PROJECT FINAL REPORT

Grant Agreement number: 244356

Project acronym: NEXTGEN

Project title: Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options

Funding Scheme: CP

Period covered: from 01/04/2010 to 31/03/2014

Name of the scientific representative of the project's co-ordinator[1], Title and Organisation: Dr Pierre TABERLET, CNRS

Tel: + 33 (0)476 63 54 33

Fax: +33 (0)476 51 42 79

E-mail: pierre.taberlet@ujf-grenoble.fr

Project website[2] address: http://nextgen.epfl.ch/

Table of contents

1. Final publishable summary report

1.1 Executive Summary

We are currently experiencing a dramatic loss of farm animal. A significant number of breeds already disappeared in developed countries, and many are presently endangered. The same process is now progressively taking place in Africa and Asia. Based on whole genome data, the NEXTGEN global objective consists to develop cost-effective optimized methodologies for preserving farm-animal biodiversity, using cattle, sheep, and goats as model species.

More specifically, NEXTGEN will:

- develop innovative bio-banking methods based on freeze-dried nuclei;

- produce whole genome data in selected populations;

- develop the necessary bioinformatics approaches; focusing on the identification of genomic regions under recent selection (adaptive / neutral variation);

- provide guidelines for studying disease resistance and genome/environment relationships in a spatial context;

- assess the value of wild ancestors and breeds from domestication centers as genetic resources.

The tissue sampling for the genetic analyses has been carried out based on a grid system covering the whole country for Uganda (cattle) and for Morocco (sheep and goats). Such an innovative sampling approach opens new perspectives at the data analysis stage, as many different hypotheses can be tested using the same dataset.

A total of 447 whole genomes of sheep (205), goats (208), and cattle (34) have been re-sequenced, with coverage of at least 10x. Additionally, a total of 1009 animals, mainly cattle from Uganda, have been genotyped with DNA chips.

The main results are:

- new bio-banking perspectives based on freeze-dried nuclei;

- a collection of computer programs specifically developed for preserving farm animal biodiversity, either by optimizing the selection of individuals for breeding and biobanking (SELCAPRE program), or by identifying genes/environment relationships (Samβada program);

- precise description of the geographical molecular diversity of cattle from Uganda, with the identification of selection signatures associated to disease tolerance/susceptibility;

- selection signatures associated to environmental parameters across Morocco for sheep and goats;

- precise assessment of the potential of different populations as genetic resources, including industrial breeds, traditional breeds within or outside of the domestication centers, and wild ancestors;

- general guideline for preserving farm animal biodiversity.

1.2 Summary description of project context and objectives

1.2.1 Farm animal genetic resources (FAnGR) are being lost at an unprecedented rate

There is a growing awareness that threats to biodiversity are increasing, whether measured in terms of extinction rate, destruction of ecosystems and habitat, or loss of genetic diversity within the species utilized in agriculture.

During the last century, the European livestock sector has undergone striking changes as large-scale production expanded. The formulation of the modern breed concept during mid-1800s (Porter 2002) and its application to breeding and husbandry practices led to the formation of well-defined breeds, exposed to intense anthropogenic selection. The progress of livestock management practices, the introduction of artificial insemination and embryo transfer, the improvements in feed technology and the use of vaccines and therapeutics against endemic diseases have fostered the diffusion of industrial breeding. This has led farmers to progressively substitute the less productive, locally adapted, autochthonous breeds with highly productive cosmopolitan breeds and to progressively abandon agriculture in marginal areas (Taberlet et al. 2008). Therefore a significant number of cattle, sheep, and goat breeds already disappeared and many are presently endangered (FAO 2007). The same process is now progressively taking place in Africa and Asia.

Considering the lack of information and the unprecedented rate of extinction, the clear possibility exists that a high number of breeds are being, and will be lost in the near future, before their characteristics can be studied and their potential evaluated. This is particularly worrying in the present scenario because of uncertainties due to rapid climate change, increasing and differentiating market demand and human demographic expansion (FAO 2008). In these conditions it is more strategically important than ever to preserve as much the farm animal diversity as possible, to ensure a prompt and proper response to the needs of future generations. Sustainable management and conservation of FAnGR requires a comprehensive knowledge of breeds, including population size, geographic distribution, production performance, other functional characteristics and, most of all, on the accurate assessment and management of the within- and between-breed genetic diversity.

In the NEXTGEN project, large SNP (Single Nucleotide Polymorphism) panels and high-throughput sequencing will be used to assess livestock neutral and functional genetic diversity with levels of precision never previously achieved and to develop a core molecular dataset which will provide a long-term resource for developing methods for effective conservation of livestock biodiversity.

1.2.2 NEXTGEN objectives

In this context, NEXTGEN proposes the bold step of using whole genome data to develop and optimize conservation genetic management of livestock diversity for the foreseeable future. The rationale for choosing whole genome data is to "future-proof" DNA-based analysis in livestock conservation against the recent changes in technology and analysis. Thus, in the context of whole genome data availability, our global objective is to develop cost-effective optimized methodologies for preserving farm-animal biodiversity, using cattle, sheep, and goats as model species.

More specifically, NEXTGEN will:

- **produce whole genome data in selected populations of cattle, sheep, and goats:** comparing and contrasting industrial breeds from Europe, local breeds from Europe, Africa, and Middle East, and from sheep and goat wild ancestors. *These data will be obtained from species and populations experiencing different intensities of agriculture, different landscapes and very different climates in terms of rainfall and temperature*;

- **transfer, adjust and enhance the bioinformatics methodologies and infrastructure**: transferring methods developed within the 1000 human genome project to farm animals (cattle, sheep, goats). *This approach will enable efficient analysis and data mining from large-scale whole genome population projects, and will assist those interested in studying whole genome diversity of domestic populations to utilise appropriate tools*;

- **develop tailored methodologies for comparative genome analysis in cattle, sheep, and goats**: this will include analysis of nucleotide diversity and detecting signatures of selection along the genome (to distinguish neutral versus adaptive variation). *These methodologies will be easily transferred to study local adaptation in any kind of organisms, providing useful tools is an area of growing interest for the scientific community.*

- **develop genomic methods for the identification and mating of animals to optimize selection response and maintenance of genetic variability**: methods will consider genetic gain and contribution of founders and will be developed for both pure and crossbreeding programs. *This approach will greatly improve the maintenance of genetic diversity in traditional selection programs.*

- **to develop approaches, based on whole-genome data, for the selection of animals for bio-banking**: methods will consider various sources of information, including genomic, phenotypic, pedigree and geographic data and their combination. *This innovative approach will optimize the selection of individuals for bio-banking,*

- **develop new bio-banking technologies**: a freeze-drying technology will be adapted to cells and female gametes (oocytes). *This low cost methodology will greatly simplify the establishment and maintenance of gene banks, which immortalise cells lines/gametes from rare breeds, compared to current resource intensive cryo-conservation methods using liquid nitrogen.*

- **provide recommended methodologies for preserving farm animal biodiversity integrating new genome data:** by comparative analysis of different conservation strategies. *This approach will lead to the development of new policies enabling local socio-economic constraints to be incorporated.*

- **explore new strategies to identify disease resistance genes**: integrating and comparing information on the geographic distribution of selective sweeps and the prevalence of target diseases. *This approach will lead to the development of new strategies for detecting genomic regions and genes controlling traits, which are very difficult or very expensive to identify with other experimental approaches*

- **design and validate a methodology for studying genome/environment relationships**: by sampling sheep and goats using a grid system over an area of traditional breeding (relatively undisturbed by the recent spread of industrial breeds) across contrasting environments, by producing whole-genome data for these samples, and by analyzing the results within a GIScience context. *This sampling strategy will open new avenues at the data analysis stage.*

- **assess the potential of breeds (cattle, sheep, goats) from domestication centres as genetic resources:** by comparative analysis of genomic diversity of local breeds from the original centres of domestication with local breeds in Europe, Uganda, and Morocco, and with

industrial breeds in Europe. *This analysis will clarify the conservation priority that should be given for breed from these 'cradles of agriculture'.*

- establish the relevance of wild ancestor species as genomic resources: by comparative analysis of genomic diversity in centres of domestication between local breeds and wild populations. *This will establish the genome-level changes that accompanied domestication and will characterise additional variation present in the wild relatives, which is potentially amenable for future exploitation.*

- assess the performance of a surrogate genome data source compared with whole genome sequence data for assessing biodiversity: by comparing the results from unbiased SNP panels with whole genome data for their ability to estimate coalescence times and signatures of selection in a defined set of breeds. *This will establish whether it is viable to use a 'surrogate set of SNPs to accurately approximate whole genome processes, an approach which could simplify the process of molecular biodiversity assessment.*

- **carry out high quality training for developing research capabilities in ICPC, ACP, and European countries in farm animal conservation genomics**: by organizing several training workshops in ICPC and ACP countries, by promoting cooperative PhD programs involving ICPC and European countries, by encouraging staff exchange among partners. *This strategy is designed to maximize the capacity-building component of NEXTGEN.*

- **implement efficient dissemination of improved methodologies**: via translational activities towards non-specialists (industry, breeders, stakeholders). *This strategy is designed to maximize the NEXTGEN impact on end-users.*

1.3 Description of the main S&T results/foregrounds

1.3.1 Preliminary considerations

In order to optimize the large-scale sequencing within the NEXTGEN project, we proposed to subcontract the sequencing part to the Genoscope (French National Sequencing Center). We initiated this process three months before the expected start of the sequencing, but it took a total of 12 months for obtaining the green light from the Commission for this subcontracting. Thus, the sequencing started 9 months later than expected. In order to have enough time to properly analyze the huge dataset produced (that is larger than the dataset produced during the 1000 human genome project), we requested a six months extension. Unfortunately, this extension was rejected. As a consequence, the results presented here correspond to the analyses that have been completed at the end of month 48. The different partners involved in the NEXTGEN project will continue to analyze the data after the official end of the project, and will forward to the Commission the scientific papers that will result from the analyses carried out after the official end of the NEXTGEN project.

1.3.2 The different bio-informatic tools developed or used for the NEXTGEN project

The NEXTGEN project aims to estimate intra-specific biodiversity of three farm animals: sheep (*Ovis aries*), goat (*Capra hircus*) and cow (*Bos taurus*) using high throughput molecular techniques. Among them, next generation sequencing was used to sequence more than 400 individuals belonging the three cited species. This leads to the production of approximately 30 Tera bytes of raw data. The analysis of such an amount of data requires selecting and developing a set of efficient software.

1.3.3 Processing of the raw sequences

Raw sequence storage

The European Bioinformatic Institute (EBI) at Hixton - UK take the responsibility of raw sequence storage through the Sequence Read Archive (SRA, Leinonen *et al*. 2011) division of the European Nucleotide Archive (ENA). This solution ensures a permanent and reliable storage of the NEXTGEN raw data. Moreover SRA-ENA provides an efficient public access to the NEXTGEN raw data set for the research community.

De novo assembly of the wild sheep (*Ovis orientalis*) and wild goat (*Capra aegagrus*) genomes

The NEXTGEN project provides genomic data for the domestic sheep (*Ovis aries*), goats (*Capra hircus*) and cows (*Bos taurus*) but also for the wild species *Ovis orientalis* and *Capra aegagrus* domesticated about 10,500 years ago and that can be considered respectively as the wild ancestor of sheep and goats. To check the potential deep differences between the genome structure of the domesticated animal and of their corresponding wild species, the NEXTGEN consortium realized a deep shotgun sequencing of the *Ovis orientalis* and *Capra aegagrus* genomes (> 100x sequencing depth). The *de novo* assembling of these genomes was achieved on the EBI computational facilities using the Cortex and AllPaths (Gnerre *et al.* 2011) programs.

Mapping of the resequenced individuals on a reference genome

Most of the genome sequences were produced with an average sequencing depth of 12x, which is enough to infer individual genotype with a good accuracy but not sufficient to allow de novo assembling of the genome sequence. Each of the read set produced for each analyzed individuals following this strategy were mapped against their corresponding reference genome using the Burrows-Wheeler Aligner (BWA, Li and Durbin, 2009).

De novo assembly of the mitochondrial genomes

Even if a 12x sequencing depth is not enough for allowing *de novo* assembling of a nuclear genomes it provides a higher sequencing coverage for the mitochondrial genome. Classical assemblers most of the time failed to assemble the mitochondrial genome, despite a good sequencing depth, as the heuristics they implement is not appropriate for assembling the mitochondrial genome. To circumvent this limitation, an Organelle Assembler has been developed by CO01. The assembler software is currently a prototype and will be distributed as an open source software in the next few months. Despite its current status, the organelle assembler allowed *de novo* assembling of the mitochondrial genomes including information about copy variation number (CNV) at the D-loop locus.

Data manipulations

All the genome alignments were stored following the Binary Alignment Map format (BAM, Li *et al.* 2009). The list of all variants associated to each individual is stored following the Variant Call Format (VCF, Danecek *et al.* 2011). Consequently all the data manipulations are done using the Samtools (Li *et al.* 2009). Samtools were used directly as a package of unix programs or as a library bound to *ad-hoc* program directly implemented in C or through binding with high level languages like R (R Development Core Team, 2005) or Python.

Variant calling

The variant calling corresponding to single nucleotide polymorphisms (SNP) and small indels were achieved using the Samtools caller (Li *et al.* 2009) and GATK (McKenna *et al.* 2010).

1.3.4 Dissemination of the produced genome sequences

The complete processed dataset produced by the NEXTGEN consortium including the genome alignment of all the individuals of sheep, goats and cows have been integrated in the ENSEMBL database (Flicek *et al.* 2014, for the last release). This database is available through a web interface and will provide a user-friendly public access to the full dataset produced during the NEXTGEN project.

1.3.5 Detection of selection

Two families of approaches were used during the NEXTGEN project to detect the impact of the selection pressure on the sheep, goat or cow genomes. First, model based approaches are used to detect selection when a population structure is considered *a priori*. When we want to avoid the *a priori* division of the individuals into population, the selection was detected using correlative approaches.

Model-based approaches

These methods were applied to detect genomic signatures of selection that differentiate wild and domestic populations. In this context, several software were used to estimate population parameters that can influence the selection detection. Among them: the Pairwise Sequential Markovial Chain method (PSMC, Li and Durbin 2011) were used to estimate the demographic history of these species and the Bayesian Analysis of Population Structure (BAPS, Corander and Martinen 2006) were used to detect groups of individuals suitable for further analyses. A pipeline was developed by the NEXTGEN consortium to automatize the application of this method to each of the considered samples.

Correlative approaches

These methods are based on the correlation of allele presence and environmental variables. They are usable to detect genes potentially selected by a putative environmental factor. The first implementation of the environmental correlation approach is the spatial analysis method (SAM, Joost *et al.* 2007). To be usable on large dataset as those produced by the NEXTGEN project an efficient implementation of this method has been developed. The resulting program Samßada is open source and can be freely download (http://lasig.epfl.ch/sambada). The manuscript submitted to *Bioinformatics* can be downloaded through arXiv (http://arxiv.org/abs/1405.7658).

1.3.6 Selection of individuals based on whole genome data

Breeding simulations

Two software developed by Partner P04 and available to the NEXTGEN consortium members were used to test different breeding strategy. The first one, SelPicPop is an individual centered simulator of breeding plans to control inbreeding while enhancing productivity. The second one, is a web graphical interface named Selcapre. It allows a convenient usage of SelPicPop through a web browser and is dedicated to simulate goat breeding scheme.

Selection of individuals for bio-banking purposes

Stochastic and deterministic simulations were developed to estimate the amount of genetic material to be cryopreserved for reconstructing a population of 25 females and 25 males of reproductive age, corresponding to an effective population size of 50, accounting on information on rare alleles and on population structure (kinship). To achieve this aim, a simple algorithm was developed to select individuals carrying rare alleles and preserve the gene pool at specific loci. The algorithm steps are: (i) arrange genotyped individuals in an ordered list; (ii) starting with the first individual, compare genotype of first locus with genotypes of all others in the list; (iii) individuals with unique alleles are kept, those without unique alleles are discarded; (iv) repeat for all loci and all individuals.

1.3.7 Tissue sampling of sheep, goats, and cattle in Morocco, Iran, and Uganda

This part describes the approach used to sample sheep, goats and cattle in Morocco, Iran and Uganda and the number of individuals that have been available for sequencing/genotyping. Depending the country and the objective of the study, a specific protocol was adopted to collect samples.

1.3.7.1 Sampling in Morocco

In Morocco, samples of sheep and goats have been collected with the aim of studying local adaptation to different environments. They have been collected across a wide part of Morocco to cover a range of highly contrasted environments (~400 km^2; Northern part of Morocco with latitude between 28° and 36°; Figure 1). For this purpose, a sampling grid consisting of 198 cells of 0.5° of longitude and latitude was established. The goal was to cover a number of 160 of these cells. In each cell, a maximum of 3 unrelated individuals have been sampled by flock in 3 different flocks. Tissue samples, geo-coordinates and phenotypic traits of each individual have been collected.

1.3.7.2 Sampling in Iran

In Iran, the aim of the study was mainly to assess genetic resources in small ruminants in domestication center (north-west Iran). The protocol was based on collecting samples from ~~30~~ *O. orientalis*, ~~30~~ *C. aegagrus*, ~~60~~ sheep and ~~60~~ goats from different breeds existing in that area. Sampling of wild animals has been mainly done from hunted animals and from the stored tissues in Iranian Conservation Centers. Sampling of local breeds has been done in remote areas, in order to avoid a possible introgression from industrial breeds, as observed in areas of more intensive agriculture. As in Morocco, sampling was consisting of collecting tissue samples, geo-coordinates and phenotypic traits if relevant (for domestics).

1.3.7.3 Sampling in Uganda

In Uganda, the aim was to study the relationships between cattle genome and disease resistance. Thus, the sampling approach was based on taking tissue samples, blood, and serum from local cows from 50 cells over the whole country. A maximum of six herds were sampled per cell with the constraint of avoiding herds that are in close geographic proximity to each other as much as possible. In each sampling site, two pairs of one healthy and one sick animal have been sampled.

1.3.7.4 Standardized protocol

In the 3 countries, the sampled animals have been identified using the NEXTGEN animal ID code. This included the country, the species, the cell on the grid and the number of the individual of the concerned species within the country. Tissues have been sampled by taking three biopsies on the external and distal part of the ear (~2x2x6 mm per biopsy). The biopsies have been left in alcohol during 1 day and then transferred into silica gel for storage at ambient temperature. Blood samples have been taken for direct examination for haemoparasites. Blood for examination of antibodies have been taken for serology. The serum samples have been left to stand at ambient temperature for 1–2 hours until the clot begins to contract and then stored in a cool box in the field. In the laboratory, the whole blood has been aliquoted and thin smears prepared for examination of haemoparasites. Serum has been

subjected to ELISA testing for exposure to particular aetiologic agents using OIE standardized protocols.

