Sex and Power: sexual dimorphism in trait variability and its eco-evolutionary and statistical implications

Zajitschek, S.R.K.[1,2 *], Zajitschek, F.[1], Bonduriansky, R. [1], Brooks, R.C. [1], Cornwell, W.[1], Falster, D.S.[1], Lagisz, M. [1], Mason, J.[3], Senior, A. M. [4], Noble, D. A. W.[1,5 #], S. Nakagawa[1 * #]

[1] Evolution & Ecology Research Center, School of Biological, Earth, and Environmental Sciences, University of New South Wales, Sydney, 2052, NSW, Australia.

[2] Liverpool John Moores University, School of Biological and Environmental Sciences, James Parsons Building, 3 Byrom Way, Liverpool L3 3 AF, UK.

[3] European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

[4] University of Sydney, Charles Perkins Centre, School of Life and Environmental Sciences, School of Mathematics and Statistics, Sydney, 2006, NSW, Australia.

[5] Research School of Biology, Australian National University, Canberra, 2601, ACT, Australia.

# equal contributions as senior authors.

*corresponding authors:

S.Z. (susi.zajitschek@gmail.com) and S.N. (s.nakagawa@unsw.edu.au)

ORCID IDs:

SZ 0000-0003-4676-9950; FZ 0000-0001-6010-6112; RB 0000-0002-5786-6951; RCB 0000-0001-6926-0781; WC 0000-0003-4080-4073, DSF 0000-0002-9814-092X, ML 0000-0002-3993-6127; JM 0000-0002-2796-5123; AS 0000-0001-9805-7280; DN 0000-0001-9460-8743; SN 0000-0002-7765-5182

ABSTRACT

Biomedical and clinical sciences are experiencing a renewed interest in the fact that males and females differ in many anatomic, physiological, and behavioral traits. Sex differences in trait variability, however, are yet to receive similar recognition. In medical science, mammalian females are assumed to have higher trait variability due to estrous cycles (the 'estrus-mediated variability hypothesis'); historically in biomedical research, females have been excluded for this reason. Contrastingly, evolutionary theory and associated data support the 'greater male variability hypothesis'. Here, we test these competing hypotheses in 218 traits measured in >26,900 mice, using meta-analysis methods. Neither hypothesis could universally explain patterns in trait variability. Sex-bias in variability was trait-dependent. While greater male variability was found in morphological traits, females were much more variable in immunological traits. Sex-specific variability has eco-evolutionary ramifications including sex-dependent responses to climate change, as well as statistical implications including power analysis considering sex difference in variance.

Keywords

Sex inequality, gender difference, sexual selection, meta-regression, power analysis

INTRODUCTION

Sex differences arise because selection acts on the two sexes differently, especially on traits associated with mating and reproduction (1). Therefore, sex differences are widespread, a fact which is unsurprising to any evolutionary biologist. However, scientists in many (bio-)medical fields have not necessarily regarded sex as a biological factor of intrinsic interest (2–7). Therefore, many (bio-)medical studies have only been conducted with male subjects. Consequently, our knowledge is biased. For example, we know far more about drug efficacy in male compared to female subjects, contributing to a poor understanding of how the sexes respond differently to medical interventions (8). This gap in knowledge is predicted to lead to overmedication and adverse drug reactions in women (9). Only recently have (bio-)medical scientists started considering sex differences in their research (10–16). Indeed, the National Institutes of Health (NIH) have now implemented new guidelines for animal and human research study designs, requiring that sex be included as a biological variable (2, 17, 18).

[Figure 1 here]

When comparing the sexes, biologists generally focus on mean differences in trait values, placing little or no emphasis on sex differences in trait variability (see Fig. 1 for a diagram explaining differences in means and variances). Despite this, two hypotheses exist that explain why trait variability might be expected to differ between the sexes. Interestingly, these two hypotheses make opposing predictions.

First, the "estrus-mediated variability hypothesis" (Fig. 2), which emerged in the (bio-)medical research field, assumes that the female estrous cycle (see for example 6, 19) causes higher variability across traits in female subjects. A wide range of labile traits are presumed to co-vary with physiological changes that are induced by reproductive hormones. High variability is, therefore, expected to be particularly prominent when the stage of the estrous cycle is unknown and unaccounted for. This higher trait variability, resulting from females being at different stages of their estrous cycle, is the main reason for why female research subjects are often excluded from biomedical research trials, especially in the neurosciences, physiology and pharmacology (18). Female exclusion has traditionally been justified based on the grounds that including females in empirical research leads to a loss of statistical power, or that animals must be sampled across the estrous cycle for one to make valid conclusions, requiring more time and resources.

Second, the "greater male variability hypothesis" suggests males exhibit higher trait variability because of two different mechanisms. The first mechanism is based on males being the heterogametic sex in mammals. Mammalian females possess two X chromosomes, leading to an 'averaging' of trait expression across the genes on each chromosome. In contrast, males exhibit greater variance because a single gene on the X chromosome is likely to lead to more extreme trait values (20). The second mechanism is based on males being under stronger sexual selection (21–23). Empirical evidence supports higher variability of traits that are sexually

selected, often harbouring high genetic variance and being condition-dependent, which makes sense as 'condition' as a trait is likely to be based on 1000s of loci (24, 25). Thus, higher genetic and, thus, phenotypic variance resulting from sexual selection is less general because it is only expected to apply to sexually selected traits. In mammals, it is likely that both mechanisms are operating concomitantly. So far, the "greater male variability hypothesis" has gained some support in the evolutionary and psychological literature (20, 26).

[Figure 2 here]

Here we conduct the first comprehensive test of the greater male variability and estrus-mediated variability hypotheses in mice (Fig. 2; cf. , 20, 27–31), examining sex differences in variance across 218 traits in 26,916 animals. To this end, we carry out a series of meta-analyses in two steps (SI Appendix Fig. S1.1). First, we quantify the natural logarithm of the male to female coefficients of variation, CV, or relative variance (lnCVR) for each cohort (population) of mice, for different traits, along with the variability ratio of male to female standard deviations, SD, on the log scale (lnVR, following 32, see Fig. 1). Then, we analyze these effect sizes to quantify sex bias in variance for each trait using meta-analytic methods. To better understand our results and match them to previously reported sex differences in trait means (4), we also quantify and analyze the log response ratio (lnRR). Then, we statistically amalgamate the trait-level results to test our hypotheses and to quantify the degree of sex biases in and across nine functional trait groups (for details on the grouping, see below). Our meta-analytic approach allows easy interpretation and comparison with earlier and future studies. Further, the proposed method using lnCVR (and lnVR) is probably the only practical method to compare variability between two sexes within and across studies (32, 33), as far as we are aware. Also, the use of a ratio (i.e. lnRR, lnVR, lnCVR) between two groups (males and females) naturally controls for different units (e.g., cm, g, ml) and also for changes in traits over time and space.

