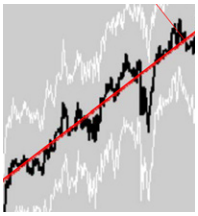
Genomic selection in commercial pig breeding



Egbert F. Knol,* Bjarne Nielsen,† and Pieter W. Knap‡

*Topigs Norsvin, Beuningen, the Netherlands

†SEGES Videncenter for Svineproduktion, Copenhagen, Denmark

‡Genus-PIC, Schleswig, Germany

Implications

- Leading pig breeding companies have implemented single-step evaluation for genomic selection. Overall, this increases EBV accuracies by half.
- Further improvement requires focus on trait- and line-specific QTLs, exploitation of crossbred performance for non-additive genetic effects, and training for hard-to-measure traits.
- The multi-breeding-company multi-line crossbred pig production structure limits very high accuracies. Increased genotyping and phenotyping will lead to improved training data and may increase accuracies by two-thirds rather than half.
- Novel technologies will allow for genotyping all selection candidates, reducing the generation interval and emphasizing the need for inbreeding control, more efficient breeding structures, and higher selection intensities.

Key words: DNA marker, genome-wide prediction, industry, SNP, swine

Introduction and Outline

Genomics covers the use of DNA technology and is generally applicable to the whole range of bacteria, yeast, plants, vertebrates, livestock species, and humans. Genetic improvement in livestock species aims at increasing sustainability and efficiency of animal products. Pig production centers around the use of crossbred animals (just like laying hens and broiler chickens) and, in that sense, differs from dairy cattle, where emphasis still tends to be on improvement of purebred performance.

In this paper, we give a narrative on genomic selection in commercial pig breeding to describe the current status and potential future and to highlight differences with other species; some examples will clarify status and developments.

Historical Developments

Marker-assisted selection

DNA technology was first applied in commercial pig breeding in the early 1990s when the Hal-1843 marker test became available for marker-assisted selection (**MAS**) against a recessive mutant allele that causes malignant hyperthermia (which is often lethal) in stressful conditions, and poor meat quality. This first marker test quickly became immensely popular for three reasons. First: susceptibility to this disorder in the pig is controlled by this single gene only. Second: conventional selection against this disorder involved halothane testing and is therefore labor-intensive. Third: because the target allele is recessive, conventional selection involves progeny testing and is therefore slow and expensive. In other words, Hal-1843 provided a simple and 100% effective alternative to an expensive and cumbersome procedure. Another early DNA marker linked to meat quality is the PRKAG3 locus (Leroy et al., 1990), which influences curing and cooking yield (*Rendement Napole*) particularly in Hampshire pigs.

A few other single-gene traits, such as resistance to *Escherichia coli*-induced diarrhea, got their own DNA markers with similarly successful results. Most of them are still in use.

But most livestock production traits are of a quantitative nature (i.e., not all-or-nothing events) and are controlled by large numbers of genes that were dubbed quantitative trait loci (QTLs) by Geldermann (1975). The first available QTL marker in pigs (Rothschild et al., 1996) targeted an estrogen receptor locus and was found to control roughly 12% of the phenotypic standard deviation of litter size. Several other major genes followed, but the conventional notion of the infinitesimal model in animal breeding (Fisher, 1930) was quickly confirmed: most QTL have only small effects on the traits of interest, and the probability of ever finding them is strongly constrained by experimental sample size (e.g., Hayes and Goddard 2001; Visscher 2008). As a consequence, the initial vision of getting the main breeding goal traits under control via MAS on their underlying QTLs had to be abandoned.

Marker-assisted BLUP

Instead, since the late 1990s, DNA marker genotypes were included into the conventional BLUP analyses following Fernando and Grossman (1989): add the marker genotype (0, 1, or 2, for an animal) as a fixed effect to the statistical model for a trait, obtain the BLUP solutions for the additive polygenic effect as before, and also obtain the properly adjusted BLUE solution for the marker's allele substitution effect; multiply this BLUE by 0, 1, or 2 (specific for the animal) and add the result to the animal's BLUP to obtain its final marker-enhanced EBV. A logical next step was to treat the marker genotypes as semi-random effects, making use of several different shrinkage strategies all based on the marker heritability (e.g., Tsuruta et al., 2001); by 2007, breeding value estimation packages such as PEST (Neumaier and Groeneveld, 1998) supported this strategy as part of their internal calculations. At that time, a typical genetic evaluation run for a production trait would involve up to 30 markers.

