Selecting Thoroughbreds Today

B. Langlois

Introduction

This is a paper whose aim is to infer a logical modern development of Thoroughbred breeding taking into account the new molecular tools now available. To do this, we begin with the bibliographical description of the state of the art of Thoroughbred breeding.

I-State of the art.

Selection of the English horse for and by racing results instead of a lovely appearance as was previously the case was an incontestable innovation and success of British zootechny. During the XIXth century, mean times in races continuously decreased. DARWIN himself noticed that selection had led the English race horse to surpass the Arab horse in speed and wither height from which it was derived. He also noted that a limit will certainly be reached because of physical constraints determined by the strength of muscles and resistances connected with the weight of the animal.

During the XXth century, this decrease in mean racing times did not continue despite persevering in selection. Mass selection summarized by the well known adage “breed the best with the best and hope for the best” did not succeed anymore to improve mean speed of Thoroughbreds.

However, different measures of performances such as transformed earnings, general handicap values mathematized as performance rates which led to the advanced use of rankings in races, still exhibit high correlations between relatives as seen in the latest bibliographical synthesis by Thiruvenkadan *et al.* (2009).

Despite this fact, implementation of the techniques of quantitative genetics leading to breeding value estimations did not succeed in convincing breeders. First, a very speculative economy has led to very little interest in cooperative actions aiming to increase the mean performance of the population. Indeed at the breeder’s level these techniques have too much imprecision to convince. Second, I would add that applying quantitative genetics theories supposes respecting the conditions of panmixia, fitting a genetic model on the phenotypic

---

INRA, BIGE-GABI, 78 352, Jouy-en-Josas, France

data where the genetic correlation between relatives can be assessed to be two times their coefficient of relationship calculated from the available pedigree. Rigorously this supposes no genetic drift and no inbreeding, random mating and no selection. Departures from these conditions are more or less supported but for the Thoroughbred it is difficult to account for remote inbreeding (Mahon G.A.T. Cunningham E. P. 1982), selection and non random mating. Selection can generally be accounted for when the data describe the process. But national files in international breeding make this condition difficult to match.

II- What recently happened?

In February 2007, the horse genome was sequenced. Scientific publication came later (Wade *et al.* 2009). In 2008, this knowledge rapidly allowed a 50 K illumina bead chip to be developed and made available on the open market. Fifty thousand bi allelic markers can then be routinely revealed for genetic analyses. These are evenly dispersed along the 32 chromosomes of the horse except the Y. The horse genome was evaluated to be 2.7 Gb (2.7 billions base pairs). If you estimate that one centimorgan is approximately equal to one million base pairs (1Mb), then markers are available every 0.54 centimorgans on average.

A technique then became available to mark the genome of the horse with markers every 0.5-0.6 centimorgans. If the transmission of these markers is followed they can be associated with hereditary characters segregating in the same way. This genome scan remains an approximate image of the real genome; however, it gives a considerably sharper image than what we had before by just inferring the genotype from the phenotype. We can add that bead chips of 130 KSNP are being prepared making this kind of tool sharper and sharper for any genetic analysis. In addition, high throughput sequencing is developing quickly. Several billion base pairs can be analyzed in only one week. They are later reassembled by bioinformatics techniques. We are then in the situation to completely solve the problem of the inaccuracy of the definition of the genome. This will open a new large area of research to discover how a genome is translated in a phenotype (transcriptomic and even epigenetic). Infinity of interactions and feedbacks are possible leading to a vast range of new knowledge.

III- What can we still do?

As the image of the genotype is becoming sharper, we can imagine mastering the effect of heredity on fewer numbers of animals than before. Since the statistical inference of the genotype from the phenotype uses the law of great numbers, it is expected to greatly reduce this condition when the genotype is directly available. However a precise preliminary analysis is needed to identify the real value of different genotypes. This will therefore not be available immediately.

On the contrary, we have seen (Mahon and Cunningham 1982, Cunningham *et al.* 2001) that there is a problem in Thoroughbreds to master real inbreeding and/or parentage. We have shown how the great amount of markers available through the SNP bead chips allowed inferring the relationship of two individuals through the situation of identity of their molecular markers (Langlois 2003.2006. 2007).We only need to know the allele frequency of each SNP in the Thoroughbreds and to choose a panel sufficiently spaced in base pairs to be considered as independent. SNP with MAP>0.40 should also be preferred. Genotyping 50-60 randomly chosen Thoroughbreds should be sufficient to get the allele frequencies (Langlois 2009).

IV- New prospects for the coming years

Most of the genome regions are probably not concerned by the selection of the Thoroughbred. It is therefore important to identify the regions of the horse genome concerned by the characters that are being selected in the Thoroughbred. Two approaches are possible, the intra-breed analysis and the inter-breed. The two are genotyping consuming, that means expensive. The first one also developed under the term “genomics” is politically difficult to develop; the second is more neutral for breeder organizations. I therefore support this second approach based on the detection of signatures of selection that gives more knowledge about different orientations of selection that is expected for the same amount of genotyping. Moreover loci under balancing selection favoring heterozygous can also be detected (Flori *et al.* 2009, Gautier *et al.* 2009) which is a first step to surpass the common strictly additive genetic model used in the intra-breed approach.

Detection of QTL in the first case or detection of signatures of selection in the second will determine the precise zones of the genome to be more deeply analyzed on a physiological point of view by reference mainly to the human genome. In my opinion, this is a necessary study aimed at reducing the dimension of the problem of relationships between genotype and phenotype. Indeed we have at present more variables than equations and proposed solutions are making hypotheses only to make solutions possible. This is not fully satisfactory.

Knowing these precise zones should also change the manner of apprehending parentage in general or relative to the characters under selection. This would also be a kind of revolution.

Conclusion

New molecular genetics tools are bringing forth new knowledge for conducting selection in the Thoroughbred.

The cost of genotyping with 50K SNP bead chips is low when compared to the prize of the animals.

The expected gain in accuracy for estimating breeding value is high. This would be of interest to private owners interested in stud farm selection but not on genetic progress neither at the breed level nor the country level.

A better management of parentage and inbreeding can still be proposed. We know that current tools based on pedigree are not sufficient to manage an existing problem of remote inbreeding.

Preliminary studies are needed to propose efficient genomic tools. We need to identify the regions of the genome involved in the selection process.

For that determination, an interbreed approach is probably the most adapted. It will give more information and allow making the first steps change from the additive genetic model which is efficient but is in fact more statistic than genetic.

References

Cunningham E.P. *et al.* 2001. Animal Genetics 32, 360-364.

Flori L. *et al.* 2009. Plos one,4, (8) e6595.

Gautier M. *et al.* 2009. BMC Genomics,10;550,1471-2164.

Langlois B. 2003. Wageningen Pers, EAAP Publication,116 ,35-53.

Langlois B. 2006. 8th Wcgalp, Belo Horizonte, MG, Brazil, August 13-18, 2006, Communication N° 08-02.

Langlois B. 2007. 58th Ann. Meet. of EAAP, Session S.29, Dublin, Ireland, August 26th-29th 2007, 7p..

Langlois B. 2009. 60th Ann. Meet. of EAAP, horse commission S.19 , Barcelona, Spain 25th august 2009. Abstract n°4549

Mahon G.A.T. Cunningham E. P. 1982. Livest. Prod. Sci.9, 743-754.

Thiruvenkadan A.K. *et al.* 2009. Livest. Sci.121,308-326.

Wade C.M. *et al.* 2009. Science 326,865.