RESULTS

Data characteristics and workflow

We used a dataset compiled by the International Mouse Phenotyping Consortium (34) (IMPC, dataset acquired 6/2018). To gain insight into systematic sex differences, we only included data of wildtype-strain adult mice, between 100 and 500 days of age. We removed cases with missing data, and selected measurements that were closest to 100 days of age (young adult) when multiple measurements of the same trait were available. To obtain robust estimates of sex differences, we only used data on traits that were measured in at least two different institutions (see workflow diagram, SI Appendix Fig. S1.1 A).

Our data set comprised 218 continuous traits (after initial data cleaning and pre-processing; SI Appendix Fig. S1.1 A-D). It contains information from 26,916 mice from 9 wildtype strains that were studied across 11 institutions. We combined mouse strain/institution information to

create a biological grouping variable (referred to as "population" in SI Appendix Fig. S1.1 B; see also Table S6.1 for details), and the mean and variance of a trait for each population was quantified. We assigned traits according to related procedures into functionally and/or procedurally related trait groups to enhance interpretability (referred to as "functional groups" hereafter; see also SI Appendix Fig. S1.1 G). Our nine functional trait groups were behaviour, morphology, metabolism, physiology, immunology, hematology, heart, hearing and eye (for the rationale of these functional groups and related details, see Methods and SI Appendix Table S6.3).

[Figure 3 here]

Testing the two hypotheses

We found that some means and variabilities of traits were biased towards males (i.e. 'male-biased', hereafter; "turquoise" shaded traits, Fig. 3), but others towards females (i.e. 'female-biased', hereafter; "orange" shading, Fig. 3) within all functional groups. These sex-specific biases occur in mean trait sizes and also in our measures of trait variability. There were strong positive relationships between mean and variance across traits (r > 0.94 on the log scale; SI Appendix Fig. S2.1), and therefore, we report the results of lnCVR, which controls for differences in means, in the main text. Results on lnVR are presented in the electronic supplementary material (SI Appendix Fig. S5.1 and S5.2).

There was no consistent pattern in which sex has more variability (lnCVR) in the examined traits (left panel in Fig. 3A). Our meta-analytic results also did not support a consistent pattern of either higher male variability or higher female variability (see Fig. 3B, left panel: "All" indicates that across all traits and functional groups, there was no significant sex bias in variances; lnCVR = 0.005, 95% confidence interval, 95% CI = [-0.009 to 0.018]). However, there was high heterogeneity among traits (I^2 = 76.5 %, SI Appendix Table S6.4; see also SI Appendix Table S6.5), indicating sex differences in variability are trait-dependent, corroborating our general observation that variability in some traits was male-based but others female-biased (Fig. 3A).

As expected, specific functional trait groups showed significant sex-specific bias in variability (Fig. 3B). The variability among-traits within a functional group was lower than that of all the traits combined (SI Appendix Table S6.4). For example, males exhibited an 8.05% increase in CV relative to females for morphological traits (lnCVR = 0.077; CI = [0.041 to 0.113], I^2= 67.3%), but CV was female-biased for immunological traits (6.59% higher in females, lnCVR = -0.068, CI =[-0.098 to 0.038], I^2 = 40.8%) and eye morphology (7.85% higher in females, lnCVR = -0.081, CI =[-0.147 to (- 0.016)], I^2 = 49.8%).

The pattern was similar for overall sexual dimorphism in mean trait values (here, a slight male bias is indicated by larger "turquoise" than "orange" areas; Fig. 3A, right and Fig. 3B, lnRR: "All", lnRR = 0.012, CI = [-0.006 to 0.31]). Trait means (lnRR) were 7% larger for males (lnRR = 0.067; CI = [0.007 to 0.128]) in morphological traits and 15.3 % larger in males for metabolic traits

(lnRR = 0.142; CI = [0.036 to 0.248]). In contrast, females had 5.59 % [lnRR = 0.057, CI = [-0.107 to (-0.007)] larger means than those of males for immunological traits. We note that these meta-analytic estimates were accompanied by very large between-trait heterogeneity values (morphology I^2 = 99.7%, metabolism I^2 = 99.4%, immunology I^2 = 96.2; see SI Appendix Table S6.4), indicating that even within the same functional groups, the degree and direction of sex-bias in the mean was not consistent among traits.

[Figure 4 here]

DISCUSSION

We tested competing predictions from two hypotheses explaining why sex-biases in trait variability exist. Neither the 'greater male variability' hypothesis nor the 'estrus-mediated variability' hypothesis explain the observed patterns in sex-biased trait variation on their own. Therefore, our results add further empirical weight to calls that question the basis for the routine exclusion of one sex in biomedical research based on the estrus-mediated variability hypothesis (3, 5–7, 30). It is important to know that for each trait we estimated the mean effect size (i.e. lnCVR) over strains and locations. As such, our results may not necessarily apply to every group of mice, which may or may not result in stronger support for either of the two hypotheses.

Greater male variability vs. estrus-mediated variability?

Evolutionary biologists commonly expect greater variability in the heterogametic sex than the homogametic sex. In mammals, males are heterogametic, and hence are expected to exhibit higher trait variability compared to females, which is also consistent with an expectation from sexual selection theory (20). Our results provide only partial support for the greater male variability hypothesis, because the expected pattern only manifested for morphological traits (see Fig. 3 & 4). This result corroborates a previous analysis across animals, which found that the heterogametic sex was more variable in body size (20). However, our data do not support the conclusion that higher variability in males occurs across all traits, including for many other morphological traits.

[Figure 4]

The estrus-mediated variability hypothesis was, at least until recently (6, 13), regularly used as a rationale for including only male subjects in many biomedical studies. So far, we know very little about the relationship between hormonal fluctuations and general trait variability within and among female subjects. Our results are consistent with the estrus-mediated variability hypothesis for immunological traits only. Immune responses can strongly depend on sex hormones (35, 36), which may explain higher female variability in these traits. However, if estrus status affects traits through variation in hormone levels, we would expect to also find

higher female variability in physiological and hematological traits. This was not the case in our dataset. Interestingly, however, eye morphology (structural traits, which should fluctuate little across the estrous cycle) also appeared to be more variable in females than males, but little is known about sex differences in ocular traits in general (37, 38). Overall, we find no consistent support for the female estrus-mediated variability hypothesis.

