CHAPTER 2

LITERATURE REVIEW

2.1 Thai native pigs

Thai native pigs are classified as lard type pigs. They grow slowly and their reproductive rate is low. They, however, adapt well to hot and humid climates, tolerate low quality feed, and probably are resistant to foot and mouth disease and internal parasites. Thai native pigs are generally classified into four “breed lines” according to physical appearances and regions where they are predominant such as Raad, Puang, Hainan, and Kwai (Figure 2.1). These breed lines were common 30 to 40 years ago (Rattanaronchart, 1994). However, it was difficult to determine real characteristics specific for each breed line.

2.1.1 Classification of Thai native pigs

First, pigs in Lower Northeastern Thailand belong to the "Raad" breed, in some places alternatively called "Ka Done". The Raad pig is the smallest pig among the Thai native pigs, with a mature body weight of 60 to 70 kg. It has a black hair coat color, a small head, small and erect ears, a long and straight snout, and a short body. Raad pigs are relatively low in prolificacy, with five pairs of teats and a litter size of about six. Secondly, pigs in the Upper Northeastern Thailand are "Puang" breed. The Puang pig is much bigger than the Raad pig; its mature body weight may reach 120 to 130 kg. They are black, with large thick ears, and a rough and wrinkled skin. They are similar to the Chinese Taihu pig. However, they are excellent suitable for mating with the Raad breed. Thirdly, pigs in Central, Eastern, Western, and Southern Thailand are "Hainan" breed (pigs mainly raised by Hainanese immigrants). The mature body weight of a Hainan pig is 110 to 120 kg. Hainan pigs have generally a black and a white hair coat color, with black color at the head, back, and rump. They are white on the belly and legs. The back is

concave, and the belly is drooping. They have short and a straight snout, small and erect ears, and a weak pastern. They are the most prolific Thai native pigs with six pairs of teats and eight piglets per litter. Finally, pigs in Northern Thailand are called "Kwai" (Kwai means buffalo or big in Thai). Kwai pigs are the biggest animals among the Thai native pigs. Their mature body weight is 130 to 150 kg. Their appearance is similar to Hainan pigs except they have a black hair coat color, with white legs, longer and straight snouts and larger ears (Rattanaronchart, 1994).

Figure 2.1 Thai native pigs: (A) Raad, (B) Puang, (C) Hainan, and (D) Kwai.

2.1.2 Situation of Thai native pigs

At present, the numbers of native pigs in Thailand has largely decreased due to the popularity imported commercial breeds which have a higher productive performance. Since 20 to 30 years, many Thai native pigs have been mated with European commercial breeds (e.g. Large White, Duroc, and Hampshire) as well as with the Meishan breed to improve economic traits, such as Average daily gain

(ADG), Feed conversion ratio (FCR), reproductive performance and meat percentage. In the recent past, more and more effective breeding programs have been implemented and have led to an emphasis on a few specialized stocks. Breeds that are less suited to current needs tend to see their numbers declined. Currently, the native pigs are replaced by pure European pig breeds especially by the pig industry. Thai native pigs in the lowlands have decreased in number due to its fewer than 500 breeding females and less than 10 herds (Maijala, 1992 cited by Rattanaronchart, 1994). Thai native pigs are still rised by some distrinct groups of people in the highland and in some villages which will presumably be a major viable genetic resource of Thai native pigs for the future use (Theera and Choke, 1980). Because only small groups of native pigs have been raised, on generally known risk has exited to lose in genetic diversity caused by genetic drift, inbreeding problems and the introduction of crossbred pigs into villages without any breeding plan. A survey of Thai native pigs in 1994 found that they have been endangered or nearly extinct (Rattanaronchart, 1994).

