Impact of trait exaggeration on sex-biased gene expression and genome architecture in a water strider

Authors: William Toubiana¹, David Armisen^{1,2}, Corentin Dechaud¹, Roberto Arbore^{1,3} and Abderrahman Khila^{1*}

Author affiliations:

¹ Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Claude Bernard Lyon1, CNRS UMR 5242, Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France

² Present address: INRA-Université Clermont Auvergne, Clermont-Ferrand, France.

³ Present address: Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal.

* Corresponding author: abderrahman.khila@ens-lyon.fr

Keywords: Sexual dimorphism, sexual selection, sex-biased genes, scaling relationships, genomic clusters

Abstract

Exaggerated secondary sexual traits are widespread in nature and often evolve under strong directional sexual selection. Although heavily studied from both theoretical and empirical viewpoints, we have little understanding of how sexual selection influences sex-biased gene regulation during the development of sex-specific phenotypes, and how these changes are reflected in genomic architecture. This is primarily due to the lack of a representative genome and transcriptomes to study the development of secondary sexual traits. Here we present the genome and developmental transcriptomes, focused on the legs of the water strider *Microvelia longipes*, a species where males exhibit strikingly long third legs used as weapons. The quality of the genome assembly is such that over 90% of the sequence is captured in 13 scaffolds. The most exaggerated legs in males were particularly enriched in sex-biased genes, indicating a specific signature of gene expression in association with sex-specific trait exaggeration. We also found that male-biased genes showed patterns of fast evolution compared to non-biased and female-biased genes, indicative of directional or relaxed purifying selection. Interestingly, we found that female-biased genes that are expressed in the third legs only, but not male-biased genes, were over-represented in the X chromosome compared to the autosomes. An enrichment analysis for sex-biased genes along the chromosomes revealed that they can arrange in large genomic regions or in small clusters of two to four consecutive genes. The number and expression of these enriched regions were often associated with the exaggerated legs of males, suggesting a pattern of common regulation through genomic proximity in association with trait exaggeration. Our findings shed light on how directional sexual selection drives sex-biased gene expression and genome architecture along the path to trait exaggeration and sexual dimorphism.

Introduction

Sexual dimorphisms, or phenotypic differences between males and females of the same species, is one of the most common sources of phenotypic variation in nature^{1,2}. Understanding how this process is regulated in a sex-specific manner at the genomic level, despite the shared genome between males and females, still poses an important challenge³. Differences in gene expression have emerged as a common mechanism to explain phenotypic differences among individuals sharing almost the same genome^{4,5}. In the last decade, a large number of studies have characterized genes with sex-biased expression in a variety of species, leading to an emerging framework attempting to link sex-biased gene expression to phenotypic divergence of the sexes⁴⁻⁶. However, these studies have mostly focused on adult gonads or whole-body transcriptomic datasets, which are unsuited to understand how secondary sexual characters are built during development^{4,5}. Among the countless examples of sexual dimorphism, some species have evolved extreme patterns whereby males, generally, develop such drastic phenotypes that they appear exaggerated compared to homologous traits in the other sex or to other body parts⁷⁻¹⁰. These growth-related sexual traits have received lots of attention in developmental genetics, but we still lack a general understanding of the genomic regulation underlying their development^{7,11-21}.

We aimed here to assess the effect of sexual selection on the sex-specific regulation of gene expression and genome architecture in the water strider *Microvelia longipes* (Heteroptera, Gerromorpha, Veliidae), an emerging model in the field of sexual selection and trait exaggeration²² (see also companion paper Toubiana et al.). *M. longipes* is a hemimetabolous insect that displays a striking case of sex-specific exaggerated trait where some males develop extremely long rear legs compared to females. The length of the rear legs (third legs) in males is under strong directional sexual selection and these legs are used as weapons to kick opponents away from the sites where females mate and lay eggs²². Such directional selection is associated with the evolution of disproportionate growth (i.e. hyperallometry) in male rear legs. Here we study the genomic regulation underlying the elaboration of this exaggerated phenotype in order to shed light on the role of sexual selection in shaping genome evolution. We generated a high-quality genome of *M. longipes*, with chromosome-scale resolution, and compared the expression, molecular evolution and genomic location of sex-biased genes in the three pairs of legs at a developmental stage where the legs diverge between the sexes (companion paper Toubiana et al.). Combined, our approach first identified signatures of trait exaggeration in terms of sex-biased gene expression patterns and sequence evolution. Second, it characterized chromosomes and genomic regions that are enriched in the sex-biased genes associated with the directional sexual selection applying to male exaggerated legs in *M. longipes*.

Results

De novo assembly and automatic annotation of *M. longipes* genome

To study the genetic mechanisms underlying extreme growth of male legs, we generated *de novo* the genome of *M. longipes* (Figure 1A) using lines established from a French Guiana population that were inbred through 15 sib-sib crosses²². Next generation sequencing and k-mer frequency distribution in raw sequencing reads estimated *M. longipes* genome size to about 0.67Gb (Supplementary table 1; see material and methods). Genome assembly combined multiple mate-pair Illumina libraries, PacBio, and Dovetail Hi-C/Hi-Rise libraries²³⁻²⁵ (Supplementary table 1; see Material and Methods). The final assembly generated chromosome-length scaffolds with scaffold N50=54.155 Mb and contig N50=216.72 Kb (Supplementary table 1). Over 90% of *M. longipes* genome is represented in the thirteen largest scaffolds (Supplementary table 1; see material and methods).

We then used automatic genome annotation, supported by *de novo* transcriptome-based gene models, to build the gene set of *M. longipes* (see material and methods). This analysis predicted 26 130 genes and 27 553 transcripts. BUSCO analysis, based on the 2018 insect dataset²⁶, revealed that 96% of gene models are present; among these 92% are complete, 3% are fragmented and 1% are duplicated (Supplementary figure 1). We therefore conclude that *M. longipes* genome is more complete than most available insect genomes²⁷.

Variation in gene expression explains differences in leg length

In *M. longipes*, the most obvious difference between legs is reflected in their size (Figure 1B)²². Principal component analysis (PCA) of adult male and female leg length from two isogenic lines, selected for differences in absolute leg length²², revealed that the first major component of variation encompassed 97% of the total variation and separated samples based on serial homolog in both sexes, with male third legs covering most of the divergence (Figure 1B). The second and third major axes of variation discriminated males from the two lines and sexes, respectively, although they only contributed 2% to the total variation (Figure 1B-C).