In line with our findings, recent studies have refuted the prediction of higher female variability (6, 13, 19, 30, 31). For example, several rodent studies have found that males are more variable than females (6, 13, 30, 31, 39, 40). Further studies should investigate whether higher female variability in immunological traits is indeed due to the estrous cycle, or generally because of greater between-individual variation (cf. Fig. 2).

In general, we found many traits to be sexually dimorphic (Fig. 4) in accordance with the previous study, which used the same database (4). Although the original study also provided estimates for sex differences in traits both with and without controlling for weight (we did not control for weight; cf. , 41). More specifically, males are larger than females, while females have higher immunological parameters (see Fig. 4). Notably, most sexually dimorphic trait means also show the greatest differences in trait variance (Fig. 3 & Fig. 4). Indeed, theory predicts that sexually selected traits (e.g., larger body size for males due to male-male competition) are likely more variable, as these traits are often condition dependent (24). Therefore, this sex difference in variability could be more pronounced under natural conditions compared to laboratory settings. This relationship may explain why male-biased morphological traits are larger and more variable.

Eco-evolutionary implications

We have used lnCVR values to compare phenotypic variability (CV) between the sexes. When lnCVR is used for fitness-related traits, it can signify sex differences in the 'opportunity for selection' between females and males (24). If we assume that phenotypic variation (i.e. variability in traits) has a heritable basis, then large ratios of lnCVR may indicate differences in the evolutionary potential of each sex to respond to selection, at least in the short term (42). We note, however, that in our study, lnCVR reflects sex differences in trait variability within strains, such that the variability differences we observe between the sexes may be partially the result of phenotypic plasticity.

Demographic parameters, such as age-dependent mortality rate (43), are often different for each sex. Indeed, recognition of this fact has resulted in population dynamic models taking these widely observed sex-differences into account (44, 45). For example, a study on European sparrowhawks found that variability in mortality was higher in females compared to males (46). In this species, sex-specific variation affects age-dependent mortality and results in higher female life expectancy. As such, explicitly modelling sex difference in trait variability could lead to different conclusions compared to traditional modelling approaches.

Statistical and practical implications

It is now mandatory to include both sexes in biomedical experiments and clinical trials funded by the NIH, unless there exists strong justification against the inclusion of both sexes (18, 47). In order to conduct meaningful research and make sound clinical recommendations for both male and female patients, it is necessary to understand not only how trait means, but also how trait variances differ between the sexes. If one sex is systematically more variable in a trait of interest than the other, then experiments should be designed to accommodate relative differences in statistical power between the sexes (which has not been considered before, see 3, 5–7). For example, given a limited number of animal subjects in an experiment measuring immunological traits, a balanced sex ratio may not be optimal. Female immunological traits are generally more variable (i.e. higher CV and SD). If we assume that responses to an experimental treatment will be similar between the sexes for this functional trait group, we will require more females to achieve the same statistical power as for the males.

To help researchers adjust their sex-specific sample size to achieve optimal statistical power, we provide an online tool (ShinyApp; https://bit.ly/sex-difference). This tool may serve as a starting point for checking baseline variability for each sex in mice. The sex bias (indicated by the % difference between the sexes) is provided for separate traits, procedures, and functional groups. These meta-analytic results are based on our analyses of more than 2 million rodent data points, from 26,916 individual mice. We note that, however, variability in a trait measured in untreated individuals maintained under carefully standardized environmental conditions, as reported here, may not directly translate into the same variability when measured in experimentally treated individuals, or individuals exposed to a range of environments (i.e. natural populations or human cohorts). Further, these estimates are overall mean differences across strains and locations. Therefore, these may not be particularly informative if one's experiment only includes one specific strain. However, we point out that our estimates may be useful in the light of a recent recommendation of using 'heterogenization' where different strains are mixed to increase the robustness of experimental results (48). Also, even in the case of using a particular strain, our tool can provide potentially useful benchmarks.

Importantly, when two groups (e.g., males and females) show differences in variability, we violate homogeneity of variance or homoscedasticity assumptions. Such a violation is detrimental because it leads to a higher Type I error rate. Therefore, we should consider incorporating heteroscedasticity (different variances) explicitly or using robust estimators of variance (also known as 'the sandwich variance estimator') to prevent an inflated Type I error rate (49), especially when we compare traits between the sexes.

Conclusion

238 We have shown that sex biases in variability occur in many mouse traits, but that the directions
239 of those biases differ between traits. Neither the 'greater male variability' nor the 'estrus-
240 mediated variability' hypothesis provides a general explanation for sex-differences in trait
241 variability. Instead, we have found that the direction of the sex bias varies across traits and
242 among trait types (Fig. 3 & 4). Our findings have important ecological and evolutionary
243 ramifications. If the differences in variability correspond to the potential of each sex to respond
244 to changes in specific environments, this sex difference needs to be incorporated into
245 demographic and population-genetic modelling. Moreover, in the (bio-)medical field, our
246 results should inform decisions during study design by providing more rigorous power analyses
247 that allow researchers to incorporate sex-specific differences for sample size. We believe that
248 taking sex-differences in trait variability into account will help avoid misleading conclusions and
249 provide new insights into sex differences across many areas of biological and bio-medical
250 research. Ultimately, such considerations will not only better our knowledge, but also close the
251 current gaps in our biased knowledge (50).

252 METHODS

253 *Data selection and process*

254 The IMPC (International Mouse Phenotyping Consortium) provides a comprehensive catalogue
255 of mammalian gene function for investigating the genetics of health and disease, by
256 systematically collecting phenotypes of knock-out and wild type mice. To investigate
257 differences in trait variability between the sexes, we only considered the data for wild-type
258 control mice. We retrieved the dataset from the IMPC server in June 2018 and filtered it to
259 contain non-categorical traits for wildtype mice. The initial dataset comprised over 2,500,000
260 data points for 340 traits. In cases where multiple measurements were taken over time, data
261 cleaning started with selecting single measurements for each individual and trait. In these cases,
262 we selected the measurement closest to "100 days of age". All data are from unstaged females
263 (with no information about the stage of their estrous cycle). We excluded data for juvenile and
264 unsexed mice (SI Appendix Fig. S1.1 A; this data set and scripts can be found on
265 https://bit.ly/code-mice-sex-diff; raw data: https://doi.org/10.5281/zenodo.3759701).

