The Z chromosome is enriched for sperm proteins in two divergent species of Lepidoptera

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0068.R1
Manuscript Type:	Article
Date Submitted by the Author:	01-Aug-2017
Complete List of Authors:	Mongue, Andrew; University of Kansas, Ecology and Evolutionary Biology Walters, James; University of Kansas, Ecology and Evolutionary Biology
Is the invited manuscript for consideration in a Special Issue? :	Ecological Genomics
Keyword:	sex chromosomes, sexual antagonism, Lepidoptera, sperm, female heterogamety

SCHOLARONE™
Manuscripts

**The Z chromosome is enriched for sperm proteins in
two divergent species of Lepidoptera**

Andrew J. Mongue and James R Walters*

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA

*Author for correspondence: James R Walters, Department of Ecology and Evolutionary Biology,
University of Kansas, Lawrence, KS, USA

phone: 785-864-4361

email: jrwalters@ku.edu

Draft

Abstract

Genes that promote sexual conflict, such as those with a sex-limited fitness benefit, are expected to accumulate differentially on sex chromosomes relative to autosomes. Few tests of this hypothesis exist for male homogametic (ZZ) taxa, however, and most use RNA expression data to identify such genes. Here, we employ a different identification method by using proteomic analysis of sperm cells to identify genes with a sex-limited benefit. We tested for a bias in genomic location of sperm protein genes in two species of Lepidoptera. An excess of sperm protein genes was identified on the Z chromosomes of both the Carolina sphinx moth (*Manduca sexta*) and the monarch butterfly (*Danaus plexippus*). Taking into consideration a Z-autosome fusion in monarchs, we discover that the ancestrally sex-linked portion of the genome is the source of this enrichment while the newly sex-linked portion still appears similar to autosomes in relative abundance of sperm protein genes. Together, these results point to an enrichment of male-beneficial genes on the Z chromosome and demonstrate the usefulness of proteomic datasets in sexual conflict research.

Keywords

sex chromosomes, sexual antagonism, sexual conflict, sperm, female heterogamety, Lepidoptera

Introduction

All living organisms exist as partnerships of genes that can only function in conjunction with each other, but there are many times when a gene's function does not benefit the organism in which it resides (Michod and Roze 2001). In some cases, the interests of the genetic element and the organism as a whole are always at odds, as is the case with transposable elements, sequences that exist solely to copy themselves at the expense of their host genome (Levin and Moran 2011). Another prominent example of such genetic conflict is found in the set of genes that are responsible for sexual dimorphism in gonochorous (*i.e.* unisexual) organisms. Genetic variants promoting maleness are maladaptive in females and vice versa, but both must be maintained in the population to ensure the continued function of both sexes (Cox and Calsbeek 2009; Van Doorn 2009).

One solution to this conflict can be found in species with chromosomal sex determination (reviewed in Chapman *et al.* 2003). Genes on autosomes have equal representation in both sexes, but those located on sex chromosomes either have representation in only one sex (such as those found on Y or W chromosomes) or a biased representation being found in one sex two thirds of the time (in males in the case of the Z and females in the case of the X). Thus sex chromosomes present an opportunity for sexually antagonistic alleles to minimize negative selective pressure that would come from being found in the opposite sex (Rice 1984). Sex chromosomes therefore make appealing targets for studies of sexually antagonistic gene evolution. In particular, theory predicts genetic elements with sex-specific fitness effects (*i.e.* sexually antagonistic loci) should accumulate differentially on the sex chromosomes relative to autosomes.

To identify putatively antagonistic loci, many studies use sex-biased gene expression to identify genes with sexually antagonistic effects (Connallon and Clark 2010; Parsch and Ellegren 2013). In such studies, the assumption is that sex-biased expression typically reflects sexually antagonistic evolution. However,

this is not necessarily the case, particularly for sex chromosomes, where sex-biased expression may reflect uncompensated differences in chromosomal copy number between the sexes, *i.e.* incomplete dosage compensation (Mank 2013). Furthermore, it is often the case that the majority of sex-biased gene expression arises in gonadal tissue (Gu and Walters in press; Mank *et al.* 2008; Allen *et al.* 2013). However, in gonads the sex chromosomes often experience special modulations of gene regulation that may result in distinct patterns of sex-biased expression unrelated to sexual antagonism (Turner 2015; Larson *et al.* 2016). Therefore, transcriptome-based studies of sexual antagonism are informative, but are often also complicated by dosage effects and idiosyncrasies of meiotic gene regulation.

