BIOLOGY OF REPRODUCTION **60,** 32–41 (1999)

Sperm Subpopulations in Boar (*Sus scrofa*) and Gazelle (*Gazella dama mhorr*) Semen as Revealed by Pattern Analysis of Computer-Assisted Motility Assessments

Teresa Abaigar,[2] William V. Holt,[1,3] Robin A.P. Harrison,[4] and Gabriel del Barrio[2]

Estación Experimental de Zonas Aridas,[2] General Segura 1, 04001 Almeria, Spain
Institute of Zoology,[3] Zoological Society of London, Regent's Park, NW1 4RY London, United Kingdom
The Babraham Institute,[4] Babraham, Cambridge, CB2 4AT, United Kingdom

ABSTRACT

The aim of this study was to test the suitability of "pattern analysis" for the exploration of data provided by computer-assisted semen analysis methods. Data sets derived from the examination of boar sperm responses to bicarbonate and caffeine (measurements on 3208 spermatozoa) and from studies of semen cryopreservation in Mohor gazelles (7278 spermatozoa) were reanalyzed. A nonhierarchical classification method was used to generate initial subgroups of spermatozoa (9 for boar, 13 for gazelle). The subgroup centroids were fused, yielding three boar sperm subpopulations and four gazelle sperm subpopulations distinguished by sperm behaviors. Bicarbonate and caffeine both induced major transitions ($p < 0.0001$) of boar sperm behavior, detected as shifts in group membership (from group 2, i.e., active but nonlinear movement, into group 1, i.e., linear, rapid movement). Some spermatozoa (approximately 3%) were refractory to both caffeine and bicarbonate. The gazelle sperm subpopulation structure was affected by the inclusion of equex (sodium triethanolamine lauryl sulfate) in the cryoprotective diluents. Equex suppressed the appearance of spermatozoa with erratic behavior ($p < 0.0001$; high curvilinear velocity, low linearity, low straight-line velocity) after cryopreservation. The proportion of these erratic spermatozoa was positively correlated with animal age ($r = 0.68$, $p = 0.029$). Pattern analysis revealed novel aspects of the data not seen in the original investigations and usefully supplemented the more standard data analysis approaches.

INTRODUCTION

Evidence from a number of sources strongly suggests that different subpopulations of spermatozoa coexist within any typical mammalian ejaculate (see, for example, [1, 2]). These are thought to owe their origins to variations in the assembly of individual spermatozoa during spermatogenesis as well as to differential maturational status and age through mixing in the epididymis. Studying these subpopulations by flow cytometry, through which different attributes are detectable using fluorescent probes, has proved useful in understanding processes such as capacitation [3]. Analogous approaches to the study of sperm motion in separate subpopulations have not, however, been widely adopted despite the obvious potential currently offered by computerized motion analysis systems.

In theory, computer-assisted semen analysis (CASA) systems permit the study of motion characteristics in sperm subpopulations to an unprecedented degree of sophistication. As the trajectories of individual spermatozoa are determined by their flagellar function, characteristics such as sperm velocity, flagellar beat frequency, and amplitude should faithfully reflect the physiological status of individual cells. Large sets of data derived from hundreds or thousands of individual sperm measurements are routinely provided by CASA systems, so the problem of dissecting them should be amenable to established methods of multivariate analysis. This valuable potential of CASA has, however, been rather neglected.

Many publications aimed at identifying motion parameters that best correlate with fertility in both humans and animals have reported average values with an indication of variance [4–9]. This approach diminishes the informative value of the data by making the tacit assumption of normal distributions for all variables. Examination of distributions often reveals highly skewed and bimodal population structures, and it is not surprising that results have often been disappointing.

A few authors have, however, investigated multivariate approaches to sperm motion analysis [10–12]. Davis et al. [11] used a complex iterative multiple-regression technique to analyze human sperm motion parameters, straight-line velocity and linearity, prior to freezing, and suggested that, with this information, they could predict the percentage of sperm that remain viable after thawing. Their predictions were based on knowledge of the difference in straight-line velocity between the subpopulation with the highest value and the subpopulation with the lowest value in each prefreeze specimen. This presents an interesting prospect, as it implies a relationship between the vigor of sperm motion and the cell's resistance to cryoinjury.

Results of this nature [10–12] suggest that it would be possible to use similar techniques for analyzing data from cryopreservation experiments as an alternative or an adjunct to the more common approach of analysis of variance. Indeed, the same principle could apply to any experiment in which sperm motion characteristics are used as a measurable end-point.

The present investigation was undertaken to explore the information provided by pattern analysis of the heterogeneity of sperm motion, as a slightly different approach to the multiparametric analysis of CASA-derived data. The general aim of pattern analysis is data exploration. This is achieved by simplifying the structure of complete data sets obtained from CASA studies, thus reducing both the number of cases (individual observation) and measured descriptors but without losing their information content. Using the computer program PATN [13], the original observations are classified into multidimensionally defined groups, where the group characteristics distinguish different physiological states. The relative sizes of the different groups in response to experimental treatments or other sources of variation can be used as inputs to hypothesis-testing statistics, such as ANOVA or chi-square tests. The PATN program itself represents a collection of statistical modules that are based on well-established statistical principles. Further information about the specific procedures used in PATN can be found

Accepted August 18, 1998.
Received April 21, 1998.
[1]Correspondence. FAX: 44 171 586 2870; e-mail: w.holt@ucl.ac.uk

in the program manual [13] and in standard texts on multivariate statistics (for example, [14]).

To examine the benefits of using pattern analysis with CASA data, we have reanalyzed data from two contrasting experimental studies in which sperm motion measurement was an important analytical component. The first study, of boar spermatozoa, represents a preliminary investigation of the stimulatory effects of bicarbonate and caffeine on sperm movement. Subjective examination of the sperm trajectories suggested that caffeine and bicarbonate elicited different qualities of motion, best described as ''smooth and snaky'' and ''fast but jerky'' for the two stimulators, respectively. However, not all cells in the samples appeared to respond, and this presented an ideal opportunity for comparing the different types of analysis and determining the heterogeneity of the motility response.

The second study was part of an investigation of semen cryopreservation in the Mohor gazelle (*Gazella dama mhorr*) that examined egg yolk and detergent effects on the cryosurvival of spermatozoa. Information from this study was published previously [15], but the CASA data were examined only through comparisons of the summarized treatment means derived from individual data sets. These data have been reassessed for the treatment effects. However, as 10 animals were used in the study, it was possible also to examine sperm heterogeneity in different individuals, relating the responses to variables such as age, body weight, and inbreeding coefficient.

MATERIAL AND METHODS

Boar Semen: Treatments and Experimental Procedures

Semen samples from four boars (one ejaculate per boar) were used in this experiment. Two of the boars, a Landrace and a Large White, were from a commercial artificial insemination (AI) center (JSR Healthbred Ltd., Thorpe Willoughby, UK) and were thus from lines selected for their high fertility as well as other genetic traits. The two other boars, Large Whites, were used as semen donors at the Babraham Institute (Babraham, Cambridge, UK); both produced normal semen and were presumed fertile, although they were not used for matings.

Spermatozoa were received from both the pig breeding center and Babraham after 24-h storage at ambient temperature (20–24°C) in a holding medium (Beltsville Thawing Solution; BTS, 37 g/L glucose monohydrate anhydrous, 6 g/L sodium citrate, 1.25 g/L sodium hydrogen carbonate, 1.25 g/L EDTA-disodium, 0.75 g/L KCl, pH 7.2). The samples were recovered from the BTS medium prior to experimentation using a Percoll (Sigma-Aldrich Company Ltd., Poole, UK) washing technique developed for boar spermatozoa [16]. This step was included as a standard procedure designed to ensure removal of gel particles and other debris that encourage sperm agglutination during subsequent incubations. Samples were resuspended in a bicarbonate-free, modified Tyrode's-based medium [17] buffered with Hepes (Sigma-Aldrich) and containing 5 mg/ml BSA (Sigma-Aldrich).

