Research Networking Programme
NEWSLETTER No. 3 August 2012

ESF GENOMIC-RESOURCES

Standing Committee for Life, Earth and Environmental Sciences (LESC)

Introduction

Last June, we reached half of the ESF GENOMIC-RESOURCES Research Networking Programme (2010-2014). During the first two years of the action, science meetings (workshops, summer schools) and exchange visits have been organized for young researchers in the domains of the characterization, the economic evaluation, and the use of Farm Animal Genomic Resources. This permitted to bring to the fore innovative ideas and new strategies for the management of FAnGR.

During the second half of the project, the goal of the steering committee is to organize activities dedicated to the promotion of the collaboration to take place between animal scientists and socio-economists. Indeed, the preserving of livestock resources is gradually becoming an institutional or a public service rather than a purely scientific issue. In this context, the role of GENOMIC-RESOURCES will be to concretize interactions between the natural, agricultural and social sciences by means of interdisciplinary events to address the whole system and to reach a comprehensive understanding of the complexity of the reality of farm (animal) management.

But interdisciplinarity between social and natural science is still perceived as a challenge: the scientific practices are different, mutual understanding is time consuming, projects are more risky and in competition with well-defined monodisciplinary approaches. And another challenge is the difficulty to properly define what interdisciplinarity means. Is it the juxtaposition of natural and social science outputs within the same project or does it require specific organization and training to address the same issue with a combination of skills?

Next October in Munich, an extended GENOMIC-RESOURCES steering committee - several sociologists and economists have been invited - will plan the organization of a sandpit meeting to take place in 2013 and whose role will be to propose a combination of activities to favor interactions within and among young and more experienced scientists of different disciplines in an open-minded atmosphere. The sandpit meeting will be organized on three of four days and limited in size (no more than 50 people).

The aim of the Munich preparation meeting will be to get the views of both communities (social and natural sciences) on the interest of this meeting and on the format. The main objectives will be to challenge the rationale of the sandpit meeting, to discuss and complete a first set of activities proposed, to identify target communities and potential speakers. More information in the next GENOMIC-RESOURCES newsletter...

What's new ?

Travel grants

In December 2011, GENOMIC-RESOURCES decided to fund 6 exchange visit grants (representing a total of €24,650) to take place between February and July 2012. In the present issue, we present the final report of L. Czegledi, M. Ferencakovic, A. Frkonja, S. Mucha, Y.

Utsunomiya, and B. Gutierrez-Gil (see from page 3). We received a total 8 applications for a total amount of €30,750, very few compared to the 32 applications (representing a total amount of €143'000) received the previous year.

Science meetings 2012

GENOMIC-RESOURCES will fund two science meeting in 2012.

First, a Summer school entitled **"Livestock Conservation Genomics: Data, Tools and Trends"** will be organized by Professors Ino Curik and Johann Soelkner at Hotel Luna, Pag Island, Croatia, from October 1st to October 7, 2012.

The age of high throughput genotyping and whole genome sequencing provides new views on old problems and allows addressing problems that were considered infeasible so far. The summer school (6 days of intensive scientific program + 1 day excursion) aims to provide researchers with a toolbox (methods and software) to address research questions related to effective population size, population structure, inbreeding and crossbreeding and other population genetic concepts. It will consist of lectures, practical computer exercises and paper discussions.

The summer school is targeted at PhD students and young scientists willing to start working on genomic data analysis in livestock.

Priority will be given to applicants who come from countries that financially support the program (Austria, Belgium, Croatia, Finland, France, Germany, Netherlands, Norway, Sweden, Switzerland and United Kingdom). Applicants are expected to submit a CV, a PhD or recent work summary up to 300 words, and a motivation letter for participation. A total number of 8 grants will be provided by the organizers. Each grant will cover school fee (€ 600) plus travel costs up to € 300. In case travel costs are lower, only the actual costs will be reimbursed. Each application will be evaluated separately. Grants will be given according to the ranking provided by the grant evaluation panel.

All participants are required to bring their own laptop for the practical workshops. Data, software and programs will be supplied by the organizers at the beginning of the school. In addition, participants are required to prepare and submit in advance a presentation based on a scientific publication related to the topics covered by the school for the paper discussions that will take place during the week (a choice of 3 papers will be supplied by the organizers in advance).

Summer school fee: € 600; participants will get access to the program with full accommodation in a double room for 7 days (breakfast, lunch and dinner), course materials, tea/coffee breaks and excursion.

Detailed information available here:
http://www.agr.unizg.hr/files/lcd_2012.pdf

A Workshop entitled **"Genotype-by-Environment Interactions and Adaptation of Farm Animals on Phenotypic and Molecular Level"** will be organized by Dr. Anne Praebel (NordGEN) on November 7 and 8, 2012, in Tuusula, Finland. The workshop will focus on the effects of climate change on animal genetic resources (AnGR) relative to prioritizing resource allocation in active populations and conservation, incorporating genomic selection signatures on AnGR conservation schemes as well as development of new animal breeding and conservation strategies. It will provide a multidisciplinary view merging state of the art knowledge on quantitative, population and molecular genetics within both

breeding and natural populations to increase the understanding of climate change impacts, adaptation and mitigation with respect to the sustainable use of farm AnGR.

The workshop is organized in three sessions each having three or four invited speakers: a) Genotype-by-Environment Interactions; b) Farm Animal Adaptation on Phenotypic and Genetic Level; and 3) Genotype-by-Environment Interactions in Practical Breeding and Conservation Programs.

In addition, a poster session will be arranged.

Detailed information available here: http://www.nordgen.org/index.php/en/content/view/full/2114/

Exchange Visits 2011-2012: final report of the 6 funded projects

We carried out a light publishing work on the following texts. But the content corresponds to the final reports transmitted by the recipients of the grant. All reports can be downloaded from the GENOMIC-RESOURCES website.

Project N.1

Simulation study of the effect of gaps in pedigrees on effectiveness of conservation of genetic diversity

Sebastian Mucha, From Poznan University of Life Sciences, Poland hosted by Wageningen University, The Netherlands.

Purpose of the visit

Preservation of biodiversity is essential for the quality of life of future generations. According to the World Conservation Monitoring Centre (WCMC), 905 species of plants and animals died out and about 16.928 are threatened with extinction. Development of effective and sustainable breeding programs is an important component of preservation of biodiversity. Additional research is needed on restriction of inbreeding levels resulting from breeding programs within breeds. High levels of inbreeding increases the risk of extinction of a breed.

Inbreeding is the main factor reducing genetic diversity within small populations of livestock as well as wild animals. High levels of inbreeding have a negative impact both on reproduction and survival of individuals. Further, it may lead to elevated incidence of genetic diseases and disorders along with increased mortality. One of such examples is brachygnathy among deer where a lower jaw is shorter than its upper counterpart. Moreover inbreeding leads to loss of alleles which results in increased homozygosity. This reduces adaptive capacity of populations and makes them more vulnerable to changes in environment. Breeds with an inbreeding rate above 1% per generation, are considered as critical by the FAO. Improvements in genetic management (breeding program) are needed for such populations to counteract the effects of small population size. Mating of closely related individuals leads to elevated inbreeding levels. Therefore, knowledge on coancestry between individuals plays a crucial role in genetic management. Degree of coancestry can either be estimated from marker or pedigree data. In the latter case it is essential to have accurate pedigree records over a long period of time. This information is often not available.

In captive populations, breeders decide which animals to mate in order to achieve the goal in terms of genetic diversity and characteristics of a population. In case of deer populations in zoos, for example, breeding coordinators aim at conservation

of the widest amount of genetic diversity. Unfortunately, often they lack a clear selection criteria to implement this objective due to the lack of accurate information on pedigrees and, therefore, on degree of coancestry. Moreover in some species accurate record keeping is a problem. Those issues can disrupt the conservation effort and may lead to problems with fitness of the populations in the future.

During my visit to Wageningen University I wanted to focus on an important part of my PhD research. It was a simulation study aiming at investigation of the impact of pedigree errors on conservation of genetic diversity. Moreover I aimed to compare three mating strategies with respect to their efficiency in conservation of genetic diversity. This should give an insight into the impact of gaps on the estimates of inbreeding level in the deer populations maintained in zoo gardens and the opportunities for using various breeding schemes to overcome such problems. The results should provide a better insight into genetic management of small populations which is important both for wild species such as deer as well as livestock populations.

Description of the work

Three populations resembling real deer populations maintained in zoo gardens were simulated. They differed in size as the first was a big population of 1200 (pop20), the second of 240 (pop6) and the third of 60 (pop3) individuals born each generation. Populations were divided into a number of subgroups reflecting the real number of zoo gardens (20, 6 and 3, respectively). They went through 20 generations of mating based on three scenarios. For the purpose of the simulation it was assumed that animals reproduced only once and died afterwards (discrete generations). In order to maintain a constant population size every female was used for mating and produced two offspring. It was assumed that the ratio of male to female offspring born each generation (sex ratio) was 1:1. There was no selection in females and only males were subject to selection. Four different mating ratios of males to females (1:1, 1:2, 1:5 and 1:10) were used. Single trait related to fitness, with a heritability of 0.3 and variance of 1.0, was included in the simulations.

