Consequences of Asexuality in Natural Populations: Insights from Stick Insects

Jens Bast,*[1] Darren J. Parker,[1,2] Zoé Dumas,[1] Kirsten M. Jalvingh,[1] Patrick Tran Van,[1,2] Kamil S. Jaron,[1,2] Emeric Figuet,[3] Alexander Brandt,[4] Nicolas Galtier,[3] and Tanja Schwander[1]

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Institute of Evolutionary Sciences, University of Montpellier, CNRS, IRD, EPHE, Montpellier, France
[4]Johann-Friedrich-Blumenbach Institute of Zoology and Anthropology, University of Goettingen, Goettingen, Germany

*Corresponding author: E-mail: mail@jensbast.com.
Associate editor: Stephen Wright

Abstract

Recombination is a fundamental process with significant impacts on genome evolution. Predicted consequences of the loss of recombination include a reduced effectiveness of selection, changes in the amount of neutral polymorphisms segregating in populations, and an arrest of GC-biased gene conversion. Although these consequences are empirically well documented for nonrecombining genome portions, it remains largely unknown if they extend to the whole genome scale in asexual organisms. We identify the consequences of asexuality using de novo transcriptomes of five independently derived, obligately asexual lineages of stick insects, and their sexual sister-species. We find strong evidence for higher rates of deleterious mutation accumulation, lower levels of segregating polymorphisms and arrested GC-biased gene conversion in asexuals as compared with sexuals. Taken together, our study conclusively shows that predicted consequences of genome evolution under asexuality can indeed be found in natural populations.

Key words: evolution of sex, parthenogenesis, polymorphism, purifying selection, biased gene conversion, *Timema*.

Introduction

The absence of recombination is a common feature in eukaryotic genomes and it can span over a few hundred bases, entire chromosomes and, as in the case of asexual organisms, over complete genomes. The reasons for an absence of recombination can vary. For example, centromeric regions do not recombine because of their mechanistic role as the attachment points for kinetochores during meiosis (Talbert and Henikoff 2010). Recombination can be reduced in genome portions with nearby transposable elements as a consequence of heterochromatin formation through local epigenetic silencing (Castel and Martienssen 2013), or because of extensive structural differences between homologous chromosomes, a trait notably characterizing many supergenes and sex chromosomes (Bachtrog 2013; Schwander et al. 2014). Finally, meiotic recombination is completely lacking in asexual organisms because of clonal reproduction.

Theory predicts two major consequences for selection on genome portions and complete genomes that lack recombination. First, selection is less effective, because physical linkage among loci hinders selection's ability to act upon loci independently (Muller 1964; Hill and Robertson 1966; Felsenstein 1974; Keightley and Otto 2006). This should translate into decreased rates of adaptation and increased accumulation of mildly deleterious mutations, including point mutations and

repetitive elements. A related second consequence is that population levels of polymorphisms change (Hill and Robertson 1966; Charlesworth et al. 1993). Whether the lack of recombination increases or decreases polymorphism depends on the relative importance of multiple counteracting factors, and especially the level of heterozygosity in the nonrecombining region. A classical prediction is that heterozygosity (and thus polymorphism) increases over time in the absence of recombination, as allelic sequences diverge independently of each other (Birky 1996). Polymorphism levels can therefore be higher in nonrecombining than recombining genomes and genome portions (Balloux et al. 2003). In contrast, mitotic gene conversion and selection at linked sites (background selection) lead to a loss of heterozygosity and diversity in the absence of recombination, which can translate to lower polymorphism in recombining than nonrecombining genome portions (Hill and Robertson 1966; Charlesworth et al. 1993).

In addition to selection-driven consequences of linkage, mechanistic effects that are independent of selection or drift can differ between recombining and nonrecombining genome portions. For example, recombination per se is thought to be mutagenic, which can affect polymorphism levels and the rate of divergence (Webster and Hurst 2012). Finally, base compositions are expected to differ between recombining and nonrecombining genomes and genome portions because

Article

recombination-associated gene conversion will tend to increase GC content. Gene conversion is a molecular mechanism associated with DNA double-strand breaks and mismatch repair during recombination, where one allele is replaced by its homolog (van den Bosch et al. 2002). In many species, gene conversion is nonrandom with respect to the base composition of an allele; the bases A and T are more likely to be converted into G and C than the reverse (Marais 2003; Mugal et al. 2015). In the absence of meiotic recombination, such GC-biased gene conversion (gBGC) should occur at greatly reduced rates, and the base composition of neutral sequences should shift towards the mutational equilibrium (Galtier et al. 2018). There is accumulating empirical evidence for several of the above mentioned consequences in nonrecombining genome portions of sexual organisms. For instance, studies of Y-chromosomes revealed lower levels of neutral polymorphism, increased rates of deleterious mutation accumulation and lower rates of adaptive evolution compared with recombining parts of genomes (Bachtrog and Charlesworth 2002; Skaletsky et al. 2003; Bachtrog 2013). Furthermore, as a result of reduced gBGC, nonrecombining genome portions typically feature low GC content (Spencer et al. 2006; Duret and Galtier 2009; Pessia et al. 2012; Glémin et al. 2015).

In contrast to these nonrecombining, "asexual" genome portions in sexual genomes, the consequences of a genome-wide lack of recombination in natural populations of asexual species remain less clear. In particular, the effects of asexuality on segregating polymorphisms and gBGC remain largely untested. The reduced effectiveness of selection in asexuals was investigated in several studies via analyses of deleterious mutation accumulation, but with mixed results. Most studies based on few individual genes indeed found increased rates of mutation accumulation in asexual as compared with sexual lineages, but others reported no differences (reviewed in Hartfield 2016). Similarly, the three available genome-wide studies all reported different patterns, including one case where selection was more effective in asexual than sexual species (Hollister et al. 2015; Ament-Velásquez et al. 2016; Brandt et al. 2017). Given these conflicting results, additional whole-genome studies are warranted.

Timema stick insects are an ideal group for identifying signatures of asexual genome evolution. Because asexuality is a lineage level-trait, independently derived asexuals are required to disentangle consequences of asexuality from species-specific factors. Seven such independently derived asexuals are known in *Timema*, with an ecologically equivalent sexual sister species at hand for comparison (fig. 1). Sexual and asexual sister species are morphologically alike, live and feed on similar host plants and have a generation time of one year (Law and Crespi 2002). The asexual lineages are obligate apomicts, diploid, not of hybrid origin and differ in age, ranging from 0.2 to 2 My (Schwander and Crespi 2009; Schwander et al. 2011).

