**Appendix 1: Pre-proposal format**

**The use of comparative genomics in animal breeding**

X Animal Breeding and Genetics Ο Adaptation Physiology

Ο Animal Nutrition Ο Quantitative Veterinary Epidemiology

Ο Animal Production Systems Ο Human and Animal Physiology

Ο Cell Biology and Immunology Ο Experimental Zoology

Ο Aquaculture and Fisheries Ο Aquatic Ecology and Water Quality Management

Ο Other, namely...

**1a. Details of the applicant**

Name: Fernando

Address: Hoevestein 253 13B, 6708 RZ, Wageningen

Nationality: Spanish

Gender: Male

Previous degree: Master in Agricultural Engineering

Obtained at: Valladolid University (Spain)

Main subject: Animal Breeding and Genetics

**1b. Institute where the research will be executed**

Name: Animal Breeding and Genetics Chair Group (Wageningen University)

Address: Droevendaalsesteeg 1, Gebouw 107. 6708 PB Wageningen

### Name and titles of intended supervisors: Dr.ir. Hendrik-Jan Megens, Pr.dr. Martin Groenen

**2. Popular summary of the application**

2a. summary of the intended research (max 0.5 page, min 10 pts)

In livestock, animals with the best genetic value are selected as parents for the next generation. This “artificial selection” enables the spread of favourable genetic material across the populations. Before the development of sequence technology, the genetic value of animals used to be estimated from performance and pedigree records. Nowadays, the DNA sequence (genome) of any animal can be known, and we provide acceptable estimates of genetic values at early ages.

Since the DNA contains the instructions for the cell’s functioning, information derived from the genomes of individuals is very valuable. However, sequence data are massive, and it is very difficult to know what regions in the genome cause what function. At present, sequence information is merely used as a heuristic tool to compare animals. Because we know the DNA of “good” and “bad” animals for a particular phenotypic trait, we can assemble a “most desired genome”. Then, we can score the animals based on how much their genome resembles that particular genome. This approach provides reasonably accurate results, but it is possible to make even better use of genomic information.

The genome contains the instructions for life. Therefore, if the location and function of each DNA element were known, we might be able to predict the genetic value of animals with perfect precision. At present, however, the regulatory networks of the DNA remain mostly unknown. Knowledge about the relative importance of certain DNA elements is accessible, though. For instance, the DNA segments that are well preserved across species (the so-called conserved regions) are expected to be involved in important biological mechanisms. In practice, only the DNA elements that show variation within a population are useful for genetic merit estimation. Among the elements with variation, we expect that those that are better preserved across species are more likely to be functional, and therefore are more important. The aim of this project is to investigate whether accounting for the degree of constraint of DNA elements contributes to a better estimation of genetic value of pigs (*sus scrofa*).

2b. summary of the societal significance and utilization (max 0.5 page, min 10 pts)

The aim of the proposed project is to develop methodologies that improve genomic prediction (GP). This means that we may make a step forward in the comprehension of the expression sequences underlying the genome. Applications include disease discovery in humans and genetic improvement of animals and plants. When genetic diseases are discovered at early stages, it becomes easier to deal with them. The project aims to provide knowledge about the relative importance or even the functionality, of certain DNA elements. An important aspect is that part of this information can be transferred across species (i.e. from pigs to humans). Thus, genomic annotations such as provided in this project may be used for drug development in humans.

Today, the goal of agriculture is to produce more nutritious food in a sustainable way. With GP, it is possible to select, as parents, the animals/plants with favourable genetic material at an earlier stage. In this way, the genetic pool of the populations can be improved. In the farms, accurate GP of animals is used to produce more nutritious food by housing fewer animals.

Another important aspect of the proposed project is that the methodologies to be developed are expected to be more effective in multi-breed genomic prediction. This implies that particularly the local breeds may benefit. Such methodologies will make GP less dependent on the number of individuals. Local breeds are good for the environment and run on low input, and they constitute important sources of genetic material. Nevertheless, they are threatened by progressive marginalization in favor of commercial lines. By making better use of the local breeds, we can guarantee their existence.