Despite these potential caveats, research concerning sex-biased expression supports the predictions of biased genomic distribution in male-heterogametic (XX/XY) systems: dominant female-benefiting genes should be overrepresented on the X chromosome (Rice 1984). This pattern has been observed in both vertebrates (e.g. mice: Khil *et al.* 2004) and invertebrates (e.g. fruit flies: Parisi *et al.* 2003), strengthening the conclusion that these dynamics are the result of consistent evolutionary pressures rather than contingency in the origin of a sex chromosome system, *i.e.* that the precursor to the sex chromosome happened to be enriched for sex-biased genes by mere chance.

Reciprocally, research in female-heterogametic (ZZ/ZW) taxa suggests that the Z is enriched for male-biased genes, indicating that male-benefiting genes have preferentially accumulated there (Storchová and Divina 2006; Ellegren 2009). This research was initially carried out mainly in birds, which represent just a single WZ/ZZ evolutionary event (Griffiths *et al.* 1998), and is complicated by incomplete dosage compensation (Ellegren *et al.* 2007; Itoh *et al.* 2010; Uebbing *et al.* 2013). Gene expression analysis in an independent set of WZ/ZZ taxa, Lepidoptera (butterflies and moths), offers further support for a sex-chromosome-bias in sexually antagonistic genes (Arunkumar *et al.* 2009; Walters and Hardcastle 2011; Walters *et al.* 2015). Once again however, how dosage effects impact this system is not yet settled (Gu

and Walters in press; Harrison *et al.* 2012; Smith *et al.* 2014). Nonetheless, at least in birds, research using approaches that circumnavigate the issue of dosage compensation does indicate an abundance of male-benefitting and/or female-detriment genes on the chicken Z (Mank and Ellegren 2009; Ellegren 2011).

Here we test this prediction of a sex-linked bias in the location of sexually antagonistic loci in Lepidoptera using a set of genes identified via proteomic analysis that does not depend on gene expression levels and avoids potential complications due to sex chromosome dosage effects. Recently, efforts to understand the reproductive biology of Lepidoptera have produced two sperm proteomic datasets, one in *Manduca sexta*, the Carolina sphinx moth (Whittington *et al.* 2015), and the other in *Danaus plexippus*, the monarch butterfly (Whittington *et al.*, forthcoming). We assert that these sets of proteins present in sperm, a male-limited cell type, can serve as a simple heuristic for identifying genes with potentially male-limited benefits. Sperm are produced in males and function only to pass on the male genome to offspring, thus their selective background is decidedly male. With this rationale, we examined the set of proteins expressed in lepidopteran sperm cells for evidence of a Z-bias in genomic location of male benefiting genes, as predicted by sexual antagonism theory.

The monarch and Carolina sphinx moth share the WZ/ZZ sex determination system (Traut *et al.* 2007) but last shared a common ancestor over 100 million years ago (Heikkila *et al.* 2012). Notably, in the time since that divergence, the lineage leading to monarchs has undergone a major chromosomal fusion between the Z and an autosome, yielding a bipartite Z chromosome with an "ancestral-Z" segment homologous to the Z in other Lepidoptera and a "neo-Z" segment homologous to an autosome in other species (Mongue *et al.* 2017). This large movement event of genes onto the monarch Z presents a unique opportunity to compare accumulation of sex-biased genes on a sex chromosome with a more complicated history.