To examine sperm motility under different incubation conditions, the spermatozoa were further diluted to a concentration of approximately 20×10^6/ml in modified Tyrode's medium supplemented with a) no additional components, b) 15 mM $NaHCO_3$, or c) 1 mM caffeine. Tubes were incubated for 10 min at 39°C before sampling for CASA analysis. To maintain osmolarity, NaCl was suitably reduced in the supplemented media; to maintain pH at 7.4, the bicarbonate-containing medium was kept in equilibration with 5% CO_2 in air throughout incubation [18]. Subsamples of sperm suspension (25 μl) were placed on agar-coated glass slides for motility assessment and video recording. (See Holt et al. [19] for details of the video recording method.)

Gazelle Semen: Treatments and Experimental Procedures

The study was conducted using adult Mohor gazelles (*Gazella dama mhorr*) from the breeding herd maintained and closely managed by the Estación Experimental de Zonas Aridas (see [15] for details of the semen collection and freezing procedures). In brief, semen samples (one ejaculate from each animal) from 10 gazelles (replicates) had been frozen during the course of an investigation into semen cryopreservation in this species. For motility assessment the spermatozoa were not selected using ''swim-up'' or any similar procedure.

The main objectives of this study were to compare the effects of three concentrations of egg yolk (5%, 10%, and 20%), in the presence or absence of 0.5% equex (sodium triethanolamine lauryl sulfate), as components of cryopreservation media. These were tested by measuring sperm motion characteristics of the frozen semen. Straws of semen (0.25 ml) were thawed for 30 sec at 37°C in a water bath and diluted with 0.75 ml of the modified Tyrode's medium described above (lacking bicarbonate). Aliquots (5 μl) were removed for motility assessment and placed in a chamber approximately 30 μm in depth. Video recordings (3 min in length) of each sample were made using an Olympus (Tokyo, Japan) BH2 microscope fitted with a positive phase-contrast objective (×10) and heated stage (37°C). (Details of the recording procedure have been reported previously [15].)

Sperm Motion Analysis

The quantitative sperm motion descriptors were measured using a Hobson Sperm Tracker (Hobson Tracking Systems, Sheffield, UK). Detailed information, including the setup parameters for gazelle spermatozoa, has been published previously [15]. The ''search radius'' and ''minimum track points'' settings for boar spermatozoa were 5.9 μm and 50 frames, respectively. The gazelle spermatozoa were examined using a 25-Hz sampling frequency, and the boar spermatozoa were examined using a 50-Hz sampling frequency.

The measured descriptors of sperm motion were curvilinear velocity (VCL), average path velocity (VAP), straight-line velocity (VSL), amplitude of lateral head displacement (ALH), beat cross frequency (BCF), linearity (LIN), straightness (STR), and time (TIME). Definitions of all these descriptors except TIME can be found in previous publications [20, 21]. The parameter TIME is the duration of an individual sperm track measured in tenths of a second.

Classification, Ordination, and Identification of Sperm Subpopulations

The multivariate analyses were carried out in a number of sequential steps (see Table 1 for a summary of the method). First a non-hierarchical classification of spermatozoa (step 1; Table 1) was undertaken using the ALOC (''allocation'') algorithm [13, 22]. In this procedure the objects (spermatozoa) are allocated to groups according to the magnitudes of their descriptors. The resultant number of groups

TABLE 1. Summary of the statistical strategy used in the PATN analysis.

Step	Process	PATN module or other method*	Main aim of step
1	Nonhierarchical classification	ALOC (Apply Bray and Curtis association threshold, 0.3)	Reduce the number of original objects to produce subgroups. Analysis uses untransformed data.
2	Hierarchical classification	Flexible UPGMA	Cluster the subgroup centroids and arrange as a dendrogram.
3	Hierarchical classification	FUSE	Further simplify the data structure by fusing components of the dendrogram. Provides an objective means of cutting the dendrogram.
4	Ordination (1)	PCoA	Reduce the number of descriptors needed to identify the data structure. Typically 2 or 3 resultant variables (x, y, z axes).
5	Ordination (2)	PCC	Check correlation of the new x, y, and z axes with the original descriptors.
6	Hypothesis testing	ANOVA and chi-square tests	Compare means of derived subsets of spermatozoa (log-transformed data) and frequencies of subsets, in relation to original experimental designs.

* The PATN modules are described in more detail within the text.

depends upon the value of an initial association threshold (Bray and Curtis association index) that must be chosen at the beginning of the analysis. The value chosen defines the resolution at which the multidimensional space defined by the descriptors is sampled. In the present study, the threshold association index was set at 0.3 to yield relatively small numbers of subgroups: 9 and 13 for the boar and gazelle data, respectively. These subgroups were then expressed mathematically as centroids, and the centroids were progressively fused into fewer groups using the Sequential Agglomerative Hierarchical-Combinatorial (SAHN) flexible unweighted pair groups median average (UPGMA) algorithm (step 2; Table 1). Figure 1 shows the dendrograms resulting from the fusion process. Final groups were obtained by cutting the dendrogram at the appropriate distance recommended by the PATN program using the FUSE procedure (step 3; Table 1). In general, the final number of groups is set to the higher integer nearest to the square root of the respective number of subgroups. However, the user retains the option to specify a larger or smaller number of subgroups if so desired.

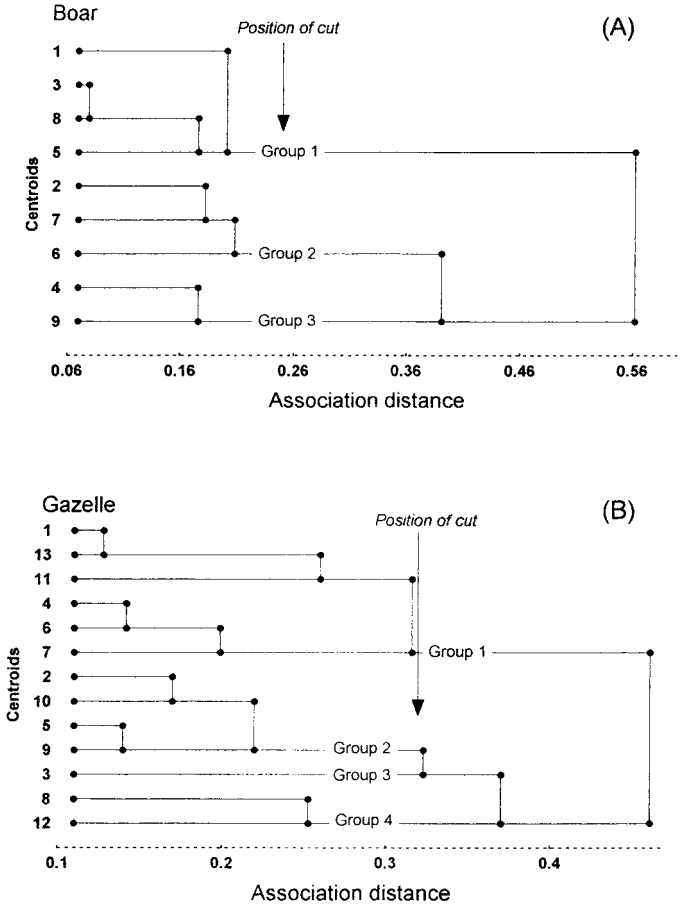

FIG. 1. Dendrograms derived from boar (**A**) and gazelle (**B**) data. Numbers to the left of the plots represent group centroids formed by the fusion of cases (individual spermatozoa). Final reduction of the number of clusters was achieved by cutting the dendrograms at the positions shown along the association distance axis. See *Materials and Methods* for explanation of association distance.