Beside the biological samples, several other information have been collected for each animal or sampling site. Thus, geographical coordinates of the sampling site have been reported. Longitude-latitude/WGS84 was the chosen reference coordinate system and the coordinates have been indicated in decimal degrees. Information like date, sex, age, breed and several morphometric traits of the animal (horn length, body length, …) have been collected. Also, individual pictures of animals have been taken.

1.3.7.5 Number of available animals for sequencing/genotyping

At the end of the project, in Morocco, 1283 goat samples from 162 cells (average of 7.92 goats/cell) and 1412 sheep samples from 164 cells (average of 8.61 individuals/cell) have been collected and stored in the sample bank in Beni Mellal, Morocco (Figure 1).

In Iran, 62 sheep, 65 goats, 19 *O.orientalis*, 25 *C. aegagrus* and 4 *O. vignei* have been sampled. Additionally, tissues from 8 cattle have been collected from the sampling area in this country.

In Uganda, Biological samples in form of three ear nicks per cow, whole blood, and serum have been collected from 906 local cows from 50 cells across the country.

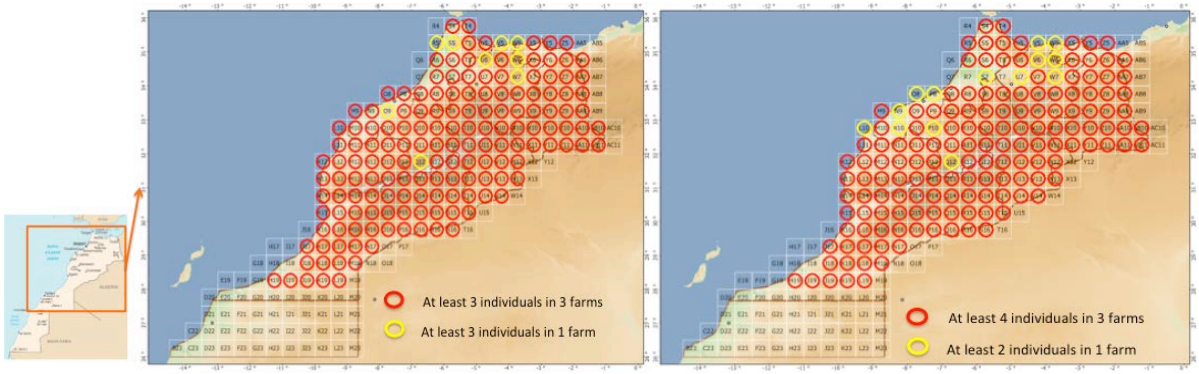


***Figure 1**. Maps of Morocco showing the distribution of sampled goats (right) and sheep (left).*

1.3.8 Biobanking technologies

1.3.8.1 Using freeze-dried cells to replace expensive cryopreservation

Procedures for cryo-storage of spermatozoa, embryos and cell lines are widely used in research, animal breeding and biomedicine. Current methods for cryopreservation are straightforward and efficient, with a 50% to 60% recovery rate after thawing. However, long-term storage is very expensive requiring a continuous supply of liquid nitrogen (Carter 1991). Therefore, alternative solutions capable of at least the same efficiency, but with lower maintenance costs, are attractive. The freeze-drying approach has been tested by Partner P07 who have recently shown that the freeze-dried somatic cells stored at room temperature for 5 years in a cardboard box maintain nuclear viability (Loi *et al.* 2008).

1.3.8.2 Main experiments and results

Assessing the DNA function of dry cells before and after nuclear transplantation of the dry cells into enucleated oocytes

An in depth, multidisciplinary series of studies, ranging from ultrastructure (TEM), immune-fluorescence, molecular biology and experimental embryology essays (nuclear transfer) has provided relevant insights on the effects of dry storage on functional properties of DNA.

The main findings are:

- An extraordinary good preservation of nuclear structure after dry storage;
- A good proportion of cells with intact DNA after re-hydration;
- An unexpected, highly redundant DNA repairing capacity of the oocyte;
- Normalcy of cloned embryos derived from nuclear transfer of dry cells.

Exploiting a new class of Dry-protectant: Late Embryogenesis Abundant (LEA) proteins for inducing dry-tolerance.

Anticipated by an exhaustive study, Partner P07 decided to exploit Late Embryogenesis Abundant (LEA) proteins to induce dry tolerance in somatic cells. The LEA proteins that have been analyzed are originating from the following organisms:

1) *Artemia franciscana*, GenBank FJ592175.1 (Artemia); targets mitochondria

2) *Zea mays*, GenBank NM_001111949.1; binds to membranes

3) *Triticum aestivum*, GenBank L29152.1 (WCOR410); permeates nucleus and cytoplasm.

The gene sequences were inserted into transfection vectors and transfected into primary cultures of sheep fibroblasts. The gene products were later detected at proper sub-cellular localization by epi-fluorescence and confocal microscopy (Figure 2). The LEA transfected cells were de-hydrated at room temperature, and monitored for viability at different time frames. The results are summarized in Figure 3.

Sheep oocytes are very sensitive to dry conditions; only chromosomes retain viability (upon nuclear transplantation upon enucleated, fresh oocytes) while the all structure is irreversibly. Hence, on the basis of the positive outcomes of LEA proteins on cells, Partner P07 have produced all 3 recombinant LEA proteins and assessed eventual function upon injection in sheep oocytes (*in vitro* activation and development to blastocyst stage). The LEA proteins are perfectly tolerated and do not interfere with normal embryo development (Figure 4).

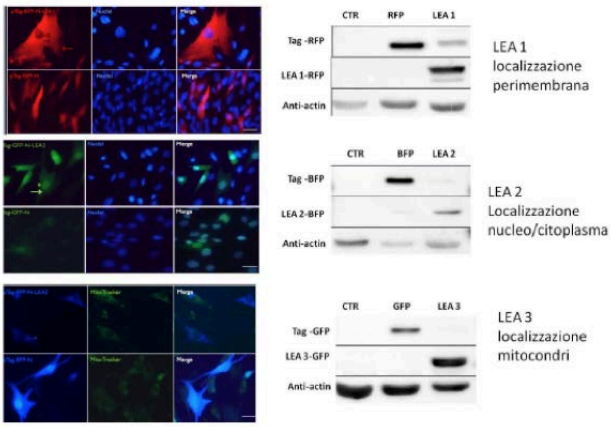


Figure 2. *Localization of the expression of three LEA proteins in ovine somatic cells after transfection. First row: expression of LEA 1: expression (PCR), the proteins is expressed in membranes. Second row: Expression of LEA 2: expression (PCR) and at nuclear/cytoplasmatic level. Third row: Expression of LEA 3: expression (PCR), the protein is localized at mitochondria levels.*

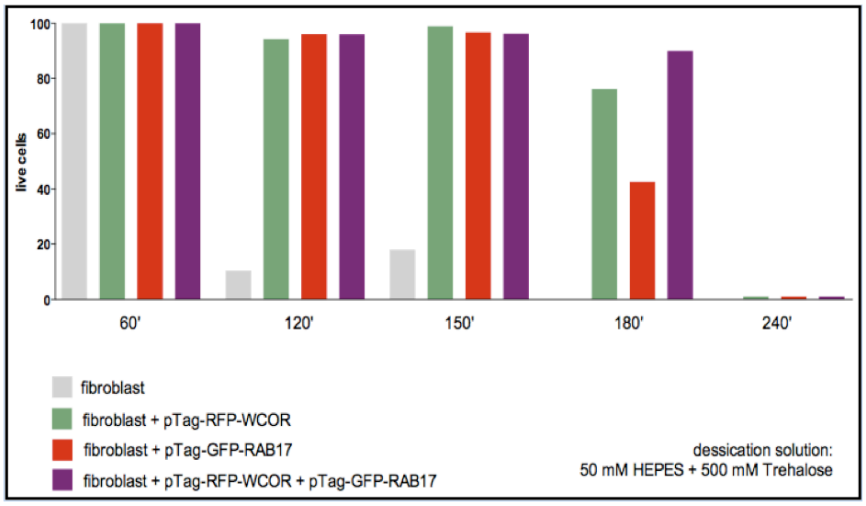
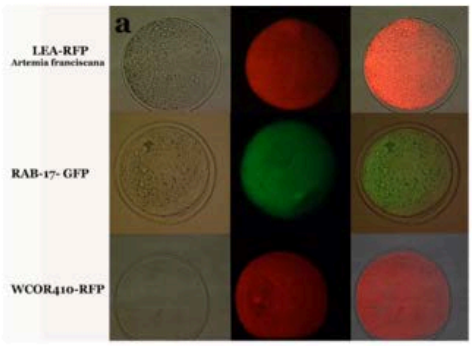
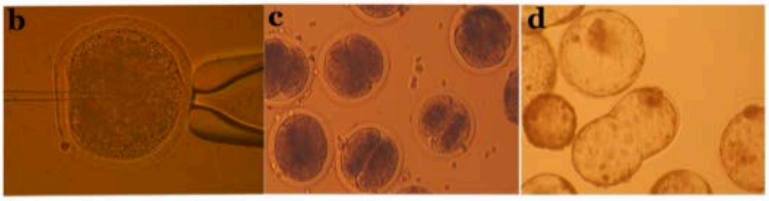


Figure 3. *Dry tolerance in sheep fibroblasts expressing single LEA proteins genes, of two Lea protein genes. The results shown that LEA proteins alone, but better in combination, induce dry tolerance in somatic cells, although for short period of time.*



In vitro parthenogenetic development of oocytes injected with LEAs.

LEAs injection	Oocytes Cultured	Cleavage	Morula-Blastocyst/Cleavage
LEA-RFP artemia franciscana	25	15/25 (60%)	5/15 (33.3%)
RAB-17-GFP	28	22/28 (78.6%)	8/22 (36.4%)
WCOR-410-RFP	31	20/31 (64.5%)	7/20 (35%)



LEA proteins inside oocytes: a) oocytes injected with fluorescent LEAs, LEA-RFP (Artemia franciscana), RAB 17-GFP, WCOR410-RFP; table: In vitro parthenogenetic development of oocytes injected with LEAs; b) oocytes injected with LEAs, c) embryos at 2 cell stage after in vitro activation of injected oocytes; d) blastocysts following in vitro activation of injected oocytes.

Figure 4. *Results of the experiments on oocytes: LEA proteins are perfectly tolerated by the oocytes and they could protect oocytes during the dry processing.*

1.3.9 The genetic data produced within the NEXTGEN project

NEXTGEN produced whole genome sequences (WGS) at 10 X coverage via subcontracting with the Genoscope (French Sequencing Centre - CEA, Evry France) using the Illumina Highseq® technology. An automatic procedure was set up for the transfer of genome data from the Genoscope to the EMBL-EBI Vertebrate Genomics group (Cambridge, UK), which was in charge of data management and genome assembly. The data will be publicly available from September 2014 in the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena).

WGS were produced for indigenous livestock breeds/populations (i.e., Moroccan and Iranian sheep and goats, Iranian and Ugandan cattle) and industrial breeds (Saanen and Alpine goats), as well as for wild relatives (Asian Mouflon – *Ovis orientalis* and *O. vignei*, and bezoars – *Capra aegagrus*). For industrial breeds, complementary data were provided by the Sheep Genome Consortium, -CSIRO (J. Kijas, CSIRO/IGSC), IBBA-CNR project (A. Stella), and complementary samples by French INRA (G. Tosser-Klopp, INRA).

In addition to the WGS produced, genotyping were carried out using SNP Illumina® beadchips (OvineSNP50, caprineSNP50, BovineHD and BovineSNP50). For sheep, goats and wild species the SNP typing was done on a subset of individuals for which we produced WGS. This allowed both quality control of the WGS and assessment of the performance of the SNP beadchip as a surrogate of WGS for characterizing the polymorphism of wild species and local breeds. For cattle, the SNP typing gave the core dataset that was used to study the adaptation in Ugandan cattle (the efficiency of the Illumina beadchip for genotyping African

breeds has already been demonstrated in cattle). All genotypes and WGS produced are detailed in the Table 1 below.

Table 1. *Number of samples genotyped and genotyping method. WGS: Whole Genome Sequence, Illumina® beadchips (OvineSNP50, caprineSNP50, BovineHD and BovineSNP50).*

Sample	# of WGS at 10x (except references at 95x)	# SNP typing and Illumina® BeadChip used
Reference Mouflon	1	
Reference bezoar	1	
Ugandan cattle	26	805 BovineSNP50 + 102 BovineHD
Iranian Cattle	8	
Moroccan sheep	162	30 OvineSNP50
Moroccan goats	163	30 caprineSNP50
Iranian sheep	20	18 OvineSNP50
Iranian goats	20	9 caprineSNP50
Bezoars (*C. aegagrus*)	20	7 caprineSNP50
Asian Mouflon (*O. orientalis*)	16	8 OvineSNP50
Urial (*O. vignei*)	6	
'Industrial' goats	4	
TOTAL	447	

1.3.9.1 How to optimize the selection of individuals for breeding and biobanking?

Balancing selection and conservation is crucial in local livestock species. Genomic information could aid towards this goal when breeding scheme are constructed to optimize its application. Optimum contribution (OC) selection is efficient in controlling inbreeding and maximizes genetic gain. NEXTGEN has applied OC methods to develop a web based decision aid tool for balancing selection and inbreeding rate in population of goats: SELCAPRE (Figure 5).

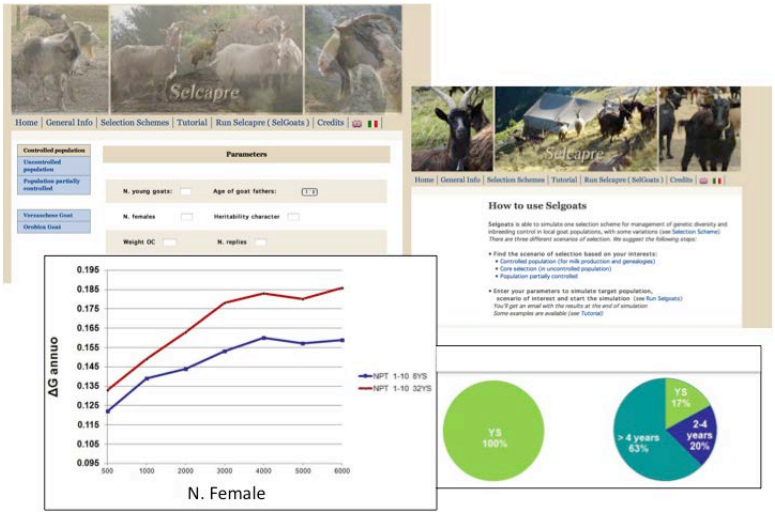


Figure 5. *Example of SELCAPRE screenshot.*

SELCAPRE analyses selection scheme for management of genetic diversity and inbreeding control in local goat populations and provides estimates of achievable results in terms of genetic progress, given a fixed inbreeding rate, given a choice of parameters defining alternative breeding plans (e.g. population size, use of sires, availability of phenotypic, genomic and pedigree information).

In livestock local breeds, selection is often hampered by small population size, incorrect animal identification, inadequate animal performance and pedigree recording, and organizational shortcomings. The use of two or more sires with natural insemination in a single flock, as often observed, does not allow to identify unambiguously assign paternity of newborns without using the use of genetic markers. The limited use of artificial insemination often results into insufficient connections to allow for across-flock genetic evaluation.

The introduction of exotic trans-boundary, more productive, breeds can result in failures because of their poor adaptability to the harsh conditions of the extensive farming environment. Alternatively, selection within local breeds has the potential to balance genetic improvement in productive and adaptation traits accounting for and in traits associated with adaptation to the environment and the local production system and can contribute to economic sustainability of local breeds being a viable livelihood option for farmers who maintain them farming. (FAO, 2013).

Despite the importance of small ruminant farming in Europe, and the need for their conservation and sustainable utilization the available information on genetic programs for local breeds farmed under low input and low technology production systems is scarce.

In such production systems, NEXTGEN is proposing the application of schemes where genetic improvement can be generated in a small fraction of the population, the nucleus, and then disseminated to the whole population (Figure 6.). Within the nucleus, trait and pedigree recording and genomic typing can be carried out at limited cost and organizational effort, and breeding strategies based on sire identification, such as the use of genomics, artificial insemination or of a single sire per flock, can be implemented, making allowing more reliable breeding values estimation genetic evaluations possible. The nucleus population can be an institutional flock in an experimental or public station, or be constituted by two or more coordinated farmer flocks.

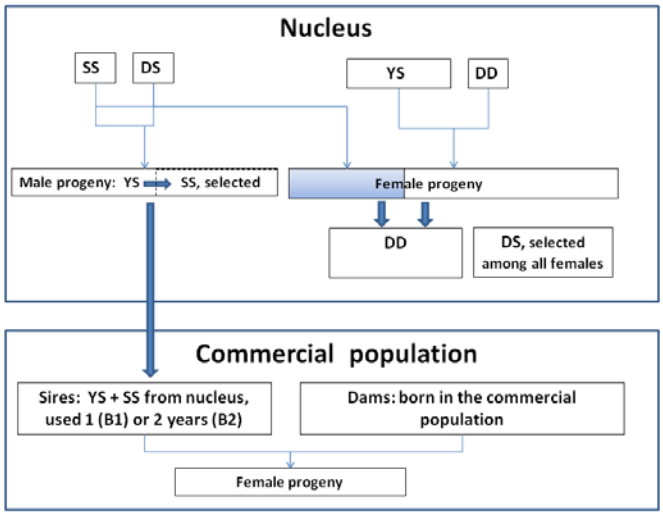


Figure 6. *Breeding scheme: nucleus and commercial population. SS: sires of sires; YS: young sires; DS: dams of sires; DD: dams of dams. B1 and B2 refer to different use of sires from the nucleus in the commercial population, one and two years, respectively.*

The whole population is divided in two tiers: the closed nucleus where selection (traditional and genomic) is carried out, and the commercial population that receives genetically superior sires from the nucleus. Migration from nucleus to the commercial population is restricted to males. Genotype information are collected on sires and dams within the nucleus and used as an aid to preserve

The selection scheme proposed here is efficient for the genetic improvement of local small ruminant populations farmed in low input production systems with low technological level. In the commercial population no pedigree and performance recording is requested, but only the homogeneous use of sires coming from the nucleus. Migration from the nucleus to the population is restricted to males, and assumes no artificial insemination in the population. OC selection in the nucleus requires good pedigree and performance recording, however the adoption of a young sire scheme would facilitate selection even at low organizational levels. In case the optimum nucleus size cannot be adopted, a smaller sub-optimal nucleus breeding structure is a convenient start that allows to begin the selection structure in the breed and to achieve some genetic gain. Later, as conditions will allow, the nucleus could be progressively enlarged. Whenever the nucleus flocks will not reflect the management conditions of the population farms receiving sires from the nucleus, appropriate considerations to avoid wastage of selection efforts should be done.

