Progress in porcine genomics

Chris Moran

Centre for Advanced Technologies in Animal Genetics and Reproduction (Reprogen), Faculty of Veterinary Science, University of Sydney, NSW 2006, Australia

Introduction

In this review, I will discuss the current status of porcine genomic studies and progress towards a complete, publicly available genome sequence for the pig. I will also describe an example of a successful European gene identification project which has proceeded all the way from broadly mapped quantitative trait locus (QTL) in the pig, through positional candidate gene identification to identification the mutation - the so-called quantitative trait nucleotide (QTN) - responsible for an effect on growth and muscle deposition. Finally I will briefly describe some recent work in my laboratory which has eliminated a candidate gene for androstenone-related boar taint.

Progress on porcine genome sequencing

1. The Danish Chinese collaboration

The Royal Veterinary and Agricultural University, the Danish Institute of Agricultural Sciences and the National Committee of Pig Breeding, Health and Production from Denmark and the Chinese Academy of Science and especially the Beijing Genomics Institute in China have been collaborating since 2001 on a major genome and EST sequencing project. This project explicitly aims to apply for, hold and exploit patents on sequences or partial sequences of genes and to transfer useful research results and technology into the Chinese and Danish pig industries. However all sequences are supposed to be publicly released after 6 months delay. At this stage, as far as I am aware, no sequences from this project have been released and, as far as can be assessed, progress on the genome sequence has been disappointingly slow. It is rumoured that a one or two pass genome sequence has been produced. Professor Lars Bolund in recent discussions at the Beijing Genomics Institute indicated that publication of and release of the genomic data may occur reasonably soon. So in summary, this porcine genome sequence is incomplete and unavailable.

2. Public Consortium sequencing project

At the 2004 International Society of Animal Genetics (ISAG) meeting held in Tokyo on September 10-16, progress and problems with the public pig genome sequencing effort were discussed. Unfortunately for reasons which are unclear, the pig appears to have fallen off the queue for the NIH/National Human Genome Research Institute supported program of mammalian genome sequencing, despite previous assurances that it would join the next round. This is puzzling when one considers the fact that the pig is the main source of animal protein consumed worldwide, is an excellent and widely used

biomedical model species, is likely to become of even greater biomedical significance if xenotransplantation becomes established as a medical therapy and has all the necessary resources in place for completion of a genome sequence including very detailed genetic maps and BAC contigs. Instead the NIH has recently committed substantial resources to sequencing a diverse array of other mammalian species, including the African savannah elephant (*Loxodonta africana*), the European common shrew (*Sorex araneus*), the European hedgehog (*Erinaceus europeaus*), the guinea pig (*Cavia porcellus*), the lesser hedgehog tenrec (*Echinops telfairi*), the nine-banded armadillo (*Dasypus novemcinctus*), the rabbit (*Oryctolagus cuniculus*), the domestic cat (*Felis catus*), since it is an important medical model for studying disease, and the orangutan (*Pongo pygmaeus*), as another primate closely related to humans. At this stage, the human, mouse, rat, chicken and dog genomes have been sequenced and the kangaroo, the cow, the grey short-tailed opossum are in the pipeline. Indeed the unannotated but complete bovine genome sequence was publicly released on October 7 2004.

It is believed that the NIH may have formed the opinion that alternative sources of funding would become available for generating a porcine genome sequence and for this reason, dropped it from its priority list, despite supporting the bovine genome sequence in the previous round. Kellye Eversole, the head of the Alliance for Animal Genome Research, a non-profit organization established to secure increased funding for livestock, poultry, and companion animal genomics research, attended the ISAG meeting and discussed ways for raising money to support a 3X BAC skim and 3X shotgun coverage of the porcine genome. Because of the excellent nature of the porcine genomic resources, sequencing minimally overlapping BAC clones from the pig would provide less redundant sequence than in other species and would therefore provide very useful coverage. Ms Eversole is looking for commitments from any possible source to help fund the porcine genome effort.

The porcine genome news at the ISAG meeting was not all bad. John Beever reported that by systematically mapping BAC end sequences, the IMpRH porcine radiation hybrid map has now been completed with only two unspannable gaps. Effectively this means that any novel sequence can be mapped to about 100 kb precision in any laboratory in the world with a PCR machine and access to the IMpRH panel DNA.