99 Methods

100 Analyzing sperm protein genes for Z-linked bias in genomic location required categorizing all predicted
101 genes as being 1) part of the sperm proteome or not and 2) Z-linked or autosomal. Putatively male-
102 benefitting genes were established from the published sperm proteomes. In these studies, proteins
103 were identified based on mass spectrometry analyses of sperm in the Carolina sphinx moth and
104 monarch butterfly (Whittington *et al.* 2015, 2017). Both of these species have published draft genome
105 assemblies providing a comprehensive set of protein-coding loci to assess for chromosomal linkage
106 (Zhan *et al.* 2011; Kanost *et al.* 2016). Additionally, in both species, chromosomal linkage is already
107 predicted for most genes. For the Carolina sphinx moth genome assembly, chromosomal linkage of
108 scaffolds was predicted by ortholog comparison to *Bombyx mori*, another lepidopteran species with
109 genome scaffolds genetically mapped to linkage groups (Kanost *et al.* 2016). Chromosomal linkage of
110 genes (*i.e.* macrosynteny) is extremely well conserved among Lepidoptera, so this ortholog-based
111 approach is generally robust (Ahola *et al.* 2014; Davey *et al.* 2016; Kanost *et al.* 2016). Furthermore, in
112 the case of Carolina sphinx moth, fluorescence in situ hybridization (FISH) mapping relative to *Bombyx*
113 has revealed no large-scale transitions in sex chromosome identity (Yasukochi *et al.* 2009). Similar
114 orthology-based chromosomal assignment analyses were carried out for the monarch butterfly as part
115 of an investigation of a recent Z-autosome fusion in the genus *Danaus*; this same study further assessed
116 scaffold Z-linkage via an independent method comparing differences in depth of Illumina sequencing
117 coverage between male and female samples, thus distinguishing between the ancestral and "neo"
118 portions of the monarch Z chromosome (Mongue *et al.* 2017).

119 To assess a chromosomal bias in the location of sperm proteins, all genes were scored on two simple
120 metrics: gene class and gene location. In both species, genes were labeled as either part of the sperm
121 proteome dataset or not. Gene location categories varied slightly between species. In the monarch,

genes were assigned to one of three categories: autosomal, ancestrally Z-linked, or neo-Z linked, to reflect the evolutionary history of the Z chromosome in this genus. In the Carolina sphinx moth, genes from the sperm proteome were assigned to one of two classes: autosomal or Z-linked, to be consistent in comparison to the monarch. Raw gene counts in each category are presented in Table 1.

If the sperm protein genes do not contribute to sexual conflict, their relative abundance should be the same on the Z as on the autosomes, *i.e.* they should be distributed in proportion to their overall frequency in the genome. The proteomic datasets have not been curated to remove genes that play roles in the nucleus, for instance, and thus are almost certainly not involved in sexual conflict. Thus, our heuristic overestimates the number of genes involved in sexual conflict, making this assay conservative and decreasing the chances of a false positive result. With gene locations and classes annotated, deviation of the Z chromosome's composition from autosomes was analyzed with a χ^2 test for independence of sperm protein distribution.

Finally, to test for the effects of shared ancestry on the composition of the Z chromosome's genes, sperm genes were split into two groups: those unique to one species' proteome and those shared between species (*i.e.* orthologs or close paralogs identified in the sperm proteomes of both species). This partitioning was motivated by the concern that including genes with shared ancestry and function create non-independence in the two datasets and could bias our results. Orthology predictions were taken from Whittington *et al.* (2017) and reflect the default settings from the *proteinortho* software (Lechner *et al.* 2011). In this case we considered sperm proteins to be shared between species if they were reported as one-to-one orthologs, or were included in distinct one-to-many (e.g. co-orthologs) and many-to-many (e.g. unresolved paralogy) relationships. Overwhelmingly these cases of shared sperm proteins reflected one-to-one homology, but also included a handful of sperm genes reported in paralogous groupings (n=24 in monarchs, and n=17 in the Carolina sphinx). Differences in distribution of

shared compared to unique sperm genes were assessed with χ^2 tests. All analyses were performed with statistical program R (R Development Core Team 2015).