We will use genomic data from breeds of pig (*sus scrofa*) because they are available for this project. By re-utilizing tools developed in previous research and by re-using annotations that are freely available, the proposed project will contribute to the improvement of these resources.

**3. Has the student been involved in the subject before?** (max 0.5 page, min 10 pts)

This project combines two disciplines in animal science: animal breeding and genomics. The experiences described here will show that I have a good understanding of both areas of research.

I have done a Master’s thesis titled “Selection of a subset of loci for local bovine breeds identification”. In this, I performed a genome-wide association study (GWAS) (qualification 9.1). Further, I have completed another Master’s thesis relating to pig breeding (“Genetic correlation between sow longevity and litter size variability” – qualification 7). I have followed all courses of the Animal Breeding and Genetics chair group at Wageningen University (ABG-WU) with an average score of 7.8. These courses include: Genetic Improvement in Livestock; Genomics; and Modern Statistics. I have a solid background in statistics and I have worked as professor’s aid in the “Advanced Statistics” course. Before the start of the project, I intend to follow the “Advanced Bioinformatics” course and do a major thesis in comparative genomics.

There are several reasons why I am very enthusiastic about this project. First, it combines two of my main interests (genomic selection and comparative genomics). Second, it offers me the opportunity to collaborate with some eminent scientists at ABG-WU. Third, the research may be helpful for local breeds. The genetic improvement of local breeds is of great interest for me, and I have two communications accepted in International Congresses of research applied to this purposes.

As knowledge in functional genomics increases, professionals with a solid background in quantitative genetics and genomics will be in demand. The convergence between the two disciplines will enable the valuable efforts in the field of genomics to be put into practice.

**4. Detailed description of research area and research plan** (**max 4 pages, min 10 pts**)

**Background**

Nowadays, genomic prediction (GP) is the technique preferred by breeding organizations to estimate the breeding value of animals. The development of dense panels of single nucleotide polymorphism (SNP) markers provides good insight on the genetic variation of animals. Subsequently, GP has become accurate. Currently, GP does not seem to improve with more dense panels of markers or whole-genome sequence data [1]. The reason is that our proficiency to generate data from the genome has exceeded our ability to process it. The number of potential explanatory variables (SNPs) and their possible interactions is now much larger than the number of animals (i>>n). It is important to overcome the problem an of excessive number of markers in order to accurately predict genetic diseases in humans, or breeding values in animals and plants.

Pre-selection of variants is a renewed topic in genetics. The expected number of SNPs with a high effect on the phenotypic traits is expected to be low. Therefore, if the SNPs were adequately selected, data processing would be simpler. Current researches are aiming to infer the causality of the SNPs on the phenotypes. This allows to subset SNPs, and also to prioritize them. The SNPs can receive different values in the prediction models based on their estimated causal effect. These values are called “priors”, and when estimated accurately, the accuracy of GP can increase. Attempts to define priors based on inferred causality have not been very successful [1]. The reason is that understanding causality is very difficult. A more straightforward approach is to define priors based on the relative importance of the SNPs. These priors refer to the probability that the SNP is causal and the estimated size of the effect. Values of relative importance are more robust than values of causal effect. Moreover, they can provide some insight on the functionality of DNA elements. “Comparative genomics” can be used to infer this relative importance.

In comparative genomics, the genome of different species is compared. Hence, it is possible to find genomic segments that remain constant through evolution (constrained elements). These are expected to be involved in important biological functions [2]. In an era of ample sequence data and with the progress in functional annotation, comparative genomics is emerging as a primary research tool in genomics [3]. In livestock species, there are annotations derived from comparative genomics analyses (conservation scores). To our knowledge, comparative genomics has never been considered to deal with the problem of i>>n. The proposed project aims to investigate the potential of comparative genomics to estimate priors for the SNPs, subset markers with high relative importance and increase the accuracy of GP. This will enable a more accurate selection of animals to be made according to our interests. A second aim is to use GP gain understanding on the regulatory networks underlying the genome. This information could be used in further researches to infer causality.

