Behavioral Ecology and Sociobiology (2020) 74:106
https://doi.org/10.1007/s00265-020-02889-y

ORIGINAL ARTICLE

The development, evaluation, and illustration of a timeline procedure for testing the role of sperm competition in the evolution of sexual traits using paternity data

R. Robin Baker[1,2] · Todd K. Shackelford[3]

Received: 10 July 2020 / Revised: 30 July 2020 / Accepted: 3 August 2020

Abstract

The empirical study of the role of sperm competition in the evolution of sexual traits has historically been problematic through the inability either to measure sperm competition levels directly in the present or to reconstruct changes in the evolutionary past. Here, we develop and test a procedure based on paternity data that potentially permits both. For our pilot study, we use the rate of change of the seminal protein gene *SEMG2* for catarrhine primates published by Dorus et al. (Nat Genet 36:1326–1329, 2004). From their data, Dorus et al. proposed a two-part hypothesis: (1) sperm competition plays a role in the evolution of the *SEMG2* gene and (2) higher levels of sperm competition generate more positive selection for change in *SEMG2* than lower levels. Dorus et al. were limited, however, by being able to use only proxy measures of sperm competition and only seven "recent" segments of catarrhine primate phylogeny. Here, we develop a "timeline procedure" that permits the Dorus hypothesis to be tested using data from across the whole of catarrhine phylogeny. Our analysis supports part (1) of the Dorus hypothesis but questions part (2), suggesting instead that changes in level of sperm competition have a more powerful influence on the rate of evolution of traits than the level of sperm competition itself. We conclude that the timeline procedure developed here could be a valuable investigative tool in the role of sperm competition in the evolution of sexual traits measured over evolutionary time such as *SEMG2*.

Significance statement

The empirical study of the role of sperm competition in the evolution of sexual traits has historically been problematic through the inability to measure sperm competition levels directly in the present and to reconstruct changes in the evolutionary past. Here, we test a "timeline procedure" based on paternity data that potentially permits both. For our pilot study, we use the rate of change of the seminal protein gene *SEMG2* for catarrhine primates published by Dorus et al. (Nat Genet 36:1326–1329, 2004). Whereas Dorus et al. were limited to using proxy measures for only "recent" segments of catarrhine primate phylogeny, our method permits direct measures of sperm competition to be applied across the whole of Catarrhine phylogeny. We conclude that this new procedure could be valuable for investigation of the role of sperm competition in the evolution of a wide range of sexual traits.

Keywords Paternity data · Sperm competition · Sexual trait evolution · Catarrhini · Timeline procedure

Communicated by E. Huchard

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00265-020-02889-y) contains supplementary material, which is available to authorized users.

✉ Todd K. Shackelford
shackelf@oakland.edu

[1] School of Biological Sciences, University of Manchester, Manchester, UK

[2] Present address: Hard Nut Books Ltd, London, UK

[3] Department of Psychology, Oakland University, Rochester, MI, USA

Introduction

Fifty years ago, focusing on insects, Parker (1970) identified the phenomenon of sperm competition (i.e., the competition between sperm from multiple males to fertilize the egg(s) produced by a single female). Since then, students of sexual traits have demonstrated the importance of this competition as an evolutionary force for a range of animals, including primates (e.g., Smith 1984). These early studies concentrated on the male contribution to sperm competition. More recently, however, the perspective has widened to include the post-copulatory influence of the female and female reproductive

Published online: 09 August 2020

tract, an influence known as cryptic female choice (Eberhard 1996; Lüpold and Pitnick 2018).

Despite theoretical progress in certain areas (Parker 2016), the empirical study of the influence of sperm competition in evolutionary processes has encountered major difficulties. The most notable are an inability to measure sperm competition in real terms (i.e., males/conception) in the present and to estimate figures for the evolutionary past.

In primates, early empirical studies of sperm competition as a factor in the evolution of sexual characteristics concentrated on traits such as relative testes size (Short 1979) that could be assessed at only one point in time, the present. Recently, however, an increasing number of molecular studies have been conducted that measure the change in a trait, such as a seminal protein gene, over time (e.g., Dorus et al. 2004). Some of these changes seem to relate to current sperm competition level (e.g., Dorus et al. 2004) but others do not (e.g., Good et al. 2013). However, as Wong (2010, 2014) has noted, a lack of correlation could result from the methods of measurement. Not only were proxy measures of sperm competition used (i.e., socio-sexual structure and/or relative testes size), but also those measures reflected the modern outcome of evolution, not the changes in sperm competition that occurred during the time over which the molecular changes occurred. If such traits are to be evaluated appropriately in relation to sperm competition, there is a need for a procedure that allows the changes in sperm competition level along different segments of individual phylogenetic lineages to be expressed numerically (for the purposes of this article, a segment is defined as that section of a lineage from one specified internal phylogenetic node to another or from a specified internal node to an extant or extinct single species).

In two recent articles, we developed a method for the direct measurement of current levels of sperm competition using paternity data (Baker and Shackelford 2018a, b). The new metric has the advantage that it measures the level of sperm competition at the moment bouts conclude at conception. The metric therefore automatically weights not only all the male contributions but also cryptic female choice. For a full description, see Baker and Shackelford (2018a). Now, we build on this method to develop a numerical procedure for estimating the changes in level of sperm competition in the evolutionary past. In combination, these two methods allow the relating of such changes to changes in the expression of a trait. We refer to this procedure as a "timeline procedure."

Application of the timeline procedure to a particular phylogenetic group requires three sets of data: (1) paternity data for a sufficient number of representative species; (2) a consensus molecular phylogeny; and (3) sufficient measures of evolutionary changes in a trait. For this pilot article, we have chosen to develop, test, and illustrate the procedure using the catarrhine primates as a test group and the molecular structure of the seminal protein gene, *SEMG2*, as a test trait.

The ejaculate of all catarrhines appears to coagulate soon after ejaculation, showing at least some trace of gelatinous consistency or coagulation (Dixson and Anderson 2002). The two genes *SEMG1* and *SEMG2* are expressed in the seminal vesicles and encode nearly half of the protein in the ejaculate (Hurle et al. 2007). After ejaculation, these *SEMG*-encoded proteins undergo cross-linking to become the principal structural component of the coagulum (Robert and Gagnon 1999). The proteins encoded by *SEMG1* (Robert and Gagnon 1999) and at least a portion of *SEMG2* (Dorus et al. 2004) are thought also to be involved in the inhibition of sperm motility.

SEMG1 and *SEMG2* have been investigated from an evolutionary perspective. Such studies have included the calculation of the number of nonsynonymous substitutions (d_N) (which change the amino acid sequence) and synonymous substitutions (d_S) (which do not change the amino acid sequence). The ratio $(d_N)/(d_S)$ (termed ω) is then calculated as a measure of the rate of protein evolution scaled to mutation rate (Yang 1998).

In this pilot article, for the purposes of the development, testing, and showcasing of the timeline procedure, we focus on the convenient dataset published by Dorus et al. (2004) which also provides a benchmark analysis against which the timeline procedure can be judged. This dataset comprises values of ω for 19 different segments of catarrhine phylogeny, some internal, some terminal. Dorus et al., however, could only analyze seven "recent" segments.

