

SCIENTIFIC REPORTS



OPEN

No signal of deleterious mutation accumulation in conserved gene sequences of extant asexual hexapods

Alexander Brandt¹, Jens Bast², Stefan Scheu¹, Karen Meusemann^{3,4}, Alexander Donath⁴, Kai Schütte⁵, Ryuichiro Machida⁶ & Ken Kraaijeveld⁷

Loss of sex and recombination is generally assumed to impede the effectiveness of purifying selection and to result in the accumulation of slightly deleterious mutations. Empirical evidence for this has come from several studies investigating mutational load in a small number of individual genes. However, recent whole transcriptome based studies have yielded inconsistent results, hence questioning the validity of the assumption of mutational meltdown in asexual populations. Here, we study the effectiveness of purifying selection in eight asexual hexapod lineages and their sexual relatives, as present in the 1K Insect Transcriptome Evolution (1KITE) project, covering eight hexapod groups. We analyse the accumulation of slightly deleterious nonsynonymous and synonymous point mutations in 99 single copy orthologue protein-coding loci shared among the investigated taxa. While accumulation rates of nonsynonymous mutations differed between genes and hexapod groups, we found no effect of reproductive mode on the effectiveness of purifying selection acting at nonsynonymous and synonymous sites. Although the setup of this study does not fully rule out nondetection of subtle effects, our data does not support the established consensus of asexual lineages undergoing 'mutational meltdown'.

The ubiquitous prevalence of sex among eukaryotes is surprising given that sexual reproduction involves manifold evolutionary costs as compared to obligate asexuality¹⁻³. One prediction for the benefit of sex in the long-term is the increased effectiveness of purifying selection⁴. The rationale is that segregation, recombination and outcrossing enable the uncoupling of linked loci with different selection coefficients, such that selection can act on different loci independently⁵. This accelerates adaptation and the purging of slightly deleterious mutations and facilitates the restoration of least loaded genotypes that are continuously lost by drift⁵⁻⁸. Asexual lineages lack these benefits and are therefore predicted to succumb to 'mutational meltdown'⁹.

A multitude of studies have tested the prediction of impeded effectiveness of purifying selection (i.e. selective removal of deleterious mutations) in non-recombining genomic regions, such as mitochondria or (neo-) Y chromosomes as well as different lineages of asexual eukaryotes¹⁰⁻¹³. Their results have led to the established consensus that slightly deleterious mutations accumulate in the absence of sex. However, many of the studies that have investigated purifying selection in asexual species were based on only few individual genes and recent studies based on whole transcriptome comparisons between asexual and related sexual lineages did not find consistent support: while accumulation of slightly deleterious mutations was found in asexual *Timema* stick insects,

¹University of Göttingen, JF Blumenbach Institute of Zoology and Anthropology, Untere Karspüle 2, D-37073, Göttingen, Germany. ²University of Lausanne, Department of Ecology and Evolution, UNIL Sorge, Le Biophore, CH-1015, Lausanne, Switzerland. ³University of Freiburg, Biology I, Evolutionary Biology & Ecology, Hauptstraße 1, D-79104, Freiburg, Germany. ⁴Center for Molecular Biodiversity Research (ZMB), Zoological Research Museum Alexander König, Adenauerallee 160, D-53113, Bonn, Germany. ⁵University of Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Department of Biology, Institute of Zoology, Research Unit Animal Ecology and Conservation, Martin-Luther-King-Platz 3, D-20146, Hamburg, Germany. ⁶Sugadaira Research Station, Mountain Science Center, University of Tsukuba, 1278-294, Sugadaira Kogen, Ueda, Nagano, 386-2204, Japan. ⁷University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics, Science Park 904, 1090 GE, Amsterdam, The Netherlands. Correspondence and requests for materials should be addressed to A.B. (email: abrandt3@gwdg.de)

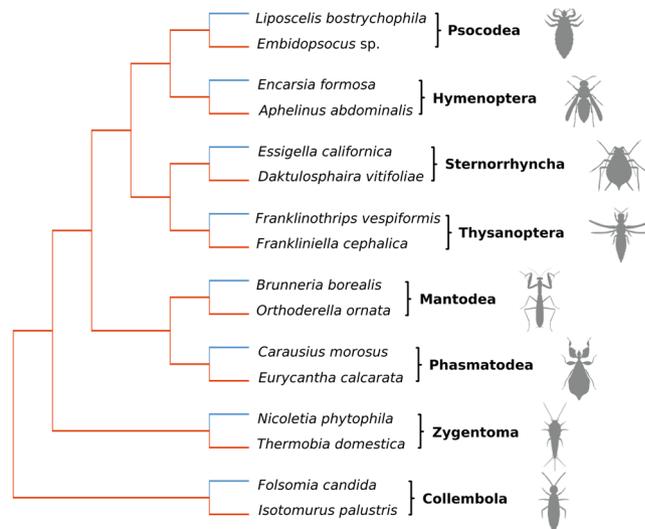


Figure 1. Cladogram of 16 hexapod species analysed in this study. The cladogram was manually built based on the phylogeny published by Misof *et al.* 2014 (see Methods)³³. The taxon sampling includes eight asexual lineages along with their closest sexual relatives as present in 1KITE, covering eight hexapod groups. Sexual and asexual lineages are depicted in red and blue, respectively. Silhouettes courtesy of Hans Pohl.