2.2 Molecular evolution and phylogenetic analysis

Historically, taxonomists have categorized different species primarily on the basis of the phenotype. Initially, morphological differences were used to construct evolutionary trees in which species that are more similar in appearance tended to be placed closer together in the tree. In addition, species have been categorized based on differences in physiology, biochemistry, and even behavior. However, because evolution involves genetic changes it makes sense to categorize species based on the properties of their genetic material. In term of a phylogenetic tree reconstruction, those species that are evolutionary closely related to each others are expected to have greater similarities in their genetic material that are distantly related species. Molecular evolution is the process of evolution at the scale of genetic material and protein. This field emerged in the 1960's as researchers from molecular biology, evolutionary biology and population genetics built upon ground-breaking discoveries in the studies of sequences of nucleic acids, the structures of proteins and their functions. In past few decades, molecular genetics has greatly facilitated our

understanding of speciation and evolution (Page and Holmes, 1998). Differences in nucleotide sequence are regarded as quantitative and can be analyzed using mathematical principles in conjunction with computer programs. Evolutionary changes at the DNA level can be objectively compared among different species to establish evolutionary relationships. Furthermore, this approach can be used to compare any two existing organisms, no matter how greatly they may differ in their morphological traits (Brooker, 2005). For example, we can compare humans and pigs, and bacteria, on the DNA level wheas such comparisons would be very difficult at a morphological level.

2.2.1 Genome and sequence evolution

Phenotypes of living organisms are always a result of the genetic information that they carry and pass on to the next generation and of the interaction with the environment. The genome, carrier of this genetic information, is in most organisms deoxyribonucleic acid (DNA), whereas some viruses have a ribonucleic acid (RNA) genome. Molecular sequences are coded as strings of literals from an alphabet representing the order in which the building blocks are connected in the molecules. Nucleotide sequence data (DNA and RNA) are coded with an alphabet of the four nucleotides A, G, C, and T in DNA or U in RNA. (Li´ebecq, 1992). Nowadays nucleotide sequences can be efficiently obtained using fast (automated) sequencing techniques (Sanger *et al.*, 1977).

Part of the genetic information in DNA is transcribed into RNA, either mRNA, which acts as a template for protein synthesis rRNA, which together with ribosomal proteins constitutes the protein translation machinery or tRNA, which offers the encoded amino acid. The genomic DNA also contains elements, such as promotors and enhancers that orchestrate the proper transcription into RNA. A large part of the genomic DNA of eukaryotes consists of genetic elements, such as introns, alu-repeats, the function of which is still not entirely clear. Proteins, RNA, and to some extent DNA, through their interaction with the environment, constitute the phenotype of an organism (Salemi and Vandamme, 2004). Before the division of any cell, plastid, or mitochondrion its genome has to be replicated to be inherited to the daughter cells or organelles. In spite of a proof reading machinery the process of copying is not error-free

(Turner *et al.*, 2000). Additionally, damages are introduced to the DNA by mutagens such as certain chemicals or UV light. Hence, the genomic sequence accumulates mutations as traces of evolutionary development.

Although more complex mutations like rearrangements, duplications, and inversions are possible on the chromosome level, only point mutations affecting single spots of the DNA are commonly considered (Page and Holmes, 1998). Point mutations are substitutions (the exchange of one character state by another), insertions (inserting one or more characters) and deletions (deleting one or more characters) (Figure 2.2). Insertions and deletions cannot be distinguished when examining two contemporary sequences (e.g., s_2 and s'_2 in Figure 2.2). To decide whether an insertion or deletion has happened, one needs information on the sequence of the common ancestor (s_1 in Figure 2.2) which, however, is usually unknown. For that reason insertions and deletion are generalized as indels. Indels and substitutions generally apply to nucleotide as well as protein data. Although mutations happen at the genomic level, they can affect the amino acid sequence translated from protein coding genes. Nucleotide substitutions in protein coding genes that do not result in an amino acid substitution are called synonymous or silent mutations in contrast to non-synonymous mutations (Page and Holmes, 1998). Incorporation errors replacing a purine with a purine and a pyrimidine with a pyrimidine are for steric reasons more easily made. The resulting mutations are called transitions. Transversions, purine to pyrimidine changes and the reverse, are less likely. When resulting in an amino-acid change, transversions often have a larger impact on the protein than transitions. There are four possible transition errors (A↔G, C↔T) and eight possible transversion errors (A↔C, A↔T, G↔C, G↔T) therefore, if a mutation would occur randomly, a transversion would be two times more likely than a transition. However, in many genes, transitions are twice as more likely to occur than transversions, which is used as default substitution parameter in substitution models that can score transitions and transversions differently (Salemi and Vandamme, 2004).

