# THESIS APPROVAL

## GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Philosophy (Agricultural Biotechnology)
**DEGREE**

Agricultural Biotechnology
**FIELD**

Interdisciplinary Graduate Program
**PROGRAM**

**TITLE:** Molecular Markers Identifying Genes Controlling Certain Fatty Acids in Soybean (*Glycine max* (L.) Merr.)

**NAME:** Miss Jidapa Moongkanna

**THIS THESIS HAS BEEN ACCEPTED BY**

**THESIS ADVISOR**
( Professor Peerasak Srinives, Ph.D. )

**COMMITTEE MEMBER**
( Assistant Professor Sutkhet Nakasathien, Ph.D. )

**COMMITTEE MEMBER**
( Mr. Theerayut Toojinda, Ph.D. )

**COMMITTEE MEMBER**
( Ms. Pratuengsri Sinchaisri, Ph.D. )

**GRADUATE COMMITTEE CHAIRMAN**
( Associate Professor Pongthep Akratanakul, Ph.D. )

**APPROVED BY THE GRADUATE SCHOOL ON** .....

**DEAN**
( Associate Professor Gunjana Theeragool, D.Agr. )

THESIS

# MOLECULAR MARKERS IDENTIFYING GENES CONTROLLING CERTAIN FATTY ACIDS IN SOYBEAN (*Glycine max* (L.) Merr.)

JIDAPA MOONGKANNA

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy (Agricultural Biotechnology)
Graduate School, Kasetsart University
2011

Jidapa Moongkanna 2011: Molecular Markers Identifying Genes Controlling Certain Fatty Acids in Soybean (*Glycine max* (L.) Merr.). Doctor of Philosophy (Agricultural Biotechnology), Major Field: Agricultural Biotechnology, Interdisciplinary Graduate Program. Thesis Advisor: Professor Peerasak Srinives, Ph.D. 132 pages.

Soybean is an economic crop used as a main source of vegetable oil. Quality and quantity of oil in soybean seed is important as a raw material determining manufacturing cost. Oil quality is judged from its fatty acid composition which may affect human health. Assessment of oil quality in a breeding program is expensive and time-consuming, thus it is useful to identify molecular markers linked to genes controlling oil content and synthesis of certain fatty acids. The linked markers would be helpful in selection of soybean lines for fatty acid traits. The aim of this research was to develop molecular markers linking to genes controlling seed traits and total oil content in soybean. To identify the markers, an $F_{2:3}$ population comprising 186 families was developed from a cross between Pak Chong 2 and Laos 7122. The population was genotyped by 159 polymorphic SSR markers, and seeds were determined for oil content by hexane extraction method and fatty acid profile by gas chromatography. QTL analysis was done by a simple regression method and composite interval mapping. The results revealed that, from 159 polymorphic markers, 138 were grouped into 30 linkage groups, covering 1,921.1 cM of soybean genome, and 21 remained unlinked. There were 20 putative QTLs locating on 7 linkage groups including A1, C2, D2, E, G, H and O that found associating with number of nodes on main stem per plant, seed length, seed width, 100-seed weight, total oil content, percentages of palmitic, stearic, oleic, linoleic and linolenic acids.

Student's signature Thesis Advisor's signature ___ / ___ / ___

# ACKNOWLEDGEMENTS

First of all I owe a lot to my advisor, Prof. Dr. Peerasak Srinives for his kind suggestions through out this study. I am also grateful to Assist. Prof. Dr. Sutkhet Nakasathien, Dr. Theerayut Toojinda and Dr. Pratuangsri Sinchaisri for their advice and suggestions on this work and laboratory support. I also gratefully thank Assoc. Prof. Dr. Apichart Vanavichit from Graduate School for valuable comment and suggestion.

I would like to acknowledge Prof. Dr. Vincent R Pantalone, Mrs. Deborah Landau Ellis from the University of Tennessee, Department of Plant Sciences for laboratory facility and technical suggestion. I also appreciate Prof. Dr. Joseph W Burton, Dr. Prachuab Kwanyuan, Mr. William P Novitzky from the USDA-ARS, Raleigh, NC for laboratory facilities and the fatty acid analysis; Dr. Jim Holland and Dr. Andrea Cardinal from North Carolina State University, Raleigh, NC for their help.

This thesis was supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education (AG-BIO/PERDO-CHE) and a scholarship from the Royal Golden Jubilee Ph.D. Program under the Thailand Research Fund.

Most importantly, I would like to thank my parents, brother and sister for their constant love and understanding and financial support through out my study.

Jidapa Moongkanna

December 2011

# TABLE OF CONTENTS

**Page**

# LIST OF TABLES

# LIST OF TABLES (Continued)

# LIST OF FIGURES

# LIST OF FIGURES (Continued)

# LIST OF FIGURES (Continued)

# LIST OF FIGURES (Continued)

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| ACCase | = | acetyl CoA carboxylase |
| ACP | = | acyl-carrier protein |
| AFLP | = | amplified fragment length polymorphism |
| ANOVA | = | analysis of variance |
| AOAC | = | Association of Official Analytical Chemists |
| AOCS | = | The American Oil Chemists Society |
| ATP | = | adenosine triphosphate |
| avg | = | average |
| $BF_3$ | = | boron trifluoride |
| BHT | = | butylated hydroxytoluene |
| bp. | = | boiling point |
| CIM | = | composite interval mapping |
| cM | = | centimorgan |
| cm | = | centimeter |
| $CO_2$ | = | carbon dioxide |
| CRD | = | completely randomized design |
| CTAB | = | cetyl trimethyl ammonium bromide |
| CV | = | coeffeiciant of variability |
| cv | = | cultivar |
| C16:0 | = | palmitic acid |
| C16:3 | = | palmitoleic acid |
| C18:0 | = | stearic acid |
| C18:1 | = | oleic acid |
| C18:2 | = | linoleic acid (C18:2 $^{\Delta 9, 12}$) |
| C18:3 | = | linolenic acid (C18:3, n-6; C18:3 $^{\Delta 9, 12, 15}$) |
| ºC | = | degree Celsius |
| $dH_2O$ | = | de-ionized water |
| DMRT | = | Duncan's multiple range test |
| DNA | = | deoxyribonucleic acid |

## LIST OF ABBREVIATIONS (Continued)

| | | |
|---|---|---|
| dNTPs | = | deoxyribonucleotide triphosphate |
| EDTA | = | ethylenediamine tetraacetic acid |
| ELSD | = | evaporative light scattering detector |
| ER | = | endoplasmic reticulum |
| EtBr | = | ethidium bromide |
| EtOH | = | ethanol |
| FAME | = | fatty acid methyl ester (s) |
| FAS | = | fatty acid synthase |
| FID | = | flame ionized detector |
| GC | = | gas chromatography |
| GLC | = | gas liquid chromatography |
| g | = | gram(s) |
| ×g | = | times gravity |
| h | = | hour |
| HDL | = | high density lipoprotein |
| HPLC | = | high performance liquid chromatography |
| HRT | = | hormone replacement therapy |
| KOH | = | potassium hydroxide |
| λ | = | lambda |
| l | = | litre |
| LG | = | linkage group (s) |
| LOD | = | log of odds |
| LDL | = | low density lipoprotein |
| MS | = | mass spectrometer |
| MΩ | = | mega-ohm ($10^6$ Ω) |
| m | = | meter |
| mg | = | milligram |
| mm | = | millimeter |
| mM | = | millimolar |
| ml | = | mililitre |

## LIST OF ABBREVIATIONS (Continued)

| | | |
|---|---|---|
| min | = | minute |
| M | = | molar |
| ng | = | nanogram(s) ($10^{-9}$g) |
| nm | = | nanometer |
| $Na_2SO_4$ | = | sodium sulfate |
| NIR | = | near infrared resonance |
| NMR | = | nuclear magnetic resonance |
| O/N | = | overnight |
| PAGE | = | polyacrylamide gel electrophoresis |
| PCR | = | polymerase chain reaction |
| PVE | = | phenotypic variance explained |
| PVP | = | polyvinylpyrrolidone |
| QTL | = | quantitative trait locus/loci |
| RAPD | = | random amplified polymorphic DNA |
| RFLP | = | restriction fragment length polymorphism |
| RI | = | reflective index |
| RP-HPLC | = | reversed phase-HPLC |
| rpm | = | revolutions per minute |
| s | = | second |
| SCARS | = | sequence characterized amplified regions |
| SNP | = | single nucleotide polymorphism |
| SSCP | = | single-strand conformation polymorphism |
| SSR | = | simple sequence repeat |
| STS | = | sequence tagged site |
| TAE | = | Tris acetate-EDTA buffer |
| TBE | = | Tris borate-EDTA buffer |
| TE | = | Tris-EDTA buffer |
| TEMED | = | *N*,*N*,*N'*,*N'*-Tetramet |
| USDA-ARS | = | The United States Department of Agriculture-Agricultural Research Service |

## LIST OF ABBREVIATIONS (Continued)

| | | |
|---|---|---|
| UV | = | ultraviolet |
| V | = | Volt |
| viz | = | videlicet |
| vs | = | and |
| v/v | = | volume by volume |
| W | = | Watt |
| α | = | alpha- |
| β | = | beta- |
| β-ME | = | β-mercaptoethanol |
| γ | = | gamma- |
| δ | = | delta- |
| µl | = | microlitre |
| µm | = | micrometer |
| 17:0 me | = | heptadecanoic acid methyl ester |

# MOLECULAR MARKERS IDENTIFYING GENES CONTROLLING CERTAIN FATTY ACIDS IN SOYBEAN (*Glycine max* (L.) Merr.)

## INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is the most important economic crop in the world for vegetable oil and feeds. There are many co-products or by-products available from oil extraction process such as hull, soybean cake, lecithin, vitamin E and soapstock (Roa *et al.,* 1998; Blasi *et al.,* 2000; Bruce *et al.,* 2006). The ranges of oil and protein are 19.0-23.5% and 34.9-39.6%, respectively. Soybean accumulates oil in cotyledons in triacylglycerol form. Fatty acid composition in soybean seed, although dependent of varieties, constitutes roughly 11% palmitic (C16:0), 4% stearic (C18:0), 24% oleic (C18:1), 54% linoleic (C18:2) and 7% linolenic acid (C18:3) (Liu, 1997; Hildebrand *et al.,* 2008). The composition affects nutritional value, flavor and stability of the oil.

Annually, Thailand imports over 3 million tons of soybean seed and cake from the US, Brazil, Argentina and China (Office of Agricultural Economics, 2010), because Thailand cannot produce enough to meet with the domestic demand. Approximately 85% of soybean produced in Thailand is used in vegetable oil industry. The main problem of Thai soybean itself is low yielding and medium oil content. The breeding methods which have been used to improve soybean cultivars are conventional, and selection techniques are based on observed phenotypes.

The recent molecular marker techniques have been exploited as a tool for plant breeders to improve desirable traits more efficiently. These techniques could save time and money as compared to conventional phenotypic selection in the field or when applied to selection of biochemical traits which are expensive to analyze as in the case of total oil and fatty acid content. However, during marker development, actual chemical analyses must be performed to obtain a standard phenotyping of a genetic mapping population. Since several steps of extraction and determination of

oil must be done during genetic study and breeding, an optimum method using basic instrument in the laboratory should be determined to save time and money in a large scale operation. The desirable method should be the one that saves time and money and is easy to do with a large number of samples. In breeding programs, plant breeders want to save seeds for planting in the successive seasons, and thus they can spare only as small portion of seeds as possible for any destructive chemical analyses.

Since the polyunsaturated fatty acids become rancid in a short time and not desirable for human consumption. Linolenic acid is the first target fatty acid that plant breeders want to modify through increasing oleic acid and decreasing linolenic acid. The objective of this research was to identify molecular markers linking to genes controlling oil content, some agronomic and seed traits, and 5 major fatty acids in soybean, viz. palmitic, stearic, oleic, linoleic and linolenic acids for use in breeding program.

# OBJECTIVE

1. To compare between different analytical methods and identify a suitable protocol for analyses of total oil content and correlation of fatty acid profile in soybean seed.

2. To develop molecular markers linked to genes controlling some agronomic traits, total oil content and 5 major fatty acids composition in soybean seed.

# LITERATURE REVIEW

## 1. Soybean

### 1.1 Soybean

Soybean (*Glycine max* (L.) Merr.) is the most important vegetable oil and feeds crop in the world. Approximately 85% of soybean produced in Thailand is used in vegetable oil industry. Also, soybean cake, the major by-product of the extraction process, is used as animal feed. About 12 % is used as a source of protein in the diet of Thai people. The remaining 3 % is used as seed.

The planted area has expanded to the lower part of the northern area, and was later extended to the north-eastern region and Central Plains. Approximately 70-78 % of the soybean planted area is in the northern provinces, mainly in Chiang Mai (dry season crop) and Sukhothai (rainy season crop). Annual production of Thailand has been able to supply only 20-30% of the domestic demand. Therefore, Thailand imported over 3 million tones from any countries at cost of approximately 25,000 million baht in 2010 (Office of Agricultural Economics [OAE], 2010). Increase the productivity by releasing high yielding cultivars, for example, SJ.4, SJ.5, Sukhothai 1, Sukhothai 2 and Chiang Mai 60 (CM 60), together with improved cultural practices was not enough. The main problem is low yield and medium oil content germplasm.

### 1.2 Soybean seed composition

Soybean seed component is about 40% protein, 20% oil, 17% cellulose and hemicellulose, 7% sugar, 5% fiber and 6% ash. Soybean seed had oil and protein content in the range 19.0-23.5% and 34.9-39.6%, respectively. The soybean is a good source of quality protein when compare with other protein foods. Soybean oil is rich in polyunsaturated fatty acids and contains no cholesterol. Also, it contains bioavailability of calcium, iron, zinc, phosphate, magnesium, vitamin B and folate.

## 1.3 Soybean process in oil industry and their product

Soybean oil process including of cleaned, cracked, de-hulled and rolled into flakes which ruptures the oil cells for efficient extraction. The crude oil is removed with solvents or screw process. Then, refined soybean oil to the products, such as margarine, salad dressings, and cooking oils. After soybean oil is extracted, the flakes are toasted and ground to soybean meal for used as a high protein component of animal feeds. Moreover, for human consumption products for example soy flour, soy protein concentrate, and soy protein isolated.

There are many co-products or by-products available from oil extraction process such as hull, soybean cake, lecithin, vitamin E and soapstock (Rao, *et al*., 1998, Blasi *et al.*, 2000, Bruce *et al*., 2006). Soybean meal is a common source of protein-rich meal for animal feed. Also, soybean products have soy texture and mouth feel as meat product. Its oil is also used in the food applications /industry including baking, frying, salad dressings, margarine, sandwich spreads, mayonnaise, non-dairy creamers, whipped toppings, snack foods and process foods. In addition, soybean oil is one of the major sources of vitamin E.

### 1.3.1 Vitamin E

Whole soybean is a good source of vitamin E which can remove from oil process over 30%. Tocopherols are present in cereal kernels (such as wheat, corn, etc), oil crop seed (sunflower), certain tropical fruits (oil palm, coconut, etc) and green parts of higher plants. Vitamin E is powerful fat-soluble antioxidant. The main biological function of vitamin E is protects the polyunsaturated fatty acid against peroxidation. Tocopherols are natural antioxidant. It has biological activity may reduce the risk of heart-vascular disease and cancer (Burton *et al.,* 1990; Borek, 2004).

Tocochromanol including tocopherol and tocotrienols, both of them show vitamin E activity. Tocopherols and tocotrienols can separate from the

degree of saturation of the side chain, which is saturated in the tocopherols and unsaturated in the tocotrienols. The derivative of tocochromanol is classified from the number and position of methyl substitution in the chromanol ring including, alpha- (α), beta- (β), gamma- (γ) and delta- (δ).

1.3.2 Lecithin

Lecithin is extracted from soybean oil in refining process and used in the pharmaceutical process to protective coatings. Lecithin is a natural emulsifier, for example, used to keep the chocolate and cocoa butter in a candy bar from separating. In addition, it is a lubricant. Soybean oil is also a natural good source of vitamin E from refining process recover. Soybean oil also contains lecithin which lowers blood levels of cholesterol.

1.3.3 Isoflavones

Soybean has chemical compounds called phytoestrogens. Phytoestrogens have chemical structure similar to the estrogens that can treat the symptoms of menopause. Isoflavones, which have estrogen like properties, are the active ingredients in soybean. There are 3 major isoflavones in soybean including daidzein, genistein and glycitein. The symptoms of post menopausal women are hot flashes, insomnia, nervousness, melancholia, headaches, weakness, vertigo. However, hormone replacement therapy (HRT) can reduce these symptoms, it possible increase the risk of breast cancer. Soy products are a great deal of alternative treatment for menopausal women because they can reduce some metabolic change. Soy products have not only benefit to women's health but also reduce the risk in prostate cancer in men. Cancer needs nutrients via blood vessels. If new blood vessels cannot be growth, the tumor growth will mot be develop or will be stopped. Soy isoflavone may be preventing cancerous cells from growing and spreading by apoptosis process. Moreover, soybean seed compound such as isoflavone showed significant effect to reduce the risk from heart disease, osteoporosis, and breast

cancer. Soybean sprouts, miso, tofu and tempeh have high isoflavone but soy source has low level.

1.4 Soybean oil

Soybean oil share 30% of the world's vegetable oil market. Crude soybean oil contains about 88% neutral lipids, 10% phospholipids and 2% glycolipids. Moreover, oil quality judged from its fatty acid composition, is important for human health.

Soybean oil contains the polyunsaturated fatty acids which reduce blood levels of LDLs and phytosterols. Rich polyunsaturated fatty acids in oil should not be heated because they have grater tendency than other oil toward oxidation and rancidity. Oxidized oil contain free radicals that promote arterial damage, cancer, inflammation, premature aging of cells and tissues. Also, rancid foods and oils develop highly reactive chemicals which produce unpleasant and obnoxious odors and flavors, and destroy nutrients in food. Under some conditions, rancidity and the destruction of vitamins, occurs very quickly.

Hydrogenation process can solve this problem by added hydrogen atom to make the oil more stable, either solid at room temperature or still liquid but more resistant to oxidation. For example, *trans*-fat is found in vegetable shortenings and some margarine, crackers, cookies, snack foods and other foods for increasing the shelf life and the flavor stability of oil. This process changes the chemical composition by forms monounsaturated fats with different physical arrangement. These *trans*-monounsaturated fatty acids are similar in stability to saturated fatty acids. However, many studies concern to the pointed of possible negative health effects of *trans*-fatty acids may behave like saturated fat in the body. In the nature, most polyunsaturated fatty acid is in the *cis*- form but *trans*- fats, the hydrogen atoms exist on the opposite sides of the carbon chain. This causes the fatty acid to be a straighter chain, more a saturated fatty acid than a monounsaturated or polyunsaturated fatty acid, which has bends in the chain. The stereoisomer *cis*- and

*trans*- of oil has been concern because there are many research revealed the effects of oil consumption and diseases, such as cardiovascular, breast or colon cancer. *Trans*-fat drives up the LDL cholesterol which increases the risk of coronary artery heart disease and stroke. The U.S. Food and Drug Administration will soon require *trans*-fatty acids to be included on food labels), the soybean agricultural industry has developed a new heart-healthy soybean, which when processed into oil will lower *trans*- fatty acids content to respond to this problem.

1.5 Major composition of fatty acid in soybean

Fatty acid composition in soybean seed depends on varieties. The average of fatty acid component of commercial soybean oil is 11% palmitic (C16:0), 4% stearic (C18:0), 24% oleic (C18:1), 54% linoleic (C18:2) and 7% linolenic acid (C18:3). The fatty acid composition of oil is variable and depending on the genetic and environment. Fatty acid composition in the oil affect to nutrient, flavor and stability. The objective to improve oil quality depending on the application of the consumption, such as, for cosmetic industry, increasing of saturated fatty acid in desirable.

1.5.1 Palmitic acid

Palmitic acid is not essential fatty acid because human body can synthesize by themselves. Palmitic acid in soybean seed is high percentage of undesirable saturated lipid because it associated with cholesterol, LDL and heart disease. The consumption of low palmitic acid oil may be able to reduce the risk of coronary disease.

1.5.2 Stearic acid

The stearic is saturated fatty acid and not essential fatty acid because human body can synthesize through metabolism.

1.5.3 Oleic acid

Oleic acid is a monounsaturated fatty acid synthesized from the metabolites of the body. It is less oxidized during frying and storage than linolenic acid.

1.5.4 Linoleic acid

Linoleic acid is an essential fatty acids because human cannot synthesize them and have to obtain from foods.

1.5.5 Linolenic acids

Linolenic acid is an essential fatty acids because human cannot synthesize them and have to obtain from foods. In plant, Linolenic acid is synthesized from the desaturation of linoleic acid (C18:2) and elongation of C16:3. It is essential in photosynthesis and pollen development, thus it cannot be eliminated from the seed oil (McConn and Browse, 1996). There are many form of linolenic acid. Gamma-linolenic acid (C18:3, *n*-6) is an essential fatty acid for human and a precursor and intermediate for biosynthesis compound such as prostaglandins, prostacyclins and thromboxanes. In addition, it must be supplied from the diet.

Linolenic acid is a polyunsaturated fatty acid and easily oxidized at three double bonds in its structure. This is a major cause of rancidity and short shelf life in soybean oil. The oxidized oil produces free radicals which may promote cancer, arterial damage, inflammation, and premature aging of cell and tissue. In soybean oil industry, oil additive and hydrogenation were applied to prevent linolenic acid from being oxidized. Since saturated fatty acids have *cis*-configuration of oleic acid, the hydrogenation process converts double bonds to *trans*-saturated fatty acid. These products may increase the risk of cancer, decrease low density lipoprotein, and cause heart diseases (Hu *et al.,* 1997; Lichtenstein *et al.,* 1999 and Mazur *et al.,* 1999).

Linolenic acid was the first target that geneticists selected to modify because this fatty acid is the most easily oxidize and is considered to be a major cause of flavor problems in soybean oil. Also, the polyunsaturated fatty acid become rancid in a short time, so it is not desirable for human consumption. From this reason, plant breeders want to increase oleic acid and decrease linolenic acid to improve oil quality for human consumption.

## 2. Lipids

Lipid is a substantial chemical reserve of free energy. Lipids are usually stored as triacylglycerols (TGs) with fatty acids esterified to the glycerol backbone. TGs are frequently referred to as neutral lipids because of their non-polar nature. The monounsaturated fatty acids and polyunsaturated fatty acids are *cis*- isomer in the nature, meaning that the hydrogens that are bonded to the carbon atoms at the point of unsaturation are on the same side of the carbon chain.

Four principal types are found in plants: triaclyglycerols, phospholipids, galactolipids and a sulfolipid. The same fatty acids found in the oil are also major constituents of cell membranes, where they have an important function in the physiology and development of plants Thus, although oil is normally a seed storage product, its constituent fatty acids are synthesized constitutively in all cells (Somerville *et al.,* 2000).

### 2.1 Lipid in oil seed

Plant seed store oils triacylglycerol as energy sources for germination and post-germination growth of seedling. The oils are preserved in small discrete intracellular organelles called oil bodies. An oil body contains an oil matrix surrounded by a layer of phospholipids embedded with abundant oleosins and some minor protein. Oleosins are unique seed oil proteins and possess the longest hydrophobic segment among the natural proteins. One protein is an unique calcium

binding protein and thus named caliosin, while the other protein is a sterol-binding dehydrogenase and thus termed steroleosin.

Oil is composed of triacylglycerols with three same or different fatty acids esterified to the glycerol backbone. Soybean oil, like most edible oils is composed of five common fatty acids: palmitate (16:0), stearate (18:0), oleate (18:1), linoleate (18:2) and linolenate (18:3). A major fraction of the fatty acids in soybean is the polyunsaturated fatty acids linoleic acid. Linoleic acid ($18:2^{\Delta 9, 12}$) and a-linolenic acid ($18:3^{\Delta 9, 12, 15}$) are essential fatty acids because animals cannot synthesis delta12- and delta 15- double bonds so they must be obtained directly from the diets. The fatty acid composition and distribution triglyceride molecule largely determine oil quality nutritional value, flavor and physical properties, such as oxidative stability and melting point (Somerville *et al.,* 2000; Clemente and Cahoon, 2009).

2.2 Fatty acid synthesis

Acetyl-CoA is initial substrate for synthesis of the carbon backbone of all fatty acids. The enzymes involved in this synthesis are acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS). The name fatty acid synthase refers to a complex of several individual enzymes that catalyse the conversion of acetyl-CoA and malonyl-CoA to 16:0 and18:0 fatty acids. Acyl-carrier protein (ACP) an essential protein cofactor, is generally considered a component of FAS. 18:3 plants synthesize most or all of their lipids in ER, whereas 16:3 plants utilize biosynthetic pathways in the plastid as well. Acetyl Co-A is a precursor in fatty acid synthesis pathway of both saturated and unsaturated fatty acid, particularly palmitic acid is an important intermediate for the synthesis of oleic, linoleic and linolenic acids (Miquel *et al*., 1995). Triacylglycerols are synthesized by acyltransferase that catalyze by add fatty acids to the glycerol backbones (Somerville *et al.,* 2000; Shibata *et al.,* 2008).

2.3 Fatty acid

Unsaturated fatty acids are found in membrane tissue and are also important building blocks for other compounds in the body such as prostaglandins. The change from single to double or triple bonds has an effect on the structure, chemistry and function in the body of the fatty acid. The structure, such as, cell wall membranes contain large quantities of polyunsaturated fatty acids. Some fatty acids are essential for the body. It means that the body cannot function without them. The body cannot synthesize two essential fatty acids, they must be provided by the food. These two acids are linoleic acid and alpha linolenic acid. They are essential for the health, and from them other essential fatty acids are made.

The major component in soybean oil is the polyunsaturated oil. The polyunsaturated fatty acids reduce blood level of LDLs but HDLs too. Monouasaturated fatty acid, also lower blood cholesterol levels but saturated fatty acids raise blood cholesterol levels, one of the major risk factors for heart disease. Soybean oil also contains lecithin which lowers blood levels of cholesterol.

2.4 Important of fatty acid

Because fatty acids are substantially more reduced organic molecules than carbohydrates, fatty acid oxidation has a higher potential for producing energy. Furthermore, triacylglycerols are largely hydrophobic and exist in an essentially anhydrous environment. Carbohydrate, however, are hydrophilic, and the water of hydration adds substantially to their mass. On a mass basis, the ATP yield from catabolism to $CO_2$ and $H_2O$ is approximately twice from triacylglycerols than carbohydrate. Thus, the carbon and energy required for seed germination are often stored in the form of triaclyglycerols rather than as starch.

**3. Total oil analysis**

Total oil content was determined by many techniques.

3.1 Soxhlet extraction

Soxhlet extraction method is standard method in AOCS, AOAC. This technique required special apparatus and extracted by organic solvent. The sample must be dry before analyze. The sample size around 1-6 g depending on the apparatus, add porous boiling stone before extract with solvent, i.e. petroleum ether (bp. 40-60ºC), then refluxed continuously for 48 h. Also, 0.01% (w/v) butylated hydroxytoluene (0.01% BHT) was added in the solvent to inhibit the oxidative degradation or antioxidant. Crude oil extraction was dried using anhydrous sodium sulfate (anhydrous $Na_2SO_4$) to remove any residual water. The residual solvent was removed by flushing with nitrogen.

3.2 Solvent extraction

There are many solvent can used to extract oil from the sample. The solvent was chosen depending on lipid composition in the sample, so the solvent system should compatible for lipid component. These are the example of other research used this technique. Takagi *et al.* (1985) compared solvent mixtures for soybean lipid extraction among 4 procedures. The result show that a chloroform: methanol (2:1, v/v) extraction method which developed by Folch *et al.* (1957) could be recommended as short-time extraction and should be use as routine. Moreover, the chloroform: methanol (2:1, v/v) was extracted as same as Soxhlet extraction with benzene: ethanol (4:1, v/v). However, there was some interfere in Soxhlet extraction with diethyl ether. In other research used other organic solvent system to extraction oil for example chlorinated solvent was not flammable but it was corrosive. Alcohols are flammable and vapor at the low temperature but can extract oil in the intermediate range (Beckel *et al,* 1948). Dahmer *et al*. (1989) applied the method by cut a fragment of cotyledon tissue on the opposite side of the embryonic axis of a single seed to analyze for fatty acid composition. The rest of the seed was then used for planting. In a soybean breeding program, a small modification in protocol steps for oil extraction can lead to a large saving of budget and reduction of toxic waste used in the analyses.

3.3 Nuclear Magnetic Resonance (NMR)

NMR technique analyze by using their chemical and physical properties of the sample. This method determined by integrated area per proton and give equation for determine the unsaturated fatty acids' amount. The result of $^{1}$H-NMR was the spectrum of fatty acids.

3.4 Near Infrared Reflectance (NIR) spectroscopy

NIR is used to analyze the composition in sample, such as moisture, protein and oil content. This technique required calibration curve for predict the composition in the sample from the equation. Also, it can use to analyze from many type of material for example whole seed, single seed, sampling of seed, and ground material. There are many research used NIR, such as, Pallot *et al.* (1999) used NIR to analyze fatty acid composition from the whole seed of brassicas. From this research, they reported that these technique need to develop the calibration curve from each fatty acid composition of the sample. Also, it save time for analyze thousands samples. Hurburgh (1994) used NIR to analyze moisture, protein and oil content. Velasco *et al*. (1999) screened quality trait from single seed of rapeseed by NIR.

**4. Fatty acid analysis**

There are many techniques to analyze fatty acid composition.

4.1 Gas chromatography (GC)

The fatty acid composition of lipids can be determined by using gas chromatography to separate the methylated derivatives of the fatty acids. GC analysis of oil revealed various fatty acids in the oil. The most common recorded for soybeans in the order of abundance were linoleic (48.3-52.0%), oleic (25.4-28.9%), palmitic (10.6-11.2%), linolenic (5.1-6.1%) and stearic (4.4-5.6%). FAMEs were identified by comparison the retention time with known standard and an internal standard was

added in each sample such as, heptadecanoic acid methyl ester (17:0 me). The percentage of fatty acid was obtained by dividing the peak area of the individual fatty acid by the sum of all peak areas obtained for fatty acids.

Boron trifluoride ($BF_3$) is a catalyst for transesterification. $BF_3$ is colorless toxic gas, corrosive, shorten shelf life and always store in the refrigerator temperature. When $BF_3$ was used, the suitable material was chosen to handle it, i.e., stainless steel. The sample size in the preparation of FAME by boron trifluoride in methanol method is important. From the research recommended at least 350 mg of lipid was used for obtained overall recovery of 98%. Other procedure for transesterification was used sodium methoxide, potassium hydroxide in methanol with heat (Shibata et al., 2008).

Sukhija and Palmquist (1988) suggested the one-step extraction-transesterification protocol which rapid, simple, convenient and quantitative to analyze total fatty acid content and composition of feedstuffs and feces by gas chromatography. The method can also be applied in milk, soap sample, and oil seed. It requires only 5 ml of benzene: methanol: acetyl chloride (2.0:2.7:0.3, v/v/v) to extract ≤ 0.5 g sample, thus it can be used for routine analysis of a large number of samples from various materials. The authors also compared the one-step extraction-transesterification (benzene and chloroform) and Soxhlet extraction by petroleum ether to analyze fatty acid in alfalfa pellets, whole cottonseed and soybean seed. The total fatty acid content of oilseed was higher with chloroform than benzene and Soxhlet extraction. While total fatty acid content from the one-step protocol of alfalfa pellets was higher than Soxhlet extraction, both solvents (benzene and chloroform) gave equally good result.

