EUROPEAN COMMISSION

H2020 Workshop: Animal genomics

Workshop Report

18/04/2017

Contents

The conclusions, recommendations and opinions presented in this report reflect the opinion of the participants and do not necessarily reflect the opinion of the Commission.

1. Introduction

Animal breeding has played an important role in increasing the production efficiency of animals, in a cumulative way, by on average 2% increase per year in the last 10 years. An increase of the global demand for food of animal origin is expected. Farmed animal production, however, faces a number of challenges, including the impact of climate change on animals and agricultural production systems and the impact of livestock on the environment and climate. There are further challenges concerned with the increasing scarcity of natural resources and feed, or concerns about animal welfare, anti-microbial resistance and genetic diversity.

In the last 10 years, the EU funded a number of projects focussed, wholly or partly, on animal genomics, breeding and genetic resources. The aim of the workshop was to look at the outputs of some of these EU funded projects and at related initiatives to get an idea of the state of play in the domain in order to contribute to the development of a roadmap for future research in this field. It did not dig into every aspect of animal genomics, breeding and genetic resources, e.g. epigenetics, new breeding techniques, reproduction techniques were hardly discussed. The workshop was focussed on terrestrial farmed animals and was structured around 3 areas: animal breeding; genetic resources; and genotype to phenotype. The EU funded projects were clustered around these areas.

Participants in the workshop included: representatives of EU funded projects, EU stakeholders, (mostly industry), some non-EU scientists to contribute to the international dimension and additional experts who gave opening speeches and wrapped up each session.

Presentations are available separately.

2. Animal breeding

The introductory speech by Theo Meuwissen highlighted that animal breeding is one of the tools to address the challenges faced by farm animal production, which demand particularly efficient use of resources, rapid genetic change/improvements, increased robustness of animals, efficient production under changing circumstances. Precision agriculture can help in the efficient use of critical resources, and precision animal breeding is part of this.

Genomic selection was first introduced in dairy cattle, but has now been implemented in the selective breeding in other farmed animals, including aquaculture species. Genomic selection speeds up the selection process dramatically, when the generation interval can be reduced, especially in dairy cattle where genomic selection has virtually replaced progeny testing. Reproductive technology can help this approach when parents can produce many gametes or embryos at a young age.

Furthermore genomic selection can address issues linked to genotype-by-environment interactions. In doing so, it can enable selection to be focussed onproduction under practical environments (including low input), selecting elite animals best suited under these conditions or selecting for traits not measured on elite animals (meat quality, disease resistance).

Precision breeding can improve genome prediction coming from the inclusion of functional information and increase accuracy of genomic estimated breeding value (GEBV), can avoid deleterious side effects from breeding (in particular when selecting for broad breeding goals). It can manage genetic variation (maintain genetic resources). Precision breeding could also be used in the future in non-mainstream breeds.

With precision breeding, a GEBV above 90% can be reached for a broad spectrum of traits in dairy cattle. Moreover it ensures rapid genetic improvement and efficient allocation of animals to different environments/markets. But usually it requires large scale genotyping and phenotyping to make genomic selection work across large genetic distances. Indeed, "In the age of genomic selection, phenotype is king" (Mike Coffey), meaning that measuring and recording of appropriate phenotypes is critical for genomic selection to function accurately.

Because genomic selection requires large reference populations, with many phenotyped and genotyped animals, big breeds are more easy targets. It is even more important for a GEBV to reach above 90%.
In such a situation, there is no obvious need for high density genotyping nor genome sequencing. Methods are already in place for genomic prediction: G-BLUP or SNP-BLUP
Still, issues exist for traits with large-scale recording difficulties, as a limited number of records leads to low accuracy. If the breeding goal is not extensive enough, no precision breeding can be achieved, as precision breeding also implies breeding without deleterious effects (the latter being avoided by applying an extensive breeding goal).

For non-mainstream breeds, with small (reference) populations, across breed genomic prediction is an option. The reference populations can arise from cooperation across breeds or breed crosses. An additional hurdle is that it requires non-linear methods for genomic prediction. In such situation, high density genotyping / whole genome sequence data are needed. This approach can lead to discovery of causal variants or small causal segments and increase knowledge of the biological mechanisms behind the traits and is beneficial for traits that are difficult to record on a large scale, and for applying extensive breeding goals. Additionally, this approach utilises and nurtures animal genetic resources.

Phenotypic variation is produced through a complex web of interactions between genotype and environment, and such a 'genotype–phenotype' map is inaccessible without the detailed phenotypic data that would allow these interactions to be studied. Our ability to characterize phenomes — the full set of phenotypes of an individual — lags behind our ability to characterize genomes[1]. Phenomics, the area of biology concerned with the measurement of phenomes is required to predict many traits. These data could be combined with genotypes to get better GEBV, for a broader spectrum of traits. Robustness is certainly an area of interest that deserves more attention, as it captures the ability of animals to maintain production under disturbing circumstances.

---

[1] Houle, David; Govindaraju, Diddahally R.; Omholt, Stig (2010). "Phenomics: the next challenge". Nature Reviews Genetics. 11 (12): 855–866.

Managing genetic variation aims to maximise genetic progress while restricting the inbreeding at the level of the DNA (using high density Single Nucleotide Polymorphism (SNP) data) to maintain the ability for genetic change in future generations.

In order to maintain fast genetic change, the following perspectives need to be considered:

- looking for high accuracy in genomic selection and using novel reproductive techniques to shorten generation intervals
- using genomic selection by introgression, i.e. combining it with animal genetic resources.
- looking into animal genetic resources, within breed and across breeds.
- Using widespread trait recording

Introgression consists in bringing in an interesting trait from a donor breed (e.g. disease resistance) into a breed which has a better total genetic merit. After a few generations, the cross could outperform the total merit that the more performing breed would have reached without the introgression. Introgression requires a good understanding of the genetic background of the trait of interest. The number of generations of backcrossing prevents application of introgression but this could be overcome by exploiting a combination of genomic selection and advanced reproduction.

In order to enable/facilitate livestock phenomics and the required widespread trait recording:

- Recording techniques exist, such as mid infra-red (MIR), chromatography, gene-expression data, automatic milking systems and feed-intake stations, but additional/novel recording technologies are necessary to record practical data on a large scale (whether in reference populations or in farms) and on novel phenotypes;
- An option is to record indicator traits instead of actual traits (high genetic correlation needed, or compensation of lower correlation by large scale recording);
- Genotyping of animals should be widespread, i.e. recorded animals are genotyped.