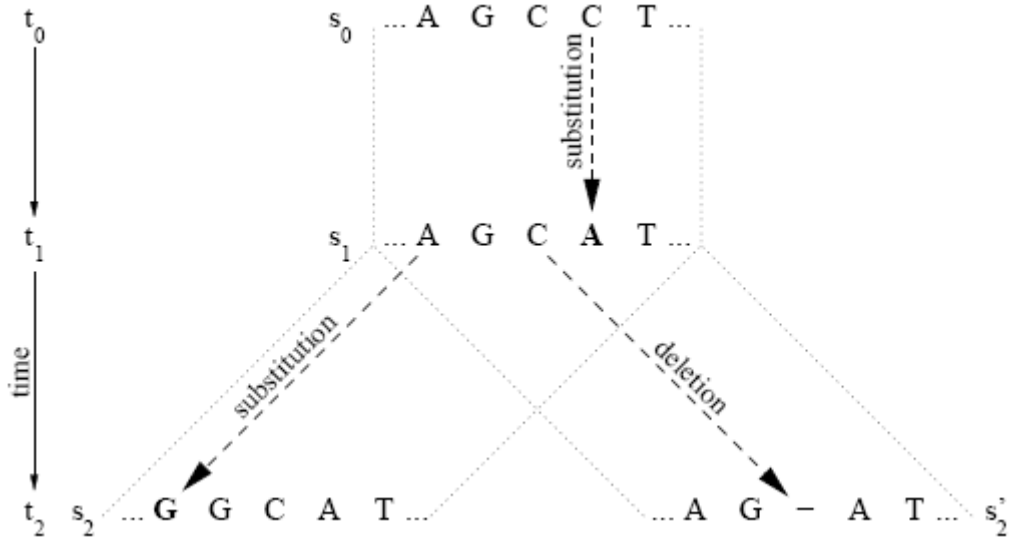

Figure 2.2 An example for sequence evolution. Sequence s_0 evolves to s_1 in time t_1-t_0. Then s_1 evolves divergently to s_2 and s'_2 after a speciation/ duplication event in time $t_2 - t_1$ (Page and Holmes, 1998).

2.2.2 Multiple sequence alignment

Homologous sequences, that means, sequences related by a common ancestor sequence, are typically presented in a multiple sequence alignment (MSA). An alignment is a data matrix in which homologous characters, also called sites, of the sequences are aligned in the same column (Page and Holmes, 1998). To present indels, an additional gap character '-' is used. Obtaining the correct alignment is easy for closely related species and can even be done manually using a word processor. The more distantly related the sequences are, the trickier it is to find the best alignment (Salemi and Vandamme, 2004). Sequence alignments serve as input to almost every sequence analysis and programs like ClustalW (Thompson *et al.*, 1994) are available to compute an MSA for a collection of sequences . However, there is still no standardized approach how to treat indels in the subsequent analysis. Typically, columns with gaps are either discarded from the alignment, or indels are considered as wildcard characters substituted by a distribution of character states.

2.2.3 Methods to infer phylogenetics

Phylogenetics is the study of relationships among various groups of organisms (e.g., species, population) that descend from a common ancestor. To study phylogenetic relationships, very diverse biological data are used like morphological characters, binary data, genomic gene order, nucleotide, and protein sequence data. The first data used were morphological characters of the species under interest. Morphological characters have the advantage to be easily obtainable by eye or by microscopy without molecular laboratory work. In the early times of molecular phylogenetics, coarse grain genetic data like binary characters of presence and absence of restriction sites played an important role. Recently, gene order data has been used to reconstruct phylogenetic relationships from the order of genome rearrangements, so-called breakpoint phylogenies (Blanchette *et al.*, 1997).

