

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**Departamento de Genética**



**TESIS DOCTORAL**

**Estudio de la función y regulación del gen ovino *HSP90AA1***

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Judit Salces Ortiz**

Directora

M<sup>a</sup> Magdalena Serrano Noreña

**Madrid, 2015**



UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE GENÉTICA

TESIS DOCTORAL

**Estudio de la función y regulación del gen ovino  
*HSP90AA1***

Memoria presentada

por

Judit Salces Ortiz

Licenciada en Biotecnología,

para optar al grado de Doctor en Biología con Mención Europea

por la Universidad Complutense de Madrid







UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE GENÉTICA

**Estudio de la función y regulación del gen ovino  
*HSP90AA1***

TESIS DOCTORAL

JUDIT SALCES ORTIZ

Vº Bº Directora

La doctoranda

Mª Magdalena Serrano Noreña

Judit Salces Ortiz

MADRID, 2014



*Este trabajo se ha realizado en el Departamento de Mejora Genética Animal del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) de Madrid.*

*Ha sido posible gracias a una beca FPI-INIA asociada al proyecto RTA2009-00098-00-00 (Subprograma de Proyectos de Investigación Fundamental orientada a los Recursos y Tecnologías Agrarias en coordinación con las comunidades autónomas) y al convenio de colaboración CC13-071 entre el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) y El Centro de investigación y Tecnología Agroalimentaria de Aragón (CITA).*





*A mi Familia.*

*A mi P.A.*



No estaría ahora mismo escribiendo estas palabras si Malena no me hubiese brindado una fabulosa oportunidad hace ya algo más de cuatro años. He aprendido muchísimo en este tiempo, no sólo en el campo científico, que ha sido mucho, sino de escalada, montañas varias, vinos, ... en fin, que eres un pozo de sabiduría!!!! No solo has sido una gran directora de tesis, sino un gran apoyo al que en muchos momentos he podido recurrir, de verdad, Gracias. Como no, a mi Mariflor particular, mi maestra en el labo, con la que he podido contar no solo en aspectos *laboratoriales*, sino en muchas ocasiones dándome puntos de vista en circunstancias más trascendentales que no habría tenido en consideración si no fuera por ti.

A mi clon, poco tengo que decirte que no sepas. Gracias por todo, por estar ahí siempre, porque los años pasan y últimamente los kilómetros nos separan, y sin embargo solo estás a un telefonazo cuando empiezo a panicar ;)

A mi *hermana mayor* en este mundillo, Car, porque siempre estás ahí para todas mis dudas y cuestiones varias, como te echo de menos!! A Silvia y Fer, porque estos últimos meses he estado desvariando y nunca me habéis hecho sentirme como una loca en el despacho, gracias por aguantarme!!! :P. A Ana, María Saura, Estefânia, María Muñoz, Óscar, porque esos cafetines que a veces se hacían más largos de lo habitual enfrascados en debates varios son inolvidables. A Fabián, porque siempre encuentras un momento para escuchar mis alegrías y mis penas, poniendo un comentario sarcástico que siempre me hace sacar una sonrisa. A Cris, porque voy a echar mucho de menos nuestras charlas matutinas en el bus. A Carmen R. y Luis porque si no hubiese sido por vosotros, probablemente no hubiese podido tener la oportunidad de conocer a Malena. A Beatriz, María Gil y Jesús siempre pendientes de mis progresos. A Manuel, M<sup>a</sup> Jesús, Almudena, Rubén, Raquel, Cristina O.,... dando ánimos en esta última etapa.

A José Carlos, por darme la oportunidad de desplazarme a su laboratorio más tiempo de lo esperado y aún así, siempre dispuesto a ayudar. A Javi y Fonso, por conseguir sacarme mi lado más friki ;). A Laura, porque mi estancia en Santander no hubiese sido lo mismo sin ti, gracias por el apoyo pequeña Chef!

A Jorge, por sus continuos consejos, tanto en España como en la mismísima China. Por estar siempre dispuesto a corregir todos los trabajos que le he enviado en este tiempo o incluso implicarse de lleno en decisiones de última hora con título de tesis de por medio incluido... ☺.

I would like to thank James Kijas to give me the opportunity to work in his group during several months, because it was a fantastic experience, both professional and personal.

A Salo, Esteffi y Sergio: mi familia madrileña, por nuestras cenas brindando con una buena copa de vino sin venir a cuento, por nuestras escapaditas familiares con nuestra *pequeña*

Esther adoptiva, nuestras confianzas escondidas en la cocina, y tantos y tantos pequeños momentos familiares que me dejó en el tintero.

A las chicas de la uni, Ana, Blan, Pili, Patri, Ade, nos van separando los postdocs que surgen en lugares lejanos y demases, pero después de 10 años seguimos buscando un hueco para vernos.

A Jose *Farmacia*, Adri y Ángela porque mis visitas a casa nos serían las mismas sin teneros allí siempre.

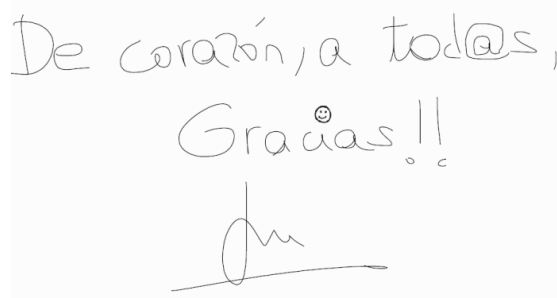
A mis tíos y primos que siempre habéis estado pendientes de mis progresos incluso estando preparados para ofrecer sus ovejas si hubiese sido necesario para mi trabajo! ☺ Y en especial a mi tía Eva, que me ha dado cobijo este último año y me ha aguantado mis malos humos cuando la frustración se cernía sobre mí.

A Pol y Dave, porque podéis llegar a ser mi peor pesadilla y sin embargo no sabría vivir sin vosotros.

Pero sin duda a quien más agradecida estoy es a mis padres. A mi madre, ejemplo de perseverancia y energía positiva. Porque en mis últimas épocas de bipolaridad me has sabido comprender cuando yo ni siquiera lo hacía.

A mi padre, porque siempre estás orgulloso de todo lo que hago y me empujas a superarme cada día.

Y como no, a Ernesto, gracias por estar siempre ahí, tanto para celebrar los buenos momentos como para consolarme en los malos. Por recorrerte medio mundo para estar conmigo y estar siempre dispuesto a ayudarme.



De corazón, a tod@s,  
Gracias!!  
dm





# Contents

Resumen	19
Summary	27
General introduction	35
Objetivos	59
Aim of the thesis	63
Chapter 1: Gene expression analysis	67
Chapter 2: Functional study and epigenetic marks	97
Chapter 3: An adaptive role gene	127
Chapter 4: From Genotype to Phenotype	163
General discussion	181
Conclusiones	195
Conclusions	199
References	203
Appendix I: Supplemental material	225
Appendix 2: Additional research work	247





The chapters in this Thesis correspond to the following scientific publications:

Los capítulos de esta Tesis se corresponden con las siguientes publicaciones científicas:

#### Chapter 1.

- Salces-Ortiz J, González C, Moreno-Sánchez N, Calvo JH, Pérez-Guzmán MD, Serrano MM. 2013. **Ovine *HSP90AA1* expression rate is affected by several SNPs at the promoter under both basal and heat stress conditions.** PLoS One. 24;8(6):e66641
- Salces-Ortiz J, Ramón M, González C, Pérez-Guzmán MD, Garde JJ, Calvo JH, Serrano MM. **Differences in the ovine *HSP90AA1* gene expression rates caused by two linked polymorphisms at its promoter affect rams sperm DNA fragmentation under environmental heat stress conditions.** PLoS One (submitted).

#### Chapter 2.

- Salces-Ortiz J, González C, Bolado-Carrancio A, Rodríguez-Rey JC, Calvo JH, Serrano MM. ***HSP90AA1* gene promoter: functional study and epigenetic modifications.** Under preparation.

#### Chapter 3.

- Salces-Ortiz J, González C, Martínez M, Mayoral T, Calvo JH, Serrano MM. **Looking for adaptation footprint in the *HSP90AA1* ovine gene. Learning from the Wild.** BMC Evolutionary Biology journal (submitted).

#### Chapter 4.

- Ramón, M, Salces-Ortiz J, González C, Pérez-Guzmán MD, Garde JJ, García-Álvarez O, Maroto-Morales Calvo JH, Serrano MM. 2014. **Influence of the Temperature and the Genotype of the *HSP90AA1* Gene over Sperm Chromatin Stability in Manchega Rams.** Plos One 9(1) e86107
- Salces-Ortiz J, Ramón M, González C, Pérez-Guzmán MD, Garde JJ, Calvo JH, Serrano MM. **Differences in the ovine *HSP90AA1* gene expression rates caused by two linked polymorphisms at its promoter affect rams sperm DNA fragmentation under environmental heat stress conditions.** PLoS One (submitted).



# Resumen



## **Estudio de la función y regulación del gen ovino *HSP90AA1***

Las consecuencias visibles, ya hoy en día, producidas por el cambio climático y el aumento en la demanda de alimentos debido al crecimiento de la población mundial, hacen necesario un cambio en las estrategias de manejo de las especies ganaderas. Entre las diferentes estrategias a tomar en consideración, la búsqueda de animales más adaptados a las nuevas condiciones ambientales existentes ha cobrado gran importancia.

El aumento de la temperatura global, es uno de los fenómenos más notables ocurridos en las últimas décadas. De hecho, las temperaturas medias globales han experimentado un aumento, prediciéndose incrementos de 1.8 hasta 3.9°C para el año 2100. Este hecho junto con el aumento de zonas desérticas en el mapa global, hacen necesaria la búsqueda de especies rústicas y adaptadas a estos ambientes, cuya capacidad productiva no se vea mermada por dichos factores.

La especie ovina fue una de las primeras especies domesticadas adaptada a climas agrestes y áridos de la que existen más de 1400 razas, repartidas prácticamente por todo el planeta. Este pequeño rumiante generalmente se explota por su triple aptitud productiva, tanto de lana, leche como carne, aunque existen razas que han sido seleccionadas específicamente para alguna de estas características. Sus condiciones de manejo van, desde sistemas extensivos o semi-extensivos en zonas rurales donde los ganaderos manejan rebaños con un reducido número de cabezas, hasta grandes empresas productoras con sistemas de producción muy intensivos y tecnificados.

Todos los organismos vivos poseen una serie de mecanismos para afrontar los cambios que se producen en el medio ambiente en el que se desarrollan, y especialmente para los cambios de temperatura. La respuesta al estrés térmico (HSR, por sus siglas en inglés), hace frente a los cambios que se producen en el organismo como consecuencia del aumento de temperatura para garantizar la viabilidad celular. La respuesta a estrés térmico supone la activación de una serie de genes, entre los cuales, aquellos que codifican las proteínas de shock térmico (HSP, heat shock protein), adquieren un importante protagonismo. El papel de las proteínas de estrés térmico, entre otros, es la restauración de la homeostasis celular que se ve profundamente alterada como consecuencia del estrés por calor. En concreto, la proteína de shock térmico de 90 kDa alfa ( $Hsp90\alpha$ ) realiza esta función identificando proteínas cuya

conformación se ve alterada por el aumento de la temperatura para su posterior reparación o destrucción. Además, forma parte de un complejo de proteínas activadoras del Heat shock factor 1, un factor de transcripción clave en la respuesta a estrés. El gen *HSP90AA1* que codifica la Hsp90 $\alpha$  presenta niveles de transcripción moderados en condiciones termo neutras. Bajo condiciones de estrés térmico, la transcripción del gen es inducida hasta niveles tres veces superiores a los existentes en condiciones basales. El gen *HSP90AA1* ha sido caracterizado y mapeado en la especie ovina, concretamente en la raza Manchega. Dicha raza, de aptitud lechera, es criada principalmente en la región de Castilla-La Mancha donde las temperaturas en verano son muy elevadas. Estas condiciones hacen de esta raza un excelente modelo animal para el estudio de la adaptación a condiciones de alta temperatura ambiental.

El presente estudio, se ha focalizado en la caracterización de la región promotora del gen ovino *HSP90AA1*, donde se ha detectado la existencia de un gran número de polimorfismos, de los cuales al menos uno de ellos, g.660G>C, ha sido asociado previamente a las diferencias de expresión del gen bajo condiciones de estrés térmico.

En estudios anteriores, se observó la relación de uno de estos polimorfismos, un SNP (del inglés, Single Nucleotide Polymorphism) que consiste en una transversión de citosina a guanina, situado 660 pares de bases antes del inicio de la transcripción (g.660G>C), con diferencias en el tiempo de incubación del scrapie. Animales portadores del genotipo CC<sub>.660</sub> presentaban un mayor el tiempo de incubación que aquellos portadores del genotipo GG<sub>.660</sub>. Posteriormente, en un pequeño grupo de animales en los que se tomaron muestras de sangre en momentos de distinta temperatura ambiental, se observó una diferencia en el patrón de expresión del gen *HSP90AA1* bajo condiciones de estrés térmico, de tal manera que los animales portadores del genotipo CC<sub>.660</sub> presentaban una mayor tasa de transcripción que aquellos portadores de CG<sub>.660</sub> y GG<sub>.660</sub>.

Para corroborar estos resultados previos realizados con una muestra limitada de animales y para analizar el posible efecto de otros polimorfismos existentes en la región promotora del gen *HSP90AA1*, en el trabajo actual y como uno de los objetivos de la presente tesis, se realizaron nuevos estudios de expresión génica con un tamaño muestral superior y una metodología más compleja (Capítulo I). Los resultados de este trabajo permitieron constatar que el SNP g.660G>C tiene un efecto de la misma

magnitud sobre la tasa de expresión del gen, tanto en condiciones de termo neutralidad como de estrés por calor. Además, la inserción de una citosina 668 pares de bases antes del inicio de la transcripción (g.667\_668insC) produce unas diferencias de transcripción del gen de mucha mayor magnitud y sólo en condiciones de elevada temperatura. Así, en los animales que portan la inserción en homocigosis la expresión del gen se ve aumentada hasta 3 veces respecto a la de aquellos animales heterocigotos y sin presencia de inserción.

El promotor del gen *HSP90AA1* ha sido caracterizado como un promotor híbrido que no solo contiene la caja TATA, sino que además posee una isla CpG reguladora (Capítulo 2). Este hecho confiere a este tipo de promotores una mayor plasticidad respecto a la regulación de su transcripción. Además, dentro de esta isla CpG es donde se encuentran los polimorfismos detectados, siendo uno de ellos, g.660G>C, susceptible de metilación. Este hallazgo, podría explicar las diferencias en expresión observadas ya que también es posible la existencia de una regulación epigenética adicional a la regulación canónica del gen. Dado que los polimorfismos asociados a los cambios de expresión del gen, observado en el Capítulo I (g.660G>C y g.667\_668insC) se encontraban en los boques de ligamiento con otros polimorfismos, se llevaron a cabo experimentos *in vitro* para tratar de determinar la mutación responsable de los cambios de transcripción observados en el capítulo previo. Mediante estos experimentos *in vitro*, se pudo constatar que la combinación de tres citosinas, g.667\_668insC, g.666\_667insC y C<sub>-660</sub>, son responsables de la mayor tasa de transcripción del gen bajo condiciones de estrés térmico.

En el Capítulo 3, se procedió a determinar la presencia de los polimorfismos detectados previamente en la raza Manchega, en otras razas ovinas y especies de la subfamilia Caprinae, a la cual pertenece el género *Ovis*, además de en dos especies más alejadas correspondientes a la familia Bovidae, y su correlación con las variables climáticas de las regiones en que se desarrollan dichas razas y especies. A dicho fin, se caracterizaron los polimorfismos del promotor del gen *HSP90AA1* en 31 razas ovinas distribuidas por Europa, Asia y África, en 9 especies de la subfamilia Caprinae, además de en dos especies más alejadas, incluidas en el taxón Bovini. Se ha podido constatar la correlación entre la frecuencia de los alelos de aquellos polimorfismos asociados a cambios en la tasa de expresión del gen dependientes de genotipo y condiciones ambientales (g.660G>C y g.667\_668insC) con las variables climáticas existentes en las distintas regiones de origen de las distintas razas ovinas analizadas. Además, se ha



podido dilucidar el alelo salvaje en muchos de los polimorfismos identificados en el promotor ovino. En el caso del SNP g.660G>C se ha identificado el alelo C como el salvaje tanto por su frecuencia en las razas analizadas como por su susceptibilidad de metilación alelo-específica (Capítulo 2). En el caso de la inserción de citosina en la posición -668 (g.667\_668insC), la delección parece ser el alelo salvaje en la especie ovina. Sin embargo, los resultados obtenidos en las especies procedentes de la subfamilia Caprinae no son tan esclarecedores.

La comparación de la región promotora de las distintas especies de rumiantes estudiadas, ha permitido constatar que muchos de los polimorfismos detectados en la especie ovina se remontan a la formación de la subfamilia Caprinae, aunque su ventaja selectiva solo ha podido ser constatada en la actual especie ovina. También han conducido a un novedoso hallazgo, en concreto, el elevado número de polimorfismos compartidos por las especies *Ovibos moschatus* y *Ovis aries*, lo que podría contribuir a dilucidar la controvertida posición del *O. moschatus* en la subfamilia Caprinae. Sorprendentemente, esta especie que actualmente se encuentra confinada a regiones árticas, presentan una buena adaptación a estrés térmico, al menos en lo que se refiere a la frecuencia de alelos de respuesta a dicha fuente de estrés en el gen *HSP90AA1*. Una posible hipótesis sobre este hecho, sería la herencia de dichas mutaciones de un ancestro lejano de dicha especie, el *Praeovibos*, cuyos fósiles han sido encontrados en yacimientos arqueológicos localizados en ambientes mucho más templados que el actual ecosistema de *Ovibos moschatus*. Contrariamente a lo esperado, en una especie típica de zonas desérticas con elevadas temperaturas, *Ammotragus lervia*, el polimorfismo más asociado a la respuesta a estrés térmico (g.667\_668insC) se encuentra fijado para el alelo desfavorable ligado a menores tasas de expresión del gen bajo condiciones de estrés por calor.

Finalmente, y como resultado más directamente aplicable al sector de producción ovina, se han podido asociar los genotipos responsables de las diferencias de expresión del gen *HSP90AA1*, con las diferencias en el fenotipo de un carácter reproductivo de los machos de esta especie, en concreto, con la tasa de fragmentación del ADN espermático en machos de la raza Manchega, estudio desarrollado en el Capítulo 4. En este trabajo, se ha observado que existen fases en el proceso de la espermatogénesis, en las que la temperatura ambiental es un factor crítico en el desarrollo de las células germinales masculinas. En concreto, se ha constatado que existe un mayor índice de fragmentación del ADN de forma alelo-

---

---

específica durante la fase de espermatocitogénesis. Esta fragmentación es casi cuatro veces superior en animales portadores del genotipo combinado ID<sub>-668</sub>CG<sub>-660</sub>, frente a los animales II<sub>-668</sub>CC<sub>-660</sub> por cada unidad de incremento del índice de temperatura-humedad (THI) a partir de un umbral de THI igual a 22.

Los resultados obtenidos a través de los distintos enfoques abordados nos han permitido, en primer lugar, determinar cuáles de los 11 polimorfismos estudiados están asociados a las diferencias en la tasa de expresión del gen, tanto bajo condiciones de estrés térmico como en situaciones de termoneutralidad. En segundo lugar, conocer la compleja estructura del promotor del gen *HSP90AA1* que revela la existencia de un doble modo de regulación de la transcripción. Uno de ellos, constitutivo responsable de la expresión del gen bajo condiciones basales termoneutras y otro que se activa en situaciones de estrés térmico cuando se requiere una sobreexpresión del gen. En tercer lugar, los polimorfismos descritos en la raza ovina Manchega no son exclusivos de esta raza, sino que se encuentran presentes, en las 31 razas ovinas estudiadas y en varias especies de la subfamilia Caprinae más o menos próximas a la especie *Ovis aries*. Además, la frecuencia de los polimorfismos asociados a la respuesta a estrés térmico está altamente correlacionada con las variables climáticas de las regiones de origen de cada una de las razas estudiadas, revelando la importancia de este gen en la adaptación de los animales a las condiciones de su entorno a pesar del proceso de domesticación al que esta especie ha estado sometida. Por último, se ha constatado que el genotipo combinado del SNP g.660G>C y la inserción g.667\_668insC tiene un papel importante en las diferencias observadas en la tasa de fragmentación del ADN espermático bajo condiciones de estrés por calor. Estas diferencias en fragmentación dependientes de genotipo y temperatura son mayores cuando el estrés térmico se produce en la fase de espermatocitogénesis, fase en la cual las células poseen el máximo potencial transcripcional.



# Summary



## **Estudio de la función y regulación del gen ovino *HSP90AA1***

There are currently appreciable consequences caused by climate change associated with agricultural and livestock. Moreover, there is an increase on food demand due to the growth of world population which is expected to continue increasing. Thus, it seems necessary a change in the management strategies of livestock species to cope those demands. Among the different strategies to consider, searching for more adapted animals to the new environmental conditions has become very important.

The increase of global temperature, is one of the most notable changes occurred in recent decades. Indeed, average global temperatures have risen considerably, and it has been predicted to reach increases of 1.8–3.9°C by 2100. Along with the increase of desert areas on the world map, it makes necessary to search for rustic species better adapted to these environments and whose productive capacity would not be impaired by these factors.

Among livestock animals, sheep (*Ovis aries*), were one of the first domesticated species. Sheep were first reared for access to meat before human mediated specialization for wool and milk commenced 4,000–5,000 years ago. It is a species adapted to arid and harsh climates. There are more than 1400 sheep breeds spread almost all over the planet. Sheep are probably the most versatile of the domestic animal species and it is appreciated by its triple-purpose, capable of providing wool, milk and meat. Even though, nowadays most of breeds are managed to one of the productive purposes. Its management ranges from extensive or semi-extensive systems in rural areas where farmers have a small number of animals to large companies producing very intensive systems and technologically advanced production.

Furthermore, all living organisms have a number of mechanisms to deal with changes that occur in the environment in which they develop, especially against temperature changes. The heat stress response (HSR) is the mechanism in charge to face the changes that occur in the organism as a result of increased temperature to ensure cell viability. It was first mentioned in 1962 by Ferruccio Ritossa who discovered in the salivary glands from *Drosophila melanogaster* the formation of new and rapidly synthesized RNA when they were exposed to high temperatures. The HSR involves the activation of a number of genes including those that encode heat shock

proteins (hsp), proteins which play an important role. The role of hsps, among others, is the restoration of cellular homeostasis that is profoundly altered after heat stress events. Specifically, the heat shock protein of 90 kDa alpha (Hsp90 $\alpha$ ) performs this function by identifying proteins whose conformation appears altered by temperature increase and manages its subsequent repair or destruction. It also takes part of a protein complex in charge of the activation of Heat Shock Factor 1, a key transcription factor in the response to stress. The gene encoding the Hsp90 $\alpha$ , *HSP90AA1*, presents moderate transcription levels in thermo-neutral conditions. Under heat stress conditions, gene transcription is induced up to three times higher levels than those in basal conditions. The *HSP90AA1* gene has been characterized and mapped in sheep, particularly in the Manchega breed. The Manchega sheep breed is mainly a dairy sheep breed raised primarily in the region of Castilla-La Mancha where summer temperatures are high. These conditions make this breed an excellent model to study adaptation to high environmental temperature conditions.

The present study is focused on the characterization of the promoter region of the *HSP90AA1* ovine gene where there have been observed numerous polymorphisms. At least one of them (g.660G>C) has been previously associated with differences in expression of the gene under heat stress conditions.

Previous studies described the relationship of a SNP (Single Nucleotide Polymorphism) consisting of a transversion from cytosine to guanine located 660 base pairs upstream of the transcription start site (g.660G>C), with differences in the incubation time of scrapie. It was observed that CC<sub>-660</sub> animals had higher incubation periods of the disease than those with the GG<sub>-660</sub> genotype. Subsequently, in a small group of animals biological samples were taken at different time points of environmental temperature. Differences in the expression pattern of the *HSP90AA1* gene under thermal stress conditions were observed. Animals carrying the genotype CC<sub>-660</sub> had a greater rate of transcription comparing with those carrying CG<sub>-660</sub> or GG<sub>-660</sub>.

To corroborate these previous results using a limited sample of animals and to analyze the possible effects of other existing polymorphisms at the promoter region of the *HSP90AA1* gene this thesis was initially carried out. In the current job, further studies were performed with a higher sample size and more complex methodology (Chapter 1). The results from this work led to demonstrate that the g.660G>C SNP

has the same magnitude effect on gene expression rate, as much under thermal neutrality as heat stress. In addition, a cytosine insertion 668 base pairs upstream of the transcription start site (g.667\_668insC) produces much greater magnitude differences in transcription and only under high temperature conditions. Thus, in animals carrying homozygous insertion, gene expression is increased with respect to those without cytosine presence or carrying heterozygous insertion.

The *HSP90AA1* gene promoter has been characterized as a hybrid promoter that contains not only the TATA box, but also a regulatory CpG island (Chapter 2). This fact confers to this type of promoters a greater plasticity regarding to the regulation of transcription. The polymorphisms previously described are located within this CpG island. One of them, g.660G>C, is susceptible to methylation. This finding might explain the observed differences in expression as the existence of an additional epigenetic regulation to the canonical gene regulation is also possible. Furthermore, it was also shown *in vitro* the role of candidate polymorphisms independently. Moreover, it has discarded the possibility that some other detected polymorphisms linked to g.660G>C and/or g.667\_668insC could be the responsible for the observed expression changes described in Chapter 1. It has also been observed *in vitro* that the combination of three cytosines, g.667\_668insC, g.666\_667insC and C<sub>-660</sub>, one following the other, has greater effect in the expression pattern under heat stress.

In Chapter 3, we proceeded to determine the presence of polymorphisms previously detected in the Manchega breed in other sheep breeds and species of the Caprinae subfamily, to which the *Ovis* genus belongs. Additionally, two more distant species corresponding to the Bovidae family were also studied. We have analyzed the correlation between genotype frequencies of such species with the climatic regions where they belong. With that purpose, we have characterized the polymorphisms from the *HSP90AA1* gene promoter in 31 sheep breeds distributed in Europe, Asia and Africa, 9 species of the subfamily Caprinae and the addition of two species belonging to the Bovini taxon. It has been noticed the correlation between the frequency of those polymorphisms previously related with differences in expression rates (g.660G>C, g.667\_668insC) and the climatic variables from the different breeds and species regions of origin. Different breeds from same latitudes and, therefore, with similar climatic conditions were clustered together, depending on genotype and environmental conditions where they are reared. That showed the relationship between allele frequencies between certain polymorphisms crucial to high temperatures adaptation.



In addition, we were able to elucidate the wild allele in many of the polymorphisms identified in the ovine promoter. In the case of SNP g.660G>C, C has been identified as the wild allele by its distribution in the different sheep breeds and the Caprinae species here analysed as well as based on its allele-specific methylation pattern (Chapter 2). The deletion is the wild-type allele of the g.667\_668insC. In addition, the current appearance of an extra cytosine (g.666\_667insC) has been a mutation that could confer a selective advantage in adaptation to increasingly hot environments. In this case, the wild allele of this insertion, g.666\_667insC, is also the deletion.

Moreover, the results obtained from the comparison of different ruminant species promoter's region have allowed a deep study of the presence and frequency of those polymorphisms critical for heat environments in the Caprinae subfamily. It includes the study of the promoter region in closer species to *Ovis aries*, which are reared in warm environments. In addition to that, the analysis has led us to a novel finding: the high number of shared polymorphisms between *Ovibos moschatus* and *Ovis aries*. This fact may help to clarify the controversial position of *Ovibos moschatus* in the Caprinae subfamily. Surprisingly, this species which is currently confined to arctic regions, is well adapted to heat stress, at least with respect to the frequency of alleles in response to that source of stress. One possible hypothesis for this fact would be the inheritance of these mutations from a distant ancestor of this species, *Praeovibos*, whose fossils have been found in archaeological sites located in more temperate than the current ecosystem of *Ovibos moschatus*. Contrary to expectations, in a typical species of desert areas with high temperatures, *Ammotragus lervia*, the polymorphism most associated with the response to heat stress (g.667\_668insC) contains fixed the unfavorable allele linked to lower rates of expression of the gene under heat stress conditions.

Finally, and as a result more directly applicable to the ovine production sector, it has been able to correlate expression differences of dependent *HSP90AA1* gene genotype and climatic conditions, with differences in the phenotype of a reproductive character in Manchega breed rams. Specifically, differences in genotype dependent expression have an effect on the sperm DNA fragmentation rate in those animals (described in Chapter 4). In this study, it has been observed that there are stages in the process of spermatogenesis in which the environmental temperature is a critical factor during the development of male germ cells. Specifically, it was found that there is an allelic-dependent DNA fragmentation rate during the spermatocytogenesis. This

fragmentation is almost four times higher in animals carrying the combined genotype ID-668CG-660, compared to II<sub>-668</sub>-CC<sub>-660</sub> animals. Differences observed for each unit of increase of temperature-humidity index (THI) from a value threshold of THI=22.

The results obtained through the different approaches here summarized, have allowed us, firstly, to determine which of the 11 studied polymorphisms are associated with differences in gene expression rate, under both thermal stress and thermoneutral conditions. Secondly, to know the complex structure *HSP90AA1* gene promoter has, revealing the existence of a dual mode of transcription regulation. One of them responsible for the constitutive expression of the gene under thermoneutral conditions and the other, activated in situations when thermal stress effects requires gene overexpression. Thirdly, polymorphisms described in Manchego sheep breed are not unique to this breed; they are present not only in the 31 sheep breeds studied, but also in several species of the subfamily Caprinae so close to the *Ovis aries* species. Moreover, the frequency of the polymorphisms associated with response to heat stress is highly correlated with climatic variables in the regions of origin of each of the breeds studied, revealing the importance of this gene in animal adaptation to their environment conditions. This adaptation is clear in sheep, despite the domestication process to which this species has been subjected. Finally, it was found that the combined genotype of g.660G> C and g.667\_668insC has an important role in sperm DNA fragmentation rate, where clear differences were observed under heat stress conditions. These differences, genotype and temperature –dependent, are also higher when thermal stress was produced during the spermatocytogenesis stage. This stage is the one in which the cells have the maximum transcriptional potential.



# **General introduction**



## I. Climate change and global warming

Many impacts of global warming are already detectable. Glaciers retreat, the sea level rises, the tundra thaws, the desert areas on the world map increases, hurricanes and other extreme weather events occur more frequently. Moreover, specialist species as penguins, polar bears or corals, which are adapted exclusively to a specific ecological niche struggle to survive. Experts anticipate even greater increases in the intensity and prevalence of these changes as the 21st century brings rises in greenhouse gas emissions (Koneswaran et al. 2008). The five warmest years since the 1890s were 1998, 2002, 2003, 2004, and 2005 ((NOAA) 2011). Indeed, average global temperatures have risen considerably, and the Intergovernmental Panel on Climate Change (Lenny Bernstein et al. 2007) predicts increases of 1.8–3.9°C by 2100. These temperature rises are much greater than those seen during the last century, when average temperatures rose only 0.06°C per decade (Kicza 2011). Since the mid-1970s, however, the rate of increase in temperature rises has tripled. Even though high efforts have been made to control and study the way to stop or at least alleviate worldwide temperature increment, it does not seem to be working so new strategies for environmental adaptation must be carry out.

“For a long time, nearly all climate policies have focused on mitigation. Now, with some change in climate inevitable there has been a shift in emphasis to adaptation. Investments in adaptation, which can help reduce exposure to climate impacts and may also lessen uncertainty in the assessment of possible and probable impacts...” (David Victor 2014).

This current concern about global warming and its effects over agricultural and livestock production systems have opened novel scientific opportunities to study the adaptation of organisms to new and harsher environmental conditions. Only this year for example, and up to date (9/11/2014) there have been published 2625 scientific articles in PubMed website using *environmental adaptation* as key words being most of them related to high temperatures adaptation.

In this context, the study of the genetic basis of traits linked with adaptation and fitness has great importance. Heat is one of the main sources of stress which has an important impact in livestock production. The genetic variability underlying animal's thermo tolerance could be exploited in livestock breeding programs to achieve animals that could cope with the effects of heat stress over productive and functional traits.

## 2. Ovine species

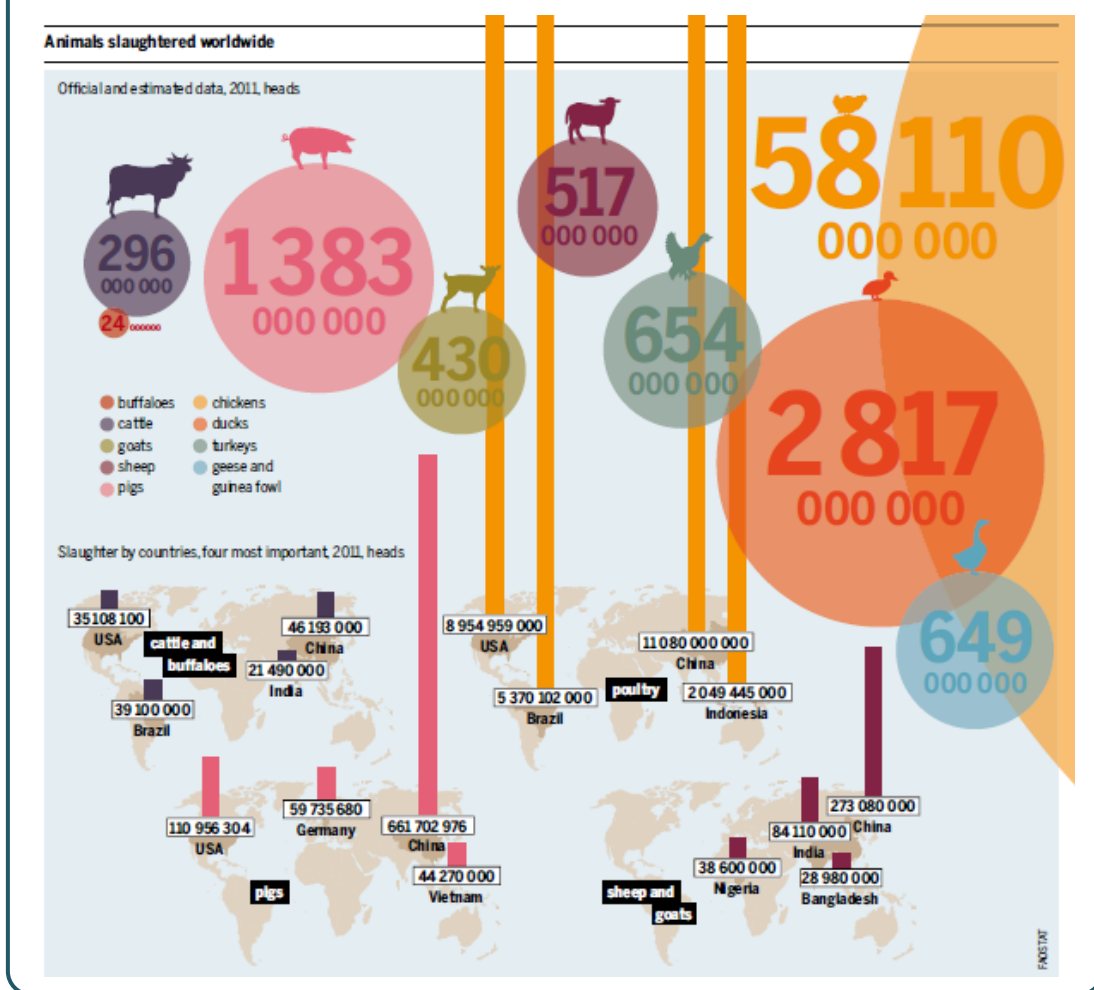
Between 9000 and 5000 BC the human race underwent to the so-called Neolithic Revolution (Weisdorf 2005). In this time we have records data of the first settlements (Braidwood 1960, MacNeish 1992, Olsson 2001, Weisdorf 2003). Human communities completely transform their way of life, from hunting and gathering through the transition to agriculture and the emergence of farming. The spread of the Neolithic process found its strength in the fusion of two complementary food production strategies developed independently, cereal growing and small livestock rearing, centered respectively on the Levant and the Taurus–Zagros zone, which merged during the second half of the eighth millennium BC (Pedrosa et al. 2005).

Among livestock animals, sheep (*Ovis aries*), were one of the first domesticated species (Pedrosa et al. 2005). Sheep were first reared for access to meat before human mediated specialization for wool and milk commenced 4,000–5,000 years ago (Chessa et al. 2009). The form of the wild sheep ancestral population and the number of times and the process of its domestication remain unknown, as does its genetic contribution to at least 1400 breeds (Scherf 2000) currently recognized in today's agricultural systems. There are several works based on mtDNA variation which have tried to elucidate the origin of sheep domestication (Tapio et al. 2006).

Sheep are probably the most versatile of the domestic animal species. They are widely distributed throughout the world due to its high plasticity and adaptability to withstand poor nutrient diets and tolerance to extreme climatic conditions. They are kept mainly for the production of meat, wool and milk.

Small ruminants are a source of meat especially in developing countries where the climate makes difficult the management of any other livestock species. Small-stock husbandry plays a very important role as it allows for the exploitation of the low potential rural areas and wastelands. It is hard to obtain accurate data pertaining to sheep and goat meat consumption, as the animals are often slaughtered on-farm and consumed locally. The most widely-consumed meats in the world (in order) are pork, poultry, and beef. Nevertheless, lamb and mutton consumption represent primary animal-sourced protein for regions of North Africa, the Middle East, India, and parts of Europe. Moreover, lamb consumption has increased from 1965 to 2007 (FAO's latest numbers). Asia and Mexico experienced increases in consumption primarily because of increasing personal incomes and urbanization. Total consumption in these regions was also influenced by population growth (Fig. 1) (Timon 1985).

**Figure 1.** Livestock Worldwide consumption. Small ruminant consumption is mainly centered in developing countries. From Meat Atlas, 2014 (Stiftung 2014).



As ovine production, not only meat source but also wool and milk production are a big economic source for developing countries, during the last 20 years extensive scientific progress has been made towards increasing the efficiency of small ruminants production. However, research findings have not been fully tested or adopted by the farmers either because some of the data obtained in developed temperate countries are not appropriate for the developing countries (semiarid, arid and tropics) or there is a weakness in the institutional frame work for providing technical advice. Weaknesses in providing credit for the application of new technology and lack of organization of the market for the protection of the animal production also inhibit adoption of new methods (Timon 1985).



In addition, also developed Mediterranean countries as Greece, Italy and Spain are linked to historical and cultural use of ovine products. From farming systems with dominant extensive grazing situations, specific technologies and conditions for slaughtering as well as for the transformation processes of cheese-making and its maturing; they are also characterized by traditional nutritional habits of the consumers (Boyazoglu et al. 2001). One example is the Manchega breed, whose derived cheese is one of the most famous in the world, due to its strong flavour and traditional production in esparto grass molds which imprint the characteristic zigzag pattern along the side of the cheese.

### 3. Regulation of gene expression in eukaryotes: promoter's function

Eukaryotic genes can generally be divided by a proximal regulatory region, called promoter and a codificant region, which is compound by translated (exons) and non-translated (introns) regions. There also exist distal regulatory regions, potentiating or repressing transcription. They are usually located 5' of the gene, even though it has also been described regulatory regions located at 3' and within introns.

A canonical promoter structure is defined as a regulatory DNA sequence located 5' from a gene, which contains specific sequences recognized by proteins called transcription factors (TFs). These proteins help to recruit RNA Polymerase II (RNAP II) to catalyze the transcription process. There are many proteins described in eukaryotic cells, which binds to DNA and are also necessary to RNAP II recognition site.

In fact, it is estimated that over 5% of eukaryotic genes encode TFs, which gives us an idea of the importance of these proteins in gene expression. It also shows how complex the regulation of the transcription process can be (Babu et al. 2004). Eukaryotic promoters are highly diverse, making difficult to characterize them (Gagniuc et al. 2012). In the basic structure of a eukaryotic promoter we can distinguish (Fig. 2):

- **The core promoter**, is the best-characterized transcriptional regulatory sequence in complex genomes because of their predictable location which can extend 35 bp 5' or 3' of the transcription start site (TSS). It is the region of the promoter needed for transcription, that is, those elements that interact directly

with components of the basal transcription machinery. It contains the TSS site (+1), the RNAP II binding site, and general transcription factor binding sequences (Muhlbacher et al. 2014).

- **The proximal promoter**, which is located 5' from the core promoter. It contains primary regulatory elements, that is, specific sequences of the transcription factor binding site (Soltani et al. 2014).
- **The distal promoter**, which contains additional regulatory elements. It tends to have weaker effects on gene expression elements located closer to where transcription begins. The regulator elements, despite being located several Kilobases away from the transcription start site may exert their modulator function through other proteins or altering DNA topology and interacting directly with the transcription machinery (van Arensbergen et al. 2014).

Until recently it was believed that core promoters were unchanged, but now we know they have a high structural and functional variability, which appears to contribute to the regulation of differential gene expression, since combining different regulatory elements generates different promoters (Smale 2001). The utility of the diversity is the ability to integrate all existing information on the status of the cell where it resides, and altering the ratio of gene transcription under conditions that occur in the cell in a particular circumstance.

There is another level of regulation and signal integration through the huge variety of TFs that can bind specifically to these DNA sequences and altering the expression levels of the genes. Normally, it is considered that the sequences contained in the promoter to which no TFs can be joined, are non-functional sequences. However, in some cases these nucleotides can influence local chromatin conformation, indirectly playing a regulatory role, and by this way affecting the union of TFs with their target sequences and thus, the genetic expression (Berger 2007, Clapier et al. 2009).

### **3.1 Promoter's regulatory elements.**

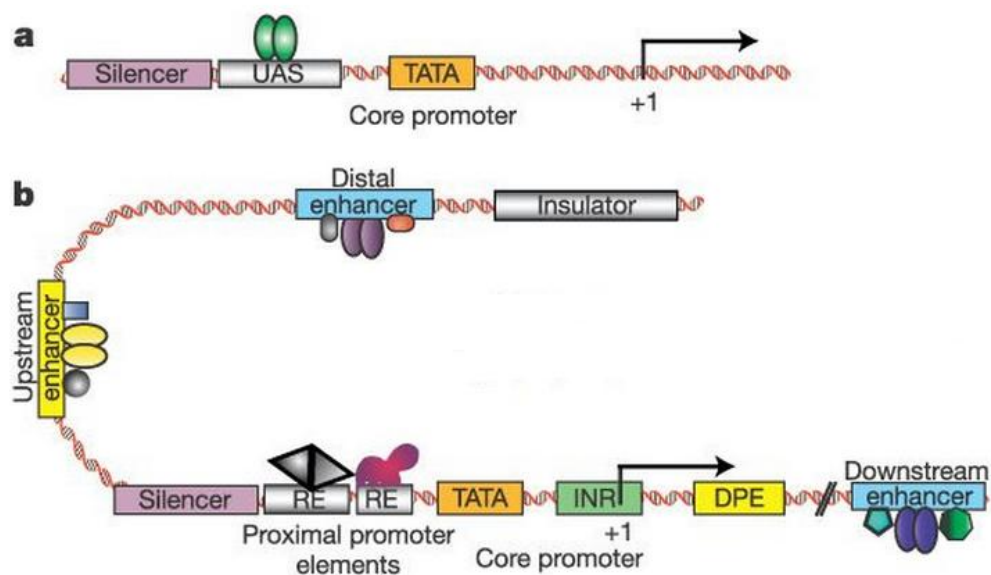
The promoter's regulatory sequences are typically short sequences present in the vicinity of the genes or in introns. Currently, the systematic knowledge of these

sequences and how they act among gene regulation in complex sensitive networks to exogenous signals is very low and it is beginning to develop through studies of comparative genomics, bioinformatics and systems biology. The identification of regulatory sequences is based in part on the search of evolutionarily conserved non-coding regions (Dubchak et al. 2000, Santini et al. 2003). Some of them have been described as regulatory elements located mostly in the core promoter, several of which are listed below (see Fig. 2):

**Figure 2.** General diagram of eukaryotic gene regulation (Levine et al. 2003).

**a.** Simple eukaryotic transcriptional unit. A simple core promoter (TATA), upstream activator sequence (UAS) and silencer element spaced within 100–200 bp of the TATA box that is typically found in unicellular eukaryotes.

**b.** A complex arrangement of multiple clustered enhancer modules interspersed with silencer and insulator elements which can be located 10–50 kb either upstream or downstream of a composite core promoter containing TATA box (TATA), Initiator sequences (INR), and downstream promoter elements (DPE).



- **TATA box:** this element is present in many eukaryote promoters. It is flagged 5', about 30 bp of the TSS (+1). It has been observed that the first RNA nucleotide synthesized is usually an A in 50% of cases. The TATA box, has the following consensus sequence: 5' TATA(A/T)AA(G/A) 3', used for the binding of the TATA binding protein (TBP) which recruitment is critical for the onset of transcription. It has the ability to position and orient RNAP II on the DNA for transcription initiation (Muhlbacher et al. 2014). The basal transcriptional

complex (BTC) is constituted by the TATA box, the RNAP II and the TBP associated factors (TAFs), including all essential transcription factors for anchoring and recognition of the TSS. It has been found that this complex has a very low basal transcriptional activity, and that other factors apart from TSS must stimulate the transcription. A gene may have more than one BTC, each of which begins in a transcription different position. This allows obtaining different products of the same gene (Maston et al. 2006).

- **Initiator (Inr)**, it is a feature of many genes lacking TATA box which supplies its functions, but in some promoters it is located also coupled to the TATA box (Maston et al. 2006).
- **3' promoter element** or **DPE** (Downstream Promoter Element): we can find it in promoters that have lost the TATA box in positions from +28 to +32. It is believed that the presence of DPE is an alternative mechanism by which TFIID transcription factor binds DNA (Maston et al. 2006).
- **5'UTR**: after the BTC binds to DNA, a second contact 30 base pairs downstream is established. This second contact is the proper TSS. This point is determined by the BTC and the particular composition of each promoter, which will make specific proteins to be bound to the nucleotide sequence and interact with the basal transcription complex, after that, the gene will start to express. The distance between sites of transcription initiation and translation differ considerably among genes, ranging from 101-104 bp. The **non-translated 5' region** or **5'UTR** may contain introns which alters post-transcriptional length, although the effect of the 5'UTR variability, is not known (Araujo et al. 2012).
- **CAAT box**: It is usually about 75-80 bp from the transcription start site. It serves as an anchor for the TF binding to the DNA sequence (Carninci et al. 2006).
- **CpG Islands (CGIs)**: are rich CGs sequences of 20-50 bp, which are generally place about 100-200 bp from the TSS generally lacking TATA box or Inr. Given the low frequency of CG nucleotides in the genome vertebrates, it is believed

that the presence of these sequences in this location is not due to chance; and furthermore, it is known to be the DNA methylation sites, suggesting playing a role in the regulation of transcription. CGI promoters turn out to have distinctive patterns of transcription initiation and chromatin configuration (Deaton et al. 2011). Approximately 70% of annotated gene promoters are associated with a CGI, making this the most common promoter type in the vertebrate genome (Saxonov et al. 2006). Customarily, CGIs are often associated with the regulation of housekeeping genes, developmental regulator genes and some tissue-specific genes expression (Deaton et al. 2011). The remaining genome contains disperse and generally methylated CpGs .

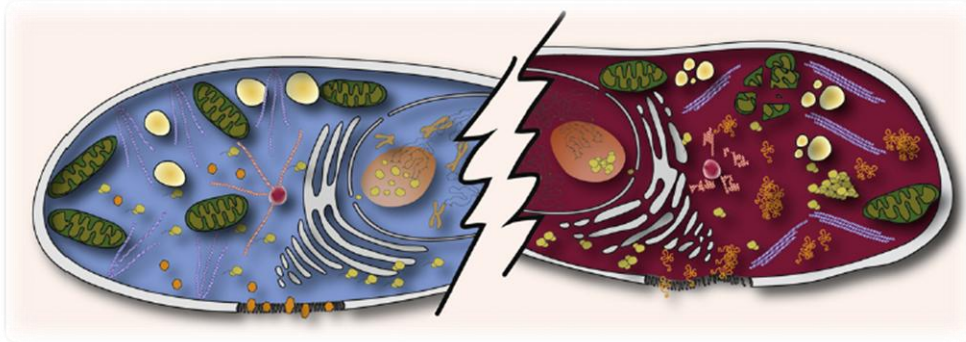
- **Enhancers** are distal regulatory elements which are located several Kilobases away from the gene they control. They can be found in 5' of the regulated gene, introns and 3'-regions from the coding region gene. They tend to be specific to cell type. They are composed of several elements that contribute to increase the expression of a gene regardless its position or orientation, without being confined to a particular gene to perform its function (van Arensbergen et al. 2014).
- **Silencers** are negative regulatory elements that play transcendental functions in spatial and temporal expression of genes. They may be located 3' or 5', in introns or exons. Silencers can exert their function directly, joining the DNA sequences and preventing the binding of activating factors to RNAP II or DNA; or indirectly by recruiting for example histone deacetylase activities (Muhlbacher et al. 2014).
- **TFIIB recognition element** or **BRE**, which is situated immediately 5' of the TATA box of some promoters. It seems to repress transcription in eukaryotes, while activates the expression of genes in less evolved organisms (Carninci et al. 2006).
- **Insulators** or **insulator sequences** are long-acting repressors. They repress inadequate interactions which facilitates the regulation of transcription (Carninci et al. 2006).

#### 4. The heat shock response and heat shock proteins

Living cells can function only within a narrow range of stressful conditions as temperature, oxygen supply, pH, ion concentrations or nutrient availability (Fulda et al. 2010). Anyhow, these conditions are not always stable, and organisms must survive in an environment where these and other conditions can vary regularly. For all living organisms, temperatures only moderately above the respective optimum growth temperature represent a challenging problem for survival (Fig. 3). Changes in the optimal temperature range, can compromise structure and function viability. By this way, organisms have adopted mechanisms to cope these changes. One of the most powerful adaptation mechanisms is the heat shock response (HSR), which was first mentioned in 1962 by Ferruccio Ritossa (Ritossa 1962). Ritossa discovered in the salivary glands from *Drosophila melanogaster* the formation of activated chromosome puffs when they were exposed to high temperatures (Ritossa 1996) (Fig. 4). These puffs resulted to be new RNA synthesized much rapidly than he had observed before (Ritossa 1996). Nowadays we know that the HSR modulates the transcription of more than 1000 genes. Encoding factors participating in protein folding, degradation, transport, RNA repair, metabolic pathways, and so forth, are upregulated under HS conditions.

The HSR is a highly conserved program of changes in gene expression (Fig. 5) that results in physiological and metabolic adaptation to the new conditions. But not only it is one of the most ancient molecular pathways but also, it is very similar under different stress conditions such as oxidative stress or cold shock. The physiological impacts of heat stress includes changes both at the cellular and at the molecular levels (Velichko et al. 2013).

**Figure 3. Effects of heat shock on the organization of the eukaryotic cell.** An unstressed eukaryotic cell (left) is compared to a cell under heat stress (right). Heat stress leads to damage to the cytoskeleton, including the reorganization of actin filaments (blue) into stress fibers and the aggregation of other filaments (microtubuli, red). Organelles like the Golgi and the endoplasmic reticulum (white) become fragmented and disassemble. The number and integrity of mitochondria (green) and lysosomes (yellow-white gradient) decrease. The nucleoli, sites of ribosome (yellow) assembly, swell, and large granular depositions consisting of ribosomal proteins become visible. Large depositions, the stress granula (yellow), resulting from assemblies of proteins and RNA, are found in the cytosol in addition to protein aggregates (hexagonal versus spaghetti style, orange). Finally, there are changes in the membrane morphology, aggregation of membrane proteins, and an increase in membrane fluidity. Together, all these effects stop growth and lead to cell-cycle arrest as indicated by the noncondensed chromosomes in the nucleus (Richter et al. 2010).



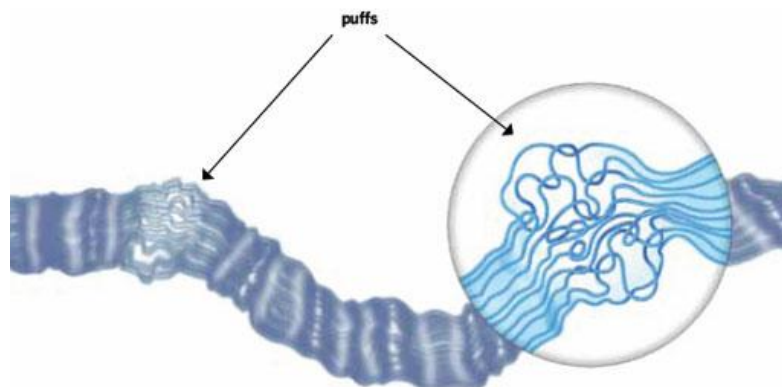
### Cellular level response

- **Cell membranes.** Heat shock affects cell organization and cellular components. As response to stress cytoplasmic membrane fluidity is increased caused by an increment in the saturation level of lipid fatty acids (Torok et al. 2014). At the same time, saturated lipids seem to induce expression of HSP genes whose accumulation rigidify the membrane and restore its normal fluidity. Membrane fluidization also destabilizes cell surface morphology, for instance disruption of intercellular interactions, changes in the surface charge, and changes in the membrane potential of the cell due to  $\text{Na}^+/\text{H}^+$ -ion exchange channels and  $\text{K}^+$ -ATPase dysfunction. Under HS conditions, the entrance of extracellular calcium ions leads to the activation of a cascade of signals which includes the activation of HSP72 and HSP90 genes (Velichko et al. 2013).
- **Internal organization of the cell.** In eukaryotes, one of the major damages observed in response to stress conditions are defects of the cytoskeleton. An

elaborate signaling pathway has been elucidated, linking outer membrane transmembrane proteins that serve as putative heat sensors (Verghese et al. 2012). Mild heat stress leads to the reorganization of actin filaments into stress fibers, while severe heat stress results in the aggregation of vimentin or other filament-forming proteins, leading to the collapse of intermediary, actin, and tubulin networks (Richter et al. 2010). It has also been observed the loss of the correct localization of organelles and a breakdown of intracellular transport processes (Welch et al. 1985).

- **Cell cycle arrest.** Heat shock induces transient arrest in  $G_0/G_1$  due to a reduction of transcript levels of non-essential proteins to reduce protein missfolding caused by the elevated temperatures. It has been suggested that this fact is not a direct physiological phenomenon but rather a signaled event (Verghese et al. 2012) to activate the responsible mechanisms to restore cell homeostasis and implies that the cell does not recognize temperature directly.

**Figure 4.** Ritossa's "heat shock" puffs, which by light microscopy looked like cotton balls compressed between sections of tightly packed chromosomes, proved to be new sites of transcription. Thom Graves©



- **Metabolic reprogramming.** Quiescent cells that have exited the cell cycle concomitantly acquire substantial heat shock resistance in a process linked to nutrient availability. In 1983 it was already shown that starving cells ( $G_0$  phase) are significantly more thermotolerant than exponentially dividing populations (Paris et al. 1983).



- **Cell death.** Massive destruction and reorganization of cellular components lead to a general decrease in cell viability and to cell death. Induction of cell death depends on the strength and duration of HS. Under conditions of acute hyperthermia over 45.5 °C, cells die exclusively through necrosis (Velichko et al. 2013). Two modes of cellular death have been identified in cell populations subjected to heat stress below this temperature: rapid mode and slow mode (Velichko et al. 2013). The first one, results in apoptosis induction. The probability of inducing apoptosis depends not only on the intensity of the HS but also on the cell type. The slow mode of cell death is the result of delayed consequences of HS, such as centrosome damage, cell division aberrations, and mitotic catastrophe.

### **Molecular level response**

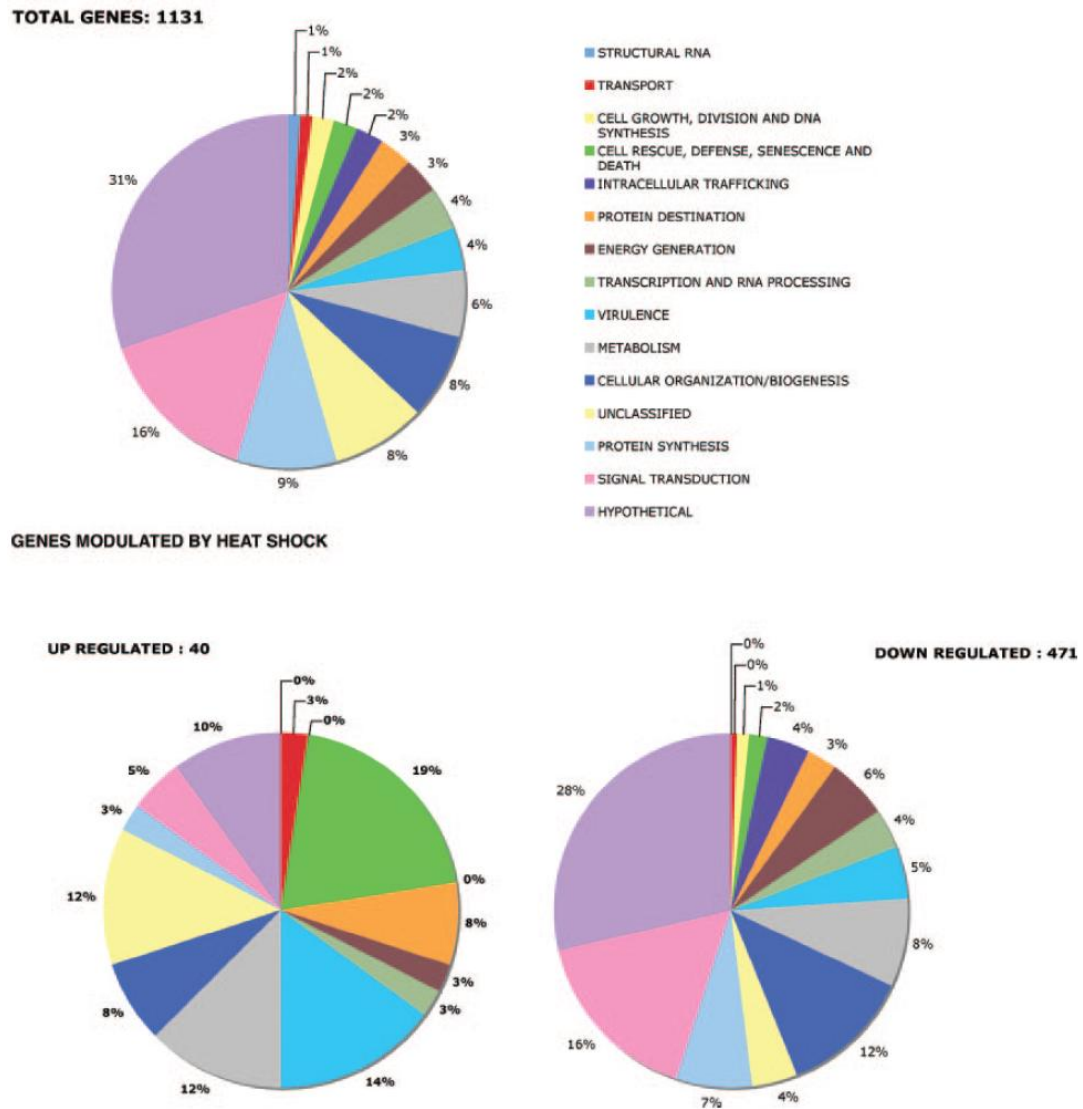
- **DNA replication.** HS inhibits multiple processes associated with DNA replication, including the initiation of new replicons, the elongation of replication forks, and the maturation of chromatin. DNA replication is thermosensitive and is inactivated during prolonged heat stress (Velichko et al. 2012). Many research studies experiments show that extracts from cells exposed to HS are less replicatively active than control cell extracts. Nevertheless, even after prolonged HS at 44 °C, cell extracts retain relatively high level of replication activity.
- **DNA damage response.** Different DNA repair systems are targets of HS. The system of base excision repair (BER) is impaired by hyperthermia through inactivation of DNA-polymerase  $\beta$ , one of the key enzymes of BER. In eukaryotic cells, such DNA lesions are removed through nucleotide excision repair (NER). Hyperthermia has a great impact on this system. In addition, all organisms possess systems to repair DNA double stranded breaks (DSBs) (Karpenshif et al. 2012). Two competing pathways are used by higher eukaryotes for DSB repair: homologous recombination (HR) and non-homologous end joining (NHEJ). Even though, elevated temperatures block the removal of DSBs, but it is still not well understood how this process is carried out (Turner et al. 2014).

