CHAPTER II

LITERATURE REVIEW

1. Molecular Genetics

Molecular genetics is the science of the biological properties of DNA (deoxyribonucleic acid). DNA nucleotide sequence, encoded biological information, function, regulation and structure are the subjects of molecular genetics. Genome studies with molecular marker technology come from the technology of molecular genetics applied to the traditional problems of classical genetics.

1.1 DNA Structure

The subunits of DNA are four basic molecules called nucleotide. The nucleotides have different nitrogenous bases attached to a phosphorylated five carbon sugar are denoted by those base: adenine(A), guanine(G), cytosine(C) and thymine(T). The A and G bases are similar in structure called purines. The T and C bases are similar and are called pyrimidines (Figure 1). The nitrogenous bases for each strand are held together by the sugar-phosphate backbone. The backbones run in opposite direction, one is in the direction of 5' to 3' and the other is antiparallel in direction 3' to 5'. The two strands of DNA are held together by hydrogen bonds between purines and pyrimidines. The hydrogen bonds are always formed in pairs, G-C and A-T described by the Watson -

Crick rules. Each is called a base pair (bp). The two strands of DNA twist to form a double helix structure (Waterson., *et al.*, 1988 ; Liu., 1997).

Genetic information is encoded by the sequence of base pairs in the DNA strands. The two strands of DNA are complementary, so the base pair sequence of one strand can be inferred from the other strand. It is the convention to write a DNA sequence for only a single strand and in the direction of 5' to 3'. One of the fundamental functions of DNA is to duplicate itself before cell division (Figure 2). The process of DNA replication parallels chromosome duplication. The two strands of the parental helix unwind and each strand direct the synthesis of complementary strand. The process results in two identical double strands of DNA. A complex of additional enzyme associated with DNA polymerises are involved in the DNA replication process (Waterson, *et al.*, 1988 ; William *et al.*, 1997).

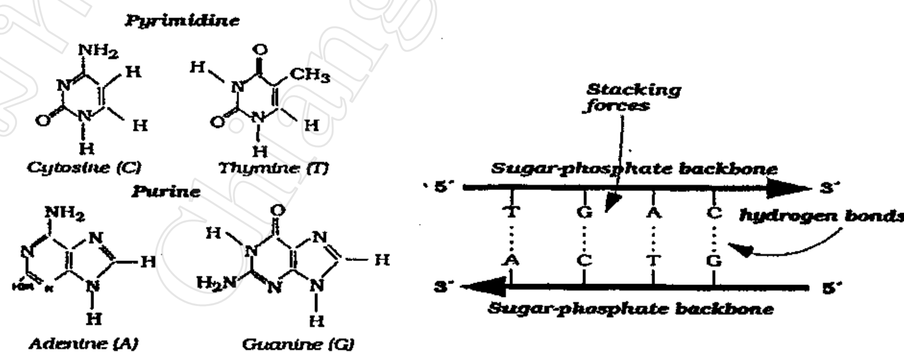


Figure 1 The nitrogenous bases for the four types of nucleotides and DNA structure. The nitrogenous bases for each strand are held together by the sugar-phosphate backbone. The back bones are antiparallel. Two strands of DNA are held together by hydrogen bonds between purines and pyrimidines (Liu.,1997).

Figure 2 DNA replication. The two stands of the parental helix unwind and each of the strands directs the synthesis of a complementary strand (Liu., 1997).

1.2 Gene Expression

DNA can be considered as an information for biological development. The information can be duplicated and transferred from parental cells to daughter cells in the process of cell division. The consequence of the information transfer is the inheritance of the information and a program for expression from generation to generation. DNA is also the information resource for protein synthesis. Through control of production in the cells, DNA controls biochemical pathways, biological development of enzymes and other proteins and responses to change in the environment (Liu., 1997).

The process from DNA to protein starts with a tempolary copy of the information through the transcription of DNA to ribonucleic acid (RNA). Only small fraction of the total DNA in cells is coding DNA which is transcribed to the functionally mature RNA. This process is initiated by RNA polymerase. Only one strand of the double helix is

transcribed and the RNA has a sequence complementary to that strand. The transcription process is highly regulated. Several kinds of RNA polymerases and associated protein transcription factors regulate the specificity and rate of transcription. Promoters, which are regions of DNA that include RNA polymerase binding start sites and transcription start sites, are located 5' to genes and control transcription initiation and level of expression. Some of the non-coding DNA, such as centromeres or telomeres, may have specific chromosome functions. This DNA may have indirect effects on biological functions (William *et al.*, 1997 ; Farell., 1993).

The DNA containing the information to be expressed is transcribed into messenger RNA (mRNA). Not all of the RNA transcript is translated. An RNA splicing process cleaves the pre-mRNA at specific sites and ligates the exons, producing functional mRNA. The sequence corresponding to introns (non-coding sequence) are discarded. The mature functional mRNA moves from the nucleus to the cytoplasm to provide information for protein synthesis. The final structure of a mRNA may vary between organs and tissues, depending upon how the genes were spliced together to create a mRNA. Spliced variants may produce proteins with substantially different function in individual cells, organs and tissue. The process of converting the information encoded in the mRNA into protein is called translation. The groups of three nucleotides in mRNA coding for each amino acid are called codons. The process of translation from RNA to protein polypeptide is highly regulated and many components are involved in the process. The most important RNA is the transfer RNAs (tRNAs). tRNAs are small molecules that attach to and activate specific amino acids. There is one or more specific

tRNAs for each amino acid. This tRNAs bind to the codon by means of complementary base pairing and allow each amono acid to be added at the correct position in a growing peptide chain. The tRNA actually translate the genetic code (Voet *et al.*, 1990 ; Farell., 1997).

1.3 Reading Frame

Each codon is three nucleotides long, therefore, in any sequence of RNA, there are three possible ways to translate the mRNA sequence to amino acid sequence, depending on the starting point for initiation of the translation. These three ways are called reading frames. Only one initiation site and its corresponding reading frame is used for any protein. Other starting sites and reading frames would lead to the synthesis of completely different proteins (Liu., 1997).

