Electronic Supplementary Material

The cooperative sex: Sexual interactions among female bonobos are linked to increases in oxytocin, proximity and coalitions.

Liza R. Moscovice, Martin Surbeck, Barbara Fruth, Gottfried Hohmann, Adrian V. Jaeggi and Tobias Deschner.

1. Additional details of methods

1.1 Behavioural data collection: Assigning contexts to sexual events

Each sexual event was classified into one or more contexts, which were not mutually exclusive (see Table S1). Feeding contexts were assigned to any sexual events that occurred in patches of confirmed bonobo food species, and where at least one of the participants was foraging immediately prior to the sexual interaction. The majority of sexual events in feeding contexts occurred in the canopy of a feeding tree, but a subset occurred terrestrially. A subset of feeding events (n = 132, 16.7%) also occurred in food sharing contexts, during which one or more individuals shared access to a potentially monopolizeable food source (most commonly meat from duikers (*Cephalophu*s sp) or fruit of *Treculia africana* or *Anonidium mannii*). An aggressive context was assigned when sexual events occurred either during, or within one minute following the end of, an aggressive interaction that was witnessed by both sexual partners, whether or not they were directly involved. Changes in party composition were assigned to all sexual interactions that occurred within five minutes following a fusion event, during which individuals not seen in the previous party scan joined the group. The remaining sexual events occurred during play, while traveling or resting, or involved 'furtive sex', in which individuals left their party to have a sexual interaction and then returned to their party afterwards (see Table S1). We included all sexual interactions that occurred in feeding contexts in the analyses, but we controlled in models for any potential influences of other events that occurred in addition to feeding (e.g. food sharing, aggression, or fusions) on the response.

1.2 Additional assay validations

When there was sufficient volume, we divided each urine sample directly after collection into two 500 µl aliquots and stored one aliquot in a cryo vial containing 100 µl of 0.5 N phosphoric acid, to test whether preserving samples in acid helped to avoid degradation, as suggested by Crockford and colleagues (2013). We found no differences in measured uOT concentration between samples that were aliquoted either with or without phosphoric acid and otherwise processed identically (Wilcoxon, T+ = 264, n = 29, p = 0.33). We thus present results from the larger dataset of samples collected without phosphoric acid.

In addition to tests of parallelism and accuracy, we also established an immunogram via fractionation with High Performance Liquid Chromatography (HPLC) to compare immunoreactivity in bonobo urine and a synthetic OT standard. We first extracted a pooled bonobo urine sample and an oxytocin standard sample that was prepared by diluting the standard provided by the commercially available enzyme immunoassay kit (Assay Designs; ADI-901-153) to a 500 pg ml^{-1} concentration. We extracted both samples following the protocol detailed in the main text. We then ran 100 µl of the extracts from the pooled bonobo sample and OT standard over a Waters Alliance 2695 HPLC equipped with a Gemini C18 column (Phenomenex, Torrance, CA, USA) with a flow rate of 0.2 ml/min using a gradient of eluent A (5% acetonitrile with 0.25% formic acid) and eluent B (95% acetonitrile with 0.25% formic acid). We collected 15 fractions of 600 µl every 3 minutes with a Waters Fraction Collector 3 (Waters, Milford, MA, USA). Fractions were lyophilized overnight and kept frozen at -20°C until further analysis via enzyme immunoassay as described in the main text.

We determined the immunoreactivity (IR) in each fraction obtained from the HPLC based on the EIA plate measurements (see Fig S1). We first determined in which of the OT standard fractions IR was found. We then determined for the bonobo pooled sample the amount of IR that occurred in the same fractions in which IR was also present for the OT standard. The sum of IR in these co-occurring fractions was labelled ‘explained IR’, referring to IR that overlapped with the standard and therefore most probably stems

from OT or one of its metabolites. We then determined the amount of IR found in fractions of the bonobo pooled sample where no IR was found in the OT standard. The sum of IR in these fractions was labelled ‘unexplained IR’. We then calculated the percentage of unexplained IR (%) as: ((total IR – sum of IR in OT Standard) / total IR) * 100. The percentage of explained IR (%) was then calculated as: 100- percentage of unexplained IR. We found that 67.5% of the IR detected in the pooled bonobo urine sample occurred in the same fractions in which IR was detected in the synthetic OT sample (‘explained IR’, see Fig. S1). The IR validation confirms that the assay captures the majority of OT and its metabolites and is suitable for measuring OT secretion in bonobo urine