- **Heat shock proteins (HSPs).** The induction of these molecular chaperones in the HSR is a key factor. HSFs are indispensable for stress-dependent activation of the expression of certain genes. Moreover, the present thesis will be focused in a HSP gene, by this way this issue deserves a separate section where it will be extensively described (see section 4.1).
- **Transcription, splicing and translation.** An important part of the heat shock response is the massive and extensive downregulation of gene expression, as a way of reducing energetic costs under stress (Alexandre et al. 2014) (see Fig. 5). Anyhow, as commented previously, a number of genes result upregulated during HSR, including the HSPs which at the same time, are necessary to activate certain stress-dependent genes. Moreover, proteins from the posttranslational modification system are also increased as well as those that compose the proteasome degradation system, membrane transport and general metabolism (Fig. 5). Nowadays, it has been observed that the regulation of gene expression during hyperthermia can be due to noncoding RNAs which can be either repressors or activators. But, gene expression regulation can also occur at the post-transcriptional level where the mRNA from its own gene produces stabilization of mRNA rather than an increase in transcription. RNA processing is also a subject for temperature inactivation. It has been observed that incubation of cells at sublethal temperatures leads to reversal inhibition of mRNA splicing. For example, in HeLa cells exposed to HS, inactivation of splicing is linked to the destruction of small nuclear ribonucleoprotein complexes (snRNP) (Bond 1988, Shukla et al. 1990). At the same time, the inhibitory effect of HS on RNA processing is not limited to destruction. It turns out that the inhibition of splicing during hyperthermia can be regulated.

Many studies have addressed the heat shock response on a genome-wide level using differential display, transcriptional profiling or proteomic approaches in a variety of cells and organisms (Eisen et al. 1998, Larkindale et al. 2008, Tabuchi et al. 2008, Matsuura et al. 2010). These studies showed that roughly 50–200 genes are significantly induced in different model organisms (Verghese et al. 2012), being most of them those codifying the so-called molecular chaperones (Fig.5). In the same work,

they observed that the conservation of specific genes across species is low with the exception of these molecular chaperone genes.

**Figure 5.** Categories of genes whose transcription rate is modulated by heat shock treatment (Weber et al. 2006).



#### 4. I Heat shock proteins

As commented above, multiple endogenous pathways are engaged in restoring cellular homeostasis. One well-characterized mechanism that facilitates protein folding and guard the proteome from the dangers of misfolding and aggregation is the heat shock family of stress proteins, or HSPs (Kampinga et al. 2009). They are expressed as response to adverse environmental or chemical stresses, such as heat or cold shock, hypoxia, salinity, heavy metals and pathophysiological situations and play important role

in cell survival (Feder et al. 1999). Insights into the mechanisms underlying HSP function is provided by two ways: HSPs take part of the basal protein folding machinery but also, HSP chaperones repair denatured proteins or promote their degradation after heat shock.

HSPs belong to multigene families that range in molecular size from 10 to 150 kDa and are found in all major cellular compartments. Their assigned name is based on their molecular weight (Kampinga et al. 2009).

Hsp90s (90kDa) account 1–2% of total cytosolic proteins present in unstressed cells (Taipale et al. 2010), making it one of the most abundant cellular proteins. When cells are heated, the fraction of heat shock proteins increases to 4–6% of (Crevel et al. 2001). Their function is dependent on the interaction with many co-chaperones (Jackson 2013). They either prevent aggregation of newly synthesized or misfolded proteins, assisting in their proper folding, or direct them for proteasomal degradation (Pearl et al. 2008, Zuehlke et al. 2010). Their client proteins are involved in signal transduction, transcription and apoptosis (Brown et al. 2007, Hartson et al. 2012). In recent years, many studies have focused on the role of this family in cancer (Whitesell et al. 2005, Trepel et al. 2010). HSP90 expression has been associated with many types of tumors including breast cancer, pancreatic carcinoma, human leukemia, systemic lupus erythematosus, as well as multidrug resistance (Csermely et al. 1998). HSP90 inhibition provides a recently developed, important pharmacological platform for anticancer therapy (Sreedhar et al. 2004).

HSP90 family is highly conserved across kingdoms except in archea where it does not exist a proper HSP90 gene, but an HSP90-like one and viruses where there is no gene homologue. Chen and co-workers (Chen et al. 2006) proposed a new nomenclature system for the HSP90 family based on the phylogeny of their proteins and the cell compartments. They divided the gene family into five subfamilies named HSP90A (cytosolic), HSP90B (Endoplasmic reticulum), HSP90C (chloroplast), TRAP (mitochondria) and HTPG (bacterial homolog). HSP90A was further divided into two classes: HSP90AA for conventional Hsp90-alpha and HSP90AB for Hsp90-beta.

Hsp90 is required for activity of all clients, but there are differences in when and how Hsp90 is required. Hsp90 is continually required to maintain steroid hormone receptors in a nearly completely folded conformation capable of ligand binding. Some, but not all, protein kinases become unstable soon after Hsp90 function is reduced.

However, Hsp90 is continuously required for the activity, but not stability of the v-src kinase, and the p56-lck kinase requires Hsp90 only during synthesis. A conserved function of Hsp90 is to promote assembly of multiprotein complexes (Johnson 2012).

#### **4.1.1 HSP90AA1**

The genes encoding Hsp90 family members have undergone multiple duplication events. Most eukaryotic species contain at least two genes encoding highly homologous isoforms of cytoplasmic Hsp90 (HSP90A), which likely arose from separate gene duplication events (Gupta 1995). Hsp90AA and AB are 76% identical and are the consequence of a gene duplication about 500 million years ago (Krone et al. 1994). The major differences between them is that HSP90AA gene is inducible while HSP90AB is constitutively expressed (Chen et al. 2005).

Functional promiscuity in HSP90A is promoted by its history of duplication events, which provides raw genetic material to evolve novel gene and gene functions (Qian et al. 2009). Expression divergence after *HSP90A* gene duplication has been reported previously; in a number of animal species, *HSP90AA* is strongly upregulated in response to elevated temperatures, whereas *HSP90AB* is not induced during heat shock (Meng et al. 1993, Krone et al. 1994, Carretero-Paulet et al. 2013). In *Arabidopsis thaliana*, where four *HSP90A* genes have been reported, only one is heat inducible, whereas the rest are constitutively expressed (Carretero-Paulet et al. 2013). Similar evidence of expression divergence between *HSP82* and *HSC82* has been found in *Saccharomyces cerevisiae* (Borkovich et al. 1989).

Carretero-Paulet and colleagues (Carretero-Paulet et al. 2013) pointed out that over 55 species studied in their work, 22 showed *HSP90A* multigene families which were formed by up to six members. Phylogenetic analysis confirmed the occurrence of independent HSP90A duplication events in higher order eukaryotic lineages, such as vertebrates, fungi species from the *Saccharomyces* group, and seed plants.

Focusing on the inducible form, *HSP90AA*, which is in this memory of major interest, in mammalian cells there are two or more genes encoding cytosolic Hsp90 homologues (Chen et al. 2006). For example, in human there are 6 gene variants (*HSP90AA1-HSP90AA6*) (Chen et al. 2005), while in sheep it is only described one (*HSP90AA1*).

It is *HSP90AA1* in sheep, located in OAR 18, by which this work is about, and more accurately about its expression pattern which is mainly due to some features existing at its promoter region. Its promoter is highly polymorphic (Marcos-Carcavilla et al. 2008), which could be the clue of the variable behaviour of the gene, at least in sheep.

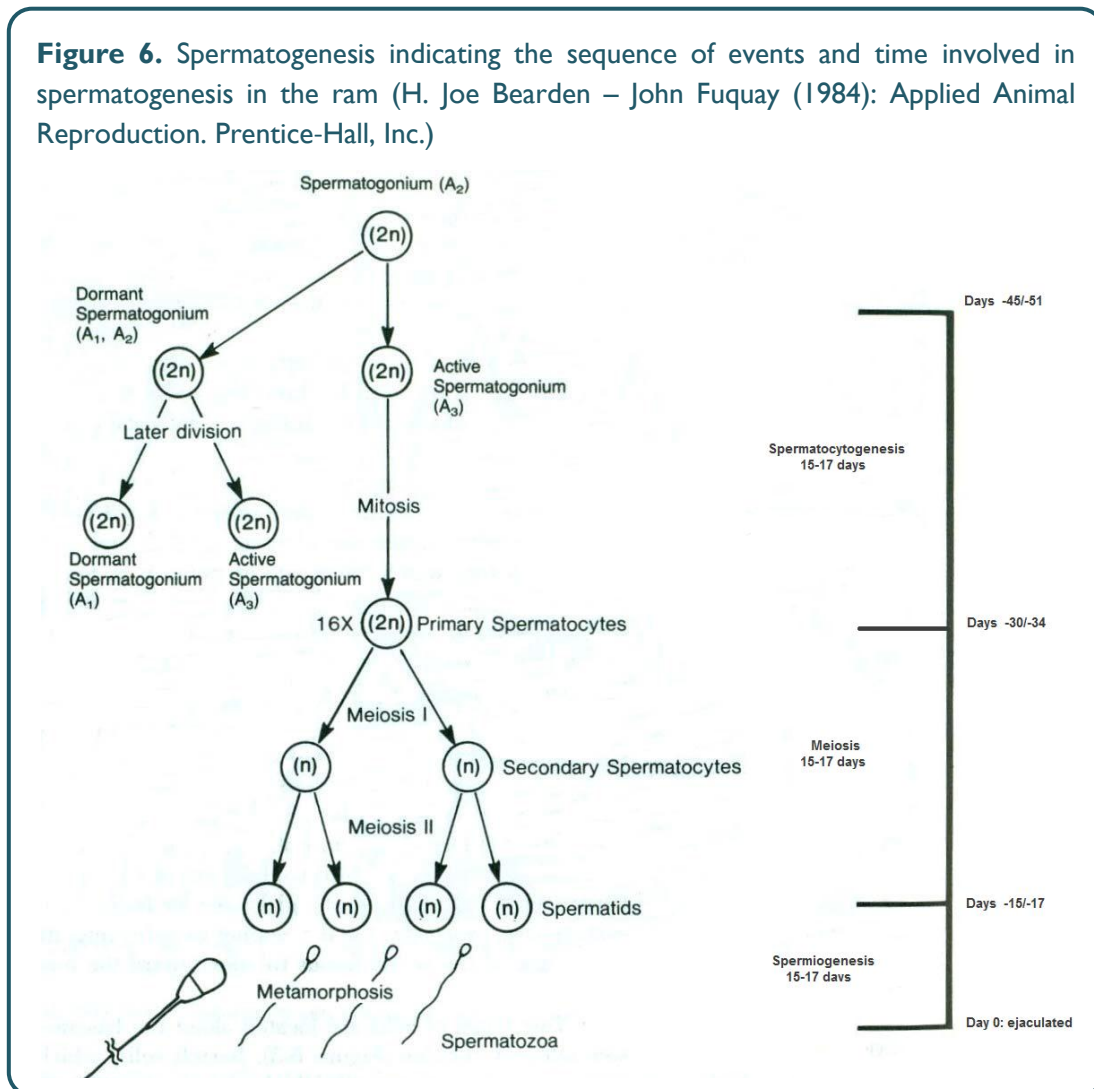
## **5. Germ cell development in males: Spermatogenesis**

Spermatogenesis is the process by which sperm cells are produced. It is a complex differentiation program produced in the scrotum that starts after birth when spermatogonial stem cells enter the differentiation pathway. The process starts (spermatocytogenesis, Figure 6) when differentiating spermatogonia undergo subsequent mitotic divisions to expand the colony of differentiating germ cells. After the proliferation phase, germ cells initiate meiosis I as spermatocytes during which homologous chromosomes pair and synapse and homologous recombination occurs.

Meiosis is a type of cell division that is performed only by germ cells to form haploid gametes. The switch from mitosis to meiosis exhibits distinct sex-specific difference in timing, with female germ cells entering meiosis during fetal development and male germ cells at puberty when spermatogenesis is initiated. During early fetal development, potential primordial germ cells migrate to the forming gonad where they remain sexually indifferent until the sex-specific differentiation of germ cells is initiated by signals produced by the somatic cells. This irreversible step in gonadal sex differentiation involves the prevention of meiosis in the germ cells of fetal testes (Jorgensen et al. 2014).

The process continues at puberty, resulting from the increased levels of luteinizing hormone, follicle-stimulating hormone and testosterone in rams (Courrot et al. 1981). Manchega ram lambs reach puberty between 5 and 8 months of age, when they reach 50 to 60 percent of their mature weight. Subsequently, spermatocytes undergo a reduction division to split the sister chromosomes into two cells to generate secondary spermatocytes. These cells divide again very quickly, and the resulting haploid round spermatids commence the differentiation phase (spermiogenesis) to construct sperm-specific structures, such as a flagellum and an acrosome, as well as to reshape the nucleus and compact the chromatin with the help of sperm-specific protamines that replace most of the histones (Conwell et al. 2003).

In this stage, unique and dynamic histone exchanges occur resulting in a dense compacting (toroid) that gives protection against exogenous assault to the sperm DNA (Barratt et al. 2010). The timing and types of histone exchanges defines the particular stages of spermatogenesis (Makino et al. 2014). Spermatozoa, which are released into



the tubular lumen continue their journey to the epididymis for final maturation and storage.

### 5.1 Gene expression during spermatogenesis

In the mid-Sixties, Olivieri and Olivieri (Olivieri et al. 1965) demonstrated RNA transcription during mitosis and early meiosis but no transcriptional activity during postmeiosis was observed. Based on these findings and other similar studies, it was generally thought that there was absence of transcriptional activity during the

postmeiotic phase and there was only a repository of mRNAs which were lately translated.

New recent works demonstrated mRNA accumulation in postmeiotic spermatids (Barreau et al. 2008) and direct evidence for the production of nascent RNA in primary spermatocytes during postmeiosis, even though in lower dosis (Vibrantovski et al. 2010).

However, even though sperm also carry thousands of different RNAs, sperm RNA is unlikely to be transcribed from sperm nuclear DNA because of the changes in chromatin structure that occur when protamines replace histones during sperm DNA compaction. The RNA population carried by sperm is large and varied. It includes messenger RNA, microRNA, interference RNA, and antisense RNA (Dadoune 2009). These include transcripts for heat shock proteins, cytochrome P450 aromatase, and a range of receptors, including odor receptors. They are relics of what has been expressed during spermatogenesis. Not only genes needed at that moment for normal development or induced by alteration of normal conditions. But also, those RNAs needed in later stages (sperm-specific transcripts) where gene expression is inactive but translational activity in the sperm cells has been observed (Fischer et al. 2012). In addition, there is evidence that those RNAs contribute to fertilization and to embryo development (Boerke et al. 2007).

## **5.2 Effects of heat on male germ cells**

The function of sperm is to safely transport the haploid paternal genome to the egg containing the maternal genome. An optimal scrotal temperature is critical in the development of correct mature spermatozoa as heat has an adverse effect on mammalian spermatogenesis and eventually leads to sub- or infertility (Kim et al. 2013a). That is why, in most mammals, the scrotum is located outside the body cavity.

In rams, published data have demonstrated an adverse effect of experimental heat stress on sperm motility, morphology, and fertilizing capacity (Malama et al. 2013). The most significant consequence of heat stress on testis is the loss of germ cells via apoptosis. Sperm quality has also shown to suffer, with a reduction in progressive sperm motility and a significantly lower *in vitro* fertilization rate of oocytes by sperm from heat-shocked males. However, germ cell apoptosis in response to heat stress occurs in a developmental stage-specific manner and some defective cells achieve to



avoid this defensive stage and continue along the spermatogenesis process (Liu 2010). These damaged spermatocytes avoid heat-induced apoptosis and develop to mature spermatozoa with defects in their chromatin integrity.

In addition, sperm damage is influenced by the stage at which germ cells are exposed to stress. It has been shown that the most affected germ cell types are the primary spermatocytes through Meiosis I (specifically, pachytene and diplotene) and the early round spermatids (Setchell 1998). In addition, Hales and coworkers (Hales et al. 2005) observed in an experiment with male rats, that after chemical high dose exposure heat shock proteins were expressed predominantly in both pachytene spermatocytes and round spermatids. They proposed that such gene expression induction may suggest that the cell activates mechanisms of defense trying to cope with the damage induced by the chemical agent.

## **6. Background of the present study**

### **6.1 Scrapie**

Transmissible Spongiform Encephalopathies (TSEs) are a group of fatal neurodegenerative diseases that affect humans and various animal species. The main TSE affecting animals are scrapie sheep and goats, bovine spongiform encephalopathy (BSE), transmissible mink encephalopathy, Chronic Illness Cachectic deer and elk, and feline spongiform encephalopathy. This group of diseases share several common features, among which is the possibility of being transmitted to individuals of the same or other species, long incubation periods, a progressive clinical picture evolves slowly without remission, and microscopic lesions characterized by a spongiform degeneration or localized vacuolization in the central nervous system (Imran et al. 2011).

The first reported TSE is scrapie sheep, described in Spanish merino sheep in the United Kingdom in 1732 (Liberski 2012). Since then, scrapie research was intensified as many cases were identified and the sheep industry underwent substantial financial losses. These losses prompted also studies on the true nature of the infectious agent. The conversion of a normal cellular protein ( $\text{PrP}^c$ ) into a pathological isoform ( $\text{PrP}^{\text{sc}}$ ) as key event of TSE pathogenesis was postulated in 1985 (Oesch et al. 1985). The common and specific feature of TSEs is the accumulation of  $\text{PrP}^{\text{sc}}$  in the central nervous system, an abnormal isoform of a normal cellular protein called  $\text{PrP}^c$ .

However, the prion theory is still debated since PrP<sup>Sc</sup> is not always infectious and the phenomenon of strains is still an enigma (Christine Fast D.V.M. 2013, Gasperini et al. 2014).

The number of variables influencing the susceptibility to scrapie are high and depend not only on the genotype of the host and the infectious agent but also on individual flocks, breeds and geographical location, without forgetting dose and route of inoculation effects (Christine Fast D.V.M. 2013). It has been published different research lines not only centered in the study of the PRNP gene, the one that codifies PrP<sup>C</sup> and PrP<sup>Sc</sup> proteins as 21% of the variability on response does not depend on this gene (Diaz et al. 2005). These works identified different QTLs associated with the variability of scrapie incubation time in mouse (Manolakou et al. 2001, Lloyd et al. 2002, Moreno et al. 2003).

One of the candidate genes obtained from the regions enclosed previously was *HSP90AA1*. As it has been mentioned before, one of the principal roles of the protein codified by this gene, Hsp90 $\alpha$ , is to repair aberrant proteins. Hsp90 $\alpha$  is involved in the proper conformation of other proteins assisting in the folding of nascent proteins and in the refolding of damaged ones. Hsp90 $\alpha$  also prevents them against aggregation, which is the principal problem found in the scrapie disease.

Marcos-Carcavilla and coworkers (Marcos-Carcavilla et al. 2008, Marcos-Carcavilla et al. 2010a) identified 34 polymorphisms (12 at the coding region, 14 at the promoter -Figure 7- and 8 at the intron 10) at the *HSP90AA1* gene. One of them, located at the promoter region of *HSP90AA1* gene (g.660G>C) was associated with differences in the scrapie incubation period.

## **6.2 Heat stress adaptation**

Moreover, in another work Marcos-Carcavilla and colleagues (Marcos-Carcavilla et al. 2010b) also identified that the same polymorphism located at the promoter region of the *HSP90AA1* gene was associated with differences in gene expression under different thermal environments and with the adaptation pattern of different sheep breeds to the thermal conditions in where they are reared. However, the study was based on a limited number of animals and on standard basic statistical methods used to analyze quantitative real-time PCR (qPCR) data to assess differences in expression levels of alternative genotypes under control and heat stress conditions.

Further on, also a new INDEL (insertion/deletion) (g.703\_704insAA, Figure 7), was discovered short-after (Oner et al. 2012).

**Figure 7.** Sequence of the *HSP90AA1* ovine promoter and the polymorphisms detected at it. Intronic sequence in lower case and exons in capital letters. SNPs are in square brackets. The 7 SNPs of interest (polymorphic) (g.660G>C, g.601A>C, g.528G>A, g.524G>T, g.522A>G and g.444A>G) are in grey. INDELs are in brackets and squared. Putative methylated sites are circled. Initiation of transcription (TATA box) and translation (ATG) in bold. A HSE already detected is underlined. Modified from Marcos-Carcavilla and coworkers (Marcos-Carcavilla et al. 2008).

```

-1138  cctggcgggctaccgtcgcagagttggacacgactgagcgactaagcacagcacaagtaggta
        caaaggctcaaagaactaagcaaaggccgggagaatcgggactgtgtccccttttaacattcc
        ccgcgccaccgtcccctgccttgggcgcaatcgggcctaaggacaggacgagtagccatactg
        gagtggcgggggaatggtggttgttccaggggccacctctggttcagaaccagccgaatagac
        ccacgaggctctggagtcccccctctgctccccctgcacacacgctggtcaccctcgcctcacag
        accgctgcaccccacaaccaggctggggctctgagcctcgcgcttaacgaggetctggcaggca
        cttgttgcgcgaatcctagggccctgaacctataacgctgcccctggg(aa)agagatgctc
        cctagaccctctaatacgtgcccga(c)(c)cccc[G/C]aaacctaaacatggacctg
        agcctgcatcaatgccccaaagcatatacctgtgaca[A/C]gggactggaccctccccggc
        ttgatcccctaagcgcggcctcaaggctcgaacctagcccaggccc[G/A]tag[G/T]c[
        A/G]gggg(g)ccccgatgctgggggatccggagcgcacactatctggggacacgctt[G/T
        ]aaccaccgagagtttactttct[A/G]gacttccagagctgctggcagggcgatcagccc
        ccggggcctggacagaggtttgggtcgggggagcgttaggcccgcgcgccaagcggggttg
        cgggccagcgggtgtagggccggcggggcgcagaacccgc[A/G]cgccggg[A/G][C/T]ag
        gtagggctgggaacggcggcaccggaaggcgcgagaccgc[C/G]cggaacgctccggatagcg
        caggcgcactgcgccgggcgggcacgcttggcacgcaggcgcctgaggcgtgggagcgcgc
        ggccggccccgagagggggcgggcggcctccagtgcgcaggcgcaggcgcggccgctccg
        gcggttggggagggttcttcggaaggttcgggaggetctctggaaaggcgcgcgctgggc
        gggccccctctatataaaggcgcggggcggcgcAGTTGCCTCGCGTCCCGGTGCT
        GCGTCGCTGCGGAGCCCGTGTAGTCTCCTCAGCCAAGgtgaggggaggggtttaggggacg
        cgggatccacggcgggtggcgggcggcaggcagccccggagccccctgaggccggcggg
        cccggggcggcctgattaaccggtaggcctcgggcctgcggcggctctcccgtcccctcccc
        ccgcccctcccccgctcgggagcagtagggccccgggctccgggctcgtcgggatggtga
        gggccggggtctgcggccccgaggggggtccccggcgcgggtgggggcggccggcggcgcg
        ggcggcctgcgccggcgtgggcggcggcgcagcggcggcggcggccccgggaatgaggtc
        atccttgcagccggcctcccgcgcggcggcctcgggagcgggagcgcgagcgcggcggg
        gaggggctccggcgggtggcctcacgggctccgcttccccggcagATGCCCGAGGAGACCAGG
    
```

# Objetivos



El objetivo general de la tesis ha consistido en la profundización del conocimiento de la estructura del promotor del gen ovino *HSP90AA1* que codifica la proteína de estrés térmico Hsp90 $\alpha$  y de la implicación de los polimorfismos presentes en dicho promotor sobre la regulación de su transcripción en el contexto de la respuesta al estrés por calor en la especie ovina y su posible efecto sobre un carácter reproductivo en machos de esta especie. Este objetivo general se ha llevado a cabo abordando los siguientes objetivos específicos:

1. Análisis de expresión génica utilizando un número de muestras representativo de la población para poder valorar el efecto que los diferentes polimorfismos presentes en dicho promotor tienen sobre la expresión génica bajo condiciones de estrés térmico.
2. Estudio de la estructura del promotor, mediante la identificación de todos los elementos implicados en la regulación de su transcripción. Estudio funcional *in vitro* de los polimorfismos candidatos en relación con las diferencias de expresión observadas y análisis de las marcas epigenéticas, posibles co-responsables de dichos cambios de expresión observados.
3. Estudio de la distribución de los polimorfismos detectados en el promotor del gen *HSP90AA1* en 31 razas ovinas originarias de Europa, Asia y África para determinar la posible correlación de sus frecuencias con las variables climáticas de sus regiones de origen, y búsqueda del origen de dichas mutaciones en especies de la subfamilia Caprinae relacionadas filogenéticamente con la especie ovina.
4. Estudio de asociación para determinar el efecto que las diferencias de expresión pudieran causar sobre un carácter reproductivo ligado a la fertilidad de los machos de la especie ovina, en concreto, la fragmentación de ADN espermático.



# **Aim of the thesis**





The overall aim of the thesis was to deepen the knowledge of the ovine *HSP90AA1* gene promoter structure, encoding the heat stress protein Hsp90 $\alpha$ . Also, to study the involvement of several polymorphisms present at its promoter region in the regulation of transcription in response to heat stress in sheep. This general objective has been conducted to address the following specific objectives:

1. To perform a gene expression analysis using a representative population sample number in order to assess the role of several polymorphisms present at the promoter region in gene expression under heat stress conditions.
2. Study the structure of the promoter, by identifying all the elements involved in the regulation of transcription. To carry out an *in vitro* functional study of candidate polymorphisms related with the expression differences previously observed. To analyze epigenetic marks, possible co-responsible of the observed expression changes.
3. Study the distribution of the polymorphisms identified at the *HSP90AA1* promoter in 31 sheep breeds from Europe, Asia and Africa. To determine the possible correlation of their frequency with climatic variables in their regions of origin. In addition, to study the evolvability of these mutations in species from the Caprinae subfamily phylogenetically related to sheep.
4. To develop an association study to determine the effect that differences in expression could cause on a reproductive character linked to the male fertility in sheep, namely sperm DNA fragmentation rate.



# Chapter I

## Gene expression analysis:

Ovine *HSP90AA1* expression rate is affected by several polymorphisms at the promoter under both basal and heat stress conditions



## INTRODUCTION

The genetic variability underlying animal's thermo tolerance could be exploited in livestock breeding programs to achieve animals that could cope with the effects of heat stress over productive and functional traits. Among the livestock animals, sheep is an interesting biological material to study the genetic basis of thermo-tolerance. There is some literature about heat stress effects over physiological and productive traits in cattle (Morton et al. 2007, Aguilar et al. 2009, Sanchez et al. 2009) and sheep (Finocchiaro et al. 2005, Marai et al. 2007, 2008, Sevi et al. 2012). Also, at the molecular level, genes involved in the heat stress response have been described (Favatier et al. 1997, Trinklein et al. 2004, Collier et al. 2008), (Charoensook et al. 2012). Among them, those encoding heat shock proteins have been the most studied. However, in sheep there are few works regarding this topic.

As it was detailed previously, a prior work (Marcos-Carcavilla et al. 2010b) pointed out a candidate SNP (g.660G>C) located at the promoter region of *HSP90AA1* to be related with differences in the transcription rate of this gene. However, this work had several methodological and sample size limitations.

In general, qPCR has provided a powerful tool for quantifying gene expression. Nevertheless, it makes necessary to carefully consider some technical and analytical factors to ensure reproducible and accurate measurements and not lead to misinterpretations (Dheda et al. 2005). However, commonly, several of these essential procedures have been widely ignored. Those technical factors include the initial sample amount, RNA recovery and RNA purity and integrity (Fleige et al. 2006a), (Bustin et al. 2009) among others. Some factors considered as analytical are the selection of the suitable housekeeping gene(s) (HK), the experimental design (Auer et al. 2010), the statistical method used, etc.

Traditional statistical analyses have been restricted to pair-wise comparisons of treatments in which Cq (quantification cycle) values of GOIs (genes of interest) were previously normalized using standard HKs. This kind of approach does not allow to include technical and biological effects having influence over gene expression data. The joint analysis of GOIs and HKs data can lead to a better partition of such sources of variation (Steibel et al. 2009) and allows checking HK stability and subsequent normalization of GOIs simultaneously. Mixed model methodology makes possible this kind of approach, giving the possibility of including systematic and random effects, and interactions among them. They constitute a powerful tool in qPCR analyses including more than two treatments and multiple experimental factors (Lim et al. 2012), (Arceo et al. 2012).

The objectives of this chapter were to 1) simultaneously select the best HK and analyze expression data using a mixed model statistical approach that includes technical and biological sources of variation; 2) determine the linkage disequilibrium (LD) among all the polymorphisms detected at the gene promoter to establish LD cosegregating blocks; 3) study gene expression differences observed for alternative genotypes of the g.660G>C, g.703\_704insAA, g.601A>C, g.528G>A, g.524G>T, g.522A>G, g.468G>T and g.444A>G polymorphisms and from some INDELS not yet detected and studied (g.667\_668insC, g.666\_667insC, g.516\_517insG) using more animals and a wider range of climatic conditions than those in (Marcos-Carcavilla et al. 2010b).

## MATERIALS AND METHODS

### Linkage disequilibrium analysis

#### Animal material, nucleic acid isolation, DNA amplification and SNPs genotyping.

Peripheral whole blood samples were collected from 103 animals of the Manchega Spanish sheep breed in order to analyse linkage disequilibrium among the 11 polymorphisms of interest located at the *HSP90AA1* promoter. Animals were grouped in 48 parent-offspring trios. Trios consist of 10 sires, 48 dams and 1 offspring per pair (all females). Three of the dams were also daughters from another trio. Genomic DNA was extracted from lymphocytes according to the salting out procedure (Miller et al. 1988). The polymerase chain reaction was performed from 100 ng of genomic DNA using CERTAMP complex amplifications kit chemistry (Biotools, Madrid, Spain) with specific primers (Forward: 5'CGAGGCTCTGGCAGGCACTTGTTG3' and Reverse: 5' GCCGCCGTTCCCA GCCCTACCT 3'). A 499bp fragment of the promoter containing the g.660G>C SNP and 7 more polymorphisms (g.703\_704insAA, g.601A>C, g.528G>A, g.524G>T, g.522A>G, g.468G>T and g.444A>G) was obtained. The resulting PCR fragment was purified with ExoSAP-IT (USB Corporation, OH, USA) and sequenced with specific primers (shown above).

Due to problems in genotyping and reading, INDELS, g.667\_668insC, g.666\_667insC, g.516\_517insG were genotyped with additional primers (Sequencing g.667\_668insC and g.666\_667insC: 5'GCTAGGTTTCGAGCCTTGAGG3' and for sequencing g.516\_517insG: 5'AAGCGTGTCCCCAGATAGTG3').

#### Linkage disequilibrium estimation.

PLINK software (Purcell et al. 2007b) (<http://pngu.mgh.harvard.edu/purcell/plink/>) was used to estimate linkage disequilibrium among all pairs of the 11 polymorphisms measured as  $r^2$ , the squared correlation based on genotypic allele counts (Hill et al. 1968). Hardy-Weinberg equilibrium exact test and observed and expected heterozygosities for each polymorphism were also calculated using PLINK.

### **Detection of putative transcription factor (TF) binding sites in ovine *HSP90AA1* promoter**

Putative TF binding sites were predicted using TESS (Schug 2008) (keeping default settings) and ALGGEN-PROMO (Messeguer et al. 2002), (Farre et al. 2003) (limiting to mammal transcription factors) softwares. See Table S1.

### **Expression analysis**

#### *Animal material*

In order to confirm the association of the *HSP90AA1* polymorphism (g.660G>C) with the adaptation to different thermal conditions in sheep previously described by Marcos-Carcavilla and coworkers (Marcos-Carcavilla et al. 2010b), 428 unrelated rams of Manchega Spanish sheep breed were genotyped (same protocol and primers as described in (Marcos-Carcavilla et al. 2010a)). All animals belonged to an artificial insemination centre, and therefore they were reared under the same environmental and management conditions. A total of 120 out of 428 rams were selected based on their genotype: 40 CC<sub>.660</sub>, 40 CG<sub>.660</sub> and 40 GG<sub>.660</sub>. Genomic DNA from these 120 animals was used to genotype the previously defined 499bp amplicon of the *HSP90AA1* promoter. Genotype frequencies are shown in Table 1.

Peripheral whole blood samples from the 120 rams selected, were collected in 4 time points, corresponding to different climatic conditions in a dry region of central Spain (Ciudad Real). The 4 time points were in March, when environmental temperature conditions are mild, and in July and August (2 sample collections August 1 and August 2) when heat stress temperatures occur. Hereafter, we will refer to the March collection as the control. The temperature humidity index (THI) equation proposed by Marai et al. (Marai et al. 2007) was used as another indicator of thermal stress. This index combines both temperature and relative humidity. The environmental parameters for the 4 collections are shown in Table 2.



**Table I.** Genotype frequencies of the polymorphisms located at the *HSP90AA1* promoter in 120 rams of Manchega sheep breed.

	g.703_704insAA	g.667_668insC	g.666_667insC	g.660G>C	g.601A>C	g.528G>A	g.524G>T	g.522A>G	g.516_517insG	g.468G>T	g.444A>G	N	Freq (%)
ID	DD	DD	CG	CC	GG	TT	GG	ID	TT	GG		2	1,64
ID	DD	DD	CG	CC	AG	TT	AG	DD	TT	GG		5	4,10
DD	DD	DD	GG	CC	GG	TT	GG	DD	TT	GG		29	23,77
II	DD	DD	CC	CC	AA	TT	GG	II	TT	GG		2	1,64
ID	DD	DD	CG	CC	AG	TT	GG	ID	TT	GG		12	9,84
II	DD	DD	CC	CC	AG	TT	GG	ID	TT	GG		2	1,64
II	DD	DD	CC	CC	AG	TT	AG	ID	TT	GG		1	0,82
DD	DD	DD	GG	CC	AG	TT	GG	DD	TT	GG		9	7,38
ID	DD	DD	CG	CC	AG	TT	GG	DD	TT	GG		2	1,64
II	DD	DD	CC	CC	GG	TT	GG	ID	TT	GG		1	0,82
ID	DD	DD	CG	AC	AG	GT	GG	DD	GT	GG		4	3,28
ID	DD	DD	CG	CC	AA	TT	GG	ID	TT	GG		2	1,64
II	DD	DD	CC	CC	AA	TT	GG	ID	TT	GG		7	5,74
ID	DD	DD	CG	CC	GG	TT	GG	DD	TT	GG		2	1,64
II	DD	DD	CC	AC	AA	GT	GG	ID	GT	GG		1	0,82
II	DD	DD	CC	CC	AA	TT	AG	ID	TT	GG		1	0,82
ID	DD	DD	CG	AC	AA	GT	GG	DD	GT	GG		1	0,82
ID	DD	DD	CG	CC	AG	GT	GG	DD	GT	GG		1	0,82
II	DD	DD	CC	CC	AA	GT	GG	ID	GT	GG		1	0,82
ID	DD	DD	CG	CC	AA	TT	GG	ID	TT	AG		1	0,82
DD	DD	DD	GG	CC	AG	TT	GG	DD	TT	AG		2	1,64
II	DD	DD	CC	CC	AA	TT	AG	DD	TT	GG		1	0,82
II	DD	DD	CC	CC	AA	TT	GG	ID	GT	GG		1	0,82
DD	DD	DD	GG	CC	AA	TT	GG	DD	TT	AG		1	0,82
DD	DD	DD	GG	CC	AA	TT	GG	DD	TT	GG		1	0,82
II	DD	DD	CC	CC	AA	TT	GG	ID	TT	AG		1	0,82
II	ID	DD	CC	CC	AA	TT	GG	ID	TT	GG		10	8,20
II	ID	DD	CC	AC	AA	GT	GG	DD	GT	GG		1	0,82
ID	ID	DD	CG	CC	AG	TT	GG	DD	TT	GG		7	5,74
II	ID	DD	CC	AC	AG	GT	GG	DD	GT	GG		1	0,82
II	ID	DD	CC	CC	AG	TT	AG	DD	TT	GG		1	0,82
II	ID	DD	CC	AC	AA	TT	GG	ID	TT	GG		1	0,82
II	ID	DD	CC	CC	AA	TT	AG	DD	TT	GG		1	0,82
II	ID	DD	CC	CC	AA	TT	GG	DD	TT	GG		1	0,82
II	II	DD	CC	CC	AA	TT	GG	DD	TT	GG		5	4,10
II	ID	ID	CC	CC	AA	TT	GG	DD	TT	AA		1	0,82

**Table 2.** Climate parameters existing at day of blood samples collection.

blood collection date	AvT	MaT	MiT	Rh	Rhmax	Rhmin	THlavr	THlmax	treatment ID
23/03/2010	11.6	19.9	3.8	69.0	92.6	37.2	11.87	19.77	Control
05/07/2010	26.8	35.0	16.8	39.4	63.9	19.8	24.47	32.69	July
03/08/2010	24.7	34.4	16.6	49.4	89.3	21.0	23.08	33.74	August 1
09/08/2010	27.3	33.8	22.2	50.0	71.5	28.2	25.30	32.09	August 2

From: Manzanares (Ciudad Real) Meteorological Station, coordinates 654m-38° 59'47N-03° 22'23W (<http://crea.uclm.es/siar>)

AvT = average temperature (°C)

MaT = maximum temperature (°C)

MiT = minimum temperature (°C)

Rh = relative humidity (%)

Rhmax = maximum relative humidity (%)

Rhmin = minimum relative humidity (%)

THlavr = THI calculated with the average temperature and relative humidity

THlmax = THI calculated with the maximum temperature and the maximum relative humidity

Temperature humidity index (THI) calculated as  $THI = T^{\circ}C - ((0.31 - 0.31RH) (T^{\circ}C - 14.4))$ . T = temperature in °C; RH = relative humidity in %/100 (Marai et al. 2007).

#### Total RNA isolation and cDNA synthesis

Total RNA was isolated from 9 ml of whole blood using the LeukoLock™ kit (Ambion, Inc., TX, USA), following manufacturers instructions. RNA concentration was determined using a NanoDrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies, Inc., DE, USA). Degradation of RNA samples was assessed with the Agilent 2100 bionalyzer (Agilent Technologies Hewlett-Packard-Str.8 76337 Waldbronn, Germany) in RNA Nano Chips, following manufacturers instructions. RIN (RNA Integrity Number) values were obtained. cDNA was synthesized using the ImProm-II™ Reverse Transcription System (Promega Corp., WI, USA).

#### Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

qRT-PCR was performed on all samples collected. Three HKs were tested, *MDHI*, *SDHA* and *HSP90ABI*. *MDHI* and *SDHA* became the most stable HK pair for the heat stress response in sheep under similar conditions (Serrano et al. 2011). Also the *HSP90ABI* gene was included as HK candidate since its expression is ubiquitous, less inducible and more constitutive than that of the *HSP90AAI* gene (Csermely et al. 1998), (Deuerling et al. 2003). Primers were designed with NetPrimer software (Biosoft International, CA, USA), and are listed in Table 3 together with amplicon sizes

and CG content. Primers were designed avoiding possible genomic DNA amplifications. *In silico* specificity of the amplicons was screened by BLAST searches.

qRT-PCR amplification reactions were performed from 100 ng of cDNA using LightCycler® 480 SYBR Green I Master kit (Roche, Switzerland). Reactions were run in triplicate on a LightCycler® 480 (Roche, Switzerland) following manufacturer's cycling parameters. Dissociation curves were performed for each gene to check primer specificity and to confirm the presence of a unique PCR product. The corresponding mRNA levels were measured and analyzed by their Cq.

To estimate PCR efficiencies, standard curves based on 6 serial dilutions (1/20 from a departure concentration of 50 ng/μl) of a cDNA stock (a cDNA mixture of more than 121 samples accounting for the 3 genotypes and the 4 time points) were performed. Efficiencies (E) were calculated from the slope of curves as in Rasmussen and coworkers (Rasmussen 2001). Estimated E for each gene are shown in Table 3.

**Table 3.** Primers and efficiencies of the qPCR reactions.

Gene	Forward primer (5'-3') Reverse primer (5'-3')	Amplicon size (bp)	Efficiencies	Amplicon % bases and GC content
<i>HSP90AA1</i>	CCACTTGGCGGTCAAGCATT AAGGAGCTCGTCTTGGGACAA	80	1.951	A/22.50 G/25.00 C/31.25 T/21.25 GC content 47.50
<i>MDHI</i>	GGTCAAATTGCATATTTCACTACTA ACCATCCAGGACACCCATCAT	117	1.883	A/23.07 G/20.51 C/32.48 T/23.93 GC content 43.58
<i>SDHA</i>	GGCATCCCCACCAACTACA TACACCACCTCAAAGCCCCG	134	2.000	A/35.55 G/29.62 C/17.03 T/17.77 GC content 65.17
<i>HSP90AB1</i>	TACATCACTGGTAAGAGCAAAGA TACACCACCTCAAAGCCCCG	81	1.950	A/37.03 G/18.52 C/22.22 T/22.22 GC content 55.55

## Statistical procedures

### Statistical analysis of RIN values

A mixed model was fitted by using the MIXED procedure of the SAS statistical package (Littell 2006) for determining factors affecting RIN values. RIN values of all samples were included as a dependent variable. Fixed effects included were g.660G>C genotype (G) - 3 levels: CC, GC and GG -; date of collection (D) - 4 levels: Control, July, August 1 and August 2 -; group of sample processing (GP) - 4 levels corresponding to the barn where a group of animals were located and sampled - and the interaction date of collection x group of sample processing (DxGP) were included as fixed effects. The barn needs to be included because it is related to the period of time between samples collection and processing. The animal (A) was included as random effect. Goodness of fit statistics AIC (Akaike's Information Criterion) and BIC (Schwarz's Bayesian Criterion) were used as criteria for model selection. A type III fixed effects test was used to determine significance of the effects included in the model.  $P < 0.05$  was established as threshold for statistical significance.

### HK selection

HK selection among *HSP90ABI*, *MDHI* and *SDHA* genes followed the strategy from Serrano et al. (Serrano et al. 2011), including also the GOI in the analysis. As amplification efficiencies of some genes were  $< 2$  ( $< 100\%$ ), Cq data were transformed using the equation proposed by Steibel et al. (Steibel et al. 2009) to rescale Cq values.

The equation of the mixed model used was the following:

$$y_{oijkmr} = M_o + T_i + G_j + P_k + b(RG)_{imj} + S_{im} + A_m + MTG_{oij} + e_{oijkmr} \quad \text{I}$$

where  $y_{oijkmr}$  is the transformed Cq data of the  $j^{\text{th}}$  gene, from the  $r^{\text{th}}$  well, in the  $k^{\text{th}}$  plate, collected from the  $m^{\text{th}}$  animal under the  $i^{\text{th}}$  treatment;  $M_o$  is the fixed effect of the  $o^{\text{th}}$  genotype;  $T_i$  is the fixed effect of  $i^{\text{th}}$  treatment;  $G_j$  is the fixed effect of the  $j^{\text{th}}$  gene;  $P_k$  is the effect of the  $k^{\text{th}}$  plate;  $b(RG)_{imj}$  is the interaction between the RIN value of the  $m^{\text{th}}$  sample and the  $j^{\text{th}}$  gene, and  $b$  is the regression coefficient of RIN x gene variable on Cq;  $S_{im}$  is the random effect of the biological sample ( $\sim (0, 2)$ );  $A_m$  is the random effect of the animal from where samples were collected ( $\sim (0, 2)$ );  $MTG_{oij}$  is the random interaction effect among the  $o^{\text{th}}$  genotype, the  $i^{\text{th}}$  treatment and the  $j^{\text{th}}$  gene ( $\sim (0, 2)$ );  $e_{oijkmr}$  is the random residual. Gene specific residual variance (heterogeneous residual) was fitted to the gene by treatment effect ( $\sim (0, 2)$ ).

Expression stability values were obtained by calculating the Mean Square Error (MSE), which was defined as in (Serrano et al. 2011).

### Analysis of expression results

Statistical analysis of gene expression was carried out following the method proposed by Steibel et al. (Steibel et al. 2009). As amplification efficiencies for *HSP90AA1*, *HSP90ABI* and *MDHI* genes were  $< 2$ , Cq data were transformed as aforementioned. The mixed model fitted was:

$$y_{oijkmr} = MTG_{oij} + P_k + b(RG)_{imj} + S_{im} + A_m + e_{oijkmr} \quad \mathbf{2}$$

where effects were as in model 1, except that in this case the MTG factor was included in the model as fixed effect and the residual variance was heterogeneous for the gene effect ( $\sim (0, \sigma^2)$ ).

To test differences,  $diff_{GOI}$ , in the expression rate of alternative genotypes and to obtain fold change (FC) values from the estimated MTG differences, the approach suggested in (Steibel et al. 2009) was used. Significance of  $diff_{GOI}$  estimates was determined with the  $t$  statistic.

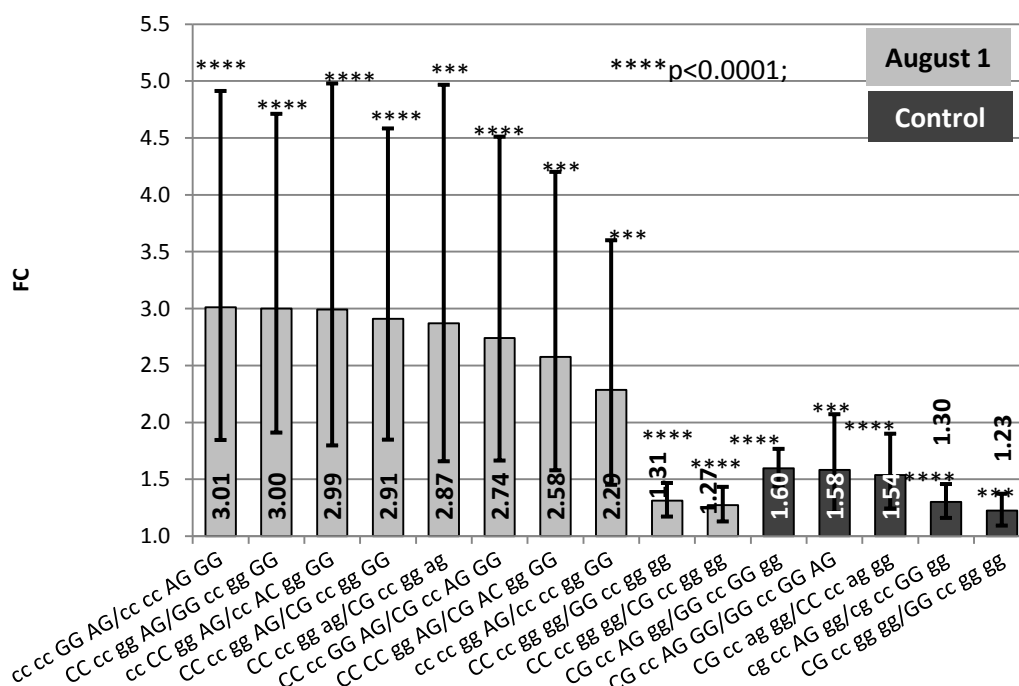
Also asymmetric 95% confidence intervals (up and low) were calculated for each FC value by using the standard error (SE) of  $diff_{GOI}$ :

$$FC_{UP} = 2^{-(diff_{GOI} - 1.96 \times SE)} \quad \mathbf{3}$$

$$FC_{LOW} = 2^{-(diff_{GOI} + 1.96 \times SE)} \quad \mathbf{4}$$

Only contrasts between genotypes expression data that remained significant after the Holm-Bonferroni correction and with a  $FC > 1$  are going to be discussed. FCs are graphically represented in Figures 1, 2, 3, 4, 5 and 6 where segments indicate 95% confidence interval.

**Figure 1.** Fold change (FC) for the contrast among alternative genotypes G/C<sub>-660</sub>-C/A<sub>-601</sub>-G/A<sub>-522</sub>-A/G<sub>-444</sub> of the *HSP90AA1* promoter within each treatment normalized by *HSP90AB1*. Segments indicate the 95% confidence interval (FC<sub>up</sub>-FC<sub>low</sub>). In abscissa the FC, in ordinate genotype contrasts. Asterisk over each bar indicates the significance level of the contrasts.



## RESULTS

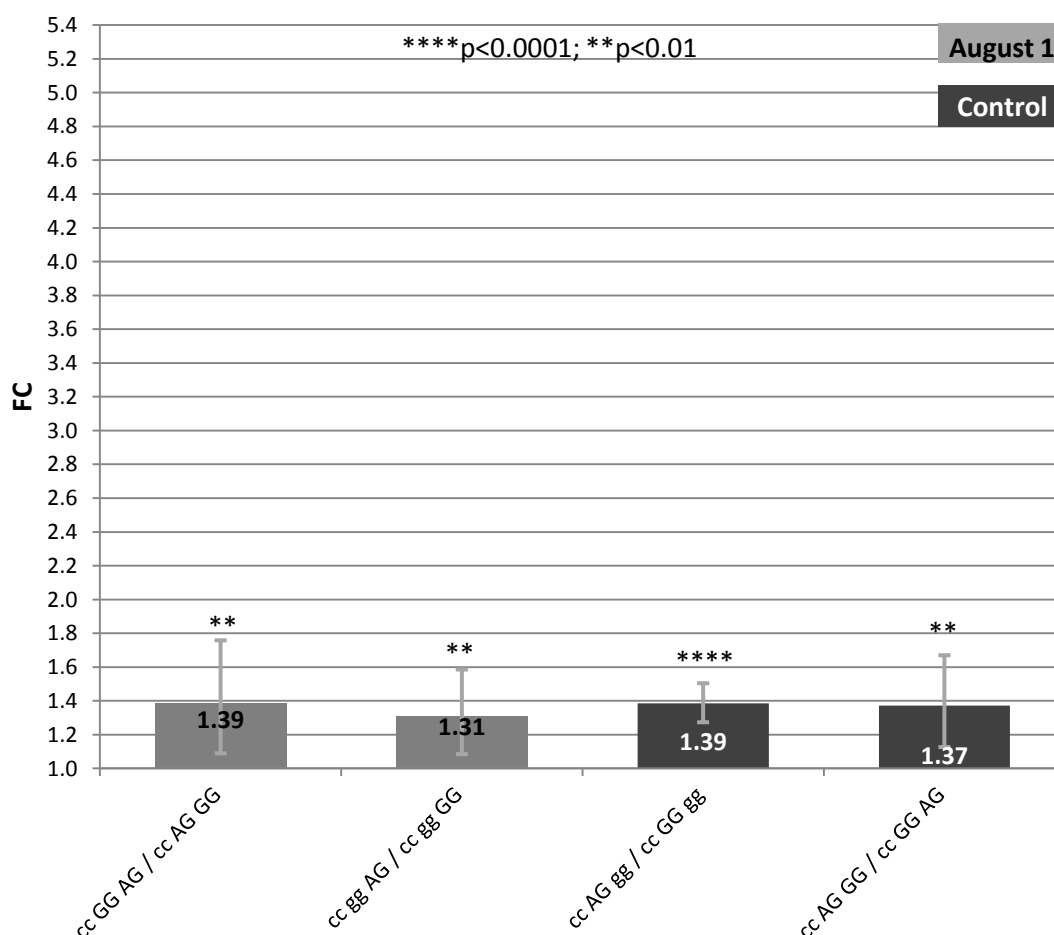
### Linkage disequilibrium estimation among 11 polymorphisms of the *HSP90AA1* promoter

Results from Hardy-Weinberg equilibrium exact test and expected and observed heterozygosities are shown in Table S2 including the total number of final polymorphisms found. No deviations from the Hardy-Weinberg equilibrium were observed for any of the SNPs genotyped on population composed by trios. The average expected and observed heterozygosities were 0.27 and 0.32, respectively. The polymorphisms with the lowest allele frequencies were g.522A>G, g.666\_667insC and g.444A>G (0.018, 0.036 and 0.053, respectively).

Table 4 shows the matrix of  $r^2$  values among the 11 polymorphisms detected in the *HSP90AA1* promoter. Two linked blocks were observed. g.703\_704insAA-g.660G>C-g.528A>G, with linkage from 54% to 98% linkage and g.601A>C-g.524G>T-g.468G>T from 61% to 90%. Moderate LD values (from 31% to 33%) were found

between pairs g.515\_516insG-g.528A>G, g.515\_516insG-g.660G>C and g.515\_516insG-g.703\_704insAA-g.666\_667insC showed a 40% of LD with g.444A>G-g.667\_668insC showed LD values from 19% to 26% with the SNPs g.528A>G, g.660G>C and the INDEL g.703\_704insAA. Among g.667\_668insC, g.666\_667insC, g.516\_517insG very low  $r^2$  values were found.

**Figure 2.** Fold change (FC) for the contrasts among alternative genotypes A/C<sub>-601</sub>-G/A<sub>522</sub>-A/G<sub>-444</sub> of the *HSP90AA1* promoter within each treatment normalized by *HSP90AB1*. Segments indicate the 95% confidence interval ( $FC_{up}$ - $FC_{low}$ ). In abscissa the FC, in ordinate genotype contrasts. Asterisk over each bar indicates the significance level of the contrasts.



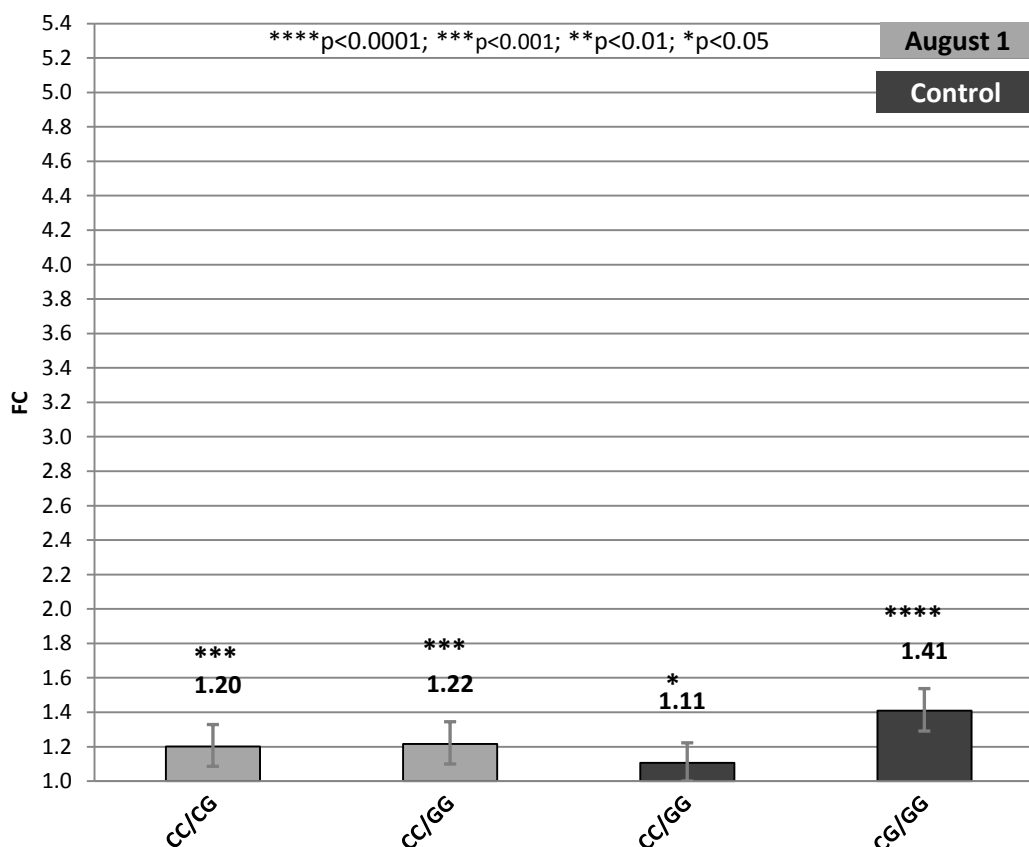
### Statistical analysis of RIN values

The model including the interaction DxGP as fixed effect and the animal random, showed the lowest values for the goodness of fit criteria (AIC and BIC).

Estimated animal and residual variances were 0.11 and 2.25, respectively. Type III fixed effects test showed a highly significant ( $p < 0.0001$ ) effect of DxGP on RIN values but no significant effect were observed for G, D and GP on the trait.

Thus, as RIN values only depend on the order in which samples were processed after their collection, it can be included as a systematic effect in the statistical model used to analyse expression data.

**Figure 3.** Fold change (FC) for the contrast among alternative genotypes G/C<sub>-660</sub> of the *HSP90AA1* promoter within each treatment normalized by *HSP90ABI*. Segments indicate the 95% confidence interval ( $FC_{up}-FC_{low}$ ). In abscissa the FC, in ordinate genotype contrast. Asterisk over each bar indicates the significance level of the contrasts.