To benchmark and showcase the novel aspects of our timeline procedure, we proceed in three steps: (1) we repeat the analysis of the seven "recent" segments defined by Dorus et al. (2004) but substitute their proxy measure of sperm competition level with our point-in-time measure from paternity data; (2) we then repeat this analysis yet again but this time use the timeline procedure to replace the point-in-time levels of sperm competition with change-over-time measures, thus removing the mismatch in data type (see Wong 2014); and (3) we apply the timeline procedure to all the phylogenetic segments defined by Dorus et al., internal as well as terminal, to show the variety of analyses that the timeline procedure affords.

Methods

All field data used in this article were collected and published by other authors for their own purposes with no knowledge of how they were to be used here. To minimize bias, the subsequent selection, recording, and analysis of these data used blinded methods wherever relevant.

Species referenced

A full taxonomic list of the names of all 32 species/subspecies referenced in this article is provided in Table E-1 in the Electronic Supplementary Material file (ESM_1.pdf) that accompanies this article (Online Resource 1).

All table and section numbers in Online Resource 1 are prefixed with E-. References such as Table E-1 and §E-1 etc. therefore all direct to Online Resource 1. Tables, figs, and sections referenced without this prefix are located in the published article.

Paternity data

The collection of paternity data from populations of primates involves obtaining genetic material from a sample of individuals and their potential parents, and then assigning paternity. A full list of sources, details, and raw paternity data for the 83 study groups that form the basis of this article is given in Table E-2. The criteria for selection of studies are as in Baker and Shackelford (2018a, b). These criteria plus full details of the additions, updates, and minor differences between the data used here and in these earlier publications are itemized in §E-1.

Calculation of current levels of sperm competition from paternity data

The procedures and equations used here were developed, described, and discussed by Baker and Shackelford (2018a, b). Only a summary of the main terms and the equations themselves is provided here.

Main data: designated-male and other-male paternity

The paternity of each female's offspring is expressed with respect to an individual male, here termed the *designated male*, chosen usually by the original field researchers from within that female's range of potential mating partners. Ideally, the designated male has a high (preferably ~ 100%) probability of having sperm present inside the female when she conceives. If the female conceives to her designated male, the offspring is scored as a case of designated-male paternity. All other males within a female's range of potential mating partners are here termed *other males*. The designated male for one female can sometimes be an other male for another female. In this article, we express the paternity data for a study group as the percentage of young with other-male paternity (OMP %) (Table E-2).

Subsidiary data

Although designated-male paternity and other-male paternity are the primary data for further analysis, the formulae below ("Equations") also require values for two subsidiary variables (i.e., subsidiary in the sense that if no usable field data exist, default values can be allocated). The variables are (1) percentage of other-male matings that are multiple matings (POM, default = 100, meaning that all other-male matings are multiple matings insofar as the sperm transferred encounter sperm from at least the designated male) and (2) fertilization bias (i.e., the ratio of observed to randomly expected "wins" by designated males during sperm competition) (FB, default = 1.0, meaning that the sperm set from the designated male and the sperm set of the average other male have an equal chance of fertilizing the egg). When FB > 1.0, this implies that the sperm set of the designated male has the greater chance of fertilizing the female's egg. When FB < 1.0, the converse is true.

A full description of these variables plus a discussion of caveats and their contribution to the signal-to-noise ratio in the calculation of sperm competition levels is provided in Baker and Shackelford (2018a, b). The precise procedure for calculating fertilization bias, however, has of necessity changed over the three publications. The current article raises a particular challenge because it accommodates variation in fertilization bias between different study groups of the same species.

A detailed description of the variation in methodology for calculating fertilization bias across the three articles, the field data used in this article (Table E-3), and a statistical comparison of the results obtained in the three studies (Table E-4) are all presented in §E-2.

Equations

The three equations used to calculate the level of sperm competition from paternity data in this article are:

$$\mathrm{FSC} = \{100\}*\{[\mathrm{OMP}]*[\mathrm{POM}/100]*[1+(\mathrm{FB}/(\mathrm{ISC}-1))]\}/\{\mathrm{DMP}+\mathrm{OMP}\} \quad (1)$$

$$\mathrm{ISC} = [\{\mathrm{FB}\}/\{[((\mathrm{DMP}+\mathrm{OMP})*\mathrm{FSC})/(\mathrm{OMP}*\mathrm{POM})]-1\}]+1 \quad (2)$$

$$\mathrm{LSC} = 1+[(\mathrm{FSC}/100)*(\mathrm{ISC}-1)] \quad (3)$$

where FSC is the frequency of sperm competition (i.e., the proportion of offspring, range 0–100%, whose conception follows a bout of sperm competition between two or more males); DMP is the number or percentage of designated-male paternities; OMP is the number or percentage of other-male paternities; POM is the percentage of other-male matings that are multiple matings; ISC is the intensity of sperm competition (i.e., the number of males, range 2 to *n*, and including the designated male, which have sperm inside the female at

conception); FB is the fertilization bias; and LSC is the level of sperm competition (i.e., the mean number of males, range 1 to *n*, whose sperm are present at each conception, including occasions when no sperm competition occurs).

Although we can conceive of situations in which it may be of interest to analyze FSC and ISC independently, we anticipate that in most studies of sperm competition, as here, LSC will be the variable most often used.

A full account of the development, rationale, and procedural use of these formulae along with a detailed exposition of caveats, signal-to-noise ratio, relationship to theoretical formulations by others (e.g., Parker 2016), and other considerations is provided in Baker and Shackelford (2018a, b).

Sampling bias and correction

Although the basic units for analysis in this article are the 83 study groups in Table E-2, a final analysis requires a value for each parameter at the level of either the species or subspecies, as appropriate. It would be inappropriate, however, simply to calculate an unweighted average for each species from the different study groups for that species. None of the study group values for OMP % presented in Table E-2 was calculated from primary samples that involved structured sampling designed to yield data representative of a population. We have discussed this issue previously (Baker and Shackelford 2018a) and there employed such corrective procedures as possible whenever these could lead to a reduction in sampling bias. Here, as in Baker and Shackelford (2018b), we use up to three factors (depending on data availability) to achieve such a reduction in bias for the measure of sperm competition level (Table E-5).

Reconstruction of past levels of sperm competition for the Catarrhini

Considerable progress has recently been made not only in compiling molecular phylogenetic trees but also in using these trees for statistical analyses that control for phylogenetic artifacts and uncertainty (e.g., Felsenstein 1985; Lutzoni et al. 2001; Pagel and Lutzoni 2002). Procedures have also been developed (e.g., COEVOL (Lartillot and Poujol 2011), Forward Genomics (Hiller et al. 2012), and PhyloAcc (Hu et al. 2019)) that attempt to identify associations between patterns of convergent evolutionary rate shifts and convergent changes in environment or traits.

Such methods undoubtedly signpost the future of phylogenetic analysis. At the same time, it is acknowledged (e.g., Kowalczyk et al. 2018) that there is still room for improvement, particularly in ease of both visualization and downstream analysis. Attempts to remedy the situation are ongoing (e.g., RERconverge; Kowalczyk et al. 2018), but it remains the case that no method is suitable for all situations, and the situation reported in this article is an example.

The *SEMG2* data provided by Dorus et al. (2004) yield information for a variety of internal and terminal phylogenetic segments. To replicate their study properly using our measure of sperm competition level, it is necessary to isolate the change in sperm competition level for those specific segments. Even when data are not collected in this way, it will often be of interest to delve into the evolutionary past to compare changes at specific stages of phylogeny. The method developed here allows such investigations, as this article illustrates.

