1   Research article

2   # Disentangling the causes for faster-X evolution in aphids

3   Jaquiéry J[1*], Peccoud J[2], Ouisse T[1], Legeai F[3,4], Prunier-Leterme N[3], Gouin A[3,4], Nouhaud
4   P[3,5], Brisson JA[6], Bickel R[6], Purandare S[7], Poulain J[8], Battail C[9], Lemaitre C[4], Mieuzet L[3], Le
5   Trionnaire G[3], Simon JC[3$], Rispe C[10$]

7   [1] Université de Rennes 1, CNRS UMR 6553 ECOBIO, Avenue du Général Leclerc, Campus
8   de Beaulieu, 35042 Rennes Cedex, France

9   [2] Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe
10   Ecologie Evolution Symbiose, TSA 51106, 86073 Poitiers Cedex 9, France

11   [3] IGEPP, INRA, Domaine de la Motte, F-35653 Le Rheu cedex, France.

12   [4] INRIA Centre Rennes - Bretagne Atlantique, GenOuest, Campus de Beaulieu, 35042
13   Rennes, France.

14   [5] Institute of Population Genetics, Vetmeduni Vienna, Vienna, Austria
15
16   [6] University of Rochester, Department of Biology, Rochester, NY 14627 USA
17
18   [7] Department of Entomology, Ohio Agricultural Research and Development Center, Ohio
19   State University, Wooster, OH 44691

20   [8] Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 2, rue
21   Gaston-Crémieux, CP 5706, 91057 Evry, France

22   [9] Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Centre National de
23   Génotypage (CNG), 2, rue Gaston-Crémieux, CP 5706, 91057 Evry, France

24   [10] BIOEPAR, INRA, ONIRIS, La Chantrerie, F-44307 Nantes, France

25   [$] joint last authors

27   ***Author for Correspondence:** Julie Jaquiéry, Université de Rennes 1, CNRS UMR 6553
28   ECOBIO, Avenue du Général Leclerc, Campus de Beaulieu, 35042 Rennes Cedex, France,
29   tel. ++33 2 23 23 56 92, fax ++33 2 23 23 50 26, Julie.Jaquiery@univ-rennes1.fr

30   **Running title:** X chromosome evolution in aphids

1

Abstract

Faster evolution of X chromosomes has been documented in several species and results from the increased efficiency of selection on recessive alleles in hemizygous males and/or from increased drift due to the smaller effective population size of X chromosomes. Aphids are excellent models for evaluating the importance of selection in faster-X evolution, because their peculiar life-cycle and unusual inheritance of sex-chromosomes lead to equal effective population sizes for X and autosomes. Because we lack a high-density genetic map for the pea aphid whose complete genome has been sequenced, we assigned its entire genome to the X and autosomes based on ratios of sequencing depth in males and females. Unexpectedly, we found frequent scaffold misassembly, but we could unambiguously locate 13,726 genes on the X and 19,263 on autosomes. We found higher non-synonymous to synonymous substitutions ratios *(dN/dS)* for X-linked than for autosomal genes. Our analyses of substitution rates together with polymorphism and expression data showed that relaxed selection is likely to contribute predominantly to faster-X as a large fraction of X-linked genes are expressed at low rates and thus escape selection. Yet, a minor role for positive selection is also suggested by the difference between substitution rates for X and autosomes for male-biased genes (but not for asexual female-biased genes) and by lower Tajima's D for X-linked than for autosomal genes with highly male-biased expression patterns. This study highlights the relevance of organisms displaying alternative inheritance of chromosomes to the understanding of forces shaping genome evolution.

2

Introduction

Sex chromosomes are major players in evolution. Besides their role in sex determination, sex chromosomes contribute to genomic conflicts (Rice 1984; Meiklejohn and Tao 2010; Soh et al. 2014), genetic incompatibilities and reproductive isolation (Coyne and Orr 2004; Saether et al. 2007; Kitano et al. 2009; Johnson and Lachance 2012). A pair of sex-determining chromosomes typically evolves toward reduced recombination rates (crossing overs), which eventually causes one of the sex chromosomes to gradually lose most of the chromosomal regions (loci) present in the alternate one (Charlesworth et al. 2005). These loci will thus be found in single copy in the sex that carries the degenerate, smaller sex chromosome. When the heterogametic sex is the male, sex chromosomes are denoted X and Y (e.g., in mammals), whereas when it is the female, sex chromosomes are noted W and Z (e.g., in birds). Alleles of loci present only on the X (W) are more exposed to selection in individuals of the heterogametic sex, facilitating the fixation of beneficial mutations and the purging of deleterious ones (Charlesworth et al. 1987). On the other hand, because males (XY) bear and transmit a single X chromosome, effective population size is smaller for the X compared to autosomes (Wright 1931; Caballero 1994; Caballero 1995). This increases the rate of fixation of slightly deleterious mutations on the X by genetic drift (Kimura 1983) (the same principles apply to ZW systems, so we ignore these in the following). Consequently, X-linked genes may evolve faster than autosomes ("faster-X" evolution) due to higher positive selection (rate of fixation of beneficial mutations) and/or higher genetic drift (rate of fixation of slightly deleterious mutations) (Vicoso et al. 2009, Mank et al. 2010). Faster evolution of X-linked proteins is supported by observations in a large panel of species (e.g., *Drosophila*, nematodes, mammals, birds, see Meisel and Connallon 2013 for a review). In some, a preponderant effect

3

78   of drift was demonstrated (Mank et al. 2010; Avila et al. 2014), while positive selection

79   would play a predominant role in other species (Baines et al. 2008; Hvilsom et al. 2012;

80   Langley et al. 2012; Mackay et al. 2012; Kousathanas et al. 2014; Sackton et al. 2014; Avila

81   et al. 2015).

82   In this context, organisms with atypical inheritance of sex chromosome can greatly

83   facilitate inferences about the processes contributing to the evolution of sex chromosomes

84   (Bachtrog et al. 2011). Among those, aphids, which present X0 males and XX females, have

85   equal effective population sizes for X chromosomes and autosomes (Jaquiéry et al. 2012a).

86   This eliminates one possible confounding factor of faster-X evolution, the smaller effective

87   population size of the X, helping to properly test its causes. Aphids reproduce by cyclical

88   parthenogenesis, such that males and sexual females constitute only a short part of their life-

89   cycle, which is dominated by apomictic parthenogenetic (clonal) XX females (figure 1).

90   Males are produced asexually via the elimination of one X from the germ line (Wilson et al.

91   1997; Caillaud et al. 2002). As a result, X-linked recessive alleles are exposed to selection in

92   male aphids, just like in other X0 or XY males. However, because all sexually-produced

93   aphid eggs are XX females, all the progeny inherit their X from males and sexual females in

94   equal proportions. This difference from other heterogametic systems, where progeny present

95   even sex ratios, has deep consequences for the evolutionary trajectory of the aphid X

96   chromosome (Jaquiéry et al. 2012a). Furthermore, in contrast to standard systems, variance in

97   reproductive success between sexes, population expansion, bottlenecks and sex-biased

98   dispersal should not differentially affect sex chromosomes and autosomes in aphids (Jaquiéry

99   et al. 2012a). Mutation and recombination rates are also expected to be equal across

100   chromosomes because of their similar mode of inheritance and the complete absence of

101   crossing overs in males (Jaquiéry et al. 2012a). These similarities between X chromosomes

4

102   and autosomes make aphids exceptionally useful to pinpoint the causes of faster-X evolution,

103   since the factors mentioned above need not be accounted for. Still, a notable difference

104   between X and autosome in aphids is the theoretical propensity of the X to accumulate

105   sexually antagonistic mutations beneficial for males and detrimental to asexual females,

106   which also is the consequence of cyclical parthenogenesis combined with the inheritance of

107   the X (Jaquiéry et al. 2013).

