Genomic predictions and genomic selection

C. Dadousis[1] and Z. Ampas[2]

[1] Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

[2] Department of Agricultural Development, Democritus University of Thrace, Greece

Abstract

Some decades ago, the idea of predicting production characteristics of animals (e.g. milk yield) solely based on DNA information would appear as science fiction. However, the rapid progress in molecular biology (e.g. reading the whole genome, DNA microarrays, etc.) in the recent decades has led not only to the development of this technology but to its immediate application in breeding programs as well. In 2001 it was shown, based on simulations, that 50,000 molecular markers would be adequate to achieve accuracies of breeding values close to those of progeny testing. This process which, with the use of molecular markers and "simple" (linear) statistical models, enables the investigation of the genetic background of an individual, aiming in finding the genetic background and to select the best individuals for reproduction, is what we call as ***genomic selection*** (***GS***). The hitherto theoretical data suggest that the *genomic selection* can provide a dynamic up to twofold benefit in the genetic gain of genetic improvement of production animals. The implementation of *GS* has already started in breeding cooperatives around the world, mainly in dairy cows. The cost of genotyping per animal currently stands at ~ $ 100. Nevertheless, with the continuous progress of molecular technology it is expected in the next few years the cost of genotyping to be proportionate to the cost of a simple blood test. The technology of genomic predictions has been recognised as (r)evolution in animal and plant breeding programs while also finds fertile ground in human studies, particularly for predicting diseases.

Introduction

If the 20th century was marked by breakthroughs in physics, the 21st century is already considered as the "century of biology", characterized by significant progress in understanding the basic unit of life, DNA. Almost half a century after the discovery of the structure of DNA by Watson and Crick, the science of biology has succeeded in sequencing the entire genome of organisms, in comparing genomes of different organisms, in genotyping individuals with thousands or millions of molecular markers (or even complete sequencing), etc. all at relatively low cost. Since the late 90's began the sequencing of genomes of various animal and plant species with the breakthrough of sequencing the human genetic code in 2001[1, 2]. A new page in the genetic improvement of cattle started with the complete sequencing of the bovine genome in 2009 [3] (Figure 1).

The new possibilities of the new knowledge and technology from the field of genomics as well as the benefits in breeding programs became quickly apparent and within a relatively short time the new technology became applicable in practice, placing genomic predictions and DNA analysis in the daily life of breeding companies [4-8].

The aim of this article is to make known, in brief, the *genomic selection(GS)* methodology and its potential to farmers and the wider public.

A short historical note in animal breeding

To illustrate the potential of *GS* it would be of some help to firstly have a short historical overview of the breeding programs.

Quantitative genetics theory

The origins of the genetic improvement of plants and animals are traced back to the theory of *quantitative genetics*, which was mainly developed by Ronald Fisher [9] and Sewall Wright [10]. The quantitative genetics theory is a major tool of studying the phenotypic diversity of species. Fisher's theory and the model to study phenotypic diversity are simple in their base:

$$phenotype = genetic\ value + residual$$

The genetic background for each trait, however, is not directly seen (i.e. observation on DNA), but can be assessed through measurements on performance (e.g. milk yield) of the individuals itselves as well as their relatives and population parameters like variance components and heritability. Moreover, the theory assumes that we do not know how different genes work together to produce a particular phenotype (genetic architecture), but suggests that each gene has a very small effect (infinitesimal model) for each trait and the effects show an additive action. Therefore, the theory of quantitative genetics and hence animal breeding have not (or at least do not require) any information on i) the effect of genes, ii) their number and iii) their position on the genome (as opposed to molecular genetics).

History of animal breeding

Animal breeding aims in ‘t*he development of animals, that will result in animal products economically more advantageous under the current environmental, social and economic conditions*’’ [11].

A breeding programme is an organization or a system in which 1) information of potential breeding animals is collected and used to estimate the breeding values (EBV; an estimate of the additive genetic merit for a particular trait that an individual will pass on to its descendants) and 2) genetically superior animals are selected and mated to produce the next generation [12].

The starting point of a breeding program is the definition of the breeding goal [11]. A breeding goal can be defined as “*the development of future animals that can produce in a more effective way under the future production conditions*”. Important tools for the achievement of the breeding goal are [13]:

- Identification of all individuals
- Pedigree construction
- Recording system of the desirable traits
- Establishment of reproduction techniques

One of the most remarkable changes in (dairy cattle) breeding programs was the implementation of progeny test. Progeny test was firstly implemented in Denmark and very soon spread out all over the world [14].

However, the first revolution in animal breeding came with the use of artificial insemination (A.I.). Artificial insemination aided in the rapid expansion of higher genetic material to the entire population. The history of A.I. closes over two centuries, while the widespread use in farms about 80 years [15].

