

# **Molecular Strategies in the Analysis of the Porcine Genome**

Dissertation  
zur Erlangung des Doktorgrades  
der Fakultät für Agrarwissenschaften  
der Georg-August-Universität Göttingen

vorgelegt von  
**Kefei Chen**  
geboren in Jiangxi/P. R. China

Göttingen, Dezember 2003

D7

1. Referent: Prof. Dr. Dr. Bertram Brenig

2. Korreferent: Prof. Dr. Henner Simianer

Tag der mündlichen Prüfung: 05 February 2004

*To my family*

---

|   |    |
|---|----|
| <b>TABLE OF CONTENTS</b>  | 1  |
| <b>1 LIST OF PUBLICATIONS</b>   | 2  |
| <b>2 ABSTRACT</b>   | 3  |
| <b>3 INTRODUCTION</b>   | 4  |
| <b>3.1 Tools for genome analysis</b>  | 4  |
| 3.1.1 Molecular marker systems  | 4  |
| 3.1.2 Isolation of microsatellites  | 6  |
| 3.1.2.1 'Traditional' strategy of microsatellite isolation                          | 7  |
| 3.1.2.2 'Enrichment' strategy of microsatellite isolation                           | 8  |
| 3.1.2.2.1 Randomly amplified polymorphic DNA (RAPD)-based                           | 8  |
| 3.1.2.2.2 Primer extension  | 8  |
| 3.1.2.2.3 Selective hybridization   | 9  |
| 3.1.2.3 Direct clone-sequencing strategy of microsatellite isolation                | 10 |
| 3.1.3 Characterization of genes   | 11 |
| 3.1.3.1 Large-insert libraries in pig genome mapping                                | 11 |
| 3.1.3.2 Strategy of positional candidate gene approach                              | 13 |
| <b>3.2 Application of DNA markers in pig genetics</b>                               | 14 |
| 3.2.1 Phylogenetics   | 15 |
| 3.2.2 Quantitative trait locus (QTL) analysis and disease survey                    | 17 |
| 3.2.2.1 Strategies to identify QTLs   | 17 |
| 3.2.2.2 Comparative mapping and functional analysis                                 | 19 |
| 3.2.2.3 Monogenic disease   | 20 |
| 3.2.2.4 Disease resistance and immune response traits                               | 21 |
| <b>3.3 The phosphoglycerate kinase (PGK) gene family</b>                            | 23 |
| 3.3.1 Members of the <i>PGK</i> gene family   | 23 |
| 3.3.1.1 Functional loci and pseudogenes of the <i>PGK</i> gene family               | 23 |
| 3.3.1.2 Molecular evolution of the <i>PGK</i> gene family by retrotransposon events | 24 |
| 3.3.2 Biological/physiological function of the PGKs                                 | 27 |
| 3.3.2.1 Expression and encoded proteins of <i>PGK</i> genes                         | 27 |
| 3.3.2.2 <i>PGK</i> transcript switch and <i>PGK</i> methylation                     | 28 |
| 3.3.3 Direct effects and associations of the PGKs                                   | 29 |
| 3.3.3.1 PGK and diseases  | 29 |
| 3.3.3.2 PGK and male reproduction   | 29 |
| <b>4 REFERENCES</b>   | 31 |
| <b>5 ACKNOWLEDGEMENTS</b>   | 40 |
| <b>6 CURRICULUM VITAE</b>   | 41 |

## 1 LIST OF PUBLICATIONS

This thesis is based on the following papers:

### --- Peer-Reviewed Journals

1. Chen KF, Beck J, Huang LS, Knorr C, Brenig B (2004) Assignment of the phosphoglycerate kinase 2 (*PGK2*) gene to porcine chromosome 7q14-q15 by fluorescence *in situ* hybridization and by analysis of somatic cell and radiation hybrid panels. *Anim Genet*, 35: 71-72.
2. Chen KF, Beck J, Huang LS, Knorr C, Brenig B (2004) Assignment of the phosphoglycerate kinase 1 (*PGK1*) gene to porcine chromosome Xq12-q13 by fluorescence *in situ* hybridization and confirmation by hybrid panel analyses. *Anim Genet*, doi: 10.1111/j.1365-2052.2004.01092.x.
3. Chen KF, Knorr C, Bornemann-Kolatzki K, Huang LS, Brenig B (2004) Isolation and characterization of the highly polymorphic porcine STS-marker *S0719* adjacent to the 5' region of the phosphoglycerate kinase 2 (*PGK2*) gene. *Revised version in preparation*.
4. Chen KF, Knorr C, Moser G, Gatphayak K, Brenig B (2004) Molecular characterization of the porcine testis-specific phosphoglycerate kinase 2 (*PGK2*) gene and its association with male fertility. *Revised version in preparation*.
5. Chen KF, Knorr C, Bornemann-Kolatzki K, Huang LS, Brenig B (2004) Targeted isolation of microsatellites and their application in phylogenetic analyses of European and Chinese pig breeds. *Revised version in preparation*.
6. Gatphayak K, Knorr C, Chen KF, Brenig B (2004) Structural and functional analysis of the porcine *FUS2* gene. *Gene*, accepted for publication.

### --- Congress Contribution

1. Chen KF, Knorr C, Bornemann-Kolatzki K, Brenig B (2003) Microsatellite isolation and phylogenetic analyses of European and Chinese pig breeds. *Proceedings of the DGfZ/GfT convention*, September 17/18 2003, Goettingen, Germany, B10.

## 2 ABSTRACT

The main goal of genome research in livestock is to map and characterize trait loci controlling various phenotypic characters like disease resistance, reproduction and other production traits of economic importance (e.g. meat quality). Although identification and mapping of functional genes and quantitative trait locus (QTL) regions associated with traits of interest increased significantly in pigs, the number of genes mapped is still small compared with human and mouse. Comparative analysis and bioinformatic tools combined with genomic and post-genomic techniques enable to achieve greater progress in the analysis of the porcine genome. The present thesis focuses on structural and functional analysis of genes of the porcine phosphoglycerate kinase (*PGK*) family which have previously been identified as candidate genes for inherited male fertility along with other inherited disorders in human and mouse. The second part of the thesis describes the isolation of microsatellite markers either from specific chromosomal region of interest, or from large-insert library clones harboring targeted trait genes, and the application of these informative markers in phylogenetic studies.

The porcine functional *PGK* genes, i.e. the ubiquitously expressed X-linked *PGK1* gene and the autosomally determined testis-specific *PGK2* gene, were identified and mapped on *SSCXq12-q13* and *SSC7q14-q15*, respectively. The 1665 bp of full-length *PGK2* cDNA were compiled using the modified rapid amplification 5'-RACE and 3'-RACE information. The results of genomic and cDNA sequences of the porcine *PGK2* gene demonstrated that it is a single-exon intronless gene with a complete open reading frame of 1251 bp encoding a PGK protein of 417 amino acids. Real time PCR results showed that *PGK2* mRNA was solely expressed in the testis. There was a lower amount of *PGK2* expression in the testis from a herniated boar and a very small amount of *PGK2* expression in the testis of an 8-week-old piglet compared to an adult boar. Two SNPs of *PGK2* gene (SNP-A: T427C; SNP-B: C914A) resulting in amino acid substitutions (SNP-A: Ser<sup>102</sup>-Pro<sup>102</sup>; SNP-B: Thr<sup>264</sup>-Lys<sup>264</sup>) were detected and genotyped among six pig breeds. The SNP-A mutation of proline could lead to the loss of a casein kinase II (CK2) phosphorylation site in the PGK2 peptide. Association analyses between the *PGK2* gene and some traits of sperm quantity and quality were performed. The results showed that SNP-B has a significant effect on semen volume in the Pietrian breed ( $p = 0.08$ ).

Fifteen (GT)<sub>n</sub> porcine microsatellites (*S0701* – *S0715*) were isolated from nine PAC clones by an improved efficient and rapid flanking sequence-based approach. Eight of the newly developed microsatellites were gene-associated, and the other seven microsatellites within the poorly covered chromosome region *SSC16q* were isolated to refine this region. The microsatellite markers were genotyped for a panel of 245 unrelated animals representing 15 European and Chinese indigenous pig breeds. A set of eleven microsatellite loci with a total of 136 alleles was used to analyze the phylogenetic relationships between the pig populations. Breed differentiation was significant as shown by the high average fixation index ( $F_{ST} = 0.271$ ). Neighbor-joining (NJ) tree and unweighted pair group method with arithmetic mean (UPGMA) dendrogram were constructed. The results revealed that European and Chinese indigenous pig breeds had diverged into two distinct groups, apart from the German Göttingen Minipig and the European wild boar, which clustered into the branches of Chinese pig populations.

### 3 INTRODUCTION

The pig was one of the first animals domesticated over 7,000 years ago (Rothschild and Ruvinsky, 1998). The adaptation to different environmental conditions and phenotypic selection by human beings has created a wide diversity of pig breeds. As the 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 (Ollivier *et al.*, 2001). Nowadays, the pork is the major red meat consumed (43%) worldwide and the pig can also be considered as an important model system for human health. The pig industry has made remarkable progress in the last 40 years: each sow produces 50% more progeny per year, which eat one third less feed in order to produce 33% more lean meat (Plastow, 2003). Genetics and breeding has played an increasingly significant role in the pig industry, which is accomplished by using combined approaches of molecular genetics and quantitative genetics enabling the pig industry to reduce the incidence of diseases, and to respond more rapidly to the changing demands of consumers.

In the early 1990s, several international gene mapping projects in livestock were launched including the PiGMaP gene mapping project initiated in Europe (Archibald *et al.*, 1995), two efforts (USDA-ARS; USDA-CSREES) initiated by the USDA in the USA, and the AGP animal genome research program by the NIAS-STAFF in Japan to unravel the pig genome. The main goal of genome research in livestock is to map and characterize trait loci that control various phenotypic characters (Andersson, 2001). To date, great progress has been achieved as more functional genes and QTL regions associated with traits of interest continue to be identified and mapped in pigs, however, the number of genes mapped is still small compared with the situation in humans and mice.

#### **3.1 Tools for genome analysis**

##### **3.1.1 Molecular marker systems**

To evaluate qualitative and quantitative information on diversity molecular genetic markers are essential aspects of both fundamental and applied fields in biology: ecology, evolutionary biology, taxonomy, agronomy, breeding, conservation and forensic identification are mentioned as examples (Karp and Edwards, 1997). There are a variety of genetic markers applied in genome analysis. In the following focus is laid on six DNA marker systems including restriction fragment

length polymorphism (RFLP), single nucleotide polymorphism (SNP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), minisatellite and microsatellite.

Genetic markers can be generally grouped into two types based on their association with functionality (O'Brien, 1991): Type I markers are these DNA segments encoding for expressed DNA sequences which possess a relatively low degree of polymorphism but high evolutionary conservation, whereas type II markers usually have no identifiable biological function but they are highly polymorphic and not well conserved between species. RFLP and SNP are the most common type I markers in genetic mapping. Type II markers mainly consist of RAPD, AFLP, minisatellite and microsatellite. Alternatively, Dodgson *et al.* (1997) distinguished DNA-based markers into the groups of clone/sequence-based (CSB) markers and fingerprint (FP) markers (Dodgson *et al.*, 1997). CSB markers require the isolation of a cloned DNA fragment and often determination of some, if not all, of its DNA sequence. RFLP, microsatellite and SNP are included in this category. FP markers comprising RAPD, AFLP and minisatellite require no previous knowledge of isolation of a cloned DNA fragment or the sequence of the polymorphic region. Taking both hypotheses into account, FP markers are inherently type II markers while CSB markers include both type I and type II markers. A detailed comparison of the six main genetic markers is listed in Table 1.

**Table 1:** Summary of main genetic markers (modified based on Dodgson *et al.*, 1997; O'Brien 1991).

|                        | RFLP                 | Microsatellite       | SNP                  | RAPD        | Minisatellite   | AFLP        |
|------------------------|----------------------|----------------------|----------------------|-------------|-----------------|-------------|
| Type of loci (Dodgson) | Clone sequence-based | Clone sequence-based | Clone sequence-based | Fingerprint | Fingerprint     | Fingerprint |
| Type of loci (O'Brien) | I and II             | II > I               | I and II             | II          | II              | I and II    |
| Genome distribution    | Ubiquitous           | Ubiquitous           | Ubiquitous           | Ubiquitous  | Heterochromatin | Ubiquitous  |
| PIC*                   | Low                  | High                 | Low                  | Moderate    | High            | Moderate    |
| Typical allele no.     | 2                    | 2 - 30               | 2                    | 2           | 2               | 2           |
| Inheritance mode       | Co-dominant          | Co-dominant          | Co-dominant          | Dominant    | Dominant        | Dominant    |
| Reliability            | High                 | High                 | High                 | Low         | High            | Moderate    |
| Speed of assay         | Low                  | High                 | High                 | High        | Low             | High        |
| Initial investment     | Moderate             | High                 | High                 | Low         | Moderate        | Low         |
| Automation             | Difficult            | Yes                  | Yes                  | Yes         | Yes             | Yes         |
| Development costs      | Medium               | High                 | Medium               | Low         | High            | Low         |
| Running costs          | Medium               | High                 | Medium               | Medium      | High            | Medium      |

\*PIC = polymorphism information content

### 3.1.2 Isolation of microsatellites

Microsatellites also called simple sequence repeats (SSR) are tandemly repeat motifs of 1-6 bases that are known to be ubiquitous in all eukaryotic genomes analysed to date. They are present with a high probability in non-coding regions (Litt and Luty, 1989; Tautz, 1989; Weber and May, 1989). Specific PCR reactions are performed using genomic DNA as template and two specific primers flanking the repeats. The PCR products that differ in a few nucleotides among individuals are separated through high-resolution electrophoresis by radioactively or fluorescently labelled primers. The most common dinucleotide motif in mammals is (CA)<sub>n</sub>. In pigs the number was estimated between 65,000 and 100,000 copies, highly distributed throughout the genome (Winterø *et al.*, 1992; Rohrer *et al.*, 1994). The mutation rate of microsatellites is thought to be high - a rate of up to 10<sup>-2</sup> was observed per generation (Bruford and Wayne, 1993) -and there are often large numbers of alleles that vary in size at a single locus (Beuzen *et al.*, 2000). Microsatellites have the advantages of being abundant, multiallelic, highly polymorphic, co-dominant and assayable by PCR (Dodgson *et al.*, 1997). Therefore, microsatellites have been extensively applied in animal genome mapping, biodiversity and phylogenetic studies. The main disadvantages involve the high cost of generation of microsatellites as microsatellites need to be developed *de novo* for each new species in most cases. In addition, detection of microsatellites sometimes suffers from an inherent weakness in accurately determining the fragment size (Beuzen *et al.*, 2000).