To test whether these phenotypic differences correlate with variation in gene expression, we sequenced the leg transcriptomes of males and females from these lines at the 5th nymphal instar – the developmental stage where we observed a burst of growth (See companion paper, Toubiana et al.). If a strict correlation existed between leg length and gene expression, we should predict samples to cluster by serial homolog, then by line, and finally by sex. Instead, the three first major axes of variation in the leg transcriptomes clustered samples based on line (Figure 1D). The line effect accounted for about 60% of the total variation in gene expression, thus potentially hiding signals associated with differences between legs and sexes. We therefore corrected for this line effect, using within class analysis, and generated a new PCA that now separates the sexes in PC1 (28.3% of variation in gene expression), and the serial homologs of the same sex in PC3 (10.3% of the variation in gene expression) (Figure 1E). We conclude that the three main components involved in leg length variation were retrieved in our transcriptome datasets. Yet, conversely to leg morphologies, homologous legs from the two sexes are now more divergent in terms of gene expression than serial homologs from the

same sex. We hereafter focus on the effect of sex on gene expression as it represents a major factor underlying leg exaggeration through differences in allometric coefficients²² (Supplementary figure 2).

Leg exaggeration and sex-biased gene expression

The legs of *M. longipes* males and females differ in their scaling relationships and degree of exaggeration²² (Figure 1A-C, Supplementary figure 2). To determine more specifically the patterns of gene expression underlying the observed sexual dimorphism in scaling relationships, we compared expression profiles of homologous legs between the sexes. All three legs consistently showed over twice as many female-biased than male-biased genes, with the third legs having the highest total number of sex-biased genes (Figure 2A-C). Interestingly, the average degree of differential expression (log2 Fold Change) in male-biased genes correlated with the patterns of leg growth such that the second and the third legs, which are hyper-allometric, showed higher degree of male-biased expression than the first leg, which is iso-allometric (Figure 2D). This pattern was consistent with the general overexpression of male-biased genes in the two exaggerated legs, especially the third, compared to the first legs (supplementary figure 3). This correlation was however absent for female-biased genes (Figure 2D; supplementary figure 3). More than two thirds of the male-biased genes in the first and second legs were shared with the most exaggerated legs, whereas about two thirds of male-biased genes in the third leg were not shared with the other legs (Figure 2E). A hierarchical clustering analysis separated males' third legs from the other serial homologs in both sexes, confirming that on average the sexual dimorphic expression of these 354 genes is restricted to the most exaggerated leg (Figure 2 F). Furthermore, the third legs showed a high number of female-biased genes, despite the lack of exaggeration in females, suggesting that the development of this extreme sexual dimorphism may also result from the active regulation of specific genes in female's third legs or their active repression in male's third legs. Altogether, these results show that the exaggerated third legs of males display unique patterns of sex-biased genes, in terms of number and/or levels of expression, compared to the two other serial homologs.

Male exaggerated legs are enriched in both leg- and sex-biased genes.

Theory predicts that pleiotropy may constrain the evolution of sex-biased gene expression²⁸. Our analysis of sex-biased genes reports a possible crosstalk between tissue and sex regulations in the context of trait exaggeration, which in turn may relax genetic constraint on sex-biased genes. Alternatively, sexual dimorphism may result from post-transcriptional regulation, possibly resulting in rather broad expression of sex-biased genes²⁹. To test these hypotheses, we combined our list of leg-biased genes (genes differentially expressed between the legs of the same sex, see companion paper Toubiana et al.) with the list of sex-biased genes (genes differentially expressed between the same leg in males and females), and performed a four-dimension comparison of fold-change between first and third legs as serial homologs in the Y-axis and males and females in the X-axis (Figure 3).

Interestingly, genes that are up regulated in male compared to female third legs also tend to be up regulated in male third legs compared to first legs (Fisher's exact test; p -value < 0.05) (Figure 3A). Conversely, female-biased genes in the third legs tend to be down regulated in males' third legs compared to the first legs (Fisher's exact test; p -value < 0.05) (Figure 3A). These patterns were however absent when we analyzed the genes that are differentially regulated in female serial homologs (Fisher's exact tests; p -values > 0.05) (Figure 3B). Likewise, these associations were lost when we looked at sex-biased genes in the first legs. Male-biased genes in the first legs tend to be under-represented among up regulated genes in male first legs (Fisher's exact test; p -value < 0.05) (Figure 3C). Moreover, female-biased genes did not show any leg-biased tendency in males (Fisher's exact test; p -value > 0.05) (Figure 3C). We obtained similar results when we selected leg-biased genes in females, except for the up-regulated genes in the first legs that tend to be female-biased (Fisher's exact test; p -value < 0.05) (Figure 3D).

In the second legs, which are mildly exaggerated in males, we also recovered enrichment between leg-biased genes in males and sex-biased genes in the second legs, though with fewer genes than the third legs (Fisher's exact tests; p -value < 0.05) (Supplementary figure 4). Overall, we found an enrichment of genes with both leg- and sex-biased expression that was particularly higher in male exaggerated third legs. This crosstalk highlights possible modularity by which leg-biased genes would have more easily acquired sex-biased expression to promote the exaggerated growth of male third legs without affecting other organs.

Male-biased genes evolved fast

We have shown that the pattern of sex-biased expression in *M. longipes* legs correlated in several aspects with the elaboration of the exaggerated third legs in males. In several species, sex-biased genes display higher rate of evolution compared to unbiased genes but relatively little is known about their sequence evolution in the context of trait exaggeration. We classified expressed genes in *M. longipes* legs based on their sex-biased expression pattern (male-biased, female-biased and unbiased, respectively) and compared their sequence evolution with four other *Microvelia* species (Figure 4; Supplementary table 2). We generally found that male-biased genes evolved faster than female-biased and unbiased genes; female-biased genes being the slowest evolving (Figure 4; Supplementary table 2). This pattern was true for all species-pair comparisons and remained consistent across legs or when we separated X-linked genes from autosomal genes (Figure 4; Supplementary table 2). Conversely to what has been observed in several organisms, we failed to detect any pattern of faster-X evolution (Supplementary table 2).