Regarding the mathematical (statistical) part, we have to refer to the major contribution of Henderson for creating a statistical model known as best linear unbiased prediction (BLUP) [16-21]. BLUP uses all available information, through the additive genetic relationships of individuals, for an accurate prediction of the genetic value of the animal. For a short overview of the history of statistical models that have been used in breeding animals see [22].

It should be noted, though, that progeny test alone did not contribute a lot to the genetic progress. It was the progeny testing in connection to A.I. that resulted in a fast genetic progress [23][24].

Genetic improvement of a whole population, which may consist of millions of animals, is usually impractical in all animals in the population, due to a demanding organization and high costs. Therefore, breeding programs are usually constructed in a hierarchical - pyramidical way, where animals are split into different groups and receive different controls (Figure 2).

In the small group of animals at the top of the pyramidical structure (nucleus) all selective processes are implemented. Then, the genetic superiority of the nucleus is transferred to the rest of the population (commercial) [25]. Usually, between the nucleus and the commercial part an intermediate group is included, the multipliers part [11].

The contemporary breeding programs of (dairy) cattle are based on progeny testing of male and assortative mating [26]. The breeding program is organized in a pyramidical structure and the genetic superiority is transferred to the commercial population through the males and A.I. [25].

An effective dairy breeding program should include:

- Progeny test
- Accurate estimation of EBV
- Identification of the genetical superior males and females and organize assortative matings with the appropriate caution on the inbreeding level
- Fast distribution of the superior genetic material to the whole population

However, progeny test is a very intensive and costly technique requiring long time for results of the elite animals. For example, it takes, approximately, 6 to 7 years for a dairy young bull to be proved as superior and selected for breeding. During all this time the animals are kept in the breeding stations and fed, waiting for the results of progeny test.

The use of molecular markers in animal breeding

The idea of using molecular markers in animal breeding was around for quite some time [27]. During the '90s there was an attempt to incorporate genomic markers in breeding values predictions. This is what is called Marker Assisted Selection (MAS) and the theoretical background was firstly introduced by Fernando and Grossman in 1989 [28]. In MAS genetic markers and BLUP model are combined for the prediction of the EBV of the animals.

The research for the identification and location of genomic regions that control quantitative traits (quatitative trait loci; QTL) (e.g. mik yield) was quite enthousiastic at the beginning and lead to a pleithora of studies, mainly during the '90. Thus, QTL databases were developed (e.g. Animal QTL database; http://www.animalgenome.org/cgi-bin/QTLdb/index) in which someone can freely search for already identified QTL on desirable traits. However, specific limitations of MAS such as i) small number of markers available at that time, ii) low fraction of genetic variation explained by QTLs and iii) the fact that results were only descriptive for the specific sire family where QTL analysis had been performed, resulted in difficulties in implementing MAS in practice. For more information in MAS see [29-31].

Genomic selection

The limitations of MAS have been overcome by the recent developments in molecular technology. Nowadays, it is provided the possibility of sequencing the entire genome and discover a large number of genetic markers in the form of single polymorphisms (Single Nucleotide Polymorphisms, SNPs). Thus, an animal can be genotyped for tens or hundreds of thousands (or millions) SNPs at a cost of ~100$ (Figure 3).

Microarrays allow us to conduct checks for associations with production traits in many more parts of the genome. In this way, we have the possibility of exploring the entire genome and quantify at a larger extend the genetic variability among individuals.

In 2001, Meuwissen, Hayes and Goddard [32] showed through simulations that 50,000 genome-wide dense markers, equally distributed along the genome, can adequately be used to predict breeding values for animals with a considerably high accuracy. This idea is what is called ***Genomic Selection*** (***GS*; also known under various names such as whole genome enabled predictions; genomic predictions; etc.**). In *GS*, DNA information is used to predict the genetic merit of young animals. The key point in *GS* is that with a genome-wide panel of dense markers all quantitative trait loci (location of a gene on the chromosomes that affects a quantitative trait; QTLs) are in linkage disequilibrium (non-random association of alleles at two or more loci; LD) with at least one marker.

The idea of *GS* is quite simple (Figure 4). In practice, *GS* involves two steps. First, the effect of each marker (SNP) is estimated in a reference population consisting of animals with both known phenotypes and marker genotypes. In the second step, genomic breeding values (GEBV) of young animals are calculated by using only their marker information, and subsequently ranked for selection (Figure 5). It should be noted that *GS* does *not* cancel the phenotypic recording; on the contrary, accurate phenotyping is a key point to develop accurate estimates of SNP effects and thereby accurate genomic predictions of GEBV. In addition, *GS* has nothing to do with genes!! Phenotypes (or EBVs) are simply regressed on all available SNPs in a (non-) linear model.

