




Male and Female Nipples as a Test Case for the Assumption that Functional Features Vary Less than Nonfunctional Byproducts

Ashleigh J. Kelly¹ · Shelli L. Dubbs¹ ·
Fiona Kate Barlow¹ · Brendan P. Zietsch¹ 

Received: 19 April 2018 / Revised: 16 May 2018 / Accepted: 17 May 2018 /
Published online: 26 June 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Objectives Evolutionary researchers have sometimes taken findings of low variation in the size or shape of a biological feature to indicate that it is functional and under strong evolutionary selection, and have assumed that high variation implies weak or absent selection and therefore lack of function.

Methods To test this assumption we compared the size variation (using a mean-adjusted measure of absolute variability) of the functional human female nipple (defined as the nipple-areola complex) with that of the non-functional human male nipple.

Results We found that female nipples were significantly *more* variable than male nipples, even after controlling for body mass index, testing-room temperature, bust size in women, and chest size in men.

Conclusions Morphological variation in a feature should not be used by itself to infer whether or not the feature is functional or under selection.

Keywords Functional adaptation · Selection strength · Byproduct · Nipple · Variation

Introduction

A major goal in evolutionary biology is to understand the extent to which various biological features are products of adaptation (in response to Darwinian selection), or of

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40750-018-0096-1>) contains supplementary material, which is available to authorized users.

✉ Brendan P. Zietsch
zietsch@psy.uq.edu.au

¹ School of Psychology, The University of Queensland, Room 457, McElwain Building (24A), St Lucia, Australia

other evolutionary processes such as drift (random changes) or developmental constraint. Simplified, the aim is to distinguish functional adaptations from nonfunctional byproducts (or ‘spandrels’; Andrews et al. 2002; Gould and Lewontin 1979).

One way this aim has been pursued is by comparing within-species variability of features. This approach is based on the premise that functional structures are less variable than nonfunctional structures; it is thought that low variation in a feature within a population indicates that variation in the feature has been winnowed away by strong selection (implying function), whereas high variation indicates that the feature has been under weak or no selection (implying no function). This idea is often attributed to Fisher’s fundamental theorem of natural selection (Fisher 1930), although the theorem’s original formulation is quite different (“The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.” p. 35), and its meaning and applicability to real populations have been controversial (Frank and Slatkin 1992; Okasha 2008; Lessard 1997; Price 1972). Nevertheless, researchers have used differences in variation in the physical size of features to infer their functionality or lack thereof.

One such example concerns the female orgasm, which could have developed as a byproduct of the clearly functional male orgasm (Gould 1987; Lloyd 2005; Symons 1979; Wallen and Lloyd 2008), or could alternatively serve its own adaptive function (e.g. in mate choice; Puts et al. 2012a, b). Within this debate, the greater length variability of the human clitoris versus penis has been taken as evidence that the human female orgasm is a functionless developmental byproduct of the functional male orgasm (Wallen and Lloyd 2008). Note, however, that the authors demonstrated no connection between men’s capacity to orgasm (and thus potentially impregnate a mate) and phallus size. Instead, it was simply presumed that functional features necessarily have less variance in physical size than non-functional features. In a related example, lower variability in human penis girth versus length has been used as evidence that a selective pressure of female choice is stronger on penis girth than on its length (Apostolou 2015).

However, such inferences are only valid if the premise is consistently true: that is, functional/strongly selected features are less variable in size than nonfunctional/weakly selected features. This assumption is falsifiable, given a consensually agreed upon feature that is functional in one sex, but not the other. Human nipples are such a feature. Gould, who first coined the term ‘spandrel’ (Gould and Lewontin 1979), has cited men’s nipples as a clear example of a functionless developmental byproduct of the strong selective pressure on women to have (functional) nipples (Gould 1987). That men’s nipples are spandrels is agreed upon by those who cite lack of variability as evidence of trait functionality (e.g. Wallen and Lloyd 2008). Not only are male nipples a *good* example of a spandrel, they are to our knowledge the *only* example of an uncontested spandrel of which a functional version of the feature is carried by members of the same species. This provides a rare test case for relative size variability of functional versus nonfunctional features, since factors other than functionality (e.g. type of trait, dimensionality of trait, species in which it is observed) are constant.