In livestock science, collections of germplasm and tissue are built for different objectives. While the main function of gene banks is conservation of animal genetic resources for use in the medium or long term, the material stored may also be used for other purposes, e.g. to decrease inbreeding in a population following a genetic bottleneck, by introducing genetic diversity into in vivo populations; to provide flexibility to the livestock industry to change selection goals or to comply to new regulations or changing farming conditions, as in climate change.

One common reason for establishing a gene bank is to provide the possibility of recreating breeds or breeding lines if they are lost as the result of an extreme event that causes breed destruction. Storage of germplasm for this purpose is typically long term, and does not involve frequent use of the stored material or necessitate regular updating of the collection. Gene banks should sample enough animals to capture rare alleles within the respective population, and thereby ensure that their collections cover the range of phenotypes needed in order for them to be used for corrective mating or as a basis for introducing the genotypes needed for adapting breeds to future market demands. When reconstituting a breed from germplasm collections, significant attention must be given to the mating plan, so that after backcrossing has been completed the genetic relationships are minimized and the constant effective number (Ne) is maintained.

NEXTGEN has developed a pipeline for selection of individuals for cryopreservation with the aim of conserving all the alleles and maximizing average kinship calculated at neutral loci and loci relevant to adaptation, after population reconstruction. Two main strategies are applied. The first model assumes that animals have being genotyped with SNP panel spanning across the genome The second model applies when molecular markers are not an option due to costs and pedigree information are available. The application of genetic contribution theory is applied to select the least-related group of germplasm donors.

When genomic information is available, application of strategies to maximize the genetic variation within the group of selected donors result in better breed reconstruction process: lower inbreeding levels are reached and higher genetic similarity to the original breed is ensured. However, if genomic analysis is hampered by costs and logistics, pedigree information may be efficiently used to select individuals for cryopreservation. Finally, when no reliable animal recording is available and resources are insufficient for the use of

molecular information, donors should be carefully chosen based on their geographical location, phenotype and herd history.

1.3.10 Farm animal biodiversity and disease resistance in Uganda

In the course of NEXTGEN project, the genotypic profiles of Ugandan cattle obtained with SNP chip marker panels were analyzed with different approaches to highlight i) the geographical distribution of molecular diversity, the genomic structure and the level of admixture of cattle populations; ii) to identify selection signatures associated to different levels of exposure of animals to disease challenges and therefore likely associated to disease tolerance/susceptibility.

The analyses on genomic diversity in Ugandan cattle were performed on the 54K and 800K SNP datasets and included the calculation of expected (He) and observed (Ho) heterozygosity, and an analysis of population structure with Admixture software (http://www.genetics.ucla.edu/software/admixture/). The calculated values of Ho, He and the geographical distribution of the different genomic components identified by Admixture software were plotted on the map of Uganda at the different scale levels: sampling grid cells, districts, agro-climatic zones and whole country (see Figure 7 and 8 for some examples). The overall analysis of population structure revealed high levels of admixture in Ugandan bovines, together with the occurrence of introgression from the Ankole taurine gene pool into almost all the indicine cattle.

The overlap of the maps of genomic components and heterozygosities at the cell grid level, highlighted areas where some minor genomic components, rare or absent elsewhere in the country, occur at high frequency and where Ho is higher than expected, thus highlighting a probable introgression from still unidentified gene pools. In fact, when the Admixture analyses have been performed including the reference breeds from outside Uganda, the minor genomic components found in Uganda could not be assigned to any of the reference populations. A comparison with a wider reference breed set will be performed as a next step.

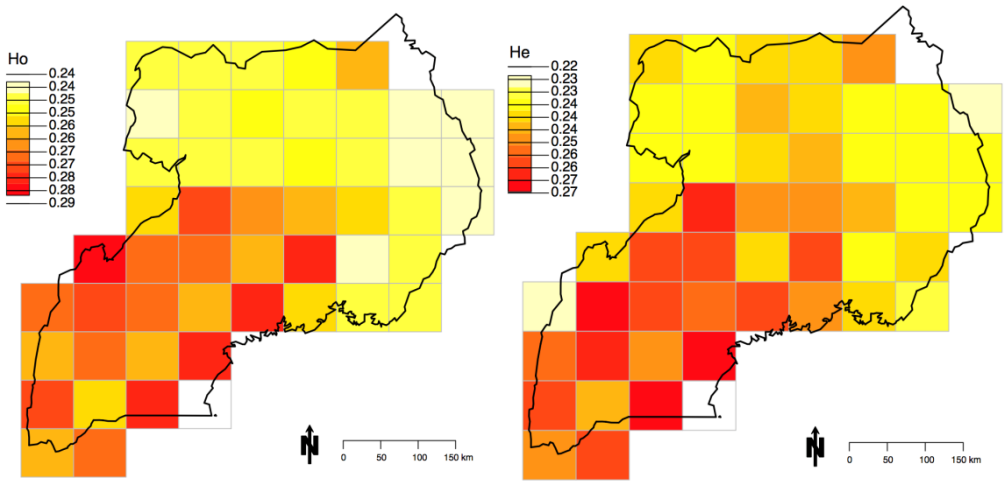


Figure 7. *Observed (Ho, left panel) and expected (He, right panel) heterozygosity values plotted at the grid cell level.*

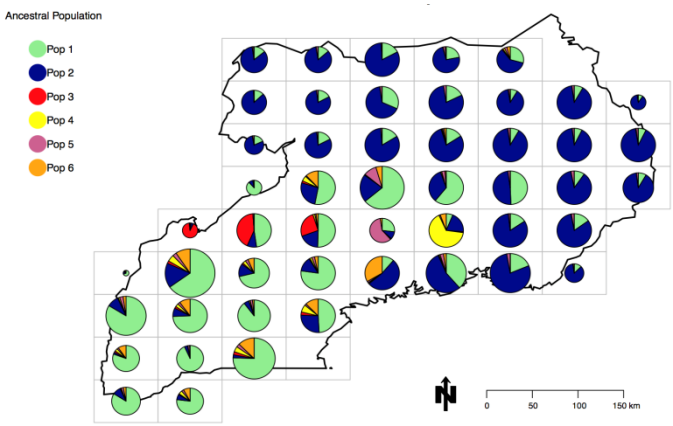


Figure 8. *Results of the clustering procedure performed with the software Admixture plotted on the map of Uganda at the level of grid cells. The example shows the distribution of the genomic components identified when 6 ancestral populations are hypothesized. The results have been averaged and represented as pie-charts of size proportional to the number of individuals sampled.*

Molecular data have also been used to investigate the geographic patterns of diversity in genomic regions flanking candidate genes known to be involved in immune response. An extensive survey of the scientific literature identified 41 candidate genes involved in host resistance to pathogens or in the immune response in cattle or other species (livestock, mouse, humans). Once identified, these genes have been mapped on the bovine genome by online bioinformatic tools and databases. Then, the genotypes of animals at SNP markers located inside or close to genes identified, were extracted from the 50K and HD SNP panels and compared to those in reference populations genotyped in previous projects or having publicly available SNP data. The results were subsequently plotted on the geographical map of Uganda at different scale levels. Figure 9 displays as an example the results obtained for the IL8-Interleukin 8 gene region.

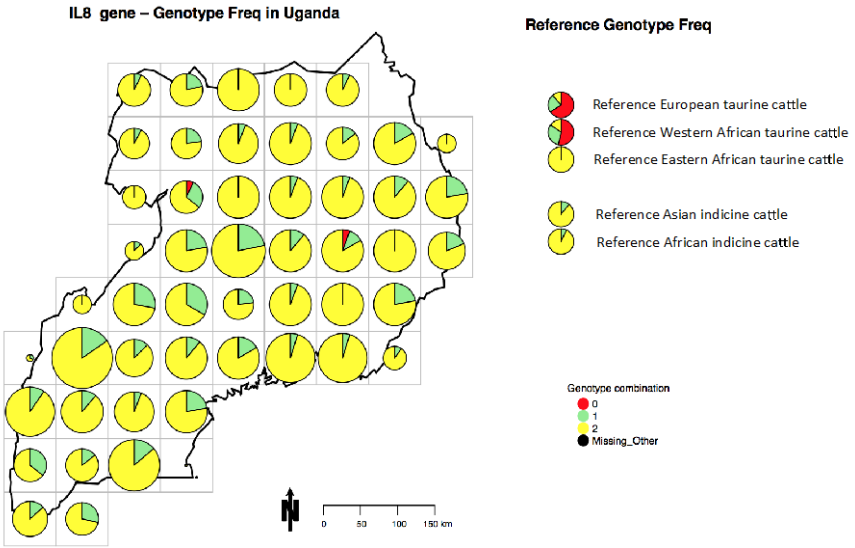


Figure 9*. Frequency of the genotypes scored nearby the IL8 gene. The data are represented at the level of grid cells. The color/number/genotype correspondence is the following: red/0 = homozygote for reference allele; green/1 = heterozygote; yellow/2 = homozygote for the alternative allele.*

These analyses allowed identifying regions around 8 disease resistance and immune response candidate genes showing differences in the geographical distribution of the different genotypes. As a general trend, the African cattle possess at these genes genotypes absent or rare in the European reference breeds and show evidence of varying levels of introgression from Asian zebu. Ugandan cattle, in particular, differ from other bovines from Western Africa, often sharing genotypes with Asian and Eastern African indicine breeds.

To design a proper approach to search for selection signatures, information on laboratory analyses conducted at Makerere University was used to draw maps of disease prevalence in Uganda. In particular, at Makerere, Ugandan cattle blood samples have been tested for the presence of Theileria parva, the parasitic protozoan responsible for East Coast Fever disease, and of Brucella abortus, the bacterium responsible for brucellosis. The data on disease prevalence, calculated as the proportion (percentage value) of sick animals within a sample, have been used in a spatial context to produce synthetic maps at different levels of scale mentioned above.

According to the maps the distribution of both diseases has a geographic component. In particular, East Coast Fever prevalence seems to be higher in central/southern Uganda, while and Brucellosis in the northeastern regions of the country (Figure 10).

Since a large amount of data on diseases prevalence has been made available only close to the end of the project, the results on the relationships between disease prevalence, genomic diversity and selection signatures described here are only preliminary. More detailed analyses, and in particular genome-wide selection signature approaches aided by whole genome sequencing data, are still in progress and will be finalized in the months following the end of the project. Information from the different levels of information will be integrated to jointly evaluate the spatial distribution of molecular variation, selection signatures and disease prevalence data, by estimating the probability of the presence of a specific disease over the country, based on the variation of environmental parameters and genomic information.

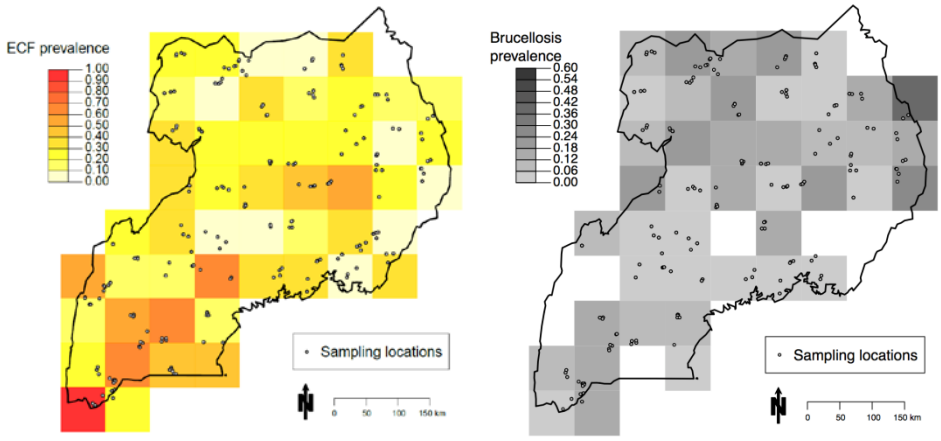


Figure 10. *Maps of disease prevalence in Uganda at the cell grid level. ECF (left panel); Brucellosis (right panel).*

1.3.11 Adaptation to different environments in Morocco

One of the main goals of NEXTGEN was to try to understand the mechanisms underlying the relationships between the environment and genome of small ruminants by: (i) sequencing the whole genome of samples of sheep *(Ovis aries*) and goat (*Capra hircus*) collected in Morocco across a steep environmental gradient from the North of the country towards the

South and covering the range over the mountainous areas and (ii) adapting the available bioinformatics tools and developing new ones to study local adaptation using the whole genome sequence data.

1.3.11.1 Samβada – software to carry out landscape genomic analyses

The software Samβada was developed with the aim of enabling users to carry out geospatially explicit tests for selection on genomic data in a landscape genomics approach where genetic markers are related to environmental variables collected on the individuals' sampling sites. Samβada implements univariate and multivariate models to predict the distribution of polymorphic variants on the basis of environmental variables. Additionally, Samβada also carries spatial autocorrelation analyses to identify whether patterns observed in the spatial distribution of genetic data reflects kin relationship between neighbouring individuals. It is a standalone application written in C++ and was developed using the Scythe Statistical Library for matrix computation and probability distributions, as well as for the development of the application programming interface. Samβada is distributed under an open source GNU General Public License.

1.3.11.2 Signatures of selection for local adaptation in sheep and goats

After sequencing, mapping, variant calling and filtering, a total of 160 sheep and 161 goat Moroccan genomes with ~39 million and ~32 million of variants respectively were selected for the analyses of signatures of selection.

Population structure was assessed in the data using 2 different approaches: (i) a principal component analysis (PCA) and (ii) an ancestry estimate analysis with the software sNMF (Frichot *et al.*, 2014). Both analyses showed a very weak structure in Moroccan individuals. The first and second PCA components explained less than 2% of variation in both species (Figure 11) and sNMF showed that the data were better explained by the presence of a single cluster in each species.

To study local adaptation, two approaches were chosen: (i) On one hand population genetic approaches that identify candidates under selection on the basis of deviations from the neutral allele frequency spectrum and changes in linkage disequilibrium after positive selection (i.e. using XPCLR (Chen *et al.*, 2010), SweeD (Pavlidis *et al.*, 2013), iHS (Voight *et al.*, 2006)). For some of these analyses it was necessary to define populations that were compared against each other (e.g. XPCLR, iHS); therefore for each variable tested we selected the animals occurring on the extremes of the environmental gradient (e.g. for the variable altitude we compared the 25 sheep occurring at lowest altitude – less than 219 meters – and the 25 sheep occurring at the highest altitude – 1433 meters or higher). (ii) On the other hand, the data were analysed using correlative approaches between SNPs and environmental variables collected for each point on the sampling grid of Morocco. For this purpose the software Samβada and LFMM (Frichot *et al.*, 2013) were used.

The preliminary results with XPCLR identified several candidate genes and regions under selection for different environmental parameters. In sheep, several selective signals were identified for low/high altitude (e.g. on chromosome 20; Figure 2), slope, temperature annual range, precipitation in March and the mean temperature of July. Similarly, in goats, strong signals were identified for several parameters, such as, the mean temperature of July (e.g. 2 signals detected on chromosome 18; Figure 3), temperature annual range, altitude, precipitation in March and the mean temperature of the warmest quarter. Further analyses using the other environmental parameters are currently in progress. Among the selection signals found so far by XPCLR, several genes were identified such as the sheep gene *GMDS*

for altitude (Figure 12: Chromosome 20: 50,065,679-50,397,477) or the goat genes *AGRP* and *CTCF* for the mean temperature of July (Figure 3). However, some other identified signals have not been linked to any known gene in sheep and goat genomes. Top lists of candidate genes showing selection signatures for each parameter in each species are being developed.

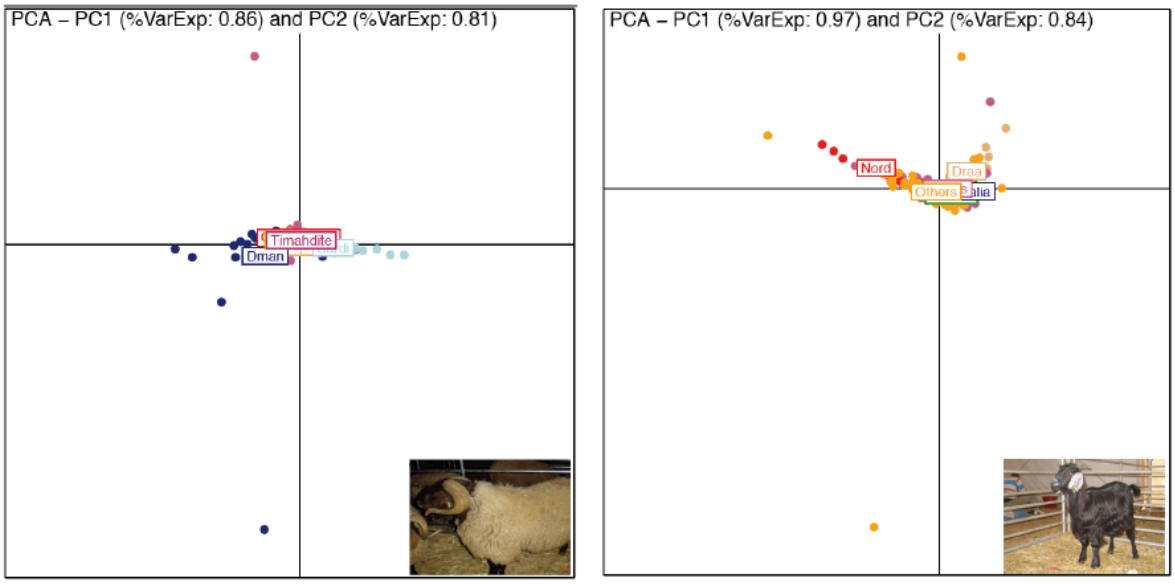


Figure 11. *Distribution of the 160 sheep (left) and 161 goats (right) according to the first two principal components (PC). The variance explained by each PC is mentioned in the top of each graph. Colours distinguish the phenotypic groups of individuals (i.e. breeds or populations)*

The preliminary correlative analyses with Samβada identified six loci under selection in sheep and five in goat. The loci identified in sheep were significant for statistical association models involving the variables longitude and precipitation in the third and ninth month. Contrastingly, for goat the identified SNPs were associated to the variables slope, aspect, curvature, precipitation in the third month and sunshine duration on the 21st of June. These loci seem very few when the total set of loci is considered. However, the False Discovery Rate approach used to select models is still under development and this result is likely to change, as the method is refined.