At this point the reduced data sets defined from that hierarchy still referred to groups of centroids rather than individual observations. The characteristics of the individual spermatozoa represented by the centroids were therefore used to calculate new characteristics describing the newly formed sperm groups.

The groups of spermatozoa thus defined were then examined by Principal Coordinates Analysis (PCoA) (step 4; Table 1), which reduced the number of variables (descriptors) defining the groups of observations to a small number of ''vectors'' (combinations of the original motion descriptors). The groups were now expressed in terms of the newly derived vector variables. Principal Axis Correlation (PCC) (step 5; Table 1), a multiple linear regression program designed to see how well a set of attributes can be fitted to an ordination space, was now applied in order to see how well the original groupings (based upon the original descriptors) fitted into the space formed by the vector axes derived from the PCoA [23].

Untransformed data were used for all the steps of the multivariate analysis (steps 1–5; Table 1).

Hypothesis Testing

The final outcome of the pattern analysis was the definition of various subpopulations of spermatozoa within any given sample, distinguished by their multidimensional motion characteristics. The summary statistics of the finally derived groups of spermatozoa were calculated and examined for their physiological relevance, and the log-transformed values [24] were compared using one-way ANOVAs (step 6; Table 1).

Within different samples, the relative frequencies of spermatozoa belonging to each group provided a new variable for the examination of experimental effects and differences between individuals. These untransformed relative frequencies were examined by chi-square (step 6; Table 1) tests and ANOVA using Statistica for Windows (Statsoft UK, Letchworth, UK). To complement the statistical anal-

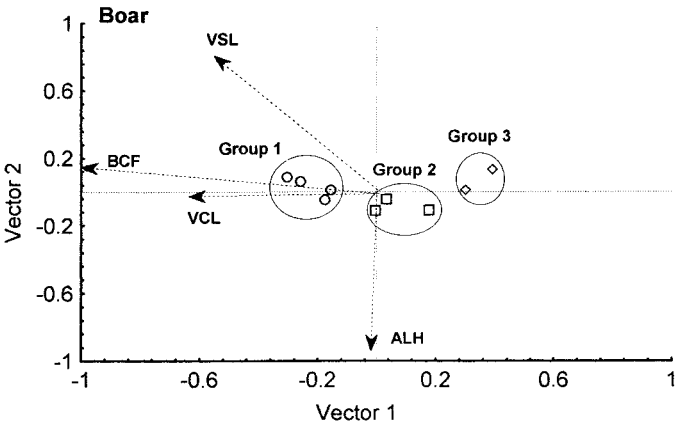

FIG. 2. Plot showing the positions of boar sperm subpopulations in the multidimensional ordination space defined by the first two ordination vectors. Open symbols correspond to the individual centroids shown in Figure 1A, and their allocation to groups is as indicated. Arrows indicate vectors representing individual sperm motion descriptors; these values can be considered as increasing as they tend away from coordinate 0,0 (spatial center).

yses, the effects of treatments within the experimental designs upon individual sperm motion descriptors were examined in standard one- or two-way ANOVAs. For the latter ANOVAs, values were logarithmically transformed prior to analysis [24].

RESULTS

Interpretation of the Multivariate Analyses

The boar sperm data matrix consisted of 3208 observations (individual spermatozoa) and 4 variables (motion descriptors VCL, VSL, BCF, and ALH); these descriptors were chosen as they were not highly intercorrelated. The gazelle data consisted of a matrix of 7278 observations (individual spermatozoa) and 8 variables (all 8 motion parameters measured). In this case there was no a priori selection of descriptors.

Nine group centroids were initially obtained from the nonhierarchical classification of boar spermatozoa by the ALOC algorithm. These group centroids were reduced to three using the SAHN-UPGMA procedure by cutting the dendrogram at an appropriate point on the association distance axis (Fig. 1). Three groups emerged if the cut was placed anywhere between about 0.18 and 0.3 along the horizontal axis of the dendrogram. By the same procedure, 13 group centroids were obtained after the nonhierarchical classification of the gazelle sperm data. These group centroids were further reduced to 4 by cutting the dendrogram at the appropriate level.

Figures 2, 3A, and 3B show the locations of the resultant sperm subpopulations in ordination space. Each symbol represents one of the initial 13 (gazelle) or 9 (boar) groups, and membership of the combined groups is indicated. The arrows represent vectors indicating the contribution of individual descriptors to the position of each centroid in ordination space. For clarity of presentation not all of these vectors are represented as arrows; some are shown as filled symbols plotted at the appropriate coordinates. For the boar data (Fig. 2), vector 1 provided very good discrimination between the groups, accounting for 83.5% of the variance. Those points with negative scores, situated on the left of the plot, are associated with high VCL and BCF. Relatively little extra discrimination, only 9.7% of the variance, is

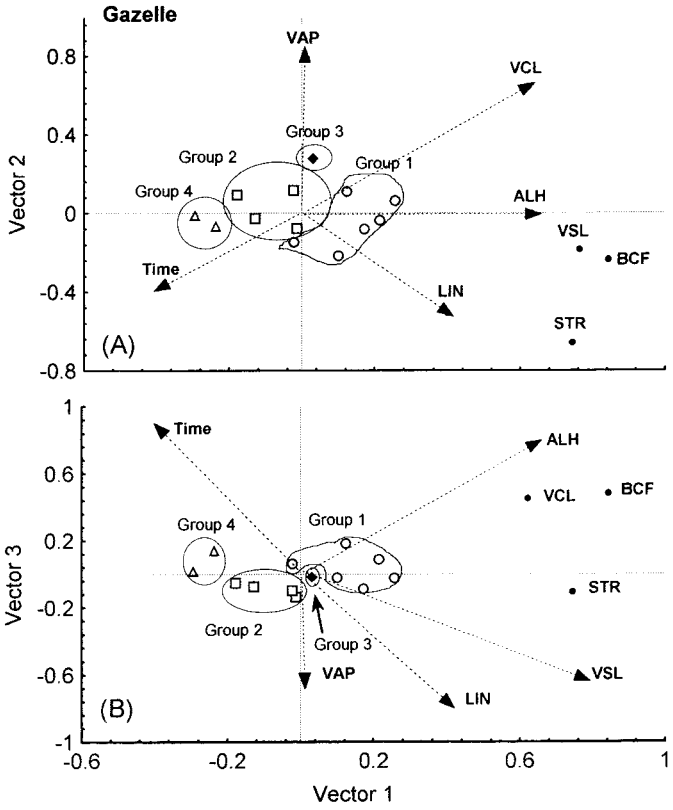

FIG. 3. Plots showing the positions of gazelle sperm subpopulations in the multidimensional ordination space defined by **A**) the first two ordination vectors and **B**) vectors 1 and 3. Open symbols correspond to the individual centroids shown in Figure 1B, and their allocation to groups is as indicated. Arrows indicate vectors representing the sperm motion descriptors; these values can be considered as increasing as they tend away from the spatial center (coordinate 0,0). For the purposes of clarity not all descriptors are represented by arrows; appropriately labeled closed symbols indicate the other vector coordinates.

provided by vector 2; but negative scores indicate increasing ALH values, and probably erratic movement.

For the gazelle sperm there was clearly discrimination between the groups on the first two ordination vectors (Fig. 3A), although some overlap occurred. The first two ordination vectors accounted for 80.2% of the variance. Points with positive scores on vector 1, and located on the right of the plot, were dominated by high values of ALH, VSL, and BCF, i.e., by spermatozoa showing high straight-line velocity and high beat frequency. Those scoring negatively on vector 1 were characterized by high values of TIME and thus represented slow-moving spermatozoa. High positive scores on vector 2 indicate high curvilinear velocity (VAP, VCL). Vector 3 accounted for another 15.7% of the variance but did not provide good discrimination between the groups. This axis mainly discriminates on the basis of speed and linearity; points with high positive TIME and ALH scores are contrasted against high VAP and LIN values, which score negatively.