Oenothera evening primroses and *Boechera* rockcress, it was absent in *Lineus* ribbon worms as well as four aphid species and, opposite to predictions, reduced in asexual as compared to sexual oribatid mites^{14–19}. Moreover, all whole transcriptome based studies found excessive variation among genes, and hence doubts have been raised about the robustness of inferences drawn from single gene analyses²⁰. These conflicting results highlight the need for more studies of larger gene sets along with broader taxonomic sampling to infer whether or not accumulation of deleterious mutations is indeed a consequence of asexual reproduction.

Here, we study the effectiveness of purifying selection in obligately asexual and related sexual hexapod lineages covering eight hexapod groups using transcriptome data generated by the 1KITE project (1 K Insect Transcriptome Evolution, www.1kite.org). As parthenogenesis is a lineage-level trait, we use the term ‘lineage’, whenever to distinguish between the two reproductive modes (sexual and asexual) within a hexapod group, from here on. With the term ‘hexapod group’ we refer to one of the following analysed taxa: Collembola, Zygentoma, Phasmatoidea, Mantodea, Thysanoptera, Sternorrhyncha, Hymenoptera and Psocodea, respectively.

We compared the accumulation of slightly deleterious mutations in nuclear protein-coding genes between eight asexual lineages and their sexual relatives as available from 1KITE (see Fig. 1). To this end, we first inferred divergence at nonsynonymous sites normalised for background mutation rates (dN/dS), and the potential ‘deleteriousness’ of nonsynonymous substitutions. Second, we investigated selection on Codon Usage Bias (CDC), because selection also acts at synonymous sites²¹. We based all analyses on 99 single copy orthologues which we found to be under purifying selection in the analysed lineages. We found extensive variation in dN/dS and CDC among genes and between hexapod groups, but no overall difference between reproductive modes.

Methods

Species selection. We searched for obligately asexual lineages within the 1KITE species list based on primarily van der Kooij & Schwander 2014²², Vershinina & Kuznetsova 2016²³ and the Tree of Sex database²⁴. We found eight hexapod groups with obligately asexual lineages represented by the following species: *Folsomia candida* (Collembola), *Nicoletia phytophila* (Zygentoma), *Carausius morosus* (Phasmatoidea), *Brunneria borealis* (Mantodea), *Franklinothrips vespiformis* (Thysanoptera), *Essigella californica* (Sternorrhyncha), *Encarsia formosa* (Hymenoptera) and *Liposcelis bostrychophila* (Psocodea). In all analysed species, parthenogenesis is thelytokous. In *C. morosus* and *E. californica* it is obligately apomictic whereas in *F. candida*, *F. vespiformis* and *E. formosa* parthenogenesis is induced by bacteria of the genus *Wolbachia* and in *L. bostrychophila* by the genus *Rickettsia*^{25–31}. We found no information concerning the mode of parthenogenesis in *B. borealis* and *N. phytophila* in the literature. Further, we selected per hexapod group the closest sexual relatives to the asexual lineages present in 1KITE, namely *Isotomurus palustris* (Collembola), *Thermobia domestica* (Zygentoma), *Eurycantha calcarata* (Phasmatoidea), *Orthoderella ornata* (Mantodea), *Frankliniella cephalica* (Thysanoptera), *Daktulosphaira vitifoliae* (Sternorrhyncha), *Aphelinus abdominalis* (Hymenoptera) and *Embidopsocus* sp. (Psocodea). For information on first describers of species used, see Supplementary Table S1.

Data acquisition. We downloaded published and most current transcriptome assemblies from Genbank and TSA of the following species: *A. abdominalis*, *E. formosa*, *E. californica*, *F. candida*, *F. cephalica*, *L. bostrychophila*, *T. domestica* and *Xibalbanus cf. tulumensis*^{32–34}. For remaining species, we obtained unpublished transcriptome assemblies from the 1KITE consortium (RNA extraction, cDNA library generation, sequencing, assembly and contaminant removal was done as described in detail in Misof *et al.* 2014 and Peters *et al.* 2017)^{33,35}. Assemblies

were used as input for identification of orthologous sequences among all studied species. For information on References, BioSample IDs and Bioproject IDs of assembled transcriptomes, see Supplementary Table S1.

Orthologue detection and alignment processing. To infer orthologue protein-coding genes among the 16 hexapod species, we used the Orthologous Matrix (OMA) version 2.1.1 installed on the Vital-IT computing cluster³⁶. To generate the input for OMA, we extracted long Open Reading Frames (ORFs) and predicted likely coding regions with Transdecoder version 2.0.1 using default options for each transcript and species³⁷. Amino acid sequences of likely coding regions of each of the 16 species were passed to OMA together with an unrooted cladogram built by reducing a published phylogeny of hexapods to the eight hexapod groups analysed in this study³³, which resulted in 286 orthologues shared among all analysed species. We aligned the amino acid sequences (ORFs) of each of the 286 orthologues with M-Coffee using a combination of different alignment methods creating a consensus alignment (option `clustalw2_msa muscle_msa kalign_msa mafft_gins_msa t_coffee_msa`)³⁸. Based on these, we generated corresponding codon alignments from the original nucleotide sequences with T-Coffee version 11.00.8³⁸. As orthologue detection based on naturally incomplete datasets such as transcriptome data can lead to the detection of false positives, e.g. due to loci with paralogous sequences, we tested all protein sequences of the 286 orthologues of all 16 species for their presence in a precompiled set of orthologues of insects (`insecta_hmmer3.1`; www.deep-phylogeny.org/hamstr/download/datasets/hmmer3) using HamStR version 13.2.6³⁹. We only kept multiple sequence alignments of orthologues for which all species had an equivalent in the precompiled orthologue set leaving a set of 153 loci for further analyses. We curated these multiple sequence alignments using Gblocks version 0.91b with sequence type set to codons ($t = c$) and minimum block length set to 4 ($b4 = 4$)⁴⁰.