### 4.2 High Performance Liquid Chromatography (HPLC)

HPLC (High Performance Liquid Chromatography) is a kind of liquid chromatography to separate any liquid mixture. Reversed phase HPLC (RP-HPLC) used mobile phase as hydrophilic (water-loving) while the bonded phase is

hydrophobic (water-hating). Compounds that are more polarity prefer the polar mobile phase and move through the column more quickly. Compounds that are non-polar tend to prefer the non-polar bonded phase and move through the column more slowly. In addition this technique can used to analyze Alpha-tocopherol that detected by HPLC-MS, HPLC-UV, HPLC-fluorescence.

High performance liquid chromatography with evaporative light scattering detector (HPLC-ELSD) could detect phospholipids, saturated and unsaturated fatty acid nature. ELSD is a kind of mass detector which nebulize the solute component as fine droplets. While, the detector detect the amount of scattered light from the droplet when the laser illuminate to the droplets (Brouwers *et al*., 1998). Christie (2011) would not recommend an ELSD for quantitative analysis of sample fatty acid derivatives because there is high volatility. Moreover, this technique required ultrapure water.

HPLC-UV was used to analyze lipid content. However, the double bonds of fatty acid could absorb UV, molar extinction coefficient was varied depending on fatty acid composition. Lipid can absorbed UV in the 190-210 nm range. Also, the specific wavelength has not known. Lipid solvents such as chloroform and ethyl ether are strong UV absorption properties.

**5. TG analysis**

From other research, triacylglycerol composition was determined by reverse pressure high resolution liquid (RP-HPLC) with a refractive index detector and used propionitrate as mobile phase**.** In addition, triacylglycerol component was analyzed by GC-FID from fatty acid methyl ester. From this method, triacylglycerol was prepared to methyl ester from by using methylation with KOH in methanol or other procedure before analysis.

The most common methodology used to analyze TG is HPLC because GLC presents some difficulties due to the low volatility of these compounds, the injection

system problem and the effect of the temperature to the column (Stolyhwo *et al.,* 1985). The reverse phase HPLC system eluted with non-aqueous solvent mixture was used to separate TGs. The separation based on the length of fatty acid chain and double bond in the molecule/ the degree of unsaturation.

TGs analysis was determined the stereospecific distribution of fatty acid on the glycerol molecule which controlling by genetic. This is the advantage when compare to fatty acid profile (Amaral *et al.,* 2004). Also, fatty acid composition was calculated from TG composition assuming that each fatty in a TG molecule represent one-third of the percentage of that molecule in the total TGs.

There are several detector have been used coupled RP-HPLC, for example, the ultraviolet (UV), the refractive index (RI), mass spectrometer (MS) and evaporative light-scattering (ELSD) detector.

TGs have weak chromophores. UV detector is not suitable because some solvent can absorb UV such as acetone (Stolyhwo *et al*., 1985; Andrikopoulos, 2002).

RI detector cannot be used in gradient eluent system, is susceptible to temperature and has a poor sensitivity (Stolyhwo *et al.,* 1985; Andrikopoulos, 2002).

MS detector is expensive and required the technical skill but this technique can identify the compound.

ELSD can determine nonvolatile compound, can used with gradient system elution, higher sensitivity than RI detector. This technology can analyzed only semi-quantitative evaluation because nonlinear of the signal (Stolyhwo *et al,* 1985).

**6. DNA extraction**

There are many key points of protocols to extract DNA from soybean. Briefly, young trifoliate leaves used fresh or lyophilized leaf tissue was ground to a

fine powder for extract DNA since it may contain less polyphenolic and terpenoid compounds than older tissue. The key of plant materials for high quality DNA extractions is to properly the tissues for extraction. There are many ways to preserve material, for example, kept algae material was freshly collected, frozen at -80ºC or silica gel preserved. Other methods keep the sample after grinding into the powder and add extraction buffer. By keeping the temperature below 0ºC, the oxidizing enzymes are inactive during this step. If large amounts of plant sample simultaneously process in the same time, the sample can be stored in a -20ºC freezer

The extraction buffer contains high amounts of Polyvinylpyrrolidone (PVP) and β-mercaptoethanol (β-ME) which prevent oxidation of the secondary metabolites in the disrupted plant material. Di-sodium ethylenediamine tetraacetic acid (EDTA) is chelator. CTAB is used as a detergent in the extraction buffer to separate polysaccharide from DNA. Polyvinylpyrrolidone to bind the polyphenolic compounds (Kim *et al.,* 1997).

Chloroform extraction use to eliminate co-precipitation of proteins and polysaccharides. An upper aqueous phase contains the DNA, and a lower chloroform phase contains some degraded proteins, lipids, and many secondary compounds. The interface between these two phases contains most of the cell debris, many degraded proteins (Doyle and Doyle, 1990). A wide bore pipette is used because DNA in solution is a long, skinny molecule that is easily broken.

Polyphenols and polysaccharides bind firmly to nucleic acids during DNA isolation and interfere with subsequent reactions. The good quality DNA should free from contaminating proteins, polysaccharides, and colored pigments.

**7. Molecular Marker**

Molecular markers are DNA sequences that are located at unique positions on the plant's chromosomes and that can serve as identification tags for neighboring genes. Genetic markers are used as a flag to identify the specific location of a genetic

trait of interest on a chromosome. By flagging the desired trait, plant breeders can breed plants more efficiently. Also, it is a tool of genetic research to understand the inheritance and interaction among genes or alleles controlling quantitative traits, such as used to study the genetic basis of accumulate ion compound in soybean seed (Rajcan *et al.,* 2005).

Restriction fragment length polymorphism (RFLP) for soybean were introduced in 1980 (Keim *et al.,* 1989). Next, random amplified polymorphic DNA (RAPDs) and DAP were developed in 1990. Simple sequence repeat (SSRs) marker was used in 1992 (Akkaya *et al.,* 1992). RFLP is the most reliable polymorphism which can be used for accurate scoring of genotypes. Also, RFLPs are co-dominant and can identify a unique locus. (Mohan *et al.,* 1997). Lin *et al.* (1996) found that amplified fragment length polymorphism (AFLP) is the most efficient technique in detecting polymorphism in soybean when compare with RFLP and RAPD. There are many advantage of AFLP including high reproducibility, rapid generation and high frequency of identifiable polymorphisms so this technique was used to identifying polymorphisms and for determining linkages by analyzing individuals from a segregating population. However, AFLPs are expensive to generate as the bands are detected by silver straining, fluorescent dye or radioactivity.

The sequence characterized amplified regions (SCARs) have the advantage of being inherited in a co-dominant fashion in contrast to RAPDs which are inherited in a dominant manner. The polymorphism of SSR based on the number of repeat units in a defined region of the genome being investigated. The number and composition of microsatellite repeats differ in plants and animals. This type of polymorphism is highly reproducible. These primers are very useful for rapid and accurate detection of polymorphic loci and the information could be used for developing a physical map based on these sequences for detecting SSRs have been developed in plants (Mohan *et al.,* 1997). Single-strand conformation polymorphism (SSCP) is a powerful and rapid method but it can use with relatively short DNA fragments. However, SSCP can identify the heterozygisity of the DNA fragment in DNAs of

same molecular weight and can detect the a few nucleotide bases changing. In plants, SSCP is not well developed.

**8. Mapping and Quantitative Trait Loci (QTL)**

The genetic linkage map was construct since the early 1990s. Genetic linkage maps of soybean including the maps from RFLP, AFLP and microsatellite markers or SSR marker are available (Cregan *et al*., 1999).

Molecular markers are used in developing genetic maps of the chromosome, on which individual genes affecting specific traits can be located because of their physical proximity to specific markers. Plant breeding can use the understanding of the genetic basis of protein and oil to improve breeding strategies to manipulate the interesting trait. Molecular markers are a tool for identifying chromosomal region related to particular traits. Linkage map gave the genetic information that maybe valuable for improve the interesting traits and have been useful to detect QTL of soybean genome. The physical proximity is determined by linkage studies, where progenies of crosses between two individuals are analyzed for association of specific markers with desired traits. Mapping and sequencing of plant genomes would help to elucidate gene function, gene regulation and their expression (Mohan *et al.,* 1997).

Marker locations that are associated with the expression of quantitative traits are called QTL. The identification of QTLs is very important for the use of molecular breeding in accelerating and improving the rate of success of our seed company breeding program. Plant breeder used QTL mapping approach to identify genomic region linked to the interesting traits with large parental differences for the trait under study.

**9. Plant breeding for improving soybean varieties**

Plant breeder goal is to develop soybean varieties that suitable for health and food application and oil industry (Rajcan *et al*., 2005). The ideal edible oil of

nutritionist should contain about 3% of linolenic acid. 'Soyola' is soybean variety which reduced linolenic acid content and is not transgenic modified soybean. The soybean oil from Soyola no needs hydrogenation process. Also, it is useful for cooking and longer shelf-life.

Other research team developed soybean by conventional breeding methods using three genes that individually reduce linolenic acid from 7% to about 3.5-5.0%. The genes were designated from *fan1 (A5)*, *fan2*, and *fan3*. The combination of three independent genes, the linolenic acid of the oil was reduced to 1%. To develop commercial varieties, these lines were crossed to the best conventional varieties available and self-pollinated for several generations to obtain true 1% linolenic acid with important agronomic traits. The lines were evaluated for several years in field tests. Finally, they were identified, IA 2064 and IA 3017.

There are some QTL mapping for seed protein, oil content and seed weight (Diers *et al.*, 1992; Lee *et al.*, 1996; Mian *et al.*, 1996; Sebolt *et al.*, 2000). Burton (1987) reported that seed protein and oil contents in soybean seed have been negatively correlation, so increasing of protein and oil content in soybean seed have a limited. The quantitative traits including protein, oil and seed size was controlling by multiple genes. It is difficult to improve soybean variety which has high protein together with high oil level because the negative correlation between protein and oil.

# MATERIALS AND METHODS

## Materials

### Plant material

Sixty-five diverse soybean accessions obtained from Laos, Thailand and the US (Appendix Table 1).

### Equipments for planting and crossing to develop population

1. plastic tag
2. plastic bag
3. paper bag
4. bamboo stick
5. germination paper
6. squeeze bottle
7. ruler
8. digital vernier caliper
9. germination tray
10. forceps

### Chemicals for planting and crossing to develop population

1. fertilizer
2. pesticide
3. fungicide
4. 70% alcohol

**Equipments for laboratory**

1. pH meter:
2. spectrophotometer: Beckman DU® 530 UV/VIS Life Science, Beckman Coulter
3. analytical balance (4 digits): Sartorius, Germany
4. hot air oven: Memmert, Germany
5. vortex mixer: Vortex Genie 2, Scientific Industires, Inc., USA
6. agarose gel electrophoresis apparatus: Gelmate 2000
7. polyacrylamide gel electrophoresis apparatus: Bio-Rad, Sequi-Gen GT system.
8. power supply: Consort
9. refrigerated tabletop centrifuge: Universal 32R, Hettich Zentrifugen, Tuttlingen, Germany.
10. freeze dryer:
11. deep freezer (-80ºC):
12. freezer (-20ºC): Sanyo
13. PCR machine: T1 Thermocycler, Biometra, MJ Research, Inc. Watertown, USA.
14. gel documentation: Syngene, Genius
15. waterbath: Memmert, Germany
16. laboratory shaking machine: NUVE SL350
17. Chrompack CP 9001 gas chromatograph (Chrompack, Middelburg, Netherlands).
18. CP-Sil 88 for FAME capillary column of 100 m long, film thickness of 0.20 µm, inner diameter of 0.25 mm (Varian, J&W Scientific, CA).

19. flame ionized detector
20. Nuclear Magnetic Resonance
21. Soxhlet extraction apparatus
22. autoclave: Sanyo MLS-3780
23. 2 arm balance
24. analytical balance (2 digits):
25. magnetic bar and stirrer (AGR VELP® scientifica)
26. hot plate
27. microwave (Empire)
28. electric blender (RT-02A)
29. desiccator
30. aluminum foil
31. aluminum tray
32. brush
33. moisture can (aluminum can)
34. funnel
35. mortar and pestles
36. spatulas
37. toothpicks
38. volumetric flasks
39. beakers
40. cylinders
41. glass ware
42. parafilm
43. 96 wells PCR plate with silicone lid
44. 1.5 ml microcentrifuge tube
45. 16 × 150 mm screw cap test tube
46. 16 × 100 mm test tube
47. Pasteur pipette and silicone bubble
48. amber vial with Teflon lined seal cap
49. filter paper
50. rack

51. ice boxes

**Chemicals and reagents for laboratory**

All chemicals and solvents used in this study were of analytical grade for chemical analysis and molecular grade for molecular analysis.

1. silica gel
2. acetone
3. *n*-hexane was purchased from JT Beaker
4. methanol (MeOH) was from Fisher
5. chloroform ($CHCl_3$) Merck
6. acetic acid ($CH_3COOH$) were from Merck
7. boron trifluoride ($BF_3$)
8. sodium sulfate anhydrous (anhydrous $Na_2SO_4$)
9. The mixed standard included methyl esters of 1.0% myristic acid (C14:0), 4.0% palmitic acid (C16:0), 3.0% stearic acid (C18:0), 45.0% oleic acid (C18:1, *cis*-9), 15.0% linoleic acid (C18:2, *cis*-9,12), 3.0% linolenic acid (C18:3, *cis*-9,12,15), 3.0% arachidic acid (C20:0), 3.0% behenic acid (C22:0), 20.0% erucic acid (C22:1, *cis*-13), and 3.0% lignoceic acid (C24:0). The standard and mixture of standard fatty acid methyl esters (FAMEs) were from Sigma-Aldrich Chemical, GmbH (Steinheim, Germany).
10. liquid nitrogen
11. Tris-base (2-amino-2-hydroxymethyl-propane-1, 3-diolhydroxymethyl-aminomethane) or Tris –HCl pH 8.5
12. boric acid ($B(OH)_3$)
13. ethylene diaminetetraacetic acid (EDTA or Fe-EDTA)
14. cetyl trimethyl ammonium bromide (CTAB)
15. β-mercaptoethanol or 2-mercaptoethanol ($HSC_2H_4OH$)
16. RNase A

17. polyvinylpyrrolidone (PVP)
18. absolute ethanol
19. λ DNA marker
20. isoamyl alcohol
21. isopropanal
22. sodium chloride (NaCl)
23. agarose
24. ethidium bromide (EtBr)
25. *Taq* DNA polymerase, magnesium chloride ($MgCl_2$) and $(NH_4)_2SO_4$ buffer (Fermentas, Lithuania).
26. dNTP set: dATP, dGTP, dCTP, dTTP (Fermentas, Lithuania)
27. forward-reverse primer
28. mineral oil
29. 95% ethanol
30. ultrapure water or 18 MΩ water
31. urea ($(NH_2)_2CO$)
32. *N*,*N*,*N*',*N*'-Tetramethylethylenediamine (TEMED)
33. ammonium persulfate ($(NH_4)_2(SO_4)_2$ or APS)
34. glass binding
35. bisilane or clearview solution
36. polyacrylamide gel solution (19:1)
37. silver nitrate ($AgNO_3$)
38. acetic acid ($CH_3COOH$)
39. sodium carbonate ($Na_2CO_3$)
40. sodium thiosulfate ($Na_2S_2O_3$)
41. xylene cyanol
42. bromphenol blue
43. 98% formamide
44. formaldehyde (HCHO)
45. Φ174 DNA marker (Fermentas, Lithuania)
47. sodium hydroxide (NaOH)

## Methods

### Part I: Screening and selection of germplasm

#### 1. Propagate germplasm

Sixty-five diverse soybean accessions obtained from Laos, Thailand and the US (Appendix Table 1) were grown together once season to refresh the seeds at Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand. Upon maturity, individual accessions were harvested and hand-threshed. The fully mature intact seeds were packed in polyethylene bags. The seed samples were kept in cold room at 4°C for plant in the next generation.

#### 2. Sampling and preparation soybean seed sample

Soybean seed was sampled and cleaned following AOCS method Ac 1-45 (AOCS, 1993). Each accession produced at least 250 g of seeds. Ground soybean seed was prepared by grinding the seeds finely in an electric grinder. The samples were sealed in polyethylene bags and kept in a desiccator with silica gel at room temperature.

#### 3. Determination of seed moisture content and seed preparation

Seed moisture content was determined in all 65 soybean accessions following AOCS method Ac 2-41 (AOCS, 1993), using 4 samples per accession. The averaged moisture percentage was used to calculate the seed dry weight.

Briefly, three grams of ground soybean seed was weighed into a moisture can which known weight. The sample was placed in a hot air oven at 130°C for 3 h, then cooled in a desiccator and weighed once. The sample was put in the oven for another hour, cooled and weighed again. The process was repeated until a constant weight

was obtained. The percentage of moisture content was calculated using the following formula.

$$\% \ moisture = 100 \times \frac{(weight\ before\ roasting - weight\ after\ roasting)}{weight\ before\ roasting}$$

$$= 100 \times \frac{lost\ weight\ (g)}{fresh\ weight\ (g)} \quad \text{----------(1)}$$

lost weight (g) = weight before roasting – weight after roasting
fresh weight (g) = weight before roasting

**4. Comparison of total oil extraction methods**

In this part, total oil was extracted from 3 standard soybean varieties, viz. CM 60, KUSL 20004 and SJ 5 to compare among the five different methods.

4.1 Soxhlet extraction method (AOCS, 1993 method Ac 3-44).

The standard soybean varieties were extracted total oil by Soxhlet extraction apparatus. In each extraction, 2 g of ground soybean seed was wrapped in a filter paper, put into extraction thimble and extracted by petroleum ether (bp. 40-60ºC). The sample was refluxed continuously for 5 h, then, the solvent was removed by evaporator until no petroleum ether odor was detected. The oil contents of the samples were determined. This method was used as the standard to compare with the other methods.

4.2 Chloroform: methanol extraction method.

Total oil was extracted by chloroform: methanol (1:2, v/v) modified from Bligh and Dyer (1959). Briefly, 0.5 g of ground soybean seed was weighed into 16 × 100 mm test tube, then added with 4 ml of 1:2 (v/v) chloroform: methanol, mixed well with vortex-Genie 2 (600-3,200 rpm; Scientific Industries, Inc. NY, USA)

for 15 min and centrifuged at 2,500-3,000 rpm for 3 min, the supernatant was transferred into another 16 × 150 mm screw cap glass tube. These extraction steps were repeated three times. Four milliliter each of chloroform and de-ionized water ($dH_2O$) were added in each tube in the order, mixed well by vortex for 15-30 s and left them until the solvent was separated. The lower phase was aspirated and combined into a 16 × 100 mm pre-weighed test tube.

Crude oil extracted (oil dissolve in organic solvent) was dried under nitrogen gas and kept in a desiccator, weighed and calculated into oil percentage. The oil percentage was calculated from the formula

$$\%\ total\ oil = 100 \times \frac{oil\ weight}{ground\ soybean\ seed\ weight} \quad \text{----------(2)}$$

oil weight (g) = oil weight after evaporate organic solvent (g)
= test tube with crude oil weight (g) – test tube weight (g)

4.3 Chloroform: methanol: acetic acid extraction method.

Extraction was done as in the chloroform: methanol extraction method, but the solvent was comprised 1:2:0.75 (v/v/v) of chloroform: methanol: 0.15 M acetic acid.

4.4 Hexane extraction method.

This method required 0.5 g of ground soybean seed weighed in a 16 × 100 mm test tube, then added with 4 ml of *n*-hexane, mixed by vortex for 15 min and incubated at room temperature for 3 h. The solution was centrifuged at 2,500-3,000 rpm for 3 min and the supernatant was aspirated to a new 16 × 100 mm pre-weighed test tube. The extraction steps were repeated 3 times and, thus made the total oil hexane volume 12 ml. The supernatant from each cycle was transferred to the same 16 × 100 mm pre-weighed test tube, thus made the total hexane volume

12 ml. The last extraction cycle was incubated at least 16 h (overnight or O/N) to ensure of complete extraction. The crude oil extracts were combined and dried under nitrogen gas and kept in a desiccator (Figure 1), then weighed and calculated to oil percentage.

**Figure 1** Organic solvent extraction by *n*-hexane.

4.5 Nuclear Magnetic Resonance (NMR) technique.

Oil content was measured with a Maran pulsed NMR instrument (Resonance Instruments Ltd., Whitney, Oxfordshire, United Kingdom). This technique was advocated by, USDA-ARS (W.P. Novtizky, 2005 per com.) operated at North Carolina State University, USA. The principle is based on utilizing the area per

proton as determined by an integration equation, resulting in a standard equation for determining the amount of oil.

**5. Optimization of hexane extraction condition for total oil**

The extraction steps followed hexane extraction method as explained above, but the extraction conditions were varied as treatments in this experiment as followed.

5.1 Comparison of sample weight.

An experiment compared between sample weight (0.5 vs 1.0 g) using fine seed powder of soybean cv CM 60. The total hexane volume (8 vs 12 ml), and incubation time in the last extraction cycle (3 h vs O/N) were also compared. Each sample was placed in a 16 × 100 mm test tube, then extracted by the above volumes of total hexane (Figure 2) by adding 4 ml in each cycle. The solution was shaken by vortex for 15 min and left them at room temperature for 3 h to extract the oil. The extraction process was repeated and in the last step incubated 3 h and overnight as in the hexane extraction method.

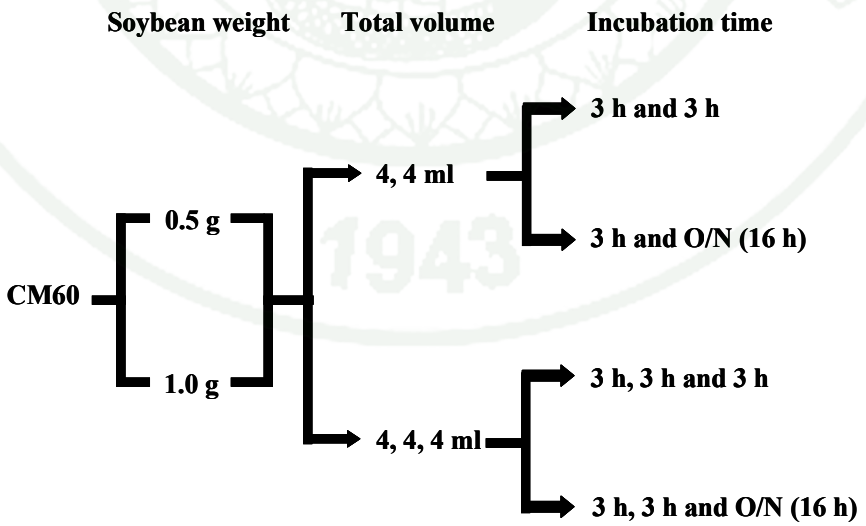



**Figure 2** Treatment combinations for comparing the effects of sample size, hexane volume and incubation time on oil extracted from soybean seed cv CM 60**.**

5.2 Comparison of sample types.

This experiment compared between fine vs coarse powder of 0.5 g samples, using seed from cv CM 60, KUSL 20004 and SJ 5. The fine sample was prepared by electrical grinder, while the coarse sample was prepared by breaking soybean seed with a hammer to obtain the diameter size of about 0.1-0.3 cm. Total oil was extracted using hexane as mentioned in the hexane extraction method above. All samples were extracted by 12 ml of total hexane in three cycles. The operation in each cycle included shaking by vortex for 15 min and incubated for 3 h. After shaking in the last cycle, the samples were incubated for at least 16 h. The oil percentage was calculated from total crude oil in all cycles and compared between total oil from fine vs coarse sample across soybean varieties (Figure 3).

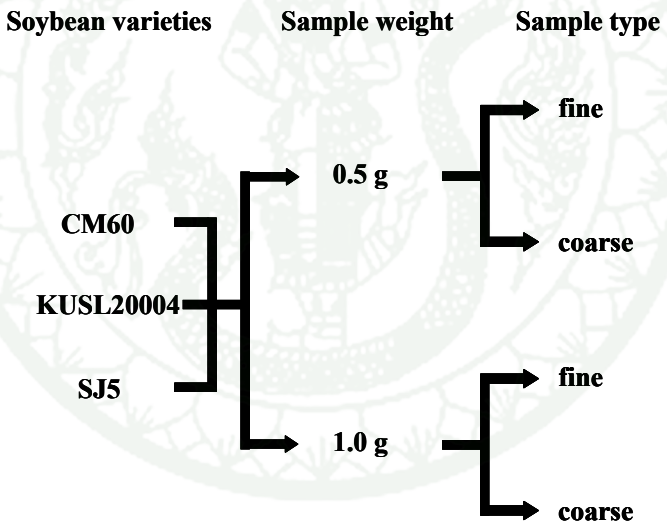



**Figure 3** Treatment combinations for comparing the effects of sample size and sample type on oil extracted from seed of three soybean varieties.

5.3 Comparison between shaking vs no shaking of samples.

Ground samples of 0.5 g each from the three soybean varieties were extracted by total hexane volume of 4 and 5 ml incubated at room temperature overnight vs 6 (2, 2, 2), 9 (3, 3, 3), 12 (4, 4, 4) and 15 (5, 5, 5) ml with the last extraction was done overnight. Each accession was prepared into 4 samples and extracted by the same hexane extraction method above. The samples were divided into 2 sets of the same varieties and incubation condition. One set was then shaken while the other was not shaken. Oil percentage was compared between shaking vs no shaking (Figure 4).

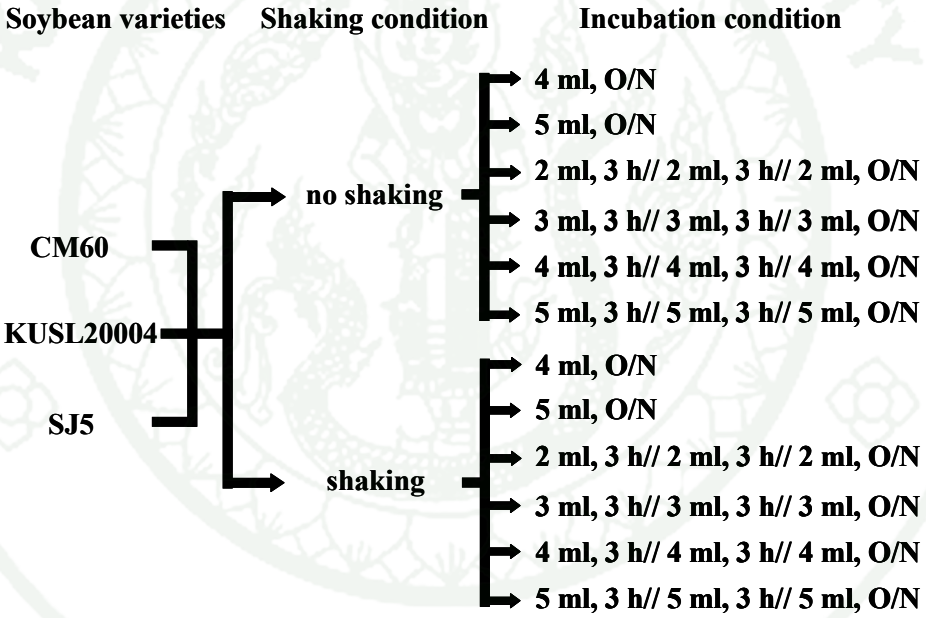



**Figure 4** Treatment combinations for comparing the effect of shaking and no shaking conditions on oil extracted from seeds of three soybean varieties.

5.4 Comparison between final extraction times.

Ground soybean seed samples of 0.5 g each from the three soybean varieties were extracted three times by *n*-hexane. In each extraction cycle, the sample was shaken by vortex with 3 vs 4 ml hexane for 15 min and left at room temperature for 3 h, centrifuged at 2,500-3,000 rpm for 3 min, transferred all crude extract from each extraction cycle into the same 16 × 100 mm pre-weighed test tube. The process was repeated two more times with the last extraction cycle was incubated for 3 h compared with overnight as shown in Figure 5.

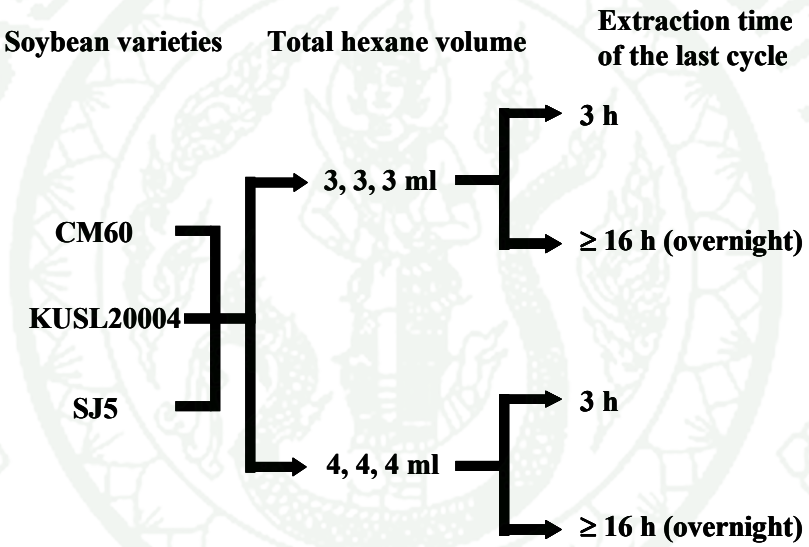



**Figure 5** Treatment combinations for comparing between the total hexane volume and the extraction times of the last cycle (3 h vs ≥ O/N) on oil extracted from seeds of three soybean varieties.

**6. Correlation of fatty acid composition as esterification by 2 methods.**

The fatty acid composition of lipids can be determined by using gas chromatography to separate their methylated derivatives. FAMEs were identified by comparing with known standards. Fatty acid composition in each sample was compared base on the same FAMEs reaction.

6.1 The American Oil Chemists Society's method (AOCS method)

The methods Ce 2-66 and Ce 1-62 of AOCS (1993) were used. Briefly, 0.1-0.25 g of crude oil sample was placed in 50 ml flask, added 4 ml of 0.5 M methanolic sodium hydroxide and a boiling chip. The solvent mixture was heated on waterbath for 5-10 min, then added 5 ml of 14% $BF_3$-methanol reagent, 5 ml hexane and 15 ml saturated sodium chloride, shook well with vortex for 30 s and left them at the room temperature to separate the upper layer which is FAMEs. Then 1 µl of FAMEs can by injected to capillary column of GC. Fatty acid profile was determined using Chrompack CP 9001 gas chromatograph (Chrompack, Middelburg, Netherlands) with a 100 m, film thickness of 0.20 µm, inner diameter of 0.25 mm of CP-Sil 88 for FAME capillary GC column (Varian, J&W Scientific, CA), using helium as the carrier gas, injector temperature of 270ºC, flame ionization detector temperature of 280ºC and a split ratio of 10:1. The GC column was programmed from 200 to 225ºC with 5ºC $min^{-1}$ and then kept at 225ºC.

