Taking advantage of the use of supervised learning methods for characterization of sperm population structure related with freezability in the Iberian red deer

ARTICLE *in* THERIOGENOLOGY · FEBRUARY 2012

Impact Factor: 1.8 · DOI: 10.1016/j.theriogenology.2011.12.011 · Source: PubMed

9 AUTHORS, INCLUDING:

Manuel Ramon
Health Sciences Institute, Castile-La Mancha
43 PUBLICATIONS **210** CITATIONS
SEE PROFILE

Felipe Martínez-Pastor
Universidad de León
145 PUBLICATIONS **1,395** CITATIONS
SEE PROFILE

Ana J. Soler
University of Castilla-La Mancha
88 PUBLICATIONS **1,387** CITATIONS
SEE PROFILE

Julian Garde
University of Castilla-La Mancha
153 PUBLICATIONS **2,394** CITATIONS
SEE PROFILE

Revised

TAKING ADVANTAGE OF THE USE OF SUPERVISED LEARNING METHODS FOR CHARACTERIZATION OF SPERM POPULATION STRUCTURE RELATED WITH FREEZABILITY IN THE IBERIAN RED DEER.

MANUEL RAMÓN,[a,b] FELIPE MARTÍNEZ-PASTOR,[c,d] OLGA GARCÍA-ÁLVAREZ,[e] ALEJANDRO MAROTO-MORALES,[a] ANA JOSEFA SOLER,[a] PILAR JIMÉNEZ-RABADÁN,[e] MARIA ROCÍO FERNÁNDEZ-SANTOS,[a] RODOLFO BERNABÉU,[f] JOSÉ JULIÁN GARDE[a,*]

[a]*Biology of Reproduction Group, National Wildlife Research Institute (IREC) UCLM-CSIC-JCCM, 02071, Albacete, Spain.*

[b]*Department of Medicine and Animal Surgery, Faculty of Veterinary Medicine, University of Murcia, 30071, Murcia, Spain.*

[c]*ITRA-ULE, INDEGSAL, University of León, 24071, León, Spain.*

[d]*Molecular Biology (Cell Biology), University of León, 24071, León, Spain.*

[e]*CERSYRA, Castilla-La Mancha, 13300, Valdepeñas, Spain.*

[f]*Escuela Técnica Superior de Ingenieros Agrónomos, University of Castilla-La Mancha, 02071, Albacete, Spain*

[*] Corresponding author. Tel: +34-67-599200; fax: +34-67-599238.

E-mail address: julian.garde@uclm.es (J. J. Garde)

ABSTRACT

Using Iberian red deer as model, this study presents a supervised learning method, the Support Vector Machines (SVM), to characterize sperm population structure related with freezability. Male freezability was assessed by evaluating motility, membrane status and mitochondrial membrane potential of sperm after a freezing-thawing procedure. The SVM model was generated using sperm motility information captured by computer-assisted sperm analysis (CASA) from thawed semen, belonging to 6 stags with marked differences on their freezability. A total of 1369 sperm tracks were recorded for seven kinematic parameters and assigned to four motility patterns based on them: weak motile, progressive, transitional and hyperactivated-like. Then, this data were split in two sets: the training set, used to train the SVM model, and the testing set, used to examine how the SVM method and three other unsupervised methods, a non-hierarchical, a hierarchical and a multi-step clustering procedures, performed the sperm classification into subpopulations. The SVM was revealed as the most accurate method in the characterization of sperm subpopulations, showing all the sperm subpopulations obtained in this way high significant correlations with those sperm parameters used to characterize freezability of males. Given its superiority, the SVM method was used to characterize the sperm motile subpopulations in Iberian red deer. Sperm motile data from frozen – thawed semen belonging to 25 stags were recorded and loaded into the SVM model. The sperm population structure revealed that those males showing poor freezability were characterized by high percentages of sperm with a weak motility pattern. In opposite, males showing good freezability were characterized by higher percentages of sperm with a progressive and hyperactivated-like motility pattern and lower percentages of sperm with a weak motile pattern. We also identified a sperm subpopulation with a transitional

motility pattern. This subpopulation increased as the freezability of males improved, and may be used as indicative of overall sperm motility.

Keywords: sperm subpopulations, support vector machines, sperm freezability, Iberian red deer.

Funding/Acknowledgements

This study was supported in part by the Spanish Ministry of Science and Innovation (Project AGL2007-60271/GAN), and by the Education and Science Council of the Junta de Castilla y León, Spain (Project LE019A10-2). Felipe Martínez-Pastor was supported by the Ramón y Cajal program (Ministry of Science and Innovation, Spain).

1. Introduction

One of the most recognized characteristic of the mammalian ejaculates is its heterogeneity, reflected in the presence of different sperm subpopulations [1-9]. This heterogeneity has been related to different key issues of male reproductive performance. Thus, it has been found that ability to undergo capacitation and fertilize may vary depending on the subpopulation under consideration [10,11], and that freezability may vary significantly among sperm subpopulations [9,12]. Therefore, it is fundamental to obtain a prior deep knowledge of the population structure of semen, in order to study the relationships between the quality and the reproductive performance of sperm samples.

Cell cryopreservation has become an indispensable tool in biology. Biological materials can be safely kept and used after a very long period of time. In the case of spermatozoa, cryopreservation is used not only in research, but also in livestock management and in the conservation of wild and domestic species, as a complementary tool for managing live animals and preserving their genetic diversity. Sperm cryopreservation combined with artificial insemination (AI) is the assisted reproductive technology (ART) which possibly has been increasingly applied to deer species too [13].

One important problem for standardizing sperm cryopreservation protocols is that sperm from different individuals exhibit significant different responses to the same freezing treatment [14-16]. Thus, males may show different freezability depending on their sperm population structure. Therefore, it is of interest to identify those characteristics that favor the freezability of spermatozoa, and to characterize the distribution of sperm subpopulations of males as a way to predict their freezability.

Different statistical procedures have been used for the definition and identification of sperm subpopulations. Martinez-Pastor et al. [17] provides references and a general

discussion on the classification methods commonly used to identify sperm subpopulations. In the last years, other analytical and statistical methods for cluster analysis have been developed, which could improve the current analyses of sperm subpopulations. *Data Mining* and *Machine Learning* disciplines are becoming increasingly important tools that provide useful methods to reach those objectives. In a general way, we could say that the aim in *Data Mining* and *Machine Learning* is to design computer programs to solve a task not based on predefined rules provided by the user, but using relations that they 'learned' from the information, data or feedback that they receive [18].

The learning processes can be roughly categorized as unsupervised or supervised. In *Unsupervised Learning*, there is no outcome measure; we observe only the features and the goal is to describe the associations and patterns among a set of input measures. As examples of unsupervised learning methods, we have the hierarchical and non-hierarchical clustering methods, among others, which are the preferred methods currently used in sperm subpopulations analyses. In *Supervised Learning*, the goal is to predict the value of an outcome measure based on a number of input measures, so the presence of the outcome variable guide the learning process. Data is usually split in two sets: the *training set* and *test set*. Training set of data is used to observe the outcome and feature measurements for a set of objects. Using this data we build a prediction model, or learner, which will enable us to predict the outcome for new unseen objects, the test set. A good learner is one that accurately predicts such an outcome. As supervised methods we can found several references on literature: nearest neighbor methods [19], logistic regression [20], decision trees [21], support vector machines [22] or neural networks [23] among others. An extended explanation of these methods and other supervised methods is given by [18].

In this study we used the Support vector machines (SVM) procedure. SVM's are a set of related supervised learning methods used for classification and regression [18,22,24]. SVM has been successfully used for several purposes [25,26]. In the field of biological sciences, we can found references about the use of this procedure for protein sequence comparisons [27,28], classification of genes and proteins [29,30], microarray gene expression analysis [31] or cancer classification [32,33]. As other supervised learning techniques, the SVM procedure involves separating data into *training* and *testing sets*. Each instance in the training set contains one "target value" (i.e. the class labels) and several "attributes" (i.e. the features of the observed variables). The goal of the SVM is to produce a model, based on the *training data*, which predicts the target values of the test data only given the test data attributes. Thus, having learned the features of one class, the SVM could recognize new objects as members or non-members of that class based on their attributes.

The purpose of this study has been to characterize those sperm subpopulations based on motile characteristics that could be related with the freezability of males in the Iberian reed deer, using the SVM methodology. For this study, semen samples from Iberian red deer were used as model. The advantage of using wild animals is that males are not artificially selected for fertility. Thus, they are expected to exhibit considerable diversity in sperm characteristics and fertility, as well as being an excellent model to study the eventual associations between sperm characteristics and reproduction performance.