266 *Grouping and effect size calculation*

267 We created a grouping variable called "population" (SI Appendix Fig. S1.1 B). A population
268 comprised a group of individuals belonging to a distinct wild-type strain maintained at one
269 particular location (institution); populations were identified for every trait of interest. Our data
270 were derived from 11 different locations/institutions, and a given location/institution could
271 provide data on multiple populations (see SI Appendix Table S6.1 for details on numbers of
272 strains and institutions). We included only populations that contained data points for at least 6
273 individuals, and which had information for members of both sexes; further, these populations
274 for a particular trait had to come from at least two institutions to be eligible for inclusion. After

this selection process, the dataset contained 2,300,000 data points across 232 traits. Overall, we meta-analysed traits with between 2–18 effect sizes (mean = 9.09 effects, SD = 4.47). However, each meta-analysis contained a total number of individual mice that ranged from 83/91 to 13467/13449 (males/females). While a minimum of N = 6 mice were used to create effect sizes for any given group (male or female), in reality samples sizes of male / female groups were much larger (males: mean = 396.66 (SD = 238.23), median = 465.56; females: mean = 407.35 (SD = 240.31), median = 543.89). We used the function *escalc* in the R package, *metafor* (51) to obtain lnCVR, lnVR and lnRR and their corresponding sampling variance for each trait for each population; we worked in the R environment for data cleaning, processing and analyses (52, version 3.6.0; for the versions of all the software packages used for this article and all the details and code for the statistical analyses, see the electronic supplements). As mentioned above, the use of ratio-based effect sizes, such as lnCVR, lnVR and lnRR, controls for baseline changes over time and space, assuming that these changes affect males and females similarly. However, we acknowledge that we could not test this assumption.

Meta-analyses: overview

We conducted meta-analyses at two different levels (SI Appendix Fig. S1.1 C-J). First, we conducted a meta-analysis for each trait for all three effect size types (lnRR, lnVR and lnCVR), calculated at the 'population' level (i.e. using population as a unit of analysis). Second, we statistically amalgamated overall effect sizes estimated at each trait (i.e. overall trait means as a unit of analysis) after accounting for dependence among traits. In other words, we conducted second-order meta-analyses (53). We used the second-order meta-analyses for three different purposes: A) estimating overall sex biases in variance (lnCVR and lnVR) and mean (lnRR) in the nine functional groups (for details, see below) and in all these groups combined (the overall estimates); B) visualizing heterogeneities across populations for the three types of effect size in the nine functional trait groups, which complemented the first set of analyses (SI Appendix Fig. S1.1 I, Table S6.6); and C) when traits were found to be significantly sex-biased, grouping such traits into either male-biased and female-biased traits, and then, estimating overall magnitudes of sex bias for both sexes again for the nine functional trait groups. Only the first second-order meta-analysis (A) directly related to the testing of our hypotheses, results of B and C are found in the supplemental tables and Figures, and reported in our freely accessible code.

Meta-analyses: population as an analysis unit

To obtain degree of sex bias for each trait mean and variance (SI Appendix Fig. S1.1 C), we used the function *rma.mv* in the R package *metafor* (51) by fitting the following multilevel meta-analytic model, an extension of random-effects models (sensu 54):

$$ES_i \sim 1 + (1 \mid Strain_j) + (1 \mid Location_k) + (1 \mid Unit_i) + Error_i,$$

where 'ES_i' is the *i*th effect size (i.e. lnCVR, lnVR and lnRR) for each of 232 traits, the '1' is the overall intercept (other '1's are random intercepts for the following random effects), '$Strain_j$' is

312 a random effect for the *j*th strain of mice (among 9 strains), 'Location$_k$' is a random effect for
313 the *k*th location (among 11 institutions), 'Unit$_i$' is a residual (or effect-size level or 'population-
314 level' random effect) for the *i*th effect size, 'Error$_i$' is a random effect of the known sampling
315 error for the *i*th effect size. Given the model above, meta-analytic results had two components:
316 1) overall means with standard errors (95% confidence intervals), and 2) total heterogeneity
317 (the sum of the three variance components, which is estimated for the random effects). Note
318 that overall means indicate average (marginalised) effect sizes over different strains and
319 locations and total heterogeneities reflect variation around overall means due to different
320 strains and locations.

321 We excluded traits which did not carry useful information for this study (i.e. fixed traits, such as
322 number of vertebrae, digits, ribs and other traits that were not variable across wildtype mice;
323 note that this may be different for knock-down mutant strains) or where the meta-analytic
324 model for the trait of interest did not converge, most likely due to small sample size from the
325 dataset (14 traits, see SI Appendix, for details: Meta-analyses; 1. Population as analysis unit).
326 We therefore obtained a dataset containing meta-analytic results for 218 traits at this stage, to
327 use for our second-order meta-analyses (SI Appendix Fig. S1.1 D).

328 ***Meta-analyses: accounting for correlated traits***

329 Our dataset of meta-analytic results included a large number of non-independent traits. To
330 account for dependence, we identified 90 out of 218 traits, and organized them into 19 trait
331 sub-groups (containing 2-10 correlated traits, see SI Appendix Fig. S1.1 E). For example, many
332 measurements (i.e. traits) from hematological and immunological assays were hierarchically
333 clustered or overlapped with each other (e.g., cell type A, B and A+B). We combined the meta-
334 analytic results from 90 traits into 19 meta-analytic results (Fig. 3F) using the function *robu* in
335 the R package, *robumeta* with the assumption of sampling errors being correlated with the
336 default value of r = 0.8 (55). Consequently, our final dataset for secondary meta-analyses
337 contained 147 traits (i.e. the newly condensed 19 plus the remaining 128 independent traits,
338 see SI Appendix Fig 1.1 , Table S6.2), which we assume to be independent of each other.

339

340 ***Second-order meta-analyses: trait as an analysis unit***

341 We created our nine overarching functional groups of traits (SI Appendix Fig. S1.1 G) by
342 condensing the IMPC's 26 procedural categories ("procedures") into related clusters. The
343 categories were based on procedures that were biologically related, in conjunction with
344 measurement techniques and number available traits in each category (see SI Appendix Table
345 S6.3 for a list of clustered traits, procedures and grouping terms). To test our two hypotheses
346 about how trait variability changes in relation to sex, we estimated overall effect sizes for nine
347 functional groups by aggregating meta-analytic results via a 'classical' random-effect models
348 using the function rma.uni in the R package *metafor* (51). In other words, we conducted three

349 sets of 10 second-order meta-analyses (i.e. meta-analyzing 3 types of effect size: lnRR, lnVR and
350 lnCVR for 9 functional groups and one for all the groups combined, SI Appendix Fig. S1.1 H).
351 Although we present the frequencies of male- and female-biased traits in Figure 3A, we did not
352 run inferential statistical tests on these counts because such tests would be considered as vote-
353 counting, which has been severely criticised in the meta-analytic literature (56).