Microsatellite markers became soon after the first description the workhorse to establish genome maps in swine (Archibald *et al.*, 1995; Rohrer *et al.*, 1996) and were subsequently used for the identification of chromosomal regions associated with (quantitative) traits (Andersson *et al.*, 1994; Cassady *et al.*, 2001). To date, more than 1,000 porcine microsatellites have been isolated in cooperative international efforts (Rothschild and Plastow, 1999). Although the current swine linkage maps contain an overall good coverage with an average interval of 2.23 cM (Rohrer *et al.*, 1996), there are still efforts going on to enhance SSR marker densities on so far poorly covered regions or chromosomes like *SSC18* (Campbell *et al.*, 2001). Also, as high-resolution genetic linkage maps are indispensable for positional cloning of economically important quantitative trait loci (QTLs) or major genes (Matisse *et al.*, 1994), microsatellites are also specifically developed in fine mapping approaches to resolve regions associated with important traits (Milan *et al.*, 2000).

Recently, microsatellites have been also widely used - besides the mtDNA D-loop region - to conduct phylogenetic studies in pigs (Paszek *et al.*, 1998; Laval *et al.*, 2000).

Microsatellites have been traditionally isolated from partial genomic libraries selected for small insert size by screening a large number of clones through colony hybridization with repeat-containing probes (Figure 1) (Rassmann *et al.*, 1991). Generally, this method is quite tedious and inefficient, especially for species with low microsatellite frequencies such as birds or plants. According to *Molecular Ecology* primer notes from 1999 to March 2001, the number of positive clones from traditional isolation protocols was in the range of 0.13% to 4.5% with the average of 1.67% in mammals (Zane *et al.*, 2002). Therefore, in order to reduce the time invested in microsatellite isolation and to significantly increase yield, several alternative strategies have been attempted and most of them can be summarized as enhanced protocols with microsatellite-enriched libraries. In fact, these isolation strategies have focused on the improvement of the efficiency of library construction and screening for microsatellite isolation, such as RAPD based, primer extension, selective hybridization and AFLP based selective hybridization methods. As shown in a survey of papers in *Molecular Ecology* primer notes (1999 - March 2001), one third of reports had used enhanced protocols for the isolation of microsatellite loci (Zane *et al.*, 2002). On the other hand, to isolate further microsatellites from known clones of interest and thus to increase microsatellite coverage on the genetic map in predefined regions is also lately receiving massive attention in genetic studies (Wu *et al.*, 1996; Hayden and Sharp, 2001; Fujishima-Kanaya *et al.*, 2003). In Zane *et al.*'s review paper (2002), several methods of microsatellite isolation were discussed and compared with each other. These methods often present slight differences from one another and frequently have not been extensively tested (Zane *et al.*, 2002).

### 3.1.2.1 'Traditional' strategy of microsatellite isolation

Genomic DNA is fragmented by restriction enzymes or sonication and the small DNA fragments in a range from 300 - 700 bp are usually size-selected prior to construct a partial genomic library using a common plasmid vector either directly or after ligation to specific adaptors (Rassmann *et al.*, 1991). Thousands of recombinant clones can be screened for the presence of microsatellite repeat sequences by Southern hybridization. Radioactive probes (e.g.  $^{32}\text{P}$ ,  $^{33}\text{P}$ ) are generally more sensitive, but the manipulation of radioactive materials and their short half life led to the limited use of radiolabelled probes. In contrast, non-radiolabelled probes (e.g. digoxigenin) have become

more popular since the efficiency has greatly improved. After identification of repeat containing clones, specific primers are designed based on the sequencing results and PCR conditions are optimized to amplify microsatellites among individuals.

### *3.1.2.2 'Enrichment' strategy of microsatellite isolation*

#### *3.1.2.2.1 Randomly amplified polymorphic DNA (RAPD)-based*

RAPD-based amplification of unknown microsatellites has been proposed using either repeat-anchored random primers or RAPD primers which produce large number of small size fragments by non-library PCR amplification (Wu *et al.*, 1994; Cifarelli *et al.*, 1995; Richardson *et al.*, 1995). The idea behind this approach is that RAPD fragments seem to contain more frequently microsatellite repeats than random genomic clones. PCR fragments can then undergo subsequent Southern hybridization with microsatellite probes and selective cloning of positive bands or through the cloning of all RAPD products followed by screening of arrayed clones (Ender *et al.*, 1996; Lunt *et al.*, 1999). However, the RAPD-based approach has not frequently been used to isolate microsatellites because no sequence information of the regions flanking the microsatellite can be obtained.

#### *3.1.2.2.2 Primer extension*

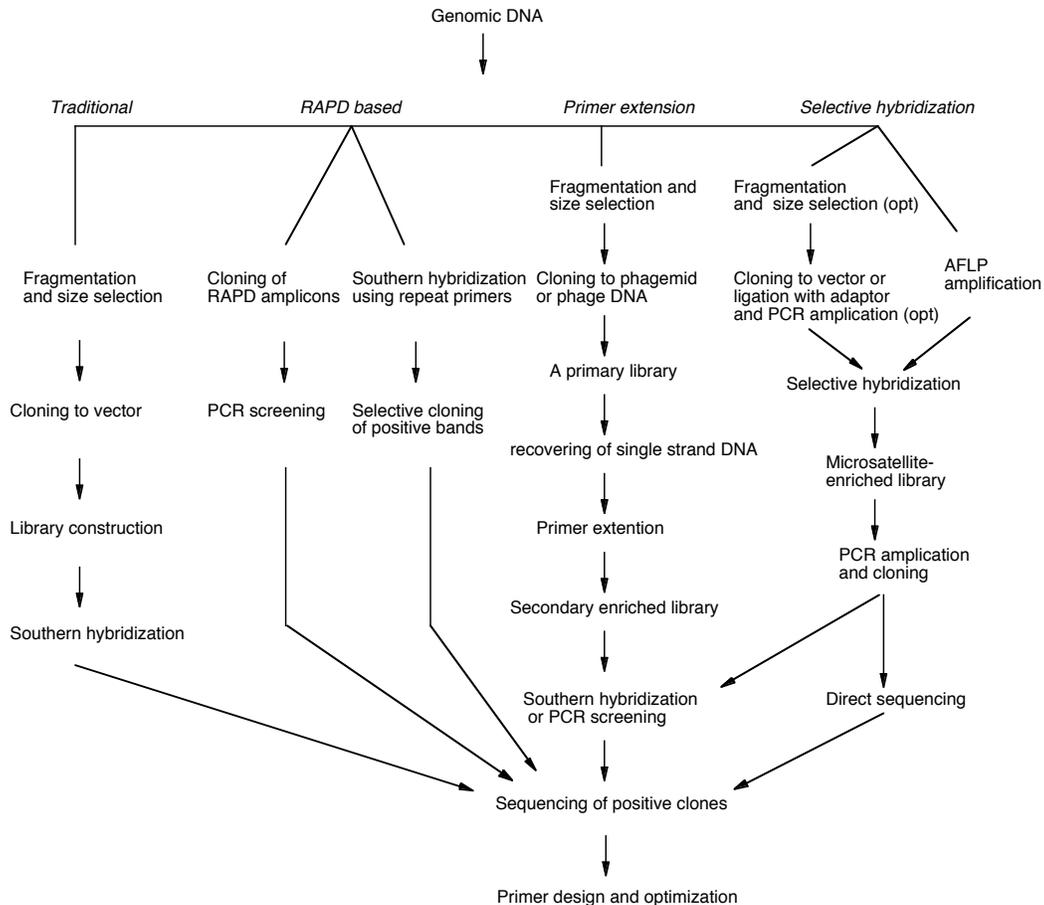
The primer extension strategy has been proposed for the production of libraries enriched with repeat motifs (Ostrander *et al.*, 1992; Paetkau, 1999). For this approach, the construction of a primary library is essential to obtain a pool of single-strand circular DNA for subsequent enrichment, in which DNA fragments are inserted into phagemids or phage vectors. Following this, a secondary enriched library is produced for specific microsatellite repeats using a primer extension reaction. The resulting library is highly enriched with repeat containing inserts and the following screening steps are similar to the traditional isolation strategy. The primer extension strategy has been proven to be useful for the isolation of dinucleotide repeat microsatellites. For example, the enrichment of (AC)<sub>n</sub> repeats could yield from 40% to 50% up to 100% of positive clones (Ostrander *et al.*, 1992; Paetkau, 1999), however it is unclear if this approach is also effective for the isolation of tri- and tetranucleotide repeats. Since it has been found that 0 - 25% positive clones could be isolated using a tetranucleotide repeat primer in the extension step, the usefulness of the method to gain tri- or tetranucleotide containing microsatellites needs more investigation (Paetkau, 1999). In addition, to produce a secondary enriched library only a limited

portion of the genome can be cloned, so the population of inserts undertakes a severe bottleneck that results in loss of rare repeat motifs (Zane *et al.*, 2002). In total, this protocol involves a rather large number of steps which might limit the application of the primer extension strategy for microsatellite isolation.

#### 3.1.2.2.3 Selective hybridization

The selective hybridization strategy is relatively simple and straightforward compared with other methods (Karagyozov *et al.*, 1993; Armour *et al.*, 1994; Kijas *et al.*, 1994). The first step of this strategy produces small genomic fragments which are ligated to vectors or adaptors. After selective hybridization with the repeat nucleotide probe and several washing steps to remove nonspecific binding, microsatellite-containing DNA can be eluted and enriched by PCR amplification. The use of a biotinylated probe in the liquid medium is generally preferable compared with the probe crosslinked to a nylon membrane in the process of the selective hybridization step. Finally, the recovered DNA fragments are cloned and screened for microsatellites by Southern blotting, PCR or direct sequencing. The overall enrichment efficiency of the selective hybridization strategy is in a range of 20% to 90% with respect to di-, tri- and tetranucleotide probes (Zane *et al.*, 2002). Therefore, this approach may be rather efficient and widely applicable.

In addition, Zane *et al.* (2002) suggested a microsatellite isolation strategy based on AFLP-based selective hybridization, which is similar to the selective hybridization strategy except that this method uses AFLP amplification rather than traditional library construction before. Genomic DNA is simultaneously digested with *MseI* and ligated to the *MseI*-AFLP-adaptor followed by PCR. Amplified DNA fragments are selectively hybridized with biotinylated oligonucleotide probes using streptavidin coated beads. The isolation steps are identical to the selective hybridization procedure. The percentage of clones containing dinucleotide repeats varies from 50% to 95% in a variety of species demonstrating a high microsatellite isolation efficiency (Zane *et al.*, 2002).



**Figure 1:** Schematic overview of microsatellite isolation strategies including traditional, RAPD-based, primer selection and selective hybridization methods (opt=optional step).

### 3.1.2.3 Direct clone-sequencing strategy of microsatellite isolation

Direct clone-sequencing strategies to isolate microsatellites have become more of interest because of the need to generate microsatellites from specific large-insert clones which harbor either functional candidate genes or, which can be used to enrich the already known microsatellite density on so far poorly covered chromosomes.

Recently, Fujishima-Kanaya *et al.* (2003) isolated some gene-associated dinucleotide microsatellites from BAC clones using a direct-sequencing method. Chen *et al.* (2004a) independently developed another rapid and efficient technique to isolate new microsatellite markers from PAC clones. The main difference between the two methods is the larger number of repeats contributing to their primers - (CA/GT)<sub>10</sub> instead of (CA/GT)<sub>8</sub> - and the increased number of nucleotides (n = 3) at the 3' end following the repeat motif (e.g. CGA/GCT). Fujishima-Kanaya *et al.* (2003) also determined the primary sequence of the large-insert clone by

setting up two sequencing reactions using a CA-repeat primer as the forward primer and a GT-repeat primer as the reverse primer. So, their averaged initial number of primers to determine sequence information was four compared with two oligonucleotides used by Chen *et al.* (2004a).

### *3.1.3 Characterization of genes*

#### *3.1.3.1 Large-insert libraries in pig genome mapping*

Large-insert genomic libraries are required for the straightforward isolation and characterization of genes accomplished by construction of large-insert contigs. Large-insert libraries constructed in bacterial artificial chromosome (BAC), P1-derived artificial chromosome (PAC) and yeast artificial chromosome (YAC) vectors have been gradually established in livestock and extensively used in pigs for positional cloning, physical mapping and functional analysis similar to their impacts in humans and mice (Alexander *et al.*, 1997; Rogel-Gaillard *et al.*, 1997; Rogel-Gaillard *et al.*, 1999; Al-Bayati *et al.*, 1999; Anderson *et al.*, 2000; Suzuki *et al.*, 2000). A comparison of the characteristics of the three vector cloning systems is shown in Table 2 (Amemiya *et al.*, 1999; Giraldo and Montoliu, 2001). The most notable feature of these systems relative to conventional lambda and cosmid systems is their greatly increased insert sizes which may contain full-length genes and therefore allow for expedient characterization of gene structure and functions.