1.3 Additional details of statistical analyses

To determine whether there were differences in uOT concentrations more generally following any sexual events compared with baseline samples, we fit a gaussian linear mixed model (LMM) to test whether variation in uOT concentrations was influenced by the type of sample (post-sex vs baseline). We controlled for additional predictors that may influence uOT concentrations, including: 1) Female reproductive state (cycling, pregnant or lactating), which is associated with variation in uOT in other species (Borrow & Cameron, 2012); 2) Sexual swelling stage, since changes in sexual swelling stage may correspond with changes in oestrogens, which have excitatory effects on OT synthesis and OT receptor activation (Lim & Young, 2006); 3) Female dominance rank, since higher-ranking females are likely to have more control over their sexual interactions, and sexual interactions may have different value for females of different ranks (Clay & Zuberbühler, 2012) and 4) Hour of sample collection, due to evidence that hormone release may be influenced by circadian rhythms (reviewed in: Kim, Jeong, & Hong, 2015). We included subject and date of sample collection as random effects and included random slopes of test predictors within subject.

To categorize female reproductive state, pregnancy was typically detected within the first trimester using pregnancy test kits (HCG-S pregnancy test strips, Verify

Diagnostics), or was estimated following birth, using an average gestation length of 224 days (Drews et al., 2011). Lactation was defined for four years post-partum, based on evidence that most young are weaned by that age (de Lathouwers & van Elsacker, 2006; Oelze et al. 2016). To categorize sexual swelling stage, swellings were visually assessed daily for all mature females and scored on a 1- 4 scale of increasing tumescence, based on each female's relative swelling size and turgidity, following Douglas and colleagues (2016). To determine dominance rank, we calculated rankings based on all aggressive interactions during the study period using the ADAGIO algorithm, which is appropriate for assessing dominance in social groups without strong linear dominance structures (Douglas, Ngoma & Hohmann, 2017).

Citations

Borrow, AP, Cameron, NM (2012). The role of oxytocin in mating and pregnancy. *Hormones and Behavior*, *61*(3), 266–276.

Clay Z, Zuberbühler K (2012) Communication during sex among female bonobos: Effects of dominance, solicitation and audience. *Sci Rep* 2:1–8.

Crockford, C, Wittig, R, Langergraber, K, Ziegler, T, Zuberbühler, K, Deschner, T (2013). Urinary oxytocin and social bonding in related and unrelated wild chimpanzees. *Proceedings of the Royal Society B* 280: 20122765.

Douglas PH, Hohmann G, Murtagh R, Thiessen-Bock R, Deschner T (2016) Mixed messages: Wild female bonobos show high variability in the timing of ovulation in relation to sexual swelling patterns. *BMC Evol Biol* 16(1):140.

Douglas PH, Ngomo AN, Hohmann G (2017) A novel approach for dominance assessment in gregarious species: ADAGIO. *Anim Behav* 123:21–32.

Drews B, Harmann LM, Beehler LL, Bell B, Drews RF, Hildebrandt, TB (2011) Ultrasonographic monitoring of fetal development in unrestrained bonobos (*Pan paniscus*) at the Milwaukee County Zoo. *Zoo Biol* 30(3):241–253.

Kim TW, Jeong JH, Hong SC (2015) The impact of sleep and circadian disturbance on hormones and metabolism. *Int J Endocrinol* 2015:1–9.

de Lathouwers M, van Elsacker L (2006) Comparing infant and juvenile behaviour in

bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*): A preliminary study. *Primates* 47(4):287–293.

Lim MM, Young LJ (2006) Neuropeptidergic regulation of affiliative behaviour and social bonding in animals. *Horm Behav* 50(4):506–517.

Oelze VM, Douglas PH, Stephens CR, Surbeck M, Behringer V, Richards, MP *et al.* (2016) The steady state great ape? Long term isotopic records reveal the effects of season, social rank and reproductive status on bonobo feeding behaviour. *PLoS One* 11(9):e0162091.

2. Supplementary Figures and Tables

Fig S1. Comparison of immunoreactivity in a pooled sample of bonobo urine and a synthetic OT standard provided by the commercially available EIA kit (Assay Designs no. ADI-901-153). Samples were run through High Performance Liquid Chromatography (HPLC) and then measured via enzyme immunoassay.