Results

In the Carolina sphinx moth, sperm proteins are twice as abundant on the Z as on the autosomes, making up 9.6% of genes on the Z but only 4.9% of autosomal genes ($\chi_1 = 25.1$, $p < 0.00001$). Indeed, the Z has the highest proportion of sperm proteins of any of the twenty-eight chromosomes in the genome (Figure 1A). Finally, 42% (288/685) of the sperm proteins in the sphinx moth proteome were shared with the monarch sperm proteome (Figure 2, Table 1, Whittington *et al.* 2017). There was no difference in Z- vs.-autosome distribution between these shared loci and genes unique to the Carolina sphinx moth sperm proteome ($\chi_1 = 0.03$, $p = 0.870$).

The monarch sperm proteome shows a similar bias in sperm gene distribution: with 4.5% of genes on the Z as a whole found in the sperm proteome and significantly fewer, 3.7%, of autosomal genes contributing to sperm ($\chi_1 = 6.8$, $p = 0.009$). As noted above however, the Z in *Danaus* butterflies has a more complicated evolutionary history, containing both ancestrally and newly sex-linked segments. Because of this, we also carried out analyses splitting the Z into these two segments (Mongue *et al.* 2017). This analysis also suggested a significant bias in sperm protein distribution ($\chi_2 = 28.3$, $p < 0.00001$). Post hoc pairwise analysis revealed that the main signal for the difference is attributable to the ancestral Z. 7.7% of the ancestral Z's gene content appears in the sperm proteome, compared to 3.7% of autosomal genes and only 2.3% of the neo-Z; both of which were significant results even after Bonferroni correction (ancestral vs. autosomal: $\chi_1 = 23.8$, $p < 0.00001$; ancestral vs. neo: $\chi_1 = 14.4$, $p = 0.0001$). While the neo-Z displays lower proportional representation in the sperm proteome than the autosomes, this difference was not significant ($\chi_1 = 2.3$, $p = 0.13$; Figure 1B).

Comparing between species, 51% (299/584) of monarch sperm genes are shared with the Carolina sphinx moth (Figure 2, Table 1). Looking first at the unpartitioned Z chromosome, there was no difference in the distribution of shared sperm genes compared to those unique to monarchs ($\chi_1 = 0.007$, $p = 0.94$). The more complex view of the monarch Z, accounting for the ancestral and neo-Z, also revealed no difference in the distribution between shared versus unique genes ($\chi_2 = 1.46$, $p = 0.48$).

Discussion

Our results indicate that the Z chromosome in two distantly related species of Lepidoptera is enriched for sperm protein genes. These results are consistent with observations from gene-expression studies in female-heterogametic taxa (Storchová and Divina 2006; Arunkumar *et al.* 2009; Walters *et al.* 2015; Gu *et al.* 2017) and established theory that the homogametic sex chromosome should accumulate dominant mutations that benefit males (Rice 1984). We have made the simplifying assumption that such mutant alleles can be identified by their expression in a male-limited cell type. In reality however, some proteins found in sperm cells may have sex-neutral benefits, such as nuclear and structural proteins; their corresponding genes are not predicted to be preferentially located on sex chromosomes. Thus, this assumption would decrease the power to detect a bias in sperm-specific gene location. Nevertheless, we see significant enrichment of the Z chromosome in the sperm proteomes of both the Carolina sphinx moth and the monarch butterfly (Figure 1A&B); genes expressed in sperm cells are more likely to be found on the Z chromosome than on autosomes. These results from lepidopteran sperm reflect the inverse of expectations from the sperm proteomes of some male-heterogametic species. For instance, sperm proteins are under-represented on the X chromosome in Rhesus macaque (Skerget *et al.* 2013), but this pattern is less consistent in other XX/XY taxa.

A similar result was also initially reported for *Drosophila melanogaster* (Dorus *et al.* 2006), though an expanded assay in this species did not replicate this result (Wasbrough *et al.* 2010), nor did an RNA-

based investigation of these flies (Meiklejohn and Presgraves 2012). However, intersecting these sperm proteome results with a nuanced dissection of gene expression indicates a possible dearth of X-linked sperm genes with testis-specific expression, depending on the gene-set analyzed (Meisel *et al.* 2012). More recently, another species of fly, the stalk-eyed fly, was observed to have a masculinized X chromosome (Baker *et al.* 2016). Thus, the pattern of sex-biased gene distribution is still far from clear; these sperm proteomic results from monarch and Carolina sphinx represent an important complement to similar analyses in XX/XY species by testing the same hypotheses in a set of independent sex chromosomes.