Thus, in the present paper, we compare the coefficient of variation (a mean-adjusted measure of absolute variability) of men’s and women’s nipples, with and without controlling for potentially correlated body measures (bust size for women, chest size for men). If degree of between-individual variation in the size of a feature can be used

to infer the feature's functionality and selection history, then human female nipples should be less variable than male nipples.

Methods

For the purposes of brevity we refer to the entire nipple-areola complex as 'nipple(s)'. This complex was chosen because it can be reduced to a 2-dimensional area measurement (see Fig. 1), and includes a range of adaptive features (arguably more than the nipple bud itself). It contains the tubercles of Morgagni and Montgomery glands (Montgomery 1837; Rosen 2001) which secrete milk and oils (Kopans 2007; Montgomery 1837; Nicholson et al. 2009), with an "abundant lymphatic system" known as the "subareolar or Sappey plexus" (Nicholson et al. 2009, p. 511). The areola glands have a lubricating function that physically eases breastfeeding (An et al. 2010), and have been described as "miniature lacteal units combined with sebaceous units" (Smith et al. 1982). The areola secretions are an olfactory stimulant to infants (Doucet et al. 2009; Porter and Winberg 1999), and help initiate suckling and are positively related to infant growth (Doucet et al. 2012). Additional blood vessels relative to the breast increases areola temperature, diffusing the scent (Porter and Winberg 1999) and physically orienting infants (Doucet et al. 2012; Zanardo and Straface 2015). The dense skin of the nipple protects the breast against saliva and sucking (Perkins and Miller 1926; Zanardo and Straface 2015), and may be a visual target for infants (see Doucet et al. 2007). None of these features are useful in men, who do not breast feed.

Participants

Australian undergraduates (33 male, 30 female) aged between 18 and 33 ($M=20.16$, $SD=2.74$) participated in exchange for course credit. Informed consent was obtained from all individual participants included in the study. Note that participation was completely voluntary, and participants were informed of the requirements of the study

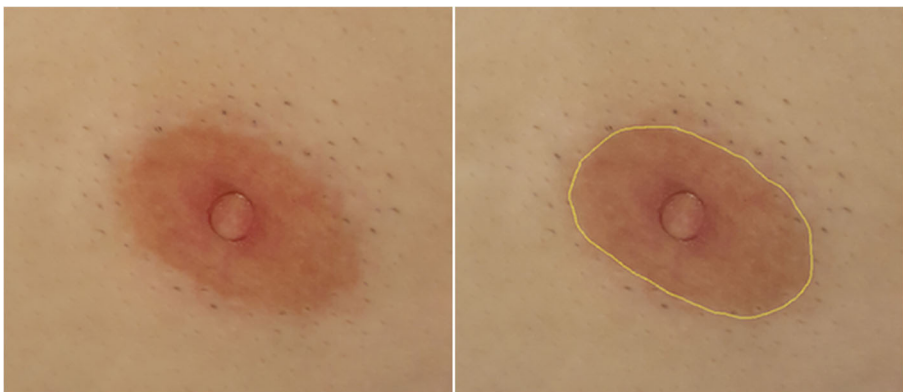


Fig. 1 Example of male nipple scan and tracing for size measurement. Image from volunteer not involved in study with written permission for publication

prior to signing up for participation. There were multiple different studies that students could choose to participate in for credit, or they could complete an alternative to *any* study participation and still receive full class credit. Participation was dependent on being 18 years or older, having natural breasts or pectorals (i.e. no breast or chest implants, or surgeries on those areas), and for females, having never experienced *any* stage of a pregnancy. The majority of the sample were Caucasian (82.5%), with the rest Asian (9.5%), African (1.6%), or Other/Non-Specified (6.4%).

Measures

Nipple area (mm²) was an average measurement taken from two images of each participant's chest. In private, the bare chest was pressed against a vertical A3 flatbed scanner for two scans. Two of the authors, Ashleigh Kelly and Shelli Dubbs, later independently measured the de-identified images on-screen by tracing the nipple structure (see Fig. 1) using the freehand tool in ImageJ (www.imagej.net). Freehand tool settings were calibrated identically for all measurements, based on a scale physically affixed to the scanner (hence present in all images).