108   Empirical analyses on a small subset of genes of the pea aphid (*Acyrthophon pisum*)

109   showed that X-linked genes evolve faster than autosomal genes (Jaquiéry et al. 2012a) and

110   that genes expressed predominantly in males (hereafter "male-biased" genes) predominantly

111   locate on the X (Jaquiéry et al. 2013). Subsequent genome-wide analyses did not however

112   support faster-X evolution (Purandare et al. 2014), and mitigated the degree of enrichment of

113   male-biased genes on the X (Purandare et al. 2014; Pal and Vicoso 2015). These

114   discrepancies however likely stem from the fact that these two studies did not assign

115   individual genes to chromosome types, but entire scaffolds, which present assembly errors (as

116   shown by Bickel et al. 2013 and suggested by Jaquiéry et al. 2013). Misassignment of genes

117   to chromosomes would indeed artificially decrease the contrast between X-linked and

118   autosomal genes.

119   Here, we aimed to go past these shortcomings in order to fully disentangle the causes

120   for faster-X evolution in aphids. For this, we first attributed genes to the X or to autosomes at

121   the scale of the entire genome in the pea aphid. On a large set of genes, we combined

122   estimates of substitution rates at the interspecific level with polymorphism data in pea aphid

123   populations and gene expression levels in the various genders and morphs. This allowed

124   assessing how relaxed selection (genetic drift) and adaptation contribute to the faster

125   evolution of the X chromosome in this system.

5

Material and Methods

Assignment of scaffold regions to the X and autosomes

Full-genome sequencing of females and males

An asexual aphid mother has the same diploid autosomal genome as her sons, but has two X chromosomes instead of just one (figure 1). We took advantage of this XX/X0 system to assign the pea aphid genome sequence (Acyr 2.0, Genbank accession GCA_000142985.2, IAGC 2010) to the X or to the autosomes by comparing sequencing depth along assembled scaffolds between mapped reads from females and from males of the same parthenogenetic lineage (clone). DNA from five asexual females and five winged and five wingless males of clone P123 (Simon et al. 2011) was extracted with the Qiagen DNeasy Blood and Tissue kit, following the manufacturer's protocol. Wing male polymorphism in this clone was used to determine the X copy that each male carried, based on the knowledge that the locus that controls this trait is X-linked (Caillaud et al. 2002) and is heterozygous in clone P123 (Frantz et al. 2010). Each individual was genotyped at seven polymorphic microsatellite markers (Peccoud et al. 2008) to confirm its identity. One of those markers, which is known to be X-linked (Caillaud et al. 2002), allowed us to confirm the nature of the X copy inherited by each male. Three DNA extracts were obtained: one from five females and one per genotype of males, using five individuals in each. These samples were sequenced on the Illumina Hiseq 2000 platform yielding 100 bp pair-end reads at ~43X coverage for the females sample and 25 to 30X coverage for each male type. Reads from each sample were mapped on scaffolds of the pea aphid genome assembly (Acyr 2.0) and on genome sequences of the bacterial symbionts of this pea aphid clone using the method described in Gouin et al. (2015) using Bowtie 2 (Langmead and Salzberg 2012) with proper insert sizes and parameters set as

6

149   default. Depth of coverage at each nucleotide position of the reference genome was recorded

150   and single nucleotide polymorphisms (SNPs) were identified using GATK's Haplotype Caller

151   (McKenna et al. 2010; DePristo et al. 2011). The raw sequence data has been deposited in the

152   SRA division of Genbank (project accession: ERP022905 and PRJNA385573).

153

154   *Comparison of sequencing depth between males and females*

155   The following analysis was performed in R (R Development Core Team 2015). We analyzed

156   genome positions covered by 20 to 70 reads in the asexual female sample, a range chosen to

157   eliminate regions with low-coverage and regions with suspiciously high coverage (potentially

158   duplicated or repeat-rich regions). Since overall coverage was slightly higher for one of the

159   male types (~30X) than for the other (~25X), we normalized the depth of coverage data of the

160   second male type (multiplying coverage estimates by a 30/25 ratio). We then averaged depth

161   of coverage at each base position over male types. The ratio of median coverage depth of

162   males to median coverage depth of the female sample was calculated on 10-kilobases (kb)

163   scaffold windows sliding by 2-kb steps. A single window was used for scaffolds shorter than

164   10 kb. We expect the ratio of median coverage depth to be twice larger for autosomal regions

165   than for the X chromosome. Accordingly, this ratio had a clearly bimodal distribution

166   (supplementary figure S1), with modes at 0.34 and 0.66. We assigned a 10-kb window to the

167   X if its ratio ranged between 0.2 and 0.445, to autosomes if it ranged between 0.53 and 1,

168   whereas the region was tagged as "ambiguous" if it ranged between 0.445 and 0.53. Windows

169   assigned to the same chromosome type and which were separated by less than four

170   consecutive "ambiguous" windows were aggregated into a scaffold region we call a "block".

171   A whole block, including its "ambiguous" windows, was assigned to the corresponding

172   chromosome type.

7

173

174 *Comparison of male and female genotypes*

175 Inheritance of single nucleotide polymorphisms (SNPs) also informs on the type of

176 chromosome carrying a scaffold block. In fact, SNPs that are heterozygous in females but are

177 also heterozygous in males are necessarily located on autosomes. Conversely, SNPs which are

178 heterozygous in females but homozygous in males must be on the X. This SNP-based

179 approach is however expected to be less powerful than the depth of coverage-based method

180 for genomic regions with low heterozygosity. Thus, we only used SNP data to validate X/A

181 assignments based on depth of coverage ratio (supplementary figure S2). A position was

182 determined as heterozygous if the rarest allele was represented in at least 25% of reads,

183 otherwise it was considered as homozygous. Assignment of SNPs to chromosome types was

184 performed according to the genotypes of males, as described above. SNPs showing

185 inconsistent genotypes (e.g., females and males of one type are both heterozygous while

186 males of the other type are homozygous) were not assigned. SNP-based assignments were

187 then visually compared to assignments based on depth of coverage (supplementary figure S2).

188

189 **Assignment of predicted genes to the X and autosomes**

190 We used the 36,990 genes that constitute the gene prediction v2.1 for the Acyr 2.0 genome

191 assembly available at http://bipaa.genouest.org/is/aphidbase/. Each of these genes was

192 determined as X-linked or autosomal if the full length of its coding sequence (CDS) was

193 comprised in a single scaffold block or was spread over several scaffold blocks assigned to

194 the same type (either X, or A). Genes that could not be unambiguously assigned (mainly

8

195   because they located on "ambiguous" blocks) were removed from further analyses. We also

196   excluded 589 predicted genes that corresponded to rRNA (non-coding DNA).

197

Sex-biased gene expression

199   We used the eight RNAseq libraries from Jaquiéry et al. (2013) to characterize gene

200   expression patterns between morphs. Briefly, these eight libraries correspond to whole

201   insects, with three male libraries, three parthenogenetic female libraries and two sexual

202   female libraries – different libraries in each morph representing biological replicates – using

203   adults of a single clone of *A. pisum* (clone LSR1). Details regarding aphid rearing, library

204   preparation and sequencing are provided in Jaquiéry et al. (2013). Libraries were mapped on

205   Acyr 2.0 as described previously. The number of reads covering each CDS was then counted.

206   Read counts were normalized with the R package DESeq with default parameters (Anders and

207   Huber 2010). For each gene, the effect of the morph (a three-level factor comprising male,

208   sexual female and asexual female) on expression was tested with a GLM (R package MASS,

209   Venables et al. 2002) with a quasi-poisson distribution of residuals, considering the different

210   libraries for each morph as replicates. *P*-values were corrected for multiple testing using the

211   Benjamini-Hochberg method implemented in R. Genes differentially expressed between

212   morphs ($p<0.05$ after adjusting for multiple testing) were then categorized according to their

213   pattern of expression in the different morphs as described in table 1.