Sex-biased gene expression and genome architecture

Theory predicts that sexual selection can be an important driver of genome evolution^{30,31}, and we sought to test this prediction by analyzing the distribution of sex-biased genes along the genome of *M.*

longipes. First, we identified the scaffold that corresponds to the X chromosome (see material & methods). Interestingly, our analysis detected enrichment in the X chromosome with female-biased genes of the third legs, but not the two other legs, compared to the autosomes (Figure 5A). The percentage of female-biased genes between the X chromosome and the autosomes in the different legs confirmed that the enrichment observed was caused by an accumulation on the X chromosome of genes specifically biased in the third legs of females (Figure 5A). In contrast, we did not find any significant under- or over-representation of male-biased genes from any of the three legs on the X chromosome (Figure 5A). Because of the known effect of dosage compensation on the expression of genes located on the X chromosome^{4,32-34}, we compared the levels of expression of all X chromosome genes between the sexes. This analysis failed to detect any significant global difference in expression of these genes between males and females, regardless of the legs (Supplementary figure 5).

Another potential effect of sexual selection on genome architecture is through rearrangements of genes or large genomic regions within chromosomes³⁵⁻⁴⁰. We therefore performed a fine-scale visualization of the sex-biased genes along the thirteen largest scaffolds covering *M. longipes* genome (Figure 5B; Supplementary figure 6). We recovered few large genomic regions of 2 Mb significantly enriched in sex-biased genes (Supplementary figure 6; but see Material and methods). These include a total of 100 sex-biased genes (about 2% of the total number of sex-biased genes in all three legs), indicating that only a fraction of sex-biased genes arranges in such large genomic regions (Figure 5B; Supplementary figure 6). Among these, three large enriched regions located on scaffolds #2 and #8 contained a total of 36 sex-biased genes in the third legs (11 female-biased and 25 male-biased) (Figure 5B). Interestingly, two of these regions were specific to the third leg whereas the third indicated an enrichment of male-biased genes that was common to the three pairs of legs but with a higher degree of differential expression in the third legs (Figure 5B). In these regions, we could notably identify several unknown genes (10 out of 36 genes) including a cluster of four that were all strongly male-biased. Protein motif prediction, using Pfam, revealed a conserved domain of several transmembrane motifs in these four protein-coding genes.

Finally, we looked for small clusters of consecutive genes with similar patterns of expression in an attempt to assess common regulation⁴¹. We found that over 15% of male-biased and over 20% of female-biased genes arrange in clusters of two to four genes in the third legs, while only about 8.5% and 10% respectively are expected under a null hypothesis of random gene order (p -value < 0.05; see material and method) (Figure 5C; Supplementary figure 7). More specifically, we found up to seven clusters of four consecutive sex-biased genes in the third legs while only a maximum of two of them were expected by random permutation (Figure 5C). In the second pair of legs, we also found that about 10% of the male- and female-biased genes are arranged in clusters of at least two genes, while 2 to 3% were expected by random permutation (p -value < 0.05; Figure 5C; Supplementary figure 7). In

comparison with the third legs, clusters of male- and female-biased genes did not exceed two and three consecutive genes, respectively (Figure 5C). Male-biased genes in the first legs did not show any enrichment in clusters, and only one such cluster of three genes was detected (p-value > 0.05; Figure 5C; Supplementary figure 7). However, we found an enrichment of female-biased gene clusters, including 2 clusters of 3 consecutive genes (Figure 5C; p-value < 0.05; Supplementary figure 7).

Molecular function of sex-biased genes

Finally, we aimed to determine the molecular function of the sex-biased genes in our dataset. Gene ontology (GO) term analyses revealed enrichment in translation, metabolic processes and Wnt signaling pathways for the male-biased genes in the third legs (Supplementary table 3). The “translation” GO term uncovered enrichment for several ribosomal proteins also known to play an essential role in cell proliferation in response to ribosomal stress⁴². We also identified enrichment in molecular functions such as transferase activity indicative of possible post-transcriptional regulation differences between the two sexes. Female-biased genes in the third legs were enriched in various functions such as transcription factor, kinase, or GTPase activities that are probably involved in regulating biological processes such as transcription, metabolism, or signal transduction (Supplementary table 3).

Discussion:

Uncovering the genomic regulation underlying the process of phenotypic divergence between males and females is central to our understanding of morphological evolution^{5,8,30,43}. As an emerging model, *Microvelia longipes* offers exciting life history and ease of experimental manipulation to study how sexual selection can drive morphological and genomic adaptation²² (also see companion paper Toubiana et al.). In a previous study, we found that the evolution of male third leg exaggeration was associated with intense competition between conspecific males to dominate egg-laying sites²². The current study sheds light on the regulatory processes, both developmental and genomic, underlying this sex-specific exaggeration.

We identified a signature of leg exaggeration among sex-biased genes. Consistent with studies of sex-biased gene expression^{5,44-46}, we found that the degree of sexual dimorphism is associated with different patterns of expression among these genes (Figure 2&3). In our dataset, the most exaggerated legs mobilized more differentially expressed genes between the sexes and a higher degree of differential expression, especially in male-biased genes. Consistent with their potential role in driving the evolution of sexual dimorphism, male-biased genes were also relatively fast-evolving in comparison to female-biased and unbiased genes. This pattern is however common to all three legs, raising the question of pleiotropy in constraining adaptive evolution between the sexes. We found that a large proportion of sex-biased genes, especially in the third legs, displayed also tissue-specific expression (Figure 3A). Along with other studies showing less pleiotropy for sex-biased than unbiased

genes^{47,48}, our results point out modularity as possible regulatory mechanism whereby tissues can evolve biased expression freely and acquire sex-specific phenotypes with little deleterious effects.