doi:10.2527/af.2016-0003

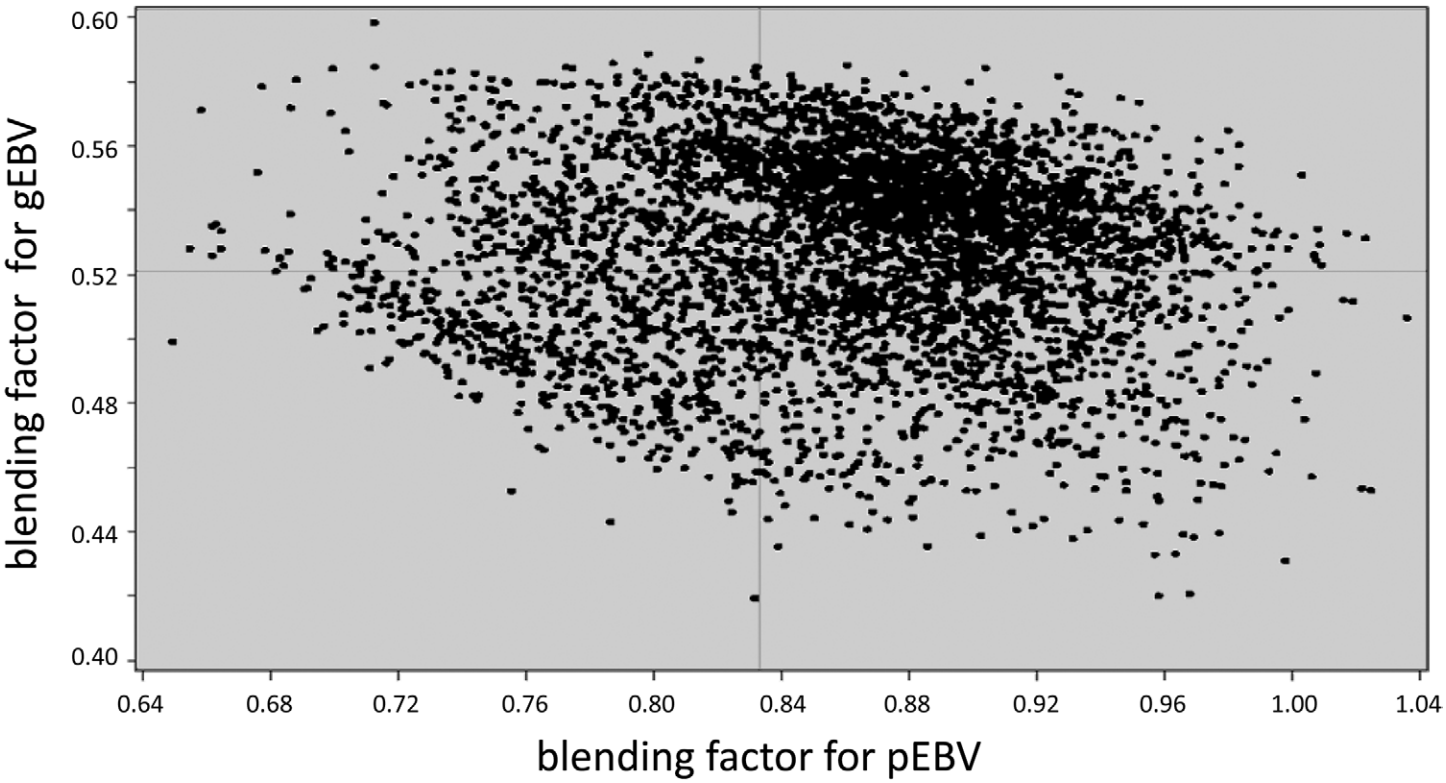


Figure 1. An example of blending factors for the conventional polygenic BLUP EBV (pEBV) and the genomic EBV (gEBV) for litter size in a single pig line. Each datapoint represents an animal with its specific accuracies of the two EBVs. Note that the two axes have the same scale: the pEBV blending factor is always larger than the one for the gEBV, and its range is twice as wide. This reflects the limited accuracy of this 2011 gEBV, based on a panel of 196 markers.

SNPs: Anonymous markers

Around 2009, a new class of DNA markers (SNPs) became commercially available, which introduced the principle of very large numbers of anonymous markers (the currently most commonly used SNP chip in pig breeding holds 64,232 markers; typically about two-thirds of these pass quality control tests and qualify for usage in analyses), most of them without a clearly known biological function in sharp contrast to the previous version described above: those earlier markers had typically been developed using the *candidate gene approach*, deliberately exploring associations to genes with known functions in the regulation of the target trait.

In the meantime, Meuwissen et al. (2001) had described how to use Bayesian statistics to relate all those markers to the target trait phenotype simultaneously. With sufficient marker density, this technique produces properly adjusted estimates for all the marker allele substitution effects, and these can be matched to an animal's marker genotypes (0, 1, and 2 as above) into a genomic EBV (**gEBV**).

Genome-wide prediction and genomic selection

Following developments in dairy cattle breeding (e.g., Harris and Johnson, 2009), the gEBV can be combined (*blended*, making use of blending factors that represent each element's statistical accuracy) with the conventional BLUP value that takes no account of genomic information (a polygenic EBV: **pEBV**) into the final marker-enhanced EBV. See Fig. 1 for an example from 2011. Nielsen et al. (2010) showed that blending SNP information into the BLUP evaluation increases the reliability of EBVs for both genotyped and non-genotyped animals.

The Meuwissen method calculates the gEBV without worrying about the individual significance of each marker: all the marker effect estimates (up to 64,232 in the above example) are included, implicitly assuming that the (very many) nonsignificant ones will cancel each other out. Again, this is in sharp contrast to the earlier approaches, where significance testing was performed very rigorously as is common practice in genome-wide association studies (**GWAS**).

An alternative approach makes use of marker panels that hold a relatively small subset (up to a few hundred, as in Fig. 1) of the full anonymous SNP set, selected based on the significance of their effects on the target trait although the biological function of their marked QTLs is still unknown (and typically not explored). Such panels are trait specific and usually population specific; examples were described by Otto et al. (2007) for meat quality, Deeb et al. (2010) for litter size and for scrotal hernia, and Mathur et al. (2011) and Jafarikia et al. (2015) for boar taint. Note that if the subset has been selected properly, and if the Meuwissen method functions as it should, then both approaches should have the same predictive power because they exploit the same subset of significant markers; the main advantage of the small panel is that it is cheaper to genotype a few hundred markers than several tens of thousands. The main disadvantage is that the panel has to be generated as such, and it will work for one trait in one line only; with the decreasing cost of genotyping (Fig. 2), the larger panels gain in popularity because of ease and uniformity in database handling.

Estimating the SNP effects: Reference populations

All the above approaches rely on the estimation of marker allele substitution effects by regressing the target trait phenotype (or in a more sophisticated

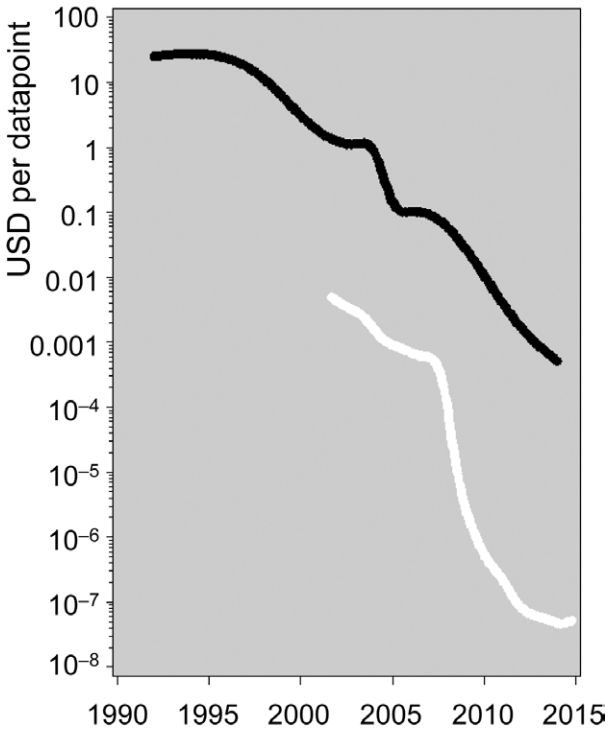


Figure 2. Timetrends of the cost of genotyping, in US dollars per datapoint (one marker, one individual). Black line: commercial genotyping of pig DNA. White line: human full genome sequencing (data from www.genome.gov/sequencingcosts). Note that the vertical axis is logarithmic.