Below are the main achievements/challenges identified by the EU projects in that session, in relation to genetics/breeding/phenomics :

ECO-FCE ("A whole-systems approach to optimising feed efficiency and reducing the ecological footprint of monogastrics") - http://www.eco-fce.eu/ - presented by Klaus Wimmer
The objectives were: to use current knowledge and investigating new feeding strategies; to identify genetic factors for improved feed efficiency; to develop new statistical models to select for improved feed efficiency; to advance the understanding of common factors attributable to high and low feed efficiency potential
Among the main achievements (genetics/breeding):

- Novel markers and positional functional candidate genes for chicken. Development of an industry directed panel of SNP markers for genomic selection in pigs
- Modelling confirmed that improved feed efficiency was better achieved through selection for Residual Feed Intake (RFI) than for Feed Conversion Ration (FCR). The

Genomic Feature Model (GFBLUP), using information from the project, resulted in it having a better predictive ability than the standard GBLUP

Some key areas of future potential:

- Genotype x Environment interactions
- Integrating Multi layers of 'Omics' into breeding models

Prolific (Pluridisciplinary study for a robust and sustainable improvement of fertility in cows) - www.euprolific.eu – presented by Joelle Dupont

The objectives were:

- To develop models to support on farm decision at different levels: animal fertility, herd management, and socio-economic impact for the farm and the farmer.
- To identify genes and pathways involved in the adaptation of the reproductive function to different environmental conditions, especially low input feeding systems
- To identify the functional quantitative trait nucleotides for days till first luteal activity (based on progesterone measures) and estimate genomic breeding values using whole sequence information on individuals
- To study the adaptative response of animals to different feeding systems and management strategies
- To develop decision support tools to optimise the timing of reproductive management decisions, improve the rate of successful inseminations, and provide reproductive performance benchmarking

Main achievements:

- List of genes and network of genes differentially Expressed/deregulated under metabolic imbalance or exposure to pathogens, as functional information to orientate future genetic studies (as well as information to design new diagnosis and prognosis tools, basic information for future research on health and resilience
- Identification of a new trait: commencement of luteal activity (CLA) (higher heritability and repeatability than traditional fertility traits). CLA could be a good trait to select on without impairing milk production
- Identification of genetic variants associated with endocrine fertility traits through GWAS. Detected variants can be used in breeding programmes to enhance genomic selection

Main lessons learnt and challenges:

- To use the models to test research hypotheses is hardly affordable from a systemic point of view by the experimental method (for ex. to quantify the combined effects of milk production and nutritional status on all components of the reproductive responses and to estimate the effects on variability intra herd)
- To design new diagnosis and prognosis tools form omics data

ROBUSTMILK (Innovative and practical breeding tools for improved dairy products from more robust dairy cattle) - http://www.robustmilk.eu/ - presented by Roel Verkamp

The main objective of the project was to develop new practical technologies to allow breeders to re-focus their selection to include milk quality and dairy cow robustness and to evaluate the consequences of selection for these traits taking cognisance of various milk production systems. Specific objectives were: to develop a common database for research

herds; phenotypic tools (MIR); statistical tools dealing with Genotype-by-Environment (GxE);
genomic tools (statistics and Quantitative Trait Loci -QTL); to integrate and disseminate knowledge on the consequences of selection practices on robustness and milk quality.

Achievements:

- A common database: database and protocols, but also an acceptable “way of working”: follow up through gDMI initiative and ICAR working group Feed&Gas.
- Phenotypic measurements tools: An agreement for fatty acids between countries, feed intake measured using MIR
- Statistical tools on GxE interaction delivered, but relatively low impact in cattle (yet)
- Genomic tools: Statistics multitrait genomic prediction, GxE; Research herds got genotypes -> stimulation

As part of the dissemination, 6 papers made up a special issue of Advances in Animal Biosciences Journal (Juy 2013)
Lessons learnt and challenges:

- There were no direct industry partners, but very close links were present and application orientation was present. RobustMilk served as a breeding ground for innovation;
- The project created a European momentum in cattle breeding, and collaboration is continuing today in other initiatives; MIR, feed intake, GHG. Rest of the world was following on this subject. It is important to keep lead and momentum at EU level.

Gene2Farm (Next-generation European system for cattle improvement and Management) - http://www.gene2farm.com/ - presented by Filippo Biscarini
The main objectives of the project were to develop methods and tools to apply genomic selection in (relatively) smaller cattle breeds (mainly dairy cattle); to generate whole-genome sequence data and High-Density SNP data.
The achievements were:

- Models and tools for within-, across- and multiple-breed genomic selection.
- Around 200 whole genome sequences, over 6000 genotypes from 14 dairy, beef and dual-purpose European cattle populations.
- Among the main tools developed: SNPchiMp for SNP data conversion; "Zanardi": a suite for imputation, data format conversions, quality control, inbreeding...

Lessons learnt:

- There is potential for genomic selection in smaller European cattle breeds: across-breed / multiple-breed genomic selection
- Genomics is much more than just genomic selection:
    - Estimate inbreeding
    - Mate planning (e.g. for conservation)
    - Traceability
    - Identifying carriers of mutation
    - Tracking genes
- SME financing scheme: some specific issues can be hard to manage (SME financing RTD, VAT issue with within-country operations)

FECUND (Optimisation of early reproductive success in dairy cattle through the definition of new traits and improved reproductive biotechnology) - http://www.fecund-project.eu/ - presented by Filippo Biscarini
The objectives were to gain knowledge on fundamental aspects of cattle reproduction; identify genetic loci, proteins, miRNAs, biomarkers, novel phenotypes associated with cattle fertility; evaluate potential applications to cattle breeding and farming
The achievements were:

- Two large experiments: energy balance (low vs high), genetic (low vs high
- fertility heifers)
- Lots of data (Next Generation Sequencing) generated on fundamental aspects of reproduction (SNP data, phenotypes, RNA-seq, miRNA, proteomics, biomarkers, metabolites …) from oocytes, embryos, oviducts etc ...
- Tools developed: e.g. bioinformatics pipeline for automated gene functional and pathway analysis (at PTP Science Park)