6.2 The United States Department of Agriculture, Agricultural Research Service (USDA-ARS)'s method

The chemicals used in this method are less toxic and have longer shelf-life than boron trifluoride ($BF_3$) which was used as a catalyst in methylation process. This technique was advocated by, USDA-ARS (W.P. Novtizky, 2005 per com.). Briefly, crude soybean oil 20 µl, added 100 µl of sodium methoxide, then adjusted the total volume to 1 ml with hexane, sealed the cap and mixed the sample. The sample was incubated at room temperature for 1 h 30 min. The FAMEs were analyzed with an Agilent 6890 gas chromatograph with a DB-23 capillary column (30 m × 0.53 mm; 0.5 µm film thickness), using the run of 7.5 min, flame ionization detector temperature of 275°C, injector temperature of 250°C, and the split ratio of 10:1. The oven temperature was set at 200°C with the helium head pressure at the flow rate of 6.7 ml $min^{-1}$.

**7. Statistical analysis in chemical analysis**

Means of total oil obtained from different methods were compared by an *F*-test at $P \leq 0.05$ from the analysis of variance (ANOVA) in either single factor or factorial arrangement in a completely randomized design with 4 replicates. Optimization condition for total oil content was conducted by factorial arrangement in a completely randomized design (CRD) with 4 determinations. The treatment means were compared for their difference by Duncan's multiple range test (DMRT) at $P \leq 0.05$ significant level. Fatty acid compositions obtained from each method were compared. The correlation of fatty acid composition was compared in each individual method. The R software version 2.8.1 (R Development Core Team, 2006) was used to analyze the data.

## Part II: development of $F_{2:3}$ population

**1. Screening total oil content in 65 soybean accessions for choosing parental lines**

The parents were chosen based on their difference in total oil content together with fatty acid composition**.** The suitable methods from part I were used to analysis.

**2. Crossing and developing population**

The crossing was designed by a big different of total oil and fatty acid composition. After, crossing, $F_1$ DNA were tested for confirm $F_1$ hybrid by extracted $F_1$ DNA from $F_1$ leaves and amplified PCR product using co-dominant SSR primer that showed polymorphic between female and male parental line. The PCR products were separated on 1.0% agarose gel run for 30 min at 100 V in 1x TBE buffer, stain with ethidium bromide and photographed under UV light.

The soybean plant was generated until $F_3$ plant. The $F_4$ seeds on $F_3$ plant were harvested and used for chemical analysis, consisting of total oil content and fatty acid

composition as phenotypic data. The seed were put together with silica gel in a dark plastic box and kept in cold room at 4°C until used.

Figure 6 show the diagram for developing $F_{2:3}$ population and the step of how to collect the phenotypic and genotypic data from the population.

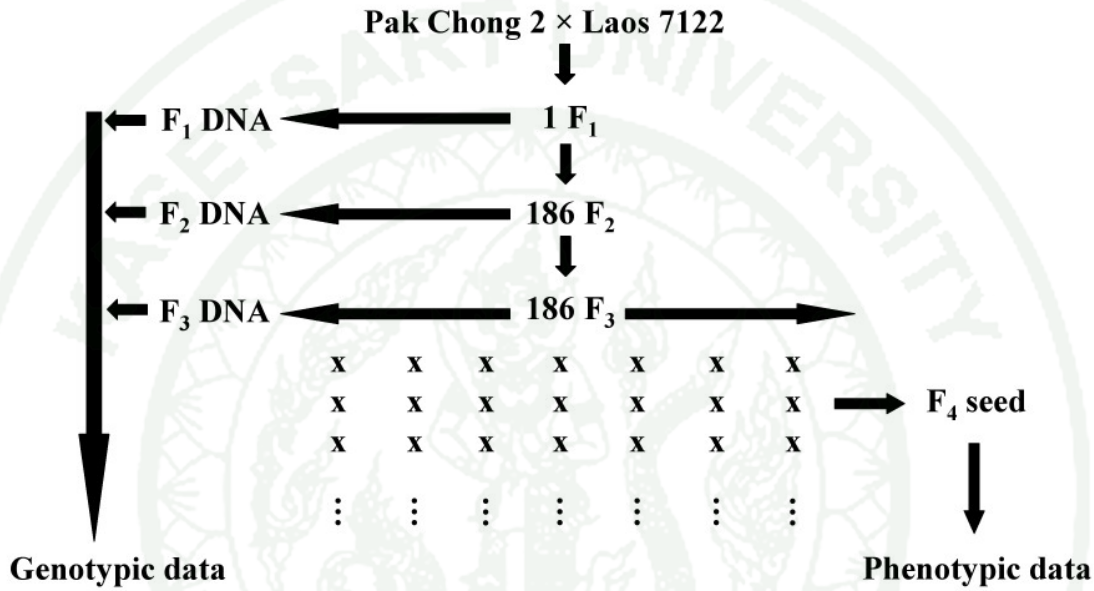



**Figure 6** Developing of $F_{2:3}$ population from crossing between female and male parental line (Pak Chong 2 and Laos 7122, respectively) differential in total oil content and fatty acid composition.

**3. Parental screening by SSR primer on 5% denaturing urea polyacrylamide gel electrophoresis**

Four hundred and fifty SSR markers were used to screen in parental line of male and female on 5% denaturing urea polyacrylamide gel electrophoresis (denaturing urea PAGE). The polymorphic SSR markers were later used in $F_{2:3}$ population.

## Part III: QTL analysis

### 1. Phenotypic analysis

Sampling an $F_{2:3}$ population comprising 186 families was developed from crossing between Pak Chong 2 and Laos 7122 and grown in the field of Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand. Seeds from each $F_2$ plant were sown and 5 $F_3$ plants were sampled as one $F_{2:3}$ lines. $F_4$ seeds from individual $F_3$ plants were used. We ground $F_4$ seed samples to analyze total oil and fatty acid profile followed the result of optimization of hexane extraction condition for total oil extraction and analytical methods for fatty acid compositions. The data of some agronomic and seed traits were collected from $F_3$ families.

#### 1.1 Total oil content

Total oil content was determined from ground $F_4$ seed samples using hexane extraction method which had been optimized for 0.5 g ground soybean seed. A suitable condition for hexane extraction in 16 × 100 mm test tube was used to screen the population. Briefly, each sample was extracted by 3 ml hexane volume by shaking for 15 min and incubated for 3 h in each cycle. The extraction process was repeated 3 times and incubated overnight in the last step.

#### 1.2 Fatty acid composition

We analyzed fatty acid profile from ground $F_4$ seed samples from each $F_3$ plant using gas chromatograph with flame ionization detector (GC-FID) recommended by USDA-ARS method followed fatty acid composition analysis in the previous experiment. Briefly, 20 µl of crude soybean oil, added 100 µl sodium methoxide, then adjusted the total volume to 1 ml with hexane, sealed the cap and mixed the sample. The sample was incubated at room temperature for 1 h 30 min. Fatty acid profile was determined using Chrompack CP 9001 gas chromatograph (Chrompack, Middelburg, Netherlands) equipped with a CP-Sil 88 for FAME

capillary column of 100 m long, film thickness of 0.20 µm, inner diameter of 0.25 mm (Varian, J&W Scientific, CA), using helium as the carrier gas, injector temperature of 270ºC, flame ionization detector temperature of 280ºC and a split ratio of 10:1. The GC column was programmed from 200 to 225ºC with the temperature increment of 5ºC $min^{-1}$ and then kept at 225ºC. The fatty acid compositions were expressed as percentages of total fatty acids.

1.3 Some agronomic and seed traits

Data on agronomic traits were collected from 5 random plants of each $F_3$ families.

1.3.1 Plant height (m) was measured from the ground to the tip of the central axis on main stem when 95% of plants in the plot attained maturity.

1.3.2 Number of nodes on main stem per plant (node) was counted and averaged data from the 5 plants were used in analysis.

1.3.3 Seed length (mm) was measured in mm from ten seeds each of the 5 plants.

1.3.4 Seed width (mm) was measured in mm from ten seeds each of the 5 plants.

1.3.5 One hundred seed weight (g) was measured from 100 seeds per plant from each of the 5 plants.

**2. Genetypic analysis**

2.1 DNA isolation

Two to three grams of young trifoliolate leaves were collected in bulk from each parental line and individual $F_2$ plants. Genomic DNA was extracted by modified CTAB method recommended by Lodhi *et al*. (1994). Briefly, young fresh leaves were ground to fine powder in liquid nitrogen by pre-chilled mortar and pestle. The powder was transferred to a 1.5 ml microcentrifuge tube. Added 700 µl of extraction buffer [2% CTAB, 0.1 M Tris-HCl (pH 8.0), 1.4 M NaCl, 20mM EDTA (pH 8.0), 2% PVP and 2% β-mercaptoethanol]. The solution was mixed thoroughly and the tubes were incubated in waterbath at 65ºC for 1 h, removed them every 15 min and vortex again. Then, added 700 µl of chloroform: isoamyl alcohol (24:1, v/v) in the tube and mixed them. The suspension was centrifuged at 8,000×g (12,000 rpm) for 30 min. The supernatant was transferred to a new microcentrifuge tube and the chloroform: isoamyl alcohol (24:1, v/v) extraction repeated once again. The upper layer was transferred to a new tube and 0.5 volume of 5 M NaCl was added. Then, added an equal volume of isopropanol and gently inverted the tube. The microcentrifuge tube was placed at -20ºC for 1 h and then centrifuged at 8,000×g (~12,000 rpm) for 15 min. The supernatant was poured off. The pellet was washed with 500 µl of 70% ethanol twice. After centrifuged at 8,000× g (12,000 rpm) for 20 min. the supernatant was discarded and the pellet was dried in the incubator at 37ºC for 30 min or left them at room temperature until dried. The dried pellet was re-suspended in 100 µl of TE buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)] with 10 µl of 1 mg/ml RNase A and incubated at 37ºC for 1 h.

2.2 DNA concentration determination by agarose gel electrophoresis

Electrophoresis apparatus including tray and comb was set before prepared the gel. Fifty milliliter of 1.0% agarose gel was prepared from 0.5 g of agarose in 50 ml of 1x TBE buffer. The volume of agarose solution was depending on the tray volume. Then, melted the 1.0% agarose solution and cooled at room

temperature. When the agarose solution was warm, the gel was poured in the tray. Waiting for the 1.0% agarose gel set for 30 min, then 1x TBE buffer was poured on the surface of the gel before took the comb out. Then, bring the tray with agarose gel put in the electrophoresis tank which contained 1x TBE buffer. Poured 1x TBE buffer until the agarose gel was submerged under the buffer. Two microlitre of unknown DNA samples and DNA marker were mixed with 2 µl of 6x loading buffer on the parafilm before loaded in the well. Electrophoresis tank was connect with the power supply and run at 100 V in 1x TBE buffer for 30 min (Sambrook and Russell, 2001), stained with ethidium bromide and photographed under UV light using a gel documentation (SYNGENE, Genius). Genomic DNA was compared with a known concentration of standard λ DNA. Finally, the stock DNA solution was adjusted to the working concentration of 10 ng/µl with free DNase/RNase free water and stored at -20ºC until use.

2.3 Polymerase chain reaction (PCR)

The total volume of each PCR reaction was 10 µl comprising 1 µl of 1x *Taq* buffer with $(NH_4)_2SO_4$, 2 µl of 20 ng/µl DNA, 2 µl of 2.5 µM of forward and reverse primers, 0.8 µl of 20 mM $MgCl_2$, 2 µl of 2 mM of each dNTP, 1 unit *Taq* DNA polymerase (Fermentas, Lithuania). The PCR reaction was conducted in a thermocycler of MJ Research model PTC-100$^{TM}$ (MJ Research, Inc., Watertown, USA.). The reactions were pre-denatured at 94ºC for 2 min and denaturing at 94ºC for 30 s. The cycle was repeated 35 times, then annealing for 30 s at 47-55ºC, depending on SSR primers, elongation at 72ºC for 1 min and the final elongation was held at 72ºC for 10 min.

2.4 Denaturing urea polyacrylamide gel electrophoresis

The denaturing urea polyacrylamide gel was cast using the Bio-Rad 38 × 50 cm of Bio-Rad Sequi-Gen GT system electrophoresis cell (Bio-Rad). Five percentage of denaturing urea polyacrylamide gel electrophoresis was prepared 2 step including prepared apparatus and solution for pre-cast gel. One set of 38 × 50 cm

glass and chamber was cleaned by 95% ethanol 3 times each. Then, cleaned the glass with 1 ml of glass bond solution 1 time followed with 95% ethanol 3 times. The chamber was cleaned with 1 ml of clearview solution. Put the chamber laid down before placed 2 spacers on the chamber at the left and right side, then placed the glass on the top. The glass and chamber was assembled together with the clamps at left and right side. Poured 90 ml of 5% polyacrylamide solution in the squeeze bottle then added 1 ml of 10% (w/v) ammonium persulfate and 100 µl of TEMED, then, mixed well. The polyacrylamide solution was poured into the space between the glass and chamber. Left it at the room temperature for at least 1 h 30 min before used. The pre-cast gel was assembled with buffer chamber and chamber's cover. The 0.5x TBE buffer was added in the lower buffer chamber and inside pre-cast chamber submerge the gel surface. An 0.4 mm of one hundred wells comb was placed on the polyacrylamide gel surface. Loading sample was prepared by mixed 10.0 µl of denaturing dye [0.02 % xylene cyanol and 0.02 % bromphenol blue, 98 ml of 98% formamide and 10 mM EDTA (pH 8.0)] with 10.0 µl of PCR products together for denatured sample. Power supply was set for pre-run polyacrylamide gel at 70 W for 1 h (the temperature around 45ºC). Two microlitre of amplified PCR products of each sample was loaded and compared with 2 µl of 50 ng/µl Φ 174 DNA marker and separated by electrophoresis on 5% denaturing urea polyacrylamide gels in 0.5x TBE at constant power 70 W for 2-3 h depending on sample size 100-700 base pairs in range.

The denaturing urea polyacrylamide gel was taken out from the electrophoresis system, then separated the chamber from the glass which attached with polyacrylamide gel. The polyacrylamide gel was stained with silver staining solution following this protocol. First, the polyacrylamide gel was fixed by soaked in 2.5 l of 10% acetic acid with shaking 50 rpm for 20 min. Next, the gel was washed twice in de-ionized water ($dH_2O$) with shaking 50 rpm for 3 min. Third, the gel was soaked in 2.5 l of silver staining solution with shaking 50 rpm for 30 min, washed by $dH_2O$ a few second before staining. Forth, the gel was soaked in 1 l of developing solution until the band appeared, stopped reaction by soaked in 2.5 l of 10% acetic

acid for a few min. Finally, the gel was placed in 2.5 l of $dH_2O$ before dried in the fume hood. The gel was visualized and scored.

2.5 Screening in parental lines and population by SSR primer

SSR markers were synthesized according to the sequences published on the Soybase website (http://www.soybase.org). Four hundred and twenty-eight SSR primers were used to survey for polymorphism among the parental lines. The polymorphic markers were later used in the $F_{2:3}$ population (Figure 7). The products of amplification were separated on 5% denaturing urea polyacrylamide gels in 0.5x TBE buffer. The polyacrylamide gel was cast using the Bio-Rad 38 × 50 cm gel apparatus. The electrophoresis was performed at a constant power of 70 W for 3-4 h depending on PCR product size, then stained with silver staining solution and visually scored.

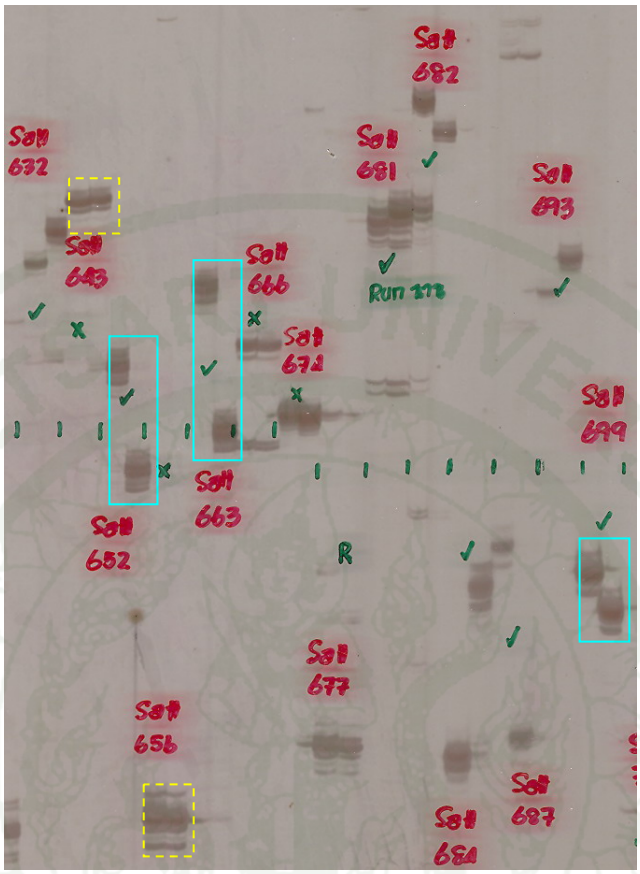



**Figure 7** Parental survey on 5% denaturing urea PAGE in 0.5x TBE buffer stained with silver staining solution. Satt643 and Satt656 showed monomorphic marker and Satt652, Satt663and Satt699 showed polymorphic marker.

2.6 Scoring genotypic data

The polymorphic SSR data were used to construct a linkage map. The marker data were scored in each SSR loci, giving A for homozygous alleles inherited from the female parental line (Pak Chong 2), giving B for homozygous alleles inherited from the male parental line (Laos 7122), giving H for heterozygous alleles from both parents, C for homozygous or heterozygous allele from male parental line (Laos 7122), D for homozygous or heterozygous allele from male parental line (Pak Chong 2) and – for missing data. Only the primers with clear and repeatable in

the parents were used to screen the $F_{2:3}$ population. Figure 8 presented the marker data was scored in each $F_2$ individual.

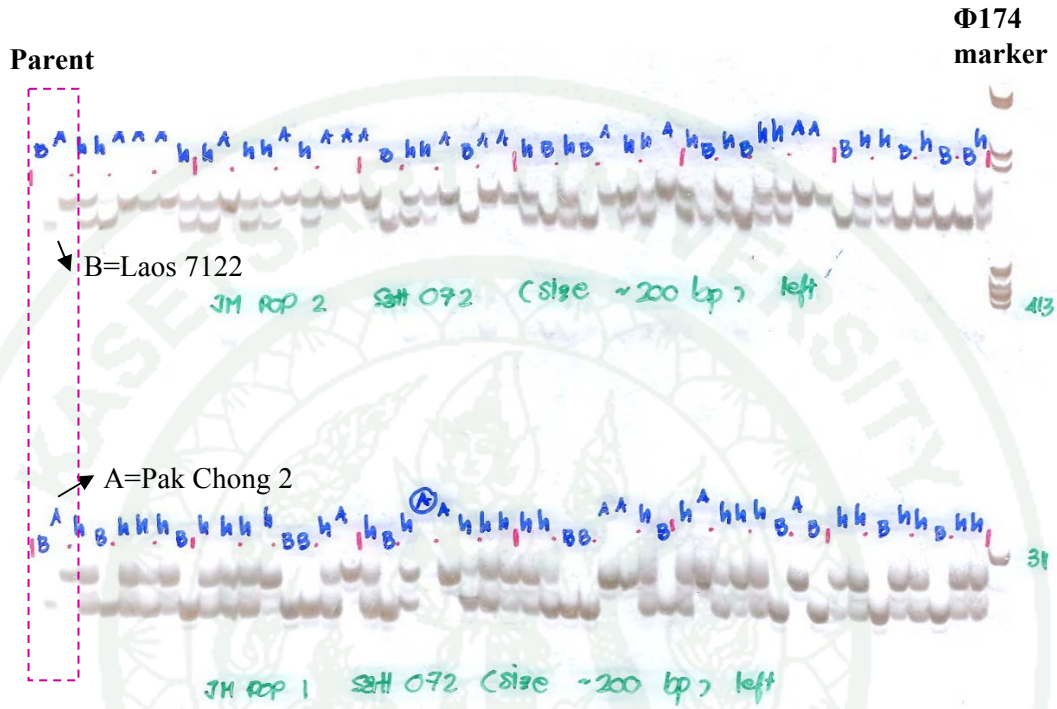



**Figure 8** Scoring co-dominant SSR markers in $F_{2:3}$ population of Pak Chong 2 and Laos 7122 using Satt072. The data was scored of each SSR loci by A = the same pattern band as Pak Chong 2, B = the same pattern band as Laos 7122, H or h = heterozygous allele and - = missing data.

## 3. Data analysis

### 3.1 The Chi-square test

A Chi-square test was used to test for goodness of fit against a 1:2:1 ratio in individual SSR primer. These primers were used to construct the linkage map analyzed by R program version 2.8.1 (R Development Core Team, 2006).

3.2 Linkage analysis

The marker loci that fitted with the ratio were used to construct the linkage map using JoinMap 3.0 software (Van Ooijen and Voorrips, 2001). The parameters included the map distance in Kosambi function (Kosambi, 1944). The mapping analysis was conducted using a LOD score of greater than 3.0 and the maximum distance of 50 cM. The error detection probability level was set at 5%.

3.3 Simple regression analysis

All SSR markers were preliminarily tested by simple regression analysis (R software version 2.8.1, 2006).

3.4 QTL analysis

QTL analysis was performed by composite interval mapping (CIM) using Window QTL Cartographer 2.5 software (Wang *et al.,* 2011). One thousand permutation tests were performed on each trait for the empirical logarithm of the odd (LOD) threshold at the significant level 0.05. Also, 2,500 permutation tests at significant level 0.01 was performed. The QTL was considered at the position where LOD score exceeded the corresponding significant threshold of every 1.0 cM between the adjacent linked markers.

All experiments in this research were shown in Figure 9.

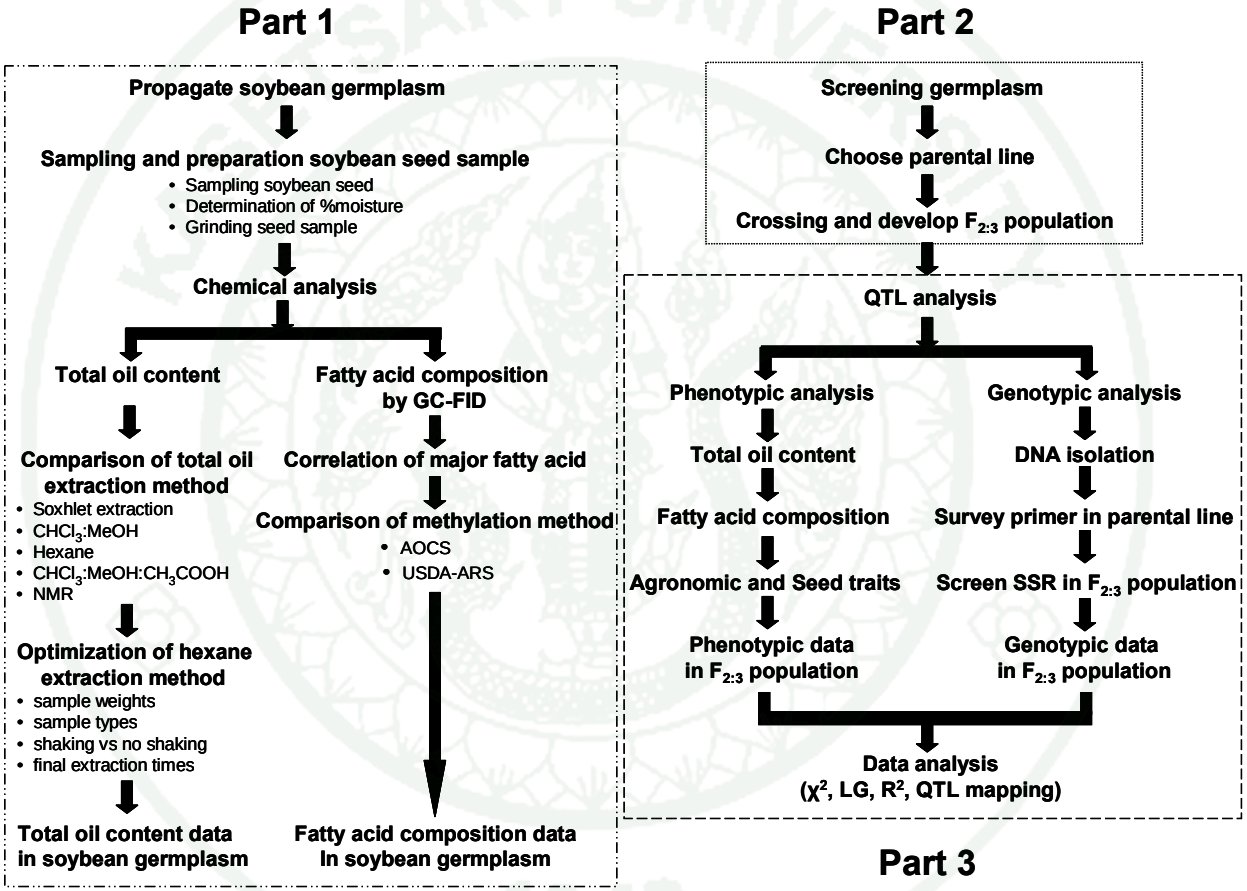



**Figure 9** Overall research

# RESULTS AND DISCUSSION

## Results

### Part I: Screening and selection of germplasm

### 1. Determination of seed moisture content and sample preparation

Seed moisture was used to calculation of total oil content in seed dry weight. Figure 10 and 11 showed that seed moisture content average was 6-10.5% in range from AOCS's method and 2-5% in range from NMR. Normally, soybean germplasm should keep seed moisture under 13%. Ground soybean seed cannot be kept for a long time because it is easy to damage by oxidation. For long term storage, soybean samples should be kept in the form of intact seed.

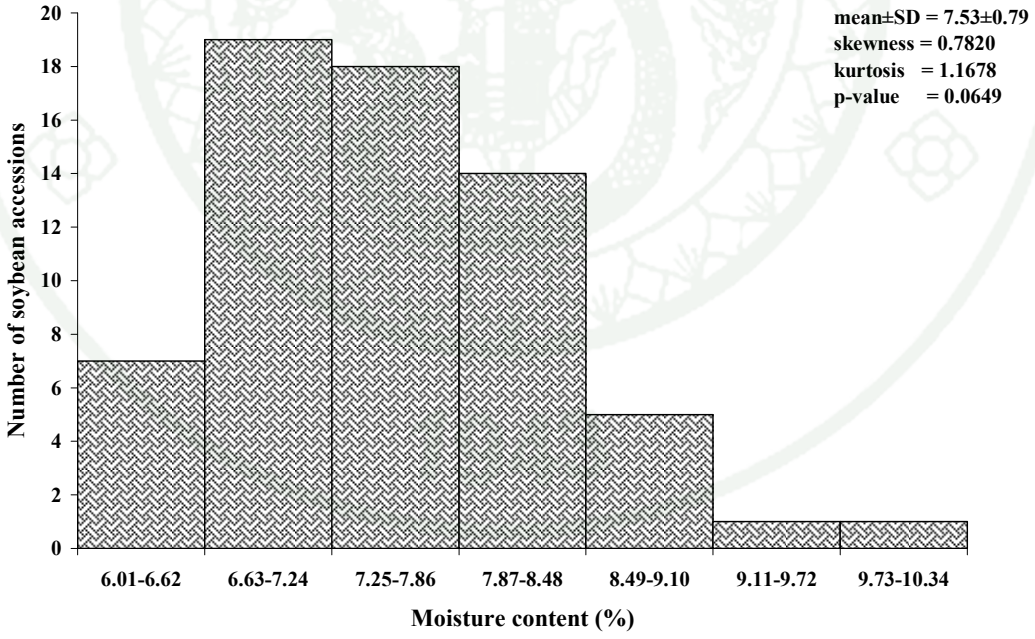



**Figure 10** The distribution of moisture content in 65 soybean accessions analyzed by AOCS method.

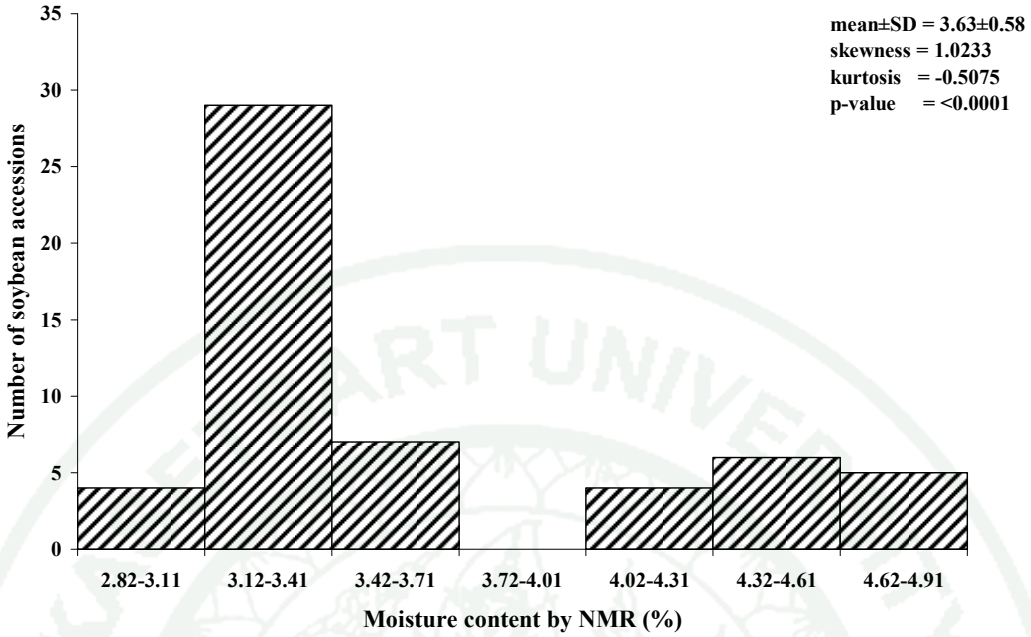



**Figure 11** The distribution of moisture content in 65 soybean accessions analyzed by NMR technique.

## 2. Comparison of total oil extraction methods

Table 1 shows mean percentage of oil content in three soybean varieties as determined by five extraction methods. The results could be classified into three groups. The first group comprised the method of hexane extraction, chloroform: methanol extraction, and NMR technique. All these three methods recovered the highest percentage of oil (over 20 % based on dry weight). The second and third groups were Soxhlet extraction and chloroform: methanol: acetic acid extraction, respectively. Although there was no different in oil percentages extracted by three methods of the first group, the NMR technique has an advantage that it produces no waste and uses no solvent after developing a standard equation for oil prediction. However, the machine is expensive. The data on oil content of 65 soybean varieties also showed no different between hexane extraction method and NMR technique. Comparison between Soxhlet extraction and hexane extraction methods showed that Soxhlet extraction took a long time (5 h per sample) to extract an oil sample. An experiment designed for 2 determinations per sample could accommodate only

3 samples per day, providing that the instrument can analyze 6 samples at a time. For hexane extraction, more samples could be analyzed per day using the basic instrument in the laboratory (centrifuge, two arm balance, vortex mixture or orbital shaker and 4 decimal point balance). Thus the hexane extraction method was further used to analyze oil content of soybean seed in the experiments on optimization. Hexane is suitable for extracting neutral lipid (triglyceride). Yet, it is used in vegetable oil industry, cheap, less toxic than chloroform and easier to handle its waste.