Comparison of shared and unique genes in the proteome

It is worth noting that the monarch and Carolina sphinx moth show substantial overlap in sperm gene content (Whittington *et al.* 2017), with about half of the sperm proteins in each species found to have orthologs or close paralogs in the sperm proteome of the other (Figure 2, Table 1). This overlap creates a degree of pseudo-replication in the data, since shared sperm genes likely share the same location, e.g. on the Z chromosome, and thus may not reflect two independent events to the Z but rather one movement or origin in the common ancestor of the two species studied here. Thus, the shared sperm proteins give us only one inference of Z-bias.

On the other hand, about half of sperm genes are unique to one or the other species. These genes were either retained (e.g. lost in one lineage) or gained sperm function in the 100+ million years of evolution following the divergence of these two lineages from their last common ancestor (Heikkila *et al.* 2012). Unique sperm genes from both monarchs and the Carolina sphinx moth show the same distribution as the orthologous portion of the sperm proteome. Doubtless the long period of independent evolution has been sufficient to fix or preserve more male-benefitting alleles on the Z than the autosomes in both

species, ensuring that the Z-bias persists despite substantial gene turnover in the sperm proteome between these two Lepidoptera, including a major chromosomal fusion event.

The neo-Z in monarchs

Although we observed an enrichment of the whole monarch Z for sperm genes, the relatively low content (4.5% in monarchs compared to the 9.6% in the Carolina sphinx moth) and unique evolutionary history of this chromosome led us to examine it in more detail. Contrasting the ancestrally Z-linked segment and the newly sex-linked neo-Z segment made an obvious point of comparison. Considered separately, the two segments are quite distinct in their representation in the sperm proteome; as reported above, over 7% of genes on the ancestral-Z are present in the sperm proteome. This greater degree of enrichment is more in line with the result from the Carolina sphinx moth Z, but it is obscured by the low sperm gene content of the neo-Z. Indeed, this neo-Z segment is homologous to chromosome 16 of the sphinx moth, which notably has the lowest sperm gene proportion among sphinx moth autosomes.

The time of fusion between this autosome and the Z chromosome is still uncertain. The neo-Z segment appears to be older than the genus *Danaus* but not as old as the subfamily, Nymphalinae (Mongue *et al.* 2017). Although these divergences bracket a wide swath of time, from 5 million (Lushai *et al.* 2003) to 50 million years (Heikkila *et al.* 2012), the lack of accumulation of sperm genes onto the neo-Z implies a more recent origin within this range.

Broadly considered, sperm proteomic data are becoming increasingly common in a wide range of vertebrate and invertebrate taxa (e.g. Johnston *et al.* 2005; Byrne *et al.* 2012; Ma *et al.* 2014; Soler *et al.* 2016), yet such data are seldom applied to assess predictions of sexual conflict, as we have done here. Importantly, this approach offers an alternative to RNA-based expression for identifying sexually

234 antagonistic loci. If sexual conflict shapes genomic architecture, the patterns we report here will surely
235 be borne out in other taxa.

236 Acknowledgements

237 We thank Tim Karr, Steve Dorus, Desiree Harpel, and Emma Whittington for their important
238 contributions in laying the foundation for the analysis presented here. AJM thanks Megan Hansen for
239 comments and feedback on this manuscript. JRW thanks members of the Ecological Genomics program
240 at Kansas State University for the invitation to present at the 2016 Ecological Genomics conference and
241 to subsequently participate in this special journal issue. This work was supported by research start-up
242 funds provided by the University of Kansas to JRW.

Draft

References

Ahola V., Lehtonen R., Somervuo P., Salmela L., Koskinen P., *et al.*, 2014 The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* 5: 1–9.

Allen S. L., Bonduriansky R., Chenoweth S. F., 2013 The genomic distribution of sex-biased genes in *Drosophila serrata*: X chromosome demasculinization, feminization, and hyperexpression in both sexes. *Genome Biol. Evol.* 5: 1986–1994.