**Table 2:** Comparison of the BAC, PAC and YAC cloning systems (adapted from Amemiya *et al.* 1999; Giraldo & Montoliu 2001).

|                               | BAC                     | PAC                     | YAC  |
|-------------------------------|-------------------------|-------------------------|--|
| Host cells                    | <i>Escherichia coli</i> | <i>Escherichia coli</i> | <i>Saccharomyces cerevisiae</i>                |
| DNA topology of recombinants  | Circular, supercoiled   | Circular, supercoiled   | Linear   |
| Insert size                   | Up to 300 kb            | Up to 300 kb            | Up to 1 – 2 Mb                                 |
| Transformation method         | Electroporation         | Electroporation         | Spheroplast transformation                     |
| Selection for vector          | Chloramphenicol         | Kanamycin               | Drop out media (lacking tryptophan and uracil) |
| DNA yield                     | Medium                  | High                    | Low  |
| Stability                     | Very stable             | Very stable             | Variable, but can be very unstable             |
| Degree of chimerism           | Very low (<5% clones)   | Very low (<5%)          | Variable, but can be very high (>50%)          |
| Purification of intact insert | Easy                    | Easy                    | Relatively difficult                           |
| Direct sequencing of insert   | Relatively easy         | Relatively easy         | Difficult                                      |
| Insert rearrangement          | Very rare               | Very rare               | Yes  |

The BAC system is bacterial plasmid-based and the vectors are essentially an F-factor replicon with a chloramphenicol resistance gene (Shizuya *et al.*, 1992). BAC vectors can maintain genomic DNA fragments of up to 300 kb size. Very little or no rearrangements of the inserts and

hardly any chimeric clones have been observed. BACs appear to be quite stable and can easily be handled and characterized. The PAC cloning system uses a P1 bacteriophage replicon and a kanamycin resistance gene (Ioannou *et al.*, 1994). It can accommodate insert fragments of up to 300 kb and exhibits nearly all the advantages of BACs, and overcomes the drawback of low amount recombinant DNA, but the disadvantage of the PAC system is the larger vector size compared to BACs. The YAC cloning system uses two vector arms, both containing yeast selectable markers and telomere sequences, and one contains a centromere and a yeast autonomously replicating sequence (ARS) (Burke *et al.*, 1987; Nelson and Brownstein, 1994). The development of YAC permits cloning of extremely large fragments of more than 1 Mb size, however, disadvantages of the YAC system including a high degree of chimerism, inherent instability, and an overall difficulty in handling limits its usefulness.

Several large-insert genomic libraries have been constructed for the use in pig gene mapping and have been successfully used to isolate known genes and/or to generate contigs of genomic regions. Their relevant details regarding the respective libraries are provided in Table 3. The average insert sizes of these libraries range from 589 kb down to 119 kb in the order of YAC > BAC > PAC. All the libraries represent from three to five haploid genome equivalents.

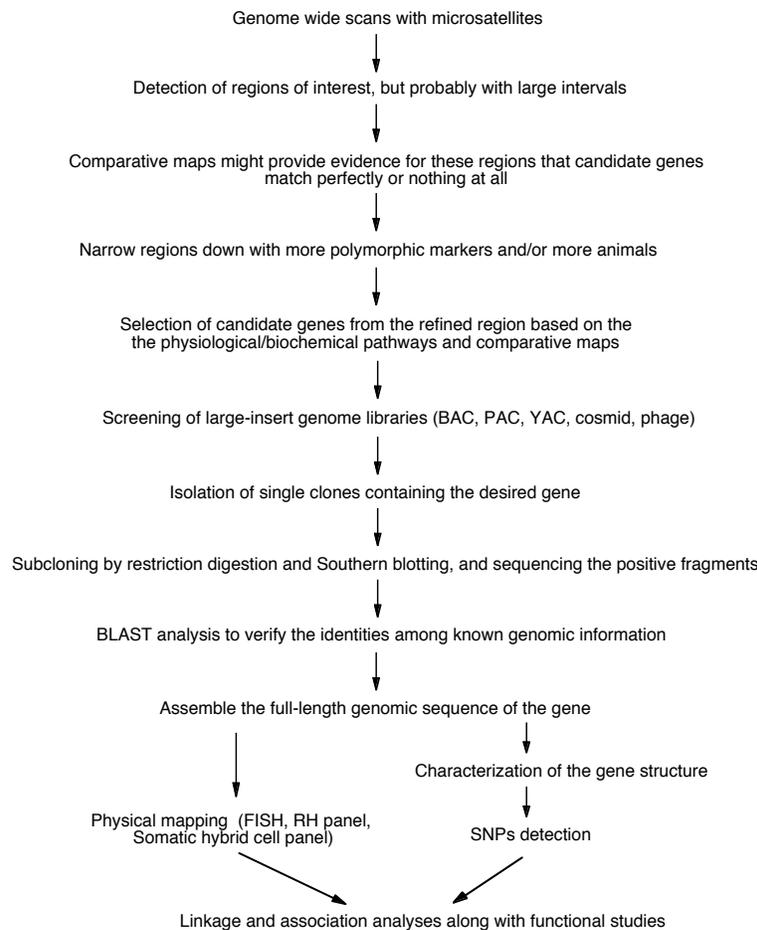
**Table 3:** Large-insert libraries used in pig genome mapping.

|                            | BAC                                 |                                     |                                     |                                 | PAC                            | YAC                                 |                                     |
|----------------------------|-------------------------------------|-------------------------------------|-------------------------------------|---------------------------------|--------------------------------|-------------------------------------|-------------------------------------|
| <b>Institution</b>         | INRA, France                        | CHORI, USA                          | Roslin Institute, UK                | NAI, Japan                      | IVM, Göttingen Germany         | USDA, USA                           | INRA, France                        |
| <b>Pig breed</b>           | Large White male                    | Duroc female                        | Large White/Meishan boar            | Large White/Landrace/Duroc boar | German Landrace male           | Meishan boar                        | Large white male                    |
| <b>Tissue</b>              | Skin fibroblasts                    | White blood cell                    | Blood lymphocytes                   | Kidney                          | White blood cell               | Blood lymphocytes                   | Blood lymphocytes                   |
| <b>Vector</b>              | pBeloBACII                          | pTARBAC1.3                          | pBeloBACII                          | pBAC-Lac                        | pCYPAC2                        | pYAC4                               | pYAC4                               |
| <b>Host strain</b>         | DH10B                               | DH10B                               | DH10B                               | DH10B                           | DH10B                          | AB1380                              | AB1380                              |
| <b>Restriction enzyme</b>  | <i>HindIII</i>                      | <i>MboI</i>                         | <i>HindIII</i>                      | <i>HindIII</i>                  | <i>MboI</i>                    | <i>EcoRI</i>                        | <i>EcoRI</i>                        |
| <b>No. of Total clones</b> | 107,520                             | 98,229                              | ~ 100,000                           | 103,488                         | 90,240                         | 33,120                              | 33,792                              |
| <b>Insert size</b>         | 135 kb                              | 173 kb                              | 150 kb                              | 133 kb                          | 119 kb                         | 589 kb                              | 280 kb                              |
| <b>Coverage genome</b>     | 5 ×                                 | 5.7 ×                               | 5 ×                                 | 4.4 ×                           | 3.2 ×                          | 5.5 ×                               | 3 ×                                 |
| <b>Library screening</b>   | PCR on pools                        | PCR on pools (High density filters) | PCR on pools (High density filters) | PCR on pools                    | PCR on pools                   | PCR on pools (High density filters) | PCR on pools                        |
| <b>Reference</b>           | Rogel-Gaillard <i>et al.</i> , 1999 | de Jong <i>et al.</i> , 2002*       | Anderson <i>et al.</i> , 2000       | Suzuki <i>et al.</i> , 2000     | Al-Bayati <i>et al.</i> , 1999 | Alexander <i>et al.</i> , 1997      | Rogel-Gaillard <i>et al.</i> , 1997 |

\* <http://bacpac.chori.org/porcine242.htm>

### 3.1.3.2 Strategy of positional candidate gene approach

As discussed in later chapters, there are strategies including candidate genes, positional cloning and QTL genome scans to isolate and characterize genes of interest. In this section, the combined strategy of positional cloning and candidate gene approach, the positional candidate gene approach as used in our studies is illustrated (Figure 2).



**Figure 2:** Flow chart of the positional candidate gene strategy

Genome wide scans of QTL mapping normally result in some large chromosomal regions associated with effects on traits of economical importance or on other traits. If the QTL regions are too large to imply direct sequencing and positional cloning, e.g. 10 Mb - 40 Mb (1 Mb = 1,000,000 bp = 1 cM), it is necessary to narrow down QTL regions with large intervals. A finemapping approach normally starts with more animals and/or the inclusion of more suitable markers. However, it might be difficult to find enough and suitable markers in a region of special interest. Thus the approach proposed by Chen *et al.* (2004a) to use already known markers in a

region to isolate additional large-insert library clones to develop some more polymorphic markers is a promising strategy. In addition, analyses of candidate genes in such chromosomal regions can be an alternative or supplement to narrow down QTL regions and further to analyse the causative genes (Collins, 1995). To determine the genomic structure of genes, isolated clones of interest are sequenced and the data processed with bioinformatic tools. After the full-length genomic sequence is assembled, the genomic structure of the gene (exon-intron boundaries, translation part, start codon, poly A) can be determined and all possible SNPs in the coding and non-coding regions can also be identified (e.g. by comparative sequencing of individuals of different breeds or phenotypes). The physical localization of the clone can be mapped alternatively by fluorescent *in situ* hybridization (FISH), somatic cell hybrid panel and radiation hybrid panel analysis. SNPs are detected to evaluate genomic differences between the so-called wild-type and the mutant-type. If an observed SNP is consistent among all mutant animals, this change is probably responsible for the mutant phenotype. Likewise, if a particular nucleotide substitution leads to an obvious non-permissive amino acid change or of a conserved promoter element, this SNP holds the possibility of affecting the structure, and then the function of the corresponding protein (Andersson, 2001). In addition, the characterization of positional candidate genes of desirable traits requires further linkage, association and functional studies. Gene expression such as RT-PCR and Northern blotting should be investigated among the wild-type and mutant animals.

### **3.2 Application of DNA markers in pig genetics**

With the development of molecular markers porcine genomic maps have been largely enriched in the last few years. The pig genetic linkage map at Roslin (Scotland) has entries for over 3,000 loci including more than 1,000 genes (<http://www.thearkdb.org/browser?species=pig>). A developing AFLP map with about 3,000 AFLP markers is well underway and will provide a great contribution input to map density when finished. However, while the average distance between markers is about 3 - 5 cM, some large gaps still exist in the pig linkage map. The physical map for pigs as for other farm animals lagged behind at the beginning. With the use of a somatic cell hybrid panel (Yerle *et al.*, 1996) and a 7,000 rad (IMpRH) or recently of a 12,000 rad (IMNpRH2) radiation hybrid panel (Yerle *et al.*, 1998; Yerle *et al.*, 2002; Hawken *et al.*, 1999), the physical map has been growing rapidly and contains now over 3,000 genes and markers.

DNA markers can be applied in a variety of different applications, but here the focus will be on the two major basic ways in which they are used in modern pig genetics: (1) as indicators of genetic diversity and in phylogenetic studies, and (2) as genetic markers for QTL mapping and for tagging traits of interests, with special emphases on porcine disease research.

### 3.2.1 Phylogenetics

The objectives of phylogenetic studies are to reconstruct the correct genealogical ties between species and to estimate the time of divergence between two organisms since they last shared a common ancestor. In phylogenetic studies, a phylogenetic tree is commonly employed to illustrate the evolutionary relationships among a group of organisms.

To date, there are likely over 600 pig breeds or lines worldwide of which two thirds reside in China and Europe and over 200 of them are considered as endangered or critical. In order to meet future challenges in the agricultural and food industries, special efforts are required to conserve the genetic resources. Therefore, phylogenetic studies in pigs aim to evaluate the genetic uniqueness and diversity of pig breeds to assist in developing a rational plan for breed conservation programmes (Ruane, 1999). Ruane (1999) has suggested a set of criteria in an attempt to choose specifically breeds for conservation. The degree of endangerment and genetic uniqueness of the breed are two of seven essential criteria discussed. In addition, the origin and history of domestic pigs can also be explained by phylogenetic analyses. Independent domestication has occurred from wild boar subspecies in Europe and Asia, and introgression of Asian germplasm into European domestic breeds took place during the 18<sup>th</sup> and early 19<sup>th</sup> centuries (Giuffra *et al.*, 2000). The divergence of ancestors for European and Chinese Meishan domestic pigs was estimated around ~500,000 years ago using results of mitochondrial DNA.

On the basis of allele frequencies of DNA markers, pairwise genetic distances between breeds can be calculated and applied in diversity analyses among different populations and phylogenetic tree construction. Among at least ten commonly used distance parameters, Cavalli-Sforza and Edward's chord distance ( $D_C$ ) and Nei *et al.*'s  $D_A$  distance generally show higher correct topology values. Nei's standard distance ( $D_S$ ) and Goldstein *et al.*'s  $\delta_\mu^2$  are more appropriate to estimate evolutionary times applying computer simulation (Takezaki and Nei, 1996). In the European Community Pig Biodiversity I project (PigBioDiv 1) (<http://databases.roslin.ac.uk/pigbiodiv>), that was completed in 2000, genetic distances represented by Reynold's  $D_R$  and Nei's  $D_S$  were

used in genetic diversity studies of European pig breeds (SanCristobal *et al.*, 2003). However, there is still no clear consensus opinion regarding which genetic distances are best to estimate phylogenetic relationships among domestic animals as they have undergone complicated genetic changes since the time domestication has started. It is likely that genetic drift and mutation are important factors affecting breed domestication. Moreover, the efficiencies of genetic distance measurements in phylogenetic construction are influenced by the number of animals and molecular loci sampled and by the method used for tree construction such as the neighbour-joining (NJ) and the unweighted pair group method with an arithmetic mean (UPGMA) (Takezaki and Nei, 1996).