**Perspective of improvement**

The scope of improvement of GP using priors remains unknown. To our knowledge, the increase in accuracy of GP using priors ranges between ~1% [4] and ~22% [5] (ignoring studies with simulated data). These estimates of accuracy strongly depend on many factors (i.e. the traits of study, the breeds, the population size in the phase of model trainning and in the validation, the models considered…). Pérez-Enciso et al. [1] reported that the use of priors is required to fully utilize the advantages of whole-genome data. Among these advantages: more accuracy, more efficiency in small populations (local breeds) [6], and more protection against changes of generations within the populations. However, the improvement of GP using priors strongly depends on the accuracy of the priors [1]. Therefore the assignment of priors should be a meticulous process. Some positive aspects of this project are that comparative genomics has been successfully used for functional annotation in human medicine [3]; that comparative genomics can be particularly powerful in livestock because breeds within the same species are often very divergent (due to breeding practices); and that the quality of the priors will be verified with genome-wide association studies (GWAS).

**Formulation of the problem**

The main objective of this project is to develop methodologies that can optimally utilize data derived from comparative genomics between species and between breeds to improve the accuracy of GP. The efficient use of priors is expected to improve genomic prediction (GP). The main research questions are: How can we utilize comparative genomics to reduce the problem of i>>n? Moreover, can GP be used to gain understanding on the regulatory networks underlying the genome? This research question is further divided into four objectives:

1. How can we integrate GWAS and conservation scores to estimate priors for GP?

2. How can we assign breed-specific priors using comparative genomics?

3. How can we re-estimate the relative importance of the SNPs using results from GP?

4. What DNA elements are more affected by breeding?

**Methodology**

In this section, the activities are described. Each activity corresponds to one objective. Activities 1 and 3 will result in one paper each, whereas Activity 2 will result in 2 papers. We will develop methodologies for pig populations (*sus scrofa)* because muti-breed data is available for this project.

**Activity 1: Combining conservation scores and GWAS to estimate priors**

The objective of this activity is to use two independent sources of information: statistical association (GWAS) and inter-species conservation scores to obtain priors for GP. These priors can then be used to deal with the problem of i>>n. Conservation scores will be obtained from Genomic Evolutionary Rate Profiling (GERP scores) [2]. These scores are computed by quantifying substitution deficits in constrained elements. GERP scores for *sus scrofa* are available in USCS (University of California Santa Cruz) (http://genome.ucsc.edu).

**Combining GERP scores with GWAS**

We will integrate the GERP scores and GWAS into a single score, This enables: (1) to validate two independent sources of information (GERP and GWAS), (2) to define priors based on both sources of information.

We will choose 4 traits among the broad spectrum in pig breeding. One from each class: fertility, meat quality, conformation and type. The difference between the traits may be worthwhile to consider (i.e. fertility traits may be subject to natural selection). Second, we will perform GWAS for the elected traits. For this, sequence data (or imputed to sequence) that is available for this project will be pooled (~20000 animals from 19 breeds of *sus crofa).* The number of genotypes is substantially larger for the commercial lines, but GWAS is expected to be valid for all breeds [6].

Data for commercial lines will be provided by Topigs-Norsvin (Table 1). Whole-genome sequence data for local breeds is available for research groups in the ABG-WU (19 breeds) [7]. Data in Table 1 was imputed to sequence. It is expected that the number of genotyped individuals in both datasets will increase.

Table 1. Number of genotyped animals in the commercial lines. Data for GWAS.

|  |  |  |
| --- | --- | --- |
| **Breed** | **Number of individuals** | **SNP chip** |
| Large White | 3565 | 60-80-660K and imputed to sequence |
| Dutch Landrace | ~3100 | 80-660K and imputed to sequence |
| Norwegian Landrace | ~7600 | 80-660K |
| Duroc | ~5000 | 80-660K and imputed to sequence |
| Large White | 183 | 80K and sequence |
| Dutch Landrace | 151 | 80K and sequence |
| Duroc | 162 | 80K and sequence |

To combine SNP effect from both information sources, we will prioritize the SNPs based on the degree of overlap between GWAS and the GERP scores. This step involves efforts to find suitable thresholds. We will define different criteria of overlap GWAS-GERP for both approaches. With each criteria, we will estimate the effect of the markers as explained in Gao et al. [8] and we will verify by estimating the increment in accuracy of GP.