The GEBV of the individuals is calculated through the following model:

$$y_i = \mu + \sum_{j=1}^{p} x_{ij} b_j + e_i \text{ where } e \text{ are iid } \sim N(0, \mathbf{I}\sigma_e^2),$$

y is a vector of phenotypic records (or EBVs), μ is the overall mean, x is the code of genotype for SNP j and b is the additive effect of SNP j.

This implies that at birth we can have an estimation of the genetic merit of the young bulls which further helps to the correct ranking. Therefore, the progeny testing period of 6-7 years is overcome.

In animal breeding programs the predicted genetic gain (ΔG) can be calculated from the following equation, known as the breeders' equation [33]:

$$\Delta G_T = \frac{\sigma_A \times i \times r_{AI}}{T}$$

1. accuracy of the index, r_{AI}.
2. generation interval, T
3. selection intensity, i.
4. genetic standard deviation, σ_A.

The major contribution of *GS* in animal breeding programs is referred to:

- reduction of generation intervals
- increase of the accuracy of EBV at a young age and for difficult to be measure traits (e.g. carcass quality) as well as traits with low heritability (e.g. somatic cell score).

Schaeffer [34] showed that *GS* could substantially increase the rate of genetic progress in dairy cows, allowing the selection of bulls for breeding at a younger age based on the GEBV and without having to wait for progeny testing. More specifically, Schaeffer explored the possibility of applying *genomic selection* in Holstein cattle population in Canada and investigated the economic as well as genetic benefits. The

results showed a large decrease in the interval when selecting bulls (sire bulls) from 6.5 years to 1.75 while the total space of selection (i.e. the four pathways of selection) decreased from 21.75 to 9.75 years. Shaeffer [34] and Pryce et al [35] concluded that *GS* has the potential of double genetic gain compared to conventional progeny schemes. Therefore, *GS* has fairly been described as **the most promising molecular application in livestock** [36]. Just to pin point the importance of the new technique, during the world congress on genetics applied to animal production that takes place every four years (thus also known as the "Olympic Games" of animal breeding) in Leipzig (Germany), more than one third of the presented studies (276 studies out of a total of 846) deal with *GS* [37].

Genomic selection has already been implemented in breeding programs worldwide, mainly of dairy cows [4-8], while an extensive literature can also be found on other animal species like sheep [38, 39], goats [40, 41], poultry [42-44], swine [45], as well as in plants [46, 47] and forestry species [48, 49].

Genomic predictions in humans?

The idea of predicting the genetic background of individuals solely based on DNA information is not limited in animal and plant species, but can also be extended in human studies [50, 51]. For instance, in a recent study a genomic prediction method was used to predict the possibility of skin cancer in humans with encouraging results [52]. Therefore, the method is expected to contribute in a better understanding of the genetic background of quantitative traits as well as human diseases.

A short but important note here: the term *genomic selection* was firstly used for animal breeding programs, where genetically superior animals that have been identified through DNA information, can be selected and mated to produce the next generation. Therefore, the term *GS* is descriptive for the whole stage of selection. The statistical methodology used with the use of molecular markers to predict the genetic merit of the individuals could be better defined as "*genomic predictions (GP)*" or "*whole genome enabled predictions*", etc., term that better fits for human studies.

As already mentioned above, *GP* do not require gene detections or any other kind information on genes. At this point, a small parenthesis could be opened to refer to new knowledge obtained the last few years on human genomics. For example, with

the accumulation of new knowledge derived from the ENCODE Project Consortium (ENCODE) even the basic definition of a *"gene"* has been questioned leading to a broader use of the term [53] (http://www.youtube.com/playlist?list=PL1ay9ko4A8slDIOZvtYjTys_BTDc0klkS). Moreover, it has already been demonstrated and realized that identification of *genes* is not as an easy topic as has been, till recently, thought [54-57].

Some interesting questions on *GS*

How many animals do we need to genotype;

In dairy cattle it seems that a reference population with at least 10,000 phenotyped and genotyped bulls is required to achieve considerably high prediction accuracies. This has led to the establishment of big consortia between breeding cooperatives in Europe to join forces in a common reference population with ~20,000 genotyped bulls.

How many SNP are needed?

This depends on various factors. However, it seems that the 50k SNP works fine in cattle for predictions within breeds. The use of 777k for the moment didn't contribute to any increase in the prediction accuracies.

What is the most restricted factor for genomic predictions, the number of individuals or the number of SNP?

It has been demonstrated that prediction accuracies are increasing much faster with an increase in the number of individuals rather an increase in the number of markers. Note that from statistical point of view the large p small n ($n<<p$) is an unsolved and challenging problem.