2. Materials and Methods

2.1. Animals and sperm collection

Animal manipulations were performed in accordance with the Spanish Animal Protection Regulation, RD1201/2005, which conforms to European Union Regulation

2003/65. This study included a total of 25 Iberian red deer (*Cervus elaphus hispanicus*) stags hunted during the mating season in the south of Spain, coinciding with their reproductive season (end of September to December) [34]. Both testes and epididymes were removed (in the scrotum) and transported at 20-21 ºC to the laboratory. Elapsed time between animal death and sperm analyses ranged from 3 to 6 hours [35]. At the laboratory, testes and epididymides were removed from the scrotum. Spermatozoa were collected by cutting the distal proportion of the epididymides with a surgical blade, and diluted in PBS (pH 7.5; 320 mOsm/kg). After dilution, sperm motility was assessed subjectively and only those epididymal semen samples with a minimum quality were cryopreserved, as a way to assure that all sperm samples showed good quality before freezing. Thus, only semen samples with a sperm motility subjectively assessed over 80% (SM; 0 – 100%) and a quality movement over 4 (QM; on a scale of 0 – 5, where 0 is no motility and 5 is vigorous progressive movement) were freeze. Cryopreservation was performed as described by Soler et al. [16], and frozen semen was stored in liquid nitrogen (-196 ºC) for a minimum period of 6 months before thawing. Thawing was performing by dropping the straws in a water bath with saline serum at 37 ºC for 20 s.

2.2. Assessment of thawed sperm quality

After thawing, semen samples were incubated in a water bath at 37 ºC during 2 hours previously to semen quality assessment. Percentage of motile spermatozoa (SM) and the quality of movement (QM) were subjectively assessed and a resume measure, the *Sperm Motility Index* (SMI) was calculated as described by Comizzoli et al. [36]:

$$SMI = \frac{SM + (20 \times QM)}{2} \quad [1]$$

154

155 Membrane stability with YO-PRO-1, the viability with Propidium Iodide (PI) and the
156 mitochondrial membrane potential with Mitotracker Deep Red were assessed by flow
157 cytometry [37]. Thus, the YO-PRO-1–/PI– ratio is the proportion of viable spermatozoa with
158 a stable membrane, the YO-PRO-1+/PI+ ratio is the proportion of death spermatozoa and the
159 Mitotraker+/YO-PRO-1– ratio is the proportion of spermatozoa with high mitochondrial
160 membrane potential.

161 We also objectively assessed the motility characteristics of sperm by Computer Assisted
162 Semen Analysis (CASA) as described in Martínez-Pastor et al. [38]. Analyses were carried
163 out using the Sperm Class Analyzer software (SCA® 2002, Microptic, Barcelona, Spain) and
164 the following motility descriptors were recorded: curvilinear velocity (VCL, μm/s), average
165 path velocity (VAP, μm/s), straight line velocity (VSL, μm/s), linearity (LIN, %), straightness
166 (STR, %), amplitude of lateral head displacement (ALH, μm) and beat cross frequency (BCF,
167 Hz). A total of 6542 spermatozoa belonging to the 25 stags were recorded, with a minimum of
168 200 spermatozoa per male being assessed.

169

170 **2.3. Characterization of sperm motile subpopulations**

171 **2.3.1. Classification methods**

172 This study made use of four different classification procedures: the non-hierarchical
173 (*k-means*) clustering, the hierarchical clustering, and a multi-step procedure that used both
174 clustering methods jointly [12] as unsupervised methods, and the support vector machines
175 procedure as supervised learning method.

176

177 2.3.1.1. *Non-hierarchical (k-means) clustering*. A non-hierarchical clustering
178 method performing a disjoint cluster analysis on the basis of Euclidean distances has been
179 used. The optimal numbers of clusters to keep, four in this study, was set by using the
180 Silhouette Average Width (SAW) criterion [39]. The *kmeans* function from the STATS R
181 package [40] was used.
182
183 2.3.1.2. *Hierarchical clustering*. In this study, distance matrix was computed by
184 using the *Euclidean* distance measure and the *Ward's* minimum variance method was set to
185 classify the data. To determine the final number of clusters, the *Hubert* Γ coefficient [41]
186 criterion was considered. To perform the analysis, we used the *hclust* function in the STATS
187 package [40].
188
189 2.3.1.3. *Multistep procedure*. Non-hierarchical and hierarchical clustering
190 methods were used jointly in a multi-step procedure. Multi-step procedures have been used
191 successfully to classify sperm subpopulations [7,12]. Generally, non-hierarchical methods are
192 employed as the first step. The clusters produced by the non-hierarchical method are then
193 merged in the second step by an agglomerative hierarchical method. The first step may also
194 be used to identify outliers or special clusters, allowing continuation to the second step with
195 an optimized set of clusters. In this study, a total of 10 clusters were obtained from the non-
196 hierarchical step and the merged in a final number of 4 clusters in the hierarchical clustering
197 step. The criteria considered to define the optimal number of clusters in each step were then
198 same describe above for the non-hierarchical and hierarchical clustering methods.
199

2.3.1.4. *Support vector machines*. The SVM procedure involved the split of data into a training set used to train the SVM and a testing data set to evaluate the accuracy of these SVM. This testing set was used to evaluate the accuracy of the other non supervised methods as well. The SVM equation obtained from the training step were used to characterize the different sperm subpopulations in the Iberian red deer. A further explanation of these steps is given below.

2.3.2. Training and testing data sets

In order to identify the sperm subpopulation structure that best correlates with the freezability of males, individual sperm tracks from 6 stags with marked differences on their freezability were assessed (Table 1 and Figure 1, closed circles). A total of 1369 sperm tracks were recorded for the seven kinematic parameters defined above (section 2.2) and assigned to four different motility patterns, as described in Goodson et al. [42] (Table 2 and Figure 2). Then, this database was split and used to generate the SVM equations and to test how the SVM method and other clustering procedures perform the spermatozoa classification into subpopulations. Thus, the *training data set* consisted of 720 sperm tracks (120 per male) randomly chosen and was used to generate the SVM equations, while the *testing data set* consisted of the other 649 sperm tracks and was used to test the performance of the clustering procedures. The kinematic parameters VCL, VSL, VAP, ALH and BCF were loaded into the SVM procedure to generate SVM equations that were able to distinguish among sperm belonging to different subpopulations. We discarded the use of LIN and STR parameters because of they are linear combination of the other motility parameters. Once, the SVM equations were constructed, the accuracy of the classification of spermatozoa into different subpopulations by the SVM procedure, as well as, by the k-means, hierarchical and multi-step

clustering procedures were tested using data in the *testing set*. For the latter clustering methods, the same five kinematic traits, VCL, VSL, VAP, ALH and BCF were used.

2.3.3. Characterization of sperm motile subpopulations in the Iberian red deer

Finally, we examined the sperm population structure of the Iberian red deer using the SVM equations obtaining from the *training set*, with the aim to find a subpopulations distribution being related with freezability of males. Thus, 25 Iberian red deer males showing different freezability were used, and a total of 6542 sperm tracks were recorded and used to characterize the sperm motile subpopulations.

2.4. Statistical analysis

All statistical analyses in this study were conducted with the R statistical software [41]. To implement the SVM methodology, we used the **svm** function from the **e1071** R package [43]. Package e1071 provides an interface to **libsvm** [44], a robust and fast implementation of the most popular SVM formulations (*C* and ν classification, ϵ and ν regression, and novelty detection).

Results obtained from the use of different classification methods to characterize the subpopulation in the test set were presented graphically (Figure 3) and as a confusion matrix (Table 3). In this matrix, each row represents the instances in a predicted class, while each column represents the instances in an actual class. Values on diagonal (in bold) represents events that have been well-classified. To evaluate how different methods performed the classification, the overall accuracy rate has been calculated. The overall accuracy is defined as the sum of the diagonal of the confusion matrix divided by the total number of events. For this study, accuracy has been presented as an error rate, that is, 1 minus the overall accuracy. In

addition, correlations between the sperm parameters used to determine the freezability of a male (that is, the SMI and the cytometry parameters) and the percentages of subpopulations of males were calculated for each classification method. This allowed us to examine which of these methods perform a sperm subpopulation characterization that best correlates with freezability of males.

The method that performed the most accuracy classification, in this study the SVM method, was then used to characterize the sperm subpopulations distribution in the Iberian red deer. Once sperm population structure was characterized, we performed a principal component analysis (PCA) to examine the relations between the different sperm subpopulations defined for the Iberian red deer and to explore how an overall measure of the sperm population distribution could be useful to characterize freezability of males.