2. Genome

2.1 Genome Structure

Genome size can be quantified in terms of molecular mass and number of nucleotide pairs. The important ratio is the kilobase per recombination unit (cM). It is an indicator for how precisely a target gene can be located in terms of physical distance through linkage mapping approaches. The smaller ratio, the more precisely a specific gene can be located physically. This knowledge is important for cloning genes based on their genome locations (Liu.,1997).

At the molecular level, chromosome are composed of DNA and proteins. The DNA carries the genetic information while the protein components provide enzymatic and structural function important to the replication, recombination and segregation of the chromosome (Voet *et al.*, 1990). The main part of the chromosome contains a mixture of coding sequences (exon), intragenic (intron) and intergenic non-coding sequences, regulatory sequences, tandem repeats (mainly minisatellites and microsatellites) and other dispersed repeats. Tandem repeats have been used in genomic research such as minisatellite markers, commonly referred as variable length repeat (VNTR) (Ponsuksili *et al.*, 1996).

2.2 Source of Genome Variation

Chromosome Rearrangement

Chromosome rearrangement is one important source of genome variation. Either due to abnormal pairing or recombination during meiosis, such rearrangements are usually associated with deleterious effects. Rearrangements can be described as insertions, deletions, inversions, translocations or duplications (Ponsuksili *et al.*, 1996).

Point Mutation

Polymorphism can also be created by point mutation, which is the substitution of one nucleotide for another. Nucleotide substitution is a common type of mutation and can be classified as a transition or a transversion mutation. A transition results in substitution of a purine nucleotide by a purine nucleotide, or of a pyrimidine by a

pyrimidine. Conversion of an A to a G is an example of transition. Transversion results in substitution of a purine nucleotide by a pyrimidine or vice versa. Conversion of an A to a C is an example. Another commonly observed mutation is called a framshift mutation. This mutation is created by insertion or deletion of one or more (except multiples of three) nucleotides and will cause a shift of the reading frame during the translation of mRNA to protein, usually resulting in a truncated, non-functional protein (Liu., 1997 ; Cater and Morgan., 1999).

Polymorphism

Polymorphism is a variation in excess of mutation-selection balance expectation (Cater and Morgan., 1999). The polymorphisms are caused by heterozygote advantage, frequency dependent selection, heterogeneous environment, transition between selective conditions on evolutionary time scale and neutral mutation. The Human Genome Research Institute (HGIS) expected that Single nucleotide polymorphisms (SNPs) located in expressed mRNAs. SNPs are used to speed the identification of genes that give rise to inherited genetic diseases (Human genome science., 2000).

Polymorphism is defined as detectable heritable variation at a locus. No matter what level of genome variation is used, the polymorphism and informativity of the marker can be defined in the same way. For a single experimental population produced from a controlled cross, the number of alleles and population frequencies of alleles may be accurately determined. A genetic marker has to be a polymorphic marker but a polymorphic marker may not be a genetic marker. A genetic marker may be

operationally defined as a heritance polymorphic marker with clear genetic interpretation and repeatability. Genomic analysis using genetic markers should be based on well eatablished genetic models. If the underlying genetics of marker are not clear, then the analysis may be misleading. It is also important that the marker assay is repeatable at different times in the same or different laboratories. Different type of markers may identify different polymorphisms. The genetic interpretation of a marker strongly depends on the sequence complexity of the genome and the kind of variation the marker identifies (Liu., 1997).

3. Biological Techniques in Genomics

The gene mapping approach has revolutionized the process of finding genes compared to the traditional approachs, without knowing what QTLs are. QTL mapping is limited in terms of fundamental biology and knowledge of mechanism behide the phenotypes. Not knowing what QTL are may led to wrong genetic models, imprecise definitions, makes complete genetic models difficult to construct and QTL mapping inefficient in practice.

3.1 Genetic Mapping

A genetic or linkage map is an abstract model of the linear arrangement of a group of genes and markers. Genetic map is developed from the inheritance or segregation patterns of marked chromosomal fragments. Because this maps are built up by looking at co-segregation between genetic markers, it makes a lot of senses for several groups to

collaborate in mapping different markers with the same set families. The gene can be the traditionally defined Mendelian factor or a piece of DNA identified by a known function or by means of a biochemical assay (Wimmer *et al.*, 1999). Genetic mapping requires pedigrees in which the relationship are known and polymorphic genetic markers. The marker can be a cytological marker, a variant based on a change in a known gene or protein or a piece of DNA without known function. A gene with known function can be considered a marker if it contains detectable variation. Both the gene and the marker should have simple inheritance that can be followed through generations. Recombination or crossing over can occur between pairs of homologous chromosomes that some of the chromosomes passed onto the next generation will be part paternal and part maternal in origin. A genetic map is based on homologous recombination during meiosis, so a genetic map is also a meiotic map. If two or more markers are located close together on a chromosome, their alleles are usually inheritance together through meiosis (Ponsuksili *et al.*, 1996 ; Archibald and Haley., 1998).

One of the most important applications of genetic maps is to locate specific genes of interest, such as controlling traits of economic importance in animals. Most of these traits controlled by more than one gene, involving in a quantitatively varying phenotype, the loci are commonly described as Quantitative Trait Loci (QTL) and the mapping procedure is called QTL mapping (Liu., 1997).

3.2 Physical Mapping

One goal of genomic analysis is to obtain DNA sequence and information on how the DNA is transcribed and translated in development. Complete genome DNA sequences have already been obtained for several bacteria, since their genomes are small compaired to higher organisms. However, for most animal species, it is not yet practical to obtain complete direct correlations or associations needed between traits and DNA sequences in order to understand the genes controlling complex traits. Genetic mapping is one way to make the associations to bridge this gap. Although the genetic markers may be generated using recombinant DNA techniques or polymerase chain reaction, genetic maps are still made by traditional Mendelian analysis. (Liu., 1997 ; Archibald and Haley., 1998).