Due to its technical difficulty and a too low number of cell available, the experiment on differential methylation (epigenetics) from oocytes was not performed so far and is being replaced by a quantification of DNA methylase (e.g. by PCR)
Lessons learnt (on project management): it is very complex to carry out two large experiments simultaneously; it is better to split in order to be able to deliver, and then compare the results

EADGene ("European Animal Diseases Genomics Network of Excellence") and **EADGene_S** (strengthening the implementation of durable integration of EADGENE) - http://www.eadgene.info – presented by Marie-Hélène Pinard-van der Laan
The main goals were to share, produce, update data and genomic and bioinformatics tools and methodologies for joint collaborative research
Achievements: A virtual laboratory, common genomic tools, joint research, technology transfer:

- Provision + annotation of chicken, pig, bovine microarrays
- Provision of SNP panels for large scale genotyping
- Development and distribution of bioinformatics tools
- Promotion and organisation of workshops, courses and conferences
- Exchange of scientists
- Collaborative research on many major pathogens
- Comparative transcriptomics & meta-analyses in hosts and pathogens
- Sequencing and annotation of animal and microbial genomes
- Mathematical models of disease
- Club of Interest: over 70 industry members
- Collaborations between research and industry
- Guidelines for addressing Intellectual Property issues
- Ethical Guidelines

Challenges:

- Integrating interactions with environment, e.g. microbiome
- Improving impact of advanced genomic tools on higher scales (gene -> population -> territories), developing e.g. epidemiogenetics
- Revisiting genetic architecture of health traits, exploring e.g. Epigenetics and immune traits, improved predictive understanding of phenotypic expression

- Renewing breeding strategies with true “Integrated management of health” approach
- Better balance (in projects) between “blue sky” FAANG type research vs “applied research”
- Uptake of genomic tools/methodologies by industry through better knowledge exchange of science & constraints (e.g. trade –offs) towards innovation

Lessons learnt:

- To build a critical mass and ensure flexibility are essential elements. “Genomics” was a learning process
- Genomics is part of multi(inter)disciplinary research
- Sharing knowledge & data in a structured way; for this common data/interest and elements of trust are required
- Researchers have as much more to learn from Industry and application than the reverse. The best model of partnership still needs to be designed

During the discussions around the EU funded projects' outputs and the breakout out sessions, the following aspects were discussed:

- EU projects enabled significant progress in knowledge but also concrete exploitation of results, not least on genomic selection, but also breeding for feed efficiency or mid-infra-red measurement of milk fatty acids in cattle.
- As regards feed efficiency, Residual Feed Intake (RFI) seems to be a better trait than Feed Conversion Ratio (ECO-FCE project) in monogastrics. Our understanding of feed efficiency in cattle is limited. A better understanding is crucial to reduce the ecological impact of ruminants.
- With the fast advance in high throughput ‘omics technologies, many projects underestimated the size of data generated and in a number of cases, interpretation was not completed by the end of the projects. The pooling of data and the methodology to exploit them are underestimated challenges
- While breeding (i.e. selection of animals with useful traits) is close to implementation, it exploits the outputs on upstream research. Therefore, fundamental research should also be performed to allow later exploitation. For instance, FAANG is seen as crucial, although it cannot be directly linked to a societal issue. It is difficult to address the whole path from basic research to breeding application in a single project lasting a few years. It is important to invest sufficiently in fundamental research which has a potential for application.
- Encouraging the sharing of data is essential, but issues of ownership/IPR can be critical. Data sharing can be easier to implement in pre-competitive research projects, but the borders are hard to define.
- It might be useful to design a meta-project to facilitate better exploitation of the wealth of data that is collected
- Next to large projects, smaller focused projects are also needed

Conclusions of the session on animal breeding by Theo Meuwissen

There should be a clear difference in research projects:

- Applied research: it results directly in exploitation and thus should ideally involve multi-stakeholders.

- Fundamental research: it results indirectly in exploitation via later projects

This could be achieved by selecting several projects from a call, with some more fundamental, some more societal directed. An infra-structure could be set up for them to interact.

In order to favour exploitation of results, there should be back and forth communication between industry needs and technology development. Further, it is important to acknowledge differences between species due to not only biological differences between species but also the organisation of the production pyramide.

The priorities for research are:

- Phenomics (high resolution/large scale phenotyping), including deep phenotyping, molecular markers, using precision technologies
- Across breed genomic selection in monogastrics and ruminants:
    - High resolution genotyping
    - Increased biological knowledge (such as FAANG initiative)
    - Non-linear models for genomic selection exploiting biological knowledge
    - design of reference populations
- Reproduction technology
    - Supplying many embryos in ruminants (for intense genomic selection-step)
    - Young parents (for short generation interval)

Although epigenetics is an important component, it is perceived that the impact on genetic improvement at this point of time is limited. Regarding epigenetics it is important to better understand transgenerational epigenetic inheritance: transmittance of information from one generation to the next that affects performance of offspring.

3. Genetic resources

While focussing mainly on ruminants, the introductory speech by Paolo Ajmone Marsan, highlighted that the present livestock diversity is the result of thousands of years of evolution, with domestication (starting around 8000 years ago for cattle), combined with colonisation, migration, trade that led to introgression, influencing natural and anthropogenic selection, in addition to natural mutations and drifts. The concept of breed was born around 1900.

Thousand local breeds are adapted to very different and harsh environments, spanning mountain to desert areas and extreme conditions in terms of humidity, temperature and elevation.

As stressed by FAO, farm animal genetic resources (FAnGR) are important as

- They are an essential part of biological basis for world food security
    - one billion people rely directly on livestock for major proportion of livelihood
- It is a diverse resource base critical to contribute to eradicate world hunger
    - adaptation
    - “raw material” for breeders to make genetic improvement
- They are an international public good
    - Cultural and social value
    - Ecosystem services

The global trends in developing as well as developed economies tend to lead to loss of local germplasm, in particular by importation of exotic germplasm and cross breeding in developing and emerging economies or by the efficiency objectives in livestock production in developed economies. In both contexts, there is an issue of sustainable intensification and its impact on farm animal genetic resources. Issues of environmental impact, food quality for human health, typical products, animal welfare, precision farming, organic farming are increasingly raised.

