CHAPTER II

LITERATURE REVIEW

2.1 Family Fagaceae

Fagaceae is botanically described as follows (Chengjiu et al., 1999). *Trees or rarely shrubs, monoecious, evergreen or deciduous. Stipules usually early deciduous. Leaves alternate, sometimes false-whorled in Cyclobalanopsis. Inflorescences unisexual or androgynous with female cupules at the base of an otherwise male inflorescence. Male inflorescences a pendulous head or erect or pendulous catkin, sometimes branched; flowers in dense cymules. Male flower: sepals 4-6(-9), scalelike, connate or distinct; petals absent; filaments filiform; anthers dorsifixed or versatile, opening by longitudinal slits; with or without a rudimentary pistil. Female inflorescences of 1-7 or more flowers subtended individually or collectively by a cupule formed from numerous fused bracts, arranged individually or in small groups along an axis or at base of an androgynous inflorescence or on a separate axis. Female flower: perianth 1-7 or more; pistil 1; ovary inferior, 3-6(-9)-loculed; style and carpels as many as locules; placentation axile; ovules 2 per locule. Fruit a nut. Seed usually solitary by abortion (but may be more than 1 in Castanea, Castanopsis, Fagus, and Formanodendron), without endosperm; embryo large.* This family includes 7 - 12 genera and 600 - 1000 species distributing worldwide except for tropical and southern Africa, dominating forests in the temperate, seasonally dry regions of the Northern Hemisphere, with a centre of diversity found in tropical Southeast Asia, particularly at the generic level (Soepadmo, 1972; Scoggan, 1978; Mabberley, 1997; Chengjiu et al., 1999, Manos et al., 2001). Many species are important timber trees. Nuts of *Fagus, Castanea*, and of most *Castanopsis* species are edible, and oil is extracted from nuts of *Fagus* (Chengjiu et al., 1999). The most economically important genus is *Quercus*, on account of its valuable timbers. Commercial cork is obtained mostly from the cork oak, *Quercus suber* Linn. Sweet

chestnuts are the edible nuts of the Spanish chestnut, *Castanea sativa* Mill., and the American species *Castanea dentate* (Marsh) Borkh., which has been ravaged by a blight disease. The Japanese chestnut, *Castanea crenata* Sieb. & Zucc., seems to be resistant to this disease. Seeds of the beech *Fagus sylvatica* Linn. are edible, and in some regions are an important food for pigs (Hutchinson, 1964).

In Thailand, this family comprises four genera (Smitinand, 1980; Gardner et al., 2000; Forman, 1964; Nixon and Crepet, 1989): *Castanopsis* (D. Don) Spach. (chestnuts), *Lithocarpus* Blume (stone oaks), *Quercus* L (oaks), and *Trigonobalanus* Forman. *Castanopsis* is evergreen tree, includes about 120 species distributing in tropical and subtropical Asia. *Lithocarpus* is evergreen tree, rarely shrub, and includes about 300 species distributing mainly in Asia. *Quercus* is evergreen or deciduous tree or shrub, and includes about 300 – 600 species distributing in northern Africa, Asia, Europe, North America, and South America (Soepadmo, 1972, Chengjiu et al., 1999). *Trigonobalanus* includes 3 species; the type species *T. verticillata* Forman from Celebes, Borneo and Malaya, *T. doichangensis* (Camus) Forman from northern Thailand, and *T. excelsa* Laxano from Columbia, South America (Forman, 1964; Nixon and Crepet, 1989). The habitats of Fagaceae are relatively diverse – while being most common in Hill evergreen forests they are also found in Dry dipterocarp forests at lower altitudes and Mixed deciduous forests with open grassland (Gardner et al., 2000).