### Best HKs

Table 5 shows MSE values obtained for each gene within treatments and across genes. *HSP90ABI* was in all cases the most stable gene, followed by *HSP90AA1*. Therefore, *HSP90ABI* was selected as the only HK to normalize the expression results



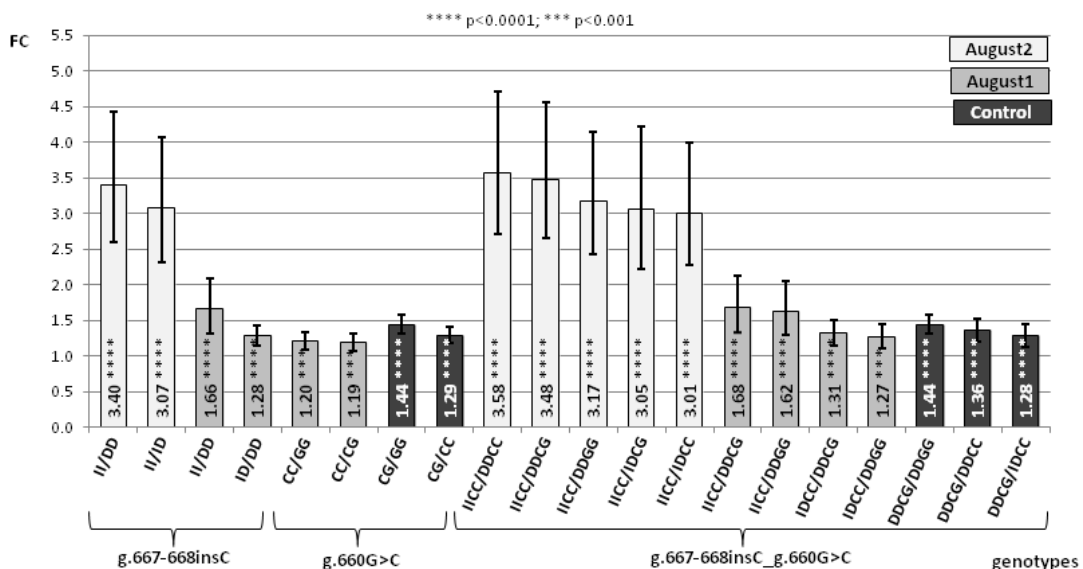
of *HSP90AA1*. The highest stability values for all genes corresponded to samples collected in August 2 and lowest stability values corresponded to the control samples.

### Environmental conditions and statistical analysis of gene expression

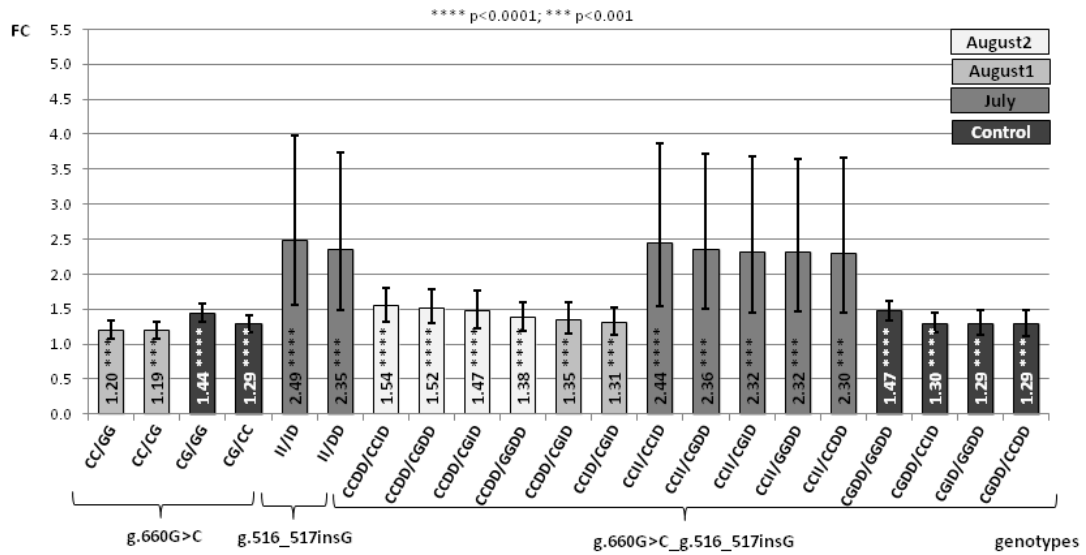
As it is shown in Table 2 the maximum temperatures for days of samples collection in July, 35.0°C, August 1, 34.4°C and August 2, 33.8°C, exceeded the sheep thermoneutral zone (Curtis 1983), but this is not the case for the average temperatures, 26.8°C, 24.7°C and 27.3°C, respectively. For the three time points, average and maximum THI values occurred in the zones of severe and extreme heat stress (Marai et al. 2001). The THI<sub>max</sub> was one unit higher in August 1 than in July or August 2. For the control time point, temperatures and THI values indicated no heat stress conditions.

Raw Cq values for all genes in each treatment are shown in Figure 7. Under control conditions, Cq values for the *HSP90AA1*, *HSP90AB1*, *MDHI* and *SDHA* genes were 26.2, 26.6, 29.4 and 30.3, respectively. Smaller Cq values were observed for all genes in samples collected under high temperatures (July and August 1 and 2). They were 25.7 for both chaperones and 28.8 and 29.7 for *MDHI* and *SDHA*, respectively. Variability in the expression rate of all genes was higher in samples collected under control conditions.

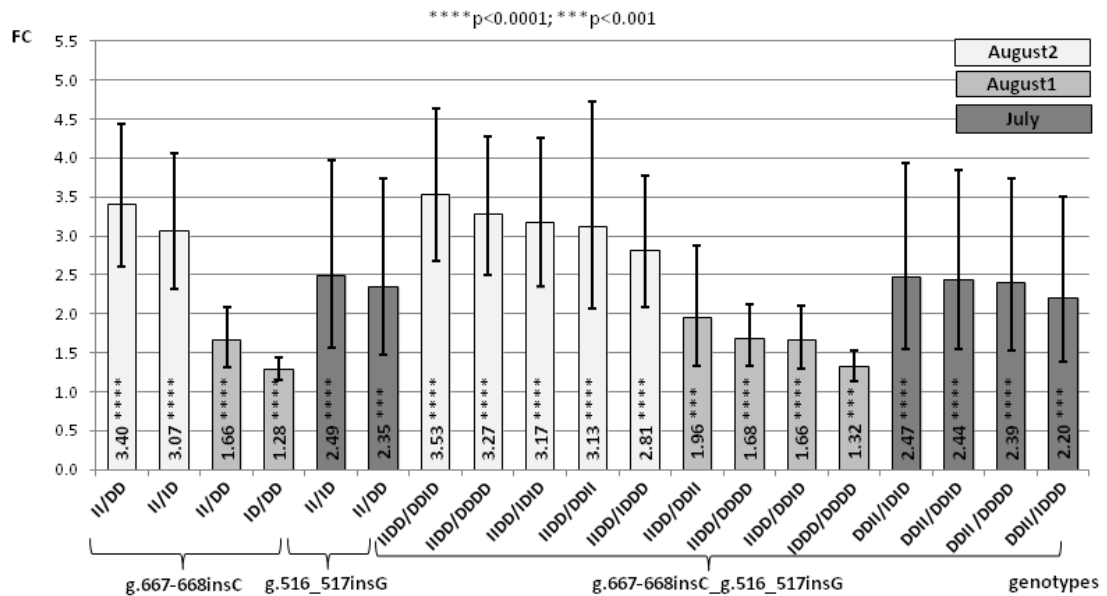
**Figure 4.** Fold change (FC) for the contrast among alternative genotypes of g.667\_668insC, g.660G>C and combined genotypes g.667\_668insC-g.660G>C of the *HSP90AA1* promoter within each treatment normalized by *HSP90AB1*. Segments indicate the 95% confidence interval (FC<sub>up</sub>-FC<sub>low</sub>). In abscissa the FC, in ordinate genotype contrasts (I = insertion; D = deletion).



**Figure 5.** Fold change (FC) for the contrast among alternative genotypes of g.660G>C, -516insG and combined genotypes g.660G>C\_-516insG of the *HSP90AA1* promoter within treatment normalized by *HSP90ABI*. Segments indicate the 95% confidence interval ( $FC_{up}-FC_{low}$ ). In abscissa the FC, in ordinate genotype contrasts (I = insertion; D = deletion).



**Figure 6.** Fold change (FC) for the contrast among alternative genotypes of g.667\_668insC, -516insG and combined genotypes g.667\_668insC\_-516insG of the *HSP90AA1* promoter within each treatment normalized by *HSP90ABI*. Segments indicate the 95% confidence interval ( $FC_{up}-FC_{low}$ ). In abscissa the FC, in ordinate genotype contrasts (I = insertion; D = deletion).



**Table 4.** Linkage disequilibrium values ( $r^2$ ) among the 11 polymorphisms detected in the *HSP90AA1* promoter.

	g.444A>G	g.468G>T	g.515_516insG	g.522A>G	g.524G>T	g.528A>G	g.601A>C	g.660G>C	g.666_667insC	g.667_668insC	g.703_704insAA
g.444A>G	1										
g.468G>T	0.0057	1									
g.515_516insG	0	0.0028	1								
g.522A>G	0.0046	0.0082	0.0188	1							
g.524G>T	0.0051	<b>0.9008</b>	0.0087	0.0074	1						
g.528A>G	0.0449	0.0269	<b>0.3331</b>	0.0106	0.0181	1					
g.601A>C	0.0046	<b>0.6189</b>	0.0057	0.0066	<b>0.6888</b>	0.0205	1				
g.660G>C	0	0.032	<b>0.3143</b>	0.0382	0.0225	<b>0.5842</b>	0.0246	1			
g.666_667insC	<b>0.3982</b>	0.0008	0.0041	0.0007	0.0008	0.0123	0.0007	0.0127	1		
g.667_668insC	0	0.0031	0.0006	0.0008	0.0018	<b>0.1998</b>	0.001	<b>0.2513</b>	0.0156	1	
g.703_704insAA	0	0.0352	<b>0.335</b>	0.0412	0.025	<b>0.6006</b>	0.0271	<b>0.9405</b>	0.0132	<b>0.2644</b>	1

**Table 5.** Minimum square error (MSE) within and across treatments for *HSP90AA1*, *HSP90AB1*, *MDHA* and *SDHA* genes.

Treatment	Gene	maximum mse	
		MTG within treatment	maximum mse MTG within gen
August 1	<i>HSP90AB1</i>	0.92	
August 1	<i>HSP90AA1</i>	1.58	
August 1	<i>SDHA</i>	1.77	
August 1	<i>MDH1</i>	6.60	
August 2	<i>HSP90AB1</i>	0.55	
August 2	<i>HSP90AA1</i>	0.81	
August 2	<i>SDHA</i>	1.72	
August 2	<i>MDH1</i>	1.77	
July	<i>HSP90AB1</i>	1.25	
July	<i>HSP90AA1</i>	1.55	
July	<i>SDHA</i>	2.41	
July	<i>MDH1</i>	2.71	
control	<i>HSP90AB1</i>	1.62	1.62
control	<i>HSP90AA1</i>	1.92	1.92
control	<i>SDHA</i>	2.63	2.63
control	<i>MDH1</i>	2.92	6.60

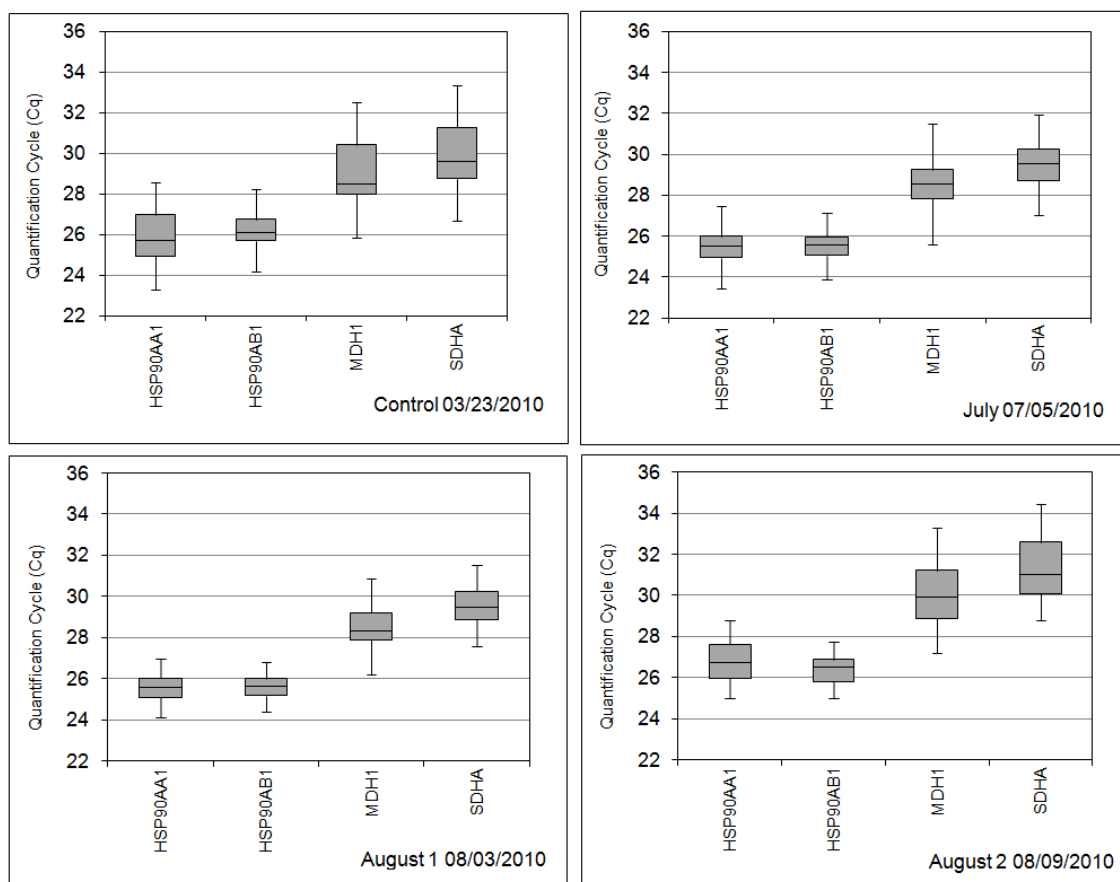
### Overall outcomes from fitting the mixed model

Test type III of fixed effects shows high significant F values ( $p < 0.0001$ ) for the MTG, P and b(RG) effects in all sets of genotypes. Estimates of the P effect levels (74)

averaged 0.73 and ranged between 0.01 to 1.41 Cq. Regression coefficients estimates relating covariation of Cq and RIN values for each gene, were -0.28, -0.47, -0.55 and -0.67 for *HSP90ABI*, *HSP90AAI*, *MDHI* and *SDHA*, respectively.

Estimates of animal, sample and residual variances were very similar for the 3 sets of genotypes studied. Animal variance was very small and ranged between 0.03 and 0.05. Sample variance (1.40) was 2.8 times higher than the animal effect one. *HSP90ABI* showed a residual variance (0.0086) 34 times lower than the one of *HSP90AAI* (0.27) and 53 times lower than the one of *SDHA* (0.46). *MDHI* had the highest residual variance (1.34).

**Figure 7. Distribution of Quantification Cycle (Cq) values for the target gene *HSP90AAI* and the reference genes *HSP90ABI*, *MDHI* and *SDHA*.** They were obtained by qPCR from samples collected at three different time points (Control, July, August 1 and August 2) along the year. Boxes show the range of Cq values within each gene and treatment; the centre line indicates the median; extended vertical bars show standard deviation of the mean.



In this chapter, expression results are organized in two different sections in concordance with the chronological procedure carried out.

### **I. Expression studies of 8 polymorphisms constituting highly linkage block**

Based on linkage disequilibrium results, 3 sets of genotypes were selected to carry out expression studies. Combined genotypes considered in separate analyses were: 1) g.660G>C-g.601A>C-g.522A>G-g.444A>G; 2) g.601A>C-g.522A>G-g.444A>G; 3) g.660G>C.

*1) Combined genotype g.660G>C-g.601A>C-g.522A>G-g.444A>G.* Figure 1 shows expression differences estimated among g.660G>C\_g.601A>C\_g.522A>G\_g.444A>G genotypes under the different climatic conditions considered.

For samples collected in August 1, CC<sub>-660</sub>CC<sub>-601</sub>GG<sub>-522</sub>AG<sub>-444</sub> showed the highest expression rate (differences in FC from 2 to 3) when comparing with the other 8 genotypes. Seven of these 8 genotypes were GG<sub>-444</sub>, highlighting the importance of this position for the expression rate of the gene under heat stress conditions independently of the other three SNPs. Two significant contrasts pointed out the effect of g.660G>C in terms of expression efficiency. CC<sub>-660</sub> showed differences in FC of 1.27 and 1.31 when comparing with CG<sub>-660</sub> and GG<sub>-660</sub>, respectively. It is important to note that contrasts showing differences in FC from 2.3 to 3.0 were those with wider confidence intervals which indicate higher estimated standard errors. This is due to the fact that in many of these contrasts, composed genotypes are in very low frequencies. In some cases only one animal exhibit the genotype compared. For comparisons where FC differences ranged between 1.2 and 1.3, confidence intervals were narrower due to the higher frequencies of the genotypes compared.

For samples collected in July and August 2, no comparisons between alternative genotypes were significant after the Holm-Bonferroni correction.

For the 5 significant contrasts of control samples the loose of the effect over the expression rate of the g.444A>G in the gene promoter and the change in the expression rank of the g.660G>C genotypes were the most important effects observed. In 4 contrasts the GC<sub>-660</sub>CC<sub>-601</sub>AG<sub>-522</sub>GG<sub>-444</sub> genotype showed differences in FC from 1.3 to 1.6 when compared with alternative genotypes of g.660G>C and g.522A>G. The superiority in terms of expression rate seems to depend in this case on the combination of g.660G>C and g.522A>G genotypes. Thus, the double

heterozygous  $GC_{-660}AG_{-522}$  showed higher expression levels than  $GG_{-660}GG_{-522}$  (1.58-1.60),  $CC_{-660}AG_{-522}$  (1.54) and  $GC_{-660}GG_{-522}$  (1.30). Finally, the  $g.601A>C$  did not seem to be involved in the regulation of the basal expression of the gene. In this case, differences in the confidence intervals were smaller since genotypes compared had higher frequencies.

2) Combined Genotype  $g.601A>C$ - $g.522A>G$ - $g.444A>G$ . To determine the impact of  $g.660G>C$  over the expression rate of the gene under the different environmental conditions we have tested expression differences among genotypes of three SNPs excluding  $g.660G>C$ .

Figure 2 shows results for contrasts among  $g.601A>C$ - $g.522A>G$ - $g.444A>G$  genotypes under the different climatic conditions considered.

Regarding the high decrease in the magnitude of the FC differences for the 4 significant comparisons, we can confirm the critical influence of  $g.660G>C$  over the gene expression rate. Surprisingly, higher differences in FC were detected for contrasts in control than in August, revealing a more important effect of these SNPs on the basal expression of the *HSP90AA1* gene.

In samples collected in August 1, the two significant contrasts with the highest expression rate implied the  $AG_{-444}$  genotype in all cases. Again,  $AG_{-444}$  showed higher expression rate (FC=1.3-1.4) over  $GG_{-444}$  under heat stress conditions.  $g.601A>C$  and  $g.522 A>G$  genotypes did not show any clear effect. No significant contrast was found among genotypes from samples collected in July.

Once more, under mild temperatures the  $g.444A>G$  lost its effect over the *HSP90AA1* expression which appeared under heat stress conditions. As in the case above mentioned,  $AG_{-522}$  showed a positive effect over the expression of the gene comparing with  $GG_{-522}$  (FC=1.37-1.39).

3) Genotype  $g.660G>C$ . Figure 3 shows results for contrasts among  $g.660G>C$  genotypes under the different climatic conditions considered.

Contrasts of this genotype showed lower differences in FC than that observed for the previous 2 sets of genotypes. Two comparisons were significant in samples collected in August 1. Differences in FC were 1.22 and 1.20 for the contrasts  $CC_{-660}$  vs.  $GG_{-660}$  and  $CC_{-660}$  vs.  $CG_{-660}$  respectively, showing the superiority of the  $CC_{-660}$

genotype over the other genotypes in terms of gene expression rate under heat stress conditions.

In July samples, no comparison had statistical significance. However, under control temperatures, 2 contrasts showed significant differences in FC, CC<sub>-660</sub> vs. GG<sub>-660</sub> with a FC equal to 1.11 and GC<sub>-660</sub> vs. GG<sub>-660</sub> with a FC of 1.41.

Thus, when considering only g.660G>C, results indicated no differences in *HSP90AA1* expression rate across treatments, but the existence of such differences across genotypes.

In order to understand better the results obtained, contrasts involving g.601A>C and g.522A>G, were carried out (Figure S2 and Figure S3). In the first analysis, alternative genotypes of g.601A>C-g.522A>G were compared. Only one significant contrast was found under control conditions. A FC value of 1.4 was observed for the contrast CC<sub>-601</sub>AG<sub>-522</sub> vs. CC<sub>-601</sub>GG<sub>-522</sub>, confirming the effect of g.522A>G over the basal expression of the gene under mild temperatures. g.601A>C, did not seem to play any clear role in the *HSP90AA1* expression rate.

In the second additional analysis, genotypes of g.660G>C-g.522A>G were compared to elucidate the relevance of g.522A>G under heat stress conditions. In this case, significant contrasts were found for control samples and those collected in August 1. The importance of g.660G>C in August 1 was again revealed. CC<sub>-660</sub> had higher expression rate than GC<sub>-660</sub> and GG<sub>-660</sub> (FC = 1.24 and 1.27, respectively). Under control conditions the effect of g.522A>G was clear (FC = 1.28 at least). Changes in the behavior of g.660G>C under control conditions were observed as well. In this case, GC<sub>-660</sub> was superior to CC<sub>-660</sub> and GG<sub>-660</sub> as it was showed in previous analyses.

## 2. New approach: INDELS not previously studied

In the previous part of this chapter, we aimed to confirm the results obtained previously (Marcos-Carcavilla et al. 2010b), that showed an association of the g.660G>C SNP of the ovine *HSP90AA1* gene promoter with the expression levels of this gene under different environmental temperatures, using a higher amount of samples and treatments. In this new approach, we aimed to test the effect of some INDELS in the expression of the gene under alternative climatic conditions. After obtaining reliable genotyping results from three more polymorphisms (g.667\_668insC,

g.666\_667insC and g.516\_517insG), an extension of the study was carried out following the same steps.

### **Gene expression analysis for isolate and combined genotypes of INDELS and g.660G>C**

New analysis were performed including the candidate polymorphism already detected in the previous section, g.660G>C. Since among the rams used in this expression assay only one was heterozygous for the g.666\_667insC, this mutation was not included in the statistical analyses to test expression and differences (See Table 1).

*1) Contrasts between alternative genotypes of the g.667\_668insC.* Figure 4 shows the results for contrasts comparing cytosine insertion genotypes. Only high statistically significant contrasts were showed. We could observe differences in expression between genotypes only for samples collected in August 1 and August 2, when maximum environmental temperatures exceeded 33°C (Table 1). The homozygote genotype for the insertion (II<sub>.668</sub>) showed much higher expression rates ( $p < 0.0001$ ) than the heterozygote ID<sub>.668</sub> (FC=3.07) and the homozygote DD<sub>.668</sub> (FC= 3.40) for samples collected in August 2 (average temperature =27.3°C, maximum temperature =33.8°C and minimum temperature=22.2°C). For samples collected in August 1 (average temperature=24.7°C, maximum temperature=34.4°C and minimum temperature=16.6°C) lower differences among genotypes than in the previous case were found. Thus, animals carrying the II<sub>.668</sub> genotype showed higher expression rate ( $p < 0.0001$ ) than those with DD<sub>.668</sub> (FC=1.66). In this case also the heterozygote (ID<sub>.668</sub>) had significant ( $p < 0.0001$ ) higher expression levels than the DD<sub>.668</sub> one (FC=1.28) but no differences were observed between II<sub>.668</sub> and ID<sub>.668</sub> genotypes.

*2) Contrasts between alternative genotypes of the g.516\_517insG.* Figure 5 shows the results for contrasts comparing the guanine insertion genotypes. Only high statistically significant contrasts were showed. We could observe differences of expression between genotypes only for samples collected in July (average temperature = 26.8 °C; maximum temperature = 35.0 °C; minimum temperature = 16.8 °C). The II<sub>.516</sub> genotype showed higher expression rate than ID<sub>.516</sub> (FC=2.49,  $p < 0.0001$ ) and DD<sub>.516</sub> (FC=2.35  $p < 0.001$ ) genotypes. However, as there were only two animals carrying the II<sub>.516</sub> genotype the standard errors of the estimates were so large.

*3) Contrasts of combined genotypes g.667\_668insC-g.660G>C.* Figure 4 shows high significant contrasts among the existent combined genotypes of both polymorphisms.



To facilitate the comparison with results obtained for the isolate g.660G>C in the first section of results, significant contrast for genotypes of this polymorphism was also included. In this case, for samples collected under the most extreme heat stress environmental conditions (August 2) gene expression of the combined genotype seem to be controlled by the genotype of the g.667\_668insC. Thus the II<sub>-668</sub>CC<sub>-660</sub> genotype showed higher expression rates than the DD<sub>-668</sub>CC<sub>-660</sub> (FC=3.58) and the ID<sub>-668</sub>CC<sub>-660</sub> (FC=3.01). Under climatic conditions existing in August 1, lower effect of the g.667\_668insC over gene expression differences than those observed in August 2 were obtained. In this case similar FC (1.6) was found in contrasts between II<sub>-668</sub>CC<sub>-660</sub>/DD<sub>-668</sub>CG<sub>-660</sub> and II<sub>-668</sub>CC<sub>-660</sub>/DD<sub>-668</sub>GG<sub>-660</sub>. For the contrast ID<sub>-668</sub>CC<sub>-660</sub>/DD<sub>-668</sub>CG<sub>-660</sub> and ID<sub>-668</sub>CC<sub>-660</sub>/DD<sub>-668</sub>GG<sub>-660</sub>, FC ranged from 1.27 to 1.31. Under mild environmental temperatures (Control) the effect of the g.667\_668insC genotypes over gene expression differences was lost, and were the genotypes of the SNP g.660G>C those revealing expression differences. For Control samples, animals carrying the CG<sub>-660</sub> genotype showed higher expression levels than those with the GG<sub>-660</sub> (FC=1.44) and CC<sub>-660</sub> (FC=1.28 to 1.36) independently of the g.667\_668insC genotype, as it has been previously observed in the first section of results.

4) Contrasts of combined genotypes g.660G>C-g.516\_517insG. Figure 5 shows high significant contrasts among the existent combined genotypes of both polymorphisms. To facilitate the comparison with results obtained for the isolate g.660G>C, significant contrast for genotypes of this polymorphism was also included. In significant contrasts from August 1 and 2, except in one case (CC<sub>-660</sub>DD<sub>-516</sub>/CC<sub>-660</sub>ID<sub>-516</sub>), the genotype of g.660G>C seems to be the responsible of differences in the expression rate observed. Thus animals carrying the CC<sub>-660</sub> genotype independently of the g.516\_517insG one, showed higher expression levels than those with the CG<sub>-660</sub> (FC from 1.31 to 1.52) and GG<sub>-660</sub> (FC=1.38). For the environmental conditions occurred when Control samples were collected, also genotypes of the g.660G>C were responsible of differences observed in the expression rate of the gene (CG<sub>-660</sub>>GG<sub>-660</sub> FC=1.3 to 1.4 and CG<sub>-660</sub>>CC<sub>-660</sub> FC=1.3) as in previous analyses.

5) Contrasts of combined genotypes g.667\_668insC-g.516\_517insG Figure 6 shows high significant contrasts among the existent combined genotypes of both polymorphisms. To facilitate the comparison with results obtained for the isolate g.667\_668insC and g.516\_517insG significant contrast for genotypes of these polymorphisms were also included. In those contrasts belonging to August 1 and August 2 treatments, the preponderance of the g.667\_668insC in the composed

genotypes *g.667\_668insC-g.516\_517insG* was clear. Thus, homozygous  $II_{.668}$  genotype showed higher expression levels than the heterozygous  $ID_{.668}$  ( $FC=2.81-3.17$ ) and the homozygous  $DD_{.668}$  ( $1.66-3.53$ ), independently of *g.516\_517insG* genotypes. Also heterozygous  $ID_{.668}DD_{.516}$  showed higher expression rate than the  $DD_{.668}DD_{.516}$  ( $FC=1.32$ ). Results from July are closer to those obtained when considering the *g.516\_517insG* alone. It is important to emphasize that in this case many significant contrasts had high standard errors because a scarce number of animals of a particular genotype in some comparisons. These are the cases in which the  $II_{.668}$  and the  $II_{.516}$  genotypes were compared, since only five animals have the  $II_{.668}$  genotype and only two the  $II_{.516}$  one.

## DISCUSSION

In this study, we aimed to confirm the results obtained previously (Marcos-Carcavilla et al. 2010b), that showed an association of the *g.660G>C* SNP located at the ovine *HSP90AA1* gene promoter region with the expression levels of this gene under different environmental temperatures using a suitable amount of samples. We also aimed to increase the information available of the biological process underlying this type of stress response. With these purposes we have studied new polymorphisms found at the promoter region, which would affect the expression rate of the gene not only in heat stress events as it was firstly thought, but also modulating its basal expression.

### RIN effect

Best conservation and minimum degradation processes are critical points when sampling commercial livestock animals for expression studies. The degree of RNA degradation in the samples affects gene expression measurements. We have established that RIN values depended neither on the source of biological sample (the animal) nor on the environmental conditions surrounding samples collection. The only factor having a significant effect on RIN values was the period of time occurring between blood extraction and blood processing in LeukoLock™™ platforms for each time point (DxGP). The higher was this period of time, the higher the RNA was degraded (lower values of RIN). Therefore, we proposed to include RIN values as a fixed effect or as a covariate in the statistical model used to analyze expression

differences. In fact, the same results were obtained using both approaches (data not shown).

RIN values affect Cq of samples depending on amplicon size (Fleige et al. 2006b). The higher is the amplified DNA fragment the higher is the probability to be broken down. The length of the amplified product was more correlated with RIN values than expected. Amplicon sizes were in the range of 70 to 250 bp (Table 3) for which Fleige and Pfaffl (Fleige et al. 2006a) indicates a more or less independence of qPCR products and RNA quality. Furthermore, DNA CG content did not seem to affect RNA stability as it has been previously described, where lower CG degree content was correlated with higher RIN values (Opitz et al. 2010). *SDHA* amplicon was the one with the highest CG content (65%) but was the most affected by RNA degradation. Only for *MDHI*, which has the lowest CG content and a high RIN effect, this relationship seems to be true. These results reveal that the effect of RNA integrity over both the GOI and the HK should be taken into account in expression analyses.

### **HK selection**

A crucial aspect revealed in this work, is the need to test the stability of the candidate HKs and the GOI simultaneously. *MDHI* and *SDHA* were previously selected (Serrano et al. 2011) among 16 candidates tested, as the most stable pair in similar conditions to those evaluated here. *HSP90AA1* was not included in that experiment. In the present work, we have verified that the GOI, *HSP90AA1*, is much more stable than the two previously selected HKs, *MDHI* and *SDHA*. Therefore, none of them can be used to normalize the GOI expression data. The constitutive counterpart of the *HSP90AA1* gene, *HSP90AB1*, showed the best stability values within and between treatments (Table 4), and it was chosen as the HK to normalize the expression data of the *HSP90AA1*. Differences in the stability of both chaperone genes might be due to the effect of the polymorphisms existing at the promoter of the *HSP90AA1* (Deuerling et al. 2003) and to the inducible behaviour of this last gene.

### **qRT-PCR experimental design and statistical methods for expression data analyses**

When a great number of samples and treatments are included in a qPCR study the experimental design is important since qPCR plates have a limited capacity (96 or 384 wells). In our design, plates contained a randomized set of animals, treatments, genotypes, RIN values and genes to avoid estimation biases. The repetition of one or

more samples in all plates connects the plate's system allowing to remove technical nuisance from this source of variability and to compare results from all plates. We have confirmed that the plate effect is an important source of variability since differences in Cq among plates can reach values up to 1.4.

Traditional statistical methods to analyze qPCR data was restricted to pair-wise comparisons of treatments in which expression data from GOIs are previously normalized with one or more HKs. This kind of approach does not include systematic nor random effects and their interactions that could affect expression results. In the linear mixed model used in this study, GOI and HKs data are simultaneously analyzed. This model includes different sources of biological and technical variation (i.e. plate, RIN, genotypes, genes, and interactions among them) as fixed or random effects. Fitting this model let us checking HK stability, normalization of GOI data with the most stable HK(s) and test the linear hypothesis of the existence of different expression levels of the *HSP90AA1* gene depending on the genotype of the mutations located at its promoter and on diverse environmental conditions.

### **Environmental conditions and gene expression**

Sheep are believed to be one of the most resistant species to climatic extremes, especially to high environmental temperatures. Environmental conditions in Ciudad Real often exceed sheep thermo neutral zone which is comprised between 5°C and 25°C (Curtis 1983). As expected, expression results differed between heat stress and control conditions. However, in the first section of results, differences in expression rate among genotypes were generally observed in samples collected in August but not in July. The scarce differences in climatic parameters existing between August and July collects did not explain the observed differences between these time points in terms of FC. The higher THI<sub>max</sub> values at collection time and during 5 days before collection in August than in July was thought to be the clue to such differences. Other environmental factors here unknown such as wind speed, number of hours over the comfort temperature, insulation, etc. included in Fanger's comfort equation (Fanger 1970/1982) would have also contributed in such differences.

Significant differences among g.601A>C\_g.522A>G\_g.444A>G and g.660G>C genotypes, found for August 1 and control samples but not for July or August 2 would be explained by the existence of a transition in the expression state of the gene between the basal transcription and the heat stress response. Changes in the expression ranking of g.660G>C observed between control and August 1 samples, and

also for other SNPs in a less evident way, would support this hypothesis. Also, since the heat stress response is not a permanent state, in terms of gene expression, even when heat shock conditions are still present, acclimatization processes cannot be discarded as possible source of differences found in samples collected in July and August in the first section of results (Basu et al. 2002).

### **Expression analysis and genotype comparisons**

Our initial hypothesis was that differences in the expression rate of the gene with different *g.660G>C* genotypes would be observed only under heat stress conditions (Marcos-Carcavilla et al. 2010b). Surprisingly, after including a higher amount of samples and a set of 10 additional polymorphisms located also at the *HSP90AA1* promoter, the existence of expression differences under heat stress and thermoneutral situations was confirmed as well. Thus, polymorphisms located at the *HSP90AA1* promoter seem to affect not only its expression rate as response to heat shock but also its basal transcription levels.

Differences in the expression rate found for the contrasts among alternative genotypes for the polymorphisms studied here suggest that the transcription of this gene may be multiply regulated by cross-talk of various transcription factors, as it was pointed out for this gene in human (Csermely et al. 1998). Although much of the heat-induced gene expression can be explained by HSF1 (heat shock factor 1), a perfect correlation between its binding and induction has not been found (Trinklein et al. 2004). Signal transduction cascades activated by p53, Jak and Ras pathways via HSF1 binding to the heat-shock response element (HSE) and integrating to modulate HSP transcription have been reported (Stephanou et al. 2011). Additional positive or negative factors may modulate the transcriptional induction of HSF1-bound genes. Moreover, eukaryotic gene expression is tightly regulated at many levels, and can vary its regulation complexity (Lemon et al. 2000). The core promoter (TATA box, initiator –INR- and downstream promoter element –DPE-), is the essential part. Next to the core, proximal enhancers as cis-control elements (i.e. CCAAT box, GC box, B recognition element (BRE) and STRE elements) might be acting. Upstream, distal enhancers (hormone responsive elements –HRE- and nuclear factor element –NFE-) and a huge diversity of regulators that recruit a cascade of more transcription factors contribute to gene transcription regulation (Thanos et al. 1995), (Blau et al. 1996). The polymorphisms studied in this work are located enough upstream to the beginning of the transcription initiation to consider them as binding sites of these co-regulators, or

distal enhancers that do not directly activate the transcription of the gene but modulate its expression.

The role of the SNP g.660G>C in the transcription of the gene under heat stress has been confirmed through analysis in which only this mutation is tested. In addition, results from the analyses of composed genotypes g.660G>C-g.601A>C-g.522A>G-g.444A>G and g.601A>C-g.522A>G-g.444A>G revealed a cooperative relationship among several SNPs in terms of transcription efficiency. Thus, alternative genotypes of g.660G>C-g.444A>G seem to affect the expression of the gene in response to heat stress and those of g.660G>C-g.522A>G the basal transcription of *HSP90AA1*, which may occur under climatic conditions comprising comfort temperatures. Under heat stress conditions, the superiority of CC<sub>-660</sub> over GC<sub>-660</sub> and GG<sub>-660</sub> and of GC<sub>-660</sub> over GG<sub>-660</sub> indicated an additive effect for this mutation. However, for the control samples, GC<sub>-660</sub> was superior to CC<sub>-660</sub> and GG<sub>-660</sub>. The effect of g.444A>G was less clear due to the low frequencies of the AG<sub>-444</sub> and AA<sub>-444</sub> genotypes; however, more clear conclusions can be extracted based on the results obtained from contrasts involving g.522A>G.

Several putative TFs have been predicted (Table S1) for the presence of C<sub>-660</sub> and A<sub>-444</sub>. Some TFs that could co-activate gene expression as distal enhancers only with CC<sub>-660</sub> were NFI/CTF (Nuclear factor I or CCAAT box-binding transcription factor) and VDR (Vitamin D receptor) together with RxR-alpha (Retinoid X receptor Alpha). The last two TFs form a heterodimer which attracts a complex of co-activators proteins. This complex links the heterodimer to the initiation complex formed at the TATA box, promoting the transcription machinery (Bikle 2010). Both TFs bind putatively at the sequence around g.660G>C, and VDR only when this position is C<sub>-660</sub>. For AG<sub>-444</sub> a heat shock element that could bind a heat shock factor was predicted for the presence of the A<sub>-444</sub> (Csermely et al. 1998).

The presence of g.703\_704insAA (Oner et al. 2012) completely linked, at least in this breed with g.660G>C, must also be considered in the expression regulation of the gene under heat stress conditions. Thus, CC<sub>-660</sub> animals are also homozygous for the AA<sub>-704</sub> insertion (II<sub>-704</sub>; AA/AA), CG<sub>-660</sub> animals are heterozygous for that INDEL (ID<sub>-704</sub>; AA/--) and GG<sub>-660</sub> animals are homozygous for the AA deletion (DD<sub>-704</sub>; --/--). g.703\_704insAA is located within a putative glucocorticoid receptor (GR) transacting factor binding site. The AA deletion (D<sub>-704</sub>) created a GR transcription site. It has been pointed out that glucocorticoids can suppress the heat shock response in stressed cells

by inhibiting the action of HSF1 (Wadekar et al. 2001). Therefore, this mutation would be the responsible of the expression differences observed for g.660G>C. Because of the high linkage disequilibrium between the g.660G>C and g.528G>A ( $r^2=0.95$ ) the possible effect of g.528G>A over the transcription rate of the gene under heat stress conditions is masked by the first, and therefore no conclusions could be extracted from this position.

Under control conditions, g.522A>G seems to have a predominant effect over the transcription rate of the gene, being AG<sub>-522</sub> (AA<sub>-522</sub> was not found in these samples) superior than GG<sub>-522</sub>. Due to the proximity of this mutation to g.528G>A and g.524G>T, TF binding sites were predicted for a sequence containing the three SNPs (Table S1). Several putative TF binding sites linked to A<sub>-522</sub> and T<sub>-524</sub> were found. Among them, the stress response element (STRE) (Csermely et al. 1998) and the JunD (functional component of the API -activator protein I- transcription factor complex) related with transcription coactivator activity, oxidative stress response (Mendelson et al. 1996) and spermatogenesis (Thepot et al. 2000) seemed to be closer to the *HSP90AA1* functions. c-Fos stimulates transcription of genes containing AP-I regulatory elements and was predicted for the sequence AtagTcA for the g.528G>A, g.524G>T and g.522A>G SNPs. In our samples animals with AG<sub>-522</sub> were always CC<sub>-601</sub> TT<sub>-524</sub> TT<sub>-468</sub> and in most cases (70%) CC<sub>-601</sub> AG<sub>-522</sub> TT<sub>-524</sub> TT<sub>-468</sub>. Two putative TF binding sites for C<sub>-601</sub> were predicted. HES-I (hairy and enhancer of split-I) (Yan et al. 2002), which can act as a repressor or activator, and USFI (Upstream stimulatory factor I) (Kumari et al. 2001), that has been found to be involved in the stress-activated signaling cascade (Galibert et al. 2001) and in the cessation of Sertoli cell proliferation and differentiation to spermatozooids (Wood et al. 2009). For the SNP g.468G>T one interesting homolog of the human ZNF395 binding SpI was found for maize (Lal et al. 2001) linked to the response to oxidative stress (Table S1).

In the second part of results, alternative genotypes of the g.667\_668insC have been here directly associated with the highest differences in the transcription rate of the gene under heat stress environmental conditions but no under thermoneutral ones. Thus animals carrying the II<sub>-668</sub> genotype showed higher transcription rates than those with ID<sub>-668</sub> (FC=3.07) and DD<sub>-668</sub> (FC=3.40) genotypes for samples collected in August 2. However, much lower differences among these genotypes (II<sub>-668</sub> DD<sub>-668</sub> FC= 1.66 and ID<sub>-668</sub> DD<sub>-668</sub> FC=1.28) for samples collected in August 1 and no differences for samples collected in July were observed. Despite maximum and average temperatures of these three collection dates are quite similar some differences should be the clue

for the results obtained in each of them. In particular, the minimum temperature in August 2 (22.2 °C) was quite higher than those of August 1 (16.6°C) and July (16.8°C). The minimum temperature affects one important variable related with the heat stress response, the daily thermal width (TW) which is the difference between the maximum and minimum temperatures occurring along the day. TW values were 18.2°C, 17.8°C and 11.6°C in July, August 1 and August 2 collection dates, respectively. Therefore, it seems that the heat stress response, in terms of over expression of genes involved in this metabolic pathway, more depends on the daily temperature pattern than on the magnitude of the maximum temperature reached, that is, how long is the period of time in which the environmental temperature exceeds a thermoneutral threshold. The magnitude and duration of the stress response is proportional to the dose or severity of the perturbation (Gasch et al. 2000). These facts would explain the differences in gene expression between genotypes found in samples collected in July, August 1 and August 2. Also, differences observed in the relative humidity among collections from July (39%), August 1 (49%) and August 2 (50%). can contribute to the differences in transcription observed, since the joint effect of heat and humidity become as a larger stressor source.

The single effects of these two mutations (g.667\_668insC and g.660G>C) increased when the combined genotype of both polymorphisms was considered (Figure 4) supporting the hypothesis of a combined action in modulating the gene transcription changes. The combined  $I_{-668}CC_{-660}$  genotype showed the highest expression levels in comparison with the remaining existent genotypes (FC from 1.27 to 3.58) under heat stress conditions. It is important to remark that in the sheep breed here studied (Manchega) these two polymorphisms showed a LD of 25%. The  $I_{-668}C_{-660}$  haplotype has a frequency of 0.14 while the  $D_{-668}C_{-660}$  and the  $D_{-668}G_{-660}$  have frequencies of 0.37 and 0.49, respectively. The haplotype  $I_{-668}G_{-660}$  does not exist in the ovine species (836 animals from 31 different sheep breeds from different locations of Europe, Africa and Asia genotyped, Chapter 3). Therefore we cannot completely distinguish the effect of each polymorphism by itself. However, due to its proximity in the DNA sequence, the most likely thing is a synergistic effect of both mutations.

A putative binding site for the Sp1 (specificity protein 1) transcription factor has been predicted for the sequence constituted by these two mutations (Supplemental Table I). Sp1 is a zinc finger transcription factor that binds to GC-rich motifs of many promoters and is involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin



remodeling. The highest binding affinity of SpI was found for the sequence containing the I<sub>-668</sub> and C<sub>-660</sub> alleles. This affinity decreases for the D<sub>-668</sub>C<sub>-660</sub> haplotype and disappear for the D<sub>-668</sub>G<sub>-660</sub> combination. The high binding affinity of SpI for the haplotype I<sub>-668</sub>C<sub>-660</sub> could be the explanation of the higher expression rate observed for the homozygous II<sub>-668</sub>CC<sub>-660</sub> genotype under heat stress conditions than for the remaining ones.

Based on the differences in gene expression observed in the last assay for the combined genotype g.667\_668insC-g.660G>C, and in the fact that both polymorphisms showed a moderate LD (25%), the effect of the g.660G>C SNP observed previously in the first assay must be probably due to the action of both polymorphisms together. That was the reason for the low expression contrasts firstly shown. In the second assay of this chapter, differences in transcription rate of the gene due to the combined genotypes of both polymorphisms were isolated. By this way we could conclude that the expression of the gene under thermoneutral temperatures is regulated by the g.660G>C SNP and that the high upregulation observed under heat stress events is mainly due to a cooperative effect of g.667\_668insC and g.660G>C polymorphisms.

# **Chapter 2**

## **Functional study and epigenetic marks**

**Ovine *HSP90AA1* gene promoter: functional study and epigenetic modifications**



### INTRODUCTION

Promoters are sequence elements that interact with a significant number of transcription factors and epigenetic modifications to regulate gene expression (Landolin et al. 2010). Moreover, they are responsible for the integration of different favorable mutations including those which result beneficial for environmental condition changes (Gagniuc et al. 2013).

Although the transcriptional enhancement of *HSP90 $\alpha$*  is mainly due to HSF1 (heat shock factor 1), many other heat stress related transcription factors also modulate gene expression in response to environmental stress and could act simultaneously by cross-talk (Pirkkala et al. 2001, LeBlanc et al. 2012, Guisbert et al. 2013). Polymorphic changes (INDELs, SNPs, etc) at gene promoters can cause alterations in its expression levels if they affect binding motifs of transcription activators or repressors. Mutations can alter the binding sites of transcription factors blocking totally or partially their binding to the DNA sequence but also creating or destroying DNA motifs of epigenetic changes, i.e. CpG sites. Alterations of the gene transcription levels can have consequences at the phenotypic level in characteristics related with the functional pathways in which genes are involved.

In Chapter 1 differences in the expression of the *HSP90AA1* ovine gene depending on some polymorphisms and environmental conditions have been assessed. However, due to the linkage disequilibrium (LD) existing among some of the candidate polymorphisms it is not possible to distinguish the causal mutation producing such changes in the transcription level of the gene. Moreover, some of the polymorphisms located at the gene promoter (five SNPs) are susceptible of allele-specific methylation which might be other alternative mechanism regulating gene transcription levels.

Marcos-Carcavilla and coworkers (Marcos-Carcavilla et al. 2008) described the *HSP90AA1* promoter as a rich CG region, which opens the possibility that this highly inducible gene could have an alternative regulatory region apart from the identified TATA-box. Although the *HSP90AA1* gene is ubiquitous expressed there are tissues as testes and brain where the highest expression rates of this gene were found (Csermely et al. 1998). Therefore, it is possible that the gene's regulatory pathway would not be the same in all tissues.

Therefore, the aim of this chapter was: 1) to develop a functional study of the *HSP90AA1* promoter to detect which polymorphism(s) is(are) responsible(s) of

temperature dependent changes in the transcription rate of the gene previously detected; 2) to determine the allele-specific methylation pattern of susceptible polymorphisms; 3) to study the promoter structure and epigenetic marks in different tissues.

## MATERIAL AND METHODS

### Animal material and nucleic acid isolation.

Samples from different tissues, ages and ovine breeds were used. Sample identification, tissue, breed, age, genotype and extraction kit method are described in Table I. Several biological samples were those used previously in Chapter I. In addition, extra samples were obtained for this specific study (Table I).

**Table I.** Identification of samples used in the study.

**MCAB:** Manchega Control Adult Blood, **MHSB:** Manchega Heat Stress Adult Blood and **MBT:** Manchega Blood Trios (Blood samples from the same animals used in Chapter I). **RYH:** Rasa Aragonesa Young Heart, **MAB:** Manchega Adult Brain, **MAL:** Manchega Adult Liver, **MYO:** Manchega Young Ovary, **RYT:** Rasa Aragonesa Young Testicle, **MYT:** Manchega Young Testicle, **MAT:** Manchega Adult Testicle and **MAS:** Manchega Adult Sperm (Tissue samples collected specifically for this experiment).

**Extraction kits:** Salting out (Miller et al. 1988), Genra Puregene DNA Purification Kit protocol (Genra, Minnesota, USA) and MasterPure DNA Purification Kit protocol (Epicentre, Wisconsin, USA).

ID	TISSUE	BREED	AGE	NUMBER	GENOTYPE (g.660G<C)	SAMPLING SITE	EXTRACTION KIT
<b>MCAB</b>	BLOOD-CONTROL	MANCHEGA	ADULT	120	CC,GG,CG	CERSYRA	SALTING OUT
<b>MHSB</b>	BLOOD-HEAT STRESS	MANCHEGA	ADULT	16	CC,GG,CG	CERSYRA	SALTING OUT
<b>MBT</b>	BLOOD-TRIOS	MANCHEGA	ADULT	6	CC,GG,CG	CERSYRA	SALTING OUT
<b>RYH</b>	HEART	RASA	90 DAYS	8	CC,CG	SLAUGHTERHOUSE	GENTRA
<b>MAB</b>	BRAIN	MANCHEGA	ADULT	1	GG	SLAUGHTERHOUSE	GENTRA
<b>MAL</b>	LIVER	MANCHEGA	ADULT	1	GG	SLAUGHTERHOUSE	GENTRA
<b>MYO</b>	OVARY	MANCHEGA	90 DAYS	3	CC,GG,CG	SLAUGHTERHOUSE	MASTERPURE
<b>RYT</b>	TESTICLE	RASA	90 DAYS	8	CC,CG	SLAUGHTERHOUSE	GENTRA
<b>MYT</b>	TESTICLE	MANCHEGA	90 DAYS	1	GG	SLAUGHTERHOUSE	MASTERPURE
<b>MAT</b>	TESTICLE	MANCHEGA	ADULT	1	GG	SLAUGHTERHOUSE	GENTRA
<b>MAS</b>	SPERM	MANCHEGA	ADULT	9	CC,GG,CG	CERSYRA	MASTERPURE

### PCR and genotyping

The polymerase chain reaction was performed and the resulting PCR fragments were sequenced as in Chapter 1. A promoter fragment of 495 pb was obtained, containing 11 SNPs (g.660G>C, g.601A>C, g.528A>G, g.524G>T, g.522A>G, g.468G>T, -444A>G, -304A>G, -296A>G, -295C>T, -252C>G) and 4 INDELS (g.704\_705insAA, g.667\_668insC, g.666\_667insC, g.516\_517insG). Primers used in the PCR and PCR conditions are previously described in Chapter 1.

Five SNPs had putative methylation sites, depending on their alternative genotypes: g.660G>C, g.601A>C, g.528A>G, g.522A>G, g.304A>G.

### Electrophoretic mobility shift assay (EMSA)

A double-stranded probe with the sequences of the SNPs was used to determine the differences in binding due to genotype and methylation. Oligo sequences are shown in Supplementary Table 1. The forward oligonucleotide was labeled in 5' with IRDye700 (Li-Cor Biosciences, Lincoln, NE, USA) during the synthesis. We compared cell extracts binding between genotypes and temperature culture cells. For the competition experiments, an excess of unlabeled unmethylated probe was added to the mix prior to the addition of the labeled oligonucleotide. Band intensities were quantified in an Odyssey Infrared Imaging System (Li-Cor Biosciences). The inverse of band intensity versus the excess of unlabeled oligonucleotide was represented, and the slope of the resulting straight line indicated the affinity of each probe for the proteins in the nuclear extract.

Nuclear extracts were obtained using the Nuclear Extract kit from Active Motif (California, USA) following manufacturer's instructions. Protein quantification was obtained by Bradford method (Bradford 1976).

***In vitro methylation of probes:*** EMSA oligonucleotide probes used to compare methylation patterns (Supplementary Table 1) were incubated with *SssI methylase* (New England Biolabs, Ipswich, MA, USA) by double methylation. 1 µg of DNA was incubated with 1 µl of *M.SssI CpG methyltransferase* (4 U/µl, NEB, Ipswich, MA), 1 µl of S-adenosylmethionine (SAM, 32 mM) and 2 µl of 10×NEBuffer 2 in a 20 µl reaction volume at 37°C overnight, followed by incubation with freshly added 1 µl of *M.SssI* (4 U/µl), 1 µl of SAM (32 mM), 0.5 µl of 10×NEBuffer 2 and 2.5 µl of water at 37°C for 4 hr, then at 65°C for 20 min. To determinate the efficiency of methylation we used

Sau3AI and MspJI restriction enzymes (New England Biolabs, Ipswich, MA, USA) following manufacturer instructions.

### **Promoter-reporter constructs**

We PCR amplified from -1449 to +61 (with the TSS defined as +1) of the ovine *HSP90AA1* proximal promoter from genomic DNA. KpnI and XhoI restriction sites were introduced into the 5' ends of the primers to enable directional ligation into the same sites in pGL3-Basic (Promega). The PCR products were previously cloned in the pGEM®-T basic vector (Promega) to remove polyA generated during PCR. The *HSP90AA1* promoter fragment cloned and pGL3-Basic vector were digested with restriction enzymes, gel purified and ligated together with T4 ligase. Sequences of all plasmids were verified by sequencing. Relevant primer sequences are presented in Supplementary Table 2.1

Site directed mutagenesis of the genotypes not carried out by any of the animals previously sequenced (animals from Chapter 1 and 3) were obtained by overlap extension PCR and confirmed by sequencing. Primer sequences used are shown in Supplementary Table 2.1

### **Cell culture**

As model system, we chose human HepG2 hepatoma cell line, where HSP90 $\alpha$  levels and regulatory mechanisms are well characterized. They were used for *in vitro* experiments and EMSA extracts. They were maintained in culture with Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (FBS) and antibiotics.

Cells were plated at 5000 cells/cm<sup>2</sup> until 80% confluence. Then media was removed and they were washed with PBS 1x to be used.

Cell culture plates were treated with two different temperature treatments. Basal temperature treated cells were incubated at 37°C and 10% CO<sub>2</sub>.

Heat stress treated cells were incubated 2 hours at 42°C and 1 hour at 37°C before nuclear extract procedure was made always providing 10% CO<sub>2</sub>.

### **Transient transfections and luciferase reporter assay**

HepG2 cells were transiently transfected in 6-well plates using jetPEI transfection reagent (PolyPlus transfection SA, Illkirch, France) according to the

manufacturer instructions. For transfections, 2 µg/well of each reporter vector were used. A Renilla gene (0.1 µg) served as an internal control for transfection efficiency. After 48 hours, cells were lysed with Passive lysis buffer (Applied Biosystems) and luciferase activity was measured with the Dual-Glo luciferase assay system (Promega) following the manufacturer's instructions. Three independent experiments were carried out for each construction and experimental condition.

For cell culture plates that were heat shock induced, were incubated 2 hours at 42°C and 1 hour at 37°C before cell lysis.

### **Bioinformatics analysis of CpG islands and identification of Short Interspersed Elements**

The nucleotide sequence surrounding the transcription start site (TSS) of the *HSP90AA1* gene was explored by 3 putative CpG island softwares: Methyl Primer Express (Applied Biosystems, Bedford, MA, USA); EMBL-EBI (<http://www.ebi.ac.uk/tools/emboss>); USCnorris (<http://www.uscnorris.com/cpgislands2/cpg.aspx>)

To differentiate between transposable elements and CpG islands we used the software: RepeatMasker 4.0.5 (2014) version of Repbase 19.03 (for all species different of human).

### **DNA methylation analysis by sequencing**

The methylation status was determined using sodium bisulfite treatment. Bisulphite treatment was performed with ≤2µg of whole blood genomic DNA from 120 different animals in control temperature, 16 of them also under heat stress conditions, 6 from trios families, 23 samples of genomic DNA from different tissues and 9 samples of genomic DNA from sperm using Epiect plus Bisulphite Conversion (Qiagen, Valencia, CA USA) and EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA USA) following the guidelines of the manufacturer (see Table 1).

Genomic DNA and DNA after bisulphite treatment concentrations were determined using a NanoDrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies, Inc., DE, USA).

Due to the bisulphite treatment, DNA strands are no longer complementary and therefore the top and bottom strands from the fragment of interest were amplified and analyzed separately (Hajkova et al. 2002). Primers were designed



manually with the assistance of “Methyl Primer Express” software (Applied Biosystems, CA, USA) to replace all Cs for Ts except at CpG sites and NetPrimer software (Biosoft International, CA, USA) to avoid possible hairpin structures and primer dimmers and cross-dimers. Furthermore, primers were designed avoiding CpG in their sequence except two of them. In some cases, primer design was considered possible non-CpG methylation (Warnecke et al. 2002). Primers are listed in (Supplementary Table 2.2) together with amplicon sizes, T<sub>m</sub> and PCR kits used in each amplification fragment.

The polymerase chain reactions were carefully optimized from 150 -200 ng of bisulphite-treated DNA (Supp. Table 2.2). The resulting PCR fragments were purified with ExoSAP-IT (USB Corporation) and High Pure PCR Product Purification kit (Roche Diagnostics, Indianapolis, IN, USA) and sequenced with specific primers (Supp. Table 2.2).

Bisulphite sequencing has associated technical difficulties and potential artifacts. This may involve the formation of stable secondary structure around the methylated CpG site, creating a localized region of dsDNA and preventing access by bisulphate (Warnecke et al. 2002). Accordingly, we resolved it by first, re-sequencing and second, RFLP analysis with restriction enzymes: TaqI, BtsIMul and BstBI.

### **Statistical analysis**

Differences in EMSA band intensities were analyzed with Image J software using the Gel Analysis method. Average intensity bands were scored as the inversed of the most intensity band, which was used as reference.

Results from Luciferase assays were analyzed using the GLM procedure of the SAS statistical package (SAS/STAT® statistical package) fitting a model in which the dependent variable was the Relative Luminescence Units (LRU) obtained for each gene and genotype in each transfection. Transfection, replicate nested to genotype and genotype, were included in the model as fixed effects. Least Square Means, 95% confidence intervals and t test for means comparisons were calculated.

## **RESULTS**

### Functional study

#### Differences in transcription factor binding affinities

We investigated the formation of protein complex binding on the polymorphic sequence of the promoter, using a purified protein assay system to elucidate the differences in gene expression ratios. We designed labeled oligonucleotides for the alternative genotypes of a total of ten polymorphisms. Three of them, completely linked in the Manchega breed (SNPs g.601A>C, g.524G>T and g.468G>T), did not show any expression differences in the previous chapter. We could confirm this fact analyzing each SNP independently by electrophoretic mobility shift assay (EMSA) under control and heat stress treated cell extracts. There were no differences in protein complex binding among their genotypes, or they were unspecific.

We used the same assay procedure to evaluate g.444A>G and g.522A>G which showed in Chapter 1 subtle differences in gene expression, the first one under heat stress and the second one under thermoneutral conditions. EMSA assays in both cases did not show any specific protein affinity to any of the alleles from those polymorphisms.

