ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

ORIGINAL ARTICLE

Correspondence:
M. Cruz Gil Anaya, Research Team of Intracellular Signaling and Technology of Reproduction (SINTREP), Faculty of Veterinary Medicine, University of Extremadura, Avda. de la Universidad s/n, 10003 Cáceres, Spain.
E-mail: crgil@unex.es

Keywords:
Bayesian networks, motility analysis, semen, sperm subpopulations, tench

Received: 8-Apr-2015
Revised: 26-May-2015
Accepted: 10-Jun-2015

doi: 10.1111/andr.12071

A new Bayesian network-based approach to the analysis of sperm motility: application in the study of tench (*Tinca tinca*) semen

[1]M. C. Gil Anaya, [2]F. Calle, [2]C. J. Pérez, [1]D. Martín-Hidalgo, [3]C. Fallola, [1]M. J. Bragado, [1]L. J. García-Marín and [4]A. L. Oropesa

[1]Research Team of Intracellular Signaling and Technology of Reproduction (SINTREP), [2]Biostatistics Unit, Department of Mathematics, Faculty of Veterinary Medicine, University of Extremadura, Cáceres, [3]Aquaculture Center, Dirección General de Medio Natural, Gobierno de Extremadura, and [4]Toxicology Area, Animal Health Department, Sciences Faculty, University of Extremadura, Badajoz, Spain

SUMMARY

In this study a Bayesian network (BN) has been built for the study of the objective motility of *Tinca tinca* spermatozoa (spz). Semen from eight 2-year-old sexually mature male tenchs was obtained and motility analyses were performed at 6–17, 23–34 and 40–51 s after activation, using computer-assisted sperm analysis (CASA) software. Motility parameters rendered by CASA were treated with a two-step cluster analysis. Three well-defined sperm subpopulations were identified, varying the proportion of spermatozoa contained in each cluster with time and male. Cluster, cinematic and time variables were used to build the BN to study the probabilistic relationships among variables and how each variable influenced the final sperm classification into one of three predefined clusters. Both network structure and conditional probabilities were calculated based on the collected data set. Results shown that almost all the variables were directly or indirectly related to each other. By doing probabilistic inference we observed that the cluster distribution corresponded to the definition provided by the cluster analysis. Also, velocity and time variables determined the cluster to which each spermatozoon belonged with a high degree of accuracy. Thus, BNs can be applied in the study of sperm motility. The construction of a BN that include fertility data opens a new way to try to clarify the roles of motility and other sperm quality indicators in fertilization.

INTRODUCTION

In aquaculture, as in other production systems, one of the limiting factors of reproductive success is the sperm quality, which is influenced by several factors (Bobe & Labbé, 2010; Fauvel *et al.*, 2010). Evaluations of sperm quality are needed to increase the efficiency of artificial insemination, particularly in commercial aquaculture (Rurangwa *et al.*, 2004). One of the most commonly used indicators for the assessment of sperm quality in all species is motility because this parameter reflects several attributes of sperm functionality, including the integrity of the plasma membrane and the availability of ATP (and thus mitochondrial functionality), and the functionality of the flagella (Beirão, 2011). It is accepted that highly motile spermatozoa have greater likelihoods of fertilizing eggs. For example Cosson *et al.* (2008b) observed that decreases in the motilities of sea bass and turbot spermatozoa are accompanied by declines in fertilization abilities.

In fish, evaluations of the percentages of motile spermatozoa and/or the durations of their motility have been performed using various methods, such as phase contrast and dark-field microscopy and computer-assisted sperm analysis (CASA) systems (reviewed by Rurangwa *et al.*, 2004). The use of CASA systems is preferable to the use of subjective visual estimations of motility because these systems provide more reliable, repeatable and objective measurements of sperm movement. These systems are well adapted to the sperm cells of fish, which are characterized by a short period of motility after activation in the majority of the species used for aquaculture (approximately 1 min). Wilson-Leedy & Ingermann (2007) developed a CASA system based on open source software for the characterization of zebrafish sperm motility parameters. CASA systems have proven to be valuable tools in aquaculture for studies of the effects of pollutants on motility (Kime *et al.*, 1996; Rurangwa *et al.*, 1998; Dietrich *et al.*, 2010), the improvement of cryopreservation protocols (Rurangwa *et al.*, 2001; Beirão *et al.*, 2011) and the choices of

extenders (Rurangwa *et al.*, 2001; Martínez-Páramo *et al.*, 2012) and activators of motility (Kime & Tveiten, 2002; Martínez-Pastor *et al.*, 2008; Kanuga *et al.*, 2012). Because CASAs record the movement of each sperm cell, large amounts of data are collected from each semen sample. The application of cluster analysis to such data provides more information that can be acquired solely based on the average values provided by CASA systems. Cluster analyse group spermatozoa with similar motility characteristics, which allows analyses of different sperm subpopulations within a single sample. Variations in the distributions of these subpopulations across several animal species have been associated with sperm freezability (Martínez-Pastor *et al.*, 2005; Flores *et al.*, 2009; Muino *et al.*, 2009; Beirão *et al.*, 2011), individual variations between ejaculates and males (Nunez-Martinez *et al.*, 2006), sperm fertility (Quintero-Moreno *et al.*, 2003; Beirão *et al.*, 2011) and the effects of ejaculation (Contri *et al.*, 2012) or the activation solution (Kanuga *et al.*, 2012).The presence of a rapid and linear subpopulation has been proposed to be an indicator of seminal quality (Martínez-Pastor *et al.*, 2005, 2011; Ferraz *et al.*, 2014); thus, analyses of motility based on sperm subpopulations may enable the identification of males with good sperm quality. However, despite the advantages provided by the information obtained trough cluster analysis, this method is still rarely applied to the analysis of the motility of fish spermatozoa.

Analysis and data modelling techniques have been improved with great interest because they allow the extraction of information regarding associations between variables, patterns, correlations, etc. One of these techniques, specifically within the probabilistic expert systems that are integrated in artificial intelligence, is probabilistic modelling with Bayesian networks (BNs). In recent years, BNs have shown their potential as models of knowledge representation that include uncertainty (Daly *et al.*, 2011). A BN defines the structure of the relations of dependency and joint probability distribution for a set of random variables. The use of BNs is particularly interesting for both the relational structures that might be discovered among variables and the probability distributions that this information provides, which may be used to calculate and update the marginal probabilities based on the information provided to the network (i.e. evidence); that is these networks allow the investigator to perform processes of inference (Jensen & Nielsen, 2007). This data modelling technique uses a graphical representation to explain the model, which makes it a very attractive knowledge representation tool in addition to being an intuitive, compact and robust representation.

Because BN models describe independence/dependence relationships between variables, these networks can be applied to almost any type of problem. BNs are currently being used in areas such as economics (Ngai *et al.*, 2011), biology (Wong & Li, 2006) and medicine (Yet *et al.*, 2014). In the area of human medicine, BNs are effective and reliable tools that are used in decision-making processes related to certain pathologies and aid the diagnoses and choices of medical treatments (Lucas *et al.*, 2004). In contrast, the application of BNs in the field of veterinary medicine remains relatively limited, but BNs have been found to be powerful and intuitive tools for investigations of animal health data (Lewis *et al.*, 2011; Ward & Lewis, 2013). Analyses of animal health data may involve rather complicated tasks because of the nature of such data; that is many variables are interrelated. Classical statistical methods cannot discriminate between direct and indirect associations, and BNs are thus used for this purpose (Korb & Nicholson, 2010). In this study, we applied BNs in the area of animal reproduction; specifically, this study provides the first application of BNs to the study of the objective motility of spermatozoa. The fish *Tinca tinca* was chosen as an animal model for two main reasons: few studies about sperm motile subpopulations have been conducted in fish, and because post-activation time (variable included in this study for the construction of the BN) is an important factor for motility parameters in the case of fish spermatozoa.

The aim has been to assess the direct and indirect relationships between the variables of motility and other variables of interest (cluster and time) to classify semen samples into different groups. We previously conducted a cluster analysis to identify sperm subpopulations in an attempt to broaden the understanding of the physiology of fish spermatozoa and to serve as a basis for other studies. This study concludes with the proposal of future applications of BNs in the field of sperm quality analysis.