Species diversity of Fagaceae in Thailand is substantial, as to be expected from this region of South East Asia. The many local names (over 30) and diverse habitats reflect high species diversity. Three new species and a new variety have recently been described during the work on revision of Fagaceae for the Flora of Thailand (Phengklai, 2004). In practice, identification of species in the oak group is known to be difficult, due to many reasons including extensive intraspecific morphological variation, ecological adaptation, clinal differentiation, hybridisation and/or introgression. These phenomena have been well documented especially in oaks (*Quercus*) of Europe and North America (Schwarz, 1964; Scoggan, 1978;

Dumolin-Lapegue et al., 1997; Montalvo et al., 1997; Dow and Ashley, 1998; Streiff et al., 1998; Bellarosa et al., 2005; Gonzalez-Rodriguez et al., 2005). The levels of genetic diversity within *Quercus* of Eastern Asia also appear to be relatively high, due to hybridisation and gene flow mediated by long-distance wind pollination (Huang et al., 2002; Kanno et al., 2004; Chung et al., 2005). According to Flora Europaea (Schwarz, 1964), most species within each subgenus of *Quercus* are interfertile and hybrids are therefore common in regions where related species grow together, and much of the intraspecific variation is due to introgressive hybridisation. The genus *Quercus* is known to be difficult to taxonomists, and this is reflected in a different number of species recognized by different authors. Linnaeus proposed 12 species and since then the number of oak taxa has increased to several hundreds of separate species, plus many forms, varieties and subspecies. For example, two commercially important white oak species of Europe, *Q. robur* and *Q. patraea*, are occasionally proposed to be combined into a single species *Q. robur*, but generally they are accepted as separate "morphological-ecophysiological" species. Biologically they are definitely not separate because they are interfertile, as most oak species are. With this pair of species, compiling evidence based on biochemical (isozymes) and molecular (DNA) markers support that any differences, allelic or genomic, are too small for differentiation at species level (Gomory et al., 2001; Muir et al., 2001; Zoldos et al., 2001). Another example of taxonomical discrepancy in Fagaceae: a Eurasian beech (*Fagus*) could be identified correctly using combined evidence from nuclear genes, morphology and the fossil record (Denk et al., 2002). Species in other Fagaceae genera have also been subjected to taxonomic revision based on new evidence. For example, the three species of *Trigonobalanus* were segregated into three monotypic genera based on fossil evidence and phylogenetic analysis, thus proposing two new genera *Formanodendron* and *Colombobalanus* (Nixon and Crepet, 1989). As new methods for identifying genetic and evolutionary relationships among species or populations, especially molecular markers, have become available, certain taxonomic questions can now be answered and complex processes that have shaped the species diversity can be explained.

2.2 Molecular methods in genetic diversity studies

The first molecular markers to be established were allozymes. The principle of allozyme markers is that protein variants in enzymes can be distinguished by gel electrophoresis according to differences in size and changes caused by amino-acid substitutions. To visualize the allozyme bands, the electrophoretic gels are treated with enzyme-specific stains that contain substrate for the enzyme, cofactors and an oxidized salt (Schaal and Olsen, 2000; Schlotterer, 2004). Allozymes have been used as molecular (biochemical) markers for studying genetic variation within species as well as its spatial distribution, and estimating levels of gene flow among populations (Newton et al., 1999; Schaal and Olsen, 2000; Schlotterer, 2004). The early studies of genetic structure and diversity in oaks were primarily based on isozymes or allozyme variations. Spatial genetic structure of *Q. acutissima* Carruth in an undisturbed forest landscape on Oenaro Island in southern Korea was investigated using multilocus allozyme genotypes, and the study indicated gene flow among populations (Chung et al., 2002). A fine-scale genetic structure of oak forests in Europe were studied using isozymes, or together with other markers (Berg and Hamrick, 1995; Streiff et al., 1998; Finkeldey, 2001; Gomory et al., 2001). Well-documented enzyme systems can provide unequivocal measures of allele frequencies, and therefore they serve as valuable indicators of gene transfer within and among populations.

