**SUPPLEMENTARY METHODS**

**(a) Measuring social connectedness.** To measure variation in female affiliative behavior, we constructed two individual-based, age-specific indices of female social connectedness: one to adult females (SCI-F) and the other to adult males (SCI-M). These indices compared the strength of each female’s social connectedness relative to all other adult females alive in the population in the same year [1]. We based these estimates of affiliative social connectedness on grooming interactions, which both maintain and strengthen social bonds in baboons and other primates [2, 3].

Given the data available for our population, estimates of social connectedness could have been derived from either: i) hourly rates of grooming for each female calculated from focal animal sampling data, or ii) relative grooming frequencies between females calculated from *ad libitum* behavioral data [4, 5]*.* Specifically, thefocal animal sampling data were collected on all adult females and juveniles during 4 hours of each 5-hour visit to the social group. Focal animals were observed for 10-minute intervals in the order dictated by a randomized list. During these intervals, observers recorded all occurrences of grooming that involved the focal animal, allowing hourly rates of grooming to be calculated. In contrast, the *ad lib* data were collected throughout the entire 5-hour visit to the group. During *ad lib* sampling, observers recorded all observed grooming interactions between any group members, allowing relative frequencies of grooming to be calculated.

We ultimately chose to use the *ad lib* data in our analysis because they provided a substantially denser data set than the focal data for analyzing female social interactions. In particular, grooming events between adult baboons were too infrequent to be adequately captured during focal animal sampling data. Specifically, the hourly rate of grooming between adult females calculated from focal sampling data was 0.37 events per hour (mean of 6.3 hours of focal animal sampling data per female per year; range = 0.33 to 20.50 hours). Hence, we only observed ~2.3 grooming interactions per female-year during focal animal sampling. In contrast, the *ad lib* data set included approximately 70,000 total grooming interactions between adult females, with an average of 55.8 interactions per female per year. Hence, the *ad lib* data contained over 20 times the number of interactions found in the focal sampling data, making *ad lib* data the preferred data source in terms of data richness. Notably, if we had based our analysis on focal sampling data, in some years a subset of females would have appeared to never participate in grooming interactions, simply as a consequence of data sparsity. In comparison, we always captured grooming interactions for these females in our *ad lib* data.

However, a possible concern with using *ad lib* data is that such data might provide biased estimates of relative grooming frequencies if observers inadvertently spent more time watching some animals than others [4]. For example, animals that regularly occupied more central positions in the group might have been observed more often than animals that regularly occupied peripheral positions. Our sampling protocol was designed to avoid this type of bias as observers collected the great majority of *ad lib* data during simultaneous focal sample collection (with the order of subjects predetermined by a random rotation list). Thus, observers constantly moved to new locations in the group, rarely spending more than 10 minutes in any given location. This protocol led to roughly even behavioral sampling across individuals. Moreover, baboon groups in Amboseli are not highly centralized, and the members of a given group (20 to 100 animals) are often scattered across a relatively wide area (up to 1km; Alberts unpublished data). Thus, animals (particularly adult females) are rarely consistently “central” because there is often no single “central” location within a group. Finally, rates of grooming from focal sampling were significantly correlated with rates of grooming derived from the *ad lib* data (Spearman’s rho = 0.293, p = 5.55 x 10-16, restricting the data set to female-years with at least 20 focal animal samples per female). Notably, the relatively low correlation coefficient in this relationship can be explained almost entirely by the sparse level of sampling in the focal data: re-sampling the *ad lib* data to the same level of sparsity (~2.8 interactions per female per year for female-years with at least 20 focal animal samples) produces a mean rho of a similar magnitude even between two subsampled *ad lib*-based data sets (mean rho = 0.4 across 1000 iterations; range = 0.26 – 0.50).

*Calculation of affiliative social connectedness from* ad lib *data*. The *ad lib* grooming data were used to calculate, for each female, in each year of her life, (i) the observed daily frequency with which each female was observed grooming other adult females and adult males, and (ii) a parallel calculation for the frequency with which she was observed to receive grooming from adult females and males. One additional challenge with using *ad lib* behavioral data to estimate relative grooming frequencies was that the number of grooming observations per animal (and hence the apparent rate of grooming) was higher in smaller social groups than in larger groups. This difference arises as a simple artifact of having a constant number of observers regardless of group size. To correct for this artifact, we therefore regressed daily rates of recorded grooming (given or received) for all adult females alive in the population during the same year against a measure of observer intensity (Figure S1). Our measure of observer intensity came from focal animal sampling of adult females, which directly reflects observer effort on the collection of *ad lib* grooming interactions [4]. Observer effort was calculated as the number of point samples (there are 10 point samples per focal animal sample) per adult female per day [5].

Each annual value of SCI-F for each female was calculated as the mean value of her residuals from two regressions: (i) the regression between observer intensity and daily rates of grooming given to adult females for all females alive in the population in the same year (e.g. Figure S1), and (ii) the regression between observer intensity and daily rates of grooming received from adult females for all adult females alive in the population during the same year. Likewise, each annual value of SCI-M for each female was calculated as the mean value of her residuals from two regressions: (i) the regression between observer intensity and daily rates of grooming given to adult males for all adult females alive in the population during the same year, and (ii) the regression between observer intensity and daily rates of grooming received from adult males for all adult females alive in the population during the same year. Positive SCI values represent females with relatively high rates of grooming for the population in that year; negative values represent females with relatively low rates of grooming for that year.

