

# Convergent consequences of parthenogenesis on stick insect genomes

Kamil S. Jaron<sup>1, 2, 3, a</sup>, Darren J. Parker<sup>1, 2, a</sup>, Yoann Anselmetti<sup>4, 5, a</sup>, Patrick Tran Van<sup>1, 2</sup>, Jens Bast<sup>1, 6</sup>, Zoé Dumas<sup>1</sup>, Emeric Figuet<sup>4</sup>, Clémentine M. François<sup>4, 7</sup>, Keith Hayward<sup>1, 2</sup>, Victor Rossier<sup>1, 2</sup>, Paul Simion<sup>4, 8</sup>, Marc Robinson-Rechavi<sup>1, 2, b</sup>, Nicolas Galtier<sup>4, b</sup>, Tanja Schwander<sup>1, b</sup>

<sup>1</sup>Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland;

<sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland;

<sup>3</sup>Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL

<sup>4</sup>ISEM - Institut des Sciences de l'Evolution, Montpellier, France;

<sup>5</sup>Current address: CoBIUS lab, Department of Computer Science, University of Sherbrooke, Sherbrooke, Canada;

<sup>6</sup>Current address: Institute for Zoology, University of Cologne, Köln, Germany;

<sup>7</sup>Current address: Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France;

<sup>8</sup>Current address: Université de Namur, LEGE, URBE, Namur, 5000, Belgium

<sup>a</sup> shared first authorship

<sup>b</sup> equal contribution

Correspondence should be addressed to [DarrenJames.Parker@unil.ch](mailto:DarrenJames.Parker@unil.ch), [Kamil.Jaron@ed.ac.uk](mailto:Kamil.Jaron@ed.ac.uk), or [Tanja.Schwander@unil.ch](mailto:Tanja.Schwander@unil.ch)

## 30 Abstract

31 The shift from sexual reproduction to parthenogenesis has occurred repeatedly in  
 32 animals, but how the loss of sex affects genome evolution remains poorly  
 33 understood. We generated *de novo* reference genomes for five independently  
 34 evolved parthenogenetic species in the stick insect genus *Timema* and their closest  
 35 sexual relatives. Using these references in combination with population genomic  
 36 data, we show that parthenogenesis results in an extreme reduction of  
 37 heterozygosity, and often leads to genetically uniform populations. We also find  
 38 evidence for less effective positive selection in parthenogenetic species, supporting  
 39 the view that sex is ubiquitous in natural populations because it facilitates fast rates  
 40 of adaptation. Contrary to studies of non-recombining genome portions in sexual  
 41 species, genomes of parthenogenetic species do not accumulate transposable  
 42 elements (TEs), likely because successful parthenogens derive from sexual  
 43 ancestors with inactive TEs. Because we are able to conduct replicated comparisons  
 44 across five species pairs, our study reveals, for the first time, how animal genomes  
 45 evolve in the absence of sex in natural populations, providing empirical support for  
 46 the negative consequences of parthenogenetic reproduction as predicted by theory.

47

## 48 Introduction

49 Sex: What is it good for? The reason why most eukaryotes take a complicated  
 50 detour to reproduction, when more straightforward options are available, remains a  
 51 central and largely unanswered question in evolutionary biology (1, 2). Animal  
 52 species in which parthenogenetic reproduction is the sole form of replication typically  
 53 occur at the tips of phylogenies and only a few of them have succeeded as well as  
 54 their sexually reproducing relatives (3). In other words, most parthenogenetic  
 55 lineages may eventually be destined for extinction. These incipient evolutionary  
 56 failures, however, are invaluable as by understanding their fate something may be  
 57 learned about the adaptive value of sex.

58

59 Parthenogenesis is thought to be favored in the short term because it generates a  
60 transmission advantage (4, 5), as well as the advantage of assured reproduction  
61 when mates are scarce (6, 7). The short-term benefits of parthenogenesis, however,  
62 are believed to come along with long-term costs. For example, the physical linkage  
63 between loci it entails can generate interferences that decrease the efficacy of  
64 natural selection (e.g. (8–10), reviewed in (11)). This is expected to translate into  
65 reduced rates of adaptation and increased accumulation of mildly deleterious  
66 mutations, which may potentially drive the extinction of parthenogenetic lineages.