Although the situation of a number of breeds is not well known in a number of regions, the world's livestock diversity remains at risk, according to the 2015 FAnGR status report of FAO. The genetic erosion is linked to a number of issues, in particular indiscriminate cross-breeding, introduction/increased use of exotic breeds, weak policies or institutions, lack of profitability and competitiveness, production system intensification, diseases/disease management, loss of pastures or production environment, poor control of inbreeding…

There are layers of diversity from epigenome, to genome, transcriptome, proteome, metabolome, microbiome, which together with environmental factors, determine different phenotypes. There is structural diversity, such as Copy Number Variations (CNVs), variation in functional sequences comprising both coding and regulatory sequences. The selective pressure has an impact not only on targeted genes/SNPs but also neutral variation. It is still difficult to identify causative Quantitative Trait Genes and Quantitative Trait Nucleotids and genomic selection is difficult to apply.

Ambitious projects (the 1000 genomes project in humans, the 10,000 genomes project in vertebrates and the 1000 bull genomes project) aim to sequence (or re-sequence) a number of (key) individuals. FAANG aims to address notably the regulatory sites.

This increased capacity to decipher the diversity and underlying biological mechanisms will allow drawing a much more accurate picture and more precise/accurate breeding and management of FAnGR.

The introductory speech put forward a number of questions about FAnGR:

- Characterization
  - We refer to genomic and phenotypic characterisation: what about intermediate phenotypes and other –omics? Interdisciplinarity? Go to the biology? Environment? Trait ontology? Data conservation and sharing? Training?
- Utilization
  - Genomic selection? Crossbreeding? In research? In other use? What are the effects of the Nagoya protocol?
- Conservation
  - Which populations? How about wild relatives? How to prioritize them? Role of cryobanks?
- Communication
  - has communication from science to decision makers been effective? New modes?

Can a real synergy between local and cosmopolitan germplasm be achieved? A suggestion was that by introducing production alleles from "industrial breeds" into "local germplasm", and adaptative alleles from local germplasm (e.g. slick hair locus for thermotolerance) into industrial breeds, it could respectively make production more sustainable and FanGR conservation more sustainable. The question of the possible role of new biotechnologies (CRISPR genome editing) in this was also raised.

Below are the main achievements/challenges identified by the EU projects in that session, in relation to genetics/breeding/phenomics :

LowInputBreeds (development of integrated livestock breeding and management strategies to improve animal health, product quality and performance in European organic and *low input* milk, meat and egg production) - http://www.lowinputbreeds.org/home.html - presented by Gillian Butler
The rational for this project was that animal breeding focuses on intensive production systems and is dominated by big businesses; selection is essentially on performance traits and functional traits have low priority; modern genotypes are successful, if supported by high inputs; Organic and low-input systems need robust animals; they represent a diverse and relatively small market
Key findings were presented in dairy cattle:

- For many of low input traits, genomic prediction is promising & provides more accurate breeding values
- No major differences in *'gene relationships*' between Swiss organic, high input Brown Swiss and German Holstein/Friesian populations
- Concentrations of the nutritionally desirable Fatty Acids are inversely related to highly selected US BS genetics
- Confirming crossbreeding offers potential in low-input & organic systems with considerable variation between individual cows

Work was also performed in sheep, pigs and poultry; more info can be found on www.lowinputbreeds.org and http://organiceprints.org/

Selsweep (Molecular characterization of genetic factors in the pig under selection during speciation, domestication and breeding) - presented by Martien Groenen
The objectives were to give insight into genes and regions contributing to speciation of Suids; to study selective sweeps and selective pressure in pigs (at three levels); to give insight into the contribution of the different kinds of variation (SNP, CNV) during selection; to design a 500K SNP chip for pigs.
Achievements:

- The project generated WGS data for >300 sequenced individual pig genomes
- As regards suid evolution, it provided high confidence phylogeny of Suids; evidenced extensive admixture after speciation; showed a 1 Million year divergence between European and Asian wild boar; identified regions under selection
- As regards domestication and selection, there was independent domestication of European and Asian pigs; gene flow from wild to domestics after domestication; extensive gene flow from Asian into EU domestic pigs (200 years ago); the project identified multiple genes under selection in EU and Asian pigs and wild boar and showed selection of Asian variants after introgression
- As regards resources, the project notably identified over 40 million variants in pigs; it designed a 650K High Density SNP chip (Affymetrix) ; it identified potential functional variants for relevant traits; it provided a catalogue of potential deleterious alleles
- Among the limitations, the biological insight in selective sweeps was mostly limited to protein coding genes (and non-synonymous variants). In connection with FAANG, efforts to improve the annotation of the pig genome were initiated after the end of Selsweep, with preliminary results available for miRNAs and lnRNAs.

Treasure (Diversity of local pig breeds and production systems for high quality traditional products and sustainable pork chains) – Horizon 2020 project recently started - http://treasure.kis.si/ - presented by Luca Fontanesi
The objectives of the project:
Description and evaluation of local pig breeds, with an emphasis on untapped ones will be performed using novel genomic tools (Work Package 1). Performance of local pig breeds will be evaluated in contrasted agro-geo-climatic conditions and production systems (indoor, outdoor, organic). Focus will be on pig feeding and management strategies and on the use of locally available feeding resources. Intrinsic quality of traditional and new regional high quality pork products and attitudes of consumers from various market areas will be assessed; in particular the motives for the choice and willingness to pay such products. Marketing strategies will be addressed in particular short chain distribution channels.
Work-package 1 (phenotypic and genetic characterization of regional autochthonous pig populations in Europe) is addressing 20 breeds with the aim to describe the singularity of local pig breeds at phenotypic, genomic and functional level; to evaluate the adaptation of local pig populations to agro-climatic conditions; to develop new methodologies for the management of the local pig populations; and to develop DNA based tools for animal and meat traceability and authentication.
The achievements so far were : identification of genes affecting the presence/absence of hairs in Casertana breed; identification of genes affecting the shape of the tail in pigs; identification of a gene marker for product authentication of Cinta Senese pigs; genomic inbreeding coefficients in Italian local pig breeds; a DNA repository of local pig breeds; a genotyping database of local and cosmopolitan pig breeds.