version, high-accuracy conventional EBVs for that same trait: Garrick et al., 2009; Ostersen et al., 2011) on the marker genotypes (Whittaker et al., 1996). This requires a *reference* or *training population* of genotyped animals with trait phenotype records (or high-accuracy EBVs, typically based on progeny testing)—likely, different reference populations for different groups of traits (e.g., growing pigs vs. sows). These marker effect estimates are then used to calculate gEBVs for genotyped selection candidates from the main breeding population; those are typically not part of the reference population to avoid double counting, and they don't need to have a phenotype record.

This creates a dilemma: marker effects must be estimated in the reference population, and this requires genotyping (Henryon et al., 2015). But selection candidates in the main breeding population must be genotyped too so that their gEBVs can be calculated. Because genotyping costs money, the system must be optimized: sufficient genotyping in the reference population to make the marker effect estimates reliable enough, and sufficient genotyping of selection candidates to support a good selection intensity: *strategic genotyping*. A key element here is the genetic relationship of selection candidates to the reference population: the reliability of the gEBV is directly proportional to this parameter (e.g., Clark et al., 2012). Therefore, the reference population will have to be updated on a regular basis to keep the marker effect estimates in sync with new generations of the breeding population.

So, the reference populations and the main breeding population are not the same, and this can be exploited by keeping the reference pigs in commercial production farms—with infectious, nutritional, climatic, and social conditions much closer to the "real world" than what the typical nucleus farm with its high biosecurity has to offer. That way, the nucleus population can be directly selected for performance in commercial conditions without actually housing it there because the markers have been trained on that: an extension of the conventional combined crossbred and purebred selection (**CCPS**) schemes described by Casey et al. (2006). It would then be logical to have a *crossbred* reference population (e.g., Dekkers 2007; Zeng et al., 2013) and use the marker effect estimates in both parental lines—although that halves the genetic connection of the reference population to each of those lines, so the reference population will have to be larger to compensate for that, and even more regularly updated. Other obvious factors that influence the reliability of such a gEBV for commercial performance are the purebred/nucleus versus crossbred/commercial genetic correlation of the trait, and the volume and recording quality of purebred versus crossbred phenotypes (Ibáñez-Escriche et al., 2014; Hidalgo et al., 2015). In other words, the optimum information design is likely to be line and trait specific.

Single-step evaluation

Still, blending the gEBV and the non-genomic pEBV takes place after calculating them separately, and that is awkward because they will need standardization. This was remedied by Legarra et al. (2009) and Christensen and Lund (2010) who proposed a high-marker-density method to do the blending inside the mixed-model equations; see Aguilar et al. (2010) for useful detail. This *single-step evaluation* pulls all the phenotypic, genomic, and pedigree information together into a single model, and it can successfully deal with EBVs for purebreds and crossbreds in a CCPS-based system (Christensen et al., 2014).

An important element here is that the marker effect estimation does take place internally, but the estimates do not form explicit output of the system. With that, the concept of a reference population as such becomes fuzzy, particularly in CCPS-based systems where entry of the relevant phenotypic data (i.e., training data) is a regular routine. Strategic genotyping remains important to ensure selection of the best candidates for the next generation (Henryon et al., 2015).

Genomic relationship-based prediction

At the same time, Hayes et al. (2009) showed that for a system with high enough marker density and with normally distributed QTL effects, the gEBV can be interpreted in two equivalent ways: 1) as above: the result of combining the estimated marker effects with the animal's marker genotypes, or 2) as the result of a BLUP evaluation where the relationship matrix among the animals in the system contains genomic relationships (based on how many marker alleles two animals have in common) instead of pedigree-based ones. Option 2 is delightfully simple, mainly because it is trait independent; it is much easier to explain (particularly to non-geneticists) and has become the standard mode of interpretation.

Although the datapoint cost of genotyping has come down more than 10,000-fold since 1990 (Fig. 2) and is still falling rapidly, the number of datapoints (markers and animals) demanded by the recent methods is growing at least as fast as that—so the approach continues to be costly, certainly when compared with the much more large-scale sequencing technologies in the human field. Much of this was resolved by the *imputation* methodology proposed by Habier et al. (2009), where low-density genotypes are scaled up to high-density ones. Fig. 3 (based on Huang et al., 2012) shows the associated cost-benefit relationships.

Comparisons to Other Species Systems

Following from the above, the most advanced genomic selection applications in commercial pig breeding at this moment (late-2015) are single-

step evaluations based on imputed genotypes, to produce gEBVs that are used for selection the same way as before. There are some important differences with the dairy cattle and poultry systems, as follows.

Bacon and cheese

Genetic improvement depends on existing genetic variation, selection intensity (i), generation interval (L), and EBV accuracy. For the sake of discussion, genetic variation is assumed constant, i/L is a matter of design, and EBV accuracy has to do with accurate phenotyping and optimum genetic evaluation. This basic concept still fully holds in the genomic age, but the *optimization* of genetic improvement is different.

Generation interval versus accuracy. The most accurate EBVs come from progeny testing, and this must be balanced against time. In dairy cattle breeding, the balance has always been toward high accuracy, whereas in pig breeding, the balance tipped about 30 yr ago toward a short generation interval. An important factor here is that cattle AI produces significantly more progeny than pig AI, allowing for a higher male selection intensity in cattle. Genomics improve accuracy most at a young age; in dairy cattle breeding, this allows for a dramatic shortening of the generation interval with marginally lower accuracies, but in pig breeding, the benefit comes from a considerable increase in accuracy at the same age of selection as before. This contrast relates to the type of traits that dominate the breeding goals, as follows.