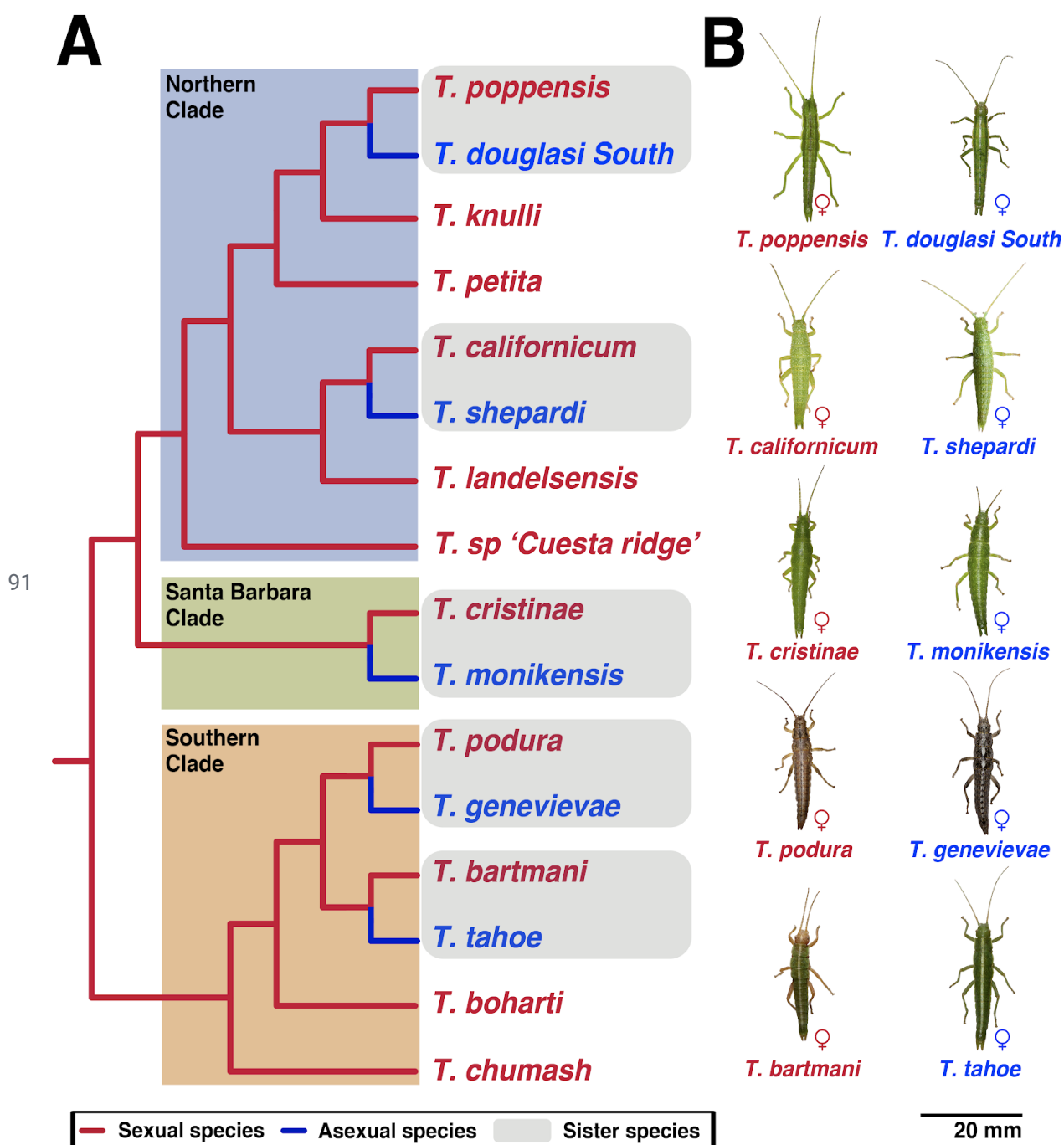
67

68 In addition to these predicted effects on adaptation and mutation accumulation,  
69 parthenogenesis is expected to drive major aspects of genome evolution. A classical  
70 prediction is that heterozygosity (i.e., intra-individual polymorphism) increases over  
71 time in the absence of recombination, as the two haploid genomes diverge  
72 independently of each other, generating the so-called “Meselson Effect” (12, 13).  
73 Parthenogenesis can also affect the dynamics of transposable elements (TEs),  
74 resulting in either increased or decreased genomic TE loads (14–16). Finally, some  
75 forms of parthenogenesis might facilitate the generation and maintenance of  
76 structural variants, which in sexuals are counter-selected due to the constraints of  
77 properly pairing homologous chromosomes during meiosis (17).