These results were encouraging since they allowed identifying several genes/markers under selection across the Moroccan landscape. The combination of the results of different approaches would validate genes showing signatures of selection. The functional annotation of these candidate genes and the study of other environmental parameters are currently underway. These analyses will help determining whether the candidate genes/loci identified in each species reflect the same or similar metabolic pathways.

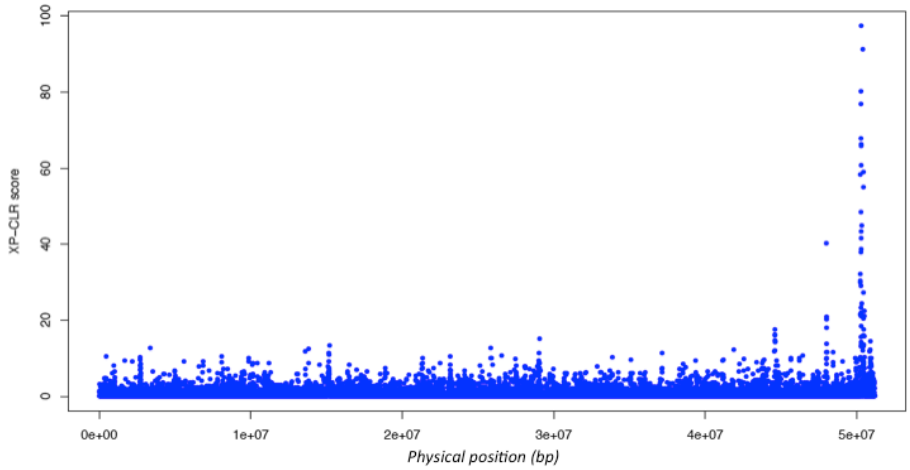


Figure 12. *Selection signatures in sheep across chromosome 20 related to differences in altitude. 25 low altitude animals were compared to 25 high altitude animals. The high peak of XPCLR score correspond to the gene GMDS*

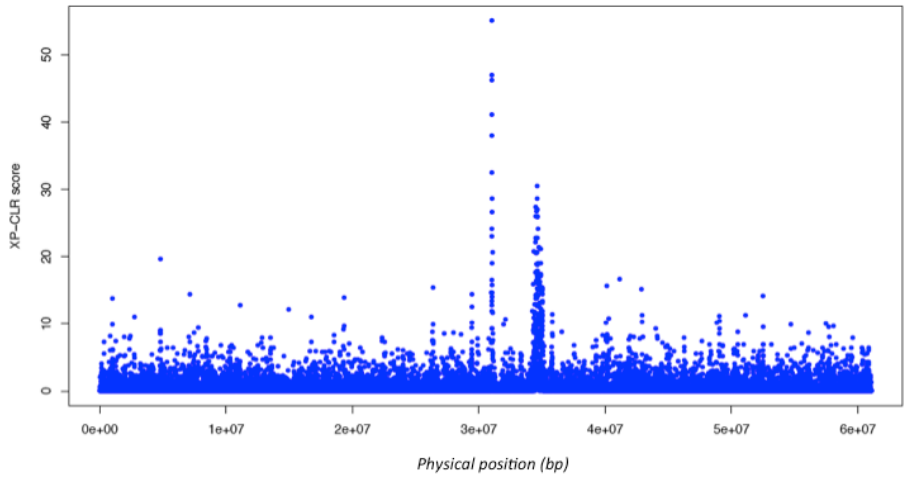


Figure 13. *Selection signatures in goats occurring in high mean temperature of July across chromosome 18. 15 goats occurring in low temperature in July were compared to 15 goats occurring in high temperature during that month. The high peak of XPCLR score does not correspond to any known gene and the second peak spans several genes (e.g. AGRP, CTCF)*

1.3.11.1 Wild ancestors versus local and industrial breeds as genetic resources

The domestication of sheep (*Ovis aries*) and goats (*Capra hircus*) happened around 10.500 years ago in the Middle East from the wild species mouflons (*Ovis orientalis*) and bezoars (*Capra aegagrus*). Due to human selection, the level of genetic variation has been probably reduced in the domestic animals compared to the wild animals. Moreover, along with the emergence of the concept of breed, selection was progressively intensified in the last 200 years. It is thus likely that traditionally-managed populations present more genetic variation than industrial breeds. It is therefore a major concern to assess the impact of both selection processes on the genetic resources of sheep and goat and to determine whether the wild species and the traditional populations may represent genetic resources for future breeding options.

1.3.12.1 Sampling

As shown in Table 2, we analyzed a dataset including individuals from Iran and Morocco representing local breeds or populations and individuals representing sheep and goat industrial breeds. We also used samples representing the two wild species mouflons and bezoars from Iran.

Table 2. *Samples collected. The species, data origin, code and number of individuals is given for each sample.*

Species	Origin	Code	Sample size
O. orientalis	Iran	IROO	16
O. aries	Iran	IROA	20
O. aries	Morocco	MOOA	20
O. aries	Industrial breeds	indusOA	20
C. aegagrus	Iran	IRCA	19
C. hircus	Iran	IRCH	20
C. hircus	Morocco	MOCH	20
C. hircus	Industrial breeds	indusCH	10

1.3.12.2 Results

When looking at the genetic structure (Figure 14; K = 1 and 2), the wild and the domestic species were first detected as two distinct genetic pools, which were then sub-divided when increasing the number of clusters. While different groups were detected within the wild animals, the domestic animals were separated among Iranian, Moroccan, and industrial breeds (Figure 14; K = 3 to 5).

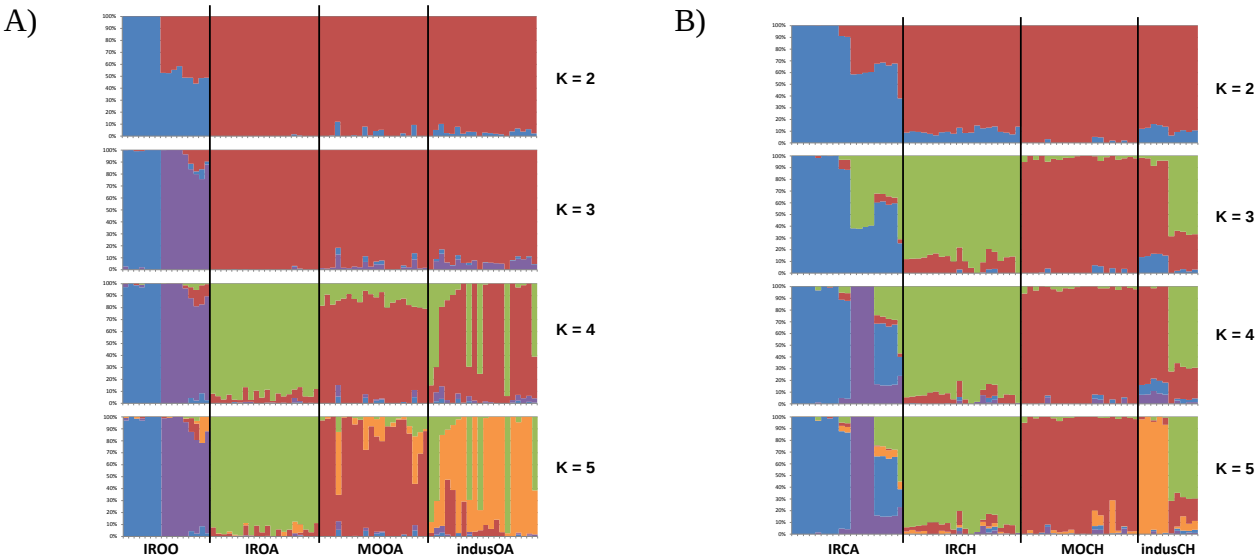


Figure 14. *Genetic structure among A) Ovis and B) Capra samples for increasing number of clusters K from 2 to 5 using sNMF. Each individual is represented by a bar where the different colors represent the assignment probabilities to the K clusters.*

Sheep and mouflons

Among *Ovis* samples, the wild ancestors showed a higher number of polymorphic variants (30.1 million) and higher nucleotide diversity (pi = 0.210) compared to the domestic animals (see below). The variants correspond to both SNPs and indels.

The Iranian mouflons were separated in two groups at K = 3 clusters (Figure 14), with different levels of genetic diversity. For one group, the individuals were sampled on an island where the population was established probably from a few individuals introduced several decades ago for hunting purposes. The 7 individuals all present high levels of inbreeding (mean F = 0.27) and they all show high relatedness among them (IBS = 0.913 on average) compared to the other 9 individuals (mean F = 0.10 and IBS = 0.842 on average). They also harbor lower nucleotide diversity (pi = 0.114 against 0.171) and a lower number of polymorphic variants (14.3 million against 28.5 million). This group thus clearly does not represent the genetic diversity of other Iranian mouflons. Comparatively, the other group of Iranian mouflons seems to possess a high level of genetic resources, even higher than found in the domestic samples. When comparing domestic sheep samples together, the levels of nucleotide diversity were quite homogeneous ranging from 0.139 for the Iranian sample to 0.141 and 0.145 for the industrial breeds and the Moroccan sample respectively. The number of polymorphic markers was slightly higher in Morocco (27.2 million variants) compared to the two other samples (around 25 million variants each). Thus, it seems that genetic resources were not reduced during spread from Iran to Morocco and during the industrial breeding in sheep. However, the industrial breeds showed rather high inbreeding values (mean F = 0.20) compared to the Iranian and Moroccan individuals (mean F = 0.15 and 0.16 respectively). This potentially indicates that each breed experienced a loss of diversity, which is however well preserved at the worldwide scale. Consequently it seems that taking together the traditionally-managed and industrial domestic sheep breeds have relatively well preserved genetic resources, but still lower than the wild species.

Goats and bezoars

Among *Capra* samples, the highest level of nucleotide diversity was found within Iranian goats (pi = 0.125) followed by Moroccan goats (pi = 0.118), Iranian bezoars (pi = 0.109) and industrial breeds (pi = 0.092). The pattern was globally the same for the number of polymorphic sites.

From the results of the genetic structure, the 19 Iranian bezoars could be subdivided in three geographic groups. For two groups, the number of polymorphic sites and the levels of nucleotide diversity were rather low (respectively 6.5 and 11.4 million variants and pi = 0.077 and 0.099), certainly due to genetic drift caused by isolation from other populations. The third group of 5 individuals showed a rather high level of nucleotide diversity and a high number of polymorphic sites compared to its small sample size (pi = 0.121 and 12.9 million variants). The maintenance of genetic diversity in this third group may be explained by the possible hybridization with domestic animals, as shown by the admixture with the cluster representing the Iranian goats.

The domestic goats have experienced increasing selection intensity from Iranian and Moroccan traditional populations to industrial breeds. While respectively 19.2 and 21.7 million variants were found in the Iranian and Moroccan populations, the industrial breeds showed only 11.2 million variants. This sample also showed higher inbreeding (mean F = 0.24) compared to Iranian and Moroccan goats (mean F = 0.08 and 0.15 respectively). The number of breeds representing the industrial sample is lower than in *Ovis* but at least in these 4 breeds the intensity of selection has leaded to an important erosion of genetic diversity. This result suggests that the genetic resources present in Iranian and Moroccan goats, and to a

lower extent in Iranian bezoars, could be helpful to restore the potential of adaptation of the industrial breeds in the future.

1.3.12 Conclusion

We are currently witnessing a dramatic loss of biodiversity at an unprecedented rate. Two major advances in the last two hundred years have had a major impact on the diversity of livestock. Namely, the implementation of the breed concept and the introduction of artificial insemination, which have helped farmers and breeders to increase the quality and amount of product, by identifying animals carrying valuable traits and focusing on them for breeding. While this approach seems sensible from a farmer's or breeder's perspective, from the conservation stand point it poses challenges for the maintenance of biodiversity. In particular, selection results in derived populations losing a substantial part of their genetic variation and adaptive potential. Additionally, the replacement of locally adapted indigenous breeds by breeds that seem to provide immediate gains (e.g. higher milk production) may result in the loss of valuable adaptive genomic resources. Consequently identifying methodologies that can be used to delineate recommendations on using genomics to evaluate the distinctiveness and genomic value of livestock resources is important in light of current breeding practices and environmental challenges such as sustainable intensification needs and global change. In this context NEXTGEN used new generation genomic and reproductive technologies to develop innovative approaches and characterize Farm Animal Genetic Resources, producing 1355 Whole Genome datasets (including 447 whole genome sequences and 907 SNP-Chip datasets).

Based on Whole Genome Sequences, comparison of the distribution of genomic variation between wild relatives and domestic species suggests that both wild populations and indigenous breeds represent an especially valuable genetic reservoir for the future. Higher genomic diversity was found in wild *Ovis orientalis* compared to its domestic counterpart (*O. aries*). Domestic sheep (breeds in the domestication center, Morocco and cosmopolitan breeds) have similar levels of polymorphism indicating a parallel loss in genetic diversity since the domestication process (~30 million SNPs for mouflon and ~25 million for domestics). In contrast, the wild bezoar (*Capra aegagrus*) has reduced genetic variation (~17 million SNPs) when compared to its domestic counterpart (*C. hircus*), irrespective of whether it is compared with the Iranian or Moroccan samples studied by the consortium (~21 and ~19 million SNPs respectively). However, all indigenous breeds and wild populations (even the ones with a level of inbreeding comparable to that of industrial breeds) have a high number of alleles not found in domestics (> 10 million in both *Ovis* and *Capra*) that may also provide a fund of variants of an adaptive nature.

Analyses were carried out to use genome resequencing data and SNP arrays to identify genetic variants involved in local selection. While the bulk of these analyses are ongoing, first results show that the Moroccan animals (both sheep and goat), and Ugandan cattle carry strong signatures of natural selection in their genome. The Moroccan sheep and goat dataset was queried with population genetic approaches and landscape genetics, and both approaches identified SNPs involved in adaptation to gradients of environmental variables (e.g. the gene *GMDS* related to altitude in sheep, or the genes *AGRP* and *CTCF* related to temperature in goats). Contrary to the lack of population structure observed in the Moroccan animals, Ugandan cattle could be easily divided into two major groups comprised of Ankole (*Bos taurus*) and Zebu animals (*Bos indicus*). Analyses of signatures of selection identified markers such as HM-28, showing a genotype which distribution of reflects the habitat of the pathogen *Trypanosoma brucei rhodesiense* responsible for sleeping sickness. These preliminary results make it apparent that the genome-wide data produced by the consortium carries valuable

information regarding the evolutionary processes that have affected the distribution of genetic variation in these species, both in terms of selection and demography. These data, when finalized, will be used to assess the distribution of locally adaptive versus common genetic variants to enable prioritization of animal genomic resources, which will balance neutral and selected genetic variants in a geospatial context.

Additionally, simulation studies were carried out to identify alternative methods of genome-assisted breed conservation. Guidelines were defined to optimize breeding strategies that conserve both neutral and adaptive variation. In this context, a simulation approach showed that a nucleus population (with controlled inbreeding and optimal contributions from each sire) is likely to genetically improve a local breed that it augments, and which is used for commercial purposes. For this approach to work efficiently, it is necessary to have good pedigree and performance recording for the animals in the nucleus and to specify the size of the nucleus in a breed and farming area context. An alternative simulation based approach designed to identify animals for biobanking showed that genomic information significantly improves the chances of reconstructing breeds from a selected group of individuals over the absence of genetic data, i.e. it can maximise the conserved genetic variation (including rare alleles) and minimises inbreeding. However, because collecting genetic data can be costly, in its absence any available information on the animals' pedigree can be efficiently used, although the reconstituted breed may not harbour particular aspects of the original's breed genetic variation (e.g. rare alleles which may have an adaptive advantage).

Finally, NEXTGEN demonstrated the functionality of chromosomes isolated from lyophilized oocytes, which direct early embryogenesis upon injection into fresh previously enucleated oocytes. This unprecedented finding supports the use of freeze- drying, a technically easy and low-cost strategy, for the storage in bio-banks of cell samples and gametes for biodiversity preservation.

Thus, besides stressing the role of indigenous breeds and wild relatives to act as reservoirs of neutral and adaptive diversity, NEXTGEN developed new methods for the identification genomic resources, the choice of appropriate breeding strategies through simulation approaches and the identification of individuals of interest for conservation purposes, for instance via bio-banking.

1.3.13 Cited references

Carter TH (1991) Biotechnology, economics, and the business of blood. *Biotechnology*, **19**, 3-30.

Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. *Genome Res*earch, **20**, 393-402

Danecek P, Auton A, Abecasis G, *et al.* (2011) The Variant Call Format and VCFtools. *Bioinformatics* **27**, 2156-2158.

FAO (2007) *The state of the world's animal genetic resources for food and agriculture*. Commission on genetic resources for food and agriculture. Food and agriculture organization of the United Nations, Rome.

FAO (2008) High-level conference on world food security: the challenges of climate change and bioenergy. Food and agriculture organization of the United Nations, Rome. (http://www.fao.org/foodclimate/hlc-home/en/)

FAO (2013) *Sustaining livestock diversity*. Food and agriculture organization of the United Nations, Rome.

Flicek P, Amode MR, Barrell D, *et al.* (2014) Ensembl 2014**,** *Nucleic Acids Research*, **42**, D749-D755.

Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics*, **196**, 973-983.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699

Gnerre S, Maccallum I, Przybylski D, *et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of USA*, **108**, 1513-1518

Joost S, Bonin A, Bruford MW, *et al.* (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Leinonen R, Sugawara H, Shumway M (2011) The Sequence Read Archive. *Nucleic. Acids Research*, **39**, D19-D21, doi:10.1093/nar/gkq1019.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25**, 1754-1760.

Li H, Durbin R (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475**, 493-496.

Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Loi P, Matsukawa K, Ptak G, *et al.* (2008) Freeze-dried somatic cells direct embryonic development after nuclear transfer. *PLoS ONE*, **3**, e2978.

McKenna A, Hanna M, Banks E, *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.

Pavlidis P., Zivkovic D., Stamatakis A., Alachiotis N. (2013) SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, **30**, 2224-2234.

Porter V (2002) *Mason's world dictionary of livestock breeds, types and varieties. 5th edition* CABI Publishing, Wallingford, UK.

R Development Core Team (2005) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, Negrini R, Ajmone-Marsan P (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology*, **17**, 275-284.

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

1.4 Potential impact and the main dissemination activities and exploitation of results

1.4.1 Strategic impact

1.4.1.1 Added value of collaborations between different scientific disciplines

Innovation is fostered by information gathered from new connections, and particularly from connections among different scientific disciplines. Scientists of the NEXTGEN consortium are from diverse scientific disciplines such as conservation genetics, genomics, bioinformatics, geographic information science, veterinarian sciences, and agricultural sciences. In this context, it is clear that the close collaborations established among the disciplines during the course of the project represents an important added value, and will undoubtely lead to further collaborations in the future.

1.4.1.2 Impact on capability-building

NEXTGEN work plan gave a large place to capability-building, and thus expect a high impact at that level. Different actions were implemented to specifically enhence capability-building:

(i) two open workshops have been held in Morocco and Uganda, involving hundreds of participants in total;

(ii) at least three PhD students from the three ICPC countries were under co-supervision (cooperative PhD programmes), starting their PhD within WP1.2 (sampling) in their own country, and then staying two years in a European laboratory for participating to the data analysis;

(iii) courses for all partners during general meetings, with the specific goal of bridging the gap between disciplines (Task T3.1.1);

(iv) exchange of scientists, post-docs, PhD students, technicians among different partners having different scientific backgrounds when necessary according to the work plan.

1.4.1.3 Impact on management of farm animal biodiversity and genetic resources

Most of the industrial breeds come from Europe, and a large part of the potential genetic resources lies outside of Europe. In such a context, it was extremely important to take into account the farm animal biodiversity at the world-wide level, and to work within an international context, involving the appropriate ICPC countries. NEXTGEN fully exploited this opportunity, and involved three key partners from ICPC countries, Iran for having access to sheep and goats wild ancestors and to local breeds from the domestication centre, Uganda for allowing to work efficiently on disease resistance in cattle, and Morocco for the unique chance of developing a sound landscape genetics approach in sheep and goats. There is no doubt that the involvment of these countries has been an important added value to NEXTGEN and will have an important impact of the conservation strategies that will be established by the different countries.

The main goal of NEXTGEN was to produce optimized tools and methods to assess farm animal biodiversity and genetic resources. More specifically, NEXTGEN provided precise methodology for studying the biodiversity aspect of disease resistanc and the relationships between genome and environment.

Beside optimized tools and methods, NEXTGEN also produced key results on the management of genetic resources. The value of wild ancestors of sheep and goat as genetic resources has been demonstrated, as well as the value of cattle, sheep, and goats from the domestication centre in the Middle East. Surprisingly, traditionnal sheep and goat populations from Morocco also harbor a high level of genetic diversity and have a high conservation value. This is not the case for industrial breeds.

Therefore, the NEXTGEN outputs will have a strong impact on governemental and non-governemental organizations in charge of preserving farm animal biodiversity and on managing the genetic resources, such as (i) the Food and Agriculture Organization of the United Nations (FAO) or breeder associations for the domestic breeds, and (ii) the International Union for Conservation of Nature (IUCN) or the World Wide Fund for Nature (WWF) for wild ancestors.

1.4.1.4 Impact on farm animal breeding and biobanking

The bioinformatic tools developed during the course of NEXTGEN has been especially designed to optimize the selection of individuals both for breeding and for biobanking, according to criteria related to the importance of neutral variation.

The novel approach for biobanking based on freeze-dried somatic cells stored at room temperature opens unprecedented opportunities for alternative biobanking conservation of endangered/rare breeds and will have a large impact on technological related areas.

The surrogates for whole genome data (a set SNPs producing unbiased results compared with whole genome sequences) will also be an important output for breeders and for the industry if the goal is to preserve as much diversity as possible. Finally, the 447 whole genome sequences produced will constitute a very valuable resource for the whole scientific community working on farm animals.

1.4.1.5 Additional impact on conservation and evolutionary biology

To our knowledge, NEXTGEN was the first project in the area of conservation genetics that proposes a comparative analysis of whole genome data at the intraspecific level. Therefore, the project gathered data on an unprecedented scale on all major types of genetic variation in the genome af cattle, sheep and goats. We indeed expect a high impact, far beyond the farm animal scientific community mainly on conservation and evolutionary biology.

First, the development of bioinformatic methods and tools for handling whole genome sequence data for a conservation purpose is of general interest for other studies that intend to use whole genome data in conservation genetics.

Second, the project used high troughput technologies at the upper limit of those available to generate an enormous amount of molecular information made accessible to the scientific community trough public databases that will provide the foundation for further extensive studies concerning genetic variation.

Third, the development of methods to identify genomic regions under selection allows the application of the same approach to study local adaptation in any kind of organisms, provided that several whole genome sequences will be available. The study of the mechanisms responsible of local adaptation is an area of growing interest from the scientific community, and it is clear that it will be boosted by the current improvement in sequencing technology.

Fourth, the sampling approach implemented in NEXTGEN in Uganda and in Morocco allows landscape genetic analyses based on whole genome data (either whole genome

sequence, or large SNP panels). Sampling using a grid system opens new opportunities at the data analysis stage, and even allows analyses comparable to classical associations studies.

Fifth, the conservation strategy elaborated within NEXTGEN, i.e. a conservation strategy based on whole genome data and taking into account adaptive aspects, can also be applied to wild plant and animal species. It can help the prioritization of populations in order to maximize the neutral and adaptive diversity that will be preserved. With the progress in DNA sequencing and the decrease of the cost per genome, it is likely that conservation strategies based on whole genome data will be the gold standard in few years. NEXTGEN pionniered this research area, and thus will have a large impact on the scientific community.

Finally NEXTGEN provided an outstanding example of 'European research excellence' and support Europe as the world-leader in the field of sustainable use of livestock and conservation of biodiversity resources.

1.4.2 Dissemination of project results

Due to the late availability of the sequence data, we had first to focus on the data analysis during the last months of the project, postponing most of the dissemination activities after the official end of the project.

Results from research conducted within the NEXTGEN consortium were disseminated both within the consortium to improve the knowledge of its members and outside the consortium towards the scientific community, the general public, and commercial organisations. Knowledge has been disseminated through a variety of supports (public web site, scientific papers, oral communications and posters during meetings, special actions towards industry, breeders, and stakeholders etc.), ensuring that the results are exploited and understood by our target groups identified above. We plan to organize a final open meeting in Morocco after the official end of the project, most probably early in 2015. At that time the data will be fully analyzed, and the dissemination towards scientists and stakeholders will be efficient.

Dissemination to the scientific community seems to be the easiest task. The scientists involved in NEXTGEN have an excellent practice of publishing in leading journals. We are very confident that the dissemination towards other scientists will be very efficient. In addition, to ensure that the potential impact of NEXTGEN project will be fully realised, as soon as possible, the consortium will share the resources i.e. samples, protocols and analytical methods and will release the molecular data gathered through freely accessible or public databases (see above).

Given the relevance of the topics addressed by NEXTGEN, the dissemination activities are central, not only within the scientific community, but particularly for the general public audience. The dissemination towards the public will be promulgated mainly via the NEXTGEN web site. If wildlife conservation is very well promoted by organizations like Green Peace, the World Wide Fund for Nature (WWF), or the International Union for Conservation of Nature (IUCN), the problem of farm animal biodiversity has a much lower public profile. The preservation of genetic resources in domestic animals does not have the same image for the public as preserving the giant panda or whales. However farm animals represent an important source of protein, work-power and companionship for mankind, and preserving their genetic resources is equivalent to preserving our future. Another way to reach the public consists to launch a press release each time that an important scientific result is produced, leading to scientific popularization via magazines. Many members of the consortium have regular and intensive contacts with the media, like scientific magazines, but also radio and television at national and international level. These contacts hold a great potential for dissemination.

Finally, the technological transfer towards industry, breeders, and stakeholders is leaded by partner P04 (SME), with the help of partners P02 and P03.

1.4.3 Management of intellectual property

The NEXTGEN position in this area is very simple, and consists to make all of the knowledge and data produced as part of NEXTGEN freely available. It is therefore NEXTGEN's intent that all genomic DNA sequence data generated by the project be released and placed in the public domain where it will be available. In order to implement this policy, the sequences will be available via the Ensembl database maintained by partner P05.

In accordance with the standard 'Fort Lauderdale Principles' (http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf), data are made available to the community under the "Responsible use" and in a way that considers the roles and responsibilities of data producers, data users, and funders of "community resource projects", and propose a balance between the interests of scientific community in rapid access to data and the needs of data producers to receive recognition for their work. "Responsible use" was defined as allowing the data producers to have the opportunity to publish the initial global analyses of the data, that will also ensure that the data generated will be fully described.

Since Partner P05 (EMBL) is actively involved in data storage and handling, NEXTGEN also agrees with the general rules for EBI services users stated in "Terms of Use of the EBI Services" (http://www.ebi.ac.uk/Information/termsofuse.html).

In the same way, the management of intellectual property concerning the genetic resources is very simple and is regulated by the Rio Convention (United Nations 1993): the genetic resources identified whithin NEXTGEN will remain the property of the country of origin. All the countries involved in NEXTGEN as well as the European Union signed the text of this convention in 1992.

1.5 Project public website and contact

The website provides the following information:

1. General information on the project
2. The project presentations
3. The official documents of the project when they are public
4. The work done for each activity of the project through descriptions pages
5. The public data and documents issued by the NEXTGEN work

Website address: http://nextgen.epfl.ch/

Contact: Stéphane Joost, EPFL, Switzerland, **Stephane.Joost@epfl.ch**

List of all beneficiaries with the corresponding contact name and associated coordinates

Code	Beneficiary name	Short name	Country	Team Leader	Email
CO01	Centre National de la Recherche Scientifique	CNRS	France	Pierre Taberlet	pierre.taberlet@ujf-grenoble.fr
P02	Cardiff University	CU	United Kingdom	Mike Bruford	BrufordMW@Cardiff.ac.uk
P03	Università Cattolica del Sacro Cuore	UNICATT	Italy	Riccardo Negrini	riccardo.negrini@unicatt.it
P04	Parco Tecnologico Padano	PTP	Italy	Alessandra Stella	alessandra.stella@tecnoparco.org
P05	European Molecular Biology Laboratory - European Bioinformatics Institute	EMBL	Germany	Paul Flicek	flicek@ebi.ac.uk
P06	Ecole Polytechnique Fédérale de Lausanne	EPFL	Switzerland	Stéphane Joost	Stephane.Joost@epfl.ch
P07	Università Teramo	UNITE	Italy	Pasqualino Loi	ploi@unite.it
P08	Makerere University	MAK	Uganda	Charles	cmasembe@muienr.mak.ac.ug
P09	Institut National de la Recherche Agronomique	INRA-Mor	Morocco	Aziz Fadlaoui	fadlaouiaziz@yahoo.fr
P10	Gorgan University of Agriculture and Natural Resources	GAU	Iran	Hamid-Reza Rezaei	hamid.r.rezaei@gmail.com
P11	Commonwealth Scientific and Industrial Research Organisation	CSIRO	Australia	James Kijas	James.Kijas@csiro.au

2. Use and dissemination of foreground

SECTION A (PUBLIC)

A). PUBLIC FINALISED PROJECT OUTPUTS

Beside the papers and manuscripts listed below, we envisage further publications in at least three high profile papers:
(i) the tracking of the domestication genes by comparing genomes of local breeds of sheep and goats in the domestication center with the relevant wild ancestor (to be submitted to Science or Nature);
(ii) the landscape genomics results in Morocco; gene/environment relationships (to be submitted to Science or Nature);
(iii) the synthesis about the strategy for preserving the genetic diversity of farm animals (to be submitted to Conservation Biology).

Several international projects are already using (or going to use soon) NextGen data (Goat ADAPTmap Project, International Sheep Genome Consortium, International Goat Genome Consortium, the 1,000 bull project).

European projects will also take advantage of NextGen samples and data to characterize livestock genomics resources and their use in setting up breeding strategies in the context of climate change. One project is already financed (i.e., ClimGen - FACCE ERA-Net Plus « Climate Smart Agriculture ») and other are submitted (e.g., GenResUse submitted to phase II - H2020 "Traditional resources for agricultural diversity and the food chain" - SFS- 07a-2014). Moreover, We are already in contact with J. McEwan (International Sheep Genomics Consortium, AgResearch, New Zealand) for sharing variants in order to contribute to elaborate a new LD Chip for sheep genotyping.

Note that all the sequence data produced during the course of the project are publicly available (the 448 whole genome data for cattle, sheep, and goats were released on 1st September 2014), with the following data usage agreement:

Use of the NextGen Project data

NextGen project data are being released early, prior to their publication, in the expectation that they will be valuable for many researchers. In keeping with Fort Lauderdale principles, data users may use the data for many studies, but are expected to allow the data producers to make the first presentations and to publish the first papers with global analyses of the data.

Global analyses of Project data

The NextGen project will publish global analyses of the sequence data and quality, SNPs, structural variants, STRs, microsatellites, and population genetic phenomena such as breed and population comparisons, mutation rates, signals of selection, and the association of loci with phenotypes and environmental variations (landscape measurements). Talks, posters, and papers on all such analyses are to be published first by the NextGen Consortium, by approved presenters on behalf of the Project, with the Consortium as author. After the NextGen consortium has published analyses, then researchers inside and outside the Project are free to present and publish using the Project data for these analyses.

Methods development using Project data

Researchers who have used small amounts of Project data (<= one chromosome) may present methods development posters, talks, and papers that include these data prior to the first major NextGen Project paper, without needing Project approval or authorship, although the Project should be acknowledged. Methods presentations or papers on global analyses or analyses using large amounts of Project data, on topics that the Consortium plans to examine, would be similar to large-scale analyses of Project data: researchers within the Project may make presentations or submit papers at the same time as the main Project presentations and papers, and others could do so after the Project publishes analysis papers.

Population comparisons using Project data

Researchers may use Project data as controls or additional information for comparisons with their samples from other populations, prior to the major Project paper being published, as long as the analyses that the Project plans to do are not included. These are not Project studies, and the Project should not be listed as an author.

Permission and acknowledgements

Researchers who have questions about whether they may make presentations or submit papers using Project data, or whether to include the NextGen Consortium as an author, may contact Pierre Taberlet (pierre.taberlet@ujf-grenoble.fr)

The NextGen project should be acknowledged in publications and posters with the following text:

This study makes use of data generated by the NextGen Consortium. The European Union's Seventh Framework Programme (FP7/2010-2014) provided funding for the project under grant agreement no 244356 - "NextGen".

A1). Scientific Publications

LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	**D.O.I.**	**Title**	**Author(s)**	**Title of the periodical or the series**	**Number, date or frequency**	**Publisher**	**Place of publication**	**Date of publication**	**Relevant pages**	**Open access is/will be provided to this publication?**
1	10.1089/cell.2013.0032	Sheep: the first large animal model in nuclear transfer research.	Loi P, Czernik M, Zacchini F, Iuso D, Scapolo PA, Ptak	Cellular Reprogramming	quarterly	Bentham	USA	-	367.373	No
2	-	"General and specific hurdles of Somatic Cells Nuclear Transfer in wild animals: a realistic view and some proposals for improvement"	Pasqualino Loi (Lino Loi & Grazyna Ptak)	Book	2010	Life Sciences Publisher	Poland	2010	88-94	No
3	-	Constitutive deregulation of DNA methyltransferases (DNMT1) throughout development of embryo derived from prepubertal	Iuso D, Zacchini F., Fidanza A., D'Agostino A., Toschi P., Failla C., Loi P., Ptak G.,	Chromosome Research	monthly, vol. 18	Springer	Berlin	2010	737	No

		sheep embryos								
4	-	Transplantation of nucleoli into human zygotes: not as simple as expected?	2. Josef Fulka, Jr., Alena Langerova, Helena Tyllerova, Pasqualino Loi, Stanislava Martinkova and Helena Fulka	Journal of Assisted Reproduction Genetics, IF. 1,23, in press	monthly	Springer	USA	2011	May;28(5):385-9	No
5	-	Non-living genomes: trash or recycle?	Pasqualino Loi, Josef Fulka Jr, Thomas Hildebrand and Grazyna Ptak	Reproduction	monthly	Society for the Study of Reproduction, Cambridge, UK	UK	2011	Oct;142(4):497-503	No

6	-	Biological time machines: a realistic approach for cloning an extinct mammal. Endangered Species research. In press. IP 1.563	Pasqualino Loi, Teruhiko Wakayama , Josef Fulka Jr and Grazyna Ptak	Endangered Species research. In press. IP 1.563	monthly	Inter-Research Science CenterInter-Research Science Center (IR)	Germany	2011	227-223	No
7	-	Current State of Art of Large Animal Cloning: Any Lesson for Future Progress?	Loi P & Ptak G	OOCYTE MATURATION AND FERTILIZATION: A long history for a short event	e-book	Bentham books	Bentham e books http://www.benthamscience.com/ebooks/9781608051823/index.htm	2011	149-155	Yes
8	-	Efficient Production and Cellular Characterization of Sheep Androgenetic Embryos	Federica Zacchini, Marta Czernik , Fiorella di Egidio, Pier Augusto Scapolo, Pasqualino Loi and Grazyna Ptak.	Cellular Reprogramming	montlhy	Mary Anny Liebert srl publisher	USA	2011	[Epub ahead of print]	No
9	-	Cloning the mammoth: a complicated	Pasqualino Loi, Joseph Saragustry	Reproductive Biology and	Book	Springer Science	UK	2013	0000-0000	No

		task or just a dream?	1, Grazyna Ptak	Integrated Conservation Science						
10	-	"Expression and subcellular organization of de novo DNA methyltransferases (Dnmts) in adult and prepubertal oocytes and embryos in sheep." SRF 2010, July 11th-13th, Nottingham, UK	Czernik M, Zacchini F, Di Egidio F, Loi P, Ptak G	Society for the Study of Reproduction - poster presentation	yearly	Society for the Study of Reproduction	Cambridge, UK	2010	0000-0000	No
11	-	Interspecies Somatic Cell Nuclear Transfer: a salvage tool seeking first aid. , in press IF 2.3	Loi P. Jacek A. Modlinski and Grazyna Ptak	Theriogenology, 2011	monthly	Elsevier	USA	2011	Oct;142(4):497-503	No