The use of allozymes has led to more than thirty years of insight into how plant populations evolve. However, inferring population structure solely from allele frequencies has its limitation. In a review of 322 allozyme studies of woody taxa, Hamrick et al. (1992) concluded that tree species maintain more intraspecific variation than other life forms but generally display less variation between populations, particularly in the case of widespread species with outcrossing systems and wind- or animal-dispersed seeds. Such life history and ecological traits only accounted for one third of the variation between species for the genetic parameters considered, indicating that much of this variation remained unexplained. To explore this huge gap, there is clearly a need for better markers than the allozyme systems.

Recent years have witnessed the development of an array of molecular tools that can be used to assess DNA variation directly, most notably by refinements of the PCR (polymerase chain reaction) technique. But the first of molecular methods, so called restriction fragment length polymorphism (RFLP), was developed in the early eighties following the discovery and isolation of restriction endonucleases (Botstein et al., 1980; reviewed in Schlotterer, 2004). The principle of RFLP is that total genomic DNA is digested with restriction enzymes and the resultant fragments are separated by gel electrophoresis, after which they are probed with specific DNA fragments by Southern blotting and hybridisation. RFLP allows DNA variation to be assayed because single base substitutions in the recognition sequence of a restriction enzyme can change the pattern of resulting restriction fragments. As RFLPs are co-dominant markers, they are highly applicable in map-based plant improvement programs and are reliable markers for characterising genetic diversity of a given species or population (Karp et al., 1998). However, the greatest obstacle for a wider range of applications of RFLP analysis is the requirement for suitable hybridisation probes to detect the polymorphisms.

The discovery of PCR technology (Mullis, 1990) has made molecular studies much more powerful, simple and practical. Only small amount of DNA sample is required (nanograms, as for all PCR methods), thus allowing accessibility to a wider range of materials and can include a much larger sample size suitable for investigating genetic structure at population levels. A number of PCR-based methods have been developed and are proven useful for assessing genetic variation, biodiversity and genetic studies of populations (Karp et al., 1998; Newton et al., 1999; Tienderen et al., 2002; Zhang and Hewitt, 2003). Two types of markers can be generated: co-dominant, single and known markers (e.g. PCR-RFLP, SSR) and dominant multilocus markers (e.g. RAPD, ISSR, AFLP). The conventional RFLP approach has been made more simple by PCR amplifying particular genes and the amplified fragments, which are digested with restriction enzymes, can be analyzed directly without Southern blotting and hybridisation. Phenetic or cladistic analysis can also be performed after sequencing this type of fragments, thus making the method highly sensitive. This

PCR-RFLP approach, with or without sequencing, has been most often applied to study polymorphism in the chloroplast genomes of plants, although mitochondrial or nuclear genes are also examined. The approach based on PCR-RFLP of chloroplast (cp) genes has been used extensively to reveal genetic variation in oaks and other members of Fagaceae, and as chloroplast genes are mainly maternally inherited in this plant group the studies can infer long-range migration of species such as northward postglacial migration in Europe (Ferris et al., 1995; Dumolin-Lapegue et al., 1997; Petit et al., 1997; Taberlet et al., 1998; Csaikl et al., 2002). In recent years, SSR markers (microsatellites or simple sequence repeats) from the nuclear or organelle genomes of Fagaceae have become more available and the studies have therefore added significantly to the understanding of genetic diversity and population structures of this plant group (Dow and Ashley, 1996; Streiff et al., 1998; Hornero et al., 2001; Yamamoto et al., 2003; Blakesley et al., 2004). Using appropriate molecular techniques, especially the PCR-RFLP, to a variety of tree species has highlighted a far higher degree of population differentiation than indicated by previous isozyme analyzes, a result consistent with theoretical predictions (Newton et al., 1999).

