CHAPTER II

BACKGROUND

2.1 Morphological characteristics of tobacco

The tobacco plant (*Nicotiana tabacum* L.) is of the genus *Nicotiana,* one of the large genera of the family Solanaceae. Tobacco is therefore a close relative of eggplant, pepper, petunia, potato, tomato and nightshade (Arslan and Okumus, 2006). The origins of *N. tabacum* are obscure. It is not known from the wild and appears to be a hybrid plant, deliberately selected by humans a long time ago. Indeed, *N. tabacum* is an amphidiploid species (2n=48) resulting from an interspecific cross between two wild forms, *N. sylvestris* (2n=24) and *N. tomentosiformis* (2n=24) (Ren and Timko, 2001; Yukawa et al., 2005). Domesticated tobacco has been cultivated for several thousand years and was widely used by the indigenous peoples in the Americas for medicinal and ceremonial purposes (Ren and Timko, 2001). It can be used as an organic pesticide in the form of nicotine tartrate and also used in some medicines. In consumption it may be in the form of smoking, chewing, snuffing or dipping tobacco. Moreover, tobacco plants are also used in plant bioengineering while some of the 60 *Nicotiana* species are grown as ornamentals (Brücher, 1989).

About morphological characteristics of tobacco, *N. tabacum* is a robust annual little branched herb up to 2.5 metres high (Figure 2.1 A) with large green leaves and long trumpet shaped white–pinkish flowers. All parts are sticky, covered with short viscid-glandular hairs which exude a yellow secretion containing nicotine (Brücher, 1989). The tobacco plant produces between 10 and 20 broad leaves with very varied in size of the leaves (Figure 2.1 B). The lower leaves are the largest up to 60 cm long with shortly stalked or unstalked. The shape of these lower-leaves is oblonge-elliptic with shortly acuminate at the apex and decurrent at the base. The following leaves decrease in size. The upper ones are sessile and smallest with oblong-lanceolate or elliptic shapes (Brücher, 1989). The plant terminal has many flowered inflorescenses (Figure 2.1 C). The floral tubes are 5-6 cm long and 5 mm in diameter with yellowish-white

coloured. The calyx and throat are expanded and have 5 narrowly triangular lobes. The floral lobes are broadly triangular. The corolla is white-pinkish with pale violet or carmine coloured at tips. The persistent calyx and a short apical beak, about 2 cm long, have a capsular ovoid or ellipsoid surrounded. Seeds are numerous and very small (Figure 2.1 D), with ovoid or kidney shaped and brown (Albert, 1996).

Figure 2.1 Morphological characteristics of tobacco. (A) Tobacco plants. (B) A tobacco leaf. (C) Tobacco flowers. (D) Tobacco seeds.

2.2 History of tobacco

Tobacco grows natively in North and South America, and it is believed that tobacco began growing in the Americas about 6,000 B.C. (Davis and Nielsen, 1999). Archaeological discoveries date the earliest use of tobacco to around the 1st century B.C. when the American Indians or the Maya, a highly cultured people in Central America, smoked the tobacco leaf in sacred and religious ceremonies and began using tobacco in many different ways, such as in medicinal practices (Davis and Nielsen, 1999; Grehan, 2006). In 1492, Christopher Columbus encountered a native man on San Salvador, at Cuba, with dry tobacco leaves and was first to note the use of tobacco to cure many illness. Soon after that, sailors brought tobacco back to Europe and the plant was being grown all over Europe (Davis and Nielsen, 1999). Tobacco spread to other European countries and then to Asia. It soon arrived in the Middle East and Africa where it was welcomed with the same enthusiasm with which coffee had been greeted in the 17th century (Grehan, 2006). In Asia, tobacco was first introduced to the Philippines and then to India, China and Indonesia, respectively (Albert, 1996).

Within 150 years of Columbus's finding "strange leaves" in the new world, tobacco was being used around the globe. Its rapid spread and widespread acceptance characterise the global addiction to the plant. In the 18th century, the mode of delivery of tobacco changed to snuff held sway. The 19th century was the age of the cigar and the rise of the manufactured cigarette was seen in the 20th century, bringing a greatly increased number of smokers, as shown in Figure 2.2 (Mackay and Eriksen, 2002; Burns, 2007). During the 20th century, tobacco has become one of the most economically important agricultural crops in the world. Indeed, farmers in over 100 developed and developing nations (Figure 2.3) rely upon tobacco production to provide a stable source of cash income. The processing, manufacture and distribution of tobacco products provide employment for thousands of other people throughout the globe (Davis and Nielsen, 1999).

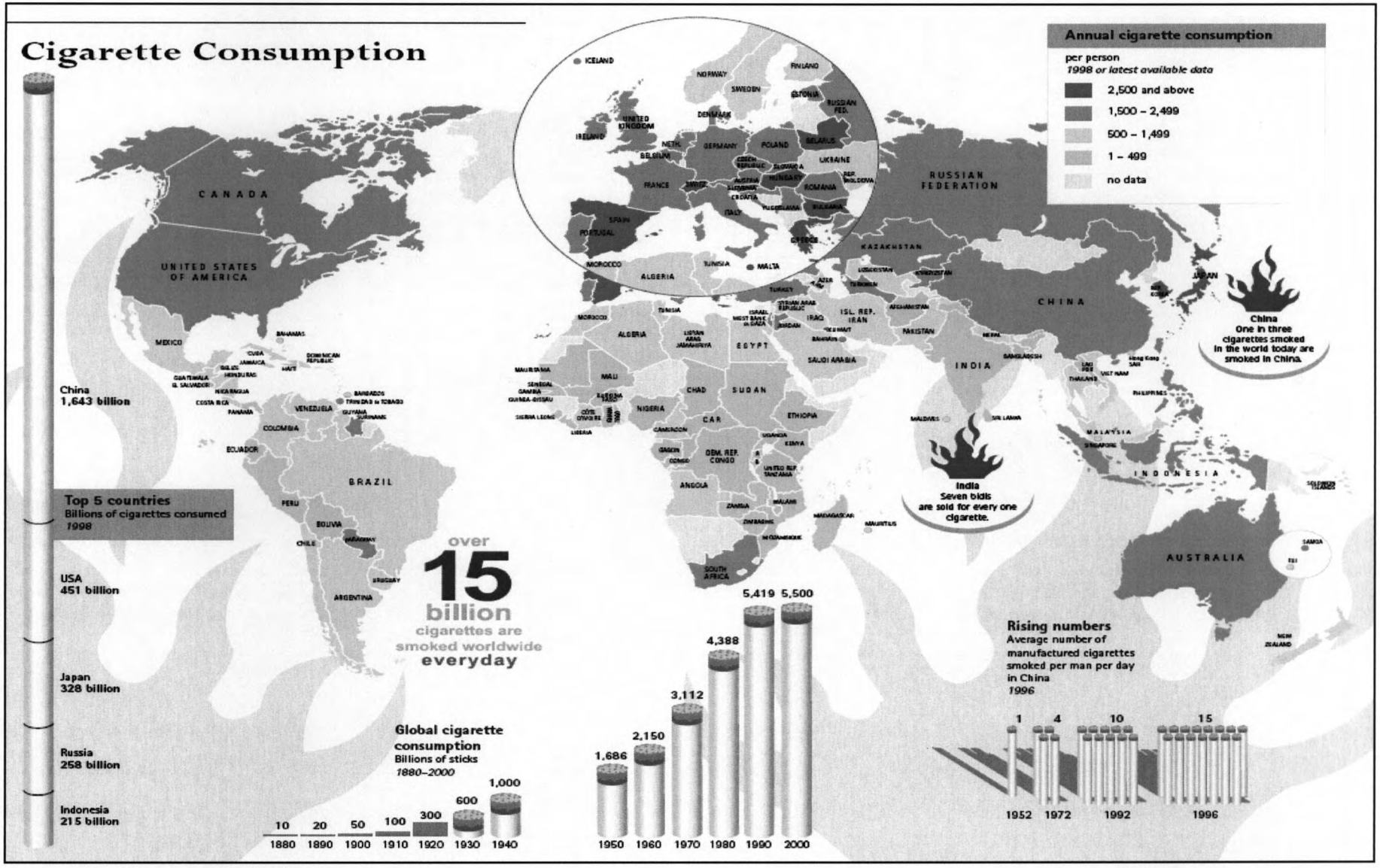


Figure 2.2 Global consumption of cigarettes has been rising steadily (Mackay and Eriksen, 2002).

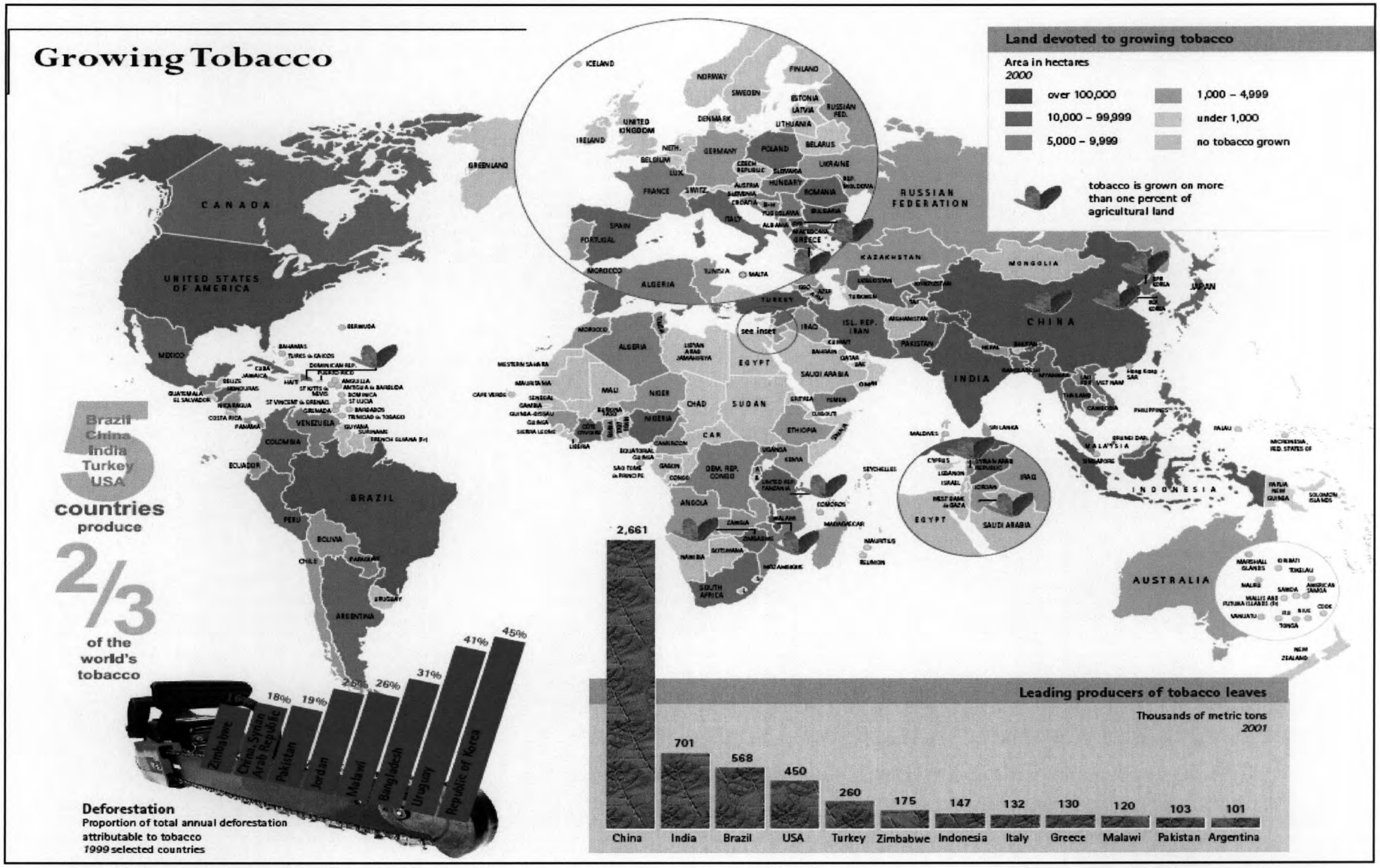


Figure 2.3 Growing tobacco plants around the world (Mackay and Eriksen, 2002).

2.3 Tobacco cultivation

Tobacco is cultivated annually, and can be harvested in several ways. Tobacco cultivation is similar to other agricultural products. At first, tobacco seed was usually scattered onto the soil. Today, however, tobacco is sown in cold frames or hotbeds, as their germination is activated by light (Tobacco Research Board, 1981). After the plants have reached relative maturity about 15 cm in height, they are transplanted into the fields in which a relatively large hole is created in the tilled earth with a tobacco peg (Walker, 1980; Walker, 1981). Transplants should not be too large as the number of harvestable leaves may be reduced as a result of early flower initiation. It is important that most of the root system is retained at pulling and the entire plant maintained in a turgid condition after pulling to transplanting (McCants and Woltz, 1967). Seedlings need to be placed into moist soil in the field with the terminal bud at least 2 to 3 cm above ground level.

When planting under hot and dry conditions, seedlings should be hardened by withholding water at least 14 days before planting. Hardened seedlings have more starch in the stem and roots and have greater dry mass and root volume than non-hardened seedlings, although excessive hardening may result in a lower starch concentration. The enhanced growth effect is not measurable by about 6 weeks after transplanting, and the yield and grade index of cured leaves are not affected (Tobacco Research Board, 1983; 1985; 1986; 1987; 1990). Under climatic conditions that are suitable for premature flowering, i.e. short daylength and/or low temperatures, the removal of portions of leaves by clipping significantly increases the number of harvestable leaves and yield of tobacco (Pfeiffer et al., 1990). As the industrial revolution took hold, harvesting wagons used to transport leaves were equipped with man-powered stringers, an apparatus which used twine to attach leaves to a pole. In modern times large tobacco fields are harvested by a single piece of farm equipment, although topping of the flowers (topping always refers to the removal of the tobacco flower before the leaves are systematically removed and, eventually, entirely harvested) and in some cases plucking of immature leaves are still done by hand.

2.4 Types of tobacco

Tobacco types are usually categorised according to methods of curing (Akehurst, 1981; Goodman, 1993). The latter distinction refers to colour as well as quality. Table 2.1 below shows the main methods of curing and types referred to in the text.

Table 2.1 Types of tobacco are classified by the main of curing methods (Akehurst, 1981).