Today the vast majority of data in phylogenetic reconstruction is biological sequence data like nucleotide or protein sequences as stored in the public databases. One major advantage of biological sequence data is the increase of phylogenetically relevant information due to the large number of characters, that is, sites or residues that can be obtained by sequencing. The methods for construction of phylogenetic tree from molecular data can be classified into two types depending on the type of data used. First according to whether the method uses discrete character states or a distance matrix of pairwise dissimilarities, and second according to whether the method clusters operational taxonomic units (OTUs) stepwise, resulting in only one best tree, or considers all theoretically possible trees. Table 2.1 lists the currently most used phylogenetic tree construction and tree analysis methods, classified according to the strategy used: character state or distance matrix, exhaustive search or stepwise clustering.

All methods use particular evolutionary assumptions, which do not necessarily apply to the data set. Therefore, it is important to realize which assumptions were made when evaluating the best tree given by each method. Regarding types of data, unweighted pairgroup method using arithmetic averages, Neighbor joining, Fitch-Margoliash, KITCH are distance matrix methods whereas maximum parsimony and maximum likelihood are character-state methods. If, however, classification of phylogenetic reconstruction is based on strategy of

the method to find the best trees. Generally, DNA sequences need to be aligned for identification of the homologous positions among different sequences. The computer programmes such as PHYlIP (Felsenstein, 1995), can be used for construction of the phylogenetic tree. The phylogenetic approach is then chosen on the basis of data as illustrated by Figure 2.3.

Table 2.1 Most used phylogenetic analysis methods and their strategies.

	Exhaustive search	Stepwise clustering
Character State	Maximum parsimony (MP) Maximum likelihood (ML)	
Distance Matrix	Fitch-Margoliash	UPGMA Neighbor-joining (NJ)

Source: Adapted from Salemi *et al.* (2004)