Group Characteristics

Tables 2 and 3 show the mean (± SD) values for motion parameters of the groups thus defined by the multivariate analyses of the boar and gazelle data, respectively. To explore the differences between the groups, conventional ANOVA analyses were undertaken. Significant differences ($p < 0.05$) between groups were found for all parameters. Ta-

TABLE 2. Means (± SD) of subsets of the boar data arranged by treatment, assigned group, and origin (n = 4 boars, 1 ejaculate per animal).*

Source of variation	No. of sperm in subset	VCL (μm/sec)	VSL (μm/sec)	BCF (Hz)	ALH (μm)
By treatments					
Control	936	139.0 ± 50.4[a]	44.6 ± 40.6[a]	12.5 ± 7.9	10.0 ± 10.0[a]
Bicarbonate	1095	140.7 ± 37.7[c]	91.4 ± 44.5[b]	14.9 ± 6.8	6.4 ± 5.4[b]
Caffeine	1177	178.4 ± 40.9[b]	104.5 ± 44.1[b]	17.3 ± 6.6	7.8 ± 5.3[c]
By assigned groups					
Group 1	2354	174.7 ± 31.7[a]	106.3 ± 34.6[a]	17.9 ± 5.5[a]	7.1 ± 3.9[a]
Group 2	622	112.3 ± 24.5[b]	20.1 ± 15.7[b]	8.9 ± 5.5[b]	12.9 ± 12.7[b]
Group 3	232	58.6 ± 12.5[c]	8.7 ± 6.2[c]	2.4 ± 3.3[c]	4.2 ± 6.6[c]
By animal origin					
AI Center	1895	162.5 ± 44.3[a]	90.0 ± 46.4[a]	16.4 ± 6.9[a]	7.4 ± 5.7[a]
Research Institute	1313	142.2 ± 47.4[b]	71.8 ± 52.9[b]	13.1 ± 7.5[b]	8.8 ± 8.8[b]

* While the raw data are shown, treatment comparisons were evaluated by ANOVA after logarithmic transformation of the data.
[a–c] Different superscripts indicate significant differences within subsets within columns ($p < 0.05$).

bles 2 and 3 also show comparisons between data variously partitioned into subsets and examined by conventional ANOVAs for experimental treatment effects.

Boar Data

The CASA system is capable of tracking nearly all spermatozoa in the field of view; however, if spermatozoa are completely immotile (i.e., no flagellar movement whatever) they will not be detected and consequently are excluded from the measurements. In the boar sperm experiments reported here, very few immotile sperm were present, and therefore the sperm groups defined by the multivariate analysis represented genuine motility subpopulations within the overall populations (i.e., the Percoll-washed samples). Within the data set, three types of sperm motion were recognizable by pattern analysis. Group 1 represented those sperm with highly progressive movement (high VSL) and with vigorous flagellar action (high BCF); group 2 represented spermatozoa showing an active type of movement (high VCL) but with considerably reduced forward progression (lower VSL and higher ALH). Group 3 appeared to represent a slow and possibly degenerate class of cells, for which all motion parameters were significantly lower than in either group 1 or 2.

Examination of the experimental treatment effects upon the allocation frequencies of individual spermatozoa to groups 1, 2, or 3 (Fig. 4) revealed that both caffeine and bicarbonate induced a major shift of spermatozoa out of group 2 and into group 1 ($\chi^2 = 496.5$, 4 *df*, $p < 0.0001$). A small proportion of group 2 spermatozoa remained unaffected by the stimulatory treatments. Although group 3 was relatively small in the control treatment (3.6% of the total measured sperm population), even this was reduced in response to caffeine and bicarbonate (to 1.3% and 2.4% of the total data set, respectively). Nearly half of the control spermatozoa were classified as group 1. This effect was mainly attributable to the highly asymmetric partitioning of group 1 spermatozoa, the most progressive, between ejaculates from the AI center and the research institute ($\chi^2 = 157.6$, 2 *df*, $p < 0.0001$). Overall, approximately twice as many group 1 spermatozoa originated from the AI center boars as from the research institute boars (Fig. 5).

Conventional ANOVA confirmed that spermatozoa responded to bicarbonate and to caffeine by showing considerably increased velocity (Table 2). Average values for VCL and VSL were both increased ($p < 0.001$), with simultaneous reductions in ALH. Caffeine stimulation produced significantly higher VCL and ALH ($p < 0.001$), but not VSL, than bicarbonate. These results indicate that the cells adopted a more progressive type of movement when exposed to these effectors; but while bicarbonate induced smooth forward progression, caffeine produced more lateral deviation from the average path, resulting in a jagged trajectory. Subjective examination of the video recordings

TABLE 3. Means (± SD) of subsets of the gazelle data arranged by treatment and assigned group (n = 10 gazelles, 1 ejaculate per animal).*

Source of variation	No. of sperm in subset	VCL (μm/sec)	VAP (μm/sec)	VSL (μm/sec)	BCF (Hz)	ALH (μm)	LIN (%)	STR (%)	Time (0.1 sec)
By assigned group									
Group 1	2630	127.8 ± 48.7[a]	53.0 ± 21.8[a]	41.3 ± 19.9[a]	1.51 ± 1.0[a]	31.5 ± 21.3[a]	35.8 ± 18.3[a]	78.4 ± 22.2[a]	9.9 ± 7.4[a]
Group 2	3446	87.9 ± 21.1[b]	87.3 ± 20.9[b]	23.0 ± 11.9[b]	0.02 ± 0.24[b]	0.15 ± 1.79[b]	27.6 ± 15.1[b]	27.8 ± 15.2[b]	7.6 ± 3.3[b]
Group 3	732	160.0 ± 26.2[c]	158.9 ± 27.1[c]	22.0 ± 14.4[c]	0.03 ± 0.32[b]	0.20 ± 1.94[b]	13.8 ± 9.3[c]	13.9 ± 9.3[c]	7.1 ± 3.0[c]
Group 4	470	61.0 ± 15.1[d]	51.9 ± 11.7[d]	4.6 ± 2.9[d]	0.08 ± 0.46[b]	0.63 ± 3.55[b]	8.2 ± 5.0[d]	10.0 ± 6.5[d]	19.9 ± 21.9[b]
By egg-yolk concentration									
5%	2530	108.3 ± 42.7[c]	80.5 ± 39.3[a]	27.8 ± 18.5[a]	0.57 ± 0.95[a]	11.5 ± 19.4[c]	27.3 ± 17.3[b]	43.3 ± 32.1[c]	9.2 ± 8.0[c]
10%	2242	98.1 ± 36.9[b]	76.1 ± 33.9[b]	26.2 ± 18.6[b]	0.49 ± 0.90[b]	9.1 ± 17.4[b]	27.8 ± 18.6[b]	41.3 ± 32.0[b]	10.9 ± 11.5[b]
20%	2506	116.0 ± 47.9[a]	82.5 ± 39.2[a]	30.7 ± 18.5[a]	0.62 ± 1.01[a]	13.7 ± 21.9[a]	28.6 ± 16.6[a]	45.8 ± 31.4[a]	7.5 ± 2.1[a]
By presence of equex									
Present	3554	114.6 ± 45.6[a]	84.6 ± 39.4[a]	30.2 ± 18.9[a]	0.59 ± 0.98	12.7 ± 20.8[a]	28.4 ± 17.4[a]	44.3 ± 32.0	7.4 ± 1.7[a]
Absent	3724	101.4 ± 40.4[b]	75.3 ± 35.6[b]	26.5 ± 18.2[b]	0.54 ± 0.94	10.3 ± 18.7[b]	27.5 ± 17.6[b]	42.8 ± 31.7	10.9 ± 10.9[b]

* While the raw data are shown, treatment comparisons were evaluated by ANOVA after logarithmic transformation of the data.
[a–d] Different superscripts indicate significant differences within subsets within columns ($p < 0.05$).

