Sponsors

University of Minnesota
College of Veterinary Medicine

College of Agricultural, Food and Environmental Sciences

Extension Service

Swine Center

Use of molecular genetics to influence production

L. Ollivier

INRA-SGQA, 78352 Jouy-en-Josas cedex, France

Introduction

Recent developments in molecular genetics have led to a spectacular increase in the information available on the genome of our domestic species, and, consequently, new ways to influence production can now be seen. Genetic maps in particular have remained rudimentary until the early 1990s. Since then, the identification of genetic markers based on DNA sequence variation has resulted in a rapid increase in genome coverage in most species. This has also been the case for the pig genome, where several hundred genetic landmarks have up to now been identified and provide an almost complete genome coverage.

To evaluate the contribution of molecular genetics to pig improvement, it may be useful to refer to the general basis of sound breeding plans. These are based on the optimal exploitation of *reproductive capacity*, the efficient use of *breeding value evaluation* tools, and an adequate management of *genetic diversity*. Molecular genetics, as we shall see, is expected to contribute most to the efficiency of selection and to the management of genetic variability.

The pig genome: An overview

There are various ways to describe the genome of a given species. If we limit ourselves to the genome included in the cell nucleus, the nuclear genome, three ways of describing it and even measuring it exist. Each method implies the use of a specific instrument and a corresponding unit of measurement.

The *cytogenetic length* was the first known measurement, from visual observation of the chromosomes at particular stages of cell division. The measurement instrument is the microscope and the unit is the micron ($1\mu=10^{-6}$ meter). The exact number of pig chromosomes is known to be 2n=38, and, with the advent of the banding techniques in 1970, a precise identification and measurement of each chromosome was made possible.

The *physical length* of the pig genome has been evaluated much more recently. It is based on a modern technique known as flow cytometry. The instrument used is a cell sorter, and the unit is the base pair (bp), or a multiple of it such as a megabase = 10^6 bp. Bps are the elementary units of the DNA molecule, and their arrangement determines the hereditary message specific to each individual and transmitted from one generation to the next. By using a cell sorter, which is able to separate the chromosomes according to their DNA content, it can be said that the genome weight is being estimated, rather than its size. Actually, the pig genome weighs about 2.7×10^9 bp, not too far below the 3.2×10^9 bp of the human genome.

The *genetic length* is obtained by counting recombination events between pairs of genes. This has been done by geneticists since the early times of genetics, in various types of crosses. Distances between genes are obtained by using a scale based on recombination percentage, the unit being the centimorgan (cM) corresponding to one percentage point of recombination. Contrary to physical or cytogenetic tools of measurement, which cover the whole genome, the genetic tool only applies to genes which can be identified on the genome, and genetic length is therefore closely dependent on the number of genes known. Genome length consequently tends to increase over time, as more and more genes become known. The genetic length of the human chromosome 1 has increased by nearly 60 percent between 1976 and 1991. A similar evolution, however in a shorter period of time, has been observed in farm animal species. In the pig, for example, the genome length evaluated in 1996 was 22% above a previous estimate of 1994. The estimated length in 1996 was about 2300 cM, markedly below the human genome evaluated to be about 3800 cM long.

The correspondence between those three evaluations of genome size is somewhat loose and there is no strict proportionality between the three scales considered. One reason in that far from all DNA specifies development information. In fact, the largest part of DNA is not expressed (so-called repetitive DNA) and supposedly has no functional role. The total number of functional genes (or single-copy DNA) is evaluated to be somewhere between 50,000 and 100,000. A complete identification of these genes would in itself be a formidable task. In humans, which is the best known species, only about 5% of them have so far been shown to be involved in a recognized function. In the pig, as in the other farm animal species, molecular geneticists have to restrict their ambition to the study of a very limited number of genes "useful" in production. It should be added that the extranuclear genome (or mito-

chondrial genome) is also under current investigation. A recent account of the status of pig molecular genetics has been provided by C. Moran.[1]