Arunkumar K. P., Mita K., Nagaraju J., 2009 The silkworm Z chromosome is enriched in testis-specific genes. *Genetics* 182: 493–501.

Baker R. H., Narechania A., DeSalle R., Johns P. M., Reinhardt J. A., *et al.*, 2016 Spermatogenesis drives rapid gene creation and masculinization of the X chromosome in stalk-eyed flies (Diopsidae). *Genome Biol. Evol.* 8: 896–914.

Byrne K., Leahy T., McCulloch R., Colgrave M. L., Holland M. K., 2012 Comprehensive mapping of the bull sperm surface proteome. *Proteomics* 12: 3559–3579.

Chapman T., Arnqvist G., Bangham J., Rowe L., 2003 Sexual conflict. *Trends Ecol. Evol.* 18: 41–47.

Connallon T., Clark A. G., 2010 Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution (N. Y.)*. 64: 3417–3442.

Cox R. M., Calsbeek R., 2009 Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am. Nat.* 173: 176–187.

Davey J. W., Chouteau M., Barker S. L., Maroja L., Baxter S. W., *et al.*, 2016 Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3 Genes | Genomes | Genet.*: g3-115.

Dorus S., Busby S. a, Gerike U., Shabanowitz J., Hunt D. F., *et al.*, 2006 Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat. Genet.* 38: 1440–1445.

Ellegren H., 2009 Genomic evidence for a large-Z effect. *Proc. Biol. Sci. / R. Soc.* 276: 361–366.

Ellegren H., 2011 Emergence of male-biased genes on the chicken Z-chromosome: Sex-chromosome contrasts between male and female heterogametic systems. *Genome Res.* 21: 2082–2086.

Ellegren H., Hultin-Rosenberg L., Brunström B., Dencker L., Kultima K., *et al.*, 2007 Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* 5: 40–52.

Griffiths R., Double M. C., Orr K., Dawson R. J. G., 1998 A DNA test to sex most birds. *Mol. Ecol.* 7: 1071–1075.

Gu L., Walters J. R., Knipple D. C., 2017 Conserved patterns of sex chromosome dosage compensation in the Lepidoptera (WZ/ZZ): insights from a moth neo-Z chromosome. *Genome Biol. Evol.* 9: 802–816.

Harrison P. W., Mank J. E., Wedell N., 2012 Incomplete sex chromosome dosage compensation in the indian meal moth, *Plodia interpunctella*, based on de novo transcriptome assembly. *Genome Biol. Evol.* 4: 1118–1126.