Is the prediction across heterotic groups possible?

No. For the moment it seems unlikely and prediction accuracies are close to zero. To be more precise, the question is if it is possible for e.g. to train our model in Brown Swiss cattle and predict Holstein. Maybe the topic is a bit more complicated, for e.g. what is the right question to ask "*why we cannot predict across heterotic groups?*" or "*why should we be able to predict across heterotic groups?".* Let's use an analogy: Imagine we have a population of Chinese people. We take measurements on their height and we genotype them. Is it possible then to predict the height of genotyped Dutch people? Note, however, that all studies that have been carried out so far in plants and animals have been restricted on the amount of observations (few thousands of individuals at maximum).

Are we searching for genes in *GS*?

No. We just make use of all available information (all SNP available) in the genome. The assumption and the underlying theory is that all the genome has a function. However, in

case of identification of genomic regions that contribute a lot to the variability of one trait the information could theoretically be used and incorporated into the statistical prediction models.

What is so exciting about *GS*?

- It doesn't require an extra theory. The theory of *quantitative genetics* perfectly fits to the new discoveries.
- It is easy to apply and in reasonable cost.
- It seems to be profitable for the breeding organizations. It is a technique that offers the opportunity for even double genetic gain (first time in the history of breeding).
- Deepens (but perhaps also complicates) our knowledge around the connection between genomes and phenotypic variability.

Perhaps, though, the most exciting about *GS* is that it is a new (r)evolutionary technique that does not only stay in theory but it is applicable in practice. Perhaps, it complicates a bit the way we thing about genome functionality and *genes* but it give us the unique opportunity to predict based on the genome and early in life of the individuals. We may not understand the underlying mechanism but we can predict the outcomes. In terms of physics it can be considered as an "*effective theory"*. As Prof. Daniel Gianola use to say in his lectures *"are you going to refuse your dinner because you do not understand the digestive system…?".*

References

1. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** Nature Genetics 2001, **409**(6822):860-921.

2. Venter JC, Adams MD, Myers EW, et al. : **The Sequence of the Human Genome.** Science 2001, **291**(5507):1304-1351.

3. The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL, Worley KC: **The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution.** Science 2009, **324**(5926):522-528.

4. Berry PD, Kearney F, Harris BL: **Genomic Selection in Ireland.** In *Proc of the Interbull International Workshop – Genomic Information in Genetic Evaluations; Uppsala, Sweden*. 2009. Bulletin no. 39

5. de Roos APW, Schrooten C, Mullaart E, van der Beek S, de Jong G, Voskamp W: **Genomic Selection at CRV**. In *Proc of the Interbull International Workshop – Genomic Information in Genetic Evaluations; Uppsala, Sweden*. 2009. Bulletin no. 39.

6. Ducrocq V, Fritz S, Guillaume F, Boichard D: **French report on the use of genomic evaluation**. In *Proc of the Interbull International Workshop – Genomic Information in Genetic Evaluations; Uppsala, Sweden*. 2009. Bulletin no. 39.

7. Wiggans GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF, Chesnais JP, Schenkel FS, Tassell CPv: **Genomic evaluations in the United States and Canada: a collaboration**. In *Proceedings of International Commitee of Animal Recording, 16–20 June 2008, Niagara Falls, NY, 6pp.*

8. Loberg A, Dürr JW: **Interbull service on the use of genomic information**. In *Proc of the Interbull International Workshop – Genomic Information in Genetic Evaluations; Uppsala, Sweden*. 2009. Bulletin no. 39.

9. Fisher RA: **XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.** Trans R Soc Edinb 1919, **52**(02):399-433.

10. Wright S: **Correlation and causation.** Journal of agricultural research 1921, **20**(7):557-585.

11. Ρογδάκης Ε: *Γενετική Βελτίωση Αγροτικών Ζώων:* Εκδόσεις Σταμούλης; 2008.

12. Groen A, Van Arendonk J: *Breeding Programmes. Lecture notes for E250 – 210:* Wageningen, The Netherlands: Wageningen University; 1995.

13. Άμπας Ζ: *Πανεπιστημιακές σημειώσεις στο μάθημα "Αειφορική διαχείριση γενετικού υλικού". Πρόγραμμα μεταπτυχιακών σπουδών:* Δημοκρίτειο Πανεπιστήμιο Θράκης; 2009.

14. Johansson I: **Progeny Testing Methods in Europe1.** Journal of dairy science 1960, **43**(5):706-713.

15. Vishwanath R: **Artificial insemination: the state of the art.** Theriogenology 2003, **59**(2):571-584.

16. Henderson CR: *Application of Linear Models in Animal Breeding:* Guelph: University of Guelph; 1984.

17. Henderson CR: **Best Linear Unbiased Prediction of Nonadditive Genetic Merits in Noninbred Populations.** Journal of Animal Science 1985, **60**(1):111-117.

