II. REVIEW OF LITERATURE

In spite of jasmine’s high export demand and use in perfumery and cosmetics industry, there have been limited efforts to improve this crop and hence the availability of literature is also very scanty. An attempt has been made to review the available literature related to jasmine improvement. In this chapter, the literature pertaining to the various aspects of the present investigation is reviewed. Similar research carried out in other perennial plants including plants belonging to Oleaceae family has also been reviewed.

2.1. Importance:

Jasmine is a highly valued ornamental plant for home gardens and commercial cultivation. Flowers and buds are used for making garlands, bouquets and for religious offerings, while veni is used as hair adornment. The flowers are also used for the production of perfumed hair oils and attars. Jasmine essential oil has a sweet and floral aroma. It is regarded as unique, as it blends well with other floral extracts and which is highly valued throughout the world for its high grade perfumes, which is used in soap and cosmetic industries and in flavouring mouth wash liquids. The flowers should preferably be picked at night for extraction of essential oil. Jasmine fragrance is said to give a feeling of optimism, confidence and euphoria, and is helpful against depression, nervous exhaustion and stress related conditions. Jasmine is also used for catarrh, coughs, laryngitis, dysmenorrhea, labor pains, uterine disorders and many skin problems.

India is the largest exporter of jasmine oil in the world accounting for over 40 per cent of total world export. India earned US$ 329 million by exporting of essential oil in the year 2008-09. Rose oil (US $ 3432 per 500g) commands highest price in the international essential oil markets and followed by jasmine oil (US $ 527 per 500g).
The leading export markets for Indian jasmine oil are France, accounting for over 36 per cent of total jasmine oil exports from India, followed by UAE, USA, Germany and UK (Indian Horticulture Database, 2009). Other jasmine producing countries are UAR, Morocco, Italy, Algeria etc. (Panda, 2006).

Among the varieties of jasmine, *Jasminum grandiflorum* (Pichi) is cultivated in more than 60% of the area followed by *Jasminum auriculatum* (Mullai). Tamil Nadu is the leading producer of jasmine in the country with an annual production of 77,247 tonnes from a cultivated area of 9360 hectare (Singh, 2006). Karnataka is the second highest producer of jasmine flowers. In the year 2004 Karnataka has produced 20,244 tonnes of jasmine flower from 3,451 hectare earning 8,265 lakh rupees. The Tigala community near Devanahalli and Chickaballapur are extremely good at growing flowers (Banumathy and Devi, 2004).

### 2.1.1. Origin and distribution

Jasmines are native of tropical and sub-tropical regions. The Arabian or Tuscan jasmine (*J. sambac*) is considered as native of the East Indies. But contrary opinions are also found to indicate its original home being the region of west India. The distribution of the genus is wide but majority of the species are centered on India, China and Malaysia (Anonymous, 1959). The Spanish jasmine or Catalonion jasmine (*J. grandiflorum*) is sometime referred to as *J. officinale* var. grandiflorum. The Royal jasmine or common white jasmine or Poet's jasmine (*J. officinale*) is considered to be of Persian origin. *J. auriculatum* is distributed in the western peninsula of India. It is native of south India and central provinces. The genus Jasminum is reported to comprise of around 500 species (Bailey, 1958). A critical analysis of these species revealed the number of true species to be only 89, of which 40 inhabit in Indian sub-continent (Veluswamy et al., 1975). The commercial cultivation is confined to only a few species.
viz., *J. grandiflorum, J. sambac, J. auriculatum* and *J. pubescens*. These species are commercially grown in various parts of India particularly Coimbatore, Madurai, Dindigal, Athoor, Chennai, Kalupathi, Chinniparakundram (Tamil Nadu), Bangalore, Mysore and many places in Karnataka, Andhra Pradesh, Uttar Pradesh and some parts of Bihar and West Bengal (Singh, 2006).

### 2.1.2. Taxonomy

Jasmine belongs to the family Oleaceae, order Oleales and genus *Jasminum*. Jasmine is a climbing, trailing and erect flowering shrub. There are both deciduous and evergreen species in jasmine. Leaves are opposite and alternate, simple, trifoliate or pinnate; leaflets are entire. Flowers are white, yellow or rarely reddish, sometimes solitary, more often in cymose clusters of three to many, usually fragrant; 2 loculed with 1 to 4 erect ovules. Fruit is a berry, rarely with separate capsules and each having 2 seeds. The main beauty and uniqueness of jasmine is its fragrance, which cannot be imitated by any known synthetic aromatic chemical and has a unique status in the perfume world.

One of the earliest accounts of descriptive studies in 43 jasmine species was accomplished by Hooker (1882), who divided *Jasminum* into two main groups as below

i) With simple leaves, calyx pubescent or glabrous, subulate or short.

ii) With compound leaves, either trifoliate or imparipinnate.

Engler and Prantl (1897) cited about 160 species of jasmine in the tropical and sub-tropical regions of Asia, Africa and Australia and over 40 in India. They considered that the simple leaf was only the transformed terminal leaflet of the imparipinnately compound leaf. They grouped jasmine into four groups- Unifoliata, Trifoliata,
Alternifolia and Pinnatifolia. Gamble (1936) described 20 species of jasmine occurring in Chennai, Tamil Nadu, India and classified them according to Hooker’s classification. In his study, simple leaved species *viz.*, *J. sambac* and *J. pubescens* were categorized under the group calyx pubescent and long. Compound leaved species *J. auriculatum* was grouped under trifoliate leaves, calyx lobes small. *J. grandiflorum* was categorized under leaves imparipinnate and opposite. Bailey (1947) gave the taxonomic description of 23 species as well as key for identification. Raja (1953) gave brief description of 8 south Indian species of jasmine. Wight (1963) gave taxonomic description of 21 species of jasmine with elaborate description of *J. sambac*, *J. grandiflorum* and *J. auriculatum*.

2.1.3. Cytology of jasmine

The genus with basic chromosome number of n =13 generally consists of forms with 2n = 26. Taylor (1945) in his work on cytotaxonomy and phylogeny determined the basic chromosome number of this genus as X=13. Krishnaswamy and Raman (1948) reported diploids, triploids and tetraploids in the genus *Jasminum*. Polyploidy has played an important role in origin of new cultivars in ornamental and essential oil bearing plants. It may be spontaneous in nature or artificially induced by Colchicines. Krishnaswamy and Raman (1948) reported ‘Iruvantige’ and ‘Yellusuntu-mallige’ as diploids (2n=26) and ‘Dundu-mallige’ as a triploid (2n=39) in jasmine. Spontaneous triploid (2n=39) were also reported in *J. ilicifolium* (Taylor, 1945), *J. nitidum* (Taylor, 1945), *J. primulinus* Henst (Krishnaswamy and Raman, 1948), *J. autumnale* (Sharma and Sharma, 1958), *J. sambac* (Sharma and Sharma, 1958) and *J. grandiflorum* (Murthy and Khanna, 1971). Spontaneous tetraploidy (2n=52) has been reported in *J. calophyllum* Wall and *J. flexile* Vahl (Raman, 1955). Natural occurrence of higher ploidy was also observed in *J. sambac* cv. Gundumalli (2n = 39), cultivated variety of *J. flexile* (2n =52), *J. primulinum* (2n =39) and *J. angustifolium* (2n =52) (Singh,
Sambandhmurthy and Khader (1982) reported that induced tetraploidy in *J. grandiflorum* did not reveal superiority.

### 2.1.4. Speciation in jasmine

Variation of hereditary characters by gene mutation and recombination, which form the raw material of evolution, and operative through sexual reproduction, can be ruled out. At the same time, there are about 500 species of Jasmine and several varieties exist today. Therefore, it has been a subject of investigation to many scientists to know the mode of speciation in this genus. These investigations on hereditary variation through alteration in the somatic chromosome complement of *Jasminum* species offer a clue to solve the problem of the mechanism of speciation in vegetatively reproducing plants. The phenomenon of cytomixis is characterized by the migration of chromatin/chromosomes between the proximate meiocytes through cytoplasmic channels or intercellular bridges. In the PMC (pollen mother cell), since there is no other division other than the meiotic division, further shuffling of chromosome could not be expected. Here the meiotic product either degenerates or become non-functional. On the other hand, in the meristematic tissue of shoot bud, repeated mitotic divisions are likely to involve a series of cytomixis, shuffling of chromosome and consequently stable genetic constitution. It results in different ploidy levels but most commonly aneuploidy. When one of these cells with altered chromosome number undergoes normal mitotic divisions, it undergoes a homogenous but heteroploid tissue. A shoot arising from such a tissue when propagated vegetatively after continuous natural selection gives rise to a new variety or species (George and Geethamma, 1985).

Sharma and Sharma (1958) stated that speciation in *Jasminum* spp has been principally affected by the structural alteration of chromosomes. Continuous accumulation of these structural changes has been the main reason for the origin of new
species, although polyploidy too has assisted to some extent initially, in the speciation of the genus. Gupta and Sharma (1972) reported that polyploidy and gene mutation played an important role in speciation of the genus and there is great scope for evolving new and improved species of jasmine. Jasmine is highly domesticated and propagated almost entirely by vegetative means. Sexual reproduction is obsolete and almost absent in many of the jasmine species, many being either pollen sterile. A study by George and Geethamma (1983) showed that the sterility of the pollen is ultimately due to some defective gene function in their meiotic cells causing various meiotic abnormalities, including cytomixis which occurs spontaneously and abundantly in their pollen mother cells.