To improve the reliability of phylogenetic results, a guideline for genetic distance projects was published by an FAO working group in 1993, recommending that at least 25 animals per breed should be used and that 20 to 30 microsatellite loci, with four to ten alleles each, should be typed (Barker *et al.*, 1993). More recently, several other recommendations have been made public for genetic diversity studies including the use of 2 - 5 microsatellites per chromosome and genotyping of 50 animals (25 of each sex) that are unrelated (Barker *et al.*, 1998). Nevertheless, these guidelines are sometimes ignored and results are published on lower requirements because of the limited budget and availability of some endangered breeds (Paszek *et al.*, 1998). The implication of genetic distances is also very important in genetic biodiversity studies between pig breeds. Diversity measurements such as PIC, effective allele number ( $N_E$ ), expected heterozygosity, fixation indices of loci, Nei's measure of diversity and allelic richness have been also considered in phylogenetic studies (Petit *et al.*, 1998).

So far, a number of molecular markers have been used for genetic diversity and phylogenetic analyses. These include microsatellites, minisatellites, AFLPs, SNPs, large scale or directed sequencing, mitochondrial DNA (mtDNA) and Y chromosome genotyping (Rothschild, 2003a). Microsatellite markers have been largely used in phylogenetic studies and have some value to measure differences within breeds, however due to their neutral properties, they are poorly correlated with phenotypic changes due to selection. Very recently the use of gene markers has attracted more researchers as variation in these allele frequencies may provide information related to the functional differences between breeds. The phylogenetic studies using major gene markers or SNPs associated with traits of interest would be most relevant for breed conservation and

provide potential breeds efficiently for the future production market. Despite the limited information of gene markers, Ciobanu *et al.* (2001) examined the genetic variation in two conserved local Romanian pig breeds using type I DNA markers. Moreover, mitochondrial DNA maternally inherited is useful for tracing the maternal linkages in populations. Y chromosome consists of a euchromatic non-recombining region (NRY), which makes up 95 percent of its length, flanked by two pseudoautosomal regions (Cooke *et al.*, 1985) and its genotyping can be used for measuring breed history and phylogenetic origins.

The largest ongoing project on biodiversity studies of pig breeds is the European Community Pig Biodiversity II project (PigBioDiv 2), which will evaluate and compare genetic diversity among at least 100 pig breeds originated from China and Europe (Blott *et al.*, 2003). The project will not only determine the relationships between breeds by estimating genetic distances based on microsatellite markers and haplotypic relationships from mtDNA and Y chromosome polymorphisms, but also determine functional differences among breeds by characterizing trait gene loci and QTL regions.

In addition, there are still some technical problems that limit the application of molecular markers in phylogenetic studies (Rothschild, 2003a). Methods such as microsatellites and AFLPs are constrained generally by gel electrophoresis resulting in low throughput. The use of microsatellites or SNPs requires previous identification of a polymorphism, and recognition of the polymorphisms based on size separation can be different among labs, which makes results incompatible and misleading.

### 3.2.2 Quantitative trait locus (QTL) analysis and disease survey

#### 3.2.2.1 Strategies to identify QTLs

Many economically important traits, such as growth rate, meat quality, milk production, litter size and disease resistance are supposed to be multifactorially influenced by polygenic and environmental factors. The main QTL mapping projects in pigs were summarized by Bidanel and Rothschild (2002) as shown in Table 4. A QTL is defined as a region of the genome covering one or more genes that affect a quantitative trait (Geldermann, 1975). There are usually two main strategies for finding quantitative trait loci including association analysis along with candidate gene typing and genome scans based on linkage mapping (Beuzen *et al.*, 2000; Andersson, 2001; Dean, 2003; Rothschild, 2003b).

**Table 4:** Main QTL mapping programmes in pigs (Bidanel & Rothschild 2002)

| Country     | Institution                       | Population <sup>a</sup> | Animals | Markers                             | Traits <sup>b</sup> |
|-------------|-----------------------------------|-------------------------|---------|-------------------------------------|---------------------|
| Belgium     | Liege University                  | LW × PI                 | 1032    | 11 (SSCr 2)                         | GR, CC              |
| France      | INRA                              | MS × LW                 | 1103    | 137                                 | GR, CC, MQ, RE, BE  |
| Germany     | Hohenheim University              | PI × (MS or WB)         | 979     | 121                                 | GR, CC, MQ          |
| Japan       |                                   | MS × PM                 | 215     | 318                                 | GR, CC              |
| Netherlands | Wageningen University             | MS × LM                 | 800     | 127-132                             | GR, CC, MQ          |
| Norway      | Agricultural University of Norway | (DU × LR) × LW          | 305     | 11 (SSCr 4); 9 (SSCr 6); 9 (SSCr 7) | MQ                  |
| Scotland    | Roslin Institute                  | MS × LW                 | 390     | 9 (SSCr 4)                          | GR, CC              |
| Spain       | IRTA-INIA                         | LR × IB                 | 500     | 7 (SSCr 6)                          | GR, CC, MQ          |
| Sweden      | University of Uppsala             | WB × LW                 | 191     | 117-236                             | GR, CC, MQ, DR, ... |
| USA         | Iowa State University             | CH × YO                 | 294     | 5 (SSCr 4); 10 (SSCr 7)             | GR, CC, MQ          |
| USA         | Iowa State University             | BE × YO                 | 500     | 125                                 | GR, CC, MQ          |
| USA         | University of Minnesota           | MS × LW                 | 298     | 119                                 | GR, CC              |
| USA         | USDA                              | MS × SL                 | 540     | 157                                 | GR, CC, RE          |
| USA         | University of Nebraska            | ORL × CL                | 306     | 151                                 | RE                  |
| USA         | University of Wisconsin           | HGL × LGL               | 240     | 75                                  | GR                  |

<sup>a</sup>BE = Berkshire; CH = Chinese breeds; IB = Iberian breed; CL = control line; HGL, LGL = line selected for high and low growth rate, respectively; LR = Landrace; LW = Large White; MS = Meishan; ORL = line selected for ovulation rate and embryo survival; PI = Pietrain; SL = synthetic line; WB = wild boar; YO = Yorkshire

<sup>b</sup>BE = behavioral traits; CC = carcass composition; DR = disease resistance; GR = growth; MQ = meat quality; RE = reproduction

Candidate gene analyses have been employed to investigate a variety of traits whose physiology is well understood. This approach can be very powerful and can detect loci even with small effects provided that the candidate gene is the causative gene. To evaluate all possible candidate genes for the trait of interest seems, however, too time-consuming and therefore impractical. In farm animals, linkage disequilibrium is more frequently observed (Farnir *et al.*, 2000) which can increase the probability of finding true associations between genetic markers and trait loci, but it also elevates the risk of false-positive associations.

Genome wide scans are useful to resolve the genetics of traits for which hardly anything is known. Genome scans may also reveal previously unsuspected loci as candidate genes for well understood traits. In order to increase the statistical power in QTL mapping, artificial crosses are often made between divergent populations, e.g. European and Chinese pig breeds intercross (Spelman and Bovenhuis, 1998) or progeny testing using the granddaughter design in dairy cattle (Weller *et al.*, 1990; Heyen *et al.*, 1999). In principle, a sufficiently large number of animals in properly designed populations in combination with a set of polymorphic markers evenly spaced

across the genome (high resolution map) are also required to map a QTL with a major effect. In addition, several other factors can influence the chance to detect a QTL, including the size of the effect, the frequency of the alleles, the heritability of the trait, and the chosen thresholds in the statistical analysis (Haley, 1995). Recently, another approach of using a denser collection of microsatellite markers within and nearby genes involving functional variants has been developed for the human genome project and proved to have an increased power (Collins *et al.*, 2003). Moreover, mappers have to be aware that any potential polymorphisms of the gene must later be confirmed in commercial populations before their wide application in marker-assisted selection in breeding programmes.

The chromosome regions of most QTLs are too imprecise for direct positional cloning (Knott and Haley, 1992; Andersson *et al.*, 1994), and thus positional candidate gene approaches are extensively used to identify the causal genes in QTL regions. So far, there is only one example of successful positional cloning known: *PRKAG3* in porcine skeletal muscle is associated with excess glycogen content (Milan *et al.*, 2000). A missense mutation of R222Q in *PRKAG3* causes poor porcine meat processing quality. Once a QTL has been mapped, the potential candidate genes will be investigated to narrow the region of interest by a stepwise process. By comparing the coding sequences and expression levels of positional candidate genes in populations, the unrelated genes might be eliminated and the causative quantitative trait nucleotide (QTN) can be discovered. QTL mutations were recently also found in regulatory rather than in coding regions. Results from *Drosophila* suggest that some trait loci are too difficult to prove causal relationships with QTLs leading to the demand that all gene regions need careful analysis (Mackay, 2001). This postulate has been further confirmed by the finding that a regulatory mutation of *IGF-2* (*IGF2*-intron3-nucleotide 3072) causes a major QTL effect on muscle growth in the pig (van Laere *et al.*, 2003).

#### 3.2.2.2 Comparative mapping and functional analysis

The large numbers of expressed sequence tags (ESTs) from cDNA clones can assist comparative mapping and large-scale expression analysis using array technologies. At present, there are approximately 171,920 porcine ESTs (update to 14<sup>th</sup> November 2003) deposited in Genbank ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)). The largest EST project published to date was initiated by the USDA. As an example, the PathoCHIP project funded by the

European community (<http://www.pathochipproject.com>) aims to use spotted cDNA array for disease organism and immune response gene studies. International cooperative efforts will therefore largely advance the pig transcriptome and functional analyses. Comparative mapping is also an important approach in farm animal genome studies as it allows extensive exploitation of human genetic map. Chromosome painting has been successfully used to establish regions of conserved synteny between mammalian animals and humans with human chromosome-specific probes (Chowdhary *et al.*, 1998). Gene expression researches may be particularly valuable to detect imprinted genes (de Koning *et al.*, 2000). Imprinting is the phenomenon in which only one of the two alleles of a gene may be expressed either on random events or origin of parent effect (Hall, 1997). The above mentioned *IGF-2* linked QTL has been reported to be genetically an imprinted locus affecting body composition in pigs (Georges *et al.*, 2003; van Laere *et al.*, 2003). Therefore, an improved strategy might be to develop a haplotype map of the whole genome which is useful to design the second generation genome scans (Weiss and Clark, 2002).

In the last few years, the mutagenesis strategy has also been proposed and offers a different perspective in finding genes determining complex traits (Nadeau and Frankel, 2000). Mutagenesis which considers all genes in a pathway will identify rare mutations causing disease. In contrast, QTL mapping is more likely to identify genes that encode regulatory proteins or rate-limiting enzymes (Korstanje and Paigen, 2002).

### 3.2.2.3 Monogenic diseases

Simple monogenic diseases in farm animals have been identified and are catalogued in the Online Mendelian Inheritance in Animals (OMIA) database (<http://morgan.angis.su.oz.au/Databases/BIRX/omia/>) (Table 5). Some diseases may not be apparent to breeders or even associated with benign traits in affected animals, and thus inadvertently selected in animal breeding. Hyperkalaemic periodic paralysis is an example of an autosomal dominant disease in horses. Affected individuals have more muscle because the biochemical defect is due to a mutation in the sodium channel of muscle tissue (Rudolph *et al.*, 1992). An important autosomal recessive disease in cattle is the bovine leucocyte adhesion deficiency (*BLAD*), a severe immunodeficiency syndrome caused by a missense mutation that prevents the production of effective *CD18*. The disease is widely spread in the US Holstein-Friesian cattle because one popular sire with a huge number of progeny was the carrier for this mutation (Shuster *et al.*, 1992). For these cases in

which molecular basis has been revealed, diagnostic tests such as a PCR-RFLP screening test have been developed and applied in breeding programmes to increase animal welfare, to reduce the incidence of diseases and to improve production efficiency.

**Table 5:** Summary of the OMIA database contents for major species (latest update: 16<sup>th</sup> October 2003).

|  | Cat | Cattle | Chicken | Dog | Emu | Fox | Goat | Horse | Pig | Quail | Rabbit | Sheep | Turkey | Total |
|--|-----|--------|---------|-----|-----|-----|------|-------|-----|-------|--------|-------|--------|-------|
| Disorders/traits                                 | 263 | 357    | 174     | 451 | 4   | 5   | 66   | 184   | 203 | 34    | 48     | 179   | 28     | 1996  |
| Single-locus disorder/traits                     | 38  | 56     | 63      | 100 | 2   | 3   | 8    | 26    | 33  | 19    | 12     | 59    | 8      | 427   |
| Disorder/traits (causative mutation identified)* | 7   | 27     | 11      | 38  | 1   | 2   | 5    | 9     | 11  | 2     | 3      | 9     | 1      | 126   |
| Potential animal models for a human disorder     | 123 | 117    | 34      | 207 | 2   | 3   | 24   | 86    | 65  | 9     | 27     | 62    | 3      | 762   |

\*Disorders/traits for which the causative mutation has been identified at the DNA level

#### 3.2.2.4 Disease resistance and immune response traits

Different strategies have been used to identify the genes responsible for multifactorial diseases. Compared with QTL mapping for other traits of economical importance, little progress has been made so far for immune response and disease resistance, primarily owing to the lack of pedigree data with informative disease records and the difficulty to generate dedicated resource populations with a high incidence for the disease. The ryanodine receptor (*RYR1*) encoding the calcium release channel in skeletal muscle was identified as a functional candidate gene. A single missense mutation of R614C was revealed to be responsible for susceptibility to malignant hyperthermia in pigs, which could be triggered by stress or exposure to the anaesthetic gas halothane. This mutation was also associated with increased leanness and appeared to be inadvertently selected by breeders (Fujii *et al.*, 1991). If the genes associated with human or mouse diseases are well studied, candidate gene approaches can be used to identify the genes responsible for these diseases in farm animals and then lead to their functional and association analyses, such as natural resistance associated macrophage protein 1 (*NRAMP1*) and clotting factor IX (Sun *et al.*, 1998; Singer *et al.*, 1996).