Inter-rater reliability was computed for nipple measurements using the *irr* Package (Version 0.84). As recommended by Hallgren (2012), two-way average-measures intra-class correlations (ICC) were assessed against an absolute standard (i.e. that nipple area ratings would be similar in absolute value) using a mixed model (with our two raters treated as fixed effects). For nipple ratings, $ICC = .99$, indicating near-perfect agreement between the raters. For each participant, nipple area rating was averaged across raters, and this variable was used for all subsequent analyses. Inter-scan reliability was computed similarly (with nipple area averaged across left & right, and across raters). Again, results indicated excellent reliability between scans 1 and 2 ($ICC = .99$).

Weight in kilograms was measured standing and barefoot minus all jewellery and metal, using Tanita BC-575 Body Composition Scales.

Body fat percentage was measured as per weight. To ensure accuracy, pre-tests were conducted comparing the Tanita scales to a highly accurate ImpediMed DF50 bio-impedance medical analysis device (for support for use of both bioelectrical impedance and ImpediMed device, see Lukaski et al. 1985; Segal et al. 1988). The devices showed good relative agreement on body fat percentage ($r = .96$, $p < .001$), and absolute agreement was within 3%. Pateyjohns et al. (2006) performed similar tests using Tanita brand scales and the DF50 device, and came to similar conclusions regarding the accuracy of Tanita scales for group comparisons. As the ImediMed device required a supine position, strict test and pre-test conditions, and multiple electrodes, it was impractical to use for our purposes.

Other body measurements were taken in centimeters using a tape measure and standardized procedures (Fig. 2 shows anatomical placement). Although participants were given the option of placing the tape themselves, nearly all opted to allow the researcher to place it (excl. bust). For sensitive measurements (e.g. bust), the option to hold the tape was taken by a small percentage of participants, with the researcher requesting adjustments to placement and tension, and reading the value. All measurements were taken while clothed and wearing a non-padded bra (females only).

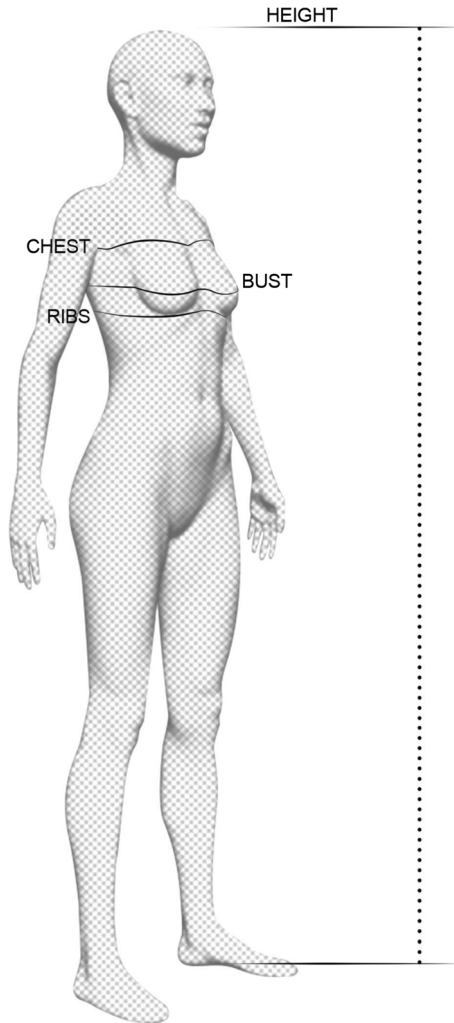


Fig. 2 Measurements taken on both genders, excluding ribs and bust (females only). All were circumferences excluding height. Image by JBamaya Design (JBamaya.com)

- *Height* was measured barefoot, pressing the back and heels against a wall.
- *Chest* circumference was taken around the chest and back, directly under the armpits and above the bust.
- *Ribs* circumference was taken for females only, just under the bust and around the back as part of a bra size measurement.
- *Bust* circumference was taken for females only, at the fullest part of the breasts and around the back, typically aligned with the nipples.

An array of other body measurements were gathered, yet most were either redundant or uncorrelated with nipple area once analyzed (hence excluded).

Room *temperature* was measured in degrees Celsius using a digital thermometer placed directly above the scanner.

Procedure

The study was advertised as an investigation into sexual dimorphism in humans, with procedures explicitly disclosed prior to participants signing up for participation in the study, and again during participation (at multiple times, to allow for withdrawal at any stage). Body measurements were taken while clothed with bulky items removed, and feet were cleaned for scale measurements. Participants also undertook a questionnaire, confirming gender and ethnicity. They were then instructed on using the scanner, taking supervised images of their hands for practice before being left in private to undertake chest scans.

