**COURSE GUIDE**

**ANP 507**
**ANIMAL BREEDING AND LIVESTOCK IMPROVEMENT (2 Units)**

**Course team** Dr. C. E. Isidahomen (Writer/Developer)
Ambrose Alli University Ekpoma

**NATIONAL OPEN UNIVERSITY OF NIGERIA**

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng



Printed 2023

ISBN: 978-978-058-416-2

**CONTENTS** **PAGES**

## INTRODUCTION

The course ANP507 (Animal Breeding and Livestock Improvement), is a two (2) credit unit course designed for 500 level undergraduate students pursuing a degree in Agricultural Science. The course is expected to provide a good knowledge base for the future manpower for genetic improvement of Nigerian livestock resources towards a sustainable production of livestock. It explains the rudimentary Production traits, their measurement and evaluation, selection for breeding for improvement of livestock performance. Breeding systems and selection methods. Performance testing, progeny testing.. The course also provides basic knowledge for biotechnological. Identifying and incorporating genetic markers and major genes in animal breeding programmes. DNA tests and segregation analysis for genetic disorders. Determining associations between genetic markers and quantitative test loci (QTL) applications in livestock genetic improvement and the development of genotypes that are adapted to Nigerian environment. The course will provide a basic foundation for students intending to take up Animal breeding and Livestock Improvement as a Career in the future. The course is divided into seven (7) units with unit one to seven units. Each unit begins with a clear introduction and statement of objectives followed by the main content. The conclusion, summary and references (for further reading) were also provided for each unit. Tutor marked assignments were provided for each unit to enable you attempt some questions on the topics treated for onward submission to your tutor. The Course Guide provides you with access to brief information and overview of the course content, course duration, what you are expected to know in each unit, what course material you need to use and how you can systematically go through the course materials.. Thus, we intend to achieve the above through the following broad aim and other specific objectives.

## COURSE AIM

The major aim of this course is to treat the fundamental principles of animal genetics and breeding through the highlights of the basic knowledge of breeding and livestock improvement. Using Production traits, their measurement and evaluation, selection for breeding for improvement of livestock performance. Breeding systems and selection methods. Performance testing, progeny testing. Identifying and incorporating genetic markers and major genes in animal breeding programmes.DNA tests and segregation analysis for genetic disorders. Determining associations between genetic markers and quantitative test loci (QTL).

**MAIN COURSE**

**CONTENTS** **PAGE**

# MODULE 1 PRODUCTION TRAITS, THEIR MEASUREMENTS AND EVALUATION

The purpose of animal breeding is not to genetically improve individual animals once an individual is conceived; it is a bit late for that-but to improve animal **populations,** to improve future generations of animals. To this task breeders bring two basic tools: selection and mating. Both involve decision making. In selection, we decide which individuals become parents, how many offspring they may produce, and how long they remain in the breeding population. In mating, we decide which of the males we have selected will be bred to which of the females we have selected. This chapter examines both kinds of decisions from a broad perspective. Cattle Beef/milk productions. (Pregnancy, Calving ease, Birth weight, Weaning weight ,Yearling weight, Mature weight, Hip height, Pelvic area, Feed conversion (feed per gain),Scrotal circumference, Breeding soundness, Back fat thickness, Days dry, Calving interval, Milk yield ,Fat in milk (%) ,Protein in milk (%)) Horses (Wither height, Mature weight, Time to trot 1 mile Time to run ~ mile Time to run 1 mile Weight started (draft) Cutting score, Placing (in a race or show) Winnings), Swine (Pregnancy, Litter size (number born alive) Litter size (number weaned) Weaning weight, 21-day litter weight Days to 230 **lb,** Feed conversion (feed per gain) Loin eye area Back fat thickness),Sheep(Pregnancy, Number born, Birth weight ,60-day weaning weight Yearling weight, Loin eye area, Grease fleece weight, Clean fleece weight Staple length, Breeding soundness, Poultry(Number of eggs in first year (layers) Egg weight (layers),Hatchability (chickens),Feed conversion ratio (broilers), Hot carcass weight (broilers), Mature body weight (broilers), Shank length (turkeys) Breast weight (broilers).

# MODULE 2 SELECTION FOR BREEDING FOR IMPROVEMENT OF LIVESTOCK PERFORMANCE

**CONTENTS**



## 1.0 INTRODUCTION

### 1 Measuring Performance

In order to select animals, we must first measure performance (phenotype) on eligible candidates for selection. Systematic measurement of performance in a population is called **performance testing.** Performance testing programs vary among species and breeders within species. A progressive beef cattle breeder's program, for example, might include the recording of birth date, calf birth weight and calving ease score at calving time; weaning date, calf weaning weight, cow weight, and cow pregnancy status at weaning time; feed intake from weaning to yearling age; and weigh date, yearling weight, hip height, pelvic dimensions, back fat thickness or ultrasound measurements, scrotal circumference, and breeding soundness score at yearling time.

Performance Testing: Systematic measurement of performance (phenotype) in a population.

Performance testing programs are widespread in traditional livestock species (beef and dairy cattle, swine, poultry, and sheep) in developed countries. Seed stock producers commonly take part in such programs, reporting the data they record to breed associations or government agencies. Commercial producers may do performance testing also. However, because of the labor and expense involved in recording performance data, commercial testing programs are typically less elaborate than seed stock programs. Listed in Table 1.1 are traits commonly measured in several species?

## 2.0 OBJECTIVES

- You will understand what production traits means
- You will understand their measurement and evaluation

## 3.0 MAIN CONTENT

## 3.1 Measuring Performance

In order to select animals, we must first measure performance (phenotype) on eligible candidates for selection. Systematic measurement of performance in a population is called **performance testing.** Performance testing programs vary among species and breeders within species. A progressive beef cattle breeder's program, for example, might include the recording of birth date, calf birth weight and calving ease score at calving time; weaning date, calf weaning weight, cow weight, and cow pregnancy status at weaning time; feed intake from weaning to yearling age; and weigh date, yearling weight, hip height, pelvic dimensions, back fat thickness or ultrasound measurements, scrotal circumference, and breeding soundness score at yearling time.

**Performance Testing: Systematic measurement of performance (phenotype) in a population.**

Performance testing programs are widespread in traditional livestock species (beef and dairy cattle, swine, poultry, and sheep) in developed countries. Seed stock producers commonly take part in such programs, reporting the data they record to breed associations or government agencies. Commercial producers may do performance testing also. However, because of the labor and expense involved in recording performance data, commercial testing programs are typically less elaborate than seed stock programs. Listed in Table 1.1 are traits commonly measured in several species?

**Selection Using Information on Relatives**

Most animal breeders are unlikely to limit themselves to individual performance information alone in making selection decisions. They will use information on relatives as well. For example, when a dog breeder purchases an eight-week-old puppy from another breeder, she probably does not base her choice on just the conformation and personality characteristics evident in such a young puppy. She wants to evaluate those same traits in the

littermates, the dam, and the sire. She might want to see a copy of the puppy's extended pedigree to learn more about its ancestors. Similarly, when beef cattle breeders evaluate a sire to use via artificial insemination (A.I.), they look further than the sire's own performance for growth rate. They want to know something about the growth performance of his progeny.

**Dam: A female parent. Sire: A male parent**

The above examples illustrate the use of two different types of information (data) on relatives: pedigree data and progeny data. By examining the young puppy's parents, littermates, and extended pedigree, the dog breeder is using pedigree data. She is trying to learn something about the genes made available to the puppy through its parents. Beef cattle breeders, on the other hand, are using progeny data. They are trying to learn something about an A.I. sire's genes by evaluating the performance of his offspring.

**Pedigree Data: Information on the genotype or performance of ancestors and (or) collateral relatives of an individual.**

As the above examples should make clear, the information used to make Selection decisions can be subjective, objective, or something in between. The pedigree data used by the dog breeder are, for the most part, subjective. The puppy's Papers may include some semi objective information on show championships won by ancestors, but the breeder's observations on conformation and personality are essentially subjective in nature. In contrast, the progeny data used by beef cattle breeders are relatively objective. They consist of carefully measured (we hope) weights of animals taken at specific ages.

**Table 1.1** Commonly Measured Traits

| **Species** | **Trait** |
|---|---|
| Cattle (beef): | Pregnancy |
| | Calving ease |
| | Birth weight(kg) |
| | Weaning weight(kg) |
| | Yearling weight(kg) |
| | Mature weight(kg) |
| | Hip height(cm) |
| | Pelvic area (cm) |
| | Feed conversion (feed per gain) |
| | Scrotal circumference(cm) |

| | |
|---|---|
| | Breeding soundness |
| | Back fat thickness (cm) |
| Cattle (dairy) | Days dry |
| | Calving interval Milk yield |
| | Fat in milk (%) |
| | Protein in milk (%) |
| Horses: | 1 Wither height |
| | 2 Mature weight |
| | 3 Time to trot 1 mile |
| | 4 Time to run ~ mile |
| | 5 Time to run 1 mile |
| | 6 Weight started (draft) |
| | 7 Cutting score |
| | 8 Placing (in a race or show) |
| | 9 Winnings |
| Swine | 1 Pregnancy |
| | 2 Litter size (number born alive) |
| | 3 Litter size (number weaned) |
| | 4 Weaning weight |
| | 5 21-day litter weight |
| | 6 Days to 230 **lb** |
| | 7 Feed conversion (feed per gain) |
| | 8 Loin eye area |
| | 9 Back fat thickness |
| Poultry | 1 Number of eggs in first year (layers) |
| | 2 Egg weight (g) (layers) |
| | 3 Hatchability (%) (chickens) |
| | 4 Feed conversion ratio (broilers) |
| | 5 Hot carcass weight(kg) (broilers) |
| | 6 Mature body weight(kg) (broilers) |
| | 7 Shank length (cm) (turkeys) |
| | 8 Breast weight (kg) (broilers) |
| Sheep: | 1 Pregnancy |
| | 2 Number born |
| | 2 Birth weight |
| | 3 60-day weaning weight |
| | 5 Yearling weight |
| | 6 Loin eye area |
| | 7 Grease fleece weight |
| | 8 Clean fleece weight |
| | 9 Staple length |
| | 10 Breeding soundness |

## 6.0 TUTOR MARKED ASSIGNMENT

1. Explain how selection causes changes in the performance of future generations of a population
2. Why is a selection generally more effective for highly heritable traits than for lowly heritable ones?