2.1.4.1. Species

Some important species of jasmine are *J. auriculatum*, *J. beesianum* (Rosy Jasmine), *J. calophyllum*, *J. dichotomum* (Gold Coast Jasmine), *J. dispermum*, *J. faveri* or *J. caudatum*, *J. humile*, *J. bignoniaceum* (Yellow jasmine or Italian jasmine), *J. multiflorum*, *J. nitidum*, *J. officinale*, *J. parkeri*, *J. polyanthum*, *J. primulinum*, *J. pubigerum*, *J. revolutum*, *J. rex* (Kings Jasmine), *J. sambac*, *J. smilacifolium*, *J. stephanense*, *J. trinerve*, *J. volubile* or *J. gracile* (Australian or Wax Jasmine) and *J. Wallichianum*. Some commercially important species are described below.

**J. auriculatum**: It is a shrub with shiny leaves and auricles. Flowers are white, sweet scented borne in pubescent, compound, many flowered flax cymes. Corolla lobes are elliptical and fruit is black. Flowers are used for production of perfumes.

**J. grandiflorum**: This jasmine is commonly called as Spanish Jasmine. It is a woody bush with pinnate leaves bearing 3 to 5 leaflets of equal size. It bears large, white flowers with delightful fragrance,
borne in lax axillary or terminal cyme. It is extensively used for preparation of garland and decorative bunches.

**J. sambac:** It is popularly known as Arabian or Tuscan Jasmine. It is an evergreen twiner having simple cordate to oblong dark green leaves. Flowers are white, fragrant and small with three forked cymes. Essential oil is extracted for perfumery industry. The flowers are used for making garland, bouquet etc.

**J. officinale:** It is commonly termed as White Jasmine or Poet’s Jasmine. It is a deciduous climber bearing stalkless leaflets. Flowers are white and fragrant, borne in loose clusters.

**J. multiflorum:** It is commonly called as Fussy/ Downy/Star Pinwheel Jasmine. It is a strong woody vine bearing simple, opposite leaves. Flowers are sessile, fragrant, large, white and borne in terminal umbels, having 6 to 9 petals.

**J. calophyllum Vahl:** It is known as Pandal Malli in local language. It is a profuse bloomer producing scented white flowers. Plant blooms throughout the year. Plants are generally free from pest and disease attack.

**J. arborescens:** It is known a Muta Bela and Neba Mallinga in Karnataka. Flowers are white and have strong fragrance. It is also grown in West Bengal and Uttar Pradesh.

**J. flexile Vahl:** It is also known as *J. caudatum* Wall. It is a profuse flowering species grown widely in home garden for its fragrant flowers. Plant blooms throughout the year. Cultivar is free from insect, pest and disease.

**J. humile L:** It is also known as Semmalligai and Pelli Chameli. It produces yellow flower and is also known as Yellow Jasmine. Flowers are fragrant and are used in the perfume industry.
**J. pubescens Wild:** It is known as Kundo and Kunda in local language. It is an ornamental plant with slightly fragrant, white coloured flower and it blooms throughout the year. Flowering is in its peak during winters. It is grown in North India.

### 2.1.4.2. Varieties

There are many varieties of jasmine which are commercially grown in India. Some are local varieties and some are released from Tamil Nadu Agricultural University, Coimbatore and Indian Institute of Horticultural Research, Bangalore. Following are some commonly cultivated varieties of jasmine.

**Jasminum auriculatum:** CO-1 Mullai, CO-2 Mullai, Pari Mullai, Long Point, Short Point.

**J. grandiflorum:** CO-1 Pitchi, CO-2 Pitchi, and Arka Surabhi.

**J. sambac:** Double Mohra, Gundumali, Iruvatchi, Kasturimalli, Madanbari, Oosimalli, Ramabanam, Single Mohra, Soojimalli, Khoya, Khoya Large, Arka Aradhana, Bale Japani, Butt Mohra.

**J. multiflorum:** Arka Arpan

### 2.1.5. Crop improvement in jasmine:

Bud sports, induced mutation and polyploidy has played important role in development of new cultivars of Jasmine. Details are given below.

**Mutation Breeding:**

i) **Spontaneous Mutant:** Rao and Krishnan (1980) reported bud sport in *J. auriculatum* Vahl. with increased bloom size. Analysis of spontaneous mutant revealed that the mutant excelled in length and width of floral bud, length of corolla tube, diameter of open flower,
number of petals, length, and width of petal and weight of 100 floral buds.

**ii) Induced Mutant:** Nambisan *et al.* (1980) reported two induced mutant in *J. grandiflorum*. One mutant was dwarf in nature while the other was resistant to *Cercospora jasminicola*. Flowers of mutants yield lower concrete percentage. Effect of gamma irradiation on cuttings of *J. grandiflorum* has been studied by Kumar *et al.* (1983). They have irradiated rooted cutting of variety CO-1 Pitchi by 5 to 3 Krad. Reduction in sprouting and rooting percentage was observed after irradiation in comparison to control. Devaiah and Srivastava (1989) reported that LD50 for *J. grandiflorum* var. 'Pink Pin' is close to 2.5 Krad, 0.5 Krad for 'Pink Thrum', close to 2.5 Krad for *J. flexile*, close to 1 Krad for *J. calophyllum* Wall and 2 Krad for *J. sambac* Ait 'Gundumullai. Percentage of rooting, number of roots per cuttings, and, length and thickness of roots decreased with increase in intensity of gamma irradiation.

### 2.1.6. Propagation:

#### 2.1.6.1. Climate and soil

Jasmines are grown in different climatic conditions. Some species are hardy and can be grown in adverse climatic conditions. But, in general, mild tropical climate is good for proper plant growth and flower production. Jasmine can be grown on almost all types of soil, but gives best results when grown in well drained rich loamy soil.

#### 2.1.6.2. Vegetative propagation

Jasmine is commonly propagated by vegetative means i.e. cutting, layering, budding and grafting and by suckers. It can be multiplied by cutting and rooting, which take 50-70 days. Large scale multiplication of uniform and disease free plant can be achieved by tissue culture method. Propagation through seed, seed setting and its
germination is a rare phenomenon; however it depends on genotype and surrounding environment. Sterility and meiosis in five species of jasmine has been studied by Karmakar and Srivastava (1986). Extensive studies on seed set and germination has been carried out at IIHR, Bangalore. Pollen sterility and non-viability may be the causes for poor seed setting (Dadlani et al. 1988).

2.1.6.3. Harvesting, yield and income

Flowering in jasmine starts from second year or sometime earlier, but economic yield is generally obtained from third year. The stage of flower harvest depends on the type of flower use. For fresh flower, fully developed unopened flowers are picked early morning; while for extraction of concrete only fully opened are chosen. Flower yield in jasmine varies from species to species and it also depends on cultural practices and management of cultivation. Flower yield of *J. auriculatum* is 4,636-9,022 kg/ha; for *J. grandiflorum* it ranges from 4,329- 10,144 Kg/ha and in *J. sambac* it ranges from 739-8,129 kg/ha.

The sale price of the crop varies from Rs.20 to Rs.70 per kg, depending on the season, and an average price of Rs.25 per kg is assumed for working out the economics. Accordingly, the gross income per acre would be Rs.18,750 per acre in first year, Rs.50,000 in second year, Rs.62,500 in third year and Rs.87,500 from fourth year onwards.

2.1.6.4. Disease and Insects associated with Jasmine:

Jasmine is affected by fungal diseases like leaf blight, rust, wilt and viral diseases like mosaic and phyllody. It is also affected by insects like bud worm (*Hendecasis duplifasciallis*), mites, plant hopper, citrus whitefly etc. Root knot nematode (*Meloidogyne incognita*) infest *J. sambac* and *J. flexile* causing small swelling and
enlarged rootlets leading to conspicuous pale yellowish leaves and die back.

2.1.7. Jasmine essential oil

2.1.7.1. Essential oil production in plant

From the vast number of species of plants that are known, about 3000 essential oils have been well identified, though only some 150 have been exploited for commercial production. The most odoriferous plants are found in the tropics, where the solar energy is greatest. Plant volatile oils are synthesised, stored and released to the environment by a variety of epidermal or mesophyll structures, whose morphology tends to be characteristic of the taxonomic group. These structures in leaves, roots, stem, floral part and fruits include oil cells, secretory glands (eucalyptus), glandular hairs or trichomes (labiates). Epidermal hairs (including root hairs) or glands are cytochemically divisible into secreting and non-secreting types, depending on whether or not, respectively. They accumulate specialised secondary phytochemicals within them. The essential-oil bearing plants are unique in possessing specialised secretory type epidermal hairs or trichomes to synthesise and accumulate large quantities of these compounds. These epidermal appendages (glandular trichomes, glandular hairs, resin ducts etc. occur in different plant parts from flowers to roots with special individual attributes (Sangwan et al., 2001).

The constituents of plant essential oils fall into two entirely distinct chemical classes, terpenoids and phenylpropanoids. Although terpenes represent the major components, occurring much more frequently and abundantly, whenever phenylpropanoids are present they provide indispensable and significant flavour and odour to the oil. It is well established that terpenoids are synthesised from five carbon units of isopentenyl pyrophosphate (IPP) and its isomer,
dimethylallyl pyrophosphate (DMAPP). Phenylpropanoids originate through the shikimate pathway [Sangwan et al., 2001].