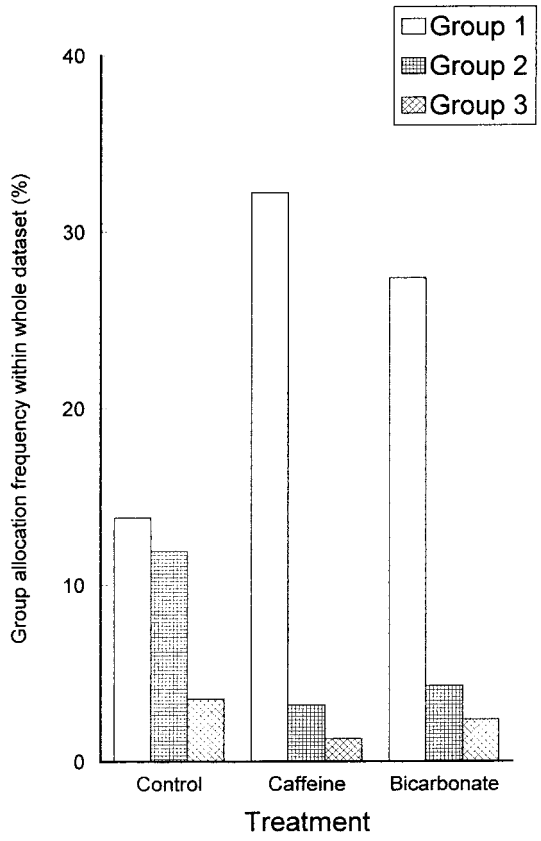

FIG. 4. Relative proportions (%) of boar spermatozoa in the whole data set classified as groups 1, 2, or 3 in relation to treatment with **A**) no added effectors, **B**) caffeine, or **C**) bicarbonate. Group 1 was significantly increased, and group 2 decreased, by caffeine and bicarbonate ($\chi^2 = 496.6$, 4 *df*, $p < 0.0001$).

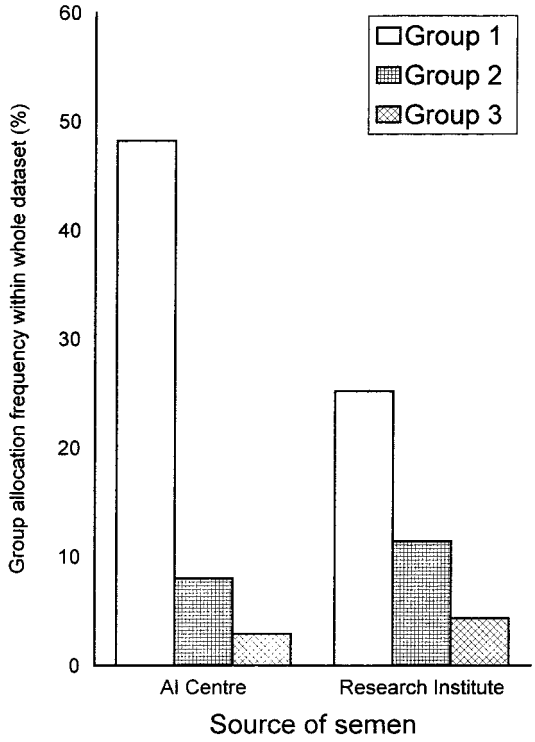

FIG. 5. Relative proportions (%) of boar spermatozoa in the whole data set classified as group 1, 2, or 3 in relation to source of semen. Group 1 was significantly higher, and group 2 lower, in boars from the AI center compared to the research institute ($\chi^2 = 157.6$, 2 *df*, $p < 0.0001$).

confirmed this interpretation. The source of boar semen also had a significant impact upon the mean sperm motion parameters; sperm samples from the AI center showed significantly higher values for VCL, VSL, and BCF ($p < 0.001$).

Graphical support for the PATN analysis results was obtained by examining two-dimensional scatterplots of LIN vs. VAP (Fig. 6, A–C). These plots confirmed that both bicarbonate and caffeine elicited a major, but variable, shift of points from the bottom left of the graphs (slow, nonlinear motion) toward the top right (rapid, linear motion). The two research institute boars (Babraham 1 and 2) showed this effect much more markedly than did those from the AI center (Landrace and Large White).

Gazelle Data

As the gazelle spermatozoa were frozen and thawed, and not subjected to methods for selecting motile cells, there were variable proportions of immotile cells in the samples (overall mean ± SD, 31.7 ± 22.2%). The percentage motility data have been reported previously [15]; in essence, increasing egg yolk concentrations resulted in decreasing proportions of motile cells.

Interpreting the differences between pattern-derived groups of spermatozoa was less straightforward than for the boar data, presumably reflecting the poorer resolution achieved by the analysis. Groups 1 and 4, being farthest apart in ordination space (Fig. 3A), were easily distinguished. Group 1 spermatozoa (36% of the data set) had extremely high, but also highly variable, mean ALH values (31.5 ± 21.3 μm). This indicates nonlinear but rapid motion, an interpretation supported by the high VCL/VSL ratio and the relatively low mean LIN value (35.8%). Group 4, being associated with the highest TIME and extremely low VSL values, made slow progress, staying in the measurement window for the longest duration. Groups 2 and 3 (47% and 10% of the data set, respectively) seem somewhat similar, except that group 3 showed higher VCL and VAP values than group 2. Combined with the lower LIN values, this suggests that spermatozoa in group 3 were less progressive, and more erratic, than those in group 2.

Whereas the pattern analysis produced evidence for a complex subpopulation structure within the data, conventional ANOVA, while valuable in its own right, missed this point completely. Thus, when the data were partitioned with respect to egg yolk concentration (Table 3), there was little indication that egg yolk affected motion parameters. Where statistically significant differences were detectable, they were rather small. Nevertheless, significant differences due to egg yolk concentration were observed in the distribution of sperm subpopulations ($\chi^2 = 127.2$, 6 *df*, $p < 0.0001$; Fig. 7). Most spermatozoa were classified as groups 1 or 2, with group 2 being most frequently observed in every treatment. Group 4 slow-moving spermatozoa were more frequently observed in the 10% egg yolk diluent than either the 20% or 5%. In contrast, group 3, which denoted spermatozoa with the highest values of VCL, were more abundant in diluents with 20% and 5% egg yolk. This pattern of response is difficult to explain, but it may be due to the detrimental effects of egg yolk on the overall percentage motility of gazelle spermatozoa [15].

