PUBLISHABLE SUMMARY

Summary of the context and overall objectives of the project (For the final period, include the conclusions of the action)

The future of a sustainable dairy industry is dependent on increased efficiency of production per cow. Economic analysis shows that profitability of dairy enterprises increases with increased longevity and fertility, however diseases as well as adaptation to different environmental conditions are heavily affecting the average life span in cattle. Dealing with these interlinked problems requires genetic selection for more fertile and robust and so long-lasting cows.

The first objective was for the fellow to obtain intensive training on using diverse bioinformatic methodologies (ranging from genome alignments to detection of evolutionary conserved elements) enabling two main comparative genomics projects to be carried out: i) enrichment for conserved non-coding elements and transcription factor binding sites in ruminant genomes, ii) whole genome resequencing of two native Russian cattle breeds.

The aim, in the first project, was to explore the potential role of conserved non coding elements in ruminants' gene regulation and phenotypic traits consequently affected. Genomic comparisons of transcription factor binding sites enrichment pattern, among several mammalian species, was performed, and few ruminants specific motifs were predicted to be in regulatory domain of genes involved in reproduction.

In the second project, two native Russian cattle breeds, very well adapted to harsh (cold) climate and also resistant to various pathogens, were deeply investigated. Whole genome resequencing of Yakut and Kholmogory Russian native cattle breeds was performed in order to identify signature of selection in genomic regions potentially responsible for adaptation to harsh climate as well as disease resistant. Detection of genomic region under selection pinpointed potential causative SNPs which might be used further as means for adaptation to harsh climates. In addition, they can be used to improve other European breeds adaptation to the "fast acting" climate change, as well as disease resistance, leading to long lasting cows.

Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far (For the final period please include an overview of the results and their exploitation and dissemination)

Objective 1: Bioinformatics training

During the first year the fellow has been learning both bash and python programming languages, and she obtained training on using bioinformatic tools ranging from genome alignments to detection of evolutionary conserved elements. This Marie Curie fellowship has allowed the fellow to acquire new skills and become confident in using a diverse range of bioinformatic tools and she also had the opportunity to analyse big data sets and perform various analysis, such as genome alignments (NGS – next generation sequencing), SNP detection and annotation.

Objective 2: Data preparation

The fellow was able to prepare databases of phenotypic and genomic data on different cattle population to downstream analysis like detection of conserved elements, multiple genome alignment, SNP detection. She became confident in accessing data from different repositories (NCBI, UCSC, Ensembl), and in modifying the data format according to software requirements.

Objective 3: Data Analysis & Comparative genomics

i) enrichment for conserved non-coding elements and transcription factor binding sites in ruminant genomes.

A total of 28 mammalian species, obtained from public repositories, were selected and divided into three clades: Ruminants (10 species), Cetartiodactyls (5 species), and other Mammals (13 species). Clade-specific conserved non coding elements (CNEs) were defined from the multiple genome alignment . A transcription factor binding sites (TFBSs) scan was performed predicting over 900 million of TFBSs along the genome. Analysis of their potential involvement in the gene regulation was performed focussing on TFBSs predicted to be in clade-specific regulatory domains overlapping CNEs. Comparisons of TFBSs in CNEs in regulatory domains among the three clades highlighted 56, 15, and 5 unique motifs in mammals, cetartiodactyls, and ruminants, respectively. GO analysis of clade specific regulatory domains highlighted the ancestral nature of the mammalian CNEs, as well as ruminants only GO terms enrichment. Gene regulation/TFBSs pattern in the ruminants clade might have evolved a different TFBSs patterns if compared to their ancestral mammalian clade. However, expression data on many individuals are indeed needed in order to better understand and explore their potential role in gene regulation and phenotypic traits affected.

ii) whole genome resequencing of two native Russian cattle breeds.