While seeking to harness the power of the Bayesian methodology at the heart of recent methods, we also needed a procedure that (1) is simple and transparent enough to benchmark our approach against historical data and analyses; (2) allows easily visualized downstream analysis to generate new hypotheses; and, in particular, (3) will allow us in the future easily to switch from testing hypotheses generated via traits based on molecular change against traits that are not. The timeline procedure now to be described (Fig. 1) meets all these requirements.

First, we use Version 3 of the 10kTrees Website (see Arnold et al. 2010) to generate the consensus phylogeny, in Nexus format, for the primate species for which we have relevant data. This phylogeny (with branch lengths) has been sampled from a Bayesian phylogenetic analysis of genetic data and is illustrated in Fig. 2. For analyses, we use version 3.0.1 (November 2017) of BayesTraits (Pagel and Meade 2017).

Among the procedures available in BayesTraits is the facility to estimate the numerical value of a trait, in this case sperm competition level, not only at the root node but also at particular nodes internal to the phylogeny. The former is the phylogenetically corrected mean of the data, whereas an internal node is the most recent common ancestor of a specified group of species (Pagel et al. 2004). To calculate the latter, we use the random walk model and Markov chain Monte Carlo (MCMC) method.

A full reconstruction of potential levels of sperm competition at nodes throughout catarrhine evolution (insofar as can be calculated from the species for which we have paternity data) using BayesTraits is presented alongside the phylogeny in Fig. 2.

For simplicity, Fig. 2 shows only the *levels* of sperm competition at different nodes and current endpoints. These levels, however, also provide the means to quantify *changes* from the beginning to the end of particular lineage segments. For example, the difference between the current sperm competition level (s) for a species and the level (g) of that species' generic ancestor is a measure of the change in sperm competition over that defined segment of the lineage. If the change is expressed as $s - g$, then positive changes signify that sperm competition level has increased over that segment of the lineage and

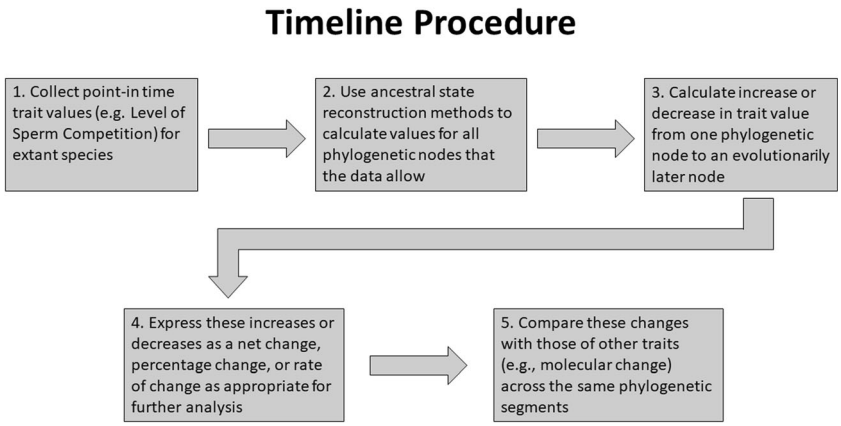

Fig. 1 The five steps of the timeline procedure

negative values that the level has decreased. This procedure can be applied to the segment between any two nodes for which paternity data allow calculation of the two nodes themselves. This yields an advantage over most alternative methods because ancestral segments can then transparently be analyzed in the same way as terminal segments.

It is a moot point whether the changes in sperm competition level calculable from Fig. 2 should be expressed as net change, proportional change, or rate of change. However, as an exploratory article such as this is not particularly threatened by multiple hypothesis testing, we prefer to present analyses covering all viewpoints, thus providing a basis for future discussion, rather than to exclude one approach or another in advance. The different measures reflect different nuances of change in level and it could be of interest how those nuances compare in practice. The data necessary to calculate all three measures of change in sperm competition level over the phylogenetic segments for which Dorus et al. (2004) measured changes in the seminal protein gene *SEMG2* are collected in Table E-6.

We should stress that whichever measure of change is used, our timeline procedure, like all methods based on Brownian

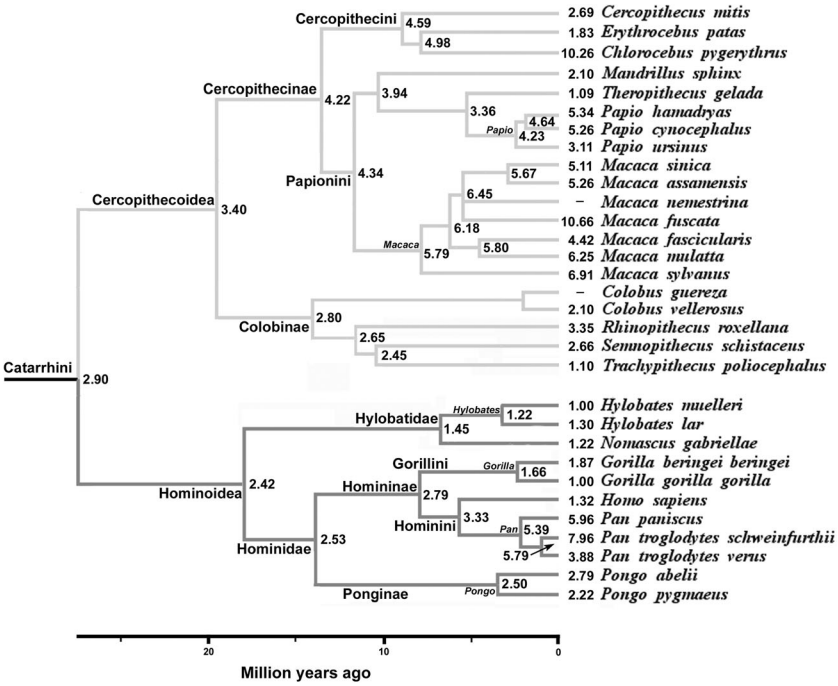

Fig. 2 Constructed phylogenetic timeline of level of sperm competition for 31 species of the Catarrhini. Tree organization and branch lengths extracted from Version 3 of the consensus molecular tree for primates in 10kTrees (Arnold et al. 2010). Level of sperm competition at nodes and tips expressed as mean number of males with sperm in a female's ampulla at conception. Values for extant species are calculated from paternity data (Baker and Shackelford 2018a, b). Values for root or internal nodes (the most recent common ancestor of the species evolutionarily downstream) are calculated using the random walk model and Markov chain Monte Carlo (MCMC) method offered in version 3.0.1 of BayesTraits (Pagel and Meade 2017). The latter values are the mean of 20,000 iterations over 20 separate runs. Although for display purposes the figures are rounded to two decimal places, all calculations use raw data expressed to 5 decimal places. —, no paternity data available for species

movement, assumes a uniform change in trait value along a segment. So, too, do a multitude of other measures, such as the ω measure for molecular change used here. Like all these others, therefore, the timeline procedure could mask a series of evolutionary bursts and constraints during the course of a phylogenetic segment. Unlike other methods, however, the timeline procedure (when data allow calculations for multiple nodes along a lineage) permits longer segments to be broken into shorter segments. This potentially allows analysis of the rate and nature of changes in sperm competition level in these shorter segments with changes in value of some other trait, such as ω, over the same segments. The timeline procedure, therefore, offers the potential for a more transparent, flexible, and finely grained analysis of evolution across a phylogeny than other methods.