Two female experimenters were present with the participant at all times except during chest scans. Anonymity, comfort, and consent were of utmost importance, with participants assigned a secret code, which they used as a reference when undertaking their scans. Withdrawal was offered multiple times, and any hint of discomfort required the experimenters to abandon the chest scans and confirm desire to continue with other segments. Ethical approval was obtained for all procedures.

Results

Preliminary Statistics

All analyses were performed in RStudio (Version 0.99.489; R Version 3.2.2). The dataset supporting this article has been uploaded as supplementary material. Descriptive statistics for women and men are displayed in Tables 1 and 2, respectively. Note that ambient temperature was not significantly correlated with nipple measurements. Nipple size variation had low and nonsignificant skew (.37 for men, .39 for women).

Main Analyses

Tables 1 and 2 reveal large mean differences between the area of female and male nipples, such that male nipples were on average 36% the size of female nipples. To compare variances of samples with different means, the variances are scaled to the sample means by calculating coefficients of variation (CVs) for each sample, expressed as percentages (Pearson 1896; Broberg 1999, 2016; as per Apostolou 2015; Lynch 2008; Wallen and Lloyd 2008). The CV is the ratio of the standard deviation divided by the mean, multiplied by 100 to yield a unitless percentage of variance. Significance (p) of the sex difference was computed via Studentized bootstrapping (a procedure that does not assume data normality) using Broberg's `rsd.test` function within the SAGx R package (see Broberg 1999, 2016, for details on testing the significance of differences in CVs). Note that the results are substantively the same if a simple F-test is used as per Apostolou 2015, and Wallen and Lloyd 2008.

In raw terms, the CV of female nipples was significantly higher than that of male nipples (42% versus 27%, $p < .001$). After accounting for bust size in women, and chest size in men, the CV for female nipples was still higher than male nipples (38% for women, versus 25% for men, $p < .001$). Results were substantively the same when

Table 1 Women's uncentered means, standard deviations, ranges and intercorrelations

Variable	Mean	SD	Range	1.	2.	3.	4.	5.	6.	7.	8.
1. Age (18 to 33)	19.4	1.5	5								
2. Nipple (mm ²)	682.1	287.8	1148.4	.32							
3. Weight (kg)	63.1	9.8	44	.20	.29						
4. Body fat (%)	25.4	7.7	30	.11	.40*	.69***					
5. Height (cm)	168.7	7.8	41	−.09	−.15	.54**	.01				
6. Chest (cm)	87.1	5.5	24.9	.14	.28	.82***	.52**	.34			
7. Ribs (cm)	76.3	9.9	47.8	.07	.25	.82***	.72***	.42*	.78***		
8. Bust (cm)	87.8	6.7	27	.24	.37*	.53**	.59***	.15	.47*	.11	
9. Temp (°C)	24.0	1.3	5.6	.37*	.26	.16	.06	−.01	−.02	−.02	.19

* $p < .05$, ** $p < .01$, *** $p < .001$. "Temp" = Temperature

accounting for bust/chest size *and* women's and men's respective body fat percentages (38% versus 25%, $p < .001$).

Discussion

Female nipples are functional whereas male nipples are nonfunctional byproducts; therefore, our finding that women's nipples are significantly more variable in size than men's nipples demonstrates that high morphological variability of a feature does not necessarily imply a lack of function or of historical selection, as has previously been assumed.

This finding is novel in that a comparison of variability had not previously been made between clearly functional versus nonfunctional byproduct versions of the same feature. However, previous influential studies have made other relevant comparisons of variability. Pomiankowski and Moller (1995) compared the variability of features that are functional in both sexes but are exaggerated as a sexual display in males (e.g. bird tails). They found that the sexually selected features showed greater variability. Houle (1992) compared variability across different kinds of traits completely, finding that the

Table 2 Men's uncentered means, standard deviations, ranges and intercorrelations

Variable	Mean	SD	Range	1.	2.	3.	4.	5.	6.
1. Age (18 to 33)	20.9	3.4	15						
2. Nipple (mm ²)	242.8	66.9	271.8	.27					
3. Weight (kg)	78	9.2	44.3	.10	.32				
4. Body fat (%)	14.8	4.1	17.7	.43*	.28	.60***			
5. Height (cm)	183	6.2	25	−.24	.15	.52**	−.23		
6. Chest (cm)	98.4	6.8	28.9	.14	.42*	.74***	.55**	.38*	
7. Temp (°C)	24.2	1.2	4.8	.04	.28	.09	.03	.02	−.07