Here, we tested whether the lack of recombination in *Timema* asexuals is associated with 1) changes in the strength of gBGC; 2) changes in the amount of genetic variation within populations; and 3) increased rates of deleterious mutation
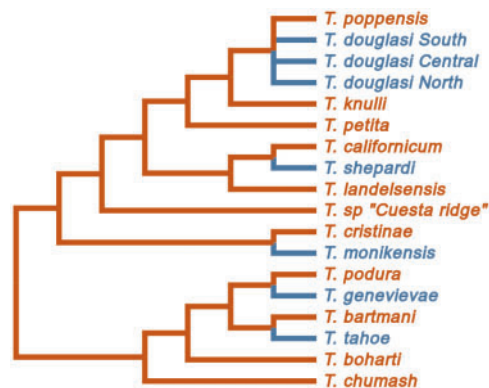

Fig. 1. Phylogeny of *Timema* species (sexual species phylogeny redrawn from Riesch et al. 2017; asexual lineages added from Schwander et al. 2011). Sexual species are colored red, asexual species blue. For this study, five sex–asex species pairs were used (for *Timema douglasi* only the southern lineage).

accumulation and more segregating deleterious variants. We generated de novo transcriptomes for ten *Timema* species (see Materials and Methods for details), including transcriptomes of five asexual lineages (*T. tahoe, T. monikensis, T. douglasi* "South," *T. shepardi*, and *T. genevievae*), and their sexual sister-species (*T. bartmani, T. cristinae, T. poppensis, T. californicum*, and *T. podura*).

Results

De novo Transcriptome Assembly and Ortholog Prediction

Depending on the species, the quality-filtered assemblies contained between 31,747 and 45,655 transcripts with a N_{50} between 1,301 and 1,836 bp. Open reading frames (ORFs) were identified in 7,329–10,436 of the transcripts, and these ORFs were used to determine orthologs shared between the two species of a pair (between 5,329 and 5,908 per pair) as well as orthologs shared across all ten species (3,010 orthologs; see Materials and Methods for details, supplementary data 1 and supplementary tables 1 and 2, Supplementary Material online).

The ages of the asexual lineages were previously estimated based on intra and interspecific divergence at a single mtDNA locus (Schwander et al. 2011). We used orthologs within each species-pair to update the ranking of asexuals from youngest to oldest, with interspecific divergence as a proxy for lineage age (fig. 2).

To evaluate the robustness of the 3,010 ten-species orthologs, we screened each transcriptome for variants of these orthologs. Variants are transcripts with the same ORF, but that contain sufficiently many SNPs, deletions or insertions of variable length in any part of the ORF to result in separately assembled transcripts in the de novo assembly (our ortholog detection pipeline retained the longest ortholog if multiple options are available, see Materials and Methods for details). Surprisingly, the proportion of orthologs with variants was significantly higher for the asexual than sexual *Timema* species (mean asex:
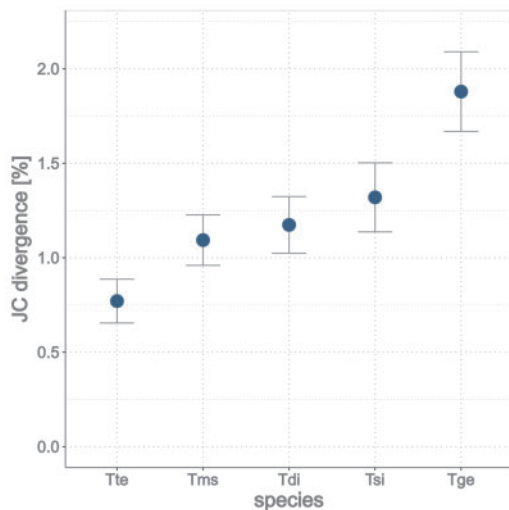
0.076, mean sex: 0.013; paired *t*-test, $t = 7.29$, $df = 4$, $P < 0.001$; supplementary data 2, Supplementary Material online). Given this systematic difference between sexual and asexual species, we evaluated whether ortholog variants represent transcripts from multiple gene copies by mapping genomic reads from single individuals generated in a different study (SRR5248877–SRR5248936) to our 3,010 ten-species orthologs (see Supplementary Material online). If ortholog variants were generated by gene duplications, the coverage for orthologs with variants would be higher than the coverage for orthologs without variants (i.e., an ortholog with two copies in the genome would have twice the coverage relative to an ortholog with single copy). Our analyses revealed that ortholog variants do not stem from gene duplicates, as the coverage for orthologs was the same, independently of whether an ortholog had a variant or not in a given transcriptome (supplementary fig. 1, Supplementary Material online). By consequence, ortholog variants are most likely generated by transcriptional noise and error, or perhaps by different alleles (see discussion for details). Indeed half of the variants (725 out of 1,353) differed in length from each other (up to 7-fold, mean length difference: 13%) and 27% of the variants further featured SNP divergences (mean pairwise distance 1.55%). While the enrichment of orthologs with variants in asexuals is interesting per se, it is important to note that these variants did not affect our subsequent analyses and results. This is because all polymorphism, pN/pS and dN/dS analyses were based on one "best" ortholog variant (i.e., longest) and estimates of segregating polymorphism were conservative, as the presence of variants could potentially inflate polymorphisms in asexuals compared
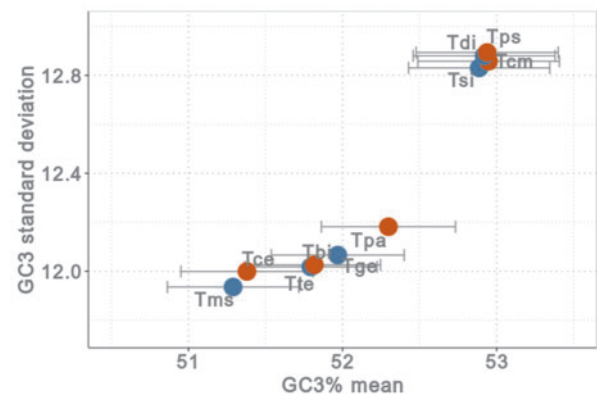
with sexuals but sexuals had more polymorphisms than asexuals (see below).