To date, QTL scans for immune response and disease resistance are limited in number. Four QTL regions with significant effects on immune capacity were detected on four chromosomes respectively in a Large White × Wild boar cross (Edfors-Lilja *et al.*, 1998). In 2000, additional four QTLs were found on chromosomes 2, 6, 8, and 12 associated with stress-induced alterations

of immune functions in which stress of mixing and transportation was imposed on three-month old pigs (Edfors-Lilja *et al.*, 2000). Two alpha (1, 2) fucosyltransferase genes (*FUT1*, *FUT2*) on porcine chromosome *6q11* have been identified that are closely linked to the blood group inhibitor (S) and *E. coli* F18 receptor (*ECF18R*) loci (Meijerink *et al.*, 1997). Presumably, mechanisms of *FUT1* or *FUT2* gene products are responsible for bacterial adhesion and need to be further determined. A QTL region affecting susceptibility to K88 *E. coli* infection is localized on chromosome 13 (Edfors-Lilja *et al.*, 1995). In addition, Reiner *et al.* (2002) detected QTLs for resistance/susceptibility to pseudorabies virus (*PrV*) in swine by a genome-wide scan (85 microsatellite markers evenly spaced on the 18 pig autosomes and on the pseudoautosomal region of the X chromosome) using a resource family (European Large White × Chinese Meishan pigs) (Reiner *et al.*, 2002). The results revealed that after being challenged, the pure-bred Large White pigs, the F<sub>1</sub> and three-quarters of the F<sub>2</sub> animals, but none of the Meishan pigs, developed the neurological symptoms and died or were euthanized. QTLs for appearance/non-appearance of neurological symptoms were found on chromosomes 5, 6, 9 and 13; QTLs for rectal temperature after *PrV* challenge were found on chromosome 2, 4, 8, 10, 11 and 16. Therefore, fine mapping and association analyses of positional candidate genes are required substantially to identify the causative QTN or genes.

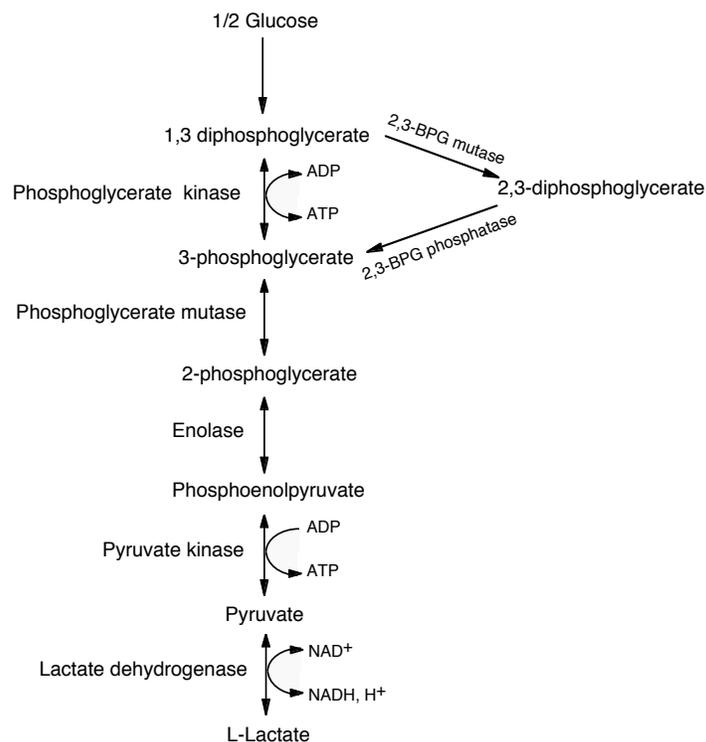
Due to the acceleration of identification of genes from QTL studies, it is suggested that QTL mapping is a promising technique. The harvesting era of QTL genes is beginning (Korstanje and Paigen, 2002). Functional studies and comparative mapping will facilitate the identification of genes associated with immune response and disease resistance. In addition, QTL mapping can be further enhanced by the increasing availability of new markers, the efficiency of genotyping technology, and improved statistical methods for analysing QTL data.

### 3.3 The phosphoglycerate kinase (PGK) gene family

#### 3.3.1 Members of the PGK gene family

##### *3.3.1.1 Functional loci and pseudogenes of the PGK gene family*

Phosphoglycerate kinase (EC: 2.7.2.3) catalyzes the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate generating one molecule of ATP in the glycolytic pathway, formally known as the Embden-Meyerhof pathway, which converts glucose (or fructose) to pyruvate (Figure 3) (Devlin, 1997). Phosphoglycerate kinase does also modulate sperm metabolism during epididymal transport (Salisbury *et al.*, 1977). It is also critical for tumor expansion and metastasis through the angiogenic process as a disulfide reductase (Lay *et al.*, 2000). Hybridization results of a full-length cDNA encoding for phosphoglycerate kinase indicated the existence of a small family of dispersed *PGK* or *PGK-like* genes in man (Michelson *et al.*, 1983; Michelson *et al.*, 1985; McCarrey and Thomas, 1987): an ubiquitously expressed, X-linked form referred to as *PGK1* (*PGKA*) and a testis-specific, autosomally determined isoform referred to as *PGK2* (*PGKB*) on chromosome *6p21.1-p12* are known besides two pseudogenes. *PGK1P1* is located on *Xq11-q13*, proximal to the expressed *PGK1* gene (Michelson *et al.*, 1985; Willard *et al.*, 1985), whereas *PGK1P2* is assigned to chromosome 19 (Willard *et al.*, 1985; Gartler *et al.*, 1986).



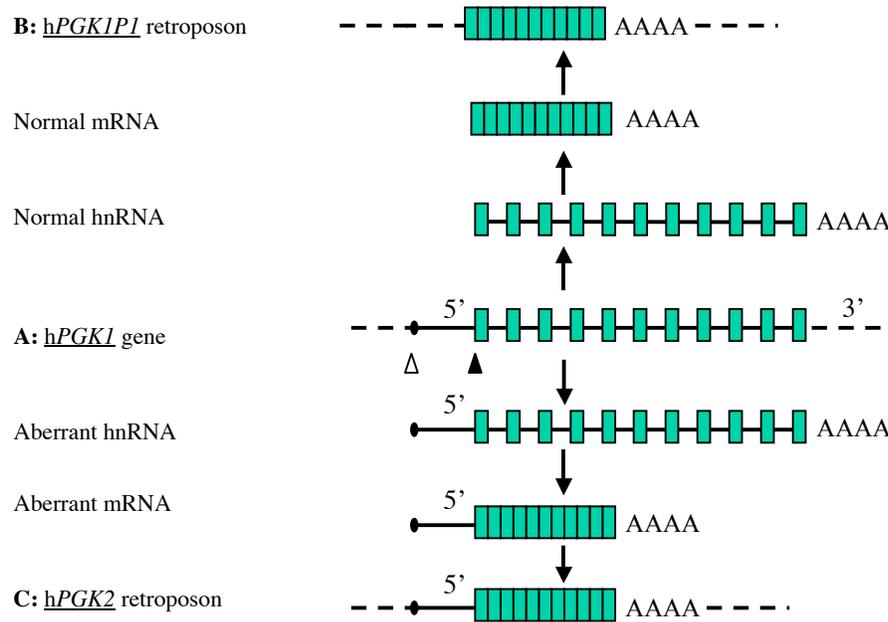
**Figure 3:** The glycolytic pathway (Devlin 1997) (BPG=bisphosphoglycerate).

PGK is a highly-conserved enzyme in all eukaryotic species, including yeast as the human-yeast chimeric PGK molecules are also functional (Mas *et al.*, 1986). Therefore, it is reasonable to assume that similar functional regions have been conserved in the human PGK proteins. Two isozymes of PGK, each encoded by a separate gene, are differentially expressed during spermatogenesis in mammals (Kramer and Erickson, 1981; VandeBerg, 1985; McCarrey *et al.*, 1996). The PGK1 isozyme is encoded by the X-linked *PGK1* gene and is expressed in premeiotic spermatogenic cells as well as in all somatic cells and in oogenic cells (VandeBerg, 1985), while the PGK2 isozyme is exclusively expressed in meiotic and postmeiotic spermatogenic cells, which was proposed to compensate for reduced expression of the *PGK1* gene in these cells (McCarrey and Thomas, 1987; McCarrey, 1990). By analysis of the cDNA and the encoded proteins of these two functional *PGK* genes, it was suggested that none of the differences in the amino acid sequence between the PGK1 and PGK2 protein contributes to any likely change of catalytic function of these two isozymes. However, a potential difference exists between them suggesting that the PGK2 protein is more stable *in vivo* than the PGK1 protein (McCarrey *et al.*, 1996). To identify some associated genes or markers for male fertility and/or male fertility dysfunction, the porcine functional *PGK* genes were identified by Chen *et al.* (2004b) and Chen *et al.* (2004c). The X-linked *PGK1* on *SSCXq12-q13* and the testis-specific *PGK2* on *SSC7q14-q15*, respectively, coincide with the homology between the X chromosomes in human and pig, and the homology between the p arm of human chromosome 6 and q arm of pig chromosome 7.

### 3.3.1.2 Molecular evolution of *PGK* gene family by retrotransposon events

Analysis of the nucleotide sequence of *PGK2* and *PGK1P1* suggested that they originated from the “housekeeping” gene *PGK1* more than 100 million years ago by RNA-mediated gene duplication. It is hypothesized that the *PGK2* gene might then be a transcribed retrotransposon (Boer *et al.*, 1987; McCarrey, 1990). A comparison of the human and murine *PGK* genes indicates that *PGK2* has evolved more rapidly than the *PGK1* gene since the two genes diverged early in mammalian evolution, but the lack of introns in *PGK2* may have diminished the inter-exon variation. Nucleotide sequence characteristics indicate that the *PGK2* retroposon initially included a copy of the endogenous *PGK1* “housekeeping” 5'-regulatory sequences, which facilitated the initial expression of this retroposon (McCarrey, 1987). Therefore, it is implied that the *PGK2* gene initially carried a “housekeeping” promoter sequence that directed ubiquitous

low-level transcription, and has subsequently evolved a tissue-specific promoter that directs high-level transcription (McCarrey, 1990) (Figure 4). With the comparison of the *PGK* genes in human and mouse (Table 6), it was shown that these genes share high homology.



**Figure 4:** The origin of human *PGK* retroposons (McCarrey 1990). The transcript initiated at the normal start point (solid arrowhead) in *PGK1*; The functional *PGK2* gene is shown originating by similar processing of an aberrant transcript of *PGK1* that initiated at an upstream start point (open arrowhead) and included the endogenous *PGK1* 5'-promoter sequence. hnRNA= heteronuclear RNA.

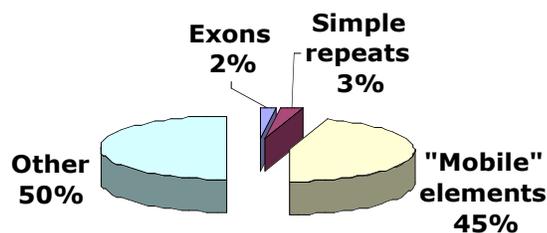
**Table 6:** *PGK* gene region homologies in human and mouse (%) (McCarrey 1990).

| Genes compared                 | Promoter (200 bp) | 5'-UTR (90 bp) | Coding sequence (1251 bp) |      | 3'-UTR (200 bp) |
|--------------------------------|-------------------|----------------|---------------------------|------|-----------------|
|                                |                   |                | N.A                       | A.A  |                 |
| h <i>PGK1</i> /m <i>PGK1</i>   | 48                | 65.2           | 92.9                      | 97.6 | 55.8            |
| h <i>PGK2</i> /m <i>PGK2</i>   | 63.9              | 44.8 (67.2)*   | 84.9                      | 86.1 | 56.0            |
| h <i>PGK1</i> /h <i>PGK2</i>   | 44.1              | 46.2           | 85.3                      | 87.1 | 55.9            |
| m <i>PGK1</i> /m <i>PGK2</i>   | 36.1              | 24.7 (40.7)*   | 79.2                      | 84.0 | 50.5            |
| h <i>PGK1</i> /h <i>PGK1P1</i> | 30.9              | 74.4           | 94.1                      | 90.0 | 98.5            |

\* Values in parentheses are based on maximized alignments with deletions excluded. N.A. = nucleic acid, A.A. = amino acid

Evolutionary biologists hypothesized that the earliest life forms contained self-replicating RNA genomes. The advent of polymerases that make DNA copies of RNA templates allowed the conversion of information from unstable ribose-based polymers to more stable deoxyribose-based polymers through the process of reverse transcription. In this way, reverse transcription appears to have a pivotal role in the formation of the first DNA genomes (Kazazian and Moran, 1998). It

is now apparent that reiterative rounds of reverse transcription served to expand both the size and complexity of the mammalian genomes. Retrotransposons of any sequence are potentially threatening to the normal genome function, however, the mechanisms controlling this process remain poorly understood (Curcio and Garfinkel, 1999; Moran *et al.*, 1999; Werle-Schneider *et al.*, 1999). As a result of the human genome project, we now know that only approximately 2-3% of the human genome is comprised of exonic sequences. The remainder, so-called “junk DNA”, is composed largely of introns, simple repeat sequences and mobile elements or their remnants (Figure 5). Mobile elements fall into three major classes: DNA-based transposable elements, autonomous retrotransposons and non-autonomous retrotransposons (Kazazian and Moran, 1998). The non-autonomous retrotransposon is composed mainly of *Alu* elements and processed pseudogenes, both of which end in a poly(A) tail and lacking protein-coding capacity. Their mobilization requires a cellular source of reverse transcriptase, which is mostly likely encoded by retrotransposition-component long interspersed nuclear elements (LINEs or LIs). Mobile elements were firstly recognized as potential causal agents of human disease in 1988 when two separate insertions of truncated LIs retrotransposon were found to disrupt the factor VIII gene, resulting in haemophilia A (Kazazian *et al.*, 1988). Recently, six additional retrotransposed LI insertion were subsequently found, while five of them occurred either in the germ line or during the early development (Woods-Samuels *et al.*, 1989; Miki *et al.*, 1992; Narita *et al.*, 1993; Holmes *et al.*, 1994).