Significant differences in sperm group allocation frequencies due to the inclusion of equex were observed (χ^2

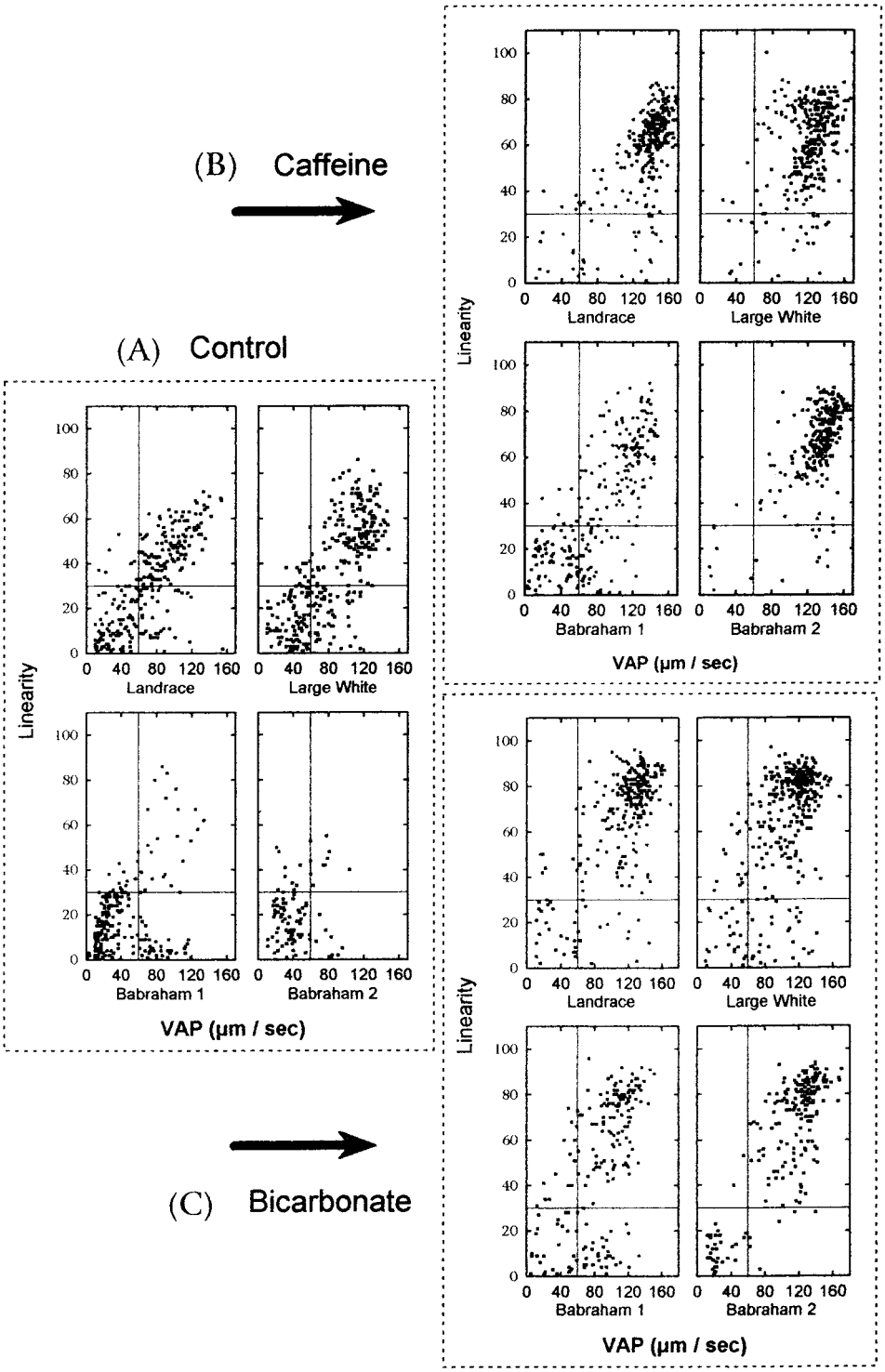

FIG. 6. Two-dimensional scatterplots showing changes in VAP and LIN of boar spermatozoa in response to caffeine and bicarbonate. Each plotted point represents a single spermatozoon, and each plot represents within-boar data. Landrace and Large White refer to boars from the AI center, while Babraham 1 and Babraham 2 are the boars from the Babraham Institute. **A** represents spermatozoa in Tyrode's medium without supplementation. **B** and **C** represent sperm motion in the same samples after 10-min incubation in caffeine- or bicarbonate-containing medium, respectively.

= 116.8, 3 *df*, $p < 0.0001$). Approximately equal proportions of groups 1 and 2 spermatozoa were present regardless of equex. However, the inclusion of equex strongly suppressed the appearance of group 3 spermatozoa within the samples. The characteristics of group 3 would be most consistent with erratic behavior, i.e., high VCL, low VSL, and low LIN, indicating that the detergent exerted a subtle protective effect on the swimming behavior of a sizable proportion of spermatozoa. This difference in subpopulation structure resulted in significantly higher sperm velocity parameters due to equex when the data were examined by conventional ANOVA (VCL, VAP, and VSL; $p < 0.001$); significantly higher ALH and LIN values were also detected ($p < 0.05$). The underlying explanation for these differences is, however, by no means clear from the ANOVA alone.

Significant differences in the distributions of the 4 groups of spermatozoa between the various individual semen samples ($\chi^2 = 1086.3$, 27 *df*, $p < 0.0001$) were observed. To analyze the factors responsible for this interejaculate variation, we examined attributes of the 10 animals represented in the study (i.e., age, weight, inbreeding

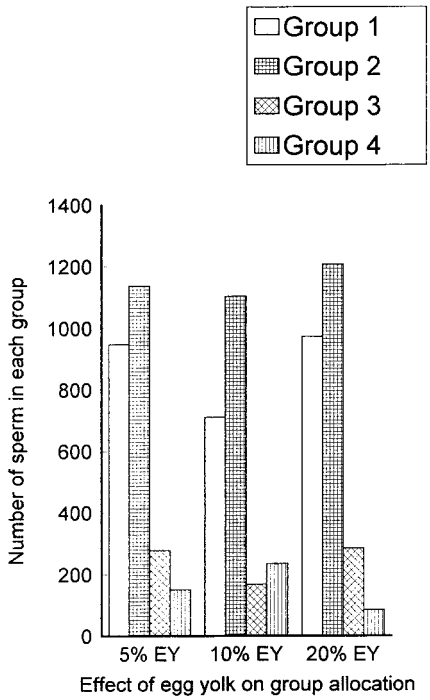

FIG. 7. Numbers of gazelle spermatozoa classified as groups 1–4, partitioned in relation to the concentration of egg yolk. Group 3 was least numerous, and group 4 most numerous, in the presence of 10% egg yolk ($\chi^2 = 116.8$, 3 *df*, $p < 0.0001$).

coefficient, and herd management condition [i.e., whether maintained with females, other males, or individually]). Further details of these factors have been published previously [15]. The proportion of group 3 spermatozoa in ejaculates (erratic movement) was positively correlated with animal age ($F(1,8) = 7.03$, $p = 0.029$, $r = 0.68$). There was some evidence that the proportion of group 4 spermatozoa (lowest quality) was positively correlated with inbreeding coefficient, but this was due mainly to a single data point. No significant effects due to weight and herd management condition of the animals were found. It is important not to overinterpret these findings in terms of interanimal variability, as only one ejaculate was studied from each animal; however, repeated sampling of the same animals was not possible or practical within the design of the current experiment.

DISCUSSION

Statistical Issues

Although many CASA studies have now been published, the results have in many cases been disappointing or uninformative. To some extent this is probably due to inappropriate analysis of the multivariate data sets, most reports having presented mean values (± standard deviations or standard errors), with analyses based on parametric statistics. Examination of typical CASA data shows, however, that the ''normal'' statistical distribution is rarely encountered; bimodal and skewed distributions often occur, together with remarkably wide ranges that reflect different sperm subpopulations within a sample. Such ranges tend to mask treatment effects if parameters are examined individually, because the standard deviations are too high to permit their detection by variance-based statistical tests. This problem is caused partly by the existence of sperm subpopulations; even if some sperm subpopulations are affected by a treatment, other nonresponding subpopulations may confound attempts to analyze for increased or decreased population means. One solution to this problem is to evaluate the subpopulation structure of the semen samples before undertaking hypothesis testing, using the multivariate nature of the CASA data to aid in recognition of different types of spermatozoa within the overall sample. The advantage of undertaking this type of analysis has been recognized previously; however, the few authors who have examined ways to characterize the subpopulations have mainly relied on the K-means cluster analysis approach. In the present study an alternative method of subpopulation analysis was explored, and the results of two different types of investigation have been presented. These analyses were of interest both for their new insights into the experimental results themselves and also for the general demonstration of the statistical approach.