Statistics

The timeline procedure, data independence, and analysis

The analytical problems raised by phylogenetic inertia are well known, and many elegant solutions and calculation aids have been developed, such as the BayesTraits collection used here (Pagel and Meade 2017). General issues that relate to the timeline procedure are described in §E-3.

Non-phylogenetic procedures

All standard parametric and non-parametric analyses in this article were generated using the Real Statistics Resource Pack software (Release 5.4; Zaiontz 2018). Tests of the difference between two dependent correlations with one variable in common used Lee and Preacher (2013), following Steiger (1980). Tests of the difference between two independent correlations used Preacher (2002), following Cohen and Cohen (1983).

When calculating probability values, we use the following protocol. If there is no clear direction to the hypothesis being tested, we present 2-tailed values. When there is a clear direction to the hypothesis being tested, we do not present 1-tailed p values but instead follow Rice and Gaines (1994) and present p_{dir} values with γ/α set to 0.8.

Phylogenetic procedures

Correlation analyses with phylogenetic correction were performed for extant species as described by Pagel and Meade (2017). For the primate species for which we have relevant data, we use Version 3 of the 10kTrees Website (Arnold et al. 2010) to generate the consensus phylogeny in Nexus format. We then use the Independent Contrasts facility in version 3.0.1 (November 2017) of BayesTraits (Pagel and Meade 2017) for correlation analysis. The generated statistic (*logBF*) is then interpreted as showing either "no positive evidence," "positive evidence," "strong evidence," or "very strong evidence," following Raftery (1996).

Phylogenetic studies using Bayesian processes often default to using log-transformation of quantitative variables, and where appropriate, we follow this protocol here.

Steps in evaluating and using the timeline procedure

Benchmarking

The first step in evaluating the timeline procedure is to benchmark the relationship between ω and the traditional proxy measure of sperm competition. We then compare this benchmark with that obtained using the new metric (mean males/conception) derived from paternity data (Baker and Shackelford 2018a, b).

Dorus et al. (2004) published values of ω for *SEMG2* for 19 segments of catarrhine phylogeny. They were unable, however, to use their full dataset to test for an association with sperm competition level because they had measures of sperm competition (in the proxy form of relative testes size) only for segments at the phylogenetic tips. To circumvent this problem, Dorus et al. reduced their data for ω to just seven such recent segments (Table E-6, selection A).

The measurements of relative testes size used by Dorus et al. (2004) were residuals from a regression of testes weight on body weight (both log-transformed) published by Anderson and Dixson (2002). When plotted against ω, these residuals show a positive correlation for which the r^2 value was 0.52 and to which Dorus et al. attached a (one-tailed) p value of 0.035 ($n = 7$). A (positive) r^2 value of 0.52 as reported by Dorus et al. translates into a correlation coefficient of $r = 0.721$ ($p_{\mathrm{dir}} = 0.044$; $n = 7$). This is the historic benchmark value, therefore, against which any other correlation coefficients for the same seven segments can be judged.

For the next benchmark, we replace the proxy measure of relative testes size with the new paternity data metric for sperm competition level (mean males/conception) (Table E-6, selection A, "End" column). This gives a figure of $r = 0.744$.

Evaluation of benchmarks in the absence of a known "true" correlation

The benchmarking exercise just described results in two correlation coefficients (0.721 and 0.744; see Table 1, row 1; "point-in-time" columns) for which the comparison of main interest is relative performance. It is a simple matter to calculate the probability of each being different from zero. It is also a simple matter to calculate the probability of the two being different from each other (see "Non-phylogenetic procedures" and "Benchmarks"). Beyond that, though, further interpretation is problematic.

Table 1 Summary of correlation coefficients between ω and measures of level of sperm competition

Row ID	Data selected	*n*	Point-in-time analyses		Change-over-time analyses		
			RTS	LSC	Net	%	rate
ω—untransformed							
1	A—full data set	7	0.721 *	0.744 *	0.737 *	0.745 *	0.899 ***
2	A—without data caveat	6	n/a	0.741 ^{ns}	0.718 ^{ns}	0.733 ^{ns}	0.911 ***
ω—log-transformed							
3	A—full dataset	7	0.780 **	0.798 **	0.797 **	0.804 **	0.888 ***
4	A—without data caveat	6	n/a	0.804 *	0.783 *	0.799 *	0.909 ***
5	B—full data set	15	–	0.291 ^{ns}	0.528*	0.562 **	0.445 ^{ns}
6	B—without data caveats	12	–	0.486 ^{ns}	0.633**	0.676 ***	0.594 *
7	B—Hominoidea	9	–	(0.567 ^{ns})	0.799 ***	0.811***	0.792 ***
8	B—Cercopithecoidea	6	–	(− 0.163 ^{ns})	0.289 ^{ns}	0.314 ^{ns}	0.168 ^{ns}
9	B—internal	7	–	(0.377 ^{ns})	0.394 ^{ns}	0.439 ^{ns}	0.425 ^{ns}
10	B—terminal	8	–	(0.054 ^{ns})	0.552 ^{ns}	0.627 ^{ns}	0.390 ^{ns}
11	Since 10 mya	8	–	(0.322 ^{ns})	0.728 *	0.769 **	0.622 ^{ns}
12	Length < 10 my	11	–	(0.192 ^{ns})	0.575 *	0.612 *	0.430 ^{ns}
13	Length > 10 my	4	–	(0.913 ^{ns})	0.976 **	0.950 *	0.958 *

Data listed are Pearson's parametric correlation coefficients (*r*) for a variety of analyses of association between ω (or logω) and different measures of sperm competition. Values in parentheses are listed for completeness and show that analyses that compare *log*ω with levels rather than changes in level of sperm competition are not significant. These data, though, are not discussed further in the text. *Row ID*, row number for ease of reference from main text. *Data selected* (for details, see Table E-6). *A*, *B*, and caveats as itemized in Table E-6. *Hominoidea*, all segments for which data exist between the catarrhine root node and extant hominoids; *Cercopithecoidea*, all segments for which data exist between the catarrhine root node and extant cercopithecoids; *internal*, segments that end at an internal node; *terminal*, segments that end with an extant species; *since 10 mya*, segments that both start and end within the last 10 million years; *length < 10 my*, segments shorter than 10 million years; *length > 10 my*, segments longer than 10 million years. *n*, number of segments. *Point-in time analyses. RTS*, sperm competition measure based on relative testes size from Dorus et al. (2004); *LSC*, level of sperm competition measure (mean males/conception) from paternity data, using values from Table E-6, "End" column. *Change-over-time analyses*. Change in level of sperm competition (mean males/conception) from timeline procedure (see Figs. 1 and 2), expressed as either *net change* (from "change" column in Table E-6), *% change*, or *rate of change*, calculated from data, and as described, in Table E-6

ω rate of molecular change in the seminal protein gene, *SEMG2* (from Dorus et al. (2004); – not calculable; *n/a* not applicable

P_{dir} values: ^{ns} not significant; *< 0.05; **< 0.02; ***< 0.01

Ideally, the two values would be used to answer the question of which measure of sperm competition level—relative testes size or males/conception—is "better," but this cannot be done. The default assumption is that the stronger the correlation the more accurate the measures that produce the correlation, but this is not necessarily the case. The "true" correlation between ω and level of sperm competition for a sample size of 7, as here, is unknown. If the "true" *r* value were very high, approaching 1.0, then the males/conception measure, as it yields a nominally higher correlation coefficient, could be considered marginally better. However, if the "true" value were low (say < 0.7), then the converse would hold and the relative testes size measure could be considered to have the edge. Without knowing the "true" value, no judgment can be made on the matter.