Gene mapping: Linkage map and cytogenetic map

The first objective of gene mapping is to assess the position on the genome of each individual gene relatively to a given set of identified genes. The respective positions of the genes are represented on a linkage map, using a scale expressed in centimorgans (cM). As recombination events can only be observed between loci showing some genetic variation (polymorphic loci), linkage maps are most efficiently established from crosses between breeds genetically very different from each other, such as Chinese breeds or Wild Boar against western breeds used in the recent mapping efforts in Europe (PiGMaP Consortium) and the USA (USDA-MARC).

Cytogenetic maps are derived from various laboratory techniques such as *in situ* hybridization or hybrid cell panels. These techniques do not require observing recombinations between genes and therefore allow assigning chromosomal positions to genes which are not polymorphic. The present situation of the linkage and cytogenetic maps of the pig has been recently reviewed.[2,3] The number of points (loci) located on the pig linkage map now exceeds 1800, and more than 600 loci are cytogenetically located.[4]

Genes of interest in production: Monogenic and polygenic traits

Only a limited fraction of the genes (or DNA variation) in the pig genome is of interest to the pig producer. For many years, the only genes of interest were those responsible for visible traits, usually monogenic, such as color, morphology, or abnormalities. In contrast, performance traits, which vary in a continuous manner, were assumed to be jointly determined by several genes impossible to identify individually. Such traits, called polygenic, could however be acted upon through quantitative genetics methods. In fact, genetic improvement still relies essentially on selection and crossbreeding schemes based on the polygenic model of inheritance. The precise evaluation of the number of genes responsible for a quantitative trait is difficult because only the total genetic variation can be evaluated and individual genes vary in the sizes of their effects. Though in some instances large effects of individual genes, called *major genes*, can be evidenced (e.g. the halothane gene for lean content), the general rule is that of several genes of small effects, not individually measurable—so-called *minor genes*.

With the development of molecular genetics, a new category of genes, called *marker genes*, has entered the scene, and should now attract the attention of the pig producer. As indicated previously, the number of those markers positioned on the pig genome is close to 2000. Most of them belong to the category of "anonymous" microsatellites, made of short DNA sequences repeated a variable number of times. **Table 1** summarizes the properties of the marker genes as compared to the genes responsible for quantitative traits, or quantitative trait loci (QTL).

QTL mapping is progressively resolving the question so far unanswered about the actual number of genes implied in a quantitative trait and of their relative importance. By using a network of markers covering the entire genome, it is now possible to position QTL on the genome and simultaneously to estimate their effects. However, large numbers of individuals are needed for detecting QTL and the chances of detection closely depend on the importance of the QTL effect. Genes with small effects will remain undetected. However QTL mapping allows us to detect gene effects which could not have been detected otherwise. QTL mapping may thus be compared to a magnifying-glass, displacing the limit between major and minor genes towards a progressively lower and lower limit. In fact the distinction is no more to be made between major and minor genes, but rather between detected and undetected genes. **Table 2** gives some examples of

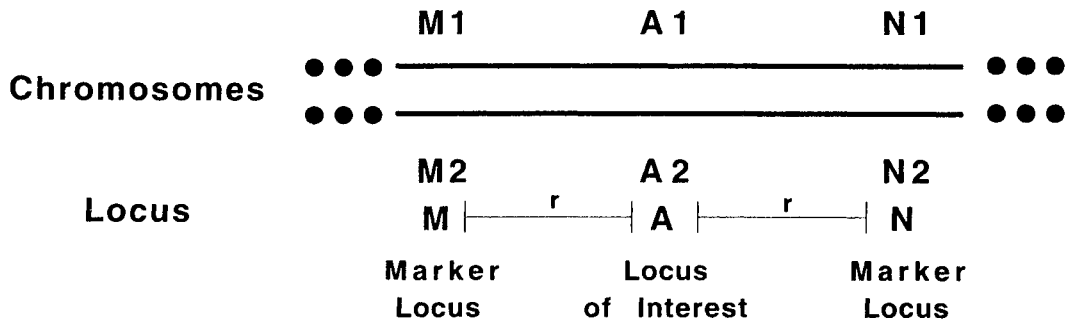
TABLE 1: Comparative properties of two categories of loci