Trait-related differences. Milk production traits form the dominant objective in most dairy cattle populations: these are single-sex late-in-life traits, so that selection decisions must be made before phenotypes are available, particularly in males. Pig breeding is ultimately about a low production cost of quality pork. Many relevant traits can be recorded on selection candidates of both sexes before selection takes place (growth rate, feed intake, mortality, and ultrasound body composition). Other traits like meat quality require the slaughter of relatives but can be recorded at that same time. The similarity to the cattle system is in the relevance of lowering the cost price of piglets through improvement of reproduction traits: this creates the same single-sex and late-in-life issues as in milk production.

High-value bulls. Dairy bulls are conventionally "proven" at a high age (when their daughters are at the end of their first lactation), so the investments in the system are huge. A bull has to be reared, a large number of semen doses must be collected, and then progeny test commences with lots of waiting time. The cost of genotyping or even sequencing such animals is very low compared with the conventional testing costs. This situation is very different in pigs where the production cost of the animal is relatively low and the pool of selection candidates contains all the tested, pedigreed, and purebred male piglets—this makes the cost of genotyping a more important issue.

Population stratification and power of analysis. Successful application of genomics depends on the size and structure of the training population and its relatedness to selection candidates. The dominant dairy breed Holstein is mostly used purebred, giving rise to huge numbers of more or less closely related cows around the world (all of them selection candidates). Potentially all of these contribute phenotypes to the reference population; genotypes enter the equation through their sires or through (low density) genotyping of females themselves. The first genomic applications in Holstein involved about 1,500 progeny-tested bulls, each with easily more than 100 daughter phenotypes. The second step involved 4,000 bulls, followed by collaboration between countries and organizations to achieve 16,000 bulls as a reference population.

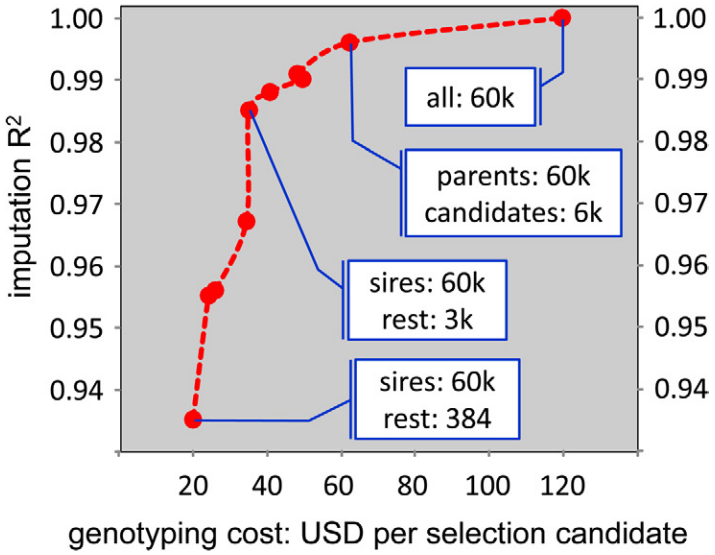


Figure 3. Cost and benefits of imputing high-density (64,232 markers: 60k) genotypes of selection candidate pigs, for several scenarios of genotyping various subgroups of the population at different densities (384, 3,071 or 5,963 markers: 384, 3k, 6k) and scaling them all up to 60k by imputation. Cost in US dollars per selection candidate (based on 2011 genotyping cost), benefits in terms of the reliability of the imputed 60k genotype. Data from Huang et al (2012).

By contrast, pig breeding is organized by separate herdbook societies and breeding companies; phenotype availability is therefore highly fragmented. The typical large pureline pig population has about 2,000 sows—with about 50 sires selected per year, this produces 40 daughters per sire on average; it would then take 30 yr of data recording to achieve the initial Holstein setup. The crossbred end products derive from multiple breeds and lines (pig breeding programs involve one to four sirelines and two to five damlines); the reference populations for sirelines and damlines comprise crossbred commercial slaughter pigs and ditto sows, respectively. In the current stratified situation, collaboration between programs is hardly feasible—GWAS and gBLUP efforts are therefore much more challenging in terms of costs and income, triggering alternative approaches.

Male domination versus female repeated records. The average sow at parity 5 has 5 × 15 = 75 progeny. A nucleus population with 2,000 sows and a replacement rate of, say, 40% then generates a yearly training population of 750 sows with 75 progeny, half the size of the basic situation in dairy cattle. In pigs, analysis is mostly driven from the maternal side where dairy cattle is a paternally driven system (see Lillehammer et al., 2011, for simulations).

Bacon and eggs

Costs. In cattle, the production cost of the animal is considerable and in pigs a bit less so, but the slaughter value and hence the production cost of poultry (particularly broiler chickens) is very low, and genotyping an individual has to compete, cost-wise, with phenotyping several individuals—thus balancing higher selection intensity against higher accuracy.

Structure. The production column of broilers and layers involves at least one extra pureline multiplication step; this makes the tracking and tracing of QTL at the crossbred commercial level even more challenging than it is in pigs. This is theoretically solved by tracing haplotype blocks back to the line of origin and then selecting the pure lines on the presence of the favorable blocks.

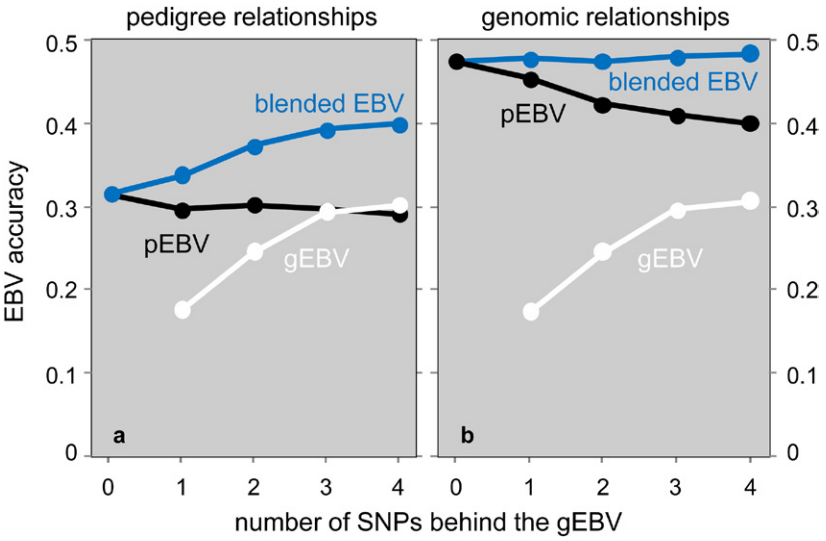


Figure 4. EBV accuracies for teat number in a pig line, based on pedigree (a) versus genomic (b) relationship coefficients. pEBV: polygenic EBV from conventional BLUP analysis; gEBV: genomic EBV based on 1, 2, 3, or 4 most significant markers fitted explicitly (which diverts genetic variation away from the pEBV). The blended EBV is a weighted combination of pEBV and gEBV. Data from Lopes et al. (2016).