18. Henderson CR: **Best linear unbiased estimation and prediction under a selection model.** Biometrics 1975, :423-447.

19. Henderson CR: **Estimation of Variances and Covariances under Multiple Trait Models.** J Dairy Sci 1984, **67**(7):1581-1589.

20. Henderson CR: **Best Linear Unbiased Prediction of Breeding Values Not in the Model for Records.** J Dairy Sci 1977, **60**(5):783-787.

21. Henderson CR: **Best Linear Unbiased Prediction Using Relationship Matrices Derived from Selected Base Populations.** J Dairy Sci 1985, **68**(2):443-448.

22. Balding DJ, Bishop M, Cannings C: *Handbook of statistical genetics:* John Wiley & Sons; 2008 (p. 678 - 717).

23. Foote RH: **The history of artificial insemination : selected notes and notables.** J Anim Sci 2002, **80**:1-10.

24. Butler LJ (Ed): *Proceedings of the The Potential for Growth in the California Dairy / Forage Industry and Implications of the 2002 Farm Bill: Modesto, California.* 31st California Alfalfa and Forage Symposium; 2002.

25. Kinghorn B, Van der Werf J, Ryan M: *Animal breeding: Use of new technologies.* Post graduate foundation in veterinarian science. University of Sydney; 1999.

26. Sanna SR, Casu S, Carta A (Eds): *Proceedings of the Breeding Programmes in Dairy Sheep: August 19-23, 2002; Montpellier, France.* 7th World Congress on Genetics Applied to Livestock Production; 2002.

27. Neimann-Sorensen A, Robertson A: **The Association between Blood Groups and Several Production Characteristics in Three Danish Cattle Breeds.** Acta Agriculturae Scandinavica 1961, **11**(2):163-196.

28. Fernando RL, Grossman M: **Marker assisted selection using best linear unbiased prediction.** Genetics Selection Evolution 1989, **21**(4):467-477.

29. Boichard D, Fritz S, Rossignol MN, Boscher MY, Malafosse A, Colleau JJ (Eds): *Proceedings of the Implementation of Marker-Assisted Selection: Practical Lessons*

from Dairy Cattle: 2002; Montpellier, France. 7th world congress of genetics applied to livestock production, Communication no. 22-03; .

30. Food and Agriculture Organization of the United Nations: **MARKER-ASSISTED SELECTION. Current status and future perspectives in crops, livestock, forestry and fish.** *F A O* 2007, **Rome:**.

31. Khatkar M, Thomson P, Tammen I, Raadsma H: **Quantitative trait loci mapping in dairy cattle: review and meta-analysis.** Genetics Selection Evolution 2004, **36**(2):163-190.

32. Meuwissen T, Hayes B, Goddard M: **Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.** Genetics 2001, **157**(4):1819-1829.

33. Falconer DS, Longman TFCM: *Introduction to quantitative genetics:* 4th ed. Pearson Prentice Hall; 1996.

34. Schaeffer L: **Strategy for applying genome-wide selection in dairy cattle.** J Anim Breed Genet 2006, **123**:218-223.

35. Pryce JE, Goddard ME, Hayes BJ (Eds): *Proceedings of the Breeding Schemes for Dairy Cows Under Genomic Selection – what can we do? August 1-6; Leipzig, Germany.* 9th World Congress on Genetics Applied to Livestock Production; 2010.

36. Sellner EM, Kim JW, McClure MC, Taylor KH, Schnabel RD, Taylor JF: **Board-invited review: Applications of genomic information in livestock.** Journal of Animal Science 2007, **85**(12):3148-3158.

37. Habier D: **More than a third of the WCGALP presentations on genomic selection.** J Anim Breed Genet 2010, **127**(5):336-337.

38. Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc J-, Barillet F, Robert-Granié C, Manfredi E: **Genomic selection in the French Lacaune dairy sheep breed.** J Dairy Sci 2012, **95**(5):2723-2733.

39. Salaris S, Casu S, Usai MG, Fresi P, Carta A: **Evaluating the accuracy of the genetic ranking of rams in the selected population of the Sarda dairy sheep breed.** ICAR Technical Series 2010, **14**:19-22.

40. Carillier C, Larroque H, Palhière I, Clément V, Rupp R, Robert-Granié C: **A first step toward genomic selection in the multi-breed French dairy goat population.** J Dairy Sci 2013, **96**(11):7294-7305.

