & Read, 1996).

Microsatellites can be isolated by various methods. The development of microsatellite markers requires the construction of a genomic library. Briefly, the construction of a genomic library involves cloning the DNA of a specific species (e.g. chicken DNA) as follows: firstly, the genomic DNA is digested using restriction enzymes, which yield small DNA fragments. The DNA fragments are then cloned into vectors such as phages or plasmids, which allow proliferation in bacterial cells. The next step involves hybridization where thousands of clones are screened with synthetic polynucleotides such as $(TG)_{13}$, $(CAC)_5$, and $(GAT)_4$ which are labeled radio-actively with $^{32}P\gamma$-ATP. Positive clones are then isolated and sequenced (Crooijmans *et al*, 1993). The sequence information is used to synthesize PCR primers, which are then also tested on a panel of unrelated animals. This step is essential to ensure that primers work optimally and that there is no cross reactions (Crooijmans *et al*, 1993; Crooijmans *et al*, 1997). Primers are developed in pairs and usually labeled with a fluorescent dye for application on automated sequencing machines.

Microsatellites have been mapped for various species, including humans, mice, fruit flies, cattle, sheep, pigs and chickens (Goldstein & Pollack, 1997). As a consequence these elements have become most valuable markers in studies on genetic variability, parentage verifications and genome mapping projects.

2.4.2.3 Minisatellites

Minisatellites were first described as hypervariable tandem repeats, when found in the human genome. They were found to be longer repeats than microsatellites, consisting of

up to 200bp. This led to the use of the term Variable Number of Tandem Repeat loci (VNTR's) in reference to repetitive units that include mini- and microsatellites. Minisatellites are also well distributed through the eukaryotic genome (Bruford & Wayne, 1993). Minisatellite markers have certain limitations, as they do not uniformly mark the genome, the marker fragment is difficult to clone and they are dominant markers if the repeat is used as a probe, which reduces the potential information for genotyping (Dodgson *et al.*, 1997).

2.4.2.4 Random Amplified Polymorphic DNA (RAPD)

Random amplified polymorphic DNA (RAPD) is also a technique that can be applied in evaluation of genetic diversity (Smith *et al.*, 1996; Nicholas, 1996). The RAPD technique is based on the use of a number of short, arbitrary primers in one PCR reaction. These random primers will generate several amplification products, which will differ in size and may be characterized by simple agarose electrophoresis. RAPD markers have the advantage that they can be obtained at a reasonable cost and will generally amplify a range of fragments of most DNA and show polymorphism. The polymorphic bands obtained from a RAPD can be cloned for further analysis. A major disadvantage is that the RAPD primers are very sensitive to PCR conditions and may lead to poor reproducibility (Dodgson *et al.*, 1997).

Minisatellites, Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) have also been used widely for parentage determinations, genomic diversity and measurement of kinship (Gilbert *et al.*, 1991).

2.4.2.5 Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is a DNA fingerprinting technique that is based on the detection of DNA fragments, subjected to restriction enzymes, followed by selective PCR amplification. The DNA is cut with two restriction enzymes and double stranded adapters are then ligated to the ends of the DNA fragments to

generate the template for PCR. The specific adapter, ligated to the DNA fragment, determines the sequence of AFPL primers. Fragments are then separated and analyzed using gel electrophoresis (Vos *et al.*, 1995).

Two restriction enzymes are used, the one usually a rare and the other a frequent cutter. There are various reasons for using two enzymes. The frequent cutter is important for producing small fragments that will amplify well and have an optimal size for denaturation on a gel, while the rare cutter will reduce the number of fragments to be amplified. A further reduction in the number of fragments to be amplified is achieved by extending the PCR primers with 1 to 3 bp. This results in a further four fold reduction for every bp extension of the primer. Two enzymes will optimize the amplification and a large number of fingerprints may be produced with a small number of primers (Vos *et al.*, 1995).

The AFLP technique can be done at a reasonable cost and has extensively been used, particularly in the genome mapping of plants. It also has the advantage of a higher reproducibility than RAPD (Vos *et al.*, 1995).

2.4.2.6 Single Nucleotide Polymorphisms (SNP's)

SNP refers to the substitution of one nucleotide for another. It could also be an addition or deletion of one or more nucleotides, causing the polymorphism (Beuzen *et al.*, 2000). SNP's are bi-allelic markers, indicating a specific mutation (polymorphism) in two alleles only of a population. In order to obtain information from a SNP marker, which is similar in complexity to that obtained from a microsatellite marker, at least five SNP markers will be required (Beuzen *et al.*, 2000).