Evolutionary rates

215   To assess substitution rates in X-linked and autosomal genes, sequences from another aphid

216   species were necessary. *Acyrthosiphon svalbardicum* was chosen, ensuring sufficient genetic

217   divergence, and reducing the risk of mutational saturation and of chromosomal

9

218   rearrangements between the two species. Asexual females of *A. svalbardicum* were collected

219   in Svalbard in 2009, and were then reared in the lab under 10:14 light:dark and 15°C on

220   *Dryas octopetala*. Ten females were then frozen into liquid nitrogen and kept for subsequent

221   RNA extraction performed using the RNeasy plant mini kit (Qiagen) according to

222   manufacturer's instructions. Two separate RNA extractions of 5 adults were performed. RNA

223   quality was checked on Bioanalyzer (Agilent) and quantified on Nanodrop (Thermo

224   Scientific). One sample made of a pool of 2 µg of the two independent RNA extractions was

225   subsequently sent to GATC Company for RNA paired-end sequencing. The raw sequence

226   data has been deposited in the SRA division of Genbank (project accession: PRJNA385897).

227   A *de novo* transcriptome assembly for *A. svalbardicum* was obtained following the

228   methods of Rispe et al. (2016). Low quality parts of the reads were trimmed from the right

229   ends with prinseq-lite (http://prinseq.sourceforge.net/) when the mean of phred score in a 20-

230   bp window was below 20. Reads longer than 20 bp after trimming were re-organized by pairs

231   (orphans were suppressed) and assembled with Trinity (Grabherr et al. 2011) using default

232   parameters. Coding regions were predicted using FrameDP (Gouzy et al. 2009). Reciprocal

233   BLASTN (Altschul et al. 1990) searches between CDSs of *A. svalbardicum* and of *A. pisum*

234   were carried out with an e-value threshold of 10^{-8}. The following steps were performed with

235   an R script. Reciprocal best hit criterion was used to identify putative orthologous genes

236   between the two species. These were aligned by the pairWiseAlignment function of the

237   Biostrings package (Pages et al. 2016). Indels were inspected to flag CDS regions where the

238   two species did not present the same reading frame. Bases in these regions were replaced with

239   Ns, and were therefore trimmed by the Gblocks program (Castresana 2000; Talavera and

240   Castresana 2007), alongside regions of unreliable alignment. We then estimated pairwise

241   synonymous (*dS*) and nonsynonymous (*dN*) substitution rates for each gene, using the codon-

       10

242   based method of Li (1993), as implemented in the R package seqinR (Charif and Lobry

243   2007). Only the 9,696 genes (out of 9,924) with an alignment length of >90 nucleotides, *dN*

244   <0.3 and *dS* <2 were kept. We also truncated *dN/dS* ratios to a maximal value of 2.5.

245

Estimates of selection intensity based on intraspecific polymorphism

247   Polymorphism data for *A. pisum* was obtained from 60 genotypes originating from

248   three *Medicago sativa* fields located in France and Switzerland (Jaquiéry et al. 2012b). These

249   fields can be considered to harbour a single large population of pea aphids (Jaquiéry et al.

250   2012b, Peccoud et al. 2009a). DNA was individually extracted from four asexual females of

251   each clone using the method described above. Because the approach described below does not

252   require reconstructing allele sequences or individual genotypes, sequencing the pooled

253   individuals (Gautier et al. 2013) was used to save costs. After RNAse treatment on each

254   sample and DNA dosage with Pherastar, DNA samples were pooled to attain equimolar

255   proportions. Paired-end libraries were then sequenced on two lanes of Illumina HiSeq 2000

256   using the Illumina Sequencing Kit v3 (producing 100-bp reads) by Beckman Coulter

257   Genomics (Danvers MA, USA). This yielded ~85X of sequencing coverage, hence an

258   expectation of 0.71X per individual chromosome. Reads were mapped on Acyr 2.0 and

259   symbiont genome sequences as described previously. The two alignment (BAM) files (one

260   per sequencing lane) were filtered from PCR duplicates using SAMtools rmdup (Li et al.

261   2009) and reads realigned near indels using the Genome Analysis Toolkit (McKenna et al.

262   2010). The raw sequence data have been deposited in the SRA division of Genbank (project

263   accession: PRJNA385905).

11

264   The two BAM files were merged and converted as pileup format using SAMtools

265   (options -B -Q 0 –R) (Li et al. 2009). A modified estimator of Tajima's D (Tajima

266   1989) which takes into account sequencing errors (Achaz 2008) was then calculated from

267   this mpileup with Popoolation 1.2.2 (Kofler et al. 2011), after subsampling at a uniform

268   coverage (subsample-pileup.pl, options : --target-coverage 30 --max-coverage 120 --method

269   withoutreplace). Computations were performed for each gene including introns (Variance-at-

270   position.pl --pool-size 120). Tajima's D allows evaluating the type of selection at work, since

271   selective sweeps and/or purifying selection tend to decrease it, and balancing selection tends

272   to increase it.

273   The McDonald and Kreitman (1991) approach, which compares fixed mutations to

274   polymorphic mutations in CDS, was adopted to further evaluate selection pressures on these

275   different categories of genes, using the DoS estimate (Direction of Selection, Stoletzki and

276   Eyre-Walker 2011). Positive, null and negative values of DoS respectively suggest adaptive

277   evolution, neutral evolution, and purifying selection. Fixed mutations between species were

278   counted from alignments we previously generated for *A. svalbardicum* and *A. pisum* CDSs.

279   We restricted the analysis to regions of reliable alignments, as given by the Gblocks txts

280   outputs. In these regions, we called SNPs on the BAM files with LoFreq (Wilm et al. 2012),

281   which offers a good compromise between speed, sensitivity and accuracy in pools of multiple

282   individuals (Huang et al. 2015). We used SAMtools mpileup (Li et al. 2009) to assess depth

283   of coverage at all positions in these regions, polymorphic or not. We instructed mpileup to

284   discard reads with mapping quality 0. The following was done in R. We discarded all

285   positions covered by less than three reads (both BAM files combined). At each SNP, the

286   number of polymorphic mutations was the number of different bases (alleles) found in the pea

287   aphid population minus one. A fixed difference was counted if no base was shared at a

12

288   position between the pea aphid population and *A. svalbardicum*. The number of polymorphic

289   non-synonymous mutations per codon was taken as the number of amino acids found in the

290   pea aphid populations for that codon minus one. To count the number of fixed non-

291   synonymous differences per codon, we considered that a codon might differ between the two

292   species by up to three mutations. Any of these may involve a change in protein sequence that

293   we cannot ascertain without knowledge on the order of appearance of the mutations. We

294   adopted parsimony and considered the minimum number of mutations required between the

295   two codons. If several codons were present in the pea aphid population (due to a SNP), we

296   considered the minimum number of coding changes that any pair of codons between the pea

297   aphid and *A. svalbardicum* involves. For all these counts, we discarded rare codons showing

298   more than one SNP, because the actual codons (and amino acids) present in the pea aphid

299   population cannot be determined without phasing. We counted the following for each gene:

300   the number of polymorphic non-synonymous changes (Pn), the number of all polymorphic

301   changes minus Pn (which is the number of polymorphic synonymous mutations, noted Ps),

302   the number of fixed non-synonymous differences (Dn), the number of all fixed differences

303   minus Dn (which is the number of fixed synonymous changes, noted Ds). DoS was then

304   calculated as $Pn/(Pn+Ps) – Dn(Dn+Ds)$ by concatenating all genes from a given class of

305   expression to avoid divisions by zero, due to genes without polymorphism or divergence, as

306   done in Burgarella et al. (2015). We considered only genes whose average depth of coverage,

307   as given by mpileup, was between 20 and 150 (expected coverage was ~85) to avoid

308   including genes presenting multiple collapsed copies that could artificially inflate

309   polymorphism.

310

13

311 **Statistical analyses**

312 Differences in expression levels between X-linked and autosomal genes in the different

313 morphs, as well as differences in *dN/dS*, *dN* and *dS* between X-linked and autosomal genes

314 were tested with Mann-Whitney U tests. The latter analysis was done on all genes, and on

315 genes grouped based on average expression over the three morphs. To evaluate evolutionary

316 forces responsible for faster-X evolution, we then compared *dN/dS*, Tajima's D and DoS

317 between X-linked and autosomal genes for classes of genes with different expression patterns

318 (unbiased, male-biased, sexual female-biased and asexual female-biased genes). For biased

319 genes, we considered different fold changes in expression (2- to 5-fold, and > 5-fold).