## 7.0 REFERENCES/FURTHER READING

Richard, M Bourdon (2000).Understanding Animal Breeding 2nd Edition. Pp 1-538

# MODULE 3 BREEDING SYSTEMS AND SELECTION METHODS

## CONTENTS



## 1.0 INTRODUCTION

Selection: The process that determines which individuals become parents, how many offspring they produce, and how long they remain in the breeding population. Selection is the process that determines which individuals become parents, how many offspring they produce, and how long they remain in the breeding population. Most of us are familiar with the term **natural selection.** Natural selection is the great evolutionary force that fuels genetic change in all living things. The term conjures up visions of fossil records, species creation, gradual anatomical and physiological changes, and mass extinctions. We commonly think of natural selection as affecting wild animals and plants, but in fact it affects both wild and domestic species. All animals with lethal genetic defects, for example, are naturally selected against-they never live to become parents. Animal breeders cannot ignore natural selection, but the kind of selection of primary interest to them is called **artificial selection;** selection that is under human control. Artificial selection has two aspects: **replacement selection** and **culling.** In replacement selection we decide which individuals will become parents for the first time. Replacement selection gets its name from the fact that we select new animals to replaceparents that have been culled. These new animals are termed replacements.

Animal breeders cannot ignore natural selection, but the kind of selection of primary interest to them is called **artificial selection;** selection that is under human control. Artificial selection has two aspects: **replacement selection** and **culling.** In replacement selection we decide which individuals will become parents for the first time. Replacement selection gets its name from the fact that we select new animals to replace parents that have been culled. These new animals are termed replacements.

**Natural Selection: Selection that occurs in nature independent of deliberate human control.**

We normally think of replacements as being young animals. When you choose the pups in a litter, the lambs in a flock, or the calves in a herd to be kept for breeding purposes, you practice replacement selection with young animals. Broadly speaking, however, replacement selection need not be confined to young animals. If you were a dairyman and you chose to use for the first time a well known bull via artificial insemination (A.I.); you would still be practicing replacement selection. The bull is not young, nor will he be a parent for the first time, but he will be a parent for the first time in your herd.

## 2.0 OBJECTIVES

- You will understand breeding systems
- You will understand the selection methods
- You will understand the merit and demerit breeding systems and selection methods

## 3.0 MAIN CONTENT

## 3.1 Selection methods

This is the methods used by breeders to make long-term genetic change in animals.

(a) Tandem: is selection for one trait at n time improved, then for another? This is the most efficient method if only one trait needs improvement. With more traits it is inefficient.
(b) Independent culling: can be applied for two or more characters.

A minimum level is established for each trait below which animals are culled. Animals which satisfy requirements for all traits are retained. The number of animals which can be kept decreases as the number of traits under consideration increases.

(c) Selection index: is the most efficient method. It uses one single value for any number of traits, each of which is weighted by its economic value. The value I equal the sum of traits each of which is multiplied by a certain factor b (regression coefficient).

$I = b_1X_1 + h_2X_2 + \ldots\ldots + {}_1b_nX_n$

To construct a selection index information is needed on the phenotypic and genotypic covariance) gross and genetic correlation) between each pair of traits, and then of course the economic values (b).
All three methods of selection are also called mass or individual selection because selection is based on the individuals record. Mass selection is simple, easy and effective with high heritability.

**Selection systems:** Besides individual records information on relatives is used or instead. This applies when traits are measurable in every individual (growth rate). Records on relatives must be used in two cases.

(i) For sex-limited traits (milk production, egg production);
(ii) For data becoming available at slaughter (carcass traits).
(a) Family selection: Is used when comparing average performance of families (litters) and selecting whole families, no matter whether there are outstanding or poorer individuals in the group. It is used when heritability of trait (s) is low, when there is little variation in the common environment, when families are large. However, intensity of selection is lowered, much space is needed for many animals, and some danger of inbreeding is involved Figure : shows family selection (From Johnsson-Rendel, 1968)

| Family | Average daily gain in grams | x |
|---|---|---|
| 1 | A B CD | 645 |
| 2 | E F G H | 675 |
| 3 | I J K L | 666 |
| 4 | M N O P | 680 |
| 5 | Q R S T | 656 |
| | „600„625„650„675„700,725,750 | 656 |

If 8 animals are selected, the two best families (2 and 4) are taken. Questions: What was the selection differential?

(b) Within Family Selection: (From Falconer, 1970).If 10 animals are needed, you select the two top individuals in each family, no matter how high or low the family average is.
Note: Individual or mass selection would be the choice or the best performers.

(c) Combined selection: applies both procedures.(d) Sib selection: uses information from sib (full brother or sister) average excluding the individual under consideration, while in family selection it is included. The above systems are based on information from contemporaries or collateral relatives.

(e) Progeny or offering averages: can be used for selection of a parent. It is widely applied in progeny testing (to be discussed later). This system lengthens the generation interval, involves large numbers of animals and is time – and space consuming.

(f) Pedigree selection: can only be used together with other information, or if alone in the case no other information available. Ancestors‟ performance records become less important the farther they are removed in the pedigree.

(g) Recurrent selection: is a scheme which selects for combing ability by testing crosses of individuals form one population (line) against those forms a tester line which has a proven combining ability. The tester line is usually high inbred. Individuals in the line tested are selected on the basis how well they cross. Crossbreds, but this will depend on the purebreds used; some do not combine well.

(h) Reciprocal recurrent selection (RRS): involves animals of two populations tested against each other by reciprocal crossing, improving each line simultaneously on the crossbred

performance. There is no tester line. The two lines need not be highly inbred. But again purebreds are selected, not crossbreds.

## 7.0 REFERENCES/FURTHER READING

Falconer, D.S. (1970). Introduction to Quantitative Genetics. Oliver and Boyd.

Johansson, I. and J. Rendel. (1968). Genetics and animal Breeding. Oliver and Boyd.

Lerner, I..M. 1958. The Genetic Basis of Selection. Campanand Hill.

Richard, M Bourdon (2000).Understanding Animal Breeding 2nd Edition. Pp 1-538

# MODULE 4 PERFORMANCE TESTING AND PROGENY TESTING

## CONTENTS



## 1.0 INTRODUCTION

The performance test on individuals and on brothers and sisters is used to determine which animal will be selected for breeding purposes. These male animals are then progeny tested as the final test of breeding value. One of the most important consequences of performance testing is that it leads to the scrutinization of the relative importance of the different traits for which we select. This not only applies to measurable but also to immeasurable attributes. In pigs for instance, performance testing has highlighted the importance of good legs and the enormous problem the pig industry has in this respect. The same applies to excessive folds in Merino sheep in harsh environments. Animal breeding is extremely complex in the sense that different characteristics, some positively correlated, some negatively correlated and others un-correlated, some very important, others less important, make up the total economic and breeding worth of an animal. Without figures it is virtually impossible to select sensibly for such a very complex combination. Also bear in mind that the economic value and the breeding value of an animal is not necessarily the same thing, as some traits are readily passed on to the offspring while others have a low heritability.

## 2.0 OBJECTIVES

- You will understand performance testing and procedures involved
- You will understand the merit and demerit of performance testing
- You will understand progeny testing and its procedures
- You will understand the merit and demerit of performance testing

## 3.0 MAIN CONTENT

### 3.1 Performance testing

Performance testing today forms the basis of breed improvement of nearly all kinds of livestock in all the developed Western countries of the world. In South Africa, too, performance testing is being accepted readily as an indispensable aid in animal improvement. In fact, as far as methods and techniques are concerned, South Africa can be regarded as a world leader in many aspects.

The obvious reason for the success obtained by using performance testing is that it leads to more accurate selection of superior breeding stock. This, however, is only part of the whole story and it must be stressed that performance testing merely supplies data and that breeders differ in their ability to utilize these data in the same way that they differ in their ability to select animals efficiently without objective measurements. Given the best measuring technique and data processing system, the incapable breeder could still make a terrible hash of his breeding enterprise. Performance testing invariably has, as a first consequence, the formulation of breeding objectives which are sensible, realistic and based on fact. Without performance testing the stud breeder could easily fail to appreciate the exact needs of his final customer, the commercial producer. An example of how easily this can happen is the fact that many beef cattle breeders insisted on selecting for traits such as coat colour, shape of horns, etc., while the commercial producer's needs shifted to economically important properties such as fertility and growth rate.

#### 3.1.1 Performance Testing Procedure

The performance testing procedure in Irish Cattle Breeding Federation is as follows;

1. The bulls are made to enter a pre-entry isolation where they are clipped, dosed, treated for lice and sorted into pens based on breed and weight. For IBR- Irish BreedingResearch, the

pre-entry isolation period is 30 days and during which the bulls are weighed if they meet the testing requirements.

2. The bulls are housed primarily indoors and they also have access to outdoor pen.
3. The bulls are fed ration at a less than ad lib rate until such time as they have become acclimatized to the meal. This is very important to maintain the health of the animal with a change in diet. They are thereafter moved to ad lib feeding, and also given a ration of hay twice a day. They are however provided with ad lib water.
4. Health checks of the entire herd are carried out three times daily.
5. Bulls are provided with vitamins and mineral lick, and a mineral revitalize dose if required‟
6. During the test the animals are weighed on a three-weekly basis to assess their growth rate.
7. The bulls are washed every four to six weeks during the test, depending on weather.
8. Bulls are exercised on an open pasture paddock twice a week for a three-hour period each time if deemed necessary.

   All the traits that the animals are been tested for are measured, combined at the end of the test and stored in the ICBF database‟

### 3.1.2a: Advantages of performance testing

1. The results allow a genetic profile of all animals with records and related to be computed.
2. When performance test commence on the bulls, their new information can help to improve their indexes in an equal measure.

### 3.1.3b: Disadvantage of Performance Testing

1. Despite the test being carried out under favourable environmental conditions, the average genetic merit of the group tested generally remained similar before and after the performance test.

