Pervasive relaxed selection on spermatogenesis genes coincident with the evolution of polygyny in gorillas

Jacob Bowman[1,*], Neide Silva[2,3,*], Erik Schüftan[4,7], Joana M. Almeida[2], Rion Brattig-Correia[3], Raquel A. Oliveira[3,5], Frank Tüttelmann[4], David Enard[6**], Paulo Navarro-Costa[2,3,**], and Vincent J. Lynch[1,**]

[1] Department of Biological Sciences, University at Buffalo, SUNY, 551 Cooke Hall, Buffalo, NY, USA.

[2] EvoReproMed Lab. Environmental Health Institute (ISAMB), Associate Laboratory TERRA, Faculty of Medicine, University of Lisbon, Portugal

[3] Gulbenkian Science Institute, Oeiras, Portugal

[4] Institute of Reproductive Genetics, Centre of Medical Genetics, University of Münster, Münster, Germany

[5] Católica Biomedical Research Centre, Católica Medical School, Lisbon, Portugal

[6] Department of Ecology and Evolutionary Biology. University of Arizona, Tucson, AZ, USA

[7] Present address: Department of Pediatric Hematology and Oncology, University Hospital Münster, Münster, Germany

* and ** Equal contribution

*Correspondence: vjlynch@buffalo.edu, navarro-costa@medicina.ulisboa.pt, denard@arizona.edu

Abstract

Gorillas have a polygynous social system in which the highest-ranking male has almost exclusive access to females and sires most of the offspring in the troop. Such behavior results in a dramatic reduction of sperm competition, which is ultimately associated with numerous traits that cause low efficacy of gorilla spermatogenesis. However, the molecular basis behind the remarkable erosion of the gorilla male reproductive system remains unknown. Here, we explored the genetic implications of the polygynous social system in gorillas by testing for altered selection intensity across 13,310 orthologous protein-coding genes from 261 Eutherian mammals. We identified 578 genes with relaxed purifying selection in the gorilla lineage, compared with only 96 that were positively selected. Genes under relaxed purifying selection in gorillas have accumulated numerous deleterious amino acid substitutions, their expression is biased towards male germ cells and are enriched in functions related to meiosis and sperm biology. We tested the role of gorilla relaxed genes previously not implicated in male reproductive function using the *Drosophila* model system and identified 41 novel spermatogenesis genes required for normal fertility. Furthermore, by exploring exome/genome sequencing data of infertile men with severe spermatogenic impairment, we found that the human orthologs of the gorilla relaxed genes are enriched for loss-of-function variants in infertile men. These data provide compelling evidence that reduced sperm competition in gorillas is associated with relaxed purifying selection on genes related to male reproductive function. The accumulation of deleterious mutations in these genes likely provides the mechanistic basis behind the low efficacy of gorilla spermatogenesis and uncovers new candidate genes for human male infertility.

Introduction

Sexual selection is a mechanism of evolutionary change based on direct interactions within and between males and females for access to reproductive opportunities. Darwin first proposed this mechanism to explain paradoxical traits that

Natural Selection seemed unable to account for because of their obvious fitness costs. Sexual selection can take many forms and is associated with diverse behavioral and morphological characteristics. Among the most common is male-male competition, in which males directly compete for access to females. This intrasexual selection is associated with the evolution of traits that facilitate competition, such as antler in Cervidae (deer) and horn in Artiodactyla (even-toed ungulates). Likewise, female choice, in which females actively select mates, also drives the evolution of male traits such as elaborate behaviors and ornaments, including gaudy feather colors, sizes, and shapes, as observed in peacocks and birds-of-paradise. Conversely, cryptic female choice, in which females exert physical or chemical control over fertilization, implantation, or otherwise reduce the rate at which offspring are produced (Eberhard, 1996; Firman et al., 2017; Parker, 1970), is associated with the evolution of female traits such as anatomical features that allow female control over cellular processes occurring in the reproductive tract (Brennan et al., 2007), and the Bruce effect, in which females terminate pregnancy when exposed to an unrelated male (Bruce, 1959).

In polygamous (multimale–multifemale) species, multiple males frequently mate with a single female in a relatively short time, leading to competition not only for the opportunity to mate but for sperm to fertilize the egg ("sperm competition") and for females to select the best sperm from among the males she has mated with ("cryptic female choice"). In these mating systems, sperm from different males compete within the female reproductive tract (Dixson, 2012; Short, 1979). This can affect the behavior, morphology, and functional anatomy of male and female reproductive systems, among other functions. In primates with high sperm competition, for example, sexual selection has led to the evolution of larger testis relative to body weight (Harcourt, 1997, 1995; Harcourt et al., 1981), high sperm production rates, large sperm reserves, and large numbers of sperm per ejaculate (Møller, 1998). In addition, the *vas deferens* are shorter and have a thicker muscular structure in males from species with polygamous mating systems than in males from single-partner mating systems, presumably to facilitate ejaculation (Anderson et al., 2004). Male gametes in polygamous mating systems also tend to have a larger sperm midpiece volume (Anderson and Dixson, 2002) and higher swimming speed than those found in single-partner mating systems (Gómez Montoto et al., 2011). Female-female competition for mating opportunities can also alter the intensity

87 of male-male competition, selecting longer sperm and less variable sperm traits (Lipshutz
88 et al., 2023).

89 These and other data indicate that the action of sexual selection on the structure
90 and function of the male reproductive system, semen, and sperm traits is particularly
91 intense in polygamous primate species. In contrast, some species have reproductive
92 systems in which nearly all male-male competition occurs before mating, and there is
93 little to no sperm competition. Most gorillas, for example, live in groups with age-graded
94 dominance structures where the oldest, highest-ranking (alpha) silverback male is
95 dominant or in groups with multiple females and a single resident adult male. In both
96 systems, the dominant or sole adult male has nearly exclusive mating access to in-group
97 females (effective polygyny) and sires the majority, but not necessarily all, offspring
98 (Bradley et al., 2005; Inoue et al., 2013; Nsubuga et al., 2008). Thus, in most gorillas,
99 competition between males for reproductive success occurs before copulation, through
100 social and group dynamics, rather than between sperm from different males within the
101 female reproductive tract.

102 Consistent with this pattern of male-male competition, sexual selection in gorillas
103 has led to the evolution of very large bodies and behaviors to protect their reproductive
104 access rather than post-copulatory traits that facilitate sperm competition. Indeed, the
105 dramatic reduction in sperm competition is associated with the evolution of many derived
106 traits in the gorilla male reproductive system (**Box 1**), including relatively small testicles
107 with few spermatogenic cells (Harcourt et al., 1981; Hill and Harrison-Matthews, 1949),
108 low sperm counts (Fujii-Hanamoto et al., 2011; Gould, 1990; Martinez and Garcia, 2020),
109 and a large proportion of morphologically abnormal (Martinez and Garcia, 2020; Platz et
110 al., 1980; Seuanez et al., 1977) and immotile sperm (Martinez and Garcia, 2020). Along
111 with these phenotypic changes, gene expression patterns have diverged in the gorilla
112 testis (Yapar et al., 2021), and several genes with male reproductive functions have
113 become nonfunctional (pseudogenes). Examples of these include the testis-specific
114 histone H3 (*H3t*), which plays a role in orchestrating the histone-protamine transition,
115 several genes involved in semen coagulation and liquefaction, including *TGM4*, *KLK2*,
116 *SEMG1,* and *SEMG2* (Carnahan and Jensen-Seaman, 2008; Clark and Swanson, 2005;
117 Jensen-Seaman and Li, 2003), and *TEX14* which encodes an intercellular bridge protein

118 essential for spermatogenesis (Greenbaum et al., 2006; Krausz et al., 2020; Scally et al.,
119 2012).

120 These data illustrate the pervasive impact sexual selection can have on
121 morphology, physiology, behavior, genetics, and the genome. However, the true extent
122 of the latter remains largely unknown. Here, we used a suite of evolutionary methods to
123 characterize the strength and direction of selection acting on 13,310 protein-coding
124 genes across 261 mammals, explicitly testing for genes that experienced adaptive
125 evolution and relaxed purifying selection in the gorilla lineage. We identified only 96 genes
126 that experienced an episode of positive selection but 578 that evolved under relaxed
127 selection intensity in gorillas. These 578 genes have accumulated putatively deleterious
128 amino acid substitutions, are preferentially expressed during male germ cell
129 development, and are enriched in functions related to testis and sperm development.
130 Taking advantage of the conservation of the genetic program responsible for male germ
131 cell development across metazoans (Brattig-Correia et al., 2024), we functionally
132 characterize the reproductive functions of genes with relaxed selection in gorillas using a
133 high throughput *Drosophila* screen and identify 41 new spermatogenesis genes. Finally,
134 we show that the human orthologs of gorilla relaxed genes are enriched for loss-of-
135 function variants in men with severe spermatogenic impairment. These data illustrate the
136 complex relationship between life history traits, such as mating system type and form of
137 sexual selection, and the evolution of reproduction-related genes. Furthermore, the deep
138 conservation of genes with spermatogenic functions across species suggests that genes
139 under relaxed selection intensity in gorillas can reveal unknown aspects of male germ cell
140 development in general and male infertility in humans.

141

142 **Results**

143 **Assembling alignments of orthologous coding genes from 261 Eutherians**

144 We used a reciprocal best BLAT hit (RBBH) approach to assemble a dataset of
145 orthologous coding gene alignments from the genomes of 261 Eutherian ("Placental")
146 mammals (**Figure 1A**). We began by using RBBH with query sequences from human
147 CDSs in the Ensembl v99 human coding sequence dataset and BLAT searching the

genomes of 260 other Eutherians, followed by searching the top hit to the human query back to the human genome; we used BLAT matching all possible reading frames, with a minimum identity set at 30% and the "fine" option activated (Kent, 2002), and excluded genes with fewer than 251 best reciprocal hits out of the 261 (human+other mammals) species included in the analysis. Orthologous genes were aligned with Macse v2 (Ranwez et al., 2018), a codon-aware multiple sequence aligner that identifies frameshifts and readjusts reading frames accordingly. Alignments generated by Macse v2 were edited by HMMcleaner with default parameters (Di Franco et al., 2019) to remove species-specific substitutions that are likely genome sequencing errors and "false exons" that might have been introduced during the Blat search. Finally, we excluded incomplete codons and the flanks of indels, which usually have more misaligned substitutions. We thus generated a dataset of 13,491 orthologous coding gene alignments from the genomes of 261 Eutherian mammals, corresponding to 62.7% of all protein-coding genes in the gorilla genome. Of the 13,491 alignments, 13,310 had an identifiable HUGO gene symbol and were used in all subsequent analyses.

Few genes experienced episodic positive selection in gorillas

We first used ABSREL to identify genes with evidence of positive selection in the gorilla lineage (**Figure 1B**). The ABSREL model is a variant of the branch-sites random effect likelihood (BSREL) model of coding sequence evolution (Kosakovsky Pond et al., 2011; Smith et al., 2015), which allows for variation in d_N/d_S (ω) rates across lineages and sites, selects the optimal number of rate categories for each gene, and accommodates synonymous rate variation across sites [S] (Pond and Muse, 2005; Wisotsky et al., 2020), multi-nucleotide mutations per codon [MH] (Lucaci et al., 2021), and both synonymous rate variation and multi-nucleotide mutations [SMH]. Positive selection is inferred for a gene when the c-AIC selected best-fitting model includes a class of sites with $\omega>1$ and a likelihood ratio test $P\leq0.05$ compared to a null model that does not allow for $\omega>1$ (**Figure 1B**). The base ABSREL model identified 240 genes with a class of sites with $\omega>1$ in the gorilla lineage. However, after selecting the best [S], [MH], and [SMH] model, only 96 genes were identified with a class of sites with $\omega>1$ at $P\leq0.05$ (**Figure 1C**). Of these 96 genes, 26 were positively selected only in the gorilla lineage (**Figure 1D**). Gene ontology

178 enrichment analysis did not identify any statistically significant over-representation of GO
179 terms or pathways (FDR q-value≤1), which likely reflects low statistical power associated
180 with a small gene set.

181 **Hundreds of genes experienced relaxed selection intensity in gorillas**

182 We used RELAX to identify genes with evidence of relaxed selection intensity in
183 the gorilla lineage (**Figure 1E**). The RELAX model is also a variant of the BSREL model of
184 coding sequence evolution, which includes a selection intensity parameter (K) that scales
185 the distribution of d_N/d_S rate categories (ω_1^k, ω_2^k, ω_3^k). When selection is relaxed, d_N/d_S
186 rates move toward one and/or the proportion of sites increases in rate classes with d_N/d_S
187 values close to 1 (Wertheim et al., 2015). Relaxed selection intensity is inferred for a gene
188 when it includes a class of sites with K<1 in the foreground lineage compared to
189 background branches with a likelihood ratio test P≤0.05 (**Figure 1E**). In contrast,
190 increased selection intensity is inferred for a gene when it includes a class of sites with
191 K>1 at P≤0.05. We thus identified 578 genes with relaxed selection intensity (K<1) and
192 144 genes with increased selection intensity (K>1) in the gorilla lineage (henceforth
193 referred to as K<1 and K>1 genes, respectively, **Figure 1F**). K≠1 genes had slightly longer
194 coding sequence lengths, transcript lengths, and genome spans but similar GC content
195 as the background gene set; this indicates that the RELAX test likely has more statistical
196 power to detect deviation from K≈1 in longer genes (**Figure 1 – figure supplement 1**).
197 The 4-fold disparity between K<1 and K>1 genes is statistically significant (Yates'
198 corrected Chi-square $P<1.0\times10^{-4}$), indicating more genes have experienced an episode
199 of relaxed than intensified selection in the gorilla lineage.

200 To infer whether the altered selection intensity on these 722 genes is unique to
201 gorillas or also occurs in other mammalian lineages, we ran a version of RELAX (RELAX-
202 Scan) that iteratively tests for K≠1 on each branch compared to all other branches
203 (Kosakovsky Pond, 2021). Of the 578 genes with gorilla K<1, 32.7% (189/578) was K<1

204 in at least one other lineage but never K>1, 50% (289/578) were K<1 and K>1 in at least
205 two other lineages, and 13.3% (77/578) were K>1 in at least one other lineage but only
206 K<1 in gorilla; the remaining 23 genes failed to converge on parameter estimates in the
207 RELAX-Scan model. Similarly, of the 144 gorilla K>1 genes, 59% (85/144) were K<1 and
208 K>1 in at least one other lineage, 25% (36/144) were K<1 in at least one other lineage but
209 were only K>1 in gorilla, and 16% (23/144) were K>1 in at least one other lineage but
210 never were K<1 (**Figure 1G**). Thus, we conclude that while the vast majority of protein-
211 coding genes evolve under similar selection intensity in gorillas and other Eutherian
212 mammals (94.6%), a small percentage of genes in gorillas experienced either an episode
213 of relaxed (4.3%) or intensified (1.1%) selection.

214 **Gorilla K<1 genes harbor deleterious amino acid substitutions**

215 To explore the putative functional consequences of amino acid changes in genes
216 under relaxed purifying selection in the gorilla lineage, we reconstructed the ancestral
217 Homininae sequence for each gorilla K≠1 gene with IQTREE2 (Minh et al., 2020) and used
218 these ancestral sequences to identify 20,230 amino acid changes in the gorilla lineage
219 (**Figure 1A**). We then used a fixed effect likelihood (FEL) model to estimate the strength
220 and direction of selection (Kosakovsky Pond and Frost, 2005) acting on each codon site
221 (across mammals) in gorilla K≠1 genes. In the FEL model, codons with $d_N/d_S<1$ evolve
222 under selective constraints (i.e., selection against amino acid changes). Thus,
223 substitutions at these sites are likely deleterious. Concurrently, codons with $d_N/d_S>1$
224 evolve under diversifying selection (i.e., selection favors amino acid changes), and
225 codons with $d_N/d_S\approx1$ evolve under very weak or absent selective constraints. We found
226 that 75% (2,756/3,695) of gorilla-specific amino acid changes in gorilla K<1 genes
227 occurred at sites with $d_N/d_S<1$, whereas only 5% (187/3,695) occurred at sites with d_N/d_S
228 > 1, and 20% (752/3,695) occurred at sites with $d_N/d_S\approx1$ (**Figure 2A**). Similarly, 53%
229 (112/210) of gorilla-specific amino acid changes in genes with K>1 occurred at sites with
230 $d_N/d_S<1$, only 8% (17/210) occurred at sites with $d_N/d_S>1$, and 39% (81/210) occurred at

231 sites with $d_N/d_S \approx 1$ (**Figure 2A**). These data indicate that most amino acid changes in
232 gorilla K≠1 genes, especially those under relaxed selection intensity, occur at sites that
233 evolve under purifying selection and are thus likely deleterious.

234 To explore this possibility, we first compared the physicochemical properties of
235 gorilla-specific substitutions in K≠1 genes and found that many have negative BLOSUM
236 scores, which indicates they are relatively uncommon amino acid changes and likely
237 function-altering (**Figure 2B**). Next, we used PolyPhen-2 (Adzhubei et al., 2013) to infer
238 the functional consequences of amino acid substitutions in the gorilla lineage; PolyPhen-
239 2 predicts the functional impact of amino acid substitutions using structural data, the
240 physicochemical properties of amino acids, and comparative evolutionary data, and
241 assigns each substitution a score in the range of 0 (benign) to 1 (probably deleterious).
242 Out of the 20,753 gorilla-specific amino acid changes, 20,230 had PolyPhen-2
243 predictions. The mean PolyPhen-2 score of K≈1 genes was 0.21 (95% CI 0.21-0.22),
244 while the mean score for K<1 genes was 0.29 (95% CI 0.28-0.30), and for K>1 genes was
245 0.28 (95% CI 0.22-0.33; **Figure 2C**). There was a statistically significant difference in the
246 mean score between the K≈1 and K<1 sets (Holm adjusted Welch's *t*-test $P=1.22\times10^{-7}$)
247 and the K≈1 and K>1 sets (Holm adjusted Welch's *t*-test $P=0.09$). A larger proportion of
248 substitutions were predicted to be "probably damaging" for the K<1 gene set (13.38%,
249 450/3,253) compared to K≈1 genes (10.69%, 1,791/16,758; binomial test $P<1.00\times10^{-6}$),
250 but fewer substitutions were "probably damaging" for the K>1 gene set (4.57%, 10/219)
251 than for the K≈1 (binomial test $P=6.10\times10^{-4}$; **Figure 2D**). Thus, we conclude that amino
252 acid substitutions in gorilla K<1 genes show an increased propensity to be deleterious (or
253 at least functionally altering) compared to K>1 and K≈1 genes.

254 **Gorilla K<1 genes are enriched in sperm-related functions**

255 We next explored the possible functional significance of altered selection intensity
256 on gorilla genes. We started by determining which GO biological process terms were

257 enriched in the K≠1 gene set compared to the 13,310 background set using the over-
258 representation (ORA) test implemented in WebGestalt (Liao et al., 2019). We observed
259 that 29% (35/119) of terms enriched among genes with K<1 were related to reproduction,
260 meiosis, or sperm biology (**Figure 3A**). In contrast, only 2% (1/45) of terms enriched
261 among genes with K>1 were related to these processes. This 14.5-fold disparity is
262 statistically significant (one-sided Fisher's exact test P=4.0x10^{-4}), indicating a functional
263 bias in the type of genes that experience relaxed selection intensity in the gorilla lineage.
264 Among the GO terms related to spermatogenesis for which the K<1 genes are enriched
265 were "male gamete generation" (hypergeometric P=1.0x10^{-3}), "spermatogenesis"
266 (P=2.0x10^{-3}), "synaptonemal complex assembly" (P=3.0x10^{-3}), "spermatid development"
267 (P=3.0x10^{-3}), "meiotic chromosome condensation" (P=0.012), and "flagellated sperm
268 motility" (P=0.026), as well as "female meiosis chromosome segregation" (P=1.4x10^{-3}).
269 The latter may reflect the involvement of some K<1 genes in the core meiotic program
270 common to both sexes (Villeneuve and Hillers, 2001). Genes with K<1 were also enriched
271 in numerous GO cellular component terms related to the cytoskeleton, cilia,
272 centrosomes, septin ring complexes, and the cytoskeleton (**Figure 3B**). In contrast, K>1
273 genes were not enriched in cellular component terms related to these structures.

274 To further characterize the association of gorilla K≠1 genes with male reproductive
275 functions, we analyzed their expression pattern across different cell types and tissues.
276 Since comprehensive gene expression data are not available for most gorilla organ
277 systems, we took advantage of the fact that humans and gorillas diverged recently (Scally
278 et al., 2012) to use human transcriptomic datasets as a proxy for the expression pattern
279 of their gorilla orthologs. Specifically, we used WebCSEA (Dai et al., 2022) to test for
280 enriched expression of the human orthologs of the K≠1 genes across a panel of 111
281 single-cell RNA-seq datasets from 61 different somatic tissues representative of various
282 human organ systems except testis, which is not included in the WebCSEA dataset. While
283 the expression of K≠1 genes was enriched in numerous tissues, the most notable case

284 was the prostate, where the expression of the K<1, but not of the K>1 genes, was
285 significantly enriched, particularly in the prostate parenchyma cells (**Figure 3C**). Given the
286 high enrichment scores observed in the seminal fluid-producing epithelial cells of the
287 prostate, it is likely that the enriched expression of K<1 genes in this organ has a
288 functional impact on the level of male reproductive fitness. Next, we used a published
289 gorilla testis single nucleus RNA-Seq dataset (Murat et al., 2023) to explore the
290 expression pattern of K≠1 genes. We observed that the expression of K<1 genes was
291 enriched both in meiotic (leptotene, zygotene, and pachytene spermatocytes) and post-
292 meiotic (early and late round spermatids) cells at FDR q-values≤0.10 (**Figure 3D**). This
293 observation is particularly noteworthy as, in the mammalian germ line, the expression of
294 spermatogenesis genes is activated from meiosis onwards (Alavattam et al., 2019;
295 Maezawa et al., 2020).

296 Finally, we tested if proteins encoded by gorilla K≠1 genes were enriched in
297 proteome datasets generated from seminal plasma (Wu et al., 2019), whole sperm cells
298 (Wang et al., 2016), sperm tails (Amaral et al., 2013), sperm nuclei (de Mateo et al., 2011),
299 the soluble and insoluble fractions of sperm chromatin (Castillo et al., 2014), sperm
300 perinuclear theca (Zhang et al., 2022), sperm acrosomal matrix (Guyonnet et al., 2012),
301 sperm centrioles (Firat-Karalar et al., 2014), and accessory structures of the sperm
302 flagellum including the fibrous sheath, outer dense fibers, and mitochondrial sheath (Cao
303 et al., 2006). We found that proteins encoded by K<1 genes were significantly enriched
304 in sperm cells, either in whole cells or in several of its subcellular structures (**Figure 3E**).
305 The highest enrichment was found in the annulus, an essential element for the correct
306 function and organization of the sperm tail (Lehti and Sironen, 2017). The annulus is
307 supported by a core complex of septin ring units (Kuo et al., 2015), one of the enriched
308 terms found in the GO analysis of the K<1 genes (**Figure 3A-B**). Consistent with our
309 observation that the expression of K<1 genes was enriched in the prostate, we also
310 detected an overrepresentation of proteins encoded by the K<1 genes in the seminal

311 plasma (FDR q-value≤0.10). In contrast, proteins encoded by the K>1 genes were
312 significantly enriched only in whole sperm, the perinuclear theca, and the sperm tail (FDR
313 q-value≤0.10 for all cases; **Figure 3E**). These data indicate that genes with relaxed
314 selection intensity in the gorilla lineage are preferentially involved in male reproductive
315 functions, tend to be expressed at the meiotic and post-meiotic germ cell stages, and
316 are particularly enriched in multiple components of mature sperm and the seminal
317 plasma.

318 **Gorilla K<1 genes are associated with male reproductive impairment in mice**

319 We next tested if gorilla K<1 genes harboring putatively deleterious amino acid
320 substitutions have been previously associated with male reproductive phenotypes. As
321 we observed for all K<1 genes, the subset defined only by those with "probably
322 damaging" substitutions was enriched in sperm-related GO terms (**Figure 4A**). Gorilla
323 K<1 genes with "probably damaging" substitutions were enriched in many mouse
324 knockout phenotypes related to sperm biology annotated in the Mouse Genome
325 Database (Blake et al., 2021). Multiple phenotypes associated with sperm motility were
326 significantly enriched, as were infertility-associated defects such as asthenozoospermia,
327 teratozoospermia, and morphological abnormalities of the sperm tail (**Figure 4B**). The
328 expression of gorilla K<1 genes with "probably damaging" substitutions was particularly
329 enriched in meiotic cells (**Figure 4C**), and their proteins were over-represented (FDR q-
330 values≤0.10) in male gametes (**Figure 4D**). Curiously, this subset of K<1 genes was also
331 enriched in mouse knockout phenotypes related to the assembly and function of cell
332 protrusions across different developmental contexts, such as defects in hair cells and
333 enterocytes, which are characterized by extensive apical cell surface specializations.
334 These data highlight the functional significance of gorilla K<1 genes for male reproductive
335 function and suggest that deleterious mutations in these genes may cause at least some
336 of the adverse sperm traits in gorillas.