Accumulation of nonsynonymous mutations. As a measure of purifying selection, we analysed the divergence at nonsynonymous sites normalised for background substitution rates (dN/dS) using CodeML as implemented in PAML version 4.9⁴¹. For this, we first manually constructed an unrooted cladogram of the 16 analysed species based on a published hexapod phylogeny³³. To exclude orthologues that were under positive selection, we ran BUSTED as implemented in HYPHY version 2.3.10⁴² providing as input the unrooted cladogram described above and the multiple sequence alignments of the 153 orthologues. We found 54 loci showing signatures of episodes of positive selection, which left 99 loci for subsequent analyses (for GenBank Accession numbers, see Supplementary Table S2). CodeML relies on a Maximum Likelihood framework to estimate the goodness of fit of a codon substitution model to a sequence alignment and an unrooted species tree with gene-specific branch lengths for inference of branch-specific dN/dS ratios. We calculated per-gene branch lengths for the unrooted cladogram based on each of the 99 codon alignments, accordingly, using RAxML version 8.2.8⁴³ with GTRGAMMAI (with four GAMMA rate categories) set as model of sequence evolution. We modified a custom script used by Brandt *et al.*¹⁵ (see Supplementary information) to pass the loci-specific branch lengths and fixed species tree together with each codon alignment to CodeML for divergence rate estimations. Due to the 1KITE taxon sampling, the asexual and sexual lineages of each hexapod group used in our analysis are likely not natural sister lineages (i.e. not the closest extant relatives) and therefore the time of transition to asexuality (or the split from the closest sexual relative) was unknown. As this could potentially lead to an overestimation of dN/dS under a model restricted to only two rates (one for asexual and one for sexual branches), we chose a free model allowing for different dN/dS ratios, one for each branch in the tree. For between-species comparisons of dN/dS ratios, we excluded all dN/dS ratios of internal branches and four terminal branch dN/dS ratios that were > 1 , indicating positive selection acting at one branch, from statistical analyses. We then tested whether branch-specific dN/dS ratios differed according to (I) gene (II) reproductive mode, or (III) hexapod group using a permutation ANOVA with 5,000 bootstrap replicates (available at <https://gist.github.com/KamilSaron/358c997698b67486be47d4e8eef2921d>)^{44,45}. Differences in dN/dS can be driven by differing synonymous substitution rates and, in the long-term, different levels of saturation at synonymous sites. Given the old age of the splits between sexual and asexual lineages (~40 myo for Phasmatodea – ~160 myo for Zygentoma)^{33,46} we tested for differences in branch-specific dS as described above. To infer whether or not sexual and asexual lineages within individual hexapod groups differed in dN/dS ratios, we compared between reproductive modes using Wilcoxon signed-rank tests. All statistical analyses were done in R version 3.4.4⁴⁷.

'Deleteriousness' of nonsynonymous mutations. To infer the 'deleteriousness' of nonsynonymous substitutions we analysed hydrophobicity changes from ancestral to replacement amino acids along the terminal branches of the phylogenetic tree. Hydrophobic interactions are the main determinants of the 3D conformation of proteins and thus an indicator of protein stability⁴⁸. Inference of ancestral amino acids relies on the presence of an outgroup sequence included in the input amino acid alignments and the phylogenetic tree (cladogram with loci-specific branch lengths) used in analyses with CodeML. Therefore, we first searched for orthologues shared among all 16 hexapod species plus the crustacean *Xibalbanus cf. tulumensis* (previously *Speleonectes cf. tulumensis*), as a representative of Remipedia, the sister-group of hexapods^{32,33}. For this, we predicted ORFs from the assembled transcriptome of *X. cf. tulumensis* and checked the ORFs of *X. cf. tulumensis* for presence of orthologues in the precompiled orthologue set of insects as described above. We found 73 ORFs of *X. cf. tulumensis*, each of them orthologous to one of the previously inferred 99 clusters of orthologues of the 16 hexapod species (for GenBank Accession numbers, see Supplementary Table S2). We again aligned the amino acid sequences (ORFs, including now the sequences of *X. cf. tulumensis*) of the 73 clusters of orthologues and subsequently generated corresponding codon alignments from the original nucleotide sequences as described above. Further, we added *X. cf. tulumensis* to the unrooted manually constructed cladogram used for the calculation of branch lengths for dN/dS ratio analyses as outgroup to all hexapods. Using this tree as a fixed topology, we calculated branch lengths for each orthologue locus individually as described above for analyses of nonsynonymous

Measure	Hexapod group							
	Coll.	Zyg.	Phas.	Man.	Thys.	Stern.	Hym.	Psoc.
dN/dS	V = 1886	<u>V = 3786</u>	V = 1747	V = 1263	V = 2008	V = 2382	V = 2845	V = 2056
	P = 0.107	<u>P < 0.001</u>	P = 0.381	P = 0.585	P = 0.314	P = 0.179	P = 0.092	P = 0.407
CDC	<u>V = 1191</u>	V = 4198	<u>V = 1832</u>	V = 2593	V = 2040	V = 3562	V = 2471	V = 2737
	<u>P < 0.001</u>	P < 0.001	<u>P = 0.025</u>	P = 0.682	P = 0.129	P < 0.001	P = 0.99	P = 0.361