354 Author contributions

355 SN conceived the initial idea, and all contributed to furthering the idea and the design of the
356 study. SRKZ, along with FZ, led the analyses and writing with inputs from all authors. DWAN
357 created the Shiny application. Apart from SRKZ, FZ, DWAN and SN, all authors have contributed
358 equally, yet uniquely, and are listed in alphabetical order.

359 SRKZ, ML and SN were all supported by the Australian (ARC) Discovery Grant (DP180100818).
360 JM was supported by EMBL core funding and the NIH Common Fund (UM1-H G006370). AMS
361 was supported by an ARC fellowship (DE180101520).

362 REFERENCES

363 1. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (John Murray, 1871).

364 2. J. A. Clayton, Studying both sexes: a guiding principle for biomedicine. *FASEB J.* **30**, 519–524
365 (2015).

366 3. K. L. Flanagan, Sexual dimorphism in biomedical research: a call to analyse by sex. *Trans R*
367 *Soc Trop Med Hyg* **108**, 385–7 (2014).

368 4. N. A. Karp, *et al.*, Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat.*
369 *Commun.* **8**, 15475 (2017).

370 5. S. L. Klein, *et al.*, Opinion: Sex inclusion in basic research drives discovery. *Proc. Natl. Acad.*
371 *Sci.* **112**, 5257–5258 (2015).

372 6. B. J. Prendergast, K. G. Onishi, I. Zucker, Female mice liberated for inclusion in neuroscience
373 and biomedical research. *Neurosci Biobehav Rev* **40**, 1–5 (2014).

374 7. R. M. Shansky, C. S. Woolley, Considering sex as a biological variable will be valuable for
375 neuroscience research. *J. Neurosci.* **36**, 11817–11822 (2016).

376 8. A. Nowogrodzki, Inequality in medicine. *Nature* **550**, S18–S19 (2017).

377 9. I. Zucker, B. J. Prendergast, Sex differences in pharmacokinetics predict adverse drug
378 reactions in women. *Biol. Sex Differ.* **11**, 32 (2020).

379 10. D. M. Dorris, J. Y. Cao, J. A. Willett, C. A. Hauser, J. Meitzen, Intrinsic excitability varies by
380 sex in prepubertal striatal medium spiny neurons. *J. Neurophysiol.* **113**, 720–729 (2015).

11. C. Ingvorsen, N. A. Karp, C. J. Lelliott, The role of sex and body weight on the metabolic effects of high-fat diet in C57BL/6N mice. *Nutr. Amp Diabetes* **7**, e261 (2017).

12. C. M. Robinson, Y. Wang, J. K. Pfeiffer, Sex-dependent intestinal replication of an enteric virus. *J. Virol.* **91** (2017).

13. B. L. Smarr, A. D. Grant, I. Zucker, B. J. Prendergast, L. J. Kriegsfeld, Sex differences in variability across timescales in BALB/c mice. *Biol. Sex Differ.* **8** (2017).

14. A. A. Ahmad, M. D. Randall, R. E. Roberts, Sex differences in the role of phospholipase A2-dependent arachidonic acid pathway in the perivascular adipose tissue function in pigs. *J. Physiol.* **595**, 6623–6634 (2017).

15. R. W. Foltin, S. M. Evans, Sex differences in the anorexigenic effects of dexfenfluramine and amphetamine in baboons. *Exp. Clin. Psychopharmacol.* **26**, 335–340 (2018).

16. L. P. Thompson, L. Chen, B. M. Polster, G. Pinkas, H. Song, Prenatal hypoxia impairs cardiac mitochondrial and ventricular function in guinea pig offspring in a sex-related manner. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **315**, R1232–R1241 (2018).

17. J. A. Clayton, F. S. Collins, Policy: NIH to balance sex in cell and animal studies. *Nature* **509**, 282–3 (2014).

18. NIH, Consideration of sex as a biological variable in NIH-funded research. **Notice NOT-OD-15-102** (2015).

19. A. K. Beery, I. Zucker, Sex bias in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* **35**, 565–572 (2011).

20. K. Reinhold, L. Engqvist, The variability is in the sex chromosomes. *Evolution* **67**, 3662–8 (2013).

21. A. Pomiankowski, A. P. Moller, A resolution of the lek paradox. *Proc. R. Soc. LondonSeries B-Biol. Sci.* **260**, 21–29 (1995).

22. J. J. Cuervo, A. P. Møller, Phenotypic variation and fluctuating asymmetry in sexually dimorphic feather ornaments in relation to sex and mating system. *Biol. J. Linn. Soc.* **68**, 505–529 (1999).

23. J. J. Cuervo, A. P. Møller, Components of phenotypic variation in avian ornamental and non-ornamental feathers. *Evol. Ecol.* **15**, 53–72 (2001).

24. L. Rowe, D. Houle, The lek paradox and the capture of genetic variance by condition dependent traits. *Proc. R. Soc. LondonSeries B-Biol. Sci.* **263**, 1415–1421 (1996).

25. J. L. Tomkins, J. Radwan, J. S. Kotiaho, T. Tregenza, Genic capture and resolving the lek paradox. *Trends Ecol. Evol.* **19**, 323–328 (2004).

414 26. A.-C. Lehre, K. P. Lehre, P. Laake, N. C. Danbolt, Greater intrasex phenotype variability in
415 males than in females is a fundamental aspect of the gender differences in humans. *Dev.*
416 *Psychobiol.* **51**, 198–206 (2009).

417 27. W. Johnson, A. Carothers, I. J. Deary, Sex differences in variability in general intelligence: a
418 new look at the old question. *Perspect. Psychol. Sci.* **3**, 518–531 (2008).

419 28. L. V. Hedges, A. Nowell, Sex differences in mental test scores, variability, and numbers of
420 high-scoring individuals. *Science* **269**, 41–45 (1995).

421 29. Y. Itoh, A. P. Arnold, Are females more variable than males in gene expression? Meta-
422 analysis of microarray datasets. *Biol. Sex Differ.* **6** (2015).

423 30. J. B. Becker, B. J. Prendergast, J. W. Liang, Female rats are not more variable than male rats:
424 a meta-analysis of neuroscience studies. *Biol. Sex Differ.* **7**, 34 (2016).

425 31. A. K. Beery, Inclusion of females does not increase variability in rodent research studies.
426 *Curr. Opin. Behav. Sci.* **23**, 143–149 (2018).