A physical map is another mechanism that can be used to make associations between traits and DNA sequences. A physical map can be the traditional cytogenetic chromosome map, based on chromosome structure or banding patterns observed using modern cytogenetic chromosome mapping. The commonly described physical map is one which contains ordered overlapping cloned DNA fragments. The cloned DNA fragments are usually obtained using restriction enzyme digestion. The restriction map is a common form of a physical map of a chromosomal segment, but the ultimate physical map is the DNA sequence. In general, when making the restriction physical maps, the larger the fragment, the better the analysis. This is usually achieved by using rare-cutting endonucleases and partial digestion. A procedure to select fragment with the desired

large size is applied after the digestion, using centrifugation or electrophoresis (Ponsuksili *et al.*, 1996 ; William *et al.*, 1997).

3.3 DNA Vector

To propagate and maintain the DNA fragments, it is necessary to insert the DNA fragments into cloning vectors. A vector is a cloning vehicle for replication the fragments. Commonly used vectors are phage, cosmids, yeast artificial chromosomes (YAC) and bacterial artificial chromosomes (BAC). The important considerations in choosing a vector system are the size of DNA fragment that can be inserted into the vector and the stability of the product. For phage, cosmids, YAC and BAC, the average sizes of the insert are 5-25 kb, 35-45 kb, 200-2000 kb and <300 kb, respectively. Smaller inserts are usually more stable than larger (Liu, 1997). The size of the inserts determines how many clones are needed to include essentially every sequence from a genome in the genomic library (Sambrook *et al.*, 1989). Once a genome has been disassembled and parts (cloned fragments) are in a storage bank (genomic library), the next step of physical mapping is to determine a linear order of the original genome before the digestion. There are several methods for assembling physical maps. The goal is to establish which clones overlap. The ultimate way to assemble a physical map is by sequencing all the clones (Figure 4) (Liu., 1997., William *et al.*, 1997).

3.4 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a DNA synthesis technique that amplifies specific regions of DNA that lie between two sites defined by the complementary sequences of two specific primers. A series of synthetic reaction are catalysed by a DNA polymerase. PCR based procedures to obtain DNA fragment profiles have a higher sensitivity and are less time consuming than the hybridisation based procedures (Clauster., 2001).

The principle of PCR is illustrated in Figure 3. The template DNA is first denatured by heating in the presence of a large molar excess of each of the two oligonucleotides and four deoxyribonucleoside triphosphates (dNTPs) in an appropriate buffer solution. The reaction mixture is then cooled to a temperature that allows the oligonucleotide primers to anneal to their target sequences, after which the annealed primers are extended with thermal stable *Tag* DNA polymerase (William *et al.*, 1997). The DNA polymerase is inactivated at the temperatures that are required to denature DNA (92-97°C), under normal reaction conditions, the amount of *Tag* DNA polymerase becomes limiting after 25-30 cycles of amplification (Kolmodin and William., nodate). For the first cycle of amplification, two DNA strands extend from each of the primers along the original templates without defined termination sites. In cycle two, the primers initiate synthesis on the products of the first cycle and two new strands are synthesized. The second strand must terminate at the original primer sites where the new template ends. Further amplification is therefore restricted to the sequence between sites defined by the primer. As further cycles proceed, the number of DNA fragment increase

exponentially. After an amplification, level of approximately 10^6 copy has been obtained. For requirement of amplification conditions, such as annealing temperature, primer length, G : C content of primers and other aspects of the protocol may vary. The polymorphisms identified using PCR based markers where sequence is known represent variation in segment length between the two primers (Clauster., 2001).

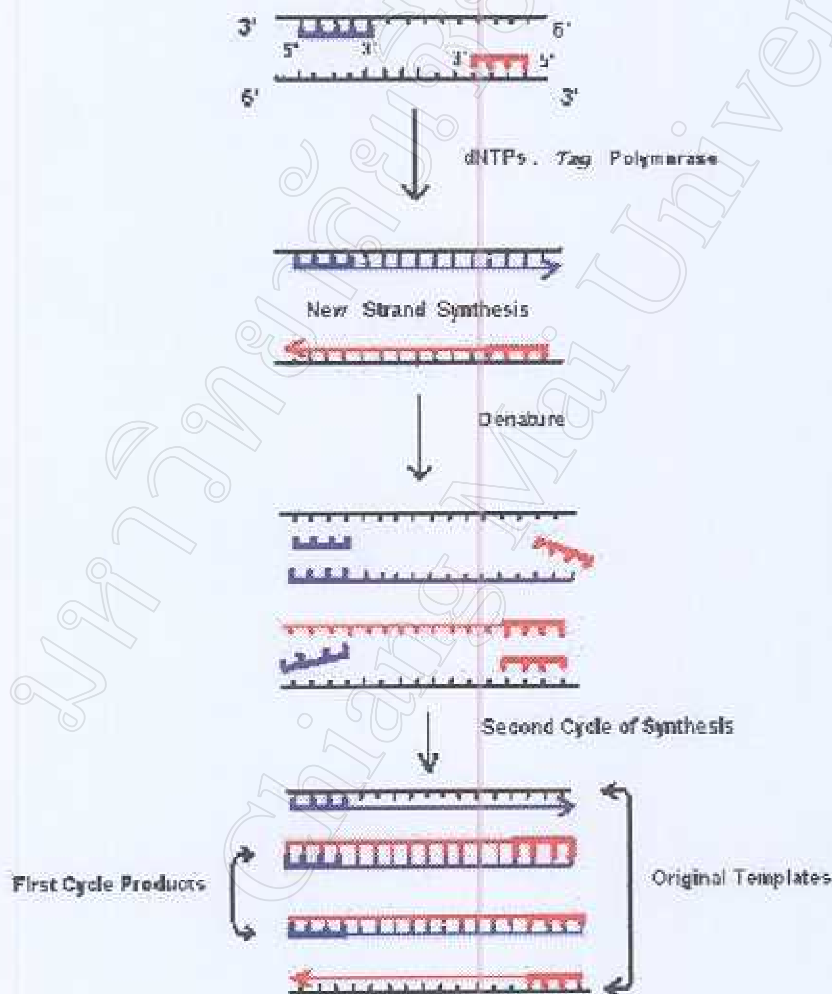


Figure 3 Principle of polymerase chain reaction (Wirairat., 1997).