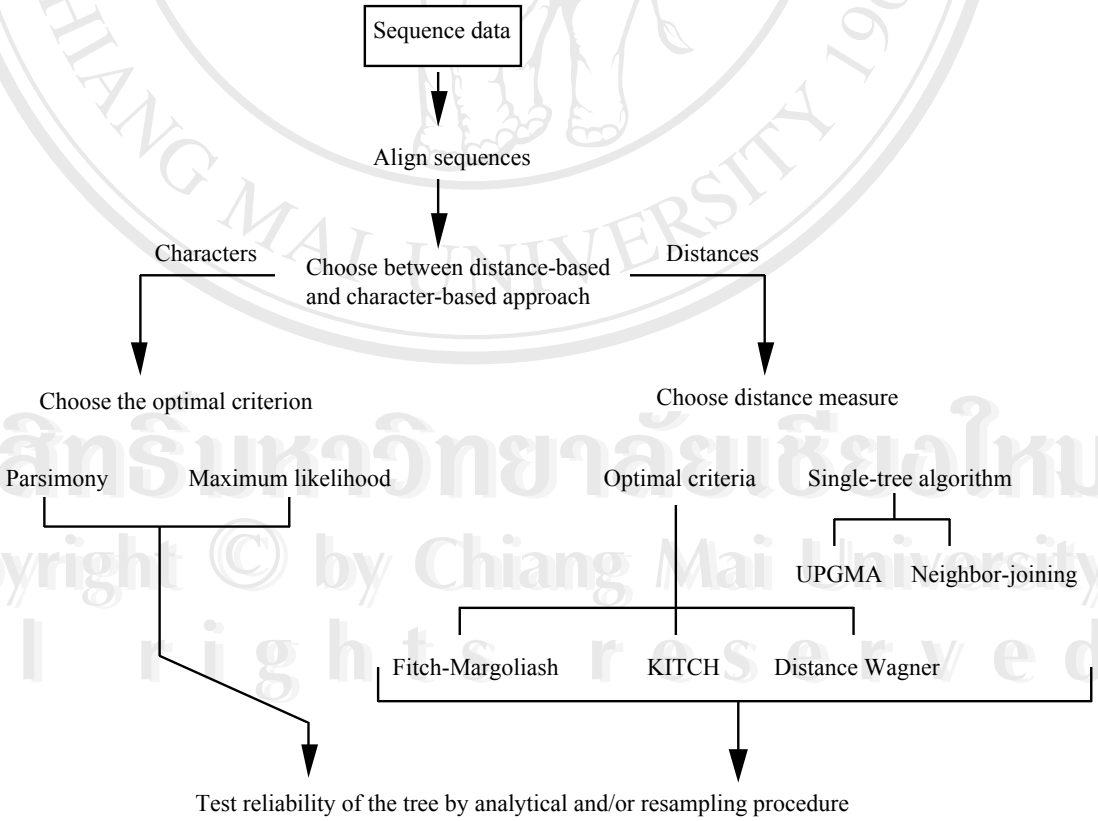

Figure 2.3 Steps to analyse DNA sequencing data to assess the phylogenetic result.

2.3 Genetic relationships of pigs

2.3.1 Taxonomy of the pig

Estimates based on paleontological evidence put the origin of the pig in the Miocene epoch about 40 million years ago. During this period, pigs browsed in the forest of Mongolia and were present in Europe, Asia, New Guinea, the East Indies and Africa. The wide geographic distribution of the pig at these early stages of its evolution was associated with a high degree of variation in its food supply, physical and climatic descendants of the European wild pig (*Sus scrof*a) and the Eastern Asiatic banded pig (*Sus vittatus*) a descendant of *Sus crystatus*. The taxonomy classification of the pig is shown in Table 2.2.

Domestic pigs and wild boars (Europe, Asia) have been classified in the order Artiodactyla (even-toed ungulates), which also embraces ruminant animals, including cattle, deer, goat and sheep. Pigs are non-ruminant and form a separate suborder-Suiformes-which includes hippopotamuses and peccaries. Within Suiformes, pigs are in the family Suidae, composed of five genera (*Sus*, *Babirussa*, *Hylochroerus*, *Phacochoerus* and *Potomochoerus*). Genus *Sus* contains seven species, including *scrofa* (European and Asian domestic pigs, wild boar) and *crytatus* (Indian crested pig or Asiatic wild pig) from which domesticated swine are believed to have originated. The banded pig (*Sus vittatus*) a descendant of *Sus crytatus*, is numerous among modern wild swine in Malaysia (Pond and Mersmann, 2001).

2.3.2 Diversity of Asian and European pig breeds

The European domestic pig (*Sus scrofa*) and the Asian domestic pig (*Sus indicus*) are recognized as the two major forms of modern domestic pig (Figure 2.4). It is well documented that Asian pigs were used to improve European pig breeds during the eighteenth and early nineteenth century, but to what extent Asian pigs have genetically contributed to different European pig breeds is unknown (Giuffra *et al.* 2000).

Table 2.2 Taxonomy of the pig.

Phylum	Chordata
Class	Mammalia
Order	Artiodactyla (even-toed ungulates)
Suborder	Suiformes—Ancodonta (hippopotamuses, pigs, peccaries)
Infraorder	Suina—Tayassuidae (peccesries)
Family	Suidae — hogs, pigs
Genus	*Sus*
	S. batatus—beared pig (Malaya, Sumatra, Borneo)
	S. celebenis—Sulawesi warty pig (Sulawesi)
	S. crytatus—Indian crested pig or Asiatic wild pig (India)
	S. salvanius—pygmy hog (Southeast Nepal, Assam)
	***S. scrofa*—domestic pigs, wild boar (Europe, Asia)**
	S. verucosus—Javanese warty hog (Java, Sulawesi, Philippines)
	S. vittatus—banded pig (Malay Archipelago)
Genus	*Babirussa* *B. babyrussa*—babirusa (East Indonesia)
Genus	*Hylochoerus* *H. meinertzhageni*—giant forest hog (Central America)
Genus	*Phacochoerus* *P. aethiopicus*—warthog (Sub-Saharan Africa)
Genus	*Potomochoerus* *P. porcus*—bushpig or Red River hog (sub-Saharan Africa, Madagascar)

Source: Adapted from Pond and Mersmann (2001)