2.1.7.2. Jasmine concrete and absolute in perfumery industry:

Jasmine concrete and its absolute is an invaluable item in perfumery industry. It is used in highly expensive perfumes and a considerable quantity of jasmine concrete has been produced in India for the last 20 years. Jasmine concrete is the product obtained by solvent extraction of the fresh harvested jasmine flowers. It is also interesting to note that it takes 8,000 carefully hand-picked blossoms to produce 1 gram (about 1 ml) of jasmine concrete. It has a semisolid consistency and contains the volatile perfumery principles of the flowers, waxes and colouring matters. Concretes are not widely used in perfumery in their native form but are generally converted into an alcohol-soluble volatile concentrate known as an absolute, i.e. they have to be extracted with alcohol. It is a reddish-brown, free flowing liquid and represents the concentrated form of jasmine perfume. Although there are several methods of preparation of jasmine perfumes, solvent extraction method is generally employed. The solvent used for extraction is hexane. Two varieties of the flowers of scented jasmine are preferred for the production of the concretes. They are *Jasminum grandiflorum* (Chameli) and *J. sambac* (Gundumalle). The yield of concrete from the grandiflorum flowers is about 0.25 - 0.30 % and that from *J. sambac* flowers is about 0.15 - 0.18 %. The absolute content in the concretes varies from 45 - 55 % depending on the extraction technique and also on the place of cultivation of the plants.

2.1.7.3. Chemical composition of the absolutes of jasmine

The main constituents of *J. grandiflorum* absolute are benzyl benzoate, phytol and isophytol. The minor constituents are eugenol, cis-jasmone, n-acetyl methyl anthranilate, jasmine lactone and cis-
and trans-methyl jasmonates. These constituents are responsible for the characteristic fragrance of *J. grandiflorum* absolute. The main constituents of *J. sambac* absolute are benzyl alcohol, linalool, cis-3-hexenyl benzoate, indole, (E)-farnesene and methyl anthranilate. The minor odorous principles of the *J. grandiflorum* absolute are also present in the *J. sambac* absolute. The characteristic fragrances of the two absolutes are distinctly different because of the differences in their chemical composition.

### 2.2. Morphological studies in jasmine

Improvement of any crop depends on the availability of genetic variation in the germplasm of the crop and heritability of the traits. Effective selection of important traits from cultivated and wild germplasm is an age old process followed for crop improvement. Detection and estimation of genetic variability in working collections (germplasm or population) is a prerequisite in any crop improvement programme. The variability present in any population is due to genetic and environmental factors. The relative contribution of many factors for total variability determines genetic gain possible through selection.

Much work has been done by selection among different species of jasmine and their clones at Tamil Nadu Agricultural University, Coimbatore. Sundaraj *et al.* (1967) were able to select a high yielding clone “Parimullai” in the species *J. auriculatum*. This selection was reported to be resistant to gall mite which is a major problem in jasmine. A triploid form of *J. grandiflorum* with chromosome number 2n= 39 was reported by Khan *et al.* (1969). The triploid was found to be more vigorous with longer leaves, longer petals, peduncle, corolla tube and increased size and thickness of buds compared to the diploid cultivars. But flower yield, total concrete and percentage of absolute were found to be lower in case of triploid than diploids (Murthy and Khanna, 1971). Khan *et al.* (1969) observed the size of open flowers in *J. nitidum* to be larger than those of *J. auriculatum* and cultivars
“Iruvatchi” and “Gundumalli” of *J. sambac*. They also reported that there was no significant enhancement in flower number per plant, flower size or yield in the tetraploid *J. rigidum* over diploids. Khan and Muthuswamy (1969) separated a population of *J. auriculatum* into five groups of morphological variants namely long pointed bud, medium pointed bud, short pointed bud, long round bud and short round bud. Raman *et al.* (1969) while studying the flowering habit and flower yields of some *Jasminum* spp., found *J. grandiflorum* to give highest flower yield followed by *J. flexile*. However, largest flowers and buds were found on *J. sambac* var. Double Mohra and Big Double. Khan *et al.* (1970) has reported fifteen morphological variants in *J. sambac* and concluded that the variety Madanban was the best followed by Gundumalli and Ramabanam for various economically important characters like shape of bud, length of pedicel, length of corolla tube, diameter of flower, number of flowers per plant and time taken for a bud to open up completely. Muthuswamy *et al.* (1972) reported that a lot of variation in length of flower bud, length of corolla tube and diameter of flower bud could be found in seedlings obtained through open pollination in *J. auriculatum*.

Thangaraj (1977), while studying the variability and genetic improvement in *J. auriculatum* found that the maximum genetic potentiality of seedling originated plants was present in case of flower bud length, flower bud diameter and flower yield. The best seedlings for yield, weight of flower buds and diameter of flower buds were secured from the mother clone Short Round. In further studies, Thangaraj *et al.* (1982) have reported high variability for flower bud weight, corolla tube length and flower bud diameter. The environmental influence was reported to be very high for yield and low for other floral characters. Muthukrishnan and Pappiah (1980) reported the isolation of CO-1 Mullai in *J. auriculatum*, which has desirable flower bud character like bold and long corolla tube. Veluswamy (1980) observed that the clone of *J. grandiflorum* J.g.-3,
gave high yield of 10.15 tonne of flower buds per hectare, and concrete yield per hectare of 29.42 kg with concrete recovery of 0.29%. This clone was later named as CO-1 Pitchi (Sambandamurthy and Khader, 1982).

Distyly in flowers of *J. pubescens* was reported by Srivastava and Karmakar (1985). Plants producing white flowers were long styled (pin) whereas those with pink flowers were short styled (thrum). They observed that pin type possessed higher bud weight and longer corolla tube length but these flowers lacked fragrance and indole compound. On the other hand, thrum type excelled pin type by having high fragrance and indole compounds in flower and also higher yields per day during the blooming season.

### 2.3. Diversity studies in jasmine based on morphological traits:

Highly inherited characters are controlled by one or few genes. Additive gene action forms the basis of heritable part of genetic variation in a variable population. A character under study will respond better to continuous selection if other components of genetic variability namely dominance and epistasis (non-additive gene action) are absent or very negligible (Panse, 1957).

More (1980) studied correlation, regression coefficient and path analysis of eight characters comprising yield and its components in jasmine. He observed number of flowers to be significantly correlated with yield. Weight of hundred flowers had a significant and positive correlation with bud diameter, corolla tube length and bud shape index. Number of flowers had the maximum direct effect on yield followed by total floral bud length and floral bud length excluding the corolla tube.

Positive association of number of primary laterals and the length of style with yield in *J. auriculatum* was observed by Thangaraj et al. (1980). They also found that weight of flower buds, length of
internodes and numbers of days taken for flowering are directly related with corolla tube length.

While studying the correlation and path analysis in pin and thrum types of open pollinated progeny in *J. auriculatum*, Karmakar and Srivastava (1986) observed flower yield of the plant to be significantly correlated with diameter of flowers, petal length, bud length, 100 bud weight, number of flowering branches per plant and number of flowers per plant with flower yield in pin type. They concluded that 100 bud weight and number of flowers per plant in case of thrum type and number of flowers per plant in case of pin type to have major contribution to yield.

### 2.4. DNA marker based diversity studies in jasmine and related plants in Oleaceae family

The molecular markers are one of the versatile tools in Molecular Biology and Genetic Engineering. With the advent of molecular markers, new generation of markers have been introduced over the last two decades, which has revolutionized the entire scenario of biological sciences. Ever since their development, they are constantly being modified to enhance their utility and to bring about automation in the process of genome analysis.

The term ‘marker’ implies that the loci detected are anonymous and phenotypically neutral. These DNA based makers differentiate the organisms at DNA level and are inherited in simple Mendelian fashion (Waltson, 1993). They are particularly important for genetic resource management, as well as for the rational use of genetic resources in selection programs. An ideal genetic marker should be polymorphic, multi-allelic, codominant, non-epistatic and it should be insensitive to environment. Recent advances in the field of molecular biology have provided tools such as DNA markers, which can detect differences in genetic information carried by two or more individuals.
These makers at present are applied in a number of studies viz. forensic studies, identifying genes responsible for disease resistance, evolutionary linkage mapping, paternity testing, map based cloning and genetic diversity studies.

Molecularly, polymorphism is of two types *viz.* sequence polymorphism and polymorphism due to number of repeated sequences. There are different types of DNA markers and many more are being discovered. There are two important types of DNA markers namely hybridization based markers and PCR based makers.

DNA-based molecular markers such as randomly amplified polymorphic DNA (RAPD), Restriction fragment length polymorphisms (RFLPs), Amplified fragment length polymorphisms (AFLPs), Simple sequence repeats (SSRs) or microsatellites, Sequence tagged sites (STS) and Single nucleotide polymorphism (SNP) have been used for fingerprinting of varieties and to differentiate among genotypes at species and subspecies level of the DNA markers RAPD, SSR and AFLP have been widely used for genetic diversity assessment. The literature on the use of the DNA markers for genetic diversity assessment in jasmine is reviewed and presented here.

### 2.4.1. RAPD Marker study in jasmine

A preliminary molecular work has been done using RAPD in which 32 cultivars of *Jasminum* spp. were screened using one hundred and forty random ten base long primers. Thirty five primers gave atleast 5 bands/primer. Eight of thirty five primers gave on an average ten strong repeatable bands and chosen for measuring diversity. From these primers 134 bands were amplified and number of polymorphic bands ranged from 13-26. Dissimilarity matrix based on Squared Euclidean Distance and clustering by Ward’s coefficient indicated a moderate diversity among the 32 jasmine cultivars (Mukundan, 2000).
2.4.2. AFLP Marker studies in Oleaceae family

Amplified Fragment Length Polymorphism (AFLP) involves digestion of DNA with the restriction enzymes as in case of RFLP and PCR amplification. AFLP technique combines the strengths of two methods, the reproducibility of restriction fragment analysis and the power of the PCR (Mueller and Wolfenbarger, 1999). PCR amplification of restriction fragments is carried out using oligonucleotide adapters (Vos et al., 1995). It is relatively expensive (not on a data point basis), fast and reliable method that generates hundreds of informative genetic markers (Vos et al., 1995; Vos and Kuiper, 1997; Hillis et al., 1996). They are employed in assessing genetic diversity in many perennials like olive (Rosa et al., 2003) pinus (Lerceteau and Szmidt, 1999) and Azadiracta indica (Singh et al., 1999).