Comparing the three pairs of legs in *M. longipes* offers a unique opportunity to understand the genetic mechanisms underlying sex-specific structures as they present different types and degrees of sexual dimorphism, likely reflecting distinct selective pressures²²; (Toubiana et al. companion paper). For example, the development of the sex-combs, which are known to be under stabilizing selection^{49,50}, occurs in the first legs of males during the 5th instar⁵¹. In contrast, the exaggerated growth of male third legs, which also occurs during the 5th nymphal instar (see companion paper Toubiana et al.), is an example of directional sexual selection that is absent, or at least strongly reduced, in the two other male legs²². We therefore used *M. longipes* legs as comparison to link the development of various sexual dimorphisms with the regulation and genomic location of sex-biased genes in the three pairs of legs. The X chromosome, for example, has been hypothesized to be a genomic hotspot for sexual selection where female beneficial dominant mutations and male beneficial recessive mutations are expected to accumulate^{30,32,33}. However, interpreting the representation of sex-biased genes on the X chromosome is often influenced by dosage compensation^{4,32,33}. In *Drosophila* for example, the scarcity of male-biased genes on the X chromosome was suggested to result, at least partially, from dosage compensation. Our analyses detected a significant enrichment of the X chromosome with female-biased genes in the third legs, but not male-biased genes. This suggests that dosage compensation also operates in all legs of *M. longipes*, and is therefore unlikely to be responsible for the enrichment observed. Therefore, it is possible that this ‘feminized’ X chromosome represents a mechanism for the resolution of sexual conflict during the evolution of extreme sexual dimorphism in *M. longipes* third legs.

Previous studies have reported large genomic regions and profound genomic rearrangements (e.g. large chromosome inversions) in association with sexually dimorphic characters^{35,37-40}. In contrast, we found relatively few large but many small clusters of sex-biased genes in *M. longipes* genome (Figure 5). Moreover, these small and large enriched regions seem to be associated with the extreme elongation of male third legs, in terms of number, specificity or degree of differential expression, and may highlight some important genes and regulatory processes involved in sex-specific trait exaggeration. It is also important to note that previous studies on genomic clusters of sex-biased genes were primarily conducted on primary sexual organs, such as ovaries and testes³⁵⁻⁴⁰. These tissues are highly complex, often express more sex-biased genes than secondary sexual traits and their evolution is considered to be under natural selection^{1,45,52-54}. Moreover, analyzing gene expression in these adult tissues does not capture the sex differences that are established during development. In the case of ontogenetic sexual dimorphism, it is expected that sexual selection will act on developmental regulatory processes^{5,11,55,56}. In this regard, our results offer the opportunity to test more accurately the

role of sexual selection on gene and genome evolution by directly linking the development of sexual dimorphism with patterns and genomic locations of sex-biased genes in the three pairs of legs.

Material and methods

Population sampling and culture

A *Microvelia longipes* population was collected during fieldwork in French Guyana in Crique Patate near Cayenne. The bugs were maintained in the laboratory at 25°C and 50% humidity in water tanks and fed on crickets. Inbred populations were generated as described in ²².

Statistics and leg measurements

All statistical analyses were performed in RStudio 0.99.486. For the PCA analysis on leg length, we used twenty males and females from each inbred population and measured them with a SteREO Discovery V12 (Zeiss) using the Zen software.

Sample collection, assembly and annotation of the *M. longipes* genome

Hundreds of individuals (males and females mixed) were collected from three inbred populations and frozen in liquid nitrogen before DNA extraction. Genomic DNA was extracted and purified using the Genomic-tip 20/G DNA extraction kit from Qiagen. Genome sequencing, using a mix of Illumina mate pairs and PacBio libraries, was performed at the Beijing Genomics Institute. Chromosome-length scaffold assembly was performed by Dovetail Genomics using Hi-C/Hi-Rise libraries. Supplementary table 1 summarizes the sequencing strategy employed.

The genome sequence was polished using Illumina libraries (Supplementary table 1) and Pilon ⁵⁷. Three different automatic annotation strategies, namely Braker, Maker and StringTie were tested to annotate the genome ⁵⁸⁻⁶⁰. These annotations were based on the leg transcriptomic dataset generated in this study (36 samples in total), a transcriptome from whole-body individuals collected at all developmental stages (1 sample) and a transcriptome from a third inbred population not mentioned in this study (18 samples). Braker and Maker pipelines also performed de novo automatic annotations.

Maker and Stringtie annotations yielded lower BUSCO quality and manual quality assessment using JBrowse revealed a relatively high number of gene fragmentations that were poorly supported by the alignments. We therefore used Braker annotation for further analyses (Supplementary figure 1).

For Braker annotation, we used Hisat2 alignment files from each transcriptomic sample to train Augustus with UTR option. Final annotation includes 26,130 genes and 27,553 transcripts.

Sample collection and preparation RNA-sequencing

We collected leg tissues from male and female 5th nymphal instars (two days after molting within a six hour time window) that belonged to two inbred populations that differ in average size (see ²²). All

individuals were raised in the same laboratory condition and fed with nine fresh crickets every day until the 5th instar. Individuals from the same inbred population were raised in the same water tank. The three replicates of each condition (lines, sexes and legs) correspond to a pool of 20 individuals chosen randomly (Supplementary figure 2). The dissection of the three pairs of legs, dissociated from the thorax, was performed in RNAlater (Sigma) using fine needles; each pair of legs was incubated immediately on ice in tubes filled with TRIzol (Invitrogen). RNA extractions were performed according to manufacturer protocol. The concentrations were assessed using the Qubit 2.0 Fluorometer (Invitrogen). Quality of RNA samples, library construction and sequencing were performed by Beijing Genomics Institute. The samples were sequenced using HiseqXten sequencing technology with a paired-end read length of 150bp.

Transcriptome assembly, mapping and normalization

Read quality was assessed with FASTQC version 0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/download.html>), and trimmed with TRIMMOMATIC version 0.32. Specifically, reads were trimmed if the sliding window average Phred score over four bases was <15 and only reads with a minimum length of 36bp were kept. Braker annotation was used as reference for read alignment and the transcriptome quantification. We obtained around 90% alignment rate on the genome and about 72% of uniquely mapped reads using Hisat2 method (Supplementary table 4). The latter condition was used for the estimation of transcript abundances and the creation of count tables (raw counts, FPKM and TPM tables) were performed using the StringTie pipeline (Supplementary table 5)^{60,61}. The abundance of reads per gene was finally calculated by adding the read counts of each predicted transcript isoforms.