3. Results

3.1. Assessment of thawed sperm quality

After the freezing-thawing procedure, semen samples of Iberian red deer showed, on average (mean ± SEM), a SMI of 39.3 ± 2.9 %, a percentage of YO-PRO-1–/PI– and YO-PRO-1+/PI+ sperm of 41.2 ± 2.9 % and 39.7 ± 2.6 %, respectively, and Mitotracker+/ YO-PRO-1– of 39.0 ± 3.0 % (Table 1). Highly significant differences were observed between the two groups of males with different freezability. Thus, males with poor freezability showed a SMI of 15.0 ± 2.5 %, a YO-PRO-1–/PI– of 20.7 ± 1.4 %, a YO-PRO-1+/PI+ of 61.6 ± 2.3 % and a Mitotracker+/YO-PRO-1– of 20.2 ± 1.3 %, while for the group of three males with good freezability, a percentage of 52.5 ± 1.4 % for the SMI, a 55.8 ± 0.3 % and a 25.1 ± 1.7 % for the YO-PRO-1–/PI– and the YO-PRO-1+/PI+, respectively, and a 55.8 ± 4.0 % for the Mitotracker+/YO-PRO-1–, were observed.

272

273 **3.2. SVM model training**

274 The 1369 sperm tracks from 6 stags with large differences on their freezability were
275 manually classified into subpopulations based on their motility characteristics observed at 2
276 hours of incubation after thawing (Table 2 and Figure 2). Four motility patterns were clearly
277 defined: a subpopulation of ***weak motile*** spermatozoa (SP1) characterized by its low velocity
278 (VAP = 14.13 ± 0.36 μm/s) and very low linearity (LIN = 24.29 ± 0.51 %) with low lateral
279 head displacement from the path of movement (ALH = 1.72 ± 0.03 μm); a subpopulation of
280 spermatozoa with a vigorous ***progressive*** movement (SP2: VAP = 78.41 ± 1.88 μm/s; LIN =
281 56.12 ± 0.79 %; ALH = 3.97 ± 0.11 μm); a ***transitional*** subpopulation (SP3) that showed
282 decreasing speed (VAP = 61.02 ± 0.95 μm/s) and linearity (LIN = 31.17 ± 0.50 %) comparing
283 with SP2, but with an increasing lateral head movement (ALH = 4.69 ± 0.09 μm); and a
284 ***hyperactivated-like*** subpopulation (SP4) characterized by fast spermatozoa (VAP = 102.01 ±
285 0.98 μm/s) with low linearity (LIN = 20.48 ± 0.49 %) and a considerable lateral head
286 movement (ALH = 6.42 ± 0.10 μm). Total sperm tracks were characterized as follows (mean
287 ± SD): 55.0 ± 27.2 % as weak motile, 10.2 ± 7.7 % as progressive, 18.8 ± 11.3 % as
288 transitional and 16.0 ± 16.4 % as hyperactivated-like. When we compared between males
289 with poor and good freezability, the sperm distribution was (mean ± SD): for the poor
290 freezers, 77.7 ± 17.2 % as SP1, 3.8 ± 1.8 % as SP2, 16.6 ± 17.2 % as SP3 and 2.0 ± 2.1 % as
291 SP4; for the good freezers, 32.4 ± 5.4 % as SP1, 16.6 ± 4.8 % as SP2, 21.0 ± 2.7 % as SP3
292 and 30.0 ± 9.0 % as SP4. Differences between both poor and good freezers in the distribution
293 of all the subpopulations were significant ($p < 0.05$). Characterization of sperm into these four
294 subpopulations can be graphically observed on Figure 2A, as well as for the group of males
295 with poor (Figure 2B) and good (Figure 2C) freezability, separately. Once sperm tracks were

visually assigned to the different subpopulations based on their motility patterns, the database was split into the *training set* and the *testing set*. A total of 720 sperm tracks (120 per male) were randomly chosen to be the *training set* and then used to generate the SVM equations.

3.3. Accuracy of sperm classification by different mathematical approaches

A *testing set* including 649 sperm tracks belonging to the 6 stags with different freezability were used to test the accuracy of the SVM method and the k-means, hierarchical and multi-step clustering procedures. This *testing set* included 267 weakly motile, 91 progressive, 119 transitional and 172 hyperactivated-like sperm tracks. Table 3 presents the how different statistical procedures performed the sperm classification. The SVM method performed the most accurate classification showing an error rate of only 0.049 (~ 5 %). Thus, almost all the sperm were well-classified into their subpopulations using the SVM model generated from data in the *training set* (Figure 3A). The sperm characterization obtained using the non-hierarchical clustering and the multi-step methods were quite similar, with error rates of 0.197 and 0.216, respectively. These two methods well-identified the sperm with a weakly movement, and the transitional and hyperactivated-like subpopulations, but showed problems when identifying the sperm characterized by progressive movement (Figures 3B and 3D). The hierarchical clustering was the method that performed the less accurate classification, with an error rate of 0.537. This method showed an erratic classification of sperm, with more than half of sperm tracks being miss-classified (Figure 3C). Correlations between sperm parameters used to evaluate the freezability of a male and percentages of different subpopulations of males were calculated for each classification method (Table 4). The method showing the highest correlations was the SVM method, being the four subpopulations significantly correlated ($p < 0.05$) with the SMI and the sperm parameters

320 evaluated by flow cytometry. Meanwhile, non-hierarchical clustering methods showed lees
321 strong correlations and only the SP4 showed significant correlations with the sperm
322 parameters used to determine the freezability of males. Therefore, the SVM method resulted
323 to be clearly superior to the unsupervised clustering methods when looking for a
324 subpopulations distribution that best correlates with freezability of males.
325

326 **3.4. Characterization of the sperm motile subpopulations in the Iberian red deer**

327 Finally, we made used of the SVM model previously obtained to characterize the
328 motile subpopulation structure in the Iberian red deer and related with its freezability. The
329 6542 sperm tracks recorded from the 25 Iberian red deer stags were loaded into the SVM
330 model and then classified into four subpopulations, obtaining the following average
331 distribution within a male (mean ± SD and range): 56.0 ± 19. 7 % [27 – 92 %] of sperm tracks
332 were classified as weak motile; 10.8 ± 6.8 % [1 – 24 %] were classified as progressive; 20.8 ±
333 9.5 % [5 – 41 %] were classified as transitional; and 11.4 ± 11.6 % [0 – 40 %] were classified
334 as hyperactivated-like. The characteristics of the four motility patterns identified in the Iberian
335 red deer (Tabla 5) were similar to those described in the *training set* (Table 2).

336 To explore the relations between these four sperm subpopulations, we performed a
337 principal component analysis. We retained the first two principal components based on Kaiser
338 criterion (Figure 4). The first principal component accounted for 64.1 % of the total variance,
339 and which could be interpreted as an indicator of non-vigorous movement, so the greater this
340 value is, the less vigorous the movement is. This principal component allowed differentiating
341 between males with higher percentages of SP1, and males with higher percentages of SP2 and
342 SP4, the latter closely related (Figure 4). The subpopulation with a transitional motility
343 pattern (SP3) was mainly reflected on the second principal component which accounted for

344 25.7 % of the total variance. This second component allows differentiating among males with
345 an average freezability, and could be interpreted as indicator of overall motility. Thus, males
346 with higher percentages of motile sperm showed greater values for this second component.
347
348 4. **Discussion**
349 In the present study, we characterize the population structure of motile epididymal
350 spermatozoa in the Iberian red deer, and we suggest that this distribution could help to explain
351 the sperm freezability of different males. Contrasting with the statistical methods commonly
352 used for the characterization of sperm subpopulations, we propose a supervised learning
353 method, the support vector machines (SVM) procedure, and we show the superiority of this
354 method over traditional ones.
355 This study has been aimed to characterize the sperm population structure in the Iberian
356 stags by finding some subpopulations based on motile characteristics that maximize the
357 correlation with the freezability of these males. A number of studies have addressed the
358 characterization of sperm motile subpopulations in thawed samples in several species
359 [6,9,12,44-48]. Most of them have used unsupervised statistical methods [17], so the
360 characterization of sperm subpopulation has been conducted without considering any prior
361 information on freezability of males. This could lead to the sperm population structure
362 defined by using those clustering methods was no optimal. To our knowledge, few references
363 are found in the literature on the use of supervised learning methods for sperm analyses. For
364 instance, Holt [50] used discriminant analysis (a supervised classification system) to assign
365 cluster memberships to unclustered datasets, using an initial dataset that had been classified
366 using cluster analysis. In other two studies, Vulcano et al. [51] and, more recently, Goodson et

al. [42] examined sperm motility patterns under capacitating conditions for ram and for mouse sperm, respectively, using the SVM methodology.

Here, we examine the sperm distribution of males showing different freezability as a first step in the identification of sperm subpopulations, and then use this information to generate a SVM model for the characterization of other semen samples, different to those used to generate the model. Thus, this study proposes the use of different sources of information, in this case the SMI, membrane integrity and mitochondrial activity, as a prior knowledge for the characterization of sperm subpopulations. Individual sperm tracks from 6 stags with marked differences on their freezability have been assessed and assigned to different motility patterns, as described in Goodson et al. [42]. This information has been then used to generate a SVM model that clearly identifies and quantifies four distinct patterns of sperm movement in populations of Iberian red deer sperm: weak motile, progressive, transitional and hyperactivated-like.