Table 1. V and P values of within-hexapod group comparisons of dN/dS and CDC between reproductive modes. Values were inferred by comparing dN/dS and CDC per gene between the sexual and asexual lineages in each of eight hexapod groups using Wilcoxon signed-rank tests. Underlined and bold measures indicate more effective purifying selection in sexual species (underlined) and asexual species (bold), respectively (see Fig. 2a,b). Coll.: Collembola; Zyg.: Zygentoma; Phas.: Phasmatodea; Man.: Mantodea; Thys.: Thysanoptera; Stern.: Sternorrhyncha; Hym.: Hymenoptera; Psoc.: Psocodea.

mutation accumulation, and then translated the curated alignments into amino acids using EMBOSS version 6.6.0⁴⁹. To predict ancestral amino acid sequences for each internal node in each inferred tree from each gene, we passed the 73 amino acid codon alignments individually with its respective species trees (and estimated loci-specific branch lengths) to CodeML using the modified custom script mentioned above (see Supplementary information). We determined the strength of hydrophobicity changes (Hydrophobicity Scores; HS) for each of the amino acid transitions along the terminal branches of the species tree using a hydrophobicity scoring (HS) matrix^{44,50}. HS indicates the 'deleteriousness' of a nonsynonymous mutation by measuring the strength in hydrophobicity change from ancestral to replacement amino acid. The lower the HS, the stronger is the change in hydrophobicity and, hence, the deleteriousness of the underlying nonsynonymous mutation. We compared values of HS between the two reproductive modes using Generalised Linear Mixed Models (GLMM) implemented in the R package lme4 with gene nested in species set as random effect, correction for overdispersion and Poisson distribution fitting⁵¹.

Accumulation of synonymous mutations. Synonymous mutations are generally assumed to be neutral but can be subject to purifying selection because different codons can influence the speed and accuracy of translation²¹. Hence, we also analysed the effectiveness of selection acting on Codon Usage Bias (CUB). First, we inferred the existence of CUB for each of the 99 orthologues of the 16 species using the Effective Number of Codons (Nc) as measure with the software codonW version 1.4^{52,53}. The Nc specifies the deviation of observed codon usage from equal usage of all codons ranging from 20 (each amino acid is encoded by one codon only; strong CUB) to 61 (equal use of all possible codons; no CUB). Next, we inferred selection on CUB using the Codon Deviation Coefficient (CDC)⁵⁴. Measurement of CDC allows for cross-species comparisons by correcting for background nucleotide composition and is particularly robust because, unlike dN/dS ratio analyses, it does not rely on likelihood and branch length estimates. The CDC represents the deviation of expected CUB based on observed positional GC and purine contents from observed CUB, ranging from 0 (no deviation; no detectable selection on CUB) to 1 (maximum deviation; strong selection on CUB). We estimated CDC for the processed alignments using Composition Analysis Toolkit version 1.3⁵⁴ and analysed it for an effect of I) gene, II) reproductive mode, and III) hexapod group and inferred within-group differences as described above for statistical analyses of dN/dS ratios.

Results

Accumulation of nonsynonymous mutations. We estimated nonsynonymous to synonymous divergence (dN/dS) along individual branches of a phylogenetic tree comprising eight asexual and eight sexual hexapod species for 99 single-copy orthologous protein-coding genes under purifying selection (see Fig. 1; Methods). In genes under purifying selection, nonsynonymous mutations have likely deleterious effects, hence, a higher dN/dS ratio indicates less effective purifying selection⁵⁵. Consistent with the expectation for loci under purifying selection, dN/dS ratios at terminal branches were generally low (mean dN/dS = 0.032). Contrasting the established consensus on deleterious mutation accumulation in asexual lineages, there was no difference in dN/dS when compared between sexual and asexual branches over all hexapod groups (gene effect $P < 0.001$, reproductive mode effect $P = 0.488$, hexapod group effect $P = 0.048$, interaction reproductive mode * hexapod group $P = 0.145$; Permutation ANOVA). The absence of a difference between reproductive modes was not driven by a difference in dS (gene effect $P < 0.001$, reproductive mode effect $P = 0.278$, hexapod group effect $P < 0.001$, interaction reproductive mode * hexapod group $P = 0.004$; Permutation ANOVA). There was significant among-gene variation in dN/dS (range 0–0.454; variance 1.04×10^{-3}) and a significant difference in dN/dS among hexapod groups. To detect whether or not the effectiveness of purifying selection differed between reproductive modes within each hexapod group, we compared dN/dS between reproductive modes in each hexapod group on a per-gene basis. For Zygentoma, we found significantly lower per-gene dN/dS in the sexual as compared to the asexual terminal branch indicating more effective purifying selection for the sexual branch (see Table 1; Fig. 2a red box).