Type	Method of curing
Burley	Air-cured
Maryland	Air-cured
Cigar	Air-cured
Oriental/Turkish	Sun-cured
Virginia	Fire-cured
Bright/Virginia	Flue-cured

- **Burley**, this tobacco type first appeared as a mutant in 1864 in a tobacco field in Ohio. It has a very light nature, blends well with other tobacco types and has very high absorbent characteristics. First used in chewing tobacco, it is now the essential filler of the American blend cigarette, accounting for one-third of its composition.

- **Virginia**, this type also referred to as Bright inside the United States. It was developed in the 19th century, in North Carolina, together with the process of flue-curing. It is light-bodied, bright yellow in colour and relatively low in nicotine. Almost all of this tobacco is destined for cigarette production. The terms 'Bright' and 'flue-cured' are synonymous. The 'Virginia' cigarette is composed of 100 percentages (%) flue-cured tobacco.

- **Turkish**, this type is also known as Oriental. It produces a very aromatic product, and is used primarily in American blend cigarettes, accounting for about 15% of the tobacco mixture. It is mostly grown in the eastern Mediterranean and Black Sea coastal areas, its traditional home. The distinctiveness of Turkish tobacco is the result of specific environmental conditions and the chemical changes caused by sun-curing on the leaf.

2.5 Tobacco products

The tobacco consumptions have rapidly spread around the world. Cigarettes account for the largest share of manufactured tobacco products in the world, 96% of total sales. Except for chewing tobacco in India and smoking of kreteks in Indonesia, cigarettes are the most common method of consuming tobacco throughout the world. The invention of the cigarette-rolling machine in 1881 accelerated the tobacco pandemic by mass-producing pocket sized packets of cigarettes. Unlike tediously hand-rolled cigarettes and bulky water pipes, manufactured cigarettes offer a convenient and portable method to maintain addiction, even while driving a motor vehicle, working in a factory, or taking a stroll (Mackay and Eriksen, 2002).

There are many products manufactured wholly or partly from tobacco and intended for use by smoking, inhalation, chewing, sniffing or sucking (Mackay and Eriksen, 2002).

1) Smoking tobacco

1.1) **Bidis** or **Beedi** (Figure 2.4 A) consist of a small amount of sun-dried, flaked tobacco, hand-wrapped in dried temburni or tendu leaf (*Diospyros* sp.) and tied with string. Despite their small size, bidis deliver more tar and carbon monoxide than manufactured cigarettes because users are forced to puff harder to keep bidis lit. Bidis are found throughout South Asia and are the most heavily consumed smoked tobacco products in India.

1.2) **Cigars** (Figure 2.4 B) are tightly rolled bundles of air-cured and fermented tobacco which a tobacco-leaf wrapper. The long aging and fermentation process produces high concentrations of carcinogenic compounds that are released on combustion. Cigars come in many shapes and sizes.

1.3) **Cigarettes** (Figure 2.4 C) are the most commonly consumed tobacco products worldwide. They consist of shredded or reconstituted tobacco, processed with hundreds of chemicals and rolled into a paper-wrapped cylinder. Usually tipped with a cellulose acetate filter, they are lit at one end and inhaled through the other.

1.4) Gutka is a preparation of crushed betel nut, tobacco, and sweet or savory flavorings. It is manufactured in India and exported to a few other countries. It is sold across India in small, individual-size packets.

1.5) Pipe smoking (Figure 2.4 D) typically consists of a small chamber (the bowl) for the combustion of the tobacco to be smoked and a thin stem (shank) that ends in a mouthpiece (the bit). Shredded pieces of tobacco are placed into the chamber and ignited.

1.6) Water pipe (Figure 2.4 E) also called hookah, shisha, narghile or hubble-bubble, is a single or multi-stemmed (often glass-based) pipe and is operated by water filtration and indirect heat. Flavored tobacco is burned in a smoking bowl covered with foil and coal. The smoke is cooled by filtration through a basin of water and consumed through a hose and mouthpiece. Originally from India, the water pipe has gained immense popularity, especially in the middle east.

1.7) Kreteks (Figure 2.4 F) are cigarettes made with a complex blend of tobacco, cloves and a flavoring "sauce". They are widely smoked in Indonesia. They may contain a wide range of exotic flavorings and eugenol, which has an anesthetic effect, allowing for deeper and more harmful smoke inhalation.

1.8) Roll-your-own (RYO) cigarettes (Figure 2.4 G) are cigarettes hand-filled by the smoker from fine-cut, loose tobacco rolled in a cigarette paper. RYO cigarette smokers are exposed to high concentrations of tobacco particulates, tar, nicotine and tobacco-specific nitrosamines (TSNAs), and are at increased risk for developing cancers of the mouth, pharynx, larynx, lung and esophagus.

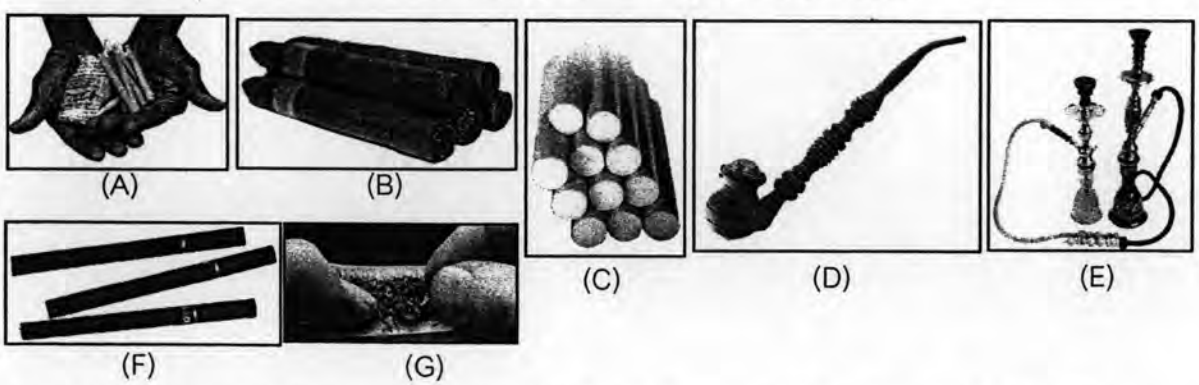


Figure 2.4 The products of smoking tobaccos. (A) Bidis. (B) Cigars. (C) Cigarettes. (D) A pipe. (E) Water pipes. (F) Kreteks. (G) Roll-your-own cigarette.

(source: www.tobaccoatlas.org/typesoftobacco.html)

2) Smokeless tobacco

2.1) Chewing tobacco (Figure 2.5 A) is oral smokeless tobacco products placed in the mouth, cheek, or inner lip and sucked (dipped) or chewed. Tobacco pastes or powders are similarly used, placed on the gums or teeth. Sometimes it is referred to as "spit tobacco" because users spit out the built-up tobacco juices and saliva. This mode of tobacco consumption became associated with American baseball players during the 20th century.

2.2) Snuff is a generic term for fine-ground smokeless tobacco products. Originally the term was referred only to dry snuff, a fine tan dust popular mainly in the 18th century. Snuff powder originated in the UK town of Great Harwood and was famously ground in the town's monument prior to local distribution and transport further up north to Scotland. There are three major varieties which include European (dry), India (creamy) and American (moist); although American snuff is often referred to as dipping tobacco.

- **Dry snuff** (Figure 2.5 B) is powdered tobacco that is inhaled through the nose and absorbed through the nasal mucosa or taken orally. Once widespread, particularly in Europe, the use of dry snuff is in decline.

- **Creamy snuff** is tobacco paste, consisting of tobacco, clove oil, glycerin, spearmint, menthol and camphor. It is sold in a toothpaste tube and marketed mainly to women in India.

- **Moist snuff** (Figure 2.5 C) is a small amount of ground tobacco held in the mouth between the cheek and gum. Manufacturers are increasingly prepackaging moist snuff into small papers or cloth packets to make the product more convenient.

2.3) Dipping tobacco is a form of smokeless tobacco. Dip is occasionally referred to as "chew", and because of this, it is commonly confused with chewing tobacco, which encompasses a wider range of products. A small clump of dip is 'pinched' out of the tin and placed between the lower or upper lip and gums.

2.4) Snus is moist powder tobacco product that is consumed by placing it under the upper lip for extended periods of time. It is a form of snuff that is used in a

manner similar to American dipping tobacco, but typically does not result in the need for spitting.

(A) (B) (C)

Figure 2.5 The products of smokeless tobaccos. (A) Chewing tobacco. (B) Dry snuff. (C) Moist snuff. (source: www.tobaccoatlas.org/typesoftobacco.html)

2.6 Other non-smoking uses for tobacco

Not everyone growing tobacco would smoke it. Tobacco plants have rich green foliage which makes it a very attractive plant addition to any garden. Most varieties also have very nice flowers which provide a brilliant contrast to their large leaves. Of particular value as ornamental garden plants are the rose-tobacco, and jasmine-tobacco which grow abundant flowers (Brücher, 1989). Additionally, many organic gardeners use it to create pesticides for their gardens that are completely biodegradable and very effective. There are specific varieties best suited for this like mountain-tobacco which also makes good filler for other herbal pesticides. It is produced by boiling strong tobacco in water, or by steeping the tobacco in water for a longer period. When cooled, the mixture can be applied as a spray, or 'painted' on to the leaves of garden plants, where it is deadly to insects.

Moreover, tobacco also has a long history of use by medical herbalists as a relaxant, though since it is a highly additive drug it is seldom employed internally or externally at present (Chevallier, 1996). The leaves act as antispasmodics, discutients, diuretics, emetics, expectorants, irritants, sedatives and sialagogues (Weiner, 1980). They are used externally in the treatment of rheumatic swelling, skin diseases and scorpion stings (Chopra, 1986). The plant should be used with great caution because it is addictive narcotic when taken internally. The active ingredients can also be absorbed

through the skin. Wet tobacco leaves can be applied to stings in order to relieve the pain (Weiner, 1980). They are also a certain cure for painful piles. A homeopathic remedy is made from the dried leaves. It is used in the treatment of nausea and travel sickness (Castro, 1990).

The regions that have histories of use of *N. tabacum* include (Chevallier, 1996):

1) **Brazil**: where the leaves are heated and the juice is squeezed out, mixed with ash from bark of *Theobroma subircanum* or other *Theobroma* species to make an intoxicating snuff. The leaf juice is taken orally to induce vomiting and narcosis.

2) **Colombia**: Fresh leaf is used as poultice over boils and infected wounds; the leaves are crushed with oil from palms and used as hair treatment to prevent baldness.

3) **Cuba**: Extract of the leaf is taken orally to treat dysmenorrheal.

4) **East Africa**: Dried leaves of *N. tabacum* and *Securinega virosa* are mixed into a paste and used externally to destroy worms in sores.

5) **Ecuador**: Leaf juice is used for indisposition, chills and snake bites and to treat pulmonary ailments.

6) **Fiji**: Fresh root is taken orally for asthma and indigestion. Fresh root is applied ophthalmically as drops for bloodshot eyes and other problems; seed is taken orally for rheumatism and to treat hoarsness.

7) **Guatemala**: Leaves are applied externally by adults for myasis, headache and wounds. Hot water extract of the dried leaf is applied externally for ring worms, fungal diseases of the skin, wounds, ulcers, bruises, sores, mouth lesions, stomartitis and mucosa. Leaf is orally taken for kidney diseases.

8) **Haiti**: Decoction of dried leaf is taken orally for bronchitis and pneumonia.

9) **India**: Juice of *Securinega leucopyrus* is mixed with the dried leaves of *N. tabacum* and applied externally for parasites. Fresh leaves are mixed with corn-cob or *Amorphophallus paeonifolium* to treat asthma.

10) **Iran**: Infusion of the dried leaf is applied externally as an insect repellent; ointments made from crushed leaves are used for baldness, dermatitis and infectious ulceration and as a pediculicide.

11) United States: Extract of *N. tabacum* is taken orally to treat tiredness, ward off diseases and quiet fear.

12) Tanzania: Leaves of *N. tabacum* are placed in the vagina to stimulate labor.

2.7 Tobacco cultivars growing in Thailand

Tobacco possibly first arrived in Thailand in the late 16th century, although there is no strong evidence when it was imported and by whom. There is only the writing of Damrong Rajanubhab, H.R.H. Prince of Siam, in the Journal of the Siam Society during 1862-1943. He mentioned the chronicles of Monsieur De La Loubere, a French envoy to Ayutthaya kingdom or Siam (Thailand) in 1668 of the King Narai era. Such chronicles said that Monsieur De La Loubere found both Thai men and women already smoking Ya-chun, or strong tobacco, in every place. The strong tobaccos were imported from Manila (in Philippines archipelago) and China, and some of them were also grown in the country (ประกิต และ กรองจิต, 2547; วรวิชย์ และคณะ, 2549). In 1917 of the King Chulalongkorn era, the first cigarette company held by British owner was founded in Thailand. This company first produced rolled tobaccos only by hand. Soon after that, a blunt-bottom cigarette was invented by Singhanart Ratchadurongrit, a royal relative of the King Mongkut. Cigarette-producing machines were soon imported from Germany in the King Rama VI era and resulted in widespread cigarette smoking across the country (ประกิต และ กรองจิต, 2547).

In 1933, Maejo Experiment Station was founded in Chiang Mai province, northern of Thailand, to serve as research and development posts for tobacco leaves. The station was conducted by Phra Chuang Kashetra who planted the first two imported tobacco cultivars (Joiner and White Burley cultivars) over there (สงวน, 2549). Thereafter, imported Virginia cultivars were firstly introduced into Thailand by British American Tobacco (B.A.T) company in 1935. Soon after that, the Royal Thai Government bought the tobacco-production company from Burapha Yasup limited partnership and founded the Thailand Tobacco Monopoly in 1939, under Department of Excise, Ministry of Finance. After that, the Government brought all other tobacco-operations of Guanghok

company, Hoffun company and British American Tobacco (B.A.T) company. These tobacco companies were also merged together into Thailand Tobacco Monopoly (ประกิต และ กรองจิต, 2547; วรวิชย์ และคณะ, 2549). Moreover, several regional tobacco stations were set up around the country to control and buy tobacco leaves from local farmers (วรวิชย์ และคณะ, 2549).

Nowadays, tobacco is one of Thailand's important economic crops. These plants are mostly grown in the northern region and also in the South, the Northeast and the Central. Tobacco cultivars grown in Thailand can be separated to two major groups: local cultivars and imported cultivars. The latter is further separated to three minor groups: Virginia, Turkish and Burley (วรวิชย์ และคณะ, 2549).