In the domestic animal diversity information system (DAD-IS) of the Food and Agriculture Organization (FAO), Europe shares a large proportion of the world pig population (about 30%) as well as the genetic diversity (about 37% of the breeds in the world inventory). The origin, genetic variation, evaluation and conservation of domestic pigs have attracted an increased worldwide attention as their genetic diversity rapidly erodes. The necessity to maintain diversity and to develop alternative stocks to meet a wide variety of production/market conditions is recognized, as well as the ensuing need to establish sound conservation programs (Laval *et al.*, 2000). Biodiversity is large concept that all countries now emphasize on. The studies of their genetic information, the development of their economic traits and the preservation of native breeds in each part of the world are very important for the development of genetic resources (Frankham *et al.*, 2002).

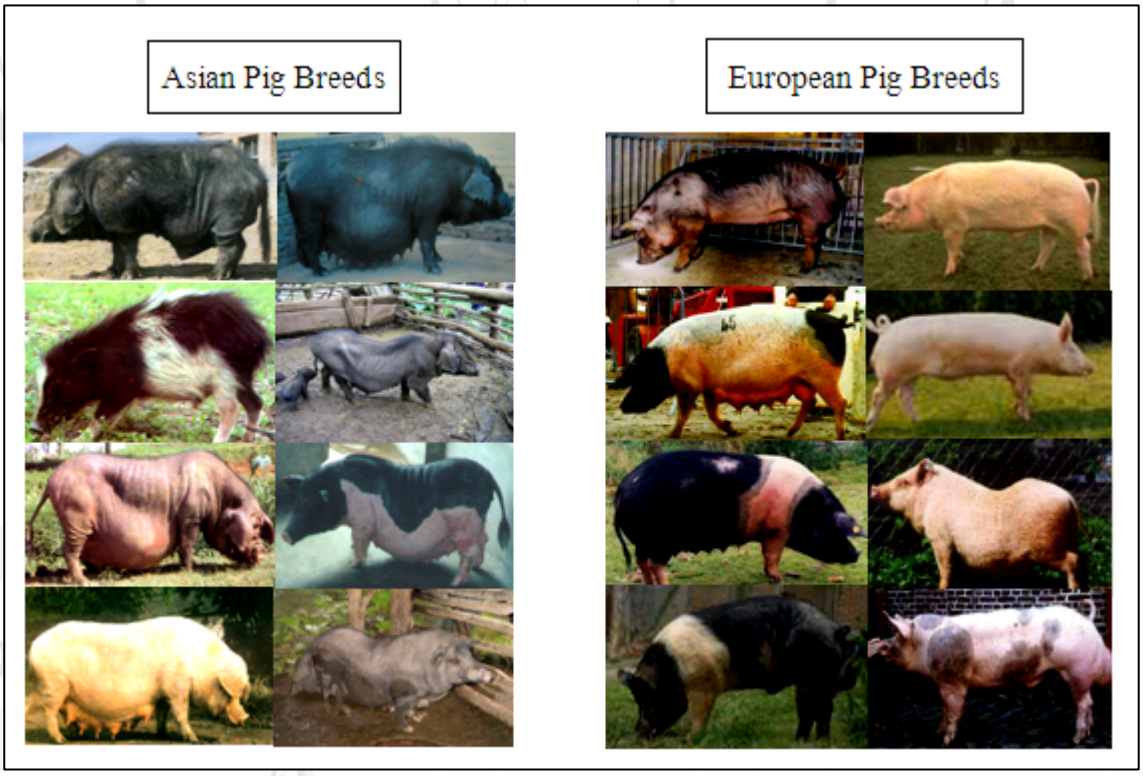

Figure 2.4 Diversity of Asian and European pig breeds.

At present, studies of genetic distribution and diversity of Thai native pigs provide no clear information. Using phenotypes or allozyme electrophoresis can not analyze the genetic difference at the same level as with molecular genetic method. (Chen, 2003). Recently, molecular techniques have been widely used to analyze phylogeny in pigs. The most popular techniques were single strand conformation polymorphism (SSCP; Takeda *et al.*, 1995), microsatellites, AFLP and mitochondrial DNA sequence polymorphisms analysis. Especially DNA sequence analysis is mainly because of the evolutionary information that can be drawn from sequence data. By comparing DNA sequences, one can derive evolutionary relationships, levels of variability and geographical sub structuring within and between groups of animals.