## 3.2 Progeny testing

Progeny testing is a process by which a sire‟s genetic merit is measured through the performance of his progeny. When the progeny are evaluated, the genetic merit of the sire is more accurately assessed and which provide the opportunity to use the high-ranking sires in the breeding programmes with confidence. The basis of breeding programmes is the identification of superior individuals and their widespread use within the population. It is a two-stage operation in

which the superior individuals are first identified and are then used as seed stock for future generations. Preselection of candidates for progeny testing can be done by means of individual performance testing which might increase efficiency and reduce costs in breeding programmes

Progeny testing in animal breeding is used to determine the true breeding value of an animal especially males which are used extensively for propagation of best germplasm. The extensive use of artificial insemination in domestic animals has helped in increasing the selection intensity on the male animals. This selection tool is usually used for characters that are sex-limited, expressed after death (meat characteristics) and usually with low heritability, for example milk or egg production in females. A bull for example cannot be assessed for milk production, however the performance of its female offspring‟s can be used to determine the use of the animal for future crosses. A progeny test is performed by mating the male with a number of females to produce many progenies in different environment and over a long time period involving different seasons to nullify the impact of season, management, environment in breeding value estimation. The average performance of the offspring is then found, giving a measure of the male's respective value to the breeder.
In animals the progeny testing could be conducted in a large herd or involving associated herds or in the field in farmers place. The field based progeny testing is highly required when the selected bulls are to be distributed in a large area, to many farmers in different environments.

### 3.2.1: Progeny Testing Procedure

A typical breeding design for a Progeny Testing Programme that can be implemented in smallholder production situations as researched and documented by the National Dairy Development Board is as follows:
A certain number of young sires produced using the very best dams and sires are put under test. Adequate number of test doses of bulls put to test are distributed in selected herds/villages to ensure that at least 80 to 100 complete first lactation record of daughters per bull in as many herds/villages as possible are made available for estimating breeding values of bulls with a very high reliability. The very best 1-10% of progeny tested bulls and the very best 1 to 10% of recorded cows are used for producing the next generation of young bulls. The young bulls are again put to test and the cycle is repeated. The top 10 to 15% of the progeny tested bulls are usedfor production

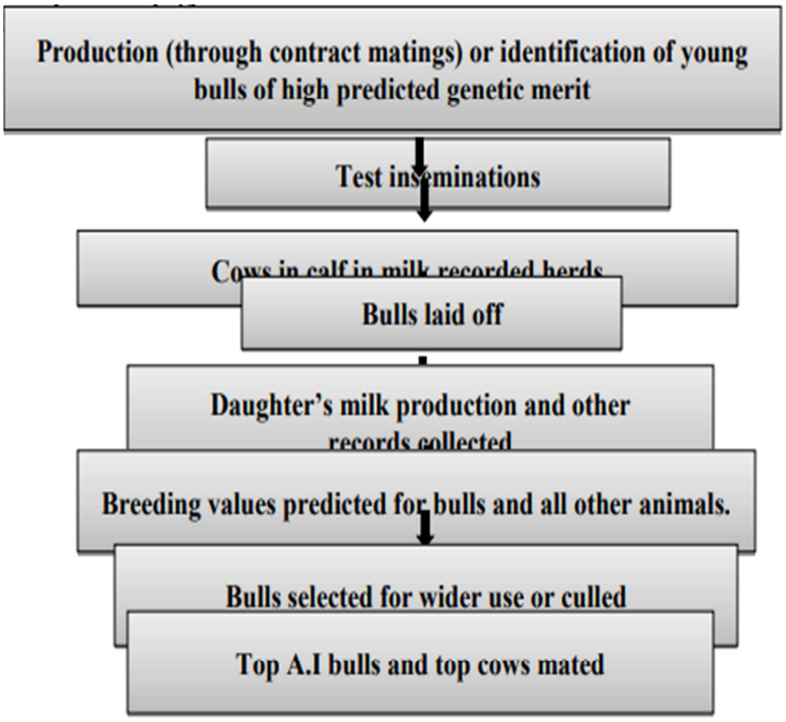



**Figure 1**: Main features of progeny testing programmes for dairy cow bulls (Andrabi and Moran,2007)

### 3.2.3: Factors that affect the usefulness of Progeny Testing

On the standpoint of genetic progress expected from selection using progeny testing, the usefulness of progeny testing can be greatly influenced by some factors and the most important of them are: **Age and Rate of reproduction**. These two factors have been reported to have their effect on the generation interval of the results from the use of progeny test and which can offset the advantage of more accurate selection and reduce the rate of improvement obtained.

**Advantages of Progeny Testing**

1. Genetic merit of sire is more accurately assessed
2. Provide opportunity to use high-ranking sires in the breeding programme with confidence.

**Disadvantages of Progeny Testing**

1. Progeny testing programmes are very long-term.
2. It involves high cost
3. It requires high level of technical and professional skills

4. Strong field artificial insemination and quality semen production infrastructure are needed.
5. Could only be entrusted to institutions having requisite experience and skills and financial resources.

## 4.0 CONCLUSION

Conducting a field based performance and progeny testing especially in small holder production systems of Asia and Africa require huge resources both financial and infrastructural - a large AI network, robust and dynamic data collection and analysis system. Usually the breeding companies conduct progeny testing of their bulls so that they can be commercially promoted. But when the breeding organizations are Government controlled (eg.India), the onus of conducting the testing also lies with them if required genetic improvement is to be achieved

## 5.0 SUMMARY

- Performance and progeny testing are very crucial to any meaningful animal breeding programmes and leads to more accurate selection of superior breeding stock
- Preselection of candidates for progeny testing can be done by means of individual performance testing which might increase efficiency and reduce costs in breeding programmes
- The trait to evaluate depends on the breeding objectives for example, beef cattle breeders may select for traits such as coat colour, shape of horns, etc., while the commercial producer's focus will be on economically important traits such as fertility and growth rate
- The progeny test is needed most for traits which cannot be expressed in one sex and for traits which are but slightly hereditary.
- The bases for estimating breeding value are pedigree, own performance, and progeny test. As fast as some selection is practiced on one of these bases, the possibilities for further progress by additional selection on the same basis rapidly diminish and correspondingly increased attention should be given to one of the other bases.

## 6.0 TUTOR MARKED ASSIGNMENT

1. Differentiate between progeny and performance testing
2. What are the procedures in performance testing?
3. What are the advantages and disadvantages of progeny testing?

4. Itemise the steps involved in progeny testing using specific example

## 7.0 REFERENCES/ FURTHER READING

Ström, H, and Philipsson, J. 1978 Relative importance of performance tests and progeny tests in horse breeding. Livestock Production Science, Volume 5 , Issue 3 , 303 – 312

Lush J. L. (1935). Progeny Test and Individual Performance as Indicators of an Animal's

Breeding Value. Journal of Dairy Science Volume 18, Issue 1, January 1935, Pages 1-19

# MODULE 5 IDENTIFYING AND INCORPORATING GENETIC MARKERS AND MAJOR GENES IN ANIMAL BREEDING PROGRAMMES.

## CONTENTS



## 1.0 INTRODUCTION

Recent developments in molecular biology and statistics have opened the possibility of identifying and using genomic variation and major genes for the genetic improvement of livestock. During the last five decades, the application of methods based on population genetics and statistics allowed the development of animals with a high productive efficiency.

These systems are based on simplified models of genic actionthat assume a large number of or genes with small individual effects in the expression of the phenotype (polygenes) and emphasizes the average genic effects (additive effects) over their interactions. The basis is predicting the breeding values of the animals using phenotypic and genealogical information. Molecular techniques allow detecting variation or polymorphisms exists among individuals in the population for specific regions of the DNA. These polymorphisms can be used to build up genetic maps and to evaluate differences between markers in the expression of particular traits in a family that might indicate a direct effect of these differences in terms of genetic determination on the trait. More probably, the can prove some degree of linkage of the QTL effecting the trait and the marker. Recently, methods have been

developed to detect the presence of major genes from the analysis of pedigreed data in absence of molecular information

## 2.0 OBJECTIVES

- You will understand molecular markers and major genes
- You will understand different molecular markers
- You will understand how to incorporate molecular markers and major genes in animal breeding programmes

## 3.0 MAIN CONTENT

### 3.1 Molecular markers

A molecular marker is a gene or DNA sequence with a known location on a chromosome and associated with a particular gene or trait. It can be described as a variation, which may arise due to mutation or alteration in the genomic loci that can be observed. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), or a long one, like mini & micro satellites. Recent years have witnessed a great interest towards molecular markers, revealing polymorphism at the DNA level, as they play an important role in animal genetics studies. When differences in DNA occur within genes, the differences have the potential to affect the function of the gene and hence the phenotype of the individual. Genetic markers which have been used a lot in the past include blood groups and polymorphic enzymes. We have relatively few such markers, but this has been overcome with the advent of new types of markers.

However, most molecular markers are not associated with a visible phenotype. The main types of molecular markers are VNTRs, RFLPs and RAPDs, AFLPs and SNPs.

**3.1a: Variable number tandem repeat** (VNTR‟s ) are scattered at various locations in the genome and are regions that are highly variable. These regions contain a type of DNA sequence called Variable Number Tandem Repeat which are multiple copies of a sequence of base pairs arranged in head to tail fashion. For example, a frequently found tandem repeat is CA, and one strand containing this type of repeat reads CACACA….. , notated as (CA)n. The other strand would read GTGTGT… In this example, the number of repeating basepairs is two, but it can be more. When the repeating unit is less than four, the VNTR is called a microsatellite and when the repeating unit is longer it is a minisatellite.

**Microsatellites** are DNA regions with variable numbers of short tandem repeats flanked by a unique sequence. Microsatellites make good genetic markers because they each have many different 'alleles' - ie. There can be many different lengths of the repeat region. An allele is defined by the number of repeats there are at the same location. With many alleles, most individuals are heterozygous, giving power to note association between marker allele and performance in progeny inheriting a favorable linked QTL allele. Through the PCR reaction (see below), which uses the unique sequences either side of the repeat sequences as primer binding sites, microsatellite DNA can be specifically amplified. The alleles an individual carrier at a particular microsatellite locus can then be determined by accessing the size of the amplified fragment through agarose gel electrophoresis.

**3.1b: Restriction Fragment Length Polymorphisms' (RFLP's):** here restriction enzymes enzymes cut DNA wherever they find the appropriate nucleotide sequence (eg. Eco R1 cuts at the 'recognition sequence' GAATTC). If there is a mutation at this sequence, no cut is made and the resulting DNA fragment is longer. Also mutation to give a new recognition sequence gives a pair of shorter fragments. Genetic differences (polymorphisms) of this type are known as Restriction Fragment Length Polymorphisms.