2.7.1 Imported tobacco cultivars

Since 1939 Thailand Tobacco Monopoly has imported finest tobacco seeds from many foreign tobacco companies for experiment and development of tobacco cultivation in Thailand (วรวิชย์ และคณะ, 2549). Such imported tobacco seeds are regularly sent to their regional stations for germination. The staff of each station then distributes the plant seedlings to local farmers and also provides knowledge to the farmers step-by-step: from growing techniques to harvest and curing processes. Moreover, Thailand Tobacco Monopoly also provides financial support and other else to their contact farmers.

The imported tobacco cultivars are usually classified to three cultivar groups - Virginia, Burley and Turkish - by different curing methods and tobacco barn designs. Curing is the process for drying freshly harvested tobacco with partially or fully controlled temperature and moisture schedules that preserve the inherent quality of tobacco leaves during drying (Davis and Nielsen, 1999).

2.7.1.1 Virginia cultivar group (or flue-cured tobacco)

Virginia cultivar, also called Bright or flue-cured tobacco, derives its name from the unique curing process used to produce lemon to orange coloured cured

leaves (Figure 2.6 A) with high sugar contents and smooth smoking property (Davis and Nielsen, 1999). The total world tobacco production is classified as approximately 80% of the flue-cured tobacco, which is the major raw material for use in blended cigarettes (Stevens et al., 1996). Cigarettes made totally or primarily with flue-cured tobacco are quite popular in China, Europe and Asia (Davis and Nielsen, 1999). This Virginia cultivar-group represents 90% of tobacco cultivation in Thailand. These cultivars are mostly grown in the upper northern region (for example, Chiang Rai, Chiang Mai, Lumpang, Lamphun, Phrae and Nan provinces) and the northeastern region (Nong Khai and Nakhon Phanom) of the country (Figure 2.6 B). Popular cultivars of Virginia tobaccos grown in Thailand are Coker187 Hicks, Coker347, Coker411 and K326 (วรวิชย์ และคณะ, 2549).

Flue-curing for leaves of Virginia cultivars occurs in tightly constructed convention-barns (Figure 2.6 C) with artificial heat beginning at about 35°C and ending at about 75°C over a five-to-seven day period for completion. About 97% of the moisture is removed from the harvested leaf (Peele et al., 1995). Nowadays, the flue-curing technology for Virginia cultivars is improved by constructing closed curing-bulks (Figure 2.6 D) instead of the curing barn. The heat source of the curing-chambers is controlled by hot-steam flue for efficient quality and quantity of the cured tobaccos.

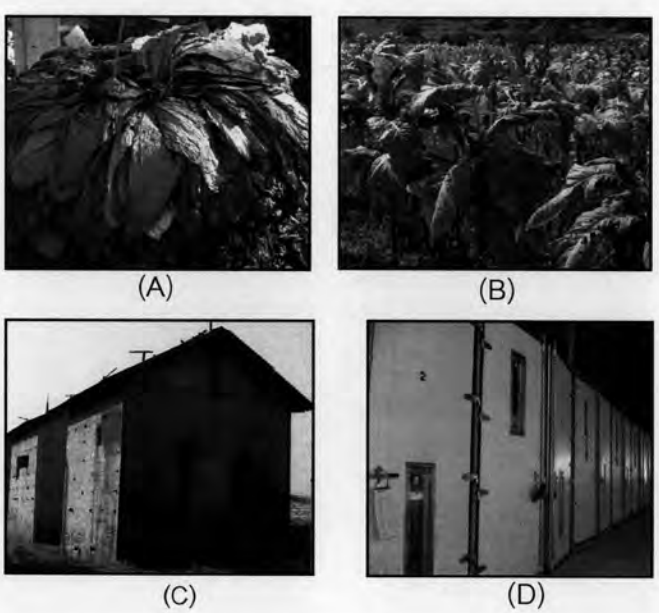


Figure 2.6 Virginia tobacco cultivars. (A) Dried Virginia leaves after passed flue-curing process. (B) Virginia tobacco cultivars grown in Chiang Rai province. (C) Flue-curing convention barns. (D) Flue-curing bulks.

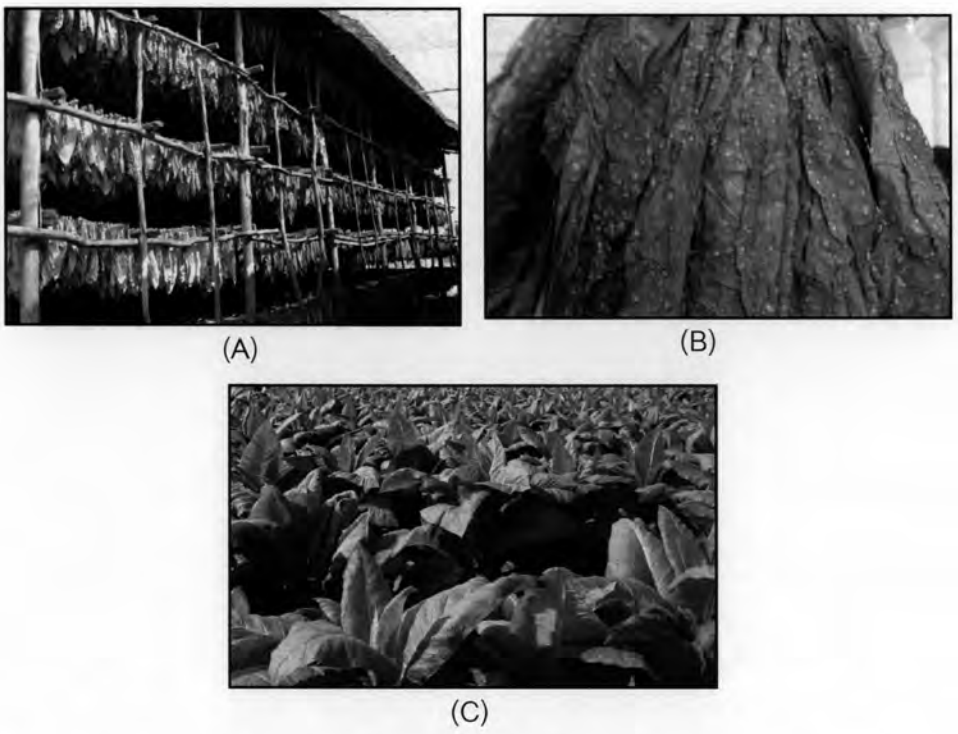
2.7.1.2 Burley cultivar group (or light air-cured tobacco)

Most Burley-type tobacco also becomes a component in the manufacturing of blended cigarettes, though inferior to Virginia cultivars. Light air-cured tobacco is common in many locations all over the world. Production practices vary depending on local weather patterns, socio-economic status of the growers and the degree of industrialisation of the region (Davis and Nielsen, 1999). Even in developed countries, light air-cured tobacco production is a labor-intensive enterprise when compared with other crops. The common factor is that curing is primarily without artificial sources of heat and humidity. Because air-cured tobacco is not cured under a prescribed set of temperatures and humidities, the end product may differ considerably from one location to another and from year-to-year (Walton and Henson, 1971; Palmer and Calvert, 1993).

In curing method, Burley tobaccos are air-cured by being hung in well-ventilated barns (Figure 2.7 A) and allowed to dry over a period of four-to-eight weeks (Davis and Nielsen, 1999). This curing process is slowly carried on under virtually natural conditions in which the temperature, humidity, air flow and sunlight can be carefully controlled (Jeffrey, 1940; Jeffrey, 1946). There are three stages in the curing process. The first stage, or the green stage, is short in duration, lasting for two-to-five days. During this stage the chlorophyll degrades, giving way to the second, the yellow stage. Soon after that, tobacco leaf colour is changed to yellow. The yellow pigments of the second curing stage degrade more slowly than the chlorophyll. Then the leaf is changed to medium brown in the final stage (Figure 2.7 B), follows as the yellow pigment degrades and the leaf dies (Jeffrey, 1940; Massie and Smiley, 1974; McKee and Conrad, 1994). These cured tobaccos produce low in sugar with sweet flavour and high nicotine (Davis and Nielsen, 1999).

In Thailand, Burley cultivars are mostly grown in the lower-northern region (for example, Sukhothai and Phetchabun provinces) (Figure 2.7 C). This Burley cultivar-group represents 5% of tobacco cultivation in Thailand, and almost all of 15-45% tobacco leaves to be processed into cigarettes are also of Burley cultivar group.

Examples of Burley cultivars usually grown in Thailand are Burley21, Burley37, KY14 and B1 special (วรวิชย์ และคณะ, 2549).



(A) (B)

(C)

Figure 2.7 Burley tobacco cultivars. (A) Burley leaves hung in an air-curing barn. (B) Dried Burley leaves after passed air-curing process. (C) Burley tobacco cultivars growing in the crop field in Phetchabun province.

2.7.1.3 Turkish cultivar group (or oriental tobacco)

Turkish cultivars (Figure 2.8 A), or oriental tobaccos, generally have very short leaves, two or three inches wide (Figure 2.8 B). The small plant- and leaf-size characteristics of today's Oriental tobacco cultivar, as well as its unique aromatic property, is a result of the plant's adaptation to poor soil and stressful climatic conditions in which it developed over the early 17^{th} centuries (Davis and Nielsen, 1999). Commercial interest in Turkish tobacco began in the early 20^{th} century when Turkish cigarettes became widely popular in Europe (Wolfe, 1962). Cultivation of Turkish tobacco today is still concentrated in the geographical regions where it originated: near the coastal areas of the eastern Mediterranean, Aegean, Marmara and Black Sea. In addition to these types, Turkish cultivars are also grown in Italy, Albania, Syria, Iran,

Lebanon, the southern republics of the Former Soviet Union and Thailand. In Thailand, these cultivars are mostly grown in the Northeast (for example, Roi Et, Maha Sarakham, Nakhon Phanom, Khon Kaen and Kalasin provinces) (วรวิชย์ และคณะ, 2549). Their leaves contain high aroma and sugar, but low nicotine.

Once sufficiently wilted, the tobaccos are hung directly in the sun for curing (Figure 2.8 C). There are a number of various techniques for this process, each unique to the specific cultivar and location. In general terms, the strings of leaves are hung out on racks of varied construction which are situated in such a way that they receive direct sunlight during the day and are protected from harsh winds (Wolfe, 1962). These racks are usually built just far enough off the ground so that the draped leaves do not touch the ground or stones below, but can still benefit from the heat reflected off the ground. At night the racks are either covered with nylon sheets or the strung leaves are moved to a covered area to protect them from the cooler night air and morning dew. The whole process takes from 12 to 30 days (Davis and Nielsen, 1999). Once curing is completed, the tobaccos are dry and very brittle and must be handled with care to prevent breakage and loss. The sun's direct heat fixes the leaves at a yellow to orange colour (Figure 2.8 D) resulting from the high sugar content in the leaves. The best known cultivars of Turkish tobaccos are Samsun, Basma and Izmir (Murray, 2008).

(A) (B) (C) (D)

Figure 2.8 Turkish tobacco cultivars. (A) Turkish tobacco cultivars grown in Nakhon Phanom province. (B) A leaf of Turkish cultivar. (C) Leaves hung out on racks and drying in the sun. (D) Dried Turkish leaves after sun-curing process.

2.7.2 Local cultivar group (or air-and-sun cured tobacco)

Local tobaccos usually have high stalks with large leaves drooping off the main stem, similar to Virginia and Burley imported cultivars. Local cultivars are grown in many cultivation crops around Thailand (Figure 2.9 A) such as the northern region (for example, Phrae, Phayao and Nan provinces), the southern region (Nakhon Si Thammarat), the central region (Kanchanaburi, Suphan Buri and Lop Buri) and the northeastern region (Nakhon Phanom, Nong Khai and Ubon Ratchathani). Tobacco farmers usually cultivate the plants in lightly sandy loams near riverside areas. During cultivation of the local tobaccos, the terminal buds are picked, an operation known as topping method, in order to keep the leaf strength (Figure 2.9 B).

The air-and-sun curing process for local cultivars includes three steps. First of all, the leaves are air-cured for about four-to-five days after harvested, then cut into one-to-two inch strips by hand or a cutting-machine, and sun-dried for about two days on wooden racks (Figure 2.9 C). Local tobaccos are usually used to make "roll-your-own (RYO) tobacco", a dried local tobacco product (Figure 2.9 D). Examples of popular cultivars of the local tobaccos in Thailand are Hangkai (in Phayao province), E-dum (Phetchabun), Petkhangsink (Sukhothai) and Kan (Suphan Buri) (วรวิชย์ และคณะ, 2549).

(A) (B) (C) (D)

Figure 2.9 Local tobacco cultivars. (A) Local cultivar grown in the crop field in Nakhon Phanom province. (B) The terminal bud of local cultivar is picked. (C) Local tobaccos are sun-dried on wooden racks. (C) Roll-your-own (RYO) tobacco after air-curing processed.

2.8 Molecular markers in plant

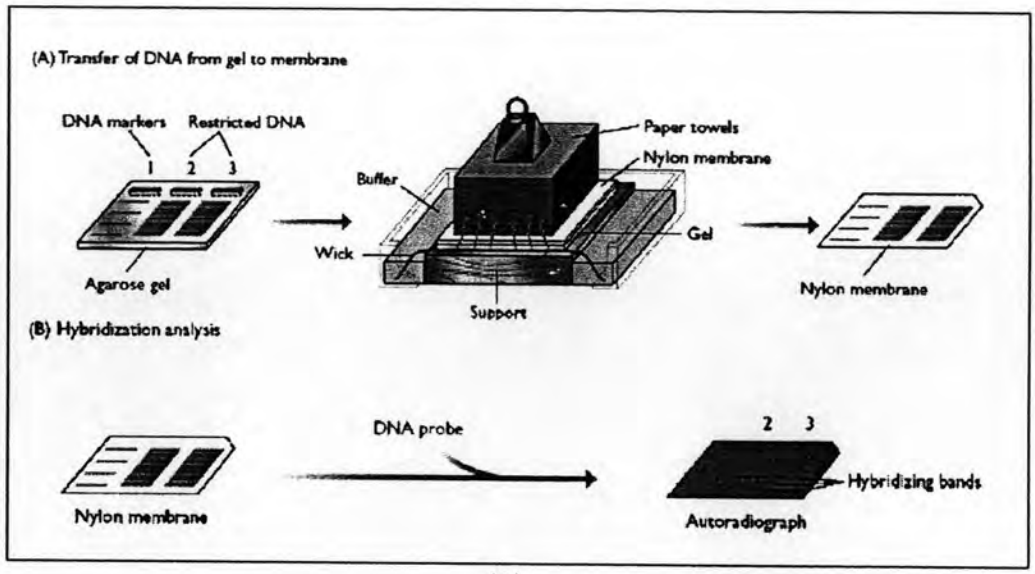
Recently, analyses of genomic DNA using molecular biology techniques have been increasing used to solve difficulty in plant genetic studies. Precise genotypic identification and assessment of genetic diversity germplasms are now very important for plant breeding purpose (Kuras et al., 2004). Plant breeding is based around the identification and utilisation of genetic variation. A major cost and logistical issue in plant breeding are the actual number of lines that need to be carried through the evaluation and selection phases of a program. However, its field trials can be expensive and evaluation of some traits, such as quality and yield stability can be expensive to assess. From such difficulties molecular markers have introduced and proved to be a powerful tool in improving the conventional plant breeding method and there are now many examples available to show the efficacy of such markers (Langridge and Chalmers, 2004).