2.3.3 mtDNA sequence analyses

Mitochondrial DNA (mtDNA) has been widely used for phylogenetic studies for several reasons (Kierstein *et al.*, 2004). First, evolution of mammalian mtDNA occurs primarily as single base pair substitutions, with only infrequent major sequence rearrangements. Secondly, the rate of mtDNA evolution appears to be as much as 10 times faster than that of nuclear DNA (Brown *et al.*, 1979). Thirdly, mtDNA is maternally inherited, haploid and non-recombining (Avise, 1994). These studies facilitate the use of mtDNA as a tool for determining relationships among individuals within species and among closely related species with recent times of divergence. Randi *et al.* (1996) used *cytB* polymorphism for evolutionary analysis of the *suiformes* and also to determine relationships among some *Sus scrofa* populations. However, in mtDNA sequence the Displacement loop (D-loop) region or so called the control region is known to be more variable in sequence than other regions and thus has been frequently used for phylogenetic analyses of closely related groups (Kim *et al.*, 2002a). A few phylogenetic studies on pigs have been carried out using mtDNA D-loop sequence variations (e.g. Okumura *et al.*, 1996; Giuffra *et al.*, 2000) as well as restriction fragment length polymorphism (RFLP) of mtDNA (e.g. Lan and Shi, 1993; Huang *et al.*, 1999) The function and location of porcine mtDNA are displayed in Figure 2.5 and Table 2.3.

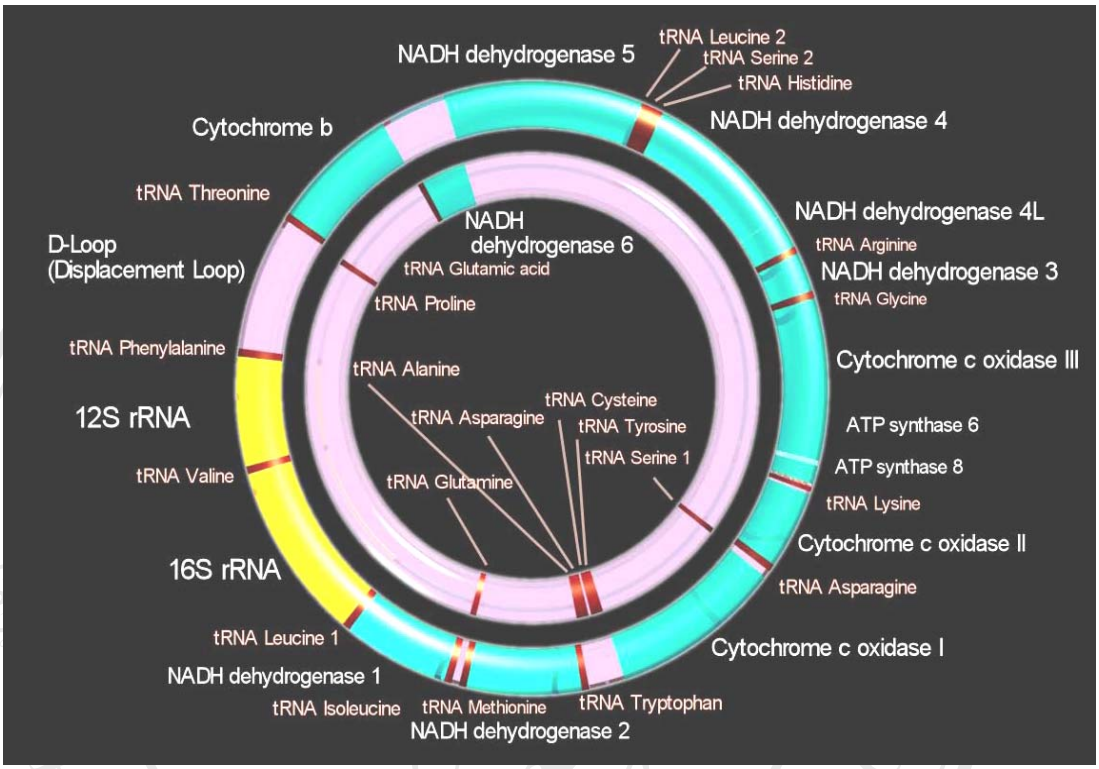

Figure 2.5 The mtDNA genome and function locations (Brooker, 2005)