SNP's are also found in coding regions directly associated with the protein function and the inheritance pattern is more stable, making them more suitable markers for selection over time (Beuzen *et al.*, 2000). Most RFLP and AFLP markers are also the result of a

SNP in a restriction enzyme recognition site, which confirms the importance of SNP markers.

Microsatellite markers are currently the most reproducible and applicable method for genotyping and genome mapping in farm animals. It is however envisaged that the use of SNP's will increase as soon as optimization of arrays and application for specific species are confirmed.

2.5 Mapping of the chicken genome

The chicken genome has 39 pairs of chromosomes, eight large macro-chromosomes, 30 pairs of micro-chromosomes and two sex chromosomes (♀ ZW, ♂ ZZ) (Bitgood & Somes, 1990). The size of the chicken genome is estimated at 1.2 x 10^9 bp, which is small in comparison to the mammalian genome of 3 x 10^9 bp (Stevens, 1986; Crooijmans, 2000). The macro-chromosomes and micro-chromosomes account for approximately 130Mb and 12.5Mb respectively (Fillion, 1998). The chromosome lengths were found to be linear to the DNA content, the larger marco-chromosomes (one to eight) have the highest DNA content and represent up to 82% of the haploid genome together with the sex chromosomes (Smith & Burt, 1998). It was reported by Fillion (1998) that the micro-chromosomes represent up to 30% of the genome. The micro-chromosomes have a high gene density, even higher than the macro-chromosomes, which confirms their genetic importance (McQueen *et al.*, 1998).

The chicken is an ideal model for genetic mapping technology, because of its economic importance, a short generation interval and the potential to generate large full sib families (Crooijmans, 1993 *et al*; Fillon, 1998). Hutt published the first attempt for a genetic map of the chicken during 1936. Since then three different linkage maps were developed. The International Workshop on the Poultry Genome initiated the detection of polymorphic markers towards the mapping of the chicken genome. The choice and establishment of the reference families (East Lansing and Compton families) were also decided in this process (Crittenden *et al.*, 1993). The Compton population was

established at the Institute for Animal Health in Compton in the United Kingdom and the first map based only on DNA markers were published by Bumstead & Palyga (1992). The Compton line was based on a cross between an outbred and inbred White Leghorn line. These lines differ in susceptibility to different diseases. A single F_1 female was backcrossed to the male to generate the progeny for the reference family (Bumstead & Palyga, 1992). The East Lansing family was initiated in 1988 at the Michigan State University and was bred from a cross between a single, inbred male of Jungle Fowl origin and females from a highly inbred line of White Leghorn, to produce the F_1-progeny (Crittenden *et al.*, 1993; Cheng 1997; Cheng *et al.*, 1998). F_1-males were then backcrossed with the inbred Leghorn females to produce the BC1 (F_2-generation). A chicken genome map was constructed consisting of RFLP, RAPD and CR1 markers (Cheng 1997; Cheng *et al.*, 1998). The Wageningen Agricultural University generated a third population of 10 full sib families, using two commercial broiler lines from Nutreco BV. These lines are originally from the Plymouth Rock breed (Groenen *et al.*, 1998). The map generated was based primarily on microsatellite and AFLP markers (Herbergs *et al.*, 1999). A consensus linkage map combining the genotyping data of the East Lansing, Compton and Wageningen reference populations have been completed, consisting of 1889 loci, covering approximately 3800cM (Groenen *et al.*, 2000). The Chicken Database listed 586 genes, 2349, loci, 1251 microsatellites and 2959 available primers during September 2000 (www.Ri.bbsrc.av/chickmap/chickgbase/html).

As the commercial poultry industry is focused on improvement of quantitative traits such as growth, reproduction and disease resistance, genome mapping is aimed at discovering genetic markers, genes and Quantitative Trait Loci (QTL) (Cheng, 1994; Lamont *et al.*, 1996). These discoveries will also aid in assays for marker-assisted selection (MAS). Traits associated with growth have been so far been empasized and the most likely position for a QTL affecting body weight was found to be on chromosome 1 at 240cM (Groenen *et al.*, 1998; Van Kaam *et al.*, 1998). The HMGI-C gene also has been under investigation as a candidate gene for autosomal dwarf gene (*adw*) in the chicken, which is of importance in the broiler breeding stock programs (Ruyter-Spira *et al.*, 1998). Quantitative trait loci affecting the susceptibility to Marek's Disease are also being

studied. Vallejo *et al.* (1998) have indicated the mapping of a non-major histo-compatibility complex QTL affecting the susceptibility to Marek's disease.