3.5 DNA Sequencing

The ultimate goal of genome research is the determination of complete DNA sequences and the biological roles or effects of these sequences. Genetic mapping and physical mapping can be considered to be early steps toward reaching that goal. Development of automated, high throughput DNA sequencing technology has been a focus of biological research, both public and private. Even through many improvements have been made, the most widely used method of DNA sequencing is still based on procedures developed by Sanger and co-worker in 1970s. The basic strategy is to create a population of DNA molecules differing in length by one nucleotide increments. These molecules are then separated by electrophoresis in polyacrylamide gels that resolve the single nucleotide differences (Sambrook *et al.*, 1989).

3.6 Genomic Informatics

Genomic DNA is an information bank of DNA sequences. The purpose of genetic mapping, physical mapping and DNA sequencing is to explore how the information is organized and how it is used in growth, reproduction and development. Genomics is becoming an information science as the genetic information stored in cells in DNA and transferred from molecules to computers. The amount of data on genetic maps, physical maps and DNA sequences stored in computers is growing exponentially. The well-organized data bank that exists as DNA molecules is not necessarily well organized when entered into computers. Genomic informatics has three functions: information management, analysis and communication. Several genome databases are

now available for different organisms, including *E.coli*, yeast, nematode (*C. elegans*), *Drosophila*, mouse, human (CMG-Mutation Detection., 1998). GenBank is a DNA sequence database maintained by the U.S. National Center for Biotechnology Information (NCBI). EMBL is the European Molecular Biology Laboratory database. DDBJ is the DNA Database of Japan. The current genomic data banks have the function of bookkeeping, communication and limited analytical functions (Liu., 1997).

3.7 Traits Maps and Sequences

It is not possible to directly relate most complex traits with the already massive DNA sequence databases, because little is known about the molecular identity of most genes controlling complex traits. However, DNA sequence is the ultimate information resource for the traits. Reverse genetics, from protein to DNA, is an elegant way to isolate genes with simple functions. Many genes have been identified and cloned using the reverse genetics approach. The coordination among genetic and physical maps and DNA sequences may provide more efficient ways to isolate genes with simple functions and possibly to isolate gene controlling complex traits (Figure 4). Genetic and physical maps are bridges between complex traits and DNA sequences (Liu., 1997).

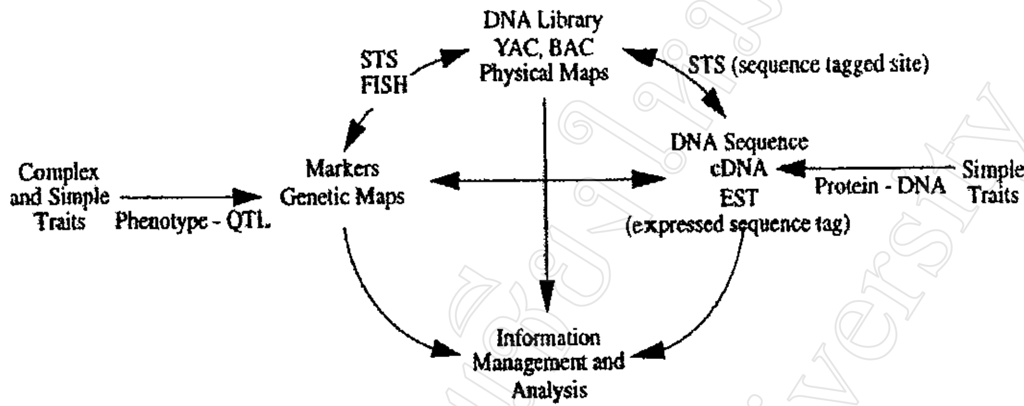


Figure 4 Connecting traits, maps and DNA sequence (Liu, 1997).

4. Mapping Populations

Selection of population for genetic mapping involves choosing parents and determining a mating scheme. Decisions on selection of parents and mating design, as well as the type of markers, should depend upon the objectives of the experiment (Archibald and Haley., 1998).

Parents of a mapping population must have sufficient variation for the traits of interest at both the DNA sequence and the phenotypic level. The variation at the DNA level is essential to trace recombination events. The more DNA sequence variation exists, the easier it is to find polymorphic informative markers. When the objective of the experiment is to search for genes controlling a particular trait, genetic variation of the trait between the parents is important. If the parents are greatly different at the phenotypic level for a trait, there is a reasonable chance that genetic variation exists

between the parents, although uncontrolled environmental effects could create large phenotypic variation without any genetic basis for the effects. Allele with large effects segregate in the intercross, a relatively high marker heterozygosity improves the information content compare to within population studies, and high heterozygosity at the QTL, the regions of the genome influencing quantitative traits (Anderson., 1998). However, lack of phenotypic variation between the parents does not mean that there is no genetic variation. Different sets of genes could result in the same phenotype. In some cases, variation within individuals, based on a high level of heterozygosity, may be exploited in controlled crosses. A highly heterozygous individual may be treated as an F1 in some experimental designs. Different sets of genetic markers have different levels of resolution for detecting genomic variation. For some species, little genomic variation exists in natural populations. In addition to consistent linkage phases make the statistical analyses powerful, a disadvantage from a practical breeding perspective is that the QTL detected may not be segregating within the commercial populations of interest. A population with overall inferior performance can contain genes that enchance performance or product quality (Liu., 1997). Thus, some of the detected QTL may be of great interest for introgression into commercial populations. However, technology is available to detect even a single base change in a DNA fragment of interest such as by direct DNA sequencing (CMG-Mutation Detection., 1998).