Reduced gBGC in Asexuals

Because meiotic recombination has stopped in asexuals, gBGC should be strongly reduced if the process was at work in *Timema*. We tested for evidence of this mechanistic consequence of asexuality via various approaches. First, we analyzed GC content at third codon positions (GC3 content), which are most likely neutral. In agreement with the hypothesis that gBGC favors the increase of GC content and generates variation among genes experiencing different recombination rates, sexual species within each pair had a higher GC3 content (Wilcoxon signed-rank test, $P < 0.01$ in all pairs) and higher variance among genes than asexual species (fig. 3). We also compared per gene GC3 in each of the sexual–asexual species pairs. We calculated the proportion of genes in which GC3 was higher in the sexual than in the asexual species, and found that in all five pairs this proportion was above 0.5. Interestingly, the proportion was maximal (0.62) in the oldest pair (*T. podura* vs. *T. genevieve*), minimal (0.51) in the youngest pair (*T. bartmani* vs. *T. tahoe*), and intermediate in the other three pairs (0.52, 0.54, 0.55). Then we investigated the substitution processes that shaped base composition in *Timema* using maximum likelihood approaches. We estimated branch-specific equilibrium GC3 using the nonhomogeneous maximum likelihood method developed by Galtier and Gouy 1998. Similar to the results from the previous approach, the asexual species were characterized by a lower equilibrium GC3 than their sexual counterparts (likelihood ratio test, $P < 2.2e\text{-}16$ between stationary and nonstationary per-branch model of sequence evolution), meaning that current substitution processes drive GC3 towards lower values in asexuals than in sexuals. We mapped synonymous substitutions along the *Timema* phylogeny and estimated the per-lineage number of weak-to-strong (A/T → G/C), strong-to-weak (G/C → A/T), and GC-conservative (A ↔ T, G ↔ C) substitutions. As expected
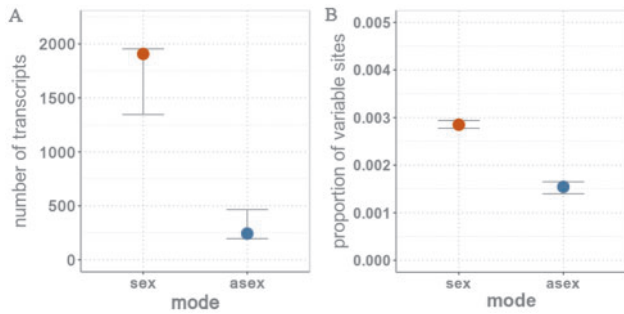
FIG. 4. Medians (and 95% CI with 5,000 bootstrap replicates) of (A) number of transcripts containing SNPs and (B) proportion of variable sites per transcript for sexual (red) and asexual (blue) *Timema* species.
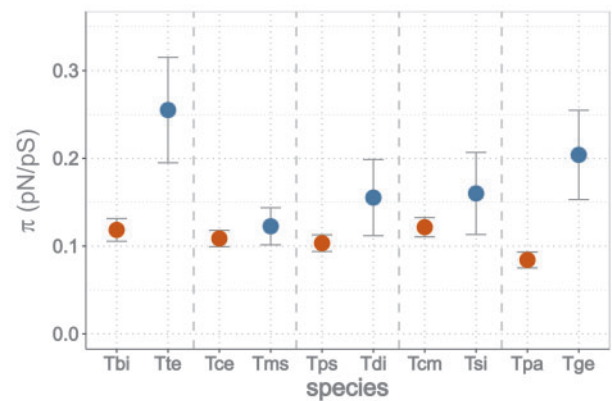

FIG. 5. Means (and 95% CI) of nonsynonymous segregating polymorphisms (pN/pS) of sexual (in red) and asexual (in blue) *Timema* species. Tbi: *T. bartmani*, Tte: *T. tahoe*, Tce: *T. cristinae*, Tms: *T. monikensis*, Tps: *T. poppensis*, Tdi: *T. douglasi*, Tcm: *T. californicum*, Tsi: *T. shepardi*, Tpa: *T. podura*, Tge: *T. genevievae*. Species pairs are ranked by age of the asexual lineage (youngest to oldest).

under gBGC arrest, asexuals always harbored a higher proportion of strong-to-weak substitutions and a lower proportion of weak-to-strong substitutions than to sexuals (supplementary data 3, Supplementary Material online). Finally, we calculated GC-content in UTR regions and correlated it to GC3. We obtained a strong and significantly positive correlation between GC_UTR and GC3 in all ten species (r between 0.37 and 0.44, $P < 0.001$ in all cases), demonstrating that GC3 is governed by evolutionary forces independent of the process of translation—most likely gBGC. In combination, the results from these different approaches strongly support the hypothesis of a reduced impact of gBGC in *Timema* asexuals.

Reduced Levels of Polymorphisms in Asexuals

Polymorphism levels at individual loci can be higher or lower in asexual than sexual populations, depending on the amount of heterozygosity in asexuals and the number of different clones (genotype diversity). To quantify the amount of segregating polymorphisms in sexuals and asexuals, we identified SNPs in transcripts by mapping the read data derived from three pooled individuals (all collected from the same population) back to the 3,010 ten-species orthologs that include only one "best" variant per ortholog (see Materials and Methods for details). Since alternative variants of each ortholog were not included, reads corresponding to these variants mapped to the "best" ortholog. Specifically, for 63% of the orthologs, the sum of reads corresponding to the different variants was exactly identical to the number of reads mapping to the best ortholog; for 93%, the read numbers differed by 5% or less (supplementary fig. 2, Supplementary Material online). As expected given the larger number of variants in asexuals than sexuals, mapping quality was somewhat lower in asexuals than sexuals, however not significantly so (paired *t*-test, $t = 1.3$, df = 7.9, $P = 0.215$). We found that a higher proportion of transcripts contained SNPs in sexuals than asexuals (paired *t*-test, $t = 13.46$, df = 4, $P < 0.001$; fig. 4A), in spite of the enrichment for ortholog variants in asexual transcriptomes. Similarly, among these transcripts that contained SNPs, the proportion of variable sites per transcript was higher for sexuals than asexuals (paired *t*-test, $t = 3.850$, df = 4, $P = 0.018$; fig. 4B).