78

79 We tested these predictions by comparing the genomes of five independently  
80 derived parthenogenetic stick insect species in the genus *Timema* with their close  
81 sexual relatives (Figure 1). These replicate comparisons allowed us to solve the key  
82 problem in understanding the consequences of parthenogenesis for genome  
83 evolution: separating the consequences of parthenogenesis from lineage specific  
84 effects (17). *Timema* are wingless, plant-feeding insects endemic to western North  
85 America. Parthenogenetic species in this genus are diploid and of non-hybrid origin  
86 (18) and ecologically similar to their sexual relatives. Previous research, based on a  
87 small number of microsatellite markers, has suggested that oogenesis in  
88 parthenogenetic *Timema* is functionally mitotic, as no loss of heterozygosity between  
89 females and their offspring was detected (18).

90



**Figure 1.** Multiple, independent transitions from sexual to parthenogenetic reproduction are known in the genus *Timema* (19), each representing a biological replicate of parthenogenesis, and with a close sexual relative at hand for comparison

**A.** Phylogenetic relationships of *Timema* species (adapted from (19, 20)). **B.** Species sequenced in this study. Photos taken by © Bart Zijlstra - [www.bartzijlstra.com](http://www.bartzijlstra.com).

## 98 *De novo* genomes reveal extremely low heterozygosity in 99 parthenogenetic stick insects

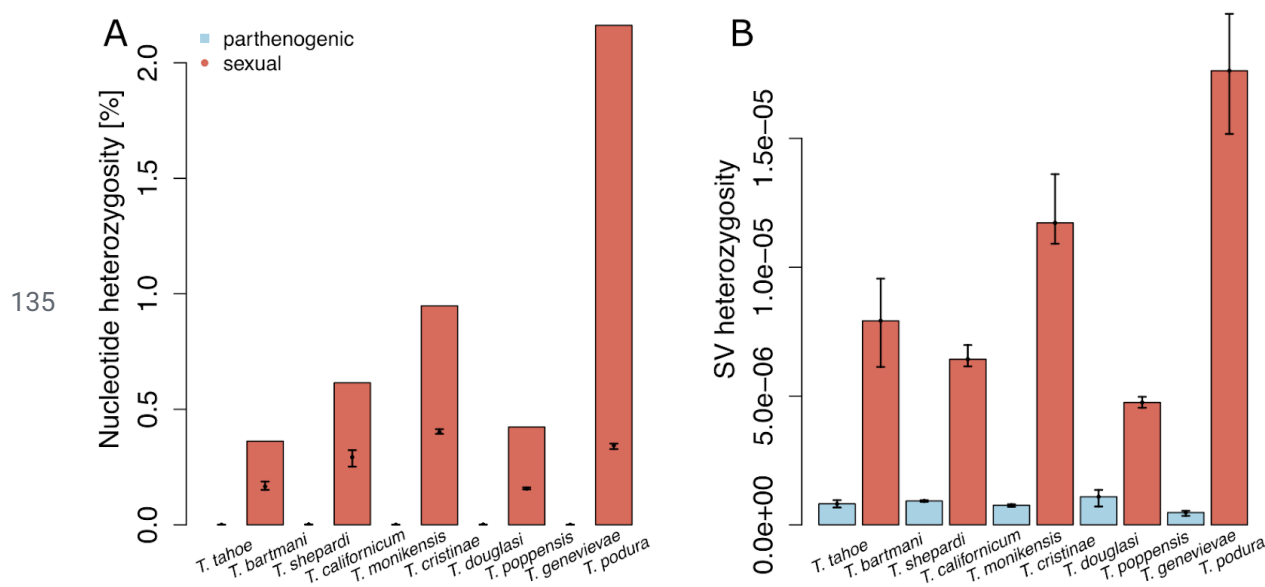
100 We generated ten *de novo* genomes of *Timema* stick insects, from five  
101 parthenogenetic and five sexual species (Figure 1, SM Tables 1, 2). Genomes were  
102 subjected to quality control, screened for contamination, and annotated (see  
103 Methods, SM text 1). The final reference genomes were largely haploid, spanned  
104 75-95% of the estimated genome size (1.38 Gbp (21)), and were sufficiently  
105 complete for downstream analyses, as shown by the count of single copy orthologs  
106 conserved across insects (96% of BUSCO genes (22) detected on average; SM  
107 Table 3). A phylogeny based on a conservative set of 3975 1:1 orthologous genes  
108 (SM Table 4) corroborated published phylogenies and molecular divergence  
109 estimates in the *Timema* genus (SM Figure 1). Finally, we identified 55 putative  
110 horizontal gene acquisitions from non-metazoans, and they all happened well before  
111 the evolution of parthenogenesis (SM text 2).