41. Shumbusho F, Raoul J, Astruc JM, Palhiere I, Elsen JM: **Potential benefits of genomic selection on genetic gain of small ruminant breeding programs.** J Anim Sci 2013, **91**(8):3644-3657.

42. Abdollahi-Arpanahi R, Pakdel A, Nejati-Javaremi A, Moradi-Shahrbabak M, Morota G, Valente BD, Kranis A, Rosa G, Gianola D: **Dissection of additive genetic**

variability for quantitative traits in chickens using SNP markers. J Anim Breed Genet 2014, .

43. Fulton JE: **Genomic selection for poultry breeding.** Anim Front 2012, **2**:30-36.

44. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P, Hocking PM, Fife M, Salmon N, Fulton J, Strom TM, Haberer G, Weigend S, Preisinger R, Gholami M, Qanbari S, Simianer H, Watson KA, Woolliams JA, Burt DW: **Development of a high density 600K SNP genotyping array for chicken.** BMC Genomics 2013, **14**:59.

45. Albers GAA: **Genomic selection in poultry and pig breeding: a breakthrough technology?** Advances in Animal Biosciences 2010, **1**(01):359-359.

46. Albrecht T, Wimmer V, Auinger H, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C: **Genome-based prediction of testcross values in maize.** Theor Appl Genet 2011, **123**(2):339-350.

47. Heslot N, Yang HP, Sorrells ME, Jannink JL: **Genomic selection in plant breeding: a comparison of models.** Crop Sci 2012, **52**:146-160.

48. Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J: **Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats.** Plant Gen. 2011, **4**(2):132-144.

49. Grattapaglia D, Resende MD: **Genomic selection in forest tree breeding.** Tree Genetics & Genomes 2011, **7**(2):241-255.

50. de lC, Gianola D, Allison DB: **Predicting genetic predisposition in humans: the promise of whole-genome markers.** Nat Rev Genet 2010, **11**(12):880-886.

51. de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D: **Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor.** PLoS Genet 2013, **9**(7):e1003608.

52. Vazquez AI, de los Campos G, Klimentidis YC, Rosa GJM, Gianola D, Yi N, Allison DB: **A Comprehensive Genetic Approach for Improving Prediction of Skin Cancer Risk in Humans.** Genetics 2012, **192**(4):1493-1502.

53. Stamatoyannopoulos JA: **What does our genome encode?** Genome Res 2012, **22**(9):1602-1611.

54. Ioannidis J: **Genetic associations: false or true?** Trends Mol Med 2003, **9**(4):135-138.

55. Pavlidis P, Jensen JD, Stephan W, Stamatakis A: **A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans.** Molecular Biology and Evolution 2012, .

56. Ioannidis JP: **Why most discovered true associations are inflated.** Epidemiology 2008, **19**(5):640-648.

57. Ioannidis JP: **This I believe in genetics: discovery can be a nuisance, replication is science, implementation matters.** Frontiers in genetics 2013, **4**.

58. Δαδούσης Χ: **Αξιολόγηση Προγραμμάτων Γενετικής Βελτίωσης στο Πρόβατο Φυλής Χίου με τη Χρήση Γονιδιωματικής Επιλογής.** *MSc.* Δ.Π.Θ., Ορεστιάδα; 2010.

Figure 1 History of whole genome sequencing of various species.