During the discussions around the EU funded projects' outputs and the breakout out sessions, the following aspects were discussed:
The industry is interested in conservation of genetic diversity from the angle of inbreeding. One group stressed that there is no scientific evidence that intensive selection has reduced robustness. It was also recognised that genetic diversity may be more important for fitness and health traits. An important issue is the interpretation of genetic differences to identify what is most worth preserving.
Genetic resources can be used for mainstream animal breeding and to improve local breeds.

- For breeding: combining genomic selection and introgression. Genetic information (rare phenotypes, related genes) from FAnGR could be obtained and applied in commercial breeds, possibly by gene editing in commercial breeds. The issue of acceptability of new technologies (i.e. New Breeding Techniques) was raised. Phenotyping of local breeds and looking at environmental metadata would enable disentangling G x E aspects, case-by-case.
- For local breeds, a special genomic toolbox is needed. On case-by-case, breeding tools should be targeted and adapted to the different local situations. The biological mechanisms affecting local breeds should be studied. In one group, the question was raised about how genomics can deliver as well for small breeds.

There is a need for relatively basic research to study introgression between breeds and connections between variation and function. It is certainly difficult to connect this knowledge to phenotypes. New methodology for introgression should be developed
Poultry and pigs may be more suited and industry interested in trying new lines by introgression and in validating genotypes that carry advantageous traits (e.g. disease resistance)

A concern was raised about the possible gaps in between identifying a genetic marker and how far this can be used by breeders.

Genomic and phenotypic characterization of FAnGR is essential, but intermediate phenotypes and other –omics may be equally important, in particular for understanding the biological mechanisms and the environmental influence. Adaptative trait ontology (role of ICAR, FAO) is necessary and interdisciplinary research will be required.

The issue of how to correct under-utilisation of local breeds was raised, as well as more cross-cutting issues, like data conservation and sharing, training.

The purpose of the workshop was not to discuss specially NBTs, but a remark was made that NBTs may not be a game changer, as traits are already present. Use of a pool of FAnGR might better to be used.

Conclusions of the session on genetic resources by Paolo Ajmone Marsan

- A key word is adaption (e.g. to environment, diseases, stress, feed), whether for commercial or for local breeds.
- A substantial amount of data are already available and they should be further exploited. However, "we are only scratching the surface of these adaptive traits"
- Research should be focussed on further genomic characterisation (e.g. sequencing and selection signatures; species genomics) should be performed and biological understanding should be pursued (molecular phenotypes).
- Useful genes may be used in commercial breeds by introgression or gene editing technology.
- Targeted molecular breeding tools should be developed/improved in FAnGR having different sizes and in different species, for the purposes of diversity conservation and/or genetic improvement.
- As regards conservation of local breeds, there may be a need to discuss/agree on priority methods or gold standards. The under-utilisation of locals breeds needs to be addressed, including policy and socio-economic aspects (e.g. promoting conservation, brands, organic market)
- A comprehensive view is needed when looking at industrial versus local breeds, or genetic resources versus gene editing. A comment was made that genes could be introgressed/edited in e.g. Holstein but Holstein may not have all the best traits. There are certainly many opportunities for FAnGR.

4. From Genotype to Phenotype

The introductory speech by Alan L. Archibald "From phenotype to genotype and back again – animal genomics enabling prediction" illustrated not only the relation between genotype and phenotype but the iterative process required for related research.

The study of the Genotype – Phenotype relation aims to predict outcomes (e.g. susceptibility to cancer; performance of daughters of elite dairy bull; susceptibility to disease). It is a necessity for animal breeding.

Progress in genomics was very fast after DNA was sequenced in 1977, enabled by advances in quantitative analysis. While the human genome project was launched in 1990, The PiGMaP project started in 1991. In the nineties, emerged the quantitative trait locus (QTL) mapping, and marker assisted animal selection. In 2001, while the draft human genome sequence was released (human genome sequence was "finished" in 2003), genomic animal selection was proposed (and it was implemented soon after in dairy cattle).

Thanks to breeding, there were tremendous increases in productivity with intensification of (animal) agriculture.

In the livestock domain, successes were made in identifying genes associated to certain traits (e.g. DGAT1 for milk yield; MSTN for sheep muscling) enabling more sophisticated selective breeding. But working on one gene at a time was slow and relatively inefficient.

Since 2001, draft reference genome sequences have been established for a number of economically important animal species (notably Chicken in 2004, cattle and horse 2009, pig 2012, goat and duck 2013, sheep 2014). Substantial investments have been made in order to more fully characterise the human genome and the genome of mice and other model organisms through the ENCODE project (ENCODE 1% (2003) followed by ENCODE genome wide (2007)). In parallel with the sequencing of farmed animal genomes the genomic tools that are essential for the application of genomic selection, e.g. a variety of SNP-chips (SNP = Single Nucleotide Polymorphism), notably HD SNP chips (notably 750K in bovine 2010, 600K in chicken 2012, 650K in pigs 2015).

Genomic selection underpins faster and more accurate genetic improvement and is being implemented in all major farmed animal sectors in the developed world.

For instance, in the USDA dairy cattle genomic evaluation, the accuracy of breeding for milk yield reached 0.86 with genomic selection, compared to 0.51 for breeding based on pedigree (source George Wiggans, USDA-Beltsville).

As currently implemented, genomic selection relies upon linkage relationships rather than linkage disequilibrium. Thus, the predictive value of the marker alleles decays rapidly in subsequent generations as the linkage phase between the marker alleles and the trait alleles is broken by recombination events. As a result predictions are not robust in unrelated individuals, for example, between breeds. In future genomic selection will evolve to operate as originally envisaged by exploiting population-wide linkage disequilibrium so that the relationships between marker and trait alleles are more robust over time and across populations.

It is foreseen that in the next five years, sequence data will likely supplant SNP genotypes. 2 approaches are considered: sequencing individuals, as in the 1000 bull genomes project (but expensive even at 1000$ per genome). There are already multiple (aligned) livestock genomes, some of which resulting from/contributed by EU funded projects. But sequencing of a few individuals may not be that useful. The other option is sequencing whole populations with the aim to get sequence data for 100,000 to 1 million individuals at low coverage (if cost would be around 10$ per genome). This approach would exploit pedigree structures in managed population, with imputation from low sequence coverage.