**3.1c: Random Amplified Polymorphic DNA (RAPD) markers** are DNA fragments generated in PCR reactions that use a single short primer (in normal PCR a primer-pair is used). The primer must be complementary to sequences that are on opposite strands within a small number of base pairs (say 2000). The DNA strand between these two sites is amplified in a PCR. Polymorphism is determined by individuals who have mutations at those sites, and therefore will not show a product on the gel. The advantage of RAPD‟s is that we do not need to know the DNA sequence of the species studied. A primer has a certain chance of randomly generate a PCR product. Hence, RAPDs are cheap markers to develop. The disadvantage is that RAPDs either give or do not give a product and therefore, we cannot distinguish between homo- and heterozygotes.

**3.1d: Amplified Fragment Length Polymorphism (AFLP)** is based on PCR amplification of selected restriction fragments. Like RAPDs, AFLPs require no prior knowledge of DNA sequences (unlike microsatellites). The advantage of AFLPs over RAPDs is that they are more reliable and reproducible (depend less on DNA quality and lab conditions). Also, the number of polymorphic loci (molecular markers) that can be detected is 10-100 times greater with AFLPs than with microsatellites or RAPDs

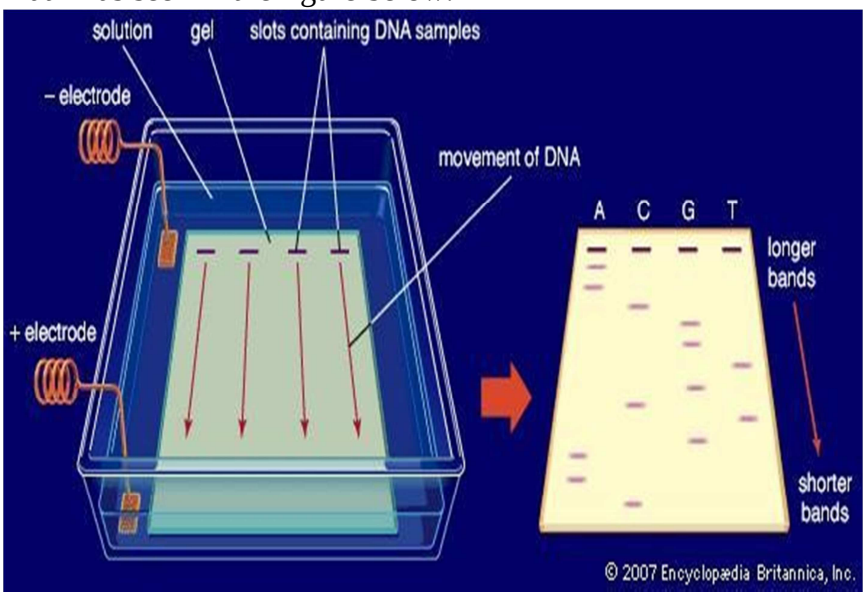
**3.1e: Single Nucleotide Polymorphisms** are based on single base pair polymorphisms. A SNP is a position at which two alternate bases occur at appreciable frequency. In humans they may number greater than one in a thousand base pairs. SNPs can be detected by a number of methods, however a relatively new technology, using DNA chips, can be used for large scale screening of numerous samples in a minimal amount of time.

### 3.1.1 Identifying molecular markers

Molecular techniques (such as polymerase chain reaction (PCR) or restriction enzyme digestion, followed by gel electrophoresis) can be used to identifying different alleles resulting from DNA polymorphisms. Different alleles from a VNTR will have different size and similarly, RFLP‟s have different sizes (as defined by their name!)

**Gel electrophoresis**.

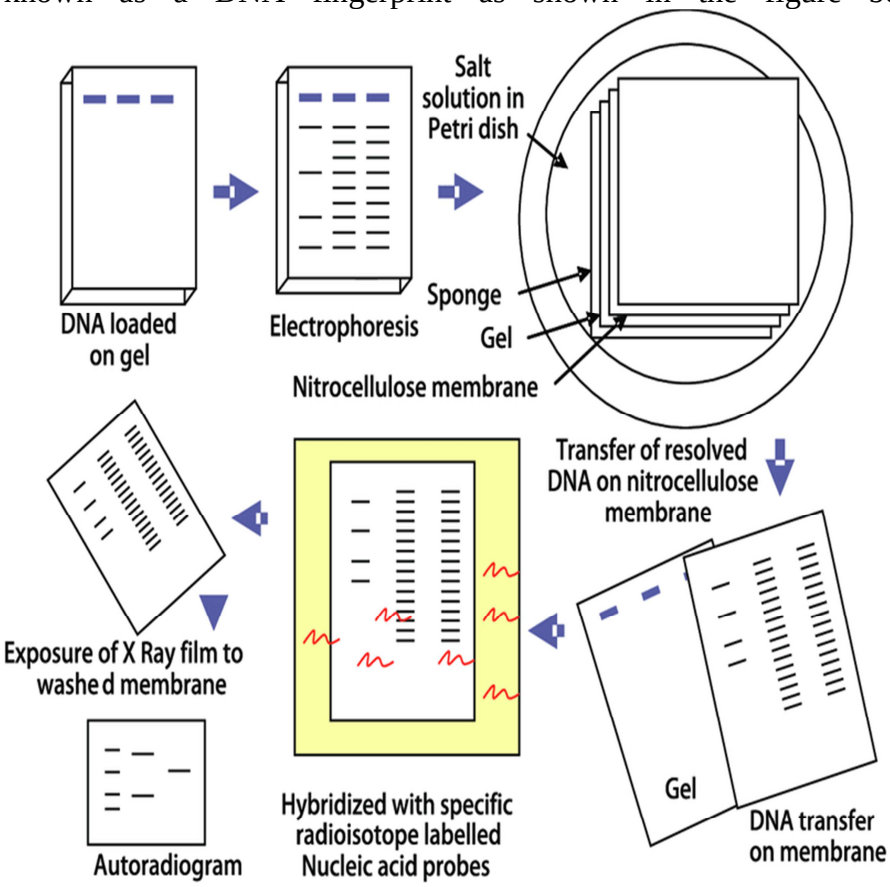
Gel electrophoresis separates DNA according to size. A gel is essentially a slab of gelatinous material. DNA is applied to 'wells' at the top of the gel (which is submerged in a tank containing some buffer), and an electrical current applied. DNA is negatively charged and is drawn towards the positive electrode. Smaller fragments will move down the gel faster, as it is easier for them to move through the gel matrix as seen in the figure below.




### Southern blot

A southern blot involves the transfer of DNA from a gel (where it has been separated according to size) to a special type of membrane. The

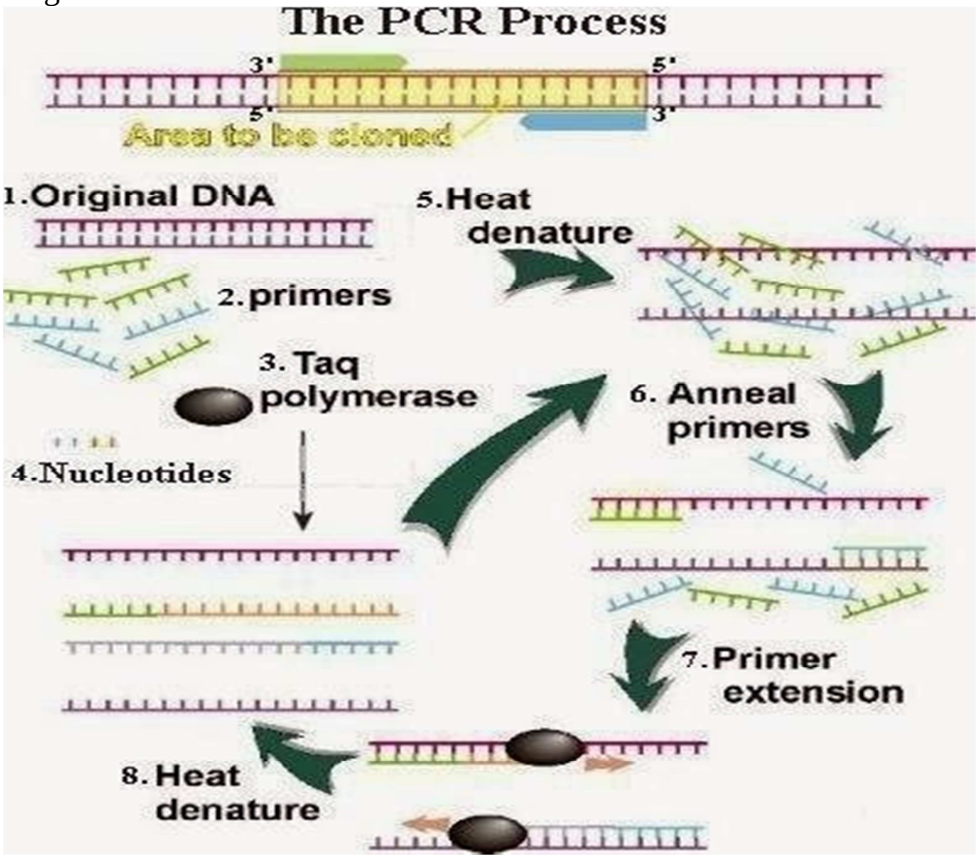
DNA on the membrane (which is in a denatured or single stranded state) is exposed to a probe. A probe is a short sequence of DNA that is complimentary to, and thus binds to, a DNA sequence of interest. Probe bound to the membrane is then visualized: this can be achieved by labeling the probe with radiation and exposing the membrane to X-ray film. A Southern Blot will usually show the alleles of VNTR‟s on all chromosomes, giving a complex pattern known as a DNA fingerprint as shown in the figure below.




**Polymerase chain reaction (PCR)**

The polymerase chain reaction (PCR) amplifies a specific region of DNA as defined by two primer sequences. It can thus be used to examine one particular region of the genome. Because many copies of one specific section of the genetic material are generated, it is possible to use this technique with very, very small amounts of DNA as starting material (e.g. a single hair root or a small blood stain).PCR is a three stage process. Firstly the DNA is denatured (made single stranded), secondly the primers bind or anneal to their complementary sequence, and thirdly the primers are extended by the addition of nucleotides complementary to that on the template sequence (this requires the action of an enzyme called DNA polymerase). This three stage

process is then repeated 20-40 times as depicted in the following diagram.