**Table 1** Total oil percentage extracted from ground soybean seed samples as determined by 5 methods.

| Varieties | Extraction methods | | | | |
|---|---|---|---|---|---|
| | Soxhlet | 1$CHCl_3$:2MeOH | 1$CHCl_3$:2MeOH:0.75 of 0.15 M $CH_3COOH$ | Hexane | NMR |
| CM 60 | 19.71 | 20.75 | 3.35 | 21.60 | 20.57 |
| KUSL 20004 | 19.16 | 21.13 | 3.31 | 21.62 | 20.95 |
| SJ 5 | 17.93 | 20.67 | 2.95 | 20.48 | 19.84 |
| Average[1/] | 18.93 b | 20.85 a | 3.20 c | 21.23 a | 20.45 a |

[1/] Means of the methods followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

**Table 2** Advantages and disadvantages of three analytical methods for total oil in soybean seed.

| Items | Soxhlet apparatus | Organic solvent extracted in test tube | NMR |
|---|---|---|---|
| Amount of seed sample | 3-5 g | 0.5 g | whole seed or ground sample |
| Instrument | specific | basic | advance and specific |
| Effect to sample | destructive | destructive | non-destructive |
| No. samples determined per day | a few | many | many |
| Solvent volume per sample | 150 ml | 12 ml | non |
| Waste volume | high | low | non |
| Solvent cost | medium | low | non |
| Instrument cost | medium | Low to medium | high |
| Technical requirement | skill | skill | standard curve |

**3. Optimization of hexane extraction condition**

3.1 Comparison of sample weight.

This experiment used 0.5 and 1.0 g ground soybean seed samples of CM 60 to be extracted with total hexane volumes of 8 and 12 ml, and varied time of the last extraction (3 h vs ≥ 16 h). The overnight and 3 h in the final extraction step gave the same result (Table 3). The ground sample weight of 0.5 g was more suitable for the test tube condition than 1.0 g sample, as the smaller samples can be extracted more thoroughly in a shorter time by a given solvent volume. The results showed interaction between sample weight and volume (Table 4 and Figure 12). Averaged across two sample sizes, 12 ml of total hexane volume extracted more total oil than 8 ml (Table 4 and Figure 12). Thus 0.5 g sample was used throughout the later experiments. No other difference in either single or compound factors was identified in this experiment.

**Table 3** ANOVA to compare sample weight, hexane volumes and incubation times of the final extraction.

| Source | df | MS |
|---|---|---|
| Sample Weight | 1 | 1.6607** |
| Hexane Volume | 1 | 1.8963** |
| Incubation Time | 1 | 0.1188 ns |
| Sample Weight × Hexane Volume | 1 | 0.3300** |
| Sample Weight × Incubation Time | 1 | 0.0038 ns |
| Hexane Volume × Incubation Time | 1 | 0.0140 ns |
| Sample Weight × Hexane Volume × Incubation Time | 1 | 0.0871 ns |
| Error | 24 | 0.0337 |
| Total | 31 | |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.

ns = non- significant difference

**Table 4** Total oil percentage extracted from 2 sample sizes and 2 total hexane volumes.

| Ground soybean sample (g) | Total hexane volume (ml) | | Average[1/] |
|---|---|---|---|
| | 4, 4 | 4, 4, 4 | |
| 0.5 | 20.67 | 20.95 | 20.81 a |
| 1.0 | 20.01 | 20.70 | 20.36 b |
| Average[1/] | 20.34 b | 20.83 a | |

[1/] Means of the sample sizes and hexane volumes followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

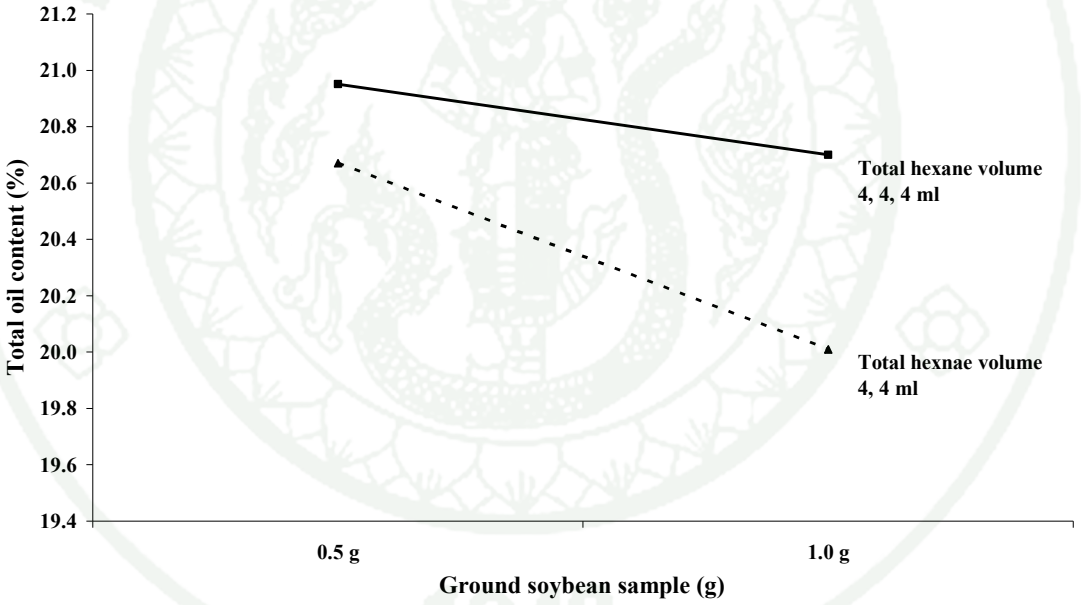



**Figure 12** Interaction between total hexane volume and ground soybean weight affecting the amount of extracted soybean oil.

3.2 Comparison of sample types.

When fine and coarse seed samples were extracted with the same solvent and condition, the results showed interaction between sample type and variety (Table 5). Also, the fine samples gave over 3 times more total oil than the coarse samples (Table 6 and Figure 13). Thus the fine samples were used in the later experiments.

**Table 5** ANOVA of comparison between fine and coarse of sample

| Source | df | MS |
|---|---|---|
| variety | 2 | 2.9028** |
| type | 1 | 674.2089** |
| variety × type | 2 | 2.6455** |
| Error | 15 | 0.0478 |
| Total | 20 | |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.
ns = non- significant difference

**Table 6** Total oil percentage extracted from different sample types and varieties.

| Sample type | Varieties | | | Average[1/] |
|---|---|---|---|---|
| | CM 60 | KUSL 20004 | SJ 5 | |
| Fine | 18.20 | 19.48 | 18.94 | 18.87 a |
| Coarse | 5.51 | 4.94 | 7.46 | 5.96 b |
| Average[1/] | 11.85 b | 12.20 b | 13.19 a | |

[1/] Means of the sample types and varieties followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

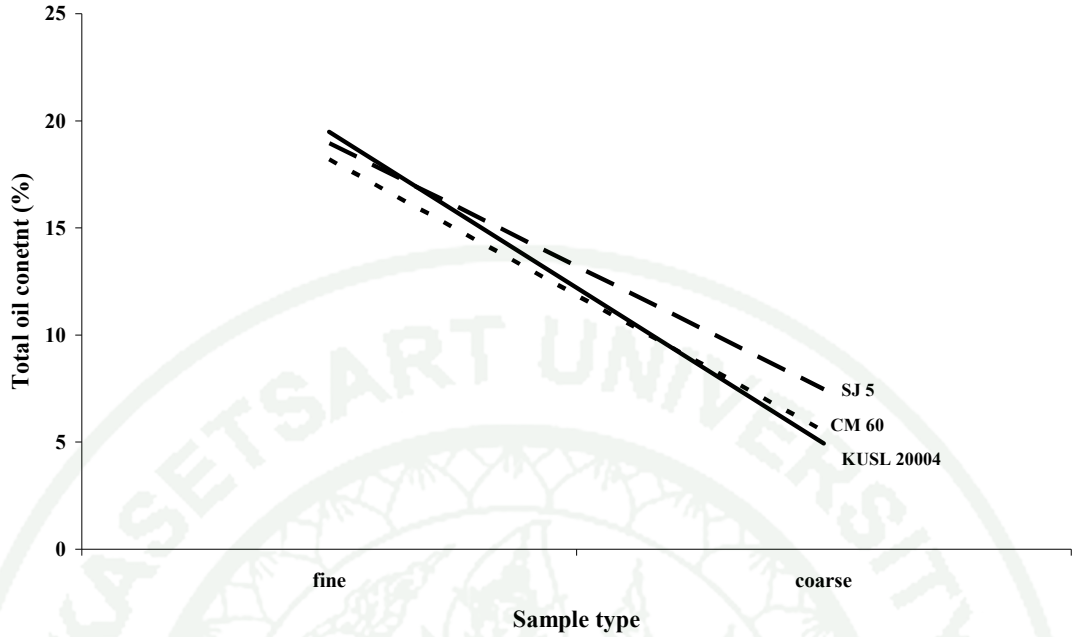



**Figure 13** Interaction between fine and coarse of soybean type affecting the amount of extracted soybean oil

3.3 Comparison between shaking vs no shaking of samples.

Based on the amount of oil extracted from 0.5 g fine ground seed samples of each treatment, shaking resulted in more oil extracted than no shaking and 15 ml total hexane volume gave the best oil yield as would be expected (Table 7). However, a similar amount of total oil was obtained from the total hexane volume between 9 to 15 ml (Table 8, 9 and Figure 14). Thus we recommended using 9 ml total hexane to save chemicals and extraction cost. In addition, extraction cycle was another factor influencing the results. When 2 extraction steps were compared using the same hexane volume between 4 ml with 2 cycles and 8 ml in one cycle, the result showed that 2 extraction cycles resulted in more extracted oil.

**Table 7** ANOVA of comparison between shaking and no shaking of sample.

| Source | df | MS |
|---|---|---|
| variety | 2 | 5.5418** |
| shaking | 1 | 237.0169** |
| treatment | 5 | 58.6122** |
| variety shaking | 2 | 0.1145 ns |
| variety × treatment | 10 | 0.0951* |
| shaking × treatment | 5 | 2.0935** |
| variety × shaking × treatment | 10 | 0.0650 ns |
| Error | 102 | 0.04844 |
| Total | 137 | |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.
ns = non- significant difference.

**Table 8** Total oil percentage extracted from different incubation shaking conditions.

| Incubation condition | Shaking condition | | Average[1/] |
|---|---|---|---|
| | no shaking | shaking | |
| 4 ml, O/N | 13.97 | 17.24 | 15.53 e |
| 5 ml, O/N | 14.13 | 17.63 | 15.88 d |
| 2 ml, 3 h// 2 ml, 3 h// 2 ml, O/N | 17.39 | 19.82 | 18.55 c |
| 3 ml, 3 h// 3 ml, 3 h// 3 ml, O/N | 17.69 | 19.99 | 18.84 b |
| 4 ml, 3 h// 4 ml, 3 h// 4 ml, O/N | 17.90 | 19.97 | 18.87 ab |
| 5 ml, 3 h// 5 ml, 3 h// 5 ml, O/N | 17.84 | 20.00 | 18.98 a |
| Average[1/] | 16.45 b | 19.10 a | |

[1/] Means of the incubation and shaking conditions followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

**Table 9** Interaction between incubation condition and varieties from the experiment of comparison shaking and no shaking of sample.

| Incubation condition | Varieties | | | Average[1/] |
|---|---|---|---|---|
| | CM 60 | KUSL 20004 | SJ 5 | |
| 4 ml, O/N | 15.71 | 15.58 | 15.33 | 15.53 e |
| 5 ml, O/N | 16.16 | 15.86 | 15.63 | 15.88 d |
| 2 ml, 3 h// 2 ml, 3 h// 2 ml, O/N | 18.86 | 18.68 | 18.13 | 18.55 c |
| 3 ml, 3 h// 3 ml, 3 h// 3 ml, O/N | 19.10 | 19.11 | 18.38 | 18.84 b |
| 4 ml, 3 h// 4 ml, 3 h// 4 ml, O/N | 19.21 | 19.07 | 18.62 | 18.87 ab |
| 5 ml, 3 h// 5 ml, 3 h// 5 ml, O/N | 19.21 | 18.84 | 18.56 | 18.98 a |
| Average[1/] | 18.09 a | 17.76 b | 17.42 c | |

[1/] Means of the incubation conditions and soybean varieties followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

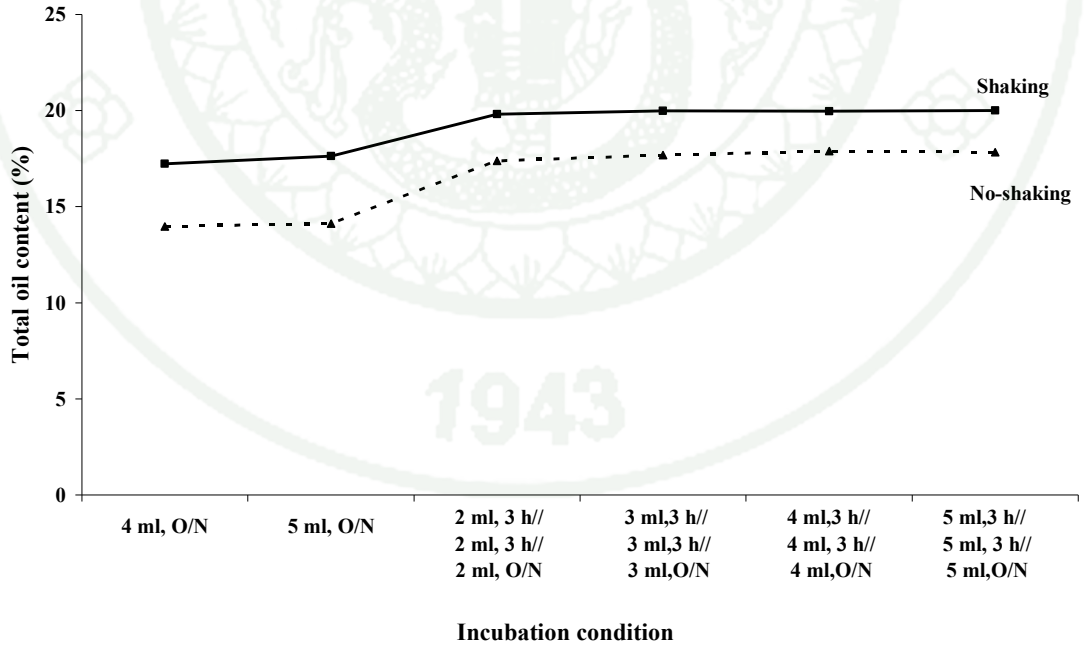



**Figure 14** Interaction between shaking and incubation times affecting amount of extracted soybean oil.

3.4 Comparison between the final extraction times.

The result from our experiment showed that the ratio of soybean sample and solvent volume 1:6 to 1:8 was an important factor in solvent extraction. When hexane extraction was incubated O/N the amount of oil was higher than 3 h extraction, while 9 ml total hexane volume gave the same amount of oil as extracted by 12 ml total hexane volume (Table 10). Thus 9 ml of total hexane is recommended here to save cost and reduce extraction waste. The result revealed that 6 fold of solvent (1:6 of soybean sample: solvent) was sufficient than 8 fold to extract oil in this experiment, with an interaction detected between varieties and the final incubation time (Table 10). Incubation overnight resulted in the highest oil content (Table 11 and Figure 15). In this experiment, CM 60 and KUSL 20004 gave significantly higher total oil than SJ 5.

**Table 10** ANOVA to compare soybean varieties extraction times and extraction volumes in the last cycle

| Source | df | MS |
|---|---|---|
| variety | 2 | 1.2586 ** |
| overnight | 1 | 7.6946 ** |
| volume | 1 | 0.0043 ns |
| variety × overnight | 2 | 0.3033 ** |
| variety × volume | 2 | 0.0424 ns |
| overnight × volume | 1 | 0.0006 ns |
| variety × overnight × volume | 2 | 0.0080 ns |
| Error | 35 | 0.0232 |
| Total | 46 | |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.

ns = non- significant difference

**Table 11** Total oil percentage extracted from 3 soybean varieties using 2 different final extraction times in the last step.

| Extraction time | Varieties | | | Average[1/] |
|---|---|---|---|---|
| | CM 60 | KUSL 20004 | SJ 5 | |
| 3 h | 19.16 | 19.39 | 18.97 | 19.17 b |
| ≥ 16 h | 20.28 | 20.13 | 19.54 | 19.98 a |
| Average[1/] | 19.72 a | 19.74 a | 19.25 b | |

[1/] Means of the final extraction times and varieties followed by the same letter are not significant difference as compared by Duncan's Multiple Range Test (DMRT) at $P \leq 0.05$.

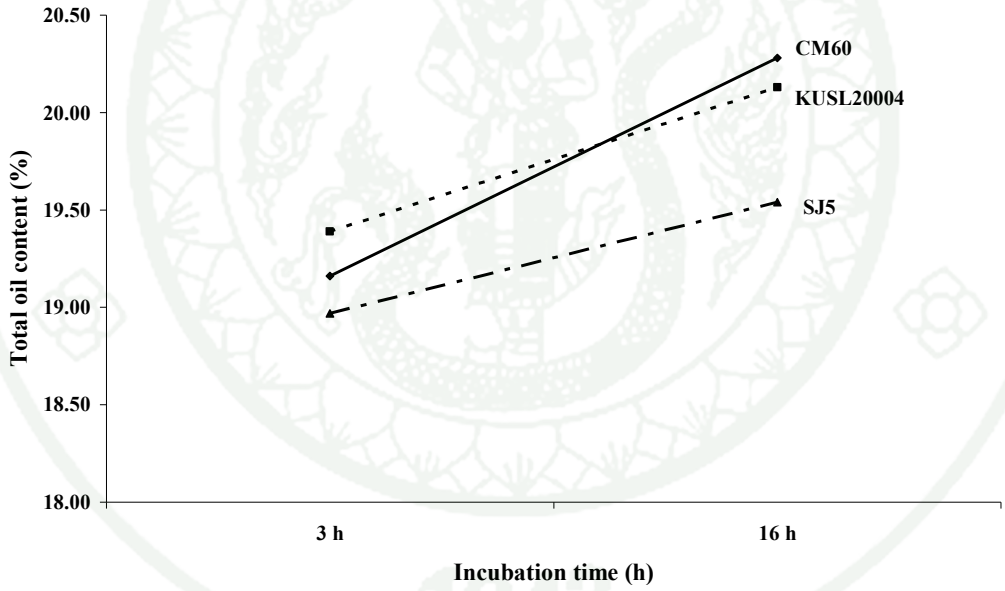



**Figure 15** Interaction between varieties and incubation times affecting amount of extracted soybean oil

**4. Correlation between fatty acid compositions obtained from 2 esterification methods**

The results of fatty acid analysis from both AOCS and USDA-ARS esterification methods indicated that there were five major fatty acids in soybean seed, viz. palmitic, stearic, oleic, linoleic and linolenic acids, same as those reported earlier (Dahmer *et al*., 1989; Wilson, 1996; Wilson, 2004; Hou *et al*., 2006; Shibata *et al*., 2008). Based on coefficient of variability (CV), more variation among accessions was identified in the compositions of stearic, oleic and linolenic acids; medium variation in linoleic acid and less variation in palmitic acid (Table 12). The correlations between the amounts extracted from both methods were medium for palmitic, stearic and linolenic acids and high for oleic and linoleic acids (Table 13 and 14).

The result of correlation among fatty acid composition in each method (Table 13 and 14) revealed that palmitic acid was positively correlated with saturated fatty acids (AOCS method: r = 0.86 and USDA-ARS method: r = 0.79), but negatively correlated with unsaturated fatty acids (AOCS method: r = -0.86, and USDA-ARS method: r = -0.79). Oleic acid content was negatively correlated with linoleic acid content (AOCS method: r = -0.93, and USDA-ARS method: r = -0.99), because oleic acid is the intermediate of fatty acid pathway, during changing from palmitic to stearic, oleic, linoleic and linolenic acids in the respective order. Thus the amount of oleic acid decreased when linoleic acid increased.

**Table 12** Percentage of major fatty acids in oil extracted from 65 soybean samples and analyzed by AOCS and USDA-ARS methods. T-values of a paired-comparison and correlations between fatty acid percentages obtained from both methods were also shown.

| | Mean $_{AOCS}$ | $CV_{AOCS}$ (%) | Mean $_{USDA-ARS}$ | $CV_{USDA-ARS}$ (%) | t-value | r | $R^2$ |
|---|---|---|---|---|---|---|---|
| Palmitic acid (C16:0) | 13.49 ± 1.39 | 10.3 | 11.48 ± 0.73 | 6.3 | -11.18* | 0.54** | 0.29 |
| Stearic acid (C18:0) | 2.61 ± 0.79 | 30.3 | 3.26 ± 0.55 | 16.9 | 5.48* | 0.42** | 0.18 |
| Oleic acid (C18:1) | 27.83 ± 6.32 | 22.7 | 32.49 ± 7.22 | 22.2 | 4.00* | 0.83** | 0.69 |
| Linoleic acid (C18:2) | 48.69 ± 5.70 | 11.7 | 46.54 ± 6.47 | 13.9 | -2.05* | 0.85** | 0.72 |
| Linolenic acid (C18:3) | 7.03 ± 1.80 | 25.6 | 6.05 ± 0.90 | 14.9 | -4.02* | 0.52** | 0.27 |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.

t .05 at df 64 = 1.66, r .01 at df 63 = 0.32

**Table 13** Correlation between total oil content extracted by hexane extraction method and fatty acid composition by AOCS method from 65 soybean varieties.

| AOCS | Stearic acid | Total saturated fatty acids | Oleic acid | Linoleic acid | Linolenic acid | Total unsaturated fatty acids | %Total oil $_{hexane}$ |
|---|---|---|---|---|---|---|---|
| Palmitic acid | -0.09 | 0.86** | -0.10 | -0.15 | 0.06 | -0.86** | -0.08 |
| Stearic acid | | 0.42 | -0.44 | 0.45* | -0.26 | -0.42 | 0.51** |
| Total saturated fatty acids | | | -0.30 | 0.08 | -0.08 | -1.00** | 0.17 |
| Oleic acid | | | | -0.93** | -0.28 | 0.31 | -0.28 |
| Linoleic acid | | | | | 0.02 | -0.09 | 0.38 |
| Linolenic acid | | | | | | 0.08 | -0.37 |
| Total unsaturated fatty acids | | | | | | | -0.19 |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.

**Table 14** Correlation between total oil content analyzed by NMR and fatty acid composition by USDA-ARS method from 65 soybean varieties.

| USDA-ARS | Stearic acid | Total saturated fatty acids | Oleic acid | Linoleic acid | Linolenic acid | Total unsaturated fatty acids | %Total oil $_{NMR}$ |
|---|---|---|---|---|---|---|---|
| Palmitic acid | 0.04 | 0.79** | -0.10 | 0.00 | 0.02 | -0.79** | -0.18 |
| Stearic acid | | 0.65** | -0.68** | 0.65** | 0.05 | -0.65** | 0.50** |
| Total saturated fatty acids | | | -0.49 | 0.40 | 0.04 | -1.00** | 0.17 |
| Oleic acid | | | | -0.99** | -0.36 | 0.49* | -0.28 |
| Linoleic acid | | | | | 0.24 | -0.40 | 0.35 |
| Linolenic acid | | | | | | -0.04 | -0.41 |
| Total unsaturated fatty acids | | | | | | | -0.17 |

*, ** = significant difference at $P \leq 0.05$ and 0.01 level of probability, respectively.

## Part II: Development of $F_{2:3}$ population

### 1. Screening total oil content in 65 soybean accessions for choosing parental lines

From the pervious experiment, we compared the total oil extraction method that gave the best result to extract oil content and optimized the suitable condition. Then, we used hexane extraction method to extract total oil content and USDA-ARS method to methylated fatty acid to FAME and analyzed by GC-FID.

This figure 16 and 17 showed histogram of total oil content extracted by hexane extraction method from 65 soybean varieties (the raw data of total oil content was presented in Appendix Table 2). Their results revealed that the range of total oil was 16 to 24 % and soybean had oil content about 20%. Soybean seed could classify soybean group depending on total oil content and fatty acid composition. This information was used to make decision to choose parental line

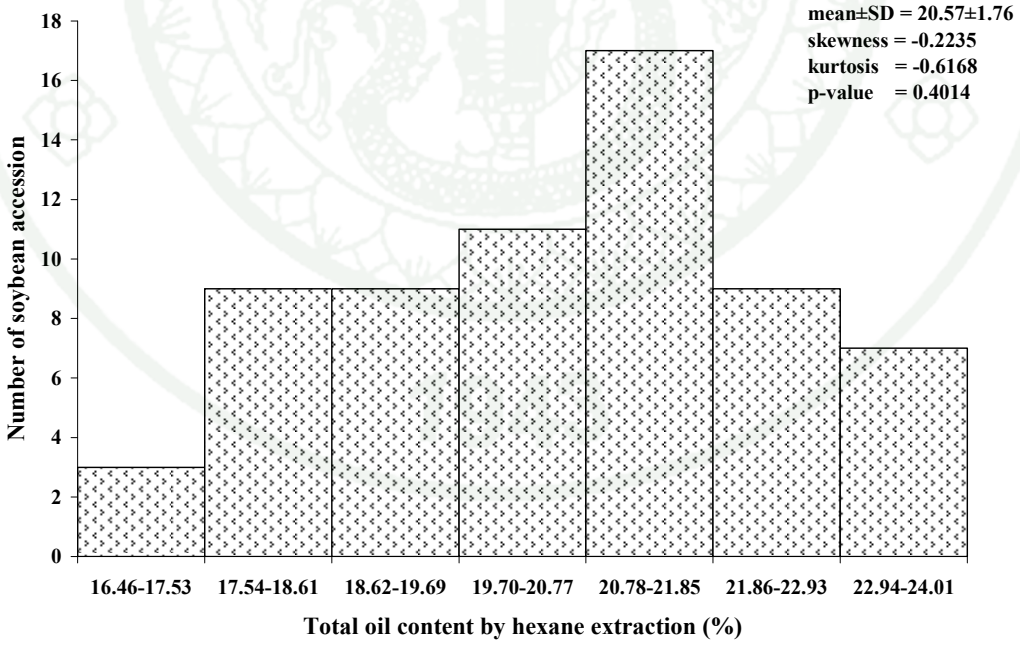



**Figure 16** Frequency distribution of total oil content in 65 soybean varieties analyzed by hexane extraction method.

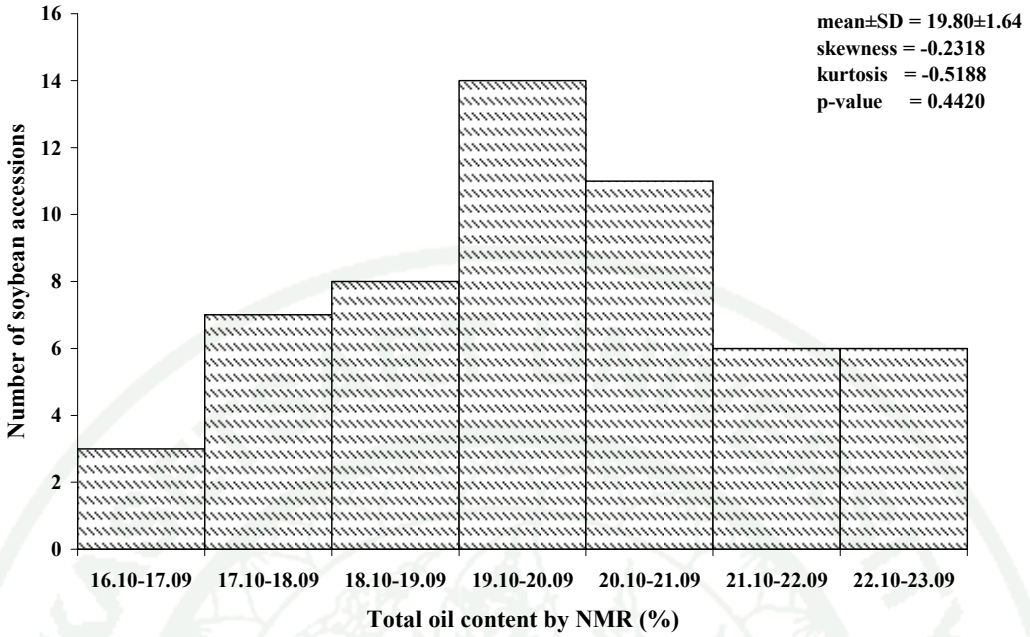



**Figure 17** Frequency distribution of total oil content in 65 soybean varieties analyzed by NMR technique.

The fatty acid composition varied depending on soybean accessions used in our study (the raw data of fatty acid composition was presented in Appendix Table 3). AOCS method gave significantly higher stearic and oleic acids but less palmitic, linoleic and linolenic acids (Figure 18 and19).

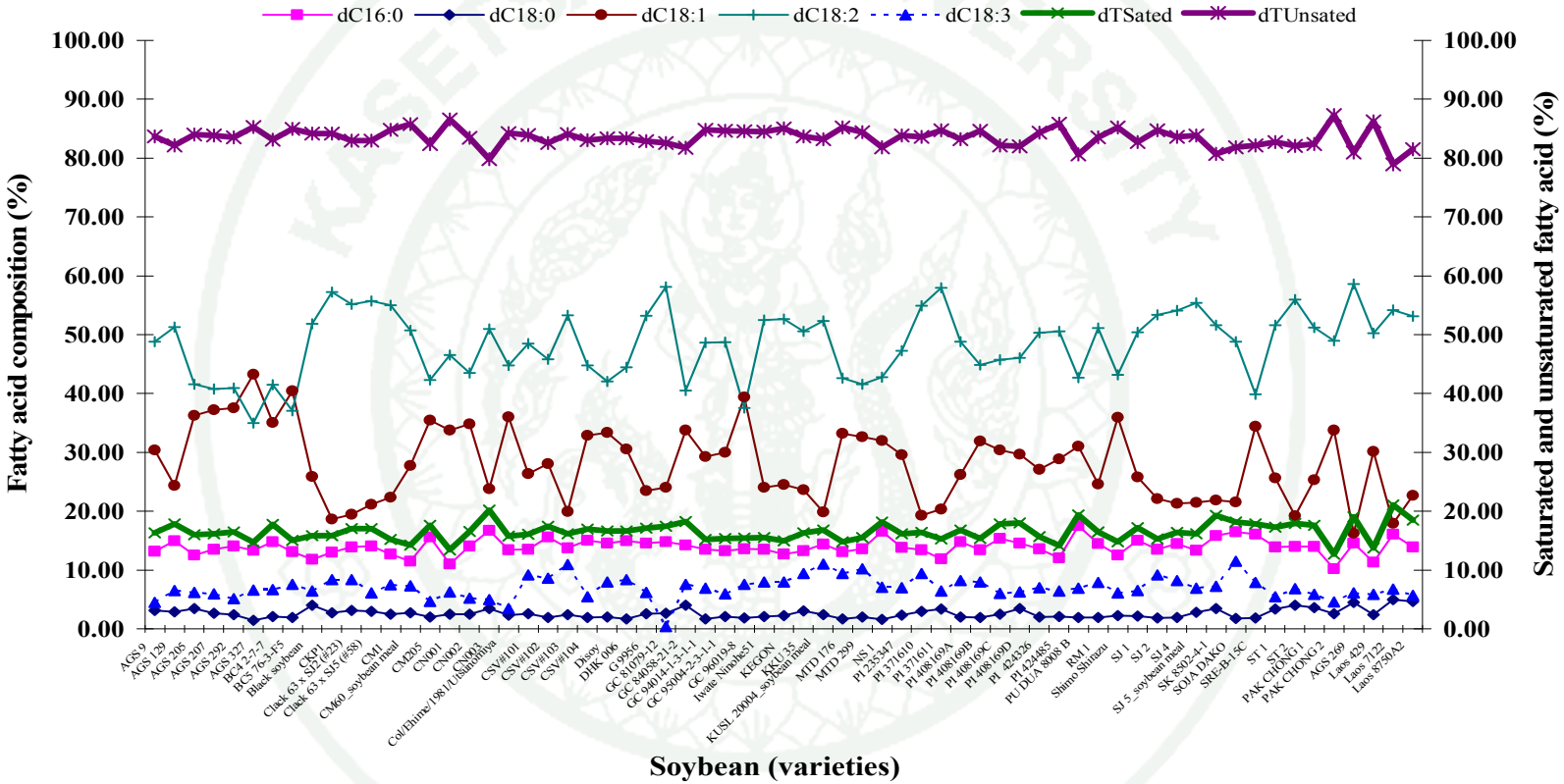



**Figure 18** Fatty acid composition determined from AOCS method in soybean seeds from 65 accessions: dC16:0=palmitic acid, dC18:0=stearic acid, dC18:1=oleic acid, dC18:2=linoleic acid, dC18:3=linolenic acid, dTSaturated=total saturated fatty acids and dTUunsaturated=total unsaturated fatty acids.