From QTL to QTN

Ideally all QTL mappers would like to identify the mutation(s) in a specific gene underlying the QTL effects they have identified. This would not only provide the best possible genetic test for exploiting the favourable allele in marker assisted selection but would also illuminate the biology underlying the biological effect, since the gene could then be related to a metabolic pathway or signalling system which might be amenable to alternative forms of manipulation to improve productivity. Quite a few examples of such successful molecular discovery in domestic animals are now available, initially for genes of large effect such as the double muscling phenotype in cattle (Grobert et al., 1997), the Booroola fecundity gene in sheep (Galloway et al. 2000) and the RN gene for excess glycogen content causing acidity in processed pork (Milan et al., 2000) but extending more recently to genes originally discovered in QTL scans such as the DGAT1 locus affecting fat content of milk in dairy cattle (Grisart et al., 2002).

1. Discovery of a QTL for muscle growth on pig chromosome 2

In 1999, back-to-back papers in *Nature Genetics* by Jeon et al. and Nezer et al. reported the independent identification in Sweden and Belgium of a QTL mapping to chromosome 2 with a substantial effect on muscle and fat deposition in the pig. It explained 15–30% of the phenotypic variation in muscle mass and 10–20% of the variation in back-fat thickness in the resource pedigrees in which it was mapped. The QTL was discovered in routine genome scans for QTLs for numerous traits being performed in many laboratories throughout the world at that time.

2. Imprinting a clue to QTL identity

However this QTL was unusual from the start since it was imprinted. The effect of the QTL alleles depended on whether they were inherited from the mother or the father. In this case, only the paternally inherited allele is expressed. Suspicion immediately fell on the insulin-like growth factor 2 (IGF2) locus, since comparative mapping indicated it lay in the relevant region of porcine chromosome 2. IGF2 was known to be similarly imprinted in humans, mice and pigs and on *a priori* biochemical grounds, IGF2 might be expected to have an effect on muscle growth. Amarger et al. (2002) conducted a detailed comparative analysis of the structure and properties of the IGF2 and adjacent loci in the pig in he lead up to the identification of the mutation.

Subsequent mapping refined the position of the QTL to an interval of 250 kb at the tip of chromosome 2, containing the IGF2 locus (Nezer et al., 2003). An intensive analysis of mutations in IGF2 and surrounding loci commenced. A number of animals were available with or without the growth-enhancing QTL allele, but there were numerous mutations both within IGF2 and surrounding genes to sift through. The objective was to find in these animals a DNA sequence variant which matched exactly the inheritance of the growth enhancing QTL allele – this was termed “mapping by haplotype sharing”. Van Laere et al. (2003) finally reported the full characterisation of the mutation in IGF2 responsible for the QTL effect in late 2003. This was extremely important and useful from a practical sense as it means that an accurate and direct test is now available for the favourable allele at this locus.

3. Intronic mutation responsible for growth enhancing QTL effect

However at the scientific level, it revealed something profoundly important as well. The mutation responsible for the effect did not lie in a protein encoding region of the gene as one might expect, nor did it lie in the 5” upstream regulatory region of the gene, where one might also expect an effect of transcription of the gene to influence the growth phenotype. Instead it was actually found to lie within intron 3 of IGF2. An intronic location had been deemed so unlikely that in fact the causative mutation was ignored for some time after it was first discovered and was not given serious consideration until the genetic evidence from analysis of haplotype sharing compelled it, having eliminated all other DNA sequence variants. It was also shown that the mutation prevents the binding of a protein factor believed to be a repressor of transcription. Pigs inheriting the mutation from their sire have a threefold increase in IGF2 messenger RNA expression in postnatal muscle, but not in other tissues, and this is believed responsible for the enhanced growth effect.

From the perspective of pig QTL mappers like myself, the outcome of these studies provides confirmation that we will eventually discover the genes affecting economically important traits in pigs. In the case of this QTL, the imprinted nature of the effect greatly assisted in homing in on a very good candidate locus. However it also illustrates that biology is complex and still poorly understood and that studies such as these will also discover new biological principles and new opportunities for interventions in manipulating growth and other characteristics.

Eliminating a candidate gene for a boar taint QTL

Since castration of male pigs is not routinely practiced in Australia, boar taint due to accumulation in fat tissues of a sex pheromone, androstenone (5α-androst-16-en-3-one), is a potentially serious problem. Boar taint is the undesirable off-odour and off-flavour emanating from cooked meat of the male pig (Brooks and Pearson, 1986). There are several causes, with androstenone usually being the major one. Androstenone is a steroid pheromone and is related to, and synthesised in parallel with, steroid hormones (Claus, 1979).