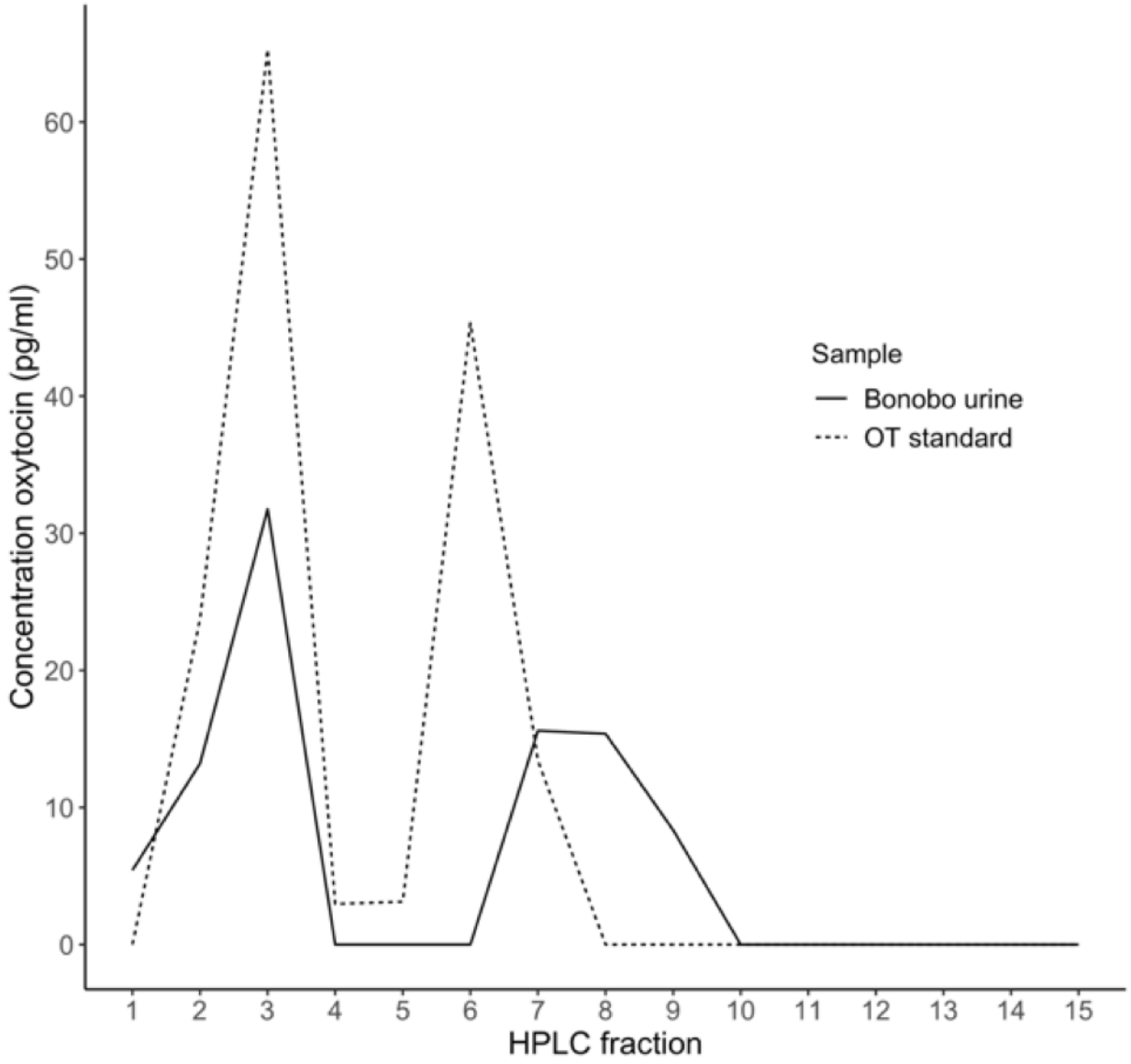


Fig S2. Comparison of the number of dyads of each sex combination that were observed having sexual interactions at least once during the study period, relative to expected based on the proportions of each dyadic sex combination in the community (considering all mature, unrelated dyads).