Another important application of genome mapping is identifying genetic markers that are associated with disease resistance. For example, genotypes with a lower susceptibility to Marek's disease, were identified by Lamont *et al.* (1996). Markers are also mapped for application in evolutionary studies. A further important aspect of genome mapping is the comparison of maps of different species, to identify DNA markers that amplify in both species. Already, comparative mapping between chickens and humans is expanding as more chicken genes are found with a human homologue. A total of 154 autosomal conserved segments have been identified between the chicken and human which may be very useful in human genetic and disease studies (Schmid *et al.*, 2000). The most exciting application will surely be in what is referred to as functional genomics, where a connection is sought between the genotype and the physiological mechanism involved in the final expression of the trait (Dodgson & Cheng, 1999).

2.6 Genetic markers and variability in chickens

Genetic variability in chickens has mostly been studied by using fingerprint markers such as DFP's and RAPD (see Table 2.2). DFP's also have been applied in evaluation of potential QTL's in chickens (Dunnington *et al.*, 1993; Lamont *et al.*, 1996). Various commercial strains and indigenous types have been studied. Both types of markers were found to be useful tools in the estimation of genetic variability.

Until recently genetic characterization of chicken breeds, especially native and indigenous populations were limited. However, it can be expected that this situation will rapidly change with more microsatellite markers becoming available. Over a thousand microsatellite markers are available which is more than adequate for biodiversity studies.

Table 2.2 Fingerprint markers used in estimation of genetic variability in chickens

Aim of study	Breed/line of chicken	Markers used	Reference
DFP's for individual identification and linkage studies in poultry	Broilers: Cornish x White Rock breed, Layers: Leghorn Muscovy Duck, Turkey & Goose	DFP's	Hillel *et al.*, 1989
Genetic variation	Two White Leghorn lines, 3 commercial broilers lines, Rhode Island Red, & 8 exotic breeds	DFP's	Ponsuksili *et al.*, 1994
Determination of genetic distance	French broiler breed, random bred and inbred lines of White Leghorn	DFP's	Kuhnlein *et al.*, 1989
Genetic characterization of chicken lines	Ten lines of White Leghorn, 2 Fayoumi & 1 Spanish breed	DFP's & RAPD	Plotsky *et al.*, 1995
Genetic variability among layers and their correlation with performance	Nine different lines bred from commercial strains imported form Romania & USA	DFP's	Meng *et al.*, 1996
Relatedness & diversity in Chickens & turkeys	Rhode Island Red, White Plymouth Rock, single comb White Leghorn & Araucona & Turkeys	RAPD	Smith *et al.*,1996
Estimation of relatedness in White Leghorn lines	White Leghorn	RAPD	Deepak *et al.*, 1998

In a study of genetic variability with native lines from Egypt, India, Indonesia and Thailand, a higher heterozygosity was found among the lines using microsatellite markers, compared to DFP's (Ponsuksili *et al.*, 1996). Microsatellite markers have been successfully applied in characterization of waterfowl (Fields & Scribner, 1997) and are frequently used in studying genetic variability in other mammalian species such as sheep, pigs and cattle (Buchanan *et al.*, 1994; Van Zeveren *et al.*, 1995; MacHugh *et al.*, 1997). Microsatellite polymorphism has been found to be applicable for determination of genetic variation in commercial broiler and layer lines (Crooijmans *et* al., 1996b). Vanhala *et al.* (1998) also applied microsatellite markers for evaluation of variability and genetic distance in commercial egg and broiler lines. Genetic relationships among Japanese native chicken breeds based on eight microsatellite markers were also shown (Takahashi *et al.*, 1998). Microsatellite markers were also shown to be accurate and reliable for studies of genetic biodiversity in a study by Zhou & Lamont (1999) on genetic characterization of highly inbred chicken lines. There is a tendency towards

microsatellite markers as the preferred markers in genetic studies of chicken and other mammalian species.

2.7 Measurement of genetic variation

Animal populations differ in size and may change as they are continuously subjected to the forces of natural selection, migration and mutation. In livestock populations, artificial selection plays a major role in population changes. Selection, migration and mutation all may lead to non-random or directional changes in the allele frequencies of the population (Hartl, 1988).