Dominant multilocus markers, on the other hand, have been popular in genetic diversity studies, due to the simplicity of some of the methods. They are particularly useful in studying species that have received very little attention for being non-crop species, as the methods generally do not require prior sequence information. The method so called random amplified polymorphic DNA (RAPD, Williams et al., 1990), in particular, was developed initially for mapping of genomes of soybean and other crop plants. RAPD method uses arbitrary 10 base-primers to amplify DNA randomly, probably over the whole genome. It generates very high polymorphism, hence suitable for fingerprinting cultivars, clones and genotypes (e.g. Sigurdsson et al., 1995). The RAPD method has been used successfully in characterising genetic variation and phylogeography of tree species, for example Costa Rican trees including cedar (Gillies et al., 1997; Schierenbeck et al., 1997) and Pan-Arctic juniper (Adams et al., 2003). A more targeted PCR method based on semi-arbitrary primers containing repetition of very short (1-4 bp) sequences, called inter-simple sequence

repeats (ISSR, Zietkiewicz et al., 1994), can detect polymorphisms in the microsatellites and the flanking regions. This fingerprinting method has been used successfully to characterize genetic diversity of several tree species, such as Chilean *Nothofagus* (Mationi et al., 2002). The method called amplified fragment length polymorphism (AFLP; Vos et al., 1995) combines restriction enzyme digestion and PCR technology in such a way that a single genotype can reveal 50 – 100 polymorphic AFLP fragments, hence a very large dataset for marker mapping or phenetic analysis. The AFLP method used with a large-scale field screening of *Fusarium*-fungal disease resistance in wheat has successfully identified a major QTL (quantitative trait loci) that has potential of making wheat crop much more profitable (Guo et al., 2003). The same approach has been used to identify QTLs for leaf blight resistance in rubber trees (Lespinasse et al., 2000). The AFLP method has also been used to characterize inter- and intraspecific variation in tree species, for example Indonesian oil palm (Purba et al., 2000) and Asian chestnuts (Yamamoto et al., 1998). Dominant multilocus markers, i.e. RAPD, ISSR and AFLP, have not been applied to study phylogeographic and genetic variation in Fagaceae effectively. As these polymorphisms can be too extensive at the individual level, they may not reflect higher levels of differentiation or evolutionary history of the species. Nevertheless, these markers can be more suitable in situations where current population structure is more accessible and where the molecular facilities can only support small-scaled studies (Newton et al., 1999). In addition, the identification of management units of conservation, such as those obtained using these multilocus markers, can be of practical importance for informing the transfer of germplasm within and between regions, as can occur in forest reforestation or restoration activities.

The genetic diversity can be assessed using markers directly targeted at specific genes, gene families or particular sequence classes, some of which could infer changes in functional or ecologically adaptive traits (Tienderen et al., 2002). Site-targeted PCR markers include (1) single nucleotide polymorphisms (SNP) of selected genes involved in key processes of plant metabolism (Tienderen et al., 2002), (2) the use of conserved sequence motifs as anchors for sequence-specific amplification

polymorphisms (SSAP; Waugh et al., 1997), (3) IRAP and REMAP retrotransposon-based DNA fingerprinting techniques (Kalendar et al., 1999; Leigh et al., 2003), and (4) internal transcribed spacer regions (ITS) of the nuclear ribosomal genes (Alvarez and Wendel 2003) or other genes. Sequenced data from ITS regions of the nuclear rDNA and ITSs of chloroplast genes revealed a large-scale phylogeography of *Lithocarpus* of SE Asia (Cannon and Manos, 2003). Site-targeted PCR markers will undoubtedly be the tools for diversity studies in the future as they provide the possibility of ecological inference, and together with the microarray technique new and important genes can be identified.

2.3 Molecular cytogenetics in genome studies

The first period of cytogenetics started when W.S. Sutton published his paper "The chromosomes in heredity" in 1903 (Tagarelli et al., 2003). The conclusion from this and Sutton's other publications is that chromosomes contain all the genetic information, that chromosomes have individuality, that they occur in pairs, with one member of each pair contributed by each parent, and that the paired chromosomes separate from each other during meiosis (Tagarelli et al., 2003). Since then, a number of importance researches in cytogenetics have been published, for example the demonstration that cytological recombination of marked chromosomes correlated with recombination of gene (Creighton and McClintock, 1931) and the discovery that normal human cells contain 46 chromosomes (Tjio and Levan, 1956), which lead to the beginning of human cytogenetic diagnostics (Hulten, 2002). The discovery of genetic codes by Watson and Crick (1953) has opened up a new discipline called molecular cytogenetics, whereby a single gene can be visualised in a microscope.