The key feature of AFLP-PCR is its capacity for the simultaneous screening of many different DNA regions distributed randomly throughout the genome. In essence, the AFLP technique involves the digestion of DNA template with specific restriction enzymes, followed by the ligation of a specially designed adapter onto the sticky ends of the sample DNA, using amplification of subsets of genomic restriction fragments (Vos et al., 1995, Vos and Kuiper, 1997). The results are highly informative fingerprints. Fingerprinting is the identification of individuals based on a pattern of DNA markers that can be detected in the genomic DNA of an individual (Fairbanks and Andersen, 1999).

Fingerprints are an increasingly popular method for crop diversity studies through estimation of genetic distances. Ideally, a fingerprinting technique does not require any kind of investment on sequence analysis, primer synthesis or characterization of DNA probes.

Currently, AFLP markers are also used in biodiversity studies, population and conservation genetics, and QTL mapping (Mueller and
Wolfenbarger, 1999). AFLP markers are also suitable for analysis of relatedness, parentage, mating frequency and other genetic parameters (Vos et al., 1995). The use of AFLP markers is based on their advantages: they are highly sensitive and show a high level of polymorphism, are highly reproducible, and widely applicable. Moreover, AFLPs just require the use of generic primers to produce a practically unlimited number of markers (Vuylsteke et al., 2000). One of the strongest advantages of these markers is their capacity to resolve extremely small genetic differences. For instance, AFLP markers have been used to distinguish near isogenic lines of soybeans that differ only in a single small region of the entire genome (Maughan et al., 1996). Among other utilities of the AFLP technique is its capacity of generating markers from any organism with DNA, without any prior knowledge about the genomic make-up of the organism. For example, AFLP markers have been efficiently used in bacteria (Huys et al., 1996; Keim, 1997), fungi (Rosendahl and Taylor 1997; Gonzalez et al., 1998; Majer et al., 1998), nematodes (Semblat et al., 1998), vertebrates (Otsen et al., 1996; Ajmone et al., 1998; Liu et al., 1998), cultivated crops (Maughan et al., 1996; Mackill et al., 1996; Menz et al., 2004) and trees (Gaiotto et al., 1997; Arens et al., 1998; Paglia and Morgante, 1998). Another great advantage of AFLP markers is that their error levels are less than those of any other molecular markers. Moreover, AFLP amplifications are performed under high selection at high stringency conditions, eliminating any artifactual variation that is commonly seen in RAPD-PCR markers (Vos et al., 1995). It requires a minimal amount of DNA, and partially degraded samples can be used compared with other molecular markers that require high amounts of DNA (Vos et al., 1995).

The first linkage map of the olive (Olea europaea L.) genome has been constructed using random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP) as dominant markers and a few restriction fragment length polymorphisms (RFLP)
and simple-sequence repeats (SSR) as co-dominant markers. Ninety-five individuals of a cross progeny derived from two highly heterozygous olive cultivars, Leccino and Dolce Agogia, were used by applying the pseudo test-cross strategy. From 61 RAPD primers 279 markers were obtained – 158 were scored for Leccino and 121 for Dolce Agogia. Twenty-one AFLP primer combinations gave 304 useful markers – 160 heterozygous in Leccino and 144 heterozygous in Dolce Agogia. In the Leccino map 249 markers (110 RAPD, 127 AFLP, 8 RFLP and 3 SSR) were linked. This resulted in 22 major linkage groups and 17 minor groups with fewer than four markers. In the Dolce Agogia map, 236 markers (93 RAPD, 133 AFLP, 6 RFLP and 4 SSR) were linked; 27 major linkage groups and three minor groups were obtained. The total distance covered was 2,765 cM and 2,445 cM in the Leccino and Dolce Agogia maps, respectively. The mean map distance between adjacent markers was 13.2 cM in Leccino and 11.9 cM in Dolce Agogia, respectively. Both AFLP and RAPD markers were homogeneously distributed in all of the linkage groups reported. The stearoyl-ACP desaturase gene was mapped on linkage group 4 of cv. Leccino (Rosa et al., 2003).

In *Olea europaea* (Sensi et al., 2003), a highly consistent AFLP banding pattern was found by use of the automated ALFexpress II DNA sequencer. A non-radioisotopic AFLP method was optimized for European beech (*Fagus sylvatica* L.), a member of Oleaceae family and was found to be reproducible. In this study, AFLP technique was optimized and reproducibility tested for different starting sample concentration by use of an automated DNA sequencer (ALFexpress II, Amersham Pharmacia). Sequences of AFLP adapters and primers were based on Vos et al. (1995).

The genetic structure of wild and cultivated olive tree populations was evaluated by amplified fragment length polymorphism (AFLP) markers at a microscale level in one continental and two insular Italian regions. The observed patterns of genetic
variation were able to distinguish wild from cultivated populations and continental from insular regions. Island oleasters were highly similar to each other and were clearly distinguishable from those of continental regions. Ancient cultivated material from one island clustered with the wild plants, while the old plants from the continental region clustered with the cultivated group. On the basis of these results, they showed that olive trees have undergone a different selection/domestication process in the insular and mainland regions. The degree of differentiation between oleasters and cultivated trees on the islands suggests that all cultivars have been introduced into these regions from outside, while the Umbrian cultivars have originated either by selection from local oleasters or by direct introduction from other regions (Baldoni et al., 2006).

Casas et al. (2006) studied genetic structure and evolutionary patterns of the wild olive tree (Olea europaea L.), a member of Oleaceae family with AFLP fingerprinting data in Eurasia and Africa and the Mediterranean basin. Two statistical approaches (Bayesian inference and analysis of molecular variance) were used to analyse the AFLP fingerprints. The AFLP procedure followed was that of Vos et al. (1995) with the Odi Wcations of Schönswetter et al. (2004) and the primer combinations of Angiolillo et al. (1999).

Yan et al. (2008) carried out studies on genetic diversity and genetic relationships of 46 Osmanthus fragrans cultivars, collected from Hubei, Zhejiang and Guangxi provinces in China, by the technique of amplified fragment length polymorphism (AFLP). Ten primer combinations were used which generated 436 scorable bands including 269 polymorphic. It indicated that each primer combination generated 26.9 polymorphic bands. Genetic similarities were obtained using simple matching (SM) coefficients, and a dendrogram of the 46 cultivars was established by UPGMA clustering method. The high level genetic variations in 46 O. fragrans cultivars were proved by the SM coefficient value from 0.69 to 0.87. The cluster analysis suggested that
the 46 *O. fragrans* cultivars could be divided into five groups, and this result was not absolutely consistent with the morphology-based traditional classification. The cluster analysis showed that there were close genetic relationships among cultivars of the same flower color, and the geographic origin of *O. fragrans* was correlated with the analysis cluster results at a certain level. Cluster analysis also indicated that Guangxi cultivars were distinct from those of Hubei and Zhejiang.

Rao *et al.* (2009) reported AFLP profiling of 70 olive genotypes that represents varieties mainly used for production of PDO olive oil in the Campania region of Italy. AFLP analyses provided clear genetic differentiation between cultivars. Genetic distances were also calculated based on 40 phenotypic traits. For morphological traits, dissimilarity between cultivars was computed using the simple matching coefficient (Sneath and Sokal, 1973). The cultivars were then clustered by the unweighted pair-group method with arithmetic averages (UPGMA). A co-phenetic value matrix of the UPGMA clustering was used to test goodness of fit. AFLP markers were generated as described by Vos *et al.* (1995) with modifications. Ten primer combinations were used for selective amplification. These primers were *Eco RI*- AAC combined with *Mse I*- caa, *Mse I*- CAC, *Mse I*- CAG, *Mse I*- CAT, *Mse I*- CTA, *Mse I*- CTC, *Mse I*- CTG and *Mse I*- CTT. Genetic similarities were evaluated by calculating the Dice coefficient (Dice, 1945) and samples were clustered by the UPGMA method. Goodness of fit of the clustering to the similarity matrix was tested for the data. Morphological and AFLP based genetic distances yielded different hierarchical pattern, although the two data sets were both useful for assessing the presence of genetic variation.

**2.4.3. SSR Marker studies in Oleaceae family**

Litt and Lutty (1989) coined the term microsatellites, which are otherwise called as Simple Sequence Repeats (SSRs). They consist of
tandem repeats of simple motif sequences, usually one to five bases, which are amplified by PCR. Tandem repeats of many simple sequence motifs, in particular the dinucleotide repeats are abundant in most eukaryotic genomes, and are distributed throughout these genomes in dispersed locations. These microsatellite repeats are flanked by unique sequences, occurring only once in the genome. Microsatellites are highly informative owing to their high degree of polymorphism and co-dominance.

These are a class of repeat sequences that are comprised of tandem repeats of short, core sequences dispersed throughout the genome. When these core sequences are hybridized to restriction enzyme digested genomic DNA, they detect several hyper variable loci simultaneously. Microsatellites are quite attractive as the assay is PCR based. It is sufficient to merely separate the amplification products by electrophoresis. If the DNA sequence information is made available, individual laboratories can synthesize their own oligonucleotide. (CA)<sub>n</sub> repeat is one of the most frequently occurring microsatellite in human and many mammalian genomes. In contrast (AT)<sub>n</sub> is more abundant in plants.