Molecular markers have been taken, in recent years, to refer to assays that allow the detection of specific sequence differences between two or more individuals. Isoenzyme and other protein-based marker systems represent molecular markers and were in wide use long before DNA markers became popular (Langridge and Chalmers, 2004). However, DNA-based markers are more powerful tools for unbiased estimation of genetic evaluation and breeding, identification of cultivars, phylogenetic evaluation and study of genetic diversity (Souframanien and Gopalakrishna, 2004). The greatest advantage of molecular marker techniques is their capability to detect genetic diversity at a higher level of resolution than other conventional methods. The DNA-based assays are usually accurate, speedy and not effected by environmental conditions. Furthermore, molecular information can be obtained from little amounts of plant materials (Sergio and Gianni, 2005).

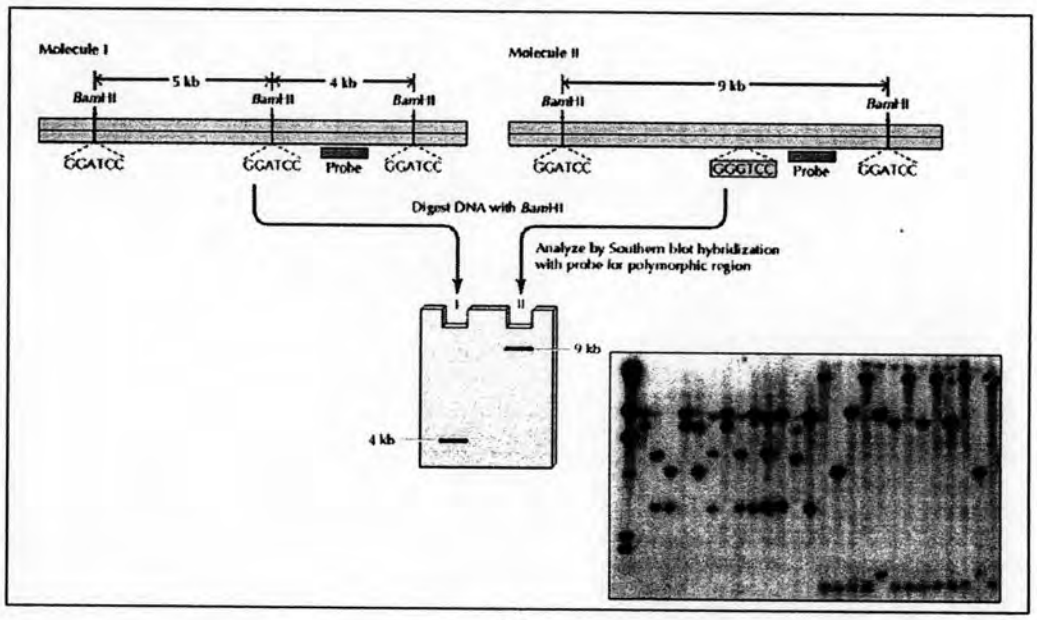
2.8.1 DNA fragment markers

Different types of molecular markers using DNA fragments, such as Restriction Fragment Length Polymorphism (RFLP) (Botstein et al., 1980), Random Amplified Polymorphic DNA (RAPD) (Williams et al., 1990), Amplified Fragment Length Polymorphism (AFLP) (Vos et al., 1995) and Simple Sequence Repeat (SSR) (Powell et al., 1996) have been developed and increasingly used as modern techniques to distinguish genotypes of plants. These DNA fragment markers have been successfully used in polymorphism analysis, cultivar identification and phylogenetic evaluation in many crop plants. Each marker technique has its own advantages and disadvantages. Common benefits from most markers include rapid analyses, highly informative results and being independent on environmental factors.

One of the earliest types of DNA-based molecular markers, RFLPs (Figure 2.10), is widely used to indirectly detect genetic variation at the DNA level. It examines size variation of specific DNA fragments after digested with a specific endonuclease. Such size variation is made visible by Southern hybridization with a labeled probe (Botstein et al., 1980; Tanksley et al., 1989). RFLP markers had been utilised as a basis method for most works in plant genetic studies for many years. However, the process requires too much amount of DNA and is too impractical to be of great use. The requirement of radioactive isotope makes the analysis even relatively expensive and hazardous (Tanksley et al., 1989; Davis and Nielsen, 1999). Moreover, the assay is time-consuming and labour-intensive and only one out of several markers may be polymorphic, which is highly inconvenient especially for crosses between closely-related species. Their inability to detect single base changes restricts their use in detecting point mutations occurring within the regions at which they are detecting polymorphism (Tanksley et al., 1989).



(A)



(B)

Figure 2.10 Restriction Fragment Length Polymorphism (RFLP) technique. (A) A Southern hybridization method. (B) A principle of RFLP and an example of RFLP band pattern. (Source: http://www.clas.ufl.edu/users/magitz/notes/Lect9-Molecular_markers.pdf).

In 1985, a polymerase chain reaction (PCR) technique was introduced and revolutionised molecular marker methods (Saiki et al., 1985). The PCR-based markers involve an *in vitro* amplification of a particular DNA sequence with specific or arbitrary primers. PCR basically amplifies DNA template sequence by primers of

complementary DNA strands to high copy numbers (Weising et al., 1995). The amplified products are then separated by electrophoresis.

The basic protocol of PCR (Saiki et al., 1985) is simple (Figure 2.11): (1) double-stranded DNA is denatured at high temperature to form single strand (template); (2) short oligonucleotide primers bind at a lower annealing temperature to the single strand complementary templates; (3) the temperature is raised for synthesis of targeting sequences by primer extension; and (4) the newly synthesised double-strand DNA target sequence are denatured at high temperature, and the cycle is repeated. The amplified target DNA can be increased exponentially as every cycle has the potential to double the amount of target DNA from the previous cycle if there is sufficient amount of polymerase, primers and nucleotides in the solution.

PCR is the powerful technique in plant molecular genetic and systematic studies because sometimes the source of DNA was from herbarium specimens ranged quality from normal to several degraded (Doyle et al., 1995).

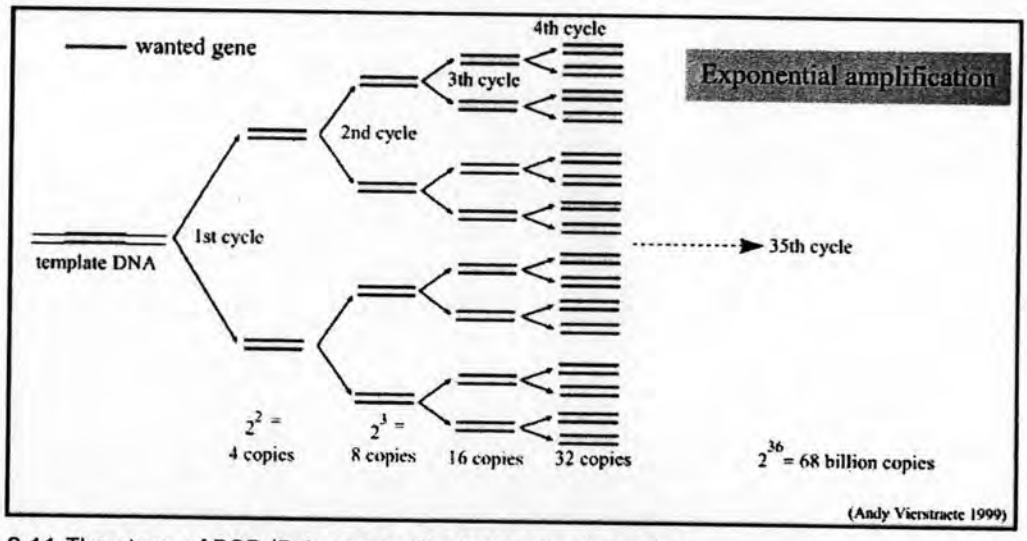


Figure 2.11 The steps of PCR (Polymerase Chain Reaction) reaction.
(Source: http://users.ugent.be/~avierstr/principles/pcrcopies).

With the development of PCR techniques, molecular marker systems based on PCR have emerged as major molecular tools for various genetic analyses. PCR-based DNA markers have been very useful in assessment of biodiversity, study of plant populations and crop cultivar identification. The most common techniques of these PCR-based systems are RAPD, AFLP and SSR. Each marker technique has its own advantages and disadvantages.

In 1990, William et al. reported a novel identification technique based on random amplification of genomic DNA sequences by PCR using single, short arbitrary primers (usually ten base-pairs). This technique is commonly called RAPD-PCR (Figure 2.12). RAPD priming sites are thought to be randomly distributed throughout a genome and the DNA polymorphisms in these priming regions result in differing amplified products. The application of RAPDs and their related modified markers in variability analysis and individual-specific genotyping has largely been carried out (Morgan et al., 1993; Mbwana et al., 2006). Some other advantages of this technique include no knowledge of the plant's DNA needed, its ability to scan and detect extensive polymorphism, its simplicity and rapidity, no radioactivity used, and very small amounts of genomic DNA required. However, it is less popular due to problems such as poor reproducibility, faint or fuzzy products over time as well as between laboratories, and difficulty in scoring bands which lead to inappropriate inferences (Thormann et al., 1994; Arcade et al., 2000).

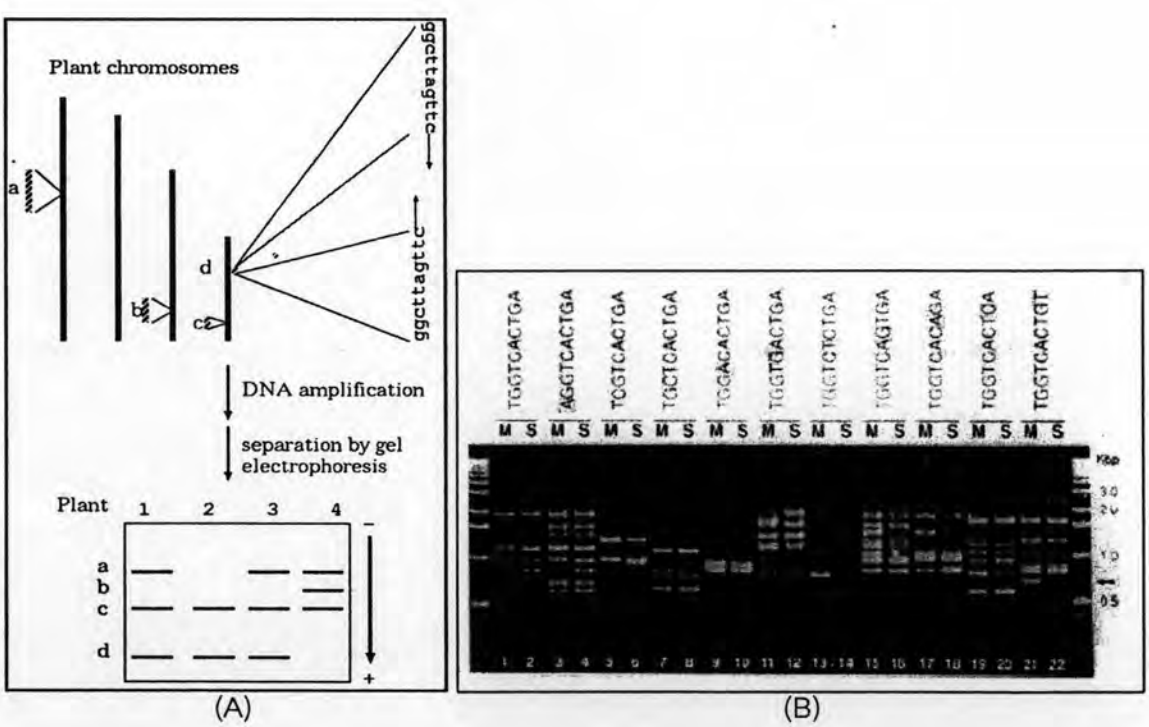
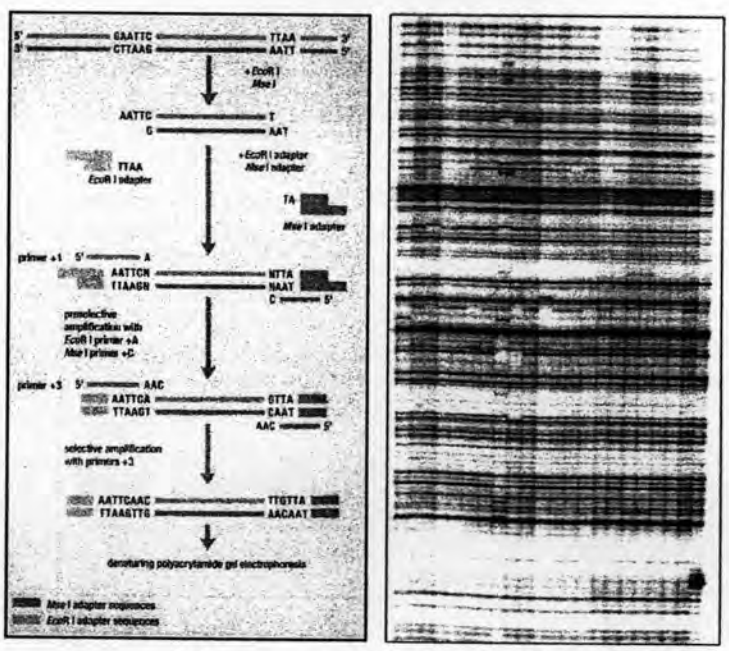


Figure 2.12 Random Amplified Polymorphic DNA (RAPD) technique. (A) A principle of RAPD (Source: http://www.umanitoba.ca/faculties/afs/plant_science/course). (B) An example of a RAPD band pattern (Source: http://www.clas.ufl.edu/users/magitz/notes/Lect9-Molecular_markers.pdf).