12	-	“Epigenetic Mechanism in mammals and their effects on cloning procedures,	Pasqualino Loi, Robert Feil & Grazyna Ptak	In: Lost Sex: The evolutionary Biology of Parthenogenesis” Eds. Van Dijk, Martens, Schon Schoen. ”.	Chapter 26,	Springer Press	The Netherlands	2009	559-579	No
13	-	“Expression and subcellular organization of de novo DNA methyltransferases (DNMTs) in adult and prepubertal oocytes and embryos in sheep”	Czernik M, Zacchini F, Toschi P, Fidanza A, Loi P, Ptak G	Conference on "Basic and advanced mammalian reproductive technologies " poster presentation	June 2010 – Jastrzebiec n/Warsaw, Poland	Organized within the Centre of Excellence ANIMBIOGEN in EU–	Poland	2010	0000-0000	No
14	-	Gene expression/phenotypic abnormalities in placental tissues of sheep clones: insurmountable block in cloning progress?	Pasqualino Loi & Grazyna Ptak	in “Epigenetics and Human Health”Ed. Drs Saadi Kochbin & Sophie Rousseaux,	Book Chapter	Springer Press	France	2010	in press	No

15	-	Genomic stability of lyophilized sheep somatic cells before and after nuclear transfer.	Iuso D, Czernik M, Di Egidio F, Sampino S, Zacchini F, Bochenek M, Smorag Z, Modlinski JA, Ptak G, Loi P.	PLoS ONE, IP 4,5	Epub 2013 Jan 8	PLoS ONE; e-pubblication	USA	2013	2013;8(1) :e51317.	Yes
16	-	Somatic cell nuclear transfer using lyophylized cells.	Pasqualino Loi, Domenico Iuso, Marta Czernik, Federica Zacchini & Grazyna Ptak	Methods in Molecular Biology"	In press	Humana Press	233 Spring Street, 6th Floor, New York, NY 10013-1578 USA	2013	0000-0000	No
17	-	Cloning endangered species	Pasqualino Loi1, Jacek Modlinski, 2 & Grazyna Ptak1,2	Principles of Cloning, 2nd Edition	Book	Elsevier	Elsevier/Academic Press Life Sciences 225 Wyman Street Waltham, Massachusetts 02451, USA	2013	0000-0000	No
18	10.1089/cell.2013.0051	A simplified approach for oocyte enucleation in mammalian cloning	Iuso D, Czernik M, Zacchini F, Ptak G, Loi P	Cellular Reprogramming	Quarterly	Bentham	USA		490-94	No
19	10.1016/j.tibtech.20	Towards storage	Loi P, Iuso	Trends in	monthly	Elsevier	USA		688-95	No

	13.09.004	of cells and gametes in dry form	D, Czernik M, Zacchini F, Ptak G.	Biotechnology						
20	10.1371/journal.pone.0051317.	Genomic stability of lyophilized sheep somatic cells before and after nuclear transfer. .	Iuso D, Czernik M, Di Egidio F, Sampino S, Zacchini F, Bochenek M, Smorag Z, Modlinski JA, Ptak G, Loi P.	PLoS ONE	Continuously	PLOS group	USA		1055-8	No
21	10.1371/journal.pone.0033027	Embryonic diapause is conserved across mammals. P	Ptak GE, Tacconi E, Czernik M, Toschi P, Modlinski JA, Loi	PLoS One	2012;7(3): e33027	Public Library of Science	USA	12/03/2012	1-10	Yes
22	10.1016/j.theriogenology.2011.01.016	Interspecies somatic cell nuclear transfer: a salvage tool seeking first aid	Loi P, Modlinski JA, Ptak G	Theriogenology	2011 Jul 15	Elsevier Inc	USA	15/06/2011	217-228	No
23	10.1371/journal.pone.0051317	Genomic stability of lyophilized sheep somatic cells before and after nuclear transfer.	Iuso D, Czernik M, Di Egidio F, Sampino S, Zacchini F, Bochenek M, Smorag	PLoS One	2013;8(1):e51317	Public Library of Science	USA	08/01/2013	1-10	Yes

			Z, Modlinski JA, Ptak G, Loi P							
24	10.1530/REP-11-0063	Genome of non-living cells: trash or recycle?	Loi P, Fulka J Jr, Hildebrandt T, Ptak G.	Reproduction	Oct;142(4):	BioScientifica Ltd.	Cambridge UK	01/10/2011	497-503	Yes
25	-	Efficient production and cellular characterization of sheep androgenetic embryos	Zacchini F, Czernik M, Iuso D, Toschi P, di Egidio F, Scapolo PA, Loi P, Ptak G	Cell Research	Vol. 13	Nature Publishing Group	USA	01/11/2011	495-502	No
26	-	Biological time machines: a realistic approach for cloning an extinct mammal	Loi P, Teruhiko Wakayama, Josef Fulka Jr and Grazyna Ptak.	Endangered Species Research	VOL 14	Inter-Research	Germany	23/09/2011	227-233	Yes
27	10.1093/humrep/des397	Post-implantation mortality of in vitro produced embryos is associated with DNA methyltransferase 1 dysfunction in sheep placenta	Ptak GE, D'Agostino A, Toschi P, Fidanza A, Zacchini F, Czernik M, Monaco F, Loi P	Human Reproduction	Vol. 28	Oxford University Press	Oxford, UK	28/02/2013	298-305	No

28	10.1093/humrep/der454	A short exposure to polychlorinated biphenyls deregulates cellular autophagy in mammalian blastocyst in vitro	Ptak G, Zacchini F, Czernik M, Fidanza A, Palmieri C, Della Salda L, Scapolo PA, Loi P	Human Reproduction	vol. 27,	Oxford University Press	Oxford, UK	27/04/2012	p. 1034-1042,	Yes
29	10.1002/jcb.24310.	Differentiation potential and GFP labeling of sheep bone marrow derived mesenchymal stem cells	Czernik M, Fidanza A, Sardi M, Galli C, Brunetti D, Malatesta D, Della Salda L, Matsukawa K, Ptak G, Loi P	Journal of Cellular Biochemistry	114(1):134-43	Wiley-Liss Inc.	USA	08/01/2013	1-10	No
30	-	Utilization of the Scythe C++ open source library for statistical geocomputation in livestock landscape genomics http://infoscience.epfl.ch/record/181996	S. Stucki, S. Agha, M. Li and S. Joost	Proceedings of the Open Source Geospatial Research and Education Symposium (OGRS)	2	Lulu.com	London	24/10/2012	186-194	Yes

.

A2). Presentations given at conferences

A2.1). Presentations – Talks-plenary lectures

NO.	Type of activities	Main leader	Title	Date	Place	Type of audience	Size of audience	Countries adressed
1	International congress	P07 - UNITE	First International congress on the restoration of endangered and extinct animals	17/05/2010	-	young and senior scientific staff, PhD students, undergraduate students, local Media (Polish television)	100	Poland, Italy, France, Belgium, United Kingdom, Ukraine, Netherlands, Denmark
2	International symposium	P07 - UNITE	The first International congress on Controverises in cryopreservation of stem cells, reproductive cells, tissue and Organs "CRYO"	27/05/2010	-	gynecologists, scientists, technicians	500	Spain, Italy, France, Germany, Poland, Australia, Switzerland, Turkey, Israel, and many more.
3	international meeting	P07 - UNITE	"Born on Ice: an evolution of parenthood"	15/10/2010	-	gynecologist, cryobiologists, developmental biologists	100	USA; Italy, Spain, Israel, UK.

4	International congress	P07 - UNITE	First International congress on the restoration of endangered and extinct animals	17/05/2010	-	young and senior scientific staff, PhD students, undergraduate students, local Media (Polish television)	100	Poland, Italy, France, Belgium, United Kingdom, Ukraine, Netherlands, Denmark
5	invited speaker to the Ovarian Club Meeting: from basic research to practical practice. Barcelona, Spain, Novembre 3-6, 2011.	P07 - UNITE	Freeze Dried Oocytes and "Micro-Cytplasts"	03/11/2011	Barcelona, Spain	Researchers, Obstetricians, Medical Doctors (Genealogists)	150	All Europe, plus Japan, USA, Russia, China
6	Invited speaker to the Ovarian Club: The oocyte, from Basic Research to Clinical Practice.	P07 - UNITE	Nuclear Transfer	06/11/2011	Barcelona, Spain	Researchers, Medical Doctors, gynecologists	150	All Europe, plus Japan, China, USA, Russia
7	TedxLakeComo 2012	P07 - UNITE	- The Noha's Arch of freeze dried cells	10/11/2012	Lake como, Italy	open access meeting, student, scientist, normal people	500	Italy

8	invited talk at Leibniz Institute for Zoo and Wildlife Research (IZW), Berlin	P07 - UNITE	Nuclear transfer of lyophilized cells: work in progress	23/11/2012	Berlin	Scientists	50	Germany
9	Controversies in Cryopreservation of Stem Cells, Reproductive Cells, Tissue & Organs (CRYO) Berlin, Germany, March 21-23, 2013	P07 - UNITE	LEA proteins and desiccation tolerance of mammalian oocytes	22/03/2013	Berlin	Scientists	250	Spain, Italy, France, Germany, Poland, Australia, Switzerland, Turkey, Israel, and many more.
10	Conference	M. Bruford	Farm Animal Genetic Resources	22-23/11/2012	Chile	Scientific community (higher education, Research) - Civil society - Policy makers - Medias	100	-
11	Conference	M. Bruford	Livestock genomic resources and conservation	30/8/2013	Italy	Scientific community (higher education, Research)	50	-
12	Conference	M. Bruford	Farm animal genomics	7/6/2012	Netherlands	Scientific community (higher education, Research)	15	-
13	Conference	M. Bruford	Farm animal genetic resources and global change	19-20/9/2013	Spain	Scientific community (higher education, Research)	100	-
14	Conference	M.	Wildlife and	9-	UK	Scientific	100	-

		Bruford	Domestic animal genomics	12/09/2013		community (higher education, Research)		
15	Conference	P Orozco-terWengel	Investigating the evolutionary history of Ovis aries using next generation sequencing	9-12/09/2013	UK	Scientific community (higher education, Research)	100	-
16	workshop	P Orozco-terWengel	Genomics for livestock conservation	18/07/2013	Hungary	Scientific community (higher education, Research)	20	-
17	Presentation	Università Cattolica del S. Cuore di Piacenza	Introgression of European *Bos taurus* genome in Ugandan taurine and zebuine cattle breeds.	26-30/08/2013	64th Annual Meeting of the European Association for Animal Production EAAP, Nantes (France)	Scientific community (Research) Industry Policy makers	500	Europe
18	Presentation	Università Cattolica del S. Cuore di Piacenza	Introgression of European *Bos taurus* genome in Ugandan taurine and zebuine cattle breeds.	11-13/06/2013	20th ASPA Congress, Bologna (Italy)	Scientific community (Research) Industry Policy makers	300	Italy

19	Presentation	Università Cattolica del S. Cuore di Piacenza	La biodiversità nel comparto agrozootecnico	05/12/2013	"Le produzioni tradizionali: una via indiretta per la salvaguardia del territorio", Portici (Italy)	Scientific community (Research), Policy makers, Medias	100	Italy
20	Presentation	Università Cattolica del S. Cuore di Piacenza	New approaches to investigate the genetic basis of animal adaptation to different environments	27/10/2010	Animal Farming and Environment Interaction in Mediterranean Region, Zadar (Croatia)	Scientific community (Research)	300	Europe
21	Presentation	Università Cattolica del S. Cuore di Piacenza	Livestock genomics: assess the present to understand the past and drive the future of animal breeding	09/03/2011	Special Animal Genetics Seminar, Zurich (Switzerland)	Scientific community (Research)	200	Europe
22	Presentation	Università Cattolica del S. Cuore di Piacenza	Non conventional strategies to identify markers associated with difficult traits	08/02/2010	Consultant meeting on genetic variation on the control of resistance to infectious diseases in small ruminants for improving animal productivity, Vienna (Austria)	Scientific community (Research)	60	South West Asia

23	Presentation	Università Cattolica del S. Cuore di Piacenza	The DNA handbook of livestock history	28/08/2013	Araçatuba (Brazil)	Scientific community (Research)	150	Brazil
24	Presentation	Università Cattolica del S. Cuore di Piacenza	La domesticazione dei bovini - Bovine domestication	13/12/2013	Seminar organized by the Università di Udine, Udine (Italy)	Scientific community (Research)	80	Italy
25	Presentation	Università Cattolica del S. Cuore di Piacenza	Global Consortium on animal genome mapping	12/02/2013	One Day International Seminar, Bogor (Indonesia)	Scientific community (Research)	300	Europe
26	Presentation	Università Cattolica del S. Cuore di Piacenza	Use of molecular data for the characterization of animal genetic resources	17-19/05/2011	FAO Regional Workshop “Characterization, Inventory and Monitoring of Animal Genetic Resources”, Tartu (Estonia)	Scientific community (Research)	150	Europe
27	Presentation	Università Cattolica del S. Cuore di Piacenza	La genomica al servizio del miglioramento genetico dei bovini	18/10/2012	Cremona (Italy)	Scientific community (Research)	400	Italy
28	Presentation	P07 UNITE	Cellular and epigenetic characterization of monoparental (andro/partheno genetic) ovine	2010	Conference on "Basic and advanced mammalian reproductive technologie	Scientific community (Research	-	Warsaw

			blastocysts					
29	Presentation	P07 UNITE	“Retarded placental vascularization of in vitro produced sheep embryos during implantation is associated with pregnancy loss	June 2010	Conference on "Basic and advanced mammalian reproductive technologies" , organized within the Centre of Excellence ANIMBIOGEN in EU	Scientific community (Research	-	Jastrzebiec n/Warsaw, Poland
30	Presentation	P07 UNITE	Expression of DNA methyltransferases 3 family (DNMT3s) in monoparental and biparental sheep embryos	11-13/07/2010	Proceedings of the Society for Reproduction and Fertility. Society for the Study of Reproduction	Scientific community (Research	-	Nottingham, UK
31	Presentation and Poster	P07 UNITE	“Epigenetic immaturity of oocytes and embryos derived from prepubertal sheep” SRF	11-13/07/2010	Annual Meeting of the Society for the Study of Reproduction,	Scientific community (Research	-	Cambridge, UK
32	Presentation	P07 UNITE	Altered expression of DNMT1 and related proteins in sheep placentas from IVP embryos	June 2010	Conference on "Basic and advanced mammalian reproductive technologies"	Scientific community (Research	-	Warsaw, Poland

			during early gestation”		organized within the EU Centre of Excellence ANIMIBIOGEN;			
33	Presentation and Poster	P07 UNITE	“Expression and subcellular organization of de novo DNA methyltransferases (DNMTs) in adult and prepubertal oocytes and embryos in sheep”	June 2010	Conference on "Basic and advanced mammalian reproductive technologies. Organized within the Centre of Excellence ANIMBIOGEN in EU	Scientific community (Research	-	Warsaw, Poland
34	Presentation	P07 UNITE	Lyophilized and rehydrated metaphase ii (mii) ovine chromosomes maintain functionality upon transfer in fresh mii oocytes.	21-24/10/2012	American Association Reproductive Medicine, SRM meeting , San Diego, California	Scientific community (Research	2012	San Diego, California, USA

A2.2). Presentations – Posters

NO.	Type of activities	Main leader	Title	Date	Place	Type of audience	Size of audience	Countries adressed
1	Conference	EMBL-EBI	Poster: The NextGen Project, Identifying Farm Animal Genetic Resources using Whole Genome Data	8/05/2013	Cold Spring Harbor	Scientific community (higher education, Research) - Civil society - Policy makers - Medias	300	-
2	Conference	EMBL-EBI	Poster: The NextGen Project: Whole Genome Data to Optimize Methods to Maintain Farm Animal	3/09/2013	Nottingham University	Scientific community (higher education, Research) - Civil society - Policy makers - Medias	250	
3	Conference	EMBL-EBI	Poster: The NextGen Project: Whole Genome Data to Optimize Methods to Maintain Farm Animal	13/01/2014	San Diego	Scientific community (higher education, Research) - Civil society - Policy makers - Medias	500	-
4	Poster	Università Cattolica del S. Cuore di Piacenza	Analysis of B. taurus and B. indicus admixture in Uganda as revealed by the Illumina BovineSNP50 Genotyping BeadChip.	12-16/01/2013	XXI Plant and Animal Genome Meeting, San Diego, California (USA)	Scientific community (Research)	1500	international
5	Poster	Università Cattolica del S. Cuore di Piacenza	Progetti di cooperazione internazionale per la caratterizzazione e conservazione della biodiversità zootecnica	23/11/2011	Milano (Italy)	Scientific community (Research)	80	Italy

6	Poster	P07 - UNITE	Production of bovine blastocysts by nuclear transfer using freeze-dried fibroblast cells	22–26 July 2013	SSR's 46th Annual Meeting 22–26 July 2013	Scientific community (Research)	-	Montréal, Québec, Canada
7	Poster	P07 - UNITE	Upregulation of autophagy genes (ATGs) in developing chorioallantoic tissues (CA) of embryos produced by Assisted Reproductive Technologies (ART)"	11-1/07/2010	Annual Meeting of the Society for the Study of Reproductio n	Scientific community (Research)	-	Nottingham, UK

The NextGen Project
Next generation methods to preserve farm animal diversity

Laura Clarke, Ian Streeter, Pablo Orozco ter Wengel, Holly Zheng-Bradley, David Richardson, Paul Flicek and The NextGen Consortium,
EMBL-EBI, Cambridge, United Kingdom

Background

NextGen project is an EU FP7 project that aims to develop methods to preserve farm animal biodiversity by optimizing present and future breeding options.

The project is recording data on all major types of genetic variation in the genome of cattle, sheep and goats. These data will aid the design of breeding programs that maximise genetic progress in livestock populations while maintaining diversity, and will aid the choice of animals for biobanking to maximise neutral and functional genetic diversity in stored material.