*Figure 5: Genome Anatomy*

Most pseudogenes have multiple features that confirm their non-functional status. However, there are genes that have many features of pseudogenes, but with the recognition that these genes are potentially functional. Accordingly, experimental design and interpretation across the whole field of molecular genetics must take pseudogenes into careful consideration (Mighell *et al.*, 2000). Processed pseudogenes are probably generated by RNA polymerase II rather than the RNA

polymerase III transcription responsible for other retrotransposons such as either *Alu* or LINE repeats (Maestre *et al.*, 1995).

### 3.3.2 Biological/physiological function of the PGK

#### *3.3.2.1 Expression and encoded proteins of PGK genes*

The mammalian testis is a common location for expression of tissue-specific isozymes of enzymes expressed ubiquitously in somatic cell types (Goldberg, 1977; Hecht, 1993). In many cases, these isozymes are encoded by several autosomal loci. This situation appears to reflect unique requirement in somatic isozyme, thus mandating the need for a specialized testis-specific isozyme. One of the best characterized examples is the *PGK* gene system, encoding the PGK protein which is a two-domain protein; each domain is composed of six repeats of an *alpha/beta* structural motif. In this system, expression of the X-linked somatic form of the enzyme (PGK1) is repressed during the spermatogenesis due to transcriptional inactivation of the single X chromosome in meiotic spermatocytes (McCarrey *et al.*, 1992a; McCarrey *et al.*, 1992b; Kumari *et al.*, 1996). Thus, it has been suggested that the autosomal gene encoding the testis-specific isozyme (PGK2) may be expressed in these cells simply to compensate for this loss, rather than to provide any specialized function in the late spermatozoa (McCarrey, 1994). This suggestion was also supported by enzymological studies of PGK isozymes in humans, mice, and other mammals, and from the evidence of no obvious differences of the molecular size, substrate affinity, energy of activation, coenzyme specificity, catalytic properties, thermal denaturation, as well as  $K_m$  and  $V_{max}$  values between these isozymes (Pegoraro and Lee, 1978; VandeBerg, 1985).

By analysis of the specific activities and synthesis of the ubiquitous isozyme PGK1 and the testis-specific isozyme PGK2 in spermatogenic cells of the mouse, it was demonstrated that there is a five-fold increase of total PGK specific activity between immature and adult testes. The increase in total PGK is entirely due to the appearance and increase of the PGK2 isozyme (Kramer and Erickson, 1981; Erickson, 1990). The occurrence of the transcripts produced from the ubiquitously expressed X-linked *PGK1* gene and the testis-specific autosomal *PGK2* gene during spermatogenesis in mouse, revealed that the transcription of *PGK2* is initiated coincidentally with the onset of meiosis in male germ cells, and then continues to increase in the later spermatocytes and postmeiotic round spermatids.

### 3.3.2.2 *PGK transcript switch and PGK methylation*

The spermatogenic pathway consists of complex developmental events, such as differentiation of spermatogonial stem cells, meiotic cell division, and morphogenic maturation of sperm (Hecht, 1986). Many genes undergo altered expression during this process, and the resulting repertoire of testis-specific proteins are believed to be required both for advancing the spermatogenic pathways and for maintaining the architectural organization of germ cell. Although the molecular basis for the spermatogenic pathway remains to be elucidated, changes in gene expression at both the transcriptional and translational levels are proposed to play important roles (Willison and Ashworth, 1987; Prospst *et al.*, 1988; Erickson, 1990). In the testis, PGK1 exists in spermatogenic cells only during the early stages of spermatogenesis, and the PGK isozyme switches from PGK1 to PGK2 as the pathway proceeds (Kramer, 1981). It was later proposed that the isozyme switch of murine PGK occurs at the transcriptional level during the pachytene spermatocyte stage (Goto *et al.*, 1990; Singer-Sam *et al.*, 1990; McCarrey *et al.*, 1992a). The spermatogenic expression of the murine *PGK2* gene was thought to be regulated by both positive and negative *cis*-acting DNA elements (Goto *et al.*, 1991; Goto *et al.*, 1993; Mizuno *et al.*, 1992). Some of the *cis*-elements could be responsible, at least partly, for the PGK transcript switch in the testis. The negative *cis*-element functions as a silencer (Mizuno *et al.*, 1992), which consists of two separate DNA regions (Nishiyama *et al.*, 1994). By analysis of the upstream region of the *PGK2* gene, the proper expression of the mouse *PGK2* gene requires an upstream region longer than that required by the human gene suggesting the presence of one or more negative *cis*-elements that suppress murine *PGK2* expression in non-testicular organs (Ando *et al.*, 2000).

DNA methylation is thought to play a role in the regulation of tissue-specific genes in animal cells (Cedar, 1988). Most tissue-specific genes are fully methylated in sperm and in almost all somatic tissues of the adult organism (Yisraeli and Szyf, 1984). In the respective tissue of expression, however, these tissue-specific genes undergo a striking “demethylation” that could either be necessary for gene transcription or represent a secondary response to induction which serves to maintain the new active state (Ariel *et al.*, 1991). *PGK2*, a non-CpG-island containing gene is highly methylated in nonexpressing cell types but undermethylated in pachytene spermatocytes and round spermatids, where it is actively transcribed. At later stages of spermatogenesis, however, the gene becomes remethylated and thus acquires the full modification pattern in sperm DNA. It was revealed that the sperm DNA emerging from the testis

does not contain any germ-line-specific unmethylated sites and thus carries the methylation pattern typical for that in somatic tissues (Ariel *et al.*, 1991). A further investigation observed a specific demethylation event in the *PGK2* gene in prospermatogonia at about time of birth. Approximately 10 days before the onset of transcription which occurs for the first time in primary spermatocytes, the *PGK2* gene was unmethylated in adult spermatogenic testicular cells, but was remethylated in mature spermatozoa in the vas deferens. Surprisingly, this remethylation is part of the process of sperm maturation which occurs in the epididymis (Ariel *et al.*, 1994).

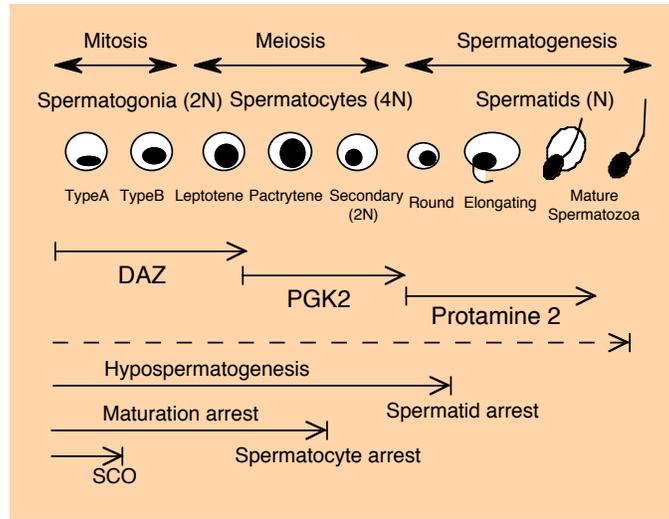
### 3.3.3 Direct effects and associations of the PGKs

#### *3.3.3.1 PGK and diseases*

PGK1 isozyme deficiency was found to be associated with hemolytic anemia, mental disorder and myoglobinuria in human (Fujii and Yoshida, 1980; Tonin *et al.*, 1993). Recently, Lay *et al.* (2000) found that the administration of phosphoglycerate kinase to tumour-bearing mice caused an increase in plasma levels of angiostatin, and a decrease in tumour vascularity and rate of tumour growth (Lay *et al.*, 2000). It was indicated that phosphoglycerate kinase not only functions in glycolysis, but is also secreted by tumour cells and participates in the angiogenic process as a disulphide reductase.

#### *3.3.3.2 PGK and male reproduction*

Studies for PGK2 deficiency are limited to date, mainly due to its solely expression in testis, but the concerns for the testis-specific PGK2 are now growing for its putative function related to male reproductive traits (e.g. sterility and fertility) in mammals. The mature spermatozoa requires significant amounts of phosphoglycerate kinase to participate in the metabolism of fructose which is an essential component in semen and which is secreted in both the male and female reproductive tracts (Mann, 1964). Moreover, the inactivation of the single X-chromosome in spermatogenic cells before meiosis may have contributed to the need for a functional autosomal *PGK* locus (McCarrey and Thomas, 1987). According to the presence or absence of different gene transcripts associated with various spermatogenic defects, *DAZ* (Deleted in AZoospermia), *PGK2* and protamine 2 genes have been detected by RT-PCR in patients with nonobstructive azoospermia (Figure 6). The PCR diagnosis was employed to predict the testicular spermatozoa (Song *et al.*, 2000).



**Figure 6:** Nuclear status and gene expression during the spermatogenesis. (*PGK2* gene is transcribed in the pachytene spermatocyte.) (Song *et al.* 2000).

Koizumi *et al.* (1990) concluded that the relative *PGK2* quantity provided information on the testicular development and is therefore useful as an indicator of testicular function, such as sperm production capacity. A good correlation was observed between the relative *PGK2* quantities and sperm/spermatid counts. This correlation was also supported by further evidence that in case of sterility-associated with chromosomal abnormality (chromosomal translocation between chromosome X and chromosome 16), a relative *PGK2* quantity was not detected in any of the seven abnormal adult mice. Recently, both the equine and the murine *PGK2* gene were identified within the cysteine-rich secretory protein (*CRISP*) gene family, which are expressed in the male genital tract and that are thought to be involved in sperm-egg fusion (Kasahara *et al.*, 1989; Richards *et al.*, 1989; Giese *et al.*, 2002). Therefore, these genes are supposed to be of interest as candidate genes for inherited male fertility dysfunction and as putative QTL region for male fertility traits.

#### 4 REFERENCES

- Al-Bayati HK, Duscher S, Kollers S, Rettenberger G, Fries R, Brenig B (1999) Construction and characterization of a porcine P1-derived artificial chromosome (PAC) library covering 3.2 genome equivalents and cytogenetical assignment of six type I and type II loci. *Mamm Genome*, **10**, 569-572.
- Alexander LJ, Smith TPL, Beattie CW, Broom MF (1997) Construction and characterization of a large insert porcine YAC library. *Mamm Genome*, **8**, 50-51.
- Amemiya CT, Zhong TP, Silverman GA, Fishman MC, Zon LI (1999) Zebrafish YAC, BAC, and PAC genomic libraries. *Methods in Cell Biology*, **60**, 235-258.
- Anderson SI, Lopez-Corrales NL, Gorick B, Archibald AL (2000) A large-fragment porcine genomic library resource in a BAC vector. *Mamm Genome*, **11**, 811-814.
- Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet*, **2**, 130-138.
- Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, *et al.* (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science*, **263**, 1771-1774.
- Ando H, Haruna Y, Suzuki M, Yamada S, Okabe M, Nakanishi Y (2000) Ectopic activation of the transcription promoter for the testis-specific mouse P<sub>gk</sub>-2 gene on elimination of a cis-acting upstream DNA region. *Dev Growth Differ*, **42**, 385-393.
- Archibald AL, Haley CS, Brown JF, Couperwhite S, McQueen HA, Nicholson D, Coppieters W, Van de Weghe A, Stratil A, Wintero AK, *et al.* (1995) The PiGMap consortium linkage map of the pig (*Sus scrofa*). *Mamm Genome*, **6**, 157-175.
- Ariel M, McCarrey J, Cedar H (1991) Methylation patterns of testis-specific genes. *Proc Natl Acad Sci USA*, **88**, 2317-2321.
- Ariel M, Cedar H, McCarrey J (1994) Developmental changes in methylation of spermatogenesis-specific genes include reprogramming in the epididymis. *Nat Genet*, **7**, 59-63.
- Armour JA, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. *Hum Mol Genet*, **3**, 599-565.
- Barker JSF, Bradley DG, Fries R, Hill WG, Nei M, Wayne RK (1993) An integrated global programme to establish relationships among the breeds of each animal species. *FAO Animal Production and Health Paper*, FAO, Rome.
- Barker JSF, Hill WG, Bradley D, Nei M, Fries R, Wayne RK (1998) Measurement of domestic animal diversity (MoDAD). *Original Working Group Report*, FAO, Rome.
- Beuzen ND, Stear MJ, Chang KC (2000) Molecular markers and their use in animal breeding. *The Veterinary J*, **160**, 42-52.
- Bidanel JP, Rothschild MF (2002) Current status of quantitative trait locus mapping in pigs. *Pig News and Inform*, **23**, 39N-54N.
- Blott S, Andersson MG, SanCristobal CC, Cardellino R, Li N, Huang L, Li K, Plastow G, Haley C (2003) Characterisation of genetic variation in the pig breeds of China and Europe – the PigBioDiv2 project. *Arch Zootec*, **52**, 207-217.