Comparative transcriptomics: analyses of variance

Initially the transcriptomic approach was performed on three levels of comparisons; namely the lines, the sexes and the legs (Supplementary figure 2). The first three axes of variation in gene expression explained 57.1% of the total variation and separated the two inbred populations (Figure 1D). This confirms the genetic similarity that exists between individuals of the same inbred population. In order to correctly assess the influence of sex and leg comparisons on gene expression variance, we corrected for the line effect using a Within-Class Analysis⁶². After correction, the first major axis of variation separated male and female conditions, while PC3 explained the variation between legs (Figure 1E).

Identification of sex-biased genes

We first filtered transcripts for which expression was lower than 2 FPKM in more than half of the samples after combining the two inbred populations (12 samples total). Transcripts with average expression that was lower than 2 FPKM in both males and females were also discarded. The number of reads per “gene” was used to identify differences in expression among the different conditions of

interest using DESeq2⁶³. Differential expression analyses between males and females were performed on the two lines combined as we aimed to identify genes involved in male third leg exaggeration, which is a common feature to both lines. We repeated the differential expression analyses in lines separately and generally found high expression similarities between the two lines (data not shown). The differential expression analysis was also corrected for the line effect and we called sex-biased genes any gene with a fold-change > 1.5 and a Padj < 0.05. Other differentially expressed genes that do not fit such criteria (e.g. fold-change > 0 and a Padj < 0.05) are specified in the appropriate section.

Interaction between leg and sex regulations

In order to detect a possible interaction between leg and sex regulations, we combined our list of sex-biased genes with another list of genes that were identified as differentially expressed between legs of the same sex (i.e., leg-biased genes) (see companion paper Toubiana et al.). Using Fisher's exact tests, we then identified possible enrichment of genes with both sex- and leg-biased expression among the genes expressed within each tissue.

We also used the interaction model implemented in DESeq2. For this, we first filtered for lowly expressed transcripts by removing all transcripts for which the expression was lower than 2 FPKM in more than two-thirds of the samples (36 samples total). This filtering process leaves 9364 transcripts for the differential expression analysis. The interaction model between legs and sexes, after correcting for the line effect, revealed 2 genes for the third and first leg comparison and no gene for the third and second leg or the second and first leg comparisons. When we looked at these two genes in the differential expression analyses without the interaction effect, we found that one of them (g7203) was detected as male-biased in the third leg but not leg-biased whereas the second gene (g23967) was both male-biased and upregulated in the third leg compared to the first.

Hierarchical clustering

Average expressions of sex-biased genes in the different tissues were clustered using Euclidean clustering in the R package PVCLUST version 1.3-2⁶⁴ with 1000 bootstrap resampling. Heatmaps and clustering were performed using the log2(TPM) average expression of each gene from each tissue. Heatmaps were generated using the R package GPLOTS version 3.0.1.1.

Sex-biased gene distribution between chromosomes

Sex chromosome identification

Sex in the Gerromorpha is genetically determined and established by either the XX/XY or XX/X0^{65,66}. In *M. longipes*, Illumina genomic sequencing containing only males was used to align genomic reads against *M. longipes* genome and extract the genomic coverage of each scaffold. The scaffold 1893 was the only scaffold among the 13 biggest scaffolds (more than 90% of the genome) that presented twice less coverage than the other scaffolds. To finally assess the identity of the X chromosome in *M.*

longipes, we monitored the gene expression and found that the scaffold 1893 included both male- and female-biased genes, excluding this scaffold to be the Y chromosome. We also looked for a possible Y chromosome by identifying scaffolds with similar genomic coverage as the X chromosome but containing genes with only male-biased expression. We did not find any among the fifty largest scaffolds, suggesting that *M. longipes* has a XX/X0 sex determination system or presents a highly degraded Y chromosome.

Genomic distribution of sex-biased genes

We identified the genomic location of each gene and selected genes with a fold change superior to 1.5 between males and females as sex-biased genes ($\text{Padj} < 0.05$). Over- or under-representation of sex-biased genes in the X chromosome (scaffold 1893) compared to the autosomes (12 other largest scaffolds) was tested using Fisher's exact tests.

Estimation of dosage compensation

To compare the average level of gene expression between males and females in the X chromosome we first selected expressed genes with $\text{FPKM} > 2$ in at least half of the samples (12 samples per leg). We also averaged gene expressions between replicates and lines before testing for differences in expression (Wilcoxon tests on the $\log_2(\text{FPKM})$).

Detection of large sex-biased gene regions

To detect large chromosomal regions enriched in sex-biased genes we developed a bootstrapping method based on sliding windows of 2 Mb with a step size of 100 kb (Supplementary figure 8). Gene density calculation revealed that on average, genes are found every 20 kb in *M. longipes* genome. This pattern was homogeneous among chromosomes (Supplementary table 6). We therefore split each chromosome into bins of 100 kb and generated sliding windows of 2 Mb (20 bins) to include approximately 100 genes per window in the analysis (Supplementary figure 8B, Supplementary table 6). We used two scaffolds, one scaffold with two enriched regions (scaffold 2) and a scaffold with no enriched region (scaffold 1914), to repeat the analysis with smaller regions (1 Mb, 500 kb, 250 kb and 120 kb). We found similar results in both scaffolds, regardless of the size of the region, indicating that our analysis is statistically robust and is not missing information.

Fold-change reassignment and gene position

From the DESeq2 analyses, all expressed genes were associated with a $\log_2$ fold change (Log_2FC) and a p-value (Padj). Unexpressed genes ($\text{FPKM} < 2$) were assigned a $\log_2\text{FC}$ of 0 and a p-value of 1. Among the expressed genes, we switched the $\log_2\text{FC}$ to 0 for the unbiased genes ($\text{Padj} > 0.05$), in order to directly assess sex-biased genes based on $\log_2\text{FC}$ values (Supplementary figure 8A).

In a second step, we merged the dataset on sex-biased expression with the gene positions (Supplementary figure 8A).

Genome-wide detection of sex-biased gene regions

A mean log2FC was calculated for each window and reported along the chromosomes to reveal genome-wide regions of sex-biased genes (Supplementary figure 8B).

Bootstrapping method

To test whether these regions are significantly enriched in male or female-biased genes, we developed a bootstrap approach (Supplementary figure 8C). As the mean expression level of a gene influences the log2FC value (i.e. genes with low expressions are more likely to have high log2FC values and genes with high expression are more likely to be differentially expressed), we created 5 categories of genes based on their expression levels (baseMean values from DESeq2 tables). We then reassigned randomly, within each category, the log2FC at each gene position in the genome. This step was performed 100 000 times, therefore generating 100 000 random log2FC profiles.

