χ^2=484.4, df=91, p<0.0001). FA was done using the correlation matrix and principal axis factoring; the factors were extracted based on eigenvalue >1 and scree-plot. The solution was Varimax rotated with Kaiser normalization. The analysis was repeated with an oblique (direct Oblimin) rotation.

Thirdly, I calculated the factor scores for the individuals with the regression method based on the final FA solution. These scores were compared across the zoos with one-way ANOVA or Kruskal–Wallis (K–W) tests, depending on the data distribution. Factor scores were also compared between males and females with independent samples t tests or Mann–Whitney U tests, depending on the data distribution. Analyses were done with SPSS 16.0 and 19.0.

Results

Nearly all tested potential personality variables were moderately or highly repeatable across all zoos (Table 2). Only one variable (frequency of aggression received) failed to show repeatability in two of the three zoos. Nine variables were repeatable in all three zoos and in both long-term and short-term AR data. Six additional variables were repeatable in all three zoos, but this was true only for the long-term data in AR (frequency of being approached, frequency of play invitations accepted and grooming diversity), or were repeatable in two out of the three zoos (activity, frequency of grooming received and frequency of play initiated). Thus, 15 behaviour variables were deemed repeatable based on sufficiently high scores in at least two out of the three zoos.

There were no consistent differences in repeatability values of indoor and outdoor observations of AR data (Wilcoxon signed ranks (WSR), z=1.48, p=0.14, N=16; data not presented). There was also no temporal difference in the repeatability of AR data. The long-term repeatability that covered nearly continuous observations across 3 years had repeatability that were similar overall to data from two consecutive winter observation periods (WSR, z=1.89, p=0.059, N=16). Long-term repeatability was higher than short-term repeatability for some variables, whilst short-term repeatability was higher than long-term repeatability for others (Table 2).

The 15 variables that were repeatable in at least two of the three zoos were included in the factor analysis of intercorrelational structure. The frequency of received aggression was excluded as an unrepeatable variable. After the initial run, the frequency of submissive behaviour was excluded due to poor loading in any factor (maximum loading 0.26) and poor communality (0.15). The remaining 14 variables were analysed again with FA, and 5 factors that explained 77.3% of variance were extracted. These were orthogonally rotated, and the solution is presented in Table 3.

The first factor explained 25.0% of the variance. It included strong loadings of the frequency of grooming given and received, the number of individuals in close proximity and the frequency of being approached with a neutral or positive response by a focal subject. These behaviours are sociopositive, reflecting relationship maintenance and sociability. Consequently, this factor was labelled *sociability*. The second factor explained 17.2% of the variance. Variables that loaded on this factor were the frequency of approaching others (non-aggressively), frequency of initiating and receiving play and frequency of short duration affinitive behaviours such as kiss, embrace and sexual behaviours. Therefore, this factor was labelled *positive affect*. The third factor explained 15.2% of the variance and had loadings from the two grooming indices, diversity and density. As they both loaded positively on the factor, and higher GDI indicates more equitable distribution of grooming given, this indicates that a high number of grooming partners correlated with a more equal distribution of grooming effort amongst them. Therefore I labelled this factor *equitability*. The fourth factor, which explained 11.4% of the variance, had high loadings from SDB. As SDBs are considered indicators of anxiety, I labelled this factor as *anxiety*. Finally, the fifth factor explained 8.4% of the variance and had loadings of activity, the frequency of aggression and (negatively) the frequency of proximity initiated. However, this factor has to be treated with caution because aggression and proximity initiation had relatively weak loadings (Budaev 2010) (Table 2). I labelled the factor *activity* but consider it less reliable than the first four factors.

Reanalysis with an oblique (direct Oblimin) rotation did not change much the solution; correlations between the factors did not exceed ±0.21 (range −0.01–0.209) (Table 4). Three variables loaded>±0.40 on an additional factor (indicated as footnote "a" in Table 3).

The factor scores of individuals differed between males and females in all but the first factor (sociability, Mann-Whitney U test (MWU) N_{females}=55, N_{males}=20; z=0.61, p=0.541) (Fig. 1). In positive affect, equitability, anxiety and activity, the male scores were significantly higher than female scores (in all analyses N_{females}=55, N_{males}=20, positive affect, MWU, z=2.40, p=0.017; equitability t test, F=2.11, t=−3.33, p=0.001; anxiety, MWU, z=2.42, p=0.016; activity t test, F=0.23, t=−2.46, p=0.016).

The factor scores of individuals differed between the zoos in all but the fourth factor (sociability, K–W test H(2)=36.8, p<0.000001; positive affect, K–W H(2)=9.35, p=0.009; equitability, ANOVA F(74)=14.48, p<0.00001; anxiety, K–W H(2)=2.32, df=2, p=0.31; activity, ANOVA F(74)=40.43, p<0.00001) (Fig. 2). Post-hoc assessment of the differences between the zoos in sociability and positive affect scores were done with MWU (with a Bonferroni