First scenario (similar to wild conditions) assumed that a population was divided into herds. In each herd, there were dominant males with a priority for breeding. Males with the highest phenotypic value for fitness obtained the highest contributions to the next generation. Contributions of males were drawn from a gamma distribution, thus few males had large contributions and most of the males had minor or no contributions. It was supposed to mimic fight for dominance in which the fittest males produce the most offspring. Each generation animals could migrate between herds. Percentage of migrants for males was between 0 and 30% (depending on a scenario) and was much higher than for females where it was between 0 and 5%. Pattern of migration was random, thus in each generation animals could migrate to a different herd or remain in their native herd. It was assumed that both males and females migrated to the same destination herd.

In the second scenario based on optimal contribution mating (Meuwissen 1997) animals were given contributions so that both the coancestry of parents ( ) and genetic level of the parents (G) were optimized as: G=c'EBV and , where c is a vector of contributions of parents, EBV is a vector of estimated breeding values of parents, A is an additive relationship matrix of selection candidates. was calculated as where is a desired mean coancestry, ΔF is a desired rate of inbreeding. Breeding values were estimated with an animal model in ASReml package (Gilmour et al. 2008).

Optimal contributions were obtained from package Gencont (Meuwissen 2002). Given a certain mean coancestry and desired rate of inbreeding a solution for was obtained. Optimal contribution was evaluated in two different alternatives. First we performed it treating the whole population as a single herd (metapopulation). Afterwards the population was divided into herds (with no exchange of animals between the herds) and optimal contribution was performed within each herd separately. Subsequently the simulation was performed with the same division into herds, but sires were exchanged between the herds in a circular manner (similar to a breeding circle). All of the optimal contribution scenarios were repeated with 5-40% gaps introduced to pedigrees. The gaps were applied by erasing both parents of randomly chosen individuals.

The third breeding scheme was based on rotational mating. Breeding circles were formed so that the first herd provides sires for the second herd, the second herd provides sires for the third herd and the last herd provides sires for the first herd. This method does not require pedigree information and thus is not sensitive to missing pedigree information. In this scenario contributions of males to the next generation were based on a fixed mating ratio as described before, or obtained from a gamma distribution. In each case only the fittest males contributed to the next generation.

Simulation were performed using a custom program written in R language. At the end of each simulation, inbreeding coefficient, mean kinship, genotypic and phenotypic mean were calculated per each generation. All of those parameters were calculated on two levels. First calculations were performed for a population as a whole. Subsequently all parameters were calculated separately for each herd from a particular population. This served as a basis for comparison of scenarios.

Description of the main results obtained

The highest genetic progress in all scenarios was observed with the highest mating ratio of 1:10 and the smallest with mating ratio of 1:1. In the rotational scenario the highest genetic level was 2.55 SD for the smallest population (pop3), and 3.56 SD for the medium population (pop6). In the wild mating scenario a similar trend was observed, with 2.31 SD, and 2.95 SD for pop3 and pop6, respectively (Table 1). Optimal contribution performed in a metapopulation resulted in the highest genetic gain of all scenarios, with 2.93 SD and 4.09 SD for pop3 and pop6, respectively. If the population was subdivided into herds, and optimal contribution was performed within each herd separately with no exchange between the herds, genetic gain was smaller. In case of the 1:10 mating ratio it was close to zero, which indicates that it was not possible to achieve genetic gain and restrict inbreeding at the same time. On the other hand when optimal contribution was performed within herds but there was exchange of sires between the herds on rotational basis, genetic level reached 2.49 SD and 1.08 SD for pop3 and pop6, respectively.

Inbreeding rate (ΔF) was larger with increasing mating ratio, with the highest values for the ratio of 1:10. Overall, the lowest increase in inbreeding was observed in rotational mating scenarios, where ΔF (for 1:10 mating ratio) reached 3.25% and 0.97% for pop3 and pop6, respectively. In the wild mating scheme inbreeding rate was lower reaching 1.48% and 0.49% for pop3 and pop6, respectively. The highest ΔF was in the optimum contribution scenarios. It reached 5.58% and 2.53% for pop3 and pop6 treated as a metapopulation. When populations were divided into herds and optimal contribution was performed within herds with no exchange of animals inbreeding rate reached 10.99% and 4.32%, respectively. If rotational exchange

of sires was permitted, ΔF was 3.24% and 0.81% for pop3 and pop6, respectively.

Table 1. Phenotypic level (P), genotypic level (G), inbreeding (F), inbreeding rate (ΔF), mean kinship (Kin), minimum and maximum inbreeding rate (ΔFw) and kinship (Kinw) within herds for the simulated populations of 3, 6, and 20 herds marked as pop3, pop6, and pop20, respectively.

Scenario	Pop	P	G	F	ΔF	Kin	ΔFw	Kinw
Rotational	pop3[a]	2.55	2.57	0.45	3.25	0.49	0.43-3.26	0.55-0.55
	pop6[a]	3.56	3.56	0.18	0.97	0.17	0.11-0.99	0.26-0.26
Wild	pop3[b]	2.31	2.31	0.27	1.48	0.24	1.41-3.74	0.31-0.32
	pop6[b]	2.95	2.96	0.11	0.49	0.08	0.47-2.33	0.14-0.14
Optimal contribution metapopulation	pop3[a]	2.93	2.91	0.67	5.58	0.68	-	-
	pop6[a]	4.09	4.10	0.38	2.53	0.38	-	-
Optimal contribution subdivided population	pop3[a]	0.15	0.14	0.89	10.99	0.30	3.45-10.99	0.91-0.91
	pop6[a]	-0.01	-0.01	0.64	4.32	0.10	1.99-5.54	0.58-0.62
Optimal contribution with rotation of sires	pop3[a]	2.49	2.49	0.45	3.24	0.49	0.96-3.25	0.55-0.55
	pop6[a]	1.08	1.07	0.16	0.81	0.14	0.31-1.01	0.22-0.22

[a] *Mating ratio of 1:10*
[b] *5 % of dams and 30% of sires migrate*

Similarly to ΔF, mean kinship was higher with increasing mating ratio. The highest mean kinship was obtained for the 1:10 mating ratio and those values are used for comparison of scenarios. Mean kinship, calculated over the whole population, in rotational mating schemes reached 0.49 and 0.17 for pop3 and pop6, respectively. Kinship within herds was similar for all herds from a given population, and reached 0.55 and 0.26 for pop3 and pop6, respectively. Optimal contribution scheme performed in metapopulations, performed worse than rotational mating. It resulted in mean kinship of 0.68 and 0.38 for pop3 and pop6, respectively. When populations were divided into herds and optimal contribution was preformed within herds with no exchange of animals, mean kinship of the whole population was lower. It reached 0.30 and 0.10 for pop3 and pop6, respectively. Even though overall mean kinship was low, within each herd kinship increased with time and reached 0.91 and 0.62 for pop3 and pop6, respectively. Better result were obtained when populations were managed with optimal contribution within herds and rotational exchange of males between herds in each generation. In this scenario mean kinship of the whole population after 20 generations was 0.49 and 0.14 for pop3 and pop6, respectively. Kinship within herds was also restricted in this situation and was equal to 0.55 and 0.22 for pop3 and pop6, respectively. Wild mating scenario turned out to be efficient in terms of restriction of kinship. Kinship in scenarios with 5-30% migration was similar and achieved higher values than in the scenario with no migration. Overall mean kinship for the scenarios with migration was 0.24 and 0.08 for pop3 and pop6, respectively. With no migration kinship reached 0.18 and 0.06, respectively. On the other hand kinship within herds was the highest in the scenario with no migration reaching 0.53 and 0.37 for pop3 and pop6, respectively.

Table 2 Inbreeding rate (ΔF), mean kinship (Kin), minimum and maximum inbreeding rate (ΔFw) and kinship (Kinw) within herds based on the observed pedigrees (with gaps), along with analogous parameters calculated based on a true pedigree (without gaps) marked as ΔFc, KinC, ΔFwC, and KinwC fot simulated populations of 3, 6, and 20 herds marked as pop3, pop6, and pop20, respectively.

Scenario	Pop	ΔF	Kin	ΔFw	Kinw	ΔFc	KinC	ΔFwC	KinwC
Optimal contribution, metapopulation, 25% gaps	Pop3[a]	0.03	0.06	-	-	5.88	0.69	-	-
	Pop6[a]	0.07	0.03	-	-	2.75	0.38	-	-
Optimal contribution, subdivided population with no exchange of animals, 25% gaps	Pop3[a]	-0.06	0.05	-0.20 0.74	0.13 0.15	11.49	0.31	11.46 11.55	0.92 0.92
	Pop6[a]	0.04	0.02	-0.03 0.92	0.09 0.09	7.30	0.13	6.96 7.38	0.77 0.78

[a] *Mating ratio of 1:10*

Pedigree gaps had a strong effect on effectiveness of conservation programs. With increasing percentage of pedigree gaps, difference between the observed and true kinship was larger. In case of the smallest population (pop3) maintained as a metapopulation with 5% gaps, observed kinship was equal to 0.09, whereas the true kinship was 0.22. When the amount of gaps increased to 25% the difference was even larger as the observed kinship was 0.06 and the true kinship was 0.69 (Table 2). Similar differences between true and observed kinship were present regardless of the mating ratio. Pedigree gaps had a smaller effect on estimates of kinship when a population was divided into herds. In such case the observed kinship for pop3 was 0.05 whereas the true kinship was 0.31. In all cases there was also a large difference between the scenarios with and without gaps in terms of inbreeding rate. It had strong fluctuations in the scenario with gaps, whereas it was stable when optimal contribution was performed with complete pedigrees.