## 3.2 Major genes

**Major gene** is a gene with pronounced phenotype expression and characterizes common expression of oligogenic series, that is, a small number of genes that determine the same trait. Major genes control the discontinuous or qualitative characters in contrast of minor genes or polygene‟s with individually small effects. Major genes segregate and may be easily subject to Mendelian analysis. The gene categorization into major and minor determinants is more or less arbitrary. Both of the two types are in all probability only end points in a more or less continuous series of gene action and gene interactions. Recent developments in molecular biology and statistics have opened the possibility of identifying and using genomic variation and major genes for the genetic improvement of livestock. The detection of major genes using mixture models with segregation analysis can direct the work of identification of DNA marker genotypes towards populations and characteristics with greater probability of detecting a QTL.

The present trend indicates that molecular, pedigree and phenotypic information will be integrated in the future through mixture models of segregation analysis that might contain QTL effects through the markers, polygenic inheritance and uses powerful and flexible methods of estimation such as Gibbs Sampling.

Recently, methods have been developed to detect the presence of major genes from the analysis of pedigreed data in absence of molecular information. These methods, based on mixture models and segregation analysis, allow to direct the potentially expensive and time consuming genotyping activities towards populations and characteristics with a greater probability of being controlled by a QTL and to optimize the collection of molecular data

### 3.2.1: Detection and use of major genes

In the last ten years statistical methodologies of detection of major genes based on pedigree and phenotypic information on populations have been developed for animal populations. These methods are based on the use of mixed models and segregation analysis to fit the data to a mixture genetic model that includes in addition to the polygenic effects, those of a biallelic major gene. Calculation is performed in two stages; firstly genotype probabilities are obtained, then major gene, fixed effects and polygenic effects are fitted and used to recalculate new parameters by regressing phenotypes on estimated probabilities. Calculation is iterated upon convergence. Segregation analysis allows inferring the unknown genotypes from the probabilities of transmission of the gene given the phenotype of the individual and their relatives.In mixture models, regression and Gibbs sampling estimation approaches have been implemented to obtain estimates of the major gene effects and allelic frequency. Meuwissen and Goddard (1997) evaluated the effect of including different proportions of individuals genotyped for a QTL in a mixture model that is based on the analysis of segregation of Kerr and Kinghorn (1996) and a regression approach which uses the estimated genotype probabilities as weights in the estimation process. Unbiased estimates of QTL effect and frequency were obtained in absence of information on the genotype of the QTL, but some improvements in the precision of the estimates were observed as the proportion of genotyped individuals increased. The main limitation of this method is that the genetic hypothesis is generally limited (one biallelic locus), thus, the presence of more alleles could not be detected. Also, the location of the locus in the genome, in absence of markers, remains unknown. Mixture models can be modified to include markers associated with the QTL, instead of the direct effect of the QTL in addition to the information of the pedigree and the phenotype. This is achieved by modifying the additive numerator relationship matrix (A),

according to the conditional probabilities of transmission of the given QTL the information of the markers. These developments can make possible to evaluate the likelihood of the model or another fitting criterion, to prove the relation between the markers and the QTL in population animals with outbreed mating structures. They also may increase the possibilities of making MAS in animal populations when incomplete information exists on the genotypes of the animals for the QTL or markers so that the use of the genomic information is optimized. Scientist has developed a method to evaluate the amount of genomic information that it allows maximizing a function of economic utility for the analysis of QTL with mixture models. Major genes have been detected using these methods for carcass characteristics in pigs based on a mixture model of inheritance and Gibbs Sampling. Also, important effects of major genes have been detected using Findgene software for several carcass characteristics in cattle and for parasite resistance in sheep. This methods that make use of information currently available in many animal populations, are an option for a preliminary screening for major genes that can contribute to rationalize the use of expensive QTL-marker linkage estimation experiments.

### 3.2.2 Incorporating genetic markers and major genes in animal breeding programmes. Marker assisted selection (MAS)

The addition of genomic information to phenotypic information to increase the selection response to the traditional method is known as Marker-Assisted Selection (MAS). The concept of Marker Assisted Selection (MAS) utilizing the information of polymorphic loci as an aid to selection was introduced as early as in 1900. The method where marker genes used to indicate the presence of desirable genes is called as marker assisted selection. Marker assisted selection (MAS) is indirect selection process where a trait of interest is selected not based on the trait itself but on a marker linked to it. The purpose is to combine all genetic information at markers and QTL with the phenotypic information to improve genetic evaluation and selection. The advantage of using MAS is that the effect of genes on production is directly measured on the genetic makeup of the animal and not estimated from the phenotype. The integration of two selection methods, i.e., traditional or conventional selection methods with molecular genetics methods beneficial to the selection response. Multiple estimated QTL effects and multiple trait selection could help to make better decisions regarding the use of MAS in animal improvement. Combined with traditional selection techniques, MAS has become a valuable tool in selecting organisms for desirable traits. MAS is expected to increase genetic gain compared to traditional breeding programs and reduce the cost of progeny testing by early selection

of the potential young bulls. The application of MAS in breeding programmes depends on the knowledge of breeders about variable marker information from animal to animal and the different effects on multiple traits and his ability to spend in genotypic information that helps in improve their commercial breeding activities. MAS also provide an apparently possible approach to selection for genetic disease resistance animals. In the future to make MAS effective in large breeding populations, the availability of large-scale genotyping methods and infrastructure that allows the generation of hundreds of thousands of molecular data at a reasonable cost will be necessary.

**3.2.2a: Marker assisted introgression:** An application that has been mentioned in the literature is the introgression a major gene in another population by means of backcrosses assisted by molecular markers. In this case, it does not seem to exist advantage in using single genetic marker information, in comparison with the use of only phenotypic information when the characteristic is continuous and the considered genetic effects are additives. Nevertheless, it seems feasible that using a dense map that involves many chromosomal regions and with more than one allele of interest, the time for fixation of the major genes can be reduced.

An example of introgression in pigs breeding is the introduction of litter size genes from the Meishan breed into Western pig breeds. The possible gains from such strategies depend heavily on the gene effect and the frequency in the commercial lines. Introgression is expensive, as it involves several generations of backcrossing to the desired genotype, while keeping a desired haplotype from the introgressed QTL. At the same time markers can be used to select against haplotypes for background genes from the imported line. This generally speeds up the introgression process and reduces the number of generations needed to arrive at the desired genotype (possibly in two generations).

Marker assisted selection can also be used in crosses of lines of about equal economic value. In that case, population wide linkage disequilibrium can be exploited, giving potentially large increases in response (Lande and Thompson, 1990). Genetic evaluation models can have a significant effect on the achieved genetic response, models with random marker (haplotype) effects being superior, because the approach takes better account of the uncertainty of certain haplotype effects.

## 4.0 CONCLUSION

A rational use of the molecular methodologies requires thesimultaneous optimization of selection on all the genes affecting important traits in the population. The maximum benefit can be obtained when these

techniques are used in conjunction with reproductive technologies like the artificial insemination, and collection and production in vitro ofembryos to accelerate the genetic change.There is a danger associated with a potentially inadequate use of QTL information, giving an excessively high emphasis to simple molecular information in detriment of the overall economic gain through all traits and their polygenic effects in the population. Dissemination of the information to the industry is therefore a complexissue concerning QTL effects and molecular markers

## 5.0 SUMMARY

- The characteristics on which the application of the MAS canbe effective are those that are expressed late in the life of theanimal, or those that are controlled by a few pairs of alleles
- Because of its high cost, the use of MAS could be justified, in animal nuclei that allow dilution of the costs when germ plasm is extensively used towards the commercial population. Also in those characteristics in which the procedures of conventional selection have reached their limits in efficiency or the results have been not satisfactory
- Before the molecular information on the QTL which control the characteristics of economic interest is generated, the detection of major genes using segregation analysis could direct the work of identification of genotypes towards populations and characteristics with greater probability of detecting a QTL using molecular markers

## 6.0 TUTOR MARKED ASSIGNMENT

1. What are molecular markers?
2. Describe some molecular markers you know
3. What are major genes?
4. How are these markers incorporated in animal breeding programmes?

## 7.0 REFERENCES/ FURTHER READING

Kinghorn, B. & van der Werf, J. (2000). IDENTIFYING AND INCORPORATING GENETIC MARKERS AND MAJOR GENES IN ANIMAL BREEDING PROGRAMS. University of New England Armidale, Australia. Chapter 12. Belo Horizonte (Brazil) 31 May –5 June 2000

Montaldo, H. H. and Meza-Herrera, C. A. (1998). Use of molecular markers and major genes in the genetic improvement oflivestock. Journal of Biotechnology ISSN: 0717-3458. Vol.1 No.2, Issue of August 15, 1998.

# MODULE 6 DNA TESTS

## CONTENTS



## 1.0 INTRODUCTION

Almost every cell in every living organism contains a cell-nucleus that holds the pairs of chromosomes that make up the genetic material. Each chromosome has within it the DNA (deoxyribonucleic acid) that makes up hundreds of thousands of genes. Many of these genes make proteins that have a number of roles in the body. Some proteins are structural and make up tissues like bones and muscles. Proteins called enzymes are involved in chemical reactions like breaking down the ingested food. While others are like little messengers that send signals around the system, these proteins are known as hormones.All individuals within a species share the same set of genes but the precise DNA sequence of these genes differs slightly between individuals (by about 0.1-0.2%). While these differences account for things like differing hair, eye and skin colour, they can also be the cause of genetic disease or disease susceptibility. A disease causing change in the DNA of a gene is called a mutation. DNA or genetic tests are such procedures used to uncover mutation or alterations in genes that could lead to genetic disorders.

Many disorders in animals are observed more frequently in certain breeds and within breeds more often in the same families. Familiarity is assumed for a disorder when families are observed with more than one affected family member. Familial disorders may have a genetic contribution. The same is often claimed for disorders which show a breed disposition. On the other hand, genetically caused diseases may not necessarily lead to breed differences in incidence but will contribute to variation among families within breeds. A useful starting point for answering the question whether a disorder is inherited

is by drawing pedigrees to provide an initial impression of the distribution of affected and non-affected animals and how frequently the disorder is transmitted from one generation to the next.