427 32. S. Nakagawa, *et al.*, Meta-analysis of variation: ecological and evolutionary applications and
428 beyond. *Methods Ecol. Evol.* **6**, 143–152 (2015).

429 33. A. M. Senior, W. Viechtbauer, S. Nakagawa, Revisiting and expanding the meta-analysis of
430 variation: The log coefficient of variation ratio. *Res Synth Methods*, e176 (2020).

431 34. M. E. Dickinson, *et al.*, High-throughput discovery of novel developmental phenotypes.
432 *Nature* **537**, 508 (2016).

433 35. M. Zuk, K. A. McKean, Sex differences in parasite infections: Patterns and processes. *Int. J.*
434 *Parasitol.* **26**, 1009–1024 (1996).

435 36. C. Grossman, Possible underlying mechanisms of sexual dimorphism in the immune
436 response, fact and hypothesis. *J. Steroid Biochem.* **34**, 241–251 (1989).

437 37. H. Wagner, B. A. Fink, K. Zadnik, Sex- and gender-based differences in healthy and diseased
438 eyes. *Optom. - J. Am. Optom. Assoc.* **79**, 636–652 (2008).

439 38. A. Shaqiri, *et al.*, Sex-related differences in vision are heterogeneous. *Sci. Rep.* **8** (2018).

440 39. A. Fritz, I. Amrein, D. P. Wolfer, Similar reliability and equivalent performance of female and
441 male mice in the open field and water-maze place navigation task. *Am. J. Med. Genet. C*
442 *Semin. Med. Genet.* **175**, 380–391 (2017).

443 40. J. S. Mogil, M. L. Chanda, The case for the inclusion of female subjects in basic science
444 studies of pain. *Pain* **117**, 1–5 (2005).

445 41. S. Nakagawa, D. W. A. Noble, A. M. Senior, M. Lagisz, Meta-evaluation of meta-analysis: ten
446 appraisal questions for biologists. *BMC Biol.* **15**, 18 (2017).

42. T. F. Hansen, D. Houle, Measuring and comparing evolvability and constraint in multivariate characters. *J. Evol. Biol.* **21**, 1201–1219 (2008).

43. J.-F. Lemaître, *et al.*, Sex differences in adult lifespan and aging rates of mortality across wild mammals. *Proc. Natl. Acad. Sci.* **117**, 8546–8553 (2020).

44. H. Caswell, D. E. Weeks, Two-Sex Models: Chaos, Extinction, and Other Dynamic Consequences of Sex. *Am. Nat.* **128**, 707–735 (1986).

45. J. Lindström, H. Kokko, Sexual reproduction and population dynamics: the role of polygyny and demographic sex differences. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 483–488 (1998).

46. F. Colchero, A. E. Aliaga, O. R. Jones, D. A. Conde, Individual heterogeneity determines sex differences in mortality in a monogamous bird with reversed sexual dimorphism. *J. Anim. Ecol.* **86**, 899–907 (2017).

47. NIH, Enhancing reproducibility through rigor and transparency. **Notice NOT-OD-15-103** (2015).

48. B. Voelkl, *et al.*, Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* **21**, 384–393 (2020).

49. I. R. Cleasby, S. Nakagawa, Neglected biological patterns in the residuals. *Behav. Ecol. Sociobiol.* **65**, 2361–2372 (2011).

50. C. Tannenbaum, R. P. Ellis, F. Eyssel, J. Zou, L. Schiebinger, Sex and gender analysis improves science and engineering. *Nature* **575**, 137–146 (2019).

51. W. Viechtbauer, Conducting meta-analyses in R with the metafor package. *J Stat Softw* **36** (2010).

52. R Core Team, *R: A language and environment for statistical computing.* (2017).

53. S. Nakagawa, *et al.*, Research weaving: visualizing the future of research synthesis. *Trends Ecol. Evol.* **34**, 224–238 (2019).

54. S. Nakagawa, E. S. Santos, Methodological issues and advances in biological meta-analysis. *Evol. Ecol.* **26**, 1253–1274 (2012).

55. Z. Fisher, E. Tipton, H. Zhipeng, M. Z. Fisher, Package ‘robumeta’ (2017).

56. J. P. Higgins, *et al.*, *Cochrane handbook for systematic reviews of interventions* (John Wiley & Sons, 2019).

Fig. 1.

Overview of meta-analytic methods used to detect differences in means and variances in any given trait (e.g., body size in mice). The orange shading represents females (F), turquoise shading stands for males (M). The solid "dot" represents a mean trait value within the respective group. Solid lines represent standard deviation, with upper and lower bounds indicated by diamond shapes. Below, we present three types of effect sizes that can be used for comparing two groups, along with the respective formulas and interpretations. Compared to lnVR (the ratio of SD), lnCVR (the ratio of CV or relative variance) provides a more general measure of the difference in variability between two groups (mean-adjusted variability ratio).

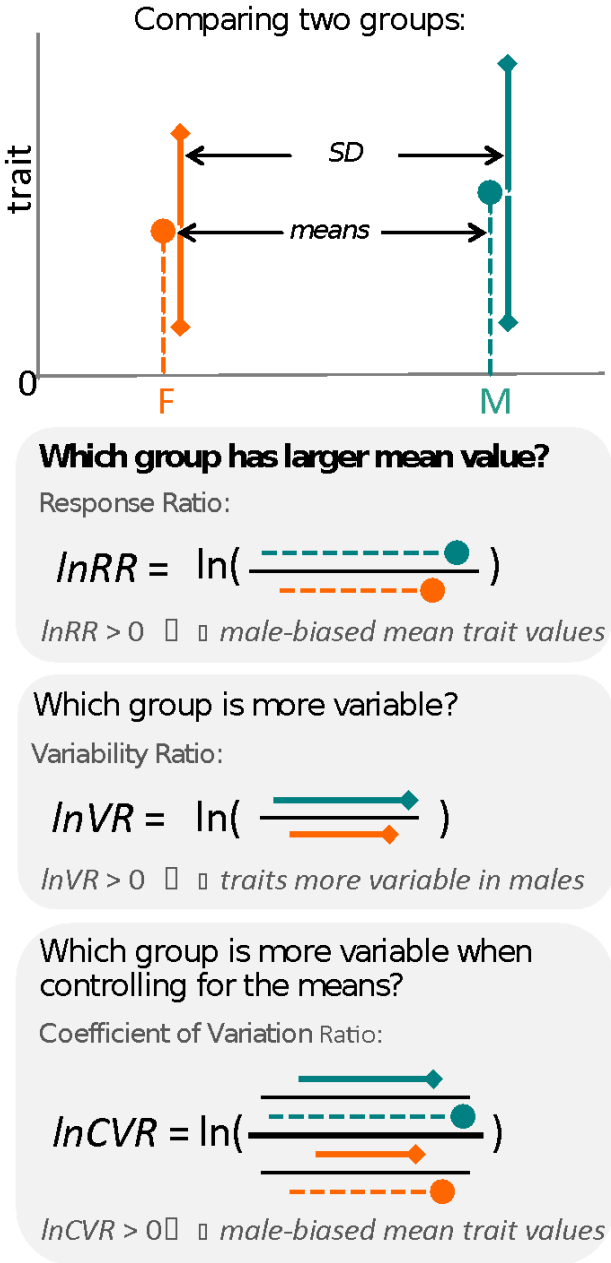


Fig. 2.