To understand ‘from sequence to consequence’ (i.e. from genome to phenome), not only the genome (e.g. SNP, CNV-copy number variants, genomic rearrangements), but also epigenome (e.g. DNA methylation, histone modification, miRNA), transcriptome (e.g. gene expression, alternative splicing, long non-coding RNA), proteome (e.g. protein expression, post-translational modification) and metabolome (metabolite profiling) are important.

Reference genomes need to be (and are being) improved, and annotation progresses, but farm animal genomes are far less well annotated than human and models organisms.

Visualization and alignment tools are under development.

Discovering functional sequences and their biological role requires expensive tools/schemes such as ENCODE, iHEC in the medical domain, or FAANG in livestock. Work in ENCODE

revealed that 80% of DNA participates in at least one biochemical RNA-and/or chromatin-associated event in at least one cell type and that promoter functionality can explain most of the variation in RNA expression. ENCODE cost was over 250million$, but prices went down considerably since then and the same activity could be performed nowadays for a fraction of that amount.
Richly annotated reference genomes, through infrastructures like ELIXIR, and tools like Ensembl, are a key shared (open access) resource for biological research, a source for effective exploitation, such as genomic enabled prediction (e.g. selective animal breeding). Because these resources are expensive to develop, international collaboration is necessary in order to share costs and minimise wasteful duplication of effort.

Ruminomics ("Connecting the animal genome, the intestinal microbiome and nutrition to enhance the efficiency of ruminant digestion and to mitigate the environmental impacts of ruminant livestock production") - www.ruminomics.eu – presented by Johanna Wilkki
Among the main achievements:

- Host effect on rumen microbiome are not predictable in digesta exchange studies, however some microbes show interesting pattern of host genetic control (e.g. Firmicutes)
- Results of Genome Wide Association Studies (GWAS) on cow genome: regions associated with CH4 emissions, archaea/bacteria proportions, individual phyla (+ milk FAs) could be used for GEBV.
- The conclusions from the 1000 cows study are that the relationships between emissions, digestion and efficiency are complex. While generally more efficient cows produce more methane per unit of feed intake, opportunities to select cows with both lower methane emissions and higher feed efficiency exist.

The future challenges as seen by Ruminomics:

- Feed efficiency and feed intake traits are still challenging to improve, because of the difficulties to collect accurate phenotype measurements from large numbers of individuals
- Understanding the role of the microbiome
- Including biological background information and causal variants to genomic evaluation would improve predictions and reduce dependence on Linkage Desequilibrium (-> across-breed)
- Difficulty to pinpoint causative variants: most SNPs in GWAS studies map outside protein-coding regions
- Cattle functional regulatory elements have not been well annotated (the FAANG initiative would help)
- Epigenetic effects (here as well FAANG would help)

GplusE ("Genotype and Environment contributing to the sustainability of dairy cow production systems through the optimal integration of genomic selection and novel management protocols based on the development") –www.GplusE.eu – presented by Miel Hostens
The aim of the project was mainly to identify the genotypes controlling biological variation in the important phenotypes of dairy cows, to appreciate how these are influenced by environmental and management factors and thus allow more informed and accurate use of genomic selection.

The objectives are:

- to develop and validate new, easily-measured phenotypes that can supplement or correlate with and replace traditional phenotypes (milk yield and composition, fertility, product quality, traditional disease and fertility records);
- to identify novel genomic markers, including causative mutations, for the key phenotypic traits (productivity, efficiency, environmental footprint, health, fertility and animal welfare) that can be used for GS;
- to generate ENCODE-like (i.e. FAANG) data for the bovine as a public resource that can be used for both GS in dairy cattle and to increase understanding of genotype/phenotype associations across all species;
- to define the key interacting problems that contribute to management of dairy cows under different systems that will allow the building and validation of an appropriate management blueprint based on Hazard Analysis & Critical Control Point (HACCP) and Evolutionary Operations principles;
- to develop new breeding and management strategies that incorporates the new genomic information attaching to the common currency phenotypic data established through the project;
- to disseminate the information widely to producers, consumers and the scientific community and to use the project to train a new generation of young scientists in the specialism of bio-informatics.

The project started in 2014 and a number of work packages are expected to produce their outputs after the 3rd year, therefore achievements are not there yet.

One main challenge for this data intensive project was the variety of people and interests, of data and their interpretation. Involvement of industry in such project is a 'must'.

Services provided by European Molecular Biology Laboratory's-European Bioinformatics Institute - presented by Ian Streeter -www.ebi.ac.uk

The mission of EMBL-EBI is to provide freely available data and bioinformatics services to all facets of the scientific community in ways that promote scientific progress.

As an ELIXIR node, EMBL-EBI supports the coordination of biological data provision throughout Europe.

There is huge demand for big data. To illustrate this: in 2015, scientists at over 5 million unique sites use EMBL-EBI websites and around 18.5 million data requests are made to EMBL-EBI websites every day. The storage capacity at EMBL-EBI is over 100 petabytes.

Among the data archives are: the European Genome-Phenome Archive; the European Variation Archive; the European Nucleotide Archive; The Protein Data Bank in Europe; Array Express...

A number of tools for integrating data are available, such as Ensembl, BioStudies; UniProt; EBI Metagenomics...

The following principles apply EMBL-EBI service provision:

Open - data and tools are freely available, without restriction. The only exception is potentially identifiable human genetic information, for which access depends on research consent agreements.

Compatible - EMBL-EBI is a world leader in the development of global bioinformatics standards, which are key to data sharing.

Comprehensive - Thanks to many data-sharing agreements, EMBL-EBI resources are comprehensive and up to date. EMBL-EBI works with publishers to ensure that biological data must be placed in a public repository and cross-referenced in the relevant publication.

Portable - All data and many of EMBL-EBI software systems can be downloaded and installed locally.
High quality - Databases are enhanced through annotation: highly qualified biologists add value to databases by incorporating features of genes or proteins from other sources, and automated annotation is subjected to rigorous quality control.
Ensembl is one of the worlds most highly used genomics resources, with strong visualization function. Several farm animals are among Ensembl's top ten most popular species.
Biostudies, which provides for description of biological studies, offers a repository for data (including metadata) which doesn't fit in other structured archives and links with other EBI resources.
EBI provides also ontology services for well structured, consistent, specific data, which enables aggregation, integration, tracking.
About metadata, EBI works with communities to agree good standards and provides tools to help the community meet those standards. In that framework, EMBL-EBI works with FAANG initiative (see below).

