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

Berichterstattung

Final Report Summary - NEXTGEN (Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options)

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 (Samada 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.

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.

Project Results:
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 SamBada is open source and can be freely downloaded (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 km²; 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.

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).

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.

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).

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.

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.

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 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.

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 Samāda – software to carry out landscape genomic analyses

The software Samāda 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. Samāda implements univariate and multivariate models to predict the distribution of polymorphic variants on the basis of environmental variables. Additionally, Samāda 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. Samāda 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 Samāda 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
The preliminary correlative analyses with Sam\textsuperscript{2}ada 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.

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 moufflons (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 moufflons and bezoars from Iran.

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).

Sheep and moufflons

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 moufflons 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 led 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 standpoint 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
Bioinformatics, 25, 1754-1760.

Potential Impact:
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 undoubtedly 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 enhance 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 involvement 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 resistance 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, traditional 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 governmental and non-governmental 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 of 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 throughput technologies at the upper limit of those available to generate an enormous amount of molecular information made accessible to the scientific community through 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 pioneered 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 within 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.

List of Websites:
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/