Bidanel et al. (1996) and Milan et al. (1998), using a Meishan x Large White F2 mapping resource, reported that a QTL for androstenone levels in entire males mapped to pig chromosome 7 in the approximate vicinity of the major histocompatibility complex (MHC). The biosynthesis of steroid hormones and pheromones is extremely complex with numerous enzymes involved and potential products. However one enzyme involved in this process, steroid 21 hydroxylase P450c21 (CYP21), was noted as a positional candidate, since it mapped to the MHC in the pig (Geffrotin et al.,1990; Chardon and Renard, 1999) as it also does in other species. The *CYP21* gene is about 3.36 kb in length and contains 9 introns and 10 exons (Burghelle-Mayeur et al. 1992)

Consequently several laboratories were keen to test CYP21 as a candidate and Mr Payam Arasta took this up as a PhD project in my laboratory. We used the US43 pedigree bred at Bunge (QAF) Meat Industries at Corowa as our mapping and candidate evaluation resource. This resource consists of 596 progeny in eight sire families with a total of 130 dams. 326 of the progeny had androstenone measurements. Two sire families had already provided significant evidence of heterozygosity for a QTL for androstenone levels on SSC7 (Kerr et al., unpublished) in the vicinity of the MHC in a similar position to the French reports. PCR primers were designed to amplify the CYP21 gene in four overlapping fragments of 1051 bp from position 458 to 1509, 914 bp from position 1344 to 2285, 940 bp from position 2164 to 3104 and 875 bp from position 2834 to 3709. These were then amalgamated. PCR products were amplified from four of the eight sires and analysed for the presence of single nucleotide polymorphisms (SNPs). Thirty six SNPs were identified in the four sires but all were either synonymous (i.e. did not change the amino acid sequence of the CYP21 protein) or were located in introns. There was some evidence from an analysis of potential transcripts using Genscan (Burge and Karlin, 1997) that the G allele at an intronic A/G SNP at position 2329 might favour an alternative transcript causing an in-frame insertion of 105bp. A TauI PCR-RFLP for this 2329A/G SNP was devised and all progeny and parents of the US43 resource were genotyped. Variance analysis of the 2329 G/A genotypes using SAS PROC GLM showed that segregation of the SNP in these sire families had no significant impact ($p=0.75$) on androstenone levels. Finally an RT-PCR assay designed to detect the alternative transcript was devised. This was

applied to boars from QAF Meat Industries of A/G genotype, but provided no evidence of the use of the alternative transcript.

We eventually concluded from this exhaustive sequencing analysis of four sires, including examination of an alternative transcript as another potential source of variation in CYP21 structure and activity, that we could confidently exclude CYP21 as the gene underlying the SSC7 androstenone QTL. Quintanolla et al. (2003) have similarly excluded CYP21 after sequencing the locus in two founder animals in their resource pedigree, one Meishan and one Large White, after finding no SNPs that caused amino acid substitutions. Therefore the search for the gene and the mutations underlying this QTL must continue in other genes in this region of SSC7. A complete porcine genome sequence will be invaluable in this and all other QTL identification tasks.

Conclusions

International efforts to complete the porcine genome sequence and make it publicly available are continuing. Despite the current delays, I anticipate that this task will be completed before the end of 2006 and possibly earlier. This will have a major impact on the way in which porcine genomics and genetic studies are performed and hopefully will lead to accelerated discovery of genes and mutations that have economically important effects. The path from QTL to QTN has already been travelled for the IGF2 effect on porcine muscle growth, illustrating both the immediate and obvious short-term benefits, namely an accurate genetic test, and the potential longer terms benefits from fundamental biological discoveries. In the case of the boar taint QTL on SSC7, we have been able to exclude an obvious candidate locus, and the task of going from QTL to QTN still lies ahead of us.

References

Amarger, V., Nguyen, M., Van Laere, A.S., Braunschweig, M., Nezer, C,. Georges, M. and Andersson, L. (2002) "Comparative sequence analysis of the insulin-IGF2–H19 gene cluster in pigs." *Mammalian Genome* 13: 388–398.