Finally, to test for the significant enrichment of gene in these regions, we compared for each bin the observed log2FC values with the log2FC values generated from the bootstrap. To call for significantly enriched region of sex-biased genes, we identified regions for which the observed log2FC value was higher (male-biased) or lower (female-biased) than the 97500 (out of 100 000) bootstrap values generated randomly (Supplementary figure 8D). We finally applied a Bonferroni correction, correcting the bootstrap values by the total number of independent windows in the genome ($n=300$), leading to a bootstrap threshold of 99992. We note that this analysis is rather stringent (possibly missing other important enriched regions), but the highlighted regions are unambiguous (low sensitivity but high robustness).

Detection of clusters of consecutive sex-biased genes

This analysis was primarily inspired from Boutanaev et al.³⁵. In short, we determined clusters by ordering genes along the genome and detecting regions of consecutive male- or female-biased genes ($P_{adj} < 0.05$). To avoid identifying clusters overlapping two different chromosomes, we performed this analysis on the thirteen largest scaffolds separately. We then tested whether the observed distribution of genes differed from a stochastic distribution by randomly assigning a genomic position to unbiased, male-biased and female-biased genes respectively. The proportion of sex-biased genes found in clusters as well as the distribution of cluster sizes was calculated by averaging 1000 iterations (Supplementary figure 8). P-values were extracted from the 95% fluctuation intervals calculated from the 1000 randomized iterations.

Gene Ontology analysis

Gene names and functions were annotated by sequence similarity against the NCBI ‘non redundant’ protein database using Blast2GO. The Blast2GO annotation was then provided to detect Gene Ontology terms enrichment (p-value < 0.05) using the default method of TopGO R package version 2.34.0.

Acknowledgements: We thank Jean-Nicolas Volff comments on the manuscript, Laurent Duret for help with dn/ds analyses and François Leulier, Kevin Parsons and Gaël Yvert for helpful discussions. This work was funded by ERC-CoG WaterWalking #616346 and Labex CEBA to A.K., a Ph.D. fellowship from *Ecole Doctorale* BMIC de Lyon to W.T, and Swiss National Foundation fellowship to R.A.

References:

- 1 Darwin, C. *The descent of man, and selection in relation to sex.* (J. Murray, 1871).
- 2 Punzalan, D. & Hosken, D. J. Sexual Dimorphism: Why the Sexes Are (and Are Not) Different. *Curr Biol* **20**, R972-R973, doi:10.1016/j.cub.2010.09.067 (2010).
- 3 Lande, R. Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters. *Evolution* **34**, 292-305 (1980).
- 4 Grath, S. & Parsch, J. Sex-Biased Gene Expression. *Annual Review of Genetics, Vol 50* **50**, 29-44, doi:10.1146/annurev-genet-120215-035429 (2016).
- 5 Mank, J. E. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**, doi:UNSP 0006 10.1038/s41559-016-0006 (2017).
- 6 Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* **8**, 689-698, doi:10.1038/nrg2167 (2007).
- 7 Lavine, L., Gotoh, H., Brent, C. S., Dworkin, I. & Emlen, D. J. Exaggerated Trait Growth in Insects. *Annu Rev Entomol* **60**, 453-472, doi:10.1146/annurev-ento-010814-021045 (2015).
- 8 Emlen, D. J. & Nijhout, H. F. The development and evolution of exaggerated morphologies in insects. *Annual Review of Entomology* **45**, 661-708 (2000).
- 9 Emlen, D. J. in *From Field Observations to Mechanisms: A Program in Evolutionary Biology.* (ed Princeton University Press) (2010).
- 10 Emlen, D. J. The Evolution of Animal Weapons. *Annu Rev Ecol Evol S* **39**, 387-413, doi:10.1146/annurev.ecolsys.39.110707.173502 (2008).
- 11 Emlen, D. J., Warren, I. A., Johns, A., Dworkin, I. & Lavine, L. C. A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. *Science* **337**, 860-864, doi:10.1126/science.1224286 (2012).
- 12 Gotoh, H. *et al.* The Fat/Hippo signaling pathway links within-disc morphogen patterning to whole-animal signals during phenotypically plastic growth in insects. *Developmental Dynamics* **244**, 1039-1045, doi:10.1002/Dvdy.24296 (2015).
- 13 Moczek, A. P. & Rose, D. J. Differential recruitment of limb patterning genes during development and diversification of beetle horns. *Proc Natl Acad Sci U S A* **106**, 8992-8997, doi:0809668106 [pii] 10.1073/pnas.0809668106 (2009).
- 14 Warren, I. A., Gotoh, H., Dworkin, I. M., Emlen, D. J. & Lavine, L. C. A general mechanism for conditional expression of exaggerated sexually-selected traits. *Bioessays* **35**, 889-899, doi:10.1002/bies.201300031 (2013).
- 15 Gotoh, H. *et al.* Developmental Link between Sex and Nutrition; doublesex Regulates Sex-Specific Mandible Growth via Juvenile Hormone Signaling in Stag Beetles. *Plos Genetics* **10**, doi:ARTN e1004098 10.1371/journal.pgen.1004098 (2014).
- 16 Gotoh, H. *et al.* Juvenile Hormone Regulates Extreme Mandible Growth in Male Stag Beetles. *Plos One* **6**, doi:ARTN e21139 10.1371/journal.pone.0021139 (2011).
- 17 Kijimoto, T. & Moczek, A. P. Hedgehog signaling enables nutrition-responsive inhibition of an alternative morph in a polyphenic beetle. *P Natl Acad Sci USA* **113**, 5982-5987, doi:10.1073/pnas.1601505113 (2016).