## 2.0 OBJECTIVES

- You will understand DNA tests
- You will understand different types of DNA tests
- You will understand segregation analysis

## 3.0 MAIN CONTENT

### 3.1 DNA tests

These are diagnostic tests that are used to confirm a diagnosis in a patient or animals with symptoms suggesting a particular genetic disease. For example, in human, a person with a movement disorder may be tested for Huntington‟s disease. The genetic information from the test is useful in the treatment, management and genetic counseling of the patient.Some genetic tests are used even when symptoms of a disease are not seen, but the genetic information may help in predicting if the person is at risk of developing, or are susceptible to a particular disease.Genetic screening tests are;

1. **Prenatal testing:** This type of testing is offered during pregnancy if there is an increased risk that the baby or progeny will have a genetic or chromosomal disorder. In some cases, prenatal testing can lessen a couple's uncertainty or help them make decisions about a pregnancy. It cannot identify all possible inherited disorders and birth defects.A good example of this is the screening in human is for Down syndrome in women over 35. Screening for Down syndrome is usually carried out by amniocentesis or chorionic villus sampling at 14 – 20 weeks of gestation.Prenatal testing is used to detect changes in a fetus's genes or chromosomes before birth.

1. **Newborn screening** is used just after birth to identify genetic disorders that can be treated early in life. In humans, millions of babies are tested each year in the United States and Europe especially for phenylketonuria (a genetic disorder that causes intellectual disability if left untreated) and congenital hypothyroidism (a disorder of the thyroid gland) and cystic fibrosis (CF). A blood sample is taken from the newborn; this blood sample is then sent to a laboratory for testing.

3. **Carrier screening** is offered to parents-to-be so that they can test if they are carriers for diseases such as cystic fibrosis (CF).Carrier testing is used to identify people who carry one copy of a gene mutation that, when present in two copies, causes a genetic disorder. This type of testing is offered to individuals who have a family history of a genetic disorder and to people in certain ethnic groups with an increased risk of specific genetic conditions. If both parents are tested, the test can provide information about a couple's risk of having a child with a genetic condition.
4. **Diagnostic testing** is used to identify or rule out a specific genetic or chromosomal condition. In many cases, genetic testing is used to confirm a diagnosis when a particular condition is suspected based on physical signs and symptoms. Diagnostic testing can be performed before birth or at any time during a person's life, but is not available for all genes or all genetic conditions. The results of a diagnostic test can influence a person's choices about health care and the management of the disorder.

5. **Preimplantation testing**, also called preimplantation genetic diagnosis (PGD), is a specialized technique that can reduce the risk of having a progeny with a particular genetic or chromosomal disorder. It is used to detect genetic changes in embryos that were created using assisted reproductive techniques such as in-vitro fertilization. In-vitro fertilization involves removing egg cells from a woman‟s ovaries and fertilizing them with sperm cells outside the body. To perform preimplantation testing, a small number of cells are taken from these embryos and tested for certain genetic changes. Only embryos without these changes are implanted in the uterus to initiate a pregnancy.

1. **Predictive and presymptomatic** types of testing are used to detect gene mutations associated with disorders that appear after birth, often later in life. These tests can be helpful to people who have a family member with a genetic disorder, but who have no features of the disorder themselves at the time of testing for example in dairy industry to guard against occurrence of mastitis. Predictive testing can identify mutations that increase a person's risk of developing disorders with a genetic basis, such as certain types of cancer. Presymptomatic testing can determine whether a person will develop a genetic disorder, such as hereditary hemochromatosis (an iron overload disorder), before any signs or symptoms appear. The results of predictive and presymptomatic testing can provide information about a

person‟s risk of developing a specific disorder and help with making decisions about medical care.

7. **Forensic testing** uses DNA sequences to identify an individual for legal purposes. Unlike the tests described above, forensic testing is not used to detect gene mutations associated with disease. This type of testing can identify crime or catastrophe victims, rule out or implicate a crime suspect, or establish biological relationships between people (for example, paternity).

### 3.1.1 Methods and techniques used in genetic testing

Before a genetic test is carried out, clinical examination is carried out and a detailed family history gotten. This will help in working out which gene may be responsible for the disease in question. In human, the patient will be referred to a genetic counsellor who can inform them about everything that is involved with genetic testing. The genetic counsellor can tell you what it means to have a particular genetic change and how this can affect individual or the family. In the case of animals, these will form a basis of whether to cull the animal or to be applying symptomatic treatment. Almost all genetic tests require a DNA sample from the patient; this is usually obtained by either a blood sample or mouthwash (buccal swab) which is then taken to a genetic testing lab for analysis.A number of techniques are used in the process of genetic testing, these include:

**1. Polymerase chain reaction (PCR) and DNA sequencing**

The polymerase chain reaction (PCR) is a method of amplifying (copying) a small amount of DNA to a larger amount so that it can be analyzed closely. The genetic code of the DNA can be determined by a method called „DNA sequencing‟. This then allows scientists to determine whether or not there is a change or mutation present in a gene of interest.

**2. Indirect gene tracking (linkage analysis)**

If the gene associated with a hereditary disease in a family is not known then linkage analysis can help in identifying the responsible gene. The technique is based on the fact that special DNA sequences that flank particular genes will travel with the gene when passed from parent to child. These DNA sequences are called „polymorphic markers‟ or „polymorphic repeat sequences‟. The closer that one of these markers is to gene the more likely it is that it is travelling with the gene. If a particular polymorphic marker is found

only in members of a family with a particular disease then it is likely that a gene located near the marker is associated with the disease.

**Advantages of genetic screening**

Through genetic testing we may be able to screen populations for diseases in order to better diagnose, treat and prevent disease. Reducing the incidence of disease has major impacts on and is of great importance to:

Families: When a child is born with a particular disease, there may be no apparent family history. With simple genetic testing for carrier status of the parents, the birth of a child with disease could have been prevented.

Health resources: The birth of a child with a genetic disorder adds stress to health systems and resources. Carrier screening programs could act as important components of the medical system in preventing disease through offering people informed reproductive choices.

**Disadvantages of genetic testing**

Many people would rather not know if they have a pre-disposition to a particular disease. People may have enough stress in their lives already to have to deal with an oncoming genetic disease, or that they are a carrier of a particular disease mutation that may affect their future children. There is also the possibility of some sort of negative stigma attached to having a carrier status for a particular disease.

If the genetic cause of a disease is identified in a patient, it does not necessarily guarantee that there is a cure for the disease. A treatment or therapy may not yet have been developed. The identification of a disease gene means a large step towards finding a cure; it is often of no great immediate benefit to the patient.

**Ethical issues associated with genetic testing**

As carrier screening would involve the screening of a possibly asymptomatic population, it inevitably raises a number of ethical issues in terms of consent, privacy and education, which need to be considered. It is important that all persons involved are appropriately educated, their consent obtained and the confidentiality of their genetic information upheld.

Prenatal genetic screening can inform parents of the health status of their unborn child. In the case of a prenatal diagnosis of disease, parents are able to assess their options and make decisions

accordingly. Foetal genetic testing does however raise the issue of abortion, often a particularly sensitive and controversial matter.

## 3.2 Segregation Analysis

General evidence for genetic contribution to a disorder is given when environmental factors can be excluded as the only responsible causes for a disorder and a significant proportion of the phenotypic variation of a disorder can be explained by genetic models. With increasing molecular genetic data, the type of gene action based on known DNA sequence variation can be characterized by individual genes and the nature of complex genetic traits can be understood much better.Segregation analysis is employed to determine whether familial data for particular disorders or other traits are compatible with specific modes of inheritance. Modes of inheritance tested in segregation analyses include **monogenic (Mendelian), digenic or polygenic models**. In addition, age of onset, sex effects and sampling scheme can be taken into account besides the specific genetic hypothesis under consideration. Simple segregation analysis tests the segregation parameter $\theta$ under a specified sampling scheme and mating type. Pedigrees used for segregation analysis may be from specifically planned mating or randomly sampled pedigrees with arbitrary structure or sampled through ascertained cases in clinics or veterinary practice. Arranged mating among animals can be more easily tested for specific modes of inheritance than pedigrees with arbitrary structure, missing data and many inbred animals. In the case of a rare disease and an autosomal dominant hypothesis, the segregation ratio $\theta$ is assumed to be 0.5 as families segregating for the trait are most likely composed by mating of heterozygous affected and homozygous non-carriers. As far as the segregation ratio is not significantly different from $\theta = 0.5$, this mode of inheritance is accepted. Different methods for estimating $\theta$ have been developed and are easily applied (Singles Method, Weinberg's General Proband Method). These simple approaches to segregation analyses often encounter problems when different mating types have to be considered and several hypotheses are more or less likely. Complex segregation analyses have been developed to allow for more factors to vary and to reduce the restrictions on assumptions to be made for the model tested. Methods used to solve the likelihood functions are based on maximum likelihood or Markov chain Monte Carlo approaches (Gibbs sampling).

### COMPLEX SEGREGATION ANALYSIS

Complex segregation analysis is based on a mathematical model that incorporates several, functionally independent components to accommodate for arbitrary mating types, different modes of

monogenic or oligogenic inheritance (major genes), to allow for polygenic variation and non-genetic variation in addition to major genes and different data types such as binary, categorical and continuous data. In addition, age of onset of a disease and sampling scheme (random pedigrees versus non-randomly selected pedigrees) can be modeled. The basic model as formulated in the Elston-Stewart algorithm was the basis for the more complex models. The Elston-Stewart algorithm included a component describing the joint distribution of genotypes of mating individuals whereby these genotypic distributions stem from a single locus with two alleles (monogenotype), a few loci with each two alleles (oligogenotype) or from a polygenotypic distribution with an infinite number of genotypes (polygenotype). The second component of the Elston-Stewart algorithm specified the relationship between the genotypes and phenotypes, separately for each genotype (penetrance function). Mathematically, the phenotype investigated is modeled as a conditional probability on the genotype underlying the model used. The simplest genetic model for a dichotomous trait and a monogenic autosomal inheritance of two alleles is then completely defined by the following genotype to phenotype relationships: $g_{AA}(1) = g_{Aa}(1) = 1$, $g_{aa}(1) = 0$ and $g_{AA}(0) = g_{Aa}(0) = 0$, $g_{aa}(0) = 1$, where the conditional probability equals unity when for the genotypes AA and Aa the phenotypic outcome is affected (=1) and for the genotype aa the phenotypic status is unaffected (=0). Similarly, if a completely penetrant recessive trait is assumed, we have the following conditional distributions: $g_{aa}(1) = 1$, $g_{Aa}(1) = g_{AA}(1) = 0$, $g_{aa}(0) = 0$, $g_{Aa}(0) = g_{AA}(0) = 1$. Two- or three-locus models give raise to much more models (phenogrammes) how the oligogenotype is related with the phenotype. If we do not wish to assume complete penetrance we can introduce for each distinct genotype or groups of genotypes a specific penetrance. For X-linked loci, the conditional distributions of phenotypes have to be defined for males and females separately. Furthermore, traits only expressed in males or females can be modeled via the penetrance parameter allowing fully expressed traits only for one sex. Just as the phenotypic distribution may be sex-dependent, so the disorder considered has a variable age of onset and thus the observation whether the disorder is expressed, depends upon the age at examination of each individual. Then the probability that an individual with a genotype AA, Aa or aa is affected by a specific age depends of the age- related susceptibility of the genotype to the disorder. When we turn to polygenotypes, we use normal distribution functions. In the case of a binary or categorical phenotype, this model corresponds to the threshold or liability model. The polygenotypes are normally distributed with genetic variance $\sigma^2_G$ and residual variance $\sigma^2_E$. An individual is affected or mildly/severely affected whose liability is

greater than the threshold. The threshold may also depend upon the genotype of an additional monogenic locus.