* $p < .05$, ** $p < .01$, *** $p < .001$. "Temp" = Temperature

types of traits presumed to be under strong selection (e.g. life-history traits such as fecundity and longevity) tend to exhibit more variability than the types of traits presumed to be under weak selection (e.g. morphological traits such as whole-body or body-part sizes). These studies suggest a complex relationship between historical selection pressure and variation in a trait, but they have not prevented researchers from taking high trait variation as indicative of lack of functionality/selection (e.g. Wallen and Lloyd 2008) – perhaps because the comparisons in the previous studies do not directly pertain to the byproduct/functional adaptation distinction. Our study, in contrast, pertains directly to this distinction. Indeed, male nipples are held as a prototype of an evolutionary byproduct (e.g. the major proponents of the byproduct explanation of female orgasm explicitly draw the analogy with male nipples; Gould 1987; Lloyd 2005; Symons 1979; Wallen and Lloyd 2008).

Given the status of women's and men's nipples as a prototypical product-byproduct, their further study could better illuminate the relation of function and selection to morphological variation. For example, our study used a 2-D measure of nipples in a relatively small sample; a whole-body, 3-D scan in a large sample could enable fine-grained measurement of variance in size and shape of different parts of the nipple and of the breasts, and if the sample was also genetically informative (e.g. comprising twins) then comparisons of genetic variance could be made as well as those of phenotypic variance.

In sum, our findings serve two main purposes. First, they discredit previous studies that have taken relatively high size variability of a feature to indicate lack of functionality. Returning to the example mentioned in the introduction, Wallen and Lloyd (2008) claimed that the greater length variation of clitorises than of penises meant that the female orgasm is a nonfunctional byproduct of male orgasm. This evidence can now be disregarded, since their own chosen analogy (male and female nipples) shows the opposite effect (the byproduct exhibits less variation in size than the adaptation). Second, our findings serve to prevent others from designing studies based on incorrect assumptions about the relationship of variability with functionality or selection. The salience and reliability of the nipple exemplar should help our findings to serve this purpose broadly.

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of Interest The authors declare that they have no conflict of interest.

References

- An, H. Y., Kim, K. S., Yu, I. K., Kim, K. W., & Kim, H. H. (2010). The nipple-areolar complex: A pictorial review of common and uncommon conditions. *Journal of Ultrasound in Medicine*, 29(1), 949–962.
- Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism—how to carry out an exaptationist program. *Behavioral and Brain Sciences*, 25(4; Discussion 504–453), 489–504.