Other LD block of polymorphisms was that constituted by g.703\_704insAA, g.660G>C and g.528A>G. In the case of the SNP g.660G>C, which has been largely suspicious as the causal mutation of the differences in gene expression observed under heat shock and thermoneutral conditions, EMSA suggested a difference in the ability of g.660G>C alleles to bind nuclear proteins under thermoneutral conditions (Figure 1). Competition experiments carried out with nonspecific oligonucleotides revealed the presence of a specific binding. The binding affinity to G<sub>-660</sub> was 42 ± 3.7% of the intensity of the C<sub>-660</sub> allele (100%) and disappeared very easily when competing with increasing amounts of an unlabeled C<sub>-660</sub> oligonucleotide. Similar binding patterns were obtained in EMSAs with heat shock treated cell extracts.

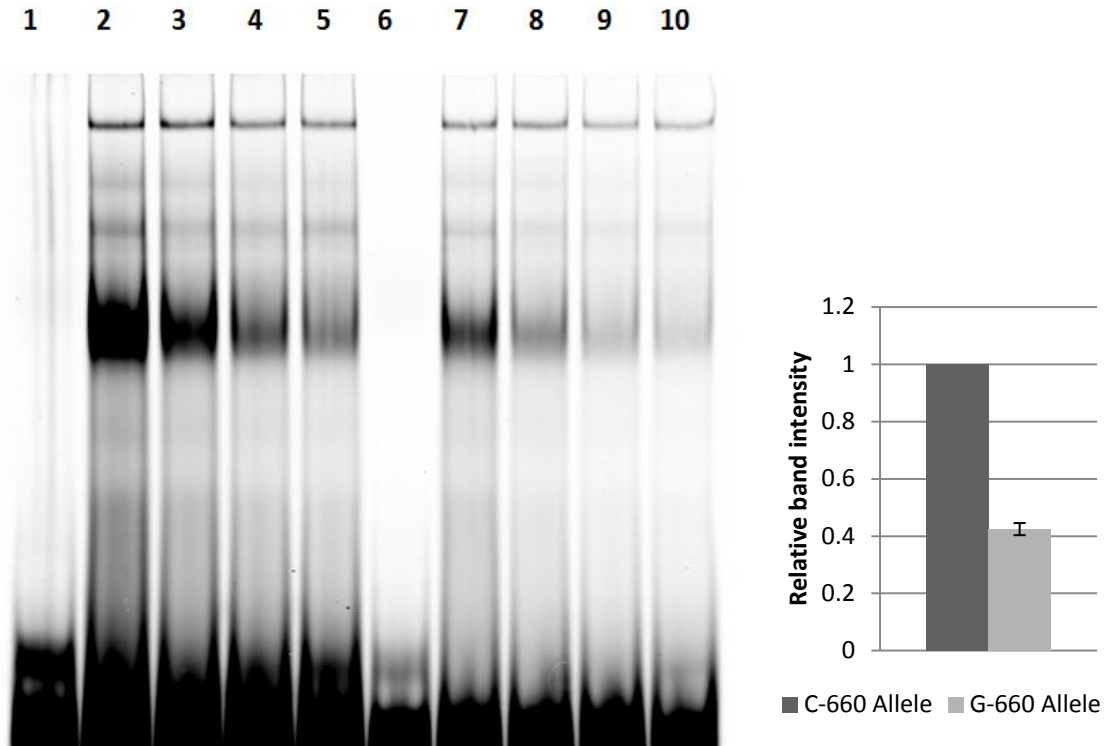
We performed the same EMSA experiment for g.703\_704insAA, and it also produced differences in band intensities (Figure 2) using cell extracts under thermoneutral conditions. In this case there is also a difference in band intensities, as the band with the double adenine insertion was 79 ± 12% of the intensity compared to the deletion one (100%). Then, as a glucocorticoid receptor has been predicted to bind the sequence with the deletion allele, we made a competition with 50 excess-fold

of cold oligo with the specific binding sequence of the glucocorticoid receptor and the band almost disappeared with a final intensity of  $33\pm 6\%$ .

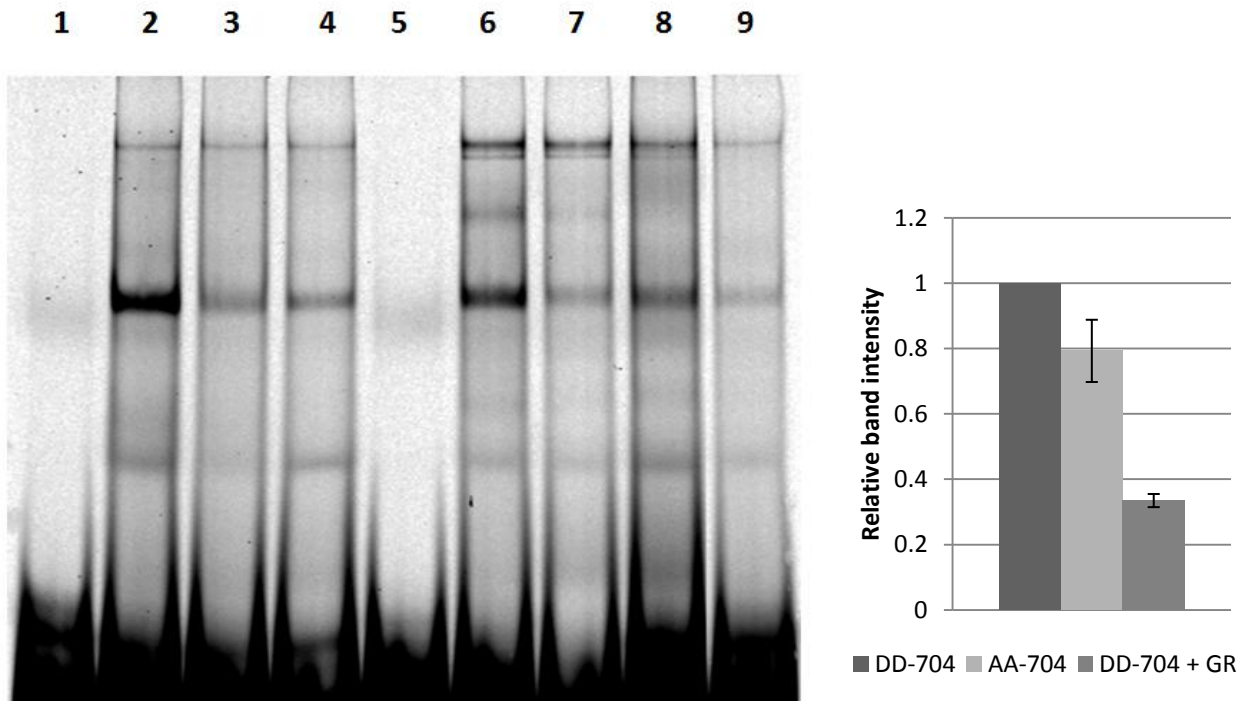
EMSA assay for the g.528A>G did not show any specific band affinity among its alternative alleles.

In addition, we also studied the INDELS g.666\_667insC and g.667\_668insC in combination with g.660G>C as they are only 6/7 pb away and there was oligo's length limitation. We compared the double deletion with one and two cytosine insertions. When cell extracts were subjected to thermoneutral temperatures the results were surprising as the double deletion and double insertion got the same band pattern independently of the SNP g.660G>C. We could observe that the sequence with an insertion, had a specific band neither present in the deletion nor in the double insertion lane (Figure 3). It seems that those proteins compete for the same sequence and their affinity could be due to the number of cytosines present in the binding sequence. EMSA from cell extracts under heat stress showed similar band intensities. In this case, compared with the control band, lower affinity was observed, probably due to the unfolding of transcription factors caused by high temperatures. In addition, in the I<sub>-668</sub>D<sub>-667</sub> lane, all complexes seem to disappear except the binding of a unique protein or complex (Supplemental Figure 1).

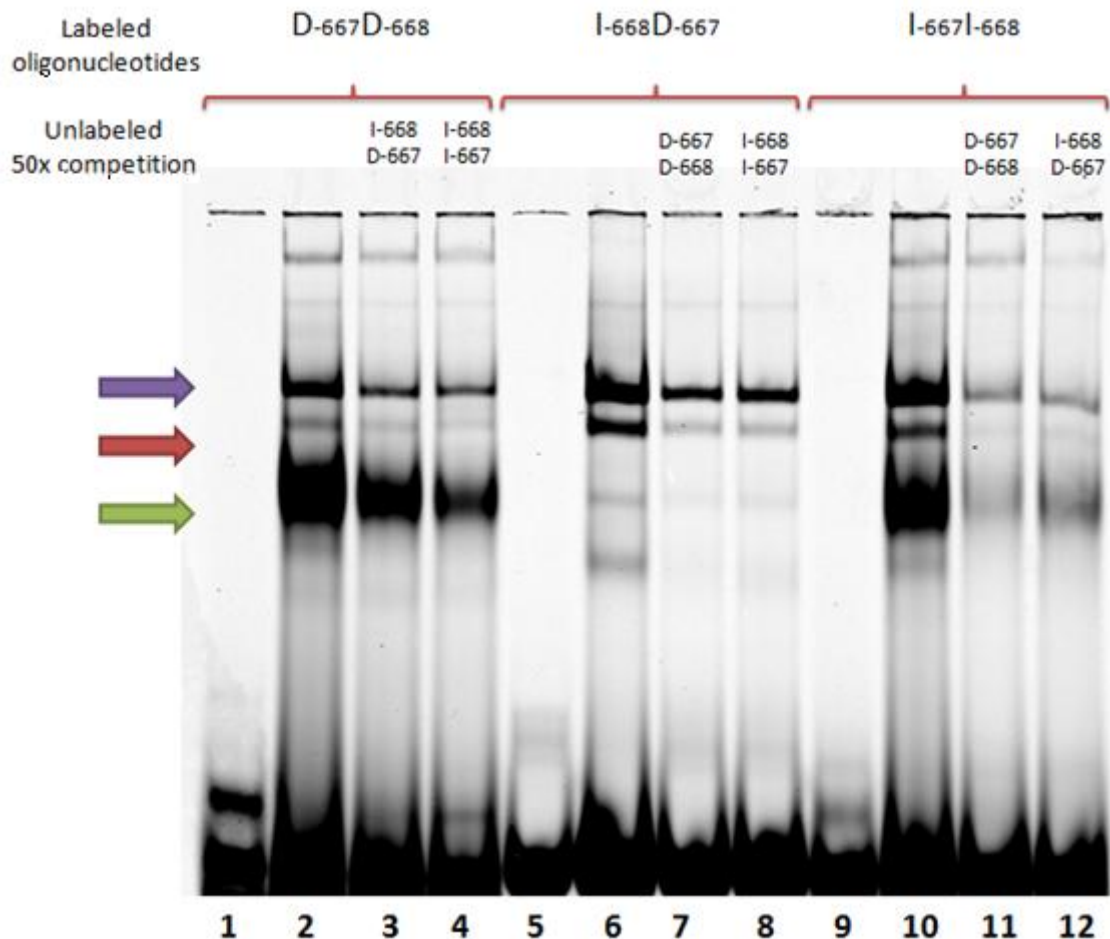
**Figure 1.** Electrophoretic Motility Shift Assay (EMSA) using nuclear extracts from HepG2 cells under thermoneutral conditions. Nuclear extracts from cultured cells were incubated with C-660-labeled oligonucleotide probe alone (*lane 2*), in the presence of increasing excess unlabeled C-660 probe (*lanes 3-10x, 4-25x, 5-50x*), G-660-labeled (*lane 7*), G-660-labeled with increasing excess unlabeled C-660 probe (*lanes 8-10x, 9-25x, 10-50x*). In *lanes 1* and *6*, nuclear extracts were not added.



**Figure 2.** Electrophoretic Motility Shift Assay (EMSA) using nuclear extracts from HepG2 cells under thermoneutral conditions. Nuclear extracts from cultured cells were incubated with the DD<sub>-704</sub>-labeled oligonucleotide probe alone (*lane 2*), in the presence of excess unlabeled DD<sub>-704</sub> probe (*lane 3*), in the presence of excess unlabeled AA<sub>-704</sub> probe (*lane 4*), the AA<sub>-704</sub>-labeled oligonucleotide probe alone (*lane 6*), in the presence of excess unlabeled DD<sub>-704</sub> probe (*lane 7*), in the presence of excess unlabeled AA<sub>-704</sub> probe (*lane 8*), DD<sub>-704</sub> in the presence of excess unlabeled a canonical GR sequence binding oligonucleotide (*lane 9*). In *lanes 1* and *5*, nuclear extracts were not added.



**Figure 3.** Electrophoretic Motility Shift Assay (EMSA) using nuclear extracts from HepG2 cells under thermoneutral conditions. Nuclear extracts from cultured cells were incubated with the D<sub>-668</sub>D<sub>-667</sub>-labeled oligonucleotide probe alone (lane 2), in the presence of excess unlabeled I<sub>-668</sub>D<sub>-667</sub> probe (lane 3), in the presence of excess unlabeled I<sub>-668</sub>I<sub>-667</sub> probe (lane 4), the I<sub>-668</sub>D<sub>-667</sub>-labeled oligonucleotide probe alone (lane 6), in the presence of excess unlabeled D<sub>-668</sub>D<sub>-667</sub> probe (lane 7), in the presence of excess unlabeled I<sub>-668</sub>I<sub>-667</sub> probe (lane 8), the I<sub>-668</sub>I<sub>-667</sub>-labeled oligonucleotide probe alone (lane 6), in the presence of excess unlabeled D<sub>-668</sub>D<sub>-667</sub> probe (lane 7), in the presence of excess unlabeled I<sub>-668</sub>D<sub>-667</sub> probe (lane 8). In lanes 1, 5 and 9, nuclear extracts were not added. The green arrow indicates the binding of a complex of proteins with more efficiency to D<sub>-668</sub>D<sub>-667</sub> and I<sub>-668</sub>I<sub>-667</sub> whereas the protein indicated with the red arrow has more efficiency to I<sub>-668</sub>D<sub>-667</sub>. The purple arrow indicates a protein that binds in all three possible genotypes, even though it seems to have more preference to I<sub>-668</sub>D<sub>-667</sub> as during competition it does not disappear.



### ***In vitro* expression assay**

Luciferase assays were performed only with those polymorphisms whose alternative genotypes showed differences in EMSA transcription binding affinities (Figure 5).

Results based on luciferase assays for the transversion g.660G>C showed that the promoter containing the G<sub>-660</sub> allele had a 17.6% decrease in its activity ( $P=0.0048$ ) comparing with that containing the C<sub>-660</sub> one under thermoneutral temperatures.

In the case of g.703\_704insAA, luciferase activity did not show significant *in vitro* expression differences between its alleles.

Meanwhile, luciferase assays involving C<sub>-660</sub>, g.667\_668insC and g.666\_667insC alternative alleles, showed significant differences after heat stress treatments. Insertion of one or both cytosines produces similar increase of promoter activation under heat stress conditions, 28.3% (0.0114) and 30.3% ( $p=0.0078$ ) respectively.

Cytosine insertions did not show differences in promoter activity under thermoneutral conditions, which agree with the expression results showed in the previous chapter.

### ***HSP90AA1* promoter structure and associated CpG island**

#### *HSP90AA1* core promoter elements and transcription factors binding sites

Next, we analyzed the ovine *HSP90AA1* gene promoter structure. Based on *Ovis aries HSP90AA1* sequence (DQ983231.1) a CpG island (CGI) was predicted associated with the 5' promoter region according to Gardiner-Garden and Frommer criteria (Gardiner-Garden et al. 1987) (ie GCm>50% and y value> 0.6, the intervals are of 200bp; Figure 6) .

The CGI is enriched with putative consensus motifs of nine SpI (M6: GGGCGGR) binding sites, two ELK-I (M3:SCGGAAGY) binding sites (one of them coincides with the Heat Shock Element HSE sequence) and one M22 (TGCGCANK) (Fig.7). These transcription factor binding sites are around the core promoter elements that would be bound by related transcription factors.

As a result we could confirm the presence of a CGI along with the previously described TATA-box (Marcos-Carcavilla et al. 2008). Methylation analysis by bisulphite sequencing allowed us to define the limits of the CpG island: -745 to +1445 respect

putative TSS, therefore, the CpG island has a length of 2199bp and includes BRE, TATA-box, Inr, TSS ( $A_{+1}$ ), the first three exons, the beginning of the fourth exon and three introns (Fig.8).

Those promoters with CGI and TATA-box structure are so-called *hybrid* promoters (Carninci et al. 2006, Deaton et al. 2011). The core promoter of *HSP90AA1* has a regulatory upstream promoter element (HSE), a consensus TFIIIB recognition element (BRE) motif which is followed by a canonical TATA-box (TATATAAG). They are located precisely at -89, -39 and -30 relative to the  $A_{+1}$ (TSS), respectively. Downstream, the core promoter contains the consensus Inr motif, which includes the TSS ( $A_{+1}$ ) and positioned relative to TATA-box in a “synergistic configuration” (O’Shea-Greenfield et al. 1992). Furthermore, a putative downstream core promoter element (DPE) consensus sequence at position +48 was found. This finding suggests that *HSP90AA1* gene promoter is DPE-less because of the big distance from the Inr.

Allele-specific DNA methylation pattern in blood

Allelic alternation of 5 SNPs of the promoter sequence could lead to methyl CpGs (660G>C, g.601A>C, g.528A>G, g.522A>G, g.304A>G). However, only one of them (g.660G>C) in its  $G_{-660}$  allele form, creates a cis-regulated allelic-specific methylation (ASM),  ${}_{me}G_{-660}$ , and it is flanked downstream by other non-polymorphic methyl CpG site at position g.632 relative to the *HSP90AA1* TSS( $A_{+1}$ ) (Fig.8). Bisulphite sequencing results showed that for the SNP g.660G>C the genotype  $CC_{-660}$  was unmethylated,  $CG_{g.660}$  hemi-methylated and  $GG_{-660}$  methylated.

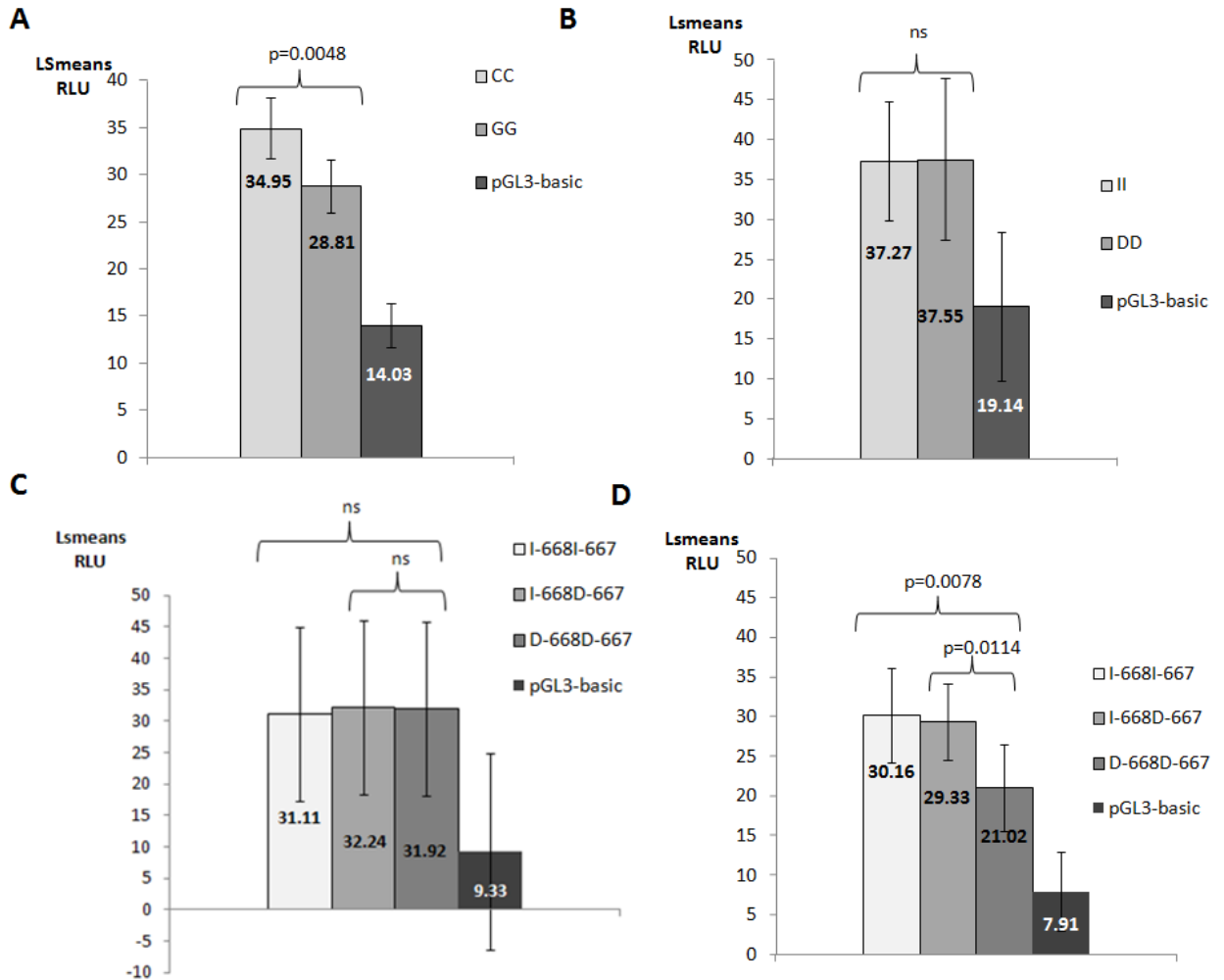
To check whether there were differences in methylation patterns in blood DNA control versus heat stress and if methylation segregates between sexes 16 animals and two family trios were analyzed, respectively. The trios genotypes for the g.660G>C SNP were: 1) ♂CC x ♀  ${}_{me}G_{me}G = C_{me}G$ ; 2) ♂ $C_{me}G$  x ♀CC =  $C_{me}G$  (Extracted from trios used in Chapter 1).

ASM in DNA from blood was independent of temperature existing when collection and independent of parental origin.

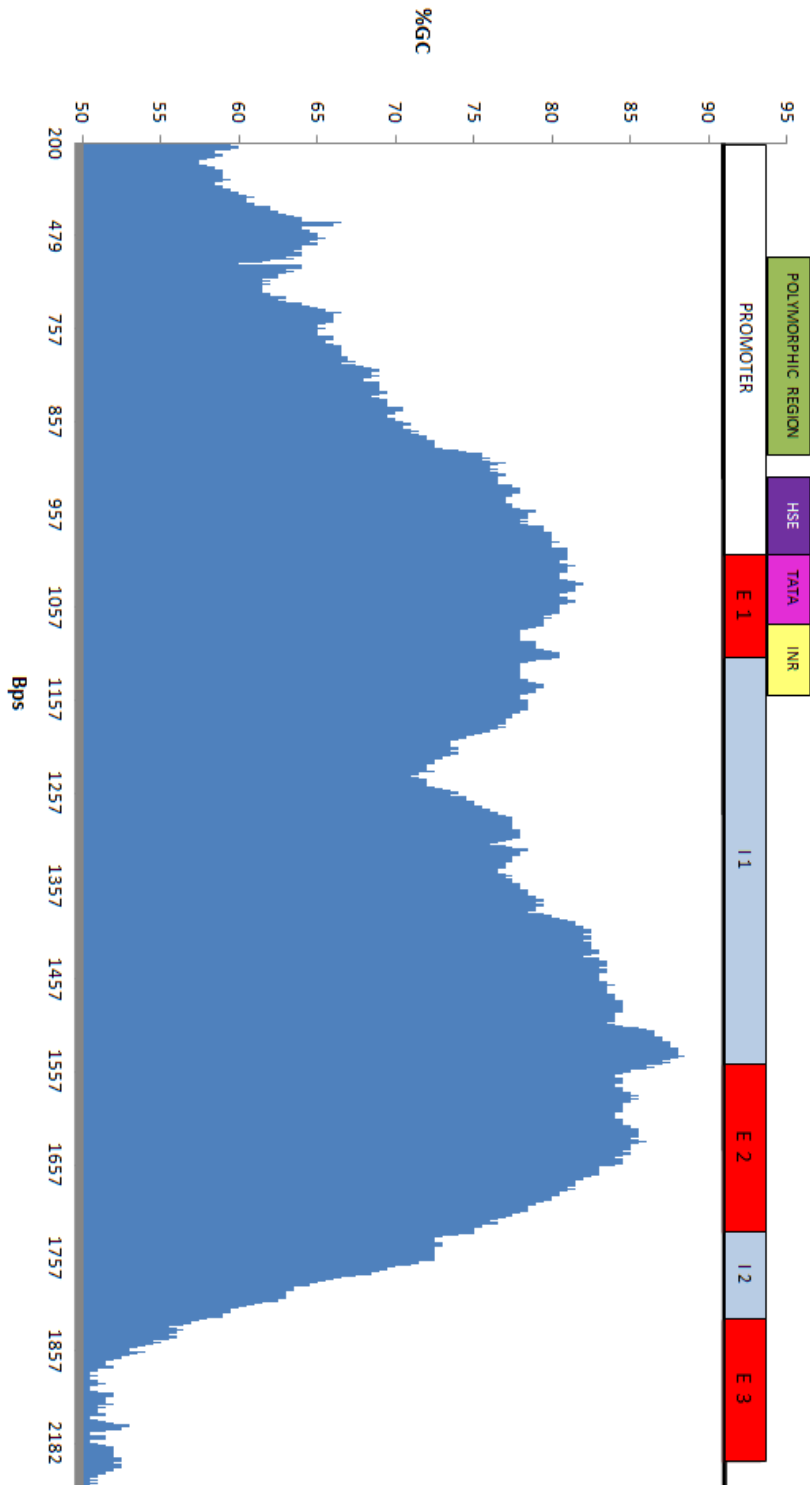
We found 5 hemimethylated CpG sites (5’ CpG island boundary) immediately upstream g.660G>C (methylated according to their alternative genotypes) and 2 hemimethylated site immediately 5’ to the first methylated CpG site (3’ CpG island boundary) (Fig. 9).



**Figure 5.** Luciferase Assays for several polymorphisms at the *HSP90AA1* promoter. Each alternative allele was transiently expressed in HepG2 cells for luciferase assays. Firefly luciferase activity was normalized to Renilla luciferase activity. Data are represented compared to pGL3-Basic and the means SD are for at least 3 replicates. **A.** g.660G>C alternative alleles under thermoneutral conditions. **B.** g.703\_704insAA alternative alleles under thermoneutral conditions. Ns: non-significant **C.** g.667\_668insC and g.666\_667insC alternative haplotypes under thermoneutral conditions. **D.** g.667\_668insC and g.666\_667insC alternative haplotypes under heat stress.



**Figure 6.** Graphic representation of ovine *HSP90AA1* gene CpG island. X axis: bases pairs. Y axis: % GC. (From: The Sequence Manipulation Suite: CpG Islands [http://www.bioinformatics.org/sms/cpg\\_island.html](http://www.bioinformatics.org/sms/cpg_island.html)). Putative CpG island of *HSP90AA1* gene spans core promoter elements, exons 1, exon 2, intron 1 and intron 2 (see top of the figure).



**Figure 7.** Description of the *HSP90AA1* promoter CpG island motifs. HSE (purple), BRE (blue), TATA-box (pink), TSS(A<sub>+1</sub>) (grey) at Inr (yellow), M6 consensus motifs (Sp1) binding sites (i.e. GC-boxes) (grey), M3 consensus motifs (ELK-1) (green), M22 consensus motif (red) and DPE motif (olive green) are highlighted.

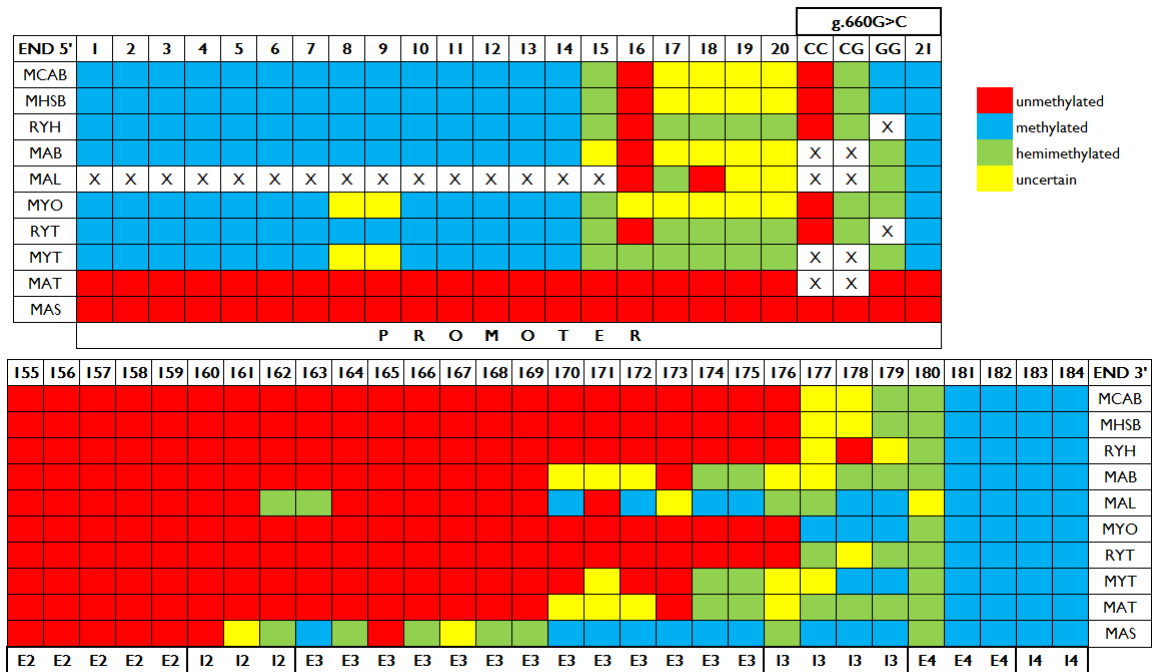
```
GGACTGGACCCCTCACCCCGGCTTTGATCCCTAAGCGCGGCTCAAGGCTCGAAACCTAGCCC
CAGGCCATAGGCGGGGGTCCCGATGCTGGGGGATCCGGAGCGCACACTATCTGGGGACACG
CTTGAACCACCGAGAGTTTACTTTTCTGGACTTCCCAGAGCTGCTGGCACGGGCATCAGCCCC
CGGGGCTGGACAGAGGTTTGGGTGGGGGAGCGTTAGGCCGACCGCGCCCAAGCGGGGTTGGC
GGCCACGCGGTGTGAGGCCGGCGGGGCGCAGAACGC CGCGCGCCGGGTAGGTAGGGCTGGGAA
CGGGGCA CCGGAAG CGCGAGACCCGCCGAACGCTCCGGATAGCGCAGGCGCACTGCGGCGG
GCGCGGGCACGCTTGGCACGCAGGCGCCCTGAGGCGTGGGAGCGCGGCGCGGCCGGAGGTT
GGGGCGGGCGCCCTCCAG TCGGCAGG GCAGGCGCGCGGCCCTCCGGCGGTTGGGGAGGGTT
TTCCGGAAGGTTCCGGAGGCTTCTGGAAAAGCGCCCGCGCTGGGCGGCCCGCTCTATATA
AGGGCGCGCGGGGGCGCGC GCCAGTT GCCTCTGCGTCCCGGTGCTGCGTCCGTCGCGTCCG
GTGCT AGTCCT T CAGCCAAGGTGAGGGGAGGCGGTTGTAGGGGACGCGGGGATCCACGGGCGG
GTGGGCGGCGGCAGGGCAGCCCCGGAGCCCCCTGAGGCCGGCGGGCCCCGGGCGGCGGTGAT
TAACCGGTAGGCCTCGGGCCTGCGGCGGTCTCCGTCCCCTCCCCCGCCCTCCCCCGCCG
TCGGGAGCGAGTGGGCCCCGGGGCTCCGGGTCTGCGGGATGGTGAGGGCCGGGGTCTGCGGC
CCCGAGGGGGTCCCCGGGCCGCGGTTGGGGCGCCCGGCCGAGCGGCGCCGTGCGGCCGG
CGTGGGCGGCGGCGGCAGCGGCGGCGCCGCCCGCGGATGAGGTCATCCTTTGTCAGCCGGC
GCCTCCGCGCGCCCCGGTCTCGGGAGCGGGAGCGAGACGCGGCGGGGAGGGTCC
```

**Figure 8.** *HSP90AA1* methylation analysis by bisulfite genomic sequencing (Primers are in light grey). Exons are dark grey shaded. CpGs of the CGI are highlighted in green and the CGI limits in blue: -745 to +1445 respect putative TSS(A<sub>+1</sub>). Uncertain methylated CpGs are in yellow. Some core promoter motifs are depicted: BRE is double underlined, TATA-box simple underlined and TSS (A<sub>+1</sub>) at the Inr highlighted in pink. <sup>m</sup>G<sub>.660</sub> and g.632 methyl-CpG are circled.

```
ggagaattcgggactgtgtcccttttaacattcccgcgccaccctccccctgcttgggccaatcggcctaaggacaggaaga
gtccccatactgtagtgccgggggaatggtggtttgttccaggccccacctctgtttcagaaccagccgaatagaccacaaggctct
ggagtccccctgtgctccccctgcacacaagctggtcaccctcgtctcacagaccctgcaccccccaaccaggctgggagctgagc
ctcggcgctaaaggctctggcaggcactgttgcggcaatctctagggccctgaacctataacgctgccccctgcaagagagatgctc
gacctagaccctctaatcgtctccgggaccctccg/c-aaaccctaaacctggaacctgagcctcgaatcaccatgcccaagcataac
ctgtgacaaggactggaccctcaccctggctttgatcccctaaaggcctcaaggctgaaaacctagccccaggccccataggc
gggggtcccagctgctggggggatccggagccacactatctggggacaacctgaaccacagagttactttctggactcccag
agctgctggcaaggaaatcagccccgggctggacagaggttgggtgggggagcttaggccaaggcccaagggggttg
gagggcagaggttgtaggcagggggccagaaaccggggccggggtaggtagggctgggaaaggagcacgaaggagag
accctcagaaacctcagatagcaggccactgaggggggggcggcaagcttggcaacaggccctgagggctgggagggc
aggggggcccaggagggggcaggcctccagtggacaggccaggccggcagctcagggttggggagggtctctcagg
aaggttagggagctctggaaaaggccctggccctggggggccctcctataataagggggggggggccagcttgc
ctctgctcccgggtgctggctggctgggagccctgctagtctcagccaaggtaggggaggaggtttagggggaagggtttaggggga
ccaaggcggggggaggaggggcagccccaggccccctgagccctgagccgggaagccaggggggccttgattaaccagtaggcc
tagggcctgggggtctcccctccccccccccccccccccctccccccctccttagggagcagtaggggcccgggctcggggtggtg
ggatggtgagggcggggctctggcccaggaggggggtcccaggccaccgggtggggggcccaggccaggcagggggcctggc
ggctggggggggggggcagggggccggccggccccggaaatgaggtcatccttgtcagcggggcctccggggcccctgctc
gggaggggagagaggggggggggggtccggcgtggcctcaaggctccagttcccagcagatgcccaggagaccag
ggccaggaccagcagatggaggaggagggtggagagctccctccaggggaaattggccagttgagtactgatcatcaaca
ctttactcaataaaagagattttctgaggaggagctgattcaaacctcctcagaagtaagtctcattggaggtgggttgtgtgt
gaaaggcancctgttttagggggaagttgtgtcatcaagggtctttttaaataagagccctggacaagatcaggtatgagagct
tgaccatcccagtaaatagactggggaaagagctgcacattaatctcatcagaaacaagcaagataaacctcaccatggtgga
cacggcattgggatgaccaaggcagatttgatataaatctgggacaccatggccaagtctgggaccaagcctcatggaggtcttg
caggctggcctgatctccatgatggccagtttgggtggggtttatcagcctatctgggtgctgagaagtgacagtcacaa
ccaaaataaaatgatgagcagatgctggggagctctctgaggagggtcctcacaagtaggagctgacaaaggtaggaccagag
aacaggtgggtgctttaggcactggagccccaaagaagggtgaagtttgggtgtagctttaaaccctgttgcctggaggggtg
atagcatgcacatgcactgtggttgggtatcttgggcaaggcccttaactgccctgagactctgccagttgtagtaaccacaagct
ctctgctcatagactttggataactgtcctagagagcaaaatagtggttttttggcctgatacggccattgtgatgaaatggc
cctcaactctggactcttccaggagaaccgctggggcggtggaaacaaaaggttattctgcatctgaaagaaagaccagactgagta
ctggaggagcgagaaataaaggagattgtgaaagaagcatctcaggttcatggctatccatcactcttttggtaggttccatgta
cctgtatgtaggttactaaagtgttgaactgaaatcctgtagtctcctggctgcttttataagagatcccactgtcactaaat
cctggcctgggtgggtcagactcctgcaaggttcccagggctgtaaatagatcctcagaatgggtgagtagcctctctggactag
gttctcaggtggatggtgtaagtgctgcagctctcttagcctttccccagtttgggagagggtgagtagattgtcctc
```

**Figure 9.** Graphic representation of the *HSP90AA1* CGI methylation pattern across different tissues.

**MCAB:** Manchega Control Adult Blood; **MHSB:** Manchega Heat Stress Adult Blood; **RYH:** Rasa Young Heart; **MAB:** Manchega Adult Brain; **MAL:** Manchega Adult Liver; **MYO:** Manchega Young Ovary; **RYT:** Rasa Aragonesa Young Testicle; **MYT:** Manchega Young Testicle **MAT:** Manchega Adult Testicle; **MAS:** Manchega Adult Sperm. Numbers in the top of the table indicate the position of the CpGs. CpGs from 22 to 154 are unmethylated (not shown in the figure). The signs at the bottom of the table indicate positions at the promoter region of the gene, exons (E) and introns (I).



Differences on the epigenetic mark patterns

To elucidate how the epigenetic mark patterns could change between tissues and ages we have compared brain, liver, testicle and sperm from adult rams and heart and testicle from young animals (90 days). Reverse strand was amplified and analyzed separately in two fragments (511bp and 418bp) to encompass the promoter region. ASM pattern in the complementary strand was confirmed in all tissues (Hajkova et al. 2002).

Differences on the epigenetic mark patterns in several tissues

According to the bisulphite sequencing results, the same border 5' CpG island DNA in heart, brain, ovary and blood were shown. In every tissue, except adult testicle and sperm, there were promoter allele-specific hemimethylation of the GG<sub>-660</sub> genotype. Regarding epigenetic marks on the gene body, germ cells (young ovary and

young testis) were free of epigenetic marks in exon 2, intron 2 and exon 3, in contrast, liver and brain had epigenetic marks that extended part of exon 3.

### Differences on the epigenetic mark patterns in two stages (young versus adult)

To examine whether there were differences in the epigenetic mark patterns between different ages, we have compared testicle from adult and young (90 days) animals and sperm from adults. We have shown a progressive loss of promoter allele-specific methylation with cell differentiation. Moreover, all epigenetic marks disappeared at the promoter of both adult testicle (where the predominant cells are spermatozoa) (Oakes et al. 2007) and mature sperm. In addition, we saw a progressive increase of epigenetic marks in the gene body of these tissues also linked with cell differentiation stage (Fig.9).

### Transcription factor binding affinities involving methylation

As above mentioned, g.660G>C creates a cis-regulated ASM. When the G<sub>-660</sub> allele is present, it forms a CG group where the cytosine can be methylated (meG<sub>-660</sub>). To investigate if this methylation could interfere with the binding of transcription factors, we performed an EMSA to compare allele G<sub>-660</sub> vs. meG<sub>-660</sub>. We observed that there was no difference in transcription factor binding pattern even though it seemed to have more initial binding affinity to meG<sub>-660</sub> (Fig.10). However, when the methylated oligo containing the G<sub>-660</sub> allele was competed with 50-fold excess of cold oligo of any of both alleles, there is the same degree of decrease in the intensity of the bands, which indicates that there was no difference in binding complexes with or without methylation. We also designed a long oligonucleotide (60pb) where two methylations, one in the CpG site formed by the G<sub>-660</sub> allele and other at the g.632CpG, were included. Even though, due to the length of the oligonucleotide, undesirable secondary structures were produced and EMSAs assays were not conclusive (data not shown).

## **DISCUSSION**

The present study was based on previous expression results obtained in the previous chapter that was limited by the high co-segregation of the polymorphisms studied. We could finally distinguish which of the polymorphisms of the linked block already selected as candidate, has some weight in the transcription mechanisms causing differences in expression. In the previous chapter, expression rate changes were

associated with alternative genotypes of one of three polymorphisms (g.703-704insAA, g.660G>C and g.528A>G) that were identified to be completely linked at least in Manchega breed.

Even though qRT-PCR is a strong, simple and widely used technology has some limitations. In our case, it was impossible to distinguish which of those three polymorphisms, was the one responsible of expression differences mainly under basal conditions. Besides confirming the g.667\_668insC role under heat shock conditions, the possible implications of an additional INDEL (g.666\_667insC) have also emerged.

### ***In vitro* expression**

In this study, and based on the knowledge that mammalian transcription factor binding sites are well conserved, we could probe the direct effect of the different polymorphism along the *HSP90AA1* ovine promoter. Functional studies performed in the present assay have shed some light on the true responsible of the differences in the promoter activity, already confirmed by quantitative expression assays. We have observed *in vitro* two INDELS, g.666\_667insC and g.667\_668insC, that produce high differences in gene expression under heat stress conditions. It is also true that SNP g.660G>C has somehow a role in gene expression activity, both under basal and heat stress conditions confirmed by qPCR studies and *in vitro* assays. As it has been previously explained, even small differences in activation of this gene, can cause great differences in the overall equilibrium of the cell homeostasis as HSP90 is one of the most abundant proteins in the cell (Taipale et al. 2010). So, changes in the expression rate during thermoneutral temperatures can also be essential to cope with other type of stresses not directly related with temperature (Sud et al. 2007, Han et al. 2009, Yang et al. 2011, Kim et al. 2013b, Tian et al. 2014). Both polymorphisms, g.703\_704insAA and g.528A>G, completely linked in Manchega rams to g.660G>C have no real effect in the expression differences observed under thermoneutral or heat stress conditions. In fact, even though, g.703\_704insAA had been hypothesized to could have somehow effect in the expression rate due to the glucocorticoid receptor binding, this fact has not been confirmed as no differences in the luciferase assay was found between alternative alleles.

In addition, we have found that two cytosine insertions are the responsible for the high upregulation of the gene as response to heat stress. Nonetheless, both insertions are not essential, as we have observed that the addition of a single cytosine (I<sub>.668</sub>D<sub>.667</sub>) is enough to cause higher gene expression levels (28.3% more than D<sub>.668</sub>D<sub>.667</sub>)

under heat stress. The addition of an extra cytosine (I<sub>-668</sub>I<sub>-667</sub>) in this sequence has similar effects, only lightly exacerbated (30.3% more than D<sub>-668</sub>D<sub>-667</sub>).

Hsp genes inducible expression is regulated by the heat shock transcription factors (HSFs) that exist as inactive proteins mostly in the cytoplasm. In response to physiological and environmental stimuli, heat stress in particular, these proteins bind to the promoter targets sequences (HSEs) triggering transcription of heat shock genes and the formation of heat shock proteins (HSPs) (Morimoto 1998, Pirkkala et al. 2001). Besides HSE, there are a number of promoter sequence structures involved in regulating *HSP90AA1* gene expression.

The transcription from TATA core promoters occur from a single site or localized cluster of sites, in contrast to transcription from CGI where it is initiated from multiple transcription binding sites (e.g. Sp1). *HSP90AA1* promoter has a *hybrid* structure, a TATA-box with CGI, which could result in a dual behaviour of the gene transcription: TATA-box+ Inr (*synergistic configuration*) and HSE, in a single core promoter can direct “strong transcription” initiation under heat stress. In addition to that, CGI may also function in concert with the “basal transcription” factors to mediate transcription initiation with or without heat stress stimuli (Butler et al. 2002). Furthermore, from the ontological point of view, *HSP90AA1* promoter also has dual behaviour with specific biological functions: TATA containing genes are more often highly regulated, such as by biotic or stress stimuli, in return, TATA-less promoter and/or associated with CGI are frequently involved in basic housekeeping (HK) processes (Basehoar et al. 2004, Kimura et al. 2006, Yang et al. 2007). Dual behavior means that *HSP90AA1* gene is more inducible than its constitutive counterpart (*HSP90AB1*) for two reasons: 1) Although *HSP90AB1* has a non canonical TATA-box associated with CGI has neither Inr nor HSE and 2) *HSP90AB1* is much more stable than several commercial HK tested in a previous work (Serrano et al. 2011)) and in Chapter 1.

### **Cis-regulated allele-specific methylation**

We have previously shown that g.660G>C alternative alleles are associated with expression changes in blood. Gene expression in animals carrying the GG<sub>-660</sub> genotype was reduced relative to those carrying the CC<sub>-660</sub> one (Chapter 1). Bisulphite sequencing confirmed the presence of one *cis*-regulated ASM caused by this transversion, which is flanked downstream by other non-polymorphic methyl CpG site at (g.632 methyl-CpG) in 120 peripheral blood samples.

As we have shown during this study, SNP g.660G>C can be target of epigenetic modifications. The effects of methylated CpG on gene activity are based in two general mechanisms that contribute gene silencing. Either methylation at promoter specific sites prevents the binding of the transcription factors, or methylation attracts methyl binding proteins (MBPs) (Nan et al. 1997). To investigate whether methylation CpG created by the transversion g.660G>C interfere with transcription factors binding, an EMSA assay was performed. As shown in Supplementary Figure I no differences in the intensity of bands in the EMSA performed to compare methylation vs nonmethylation were found. This means that, if a transcription factor binds this DNA sequence could do it independently of the methylation existence. The methylation profile, at least in this point, seems not affect protein binding. However, the repression effect of methylation, can act without altering the binding of transcription factors (Salvatore et al. 1998). Thus both mechanisms, methylation repression and TF binding are independent processes that only in few cases are directly correlated (Medvedeva et al. 2014). Thus a MBP, primary factor in most cases, can act as indirect repressor without altering the binding of transcription factors (Nan et al. 1997, Salvatore et al. 1998).

The identification of methylation-dependent transcriptional repressors is determined by the detection of proteins with affinity for methylation (Nan et al. 1997). MeCP2 is an abundant MBP with high affinity for a specific DNA sequence. It depends on the density and location of methyl-CpGs in gene promoters and interferes with regulatory components of the transcription complex (Nan et al. 1997). The sequence fragment spanning from g.660G>C to the methyl site at position g.632 could constitute two putative MeCP2 binding sites for several reasons: (i) the fragment is found in the gene promoter; (ii) methyl-CpGs are symmetrically (bottom strand was tested by sequencing); (iii) CpGs have adjacent A/T motifs (minor groove of the DNA double helix); and (iv) the distance between 2 methyl-CpGs needed for the binding of two MeCP<sub>2</sub> must be greater than 12pb (Nan et al. 1993, Klose et al. 2005) (Figure 10). On the basis of the sequence characteristics and the results obtained in the EMSA assays we can hypothesize that MeCP2 could indirectly repress transcription at this point.

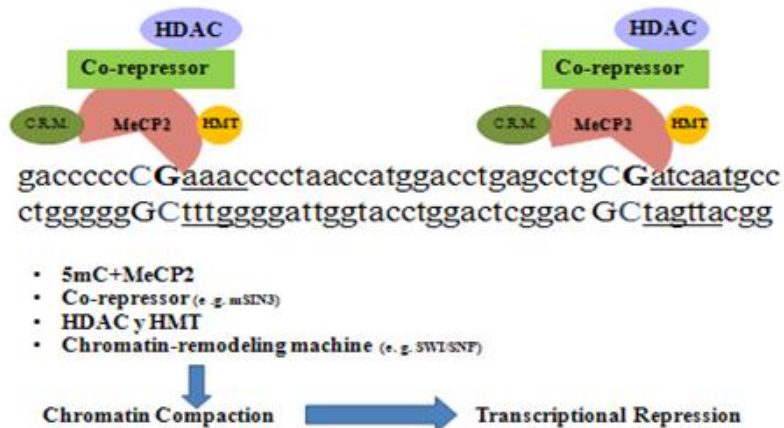
We also saw that ASM was influenced by DNA sequence in different tissues, ranging from complete association of methylation and genotype in peripheral blood to non-existent or incomplete association of methylation and genotype in sperm and other tissues, respectively. Moreover, g.632 methyl CpG disappears in adult testicle tissue and sperm. We have enough data to confirm that g.660G>C ASM is tissue-type-specific and dependent on cell differentiation (i.e. testicle tissues at different ages).



An important question is whether the apparition of g.660G>C eliminates a methylation site (wild allele would be G<sub>-660</sub>) or it creates a *de novo* methylation (wild allele C<sub>-660</sub>). We suggest the latter is the correct choice because cytosines are highly mutable to thymine in CpG dinucleotides (Coulondre et al. 1978) so it is unlikely to lead to a guanine mutation. Moreover, the C<sub>-660</sub> allele is the most common one in 31 sheep breeds, goats and other species of wild ruminants analyzed (See Chapter 3).

Allelic alternation of another 4 SNPs from the promoter region could lead to methyl CpGs g.601A>C, g.528A>G, g.522A>G and g.304A>G and to be functional cis-regulatory variants, the first one, by disrupting a CpG methylation site and, the other three, by creating a CpG methylation one. However, none of them are methylated in the genotyped tissues.

**Figure 10.** Possible mechanisms of transcriptional repression by MeCP2 (Adapted from (Wade 2005)). Interaction of MeCP2 with two methylated DNA sites results in local recruitment of chromatin-remodeling machine. This factor alters histone-DNA contacts. Furthermore, histone deacetylation by HDAC and histone methylation by HMT facilitates formation of repressive chromatin conformation and therefore, contributes to transcriptional repression.



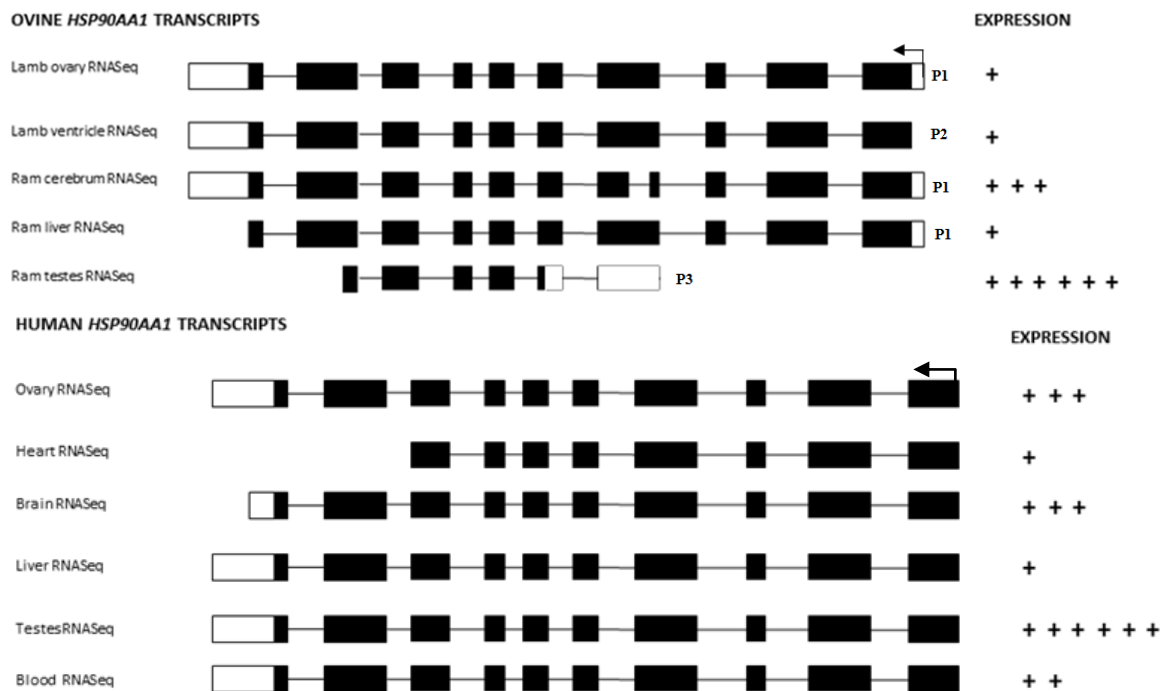
### Tissue-specific CGI epigenetic marks pattern

HSP90AA1 associated CGI is constitutively unmethylated (Feltus et al. 2003, Bock et al. 2006). It shows a differential degree of DNA epigenetic marks at the body gene and allelic-specific methylation across several tissues and development stages. These results could determine a functional role in the epigenetic control of gene expression (Meissner et al. 2008, Previti et al. 2009). However, in our case, serious

doubts are raised about correlations between changes in expression and epigenetic marks and between epigenetic marks and the tissue-specific gene expression. Firstly, analysis of CGI epigenetic mark patterns have been based on the use of DNA bisulphite sequencing technique and restriction digestion which do not distinguish between 5-methyl cytosine (5mC) and 5-hydroxymethyl cytosine (5hmC) (Huang et al. 2010, Jin et al. 2010, Nestor et al. 2010). Secondly, we could not yet perform *ad hoc* tissue-specific expression studies. Finally, DNA epigenetic mark patterns of the entire gene are still not available. Anyhow, previous works of tissue-specific transcripts and RNAseq expression of ovine and human *HSP90AA1* gene is fully available online ([http://www.ensembl.org/Ovis\\_aries](http://www.ensembl.org/Ovis_aries)) ([http://www.ensembl.org/Homo\\_sapiens](http://www.ensembl.org/Homo_sapiens)) (Fig.11). Accordingly, these disadvantages allowed us only to make descriptive and comparative studies of CGI epigenetic mark patterns without going deeper into the study of other epigenetic modifications (e.g. histone modifications): (i) 5hmC is present at the promoters and intragenic regions while 5mC predominates in intergenic regions (Jin et al. 2011); (ii) significant differences of 5hmC distribution in different tissues were observed (e.g. Brain > Liver >> Heart) (Li et al. 2011); (iii) these molecules have potentially different roles in the epigenetic regulation: 5hmC is associated with the body of transcribed genes and positively correlated with transcription levels (Nestor et al. 2012) whereas 5mC at gene promoters is involved in transcriptional repression (extensively documented e.g. (Jones et al. 2001, Klose et al. 2006).

Tissue-specific differentially *methylated* (methylated or hemimethylated) gene body regions may be associated with expression (Rakyan et al. 2008). The potential to suppress transcriptional noise or repress spurious transcription (Huh et al. 2013), interfere with transcription elongation (Lorincz et al. 2004) and repress the activity of intragenic promoters (Maunakea et al. 2010) have been described. However, the functional relationship of the gene body (intragenic regions) epigenetic mark patterns with gene expression is not clear.

**Figure 11.** Tissue-specific transcripts from ovine and human *HSP90AA1* gene. Only tissues here studied and their reverse strands are shown. The arrow indicates the direction of transcription. P1, P2 and P3 indicates the type of alternative promoter. Graphical representation of tissue expression with RNAseq total alignments (number of alignments + alignments omitted) respect to total alignments of liver (tissue with the least number of alignments) (available at [http://www.ensembl.org/Ovis\\_aries](http://www.ensembl.org/Ovis_aries) and [http://www.ensembl.org/Homo\\_sapiens](http://www.ensembl.org/Homo_sapiens)).



Our results suggest that there might be a relationship between different gene-body DNA epigenetic mark patterns and gene expression. Markedly different epigenetic profiles, such as liver and sperm (ram testes), obtained the lowest and the highest gene expression, respectively (both ovine and human expression data) (see Figure 11). Epigenetic mark changes (increase or decrease) in a particular tissue, can compromise gene expression in that tissue specifically (Hellman et al. 2007, Rakyan et al. 2008). Other phenomenon observed in this chapter, are testis and brain differences in their epigenetic profiles. Identical gene body epigenetic mark profiles and different gene promoter epigenetic marks profiles showed different expressions: highest gene expression in ram testes and mid-level gene expression in brain (both ovine and human expression data) (see figure 13 and Supplemental table 3). For the opposite case,

different gene body profiles and identical gene promoter profiles of ovary and blood tissues (only for genotype CC<sub>-660</sub> and CG<sub>-660</sub>) showed less difference in the expression levels (only human expression data) (see Figure 13). These results confirm previous genome-wide studies that showed a small positive correlation between gene body epigenetic marks and gene expression (Rakyan et al. 2008).

The functional diversity of genes generating such expression complexity is also based on alternative splicing (AS), alternative promoters (AP), AS and AP combination and their epigenetic regulation. AS produce different specific transcripts which are translated into diverse proteins with different functions and structures (Black 2000, Kimura et al. 2006). The use of AP helps to regulate and increase the transcriptional complexity of the gene (Landry et al. 2003, Kimura et al. 2006). Characteristic epigenetic patterns in intragenic regions, around exons and exon-intron frontiers, regulate tissue-specific expression from AP.

On the basis of the results obtained in previous works of tissue-specific transcripts of the ovine *HSP90AA1* gene published online (Fig.11) we identified three independent positions of TSSs, therefore, three alternative promoters. Two of them, P1 and P2, have a common downstream exon, two TSSs and have the same open reading frame (ORF). P1 is active in ovary, cerebrum and liver, whereas P2 is active exclusively in heart (from tissue data available). The last one, P3, active in testes, contains a TSS in half of the exon 5 and one alternative ORF. Furthermore, AS from *HSP90AA1* gene generates five transcript isoforms and three protein isoforms in a tissue-specific manner (Fig.11). These observations support the evidence viewed in other species that alternative promoters regulate the products of alternative splicing processes (Pecci et al. 2001).

Although *HSP90AA1* has its highest expression in brain and testis tissues (Csermely et al. 1998) we cannot consider it a tissue-specific gene, since its promoter structure is CpG rich, typical in HK genes (Meissner et al. 2008) and its ubiquitous expression (Liu et al. 2014). Instead of that, when the CGI promoter is active (non-stress conditions), gene expression regulation is tissue-specific favoured by the connection between AP, AS and epigenetic marks. The effect is manifested as alternative splicing in a tissue-specific way, where in our case testicular and brain tissue transcripts are clearly different. When considering tissue-specific differentially *methylated* regions, we saw that they are associated with intragenic regions included within CGIs. The majority of the tissue-specific differentially *methylated* regions were

located in both introns and exons of the sequence here studied: E2, I2, E3 and I3 (Fig.9). The relationship between those regions and splicing patterns when comparing sperm respect other tissues is defined as “negative” regulation because relative increased epigenetic marks seem to be associated with exclusion of several exons (E1, E2, E3, E4 and part of the E5). On the contrary, a “positive” regulation seems to occur, in our case when comparing the other tissues, as those marks seem to be associated with exons inclusion (Fig. 9) (Wan et al. 2013). These mechanisms involving TFs (Supplementary Table 3), other DNA-binding proteins and their corresponding sequence motifs containing CpGs are likely to regulate splicing synergistically with epigenetic marks (Shukla et al. 2011, Wan et al. 2013).