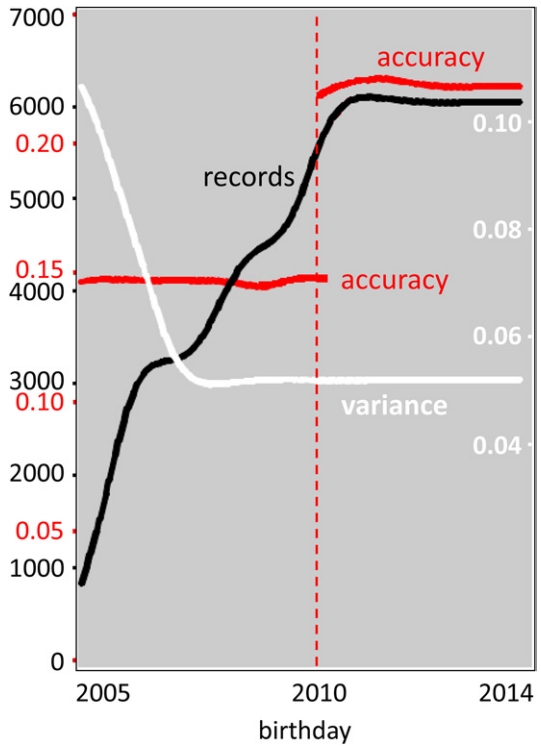


Figure 5. Post-weaning mortality recording and EBV accuracy in a pig line; spline interpolation plots through quarterly data. Black line: number of crossbred commercial CCPS records per quarter (not cumulative). White line: variance based on the mortality incidence per quarter; the recording base was extended to farms with a more stable health status during 2005 and 2006. Red line: mean accuracy of the EBV of 6-mo-old nucleus selection candidates; genomic selection was implemented in 2010.

Bacon and chips

Potato breeding (and most of plant breeding in general) is entirely different. While some production crops follow index selection for increased yield, genetic improvement of many plant species is based on clone lines, testing of clones in different environments; breeding goals aim at creating clone products with clear benefits in disease resistance and individual gene actions. The huge benefit of working with clones is that genomic evaluation can be done once very thoroughly, where repeated phenotypes can follow over time.

Application of Genomics in Pigs in Three Examples

Example 1: Teat number

Genomics add most information for traits that are difficult to improve with the conventional methods: hard to measure, late-in-life, single-sex, and/or with a low heritability. By contrast, teat number can be recorded without measurement error on newborn piglets (on time for every selection moment) of both sexes, and it has a heritability of about 0.4. Because of this favorable phenotyping structure, this is probably the trait in pig production where genetic analysis reaches its highest accuracy. Duijvesteijn et al. (2014) report on an in-depth analysis describing methodology, QTLs, comparison with literature results on microsatellite markers, and candidate genes relating the formation of teats in the pig to the formation in other species and to the regulation of vertebra number.

Figure 4 gives the mean accuracy of teat number EBVs in a single pig line, calculated from the conventional pedigree-based and genomic relationship matrixes and from these same with explicit addition of the most significant one, two, three, or four SNPs. The single most significant SNP increases the accuracy of the pedigree-based pEBV by 7%, and the four most significant ones do so by 27%. Replacing the pedigree-based relationship matrix by the genomic one gives a very interesting 50% increase. Not surprisingly, addition of SNPs to the genomic system has little effect.

It is interesting that the use of genomic information creates a substantial increase in accuracy for this trait, even if not theoretically anticipated given the favorable information structure of the conventional situation. This example shows that genomic evaluation in pigs has considerable added value; it also raises the question of how to efficiently generate relevant training information for late-in-life and/or single-sex traits.

Example 2: Post-weaning mortality

Post-weaning mortality easily qualifies as a hard-to-measure trait: a binary (0–1) trait with low incidence and heritability ($p \approx 0.05$, $h^2 \approx 0.05$), strong and erratic environmental influences, and cumbersome individual recording—but with a very high economic value (Knap, 2014). The main challenge here is proper phenotyping: any type of breeding value estimation will require very high data volumes to achieve reasonable statistical power for such a trait. The first requirement for breeding value estimation is always variation, and the variation of a binary trait is proportional to its incidence: recall that its variance equals $p \times (1 - p)$. Therefore, the ideal farm for recording this trait is very large, has a high mortality incidence, continues with that for a long time, and has staff strongly motivated toward high-quality data recording. Such farms are difficult to find; they will necessarily be commercial production farms (as opposed to the breeding company's nucleus units), and the pigs will be crossbreds. To maintain a close genetic connection to the nucleus selection candidates, the pigs will be produced from nucleus semen: a classical CCPS system.