It can be concluded that optimal contribution did not achieve satisfactory results in management of small populations. Majority of scenarios with optimal contribution led to rates of inbreeding which exceeded 1% threshold above which populations are considered as critically endangered by the FAO (1998). This can be partially explained by the mating ratios imposed in the simulations. Most likely mating ratios of 1:5 and 1:10 are too strict as they force the method to select very few males for breeding. In such a setting it could be hard to obtain an optimal solution.

Only a very complex scenario where optimal contribution was combined with rotational exchange of sires proved to be efficient in restriction of inbreeding. On the other hand a much less complicated scheme based on breeding circles proved to be efficient in the simulated populations. It was particularly more efficient in comparison with optimal contribution performed on incomplete pedigrees. Therefore rotational mating schemes can be recommended for management of genetic diversity of small populations. This strategy can be implemented both for wild animals in zoo gardens and local livestock breeds. Another advantage of rotational mating is that it can accommodate harem based structure characteristic for most of deer species.

Future collaboration with host institution

Future collaboration with Animal Breeding and Genomic Centre of Wageningen University will involve work on manuscripts describing findings of the research carried out during the exchange grant. Moreover additional analysis will be performed to obtain results for the largest population (pop20) included in the analysis. This should provide a more complete picture of the effect of using different strategies for conservation of genetic diversity.

Projected publications / articles resulting or to result from the grant

Results of the research carried out during the visit are going to be published. At the current stage of the project it is anticipated that one publication is going to be submitted. It is possible that a total of two publications will be submitted as a result of the project. First paper is going to deal with aspects of comparison between wild, rotational and optimal contribution mating. In the second paper we would like to focus on the effect of gaps in pedigrees on effectiveness of conservation of genetic diversity.

Project N.2

SNP chip data analysis for traditional Hungarian sheep and cattle breeds

Levente Czegledi, From University of Debrecen, Institute of Animal Science, Hungary hosted by Cardiff University, UK

Purpose of the visit

The aim of this exchange grant is to learn bioinformatic tools for the molecular characterization of Hungarian autochthonous domestic breeds with new molecular genetic methods and to assist in the genetic management of these breeds that has so far been lacking.

Our research group and institute will benefit from this exchange grant because of learning the state-of-art SNP based genetic analyses. Acquisition of this skill-set will dramatically improve population genetic studies at the University of Debrecen, Institute of Animal Science, which has focused only on using microsatellite markers up to now.

Molecular methods provide significant tools in population and conservation genetics for characterisation and genetic evaluation of species, breeds, varieties of wild and domestic animals. In recent years, microsatellite based analysis were dominant, but nowadays, methods based on genotyping for point mutations in the genome are more frequently used. Illumina 50k SNP chips data set provides huge amount of data of which handling, analysing was planned to carried out at the Organisms and Environment Research Division, Cardiff School of Biosciences, Cardiff, Wales. The high-density 50k-marker panel comprises approximately 50,000 single nucleotide polymorphisms (SNP) that have been identified in the bovine and ovine genome.

The results can enable in-depth analysis of the patterns of genome-wide diversity and differentiation across each breed, which can be used for genome-wide association studies and for future marker-assisted selection. Our aim was to use a variety of bioinformatic tools to analyse the high volume of data in the context of population genetics. This kind of analysis is carried out routinely within the host lab, thus Cardiff provided an ideal location for this training.

Description of the work

Before the visit I have collected blood samples from Hungarian Grey cattle and Racka sheep as well. Animals of each species were divided into 2 groups: a) moderate performance individuals, production level of these animals is as the average of the breed; b) high performance individuals, these animals had the highest weight gain in the investigated population. Data of individuals – all of them are adult females before first calving – are date of birth, sire, dam, daily gain (from birth until weaning), additional information for sheep: weight at the beginning of grazing season, at end of grazing season, at age of 1 year.

DNA was extracted from blood and its quality was checked. Samples are on Illumina 50K SNP chip for lab analyses. All the costs of SNP chip work is financed by Prof. Mike Bruford's grant.

The primary aim of the visit was to study the application of data handling and bioinformatic tools for huge DNA data analyses. Prof. Mike Bruford provided me the 50K Sheep SNP chip data of Kijas JW et al. (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biol. published in February 2012 by the members of International Sheep Genomic Consortium. This dataset consisted of more than 49000 point mutation results per individual, 2819 animals belong to 83 breeds.

At first, I have started to be familiar with UNIX system. Commands of UNIX were used to manage, organize SNP chip result files. Unix operating systems are widely used and it has a special language.

The most important and appropriate software tools were PLINK and Haploview for chip dataset analysis.
PLINK is a whole genome association analysis toolset, freely available and developed by Center for Human Genetic Research, Massachusetts General Hospital, and the Broad Institute of Harvard & MIT (Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for wholegenome association and population-based linkage analysis. American Journal of Human Genetics, 81.). The software was used in UNIX environment.
The options which were used for working on .ped and .map files of chip data are grouped as follows: basic processes on data files; selection SNPs, individuals, breeds; creating summary statistics for whole dataset and subgroups; using thresholds; stratification and clustering; association analysis procedures; linkage disequilibrium pruning.
gPLINK is a free software as well, developed by the same groups. It is a simplified, user friendly form of PLINK, it contains the most commonly used PLINK commands. One can use gPLINK by clicking, however for PLINK correct command typing is required.
Haploview is a free software to visualize the analysis of PLINK results. It was developed and available from the Broad Institute (Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15, PubMed ID: 15297300). Haploview is able to use any output file from PLINK, it can run linkage disequilibrium and haplotype analysis, permutation testing for association, SNP and haplotype association tests, haplotype population frequency estimation.

Description of the main results obtained

Main result of exchange visit is the knowledge of state-of-art bioinformatic tools. All the softwares mentioned in „description of work…" section provides the possibility to work on huge amount of SNP data. The experience I got at Mike Bruford' lab is enables me to handle and work on my own genotype data of domestic animals and the gain of these skills makes it possible to apply for research grants in the field.

Future collaboration with the host institution

Mike Bruford's lab and the Animal Genetics lab at University of Debrecen, Institute of Animal Science plan to continue the work that has been started with this grant. At present, the SNP chip data of Hungarian samples (cattle and sheep) will be analysed as a pilot study. The future collaboration will focus on domestic animal species, that could include autochthonous breeds for diversity studies or/and breeds with economic importance/intensive production for SNPs correlated to phenotype data, animal performance.

Projected publications / articles resulting or to result from the grant

The grant gave the possibility to study bioinformatic tools for genome wide mutation analyses at the host institution. It provides a strong background to apply for research grants in the future to finance the cost of consumables, materials required for large scale studies. We do believe that this type of common work will result in publications.

Project N. 3

Molecular dissection of inbreeding depression for bull fertility

Maja Ferencakovic, from the University of Zagreb (Croatia), hosted by the University of Natural Resources and Applied Life Sciences Vienna, Vienna (Austria)

Purpose of the visit

The main objective of the visit was to analyze inbreeding depression for bull fertility on a molecular level. Inbreeding depression is the reduction of the population mean for a quantitative trait such as size, fertility, vigor, yield, and fitness. Maximini et al. have shown that inbreeding depression does affect male fertility of Simmental bulls using pedigree and sperm quality data. Pedigree data are usually difficult to obtain (impossible for wild animals), they are potentially unreliable, and do not count for inbreeding arising from distant common ancestors. Even if the pedigree is well known and correct the estimates of inbreeding for single individuals can differ from expectation due to the stochastic pattern of recombination.

In contrast to pedigree Runs of homozygosity (ROH) were recently proposed as genomic measure quantifying individual autozygosity that counts for stochastic variation of recombination (McQuillan et al., 2008; Nalls et al., 2009). In addition Keller et al. (2011) proposed ROH as optimal way for indentifying ancient inbreeding that we cannot obtain from pedigree data in humans, and Ferencakovic et al. (2011) in cattle. Estimation of precise inbreeding coefficient is important for separating influences of various environmental and physiological factors on quantitative traits, from genetic factors. Genome-wide association studies (GWAS) allow us to scan entire genome for associations between common gene variants (SNPs) and phenotype. Visscher (2008) reported a total of 54 SNPs influencing human height as classical quantitative trait. Furthermore, Pryce et al. (2011) reported orthologous positions of 55 genes associated with height in four human populations located on the bovine genome. Single nucleotide polymorphisms close to eight of these genes were significantly associated with stature in cattle.

To analyze inbreeding depression from molecular perspective our goal is to test the following hypotheses:

a) is genomic autozygosity associated to male fertility ?

b) does the pedigree inbreeding coefficient accurately estimate autozygosity ?

c) does ROH inbreeding coefficient gives better insight in inbreeding depression ?

d) is inbreeding depression caused by specific chromosomal regions or is it caused by genes from the whole genome ?

Description of the work

Day after my arrival I received data bases with information related to the bull sperm quality (ejaculate volume, sperm concentration, percentage of viable spermatozoa, total number of spermatozoa, motility score) from regular measurements done in Austrian AI (Artificial Insemination) stations. There were approximately 53 000 ejaculates obtained from around 1500 bulls from three breeds (Simmental, Brown Swiss and Tyrol Grey) from the Hohenzell Station from Upper Austria, from Wieselburg, the Station from Lower Austria sampled during six years, from Gleisdorf, the Station from Styria and Birkenberg, the Station from Tyrol. First goal was to clean existing data and use only genotyped animals. This resulted with 1237 animals and 50829 ejaculates

All those animals were genotyped using the Illumina bovine SNP chip with 54 001 SNPs. After quality control we were left with 42262 SNPs that are present in all three populations.