As was previously described in human (Grunau et al. 2000) concerning CGI patterns across tissues and individuals, in this work differences in CGI epigenetic mark patterns between tissues of one individual and same CGI epigenetic mark patterns in identical tissue types from different individuals have been observed.

Additionally, our findings show different methylation patterns from the same tissue at different ages: young testicle (different cell types) versus adult testicle (the predominant cell is the spermatozoid). This phenomenon may be due to differentially expressed genes of each cell type during the spermatogenesis phases (Dormant Spermatogonium, Spermatocytogenesis, Meiosis and Spermiogenesis) and the epigenetic mechanisms implicated in the regulation of the process. This agrees with normal development and control of sperm-specific gene expression. In addition, genes that are expressed for the overall functioning of the cell (e.g. *HSP90AA1*) at initial stages have different epigenetic profiles regarding later stages (Grunau et al. 2000). In our experiment (see Fig. 9), the lack of epigenetic marks at the *HSP90AA1* CGI in sperm and adult testicle versus young testicle confirms that this CGI has not a functional role in the epigenetic control of gene expression at later stages of spermatogenesis.

Moreover, two fragments from Short Interspersed Elements (SINEs) were identified. One of them had 35 bp at the intergenic region studied, which is quite analogous to the 3' Bov-tA2 or SINE2/tRNA of ruminants (212bp) and the other one had 50 bp in the third intron of the *HSP90AA1* gene which has some analogy with MIRc SINE2/tRNA of mammals (268 bp) (Data not shown). The insertion of mobile elements near the transcription control regions contribute to the control of gene transcription. Bov-tA2 or SINE2 / tRNA of ruminants could be a good candidate as an alternative

transcriptional regulator mechanism if the entire sequence is found in this region (not yet sequenced). Anyhow, the exact transcriptional control of this intergenic SINE found remains still unknown.

We have described the epigenetic pattern shown in different tissues from young and adult animals. We have not discovered the cause that produces this epigenetic changes present in the ovine *HSP90AA1* promoter. Moreover, it seems not probable that these modifications are due because of the heat. We need to continue focusing in the specific environment where these animals are reared and contribute to these epigenetic changes.

Taking into account the results obtained, it is suggested that the transcription of the *HSP90AA1* ovine gene is regulated by a cooperative action of transcription factors (TFs) whose binding sites are polymorphic but the influence of epigenetic events should be also taken into account. At least under thermoneutral conditions where the CGI promoter is active and g.660G>C role has been highlighted. These new findings get us closer to the principal responsible of the differences of the expression profile that carries out important consequences in sheep.



# Chapter 3

## An adaptive role gene

Looking for adaptation footprint in the *HSP90AA1* ovine gene. Learning from the **Wild**.





### INTRODUCTION

Caprin evolution during the Pleistocene was a series of pulses of rapid speciation followed by longer periods of gradual change and adaptation, sometimes accompanied by extinctions (Shackleton 1997). The Subfamily Caprinae includes a widespread and diverse group of ungulates (hoofed mammals) that are most extending from the arctic to the equator. They are present in three continents including 70 different countries. Wild Caprinae were the ancestors of two of the most important species of domestic livestock - domestic sheep (*Ovis aries*) and goats (*Capra hircus*). Present day populations of wild Caprinae represent a potential source of knowledge of adaptation genetics which can be used to improve or adapt current domestic breeds to less productive conditions (Shackleton 1997).

Sheep was one of the first species to be domesticated due to its small size, docile behavior and high adaptability to very different environments. This domestication process must have involved a genetically broad sampling of wild stock and also the persistence of cross-breeding with wild populations following the initial domestication events (Kijas et al. 2012). Domestication pressure over animal's life had as consequence that natural selection loosed impact over their biological fitness giving up the turn to artificial selection imposed by humans over productive traits (wool, meat, milk). However, sheep is one of the livestock species with less intensive management systems and therefore, could have retained some genome footprint in genes related to adaptation to different environments from its wild ancestors.

Climatic factors like temperature and humidity play an important role in determining species distributions and they likely influence phenotypic variation of populations over geographic space (Hancock et al. 2011). Such variation can reveal the action of natural selection when it is correlated with variation in environmental factors over multiple independent geographic regions. Correlations between phenotype and environment may be revealed at the level of individual genetic polymorphisms, where at some loci, allele frequencies strongly differentiate populations that live in different environments (Coop et al. 2010). Such correlations can arise when selection pressures exerted by the environmental variable are sufficiently divergent between geographic locations, such that differences in allele frequency can be maintained in the face of gene flow (Lenormand 2002).

Several studies have examined the distributions of genetic variants in candidate genes for traits that vary with climate. Candidate gene approaches in humans as well as

in several other species support roles for selection at genetic variants that underlie phenotypic variation. For example, in humans, candidate gene studies have yielded evidence that variants involved in sodium homeostasis and energy metabolism are correlated with latitude and climate (Hancock et al. 2011) and those related with type 2 diabetes and obesity are strongly correlated with climate variables (Hancock et al. 2008). Also a decrease in the frequency of variants implicated in salt sensitive hypertension had been correlated with increasing distance from the equator (Thompson et al. 2004). In *Drosophila melanogaster*, variants involved in circadian rhythms, aging and energy metabolism were correlated with climate (Sezgin et al. 2004), in *Arabidopsis thaliana*, variants associated with flowering time were correlated with latitude (Stinchcombe et al. 2004), and in pines several genes contain variation were correlated with temperature (Grivet et al. 2011). In addition, evidence of selection related to climate has been shown in *Drosophila melanogaster* (Gonzalez et al. 2010), *Fagus sylvatica* (Jump et al. 2006) and *Pinus taeda* (Eckert et al. 2010) analyzing hundreds of transposable elements, AFLPs and nearly 2000 SNPs, respectively.

The heat shock response is among the most important and ubiquitous fact in nature. Heat, both quantitatively and qualitatively is one of the best inducers of Heat Stress Proteins (Hsp). However, there are only few publications on the role of Hsp90 function in species adaptation and survival under extreme conditions (Arad et al. 2010, Jarosz et al. 2010, Reddy et al. 2011). In addition, it has been exposed in Chapter 1 and 2 differences in the *HSP90AA1* transcription rate depending on genotype combination of some polymorphisms located at its promoter.

This chapter has the aim to: 1) study the relationships among the frequencies of 11 polymorphisms located at the *HSP90AA1* gene promoter and climatic and geographic variables of locations where 31 sheep breeds from Europe, Asia and Africa are reared. 2) Study the *HSP90AA1* promoter sequences in 9 species of the Caprinae and in 2 species of the Bovinae subfamilies to determine polymorphisms history and to contribute to elucidate the phylogeny of one of the most controversial subfamilies of the sub order *Ruminantia*.

## **MATERIALS AND METHODS**

### **Animal material, nucleic acid isolation, DNA amplification and SNPs genotyping.**

Animals from 31 sheep breeds from Europe, Asia and Africa and from several species of the Caprinae (9) and the Bovinae (2) subfamilies constitute the biological material of this work. Tables 1 and 2 show breeds, species, number of animals from each breed and species, location, country, continent and climatic and geographic variables.

Peripheral whole blood samples were collected in order to analyse 11 polymorphisms of interest located at the *HSP90AA1* promoter. Polymorphisms genotyped were: g.703\_704insAA; g.667\_668insC; g.666\_667insC; g.660G>C; g.601A>C; g.528A>G; g.524G>T; g.522A>G; g.516\_517insG; g.468G>T; g.444A>G. *HSP90AA1* promoter sequencing was done following the same procedure and primers as in Chapter 1.

### **Polymorphisms characterization and linkage disequilibrium estimation**

PLINK software (Purcell et al. 2007a) was used to estimate linkage disequilibrium among all pairs of the 11 polymorphisms measured as  $r^2$  in the whole sheep data and in each breed separately. Hardy-Weinberg equilibrium exact test, observed and expected heterozygosities for each breed were also calculated using PLINK.

**Table 1.** Sheep breeds, locations, countries and continents of origin and climatic and geographic variables\*.

Breed	ID	breed	Location	Country	Continent	n° animals	Lat	LON	MINaT	MAXaT	MThm	ANT	TW	TAR	MxR	MiR	HrA	HrMx	HrMi	THI	CTY
Akkaraman	AKA		Konya	Turkey	Asia	23	37.6	32.3	-0.2	23.1	30.1	11.5	23.3	31.5	41	4	60	80	40	22.02	SA
Kazakh Arkhar-Merino	ARME		Alma Ata	Kazakhstan	Asia	18	43.45	77.04	-6.1	23.6	29.4	9.1	29.7	65.3	107	26	62	77	45	22.52	SD
Assaf	AS		Haifa	Israel	Asia	30	31.86	35.21	1.4	29.9	40	20.7	13.4	53.4	148	0	68	72	60	28.36	SD
Awassi	AW		Haifa	Israel	Asia	30	31.86	35.21	1.4	27.4	40	20.7	13.4	53.4	148	0	68	72	60	26.11	SD
Bajdarak	BAJ		Irkutsk	Russia	Asia	22	51.83	107.4	-18.8	17.5	34	0	36.3	46.8	120	9	75	86	58	17.26	SA
Bni Guil	BNI		Fez	Morocco	Africa	27	33.56	4.59	9.6	27.1	35.8	17.8	17.5	53.7	85	1	67	79	56	25.8	SD
Boujad	BOUJ		Marrakech	Morocco	Africa	24	32.86	-6.96	12.2	28.3	39	19.6	16.1	28.2	41	1	58	66	47	26.49	SA
Bozakh	BOZ		Kirovabad	Azerbaijan	Asia	24	40.53	46.02	2.7	25.1	30.9	13.9	22.4	29.4	51	11	65	78	53	23.94	SA
Caucasian	CAUC		Astrakan	Russia	Asia	25	45.71	42.88	-5.5	25.2	30.6	10	30.7	21.6	25	10	70	86	54	24.2	A
Churra	Ch		León	Spain	Europe	23	42.58	-5.65	3.1	19.6	37	10.9	16.5	55.5	70	24	68	83	55	19.08	SD
Churra Lebrjiana	Cl		Jerez	Spain	Europe	26	36.75	-6.06	10.7	25.7	38	17.7	15	58.7	109	2	67	79	54	24.54	SD
Churra Tensina	Ct		Huesca	Spain	Europe	33	42.08	-0.33	4.9	23.4	37	13.6	18.5	53.4	62	20	63	81	48	22.37	SD
Daglic	DGL		Eskisehir	Turkey	Asia	24	38.5	30.3	-0.1	21.4	28.6	11.1	21.5	38.8	51	9	70	84	59	20.75	SA
Kazakh Edilbai	EDIL		Aktjubinsk	Kazakhstan	Asia	30	52.32	77.03	-14.1	22.4	39	4.7	36.5	31.1	33	19	66	82	50	21.56	SA
Ivesi	IV		Diyarbakir	Turkey	Asia	15	37.5	40.1	1.5	31	38.1	14.9	29.5	49.8	76	1	55	77	29	28.68	SA
Russian Karakul	KAR		Astradan	Russia	Asia	15	45.37	46.04	-5.5	25.2	42	10	30.7	21.6	25	10	70	86	54	24.2	A
Moldavian Karakul	KARM		Kisinev	Moldavia	Europe	15	46.93	28.75	-3.3	20.9	37	9.6	24.2	54.7	75	27	73	86	62	20.36	SD
Karabakh	KRB		Kirovabad	Azerbaijan	Asia	24	39.82	46.7	-0.4	23.6	38	11.9	24	29.4	51	11	65	78	53	22.6	SA
Karachai	KRC		Krasnojarsk	Azerbaijan	Asia	28	43.04	44.21	-1.6	18.2	37	0.8	34.2	45.0	79	12	68	76	53	17.82	SA
Karayska	KRY		Samsun	Turkey	Asia	22	41.2	36.2	6.9	22.7	26.2	14.3	15.8	69.2	89	29	72	79	65	21.98	SD
Kivircik	KVR		Bursa	Turkey	Asia	16	40.1	29	5.2	24.1	30.3	14.4	18.9	70.6	118	16	72	78	62	23.26	SD
Finnsheep	L		Jyvaskyla	Finland	Europe	30	62.31	27.17	-10	15.7	20.8	2.6	25.7	64.0	91	30	80	91	65	15.62	SD
Latxa	LX		Vitoria	Spain	Europe	41	42.88	-2.73	4.7	19.1	30	11.4	14.4	78.2	89	42	75	84	71	18.74	SD
D'Man	MAN		Ouarzazate	Morocco	Africa	26	30.93	-6.9	9.3	29.5	40	18.9	20.2	110	19	1	41	64	22	26.74	A
Spanish Merino	ME		Cáceres	Spain	Europe	29	39.29	-6.22	8.2	25.6	39	16	17.4	48.8	65	4	57	76	35	24.11	SA
Manchega	MINCH		Albacete	Spain	Europe	60	38.95	-1.85	5	24.1	38	13.5	19.1	36.7	52	9	63	79	44	22.99	SA
Olkuska	OL		Cracovia	Poland	Europe	30	49.78	21.34	-2.9	19.3	30	8.6	22.2	67.9	95	33	80	87	72	19	SD
Pramenka	PRAM		Belgrado	Servia	Europe	29	44.74	20.44	0.5	21.7	28.3	11.8	21.2	69.4	95	42	69	82	60	21	SD
Rasa Aragonesa	RA		Zaragoza	Spain	Europe	42	41.66	-1.01	6.2	24.3	35	14.6	18.1	31.4	38	15	65	77	53	23.23	SA
Saliz	SZ		Izmir	Chios	Asia	26	38.3	27.1	8.6	22.5	32.7	17.7	18.9	69.3	153	4	62	72	49	21.55	SD
Valle del Belice	VdB		Palermo	Italy	Europe	29	37.69	13.02	11.6	25.4	30.2	18.2	13.8	65.4	106	5	69	75	63	24.34	SD

\*LAT=latitude; LON=longitude; MAXaT=maximum average temperature; MThm=maximum temperature of the hottest month; MINaT=minimum average temperature; ANT=average annual temperature; TW (MAXaT-MINaT)=thermal width; TAR = total annual rainfall; MxR = maximum rainfall; MiR = minimum rainfall; HrA = relative average annual humidity (%);HrMx = maximum relative humidity (%);HrMi = minimum relative humidity (%);THI=Temperature Humidity Index (Marai et al. 2007) THI = T°C - (0.31-0.3 IRH)x(T°C-14.4) (Marai et al. 2007), T = temperature in °C, RH = relative humidity in %/100, THI<22.2=absence heat stress; 22.2>THI<23.3=moderate heat stress; 23.3>THI<25.6=severe heat stress; 25.6>THI=extreme severe heat stress; CTY= climate type (arid A=0-250mm; semi arid SA=250-500mm; semi damp SD=500-1000mm; damp D=1000-2000mm; very damp VD=> 2000 mm).

**Table 2.** Wild species from the Caprinae and Bovinae subfamilies.

Goat breeds: Guadarrama, Girgentana, Maltese, Angora, Blanca Celtibérica, cross.

Cattle breeds: Holstein, Avileña, Serrana, Pirenaica, Parda de Montaña.

Sheep breeds: Table 1.

Subfamily	Genus	Species	Name	N	Breeds	Country	Continent
Caprinae	<i>Ovibos</i>	<i>Ovibos moschatus</i>	Muskox	15		Norway	Europe
	<i>Capra</i>	<i>Capra hircus</i>	Domestic goat	57	7 breeds	Spain	Europe
		<i>Capra pyrenaica</i>	Iberian Ibex	5		Spain	Europe
	<i>Ammotragus</i>	<i>Ammotragus lervia</i>	Barbary sheep	14		Morocco	Africa
	<i>Ovis</i>	<i>Ovis aries</i>	Domestic sheep	836	31 breeds	-	-
		<i>Ovis ammon</i>	Argali	1		Mongolia	Asia
		<i>Ovis canadensis</i>	Big Horn	8		Canada	North America
		<i>Ovis vignei</i>	Urial	1			
		<i>Ovis orientalis musimon</i>	European Mouflon	31		Corsica	Europe
	<i>Rupicapra</i>	<i>Rupicapra pyrenaica</i>	Chamois	2		Spain	Europe
Bovinae	<i>Bos</i>	<i>Bos taurus</i>	Cow	43	5 breeds	Spain	Europe
		<i>Bos mutus</i>	Yak	13			

### Phylogenetic Relationship between sheep breeds

The relationship between breeds was examined using the Reynold's distance metric (Reynolds et al. 1983). Reynold's distance ( $D = -\ln(1 - F_{ST})$ ) matrix was estimated performing 90,000 permutations and a significance level of 0.05 was established. An Exact Test of population differentiation was performed to test the hypothesis of random distribution of the individuals between pairs of populations (Rousset 1995, Goudet et al. 1996) with 100,000 steps in Markov chain, 10,000 dememorization steps and a significance level of 0.05. An histogram of the number of populations which are significantly different ( $p < 0.05$ ) from a given population was generated. All analyses were made by using the ARLEQUIN 3.1 software (Excoffier et al. 2005).

A NeighborNet graph was constructed from the matrix of Reynold's distances using SplitsTree4 V4.13.1 software (Huson et al. 2006).

### Tests to detect association of loci frequencies with environmental parameters

Partial Least Square Regression (PLSR). Partial Least Squares multiple regression (PLSR) was applied to model the relationships between polymorphisms allele frequencies found in the 31 sheep breeds genotyped and a matrix describing environmental factors (14 geographical and climatic variables) as in (Fumagalli et al. 2011). The specific algorithm used to compute extracted PLSR factors was SVD

(Singular Value Decomposition). SVD is a factorization of a matrix which bases the extraction on the singular value decomposition of  $X^T Y$ .

For each polymorphism, the relationship between population allele frequency matrix (F) of dimension  $31 \times 1$  and environmental predictors matrix (M) of  $31 \times 14$  dimensions was assessed. F describes minor allele frequency (MAF) at each breed for the examined polymorphism, whereas M describes all the 14 environmental variables for each population.

In order to evaluate the fit of a model, values of explained variation,  $R^2$ , and predicted variation,  $Q^2$ , were computed as in (Fumagalli et al. 2011).  $Q^2$  provides a measure of how well a model predicts the observed data using a cross-validation procedure, which is in this case how well a model of environmental variables predicts the observed distribution of allele frequencies among breeds. If allele frequencies covary with environmental variables,  $Q^2$  will be large. Acceptable values of  $R^2$  and  $Q^2$  are totally dependent on the nature of the data. Lundstedt et al (T. Lundstedt 1998) propose  $Q^2 > 0.4$  and  $R^2 > 0.7$  as acceptable thresholds for biological data.

The number of factors chosen is usually the one that minimizes PRESS (Predictive residual sum of squares). However, models with fewer factors often have PRESS statistics that are only marginally larger than the absolute minimum. To address this, van der Voet (van der Voet 1994) proposed a statistical test for comparing the predicted residuals from different models. By applying the van der Voet's test, the number of factors chosen is the fewest with residuals that are insignificantly larger than the residuals of the model with minimum PRESS.

Uninformative variable elimination to remove those variables that are useless was made in the basis of two filter criteria: the Variable Importance in Projection values (VIP) (Chong et al. 2005) and the cumulative variance explained by the top two PLSR components. The idea behind VIP measure is to accumulate the importance of each variable being reflected by the loading weights from each component. It is generally accepted that a variable should be selected if  $VIP > 1$ , but a proper threshold between 0.83 and 1.21 can yield more relevant variables (Svante Wold 2001, Chong et al. 2005). A meteorological variable was declared important when 1) its VIP was greater than 0.83 and 2) the cumulative variance explained by that meteorological observation by the top two PLSR components was at least 40%. The criterion to assess that the elimination of uninformative variables improves the model is to compare PRESS values obtained for the complete and the reduced model (new). If  $PRESS_{new} <$

PRESS we can conclude that the elimination of uninformative variables improve modeling.

All computation were performed using the PLS procedure of the SAS 9.3 Statistical Package (Base SAS® 9.3).

Spatial Analysis Method (SAM). Other approach to assess the effect of any selection on polymorphisms across populations is the Spatial Analysis Method (SAM) developed by (Joost et al. 2007, Joost et al. 2008). SAM is based on the spatial coincidence analysis to connect genetic information with geo-environmental data. The logistic regression uses random binomial variables as response for the model, thus, each allele is set to '1' if it occurs in a given individual, and to '0' if not. Logistic regression is used to assess the significance of the models constituted by all possible marker-environmental variable pairs. The comparison of observed with predicted values is based on the likelihood ratio ( $G$ ) and Wald ( $W$ ) tests (David W. Hosmer 2000) to determine the significance of the models. For both tests, the null hypothesis is that the model with the examined variable does not explain the observed distribution better than a model with only a constant. A model is considered significant only if both tests reject the corresponding null hypothesis. To restrict the analysis to robust candidate associations the Bonferroni correction was applied and only cumulated tests in which both  $W$  and  $G$  tests were significant were used to identify associated loci (Joost et al. 2007). Computations were performed using the MatSAM v2Beta software (Joost et al. 2008).

### **Statistical analysis to detect loci under selection across populations**

Bayesian Test of Beaumont and Balding (Beaumont et al. 2004). This method evaluate the relationship between  $F_{ST}$  (population differentiation) and  $H_e$  (expected heterozygosity) describing the expected distribution of Wright's inbreeding coefficient  $F_{ST}$  vs.  $H_e$  under an island model of migration with neutral markers. This distribution is used to identify outlier loci that have excessively high or low  $F_{ST}$  compared to neutral expectations. Such outlier loci are candidates for being subject to selection (Antao et al. 2008). Low  $F_{ST}$  outliers indicate loci subject to balancing selection, whereas high outliers suggest adaptative (directional) selection (Beaumont et al. 2004). The Bayesian Test method of Beaumont and Balding was assessed using the LOSITAN (Looking for Selection In a Tangled dataset) package (Antao et al. 2008). Initially 100,000 simulations under the infinite allele mutation model were run using all populations and all unlinked loci to determine a first candidate subset of selected loci in order to remove them



from the computation of the neutral  $F_{ST}$ . After the first run, all loci outside the desired confidence interval (99%) are removed. Subsequently, a new 100,000 simulations run was developed to compute the mean neutral  $F_{ST}$ . A final run of LOSITAN using all loci was then conducted using the last computed mean. Also, a frequentist method based on moment-based estimates of  $F_{ST}$ , using the FDIST option of the LOSITAN package, was tested to compare results with the Bayesian approach.

### **Phylogenetic Relationship between species from the Caprinae subfamily**

Haplotype sequences from the different species analyzed were inferred by using PLINK software (Purcell et al. 2007a). Promoter sequences were aligned by CLUSTAL. MEGA 6 software (Tamura et al. 2013) was used to estimate nucleotide substitution models, evolutionary divergence and Tajima's Neutrality Test and to construct the ML tree. The tree was constructed considering only haplotypes with frequencies higher than 0.05. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (5000 replicates) is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the BioNJ method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 5.6803)).

## **RESULTS**

### **Polymorphism variability and test for linkage disequilibrium in sheep breeds**

Genotype and allele frequencies of the 11 polymorphisms studied in each of the 31 sheep breeds are showed in Supplemental Table 3 a) and b). Levels of polymorphism were generally high in all breeds. There were no exclusive polymorphisms in any of the breeds studied. The less polymorphic marker was the SNP g.522A>G for which the G<sub>-522</sub> allele was fixed in 18 breeds. For the INDELS g.666\_667insC and g.516\_517insG, the D (deletion) allele was fixed in nine and six breeds, respectively. The Russian Karakul breed (KAR) showed five polymorphisms fixed, while Olkuska (OL) and Sazik (SZ) four.

It is outstanding that seven polymorphisms had the MAF for the same allele in all breeds (I<sub>-668</sub>, I<sub>-667</sub>, A<sub>-601</sub>, G<sub>-524</sub>, A<sub>-522</sub>, I<sub>-516</sub>, G<sub>-468</sub> and A<sub>-444</sub>). However, the MAF for g.703\_704insAA, g.660G>C and g.528G>A polymorphisms were the AA<sub>-704</sub>, C<sub>-660</sub>, and A<sub>-528</sub> alleles in five Asian (DGL, EDIL, KAR, KRB and KRC) and one European (KARM) breeds, while for the remaining breeds were the D<sub>-704</sub>, G<sub>-660</sub> and G<sub>-528</sub> alleles (Table 3).

The Hardy Weinberg equilibrium test (Supplemental Table 1) shows all polymorphisms in HW equilibrium except the INDELS g.666\_667insC and g.703\_704insAA for all populations joined. The HWE test for each breed separately, showed that g.666\_667insC and g.660G>C were in HW disequilibrium in six (ARMA, AS, BOZ, CAUC, RA and VdB) and four (BOZ, EDIL, KAR and PRAM) populations, respectively. DGL breed showed three linked SNPs in HW disequilibrium (g.601A>C, g.524G>T and g.468G>T). The average expected (Ehet) and observed (Ohet) heterozygosity were 0.273 and 0.258, respectively, for all breeds joined. EDIL, DGL, SZ, Ct, IV and CI breeds showed Ohet values lower than 20%, while LX, KRB, MNCH, Ch, AKA and KAR breeds had values greater than 30%.

LD was estimated to obtain polymorphism linked blocks across and within breeds (Supplemental Table 2). In most breeds, similar results than those previously observed in Chapter in the Manchega (MNCH) breed, were obtained. Thus, three LD blocks can be established: 1) g.666\_667insC-g.444A>G; 2) g.703\_704insAA-g.660G>C-g.528A>G; 3) g.601A>C-g.524G>T-g.468G>T.

**Table 3.** Genotype and allele frequencies of the II polymorphisms located at the *HSP90AA1* gene in the 31 sheep breeds studied. a) Genotype frequencies. b) Allelic frequencies.

a)

Breed	g.703_704insAA		g.667_668insC		g.666_667insC		g.660Q>C			g.601A>C			g.528Q>A			g.524Q>T			g.522A>G			g.516_517insG			g.468Q>T			g.444A>G				
	ID	Breed	DD	AD	AA	II	ID	DD	CG	CC	AA	AC	CC	GG	AG	AA	GG	GT	TT	AA	AG	GG	ID	DD	GG	GT	TT	AA	AG	GG		
AKA	AKA	0.1	0.5	0.4	0.0	0.5	0.5	0.2	0.7	0.1	0.5	0.4	0.0	0.4	0.6	0.0	0.6	0.4	0.0	0.4	0.6	0.0	0.0	1.0	0.0	0.4	0.6	0.0	0.3	0.7		
Kazakh Arkhar-Merino	ARME	0.1	0.5	0.4	0.0	0.5	0.5	0.2	0.8	0.1	0.5	0.4	0.0	0.2	0.8	0.1	0.4	0.5	0.0	0.2	0.8	0.0	0.0	0.4	0.6	0.0	0.2	0.8	0.0	0.3	0.7	
Assaf	AS	0.0	0.2	0.8	0.1	0.5	0.3	0.1	0.8	0.0	0.2	0.8	0.0	0.6	0.4	0.0	0.2	0.8	0.0	0.6	0.4	0.0	0.0	1.0	0.0	0.6	0.4	0.0	0.2	0.8		
Awassi	AW	0.0	0.3	0.7	0.0	0.5	0.5	0.0	1.0	0.0	0.3	0.7	0.1	0.6	0.3	0.0	0.3	0.7	0.1	0.6	0.3	0.0	0.0	1.0	0.1	0.6	0.3	0.0	0.0	1.0		
Bajdarak	BAJ	0.3	0.5	0.2	0.0	0.1	0.9	0.0	1.0	0.1	0.6	0.2	0.0	0.4	0.6	0.1	0.6	0.4	0.0	0.4	0.6	0.0	0.1	0.9	0.0	0.3	0.7	0.0	0.2	0.8		
Bni Guil	BNI	0.2	0.6	0.3	0.0	0.4	0.6	0.0	0.2	0.7	0.2	0.6	0.3	0.0	0.2	0.8	0.1	0.5	0.4	0.0	0.2	0.9	0.0	0.2	0.9	0.0	0.2	0.8	0.0	0.3	0.7	
Boujaad	BOUJ	0.2	0.3	0.6	0.0	0.5	0.5	0.0	1.0	0.2	0.3	0.6	0.0	0.3	0.8	0.0	0.3	0.6	0.0	0.0	0.1	0.9	0.0	0.2	0.8	0.0	0.3	0.8	0.0	0.2	0.8	
Bozakh	BOZ	0.3	0.3	0.4	0.0	0.4	0.6	0.1	0.1	0.8	0.3	0.3	0.4	0.0	0.2	0.8	0.2	0.5	0.4	0.0	0.2	0.8	0.0	0.1	0.9	0.0	0.2	0.8	0.0	0.2	0.8	
Caucasian	CALIC	0.2	0.3	0.5	0.0	0.3	0.6	0.1	0.0	0.9	0.1	0.4	0.5	0.0	0.2	0.8	0.1	0.4	0.5	0.0	0.1	0.9	0.0	0.1	0.3	0.6	0.0	0.2	0.8	0.0	0.2	0.8
Churra	Ch	0.0	0.5	0.5	0.1	0.3	0.6	0.0	0.3	0.7	0.0	0.5	0.0	0.3	0.7	0.0	0.5	0.0	0.3	0.7	0.0	0.2	0.8	0.0	0.3	0.7	0.0	0.3	0.7	0.0	0.4	0.6
Churra Lebrijana	CI	0.0	0.4	0.6	0.4	0.4	0.2	0.0	1.0	0.0	0.4	0.6	0.0	0.1	0.9	0.0	0.4	0.6	0.0	0.2	0.9	0.0	0.0	1.0	0.0	0.2	0.9	0.0	0.0	1.0		
Churra Tensina	Ct	0.2	0.4	0.4	0.0	0.2	0.8	0.0	1.0	0.2	0.4	0.4	0.0	0.0	1.0	0.2	0.4	0.4	0.0	0.0	1.0	0.0	0.1	0.5	0.5	0.0	0.0	1.0	0.0	0.0	1.0	
Daglic	DGL	0.4	0.4	0.2	0.0	0.1	0.9	0.0	1.0	0.4	0.4	0.2	0.1	0.1	0.8	0.4	0.4	0.2	0.1	0.1	0.8	0.0	0.0	1.0	0.1	0.1	0.8	0.0	0.0	1.0		
Kazakh Edilibai	EDIL	0.0	0.1	0.9	0.0	0.0	1.0	0.0	1.0	0.3	0.7	0.0	0.0	1.0	0.3	0.7	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
Ivesi	IV	0.2	0.3	0.5	0.1	0.1	0.8	0.1	0.1	0.9	0.2	0.3	0.5	0.1	0.3	0.5	0.2	0.2	0.6	0.1	0.3	0.5	0.0	0.0	1.0	0.1	0.3	0.5	0.1	0.1	0.9	
Russian Karakul	KAR	0.3	0.7	0.0	0.0	0.0	1.0	0.0	1.0	0.2	0.8	0.0	0.6	0.4	0.3	0.7	0.0	0.0	0.6	0.4	0.0	0.0	0.0	1.0	0.1	0.5	0.4	0.0	0.0	1.0		
Moldavian Karakul	KARM	0.7	0.3	0.0	0.1	0.0	0.9	0.0	1.0	0.3	0.7	0.0	0.0	0.4	0.6	0.3	0.7	0.1	0.0	0.4	0.6	0.0	0.0	1.0	0.1	0.9	0.0	0.4	0.6	0.0	0.1	0.9
Karabakh	KRB	0.3	0.7	0.1	0.0	0.4	0.6	0.0	1.0	0.3	0.7	0.1	0.0	0.2	0.8	0.1	0.7	0.2	0.0	0.2	0.8	0.0	0.0	1.0	0.0	0.3	0.8	0.0	0.3	0.8		
Karachai	KRC	0.4	0.5	0.1	0.0	0.3	0.8	0.0	1.0	0.3	0.6	0.1	0.0	0.3	0.7	0.3	0.6	0.1	0.0	0.3	0.7	0.0	0.0	1.0	0.0	0.3	0.7	0.0	0.2	0.8		
Karayaka	KRY	0.1	0.4	0.6	0.0	0.2	0.8	0.1	0.1	0.9	0.1	0.4	0.6	0.1	0.3	0.6	0.1	0.3	0.6	0.1	0.4	0.6	0.0	0.0	0.3	0.7	0.1	0.4	0.6	0.0	0.1	0.9
Kivircik	KYR	0.0	0.5	0.5	0.1	0.6	0.3	0.0	1.0	0.0	0.5	0.5	0.0	0.2	0.8	0.1	0.3	0.6	0.0	0.1	0.9	0.0	0.0	1.0	0.0	0.2	0.8	0.0	0.2	0.8		
Finnsheep	L	0.2	0.5	0.3	0.1	0.4	0.5	0.1	0.4	0.5	0.2	0.5	0.3	0.0	0.0	0.9	0.1	0.5	0.4	0.0	0.0	0.9	0.0	1.0	0.0	0.0	0.9	0.1	0.4	0.5		
Latxa	LX	0.0	0.5	0.5	0.0	0.5	0.5	0.0	0.4	0.6	0.0	0.2	0.8	0.0	0.3	0.7	0.0	0.2	0.8	0.0	0.3	0.7	0.0	0.0	0.4	0.6	0.0	0.3	0.7	0.0	0.5	0.5
D'Man	MAN	0.1	0.5	0.5	0.0	0.1	0.9	0.0	1.0	0.0	0.5	0.5	0.1	0.4	0.5	0.0	1.0	0.1	0.3	0.6	0.0	0.0	1.0	0.1	0.3	0.7	0.1	0.4	0.5	0.1	0.4	0.5
Spanish Merino	ME	0.2	0.3	0.5	0.1	0.5	0.5	0.0	1.0	0.1	0.4	0.5	0.0	0.1	0.9	0.0	0.2	0.8	0.0	0.1	0.9	0.0	0.2	0.8	0.0	0.1	0.9	0.0	0.4	0.6		
Manchega	MNCH	0.1	0.5	0.4	0.0	0.4	0.6	0.0	1.0	0.1	0.5	0.4	0.0	0.3	0.7	0.0	0.5	0.4	0.0	0.3	0.7	0.0	0.0	1.0	0.0	0.3	0.7	0.0	0.1	0.9		
Olkuska	OL	0.1	0.5	0.4	0.0	0.5	0.5	0.1	0.4	0.6	0.1	0.5	0.4	0.0	0.0	1.0	0.1	0.5	0.4	0.0	0.0	1.0	0.0	0.2	0.8	0.0	0.0	1.0	0.1	0.4	0.5	
Pramenka	PRAM	0.2	0.2	0.6	0.1	0.4	0.5	0.1	0.3	0.6	0.2	0.2	0.6	0.0	0.3	0.7	0.2	0.2	0.6	0.0	0.3	0.7	0.0	0.0	1.0	0.1	0.9	0.0	0.3	0.7	0.0	0.5
Rasa Aragonesa	RA	0.1	0.3	0.6	0.1	0.4	0.5	0.1	0.1	0.9	0.1	0.3	0.6	0.0	0.2	0.8	0.1	0.2	0.7	0.0	0.2	0.8	0.0	0.1	0.3	0.6	0.0	0.2	0.8	0.1	0.2	0.8
Sakiz	SZ	0.0	0.4	0.6	0.2	0.4	0.4	0.0	1.0	0.0	0.4	0.6	0.0	0.1	0.9	0.0	0.3	0.7	0.0	0.1	0.9	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.0	1.0		
Valle del Belice	VdB	0.1	0.6	0.3	0.0	0.3	0.7	0.1	0.0	0.9	0.1	0.6	0.3	0.0	0.0	1.0	0.1	0.5	0.5	0.0	0.0	1.0	0.0	0.3	0.7	0.0	0.1	0.9	0.0	0.2	0.8	

b)

Breed	ID Breed	N	g:703_704insAA		g:667_668insC		g:666_667insC		g:660G>C		g:601A>C		g:528G>A		g:524G>T		g:522>G		g:516_517insG		g:468G>T		g:444A>G	
			D	AA	I	D	I	D	G	C	A	C	G	A	G	T	A	G	I	D	G	T	A	G
Akkaraman	AKA	46	0.39	0.61	0.24	0.76	0.24	0.85	0.39	0.61	0.24	0.76	0.33	0.67	0.2	0.8	0.04	0.96	0.02	0.98	0.2	0.8	0.13	0.87
Kazakh Arkhar-Merino	ARHE	36	0.36	0.64	0.25	0.75	0.08	0.83	0.31	0.69	0.08	0.92	0.28	0.72	0.08	0.92	0	1	0.19	0.81	0.08	0.92	0.14	0.86
Assaf	AS	60	0.1	0.9	0.4	0.6	0.15	0.85	0.1	0.9	0.28	0.72	0.1	0.9	0.28	0.72	0	1	0	1	0.28	0.72	0.12	0.88
Avsssi	AW	60	0.15	0.85	0.27	0.73	0	1	0.13	0.87	0.42	0.58	0.13	0.87	0.42	0.58	0	1	0	1	0.42	0.58	0.02	0.98
Bajarak	BAJ	44	0.52	0.48	0.07	0.93	0	1	0.45	0.55	0.18	0.82	0.34	0.66	0.18	0.82	0.07	0.93	0.02	0.98	0.16	0.84	0.09	0.91
Bhi Gull	BNI	54	0.43	0.57	0.19	0.81	0.24	0.85	0.43	0.57	0.11	0.89	0.35	0.65	0.07	0.93	0.06	0.94	0.07	0.93	0.13	0.87	0.17	0.83
Boujad	BOUJ	48	0.46	0.54	0.23	0.77	0.13	0.94	0.46	0.54	0.13	0.88	0.25	0.75	0.13	0.88	0.06	0.94	0.1	0.9	0.13	0.88	0.08	0.92
Bozakh	BOZ	48	0.48	0.52	0.23	0.77	0.17	0.85	0.48	0.52	0.1	0.9	0.4	0.6	0.1	0.9	0	1	0.04	0.96	0.1	0.9	0.13	0.88
Caucasian	CAUC	50	0.36	0.64	0.2	0.8	0.06	0.88	0.32	0.68	0.1	0.9	0.3	0.7	0.06	0.94	0	1	0.28	0.72	0.1	0.9	0.12	0.88
Churra	Ch	46	0.26	0.74	0.24	0.76	0.28	0.83	0.26	0.74	0.13	0.87	0.26	0.74	0.13	0.87	0.11	0.89	0.15	0.85	0.13	0.87	0.22	0.78
Churra Lebrjuna	Cl	52	0.21	0.79	0.6	0.4	0	1	0.19	0.81	0.06	0.94	0.19	0.81	0.08	0.92	0	1	0.02	0.98	0.08	0.92	0.02	0.98
Churra Tensina	Ct	66	0.42	0.58	0.09	0.91	0	1	0.36	0.64	0.03	0.97	0.36	0.64	0.02	0.98	0	1	0.32	0.68	0.02	0.98	0.02	0.98
Daglic	DGL	48	0.63	0.38	0.04	0.96	0	1	0.63	0.38	0.19	0.81	0.63	0.38	0.19	0.81	0	1	0.02	0.98	0.19	0.81	0	1
Kazakh Edibai	EDIL	60	0.95	0.05	0.02	0.98	0	1	0.67	0.33	0.02	0.98	0.62	0.38	0.02	0.98	0.02	0.98	0.05	0.95	0.02	0.98	0	1
Ivesi	IV	30	0.47	0.53	0.13	0.87	0.1	0.9	0.47	0.53	0.3	0.7	0.3	0.7	0.3	0.7	0.03	0.97	0	1	0.3	0.7	0.1	0.9
Russian Karakul	KAR	30	0.63	0.37	0	1	0	1	0.6	0.4	0.3	0.7	0.63	0.37	0.3	0.7	0	1	0	1	0.33	0.67	0	1
Moldavian Karakul	KARM	30	0.87	0.13	0.07	0.93	0	1	0.63	0.37	0.2	0.8	0.6	0.4	0.2	0.8	0	1	0.03	0.97	0.2	0.8	0.03	0.67
Karabakh	KRB	48	0.58	0.42	0.21	0.79	0.08	0.96	0.58	0.42	0.08	0.92	0.48	0.52	0.08	0.92	0.04	0.96	0	1	0.13	0.88	0.13	0.88
Karadai	KRC	56	0.63	0.38	0.13	0.88	0.09	0.93	0.55	0.45	0.14	0.86	0.55	0.45	0.14	0.86	0.02	0.98	0.02	0.98	0.14	0.86	0.09	0.91
Karayaka	KRY	44	0.27	0.73	0.09	0.91	0.07	0.93	0.27	0.73	0.25	0.75	0.23	0.77	0.23	0.77	0.02	0.98	0.14	0.86	0.27	0.73	0.07	0.93
Kivircik	KVR	32	0.25	0.75	0.44	0.56	0.13	0.94	0.25	0.75	0.09	0.91	0.22	0.78	0.06	0.94	0	1	0.03	0.97	0.09	0.91	0.09	0.91
Finnshp	L	60	0.42	0.58	0.28	0.72	0.42	0.72	0.42	0.58	0.05	0.95	0.33	0.67	0.05	0.95	0	1	0.02	0.98	0.05	0.95	0.28	0.72
Labra	LX	82	0.28	0.72	0.24	0.76	0.37	0.82	0.11	0.89	0.15	0.85	0.1	0.9	0.15	0.85	0	1	0.21	0.79	0.15	0.85	0.23	0.77
D'Man	MAN	52	0.31	0.69	0.04	0.96	0.08	0.96	0.29	0.71	0.27	0.73	0.02	0.98	0.23	0.77	0	1	0.21	0.79	0.27	0.73	0.31	0.69
Spanish Merino	ME	58	0.33	0.67	0.31	0.69	0.12	0.91	0.28	0.72	0.07	0.93	0.14	0.86	0.07	0.93	0.05	0.95	0.09	0.91	0.07	0.93	0.24	0.76
Manchega	MNCH	120	0.32	0.68	0.23	0.78	0.07	0.97	0.32	0.68	0.17	0.83	0.3	0.7	0.16	0.84	0.02	0.98	0.18	0.82	0.16	0.84	0.05	0.95
Olkuska	OL	60	0.33	0.67	0.28	0.72	0.4	0.75	0.33	0.67	0	1	0.33	0.67	0	1	0	1	0.08	0.92	0	1	0.28	0.72
Pramenka	PRAM	58	0.33	0.67	0.34	0.66	0.4	0.72	0.33	0.67	0.16	0.84	0.31	0.69	0.16	0.84	0	1	0.05	0.95	0.16	0.84	0.28	0.72
Rasa Aragonesa	RA	84	0.21	0.79	0.27	0.73	0.11	0.89	0.2	0.8	0.11	0.89	0.17	0.83	0.11	0.89	0.05	0.95	0.2	0.8	0.11	0.89	0.14	0.86
Sakiz	SZ	52	0.23	0.77	0.38	0.62	0	1	0.23	0.77	0.06	0.94	0.15	0.85	0.06	0.94	0	1	0	1	0.04	0.96	0	1
Valle del Belice	YDB	58	0.41	0.59	0.19	0.81	0.09	0.88	0.38	0.62	0	1	0.29	0.71	0.02	0.98	0	1	0.19	0.81	0.02	0.98	0.14	0.86

### Phylogenetic relationships between sheep breeds

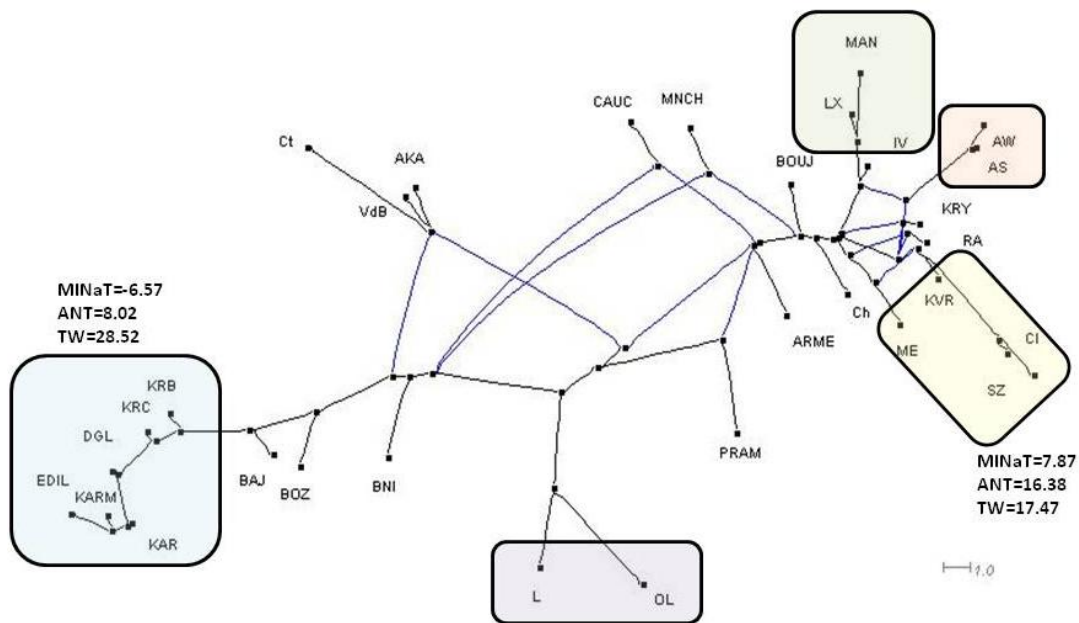
Population pairwise  $F_{ST}$ , p-values and significances and the Reynolds's distance matrix among the 31 sheep breeds studied were obtained. Average, median, maximum and minimum distances across populations were 0.0952, 0.0628, 0.6159 and 0.0000, respectively. Quartile distributions of Reynold's distances were 0.0158, 0.0579, 0.1249 and 0.6159 for the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> quartiles, respectively. Among AW, SZ, AS, CI, LX, KAR, DGL, KARM and EDIL breeds distances higher than 0.25 were observed. Distance values lower than 0.01 (even 0.00) were found among ARME, Ch, KRY, MNCH, AKA, CAUC, BNI, BOUJ and BOZ breeds.

Figure 1 shows NeighborNet graph based on Reynold's distance constructed with the ClusterNetwork splits transformation method for the 31 sheep breeds studied. The graph contains 31 taxa and 35 splits and consists of 68 vertices and 81 edges. The Fit (which is the sum of all pairwise distances in the graph divided by the sum of all pairwise distances in the given matrix, 100 times) and the LSFit (which is the least squares fit between the pairwise distances in the graph and the pairwise distances in the matrix) of the NeighborNet were 69.40 and 92.26, respectively.

The group constituted by EDIL, KARM, KAR, DGL, KRC, KRB, BAJ and BOZ breeds is outside the reticulations of the NeighborNet graph, indicating a certain degree of separation of this set from the remaining breeds. All these breeds have in common that belong to regions of West Asia and East Europe with high thermal width (arid and semiarid climates). Average, minimum and maximum distances among these breeds were 0.040, 0.000 and 0.167, respectively. The remaining breeds are included in a complex system of reticulations which indicates the existence of a genetic admixture among them (Kijas et al. 2012). Breeds from all continents are included in this complex reticulated network. AS and AW breeds are joined in the same branch, as it should be expected due to their high genetic linkage (Assaf is a synthetic breed from a cross between Awasi and Milkchaff milk breeds). From the same node come up LX, MAN and IV breeds, all of them pertaining to very different locations. KVR, SZ, ME and CI breeds come from the same node. All these breeds belong to Mediterranean regions with low thermal width and semi-dry climates. Average, minimum and maximum distances among these breeds were 0.040, 0.000 and 0.080, respectively. Also L (Finland) and OL (Poland) breeds come up from the same node of the graph constituting a certain isolated group. The central area of the NeighborNet is constituted by breeds belonging to different continents, locations and climatic areas.

Figure 2 shows the histogram of the number of significant different populations ( $p < 0.05$ ) for each of the sheep breeds studied using the Exact Test of population differentiation. The number of significant different populations ranged from 12 to 30 and the average was 22.2.

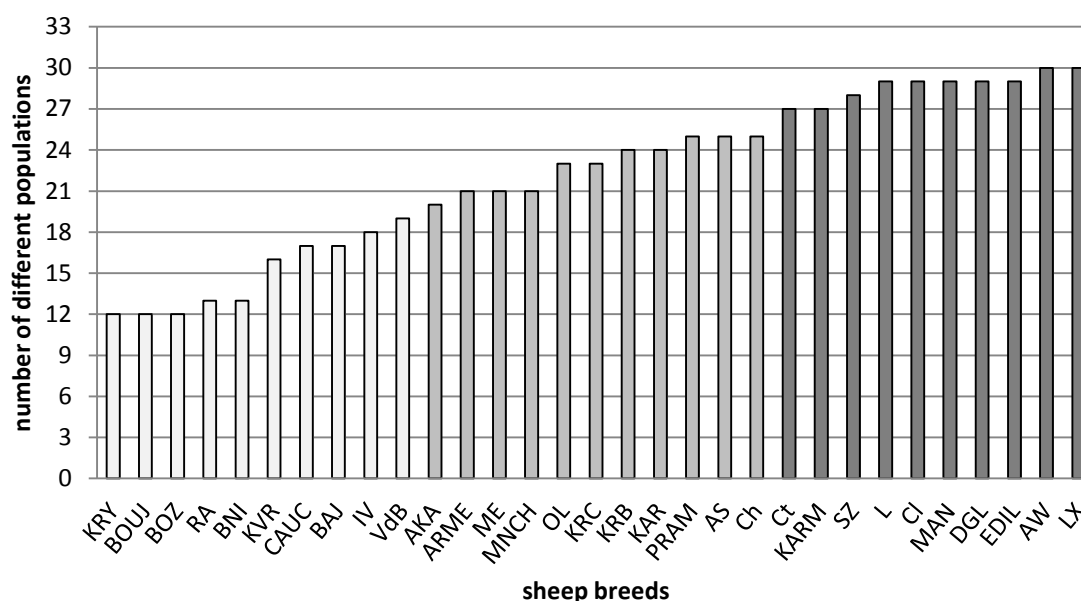
**Figure 1.** NeighborNet graph based on Reynold's distance constructed with the ClusterNetwork splits transformation method for the 31 sheep breeds studied.



### Tests to detect association of loci frequencies with environmental parameters

**PLSR.** PLSR analysis was conducted including the MAF of six polymorphisms as response variables and 14 environmental variables as predictors (CTY was not included in analyses due to its discrete nature). Polymorphisms considered were g.667\_668insC, g.522A>G, g.516\_517insG and one polymorphism of each LD block common to most breeds: g.666\_667insC, g.660G>C and g.601A>C. For all polymorphisms analyzed, the allele at lower frequency was the same in all breeds ( $I_{-668}$ ,  $I_{-667}$ ,  $A_{-601}$ ,  $A_{-522}$  and  $I_{-516}$ ). However, the  $G_{-660}$  allele of the g.660G>C SNP was the MAF in 25 from the 31 breeds studied.

**Figure 2.** Histogram of the number of significant different populations ( $p < 0.05$ ).



Dark grey = breeds differing from 27 to 30 populations; Medium grey = breeds differing from 20 to 25 populations; Light grey = breeds differing from 12 to 19 populations.

Basic statistics and Pearson and Spearman correlations among MAF and environmental variables were carried out. High negative Pearson (-0.68) and Spearman (-0.70) correlation coefficients were found between MAF of g.667\_668insC and g.660G>C ( $p < 0.0001$ ). Regarding correlations among environmental predictors high ( $\geq 0.70$ ) negative correlations (Pearson and Spearman) were found between LAT-MINaT, LAT-ANT, LON-MINaT, MINaT-TW and ANT-TW; and positive between LON-TW, MINaT-ANT and TAR-MxR. Only significant correlations among MAF and environmental predictors were found for g.667\_668insC, g.666\_667insC, g.660G>C and g.515\_516insG. Similar magnitude but opposite sign had the correlations found between g.667\_668insC and g.660G>C polymorphisms and the weather parameters MINaT, ANT, TW, TAR and MxR.

Table 4 shows VIP and percentage of variance explained by the top two PLSR (VT2) components for each environmental variable. Those variables showing VIP values greater than 0.83 and which VT2 was at least 40%, were retained for posterior analyses. With these criteria, MAXaT, HrMx, HrMi and THI variables were discarded.

A second PLSR analysis including six polymorphisms and ten environmental variables were developed. PRESS values of the complete (14 predictor variables) and reduced model (10 predictor variables) were 0.9726 and 0.9589, respectively, which indicates that the elimination of 4 useless environmental variables improve the prediction model. Reducing the number of predictors,  $R^2$  was also improved from 0.82 (14 predictors) to 0.85 (10 predictors).

Three components were retained using the optimal model determination by the leave-one-out cross validation procedure and the minimum PRESS criteria (van der Voet's test). The 72.47% of the predictor variation is already explained by just two, but only 24.50% of the response variation is achieved. Figure 3 shows Variable Importance in Projection values (VIP) and percentage of variance explained by the top two PLSR components (VT2) for each of the ten environmental predictors included in the model. Taking into account for both statistics, MINaT, ANT, TW, TAR and MiR were the predictors with the best combination of VIP and VT2 values. Their VIP values ranged from 0.90 to 1.10 and all of them accounts for more than 75% of the variance explained. However, environmental variables, as LON and MxR despite having high VIP values showed percentages of the variance explained below 50%.

$Q^2$  values obtained were calculated for the MAF of the six polymorphisms included in the PLSR model.  $Q^2$  values were 0.46, 0.47, 0.53, 0.20, 0.10 and 0.33 for I<sub>668</sub>, I<sub>667</sub>, G<sub>660</sub>, A<sub>601</sub>, A<sub>522</sub> and I<sub>516</sub>, respectively. Only for I<sub>668</sub>, I<sub>667</sub> and G<sub>660</sub>  $Q^2$  values exceed the acceptable threshold (0.4).



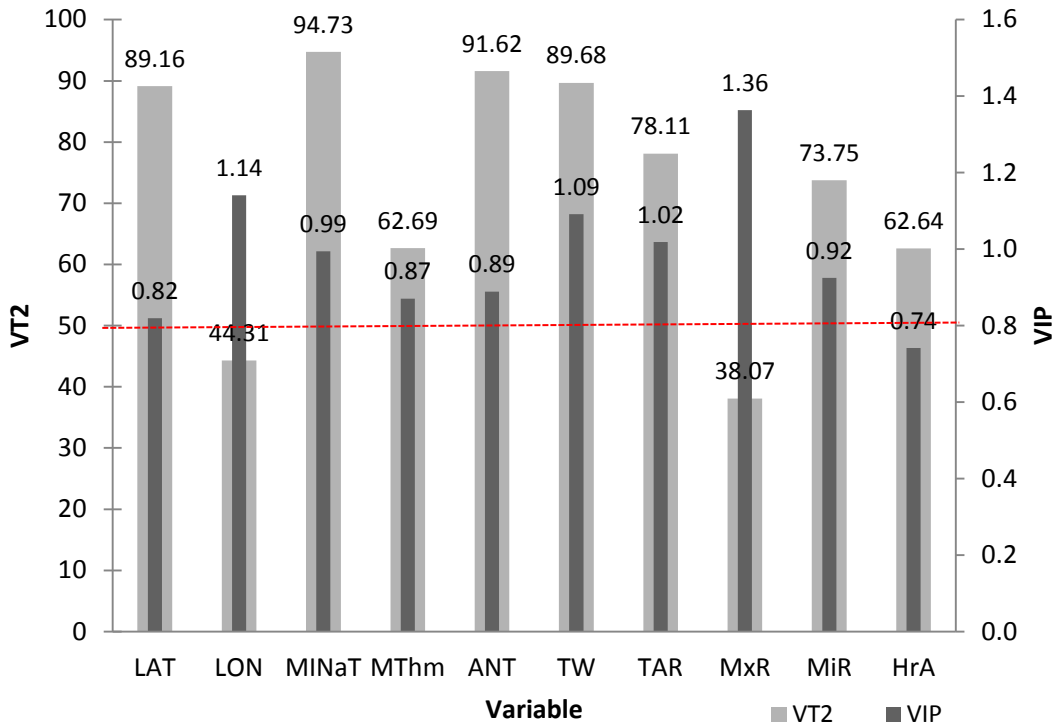
**Table 4.** VIP values and cumulative variance (VT2) explained by the top two factors. Variables discarded for posterior analysis are indicated in bold.

Variable	VIP	VT2
LAT	0.8412	87.389
LON	1.3123	40.001
MINaT	1.0642	94.862
<b>MAXaT</b>	<b>0.7268</b>	<b>84.903</b>
MThm	0.9071	54.214
ANT	0.9467	94.576
TW	1.1786	87.605
TAR	1.0767	76.934
MxR	1.5335	42.358
MiR	1.0205	66.793
HrA	0.8146	72.517
<b>HrMx</b>	<b>0.6934</b>	<b>71.635</b>
<b>HrMi</b>	<b>0.7508</b>	<b>67.205</b>
<b>THI</b>	<b>0.7357</b>	<b>82.224</b>

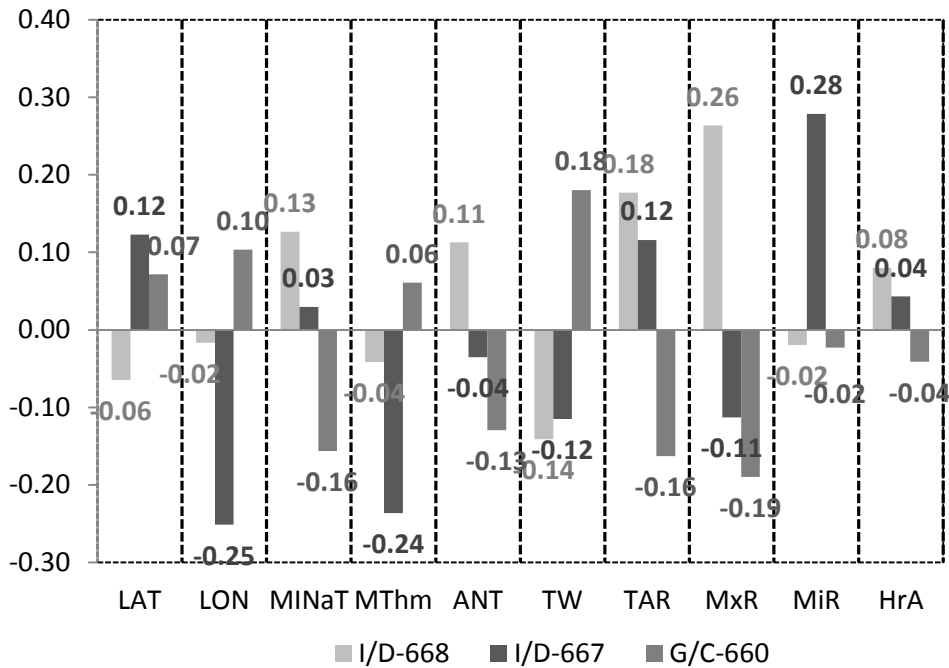
LAT=latitude; LON=longitude; MAXaT=maximum average temperature; MThm=maximum temperature of the hottest month; MINaT=minimum average temperature; ANT=average annual temperature; TW (MAXaT-MINaT)=thermal width; TAR = total annual rainfall; MxR = maximum rainfall; MiR = minimum rainfall; HrA = relative average annual humidity (%);HrMx = maximum relative humidity (%);HrMi = minimum relative humidity (%);THI=Temperature Humidity Index (Marai et al. 2007).