We have evaluated how the SVM and the non-hierarchical clustering methods performed the characterization of sperm subpopulations. The SVM has been the most accurate method, with less than 5 % of sperm being miss-classified, and being the four subpopulations obtained from this method high significant correlated with the SMI, membrane integrity and mitochondrial membrane potential used to characterize the freezability of males. By contrast, non-hierarchical methods have showed errors above 20 %, and only one of the subpopulations (SP4) have showed significant correlations with those sperm parameters used to characterize the freezability of males. The characterization provided by these methods has not taken into account information on the differences in the distribution of sperm between poor and good freezers and, for that reason, has resulted to be a little different from the characterization performed by the SVM method. Thus, although we could find differences in the sperm

391 population structure defined in this way between males showing poor and good freezability,
392 this association would be less strong than that expected from the use of the supervised
393 learning method in which information on sperm motility distribution of males have been used
394 to develop the SVM model.

395 Thus, the SVM method has been used to characterize the sperm motile population
396 structure of frozen–thawed semen in the Iberian red deer, in order to find a subpopulations
397 distribution that best correlates with freezability of males. Motility data from 25 stags
398 showing differences on freezability were recorded and loaded into the SVM model previously
399 generated in the training step. The same four subpopulations with a weak motile, progressive,
400 transitional and hyperactivated-like motility patterns have been clearly represented in the
401 sperm population. Significant differences on the distribution of sperm among these four
402 subpopulations have been observed between males with different freezability. Thus, sperm
403 characterized by a weak motility pattern (SP1) were predominant in those males with poor
404 freezability, while for those males showing better freezability, higher percentages of
405 progressive and hyperactivated-like sperm were observed. Concerning to the transitional
406 subpopulation, the percentages increased as the sperm quality at thawing increase.

407 To further explore the relations between the four sperm subpopulations, and to explore
408 how an overall measure of the sperm population distribution could be useful to characterize
409 the freezability in the Iberian red deer, we have performed a principal component analysis
410 (PCA). The PCA rendered a first principal component accounting for 64.06 % of the total
411 variance, which could be interpreted as an indicator of non-vigorous movement. This factor
412 would be very useful to discriminate between males showing great differences on their
413 freezability, but could be less efficient in differentiating between males showing an average
414 freezability. To the latter, consider the percentage of sperm belonging to the transitional

subpopulation (SP3) would be of interest. This percentage has been mainly reflected on the second principal component, which has accounted for 25.69 % of the total variance and could be interpreted as indicator of the overall sperm motility. Thus, higher values of this second component have been found in those males showing better freezability

5. Conclusions

In conclusion, Support Vector Machines (SVM) has demonstrated to be very useful tools when we look for functional correlations between spermatozoa characteristics and freezability of males. In this study, the SVM method has performed the most accurate classification, being the subpopulations distribution obtained high significant correlated with those sperm characteristics used to characterize the sperm freezability of males. The characterization of the population structure of motile spermatozoa in the Iberian red deer using a SVM method has resulted in the identification of four subpopulations characterized by different motility patterns. These subpopulations have showed different distribution among males showing differences on their quality on thawed semen and, therefore, could be useful to characterize the freezability of males.

Here, we have used the SVM method to characterize the sperm motile population structure related with freezability. However, this method can also be useful for other purposes, among which highlight the study of fertility potential of males. Sperm fertility studies are now focused on analyses that incorporate multiple variables to examine how different sperm parameters interact to determine fertility. Semen samples must be subjected in parallel to several different tests and their outcome should be subjected to multiparametric analyses in order to provide the highest level of fertility prediction. And it will be within this field where the SMV method proposed in this study will provide and important support.

439

440 **REFERENCES**

441 [1] Petrunkina AM, Topfer-Petersen E. Heterogeneous osmotic behaviour in boar sperm
442 populations and its relevance for detection of changes in plasma membrane. Reprod Fertil
443 Dev 2000; 12:297-305.
444 [2] Paasch U, Grunewald S, Fitzl G, Glander HJ. Deterioration of plasma membrane is
445 associated with activated caspases in human spermatozoa. J Androl 2003; 24:246-252.
446 [3] Perez-Llano B, Yenes-Garcia P, Garcia-Casado P. Four subpopulations of boar
447 spermatozoa defined according to their response to the short hypoosmotic swelling test
448 and acrosome status during incubation at 37 degrees C. Theriogenology 2003; 60:1401-
449 1407.
450 [4] Buffone MG, Doncel GF, Marin Briggiler CI, Vazquez-Levin MH, Calamera JC. Human
451 sperm subpopulations: Relationship between functional quality and protein tyrosine
452 phosphorylation, Hum Reprod 2004; 19:139-146.
453 [5] Esteso MC, Soler AJ, Fernández-Santos MR, Quintero-Moreno A, Garde JJ. Functional
454 significance of the sperm head morphometric size and shape for determinig freezability in
455 iberian red deer (*Cervus elaphus hispanicus*) epididymal sperm simples, J Androl 2006;
456 27:662-670.
457 [6] Nuñez-Martínez I, Moran JM, Peña FJ. A three-step statistical procedure to identify sperm
458 kinematic subpopulations in canine ejaculates: Changes after cryopreservation. Reprod
459 Domest Anim 2006; 41:408-415.
460 [7] Martínez-Pastor F, Cabrita E, Soares F, Anel L, Dinis MT. Multivariate cluster analysis to
461 study motility activation of Solea senegalensis spermatozoa: a model for marine teleosts.
462 Reproduction 2008; 135:449-459.

463 [8] Druart X, Gatti JL, Huet S, Dacheux JL, Humblot P. Hypotonic resistance of boar
464 spermatozoa: sperm subpopulations and relationship with epididymal maturation and
465 fertility. Reproduction 2009; 137:205-213.
466 [9] Muiño R., Peña AI, Rodríguez A, Tamargo C, Hidalgo CO. Effects of cryopreservation on
467 the motile sperm subpopulations in semen from Asturiana de los Valles bulls.
468 Theriogenology 2009; 72:860-868.
469 [10] Harrison R. Capacitation mechanisms and the role of capacitation as seen in eutherian
470 mammals. Reprod Fertil Dev 1996; 8:581–594.
471 [11] Petrunkina AM, Waberski D, Günzel-Apel AR, Töpfer-Petersen E. Determinants of
472 sperm quality and fertility in domestic species. Reproduction 2007; 134:3-17.
473 [12] Martínez-Pastor F, Garcia-Macias V, Álvarez M, Herraez P, Anel L, de Paz P. Sperm
474 subpopulations in Iberian Red Deer epididymal sperm and their changes through the
475 cryopreservation process. Biol Reprod 2005; 72:316-327.
476 [13] Asher GW, Berg DK, Evans G. Storage of semen and artificial insemination in deer.
477 Anim Reprod Sci 2000; 62:195–211.
478 [14] Holt WV. Fundamental aspects of sperm cryobiology: the importance of species and
479 individual differences. Theriogenology 2000; 53:47–58.
480 [15] Thurston LM, Watson PF, Holt WV. Sperm cryopreservation: a genetic explanation for
481 species and individual variation. Cryo Letters 2002; 23:255-262.
482 [16] Soler AJ, García AJ, Fernández-Santos MR, Esteso MC, Garde JJ. Effects of thawing
483 procedure on postthawed in vitro viability and in vivo fertility of red deer epididymal
484 spermatozoa cryopreserved at –196 ºC. J Androl 2003; 24:746-756.

485 [17] Martínez-Pastor F, Tizado EJ, Garde JJ, Anel L, de Paz P. Statistical Series:
486 Opportunities and challanges of sperm motility subpopulation análisis. Theriogenology
487 2011; 75:783-795.
488 [18] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining,
489 inference and prediction, second edition, Ed Springer Series in Statistics, 2009. ISBN:
490 978-0-387-84857-0.
491 [19] Weinberger KQ, Saul LK. Distance metric learning for large margin nearest neighbor
492 classification. The Journal of Machine Learning Research 2009; 10:207-244.
493 [20] Perlich C, Provost F, Simonoff JS. Tree induction vs. logistic regression: a learning
494 curve analysis. The Journal of Machine Learning Reseach 2003; 4:211-255.
495 [21] Esmeir S, Markovitch S. Anytime Learning of Decision Trees. The Journal of Machine
496 Learning Research 2007; 8:981-933.
497 [22] Scheinberg K. An efficient implementation of an active set method for SVMs. The
498 Journal of Machine Learning Research 2006; 7:2237–2257.
499 [23] Ripley BD. Pattern Recognition and Neural Networks, Cambridge University Press,
500 2008. ISBN-13: 9780521717700.
501 [24] Nobble WS. What is a support vector machine? Nature Biotechnology 2006; 24:1565-
502 1567.
503 [25] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and other
504 kernel-based learning methods. Cambridge University Press, Cambridge, UK, 2000.
505 [26] Noble WS. Support vector machine applications in computational biology, in:
506 Schoelkopf B, Tsuda K, Vert JP (Eds.), Kernel Methods in Computational Biology, MIT
507 Press, Cambridge, MA, 2004, pp. 71-92.