Estimations of genetic variation include gene diversity or heterozygosity and genetic distance when using genetic markers such as microsatellites. Phylogenetic trees can be constructed from sequence data or microsatellite data to illustrate genetic relationships among the populations studied (Nei, 1987; Weir, 1996).

2.7.1 Gene diversity (Heterozygosity)

The average heterozygosity (H) over all loci in the genome as described by Nei (1987) takes into account the number of loci and number of individuals per locus. It is assumed that sampling was done from the loci and of the genes at each locus. For a single locus h can be estimated as follows:

$$\mathbf{h = 2n\,(1 - \sum x^2_i)/(2n - 1)}$$

and for more than one locus:

$$\mathbf{H = -\sum\, h_k\, /r}$$

Where: h_k = the value of h for k_{th} locus

n = the number of individuals sampled

x^2 = the population frequency of the genotype at the locus

r = the number of loci studied

The number of individuals studied will influence the variance. Intralocus variance will be influenced by the number of individuals sampled at each locus and can be reduced by increasing the number of individuals, while the interlocus variance can be reduced by increasing the number of loci studied (Nei, 1987). This is an important aspect to consider in estimation of genetic variance and it is recommended to increase the number of loci, rather than the number of individuals (Nei, 1987; Goldstein & Pollock, 1997).

2.7.2 Measures of population structure

Population structure is often measured according to the changes in the proportion of heterozygous individuals. The complexity of heterozygosity (H) in a population can be described by using the hierarchical F-statistics, originally proposed by Wright in 1943 (Hartl 1988; Nei, 1987; Weir, 1996). There are three parameters applied for measuring reduction in H or deviations of genotype frequencies. These are the F_{IS}, F_{IT} and F_{ST} values, also referred to as Fixation Indices (Table 2.3).

The fixation index (F_{ST}) is widely used as a measure of genetic difference between populations. F_{ST} may vary between a theoretical minimum of 0 (no genetic change) to a maximum of 1 (fixation of the allele). Guidelines for interpretation suggest that 0.0 - 0.05 indicate little genetic difference, 0.05 - 0.15 moderate difference and 0.15 - 0.25 a large difference. Values above 0.25 will indicate a very large difference, but is not common (Hartl, 1988). It should be noted that low values do not imply that the difference is negligible.

Table 2.3 Equations for F-statistics (Hartl, 1988; Nei, 1987)

Parameter	Equation	Significance
$\mathbf{F_{IS}}$	$\mathbf{F_{IS}} = H_S - H_I / H_S$	Comparable with inbreeding coefficient – H relative to sub-population
$\mathbf{F_{IT}}$	$\mathbf{F_{IT}} = H_T - H_I / H_T$	Effect of sub population on total population
$\mathbf{F_{ST}}$ (Fixation Index)	$\mathbf{F_{ST}} = H_T - H_S / H_T$	Reduction in individual H relative to total population

Where: H_I = H of an individual in a sub-population
I_S = expected H of an individual in a sub population with random mating
H_T = expected H of an individual in a total population with random mating

Original statistics for measuring genetic variation was based on biochemical markers (allozyme loci) (Nei, 1987). Goldstein *et al.* (1995) and Goldstein & Pollock (1997) suggested that statistical measures for microsatellite data should be adapted due to the mutation rate, distribution of mutation size, number of repeats and asymmetry during mutation, associated with microsatellites. The major difference for microsatellite loci is a stepwise mutation process usually in one or two steps, but larger mutations have been observed. The size of the new mutant allele is often dependent on the size of the allele that has mutated (Slatkin, 1995). To prevent a biased estimate of genetic variability in the populations studied, it is critical to consider the process of mutation in microsatellite markers. An estimate $R_{ST,}$ was proposed as an alternative to F_{ST} by Slatkin (1995), where the R_{ST} is based on a fraction of the total variance of the allele size between populations.

$$\mathbf{R_{ST} = \tilde{S} - S_W / \tilde{S}}$$

Where: S_w and $\tilde{S}$ are proportional to the within-population total variances.

In simulations performed by Slatkin (1995) using F_{ST} and R_{ST}, it was concluded that R_{ST} performed better as it takes stepwise mutation into account. If the time-scale of interest for the population studied is short, F_{ST} is still an acceptable measure, but for studies over a longer period of time F_{ST} may lead to biased estimations. Pérez-Lezaun *et al.* (1997) also indicated R_{ST} as an alternative for microsatellite analyses, as F_{ST} does not take mutational relationships among alleles into consideration.