337 **Gorilla K<1 genes have evolutionarily conserved functions in spermatogenesis**

338 Our observation that gorilla K<1 genes are enriched for sperm- and testis-related
339 functions suggests that they may contain other, yet uncharacterized, spermatogenesis

340 genes that can broaden our understanding of the genetic basis of male germ cell
341 development. Unfortunately, functionally validating this hypothesis is obviously not
342 possible using traditional forward and reverse genetic approaches in gorillas, and the
343 generation and functional characterization of large numbers of mouse knockouts is not
344 feasible. Therefore, we harnessed the substantial evolutionary conservation of the
345 genetic program of spermatogenesis (Brattig-Correia et al., 2024; Murat et al., 2023;
346 Shami et al., 2020) and used a high throughput genetic screen in *Drosophila*
347 *melanogaster* to determine the reproductive functions of the fruit fly orthologs of gorilla
348 K<1 genes with still unknown roles in spermatogenesis.

349 We first identified which gorilla K<1 genes had previously been associated with
350 male fertility/spermatogenesis in humans, mice, or fruit flies. By using data from a
351 systematic review of validated monogenic causes of human male infertility (Houston et
352 al., 2021) and from the MGD (Blake et al., 2021) and Flybase (Larkin et al., 2021)
353 repositories, we identified 75 gorilla K<1 genes that were previously associated with male
354 reproductive impairment (**Supplementary dataset 1**). Given the widespread
355 transcriptional activity of male germ cells (Brattig-Correia et al., 2024) and its expected
356 contribution to increased genetic redundancy in the spermatogenic program, we
357 employed a functionality-oriented approach to associate a gene with male reproduction.
358 More specifically, all associations required phenotypical evidence linking the gene to
359 abnormal male reproductive capacity and/or some degree of spermatogenic impairment.
360 We acknowledge that the selected species represent a very small subset of the diversity
361 found in metazoan lineages. However, they are also some of the most extensively studied
362 in the context of reproductive genetics. Hence, they likely contain a sizable fraction of all
363 known genotype-phenotype correlations in male reproduction.

364 To determine if the remaining 503 genes had conserved functions in
365 gametogenesis, we used a cross-species comparative transcriptomics platform (Brattig-
366 Correia et al., 2024) to identify which of the gorilla K<1 genes had an ortholog expressed
367 in the *Drosophila* testis. This led to the identification of 144 gorilla K<1 genes without any
368 previous association with male reproductive function in any species, that were also
369 expressed in the *Drosophila* testis (**Supplementary dataset 1**). We then used *Drosophila*
370 *in vivo* RNAi to silence each gene specifically from the onset of meiotic entry onwards
371 and measured the functional consequences of this silencing for male fertility. Specifically,

372 we used the GAL4/UAS system, under the control of the *bam*-GAL4 driver (White-
373 Cooper, 2012), to ensure the germ cell-specific silencing of each gene in a developmental
374 time frame equivalent to that observed in gorilla spermatogenesis (i.e., expression
375 enrichment during the meiotic and post-meiotic stages). Of 144 tested genes
376 (corresponding to 156 *Drosophila* orthologs), we identified 41 (43 *Drosophila* orthologs)
377 that were associated with significantly decreased male reproductive fitness upon mating
378 gene-silenced males with wildtype females (average egg hatching rates below the cut-
379 off of 75%: >2 standard deviations of the mean observed in negative controls:
380 94.2±4.8%; **Figure 5A**). The frequency of K<1 genes associated with decreased male
381 fitness in this screen (27.6%) was 2.9-fold higher than the association with male fertility
382 and/or spermatogenesis observed in a random selection of 156 *Drosophila* genes (9.6%),
383 and 2.4-fold higher to a similarly random set of testis-expressed Drosophila genes
384 (11.5%), both according to Flybase gene annotation data (see **Methods**). We thus
385 conclude that gorilla K<1 genes are significantly enriched in male reproductive functions
386 in *Drosophila* (Chi-square $P<3.6\times10^{-4}$).

387 Next, we used phase-contrast microscopy to identify the earliest developmental
388 stage affected by the silencing of the 41 newly identified spermatogenesis genes (either
389 pre-meiotic, meiotic, post-meiotic, or cytologically undetectable; **Figure 5B**). By merging
390 this information with published data on the human, mouse, and fruit fly orthologs of the
391 75 previously described spermatogenesis genes, we observed that 52.6% (61/116) of
392 gorilla K<1 genes with validated functions in spermatogenesis were required for post-
393 meiotic development, i.e., spermatid development or mature gamete function (**Figure**
394 **5C**). These observations suggest that deleterious mutations in these 116 K<1 genes may
395 be an important factor for the decreased number of post-meiotic cells in the gorilla testis
396 (Fujii-Hanamoto et al., 2011) and for the high proportion of abnormal sperm cells in this
397 species (Martinez and Garcia, 2020; Platz et al., 1980; Seuanez et al., 1977). Furthermore,
398 they identify 41 novel evolutionarily conserved spermatogenesis genes that expand our
399 understanding of the male germ line genetic program.

400 **Gorilla K<1 genes are associated with human male infertility**

401 Our observations that gorilla K<1 genes are enriched in male reproductive
402 functions, particularly related to sperm biology, suggest that they may also be involved

403 in human male infertility. To explore this possibility, we performed a literature review and
404 identified 24 gorilla K<1 genes (**Table 1**) that are associated with a broad spectrum of
405 human phenotypes, such as sperm head defects, oligoasthenoteratozoospermia, and
406 azoospermia, among others (Kosova et al., 2012, 2014; Houston et al., 2021; Wang et al.,
407 2023; Dacheux et al., 2023). Next, we used Fisher's exact test (FET) to determine if human
408 orthologs of gorilla K<1 genes (*n*=568) were enriched with predicted loss-of-function
409 (LoF) variants in the Male Reproductive Genomics (MERGE) cohort (Wyrwoll et al., 2020),
410 which includes 2,100 infertile men with azoospermia (no sperm in the ejaculate, 69%),
411 cryptozoospermia (very few sperm in the ejaculate only identified after centrifugation,
412 22%), or severe oligozoospermia (<5 million/mL sperm concentration, 9%), compared to
413 gnomAD (Karczewski et al., 2020) as proxy for a cohort of fertile men. The average LoF
414 enrichment *P*-value for K<1 genes was significantly lower than 568 randomly selected
415 K≈1 genes (FET *P*=0.58 vs. *P*=0.67; Mann-Whitney *U* test $P=9.98\times10^{-5}$; **Figure 6A**). In
416 contrast, the mean enrichment *P*-value for non-LoF missense variants between K<1 and
417 the 568 randomly selected K≈1 genes was the same (FET *P*=0.58; **Figure 6A**), indicating
418 that the burden of predicted LoF variants in infertile men is greater in K<1 than K≈1 genes.
419 Indeed, the QQ-plot of expected *vs.* observed *P*-values from the burden test suggests
420 K<1 genes are enriched in low *P*-values compared to the 568 randomly selected K≈1
421 genes (**Figure 6B**), and that there is no evidence of systematic biases (such as population
422 structure, errors in analytical approach, genotyping artifacts, etc.) in the MERGE cohort
423 leading to erroneously inflated *P*-values.

424 In total, 109 K<1 genes were significantly enriched for LoF variants (FET *P*<0.05),
425 indicating these genes are likely associated with human male infertility. In contrast, 84 of
426 the 568 randomly selected K≈1 genes had a significant enrichment of LoF variants; this
427 1.3-fold disparity in the burden of LoF mutations is statistically significant (Chi-square
428 *P*=0.048). Conversely, the number of genes with no enrichment of predicted LoF variants
429 (FET *P*>0.95) was 267 for K<1 genes and 334 for K≈1 genes; this 1.25-fold disparity is
430 also statistically significant (Chi-square $P=6.8\times10^{-5}$). Of note, while the number of genes
431 significantly enriched in non-LoF missense variants was higher for K<1 (84) than K≈1 (68)

genes, the difference was not statistically significant (Chi-square P=0.16; **Figure 6A**). These data indicate that K<1 genes harbor more LoF variants, but not missense variants, than K≈1 genes in infertile men from the MERGE cohort.

To further explore if the results of the burden tests might be related to population-specific differences in the frequency of deleterious mutations, particularly of rare variants, we repeated the burden test of the gorilla K<1 *vs.* K≈1 genes selecting only those individuals from the MERGE cohort that were also scored by EthSEQ (Romanel et al., 2017) to be a near 100% match for European ethnicity (1,344 infertile men). This population was compared to 31,805 European individuals from gnomAD (excluding Finnish individuals). This analysis confirmed the previously observed higher burden of predicted LoF variants in infertile men in K<1 than K≈1 genes (FET P=0.66 vs. P=0.74; Mann-Whitney U test P=0.0001). The mean enrichment P-value for non-LoF missense variants between K<1 and the K≈1 genes deviated slightly between the two gene sets, and while effect size differences were small they are statistically significant (FET P=0.83 vs. P=0.80; Mann-Whitney U test P=0.02).

To investigate the extent to which K<1 genes are constrained within humans, we compared gnomAD gene-level constraint metrics for the human orthologs of the gorilla K<1 and K≈1 genes, including the probability of loss-of-function intolerance (pLI), and missense and synonymous Z-scores. pLI scores closer to one indicate increased intolerance to LoF variants, whereas more positive and negative Z-scores indicate increased constraint (fewer variants than expected) and relaxed constraint (more variants than expected), respectively. We found statistically significant differences between pLI (Welch's t-test P=0.01) and missense Z-scores (Welch's t-test P=0.02) for the human orthologs of gorilla K<1 and K≈1 genes (**Figure 6C**). Although the effect size differences were small, Hedge's $\hat{g}$=0.11 (95% CI 0.02-0.19) for pLI and Hedge's $\hat{g}$=0.10 (95% CI 0.02-0.19) for missense Z-scores suggest that K<1 genes are less constrained than K≈1 genes in humans. In contrast, there was no difference between synonymous Z-scores for

K<1 and K≈1 genes (Hedge's $\hat{g}$=0.03, Welch's t-test P=0.52), as expected for mutations that are unlikely to alter the function of protein-coding genes (**Figure 6C**).

These data suggest that human orthologs of gorilla K<1 genes harbor more amino acid-changing variants than K≈1 genes. To explore this possibility further, we used a variant of the McDonald-Krietman (MK) test (ABC-MK) that accounts for background selection (i.e., weak selection at linked sites) to jointly estimate the strength and rate of adaptation (α) of human genes (Murga-Moreno et al., 2023; Uricchio et al., 2019). The latter were estimated from 661 human genomes with African ancestry at the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015). We found that while the α of K≈1 genes was similar to previous estimates for human protein-coding genes (Uricchio et al., 2019), the human orthologs of the gorilla K<1 genes had a significantly lower α than K≈1 genes, suggestive of particularly weak adaptation (**Appendix Figure 1A**). These data are consistent with both recent balancing selection, which would inflate the frequency of nonsynonymous polymorphisms (P_N) and therefore decrease α, and recent relaxed selection, which would also inflate P_N/P_S compared to the long-term P_N/P_S that shapes d_N/d_S, acting on the human orthologs of gorilla K<1 genes (for a more detailed discussion of these results see **Appendix**). Thus, we conclude that the human orthologs of the gorilla K<1 genes are characterized by decreased adaptation and reduced constraint and are enriched for LoF variants in infertile men.

Discussion

Sexual selection can profoundly affect many organismal traits, including the evolution of extreme morphologies and behaviors such as long eyestalks in stalk-eyed flies (Diopsidae) and elaborately choreographed mating rituals in birds-of-paradise (Paradisaeidae). The evolution of polygyny in gorillas, a life history strategy in which a single male has nearly exclusive mating access and sires the vast majority of offspring (Bradley et al., 2005; Inoue et al., 2013; Nsubuga et al., 2008), occurred coincident with the origin of large bodies and the majority of male-male competition taking place before mating. This change in the dynamics of sexual selection is associated with a near

488 absence of sperm competition and dramatic anatomical changes in the male
489 reproductive system and sperm biology (**Box 1**), which have likely affected the genes that
490 underlie male reproductive traits. For example, several genes with male reproductive
491 functions have become nonfunctional (pseudogenes), including genes involved in the
492 histone-protamine transition (*H3t*), semen coagulation and liquefaction (*TGM4*, *KLK2*,
493 *SEMG1,* and *SEMG2*) (Carnahan and Jensen-Seaman, 2008; Clark and Swanson, 2005;
494 Jensen-Seaman and Li, 2003), and the formation of germ cell intercellular bridges (*TEX14*)
495 (Greenbaum et al., 2006; Krausz et al., 2020; Scally et al., 2012). Genes previously shown
496 to have accelerated rates of evolution in the gorilla lineage were also enriched for several
497 GO terms related to developmental processes, including male gonad development
498 (Scally et al., 2012).

499 These previous studies, however, focused on specific genes (Carnahan and
500 Jensen-Seaman, 2008; Clark and Swanson, 2005; Jensen-Seaman and Li, 2003) or used
501 models of molecular evolution that were not designed specifically to test for relaxed
502 selection intensity (Scally et al., 2012). Accelerated evolution, for example, can result from
503 both positive and relaxed selection. An expectation of relaxed selection intensity is a
504 decrease in the strength of selection acting on sites that, in other lineages, experience
505 both directional or diversifying selection and purifying selection (**Figure 1B/E**). We used
506 an unbiased, gene-wide approach to test whether the shift in male-male competition from
507 sperm competition to social dynamics was associated with changes in the direction and
508 intensity of selection acting on genes in the gorilla genome using methods specifically
509 designed to test for positive selection (ABSREL) and relaxed selection intensity (RELAX).
510 We identified 574 genes, 4.3% of all protein-coding alignments tested, that experienced
511 an episode of relaxed selection intensity in the gorilla lineage. Consistent with the shift in
512 male-male competition strategy, the expression of these genes is enriched in the testis,
513 particularly in male germ cells and in cells of the prostate that produce seminal plasma.

514 Enrichment tests depend on previously annotated gene functions and can only
515 identify known gene-function associations. Thus, they cannot discover new biology. To
516 overcome this limitation, we took advantage of the evolutionary conservation of the
517 spermatogenic program across metazoans to explore the reproductive functions of
518 gorilla K<1 genes in a high throughput genetic screen in *Drosophila melanogaster*. We
519 found that fly orthologs of 144 gorilla K<1 genes were expressed in *Drosophila* testis, of

520 which 75 had been previously functionally associated with male reproductive impairment
521 and 41 were discovered as new male fertility genes in our screen. Approximately half of
522 the 116 gorilla K<1 genes with functions in spermatogenesis (61/116; 52.6%) were
523 required for post-meiotic processes such as spermatid development and mature gamete
524 function, providing a likely functional basis for the particularly weak spermatogenic
525 output of gorillas. We also transformed humans into a model organism to study gorilla
526 biology by using the Male Reproductive Genomics (MERGE) cohort, which includes 2,100
527 infertile men with 0–5 million/mL sperm concentration (69% azoospermia, 22%
528 cryptozoospermia, 9% severe oligozoospermia), to identify K<1 genes associated with
529 male infertility. Consistent with a reproductive function in both humans and gorillas, we
530 found that the human orthologs of the gorilla K<1 genes had an increased burden of loss-
531 of-function (LoF) variants in the MERGE cohort: K<1 genes were 1.3-fold enriched for
532 genes associated with human male infertility compared to K≈1 genes. These data suggest
533 that (deleterious) amino acid changes in K<1 genes are similarly associated with low
534 sperm counts in gorillas.

535 Despite the remarkably high prevalence of male infertility, which affects 5-7% of
536 couples worldwide (Cox et al., 2022), only 30% of cases can be attributed to a specific
537 genetic cause (Tüttelmann et al., 2018). Male infertility is a complex, polygenic trait with
538 diverse genetic causes that manifest as a similar phenotype, contributing to the difficulty
539 in determining its origins (O'Flynn O'Brien et al., 2010; Shah et al., 2003). In addition to
540 identifying genes likely associated with adverse male reproductive traits in gorillas, our
541 data provide compelling candidate genes for human male infertility. Indeed, we identified
542 24 gorilla K<1 genes associated with a broad spectrum of abnormal sperm traits and
543 infertility in humans (**Table 1**), suggesting that other gorilla K<1 genes will be associated
544 with human male infertility. Conversely, mutations in many of these genes cause multiple
545 morphological abnormalities of the sperm flagella, leading to immotile sperm. This
546 suggests that gorilla-specific amino acid changes in these genes may underlie poor
547 sperm motility in gorillas.

548 Furthermore, genome-wide association studies have implicated gorilla K<1 genes
549 with sperm motility in Italian Holstein (*ATP13A2, CPLANE2, ZNF800, LCORL, PACC1,*
550 *EFHC1*) (Ramirez-Diaz et al., 2023) and Holstein-Friesian bulls (*CORIN, ARFIP1*) (Abril-
551 Parreño et al., 2023), semen volume in Assaf breed sheep (*SCAPER*) (Serrano et al., 2021)
552 and Holstein-Friesian bulls (*DCP1A*) (Hering et al., 2014), and sperm number and motility
553 in Holstein bulls (*DMRT3, ANKRD29*) (Yin et al., 2019). Collectively, these data are strong
554 circumstantial evidence that the K<1 genes have reproductive functions in gorillas and
555 that the accumulation of deleterious amino acid substitutions in these genes is at least
556 partly responsible for the decreased number of post-meiotic cells in the gorilla testis
557 (Fujii-Hanamoto et al., 2011), and the low sperm count and high proportion of abnormal
558 sperm in this species (Martinez and Garcia, 2020; Platz et al., 1980; Seuanez et al., 1977),
559 among other derived traits in the reproductive system that reduce male reproductive
560 function (**Box 1**).

561 Polygyny and reduced male post-copulatory sexual selection are found among
562 many species. Reduced sperm competition in Old World mice and rats (Murinae) is
563 associated with smaller testis mass, sperm morphologies, and signals of relaxed purifying
564 selection on protein-coding genes expressed in the late stages of spermatogenesis and
565 seminal vesicles (Kopania et al., 2025; Murat et al., 2023). These results are consistent
566 not only with our observations in gorilla, but also with studies of other mouse lineages
567 (Good and Nachman, 2005; Kopania et al., 2022), a broad selection of mammals (Murat
568 et al., 2023), and even among promiscuous Tanganyikan cichlids (*Ophthalmotilapia*
569 *ventralis*) (Morita et al., 2023). Thus, the pattern of relaxed selection intensity on genes
570 involved in male reproductive function and the associated accumulation of deleterious
571 amino acid changes we observed in gorillas may be generalizable across species with
572 reduced, or nearly absent, sperm competition.

573

574 **Ideas and speculation**

575 The absence of strong sexual selection on male reproductive functions in gorillas
576 is associated with an accumulation of numerous traits that decrease the likelihood of their
577 reproductive success (**Box 1**), including very low sperm counts and a high proportion of
578 (nearly) immotile and abnormal sperm, traits that should be counterbalanced by natural

579 selection acting to increase reproductive fitness. Our data is compatible with a scenario
580 where a near absence of sperm competition in gorillas dramatically relaxed selection on
581 genes involved in male reproduction. It is possible that the process of relaxed selection
582 has not yet reached its limits and may spread to additional genes or become even more
583 pronounced in others leading to their complete loss (pseudogenization). However,
584 additional studies are needed to infer direction and causation, as well as to characterize
585 gene losses and the ongoing selective pressures acting on gorilla genomes.

586 This study raises several interesting questions: Does the female reproductive tract
587 compensate for reduced sperm function in gorillas? While sperm are propelled by their
588 flagella, they do so with the active assistance of the female reproductive tract (Suarez
589 and Pacey, 2006; Tung and Suarez, 2021). Oviductal cilia, for example, create a fluid flow
590 that controls sperm migration to the site of fertilization by binding and holding sperm until
591 the periovulatory period, at which point they are released (Suarez, 2016; Suarez and
592 Wolfner, 2021; Yuan et al., 2021). In addition, peristalsis of oviductal muscles actively
593 moves sperm up to the ampulla, while ciliated cells that line the cervical canal likely
594 generate a fluid flow against which sperm can orient to guide their migration into the
595 uterus (Hino and Yanagimachi, 2019; Ishikawa et al., 2016; Mullins and Saacke, 1989;
596 Suarez, 1987). The composition of cervical mucus, particularly sialic acid and *O*-glycans
597 which bind sperm and reduce their ability to traverse the cervix, also affects sperm
598 transport (Abril-Parreño et al., 2022; Richardson et al., 2019). Compensatory adaptations
599 in these and other functions of the female reproductive tract may counteract the slow
600 swimming speed and weak swimming force of gorilla sperm (Nascimento et al., 2008)
601 and thus improve the likelihood of fertilization.

602 Second, is the erosion of genes with male reproductive functions
603 counterbalanced by the pleiotropic functions of these genes in other contexts? Or,
604 conversely, do the pleiotropic functions of K<1 genes impose additional costs beyond
605 reduced male reproductive function? Most genes are expressed at multiple
606 developmental stages and in multiple tissues (and thus may be functional in multiple
607 contexts), and many genes have more than one distinct function. These data suggest
608 that gorilla K<1 genes cannot be entirely free from purifying selection because they have
609 pleiotropic functions outside the male reproductive system and that relaxed selection on
610 their male reproductive functions may have effects (costs) in other contexts. For example,

611 K<1 genes are enriched in phenotypes not related to male reproduction, including those
612 associated with hair cells, which are the primary sensory receptor cells within the inner
613 ear that are essential for the perception of sound, and female meiosis, which likely reflects
614 that some of these genes function both in the male and female germ lines (Villeneuve and
615 Hillers, 2001). Thus, relaxed selection and the deleterious effects of amino acid changes
616 in gorilla K<1 genes may spread to other anatomic systems and female meiosis. These
617 pleiotropic effects may prevent the complete loss of purifying selection on K<1 genes or
618 limit the effects of deleterious amino acid changes in K<1 genes to residues that
619 contribute mostly to male reproductive function. It is also possible that the negative
620 consequences of deleterious pleiotropy become less pronounced at later stages of
621 spermatogenesis as meiotic and post-meiotically expressed genes are enriched for
622 testis-specific functions (Murat et al., 2023).

623 Finally, gorillas and humans share several sperm traits that are (likely) detrimental
624 to male reproductive fitness, including slow swimming speed and force and a high
625 proportion of abnormal sperm (**Box 1**). Similar to the signal of relaxed selection intensity
626 we observed in gorillas, human orthologs of gorilla K<1 genes harbor more amino acid
627 variants and have significantly lower ABC-MK α than K≈1 genes. These data are
628 consistent with recent balancing and/or relaxed selection; unfortunately, disentangling
629 these scenarios is extremely difficult. However, our observation that the human orthologs
630 of gorilla K<1 genes are significantly enriched in predicted LoF variants associated with
631 infertility in the Male Reproductive Genomics (MERGE) cohort suggests that balancing
632 selection, which maintains deleterious variants, is an unlikely scenario unless deleterious
633 variants are in linkage disequilibrium with adaptive variants. Thus, gorilla K<1 genes may
634 experience very recent relaxed selection intensity in the human lineage, so recent that
635 methods like the RELAX test cannot identify deviations from the null model of K≈1. If so,
636 human orthologs of gorilla K<1 genes are also likely to acquire deleterious variants that
637 reduce spermatogenic potential and male reproductive success.

638

639 **Caveats and limitations**

640 One of the major limitations of this study is that we cannot perform the kinds of
641 forward and reverse genetic methods used to establish genotype-phenotype
642 relationships and genetic causation that is standard in model organisms. For example,
643 we cannot introduce ancestral amino acid changes into gorilla genes and determine if
644 they improve sperm quality or reproductive success. Conversely, we cannot introduce
645 putatively deleterious amino acid changes from gorilla genes into a closely related
646 species and test for decreased sperm quality or reproductive success. Even if these
647 experiments were possible, male reproductive fitness and sperm quality are complex
648 polygenic traits, and the effect size of most evolutionarily relevant amino acid
649 substitutions is likely relatively small. Thus, introducing one or even a few putatively
650 deleterious amino acid changes into a closely related species or attempting to rescue
651 reproductive function with one or a few changes may not be sufficient to observe a small
652 effect on sperm function or male fitness. To circumvent this limitation, we have shown
653 that silencing *Drosophila* orthologs of gorilla K<1 genes in the testis has deleterious
654 consequences for male reproductive function and that human orthologs of gorilla K<1
655 genes are enriched in LoF mutations in infertile men from the MERGE cohort. These data
656 suggest that, in gorillas, relaxed selection in such genes may have adverse
657 consequences for male reproductive function. A potential solution to this limitation is
658 genome editing in iPSC and organoid models of spermatogenesis; however these
659 models, although promising, do not yet generate mature functional sperm *in vitro* (Easley
660 et al., 2012; Hayashi et al., 2011; Robinson et al., 2021).