The nonhierarchical classification of objects followed by a SAHN-UPGMA has several advantages over the K-means cluster analysis approach. As a nonparametric procedure, it does not require transformation of data prior to analysis because the outcome is independent of the variable distribution. Furthermore, differences in scale between different variables do not cause unequal weighting of the analysis. These two aspects are important, as changes of relationships and dependencies within the data are thereby avoided. In K-means cluster analysis, parameters such as VCL (which typically range from tens to hundreds of micrometers per second) heavily outweigh the influence of parameters such as ALH (where the scales are measured in tens and units). In K-means cluster analysis, the choice of ''seeds'' (cases chosen to form the initial groupings) can influence the allocation of subsequent cases to clusters. This problem is avoided using the ALOC approach because the resultant group structure will always be the same regardless of the seed chosen. Both types of analysis suffer from a certain amount of subjectivity in deciding upon the appropriate number of clusters that best suit the data. A number of procedures have been developed to guide this decision, e.g., examining the gains or losses in between-groups and within-groups sums of squares as each new cluster is successively added (for techniques, see Sharma [25]). The principle behind such procedures is that groups should be as homogeneous as possible (i.e., low ''within-groups'' sums of squares), but as different as possible (i.e., high ''between-groups'' sums of squares). The pattern analysis algorithms also have a major computational advantage over techniques such as nonhierarchical (K-means) cluster analysis that have been used previously; it is unnecessary to build a similarity matrix—which, if seven thousand cases are involved as in the gazelle study, requires over 100 000 megabytes of computer memory.

There was a fundamental difference between the data sets examined in this study. The boar sperm samples had not been damaged by cryopreservation and thus contained a very high proportion of motile cells. In contrast, the frozen-thawed gazelle samples contained variable proportions of immotile cells, which were excluded from the subsequent analysis but which could be considered as a valid group of their own (all descriptors set at zero). This shortcoming suggests that some advantage may be gained in future if data sets of this type are expanded prior to analysis by inclusion of the appropriate proportions of immotile spermatozoa.

Conclusions from the Analysis of Boar Data

The boar sperm experiment was originally performed to see whether bicarbonate and caffeine stimulated spermatozoa into quantitatively different types of motion, given that it was possible to detect the difference by subjective assessment. Conventional analysis of treatment means (Table 2) revealed that this was indeed possible, the increased lateral motion after caffeine stimulation being translated into higher VCL and ALH values.

Use of the multivariate approach, however, threw entirely new light on the responses to caffeine and bicarbonate. It was not apparent from the mean values alone that the control preparations contained a sizable proportion of group 1 (highly active and progressive) spermatozoa prior to stimulation. It was equally difficult to see that the stimulated samples still retained a small number of group 3 spermatozoa. This difference was more obvious from the two-dimensional scatterplots, and this graphical approach to data presentation has much to recommend it. The reasons behind the heterogenic response to stimulation cannot be explained within the present study, but the response is consistent with previous flow cytometric and biochemical data. Subpopulations of spermatozoa showing different merocyanine 540 responses to bicarbonate stimulation were noted by Harrison et al. [17]. Bicarbonate caused a rapid increase in the merocyanine 540 binding of many, but not all, boar spermatozoa while also causing elevation of intracellular cAMP concentration [26]. Cyclic AMP has long been recognized as an effector of sperm motility [27, 28], and it is therefore likely that motility stimulation is correlated with the merocyanine response. As merocyanine-binding ability is thought to be positively correlated with membrane fluidity [29], these results suggest that bicarbonate may be stimulating cAMP, membrane fluidity, and motility in rapid succession. Caffeine, a phosphodiesterase inhibitor, also elevates intracellular sperm cAMP concentrations [30]. It seems reasonable to conclude, therefore, that the motility responses seen in the present study are brought about via higher cAMP concentrations. If this is indeed the case, two interesting questions emerge. Why do some motile cells not respond to the stimulus? And why did many spermatozoa from the AI center sustain progressive motility in the absence of cAMP-inducing stimuli, even though the samples still showed significant responses to both bicarbonate and caffeine? Considerable further work will be needed to provide explanations for these questions.

Conclusions from Analysis of the Gazelle Data

In the gazelle study, four sperm subpopulations were distinguishable in the frozen-thawed ejaculates. The simplest interpretation of the subpopulation structure is that groups 1–4, respectively, represented declining levels of sperm quality as reflected by decreasing VSL values. However, contrasting values of the other associated variables indicate that the spermatozoa in each group showed different types of swimming behavior. Group 4, the smallest group but the one with the lowest velocity values, could represent a subgroup of metabolically compromised sperm, shortly destined to lose their motility altogether. This would be consistent with the relatively high proportions of immotile spermatozoa in these cryopreserved samples. Group 3 exhibited particularly high VCL and VAP values; the low LIN indicates that these cells could have been either hyperactivated or showing erratic and uncoordinated movement, but subjective assessment of the samples did not seem consistent with high frequencies of hyperactivation. Groups 1 and 2 seemed to represent progressive spermatozoa, the main difference being that while group 1 had high VSL and ALH values, group 2 showed lower values for both variables. Although there might have been merit in combining groups 2 and 3, the definition of four sperm groups served to simplify and enhance the interpretation of the treatment effects.

Analyzing the effects of egg yolk concentration in terms of the group structures revealed that most of the spermatozoa in the samples were categorized as groups 1 or 2 and that egg yolk concentration made little difference. Groups 3 and 4 were considerably less abundant in all egg yolk concentrations tested but showed reversal of frequency in the 10% yolk. Realistically this effect may be considered relatively insignificant. However, this observation is at variance with conclusions reached through examination of the data as raw means, according to which egg yolk concentration significantly affected each variable for reasons that are by no means obvious. Similar considerations also apply when the effects of equex are examined; again the frequencies of groups 3 and 4 spermatozoa were reversed by the presence of equex—but since these groups represented only a minority of cells in the samples, this may mean little. In this case, however, all of the velocity parameters were significantly higher when equex was present, which may indicate that equex had a beneficial effect in maintaining the physiological integrity of the flagellum. This would be consistent with results of other studies showing cryoprotective effects of equex on ungulate spermatozoa, although most reports concern acrosomal integrity [31–33]. No beneficial effect of equex on the subjectively estimated percentage motility scores was detected in the original data analysis [15]; the current results suggest, however, that this may have been due to poor sensitivity of the subjective analysis technique.

The significant correlation between age and the incidence of the highly active group 3 spermatozoa is of interest, albeit rather difficult to interpret. Group 3 probably represents an uncoordinated type of motion, rather than hyperactivation, suggesting that these cells are dysfunctional or at least unusually cryosensitive. Moreover, the correlation suggests that production of poor-quality spermatozoa increases with advancing age. Such a hypothesis implies that semen cryopreservation will be more successful in younger animals.

General Conclusions

The nature and origin of sperm subpopulations remain a matter for speculation. Nevertheless, there is considerable interest in the possibility that the variability reflects discrete functional, and possibly adaptive, differences. If so, the sperm subpopulations are likely to originate during spermatogenesis, when heterogeneous genotypic effects are still capable of affecting the quality of individual cells. This view is supported by extensive studies on the t-haplotype in mice [34, 35] showing that the phenotypic characteristics (impaired motility and sperm transport) of some, but not all, spermatozoa of heterozygotes are attributable to specific mutations affecting the individual spermatozoa. Similarly, observations that the frequency of morphologically abnormal spermatozoa is increased by inbreeding [36] suggest that the process of sperm modeling is directly affected during spermatogenesis in some, but not all, cells.