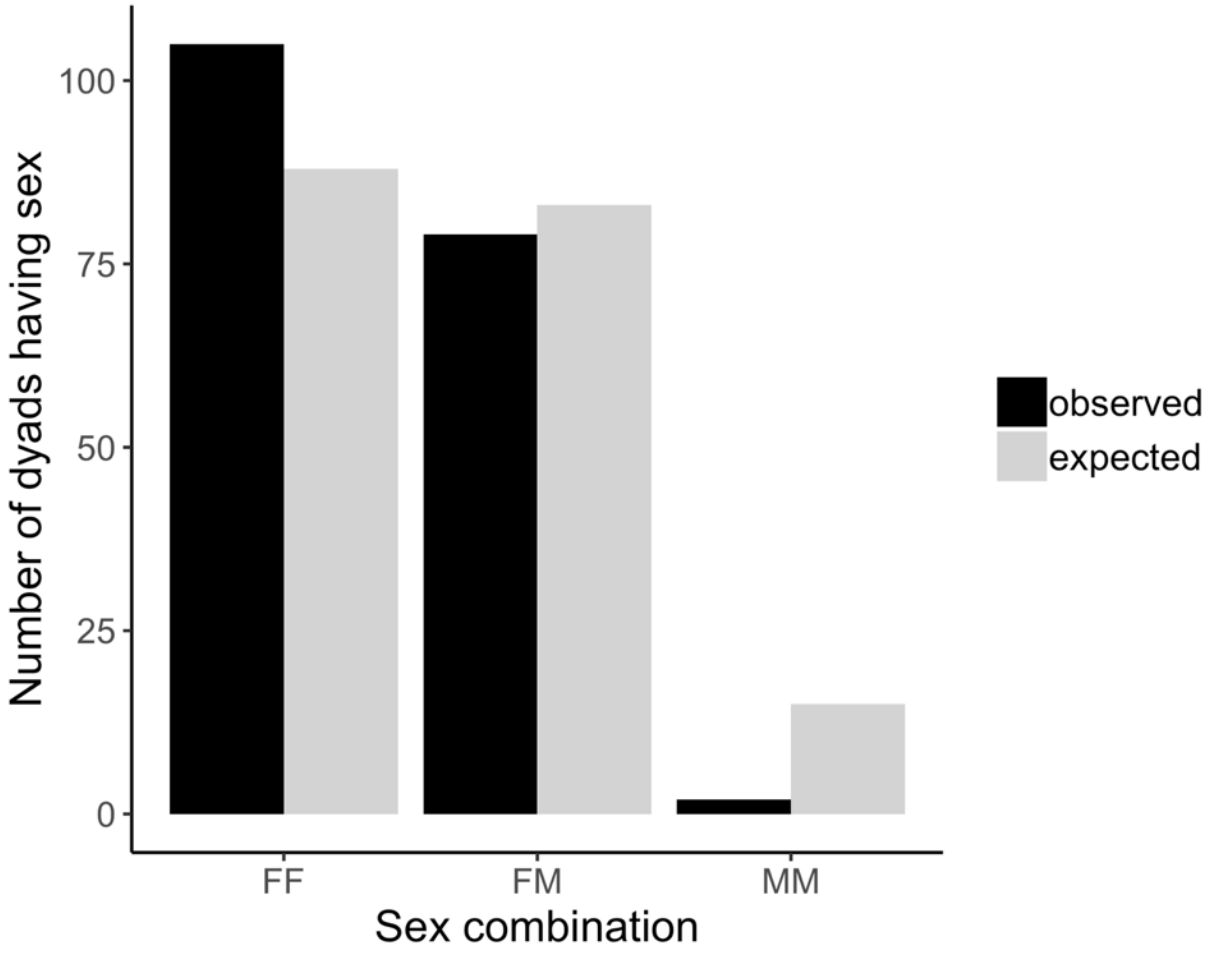
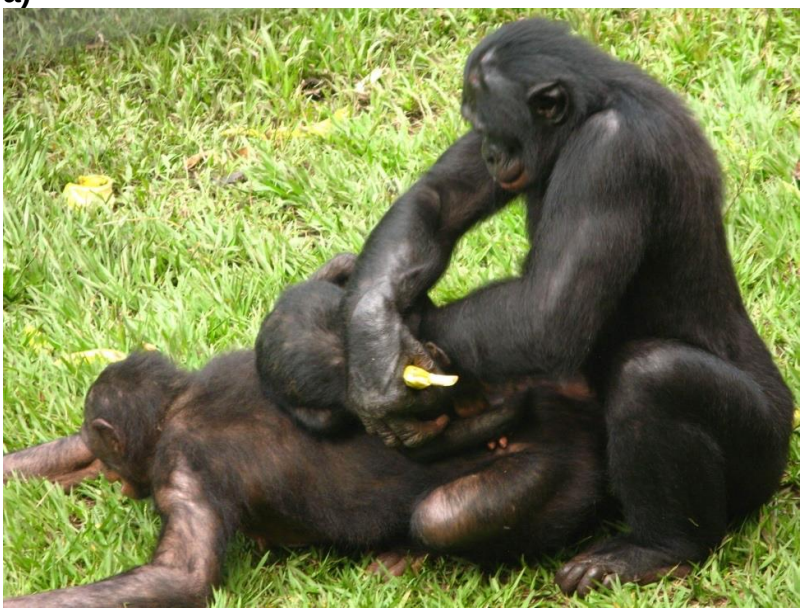