661 Importantly, we cannot establish the direction of evolutionary and developmental
662 causation (Lewontin, 2002; Svensson, 2020; Wagner, 2000). Indeed, such inferences may
663 not be possible (Wagner, 2001). For example, did large bodies evolve before and lead to
664 a polygynous mating system and, therefore, reduced sperm competition? Did the
665 formation of female groups evolve before and lead to the evolution of a polygynous
666 mating system, large bodies, and reduced sperm competition? Did the development of
667 reduced sperm function lead to the origins of a polygynous mating system? Did these
668 events occur simultaneously, or in some other order? Theoretical models suggest a
669 central role for female choice and ecological factors in the evolution of polygynous mating
670 systems (Altmann et al., 1977; Ptak and Lachmann, 2003; Verner, 1964). Previous studies

671 in pinnipeds found that sexual size dimorphism likely evolved before polygynous mating
672 systems, either as a consequence of niche partitioning or in combination with sexual
673 selection on males to enforce copulations on females; with the origins of polygyny then
674 leading to intensified sexual selection (Cullen et al., 2014; Krüger et al., 2014).
675 Unfortunately, there is little data to disentangle these scenarios in gorillas. Gorillas also
676 have reduced effective population sizes, which can lead to an increased load of
677 deleterious mutations and relaxed selection intensity. However, we do not believe that it
678 substantially affects our observations. Indeed, relatively few genes have K<1 and those
679 are enriched in sperm biology. Given that a reduced effective population size will plausibly
680 increase the load of deleterious mutations and relaxed selection across many genes, it is
681 unlikely that such a broad phenomenon would result in a specific enrichment in genes
682 related to male reproductive biology.

683 Other limitations are technical. Methods to detect statistical evidence of positive
684 selection and relaxed selection intensity, such as the ABSREL and RELAX methods we
685 used, can accommodate synonymous rate variation across sites [S] (Pond and Muse,
686 2005; Wisotsky et al., 2020) and multi-nucleotide mutations per codon [MH] (Lucaci et
687 al., 2021), but assume synonymous substitutions are (nearly) neutral; while this is likely
688 true for most synonymous substitutions, some will have negative fitness effects (Chamary
689 et al., 2006; Niu et al., 2025; Zhang and Qian, 2025). Thus, synonymous substitutions that
690 are deleterious will decrease the synonymous substitution rate, inflating the d_N/d_S rate,
691 leading to false inferences of positive selection. We note, however, that the ABSREL and
692 RELAX methods are conservative and would generally require more than one or a few
693 non-neutral synonymous sites to falsely infer positive or relaxed selection. These
694 methods also test for selection in protein-coding genes, if some genes with K<1 are in
695 the process of pseudogenization they may retain function as transcripts. In addition, we
696 note that the use of a permutation test might provide stronger support to the reported
697 enrichment in LoF mutations in K<1 genes in infertile men from the MERGE cohort.

698 Another potential limitation of our approach is that we used a single species tree,
699 rather than gene trees. This could bias ancestral sequence reconstruction (ASR), as well
700 as inferences of positive and relaxed selection intensity, if the gene and species tree are
701 different. For example, previous studies found that ~30% of the genome in the human-
702 chimp-gorilla clade had evidence for incomplete lineage sorting (ILS) (Rivas-González et

703 al., 2023; Scally et al., 2012). However, we do not believe our results are affected by ILS
704 or gene-species tree discordance. In a previous analysis using the same dataset we used
705 here, we inferred gene trees with IQTREE and a species tree with ASTRAL; while that
706 study was focused on the relationships between elephants, hyraxes, and sea cows, the
707 analyses also found that 97.7% (13108/13388) of gene trees matched the species tree
708 we used. While this estimate is higher than one might expect, given the genome-wide
709 level of ILS, ILS is lower in genes than in non-genic regions of the genome (Scally et al.,
710 2012).

711

712 **Conclusions**

713 Sexual selection can take many forms and is associated with the evolution of
714 diverse behavioral, physiological, and morphological characteristics, including extreme
715 sexual size dimorphism, polygynous mating systems, and reduced sperm competition.
716 Here, we have shown that the polygynous mating system of gorillas is associated with
717 dramatic shifts in the direction and intensity of selection acting on genes related to the
718 development and function of the male reproductive system, including traits directly
719 related to sperm competition. While functionally characterizing the consequences of
720 relaxed purifying selection on gorilla genes is technically challenging, we have shown that
721 these genes have conserved functions in *Drosophila* testis and sperm biology and are
722 enriched for loss-of-function variants in infertile men. Thus, the accumulation of
723 deleterious mutations in genes associated with male reproductive function is likely
724 directly related to the low efficacy of gorilla spermatogenesis and may also underlie
725 human male infertility.

726

727 **Methods**

728 **Generation of multiple sequence alignments**

729 The multiple sequence alignment pipeline started with Ensembl v99 human
730 coding sequences, using the longest isoforms for each gene. We then searched for the
731 orthologs of these human coding sequences within the genome assemblies of 260 other

732 mammals with a contig size greater than 30kb present in the NCBI assembly database
733 as of July 2020. We selected a minimum 30kb contig size to avoid an excessive number
734 of truncated orthologous coding sequences. To extract the orthologous coding
735 sequences (CDS) to the human CDS, we used best Blat reciprocal hits from the human
736 CDS to each other mammalian genome, and back to the human genome. We used Blat
737 matching all possible reading frames, with a minimum identity set at 30% and the "fine"
738 option activated (Kent, 2002). We excluded genes with less than 251 best reciprocal hits
739 out of the 261 (human+other mammals) species included in the analysis. In total, we
740 found 13,491 human CDS genes with best reciprocal hits orthologs in at least 250 other
741 mammals.

742 We aligned each orthologous gene with Macse v2 (Ranwez et al., 2018). Macse
743 v2 is a codon-aware multiple sequence aligner that can explicitly identify frameshifts, and
744 readjusts reading frames accordingly. This feature of Macse is particularly important
745 because erroneous indels introduced in coding sequences during genome sequencing
746 and assembly process can be common and can cause frameshifts that many aligners do
747 not take into account. These sequencing and alignment errors can result in substantial
748 misalignments of coding sequences due to incomplete codons. Note that we started from
749 the human CDS because they are likely to be of the highest quality in terms of sequencing
750 and annotations. We used Macse v2 with maximum accuracy settings.

751 The alignments generated by Macse v2 were edited by HMMcleaner with default
752 parameters (Di Franco et al., 2019). HMMcleaner is designed to remove, in a species-
753 specific fashion, "false" substitutions that are likely genome sequencing errors.
754 HMMcleaner also removes "false exons" that might have been introduced during the Blat
755 search. "False exons" are intronic segments that by chance have some similarity with
756 exons that are missing from an assembly due to a sequencing gap. When looking for the
757 most similar non-human CDS using Blat, Blat can sometimes "replace" the missing exon
758 with a similar intronic segment.

759 After using HMMcleaner to remove suspicious parts of the CDS alignments in
760 each species, we only selected those codons that are still complete to remain in the
761 alignments. In alignments of coding sequences, the flanks of indels usually include a
762 higher number of misaligned substitutions. In each separate species, we further remove

763 from the alignments the x upstream or downstream codons if more than x/2 of these
764 codons code for amino acids that are different from the consensus amino acid in the
765 whole alignment, with x varying from 1 to 20. For example, if eight of the 12 amino acids
766 just on the left side of an indel are different from the consensus in a given species, we
767 remove the 12 corresponding codons in that species. If two of the three amino acids just
768 on the right side of an indel are different from the consensus in a given species, we
769 remove the three corresponding codons in that species. Visual inspection showed that
770 the 13,491 resulting orthologous CDS alignments were of very high quality, with a very
771 low number of visibly ambiguous or erroneous segments.

772 To build the species phylogenetic tree of the 261 selected mammals, we used the
773 third (4,444) of the 13,491 orthologous CDS alignments with the lowest GC content. CDS
774 with a high GC content have been shown to give poor tree consensus in mammals due
775 to a higher number of homoplasies than in AT-rich CDS, caused by more frequent Biased
776 Gene Conversion and CpG hypermutable sites (Romiguier et al., 2013). We used
777 IQTREE2 (Nguyen et al., 2015) to generate the consensus, maximum likelihood tree from
778 the 4,444 AT rich CDS, with a GTR substitution model with six parameters (GTR-6) for
779 each CDS (Kalyaanamoorthy et al., 2017; Nguyen et al., 2015; Romiguier et al., 2013).
780 The GTR-6 model was the best fit for more than 85% of the 4,444 genes in a preliminary
781 comparison of different substitution models.

782 **ABSREL branch-site selection tests**

783 Methods to infer the strength and direction of selection acting on molecular
784 sequences, such as those to identify pervasive and episodic adaptive evolution ("positive
785 selection"), are often based on estimating the ratio of non-synonymous (d_N) to
786 synonymous (d_S) substitution rates (d_N/d_S or ω) in an alignment of homologous genes.
787 These methods can be applied to entire genes or regions within genes (Hughes and Nei,
788 1988), sites (Nielsen and Yang, 1998), branches (Messier and Stewart, 1997; Muse and
789 Gaut, 1994; Yang, 1998), or sites along a specific branch - the latter often called "branch-
790 site" (BSREL) models (Kosakovsky Pond et al., 2011; Smith et al., 2015; Yang and
791 Nielsen, 2002). A key advantage of the BSREL models is that they allow for a group of
792 sites to evolve under the action of positive selection, i.e., with $\omega>1$, while the remaining
793 sites can evolve under purifying selection ($\omega<1$) or relatively neutrally ($\omega=1$) in specific

794 branches, and thus can detect positive selection that is episodic (limited to a subset of
795 branches) and restricted to a subset of sites in a gene. Positive selection is inferred when
796 a class of sites is identified with ω>1, with a LRT P-value ≤ 0.05. Subsequently, these
797 methods have been modified (ABSREL) to allow for the number of d_N/d_S rate categories
798 to be inferred from the alignment rather than imposed *a priori* (Smith et al., 2015), as well
799 as to accommodate both synonymous rate variation across sites [S] (Pond and Muse,
800 2005; Wisotsky et al., 2020) and multi-nucleotide mutations per codon [MH] (Lucaci et
801 al., 2021). We tested for selection using the base ABSREL model, a model that accounts
802 for synonymous rate variation across sites (ABSREL[S]), a model that accounts for multi-
803 nucleotide mutations per codon (ABSREL[MH]), and a model that accounts for both
804 synonymous rate variation across sites and multi-nucleotide mutations per codon
805 (ABSREL[SMH]). Positive selection is inferred when the best-fitting ABSREL model,
806 determined from a small sample AIC test, includes a class of sites with ω>1 at a likelihood
807 ratio test (LRT) P-value ≤ 0.05.

808 **RELAX selection tests**

809 Standard methods for estimating the strength and direction of selection acting on
810 molecular sequences based on estimating the ratio of non-synonymous (d_N) to
811 synonymous (d_S) substitution rates (d_N/d_S or ω) are not suitable for detecting relaxed
812 purifying selection because they lack power for this test and can mistake an increase in
813 the intensity of positive selection for relaxation of both purifying and positive selection
814 (Wertheim et al., 2015). For example, if most sites within a protein-coding gene
815 experience purifying selection ($d_N/d_S < 1$) while only a subset of sites experience an
816 episode of relaxation in the intensity of purifying selection ($d_N/d_S \approx 1$), the average d_N/d_S
817 across sites will almost always still be below one (Wertheim et al., 2015). A relaxation in
818 the intensity of selection may also be associated with the release of some sites from the
819 action of diversifying selection, shifting these sites from $d_N/d_S > 1$ to $d_N/d_S \leq 1$. Again, the
820 average d_N/d_S across sites will be below one, and the episode of reduced selection
821 intensity masked.

822 To overcome these limitations, we used a general hypothesis testing framework
823 for detecting episodic relaxed selection intensity that utilizes codon-based maximum-
824 likelihood models that are an extension of the BSREL model (RELAX) (Wertheim et al.,
825 2015). The RELAX method is a variant of the branch-site random effects likelihood (BS-
826 REL) model (Smith et al., 2015), which allows each lineage in a phylogeny to have a class
827 of sites that evolve near neutrality ($d_N/d_S \approx 1$), a class of sites that evolve under diversifying
828 selection ($d_N/d_S > 1$), and a class of sites that evolve under purifying selection ($d_N/d_S < 1$).
829 The RELAX test is based on the differential effects that reduced selection intensity is
830 expected to have on these site-specific d_N/d_S rates: When selection is relaxed, smaller
831 d_N/d_S values increase toward 1, whereas d_N/d_S values above 1 decrease (Wertheim et al.,
832 2015). In the context of BSREL models, this trend can have two different effects: The
833 d_N/d_S values inferred for the selection categories can move toward 1 and/or the
834 proportions of sites belonging to the different classes can change in such a way that
835 more sites are assigned to categories with d_N/d_S values closer to 1. To test whether
836 selection is relaxed (or intensified) on a subset of *a priori* defined branches relative to the
837 other branches, RELAX compares a null model in which a selection intensity parameter
838 (K) is set to 1 for all branches (equivalent to the standard BSREL mode with three rate
839 classes) to an alternative model in which the test branches share a K value inferred from
840 the data. A LRT using the standard $\chi 2$ asymptotic distribution with 1 degree of freedom
841 is used to determine significance: If the LRT rejects the null model, it indicates that
842 selection on the test branches is intensified ($K > 1$) or relaxed ($K < 1$) compared with the
843 reference branches (Wertheim et al., 2015). In our implementation of the RELAX test, the
844 gorilla lineage was defined as the foreground ("test") lineage and all other lineages were
845 assigned to the background ("reference") set.

846 **RELAX-Scan selection tests**

847 The standard RELAX method described above requires that foreground ("test")
848 lineages are defined *a priori*. The RELAX-Scan analysis modifies the RELAX method to
849 iteratively scan branches to identify individual branches with evidence of
850 relaxation/intensification relative to the gene background inferred from all other branches.
851 The RELAX-Scan test uses the General Descriptive model, where variation in d_N/d_S across

852 sites and branches is handled via a k-bin distribution ($\omega_1 \leq 1$, $\omega_2 \leq 1$, $\omega_3 \geq 1$). Each branch
853 gets an additional selection intensity parameter (K), which scales the gene distribution to
854 ω_1^k, ω_2^k, ω_3^k. Like the standard RELAX test, if K < 1, we infer that selection is relaxed on
855 branch b relative to gene average (shrunk towards $\omega = 1$), otherwise it is intensified (pulled
856 away from $\omega = 1$). After fitting the general exploratory model, RELAX-Scan fits a series of
857 null hypotheses (one per branch), where K=1 is enforced at a branch, and the significance
858 of K≠1 is tested with a LRT using the standard $\chi 2$ asymptotic distribution with 1 degree.
859 Note that in this test there is no *a priori* defined hypothesis to test and P-values are
860 corrected for multiple testing using the Holm-Bonferroni procedure. The tests can be
861 carried out using one of two approaches: "fast", in which all global model parameters are
862 fixed at their maximum likelihood estimated from the general exploratory model, and
863 "slow", in which all global model parameters are refitted during each branch-level test.
864 While the "slow" approach will more accurately estimate significance, it runs very slowly.
865 In contrast, the "fast" approach runs more quickly but will generally overestimate statistical
866 significance. We ran RELAX-Scan on the 722 genes that had significant K≠1 in the gorilla
867 lineage, testing for K≠1 in each of the other 521 branches using RELAX-Scan "slow"
868 mode.

869 **Fixed Effects Likelihood (FEL) sitewise selection tests**

870 For each gene with K≠1 in the gorilla lineage, we characterized the strength and
871 direction of selection acting on each codon site with a fixed effects likelihood model (FEL)
872 that included synonymous rate variation across sites (Kosakovsky Pond and Frost, 2005).
873 This model estimates separate d_N/d_S rates for each site in an alignment and quantifies

874 the strength of selection. We used these per site d_N/d_S rates to estimate the likely
875 functional consequences of amino acid substitutions at those sites, such that
876 substitutions at sites with $d_N/d_S < 1$ are likely to be function-altering and deleterious,
877 substitutions at sites with $d_N/d_S = 1$ are likely to be non-function altering, and substitutions
878 at sites with $d_N/d_S > 1$ are likely to be function altering and adaptive.

879 **Identification of gorilla-specific amino acid changes**

880 We used IQTREE2 (Nguyen et al., 2015) to reconstruct ancestral amino acid
881 sequences of all 13,310 genes to identify gorilla-specific amino acid changes. The input
882 was each gene's MSA (in fasta format), with the following options: for IQTREE2 to
883 translate nucleotide sequences to amino acid sequences (-st NT2AA), to use the species
884 phylogeny (-te species.tree) rather than infer a gene tree, to infer the best model of amino
885 acid substitution and rate variation across sites for each gene (-m TESTONLY)
886 (Kalyaanamoorthy et al., 2017), and to infer ancestral protein sequences (--ancestral).
887 Note that identical protein sequences were maintained in the analyses (--keep-ident), to
888 ensure consistent internal node numbering across genes. After the reconstructions for
889 each gene were completed, the predicted ancestral Homininae (AncHomininae) primate
890 (i.e., the last common ancestor of *Gorilla gorilla*, *Pan troglodytes*, *Pan paniscus*, and
891 *Homo sapiens*) of each gene was compared to the gorilla sequence to identify amino acid
892 changes in gorilla. We thus identified 20,753 amino acid changes in 13,310 proteins in
893 the gorilla lineage.

894 **PolyPhen-2 amino acid classification**

895 We used PolyPhen-2 (Adzhubei et al., 2013) to classify the functional
896 consequences of amino acid changes in the gorilla lineage using the AncHomininae
897 reconstructed protein sequences as input and "wildtype" amino acid, and the 20,753
898 amino acid changes in gorilla lineage as the "mutant" amino acid. The Classifier model
899 was set to "HumDiv", and the other parameters to the GRCh37/hg19 genome, canonical
900 transcripts, and missense annotations.

901 **Over Representation Analyses (ORA)**

902 We used WebGestalt v. 2019 (Liao et al., 2019) to identify enriched gene ontology
903 (Biological Process and Cellular Component) terms and human phenotype ontology

904 (Köhler et al., 2021), DisGeNET (Piñero et al., 2020), and GLAD4U (Jourquin et al., 2012)
905 terms using over-representation analysis (ORA). Statistical significance of over-
906 representation was assessed using the set of genes with K<1 or K>1 as the foreground
907 gene set, the 13,310 genes tested for relaxed selection as the background gene set, and
908 a hypergeometric test. The minimum number of genes for a category was set to 3 and
909 the maximum to 200, and the number of categories expected from set cover was 10.
910 False discovery of terms related to reproduction was controlled with the Benjamini-
911 Hochberg false discovery rate (FDR q-value).

912 **Mouse knockout phenotype enrichment**

913 We used the Gene List Analysis and Visualization (VLAD) tool
914 (http://proto.informatics.jax.org/prototypes/vlad/) to test if genes with K≠1 were enriched
915 in the Mouse Genome Informatics database (Bult et al., 2019). Genes with K<1 or K>1
916 and probably damaging amino acid substitutions were used as the foreground gene sets,
917 with the 13,491 genes tested for relaxed selection as the background gene set, and a
918 hypergeometric test using default settings.

919 **Gorilla testis cell-type enrichment**

920 We used WebGestalt v. 2019 (Liao et al., 2019) to test if genes with K≠1 were
921 enriched in specific cell-types in the gorilla testis. Genes with K<1 or K>1 were used as
922 the foreground gene set, with the 13,310 genes tested for relaxed selection as the
923 background gene set, and a hypergeometric test using the settings described above.
924 Gene expression data for specific cell-types in the gorilla testis were generated from a
925 previously published single-nucleus RNA-Seq dataset (Murat et al., 2023). Note that we
926 did not reanalyze the gorilla single nucleus RNA-Seq data and used gene expression
927 dataset generated from the original study (in Supplementary table 4). Readers are referred
928 to Murat et al., 2023 for specifics of the single nucleus RNA-Seq data analyses.

929 **Proteome enrichment**

930 We also used WebGestalt v. 2019 (Liao et al., 2019) to test if genes with K≠1 were
931 enriched in the either the human seminal plasma (Wu et al., 2019), human whole sperm
932 proteome (Wang et al., 2016), the human sperm tail proteome (Amaral et al., 2013), the
933 human sperm nucleus (de Mateo et al., 2011), isolated soluble and insoluble sperm
934 chromatin fractions (Castillo et al., 2014), the boar perinuclear theca (Zhang et al., 2022),
935 and the mouse sperm acrosomal matrix (Guyonnet et al., 2012). Note that we did not
936 reanalyze these proteomic datasets and used protein abundance estimates generated
937 from the original studies and available in the supplementary materials for each study.
938 Genes with K<1 or K>1 were used as the foreground gene set, with the 13,310 genes
939 tested for relaxed selection as the background gene set, and a hypergeometric test using
940 the settings described above.

941 **Drosophila *in vivo* RNAi screen**

942 The *Drosophila melanogaster* orthologs of the 578 gorilla K<1 genes were
943 identified based on eggNOG orthogroups (Huerta-Cepas et al., 2019). Orthologs were
944 selected for the RNAi screen based on two criteria. First, only orthologs lacking reported
945 evidence of being functionally associated with male fertility/spermatogenesis in humans,
946 mice and fruit flies were considered. For each species, previous associations were
947 identified based on systematic review of validated monogenic causes of human male
948 infertility (Houston et al., 2021), on information available on the Mouse Genome Database
949 [all genes associated with the phenotype code MP:0001925 (male infertility); (Blake et al.,
950 2021)], and by manual curation of all reported phenotypes listed for each gorilla K<1
951 ortholog in the Flybase repository (Larkin et al., 2021). For both the mouse and fruit fly
952 databases, information reflects data available in the repositories as of April 2023. Second,
953 only orthologs that were reliably expressed in the fruit fly testis (transcripts per million >1;
954 based on the analysis of the RNAseq dataset of (Vedelek et al., 2018) were included.

955 The GAL4/UAS system was used to silence the selected 156 fruit fly orthologs
956 specifically from the onset of the mitosis-to-meiosis transition onwards (Brand and
957 Perrimon, 1993), via the use of the *bam*-GAL4 driver (White-Cooper, 2012). UAS-hairpin
958 lines targeting the selected genes were purchased from the Bloomington Drosophila
959 Stock Centre (BDSC) and the Vienna Drosophila Resource Centre (VDRC). Lines

960 previously associated with a phenotype in the literature, regardless of the tissue, were
961 preferentially chosen for this experiment. In the single case (CG33977) for which no UAS-
962 hairpin lines were available, this reagent was generated in-house following standard
963 procedures (Perkins et al., 2015). Briefly, shRNA were designed using DSIR (Vert et al.,
964 2006), with the corresponding sequences being cloned into the pWalium 20 vector, and
965 with the constructs being injected into fruit flies carrying an attp site on the third
966 chromosome (BDSC stock #24749).

967 For assessing the requirement of each tested gene for male fertility, gene-silenced
968 males (3 to 7 days post-eclosion) were mated with wildtype Oregon-R virgin females for
969 12 hours at 25ºC (Brattig-Correia et al., 2024). Laid eggs were left to develop for 24 hours
970 at 25ºC before the percentage of egg hatching was determined. This percentage (fertility
971 rate) served as a measure of the male reproductive fitness associated with each tested
972 gene. Fertility rate corresponds to the average of four independent experiments, with 25
973 - 100 eggs scored per replicate. Every batch of experiments included a negative control
974 [RNAi against the mCherry fluorophore (BDSC stock #35785) - a sequence absent in the
975 fruit fly genome], and a positive control [RNAi against Ribosomal protein L3 (BDSC stock
976 #36596) - an essential unit of the ribosome]. A cut-off of <75% fertility rate was
977 established to define impaired reproductive fitness based on the rate observed in the
978 negative controls (>2 standard deviations of the mean observed in negative controls: 94.2
979 ± 4.8%). All Drosophila lines were maintained at 25ºC in polypropylene bottles containing
980 enriched medium (cornmeal, molasses, yeast, beet syrup, and soy flour). To compare the
981 observed hit rate of our RNAi screen with that of a random selection of genes, we used
982 the Mersenne Twister pseudorandom number generator (Molbiotools.com) to unbiasedly
983 define a set of 156 Drosophila genes. An additional comparison was made to a random
984 selection of 156 Drosophila testis-expressed genes (transcripts per million >1), selected
985 using the RNAseq dataset of Vedelek *et al.*, 2018 and a random number generator
986 algorithm. Both lists were manually curated to identify all that had been previously
987 functionally associated with male fertility and/or spermatogenesis, based on information
988 available in the Flybase repository (Larkin et al., 2021). This information is available in
989 **Supplementary dataset 2**.