Carriero et al. (2002) obtained a small insert genomic library of *Olea europaea* L., highly enriched in (GA/CT)<sub>n</sub> repeats. The sequencing of 103 clones randomly extracted from this library allowed the identification of 56 unique genomic inserts containing simple sequence repeat regions made by at least three single repeats. A sample of 20 primer pairs out of the 42 available were tested for functionality using the six olive varieties whose DNA served for library construction. All primer pairs succeeded in amplifying at least one product from the six DNA samples, and ten pairs detecting more than one allele were used for the genetic characterisation of a panel of 20 olive accessions belonging to 16 distinct varieties. A total of 57 alleles were detected among the 20 genotypes at the ten polymorphic SSR loci. The remaining primer pair allowed the amplification of a single
SSR allele for all accessions plus a longer fragment for some genotypes. Considering the simple sequence repeat polymorphism, 5.7 alleles were scored on average for each of the ten SSR loci. A genetic dissimilarity matrix, based on the proportion of shared alleles among all the pair-wise combinations of genotypes, was constructed and used to disentangle the genetic relationships among varieties by means of the UPGMA clustering algorithm. Graphical representation of the results showed the presence of two distinct clusters of varieties. The first cluster grouped the varieties cultivated on the Ionian Sea coasts. The second cluster showed two subdivisions: the first sub-cluster agglomerated the varieties from some inland areas of Calabria; the second grouped the remaining varieties from Basilicata and Apulia cultivated in nearby areas. Results of cluster analysis showed a significant relationship between the multilocus genetic similarities and the geographic origin of the cultivars.

SSR and RAPD markers were used to develop DNA fingerprints of 100 accessions of olive trees (Olea europaea L.) from the National Olive Variety Assessment (NOVA) collection, located at the University of Adelaide, Roseworthy Campus (South Australia). The markers were compared for their ability to discriminate between genotypes and their level of polymorphism per marker. The results were similar for both types of markers, and it was shown that several cultivars in the collection had similar genotypes. However, nine SSRs were more discriminatory as more genotypes were identified with these, than with four RAPD markers. The DNA fingerprints collected have been used for the construction of a genetic database that was used to identify eight out of presumed Mission olive cultivar samples obtained from cultivated olives planted around the early Mission sites in the USA (Mantia et al., 2006).

Belaj et al. (2007) examined the pattern of genetic variability and genetic relationships of wild olive (Olea europaea subsp. europaea var. sylvestris) populations in the north-western Mediterranean. The
genetic variation within and between 11 wild olive populations (171 individuals) was analysed with eight microsatellite markers. Bayesian model-based clustering identified four gene pools, which was in overall concordance with the Factorial Correspondence Analysis and Fitch–Margoliash tree. Two gene pools were predominantly found in southern Spain and Italian islands, respectively, in samples gathered from undisturbed forests of the typical Mediterranean climate. The other two gene pools were mostly detected in the north-eastern regions of Spain and in continental Italy and belong to the transition region between the temperate and Mediterranean climate zones. The study indicated a degree of admixture in all the populations, and suggests some caution regarding genetic differentiation at the population level, making it difficult to identify clear-cut genetic boundaries between candidate areas containing either genuinely wild or feral germplasm.

Four invasive populations from Australia and Hawaii were characterized using eight nuclear DNA microsatellites, plastid DNA markers as well as ITS-1 sequences. Based on these data, their genetic similarity with native populations was investigated, and it was determined that East Australian and Hawaiian populations (subsp. Cuspidata) have originated from southern Africa while South Australian populations (subsp. Europaea) have mostly derived from western or central Mediterranean cultivars. Invasive populations of subsp. Cuspidata showed significant loss of genetic diversity in comparison to a putative source population, and a recent bottleneck was evidenced in Hawaii. Conversely, invasive populations of subsp. Europaea did not display significant loss of genetic diversity in comparison to a native Mediterranean population. Different histories of invasion were inferred for these two taxa with multiple cultivars introduced restoring gene diversity for Europaea and a single successful founder event and sequential introductions to East Australia and then Hawaii for cuspidate (Besnard et al., 2007).
Brito et al. (2008) analysed native olive plants from *Olea maderensis* (*O. europaea* ssp. *cerasiformis*) and *O. cerasiformis* (*O. europaea* ssp. *guanchica*), wild olives (*O. europaea* ssp. *europaea* var. *sylvestris*) and cultivated olives (*O. europaea* ssp. *europaea* var. *europaea*) with respect to genome size and microsatellite markers. The mean nuclear DNA content of *O. maderensis* was estimated as $5.97 \pm 0.191$ pg/2C, while the remaining studied taxa presented mean genome sizes ranging from 2.99 to 3.18 pg/2C. These data and the obtained simple sequence repeats (SSR) profiles, i.e., with 2–4 alleles in *O. maderensis* and a maximum of two alleles in the other taxa, enabled the identification of a new ploidy level, tetraploidy, for a species belonging to the Olea genus. Cluster analysis of the microsatellite data revealed a clear separation of each species in different clusters and a high genetic dissimilarity could be observed among genotypes belonging to different species. This work contributed to a better characterization of olive species and the obtained data can be helpful to support taxonomic studies, and to develop germplasm preservation strategies in endangered populations of *O. maderensis* from Madeira Archipelago.

Olive genetic diversity was studied using RAPD and SSR techniques, on plants growing in the Emilia territory (Reggio Emilia and Parma provinces), Northern Italy. Genotype identification comparisons were made with 8 cultivars, some of which from Central Italy. Screening was done by analysing patterns produced by 20 RAPD primers and 3 SSR primers. The dendrograms obtained from the analysis show the genetic relationship among accessions present in the Parma-Reggio Emilia district (Ganino et al., 2007).

To assess the genetic diversity in Moroccan cultivated olive, *Olea europaea* L. subsp. *europaea*, Khadari et al. (2007) performed molecular analysis of olive trees sampled in four geographic zones representing all areas of traditional olive culture. The analysis of 215 trees using 15 simple sequence repeat (SSR) loci revealed 105 alleles
distributed among 60 SSR profiles. The analysis of chloroplast deoxyribonucleic acid polymorphism for these 60 olive genotypes allowed to identify four chlorotypes: 42 CE1, one CE2, nine COM1 and eight CCK. Among the 60 SSR profiles, 52 corresponded to cultivated olive trees for which neither denomination nor characterisation is available. These local olive genotypes displayed a spatial genetic structuring over the four Moroccan geographic zones (northwest, north centre, Atlas and southwest), as pairwise Fst values ranged from 0.0394 to 0.1383 and varied according to geographic distance.

Besnard et al. (2008) carried out exhaustive study to infer putative polyploidization events and their evolutionary significance in the diversification of olive tree complex (*Olea europaea*) in north-west Africa. Representatives of six olive subspecies were investigated using (a) flow cytometry to estimate genome content, and (b) six highly variable nuclear microsatellites to assess the presence of multiple alleles at co-dominant loci. In addition, nine individuals from a controlled cross between two individuals of *O. europaea* subsp. Marocana were characterized with microsatellites to check for chromosome inheritance. Based on flow cytometry and genetic analyses, strong evidence for polyploidy was obtained in subspp. cerasiformis (tetraploid) and marocana (hexaploid), whereas the other subspecies appeared to be diploids. Lastly, abnormalities in chromosomes inheritance leading to aneuploid formation were revealed using microsatellite analyses in the offspring from the controlled cross in subsp. marocana. This study constitutes the first report for multiple polyploidy in olive tree relatives. Formation of tetraploids and hexaploids may have played a major role in the diversification of the olive complex in north-west Africa.

Forty cultivars were chosen with different chlorotypes that represent the olive RAPD-based diversity in the Mediterranean basin to know the origin of olive. They constructed a dendrogram, as reference, using 12 molecular SSR data. Cultivar origins in the
oleaster were examined by using Bayesian clustering methods. It was calculated whether some cultivars may be grouped by admixture analysis to these GRPs and how much olive cultivars admixed to the previously defined oleaster GRPs: two in the east and four in the west of the Mediterranean basin. The comparison between dendrogram and Bayesian analyses showed that when the dendrogram suggests relationships, the Bayesian method quantifies them by probabilities and proportions. The method used is applicable to microsatellite data from other crop/wild complexes (Breton et al., 2008).

Twenty-three important Ligurian olive accessions corresponding to 16 cultivars were studied using 12 SSR markers and 40 Mediterranean cultivars were included in the study in order to investigate the relationships between Ligurian and Mediterranean germplasm. All SSRs produced polymorphic amplifications. One hundred and forty-nine alleles were found in the 63 accessions analysed. Twenty-two alleles were specific to germplasm from Liguria and of these 12 were unique to single cultivars. Heterozygosity and discriminating power calculated in this regional germplasm were high (0.70 and 0.74) and not so much lower than the values in the total sample that includes cultivars from different Mediterranean countries (0.77 and 0.88 respectively). No cases of genetic identities were found between Ligurian and Mediterranean accessions. Several cases of homonyms and synonyms within the Ligurian germplasm were explained. Cluster analysis generally revealed a clear discrimination of the profiles from Liguria and Italy with respect to the cultivars from other Mediterranean countries. Only one Ligurian cultivar, “Negrea”, appeared to have a different origin, grouping with the Mediterranean cultivars. This study improved the knowledge about the Ligurian olive germplasm and highlighted the richness of olive genetic resources in small traditional areas of cultivation such as Liguria (Bracci et al., 2009).
Muzzalupo et al. (2009) used microsatellite markers (SSR) in the molecular characterization of 23 genotypes of *Olea europaea* subsp. *Europaea*. The DNA from the olive cultivars was analysed using nine preselected SSR primers (GAPU59, GAPU71A, GAPU71B, GAPU103A, UDO99-01, UDO99-12, UDO99-28 and UDO99-39) and revealed 29 alleles, which allowed each genotype to be identified. In the dendrogram, the nine primers allowed the 23 olive genotypes to be grouped into subgroups corresponding to the same cultivar denominations. SSR markers proved to be efficient and reliable for the molecular characterization of Italian olive cultivars.