Fig S3. Comparison of typical sexual positions for copulations (a) and GG-rubbing (b) in bonobos. GG-rubbing requires face-to-face contact and may facilitate clitoral stimulation. Photos from Lola ya Bonobo Sanctuary, courtesy of Zanna Clay.

a)



b)

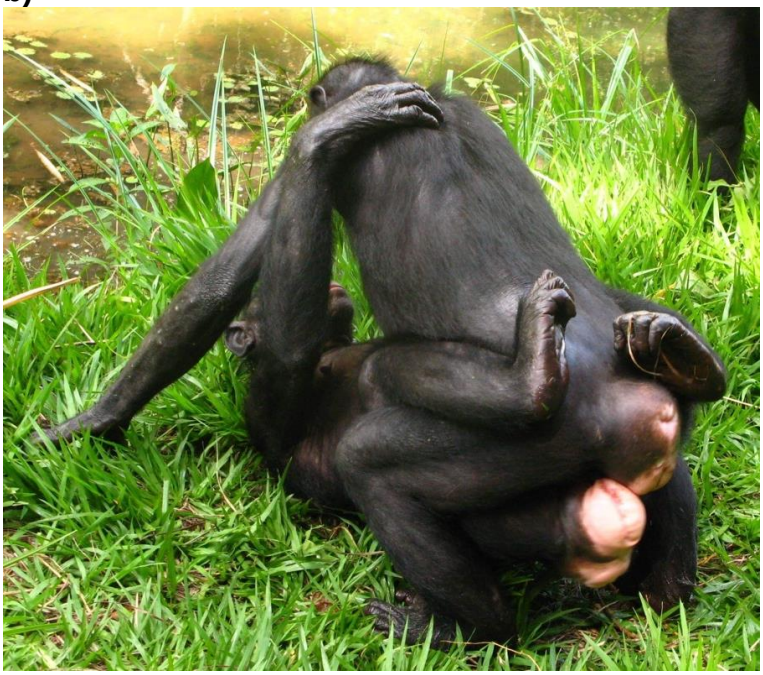


Table S1. Percent of observed copulation and GG-rubbing events that occurred in each context

Context	GG-rubbing (%) (n= 558 events)	Copulations (%) (n= 232 events)
Feeding overall;	91.9	84.1
(Feeding + food sharing)	(23.6)	(4.7)
(Feeding + change in party composition)	(7.4)	(5.2)
(Feeding + aggression)	(5.1)	(1.3)
Change in party composition	1.3	< 1
Aggression	2.7	1.3
Resting	2.9	4.3
Traveling	< 1	3.0
Play	1.1	0.0
Furtive	0.0	6.5

Table S2. Contributions of each subject (arranged by sex and age class) to the test of the sex preference hypothesis. Values indicate each subject's total number of sexual interactions in feeding contexts, as well as their proportion of sexual interactions that occurred with same-sex partners.