Cytogenetics, the discipline concerning genetic implications of chromosome structure and behavior, is often associated with mitosis and meiosis when chromosomes are visible as individual entities. These studies have resulted in numerous karyotypes and laid the basis for our present knowledge and understanding of chromosome dynamics in relation to mitosis and meiosis. Cytogenetics has been

applied to identify chromosomal rearrangements such as deletions and translocations, and map these aberrations onto the linear chromosome relative to centromeres, telomeres and chromosome bands (Tessadori et al., 2004). In plants, cytogenetics has been used primarily as tools for cytotaxonomy, as plant genomes are extremely variable. A very large part of this variation is due to hybridisation and polyploidy in plants (Leitch and Bennett, 1997; Brochmann et al., 2004; Adams and Wendel, 2005). The difference in chromosome numbers alone is often sufficient to separate species, because such difference will create meiotic pairing abnormality and infertility in the hybrids, and as a consequence restricting gene flow between species.

Plant metaphase spreads are usually prepared from meristematic cells in actively growing root tips using conventional squash-spread techniques (Schwarzacher and Heslop-Harrison, 2000). This protocol is not applicable for some trees species that do not produce roots readily from cuttings. Searching root tips in the forest ground can be an impossible task. Chromosomes from germinating seeds may not represent the same genomic constitution as their mother trees, due to introgressive hybridisation (Anamthawat-Jónsson, 2003a). Recently a method has been developed (Anamthawat-Jónsson, 2003b), whereby chromosomes can be isolated directly from leaf tissues by protoplast dropping technique similar to that described for root tips of plants (Busch et al., 1996). This method is based on enzymatic digestion of young leaf tissues (shoot tips) after which the resulting protoplasts are treated hypotonically before being dropped onto microscopic slides. This protocol is highly effective for studying cytogenetics and population genetics of plants in the field, trees in particular. It is also an excellent method for studying small chromosomes as the method produces well spread metaphases suitable for counting. The chromosome preparations have very high resolution for karyotyping purpose, as well as having been proven suitable for *in situ* hybridisation experiments.

Microscopic identification of individual chromosomes in a chromosome complement has long been based on morphological characteristics like arm length, centromere position, nucleolus organizing region at the secondary constriction and

heterochromatin pattern. With the introduction of chromosome banding techniques in the seventies, more diagnostic tools became available to discriminate individual chromosomes. For example, the technique of C-banding (Sumner et al., 1971) can be effectively applied for use with plant species with long chromosomes like cereal species including barley and rye (Anamthawat-Jónsson and Heslop-Harrison, 1990; Linde-Laursen et al., 1992). However, many plant species have karyotypes characterized by rather uniform chromosome shapes and banding patterns, or the shapes have little resolution especially if the metaphase chromosomes are very small. In such cases, recognition of individual chromosome pairs is difficult (Schubert et al., 2001). Moreover, natural hybridisation is a relatively common feature of vascular plant species and has been demonstrated to have played an important role in their evolution (Ellstrand et al., 1996; Rieseberg, 1997). Most flowering plants are polyploidy, and many are distinct in combining the diploid nuclear genomes from two or more different ancestral species or genera (allopolyploidy). Polyploidy has played a significant role in the evolution of higher plants (Leitch and Bennett, 1997). The highly powerful tools than classical cytogenetics are necessary for understanding genome organization in plant.