The candidate gene hypothesis propose that a major portion of trait quantitative genetic variation is caused by functional variation in the genes directly involved in trait development or physiology (Wimmers *et al.*, 2000). The hypothesis is supported by the

compatibility of the widespread random pleiotropy with evolvability, by a high quantitative mutation rate, by position cloning and the studies showing that mutation trait and disease phenotypes are generally caused by mutation in candidate genes (Liu., 1997).

5. Genetic Markers

Genetic markers are widely used to identify genes and regions of the genome associated with economically important traits. Three types of genetic markers have been used in genomic analysis, morphological markers, protein based markers and DNA based markers. Variation among phenotypes within a species is the raw material for genomic analysis. To be genetic marker, the marker locus has to show experimentally detectable variation among the individuals in the test population. The variation can be considered at different biological levels, from the simple heritable phenotype to detect of variation at the single nucleotide. Once the variation is identified and the genotypes of all individuals in the test population are known, the frequency of recombination events between loci is used to estimate linkage distances between markers (Ponsuksili *et al.*, 1996 ; 2000b).

Morphological markers are morphological traits such as shape, color, size or height that often have one to one correspondence with the gene controlling the traits. The morphological characters can be used as reliable indicators for specific genes and as useful as genetic markers on chromosomes. A subset of the morphological markers may be due to simple point mutation representing changes in single gene (Wutthipong *et al.*, 2000). A protein marker is protein which different in charge or size detected by using gel electrophoresis because of different allele of gene may result in proteins with different

amino acid composition, sizes or modifications. Much of the detectable protein variation identifies allelic sequence variation in the structure gene encoding the protein or at a regulatory sequence (Liu., 1997). Alternatively, some protein variation is due to post translational modification and is not useful as genetic markers (Farell., 1993).

DNA marker is typically a small region of DNA showing sequence polymorphism in different individuals within a species. Two basic approaches have been used to detect variation in the small region of DNA. The fragment can be detected by nucleic acid hybridisation which uses another fragment from the same locus which has been isolated and purified from the same related species. The previously known segment must share considerable DNA sequence homology with the fragment of interest and can be labelled and used as a probe to detect the fragment of interest by complementary base pairing. This is the approach based on the amplification of sequence using Polymerase Chain Reaction (PCR) to amplify a target fragment, two primers which flanking the target sequence designed using known sequence of the segment are needed (Liu.,1997). Microsatellites, STSs (sequence tagged sites). ESTs (expressed sequence tags), etc. have been commonly used as genetic markers based on sequence specific PCR. Markers of this type include RAPD (random amplified polymorphism DNA) and AFLP (Amplified fragment length polymorphism) (Mathuros., 1997).

6. Vinculin

In recent studies using genomic scans to detect quantitative trait loci in swine, for traits like body composition traits, fat deposition, eye-muscle area, muscling and carcass length which are the important traits in the price of market hogs considerable progress has been made. The QTL mapping of these traits has been reported for Large White and European wild boars (Anderson *et al.*,1998) and Meishan–White composite boars (Rohrer and Keele., 1998a,b), showing that there are several chromosome regions with a considerable effect on carcass traits in pigs. Genetic markers for the trait eye-muscle area and carcass length have been studied in Berlin-Bonn resource population by Ponsuksili *et al* (1999,2000b) identified genes potentially effecting carcass traits based on the differential expression of these genes in different breeds and/or individuals by differential display method. Focus was on the trait eye-muscle area. A resource family (RF) derived from a cross of Berlin Miniature Pig and Duroc and a herd of German Landrase (GL) were investigated. The trait eye-muscle area has a high heritability and high variation and can serve as a valuable model case for the evaluation of the suitability of differential displays for the detection of candidate genes. The expressed sequence tags (ESTs) are derived from cDNA clones. The advantage of using ESTs is that genomic regions which representing functional sequences are identified that are therefore more useful as genetic markers than non-functional sequences. The use of the differential display-reverse transcriptase-polymerase chain reaction (DD-RT-PCR) technique can detected ESTs which represent genes that are differentially expressed in phenotypic extremes in the trait eye-muscle area. By comparing the DD-RT-PCR banding patterns of RNA-pools of four groups of animals which were the extreme ones in the trait eye-

muscle area within a resource population, (F2 cross of Duroc and Berlin Miniature pig), and a pure bred German Landrace herd, seven bands have been analysed by sequencing and search for homology and mapping using a somatic hybrid panel. Three clones did not show any homology, two clones were homologous to an EST (SSC25004) and SINE sequence SSPRE (x64127), two showed high homology to known genes which were PTEN (Swiss Prot SP00633) and vinculin (Swiss Prot P18206), respectively. Differential expression of these ESTs has been confirmed by semi quantitative RT-PCR. (Ponsuksili *et al.*,1999; 2000b).

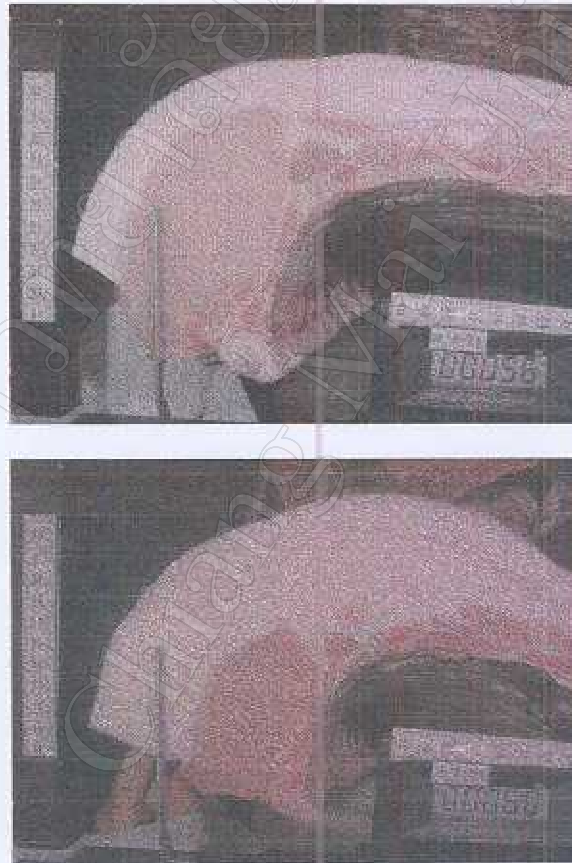


Figure 5 The measurement of loin-eye area of animal selection for the high and low performing groups to detection differential expressed gene by DD-RT-PCR (Ponsuksili *et al.*, 1999 ; 2000b).