			Sexual interactions in feeding contexts	
Subject	Sex	Age class	Total	Proportion same sex
Gw	F	adult	93	0.88
Ir	F	adult	70	0.77
Lu	F	adult	13	1.00
Ma	F	adult	80	0.96
Nn	F	adult	141	0.84
Ol	F	adult	93	0.95
Pa	F	adult	57	0.88
Po	F	adult	78	0.77
Ri	F	adult	75	0.88
Su	F	adult	131	0.67
Um	F	adult	75	0.85
Wi	F	adult	72	0.78
Zo	F	adult	114	0.82
Id	F	subadult	1	1.00
Pg	F	subadult	8	1.00
Rt	F	subadult	10	1.00
Dj	F	subadult immigrant	96	0.90
Ng	F	subadult immigrant	7	1.00
Ve	F	subadult immigrant	7	0.71
Ap	M	adult	26	0.00
Be	M	adult	37	0.05
Ca	M	adult	44	0.05
Em	M	adult	26	0.04

Ja	M	adult	23	0.00
Ro	M	adult	26	0.00
Ze	M	adult	14	0.07
Hu	M	subadult	5	0.00

Table S3. Results of a GLMM predicting the effects of dyadic sex combination (relative to FM) and Composite Relationship Index (dyaCRI) on the likelihood of staying in close proximity after sex for the subset of partners who were not in close proximity before sex. Statistically significant predictors ($p < 0.05$) are indicated in bold.

Term	est ± SE	chi-sq	df	p value
(Intercept)	- 1.15 ± 0.25			
Test predictors:				
Dyadic sex combination (FF)	**0.71 ± 0.31**	**6.25**	**1**	**0.04**
Dyadic CRI	-0.32 ± 0.26	1.83	1	0.18
Control predictors:				
Context: Feed + food sharing*	**2.02 ± 0.67**	**9.09**	**3**	**0.02**
Context: Feed + aggression*	**-0.28 ± 0.72**			
Context: Feed + change in party*	**0.96 ± 0.73**			

* Chi-square, df and p values refer to comparison with reference category: Feeding without additional contexts

Table S4. Results of a LMM predicting the effects of having any sexual interactions (copulations or GG-rubbing) on female uOT concentrations (log transformed). Trends toward significance ($0.05 \leq p < 0.09$) are indicated in bold.

Term	est ± SE	chi-sq	df	p value
(Intercept)	5.35 ± 0.22			
Test Predictor:				
Sample Type (post-sex)*	**0.31 ± 0.15**	**3.7**	**1**	**0.05**
Control predictors:				
Reproductive state (lactating)†	-0.04 ± 0.16	2.7	2	0.26
Reproductive state (pregnant)†	0.64 ± 0.35			
Swelling score 1‡	**0.61 ± 0.26**	**6.7**	**3**	**0.08**
Swelling score 2‡	**0.20 ± 0.20**			
Swelling score 3‡	**0.03 ± 0.19**			
Dominance rank	0.04 ± 0.08	0.17	1	0.68
Hour of sample collection	-0.06 ± 0.08	0.57	1	0.45

*,†,‡ Chi-square, df and p values refer to comparison with reference categories: *Baseline samples; †Cycling, ‡Swelling score 4 (most tumescent).

Table S5. Planned comparisons of changes in uOT between baseline, post-copulation and post-GG rubbing samples from the post-sex physiology model. Results are averaged over the levels of number of partners (1 vs >1) and swelling stage (1- 4). Significant contrasts ($p < 0.05$) and trends toward significance ($0.05 \leq p < 0.09$) are indicated in bold.

Contrast	est ± SE	df	t value	adj p value*
GG vs copulation (baseline samples)	-0.01 ± 0.28	103.40	-0.03	0.98
GG vs copulation (post-sex samples)	**0.64** ± 0.28	**87.53**	**2.28**	**0.05**
Copulation vs baseline	-0.05 ± 0.30	48.84	-0.17	0.98
GG vs baseline	**0.60** ± 0.20	**48.07**	**2.96**	**0.02**

* Bonferroni adjustment for four comparisons

Table S6. Results of a LMM comparing changes in uOT in samples collected following feeding and GG-rubbing compared with feeding without sexual interactions. Significant predictors ($p < 0.05$) are indicated in bold.

Term	est ± SE	chi-sq	df	p value
(Intercept)	4.34 ± 0.29			
Test predictor:				
Sample Type: Feeding + GG-rubbing*	**0.95 ± 0.10**	**82.84**	**1**	**< 0.001**
Control predictor:				
Swelling score 1†	0.65 ± 0.40	1.61	3	0.234
Swelling score 2†	0.56 ± 0.42			
Swelling score 3†	0.15 ± 0.35			

*,† Chi-square, df and p values refer to comparison with reference categories:
*Feeding without GG-rubbing; †Swelling score 4 (most tumescent).