To estimate the accuracy of GP, we will evaluate 5-10 scenarios based on statistical model and the definition of priors (i.e. criteria of subsetting, whether we make classes for the priors or not, how larger the differences between the priors are...). For this, we will make use of the methodologies developed by ongoing research in the ABG-WU. These methodologies are developed with the purpose of integrating information from the different genome annotations with GWAS and can be used in the proposed project. Integration of priors and GWAS has already been done in human medicine [3].

To accelerate the verifications, we will assess the advantages and limitations of using split-and-merge [9], Particularly, in the scenarios with a large number of SNPs, depending on increment on accuracy of GP. Once we optimized the methodologies for adequate election of criteria of GERP-GWAS overlap, we will obtain a final list of priors for the SNPs based on their estimated “relevant importance”.

If time allows, we could estimate the correlation between GERP scores and some of the commercial traits. Moreover, we could try to protect the methodologies against the effect of the different traits. Another aspect to consider is that estimating the increase in accuracy, directly from GWAS scores (ignoring GERP) could give us some insights on the added value of the integration GWAS-GERP.

**Activity 2a: Comparative genomics between breeds**

Comparative genomics will be performed for six distant breeds to estimate values of conservation between breeds. Pairwise comparisons will be made in Ensembl API, by mapping enriched ChIP-seq regions between species in a reciprocal manner using whole-genome alignments, as explained in Diego Villar et al. [10]. To estimate values of conservation, we will search for segments that show deficit of nucleotide substitution (constrained elements) and we will measure these deficits. Starting criteria for “constraint” will be 50% in length across species. Finding constrained elements in regulatory regions will be particularly challenging because they are less conserved [3]. Therefore, we may need to relax the requirement for constraint. Supportive material for the non-coding comparative genomics are the extensive RNAseq resources and other type data sets provided by ENCODE (http://encode.com). Filtering of data will be required to remove false positives/artefacts that result from the draft nature of current reference genomes and their incomplete annotation.

We will assign conservation scores to the SNPs located in constrained elements, based on:

* The degree of constraint. For instance, some SNPs may be fixed in most breeds and show variation in other few breeds. Probably the breeds that show variation acquired a new allele due to introgression from other distant breeds. Bosse et al. [11] showed that introgression signatures seem to be enriched for functionality.
* The recombination rate. If recombination rate is large, the constrained elements will be less conserved, and their relative importance will be lower.
* The location of the SNPs (i.e. coding or regulatory regions, synonymous or antonymous substitution, gene ontology...)

**Activity 2b: Obtaining breed-specific priors**

The conservation scores from Activity 2a will be combined with GWAS using tools developed in Activity 1. Hence, we will obtain breed-specific priors. Next, these breed-specific priors will be combined with the priors for all breeds of *sus crofas* (Activity 1), to obtain more complete priors that differ between breeds. For this, we will set different criteria of addition, and we will verify with increments in accuracy of multi-breed GP. To train the GP model, we will use five of the breeds considered in Activity 2a, and we will validate with the remaining breed.

**Activity 3: Re-estimating relative importance of the SNPs based on GP validation**

For this, we will use the 4 commercial lines (Table 1) because the number of individuals is only in these lines the number of animals is large enough to perform uni-breed GP. We will use the priors from Activity 1, so one unique prior for each SNP in the 4 lines. Following from Activity 1, we will calculate the increase in accuracy for the 4 commercial lines. By comparing the catalogues of variation of the different breeds and accounting for the priors in Activity 1, we will investigate which SNPs are responsible for the difference in increase in accuracy between the commercial lines. Then, accounting for the priors obtained in Activity 1, we will re-estimate the relative importance of the SNPs. These re-estimated values cannot be used for GP because they were accounted for them in the prediction models. Instead, they can be used as “feed-back” into understanding of genotype-phenotype interactions. The new values of relative importance are expected to be more reliable that the previous ones (based only on conservation). These values could be used in future research to estimate functionality. Here, we will correct as much as possible for allele substitution effects and differences in Linkage Desequilibrium between the breeds.