MATERIALS AND METHODS

Animals

Eight 2-year-old sexually mature male tenchs were used. The broodstock was assessed for maturity via abdominal compression, and spermiation was detected by sperm production (Linhart *et al.*, 2003; Rodina *et al.*, 2004). The fish had an average weight of 335.04 ± 194.07 g (mean ± SD) and an average length of 23.93 ± 5.04 cm (mean ± SD) and were bred and housed in an outdoor tank (3 × 3 × 1 m) at the Aquaculture Center 'Vegas del Guadiana' (38°53′22.1″N 6°52′37.3″W; Gobierno de Extremadura, Badajoz, Spain). The flow of water in the tank was maintained at 40 L/h to ensure that 10% of the water was refreshed each day. The daily food (1% of the mean body weight) consisted of a commercial pellet diet for cyprinids (4.5 mm; Dibaq Cyprinids, Segovia, Spain) with the following analytical constituents and additives: crude protein (42%), total ashes (13.50%), crude fibre (5%), crude fat (6%), phosphorous (1.5%), vitamin A (10 000 UI/kg), vitamin D3 (1700 UI/kg), vitamin E (200 UI/kg) and copper (11 mg/kg).

Sperm collection

The experimental design was approved by the Ethical Committee of the University of Extremadura and all of the work was performed in accordance with the ethical requirements of the current legislation (European Parliament and Council Directive 2010/63/UE). Semen samples were collected during the reproductive season. Spermiation was not hormonally stimulated. Prior to the semen collection, the fish were anaesthetized in a water bath containing 100 mg/L tricaine methane sulphonate (MS222; Sigma-Aldrich Química, S. L., Madrid, Spain). To obtain the semen, the urogenital pore was cleaned of mucus, faeces and water (Beirão *et al.*, 2009). The urogenital pore was dried with a paper towel before pressure was applied to the abdomen to collect semen using a 5-cm^3 syringe without a needle. The spermatozoa were diluted 1 : 2 (semen/extender) in a Kurokura 180 immobilizing solution (180 mM NaCl, 2.68 mM KCl, 1.36 mM $CaCl_2$•$2H_2O$ and 2.38 mM $NaHCO_3$, Panreac Química S.L.U., Barcelona, Spain; Rodina *et al.*, 2004) to prevent spontaneous

sperm activation because of possible contamination by urine (Linhart *et al.*, 2003). However, care was taken to avoid such contamination. The spermatozoa in the immobilizing solution were stored and transported on ice to the laboratory of the Faculty of Veterinary Medicine in Cáceres, Spain (transport took approximately 1 h).

Semen analyses

The eight semen samples were processed immediately after arrival to the laboratory and independently for each male. Seminal parameters evaluated were sperm concentration, motility by CASA system and viability by flow cytometry. Sperm concentration was evaluated by light microscopy (200×) in a Bürker chamber after 1/200 dilution in 10% formaldehyde solution. For the CASA motility analyses, the spermatozoa were activated with distilled water in an amount that the final semen concentration was about 30×10^6 spz/mL, in a final volume of 1 mL. For each male, two separate sperm activations were conducted for fresh samples. Immediately after dilution and rapid mixing, 2 μL of semen sample was placed in a counting chamber (Leja, Luzernestraat, The Netherlands) at room temperature (20–25 °C). Sperm motility was assessed with a microscope (Nikon eclipse E200, Tokio, Japan) equipped with a 10× negative-phase contrast objective and an attached Basler A3121 digital camera (Basler Vision Technologies, Ahrensburg, Germany). Images were captured and analysed using the Integrated System for Semen Analysis Software (Proiser R+D; Paterna, Valencia, Spain). A fully stabilized image was obtained at 6 s after motility activation. Sperm motility parameters were measured at three time intervals, 6–17, 23–34 and 40–51 s after activation. By each time interval three fields were analysed, including each of them at least 200 spermatozoa. Straight line velocity (VSL), curvilinear velocity (VCL), average path velocity (VAP), linearity (LIN = VSL/VCL), straightness (STR = VAP/VCL), beat-cross frequency (BCF), wobble (WOB), amplitude of lateral head displacement (ALH) and percentage of motile spermatozoa were the motility parameters assessed. The CASA settings were: 25 frames/s for acquisition, VCL > 10 μm/s to classify a spermatozoon as motile and 5–80 μm^2 for head area.

For the sperm viability analysis by flow cytometry, fluorescent staining using the LIVE/DEAD Sperm Viability Kit was used (Aparicio *et al.*, 2007). Briefly, 5 μL of SYBR-14 (2 μM) and 10 μL of propidium iodide (PI 5 μM) were added to 500 μL of diluted semen sample (30×10^6 cells/mL) in isotonic-buffered diluent Coulter Isoton II and incubated for 20 min at room temperature in the dark. After incubation, the cells were analysed and the percentage of viable spermatozoa was expressed as the percentage of SYBR14-positive and propidium iodide-negative spermatozoa. Flow cytometry analyses were performed using a Coulter EPICS XL-MCL flow cytometer (Beckman Coulter, Fullerton, CA, USA.) The fluorophores were excited by a 200 mV argon ion laser operating at 488 nm. A total of 10 000 gated events based on the forward scatter and side scatter of the sperm population recorded in the linear mode were collected per sample with a running rate of approximately 500 events/s. The fluorescence data were collected in the logarithmic mode and analysed using a FACStationTM and EXPOTM 32 ADC software (Beckman Coulter Inc.).

Statistical analysis

In this study, classical statistical techniques have been used, specifically, a descriptive analysis and a cluster analysis. Moreover, a Bayesian technique was used to study the relationships between the variables of interest and the conditional probabilities of the variables (BNs).

To evaluate whether there were any statistically significant differences or relationship between the variables, different hypothesis tests were conducted. The Chi-square and Fisher's exact tests were used to identify possible associations between the qualitative variables. To observe differences in the means between the different levels of a variable, ANOVAs were used when the assumptions of homoscedasticity, normality and independence were met. When these conditions were not satisfied, the Kruskal–Wallis test was applied. For pair-wise comparisons, Bonferroni tests were used except when the variables were not normally distributed; in these cases, a penalized version of the Mann–Whitney *U* test was used. The variability among males in terms of the percentages of motile spermatozoa was assessed using a *t*-test, and Bonferroni corrections were used to adjust all pair-wise comparisons. Analyses of the correlations between the kinematic variables were also conducted with the Spearman rank correlation coefficient. To perform the cluster analysis, the data were first processed through a principal component (PC) analysis. The number of PCs was decided using the Kaiser criterion, and only those with eigen values equal to or >1 were selected. Finally, a two-step cluster analysis was performed to calculate the number of clusters using the previously identified PCs (the first step utilized the BIRCH algorithm, and the second step utilized a hierarchical algorithm; Martínez-Pastor *et al.*, 2011). Both the descriptive and cluster analyses were performed with SPSS release 19 (IBM Corporation, Armonk, NY, USA).

Subsequently, using the BNs, the results were compared in terms of the relationships between variables. Only the most important aspects of BNs are included in this section because a detailed description of the basics of these techniques would exceed the scope of this study. The interested reader is referred to specialized literature (Jensen & Nielsen, 2007; Pourret *et al.*, 2008; Kjaerulff & Madsen, 2013).

Using the data obtained, a BN that included the variables VCL, VSL, VAP, LIN, STR, WOB, ALH, BCF, time and cluster was constructed. Only motile spermatozoa were considered. The variables used with BNs are normally discrete variables. In this case, the continuous variables from the CASA were grouped together to facilitate the implementation of the algorithms that were used to study the network. A BN defines the structure of the dependency relationships and the joint probability distribution for a set of random variables. The two fundamental tasks required to construct a network are obtaining the network structure (i.e. qualitative part of the network) and acquiring/updating the parameters that define the probability function of the network (i.e. the quantitative part of the network). The structure is represented by an acyclic-directed graph that consists of the following: (i) nodes (representing the variables of the model); and (ii) arches (arrows that connect two nodes, which indicate that these nodes have a relationship with a defined direction). Thus, one node being the parent of another one indicates that the former is directly related to the results of the latter. BNs not only qualitatively model knowledge but also numerically express the

strengths of the relationships between variables. The quantitative part of the model is usually specified by a table of the marginal and conditional probabilities associated with each node that indicate the likelihood of that node's state for each combination of its parents' states. If a node has no parents, its prior probability (i.e. the probability of the state of the variable in the absence of evidence) is indicated.