The mode of inheritance can be described how the genetic variability is passed on from one generation to the next and is summarized mathematically by the genotypic distributions of the offspring in dependence upon the parental genotypes. Let us assume that an individual has parents with genotypes s and t, then the conditional probabilities for the genotypes of this individual can be viewed as elements of a stochastic matrix called the genetic transition matrix, probability (P) for the individual genotype given genotypes of parents s and t, $P(g_i|g_F,g_M)$. All types of monogenic and oligogenic inheritance can be parameterized in terms of transmission probabilities. In the autosomal monogenic model with alleles A and B, the transmission probabilities are the probabilities that an individual with genotype AA, AB or BB transmits the allele A to offspring. Using the definitions for the transmission probabilities $\tau_{AA}=1$, $\tau_{AB}=0.5$ and $\tau_{BB}=0$, the probabilities for the genotype AA of the individual with parents s and t are equal to $\tau_s\tau t$, the probabilities for the genotype AB with parents s and t are equal to $\tau_s(1-\tau_t) + \tau_t(1-\tau_s)$ and the probabilities for the genotype BB with parents s and t are equal to $(1-\tau_s)(1-\tau_t)$. Extension to several unlinked loci and linked loci is straightforward. Linked loci require recombination rates among loci as further parameters. Polygenic inheritance using an additive model can be modeled through the transmission of the gamete values being 0.5 for any polygenotype. The polygenotypes of offspring are produced by the mid-parents´ values of their polygenotypic effects with variance $\sigma^2_G/2$.

Sampling scheme describes the way how individuals were selected from the population for study. Random sampling means that we take a random sample of individuals from a population and then augment this sample by including all or a random sample of relatives up to a certain degree of relatedness. When well-designed recording schemes are introduced, random samples of progeny or sibships with their ancestors can be collected. These samples can be collected in a specific geographic area which is not critical as long as individuals outside this area are not selected according to their phenotype or genotype. Rare conditions are hardly studied in random samples hence many uninformative families are collected. Typically for this situation, families are included in the study because at least one member of the family is affected. The kind of the non-random sampling procedure is characterized by the type of ascertainment. Complete ascertainment is given when a sibship enters the sample independently of the number of additional affected members. The opposite extreme to complete ascertainment is single ascertainment. The probability for an

affected individual tends to be zero to be brought into the study when there is not more than one affected family member. Incomplete multiple ascertainment is the situation between single and complete ascertainment. To ensure a valid segregation analysis, the kind of ascertainment should be identified. Methods of estimation of the segregation ratio depend on how the families have been brought into the study. A likelihood function based on the components of the segregation analysis model can be derived and maximized for the data observed. Since the likelihood function includes the different types of genetic models as well non-genetic factors, submodels can be tested against the most general model. Inferences can be performed for both continuously and categorically distributed data and genetic models that include monogenic, digenic, polygenic and mixtures of monogenic and polygenic as well as oligogenic and polygenic models. A genetic background of a trait analyses is given when the model explaining only non-genetic factors can be rejected and models including genetic components explain a significant proportion of the phenotypic variation.

A likelihood ratio test statistic is used to compare a specific null hypothesis ($H_0$) defined by a specific model (restricted model) against a most general (not restricted) model. The test statistic asymptotically follows a $\chi 2$-distribution, and significance levels can be obtained by using this distribution. Degrees of freedom are given by the difference of independently estimated parameters for the models compared. The information criterion of Akaike (AIC) can be used as an additional measure to choose the sparsest model with the best fit to the data. The model with the smallest AIC fits the data best with a minimum number of parameters but all hypotheses that cannot be rejected against the most general model using the likelihood ratio test must also be considered as possible. The AIC criterion cannot be used to exclude a hypothesis if this model was not rejected against the most general model by using the likelihood ratio test.

Complex segregation analysis is a powerful tool to detect major gene variation. Quantitative genetic models rely on the assumption of many (infinite) loci with very small and equal effects. This model is severely compromised in the presence of segregation of major genes. Extensions and improvements of algorithms made to the simple segregation models allow to estimate major genotype effects in the framework of the methodology developed for quantitative genetic analysis. Gibbs sampling can be employed to estimate non-genetic effects, genotype frequencies and their associated genotypic effects and quantitative genetic variation including all relationships of the animals. When information for genetic markers in population-wide linkage disequilibrium or mutations of genes associated with trait

variation can be included in the analysis, the genotypic distributions need no longer to be estimated and inferences on the genotypic effects are much more precise. Such genetic polymorphisms enable us to model the gene actions and their interactions in networks for complex genetic traits.

## 4.0 CONCLUSION

Complex segregation analysis is a powerful tool to detect major gene variation. Quantitative genetic models rely on the assumption of many (infinite) loci with very small and equal effects. This model is severely compromised in the presence of segregation of major genes. Extensions and improvements of algorithms made to the simple segregation models allow to estimate major genotype effects in the framework of the methodology developed for quantitative genetic analysis. Gibbs sampling can be employed to estimate non-genetic effects, genotype frequencies and their associated genotypic effects and quantitative genetic variation including all relationships of the animals.

## 5.0 SUMMARY

When information for genetic markers in population-wide linkage disequilibrium or mutations of genes associated with trait variation can be included in the analysis, the genotypic distributions need no longer to be estimated and inferences on the genotypic effects are much more precise. Such genetic polymorphisms enable us to model the gene actions and their interactions in networks for complex genetic traits

## 6.0 TUTOR MARKED ASSIGNMENT

1. What is DNA test?
2. Describe four methods of DNA test
3. What are the advantages of DNA test?
4. What is segregation analysis? Describe any method that you know

## 7.0 REFERENCES/ FURTHER READING

Kinghorn, B. DNA tests and segregation analysis for genetic disorders. Twynam Chair of Animal Breeding Technologies University of New England

# MODULE 7 DETERMINING ASSOCIATIONS BETWEEN GENETIC MARKERS AND QUANTITATIVE TRAIT LOCUS (QTL)

## CONTENTS



## 1.0 INTRODUCTION

The common animal species have a narrow genetic pool due to domestication. In contrast, theirs wild relatives as a result of genetic history and selection pressure are becoming in reservoirs of natural genetic variation. Genes associated with desired productive traits such as higher yield or disease resistance that could be lost in the breeding process can be restored using these wild species. The problem for breeder is to find the genes and find an efficient way to trace the genes and to incorporate them in breeding populations. A survey of genetic relationship using molecular markers provides polymorphism information about a germplasm pool that is useful for developing mapping and breeding populations. If quantitative traits have also been evaluated for the same accessions, then statistical associations can be sought between markers and quantitative traits. Such associations can be used to select a subset of candidate probes with enhanced potential for use in subsequent mapping experiments.

## 2.0 OBJECTIVES

- You will understand quantitative traits and quantitative trait locus
- You will understand methods of determining associations between genetic markers and quantitative trait locus

## 3.0 MAIN CONTENT

## 3.1 Quantitative trait

A quantitative trait is one that has measurable phenotypic variation within a population owing to underlying variability in genetic and/or environmental influences. A QTL is a genetic locus in which allelic variation affects variation in the observed phenotype. Generally, quantitative traits are multifactorial, meaning they are influenced by several polymorphic genes and environmental conditions. To map a QTL, its influence on a trait must be detected amid considerable "noise" from other QTLs and non-genetic sources of individual variation. This has been made feasible through the implementation of technologies to identify genetic polymorphisms throughout the genome and the development of statistical methods to map QTLs from specific genetic marker and phenotypic (i.e., trait) data. The identification of the chromosomal regions where marker allelic and phenotype variation co vary implicates the presence of a QTL. Each QTL identifies the genomic location of a gene or genes (referred to as quantitative trait genes or QTGs) affecting the trait of interest. The power of this approach was demonstrated first in plants and later in rodents, and has been used widely to identify genetic contributions to a wide variety of complex phenotypes. The observed distributions of quantitative traits can arise because the traits are influenced by many genes, which result in many possible genotypes, and also by environments. Thus the difference between the means of genotypes are unobservable because of the variability among the environments in which individuals with any particular genotype live.

**Quantitative Trait Locus Mapping**

A quantitative trait is a measurable phenotype emerging from genetic and environmental factors that is distributed in magnitude in a population rather than all or none. A quantitative trait locus (QTL) is a specific chromosomal region or genetic locus in which particular sequences of bases in DNA markers are statistically associated with variation in the trait. Several polymorphic genes and environmental conditions often influence these quantitative traits and one or many QTL(s) can influence a phenotypic trait. Inbred strains, selected lines, and other genetically specified populations have been used in studies analogous to the human population association and linkage studies described above.

The goals are first to locate a QTL harboring a gene or genes affecting the trait to be mapped, and then refine that genomic map until a single gene or genes can be implicated in the effect on the trait.

Currently, QTL fine mapping usually involves the development of congenic strains. In a congenic strain, a very small sequence of DNA on a chromosome is moved from one genetic background to another inbred strain background. An excellent discussion of QTL mapping methods discusses, in depth, the trait of alcohol withdrawal severity. Of course, each QTL generally accounts for only a small proportion of the variability in a complex behavioral trait like addiction, so this is a difficult task and cautious interpretation is warranted. The probability of success in QTL mapping depends on:

1) the heritability of the trait;
2) whether the underlying quantitative trait gene (QTG) is dominant, recessive or additive;
3) the number of genes that affect the trait;
4) whether or not their effects are interactive; and
5) most importantly, the number of subjects that can be tested (i.e., the statistical power of the mapping effort).