Regression coefficients for responses with  $Q^2$  values higher than 0.4, are shown in Figure 4. Absolute values of regression coefficients ranged from 0.02 to 0.28. Interestingly, regression coefficients of  $I_{-668}$  and  $G_{-660}$  have opposite sign for all environmental predictors, except for MiR, indicating that the MAF at these polymorphisms depends on opposite environmental and geographical circumstances. Thus, the frequency of the  $I_{-668}$  and  $G_{-660}$  alleles increases and decreases respectively, for higher values of MINaT, ANT, TAR and HrA. Otherwise, high values of LAT, LON and TW are linked to low and high frequencies of the  $I_{-668}$  and  $G_{-660}$  alleles, respectively.

**Figure 3.** Variable Importance in Projection values (VIP) and percentage of variance explained by the top two PLSR components (VT2) for each of the ten environmental predictors included in the model.



**Figure 4.** Regression coefficients for responses (polymorphisms frequencies) with  $Q^2$  values higher than 0.4



SAM. MATSAM was run for six polymorphisms (12 alleles), g.667\_668insC, g.522A>G, g.516\_517insG and one polymorphism of each LD block common to most

breeds, g.666\_667insC, g.660G>C and g.601A>C, and 14 geographic and climatic variables (LAT, LON, MINaT, MAXaT, MThm, ANT, TW, TAR, MxR, MiR, HrA, HrMx, HrMi and THI). Seven alleles at 5 loci were detected as significantly associated with at least one environmental variable with a confidence level of 99.99% (significant threshold  $ST=5.952 \cdot 10^{-5}$ ) based on cumulated results from *W* and *G* tests (Table 5). These alleles are involved in 168 significant models according to these criteria.

No association with environmental variables was found for g.522A>G alleles. MAF alleles of the remaining polymorphisms were associated with more environmental variables than those with maximum frequency, which in some cases (D<sub>-667</sub>, C<sub>-601</sub> and D<sub>-516</sub>) did not show any association. MAF alleles I<sub>-668</sub>, I<sub>-667</sub> and G<sub>-660</sub> were associated with the highest number of environmental variables, 5, 8 and 8 respectively. The environmental variables related with more number of loci were MxR, ANT, LON, MINaT, TAR and TW (4, 3, 3, 3, 3 and 3, respectively).

Supplemental Figure 2 shows correlograms of significant associations between markers and environmental variables, which differences in probability of presence of the allele between the extremes of the distribution were higher than 37%. MINaT is the environmental variable for which greatest changes was shown in the probability to find the G<sub>-660</sub> allele. A decrease in this probability from 0.9 to near 0.3 (60%) was found for G<sub>-660</sub> when MINaT increases from -22°C to 17°C. An opposite trend was observed for I<sub>-668</sub>. In this case, the likeliness to find de I<sub>-668</sub> allele increases from 0.15 to near 0.60 (42.8%) for the same rank of MINaT change. For ANT and MxR the same pattern above described was found. Thus the probability to find the G<sub>-660</sub> allele decreases from 0.85 to 0.30 (54%) and the probability to find the I<sub>-668</sub> allele increases from 0.2 to 0.5 (37.2%) when ANT increases from -2°C to 22.7°C. For MxR the probability of the G<sub>-660</sub> allele decreases from 0.70 to 0.35 (37.6%) and the probability of the I<sub>-668</sub> allele increases from 0.2 to 0.7 (45.7%) when MxR increases from -5.6 to 166.4. Finally, for TW an opposite pattern from the above described was found. So the probability of the G<sub>-660</sub> allele increases from 0.3 to 0.9 (58%) for 28 units of increment in TW (11 to 38.8), and the probability of the I<sub>-668</sub> allele decreases from 0.6 to 0.1 (44.1%) for the same rank of TW variation.

For LON, LAT and HrMx the probability to find the G<sub>-660</sub> allele increases a 45.9%, 44.5% and 38.8%, respectively, when these variables change from -18.4 to 118.8, 27.2 to 65.5 and 61.3 to 93.7, respectively.

**Table 5** – SAM cumulated test for molecular sheep data and environmental variables with a significant threshold level of 5.952-E05 (including Bonferroni correction). Grey cells with ‘I’ indicate that for this model, the null hypothesis is rejected with both the W and G test.

Marker Freq.	0.400	0.949	0.151	0.962	0.565	0.878	0.256	0.980	0.037	1.000	0.187	0.984
Marker	I <sub>-668</sub>	D <sub>-668</sub>	I <sub>-667</sub>	D <sub>-667</sub>	G <sub>-660</sub>	C <sub>-660</sub>	A <sub>-601</sub>	C <sub>-601</sub>	A <sub>-522</sub>	G <sub>-522</sub>	I <sub>-516</sub>	D <sub>-516</sub>
LAT	0	0	0	0	I	0	I	0	0	0	0	0
LON	0	0	0	0	I	I	0	0	0	0	I	0
MINaT	I	0	0	0	I	I	0	0	0	0	0	0
MAXaT	0	0	I	0	0	0	I	0	0	0	0	0
MThm	0	0	I	0	0	0	I	0	0	0	0	0
ANT	I	0	0	0	I	I	0	0	0	0	0	0
TW	I	0	0	0	I	I	0	0	0	0	0	0
TAR	I	0	I	0	I	0	0	0	0	0	0	0
MxR	I	I	0	0	I	0	0	0	0	0	I	0
MiR	0	0	I	0	0	0	0	0	0	0	0	0
HrA	0	0	I	0	0	0	0	0	0	0	0	0
HrMx	0	0	I	0	I	0	0	0	0	0	0	0
HrMi	0	0	I	0	0	0	0	0	0	0	0	0
THI	0	0	I	0	0	0	I	0	0	0	0	0

LAT=latitude; LON=longitude; MAXaT=maximum average temperature; MThm=maximum temperature of the hottest month; MINaT=minimum average temperature; ANT=average annual temperature; TW (MAXaT-MINaT)=thermal width; TAR = total annual rainfall; MxR = maximum rainfall; MiR = minimum rainfall; HrA = relative average annual humidity (%);HrMx = maximum relative humidity (%);HrMi = minimum relative humidity (%);THI=Temperature Humidity Index (Marai et al. 2007).

### Test to detect loci under selection

Bayesian Test of Beaumont and Balding. Table 6 shows expected heterozygosities ( $H_e$ ) and  $F_{ST}$  values obtained after 100,000 simulation runs of the LOSITAN and FDIST software using 31 sheep breeds and 6 unlinked polymorphisms at the *HSP90AA1* promoter. Estimated neutral  $F_{ST}$  was 0.072512. Two outlier loci were identified after 100,000 simulations and a 95% confidence interval: g.522A>G with a significant ( $p=0.02$ ) low  $F_{ST}$  value (0.02) candidate for balancing selection processes, and the g.703\_704insAA with a significant ( $p=0.99$ ) high  $F_{ST}$  value (0.14) candidate for directional selection. Expected heterozygosities ( $H_e$ ) ranged between 0.04 (g.522G>A) and 0.48 (g.703\_704insAA).

For the LD block constituted by polymorphisms g.703\_704insAA, g.660G>C and g.528A>G, only when the first of them was used in the analysis, a sign of directional selection was detected. With the frequentist approach of FDIST, only sign of balancing selection was found for the g.522A>G SNP.

**Table 6** – Expected heterozygosities ( $H_e$ ) and  $F_{ST}$  values obtained after 100,000 simulation runs of the LOSITAN and FDIST software for six unlinked polymorphisms at the *HSP90AA1* gene promoter (\* indicates significant values).

Locus	LOSITAN			FDIST		
	Het	Fst	P(Simul Fst<sample Fst)	Het	Fst	P(Simul Fst<sample Fst)
g.703_704ins	0.48	0.14	0.9957* directional	0.48	0.14	0.938
g.667_668ins	0.34	0.09	0.7658	0.34	0.09	0.475
g.666_667ins	0.17	0.07	0.5154	0.17	0.07	0.2534
g.601A>C	0.25	0.06	0.3443	0.25	0.06	0.1145
g.522A>G	0.04	0.02	0.0203* balancing	0.04	0.02	0.0174* balancing
g.516_517ins	0.16	0.09	0.7049	0.16	0.09	0.4718

### Characterization of the *HSP90AA1* promoter in species of the Caprinae and Bovinae subfamilies

Aligned sequences of a 410 pb amplicon from the *HSP90AA1* gene promoter of a total of 11 species belonging to the Caprinae and Bovinae subfamilies are shown in Supplemental Figure 1. Table 7 shows haplotype frequencies in each species studied. In *O. aries*, 36 different haplotypes were found. From them only the first four (H1 to H4) had a frequency higher than 10%. In *O. musimon* 7 haplotypes were found, all of them shared with *O. aries*. Since we have only one animal from the *O. vignei* and *O. ammon* species, only 1 and 2 haplotypes were detected in these taxa. *O. canadiensis* had no polymorphisms in the sequence analyzed and therefore a unique haplotype for this species was assessed. In *C. hircus* 24 different haplotypes were found but only one in *C. pyrenaica*. In *O. moschatus* 8 polymorphic positions were found being one of them species exclusive. Eight different haplotypes were found in this species. Finally, in *A. lervia* 3 different haplotypes were found having this species three exclusive polymorphisms.

The Tamura 3-parameter model (T92) with evolutionary rates among sites modeled by using a discrete Gamma distribution (+G) with 5 rate categories had the highest fit (lowest BIC value) among the 24 different nucleotide substitution models tested by maximum likelihood. This model was fitted to estimate Evolutionary Divergence between species sequences, to conduct the Tajima's Neutrality Test and to construct the ML tree.

Table 8 shows estimates of evolutionary divergence over sequence pairs between species. Within the Caprinae subfamily, the *R. pyrenaica* showed the highest percentage of sequence divergence with the remaining species (4.4% with species from the *Ovis* and *Ovibos* genus, 7% with species from the *Capra* genus and 7.6% with *A. lervia*. Interestingly *O. moschatus* was much closer to species of the *Ovis* genus (0.8 to 1%) than to those of *Capra*, *Ammotragus* and *Rupicapra*. As expected, very low evolutionary divergences among species of the genus *ovis* and among the species of the genus *Capra* were observed.

Table 9 shows polymorphisms detected and their frequencies in the species studied. Within the Caprinae subfamily, the species with more number of polymorphisms, SNPs or INDELS, were *C. hircus* and *O. aries*, with 14 and 11 polymorphic sites, respectively, from which only 6 were shared between them. Also, *O. moschatus* and *O. musimon* showed high number of polymorphisms in the *HSP90AA1* promoter, 8 and 7, respectively. Those species which shared more number of polymorphisms with *O. aries* (Reference species in our work) were *O. musimon* (7), *O. moschatus* (7) and *C. hircus* (6). In general, within this subfamily, polymorphisms shared among the different species had the same pattern of allele frequency, except for g.528A>G in *O. moschatus* where the A<sub>.528</sub> allele showed the highest frequency (0.93) and for g.703\_704insAA in *R. pyrenaica* where the D<sub>.704</sub> allele was the most frequent (0.75). Exclusive polymorphisms were found in *C. hircus* (8), *A. lervia* (3), *O. aries* (2) and *O. moschatus* (1). The two outgroup species from the Bovis genus (*B. mutus* and *B. taurus*) showed very few polymorphisms and did not share any mutations with the remaining species.

**Table 7** – Haplotypes inferred by PLINK for sheep breeds and wild species studied.

<i>O. aries</i>	Haplotype																																							
	ZN	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25	H26	H27	H28	H29	H30	H31	H32	H33	H34	H35	H36			
AKA	46	15	6	4	9	1	6	3	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
ARME	36	10	9	1	3	7	3	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AS	60	6	21	8	17	-	6	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AW	60	8	20	4	25	1	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BAJ	44	15	3	3	8	-	4	3	3	3	3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	
BNI	54	19	4	6	4	4	7	2	-	3	1	1	-	2	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BOUJ	48	12	9	9	6	5	2	-	-	3	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BOZ	48	19	9	4	5	2	4	3	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CAUC	50	14	7	3	3	12	5	1	-	-	-	1	2	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ch	46	12	-	6	6	6	8	-	-	5	-	-	-	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CI	52	9	30	6	3	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Ct	66	24	6	9	1	20	-	-	4	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DGL	48	30	2	6	9	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EDIL	58	33	1	2	-	-	4	12	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	
IV	30	9	1	6	9	-	3	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KAR	30	17	-	1	7	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KARM	30	17	-	2	-	-	1	4	-	-	-	-	-	-	1	-	-	3	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KRB	48	22	8	3	4	-	1	2	-	1	3	1	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KRC	54	29	5	4	8	1	2	-	1	1	1	-	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KRY	46	10	4	10	10	5	2	1	-	1	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KVR	32	6	12	4	2	1	2	2	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L	60	20	-	14	3	1	17	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LX	82	7	5	8	11	19	13	-	11	-	1	5	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MAN	52	1	-	4	11	11	2	6	-	-	8	6	-	2	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ME	58	6	15	9	4	4	4	1	1	3	8	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MINCH	194	55	39	16	36	31	6	1	-	8	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OL	60	20	3	13	-	5	15	1	1	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PRAM	58	18	6	5	9	3	14	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
RA	84	14	17	10	9	15	6	-	1	4	3	2	-	-	1	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SZ	52	8	20	17	3	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VdB	58	17	7	8	-	10	5	4	2	-	1	2	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-





Haplotypes									
<i>O. moschatus</i>	2N	Hm1	Hm2	Hm3	Hm4	Hm5	Hm6	Hm7	Hm8
OMO	30	11	5	1	8	1	1	2	1

Haplotypes				
<i>A. lervia</i>	2N	HI1	HI2	HI3
ALE	28	22	5	1

Haplotypes			
<i>B. mutus</i>	2N	Hm1	Hm2
YAK	26	24	2

Haplotypes			
<i>B. taurus</i>	2N	Ht1	Ht2
HOL	24	20	4
AVI	20	15	5
SET	10	10	-
PAR	14	13	1
PIR	18	15	3

**Table 8** – Estimates of evolutionary divergence between species (below diagonal) and their standard errors (above diagonal). Analyses were conducted using the Tamura 3-parameter model (T92+G).

	<i>A. lervia</i>	<i>C. hircus</i>	<i>C. pyrenaica</i>	<i>R. rupicapra</i>	<i>O. moschatus</i>	<i>O. aries</i>	<i>O. canadiensis</i>	<i>O. vignei</i>	<i>O. ammon</i>	<i>O. musimon</i>	<i>B. taurus</i>	<i>B. mutus</i>
<i>A. lervia</i>		0.006	0.007	0.014	0.009	0.008	0.008	0.009	0.008	0.008	0.015	0.015
<i>C. hircus</i>	0.021		0.003	0.013	0.008	0.007	0.007	0.007	0.007	0.007	0.014	0.014
<i>C. pyrenaica</i>	0.022	0.008		0.013	0.008	0.007	0.008	0.008	0.008	0.007	0.014	0.014
<i>R. rupicapra</i>	0.076	0.069	0.070		0.010	0.010	0.010	0.010	0.010	0.010	0.018	0.019
<i>O. moschatus</i>	<b>0.036</b>	<b>0.031</b>	<b>0.032</b>	<b>0.045</b>		0.004	0.004	0.004	0.004	0.004	0.014	0.014
<i>O. aries</i>	<b>0.031</b>	<b>0.026</b>	<b>0.027</b>	<b>0.046</b>	<b>0.010</b>		0.001	0.002	0.002	0.002	0.013	0.013
<i>O. canadiensis</i>	0.028	0.024	0.025	0.043	<b>0.008</b>	0.003		0.002	0.001	0.001	0.013	0.013
<i>O. vignei</i>	0.031	0.026	0.027	0.043	<b>0.009</b>	0.005	0.002		0.001	0.002	0.013	0.013
<i>O. ammon</i>	0.030	0.025	0.026	0.043	<b>0.008</b>	0.004	0.001	0.001		0.001	0.013	0.013
<i>O. musimon</i>	0.029	0.025	0.026	0.044	<b>0.008</b>	0.004	0.001	0.003	0.002		0.013	0.013
<i>B. taurus</i>	0.074	0.073	0.073	0.107	0.070	0.065	0.062	0.062	0.062	0.063		0.003
<i>B. mutus</i>	0.076	0.075	0.075	0.110	0.073	0.067	0.066	0.066	0.066	0.066	0.006	

**Table 9** – Polymorphisms and its frequencies in the wild species analyzed. In dark shadow are polymorphisms shared with the *Ovis aries* species. In light shadow are sequences of the *Bos* genus.

Species	g.703_704 insAA		g.667_668 insC		g.666_667 insC		g.660G>C		g.657A>G		g.653C>A		g.601A>C		g.571G>C		g.551A>G		g.529G>C		g.528A/T>G		g.525A>G		g.524G>T			
	-	AA	C	-	C	-	G	C	G	A	C	A	C	A	C	A	G	G	C	A/T	G	A	G	G	T	G	T	
<i>Ovis Aries</i>	0.42	0.58	0.28	0.72	0.1	0.9	0.37	0.63	0	1	0	1	0.16	0.84	0	1	0	1	0	1	0.32	0.68	0	1	0.15	0.85	0	1
<i>Ovis ammon</i>	0.5	0.5	0	1	0	1	0.5	0.5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ovis canadensis</i>	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ovis musimon</i>	0.53	0.47	0.47	0.53	0.05	0.95	0.47	0.53	0	1	0	1	0	1	0	1	0	1	0	1	0.19	0.81	0	1	0	1	0	1
<i>Ovis vignei</i>	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Capra hircus</i>	0	1	0.18	0.82	0	1	0.04	0.96	0	1	1	0	0.01	0.99	0.01	0.99	0	1	0.03	0.97	0.13	0.88	0.07	0.93	0	1	0	1
<i>Capra pyrenaica</i>	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ovibos moschatus</i>	0.14	0.86	0.43	0.57	0.4	0.6	0.1	0.9	0	1	1	0	0	1	0	1	0	1	0.3	0.7	0	1	0.93	0.07	0	1	0	1
<i>Rupicapra pyrenaica</i>	0.75	0.25	0	1	0	1	0.25	0.75	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ammotragus lervia</i>	0	1	0	1	0	1	0	1	0.82	0.18	0.04	0.96	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Bos mutus</i>	0	1	0	1	0	1	-	-	0	1	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Bos taurus</i>	0	1	0	1	0	1	-	-	0	1	1	0	0	1	0	1	0	1	0	1	0.15	0.85	0	1	0	1	0	1

Species	g.522A>G		g.516_517 insG		g.498G>C		g.482T>C		g.468G>T		g.463G>A		g.456A>G		g.444A>G		g.406A>G		g.395A>G		g.384T>G		g.320C>G			
	A	G	G	-	G	C	T	C	G	T	G	A	A	G	A	G	A	G	G	A	G	T	C	G		
<i>Ovis Aries</i>	0.02	0.98	0.12	0.88	0	1	0	1	0.17	0.83	0	1	0	1	0.15	0.85	0	1	0	1	0	1	0	1	0	1
<i>Ovis ammon</i>	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ovis canadensis</i>	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ovis musimon</i>	0	1	0.13	0.87	0	1	0	1	0	1	0	1	0	1	0.05	0.95	0	1	0	1	0	1	0	1	0	1
<i>Ovis vignei</i>	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Capra hircus</i>	0	1	0	1	0	1	0	1	0	1	0	1	0.05	0.95	0.01	0.99	0.23	0.77	0.16	0.84	0.64	0.36	0.08	0.92	0	1
<i>Capra pyrenaica</i>	0.13	0.87	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1
<i>Ovibos moschatus</i>	0.03	0.97	0	1	0	1	0	1	0	1	0	1	0	1	0.47	0.53	0	1	0	1	0	1	0	1	0	1
<i>Rupicapra rupicapra</i>	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Ammotragus lervia</i>	0	1	0	1	0.18	0.82	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Bos mutus</i>	T	T	0	1	0	1	0.92	0.08	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
<i>Bos taurus</i>	T	T	0	1	0	1	0	1	0	1	0.85	0.15	0	1	0	1	0	1	0	1	0	1	0	1	0	1

The three INDELs g.703\_704insAA, g.667\_668insC and g.666\_667insC existed simultaneously only in *O. moschatus*, *O. musimon* and *O. aries*. *C. hircus* had the two adjacent g.667\_668insC and g.666\_667insC INDELs, and *O. ammon* and *R. pyrenaica* the g.703\_704insAA. The highest frequency of the I<sub>668</sub> allele, related with heat stress resistance, was found in *O. musimon* (0.47) and *O. moschatus* (0.43) followed by *O. aries* (0.28) and *C. hircus* (0.18). The SNP g.660G>C seems to be exclusive to the Caprinae subfamily. Unfortunately, this region is a sequence of several consecutive cytosines and therefore it is difficult to know if in *bos* there is absence of cytosine at position g.660 before TSS or if the C<sub>660</sub> allele is fixed, as *bos* species only have 6 consecutive cytosines instead of the minimum 7 present in the rest of species here sequenced. In any case, C<sub>660</sub> appears to be the wild allele of the g.660G>C SNP.

Tajima's Neutrality Test (Tajima 1989) conducted for the sequences of the *HSP90AA1* promoter yields a negative D value lower than -2 (-2.56) which means an excess of low frequency polymorphisms relative to expectation. This fact could indicate a purifying selection removing alleles that diminish animal's biological fitness

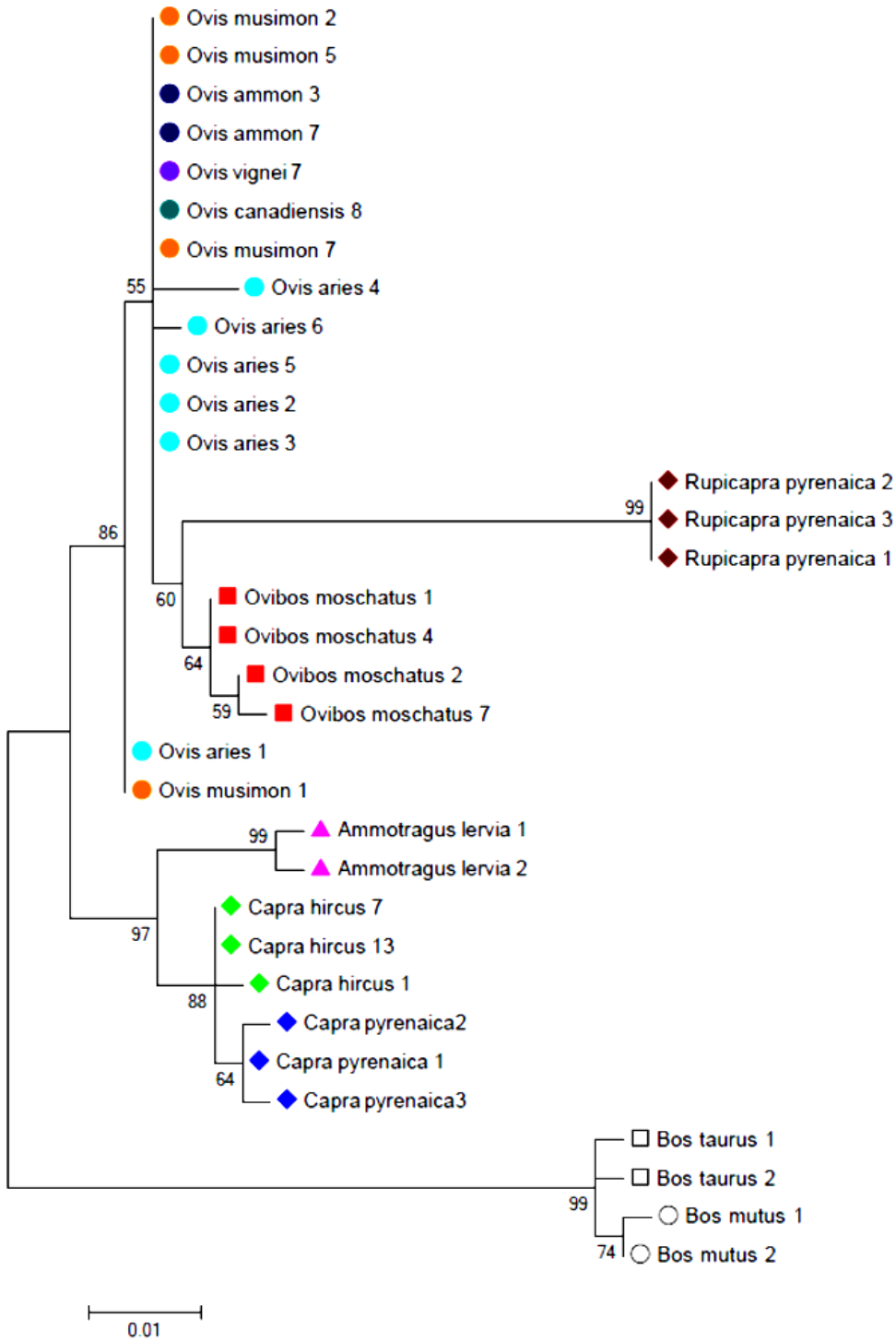
but also the presence of “young” beneficial mutations going to higher frequencies, as could be the case of the I<sub>-668</sub> cytosine insertion.

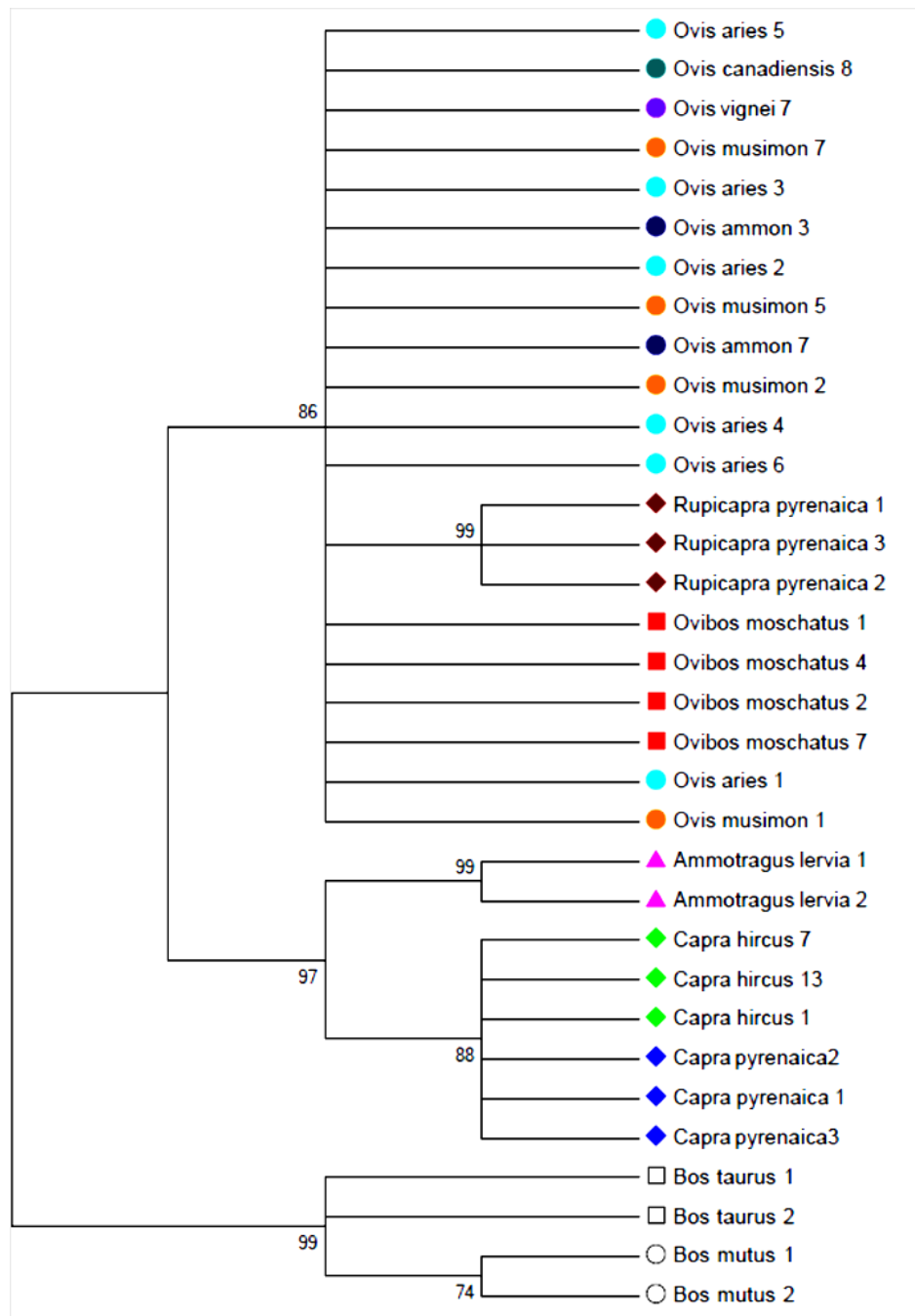
Figure 6 shows Maximum Likelihood bootstrap original and condensed trees based on the Tamura 3-parameter model and inferred from 5000 replicates. The analysis involved 33 nucleotide sequences. There were a total of 410 positions in the final dataset. Outgroup species (*B. taurus* and *B. mutus*) were located in a separate branch with a high bootstrap percentage (99). One branch were constituted by species of the *Capra* and *Ammotragus* genus (97) and other branch by *Ovis*, *Rupicapra* and *Ovibos* (86). In the ML consensus tree, *O. moschatus* was located as a sister species of *R. pyrenaica*. Species from the *Ovis* genus (*O. aries*; *O. musimon*, *O. vignei*, *O. ammon* and *O. canadiensis*) appear mixed since many haplotypes are shared among them and promoter sequences showed a high degree of similarity (Supplemental Figure 2).

## DISCUSSION

Previous studies from our group (Marcos-Carcavilla et al. 2010b), including those summarized in Chapter I, pointed out the existence of different expression profiles in sheep carrying alternative genotypes of some polymorphisms located at the *HSP90AA1* gene promoter depending on environmental temperatures. Animals carrying the II<sub>-668</sub>CC<sub>-660</sub> genotype showed higher expression rate (FC = 3.1 to 3.5) of the *HSP90AA1* gene than those with DD<sub>-668</sub>CC<sub>-660</sub>, DD<sub>-668</sub>CG<sub>-660</sub> and DD<sub>-668</sub>GG<sub>-660</sub> under heat stress environmental conditions (27°C average daily temperature and 34°C maximum daily temperature)

**Figure 6** – Molecular Phylogenetic analysis by Maximum Likelihood method (based on bootstrap, which is shown in the nod of each branch) developed with MEGA6. Original and condensed ML trees are shown, where numbers correspond to the haplotype identification.





### Relationship between polymorphism frequency in 31 sheep breeds and climatic variables

The results above described may contribute to clarify the phylogeographic relationship of some sheep breeds and the opposite correlations observed between the frequency of the  $I_{-668}$  and  $G_{-660}$  alleles and some climatic and geographic variables from the locations where they are reared. Since the  $I_{-668}$  allele is responsible of the

upregulation of the transcription of the gene under heat stress conditions (Chapter 1 and 2), high frequency of this allele is expected to be found in climates with high minimum (MINaT) and average annual (ANT) temperatures (positive regression coefficients, 0.11 and 0.13) and therefore with low thermal width (negative regression coefficient, -0.14). Despite no significant correlation was found among Total Annual Rainfall (TAR) and Maximum Rainfall (MxR) with MINaT and ANT, the frequency of the I<sub>-668</sub> allele seems to be associated with changes in these last variables. Thus the I<sub>-668</sub> allele frequency is high in climates with high TAR and MxR values. Looking at climatic variables of countries where sheep breeds analyzed are reared (Table 1) we can observe that Semi Arid (SA) regions showed greater average TW (24.87) and average ANT (13.70) and lower average MINaT (-1.15) than Semi Damp (SD) locations (average TW=18.69, ANT=11.04 and MINaT=4.47). Also in SD locations TAR and MxR values (626 and 103, respectively) are much higher than those of SA regions (372 and 58, respectively). Therefore, as SD are heater and damper than SA regions, it is possible to hypothesize that heat events accompanied with high rainfall in SD regions could be more stressful since thermal stress increases when high temperatures and high relative air humidity go together (Lally 1960).

Regarding the G<sub>-660</sub> allele in the gene promoter, opposite results to those of the I<sub>-668</sub> were observed, which agree with the transcription results above mentioned. The G<sub>-660</sub> allele is linked to the lowest expression rates of the *HSP90AA1* gene under both heat stress and mild temperature conditions. Therefore high frequencies of such allele are only expected in breeds reared in regions with low MINaT and ANT temperatures and high TW, in which heat is not a critical source of stress. Therefore, the negative association of the G<sub>-660</sub> frequency with MINaT (-0.16) and ANT (-0.13) and positive with TW (0.18) agree with such expectations. However, high frequencies of the C<sub>-660</sub> allele were found in all kind of locations but predominating in breeds reared in hot climates. This could be due to the genetic exchange that occurred during the development of modern breeds more than to adaptation processes. LD between g.667\_668insC and g.660G>C is little than 0.20 in whole breeds altogether and range between 0.001 and 0.540 across breeds, however D<sub>-668</sub> and G<sub>-660</sub> alleles are completely linked in the 836 animals genotyped, constituting the most thermo sensible haplotype.

### **Phylogenetic relationship between sheep breeds**

In the basis of Reynold's distances, two groups of breeds showing the minimum distances within group and maximum distances between groups can be established.

The first group was constituted by KRC, KRB, KAR, DGL, KARM and EDIL breeds. The second group was constituted by ME, PRAM, SZ, AS, KVR and CI breeds. Among breeds of these two groups, average, minimum and maximum distances were 0.267, 0.064 and 0.604, respectively. In these two groups of breeds, opposite frequencies of the two polymorphisms most related with gene expression differences (g.667\_668insC and g.660G>C) were observed. Thus, in the first group of breeds the average frequencies of the I<sub>-668</sub> and G<sub>-660</sub> alleles were 0.08 and 0.61, respectively. In all these breeds the G<sub>-660</sub> allele was the one with the maximum frequency and the I<sub>-668</sub> allele had a frequency <13%. In the second group of breeds, average frequencies of I<sub>-668</sub> and G<sub>-660</sub> alleles were 0.41 and 0.23, respectively. In all these breeds the I<sub>-668</sub> allele frequency was >30% and the G<sub>-660</sub> allele had a frequency <33%. Interestingly, all breeds from group 1, except KARM, are reared in SA or A climates, mainly from Asian regions. On the contrary, all breeds from group 2, except ME, are reared in SD Mediterranean climates. Average MINaT, ANT, TW, TAR and MxR were -6.57, 8.02, 28.52, 367.63 and 52.30, respectively, in group 1 and 7.87, 16.38, 17.47, 617 and 114.67, respectively, in group 2.

$Q^2$  values higher than 0.4 were only found for I<sub>-668</sub>, I<sub>-667</sub> and G<sub>-660</sub> suggesting the action of natural selection in driving the differential allele frequency distribution of these polymorphisms among sheep populations. Therefore, a correlation between genetic (allele frequencies) and environmental (climatic parameters) variables among some sheep breeds have been established which demonstrates that despite the great admixture existing among them and its domestication status, some footprints of the natural selection action can be glimpsed. This fact may be due to the general low artificial selection exerted over breeds of this species and their semi-extensive or extensive management conditions which may have retained some genes related with adaptation to environmental conditions existing in nature. Thus, breeds reared in SD climates, in which high temperatures and humidity are sources of physiological stress, have high frequency of alleles (I<sub>-668</sub> and C<sub>-660</sub>) related to higher expression rates of the *HSP90AA1* gene as response to heat stress (Chapter 1 and 2). However, low frequencies of these alleles were only found in those breeds reared in climates in which heat and humidity levels are not enough to induce a heat stress response. However, frequencies of A<sub>-601</sub>, A<sub>-522</sub> and I<sub>-516</sub> alleles ( $Q^2$  values < 0.4) are not influenced by climatic conditions and therefore its presence in the *HS90AA1* gene promoter seem to have no impact in the adaptation to environment of the ovine species. This finding was already suggested in Chapter 1 and 2 where these polymorphisms did not show

expression differences among genotypes when comparing samples collected under heat stress and thermo-neutral conditions.

In a large study where 49,034 SNPs were genotyped in 74 sheep breeds, Kijas and colleagues, (Kijas et al. 2012) found signs of directional selection in two candidate genes located at chromosome 18 ( $F_{ST}=0.428$ ), in which also the *HSP90AA1* gene is located. One of them was *ABHD2* (abhydrolase domain containing 2) which has, among other functions, a role in the response to wounding. This protein that interacts with UBC (polyubiquitin C) has a high expression rate in testicle (BioGPS) and correlates with *HSPAIL* (Heat shock protein 70 kD like). Hsp70 is a well known protein involved in the heat shock response which is part of the Hsp90 complex. Therefore, although authors (Kijas et al. 2012) recognize that the identification of adaptive alleles has not been achieved, some footprints of directional selection over genes more or less directly related with adaptive traits can be found.

When assessing evidence for an ecocline, it is crucial to control population history and structure, for accurately assessing whether a correlation between a genetic variant and geographic or climate variables is due to natural selection (Endler 1977). For example, if migration patterns correspond closely with variation in a particular climate variable, the correlations between neutral alleles and that climate variable may be high even if selection has not acted on the locus. Conversely, if selection effects are lower to that of population structure on allele frequencies, correlations may be underestimated if population history is not taken into account (Hancock et al. 2011). This is the reason why PLSR and SAM approaches cannot be used independently, without comparing results with specialized statistic methods based on population genetics theories, and focus on the analysis of genetic data as the Bayesian Test of Beaumont and Balding (LOSITAN). Thus, among all loci-environment associations detected by PLSR and SAM methods, only the frequency of two polymorphisms, the g.703\_704insAA and the g.522A>G, seem to be under the action of some selective processes. The g.703\_704insAA showed a high  $F_{ST}$  outlier which makes it a candidate to directional selective processes. The low  $F_{ST}$  outlier of the g.522A>G SNP reveals the possibility of balancing selection acting over its frequency. The g.703\_704insAA is highly linked with the g.660G>C SNP ( $r^2=0.86$  in the whole data) ranging  $r^2$  values in most breeds from 0.84 to 1. Thus, directional selection predicted for the g.703\_704insAA could be extended to the SNP g.660G>C for which differential expression of the *HSP90AA1* gene has previously been assessed depending on genotype, but not with the g.667\_668insC. The high degree of conservation in LD



phase found in this short sequence in all almost breeds, independently of their geographic origin, could indicate that high levels of gene flow have occurred between populations following domestication, as it is suggested by Kijas and coworkers (Kijas et al. 2012), but also, a selection pressure over this DNA region (Nielsen et al. 2005).

### **Characterization of the *HSP90AA1* promoter in species of the Caprinae and Bovinae subfamilies**

The Bovidae family includes more species than any other extant family of large mammals, but their phylogenetic relationships remain largely unresolved in part because it appears to represent a rapid, early radiation into many forms without clear connections among them (Hernandez Fernandez et al. 2005). Furthermore, certain morphological traits have evolved several times within the family to create evolutive convergences that obscures true relationships (Gentry 1992). The Caprinae subfamily includes bovids adapted to extreme climates and difficult terrains. Fossil records are poorly documented but the group first appeared during the upper Miocene (Gentry 1994). In a recent work, a complete estimate of the phylogenetic relationships in Ruminantia has been proposed combining morphological, ethological and molecular information (Hernandez Fernandez et al. 2005). The resolution of the supertree varies among groups and some component clades, particularly Caprinae (67.7%), are much less well resolved than others (e.g. Bovinae, 95.7%). In particular, the position of the genera *Budorcas* and *Ovibos* has been controversial, having sometimes constituted the tribe Ovibovini, and others been separated and located in different tribes. In general, the genus *Ovis* is split into a "New World" clade represented by *O. dalli* and *O. Canadensis*, and an "Old World" clade including the two sister species *O. vignei* and *O. Aries* in the same branch and *O. ammon* in a sister branch (Hassanin et al. 1998, Hernandez Fernandez et al. 2005). In our work, haplotypes from *O. vignei*, *O. canadiensis* and *O. musimon* appeared mixed with those from *O. aries*. *O. aries* and *O. musimon* share many polymorphic sites (7) as expected from the past hybridization between both species.

Ropiquet and Hassanin (Hassanin et al. 2004) using mitochondrial and nuclear DNA sequences located *A. lervia* closer to goats (*Capra*) and *O. moschatus* closer to *R. pyrenaica*. However, in recent works (Bibi 2013, Hassanin et al. 2013) *A. lervia* was close to *Rupicapra* genus within the Caprina tribe and *O. moschatus* was distant from them within the Ovibovina tribe. Our tree located *A. lervia* as a sister species of *C. hircus* and *C. pyrenaica* (bootstrap proportion = 97) and *R. pyrenaica* closer to *O.*

*moschatus* (bootstrap proportion = 60). In the work of Matthee and Davis (Matthee et al. 2001) using data from nuclear DNA, a polytomy for *C. hircus*, *O. moschatus* and *O. aries* was found. However, when analyzing nuclear DNA joined to mtDNA data, *C. hircus* and *O. aries* appear as sister species separated from *O. moschatus*. In our work we have observed a relative high similarity between *O. moschatus*, *O. aries* and *O. musimon* species regarding polymorphism sharing among them.

Although *O. moschatus* is currently restricted to Greenland and the Arctic Archipelago (Campos et al. 2010b), the highest frequencies of alleles related with the heat stress response ( $I_{-668} = 0.43$ ;  $C_{-660} = 0.90$ ) were found for the individuals analyzed in this species. Fossils of this species have occasionally been found in southwest Europe. It seems that *Ovibos* did not inhabit exclusively cold tundra during the Pleistocene (Campos et al. 2010b). *Praeovibos*, an older morphotype of *O. moschatus*, does not appear to have been restricted to inhabiting cold climates as its remains have also been identified in temperate and Mediterranean forest (Cregut-Bonnoure 1984, Rivals et al. 2009). In contrast to modern *Ovibos*, *Praeovibos* was distributed over much more southern latitudes, samples have been found as far south as France and Spain (Cregut-Bonnoure 1984, McDonald 1991, Rivals et al. 2009), which indicates that *Praeovibos* was less restricted to a specific ecological niche than *O. moschatus* (Campos et al. 2010b). Could these high frequencies of alleles related with the heat stress response found in *O. moschatus* came from its *Praeovibos* ancestor? Lent (Lent 1988) indicates that *O. moschatus* is sensitive to both climate warming and fluctuations, that is why Campos and colleagues (Campos et al. 2010b) hold these factors responsible of actual confinement of the *O. moschatus* to Greenland and the Arctic Archipelago but not a human impact. Our results regarding the polymorphisms of the *HSP90AA1* gene in this species seems to indicate that the actual *O. moschatus* is genetically well prepared to tolerate warm climates. Therefore, which could be the reasons to its actual geographic limitations? Climate change is known to affect not only animal's thermo sensitivity but also by triggering vegetation change (Barnosky et al. 2004, Rivals et al. 2009). Increasing temperature pushed the adaptive vegetation balance firmly towards bogs, shrub tundra, forest and low-nutrient acidic soils, which resulted in communities of conservative plants highly defended against herbivore and supporting a small biomass of large mammals (Guthrie 2006). Palmqvist and coworkers (Paul Palmqvist 2008) in an ecomorphological analysis of the early Pleistocene fauna of Venta Micena (Orce, Guadix-Baza basin, SE Spain), provide interesting clues on the

physiology, dietary regimes, habitat preferences, and ecological interactions of large mammals.

Regarding those polymorphisms for which our group has detected relationship with hot climates adaptation by its association with the expression rate of the *HSP90AA1* gene under heat stress conditions (g.667\_668insC and g.660G>C), it is noteworthy that they were only segregating in *C. hircus*, *O. moschatus*, *O. musimon* and *O. aries*. Since we have only one sample from *O. vignei* and two samples from *O. ammon*, we cannot extract any conclusion from these two species. It seems reasonable to hypothesize that these polymorphisms could come from an ancestral species common to the *Ovis*, *Ovibos* and *Capra* genera but not to *Ammotragus*. However, also it is possible that the evolvability of this gene may be due to its physical susceptibility to mutagenesis and therefore that the similitude/difference found in the species analyzed does not be related with their phylogenetic relations.

We have assessed that despite the domestication process occurred 11,000 years BP, sheep breeds showed some genetic footprints related to climatic variables existing in the regions where they are reared. Thus artificial selection carried out by humans to improve productive traits in this species seems to occur concurrently with natural selective forces for traits related with the adaptation to environmental conditions. Adaptation of breeds to hot climates can suppose a selective advantage to cope with global warming caused by climatic change. Polymorphisms of the *HSP90AA1* gene detected in the *Ovis aries* species can be used in selection programs to improve animals resistance to hot environments. Mutations of the ovine *HSP90AA1* gene promoter are also been found in wild species from the Caprinae subfamily, indicating a great antiquity of these mutations which can help us to elucidate how climatic conditions have evolved in the past.

# Chapter 4

## From Genotype to Phenotype

Ovine *HSP90AA1* promoter polymorphisms are related with ovine sperm DNA fragmentation



## INTRODUCTION

Among others, climate factors can have diverse and often strong effects on reproduction efficiency, with obvious consequences in animal's fitness (Grazer et al. 2012) which can result ultimately, in high economic losses for breeders (Groen AF 1997, Legarra et al. 2007). Exposure to adverse conditions of high temperature and humidity may led to a reduction of the number of spermatozoa (Jannes et al. 1998, Perez-Crespo et al. 2008) and also to an impairment of their functionality (Yaeram et al. 2006, Perez-Crespo et al. 2008), which will be accompanied by a transient period of partial or complete infertility. After heat stress, viability of the spermatozoa may not be compromised but some of them will appear with DNA damage. A reduction in sperm DNA integrity has been described in human (Paul et al. 2008), mice (Banks S 2005, Pérez-Crespo M 2008) and rams (Fleming JS 2004) as well as alterations in DNA, RNA and protein synthesis, and abnormal chromatin packing in mice (Sailer BL 1997, Banks S 2005, Pérez-Crespo M 2008) under heat stress conditions. Proper compaction and structure of DNA has been reported to have important functional roles, being essential for DNA replication and embryonic development (Ward 2010, Dominguez et al. 2011).

Two singular characteristics differentiate sperm from somatic cells: protamination and absence of DNA repair mechanisms. During spermiogenesis, protamines replace the majority of histones (Conwell et al. 2003). This dense compacting gives protection against exogenous assault to the sperm DNA (Barratt et al. 2010). DNA repair in sperm is terminated as transcription and translation stop at post-spermiogenesis, so these cells have no mechanism to repair the damage occurred during their transit through the epididymis and post-ejaculation (Gonzalez-Marin et al. 2012). Therefore, assessing levels of DNA fragmentation can be a useful tool for evaluating the effects of heat stress on sperm and its consequences on male fertility. Sperm DNA fragmentation is considered a non compensable trait which implies that the pregnancy ratio does not change when the number of sperm inseminated increases (Evenson et al. 2006, Amann et al. 2012).

The relationship between sperm DNA fragmentation index (DFI) and male fertility has been studied in humans (Evenson et al. 1999, Spano et al. 2000, Bungum et al. 2004), bulls (Ballachey et al. 1987) and boars (Didion et al. 2009) Thresholds for sub fertility were much lower for boars (6%) and bulls (14.2%) than that for humans (30%).

Recently, in rams Nordstoga et al. (Nordstoga et al. 2013) showed an association between sperm DNA integrity and the non returned rate in Norwegian cross-bred rams.

A role for Hsp90 in spermatogenesis was first described in *Drosophila melanogaster*, where males with certain transheterozygous combinations of mutant Hsp90 alleles are sterile and display a disrupted meiosis (Yue et al. 1999). In mice, a requirement of the Hsp90 $\alpha$  for spermatogenesis has been shown (Grad et al. 2010). Authors pointed out that Hsp90 $\alpha$  must be necessary at least during the first wave of spermatogenesis. In the absence of Hsp90 $\alpha$  meiosis arrests very specifically towards the end of the pachytene stage, disassembling of homologous chromosomes fail and normal diplotene spermatocytes are totally absent. Also, an absence of a comparable phenotype in Hsp90 $\alpha$  mutant females was observed (Grad et al. 2010). Also in mice (Alekseev et al. 2005) a chaperoning function of the Hsp90 protein to assist the proper folding of tNASP (testis histone binding protein) to bind linker histones, have been observed. Expression of Hsp90 and tNASP precede the expression of HI1 (histone subtype I restricted to the testis) in pachytene spermatocytes (Meistrich et al. 1985, Drabent et al. 1998). Alekseev and coworkers (2005) pointed out that after the synthesis of linker histones in the cytoplasm they are bound to a complex containing NASP and Hsp90. NASP-HI is subsequently released from the complex and translocated to the nucleus where the HI is released to DNA binding (Alekseev et al. 2003).

In previous chapters, the effect of g.660G>C and g.667\_668insC in *HSP90AA1* expression has been deeply proved. The present study try to connect differences in the transcription rate of the *HSP90AA1* gene depending on genotype by environmental circumstances with those observed in a male's reproductive trait. Specifically, we aimed to know more about the correlation between the expression patterns of the gene and the consequences in sperm DNA fragmentation levels derived from an exposure to a heat stress environment. To address these questions we have examined: 1) if heat stress has an effect on chromatin stability of ram's sperm, and 2) if a differential response to heat stress occurs based on male's g.660G>C and g.667\_668insC genotypes. For that, semen samples from males with alternative genotypes for both mutations were collected and exposed to heat during 48 h. Daily temperature and relative humidity for the 60 days prior to semen collection were

recorded and their effect on spermatozoa resistance/susceptibility to heat stress were assessed. Finally, it was examined whether there was a sperm differential response to heat stress depending on the genotype of those males analyzed.

## MATERIAL AND METHODS

### Animals

To develop the actual experiment, a total of 61 adult rams of Manchega breed were selected from the group of 120 previously used in the expression studies from Chapter 1. All males were kept in the Regional Centre of Animal Selection and Reproduction (CERSYRA) in Valdepeñas (Spain) in the same environmental conditions. Rams were trained for semen collection by artificial vagina maintaining a regimen of regular collection.

### Weather data

Castilla-La Mancha is a region located in the south of Spain which is characterized by an arid environment with low rainfall and high temperatures. Meteorological data was provided by the Irrigation Advisory Service for Farmers (SIAR) in Castilla-La Mancha. The meteorological data set consisted of hourly measures of temperature (°C) and relative humidity (%) on 245 days from March to October 2010. Daily average temperature ( $T_{ave}$ , °C), daily maximum temperature ( $T_{max}$ , °C) and daily average relative humidity (RH, %) were calculated from these hourly records. A temperature-humidity index (THI) was also calculated as proposed by Marai et al. (Marai et al. 2007) by combining daily average temperature ( $T_{ave}$ ) in °C with daily average relative humidity (RH) % 0.01:

$$THI = T_{ave} - [(0.31 - 0.31 \cdot RH) \cdot (T_{ave} - 14.4)]$$

To better understand THI values, Marai et al. (Marai et al. 2007) proposed a scale of the effect of THI values over heat stress. Thus, values of  $THI < 22.2$  assume absence of heat stress;  $22.2 \geq THI < 23.3$  means moderate heat stress;  $23.3 \leq THI < 25.6$ , severe heat stress; and  $THI \geq 25.6$  indicates extreme severe heat stress.

### Semen Samples collection

Semen collection was made by artificial vagina. A total of 8 collections per male were carried out, from March to October as a way to ensure that sperm analyses



were conducted in different weather conditions of temperature and humidity (Table 1). After collection, sperm samples were diluted in phosphate-buffered saline (PBS; pH 7.5, 310 mOsm/kg) with 0.5% bovine serum albumin and then incubated in a saline medium at 37 °C during 48 hours. The sperm chromatin stability was assessed after collection (0h) and after 24 and 48 h. Sperm incubation at 37°C has the aim to mimic the environmental circumstances existing at ewe reproductive track.

**Table 1.** Climate parameters at days of sperm samples collection.

Data from: Manzanares (Ciudad Real) Meteorological Station, coordinates 654m-38° 59'47N-03° 22'23W (<http://crea.uclm.es/siar>)

AvT = average temperature (°C)

MaT = maximum temperature (°C)

MiT = minimum temperature (°C)

Rh = average relative humidity (%)

Rhmax = maximum relative humidity (%)

Rhmin = minimum relative humidity (%)

THlavr = THI calculated with the average temperature and relative humidity

sperm collection date	AvT	MaT	MiT	Rh	Rhmax	Rhmin	THlavr	THlmax
01/03/2010	7.8	11.9	4.8	79.5	90.3	58.5	8.22	11.98
19/05/2010	18.4	26.6	8.1	46.1	82.8	24.1	17.73	25.95
29/06/2010	23.7	32.8	16.9	58.7	91.6	24.4	22.51	32.32
14/07/2010	24.9	32.6	17.7	42.2	67.6	20.1	23.02	30.77
27/07/2010	26.2	36.4	15.4	45.6	87.8	13.8	24.21	35.57
10/08/2010	27.5	36.6	19.1	48.6	77	24.3	25.41	35.02
24/08/2010	23	31.5	15.9	52.2	86.4	23.6	21.73	30.78
05/10/2010	14.4	22.7	5.6	67	96.3	31.8	14.4	22.6

THlmax = THI calculated with the maximum temperature and the maximum relative humidity

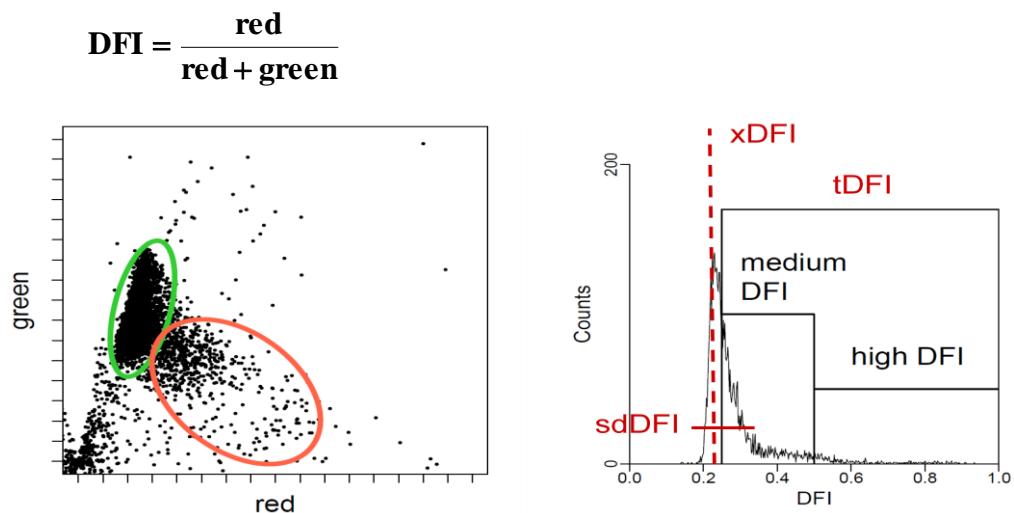
Temperature humidity index (THI) calculated as  $THI = T - [0.31(1 - RH)(T - 14.4)]$ , T = temperature in °C; RH = relative humidity/100 (Marai et al. 2007).

## DFI assessment

Chromatin stability was assessed by using the Sperm Chromatin Structure Assay (SCSA) technique (SCSA Diagnostics, Inc., Brookings, SD, USA) (Evenson et al. 2000, Evenson et al. 2002). This technique is based on the susceptibility of the sperm DNA to acid-induced denaturation *in situ* and in the Acridine Orange (AO) metachromatic acid nucleic staining. This stain fluoresces green when combined with double stranded DNA and red when combined with single stranded DNA (denatured). This technique has been used in rams with good results (Garcia-Macias et al. 2006, Garcia-Alvarez et al. 2010). García-Álvarez et al. (Garcia-Alvarez et al. 2010) provides a more in depth explanation of the procedure used to assess the sperm chromatin

stability in Manchega rams. Briefly, sperm samples were diluted with TNE (0.15 M NaCl, 0.01 M Tris HCl, 1 mM EDTA; pH 7.4) buffer at a final sperm concentration of  $2 \times 10^6$  cells/ml and flash frozen in LN2 and stored at  $-80^\circ\text{C}$  until analysis. For the analysis, the samples were thawed on crushed ice, and 200  $\mu\text{l}$  were put on a cytometry tube. Immediately, 400  $\mu\text{l}$  of an acid-detergent solution (0.08 N HCl, 0.15 M NaCl, 0.1% Triton  $\times$  100; pH 1.4) were added to the tube. After exactly 30 s, 1.20 ml of Acridine Orange (AO)-staining solution (0.037 M citric acid, 0.126 M  $\text{Na}_2\text{HPO}_4$ , 0.0011 M disodium EDTA, 0.15 M NaCl; pH 6.0,  $4^\circ\text{C}$ ) containing 6 mg/ml electrophoretically purified AO was added. Stained samples were analyzed just after 3 minutes by flow cytometry, being the excited AO used as the fluorophore. AO was excited by using an argon laser providing 488  $\mu\text{m}$  light. A total of 5000 events were accumulated for each sample. We have expressed the extent of DNA denaturation in terms of DNA Fragmentation Index (DFI), which is the ratio of red to total (red plus green) fluorescence intensity, i.e. the level of denatured DNA over the total DNA (Evenson et al. 2002). The DFI value was calculated for each sperm cell in a sample, and the resulting DFI frequency profile was obtained (Figure 1). Each sperm sample was characterized by a mean ( $\text{xDFI}$ ) and a standard deviation ( $\text{sdDFI}$ ). Total DNA fragmentation index ( $\text{tDFI}$ ) was defined as the percentage of spermatozoa with a DFI value over 25.

**Figure 1.** Sperm chromatin structure assay (SCSA) process.



### HSP90AA1 genotypes

The 61 adult rams of Manchega breed used in this study were selected on the basis of its g.660G>C genotype. Genotypes of rams were previously assessed for the expression studies (Chapter 1). Thus, sperm DNA fragmentation were characterized for twenty males of each genotype CC<sub>.660</sub>, CG<sub>.660</sub> and GG<sub>.660</sub>. Table 2 showed genotype and allele frequencies of polymorphisms in this group of rams.