Various optimization methods have been proposed to determine the most appropriate BN for a given data set. For parameter estimation (the joint probability distribution of random variables), the prior distribution for each parameter is used to obtain a posterior distribution (Berger & Bernardo, 1992). In this study, the greedy thick-thinning search algorithm (Dash & Cooper, 2004) was used for network structure discovery and inference. The Bayesian method considers a weakly informative prior distribution specified by a Dirichlet distribution with parameters equal to one, known as BDeu criteria (Cooper & Herskovits, 1992; Silander *et al.*, 2008). The maximum number of parent nodes was set to nine (one less than the number of variables) to allow for all possible relationships among the variables. This method has a low computational cost and has been applied with the GeNie/SMILE Software (Decision Systems Laboratory, Pittsburgh, PA, USA; Druzdzel, 1993).

RESULTS

General results

The average seminal volume of the evaluated fish was 0.50 ± 0.35 mL (mean $\pm$ SD), and the average sperm concentration was $4.04 \pm 1.12 \times 10^9$ spz/mL (mean $\pm$ SD). The high percentage of spermatozoa with intact plasma membranes was particularly notable ($99.3 \pm 0.02\%$, mean $\pm$ SD).

After dilution of the samples with Kurokura 180 immobilizing solution, an activation of sperm motility that was followed by declines in vigour and the percentage of motile sperm over time was predictably observed. The motile sperm percentages were 93.2% in the 6- to 17-s time period, 85.6% in the 23- to 34-s and 54.4% in the 40- to 51-s period of observation. After 51 s of evaluation, the vigour of movement was markedly declined. Regarding the VCL, the median value changed from 81.60 μm/s (69.0 and 94.10, first and third quartiles respectively) in the first time period to 22.10 μm/s (16.80 and 28.90, first and third quartile respectively) in the third time period. Table 1 shows the descriptive statistics for the CASA variables for each time period. Because all of the distributions were very asymmetrical, the medians and quartiles were selected for the analyses. There were statistically significant differences in the values of the eight CASA motility variables across the three sampling times ($p < 0.001$).

A significant variability in the percentages of motile spermatozoa was also observed between the males and increased with time. These values ranged from 89.7 to 99.7% in the first time period, from 77.3 to 94.4% in the second time period and from 35.3 to 73.3% in the third time period.

PC analysis and cluster analysis

A preliminary correlation analysis (Spearman rank correlation coefficient) of the eight CASA kinematic variables revealed strong correlations both globally and within each time period (data not shown); thus, a PC analysis was performed to reduce the dimensionality such that a smaller number of orthogonal variables could be used to explain a higher percentage of the total variance in the data. Table 2 shows the two PCs that were selected. PC1 was positively related to the velocity parameters VCL, VSL and VAP and related to a lesser degree to BCF. PC2 was positively related to LIN, STR and WOB and negatively related to ALH.

Using the two PCs, the cluster analysis identified three clusters (CLs) or subpopulations based on the values of the eight descriptors of motility provided by the CASA system. Table 3 presents the values of the motility variables for each cluster. CL1 included spermatozoa characterized by high speed (VCL, VSL and VAP) and linear trajectories (LIN and STR; i.e. the fast and linear spermatozoa, $n = 6339$, 45.8% of the sample). CL2 included spermatozoa characterized by high VCL and ALH values and non-linear trajectories (i.e. rapid, non-linear and high-ALH spermatozoa, $n = 1385$, 10.0% of the sample), CL3 included spermatozoa characterized by low speeds (VCL, VSL and VAP) and highly linear trajectories (LIN and STR; i.e. slow and linear spermatozoa, $n = 6119$, 44.2% of the sample). Non-parametric multiple comparisons revealed the existence of statistically significant differences ($p < 0.001$) between the clusters for all of the variables with the exceptions of VCL (no significant difference between CL1 and CL2, $p = 0.33$) and

Table 1 Descriptive statistics of CASA variables for each time period

Time	VCL (μm/s)	VSL (μm/s)	VAP (μm/s)	LIN	STR	WOB	ALH (μm)	BCF (Hz)
6–17 s								
Median	81.60	68.80	77.70	0.91	0.95	0.97	1.50	7.00
Q1	69.00	52.30	65.30	0.77	0.83	0.93	1.20	5.00
Q3	94.10	80.90	89.40	0.96	0.98	0.99	1.80	9.00
23–34 s								
Median	44.50	39.70	42.70	0.92	0.96	0.97	1.10	6.00
Q1	32.30	27.30	30.60	0.85	0.90	0.94	1.00	5.00
Q3	56.70	51.30	54.40	0.95	0.97	0.99	1.30	8.00
40–51 s								
Median	22.10	18.10	20.90	0.85	0.89	0.96	1.00	5.00
Q1	16.80	12.30	15.20	0.75	0.82	0.90	0.90	3.00
Q3	28.90	25.20	27.70	0.91	0.93	0.99	1.10	7.00

n (spermatozoa) = 13 843. CASA, computer-assisted sperm analysis; VCL, curvilinear velocity; VSL, straight line velocity; VAP, average path velocity; LIN, linearity; STR, straightness; WOB, wobble; ALH, amplitude of lateral head displacement; BCF, beat-cross frequency; Q1, first quartile; Q3, third quartile.

Table 2 Results of the two PC obtained in the study

	PC 1	PC 2
VCL (μm/s)	0.880	−0.432
VSL (μm/s)	0.974	0.002
VAP (μm/s)	0.912	−0.326
LIN	0.433	0.875
STR	0.370	0.805
WOB	0.371	0.653
ALH (μm)	0.353	−0.761
BCF (Hz)	0.535	0.047
Eigen value	3.427	2.715
Proportion of variance (%)	42.843	33.933
Cumulative proportion (%)	42.843	76.776

For each PC, the table shows the values of the covariance matrix, their eigen value, the variance proportion explained by the PC and its cumulative proportion. PC, principal components; VCL, curvilinear velocity; VSL, straight line velocity; VAP, average path velocity; LIN, linearity; STR, straightness; WOB, wobble; ALH, amplitude of lateral head displacement; BCF, beat-cross frequency.

BCF (no significant difference between CL2 and CL3, $p = 0.295$). Moreover, there was a significant association ($p < 0.001$) between cluster and time. Table 4 shows the results for each cluster by time. The proportions of spermatozoa contained in each cluster varied widely with time; the proportions of CL1 and CL2 spermatozoa decreased with time, and the opposite pattern was observed for CL3. In addition, the proportions of spermatozoa in the different clusters changed according to the time period. In the first time period, the majority of the spermatozoa (77.2%) were included in the first cluster. In the second time period, 55.1% of the spermatozoa were included in the third cluster, and the proportion in the first cluster substantially reduced. Finally, in the third time period, the proportion in the first cluster was residual (1.5%), and the majority of the spermatozoa were included in the third cluster (92.2%).

For each time period, there were significant differences in the proportions of spermatozoa from the different males in each cluster (Table 5).

Table 4 Number of spermatozoa for clusters and time intervals

Time	Cluster			Total
	1	2	3	
6–17 s				
No. spz	4291	917	350	5558
% of time	77.20	16.50	6.30	100.00
% of cluster	67.70	66.20	5.70	40.20
23–34 s				
No. spz	1999	264	2781	5044
% of time	39.60	5.20	55.10	100.00
% of cluster	31.50	19.10	45.40	36.40
40–51 s				
No. spz	49	204	2988	3241
% of time	1.50	6.30	92.20	100.00
% of cluster	0.80	14.70	48.80	23.40
Total				
No. spz	6339	1385	6119	13 843
% of time	45.80	10.00	44.20	100.00
% of cluster	100.00	100.00	100.00	100.00

No. spz: number of spermatozoa.

Bayesian network

Figure 1 illustrates the constructed BN. The joint conditional probability distribution, which associates all of the variables and specifies the probability of each possible combination of states for each variable, can be defined formally as follows:

$$P(\text{ALH, VSL, VCL, VAP, LIN, STR, WOB, BCF, Time, ClusterPC}) = P(\text{ALH}) * P(\text{VSL}) * P(\text{VCL}|\text{VSL, ALH}) * P(\text{LIN}|\text{VSL, ALH, VCL}) * P(\text{STR}|\text{LIN, VCL, VAP}) * P(\text{BCF}|\text{STR, VAP}) * P(\text{WOB}|\text{BCF}) * P(\text{VAP}|\text{LIN, VCL}) * P(\text{Time}|\text{VCL}) * P(\text{ClusterPC}|\text{VCL, LIN})$$

Figure 1 shows that the parameters VSL, VCL and VAP were conditionally dependent on each other. VAP depended directly on VCL, which in turn depended on VSL. In addition, the variables LIN, STR and WOB, which were derived from the velocity variables, were related to the variables that were used for their calculation; in some cases, these relations were via direct arches (e.g. LIN), whereas in others cases, the relations were indirect (e.g. WOB, which was indirectly influenced by VAP and VCL through BCF). LIN and STR were directly related and indirectly related to WOB through BCF.