A total of 20 individuals of two native Russian cattle breeds (Yakut and Kholmogory) were resequenced and the total number of variants identified following GATK Best Practices workflow, were >25 million. Population history analysis have revealed that the ancestral populations split around 100 years ago in two with effective population size of about 3000 and 8000 in Yakut and Kholmogory, respectively. In an attempt to identify genomic region under selection two different tools (hapflk and DCMS) were used and the combination will enable higher power in detecting genomic regions under selection. Moreover, potentially causative variants present in those selective regions might be further used to predict phenotypic trait potentially associated to harsh climate and disease resistance.

Objective 4: Dissemination and public engagement

The researched attended ISAG conference on 2017 with both oral and poster presentation: Transcription factor binding sites enrichment in ruminant and cetartiodactyl specific conserved non-coding elements. Laura Buggiotti*, Marta Farrè, and Denis Larkin, Royal Veterinary College, London, United Kingdom. Proceedings of the 36th International Conference on Animal Genetics, Dublin, Ireland (2017).

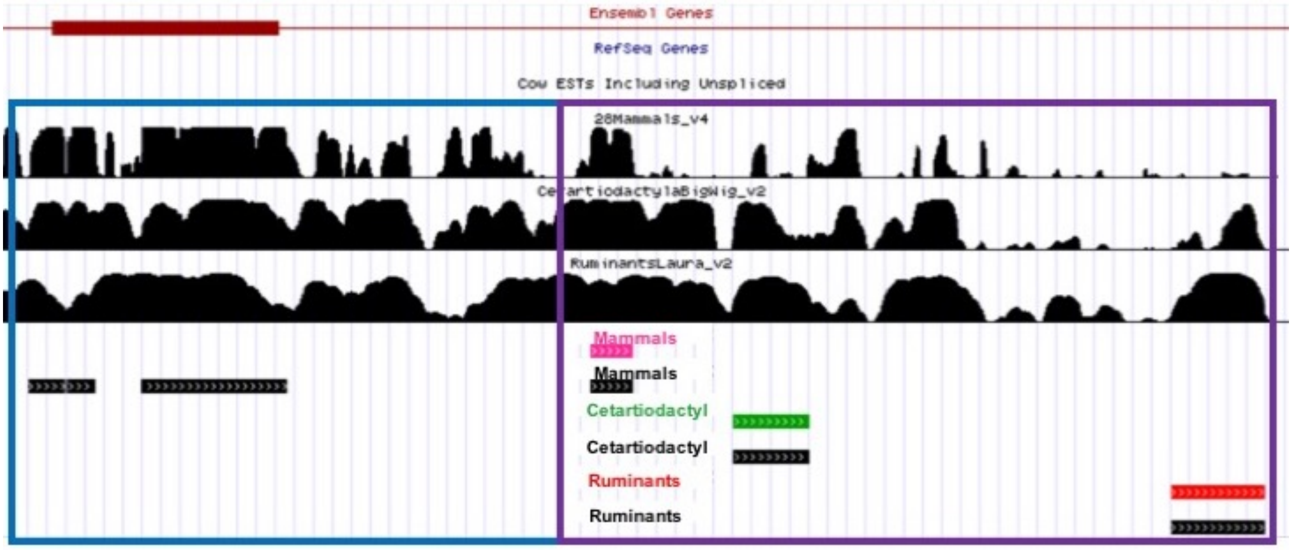
Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the project so far)

The fellow learnt state-of-the-art techniques in genetic association analysis and comparative molecular genomics becoming confident in using a diverse range of bioinformatic tools. She, indeed, acquired the knowledge to independently carry out two main projects from this Marie Curie action. The results of one project have been presented in an international conference as both oral and poster presentation and it is under preparation for submission, while the other needs some analysis to be finalized and soon to be under preparation for publication.

Address (URL) of the project's public website

Images attached to the Summary for publication

A snapshot of conserved elements (CEs) and conserved non coding elements (CNEs) along the cattle genome



CEs

Mammals: 1,192,479
Cetartiodactyls: 844,325
Ruminants: 596,993

CNEs

Mammals: 578,133
Cetartiodactyls: 483,584
Ruminants: 366,890