- 279 Heikkilä M., Kaila L., Mutanen M., Pena C., Wahlberg N., 2012 Cretaceous origin and repeated tertiary
280 diversification of the redefined butterflies. *Proc. R. Soc. B Biol. Sci.* 279: 1093–1099.
- 281 Itoh Y., Replogle K., Kim Y. H., Wade J., Clayton D. F., *et al.*, 2010 Sex bias and dosage compensation in
282 the zebra finch versus chicken genomes: General and specialized patterns among birds. *Genome*
283 *Res.* 20: 512–518.
- 284 Johnston D. S., Wooters J., Kopf G. S., Qiu Y., Roberts K. P., 2005 Analysis of the human sperm proteome.
285 In: *Annals of the New York Academy of Sciences*, pp. 190–202.
- 286 Kanost M. R., Arrese E. L., Cao X., Chen Y.-R., Chellapilla S., *et al.*, 2016 Multifaceted biological insights
287 from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem.*
288 *Mol. Biol.* 76: 118–147.
- 289 Khil P. P., Smirnova N. A., Romanienko P. J., Camerini-Otero R. D., 2004 The mouse X chromosome is
290 enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation.
291 *Nat. Genet.* 36: 642–646.
- 292 Larson B. J., Van M. V., Nakayama T., Engebrecht J., 2016 Plasticity in the meiotic epigenetic landscape of
293 sex chromosomes in *Caenorhabditis* species. *Genetics* 203: 1641–1658.
- 294 Lechner M., Findeiß S., Steiner L., Marz M., Stadler P. F., *et al.*, 2011 Proteinortho: Detection of (Co-
295)orthologs in large-scale analysis. *BMC Bioinformatics* 12: 124.
- 296 Levin H. L., Moran J. V., 2011 Dynamic interactions between transposable elements and their hosts. *Nat.*
297 *Rev. Genet.* 12: 615–627.
- 298 Lushai G., Smith D. A. S., Goulson D., Allen J. A., Maclean N., 2003 Mitochondrial DNA clocks and the
299 phylogeny of *Danaus* butterflies. *Int. J. Trop. Insect Sci.* 23: 309–315.
- 300 Ma X., Zhu Y., Li C., Xue P., Zhao Y., *et al.*, 2014 Characterisation of *Caenorhabditis elegans* sperm
301 transcriptome and proteome. *BMC Genomics* 15: 168.
- 302 Mank J. E., Hultin-Rosenberg L., Webster M. T., Ellegren H., 2008 The unique genomic properties of sex-
303 biased genes: insights from avian microarray data. *BMC Genomics* 9: 148.
- 304 Mank J. E., Ellegren H., 2009 Sex-linkage of sexually antagonistic genes is predicted by female, but not
305 male, effects in birds. *Evolution (N. Y.)* 63: 1464–1472.
- 306 Mank J. E., 2013 Sex chromosome dosage compensation: Definitely not for everyone. *Trends Genet.* 29:
307 677–683.
- 308 Meiklejohn C. D., Presgraves D. C., 2012 Little evidence for demasculinization of the *Drosophila* X
309 chromosome among genes expressed in the male germline. *Genome Biol. Evol.* 4: 1007–1016.
- 310 Meisel R. P., Malone J. H., Clark A. G., 2012 Disentangling the relationship between sex-biased gene
311 expression and X-linkage. *Genome Res.* 22: 1255–1265.
- 312 Michod R. E., Roze D., 2001 Cooperation and conflict in the evolution of multicellularity. *Heredity (Edinb.)*.
313 86: 1–7.
- 314 Mongue A. J., Nguyen P., Volenikova A., Walters J. R., 2017 Neo-sex chromosomes in the Monarch

- 315 butterfly, *Danaus plexippus*. bioRxiv.
- 316 Parisi M., Nuttall R., Naiman D., Bouffard G., Malley J., *et al.*, 2003 Paucity of genes on the *Drosophila* X
317 chromosome showing male-biased expression. *Science*. 299: 697–700.
- 318 Parsch J., Ellegren H., 2013 The evolutionary causes and consequences of sex-biased gene expression.
319 *Nat. Rev. Genet.* 14: 83–7.
- 320 R Development Core Team, 2015 R: A Language and Environment for Statistical Computing. R Found.
321 *Stat. Comput.* 1: 409.
- 322 Rice W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. *Evolution* (N. Y). 38: 735–
323 742.
- 324 Skerget S., Rosenow M., Polpitiya A., Petritis K., Dorus S., *et al.*, 2013 The Rhesus macaque (*Macaca*
325 *mulatta*) sperm proteome. *Mol. Cell. Proteomics* 12: 3052–3067.
- 326 Smith G., Chen Y. R., Blissard G. W., Briscoe A. D., 2014 Complete dosage compensation and sex-biased
327 gene expression in the moth *Manduca sexta*. *Genome Biol. Evol.* 6: 526–537.
- 328 Soler L., Labas V., Thlie A., Grasseau I., Teixeira-Gomes A.-P., *et al.*, 2016 Intact cell MALDI-TOF MS on
329 sperm: a molecular test for male fertility diagnosis. *Mol. Cell. Proteomics* M115.05291: in press.
- 330 Storchov R., Divina P., 2006 Nonrandom representation of sex-biased genes on chicken Z chromosome.
331 *J. Mol. Evol.* 63: 676–681.
- 332 Traut W., Sahara K., Marec F., 2007 Sex chromosomes and sex determination in Lepidoptera. *Sex. Dev.*
333 1: 332–346.
- 334 Turner J. M. A., 2015 Meiotic silencing in mammals. *Annu. Rev. Genet.* 49: 395–412.
- 335 Uebbing S., Knstner A., Mkinen H., Ellegren H., 2013 Transcriptome sequencing reveals the character
336 of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biol. Evol.* 5:
337 1555–1566.
- 338 Van Doorn, G. S., 2009 Intralocus sexual conflict. *Ann. N. Y. Acad. Sci.* 1168: 52–71.
- 339 Walters J. R., Hardcastle T. J., 2011 Getting a full dose? reconsidering sex chromosome dosage
340 compensation in the silkworm, *Bombyx mori*. *Genome Biol. Evol.* 3: 491–504.
- 341 Walters J. R., Hardcastle T. J., Jiggins C. D., 2015 Sex chromosome dosage compensation in *Heliconius*
342 butterflies: Global yet still incomplete? *Genome Biol. Evol.* 7: 2545–2559.
- 343 Wasbrough E. R., Dorus S., Hester S., Howard-Murkin J., Lilley K., *et al.*, 2010 The *Drosophila*
344 *melanogaster* sperm proteome-II (DmSP-II). *J. Proteomics* 73: 2171–2185.
- 345 Whittington E., Zhao Q., Borziak K., Walters J. R., Dorus S., 2015 Characterisation of the *Manduca sexta*
346 sperm proteome: Genetic novelty underlying sperm composition in Lepidoptera. *Insect Biochem.*
347 *Mol. Biol.* 62: 183–193.
- 348 Whittington E., Forsythe D., Karr T. L., Walters J., Dorus S., 2017 Contrasting Patterns Of Evolutionary
349 Constraint And Novelty Revealed By Comparative Sperm Proteomic Analysis. bioRxiv.