Three Iranian olive cultivars, Geloleh, Shengeh and Rowghani with commercial interest are distributed in 3 provinces in North of Iran. Fifty one accessions belonging to these 3 olive cultivars were screened by 13 microsatellite markers revealing high genetic variability both within and between cultivars. In total, 54 alleles were detected with a mean number of 4.2 alleles per locus. Six unique allelic patterns were observed. Heterozygosity ranged from 0.00 to 1.00 while the mean number of polymorphic information content (PIC) was 0.51. The existence of homonyms, synonyms or mislabeling as well as intracultivar polymorphism was showed by allele differences between olive accessions studied. The phenogram obtained by UPGMA clustering showed variability among as well as between cultivars (Noormohammadi et al., 2009).

*Chionanthus* is a genus of the Oleaceae family being used as an ornamental and medicinal plant. Arias et al. (2011) have created microsatellite-enriched libraries of *C. retusus*, assembled 1072 contigs, and detected 1010 repeats. The frequency of the repeats decreased exponentially with the increase in repeat length, and the most abundant motives were: AG, AC, AAG, ACC, AT and ACTC. They have screened 384 markers on 12 *Chionanthus* related taxa, characterized 57 microsatellite loci across four species of Oleaceae and characterized 195 within the species *C. retusus*, most of these
being polymorphic. Polymorphic information content (PIC) values varied from zero to 0.85, and the percentage of heterozygous loci was in a range from 24.6% to 68.4%. The SSR markers developed here could assist in the botanical characterization for breeding programs and in the industry for the quality control and authentication of varieties of these medicinal plants.

2.5. Transferability of microsatellite markers between species in plants:

In plants transferability of microsatellites between congeneric species has been demonstrated in numerous taxa, including *Actinidia*, *Brassica*, *Camelia*, *Citrus*, *Clusia*, *Olea*, *Pinus*, *Querecus* and *Vitis*. Usually, a low percentage of markers also amplifies fragments from species belonging to other genera from the same family, as has been shown in *Asteraceae*, *Brassicaceae*, *Cucurbitaceae*, *Fabaceae*, *Fagaceae*, *Limnanthaceae*, *Mimosaceae*, *Oleaceae*, *Poaceae* and *Vitaceae* (Weising *et al.*, 2009).

Van Treuren *et al.* (1997) examined the performance of 30 microsatellites from *Arabidopsis thaliana* in two species of *Arabis*. PCR products were generated by about 50% of the primers, but alleles were generally shorter, repeat numbers were smaller and the level of intraspecific variation was lower in the nonfocal compared with the focal species. Peakall *et al.* (1998) surveyed the transferability of 31 microsatellite markers from soybean to three wild relatives of the genus *Glycine* species and to several other *Fabaceae*, but only 3 to 13% were successful in other genera. Rossetto (2001) summarized the data from a large number of studies. It was observed that 58% of microsatellite markers were polymorphic within the same family and 78% within the same subgenus.

It is yet unclear why microsatellites and their flanking DNA are relatively conserved in some taxa, but not in others. For example,
Decroocq et al. (2003) showed that transferability is higher between *Vitis* species than between *Rosaceae* species. Kahru et al. (2000) and Kutil and Williams (2001) demonstrated an unusually high conservation of primer binding sites among a number of pine species over a period of more than 140 million years.

One important determinant of the extent of marker transferability across species is the source and characteristic of the library. Thus, primer binding sites are expected to be more conserved when the microsatellite flanking sequences are maintained under selective constraints. This is most obviously the case of transcribed regions. Consequently, microsatellites within genes provide good chances to design primer pairs that are more broadly applicable, e.g., within plant families. Trinucleotide repeats are the predominant type of microsatellites in exons, and frequently have been exploited as markers. More recently, microsatellite markers have been generated from EST sequences available from public databases. EST derived microsatellite markers are generally less polymorphic than genomic microsatellites, but often show an increased level of conservation among taxa (Weising et al., 2009).

Cross species transferability of SSR markers appears to be less successful in plants as compared to animals. With the plausible exception of microsatellites residing in coding regions, the evolutionary divergence time across which microsatellites are conserved in the nuclear plant genome seems to be restricted to a maximum of 15-30 million years (Shepherd et al., 2002).

### 2.6. Linkage disequilibrium

Two loci are said to be in linkage disequilibrium when the alleles at these loci are in non-random association with each other in a population. This may happen if the two loci are physically linked by being positioned close to one another on the same chromosome so
that they do not segregate randomly during meiosis. Linkage disequilibrium may also arise as an artefact of admixture of subpopulations, the presence of kin, or founder effects.

Linkage mapping has become a routine tool for the identification of QTL in plants. This method provides a high power to detect QTL in genome-wide approaches. However, its major limitations besides high costs (Parisseaux and Bernardo, 2004) are the poor resolution in detecting QTL and that only two alleles at any given locus can be studied simultaneously with biparental crosses of inbred lines (Flint-Garcia et al., 2003). An alternative, promising approach is association mapping, which has been successfully applied in human genetics to detect QTL coding for simple as well as complex diseases (Corder et al., 1994; Kerem et al., 1989). This method uses the linkage disequilibrium (LD) between DNA markers and genes underlying traits which is present in a germplasm set. Association mapping promises to overcome the limitations of linkage mapping (Kraakman et al., 2004).

The advantages of population-based association study over conventional QTL mapping in bi-parental crosses are primarily due to

1) availability of broader genetic variations with wider background for marker-trait correlations (i.e. multiple alleles evaluated simultaneously),

2) likelihood of a higher resolution mapping because of the utilization of majority recombination events in the germplasm’s developmental history,

3) possibility of exploiting historically measured trait data for associations, and

4) development of expensive and tedious bi-parental populations is not required that makes the approach timesaving and cost-effective (Hedrick et al., 1987; Gupta et al., 2005; Oraguzie et al., 2007).
Conversely, traditional QTL mapping is 1) very costly (Hedrick et al., 1987), 2) has poor resolution with the evaluation of only a few alleles (Devlin and Risch, 1995) and 3) it requires a longer research time period. Additionally, association mapping in plants, compared to human populations, has more power with the opportunity to create mapping populations with required amount of LD and diversity (Jorde, 1995).

The measurement of the LD patterns for genomic regions and specificity of LD extent among different populations or groups of the ‘target’ organisms is the important starting point to design and conduct association mapping (Jorde, 2000; Gaut and Long, 2003). LD has been quantified in several plant species (Abdallah, 2003; Whitt and Buckler, 2003) including the model organism Arabidopsis and now extended to crops such as maize, barley, durum wheat, spring wheat, sorghum, soybean, sugarcane, sugar beet and grapevine, as well as in trees such as European aspen, and loblolly pine (Devlin and Risch, 1995; Abdallah, 2003; Whitt and Buckler, 2003). These studies revealed that the genome-wide extent of LD varied across genomes and between species with the examples of longer stretches of LD in some local populations. Moreover, LD-based association mapping was successfully used in plant germplasm resources that highlighted serious influence of population structure and relatedness of individuals in conducting association mapping (Abdallah, 2003; Whitt and Buckler, 2003).

One of the initial association studies in cotton (McRae et al., 2002) reported SSR marker associations in just 56 accessions of diploid cottons.
Large-scale genome-wide association analyses of major human diseases have yielded very promising results, corroborating findings of previous candidate gene association studies and identifying novel disease loci that were previously unknown (The Wellcome Trust Case Control Consortium, 2007). The same strategy is being exploited in many plant species thanks to the dramatic reduction in costs of genomic technologies. In contrast to the widely used linkage analysis traditional mapping research in plants, association mapping searches for functional variation in a much broader germplasm context.

Association mapping enables researchers to use modern genomic technologies to exploit natural diversity, the wealth of which is known to plant geneticists and breeders but has been utilized only on a small scale before the genomics era. Owing to the ease of producing large numbers of progenies from controlled crosses and conducting replicated trials with immortal individuals (inbreds and recombinant inbred lines, RILs), association mapping in plants may prove to be more promising than in human or animal genetics.

Association mapping includes the following steps: (1) selection of a group of individuals from a natural population or germplasm
collection with wide coverage of genetic diversity; (2) recording or measuring the phenotypic characteristics (yield, quality, tolerance, or resistance) of selected population groups, preferably, in different environments and multiple replication/trial design; (3) genotyping a mapping population individuals with available molecular markers; (4) quantification of the extent of LD of a chosen population genome using a molecular marker data; (5) assessment of the population structure (the level of genetic differentiation among groups within a sampled population individuals) and kinship (coefficient of relatedness between pairs of each individuals within a sample); and (6) based on information gained through quantification of LD and population structure, correlation of phenotypic and genotypic/haplotypic data with the application of an appropriate statistical approach that reveals “marker tags” positioned within close proximity of targeted trait of interest. Consequently, a specific gene(s) controlling a QTL of interest can be cloned using the marker tags and annotated for an exact biological function (Abdurakhmonov and Abdukarimov, 2008).