ROH were calculated using PLINK software. From five ROH lengths (>1Mb, >2Mb, >4Mb, >8Mb and >16Mb) inbreeding coefficients (FROH1 , FROH2 , FROH4, FROH8, and FROH16) were calculated as sum of all ROHs of specific length divided by length of genome covered with SNPs. We have also calculated inbreeding for every autosomal chromosome separately using those 5 ROH lengths. For comparison we calculated inbreeding from pedigree for whole pedigree (FpedT) and for five generation pedigree (Fped5) using ENDOG.

Next step was estimation of inbreeding depression. For this purpose we build mixed models with a bull as a random effect and all other effects (age of the bull, semen collector, month and year of collection and number of ejaculates per bull per day), including pedigree and genomic inbreeding coefficients (covariable), as fixed effects. Furthermore, we also used models with chromosomal inbreeding coefficients (1 to 29) as well as models with combination of several chromosomal inbreeding coefficients. Models were then evaluated using Burnham & Anderson method (2002). This last part of analysis we managed to perform only for Simmental bulls from stations Hohenzell and Wieselburg but it is applicable to other breeds and stations and further analyzes are in progress.

Description of the main results obtained

Our preliminary results are giving us some orientation answers on our basic questions and also providing us idea of further work.

a) is genomic autozygosity associated with male fertility ? c) does ROH inbreeding coefficient give better insight in inbreeding depression ? d) is inbreeding depression caused by specific chromosomal regions or is it caused by genes from the whole genome ?

Here we are able to conclude that genomic autozygosity is associated with male fertility. Maximini et al. 2010 estimated inbreeding depression using pedigree and sperm quality data on almost same data set as we were. We confirmed their results and found in majority of models presence of inbreeding depression while the strength of models (pedigree versus genomic inbreeding) interchanged depending on a traits or/and population analysed (Table 1-4.) In general, models based on pedigree were comparable to models based on genomic inbreeding (runs of homozygosity) while models based on individual homozygosity (ihom – proportion of homozygous SNPs) were inferior. In contrast, for all traits analyzed, models with various chromosomal inbreeding coefficients outperformed models based on overall genomic inbreeding as well as models based on pedigree. However, while genomic inbreeding coefficients have been shown to be powerful, still much work has left to complete our understanding on architecture of inbreeding depression.

Table 1 Influence of different inbreeding coefficients (covariates) on estimation of inbreeding depression on sperm volume in AI stations Hohenzel and Wieselburg

Volume (ml) Hohenzell					
Covariate	AIC	Δ_null	b	SEE	P
F_{ROH8_chr11}	44930,401	-7,838	-3,906	1,245	0,002
F_{ROH8}	44935,516	-2,724	-9,955	4,613	0,031
F_{ROH2_chr11}	44936,239	-2,000	-2,010	1,012	0,047
F_{ROH2}	44937,174	-1,065	-6,731	3,875	0,082
F_{PED}	44938,127	-0,112	-8,814	6,104	0,149
Null	44938,239	0,000			
F_{PED5}	44939,340	1,101	-6,112	6,491	0,346
ihom	44939,951	1,711	-4,557	8,550	0,594
Volume (ml) Wieselburg					
F_{ROH2}	59298,239	-8,701	-11,592	3,544	0,001
F_{ROH8_chr10}	59301,404	-5,536	-3,423	1,251	0,006
F_{ROH2_chr27}	59301,750	-5,190	-2,631	0,984	0,008
F_{ROH2_chr26}	59301,887	-5,054	-2,251	0,851	0,008
F_{ROH2_chr16}	59302,347	-4,594	-2,514	0,983	0,011
F_{PED}	59302,379	-4,562	-17,491	6,849	0,011
F_{ROH2_chr10}	59302,578	-4,362	-2,806	1,117	0,012
F_{ROH2_chr7}	59303,208	-3,733	-2,654	1,113	0,017
F_{ROH8_chr16}	59304,766	-2,174	-2,511	1,237	0,042
F_{PED5}	59304,869	-2,072	-13,915	6,938	0,045
F_{ROH8}	59305,151	-1,789	-8,869	4,586	0,053
ihom	59305,454	-1,487	-2,376	1,281	0,064
Null	59306,940	0,000			

Table 2 Influence of different inbreeding coefficients (covariates) on estimation of inbreeding depression on sperm concentration in AI stations Hohenzel and Wieselburg

Concentration (10^9/ml) Hohenzell					
Covariate	AIC	Δ_null	b	SEE	P
F_{ROH8_chr10}	6492,085	-2,358	0,481	0,232	0,038
Null	6494,442	0,000			
F_{ROH8}	6494,793	0,350	1,292	1,014	0,203
ihom	6495,158	0,715	-2,096	1,865	0,261
F_{PED}	6495,190	0,748	-1,478	1,332	0,267
F_{PED5}	6495,685	1,243	-1,225	1,421	0,389
F_{ROH2}	6496,435	1,993	-0,070	0,851	0,934
Concentration (10^9/ml) Wieselburg					
F_{ROH2_chr20}	7113,482	-6,354	-0,506	0,175	0,004
F_{ROH8_chr26}	7114,286	-5,549	0,487	0,178	0,006
F_{ROH2_chr26}	7115,780	-4,055	0,367	0,150	0,014
F_{ROH8_chr19}	7116,091	-3,744	-0,926	0,387	0,017
F_{ROH8_chr22}	7116,270	-3,565	0,435	0,185	0,019
F_{ROH2_chr7}	7117,844	-1,991	-0,392	0,197	0,047
Null	7119,835	0,000			
ihom	7119,915	0,080	0,309	0,224	0,168
F_{PED}	7121,630	1,795	-0,550	1,223	0,653
F_{ROH8}	7121,752	1,916	0,233	0,814	0,775
F_{ROH2}	7121,795	1,960	0,126	0,637	0,844
F_{PED5}	7121,822	1,986	0,143	1,232	0,908

Table 3. Influence of different inbreeding coefficients (covariates) on estimation of inbreeding depression on sperm viable spermatozoa in AI stations Hohenzel and Wieselburg

Viable spermatozoa (%) Hohenzell					
Covariate	AIC	Δ_null	b	SEE	P
F_{ROH2_chr14}	83790,548	-8,981	-13,555	4,092	0,001
F_{ROH2_chr27}	83795,327	-4,201	-14,057	5,686	0,013
F_{ROH8_chr14}	83796,541	-2,987	-10,620	4,793	0,027
F_{PED}	83797,087	-2,442	-56,042	26,828	0,037
F_{PED5}	83797,352	-2,176	-57,643	28,465	0,043
Null	83799,529	0,000			
F_{ROH2}	83799,592	0,064	-23,669	17,181	0,168
ihom	83800,004	0,475	-46,218	37,830	0,222
F_{ROH8}	83801,492	1,963	-3,934	20,568	0,848
Viable spermatozoa (%) Wieselburg					
F_{ROH8_chr16}	71361,375	-19,174	-12,050	2,627	0,000
F_{ROH2_chr16}	71367,643	-12,906	-8,032	2,091	0,000
Null	71380,549	0,000			
F_{PED}	71380,819	0,270	-18,950	14,544	0,193
F_{PED5}	71380,833	0,284	-19,056	14,683	0,194
F_{ROH2}	71381,142	0,593	-8,901	7,585	0,241
F_{ROH8}	71381,766	1,217	-8,478	9,701	0,382
ihom	71382,515	1,966	-0,498	2,711	0,854

Table 4. Influence of different inbreeding coefficients (covariates) on estimation of inbreeding depression on sperm motility in AI stations Hohenzel and Wieselburg

Motility Hohenzell					
Covariate	AIC	Δ_null	b	SEE	P
F_{ROH2_chr14}	11024,136	-10,400	-0,408	0,117	0,000
F_{ROH2_chr20}	11027,177	-7,359	-0,347	0,115	0,003
F_{ROH8_chr20}	11030,092	-4,443	-0,415	0,166	0,012
F_{ROH8_chr14}	11032,176	-2,360	-0,279	0,136	0,040
F_{ROH8_chr23}	11032,266	-2,270	0,416	0,205	0,042
F_{ROH2_chr19}	11032,386	-2,150	0,284	0,142	0,046
F_{ROH2_chr27}	11032,459	-2,076	-0,325	0,164	0,047
F_{ROH2_chr23}	11032,484	-2,052	0,263	0,133	0,047
F_{PED5}	11033,683	-0,853	-1,347	0,811	0,097
F_{PED}	11034,100	-0,436	-1,178	0,767	0,124
Null	11034,536	0,000			
F_{ROH2}	11035,199	0,663	-0,567	0,497	0,254
ihom	11035,757	1,222	-0,956	1,101	0,385
F_{ROH8}	11036,463	1,927	-0,161	0,592	0,786
Motility Wieselburg					
F_{ROH8_chr13}	26353,654	-11,680	-2,577	0,695	0,000
F_{ROH2_chr13}	26359,971	-5,362	-1,025	0,380	0,007
F_{ROH8_chr16}	26360,262	-5,071	-0,813	0,308	0,008
F_{ROH2_chr16}	26360,816	-4,517	-0,623	0,246	0,011
F_{PED5}	26365,031	-0,302	-2,603	1,730	0,132
Null	26365,333	0,000			
F_{PED}	26366,458	1,125	-1,588	1,718	0,355
ihom	26366,636	1,303	-0,266	0,322	0,407
F_{ROH2}	26366,895	1,561	-0,585	0,895	0,513
F_{ROH8}	26367,297	1,964	-0,214	1,145	0,852

b) does the pedigree inbreeding coefficient accurately estimate autozygosity ? Here we can conclude that pedigree inbreeding coefficient does not estimate autozygosity accurately. First, this approach fails to capture the influence of relatedness among founders from the base population. Second, FPED is the expected proportion of the genome that is IBD and does not take into account the stochastic nature of recombination. Third, several studies confirm that errors in cattle pedigrees are common due to misinterpretation, misidentification and incorrect recording (e.g., Ron et al. 1996). Finally, FPED assumes that the entire genome is selection-neutral and does not account for potential bias resulting from selection (Curik et al. 2002). Here we have calculated levels of autozygosity from pedigree data and from genomic data. From genomic data we calculated FROH and FHOM. FHOM is genomic inbreeding coefficient based on the difference between

observed and expected numbers of homozygous genotypes. It is clear from Table 5 that levels of autozygosity are much higher for FROH and FHOM. FPED and FPED5 are similar to values estimated from ROH of 8 or 16Mb which are considered to be from recent inbreeding.