Arasta, P., Chen, Y., Zhang, Y., Kerr[R.] and Moran, C. "Elimination of CYP21 as a candidate gene for androstenone boar taint." (Manuscript in preparation)

Bidanel J.P., Milan D., Chevalet C. *et al.* (1996) "Chromosome 7 mapping of a quantitative trait locus for fat androstenone level in Meishan x Large White F2 entire male pigs." In: *Proceedings The European Association for Animal Production meetings*.

Brooks, R. I. and Pearson., A. M. (1986) "Steroid hormone pathways in the pig, with special emphasis on boar odor: A review." *Journal of Animal Science* 62: 632-645.

Burge, C. and Karlin, S. (1997) "Prediction of complete gene structures in human genomic DNA." *Journal of Molecular Biology* 268: 78 - 94.

Burghelle-Mayer, C., Geffrotin, C. and Vaiman, M. (1992) "Sequences of swine 21-hydroxylase gene and a portion of the opposite-strand overlapping gene of unknown function previously described in human." *Biochimica et Biophysica Acta-Bioenergetics* 1171: 153–161.

Chardon P., Renard C. and V. M. (1999) "The major histocompatibility complex in swine." *Immunological Review* 167: 179-192.

Claus, R. (1979) "The boar taint as model of a steroidal pheromone." *Acta Endocrinologica* 91, 432-433.

NIH/National Human Genome Research Institute (2004) "NHGRI adds 18 organisms to sequencing pipeline." NIH News Release 5 August 2004. http://www.genome.gov/12511858

Galloway, S.M., McNatty, K.P., Cambridge, L.M., Laitinen, M.P., Juengel, J.L., Jokiranta, T.S., McLaren, R.J, Luiro, K., Dodds, K.G., Montgomery, G.W., Beattie, A.E., Davis, G.H. and Ritvos, O. (2000) "Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner." *Nature Genetics.* 25: 279-283.

Geffrotin, C., Chardon, P., De Andres-Cara, D.F., Feil, R., Renard ,C. and Vaiman M . (1990) "The swine steroid 21-hydroxylase gene: Cloning and mapping within the swine leukocyte antigen complex." *Animal Genetic* 21: 1-13.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. and Snell, R.(2002) "Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition." *Genome Research.* 12: 222-231.

Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D,. Brouwers, B., Riquet, J,. Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., Fries, R., Hanset, R. and Georges, M. (1997) "A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle." *Nature Genetics.* 17: 71-74.

Jeon, J.-T, Carlborg, O., Tornsten, A., Giuffra, Amarger, V., Chardon, P., Andersson-Eklund, L., Andersson, K., Hansson, I., Lundstrom, K. and Andersson L. (1999) "A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus." *Nature Genetics* 21: 157–158.

Kerr, R., Chen, Y., Moran, C., Graser, H. and Luxford, B. (2001) *Marker assisted selection for profitable pigs*. Australian Pork Limited Confidential Research Report. Project Number: US43. 224 pp.

Milan, D., Bidanel, J.P., Le Roy, P. *et al.* (1998) "Current status of GTL detection in Large White x Meishan crosses in France." *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production* 26: 414-417.

Milan, D., Jeon J.T., Looft, C., Amarger, V., Robic, A., Thelander, M., Rogel-Gaillard, C., Paul, S., Iannuccelli, N., Rask, L., Ronne, H., Lundstrom, K., Reinsch, N., Gellin, J., Kalm, E., Roy, P.L., Chardon, P. and Andersson L. (2000) "A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle." *Science.* 288: 1248-1251.

Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., Karim, L., Kvasz, A, Leroy, P. and Georges M. (1999) "An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs." *Nature Genetics* 21: 155–156.

Nezer, C., Collette, C., Moreau, L., Brouwers, B., Kim, J.J., Giuffra, E., Buys, N., Andersson, L. and Georges, M. (2003) "Haplotype sharing refines the location of an imprinted QTL with major effect on muscle mass to a 250 Kb chromosome segment containing the porcine IGF2 gene." *Genetics* 165: 277–285.

Quintanilla, R., Demeure, O., Bidanel, J.P., Milan, D., Iannuccelli, N., Amigues, Y., Gruand, J., Renard, C., Chevalet, C. and Bonneau, M. (2003) "Detection of quantitative trait loci for fat androstenone levels in pigs." *Journal of Animal Science.* 81: 385-394.

Van Laere, A-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M. and Andersson, L.. (2003) "A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig." *Nature* 425: 832-836.