Many addiction-related traits have been targeted for QTL mapping studies, although very few of these QTLs have been reduced to QTGs or quantitative trait nucleotides (QTNs). The recent discovery of the addiction-relevant QTG, Mpdz, which possesses pleiotropic effects on the predisposition to severe alcohol and barbiturate withdrawal, demonstrates the power of this approach. Further studies have shown that variation in the human MPDZ gene is related to alcohol drinking. Unfortunately, QTL studies have yet to resolve to a QTG for drinking, in part, due to problems discussed above. However, three candidate genes, neuropeptide Y, α-synuclein, and CRFR2 have been associated with ethanol-seeking. Encouraging evidence shows some consistencies for alcohol and other substance-dependence phenotypes in humans and mice. The more long-term goals of QTL mapping projects are then to move to human populations for studies of the homologous or orthologous gene, and use information about the biological effects of the gene‟s product to help design therapeutic agents or other therapies.

When breeders work with a particular trait in a species, they start to work with the genetics of the trait. Many agricultural characteristics are controlled by polygene‟s and are greatly dependent of genetic x environment interactions. In an aim to work with the patterns of segregation and inheritance for breeding those traits, we think about the positions of the traits in a genetic map. Currently, when the position of a gene controlling traits is inferred we work with tools of genetic or physical mapping, depending on the information available for the species and the trait.

Traditionally it has been a challenge for breeders to work with quantitative trait loci (QTL), with the development of molecular markers technology, it has been possible to follow QTL segregation detecting markers linked to traits of interest and assessing effects, number and location of QTL in chromosomes. An alternative to QTL mapping is association mapping also called association genetics, association studies and linkage disequilibrium mapping. These two methodologies have been advocated as the method of choice for identifying loci involved in the inheritance of complex traits.Association mapping seeks to identify specific functional variants (loci, alleles) linked to phenotypic differences in a trait to facilitate detection of trait causing DNA sequence polymorphisms and selection of genotypes that closely resemble the phenotype (Oraguzie *et al.*, 2007). In order to identify these functional variants it requires high throughput markers like single nucleotide polymorphisms (SNPs).

Molecular markers are used not just to generate genetic maps but also to locate the places of interest in those maps with its incidence in the expression of the trait. That is because they are used in marker assisted selection programs. To improve the breeding methods efficiency, breeders are using markers assisted selection techniques that show great advantages compared with traditional selection methods based on phenotypic traits evaluation. Molecular techniques allow accurate selection in early stages focusing directly in its genetic base.

In order to locate QTL in a genetic map relatively few techniques have been developed, one of those is **linkage mapping**. Linkage mapping is the traditional method for QTL mapping, it implies to generate simple crosses derived populations and to estimate marker-gene recombination frequencies. Population mapping is frequently developed from diploid parental that are originated partially or completely from wild species. Such populations show only a small proportion of all the possible alleles. In contrast, another method is **association mapping** based on linkage disequilibrium (LD) concept; it is a method that exploits the diversity observed in existent cultivars and in breeding lines, without developing new populations.

Most of the important limitations for linkage mapping can be overcome using association genetics. Association genetics does not require building segregating populations and it can employ larger germplasm exploiting the natural variation that exists in the available germplasm and resolution for association could be of at least of 5 cM depending on LD decay of the species.

## 3.2 Principles of genetic mapping population

Genetic mapping is mainly employed with two aims: to identify genetic factors or loci that influence phenotypic traits and to determine recombination distance among loci. As a condition for mapping the traits to be studied must be polymorphic. One way for detecting those polymorphisms is using molecular markers.Genetic mapping by linkage is supported in genetic recombination, as condition for mapping a particular trait. This trait should be polymorphic, displaying preferably a wide variation among the individuals under study. When applying molecular markers in staid of a phenotypic trait these markers should be polymorphic as well, showing allelic variation. The selection of polymorphic markers required for QTL and single trait mapping depends on the existing knowledge regarding the species to study. In species without detailed information of its sequence the candidate gene approach may be used. This approach is based on the production of markers from gene sequences that they have been observed to take place or they are suspected that have a functional role in the selected trait.

QTL mapping begins with the gathering of genotypic and phenotypic data from a segregating population, and it is followed for a statistical analysis where all the major loci responsible of the trait variation are located. This analysis usually referred as primary QTL mapping could locate a QTL in an interval of approximately 10 to 30 cM, which may include several hundred of genes. Therefore, the genetic resolution has to be improved by assigning a QTL to the shortest chromosome segment including ideally one single gene. The final goal is the identification of DNA coding or not coding sequences responsible for QTL (QTL cloning).

## 3.3 Determining associations between genetic markers and quantitative trait locus (QTL)

Two methods have been employed for verifying the association between the shortest possible regions of a chromosome tagged using molecular markers and the value of the studied trait: **positional cloning and association mapping.** QTL cloning is difficult because of the resolution limitations, even though many QTL had been cloned since 2001 when the first QTL was cloned in but also in that year one QTL from rice was cloned as well, since this at least 20 QTL were cloned.

**Positional cloning** allows QTL resolution but it is necessary to produce a second and larger mapping population of 2000 or more F2 plants derived from a cross between two parental nearly isogenic lines with alleles functionally different in the targeted QTL. These parental lines

are called QTL-NILs (quantitative trait loci-nearly isogenic lines). The generation of these lines can be archived doing marker assisted backcrosses or iteratively identifying and selfing individuals that are heterozygous at the QTL region. The production of such NILs can last several years depending on the plant material. Other important aspects to consider are the genetic limited variability as a result of the use of only two parental. The generated population could segregate for just a fraction of many QTL that may affect the same trait in other populations. For primary QTL mapping, Monte Carlo simulations have shown that at least 200 individuals from the segregated population are required. For higher resolution, as required for positional cloning, progenies of several thousand plants are needed. For example, in the Alpert and Tanksley´s work in 1996 more than 3,400 individuals were analyzed to obtain a detailed map around a fruit weight locus in tomato.

As an alternative to positional cloning, QTL may be determined using **association mapping**. This method allows identifying a statistic association between markers or candidates loci and the overall of an analyzed phenotype within a set of genotypes (natural populations, germplasm accessions and cultivars). It is important that the plant collection contains a wide spectrum for the trait to evaluate, and in particular it is an advantage for the analysis if the collection shows up extreme phenotypes.

Five main steps exist for the association studies:

1) Selection of the population's samples,
2) Determination of the level and influence of the structure population on the sample,
3) Phenotypic characterization of the population for the interest trait,
4) Population genotyping for regions/candidate genes candidates or as a whole genome scan,
5) Assessment of the association between genotypes and phenotypes. The selection of the association test is the last step and it depends on the population's characteristics. Association mapping uses ancestral recombination and genetic natural diversity within a population to analyze quantitative traits and it is built on the base of the LD concept.

It is used to think that the terms linkage and linkage disequilibrium have similar meanings. However, although they are related, genetic linkage makes reference to the correlated inheritance of two loci through several generations because the two loci is at a sufficiently short physical distance that recombination meiotic events do not show up, and selection acts in the same way over the two loci, whereas LD refers to

the identical frequency in the presence of two alleles of different loci inside a population, and this non-random association can be caused by other factors than linkage.

Contrary to linkage mapping, where the genetic maps are created using generations of well characterized pedigrees generated from simple or multiple crossings, the LD based association studies can rely on the variation generated by the segregation in natural populations of non related individuals. It is expected that the period of time until the most recent common ancestor between two non related individuals of a population is bigger than the time presented by a population generated by a crossing, for this cause the samples used in LD mapping present more informative meiosis, generated through history, than the meiosis showed up in a traditional population mapping. Meiosis is considered informative when effective recombinations are generated, sending information from one genetic pool to other genetic pool. In this way ancestral recombination‟s can capture mixing between different populations and within this when LD is present this is important for the association assessment.

**Factors that affect LD**

LD is affected by biological factors, as the recombination and the allelic frequencies, and for historical factors that affect population size, like the selection, and bottlenecks with extreme genetic drift, selection for or against a phenotype controlled by two non linked loci (epitasis). Mating patterns and gene flow between individuals of genetically distinct populations followed by intermating can strongly influence LD.

LD decreases faster in out crossing species than selfing species, this is due to less effective recombination in selfing species where the individuals are more likely to be homozygous than in out crossing species.

In presence of a high LD a low density of markers is required in a target region. With low LD, many markers are required but the diagnostic markers resolution is higher, potentially until the level of the gene or of QTN (i.e. the quantitative trait nucleotide polymorphism responsible for the QTL effect). It is expected high variable levels of LD through the genome due to variations in recombination rates, presence of hot spots and selection, variation in recombination rate is a key factor that contributes to the variance observed in LD patterns.

Possible complications to measure LD and therefore to carry out the association mapping, can show up due to structure population in the studied sample. The influence of structure population depends on the

relationships among sampled individuals. So, populations to be employed in an association study should be classified according to the sample individual relationship. Structure population can generate statistically significant but invalid biologically associations.

Low LD levels are expected when the population is diverse and the common ancestor within the individual population is too far in time, also low LD is not distributed uniformly along all the genome and it is located in short distances around specific loci, which produce only significant cooccurrences among physically near loci, increasing mapping resolution.

## 4.0 CONCLUSION

Breeding, domestication and a limited genetic flow in many wild species have generated erosion processes and genetic drift that have produced structured populations (i.e. populations with allelic frequencies differences among sub-populations). These populations generate not functional significant associations among loci or between a marker and a phenotype, even without marker physically binding to the responsible locus for phenotypic variation.

## 5.0 SUMMARY

In Summary, different methods have already been generated; these methods make it possible to interpret results of association tests, controlling statistically the effects of stratified populations, because association studies that do not keep in mind the effects of structure population must be viewed with skepticism. All these methods are based on the use of independent marker loci to detect and correct stratified populations.

## 6.0 TUTOR MARKED ASSIGNMENT

1. What is quantitative trait locus?
2. How do you determine associations between genetic markers and quantitative trait locus?
3. State steps involved in association studies
4. What are the factors that affect linkage disequilibrium?

## 7.0 REFERENCES/ FURTHER READING

Beer S, C, et al.,. Associations between molecular markers and quantitative traits in an oat germplasm pool: Can we infer linkages?