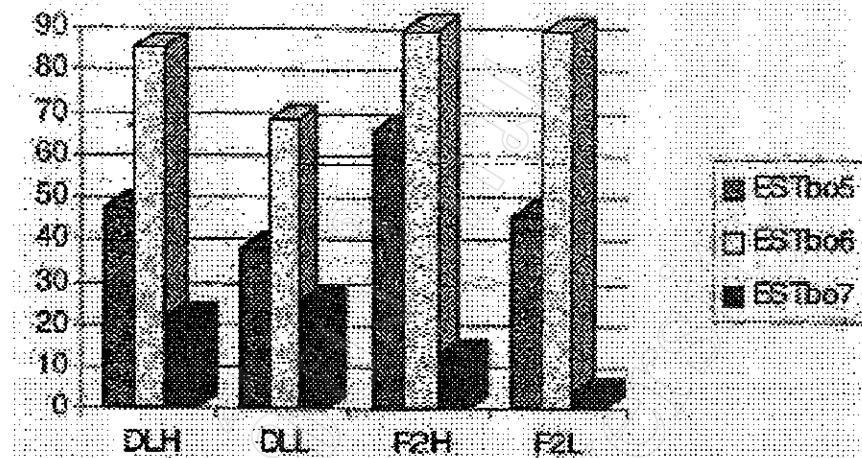


Figure 6 Comparing of RT-PCR analysis of differentially amplified fragments of the four groups GL(H), GL(L), F2(H), F2(L). Diagram indicating the intensity of the bands of the ESTbo5(EST,SSC25004), ESTbo6 (PTEN), ESTbo7 (Vinculin), relative to the intensity of the bands of beta-actin which was set to 100 % (Ponsuksili *et al.*,2000b).

From the comparson of the expression of bands of the seven ESTs (figure 6) more differences in the expression patterns were found between F2 animals than between high and low performing animals of GL. The F2 animals showed highest variation in the trait which was used to select the individuals. The variation was also high between the breeds but it was the lowest within the GL population ESTbo/, representing the vinculin gene, was found to be differentially expressed (1) between high and low performing animals of the resource population and (2) of a commercial GL herd and (3) between the breeds. Therefore, the vinculin gene is a promising functional candidate gene for the traits eye muscle area and other carcass traits (Ponsuksili *et al.*, 2000).

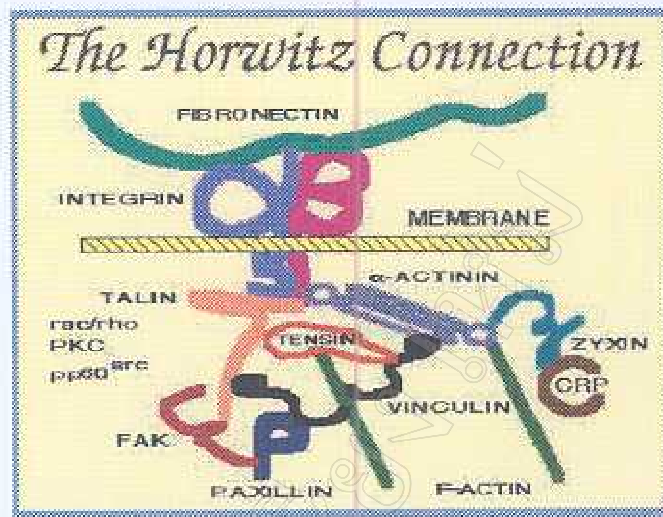
6.1 Role and Structure of Vinculin

Vinculin is a cytoskeletal protein that localizes in the cytoplasmic plaque of microfilament associated of both cell-cell and cell-extracellular matrix adherens-type junctions. It functions as one of several interacting proteins involved in anchoring F-actin to the membrane (Byrne *et al.*,1992 ; Wood *et al.*,1994 ; Volberg *et al.*, 1995 ; Mayer *et al.*, 1997). Adhesion of cells to the extracellular matrix is a critical step in such diverse biological process as normal cell growth, development and differentiation, metastasis and embryological development (Hynes., 1992). In the assembly of adhesion plaques, the beta-subunit of integrin bind to talin. Talin binds to vinculin which interacts with a-actinin and possible with itself. Since a-actinin is known to bind to and crosslink actin filaments, vinculin represents a key element in the transmembrane linkage of the extracellular matrix to the cytoplasmic microfilament system (Figure 7 and 8). Vinculin plays a role in signalling between integrins and the actin cytoskeleton, including a functional role in regulating motility of the cell and in signalling to the nucleus and other organells that regulate cell behavior (Lockusch and Riedige., 1996). Vinculin acts broaGLy in the cell as a link in the molecular bridge between the extracellular milieu and the intracellular responsive molecules and structures. It has been shown that a sequence of molecular interactions might be involved in the transmembrane assembly of adhesion plaques. Interaction between matrix proteins and their receptors is thought to provide the trigger for assembly of these specialized cellular junctions and protein phosphorylation is likely to be involved in the regulation of their stability (Weller *et al.*, 1990). The responses to signal from outside reach the actin cytoskeleton and by yet unknown mechanisms are manifested as shape and spreading changes, perhaps

brought about by change in viscosity, elasticity, and other physio-dynamic characteristics, and by other changes such as cell motion over the substrate. Xu *et al.* (1998) reported that the head domain is a part of the locomotory force of the cell, modulated by the tail and driven by the integrin or matrix conection and intact vinculin is required for normal regulation of cell behavior, suggesting that vinculin head and tail interactions control cell adhesion, spreading and locomotion. Thus, vinculin does not play a role in mechanical transduction that significant mechanical forces are not transmitted through vinculin and also that loss of vinculin at the focal adhesion site does not have significant effects in the short term on cell adhesion (Meyer *et al.*,1997). Microinjection of monoclonal antibodies (15B7) to vinculin into adherent fibroblasts disturbs cell-matrix junctions and the associated actin microfilament system (Westmeyer *et al.*, 1990).