**Activity 4. Investigating which types of DNA elements are more affected by breeding**

In this activity, the results from the previous activities are used to investigate which type of DNA elements (i.e. enhancers, promoters, coding regions...) are more subject of “selective sweep” when the population is exposed to artificial selection. “Selective sweep” is the fixation of one of the alleles in a SNP that occurs when the allele concedes advantages to the carriers. Moreover, we intend to identify which type of elements are conserved in the absence of selective sweep, and which regions are more involved in generating new variation. Conserved regions and selective sweeps are accounted for in the GERP scores, whereas more variant regions can only be detected through GP validation. SNPs with low values of conservation between breeds and high values of conservation between breeds may be candidates for selective sweep.

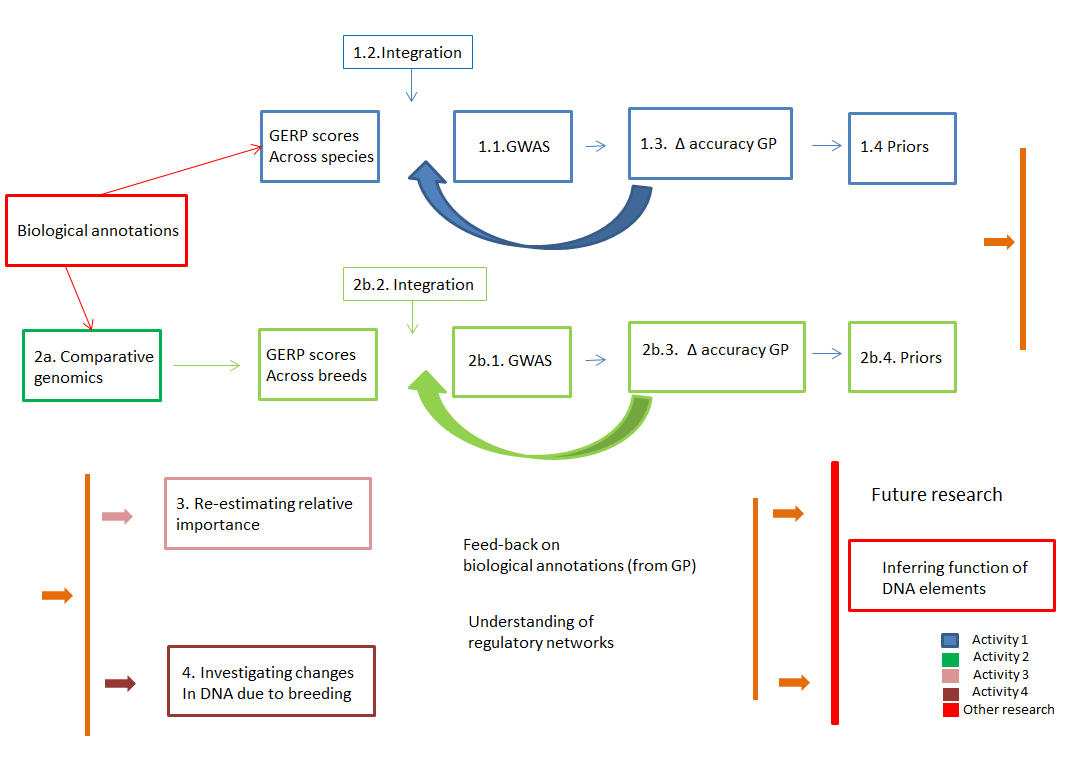


Figure . Scheme of project

**5. Timetable of the project and working programme**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Activity** | **Year 1** | | | | **Year 2** | | | | **Year 3** | | | | **Year 4** | | | |
| 1.Combining conservation scores and GWAS to estimate priors |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2a.Comparative genomics between breeds |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2b.Obtaining breed-specific priors |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3.Re-estimating relative importance using results from GP |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4.Investigating which DNA is more affected by breeding |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Finalize thesis |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**6a. Present infrastructure**

The ABG-WU and Biometris at Wageningen University have the required expertise and knowledge on genomics, statistical genetics, software and computer facilities. All analyses within this project are computational analyses of whole-genome data sets. State of the art computing facilities are available via access to High Performance Cluster (HPC). This HPC consists of 896 cores, and includes two fat nodes (64 cores each) with 1 Tb of memory for each of the 2 nodes. Furthermore, WU is connected to the Dutch Life Science Grid with access to an additional high memory machine (2 Tb).