- 540 18 Kijimoto, T., Moczek, A. P. & Andrews, J. Diversification of doublesex function underlies morph-, sex-,
541 and species-specific development of beetle horns. *P Natl Acad Sci USA* **109**, 20526-20531,
542 doi:10.1073/pnas.1118589109 (2012).
- 543 19 Lavine, L. C. *et al.* Cloning and characterization of an mRNA encoding an insulin receptor from the
544 horned scarab beetle *Onthophagus nigriventris* (Coleoptera: Scarabaeidae). *Arch Insect Biochem Physiol*
545 **82**, 43-57, doi:10.1002/arch.21072 (2013).
- 546 20 Shelby, J. A., Madewell, R. & Moczek, A. P. Juvenile hormone mediates sexual dimorphism in horned
547 beetles. *J Exp Zool B Mol Dev Evol* **308**, 417-427, doi:10.1002/jez.b.21165 (2007).
- 548 21 Ohde, T. *et al.* Rhinoceros beetle horn development reveals deep parallels with dung beetles. *Plos*
549 *Genetics* **14**, doi:ARTN e1007651
550 10.1371/journal.pgen.1007651 (2018).
- 551 22 Toubiana, W. & Khila, A. Fluctuating selection strength and intense male competition underlie variation
552 and exaggeration of a water strider's male weapon. *Proc Biol Sci* **286**, 20182400,
553 doi:10.1098/rspb.2018.2400 (2019).
- 554 23 Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin
555 interactions. *Nat Biotechnol* **31**, 1119-1125, doi:10.1038/nbt.2727 (2013).
- 556 24 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry.
557 *Nature* **456**, 53-59, doi:10.1038/nature07517 (2008).
- 558 25 Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*
559 **13**, 278-289, doi:10.1016/j.gpb.2015.08.002 (2015).
- 560 26 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing
561 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-
562 3212, doi:10.1093/bioinformatics/btv351 (2015).
- 563 27 Yin, C. L. *et al.* InsectBase: a resource for insect genomes and transcriptomes. *Nucleic Acids Res* **44**,
564 D801-D807, doi:10.1093/nar/gkv1204 (2016).
- 565 28 Hodgkin, J. Seven types of pleiotropy. *Int J Dev Biol* **42**, 501-505 (1998).
- 566 29 Williams, T. M. & Carroll, S. B. Genetic and molecular insights into the development and evolution of
567 sexual dimorphism. *Nat Rev Genet* **10**, 797-804, doi:nrg2687 [pii]
568 10.1038/nrg2687 (2009).
- 569 30 Wilkinson, G. S. *et al.* The locus of sexual selection: moving sexual selection studies into the post-
570 genomics era. *J Evolution Biol* **28**, 739-755, doi:10.1111/jeb.12621 (2015).
- 571 31 Mank, J. E. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *Am Nat*
572 **173**, 141-150, doi:10.1086/595754 (2009).
- 573 32 Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev*
574 *Genet* **7**, 645-653, doi:10.1038/nrg1914 (2006).
- 575 33 Dean, R. & Mank, J. E. The role of sex chromosomes in sexual dimorphism: discordance between
576 molecular and phenotypic data. *J Evolution Biol* **27**, 1443-1453, doi:10.1111/jeb.12345 (2014).
- 577 34 Bachtrog, D., Toda, N. R. T. & Lockton, S. Dosage Compensation and Demasculinization of X
578 Chromosomes in *Drosophila*. *Curr Biol* **20**, 1476-1481, doi:10.1016/j.cub.2010.06.076 (2010).
- 579 35 Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y. & Nurminsky, D. I. Large clusters of co-
580 expressed genes in the *Drosophila* genome. *Nature* **420**, 666-669, doi:10.1038/nature01216 (2002).
- 581 36 Dorus, S. *et al.* Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat*
582 *Genet* **38**, 1440-1445, doi:10.1038/ng1915 (2006).
- 583 37 Kim, K. W. *et al.* A sex-linked supergene controls sperm morphology and swimming speed in a songbird.
584 *Nat Ecol Evol* **1**, 1168-1176, doi:10.1038/s41559-017-0235-2 (2017).
- 585 38 Knief, U. *et al.* A sex-chromosome inversion causes strong overdominance for sperm traits that affect
586 siring success. *Nat Ecol Evol* **1**, 1177-1184, doi:10.1038/s41559-017-0236-1 (2017).
- 587 39 Kupper, C. *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nature*
588 *Genetics* **48**, 79-+, doi:10.1038/ng.3443 (2016).
- 589 40 Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff
590 (*Philomachus pugnax*). *Nature Genetics* **48**, 84-+, doi:10.1038/ng.3430 (2016).
- 591 41 Hurst, L. D., Pal, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic gene order. *Nature*
592 *Reviews Genetics* **5**, 299-310, doi:10.1038/nrg1319 (2004).
- 593 42 Zhou, X., Liao, W. J., Liao, J. M., Liao, P. & Lu, H. Ribosomal proteins: functions beyond the ribosome.
594 *J Mol Cell Biol* **7**, 92-104, doi:10.1093/jmcb/mjv014 (2015).
- 595 43 Stern, D. L. & Emlen, D. J. The developmental basis for allometry in insects. *Development* **126**, 1091-
596 1101 (1999).
- 597 44 Zinna, R. *et al.* Sexual dimorphism and heightened conditional expression in a sexually selected weapon
598 in the Asian rhinoceros beetle. *Molecular Ecology* **27**, 5049-5072, doi:10.1111/mec.14907 (2018).