Table 3 Descriptive statistics of CASA variables for each cluster

Cluster	VCL (μm/s)	VSL (μm/s)	VAP (μm/s)	LIN	STR	WOB	ALH (μm)	BCF (Hz)
1								
Median	73.70	66.60	71.00	0.94	0.97	0.97	1.40	7.00
Q1	60.50	55.20	58.30	0.89	0.93	0.95	1.20	6.00
Q3	87.60	79.20	85.00	0.96	0.98	0.99	1.60	9.00
2								
Median	77.80	26.70	65.40	0.44	0.55	0.87	1.90	5.00
Q1	37.60	10.80	28.25	0.30	0.36	0.76	1.40	3.00
Q3	95.60	44.40	84.05	0.55	0.68	0.91	2.40	7.00
3								
Median	27.60	24.00	26.50	0.88	0.92	0.97	1.00	5.00
Q1	19.60	15.90	18.60	0.80	0.86	0.92	0.90	4.00
Q3	37.00	33.40	35.60	0.93	0.95	0.99	1.10	7.00

n (spermatozoa) = 13 843. CASA, computer-assisted sperm analysis; VCL, curvilinear velocity; VSL, straight line velocity; VAP, average path velocity; LIN, linearity; STR, straightness; WOB, wobble; ALH, amplitude of lateral head displacement; BCF, beat-cross frequency; Q1, first quartile; Q3, third quartile.

Regarding the time and cluster variables, VCL was directly related to time, and this variable and LIN were directly related to the cluster variable. However, the cluster variable was influenced indirectly by all of the variables tested in the study. In general, it can be concluded that all of the variables were directly or indirectly related to each other with the exceptions of ALH and VSL, which did not share a conditional relationship. The most important relationships observed were those between the velocity, LIN and time variables.

To delve deeper into the relationships between the variables, we used one of the most important applications of BNs, that is the potential for inference or the spread of evidence. With this application, once the value (s) of some variable (s) is (are) known, the probabilities of the remaining variables can be updated (recalculated). Table 6 illustrates the evidence related to some possible (hypothetical) scenarios. Thus, if Cluster 1 is chosen as evidence (state 1 within the cluster variable), after recalculating the probability distributions of the remaining nodes under the condition that Cluster 1 has an initial probability of 100%, the spermatozoa belonging to the first cluster are more likely to exhibit medium-high speeds. The LIN and STR values are very high, and the majority of the spermatozoa belong to the first time period (6–17 s; Table 6, Scenario 1). Choosing Cluster 3 as evidence, the velocities are primarily distributed among the low values, LIN and STR tend to be more moderate and the ALH and BCF values also decrease. The majority of spermatozoa from this cluster belong to the second (23–34 s) and third (40–51 s) sampling times (Table 6, Scenario 2).

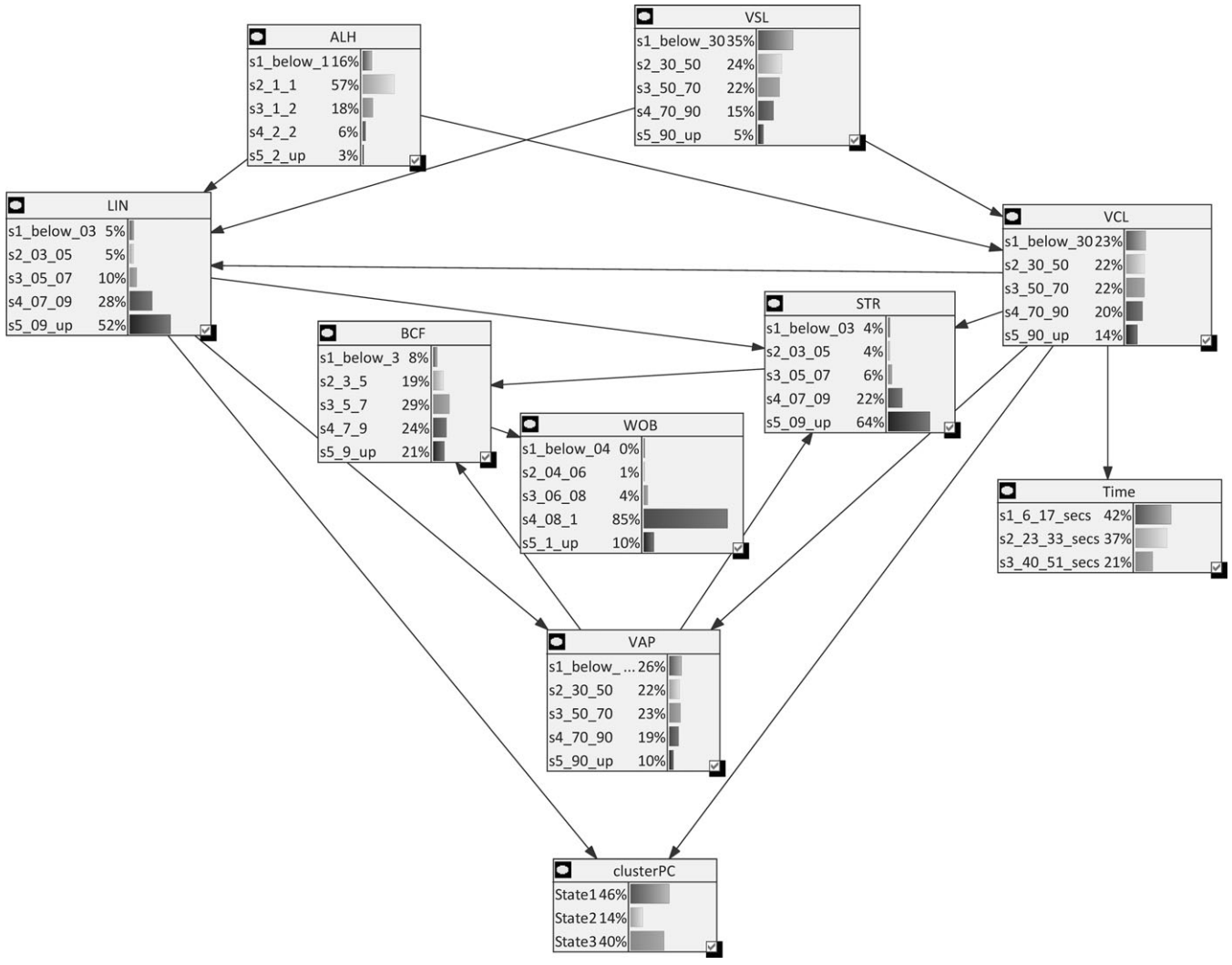
For the choice of time variable and the 6- to 17-s sampling time as evidence, the probabilities of VSL, VCL and VAP assuming values >50 μm/s are 71, 92 and 90% respectively. The probability that the LIN of the trajectory will assume a value >70% is also very high (LIN = 70%, STR = 88%; Table 6, Scenario 3). In contrast, when choosing the third sampling time period as evidence, the LIN and speed decrease (the probabilities of the VCL and VAP assuming a value below 30 μm/s are 75 and 77% respectively). The values from the first time period primarily belong to the first cluster (a probability of 73%; Table 6, Scenario 3), the values from the third time period primarily belong to the third cluster (87% probability; Table 6, Scenario 5), and the

Table 5 Evolution of each sperm subpopulation in each male through time

Time	Cluster	Male							
		1	2	3	4	5	6	7	8
6–17 s	1	80.5^{a}	78.4^{a}	80.5^{a}	77.6^{a,b}	74.3^{a,b}	70.5^{b}	79.0^{a}	77.7^{a,b}
	2	7.9^{a}	11.4^{a,c,e}	17.8^{b,d}	15.4^{b,c}	21.5^{d}	21.2^{b,d}	17.3^{b,d,e}	18.8^{b,d}
	3	11.6^{a}	10.2^{a}	1.7^{b}	7.0^{a,c,d}	4.2^{b,c}	8.3^{a}	3.8^{b,d}	3.5^{b}
23–33 s	1	22.5^{a}	28.1^{a,c}	55.6^{b,f}	30.7^{c,e}	42.6^{d,g}	36.7^{d,e}	56.2^{b}	46.6^{f,g}
	2	3.8^{a,c,d}	6.8^{a,b}	2.6^{c}	5.1^{a,c,d}	4.7^{a,c,d}	5.0^{a,c,d}	4.1^{a,c,d}	7.8^{b,d}
	3	73.7^{a}	65.0^{b}	41.9^{c}	64.2^{b}	52.6^{d,e}	58.3^{b,d}	39.7^{c}	45.6^{c,e}
40–51 s	1	0.9^{a}	0.6^{a}	1.6^{a}	0.5^{a}	3.1^{a}	1.3^{a}	1.5^{a}	1.6^{a}
	2	8.3^{a,f,g}	11.6^{a,b}	2.4^{c,e}	10.3^{a,d}	1.9^{c}	6.4^{a,e,f,g}	3.8^{c,f}	8.4^{b,d,g}
	3	90.8^{a,c,d,g}	87.8^{a,b}	96.0^{c}	89.2^{b,d,f}	95.0^{c,e,g}	92.3^{a,c,d,g}	94.7^{c,f,g}	89.9^{b,g}

Values are percentages. For each time and cluster (row), significant differences between males for the different subpopulations (clusters) are indicated by different superscripts ($p < 0.05$).