'Deleteriousness' of nonsynonymous mutations. In addition to nonsynonymous mutation accumulation in asexual hexapod lineages, purifying selection is expected to lead to more deleterious amino acid transitions in asexual lineages as reflected by stronger hydrophobicity changes from ancestral to replacement amino acids along asexual and sexual terminal branches⁵⁶. Contrasting this expectation, but consistent with our dN/dS estimates, HS were similar between asexual and sexual branches ($z = -0.152$; $P = 0.879$; GLMM; see

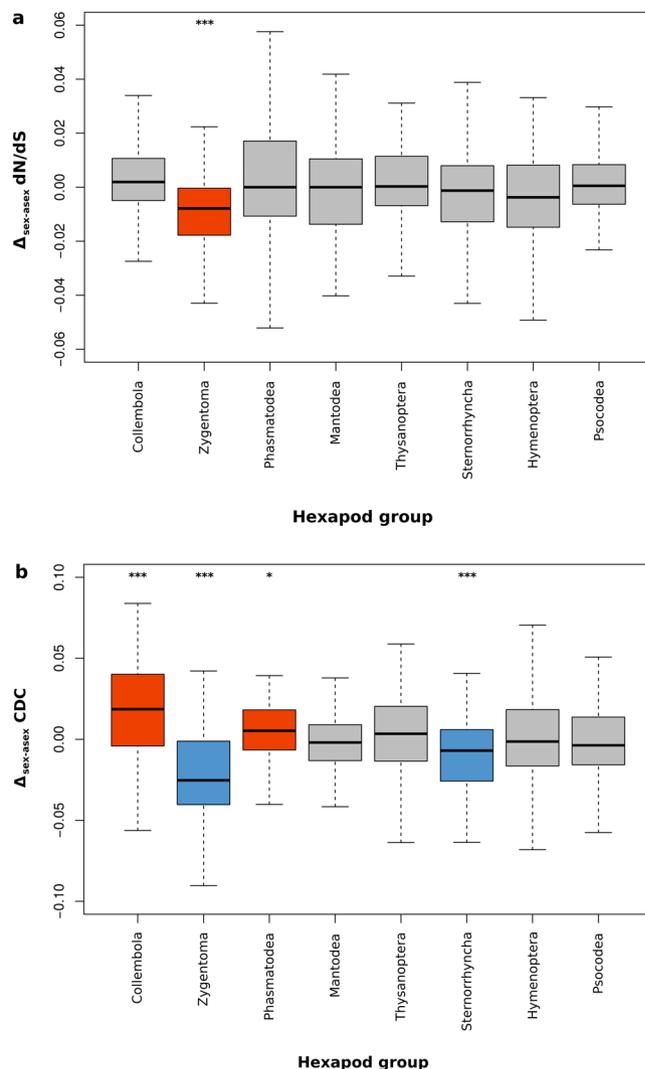


Figure 2. Per-gene differences in (a) dN/dS and (b) CDC between reproductive modes within each hexapod group. The boxplots show the distribution of per-gene differences in dN/dS between the sexual and asexual terminal branch ($\Delta_{\text{sex-asex}} \text{dN/dS}$) and of per-gene differences in CDC between the sexual and asexual species ($\Delta_{\text{sex-asex}} \text{CDC}$) of each of eight hexapod groups for 99 orthologues. For better representation, the ordinate is restricted to between -0.06 and 0.06 and -0.1 and 0.1 , respectively, and outliers outside of 1.5 times the interquartile range (whiskers) are excluded. For dN/dS seven of eight within-hexapod-group comparisons between sexual and asexual branches yielded non-significant results (see Table 1). The red plot denotes a significantly lower per-gene dN/dS in sexual as compared to asexual terminal branches of *Zygentoma* ($V = 3786$, $P < 0.001^{***}$; Wilcoxon signed-rank test). For CDC four of eight within-hexapod-group comparisons between sexual and asexual branches yielded significant results (see Table 1). There was significantly higher per-gene CDC in sexual as compared to asexual Collembola and Phasmatodea (red plots; $V = 1191$, $P < 0.001^{***}$ and $V = 1832$, $P = 0.025^*$, respectively; Wilcoxon signed-rank test) contrasting a significantly lower per-gene CDC in sexual as compared to asexual *Zygentoma* and Sternorrhyncha species (blue plots; $V = 4198$, $P < 0.001$ and $V = 3562$, $P < 0.001$, respectively; Wilcoxon signed-rank test).

Supplementary Fig. S1). Percentages of ancestral to replacement amino acid transitions with more dissimilar hydrophobicity ($\text{HS} < 90$) were similar between asexual and sexual branches (44.936% and 44.893% of all asexual and sexual transitions, respectively).

Accumulation of synonymous mutations. We assessed whether or not purifying selection on synonymous sites was less effective in asexual as compared to sexual hexapod lineages by inferring selection on CUB. To infer whether or not the species investigated use some codons preferentially over others, we first analysed the Effective Number of Codons (Nc). Nc ranges from 20 (each amino acid is encoded by one codon only; strong CUB) to 61 (equal use of all possible codons; no CUB). Overall, all species showed CUB, with Nc means of species ranging from 40.6 in *D. vitifoliae* (Sternorrhyncha) to 54.1 in *F. candida* (Collembola; see Supplementary Fig. S2) which is comparable to values of Nc found in other invertebrate groups, e.g. bivalves and nematodes^{57,58}. Afterwards, we directly inferred selection on CUB using CDC as a measure. CDC calculates the deviation of observed from

predicted CUB by correcting for background nucleotide composition ranging from 0 (no selection on CUB) to 1 (effective selection on CUB). Consistent with the results from analyses of nonsynonymous mutation accumulation and ‘deleteriousness’ of nonsynonymous mutations, there was no difference in per-gene CDC between reproductive modes (gene effect $P < 0.001$, reproductive mode effect $P = 0.283$, hexapod group effect $P < 0.001$, interaction reproductive mode * hexapod group $P < 0.001$; Permutation ANOVA). Further, there was significant among gene variation for CDC (range 0.064–0.329; variance $1.653 * 10^{-3}$) and a significant difference between hexapod groups but, unlike for dN/dS estimates, there was also a significant interaction between hexapod group and reproductive mode. As for dN/dS, we compared per-gene CDC between reproductive modes in each hexapod group, individually. Four of eight within-hexapod-group comparisons between sexual and asexual species yielded significant results (see Table 1; Fig. 2b red and blue boxes, respectively). For Collembola and Phasmatodea, there was a significantly higher per-gene CDC in sexual as compared to asexual species indicating more effective selection on CUB in sexual species, whereas for Zygentoma and Sternorrhyncha there was a significantly lower per-gene CDC in sexual as compared to asexual species indicating more effective selection on CUB in asexual species.