990 **Drosophila testicular imaging**

991 Squash preparations of freshly dissected Drosophila testes were performed as
992 previously described (Bonaccorsi et al., 2011) and examined using a phase contrast
993 microscope (Nikon Eclipse E400). Phenotypes of all silenced genes associated with
994 decreased reproductive fitness (43 in total) were assigned to one of four classes based
995 on the earliest stage in which the cytological defects were detected: 1- pre-meiotic
996 (abnormal late spermatogonia); 2- meiotic (failure to progress through meiosis
997 successfully); 3- post-meiotic (abnormal spermiogenesis and/or low numbers of mature
998 sperm); and 4- undetectable (no observable cytological defects). Images from at least 2
999 pairs of testes were acquired per genotype and were corrected for background
1000 illumination as previously described (Brattig-Correia et al., 2024). Representative
1001 examples can be found in **Figure 5B**.

1002 **Burden analyses**

1003 We compared the number of rare, putative high-impact variants in the 568 human
1004 orthologs of the gorilla K<1 genes between a cohort of infertile men and a control cohort.
1005 As a baseline, we also tested 568 human orthologs of randomly selected K≈1 genes using
1006 the same approach. More specifically, we queried variants in both gene sets in
1007 exome/genome data of 1) 2,100 infertile men with 0 - 5 million/mL sperm concentration
1008 from the Male Reproductive Genomics (MERGE) cohort (69% azoospermia, 22%
1009 cryptozoospermia, 9% severe oligozoospermia) (Wyrwoll et al., 2020), and 2) 67,961 men
1010 from gnomAD v2.1.1 (Karczewski et al., 2020). While no individual-level information is
1011 available in gnomAD that would allow us to exclude infertile individuals, gnomAD
1012 approximates a general population of average health and fertility. Stringent filtering criteria
1013 were applied: rarity (gnomAD maximum population allele frequency [AF] ≤ 1%) quality
1014 (gnomAD PASS filter or MERGE quality score ≥ 30, read depth ≥ 10, MERGE AF ≤ 1%)
1015 and predicted high impact (LoF: stop, frameshift, splice donor/acceptor and/or CADD
1016 score ≥ 20). For each of the 568 K<1 and 568 K≈1 genes, we compared the proportion
1017 of variants fulfilling the selection criteria in the two cohorts using the one-sided Fisher's
1018 exact test. The two resulting non-parametric distributions of p-values of these gene sets

1019 were then evaluated by a two-sided Mann-Whitney U test with continuity correction. All
1020 statistical analyses were performed using R statistical software, and variants were
1021 annotated for filtering with Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016).

1022

1023 **Acknowledgments**

1024 Frank Tüttelmann was supported by the Deutsche Forschungsgemeinschaft
1025 (DFG, German Research Foundation) within the Clinical Research Unit 'Male Germ Cells'
1026 (CRU326, project number 329621271). Paulo Navarro-Costa is partially supported by
1027 Fundação para a Ciência e a Tecnologia (grants EXPL/MEC-AND/0676/2021 and
1028 2024.04983.RESTART). Computational support was provided by the Center for
1029 Computational Research at the University at Buffalo.

1030

1031 **Data availability**

1032 All alignments, selection test results (ABSREL, RELAX, and FEL), and PolPhen-2
1033 scores are available from Dryad as fasta and json files, respectively
1034 (https://doi.org/10.5061/dryad.5dv41nsbc). Json files are viewable by uploading
1035 individual files to HyPhy Vision (http://vision.hyphy.org).

1036

1037 **Appendix**

1038 **Comparison to previous studies of rapid evolution in gorillas**

1039 A previous study reporting the sequencing, annotation, and analysis of the gorilla
1040 genome identified genes with evidence of rapid evolution in the gorilla lineage (Scally et
1041 al., 2012). This study included five other primates (human, chimpanzee, orangutan,
1042 macaque, and marmoset) and the widely used two-ratio test implemented in PAML v4.4c
1043 (Yang, 2007). The two ratio test compares to models, the null hypothesis that one d_N/d_S
1044 ratio applies to all lineages in the phylogeny to a two ratio alternative model in which the
1045 foreground lineage (in this case gorilla) is allowed to have a separate d_N/d_S ratio compared
1046 to all other branches which share a single d_N/d_S rate. The likelihood of these two scenarios

1047 are compared with a likelihood ratio test and a difference in selective regime is observed
1048 when the alternate model has a higher likelihood with a chi-square P-value<=1. Under
1049 this hypothesis test, one can indirectly infer that the gorilla foreground is under relaxed
1050 selection intensity if its d_N/d_S is less than the background d_N/d_S. However, this test can be
1051 biased if one or more lineages experience an increased in d_N/d_S driving up the average
1052 d_N/d_S ratio. In this scenario, selection intensity hasn't changed in the gorilla lineage even
1053 if the null model is rejected. Similarly, a decrease in d_N/d_S under the PAML model can
1054 arise because the strength of positive selection is weaker in the foreground lineage than
1055 the background lineage even if there is still positive selection acting on those sites. While
1056 the two-ratio test requires the average d_N/d_S across all sites be greater than one to infer
1057 positive selection, the branch-site test bins sites into three rate classes that constrains
1058 the first two rate classes to be $d_N/d_S \leq 1$ and one in free class that allows for $d_N/d_S > 1$.
1059 Positive selection is inferred if the likelihood of a model that fixes the third d_N/d_S rate class
1060 to be one is significantly worse than the likelihood of a model in which it is inferred from
1061 the data and when that d_N/d_S rate class is greater than one.

1062 Neither of these methods is well-suited to detect episodic relaxed selection
1063 intensity like the RELAX test (see our previous description of the RELAX test). We also
1064 note that none of the methods implemented in PAML account for synonymous rate
1065 variation across sites or multiple simultaneous codon mutations; failure to account for
1066 these pervasive mutational processes can lead to both false-positive and false-negative-
1067 inferences of positive selection in branch-site-like tests, but are expected to have less
1068 dramatic impacts on RELAX-like analyses. Of the 280 genes identified by Scally et al.,
1069 2012 with evidence of rapid evolution in the gorilla lineage from the two-ratio test, i.e.,
1070 those with a foreground d_N/d_S greater than the background d_N/d_S at LRT $P \leq 0.05$, 213 were
1071 also included in our study; 41 (19.24%) had K<1 in the RELAX test (LRT $P \leq 0.05$) and two
1072 (0.93%) had $d_N/d_S > 1$ in the ABSREL[SMH] test (LRT $P \leq 0.05$). While the overlap in genes
1073 inferred to be positively selected and under relaxed selection between our datasets is
1074 small, Scally et al., 2012 also found that rapidly evolving genes were enriched in male
1075 reproductive functions.

1076 **Human orthologs of gorilla K<1 genes may evolve under recent relaxed selection**

1077 The observed differences between the human orthologs of the K<1 and K≈1
1078 genes for pLI and missense Z-scores suggest that gorilla K<1 genes are *less* constrained
1079 than K≈1 genes in humans (**Appendix figure 1A**). This observation could result from: 1)
1080 Population-specific positive selection regimes acting on the human orthologs of the K<1
1081 genes, i.e., increasing the number of amino acid polymorphisms; 2) Long-term relaxed
1082 selection intensity in both gorillas and humans, i.e., these genes are K<1 in both species;
1083 3) Balancing selection, i.e., the active maintenance of multiple alleles of the K<1 genes in
1084 humans; or 4) Recent relaxed (rather than long-term) purifying selection in humans. To
1085 explore these possibilities, we first ran the ABSREL test on the human lineage and found
1086 that only one of the gorilla K<1 genes was identified as positively selected in the human
1087 lineage (LRT $P \leq 0.05$). Furthermore, using the RELAX test, we observed that only 24 genes
1088 were inferred to be K<1 in both human and gorilla (LRT $P \leq 0.05$). Thus, we concluded that
1089 it is unlikely that the signal for reduced constraint in humans results either from the gorilla
1090 K<1 genes having experienced an episode of adaptive evolution in the human lineage, or
1091 from long-term relaxed selection intensity in both gorillas and humans. Indeed, these
1092 results of the ABC-MK test are consistent with balancing selection acting on the human
1093 orthologs of the gorilla K<1 genes (which would inflate P_N and therefore decrease α) or
1094 with recent relaxed selection in humans (which would also inflate P_N/P_S compared to the
1095 long-term P_N/P_S that shaped the d_N/d_S).

1096 To disentangle the two scenarios, we compared our results to several studies that
1097 have used complementary methods to characterize the strength and direction of
1098 selection acting on great ape (Hominidae) genomes (Cagan et al., 2016). These included:
1099 1) The standard McDonald–Kreitman test [MK test; (McDonald and Kreitman, 1991)],
1100 which can detect lineage-specific positive selection; 2) The Hudson–Kreitman–Aguadé
1101 test [HKA; (Hudson et al., 1987)], which can detect long-term positive and balancing
1102 selection including balancing selection that predates the emergence of the Hominidae
1103 (Hedrick, 1998); 3) The Extended Lineage Sorting test [ELS; (Green et al., 2010)], which
1104 can detect lineage-specific positive selection that occurred after the divergence of an
1105 ancestral population into two species; 4) Fay and Wu's H statistic [FWH; (Fay and Wu,
1106 2000)], which can detect recent selective sweeps; and 5) The Z test (Soni et al., 2022), a

1107 more recent method to detect balancing selection based on the application of the MK
1108 test to human population data from the 1000 Genomes Project. Although the gorilla K<1
1109 genes were significantly enriched for candidate genes to be under balancing selection,
1110 only 79 of these genes were inferred to be under balancing or positive selection in gorillas,
1111 humans, or other Hominids (**Appendix figure 1B**). These data suggest that the signal for
1112 reduced constraint in humans most likely results from recent relaxed purifying selection
1113 rather than long-term balancing selection within Hominids or humans.

1114 **Comparisons to elephant seal, another species with a highly polygynous mating**
1115 **system**

1116 To determine if the patterns of reduced selection intensity we observed in gorillas
1117 may be similar to other species with a polygynyous mating system, we ran the RELAX
1118 analysis with elephant seal as the foreground. Although elephant seals are a polygynous
1119 species, they differ from gorillas in that their spermatogenesis does not undergo
1120 persistent deterioration, but instead follows a seasonal pattern (Laws, 1956). Indeed, in
1121 this species, male gamete production is upregulated during the mating season and is
1122 mostly inactive throughout the rest of the year. Of the 573 genes with K<1 in gorillas only
1123 14 also have K<1 in elephant seals, which had 350 genes with K<1 (**Appendix figure 2A**).
1124 A GO analysis of the 350 elephant seal K<1 genes does not identify enrichment in
1125 spermatogenesis-related terms. In fact, the list of GO terms is quite broad (**Appendix**
1126 **figure 2B**). A potential, if admittedly speculative, interpretation of these findings is that
1127 although polygynous, the selective pressure on elephant seal spermatogenesis is not
1128 relaxed (unlike in gorillas) because of the seasonal nature of their mating period. In other
1129 words, by having a temporally narrower window for reproductive success than gorillas,
1130 the selective constraint on male gametogenesis in seals is not weakened. Regardless,
1131 the low overlap in relaxed genes between the two tested polygynous species support the
1132 view that this reproductive strategy is probably associated with different evolutionary
1133 signatures in the genome (depending on the species), a likely reflection of the complex,
1134 nuanced and multi-factorial aspects of such strategies.

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* **526**:68–74. doi:10.1038/nature15393

Abril-Parreño L, Carthy TR, Keogh K, Štiavnická M, O'Meara C, Lonergan P, Kenny DA, Fair S. 2023. Genome-wide association study reveals candidate markers related to field fertility and semen quality traits in Holstein-Friesian bulls. *Animal* **17**:100841. doi:10.1016/j.animal.2023.100841

Abril-Parreño L, Wilkinson H, Krogenæs A, Morgan J, Gallagher ME, Reid C, Druart X, Fair S, Saldova R. 2022. Identification and characterization of O-linked glycans in cervical mucus as biomarkers of sperm transport: A novel sheep model. *Glycobiology* **32**:23–35. doi:10.1093/glycob/cwab085

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**:Unit7.20. doi:10.1002/0471142905.hg0720s76

Akbari A, Pipitone GB, Anvar Z, Jaafarinia M, Ferrari M, Carrera P, Totonchi M. 2019. ADCY10 frameshift variant leading to severe recessive asthenozoospermia and segregating with absorptive hypercalciuria. *Hum Reprod* **34**:1155–1164. doi:10.1093/humrep/dez048

Alavattam KG, Maezawa S, Sakashita A, Khoury H, Barski A, Kaplan N, Namekawa SH. 2019. Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat Struct Mol Biol* **26**:175–184. doi:10.1038/s41594-019-0189-y

Altmann SA, Wagner SS, Lenington S. 1977. Two models for the evolution of polygyny. *Behav Ecol Sociobiol* **2**:397–410. doi:10.1007/BF00299508

Amaral A, Castillo J, Estanyol JM, Ballescà JL, Ramalho-Santos J, Oliva R. 2013. Human sperm tail proteome suggests new endogenous metabolic pathways. *Mol Cell Proteomics* **12**:330–342. doi:10.1074/mcp.M112.020552

Anderson MJ, Chapman SJ, Videan EN, Evans E, Fritz J, Stoinski TS, Dixson AF, Gagneux P. 2007. Functional evidence for differences in sperm competition in humans and chimpanzees. *Am J Phys Anthropol* **134**:274–280. doi:10.1002/ajpa.20674

Anderson MJ, Dixson AF. 2002. Motility and the midpiece in primates. *Nature* **416**:496–496. doi:10.1038/416496a

Anderson MJ, Nyholt J, Dixson AF. 2004. Sperm competition affects the structure of the mammalian vas deferens. *Journal of Zoology* **264**:97.

Avidan N, Tamary H, Dgany O, Cattan D, Pariente A, Thulliez M, Borot N, Moati L, Barthelme A, Shalmon L, Krasnov T, Ben-Asher E, Olender T, Khen M, Yaniv I, Zaizov R, Shalev H, Delaunay J, Fellous M, Lancet D, Beckmann JS. 2003.

1175 CATSPER2, a human autosomal nonsyndromic male infertility gene. *Eur J Hum*
1176 *Genet* **11**:497–502. doi:10.1038/sj.ejhg.5200991

1177 Ayhan Ö, Balkan M, Guven A, Hazan R, Atar M, Tok A, Tolun A. 2014. Truncating
1178 mutations in TAF4B and ZMYND15 causing recessive azoospermia. *Journal of*
1179 *Medical Genetics* **51**:239–244. doi:10.1136/jmedgenet-2013-102102

1180 Bashamboo A, Ferraz-de-Souza B, Lourenço D, Lin L, Sebire NJ, Montjean D,
1181 Bignon-Topalovic J, Mandelbaum J, Siffroi J-P, Christin-Maitre S, Radhakrishna
1182 U, Rouba H, Ravel C, Seeler J, Achermann JC, McElreavey K. 2010. Human Male
1183 Infertility Associated with Mutations in NR5A1 Encoding Steroidogenic Factor 1.
1184 *Am J Hum Genet* **87**:505–512. doi:10.1016/j.ajhg.2010.09.009

1185 Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, Mouse
1186 Genome Database Group. 2021. Mouse Genome Database (MGD):
1187 Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res*
1188 **49**:D981–D987. doi:10.1093/nar/gkaa1083

1189 Bonaccorsi S, Giansanti MG, Cenci G, Gatti M. 2011. Preparation of live testis
1190 squashes in Drosophila. *Cold Spring Harb Protoc* **2011**:prot5577.
1191 doi:10.1101/pdb.prot5577

1192 Bonache S, Martínez J, Fernández M, Bassas L, Larriba S. 2007. Single nucleotide
1193 polymorphisms in succinate dehydrogenase subunits and citrate synthase genes:
1194 association results for impaired spermatogenesis. *International Journal of*
1195 *Andrology* **30**:144–152. doi:10.1111/j.1365-2605.2006.00730.x

1196 Bradley BJ, Robbins MM, Williamson EA, Steklis HD, Steklis NG, Eckhardt N,
1197 Boesch C, Vigilant L. 2005. Mountain gorilla tug-of-war: silverbacks have limited
1198 control over reproduction in multimale groups. *Proc Natl Acad Sci U S A* **102**:9418–
1199 9423. doi:10.1073/pnas.0502019102

1200 Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering
1201 cell fates and generating dominant phenotypes. *Development* **118**:401–415.
1202 doi:10.1242/dev.118.2.401

1203 Brattig-Correia R, Almeida JM, Wyrwoll MJ, Julca I, Sobral D, Misra CS, Di Persio
1204 S, Guilgur LG, Schuppe H-C, Silva N, Prudêncio P, Nóvoa A, Leocádio AS, Bom
1205 J, Laurentino S, Mallo M, Kliesch S, Mutwil M, Rocha LM, Tüttelmann F, Becker
1206 JD, Navarro-Costa P. 2024. The conserved genetic program of male germ cells
1207 uncovers ancient regulators of human spermatogenesis. *eLife* **13**:RP95774.
1208 doi:10.7554/eLife.95774

1209 Brennan PLR, Prum RO, McCracken KG, Sorenson MD, Wilson RE, Birkhead TR.
1210 2007. Coevolution of Male and Female Genital Morphology in Waterfowl. *PLOS*
1211 *ONE* **2**:e418. doi:10.1371/journal.pone.0000418

1212 Bruce HM. 1959. An Exteroceptive Block to Pregnancy in the Mouse. *Nature*
1213 **184**:105–105. doi:10.1038/184105a0

1214 Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Mouse Genome Database
1215 Group. 2019. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*
1216 **47**:D801–D806. doi:10.1093/nar/gky1056

1217 Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F, Prüfer K,
1218 Navarro A, Marques-Bonet T, Bertranpetit J, Andrés AM. 2016. Natural Selection
1219 in the Great Apes. *Mol Biol Evol* **33**:3268–3283. doi:10.1093/molbev/msw215

1220 Cao W, Gerton GL, Moss SB. 2006. Proteomic profiling of accessory structures
1221 from the mouse sperm flagellum. *Mol Cell Proteomics* **5**:801–810.
1222 doi:10.1074/mcp.M500322-MCP200

1223 Carnahan SJ, Jensen-Seaman MI. 2008. Hominoid seminal protein evolution and
1224 ancestral mating behavior. *Am J Primatol* **70**:939–948. doi:10.1002/ajp.20585

1225 Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballescà JL, Oliva R.
1226 2014. Genomic and proteomic dissection and characterization of the human sperm
1227 chromatin. *Mol Hum Reprod* **20**:1041–1053. doi:10.1093/molehr/gau079

1228 Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution
1229 at synonymous sites in mammals. *Nat Rev Genet* **7**:98–108. doi:10.1038/nrg1770

1230 Chen Y, Liu B, Tao S, Liu L, Gao J, Liang Y, Dong W, Zhou D. 2025. CITED2
1231 Binding to EP300 Regulates Human Spermatogonial Stem Cell Proliferation and
1232 Survival Through HSPA6. *Stem Cells Int* **2025**:2362489. doi:10.1155/sci/2362489

1233 Christensen GL, Wooding SP, Ivanov IP, Atkins JF, Carrell DT. 2006. Sequencing
1234 and haplotype analysis of the Activator of CREM in the Testis (ACT) gene in
1235 populations of fertile and infertile males. *Mol Hum Reprod* **12**:257–262.
1236 doi:10.1093/molehr/gal006

1237 Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal
1238 proteins. *PLoS Genet* **1**:e35. doi:10.1371/journal.pgen.0010035

1239 Cox CM, Thoma ME, Tchangalova N, Mburu G, Bornstein MJ, Johnson CL, Kiarie
1240 J. 2022. Infertility prevalence and the methods of estimation from 1990 to 2021: a
1241 systematic review and meta-analysis. *Hum Reprod Open* **2022**:hoac051.
1242 doi:10.1093/hropen/hoac051

1243 Cullen TM, Fraser D, Rybczynski N, Schröder-Adams C. 2014. Early evolution of
1244 sexual dimorphism and polygyny in Pinnipedia. *Evolution* **68**:1469–1484.
1245 doi:10.1111/evo.12360

1246 Dacheux D, Martinez G, Reix CEB, Beurois J, Lorès P, Tounkara M, Dupuy J-W,
1247 Robinson DR, Loeuillet C, Lambert E, Wehbe Z, Amiri-Yekta A, Daneshipour A,
1248 Hosseini S-H, Zouari R, Mustapha SFB, Halouani L, Jiang X, Shen Y, Liu C,
1249 Thierry-Mieg N, Septier A, Bidart M, Satre V, Cazin C, Kherraf Z-E, Arnoult C, Ray
1250 PF, Touré A, Bonhivers M, Coutton C. 2023. Novel axonemal protein ZMYND12
1251 interacts with TTC29 and DNAH1, and is required for male fertility and flagellum
1252 function. *eLife* **12**. doi:10.7554/eLife.87698.1

1253 Dai Y, Hu R, Liu A, Cho KS, Manuel AM, Li X, Dong X, Jia P, Zhao Z. 2022.
1254 WebCSEA: web-based cell-type-specific enrichment analysis of genes. *Nucleic*
1255 *Acids Res* **50**:W782–W790. doi:10.1093/nar/gkac392

de Mateo S, Castillo J, Estanyol JM, Ballescà JL, Oliva R. 2011. Proteomic characterization of the human sperm nucleus. *Proteomics* **11**:2714–2726. doi:10.1002/pmic.201000799

Di Franco A, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19**:21. doi:10.1186/s12862-019-1350-2

Dixson AF. 2018. Copulatory and Postcopulatory Sexual Selection in Primates. *Folia Primatol (Basel)* **89**:258–286. doi:10.1159/000488105

Dixson AF. 2012. Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans, Second Edition, Second Edition. ed. Oxford, New York: Oxford University Press.

Dixson AF. 2009. Sexual Selection and the Origins of Human Mating Systems. Oxford, New York: Oxford University Press.

Dixson AF, Moore HDM, Holt WV. 1980. Testicular atrophy in captive gorillas (Gorilla g. gorilla). *Journal of Zoology* **191**:315–322. doi:10.1111/j.1469-7998.1980.tb01461.x

Easley CA, Phillips BT, McGuire MM, Barringer JM, Valli H, Hermann BP, Simerly CR, Rajkovic A, Miki T, Orwig KE, Schatten GP. 2012. Direct differentiation of human pluripotent stem cells into haploid spermatogenic cells. *Cell Rep* **2**:440–446. doi:10.1016/j.celrep.2012.07.015

Eberhard WG. 1996. Female Control: Sexual Selection by Cryptic Female Choice. Princeton University Press. doi:10.2307/j.ctvs32rx1

Enomoto T, Matsubayashi K, Nakano M, Fujii-Hanamoto H, Kusunoki H. 2004. Testicular histological examination of spermatogenetic activity in captive gorillas (Gorilla gorilla). *Am J Primatol* **63**:183–199. doi:10.1002/ajp.20051

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405–1413. doi:10.1093/genetics/155.3.1405

Firat-Karalar EN, Sante J, Elliott S, Stearns T. 2014. Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J Cell Sci* **127**:4128–4133. doi:10.1242/jcs.157008

Firman RC, Gasparini C, Manier MK, Pizzari T. 2017. Postmating Female Control: 20 Years of Cryptic Female Choice. *Trends Ecol Evol* **32**:368–382. doi:10.1016/j.tree.2017.02.010

Fujii-Hanamoto H, Matsubayashi K, Nakano M, Kusunoki H, Enomoto T. 2011. A comparative study on testicular microstructure and relative sperm production in gorillas, chimpanzees, and orangutans. *Am J Primatol* **73**:570–577. doi:10.1002/ajp.20930

Gómez Montoto L, Varea Sánchez M, Tourmente M, Martín-Coello J, Luque-Larena JJ, Gomendio M, Roldan ERS. 2011. Sperm competition differentially

1295 affects swimming velocity and size of spermatozoa from closely related muroid
1296 rodents: head first. *Reproduction* **142**:819–830. doi:10.1530/REP-11-0232

1297 Good JM, Nachman MW. 2005. Rates of protein evolution are positively correlated
1298 with developmental timing of expression during mouse spermatogenesis. *Mol Biol*
1299 *Evol* **22**:1044–1052. doi:10.1093/molbev/msi087

1300 Gould KG. 1990. Techniques and significance of gamete collection and storage in
1301 the great apes. *J Med Primatol* **19**:537–551.