The two hypotheses ("Greater Male Variability" vs "Estrus-Mediated variability") have different underlying predictions on how variabilities influence total observed phenotypic variance (V_{total} in the figure). For Greater Male Variability, the within-subject (or within-trait) variation V_{within} could be potentially negligible, or is equal in males and females. This is illustrated as the shaded distributions around each individual mean (dashed vertical lines), which are of equal area for the males (turquoise) and females (orange). The greater value of V_{total} is driven by wider distribution of mean trait values in males compared to females (i.e. $V_{between}$, represented by a thick horizontal bar). The estrus-mediated variability hypothesis, in contrast, assumes that within-subject [or within trait] variability is much higher in females than in males (broader orange-shaded trait distributions than blue-green distributions), while the variability of the means between individuals stays the same (thick horizontal bars).

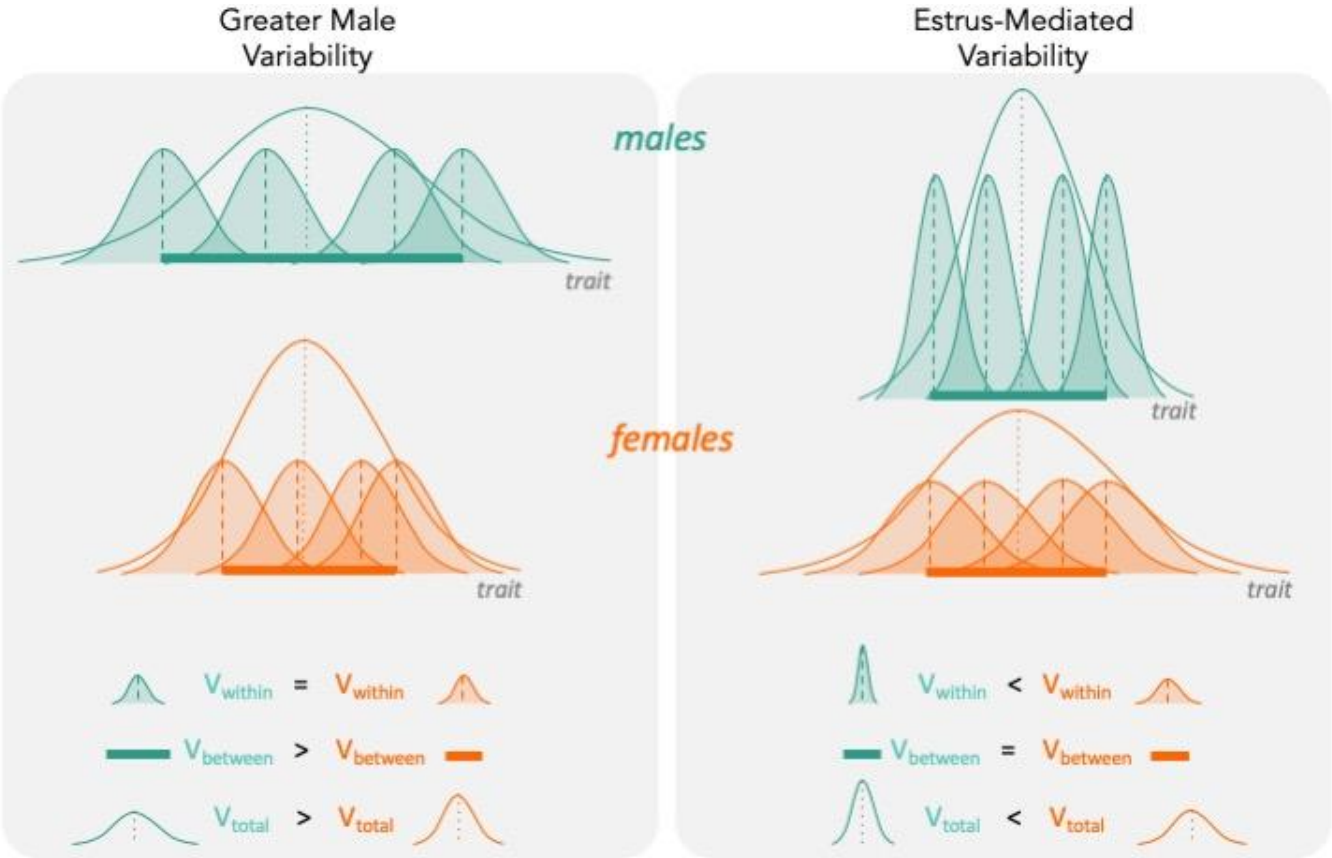
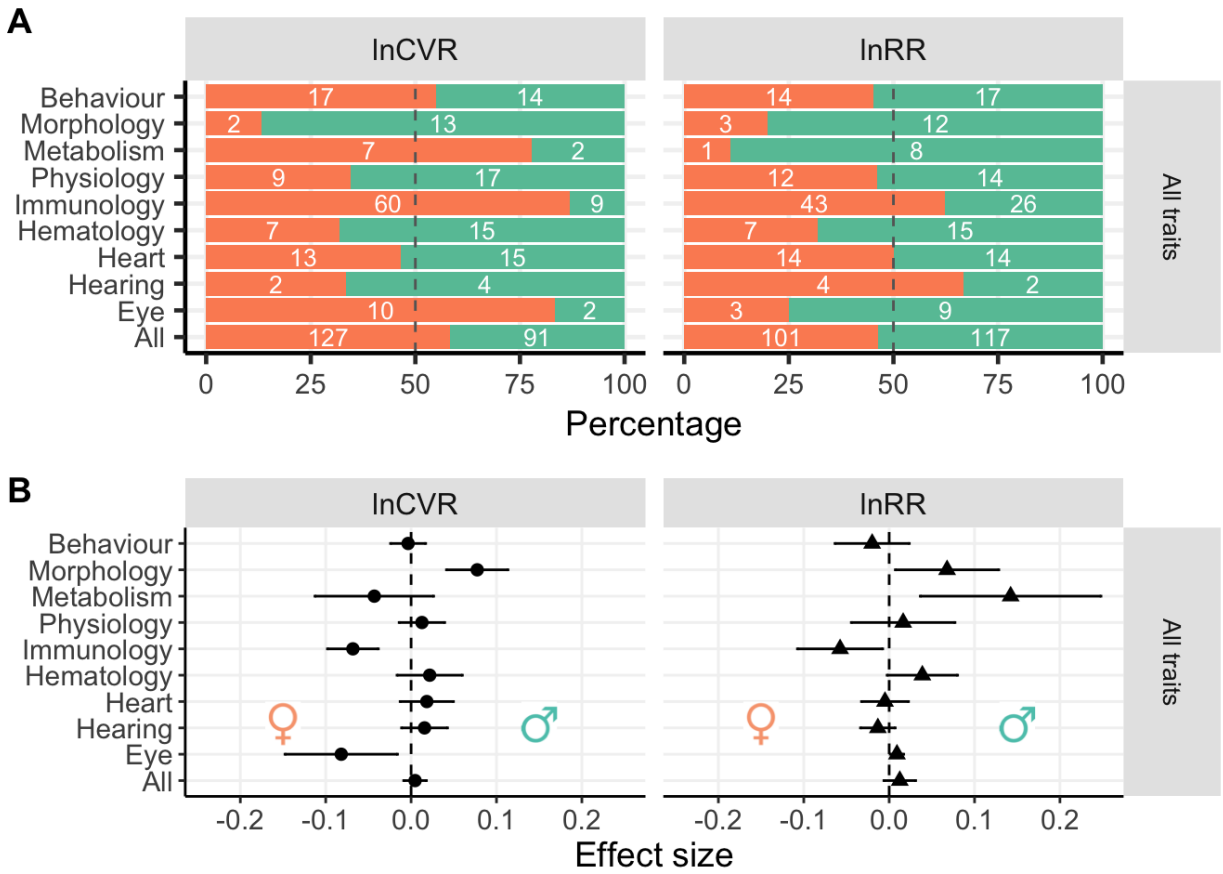