A newer system, AFLP marker (Figure 2.13), was introduced by Vos et al. in 1995. AFLP is based on the selective amplification of restriction digested genomic DNA, which can result in as many as 50-100 bands per reaction. AFLP fragments can be resolved with either a manual polyacrylamide gel electrophoresis or with the help of an automated genotyper. AFLP is considered to be more stringent than RAPD technique due to the use of longer primers, precise anchoring sequences and higher annealing temperatures (Bussell et al., 2005). AFLP technique has been widely used for genetic-mapping construction, fingerprinting, analysis of genetic relationship and genetic diversity (Loh et al., 2000; Baydar et al., 2004).

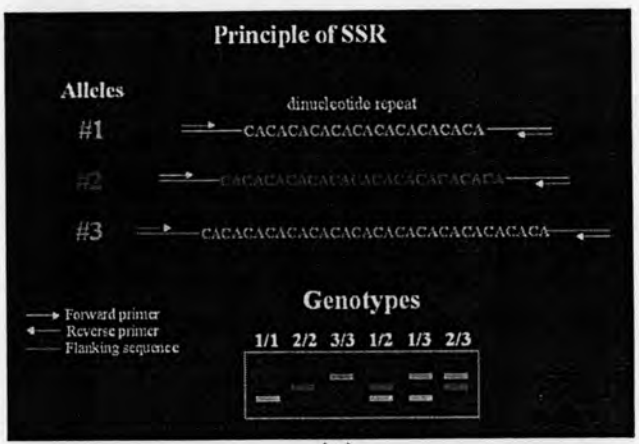
The ability of AFLP method to generate large numbers of high resolution markers in a relatively short period of time makes it very attractive. However, this technique is relatively expensive and more technically demanding than the other techniques (Davis and Nielsen, 1999). The main disadvantage of AFLP-PCR is a difficulty in identifying homologous markers (alleles), rendering this method to be less useful for studies that require precise assignment of allelic states, such as heterozygosity analyses (Mueller and Wolfenbarger, 1999).



(A) (B)

Figure 2.13 Amplified Fragment Length Polymorphism (AFLP) technique. (A) A principle of AFLP (Source: https://www.msu.edu/course/mmg/835/snapshot.afs/DNAmarkers/aflp). (B) An example of AFLP band pattern (http://www.scielo.br/img/revistas/fb/v30n2/a11fig01).

Another PCR-based marker, SSR or microsatellite (Figure 2.14), involves tandemly repeating 1-5 nucleotide units which are abundant and randomly interspersed in the plant genomes (Powell et al., 1996). The number of such repeat units has been found to change at a high frequency and allows the detection of multiple alleles. Primers designed for the sequences flanking the repeat can reveal size polymorphisms that are codominant, highly polymorphism and rapidly typed via PCR (Li, 1997; Symonds and Lloyd, 2003). Due to the larger primer size, the SSR results are more reproducible than RAPDs. However, an analysis of the SSR-PCR requires prior characterisation of the flanking sequences to allow the primer design for PCR amplification. High development cost is a major impediment to a routine application of SSR-PCR in the genetic study to identify SSR markers located in the interested chromosomal regions of non-commercial species (Hayden and Sharp, 2001; Petit et al., 2005).



(A)

(B)

Figure 2.14 Simple Sequence Repeat (SSR) technique. (A) A principle of SSR (Source: http:// sunserver.cdfd.org.in. (B) An example of SSR band pattern (Source: http://bldg6.arsusda.gov /~pooley/soy/cregan/ssr).

Recently, multiplex polymerase chain reaction (multiplex PCR) technique (Figure 2.15) has been successfully applied on genetic screening, genotyping, identification, diagnostic analysis of infectious disease and etc. (Richards et al., 1991; Oliveira and de Lencastre, 2002). Multiplex PCR is a variant of PCR in which two or more DNA fragments are amplified simultaneously. This technique amplifies multiple DNA targets using more than one pair of primers in the single reaction tube. A multiplex reaction is ideal for conserving costly polymerase and templates in short supply. Edwards and Gibbs (1994) suggested that the expense of reagents and preparation time is less in multiplex PCR than in systems where several tubes of uniplex PCRs are used. Moreover, it is also cost-saving for large scale PCR analysis. For maximum efficiency of preparation time, the reactions can be prepared in bulk, randomly tested for quality, and stored frozen without enzyme or template until used.

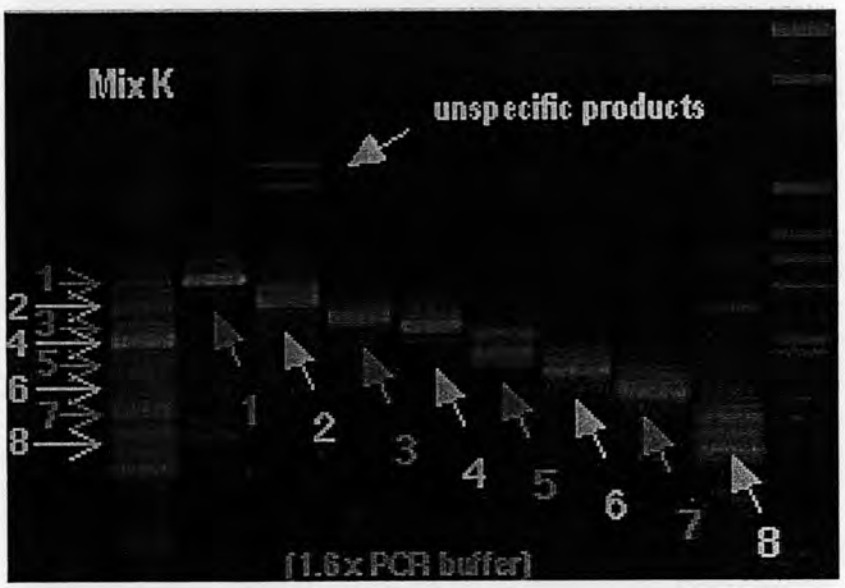
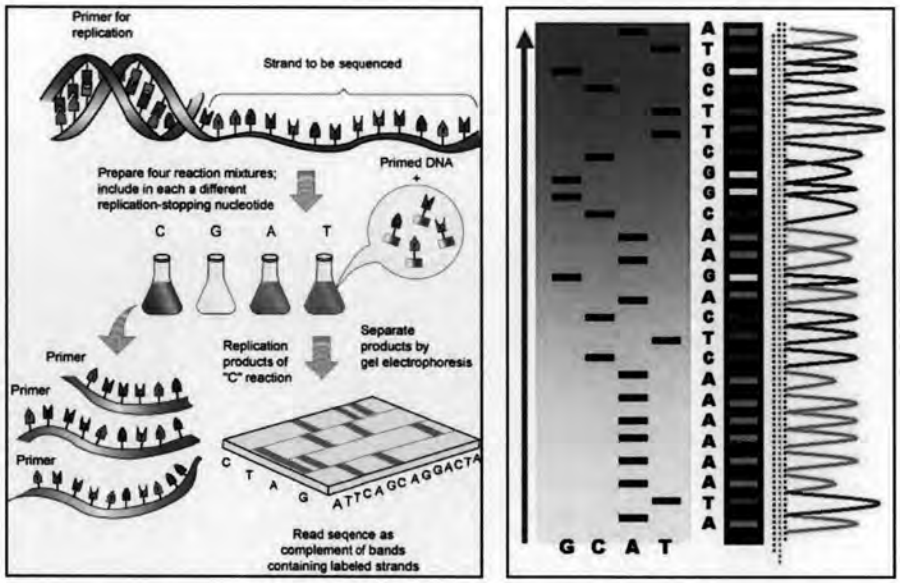


Figure 2.15 An example of multiplex PCR band pattern (lane Mix K = mixing-primer reaction and lanes 1-8 = single-locus PCR) (Source: www.med.yale.edu/genetics/ward/tavi/p10.html).

2.8.2 DNA sequencing technique

Although molecular markers using DNA fragments – RFLP, RAPD, AFLP and SSR - have been widely used in many crop plants, fragment markers also have some limitations in the data analysing step. For instance, DNA band results may not be clear enough for the analysis and some PCR amplified fragments may not be repeatable due to a low quality of the genomic DNA (Thormann et al., 1994; Arcade et al., 2000). These problems particularly lead to an uncertainty when analysing the genetic distance between very closely-related organisms, especially in the case of cultivated crops such as rice, cassava, rubber, oil palm, tobacco, etc. Therefore, to avoid such problem, DNA sequencing technique could be used as an alternative molecular marker than DNA fragment markers.

Sequencing technology (Figure 2.16) has developed rapidly over the past two decades. Knowledge of DNA sequences of genes and other parts of the genome of organisms has become indispensable for basic research studying biological processes, as well as in applied fields such as molecular systematic, phylogenetic and population genetic researchs. The advent of DNA sequencing has significantly accelerated biological research and discovery (Olmstead and Palmer, 1994). Although nucleotide sequencing is a comparatively new approach for molecular systematics, the power of this technique has ensured that DNA sequencing has become one of the most utilised molecular-marker approaches for inferring genetic relationship of individual organisms. The primary attractions of nucleic acid sequencing include the facts that (1) nucleotides are the basic units of information encoded in organism; (2) it is relatively easy to extract and incorporate information about molecular evolutionary processes into analysis; (3) sequence evolution is comparatively easy to model; and (4) the potential sizes of informative data sets are immense (Clegg and Zurawski, 1991; Olmstead and Palmer, 1994).



(A) (B)

Figure 2.16 DNA sequencing technique. (A) A principle of DNA sequencing (http://www.scq.ubc.ca /wp-content/uploads/2006/08/sequencing2). (B) An example of sequence ladder by radioactive sequencing compared to fluorescent peaks (Source: http://en.wikipedia.org/wiki/DNA_sequencing).

Recently, the number of genes and DNA regions used for molecular marker has grown rapidly. The widely-used sequences in many plant studies are internal transcribed spacer (ITS) of nuclear ribosomal DNA (nrDNA) and chloroplast DNA (cpDNA) (White et al., 1990; Chase et al., 1993; Álvarez and Wendel, 2003). ITS sequences are the most often used molecular markers in plant studies whereas cpDNA sequences have gained recognition as a useful marker since the 1990's. Out of 244 plant molecular papers published between the years 1998-2004, it was found that 66% utilised ITS data and 34% of these presented phylogenies based on ITS data alone (Álvarez and Wendel, 2003). For example, Denduangboripant et al. (2001) used ITS sequences to solve problematic sectional-classification of *Aeschynanthus,* family Gesneriaceae. Their study resulted in unique ITS sequence characteristics of some species and led to the findings that from the phylogeography *Aeschynanthus* can be divided into two clades, one occurring primarily in mainland SE Asia and the other in Malesia.

Not only the well-established ITS sequences, but sequences from other regions, especially cpDNA, also have been proved to be useful in plant genetic studies

(e.g. Chase et al., 1993; Olmstead and Sweere, 1994; Steele and Vilgalys, 1994; Mort et al., 2007). DNA from the chloroplast may be able to show genetic relatedness and geographic structure of genetic variation better than nrDNA (Mort et al. 2007). Chloroplast DNA sequences are the primary source of data for inferring plant molecular studies, rivaled only perhaps by ITS sequences in recent years (Baldwin, 1992; Baldwin et al., 1995; Álvarez and Wendel, 2003). Chloroplast genomes typically range in size from 120 to 170 kilobasepairs, and there is a relatively high degree of conservation in size, structure, gene content and linear order of the genes in many plants (Downie and Palmer, 1992). As DNA sequencing technology became available, comparative studies of cpDNA gene sequences began to accumulate sparked by the observations of Ritland and Clegg (1987) and Zurawski and Clegg (1987). Most early publications employed sequences of *rbc*L and were focused on suprageneric taxonomic questions (e.g., Chase et al., 1993). Simultaneously, noncoding regions of the cpDNA were being explored for lower level taxonomic studies (Gielly and Taberlet, 1994). Among the first regions to be exploited were the *trn*L and *trn*F intron (Taberlet et al., 1991), the *atp*B-*rbc*L region (Golenberg et al., 1993; Ehrendorfer et al., 1994; Hodges and Arnold, 1994; Manen et al., 1994) and the noncoding intron portions of the *trn*K/*mat*K region (Johnson and Soltis, 1994; Steele and Vilgalys, 1994).

2.9 Molecular marker studies in tobacco

Recently, the development of new molecular techniques such as genetic mapping with molecular markers - RAPD, AFLP, SSR and multiplex PCR - has significantly improved the possibility of overcoming the constraints that have hindered the improvement of tobacco (*N. tabacum*) yield and quality for many years. These molecular markers have also been employed to study evolutionary genetics and breeding of tobacco. Early works on molecular genetic relationships of tobaccos come from those of Filippis et al. (1996). They used RAPD-PCR to analyse the relationship between two parental species of tobaccos and six somatic hybrids produced as a result of fusion of vacuolated and evacuolated protoplasts and subsequent culture. The result showed that RAPD technology is a versatile, precise, sensitive and cost effective

method for genetic analysis of tobacco hybrids. In 2001, He et al. investigated genetic relationships of germplasms of 31 flue-cured tobacco cultivars also using RAPD, and their genetic diversity was found being rather narrow. Likewise, genetic and geographic polymorphisms of cultivated tobaccos (*N. tabacum*) in Turkey using RAPD method were studied by Arslan and Okumus (2006). These results showed that RAPD markers appear to be a valuable tool for assessing genetic diversity levels in tobacco and were highly discriminatory among accessions and geographic origins. A considerable level of genetic variability within *N. tabacum* was found in their study. Furthermore, Rossi et al. (2000) applied RAPD markers for tobacco cultivar identification and also found that tobacco cultivars appear to have low level of genetic diversity.

In addition, AFLP molecular markers had also been developed to determine genetic polymorphism and evolutionary relationships among cultivated and wild *Nictiana* species (Ren and Timko, 2001). Their result showed that genetic polymorphism existing among wild species of *Nicotiana* species was found greater than among cultivar forms. Moreover, AFLP marker is used to determine the degree of genetic diversity among 46 lines of cultivated tobacco (*N. tabacum*) and observed genetic differences among cultivated forms of tobacco by Rossi et al. (2001). This analysis indicated that the large differences observed in manufacturing quality traits may result from a relatively small number of genetic differences among cultivars, supporting the idea that such traits have a genetically definable basis and low level of genetic diversity. Other AFLP markers were generated to identify bacterial-wilt-resistant tobacco breeds (Nishi et al., 2003) and to conduct on set of 92 *N. tabacum* accessions from diverse types (flue-cured, dark air-cured, Burley, oriental, and cigar wrapper) and breeding origin to identify genetic markers associated with three disease resistance genes (blue mold, potato virus Y, and black root rot) (Julio et al., 2006).