2.7.3 Genetic distance

Genetic distance is used to explain genomic differences or similarities between two populations, which is usually a function of allele frequencies (Nei, 1987). The distance could be zero, if no differences were observed or set to a maximum of 1, if there were no common alleles found at a common locus (Weir, 1996). Genetic distance is often used to estimate time of divergence in evolutionary studies. Depending on the study, different distances may be required, as it is used in construction of phylogenetic trees. The

distance measures proposed by Nei & Roychoudhury (1974) and Nei (1978) are most widely used, where I refers to the genetic identity and D to the genetic distance (Weir, 1996). Table 2.4 summarizes a few of the distance measures that can be applied in analyzing molecular data.

Table 2.4 Distance measures

Distance	Equation	Reference
Nei standard genetic distance D_S	$\mathbf{I} = J_{XY}/\sqrt{J_X J_Y}$	Nei (1987)
	$\mathbf{D_S} = -\ln[J_{XY}/\sqrt{J_X J_Y}$	Takezaki & Nei (1996)
Nei minimum genetic distance D_M	$\mathbf{D_M} = (J_X + J_Y)/2 - J_{XY}$	Takezaki & Nei (1996)
Roger's distance D_R	$\mathbf{D_R} = \frac{1}{\gamma} \Sigma\sqrt{\frac{\Sigma^{mj}_{I}(x_{ij} - y_{y})^2}{2}}$	Takezaki & Nei (1996)
Goldstein's distance $(\delta_\mu)^2$	$(\delta_u)^2 = \Sigma^{r}_{j}(\mu_{xi} - \mu_{yj})^2/\gamma$	Takezaki & Nei (1996)

Estimations using classical measures of genetic distance are usually based on the infinite alleles model or IAM that is not based on an evolutionary model. These distances, including the chord distance (D_c) and minimum distance (D_m) make use of allele frequencies shared between populations. The distances do not increase linearly with time and are mostly useful for closely related populations or taxa. They are less accurate for estimation of larger distances (Goldstein & Pollock, 1997).

Microsatellites may have certain limitations in phylogenetic reconstruction due to the nature of their mutation process (Goldstein & Pollock, 1997). According to Bruford & Funk (1999) changes are observed in allelic frequency of microsatellites between populations and it is estimated that mutation rates may vary form 10^{-4} and $5x10^{-6}$. New alleles may occur because of polymerase slippage during replication. The majority of the mutations is in one or two steps and therefore requires alternative statistical

analysis for estimation of distance and phylogeny. For microsatellites three distances are suggested that is based on a stepwise mutation model (SMM) that allows linear increases with time.

Distance based on SMM is dependent on the variation within populations and if population sizes vary, which is often the case, there is also intra-population variance to consider. Goldstein *et al.* (1995) defined a distance measure $(\delta_\mu)^2$ that deals with the variance term and was suggested for tree construction when studying distant populations. The distance increases linearly with time, but has a lower variance. The $(\delta_\mu)^2$ distance described by Goldstein *et al.* (1995) and Goldstein & Pollock (1997) takes differences in allele size into account and a stepwise mutation in each direction of samples x and y (Table 2.4).

Phylogeny is used to describe relationships as a genetic tree. Beside the distance measure applied, the choice for the tree will depend on the aim of the tree, whether only grouping of populations are of interest or describing evolutionary relationships (Nei, 1987).

Takezaki & Nei (1996) evaluated classical distance measures and concluded that estimates for 10 loci tend to be low, irrespective of a high or low H. Phylogenetic trees based on a less than 10 loci may be questionable. The efficiency of a tree will depend on the sample error and linear relationship with time. Takezaki & Nei (1996) suggested that different distance measures might be required for construction of branches of trees. When studying current populations, where evolutionary pathways are unknown, the standard unbiased genetic distance was found to be acceptable and both the NJ and UPGMA methods can be used for tree construction.

In conclusion, it could be stated that there are a wide range of statistical programs available for the analysis of molecular genetic data (some described under 2.10). Choosing the most appropriate program will depend on the nature of specific data set and the final goal of the study. A summary of programs can be found at the following websites: http://www.lcp.ucl.ac.be and corba.ebi.ac.uk/Biocatalog/Population.genetics.