Purifying Selection is Less Effective in Asexuals

To test the prediction that selection acts less effectively on asexuals than sexuals, we focused on mutations that are expected to affect protein function and are likely deleterious. This should be the case for nonsynonymous mutations in the 3,010 orthologs shared among the ten species. Indeed, since the *Timema* species diverged ~30 Ma, the vast majority of these orthologs should be under purifying selection (or they would be too diverged to be identified as orthologs). Thus, we quantified whether more nonsynonymous mutations segregate in asexual than sexual populations and whether these mutations accumulate faster in asexual than sexual species.

We identified nonsynonymous and synonymous SNPs per transcript by mapping the read data (derived from three pooled individuals) to the 3,010 ten-species orthologs (see Materials and Methods). As expected, the ratio of nonsynonymous to synonymous polymorphisms (pN/pS) was larger in asexual than sexual populations (paired *t*-test, $t = -3.02$, df = 4, $P = 0.039$; fig. 5). This indicates that deleterious variants are rapidly removed by selection in sexual populations but persist longer in asexual populations.

To elucidate whether deleterious variants also become fixed faster in asexual than sexual *Timema* species, we estimated the ratio of nonsynonymous to synonymous divergence ($\omega = dN/dS$) along the branches of the *Timema* phylogeny, using the maximum likelihood methods implemented in PAML (Yang 2007; see Materials and Methods for details). Again, for transcripts under purifying selection (i.e., the majority of the ten-species orthologs), most nonsynonymous changes are likely mildly deleterious (Li et al. 1985) and a higher ω ratio therefore indicates a higher rate of nonsynonymous mutation accumulation and less effective purifying selection. As the *Timema* system provides evolutionary replicates, we allowed ω to vary independently for each gene and branch on the phylogeny (i.e., applying a "free model"), to take these replicates into account. Out of the 3,010 ten-species orthologs, we removed 200 without
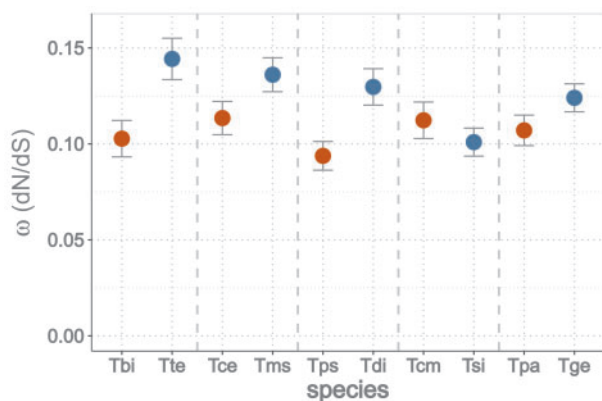
FIG. 6. Means (and 95% CI) of $\omega = dN/dS$ ratios from 2,487 orthologs in sexual (in red) and asexual (in blue) *Timema* species. Tbi: *T. bartmani*, Tte: *T. tahoe*, Tce: *T. cristinae*, Tms: *T. monikensis*, Tps: *T. poppensis*, Tdi: *T. douglasi*, Tcm: *T. californicum*, Tsi: *T. shepardi*, Tpa: *T. podura*, Tge: *T. genevievae*. Species pairs are ranked by age of the asexual lineage (youngest to oldest).

FIG. 7. Means (and 95% CI) of the codon deviation coefficient (CDC) of sexual (in red) and asexual (in blue) *Timema* species. Tbi: *T. bartmani*, Tte: *T. tahoe*, Tce: *T. cristinae*, Tms: *T. monikensis*, Tps: *T. poppensis*, Tdi: *T. douglasi*, Tcm: *T. californicum*, Tsi: *T. shepardi*, Tpa: *T. podura*, Tge: *T. genevievae*. Species pairs are ranked by age of the asexual lineage (youngest to oldest).

variation and 323 with $\omega \geq 1$ or without synonymous substitutions, which left 2,487 transcripts for comparisons. Consistent with the predictions of less effective purifying selection in asexual species, we found that asexuals accumulated nonsynonymous mutations at a higher rate than sexuals (permutation ANOVA: gene effect, species pair effect, reproductive mode effect, and interaction between pair and mode all $P < 0.001$; fig. 6). We additionally ran a simpler three-ratio model that estimates one dN/dS ratio for internal branches and one for terminal branches of each reproductive mode, which corroborated these findings (see Supplementary Material online).

Increased Substitution Rates at Synonymous Sites in Asexuals

Synonymous substitutions (dS) are commonly assumed to be neutral, and are used to scale nonsynonymous substitutions in the above-mentioned analyses. However, purifying selection can also act on synonymous sites, because different synonymous codons influence for example the speed and accuracy of translation (Hershberg and Petrov 2008). As a consequence, the frequencies by which different codons are used can differ (codon usage bias). We compared the effective number of codons (Enc), a quantitative estimate for the level of codon usage bias, as well as synonymous substitution rates (dS) between sexual and asexual *Timema* pairs. *Timema* featured Enc values that indicate weaker codon usage bias than in animals with large population sizes (e.g., nematodes), but the bias level was similar for species with different reproductive modes (gene effect $P < 0.001$, species pair effect $P < 0.001$, reproductive mode $P = 0.241$, and interaction between pair and mode all $P = 0.928$; permutation ANOVA; supplementary fig. 3, Supplementary Material online). Estimates of dS were systematically elevated in asexuals (gene effect, species pair effect, reproductive mode, and interaction between pair and mode all $P < 0.001$; permutation
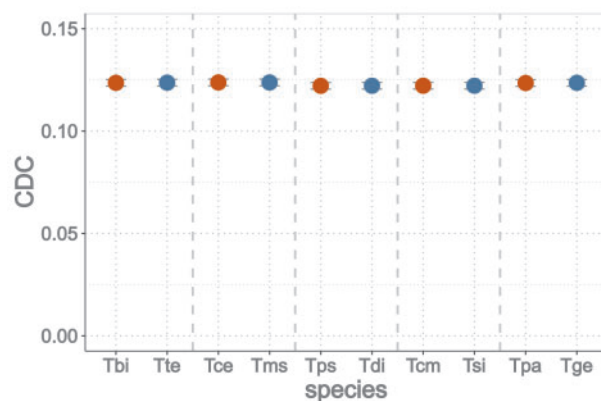
ANOVA; supplementary fig. 6, Supplementary Material online).