FAANG (Functional Annotation of ANimal Genomes) - http://www.faang.org/ - presented by Elisabetta Giuffra
FAANG is a coordinated international action to accelerate understanding of the genome to phenome link. This is important for improving fundamental understanding of biology and for Genomic Selection. For this to happen, there is a need for high quality reference genome sequences; comprehensive annotations of functional elements and variants, using comprehensive and standardised methods (so that integration of information from multiple sources is easier/possible); common infrastructures (biological, bioinformatics and database resources). The main challenges are in the limited resources available to address different species and groups (therefore the need for international cooperation) and in the variable quality of existing reference genomes and annotation (therefore the need to focus on good quality reference genomes).
This initiative is building on the successful ENCODE projects which addressed annotation of the human, mouse and some model organism genomes. Lessons from ENCODE were that there is a need for well established consortium rules and policies, technologies and experimental/data standards. They also demonstrated the added value of freely available reference datasets, and of integrating biochemical data alongside other evidence.
FAANG aims to:

- Standardize core assays and experimental protocols
- Coordinate and facilitate data sharing
- Establish an infrastructure for the analysis of these data
- Provide high quality functional annotation of animal genomes

Contributors to FAANG have to adhere to a "Data Sharing Statement".
The initial phase of FAANG aims to obtain a biological reference for main farm species. The following phase is to expand the research to different developmental stages, physiological conditions, and additional species and to work on genotype-to-phenotype interaction (looking at new breeds/phenotyped populations), with involvement of the industry.
The FAANG Data Coordination Centre is at EMBL-EBI (see above) and the Networking activities are coordinated by Roslin-Edinburgh University in UK and Delaware University on the US side. On the EU side, the network operates through a COST action: FAANG Europe (www.faang-europe.org). On US side, the network is part of an NSF funded initiative: the

Animal Genome to Phenome Research Coordination Network (www.ag2p.org). The FAANG Data Coordination Centre is funded from a grant by UK BBSRC.
Since 2015, some FAANG-pilot projects started, in connection with on-going research (few animals, several tissues/assays), while FAANG-related activities are increasingly becoming part of on-going projects (several animals, usually fewer assays). Both categories mostly include national projects and target cattle, pig, chicken and to a limited extent goat, sheep, horse and even exotic ruminants. These projects take place in EU member states (France, Germany, The Netherlands, Finland, UK), USA, Australia and Canada. For full deployment of its first phase, FAANG considers that an international and coordinated project to build the FAANG reference(s) would act as a boost for 'any' research topic and avoid redundancies/fragmentations, like notably demonstrated by ENCODE. Additionally, reinforcing the means for global coordination is presented as a key need for interconnected data analysis centres and for expansion to additional groups. Thereafter, several biology-led projects could be launched on the basis of the knowledge acquired from the first phase.
A comparable initiative has been launched for salmonids: FAASG (Functional Annotation of All Salmonid Genomes). It is coordinating the international salmonid community to acquire, standardise and share data for comprehensive mapping and characterization of the functional elements of salmonid genomes.

During the discussions around the EU funded projects' outputs, other presentations and he breakout out sessions, the following aspects were discussed:

- Reference genomes:

Richly annotated reference genomes are required.
While sequence information is used for imputation, with High Density SNP-Chips, one identified priority is to move to causative variation.

- Epigenetics (genome modification without modification of DNA sequence): separating genetic from epigenetic influences on a phenotype
- Phenotypes:

Phenotypes are multidimensional. Better defined phenotypes are required.
Trait ontologies, from simple to complex or intermediate phenotypes (e.g. chromatin status) are important for G x E study, as well as epigenetics.
In a number of cases technological developments may be required to enable e.g. collecting large quantity of on-farm data for breeding purposes.
How different phenotypes interact, or how they change with time should be taken into account.
Linking to industry and their resource population (e.g. for annotation experiments), as well as using phenotypes already collected, should be pursued, using certain cases to show the impact of increasing genomic prediction accuracy.

- Data management:

Projects generate a lot of data, whose management is an increasing challenge, for a project and after the end of a project.
Data Services are important, but commercial data is often too much protected.
Data may not be enough without related information: importance of metadata
FAANG already has a significant dataset, but one issue is the harmonisation of specimens, protocols and standards used.
A sound basis for annotation is needed, as well as FAANG type of data.

- Cost:

Cost of annotation is hard to estimate (depending on number of specimens, assays etc) but 1-1,5 million €/species on a defined type of tissues (for instance, immune cell types) would likely provide appropriate info for better understanding and increasing value of breeding.

- For non-main breeds, biological understanding is needed, also for cross-breeding

Questions were raised around the need or not for full characterisation of the transcriptome, and taking advantage of the existing human annotation.

Conclusions of the genotype to phenotype session by Alan L. Archibald

- Genomic selection (GS) has already delivered socio-economic impact; it was implemented within 7-8 years of theory development and within 1-2 years of availability of genome information and molecular tools
- Improved genomic information will allow rapid refinement and improvement of GS
- FAANG is an important initiative:
    - It intends to identify the main functional features of the genomes of domesticated animals, in terms of i) open chromatin, ii) promoters, iii) enhancers
    - Three / four species should be targeted as a priority: cattle, chicken, pigs, possibly sheep (ruminants, monogastric mammal, monogastric avian)
    - Three biological systems: muscle, immune system and gastro-intestinal tract for investigating development (e.g. time course experiments) and environmental perturbations (e.g. biotic and abiotic stresses, diet influence).
    - A rough estimation of cost is around €1.5 mio per species per biological system
- There is potential for rapid exploitation of knowledge acquired, depending on the development of new prediction equations that incorporate coding / regulatory / intergenic classification filters, as well as on user engagement, including addressing the issue of data sharing of potentially commercially sensitive data.
- Improvement in genomics-enabled animal breeding can address societal changes, such as meeting demand for animal protein in the most efficient and sustainable manner.
- Genomics-enabled selective breeding is agile and accurate, enabling early selection based on molecular genotype or changes in selection indices, increasingly informed by the understanding of the genotype-phenotype (sequence to consequence) relationship.
- Genomics-enabled breeding can help effective management of genetic resources.