Table 5. Mean, standard deviation and range of calculated inbreeding coefficients in three breeds.

	Brown Swiss			Tyrol Grey			Simmental		
	Mean	Std	Range	Mean	Std	Range	Mean	Std	Range
F_{PED}	0.048	0.020	0.001 - 0.127	0.030	0.024	0.003 - 0.169	0.014	0.013	0.000 - 0.09
F_{PED5}	0.024	0.017	0.000 - 0.106	0.018	0.023	0.000 - 0.159	0.009	0.012	0.000 - 0.085
F_{HOM}	0.138	0.036	0.047 - 0.264	0.076	0.036	0.024 - 0.239	0.066	0.026	0.001 -0.180
F_{ROH1}	**0.151**	**0.032**	0.069 - 0.273	**0.085**	**0.030**	0.041 - 0.234	**0.085**	**0.020**	0.028 -0.183
F_{ROH2}	0.125	0.032	0.046 - 0.252	0.060	0.031	0.017 - 0.213	0.052	0.019	0.009 - 0.150
F_{ROH4}	0.101	0.032	0.029 - 0.227	0.047	0.031	0.008 - 0.203	0.030	0.017	0.002 - 0.124
F_{ROH8}	0.072	0.029	0.012 - 0.194	0.035	0.03	0.000 - 0.183	0.016	0.016	0.000 - 0.102
F_{ROH16}	0.037	0.023	0.000 - 0.154	0.019	0.024	0.000 - 0.140	0.008	0.012	0.000 - 0.062

Future collaboration with host institution

The working groups of Johann Sölkner and Ino Curik have been collaborating on various issues related to inbreeding and inbreeding depression from some time. The availability of high throughput genotypes and male fertility phenotypes allows much deeper insight and tests of some of the hypotheses the two groups had pointed to in previous work.

This project represents the start of a future collaboration between these two groups as follow up of our previous work. We have just started to use genomic data as tool for understanding inbreeding and its construction.

Projected publications/articles resulting or to result from your grant

Results from this study will be partly presented at 4th International conference in quantitative genetics 17 - 22 June 2012 Edinburgh, Scotland, UK., and at the 63rd Annual Meeting of the EAAP which will take place in Bratislava from August 27 to August 31, 2012. We are also planning a paper of general interest to be published from our results.

Project N. 4

Estimating Age of Admixture in Cattle Populations based on SNP Chip Data

Anamarija Frkonja, from University of Natural Resources and Life Sciences, Vienna, (Austria) hosted by University of Liege, Liege (Belgium)

Purpose of the visit

The research visit of Anamarija Frkonja was dedicated to exploration of the methods and models to predict age of admixture in crossbred livestock populations (cattle). Such methods were already proposed for human population and rely on the sizes of chromosomes fragments inherited from the ancestral populations. However, in human populations admixture is relatively old and most often, only one event occurred. We started by investigating how accurately we can predict the "breed" origin of a small chromosome fragment (a set of consecutive markers).

SABER (Tang et al., 2006) and Step PCO (Pugach et al., 2011) were first compared. Then we developed another method based on phased haplotypes (distribution of haplotype lengths). We checked how good methods are dealing with recent admixture with high throughput single nucleotide polymorphism (SNP) data from the bovine 50K SNP Chip,

Description of the work done during the exchange visit

Age of admixture derived from autosomal part of genome can potentially provide more precise estimates of age of admixture than data from sex linked chromosomes. Furthermore, as selection is expected to have smaller effects over short distances (few generations) it is important to know age of admixture to be able to select animals for admixture mapping. Signals of selection can be obtained more clearly as age of admixture is older.

Tang et al. (2006) are using Markov-hidden Markov model (MHMM) which accounts for background LD in ancestral populations. The model estimates allele frequencies in ancestral populations is based on pure individuals and then uses that information to estimate population origin at each marker population for each admixed individual. In addition, a parameter related to age of admixtures is estimated (it relies on recombination rate between ancestral haplotypes or distribution of haplotype length within one individual). With such a parameter, it is assumed that there is a single admixture event per individual and that both chromosomes from a pair present the same pattern of admixture.

Pugach et al. (2011) described a PCA-based genome scan approach to analyze genome-wide and local admixture structure, and introduced wavelet transform analysis as a method for estimating the time of admixture.

Ways of using information about population structure is extremely relevant in the field of genome wide association studies.

Hypothesis that ancestral haplotypes break down after admixture and will be shorter with ageing of admixture event was investigated with SABER software, StepPCO, as well as scripts developed during this research visit. For analysis data of Swiss Fleckvieh and crosses of Jersey and Holstein from New Zealand were used. Swiss Fleckvieh was established about 40 years ago from crossing of a large part of the local Simmental breed with Red Holstein Friesian cattle for improved milk production. SNP Chip data for 500 animals, 101 pure Red Holstein Friesian, 91 Simmental and 308 crosses with a wide range of pedigree based levels of admixture were available. For New Zealand cattle, data of 3286 animals, 1658 Holstein Friesian, 1244 Jersey and 384 crosses were used.

Quality control was performed with PLINK (Purcell et al. 2007), excluding markers with a call rate below 0.90. STRUCTURE software (Pritchard et al., 2003) was run on SNPs from all chromosomes assuming that breeds are separated with two ancestral populations (Frkonja et al. 2012). Admixture proportions levels are then used as input data for SABER. Algorithms implemented in SABER identify ancestry blocks using two different hidden Markov models. The first approach uses a hidden Markov model to infer gene flow. The second approach incorporates a `Markov-hidden Markov model' (MHMM) to account for linkage disequilibrium present in ancestral populations. SABER requires two datafiles for calculation. One file contains genotypes and distances between markers in base pairs (bp). A file containing admixture proportion for every admixed individual should be provided as well. SABER takes a single admixed individual, and estimates, over all chromosomes, the parameters describing the admixture times of an individual, given ancestral population allele frequencies and other information. Conditional on model parameters, SABER models the ancestral states along the paternal and the maternal chromosomes as two independent and identical Markov processes. Because of the constraints imposed by an underlying genealogy, the process along each chromosome is not Markovian. The paternal side of the genealogy and the maternal side of the genealogy may have different levels of admixture, and, therefore, the two processes are not necessarily identical. They assume that matings are random with respect to ancestry, an assumption that may be violated in some populations. In both populations of cattle investigated here admixture events happened more than once and admixture on both parental chromosomes does not have to be equal.

In a first SABER analysis, 24 simple crosses according to pedigree were selected

to be able to compare true age of admixture from pedigree with estimated age of admixture calculated with SABER.

Initial values (our estimation of age of admixture) for age of admixture need to be provided to SABER. We noticed that with different starting values, posterior estimates are changing quite substantially. Part of SABER results obtained with initial values set equal to 5 and 5 generations since admixture are shown in Table 1. Results reflected the pedigree admixture process well for intermediate but not extreme crosses (> 0.90:0.10). In that case, one parent is correctly considered to be purebred while the number of generations of admixture estimated for the other only reflects the input parameter. F1 animals were sometimes not recognized by SABER and were assigned as older crosses. Results for crosses of Holstein and Jersey were similar.

StepPCO (Pugach et al. 2011) is using principal component analysis for admixture calculation and wavelet transform for calculation of age of admixture. For analysis Pugach et al. (2011) used recombination events taken from international HapMap project while we calculated cM distances according to assumption that 1 cM = 1 Mb. This is a simplification ignoring for instance recombination coldspots or hotspots. All the recoding and preparation was done with SAS software (SAS 2009). In order to run wavelet transform analysis, size and number of sliding windows need to be selected. Sliding window is taking several genotypes and according to these decides whether this window shows a recombination. The number of points which serve as a center for every window was chosen so that the windows span the entire chromosome, leaving no gaps in between. Starting from the center of each window, window is increased until the mean PC1 coordinates for the parental populations are separated by three standard deviations from each mean. The goal is to achieve a complete separation of the parental populations within each window, so there is no ambiguity in assigning chromosomal segments in an admixed genome to either ancestral population. Our attempt to differentiate breeds from small sets of consecutive SNPs was not successful. Indeed, with our population structure and our marker density, it was not possible to separate perfectly both breeds for small or intermediate chromosome fragments.