In practice, linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population. The commonly used disequilibrium measure, $r$, increases in value when specific alleles of two linked loci are found together on the same haplotype more often than expected by chance. While haplotypes are preferable for these analyses, genotypic data can also be used (Weir, 1996).

The classical definition of linkage disequilibrium (LD) refers to the non-random association of alleles between two loci or haplotypes within unrelated populations with a distinctly shared ancestry. Different quantities are used to measure LD like $D$, $r^2$, $D^2$, $D^*$, $F$, $X(2)$, $\delta$ etc. $D$ and $r^2$ are most commonly used measures of LD.

One measure of LD is $D$, calculated as (Hill, 1981),
\[ D = \text{freq}(A_1B_1)\text{freq}(A_2B_2) - \text{freq}(A_1B_2)\text{freq}(A_2B_1) \]

Where, \( \text{freq}(A_1B_1) \) is the frequency of the \( A_1B_1 \) haplotype in the population, and likewise for the other haplotypes. The \( D \) statistic is very dependent on the frequencies of the individual alleles, and so is not particularly useful for comparing the extent of LD among multiple pairs of loci (e.g. at different points along the genome). Hill and Robertson (1968) proposed a statistic, \( r^2 \), which was less dependent on allele frequencies,

\[
r^2 = \frac{D^2}{\text{freq}(A_1)\text{freq}(A_2)\text{freq}(B_1)\text{freq}(B_2)}
\]

Where \( \text{freq}(A_1) \) is the frequency of the \( A_1 \) allele in the population, and likewise for the other alleles in the population. Values of \( r^2 \) range from 0, for a pair of loci with no linkage disequilibrium between them, to 1 for a pair of loci in complete LD.

Zhao et al. (2005) recommended the \( \chi^2 \) \( (r^2) \) measure of LD for multi-allelic markers, where

\[
\chi^2 = \frac{1}{(l-1)k}\sum_{i=1}^{k}\sum_{j=1}^{m} \frac{D_{ij}^2}{\text{freq}(A_i)\text{freq}(B_j)},
\]

\[ D_{ij} = \text{freq}(A_iB_j) - \text{freq}(A_i)\text{freq}(B_j) \]

Where \( \text{freq}(A_i) \) is the frequency of the \( i \)th allele at marker A, \( \text{freq}(B_j) \) is the frequency of the \( j \)th allele at marker B, and \( l \) is the minimum of the number of alleles at marker A and marker B.

Another commonly used pair-wise measure of LD is \( D' \) (Lewontin, 1964). To calculate \( D' \), the value of \( D \) is standardized by the maximum value it can obtain:
\[ D' = \frac{|D|}{D_{\text{max}}} \]

Where, \( D_{\text{max}} = \min[\text{freq}(A1)\times\text{freq}(B2), -1\times\text{freq}(A2)\times\text{freq}(B1)] \) if \( D > 0 \),

\[ \text{else} = \min[\text{freq}(A1)\times\text{freq}(B1), -1\times\text{freq}(A2)\times\text{freq}(B2)] \] if \( D < 0 \).

The statistic \( r^2 \) is preferred over \( D' \) as a measure of the extent of LD for two reasons. If we consider the \( r^2 \) between a marker and an (unobserved) QTL, \( r^2 \) is the proportion of variation caused by the alleles at a QTL which is explained by the markers. The decline in \( r^2 \) with distance actually indicates how many markers or phenotypes are required in initial genome scan exploiting LD to detect QTL. Specifically, sample size must be increased by a factor of \( 1/r^2 \) to detect an ungenotyped QTL, compared with the sample size for testing the QTL itself (Pritchard and Przeworski, 2001). \( D' \) on the other hand does a rather poor job of predicting required marker density for a genome scan exploiting LD. The second reason for using \( r^2 \) rather than \( D' \) to measure the extent of LD is that \( D' \) tends to be inflated with small sample sizes or at low allele frequencies (McRae et al., 2002).

2.6.1. Population Structure:

The primary variable that produces genotype-phenotype misleading relationship can be due to population structure or genetic background. There are several causes of allelic association. It may be mutation, genetic drift, founder’s effect, population admixture or natural selection. All these lead to an initial association between alleles at different loci. These forces can even generate association between alleles at unlinked loci. With subsequent random mating, the association may begin to decay.

Undetected population stratification can cause spurious genotype-phenotype associations in candidate gene approaches and indirect associations via linkage disequilibrium mapping (Jorde, 1995). Further, spurious associations between genotype and
phenotype caused by population stratification must be detected among unrelated individuals in association studied, to reduce error. Clustering techniques are one approach to identify stratified populations.

2.6.2. Kinship Co-efficient (K- Matrix):

Kinship describes the probability that two homologous genes are identical by decent in a given sample. However, kinship relationships have not been considered in most plant mapping or marker assisted selection strategies. Mixed models using variance component approaches that account for kinship estimates have been exploited in animal research for over two decades (Henderson, 1984). Arbelbide et al. (2006) developed a mixed model for self pollinating plants that accounted for multiple location effects and kinship based on pedigree records.

2.6.3. Linkage disequilibrium studies in plants:

Many major crops, such as maize (Zea mays, L.), soybean (Glycine max (L.) Merr.), barley (Hordeum vulgare L.), wheat (Triticum aestivum L.), tomato (Lycopersicon esculentum Mill.), sorghum (Sorghum bicolor (L.) Moench), and potato (Solanum tuberosum L.), as well as tree species such as aspen (Populus tremula L.) and loblolly pine (Pinus taeda L.), have been studied. Association mapping studies carried out in plants is being reviewed in Table 1.

The phenotypic variation of many complex traits of agricultural or evolutionary importance is influenced by multiple quantitative trait loci (QTLs), their interaction, the environment, and the interaction between QTL and environment. Linkage analysis and association mapping are the two most commonly used tools for dissecting complex traits. Linkage analysis in plants typically localizes QTLs to 10 to 20 cM intervals because of the limited number of recombination events that occur during the construction of mapping populations and the
cost for propagating and evaluating a large number of lines (Doerge, 2002; Holland, 2007).

While hundreds of linkage analysis studies have been conducted in various plant species over the past two decades (Holland, 2007; Kearsey and Farquhar, 1998), only a limited number of identified QTLs were cloned or tagged at the gene level (Price et al., 2006). Association mapping, also known as linkage disequilibrium (LD) mapping, has emerged as a tool to resolve complex trait variation down to the sequence level by exploiting historical and evolutionary recombination events at the population level (Nordborg and Tavare, 2002; Risch and Merikangas, 1996). As a new alternative to traditional linkage analysis, association mapping offers three advantages, (i) increased mapping resolution, (ii) reduced research time, and (iii) greater allele number (Yu and Buckler, 2006). Since its introduction to plants (Thornsberry et al., 2001), association mapping has continued to gain favourability in genetic research because of advances in high throughput genomic technologies, interests in identifying novel and superior alleles, and improvements in statistical methods.

Based on the scale and focus of a particular study, association mapping generally falls into two broad categories, (i) candidate-gene association mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits; and (ii) genome-wide association mapping, or genome scan, which surveys genetic variation in the whole genome to find signals of association for various complex traits (Risch and Merikangas, 1996). While researchers interested in a specific trait or a suite of traits often exploit candidate-gene association mapping, a large consortium of researchers might choose to conduct comprehensive genome-wide analyses of various traits by testing hundreds of thousands of molecular markers distributed across the genome for association.
Table 1: Examples of association mapping studies in various plant species