Figure 1 Bayesian network structure and conditional probabilities.



values from the second time period (23–34 s) are distributed between the three clusters with Cluster 2 as the most under-represented subpopulation (Table 6, Scenario 4).

The velocity parameters VCL, VSL and VAP are interrelated and increase or decrease in unison. Evidence of these parameters simultaneously indicating high velocities (70–90 μm/s) implies that the data primarily belong to the first time period and exhibit the highest LIN (Table 6, Scenario 6). In addition, the majority of the spermatozoa (nearly 100%) are in cluster 1. In contrast, if low speeds (below 30 μm/s) are selected as evidence, the LIN decreases, the sampling time period in which the spermatozoa are evaluated is primarily the third, and the cluster of membership is three (90%; Table 6, Scenario 7). However, although the individual evidence of high

Table 6 Evidence propagation for some possible scenarios

Variables	Scenarios										
	1	2	3	4	5	6	7	8	9	10	11
VCL (μm/s)											
<30	0.00	52.49	2.39	17.61	74.64	0.00	**100.00**	0.00	0.00	0.00	92.84
30–50	5.03	45.12	5.15	40.06	24.11	0.00	0.00	0.00	0.00	0.00	6.88
50–70	40.01	2.37	19.30	35.79	1.10	0.00	0.00	0.00	0.00	0.00	0.24
70–90	35.35	0.02	41.40	6.19	0.15	**100.00**	0.00	**100.00**	81.40	73.03	0.04
>90	19.62	0.00	31.76	0.36	0.00	0.00	0.00	0.00	18.80	26.97	0.00
VSL (μm/s)											
<30	0.66	62.78	14.01	32.02	81.49	0.00	**100.00**	9.50	10.27	0.00	95.55
30–50	14.14	35.83	14.67	37.80	17.62	0.00	0.00	8.04	7.61	0.00	4.31
50–70	42.71	1.39	27.35	26.45	0.80	0.00	0.00	25.94	23.59	0.00	0.13
70–90	32.53	0.00	33.10	3.61	0.08	**100.00**	0.00	56.52	55.25	**100.00**	0.01
>90	9.96	0.00	10.86	0.12	0.00	0.00	0.00	0.00	3.27	0.00	0.00
VAP (μm/s)											
<30	0.24	56.44	3.28	22.30	77.35	0.00	**100.00**	0.53	0.00	0.00	94.18
30–50	8.58	41.50	7.12	39.74	21.53	0.00	0.00	0.41	0.00	0.00	5.57
50–70	41.41	2.05	25.23	32.66	1.00	0.00	0.00	18.96	0.00	8.53	0.21
70–90	33.69	0.01	41.68	5.05	0.12	**100.00**	0.00	80.11	**100.00**	70.25	0.04
>90	16.08	0.00	22.70	0.25	0.00	0.00	0.00	0.00	0.00	21.22	0.00
LIN											
<0.3	0.00	0.00	8.80	3.10	1.97	0.00	1.66	7.28	9.15	0.00	0.00
0.3–0.5	0.00	0.10	5.90	4.27	5.13	0.00	5.58	4.98	5.44	0.00	0.00
0.5–0.7	2.96	11.35	8.72	7.92	14.17	0.00	16.71	8.00	7.96	1.84	0.00
0.7–0.9	21.52	46.53	19.98	27.28	47.87	7.67	55.65	17.89	17.46	20.45	**100.00**
>0.9	75.52	42.02	56.61	57.43	30.86	92.33	20.40	61.86	59.99	77.70	0.00
STR											
<0.3	0.00	0.00	6.89	2.21	1.10	0.00	0.76	5.35	7.22	0.00	0.00
0.3–0.5	0.00	0.04	5.52	2.80	2.16	0.00	1.99	5.08	5.87	0.00	0.00
0.5–0.7	1.69	3.32	6.79	4.85	6.30	0.00	6.98	5.89	6.19	1.14	0.00
0.7–0.9	12.22	36.51	15.02	18.95	43.94	4.21	54.46	12.93	12.66	13.47	**100.00**
>0.9	86.08	60.14	65.77	71.20	46.50	95.79	35.80	70.75	68.06	85.39	0.00
ALH (μm)											
<1	15.63	21.45	12.47	17.26	21.08	18.11	22.14	12.30	11.97	16.03	19.74
1–1	57.91	69.96	46.40	62.06	69.66	62.79	71.34	46.61	45.38	57.03	75.62
1–2	18.80	7.86	23.83	16.53	8.11	15.17	6.04	24.79	24.10	17.86	4.42
2–2	5.53	0.65	10.87	3.22	0.91	3.37	0.43	11.33	11.86	5.97	0.22
>2	2.13	0.09	6.43	0.93	0.24	0.55	0.05	4.97	6.68	3.11	0.00
BCF (Hz)											
<3	2.28	10.77	6.33	6.08	13.77	1.14	16.70	3.23	4.64	2.39	15.30
3–5	11.13	25.17	13.06	19.68	27.66	7.19	29.94	10.67	10.22	8.37	31.16
5–7	27.85	30.77	25.95	32.26	28.96	21.86	27.22	25.16	23.44	22.69	26.97
7–9	29.29	19.35	27.32	24.51	17.48	30.80	15.55	29.97	29.50	29.71	15.86
>9	29.45	13.94	27.34	17.47	12.12	39.01	10.59	30.97	32.20	36.85	10.71
TIME (s)											
6–17	66.13	7.64	**100.00**	0.00	0.00	88.22	4.38	88.22	90.23	91.13	0.00
23–34	32.26	46.85	0.00	**100.00**	0.00	11.62	28.44	11.62	9.64	8.75	0.00
40–51	1.61	45.51	0.00	0.00	**100.00**	0.15	67.18	0.15	0.13	0.11	**100.00**
Cluster											
1	**100.00**	0.00	72.63	40.18	3.61	99.94	0.00	82.73	80.12	98.36	0.64
2	0.00	0.00	20.19	9.93	9.42	0.06	9.72	17.23	19.86	1.64	0.00
3	0.00	**100.00**	7.18	49.89	86.96	0.00	90.28	0.04	0.02	0.00	99.36

Values are percentages and indicate the probability of each state for each variable, once some evidences have been set (bold values). VCL, curvilinear velocity; VSL, straight line velocity; VAP, average path velocity; LIN, linearity; STR, straightness; ALH, amplitude of lateral head displacement; BCF, beat-cross frequency.

VCL or VAP (70–90 μm/s; Table 6, Scenarios 8 and 9 respectively) characterize the spermatozoa as belonging to both the first and second clusters, evidence of VSLs between these values characterize the spermatozoa as members of Cluster 1 with a 98.36% probability and additionally indicate that they belong to the first time period with 90% probability (Table 6, Scenario 10). Evidence of a VSL between 50 and 70 μm/s implies a 91.87% probability of belonging to Cluster 1 and the inclusion of spermatozoa from both the first and second time periods (scenarios not shown).

Regarding the LIN of the trajectory, it is particularly noteworthy that when the LIN and STR values are high (0.70–0.90) and the spermatozoa are evaluated in the third time period, there is a 99.36% probability that the spermatozoa belong to Cluster 3, and the majority will have low velocities (below 30 μm/s; Table 6, Scenario 11).