320 Statistical significance was evaluated with Mann-Whitney U tests.

321

Results

322

323 **Gene assignment to the X and autosomes**

324 Based on the ratio of depth of coverage by reads from males over reads from females, 64% of

325 the nucleotides assembled in the pea aphid reference genome (Acyr 2.0) were assigned to

326 autosomes, 31% to the X chromosome while only 5% could not be assigned (supplementary

327 figure S1). Genotypes of males at SNPs that were heterozygous in the female generally

328 confirmed the assignment from coverage depth data (supplementary figure S2), though

329 confirmation was not possible in regions lacking such SNPs. These estimates roughly

330 correspond to the expected size of the X chromosome in the pea aphid, which represents

331 ~30% of the chromosome content based on karyotypes (Mandrioli and Borsatti 2007). This

332 assignment revealed a very high rate of misassembly in Acyr 2.0: 56% of scaffolds of size

333 ≥150 kb (which represent 80% of the assembly length) comprised blocks assigned to different

14

334   chromosome types (supplementary figure S3 and table S1). Based on assigned scaffold

335   blocks, 19,263 predicted genes were located on autosomes, 13,726 on the X chromosome

336   while 4,001 genes could not be unambiguously assigned. The X chromosome contains a

337   higher fraction of predicted genes than expected from its relative size (42%, test of

338   proportion, $p<10^{-15}$).

339

340   **Gene evolutionary rates**

341   By comparing pea aphid gene sequences to transcripts sequenced from a related species (*A.*

342   *svalbardicum*), we assessed the rates of substitution. Results confirm faster-X evolution, as X-

343   linked genes present almost a twice higher non-synonymous substitution rate, dN (mean $dN_X=$

344   0.034; $dN_A=$ 0.019, Mann-Whitney U = 4839860, $p <10^{-15}$, $n = 9096$) and only slightly higher

345   synonymous substitution rate, dS (mean $dS_X=0.101$; $dS_A=$ 0.085, Mann-Whitney U =

346   6190137, $p <10^{-6}$). As a result, the evolution of X-linked genes involves more protein-

347   sequence changes (in proportion) than the evolution of autosomal genes (mean $dN_X/dS_X =$

348   0.390; $dN_A/dS_A =$ 0.237; Mann-Whitney U = 5026170, $p <10^{-15}$, figure 2A and

349   supplementary table S2).

350

351   **Causes of Faster-X evolution**

352   High ratios of non-synonymous to synonymous rates, dN/dS, can result from a decreased

353   influence of selection - and thus an increased influence of drift - on amino acid substitutions

354   (i.e., relaxed negative selection) and/or from more efficient selection of adaptive changes in

355   the protein sequence (i.e., increased positive selection). To balance these two hypotheses, we

356   used the expression levels of genes as proxies for their possible impact on fitness. Expression

15

357   levels of X-linked genes are lower than those of autosomal genes in all three main aphid

358   morphs: males, sexual females and parthenogenetic females (figure 2B). Expression level

359   averaged over the 13,726 X-linked genes ranges from ~52 reads/kb in sexual and asexual

360   females to 155 reads/kb in males, while expression level averaged over the 19,263 autosomal

361   genes varies from 575 (in asexual females) to 648 reads/kb in males. Genes that are not

362   expressed or expressed at a low level may have a reduced effect on the phenotype and may

363   therefore accumulate non-synonymous substitutions faster (reduced purifying selection). We

364   indeed observe that, for both X and autosomes, *dN/dS* ratios decrease with increasing

365   expression levels (averaged over the three morphs), and that the contrast between X and

366   autosomes tends to decline for highly expressed genes (figure 2A, supplementary table S2).

367   Slight contrast in *dN* ratios is still maintained for highly expressed genes, though

368   (supplementary table S2). Therefore, low expression levels of X-linked genes may not

369   entirely account for faster X-evolution in aphids.

370   Selection may also be relaxed in genes that are predominantly expressed in rare

371   morphs (males and sexual females), which constitute a minor fraction of the annual life cycle

372   of aphids dominated by asexual females. Relaxed selection on mutations affecting male-

373   biased genes (Purandare et al. 2014, Brisson and Nuzhdin 2008), combined with the tendency

374   of such genes to locate on the X (Jaquiéry et al. 2013; Pal and Vicoso 2015) could contribute

375   to faster-X evolution. However, the influence of X-linkage could not properly be evaluated in

376   Purandare et al. (2014) because misassembled scaffolds, rather than individual genes, were

377   assigned to chromosomes. Our new dataset of X-linked and autosomal genes unambiguously

378   confirms that the X is largely enriched in genes over-expressed in males, and to a smaller

379   extent in those over-expressed in sexual females (table 1). Like Purandare et al. (2014), we

380   observe higher *dN/dS* ratios in genes over-expressed in the rarer morphs (i.e. males and sexual

16

381   females, figure 3B-C), but not in genes over-expressed in the common morph

382   (parthenogenetic females, figure 3D) when considering X-linked and autosomal genes

383   together. Tajima's D also tends to increase in genes over-expressed in the sexual morphs

384   compared to unbiased genes (significantly so for all male-biased and for 2- to 5-fold female-

385   biased genes, figure 4B), but not in genes over-expressed in the common morph (where D is

386   significantly lower compared to unbiased genes, figure 4D), a pattern compatible with relaxed

387   selection on genes expressed mainly in the rare morphs. However, the DoS did not differ

388   significantly between these categories of genes, except for strongly female-biased genes

389   (supplementary figure S4).

390   When analyses are done by chromosome type, dN/dS ratios of X-linked genes are

391   significantly higher than those of autosomal genes for both sexual female and male-biased

392   genes (figure 3E-3H), but not for asexual females. Contrastingly, Tajima's D differs between

393   chromosome types only for strongly male-biased genes (being lower for X-linked genes,

394   suggesting more positive selection, figure 4F) and for unbiased genes (being larger for X-

395   linked genes, possibly revealing more balanced selection, figure 4E). As a result, the

396   enrichment of the X with genes expressed in the rare male morph (which have been

397   hypothesized to evolve under relaxed selection) may not strongly contribute to faster-X

398   evolution. No signal was detected between chromosome types based on the DoS index

399   (supplementary figure S4).

400   Alternatively, the faster evolution of X-linked male-biased genes compared to

401   autosomal male-biased genes could result from the fact that the former are present in

402   hemizygous state in males. Non-synonymous mutations on the X are thus more exposed to

403   selection in males, since they are not masked by other potentially dominant alleles, such that

404   adaptive mutations on the X should more rapidly and more likely reach fixation than adaptive

17

405   mutations on autosomes. This hypothesis predicts that the contrast between substitution rates

406   of X-linked genes and autosomal genes will be highest for male-biased genes, and lowest for

407   sexual- and asexual female-biased genes, because in these morphs, the X is always diploid

408   and adaptive mutations can be recessive. We indeed observe these patterns (figure 3F), and

409   the significantly lower Tajima's D for X-linked male-biased genes compared to autosomal

410   genes provides some support to this hypothesis (figure 4F).

411

Discussion

413   Here we performed a genome-wide identification of X-linked genes enabling to locate a large

414   number (13,726) and proportion (42%) of predicted genes on the X chromosome. We

415   demonstrated that these genes tend to evolve faster than autosomal genes on average,

416   confirming earlier results based on a much smaller set of coding genes (Jaquiéry et al. 2012a).

417   Faster-X evolution mainly results from low expression of a large fraction of X-linked genes,

418   which would less affect the phenotype and may accumulate non-synonymous mutations. Our

419   analyses also suggest that higher exposure of recessive X-linked alleles to selection in

420   hemizygous males might also partially contribute to faster-X evolution via more efficient

421   positive selection.