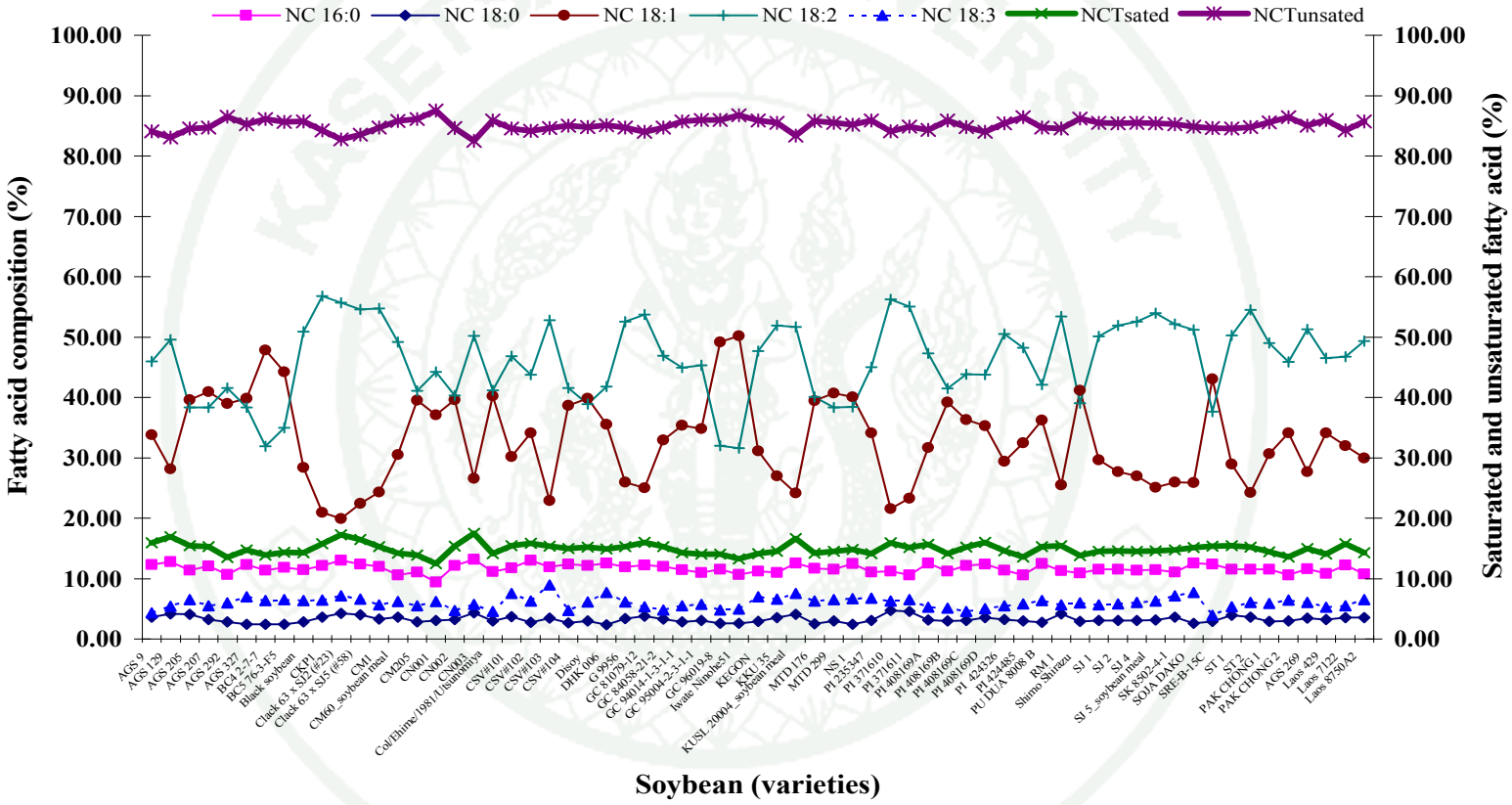



**Figure 19** Fatty acid composition determined from USDA-ARS method in soybean seeds from 65 accessions: NC16:0=palmitic acid, NC18:0=stearic acid, NC18:1=oleic acid, NC18:2=linoleic acid, NC18:3=linolenic acid, NCTsaturated=total saturated fatty acids and NCTunsaturated=total unsaturated fatty acids.

**2. Crossing and develop population**

From the result of total oil content and fatty acid composition in soybean varieties, we choose four parental lines and planed to crossing between them including CSV #103 × Laos 7122 (the interesting traits were C18:0 and C18:3), PI 371611 × Pak Chong 2 (the interesting traits were C18:1 and C18:2) and Pak Chong 2 × Laos 7122 (the interesting traits were C18:0, C18:1 and C18:3). Fatty acids were expressed as percentages of the total fatty acids. The best one of the parent was Pak Chong 2 and Laos 7122 because this parental line has more than one trait different. The fatty acid profiles of the parents were presented in Table 15. Finally, we got two seeds from one $F_1$ pod from Pak Chong 2 and Laos 7122. Then, planted them in the separate containers for generate $F_2$ seed and harvesting individual plant.

**Table 15** Total oil content and fatty acid composition in 4 soybean varieties, which choosing from 65 soybean accessions for crossing and developing $F_2$ population.

| Varieties | % Fatty acid composition in soybean seed [1/] | | | | | |
|---|---|---|---|---|---|---|
| | C16:0 | C18:0 | C18:1 | C18:2 | C18:3 | %Total oil [2/] |
| CSV #103 | 13.76 | 2.42 | 19.93 | 53.26 | 10.93 | 17.24 |
| PI 371611 | 11.90 | 3.39 | 20.30 | 57.98 | 6.43 | 23.58 |
| Pak Chong 2 | 10.19 | 2.54 | 33.77 | 48.95 | 4.55 | 18.70 |
| Laos 7122 | 16.11 | 4.99 | 17.94 | 54.21 | 6.75 | 21.59 |

[1/] Fatty acid composition was analyzed by USDA-ARS method.

[2/] Total oil content was extracted by hexane extraction method.

We planted 600 $F_2$ seeds in the tray and transplanted in separate pot. Some of them could not germinate and died. Finally, we got 300 lines and planted $F_3$ seeds in the field for collect $F_4$ seed on $F_3$ plants. Sampling an $F_2$ population comprising of 186 plants derived from a cross between Pak Chong 2 and Laos 7122 was used. $F_2$

leaves were collected and extracted DNA. These $F_2$ DNA were used to screen by polymorphic SSR markers and recorded as genotypic data.

**3. Parental screening by SSR primer on 5% denaturing urea polyacrylamide gel electrophoresis**

Four hundred and twenty-eight SSR primers were used to survey in the parental lines between female and male parental line (Pak Chong 2 and Laos 7122, respectively). Then, we found two hundred and forty SSR markers polymorphic between parents. Only one hundred and fifty-nine SSR markers showed polymorphism in 186 individual of $F_{2:3}$ population. All of these SSR were co-dominant marker.

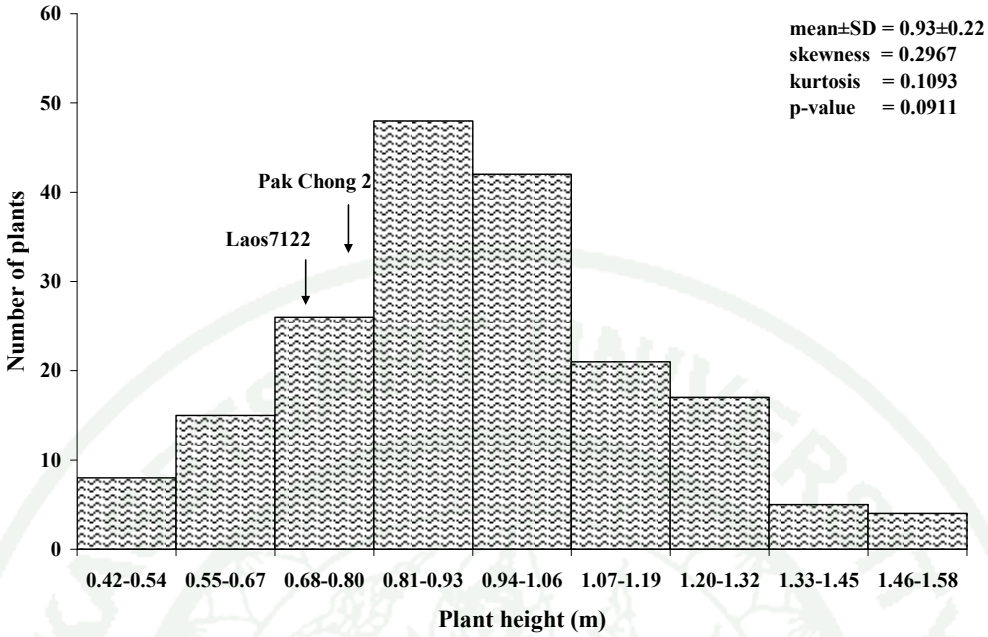
## Part III: QTL analysis

**1. Phenotypic analysis**

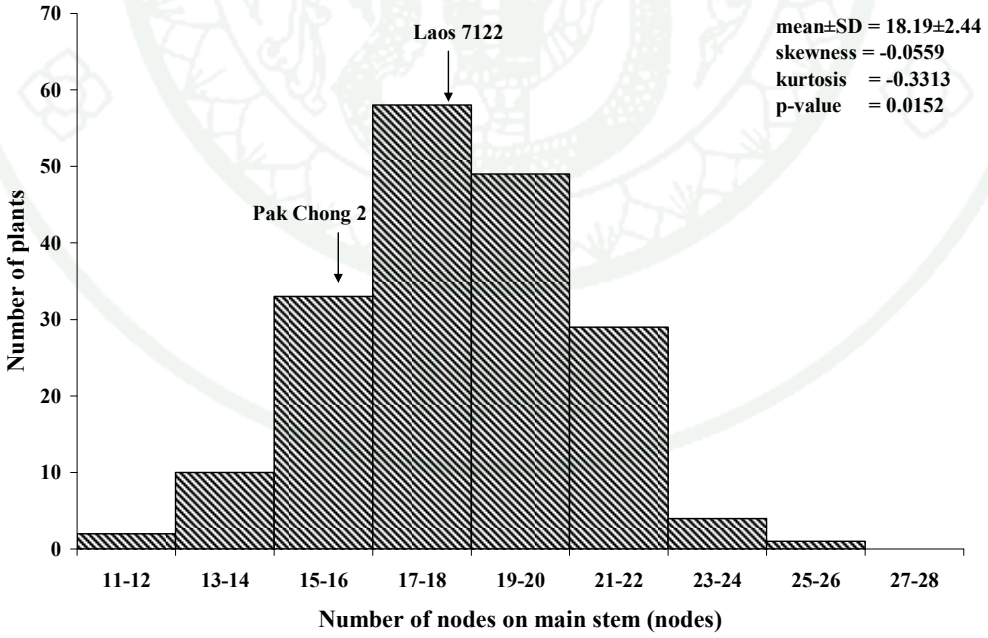
Table 16 presented the range of all traits in $F_{2:3}$ population from crossing Pak Chong 2 × Laos 7122. All traits in the $F_{2:4}$ population showed continuous distribution (Figure 20-30). Plant height, seed length, seed width, one hundred seed weight, concentration of palmitic, stearic, and linolenic acids were normally distributed, whereas the other traits skewed toward one end. The traits with normal distribution varied continuously from 0.43-1.57 m (Figure 20), 7.17-8.41 mm (Figure 22), 5.91-6.78 mm (Figure 23) and 12.52-19.62 g (Figure 24), 11.29-17.20% (Figure 26), 2.37-4.40% (Figure 27) and 3.49-6.90% (Figure 30) of fatty acid composition, respectively. The normality test of the $F_{2:3}$ population indicated that number of nodes on main stem and total oil content, oleic acid and linoleic acid were not normally distributed with the range of 12-25 nodes (Figure 21) and 15.03-25.31% (Figure 25), 16.66-50.72% (Figure 28) and 31.43-60.50% (Figure 29), respectively. We found the transgressive distortion in all traits.

**Table 16** Mean and range of plant height, number of nodes on main stem, seed length, seed width, 100-seed weight and total oil content, palmitic, stearic, oleic, linoleic and linolenic acid composition in oil of $F_{2:3}$ population from the soybean cross Pak Chong 2 × Laos 7122.

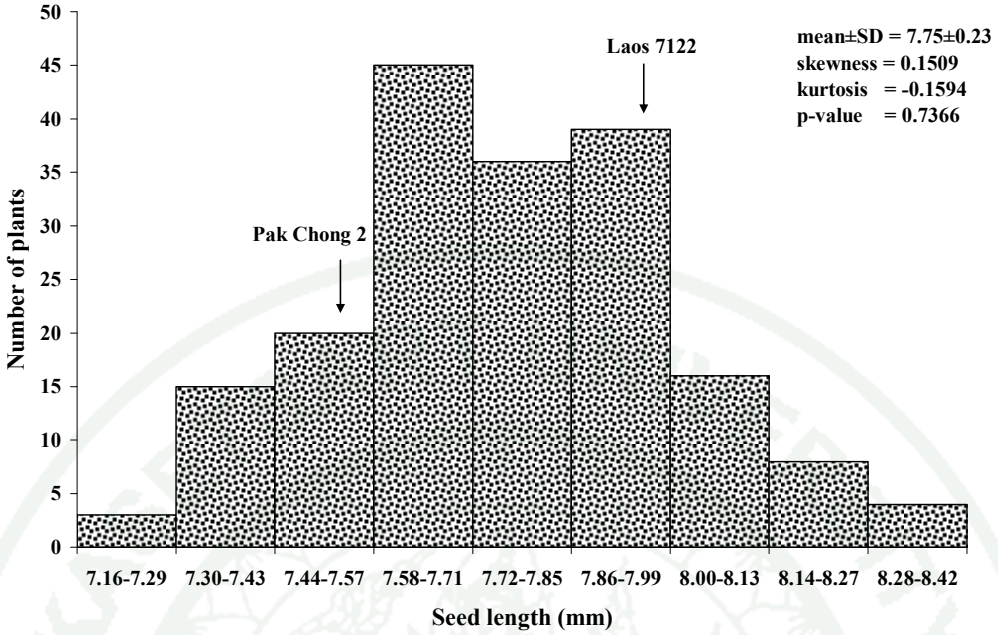
| Trait | Mean±SD | | | Population range |
|---|---|---|---|---|
| | Pak Chong 2 | Laos 7122 | $F_{2:3}$ Population | |
| Plant height (mm) | 0.77±0.03 | 0.70±0.04 | 0.93±0.22 | 0.43-1.57 |
| Number of nodes | 16±1 | 18±1 | 18±2 | 12-25 |
| Seed length (mm) | 7.51±0.05 | 7.90±0.35 | 7.75±0.23 | 7.17-8.41 |
| Seed width (mm) | 6.22±0.05 | 6.59±0.08 | 6.34±0.18 | 5.91-6.78 |
| 100 seed weight (g) | 14.49±1.02 | 17.96±0.93 | 15.56±1.41 | 12.52-19.62 |
| Total oil content (%) | 18.69±0.21 | 21.75±0.53 | 18.88±1.41 | 15.03-25.31 |
| Palmitic acid (%) | 12.57±0.03 | 13.28±0.33 | 14.27±1.14 | 11.29-17.20 |
| Stearic acid (%) | 3.38±0.11 | 3.32±0.08 | 3.10±0.34 | 2.37-4.40 |
| Oleic acid (%) | 20.89±3.23 | 33.25±2.17 | 28.81±7.11 | 16.66-50.72 |
| Linoleic acid (%) | 56.80±3.80 | 46.29±1.96 | 48.77±5.98 | 31.43-60.50 |
| Linolenic acid (%) | 6.36±0.72 | 3.86±0.02 | 5.02±0.68 | 3.49-6.90 |



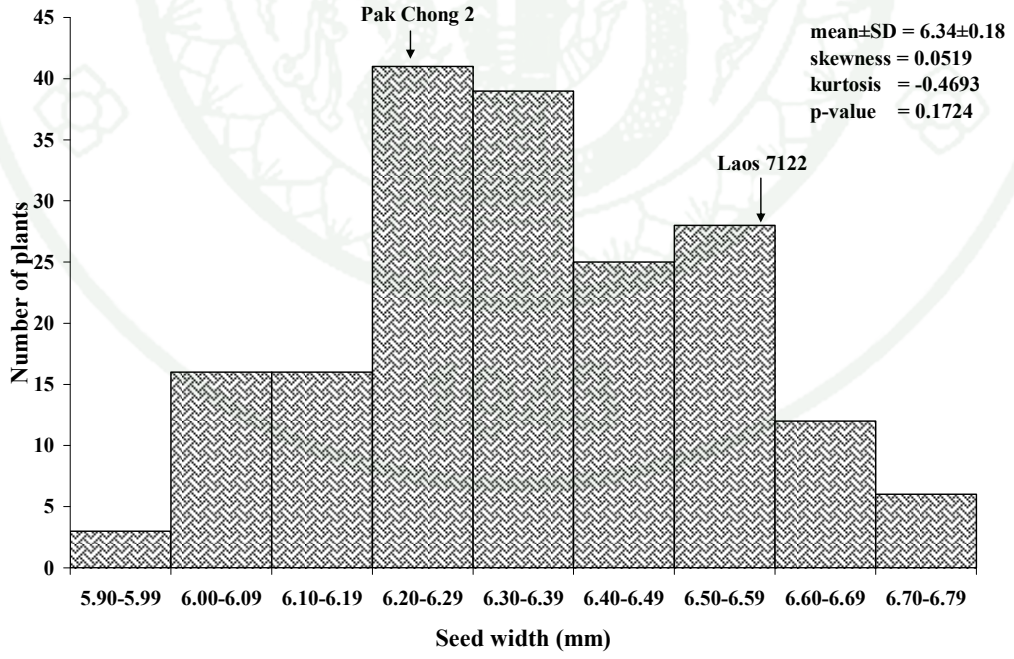

**Figture 20** Frequency distribution of plant height in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



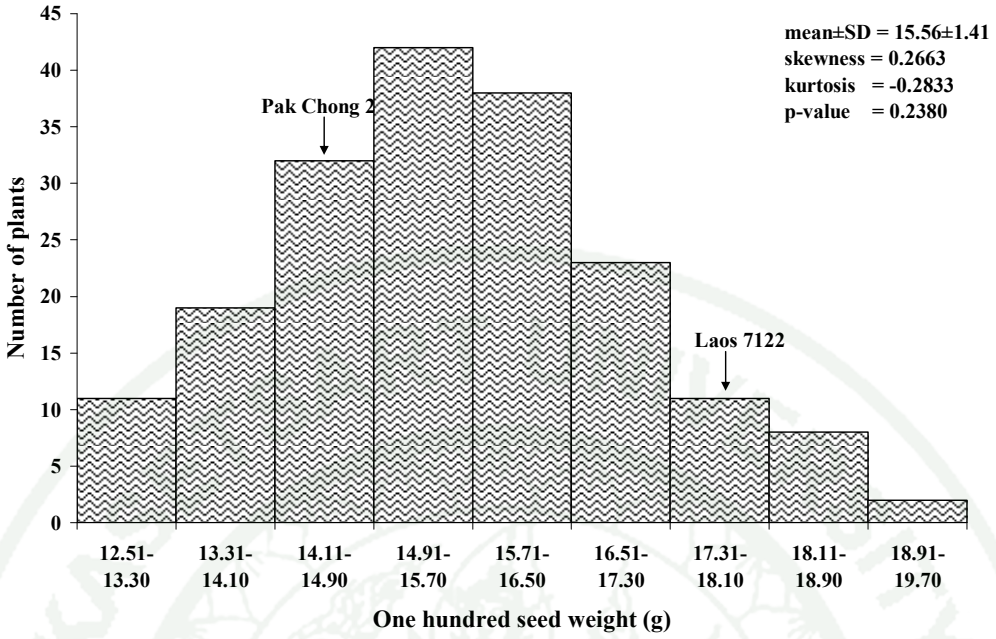

**Figture 21** Frequency distribution of number of nodes on main stem per plant in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



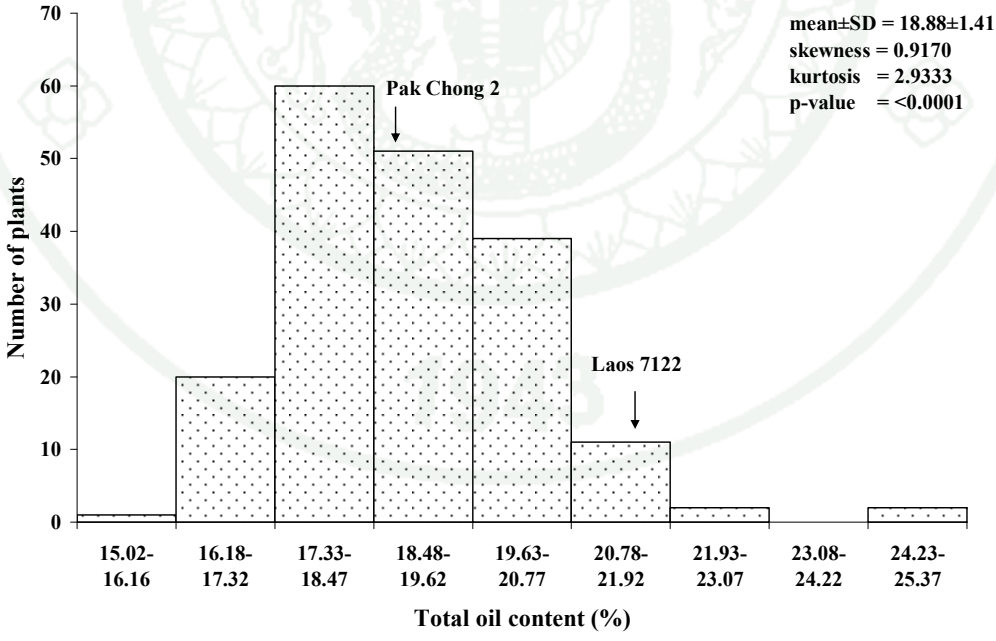

**Figture 22** Frequency distribution of seed length in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



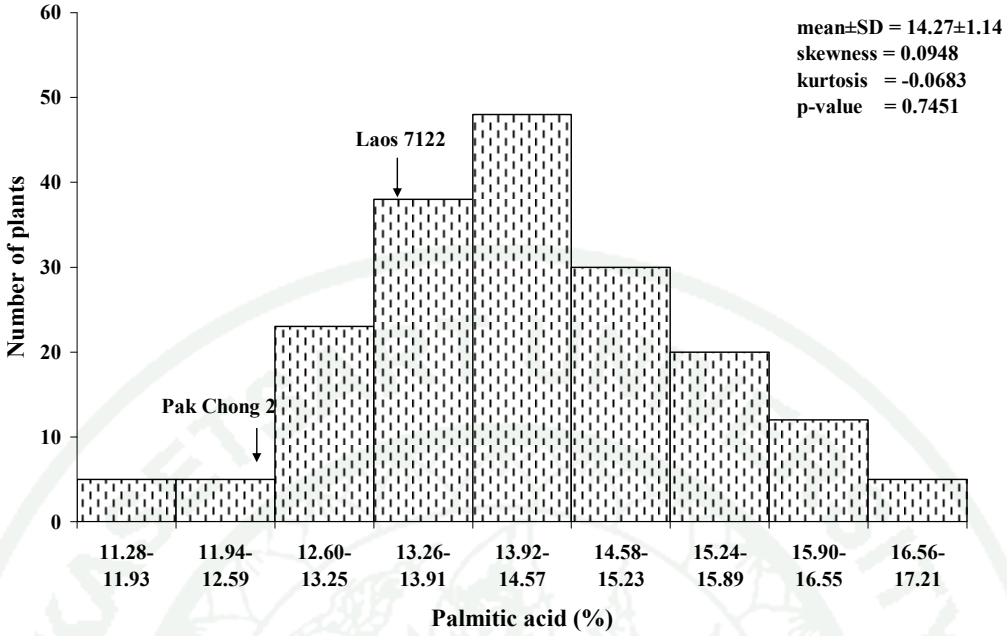

**Figture 23** Frequency distribution of seed width in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



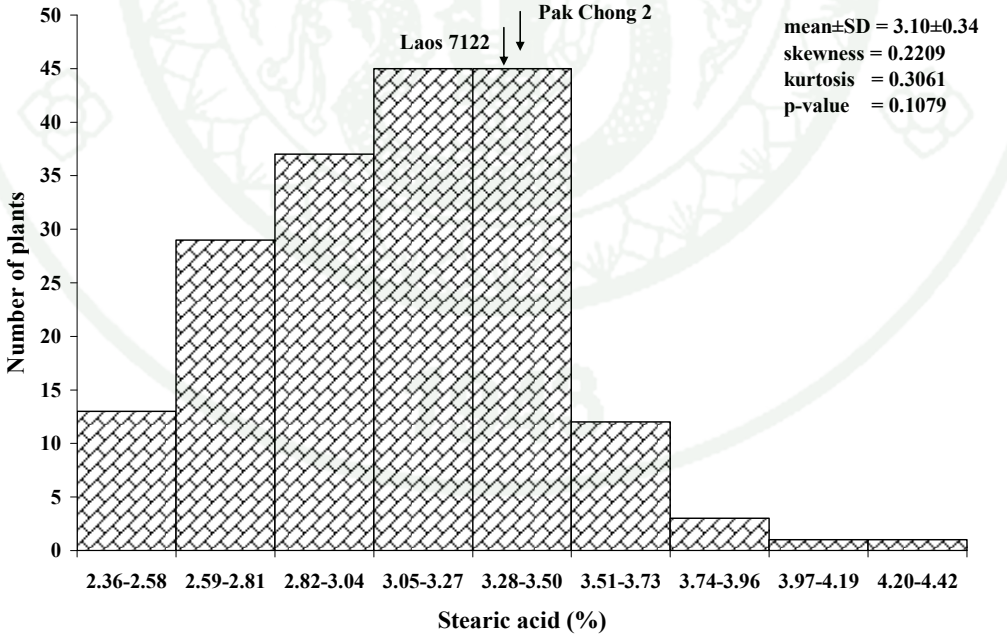

**Figture 24** Frequency distribution of 100-seed weight in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



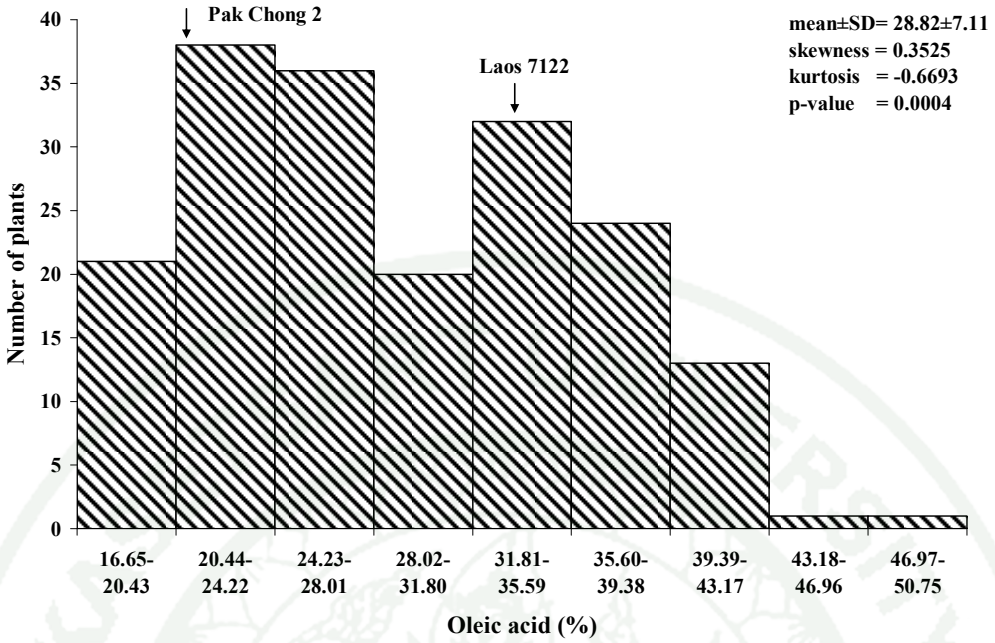

**Figture 25** Frequency distribution of total oil content in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



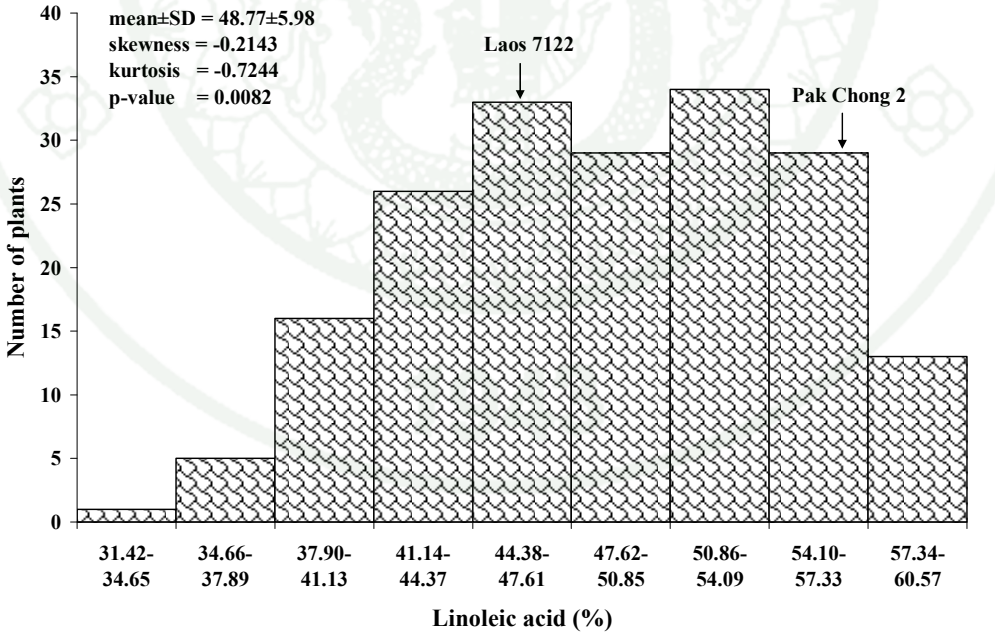

**Figture 26** Frequency distribution of palmitic acid (C16:0) percentage in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.