The ALH and BCF parameters were not decisive in characterizing the spermatozoa.

DISCUSSION

Motility analyses with CASA systems have provided one of the parameters that are used to classify spermatozoa quality (Kime *et al.*, 2001). These systems are characterized by providing information about a large number of kinematic variables and because

they assess the motility of individual spermatozoa, they generate huge data sets. To take full advantage of the obtained data, statistical techniques such as cluster analysis (or conglomerate analysis) that allow for the classification of spermatozoa into subsets (called clusters) with certain motility characteristics have been applied to provide further information about the sample. Although few such studies have been conducted in fish (Martínez-Pastor *et al.*, 2008; Beirão *et al.*, 2011; Kanuga *et al.*, 2012), many studies in other species have used cluster analysis to identify subpopulation patterns in sperm samples and have confirmed that sperm samples are heterogeneous; that is spermatozoa with different motility patterns coexist in the same ejaculates (Abaigar *et al.*, 1999; Quintero-Moreno *et al.*, 2003; Chantler *et al.*, 2004; Miró *et al.*, 2005; Nunez-Martinez *et al.*, 2006; Muino *et al.*, 2008; Dorado *et al.*, 2010; Contri *et al.*, 2012). Contri *et al.* (2012) used artificial neural networks to cluster sperm subpopulations in feline semen. In this study we have identified three different sperm subpopulations in *Tinca tinca* semen samples after <1 min of activation, varying the proportions of the different subpopulations with the post-activation time. These findings (i.e. the subpopulation patterns) are consistent with the finding of the presence of subpopulations of motile spermatozoa in other fish species (Martínez-Pastor *et al.*, 2008; Beirão *et al.*, 2009, 2011; Kanuga *et al.*, 2012) and other animal species (Martínez-Pastor *et al.*, 2011; Contri *et al.*, 2012; Ferraz *et al.*, 2014). Once sperm subpopulations are identified, the proportions of spermatozoa in each cluster are the values more often used to evaluate differences between ejaculates, males and/or treatments rather than the kinematic variables themselves (Martínez-Pastor *et al.*, 2008; Beirão *et al.*, 2011; Kanuga *et al.*, 2012). Several researchers have suggested that these subpopulations are biologically important (Quintero-Moreno *et al.*, 2003; Martínez-Pastor *et al.*, 2005, 2008), but further data are needed to confirm this claim. The presence of a 'fast and linear' subpopulation has been proposed as an indicator of high sample quality (Martínez-Pastor *et al.*, 2005, 2011) because velocity and STR are considered important parameters correlated with fertilization success (Gage *et al.*, 2004; Casselman *et al.*, 2006; Tuset *et al.*, 2008; Kanuga *et al.*, 2012); thus, reductions in these parameters may decrease fecundity (Kime *et al.*, 2001; Rurangwa *et al.*, 2004), and high proportions of spermatozoa in the fast and linear subpopulation (SPB 1 in our study) may be particularly advantageous for the selection of high-quality breeders (Beirão *et al.*, 2009; Ferraz *et al.*, 2014). In this study, the sperm subpopulation structure varied not only with time but also with the individual male, which is in accordance with the results of other studies (Martínez-Pastor *et al.*, 2008; Beirão *et al.*, 2009) and demonstrate a connection between sperm quality and male genetic quality (Fitzpatrick *et al.*, 2007). The analysis of motility based on sperm subpopulations may improve the assessment of male differences and enable the identification of males with high sperm quality. In this regard, further studies are needed.

In this study, we utilized a new approach to the analysis of sperm motility via the first application of BNs. The aim was to study the relationships between variables (including the cluster variable) and how each variable influenced the final sperm classification into one of three predefined clusters. The advantage of the use of BNs is that relationships between variables can be discovered based on the joint conditional probability distribution. These relationships would not otherwise be identified. For example in a regression analysis, the model explains the variation in the dependent variable according the variation of the independent variables, but this does not indicate that the former is directly related to one of the independent variables or whether an indirect relationship exists. The joint probability distribution for some variables is a probability distribution that gives the probability that each of the variables falls in any particular range or discrete set of values specified for that variable. That is, it does not only consider the individual probabilities of one variable, but it also considers the probability for all the variables jointly. Here the concept of dependence is involved. Two variables are dependent if the knowledge of one of them provides predictive value for the knowledge of the other. For example knowledge of the state of VSL provides predictive value for the probability of VCL (and also in the opposite direction).

The results from this study indicate that velocity parameters are related to each other and confirm that the variables that are direct transformations of the velocities (LIN, STR and WOB) are related to the variables used in their calculations. Moreover, the results of a previous study revealed strong correlations between the movement-related variables of LIN, STR and WOB. In the BN analysis, these correlations were reflected in the direct relationship between LIN and STR and the indirect relationship of WOB through BCF. These results are consistent with classical correlations that have been studied in a previous descriptive analysis in which the variables exhibited high Spearman correlation coefficients. However, we must not forget that the BN analysis does not examine correlations; rather, the relationships or influences between variables are examined through the joint conditional probability distribution, which is based on the information provided by the data.

The finding that all of the variables were related to each other with the exceptions of ALH and VSL, which were conditionally independent, confirms the redundancy of the CASA variables noted by Martínez-Pastor *et al.* (2011); indeed, some of the variables, such as VCL, VAP and VSL, convey similar information (sperm velocity), and some variables arise from other variables (e.g. LIN is the VSL/VCL ratio). A cluster analysis of CASA data precisely reduces both the dimensionality and the redundancy so that a reduced number of orthogonal variables can be used to explain the greatest percentage of the data variance (Martínez-Pastor *et al.*, 2011).

An important advantage of the use of BNs is the ability to make inferences (probability propagation). This is a process whereby new observations (evidence) are introduced, and the new probabilities of the remaining variables are subsequently updated. Using this property and entering a constant value for the cluster variable as evidence, we observed that the cluster distribution corresponded to the definition provided by the cluster analysis.

As expected, the velocity parameters were related and directly influenced the other variables. Notably, these variables determined the cluster to which each spermatozoon belonged with a high degree of accuracy. Ultimately, these variables may characterize the elements of the sample by themselves. Among these variables, VSL stands out because setting the values of this variable over a wide range (intermediate-high, 50–90 μm/s) allowed the spermatozoa to be classified into clusters.

Unlike the velocity variables VCL, VSL and VAP, high or low values of the LIN of the trajectory variables LIN, STR and WOB

were not completely decisive in terms of a spermatozoon possessing a particular velocity or being observed during a specific time period. This set of variables needed to be combined with other variables, such as velocity or time variables, to determine to which group a spermatozoon belonged. In this sense, it is notable that the spermatozoa with high STR and LIN values (70–90%) that were assessed in the third time period had a 99.3% probability of belonging to the third cluster.

In addition to a very short duration of motility, fish spermatozoa are characterized by rapid variations in the motility parameters (Rurangwa *et al.*, 2004; Cosson *et al.*, 2008a). This latter characterization can be confirmed by setting the time variable as evidence; in the 6- to 17-s time period, the highest LIN and speed values were obtained, whereas in the 40- to 51-s time period, LIN decreased, and the velocities were kept to a minimum. By itself, this variable could also categorize the spermatozoa into clusters.

On the basis of these results, we consider the velocity and time variables to be decisive measures for the characterization of *Tinca tinca* semen, and the variables related to the LIN of the trajectory complement the velocity and time variables and provide additional information about sample motility.

Using a BN, we identified the relationships between the motility, cluster and post-activation time variables. However, the applications of BNs can be extended further. Several studies in fish have shown that sperm motility and fertilizing ability are positively correlated (Lahnsteiner *et al.*, 1998; Linhart *et al.*, 2000; Rurangwa *et al.*, 2001; Butts *et al.*, 2011). The very short duration of sperm motility following activation that is observed in most farmed fish species exerts a critical influence on successful fertilization because the spermatozoa must find and enter the single point of entry, the micropyle, during this period (Kime *et al.*, 2001; Rurangwa *et al.*, 2004). In this sense, Rurangwa *et al.* (2001) found that the fertilization or hatching rate is positively correlated with the VSL or VCL parameters. A regression analysis of frozen/thawed semen samples from Atlantic cod (*Gadus morhua* L.) demonstrated significant positive relationships of VCL, VSL and VAP with fertilization success (Butts *et al.*, 2011). The applicability of CASAs would be greatly extended if the sperm motility (including analysis of the motility subpopulations) can be related to the fertilization ability, a subject of great debate. The quality of different milt samples could be compared, and those with the best fertilization rates could be predicted. As previously mentioned, several studies in mammals (Quintero-Moreno *et al.*, 2003, 2004; Martínez-Pastor *et al.*, 2005) and very few in fish (Beirão *et al.*, 2011) have found correlations between subpopulation sizes and fertilization ability, that indicate that greater numbers of spermatozoa in the fast and linear subpopulations are related to higher quality semen sample. However, definitive conclusions have not yet been drawn. In appropriately planned research studies, the construction of a BN that includes the fertility variable with the variables considered in this study would allow the assessment of the possible relationships between the variables, as well as which variables, values of these variables and subpopulations are associated with high probabilities of fertilization.