Country	Animal	Number	Sequencing Strategy
Iran	*Ovis orientalis* (mouflon)	1	High Coverage ~90x
	Capra aegagrus (bezoar)	1	High Coverage ~90x
	Ovis orientalis	15	Low Coverage ~10x
	Ovis aries	20	Low Coverage ~10x
	Capra aegagrus	20	Low Coverage ~10x
	Capra hircus	20	Low Coverage ~10x
Morocco	*Ovis aries*	164	Low Coverage ~10x
	Capra hircus	164	Low Coverage ~10x
Uganda	*Bos taurus*	25	Low Coverage ~10x
	Bos taurus	813	Bovine Illumina SNP50 Chip
	Bos taurus	102	Bovine Illumina HD Chip

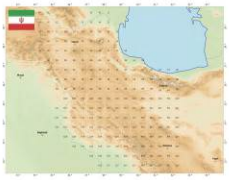
Sampling Strategy

Sheep and goat samples have been collected in Morocco and Iran and cattle samples have been collected in Uganda. Each country has been divided into sectors and 3 animals from different populations are being sampled in each appropriate sector. Cattle are being sampled in Uganda and sheep and goats in Morocco and Iran. The wild type Mouflon and Bezoar are also being sampled in Iran.

UGANDA

MOROCCO

IRAN



GPS coordinates and photos of every animal sampled are also taken to aid with downstream analysis. Tissue biopsies are taken from the external and distal part of the ear of the animal and transferred to a tube containing 70% alcohol. The day after sampling the tissue samples are dried and covered in silica gel for long term preservation

The NextGen Consortium is the work of many partners from Europe and further afield.

CNRS, Grenoble
Cardiff University
Universita Cattolica del Sacro Cuoro, Piacenza
Parco Tenologico Padano
EMBL-EBI
EPFL, Lausanne
Universita Teramo
Makerere Univesity, Uganda
INRA-Morocco
Gorgan University, Iran
CSIRO, Brisbane
Genoscope, France

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2010-2014) under grant agreement n° 244356. ("NextGen")

Aims of NextGen

- Finding signatures of local adaption in sheep, goats and cattle
- Develop methods to identify animals for biobanking
- Assess the potential of breeds from domestication centres as genetic resource for sheep and goats.
- Establish the relevance of wild ancestors as genomic resource for sheep and goats.

The project will carry out high coverage sequencing and assembly of the wild ancestors and low coverage sequencing and variant discovery on a range of animals from the three countries. We anticipate these resources will provide a good framework to allow us to answer these questions and achieve the goals of the project

Genome Assembly

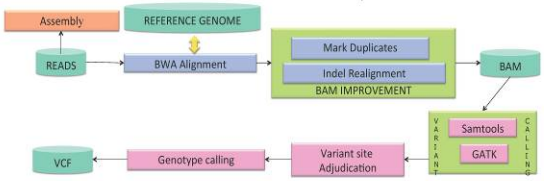
NextGen has produced two assemblies as part of our analysis process. We have generated high coverage sequence data on the *Ovis orientalis* (mouflon) and *Capra aegagrus* (bezoar) animals sampled in Iran. The samples are being sequencing following the ALLPATHS-LG[1] library strategy.

	Ovis orientalis	*Capra aegagrus*
Total length of contigs	2.43 Gb	2.45 Gb
Number of contigs	125,000	102,000
Contig N50	39.8 Kb	52.0 Kb
Total length scaffolds[1]	2.59 Gb	2.58 Gb
Number of scaffolds	6,210	6,676
Scaffold N50	2.23 Mb	1.75 Mb
Ambiguities per 10,000 bases	27.65	17.92

1. Sheep Reference is 2.62Gb, Goat Reference is 2.61 Gb

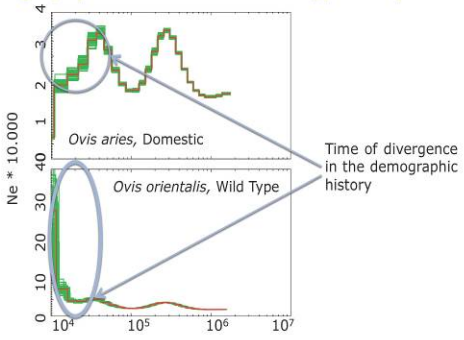
Alignment and Variant Calling

NextGen is using a standard ngs alignment and combined variant calling pipeline for low coverage genomes. We are using BWA to align the reads to the genome followed by variant calling using samtools and GATK and adjudication using GATK's Variant Quality Score Recalibration package (VQSR). So far we have been able to call variants on 8 *Ovis aries* and 8 *Ovis orientalis* samples.



	Iranian *Ovis aries*	Iranian *Ovis orientalis*
# animals	8	8
% reads mapped	96.8 %	96.2 %
Median depth of coverage (sample avg)	12.5	13.0
# SNPs (VQSR, VQSLOD > 4.0)	29.3 M	28.2 M
# Indels (VQSR, VQSLOD > 4.0)	75,000	91,000

Demography of Domestic and Wild Type Sheep



The variant data we have collected so far based in Iranian wild type and domestic samples has been used to infer the demography of the two sheep species using Pairwise Markovian Sequential Coalesence[2]. This model shows us that both species share demographic history until approximately 40,000 years ago when the domestic sheep start to decline but the wild ancestors start to expand. This may have occurred at the onset of population management of the wild animals by early herders

References

1. High-quality draft assemblies of mammalian genomes from massively parallel sequence data, S. Gnerre et al, Proc Natl Acad Sci U S A. 2011 Jan 25;108(4):1513-8. Epub 2010 Dec 27.
2. Inference of Human population history from individual whole genome sequences, H. Li & R. Durbin, Nature, 2011 July 28; 475:493-496

The NextGen Project
Next generation methods to preserve farm animal diversity

Laura Clarke, Ian Streeter, Holly Zheng-Bradley, David Richardson, Paul Flicek and The NextGen Consortium,
EMBL-EBI, Cambridge, United Kingdom

Background

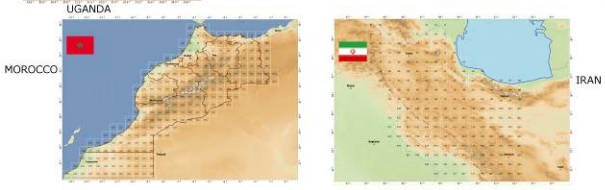
NextGen project is an EU FP7 project which aims to develop methods to preserve farm animal biodiversity by optimizing present and future breeding options.

The project is recording data on all major types of genetic variation in the genome of cattle, sheep and goats. These data will aid the design of breeding programs that maximise genetic progress in livestock populations while maintaining diversity, and will aid the choice of animals for biobanking to maximise neutral and functional genetic diversity in stored material.

Country	Animal	Number	Sequencing Strategy
Iran	*Ovis orientalis* (mouflon)	1	High Coverage ~90x
	Capra aegagrus (bezoar)	1	High Coverage ~90x
	Ovis orientalis	20	Low Coverage ~10x
	Ovis aries	20	Low Coverage ~10x
	Capra aegagrus	20	Low Coverage ~10x
	Capra hircus	20	Low Coverage ~10x
Morocco	*Ovis aries*	164	Low Coverage ~10x
	Capra hircus	164	Low Coverage ~10x
Uganda	*Bos Taurus*	40	Low Coverage ~10x

Sampling Strategy

Sheep and goat samples are being collected in Morocco and Iran and cattle samples are being collected in Uganda. Each country has been divided into sectors and 3 animals from different populations are being sampled in each appropriate sector. Cattle are being sampled in Uganda and sheep and goats in Morrocco and Iran. The wild type Mouflon and Bezoar are also being sampled in Iran.



GPS coordinates and photos of every animal sampled are also taken to aid with downstream analysis. Tissue biopsies are taken from the external and distal part of the ear of the animal and transferred to a tube containing 70% alcohol. The day after sampling the tissue samples are dried and covered in silica gel for long term preservation

The NextGen Consortium is the work of many partners from Europe and further afield.

CNRS, Grenoble
Cardiff University
Universita Cattolica del Sacro Cuoro, Piacenza
Parco Tenologico Padano
EMBL-EBI
EPFL, Lausanne
Universita Teramo
Makerere Univesity, Uganda
INRA-Morocco
Gorgan University, Iran
CSIRO, Brisbane
Genoscope, France

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2010-2014) under grant agreement n° 244356. ("NextGen")

Aims of NextGen

•Finding signatures of local adaption in sheep, goats and cattle
•Develop methods to identify animals for biobanking
•Assess the potential of breeds from domestication centres as genetic resource for sheep and goats.
•Establish the relevance of wild ancestors as genomic resource for sheep and goats.

The project will carry out high coverage sequencing and assembly of the wild ancestors and low coverage sequencing and variant discovery on a range of animals from the three countries. We anticipate these resources will provide a good framework to allow us to answer these questions and achieve the goals of the project

Genome Assembly

NextGen is aiming to produce two assemblies as part of our analysis process. We have generated high coverage sequence data on the *Ovis orientalis* (mouflon) and *Capra aegagrus* (bezoar) animals sampled in Iran. The samples are being sequencing following the ALLPATHS-LG[1] library strategy.

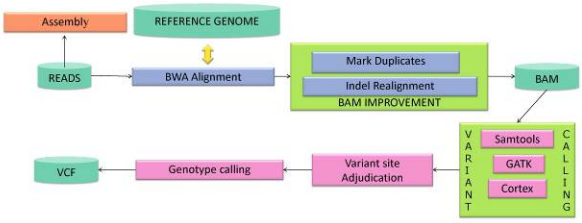
	Bezoar	Mouflon
Total length contigs	2.24 Gb	2.05 Gb
Contig N50	28.9 kb	18.1 kb
Total length scaffolds[1]	2.45 Gb	2.46 Gb
Scaffold N50	141 kb	109 kb
SNP rate	1/455	1/271

1. OARv3.1 is 2.62GB

These are statistics for our preliminary assemblies from ALLPATHS-LG. Genoscope is carrying out sequencing of additional long range libraries which we hope will improve the scaffold length

Alignment and Variant Calling

NextGen is going to use a standard NGS alignment and variant calling pipeline for all the low coverage genomes. A joint calling strategy is going to be used where all the animals from the same species and country will be called together.



Our alignment and variant calling pipeline will use BWA alignments to the appropriate reference genome. The resulting alignments will undergo BAM improvement where reads around indels are realigned to correct for mapping errors and duplicates are marked. These BAM files are then used in our variant calling pipeline. We call variants at the population level using information for all animals from the same country and species to call variants at the same time. We will use multiple different calling algorithms to produce our initial sites list including SAMtools mpileup, the GATK unified genotyper and Cortex a De Brujin graph assembly based variant detection tool. Once a union of sites is created we can run adjudication tools like GATK's VQSR and produce a final sites list to genotype all of our animals in. Our variant calls will all be distributed in VCF format.

Once we have defined our variant data set we hope to:

- Characterise the demographic history of the sheep and goats over both long and short time scales using methods like pairwise sequential Markovian coalescence and Approximate Bayesian Computation (ABC).
- Use information from the underlying demography to make coalescent simulations that can be used to define the distribution of expected values of statistical tests for the detection of natural selection

References

1. High-quality draft assemblies of mammalian genomes from massively parallel sequence data, S. Gnerre et al, Proc Natl Acad Sci U S A. 2011 Jan 25;108(4):1513-8. Epub 2010 Dec 27.

Centro di Ateneo per la Solidarietà Internazionale - CeSI

UCSC ExpoLAB

Progetti di cooperazione internazionale per la caratterizzazione e conservazione della biodiversità zootecnica

Colli L.*[1], Negrini R.[1], Nicolazzi E.L.[1,2], Milanesi M.[1], Eufemi E.[1], Mazza R.[3], Bacciu N.[1], Bomba L.[1], Murelli E.[1], Ajmone-Marsan P.[1]

1 Istituto di Zootecnica & Centro di Ricerca sulla Biodiversità ed il DNA Antico BioDNA, Facoltà di Agraria, Università Cattolica del S. Cuore, via Emilia Parmense 84, 29122 Piacenza; *Email: licia.colli@unicatt.it
2 CRSA, Consorzio di Ricerca e Sperimentazione degli Allevatori, Associazione Italiana Allevatori (AIA), Roma
3 Laboratorio Genetica e Servizi LGS, Associazione Italiana Allevatori (AIA), Cremona

Istituto di Zootecnica

BioDNA

Il progetto GLOBALDIV

GLOBALDIV "*A global view of livestock biodiversity and conservation*" è un progetto triennale (2007-2010, esteso al 2011) finanziato dalla Commissione Europea (AGRI GEN RES 067, Council Regulation (EC) No 870/2004; www.globaldiv.eu), coordinato dall'Istituto di Zootecnica della Facoltà di Agraria della UCSC di Piacenza.

Obiettivo del progetto: valutare lo stato attuale e le tendenze di fenomeni quali la perdita di biodiversità e l'erosione genetica nelle specie di animali domestici e di proporre una sintesi delle principali strategie adottate per la conservazione di tali risorse genetiche animali.

Partecipanti: i 7 partner insieme ad un gruppo internazionale costituito da 32 esperti hanno contribuito ad analizzare le criticità dalla situazione attuale e le potenzialità delle tecnologie disponibili e dei parametri socio-economici utili ad individuare le strategie ottimali per un'efficiente conservazione della diversità delle risorse genetiche animali a fronte di limitate risorse disponibili.

Il team di esperti internazionali di Globaldiv.

Poiché la perdita di biodiversità zootecnica è un fenomeno che riguarda tutto il pianeta, per il buon esito di Globaldiv sono stati coinvolti ricercatori provenienti da **Europa, Asia, Africa e Sud America** (Universidade Estadual Paulista, Brasile), insieme a numerosi **Enti ed Agenzie internazionali**, quali **l'ILRI International Livestock Research Institute** con sede in Kenia e in Cina, la **FAO Food and Agriculture Organization**, **l'EAAP European Association for Animal Production** e la **WAAP World Association of Animal Production**.

Divulgazione dei risultati: i risultati del lavoro del team di esperti di Globaldiv sono stati divulgati tramite numerosi mezzi di comunicazione:

Globaldiv Summer School 2010.

- **Summer School:** le tre Scuole Estive organizzate da Globaldiv hanno contribuito alla formazione di 109 giovani ricercatori e dottorandi provenienti da 45 diversi paesi e da tre continenti, anche grazie all'erogazione di borse di studio per studenti meritevoli, favorendo la partecipazione di candidati da paesi emergenti e in via di sviluppo. I contenuti delle 94 lezioni tenute dai docenti delle scuole estive e dei workshop di Globaldiv sono disponibili sul sito del progetto.

Il numero speciale di Animal Genetics dedicato a Globaldiv.

- **Pubblicazioni scientifiche:** un numero speciale delle rivista Animal Genetics è stato dedicato alla pubblicazione delle review scritte dagli esperti di Globaldiv. Per facilitarne al massimo la diffusione, il progetto ha supportato la pubblicazione ad accesso libero della versione elettronica degli articoli.

Workshop di Globaldiv.

- **Workshop e conferenze:** due workshop e una conferenza finale hanno contribuito a diffondere i risultati di Globaldiv anche alla comunità scientifica esterna all'ambito delle risorse genetiche animali e a tutti gli operatori coinvolti a vario livello nel processo di "decision making" del settore.

- **Newsletter:** i 19 numeri della newsletter sono stati indirizzati ad un pubblico ancora più ampio che è stato raggiunto attraverso il sito web (≈67000 downloads dal 2007) del progetto e le mailing list di FAO ed EAAP.

Le newsletter pubblicate sul sito web di Globaldiv.

Il progetto NEXTGEN

Nextgen "*Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options*" (http://nextgen.epfl.ch/) è un progetto di ricerca quadriennale finanziato dalla Commissione Europea (KBBE-2009-1-1-03: Optimization of methods to maintain farm animal biodiversity) attualmente in corso e che si concluderà nel 2014.

Obiettivi del progetto: 1) impiego delle tecnologie di nuove generazione per il sequenziamento completo dei genomi al fine di per la caratterizzare le risorse genetiche di bovini, capre e pecore in Marocco, Iran e Uganda; 2) sviluppare e proporre metodologie ottimizzate ed economicamente vantaggiose per preservare la biodiversità di questi animali domestici.

Marocco, Iran e Uganda costituiranno casi-studio ideali poiché rappresentano aree di importanza strategica per lo studio della diversità genetica di queste specie e dell'adattamento sviluppato da alcune razze di bovini in aree colpite da malattie endemiche del bestiame (Uganda).

Suddivisione dei territori di Marocco (sinistra) e Uganda (destra) in base a cui verrà effettuato il campionamento di capre, pecore e bovini previsto da Nextgen.

Partecipanti: le attività previste da Nextgen saranno rese possibili dal **network di collaboratori internazionali** coordinato dal **CNRS Centre National de la Recherche Scientifique** (Francia) e che coinvolge cinque Università ed Enti di Ricerca europei tra cui **UCSC di Piacenza**, insieme all'**Università di Makerere** (Uganda), **l'INRA Institut National de la Recherche Agronomique** (Marocco), la **Gorgan University of Agriculture and Natural Resources** (Iran), insieme a **EMBL-EBI European Molecular Biology Laboratory – European Bioinformatics Institute** (Germania) e **CSIRO Commonwealth Scientific and Industrial Research Organisation** (Australia).

Il consorzio è stato costituito in modo da permettere il raggiungimento degli obiettivi di Nextgen attraverso l'interazione multidisciplinare di ricercatori esperti in genetica della conservazione, bioinformatica, crioconservazione del materiale biologico, tecnologie di allevamento e sistemi di analisi geografica (GIScience).

UNIVERSITÀ CATTOLICA del Sacro Cuore

Convegno "Nutrire il pianeta: l'Università Cattolica e la cooperazione internazionale" – 23 Novembre 2011

A3). Origination of seminars; symposia

NO	Type of activities	Main leader	Title	Date	Place	Type of audience	Size of audience	Countries adressed
1	Seminar	P07 - UNITE	Nuclear transfer and artificial gametes	15/05/2010		Undergraduate students	20	Italy
2	Workshop	Università Cattolica del S. Cuore di Piacenza	Nextgen training workshop	24-26/01/2011	Makerere (Uganda)	Scientific community (Research)	60	Uganda
3	Workshop	Università Cattolica del S. Cuore di Piacenza	Nextgen training workshop	19-21/04/2010	Mekness (Morocco)	Scientific community (Research)	45	Morocco

A4). PhD thesis defense

“Développement d'outils de géo-calcul haute performance pour l'identification de régions du génome potentiellement soumises à la sélection naturelle - analyse spatiale de la diversité de panels de polymorphismes nucléotidiques à haute densité (800k) chez Bos taurus et B. indicus en Ouganda“.
Sylvie Stucki, supervised by Dr. Stéphane Joost, Lausanne (Switzerland), 28 02 2014
Partner responsible: P06 EPFL
Type of audience: Scientific community.