422   We demonstrate clear faster-X evolution in the pea aphid based on a large set of X-

423   linked and autosomal genes. The non-synonymous to synonymous substitution ratio (dN/dS)

424   for X-linked genes is 1.69 times greater than for autosomal genes. This clearly places aphids

425   among species showing strong contrast between evolution of X-linked and autosomal genes.

426   Excluding sex-specific genes (e.g. Torgerson 2003; Kousathanas et al. 2014) and genes that

427   escape post-meiotic silencing (Sin et al. 2012) for which X-linked sequences evolve much

428   faster than autosomal ones, the dN/dS for X-linked genes is between ~0.9 and ~1.8 times that

18

429   of autosomes in most species studied (i.e., *Drosophila*, mammals, birds and moths, review in

430   Meisel and Connallon 2013; see also Sackton et al. 2014).

431   Remarkably, the pea aphid presents the same effective population size for the X and

432   autosomes (Jaquiéry et al. 2012a), such that hemizygosity in males should be the only

433   differentiating factor affecting the evolution of genes located on different chromosome types.

434   However, our analyses revealed another key difference between X-linked and autosomal

435   genes, in that the former are on average 4 to 10 times less expressed than the latter (figure

436   2B), and expectedly show higher rates of substitution. Such negative correlations between

437   substitution rates and expression level have already been observed in several species (e.g.,

438   Drummond et al. 2005, Nguyen et al. 2015, Zhang and Yang 2015). Therefore, enrichment of

439   the X with lowly expressed genes explains faster-X evolution in aphids to a large extent.

440   Gene expression differs between chromosome types in another dimension, as the X is

441   enriched in genes that are mostly expressed in the rare morphs (i.e., males and sexual females)

442   (Jaquiéry et al. 2013, Pal and Vicoso 2015). Such genes should also evolve under more

443   relaxed constraints as they are exposed to the selective environment only during a short period

444   of the aphid life cycle (Brisson and Nuzhdin 2008, Purandare et al. 2014). However, our

445   analyses revealed that X-linked male-biased genes show lower Tajima's D than autosomal

446   male-biased genes (figure 4F-4G). As a result, enrichment of the X chromosome with genes

447   expressed in the rare male morph is unlikely to be an important contributor to faster-X

448   evolution in the pea aphid.

449   This leaves hemizygosity of the X chromosome in males, which exposes all X-linked

450   alleles expressed in males to the selective environment, as another contributing cause. This

451   hypothesis is supported by the contrast in *dN/dS* between X-linked genes and autosomal

452   genes, which is larger for male-biased genes than for sexual and asexual females-biased genes

19

453   (figure 3F-3H), and by the lower Tajima's D on male-biased X-linked genes than for those on

454   autosomes (figure 4F-4G), but not by the DoS, which shows no sign of selection.

455   Mean expression levels of X-linked genes measured in the whole body are strikingly

456   lower than those of autosomal genes, in all morphs studied. The difference we found is more

457   pronounced than in previous observations (Jaquiéry et al. 2013, Pal and Vicoso 2015)

458   probably due to more reliable gene assignments to chromosomes (we found that 10% of the

459   3,712 genes used in Jaquiéry et al. 2013 had been misassigned because of scaffold

460   misassembly). Lower expression of X-linked genes compared to autosomal genes is observed

461   in mammals (Nguyen et al. 2015), but not in *Drosophila* (Zhang and Presgraves 2016). To our

462   knowledge, no other taxon displays such a strong contrast between the X and autosomes gene

463   expression levels as does the pea aphid. This raises the question as to why genes on the X are

464   so little expressed in this species. A theoretical model (Jaquiéry et al. 2013) predicts that the

465   X chromosome is more easily invaded by sexually antagonistic alleles beneficial to males and

466   deleterious to females than autosomes. This may have favored a global decrease in gene

467   expression of this chromosome in the common morph (the asexual females) for which it could

468   be harmful. Pseudogenisation on the X chromosome would have ensued if genetic variation in

469   lowly expressed genes has little effect on fitness. Yet, this chromosome carries one third of

470   the genome and contains a higher fraction of genes than predicted by its relative size. Insights

471   into the role of sexual antagonism on the peculiar expression patterns observed here could be

472   gained by studying gene expression of this chromosome in different male and female tissues.

473   Particularly, transcriptomes of tissues subject to contrasted sex-specific selection pressures

474   (e.g. Parisi 2003; Khil et al. 2004; Yang et al. 2006; Huylmans and Parsch 2015) could help

475   examine this hypothesis.

20

476   Assignments of scaffold blocks to chromosomes revealed widespread errors in the pea

477   aphid genome assembly (Acyr 2.0). More than half of scaffolds larger than 150kb are clear

478   chimeras of X and autosomes. This is a minimal estimate for the rate of misassembly, since

479   our method only detects breakpoints between X and autosomes. Given that the X represents

480   30% of the pea aphid genome, assembly errors involving two X-linked genomic regions, and

481   those involving two autosomal regions should respectively represent 0.3^2 and 0.7^2 (58% in

482   total) of all assembly errors, a significant number of events that we could not detect.

483   Moreover, assembly errors involving fragments of less than 10 kb could not be detected given

484   the resolution of our analyses (see methods). Other lines of evidence argue for frequent errors

485   in the current pea aphid genome assembly. First, a genetic map of 305 microsatellite markers

486   (Jaquiéry et al. 2014) validates our assignment to chromosomes, and also demonstrates high

487   rates of scaffold misassembly (see supplementary file S1). Second, X-A breakpoints co-

488   localize with sharp variations in genetic differentiation (F_{ST}) between pea aphid populations

489   from different host plants, as determined by high-density genome scans (Nouhaud et al.

490   submitted). Third, synteny between Acyr 2.0 scaffolds and scaffolds from two *Myzus persicae*

491   genomes (Mathers et al. 2017, available at www.aphidbase.com) stops at X-A breakpoints

492   (data not shown). Fourth, entire scaffolds assigned by Bickel et al. (2013) to the X

493   chromosome account for 11% of Acyr 2.0, much lower than the ~30% of the genome that this

494   chromosome represents, based on karyotypes (Mandrioli and Borsatti 2007). Bickel et al.

495   (2013) explained this discrepancy by assembly errors combined with methodological bias

496   towards assigning misassembled scaffolds to autosomes (a hypothesis we confirm by the

497   comparison of scaffold assignments to chromosomes by Bickel et al. 2013 and the present

498   study, see supplementary figure S5). By contrast, our new assignments to chromosomes

499   provide estimates corresponding to the expected relative size of the X. Consequently, we

21

500   confidently conclude that the genome of the pea aphid presents considerable assembly

501   problems, to a degree that goes far beyond what current assembly pipelines typically yield

502   (e.g. Salzberg et al. 2004; Muggli et al. 2015). While the cause of these errors remains

503   undetermined, they have important drawbacks for genomic studies on a species that is

504   currently considered the model aphid, in particular those relying on the physical organization

505   of the genome, ranging from high-resolution genome scans to studies of chromatin

506   conformation, genomic rearrangements, etc. We therefore urge for a reassembly of the pea

507   aphid genome based on additional data. The results presented here should however not be

508   affected by misassembly as we were able to assign unambiguously almost 90% of the 36,990

509   predicted gene of the pea aphid. This means that at fine-grain level, that of a few kilobases to

510   few tens of kilobases (the range within which most genes are contained), the chimerism

511   problem is not a major issue.

512   In conclusion, we have shown here that faster-X evolution of proteins in the pea aphid

513   is principally explained by relaxed selection on lowly expressed genes, a class of genes that is

514   more frequent on the X chromosome than on autosomes. Exposure of X-linked recessive

515   alleles to selection in hemizygous males could play an additional, yet marginal, role in faster-

516   X evolution. Importantly, the pea aphid complex, which includes races and cryptic species at

517   different stages of divergence (Peccoud et al. 2009a, 2009b, 2014), offers excellent

518   opportunities to investigate mechanisms driving genome evolution through comparative

519   genomics. In particular, characterizing gene expression in morphs of closely and distantly

520   related species could help disentangling the role of drift and selection in the low expression of

521   the X and its masculinization.