Genomic data of commercial pig lines will be made available by Topigs-Norsvin. Most of this data has already been imputed to sequence in the ABG-WU. Sequence data of 19 pig breeds is available on a large parallel file system on the HPC (600 Tb) from previous research in the ABG-WU and it is directly accessible for our analyses.

GERP scores from the comparative analysis between 21 species is freely available as well as genome annotations (http://genome.ucsc.edu).

**6b. Affiliation with (inter)national research programmes**

Currently there are three projects from STW-Breed4Food Partnership Programme, that are doing related research in ABG-WU:

* STW-Breed4Food Partnership Programme. Project 14283: From sequence to phenotype: detecting deleterious variation by prediction of functionality.
* Method of association study for rare variants using NGS.
* Towards Precision Breeding using genomic prediction.

We intend to collaborate with these projects.

My intended supervisor: Prof Groenen is head of genomics research within the Animal Breeding and Genomics Centre (ABG-WU) at Wageningen University (WU) with a focus on the characterization of molecular variation underlying phenotypic variation in livestock species. Dr.Hendrik-Jan Megens have an excellent track record in the field of genetic architecture. In the same department, Dr. Mario Calus and Prof. Roel Veerkamp are experts in Genomic Prediction.

**7. Societal significance and utilization**

GP is an efficient technique that enables the prediction of the phenotypes based on DNA samples. This is basic for the discovery of genetic diseases in humans. Additionally, GP is also very valuable in agriculture because it enables the identification of animals/plants with favourable genes. These individuals are, then, selected as parents for the next generation in order to improve the genetic pool of the populations.

**Stating the problem**

With the development of sequence technology, the accuracy of GP has improved substantially. Progress has been epic in human medicine, agriculture and other fields such as biodiversity. Since these aspects are very important for society, it is desirable to increase the accuracy of GP as much as possible. Currently, GP does not improve even with the further development of sequence technology. This is because our ability to generate data from DNA analyses has exceeded our ability to process them. New methodologies that simplify data processing are required to solve this problem.

**Suggested solution**

The proposed project aims to develop methodologies that enable the DNA information to be prioritized. The idea is to process only the data that are informative. Several efforts have addressed this before, but results have been disappointing. The proposed project aims to consider the evolutionary marks in the pig genome to prioritize the DNA information. This approach has never been considered before, despite its potential. It is expected that the annotations based on evolutionary marks will be robust and that we will observe an increase in the accuracy of GP.

**Scope of improvement**

The problem of excessive information will certainly not be solved with this project. However, we will explore the potential of one important information source (evolutionary marks). If results are favourable, the project could point to one new direction in the GP research. Furthermore, this research will provide valuable knowledge about the relative importance and “functionality” of DNA elements. Genome annotations can, to some extent, be transferred across species. Therefore, this information could be used in any species to investigate the “causality” of DNA on the phenotype.

**Applications**

A realistic perspective is that the accuracy of GP will increase slightly. Thus, the methodologies could be applied directly to identify carriers of favourable genes in the breeds and traits considered. Such populations could become more efficient in what concerns the elected traits. The methodologies could also be applied to different breeds and traits, as well as to different plants. Subsequently, livestock and plants could respond better to the interests of society (i.e. high quality of food, individuals resistant to disease, docile temperament, high fertility, low piglets’ mortality, etc.). Further research could improve the methodologies and investigate in which conditions (i.e. species, traits, population size, etc.) they are more effective.

There is also the possibility that the accuracy of GP increases greatly according to the data analyzed. In such a case, the methodologies could be further investigated in human genetics, as well as for genetic improvement in plants and animals species.

**8**. **Legal requirements**

Has been complied with the law and legal requirements with respect to the proposed research, such as ‘Wet op Dierproeven’ and ‘DNA-recombinant legislation?