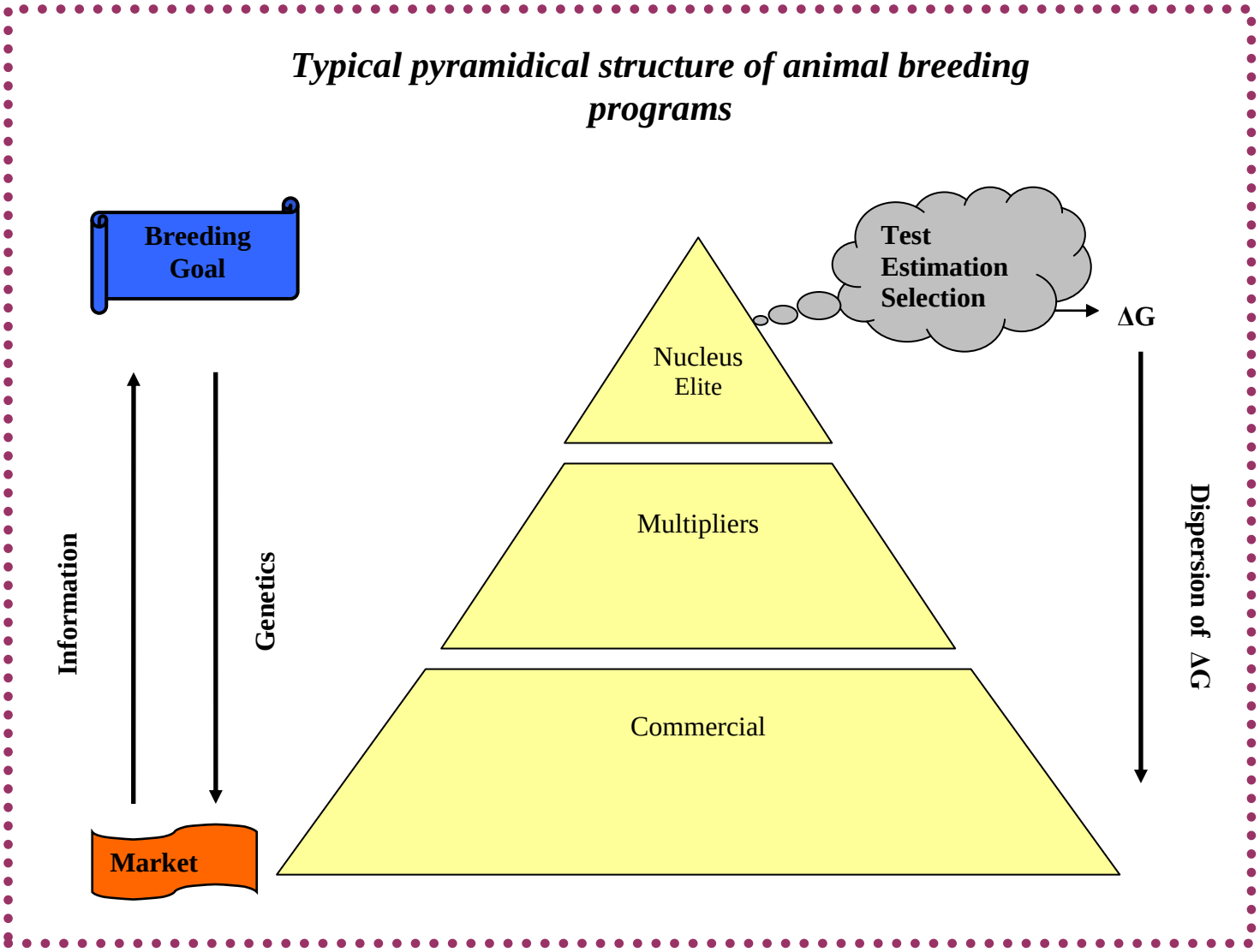
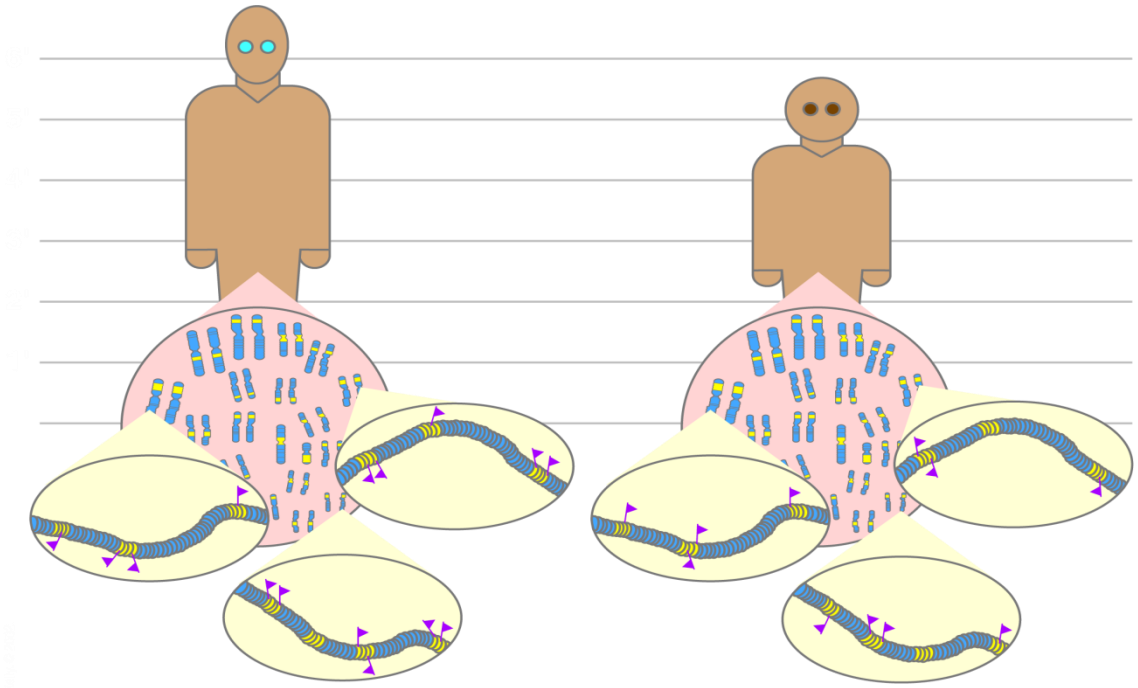


Figure 2. Description of a typical pyramidical structure animal breeding program.