Giuffra *et al.* (2000) provided comprehensive molecular analyses regarding the genetic relationships between domestic pigs and wild boars. This analysis included the mitochondrial DNA (mtDNA) *cytB* gene, the D-loop region of mtDNA, and three nuclear genes (melanocortin receptor 1 [*MC1R*], tyrosinase [*TYR*], and the glucose phosphate isomerase pseudogene [*GPIP*]. These authors presented clear evidence of the independent domestication events of European and Asian subspecies of wild boar, and their conclusion regarding these domestication events is essentially the same as that of Watanobe *et al.* (1999).

Recently, the complete mtDNA sequence of the pig was published along with its phylogenetic relationships to other animal species (Ursing and Arnason, 1998). However, no comprehensive studies have been performed so far on phylogenetic relationships of Thai native pigs using mtDNA sequence polymorphism.

Table 2.3 The location of features in the mitochondrial DNA of the pig, *Sus scrofa*[a] (Ursing and Arnason, 1998)

	Position			Codon	
Feature	From	To		Start	Stop
tRNA-Phe	1	70			
12S rRNA	71	1,032			
tRNA-Val	1,032	1,099			
16S rRNA	1,100	2,668			
tRNA-Leu (UUR)	2,668	2,742			
NADH1	2,745	3,701		ATG	TAG
tRNA-Ile	3,700	3,768			
tRNA-Gln	3,766	3,838	(L)		
tRNA-Met	3,840	3,909			
NADH2	3,910	4,953		ATT	TAG
tRNA-Trp	4,952	5,019			
tRNA-Ala	5,093	5,026	(L)		
tRNA-Asn	5,169	5,095	(L)		
Or. L-stand repl.	5,166	5,212			
tRNA-Cys	5,267	5,202	(L)		
tRNA-Tyr	5,332	5,267	(L)		
COI	5,334	6,878		ATG	TAA
tRNA-Ser (UCN)	6,950	6,882	(L)		
tRNA-Asp	6,958	7,025			
COII	7,026	7,713		ATG	T--
tRNA-Lys	7,714	7,780			
ATPase8	7,782	7,985		ATG	TAA
ATPase6	7,943	8,623		ATG	TAA
COIII	8,623	9,407		ATG	TA-
tRNA-Gly	9,407	9,475			
NADH3	9,476	9,822		ATA	TA-
tRNA-Arg	9,823	9,891			
NADH4L	9,892	10,188		GTG	TAA
NADH4	10,182	11,559		ATG	T--
tRNA-His	11,560	11,628			
tRNA-Ser (AGY)	11,629	11,687			
tRNA-Leu (CUN)	11,688	11,757			
NADH5	11,758	13,577		ATA	TAA
NADH6	14,088	13,559	(L)	ATG	TAA
tRNA-Glu	14,157	14,089	(L)		
Cyt *b*	14,162	15,301		ATG	AGA
tRNA-Thr	15,302	15,369			
tRNA-Pro	15,433	15,369	(L)		
Control region	15,434	16,679			
CSB-1	16,109	16,134			
CSB-2	16,378	16,396			
CSB-3	16,430	16,448			

[a] (L), light-strand sense; NADH1–6 and NADH4L, subunits 1–6 and 4L of nicotinamid dinucleotide dehydrogenase; ATPase6 and 8, subunits 6 and 8 of adenosine triphosphatase; COI–COIII, cytochrome *c* oxidase subunits I–III; cyt *b*, cytochrome *b*.