Gene property	Quantitative trait locus (QTL)	Marker locus
Primary effect	Unknown	type I: actual gene (coding sequence) type II: anonymous marker (noncoding sequence)
Effect on quantitative trait	Major or minor	No effect
Genotype identification	Impossible	Possible by laboratory test
Genome location	Indirectly obtained (via QTL detection)	Directly obtained (via gene mapping)

TABLE 2: QTL detected in pigs as compared to other species

Trait	Nb of QTL	Percent of phenotypic variance explained	Population investigated
Pig litter size[4]	2	1	Synthetic lines
	4	2	Large White breed
Pig growth[4]	1	12	F_2 (Wild Boar X Large White)
	2	10	F_2 (Meishan X Large White)
Mouse growth[18]	17	76	F_2 (inbred lines)
Maize yield[19]	6,8	60	F_3, F_4, and BC (inbred lines)
Human diabetes[19]	18	ND	Sibpair families

Figure 1



the QTL presently detected in the pig, compared to other species. It shows that QTL mapping in the pig is probably still in its infancy.

Application of gene mapping results to improve production efficiency: Gene hitchhiking

Theoretical investigations were initiated in the early 1960s on the role that individual marker identification could play in breeding schemes, compared to classical quantitative genetic methods relying only on measurements of the traits of interest.[5] However, the use of genetic markers in breeding schemes so far has been very limited, since the genome was not property covered. The chances of finding a marker closely linked to a gene of interest were then very low, which is no more the case now with relatively dense linkage maps. It can now be safely expected that markers sufficiently close to any gene of interest will eventually be found.

Linkage situations such as the one schematically in **Figure 1** are at the basis of marker utilization for genetic improvement. The two marker loci M and N bracket a gene of interest (A), and it is assumed that A is located half-way between M and N at the same distance r (expressed in cM) from each of the two markers.

When selection is applied to the locus M, the situation above leads to what is called a "hitchhiking effect" of M on A. This means that gametes carrying M_1 will also carry A_1, unless a recombination has occurred between M and A. When the distance (r) between M and A is short, most M_1 gametes will also carry A_1, in fact a proportion 1-r, and the A_1 gene is strongly associated with M_1 in the transmission to the following generation. The reasoning assumes that in the whole population considered, M_1 is constantly—or at least predominantly—associated with A_1. Such an association is measured by a coefficient of linkage disequilibrium (D), which may be thought of as a correlation between the two loci. D can vary from 0, i.e. linkage equilibrium, to 1 (or -1), i.e. complete linkage disequilibrium. If M1N1 gametes are selected, the association with A_1 will be enhanced. A proportion $1-r^2$ of the gametes will indeed carry A_1, since a double recombination, with probability r^2, is needed to separate A_1 from the M1N1 segment of chromosome. Thus 2 markers 20 cM apart have the same hitchhiking strength as a single marker located at 1 cM from the gene (since $1-0.10^2=1-0.01$).

Given that 20 cM is the resolution already achieved for the porcine linkage map, marker-assisted selection (MAS) along the mechanism described above has a bright future. Three categories of applications can be considered:[6] markers for single-locus traits, markers for polygenic traits, and markers for the whole genome.