**Table 2.** Genotype and allele frequencies of the g.667\_668insC and g.660G>C located at the *HSP90AA1* promoter in 61 rams used in sperm DNA fragmentation assays (I = insertion; D = deletion).

rs# <sup>1</sup>	postion at DQ983231.1	Genotype	N	Frequency
rs397514115.2	g.667_668insC	II	2	0.033
		ID	13	0.213
		DD	46	0.754
rs397514116	g.660 G>C	GG	21	0.344
		CG	20	0.328
		CC	20	0.328

rs# <sup>1</sup>	postion at DQ983231.1	Genotype	2N	Frequency
rs397514115.2	g.667_668insC	I	17	0.139
		D	105	0.861
rs397514116	g.660 G>C	G	62	0.508
		C	60	0.492

<sup>1</sup> rs# = referente SNP ID number

### Association analyses among *HSP90AA1* genotypes and sperm DNA fragmentation

To assess the effect of heat load on sperm DNA integrity of animals carrying alternative genotypes for the polymorphisms of interest, we have examined the relationship among sperm DNA fragmentation indicators and average THI, since this climatic parameter summarized the effect of both temperature and humidity. We compared the degree of sperm DNA fragmentation between genotypes in response to weather conditions from the day 60 previous to semen collection to the day of collection. The aim was to identify those stages of the spermatogenesis in which certain genotypes are more susceptible to heat stress and to identify which polymorphisms play a more important role. For this, a linear mixed-effects model

including genotype, sperm incubation time (0h, 24h and 48h), THI (Temperature Humidity Index) and its interactions as fixed effects and male as random effect were conducted. The model was as follow:

$$y_{ijkl} = \mu + IT_j + \mathbf{f}(T)_k + G \times \mathbf{f}(T)_{ik} + a_i + e_{ijkl}$$

where:

- $y_{ijkl}$ : sperm DNA fragmentation measure
- $\mu$ : global mean
- $IT_j$ : sperm incubation time (3 levels: 0, 24 and 48 h)
- $\mathbf{f}(T)_k$ : effect of THI for each of the 60 days prior sperm collection
- $G \times \mathbf{f}(T)_{ik}$ : effect of the interaction genotype-THI
- $a_i$ : male (61 levels)
- $e_{ijkl}$ : heterogeneous random residual error  $\sim N(0, \sigma_{eij}^2)$

Since it was expected that the effect of temperature on sperm DNA fragmentation measure values was revealed from a threshold, the following function was used to model the temperature effect:

$$\mathbf{f}(T)_k = \left\{ \begin{array}{l} 0, \text{ if } THI \leq k \\ b(THI - k), \text{ otherwise} \end{array} \right\}$$

where,  $k$  is the selected threshold. A threshold value of 22 was used, based on the THI heat stress categories reported by Marai and colleagues (Marai et al. 2007). As genotype ( $G_1$ ), two polymorphisms located at the promoter of the *HSP90AA1* gene were considered and three independent mixed models analyses were conducted based on the genotype included in the model: (i) the g.660G>C SNP; (ii) the INDEL g.667\_668insC; and (iii) both polymorphisms combined genotypes. Heterogeneous residual variance for the effect of sperm incubation time was also considered. Multiple comparisons among genotypes were conducted for each day, from the day 60 prior to semen collection to date of collection (N=61). Two scenarios were considered: Non Heat Stress (NHS) (sperm collections from March and May 2010) and Heat Stress (HS) (sperm collections from June, July, August and October) conditions.

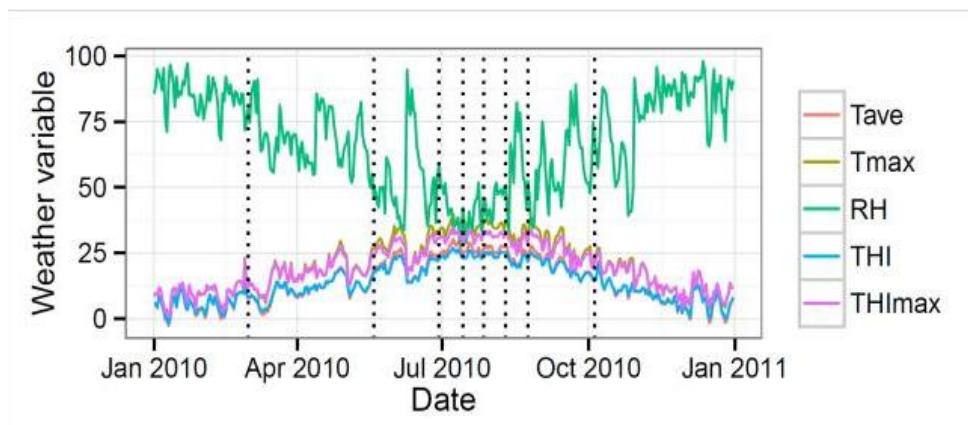
Statistical analysis was performed using the R 3.0.3 statistical language (Team 2013). For multiple comparisons analyses, Bonferroni correction was considered.

## RESULTS

### Weather data

Figure 2 shows the evolution of average ( $T_{ave}$ ) and maximum ( $T_{max}$ ) daily temperatures, average daily relative humidity (RH) and average of the Temperature Humidity Index (THI) along the period when sperm samples were collected. Average daily temperatures higher than  $25^{\circ}\text{C}$  and maximum daily temperatures higher than  $30^{\circ}\text{C}$  were observed from June to August. For the same period, minimum daily temperatures never dropped from  $10^{\circ}\text{C}$ . The highest values of RH were found from January to March (79 to 94%). However, RH higher than 70% was observed in some points of the summer season (June, July and August), probably coinciding with summer storms. There were also maximum values of RH greater than 90% in June and August. From June to August THI, ranges from 22.4 to 27.0, which include the three THI heat stress categories, moderate (22.2 to 23.3), severe (23.3 to 25.6) and extreme (25.6 and over) (Marai et al. 2007). If maximum daily THI is considered, we found days from May to August in which this parameter was in the range of extreme heat stress.

**Figure 2.** Trends of daily average ( $T_{ave}$ ,  $^{\circ}\text{C}$ ) and maximum ( $T_{max}$ ,  $^{\circ}\text{C}$ ) temperatures, relative humidity (RH, %) and average (THI) and maximum (THI $_{max}$ ) temperature humidity index along the year 2010 (Data from SIAR <http://crea.uclm.es/siar/datmeteo/>). Dotted lines are days of semen collection.

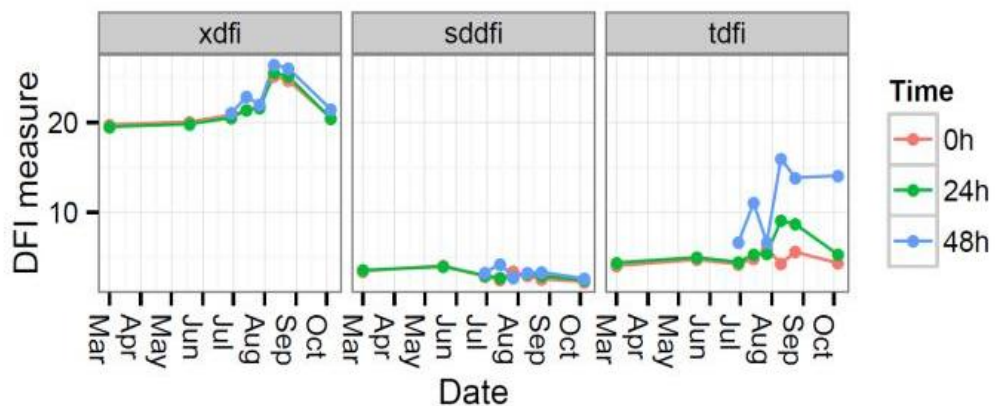


### DFI values

Figure 3 shows the evolution of xDFI, sdDFI and tDFI values with the incubation time (0h, 24h and 48h) along the period of the year from which sperm

samples were collected. There were measures at 48h of incubation time only for sperm samples collected from the end of May to October. The xDFI values for the three incubation times did not show significant changes in sperm samples collected between March and the beginning of August (xDFI around 20-21). However, their values increased to 26 in August, dropping again to values of 20-21 in October. The sdDFI values did not experiment significant changes along the year for any incubation time, showing a very stable value close to 2.6. However, for tDFI values, differences along the year among alternative incubation times were observed. Thus, tDFI values did not show important changes along the period studied (1.9) for 0h of incubation time (after semen collection), except a certain decrease observed in July (0.3). After 24h of incubation, an average tDFI value of of 2.3 was found for samples collected from March to June. tDFI value decreased to 0.7 in July, and increased to 6.8 in August, dropping to 3.0 in October. Finally, for 48h of incubation, an increase of tDFI from 4.4 to 6.4 was found for measures taken in June and July, and decreased to 2.1 for samples collected at the end of July. A clear increase of tDFI values was observed for samples collected in August, 13.7. In samples collected in October tDFI values dropped to 11.5. Incubation times of 24 and 48 h lead to same trend on tDFI values but with different magnitude, being greater when sperm samples were incubated during more time. As tDFI values showed the greatest changes depending on climatic variables existing at sperm collection dates, this was the only sperm DNA fragmentation indicator used in the association analyses.

**Figure 3.** Changes in xDFI, sdDFI and tDFI values with the incubation time (0h, 24h and 48h) along the period of the year from which sperm samples were collected.



### Sperm DNA fragmentation as a function of environmental conditions and *HSP90AA1* genotypes.

Table 3 showed raw tDFI data from three incubation times for the non-heat stress (NHS) and heat stress (HS) semen collection periods. We have observed that as temperature rises, levels of sperm DNA fragmentation also increase with clear differences between non-heat stress semen collections and those conducted under heat stress conditions.

Figure 4 shows estimates of tDFI differences values between genotypes of *g.660G>C*, *g.667\_668insC* and the combined genotype of both mutations depending on THI at each day of the period comprised between days 0 to 60 bsc (before sperm collection) for NHS and HS scenarios. Stages of the spermatogenesis process are marked. Bonferroni correction was applied to take into account for multiple tests.

When thermoneutral conditions occurred along the spermatogenesis process, no differences were shown among tDFI values and alternative genotypes of the *g.660G>C* SNP. However, significant differences in tDFI values among genotypes were detected when heat stress events occur at some stages of the spermatogenesis. Thus, the *GG<sub>.660</sub>* genotype showed higher tDFI values than those observed for *CC<sub>.660</sub>* when heat stress events occurs at periods comprised between days 29 to 34 bsc (1.14 to 1.35 folds per THI unit) and 10 to 12 bsc (1.72 folds). Only for the period 10 to 12 days bsc the *GG<sub>.660</sub>* genotype showed higher tDFI values than the *CG<sub>.660</sub>* one (1.31 folds). No difference in tDFI values were detected between *CC<sub>.660</sub>* and *CG<sub>.660</sub>* genotypes.

**Table 3.** Sperm DNA fragmentation levels (tDFI) from three incubation times for the non-heat stress (NHS) and heat stress (HS) semen collection periods. Data are mean  $\pm$  standard errors.

Semen collection period	0 h	24 h	48 h
NHS	4.36 $\pm$ 0.12	4.64 $\pm$ 0.15 <sup>a</sup>	6.67 $\pm$ 0.77 <sup>a</sup>
HS	5.03 $\pm$ 0.31	6.75 $\pm$ 0.38 <sup>b</sup>	12.27 $\pm$ 0.81 <sup>b</sup>

tDFI = total DNA fragmentation index (percentage of spermatozoa with a DNA Fragmentation Index (DFI) value over 25 %.)

NHS = non-heat stress period (semen collections from March to May)

HS = heat stress period (semen collections from June to October)

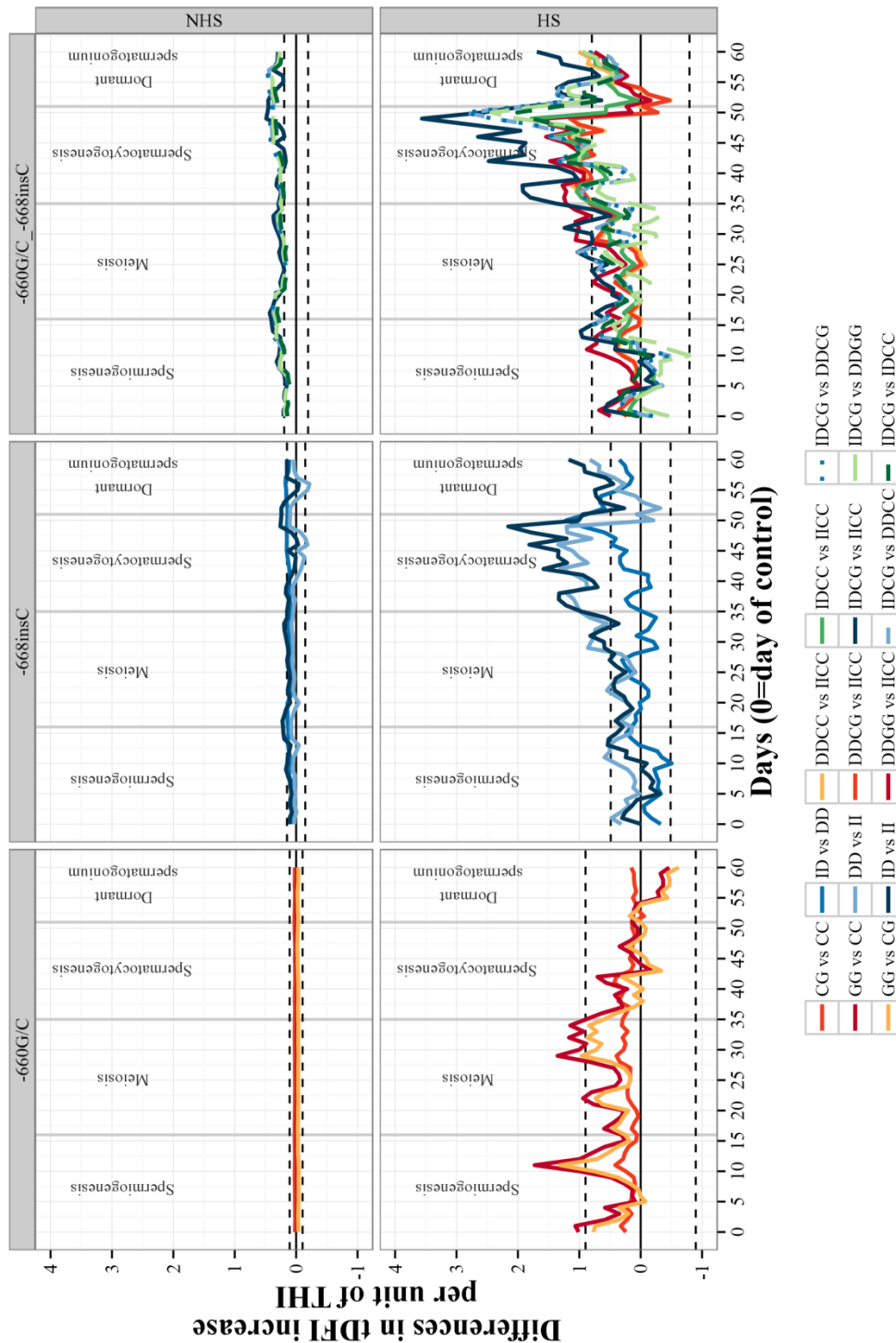
<sup>a-b</sup>: different superscript letters within column indicate significant differences between semen collection periods.

Under NHS conditions the genotype ID<sub>-668</sub> showed higher tDFI values than the II<sub>-668</sub> at some periods along the spermatogenesis process (days 15-17 bsc (0.22 folds); days 32-33 bsc (0.19 folds) and days 49-52 bsc (0.25 folds) but the magnitude of such difference was small. There were two peaks around days 46 and 56 bsc in which the II<sub>-668</sub> genotype showed higher tDFI values than the DD<sub>-668</sub> but even with lower magnitude than in the previous case (0.18 to 0.21 folds). When HS conditions took place, more convincing results were obtained. In this way, the ID<sub>-668</sub> and DD<sub>-668</sub> genotypes showed very high tDFI values than the II<sub>-668</sub> (more noticeable for the ID<sub>-668</sub> genotype than for the DD<sub>-668</sub> one) in the period comprised between days 37-49 bsc and 29-32 bsc. Thus, for the period 37-49 days bsc ID<sub>-668</sub> and DD<sub>-668</sub> genotypes showed tDFI values 1.33 to 2.15 folds and 1.20 to 1.31 folds higher than the II<sub>-668</sub>, respectively. For the period comprised between days 29 to 32 bsc differences among genotypes were smaller than those previously described (0.82 to 0.85 folds for ID<sub>-668</sub> and DD<sub>-668</sub> genotypes comparing with II<sub>-668</sub>). These differences decreased along the spermatogenesis process disappearing around day 27 bsc.

When the combined genotypes for the polymorphisms g.667\_668insC and g.660G>C were considered, a mixed pattern but mainly controlled by the INDEL mutation was found. Under NHS conditions, the genotype ID<sub>-668</sub>-CG<sub>-660</sub> was the worst in terms of tDFI values compared with the remaining genotypes in the periods comprised between days 48 to 52 bsc (0.30 to 0.48 folds), 32 to 33 bsc (0.24 to 0.37 folds) and 14 to 17 bsc (0.27 to 0.43 folds). These differences had lower magnitudes (0.13 to 0.27 folds) in the central stages of the spermatogenesis process (days 21 to 30 bsc). For the HS case (only significant contrasts are shown) much higher magnitude of differences among genotypes were found for the initial period of the spermatogenesis which comprises the period between 36 to 49 days bsc. Again the ID<sub>-668</sub>CG<sub>-660</sub> genotype was the one with the highest tDFI values (2.47 to 3.56 folds) when comparing with II<sub>-668</sub>CC<sub>-660</sub>, DD<sub>-668</sub>CC<sub>-660</sub>, DD<sub>-668</sub>GG<sub>-660</sub> and ID<sub>-668</sub>CC<sub>-660</sub> genotypes. Differences among genotypes were decreasing from day 36 bsc to the end of the spermiogenesis stage (1.24 to 1.92 folds). The genotype showing the lowest tDFI values was II<sub>-668</sub>CC<sub>-660</sub>.



**Figure 4.** Estimates of the differences of tDFI values between genotypes of the g.660G>C, g.667\_668insC and the combined genotype of both mutations depending on THI at each day of the period comprised between days 0 to 60 bsc for non heat stress (NHS) and heat stress (HS) collections. Stages of the spermatogenesis process are marked. Contrasts in the first column of the bottom legend apply to g.660G>C genotypes. Contrasts in the second column of the bottom legend apply to g.667\_668insC genotypes. Contrasts in the third to fifth columns of the bottom legend apply to combined genotypes. For reference values of tDFI under NHS and HS, see Table 2.



## DISCUSSION

The heat shock response is one of the main prosurvival activities of cells. In particular, the sensitivity of mammalian germ cells to environmental heat stress has been extensively studied (Hansen PJ 2009, Kim B 2013). Among others, cellular consequences of this stressor are protein miss-folding, DNA damaging, inhibition of DNA repair systems (Stecklein et al. 2012), and the inhibition of multiple processes associated with DNA replication (de Carcer 2004, McClellan et al. 2007, Velichko et al. 2013) and the maturation of chromatin (Zhao et al. 2005, Ruden et al. 2008). To cope with these effects, cells increase the expression of heat shock proteins (HSPs). This confers a transient protection, leading to a state that is known as thermotolerance, whereby cells become more resistant to various toxic insults, including otherwise lethal temperature elevations. Moreover, HSPs are expressed, though at lower levels, under normal conditions. This observation can be explained by the fact that HSPs are molecular chaperones for protein folding that play a central role in protein homeostasis (Freeman BC 2000, Diller KR 2006).

Whatever the mechanism of the gene transcription regulation is, it seems to be clear that genotype x environment dependent transcription rate of the *HSP90AA1* gene observed affects ram's sperm DNA fragmentation in such a way that both events can be linked. Differences in ram's sperm DNA fragmentation have been found.

By analyzing separately tDFI values of sperm samples which have been subjected or not to heat stress along the spermatogenesis process and the combined genotypes of both polymorphisms (g.660G>C and g.667\_668insC) concluding results, regarding the linkage between gene expression and sperm DNA fragmentation, have been found. Thus, when thermoneutral conditions surround the spermatogenesis process (sperm collected in March and May), differences of alternative genotypes of the g.667\_668insC and g.660G>C mutations analyzed separately were not enough to produce significant differences in sperm DNA fragmentation and only very light when the combined genotypes g.667\_668insC\_g.660G>C (Figure 4) were considered (less than 0.5 tDFI units). However, when heat stress conditions were present along or at some stages of the spermatogenesis process (sperm collected in June, July, August and October) different results were observed when considering the g.667\_668insC and g.660G>C genotypes separately or combined, not only in the magnitude of the differences in tDFI values observed among genotypes but also in the spermatogenesis stage where heat

stress has greater effect over sperm DNA fragmentation measured after 48h of 37°C heating after ejaculate collection.

Our results confirmed that both polymorphisms are involved in the effect that climatic conditions has over sperm DNA fragmentation. However, the critical stage in terms of heat stress effect over sperm DNA fragmentation differences moved to the spermatocytogenesis stage in where the maximum tDFI differences between genotypes were observed. These differences ranged from 1.27 to 1.32 folds for the DD<sub>-668</sub>GG<sub>-660</sub> vs. II<sub>-668</sub>CC<sub>-660</sub>, and from 1.92 to 3.56 folds for the ID<sub>-668</sub>CG<sub>-660</sub> vs. II<sub>-668</sub>CC<sub>-660</sub>. Differences in the gene expression rate for these same genotypes were also high (FC=1.6 to 3.1) when heat stress conditions (Chapter 1, August 2 and August 1 time points) were present. Therefore, it seems reasonable to consider that both events are correlated, but in what way?

It is well known that there are differences in the genes transcription rate in the stages (cell types) of the spermatogenesis process (Aguilar-Mahecha et al. 2001, Paul C 2008, Vbranovski et al. 2010) and also in the heat stress sensitivity of the different cell types involved (Rockett et al. 2001, Paul C 2008, Pérez-Crespo M 2008). In the spermatocytogenesis stage (mitosis), spermatogonia and primary spermatocytes have high transcription levels as occurs in other undifferentiated cells. However, these levels decay in meiotic cells (secondary spermatocytes and spermatids) and in mature spermatozoa. Moreover, cell-specific genes are transcribed at each stage of the spermatogenesis process (Aguilar-Mahecha et al. 2001). In rats, the expression levels of the *HSP90AA1* gene decrease drastically (80%) during the early phases of the spermatogenesis reaching undetectable levels in the more mature germ cells (Aguilar-Mahecha et al. 2001) as it has been observed for our group in sheep spermatozoa (data not shown). During germ cell development, different spermatogenic cell types showed remarkable variation in their susceptibility to heat stress being spermatogonia and spermatozoa the most thermotolerant cells while pachytene spermatocytes and early spermatids are more susceptible to heat (Setchell BP 2006, Grad et al. 2010).

Taking this background into account, we can make some hypothesis around the results here obtained regarding the expression levels observed for the *HSP90AA1* gene in animals carrying alternative combined genotypes of the g.667\_668insC-g.660G>C mutations and the variation of the tDFI values observed in the spermatozoa of these animals depending on heat stress events occurring along the spermatogenesis process. The highest differences in tDFI values among g.667\_668insC-g.660G>C combined

genotypes were observed when the THI threshold was exceeded during the spermatocytogenesis stage, independently of the heat stress events occurring in posterior phases of the spermatogenesis process. Heat stress at this stage may induce the expression of the *HSP90AA1* gene. Thus, unfavorable genotypes in terms of gene expression induction (ID<sub>-668</sub>GC<sub>-660</sub>, DD<sub>-668</sub>GG<sub>-660</sub>) do not produce enough mRNA (mRNAs are stored as messenger ribonucleoprotein particles (Hecht 1998)) and Hsp90 $\alpha$  protein to cope with future thermal stress which might occur in posterior stages in which transcriptional activity is reduced and cell types and molecular processes are more sensible to heat (spermatocytes in pachytene and spermatids protamination). When THI threshold was exceeded in the meiosis and spermiogenesis stages differences in tDFI values of alternative combined genotypes of g.667\_668insC-g.660G>C are much lower than those observed in the previous case described maybe due to the limited transcriptional activity of the cellular types here involved. Two peaks of higher differences corresponding to meiosis and protamination could indicate the importance of past (selective translation of stored mRNAs (Rathke et al. 2014)) and present (limited) expression rates of the *HSP90AA1* gene to protect the meiotic process and produce an optimal exchange of histones by protamines (Campos et al. 2010a, Rathke et al. 2014) to achieve an optimal spermatozoa DNA packaging.

Therefore, optimal expression rates of favorable genotypes of the *HSP90AA1* gene induced by heat stress events seem to be related with a higher ability of mature spermatozoa to cope with the effects that high temperatures exert over their DNA fragmentation when they are subjected to 37°C during 48h. This ability must consist essentially in a better packaging of the sperm DNA (efficient protamination) during the spermiogenesis process which would be favored by higher amounts of Hsp90 $\alpha$ , translated at this moment or stored in the past. In bulls, the DNA fragmentation index (DFI), has been positively correlated with the percentage of spermatozoa that showed low protamine content (Fortes et al. 2014). However, other roles of the Hsp90 $\alpha$  related with the cellular defense against other sources of stress (i.e. oxidative stress) and proteostasis maintenance (Erlejman et al. 2014) must not be discarded to preserve spermatozoa DNA from injuries.

Results here obtained lead us to question if heat stress events occurring at initial stages of the spermatogenesis process would be selectively advantageous to protect cell types of subsequent stages, which have worse heat stress response in terms of transcription ability, from injuries caused by heat or other sources of stress. Relative to this idea, it is important to remark that sheep is a short day breeder whose

favorable reproductive period begins when the days shorten (fewer hours of light). This period comes after the hottest months, and so, we could expect that those animals with a favorable genotype in terms of heat resistance were more fertile.

# **General discussion**



The present thesis was initially designed based on the results Marcos-Carcavilla and coworkers published in 2010 (Marcos-Carcavilla et al. 2010b). They showed evidences of the role of a particular polymorphism located at the *HSP90AA1* promoter region in gene expression rate differences under heat stress. However, their study was performed with a limited number of animals and only focused on a single polymorphism.

The initial objective of the study here presented was to verify the role of that candidate polymorphism described by Marcos-Carcavilla et al. (Marcos-Carcavilla et al. 2010b) and to investigate the potential effect of all the polymorphisms detected at the promoter region of these gene. Based on the results obtained new objectives were developed.

As a summary, in this work, it has been carried out a thorough study of the *HSP90AA1* ovine gene: the structure, regulation motifs, polymorphisms and epigenetic marks of its promoter region; the role of several polymorphisms in the gene expression pattern under variable environmental conditions; the correlation of those polymorphism frequencies with the adaptation of several sheep breeds to the regions where they are reared; the mutations evolution history across different species pertaining to the Caprinae subfamily; and finally their effect over a male reproductive trait, the rams sperm DNA fragmentation.

To a large extent, this study was carried out and focused in the Manchega sheep breed. This breed is located in Castilla-La Mancha where summer temperatures exceed 33°C ([www.aemet.es](http://www.aemet.es)). Therefore, this breed constitutes a good animal model to study environmental adaptation to heat stress conditions. The animals available for the experiments were all rams belonging to an artificial insemination centre. The fact that these are commercial animals led some limitations in the time where samples could be collected across the different experiments and in the type of samples collected. Rams were not always available because of their own insemination calendar, which in few occasions coincided with the hottest day. For instance, to accomplish the experiment described in Chapter I the sampling day was critical. Also, only samples from blood and sperm were available for the study.

In Chapter I, we accomplished the main objective of the thesis which was to investigate if polymorphisms at the *HSP90AA1* gene promoter region affect the expression rate of the gene depending on environmental circumstances. One important constraint that we have found when selecting housekeeping genes to



normalize expression results of our target (*HSP90AA1*) was the fact that only one candidate (*HSP90AB1*) among the 16 studied was more stable than the target (Serrano et al. 2011). As mentioned in the introduction, under heat stress events, the HSR mechanism alters the expression of many genes including those not directly related with heat stress (Weber et al. 2006) (See Fig. 5 from Introduction). Therefore, the normalization of the expression data of the *HSP90AA1* gene was made exclusively with one housekeeping, the *HSP90AB1* gene.

The first result related to this objective was that the linkage disequilibrium block constituted by the g.703\_704insAA, g.660G>C and g.528A>T polymorphisms has some effect in the expression profile of the gene (FC (Fold Change) = 1.3) independently of climatic conditions under which samples were collected. Due to the high LD existent among these polymorphisms, we could not determine which of them was the causal mutation. For the isolated SNPs g.522A>G and g.444A>G only subtle significant effects on expression were detected under thermoneutral and heat stress conditions, respectively.

In a second study one INDEL, g.667\_668insC, moderately linked ( $r^2 = 0.23$ ) with the LD block constituted by g.703\_704insAA, g.660G>C and g.528A>T showed the highest effect over the gene expression rate (FC=3.4) but only for samples collected under heat stress conditions. However, we could not determine if the effect was due exclusively to this mutation or to polymorphisms constituting the LD block mentioned above.

Our main hypotheses about how these polymorphisms could affect the transcription differences observed in the *HSP90AA1* gene were 1) polymorphisms affect transcription factor binding sites; 2) some epigenetic modifications (i.e. the methylation of a CpG produced in the DNA sequence when the G<sub>.660</sub> allele is present) could block the bind of a transcription factor or promote the bind of a repressor; 3) polymorphisms may be implied in some structural DNA features i.e G-quadruplex.

In order to deal with the two first questions (i.e. to isolate the causal mutation and determine the way of action), we carried out two studies that are described in Chapter 2: i) an Electrophoretic mobility shift assay (EMSA) and Luciferase reporter assay; ii) a study of the methylation pattern of the *HSP90AA1* gene promoter.

The EMSA assay, but mainly the luciferase assay, let us conclude that the combined genotype constituted by g.667\_668insC and g.660G>C seems to be the

main responsible of the differences in *HSP90AA1* transcription induction by heat stress. EMSA assay did not reveal any clear effect of the methylation effect regarding its influence over transcription factor binding differences. Unfortunately, we could not develop a luciferase assay to determine with accuracy if the G<sub>.660</sub> allele which creates a cis-regulated ASM has any influence over gene expression levels. Although with these type of techniques we cannot determine which transcription factor binds the sequence affected by these two mutations, *in silico* predictions indicates that a transcription factor of the SP family, probably SpI, binds DNA with more affinity when animals carrying the II<sub>.668</sub>CC<sub>.660</sub> genotype are subjected to hot environments.

Despite all these experiments provide us a high level of certainty about the *HSP90AA1* expression regulation, we cannot discard the existence of other polymorphisms or genes affecting the transcription rate of this gene. We are aware this gene is part of a complex interaction network that is involved not only in the stress response but also in numerous cellular processes such as the cell cycle, signal transduction, protein modification and folding, etc. In addition, the eukaryotic Hsp90 is a complex molecular chaperone machine associated with a large cohort of co-chaperones and therefore the appropriate working depends on the optimum equilibrium among all those components.

All these experiments provide us with a high level of certainty the mechanisms implied in the *HSP90AA1* expression regulation. However, regarding the third question, i.e. the presence of a group of adjacent cytosines promoting a structural change in the DNA conformation could not be addressed. In fact, the reverse chain was predicted to form a G-quadruplex. The G-quadruplex is an alternative DNA structural motif that is considered to be functionally important in the transcriptional regulation in the mammalian genome (Henderson et al. 2014). Most published studies have detected these structures by crystallography (Campbell et al. 2007), circular dichroism (Henderson et al. 2014), nuclear magnetic resonance (Adrian et al. 2012), X-ray crystal diffraction spectrometer (Phillips et al. 1997) and *in vivo* detection of G-quadruplex DNA, mainly based on the highly selective fluorescent (Lipps et al. 2009). Those probes need a complex operation procedure and sample preparation. Unfortunately, those requirements were not available in our case.

*HSP90AA1* has a complex and highly regulated promoter organization. Its structure is considered hybrid because it has a double mechanism for transcription regulation. Maybe because it is a highly induced gene, its transcription regulation has to

be specific to prevent much of the expression when it is not really required, but a basal regulated expression remains active by the CpG island to maintain cellular homeostasis. During Chapter 2, the whole promoter structure was deeply studied discovering a regulatory CpG island. Interesting differences in the methylation pattern between tissues were observed. Moreover, comparisons between the same tissue across different ages has also been studied. Sexual organs have different expression patterns depending on their sexual maturation stage (Laiho et al. 2013). In testes a specific splicing is produced yielding a much smaller isoform of the protein (Fig. 11 from Chapter 2), than in any other tissue studied in this work. Moreover, epigenetic marks existing in the *HSP90AA1* promoter in many tissues analyzed have been erased in the spermatozoa gene promoter. Epigenetic marks at the promoter region have disappeared in sperm, why is this reason? Transcription in sperm is low, thus the regulatory CGI would not have a real function. Why epigenetic marks exist in the body gene of the mature spermatozoa and do not in the body gene of other tissues analyzed? This fact could be to achieve a specific splicing pattern to obtain a smaller protein isoform exclusive to spermatozoa.

Finally, there were no differences in the methylation pattern between samples collected under heat stress and non-heat stress conditions, so environmental changes do not seem to be the reason why these methylations occur. However, the last reason why these epigenetic modifications exist has not been found. There are still some epigenetic changes that appear to happen without any known cause (Aguilera et al. 2010).

The experiments that constitute Chapters 1 and 2 were carried out almost exclusively with animals of the Spanish Manchega sheep breed. We then extend these experiments to other breeds of the ovine species and to related species that belong to the same taxa; i.e., the Caprinae subfamily.

Chapter 3 deals with the characterization of the *HSP90AA1* gene promoter in 31 sheep breeds distributed across different climatic regions of the world and in 9 species from the Caprinae subfamily. The main aims of this work were to: 1) assess the impact of the domestication process (artificial selection) and the environmental conditions (natural selection) over the frequencies of the *HSP90AA1* polymorphisms in the sheep breeds analyzed; 2) look for some footprint of these mutations in wild species of the Caprinae subfamily to which the *Ovis aries* belongs.

For the first aim addressed, the analysis of sheep breeds data revealed that the wild alleles of the g.667\_668insC and g.660G>C are D<sub>-668</sub> (72%) and C<sub>-660</sub> (63%), respectively. For g.667\_668insC, the favourable allele is that with the lowest frequency, indicating that its origin would be recent. For g.660G>C, the most frequent allele was the favourable one, suggesting a more ancient presence of this mutation in the genome of the ovine species.

Also, we found a highly significant association between the frequencies of these two polymorphisms with some climatic variables (minimum temperature, average annual temperature, thermal width, maximum annual rainfall, etc) from regions where sheep breeds are reared. The domestication process leads to manage and to select animals based on their productive performances. Domesticated animals do not live in completely free environments such as those where their wild counterparts live. By this way, livestock have different environmental selective pressure than wild animals. However, as mentioned above in the Introduction section, sheep is a livestock species that is managed under more extensive systems than others. Therefore, some footprints of natural selection over the gene investigated could be detected in this species. In fact, 35.5% of the sheep breeds studied are reared in hot climates and showed high frequencies of the alleles conferring an efficient heat stress response. Also, 19.4% of breeds reared in cold climates have low frequencies of heat stress favourable alleles. Thus, we can hypothesize that natural selection is acting over the *HSP90AA1* sheep gene despite the domestication that has been performed for this breed. But also, we found a group of breeds (25.8%) reared in cold climates with high frequencies of heat stress alleles and a group of breeds (19.3%) that came from hot climates with low frequencies of hot favourable alleles. This may reflect the well known phenomenon of genetic exchange that has occurred during the development of modern sheep breeds. In the first group, the biological fitness may not be compromised by hot environments, but in the second group unadaptation of animals to cope the climatic conditions of their places of origin is expected. We do not know the history of most of these breeds in terms of how and where they were originated or the migrations to which they were subjected. Some of these breeds are local and not subjected to artificial selection and some are even endangered. Despite this constraint, some evidence of the genetic adaptation of this species to the environmental conditions prevailing in the regions of the world where they live has been achieved.

For the second aim addressed in Chapter 3, we have confirmed that 9 of the 11 polymorphisms detected in the ovine *HSP90AA1* promoter are not exclusive of this species but they are present in several species of the Caprinae subfamily. However, any of them exist in the Bos species. We have assessed that the wild allele of the SNP g.660G>C is C since is the one with the highest frequency in all species studied. For the g.667\_668indC INDEL, results are more confuse since in *O. aries* the most frequent allele is D<sub>.668</sub>, however intermediate frequencies of alleles D<sub>.668</sub> and I<sub>.668</sub>, were found in *O. moschatus* and *O. musimon*. This led us to wonder if this mutation is not as new as we expected from the results obtained in the ovine species and to consider a new hypothesis. The I<sub>.668</sub> allele of the g.667\_668insC polymorphism could be more frequent in wild species of the Caprinae subfamily than in its domestic counterparts *O. aries* and *C. hircus*, because of the relaxation of natural selection resulting in the domestication process over this gene.

Surprisingly, we have found that within the *Ovis* genus, *O. moschatus* is the species sharing the highest number of polymorphisms with *O. aries* (7/11). For *O. moschatus*, which is currently restricted to Greenland and the Arctic Archipelago, very high frequencies of alleles related with the heat stress response (I<sub>.668</sub> = 0.43; C<sub>.660</sub> = 0.90) were found. These results seem to indicate that the actual *O. moschatus* is genetically well prepared to tolerate warm climates. Could the high frequencies of alleles related with the heat stress response found in *O. moschatus* be inherited from an extinct ancestor of this species? *Praeovibos*, an older morphotype of *O. moschatus*, does not appear to have been restricted to inhabiting cold climates as its remains have also been found in temperate and Mediterranean forests of France and Spain which indicates that *Praeovibos* was less restricted to a specific ecological niche than *O. moschatus* (Campos et al. 2010b). Furthermore, an opposite situation was found in *A. lervia*. This species came from the Sahara and the Magreb rocky areas and is well adapted to high temperatures and dehydration. However, the unfavourable allele (D<sub>.668</sub>) for heat stress response is fixed in this species.

Genetic divergence measures estimated locate *A. lervia* closer to species of the *Capra* genus (*C. hircus* and *C. pyrenaica*) and more distant to *Rupicapra pyrenaica* than that pointed out in a recent reconstructions of the Caprinae phylogenies (Hassanin et al. 2012; Bibi et al. 2013). *O. moschatus* was located closer to species from the *Ovis* genus and to *R. pyrenaica*. This reveals the complexity of the interpretation of the evolution of the species belonging to the Caprinae subfamily.

Finally, we investigated the effect of the *HSP90AA1* genotypes of those polymorphisms related with the heat stress response on the phenotypic variability observed depending on climatic conditions of a ram's reproductive trait, the sperm DNA fragmentation (Chapter 4). The testes of many species of mammals descend either during fetal or shortly after birth into a scrotum, where the temperature is appreciably lower than in the abdomen. The body-scrotal temperature difference is greater in an environment of 6°C and less at 40°C. This special protection of male gametes is due to the fact that Pachytene spermatocytes and early spermatids are the most susceptible cells to heat stress in the testis. The effects of heat stress over sperm cells are, among others, apoptosis and DNA damage. In addition, RNAseq studies of the *HSP90AA1* expression levels in different ovine tissues have revealed that testis and brain are the tissues with the highest transcriptional activity of the *HSP90AA1* gene.

All these arguments lead us to choose the sperm DNA fragmentation as the target phenotype to be associated with genotype x temperature transcription differences observed in the *HSP90AA1* gene. To mimic the conditions that sperm cells find during its transit along the ewes oviduct, ejaculates of rams, collected under different environmental temperatures were incubated at 37°C during 24 and 48 hours. Sperm DNA fragmentation was assessed through the Sperm Chromatin Structure Assay (SCSA) method by flow cytometry.

Our results confirmed the relationship between the combined genotype of g.667\_668insC and g.660G>C with sperm DNA fragmentation differences. The critical stage in terms of heat stress effect over sperm DNA fragmentation seems to be the spermatocytogenesis stage (35 to 50 days bsc) in where the highest differences between genotypes were observed. At this stage, spermatogonia and primary spermatocytes have high transcription levels as occur in other undifferentiated cells. However these levels decay in meiotic cells (secondary spermatocytes and spermatids) and in mature spermatozoa. The expression levels of the *HSP90AA1* gene decrease drastically (80%) during the early phases of the spermatogenesis reaching undetectable levels in the more mature germ cells.

Therefore, we can argue that spermatocytogenesis is the most heat-susceptible stage of the spermatogenesis process in terms of sperm DNA fragmentation, not due to thermosensitivity of cells at this phase but because the effect of the differences in the transcriptional activity of the gene have over subsequent stages of the spermatogenesis process. Thus, unfavourable genotypes in terms of gene expression

induction do not produce enough mRNA (mRNAs are stored as messenger ribonucleoprotein particles) and Hsp90 $\alpha$  protein to cope with future thermal stress which might occur in posterior stages in which transcriptional activity is reduced and cell types and molecular processes are more sensible to heat (spermatocytes in pachytene and spermatids protamination). This ability must consist essentially in an optimal development of the meiotic process and a better packaging of sperm DNA (efficient protamination) during the spermiogenesis process which would be favoured by higher amounts of translated Hsp90 $\alpha$  at this moment or stored in the past (spermatocytogenesis).

After globally analyzing the results obtained along the present thesis, we can conclude that the approaches here used have been useful to better understand the molecular mechanisms that undergo some aspects of the heat stress response and the consequences that sub optimal genotypes have over a phenotypic trait. In fact, new possible approaches have been opened, which this work could not still give answer but, sometimes, that's the beauty of science.

## **FUTURE PERSPECTIVES**

### **1. New technologies**

Whole genome sequencing and SNP chips genotyping have opened new possibilities in obtaining rich data base to continue studying these kind of processes. Increasing the number of animals and markers analyzed will provide to the scientific community a great amount of information: the so-called '*omics*' revolution. For instance performing association studies to identify functional polymorphisms, insertions, long non-coding RNAs or CNVs in the whole genome is now possible with an affordable estimate. The bioinformatics area has achieved great importance, as they have turned an essential tool to decipher the great amount of information obtained.

### **2. Climate change**

Large emissions of greenhouse gases has led to global warming and climate change (Lal 2013). Along this thesis we have studied some aspects of the ovine species adaptation mechanism to cope with increase temperatures, as an example of a good species to cope with climate change consequences. At the same time, ruminant livestock such as cattle, buffalo, sheep and goats contributes the major proportion of

total agricultural emission of methane, even though, cattle are the largest contributing species to enteric fermentation (Figure 1) increasing green house gases (GHG) and contributing to global warming .

However, comparing the global estimates from livestock production by species, small ruminants produce lower greenhouse emissions than even monogastric species. It is true, that global small ruminant production is smaller than any other livestock species (Figure 2), however in those regions where small ruminants production is high, as South Asia, East and Southeast Asia and Near East and North Africa regions, GHG (greenhouse gas) emissions continue being lower (Figure 2).

Increasing sustainable small ruminant production, where sheep is included, could be a good approach for obtaining high quality products with less GHG emissions and better adapted to worse environmental conditions. For developing countries investing in small ruminant production could lead to cheaper food and clothes resources. The greatest advantage of small ruminants relative to large ruminants is their low cost, small size, their suitability to small holdings and its efficient conversion of forage feeds even they are farmed in temperate, arid or semi-tropical conditions. In fact, one of FAO's active commitments is related with small ruminants in developing countries (Timon 1985).

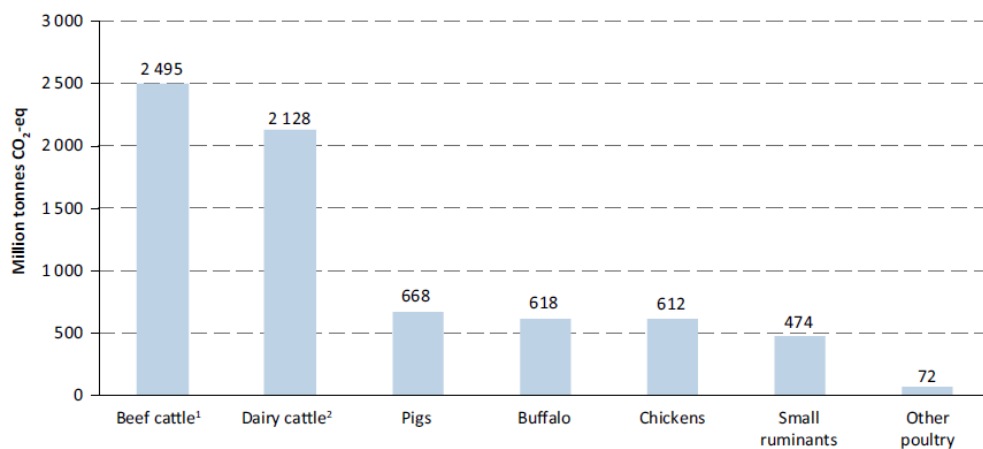
"In view of the very significant contribution of small ruminants to the economy and livelihood of peoples in almost every country around the world, and particularly in the developing countries, the Consultation strongly recommends that much greater priority and much larger investment should be made by national and international institutions in R&D and the promotion of small ruminant production."

### **3. New approaches**

One of the most important results achieved in this work has been the correlation established between differences in *HSP90AA1* expression levels depending on genotype and climatic variables existent when samples were collected with the rate of sperm DNA fragmentation of males from which these samples were obtained. In the near future our aim is to determine the impact of such variations in sperm DNA fragmentation observed over male's fertility. However this is not an easy question since the trait fertility measured as pregnancy or lambing rate is responsibility of both the dam and the sire.



Figure 1. Global estimates of emissions by species, including emissions attributed to edible products and to other goods and services, such as draught power and wool. Source: FAO's data (Gerber 2013)



<sup>1</sup> Producing meat and non-edible outputs.

<sup>2</sup> Producing milk and meat as well as non-edible outputs.

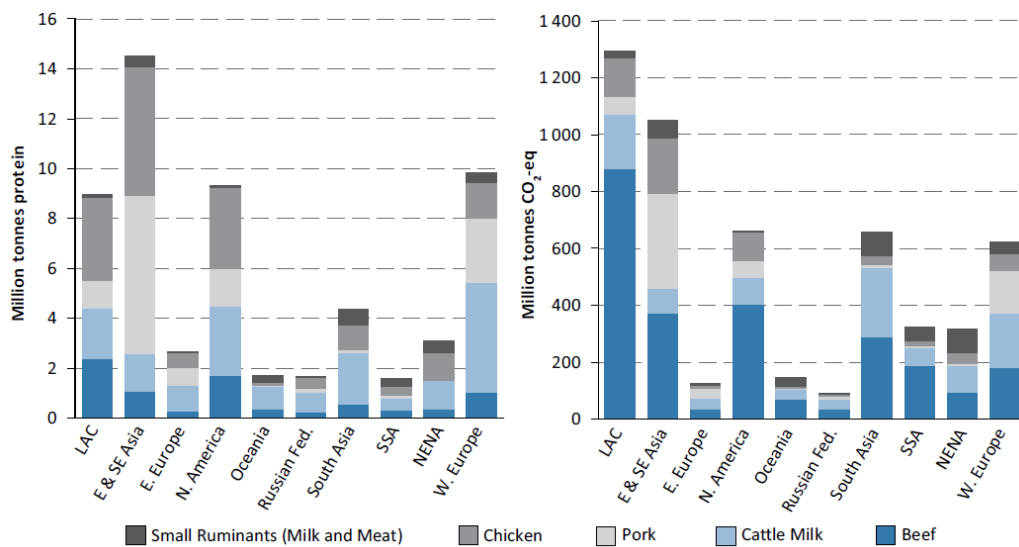
In livestock, the success of a gestation or birth has been usually related with the dam. Male's effect over pregnancy rate is not easily measurable, because the small period of time from mating in which pregnancy can be assigned to the male and the lack of accuracy technologies to make early gestation diagnosis. The availability of new approaches to measure pregnancy in sheep such as near-infrared reflectance spectroscopy (NIRS), opens up the possibility to find the way to measure the male's effect on the pregnancy rate (Andueza et al. 2014) and therefore to associate levels of sperm DNA fragmentation depending on genotype and climatic variables with males fertility.

Sperm possess a remarkable set of long-lived messenger RNAs that have been hypothesized to be required for embryogenesis. At least in human, the sperm transmits not only nuclear DNA to the oocyte but also activation factors, centrosomes, and a host of mRNAs and microRNAs during fertilization. It has been observed that several of these spermatozoal mRNAs are also found in the zygote, indicating that these transcripts may be functionally critical during embryonic development (Ostermeier et al. 2004). Thus, the health of the sperm genome and epigenome is critical for improving assisted conception rates and the birth of healthy offspring (Kumar et al. 2013). The delivery of certain sperm transcripts to the oocyte, which could translate to proteins, may be critical for the early stages of embryogenesis

and/or implantation. Moreover, some transcripts are present in spermatozoa, and not detected in the oocyte, and these should be contributed exclusively by the sperm to the egg (Kumar et al. 2012).

Finally, we do not discard that *HSP90AA1* genotype from the ewe would also have effect in pregnancy and birth success. It is however known that the expression patterns of the *HSP90AA1* gene in oocytes is not as relevant as in testes (Chapter 2).

Figure 2. Global livestock production (in Tonnes of proteins) and GHG emissions from livestock by regions. Source: FAO's data (Gerber 2013)



LAC: Latin America and Caribbean region; E. & SE. Asia: east and South East Asia; E. Europe: East Europe; N. America: North America; Russian fed.: Russian federation; SSA: Sub-Saharan Africa; NENA: Near East and North Africa region; W. Europe: West Europe



# **Conclusiones**



1. La región promotora del gen ovino *HSP90AA1* es altamente polimórfica. Los polimorfismos presentes en dicha región afectan a la regulación de la expresión génica del gen.
2. El promotor del gen *HSP90AA1* ovino es de tipo híbrido, con presencia de caja TATA e isla CpG reguladora. Este hecho, le confiere mayor plasticidad en la expresión génica y dota de mayor importancia a aquellos polimorfismos situados en la región 5' del promotor.
3. La transversión g.660G>C (rs397514116) tiene un efecto de igual magnitud sobre la tasa de expresión del gen en condiciones termoneutras y de estrés térmico. Los animales que portan el alelo G tienen una menor expresión de la chaperona inducible *HSP90AA1*.
4. La presencia de una inserción de citosina g.667\_668insC (rs397514115) en la región 5' del promotor de la chaperona *HSP90AA1* aumenta hasta tres veces la expresión del gen bajo condiciones de estrés térmico.
5. Las frecuencias de los alelos de aquellas mutaciones beneficiosas frente al estrés térmico (g.660G>C y g.667\_668insC) presentan una elevada correlación con variables climáticas de las regiones de origen de algunas de las razas ovinas estudiadas.
6. El alelo C en la transversión g.660G>C resulta ser el alelo salvaje, justificada por la metilación alelo-específica que se produce en este polimorfismo y por la frecuencia de este alelo en otras razas ovinas y otras especies rumiantes estudiadas en este trabajo.
7. El alelo salvaje de la inserción g.667\_668insC podría ser la delección de una citosina en base a los resultados obtenidos en la especie ovina. Sin embargo, el análisis de otras especies de la subfamilia Caprinae no apoyan dicha hipótesis.
8. La presencia de una inserción de citosina, g.667\_668insC, aumenta la protección del ADN espermático contra la fragmentación por efecto de la temperatura.

9. La fragmentación de ADN espermático depende de la etapa de espermatogénesis en la que la célula ha estado expuesta a estrés térmico. Siendo la fase de espermatocitogénesis la más sensible a temperaturas elevadas.

# **Conclusions**





1. The promoter region of the ovine gene *HSP90AA1* is highly polymorphic. The polymorphisms present at this region have effect in the gene expression.
2. The ovine *HSP90AA1* gene promoter is a hybrid promoter. It presents TATA-box and a regulatory CpG island. This fact confers more plasticity in the gene expression profile and it provides more relevance to the polymorphisms located at the 5' promoter region.
3. The g.660G>C transversion (rs397514116) has effect in gene expression both under thermoneutral and heat stress conditions. Animals carrying the G allele have less expression of the *HSP90AA1* inducible chaperone gene.
4. The presence of one cytosine insertion, g.667\_668insC (rs397514115.2) located at the 5' *HSP90AA1* promoter region increases up to three times the expression under heat stress conditions.
5. The frequency of both beneficial mutations against heat stress (g.660G>C y g.667\_668insC) are related with hotter climatic latitudes where most of the ovine breeds studied are reared.
6. The C allele in the g.660 G>C transversion results to be the wild type allele. It has been proved by the allele-specific methylation pattern produced in this polymorphism and the frequency and presence of both alleles in different ovine breeds and ruminant species studied in this work.
7. The wild allele of the cytosine insertion g.667\_668insC can be the deletion allele, based on the frequencies observed in the sheep breeds studied. However, not conclusive results can be obtained based on the analysis of more species from the Caprinae subfamily.
8. The presence of the g.667\_668insC cytosine insertion increases the sperm DNA protection against heat stress effects.
9. Sperm DNA fragmentation depends on the spermatogenesis stage where the cell has been exposed to heat. The most sensitive stage to high temperatures exposition is during spermatogenesis.



# References

- (NOAA), N. O. a. A. A. (2011). 2011 NOAA Satellite and Information Service. Annual report. online.
- Adrian, M., B. Heddi and A. T. Phan (2012). "NMR spectroscopy of G-quadruplexes." Methods **57**(1): 11-24.
- Aguilar-Mahecha, A., B. F. Hales and B. Robaire (2001). "Expression of stress response genes in germ cells during spermatogenesis." Biol Reprod **65**(1): 119-127.
- Aguilar, I., I. Misztal and S. Tsuruta (2009). "Genetic components of heat stress for dairy cattle with multiple lactations." Journal of Dairy Science **92**(11): 5702-5711.
- Aguilera, O., A. F. Fernandez, A. Munoz and M. F. Fraga (2010). "Epigenetics and environment: a complex relationship." J Appl Physiol (1985) **109**(1): 243-251.
- Alekseev, O. M., D. C. Bencic, R. T. Richardson, E. E. Widgren and M. G. O'Rand (2003). "Overexpression of the Linker histone-binding protein tNASP affects progression through the cell cycle." J Biol Chem **278**(10): 8846-8852.
- Alekseev, O. M., E. E. Widgren, R. T. Richardson and M. G. O'Rand (2005). "Association of NASP with HSP90 in mouse spermatogenic cells: stimulation of ATPase activity and transport of linker histones into nuclei." J Biol Chem **280**(4): 2904-2911.
- Alexandre, A., M. Laranjo and S. Oliveira (2014). "Global transcriptional response to heat shock of the legume symbiont *Mesorhizobium loti* MAFF303099 comprises extensive gene downregulation." DNA Res **21**(2): 195-206.
- Amann, R. P. and J. M. DeJarnette (2012). "Impact of genomic selection of AI dairy sires on their likely utilization and methods to estimate fertility: a paradigm shift." Theriogenology **77**(5): 795-817.
- Andueza, D., J. L. Alabart, B. Lahoz, F. Munoz and J. Folch (2014). "Early pregnancy diagnosis in sheep using near-infrared spectroscopy on blood plasma." Theriogenology **81**(3): 509-513.
- Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira and G. Luikart (2008). "LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method." BMC Bioinformatics **9**: 323.
- Arad, Z., T. Mizrahi, S. Goldenberg and J. Heller (2010). "Natural annual cycle of heat shock protein expression in land snails: desert versus Mediterranean species of *Sphincterochila*." J Exp Biol **213**(Pt 20): 3487-3495.
- Araujo, P. R., K. Yoon, D. Ko, A. D. Smith, M. Qiao, U. Suresh, S. C. Burns and L. O. Penalva (2012). "Before It Gets Started: Regulating Translation at the 5' UTR." Comp Funct Genomics **2012**: 475731.
- Arceo, M. E., C. W. Ernst, J. K. Lunney, I. Choi, N. E. Raney, T. Huang, C. K. Tuggle, R. R. R. Rowland and J. P. Steibel (2012). "Characterizing differential individual response to porcine reproductive and respiratory syndrome virus infection through statistical and functional analysis of gene expression." Frontiers in genetics **3**.
- Auer, P. L. and R. W. Doerge (2010). "Statistical Design and Analysis of RNA Sequencing Data." Genetics **185**(2): 405-U432.
- Babu, M. M., N. M. Luscombe, L. Aravind, M. Gerstein and S. A. Teichmann (2004). "Structure and evolution of transcriptional regulatory networks." Curr Opin Struct Biol **14**(3): 283-291.
- Ballachey, B. E., W. D. Hohenboken and D. P. Evenson (1987). "Heterogeneity of sperm nuclear chromatin structure and its relationship to bull fertility." Biol Reprod **36**(4): 915-925.

- Banks S, K. S., Irvine DS, Saunders PTK (2005). "Impact of a mild scrotal heat stress on DNA integrity in murine spermatozoa." Reproduction **129**.
- Barnosky, A. D., P. L. Koch, R. S. Feranec, S. L. Wing and A. B. Shabel (2004). "Assessing the causes of late Pleistocene extinctions on the continents." Science **306**(5693): 70-75.
- Barratt, C. L., R. J. Aitken, L. Bjorndahl, D. T. Carrell, P. de Boer, U. Kvist, S. E. Lewis, S. D. Perreault, M. J. Perry, L. Ramos, B. Robaire, S. Ward and A. Zini (2010). "Sperm DNA: organization, protection and vulnerability: from basic science to clinical applications--a position report." Hum Reprod **25**(4): 824-838.
- Barreau, C., E. Benson, E. Gudmannsdottir, F. Newton and H. White-Cooper (2008). "Post-meiotic transcription in *Drosophila* testes." Development **135**(11): 1897-1902.
- Basehoar, A. D., S. J. Zanton and B. F. Pugh (2004). "Identification and distinct regulation of yeast TATA box-containing genes." Cell **116**(5): 699-709.
- Basu, N., A. E. Todgham, P. A. Ackerman, M. R. Bibeau, K. Nakano, P. M. Schulte and G. K. Iwama (2002). "Heat shock protein genes and their functional significance in fish." Gene **295**(2): 173-183.
- Beaumont, M. A. and D. J. Balding (2004). "Identifying adaptive genetic divergence among populations from genome scans." Mol Ecol **13**(4): 969-980.
- Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-412.
- Bibi, F. (2013). "A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics." BMC Evol Biol **13**: 166.
- Bikle, D. D. (2010). "Vitamin D: newly discovered actions require reconsideration of physiologic requirements." Trends in Endocrinology and Metabolism **21**(6): 375-384.
- Black, D. L. (2000). "Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology." Cell **103**(3): 367-370.
- Blau, J., H. Xiao, S. McCracken, P. Ohare, J. Greenblatt and D. Bentley (1996). "Three functional classes of transcriptional activation domains." Molecular and Cellular Biology **16**(5): 2044-2055.
- Bock, C., M. Paulsen, S. Tierling, T. Mikeska, T. Lengauer and J. Walter (2006). "CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure." PLoS Genet **2**(3): e26.
- Boerke, A., S. J. Dieleman and B. M. Gadella (2007). "A possible role for sperm RNA in early embryo development." Theriogenology **68 Suppl 1**: S147-155.
- Bond, U. (1988). "Heat shock but not other stress inducers leads to the disruption of a sub-set of snRNPs and inhibition of in vitro splicing in HeLa cells." EMBO J **7**(11): 3509-3518.
- Borkovich, K. A., F. W. Farrelly, D. B. Finkelstein, J. Taulien and S. Lindquist (1989). "hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures." Mol Cell Biol **9**(9): 3919-3930.
- Boyazoglu, J. and P. Morand-Fehr (2001). "Mediterranean dairy sheep and goat products and their quality. A critical review." Small Rumin Res **40**(1): 1-11.
- Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-254.