We tested if the difference in dS between sexuals and asexuals stems from different effectiveness of selection on codon usage bias using the Codon Deviation Coefficient (CDC) as a metric (Zhang et al. 2012). This metric calculates the deviation from expected codon usage bias and accounts for per-gene background nucleotide composition, thus allowing for cross-species comparisons. A lower CDC value would indicate less effective selection on codon usage bias (see Materials and Methods), but we did not detect any differences in CDC value between reproductive modes (gene effect $P < 0.001$, species pair effect $P < 0.001$, reproductive mode effect $P = 0.595$, interaction between pair and mode $P = 0.991$; permutation ANOVA; fig. 7). Thus, elevated substitution rates at synonymous sites in asexuals are unlikely to stem from less effective selection on codon usage bias in asexuals than sexuals.

Evaluation of the Consequences of Asexuality Over Time

Given the age variation of the asexual *Timema* species (fig. 2), possible consequences of asexuality can be studied over a range of recent to old asexual organisms. For all of the measured effects of asexuality, lineage age should a priori only impact the consequences of arrested gBGC, because the return to equilibrium base composition should happen gradually over time. Indeed, the effect of arrested gBGC on GC composition was stronger for older than younger asexual *Timema* species (Pearson's product-moment correlation; $t = 3.29$, cor 0.88, $P = 0.046$, supplementary fig. 4, Supplementary Material online). There was no age effect for dN/dS (Pearson's product-moment correlation; $t = -1.03$, cor -0.51, $P = 0.686$), pN/pS (Pearson's product-moment correlation; $t = -1.23$, cor -0.13, $P = 0.829$) and dN estimates (supplementary fig. 5, Supplementary Material online), but surprisingly, dS increased with age of asexuality (Pearson's

product-moment correlation; $t = 5.83$, cor 0.96, $P = 0.010$, supplementary fig. 6, Supplementary Material online). However, any correlation (or absence thereof) should be interpreted carefully, as there are only five data points and correlates might be driven by only one old "outlier" asexual (*T. genevieve*).

Discussion

The absence of recombination is predicted to be associated with reduced GC content (because gBGC does not occur), reduced effectiveness of selection on short and long evolutionary timescales, and affect levels of neutral genetic variation segregating in populations (Muller 1964; Hill and Robertson 1966; Felsenstein 1974; Marais 2003; Keightley and Otto 2006). Using multiple, independently evolved asexual lineages, we provide for the first time empirical evidence for all of these predicted consequences, in full agreement with theory.

gBGC influences the base composition in sexual *Timema* stick insects, but is unlikely to occur (or at least is strongly reduced) in asexuals. Consistent with theory, the effect of arrested gBGC on base composition is stronger in older asexual *Timema* species compared with their sister species (supplementary fig. 4, Supplementary Material online). These findings represent the first documented case of arrested gBGC in asexual eukaryotes. Although there is extensive variation in GC content and in the strength of gBGC among sexual *Timema* species, we consistently detect lower levels of gBGC in the asexual sister lineages. Given this lineage-specific variation, sister-pair comparisons are essential to detect arrested gBGC in asexual species. This might be why all evidence for arrested gBGC in the absence of recombination thus far stems from nonrecombining parts of genomes and selfers, where appropriate comparisons are available (Glémin et al. 2006; Spencer et al. 2006; Duret and Galtier 2009; Glémin and Galtier 2012).

Asexual stick insects exhibit lower levels of segregating genetic variation within populations as compared with sexuals. This finding suggests that asexual *Timema* species are more homozygous than their sexual counterparts, as intra-individual heterozygosity should otherwise generate high genetic variation at individual loci, at least in the old asexual species such as *T. genevievae*. Low genetic diversity in asexuals is also expected because of background selection. Although asexual *Timema* populations are far from being genetically uniform, there are 8-fold more transcripts containing polymorphisms in sexual populations and within these transcripts, variants are twice as frequent (fig. 4). Thus far, comparative studies investigating segregating polymorphisms in sexual and asexual populations are complicated by confounding effects like hybridization events and/or found no differences of the amount of genetic variation between reproductive modes, likely because of young asexual lineage ages (Hollister et al. 2015; Ament-Velásquez et al. 2016). Our findings of reduced levels of standing genetic variation in asexual compared with sexual *Timema* sister species is

direct evidence that reproductive mode can strongly influence genetic diversity in natural populations.

Purifying selection is less effective in asexual than sexual *Timema*, resulting in more deleterious variants segregating in populations and a higher rate of deleterious mutation accumulation in the long term. The ratio of nonsynonymous to synonymous segregating variants is approximately two times higher in asexual than sexual populations (range 1.1–2.4, depending on the species pair). In spite of these differences, there is currently no evidence that the average fitness of asexual *Timema* females is lower than the fitness of sexual females, but fitness differences may require more time to build up and become measurable. Indeed some of the deleterious variants become fixed over time as a consequence of drift, resulting in a faster rate of deleterious mutation accumulation in asexual than sexual *Timema* species, with ω (dN/dS) on average being 1.2 times higher in asexuals than sexuals (fig. 6).

It is important to note however that ω ratios are highly variable among genes (variance = 0.043, range 0.001–0.998), with sometimes opposite effects ($\omega_{sex} > \omega_{asex}$) to the general pattern. Thus, given this variance, inferring overall genomic patterns from studies that are based on small gene numbers might result in arbitrary and wrong conclusions. Indeed most previous studies on mutation accumulation in asexual metazoans, including a previous study in *Timema*, are based on few genes (Johnson and Howard 2007; Neiman et al. 2010; Henry et al. 2012; Ollivier et al. 2012; Hartfield 2016). To date, only three studies exist that use genome-scale analyses to address mutation accumulation in asexuals, with mixed results (Hollister et al. 2015; Ament-Velásquez et al. 2016; Brandt et al. 2017). Investigated asexual lineages were either very young, and no difference in the effectiveness of purifying selection between reproductive modes was found (Ament-Velásquez et al. 2016). Or very old (tens of millions of years) and accumulating deleterious mutations at lower rates than sexual species, opposite to theoretical predictions (Brandt et al. 2017). In addition to the present study, there is thus far only one study in selfing (but functionally asexual) plants that converged with theoretical predictions of more effective purifying selection in sexual taxa compared with asexuals (Hollister et al. 2015).