Since SABER was not performing well on our crossbred data sets (with several and recent crosses) and since StepPCO was not able to correctly separate both breeds for small chromosome fragment, we investigated how well can we predict with STRUCTURE the population origin in our data set for different window sizes. We selected purebred animals and took bins of 10, 20, 50, 100 and 200 consecutive SNPs, which were tested. The MCMC procedure was run with 10,000 burnins and 10,000 iterations. Differences between estimated and real breed proportions were calculated. Minimum, maximum, mean, standard deviation of these values are reported in Table 2. In addition, accuracy of breed origin estimation (r2) is also presented in Table 2.

A new program developed during the stay was used to study age of admixture. The method models chromosomes (both copies of a pair are modelled with different parameters) as a mixture of 20 distributions. 10 distributions for each ancestral haplotype origin and one distribution for different ages of segments (from 0 to 9 generations since admixture). The program works also with HMM where different parameters are known (allele frequencies in different populations, probability of recombination as a function of time in each distribution). Therefore, only the proportion of each distribution must be estimated. Algorithm is based on Tang et al. 2006 formula which assumes that smaller recombination distance will

give proportionally bigger τ. Under a simple intermixing model and when d (distance between two markers) is measured in Morgans, τ has the interpretation of the time since admixing. Since gene flow may have occurred over many generations continuously, one should be cautious about equating τ with the admixing time. Nonetheless, this parameter provides some information regarding average time of gene flow. In this case τ was fixed for every of 20 groups. Model can be expressed by equation:

With increase of admixture proportion we will get smaller impact of distance and big τ. According to distribution classification of haplotypes it is possible to derive age of admixture for young crosses.

Running this algorithm on Holstein-Jersey crosses we obtained high correlation of averaged admixture of an animal with admixture derived with pedigree data. Spearman's correlation of pedigree derived admixture with software calculated admixture was 0.959. Mean of 20 distributions is in Table3. In that table, proportions of genome within different distributions (function of breed and age of admixture) are presented. Values of mean indicate that most of the animals are classified as young crosses.

Different approaches to the problem led us to conclusion that with a denser marker set it would improve the precision of deriving age of admixture.

The parameter, τ, approximates the average time since admixing and is of particular interest in admixture studies.

In attempt to estimate age of admixture for Swiss Fleckvieh crosses or Holstein Jersey we ran to problem that populations are not able to separate with StepPCO with 50 000 SNPs. In order to make analysis with StepPCO possible denser marker sets or even sequence data are required.

Our conclusion can be taken into account under assumptions of studying same breeds with same number of markers.

Table 1. SABER results (τ, approximate time since admixture happens) shown with proportion of Red Holstein and Simmental calculated with STRUCTURE

	Admixture proportion (STRUCTURE)	SABER age of admixture estimate		Pedigre average age of admixture	
CHEM712896043369	0,466	5-7	5-6	2,5	1,5
CHEM120029125934	0,499	4-6	4-5	4,38	4,6
CHEM712896043635	0,495	5-8	5-8	1,5	1,5
CHEM716270049817	0,513	5-7	5-6	0	3
CHEM713473108563	0,519	5-9	4-7	0	0
CHEM712028058025	0,575	5-7	4-7	2,5	2,5
CHEM120016865904	0,538	6-7	5-7	4,3	4
CHEM120021261555	0,524	3-4	3-4	0	4,6
CHEM713306111623	0,616	4-8	4-6	0	3,83
CHEM120029383716	0,999	0-11	0	0	6,25

Table 2. Separation of breeds (Red Holstein and Simmental) with different number of markers

Number of markers		Min avg. Error for breed 1	Min avg. Error for breed 2
10	Mean	0.27	0.29
10	Min	0	0.0008
10	Max	0.89	0.91
20	Mean	0.21	0.25
20	Min	0.005	0.004
20	Max	0.85	0.92
50	Mean	0.11	0.15
50	Min	0.005	0.008
50	Max	0.74	0.865
100	Mean	0.05	0.056
100	Min	0.005	0.004
100	Max	0.51	0.7878
200	Mean	0.02	0.01436
200	Min	0	0.0008
200	Max	0.09	0.2339

Table3. Distribution of haplotypes according to the length in Holstein Jersey crosses

Distribution Breed 1	1	2	3	4	5	6	7	8	9	10
Mean	0.973	0.443	0.131	0.08	0.058	0.047	0.04	0.035	0.031	0.028
Min	0	0,007	0	0	0	0	0	0	0	0
Max	0.973	0.443	0.131	0.08	0.058	0.047	0.04	0.035	0.031	0.028
Distribution Breed 2	11	12	13	14	15	16	17	18	19	20
Mean	0.991	0.443	0.134	0.084	0.062	0.05	0.043	0.037	0.034	0.031
Min	0	0,001	0	0	0	0	0	0	0	0
Max	0.991	0.443	0.134	0.084	0.062	0.05	0.043	0.037	0.034	0.031

Future collaboration

The present study showed that with 50K marker density and such cattle populations, it is difficult to correctly estimate breed origin locally (for a small genomic region). Marker windows of 100 to 200 SNPs are required (5 to 10 cM). Haplotype strategies were able to extract more of information. With such resolution, it would be difficult to correctly estimate age of admixture and

to perform accurate mapping. We also tried to develop models more appropriate for recent crosses but further investigations are required.

Projected publications/articles resulting or to result from the grant

Publication comparing SABER, StepPCO and method developed during the research is planned to be submitted to a high ranking journal after contination of the current collaboration and adding quite some extra analyses before submission. Targeted submission date is november 2011.

References

Frkonja A. et al. (2012) Prediction of breed composition in an admixed cattle population. Animal Genetics.

Pugach I, et al. (2011) Dating the age of admixture via wavelet transform analysis of genome-wide data. Genome Biology 12, R19.

Tang H. et al. (2007) Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. The American Journal of Human Genetics 81, 626-633.

Pritchard J. K., et al. (2010) Documentation for Structure software, Version 2.3

Pritchard J., et al. (2000) Inference of population structure using multilocus genotype data. Genetics 155, 945 - 59.

Purcell S, et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Project N. 5

Identification of signals of selection in the genomes of Spanish dairy sheep breeds.

Dr. Beatriz Gutierrez Gil. from University of León, León (Spain) hosted by The Roslin Institute, The University of Edinburgh (United Kingdom).

Purpose of the visit

The main objective of the visit was to perform different analyses to identify selection sweeps in dairy sheep. The Spanish group of the candidate researcher, Dr Beatriz Gutiérrez-Gil, established at the University of León (Spain), has a long tradition in the field of dairy sheep genetics and has performed several studies to gain knowledge on the quantitative and molecular genetics aspects that control the architecture of traits of interest in dairy sheep. At the molecular level, this group has performed genome scans for detection of QTL influencing milk production traits, morphological traits and disease resistance in Churra sheep, a highly specialized autochthonous breed of the region of Castilla y León. These genome scans were initially based on microsatellite genotyping (181 markers analysed; Gutiérrez-Gil et al., 2008, 2009a) and have been recently been performed using the OvineSNP50BeadChip (with about 45000 markers analysed after quality control). This medium-density high throughput SNP genotyping platform also offers the opportunity to perform selection sweeps, a complementary methodology to identify regions of the genome that have undergone selection pressure and that has not previously been implemented by the group of the University of León.

Based on previous collaborations between Dr Beatriz Gutiérrez-Gil and Dr Pam Wiener during Beatriz' postdoctoral fellowship at the Roslin Institute, Edinburgh (UK), this visit was planned as an opportunity to apply a method of selection sweep mapping that Pam Wiener has recently developed to identify selection sweeps in dairy sheep. This method is based on a regression approach for detection of regions with decreased heterozygosity (Wiener & Pong-Wong, 2011). Additionally analyses based on basic Fst and heterozygosity calculations across the genome would be performed, which would give us also the good opportunity to compare the results between methodologies and to implement the method described by Wiener and Pong-Wong (2011) on a wholegenome dataset.

Description of the work

With the purpose described above we have analyzed a number of samples from four Spanish sheep breeds analyzed with the

Illumina OvineSNP50BeadChip. Three of the breeds studied are autochthonous sheep breeds from the region of Castilla y León, where the University of León is located: Churra (dairy aptitude breed and also exploited for lamb production), Castellana and Ojalada (these two breeds are used mainly for lamb production, although the Castellana breed is also milked). Another Spanish breed included in our analyses was Rasa Aragonesa, specialized in meat production. The number of samples analysed were 96 for Churra and 24 for each of the other three breeds. After an initial quality control, we analyzed a total of 44546 SNPs distributed across the 26 autosomic ovine chromosomes

We performed three different analyses:

1) Calculations of pairwise Fsts: We estimated averaged pair-wise Fst values between Churra and each of the other three non-dairy breeds, using sliding windows of 9, 13 and 17 SNPs. Two different thresholds (quantiles of 0.005 and 0.0005) were considered for the identification of the top regions derived from this analysis.

2) Calculations of Normalized Heterozygosity: For each breed, we calculated normalized heterozygosity values averaged in sliding windows of 9, 13 and 19 SNPs individually for all the breeds included in the Fst calculations. Based on the normalized values and following Rubin et al. (2010), the estimates showing values lower than -6 were considered as significant.

3) Regression-based analysis for detection of regions of decreased heterozygosity: This method assumes that heterozygosity was at similar levels across the genome prior to the selective sweep and that an asymptotic increase in marker variation (heterozygosity) with increasing distance from a locus under directional selection will be observed. For each breed under study three different bracket sizes from the test positions (5, 10 and 20 Mbp on either side of the test positions) and three different thresholds (quantiles of 0.01, 0.005 and 0.0005) were considered for the selection of the most interesting regions revealed by this analysis.