5. General conclusion

Omics is a fast advancing science, thanks to continuous progress in technologies and data management.

EU funded projects on animal genomics, genetics and breeding have a substantial impact in progressing knowledge and its exploitation for animal production. Genomic selection is now implemented in breeding of the major farmed animal species (including dairy cattle, pigs, poultry and salmon). However, typically, its accuracy is dependent on the size of reference population and its use in small breeds is limited.

There are still large gaps in understanding the biology connected to the DNA sequences and their variation, i.e. the relationship between genome and phenome (or "from sequence to consequence") is far from being understood. To achieve this, the workshop highlighted prerequisites:

- Further work on annotation of genome and investigation of other omics (transcriptome, proteome...). International research cooperation is important to realise this.
- Better characterisation of phenotypes/traits ("in the era of genomics, phenotyping is king"). To a certain extent, "we are only scratching the surface of these adaptive traits". The industry could certainly be helpful in this domain.

Basic research projects would underpin subsequent applied projects that focus on using the newly generated knowledge.

During the workshop, it was stressed that biological understanding would facilitate genome-enabled selection (faster and more targeted breeding) and could be used for managing genetic resources, for the purpose of diversity conservation and/or genetic improvement (e.g. by introgression). However, for small populations (small breeds), specific tools would need to be developed, in particular non-linear models for genomic selection using biological data.

Additionally, reproduction technology could speed up the process through the supply of many embryos (for intense genomic selection-step) or using young parents (for reduction of generation interval).

Industry involvement is an important component to ensure that research meets the needs of the industry. Industry involvement is important not only in the formulation of the research objectives but also for the collection of data. Large data sets are available in the industry which could be made available to the research community when agreement on data sharing are in place. FAANG and structures like EMBL-EBI have an open access policy, which is welcome and important for pre-competitive research. However, the border with competitive data is not always clear and supply of / access to commercially sensitive data is an issue that needs to be addressed.

Annex I: Programme

10 October 2016

Lunch and Registration

Welcome

Animal breeding

Opening presentation: setting the scene – precision animal breeding - *Theo Meuwissen*

- project presentation **ECO-FCE** (Connecting the animal genome, gastrointestinal microbiomes and nutrition to improve digestion efficiency and the environmental impacts) – *Klaus Wimmers*
- project presentation **Prolific** (Pluridisciplinary study for a robust and sustainable improvement of fertility in cows) – *Joelle Dupont*
- project presentation **ROBUSTMILK** (Innovative and practical breeding tools for improved dairy products from more robust dairy cattle) – *Roel Verkamp*
- project presentation **Gene2Farm** (SME funding scheme - Next generation European system for cattle improvement and management)/**FECUND** - *Filippo Biscarini*
- project presentation **EADGene** (genomics, bioinformatics, animal health and animal models) – *Marie-Hélène Pinard-van der Laan*

Roundtable on animal breeding

Refreshment break

Genetic resources

Opening presentation: setting the scene - sustainable use of animal genetic resources in the "-omic" era – *Paolo Ajmone Marsan*

- project presentation **LowInputBreeds** (Development of integrated livestock breeding and management strategies to improve animal health, product quality and performance in European organic and 'low input' milk, meat and egg production) – *Gillian Butler*
- project presentation **Selsweep** (Molecular characterization of genetic factors in the pig under selection during speciation, domestication and breeding) – *Martin Groenen*
- project presentation **TREASURE** (Diversity of local pig breeds and production systems for high quality traditional products and sustainable pork chains) – *Luca Fontanesi*

17.00 – 17.45 Roundtable on genetic resources

11 October 2016

Genomics: State of play – genotype to phenotype

Opening presentation: setting the scene - from phenotype to genotype and back again - animal genomics enabling prediction – *Alan Archibald*

- project presentation **Ruminomics** (Connecting the animal genome, gastrointestinal microbiomes and nutrition to improve digestion efficiency and the environmental impacts of ruminant livestock production) – *Johanna Vilkki*
- project presentation **GPlusE** (Sustainable dairy production and identification of appropriate phenotypes: challenges and opportunities) – *Miel Hostens*
- perspective of **EMBL-EBI** services to support animal genomics – *Ian Streeter*
- perspective of **FAANG** (*Functional Annotation* of Animal Genomes) – *Elisabetta Giuffra*

Roundtable on genomics

Refreshment break

Roundtable bringing together discussions

Lunch break

Wrapping up & conclusions of the workshop

Annex II: Participants list

10-11 October 2016

1.	*AJMONE MARSAN Paolo*	*Università Cattolica del Sacro Cuore, IT*
2.	*ARCHIBALD Alan*	*The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK*
3.	*BISCARINI Filippo*	*PTP Science Park, Lodi, IT*
4.	*BUTLER Gillian*	*Newcastle University, UK*
5.	*DAVID Xavier*	*ALLICE, FR*
6.	*DUPONT Joëlle*	*INRA, FR*
7.	*FONTANESI Luca*	*University of Bologna, IT*
8.	*GIUFFRA Elisabetta*	*INRA, FR*
9.	*GROENEN Martien*	*Wageningen University & Research, NL*
10.	*HOSTENS Miel*	*University of Ghent, BE*
11.	*MEUWISSEN Theo*	*Norwegian University of Life Sciences, NO*
12.	*PINARD-VAN DER LAAN Marie-Hélène*	*INRA, FR*
13.	*PLASTOW Graham*	*University of Alberta, CA*
14.	*SHIRALI Mahmoud*	*Aarhus University, DK*
15.	*SOELKNER Johann*	*University of Natural Resources and Life Sciences, AT*
16.	*STREETER Ian*	*European Molecular Biology Laboratory - European Bioinformatics Institute, UK*
17.	*TUGGLE Christopher*	*Iowa State University, USA*
18.	*VAN ARENDONK Johan*	*FABRE TP, NL*
19.	*VAN STRATEN Miriam*	*European Forum of Farm Animal Breeders, NL*
20.	*VEERKAMP Roel*	*Wageningen University & Research, NL*
21.	*VENNEMAN Johannes*	*EFFAB/FABRE TP, NL*
22.	*VILKKI Johanna*	*Natural Resources Institute, FI*
23.	*WIMMERS Klaus*	*Leibniz Institute for Farm Animal Biology, DE*