- Boer PH, Adra CN, Lau YF, McBurney MW (1987) The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon. *Mol Cell Biol*, **7**, 3107-3112.
- Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies. *Curr Opin in Genet and Dev*, **3**, 939-943.
- Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806-812.
- Campbell EM, Nonneman D, Rohrer GA (2001) Fine mapping a quantitative trait locus affecting ovulation rate in swine on chromosome 8. *J Anim Sci*, **81**, 1706-1714.
- Cassady JP, Johnson RK, Pomp D, Rohrer GA, Van Vleck LD, Spiegel EK, Gilson KM (2001) Identification of quantitative trait loci affecting reproduction in pigs. *J Anim Sci*, **79**, 623-633.
- Cedar H (1988) DNA methylation and gene activity. *Cell*, **53**, 3-4.
- Chen KF, Knorr C, Bornemann-Kolatzki K, Huang LS, Brenig B (2004a) Targeted isolation of microsatellites and their application in phylogenetic analyses of European and Chinese pig breeds. *in preparation*.
- Chen KF, Beck J, Huang LS, Knorr C, Brenig B (2004b) Assignment of the phosphoglycerate kinase 1 (PGK1) gene to porcine chromosome Xq12-q13 by fluorescence in situ hybridization and confirmation by hybrid panel analyses. *Anim Genet*, doi: 10.1111/j.1365-2052.2004.01092.x.
- Chen KF, Beck J, Huang LS, Knorr C, Brenig B (2004c) Assignment of the phosphoglycerate kinase 2 (PGK2) gene to porcine chromosome 7q14-q15 by fluorescence in situ hybridization and by analysis of somatic cell and radiation hybrid panels. *Anim Genet*, **35**, 71-72.
- Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H (1998) Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Res*, **8**, 577-589.
- Cifarelli RA, Gallitelli M, Cellini F (1995) Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite-containing DNA clones. *Nucleic Acids Res*, **23**, 3802-3803.
- Ciobanu DC, Nagy A, Wales R, Day AE, Rothschild MF, Plastow GS (2001) Genetic variation in two conserved local Romanian pig breeds using type 1 DNA markers. *Genet Sel Evol*, **33**, 417-432.
- Collins FS (1995) Positional cloning moves from perdditional to traditional. *Nat Genet*, **9**, 347-350.
- Collins JR, Stephen RM, Gold B, Long B, Dean M, Burt SK (2003) An exhaustive DNA microsatellite map of the human genome using high performance computing. *Genomics*, **82**, 10-19.
- Cooke HJ, Brown WR, Rappold GA (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature*, **317**, 687-692.
- Curcio MJ, Garfinkel DJ (1999) New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. *Trends Genet*, **15**, 43-45.
- de Koning DJ, Rattink AP, Harlizius B, van Arendonk JAM, Brascamp EW, Groenen MAM (2000) Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci USA*, **97**, 7947-7950.
- Dean M (2003) Approaches to identify genes for complex human diseases: lessons from mendelian disorders. *Hum Mutation*, **22**, 261-274.

- Delvin TM (1997) *Textbook of biochemistry: with clinical correlations (Edition 4th ed)*. Wiley-liss, New York, USA, pp: 225-278.
- Dodgson JB, Cheng HH, Okimoto R (1997) DNA marker technology: a revolution in animal genetics. *Poultry Science*, **76**, 1108-1114.
- Edfors-Lilja I, Gustafsson U, Duval-Iflah Y, Ellergren H, Johansson M, Juneja RK, Marklund L, Andersson L (1995) The porcine intestinal receptor for Escherichia coli K88ab, K88ac: regional localization on chromosome 13 and influence of IgG response to the K88 antigen. *Anim Genet*, **26**, 237-242.
- Edfors-Lilja I, Watrang E, Marklund L, Moller M, Andersson-Eklund L, Andersson L, Fossum C (1998) Mapping quantitative trait loci for immune capacity in the pig. *J Immuno*, **161**, 829-835.
- Edfors-Lilja I, Watrang E, Andersson L, Fossum C (2000) Mapping quantitative trait loci for stress induced alterations in porcine leukocyte numbers and functions. *Anim Genet*, **31**, 186-193.
- Ender A, Schwenk K, Städler T, Streit B, Schierwater B (1996) RAPD identification of microsatellites in Daphnia. *Mol Ecol*, **5**, 437-441.
- Erickson RP (1990) Post-meiotic gene expression. *Trends Genet*, **6**, 264-269.
- Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, *et al.* (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Res*, **10**, 220-227.
- Fujii H, Yoshida A (1980) Molecular abnormality of phosphoglycerate kinase-Uppsala associated with chronic nonspherocytic hemolytic anemia. *Proc Natl Acad Sci USA*, **77**, 5461-5465.
- Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, O'Brien PJ, MacLennan DH (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*, **253**, 448-451.
- Fujishima-Kanaya N, Toki D, Suzuki K, Sawazaki T, Hiraiwa H, Iida M, Hayashi T, Uenishi H, Wada Y, Ito Y, *et al.* (2003) Development of 50 gene-associated microsatellite markers using BAC clones and the construction of a linkage map of swine chromosome 4. *Anim Genet*, **34**, 135-141.
- Gartler SM, Riley DE, Lebo RV, Cheung MC, Eddy RL, Shows TB (1986) Mapping of human autosomal phosphoglycerate kinase sequence to chromosome 19. *Somat Cell Molec Genet*, **12**, 395-401.
- Geldermann H (1975) Investigations of quantitative characters in animals by gene markers I. *Methods Theor Appl Genet*, **46**, 319-330.
- Georges M, Andersson G, Braunschweig M, Buys N, Collette C, Moreau L, Nezer C, Nguyen M, Van Laere A, Andersson L (2003) Genetic dissection of an imprinted QTL mapping to proximal SSC2. *Proceedings of the Plant and Animal Genome XI Meetings, San Diego, CA, USA*, W327.
- Giese A, Jude R, Kuiper H, Piumi F, Schambony A, Guerin G, Distl O, Topfer-Petersen E, Leeb T (2002) Molecular characterization of the equine AEG1 locus. *Gene*, **292**, 65-72.
- Giraldo P, Montoliu L (2001) Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res*, **10**, 83-103.
- Giuffra E, Kijas JMH, Amarger V, Carlborg O, Jeon JT, Andersson L (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**.
- Goldberg E (1977) Isoenzymes in the testes and speratoza. *Isozymes: Current Top Biol Med Res*, **1**, 79-124.

- Goto M, Koji T, Mizuno K, Tamaru M, Koikeda S, Nakane P, Mori N, Masamune Y, Nakanishi Y (1990) Transcription switch of two phosphoglycerate kinase genes during spermatogenesis as determined with mouse testis sections in situ. *Exp Cell Res*, **186**, 273-278.
- Goto M, Tamura T, Mikoshiba K, Masamune Y, Nakanishi Y (1991) Transcription inhibition of the somatic-type phosphoglycerate kinase 1 gene in vitro by a testis-specific factor that recognizes a sequence similar to the binding site for Ets oncoproteins. *Nucleic Acids Res*, **19**, 3959-3963.
- Goto M, Masamune Y, Nakanishi Y (1993) A factor stimulating transcription of the testis-specific P<sub>gk</sub>-2 gene recognizes a sequence similar to the binding site for a transcription inhibitor of the somatic-type P<sub>gk</sub>-1 gene. *Nucleic Acids Res*, **21**, 209-214.
- Haley CS (1995) Livestock QTLs – Bringing home the bacon? *Trends in Genet*, **11**, 488-492.
- Hall JG (1997) Genomic imprinting: nature and clinical relevance. *Annu Rev Med*, **48**, 35-44.
- Hawken RJ, Murtaugh J, Flickinger GH (1999) A first-generation porcine whole-genome radiation hybrid map. *Mamm Genome*, **10**, 824-830.
- Hayden MJ, Sharp PJ (2001) Targeted development of informative microsatellite (SSR) markers. *Nucleic Acids Res*, **29**, E44-4.
- Hecht NB (1986) Regulation of the gene expression during mammalian spermatogenesis. In: Rossant J and Pedersen RA (eds), *Experimental approaches to mammalian embryonic development*. Cambridge university press, Cambridge, pp: 151-193.
- Hecht NB (1993) Gene expression during male germ cell development. In Desjardins C and Ewing LL (eds), *Cell and Molecular Biology of the testis*. New York, USA.
- Heyen DW, Weller JI, Ron M, Band M, Beever JE, Feldmesser E, Da Y, Wiggans GR, VanRaden PM, Lewin HA (1999) A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol Genomics*, **1**, 165-175.
- Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HHJr (1994) A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet*, **7**, 143-148.
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet*, **6**, 84-89.
- Karagoyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Res*, **21**, 3911-3912.
- Karp A, Edwards KJ (1997) DNA markers: a global overview. In: Caetano-Anolles G and Gresshoff PM (eds), *DNA markers: protocols, applications, and overviews*. WILEY-VCH, NY, USA, pp: 1-15.
- Kasahara M, Passmore HC, Klein J (1989) A testis-specific gene Tpx-1 maps between P<sub>gk</sub>-2 and Mep-1 on mouse chromosome 17. *Immunogenetics*, **29**, 61-63.
- Kazazian HHJr, Moran JV (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet*, **19**, 19-24.
- Kazazian HHJr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, **332**, 164-166.

- Kijas JM, Fowler JC, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *Biotechniques*, **16**, 656-662.
- Knott SA, Haley CS (1992) Aspect of maximum likelihood internal mapping in a F2 population. *Genetical res*, **60**, 139-151.
- Koizumi A, Hamade N, Arai M, Takatoku M, Yasuhiko W, Tsukada M, Kamiyama S (1990) Electrophoresis of phosphoglycerate kinase-2 to determine testicular damage induced by ethylene glycol monomethyl ether and sterility associated with chromosomal abnormality. *Arch Toxicol*, **64**, 181-187.
- Korstanje R, Paigen B (2002) From QTL to gene: the harvest begins. *Nat Genet*, **31**, 235-236.
- Kramer JM (1981) Immunofluorescent localization of PGK-1 and PGK-2 isozymes within specific cells of the mouse testis. *Dev Biol*, **87**, 30-36.
- Kramer JM, Erickson RP (1981) Developmental program of PGK-1 and PGK-2 isozymes in spermatogenic cells of the mouse: specific activities and rates of synthesis. *Dev Biol*, **87**, 37-45.
- Kumari M, Stroud JC, Anji A, McCarrey JR (1996) Differential appearance of DNase I-hypersensitive sites correlates with differential transcription of P<sub>gk</sub> genes during spermatogenesis in the mouse. *J Biol Chem*, **271**, 14390-14397.
- Laval G, Iannuccelli N, Legault C, Milan D, Groenen M, Giuffra E, Andersson L, Nissen PH, Jorgensen CB, Beeckmann P, et al. (2000) Genetic diversity of eleven European pig breeds. *Genet Sel Evol*, **32**, 187-203.
- Lay AJ, Jiang XM, Kisker O, Flynn E, Underwood A, Condron R, Hogg PJ (2000) Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature*, **408**, 869-873.
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet*, **44**, 397-401.
- Lunt DH, Hutchinson WF, Carvalho GR (1999) An efficient method for PCR-based identification of microsatellite arrays (PIMA). *Mol Ecol*, **8**, 893-894.
- Mackay TFC (2001) Quantitative trait loci in *Drosophila*. *Nat Rev Genet*, **2**, 11-21.
- Maestre J, Tchénio T, Dhellin O, Heidmann T (1995) mRNA retroposition in human cells: processed pseudogene formation. *EMBO J*, **14**, 6333-6338.
- Mann T (1964) *The biochemistry of semen and the male reproductive tract*. Wiley, New York, USA.
- Mas MT, Chen CY, Hitzeman RA, Riggs AD (1986) Active human-yeast chimeric phosphoglycerate kinases engineered by domain interchange. *Science*, **233**, 788-790.
- Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat Genet*, **8**, 384-390.
- McCarrey JR (1987) Nucleotide sequence of the promoter region of a tissue-specific human retroposon: comparison with its housekeeping progenitor. *Gene*, **61**, 291-298.
- McCarrey JR (1990) Molecular evolution of the human P<sub>gk</sub>-2 retroposon. *Nucleic Acids Res*, **18**, 949-955.
- McCarrey JR (1994) Evolution of tissue-specific gene expression in mammals. *Bioscience*, **44**, 20-27.
- McCarrey JR, Thomas K (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature*, **326**, 501-505.

- McCarrey JR, Berg WM, Paragioudakis SJ, Zhang PL, Dilworth DD, Arnold BL, Rossi JJ (1992a) Differential transcription of P<sub>gk</sub> genes during spermatogenesis in the mouse. *Dev Biol*, **154**, 160-168.
- McCarrey JR, Dilworth DD, Sharp RM (1992b) Semiquantitative analysis of X-linked gene expression during spermatogenesis in the mouse: ethidium-bromide staining of RT-PCR products. *Genet Anal Tech Appl*, **9**, 117-123.
- McCarrey JR, Kumari M, Aivaliotis MJ, Wang Z, Zhang P, Marshall F, Vandenberg JL (1996) Analysis of the cDNA and encoded protein of the human testis-specific PGK-2 gene. *Dev Genet*, **19**, 321-332.
- Meijerink E, Fries R, Vogeli P, Masabanda J, Wigger G, Stricker C, Neuenschwander S, Bertschinger HU, Stranzinger G (1997) Two alpha (1, 2) fucosyltransferase genes on porcine chromosome 6q11 are closely linked to the blood group inhibitor (S) and Escherichia coli F18 receptor (ECF18R) loci. *Mamm Genome*, **10**, 736-741.
- Michelson AM, Markham AF, Orkin SH (1983) Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. *Proc Natl Acad Sci USA*, **80**, 472-476.
- Michelson AM, Bruns GAP, Morton CC, Orkin SH (1985) The human phosphoglycerate kinase multigene family: HLA-associated sequences and an X-linked locus containing a processed pseudogene and its functional counterpart. *J Biol Chem*, **260**, 6982-6992.
- Mighell AJ, Smith NR, Robinson PA, Markham AF (2000) Vertebrate pseudogenes. *FEBS Lett*, **468**, 109-114.
- Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler K, Vogelstein B, Nakamura Y (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res*, **52**, 643-645.
- Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, *et al.* (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science*, **288**, 1248-1251.
- Mizuno K, Goto M, Masamune Y, Nakanishi Y (1992) A silencer like cis element for the testis specific phosphoglycerate kinase 2 encoding gene. *Gene*, **119**, 293-297.
- Moran JV, DeBerardinis RJ, Kazazian HHJr (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530-1534.
- Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nat Genet*, **25**, 381-384.
- Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M (1993) Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J Clin Invest*, **91**, 1859-1860.
- Nelson DL, Brownstein BH (1994) *YAC libraries: a user's guide*. Freeman WH, New York.
- Nishiyama Y, Masamune Y, Nakanishi Y (1994) The silencer of mouse P<sub>gk</sub>-2 consists of two discrete DNA elements that individually have no effect. *Gene*, **141**, 261-266.
- O'Brien SJ (1991) Molecular genome mapping: lessons and prospects. *Curr Opin Genet Dev*, **1**, 105-111.
- Ollivier L, Wrede J, Distl O (2001) An overview of the genetic resources of pigs and their management and conservation. In: Ollivier L, Labroue F, Glodek P, Gandini G and Delgado JV (eds), *Pig genetic resources in Europe, Characterisation and conservation*. EAAP: Wageningen, pp: 5-14.

- Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proc Natl Acad Sci USA*, **89**, 3419-3423.
- Paetkau D (1999) Microsatellites obtained using strand extension: An enrichment protocol. *Biotechniques*, **26**, 690-697.
- Paszek AA, Flickinger GH, Fontanesi L, Beattie CW, Rohrer GA, Alexander L, Schook LB (1998) Evaluating evolutionary divergence with microsatellites. *J Mol Evol*, **46**, 121-126.
- Pegoraro B, Lee CY (1978) Purification and characterization of two isozymes of 3-phosphoglycerate kinase from the mouse. *Biochim Biophys Acta*, **522**, 423-433.
- Petit RJ, Mousadik AE, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conserv Biol*, **12**, 844-855.
- Plastow G (2003) The changing world of genomics and its impact on the pork chain. *Advances in pork production*, **15**, 67-71.
- Prospst F, Rosenberg MP, van de Woude GF (1988) Proto-oncogene expression in germ cell development. *Trends Genet*, **4**, 183-187.
- Rassmann K, Schlötterer C, Tautz D (1991) Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis*, **12**, 113-118.
- Reiner G, Melchinger E, Kramarova M, Pfaff E, Buttner M, Saalmuller A, Geldermann H (2002) Detection of quantitative trait loci for resistance/susceptibility to Pseudorabies virus in swine. *J General Virology*, **83**, 167-172.
- Richards S, Bucan M, Brorson K, Kiefer MC, Hunt SW, Lehrach H, Lindahl KF (1989) Genetic and molecular mapping of the Hmt region of mouse. *EMBO J*, **8**, 3749-3757.
- Richardson T, Cato S, Ramser J, Kahl G, Weising K (1995) Hybridization of microsatellites to RAPD: a new source of polymorphic markers. *Nucleic Acids Res*, **23**, 3798-3799.
- Rogel-Gaillard C, Bourgeaux N, Save JC, Renard C, Coullin P, Pinton P, Yerle M, Vaiman M, Chardon P (1997) Construction of a swine YAC library allowing an efficient recovery of unique and centromeric repeated sequences. *Mamm Genome*, **8**, 186-192.
- Rogel-Gaillard C, Bourgeaux N, Billault A, Vaiman M, Chardon P (1999) Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. *Cytogenet Cell Genet*, **85**, 205-211.
- Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. *Genetics*, **136**, 231-245.
- Rohrer GA, Alexander LJ, Hu ZL, Smith TP, Keele JW, Beattie CW (1996) A comprehensive map of the porcine genome. *Genome Res*, **6**, 371-391.
- Rothschild MF (2003a) Approaches and challenges in measuring genetic diversity in pigs. *Arch Zootec*, **52**, 129-135.
- Rothschild MF (2003b) Advances in pig genomics and functional gene discovery. *Comparative and Functional Genomics*, **4**, 266-270.
- Rothschild MF, Ruvinsky A (1998) *The Genetic of the pig*. CAB International, Oxon, UK.
- Rothschild MF, Plastow GS (1999) Advances in pig genomics and industry applications. *AgBiotechNet*, **1**, 1-8.

- Ruane J (1999) A critical review of the value of genetic distance studies in conservation of animal genetic resources. *J Ani. Breed Genet*, **116**, 317-323.
- Rudolph JA, Spier SJ, Byrns G, Rojas CV, Bernoco D, Hoffman EP (1992) Periodic paralysis in quarter horses: a sodium channel mutation disseminated by selective breeding. *Nat Genet*, **2**, 144-147.
- Salisbury GW, Hart RG, Lodge JR (1977) The spermatozoon. *Perspe Biol Med Spring*, 372-393.
- SanCristobal M, Chevalet C, Foulley JL, Ollivier L (2003) Some methods for analysing genetic marker data in a biodiversity setting – example of the PigBioDiv data. *Arch Zootec*, **52**, 173-183.
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA*, **89**, 8794-8797.
- Shuster DE, Kehrlı ME, Ackermann MR, Gilbert RO (1992) Identification of prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc Natl Acad Sci USA*, **89**, 9225-9229.
- Singer EN, Armour JAL, Jeffreys AJ (1996) Detection of an MboI RFLP at the porcine clotting factor IX locus and verification of sex linkage. *Anim Genet*, **27**, 130.
- Singer-Sam J, Robinson MO, Bellve AR, Simon MI, Riggs AD (1990) Measurement by quantitative PCR of changes in HPRT, PGK-1, PGK-2, APRT, MTase, and Zfy gene transcripts during mouse spermatogenesis. *Nucleic Acids Res*, **18**, 1255-1259.
- Song GJ, Lee H, Park Y, Lee HJ, Lee YS, Seo JT, Kang IS (2000) Expression pattern of germ cell-specific genes in the testis of patients with nonobstructive azoospermia: usefulness as a molecular marker to predict the presence of testicular sperm. *Fertil Steril*, **73**, 1104-1108.
- Spelman RJ, Bovenhuis H (1998) Moving from QTL experimental results to the utilization of QTL in breeding programmes. *Anim Genet*, **29**, 77-84.
- Sun HS, Wang L, Rothschild MF, Tuggle CK (1998) Mapping of the natural-resistance associated macrophage protein (NRAMP1) gene in pigs. *Anim Genet*, **29**, 138-140.
- Suzuki K, Asakawa S, Iida M, Shimanuki S, Fujishima N, Hiraiwa H, Murakami Y, Shimizu N, Yasue H (2000) Construction and evaluation of a porcine bacterial artificial chromosome library. *Anim Genet*, **31**, 8-12.
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389-399.
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res*, **17**, 6563-6571.
- Tonin P, Shanske S, Miranda AF, Brownell AK, Wyse JP, Tsujino S, DiMauro S (1993) Phosphoglycerate kinase deficiency: biochemical and molecular genetic studies in a new myopathic variant. *Neurology*, **43**, 387-391.
- van Laere A, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, *et al.* (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, **425**, 832-836.
- VandeBerg JL (1985) The Phosphoglycerate kinase isozyme system in mammals: biochemical, genetic, developmental, and evolutionary aspects. *Isozyme: Curr Top Biol Med Res*, **12**, 133-187.

- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet*, **44**, 388-396.
- Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet*, **18**, 19-24.
- Weller J, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci*, **73**, 2525-2537.
- Werle-Schneider G, von Brevern MC, Sylla BS, Hollstein MC (1999) De novo retrotransposition of unbiased sequences in a human breast cancer cell clone. *Genes Chromosomes Cancer*, **26**, 84-91.
- Willard HF, Goss SJ, Holmes MT, Munroe DL (1985) Regional localization of the phosphoglycerate kinase gene and pseudogene on the human X chromosome and assignment of a related DNA sequence to chromosome 19. *Hum Genet*, **71**, 138-143.
- Willison K, Ashworth A (1987) Mammalian spermatogenic gene expression. *Trends Genet*, **3**, 351-355.
- Winterø AK, Fredholm M, Thomsen PD (1992) Variable (dG-dT)<sub>n</sub>(dC-dA)<sub>n</sub> sequences in the porcine genome. *Genomics*, **12**, 281-288.
- Woods-Samuels P, Wong C, Mathias SL, Scott AF, Kazazian HHJr, Antonarakis SE (1989) Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics*, **4**, 290-296.
- Wu C, Zhu S, Simpson S, de Jong PJ (1996) DOP-vector PCR: a method for rapid isolation and sequencing of insert termini from PAC clones. *Nucleic Acids Res*, **13**, 2614-2615.
- Wu K, Jones R, Danneberger L, Scolnik PA (1994) Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Res*, **22**, 3257-3258.
- Yerle M, Echard G, Robic A, Mairal A, Dubut-Fontana C, Riquet J, Pinton P, Milan D, Lahbib-Mansais Y, Gellin J (1996) A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. *Cytogenet Cell Genet*, **73**, 194-202.
- Yerle M, Pinton P, Robic A, Alfonso A, Palvadeau Y, Delcros C, Hawken R, Alexander L, Beattie C, Schook L, *et al.* (1998) Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet Cell Genet*, **82**, 182-188.
- Yerle M, Pinton P, Delcros C, Arnal N, Milan D, Robic A (2002) Generation of a 12,000 rad radiation hybrid panel for fine mapping in pigs. *Cytogenet Genome Res*, **97**, 219-228.
- Yisraeli J, Szyf M (1984) *DNA methylation: biochemistry and biological significance*. (eds), Razin A, Cedar H and Riggs AD. Springer, New York, USA, pp: 353-378.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol*, **11**, 1-16.

## 5 ACKNOWLEDGEMENTS

It was a great opportunity to perform my PhD study at the Institute of Veterinary Medicine (IVM) in Göttingen provided by **Prof. Dr. Dr. Bertram Brenig**. I wish to express my deepest thanks to my supervisor **Prof. Dr. Dr. Bertram Brenig** for his invaluable guidance, constructive suggestion and support throughout my studies, and to my advisor **Dr. Christoph Knorr** for his fruitful discussions, encouragement and always willing to help whenever need is as well as the meticulous corrections of the manuscripts. I would also like to thank my co-supervisor **Prof. Dr. Lusheng Huang** (Jiangxi Agricultural University, China) for his recommendation, valuable suggestions and kind help, and the **Jiangxi Agricultural University** for partly supporting my study.

I am indebted to **Prof. Dr. Henner Simianer** for providing valuable information and suggestion as well as correferent. I am also thankful for **Prof. Dr. Udo ter Muelen** and **Prof. Dr. Wolfgang Holtz** for kindly being my co-examiners.

I owe equally heartfelt appreciation to those who have contributed to this work. Thank you to everyone who has shared your experience with me, and for contributing the nice and pleasant atmosphere at the lab. I especially want to express my sincere gratitude to **Mrs. Alexandra Siebels** and **Miss Viola Raupach** for their excellent technical assistance, my co-author **Dr. Gerhard Moser** (Genetic Solutions Pty Ltd, Australia) for his statistical analysis between the *PGK2* gene and phenotypic traits, my co-author **Miss Julia Beck** for her *FISH* analysis and kind help for my computer accidents, my co-author **Mrs. Kirsten Bornemann-Kolatzki** for her collaboration in microsatellite analysis, my co-author **Miss Kesinee Gatphayak** for her contribution in gene expression, **Dr. Wilhelm Wemheuer** for kindly providing the sperm data of the AI boars, valuable information and helping in collecting tissue samples, **Miss Frauke Meyer** for inputting the huge amount of data, **Miss Monique Germerodt** for preparation of some of the DNA samples in this research, **Mr. Leonard Bull** for valuable discussion and linguistic revision, **Miss Monika Mokisz** for providing some useful protocols for RNA analysis.

Many thanks are due to **Dr. Denis Milan** (INRA, France), **Dr. Gary Rohrer** (USDA, USA), **Dr. Bin Fan** (Huazhong Agricultural University, China) for their useful suggestion and valuable discussion.

Last but certainly not least, I would like to say thank you for my wife **Song Zhang** for putting up with me when I was totally stressed, everlasting love, encouragement, support and everything great or small; for **my family in China** that always supported me during these years.

---

## 6 CURRICULUM VITAE

Family Name: Chen  
First Name: Kefei  
Date of Birth: 18 January 1974  
Place of Birth: Jiangxi, P. R. China  
Marital Status: Married  
Parents: Faming Chen and Fengqing Long  
Nationality: P. R. China

### Education and Professional Experiences

1980-1985 Primary School (Ning'gang, Jiangxi, P. R. China)  
1985-1989 Middle School (Ning'gang, Jiangxi, P. R. China)  
1989-1992 High School (Ning'gang, Jiangxi, P. R. China)  
1992-1996 Bachelor degree majoring in Animal Husbandry, Jiangxi Agricultural University, Nanchang, P. R. China  
1996-1999 Master degree majoring in Animal Genetics, Breeding and Reproduction, thesis "The Primary Evaluation for the Effects of Estrogen Receptor (*ESR*) – Follicular Stimulating Hormone beta (*FSH $\beta$* ) Haplotypes on Litter Size Traits in Pig" carried out in the National Key Laboratory for Agro-biotechnologies at the China Agricultural University, Peking, and Jiangxi Agricultural University, Nanchang, P. R. China  
1999-2002 Lecturer in the Faculty of Animal Science and Veterinary Technology, Jiangxi Agricultural University, Nanchang, P. R. China  
2002-2004 PhD study majoring in Molecular Genetics, the Institute of Veterinary Medicine, Georg-August-University Göttingen