Description of the main results obtained

The three different analyses applied have yielded a large number of regions to consider depending of the breed or pair of breads analyzed, and also depending on the different parameters considered (top levels and bracket or window size). As we had previously foreseen in our project proposal, and based on the short length of the ESF funded stay (2 weeks), we did not have time during the visit to study and compare in detail the regions identified by the different analyses. This will be done by Beatriz Gutierrez from her home place of work in León (Spain) who will keep in touch with Dr Wiener through email and videoconferences to discuss the results.

Future collaboration with host institution

Since her postdoctoral stay at Roslin between 2004 and 2007, Beatriz Gutiérrez has been in contact with Dr. Pam Wiener who supervised her work during that period, mainly related to the preparation of manuscripts and meeting abstracts describing results that were generated at the end of Beatriz´s stay (Gutiérrez-Gil et al., 2009b; 2010; Wiener & Gutiérrez-Gil, 2009, Wiener et al., 2011; Gutiérrez-Gil et al., 2012). In the short term, the two researchers will continue the collaboration initiated through this short visit with the aim of discussing and interpreting the results and writing a manuscript for their publication. The analyses performed in this visit are of interest to both researchers and their corresponding research groups and they hope to extend these analyses in a future collaborative project, possibly focused on a multi-species (sheep, cattle, goats) genetic analysis of dairy traits.

Projected publications/articles

We expect to publish the results of these analyses regarding selection sweep mapping in dairy sheep in a full research article in an SCI journal. We have not determined the target journal at this point.

References

Gutiérrez-Gil B, et al. 2008. Quantitative trait loci underlying udder morphology traits in dairy sheep. J Dairy Sci. 91: 3672-3781.

Gutiérrez-Gil B, et al. 2009a. Quantitative trait loci underlying milk production traits in sheep. Anim Genet. 40: 423-434.

Gutiérrez-Gil B, et al. 2009b. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. J Anim Sci. 87: 24-36.

Gutiérrez-Gil B, et al. Anim Genet. In press doi: 10.1111/j.1365-2052.2012.02322.x. [Epub ahead of print].

Rubin CJ, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464: 587-591.

Wiener P, et al.. 2011. Information content in genome-wide scans: concordance between patterns of genetic differentiation and linkage mapping associations. BMC Genomics 26;12: 65.

Wiener P, Gutiérrez-Gil B. 2009. Assessment of selection mapping near the myostatin gene (GDF-8) in cattle. Anim Genet.40: 598-608.

Wiener P, Pong-Wong R. 2011. A regression-based approach to selection mapping. J Hered. 102: 294-305.

Project N. 6

Identifying signatures of recent selection through the comparison of indicine and taurine cattle populations of beef and dairy types

Yuri Tani Utsunomiya. Universidade Estadual Paulista - UNESP/Brazil hosted by Prof Johann Sölkner, Universität für Bodenkultur Wien - BOKU/Austria.

Purpose of the visit

1) Perform genome-wide scans for footprints of recent positive selection in four cattle breeds (Angus, Brown Swiss, Gir and Nellore) using high-density SNP data, with special focus on breed-specific signatures.

2) Benefit the student's academic formation by exposing him to state of the art analyses and different working environment.

3) Contribute to the scientific community with more knowledge in the field of signatures of natural and artificial selection that may be related to milk and meat production or taurine and indicine fitness.

Description of the work

Overview Motivated to identify footprints of recent positive selection in cattle, I sought a strategy for integrating multiple tests in genome-wide SNP data. I applied a straightforward frequentist meta-analysis approach for combining P-values across tests, targeting common, moderate frequency variants. I covered two between and two within population tests for selection sweeps, divided into three different categories: extended haplotype homozygosity, change in the allele frequency spectrum and local heterozygosity depression. I also implemented strategies for assigning relevant SNPs to genes, allowing for exploration of the biological meaning of the findings and facilitating hypothesis generation. Additionally, I performed discovery of the ancestral Bovinae allele state of over 440,000 SNPs. Samples description and quality control Genotypes for Illumina® BovineHD Genotyping BeadChip assay of Angus (ANG), Brown Swiss (BSW), Gir (GIR) and Nellore (NEL) individuals were available for prospection of selection sweeps. Details on sample size and source for each breed can be found in Table 1. Only autosome markers were included into the analyses. SNPs were removed from the dataset if they did not exhibit: (1) minor allele frequency (MAF) greater than or equal to 0.03, (2) P-value for Hardy-Weinberg Equilibrium (HWE) greater than or equal to 1 x 10-6 or (3) Call rate (CRSNP) greater than or equal to 90%. After the SNP QC, individuals exhibiting call rate (CRIND) below 90% were also removed. This procedure was performed for each breed in parallel using PLINK (Purcell et al., 2007). In order to mitigate relatedness

in the dataset, individuals were further investigated for proportions of alleles shared identically by descent using PLINK. I developed an algorithm for conservatively remove samples from potential parent-offspring, half-siblings and duplicate pairs. SNPs commonly passing QC in all four breeds were then overlapped. For each breed, a minor imputation procedure was performed with the fastPHASE software (Scheet & Stephens, 2006) to solve for remaining missing data.

Ancestral allele discovery

For the ancestral allele discovery, there were available 2 Gaur (Bos gaurus), 6 Water Buffalo (Bubalus bubalis) and 2 Yak (Bos grunniens) samples typed for the same assay. Genotypes for the three Bovinae species were pooled into a single dataset. I looked for markers with a CRSNP of 100% and with a single allele present (100% AA or 100% BB). For each case, the allele was determined as ancestral. The final SNP set was defined as markers passing QC with ancestral allele information present.

Long-range haplotype based methods

The set of methodologies cited here were calculated using the rehh package in R (Gautier & Vitalis, 2012), with minor modifications to the source code. As base for the actual tests, I calculated the integrated Extended Haplotype Homozygosity for the ancestral allele (iHHA), derived allele (iHHD) and SNP site (iES) for each marker. For the within breed test, I calculated the Integrated Haplotype Score (iHS) based on iHHA and iHHD, as described by Voight et al (2006). The scores were divided into 20 equally sized bins according to their derived allele frequencies, and then standardized to have mean 0 and variance 1. As both tails from the distribution are of interest, I derived two-sided P-values as $1-2|\varphi(iHS)-0.5|$ from the Gaussian cumulative density function. For the between breeds tests, I computed the pairwise Rsb (Tang et al., 2007) from iES. Although the statistic is also standardized, the procedure recommended by the authors does not divide scores into bins and uses the median instead of the mean. As each comparison was performed twice, just shifting the direction of selection (Breed A x Breed B, with positive values representing selection towards Breed A; and Breed B x Breed A, with positive values expressing selection towards Breed B), I derived one-sided upper tail P-values from the normal cumulative density function.

Change in the allele frequency spectrum based method

Grossman et al (2010) describe a very simple method computed as the difference in the derived allele frequency between populations, called ΔDAF. Values range from -1 to 1 and are also normally distributed. I standardized ΔDAF scores using the distribution mean and standard deviation, and retrieved one-sided upper tail P-values.

Local heterozygosity depression based method

Rubin et al (2010) defined and applied a Z-score test for local heterozygosity depression (ZHp) on whole genome sequence data of domestic chicken, which basically expresses how much the expected heterozygosity in chromosome windows deviate from the average genome heterozygosity. I adapted the approach to every SNP site and computed the observed instead of the expected heterozygosities. Values obtained resembled a normal distribution, and were standardized to produce mean 0 and variance 1. This time, negative values are of interest and the resulting site heterozygosity scores are multipled by -1 in order to switch their direction, yielding a new statistic called SHp. I derived one-sided upper tail P-values for each score obtained.

Combination of multiple tests

As all statistics approached have P-values retrieved from normal distributions with same parameters (mean 0 and variance 1), I adapted the weighted version of Stouffer method for meta-analysis of Z-transformed P-values (as reviewed by Whitlock, 2005). For each population, for each marker and each test i, the respective P-value is transformed into a Z-score by Zi = -φ-1(1-pi). Within population tests are performed only once per breed, hence their respective weight ωi is set to 1. For each comparison of between population tests, the Z-score is weighted to 1/n, where n is the number of comparisons. Then, the combined statistics of k tests, for each SNP in each breed, is defined as:

$$meta\text{-}SS = \frac{\sum_{i=1}^{k} \omega_i Z_i}{\sqrt[2]{\sum_{i=1}^{k} \omega_i{}^2}}$$

The meta-SS (stands for Meta-analysis of Selection Signals) scores are referred back to the standard normal distribution to obtain combined significance values, which are intended to address either the combination of information among different, independent tests can reject the shared null hypothesis (neutral marker) or not. Significance level for genome-wide meta-SS P-values was based on a Bonferroni threshold (0.05/nSNP).

Functional annotation

For any peak crossing the significance line, I applied three different strategies for the annotation of functional features. The first approach consists on checking if any significant SNP is intragenic via mining the Ensembl Variation 67 database with the Ensembl Biomart tool (Kinsella et al., 2011). The second strategy comprehends isolating the most significant SNP from each visually identifiable peak and mapping the closest gene to it. For that matter, I downloaded the Bovine UMD3.1 gene set from Ensembl Genes 67 database via Biomart tool and used the ClosestBed algorithm from the BedTools software (Quinlan & Hall, 2010). The third approach is a LD-based window scheme, divided into three steps.