Discussion

It has become established consensus among evolutionary biologists that sex and recombination increase the effectiveness of purifying selection, based on theoretical considerations and empirical evidence derived from a multiplicity of studies¹². Our results do not match these studies: we find no evidence for accumulation of deleterious mutations in asexual hexapod species. Overall, asexual and sexual lineages neither differed in nonsynonymous site divergences corrected for background substitution rates (dN/dS), in synonymous substitution rates (dS) potentially influencing dN/dS ratios, in the ‘deleteriousness’ of nonsynonymous mutations nor in selection acting on CUB (CDC). Comparisons between sexual and asexual lineages within each group differed for several hexapod groups which may hint at group-specific differences in the consequences of asexuality on effectiveness of purifying selection. However, here, we refrain from drawing conclusions based on the within-group comparisons because the taxon sampling of 1KITE did not allow for analysing multiple within-group replicates but restrict the discussion to the overall result of more effective selection being absent in our data of sexual hexapods: which (non-mutually exclusive) mechanisms might be responsible for the discrepancy between this finding and the established consensus?

First, analyses of purifying selection acting at nonsynonymous sites and on their ‘deleteriousness’ may have been affected due to data limitations. dN/dS ratio analyses and ancestral state reconstructions rely on branch length estimates (see Methods). Due to the limited nature of the taxon sampling of the 1KITE data set for this study, the asexual and sexual species analysed are most likely not sister species, but rather more distantly related^{46,59}. Therefore, the loss of sex did not occur with the split of the sex-asex species pair as present in the given phylogenetic tree here, such that evolution over some fraction of the asexual branch was likely sexual. If mutations occurring in asexual lineages did not accumulate at greatly increased rates compared with sexuals, a change in deleterious mutation accumulation along with the transition to asexuality might be masked by the rates that occurred in the sexual fraction of the branch. Thus, the power to detect an effect of reproductive mode on the effectiveness of purifying selection might be low. Further, the orthologue search among phylogenetically distantly related hexapod groups and the stringent control for false positive orthologues and loci under positive selection resulted in a rather small orthologue set for analysis (99 orthologue loci). This biases the analyses towards strongly conserved loci and excludes recently evolved orthologues which might differ in accumulation of deleterious mutations between reproductive modes. Also, the within hexapod group comparisons between reproductive modes for CDC did not resemble those for dN/dS; in the case of Zygentoma they even opposed them (see Fig. 2a,b; Table 1). This is surprising, given that translational selection acting at synonymous sites is assumed to be weak and effective purging of synonymous mutations likely occurs at lower rates as compared to that of nonsynonymous mutations. Hence, within hexapod group comparisons of purifying selection acting at nonsynonymous sites should not oppose those at synonymous sites.

Second, the analysed asexual lineages may have lost sex too recently to have fixed enough deleterious mutations to be detected. This was for example also assumed to be responsible for absence of deleterious mutation accumulation observed in whole transcriptome data of fissiparous *Lineus* ribbon worms and four aphid species^{14,18}. In fact, the occurrence of abundant males in some locally restricted populations of *N. phytophila*, *E. californica*, and *L. bostrychophila* analysed in this study is in line with a rather recent loss of sex¹⁸. Additionally, interference of rare, furtive, or cryptic sex with asexual genome evolution may explain the observed absence of less effective selection in asexual hexapods because rare events of sex are assumed to be sufficient to compensate for predicted consequences of asexuality⁶⁰.

A third reason for the absence of deleterious mutation accumulation in the analysed data of asexual hexapod lineages may be effective homogenising mechanisms. Gene conversion and DNA repair have been shown to maintain DNA integrity, e.g. within the human Y chromosome, higher plant chloroplasts and animal mitochondria^{61–63}. If homogenising mechanisms play a role in the analysed species remains to be investigated.

Fourth, besides reproductive mode, population size acts as a major determinant of effectiveness of selection with the speed of mutation accumulation being inversely related to population size, as shown by modelling approaches^{9,64,65}. Large population sizes have been suggested to maintain effective purifying selection in asexual organisms, e.g. in oribatid mites and polyphagous pest insects, such as scale insects^{15,66,67}. Further, large population sizes have been shown to increase the effectiveness of selection acting on CUB in a variety of animal species and on nonsynonymous mutations with large deleterious effects in *Caenorhabditis elegans*^{68,69}. As information on population sizes of the analysed species is absent, any correlation between population sizes and the observed absence of more effective purifying selection acting at nonsynonymous and synonymous sites has to remain speculative.

In conclusion, our results do not support the established consensus of reduced effectiveness of purifying selection in asexual species, contrasting earlier analyses of few individual genes in a variety of animal and plant species. However, whether or not this is due to the limited nature of our data or due to non-mutually exclusive biological

mechanisms has to remain elusive at this point. Future studies, hence, need not only to include large gene sets but also carefully selected closely related sexual and asexual focal lineages to study the effectiveness of purifying selection in asexual organisms in more detail.