X Yes Ο No

**9a. Requested budget**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **year 1** | **year 2** | **year 3** | **year 4** |
| Personnel (mm) | | 8500 | 12500 | 10000 | 10000 |
| Research costs (k€) | |  |  |  |  |
|  | Equipment | 2 |  |  |  |
|  | Consumables\* | 2 | 2.3 | 2.4 | 2.8 |
|  | Fieldwork |  |  |  |  |

\* max. € 10.000,- p/year

**9b. Explanation and/or remarks to the proposed budget:**

Consumables are:

High Performance Computing (HPC) : 8000

Open Access publications: 1500

**10. Financial assistance from (an)other source(s)**

**11. References**

1. PEREZ-ENCISO, M., RINCON, J. C. & LEGARRA, A. 2015. Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. Genetics Selection Evolution, 47.
2. COOPER, G. M., STONE, E. A., ASIMENOS, G., GREEN, E. D., BATZOGLOU, S., SIDOW, A. & PROGRA, N. C. S. 2005. Distribution and intensity of constraint in mammalian genomic sequence. Genome Research, 15, 901-913.
3. ALFOLDI, J. & LINDBLAD-TOH, K. 2013. Comparative genomics as a tool to understand evolution and disease. Genome Research, 23, 1063-1068.
4. Heidaritabar, M., Calus, M.P.L., Megens, H-J., Vereijken, A., Groenen, M.A.M., Bastiaansen, J.W.M. 2016. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. Journal of Animal Breeding and Genetics, 133(3): 167-179.
5. MACLEOD, I. M., HAYES, B. J. & GODDARD, M. E. 2014. The Effects of Demography and Long-Term Selection on the Accuracy of Genomic Prediction with Sequence Data. Genetics, 198, 1671-+.
6. Veroneze, R., Lopes, P.S., Lopes, M.S., Hidalgo, A.M., Guimarães, S.E.F., Harlizius, B., Knol, E.F., van Arendonk, J.A.M., Silva, F.F. and Bastiaansen, J.W.M. 2016. Accounting for genetic architecture in single- and multipopulation genomic prediction using weights from genomewide association studies in pigs. J. Anim. Breed. Genet, 133, 187–196.
7. HERRERO-MEDRANO, J. M., MEGENS, H. J., GROENEN, M. A. M., BOSSE, M., PEREZ-ENCISO, M. & CROOIJMANS, R. 2014. Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. Bmc Genomics, 15.
8. GAO, N., LI, J. Q., HE, J. L., XIAO, G., LUO, Y. Y., ZHANG, H., CHEN, Z. M. & ZHANG, Z. 2015. Improving accuracy of genomic prediction by genetic architecture based priors in a Bayesian model. Bmc Genetics, 16.
9. CALUS, M. P. L., BOUWMAN, A. C., SCHROOTEN, C. & VEERKAMP, R. F. 2016. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. Genetics Selection Evolution, 48.
10. VILLAR, D., BERTHELOT, C., ALDRIDGE, S., RAYNER, T. F., LUKK, M., PIGNATELLI, M., PARK, T. J., DEAVILLE, R., ERICHSEN, J. T., JASINSKA, A. J., TURNER, J. M. A., BERTELSEN, M. F., MURCHISON, E. P., FLICEK, P. & ODOM, D. T. 2015. Enhancer Evolution across 20 Mammalian Species. Cell, 160, 554-566.
11. BOSSE, M., LOPES, M. S., MADSEN, O., MEGENS, H. J., CROOIJMANS, R., FRANTZ, L. A. F., HARLIZIUS, B., BASTIAANSEN, J. W. M. & GROENEN, M. A. M. 2015. Artificial selection on introduced Asian haplotypes shaped the genetic architecture in European commercial pigs. Proceedings of the Royal Society B-Biological Sciences, 282.

**12. Abbreviations**

ABG-WU: Animal breeding and genetics Wageningen University

GERP: Genomic Evolutionary Rate Profiling

GP: Genomic prediction

GWAS: Genome-wide association study

SNP: Single nucleotide polymorphism