**(b) Imputing missing data via multiple imputations and predictive mean matching.** The data set used in the main model included missing values for one or more predictor variables in 592 of the 1,968 female-years, causing these years to be excluded from the complete case model. 90% of the missing values (533 female-years) were values for SCI-F and SCI-M, which arose from our decision to exclude 19 group-years from our long-term data during which social groups were experiencing group fissions. During group fission, the composition of the group may change on a daily basis, making it difficult to calculate group size and observer intensity. The remaining 10% of female years with missing data (N = 59) were lacking information for the female’s mother and maternal kin because the female was born before the onset of behavioral and genetic monitoring.

We imputed missing data in our final data set using multiple imputations [6] and weighted predictive mean matching as implemented via the *aregImpute* function in the *rms* package in the R statistical software program [7, 8]. Specifically, we began the imputation process by replacing each piece of missing data with a random value drawn from the observed data for that value (bootstrapping without replacement). We then fit a flexible additive model, using restricted cubic splines for continuous variables, to estimate the target variable (i.e., social connectedness, group size etc.) in all of the original observations. We then imputed each missing value of the target variable via ‘predictive mean matching’, which uses a weighted probability draw to select among a set of predicted values calculated for the missing data point’s closest ‘neighbors’. Close neighbors were those with similar values for other predictor variables (i.e., similar rank, age, number of maternal relatives etc.). The number of neighbors involved depended on the variance of the type of data being predicted. We repeated this process for all missing pieces of data in a given data set to create a single complete imputation, and we repeated the complete imputation process 50 times to create 50 imputed data sets. We fit the main Cox proportional hazards model 50 times using these data sets. The values in the main model represent model parameters averaged over the 50 model fits.

**References**

1. Runcie D.E., Wiedmann M., Archie E.A., Altmann J., Wray G.A., Alberts S.C., Tung J. 2013 Social environment influences the relationship between genotype and gene expression in wild baboons. *Philos T Roy Soc B* **368**, 20120345.

2. Silk J.B. 2012 The adaptive value of sociality. In *Evolution of Primate Societies* (eds. Mitani J.C., Call J., Kappeler P., Palombit R., Silk J.B.), pp. 552-564. Chicago, University of Chicago Press.

3. Cords M. 2012 The behavior, ecology, and social evolution of cercopithecine monkeys. In *Evolution of Primate Societies* (eds. Mitani J.C., Call J., Kappeler P., Palombit R., Silk J.B.), pp. 91-112. Chicago, University of Chicago Press.

4. Altmann J. 1974 Observational study of behavior: sampling methods. *Behaviour* **49**, 227-267.

5. Alberts S.C., Altmann J. 2011 Monitoring guide for the Amboseli Baboon Research Project: http://amboselibaboons.nd.edu/downloads/.

6. Little R.J.A., Rubin D.B. 2002 *Statistical analysis with missing data*. 2nd ed. New York, Wiley & Sons.

7. Harrell F. 2014 rms: Regression Modeling Strategies. R package version 4.1-1. http://CRAN.R-project.org/package=rms.

8. Therneau. 2014 A Package for Survival Analysis in S. R package version 2.37-7, http://CRAN.R-project.org/package=survival.

**SUPPLEMENTARY TABLES AND FIGURES**

**Table S1.** Tests of the proportional hazards assumption.

|  |  |  |  |
| --- | --- | --- | --- |
| **fixed effects** | **ρ** | **Χ2** | **p** |
| Dominance rank | 0.07 | 0.77 | 0.38 |
| Group size | -0.05 | 0.32 | 0.57 |
| Mom alive or dead | 0.00 | 0.00 | 0.98 |
| Number of adult daughters | -0.06 | 0.55 | 0.46 |
| SCI-F | 0.06 | 0.77 | 0.38 |
| SCI-M | 0.01 | 0.03 | 0.86 |
| *Global* |  | 2.03 | 0.92 |

ρ = the correlation coefficient between survival time and the residuals

**Table S2**. Results of a Cox proportional hazards model predicting female survival using female records with no missing data (i.e. the ‘complete case’ model; N = 194 females; 1,376 female-years of data; 124 censored records).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **factors** | **b** | **s.e.** | **z** | **hazard ratio** | **p** | **direction** |
| Dominance rank | -0.01 | 0.04 | -0.53 | 0.99 | 0.593 | *-* |
| Group size | -0.01 | 0.02 | -0.99 | 0.99 | 0.322 | *-* |
| Mother alive vs. dead | -0.10 | 0.32 | -0.31 | 0.91 | 0.760 | *-* |
| Has adult daughters | 0.53 | 0.405 | 1.32 | 1.71 | 0.187 | *-* |
| **SCI-F\*** | **-0.41** | **0.16** | **-2.54** | **0.66** | **0.011** | **↑ survival with ↑ SCI-F** |
| **SCI-M\*** | **-0.71** | **0.15** | **-4.86** | **0.49** | **<0.001** | **↑ survival with ↑ SCI-M** |

\* To facilitate interpretation, the coefficient (b) and hazard ratio for SCI-F and SCI-M reflect the predicted change in the hazard function for the interquartile range of SCI-F and SCI-M (i.e. between the 25th and 75th percentiles), not a change of 1 unit for these variables.