- Braidwood, R. J. (1960). The Agricultural Revolution.
- Brown, M. A., L. Zhu, C. Schmidt and P. W. Tucker (2007). "Hsp90--from signal transduction to cell transformation." Biochem Biophys Res Commun **363**(2): 241-246.
- Bungum, M., P. Humaidan, M. Spano, K. Jepson, L. Bungum and A. Giwercman (2004). "The predictive value of sperm chromatin structure assay (SCSA) parameters for the outcome of intrauterine insemination, IVF and ICSI." Hum Reprod **19**(6): 1401-1408.
- Bustin, S. A., J. Vandesompele and M. W. Pfaffl (2009). "Standardization of qPCR and RT-qPCR." Genetic Engineering & Biotechnology News **29**(14): 40-43.
- Butler, J. E. and J. T. Kadonaga (2002). "The RNA polymerase II core promoter: a key component in the regulation of gene expression." Genes Dev **16**(20): 2583-2592.
- Campbell, N. H. and G. N. Parkinson (2007). "Crystallographic studies of quadruplex nucleic acids." Methods **43**(4): 252-263.
- Campos, E. I., J. Fillingham, G. Li, H. Zheng, P. Voigt, W. H. Kuo, H. Seepany, Z. Gao, L. A. Day, J. F. Greenblatt and D. Reinberg (2010a). "The program for processing newly synthesized histones H3.1 and H4." Nat Struct Mol Biol **17**(11): 1343-1351.
- Campos, P. F., A. Sher, J. I. Mead, A. Tikhonov, M. Buckley, M. Collins, E. Willerslev and M. T. P. Gilbert (2010b). "Clarification of the taxonomic relationship of the extant and extinct ovibovids, *Ovibos*, *Praeovibos*, *Euceratherium* and *Bootherium*." Quaternary Science Reviews **29**(17-18): 2123-2130.
- Carninci, P., A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A. Semple, M. S. Taylor, P. G. Engstrom, M. C. Frith, A. R. Forrest, W. B. Alkema, S. L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S. M. Grimmond, C. A. Wells, V. Orlando, C. Wahlestedt, E. T. Liu, M. Harbers, J. Kawai, V. B. Bajic, D. A. Hume and Y. Hayashizaki (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." Nat Genet **38**(6): 626-635.
- Carretero-Paulet, L., V. A. Albert and M. A. Fares (2013). "Molecular evolutionary mechanisms driving functional diversification of the HSP90A family of heat shock proteins in eukaryotes." Mol Biol Evol **30**(9): 2035-2043.
- Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.
- Collier, R. J., J. L. Collier, R. P. Rhoads and L. H. Baumgard (2008). "Invited review: Genes involved in the bovine heat stress response." Journal of Dairy Science **91**(2): 445-454.
- Conwell, C. C., I. D. Vilfan and N. V. Hud (2003). "Controlling the size of nanoscale toroidal DNA condensates with static curvature and ionic strength." Proc Natl Acad Sci U S A **100**(16): 9296-9301.
- Coop, G., D. Witonsky, A. Di Rienzo and J. K. Pritchard (2010). "Using environmental correlations to identify loci underlying local adaptation." Genetics **185**(4): 1411-1423.
- Coulondre, C., J. H. Miller, P. J. Farabaugh and W. Gilbert (1978). "Molecular basis of base substitution hotspots in *Escherichia coli*." Nature **274**(5673): 775-780.
- Courot, M. and R. Ortavant (1981). "Endocrine control of spermatogenesis in the ram." J Reprod Fertil Suppl **30**: 47-60.

- Cregut-Bonnoure, E. (1984). The Pleistocene Ovibovinae of Western Europe: Temporo-spatial Expansion and Paleoeological Implications. Biological Papers of the University of Alaska.
- Crevel, G., H. Bates, H. Huikeshoven and S. Cotterill (2001). "The Drosophila Dpit47 protein is a nuclear Hsp90 co-chaperone that interacts with DNA polymerase alpha." J Cell Sci **114**(Pt 11): 2015-2025.
- Csermely, P., T. Schnaider, C. Soti, Z. Prohaszka and G. Nardai (1998). "The 90-kDa molecular chaperone family: structure, function, and clinical applications. A comprehensive review." Pharmacology & Therapeutics **79**(2): 129-168.
- Curtis, S. E. (1983). Environmental management in animal agriculture. Ames, Iowa 50010, Iowa State University Press.
- Charoensook, R., K. Gatphayak, A. R. Sharifi, C. Chaisongkram, B. Brenig and C. Knorr (2012). "Polymorphisms in the bovine HSP90ABI gene are associated with heat tolerance in Thai indigenous cattle." Tropical Animal Health and Production **44**(4): 921-928.
- Chen, B., W. H. Piel, L. Gui, E. Bruford and A. Monteiro (2005). "The HSP90 family of genes in the human genome: insights into their divergence and evolution." Genomics **86**(6): 627-637.
- Chen, B., D. B. Zhong and A. Monteiro (2006). "Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms." BMC Genomics **7**: 156.
- Chessa, B., F. Pereira, F. Arnaud, A. Amorim, F. Goyache, I. Mainland, R. R. Kao, J. M. Pemberton, D. Beraldi, M. J. Stear, A. Alberti, M. Pittau, L. Iannuzzi, M. H. Banabazi, R. R. Kazwala, Y. P. Zhang, J. J. Arranz, B. A. Ali, Z. Wang, M. Uzun, M. M. Dione, I. Olsaker, L. E. Holm, U. Saarma, S. Ahmad, N. Marzanov, E. Eythorsdottir, M. J. Holland, P. Ajmone-Marsan, M. W. Bruford, J. Kantanen, T. E. Spencer and M. Palmarini (2009). "Revealing the history of sheep domestication using retrovirus integrations." Science **324**(5926): 532-536.
- Chong, I.-G. and C.-H. Jun (2005). "Performance of some variable selection methods when multicollinearity is present." Chemometrics and Intelligent Laboratory Systems **78**(1-2): 103-112.
- Christine Fast D.V.M., M. H. G. D. V. M. (2013). Classical and Atypical Scrapie in Sheep and Goats. Prions and Diseases. W.-Q. Zou and P. Gambetti, Springer New York: 15-44.
- Dadoue, J. P. (2009). "Spermatozoal RNAs: what about their functions?" Microsc Res Tech **72**(8): 536-551.
- David Victor, D. Z. (2014). Introductory chapter. Climate Change 2014: Mitigation of Climate Change. online.
- David W. Hosmer, S. L. (2000). Applied Logistic Regression, John Wiley & Sons, Inc.
- de Carcer, G. (2004). "Heat shock protein 90 regulates the metaphase-anaphase transition in a polo-like kinase-dependent manner." Cancer Res **64**(15): 5106-5112.
- Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." Genes Dev **25**(10): 1010-1022.
- Deuerling, E., H. Patzelt, S. Vorderwulbecke, T. Rauch, G. Kramer, E. Schaffitzel, A. Mogk, A. Schulze-Specking, H. Langen and B. Bukau (2003). "Trigger Factor and DnaK possess overlapping substrate pools and binding specificities." Molecular Microbiology **47**(5): 1317-1328.



- Dheda, K., J. F. Huggett, J. S. Chang, L. U. Kim, S. A. Bustin, M. A. Johnson, G. A. W. Rook and A. Zumla (2005). "The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization." Analytical Biochemistry **344**(1): 141-143.
- Diaz, C., Z. G. Vitezica, R. Rupp, O. Andreoletti and J. M. Elsen (2005). "Polygenic variation and transmission factors involved in the resistance/susceptibility to scrapie in a Romanov flock." J Gen Virol **86**(Pt 3): 849-857.
- Didion, B. A., K. M. Kaspersen, R. L. Wixon and D. P. Evenson (2009). "Boar fertility and sperm chromatin structure status: a retrospective report." J Androl **30**(6): 655-660.
- Diller KR (2006). "Stress protein expression kinetics." Annu Rev Biomed Eng **8**.
- Dominguez, K., C. R. Arca and W. S. Ward (2011). The Relationship Between Chromatin Structure and DNA Damage in Mammalian Spermatozoa. Sperm Chromatin. A. Zini and A. Agarwal, Springer New York: 61-68.
- Drabent, B., C. Bode, N. Miosge, R. Herken and D. Doenecke (1998). "Expression of the mouse histone gene H1t begins at premeiotic stages of spermatogenesis." Cell Tissue Res **291**(1): 127-132.
- Dubchak, I., M. Brudno, G. G. Loots, L. Pachter, C. Mayor, E. M. Rubin and K. A. Frazer (2000). "Active conservation of noncoding sequences revealed by three-way species comparisons." Genome Res **10**(9): 1304-1306.
- Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. Gonzalez-Martinez and D. B. Neale (2010). "Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae)." Genetics **185**(3): 969-982.
- Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein (1998). "Cluster analysis and display of genome-wide expression patterns." Proc Natl Acad Sci U S A **95**(25): 14863-14868.
- Endler, J. A. (1977). Geographic Variation, Speciation and Clines, Princeton University Press.
- Erlejman, A. G., M. Lagadari, J. Toneatto, G. Piwien-Pilipuk and M. D. Galigniana (2014). "Regulatory role of the 90-kDa-heat-shock protein (Hsp90) and associated factors on gene expression." Biochim Biophys Acta **1839**(2): 71-87.
- Evenson, D. and L. Jost (2000). "Sperm chromatin structure assay is useful for fertility assessment." Methods Cell Sci **22**(2-3): 169-189.
- Evenson, D. P., L. K. Jost, D. Marshall, M. J. Zinaman, E. Clegg, K. Purvis, P. de Angelis and O. P. Claussen (1999). "Utility of the sperm chromatin structure assay as a diagnostic and prognostic tool in the human fertility clinic." Hum Reprod **14**(4): 1039-1049.
- Evenson, D. P., K. L. Larson and L. K. Jost (2002). "Sperm chromatin structure assay: its clinical use for detecting sperm DNA fragmentation in male infertility and comparisons with other techniques." J Androl **23**(1): 25-43.
- Evenson, D. P. and R. Wixon (2006). "Clinical aspects of sperm DNA fragmentation detection and male infertility." Theriogenology **65**(5): 979-991.
- Excoffier, L., G. Laval and S. Schneider (2005). "Arlequin (version 3.0): an integrated software package for population genetics data analysis." Evol Bioinform Online **1**: 47-50.
- Fanger, O. (1970/1982). Thermal Comfort.

- Farre, D., R. Roset, M. Huerta, J. E. Adsuara, L. Rosello, M. M. Alba and X. Messeguer (2003). "Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN." *Nucleic Acids Research* **31**(13): 3651-3653.
- Favatie, F., L. Bornman, L. E. Hightower, E. Gunther and B. S. Polla (1997). "Variation in hsp gene expression and Hsp polymorphism: Do they contribute to differential disease susceptibility and stress tolerance?" *Cell Stress & Chaperones* **2**(3): 141-155.
- Feder, M. E. and G. E. Hofmann (1999). "Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology." *Annu Rev Physiol* **61**: 243-282.
- Feltus, F. A., E. K. Lee, J. F. Costello, C. Plass and P. M. Vertino (2003). "Predicting aberrant CpG island methylation." *Proc Natl Acad Sci U S A* **100**(21): 12253-12258.
- Finocchiaro, R., J. van Kaam, B. Portolano and I. Misztal (2005). "Effect of heat stress on production of Mediterranean dairy sheep." *Journal of Dairy Science* **88**(5): 1855-1864.
- Fischer, B. E., E. Wasbrough, L. A. Meadows, O. Randlet, S. Dorus, T. L. Karr and S. Russell (2012). "Conserved properties of Drosophila and human spermatozoal mRNA repertoires." *Proc Biol Sci* **279**(1738): 2636-2644.
- Fleige, S. and M. W. Pfaffl (2006a). "RNA integrity and the effect on the real-time qRT-PCR performance." *Molecular aspects of medicine* **27**(2-3).
- Fleige, S., V. Walf, S. Huch, C. Prgomet, J. Sehm and M. W. Pfaffl (2006b). "Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR." *Biotechnology Letters* **28**(19): 1601-1613.
- Fleming JS, Y. F., McDonald RM, Meyers SA, Montgomery GW, (2004). "Effects of scrotal heating on sperm surface protein PH-20 expression in sheep." *Mol Reprod Dev* **68**.
- Fortes, M. R., N. Satake, D. H. Corbet, N. J. Corbet, B. M. Burns, S. S. Moore and G. B. Boe-Hansen (2014). "Sperm protamine deficiency correlates with sperm DNA damage in Bos indicus bulls." *Andrology* **2**(3): 370-378.
- Freeman BC, M. A., Song J, Kampinga HH, Morimoto RI (2000). "Analysis of molecular chaperone activities using in vitro and in vivo approaches." *Methods Mol Biol* **99**.
- Fulda, S., A. M. Gorman, O. Hori and A. Samali (2010). "Cellular stress responses: cell survival and cell death." *Int J Cell Biol* **2010**: 214074.
- Fumagalli, M., M. Sironi, U. Pozzoli, A. Ferrer-Admetlla, L. Pattini and R. Nielsen (2011). "Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution." *PLoS Genet* **7**(11): e1002355.
- Gagniuc, P. and C. Ionescu-Tirgoviste (2012). "Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters." *BMC Genomics* **13**: 512.
- Gagniuc, P. and C. Ionescu-Tirgoviste (2013). "Gene promoters show chromosome-specificity and reveal chromosome territories in humans." *BMC Genomics* **14**: 278.
- Galibert, M. D., S. Carreira and C. R. Goding (2001). "The Usf-1 transcription factor is a novel target for the stress-responsive p38 kinase and mediates UV-induced Tyrosinase expression." *Embo Journal* **20**(17): 5022-5031.
- Garcia-Alvarez, O., A. Maroto-Morales, M. Ramon, E. del Olmo, V. Montoro, A. E. Dominguez-Rebolledo, A. Bisbal, P. Jimenez-Rabadan, M. D. Perez-Guzman and A. J. Soler (2010). "Analysis

of selected sperm by density gradient centrifugation might aid in the estimation of in vivo fertility of thawed ram spermatozoa." Theriogenology **74**(6): 979-988.

Garcia-Macias, V., F. Martinez-Pastor, M. Alvarez, S. Borragan, C. A. Chamorro, A. J. Soler, L. Anel and P. de Paz (2006). "Seasonal changes in sperm chromatin condensation in ram (*Ovis aries*), Iberian red deer (*Cervus elaphus hispanicus*), and brown bear (*Ursus arctos*)." J Androl **27**(6): 837-846.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein and P. O. Brown (2000). "Genomic expression programs in the response of yeast cells to environmental changes." Mol Biol Cell **11**(12): 4241-4257.

Gasperini, L. and G. Legname (2014). "Prion Protein and Aging." Frontiers in Cell and Developmental Biology **2**.

Gentry, A. W. (1992). "The subfamilies and tribes of the family Bovidae." Mammal Review **22**(1): 1-32.

Gentry, A. W. (1994). "The Miocene differentiation of old world Pecora (Mammalia)." Historical Biology **7**(2): 115-158.

Gerber, P. J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., Falcucci, A. & Tempio, G. (2013). "Tackling climate change through livestock – A global assessment of emissions and mitigation

opportunities." Food and Agriculture Organization of the United Nations (FAO), Rome.

Gonzalez-Marin, C., J. Gosalvez and R. Roy (2012). "Types, causes, detection and repair of DNA fragmentation in animal and human sperm cells." Int J Mol Sci **13**(11): 14026-14052.

Gonzalez, J., T. L. Karasov, P. W. Messer and D. A. Petrov (2010). "Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*." PLoS Genet **6**(4): e1000905.

Goudet, J., M. Raymond, T. de Meeus and F. Rousset (1996). "Testing differentiation in diploid populations." Genetics **144**(4): 1933-1940.

Grad, I., C. R. Cederroth, J. Walicki, C. Grey, S. Barluenga, N. Winssinger, B. De Massy, S. Nef and D. Picard (2010). "The molecular chaperone Hsp90alpha is required for meiotic progression of spermatocytes beyond pachytene in the mouse." PLoS One **5**(12): e15770.

Grazer, V. M. and O. Y. Martin (2012). "Investigating climate change and reproduction: experimental tools from evolutionary biology." Biology (Basel) **1**(2): 411-438.

Grivet, D., F. Sebastiani, R. Alia, T. Bataillon, S. Torre, M. Zabal-Aguirre, G. G. Vendramin and S. C. Gonzalez-Martinez (2011). "Molecular footprints of local adaptation in two Mediterranean conifers." Mol Biol Evol **28**(1): 101-116.

Groen AF, S. T., Colleau JJ, Pedersen J, Pribyl J, et al. (1997). "Economic values in dairy cattle breeding, with special reference to functional traits. Report of an EAAP-working group." Livestock Production Science **Vol. 49**(Issue 1): Pages 1-21.

Grunau, C., W. Hindermann and A. Rosenthal (2000). "Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes." Hum Mol Genet **9**(18): 2651-2663.

- Guisbert, E., D. M. Czyz, K. Richter, P. D. McMullen and R. I. Morimoto (2013). "Identification of a tissue-selective heat shock response regulatory network." *PLoS Genet* **9**(4): e1003466.
- Gupta, R. S. (1995). "Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species." *Mol Biol Evol* **12**(6): 1063-1073.
- Guthrie, R. D. (2006). "New carbon dates link climatic change with human colonization and Pleistocene extinctions." *Nature* **441**(7090): 207-209.
- Hajkova, P., O. el-Maarri, S. Engemann, J. Oswald, A. Olek and J. Walter (2002). "DNA-methylation analysis by the bisulfite-assisted genomic sequencing method." *Methods Mol Biol* **200**: 143-154.
- Hales, B. F., A. Aguilar-Mahecha and B. Robaire (2005). "The stress response in gametes and embryos after paternal chemical exposures." *Toxicol Appl Pharmacol* **207**(2 Suppl): 514-520.
- Han, G., H. Ma, R. Chintala, D. J. Fulton, S. A. Barman and R. E. White (2009). "Essential role of the 90-kilodalton heat shock protein in mediating nongenomic estrogen signaling in coronary artery smooth muscle." *J Pharmacol Exp Ther* **329**(3): 850-855.
- Hancock, A. M., D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin, R. Sukernik, G. Utermann, J. K. Pritchard, G. Coop and A. Di Rienzo (2011). "Adaptations to climate-mediated selective pressures in humans." *PLoS Genet* **7**(4): e1001375.
- Hancock, A. M., D. B. Witonsky, A. S. Gordon, G. Eshel, J. K. Pritchard, G. Coop and A. Di Rienzo (2008). "Adaptations to climate in candidate genes for common metabolic disorders." *PLoS Genet* **4**(2): e32.
- Hansen PJ (2009). "Effects of heat stress on mammalian reproduction." *Philos Trans R Soc Lond B Biol Sci* **364**.
- Hartson, S. D. and R. L. Matts (2012). "Approaches for defining the Hsp90-dependent proteome." *Biochim Biophys Acta* **1823**(3): 656-667.
- Hassanin, A., J. An, A. Ropiquet, T. T. Nguyen and A. Couloux (2013). "Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae)." *Mol Phylogenet Evol* **66**(3): 766-775.
- Hassanin, A., G. Lecointre and S. Tillier (1998). "The 'evolutionary signal' of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny." *C R Acad Sci III* **321**(7): 611-620.
- Hassanin, A. and A. Ropiquet (2004). "Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, *Bos sauveli* Urbain 1937." *Mol Phylogenet Evol* **33**(3): 896-907.
- Hecht, N. B. (1998). "Molecular mechanisms of male germ cell differentiation." *Bioessays* **20**(7): 555-561.
- Hellman, A. and A. Chess (2007). "Gene body-specific methylation on the active X chromosome." *Science* **315**(5815): 1141-1143.
- Henderson, A., Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, Jr., D. Sen and P. M. Lansdorp (2014). "Detection of G-quadruplex DNA in mammalian cells." *Nucleic Acids Res* **42**(2): 860-869.

- Hernandez Fernandez, M. and E. S. Vrba (2005). "A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants." Biol Rev Camb Philos Soc **80**(2): 269-302.
- Hill, W. G. and A. Robertson (1968). "Linkage disequilibrium in finite populations." Theoretical and Applied Genetics **38**(6): 226-231.
- Huang, Y., W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu and A. Rao (2010). "The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing." PLoS One **5**(1): e8888.
- Huh, I., J. Zeng, T. Park and S. V. Yi (2013). "DNA methylation and transcriptional noise." Epigenetics Chromatin **6**(1): 9.
- Huson, D. H. and D. Bryant (2006). "Application of phylogenetic networks in evolutionary studies." Mol Biol Evol **23**(2): 254-267.
- Imran, M. and S. Mahmood (2011). "An overview of animal prion diseases." Virology **8**: 493.
- Jackson, S. E. (2013). "Hsp90: structure and function." Top Curr Chem **328**: 155-240.
- Jannes, P., C. Spiessens, I. Van der Auwera, T. D'Hooghe, G. Verhoeven and D. Vanderschueren (1998). "Male subfertility induced by acute scrotal heating affects embryo quality in normal female mice." Hum Reprod **13**(2): 372-375.
- Jarosz, D. F. and S. Lindquist (2010). "Hsp90 and environmental stress transform the adaptive value of natural genetic variation." Science **330**(6012): 1820-1824.
- Jin, S. G., S. Kadam and G. P. Pfeifer (2010). "Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine." Nucleic Acids Res **38**(11): e125.
- Jin, S. G., X. Wu, A. X. Li and G. P. Pfeifer (2011). "Genomic mapping of 5-hydroxymethylcytosine in the human brain." Nucleic Acids Res **39**(12): 5015-5024.
- Johnson, J. L. (2012). "Evolution and function of diverse Hsp90 homologs and cochaperone proteins." Biochim Biophys Acta **1823**(3): 607-613.
- Jones, P. A. and D. Takai (2001). "The role of DNA methylation in mammalian epigenetics." Science **293**(5532): 1068-1070.
- Joost, S., A. Bonin, M. W. Bruford, L. Despres, C. Conord, G. Erhardt and P. Taberlet (2007). "A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation." Mol Ecol **16**(18): 3955-3969.
- Joost, S., M. Kalbermatten and A. Bonin (2008). "Spatial analysis method (sam): a software tool combining molecular and environmental data to identify candidate loci for selection." Mol Ecol Resour **8**(5): 957-960.
- Jorgensen, A. and E. Rajpert-De Meyts (2014). "Regulation of meiotic entry and gonadal sex differentiation in the human: normal and disrupted signaling." Biomol Concepts **5**(4): 331-341.
- Jump, A. S., J. M. Hunt, J. A. Martinez-Izquierdo and J. Penuelas (2006). "Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*." Mol Ecol **15**(11): 3469-3480.
- Kampinga, H. H., J. Hageman, M. J. Vos, H. Kubota, R. M. Tanguay, E. A. Bruford, M. E. Cheetham, B. Chen and L. E. Hightower (2009). "Guidelines for the nomenclature of the human heat shock proteins." Cell Stress Chaperones **14**(1): 105-111.

- Karpenshif, Y. and K. A. Bernstein (2012). "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst) **11**(10): 781-788.
- Kicza, M. E. (2011). 2011 NOAA satellite and information service, Annual Report. NOAA Magazine. online, National Oceanic and Atmospheric Administration (NOAA), United States Department of Commerce.
- Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, S. Paiva, W. Barendse, E. Ciani, H. Raadsma, J. McEwan, B. Dalrymple and M. International Sheep Genomics Consortium (2012). "Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection." PLoS Biol **10**(2): e1001258.
- Kim, B., K. Park and K. Rhee (2013a). "Heat stress response of male germ cells." Cell Mol Life Sci **70**(15): 2623-2636.
- Kim B, P. K., Rhee K (2013). "Heat stress response of male germ cells." Cell Mol Life Sci **70**.
- Kim, Y. J., J. Y. Kim, A. R. Ko and T. C. Kang (2013b). "Reduction in heat shock protein 90 correlates to neuronal vulnerability in the rat piriform cortex following status epilepticus." Neuroscience **255**: 265-277.
- Kimura, K., A. Wakamatsu, Y. Suzuki, T. Ota, T. Nishikawa, R. Yamashita, J. Yamamoto, M. Sekine, K. Tsuritani, H. Wakaguri, S. Ishii, T. Sugiyama, K. Saito, Y. Isono, R. Irie, N. Kushida, T. Yoneyama, R. Otsuka, K. Kanda, T. Yokoi, H. Kondo, M. Wagatsuma, K. Murakawa, S. Ishida, T. Ishibashi, A. Takahashi-Fujii, T. Tanase, K. Nagai, H. Kikuchi, K. Nakai, T. Isogai and S. Sugano (2006). "Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes." Genome Res **16**(1): 55-65.
- Klose, R. J. and A. P. Bird (2006). "Genomic DNA methylation: the mark and its mediators." Trends Biochem Sci **31**(2): 89-97.
- Klose, R. J., S. A. Sarraf, L. Schmiedeberg, S. M. McDermott, I. Stancheva and A. P. Bird (2005). "DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG." Mol Cell **19**(5): 667-678.
- Koneswaran, G. and D. Nierenberg (2008). "Global farm animal production and global warming: impacting and mitigating climate change." Environ Health Perspect **116**(5): 578-582.
- Krone, P. H. and J. B. Sass (1994). "HSP 90 alpha and HSP 90 beta genes are present in the zebrafish and are differentially regulated in developing embryos." Biochem Biophys Res Commun **204**(2): 746-752.
- Kumar, K., D. Deka, A. Singh, P. Chattopadhyay and R. Dada (2012). "Expression pattern of PRM2, HSP90 and WNT5A in male partners of couples experiencing idiopathic recurrent miscarriages." J Genet **91**(3): 363-366.
- Kumar, M., K. Kumar, S. Jain, T. Hassan and R. Dada (2013). "Novel insights into the genetic and epigenetic paternal contribution to the human embryo." Clinics (Sao Paulo) **68 Suppl 1**: 5-14.
- Kumari, D. and K. Usdin (2001). "Interaction of the transcription factors USF1, USF2, and alpha-Pal/Nrf-1 with the FMRI promoter - Implications for Fragile X mental retardation syndrome." Journal of Biological Chemistry **276**(6): 4357-4364.
- Laiho, A., N. Kotaja, A. Gyenesei and A. Sironen (2013). "Transcriptome profiling of the murine testis during the first wave of spermatogenesis." PLoS One **8**(4): e61558.

- Lal, A., H. Peters, B. St Croix, Z. A. Haroon, M. W. Dewhirst, R. L. Strausberg, J. H. Kaanders, A. J. van der Kogel and G. J. Riggins (2001). "Transcriptional response to hypoxia in human tumors." J Natl Cancer Inst **93**(17): 1337-1343.
- Lal, I. S. R. (2013). "Agroforestry and biochar to offset climate change: a review." Agronomy for Sustainable Development **33**(1): 81-96.
- Lally, V. E., and B. F. Watson (1960). "Humiture revisited." Weatherwise **13**: 254-256.
- Landolin, J. M., D. S. Johnson, N. D. Trinklein, S. F. Aldred, C. Medina, H. Shulha, Z. Weng and R. M. Myers (2010). "Sequence features that drive human promoter function and tissue specificity." Genome Res **20**(7): 890-898.
- Landry, J. R., D. L. Mager and B. T. Wilhelm (2003). "Complex controls: the role of alternative promoters in mammalian genomes." Trends Genet **19**(11): 640-648.
- Larkindale, J. and E. Vierling (2008). "Core genome responses involved in acclimation to high temperature." Plant Physiol **146**(2): 748-761.
- LeBlanc, S., E. Hoglund, K. M. Gilmour and S. Currie (2012). "Hormonal modulation of the heat shock response: insights from fish with divergent cortisol stress responses." Am J Physiol Regul Integr Comp Physiol **302**(1): R184-192.
- Legarra, A., M. Ramon, E. Ugarte and M. D. Perez-Guzman (2007). "Economic weights of fertility, prolificacy, milk yield and longevity in dairy sheep." Animal **1**(2): 193-203.
- Lemon, B. and R. Tjian (2000). "Orchestrated response: a symphony of transcription factors for gene control." Genes & Development **14**(20): 2551-2569.
- Lenny Bernstein, P. B., Osvaldo Canziani, Zhenlin Chen, Renate Christ, Ogunlade Davidson, William Hare, Saleemul, D. K. Huq, Vladimir Kattsov, Zbigniew Kundzewicz, Jian Liu, Ulrike Lohmann, Martin Manning, Taroh Matsuno,, B. M. Bettina Menne, Monirul Mirza, Neville Nicholls, Leonard Nurse, Rajendra Pachauri, Jean Palutikof, Martin, D. Q. Parry, Nijavalli Ravindranath, Andy Reisinger, Jiawen Ren, Keywan Riahi, Cynthia Rosenzweig, Matilde, S. S. Rusticucci, Youba Sokona, Susan Solomon, Peter Stott, Ronald Stouffer, Taishi Sugiyama, Rob Swart, and C. V. Dennis Tirpak, Gary Yohe (2007). *Climate Change 2007: Synthesis Report*. R. B. Abdelkader Allali, Sandra Diaz, Ismail Elgizouli, Dave Griggs, David Hawkins, Olav Hohmeyer, and L. k. K.-B. Bubu Pateh Jallow, Neil Leary, Hoesung Lee, David Wratt. online.
- Lenormand, T. (2002). "Gene flow and the limits to natural selection." TRENDS in Ecology & Evolution **17**(4): 183-189.
- Lent, P. C. (1988). "Ovibos moschatus." Mammalian Species **302**: 1-9.
- Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-151.
- Li, W. and M. Liu (2011). "Distribution of 5-hydroxymethylcytosine in different human tissues." J Nucleic Acids **2011**: 870726.
- Liberski, P. P. (2012). "Historical overview of prion diseases: a view from afar." Folia Neuropathol **50**(1): 1-12.
- Lim, A., J. P. Steibel, P. M. Coussens, D. L. Grooms and S. R. Bolin (2012). "Differential gene expression segregates cattle confirmed positive for bovine tuberculosis from antemortem tuberculosis test-false positive cattle originating from herds free of bovine tuberculosis." Veterinary medicine international **2012**.

- Lipps, H. J. and D. Rhodes (2009). "G-quadruplex structures: in vivo evidence and function." Trends Cell Biol **19**(8): 414-422.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). SAS for Mixed Models. second Edition, Cary, NC: SAS Institute Inc.
- Liu, Y. X. (2010). "Temperature control of spermatogenesis and prospect of male contraception." Front Biosci (Schol Ed) **2**: 730-755.
- Liu, Z., L. Zhang, Y. Pu, Z. Liu, Z. Li, Y. Zhao and S. Qin (2014). "Cloning and expression of a cytosolic HSP90 gene in *Chlorella vulgaris*." Biomed Res Int **2014**: 487050.
- Lorincz, M. C., D. R. Dickerson, M. Schmitt and M. Groudine (2004). "Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells." Nat Struct Mol Biol **11**(11): 1068-1075.
- Lloyd, S. E., J. B. Uphill, P. V. Targonski, E. M. Fisher and J. Collinge (2002). "Identification of genetic loci affecting mouse-adapted bovine spongiform encephalopathy incubation time in mice." Neurogenetics **4**(2): 77-81.
- MacNeish, R. S. (1992). "The origins of agriculture and settled life."
- Makino, Y., E. Inoue, M. Hada, K. Aoshima, S. Kitano, H. Miyachi and Y. Okada (2014). "Generation of a dual-color reporter mouse line to monitor spermatogenesis in vivo." Front Cell Dev Biol **2**: 30.
- Malama, E., H. Bollwein, I. A. Taitzoglou, T. Theodosiou, C. M. Boscovos and E. Kiossis (2013). "Chromatin integrity of ram spermatozoa. Relationships to annual fluctuations of scrotal surface temperature and temperature-humidity index." Theriogenology **80**(5): 533-541.
- Manolakou, K., J. Beaton, I. McConnell, C. Farquar, J. Manson, N. D. Hastie, M. Bruce and I. J. Jackson (2001). "Genetic and environmental factors modify bovine spongiform encephalopathy incubation period in mice." Proc Natl Acad Sci U S A **98**(13): 7402-7407.
- Marai, I. F. M., M. S. Ayyat and U. M. Abd El-Monem (2001). "Growth performance and reproductive traits at first parity of New Zealand White female rabbits as affected by heat stress and its alleviation under Egyptian conditions." Tropical Animal Health and Production **33**(6): 451-462.
- Marai, I. F. M., A. A. El-Darawany, A. Fadiel and M. A. M. Abdel-Hafez (2007). "Physiological traits as affected by heat stress in sheep - A review." Small Ruminant Research **71**(1-3): 1-12.
- Marai, I. F. M., A. A. El-Darawany, A. Fadiel and M. A. M. Abdel-Hafez (2008). "Reproductive performance traits as affected by heat stress and its alleviation in sheep." Tropical and Subtropical Agroecosystems **8**: 209-234.
- Marcos-Carcavilla, A., J. H. Calvo, C. Gonzalez, K. Moazami-Goudarzi, P. Laurent, M. Bertaud, H. Hayes, A. E. Beattie, C. Serrano, J. Lyahyai, I. Martin-Burriel and M. Serrano (2008). "Structural and functional analysis of the HSP90AA1 gene: distribution of polymorphisms among sheep with different responses to scrapie." Cell Stress Chaperones **13**(1): 19-29.
- Marcos-Carcavilla, A., C. Moreno, M. Serrano, P. Laurent, E. P. Cribiu, O. Andreoletti, J. Ruesche, J. L. Weisbecker, J. H. Calvo and K. Moazami-Goudarzi (2010a). "Polymorphisms in the HSP90AA1 5' flanking region are associated with scrapie incubation period in sheep." Cell Stress Chaperones **15**(4): 343-349.
- Marcos-Carcavilla, A., M. Mutikainen, C. Gonzalez, J. H. Calvo, J. Kantanen, A. Sanz, N. S. Marzanov, M. D. Perez-Guzman and M. Serrano (2010b). "A SNP in the HSP90AA1 gene 5' flanking region is



associated with the adaptation to differential thermal conditions in the ovine species." Cell Stress Chaperones **15**(1): 67-81.

Maston, G. A., S. K. Evans and M. R. Green (2006). "Transcriptional regulatory elements in the human genome." Annu Rev Genomics Hum Genet **7**: 29-59.

Matsuura, H., Y. Ishibashi, A. Shinmyo, S. Kanaya and K. Kato (2010). "Genome-wide analyses of early translational responses to elevated temperature and high salinity in *Arabidopsis thaliana*." Plant Cell Physiol **51**(3): 448-462.

Matthee, C. A. and S. K. Davis (2001). "Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective." Mol Biol Evol **18**(7): 1220-1230.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang and J. F. Costello (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature **466**(7303): 253-257.

McClellan, A. J., Y. Xia, A. M. Deutschbauer, R. W. Davis, M. Gerstein and J. Frydman (2007). "Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches." Cell **131**(1): 121-135.

McDonald, J. N., Ray, C.E., and Harington, C.R. (1991). Taxonomy and zoogeography of the musk ox genus *Praeovibos* Staudinger, 1908, Illinois State Museum Scientific Papers. **23**: 285-314.

Medvedeva, Y. A., A. M. Khamis, I. V. Kulakovskiy, W. Ba-Alawi, M. S. Bhuyan, H. Kawaji, T. Lassmann, M. Harbers, A. R. Forrest, V. B. Bajic and F. consortium (2014). "Effects of cytosine methylation on transcription factor binding sites." BMC Genomics **15**: 119.

Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch and E. S. Lander (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-770.

Meistrich, M. L., L. R. Bucci, P. K. Trostle-Weige and W. A. Brock (1985). "Histone variants in rat spermatogonia and primary spermatocytes." Dev Biol **112**(1): 230-240.

Mendelson, K. G., L. R. Contois, S. G. Tevosian, R. J. Davis and K. E. Paulson (1996). "Independent regulation of JNK/p38 mitogen-activated protein kinases by metabolic oxidative stress in the liver." Proceedings of the National Academy of Sciences of the United States of America **93**(23): 12908-12913.

Meng, X., V. Jerome, J. Devin, E. E. Baulieu and M. G. Catelli (1993). "Cloning of chicken hsp90 beta: the only vertebrate hsp90 insensitive to heat shock." Biochem Biophys Res Commun **190**(2): 630-636.

Messeguer, X., R. Escudero, D. Farre, O. Nunez, J. Martinez and M. Alba (2002). "PROMO: detection of known transcription regulatory elements using species-tailored searches." Bioinformatics **18**(2): 333-334.

Miller, S. A., D. D. Dykes and H. F. Polesky (1988). "A simple salting out procedure for extracting dna from human nucleated cells." Nucleic Acids Research **16**(3): 1215-1215.

Moreno, C. R., F. Lantier, I. Lantier, P. Sarradin and J. M. Elsen (2003). "Detection of new quantitative trait Loci for susceptibility to transmissible spongiform encephalopathies in mice." Genetics **165**(4): 2085-2091.

- Morimoto, R. I. (1998). "Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators." Genes Dev **12**(24): 3788-3796.
- Morton, J. M., W. P. Tranter, D. G. Mayer and N. N. Jonsson (2007). "Effects of environmental heat on conception rates in lactating dairy cows: Critical periods of exposure." Journal of Dairy Science **90**(5): 2271-2278.
- Muhlbacher, W., S. Sainsbury, M. Hemann, M. Hantsche, S. Neyer, F. Herzog and P. Cramer (2014). "Conserved architecture of the core RNA polymerase II initiation complex." Nat Commun **5**: 4310.
- Nan, X., F. J. Campoy and A. Bird (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." Cell **88**(4): 471-481.
- Nan, X., R. R. Meehan and A. Bird (1993). "Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2." Nucleic Acids Res **21**(21): 4886-4892.
- Nestor, C., A. Ruzov, R. Meehan and D. Dunican (2010). "Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA." Biotechniques **48**(4): 317-319.
- Nestor, C. E., R. Ottaviano, J. Reddington, D. Sproul, D. Reinhardt, D. Dunican, E. Katz, J. M. Dixon, D. J. Harrison and R. R. Meehan (2012). "Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes." Genome Res **22**(3): 467-477.
- Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark and C. Bustamante (2005). "Genomic scans for selective sweeps using SNP data." Genome Res **15**(11): 1566-1575.
- Nordstoga, A. B., A. Krogenaes, A. Nodtvedt, W. Farstad and K. Waterhouse (2013). "The relationship between post-thaw sperm DNA integrity and non-return rate among Norwegian cross-bred rams." Reprod Domest Anim **48**(2): 207-212.
- O'Shea-Greenfield, A. and S. T. Smale (1992). "Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription." J Biol Chem **267**(9): 6450.
- Oakes, C. C., S. La Salle, D. J. Smiraglia, B. Robaire and J. M. Trasler (2007). "Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells." Dev Biol **307**(2): 368-379.
- Oesch, B., D. Westaway, M. Walchli, M. P. McKinley, S. B. Kent, R. Aebersold, R. A. Barry, P. Tempst, D. B. Teplow, L. E. Hood and et al. (1985). "A cellular gene encodes scrapie PrP 27-30 protein." Cell **40**(4): 735-746.
- Olivieri, G. and A. Olivieri (1965). "Autoradiographic study of nucleic acid synthesis during spermatogenesis in *Drosophila melanogaster*." Mutat Res **2**(4): 366-380.
- Olsson, O. (2001). "The Rise of Neolithic Agriculture." Working Papers in Economics **57**(University of Gothenburg, Department of Economics).
- Oner, Y., J. H. Calvo, M. Serrano and C. Elmaci (2012). "Polymorphisms at the 5' flanking region of the HSP90AA1 gene in native Turkish sheep breeds." Livestock Science **150**(1-3): 381-385.
- Opitz, L., G. Salinas-Riester, M. Grade, K. Jung, P. Jo, G. Emons, B. M. Ghadimi, T. Beissbarth and J. Gaedcke (2010). "Impact of RNA degradation on gene expression profiling." Bmc Medical Genomics **3**.

- Ostermeier, G. C., D. Miller, J. D. Huntriss, M. P. Diamond and S. A. Krawetz (2004). "Reproductive biology: delivering spermatozoan RNA to the oocyte." Nature **429**(6988): 154.
- Paris, S. and J. R. Pringle (1983). "Saccharomyces cerevisiae: heat and gluculase sensitivities of starved cells." Ann Microbiol (Paris) **134B**(3): 379-385.
- Paul C, M. A., Spears N, Saunders PTK, (2008). "A single, mild, transient scrotal heat stress causes DNA damage, subfertility and impairs formation of blastocysts in mice." Reproduction **136**.
- Paul, C., D. W. Melton and P. T. Saunders (2008). "Do heat stress and deficits in DNA repair pathways have a negative impact on male fertility?" Mol Hum Reprod **14**(1): 1-8.
- Paul Palmqvist, J. A. P.-C., Christine M. Janis, Borja Figueirido, Vanessa Torregrosa, and Darren R. Gröcke (2008). "Tracing the ecophysiology of ungulates and predator-prey relationships in an early Pleistocene large mammal community." Palaeogeography, Palaeoclimatology, Palaeoecology **266**(1-2): 95-111.
- Pearl, L. H., C. Prodromou and P. Workman (2008). "The Hsp90 molecular chaperone: an open and shut case for treatment." Biochem J **410**(3): 439-453.
- Pecci, A., L. R. Viegas, J. L. Baranao and M. Beato (2001). "Promoter choice influences alternative splicing and determines the balance of isoforms expressed from the mouse bcl-X gene." J Biol Chem **276**(24): 21062-21069.
- Pedrosa, S., M. Uzun, J. J. Arranz, B. Gutierrez-Gil, F. San Primitivo and Y. Bayon (2005). "Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events." Proc Biol Sci **272**(1577): 2211-2217.
- Pérez-Crespo M, P. B., Gutierrez-Adán A (2008). "Scrotal heat stress effects on sperm viability, sperm DNA integrity and the Offspring sex ratio in mice." Molecular Reproduction and Development **75**.
- Perez-Crespo, M., B. Pintado and A. Gutierrez-Adan (2008). "Scrotal heat stress effects on sperm viability, sperm DNA integrity, and the offspring sex ratio in mice." Mol Reprod Dev **75**(1): 40-47.
- Phillips, K., Z. Dauter, A. I. Murchie, D. M. Lilley and B. Luisi (1997). "The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution." J Mol Biol **273**(1): 171-182.
- Pirkkala, L., P. Nykanen and L. Sistonen (2001). "Roles of the heat shock transcription factors in regulation of the heat shock response and beyond." FASEB J **15**(7): 1118-1131.
- Previti, C., O. Harari, I. Zwir and C. del Val (2009). "Profile analysis and prediction of tissue-specific CpG island methylation classes." BMC Bioinformatics **10**: 116.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly and P. C. Sham (2007a). "PLINK: a tool set for whole-genome association and population-based linkage analyses." American Journal of Human Genetics **81**(3): 559-575.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly and P. C. Sham (2007b). "PLINK: A tool set for whole-genome association and population-based linkage analyses." American Journal of Human Genetics **81**(3): 559-575.
- Qian, W. and J. Zhang (2009). "Protein subcellular relocalization in the evolution of yeast singleton and duplicate genes." Genome Biol Evol **1**: 198-204.

- Rakyan, V. K., T. A. Down, N. P. Thorne, P. Flicek, E. Kulesha, S. Graf, E. M. Tomazou, L. Backdahl, N. Johnson, M. Herberth, K. L. Howe, D. K. Jackson, M. M. Miretti, H. Fiegler, J. C. Marion, E. Birney, T. J. Hubbard, N. P. Carter, S. Tavare and S. Beck (2008). "An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs)." Genome Res **18**(9): 1518-1529.
- Rasmussen, R. (2001). Quantification on the LightCycler. Rapid Cycle Real-time PCR, Methods and Applications S. Meuer, Wittwer, C, Nakagawara, K, eds. Springer Press, Heidelberg; : 21-34.
- Rathke, C., W. M. Baarends, S. Awe and R. Renkawitz-Pohl (2014). "Chromatin dynamics during spermiogenesis." Biochim Biophys Acta **1839**(3): 155-168.
- Reddy, P. S., V. Thirulogachandar, C. S. Vaishnavi, A. Aakrati, S. K. Sopory and M. K. Reddy (2011). "Molecular characterization and expression of a gene encoding cytosolic Hsp90 from Pennisetum glaucum and its role in abiotic stress adaptation." Gene **474**(1-2): 29-38.
- Reynolds, J., B. S. Weir and C. C. Cockerham (1983). "Estimation of the coancestry coefficient: basis for a short-term genetic distance." Genetics **105**(3): 767-779.
- Richter, K., M. Haslbeck and J. Buchner (2010). "The heat shock response: life on the verge of death." Mol Cell **40**(2): 253-266.
- Ritossa, F. (1962). "A new puffing pattern induced by temperature shock and DNP in drosophila." Experientia **18**(12): 571-573.
- Ritossa, F. (1996). "Discovery of the heat shock response." Cell Stress Chaperones **1**(2): 97-98.
- Rivals, F., E. Schulz and T. M. Kaiser (2009). "A new application of dental wear analyses: estimation of duration of hominid occupations in archaeological localities." J Hum Evol **56**(4): 329-339.
- Rockett, J. C., F. L. Mapp, J. B. Garges, J. C. Luft, C. Mori and D. J. Dix (2001). "Effects of hyperthermia on spermatogenesis, apoptosis, gene expression, and fertility in adult male mice." Biol Reprod **65**(1): 229-239.
- Rousset, R. M. a. F. (1995). "An exact test for population differentiation." Evolution **49**(6): 1280-1283.
- Ruden, D. M. and X. Lu (2008). "Hsp90 affecting chromatin remodeling might explain transgenerational epigenetic inheritance in Drosophila." Curr Genomics **9**(7): 500-508.
- Sailer BL, S. L., Jost LK, Bjordahl J, Evenson DP (1997). "Effects of heat stress on mouse testicular cells and sperm chromatin structure as measured by flow cytometry." J Androl **18**.
- Salvatore, P., G. Benvenuto, M. Caporaso, C. B. Bruni and L. Chiariotti (1998). "High resolution methylation analysis of the galectin-I gene promoter region in expressing and nonexpressing tissues." FEBS Lett **421**(2): 152-158.
- Sanchez, J. P., I. Misztal, I. Aguilar, B. Zumbach and R. Rekaya (2009). "Genetic determination of the onset of heat stress on daily milk production in the US Holstein cattle." Journal of Dairy Science **92**(8): 4035-4045.
- Santini, S., J. L. Boore and A. Meyer (2003). "Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters." Genome Res **13**(6A): 1111-1122.
- Saxonov, S., P. Berg and D. L. Brutlag (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proc Natl Acad Sci U S A **103**(5): 1412-1417.

- Scherf, B. D. (2000). WORLD WATCH LIST for domestic animal diversity. B. D. SCHERF, Food and agriculture organization of the united nations, FAO.
- Schug, J. (2008). "Using TESS to predict transcription factor binding sites in DNA sequence." Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.] Chapter 2.
- Serrano, M., N. Moreno-Sanchez, C. Gonzalez, A. Marcos-Carcavilla, M. Van Poucke, J. H. Calvo, J. Salces, J. Cubero and M. J. Carabano (2011). "Use of Maximum Likelihood-Mixed Models to select stable reference genes: a case of heat stress response in sheep." Bmc Molecular Biology **12**.
- Setchell, B. P. (1998). "The Parkes Lecture. Heat and the testis." J Reprod Fertil **114**(2): 179-194.
- Setchell BP (2006). "The effects of heat on the testes of mammals." Anim Reprod **3**.
- Sevi, A. and M. Caroprese (2012). "Impact of heat stress on milk production, immunity and udder health in sheep: A critical review." Small Ruminant Research **107**(1): 1-7.
- Sezgin, E., D. D. Duvernell, L. M. Matzkin, Y. Duan, C. T. Zhu, B. C. Verrelli and W. F. Eanes (2004). "Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*." Genetics **168**(2): 923-931.
- Shackleton, D. M. (1997). Wild Sheep and Goats and Their Relatives: Status Survey and Conservation Action Plan for Caprinae.
- Shukla, R. R., Z. Dominski, T. Zwierzynski and R. Kole (1990). "Inactivation of splicing factors in HeLa cells subjected to heat shock." J Biol Chem **265**(33): 20377-20383.
- Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg and S. Oberdoerffer (2011). "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." Nature **479**(7371): 74-79.
- Smale, S. T. (2001). "Core promoters: active contributors to combinatorial gene regulation." Genes Dev **15**(19): 2503-2508.
- Soltani, S., H. Askari, N. Ejlali and R. Aghdam (2014). "The structural properties of DNA regulate gene expression." Mol Biosyst **10**(2): 273-280.
- Spano, M., J. P. Bonde, H. I. Hjollund, H. A. Kolstad, E. Cordelli and G. Leter (2000). "Sperm chromatin damage impairs human fertility. The Danish First Pregnancy Planner Study Team." Fertil Steril **73**(1): 43-50.
- Sreedhar, A. S., E. Kalmar, P. Csermely and Y. F. Shen (2004). "Hsp90 isoforms: functions, expression and clinical importance." FEBS Lett **562**(1-3): 11-15.
- Stecklein, S. R., E. Kumaraswamy, F. Behbod, W. Wang, V. Chaguturu, L. M. Harlan-Williams and R. A. Jensen (2012). "BRCA1 and HSP90 cooperate in homologous and non-homologous DNA double-strand-break repair and G2/M checkpoint activation." Proc Natl Acad Sci U S A **109**(34): 13650-13655.
- Steibel, J. P., R. Poletto, P. M. Coussens and G. J. M. Rosa (2009). "A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data." Genomics **94**(2): 146-152.
- Stephanou, A. and D. S. Latchman (2011). "Transcriptional modulation of heat-shock protein gene expression." Biochemistry research international **2011**.

- Stiftung, H. B. (2014). Meat Atlas – Facts and figures about the animals we eat, Heinrich Böll Foundation and Friends of the Earth Europe: 68.
- Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan and J. Schmitt (2004). "A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA." Proc Natl Acad Sci U S A **101**(13): 4712-4717.
- Sud, N., S. Sharma, D. A. Wiseman, C. Harmon, S. Kumar, R. C. Venema, J. R. Fineman and S. M. Black (2007). "Nitric oxide and superoxide generation from endothelial NOS: modulation by HSP90." Am J Physiol Lung Cell Mol Physiol **293**(6): L1444-L1453.
- Svante Wold, M. S., Lennart Eriksson (2001). "PLS-regression: a basic tool of chemometrics." Chemometrics and Intelligent Laboratory Systems **58**: 109–130.
- T. Lundstedt, E. S., Lisbeth Abramo, Bernt Thelin, Asa Nyström, Jarle Pettersen, Rolf Bergman (1998). "Experimental design and optimization." Chemometrics and Intelligent Laboratory Systems **42**: 3–40.
- Tabuchi, Y., I. Takasaki, S. Wada, Q. L. Zhao, T. Hori, T. Nomura, K. Ohtsuka and T. Kondo (2008). "Genes and genetic networks responsive to mild hyperthermia in human lymphoma U937 cells." Int J Hyperthermia **24**(8): 613-622.
- Taipale, M., D. F. Jarosz and S. Lindquist (2010). "HSP90 at the hub of protein homeostasis: emerging mechanistic insights." Nat Rev Mol Cell Biol **11**(7): 515-528.
- Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Mol Biol Evol **30**(12): 2725-2729.
- Tapio, M., N. Marzanov, M. Ozerov, M. Cinkulov, G. Gonzarenko, T. Kiselyova, M. Murawski, H. Viinalass and J. Kantanen (2006). "Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas." Mol Biol Evol **23**(9): 1776-1783.
- Team, R. C. (2013). A language and environmental for statistical computing. R. F. f. S. Computing. Vienna, Austria.
- Thanos, D. and T. Maniatis (1995). "Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome." Cell **83**(7): 1091-1100.
- Thepot, D., J. B. Weitzman, J. Barra, D. Segretain, M. G. Stinnakre, C. Babinet and M. Yaniv (2000). "Targeted disruption of the murine junD gene results in multiple defects in male reproductive function." Development **127**(1): 143-153.
- Thompson, E. E., H. Kuttab-Boulos, D. Witonsky, L. Yang, B. A. Roe and A. Di Rienzo (2004). "CYP3A variation and the evolution of salt-sensitivity variants." Am J Hum Genet **75**(6): 1059-1069.
- Tian, W. L., F. He, X. Fu, J. T. Lin, P. Tang, Y. M. Huang, R. Guo and L. Sun (2014). "High expression of heat shock protein 90 alpha and its significance in human acute leukemia cells." Gene **542**(2): 122-128.
- Timon, V. M. (1985). SMALL RUMINANT PRODUCTION IN THE DEVELOPING COUNTRIES - SYNTHESIS AND RECOMMENDATIONS OF THE CONSULTATION. A. a. c. p. department.
- Torok, Z., T. Crul, B. Maresca, G. J. Schutz, F. Viana, L. Dindia, S. Piotto, M. Brameshuber, G. Balogh, M. Peter, A. Porta, A. Trapani, I. Gombos, A. Glatz, B. Gungor, B. Peksel, L. Vigh, Jr., B.

- Csoboz, I. Horvath, M. M. Vijayan, P. L. Hooper, J. L. Harwood and L. Vigh (2014). "Plasma membranes as heat stress sensors: from lipid-controlled molecular switches to therapeutic applications." Biochim Biophys Acta **1838**(6): 1594-1618.
- Trepel, J., M. Mollapour, G. Giaccone and L. Neckers (2010). "Targeting the dynamic HSP90 complex in cancer." Nat Rev Cancer **10**(8): 537-549.
- Trinklein, N. D., J. I. Murray, S. J. Hartman, D. Botstein and R. M. Myers (2004). "The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response." Molecular Biology of the Cell **15**(3): 1254-1261.
- Turner, T. and T. Caspari (2014). "When heat casts a spell on the DNA damage checkpoints." Open Biol **4**: 140008.
- van Arensbergen, J., B. van Steensel and H. J. Bussemaker (2014). "In search of the determinants of enhancer-promoter interaction specificity." Trends Cell Biol.
- van der Voet, H. (1994). "Comparing the predictive accuracy of models using a simple randomization test." Chemometrics and Intelligent Laboratory Systems **25**(2): 313-323.
- Velichko, A. K., E. N. Markova, N. V. Petrova, S. V. Razin and O. L. Kantidze (2013). "Mechanisms of heat shock response in mammals." Cell Mol Life Sci **70**(22): 4229-4241.
- Velichko, A. K., N. V. Petrova, O. L. Kantidze and S. V. Razin (2012). "Dual effect of heat shock on DNA replication and genome integrity." Mol Biol Cell **23**(17): 3450-3460.
- Verghese, J., J. Abrams, Y. Wang and K. A. Morano (2012). "Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system." Microbiol Mol Biol Rev **76**(2): 115-158.
- Vibrantovski, M. D., D. S. Chalopin, H. F. Lopes, M. Long and T. L. Karr (2010). "Direct evidence for postmeiotic transcription during *Drosophila melanogaster* spermatogenesis." Genetics **186**(1): 431-433.
- Wade, P. A. (2005). "SWItching off methylated DNA." Nat Genet **37**(3): 212-213.
- Wadekar, S. A., D. P. Li, S. Periyasamy and E. R. Sanchez (2001). "Inhibition of heat shock transcription factor by GR." Molecular Endocrinology **15**(8): 1396-1410.
- Wan, J., V. F. Oliver, H. Zhu, D. J. Zack, J. Qian and S. L. Merbs (2013). "Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs." Nucleic Acids Res **41**(18): 8503-8514.
- Ward, W. S. (2010). "Function of sperm chromatin structural elements in fertilization and development." Mol Hum Reprod **16**(1): 30-36.
- Warnecke, P. M., C. Stirzaker, J. Song, C. Grunau, J. R. Melki and S. J. Clark (2002). "Identification and resolution of artifacts in bisulfite sequencing." Methods **27**(2): 101-107.
- Weber, C., G. Guigon, C. Bouchier, L. Frangeul, S. Moreira, O. Sismeiro, C. Gouyette, D. Mirelman, J. Y. Coppee and N. Guillen (2006). "Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *Entamoeba histolytica*." Eukaryot Cell **5**(5): 871-875.
- Weisdorf, J. L. (2003). "From Foraging to Farming: Explaining the Neolithic Revolution." Discussion Papers 03-41, University of Copenhagen. Department of Economics.

- Weisdorf, J. L. (2005). "From Foraging To Farming: Explaining The Neolithic Revolution." Journal of Economic Surveys **19**(4): 561-586.
- Welch, W. J. and J. P. Suhan (1985). "Morphological study of the mammalian stress response: characterization of changes in cytoplasmic organelles, cytoskeleton, and nucleoli, and appearance of intranuclear actin filaments in rat fibroblasts after heat-shock treatment." J Cell Biol **101**(4): 1198-1211.
- Whitesell, L. and S. L. Lindquist (2005). "HSP90 and the chaperoning of cancer." Nat Rev Cancer **5**(10): 761-772.
- Wood, M. A. and W. H. Walker (2009). "USF1/2 Transcription Factor DNA-Binding Activity Is Induced During Rat Sertoli Cell Differentiation." Biology of Reproduction **80**(1): 24-33.
- Yaeram, J., B. P. Setchell and S. Maddocks (2006). "Effect of heat stress on the fertility of male mice in vivo and in vitro." Reprod Fertil Dev **18**(6): 647-653.
- Yan, B., N. Raben and P. H. Plotz (2002). "Hes-1, a known transcriptional repressor, acts as a transcriptional activator for the human acid alpha-glucosidase gene in human fibroblast cells." Biochemical and Biophysical Research Communications **291**(3): 582-587.
- Yang, C., E. Bolotin, T. Jiang, F. M. Sladek and E. Martinez (2007). "Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters." Gene **389**(1): 52-65.
- Yang, Z., C. Yang, L. Xiao, X. Liao, A. Lan, X. Wang, R. Guo, P. Chen, C. Hu and J. Feng (2011). "Novel insights into the role of HSP90 in cytoprotection of H2S against chemical hypoxia-induced injury in H9c2 cardiac myocytes." Int J Mol Med **28**(3): 397-403.
- Yue, L., T. L. Karr, D. F. Nathan, H. Swift, S. Srinivasan and S. Lindquist (1999). "Genetic analysis of viable Hsp90 alleles reveals a critical role in Drosophila spermatogenesis." Genetics **151**(3): 1065-1079.
- Zhao, R., M. Davey, Y. C. Hsu, P. Kaplanek, A. Tong, A. B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili and W. A. Houry (2005). "Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone." Cell **120**(5): 715-727.
- Zuehlke, A. and J. L. Johnson (2010). "Hsp90 and co-chaperones twist the functions of diverse client proteins." Biopolymers **93**(3): 211-217.