Furthermore, it is likely that a number of different tests must be performed to more accurately predict fertilizing ability. For example in fish, in addition to assessments of sperm motility, other parameters of the quality of fish spermatozoa that have been examined include spermatocrit, morphology, ultrastructure, and sperm density, viability, energy content and DNA state (Rurangwa *et al.*, 2004; Fauvel *et al.*, 2010). BNs could be used to reveal the relationships between these parameters and to evaluate which of them, either independently or in combination, can probabilistically predict fertility.

Because the velocity of fish spermatozoa decreases rapidly with time (Chauvaud *et al.*, 1995), the time period over which the motility of the spermatozoa can be maintained as rapid and straight could also be investigated; the longer that spermatozoa can maintain this pattern the higher the probability of a spermatozoa entering the egg (Gage *et al.*, 2004). The pattern of straight motility is important because the trajectory of fish spermatozoa is generally more curved than that in mammals, and fish spermatoza can move three-dimensionally in aqueous medium (Kime *et al.*, 2001).

One advantage of the use of BNs is the ability to incorporate data from many different sources including non-empirical data, such as bibliographical data and expert opinions. Such data can be introduced into the network at any step of the model construction (Jensen & Nielsen, 2007), which implies the possibility of the participation of different research groups in the construction and use of the network. Moreover, BNs can be applied to any animal species.

In conclusion, in this study, after identification of three sperm subpopulations with specific motility characteristics in *Tinca tinca* semen samples, a BN was constructed to identify the direct and indirect relationships between motility, cluster and time variables, which confirmed some expectations regarding some of the variables. In addition, the underlying relationships were also discovered, which provided added value to the BN analysis. Velocity and time proved to be determinant variables in the characterization of tench semen. Future applications include the construction of BNs that include fertility data, which may help elucidate the roles of motility and other biomarkers of sperm quality in fertilization and improve the efficiency of artificial insemination and the choice of broodstock.

ACKNOWLEDGEMENTS

This work was supported by Plan de Iniciación a la Investigación, Desarrollo Tecnológico e Innovación ACCVII-03, UEX; GRU10156, GRU10110 and GR15106, Gobierno de Extremadura, Spain. D. Martín-Hidalgo was supported by a PhD fellowship from Gobierno de Extremadura, Spain. The authors are particularly grateful to some colleagues of the 'Vegas del Guadiana' Aquaculture Center (H. J. Pula, M. Pascual and P. Moreno) for their technical assistance.

AUTHOR CONTRIBUTIONS

M.C.G. did laboratorial experiments, interpreted the data and wrote the manuscript. F.C. and C.J.P. designed and did the statistical study, and analysed and interpreted the data. D.M.-H. and M.J.B. contributed to laboratorial work. C.F. was responsible for selection, preparation and maintenance of animals as well as sample collection. L.G.M. contributed to interpretation of data. A.L.O. did the research design and contributed to sample collection and laboratorial work. All authors revised critically and made final approval of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could perceived as prejudicing the impartiality of the research reported.

REFERENCES

Abaigar T, Holt WV, Harrison RA & del Barrio G. (1999) Sperm subpopulations in boar (*Sus scrofa*) and gazelle (*Gazella dama mhorr*) semen as revealed by pattern analysis of computer-assisted motility assessments. *Biol Reprod* 60, 32–41.

Aparicio IM, Bragado MJ, Gil MC, Garcia-Herreros M, Gonzalez-Fernandez L, Tapia JA & García-Marín LJ. (2007) Phosphatidylinositol 3-kinase pathway regulates sperm viability but not capacitation on boar spermatozoa. *Mol Reprod Dev* 74, 1035–1042.

Beirão J (2011) Sperm quality in marine teleosts: applications to broodstock management and sperm storage in two farmed species *Solea senegalensis* and *Sparus aurata*. PhD thesis, University of León, León, Spain.

Beirão J, Soares F, Herráez MP, Dinis MT & Cabrita E. (2009) Sperm quality evaluation in *Solea senegalensis* during the reproductive season at cellular level. *Theriogenology* 72, 1251–1261.

Beirão J, Cabrita E, Pérez-Cerezales S, Martínez-Páramo S & Herráez MP. (2011) Effect of cryopreservation on fish sperm subpopulations. *Cryobiology* 62, 22–31.

Berger J & Bernardo J. (1992) On the development of reference priors. In: *Bayesian Statistics 4*(eds J Bernardo, J Berger, A Dawid & A Smith), pp. 35–60. Oxford University Press, Oxford, UK.

Bobe J & Labbé C. (2010) Egg and sperm quality in fish. *Gen Comp Endocrinol* 165, 535–548.

Butts IA, Babiak I, Ciereszko A, Litvak MK, Slowinska M, Soler C & Trippel EA. (2011) Semen characteristics and their ability to predict sperm cryopreservation potential of Atlantic cod, *Gadus morhua* L. *Theriogenology* 75, 1290–1300.

Casselman SJ, Schulte-Hostedde A & Montgomerie R. (2006) Sperm quality influences male fertilization success in walleye (*Sander vitreus*). *Can J Fish Aquat Sci* 63, 2119–2125.

Chantler E, Abraham-Peskir J & Roberts C. (2004) Consistent presence of two normally distributed sperm subpopulations within normozoospermic human semen: a kinematic study. *Int J Androl* 27, 350–359.

Chauvaud L, Cosson J, Suquet M & Billard R. (1995) Sperm motility in turbot, *Scophthalmus marimus*: initiation of movement and changes with time of swimming characteristics. *Environ Biol Fish* 43, 341–349.

Contri A, Zambelli D, Faustini M, Cunto M, Gloria A & Carluccio A. (2012) Artificial neural networks for the definition of kinetic subpopulations in electroejaculated and epididymal spermatozoa in the domestic cat. *Reproduction* 144, 339–347.

Cooper G & Herskovits E. (1992) A bayesian method for the induction of probabilistic networks from data. *Mach Learning* 9, 309–347.

Cosson J, Groison AL, Suquet M, Fauvel C, Dreanno C & Billard R. (2008a) Marine fish spermatozoa: racing ephemeral swimmers. *Reproduction* 136, 277–294.

Cosson J, Groison A-, Suquet M, Fauvel C, Dreanno C & Billard R. (2008b) Studying sperm motility in marine fish: an overview on the state of the art. *J Appl Ichthyol* 24, 460–486.

Daly R, Shen Q & Aitken S. (2011) Learning Bayesian networks: approaches and issues. *Knowl Eng Rev* 26, 99–157.

Dash D & Cooper G. (2004) Model averaging for prediction with discrete Bayesian networks. *Mach Learn* 5, 1177–1203.

Dietrich GJ, Dietrich M, Kowalski RK, Dobosz S, Karol H, Demianowicz W & Glogowski J. (2010) Exposure of rainbow trout milt to mercury and cadmium alters sperm motility parameters and reproductive success. *Aquat Toxicol* 97, 277–284.

Dorado J, Molina I, Munoz-Serrano A & Hidalgo M. (2010) Identification of sperm subpopulations with defined motility characteristics in ejaculates from Florida goats. *Theriogenology* 74, 795–804.

Druzdzel M. (1993) Probabilistic reasoning in decision support systems: from computation to common sense. PhD thesis, Cornegie Mellon University, Pittsburgh, PA, USA.

Fauvel C, Suquet M & Cosson J. (2010) Evaluation of fish sperm quality. *J Appl Ichthyol* 26, 636–643.

Ferraz MA, Morató R, Yeste M, Arcarons N, Pena AI, Tamargo C, Hidalgo CO, Muiño R & Mogas T. (2014) Evaluation of sperm subpopulation structure in relation to in vitro sperm-oocyte interaction of frozen-thawed semen from Holstein bulls. *Theriogenology* 81, 1067–1072.