“Biodiversité et relations génome-environnementchez les petits ruminants“.
Badr Benjelloun, supervised by Dr. Pierre Taberlet and Dr. François Pompanon, Grenoble (France), March 2015.
Partner responsible: CO01 CNRS
Type of audience: Scientific community.

“Title to be defined yet”.
The thesis will focus in part on Nextgen project results.
Elia Vajana, supervised by Dr. Paolo Ajmone-Marsan and Dr. Licia Colli, Piacenza (Italy), 2016.
Partner responsible: P03 UNICATT
Type of audience: Scientific community.

“Patterns of genetic diversity and adaptive selection among indigenous cattle populations of the different agro-ecological zones of Uganda”
Fredrick Kabi, supervised by Dr. Vincent Muwanika and Dr. Charles Masembe, Makerere (Uganda), date to be determined.
Partner responsible: P08 MAK
Type of audience: Scientific community.

A5). Website

- Designed and managed by P06 EPFL since the beginning of the NEXTGEN project.
- Allows an external visibility of the Project to anybody
- Can be reached at http://nextgen.epfl.ch/
- Upgraded all along the project and a web link towards other web pages.

Website statistics


NextGen project | EPFL

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

YOU ARE | BY SCHOOL | ABOUT EPFL | Directory

EPFL > NextGen project English

NEXT GENERATION METHODS TO PRESERVE FARM ANIMAL BIODIVERSITY NEXTGEN

Involved laboratories | Material for students | Dissemination | Deliverables | Data sets

Share | Print

nextgen

SEVENTH FRAMEWORK PROGRAMME

EU-funded project

KBBE-2009-1-1-03: Optimisation of methods to maintain farm animal biodiversity

Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options

NEXTGEN is the first project in the area of conservation genetics that proposes a comparative analysis of **whole genome data** at the intraspecific level. Therefore, the project will gather data on an unprecedented scale on all major types of genetic variation in the genome of **cattle**, **sheep** and **goats**. A high impact is expected, far beyond the farm animal scientific community mainly on conservation and evolutionary biology. More specifically, NEXTGEN will provide precise methodology for studying the biodiversity aspect of **disease resistance** and the relationships between genome and environment (**landscape genomics**).

NEXTGEN aims to provide the necessary tools for the exploitation of new generation genomic and reproductive technologies for Farm Animal Genetic Resources (FAnGR) characterization and conservation. A bioinformatics pipeline will be set to take advantage of whole genome sequencing and large scale marker genotyping, unprecedented in livestock species. Breeding programmes will be designed to exploit whole genome data to maximise genetic progress in livestock populations while maintaining diversity. Strategies will be developed for the optimal choice of animals for biobanking, to maximise neutral and functional genetic diversity in stored material.

New technologies offer opportunities to "think differently". Accordingly, NEXTGEN proposes three very innovative approaches that will represent a breakthrough in the characterization, valuation and conservation of genetic resources, including potential application in industrial breeds.

- Powerful spatially-explicit analysis of genome diversity will enable the identification of genomic regions associated with adaptation and disease resistance, key traits for sustainable breeding that are very difficult to investigate by linkage or association studies in natural or experimental populations. NEXTGEN establishes this new concept in two case studies: in Morocco, where adaptation in sheep and goats will be investigated in a region showing marked variation in environmental conditions, and in Uganda, where vector born diseases in well defined endemic areas have greatly affected cattle and the livelihood of local farmers.
- A novel freeze-drying approach proposed for bio-banking cells and gametes at room temperature and hence at very low cost.
- An evaluation of the potential of wild ancestors (sheep and goat) as reservoirs of genetic diversity for the respective domestic species. A case study will be carried out for sheep and goats in North-Western Iran.

The potential of NEXTGEN is maximised by the collation of an international and interdisciplinary team of top level researchers that are able to work across the borders of disciplines and rigorously explore these new ideas.

News and highlights

Nextgen final meeting in Piacenza, Italy, march 26-28, 2014
Nextgen meeting in Piacenza, Italy, October 16-18, 2013
Nextgen meeting in Pinsot, France, April 8-10, 2013
Nextgen meeting in Pinsot, France, April 2-4, 2012
Nextgen in Pinsot, France, September 19-21, 2011
Second Nextgen meeting in Piacenza, Italy, April 4-6, 2011.
Second training session in Kampala, Uganda, January 24-26, 2011.
Field trip (sampling) & 1st training session in Morocco, April 13-15, 2010
Nextgen kick-off meeting in Meknès (Morocco), April 12-13, 2010,

LINKS

The story of NEXTGEN on Horizon 2020 website
Saving animal DNA for future generations
EU FP5 Econogene
EU AGRI GEN RES Globaldiv
EU AGRI GEN RES Eureca
EU FP7 CONGRESS
ESF RNP GENOMIC-RESOURCES

FOR PARTNERS ONLY

Access to the Intranet
Sampling monitoring

http://nextgen.epfl.ch/[03/07/2014 11:05:06]

NEXT GENERATION METHODS TO PRESERVE FARM ANIMAL BIODIVERSITY NEXTGEN

Involved laboratories

Centre National de la Recherche Scientifique (CNRS)
Lab. d'Ecologie Alpine, UJF-CNRS, Grenoble, France
Dr. Pierre Taberlet, coordinator

Cardiff University (UNICAR)
Organisms and Environment Group
School of Biosciences, Cardiff, UK
Prof. Michael W. Bruford

Università Cattolica del Sacro Cuore Piacenza (UNICATT)
Laboratory of Animal Genetics, Piacenza, Italy
Dr. Riccardo Negrini

Parco Tecnologico Padano (PTP)
Polo Universitario, Lodi, Italy
Dr. Alessandra Stella

European Molecular Biology Laboratory (EMBL-EBI)
European Bioinformatics Institute, Cambridge, UK
Dr. Paul Flicek

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lab. of Geographic Information Systems (LaSIG), Lausanne, Switzerland
Dr. Stéphane Joost

Università Teramo (UNITE)
Laboratory of Experimental Embryology
Dpt of Comparative Biomedical Sciences, Teramo, Italy
Prof. Pasqualino Loi

Makerere University (MAK)
Institute of Environment & Natural Resources (MUIENR)
Kampala, Uganda
Dr. Vincent B. Muwanika

Institut National de la Recherche Agronomique (INRA-Mor)
Rabat R.P. Maroc
Dr. Aziz Fadlaoui

Gorgan University of Agriculture and Natural Resources (GAU)
Department of Environmental Sciences
Gorgan, Iran

News and highlights

Nextgen final meeting in Piacenza, Italy, march 26-28, 2014
Nextgen meeting in Piacenza, Italy, October 16-18, 2013
Nextgen meeting in Pinsot, France, April 8-10, 2013
Nextgen meeting in Pinsot, France, April 2-4, 2012
Nextgen in Pinsot, France, September 19-21, 2011
Second Nextgen meeting in Piacenza, Italy, April 4-6, 2011.
Second training session in Kampala, Uganda, January 24-26, 2011.
Field trip (sampling) & 1st training session in Morocco, April 13-15, 2010
Nextgen kick-off meeting in Meknès (Morocco), April 12-13, 2010,

LINKS

The story of NEXTGEN on Horizon 2020 website
Saving animal DNA for future generations

EU FP5 Econogene

EU AGRI GEN RES Globaldiv

EU AGRI GEN RES Eureca

EU FP7 CONGRESS

ESF RNP GENOMIC-RESOURCES

FOR PARTNERS ONLY

Access to the Intranet
Sampling monitoring

Since May 24th 2013	
Visits	1150
Visitors	658
# Countries	76
Average time	1min22
# Pages/visit	1.33

Visitors/continents	
Europe	814
Asia	138
Americas	90
Africa	88
Oceania	18

Number of visits

Time

From May 24th, 2013 to July 2nd 2014

Visits/First 10 Countries

1.	United Kingdom	327 (28,43 %)
2.	France	137 (11,91 %)
3.	Italy	124 (10,78 %)
4.	Switzerland	72 (6,26 %)
5.	United States	56 (4,87 %)
6.	Uganda	50 (4,35 %)
7.	India	32 (2,78 %)
8.	Spain	26 (2,26 %)
9.	Malaysia	25 (2,17 %)
10.	Belgium	22 (1,91 %)

Visits/ First 25 Cities

2.	Grenoble	72 (6,26 %)
3.	Bristol	62 (5,39 %)
4.	Milan	54 (4,70 %)
5.	(not set)	52 (4,52 %)
6.	Kampala	39 (3,39 %)
7.	Lodi	31 (2,70 %)
8.	Cambridge	30 (2,61 %)
9.	Ecublens	26 (2,26 %)
10.	Paris	17 (1,48 %)
11.	Didcot	15 (1,30 %)
12.	Brussels	14 (1,22 %)
13.	Lausanne	14 (1,22 %)
14.	Kuala Lumpur	13 (1,13 %)
15.	Kota Kinabalu	11 (0,96 %)
16.	Zurich	9 (0,78 %)
17.	Depok	9 (0,78 %)
18.	London	8 (0,70 %)
19.	Athens	8 (0,70 %)
20.	Nairobi	8 (0,70 %)

SECTION B (CONFIDENTIAL)

B) CONFIDENTIAL OR RESTRICTED PROJECT OUTPUTS

B1). Patents, Trademarks, Registered Designs, etc

List of applications for patents, trademarks, registered designs, etc.			
Type of IP Rights: Patents, Trademarks, Registered designs, Utility models, etc.	**Application reference(s) (e.g. EP123456)**	**Subject or title of application**	**Applicant (s) (as on the application)**
None			

B2). Exploitable foreground

NO	Type of exploitable foreground	Exploitable Foreground (description)	Confidential YES/NO	Foreseen embargo date Syntax: dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable for commercial use or any other use	Patents or other IPR exploitation (licenses)	Owner & Other Beneficiary(s) involved
1	Variant call set	Iranian Ovis orientalis, 14 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
2	Variant call set	Moroccan Ovis aries, 160 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
3	Variant call set	Iranian Ovis aries, 20 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
4	Variant call set	Iranian Ovis aries, 18 samples, ovineSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
5	Variant call set	Moroccan Ovis aries, 30 samples, ovineSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
6	Variant call	Iranian Ovis	NO	30/09/2014	Vcf file and	Livestock	Free	-	NEXTGEN

	set	orientalis, 8 samples, ovineSNP50 genotyping bead chip			ped file	genomics	dissemination		project consortium
7	Variant call set	Iranian Ovis vignei, 6 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
8	Variant call set	Iranian Ovis orientalis, 19 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
9	Variant call set	Moroccan Ovis orientalis, 160 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
10	Variant call set	Iranian Capra aegagrus, 18 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
11	Variant call set	Iranian Capra hircus, 20 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
12	Variant call	Moroccan	NO	30/09/2014	Vcf file and	Livestock	Free	-	NEXTGEN

	set	Capra hircus, 161 samples, variant site discovery			ped file	genomics	dissemination		project consortium
13	Variant call set	Moroccan Capra hircus, 30 samples, goatSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
14	Variant call set	Iranian Capra aegagrus, 7 samples, goatSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
15	Variant call set	Iranian Capra hircus, 9 samples, goatSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
16	Variant call set	Iranian Capra hircus, 20 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
17	Variant call set	Iranian Capra hircus, 5 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
18	Variant call	Moroccan	NO	30/09/2014	Vcf file and	Livestock	Free	-	NEXTGEN

	set	Capra hircus, 161 samples, genotype calls at SNP sites			ped file	genomics	dissemination		project consortium
19	Variant call set	Iranian Capra aegagrus, 22 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
20	Variant call set	Iranian Bos taurus, 8 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
21	Variant call set	Ugandan Bos taurus, 25 samples, variant site discovery	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
22	Variant call set	Ugandan Bos taurus, 102 samples, bovineHD genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
23	Variant call set	Ugandan Bos taurus, 813 samples, bovineSNP50 genotyping bead chip	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
24	Variant call set	Iranian Bos taurus, 8	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project

		samples, genotype calls at SNP sites							consortium
25	Variant call set	Ugandan Bos taurus, 25 samples, genotype calls at SNP sites	NO	30/09/2014	Vcf file and ped file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
26	Genome assembly	Iranian Capra aegagrus, de novo assembly	NO	30/09/2014	Fasta files and agp file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium
27	Genome assembly	Iranian Ovis orientalis, de novo assembly	NO	30/09/2014	Fasta files and agp file	Livestock genomics	Free dissemination	-	NEXTGEN project consortium

B4). Consortium meetings

NO	Title	Date and Venue	Partner Responsible	Type of Meeting
1	Kick Off meeting	13-14 April 2010, Meknes (Morocco)	All	consortium meeting
2	Scientific meeting	July 8-9 2010 Cambridge (UK)	All	consortium meeting
3	General meeting	5-6 April 2011, Piacenza (Italy)	All	consortium meeting
4	Scientific meeting	20-21 September 2011 Pinsot (France)	All	consortium meeting
5	General meeting	3-4 April 2012, Pinsot (France)	All	consortium meeting
6	Bioinformatics meeting	3rd July 2012, Hinxton (UK)	All	consortium meeting
7	General meeting	9-10 April 2013, Pinsot (France)	All	consortium meeting
8	Scientific Meeting	24-26 April 2013 Piacenza (Italy)	All	consortium meeting
9	Scientific Meeting	13 February 2014 Grenoble	All	consortium meeting
10	Final Meeting	24-26 March 2014 Piacenza (Italy)	All	consortium meeting

.

3 Report on societal implications

A General Information *(completed automatically when **Grant Agreement number** is entered.*

Grant Agreement Number:	244356
Title of Project:	Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options.
Name and Title of Coordinator:	Dr Pierre Taberlet

B Ethics

1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	Yes ***Yes***
2. Please indicate whether your project involved any of the following issues (tick box) :	***YES***
RESEARCH ON HUMANS	
• Did the project involve children?	No
• Did the project involve patients?	No
• Did the project involve persons not able to give consent?	No
• Did the project involve adult healthy volunteers?	No
• Did the project involve Human genetic material?	No
• Did the project involve Human biological samples?	No
• Did the project involve Human data collection?	No
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	No
• Did the project involve Human Foetal Tissue / Cells?	No
• Did the project involve Human Embryonic Stem Cells (hESCs)?	No
• Did the project on human Embryonic Stem Cells involve cells in culture?	No
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	No
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	No
• Did the project involve tracking the location or observation of people?	
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	Yes
• Were those animals transgenic small laboratory animals?	No
• Were those animals transgenic farm animals?	No
• Were those animals cloned farm animals?	No
• Were those animals non-human primates?	No
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	Yes

• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	Yes
DUAL USE	
• Research having direct military use	No
• Research having the potential for terrorist abuse	No

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator		1
Work package leaders	2	11
Experienced researchers (i.e. PhD holders)	7	18
PhD Students	1	3
Other	4	8

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	**9**
Of which, indicate the number of men:	6

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?

	Yes

6. Which of the following actions did you carry out and how effective were they?

		Not at all effective				Very effective
X	Design and implement an equal opportunity policy	○	○	X	○	○
❑	Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑	Organise conferences and workshops on gender	○	○	○	○	○
❑	Actions to improve work-life balance	○	○	○	○	○
○	Other:					

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

○ Yes- please specify

X No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

○ Yes- please specify

X No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

X Yes- please specify : Online content of lectures during open training workshops

○ No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

X Main discipline[3]: 1.5

X Associated discipline[Erreur ! Signet non défini.]: 4.1

○ Associated discipline[Erreur ! Signet non défini.]:

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	○ Yes X No

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

○ No

○ Yes- in determining what research should be performed

○ Yes - in implementing the research

○ Yes, in communicating /disseminating / using the results of the project

[3] Insert number from list below (Frascati Manual).

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	○ ○	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

- ○ No
- ○ Yes- in framing the research agenda
- ○ Yes - in implementing the research agenda
- ○ Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

- ○ Yes – as a **primary** objective (please indicate areas below- multiple answers possible)
- ○ Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)
- ○ No

13b If Yes, in which fields?

Agriculture		Energy		Human rights	
Audiovisual and Media		Enlargement		Information Society	
Budget		Enterprise		Institutional affairs	
Competition		Environment		Internal Market	
Consumers		External Relations		Justice, freedom and security	
Culture		External Trade		Public Health	
Customs		Fisheries and Maritime Affairs		Regional Policy	
Development Economic and Monetary Affairs		Food Safety		Research and Innovation	
		Foreign and Security Policy		Space	
Education, Training, Youth		Fraud		Taxation	
Employment and Social Affairs		Humanitarian aid		Transport	

13c If Yes, at which level?

- ○ Local / regional levels
- ○ National level
- ○ European level
- ○ International level

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?	**29**
To how many of these is open access[4] provided?	**8**
How many of these are published in open access journals?	**5**
How many of these are published in open repositories?	**3**
To how many of these is open access not provided?	**21**
Please check all applicable reasons for not providing open access:	
X publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal ❑ lack of time and resources ❑ lack of information on open access ❑ other[5]: ……	

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		**None**
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	**None**
	Registered design	**None**
	Other	**None**
17. How many spin-off companies were created / are planned as a direct result of the project?		**None**
Indicate the approximate number of additional jobs in these companies:		

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

❑	Increase in employment, or	❑	In small & medium-sized enterprises
❑	Safeguard employment, or	❑	In large companies
❑	Decrease in employment,	X	None of the above / not relevant to the project
❑	Difficult to estimate / not possible to quantify		

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:*
Difficult to estimate / not possible to quantify	*0* ❑

[4] Open Access is defined as free of charge access for anyone via Internet.
[5] For instance: classification for security project.

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

○ Yes X No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

○ Yes X No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

❑	Press Release	❑	Coverage in specialist press
❑	Media briefing	❑	Coverage in general (non-specialist) press
❑	TV coverage / report	❑	Coverage in national press
X	Radio coverage / report	❑	Coverage in international press
X	Brochures /posters / flyers	X	Website for the general public / internet
❑	DVD /Film /Multimedia	❑	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

❑	Language of the coordinator	X	English
X	Other language(s)		

***Question F-10*:** Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES

1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]

1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)

1.3 Chemical sciences (chemistry, other allied subjects)

1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)

1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]