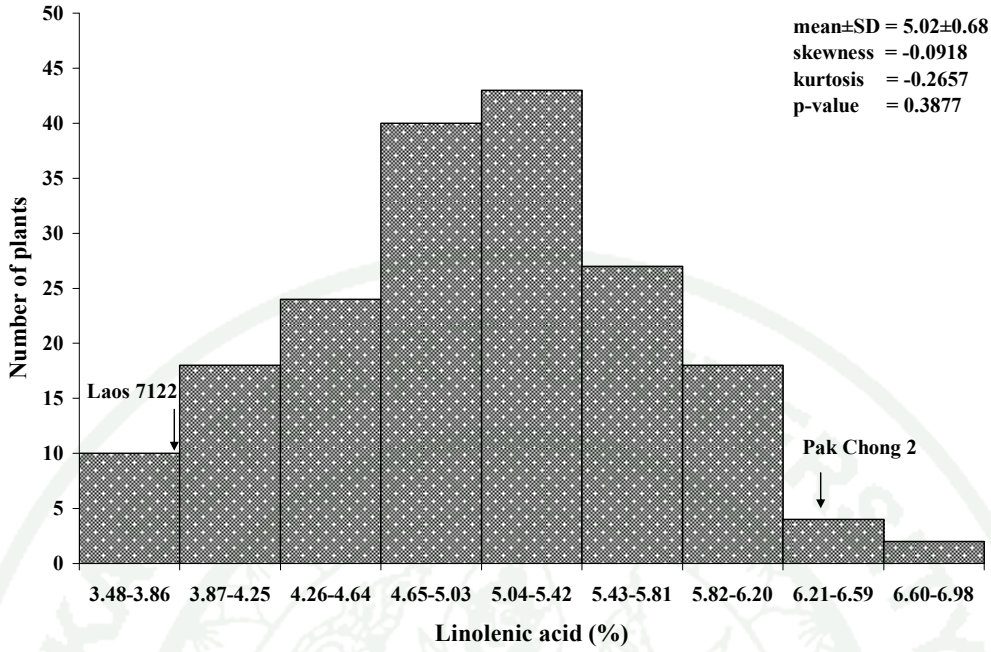

**Figture 27** Frequency distribution of stearic acid (C18:0) percentage in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.




**Figture 28** Frequency distribution of oleic acid (C18:1) percentage in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.




**Figture 29** Frequency distribution of linoleic acid (C18:2) percentage in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.




**Figture 30** Frequency distribution of linolenic acid (C18:3) percentage in $F_4$ seed from the population crossing between Pak Chong 2 and Laos 7122.

## 2. Genotypic analysis

$F_1$ leaves were collected and extracted DNA to confirm $F_1$ hybrid. Some SSR which represented the polymorphism between parents were used in this experiment. The $F_2$ population used was derived from the cross between Pak Chong 2 and Laos 7122. Parental DNA and 300 individual of $F_{2:3}$ lines were good quality because almost samples gave high concentration and purification (Figure 31). The original population consisted of 300 individual $F_{2:3}$ lines. Next, sampling to 186 lines for represent to the population. The stock DNA was diluted with 18 MΩ water for 15 ng and using for PCR reaction in other step of experiment.

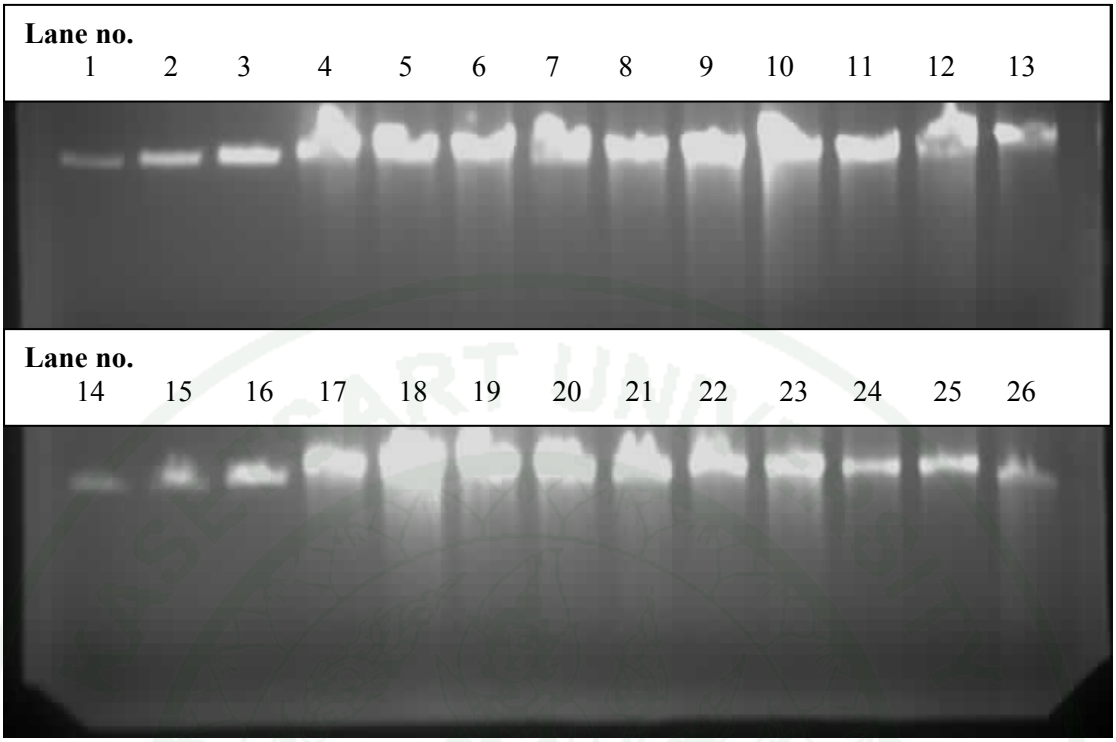



**Figure 31** Crude DNA was extracted by CTAB method and separated on 1.0% agarose gel run for 30 min at 100 V in 1x TBE buffer, stain with ethidium bromide and photographed under UV light. Lanes 1-3 and 14-16 were λ DNA concentration at 50, 100 and 200 ng, respectively. Lanes 4-13 and 17-26 were individual of $F_{2:3}$ DNA population between Pak Chong 2 and Laos 7122.

A total of 428 SSR primers were surveyed in the parental lines between female and male parental line (Pak Chong 2 and Laos 7122). The SSR markers that showed polymorphism would be used in $F_2$ population. From all SSR marker found 240 (56.07% of screened primers) were polymorphic. Of these, 159 markers (66.25% of polymorphic markers) showed polymorphism among 186 $F_{2:3}$ population.

**3. Data analysis**

3.1 The Chi-square test

On the basis of $\chi^2$ test, out of 159 SSR markers, 112 (70.44% of polymorphic markers) showed Mendelian segregation of 1:2:1, while the rest 47 markers had distortion ratios. These SSR markers were used to constructed linkage map (Figure 32).

3.2 Linkage analysis

Among 159 polymorphic SSR markers, 138 were mapped onto 30 linkage groups (LGs) and covered 1,921.1 cM of soybean genome (Figure 32). Twenty-one polymorphic markers remained unlinked. The genetic map represented approximately two third of the consensus soybean map (2,523.6 cM; Song *et al.,* 2004; Soybase, 2005). The linkage groups consist of 2 to 11 SSR markers with an average distance of 14 cM between the adjacent loci. The length of the LGs varied from 13.9 to 218.0 cM. The shortest LG was F-1 (sub-group) while the longest one was D1b. The markers were dispersed but did not cover throughout the soybean genome. There were many gaps up to 50 cM in some LGs such as C1, because the map was constructed from the markers that may not be well-spreading in the soybean genome. Eleven linkage groups were consistent with the consensus map of Cregan *et al.* (1999), including A1, B1, B2, C1, D1b, G, H, I, K, L and N. Eight linkage groups were split into 2 sub-groups, consisting of A2, C2, D2, E, F, J, M and O, while D1a was split into 3 sub-groups (Figure 32).

3.3 Simple regression analysis

The result from regression analysis (%$R^2$) presented in Table 17 and composite interval mapping presented in Table 19. Also, Table 18 presented single regression analysis of ungrouped SSR marker. Both simple regression and CIM were used to identify SSR marker associated with all traits.

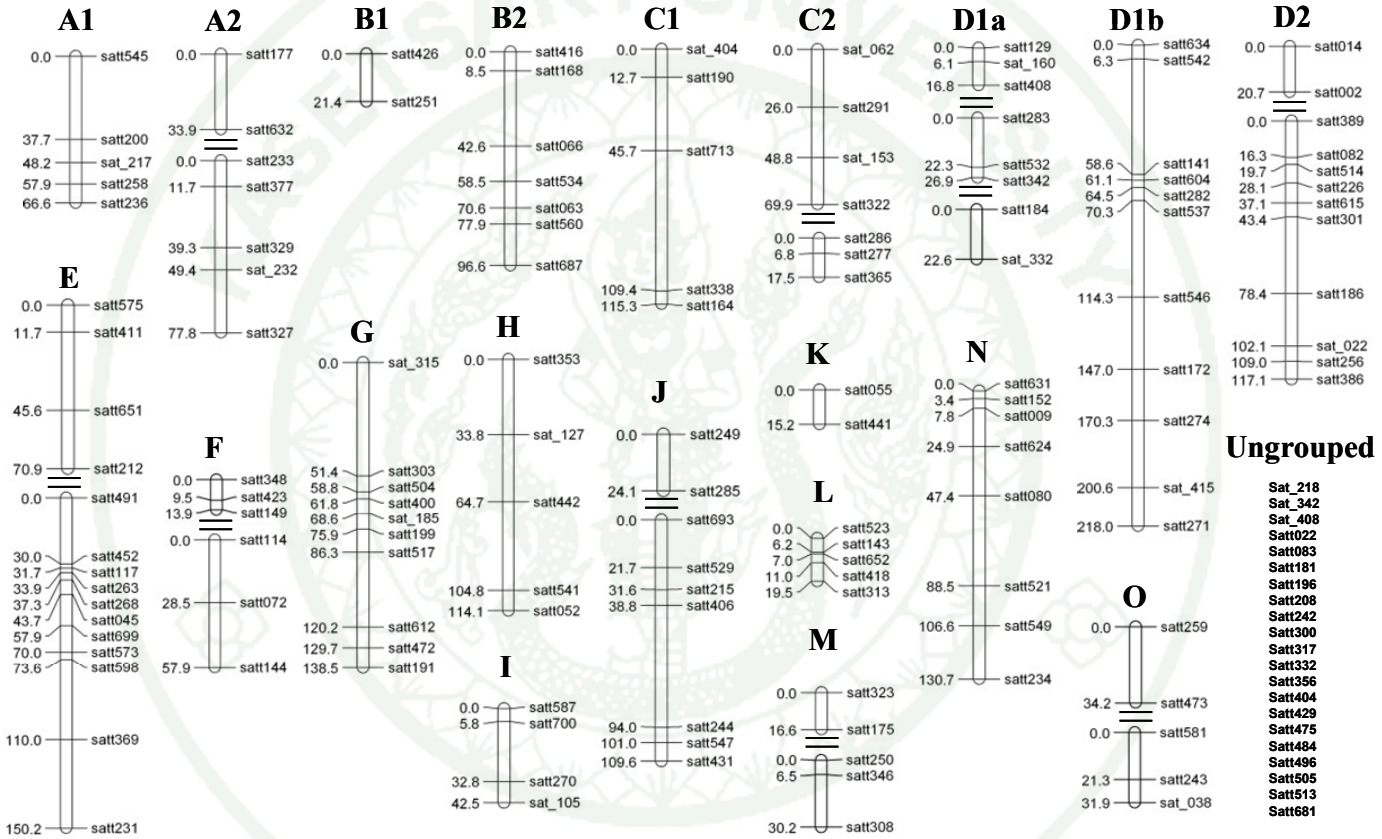



**Figure 32** Soybean genetic linkage map comprised of 138 SSR markers grouping into 30 linkage group from $F_{2:3}$ population crossing from Pak Chong 2 × Laos 7122. This map showed marker position and estimated distance (cM) on the left-hand side and marker name on the right-hand side. Twenty-one SSR markers remained ungrouped. The map length was 1,921.1 cM.

**Table 17** The candidate SSR markers linking to 11 traits were developed from $F_{2:3}$ population crossing between Pak Chong 2 and Laos 7122 calculated from single regression.

| SSR | LG | Trait (%$R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Sat_217 | A1 | ns | ns | ns | ns | ns | 21.99 | ns | 10.93 | 13.30 | 17.46 | ns |
| Satt200 | A1 | ns | ns | ns | ns | ns | 17.95 | ns | 5.10 | 12.65 | 15.57 | ns |
| Satt236 | A1 | 7.90 | ns | ns | ns | ns | 17.30 | ns | 5.78 | 12.11 | ns | ns |
| Satt258 | A1 | 6.12 | ns | ns | ns | ns | 23.92 | 5.17 | ns | 10.51 | 14.28 | ns |
| Sat_332 | A2 | ns | ns | ns | 5.97 | ns | ns | ns | ns | ns | ns | ns |
| Satt083 | B2 | ns | ns | 5.64 | 6.22 | ns | ns | ns | ns | ns | ns | ns |
| Sat_062 | C2 | ns | ns | ns | 7.36 | ns | ns | ns | ns | ns | ns | ns |
| Sat_153 | C2 | ns | 5.40 | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt277 | C2 | ns | ns | ns | ns | ns | ns | ns | 9.01 | ns | ns | ns |
| Satt286 | C2 | ns | ns | ns | ns | ns | ns | ns | 5.61 | ns | ns | ns |
| Satt291 | C2 | ns | ns | ns | 6.53 | ns | ns | ns | ns | ns | ns | ns |
| Satt365 | C2 | ns | ns | ns | ns | ns | ns | ns | 9.64 | ns | ns | ns |

**Table 17** (continued.)

| SSR | LG | Trait (% $R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Satt537 | D1b | ns | 6.19 | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt082 | D2 | ns | ns | ns | ns | ns | ns | 8.68 | ns | ns | ns | ns |
| Satt226 | D2 | ns | ns | ns | ns | ns | ns | 8.45 | ns | ns | ns | ns |
| Satt301 | D2 | ns | ns | ns | ns | ns | ns | 6.24 | ns | ns | ns | ns |
| Satt389 | D2 | ns | ns | ns | ns | ns | ns | 10.35 | ns | ns | ns | ns |
| Satt514 | D2 | ns | ns | ns | ns | ns | ns | 7.65 | ns | ns | ns | ns |
| Satt615 | D2 | ns | ns | ns | ns | ns | ns | 7.49 | ns | ns | ns | ns |
| Satt045 | E | ns | ns | ns | 6.82 | 7.30 | ns | ns | ns | ns | ns | 5.12 |
| Satt117 | E | ns | ns | ns | 6.45 | ns | ns | ns | ns | ns | ns | ns |
| Satt263 | E | ns | ns | ns | 7.20 | 6.43 | ns | ns | ns | ns | ns | ns |
| Satt268 | E | ns | ns | ns | 5.32 | ns | ns | ns | ns | ns | ns | ns |
| Satt369 | E | ns | ns | ns | 8.22 | 5.29 | ns | ns | ns | ns | ns | 5.99 |
| Satt452 | E | ns | ns | ns | 7.82 | 5.71 | ns | ns | ns | ns | ns | ns |

**Table 17** (continued.)

| SSR | LG | Trait (% $R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Satt573 | E | ns | ns | ns | 5.97 | 8.89 | ns | ns | 7.15 | ns | ns | 5.39 |
| Satt598 | E | ns | ns | ns | 5.65 | 6.72 | ns | ns | 6.91 | ns | ns | ns |
| Satt699 | E | ns | ns | 5.60 | 9.64 | 10.00 | ns | ns | ns | ns | ns | ns |
| Satt348 | F | ns | ns | 7.11 | ns | ns | ns | ns | ns | ns | ns | ns |
| Sat_185 | G | ns | ns | 5.01 | 5.06 | ns | ns | ns | ns | ns | ns | 5.33 |
| Satt191 | G | ns | ns | ns | ns | ns | ns | ns | ns | 9.05 | 7.75 | 6.39 |
| Satt199 | G | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | 5.68 |
| Satt303 | G | ns | ns | 6.97 | 8.22 | 5.76 | ns | ns | ns | ns | ns | 6.00 |
| Satt400 | G | ns | ns | ns | 5.21 | ns | ns | ns | ns | ns | ns | ns |
| Satt472 | G | ns | ns | ns | ns | ns | ns | ns | ns | 7.01 | 6.62 | 5.21 |
| Satt504 | G | ns | ns | 5.87 | 7.22 | 5.66 | ns | ns | ns | ns | ns | 5.24 |
| Satt505 | G | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | 5.72 |
| Satt517 | G | ns | ns | ns | ns | ns | ns | ns | ns | 5.26 | 5.40 | ns |

**Table 17** (continued.)

| SSR | LG | Trait (% $R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Sat_127 | H | ns | ns | ns | ns | ns | ns | ns | ns | 6.68 | 6.07 | 5.74 |
| Sat_218 | H | ns | ns | 5.40 | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt143 | L | ns | ns | ns | 5.04 | ns | ns | ns | ns | ns | ns | ns |
| Satt313 | L | 6.83 | 5.10 | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt418 | L | 6.45 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt523 | L | 5.46 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |

ns = non-significant difference

**Table 18** %$R^2$ of ungrouped SSR markers were screened in $F_{2:3}$ population from Pak Chong 2 × Laos 7122 calculated from single regression.

| SSR | Trait (% $R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Sat_218 | ns | ns | 5.39 | 3.48 | 3.77 | ns | ns | ns | ns | ns | ns |
| Sat_342 | ns | ns | ns | ns | ns | ns | ns | 2.32 | 2.23 | 2.51 | ns |
| Sat_408 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt022 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt083 | ns | ns | 5.64 | 6.21 | 3.42 | ns | ns | ns | ns | ns | ns |
| Satt181 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt196 | 3.21 | ns | ns | ns | ns | 3.83 | ns | ns | ns | ns | ns |
| Satt208 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt242 | ns | ns | ns | ns | ns | ns | ns | ns | 2.30 | ns | ns |
| Satt300 | ns | ns | ns | ns | 2.65 | ns | ns | ns | ns | ns | 2.21 |
| Satt317 | ns | ns | 2.92 | 3.07 | 2.26 | ns | ns | ns | ns | ns | ns |
| Satt332 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt356 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| Satt404 | ns | ns | ns | ns | ns | ns | 4.05 | ns | 3.92 | 3.41 | ns |

**Table 18** (continued)

| SSR | Trait (% $R^2$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Plant height** | **Number of node** | **Seed Length** | **Seed Width** | **100-Seed Weight** | **Total oil content** | **Palmitic acid** | **Stearic acid** | **Oleic acid** | **Linoleic acid** | **Linolenic acid** |
| Satt429 | ns | ns | ns | ns | ns | 2.10 | ns | 3.81 | ns | ns | 2.30 |
| Satt475 | ns | ns | ns | ns | ns | ns | 2.44 | ns | ns | 2.70 | ns |
| Satt484 | ns | ns | ns | ns | ns | ns | ns | 4.03 | ns | ns | ns |
| Satt496 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | 2.30 |
| Satt505 | ns | ns | 3.00 | 3.95 | 3.32 | ns | ns | ns | 4.58 | 4.23 | 5.72 |
| Satt513 | 2.11 | 3.46 | ns | ns | ns | ns | ns | ns | ns | ns | 2.11 |
| Satt681 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |

ns = non-significant difference

3.4 QTL analysis

Figure 31 presents seven linkage groups consisting of A1, C2, D2, E2, G, H and O, harboring 20 QTLs of all traits from this research. The result showed that no QTL conferring the contents was found on the other 13 linkage groups, viz. A2, B1, B2, C1, D1a, D1b, F, I, J, K, L, M and N. From 11 traits, no QTL of plant height but at least one QTL was detected for the other traits.

The results of simple regression analysis of plant height showed that there were two SSR markers on LG-A1 including Satt236 and Satt258 accounted for 7.90% and 6.12% of the total variation in plant height, respectively. Three markers consisting of Satt313, Satt418 and Satt523 on LG-L also linked to this trait (Table 17). However, analysis by CIM revealed no QTL for plant height. The main reason that we were unable to locate the QTL for plant height because the difference between the parents was too low (0.77 m for Pak Chong 2 and 0.70 m for Laos 7122). The QTLs for plant height were reportedly located on linkage group B1 (Zhang *et al*., 2004), N (Reinprecht *et al.*, 2006), F (Reinprecht *et al.*, 2006)**,** C2 (Zhang *et al.*, 2004; Reinprecht *et al*., 2006), D1b (Kabelka *et al.,* 2004; Reinprecht *et al*., 2006), M (Zhang *et al*., 2004) and O (Reinprecht *et al.*, 2006).

There were three markers from simple regression analysis linked to number of nodes on main stem on LG-C2 (Sat_153), D1b (Satt537) and L (Satt313), but only one QTL from CIM were found flanking by Sat_153 and Satt322 on LG-C2 and explained 9.68% of total variation in this trait. Zhang *et al.* (2004) found QTLs of this trait on LG-A2, B1, C2, F and I.

SSR markers associated with seed length were initially identified on LG-B2, E, F, G and H via simple regression analysis (Table 17). CIM analysis located two QTLs on LG-C2 (Satt291 and Sat_153) and G (Sat_315 and Satt303), explaining 8.13 and 11.10%, respectively. Only Satt303 on LG-G was found significant from both analytical methods.

On the basis of simple regression analysis, there were many SSR markers associated with seed width located on LG-A2, B2, C2, E, G and L (Table 17). By CIM, there were three QTLs for this trait locating on LG-C2 (Satt291 and Sat_153), E (Satt045 and Satt699) and G (Satt303 and Satt504). They respectively explained 10.88, 7.56 and 10.11% of the variation in seed width. Most markers which reported from CIM were the same as reported by simple regression analysis, except Sat_153.

Nine SSR markers linking to 100-seed weight were located on LG-E and G by simple regression analysis (Table 17). Linkage groups C2 and E harbor QTLs for this trait between Satt291 and Sat_153, and Satt699 and Satt573 in the order by CIM. The QTLs accounted for 10.47 and 10.69% of the variation in seed weight, respectively (Figure31 and Table 19). Only QTL on LG-E via CIM was located on the same region as reported via simple regression analysis. QTLs for seed mass were earlier reported on LG-A2 (Zhang *et al.*, 2004), B1 (Zhang *et al.*, 2004), B2 (Watanabe *et al.,* 2004), C2 (Hyten *et al.*, 2004; Watanabe *et al.,* 2004; Reinprecht *et al*., 2006), D1a (Hyten *et al.*, 2004; Panthee *et al.*, 2005), D2 (Zhang *et al.*, 2004; Panthee *et al.*, 2005; Liu *et al.*, 2007), F (Hyten *et al.,* 2004), G (Hyten *et al.*, 2004), H (Watanabe *et al.*, 2004; Liu *et al.*, 2007), I (Hyten *et al.*, 2004; Reinprecht *et al*., 2006), J (Reinpreht *et al.*, 2006), K (Hyten *et al.,* 2004; Watanabe *et al.*, 2004), L (Hyten *et al.,* 2004), M (Liu *et al.,* 2007) and O (Liu *et al.,* 2007).

Similar results from simple regression analysis and CIM were detected in two major QTLs linked to total oil content on LG-A1. One QTL locates between Satt200 and Sat_217 which explains 27.16% of the total variation, while the other locates between Sat_217 and Satt258 and accounts for 45.18%. Although the difference in oil content between parents is not large, it is possible that they are different in loci controlling this trait and thus causes transgressive segregation among the progenies. Diers *et al*. (1992) reported markers linked to oil content on LG-A. The other QTLs for oil content were found on LG-A2 (Tajuddin *et al.*, 2003), B2 (Tajuddin *et al.*, 2003), C1 (Lee *et al.*, 1996), C2 (Hyten *et al.,* 2004), D1a (Hyten *et al.,* 2004), D1b (Kabelka *et al.*, 2004; Panthee *et al.,* 2005), D2 (Hyten *et al.,* 2004), E (Lee *et al.*, 1996; Reinprecht *et al.*, 2006; Shibata *et al.*, 2008),

H (Lee *et al.*, 1996; Panthee *et al*., 2005), I (Lee *et al.*, 1996; Tajuddin *et al.*, 2003; Shibata *et al.*, 2008), J (Tajuddin *et al.*, 2003), G (Lee *et al.*, 1996; Reinprecht *et al.*, 2006), L (Hyten *et al.*, 2004; Reinprecht *et al.*, 2006), M (Tajuddin *et al.*, 2003; Hyten *et al.*, 2004) and O (Tajuddin *et al.*, 2003, Panthee *et al.*, 2005).

None of markers located on LG-A1 has been reported linking to total oil content in the past so Satt200, Sat_217 and Satt258 are novel markers. Two QTLs of total oil content linking to *OL-5.1* and *OL-5.2* were overlapped on LG-A1. Their additive effects were 0.9482% and 1.0323%, respectively. Both alleles from Pak Chong 2 increased total oil content.

The results of simple regression revealed that there were seven SSR markers linked to high palmitic acid content. Satt258 on LG-A1 explained 5.17% of the total variation of the content. Six markers including Satt082, Satt226, Satt301, Satt389, Satt514 and Satt615 on LG-D2 accounted for 8.68, 8.45, 6.24, 10.35, 7.65 and 7.49%, respectively. Study from CIM showed one QTL associating with palmitic acid on LG-D2 (Satt389-Satt082) accounted for 13.84% of the PVE (Table 19). Our work agrees with that reported earlier on LG-D2 by Hyten *et al.* (2004), although they located additional QTLs on LG-K and LG-L. The other palmitic acid QTLs reports by Diers and Shoemaker (1992), and Li *et al*. (2002) found QTLs on LG-A1, B2, J and M. Reinprecht *et al*. (2006) located one QTL on LG-N. Li *et al.* (2002) found palmitic acid content linking with Satt684 on LG-A1 and Satt175 on LG-M, while Panthee *et al.* (2006) located the content with Satt537 on LG-D1b and Satt133 on LG-A2.

SSR markers associated with high stearic acid content were on LG-A1, C2 and E (Table 17), based on simple regression analysis. There were three QTLs left when analyzed by CIM. They were located between Sat_217 and Satt258 on LG-A1, Satt699 and Satt573 on LG-E and Satt259 and Satt473 on LG-O, explaining 8.16, 8.58 and 9.87% of PVE, respectively. QTLs for stearic acid composition were earlier reported on LG-J (Diers and Shoemaker 1992; Panthee *et al.,* 2006), B2 (Brummer *et al.,* 1995; Spencer *et al*., 2003; Panthee *et al.,* 2006). More QTLs were mapped on

LG-C2 and L (Hyten *et al.,* 2004), LG-F, G and M (Reinprecht *et al.,* 2006). In addition, Spencer *et al*. (2003) found three SSR markers, including Satt070, Satt474 and Satt556 on LG-B2 associated with *Fas* locus controlling stearic acid concentration in soybean.

Eight SSR markers were found linking to high oleic acid percentage on LG-A1, G and H by simple regression analysis (Table 17). Linkage groups A1 and H harbored QTL for oleic acid content between Sat_217 and Satt258, and Sat_127 and Satt442 in the order by CIM. Their QTLs accounted for 18.52 and 11.94% PVE, respectively. Panthee *et al.* (2006) reported markers linking to oleic acid content on LG-E. The other QTLs were found on LG-A1, D2 and G (Monteros *et al*., 2008), LG-D1b and L (Hyten *et al.,* 2004), LG-I, L and O (Bachlava *et al.,* 2008).

There were seven markers from simple regression analysis associating with too high linoleic acid composition on LG-A1, G and H (Table 17), which agreed well with the results from CIM analysis. The three QTLs were located on LG-A1 (Sat_217 and Satt258), G (Satt472 and Satt191) and H (Sat_127 and Satt442) which explained 23.85, 10.42 and 10.65% of the PVE, respectively. Panthee *et al.* (2006) found a QTL on LG-E, while Hyten *et al.* (2004) reported more QTLs on LG-F and L.

Based on simple regression analysis, there were 11 SSR markers associating with high linolenic acid composition located on LG-E, G and H (Table 17), but only one QTL was detected by CIM on LG-E (Satt598 and Satt369) which accounted for 13.25% of total PVE. QTLs for linolenic acid was earlier reported on LG-B2 (Byrum *et al*., 1995), LG-C2, E, H and O (Shibata *et al.*, 2008), LG-E and K (Diers and Shoemaker 1992), LG-E and G (Panthee *et al.,* 2006), LG-F and L (Hyten *et al*., 2004), LG-E and K (Reinprecht *et al.,* 2006). In addition, Spencer *et al*. (2003) identified two SSR markers, Satt534 and Satt560, locating near *Fan* locus on LG-B2 that associated with linolenic acid content in soybean. Our finding on LG-E is in agreement with many earlier reports.

Also, the data was analyzed by 2,500 permutation tests for LOD threshold at the significant level 0.01. The results were considered under the higher LOD threshold for example, palmitic (LOD=4.3), stearic (LOD=4.4), oleic (LOD=5.7), linoleic (LOD=5.1) acid. The data showed 5 QTLs consisting *C16:0-17.1*, *C18:0-15.1, C18:1-5.1, C18:2-5.1* and *C18:2-18.1*. Ten-thousand permutation tests for LOD threshold at significant level 0.01 or 0.001 would not use for minor QTLs.