One of the most powerful tools for studying plant genomes is molecular cytogenetics, the discipline which links molecular hybridisation and cytogenetics technique together. The advent of DNA: DNA *in situ* hybridisation protocols opened the possibility to map chromatin regions of individual chromosomes on the basis of DNA sequence information. The method of *in situ* molecular hybridisation principally uses probe sequence tagged with radioisotope or a chemical reporter such as biotin. The initial step is denaturation of the target chromosomal DNA mostly in metaphase, to facilitate access of the probe to the target. This is followed by hybridisation with the probe, while complementary sequences undergo pairing. The hybridized sites are localized either through autoradiography or immunofluorescence, depending on the type of probe used. Discovery of various chemical reporter molecules, which can be tagged with the probes, and final detection through the appropriate fluorescing compound, has become a fascinating technique, widely

pursued in various laboratories (Sharma and Sharma, 2001; Schubert et al., 2001). DNA *in situ* hybridisation to chromosomal targets has become almost synonymous with fluorescent *in situ* hybridisation (FISH). This technique can be used for monitoring physical mapping of sequences to their chromosomal location, the identification and characterisation of chromosomes, or chromosome segments, and to provide indicator of recent or evolutionary rearrangements in the genome (Schwarzacher and Heslop-Harrison, 2000; Sharma and Sharma, 2001; Schubert et al., 2001). The development of genomic *in situ* hybridisation (GISH), a modification of the FISH, has provided new insights into the origin and evolution of polyploid genomes. The technique uses total genomic DNA from one species as labelled probe in hybridisation experiments to chromosomal DNA *in situ*, and sometimes together with unlabelled DNA from a related species to block common sequences between the labeled probe, the unlabelled block and the chromosomal targets (Anamthawat-Jónsson et al., 1990). GISH allows chromosomes from different parents, ancestors or genomes in hybrid plants, in allopolyploids or introgressed breeding lines to be painted (labeled) in different colours (Schwazacher et al., 1989; Anamthawat-Jónsson et al., 1996; Leitch and Bennett, 1997; Raina and Rani, 2001; Kato et al., 2005; Schubert et al., 2001; Amanthawat-Jonsson, 2001; Sharma and Sharma, 2001; Tessadori et al., 2004).

Advances in plant molecular cytogenetics have occurred essentially from studying large monocot chromosomes. For tree species, especially those with small minute chromosomes, the cytogenetic studies seldom reach beyond chromosome counting. For the family Fagaceae, most chromosome data have come from studying some species of *Quercus* in Europe. Other Fagaceae genera have not been examined cytogenetically. All species of oak (*Quercus*) examined so far have diploid chromosome 2n = 24, and having very similar karyotypes. The basic chromosome number of this genus is x = 12 (Schwarz, 1964; Ohri and Ahuja, 1990; Zoldos et al., 1999). Due to very similar genome sizes, similar karyotypes and heterochromatin organization in oak species, molecular cytogenetic techniques were used to study their genome organization. Zoldos et al. (1999) examined genome organization of

ribosomal gene and heterochromatin in eleven *Quercus* species. The fluorochrome banding and the FISH pattern have revealed a striking similarity among different species with regards to genome size and organization. The genomic differentiation between *Q. robur* and *Q. suber* were examined by representational difference analysis (Zoldos et al., 2001). Genome size and ploidy stability of some oak species were studied by flow cytometry. Their genome sizes of eight *Quercus* species ranged from 1.88 to 2.00 pc/2C. The DNA content did not vary among population of *Q. robur* but was different in some populations of *Q. petraea* (Zoldos et al., 1998). The ploidy stability of *Q. suber* is maintained among embryos, and between the embryos and their mother plants (Loreiro et al. 2005). As oaks are rich in tannins, this may have interfered with the flow sorting and the estimation of DNA content (Zoldos et al., 1998; Loreiro et al., 2005). Genome organization in the family Fagaceae has been examined only in the genus *Quercus* in Europe, and it is therefore important to have an insight into the chromosomes and genomes of other genera and other species. The study of karyotype and genome organization in taxonomically related species may be useful in understanding the phylogeny and evolution in this family.