350 Yasukochi Y., Tanaka-Okuyama M., Shibata F., Yoshido A., Marec F., *et al.*, 2009 Extensive conserved
351 synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-
352 FISH mapping. PLoS One 4: e7465.

353 Zhan S., Merlin C., Boore J. L., Reppert S. M., 2011 The monarch butterfly genome yields insights into
354 long-distance migration. Cell 147: 1171–1185.

355

356

Draft

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

Figure 1. Chromosomal distribution of sperm genes in both the Carolina sphinx moth (A) and the monarch butterfly (B). Black denotes the Z chromosome in the sphinx moth and the ancestral Z in the monarch; the monarch whole Z is colored dark grey to reflect its composition: the ancestral (black) and neo (light grey) segments. Individual autosomes are assigned numbers and shown to the right of the Z segment(s). “All autos” represents the sum of all sperm genes found across all autosomes divided by the proportion of all autosomal genes; thus, it has no variance. Autosomal numbering references mapped chromosomes in *Bombyx mori* for Carolina Sphinx (Kanost et al. 2016) and *Melitaea cinxia* for the monarch. Note the lack of a 21st chromosome in the monarch butterfly, leading to a total chromosome count of n = 30 due to the fusion of this autosomal ortholog with the Z chromosome in monarchs (Mongue et al. 2017).

Figure 2. Proportion of orthologs in sperm proteomic datasets for both the sphinx moth (left) and monarch (right). Relative ortholog proportions do not differ significantly between the Z and the autosomes for either species.

Draft

373

374

	Carolina sphinx moth		Monarch butterfly		
	Autosomes	Z Chromosome	Autosomes	Neo-Z	Ancestral-Z
Shared sperm genes	290	27	268	4	27
Unique sperm genes	339	29	257	7	21
Non-sperm genes	12253	525	13504	466	576

Table 1. Counts of sperm genes by genomic location used in the statistical inferences presented here. In initial tests of distribution, genes were analyzed without considering a shared status between species; a subsequent test evaluated distribution of shared vs. unique sperm genes. Data were taken from Whittington *et al.* (2015) for the Carolina sphinx and Whittington *et al.* (2017) for the monarch butterfly.

Genome
Carolina sphinx moth (*Manduca sexta*)

Monarch butterfly (*Danaus plexippus*)

<https://mco0.manuscriptcentral.com/genome-pubs>

Carolina sphinx moth (*Manduca sexta*)

Monarch butterfly (*Danaus plexippus*)