**Table 19** Composite interval mapping of QTLs controlling node on main stem per plant, seed length, seed width, 100-seed weight, total oil content, palmitic, stearic, oleic, linoleic and linolenic acid percentage located on 7 linkage groups were developed from $F_{2:3}$ population crossing between Pak Chong 2 and Laos 7122.

| Traits | QTL | LG | Position (cM) | Flanking markers | LOD | Additive effect | Dominant effect | PVE (%) |
|---|---|---|---|---|---|---|---|---|
| Number of node | *SN-6.1* | C2 | 53.8 | Sat_153-Satt322 | 3.7 | -1.0013 | 0.3697 | 9.68 |
| Seed length | *SeedL-6.1* | C2 | 48.8 | Satt291-Sat_153 | 3.5 | 0.0739 | -0.0811 | 8.13 |
| | *SeedL-18.1* [1/] | G | 51.4 | Sat_315-Satt303 | 3.5 | -0.0838 | 0.0962 | 11.10 |
| Seed width | *SeedW-6.1* [1/] | C2 | 34.0 | Satt291-Sat_153 | 3.6 | 0.0769 | -0.0372 | 10.88 |
| | *SeedW-15.1* | E | 56.8 | Satt045-Satt699 | 3.6 | -0.0729 | -0.0001 | 7.56 |
| | *SeedW-18.1* [1/] | G | 52.4 | Satt303-Satt504 | 3.6 | -0.0811 | 0.0187 | 10.11 |
| 100-seed weight | *SW-6.1* [1/] | C2 | 43.0 | Satt291-Sat_153 | 3.5 | 0.5389 | -0.4446 | 10.47 |
| | *SW-15.1* [1/] | E | 60.9 | Satt699-Satt573 | 3.5 | -0.6590 | -0.0088 | 10.69 |
| Total oil content | *OL-5.1* [1/] | A1 | 45.7 | Satt200-Sat_217 | 3.5 | 0.9482 | -0.5908 | 27.16 |
| | *OL-5.2* [1/] | A1 | 54.2 | Sat_217-Satt258 | 3.5 | 1.0323 | -0.4690 | 45.18 |

**Table 19** (continued.)

| Traits | QTL | LG | Position (cM) | Flanking markers | LOD | Additive effect | Dominant effect | PVE (%) |
|---|---|---|---|---|---|---|---|---|
| Palmitic acid | *C16:0-17.1* [1/] | D2 | 8.0 | Satt389-Satt082 | 3.5 | -0.5860 | -0.1059 | 13.84 |
| Stearic acid | *C18:0-5.1* | A1 | 52.0 | Sat_217-Satt258 | 3.5 | 0.1311 | 0.0512 | 8.16 |
| | *C18:0-15.1* [1/] | E | 66.9 | Satt699-Satt573 | 3.5 | 0.1443 | -0.0344 | 8.58 |
| | *C18:0-10.1* | O | 28.0 | Satt259-Satt473 | 3.5 | -0.1504 | -0.0758 | 9.87 |
| Oleic acid | *C18:1-5.1* [1/] | A1 | 51.2 | Sat_217-Satt258 | 4.6 | -4.2333 | 0.1420 | 18.52 |
| | *C18:1-2.1* | H | 43.9 | Sat_127-Satt442 | 4.6 | 3.3709 | -1.1086 | 11.94 |
| Linoleic acid | *C18:2-5.1* [1/] | A1 | 52.2 | Sat_217-Satt258 | 4.0 | 4.0191 | -0.1502 | 23.85 |
| | *C18:2-18.1* [1/] | G | 136.7 | Satt472-Satt191 | 4.0 | 2.3414 | 1.4686 | 10.42 |
| | *C18:2-12.1* | H | 42.9 | Sat_127-Satt442 | 4.0 | -2.6817 | 0.9009 | 10.65 |
| Linolenic acid | *C18:3-15.1* | E | 94.6 | Sattt589-Satt369 | 3.7 | 0.2702 | 0.2788 | 13.25 |

[1/] QTLs were detected at 2,500 permutation tests, significant level at 0.01.

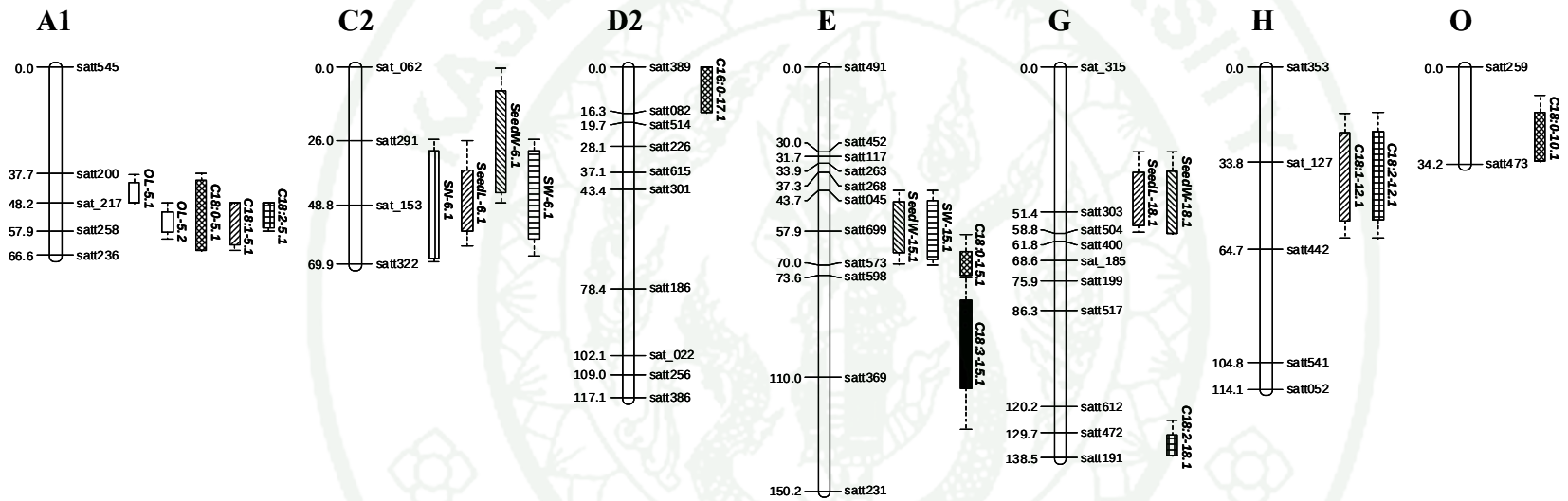



**Figure 33** QTLs located on linkage map which constructed from 138 SSR markers in $F_{2:3}$ population crossing from Pak Chong 2 × Laos 7122. Number of nodes on main stem per plant (SN), seed length (SeedL), seed width (SeedW), 100-seed weight (SW), total oil percentage (OL), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic (C18:3) acid percentage

## Discussion

### Part I: Screening and selection of germplasm

#### 1. Determination of seed moisture content and seed preparation

Sampling soybean sample following AOCS method Ac 1-45 was used at least 250 g of seed per sample. This sampling was suitable for a ton of material which process for food or feed stock. In fact, it was difficult to do in the laboratory scale because in soybean breeding program we need to save the seed for plant in the next generation especially $F_1$ or hybrid. For example, Dahmer *et al*. (1989) cut a fragment of cotyledon tissue on the opposite side of the embryonic axis of a single seed to analyze for fatty acid composition. If we choose the machine that suitable for the sample it may be useful and easy to do the laboratory, for example, electric grinder, tabletop centrifuge, analytical balance (4 digits). Moreover, we did not use all sample in the process so the sample was decrease the amount of sample in the sampling process. In addition, the ground soybean seed should be analyzed as soon as possible to minimize oxidation.

#### 2. Comparison of total oil extraction methods

Soybean seed accumulated oil seed storage for seed germination in triacylglyerol form in cotyledon. The solvent in the experiment was chosen from lipid component in the sample.

Most of methods in comparison of total oil extraction method experiment were organic solvent extraction. Some technique used specific apparatus such as Soxhlet extraction. Moreover, this technique required a lot of sample, time consuming and was not suitable for many sample. We choose the solvent from the organic solvent which used in the edible oil industry and in other reported for example, petroleum ether (bp. 40-60ºC), *n*-hexane, methanol: chloroform system. Hexane is suitable for

extracting neutral lipid (triglyceride). Yet, it is cheap, less toxic than chloroform and easier to handle its waste. Also, hexane is solvent that used in vegetable oil industry.

From the result, comparison between Soxhlet extraction and hexane extraction method showed that Soxhlet extraction take time 5 h per sample and if the experiment was design for 2 determination per sample so we could finish only 3 sample per day (if the instrument can analyze 6 sample in the same time). For hexane extraction, we can do many set of sample per day and it used the basic instrument in the laboratory (centrifuge, two arm balance, vortex mixture or orbital shaker and 4 decimal point balance).

**3. Optimization of hexane extraction condition**

This experiment we designed the condition for suitable for 16 × 100 mm glass test tube and fitted for table top centrifuge. Because of these were basic machine in the laboratory and easily to use.

3.1 Comparison of sample weight.

We wanted to compare between 0.5 and 1.0 g of sample weight for 16 × 100 mm of glass test tube. Also, this experiment was depending on the total volume of the solvent. In general, 20 times of sample was used to extract the compound from the material.

3.2 Comparison of sample types.

The soybean seed are hard seed and difficult to grinding by mortar and pestle. Also, it takes time. We looking for the electric grinding machine to grind it but almost electric grinding have a large chamber for a lot of sample. At first, we used coffee grinding but it was not fine. Finally, we used electric grinding for pearl which used in traditional medicine process but it takes time. We changed the method to use the hammer to mash the seed to small pieces and extracted the oil from the seed

sample. The small pieces of soybean seed were easily to extract the oil but it was not complete process. Thus this technique cannot use to analyze total oil content.

3.3 Comparison between shaking vs no shaking of samples.

Shaking condition was compared because these conditions have the advantage and disadvantage. Shaking condition was mixed the sample well but it takes time and difficult to precipitate. No shaking condition was take a long time to extract but easy to separate the crude extraction from the residual.

3.4 Comparison between the final extraction times.

The aim of this experiment was short period and can precipitate ground soybean sample for complete extraction.

**4. Correlation of fatty acid composition as esterification by 2 methods**

We have not attempted to compare between the methods of fatty acid methylation because the samples were not analyzed by the same machine, so only the correlation between fatty acid compositions was examined. The discrepancy between the results of both methods can also be due to biochemical changes of each fatty acid during preparation of samples. The AOCS method was conducted earlier in Thailand while the USDA-ARS method was conducted later by bringing in the same ground samples to the US and this may affect the composition of fatty acids. The rate of change may be fatty acid dependent and thus creates another interesting research topic on how the amount of these fatty acids changed while ground seed is kept in storage.
Our findings indicated that the results of the genetic and breeding study on fatty acid compositions can be different if the plant breeders employ different analytical methods. Thus a genetic linkage map constructed from different phenotyping of fatty acid contents may be influenced by the analytical methods used.

We recommended USDA-ARS method for use in a large number of samples as in soybean breeding projects because it is easier to do and the chemicals used were less toxic than the AOCS method.

## Part II: Development of $F_{2:3}$ population

### 1. Screening total oil content in 65 soybean accessions for choosing parental lines

For breeding program, we need the technique that can analyze many samples in the same time, quick and easy method to follow for save the time, cost, worker and design to plant in the next generation as soon as possible.

### 2. Crossing and develop population

The crossing from Pak Chong 2 and Laos 7122 was difficult to do and get the seed. Also, sometime it may self-pollination. If we did not know which seed was from the cross or self-pollination, we could not design to plant in the next generation. $F_1$ DNA from $F_1$ plant was test with some polymorphic SSR marker and run check on agarose gel electrophoresis to confirm $F_1$ hybrid.

## Part III: QTL analysis

### 1. Phenotypic analysis

Transgressive segregation was observed in all traits, revealing that the parents carry on alleles which contribute the effect in different direction when recombined in the hybrids. It is an important mechanism contributing to adaptive evolution (Rieseberg *et al*., 1999; Rieseberg *et al.*, 2003; Bell and Travis, 2005). Transgressive segregations beyond the lower and the higher parents were detected in all fatty acids (Table 15 and Figure 15-25), revealing the possibility in improving their contents through selection. Both parents seemed to contribute their alleles to the variation in fatty acid composition in the $F_{2:3}$ population**.**

**2. Genetypic analysis**

2.1 DNA extraction

There are other new models of spectrophotometer to determine DNA concentration such as evaluation from DNA drop which no need to dilute depending on the option of machine or using fluorescent meter to analyze.

2.2 Screening in parental lines and population by SSR primer

Only polymorphic markers between female and male parental line (Pak Chong 2 and Laos 7122, respectively) could use to screen in the population. In this experiment, we use only co-dominant SSR primer and did not use other types of marker to construct the map, so, the markers were dispersed but did not cover throughout the genome.

2.3 Gel electrophoresis and scoring phenotypic data

Polyacrylamide gel electrophoresis technique required the skill every step such as pre-cast gel, loading PCR product, developing DNA band, scoring genotypic data, especially, amplification of PCR product that affect with repeatability.

**3. Data analysis**

Regression analysis was used to identify candidate QTL which was confirm by CIM. Most markers constructed in LGs and the map position in this research corresponded well with the soybean composite map (Soybase, 2005), with some variation in linkage distance between the markers. Some markers were mapped on the same LG as the reference map with some difference in the order of the markers, especially among the closely linked markers. Gene duplication was one reason causing marker allocation in different positions (Shoemaker *et al.,* 1996; Soybase, 2005). Some markers with segregation distortion may also affect the map position.

This research reported, four QTLs consisting of *OL-5.2*, *C18:0-5.1, C18:1-5.1* and *C18:2-5.1* were co-located on LG-A1 (Figure 30). The additive effect of *C18:1-5.1* was -4.2333% while additive effect of *OL-5.2*, *C18:0-5.1* and *C18:2-5.1* was 1.0323%, 0.1311% and 4.0191%, respectively. The alleles increasing oleic acid content were from Laos 7122, while alleles increasing total oil, stearic and linoleic acid contents were from Pak Chong 2. That means if high total oil content was selected, it is expected that high oleic and linoleic acid contents would be indirectly selected at the same time. Thus it will be difficult to improve the whole profile of fatty acid in just one breeding population. Also we found *OL-5.1* overlapped with *C18:1-5.1* on LG-A1 with the additive effect of 0.9482% and 0.1311%, respectively. The alleles increasing total oil and oleic acid content were from Pak Chong 2. In addition, Diers and. Shoemaker (1992) reported three QTLs of linoleic acid linked to A170_1, A104_1 and A082_1 on LG-A1 developing from $F_2$ population. These QTL also linked to QTL of oleic acid at the same position of the composite map (soybase.org). The position of RFLP marker A170_1 (92.55 cM) was near SSR marker Satt200 (92.88 cM). Also, A082_1 (102.30 cM) was near Sat_217 (101.57 cM). The *C18:1-5.1* and *C18:2-5.1* in this research corresponded well with the reference. These QTL could use to selecting soybean line in soybean breeding program.

Three QTLs consisting of *SN-6.1, SeedL-6.1, SW-6.1* were co-located on LG-C2 with the additive effect of -1.0013 nodes, 0.0739 mm and 0.5389 g, respectively. The alleles increasing number of nodes on main stem were from Laos 7122, while those increasing seed length and seed weigh were from Pak Chong 2. *SeedW-6.1* was overlapped to these QTLs with the additive effect of 0.0769 mm (Table 17). *SeedW-15.1* and *SW-15.1* co-located on LG-E with the additive effect of -0.0729 mm and -0.6590 g respectively. The alleles that increase seed width and seed weight were both from Laos 7122. *SeedL-18.1* and *SeedW-18.1* co-located on LG-G with the additive effect of -0.0838 mm and -0.0811 mm, respectively. The alleles increased seed length and seed width were from Laos 7122. Alleles from Pak Chong 2 increased seed length, seed width and seed weight but decreased number of nodes on main stem. These relationships showed that it was

rather difficult to improve all these traits at the same time. Our research found one QTL for number of nodes on main stem, two QTL for seed length, three QTLs for seed width and two QTLs for 100-seed weight. Although many markers were found, their contribution was small and no major QTL was observed.

*C18:1-12.1* and *C18:2-12.1* were co-located on LG-H with the additive effects of 3.3709% and -2.6817%, respectively. The allele that increased oleic acid content was from Pak Chong 2, whereas the allele that increased linoleic acid content was from Laos 7122. Again, it will be difficult to improve oleic acid content from this population.

There were five independent QTLs, viz. *C16:0-17.1* on LG-D2, *C18:0-15.1* and *C18:3-15.1* on LG-E, *C18:2-18.1* on LG-G, and *C18:0-10.1* on LG-O which may be used as markers for selecting for high levels of the specified fatty acids. They also found that *C18:0-15.1* overlapped with *SeedW-15.1* and *SW-15.1* on LG-E with the additive effect of 0.1443%, -0.0729% and -0.6590%, respectively. The alleles increasing stearic acid content was from Pak Chong 2, but the alleles increasing seed width and 100-seed weight were from Laos 7122. It will be difficult to select these interesting traits at the same time.

# CONCLUSION AND RECOMMENDATION

## Conclusion

1. We found that the best methods for total oil extraction were hexane extraction, 1:2 (v/v) of chloroform: methanol, and NMR technique. While Soxhlet extraction and 1: 2: 0.75 (v/v/v) of chloroform: methanol: 0.15 M acetic acid solvent showed poorer extracting ability.

2. The hexane method was simple, fast and yet produces less analytical waste. A suitable condition for hexane extraction in 16 × 100 mm test tube was to use 0.5 g of ground sample extracted by 9 ml hexane for 3 cycles while adding 3 ml hexane in each cycle, shaking for 15 min, left them 3 h and incubating ≥ 16 h in the final step.

3. There were five major fatty acids in soybean seed, viz. palmitic, stearic, oleic, linoleic and linolenic acids. Their compositions varied depending on soybean varieties. To analyze fatty acid profile, we recommended using USDA-ARS method because the chemical in this method was toxic less than another method and shelf-life longer than boron trifluoride.

4. Two major QTLs for total oil content on LG-A1 were identified and thus can be used in breeding to improve oil content in soybean.

5. This research reports 6 major QTLs controlling high fatty acid content. The QTLs for oleic acid with 18.52% PVE and for linoleic acid with 23.85% PVE were located on LG-A1. A major QTL for palmitic acid content was located on LG-D2 with 13.84% PVE. A QTL for linolenic acid content was on LG-E with 13.52% PVE. Two co-locating QTLs were found on LG-H for oleic acid content with 11.94% PVE and for linoleic acid content with 10.65% PVE. These markers can be used to select for increasing the content of these fatty acids in soybean

## Recommendation

We should apply more SSR markers in the population to produce a fine map. Also, design new markers closely linked to QTL positions for use in the plant breeding program.

The disadvantage of this research was evaluated only one season in one year and no replication form $F_{2:3}$ population. The population should be self-pollination to the $F_6$ generation and planted in multiple location in one year for confirm the QTL.

# LITERATURE CITED


Akkaya, M. S., A. A. Bhagwat and P. B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics** 132(4): 1131-1139.

Amaral, J.S., S.C. Cunha, M.R. Alves, J.A. Pereira, R.M. Seabra and B.P.P. Oliveira. 2004. Triaclyglycerol composition of walnut (*Juglans regia* L.) cultivars: characterization by HPLC-ELSD and chemometrics. **J. Agric. Food Chem.** 52: 7964-7969.

Andrikopoulos, N.K. 2002. Chromatographic and spectroscopic methods in the analysis of triacylglycerol species and regiospecific isomers of oils and fats. **Crit. Rev. Food Sci.** 42(5): 473-505.

Bachlava, E., R.E. Dewey, J. Auclair, S. Wang, J.W. Burton and A.J. Cardinal. 2008. Mapping gene encoding microsomal ω-6 desaturase enzymes and their cosegregation with QTL affecting oleate content in soybean. **Crop Sci**. 48: 640-650.

Beckel, A.C., P.A. Belter and A.K. Smith. 1948. Solvent effects on the products of soybean oil extraction. Northern Regional Research Labolatory, Peorio, Illinois.

Bell, M.A. and M.P. Travis. 2005. Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. **Trends Ecol. Evol**. 20(7): 358-361.

Blasi, D.A., J. Drouillard, E.C. Titgemeyer, S.I. Paisley and M.J. Brouk. 2000. Soybean hull, composition and feeding value for beef and dairy cattle. Kansas State University, Research and Extension.

Bligh, E.G. and W. Dyer. 1959. A rapid method for total lipid extraction and purification. **Can. J. Biochem. Physiol.** 37: 911-917.

Borek, C. 2004. Dietary antioxidants and human cancer. **Integrative cancer therapies.** 3(4): 333-341.

Brouwers, J.F.H.M., B.M. Gadella, L.M.G. Van Golde and A.G.M. Tielens. 1998. Quantitative analysis of phosphatidylcholine molecular species using HPLC and light scattering detection. **J. Lipid Res.** 39(2): 344-353.

Bruce, K.J., L.K. Karr-Lilienthal, K.E. Zinn, L.L. Pope, D.C. Mahan, N.D. Fastinger, M. Watts, P.L. Utterback, C.M. Parsons, E.O.Castaneda, M. Ellis and G.C. Fahey, Jr. 2006. Evaluation of the inclusion of soybean oil and soybean processing by-products to soybean meal on nutrient composition and digestibility in swine and poultry. **J. Anim. Sci.** 84: 1403-1414.

Brummer, E.C., A.D. Nickell, J.R. Wilcox, and R.C. Shoemaker. 1995. Mapping the *Fan* locus controlling linolenic acid content in soybean oil. **J. Hered**. 86(3): 245-247.

Burton, G.W. and M.G. Traber. 1990. Vitamin E: antioxidant activity, biokinetics, and bioavailability. **Annu. Rev. Nutr.** 10(1): 357-382.

Burton, J.W. 1987. Soybean [Glycine max (L.) Merr.]. **Field Crops Res.** 53: 171-186.

Byrum, J.R., A.J. Kinney, R.C. Shoemaker and B.W. Diers. 1995. Mapping of the microsomal and plastid omega-3 fatty acid desaturases in soybean**. Soybean Genet. Newslett.** 22: 181-184

Christie, W.W. 2011. **Detectors for high-performance lipid chromatography of lipids with special reference to evaporative light scattering detection**. Available Source: http://lipidlibrary.aocs.org/topics/detect92/index.htm, September 2, 2011.

Clemente, T.E. and E.B. Cahoon. 2009. Soybean oil: genetic approaches for modification of functionality and total content. **Plant Physiol**. 151(3): 1030-1040.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, A. Kaya, N. Vantoai, T. Lohnes, D. Chung and J. Specht. 1999. An integrated genetic map of soybean genome. **Crop Sci.** 39: 1464-1490.

Dahmer, M.L., P.D. Fleming, G.B. Collins and D.F. Hildebrand. 1989. A rapid screening technique for determining the lipid composition of soybean seeds. **J. Am. Oil Chem. Soc.** 66 (4): 543-548.

Diers, B.W. and R.C. Shoemaker. 1992. Restriction fragment length polymorphism analysis of soybean fatty acid content. **J. Am. Oil Chem. Soc.** 69: 1242-1244.

_____, P. Keim, W.R. Fehr and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. **Thero. Appl. Genet**. 83: 608-612.

Doyle, L.L. and J.L. Doyle. 1990. Isolation of plant DNA fromfresh tissue. **Focus**. 12: 13-15.

Folch, J., M. Lees, G.H.S. Stanley. 1957. A simple method for the isolation and purification of total lipid from animal tissues**. J. Biochem.** 226: 497-509.

Hildebrand D.F., R. Li and T. Hatanaka. 2008. Chapter12: genomics of soybean oil traits. **Genomics and genetics of soybean**, pp. 185-209, Springer, New York.

Hou, G., G.R. Ablett, K.P. Pauls and I. Rajcan. 2006. Environmental effects on fatty acid levels in soybean seed oil. **J. Am. Oil Chem. Soc**. 83: 759-763.

Hu, F.B., M.J. Stampfer, J.E. Manson, E. Rimm, G.A. Colditz, B.A. Rosner, C.H. Hennekens, and W.C. Willett. 1997. Dietary fat intake and the risk of coronary heart disease in women. **N. Engl. J. Med**. 337: 1491-1499.

Hurburgh, C.R., Jr. 1994. Long-term soybean composition patterns and their effect on processing. **J. Am. Oil Chem. Soc**. 71(12): 1425.

Hyten, D.L., V.R. Pantalone, A.M. Saxto, M.E. Schmidt, and C.E. Sams. 2004. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci**. J. Am Oil Chem. Soc**. 81(12): 1115-1118.

Kabelka, E., B. Diers, W. Fehr, A. Leroy, I. Baianu, T. You, D. Neece and R. Nelson. 2004. Putative alleles for increased yield from soybean plant introductions. **Crop Sci** 44(3): 784-791.

Keim, P., R.C. Shoemaker and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. **Theor. Appl. Genet**. 77(6): 786-792.

Kim, C.S., C.H. Lee, J.S. Shin, Y.S. Chung and N.I. Hyung. 1997. A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP. **Nucleic Acids Res**. 25: 1085-1086.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. **Ann Euge**. 12: 172-175.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey and H.R. Boerma. 1996. RFLP loci associated with soybean protein and oil content across population and location. **Theor. Appl. Genet.** 93: 649-657.

Li, Z., R.F. Wilson, W.E. Rayford and H.R. Boerma. 2002. Molecular mapping genes conditioning reduced palmitic acid content in N 87-2122-4 soybean. **Crop Sci.** 42: 373-378.

Lichtenstein, A.H., L.M. Ausman, W.E. Rayford and H.R. Boerma. 1999. Effects of different forms of dietary hydrogenated fats on serum lipoprotein cholesterol levels. **N. Engl. J. Med**. 340: 1933-1940.

Lin, J.J., J. Kuo, J. Ma, J. Saunders, H. Beard, M. Macdonald, W. Kenworthy, G. Ude and B. Matthews. 1996. Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. **Plant Mol. Biol. Rep.** 14(2): 156-169.

Liu, K. 1997. Chapter 2: Chemistry and nutritional value of soybean components. **Soybean: chemistry, technology and utilization**, Aspen Publishers, Gaithersburg, MD, pp. 25-35.

Liu, B., T. Fujita, Z. Yan, S. Sakamoto, D. Xu and J. Abe. 2007. QTL mapping of domestication-related traits in soybean (*Glycine max*). **Ann Bot.** 100: 1027-1038.

Lodhi, M.A., G.N. Ye, N.F. Weeden and B.I. Reisch. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. **Plant Mol. Biol. Rep.** 12: 6-13.

Mazur, B., E. Krebbers and S. Tingey. 1999. Gene discovery and product development for grain quality traits. **Science** 285: 372-375.

McConn, M. and J. Browse. 1996. The critical requirement for linolenic acid is pollen development, not photosynthesis, in an Arabidopsis mutant. **The Plant Cell** 8(3): 403-416.

Mensink, R.P. and M.B. Katan. 1990. Effect of dietary *trans*-fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. **New Engl. J. Med.** 323(7): 439-445.

Mian, M., M. Bailey, J. Tamulonis, E. Shipe, T. Carter, W. Parrott, D. Ashley, R. Hussey and H. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. **Theor.Appl.Genet.** 93(7): 1011-1016.

Miquel, M. and J. Browse, 1995. Lipid biosynthesis in developing seeds, pp. 169-193. *In*: J. Kigel and G. Galili, (Eds.), Lipid development and germination, Dekker, New York, Ch.7.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yona, C.R. Bhatia and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. **Mol. Breeding** 3: 87-103.

Monteros, M.J., J.W. Burton, and H.R. Boerma. 2008. Molecular mapping and confirmation of QTLs associated with oleic acid content in N 00-3350. **Crop Sci.** 48: 2223-2234.

Office of Agricultural Economics. 2010. Ministry of Agriculture and Cooperatives. Amount and imported agricultural products: soybean. Amount and value of monthly imported agricultural products. Available Source: http://www.oae.go.th/statistic/import/imSY.xls, August 28, 2011.

Pallot, T.N., A.S. Leong, J.A. Allen, T.M. Golder, C.F. Greenwood and T. Golebiowski. 1999. Precision of fatty acid analyses using near infrared spectroscopy of the whole seed brassicas. *In*: N. Wratten, P. Salisbury, A. Green, B. Howlett, J. Kirkegaard, R. Mailer, G. McDonald, N. Mendham, R. Norton (editors). 10th International Rape Seed Congress. Camberra, Australia: The Regional Institute Ltd.

Panthee, D.R., V.R. Pantalone and A.M. Saxton. 2006. Modifier QTL for fatty acid composition in soybean oil. **Euphytica.** 152: 67-73.

_____, D.R. West, A.M. Saxton and C.E. Sams. 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. **Crop Sci**. 45: 2015-2022.

Pich, U. and I. Schubert. 1993. Midiprep method for isolation of DNA from plants with a high content of polyphenolics. **Nucleic Acids Res.** 21: 3328.

R Development Core Team. 2006. **R: a language and environment for statistical computing**. R foundation for statistical computing, Vienna, Austria, ISBN: 3-900051-07-0. Available Source: http://R-project.org., December 5, 2007.

Rajcan, I., G. Hou and A. Weir. 2005. Advances in breeding of seed-quality traits in soybean. **Journal of Crop Improvement.** 14(1-2): 1-2.

Rao, M.S.S., A.S. Bhagsari and A.I. Mohamed. 1998. Yield, protein, and oil quality of soybean genotypes selected for tofu production. **Plant Foods Hum. Nutr.** 52: 241-251.

Reinprecht, Y., V.W. Poysa, K. Yu, I. Rajcan, G.R. Ablett and K.P. Pauls. 2006. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. **Gonome** 49: 1510-1527.

Rieseberg, L.H., M.A. Archer and R.K. Wayne. 1999. Transgressive segregation, adaptation and speciation. **J. Hered**. 83(4): 363-372.

_____, A. Widmer, A.M. Arntz and B. Burke. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. **Phil.Trans. R. Soc. Lond. B.** 358 (1434): 1141-1147.

Sambrook, J. and D.W. Russell. 2001. Molecular cloning: A laboratory manual (3-volume set), 3rd Edition. Texas, USA.

Sebolt, A.M., R.C. Shoemaker and B.W. Diers. 2000. Analysis of quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. **Crop Sci.** 40: 1438-1444.

Shibata M, K. Takayama, A. Ujiie, T. Yamada, J. Abe and K. Kitamura. 2008. Genetic relationship between lipid content and linolenic acid concentration in soybean seeds. **Breeding Sci. 58**: 361-366.

Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). **Genetics** 144: 329-338.

Spencer, M.M., V.R. Pantalone, E.J. Meyer, D. Landau-Ellis and D.L. Hyten, Jr. 2003. Mapping the *Fas* locus controlling stearic acid content in soybean. **Theor. Appl. Genet**. 106: 615-619.

Somerville, C., J. Browse, J.G. Jaworski and J.B. Ohlrogge. 2000. Chapter 10: Lipid, pp. 456-527. *In:* B.B. Buchanan, W. Gruissem and R.L. Jones, **Biochemistry and Molecular Biology of Plants**, The American Society of Plant Physiologists, Courier Companies, Inc., USA.

Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht and P.B. Cregan. 2004. A new integrated genetic linkage map of soybean. **Theor Appl Genet.** 109: 122-128.

Soybase. 2005. **Soybase and the soybean breeder's toolbox**. Available Source: http://soybase.org., January 31, 2011.

Stolyhwo, A., H. Colin and G. Guiochon. 1985. Analysis of triglycerides in oils and fats by liquid chromatography with laser light scatting detector. **Anal. Chem**. 57: 1342-1354.

Sukhija, P.S. and D.L. Palmquist. 1988. Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. **J. Agri. Food Chem.** 36: 1202-1206.

Tajuddin, T., S. Watanabe, N. Yamanaka and K. Harada. 2003. Analysis of quantitative trait loci for protein and lipid contents in soybean seed using recombinant inbred lines. **Breeding Sci.** 53: 133-140.