Our results also suggest that the reduced effectiveness of selection in asexuals may result in an increase of transcriptional errors (or transcriptional noise). Indeed, an interesting and unexpected finding is that asexual species harbor more variants of orthologous transcripts than sexuals. Analyses using genomic data suggest that these variants do not stem from duplicated genes in the genomes. Moreover, the majority of variants is unlikely to represent different haplotypes, given that they vary considerably in the length of the ORF and the extremely low levels of segregating polymorphism in populations of *Timema* asexuals. Additional studies are required to corroborate our interpretation of higher transcriptional error rates, reduced stability or more frequent posttranscriptional alteration of transcripts in asexuals.

There is evidence that selection can also act on synonymous sites (Lawrie et al. 2013), for example through altering

the frequencies by which different codons are used (Duret 2002; Hershberg and Petrov 2008). Interestingly, rates of synonymous substitutions are increased in asexual *Timema*, compared with sexuals. The difference between sexuals and asexuals further increases with the age of asexual lineages. This pattern could be indicative of less effective selection on codon usage bias in asexual than sexual *Timema*. However, while there is weak codon usage bias overall, we could not detect differences in the strength of the bias between sexual and asexual species. This is possibly because evolution of codon bias might be too slow such that differences are too subtle to be detectable in the relatively "young" *Timema* species. Another mechanism that could potentially influence rates of synonymous substitutions is gBGC. gBGC arrest in asexuals results in a sudden change in equilibrium GC3, which under certain conditions is expected to increase substitution rates (Bolívar et al. 2016). Moreover, increased synonymous substitution rates in asexuals could also be caused by increased mutation rates.

In conclusion, this is the first genome scale study on asexual evolution that documents consequences for gBGC, polymorphism levels and effectiveness of purifying selection. This study is particularly robust, because the *Timema* system allows us to utilize evolutionary replicates of independent sexual–asexual sister pairs with different asexual lineage ages. It remains unknown and a challenge for future studies to determine if the accumulation of deleterious mutations may eventually result in the extinction of asexual *Timema* lineages and thereby contribute to the long-term maintenance of sex in this system.

Materials and Methods

Taxon Sampling and Sequencing

Timema individuals from ten different species (five sexual and five asexual) were collected in California, USA in spring 2014. For details on the sampling locations see supplementary table 3, Supplementary Material online. Prior to RNA extraction, animals were fed with artificial medium for two days to avoid contamination with gut content and then frozen at -80 °C. Total RNA was extracted from whole bodies of a pool of three individuals from the same species and location. This was done by first freezing the individuals in liquid nitrogen, followed by addition of Trizol (Life Technologies) and mechanical bead crushing (Sigmund Lindner). The homogenized tissue was then treated with Chloroform and Ethanol and the aqueous layer transferred to RNeasy MinElute Columns (Qiagen). Following the RNeasy Mini Kit protocol, potential DNA in the sample was digested, RNA eluted into water and stored at -80 °C. RNA quantity and quality was measured using NanoDrop (Thermo Scientific) and Bioanalyzer (Agilent). RNA extracts were pooled and fragmented to 120 nt for strand-specific library preparation. Single-end sequencing with a read length of 100 bp was performed on a HiSeq2000 platform at the CIG (Centre of Integrative Genomics, Lausanne, Switzerland). CutAdapt was used to remove adapter sequences from the raw reads (Martin 2011). Reads longer than 80 bp were then quality trimmed

using trimmomatic v 0.36 (Bolger et al. 2014) (first clipping leading or trailing bases with a phred score of <10 from the read, before using a sliding window from the 5′ end to clip the read if four consecutive bases had an average phred score of <20). Following quality trimming any reads <80 bp in length were discarded. For assembly, reads found to contain adapter sequence were not used, but for mapping all trimmed reads were used.

Assembly and Assembly Cleaning

All available trimmed reads per species were pooled and used as input for assembly with Trinity with the addition of the following option: −min_kmer_cov 2 (v 2.2.0; Grabherr et al. 2011). Assembly often results in an unrealistically high number of transcripts. To remove likely erroneous contigs, we applied a minimum expression filter following Moghadam et al. (2013) and Harrison et al. (2015). We mapped the pooled reads for each sample separately back to the assembly with RSEM v1.2.20 using bowtie2 v2.2.4 (Li and Dewey 2011; Langmead and Salzberg 2012). Ambiguously mapping reads were assigned to the most likely transcript by RSEM. We discarded contigs shorter than 300 bp and with low coverage (reads per kilobase per million mapped, RPKM $\leq$ 2). After filtering, all assemblies had a comparable number of isotigs and N_{50} (supplementary table 1, Supplementary Material online). For subsequent analyses in this study, the longest isoform per gene/graph was selected. ORFs were identified from the filtered transcriptome set using transdecoder v2.1 running TransDecoder.LongOrfs and TransDecoder.Predict (with strand-specific option). In cases where multiple ORFs were predicted for a single transcript the longest ORF per transcript was selected (supplementary table 1, Supplementary Material online).

Annotation and Contamination Filtering

NCBI's BLAST client (v. 2.2.30+) was used to BLAST local versions of the nt (using blastN, default options except -task blastn, -max_target_seqs 10) and nr (using blastX, default options except, -max_target_seqs 10) databases (downloaded: 07/08/2016). BLAST hits were first filtered so that any hit with an *e*-value >0.0000001 was discarded. Contigs were assigned to a Domain if $\geq$50% of BLAST hits came from a domain. In the event of a tie, the taxa with the highest *e*-value was used as a tiebreaker. Contigs annotated as "noneukaryote" were discarded prior to submission to GenBank (the number of discarded transcripts ranged from 64 to 133 among transcriptomes).

Orthologs

To identify orthologous sequences shared by any combination of the ten species, OMA v1.0.6 (Altenhoff et al. 2013) was run on the identified ORF protein sequences with default parameters and fixed species tree (fig. 1). Further, orthologs were selected that were shared between all *Timema* species, yielding 3,010 orthologous groups (with one sequence per *Timema* species in each group). To identify the variants of these orthologs from each of the ten transcriptome assemblies, a custom script, provided by Clément Train, a member

of the OMA development team, was used (see Code availability).