Figure 3. llumina 50k bovine chip.



Estimation of an effect (g) for each of the markers distributed along the genome such that their sum will produce the phenotype or the breeding value.

$$GEBV = \sum_{j=1}^{n} S_j g_j$$

$S: SNP$
$g: SNP\ effect$

Figure 4. Brief description of the basic idea behind genomic selection.

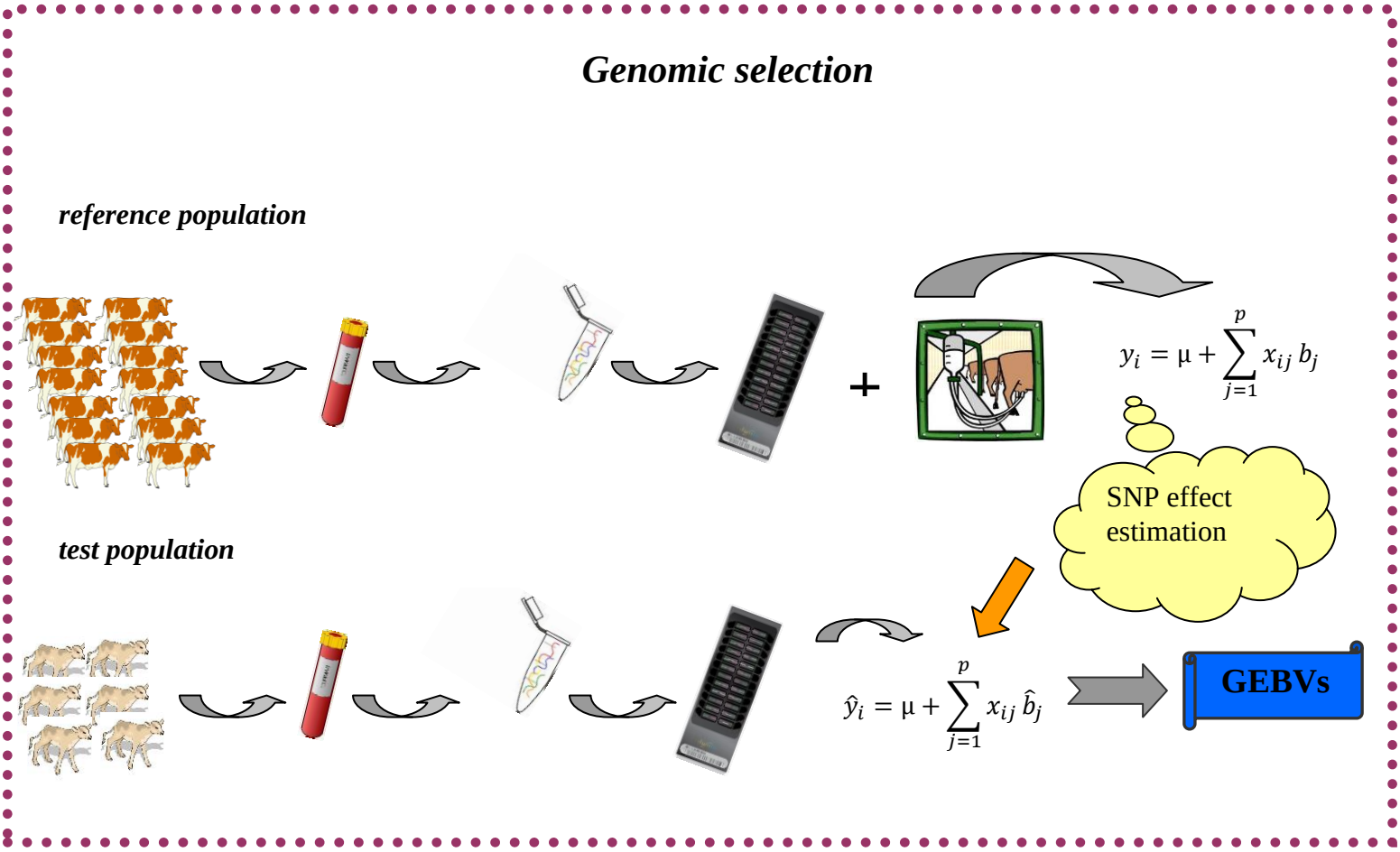


Figure 5. Brief description of genomic selection schemes.