The use of pattern analysis appears to be a relatively simple approach to the investigation of sperm subpopula-

tion structures, supplementing rather than replacing the more established types of data analysis. For sperm motility analysis it has a distinct advantage over some other approaches, such as neural net-based methods [37], which require the operator to ''teach'' the computer how to recognize different sperm types. Such approaches are more applicable where the different types of spermatozoa are already recognizable, such as in morphology assessment. The advantage of pattern analysis is that it requires no such preconceptions about sperm classification and can suggest new and unsuspected data structures. The enhanced information thereby obtained may generate new hypotheses in topics such as sperm competition theory, as well as in the more diagnostic aspects of andrology.

ACKNOWLEDGMENTS

We are grateful to JSR Healthbred Ltd. for supplying the boar semen used in this study. We also wish to thank Drs. Juana Ortiz and Paz Herraez for assistance with gazelle semen evaluation and Drs. Roberto Lazaro and Cristina Armas for statistical assistance.

REFERENCES

1. Chang MC, Hunter RHF. Capacitation of mammalian sperm: biological and experimental aspects. In: Hamilton DW, Greep RO (eds.), Handbook of Physiology. Endocrinology V. Washington: American Physiological Society; 1975: 339–351.
2. Bedford JM. Significance of the need for sperm capacitation before fertilization. Biol Reprod 1983; 28:108–120.
3. Harrison RAP. Capacitation mechanisms, and the role of capacitation as seen in eutherian mammals. Reprod Fertil Dev 1996; 8:581–594.
4. Holt WV, Moore HDM, Hillier SG. Computer-assisted measurement of sperm swimming speed in human semen: correlation with in vitro fertilization assays. Fertil Steril 1985; 44:112–119.
5. Holt C, Holt WV, Moore HDM, Reed HCB, Curnock RM. Objectively measured boar sperm motility parameters correlate with the outcomes of on-farm inseminations: results of two fertility trials. J Androl 1997; 18:312–323.
6. Marshburn PB, McIntire D, Carr BR, Byrd W. Spermatozoal characteristics from fresh and frozen donor semen and their correlation with fertility outcome after intrauterine insemination. Fertil Steril 1992; 58: 179–186.
7. Barratt CLR, Tomlinson MJ, Cooke ID. Prognostic significance of computerized motility analysis for in vivo fertility. Fertil Steril 1993; 60:520–525.
8. Krause W. Computer-assisted semen analysis systems—comparison with routine evaluation and prognostic value in male-fertility and assisted reproduction. Hum Reprod 1995; 10:60–66.
9. MacLeod IC, Irvine DS. The predictive value of computer assisted semen analysis in the context of a donor insemination programme. Hum Reprod 1995; 10:580–586.
10. Davis RO, Overstreet JW, Asch RH, Ord T, Silber SJ. Movement characteristics of human epididymal sperm used for fertilization of human oocytes in vitro. Fertil Steril 1991; 56:1128–1135.
11. Davis RO, Drobnis EZ, Overstreet JW. Application of multivariate cluster, discriminate function, and stepwise regression analyses to variable selection and predictive modeling of sperm cryosurvival. Fertil Steril 1995; 63:1051–1057.
12. Holt WV. Can we predict fertility rates? Making sense of sperm motility. Reprod Domest Anim 1996; 31:17–24.
13. Belbin L. PATN, Pattern Analysis Package. Division of Wildlife and Ecology. Canberra, Australia: CSIRO; 1993.
14. Sneath PHA, Sokal RR. Numerical Taxonomy. San Francisco: W.H. Freeman; 1973.
15. Holt WV, Abaigar T, Jabbour HN. Oestrous synchronization, semen preservation and artificial insemination in the Mohor gazelle (*Gazella dama mhorr*) for the establishment of a genome resource bank programme. Reprod Fertil Dev 1996; 8:1215–1222.
16. Harrison RAP, Mairet B, Miller NGA. Flow cytometric studies of bicarbonate-mediated Ca^{2+} influx in boar sperm populations. Mol Reprod Dev 1993; 35:197–208.
17. Harrison RAP, Ashworth PJC, Miller NGA. Bicarbonate/CO_2, an effector of capacitation, induces a rapid and reversible change in the lipid architecture of boar sperm plasma membrane. Mol Reprod Dev 1996; 45:378–391.
18. Suzuki K, Ebihara M, Nagai T, Clarke NGE, Harrison RAP. Importance of bicarbonate/CO_2 for fertilization of pig oocytes in vitro and synergism with caffeine. Reprod Fertil Dev 1994; 6:221–227.
19. Holt C, Holt WV, Moore HDM. Choice of operating conditions to minimize sperm subpopulation sampling bias in the assessment of boar semen by computer-assisted semen analysis. J Androl 1996; 17: 587–596.
20. Boyers SP, Davis R, Katz DF. Automated semen analysis. Curr Probl Obstet Gynecol Fertil 1989; 12:172–200.
21. Davis RO, Siemers RJ. Derivation and reliability of kinematic measures of sperm motion. Reprod Fertil Dev 1995; 7:857–869.
22. Belbin L. The use of non-hierarchical allocation methods for clustering large sets of data. Aust Computer J 1987; 19:31–41.
23. Belbin L. PATN User's Guide. Canberra, Australia: CSIRO; 1991.
24. Zar JH. Biostatistical Analysis. New Jersey: Prentice Hall; 1984: 718.
25. Sharma S. Applied multivariate techniques. New York: John Wiley and Sons Inc.; 1996: 493.
26. Harrison RAP. Sperm plasma membrane characteristics and boar semen fertility. J Reprod Fertil Suppl 1997; 52:195–211.
27. Tash JS, Means AR. Cyclic adenosine 3′,5′ monophosphate, calcium and protein phosphorylation in flagellar motility. Biol Reprod 1983; 28:75–104.
28. Lindemann CB, Kanous KS. Regulation of mammalian sperm motility. Arch Androl 1989; 23:1–22.
29. Stillwell W, Wassall SR, Dumaual AC, Ehringer WD, Browing CW, Jenski LJ. Use of merocyanine (mc540) in quantifying lipid domains and packing in phospholipid-vesicles and tumor-cells. Biochim Biophys Acta 1993; 1146:136–144.
30. Rees JM, Ford W, Hull M. Effect of caffeine and of pentoxifylline on the motility and metabolism of human spermatozoa. J Reprod Fertil 1990; 90:147–156.
31. Pontbriand D, Howard JG, Schiewe MC, Stuart LD, Wildt DE. Effect of cryoprotective diluent and method of freeze-thawing on survival and acrosomal integrity of ram sperm. Cryobiology 1989; 26:341–354.
32. Howard JG, Bush M, Devos V, Schiewe MC, Pursel VG, Wildt DE. Influence of cryoprotective diluent on post-thaw viability and acrosomal integrity of spermatozoa of the African elephant (*Loxodonta africana*). J Reprod Fertil 1986; 78:295–306.
33. Monfort SL, Asher GW, Wildt DE, Wood TC, Schiewe MC, Williamson LR, Bush M, Rall WF. Successful intrauterine insemination of Eld's deer (*Cervus eldii thamin*) with frozen-thawed spermatozoa. J Reprod Fertil 1993; 99:459–465.
34. Olds-Clarke P, Sego R. Calcium alters capacitation and progressive motility of uterine spermatozoa from +/+ and congenic $tw^{32/+}$ mice. Biol Reprod 1992; 47:629–635.
35. Olds-Clarke P, Wivell W. Impaired transport and fertilization in vivo of calcium-treated spermatozoa from +/+ or congenic $tw^{32/+}$ mice. Biol Reprod 1992; 47:621–628.
36. Roldan ERS, Cassinello J, Abaigar T, Gomendio M. Inbreeding, fluctuating asymmetry and ejaculate quality in an endangered ungulate. Proc R Soc Lond B 1998; 265:243–248.
37. Steigerwald P, Krause W. Estimation of sperm morphology using a new CASA system. Andrologia 1998; 30:15–22.