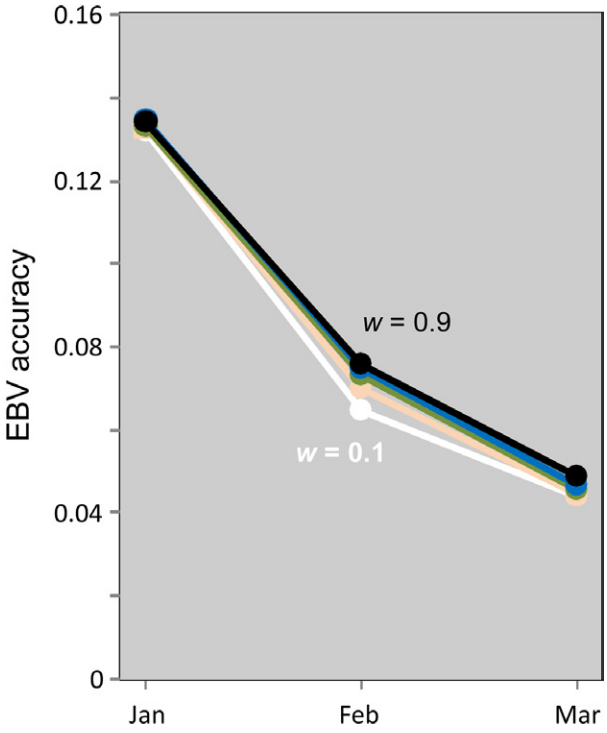


Figure 6. EBV accuracies (correlations to adjusted phenotype) for sow culling rate after first litter in a pig line, with different values for the blending factor w: higher w values give more weighting to the pEBV (based on pedigree relationships among animals) and less to the gEBV (based on genomic relationships). The reference dataset runs to December 2013; the horizontal axis shows the first three months of the subsequent validation data.

Our example gives the mean accuracy of the post-weaning mortality EBV for nucleus selection candidates, based on data recorded in such a CCPS system. In 2005, MAS was set up for this trait, using 5 to 20 markers in various lines. Genomic selection was implemented in 2010 using a dedicated small marker panel, followed by single-step evaluation in 2012. In Fig. 5, we see that this increased the EBV accuracy by 50%, from 0.14 to 0.22; for a binary trait with a 0.05 incidence and a ditto heritability, the latter value is equivalent to a progeny test on 101 progeny. Recording volume increases steadily up to late 2010; the mortality incidence and its associated variance in these records drops just as rapidly from an alarming (but for EBV purposes, highly useful) initial 10% on a few selected farms to a more representative 5% level by 2006. These opposite trends might be taken to explain the horizontal trend in the accuracies in that early period—but the incidence stabilizes by 2006 and the volume continues to increase with no apparent effect on the accuracies until the start of genomic prediction in 2010. Obviously, some other factor played a role to neutralize the increasing recording volume between 2006 and 2010.

Example 3: Blending in the single-step evaluation of culling after first litter

Single-step evaluation blends the pedigree-based relationship coefficients with the genomic ones inside the mixed-model equations; the focus on each differs by trait and by line. This example uses the trait *culling after first litter* (**CAFL**): a 0–1 trait recorded on sows in multiplier farms only—nucleus sows are culled on a selection index that disqualifies them for this trait. Hence, the nucleus has no CAFL phenotypes, and the information must be transferred to it through genetic relationships. Genomic information would be expected to be very useful here.

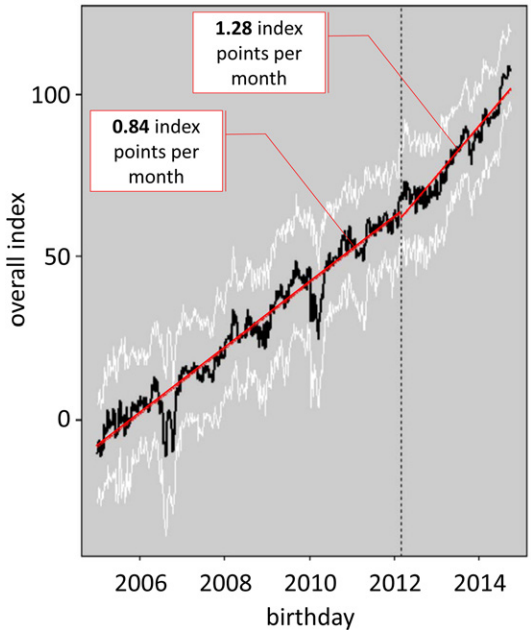


Figure 7. Genetic trend in the overall selection index of a pig line before and after implementation of genomic selection in 2012. Black line: weekly means of the index; white lines: one standard deviation interval; red lines: linear regressions before and after.

The culling rate after first litter is 20%, i.e., a variance of 0.2 × (1 – 0.2) = 0.16; the heritability is 0.25. Single-step evaluation is based here on a reference dataset that holds 319,320 CAFL records and 22,065 genotyped animals (a mix of 8k and 60k) up to 2013; records from 2014 onward serve as the validation dataset for which EBVs were produced with blending factors *w* that were varied from 0.1 to 0.9: higher *w* values focus more on the pedigree information. For the validation, *adjusted phenotypes* were obtained in a non-genomic single-trait BLUP evaluation, as the animal's EBV plus its residual value. The validation criterion was the correlation of EBV with the adjusted phenotype, by month in 2014.

Figure 6 shows how this correlation decreases over time and how the highest correlations are obtained with *w* = 0.9: genomic and pedigree relationships are best weighted 1:9 here. By contrast, in the same pig line, similar analyses for growth rate and feed conversion ratio (Christensen et al., 2012) produced optimum *w* values around 0.25: genomic and pedigree relationships are best weighted 3:1 for these traits. So, in this line, genomics provide much less added value for CAFL than for growing pig traits; similarly Guo et al. (2014) found high optimum *w* values (around 0.5) for litter size and piglet mortality in this line. Compare this to Fig. 1, where the litter size pEBV is always weighted more heavily than the gEBV. Clearly, strategic genotyping (and phenotyping) for these sow traits must follow a different pattern than for growing pig traits.

Summing up examples 1, 2, and 3

Figures 4 and 5 show a 50% increase in EBV accuracy due to a shift to genomics technology in two very different pig lines for traits as widely different as teat number and post-weaning mortality rate. Figure 6 illustrates another trait in another line where genomic information is less sig-

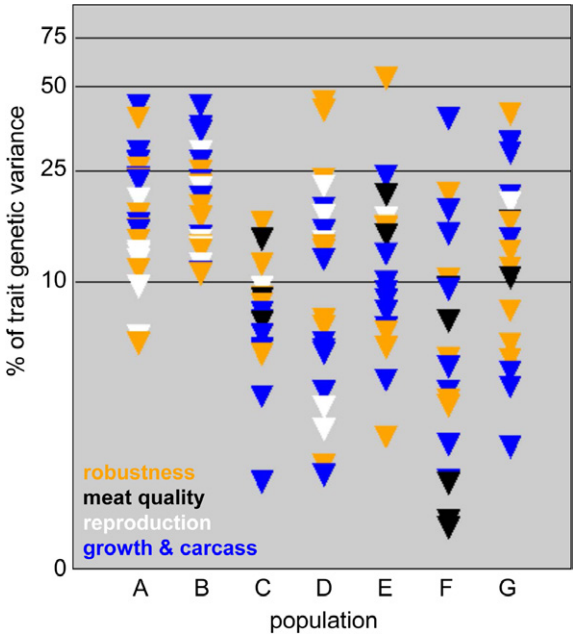


Figure 8. Percentage of trait genetic variance explained by the 1% most significant of 60k SNPs, for 31 traits (divided into four functional groups) in seven pig lines (A to G). Note that the vertical axis is logarithmic.

nificant; but Fig. 7 shows again a 50% increase in overall genetic change across eight index traits in that same line.