Although the "true" correlation between ω and sperm competition level is unknown, the current consensus (e.g., Dorus et al. 2004; Wong 2014) is that sperm competition level does have a significant influence on seminal gene evolution. As a consequence, any analysis that fails to yield a significant relationship between the two is likely to be judged to reflect a weakness in either the measures or the method. The result will tend to be considered a false, not a true, negative. Until or unless sperm competition is demonstrated to have no influence on seminal gene evolution, this is likely to remain the case. Beyond that, though, we cannot legitimately interpret further given current knowledge.

This problem of interpretation could potentially arise at a number of points in the following sections. As it happens, though, although there are occasions in our evaluation of the timeline procedure on which one method is significant and another not, there is no occasion on which two correlations for key measures or procedures are significantly different. At no point, therefore, is a judgment required over which

measure or method is the "better." It is sufficient in this article to accept that the two measures or methods being compared are equally suitable for the purpose.

Point in time and change over time

Although the two benchmark calculations described in "Benchmarking" are an important starting point in the evaluation of the timeline procedure, they are both potentially inappropriate (see Wong 2014). In both cases, they are a comparison of an across-segment measure (for ω) and a point-in-time end-taxon measure (for sperm competition). Ideally, to be comparable with ω, the metric for sperm competition should also be an across-segment change-over-time measure which has never before been available. Such a measure can be generated by the timeline procedure as in Table E-6.

The next step in evaluating the reliability of the timeline procedure, therefore, is to compare the benchmark measures described in "Benchmarking" against correlations between ω and timeline-generated measures of change in level of sperm competition across the same segments (see "Comparison of point-in-time and change-over-time metrics for sperm competition").

Use of the timeline procedure to dissect relationships across a phylogeny

Although Dorus et al. (2004) published values of ω for 19 phylogenetic segments spread across catarrhine phylogeny, the authors were unable to conduct analyses with respect to sperm competition for more than seven recent segments. Even then three of these segments required a preliminary modification of the original data in order to be analyzable. Partly this reduction in number of segments was because Dorus et al. had to reject four segments with a value of infinity. Primarily, though, it was because the authors had no means of obtaining usable measures of sperm competition for segments internal to catarrhine phylogeny (i.e., not ending with an extant species). The timeline procedure solves this latter problem and hence allows us to analyze much more of the dataset published by Dorus et al. than previously possible. Moreover, the timeline procedure removes the requirement to modify some data, thus allowing analysis of the original values.

All 15 of the usable values of ω (Table E-6, selection B) are independent mathematically (see "The timeline procedure, data independence, and analysis"). So, too, are the changes in level of sperm competition across the same 15 segments. The result is a dataset that covers catarrhine phylogeny from root to tips. This allows us to evaluate the timeline procedure's potential to dissect the relationship between sperm competition and ω at different stages and along different branches of the phylogeny. Sample sizes are inevitably low, and the evaluations presented in "Use of the timeline procedure to dissect evolution across a phylogeny" are more to illustrate potential future use than to provide firm answers.

Generating and testing of hypotheses and the use of both parametric and non-parametric statistics

A particular feature of the timeline procedure is that the measures of change, generated for individual segments across a phylogeny as above, can be separated, isolated, and then analyzed using standard statistical methods. In "Use of the timeline procedure to identify questions and to generate and test new hypotheses," we present just a few illustrations of the procedure being used to frame questions and to generate and test hypotheses. These include in "Do changes in sperm competition level generate positive, stabilizing, or relaxed selection on the evolution of SEMG2?" an example of the ease of switching to non-parametric methods when deemed necessary.

Results

The full list of phylogenetic segments to be analyzed and the appropriate data for ω and sperm competition level is provided in Table E-6.

Sections "Benchmarks" to "Use of the timeline procedure to dissect evolution across a phylogeny" involve the calculation of parametric correlation coefficients between ω and sperm competition level for several datasets taken from Table E-6. To provide an overview of these analyses, to facilitate comparisons, and to minimize the need to present significance levels and sample sizes for every comparison made below, we have collected these correlation coefficients together in a single table (Table 1).

Benchmarks

The two benchmark correlation coefficients (r) between ω and sperm competition for the seven recent phylogenetic segments defined by Dorus et al. (2004) were 0.721 (obtained by Dorus et al. using residual relative testes size as a proxy measure of sperm competition) and 0.744 (obtained by ourselves using the different metric of mean males/conception from paternity data) (Table 1, row 1, "point-in-time" columns). Both correlation coefficients are significant (but not (row 2) if the data point with a caveat (see Table E-6) is excluded). The correlations also show "very strong evidence" for a relationship between ω and level of sperm competition when a phylogenetic correction is applied ($\log BF > 10$ in all cases). However, none of the differences between r values are significant ($p_{2\text{-tailed}} > 0.5$ for any pairwise comparison).

We conclude, therefore, that (1) use of a new measure of sperm competition calculated from paternity data provides

independent support for the hypothesis (henceforth the Dorus-1 hypothesis) that level of sperm competition has an influence on *SEMG2* evolution and (2) the new measure of sperm competition is as suitable for such a study as the traditional proxy measure based on relative testes size.

Comparison of point-in-time and change-over-time metrics for sperm competition

Calculations such as performed for the two benchmarks are potentially inappropriate because whereas ω is a change-over-time measure, those for sperm competition are both point in time (see Wong 2014). The timeline procedure, however, generates a measure for sperm competition (Table E-6, "change" column) that, because it expresses change over time, is potentially more appropriate. The correlation of this measure with ω can thus be compared with the two benchmarks from "Benchmarks."

First, for direct comparison, we follow Dorus et al. (2004) and use untransformed values of ω for the same seven recent segments (Table E-6, selection A). The change in level of sperm competition across a segment can be expressed in three different ways: net, percentage, and rate. The respective correlation coefficients are 0.737, 0.745, and 0.899. All three correlations are significant (Table 1, row 1, "change-over-time" analyses) and alter only marginally (though two just lose significance) if the data point with a caveat is excluded (row 2). All correlations also show "very strong evidence" for a relationship between ω and level of sperm competition when a phylogenetic correction is applied ($\log BF > 10$ in all cases). However, no pairs of values are significantly different, either from each other or from either of the two benchmarks ($p_{2\text{-tailed}} > 0.05$ in all cases). We conclude, therefore, that (1) use of a more appropriate change-over-time metric for sperm competition still supports the Dorus-1 hypothesis, and (2) the change-over-time values generated by the timeline procedure are as suitable for investigative use as were the point-in-time measures used to calculate the benchmark values.

Next, instead of using untransformed measures of ω, we switch (see "Phylogenetic procedures") to using log-transformed measures. The r values change little (compare rows 1 and 3, and rows 2 and 4; Table 1) and all are significantly greater than zero. Still, though, no given pair of correlation coefficients is significantly different. All correlations also continue to show "very strong evidence" for a relationship between ω and level of sperm competition when a phylogenetic correction is applied ($\log BF > 10$ in all cases). We conclude, therefore, that neither the support for the Dorus-1 hypothesis nor the suitability of the timeline procedure is influenced by the log-transformation of ω.