22

References

Achaz G. 2008. Testing for neutrality in samples with sequencing errors. Genetics, 179:1409-1424.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Avila V, Campos JL, Charlesworth B. 2015. The effects of sex-biased gene expression and X-linkage on rates of adaptive protein sequence evolution in *Drosophila*. Biol. Lett. 11:20150117.

Avila V et al. 2014. Faster-X effects in two *Drosophila* lineages. Genome Biol. Evol. 6:2968–2982.

Bachtrog D et al. 2011. Are all sex chromosomes created equal? Trends Genet. 27:350–357.

Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008. Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. Mol. Biol. Evol. 25:1639–1650.

Bickel RD, Dunham JP, Brisson JA. 2013. Widespread selection across coding and noncoding DNA in the pea aphid genome. G3 3:993–1001.

Brisson JA, Nuzhdin SV. 2008. Rarity of males in pea aphids results in mutational decay. Science 319:58–58.

Burgarella C et al. 2015. Molecular evolution of freshwater snails with contrasting mating systems. Mol. Biol. Evol. 32:2403–2416.

Caballero A. 1994. Developments in the prediction of effective population size. Heredity 73:657–679.

Caballero A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139:1007–1011.

Caillaud MC, Boutin M, Braendle C, Simon J-C. 2002. A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). Heredity 89:346–352.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural approaches to sequence evolution: Molecules, networks, populations. Biological and Medical Physics, Biomedical Engineering. New York: Springer Verlag. p. 207–232.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. 130:113–146.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118–128.

Coyne JA, Orr, H. A. 2004. Speciation. Sinauer Associates. Sunderland, MA

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. 102:14338–14343.

DePristo MA et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Frantz A, Plantegenest M. & Simon JC. 2010. Host races of the pea aphid *Acyrthosiphon pisum* differ in male wing phenotypes. Bull. Entomol. Res. 100:59-66.

Gautier M et al. 2013. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Mol. Ecol. 22:3766–3779.

23

565 Gouin A et al. 2015. Whole-genome re-sequencing of non-model organisms: lessons from unmapped
566   reads. Heredity 114:494–501.

567 Gouzy J, Carrere S, Schiex T. 2009. FrameDP: sensitive peptide detection on noisy matured
568   sequences. Bioinformatics 25:670–671.

569 Grabherr MG et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference
570   genome. Nat. Biotechnol. 29:644–652.

571 Huang HW, Mullikin JC, Hansen NF. 2015. Evaluation of variant detection software for pooled next-
572   generation sequence data. BMC Bioinformatics 16:1–9.

573 Huylmans AK, Parsch J. 2015. Variation in the X:Autosome distribution of male-biased genes among
574   *Drosophila melanogaster* tissues and its relationship with dosage compensation. Genome
575   Biol. Evol. 7:1960–1971.

576 Hvilsom C et al. 2012. Extensive X-linked adaptive evolution in central chimpanzees. Proc. Natl.
577   Acad. Sci. 109:2054–2059.

578 International Aphid Genome Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon*
579   *pisum*. Plos Biol. 8:e1000313.

580 Jaquiéry J et al. 2012a. Accelerated evolution of sex chromosomes in aphids, an X0 system. Mol. Biol.
581   Evol. 29:837–847.

582 Jaquiéry J et al. 2012b. Genome scans reveal candidate regions involved in the adaptation to host plant
583   in the pea aphid complex. Mol. Ecol. 21:5251–5264.

584 Jaquiéry J et al. 2013. Masculinization of the X chromosome in the pea aphid. Plos Genet.
585   9:e1003690.

586 Jaquiéry J et al. 2014. Genetic control of contagious asexuality in the pea aphid. PLoS Genet.
587   10:e1004838.

588 Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for
589   hybrid incompatibility. Ann. N. Y. Acad. Sci. 1256:E1–E22.

590 Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. 2004. The mouse X chromosome is
591   enriched for sex-biased genes not subject to selection by meiotic sex chromosome
592   inactivation. Nat. Genet. 36:642–646.

593 Kimura M. 1983. The neutral theory of molecular evolution. Univ. press. Cambridge

594 Kitano J et al. 2009. A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–
595   1083.

596 Kofler R et al. 2011. PoPoolation: a toolbox for population genetic analysis of next generation
597   sequencing data from pooled individuals. *PloS one*, 6:e15925.

598 Kousathanas A, Halligan DL, Keightley PD. 2014. Faster-X adaptive protein evolution in house mice.
599   Genetics 196:1131–1143.

600 Langley CH et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*.
601   Genetics 192:533–598.

602 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–
603   359.

604 Li H et al. 2009. The sequence alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

605 Li WH. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J.
606   Mol. Evol. 36:96–99.

607 Mackay TFC et al. 2012. The *Drosophila melanogaster* genetic reference panel. Nature 482:173–178.

608 Mandrioli M, Borsatti F. 2007. Analysis of heterochromatic epigenetic markers in the holocentric
609   chromosomes of the aphid *Acyrthosiphon pisum*. Chromosome Res. 15:1015–1022.

24

610 Mank JE, Nam K, Ellegren H. 2010. Faster-Z evolution is predominantly due to genetic drift. Mol.
611   Biol. Evol. 27:661–670.

612 Mank JE, Vicoso B, Berlin S, Charlesworth B. 2010. Effective population size and the faster-X effet:
613   empirical results and their interpretation. Evolution 64:663–674.

614 Mathers TC et al. 2017 Rapid transcriptional plasticity of duplicated gene clusters enables a clonally
615   reproducing aphid to colonise diverse plant species. Genome Biology, 18:27.

616 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature
617   351:652–654.

618 McKenna A et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-
619   generation DNA sequencing data. Genome Res. 20:1297–1303.

620 Meiklejohn CD, Tao Y. 2010. Genetic conflict and sex chromosome evolution. Trends Ecol. Evol.
621   25:215–223.

622 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends Genet.
623   29:537–544.

624 Muggli MD, Puglisi SJ, Ronen R, Boucher C. 2015. Misassembly detection using paired-end sequence
625   reads and optical mapping data. Bioinformatics 31:i80–i88.

626 Nguyen L-P, Galtier N, Nabholz B. 2015. Gene expression, chromosome heterogeneity and the fast-X
627   effect in mammals. Biol. Lett. 11:20150010.

628 Pages H, Aboyoun P, Gentleman R, DebRoy S. 2016. Biostrings: String objects representing
629   biological sequences, and matching algorithms. R package version 2.40.2.

630 Pal A, Vicoso B. 2015. The X chromosome of hemipteran insects: conservation, dosage compensation
631   and sex-biased expression. Genome Biol. Evol. 7:3259–3268.

632 Parisi M. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression.
633   Science 299:697–700.

634 Peccoud J et al. 2008. Host range expansion of an introduced insect pest through multiple
635   colonizations of specialized clones. Mol. Ecol. 17:4608–4618.

636 Peccoud J, Ollivier A, Plantegenest M, Simon J-C. 2009a. A continuum of genetic divergence from
637   sympatric host races to species in the pea aphid complex. Proc. Natl. Acad. Sci. 106:7495–
638   7500.

639 Peccoud J, Simon J-C, McLaughlin HJ, Moran NA. 2009b. Post-Pleistocene radiation of the pea aphid
640   complex revealed by rapidly evolving endosymbionts. Proc. Natl. Acad. Sci. U. S. A.
641   106:16315–16320.

642 Peccoud J et al. 2014. Widespread host-dependant hybrid unfitness in the pea aphid species complex:
643   postzygotic isolation in pea aphids. Evolution 68:2983–2995.

644 Purandare SR, Bickel RD, Jaquiéry J, Rispe C, Brisson JA. 2014. Accelerated evolution of morph-
645   biased genes in pea aphids. Mol. Biol. Evol. 31:2073–2083.

646 R Development Core Team 2015: A language and environment for statistical computing. Vienna,
647   Austria Available from: URL http://www.R-project.org.