Future Developments

Genomic selection in pig breeding may develop along two paths that are likely to converge on the medium term. The first is via increased marker density in combination with increased genotyping and phenotyping volume (the raw statistical power approach). This is obviously a diminishing-returns system, neatly quantified by Erbe et al. (2013) for two cattle breeds, in terms of marker density and size of the reference population. Full genome sequencing would guarantee that the actual functional mutations are included in the system (Meuwissen and Goddard 2010); Hickey et al. (2014) suggest how to explore this for the next generation of genomic selection.

The second path is via a move away from the black-box approach of breeding value estimation, weighting markers into the gEBV according to their effect on the trait of interest (as estimated in GWAS analyses: the biological approach). Obviously this would be useful mainly for traits with a genetic background that deviates from the infinitesimal pattern of very many QTL, each with small effects. Figure 8 illustrates to what extent this is the case for 31 traits in seven pig lines: obviously the vast majority of cases do not really qualify (similar to the blue line in Fig. 4b which benefits from improved pedigree and from individual marker contributions, yet shows little improvement from the latter); but a subset in Fig. 8 with clear non-infinitesimal patterns comprises traits of all the categories identified here—with very line-specific patterns as expected.

An interesting approach is backsolving the genomic relationship matrix on the gEBVs to obtain estimates for the marker effects (e.g., Zhang et al., 2010; Wang et al., 2012), which would allow for blending such estimates with the regular gEBVs (Zhang et al., 2015) within a single analysis.

Heterosis plays an important role for many traits in crossbred pig production, and when gEBVs are based on crossbred performance data, it would seem obvious to include the non-additive genetic effects that cause heterosis in the statistical model: that should increase the reliability of the gEBV and reduce its bias. Su et al. (2012) and Zeng et al. (2013) present methods to deal with this in a gEBV setting. Dams and sires of two pure lines allow for the estimation of dominance effects even when progeny are not genotyped: dominance probabilities can be quantified via the combination of known sire and dam genotypes (Boysen et al., 2013).

Literature Cited

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730.

Boysen, T.-J., C. Heuer, J. Tetens, F. Reinhardt, and G. Thaller. 2013. Novel use of derived genotype probabilities to discover significant dominance effects for milk production traits in dairy cattle. Genetics 193:431–442. doi:10.1534/genetics.112.144535.

Casey, D., M. Perez, D. McLaren, and T. Short. 2006. Crossbred breeding values: Selecting for commercial performance in pigs. In: Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil. Communication 06-26.

Christensen, O.F., and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2.

Christensen, O.F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. doi:10.1017/S1751731112000742.

Christensen, O.F., P. Madsen, B. Nielsen, and G. Su. 2014. Genomic evaluation of both purebred and crossbred performances. Genet. Sel. Evol. 46:23. doi:10.1186/1297-9686-46-23.

Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J.H.J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Gen. Sel. Evol. 44:4.

Deeb, N., M.A. Cleveland, S. Forni, N. Yu, S. Sebbana, C.D. Gladney, A. Mileham, and D.G. McLaren. 2010. Genome-wide association study for scrotal hernia in commercial pig populations. In: Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Communication 0525.

Dekkers, J.C.M. 2007. Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. 85:2104–2114. doi:10.2527/jas.2006-683.

Duijvesteijn, N., J.M. Veltmaat, E.F. Knol, and B. Harlizius. 2014. High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. BMC Genomics 15:542. doi:10.1186/1471-2164-15-542.

Erbe, M., B. Gredler, F.R. Seefried, B. Bapst, and H. Simianer. 2013. A function accounting for training set size and marker density to model the average accuracy of genomic prediction. PLoS ONE 8(12):e81046.

Fernando, R.L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467–477. doi:10.1186/1297-9686-21-4-467.

Fisher, R.A. 1930. The genetical theory of natural selection. Oxford Univ. Press, Oxford.

Garrick, D.J., J.F. Taylor, and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Gen. Sel. Evol. 41:55.

Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers: 1. Methods. Theor. Appl. Genet. 46:319–330. doi:10.1007/BF00281673.

Guo, X., O.F. Christensen, T. Ostersen, Y. Wang, M.S. Lund, and G. Su. 2014. Improving genetic evaluation of litter size and piglet mortality for both genotyped and non-genotyped individuals using a single-step method. J. Anim. Sci. 93:503–512. doi:10.2527/jas.2014-8331.

Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2009. Genomic selection using low-density marker panels. Genetics 182:343–353. doi:10.1534/genetics.108.100289.

Harris, B.L., and D.L. Johnson. 2009. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93:1243–1252. doi:10.3168/jds.2009-2619.

Hayes, B., and M.E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33:209–229.

Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res., Camb. 91:47–60.
Henryon, M., P. Berg, and A.C. Sørensen. 2015. Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. Livest. Sci. 166:38–47. doi:10.1016/j.livsci.2014.06.016.
Hickey, J.M., G. Gorjanc, M.A. Cleveland, A. Kranis, J. Jenko, G. Mészáros, J.A. Woolliams, and M. Perez-Enciso. 2014. Sequencing millions of animals for Genomic Selection 2.0. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. Communication 00-00.
Hidalgo, A.M., J. W.M. Bastiaansen, M.S. Lopes, R. Veroneze, M.A.M. Groenen, and D.J. de Koning. 2015. Accuracy of genomic prediction using deregressed breeding values estimated from purebred and crossbred offspring phenotypes in pigs. J. Anim. Sci. 93(7):3313–3321.
Huang, Y., J.M. Hickey, M.A. Cleveland, and C. Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genet. Sel. Evol. 44:25. doi:10.1186/1297-9686-44-25.
Ibáñez-Escriche, N., S. Forni, J.L. Noguera, and L. Varona. 2014. Genomic information in pig breeding: Science meets industry needs. Livest. Sci. 166:94–100. doi:10.1016/j.livsci.2014.05.020.
Jafarikia, M., J. Squires, F. Schenkel, F. Fortin, S. Wyss, W. van Berkel, B. Sullivan, and R. de Wolde. 2015. Prediction of boar taint in major breeds of Canadian pigs using genetic markers. In: Proc. Banff Pork Seminar. http://www.ccsi.ca/Abstracts/2015BPS-boar-taint.pdf. (Accessed 17 July 2015.)
Knap, P.W. 2014. Pig breeding goals in competitive markets. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. Communication 007.
Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. doi:10.3168/jds.2009-2061.
Leroy, P., J. Naveau, J.M. Elsen, and P. Sellier. 1990. Evidence for a new major gene influencing meat quality in pigs. Genet. Res. 55:33–40. doi:10.1017/S0016672300025179.
Lillehammer, M., T.H.E. Meuwissen, and A.K. Sonesson. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89:3908–3916. doi:10.2527/jas.2011-4044.
Lopes, M.S., H. Bovenhuis, M. van Son, Ø. Nordbø, E.H. Grindflek, E.F. Knol, and J.W.M. Bastiaansen. 2016. Using markers with large effect in genetic predictions. Genet. Sel. Evol.
Mathur, P.K., H.A. Mulder, J. ten Napel, J.J. Windig, R.E. Crump, B. van der Fels, L. Heres, S. Bloemhof, M.S. Lopes, and E.F. Knol. 2011. Genetic solution to the problem of boar taint for production of entire males. In: Proc. Conf. Boars heading for 2018, Amsterdam, the Netherlands. Presentation 14. http://boars2018.com/conference/conferencescientific-program-and-presentations. (Accessed 17 July 2015.)
Meuwissen, T., and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185: 623–631.
Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.
Neumaier, A., and E. Groeneveld. 1998. Restricted maximum likelihood estimation of covariances in sparse linear models. Genet. Sel. Evol. 30:3–26. doi:10.1186/1297-9686-30-1-3.
Nielsen, B., T. Ostersen, G. Su, O.F. Christensen, and M. Henryon. 2010 Use of genomic SNP information in pig breeding. In: Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Communication 0522.
Ostersen, T., O.F. Christensen, M. Henryon, and B. Nielsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genet. Sel. Evol. 43:38. doi:10.1186/1297-9686-43-38.
Otto, G., R. Roehe, H. Looft, L. Thölking, P.W. Knap, M.F. Rothschild, G.S. Plastow, and E. Kalm. 2007. Associations of DNA markers with meat quality traits in pigs with emphasis on drip loss. Meat Sci. 75:185–195. doi:10.1016/j.meatsci.2006.03.022.
Rothschild, M.F., C. Jacobson, D. Vaske, C. Tuggle, L.Z. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham, and G. Plastow. 1996.The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc. Nat. Acad. Sci. USA 93:201-205.
Su, G., O.F. Christensen, T. Ostersen, M. Henryon, M.S. Lund. 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense Single Nucleotide Polymorphism markers. PLoS ONE 7(9):e45293.
Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172.

About the Authors

Egbert Frank Knol graduated from Wageningen University in 1982 and took up teaching for two years before joining the Unilever pig breeding company Dalland. "Genetics of Piglet Survival" is the title of the Ph.D. thesis he defended in Wageningen in 2002. Consolidation in pig breeding companies led to his current position of Director of Research for Topigs Norsvin International. His research is now partially coordinated within the Dutch Breed4Food consortium. Within EAAP, he is serving as Vice-President in the Commission on Pig Production.
Correspondence: egbert.knol@topigsnorsvin.com.

Bjarne Nielsen is Chief Scientist at the SEGES Danish Pig Research Centre, focusing on development and implementation of genetic and genomic models relevant for the breeding goals of DanAvl. Nielsen graduated in 1990 within agricultural sciences and obtained a Ph.D. in statistics from The Royal Veterinary and Agricultural University of Copenhagen in 1996. He has previously worked as a senior scientist at The Danish Institute of Agricultural Sciences and as scientific visitor at The Technical University of Denmark and Norwegian University of Life Sciences. He has been responsible for the experimental design, statistical analysis, and development and application of mathematical models of many biological systems, in a number of projects.

Pieter Knap graduated from Wageningen University in 1981 and has been working in commercial pig breeding ever since, based in Holland, England, Norway, Scotland, and Germany. He received a Ph.D. from Wageningen in 2000 (variation in maintenance requirements of growing pigs), and since that same year, he has served as Genetic Strategy Manager with Genus-PIC, focusing on breeding goals and the implementation of genetic improvement technology. From 2002 to 2013, he was (Vice-)President of the EAAP Commission on Pig Production.
Correspondence: pieter.knap@genusplc.com and see de.linkedin.com/in/pieter-knap-02942066

Left to right: P. Knap, E.F. Knol, and B. Nielsen.

Visscher P.M. 2008. Sizing up human height variation. Nature Genetics 40:489-490.
Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W.M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res., Camb. 94:73–83.
Whittaker, J.C., R. Thompson, and P. Visscher. 1996. On the mapping of QTL by regression of phenotype on marker type. Heredity 77:23–32. doi:10.1038/hdy.1996.104.
Zeng, J., A. Toosi, R.L. Fernando, J.C.M. Dekkers, and D.J. Garrick. 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. Genet. Sel. Evol. 45:11. doi:10.1186/1297-9686-45-11.
Zhang, Z., M. Erbe, J. He, U. Ober, N. Gao, H. Zhang, H. Simianer, and J. Li. 2015. Accuracy of whole-genome prediction using a Genetic Architecture-Enhanced variance-covariance matrix. G3 5:615-627.
Zhang, Z., J. Liu, X. Ding, P. Bijma, D.J. de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker derived relationship matrix. PLoS One 5:e12648.