508 [27] Liao L, Noble WS. Combining pairwise sequence similarity and support vector
509 machines for remote protein homology detection, in: Proceedings of the Sixth Annual
510 International Conference on Computational Molecular Biology, 2002, pp. 225-232.
511 [28] Leslie C, Kuang R. Fast string kernels using inexact matching for protein sequences.
512 The Journal of Machine Learning Research 2004; 5:1435-1455.
513 [29] Vert JP. A tree kernel to analyze phylogenetic profiles. Bioinformatics 2002; 18:276-
514 284.
515 [30] Leslie C, Eskin E, Cohen A, Weston J, Noble WS. Mismatch string kernels for
516 discriminative protein classification. Bioinformatics 2004; 20:467-476.
517 [31] Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet C, Furey TS, Ares JM, Haussler
518 D. Knowledge-based analysis of microarray gene expression data using support vector
519 machines. Proc Natl Acad Sci U.S.A. 2000; 97:262-267.
520 [32] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh
521 ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of
522 cancer: class discovery and class prediction by gene expression monitoring. Science 1999;
523 286:531–537.
524 [33] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using
525 support vector machines. Mach Learn 2002; 46:389–422.
526 [34] García AJ, Landete-Castillejos T, Garde JJ, Gallego L. Reproductive seasonality in
527 female Iberian red deer (*Cervus elaphus hispanicus*). Theriogenology 2002; 58:1553-
528 1562.
529 [35] Garde JJ, Ortiz N, García AJ, Gallego L, Landete-Castillejos T, López A. Postmortem
530 assessment of sperm characteristics of the red deer during the breading season. Arch
531 Andol 1998; 41:195-202.

532 [36] Comizzioli P, Mauget R, Mermillod P. Assessment of in vitro fertility of deer
533 spermatozoa by heterologous IVF with zona-free bovine oocytes. Theriogenology 2001;
534 56:261-274.
535 [37] Martínez-Pastor F, Fernández-Santos MR, Del Olmo E, Domínguez-Rebolledo AE,
536 Esteso MC, Montoro V, Garde JJ. Mitochondrial activity and forward scatter vary in
537 necrotic, apoptotic and membrana-intact spermatozoa subpopulations. Reprod Fertil Dev
538 2008; 20:547-556.
539 [38] Martínez-Pastor F, Martínez F, Álvarez M, Maroto-Morales A, García-Álvarez O, Soler
540 AJ, Garde JJ, de Paz P, Anel L. Cryopreservation of Iberian red deer (*Cervus elaphus*
541 *hispanicus*) spermatozoa obtained by electroejaculation. Theriogenology 2009; 71:628-
542 638.
543 [39] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster
544 analysis. J Comput Appl Math 1987; 20:53-65.
545 [40] R Development Core Team, R: A language and environment for statistical computing, R
546 Foundation for Statistical Computing, Vienna, Austria. 2010. ISBN 3-900051-07-0.
547 http://www.R-project.org.
548 [41] Haldiki M, Batistakis Y, Vazirgiannis M. Cluster validity methods, SIGMOD. Record
549 2002; 31:40-45.
550 [42] Goodson SG, Zhang Z, Tsuruta JK, Wang W, O'Brien DA. Classification of mouse
551 sperm motility patterns using an automated multiclass support vector machines model.
552 Biol Reprod 2011; 84:1207-1215.
553 [43] Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. e1071: Misc Functions of
554 the Department of Statistics (e1071), TU Wien. R package version 1.5-22. 2009.
555 http://CRAN.R-project.org/package=e1071.

556 [44] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001.
557 http://www.csie.ntu.edu.tw/~cjlin/libsvm .
558 [45] Thurston LM, Watson PF, Milejam AJ, Holt WV. Morphologically distint sperm
559 subpopulations defined by Fourier shape descriptors in fresh ejaculates correlate with
560 variation in boar semen quality following cryopreservation. J Androl 2001; 22:382-394.
561 [46] Cremades T, Roca J, Rodríguez-Martínez H, Abaigar T, Vázquez JM, Martínez EA.
562 Kinematic changes during the cryopreservation of boar spermatozoa. J Androl 2005;
563 26:610-618.
564 [47] Martínez-Pastor F, Díaz-Corujo AR, Anel E, Herraez P, Anel L, de Paz P. Post mortem
565 time and season alter subpopulation characteristics of Iberian reed deer epididymal sperm.
566 Theriogenology 2005; 64:958-974.
567 [48] Beirão J, Cabrita E, Pérez-Cerezales S, Martínez-Páramo S, Herráez MP. Effect of
568 cryopreservation on fish sperm subpopulations. Cryobiology 2011; 62:22-31.
569 [49] Dorado J, Alcaraz L, Duarte N, Portero JM, Acha D, Hidalgo M. Changes in the
570 structure of motile sperm subpopulations in dog spermatozoa after both cryopreservation
571 and centrifugation on PureSperm® gradient. Anim Reprod Sci 2011; 125:211-218.
572 [50] Holt WV. Can we predict fertility rates? Making sense of sperm motility. Reprod
573 Domest Anim 1996; 31:17.24.
574 [51] Vulcano GJ, Moses DF, Valcarcel A, de las Heras MA. A lineal equation for the
575 classification of progressive and hyperactive spermatozoa. Math Biosci 1998; 149:77-93.

FIGURE LEGENDS

Figure 1. Characterization of semen quality of males after the freezing-thawing procedure in the Iberian red deer. Numbers indicate males used in this study. Six males with marked differences on their freezability (closed circles) were used to train de SVM model and to test the accuracy of different statistical methods. (A) Relation between the proportion of viable spermatozoa with a stable membrane (YO-PRO-1–/PI–) and the sperm motility index (SMI). (B) Relation between the proportion of viable spermatozoa with a stable membrane and proportion of spermatozoa with high mitochondrial membrane potential (Mitotraker+/YO-PRO-1–).

Figure 2. Sperm motility characteristics for the 6 Iberian red deer stags used to train the SVM and to test the accuracy of different statistical methods (A). Four different motility patterns are identified: weak motile (SP1), progressive (SP2), transitional (SP3) and hyperactivated-like (SP4). Data is also represented for the group of males with poor (B) and good (C) freezability, separately.

Figure 3. Sperm subpopulation characterization of the data in the *testing set* using the SVM method (A) and the non-hierarchical (B), hierarchical (C) and multi-step clustering procedures (D). Four different motility patterns are identified: weak motile (SP1), progressive (SP2), transitional (SP3) and hyperactivated-like (SP4).

Figure 4. Males distribution in the multidimensional ordination space defined by the first two principal components from the PCA analysis. Numbers indicate males used in this study. Arrows indicate vectors representing the four motility patterns/subpopulations.

600 Table 1. Seminal parameters after freezing and thawing of spermatozoa from Iberian red deer stags.[*,**]

601

	N	SMI (%)	YO-PRO-1-/PI- (%)	YO-PRO-1+/PI+ (%)	MT+/YO-PRO-1- (%)
All males	25	39.30 ± 2.93	39.67 ± 2.63	41.22 ± 2.98	38.98 ± 3.03
Bad Freezers	3	15.00 ± 2.50	20.65 ± 1.37	61.61 ± 2.29	20.17 ± 1.27
Good Freezers	3	52.50 ± 1.44	55.76 ± 0.25	25.13 ± 1.71	55.76 ± 4.04

602 [*] Data are mean ± SEM

603 [**]All parameters differed significantly among different males' groups ($p<0.05$)

604 Table 2. Kinematics parameters for the four sperm subpopulations defined in the SVM training step.[*,†]

605

	VCL (μm/s)	VSL (μm/s)	VAP (μm/s)	LIN (%)	STR (%)	ALH (μm)	BCF (Hz)
SP1	31.44 ± 0.80	7.87 ± 0.27	14.13 ± 0.36	24.29 ± 0.51	53.25 ± 0.75	1.72 ± 0.03	3.79 ± 0.13
SP2	110.39 ± 2.85	60.93 ± 1.52	78.41 ± 1.88	56.12 ± 0.79	78.07 ± 0.84	3.97 ± 0.11	9.10 ± 0.27
SP3	114.63 ± 1.85	35.44 ± 0.79	61.02 ± 0.95	31.17 ± 0.50	58.41 ± 0.97	4.69 ± 0.09	9.52 ± 0.27
SP4	168.77 ± 1.96	34.59 ± 0.93	102.01 ± 0.98	20.48 ± 0.49	33.80 ± 0.81	6.42 ± 0.10	9.24 ± 0.22

606 [*] Data are mean ± SEM

607 [†] SP1: weakly motile; SP2: progressive; SP3: transitional; SP4: hyperactivated-like

608

609

610 Table 3. Confusion matrix for predictions in the testing set. For different classification methods, data on diagonal (in bold) represents events (no.