1302 Gould KG, Kling OR. 1982. Fertility in the male gorilla (Gorilla gorilla): Relationship
1303 to semen parameters and serum hormones. *Am J Primatol* **2**:311–316.
1304 doi:10.1002/ajp.1350020310

1305 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li
1306 H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD,
1307 Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R,
1308 Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A,
1309 Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan
1310 P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C,
1311 de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush
1312 D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D,
1313 Pääbo S. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–
1314 722. doi:10.1126/science.1188021

1315 Greenbaum MP, Yan W, Wu M-H, Lin Y-N, Agno JE, Sharma M, Braun RE,
1316 Rajkovic A, Matzuk MM. 2006. TEX14 is essential for intercellular bridges and
1317 fertility in male mice. *Proc Natl Acad Sci U S A* **103**:4982–4987.
1318 doi:10.1073/pnas.0505123103

1319 Guyonnet B, Zabet-Moghaddam M, SanFrancisco S, Cornwall GA. 2012. Isolation
1320 and proteomic characterization of the mouse sperm acrosomal matrix. *Mol Cell*
1321 *Proteomics* **11**:758–774. doi:10.1074/mcp.M112.020339

1322 Hall-Craggs ECB. 1962. The Testis of Gorilla Gorilla Beringei. *Proceedings of the*
1323 *Zoological Society of London* **139**:511–514. doi:10.1111/j.1469-
1324 7998.1962.tb01845.x

1325 Harcourt AH. 1997. Sperm Competition in Primates. *The American Naturalist*
1326 **149**:189–194.

1327 Harcourt AH. 1995. Sexual selection and sperm competition in primates: What are
1328 male genitalia good for? *Evolutionary Anthropology: Issues, News, and Reviews*
1329 **4**:121–129. doi:10.1002/evan.1360040404

1330 Harcourt AH, Harvey PH, Larson SG, Short RV. 1981. Testis weight, body weight
1331 and breeding system in primates. *Nature* **293**:55–57. doi:10.1038/293055a0

1332 Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. 2011. Reconstitution of the
1333 mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*
1334 **146**:519–532. doi:10.1016/j.cell.2011.06.052

He X, Liu C, Yang X, Lv M, Ni X, Li Q, Cheng H, Liu W, Tian S, Wu H, Gao Y, Yang C, Tan Q, Cong J, Tang D, Zhang J, Song B, Zhong Y, Li H, Zhi W, Mao X, Fu F, Ge L, Shen Q, Zhang M, Saiyin H, Jin L, Xu Y, Zhou P, Wei Z, Zhang F, Cao Y. 2020. Bi-allelic Loss-of-function Variants in *CFAP58* Cause Flagellar Axoneme and Mitochondrial Sheath Defects and Asthenoteratozoospermia in Humans and Mice. *The American Journal of Human Genetics* **107**:514–526. doi:10.1016/j.ajhg.2020.07.010

Hedrick PW. 1998. Balancing selection and MHC. *Genetica* **104**:207–214. doi:10.1023/a:1026494212540

Hering DM, Oleński K, Ruść A, Kaminski S. 2014. Genome-wide association study for semen volume and total number of sperm in Holstein-Friesian bulls. *Anim Reprod Sci* **151**:126–130. doi:10.1016/j.anireprosci.2014.10.022

Hill WCO, Harrison-Matthews L. 1949. The male external genitalia of the Gorilla, with remarks on the os penis of other Hominoidea. *Proceedings of the Zoological Society of London* **119**:363–378. doi:10.1111/j.1096-3642.1949.tb00886.x

Hino T, Yanagimachi R. 2019. Active peristaltic movements and fluid production of the mouse oviduct: their roles in fluid and sperm transport and fertilization†. *Biol Reprod* **101**:40–49. doi:10.1093/biolre/ioz061

Houston BJ, Riera-Escamilla A, Wyrwoll MJ, Salas-Huetos A, Xavier MJ, Nagirnaja L, Friedrich C, Conrad DF, Aston KI, Krausz C, Tüttelmann F, O'Bryan MK, Veltman JA, Oud MS. 2021. A systematic review of the validated monogenic causes of human male infertility: 2020 update and a discussion of emerging gene–disease relationships. *Human Reproduction Update* **28**:15–29. doi:10.1093/humupd/dmab030

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153–159. doi:10.1093/genetics/116.1.153

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**:D309–D314. doi:10.1093/nar/gky1085

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170. doi:10.1038/335167a0

Hwang JY, Nawaz S, Choi J, Wang H, Hussain S, Nawaz M, Lopez-Giraldez F, Jeong K, Dong W, Oh J-N, Bilguvar K, Mane S, Lee C-K, Bystroff C, Lifton RP, Ahmad W, Chung J-J. 2021. Genetic Defects in DNAH2 Underlie Male Infertility With Multiple Morphological Abnormalities of the Sperm Flagella in Humans and Mice. *Front Cell Dev Biol* **9**:662903. doi:10.3389/fcell.2021.662903

Inoue E, Akomo-Okoue EF, Ando C, Iwata Y, Judai M, Fujita S, Hongo S, Nze-Nkogue C, Inoue-Murayama M, Yamagiwa J. 2013. Male genetic structure and paternity in western lowland gorillas (Gorilla gorilla gorilla). *Am J Phys Anthropol* **151**:583–588. doi:10.1002/ajpa.22312

1378 Ishikawa Y, Usui T, Yamashita M, Kanemori Y, Baba T. 2016. Surfing and
1379 Swimming of Ejaculated Sperm in the Mouse Oviduct. *Biol Reprod* **94**:89.
1380 doi:10.1095/biolreprod.115.135418

1381 Jacobs SC, Beehler BA, Boese G, Story MT, Clowry LJ, Lawson RK. 1984. The
1382 prostate of the gorilla. *Prostate* **5**:597–603. doi:10.1002/pros.2990050605

1383 Jensen-Seaman MI, Li W-H. 2003. Evolution of the hominoid semenogelin genes,
1384 the major proteins of ejaculated semen. *J Mol Evol* **57**:261–270.
1385 doi:10.1007/s00239-003-2474-x

1386 Jourquin J, Duncan D, Shi Z, Zhang B. 2012. GLAD4U: deriving and prioritizing
1387 gene lists from PubMed literature. *BMC Genomics* **13 Suppl 8**:S20.
1388 doi:10.1186/1471-2164-13-S8-S20

1389 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1390 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat*
1391 *Methods* **14**:587–589. doi:10.1038/nmeth.4285

1392 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL,
1393 Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M,
1394 Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA,
1395 Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin
1396 N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH,
1397 Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L,
1398 Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S,
1399 Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo
1400 N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K,
1401 Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG. 2020. The
1402 mutational constraint spectrum quantified from variation in 141,456 humans.
1403 *Nature* **581**:434–443. doi:10.1038/s41586-020-2308-7

1404 Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* **12**:656–664.
1405 doi:10.1101/gr.229202

1406 Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky
1407 NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL,
1408 Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL,
1409 Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C,
1410 Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC,
1411 Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG,
1412 Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yüksel Z, Helbig I, Mungall CJ,
1413 Haendel MA, Robinson PN. 2021. The Human Phenotype Ontology in 2021.
1414 *Nucleic Acids Res* **49**:D1207–D1217. doi:10.1093/nar/gkaa1043

1415 Kopania EEK, Larson EL, Callahan C, Keeble S, Good JM. 2022. Molecular
1416 Evolution across Mouse Spermatogenesis. *Mol Biol Evol* **39**:msac023.
1417 doi:10.1093/molbev/msac023

1418 Kopania EEK, Thomas GWC, Hutter CR, Mortimer SME, Callahan CM, Roycroft
1419 E, Achmadi AS, Breed WG, Clark NL, Esselstyn JA, Rowe KC, Good JM. 2025.
1420 Sperm competition intensity shapes divergence in both sperm morphology and

reproductive genes across murine rodents. *Evolution* **79**:11–27. doi:10.1093/evolut/qpae146

Kosakovsky Pond SL. 2021. RELAX-scan. https://github.com/veg/hyphy-analyses/tree/master/RELAX-scan

Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**:1208–1222. doi:10.1093/molbev/msi105

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol* **28**:3033–3043. doi:10.1093/molbev/msr125

Kosova G, Hotaling JM, Ohlander S, Niederberger C, Prins GS, Ober C. 2014. Variants in DPF3 and DSCAML1 are associated with sperm morphology. *J Assist Reprod Genet* **31**:131–137. doi:10.1007/s10815-013-0140-9

Kosova G, Scott NM, Niederberger C, Prins GS, Ober C. 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet* **90**:950–961. doi:10.1016/j.ajhg.2012.04.016

Krausz C, Riera-Escamilla A, Moreno-Mendoza D, Holleman K, Cioppi F, Algaba F, Pybus M, Friedrich C, Wyrwoll MJ, Casamonti E, Pietroforte S, Nagirnaja L, Lopes AM, Kliesch S, Pilatz A, Carrell DT, Conrad DF, Ars E, Ruiz-Castañé E, Aston KI, Baarends WM, Tüttelmann F. 2020. Genetic dissection of spermatogenic arrest through exome analysis: clinical implications for the management of azoospermic men. *Genet Med* **22**:1956–1966. doi:10.1038/s41436-020-0907-1

Krüger O, Wolf JBW, Jonker RM, Hoffman JI, Trillmich F. 2014. Disentangling the contribution of sexual selection and ecology to the evolution of size dimorphism in pinnipeds. *Evolution* **68**:1485–1496. doi:10.1111/evo.12370

Kuo Y-C, Shen Y-R, Chen H-I, Lin Y-H, Wang Y-Y, Chen Y-R, Wang C-Y, Kuo P-L. 2015. SEPT12 orchestrates the formation of mammalian sperm annulus by organizing core octameric complexes with other SEPT proteins. *J Cell Sci* **128**:923–934. doi:10.1242/jcs.158998

Lanzendorf SE, Holmgren WJ, Johnson DE, Scobey MJ, Jeyendran RS. 1992. Hemizona assay for measuring zona binding in the lowland gorilla. *Molecular Reproduction and Development* **31**:264–267. doi:10.1002/mrd.1080310407

Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, Goodman JL, Gramates LS, Millburn G, Strelets VB, Tabone CJ, Thurmond J, FlyBase Consortium. 2021. FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Research* **49**:D899–D907. doi:10.1093/nar/gkaa1026

Laws RM. 1956. The Elephant Seal (Mirounga leonina, Linn.): III. The physiology of reproduction. *Falkland Islands Dependencies Survey Scientific Reports* **15**.

Lee C-H, Wu C-C, Wu Y-N, Chiang H-S. 2009. Gene copy number variations in Asian patients with congenital bilateral absence of the vas deferens. *Hum Reprod* **24**:748–755. doi:10.1093/humrep/den413

Lehti MS, Sironen A. 2017. Formation and function of sperm tail structures in association with sperm motility defects. *Biol Reprod* **97**:522–536. doi:10.1093/biolre/iox096

Lewis RW, Kim JC, Irani D, Roberts JA. 1981. The prostate of the nonhuman primate: normal anatomy and pathology. *Prostate* **2**:51–70. doi:10.1002/pros.2990020106

Lewontin RC. 2002. Directions in evolutionary biology. *Annu Rev Genet* **36**:1–18. doi:10.1146/annurev.genet.36.052902.102704

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**:W199–W205. doi:10.1093/nar/gkz401

Lipshutz SE, Torneo SJ, Rosvall KA. 2023. How Female-Female Competition Affects Male-Male Competition: Insights into Postcopulatory Sexual Selection from Socially Polyandrous Species. *Am Nat* **201**:460–471. doi:10.1086/722799

Liu M, Wang L, Li Y, Zhi E, Shen G, Jiang X, Li D, Zhao X, Ruan T, Jiang C, Wang X, Zhang X, Zheng Y, Wu B, Ou N, Zhao G, Dai S, Zhou R, Yang L, Yang Y, Liu H, Shen Y. 2024. HSF5 Deficiency Causes Male Infertility Involving Spermatogenic Arrest at Meiotic Prophase I in Humans and Mice. *Advanced Science* **11**:2402412. doi:10.1002/advs.202402412

Liu S, Zhang J, Kherraf ZE, Sun S, Zhang X, Cazin C, Coutton C, Zouari R, Zhao S, Hu F, Fourati Ben Mustapha S, Arnoult C, Ray PF, Liu M. 2021. CFAP61 is required for sperm flagellum formation and male fertility in human and mouse. *Development* **148**:dev199805. doi:10.1242/dev.199805

Lucaci AG, Wisotsky SR, Shank SD, Weaver S, Kosakovsky Pond SL. 2021. Extra base hits: Widespread empirical support for instantaneous multiple-nucleotide changes. *PLoS One* **16**:e0248337. doi:10.1371/journal.pone.0248337

Maezawa S, Sakashita A, Yukawa M, Chen X, Takahashi K, Alavattam KG, Nakata I, Weirauch MT, Barski A, Namekawa SH. 2020. Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. *Nat Struct Mol Biol* **27**:978–988. doi:10.1038/s41594-020-0488-3

Margolin DH, Kousi M, Chan Y-M, Lim ET, Schmahmann JD, Hadjivassiliou M, Hall JE, Adam I, Dwyer A, Plummer L, Aldrin SV, O'Rourke J, Kirby A, Lage K, Milunsky A, Milunsky JM, Chan J, Hedley-Whyte ET, Daly MJ, Katsanis N, Seminara SB. 2013. Ataxia, Dementia, and Hypogonadotropism Caused by Disordered Ubiquitination. *N Engl J Med* **368**:1992–2003. doi:10.1056/NEJMoa1215993

Martinez G, Garcia C. 2020. Sexual selection and sperm diversity in primates. *Molecular and Cellular Endocrinology* **518**:110974. doi:10.1016/j.mce.2020.110974

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**:652–654. doi:10.1038/351652a0

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**:122. doi:10.1186/s13059-016-0974-4

Messier W, Stewart CB. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385**:151–154. doi:10.1038/385151a0

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**:1530–1534. doi:10.1093/molbev/msaa015

Møller AP. 1998. 2 - Sperm Competition and Sexual Selection In: Birkhead TR, Møller A. P., editors. Sperm Competition and Sexual Selection. San Diego: Academic Press. pp. 55–90. doi:10.1016/B978-012100543-6/50027-1

Morita M, Satoh S, Ito T, Kohda M, Awata S. 2023. Sperm competition influences sperm quality and seminal plasma protein expression among territorial males of the courtship cichlid. *bioRxiv: The Preprint Server for Biology* 2023.09.07.556763. doi:10.1101/2023.09.07.556763

Mullins KJ, Saacke RG. 1989. Study of the functional anatomy of bovine cervical mucosa with special reference to mucus secretion and sperm transport. *Anat Rec* **225**:106–117. doi:10.1002/ar.1092250205

Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mößinger K, Brüning T, Lamanna F, Belles MR, Conrad C, Kondova I, Bontrop R, Behr R, Khaitovich P, Pääbo S, Marques-Bonet T, Grützner F, Almstrup K, Schierup MH, Kaessmann H. 2023. The molecular evolution of spermatogenesis across mammals. *Nature* **613**:308–316. doi:10.1038/s41586-022-05547-7

Murga-Moreno J, Casillas S, Barbadilla A, Uricchio L, Enard D. 2023. An efficient and robust ABC approach to infer the rate and strength of adaptation. *bioRxiv*. doi:10.1101/2023.08.29.555322

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* **11**:715–724. doi:10.1093/oxfordjournals.molbev.a040152

Nabi S, Askari M, Rezaei-Gazik M, Salehi N, Almadani N, Tahamtani Y, Totonchi M. 2022. A rare frameshift mutation in SYCP1 is associated with human male infertility. *Mol Hum Reprod* **28**:gaac009. doi:10.1093/molehr/gaac009

Nascimento JM, Shi LZ, Meyers S, Gagneux P, Loskutoff NM, Botvinick EL, Berns MW. 2008. The use of optical tweezers to study sperm competition and motility in primates. *J R Soc Interface* **5**:297–302. doi:10.1098/rsif.2007.1118

1541 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and
1542 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol*
1543 *Biol Evol* **32**:268–274. doi:10.1093/molbev/msu300

1544 Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino
1545 acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**:929–936.
1546 doi:10.1093/genetics/148.3.929

1547 Niu X, Tang W, Liu Yongshuo, Mo B, Yu Y, Liu Ying, Wei W. 2025. Prime editor-
1548 based high-throughput screening reveals functional synonymous mutations in
1549 human cells. *Nat Biotechnol* 1–13. doi:10.1038/s41587-025-02710-z

1550 Nsubuga AM, Robbins MM, Boesch C, Vigilant L. 2008. Patterns of paternity and
1551 group fission in wild multimale mountain gorilla groups. *Am J Phys Anthropol*
1552 **135**:263–274. doi:10.1002/ajpa.20740

1553 O'Flynn O'Brien KL, Varghese AC, Agarwal A. 2010. The genetic causes of male
1554 factor infertility: a review. *Fertil Steril* **93**:1–12. doi:10.1016/j.fertnstert.2009.10.045

1555 Parker GA. 1970. Sperm Competition and Its Evolutionary Consequences in the
1556 Insects. *Biological Reviews* **45**:525–567. doi:10.1111/j.1469-185X.1970.tb01176.x

1557 Perkins LA, Holderbaum L, Tao R, Hu Y, Sopko R, McCall K, Yang-Zhou D,
1558 Flockhart I, Binari R, Shim H-S, Miller A, Housden A, Foos M, Randkelv S, Kelley
1559 C, Namgyal P, Villalta C, Liu L-P, Jiang X, Huan-Huan Q, Wang X, Fujiyama A,
1560 Toyoda A, Ayers K, Blum A, Czech B, Neumuller R, Yan D, Cavallaro A, Hibbard
1561 K, Hall D, Cooley L, Hannon GJ, Lehmann R, Parks A, Mohr SE, Ueda R, Kondo
1562 S, Ni J-Q, Perrimon N. 2015. The Transgenic RNAi Project at Harvard Medical
1563 School: Resources and Validation. *Genetics* **201**:843–852.
1564 doi:10.1534/genetics.115.180208

1565 Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F,
1566 Furlong LI. 2020. The DisGeNET knowledge platform for disease genomics: 2019
1567 update. *Nucleic Acids Res* **48**:D845–D855. doi:10.1093/nar/gkz1021

1568 Platz CC, Wildt DE, Bridges CH, Seager SW, Whitlock BS. 1980.
1569 Electroejaculation and semen analysis in a male lowland gorilla,Gorilla gorilla
1570 gorilla. *Primates* **21**:130–132. doi:10.1007/BF02383831

1571 Pond SK, Muse SV. 2005. Site-to-site variation of synonymous substitution rates.
1572 *Mol Biol Evol* **22**:2375–2385. doi:10.1093/molbev/msi232

1573 Ptak SE, Lachmann M. 2003. On the evolution of polygyny: a theoretical
1574 examination of the polygyny threshold model. *Behavioral Ecology* **14**:201–211.
1575 doi:10.1093/beheco/14.2.201

1576 Ramirez-Diaz J, Cenadelli S, Bornaghi V, Bongioni G, Montedoro SM, Achilli A,
1577 Capelli C, Rincon JC, Milanesi M, Passamonti MM, Colli L, Barbato M, Williams
1578 JL, Marsan PA. 2023. Identification of genomic regions associated with total and
1579 progressive sperm motility in Italian Holstein bulls. *J Dairy Sci* **106**:407–420.
1580 doi:10.3168/jds.2021-21700

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol Biol Evol* **35**:2582–2584. doi:10.1093/molbev/msy159

Richardson L, Hanrahan JP, Tharmalingam T, Carrington SD, Lonergan P, Evans ACO, Fair S. 2019. Cervical mucus sialic acid content determines the ability of frozen-thawed ram sperm to migrate through the cervix. *Reproduction* **157**:259–271. doi:10.1530/REP-18-0547

Rivas-González I, Rousselle M, Li F, Zhou L, Dutheil JY, Munch K, Shao Y, Wu D, Schierup MH, Zhang G. 2023. Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science* **380**:eabn4409. doi:10.1126/science.abn4409

Robinson M, Witherspoon L, Willerth S, Flannigan R. 2021. A Novel Organoid Model of In Vitro Spermatogenesis Using Human Induced Pluripotent Stem Cells. *bioRxiv: The Preprint Server for Biology* 2021.06.04.447122. doi:10.1101/2021.06.04.447122

Romanel A, Zhang T, Elemento O, Demichelis F. 2017. EthSEQ: ethnicity annotation from whole exome sequencing data. *Bioinformatics* **33**:2402–2404. doi:10.1093/bioinformatics/btx165

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol Biol Evol* **30**:2134–2144. doi:10.1093/molbev/mst116

Roucher-Boulez F, Mallet-Motak D, Samara-Boustani D, Jilani H, Ladjouze A, Souchon P-F, Simon D, Nivot S, Heinrichs C, Ronze M, Bertagna X, Groisne L, Leheup B, Naud-Saudreau C, Blondin G, Lefevre C, Lemarchand L, Morel Y. 2016. NNT mutations: a cause of primary adrenal insufficiency, oxidative stress and extra-adrenal defects. *Eur J Endocrinol* **175**:73–84. doi:10.1530/EJE-16-0056

Ruan T, Zhou R, Yang Y, Guo J, Jiang C, Wang X, Shen G, Dai S, Chen S, Shen Y. 2024. Deficiency of IQCH causes male infertility in humans and mice. *eLife* **12**:RP88905. doi:10.7554/eLife.88905

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA, Gu Y, Heath P, Heger A, Karakoc E, Kolb-Kokocinski A, Laird GK, Lunter G, Meader S, Mort M, Mullikin JC, Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K, Simpson JT, Stenson PD, Turner DJ, Vigilant L, Vilella AJ, Whitener W, Zhu B, Cooper DN, de Jong P, Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA, Ryder OA, Searle SM, Warren WC, Wilson RK, Schierup MH, Rogers J, Tyler-Smith C, Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**:169–175. doi:10.1038/nature10842

Seager SWJ, Wildt DE, Schaffer N, Platz CC. 1982. Semen collection and evaluation in Gorilla gorilla gorilla. *American Journal of Primatology* **3**:13–13. doi:10.1002/ajp.1350030506

Serrano M, Ramón M, Calvo JH, Jiménez MÁ, Freire F, Vázquez JM, Arranz JJ. 2021. Genome-wide association studies for sperm traits in Assaf sheep breed. *Animal* **15**:100065. doi:10.1016/j.animal.2020.100065

Seuanez HN, Carothers AD, Martin DE, Short RV. 1977. Morphological abnormalities in spermatozoa of man and great apes. *Nature* **270**:345–347. doi:10.1038/270345a0

Shah K, Sivapalan G, Gibbons N, Tempest H, Griffin DK. 2003. The genetic basis of infertility. *Reproduction* **126**:13–25. doi:10.1530/rep.0.1260013

Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, Sukhwani M, Orwig KE, Li JZ, Hammoud SS. 2020. Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell* **54**:529-547.e12. doi:10.1016/j.devcel.2020.05.010

Short RV. 1979. Sexual Selection and Its Component Parts, Somatic and Genital Selection, as Illustrated by Man and the Great Apes* In: Rosenblatt JS, Hinde RA, Beer C, Busnel M-C, editors. Advances in the Study of Behavior. Academic Press. pp. 131–158. doi:10.1016/S0065-3454(08)60035-2

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* **32**:1342–1353. doi:10.1093/molbev/msv022

Soni V, Vos M, Eyre-Walker A. 2022. A new test suggests hundreds of amino acid polymorphisms in humans are subject to balancing selection. *PLoS Biol* **20**:e3001645. doi:10.1371/journal.pbio.3001645

Suarez SS. 2016. Mammalian sperm interactions with the female reproductive tract. *Cell Tissue Res* **363**:185–194. doi:10.1007/s00441-015-2244-2

Suarez SS. 1987. Sperm transport and motility in the mouse oviduct: observations in situ. *Biol Reprod* **36**:203–210. doi:10.1095/biolreprod36.1.203

Suarez SS, Pacey AA. 2006. Sperm transport in the female reproductive tract. *Hum Reprod Update* **12**:23–37. doi:10.1093/humupd/dmi047

Suarez SS, Wolfner MF. 2021. Cilia take the egg on a magic carpet ride. *Proc Natl Acad Sci U S A* **118**:e2108887118. doi:10.1073/pnas.2108887118

Svensson EI. 2020. O Causation, Where Art Thou? *BioScience* **70**:264–268. doi:10.1093/biosci/biaa009

Tian S, Faheem M, Satti HS, Xiao J, Zhang F, Khan TN, Liu C. 2024. A homozygous missense variant in YTHDC2 induces azoospermia in two siblings. *Mol Genet Genomics* **299**:84. doi:10.1007/s00438-024-02168-9

1663 Tu C, Cong J, Zhang Q, He X, Zheng R, Yang X, Gao Y, Wu H, Lv M, Gu Y, Lu S,
1664 Liu C, Tian S, Meng L, Wang W, Tan C, Nie H, Li D, Zhang H, Gong F, Hu L, Lu
1665 G, Xu W, Lin G, Zhang F, Cao Y, Tan Y-Q. 2021. Bi-allelic mutations of DNAH10
1666 cause primary male infertility with asthenoteratozoospermia in humans and mice.
1667 *Am J Hum Genet* **108**:1466–1477. doi:10.1016/j.ajhg.2021.06.010