Data Availability

Sequence data analysed in this study is available at NCBI GenBank under Accession numbers MH551269–MH551284, MH602437–MH602956, MH637812–MH638065 and MH799322–MH800185. Supplementary data is available for download from the digital repository DRYAD under <https://doi.org/10.5061/dryad.5501rv4> (see Supplementary Archives S1–S4; Supplementary information).

References

- Bell, G. *The masterpiece of nature: the evolution and genetics of sexuality* 1–635 (University of California Press, 1982).
- Otto, S. P. The evolutionary enigma of sex. *Am. Nat.* **174**, S1–S14 (2009).
- Lehtonen, J., Jennions, M. D. & Kokko, H. The many costs of sex. *Trends Ecol. Evol.* **27**, 172–178 (2012).
- Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
- Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
- Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964).
- Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
- Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
- Lynch, M., Bürger, R., Butcher, D. & Gabriel, W. The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344 (1993).
- Lynch, M. & Blanchard, J. L. Deleterious mutation accumulation in organelle genomes. *Genetica* **102–103**, 29–39 (1998).
- Bachtrog, D., Hom, E., Wong, K. M., Maside, X. & de Jong, P. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol.* **9**, R30, <https://doi.org/10.1186/gb-2008-9-2-r30> (2008).
- Hartfield, M. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.* **29**, 5–22 (2016).
- McDonald, M. J., Rice, D. P. & Desai, M. M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236 (2016).
- Ament-Velásquez, S. L. *et al.* Population genomics of sexual and asexual lineages in fissiparous ribbon worms (*Lineus*, Nemertea): hybridization, polyploidy and the Meselson effect. *Mol. Ecol.* **25**, 3356–3369 (2016).
- Brandt, A. *et al.* Effective purifying selection in ancient asexual oribatid mites. *Nat. Commun.* **8**, 873, <https://doi.org/10.1038/s41467-017-01002-8> (2017).
- Hollister, J. D. *et al.* Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Mol. Biol. Evol.* **32**, 896–905 (2014).
- Lovell, J. T., Williamson, R. J., Wright, S. I., McKay, J. K. & Sharbel, T. F. Mutation accumulation in an asexual relative of *Arabidopsis*. *PLoS Genet.* **13**, e1006550, <https://doi.org/10.1371/journal.pgen.1006550> (2017).
- Ollivier, M. *et al.* Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that differ by reproductive mode. *Genome Biol. Evol.* **4**, 155–167 (2012).
- Bast, J. *et al.* Consequences of asexuality in natural populations: insights from stick insects. *Mol. Biol. Evol.* **35**, 1668–1677 (2018).
- Neiman, M., Meirmans, P. G., Schwander, T. & Meirmans, S. Sex in the wild: how and why field-based studies contribute to solving the problem of sex. *Evolution* **72**, 1194–1203 (2018).
- Hershberg, R. & Petrov, D. A. Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
- van der Kooij, C. J. & Schwander, T. On the fate of sexual traits under asexuality. *Biol. Rev. Camb. Philos. Soc.* **89**, 805–819, <https://doi.org/10.1111/brv.12078> (2014).
- Vershinina, A. O. & Kuznetsova, V. G. Parthenogenesis in Hexapoda: Entognatha and non-holometabolous insects. *J. Zoolog. Syst. Evol. Res.* **54**, 257–268 (2016).
- Tree of Sex Consortium. Tree of Sex: a database of sexual systems. *Sci Data* **1**, 140015, <https://doi.org/10.1038/sdata.2014.15> (2014).
- Frati, F. *et al.* High levels of genetic differentiation between *Wolbachia*-infected and non-infected populations of *Folsomia candida* (Collembola, Isotomidae). *Pedobiologia* **48**, 461–468 (2004).
- Pijnacker, L. P. The maturation divisions of the parthenogenetic stick insect *Carausius morosus* Br. (Orthoptera, Phasmidae). *Chromosoma* **19**, 99–112 (1966).
- White, M. J. D. The chromosomes of the parthenogenetic mantid *Brunneria borealis*. *Evolution* **2**, 90–93 (1948).
- Nguyen, D. T., Spooner-Hart, R. N. & Riegler, M. Polyploidy versus endosymbionts in obligately thelytokous thrips. *BMC Evol. Biol.* **15**, 23, <https://doi.org/10.1186/s12862-015-0304-6> (2015).
- Simon, J.-C., Rispe, C. & Sunnucks, P. Ecology and evolution of sex in aphids. *Trends Ecol. Evol.* **17**, 34–39 (2002).
- Stouthamer, R. & Mak, F. Influence of antibiotics on the offspring production of the *Wolbachia*-infected parthenogenetic parasitoid *Encarsia formosa*. *J. Invertebr. Pathol.* **80**, 41–45 (2002).
- Yang, Q. *et al.* Morphological and molecular characterization of a sexually reproducing colony of the booklouse *Liposcelis bostrychophila* (Psocodea: Liposcelidae) found in Arizona. *Sci. Rep.* **5**, 10429, <https://doi.org/10.1038/srep10429> (2015).
- von Reumont, B. M. *et al.* Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Mol. Biol. Evol.* **29**, 1031–1045 (2012).
- Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
- Peters, R. S. *et al.* Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Mol. Phylogenet. Evol.* **120**, 286–296 (2018).
- Peters, R. S. *et al.* Evolutionary History of the Hymenoptera. *Curr. Biol.* **27**, 1013–1018 (2017).
- Train, C.-M., Glover, N. M., Gonnert, G. H., Altenhoff, A. M. & Dessimoz, C. Orthologous Matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference. *Bioinformatics* **33**, i75–i82 (2017).
- Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
- Wallace, I. M., O'Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
- Ebersberger, I., Strauss, S. & von Haeseler, A. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* **9**, 157, <https://doi.org/10.1186/1471-2148-9-157> (2009).
- Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
- Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
- Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Henry, L., Schwander, T. & Crespi, B. J. Deleterious mutation accumulation in asexual *Timema* stick insects. *Mol. Biol. Evol.* **29**, 401–408 (2012).

45. Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* 1–480 (Chapman & Hall/CRC, 1997).
46. Bradler, S., Cluquennois, N. & Buckley, T. R. Single origin of the Mascarene stick insects: ancient radiation on sunken islands. *BMC Evol. Biol.* **15**, 196, <https://doi.org/10.1186/s12862-015-0478-y> (2015).
47. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, <http://www.R-project.org/> (2013).
48. Pace, C. N. *et al.* Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* **408**, 514–528 (2011).
49. Rice, P. EMBOSS: the european molecular biology open software suite. *Trends Genet.* **16**, 2–3 (2000).
50. Riek, R. P. *et al.* Evolutionary conservation of both the hydrophilic and hydrophobic nature of transmembrane residues. *J. Theor. Biol.* **172**, 245–258 (1995).
51. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects models using lme4. *J. Stat. Soft.* **67**, 1–48 (2014).
52. Peden, J. CodonW version 1.4.2, <http://codonw.sourceforge.net/culong.html>.
53. Wright, F. The effective number of codons used in a gene. *Gene* **87**, 23–29 (1990).
54. Zhang, Z. *et al.* Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance. *BMC Bioinformatics* **13**, 43, <https://doi.org/10.1186/1471-2105-13-43> (2012).
55. Li, W. H., Wu, C. I. & Luo, C. C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174 (1985).
56. Sharbrough, J., Luse, M., Boore, J. L., Logsdon, J. M. Jr. & Neiman, M. Radical amino acid mutations persist longer in the absence of sex. *Evolution* **72**, 808–824 (2018).
57. Mitreva, M. *et al.* Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* **7**, R75, <https://doi.org/10.1186/gb-2006-7-8-r75> (2006).
58. Gerdol, M., De Moro, G., Venier, P. & Pallavicini, A. Analysis of synonymous codon usage patterns in sixty-four different bivalve species. *PeerJ* **3**, e1520, <https://doi.org/10.7717/peerj.1520> (2015).
59. Munro, J. B. *et al.* A molecular phylogeny of the Chalcidoidea (Hymenoptera). *PLoS One* **6**, e27023, <https://doi.org/10.1371/journal.pone.0027023> (2011).
60. Hurst, L. D., Hamilton, W. D. & Ladle, R. J. Covert sex. *Trends Ecol. Evol.* **7**, 144–145 (1992).
61. Hughes, J. F. *et al.* Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
62. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **46**, 85–94 (2006).
63. Kang, D. & Hamasaki, N. Maintenance of mitochondrial DNA integrity: repair and degradation. *Curr. Genet.* **41**, 311–322 (2002).
64. Rice, W. R. & Friberg, U. A graphical approach to lineage selection between clonals and sexuals In *Lost sex* (eds Schön, I., Martens, K. & Dijk, P.) 75–97 (Springer Netherlands, 2009).
65. Gordo, I. & Charlesworth, B. The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics* **154**, 1379–1387 (2000).
66. Ross, L., Hardy, N. B., Okusu, A. & Normark, B. B. Large population size predicts the distribution of asexuality in scale insects. *Evolution* **67**, 196–206 (2013).
67. Normark, B. B. & Johnson, N. A. Niche explosion. *Genetica* **139**, 551–564 (2011).
68. Katju, V., Packard, L. B., Bu, L., Keightley, P. D. & Bergthorsson, U. Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective population sizes. *Evolution* **69**, 104–116 (2015).
69. Galtier, N. *et al.* Codon usage bias in animals: disentangling the effects of natural selection, effective population size and GC-biased gene conversion. *Mol. Biol. Evol.* **35**, 1092–1103 (2018).

Acknowledgements

We thank Kamil S. Jaron for help with statistics, Patrick Tran Van for help with bioinformatics, Hans Pohl for permission of using silhouettes of hexapod groups and Ina Schäfer, Mark Maraun and Tamara Hartke for helpful comments on the manuscript. We also thank the 1KITE community, in particular Xin Zhou and Bernhard Misof, the 1KITE Basal Hexapod group, the 1KITE Polyneoptera group, the 1KITE Hymenoptera group and the 1KITE Paraneoptera group for collecting material, sequencing and assembly, cleaning of transcriptomes and granting us access to unpublished transcriptome data. Further, we thank Beijing Genomics Institute for funding sequencing of the 1KITE transcriptomes. This study was supported by core funding of S.S. and by DFG research fellowship BA 5800/1-1 to J.B.

Author Contributions

A.B., J.B. and K.K. conceived and designed the study. K.M., K.S., A.D. and R.M. provided transcriptome data and respective taxa specific information. A.B. and J.B. performed data analysis with the help of K.M. for identifying orthologs. A.B., J.B., K.M., K.K. and S.S. contributed to data interpretation and analyses and A.B. wrote the paper with input from all other authors. All co-authors approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41821-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019