Fitzpatrick JL, Desjardins JK, Milligan N, Montgomerie R & Balshine S. (2007) Reproductive-tactic-specific variation in sperm swimming speeds in a shell-brooding cichlid. *Biol Reprod* 77, 280–284.

Flores E, Fernández-Novell JM, Peña A & Rodríguez-Gil JE. (2009) The degree of resistance to freezing-thawing is related to specific changes in the structures of motile sperm subpopulations and mitochondrial activity in boar spermatozoa. *Theriogenology* 72, 784–797.

Gage MJ, Macfarlane CP, Yeates S, Ward RG, Searle JB & Parker GA. (2004) Spermatozoal traits and sperm competition in Atlantic salmon: relative sperm velocity is the primary determinant of fertilization success. *Curr Biol* 14, 44–47.

Jensen F & Nielsen T. (2007) *Bayesian Networks and Decision Graphs*, 2nd edn. Springer-Verlag, New York, NY.

Kanuga MK, Drew RE, Wilson-Leedy JG & Ingermann RL. (2012) Subpopulation distribution of motile sperm relative to activation medium in steelhead (Oncorhynchus mykiss). *Theriogenology* 77, 916–925.

Kime DE & Tveiten H. (2002) Unusual motility characteristics of sperm of the spotted wolffish. *J Fish Biol* 61, 1549–1559.

Kime DE, Ebrahimi M, Nysten K, Roelants I, Rurangwa E, Moore HDM & Ollevier F. (1996) Use of computer assisted sperm analysis (CASA) for monitoring the effects of pollution on sperm quality of fish; application to the effects of heavy metals. *Aquat Toxicol* 36, 223–237.

Kime DE, Van Look KJ, McAllister BG, Huyskens G, Rurangwa E & Ollevier F. (2001) Computer-assisted sperm analysis (CASA) as a tool for monitoring sperm quality in fish. *Comp Biochem Physiol C Toxicol Pharmacol* 13, 425–433.

Kjaerulff U & Madsen A. (2013) *Bayesian Network and Influence Diagrams: A Guide to Construction and Analysis*. Springer, New York, NY.

Korb K & Nicholson A. (2010) *Bayesian Artificial Intelligence*, 2nd edn. CRC Press, New York, NY.

Lahnsteiner F, Berger B, Weismann T & Patzner RA. (1998) Determination of semen quality of the rainbow trout, *Oncorhynchus mykiss*, by sperm motility, seminal plasma parameters, and spermatozoal metabolism. *Aquaculture* 163, 163–181.

Lewis FI, Brulisauer F & Gunn GJ. (2011) Structure discovery in Bayesian networks: an analytical tool for analysing complex animal health data. *Prev Vet Med* 100, 109–115.

Linhart O, Rodina M & Cosson J. (2000) Cryopreservation of sperm in common carp *Cyprinus carpio*: sperm motility and hatching success of embryos. *Cryobiology* 41, 241–250.

Linhart O, Rodina M, Bastl J & Cosson J. (2003) Urinary bladder, ionic composition of seminal fluid and urine with characterization of sperm motility in tench (*Tinca tinca* L.). *J Appl Ichthyol* 19, 177–181.

Lucas PJ, van der Gaag LC & Abu-Hanna A. (2004) Bayesian networks in biomedicine and health-care. *Artif Intell Med* 30, 201–214.

Martínez-Páramo S, Diogo P, Dinis MT, Herráez MP, Sarasquete C & Cabrita E. (2012) Incorporation of ascorbic acid and α-tocopherol to the extender media to enhance antioxidant system of cryopreserved sea bass sperm. *Theriogenology* 77, 1129–1136.

Martínez-Pastor F, Garcia-Macias V, Alvarez M, Herraez P, Anel L & de Paz P. (2005) Sperm subpopulations in Iberian red deer epididymal

sperm and their changes through the cryopreservation process. *Biol Reprod* 72, 316–327.
Martínez-Pastor F, Cabrita E, Soares F, Anel L & Dinis MT. (2008) Multivariate cluster analysis to study motility activation of *Solea senegalensis* spermatozoa: a model for marine teleosts. *Reproduction* 135, 449–459.
Martínez-Pastor F, Tizado EJ, Garde JJ, Anel L & de Paz P. (2011) Statistical series: opportunities and challenges of sperm motility subpopulation analysis. *Theriogenology* 75, 783–795.
Miró J, Lobo V, Quintero-Moreno A, Medrano A, Pena A & Rigau T. (2005) Sperm motility patterns and metabolism in Catalonian donkey semen. *Theriogenology* 63, 1706–1716.
Muino R, Tamargo C, Hidalgo CO & Pena AI. (2008) Identification of sperm subpopulations with defined motility characteristics in ejaculates from Holstein bulls: effects of cryopreservation and between-bull variation. *Anim Reprod Sci* 109, 27–39.
Muino R, Peña AI, Rodríguez A, Tamargo C & Hidalgo CO. (2009) Effects of cryopreservation on the motile sperm subpopulations in semen from Asturiana de los Valles bulls. *Theriogenology* 72, 860–868.
Ngai EWT, Hu Y, Wong YH, Chen Y & Sun X. (2011) The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decis Support Syst* 50, 559–569.
Nunez-Martinez I, Moran JM & Pena FJ. (2006) A three-step statistical procedure to identify sperm kinematic subpopulations in canine ejaculates: changes after cryopreservation. *Reprod Domest Anim* 41, 408–415.
Pourret O, Naim P & Marcot B. (2008) *Bayesian Networks. A Practical Guide to Applications*. Willey, Chichester, west Sussex, England.
Quintero-Moreno A, Miro J, Teresa Rigau A & Rodriguez-Gil JE. (2003) Identification of sperm subpopulations with specific motility characteristics in stallion ejaculates. *Theriogenology* 59, 1973–1990.
Quintero-Moreno A, Rigau T & Rodriguez-Gil JE. (2004) Regression analyses and motile sperm subpopulation structure study as improving tools in boar semen quality analysis. *Theriogenology* 61, 673–690.
Rodina M, Cosson J, Gela D & Linhart O. (2004) Kurokura solution as immobilizing medium for spermatozoa of tench (*Tinca tinca* L.). *Aquacult Int* 12, 119–131.
Rurangwa E, Roelants I, Huyskens G, Ebrahimi M, Kime DE & Ollevier F. (1998) The minimum effective spermatozoa: egg ratio for artificial insemination and the effects of mercury on sperm motility and fertilization ability in *Clarias gariepinus*. *J Fish Biol* 53, 402–413.
Rurangwa E, Volckaert FAM, Huyskens G, Kime DE & Ollevier F. (2001) Quality control of refrigerated and cryopreserved semen using computer-assisted sperm analysis (CASA), viable staining and standardized fertilization in African catfish (*Clarias gariepinus*). *Theriogenology* 55, 751–769.
Rurangwa E, Kime DE, Ollevier F & Nash JP. (2004) The measurement of sperm motility and factors affecting sperm quality in cultured fish. *Aquaculture* 234, 1–28.
Silander T, Roos T, Kontkanen P & Myllymaki P. (2008) Factorized normalized maximum likelihood criterion for learning Bayesian network structures. Proceedings of the 4th European Workshop on Probabilistic Graphical Models (PGM-08), Hirtshals, Denmark, pp. 257–272.
Tuset VM, Dietrich GJ, Wojtczak M, Slowinska M, De Monserrat J & Ciereszko A. (2008) Relationships between morphology, motility and fertilization capacity in rainbow trout (*Oncorhynchus mykiss*) spermatozoa. *J Appl Ichthyol* 24, 393–397.
Ward MP & Lewis FI. (2013) Bayesian graphical modelling: applications in veterinary epidemiology. *Prev Vet Med* 110, 1–3.
Wilson-Leedy JG & Ingermann RL. (2007) Development of a novel CASA system based on open source software for characterization of zebrafish sperm motility parameters. *Theriogenology* 67, 661–672.
Wong S & Li C. (2006) Life science data mining. In: *Science, Engineering, and Biology Informatics* (eds R Murphy, B Shapiro & C Wu), World Scientific Publishing, River Edge, NJ, USA.
Yet B, Perkins ZB, Rasmussen TE, Tai NR & Marsh DW. (2014) Combining data and meta-analysis to build Bayesian networks for clinical decision support. *J Biomed Inform* 52, 373–385.