First, every SNP crossing the significance line is defined as a 'core SNP'. Second, I walk down to proximal and distal chromosome positions calculating correlations between the core SNP and the neighbor markers, checking if they tag the core SNP or not based on r2 values. The r2 threshold adopted to declare that one marker tags the core SNP is set to 0.7. The positions of the last tag markers on both sides of the core SNP, i.e., positions from where r2 decays below the defined threshold or the tag marker distance from the core position exceeds 1 Mb, are set as the boundaries of a window. This analysis was done in PLINK. Third, this window is interpreted as a single locus, and any gene overlapping it is considered to be in LD with the core SNP and is therefore annotated. For core SNPs which no window boundaries could be determined, I included the closest gene in the vicinity to the list. As some windows may also overlap, the gene list yielded is then parsed to exclude repeated gene names and then processed in DAVID (Huang et. al, 2009-A; Huang et. al, 2009-B) for annotation of functional terms. I used the default parameters for each breed gene list, pooling together all genes annotated across the genome to reveal over-represented functional terms. Finally, I used the Enrichment Map Cytoscape plugin (Merico et al., 2010) to build networks of inter-related enriched terms based on the number of overlapping genes.

Main results obtained

Ancestral allele discovery

Assessing the outgroup species genotypes, I observed an average CRIND of 83.79%, 96.93%, 94.87% and 88.63% for Water Buffalo, Yak, Gaur and pooled data, respectively. Considering only markers perfectly typed across the pooled outgroup

samples (CRSNP = 100%), I observed a total of 559,663 (71.94%) SNPs successfully genotyped, from which 111,376 were polymorphic (MAF > 0). Hence, a total of 448,287 SNPs (56.75%) had their ancestral allele determined.

Quality control

From the initial set of 742,910 autosome markers, numbers of SNPs passing QC were 579,470, 554,826, 485,655 and 461,702 for ANG, BSW, GIR and NEL, respectively. Overlapping of the four SNP lists retrieved a final set of 281,994 markers, from which 157,702 had ancestral allele information available. Even with the drastic drop in number of SNPs, the intermarker distance mean and median were 15.94 kb and 6.43 kb, respectively, superposing the median spacing of 37 kb declared for the BovineSNP5O assay (Matukumalli et al., 2009). The number of remaining samples for each breed after duplicates and first degree relationships removal were: 24 ANG, 44 BSW, 23 GIR and 581 NEL. As NEL exhibits a sample size much larger than the other breeds, 45 individuals were sampled from the total, in order to do fair comparisons.

Identification of selection signals and overview of functional annotation

All tests for footprints of selection performed resembled normal distributions and genome-wide Z-transformed P-values were weakly correlated, satisfying the independence condition for meta-analysis. After combining P-values, the number of SNPs crossing the genome-wide significance ($P < 3.17 \times 10\text{-}7$) was: 153 for ANG, 212 for BSW, 3 for GIR and 13 for NEL. The most significant SNP was found in BSW ($P = 3.82 \times 10\text{-}12$), and is an intronic variation in Cornichon homolog 3 gene (CNIH3), located at BTA16:28478192.

In order to illustrate the potentiality of combining signals from different methodologies, P-values for each one of the individual tests for the CNIH3 region in BSW (candidate for being selected) and NEL (candidate for being neutral) are displayed in detail in Figure 2. EHH decay plots and a bifurcation diagram for the haplotypes containing the derived allele are also provided. It can be seen from BSW and NEL comparison that the composite test penalizes SNPs with little statistical support. The signal of the unusual derived allele long haplotype in BSW, revealed by the meta-SS statistic, is not detectable in NEL. The number of genes directly harboring significant SNPs was ANG 20, BSW 27, GIR 1 and NEL 3. I found 2 synonymous exonic SNPs for ANG and BSW, 1 non-synonymous variation (Ala->Thr) for a gene of the olfactory receptor family (LOC524290 - OR2W3) in ANG ($P = 7.65 \times 10\text{-}9$) and a 3'UTR variation for the KIF5C (kinesin family member 5C) gene in NEL ($P = 2.68 \times 10\text{-}7$). All other variants within genes were located in introns. Total number of genes within LD-windows included in each breed specific list was: 309 ANG, 177 BSW, 4 GIR and 14 NEL. Network of enriched terms from ANG gene list revealed three groups: (1) immune response related genes, involved with chemokine and cytokine activity; (2) transcription activity, comprehending the biosynthesis of ribonucleoproteins, transcription activation and aminoacylation of tRNA with L-histidine residuals; and (3) glucolysis and gluconeogenesis pathways. For BSW, a cluster related to post-transcriptional modifications of RNAs (mostly methylation of adenosine residuals) and another involved with Calpain were observed. A significant intronic SNP (BTA16:27801014, $P = 2.61 \times 10\text{-}7$) was detected in the Calpain 2 (m-Calpain) catalytic subunit, which may be capturing the signal of a causal untyped variant under selection. Due to a low number of genes mapped, the clustering of enriched terms for GIR and NEL retrieved no significant result. Across all lists, a total of 69 genes (13.69%) had no functional

term associated to it, being either uncharacterized proteins or novel RNAs with no functional record available.

Future collaboration with host institution

Both host and guest institutions are already collaborating in the Zebu Genome Consortium, an international initiative to characterize indicine cattle genetic resources and develop adequate strategies for using genomic information in selection schemes, mainly in Nellore and Gir breeds. This exchange visit in particular allowed for stepping up current joint activities between my advisor Prof. José Fernando Garcia and the host researcher, Prof. Johann Sölkner. After this 3 months collaboration, Prof. Sölkner has kindly agreed to become my co-advisor in UNESP-Brazil.

Projected publications/articles

The results herein described have been intensively discussed with all parties, and a draft-paper was already composed. We plan to submit this work to a medium impact journal within the next months, under the provisory title of Detecting loci under recent positive selection in dairy and beef cattle by integrating different genome-wide scan methods.

References

Gautier M, Vitalis R. rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics Advance Access. 2012.

Grossman SR et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. 327, 883-886 (2010).

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

Huang DW, Sherman BT, Lempicki RA (B). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1-13

Kinsella RJ, et al.. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford). Vol. 2011 Published online Jul 23, 2011.

Matukumalli LK, et al. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One.2009;4(4):e5350.

Merico D, et al. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation PLoS One. 2010 Nov 15;5(11):e13984.

Purcell S, et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Rubin CJ et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 464, 587-593 (2010).

Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006; 78:629–644.

Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biology. 5, e171 (2007).

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 4, e72 (2006).

Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J. EVO L. BI OL. 18 (2005) 1368–1373.

Funding

ESF Research Networking Programmes are principally funded by the Foundation's Member Organisations on an *à la carte* basis. **GENOMIC-RESOURCES** is supported by:

- Fonds zur Förderung der wissenschaftlichen Forschung (FWF), FWF Austrian Science Fund, Austria
- Fonds National de la Recherche Scientifique (FNRS), Belgium
- Fonds voor Wetenschappelijk Onderzoek - Vlaanderen (FWO), The Research Foundation - Flanders, Belgium
- Nacionalna zaklada za znanost, visoko školstvo i tehnologijski razvoj Republike Hrvatske, National Science Foundation for Science, Higher Education and Technological Development, Republic of Croatia
- Suomen Akatemia, Biotieteiden ja ympäristön tutkimuksen toimikunta, Academy of Finland, Research Council for Biosciences and Environment, Finland
- Deutsche Forschungsgemeinschaft (DFG),German Research Foundation, Germany
- Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), The Netherlands Organisation for Scientific Research,The Netherlands
- Norges Forskningsråd,The Research Council of Norway, Norway
- Forskningsrådet för miljö, areella näringar och samhällsbyggande, Swedish Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS),Sweden
- Schweizerischer Nationalfonds (SNF), Swiss National Science Foundation, Switzerland
- Biotechnology and Biological Sciences Research Council (BBSRC), United Kingdom
- Institut National de la Recherche Agronomique (INRA), France

GENOMIC-RESOURCES Steering Committee

Dr Stéphane Joost (Chair) Ecole Polytechnique Fédérale de Lausanne, Switzerland – Dr Göran Andersson, University of Uppsala, Sweden – Prof Philippe Baret Université Catholique de Louvain, Belgium – Prof Michael W. Bruford, Cardiff University, United Kingdom – Prof Nadine Buys, Katholieke Universiteit Leuven, Belgium – Prof Ino Curik, University of Zagreb,Croatia – Dr Juha Kantanen MTT Agrifood Research Finland, Finland – Dr Johannes A. Lenstra, University of Utrecht, The Netherlands – Prof Theo Meuwissen, Norwegian University of Life Sciences, Norway – Prof Jutta Roosen, Technische Universität München, Germany – Prof Johann Sölkner, University of Natural Resources and Applied Life Sciences, Austria – Dr Michèle Tixier-Boichard, INRA/AgroParisTech, France

Advisory Expert: Prof Paolo Ajmone Marsan, Università Cattolica del Sacro Cuore, Italy

Project coordination : Elena Murelli

ESF Liaison : Dr.KirstenSteinhausen, Science | leeps-rnp@esf.org, Administration

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

1 quai Lezay-Marnésia • BP 90015
67080 Strasbourg cedex • France
Tel: +33 (0)3 88 76 71 00 • Fax: +33 (0)3 88 37 05 32
www.esf.org