Alignments

Alignments of the orthologs shared between sister species were generated to calculate divergence estimates. Alignments were also generated for each of the 3,010 ten-species orthologs for analyses of dN/dS and CDC. For all types of alignments, orthologous sequences were aligned on the protein level using M-coffee for a consensus alignment from multiple software (clustalw2_msa muscle_msa kalign_msa mafftgins_msa t_coffee_msa) implemented in t-coffee v11.00.8 (Wallace et al. 2006). In almost all cases, all methods converged to a similar alignment, as indicated by a quality score > 75. Resulting protein alignments were back-translated into nucleotide alignments using t-coffee and curated with Gblocks v0.91b (type = codons; minimum block length set to 4; Talavera and Castresana 2007) to remove gap-rich and poorly aligned regions.

Distance Estimates

Pairwise Jukes–Cantor distances of curated alignments from orthologous sequences between sister species were calculated using the distmat program from the EMBOSS suite v6.6.0.0 (Rice et al. 2000).

gBGC

Ancestral GC3 was estimated for all nodes of the *Timema* tree for the 3,010 orthologs using the NHML program (Galtier and Gouy 1998). This method makes use of a nonhomogeneous and nonstationary Markov model of DNA evolution to estimate branch-specific equilibrium GC-content in a maximum-likelihood framework (Romiguier et al. 2010; Figuet et al. 2014). The substitution mapping procedure was used to calculate the number of synonymous substitutions in the terminal branch of each *Timema* species, and categorized them into weak-to-strong (A/T → G/C), strong-to-weak (G/C → A/T) and GC-conservative (others; Dutheil et al. 2012; Romiguier et al. 2012). This consisted in first fitting the YN98 codon model, as implemented in the bppML program (Dutheil and Boussau 2008), to each alignment of protein-coding sequences. The estimated model parameters were then used to map substitutions on each branch of the tree in the empirical Bayesian framework. The total number of substitutions in a branch for a given category was obtained by summing substitution counts across genes.

Polymorphism (SNPs)

To identify polymorphism from SNP data, we mapped reads from a pool of three females from the same species and population back to the 3,010 ten-species orthologs (that contained only one "best" variant for each ortholog per orthologous group) with RSEM using bowtie2 with default parameters and –fragment-length-mean 200 –fragment-length-sd 100 (Li and Dewey 2011; Langmead and Salzberg 2012). Samtools v1.2 was used to sort the mapped reads and samtools mpileup to identify SNPs. Resulting polymorphic sites were filtered using VarScan v2.3.2 mpileup2snp with a 20-fold minimum coverage cutoff, 10% minor allele freq and a minimum average phred quality of 20 (following Hollister et al. 2015). To check for mapping bias generated by the presence of variants for certain orthologs, mapping quality and read count was extracted from the RSEM output table and bam files for both the ten-species ortholog set that included the best variant only and a set that included all variants.

Divergence ω (dN/dS) Analyses

Curated back-translated nucleotide alignments of the 3,010 ten-species orthologs were used as input for dN/dS ratio estimates. To estimate gene-specific ω (dN/dS) ratios, we first calculated per-gene branch-lengths using RAxML v8.2.8 (Stamatakis 2014) given a fixed, unrooted species tree and the GTRCATI model. Gene-specific branch-lengths, respective alignments and the fixed, unrooted species tree were then parsed as input for codeml, which is included in the paml v4.8 package (Yang 2007), using a custom script that also subsequently collected the output for each gene (see Code availability). We ran both a "free" model, allowing one ω value for every branch, and a "three-rate" model, allowing one rate for all asexual and one for all sexual terminal branches as well as one rate for internal (sexual) branches to separate mutations occurring at terminal branches from internal ones. Gene and branch-specific ω values were further analyzed with R (Team and Others 2013). We restricted analyses to transcripts with overall ω values $\omega < 1$ and dS $\neq$ 0, to capture only transcripts that are under purifying selection. Within this gene set, we removed 200 orthologs that had no ω variation between species, leaving 2,487 orthologs to compare. Here, we did not analyze positive selection on a $\omega > 1$ gene set, because this would require an entirely different approach that would be based on branch-site-specific models. Given that in the current study, we only used orthologs shared across the ten species, it seems unlikely that we could detect consistently adaptive signals.

Codon Usage Bias

To infer codon usage bias we calculated the effective number of codons (Enc) per gene using codonW v1.4 (Wright 1990; Peden 2005). Enc quantifies how strongly codon usage deviates from equal usage of all synonymous codons, taking values from 20 (each amino acid is encoded by only one codon exclusively) to 61 (equal use of all possible alternative codons). Following, Enc values below 61 indicate codon usage bias.

We estimated signatures of relaxed selection on synonymous sites using the CDC (Zhang et al. 2012). CDC estimates the expected codon usage from observed GC and purine contents and calculates the deviation from observed codon usage. CDC ranges from 0 (no detectable ["relaxed"] selection on codon usage) to 1 (effective selection on codon usage). We calculated CDC using the same alignments as used for ω analyses.

Polymorphism (pN/pS) Analyses

We used the SNP data (details above under "polymorphism") to identify nonsynonymous and synonymous segregating

polymorphism in the ten *Timema* populations. For this, we identified the fold-degenerate position of resulting filtered nucleotide variants using a custom script following Li et al. (1985) from which pN, pS and (pN/pS) was calculated per gene (see Code availability).

Statistics

For statistical tests on dN/dS, dS, Enc and CDC estimates we used a permutation ANOVA (Manly 1997) with 5,000 bootstrap replicates (script available in Henry et al. 2012). For other analyses, statistical methods are given in the text.

Data Availability

Raw reads are deposited in SRA under accession codes SRR5748941–SRR5749000. The transcriptome assemblies are deposited at DDBJ/EMBL/GenBank under the BioProject PRJNA380865 with the following accession codes: GFPP00000000, GFPR00000000, GFPS00000000, GFPT00000000, GFPU00000000, GFPV00000000, GFPW00000000, GFPX00000000, GFPY00000000, and GFPZ00000000. Data will be publically released upon acceptance.

Code Availability

The custom scripts used in this study are deposited at https://github.com/jensbast/TimemaConsOfAsex/, last accessed March 16, 2018.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Author Contributions

T.S., J.B. conceived and designed the study. T.S., K.M.J. collected samples, Z.D. performed wet lab work, J.B., D.P., K.S.J., P.T.V., E.F., A.B. performed data analyses, J.B., T.S., N.G. interpreted data, J.B., T.S. wrote the paper with input from all authors.