Use of the timeline procedure to dissect evolution across a phylogeny

We stress that the following statistics are primarily illustrative, a demonstration of how the timeline procedure can be used. Sample sizes are small and correlations are sometimes not significant. The biological conclusions, therefore, should be considered preliminary. The technique, though, could have far-reaching applications.

We now switch from analyzing the seven recent segments defined by Dorus et al. (2004) to analyzing their whole usable dataset (Table E-6, selection B). In this section ("Use of the timeline procedure to dissect evolution across a phylogeny"), we are concerned only with the change-over-time analyses in Table 1.

Analysis of the whole dataset shows that two of the three measures of change in the level of sperm competition (net and percentage) have a significant positive association with logω (Table 1, row 5). If we exclude the three data points with caveats described in Table E-6 and marked in Fig. 3, the correlations alter only marginally (compare rows 5 and 6, Table 1; no differences significant, $p_{2\text{-tailed}} > 0.05$) but all are now significantly different from zero.

We conclude that extending analysis to use the whole dataset continues to provide support for the Dorus-1 hypothesis. The timeline procedure, however, has the potential to do much more than extend this support. It also allows us to begin to dissect the association of sperm competition with ω at different stages and along different branches of the phylogeny. The results are collected together in Table 1 (rows 7 to 13). Sample sizes vary from 4 to 11. All correlation coefficients are positive and there are no significant differences between

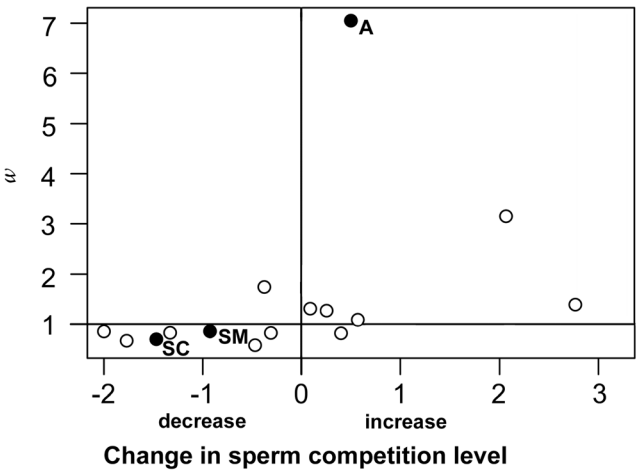

Fig. 3 Change (ω) across a phylogenetic segment in the seminal protein gene, *SEMG2*, plotted against the net change in level of sperm competition (mean males/conception) across the same segment. Open circles, data without caveats. Filled circles, data points with caveats (see Table E-6 for full details): A, outlier; SC, *Colobus vellerosus* paternity data substituted for *C. guereza*; SM, *Macaca assamensis* paternity data substituted for *M. nemestrina*

relevant pairs (e.g., rows 7/8, 9/10, and 12/13). Some correlations are significant and some are not, but as we are not seeking firm conclusions in this section, multiple hypothesis testing is not a concern. The aim is to illustrate the potential of the timeline procedure and to provide a general impression of selection across the phylogeny. When we apply a phylogenetic "correction" to the seven terminal segments (row 10), we obtain log$BF > 2$ ("positive evidence") for all three measures of change.

The primary impression is that the relationship between sperm competition and *SEMG2* evolution has remained relatively constant across catarrhine phylogeny. For all subdivisions of the data, an increase in level of sperm competition across a segment is associated with a nominally higher value of ω, sometimes significant, sometimes not. There is neither a statistical difference ($p_{2\text{-tailed}} > 0.05$) nor even a suggestion of a difference in the relationship between segments no matter whether they are internal or terminal (rows 9 and 10), are less than or greater than 10 million years long (rows 12 and 13), or have both started and finished within the last 10 million years (row 11). There is perhaps a hint worthy of future attention that the influence of sperm competition on ω may have been stronger within the Hominoidea than the Cercopithecoidea (rows 7 and 8), but the differences are not at present significant. More paternity data for more species, particularly for a wider range of cercopithecoids, are needed before this hint can be judged further.

Use of the timeline procedure to identify questions and to generate and test new hypotheses

Which is the more important in *SEMG2* evolution: level of sperm competition or change in level of sperm competition?

All analyses so far have provided support for the Dorus-1 hypothesis. We now apply the timeline procedure to the more specific hypothesis (henceforth Dorus-2) that higher levels of sperm competition generate more positive selection for change in *SEMG2* than lower levels.

Dorus et al. (2004) based this second part of their hypothesis on the fact that the correlation between ω and their relative testes size point-in-time measure of sperm competition had a positive sign ($r = 0.721$; Table 1, row 1, "RTS" column). Our re-analysis of these same species and segments (selection A) using log-transformed values of ω and a different (i.e., mean males/conception) point-in-time metric for sperm competition yielded a correlation coefficient of $r = 0.798$ (Table 1, row 3, "LSC" column). At first sight, this supports the Dorus-2 hypothesis, but further investigation using the timeline procedure reveals a problem.

The analyses shown in Table 1 (row 3) indicate that, as stated in Dorus-2, higher values of sperm competition at the end of a segment are indeed associated with higher values of ω over the course of a segment. However, the same data (row 3) show that it is equally true that higher values of ω are also associated with greater *increases in level* of sperm competition. This is because for the seven segments in selection A (Table E-6), there is a strong cross-correlation between the level of sperm competition at the end of the segment and the change in level along the segment (net, $r = 0.982$; %, $r = 0.991$; rate, $r = 0.869$; $p_{2\text{-tailed}} < 0.01$ in all cases; $n = 7$). It follows, therefore, that although selection on ω could be generated by sperm competition level (as measured at the tip and as hypothesized in Dorus-2), it could equally well be generated by the change in sperm competition level along the segment.

The timeline procedure allows the situation to be investigated further. Not only does it allow us to extend the analysis to use the full dataset (Table E-6, selection B) but it also allows us to use a more targeted analysis that can answer specific questions.

Analysis of the full dataset (see Table 1, row 5, "LSC" column) shows that there is no longer a significant correlation between logω and the *level* of sperm competition at the end of a segment. There remains, however, a significant correlation between logω and two of the different expressions of *change in level* (Table 1, row 5, "change-over-segment" columns). This hints that change in level of sperm competition across a segment may be the more important, or even the only important, factor associated with logω. However, in no case is the difference between the correlation for sperm competition level (row 5, "LSC" column) and the different correlations for change (row 5, "change-over-time" columns) significant at the $p_{2\text{-tailed}} = 0.05$ level, so a simple examination of correlation coefficients cannot settle the matter.

A more targeted approach is needed, and to illustrate the analysis, we use the data for net change in sperm competition level ($r = 0.528$: Table 1, row 5, "net" column). Residuals from the regression line of logω on the net change in sperm competition level (x males/conception) across a segment (i.e., $\log\omega = 0.113x + 0.084$; $R^2 = 0.278$; $F_{1, 13} = 5.023$; $p_{\text{dir}} = 0.028$; type III sums of squares) show no significant correlation with the end-of-segment level ($r = -0.074$; $p_{2\text{-tailed}} = 0.396$; $n = 15$). Nor does the addition of end-of-segment level add significantly ($p_{2\text{-tailed}} > 0.700$) to the 27.8% ($R^2$) variance already explained by net change in level across the segment alone (raising the explained variance to only 28.6%). Moreover, the VIF of the two factors (net change and end-of-segment value) is 1.8. This is low enough (< 5.0; Ringle et al. 2015) for us to conclude that the result is not unduly influenced by multicollinearity.