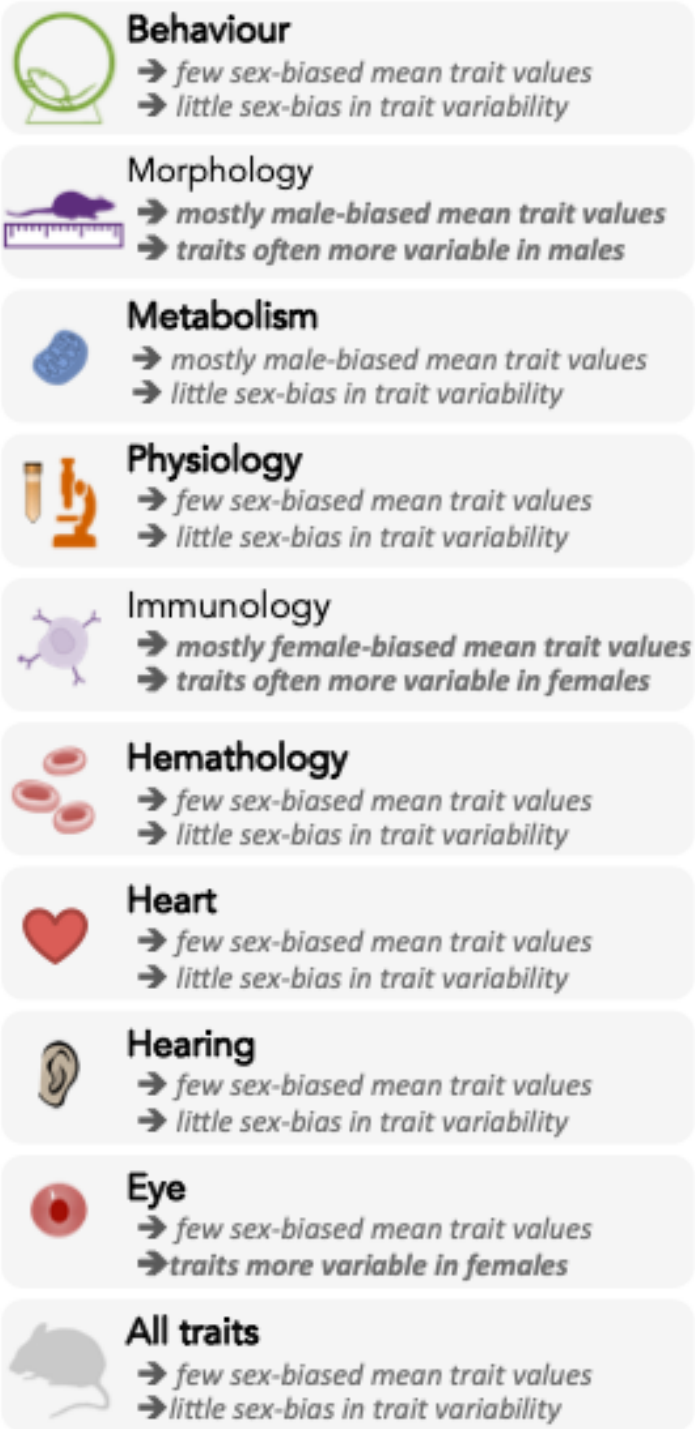
Fig. 3:

Panel A shows the numbers of traits across functional groups that are either male-biased (turquoise) or female-biased (orange; as in SI Appendix Fig. S1.1 D). The x-axes in Panel A show the overall percentages of traits, coloured shading is indicative of direction of sex-bias sex (if meta-analytic means < 0, then they are female-based whereas if they are > 0, male-based). White numbers in the turquoise bars represent numbers of traits that show male-bias within a given group of traits, number in the orange area the number of female-biased traits. Panel B shows effect sizes and 95% CI from separate meta-analysis for each functional group (SI Appendix Fig. S1.1 H). Traits that are male biased in Panel B are shifted towards the righthand side of the zero-midline (near the turquoise male symbol), whereas female bias is shifted towards the left (near orange symbol).



515 **Fig. 4.**

516 Summary of sex-differences in the mean trait values (lnRR) and variances (lnCVR) across nine
517 functional trait groups, and overall.



518