References

Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS ONE*. 8(1):e53786.

Ament-Velásquez SL, Figuet E, Ballenghien M, Zattara EE, Norenburg JL, Fernández-Álvarez FA, Bierne J, Bierne N, Galtier N. 2016. Population genomics of sexual and asexual lineages in fissiparous ribbon worms (*Lineus*, Nemertea): hybridization, polyploidy and the Meselson effect. *Mol Ecol*. 25(14):3356–3369.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 14(2):113–124.

Bachtrog D, Charlesworth B. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416(6878): 323–326.

Balloux F, Lehmann L, de Meeûs T. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164(4):1635–1644.

Birky CW. 1996. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144(1):427–437.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Bolívar P, Mugal CF, Nater A, Ellegren H. 2016. Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill–Robertson interference, in an avian system. *Mol Biol Evol*. 33(1):216–227.

Brandt A, Schaefer I, Glanz J, Schwander T, Maraun M, Scheu S, Bast J. 2017. Effective purifying selection in ancient asexual oribatid mites. *Nat Commun*. 8(1):873.

Castel SE, Martienssen RA. 2013. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet*. 14(2):100–112.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.

Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev*. 12(6):640–649.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet*. 10:285–311.

Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol Biol*. 8:255.

Dutheil JY, Galtier N, Romiguier J, Douzery EJP, Ranwez V, Boussau B. 2012. Efficient selection of branch-specific models of sequence evolution. *Mol Biol Evol*. 29(7):1861–1874.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78(2):737–756.

Figuet E, Ballenghien M, Romiguier J, Galtier N. 2014. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biol Evol*. 7:240–250.

Galtier N, Gouy M. 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol Biol Evol*. 15(7):871–879.

Galtier N, Roux C, Rousselle M, Romiguier J, Figuet E, Glémin S, Bierne N, Duret L. 2018. Codon usage bias in animals: disentangling the effects of natural selection, effective population size and GC-biased gene conversion. *Mol Biol Evol [Internet]*, doi:10.1093/molbev/msy015.

Glémin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L. 2015. Quantification of GC-biased gene conversion in the human genome. *Genome Res*. 25(8):1215–1228.

Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc Biol Sci*. 273(1604):3011–3019.

Glémin S, Galtier N. 2012. Genome evolution in outcrossing versus selfing versus asexual species. In: Anisimova M, editor. Evolutionary genomics: statistical and computational methods, Volume 1. Vol. 855. Methods in Molecular Biology. Totowa (NJ): Springer Science + Business Media. p. 311–335.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci U S A*. 112(14):4393–4398.

Hartfield M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. *J Evol Biol*. 29(1):5–22.

Henry L, Schwander T, Crespi BJ. 2012. Deleterious mutation accumulation in asexual *Timema* stick insects. *Mol Biol Evol*. 29(1):401–408.

Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet.* 42:287–299.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res.* 8(3):269–294.

Hollister JD, Greiner S, Wang W, Wang J, Zhang Y, Wong GK-S, Wright SI, Johnson MTJ. 2015. Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Mol Biol Evol.* 32(4):896–905.

Johnson SG, Howard RS. 2007. Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. *Evolution* 61(11):2728–2735.

Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* 443(7107):89–92.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–360.

Law JH, Crespi BJ. 2002. The evolution of geographic parthenogenesis in *Timema* walking-sticks. *Mol Ecol.* 11(8):1471–1489.

Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* 9(5):e1003527.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 12:323.

Li WH, Wu CI, Luo CC. 1985. 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(2):150–174.

Manly BFJ. 1997. Randomization, bootstrap and Monte Carlo methods in biology. London: Chapman and Hall.

Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19(6):330–338.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 17(1):10–12.

Moghadam HK, Harrison PW, Zachar G, Székely T, Mank JE. 2013. The plover neurotranscriptome assembly: transcriptomic analysis in an ecological model species without a reference genome. *Mol Ecol Resour.* 13(4):696–705.

Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination landscape and demography to genomic base composition. *Bioessays* 37(12):1317–1326.

Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat Res.* 1(1):2–9.

Neiman M, Hehman G, Miller JT, Logsdon JM, Taylor DR. 2010. Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Mol Biol Evol.* 27(4):954–963.

Ollivier M, Gabaldón T, Poulain J, Gavory F, Leterme N, Gauthier J-P, Legeai F, Tagu D, Simon JC, Rispe C. 2012. Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that differ by reproductive mode. *Genome Biol Evol.* 4(2):155–167.

Peden J. 2005. CodonW. Available from: http://codonw.sourceforge.net/, last accessed March 16, 2018.

Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biol Evol.* 4(7):675–682.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16(6):276–277.

Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR, et al. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nat Ecol Evol.* 1(4):0082. s41559–017–0082.

Romiguier J, Figuet E, Galtier N, Douzery EJP, Boussau B, Dutheil JY, Ranwez V. 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS ONE.* 7(3):e33852.

Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res.* 20(8):1001–1009.

Schwander T, Crespi BJ. 2009. Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in *Timema* stick insects. *Evolution* 63(1):84–103.

Schwander T, Henry L, Crespi BJ. 2011. Molecular evidence for ancient asexuality in *Timema* stick insects. *Curr Biol.* 21(13):1129–1134.

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol.* 24(7):R288–R294.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942):825–837.

Spencer CCA, Deloukas P, Hunt S, Mullikin J, Myers S, Silverman B, Donnelly P, Bentley D, McVean G. 2006. The influence of recombination on human genetic diversity. *PLoS Genet.* 2(9):e148.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577.

Talbert PB, Henikoff S. 2010. Centromeres convert but don't cross. *PLoS Biol.* 8(3):e1000326.

van den Bosch M, Lohman PHM, Pastink A. 2002. DNA double-strand break repair by homologous recombination. *Biol Chem.* 383(6):873–892.

Wallace IM, O'Sullivan O, Higgins DG, Notredame C. 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* 34(6):1692–1699.

Webster MT, Hurst LD. 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. *Trends Genet.* 28(3):101–109.

Wright F. 1990. The "effective number of codons" used in a gene. *Gene* 87(1):23–29.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.

Zhang Z, Li J, Cui P, Ding F, Li A, Townsend JP, Yu J. 2012. Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance. *BMC Bioinformatics.* 13:43.