We conclude, therefore, that the Dorus-2 hypothesis needs modification. It is not sperm competition level that is the important factor in *SEMG2* evolution, but change in level. This conclusion remains the same if we use the mean level of sperm

competition across a segment (Table E-6) instead of end-of-segment level. It also stays the same if we use percentage change or rate of change of sperm competition level (Table 1, row 5, columns "%" or "rate") instead of net change. Having led to this conclusion, the timeline procedure now allows the investigation to continue even further.

Do changes in sperm competition level generate positive, stabilizing, or relaxed selection on the evolution of SEMG2?

As a final illustration of the breadth and flexibility of the timeline procedure, we investigate the association between changes in sperm competition level and the value of ω in more detail. At the same time, we demonstrate the ease with which analysis can use non-parametric statistical methods when deemed necessary.

Although qualifications exist, values of ω that are higher than 1.0 indicate positive selection for change and values lower than 1.0 indicate stabilizing selection (Yang 1998). Values around 1.0 indicate relaxed (or no) selection. From the conclusion in "Which is the more important in SEMG2 evolution: level of sperm competition or change in level of sperm competition?" we would predict, therefore, that increases in level of sperm competition across a segment would produce values of $\omega > 1.0$ whereas decreases in level would not. Whether decreases in sperm competition should produce values of ω that are lower than 1.0 (stabilizing) or simply values not different from 1.0 (relaxed) is a moot point.

The 15 lineage segments in Table E-6, selection B, can be divided into two groups that allow these expectations to be investigated. There are those segments for which the change in sperm competition level is positive (the "increase in sperm competition" group; $n = 7$) and those for which the change is negative (the "decrease in sperm competition" group; $n = 8$). As the data in some representations of these groups depart from a normal distribution (d'Agostino-Pearson test: $p < 0.05$), we switch to the use of non-parametric procedures for one- and two-sample tests.

The distribution of values of ω and changes in levels of sperm competition (Table E-6, selection B; see also Fig. 3) are such that in the increase in level of sperm competition group, the value of ω is > 1 on 6 occasions and < 1 on 1 occasion. In the decrease group, the value of ω is > 1 on only 1 occasion and < 1 on 7 occasions (chi-squared = 8.958; $p_{2\text{-tailed}} = 0.003$; df = 1).

Comparison of the increase and decrease groups shows that the median value of ω (1.31, range 0.82–7.05) in the increase group is significantly higher than that in the decrease group (0.85, range 0.58–1.73) (Mann-Whitney test: $U_{7,8} = 9$; $z = 2.199$; $p_{2\text{-tailed}} = 0.036$). Moreover, the median ω for the increase group is significantly higher than 1.0 (Wilcoxon signed-rank test for a single sample: $T = 2$; $p_{2\text{-tailed}} = 0.042$; $n = 7$). In contrast, the median for the decrease group is not significantly lower than 1.0 ($T = 8$; $p_{2\text{-tailed}} = 0.196$; $n = 8$).

We hypothesize, therefore, that an increase in sperm competition level across a segment leads to positive selection on *SEMG2* evolution and a decrease leads to relaxed (or no) selection.

Discussion

The timeline procedure: signal-to-noise ratio, analytical power, and functionality

The timeline procedure, as developed here, builds on recently developed formulae that transform paternity data into a measure of sperm competition level in terms of mean males/conception (Baker and Shackelford 2018a, b). The signal-to-noise ratio of this measure has been discussed in detail elsewhere (Baker and Shackelford 2018b), and the conclusion was reached that the sources of noise in the base data were insufficient to prevent signal strength (e.g., interspecific variation) from showing through. In this article, however, the production of the timeline procedure (Figs. 1 and 2) has added a further source of noise: the Bayesian reconstruction of a trait across a phylogeny that spans up to ~ 30 million years.

If the additional noise generated by the reconstruction process, based on the procedures in BayesTraits (Pagel et al. 2004; Pagel and Meade 2017), had been great enough to swamp any signal in the data, there would be no reason other than chance or artifact to obtain the significant correlations shown in Table 1 in the change-over-time columns. Changes in sperm competition level and values for ω would not have correlated so strongly, nor are these variables likely to become correlated due to cryptic methodological artifacts. We therefore conclude that the signal-to-noise ratio for the timeline data in Fig. 2 is strong enough for the procedure to be a valuable investigative tool in the role of sperm competition in the evolution of sexual traits measured over evolutionary time, such as *SEMG2*.

The next important question is whether the timeline procedure as based on our new metric for sperm competition level is as powerful as could perhaps be achieved using the previous proxy metric of relative testes size (which itself suffers from noise in data collection; see discussion in Baker and Shackelford 2018b). We have argued previously that because the two measures correlate strongly, then if relative testes size is considered powerful enough for exploring a role for sperm competition in the evolution of a trait then so should the measure from paternity data (Baker and Shackelford 2018a, b). At the same time, we accepted that in principle there was one reason for considering relative testes size to be the more appropriate of the two measures. This reason was that although relative testes size itself is a point-in-time tip-taxon measure, a given relative testes size does have the advantage that it evolved over the course of a phylogenetic

segment alongside the test trait, such as the *SEMG2* gene. Potentially, therefore, relative testes size could reflect the across-segment influence of sperm competition better than the purely tip-segment measure from paternity data that we used in previous publications. The timeline procedure, however, renders the two measures more equivalent.

For the seven phylogenetic segments shown in Table E-6 (selection A), the measures of relative testes size used by Dorus et al. (2004) produced a correlation with ω of r = 0.721. By comparison, three different measures of change in sperm competition level generated by the timeline procedure yielded correlations of 0.737, 0.745, and 0.893 ("Comparison of point-in-time and change-over-time metrics for sperm competition"). There is no indication in these figures that relative testes size is any way more suitable a reflection of changes in sperm competition level than the measure that can be extracted from paternity data by the timeline procedure. Added to this observation is the fact that, as illustrated in "Use of the timeline procedure to dissect evolution across a phylogeny" and "Use of the timeline procedure to identify questions and to generate and test new hypotheses," when used as part of the timeline procedure, the males/conception measure has a much more varied functionality.

In fact, relative testes size, when measured as a residual, cannot be used to measure changes through phylogeny. Residual values are specific to the assemblage of species for which they are calculated. Add or subtract even a single species to or from the assemblage and the residual value for each species in the analysis changes. Residual relative testes sizes cannot therefore be compared through phylogeny because the assemblage of species from which the residuals could be calculated automatically changes. A residual value, say, of 2.0 for a tribal ancestor cannot be considered equivalent to a value of 2.0 for a generic ancestor. Moreover, it is not possible even to calculate a residual value at the root node because there is only one species. Only if relative testes size is measured as a percentage could it then be subjected to the timeline procedure and used for across-phylogeny analyses, a possibility that we intend to explore in a future article.