1668 Tung C-K, Suarez SS. 2021. Co-Adaptation of Physical Attributes of the
1669 Mammalian Female Reproductive Tract and Sperm to Facilitate Fertilization. *Cells*
1670 **10**:1297. doi:10.3390/cells10061297

1671 Tüttelmann F, Ruckert C, Röpke A. 2018. Disorders of spermatogenesis:
1672 Perspectives for novel genetic diagnostics after 20 years of unchanged routine.
1673 *Med Genet* **30**:12–20. doi:10.1007/s11825-018-0181-7

1674 Uricchio LH, Petrov DA, Enard D. 2019. Exploiting selection at linked sites to infer
1675 the rate and strength of adaptation. *Nat Ecol Evol* **3**:977–984. doi:10.1038/s41559-
1676 019-0890-6

1677 Vedelek V, Bodai L, Grézal G, Kovács B, Boros IM, Laurinyecz B, Sinka R. 2018.
1678 Analysis of Drosophila melanogaster testis transcriptome. *BMC Genomics* **19**:697.
1679 doi:10.1186/s12864-018-5085-z

1680 Verner J. 1964. Evolution of Polygamy in the Long-Billed Marsh Wren. *Evolution*
1681 **18**:252–261. doi:10.2307/2406398

1682 Vert J-P, Foveau N, Lajaunie C, Vandenbrouck Y. 2006. An accurate and
1683 interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7**:520.
1684 doi:10.1186/1471-2105-7-520

1685 Villeneuve AM, Hillers KJ. 2001. Whence meiosis? *Cell* **106**:647–650.
1686 doi:10.1016/s0092-8674(01)00500-1

1687 Wagner GP. 2001. What is the promise of developmental evolution? Part II: A
1688 causal explanation of evolutionary innovations may be impossible. *J Exp Zool*
1689 **291**:305–309. doi:10.1002/jez.1130

1690 Wagner GP. 2000. What is the promise of developmental evolution? Part I: why is
1691 developmental biology necessary to explain evolutionary innovations? *J Exp Zool*
1692 **288**:95–98. doi:10.1002/1097-010x(20000815)288:2%3C95::aid-
1693 jez1%3E3.0.co;2-5

1694 Wang W, Su L, Meng L, He J, Tan C, Yi D, Cheng D, Zhang H, Lu G, Du J, Lin G,
1695 Zhang Q, Tu C, Tan Y-Q. 2023. Biallelic variants in KCTD19 associated with male
1696 factor infertility and oligoasthenoteratozoospermia. *Hum Reprod* **38**:1399–1411.
1697 doi:10.1093/humrep/dead095

1698 Wang Y, Wan J, Ling X, Liu M, Zhou T. 2016. The human sperm proteome 2.0: An
1699 integrated resource for studying sperm functions at the level of posttranslational
1700 modification. *Proteomics* **16**:2597–2601. doi:10.1002/pmic.201600233

1701 Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015.
1702 RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Molecular*
1703 *Biology and Evolution* **32**:820–832. doi:10.1093/molbev/msu400

White-Cooper H. 2012. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. *Spermatogenesis* **2**:11–22. doi:10.4161/spmg.19088

Wislocki GB. 1942. Size, Weight, and Histology of the Testes in the Gorilla. *Journal of Mammalogy* **23**:281–287. doi:10.2307/1374995

Wisotsky SR, Kosakovsky Pond SL, Shank SD, Muse SV. 2020. Synonymous Site-to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection Analyses: Ignore at Your Own Peril. *Mol Biol Evol* **37**:2430–2439. doi:10.1093/molbev/msaa037

Wu Y, Yuan Y, Chen L, Wang M, Yang Y, Wang Y, Quan C, Chen D, Chen Y, Huang X, Zhou T. 2019. Quantitative Proteomic Analysis of Human Seminal Plasma from Normozoospermic and Asthenozoospermic Individuals. *Biomed Res Int* **2019**:2735038. doi:10.1155/2019/2735038

Wyrwoll MJ, Temel ŞG, Nagirnaja L, Oud MS, Lopes AM, Heijden GW van der, Heald JS, Rotte N, Wistuba J, Wöste M, Ledig S, Krenz H, Smits RM, Carvalho F, Gonçalves J, Fietz D, Türkgenç B, Ergören MC, Çetinkaya M, Başar M, Kahraman S, McEleny K, Xavier MJ, Turner H, Pilatz A, Röpke A, Dugas M, Kliesch S, Neuhaus N, Aston KI, Conrad DF, Veltman JA, Friedrich C, Tüttelmann F. 2020. Bi-allelic Mutations in M1AP Are a Frequent Cause of Meiotic Arrest and Severely Impaired Spermatogenesis Leading to Male Infertility. *The American Journal of Human Genetics* **107**:342–351. doi:10.1016/j.ajhg.2020.06.010

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586–1591. doi:10.1093/molbev/msm088

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

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19**:908–917. doi:10.1093/oxfordjournals.molbev.a004148

Yapar E, Saglican E, Dönertaş HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfs RV, Khaitovich P, Somel M. 2021. Convergent evolution of primate testis transcriptomes reflects mating strategy. *bioRxiv* 010553. doi:10.1101/010553

Yin H, Zhou C, Shi S, Fang L, Liu J, Sun D, Jiang L, Zhang S. 2019. Weighted Single-Step Genome-Wide Association Study of Semen Traits in Holstein Bulls of China. *Front Genet* **10**:1053. doi:10.3389/fgene.2019.01053

Yuan S, Wang Z, Peng H, Ward SM, Hennig GW, Zheng H, Yan W. 2021. Oviductal motile cilia are essential for oocyte pickup but dispensable for sperm and embryo transport. *Proc Natl Acad Sci U S A* **118**:e2102940118. doi:10.1073/pnas.2102940118

Zhang J, Qian W. 2025. Functional synonymous mutations and their evolutionary consequences. *Nat Rev Genet* 1–16. doi:10.1038/s41576-025-00850-1

Zhang M, Chiozzi RZ, Skerrett-Byrne DA, Veenendaal T, Klumperman J, Heck AJR, Nixon B, Helms JB, Gadella BM, Bromfield EG. 2022. High Resolution Proteomic Analysis of Subcellular Fractionated Boar Spermatozoa Provides Comprehensive Insights Into Perinuclear Theca-Residing Proteins. *Front Cell Dev Biol* **10**:836208. doi:10.3389/fcell.2022.836208

Zhang P, Jiang W, Luo N, Zhu W, Fan L. 2019. IQ motif containing D (IQCD), a new acrosomal protein involved in the acrosome reaction and fertilisation. *Reprod Fertil Dev* **31**:898–914. doi:10.1071/RD18416

Zhang X, Zheng R, Liang C, Liu H, Zhang Xiaozhen, Ma Y, Liu Mohan, Zhang W, Yang Yang, Liu Man, Jiang C, Ren Q, Wang Y, Chen S, Yang Yihong, Shen Y. 2022. Loss-of-function mutations in CEP78 cause male infertility in humans and mice. *Sci Adv* **8**:eabn0968. doi:10.1126/sciadv.abn0968

Zielen A. 2017. Primate proteomic composition of seminal plasma and prostate-specific transglutaminase activity in relation to sexual selection. *Electronic Theses and Dissertations*.

1774 **Box 1. Erosion of the male reproductive system and sperm function in gorillas**

1775 The near absence of sperm competition in gorillas is associated with the degeneration of
1776 many male reproductive system traits associated with postcopulatory male-male and
1777 sperm competition (Dixson, 2018). Anatomical changes include small testis (**Box 1—**
1778 **figure 1A**) and penis (**Box 1—figure 1B**) sizes relative to their body size compared to
1779 other primates (Harcourt et al., 1981; Hill and Harrison-Matthews, 1949). Gorillas also
1780 have abnormal testis ultrastructure, organization, and cell-type composition (**Box 1—**
1781 **figure 1C**) characterized by a smaller seminiferous tubule diameter and length and an
1782 increased proportion of interstitial (non-gametogenic) tissue than other apes (Dixson et
1783 al., 1980; Enomoto et al., 2004; Fujii-Hanamoto et al., 2011; Hall-Craggs, 1962; Jacobs
1784 et al., 1984; Wislocki, 1942). Other derived traits include a small epididymis (Dixson et al.,
1785 1980), an apparent lack of seminal vesicles (Dixson et al., 1980), and a small prostate that
1786 appears to lack a distinct cranial lobe, i.e., the coagulating gland (Jacobs et al., 1984;
1787 Lewis et al., 1981); these latter two traits are likely related to the observation that gorillas
1788 have low semen volume (Martinez and Garcia, 2020; Seager et al., 1982) and seminal
1789 plasma that lacks many proteins involved in semen coagulation and liquefaction (Lewis
1790 et al., 1981; Zielen, 2017). In addition, gorillas have a smaller sperm midpiece than many
1791 other primates [**Box 1—figure 1D**; (Anderson and Dixson, 2002)]; the midpiece is a
1792 mitochondria-enriched subcellular component of the sperm that provides energy for
1793 sperm motility. This trait is likely related to gorilla sperm having extremely low
1794 mitochondrial function (Anderson et al., 2007) and slow swimming speed and weak
1795 swimming force [**Box 1—figure 1E**; (Nascimento et al., 2008)]. Gorillas also have low
1796 sperm counts (Fujii-Hanamoto et al., 2011; Gould, 1990; Martinez and Garcia, 2020), and
1797 a large proportion of immotile (Martinez and Garcia, 2020) and morphologically abnormal
1798 sperm [**Box 1—figure 1F**; (Martinez and Garcia, 2020; Platz et al., 1980; Seuanez et al.,
1799 1977)]. Gorilla sperm also bind the egg's zona pellucida more weakly than other species
1800 (Lanzendorf et al., 1992). While many of these traits could result from abnormal hormone
1801 levels, circulating levels of steroid and protein hormones are normal and not associated
1802 with aspermatogenesis or infertility in gorillas (Gould and Kling, 1982). Therefore,
1803 abnormalities in the structure and function of the male reproductive system and sperm
1804 biology in gorillas likely result from changes in the expression or function of genes

1805 responsible for the development and function of the male reproductive system and
1806 sperm, rather than being secondary to endocrinological causes.

1807

1808 **Box 1 – figure 1. Derived male reproductive characteristics in gorillas.**

1809 **A.** Correlation between testis weight and body weight in primates. Adapted from (Harcourt
1810 et al., 1981).

1811 **B.** Correlation between erect penis length and body weight in primates. Adapted from
1812 (Dixson, 2009).

1813 **C.** Diagram of testis histology in gorilla and other primates. Note that gorillas have fewer
1814 seminiferous tubules (T) and more interstitial tissue (I) than other primates. Adapted from
1815 (Fujii-Hanamoto et al., 2011).

1816 **D.** Correlation between the sperm midpiece volume and residual testis size (relative testis
1817 size accounting for body mass). Adapted from (Anderson and Dixson, 2002).

1818 **E.** Boxplots showing the distribution of sperm swimming speeds (VCL, μms^{-1}) and escape
1819 forces (F_{esc}, pN). Adapted from (Nascimento et al., 2008).

1820 **F.** Proportion of morphologically abnormal sperm in Hominids. Adapted from (Seuanez et
1821 al., 1977).

1822

1823 **Appendix figure 1. Patterns of selection acting on human and ape orthologs of gorilla K<1**
1824 **genes.**

1825 **A.** ABC-MK test results from 1000 Genomes Project. The rate of adaptation (alpha) as a
1826 function of derived allele frequency for human orthologs of K<1 (blue line) and K≈1 genes
1827 (gray line with 95% CI).

1828 **B.** UpSet plot showing the intersection between gorilla K<1 and K>1 genes and genes
1829 inferred to be under positive and balancing selection for the ELS, HKA, FWH, standard
1830 MK, and Z tests. The inset tree shows the Hominoid phylogeny and number of individuals
1831 used for tests from Cagan et al., 2016. The total number of genes in each set, overlap with

1832 K<1 and K>1, expected overlap, enrichment ratio, hypergeometric *P*-value, and FDR q-
1833 value for enrichment of K<1 genes in each gene set are shown in a table.

1834

1835 **Appendix figure 2. Comparison between gorilla and elephant seal K<1 and K>1 genes.**

1836 **A.** Venn diagram showing gorilla and elephant seal K<1 and K>1 statistically significant (P≤
1837 0.05) genes.

1838 **B.** Top 20 gene ontology (GO) and Reactome (R-HSA-pathway name) terms in which
1839 elephant seal K<1 genes) are enriched.

1840

1841 **Table 1. Gorilla relaxed genes associated with abnormal sperm biology and**
1842 **infertility in humans.** PolyPhen-2 predictions of gorilla amino acid substitutions
1843 (probably damaging | possibly damaging | benign, or NA for unable to determine), K and
1844 *P*-values from the RELAX test, and associated human phenotypes are shown for each
1845 gene.

Gene	PolyPhen-2	K	P	Phenotype
ADCY10	3|1|9	0.053	0.012	Asthenozoospermia (Akbari et al., 2019)
STRC	2|5|10	0.305	0.018	Asthenozoospermia with deafness (Avidan et al., 2003)
CREM	NA	0.019	0.047	Azoospermia (Christensen et al., 2006)
SDHA	1|0|3	0.257	0.014	Azoospermia and oligozoospermia (Bonache et al., 2007)
PANK2	NA	0.010	0.049	Congenital bilateral absence of vas deferens (Lee et al., 2009)
NR5A1	1|0|2	0.000	0.023	Isolated spermatogenic failure (Bashamboo et al., 2010)

OTUD4	0|1|5	0.032	0.032	Kallmann syndrome (Margolin et al., 2013)
DNAH2	5|2|13	0.305	0.002	Multiple morphological abnormalities of the sperm flagella (Hwang et al., 2021)
CFAP61	0|1|13	0.130	0.005	Multiple morphological abnormalities of the sperm flagella (Liu et al., 2021)
CFAP58	2|1|3	0.104	0.007	Multiple morphological abnormalities of the sperm flagella (He et al., 2020)
ZMYND15	NA	0.021	0.001	Non-obstructive azoospermia (Ayhan et al., 2014)
SEPTIN12	3|0|2	0.118	0.019	Teratozoospermia (Kuo et al., 2015)
CEP78	0|1|7	0.048	0.050	Oligoasthenoteratospermia in combination with cone-rod dystrophy and hearing loss (Zhang et al., 2022)
DNAH10	5|7|14	0.399	0.002	Oligoasthenoteratozoospermia (Tu et al., 2021)
NNT	0|0|1	0.130	0.000	Primary Adrenal Insufficiency with precocious puberty (Roucher-Boulez et al., 2016)
DSCAML1	1|3|1	0.014	0.044	Abnormal sperm morphology and fertility (Kosova et al., 2014)
ZMYND12	0|0|3	0.000	0.042	Multiple morphological abnormalities of the sperm flagella (Dacheux et al., 2023)
KCTD19	1|0|6	0.16	0.032	Oligoasthenoteratozoospermia (Wang et al., 2023)
EP300	0|1|6	0.561	0.040	Reduced proliferation of spermatogonial stem cells (Chen et al., 2025)

HSF5	1|1|7	0.014	0.005	Meiotic arrest (Liu et al., 2024)
IQCH	2|2|10	0.012	0.026	Flagellar morphological abnormalities (Ruan et al., 2024)
IQCD	0|2|2	0.005	0.023	Impaired acrosome reaction (Zhang et al., 2019)
SYCP1	1|1|2	0.053	1.57e-10	Severe oligozoospermia (Nabi et al., 2022)
YTHDC2	0|0|3	0.296	0.040	Non-obstructive azoospermia (Tian et al., 2024)

Figure 1. Episodic adaptive evolution and relaxed selection intensity on protein-coding genes in gorillas.

A. Consensus phylogeny of 261 mammals from which protein-coding gene alignments were generated. The ape (Hominoidea) part of the tree is expanded, and the ancestral Homininae node used to reconstruct ancestral sequences is labeled. Branch lengths are scaled to the number of amino acid substitutions in each branch (see inset branch scale).

B. The ABSREL model detects episodic adaptive evolution. Under the null model of no positive selection, a gene can have one or more classes of sites with $d_N/d_S \leq 1$. When a gene experiences an episode of positive selection, most sites still have $d_N/d_S \leq 1$ but also includes an additional rate class with $d_N/d_S > 1$; positive selection is inferred with a gene includes a class of sites with $d_N/d_S > 1$ with a likelihood ratio test (LRT) P≤0.05.

C. Volcano plot showing the d_N/d_S value (Log_2) and likelihood ratio test (LRT) P-value ($-Log_{10}$) for the ABSREL test in the gorilla lineage. 96 genes were identified with 2 rate classes where one had $d_N/d_S > 1$ (LRT P≤0.05), 55 genes were identified with 2 rate classes where one had $d_N/d_S < 1$ (LRT P≤0.05), the remaining 13,336 genes had one rate class with $d_N/d_S<1$.

1863 **D.** Stacked bar chart showing the proportion of genes with d_N/d_S >1 (LRT P≤0.05) in the
1864 gorilla lineage. Pie charts show the proportion of genes with d_N/d_S >1 (LRT P≤0.05) in
1865 gorilla that are also d_N/d_S >1 (LRT P≤0.05) in at least one other lineage; 26 genes have
1866 evidence for positive selection only in the gorilla lineage.

1867 **E.** The RELAX model detects intensified and relaxed selection. When selection is relaxed,
1868 d_N/d_S rates move toward 1 and/or the proportion of sites increases in rate classes with
1869 d_N/d_S values close to 1. In contrast, intensified selection drives d_N/d_S rates away from
1870 neutrality to more extreme values in the foreground than background lineages. Each
1871 branch in the RELAX model includes a selection intensity parameter (K), which scales the
1872 distribution of d_N/d_S rate categories (ω_1^k, ω_2^k, ω_3^k) such that genes with K<1 are inferred to
1873 evolve under relaxed selection intensity, and genes with K>1 are inferred to evolve under
1874 enhanced selection intensity.

1875 **F.** Volcano plot showing the K-value (Log_2) and likelihood ratio test (LRT) P-value ($-Log_{10}$) for
1876 the RELAX test in the gorilla lineage. 578 genes were identified with K<1 and 144 with K>1
1877 (LRT P≤0.05); the remaining 12,769 genes had K≈1. The 25 genes with the most extreme
1878 K<1 and P-values and the two genes with the most extreme K>1 values are labeled.

1879 **G.** Stacked bar chart showing the proportion of genes with K<1, K>1, and K≈1 in the gorilla
1880 lineage. Pie charts show the proportion of genes with K>1 in gorilla that are also K≠1 in at
1881 least one other lineage (upper) or K<1 in gorilla that are also K≠1 in at least one other
1882 lineage (lower). Blue: proportion of genes with K<1 in at least one other lineage. Red:
1883 proportion of genes with K>1 in at least one other lineage. Purple: proportion of genes
1884 with K<1 and K>1 in other lineages.

1885 **Figure 1 – source data 1.** Genomes used in selection tests.

1886 **Figure 1 – source data 2.** List of positively selected genes from ABSREL.

1887 **Figure 1 – source data 3.** List of K<1 genes from RELAX.

1888 **Figure 1 – source data 4.** List of K>1 genes from RELAX.

1889 **Figure 1 – figure supplement 1.** Genomic features of genes with K<1 (blue) compared to all genes
1890 tested for relaxed selection intensity (red). *P*-values derived from a Chi-square test.

1891

1892 **Figure 2. Amino acid substitutions in gorilla relaxed genes are likely deleterious.**

1893 **A.** Volcano plot showing the d_N/d_S (Log_2) and likelihood ratio test (LRT) P-value ($-Log_{10}$) for
1894 codons with gorilla-specific amino acid changes in genes with K≠1 in gorilla. d_N/d_S rates
1895 and *P*-values were estimated with the fixed effect likelihood (FEL) model, the *P*-value is
1896 for the test of $d_N/d_S \neq 1$ at each codon. Sites with $d_N/d_S \neq 1$ at $P \leq 0.05$ in K<1 genes are
1897 shown in blue and K>1 genes in red.

1898 **B.** Bubble plot showing the percent of each amino acid substitution in the gorilla lineage,
1899 colored according to the BLOSUM 250 score for that transition. Note that, unlike amino
1900 acid substitution matrices, this matrix is not symmetric because the direction of amino
1901 acid substitution is inferred from the ancestral reconstruction.

1902 **C.** Stripchart showing the distribution of PolyPhen-2 scores for gorilla-specific amino acid
1903 changes in genes with K≈1 (12,769 genes), K<1 (578 genes), and K>1 (144 genes).
1904 PolyPhen-2 scores for each amino acid change are shown as jittered dots, a summary of
1905 the data for each gene set is shown as boxplot with the mean score indicated, and a violin
1906 plot reflecting the distribution of PolyPhen-2 scores. The number of amino acid
1907 substitutions is given in parenthesis for each gene set. The range of scores corresponding
1908 to benign, possibly damaging, and probably damaging mutations is highlighted. Multiple
1909 hypotheses (Holm) corrected p values ($p_{Holm\text{-}adj.}$) are shown for comparisons of K<1 and
1910 K>1 to K≈1, and K>1 to K>1.

1911 **D.** The percentage of probably damaging amino acid substitutions in K<1, K=1, and K>1
1912 genes. Binomial tests for differences between K<1 and K=1, and K=1 and K>1 genes are
1913 shown. These data indicate there is a greater percentage of probably damaging

1914 substitutions in K<1 genes, and a lower percentage of probably damaging substitutions in
1915 K>1 genes compared to K=1 genes.

1916

1917 **Figure 3. Gorilla relaxed genes are preferentially expressed from meiosis onwards and are**
1918 **enriched in sperm-related functions.**

1919 **A.** Word cloud showing gene ontology (GO) biology process terms for which the gorilla K<1
1920 genes are enriched. Note that only terms related to reproduction, meiosis, or sperm
1921 biology are shown. Inset size scale shows word size for an enrichment P-value of 0.01,
1922 words are colored according to their enrichment ratios.

1923 **B.** Bar chart showing GO cellular component terms in which genes with K<1 in gorilla are
1924 enriched. Only terms related to the cytoskeleton, cilia, centrosomes, and septin ring
1925 complexes are shown. Terms are colored according to their enrichment ratios (scale
1926 shown in panel A).

1927 **C.** Enrichment of genes with K<1 (blue) and K>1 (red) in the prostate, based on human
1928 scRNA-seq datasets (WebCSEA).

1929 **D.** Testis cell types in which the gorilla K≠1 genes are enriched. Left, UMAP plot of single-cell
1930 RNA-Seq data generated from gorilla testis. Major cell types are colored and named, the
1931 arrow indicates direction of sperm differentiation. Right, volcano plot showing the
1932 enrichment ratio (Log_2) and hypergeometric P-value ($-Log_{10}$) for the expression of genes
1933 with K≠1 in gorilla testis cell types. Cell types with statistically significant (FDR q-
1934 value≤0.10) enrichment or depletion of the K≠1 gene set are labeled (blue for enrichment
1935 of genes with K<1 and red for enrichment of genes with K>1). Inset diagram shows
1936 enrichment ratios for the expression of K<1 genes in each cell type, ordered by direction
1937 of germ cell differentiation (bottom to top). Bar widths are scaled according to the
1938 enrichment ratio in the gorilla lineage; the background gray box indicates a ratio of 1.
1939 Boxes and cell-type labels are colored according to cell population labels in the UMAP
1940 plot. Labels are in bold if the expression of K<1 genes is statistically enriched or depleted
1941 in that cell type (FDR q-value≤0.10).

1942 **E.** Enrichment ratio of proteins encoded by the K<1 (blue) and the K>1 (red) genes in different
1943 fractions of the sperm cell proteome (FDR q-value≤0.10).

1944 **Figure 3 – source data 1.** Enriched GO terms.

1945

1946 **Figure 4. Gorilla relaxed genes with probably damaging amino acid substitutions are**
1947 **associated with multiple sperm abnormalities.**

1948 **A.** Word cloud showing GO biological process and molecular function terms in which genes
1949 with K<1 and probably damaging substitutions are enriched. Note that only terms related
1950 to reproduction, meiosis, or sperm biology are shown. Inset size scale shows word size
1951 for an enrichment P-value of 0.01, phenotypes are colored according to their enrichment
1952 ratios.

1953 **B.** Word cloud showing the top 40 mouse knockout phenotypes in which genes with K<1
1954 and probably damaging substitutions are enriched. Inset size scale shows word size for
1955 an enrichment P-value of 0.01, phenotypes are colored according to their enrichment
1956 ratios.

1957 **C.** Testis cell types in which the expression of gorilla K<1 genes with probably damaging
1958 amino acid substitutions are enriched or depleted. Bar chart shows the enrichment ratio
1959 for each cell-type, ordered by direction of germ cell differentiation (bottom to top). Bar
1960 widths are scaled according to the enrichment ratio; the background gray box indicates a
1961 ratio of 1. Boxes and cell-type labels are colored according to cell population labels in the
1962 UMAP plot in Figure 2 panel D, labels are bold if the expression of genes with K<1 is
1963 statistically enriched or depleted in that cell-type at FDR q-value≤0.10.