611 of spermatozoa) that were correctly labeled. The error rate is also presented. [*,†]

612

	Non-Hierarchical				Hierarchical				Multi-step				SVM			
	actual				Actual				Actual				actual			
predicted	SP1	SP2	SP3	SP4	SP1	SP2	SP3	SP4	SP1	SP2	SP3	SP4	SP1	SP2	SP3	SP4
SP1	**239**	0	2	0	**118**	0	0	0	**256**	1	9	0	**266**	2	8	0
SP2	0	**54**	6	7	19	**42**	98	48	0	**43**	5	17	0	**81**	1	1
SP3	28	37	**83**	20	130	1	**3**	18	11	46	**90**	35	1	6	**106**	7
SP4	0	0	28	**145**	0	48	18	**137**	0	1	15	**120**	0	2	4	**164**
Error[†]	0.197				0.537				0.216				0,049			

613 [*] Error = *1 – (sum of confusion matrix diagonal / Number of observations)*

614 [†] SP1: weakly motile; SP2: progressive; SP3: transitional; SP4: hyperactivated-like

615

616

617 Table 4. Correlations among those sperm characteristics used to classify the six males in the training set as *"poor"* or *"good"* freezers and the

618 percentages of the sperm motile subpopulations of males. [*,**]

619

	SVM			
	SP1 (%)	SP2 (%)	SP3 (%)	SP4 (%)
SMI (%)	– 0.75	0.69	**0.81**	**0.81**
YO-PRO-1-/ PI- (%)	**– 0.83**	0.58	**0.89**	**0.81**
Mitotracker+/ YO-PRO-1- (%)	**– 0.94**	**0.81**	0.77	**0.93**

	NHC			
	SP1 (%)	SP2 (%)	SP3 (%)	SP4 (%)
SMI (%)	– 0.58	**0.81**	0.26	**0.81**
YO-PRO-1-/ PI- (%)	– 0.31	0.64	– 0.20	**0.81**
Mitotracker+/ YO-PRO-1- (%)	– 0.09	0.75	– 0.20	**0.93**

	HC			
	SP1 (%)	SP2 (%)	SP3 (%)	SP4 (%)
SMI (%)	– 0.58	0.53	– 0.58	**0.81**
YO-PRO-1-/ PI- (%)	– 0.31	0.12	– 0.43	**0.81**
Mitotracker+/ YO-PRO-1- (%)	– 0.09	– 0.06	– 0.03	**0.93**

	MS			
	SP1 (%)	SP2 (%)	SP3 (%)	SP4 (%)
SMI (%)	– 0.58	**0.89**	0.46	**0.81**
YO-PRO-1-/ PI- (%)	– 0.31	0.76	0.09	**0.81**
Mitotracker+/ YO-PRO-1- (%)	– 0.09	0.70	– 0.03	**0.93**

620 * Spearman correlation coefficients are presented

621 ** Significant correlations ($p<0.05$) are represented in bold.

622 Table 5. Sperm characteristics of the four motile subpopulations identified in the Iberian red deer.[*, †]

623

	N (%)[‡]	VCL (μm/s)	VSL (μm/s)	VAP (μm/s)	LIN (%)	STR (%)	ALH (μm)	BCF (Hz)
SP1	56.99 ± 19.67	34.73 ± 0.36	9.13 ± 0.13	16.41 ± 0.18	25.68 ± 0.23	53.42 ± 0.31	1.83 ± 0.02	4.42 ± 0.06
SP2	10.84 ± 6.75	105.80 ± 1.19	59.23 ± 0.64	75.27 ± 0.77	56.96 ± 0.38	79.08 ± 0.39	3.67 ± 0.04	9.82 ± 0.13
SP3	20.77 ± 9.45	114.57 ± 0.74	35.02 ± 0.33	60.40 ± 0.39	30.87 ± 0.23	58.35 ± 0.43	4.75 ± 0.04	9.74 ± 0.10
SP4	11.40 ± 11.56	159.15 ± 1.06	31.75 ± 0.52	100.53 ± 0.53	10.79 ± 0.28	31.44 ± 0.47	5.92 ± 0.05	9.72 ± 0.13

624 [*] Data are mean ± SEM

625 [†] SP1: weakly motile; SP2: progressive; SP3: transitional; SP4: hyperactivated-like

626 [‡] Summary of the percentages of subpopulations found on each male (mean ± SD).

Figure 1.

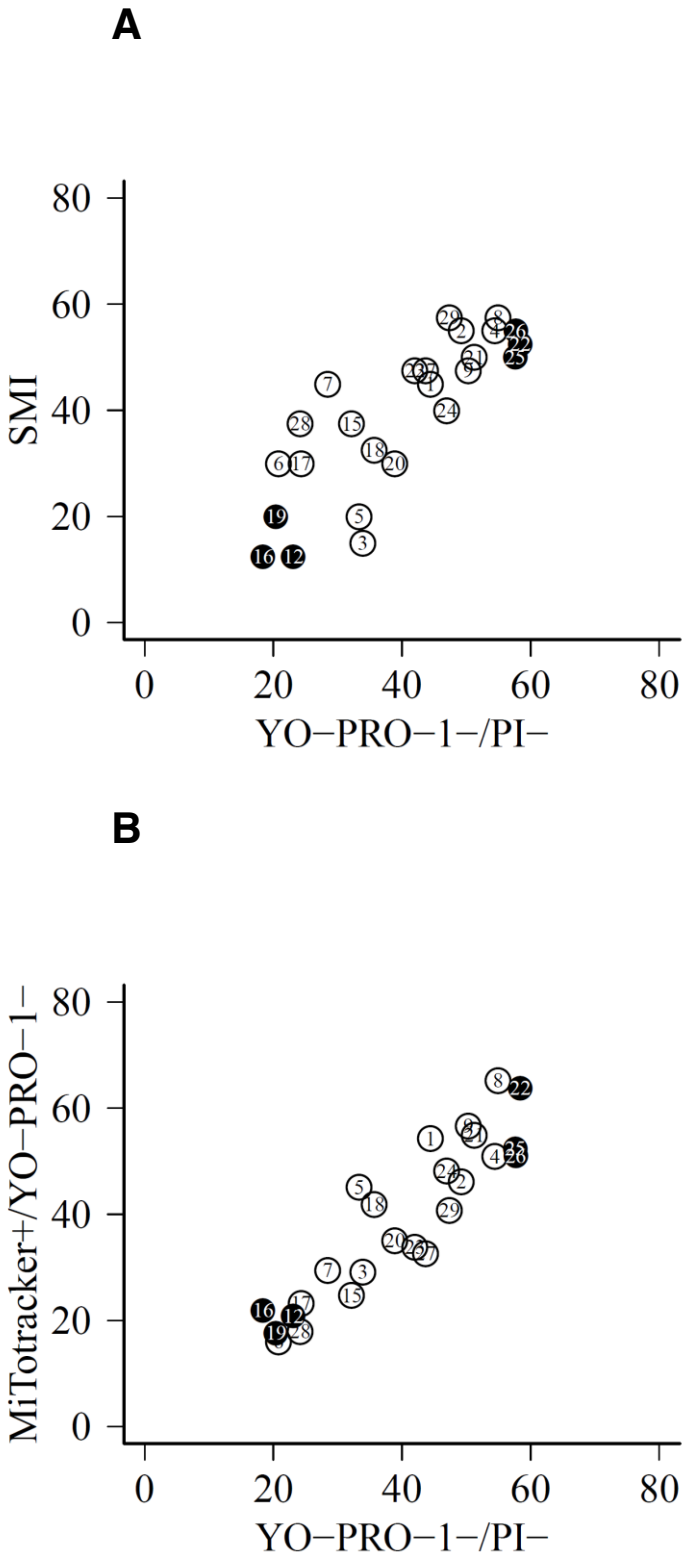


Figure 2.