648 Rice WR. 1984. Sex Chromosomes and the Evolution of Sexual Dimorphism. Evolution 38:735–742.

649 Rispe C et al. 2016. De novo transcriptome assembly of the grapevine phylloxera allows identification
650   of genes differentially expressed between leaf- and root-feeding forms. BMC Genomics 17:1–
651   15.

652 Sackton TB et al. 2014. Positive selection drives faster-Z evolution in silkmoths. Evolution 68:2331–
653   2342.

25

654 Saether SA et al. 2007. Sex chromosome-linked species recognition and evolution of reproductive
655   isolation in flycatchers. Science 318:95–97.

656 Salzberg SL, Church D, DiCuccio M, Yaschenko E, Ostell J. 2004. The genome assembly archive: a
657   new public resource. PLoS Biol 2:e285.

658 Simon J-C et al. 2011. Facultative symbiont infections affect aphid reproduction. PLoS ONE
659   6:e21831.

660 Sin H-S, Ichijima Y, Koh E, Namiki M, Namekawa SH. 2012. Human postmeiotic sex chromatin and
661   its impact on sex chromosome evolution. Genome Res. 22:827–836.

662 Soh YQS et al. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and
663   amplification on both sex chromosomes. Cell 159:800–813.

664 Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. Mol. Biol. Evol. 28:63–70.

665 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
666   Genetics 123:485–595.

667 Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and
668   ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

669 Torgerson DG. 2003. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. Mol.
670   Biol. Evol. 20:1705–1709.

671 Venables WN, Ripley BD. 2002. Modern applied statistics with S. 4th ed. New York: Springer

672 Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model.
673   Evolution 63:2413–2426.

674 Wilm A et al. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering
675   cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res.
676   40:11189-11201.

677 Wilson ACC, Sunnucks P, Hales DF. 1997. Random loss of X chromosome at male determination in
678   an aphid, *Sitobion* near *fragariae*, detected using an X-linked polymorphic microsatellite
679   marker. Genet. Res. 69:233–236.

680 Wright S. 1931. Evolution in mendelian populations. Genetics 16:97–159.

681 Yang X et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice.
682   Genome Res. 16:995–1004.

683 Zhang Z, Presgraves DC. 2016. *Drosophila* X-linked genes have lower translation rates than
684   autosomal genes. Mol. Biol. Evol. 33:413–428

685 Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. Nat. Rev. Genet.
686   16:409–420.

687

Acknowledgements

26

695   PBLAA-122658 & PA00P3-139720 to J.J.). We would like to thank the GenOuest platform

696   for the use of its cluster to perform the computational analyses.

697

Figure legends

699   **Fig. 1.** Life-cycle of the pea aphid and ploidy levels for autosomes (A) and the sex-

700   chromosome (X) (adapted from Jaquiéry et al. 2013).

701

702   **Fig. 2.** Evolutionary rates for autosomal and X-linked genes and gene expression in males,

703   sexual and asexual females. A) Evolutionary rates (*dN/dS*) are shown for all genes (expressed

704   or not in *A. pisum*) and for genes expressed at different levels (when averaged over male,

705   sexual and asexual females): lowly expressed genes (i.e. covered by < 100 reads/kb of exon

706   model); moderately expressed (from 100 to 1000 reads/kb), highly expressed genes (>1000

707   reads/kb). The number of genes per category is shown above each boxplot. B) Expression

708   level for X-linked (*n* = 13613) and autosomal genes (*n* = 18812) in males, sexual females and

709   asexual females. It should be noted that males carry only one X chromosome per cell and

710   females carry two. Significance was tested with Mann-Whitney U tests.

711

712   **Fig. 3**. Substitution rates of genes (*dN/dS*, measured between *A.pisum* and *A. svalbardicum*)

713   according to the ratios of expression levels between morphs. Panels A-D consider all genes

714   together (X-linked and autosomal), and panels E-H consider X-linked (dark grey) and

715   autosomal (light grey) genes separately. The number of genes in each class is shown above

716   each boxplot. Only genes supported by at least 100 reads/kb of exon model were retained. ub:

717   unbiased genes (padj > 0.1 for morph effect on expression), 2-5: levels of gene expression are

718   two to five times higher in the specified morph (padj <0.05 for morph effect), >5: levels of

719   gene expression are at least five times higher in the specified morph (padj <0.05).

720   Significance of differences: ns: $p > 0.05$; *$p< 0.05$, **$p < 0.01$, ***$p < 0.001$ (Mann-Whitney

721   U tests). For panels B-D, differences correspond to comparisons with genes of the "unbiased"

722   category, while X and autosomes were compared in panels F-H.

27

723   **Fig. 4.** Tajima's D according to the ratios of gene expression levels between morphs. Panels

724   A-D consider all genes together (X-linked and autosomal) and panels E-H consider X-linked

725   (dark grey) and autosomal (light grey) genes separately. Terms are defined as in Figure 3.

726   Dashed lines show median values for unbiased genes. Significance of differences was tested

727   with Mann-Whitney U tests.

728

729

730

28

Tables

Table 1. Number of X-linked and autosomal genes and frequency of X-linkage for classes of genes with contrasted patterns of expression between morphs.

Category of genes	number of X-linked genes	number of autosomal genes	Frequency of X-linkage	p value [6]
All[1]	13726	19263	0.42	10^{-16}
Low expression[2] (less than 10 reads/kb)	10995	8136	0.57	10^{-16}
Expressed[3] (at least 10 reads/kb)	2771	11127	0.20	0.0001
Unbiased[4] (>10 reads/kb & padj>0.1)	697	3355	0.17	na
2-fold male-biased[5]	1546	2245	0.41	10^{-16}
5-fold male-biased[5]	962	948	0.50	10^{-16}
2-fold sexual female-biased[5]	448	1369	0.25	10^{-10}
5-fold sexual female-biased[5]	148	407	0.27	10^{-7}
2-fold asexual female-biased[5]	244	1023	0.19	0.10
2-fold asexual female-biased[5]	93	423	0.18	0.68

[1]all predicted genes that were assigned to the X or autosomes are included
[2]genes with on average <10 reads/kb of exon (average over the 3 morphs)
[3]genes with on average ≥10 reads/kb of exon (average over the 3 morphs)
[4]genes with on average ≥10 reads/kb of exon (average over the 3 morphs) and with an adjusted p-value ≥ 0.1 when tested for morph-biased expression.
[5]A gene was included in the morph-biased category (either male-, female- or asexual-biased) if the adjusted p-value for a morph effect was <0.05 and if it was at least x-fold (2 or 5) more expressed in one of the morph compared to the two other morphs.
[6]Deviation from expectation (given by the "unbiased" category) was evaluated with a test of proportion.