- 45 Harrison, P. W. *et al.* Sexual selection drives evolution and rapid turnover of male gene expression. *P Natl Acad Sci USA* **112**, 4393-4398, doi:10.1073/pnas.1501339112 (2015).
- 46 Pointer, M. A., Harrison, P. W., Wright, A. E. & Mank, J. E. Masculinization of Gene Expression Is Associated with Exaggeration of Male Sexual Dimorphism. *Plos Genetics* **9**, doi:ARTN e1003697 10.1371/journal.pgen.1003697 (2013).
- 47 Dean, R. & Mank, J. E. Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression. *Am Nat* **188**, E74-E84, doi:10.1086/687526 (2016).
- 48 Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M. & Ellegren, H. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *Am Nat* **171**, 35-43, doi:10.1086/523954 (2008).
- 49 Ahuja, A. & Singh, R. S. Variation and evolution of male sex combs in *Drosophila*: nature of selection response and theories of genetic variation for sexual traits. *Genetics* **179**, 503-509, doi:10.1534/genetics.107.086363 (2008).
- 50 Maraqa, M. S. *et al.* Constrained evolution of the sex comb in *Drosophila simulans*. *J Evol Biol* **30**, 388-400, doi:10.1111/jeb.13015 (2017).
- 51 Crumiere, A. J. J. & Khila, A. Hox genes mediate the escalation of sexually antagonistic traits in water striders. *Biol Lett* **15**, 20180720, doi:10.1098/rsbl.2018.0720 (2019).
- 52 Andersson, M. B. *Sexual selection*. (Princeton University Press, 1994).
- 53 Mank, J. E., Hultin-Rosenberg, L., Webster, M. T. & Ellegren, H. The unique genomic properties of sex-biased genes: insights from avian microarray data. *Bmc Genomics* **9**, 148, doi:10.1186/1471-2164-9-148 (2008).
- 54 Ma, W. J. *et al.* Tissue Specificity and Dynamics of Sex-Biased Gene Expression in a Common Frog Population with Differentiated, Yet Homomorphic, Sex Chromosomes. *Genes (Basel)* **9**, doi:10.3390/genes9060294 (2018).
- 55 Khila, A., Abouheif, E. & Rowe, L. Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. *Science* **336**, 585-589, doi:10.1126/science.1217258 (2012).
- 56 Badyaev, A. V. Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol Evol* **17**, 369-378, doi:Pii S0169-5347(02)02569-7 Doi 10.1016/S0169-5347(02)02569-7 (2002).
- 57 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).
- 58 Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769, doi:10.1093/bioinformatics/btv661 (2016).
- 59 Cantarel, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188-196, doi:10.1101/gr.6743907 (2008).
- 60 Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols* **11**, 1650-1667, doi:10.1038/nprot.2016.095 (2016).
- 61 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
- 62 Baty, F., Facompre, M., Wiegand, J., Schwager, J. & Brutsche, M. H. Analysis with respect to instrumental variables for the exploration of microarray data structures. *Bmc Bioinformatics* **7**, 422, doi:10.1186/1471-2105-7-422 (2006).
- 63 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 64 Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542, doi:10.1093/bioinformatics/btl117 (2006).
- 65 Castanhole, M. M. U., Pereira, L. L. V., Souza, H. V. & Itoyama, M. M. Spermatogenesis and karyotypes of three species of water striders (Gerridae, Heteroptera). *Genet Mol Res* **9**, 1343-1356, doi:10.4238/vol9-3gmr841 (2010).
- 66 Fairbairn, D. J., Kiseliova, O. & Muir, S. Variation in chromosome numbers and the sex determination system in the Gerromorpha with special reference to the family Gerridae (Hemiptera). *Aquat Insect* **37**, 127-144, doi:10.1080/01650424.2016.1167222 (2016).

Figure legends

Figure 1: Association between variation in leg length and variation in gene expression. (A) *Microvelia longipes* in the wild. (B-C) Principal Component Analysis (PCA) on male and female leg length from the long-leg and short-leg selected inbred populations²². (B) The first PCA (Dim1) explains primarily differences between legs of the same sex while the second PCA (Dim 2) explains differences between inbred populations, specifically in males. (C) The third PCA (Dim 3) explains the differences between sexes. (D-E) Principal Component Analysis (PCA) on the whole transcriptomic dataset. (D) The three first PCAs (Dim1, 2, 3) recapitulate the variance between the Big (blue) and Small (green) lines. (E) Within-Class analysis after correcting for line effects. Dimension 1 separates sexes while Dimension 3 separates legs. The inset represents the Within-Class correction for the line effects.

Figure 2: Signature of trait exaggeration among sex-biased genes. (A-C) Comparison of gene expression (log2 TPM+1) in male and female legs. Dots highlighted in purple and blue represent genes with significant difference in expression in females and males respectively. Insets indicate the number of female- and male-biased genes in each leg. (D) Differences in fold change (Wilcoxon tests) among the sex-biased genes identified in the three pairs of legs independently. (E) Venn-diagrams of the male-biased genes identified in the three pairs of legs. Size of the diagrams is proportional to the total number of genes. (F) Hierarchical clustering (1000 bootstraps) and heatmap based on average leg expression in males and females for the genes with significant male-biased expression specifically in the third legs of males (n=354).

Figure 3: Crosstalk between leg- and sex-biased genes. Light blue dots are sex-biased genes ($FC > 0$, $Padj < 0.05$), dark blue dots correspond to leg-biased genes ($FC > 0$, $Padj < 0.05$) and red dots are genes with both leg- and sex-biased expressions. For the leg-biased genes, we only illustrated the comparisons between the third and first legs. Data on the second legs are shown in Supplementary figure 4. “NS” indicates, “Non Significant” and “Sig” indicates “Significant”. (A) Comparison between sex-biased genes in the third legs and leg-biased genes in males. (B) Comparison between sex-biased genes in the third legs and leg-biased genes in females. (C) Comparison between sex-biased genes in the first legs and leg-biased genes in males. (D) Comparison between sex-biased genes in the first legs and leg-biased genes in females.

Figure 4: Sequence evolution of genes expressed in *M. longipes* legs. We categorized genes based on their expression profile and genomic location. Boxplots indicate the pair-wise comparisons of dNdS across the three legs between *M. longipes* and each of the four other *Microvelia* species. Statistical analyses are shown in Supplementary table 2.

Figure 5: Genomic distribution of sex-biased genes in *M. longipes*. (A) Percentage of male-biased, female-biased and unbiased genes (from top to bottom) in the X chromosome and the autosomes across the three pairs of legs. Biased-distribution of the sex-biased genes on the X chromosome was estimated using the Fisher’s exact tests: * p-value< 0.05. (B) Large genomic clusters of sex-biased genes along scaffold #2 and the scaffold #8. Clusters highlighted in blue represent genomic regions enriched in male-biased genes. Clusters highlighted in purple represent genomic regions enriched in female-biased genes. Solid red frame indicates genomic clusters enriched in male- and female-biased genes specifically in the third legs. Dotted red frame indicates genomic clusters enriched in male-biased genes in all three legs but with different degrees of fold-change, recapitulating the degree of leg length exaggeration. (C) Genomic clusters of consecutive male- (blue) or female-biased (purple) genes in the three pairs of legs. Cluster size indicates the number of consecutive genes. Note that the y axis is log scaled. Error bars indicate fluctuation intervals.

Legend

- NS legs_NS sexes
- NS legs_Sig sexes
- Sig legs_NS sexes
- Sig legs_Sig sexes