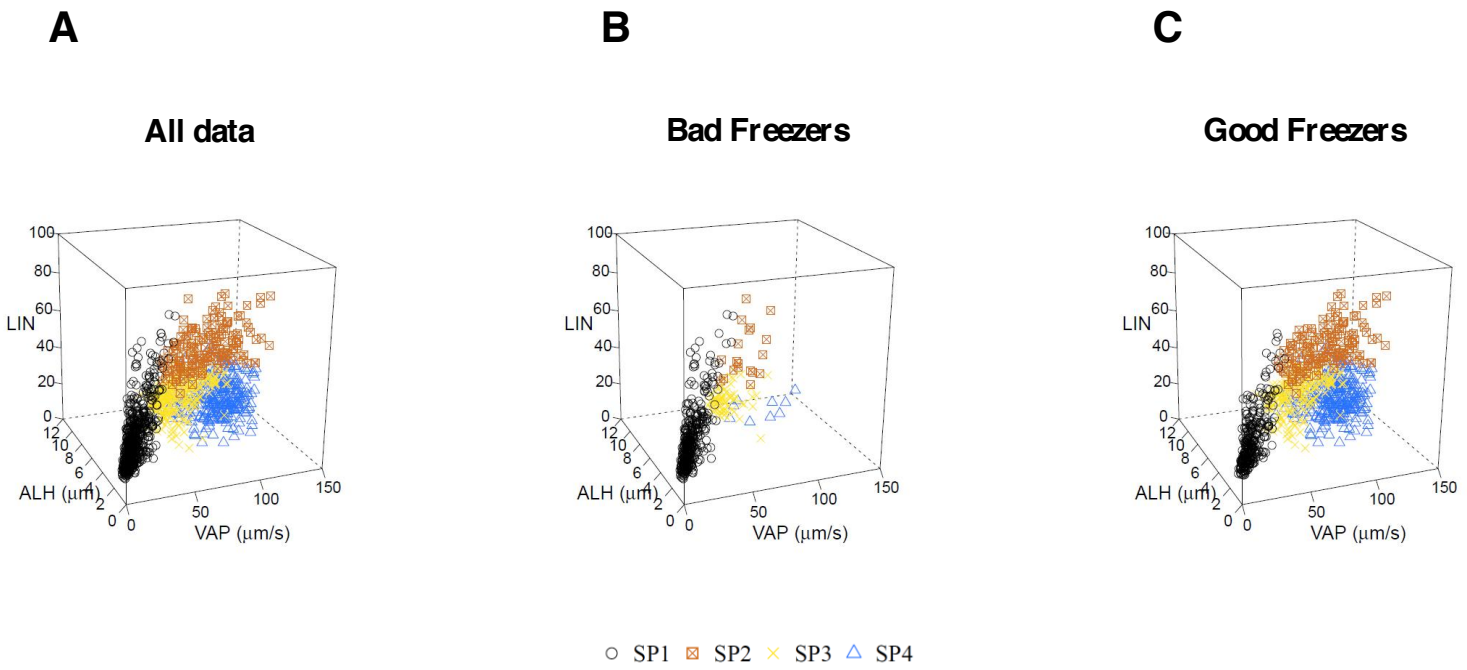

A
All data
B
Bad Freezers
C
Good Freezers
LIN
ALH (µm)
VAP (µm/s)
SP1
SP2
SP3
SP4

Figure 3.

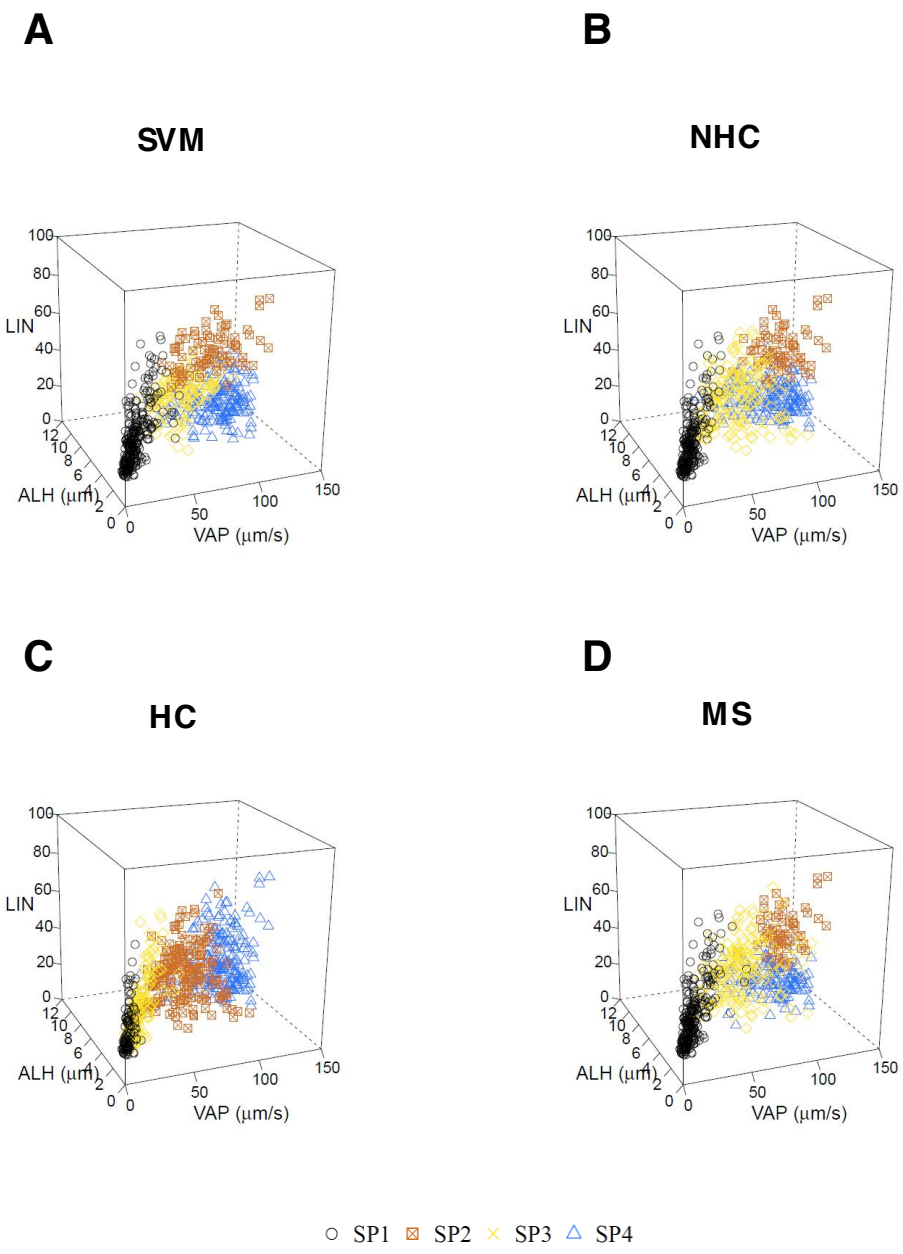
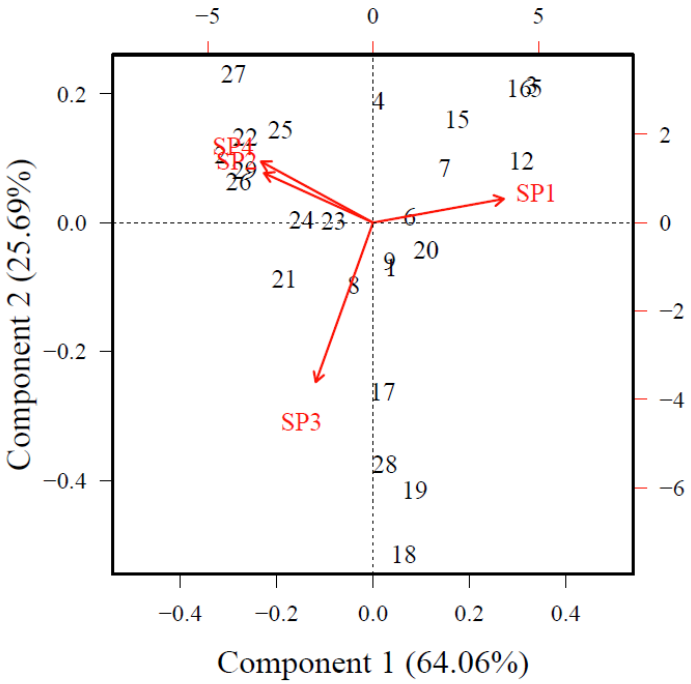


Figure 4.