Takagi, T. and T. Yanagita. 1985. Comparative studies of the solvent mixtures for the extraction of soybean lipid. **Bull. Fac. Agr. , Saga Univ**. 59: 23-26.

The American Oil Chemists' Society. 1993. **Official methods and recommended practices of the American Oil Chemists' Society**. Champaign, IL, USA, Methods Ce 2-66, Ce 1-62.

Van Ooijen, J.W. and R.E. Voorrips. 2001. JoinMap®3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen.

Velasco, L., C. Möllers and H. C. Becker. 1999. Screening for quality traits in single seeds of rapeseed by near-infrared reflectance spectroscopy. *In:* N. Wratten, P. Salisbury, A. Green, B. Howlett, J. Kirkegaard, R. Mailer, G. McDonald, N. Mendham, R. Norton (editors). 10th International Rape Seed Congress. Camberra, Australia: The Regional Institute Ltd.

Wang, S., C.J. Basten, and Z.B. Zeng. 2011. Windows QTL Cartographer 2.5_009. Department of Statistics, North Carolina State University, Raleigh, N.C. Available Source: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm, January 1, 2011.

Watanabe, S., T. Tajuddin, N. Yamanaka and K. Harada. 2004. Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines. **Breeding Sci.** 54: 399-407.

Wilson, R.F. 1996. Current research on oilseeds Institute of Food Technologists, Symposium: **Identify preserved Oils, Institute of Food Technologists**, New Orleans, LA.

_____. 2004. Seed composition, pp. 621-677. *In:* H.R. Boerma and J.E. Specht, (Eds.), **Soybeans: Improvement, Production and Uses**, ASA, CSSA, SSA, Madison, WI.

Zhang, W.K., Y.J. Wang, G.Z. Luo, J.S. Zhang, C.Y. He, X.L. Wu, J.Z. Gai and S.Y. Chen. 2004. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association wit EST markers. **Theor. Appl. Genet**. 108: 1131-1139.

# APPENDIX

**Appendix Table 1** List of 65 soybean accessions in these experiments

| ID No. | Varieties | Source |
|---|---|---|
| 1 | AGS 9 | AVRDC |
| 3 | AGS 129 | AVRDC |
| 4 | AGS 205 | AVRDC |
| 5 | AGS 207 | AVRDC |
| 6 | AGS 292 | AVRDC |
| 8 | AGS 327 | AVRDC |
| 9 | $BC_4$ 2-7-7 | AVRDC |
| 10 | $BC_5$ 76-3-$F_5$ | AVRDC |
| 11 | Black soybean | AVRDC |
| 12 | CKP 1 | AVRDC |
| 13 | Clack 63 x SJ 2 (#23) | AVRDC |
| 15 | Clack 63 x SJ 5 (#58) | AVRDC |
| 16 | CM 1 | AVRDC |
| 17 | CM 60 | AVRDC |
| 18 | CM 205 | AVRDC |
| 19 | CN 001 | AVRDC |
| 20 | CN 002 | AVRDC |
| 21 | CN 003 | AVRDC |
| 22 | Col/Ehime/1981/Utsunomiya | Japan |
| 23 | CSV #101 | AVRDC |
| 24 | CSV #102 | AVRDC |
| 25 | CSV #103 | AVRDC |
| 26 | CSV #104 | AVRDC |
| 27 | Disoy | AVRDC |
| 28 | DHK 006 | AVRDC |
| 29 | G 9956 | AVRDC |
| 30 | GC 81079-12 | AVRDC |
| 31 | GC 84058-21-2 | AVRDC |
| 32 | GC 94014-1-3-1-1 | AVRDC |
| 33 | GC 95004-2-3-1-1 | AVRDC |

**Appendix Table 1** (Continued)

| ID No. | Varieties | Source |
|---|---|---|
| 34 | GC 96019-8 | AVRDC |
| 36 | Iwate Ninohe 51 | Japan |
| 38 | KEGON | AVRDC |
| 39 | KKU 35 | AVRDC |
| 40 | KUSL 20004 | AVRDC |
| 41 | MTD 176 | AVRDC |
| 42 | MTD 299 | AVRDC |
| 43 | NS 1 | AVRDC |
| 44 | PI 235347 | 99S-1193, USA |
| 45 | PI 371610 | 97S-2011, USA |
| 46 | PI 371611 | 97S-2012, USA |
| 47 | PI 108169 A | 93U-5288, USA |
| 48 | PI 408169 B | 96U-2050, USA |
| 49 | PI 408169 C | 91S-774, USA |
| 50 | PI 408169 D | 95U-2617, USA |
| 51 | PI 424326 | 96U-2123, USA |
| 52 | PI 424485 | 92U-1380, USA |
| 53 | PU DUA 8008 B | AVRDC |
| 54 | RM 1 | AVRDC |
| 55 | Shimo Shirazu | Japan |
| 56 | SJ 1 | AVRDC |
| 57 | SJ 2 | AVRDC |
| 58 | SJ 4 | AVRDC |
| 59 | SJ 5 | AVRDC |
| 60 | SK 8502-4-1 | AVRDC |
| 61 | SOJA DAK | AVRDC |
| 62 | SRE-B-15 C | AVRDC |
| 63 | ST 1 | AVRDC |
| 64 | ST 2 | AVRDC |
| 65 | Pak Chong 1 | AVRDC |
| 66 | Pak Chong 2 | AVRDC |

**Appendix Table 1** (Continued)

| ID No. | Varieties | Source |
|---|---|---|
| 67 | AGS 269 | AVRDC |
| 68 | Laos 429 | AVRDC |
| 69 | Laos 7122 | AVRDC |
| 70 | Laos 8750 A2 | AVRDC |

**Appendix Table 2** Total oil content and moisture content in 65 soybean varieties.

| ID | Varieties | Total oil content (%) | | Moisture content (%) | |
|---|---|---|---|---|---|
| | | Hexane extraction | NMR | AOCS | NMR |
| 1 | AGS 9 | 23.32 | 22.55 | 7.64 | 4.86 |
| 3 | AGS 129 | 22.90 | 21.61 | 7.92 | 4.84 |
| 4 | AGS 205 | 21.82 | 20.36 | 7.73 | 4.77 |
| 5 | AGS 207 | 20.85 | 19.77 | 7.41 | 4.47 |
| 6 | AGS 292 | 22.12 | nd | 7.55 | nd |
| 8 | AGS 327 | 18.38 | 18.11 | 8.27 | 4.31 |
| 9 | BC4 2-7-7 | 18.12 | 17.41 | 8.30 | 4.44 |
| 10 | BC5 76-3-$F_5$ | 18.06 | 17.88 | 8.17 | 3.50 |
| 11 | Black soybean | 22.22 | nd | 6.81 | nd |
| 12 | CKP 1 | 20.44 | 18.86 | 7.33 | 4.73 |
| 13 | Clack 63 x SJ 2 (#23) | 20.27 | 18.97 | 7.53 | 4.49 |
| 15 | Clack 63 x SJ 5 (#58) | 20.90 | 19.93 | 7.49 | 4.79 |
| 16 | CM 1 | 21.53 | 20.09 | 8.04 | 4.57 |
| 17 | CM 60 | 21.60 | 20.57 | 7.34 | 4.53 |
| 18 | CM 205 | 21.40 | 20.42 | 8.40 | 4.58 |
| 19 | CN 001 | 21.24 | 20.98 | 7.97 | 4.25 |
| 20 | CN 002 | 20.43 | 19.87 | 9.01 | 4.08 |
| 21 | CN 003 | 23.97 | 22.62 | 7.65 | 4.28 |
| 22 | Col/Ehime/1981/Utsunomiya | 21.88 | 21.19 | 7.84 | 3.21 |
| 23 | CSV #101 | 18.39 | 17.49 | 8.90 | 3.40 |
| 24 | CSV #102 | 16.47 | 16.11 | 10.30 | 2.83 |
| 25 | CSV #103 | 17.24 | 16.43 | 9.26 | 3.26 |
| 26 | CSV #104 | 18.51 | 17.66 | 8.15 | 3.15 |
| 27 | Disoy | 17.07 | 16.57 | 8.38 | 3.03 |
| 28 | DHK 006 | 18.69 | 18.28 | 8.70 | 3.32 |
| 29 | G 9956 | 20.10 | 18.69 | 7.55 | 3.39 |
| 30 | GC 81079-12 | 23.11 | 22.29 | 7.41 | 3.49 |
| 31 | GC 84058-21-2 | 22.17 | 22.58 | 9.02 | 3.29 |
| 32 | GC 94014-1-3-1-1 | 19.34 | 19.53 | 8.51 | 3.19 |

**Appendix Table 2** (continued)

| ID | Varieties | Total oil content (%) | | Moisture content (%) | |
|---|---|---|---|---|---|
| | | Hexane extraction | NMR | AOCS | NMR |
| 33 | GC 95004-2-3-1-1 | 20.65 | 20.65 | 8.12 | 3.29 |
| 34 | GC 96019-8 | 21.06 | 20.77 | 8.15 | 3.44 |
| 36 | Iwate Ninohe 51 | 20.52 | nd | 8.04 | nd |
| 38 | KEGON | 17.71 | nd | 7.88 | nd |
| 39 | KKU 35 | 19.25 | 19.44 | 7.64 | 3.14 |
| 40 | KUSL 20004 | 21.47 | 20.95 | 7.14 | 3.58 |
| 41 | MTD 176 | 19.50 | 19.15 | 7.63 | 3.16 |
| 42 | MTD 299 | 19.16 | 18.52 | 7.41 | 3.37 |
| 43 | NS 1 | 18.48 | 18.02 | 7.49 | 3.27 |
| 44 | PI 235347 | 21.85 | nd | 6.97 | nd |
| 45 | PI 371610 | 23.35 | 22.44 | 6.66 | 3.25 |
| 46 | PI 371611 | 23.58 | nd | 6.60 | nd |
| 47 | PI 408169 A | 19.17 | 18.87 | 7.23 | 3.31 |
| 48 | PI 408169 B | 18.37 | 17.88 | 7.62 | 3.48 |
| 49 | PI 408169 C | 19.46 | 18.78 | 6.76 | 3.42 |
| 50 | PI 408169 D | 20.10 | 19.56 | 6.90 | 3.17 |
| 51 | PI 424326 | 20.97 | 19.70 | 6.99 | 3.21 |
| 52 | PI 424485 | 20.78 | nd | 6.99 | nd |
| 53 | PU DUA 8008 B | 18.15 | 17.90 | 8.08 | 3.07 |
| 54 | RM 1 | 20.86 | 20.86 | 6.32 | 3.32 |
| 55 | Shimo Shirazu | 19.59 | nd | 7.23 | nd |
| 56 | SJ 1 | 21.58 | 20.81 | 6.94 | 3.40 |
| 57 | SJ 2 | 20.47 | 19.75 | 7.02 | 3.35 |
| 58 | SJ 4 | 20.97 | 20.31 | 6.94 | 3.39 |
| 59 | SJ 5 | 20.25 | 19.84 | 6.90 | 3.26 |
| 60 | SK 8502-4-1 | 22.24 | 21.30 | 6.02 | 3.26 |
| 61 | SOJA DAKO | 20.57 | 20.07 | 6.86 | 3.30 |
| 62 | SRE-B-15 C | 22.34 | 21.51 | 7.16 | 3.05 |

**Appendix Table 2** (continued)

| ID | Varieties | Total oil content (%) | | Moisture content (%) | |
|---|---|---|---|---|---|
| | | Hexane extraction | NMR | AOCS | NMR |
| 63 | ST 1 | 23.33 | 21.79 | 6.76 | 3.34 |
| 64 | ST 2 | 22.34 | 21.54 | 6.62 | 3.35 |
| 65 | Pak Chong 1 | 20.96 | 19.86 | 6.62 | 3.49 |
| 66 | Pak Chong 2 | 18.70 | nd | 6.49 | nd |
| 67 | AGS 269 | 22.98 | 22.12 | 6.24 | 3.27 |
| 68 | Laos 429 | 22.13 | nd | 6.76 | nd |
| 69 | Laos 7122 | 21.59 | 20.77 | 6.92 | 3.36 |
| 70 | Laos 8750 A2 | 20.13 | 19.46 | 7.37 | 3.31 |

Total oil content and moisture content were analyzed from ground soybean samples.
nd = no data

**Appendix Table 3** Fatty acid composition in 65 soybean varieties analyzing by GC-FID.

| ID | Varieties | Fatty acid composition (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C16:0 | | C18:0 | | C18:1 | | C18:2 | | C18:3 | |
| | | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA |
| 1 | AGS 9 | 13.16 | 12.28 | 3.15 | 3.62 | 30.41 | 33.77 | 48.80 | 46.00 | 4.49 | 4.33 |
| 3 | AGS 129 | 14.93 | 12.80 | 2.91 | 4.12 | 24.37 | 28.12 | 51.31 | 49.56 | 6.48 | 5.40 |
| 4 | AGS 205 | 12.51 | 11.37 | 3.45 | 4.11 | 36.29 | 39.64 | 41.53 | 38.35 | 6.22 | 6.53 |
| 5 | AGS 207 | 13.50 | 12.09 | 2.62 | 3.19 | 37.18 | 40.93 | 40.76 | 38.33 | 5.93 | 5.46 |
| 6 | AGS 292 | 14.04 | 10.63 | 2.41 | 2.83 | 37.53 | 38.97 | 40.91 | 41.58 | 5.11 | 6.00 |
| 8 | AGS 327 | 13.31 | 12.34 | 1.43 | 2.45 | 43.28 | 39.84 | 34.96 | 38.36 | 6.62 | 7.01 |
| 9 | BC4 2-7-7 | 14.81 | 11.40 | 2.06 | 2.47 | 35.02 | 47.85 | 41.46 | 31.90 | 6.65 | 6.37 |
| 10 | BC5 76-3-$F_5$ | 13.10 | 11.88 | 1.91 | 2.45 | 40.43 | 44.22 | 37.02 | 34.98 | 7.53 | 6.48 |
| 11 | Black soybean | 11.78 | 11.46 | 4.05 | 2.84 | 25.88 | 28.42 | 51.82 | 50.93 | 6.47 | 6.36 |
| 12 | CKP 1 | 13.03 | 12.14 | 2.77 | 3.64 | 18.61 | 20.96 | 57.27 | 56.82 | 8.32 | 6.44 |
| 13 | Clack 63 x SJ 2 (#23) | 13.87 | 13.04 | 3.17 | 4.20 | 19.45 | 19.94 | 55.14 | 55.67 | 8.37 | 7.15 |
| 15 | Clack 63 x SJ 5 (#58) | 14.07 | 12.43 | 3.00 | 4.01 | 21.13 | 22.45 | 55.68 | 54.55 | 6.12 | 6.55 |
| 16 | CM 1 | 12.72 | 11.99 | 2.46 | 3.32 | 22.37 | 24.30 | 55.01 | 54.78 | 7.45 | 5.61 |
| 17 | CM 60 | 11.50 | 10.58 | 2.74 | 3.59 | 27.72 | 30.47 | 50.72 | 49.19 | 7.32 | 6.17 |
| 18 | CM 205 | 15.59 | 11.08 | 2.00 | 2.79 | 35.47 | 39.53 | 42.25 | 41.13 | 4.68 | 5.48 |

**Appendix Table 3** (continued)

| ID | Varieties | Fatty acid composition (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C16:0 | | C18:0 | | C18:1 | | C18:2 | | C18:3 | |
| | | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA |
| 19 | CN 001 | 10.98 | 9.43 | 2.46 | 3.04 | 33.76 | 37.08 | 46.56 | 44.24 | 6.24 | 6.21 |
| 20 | CN 002 | 14.04 | 12.14 | 2.48 | 3.22 | 34.77 | 39.58 | 43.49 | 40.31 | 5.22 | 4.74 |
| 21 | CN 003 | 16.74 | 13.16 | 3.46 | 4.32 | 23.82 | 26.63 | 50.96 | 50.20 | 5.02 | 5.69 |
| 22 | Col/Ehime/1981/Utsunomiya | 13.43 | 11.10 | 2.31 | 3.01 | 36.03 | 40.22 | 44.80 | 41.14 | 3.43 | 4.52 |
| 23 | CSV #101 | 13.48 | 11.76 | 2.56 | 3.68 | 26.33 | 30.16 | 48.46 | 46.85 | 9.16 | 7.55 |
| 24 | CSV #102 | 15.57 | 13.04 | 1.91 | 2.78 | 28.07 | 34.15 | 45.84 | 43.75 | 8.61 | 6.28 |
| 25 | CSV #103 | 13.76 | 11.88 | 2.42 | 3.48 | 19.93 | 22.90 | 53.26 | 52.77 | 10.93 | 8.97 |
| 26 | CSV #104 | 15.03 | 12.37 | 1.91 | 2.63 | 32.86 | 38.69 | 44.75 | 41.59 | 5.45 | 4.72 |
| 27 | Disoy | 14.59 | 12.17 | 2.03 | 3.01 | 33.39 | 39.81 | 42.02 | 38.89 | 7.97 | 6.12 |
| 28 | DHK 006 | 14.95 | 12.53 | 1.67 | 2.39 | 30.58 | 35.56 | 44.48 | 41.83 | 8.32 | 7.69 |
| 29 | G 9956 | 14.55 | 11.89 | 2.58 | 3.41 | 23.50 | 25.97 | 53.22 | 52.58 | 6.15 | 6.14 |
| 30 | GC 81079-12 | 14.77 | 12.24 | 2.64 | 3.74 | 24.06 | 25.01 | 58.12 | 53.70 | 0.41 | 5.31 |
| 31 | GC 84058-21-2 | 14.21 | 12.03 | 3.99 | 3.26 | 33.76 | 32.96 | 40.52 | 46.93 | 7.52 | 4.82 |
| 32 | GC 94014-1-3-1-1 | 13.48 | 11.47 | 1.68 | 2.79 | 29.26 | 35.34 | 48.63 | 44.91 | 6.94 | 5.48 |
| 34 | GC 96019-8 | 13.57 | 11.49 | 1.87 | 2.56 | 39.39 | 49.18 | 37.57 | 31.98 | 7.59 | 4.78 |

**Appendix Table 3** (continued)

| ID | Varieties | Fatty acid composition (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C16:0 | | C18:0 | | C18:1 | | C18:2 | | C18:3 | |
| | | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA |
| 36 | Iwate Ninohe 51 | 13.47 | 10.63 | 2.05 | 2.62 | 24.05 | 50.23 | 52.48 | 31.60 | 7.95 | 4.92 |
| 38 | KEGON | 12.71 | 11.23 | 2.22 | 2.92 | 24.51 | 31.14 | 52.62 | 47.70 | 7.94 | 7.01 |
| 39 | KKU 35 | 13.30 | 10.98 | 3.05 | 3.54 | 23.66 | 26.97 | 50.60 | 51.92 | 9.39 | 6.59 |
| 40 | KUSL 20004 | 14.37 | 12.55 | 2.42 | 4.08 | 19.86 | 24.15 | 52.32 | 51.71 | 11.03 | 7.51 |
| 41 | MTD 176 | 13.10 | 11.70 | 1.70 | 2.47 | 33.20 | 39.46 | 42.60 | 40.07 | 9.40 | 6.30 |
| 42 | MTD 299 | 13.58 | 11.53 | 1.97 | 2.98 | 32.66 | 40.68 | 41.59 | 38.32 | 10.20 | 6.50 |
| 43 | NS 1 | 16.55 | 12.45 | 1.62 | 2.40 | 32.00 | 40.05 | 42.76 | 38.41 | 7.09 | 6.68 |
| 44 | PI 235347 | 13.85 | 11.08 | 2.30 | 3.08 | 29.61 | 34.10 | 47.23 | 45.01 | 7.01 | 6.74 |
| 45 | PI 371610 | 13.42 | 11.22 | 3.00 | 4.74 | 19.27 | 21.56 | 54.92 | 56.24 | 9.38 | 6.24 |
| 46 | PI 371611 | 11.90 | 10.58 | 3.39 | 4.58 | 20.30 | 23.29 | 57.98 | 55.06 | 6.43 | 6.48 |
| 47 | PI 408169 A | 14.77 | 12.56 | 2.01 | 3.15 | 26.20 | 31.71 | 48.80 | 47.31 | 8.22 | 5.27 |
| 48 | PI 408169 B | 13.46 | 11.19 | 1.92 | 2.96 | 31.88 | 39.22 | 44.82 | 41.52 | 7.93 | 5.11 |
| 49 | PI 408169 C | 15.36 | 12.15 | 2.52 | 3.06 | 30.40 | 36.34 | 45.70 | 43.88 | 6.02 | 4.57 |
| 50 | PI 408169 D | 14.58 | 12.43 | 3.46 | 3.53 | 29.70 | 35.30 | 46.04 | 43.73 | 6.23 | 5.00 |
| 51 | PI 424326 | 13.61 | 11.38 | 2.03 | 3.21 | 27.09 | 29.41 | 50.29 | 50.47 | 6.96 | 5.53 |

**Appendix Table 3** (continued)

| ID | Varieties | Fatty acid composition (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C16:0 | | C18:0 | | C18:1 | | C18:2 | | C18:3 | |
| | | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA |
| 52 | PI 424485 | 12.05 | 10.59 | 2.06 | 2.94 | 28.88 | 32.44 | 50.55 | 48.25 | 6.45 | 5.78 |
| 53 | PU DUA 8008 B | 17.50 | 12.50 | 1.91 | 2.78 | 31.00 | 36.22 | 42.65 | 42.12 | 6.95 | 6.37 |
| 54 | RM 1 | 14.49 | 11.32 | 1.96 | 4.15 | 24.56 | 25.46 | 51.10 | 53.41 | 7.88 | 5.66 |
| 55 | Shimo Shirazu | 12.57 | 10.94 | 2.26 | 2.87 | 35.96 | 41.16 | 43.13 | 39.09 | 6.08 | 5.94 |
| 56 | SJ 1 | 15.03 | 11.51 | 2.21 | 3.03 | 25.81 | 29.65 | 50.40 | 50.13 | 6.54 | 5.68 |
| 57 | SJ 2 | 13.47 | 11.55 | 1.82 | 3.02 | 22.14 | 27.66 | 53.40 | 51.96 | 9.17 | 5.81 |
| 58 | SJ 4 | 14.44 | 11.41 | 1.95 | 3.09 | 21.31 | 26.95 | 54.06 | 52.53 | 8.23 | 6.02 |
| 59 | SJ 5 | 13.32 | 11.48 | 2.84 | 3.12 | 21.48 | 25.13 | 55.41 | 53.98 | 6.95 | 6.30 |
| 60 | SK 8502-4-1 | 15.82 | 11.09 | 3.49 | 3.63 | 21.87 | 25.94 | 51.58 | 52.18 | 7.24 | 7.16 |
| 61 | SOJA DAKO | 16.47 | 12.55 | 1.73 | 2.62 | 21.53 | 25.91 | 48.82 | 51.25 | 11.46 | 7.67 |
| 62 | SRE-B-15 C | 16.04 | 12.43 | 1.82 | 2.93 | 34.37 | 43.08 | 39.88 | 37.64 | 7.89 | 3.92 |
| 63 | ST 1 | 13.90 | 11.53 | 3.41 | 3.95 | 25.62 | 28.93 | 51.59 | 50.26 | 5.45 | 5.33 |
| 64 | ST 2 | 13.96 | 11.57 | 3.99 | 3.64 | 19.25 | 24.25 | 55.97 | 54.53 | 6.82 | 6.01 |
| 65 | Pak Chong 1 | 14.00 | 11.53 | 3.64 | 2.92 | 25.31 | 30.65 | 51.22 | 49.04 | 5.84 | 5.85 |
| 66 | Pak Chong 2 | 10.19 | 10.57 | 2.54 | 2.96 | 33.77 | 34.13 | 48.95 | 45.90 | 4.55 | 6.44 |

**Appendix Table 3** (continued)

| ID | Varieties | Fatty acid composition (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C16:0 | | C18:0 | | C18:1 | | C18:2 | | C18:3 | |
| | | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA | AOCS | USDA |
| 67 | AGS 269 | 14.56 | 11.57 | 4.54 | 3.41 | 16.22 | 27.72 | 58.60 | 51.29 | 6.10 | 6.00 |
| 68 | Laos 429 | 11.34 | 10.82 | 2.43 | 3.25 | 30.13 | 34.12 | 50.22 | 46.51 | 5.88 | 5.29 |
| 69 | Laos 7122 | 16.11 | 12.21 | 4.99 | 3.57 | 17.94 | 32.01 | 54.21 | 46.76 | 6.75 | 5.46 |
| 70 | Laos 8750 A2 | 13.90 | 10.75 | 4.63 | 3.49 | 22.64 | 29.95 | 53.11 | 49.32 | 5.74 | 6.49 |

Fatty acid composition was determined by GC-FID from 65 ground soybean seed samples. These were analyzed by AOCS and USDA-ARS's method: C16:0 = palmitic acid, C18:0 = stearic acid, C18:1 = oleic acid, C18:2 = linoleic acid, C18:3 = linolenic acid.

## Preparation of stock solution

**DNA Extraction buffer**

| Stock solution | Final concentration | volume 100 ml |
|---|---|---|
| CTAB | 2% | 2 g |
| 0.1 M Tris-HCl (pH 8.0) | 100 mM | 10 ml |
| 5 M NaCl | 1.4 M | 28 ml |
| 0.5 M EDTA (pH 8.0) | 20 mM | 4 ml |
| PVP | 2% | 2 g |
| β-mercaptoethanol | 2% | 2 ml |

**TE buffer**

10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0).

**10x TBE buffer**

10x TBE buffer prepared from 108 g of Tris-base, 55 g of boric acid, 7.4 g of EDTA and adjust volume to 1 l by $dH_2O$.

**5x TBE buffer**

54 g of Tris-base, 27.5 g of boric acid, 20 ml of 0.5 M EDTA adjusted volume to 1 l.

**5% denaturing urea polyacrylamide gel**

840 g of urea, 250 ml of 40% polyacrylamide gel: bis (19:1) solution, 200 ml of 5x TBE buffer, added de-ionized water bring volume to 2 l.

**10% ammonium persufate (10% APS)**

5 g of ammonium persulfate adjusted volume to 100 ml.

**1 l of Glass bond solution**

500 µl of Glacial acetic acid, 99.5 ml of 95% ethanol, 300 µl of glass bond

**10 mg/ml of $Na_2S_2O_3.5H_2O$**

0.1 g of Sodium thiosulfate adjusted volume by de-ionized water to 10 ml.

**2.5 l of 10% acetic acid (Fix/ stop solution)**

First, 2 l of de-ionized water, then added 250 ml of glacial acetic acid. Finally, bring volume with de-ionized water to 2.5 l.

**Denaturing dye**

0.02 % xylene cyanol and 0.02 % bromphenol blue, 98 ml of formamide, 10 mM EDTA pH 8.0.

**Φ 174 DNA marker**

50 ng/µl of Φ 174 DNA marker prepared from 20 µl of 500 ng/µl Φ 174 DNA marker mixed with 180 µl sequencing dye.

**Silver staining solution**

Silver staining solution prepared from 2.5 g of silver nitrate ($AgNO_3$) dissolved in 2.5 l of $dH_2O$ and added 3.75 ml of 37% formadehide and kept in amber bottle.

**Developer solution**

Developing solution was prepared with 30 g of sodium carbonate ($Na_2CO_3$), 1.5 ml of 37% formaldehyde, 200 µl of 10 mg/ml sodium thiosulfate ($Na_2S_2O_3$) and $dH_2O$ adjust volume to 1 l. The developing solution was kept in the refrigerator until used.

**Appendix Table 4** Special thanksgiving.

| Name | Organization |
|---|---|
| Ms. Elizabeth J Meyer,<br>Mr. David Walker and<br>for laboratory facility and technical suggestion. | The University of Tennessee,<br>Department of Plant Sciences |
| Assist. Prof. Dr. Amornrat Promboon for kind advice and helpful suggestions in total oil extraction. | Department of Bichemistry,<br>Kasetsart University |
| Assoc. Prof. Dr. Chongrak Kaewprasit for kind advice and helpful suggestions of analytical chemistry and GC technique. | Kasetsart University |
| Assist. Prof. Dr. Julapark Chunwongse for kind advice and helpful suggestions.<br><br>Dr. Jutatape Wacharachiyakup for Plant Physiol-Molecular Biology laboratory facilities.<br><br>Dr. Tanapon Chaisan for working and collect data in the field and laboratory.<br><br>Assoc. Prof. Dr. Sontichai Chanprame and<br>Assoc. Prof. Dr. Sermsiri Chanprame<br>for Plant Molecular Biology and<br>Transformation laboratory facilities. | Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen. |
| Ms. Somnuk Promdang and<br>Mrs. Sainum-oi Sawarngmek<br>for GC laboratory and their technical support. | Phytochemistry Research and Analysis Unit, Central Laboratory and Greenhouse Complex, KURDI<br>Kasetsart Univ., Kamphaeng Saen. |

**Appendix Table 4** (continued)

| Name | Organization |
|---|---|
| Mr. Wasupon Deewan for working and collect data in the field and laboratory. | Kasetsart University |
| Mr. Pirom Linla for assistance in determination the total oil. | Kasetsart University |
| Ms. Malee Panngam for working and collect data in the field and laboratory.<br><br>Mr. Prayoon Prathet for assistance in crossing soybean. | The Project on Genetics and Breeding of Field Legumes for Thailand, Kasetsart University, Kamphaeng Saen. |
| Mrs. Warunee Musch Somanas for polyacrylamide gel electrophoresis training. | |
| Mr. Sompong Chankaew and<br>Ms. Pattama Srinamngoen<br>for laboratory assistance. | |
| Dr. Prakit Somta,<br>Mr. Worawit Sorajjapinun and<br>Dr. Utumporn Sompong<br>for QTL analysis assistance and suggestion. | |
| Dr. Sukhumaporn Sriphadej for suggestion in writing process. | |
| Ms. Sarapee Tangcharoen,<br>Mrs. Suchada Rangkasiwit,<br>Mrs. Chanida Somta,<br>for their helping in this work | |

**Appendix Table 4** (continued)

| Name | Organization |
|---|---|
| Dr. Ruangchai Juwattanasomran,<br>Mr. Nyi Nyi,<br>Ms. Tarika Yimram,<br>Mr. Tanaporn Kajornpon,<br>Dr. Patcharin Tanya,<br>Mr. Tossapol Srisornthong,<br>Dr. Chontira Sangsiri and<br>Ms. Rattanakorn Krisanachandee for their helping in this work | The Project on Genetics and Breeding of Field Legumes for Thailand,<br>Kasetsart University,<br>Kamphaeng Saen. |

# CURRICULUM VITAE

**NAME** : Miss. Jidapa Moongkanna

**BIRTH DATE** : May 25, 1979

**BIRTH PLACE** : Bangkok, Thailand

**EDUCATION** :

| **YEAR** | **INSTITUTE** | **DEGREE/DIPLOMA** |
|---|---|---|
| 1999 | Chulalongkorn University | B.Sc.(2nd Class Honors) (Biochemistry) |

**POSITION/TITLE** -

**WORK PLACE** -

**SCHOLARSHIP/ AWARD** : 2000-2001. Assistant research scholarship, Center for Agricultural Biotechnology

: 2001-2011. The Royal Golden Jubilee, Thailand Research Fund