SSR or microsatellite markers were also one of molecular marker techniques studied in tobaccos. In 1999, Davis and Nielsen reported the SSR within tobacco sequences which were available in the public databases, and have revealed a number of di- and trinucleotide repeats within introns and flanking structural genes. However, levels of polymorphism of tobacco sequences in their study were not higher than those found with other methods. In addition, FISH (fluorescent *in situ* hybridisation) has been

used to physically map the chromosomes of *N. tabacum* (Parokonny and Kenton, 1995). This study revealed the chromosomal distribution of ten repetitive sequences which has been traced across the genus *Nicotiana*, section *Tomentosae*. Another microsatellite marker for tobacco genetic fingerprinting and cultivar identification has been performed by Bindler et al. (2007). They developed this maker to best discriminate between Burley, flue-cured and oriental tobacco types and 60 selected cultivars of a reference panel. Their result illustrated the use of SSR markers for tobacco type and cultivar discrimination. Furthermore, SSR marker also used to construct of a genetic map for tobacco (van der Hoeven, 2006). In that effort, SSRs were tested for the presence of polymorphisms on a panel of 16 tobacco cultivars, and a subset of SSR markers was also tested on the two ancestral species (*N. sylvestris* and *N. tomentosiformis*). These results showed low level of polymorphisms between the ancestral genes and the corresponding derived homologs in *N. tabacum* were found.

While DNA fragment markers - RAPD, AFLP and SSR - could not give highly polymorphic products in tobacco studies (Davis and Nielsen, 1999), DNA sequencing method has also been used to examine its efficiency as being molecular markers of tobacco. The early study of DNA sequencing in tobacco was started in 1992. To et al. developed PCR method for sequencing of 16S rRNA gene in tobacco chloroplast genome. Their results showed that the 750 bp amplified fragment was identical to a segment of chloroplast 16S rRNA sequence reported, indicating that their direct sequencing method is useful for the sequencing of chloroplast genome. In 2000, Aoki and Ito studied intrageneric relationships in the genus *Nicotiana* by comparison of DNA sequences of *mat*K gene of the chloroplast genome. A total of 40 taxa were examined in their study, representing 40 of approximately 70 wild species of this genus. The results support the previous hypothesis that this genus originated in South America and subsequently spread to other continents (Albert, 1996; Davis and Nielsen, 1999). The suggestion that the ancestral basic chromosome number is x=12 and that polyploidy and aneuploidy have occurred independently several times during the evolution of *Nicotiana* species was also discussed.

Additionally, the chloroplast genome of *N. tabacum* was the first genome of any flowering plant to be completely sequenced (Shinozaki et al., 1986) and has a total

length, with subsequent corrections by Olmstead et al. (1993), of 155,952 bp. Then, the updated complete gene map which comprises 155,943 bp (Figure 2.17) was reported by Yukawa et al. (2005).

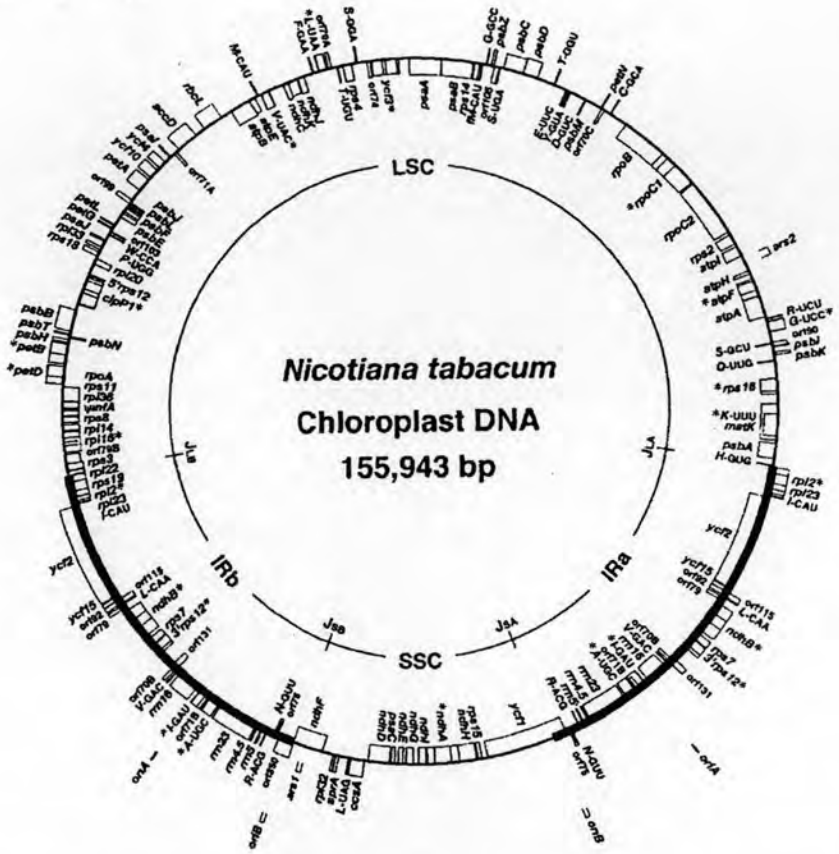


Figure 2.17 The complete gene map of the chloroplast DNA from tobacco (*N. tabacum*) reported by Yukawa et al. (2005).

Another tobacco molecular study using DNA sequencing technique has been performed in 2004. Clarkson et al. have shown the genetic relationships in the genus *Nicotiana*, with 75 naturally occurring species (40 diploids and 35 allopolyploids), inferred from sequences of multiple plastid DNA regions such as *ndh*F and *mat*K genes combined with *trn*L intron, *trn*L-F and *trn*S-G spacers. Their result revealed the line of maternal parent of the allopolyploid species of *Nicotiana* section. *Nicotiana* and the Australian endemic tribe Anthocercideae formed a sister pair, and it is likely to have evolved in southern South America, east of the Andes, and later dispersed to Africa, Australia and southwestern North America.

2.10 Highly-variable noncoding regions of chloroplast DNA and their use in tobacco genetic relationship studies.

For plant genomes, nucleotide sequences of chloroplasts have been proved to be an excellent source of data for molecular genetic relationship studies (Olmstead and Palmer, 1994; Shaw et al., 2005). Features of chloroplast DNA (cpDNA) especially useful in phylogeny reconstruction include its conservative rate of sequence and structural evolution (Wolfe et al., 1987; Palmer, 1991), its large yet manageable size relative to animal mitochondrial DNA, and the lack of independently evolving duplicated genes compared to nuclear DNA (Olmstead and Palmer, 1994). Many early publications usually focused on several coding-regions of cpDNA sequences such as *rbc*L, *mat*K, *atp*B and *ndh*F genes to elucidate genetic relationships among higher-level taxa (i.e. generic level and above) in many plants (e.g. Chase et al., 1993; Steele and Vilgalys, 1994; Jensen et al., 1995; Olmstead and Reeves, 1995).

Not only the coding sequences of chloroplasts, but noncoding cpDNA sequences (introns and intergenic spacers) have also been used in plant molecular systematic research for more than fifteen years (since the work on *trn*L-*trn*F intron of Taberlet et al. in 1991). The noncoding regions of cpDNA have been proved to be more suitable for lower-level taxonomic studies (i.e. species level and lower) than the coding regions (Gielly and Taberlet, 1994). This is under the assumption that noncoding regions should be under less functional constraint than coding regions and then provide greater levels of variation for analyses. Following these pioneering studies, the use of noncoding cpDNA regions has continually increased and is now routinely employed for studies of phylogeny at intergeneric and interspecific levels. The use of noncoding cpDNA sequences to generate plant phylogenies began in the early 1990s with the seminal publications of Taberlet et al. (1991), Morton and Clegg (1993), Clegg et al. (1994) and Gielly and Taberlet (1994). These studies were facilitated by the three chloroplast genomes that had been completely sequenced (*Marchantia polymorpha*, *Nicotiana tabacum* and *Oryza sativa*) at that time.

A survey of papers published on plants from 1995-2002 (Shaw et al., 2005) illustrates that the number of investigations employing noncoding cpDNA is rapidly

increasing. However, most of these studies (about 77%) have used some portions of either *trn*K-*mat*K-*trn*K, the *trn*L intron and/or the *trn*L-*trn*F spacer, and very few investigators have sampled from other noncoding regions of the cpDNA molecule. This illustrates that, although the number of phylogenetic investigations using noncoding cpDNA is increasing every year, the reliance on *trn*L-*trn*L-*trn*F and *trn*K/*mat*K regions is continued in spite of the fact that the phylogenetic utility of these regions is often limited with respect to others (e.g. Sang et al., 1997; Small et al., 1998).

As the majority of current phylogenetic investigations are focused at shallow phylogenetic levels, regions like the *trn*L intron, the *trn*L-*trn*F spacer and the *trn*K intron/*mat*K gene, though providing satisfactory information in some groups (Bellstedt et al., 2001; Ge et al., 2002), often yield poor resolution in other groups (e.g. Hardig et al., 2000; Klak et al., 2003; Muellner et al., 2003). To obtain additional data and provide better phylogenetic resolution, sequences from these popular regions are often coupled with other additional sequence data of cpDNA (e.g. Sang et al., 1997; Wang et al., 1999; Soltis et al., 2001; Bayer et al., 2002; Mast and Givnish, 2002). It has been clearly shown that the phylogenetic utility of different noncoding cpDNA regions within a given taxonomic group can vary tremendously (e.g. Sang et al., 1997; Small et al., 1998; Mast and Givnish, 2002; Perret et al., 2003), but choosing an appropriate cpDNA region for phylogenetic investigation is often difficult because of the paucity of information about the relative tempo of evolution among different noncoding cpDNA regions (Shaw et al., 2005).

Therefore, Shaw et al. (2005) have evaluated the amplification and phylogenetic utility of 21 different noncoding cpDNA regions (Figure 2.18) in a wide range of seed plant lineages. They have proposed that there are many more variable noncoding regions which have rarely been employed. They also showed that the most widely-used regions (such as *trn*L-*trn*F intron) are among the least variable with PIC (potentially informative character) values. Of the 21 separated noncoding cpDNA regions surveyed in this investigation, several regions consistently provided more PICs than the others across all lineages (Figure 2.19). Specifically, five intergenic spacers (*trn*D-*trn*T, *trn*S-*trn*G, *rpo*B-*trn*C, *trn*T-*trn*L and *trn*S-*trn*fM) provided more PICs than the other surveyed regions.

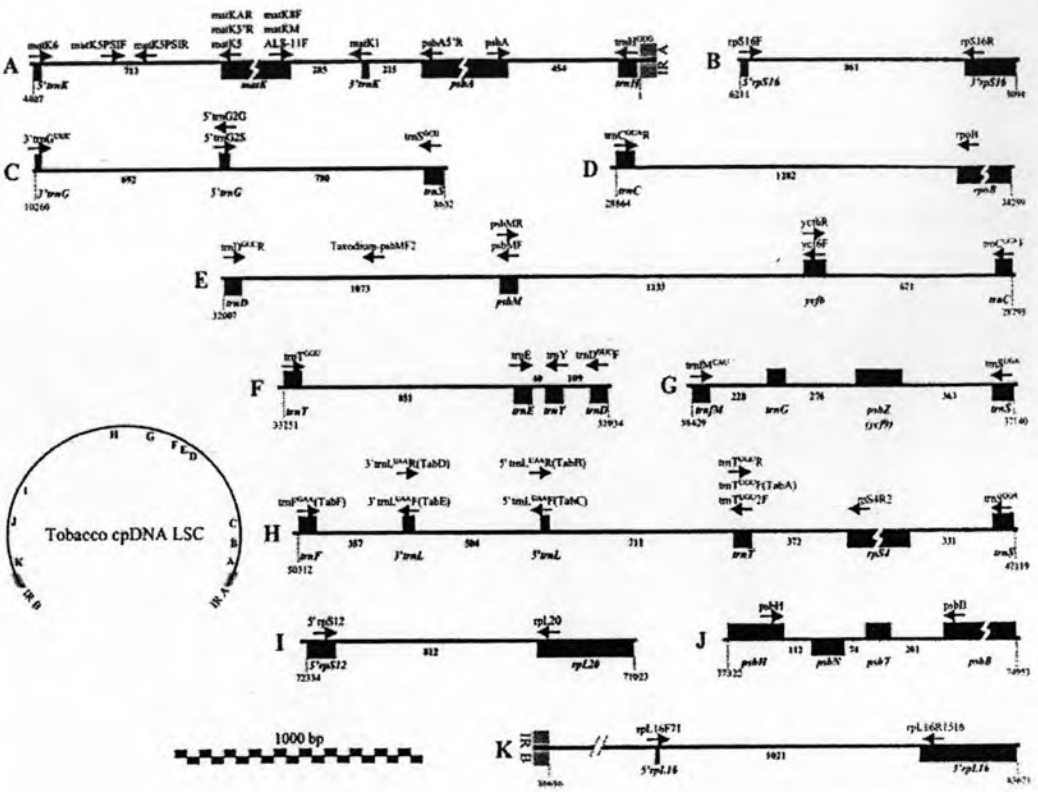


Figure 2.18 Scaled maps of the 21 noncoding cpDNA regions surveyed by Shaw et al. (2005).