Volberg *et al.* (1995) reported that the organization of actin, talin, alpha-actinin, paxillin and phosphatyrosinated components was essentially identical in the vinculin-containing and vinculin-null cells suggest that the linkage of actin to the membrane includes a net work of multiple interaction, the interaction between the various plaque proteins and each of these proteins can bind to focal contacts also in the absence of vinculin. Quantitative analysis indicated that the adhesion plaques in vinculin deficient cells stain more intensely by anti-alpha-actinin, talin and paxillin antibodies, suggesting alternative linkage mechanisms between actin filament and the membranes in adhesion plaques which may compensate for the lack of vinculin.

Figure 7 Integrins (orange) span cell membranes. They hold a cell in place by attaching at one end to molecules of the extracellular matrix or to molucules on other cells and at the other end to the cell's own scaffolding, or cytoskeleton. They connect to this scaffolding through a highly organized aggregate of molecules, a focal adhesion, that includes such cytoskeletal components as actin, talin, vinculin and a-actinin (blue) (Horwitz Lab Overview., 2000).



Fiugure 8 Proteins associated in adheren type junction (Horwitz Lab Overview., 2000).

The pattern of vinculin expression is very dynamic. In muscle, vinculin is one of 6 to 8 proteins constitutes in the filamentous structure that linked myofibrils to the sarcolemma which were named costamere, extend into the muscle cell where they encircle myofibrils to the Z-disk and run from myofibril to myofibril to link adjacent Z-disks laterally (Figure 9) (Taylor *et al.*, 1995 ; Vohra *et al.*, 1998 ; Illingworth., 2000). In smooth muscle, expression of two variants of vinculin (gamma-vinculin and meta-vinculin) is phenotype-dependent and developmentally regulated. The different between vinculin and meta-vinculin residues solely in an insert within the tail domain which comprises 68 or 69 amino acid in porcine or chicken smooth muscle, respectively (Gimona *et al.*, 1988 ; Byrne *et al.*, 1992 ; Riediger *et al.*, 1998). Vinculin is localized in neuromuscular junctions and the membrane-associated dense bodies. Vinculin expressed significantly higher concentration on myotendinous junction of *mdx* mouse's muscle, result from absence of protein dystrophin which cause of Duchenne muscular dystrophy and myopathy (Law *et al.*, 1994). During

avian development, almost all cells derived from the three primary germ layer contain approximately equal levels of vinculin. The expression of vinculin was found to increase transiently in restricted areas of ectoderm and endoderm undergoing extensive folding and vinculin was found at high levels in migrating neural crest cell. In many cell types undergoing viral transformation, vinculin becomes redistributed to rosettes or podosomes of culture cell (Byrne *et al.*, 1992 ; Moiseyeva *et al.*, 1993 ; Sigma., 2000).

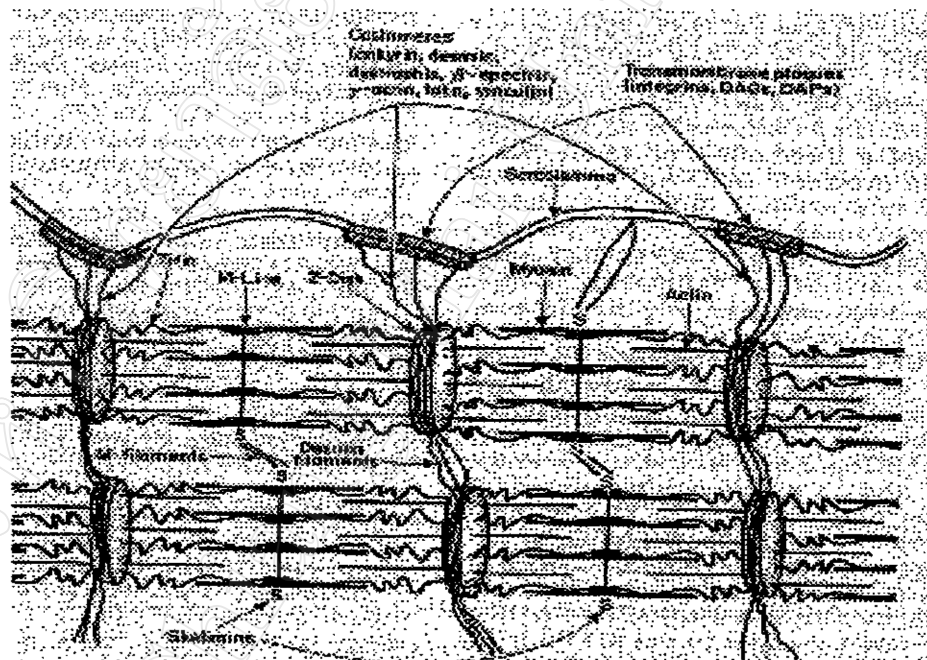


Figure 9 This structure showing the protein composition of costameres in striated muscle relative to Z-disks and the myofibrillar lattice. The costameres contained alpha-actin, vinculin, beta-spectrin, talin, and the intermidiate filament proteins, desmin and vimentin (Taylor *et al.*, 1995).