<table>
<thead>
<tr>
<th>Plant species</th>
<th>Populations</th>
<th>Sample size</th>
<th>Background markers</th>
<th>Traits</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td>Diverse ecotypes</td>
<td>95</td>
<td>104 SNPs and indels</td>
<td>Flowering time</td>
<td>Olsen et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Diverse ecotypes</td>
<td>95</td>
<td>2553 SNPs</td>
<td>Disease resistance</td>
<td>Aranzana et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Diverse accessions</td>
<td>96</td>
<td>AFLP and SNPs</td>
<td>Flowering time</td>
<td>Zhao et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Diverse accessions</td>
<td>275</td>
<td>192 flowering time htSNPs</td>
<td>Shoot branching, Flowering time</td>
<td>Ehrenreich et al., 2007, 2009</td>
</tr>
<tr>
<td>Alfalfa</td>
<td>Tetraploid population</td>
<td>190</td>
<td>71 SSRs</td>
<td>yield</td>
<td>Li et al., 2011</td>
</tr>
<tr>
<td>Barley</td>
<td>Diverse cultivars</td>
<td>148</td>
<td>139 AFLP and SSRs</td>
<td>Days to heading, leaf rust, yellow dwarf virus, rachilla hair length, lodicule size, flowering time</td>
<td>Kraakman et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Diverse cultivars</td>
<td>429</td>
<td>129 SSRs</td>
<td></td>
<td>Cockram et al., 2008</td>
</tr>
<tr>
<td>Cotton</td>
<td>Germplam lines</td>
<td>335</td>
<td>202 SSRs</td>
<td>fiber quality</td>
<td>Abdurakhmonov et al., 2009</td>
</tr>
<tr>
<td>Eucalyptus</td>
<td>Unstructured natural population</td>
<td>290</td>
<td>35 SSRs</td>
<td>Microfibril angle</td>
<td>Thumma et al., 2005</td>
</tr>
<tr>
<td>Flax</td>
<td>Diverse accessions</td>
<td>60</td>
<td>118 SSRs</td>
<td>-</td>
<td>Soto-Cerda et al., 2011</td>
</tr>
<tr>
<td>Lettuce</td>
<td>Diverse accessions</td>
<td>200</td>
<td>34 ESTs</td>
<td>Disease Resistance</td>
<td>Simko et al., 2009</td>
</tr>
<tr>
<td>Plant species</td>
<td>Populations</td>
<td>Sample size</td>
<td>Background markers</td>
<td>Traits</td>
<td>References</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Maize</td>
<td>Diverse inbred lines</td>
<td>92</td>
<td>141 SSRs</td>
<td>Flowering time</td>
<td>Thornsberry et al., 2001</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines</td>
<td>102</td>
<td>47 SNPs</td>
<td>Kernel composition</td>
<td>Wilson et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Elite inbred lines</td>
<td>75</td>
<td>133 SNPs</td>
<td>Starch pasting properties</td>
<td>Palaisa et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Elite inbred lines</td>
<td>71</td>
<td>55 SSRs</td>
<td>Sweet taste</td>
<td>Andersen et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines</td>
<td>86</td>
<td>141 SSRs</td>
<td>Flowering time</td>
<td>Szalma et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines &amp; landraces</td>
<td>375 + 275</td>
<td>55 SSRs</td>
<td>Maysin synthesis</td>
<td>Camus-Kulandaivelu et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines</td>
<td>57</td>
<td>8950 isozyme</td>
<td>Flowering time</td>
<td>Tracy et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines</td>
<td>95</td>
<td>192 SNPs</td>
<td>Kernel color</td>
<td>Salvi et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Elite inbred lines</td>
<td>553</td>
<td>553</td>
<td>Oleic acid content</td>
<td>Belo et al., 2008</td>
</tr>
<tr>
<td></td>
<td>Diverse inbred lines</td>
<td>282</td>
<td>-</td>
<td>Carotenoid content</td>
<td>Harjes et al., 2008</td>
</tr>
<tr>
<td>Peanut</td>
<td>US peanut minicore collection</td>
<td>92</td>
<td>32 SSRs</td>
<td>-</td>
<td>Belamkar et al., 2011</td>
</tr>
<tr>
<td>Plant species</td>
<td>Populations</td>
<td>Sample size</td>
<td>Background markers</td>
<td>Traits</td>
<td>References</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td><em>Pinus taeda</em></td>
<td>Unstructured natural population Lines</td>
<td>32</td>
<td>21 SSRs</td>
<td>Wood specific gravity, late wood Microfibril angle, cellulose content</td>
<td>Gonzalez-Martinez <em>et al.</em>, 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>435</td>
<td>288 SSRs</td>
<td></td>
<td>Gonzalez-Martinez <em>et al.</em>, 2007</td>
</tr>
<tr>
<td>Potato</td>
<td>Diverse cultivars</td>
<td>123</td>
<td>49 NBS markers</td>
<td>Late bright resistance</td>
<td>Malosetti <em>et al.</em>, 2007</td>
</tr>
<tr>
<td>Rice</td>
<td>Diverse land races</td>
<td>105</td>
<td>SSRs</td>
<td>Glutinous phenotype</td>
<td>Olsen and Purugganan, 2002</td>
</tr>
<tr>
<td></td>
<td>Diverse land races</td>
<td>577</td>
<td>577 SNPs</td>
<td>Starch quality Yield and its components</td>
<td>Bao <em>et al.</em>, 2006</td>
</tr>
<tr>
<td></td>
<td>Diverse accessions</td>
<td>103</td>
<td>123 SSRs</td>
<td>Stigma and spikelet characteristic</td>
<td>Agrama <em>et al.</em>, 2007</td>
</tr>
<tr>
<td></td>
<td>USDA rice core collection</td>
<td>90</td>
<td>109 SSRs</td>
<td></td>
<td>Yan <em>et al.</em>, 2009</td>
</tr>
<tr>
<td>Perennial Ryegrass</td>
<td>Diverse natural germplasms</td>
<td>589</td>
<td>26 STS and SSSs</td>
<td>Heading date</td>
<td>Skøt <em>et al.</em>, 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>96</td>
<td>506 STS and SSSs</td>
<td>Flowing time, water soluble carbohydrate</td>
<td>Skøt <em>et al.</em>, 2007</td>
</tr>
<tr>
<td>Sorghum</td>
<td>Diverse inbred lines</td>
<td>377</td>
<td>47 SSRs</td>
<td>Community resource report</td>
<td>Casa <em>et al.</em>, 2008</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>Diverse clones</td>
<td>154</td>
<td>1,068 AFLP and 141 SSRs</td>
<td>Disease resistance</td>
<td>Wei <em>et al.</em>, 2006</td>
</tr>
<tr>
<td>Plant species</td>
<td>Populations</td>
<td>Sample size</td>
<td>Background markers</td>
<td>Traits</td>
<td>References</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------</td>
<td>-------------</td>
<td>---------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Tomato</td>
<td>Wild species, landraces, vintage</td>
<td>102</td>
<td>434 PCR based markers including SNPs</td>
<td>-</td>
<td>Robbins et al., 2011</td>
</tr>
<tr>
<td></td>
<td>cultivars and varieties</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>Diverse cultivars</td>
<td>95</td>
<td>95 SSRs</td>
<td>Kernel size, milling quality, kernel weight, protein content, sedimentation volume, test weight, and starch concentration</td>
<td>Breseghello and Sorrells, 2006</td>
</tr>
<tr>
<td></td>
<td>Diverse cultivars</td>
<td>207</td>
<td>115 SSRs</td>
<td></td>
<td>Reif et al., 2011</td>
</tr>
</tbody>
</table>
2.6.4. Software used for Linkage disequilibrium studies:

A variety of software packages are available for data analysis in association mapping (Table 2). TASSEL (Trait Analysis by Association, Evolution and Linkage) is the most commonly used software for association mapping in plants and is frequently updated as new methods are developed (Bradbury et al., 2007). In addition to association analysis methods (i.e., logistic regression, linear model, and mixed model), TASSEL is also used for calculation and graphical display of linkage disequilibrium statistics and browsing and importation of genotypic and phenotypic data.

STRUCTURE software typically is used to estimate Q (Pritchard et al., 2000). The Q is an $n \times p$ matrix, where $n$ is the number of individuals and $p$ is the number of defined subpopulations.

SPAGeDi (Spatial Pattern Analysis of Genetic Diversity) software is used to estimate K among individuals (Hardy and Vekemans, 2002). K is an $n \times n$ matrix with off–diagonal elements being $F_{ij}$, a marker-based estimate of probability of identity by descent. The diagonal elements of K are one for inbreds and $0.5 \times (1 + Fx)$ for non-inbred individuals, where $Fx$ is the inbreeding coefficient.

EINGENSTRAT software is used to estimate PCs of the marker data and correct test statistics resulting from population stratification (Price et al., 2006). Other software commonly used in human association mapping includes Merlin (Abecasis et al., 2002) and QTDT (Abecasis and Cookson, 2000).

SAS software (SAS Institute, 1999) or R (Ihaka and Gentleman, 1996) often is used by advanced researchers with programming skills as the platform to develop various methods. ASREML (Gilmour et al., 2002) and MTDFREML (Boldman et al., 1993) are two of several
software packages used in animal genetics in mixed model analysis of data from a very large number of individuals.
Table 2: Common statistical software packages for association mapping.

<table>
<thead>
<tr>
<th>Software package</th>
<th>Focus</th>
<th>Website</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TASSEL</td>
<td>Association analysis</td>
<td><a href="http://www.maizegenetics.net">http://www.maizegenetics.net</a></td>
<td>Free, LD statistics, sequence analysis, association mapping (logistic regression, linear model, and mixed model)</td>
</tr>
<tr>
<td>SAS</td>
<td>Generic</td>
<td><a href="http://www.sas.com">http://www.sas.com</a></td>
<td>Commercial, standard software widely used in data analysis and methodology work</td>
</tr>
<tr>
<td>R</td>
<td>Generic</td>
<td><a href="http://www.r-project.org/">http://www.r-project.org/</a></td>
<td>Free, convenient for simulation work for researches with good programming and statistics background</td>
</tr>
<tr>
<td>STRUCTURE</td>
<td>Population structure</td>
<td><a href="http://pritch.bsd.uchicago.edu/structure.html">http://pritch.bsd.uchicago.edu/structure.html</a></td>
<td>Free, widely used for population structure analysis</td>
</tr>
<tr>
<td>EINGENSTRAT</td>
<td>PCA, association analysis</td>
<td><a href="http://genepath.med.harvard.edu/~reich/Software.htm">http://genepath.med.harvard.edu/~reich/Software.htm</a></td>
<td>Free, PCA was proposed as an alternative for population structure analysis</td>
</tr>
<tr>
<td>MTDFREML</td>
<td>Mixed model</td>
<td><a href="http://aipl.arsusda.gov/curtvt/mtdfreml.html">http://aipl.arsusda.gov/curtvt/mtdfreml.html</a></td>
<td>Free, mixed model analysis for animal breeding data, also can be used for plant data</td>
</tr>
<tr>
<td>Software package</td>
<td>Focus</td>
<td>Website</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ASREML</td>
<td>Mixed model</td>
<td><a href="http://www.vsni.co.uk/products/asreml">http://www.vsni.co.uk/products/asreml</a></td>
<td>Commercial, mixed model analysis for animal breeding data, also can be used for plant data</td>
</tr>
</tbody>
</table>

(Zhu et al., 2008)