1964 **D.** Enrichment ratio of proteins encoded by the K<1 genes with probably damaging amino
1965 acid changes in different fractions of the sperm cell proteome (FDR q-value≤0.10).

1966

1967 **Figure 5. Relaxed section intensity identifies new spermatogenesis genes.**

1968 **A.** *Drosophila in vivo* germ cell-specific RNAi screen uncovers new spermatogenic functions
1969 for orthologs of the gorilla K<1 genes. Silencing of 156 *Drosophila* orthologs expressed in
1970 the testis with no previous association with male fertility/spermatogenesis was induced at

1971 the onset of the mitosis-to-meiosis transition using the *bam*-GAL4 driver. Results reflect
1972 a total of four independent experiments (mean±standard deviation). Threshold for
1973 impaired reproductive fitness (red horizontal line) corresponds to a 75% fertility rate or >2
1974 standard deviations of the mean observed in negative (-ve) controls (in black). The red
1975 data point corresponds to the positive (+ve) control (ribosomal protein L3). Testicular
1976 phenotypes of the 43 hits were defined by phase-contrast microscopy and assigned to
1977 four color-coded classes based on the earliest manifestation of the phenotype (see B and
1978 C).

1979 **B.** Representative images of the four phenotypical classes of spermatogenic impairment
1980 defined in the Drosophila RNAi screen (see A). Phase-contrast microscopy of testis
1981 segments or of mature male gametes. Pre-meiotic phenotypes were characterized by a
1982 smaller testis with severely compromised germ cell growth and differentiation. Meiotic
1983 phenotypes were defined by an accumulation of spermatocytes (arrowheads) as well as
1984 by the lack of post-meiotic stages (arrows indicate elongating spermatid bundles). Post-
1985 meiotic (I) phenotypes corresponded to severe spermiogenesis defects, leading to
1986 irregular or altogether collapsed spermatid bundles (asterisk). Post-meiotic (II) phenotypes
1987 were characterized by a substantial reduction in the amount of mature male gametes
1988 observed, without any overt defects in spermiogenesis. For simplicity, the Drosophila
1989 gene names refer to the human ortholog (dIARS1 is CG11471, dGPAA1 is CG3033,
1990 dCDK12 is CG7597, and dAGK is CG31873). Scale bars: 20 μm.

1991 **C.** Half (61/116) of all gorilla K<1 genes involved in male fertility are functionally required for
1992 the post-meiotic stages of spermatogenesis (spermatid development and/or mature
1993 gamete function). Data include results from the Drosophila RNAi screen (n= 41; see A) and
1994 published evidence on human, mice and fruit fly orthologs (n= 75; see **Methods**).
1995 Phenotype classes follow the same color code as in A, with the addition of a class
1996 ("Other") corresponding to non-spermatogenic phenotypes reported in the literature
1997 (either unspecified or of somatic nature).

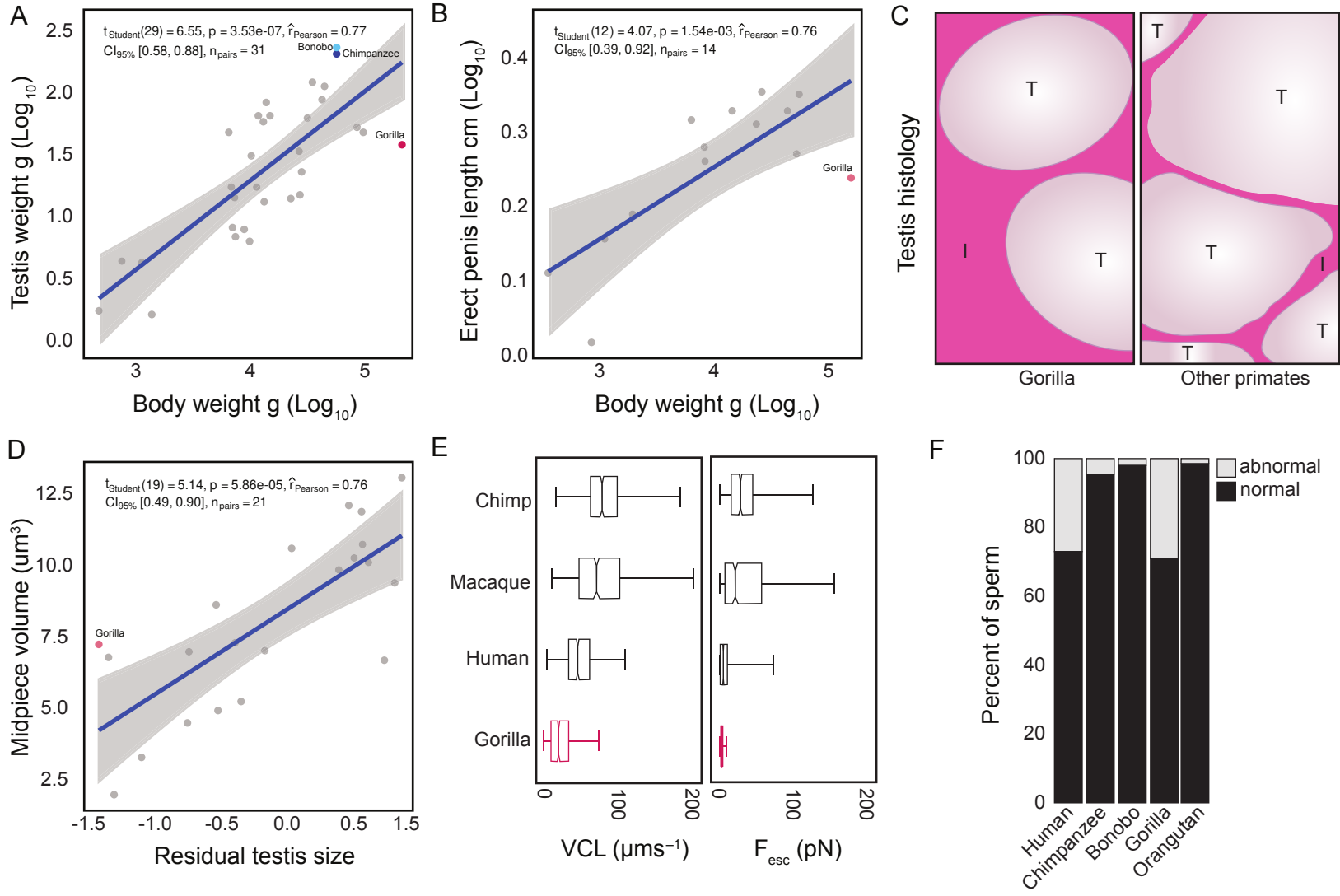
1998

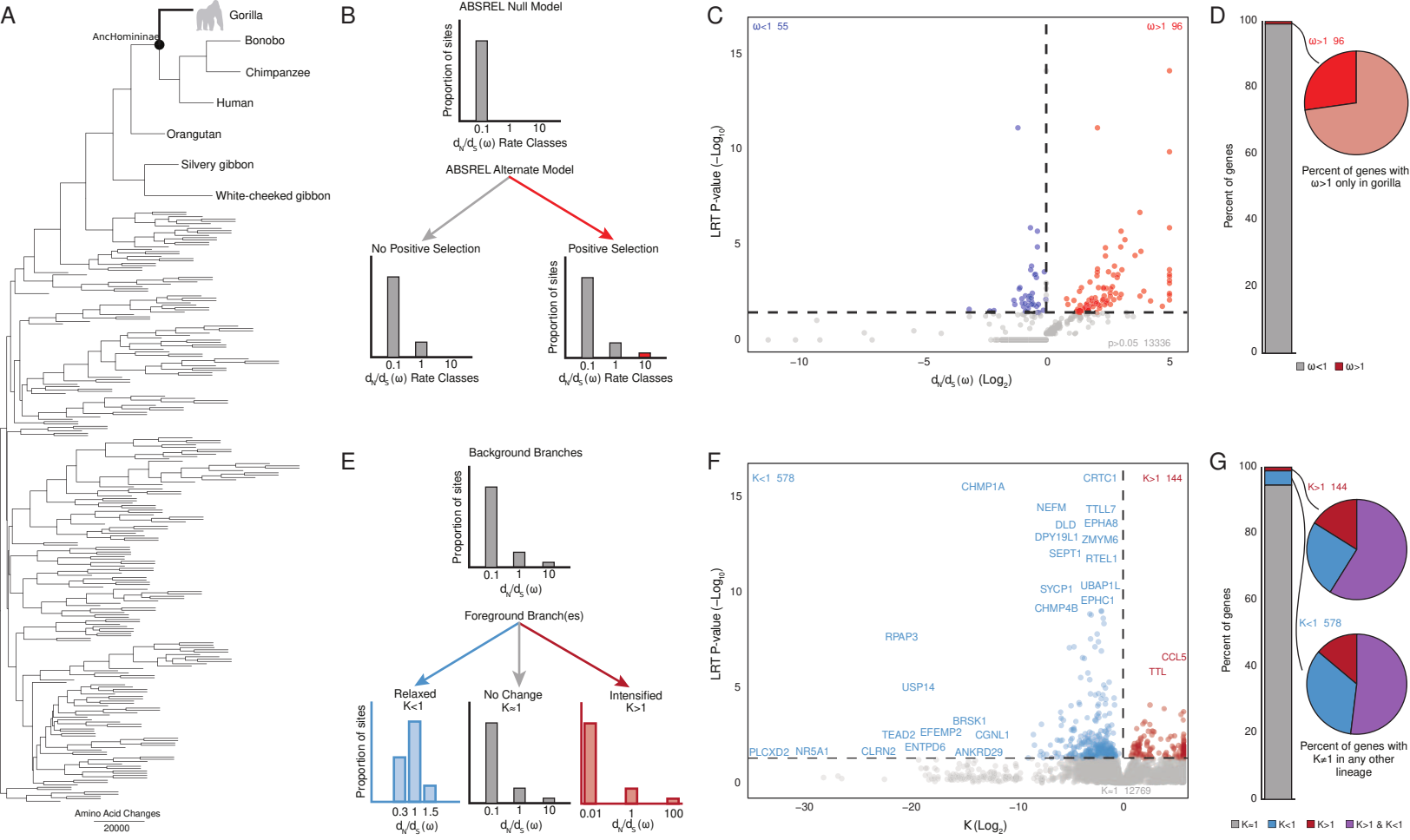
1999 **Figure 6. Gorilla relaxed genes are associated with human male infertility.**

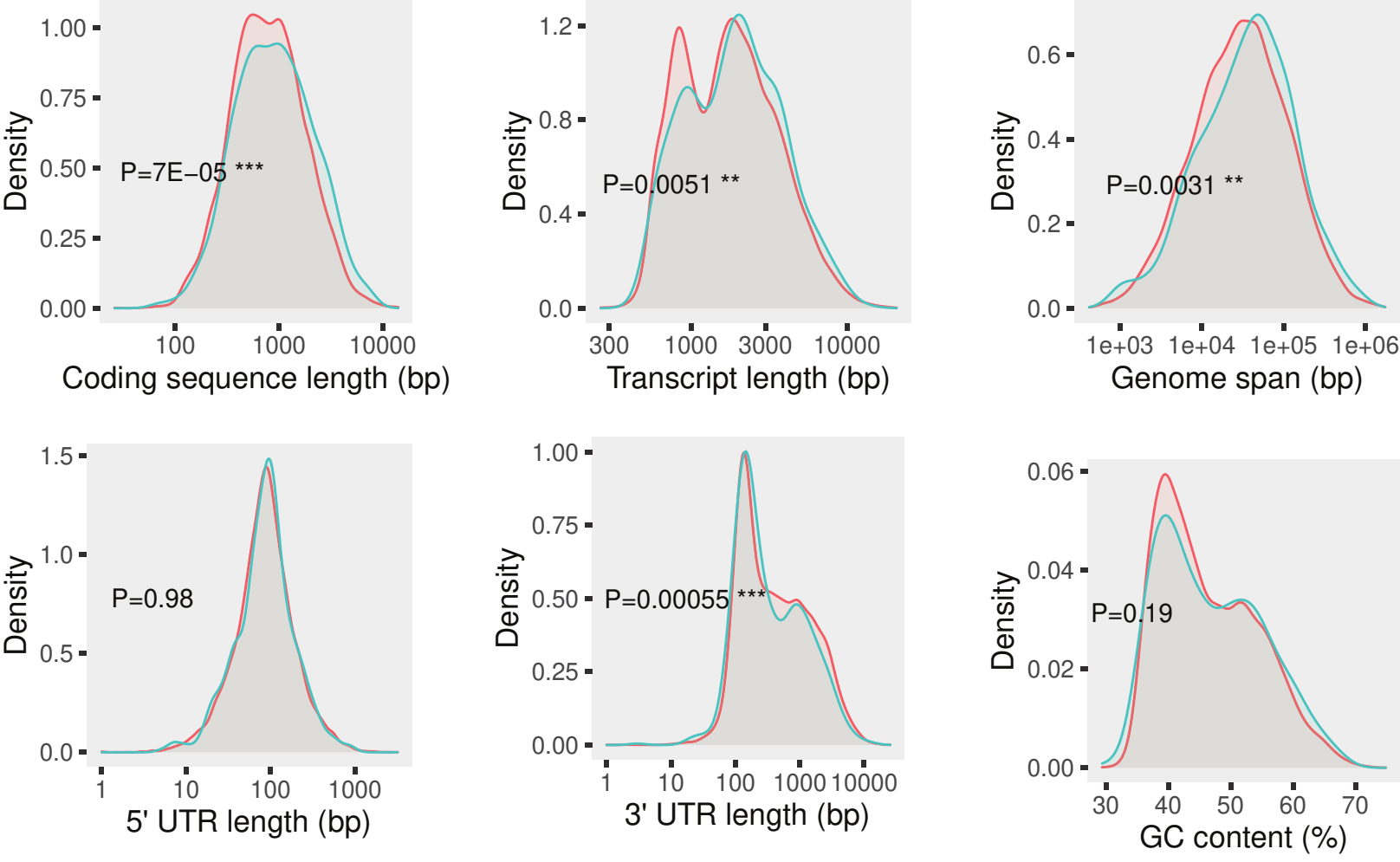
2000 **A.** Burden of predicted loss-of-function (LoF) and missense variants for human orthologs of
2001 gorilla K<1 and K≈1 genes in the MERGE cohort of infertile men. Data are shown as
2002 Fisher's exact test (FET) *P*-values for enrichment of LoF and missense variants in the
2003 MERGE cohort compared to gnomAD as proxy for a cohort of fertile men.

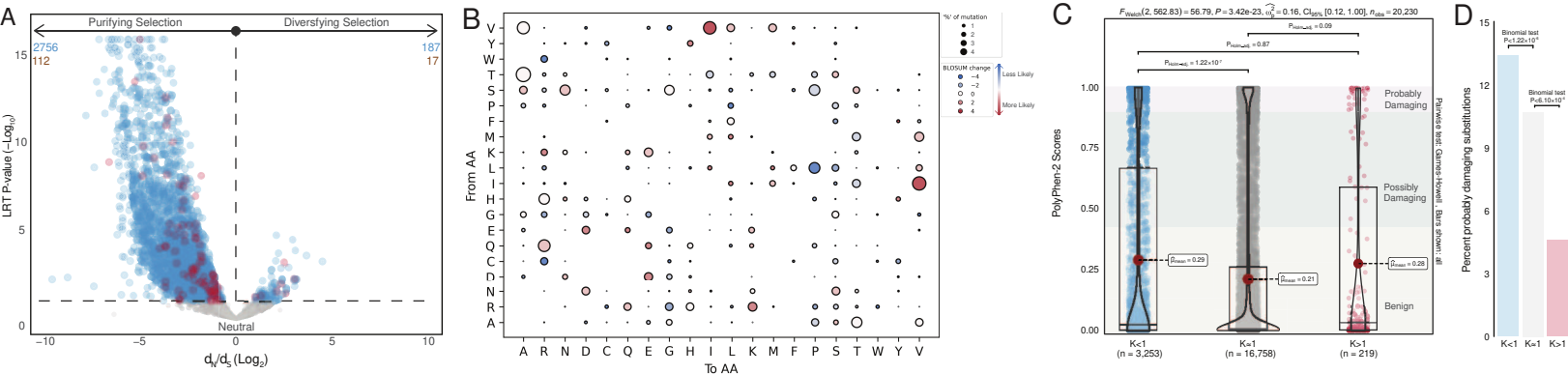
2004 **B.** QQ-plot of expected versus observed LoF FET *P*-values in the MERGE cohort.

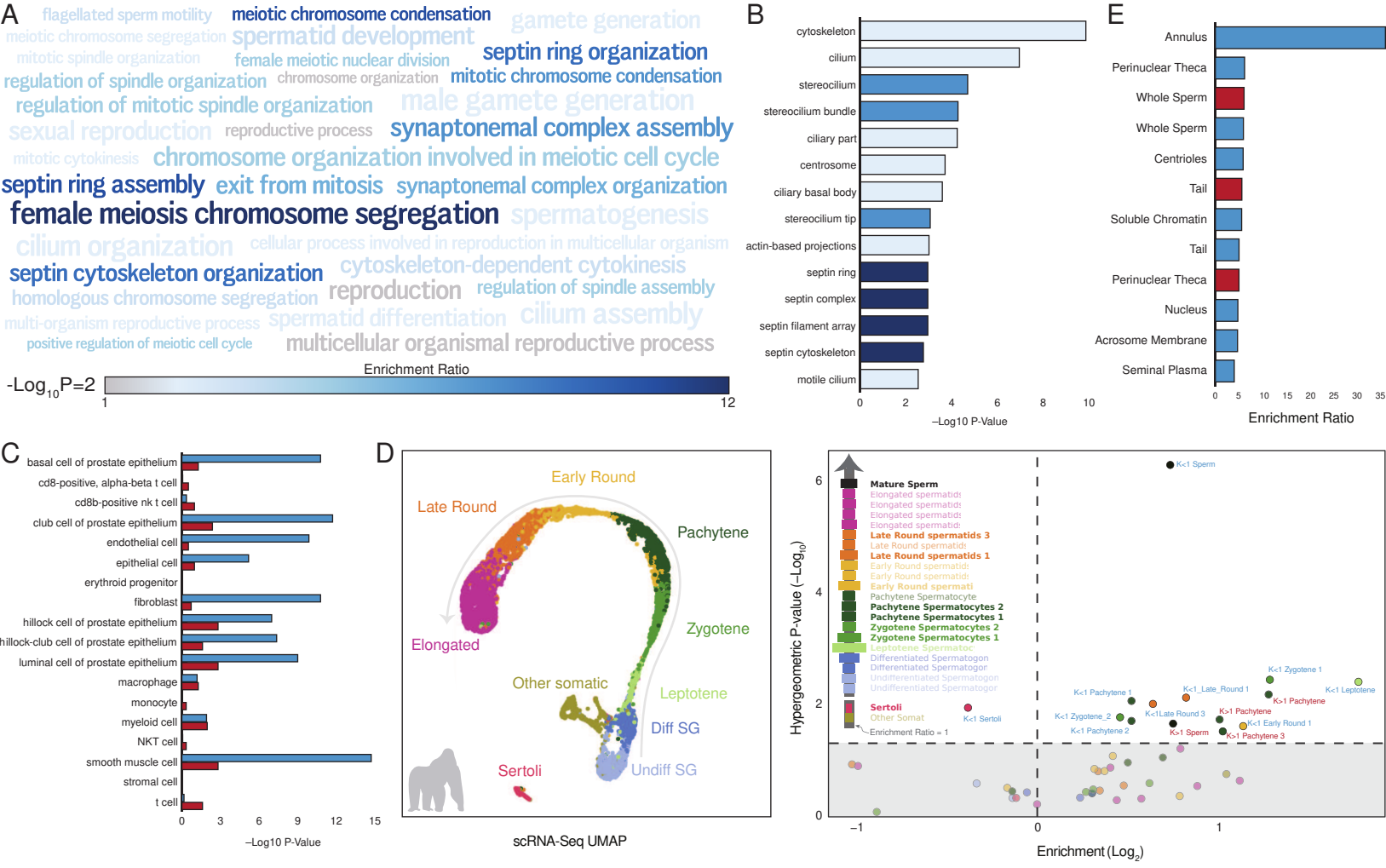
2005 **C.** Gene-level constraint metrics for human orthologs of gorilla K<1 and K≈1 genes, including
2006 the probability of loss-of-function intolerance (pLI), and missense and synonymous Z-
2007 scores from gnomAD data. pLI scores closer to one indicate more intolerance to protein-
2008 truncating variation whereas higher (more positive) Z-scores indicate more intolerance to
2009 variation. Statistics for each comparison are given above each panel.

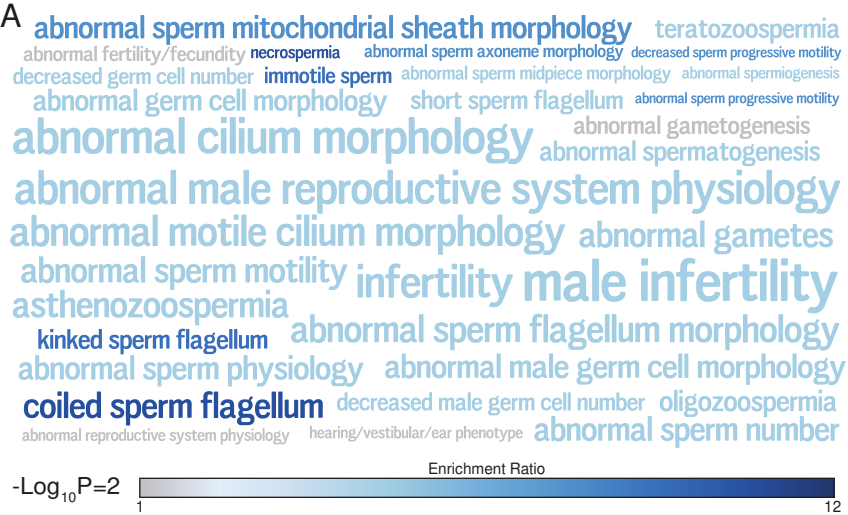

A
$t_{Student}(29) = 6.55$, $p = 3.53e-07$, $\hat{r}_{Pearson} = 0.77$
$CI_{95\%}$ [0.58, 0.88], $n_{pairs} = 31$
Bonobo
Chimpanzee
Gorilla
Testis weight g (Log_{10})
Body weight g (Log_{10})
B
$t_{Student}(12) = 4.07$, $p = 1.54e-03$, $\hat{r}_{Pearson} = 0.76$
$CI_{95\%}$ [0.39, 0.92], $n_{pairs} = 14$
Gorilla
Erect penis length cm (Log_{10})
Body weight g (Log_{10})
C
Testis histology
T
I
Gorilla
Other primates
D
$t_{Student}(19) = 5.14$, $p = 5.86e-05$, $\hat{r}_{Pearson} = 0.76$
$CI_{95\%}$ [0.49, 0.90], $n_{pairs} = 21$
Gorilla
Midpiece volume (um^3)
Residual testis size
E
Chimp
Macaque
Human
Gorilla
VCL (μms^{-1})
F_{esc} (pN)
F
abnormal
normal
Percent of sperm
Human
Chimpanzee
Bonobo
Gorilla
Orangutan

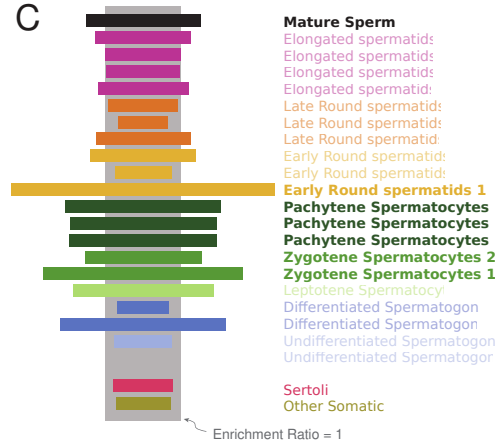

A
Gorilla
AncHomininae
Bonobo
Chimpanzee
Human
Orangutan
Silvery gibbon
White-cheeked gibbon
Amino Acid Changes
20000
B
ABSREL Null Model
Proportion of sites
0.1 1 10
d_N/d_S (ω) Rate Classes
ABSREL Alternate Model
No Positive Selection
Positive Selection
Proportion of sites
0.1 1 10
d_N/d_S (ω) Rate Classes
0.1 1 10
d_N/d_S (ω) Rate Classes
C
ω<1 55
ω>1 96
p>0.05 13336
LRT P-value (−Log10)
15
10
5
0
−10 −5 0 5
d_N/d_S (ω) (Log2)
D
100
80
60
40
20
0
Percent of genes
ω>1 96
Percent of genes with ω>1 only in gorilla
ω<1 ω>1
E
Background Branches
Proportion of sites
0.1 1 10
d_N/d_S (ω)
Foreground Branch(es)
Relaxed K<1
No Change K≈1
Intensified K>1
Proportion of sites
0.3 1 1.5
d_N/d_S (ω)
0.1 1 10
d_N/d_S (ω)
0.01 1 100
d_N/d_S (ω)
F
K<1 578
K>1 144
K=1 12769
CHMP1A
CRTC1
NEFM
TTLL7
DLD
EPHA8
DPY19L1
ZMYM6
SEPT1
RTEL1
SYCP1
UBAP1L
EPHC1
CHMP4B
RPAP3
CCL5
TTL
USP14
BRSK1
TEAD2
EFEMP2
CGNL1
ENTPD6
PLCXD2
NR5A1
CLRN2
ANKRD29
LRT P-value (−Log10)
15
10
5
0
−30 −20 −10 0
K (Log2)
G
100
80
60
40
20
0
Percent of genes
K>1 144
K<1 578
Percent of genes with K≠1 in any other lineage
K=1 K<1 K>1 K>1 & K<1