Baestead and Waterston (1991) found that mutations in the nematode vinculin gene are lethal early during larval development and muscle twitching, which preassumably requires anchorage of actin to the body wall, is notably absent because vinculin is a component of the dense bodies in the body wall muscle of the nematode. The finger-like structures project from the basal sarcolemma in the highly ordered array associated with the regular structure of striated muscle. Like the Z-disks of vertebrate striated muscle, the dense bodies bind actin filaments of opposite polarity. Their firm linkage to the sarcolemma, however, suggests that dense bodies are also analogous to focal adhesions (Francis and Watersion., 1985). The nature of the nematode dense body maybe similar to the actin-membrane attachments in nascent vertebrate muofibrils of developing striated muscle which also have the characteristics of both focal adhesion and Z-discs. Vinculin is phosphorylated on serine, threonine and tyrosine although the kinase that catalyse these even remain to be properly characterized. The evidence for limited alternative spicing of the vinculin gene seems likely that much of the molecular heterogeneity in vinculin is due to post-translational modifications (Turner *et al.*,1989). Vinculin is a member of the family of immediate early response gene whose transcription is rapiGLy activated by a wide spectrum of growth-inducing factors in a manner which is not dependent on new protein synthesis. Such genes include c-*fos* and c-*myc* as well as those that encode proteins involved in cell adhesion such as fibronectin, beta1-integrin, alpha-tropomyosin and beta-actin. Addition of serum to quiescent Swiss or Balb/c mouse 3T3 cells leads to a rapid (40-60 min.) and transient increase in transcription of the vinculin gene and an increase in vinculin mRNA and protein levels. In Balb/c 3T3 cells, transcriptional activation was

also achieved with platelet derived growth factor, fibroblast growth factor (Bellas *et al.*, 1991 ; Moiseyeva *et al.*, 1993).

6.2 Protein Structure

Vinculin is large protein of 117 Kd. (Xu *et al.*, 1998). Structurally the protein consists of an acidic N-terminal of about 90 Kd. separated from a basic C-terminal domain of about 25 Kd. by a proline-rich region of about 50 residues. The central part of the N-terminal domain consist of a variable number, 3 in vertebrates, 2 in *Caenorhabditis elegans*, of repeats of a 110 amino acids domain (Byrne *et al.*,1992). Thus, the binding domain for the cytoskeletal protein talin spans vinculin residues 1-258 and this sequences are required for targeting talin to focal adhesions (Gilmore *et al.*, 1992) and major phosphorylation sites which are the substrate of protein kinase C (PKC) (Schwienbacher *et al.*,1996). The binding site for the cytoskeletal protein paxillin has been localized to the C-terminal at a region of 50 amino acid residues 979-1028, the C-terminal residues are important in localisation of vinculin to focal adhesions (Turner *et al.*, 1990 ; Wood *et al.*, 1994).

6.3 Vinculin Gene

The sequences of nematode, rat, chicken and human vinculin have been determined (GeneBank: J04804, L13299, M87837, L04933 respectively). The complete sequences of both human and chicken embryo sequence of vinculin gene contain 1,066 amino acid and 2 proteins exhibit a high level of sequence identity, greater than 95% and the human vinculin gene had been mapped on chromosome

Copyright by Chiang Mai University

10q11.2-qter by Weller *et al.* (1990). Mulligan *et al.* (1992) used a combination of physical and genetic mapping techniques to refine this localization. Hybridisation of the vinculin cDNA probed to a human-rodent somatic hybrid panel and comfirmed by hybridizytion of the vinculin cDNA to flow-sorted translocation derivative chromosomed indicated a position of human vinculin gene is mapped in 10q22.1-q23,distal to D10S22. Moiseyeva *et al.* (1993) have shown the complete sequence of human vinculin gene that containing 22 exons ranging in size from 71 base pairs (bp) to 303 bp, average 155 bp, of coding sequence and 1848 bp of 3'-untranslated sequence including two polyadenylation signals. There is a limited correlation between exon boundaries and functional domains within the vinculin molecule. Analysis of vinculin mRNA in human uterus showed that alternative splicing of the gene is limited to exon 19 which encodes the 68 amino acids included in the muscle-specific isoform called metavinculin, porcine and chick metavinculin only differ from vinculin by inclusion of the metavinculin exon. The vinculin promoter does contain a functional serum response element. Only 5 amino acid differences between human and chick vinculin over the first 258 residues and human vinculin residues 1-258 and 820-1066 display a 54% and 61% identity with nematode vinculin. The structure of the 5' end of the gene is highly conserved between humans and chickens (Wood *et al.*, 1994). The swine vinculin sequence have only 676 bp sequenced from swine smooth muscle RNA (Gene Bank: Z19540).

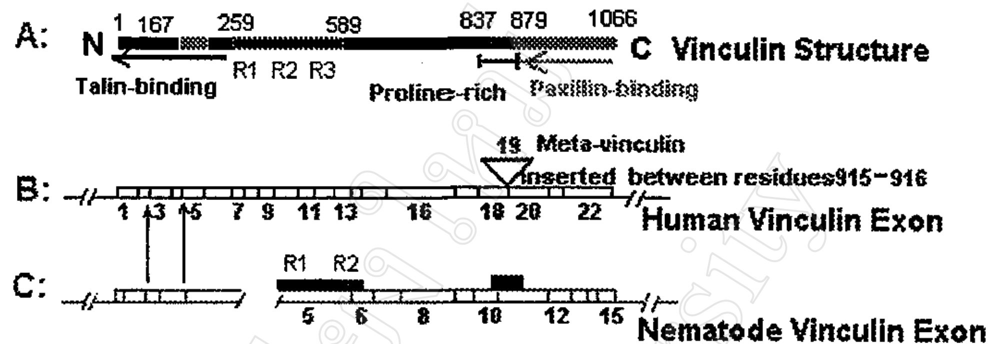


Figure 10 Aligment of a series of PCR-generated chick vinculin cDNAs with respect to the structure of the vinculin molecule. The domain structure of vinculin (1,066 amino acids) is represented schematically to show the positions of the talin-binding domain, the three 112 amino acid repeates, the proline-rich region, the V8-protease cleavage sites and the paxillin-binding region (Wood *et al.*, 1994).

The objectives of this work are:

1. To identifiy polymorphic sites in the cDNA sequence of the porcine vinculin gene.
2. To establish protocols for PCR-based genotypic at the polymorphic sites of the vinculin gene.
3. To physically map the porcine vinculin gene.