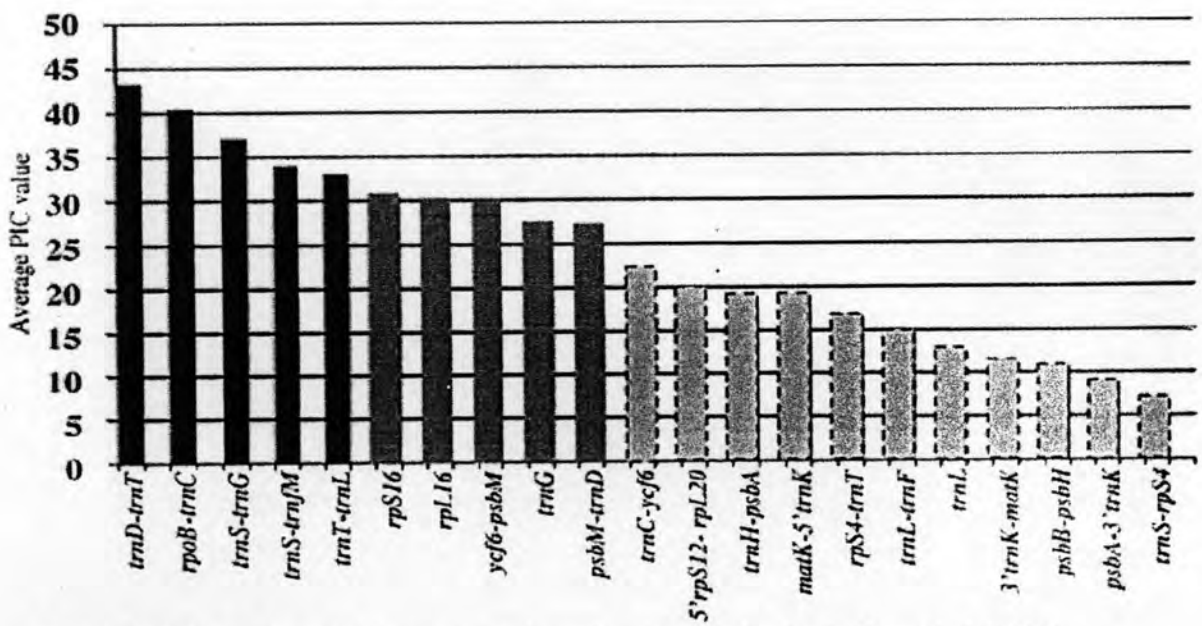


Figure 2.19 The average PIC (potentially informative character) value in Shaw et al. (2005).

The results of Shaw et al. (2005) suggested that there are predictable variations in the levels of variability between the different noncoding regions, and the more variable noncoding regions have rarely been employed, while the most widely used

regions are among the least variable. So, in 2007, Shaw et al. have further suspected that unexplored, highly variable regions of the chloroplast genome may likely exist. Using the best regions of their previous work (Shaw et al., 2005) as a baseline, they successfully identified another 13 unexplored noncoding regions (Figure 2.20) and evaluated them more thoroughly (Shaw et al., 2007). They showed the comparison of the normalised PIC values (Figures 2.21 and 2.22) poignantly illustrates this fact. These exciting results represented an evidence that the chloroplast genome may be better suited for low-level inquiry than previously thought when interpretations were based mostly on *trn*L-*trn*L-*trn*F and *trn*K-*mat*K-*trn*K data. And, their results also showed that at least nine newly explored regions - *rpl*32F-*trn*L, 5′*rps*16-*trn*Q, *ndh*C-3′*trn*V, *ndh*F-*rpl*32R, *psb*D-*trn*T, *psb*J-*pet*A, 5′*trn*K–3′*rps*16, *atp*I-*atp*H and *pet*L-*psb*E - offered the levels of variation (PICs) even higher than those of the most variable regions identified in their 2005 work. The findings of these two studies of Shaw et al. (2005 and 2007) provide an index of the relative levels of variability for plant systematics, population genetics and phylogeographic communities, with several appropriately-sized and quickly-evolving noncoding cpDNA markers.

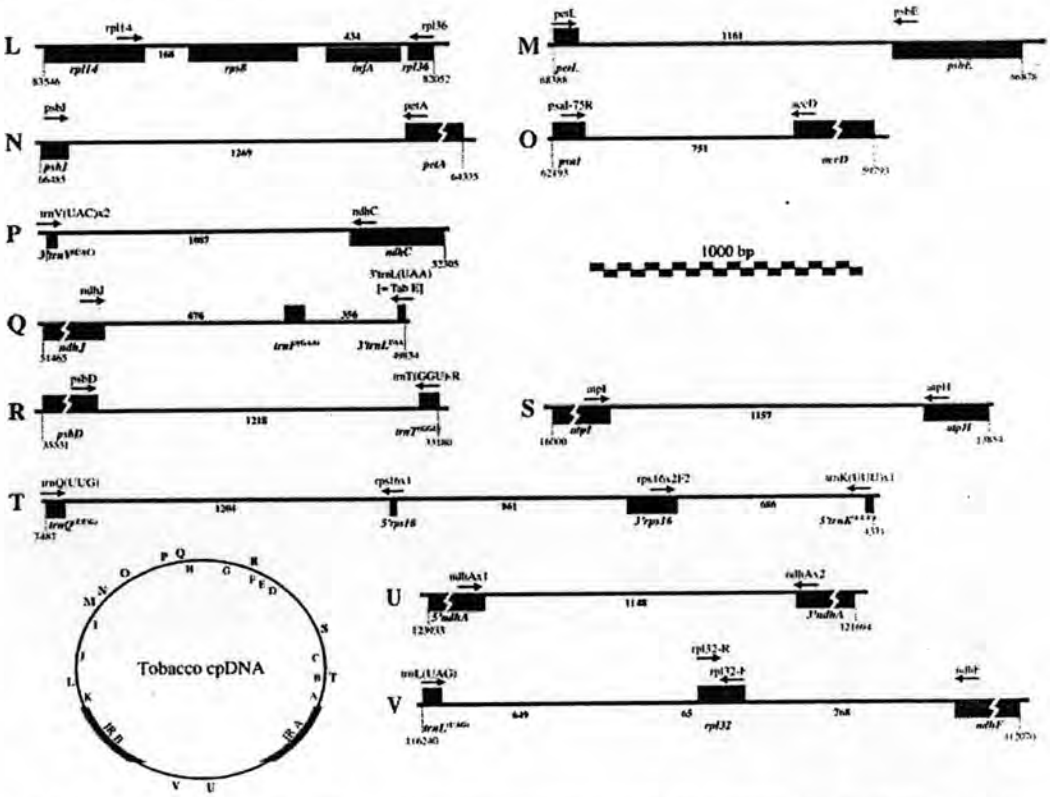


Figure 2.20 Scaled maps of the 13 unexplored noncoding cpDNA regions (Shaw et al., 2007).

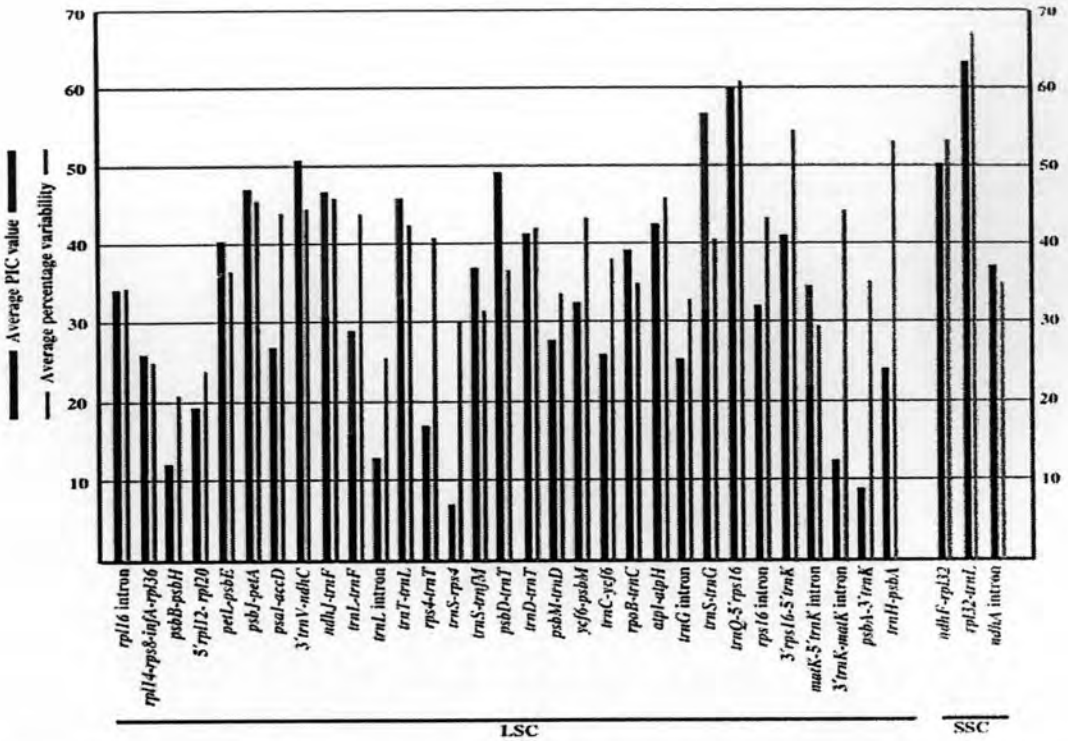


Figure 2.21 The average PIC (potentially informative character) value in Shaw et al. (2007) compared with Shaw et al. (2005).

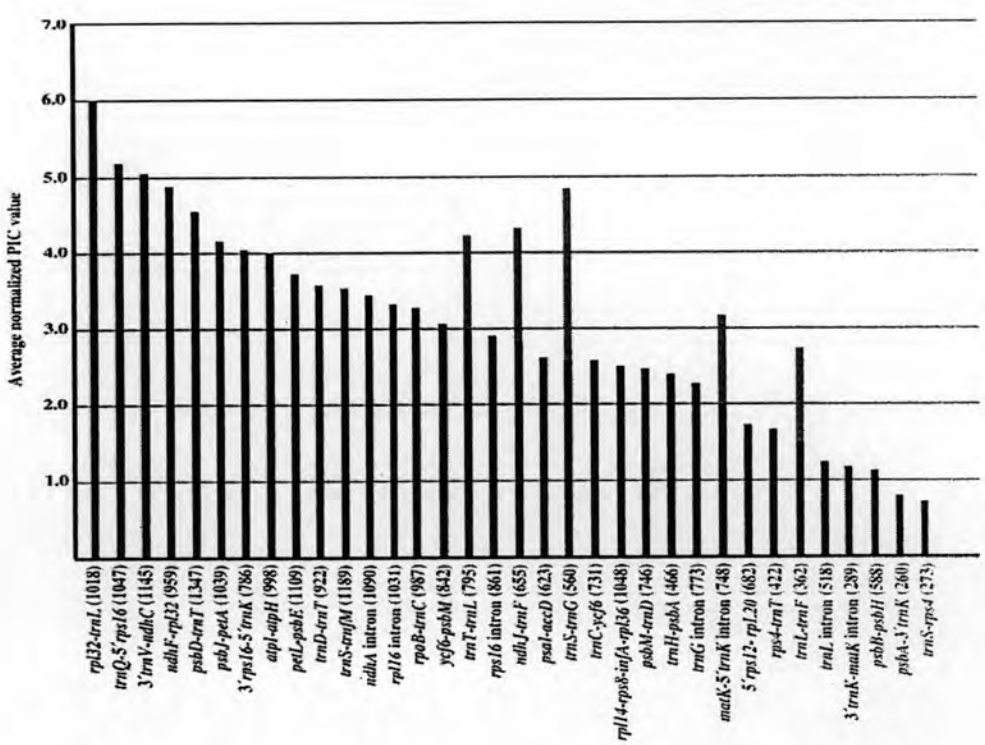


Figure 2.22 The normalised PIC value in Shaw et al. (2007) compared with Shaw et al. (2005). (Gray bars stacked on top of black bars indicate the normalised PIC value of often-combined regions).

Recently, the highly-variable cpDNA regions of Shaw et al. (2007) have been tested for molecular studies in many plants. For example, the phylogeny of Barnadesioideae (Asteraceae) has been inferred from cpDNA sequence data of *atp*I-*atp*H and *rps*16-*trn*K regions by Gruenstaeudl et al. (2009). Falchi et al. (2009) studied phylogeography of *Cistus creticus* L. using *rp*l32-*trn*L sequence of cpDNA. In addition, Sosa et al. (2009) have also used the *rp*l32-*trn*L region and another region, *ndh*F-*rpl*32, to study phylogeographic of the endemic Mexican tulip poppy *Hunnemannia fumariifolia* (Papaveraceae). Kubota and Ohara (2009) also used *ndh*F-*rpl*32 region together with *rpl*32-*trn*L region to investigate the evolution of outcrossing ancestors of *Trillium camschatcense* (Melanthiaceae). Moreover, cpDNA regions of Shaw et al. (2007) successfully developed to use as the chloroplast microsatellites (cp simple sequence repeats (cpSSRs)) primer, such as *atp*I-*atp*H, *psb*D-*trn*T, *ndh*C-*trn*V regions (Ebert and Peakall, 2009). The success of this cpSSRs primer in revealing intraspecific variation is in contrast to many of the studies that have used the universal primers for all angiosperms.

For this thesis, one objective of the study was to separate the two major groups of tobacco cultivars grown in Thailand, local and imported cultivars, which has been a major problem in both legal and technique aspects. For example, only the tobacco cultivars which have been cultivated in the country "for a long time" could be legally called "local" cultivars, although there is neither chemical nor physical standard method to determine their cultivation history. Apparently, the nine highly-variable cpDNA regions of Shaw et al. (2007) seemed to be useful for solving this problem because the regions may be able to show genetic variation within the species level with their extremely fast rate. Therefore, I decided to evaluate these highly variable regions of chloroplast genomes for their potential applicability in genetic relationship study of tobacco cultivars grown in Thailand.

2.11 Phylogenetic tree reconstruction

A phylogenetic tree (Figure 2.23, for example), also known as a phylogeny, is a diagram that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor (Maddison, 1991). Phylogenies are useful for organising knowledge of biological diversity, for structuring classifications, and for providing insight into events that occurred during evolution. Furthermore, because these trees can show descendents from a common ancestor, and because much of the strongest evidence for evolution comes in the form of common ancestry, one must understand phylogenies in order to fully appreciate the overwhelming evidence supporting the theory of evolution (Maddison, 1991; Page and Holmes, 2000).

Tree diagrams have been used in evolutionary biology since the time of Charles Darwin. Therefore, one might assume that, by now, most scientists would be exceedingly comfortable with "tree thinking"—that is, reading and interpreting phylogenies (Page and Holmes, 2000; Baum et al., 2005). However, it turns out that the tree model of evolution is somewhat counterintuitive and easily misunderstood. This may be the reason why biologists have only in the last few decades come to develop a rigorous understanding of phylogenetic trees. This understanding allows present-day researchers to use phylogenies to visualise evolution, organise their knowledge of biodiversity, and structure and guide ongoing evolutionary research (Avise, 1987; Page and Holmes, 2000). A cladogram (Figure 2.24) is a type of tree diagram that is generated through cladistic analysis to show the postulated relationships between different groups of organisms. It is a branching diagram depicting the pattern of shared similarities thought to be evolutionary novelties among a series of taxa.