Marker-assisted breeding

Single-locus traits or major genes

The objective, then, is to change gene frequency at a given locus by selecting marker genes at another locus linked to the former one. The fate of the locus of interest de-

TABLE 3: Evolution of the frequency of a recessive gene over 10 generations from direct selection or marker selection under complete or incomplete linkage disequilibrium[6]

Generation	Direct selection[1]	D[3]	Marker selection[2]			
			1		0.5	
		d[4]	1 cM	10 cM	1 cM	10 cM
0	0.20		0.20	0.20	0.20	0.20
5	0.10		0.01	0.01	0.10	0.11
10	0.07		0.00	0.01	0.10	0.11

[1]Recessive homozygotes eliminated
[2]Elimination on marker genotypes: homozygotes and half of the heterozygotes
[3]D=linkage disequilibrium
[4]d=distance between gene and marker

pends on its being hitchhiked by the marker gene. The change of gene frequency can be predicted by standard 2-locus selection theory.[7] **Table 3** shows an application to the elimination of a recessive gene. It can be seen that marker selection is more efficient than direct selection of the gene when the initial D is high, and that marker selection becomes inefficient in the long run when D is low. **Table 3** also shows that strong initial D is more important than tight linkage for efficient MAS.

Additional advantages of MAS result from the possibility of selecting both sexes for sex-limited traits, or of selecting earlier for traits expressed late in life. Meat quality traits, for which major genes exist,[8,9] and also genetic resistances to various forms of diarrhea[10,11] or immune response traits,[4] are areas of interest for future MAS in pigs.

It should also be noted here that the assistance provided by genetic markers had already been illustrated in the 1980s by the use of biochemical markers, such as glucose phosphate isomerase (GPI), closely linked to the halothane locus, for eliminating the susceptibility allele at that locus. This example also shows that the advantage of MAS depends on the knowledge at hand about the locus of interest. When the locus is perfectly identified, as is the case presently for the halothane locus, there is obviously no need to select on flanking markers.

Another use of markers is for introgressing individual genes from one breed into another. Marker-assisted introgression (MAI) has, for instance, been implemented recently for transferring the halothane-resistance gene from a Large White "donor" breed into the halothane sensitive Piétrain breed.[12] By typing the closely linked GPI locus mentioned before, a halothane-negative Piétrain strain was obtained after three backcrosses. In the future, a potential application of MAI could be to transfer prolificacy genes from highly prolific Chinese breeds into western breeds in a similar fashion.

Polygenic traits

The additional information brought in by the markers allows a more accurate selection for polygenic traits. MAS combines the classical selection index for quantitative trait (I_Q) and the marker information summarized by a molecular score taken as the sum of the effects of the marker genes on the traits considered (I_M). The MAS criterion is then a linear combination I = b1IQ+b2IM which maximizes the expected response. The accuracy of I can be shown to depend only on the accuracy of the I_Q index and on the proportion of additive genetic variance associated with the marker loci contributing to I_M.[5,13] **Table 4** shows the gain in selection accuracy for three typical situations in pig breeding. A general feature of MAS is that the expected gains are highest for traits of low heritability. In that respect, growth and carcass traits are certainly not priorities in the efforts needed to implement MAS. The predictions of **Table 4** imply the existence of linkage disequilibria between the markers and the QTL involved. As the values of D regularly decrease over generations, so does the accuracy of the marker information. Periodical re-examinations of marker-QTL associations are therefore required in order to maintain the advantage of MAS over phenotypic selection.

When linkage disequilibria do not exist, marker-QTL associations cannot be exploited and marker information is then used in a different way. The gains from MAS in such situations results from a more efficient use of the information from relatives, owing to an improved evaluation of co-ancestry coefficients for the region of the genome carrying the QTL involved.[14] The gain in accuracy in that case is considerably less than in the situations considered in **Table 4**. A route similar to that followed in dairy cattle breeding granddaughter design could also apply to pig breeding when AI is highly developed. In such a scheme, linkages between markers and litter size could be detected and then exploited by preselecting grandoffspring of sons of widely used boars on the basis of the appropriate marker genotypes.[15] However, given the present structure of the breeding programs, with a relatively large number of boars

TABLE 4: Relative accuracy of one generation marker-assisted selection compared to standard phenotypic selection for various traits in pigs, according to the proportion of additive genetic variance associated with the marker loci[6]

Trait	Litter size	Meat quality	Carcass lean
Heritability	0.10	0.30	0.50
Accuracy of standard selection	0.30	0.36	0.78
Relative accuracy of MAS			
m=0.10	1.40	1.30	1.03
m=0.30	1.98	1.76	1.11
m=0.50	2.44	2.14	1.21

being used, it is unlikely that extra responses of the same order as those expected in dairy cattle can be achieved.