- Apostolou, M. (2015). Female choice and the evolution of penis size. *Archives of Sexual Behavior*, *44*(7), 1749–1750.
- Broberg, P. (1999). Estimation of relative SD. *Drug Development and Industrial Pharmacy*, *25*(1), 37–43.
- Broberg, P. (2016). *SAGx: Statistical Analysis of the GeneChip*. R package version 1.44.0: http://home.swipnet.se/pibroberg/expression_hemsida1.html. Retrieved from <https://bioconductor.org/biocLite.R>
- Doucet, S., Soussignan, R., Sagot, P., & Schaal, B. (2007). The “smellscape” of mother’s breast: Effects of odor masking and selective unmasking on neonatal arousal, oral, and visual responses. *Developmental Psychobiology*, *49*, 129–138. <https://doi.org/10.1002/dev.20210>.
- Doucet, S., Soussignan, R., Sagot, P., & Schaal, B. (2009). The secretion of areolar (Montgomery’s) glands from lactating women elicits selective, unconditional responses in neonates. *PLoS One*, *4*(10), e7579. <https://doi.org/10.1371/journal.pone.0007579>.
- Doucet, S., Soussignan, R., Sagot, P., & Schaal, B. (2012). An overlooked aspect of the human breast: Areolar glands in relation with breastfeeding pattern, neonatal weight gain, and the dynamics of lactation. *Early Human Development*, *88*(2), 119–128. <https://doi.org/10.1016/j.earlhumdev.2011.07.020>.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Frank, S. A., & Slatkin, M. (1992). Fisher’s fundamental theorem of natural selection. *Trends in Ecology & Evolution*, *7*(3), 92–95.
- Gould, S. J. (1987). Freudian slip. *Natural History*, *96*, 14–21.
- Gould, S. J., & Lewontin, R. C. (1979). The spandrels of san Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, *205*(1161), 581–598.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorial in Quantitative Methods for Psychology*, *8*(1), 23–24.
- Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics*, *130*(1), 195–204.
- Kopans, D. (2007). Breast anatomy and basic histology, physiology, and pathology. In D. Kopans (Ed.), *Breast Imaging* (3rd ed., pp. 7–43). Philadelphia: Lippincott Williams & Wilkins.
- Lessard, S. (1997). Fisher’s fundamental theorem of natural selection revisited. *Theoretical Population Biology*, *52*(2), 119–136.
- Lloyd, E. A. (2005). *The case of the female orgasm: Bias in the science of evolution*. Cambridge: Harvard University Press.
- Lukaski, H. C., Johnson, P. E., Bolonchuk, W. W., & Lykken, G. I. (1985). Assessment of fat-free mass using bioelectrical impedance measurements of the human body. *American Journal of Clinical Nutrition*, *41*, 810–817.
- Lynch, V. J. (2008). Clitoral and penile size variability are not significantly different: Lack of evidence for the byproduct theory of the female orgasm. *Evolution and Development*, *10*(4), 396–397. <https://doi.org/10.1007/s10539-007-9103-9>.
- Montgomery, W. F. (1837). *An exposition of the signs and symptoms of pregnancy, the period of human gestation, and the signs of delivery*. London: Sherwood, Gilbert and Piper.
- Nicholson, B. T., Harvey, J. A., & Cohen, M. A. (2009). Nipple-areolar complex: Normal anatomy and benign and malignant processes. *RadioGraphics*, *29*, 509–523.
- Okasha, S. (2008). Fisher’s fundamental theorem of natural selection—A philosophical analysis. *The British Journal for the Philosophy of Science*, *59*(3), 319–351.
- Pateyjohns, I. R., Brinkworth, G. D., Buckley, J. D., Noakes, M., & Clifton, P. M. (2006). Comparison of three bioelectrical impedance methods with DXA in overweight and obese men. *Obesity*, *14*(11), 2064–2070.
- Pearson, K. (1896). Mathematical contributions to the theory of evolution: III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London, Series A*, *187*, 253–318.
- Perkins, O. M., & Miller, A. M. (1926). Sebaceous glands in the human nipple. *American Journal of Obstetrics*, *11*, 789–794.
- Pomiankowski, A., & Moller, A. P. (1995). A resolution of the lek paradox. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *260*(1357), 21–29.
- Porter, R. H., & Winberg, J. (1999). Unique salience of maternal breast odors for newborn infants. *Neuroscience & Biobehavioral Reviews*, *23*, 439–449.
- Price, G. R. (1972). Fisher’s ‘fundamental theorem’ made clear. *Annals of Human Genetics*, *36*(2), 129–140.
- Puts, D. A., Dawood, K., & Welling, L. L. M. (2012a). Why women have orgasms: An evolutionary analysis. *Archives of Sexual Behavior*, *41*, 1127–1143.
- Puts, D. A., Welling, L. L. M., Burriss, R. P., & Dawood, K. (2012b). Men’s masculinity and attractiveness predict their female partners’ reported orgasm frequency and timing. *Evolution and Human Behavior*, *33*(1), 1–9. <https://doi.org/10.1016/j.evolhumbehav.2011.03.003>.
- Rosen, P. P. (2001). *Rosen’s breast pathology* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

- Segal, K., van Loan, M., Fitzgerald, P. I., Hodgdon, J., & Van Itallie, T. (1988). Lean body mass estimation by bioelectrical impedance analysis: A four-site cross-validation study. *American Journal of Clinical Nutrition*, *47*, 7–14.
- Smith, D. M., Peters, T. G., & Donegan, W. L. (1982). Montgomery's areolar tubercle. A light microscopic study. *Archives of Pathology & Laboratory Medicine*, *106*, 60–63.
- Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.
- Wallen, K., & Lloyd, E. A. (2008). Clitoral variability compared with penile variability supports nonadaptation of female orgasm. *Evolution and Development*, *10*(1), 1–2.
- Zanardo, V., & Straface, G. (2015). The higher temperature in the areola supports the natural progression of the birth to breastfeeding continuum. *PLoS One*, *10*(3), e0118774. <https://doi.org/10.1371/journal.pone.0118774>.