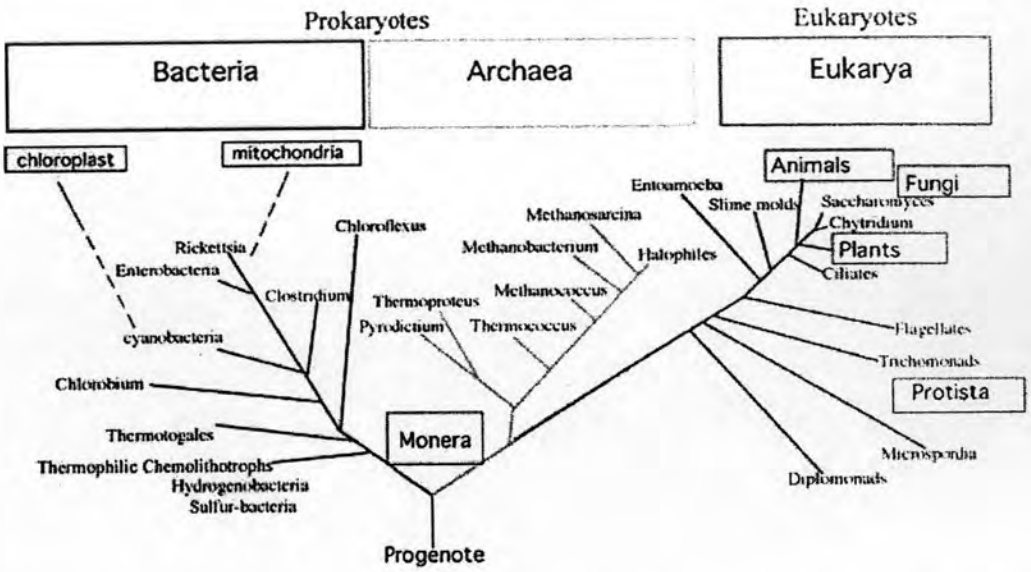


Figure 2.23 Phylogenetic tree based on rRNA data shows the separation of bacteria, archaea and eukaryotes. (source: www.snowballearth.org/lexicon.html)

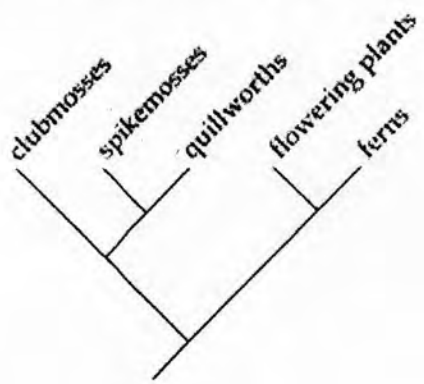


Figure 2.24 A cladogram showing the relationship between various plant groups. (source: http://www.teachersparadise.com/ency/en/media/2/2b/cladogram).

2.11.1 DNA data matrix preparation

Phylogenetic analysis of DNA sequence data begins with an alignment of two or more nucleotide sequences that are hypothesised to be homologous by using computer program such as ClustalW and ClustalX. An alignment involves determining which positions along the DNA sequence are derived from common ancestral positions. The alignment probably inserts gaps to increase or decrease the nucleotides in each DNA sequence and also to decrease nucleotide mismatching. A character-taxon data

matrix is established whenever the sequence alignment is completed (Page and Holmes, 2000).

ClustalX is a Windows interface for the ClustalW multiple sequence alignment programs. It provides an integrated environment for performing multiple sequence and profile alignments and analysing the results (Thompson et al., 1997). Input files contain the sequences in FASTA format (Figure 2.25). The sequence alignment (Figure 2.26) is displayed in a window on the screen. Pull-down menus at the top of the window allow a user to select all the options required for traditional multiple sequence and profile alignment (Thompson et al., 1997).

ClustalX performs a global multiple sequence alignment using the following steps (Ignacimuthu, 2005):

a) Perform pairwise alignment of all the sequences.

b) Use the alignment scores to produce a phylogenetic tree by neighbor-joining (NJ).

c) Align the multiple sequences sequentially, guided by the phylogenetic relationships indicated by the NJ tree.

```
>HK_ndhC-trnV
tTGGAATTGTCAAGTCATCCATAACTATTTAGTCAAAACAAGAATTCATTTTGATCGAAC
CGTCTAGTTTGCTTTGTTTATTGGTTTATTGTAGGGCATATCTCATTGCAAGATTCATCG
ACTGGAATCCGATTTTATTTCCATTATACTTATTTCCATTTTATTTAGTTAGTAGAACCT
TCTAACTATATATTACTCTTATACAAATTCTCTTGTTTCTCTTGTTTTCATCCAGGATTT
TCTCTAAAGACGGGGAATTCTAAATTAATTACTTATCTTATTTCTTCTTTAATTAGAAAT
TCTTTAAAGATTTCTATTTTTTTCTATAAATAGAATCAGGAGGTCTTTTTTCTTATTTTT
TCTTAGTGATTTAGAATAGAACAAGTAATCAAATAGAAGAGAATGTATAGGAATTTCCAT
CTCAAGATTTAGAAGATCTTGTGTTGGTATATTCCTTATTATTATTATTTAATAATAGTA
TTAGGGTTCGAATCCAGGTGACGGGGTTTTTCTTGGTTGAATACAGAAAAAGAGGACTGC
```

Figure 2.25 FASTA format of sequence input files.

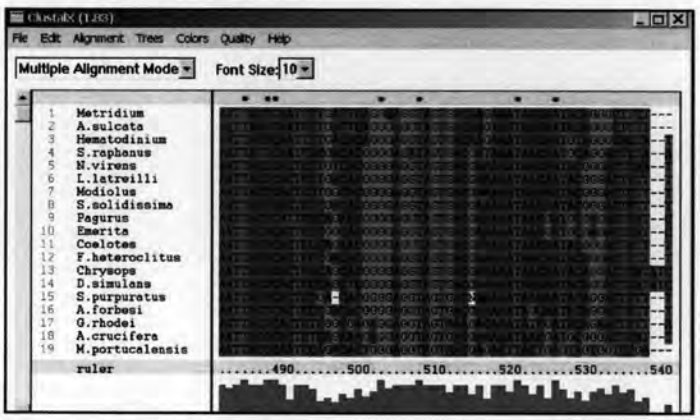


Figure 2.26 Multiple sequence alignment by ClustalX.

(source: http://www.woodrow.org/teachers/BIOLOGY03/Projects/group1/images/clustalx-picprep.gif).

2.11.2 Genetic distance analysis using neighbour-joining (NJ)

To use distance methods for phylogenetic tree building, one widely used method is the neighbour-joining or NJ technique. This method was developed by Saitou and Nei (1987) to be an efficient tree-building method that based on the minimum evolution principle (Figure 2.27). It combines less computational speed with only one resultant tree given. One of the important concepts in the NJ method is neighbours, which are defined as two taxa that are connected by a single node in an unrooted tree (Nei and Kumar, 2000). This technique is a clustering method rather than optimality method, and can not optimise a fitting criterion between tree and data. However, it is a good heuristic method for estimating a minimum evolution tree.

The neighbour-joining method is very suitable with dataset consisting descendants with largely varying rates of evolution. The very fast NJ algorithm provides a good approximation of the minimum evolution tree and is available in many software packages, such as PAUP* (Swofford, 1998). One strategy for finding the minimum evolution tree is to first compute the NJ tree, and then see if any local rearrangement of the NJ tree produces an even shorter tree. Note that this strategy does not guarantee to find the minimum evolution tree (Page and Holmes, 2000).

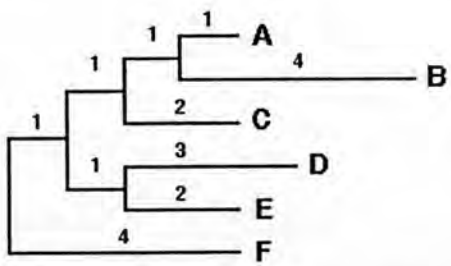


Figure 2.27 Tree-building from neighbour-joining method (Saitou and Nei, 1987).

In this thesis, PAUP* program was used to construct the NJ tree from *rpl*32F-*trn*L sequences. PAUP* (Phylogenetic Analysis Using Parsimony) version 4.0 (Swofford, 1998) is a software package for inference of evolutionary trees, for using in Macintosh (Figure 2.28, for example), Windows, UNIX/VMS, or DOS-based formats. The program can perform high-speed computational analysis of molecular, morphological

and/or behavioral data to infer phylogenetic relationships and has played a major role in evolutionary biology, conservation biology, ecology, and forensic studies. The success of PAUP* has made it the most widely used software package for the inference of evolutionary trees.

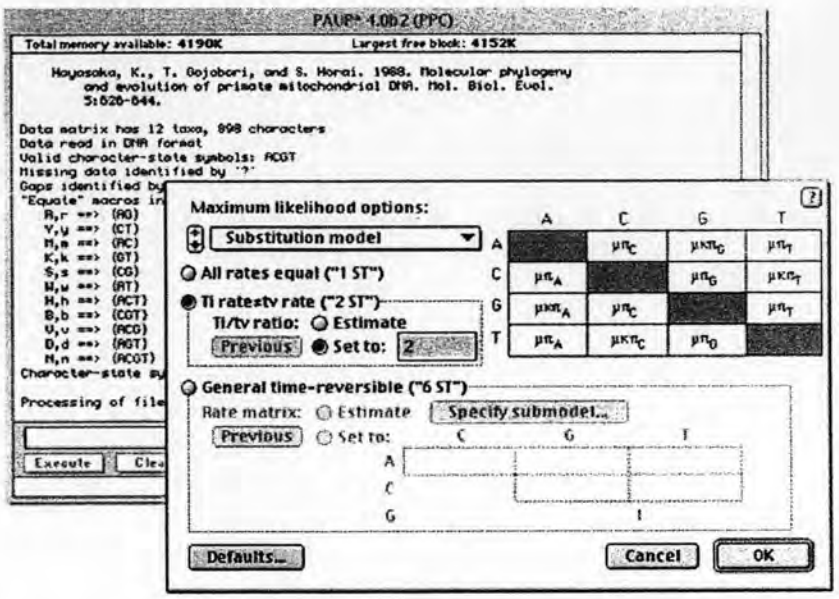


Figure 2.28 PAUP* (Phylogenetic Analysis Using Parsimony) version 4.0 used in Macintosh (source: http://paup.csit.fsu.edu/mac.html).

2.11.3 Tree evaluation

The simplest supporting measurement for individual clades, in the tree is 'branch length'. However, homoplasy (a character that specifies a different and overlapping group of taxa from another to character) makes difficulty to an interpretation of branch length as support (Holmes, 1999). Other tree support approaches aim to circumvent the problem by assessing the number of extra steps that are required before the clade is lost from the consensus tree of near-minimum-length cladograms.

One widely used tree-evaluation methods is a bootstrap supporting value (Figure 2.29), which introduced by Felsenstein (1985). The bootstrap analysis randomly samples characters with replacement (statistically resampling) to form a pseudoreplicate data-set of the same dimensions as the original data matrix. The effect is to delete some characters randomly and to reweight others randomly, with the constraint that the sum of the weights for all characters equals to the number of

characters in the matrix. A large number of pseudoreplicates is generated, typically 1,000 or more. The pseudoreplicates are then subject to phylogenetic analysis, yielding a series of bootstrap trees. Results of the bootstrap analysis could summarise using a majority-rule consenseus tree that indicates the frequency in which each bipartition, or division of the taxa into two groups, is present in the trees obtained for each bootstrap replicate. Interpretation of bootstrap values is complicated (e.g. Hillis and Bull, 1993), but clearly clades with close to 100% support could consider very robust. Group frequencies greater than 50% are indicated as posterior probability (%) nodes. The benefit of this method is that the bootstrap analysis is easy to calculate and widely used.

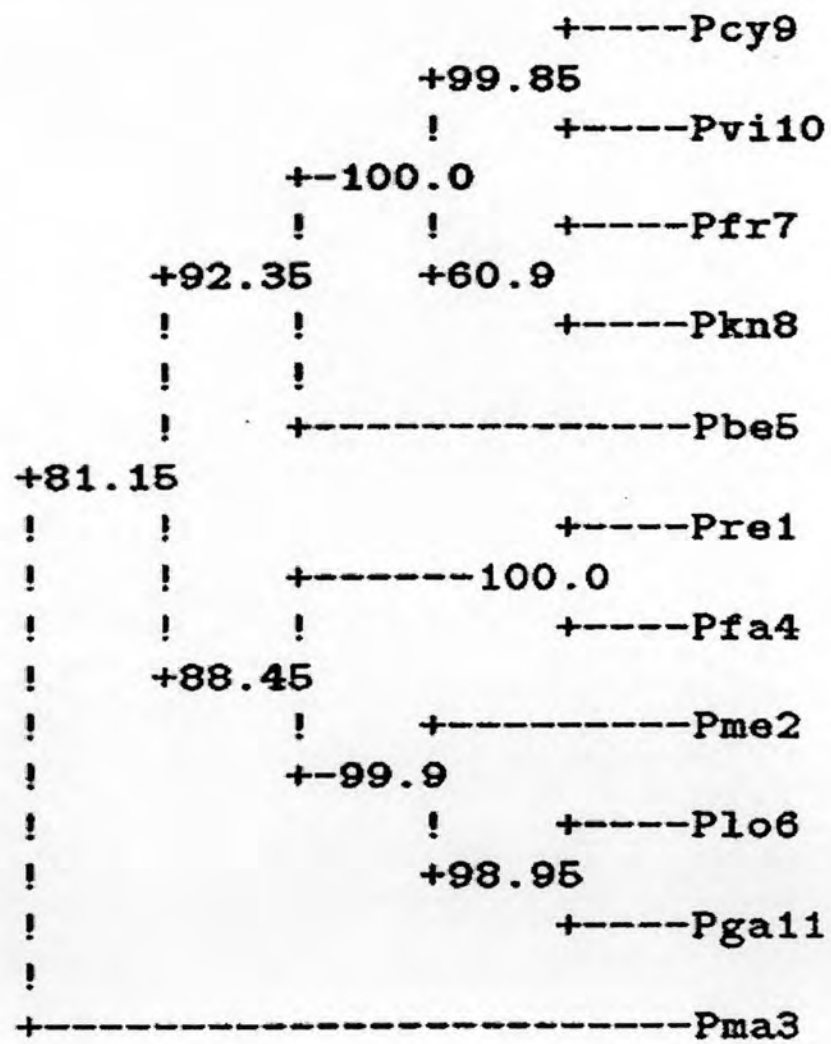


Figure 2.29 Bootstrap 50% majority-rule consenseus tree (Efron et al., 1996)