The whole genome

The assistance that markers can provide in introgressing a gene from one breed into another has been previously discussed. When the donor breed is inferior for all but the gene introgressed, there is a need to recover as quickly as possible the genomic background of the recipient breed during the repeated backcrossing scheme. Markers can be used to accelerate such a recovery by selecting individuals for maximum genomic similarity to the recipient line, a process called *genomic selection* (GS). GS should be carried out at least during three backcross generations. It is advantageous to perform GS on evenly spaced markers across the genome, such as microsatellites, compared to markers of unknown chromosomal location, such as DNA fingerprints. An optimal density of 2–3 markers per Morgan is recommended, which would mean between 50 and 80 markers for the pig genome. Under such a scheme, in order to recover 98% of the recipient genome, a gain of about 2 generations is achieved by using GS, when a proportion of 10% is selected on the markers at each generation.[16] It should be noted that a similar breeding program could be implemented for re-establishing a breed from a panel of frozen semen, if no female of the breed were available, as for instance if the breed had become extinct. GS would thus have some potential also in genetic conservation.

Heterozygosity can be directly measured for a set of marker loci and used as an indicator of the heterozygosity of the whole genome. As heterosis in a cross is proportional to the degree of heterozygosity, it should in theory be possible to predict heterosis on such a basis. One would expect the various crosses among a set of parental lines or breeds to rank in accordance with the corresponding genetic distances, for traits showing heterosis. This was confirmed in an early German study showing significant correlations between crossbred performances for heterotic traits such as litter size and survival, among a set of pig lines, and genetic distances based on 12 biochemical loci. Similar investigations would certainly deserve being pursued using the molecular markers recently developed. In a similar way, *homozygosity* can be assessed directly from markers and used to estimate realized inbreeding at the individual level. Genetic drift in a population of limited size might thus be counteracted by selecting the least inbred animals at each generation. For instance, with an effective population size of 50 the homozygosity normally reached after 25 generations could be delayed until 40 generations under such a selection scheme.

More generally, genetic markers are useful indicators of *breed diversity* and therefore should be considered in the management of the pig genetic diversity and in the definition of sound conservation programs. The evaluation of genetic diversity is based on the concept of genetic distance, which can be determined from gene frequency differences between breeds. Genetic distances can be evaluated with an accuracy which depends on the number of loci considered, on their location throughout the whole genome, on their heterozygosity, and on the number of individuals genotyped. It happens that the genetic markers presently available in the linkage maps, namely microsatellites, satisfy the above mentioned criteria, as they all are highly polymorphic and evenly distributed over the whole genome. It is considered that, in a preliminary stage, a set of 25 microsatellites typed on 25 individuals from each breed should be sufficient to provide reliable estimates of genetic distances.[17] A program for evaluating the genetic diversity of about 60 different breeds along those lines is now underway in the European Union.