Overall, therefore, we suggest that whenever sufficient paternity data are available, they provide, within the timeline procedure, a much more flexible metric than relative testes size for the investigation of a role for sperm competition in the evolution of sexual traits.

Sperm competition and the evolution of sexual traits: will absolute level or change in level generally be found to be the more important?

The main aim of this article was to develop, evaluate, and illustrate the timeline procedure as an analytical tool in the evolution of sexual traits using the *SEMG2* gene as a test trait and the Catarrhini as a source of appropriate data. In pursuance of this aim, however, we discovered that irrespective of actual level of sperm competition an increase in level generates positive selection for change in a trait whereas a decrease generates relaxed (or no) selection.

How generally in the evolution of sexual traits a change in sperm competition level is more important than the absolute level itself must await the investigation of other traits in the Catarrhini and other taxa. We might expect that when a trait (such as the ω measure of *SEMG2*) is measured as a rate, then change in sperm competition level will be the more important factor, but that when a trait (such as relative testes size) is measured as a point-in-time value, then the converse may be true. In both types of trait, however, it might be expected that absolute level and change in level of sperm competition might interact in some multi-factorial manner, though in this study of *SEMG2* we found no evidence of such an interaction ("Which is the more important in SEMG2 evolution: level of sperm competition or change in level of sperm competition?"). Analysis of a variety of traits using the timeline procedure with a view to addressing this question is an obvious next step.

Acknowledgments We thank Holger Herlyn for many helpful suggestions for the improvement of our manuscript. We also thank an anonymous reviewer for an extensive and detailed review of our original submission. It was an example of constructive reviewing at its best, and our paper is stronger and better organized as a result.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval As all data for catarrhines used here are from published studies by other authors, the protocol and procedures did not require review and approval by the IACUC or other institutional ethics committees overseeing the use of humans or other animals in research in either the USA or the UK. The study also adhered to the American Society of Primatologists Principles for the Ethical Treatment of Non-Human Primates.

Informed consent/consent to participate Informed consent is not relevant to this article.

Consent for publication Not applicable

Data availability All data generated or analyzed for this study are included in this published article and its supplementary information file.

Code availability Not applicable

References

Anderson MJ, Dixson AF (2002) Sperm competition: motility and the midpiece in primates. Nature 416:496–496. https://doi.org/10.1038/416496a

Arnold C, Matthews LJ, Nunn CL (2010) The 10kTrees website: a new online resource for primate phylogeny. Evol Anthropol 19:114–118. https://doi.org/10.1002/evan.20251

Baker RR, Shackelford TK (2018a) A comparison of paternity data and relative testes size as measures of level of sperm competition in the Hominoidea. Am J Phys Anthropol 16:421–443. https://doi.org/10.1002/ajpa.23360

Baker RR, Shackelford TK (2018b) Paternity data and relative testes size as measures of level of sperm competition in the Cercopithecoidea. Am J Primatol 80:e22937. https://doi.org/10.1002/ajp.22937

Cohen J, Cohen P (1983) Applied multiple regression/correlation analysis for the behavioral sciences. Erlbaum, Hillsdale

Dixson AF, Anderson MJ (2002) Sexual selection, seminal coagulation and copulatory plug formation in primates. Folia Primatol 73:63–69. https://doi.org/10.1159/000064784

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. Nat Genet 36:1326–1329. https://doi.org/10.1038/ng1471

Eberhard W (1996) Female control: sexual selection by cryptic female choice. Princeton University Press, Princeton

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15. https://doi.org/10.1086/284325

Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R, Fischer A, Pääbo S (2013) Comparative population genomics of the ejaculate in humans and the great apes. Mol Biol Evol 30:964–976. https://doi.org/10.1093/molbev/mst005

Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G (2012) A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep 2:817–823. https://doi.org/10.1016/j.celrep.2012.08.032

Hu Z, Sackton TB, Edwards SV, Liu JS (2019) Bayesian detection of convergent rate changes of conserved noncoding elements on phylogenetic trees. Mol Biol Evol 36:1086–1100. https://doi.org/10.1093/molbev/msz049

Hurle B, Swanson W, Green ED, Comparative Sequencing Program NISC (2007) Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. Genome Res 17:276–286. https://doi.org/10.1101/gr.6004607

Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M (2018) RERconverge: an R package for associating evolutionary rates with convergent traits. bioRxiv, 451138. https://doi.org/10.1101/451138

Lartillot N, Poujol R (2011) A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol Biol Evol 28:729–744. https://doi.org/10.1093/molbev/msq244

Lee IA, Preacher KJ (2013) Calculation for the test of the difference between two dependent correlations with one variable in common [computer software], http://quantpsy.org

Lüpold S, Pitnick S (2018) Sperm form and function: what do we know about the role of sexual selection? Reproduction 155:R229–R243. https://doi.org/10.1530/REP-17-0536

Lutzoni F, Pagel M, Reeb V (2001) Major fungal lineages are derived from lichen symbiotic ancestors. Nature 411:937–940. https://doi.org/10.1038/35082053

Pagel M, Lutzoni F (2002) Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In: Lässig M, Valleriani A (eds) Biological evolution and statistical physics. Springer-Verlag, Berlin, pp 148–161

Pagel M, Meade A (2017) BayesTraits, www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html

Pagel M, Mead A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. Syst Biol 53:673–684. https://doi.org/10.1080/10635150490522232

Parker GA (1970) Sperm competition and its evolutionary consequences in the insects. Biol Rev 45:525–567. https://doi.org/10.1111/j.1469-185X.1970.tb01176.x

Parker GA (2016) The evolution of expenditure on testes. J Zool 242:3–19. https://doi.org/10.1111/jzo.12297

Preacher KJ (2002) Calculation for the test of the difference between two independent correlation coefficients, http://quantpsy.org

Raftery AE (1996) Hypothesis testing and model selection. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman and Hall, London, pp 163–188

Rice WD, Gaines SD (1994) 'Heads I win, tails you lose': testing directional alternative hypotheses in ecological and evolutionary research. Trends Ecol Evol 9:235–237. https://doi.org/10.1016/0169-5347(94)90258-5

Ringle CM, Wende S, Becker JM (2015) SmartPLS 3. SmartPLS GmbH, Boenningstedt

Robert M, Gagnon C (1999) Semenogelin I: a coagulum forming, multifunctional seminal vesicle protein. Cell Mol Life Sci 55:944–960. https://doi.org/10.1007/s000180050346

Short RV (1979) Sexual selection and its component parts, somatic and genital selection, as illustrated by man and the great apes. Adv Stud Behav 9:131–158. https://doi.org/10.1016/S0065-3454(08)60035-2

Smith RL (ed) (1984) Sperm competition and the evolution of animal mating systems. Academic Press, London

Steiger JH (1980) Tests for comparing elements of a correlation matrix. Psychol Bull 87:245–251. https://doi.org/10.1037/0033-2909.87.2.245

Wong A (2010) Testing the effects of mating system variation on rates of molecular evolution in primates. Evolution 64:2779–2785. https://doi.org/10.1111/j.1558-5646.2010.01038.x

Wong A (2014) Covariance between testes size and substitution rates in primates. Mol Biol Evol 31:1432–1436. https://doi.org/10.1093/molbev/msu091

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573. https://doi.org/10.1093/oxfordjournals.molbev.a025957

Zaiontz C (2018) Real statistics using Excel, www.real-statistics.com