−5
0
5
0.2
0.0
−0.2
−0.4
2
0
−2
−4
−6
27
4
15
25
22
7
12
SP1
26
24
23
6
20
9
21
8
17
SP3
28
19
18
−0.4
−0.2
0.0
0.2
0.4
Component 2 (25.69%)
Component 1 (64.06%)

Figure

Figure 2. Ramón *et al.* 2011

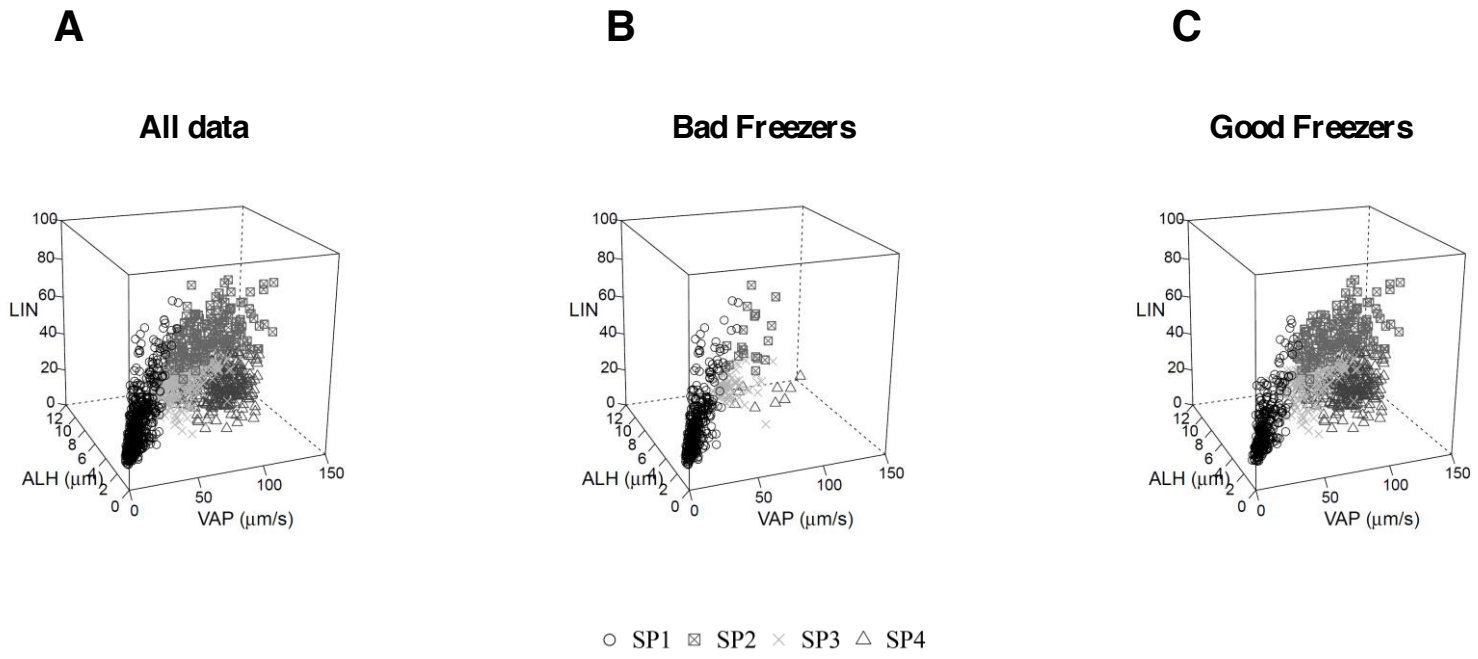
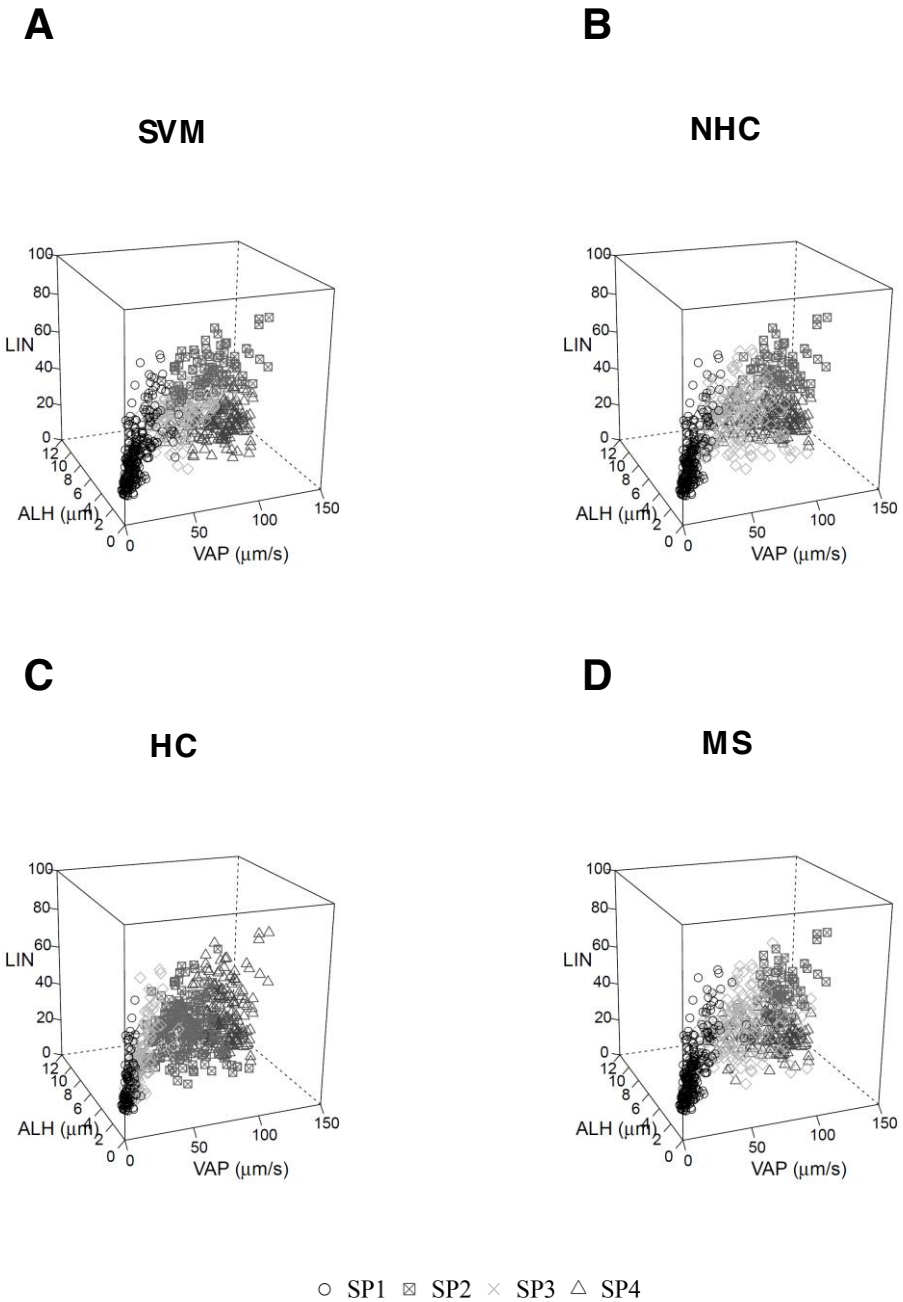


Figure 3. Ramón *et al.* 2011



Theriogenology
Author Agreement Form

This Form should be signed by all authors OR by the corresponding (or senior) author who can vouch for all co-authors. A scanned copy of the completed Form may be submitted online.
Alternatively, the completed Form may be faxed to the relevant Editor:
F. Gandolfi +3902 5031 7980; J. Kastelic +1 403 382 3156

The authors confirm the following statements:

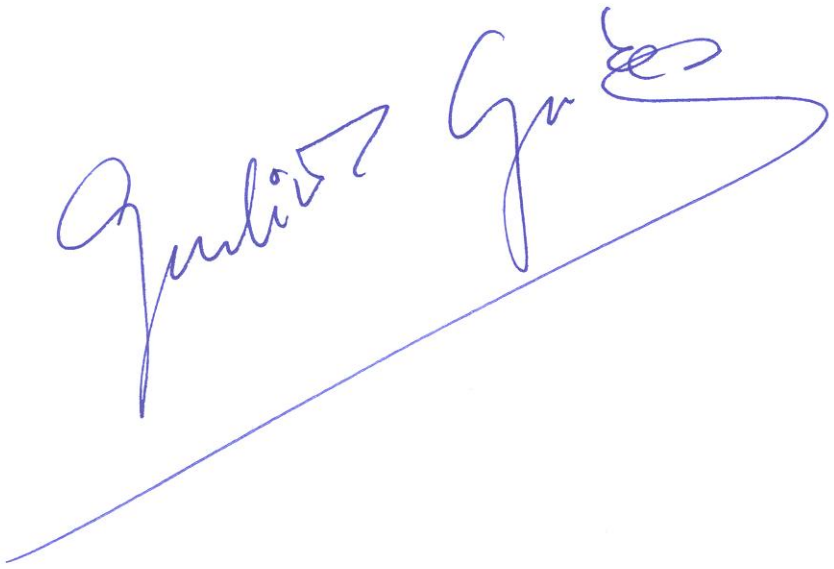
1. that there has been no duplicate publication or submission elsewhere of this work

2. that all authors have read and approved the manuscript, are aware of the submission for publication and agree to be listed as co-authors

Author Name **Signature**

JJ GARDE (Senior author)

_____



_ _____

_____