29

742 **Figures**
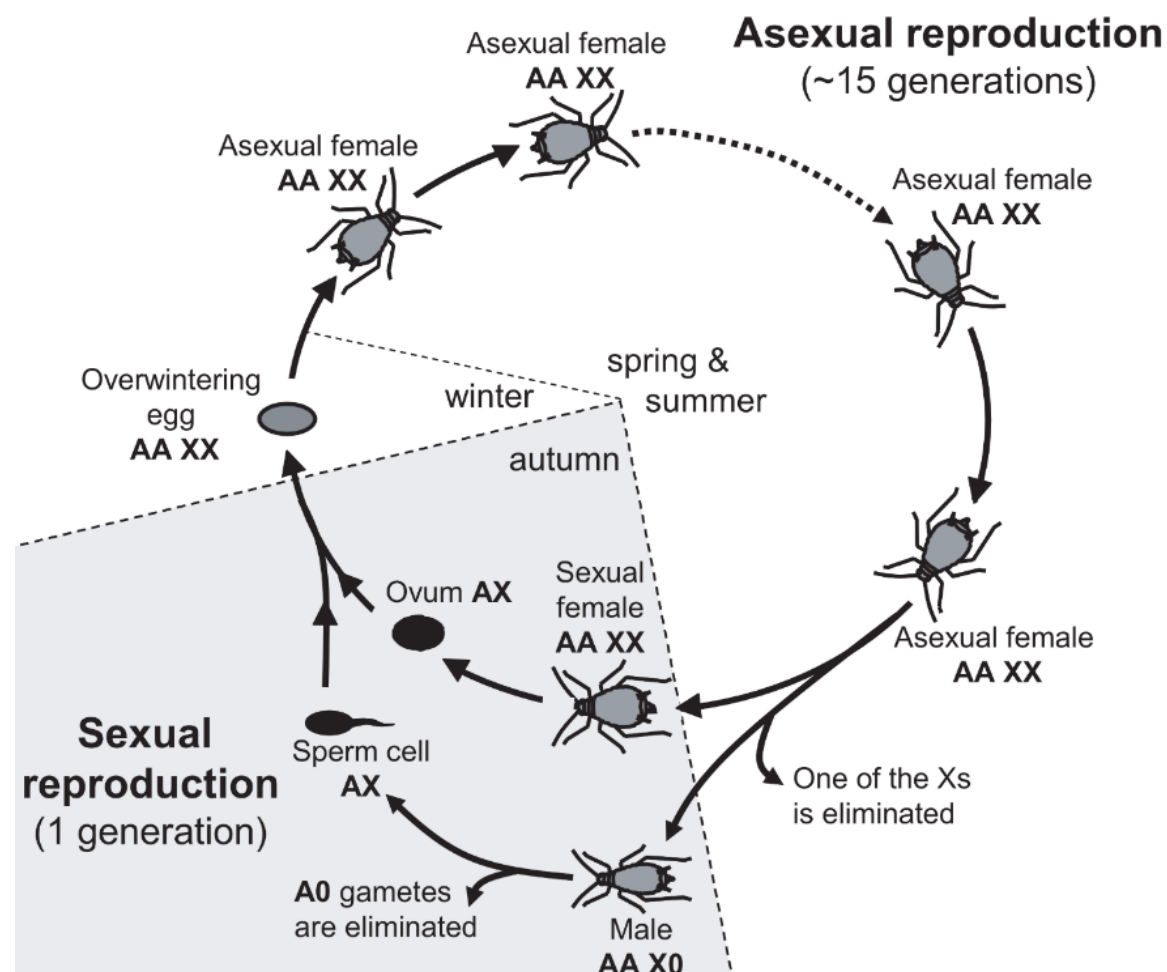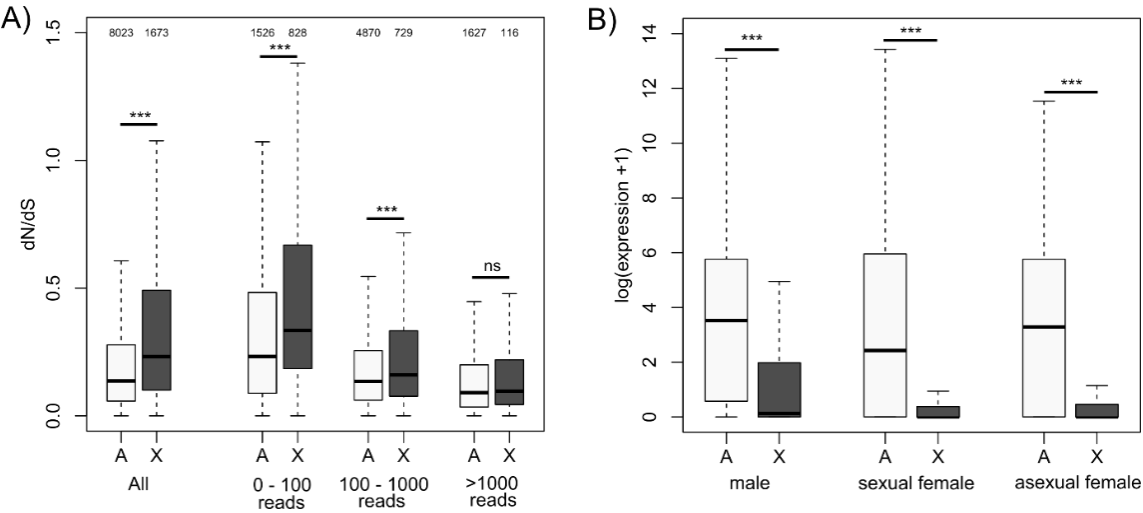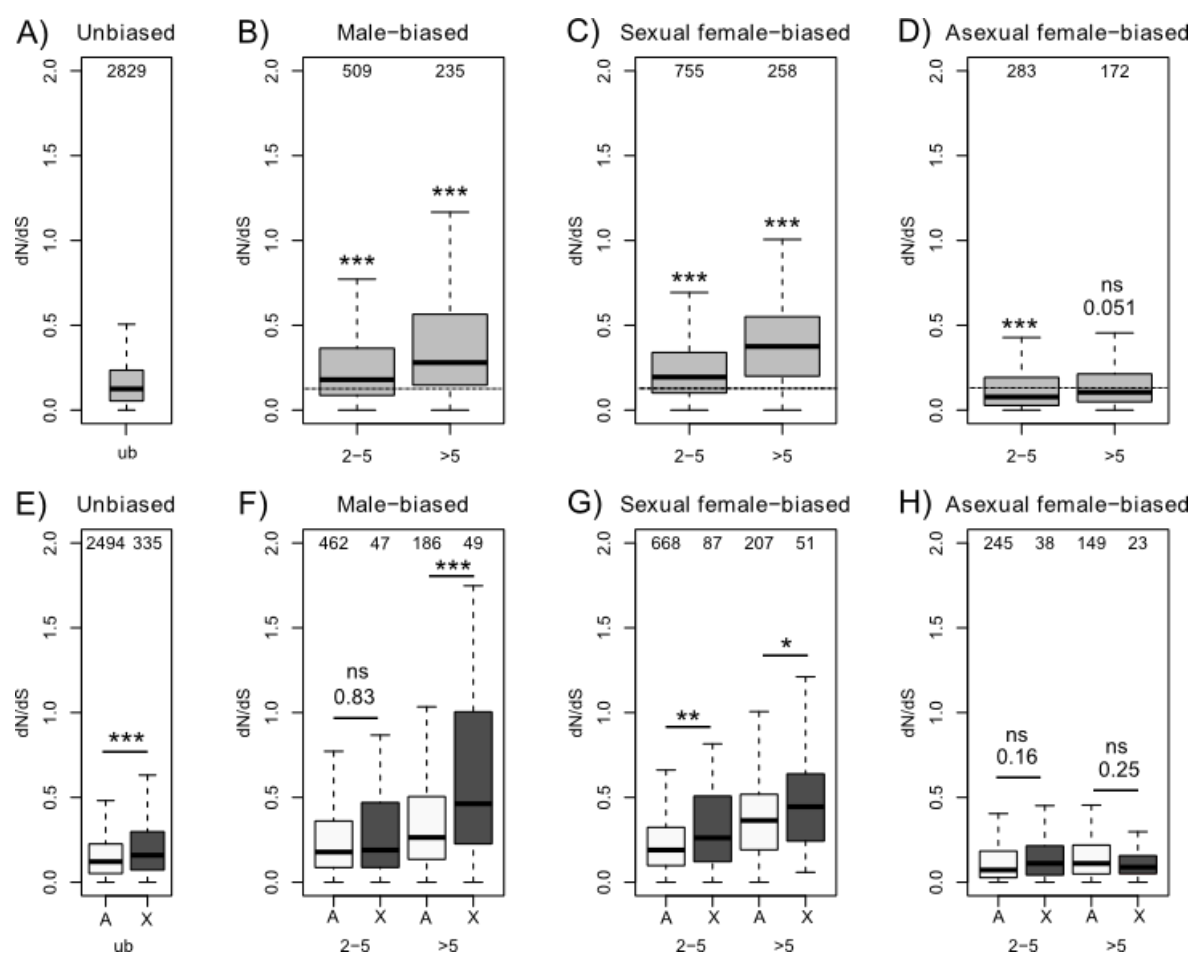
743 **Figure 1**

744

745

746

30

747 **Figure 2**

748

749

31

Figure 3

32

753 **Figure 4**

754

755

756

33