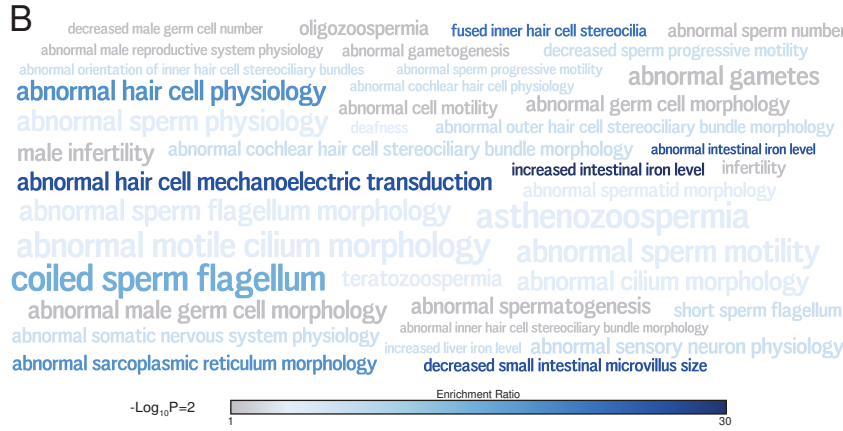

Density
P=7E−05 ***
Coding sequence length (bp)
P=0.0051 **
Transcript length (bp)
P=0.0031 **
Genome span (bp)
P=0.98
5' UTR length (bp)
P=0.00055 ***
3' UTR length (bp)
P=0.19
GC content (%)

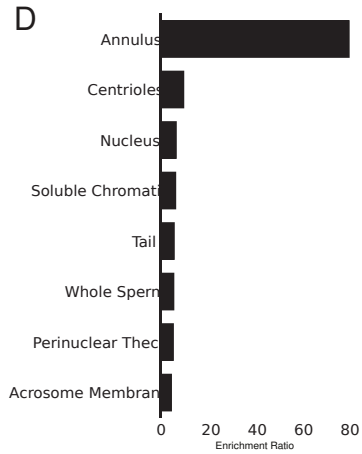

A
Purifying Selection
Diversfying Selection
2756
112
187
17
Neutral
LRT P-value (−Log10)
dN/dS (Log2)
B
From AA
To AA
% of mutation
BLOSUM change
Less Likely
More Likely
C
PolyPhen-2 Scores
Probably Damaging
Possibly Damaging
Benign
K<1 (n = 3,253)
K=1 (n = 16,758)
K>1 (n = 219)
D
Binomial test P<1.22×10⁻¹⁵
Binomial test P<8.10×10⁻¹⁵
Percent probably damaging substitutions
K<1
K=1
K>1

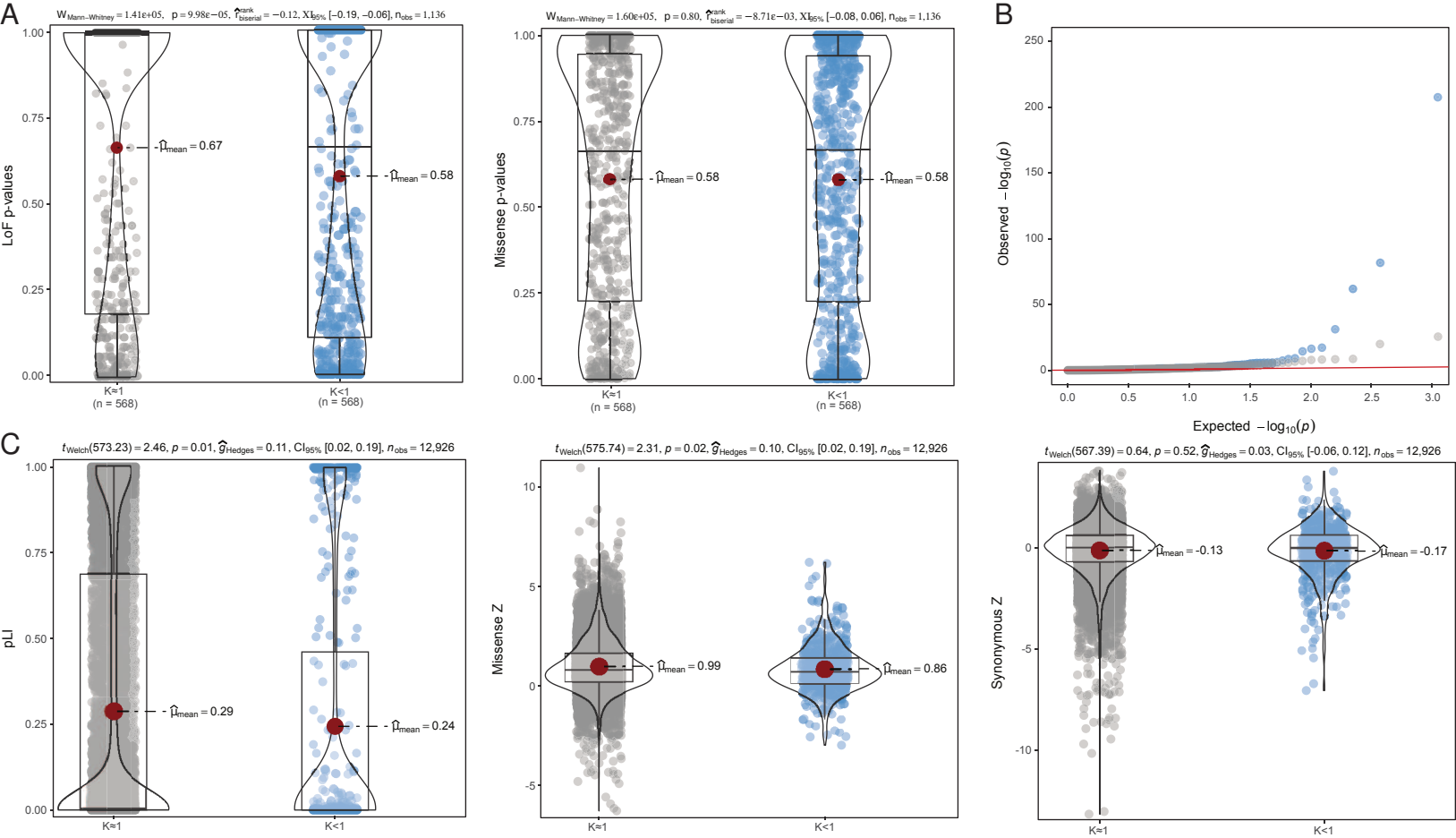

A
flagellated sperm motility
meiotic chromosome condensation
gamete generation
meiotic chromosome segregation
spermatid development
mitotic spindle organization
female meiotic nuclear division
septin ring organization
regulation of spindle organization
chromosome organization
mitotic chromosome condensation
regulation of mitotic spindle organization
male gamete generation
sexual reproduction
reproductive process
synaptonemal complex assembly
mitotic cytokinesis
chromosome organization involved in meiotic cell cycle
septin ring assembly
exit from mitosis
synaptonemal complex organization
female meiosis chromosome segregation
spermatogenesis
cilium organization
cellular process involved in reproduction in multicellular organism
cytoskeleton-dependent cytokinesis
septin cytoskeleton organization
reproduction
regulation of spindle assembly
homologous chromosome segregation
multi-organism reproductive process
spermatid differentiation
cilium assembly
positive regulation of meiotic cell cycle
multicellular organismal reproductive process
Enrichment Ratio
-Log10P=2
1
12
B
cytoskeleton
cilium
stereocilium
stereocilium bundle
ciliary part
centrosome
ciliary basal body
stereocilium tip
actin-based projections
septin ring
septin complex
septin filament array
septin cytoskeleton
motile cilium
0
2
4
6
8
10
–Log10 P-Value
E
Annulus
Perinuclear Theca
Whole Sperm
Whole Sperm
Centrioles
Tail
Soluble Chromatin
Tail
Perinuclear Theca
Nucleus
Acrosome Membrane
Seminal Plasma
0
5
10
15
20
25
30
35
Enrichment Ratio
C
basal cell of prostate epithelium
cd8-positive, alpha-beta t cell
cd8b-positive nk t cell
club cell of prostate epithelium
endothelial cell
epithelial cell
erythroid progenitor
fibroblast
hillock cell of prostate epithelium
hillock-club cell of prostate epithelium
luminal cell of prostate epithelium
macrophage
monocyte
myeloid cell
NKT cell
smooth muscle cell
stromal cell
t cell
0
3
6
9
12
15
–Log10 P-Value
D
Early Round
Late Round
Pachytene
Zygotene
Elongated
Other somatic
Leptotene
Diff SG
Sertoli
Undiff SG
scRNA-Seq UMAP
Mature Sperm
Elongated spermatids
Elongated spermatids
Elongated spermatids
Elongated spermatids
Late Round spermatids 3
Late Round spermatids
Late Round spermatids 1
Early Round spermatids
Early Round spermatids
Early Round spermati
Pachytene Spermatocyte
Pachytene Spermatocytes 2
Pachytene Spermatocytes 1
Zygotene Spermatocytes 2
Zygotene Spermatocytes 1
Leptotene Spermatoc
Differentiated Spermatogon
Differentiated Spermatogon
Undifferentiated Spermatogon
Undifferentiated Spermatogon
Sertoli
Other Somat
K<1 Sertoli
Enrichment Ratio = 1
K<1 Sperm
K<1 Zygotene 1
K<1 Leptotene
K<1 Pachytene 1
K<1_Late_Round 1
K>1 Pachytene
K<1 Zygotene_2
K<1Late Round 3
K>1 Pachytene
K<1 Pachytene 2
K>1 Sperm
K<1 Early Round 1
K>1 Pachytene 3
Hypergeometric P-value (–Log10)
6
4
2
0
–1
0
1
Enrichment (Log2)

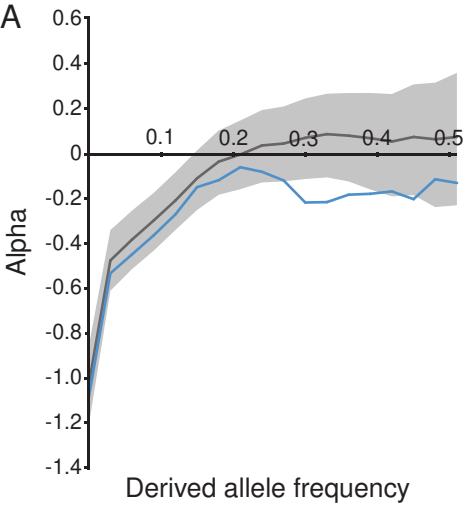

A
abnormal sperm mitochondrial sheath morphology
teratozoospermia
abnormal fertility/fecundity
necrospermia
abnormal sperm axoneme morphology
decreased sperm progressive motility
decreased germ cell number
immotile sperm
abnormal sperm midpiece morphology
abnormal spermiogenesis
abnormal germ cell morphology
short sperm flagellum
abnormal sperm progressive motility
abnormal cilium morphology
abnormal gametogenesis
abnormal spermatogenesis
abnormal male reproductive system physiology
abnormal motile cilium morphology
abnormal gametes
abnormal sperm motility
infertility
male infertility
asthenozoospermia
kinked sperm flagellum
abnormal sperm flagellum morphology
abnormal sperm physiology
abnormal male germ cell morphology
coiled sperm flagellum
decreased male germ cell number
oligozoospermia
abnormal reproductive system physiology
hearing/vestibular/ear phenotype
abnormal sperm number
-Log10P=2
Enrichment Ratio
1
12

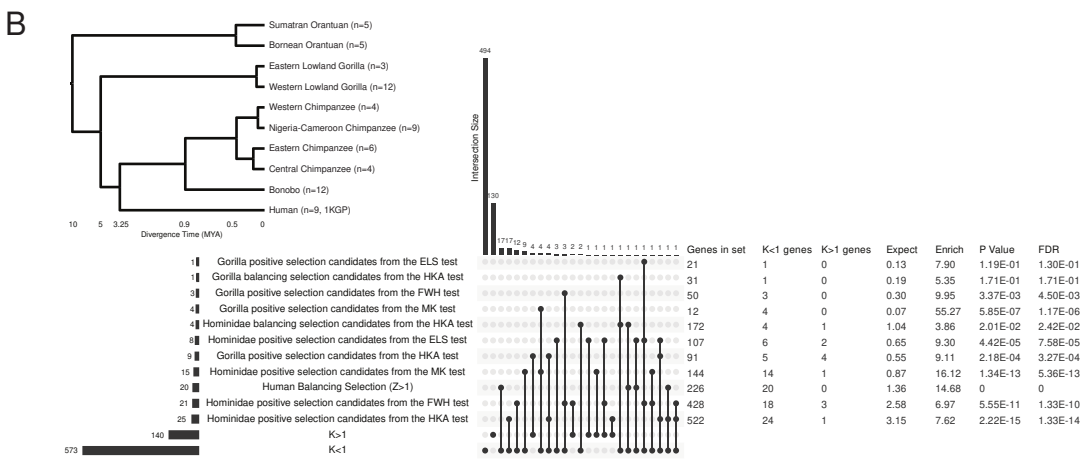

C
Mature Sperm
Elongated spermatids
Elongated spermatids
Elongated spermatids
Elongated spermatid
Late Round spermatids
Late Round spermatids
Late Round spermatid
Early Round spermatids
Early Round spermatids
Early Round spermatids 1
Pachytene Spermatocytes
Pachytene Spermatocytes
Pachytene Spermatocytes
Zygotene Spermatocytes 2
Zygotene Spermatocytes 1
Leptotene Spermatocy
Differentiated Spermatogon
Differentiated Spermatogon
Undifferentiated Spermatogon
Undifferentiated Spermatogor
Sertoli
Other Somatic
Enrichment Ratio = 1


B
decreased male germ cell number
oligozoospermia
fused inner hair cell stereocilia
abnormal sperm number
abnormal male reproductive system physiology
abnormal gametogenesis
decreased sperm progressive motility
abnormal orientation of inner hair cell stereociliary bundles
abnormal sperm progressive motility
abnormal gametes
abnormal hair cell physiology
abnormal cochlear hair cell physiology
abnormal cell motility
abnormal germ cell morphology
abnormal sperm physiology
deafness
abnormal outer hair cell stereociliary bundle morphology
male infertility
abnormal cochlear hair cell stereociliary bundle morphology
abnormal intestinal iron level
abnormal hair cell mechanoelectric transduction
increased intestinal iron level
infertility
abnormal spermatid morphology
abnormal sperm flagellum morphology
asthenozoospermia
abnormal motile cilium morphology
abnormal sperm motility
coiled sperm flagellum
teratozoospermia
abnormal cilium morphology
abnormal male germ cell morphology
abnormal spermatogenesis
short sperm flagellum
abnormal inner hair cell stereociliary bundle morphology
abnormal somatic nervous system physiology
increased liver iron level
abnormal sensory neuron physiology
abnormal sarcoplasmic reticulum morphology
decreased small intestinal microvillus size
-Log10P=2
Enrichment Ratio
1
30


D
Annulus
Centrioles
Nucleus
Soluble Chromati
Tail
Whole Sperm
Perinuclear Thec
Acrosome Membran
0
20
40
60
80
Enrichment Ratio

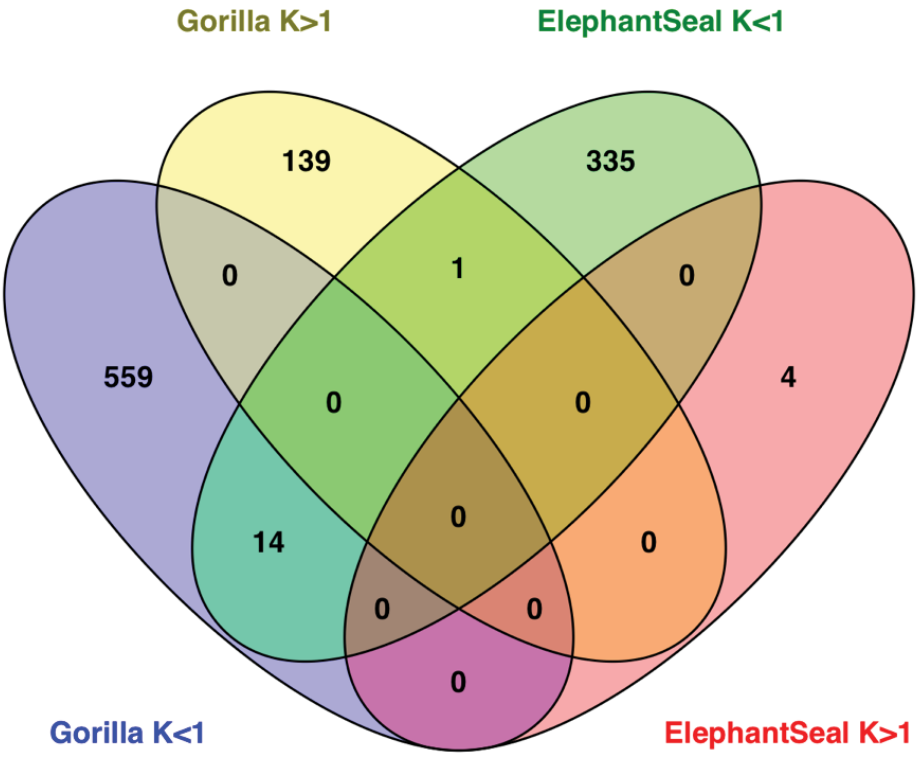

A
$W_{Mann-Whitney} = 1.41e+05$, $p = 9.98e-05$, $\hat{r}^{rank}_{biserial} = -0.12$, $CI_{95\%}$ [−0.19, −0.06], $n_{obs} = 1,136$
LoF p-values
$\hat{\mu}_{mean} = 0.67$
$\hat{\mu}_{mean} = 0.58$
K≈1 (n = 568)
K<1 (n = 568)
$W_{Mann-Whitney} = 1.60e+05$, $p = 0.80$, $\hat{r}^{rank}_{biserial} = -8.71e-03$, $CI_{95\%}$ [−0.08, 0.06], $n_{obs} = 1,136$
Missense p-values
$\hat{\mu}_{mean} = 0.58$
$\hat{\mu}_{mean} = 0.58$
B
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
C
$t_{Welch}(573.23) = 2.46$, $p = 0.01$, $\hat{g}_{Hedges} = 0.11$, $CI_{95\%}$ [0.02, 0.19], $n_{obs} = 12,926$
pLI
$\hat{\mu}_{mean} = 0.29$
$\hat{\mu}_{mean} = 0.24$
$t_{Welch}(575.74) = 2.31$, $p = 0.02$, $\hat{g}_{Hedges} = 0.10$, $CI_{95\%}$ [0.02, 0.19], $n_{obs} = 12,926$
Missense Z
$\hat{\mu}_{mean} = 0.99$
$\hat{\mu}_{mean} = 0.86$
$t_{Welch}(567.39) = 0.64$, $p = 0.52$, $\hat{g}_{Hedges} = 0.03$, $CI_{95\%}$ [-0.06, 0.12], $n_{obs} = 12,926$
Synonymous Z
$\hat{\mu}_{mean} = -0.13$
$\hat{\mu}_{mean} = -0.17$



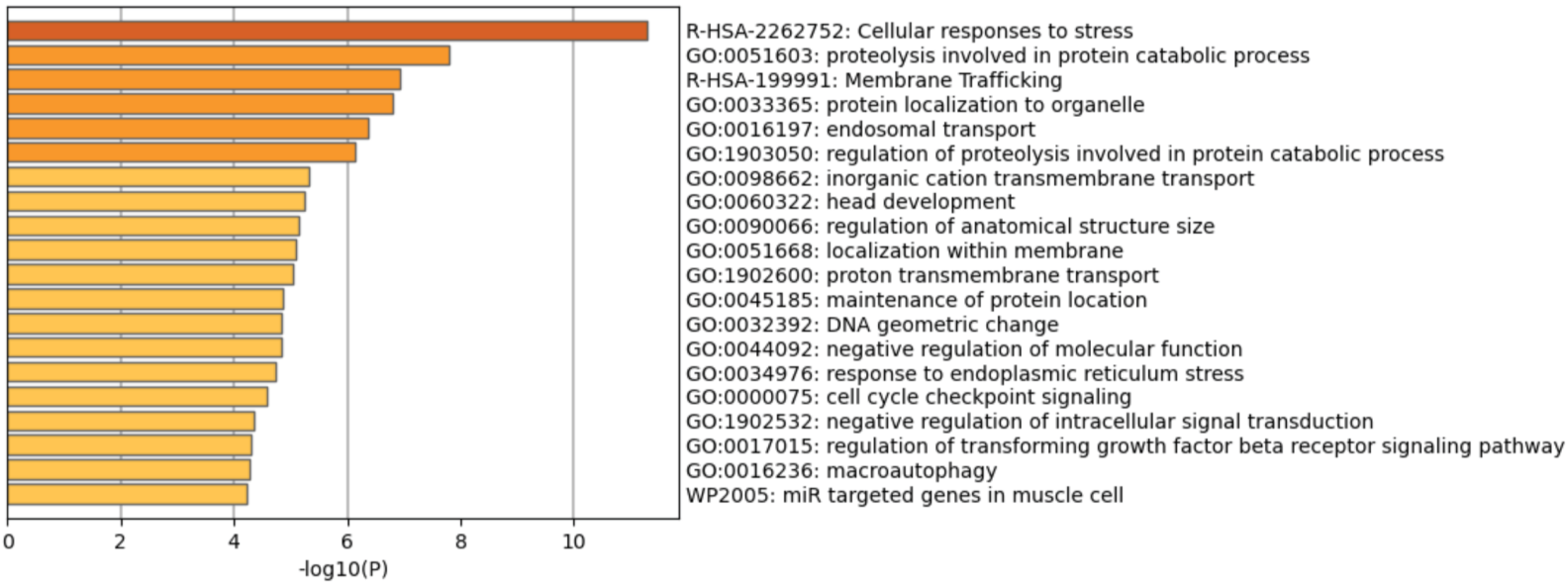


Genes in set	K<1 genes	K>1 genes	Expect	Enrich	P Value	FDR
21	1	0	0.13	7.90	1.19E-01	1.30E-01
31	1	0	0.19	5.35	1.71E-01	1.71E-01
50	3	0	0.30	9.95	3.37E-03	4.50E-03
12	4	0	0.07	55.27	5.85E-07	1.17E-06
172	4	1	1.04	3.86	2.01E-02	2.42E-02
107	6	2	0.65	9.30	4.42E-05	7.58E-05
91	5	4	0.55	9.11	2.18E-04	3.27E-04
144	14	1	0.87	16.12	1.34E-13	5.36E-13
226	20	0	1.36	14.68	0	0
428	18	3	2.58	6.97	5.55E-11	1.33E-10
522	24	1	3.15	7.62	2.22E-15	1.33E-14

A


Gorilla K>1
ElephantSeal K<1
139
335
0
1
0
559
0
0
4
14
0
0
0
0
0
Gorilla K<1
ElephantSeal K>1

B


R-HSA-2262752: Cellular responses to stress
GO:0051603: proteolysis involved in protein catabolic process
R-HSA-199991: Membrane Trafficking
GO:0033365: protein localization to organelle
GO:0016197: endosomal transport
GO:1903050: regulation of proteolysis involved in protein catabolic process
GO:0098662: inorganic cation transmembrane transport
GO:0060322: head development
GO:0090066: regulation of anatomical structure size
GO:0051668: localization within membrane
GO:1902600: proton transmembrane transport
GO:0045185: maintenance of protein location
GO:0032392: DNA geometric change
GO:0044092: negative regulation of molecular function
GO:0034976: response to endoplasmic reticulum stress
GO:0000075: cell cycle checkpoint signaling
GO:1902532: negative regulation of intracellular signal transduction
GO:0017015: regulation of transforming growth factor beta receptor signaling pathway
GO:0016236: macroautophagy
WP2005: miR targeted genes in muscle cell
0
2
4
6
8
10
-log10(P)