Conclusion

Even though numerous examples of very successful improvement plans exist, there is still scope for designing more efficient breeding schemes than those generally implemented so far. Opportunities will undoubtedly derive from combining the knowledge which has accumulated on classical, quantitative and molecular genetics. Specific challenges have to be faced for some traits, such as reproductive ability and quality of lean and fat tissue, and here much hope is founded on a better knowledge of the genome and particularly the identification of useful

genetic markers. In this area, very bright prospects can be seen from the studies realized in other species.[18,19] However, many uncertainties still remain as to the real value of marker-assisted breeding in future pig improvement programs. Difficulties come from a lack of precision in the essential parameters, such as marker effects and their distribution in various breeds. Prospects depend on the state of marker technology. An evolution towards high density genetic maps makes it possible that markers sufficiently close to any gene of interest will eventually be found and then used successfully for genetic improvement. In any case, the building up of linkage maps in the pig is a well-engaged enterprise, from which various kinds of benefits, some unknown at present, may be expected. The applications can be foreseen as complementary to present conventional breeding plans, which will essentially continue. A well-balanced approach, taking into account all opportunities, will remain advisable in any genetic improvement scheme.

References

1. Moran C. Molecular Genetics. In: Rothschild MF and Ruvinsky A, eds. *The Genetics of the Pig*. Wallingford, Oxon, UK: CAB International; 1998: 135–161.
2. Archibald AL and Haley CS. Genetic Linkage Maps. In: Rothschild MF and Ruvinsky A, eds. *The Genetics of the Pig*. Wallingford, Oxon, UK: CAB International; 1998: 265–294.
3. Chowdhary BP. Cytogenetics and Physical Chromosome Maps. In: Rothschild MF and Ruvinsky A, eds. *The Genetics of the Pig*. Wallingford, Oxon, UK: CAB International; 1998: 199–264.
4. Rothschild MF. Recent Advances in Pig Genome mapping. In: *Proceedings Symposium Series 1*. The 8th World Conference on Animal Production. Seoul, Korea; 1998: 186–192.
5. Neimann-Sorensen A and Robertson A. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agriculturae Scandinavica*. 1961; 11: 163–192.
6. Ollivier L. Genetic Improvement of the Pig. In: Rothschild MF and Ruvinsky A, eds. *The Genetics of the Pig*. Wallingford, Oxon, UK: CAB International;. 1998: 511–540.
7. Karlin J General two-locus selection models: some objectives, results and interpretations. *Theoretical Population Biology*. 1975; 7: 364–398.
8. Milan D et al. Accurate mapping of the "acid meat" RN gene on genetic and physical maps of pig chromosome 15. *Mammalian Genome*;. 1996; 7: 47–51.
9. Janss LLG et al. Identification of a single gene affecting intramuscular fat in Meishan crossbreds using Gibbs sampling. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*. 1994; 18: 361–364.
10. Guérin G et al. Evidence for linkage between K88ab, K88ac intestinal receptors to Escherichia coli and transferrin loci in pigs. *Animal Genetics*. 1993; 24: 393–396.
11. Meijerink E et al. Genetic mapping in Swiss Landrace pigs of the gene specifying receptors for F18 fimbriated Escherichia coli strains causing oedema disease and post-weaning diarrhoea. *25th International Conference on Animal Genetics*. 1996; 128.
12. Hanset R et al. Effets de l'introgression dans le génome Piétrain de l'allèle normal au locus de sensibilité à l'halothane. *Genetics Selection Evolution*. 1995; 27: 77–88.
13. Lande R and Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*. 1990; 124: 743–756.
14. Ollivier L. The accuracy of marker-assisted selection for quantitative traits within populations in linkage equilibrium. *Genetics*; 1998: 148: 1367–1372.
15. Visscher PM and Haley CS. Utilizing genetic markers in pig breeding programmes. *Animal Breeding Abstracts*. 1995; 63: 1–8.
16. Hospital F et al. Using markers in gene introgression breeding programs. *Genetics*. 1992; 132: 1199–1210.
17. Barker JSF. A global protocol for determining genetic distances among domestic livestock breeds. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*. 1994; 21: 501–508.
18. Cheverud et al. Quantitative trait loci for murine growth. *Genetics*. 1996; 142: 1305–1319.
19. Falconer DS and Mackay TFC. Introduction to Quantitative Genetics. Harlow, UK; Longman. 1996.