112

113 We estimated genome-wide nucleotide heterozygosity in each reference genome  
114 directly from sequencing reads, using a reference-free technique (genome profiling  
115 analysis (23)). These analyses revealed extreme heterozygosity differences between  
116 the sexual and parthenogenetic species. The five sexual *Timema* featured nucleotide  
117 heterozygosities within the range previously observed in other sexual species  
118 (Figure 2; (24, 25)). The heterozygosities in the parthenogenetic species were  
119 substantially lower, and in fact so low that reference-free analyses could not  
120 distinguish heterozygosity from sequencing error (SM text 3). We therefore  
121 compared heterozygosity between sexuals and parthenogens by calling SNPs in five  
122 re-sequenced individuals per species. This analysis corroborated the finding that  
123 parthenogens have extremely low ( $<10^{-5}$ ) heterozygosity, being at least 140 times  
124 lower than that found in their sexual sister species (permutation ANOVA,  
125 reproductive mode effect  $p = 0.0049$ ; Figure 2). Screening for structural variants  
126 (indels, tandem duplications, and inversions) in sexual and parthenogenetic  
127 individuals revealed the same pattern: extensive and variable heterozygosity in

128 sexual species and homozygosity in the parthenogens (Figure 2, SM text 3). Some  
 129 heterozygosity in *Timema* parthenogens could be present in genomic regions not  
 130 represented in our assemblies, such as centromeric and telomeric regions. These  
 131 regions however represent a relatively small fraction of the total genome, meaning  
 132 that for most of the genome at least, *Timema* parthenogens are either largely or  
 133 completely homozygous for all types of variants (SM text 3).

134



136 **Figure 2.** Extremely low heterozygosity in parthenogenetic *Timema* species for  
 137 different types of variants. **A.** Nucleotide heterozygosity represented by bars  
 138 indicates genome-wide estimates for the reference genomes (based on raw reads,  
 139 see Methods), heterozygosity based on SNP calls in re-sequenced individuals is  
 140 indicated by points and represents a conservative estimation of heterozygosity in the  
 141 assembled genome portions (with error bars indicating the range of estimates across  
 142 individuals) **B.** Heterozygous structural variants (SVs, reported as number of  
 143 heterozygous SVs / number of callable sites) in re-sequenced individuals (with error  
 144 bars indicating the range of estimates across individuals). Note that even though  
 145 heterozygous SNPs and SVs were called using stringent parameters, it is likely that  
 146 a large portion are false positives in parthenogenetic *Timema* (see SM text 3).

147

148 The unexpected finding of extremely low heterozygosity in *Timema* parthenogens  
 149 raises the question of when and how heterozygosity was lost. For example, the bulk

150 of heterozygosity could have been lost during the transition from sexual reproduction  
 151 to parthenogenesis (26). Alternatively, heterozygosity loss could be a continuous and  
 152 ongoing process in the parthenogenetic lineages. To distinguish these options, we  
 153 investigated the origin of the genetic variation present among different homozygous  
 154 genotypes in each parthenogenetic species. We found that only 6-19% of the SNPs  
 155 called in a parthenogen are at positions that are also polymorphic in the sexual  
 156 relative (SM Table 5). This means that most of the variation in parthenogens likely  
 157 results from mutations that appeared after the split from the sexual lineage. This  
 158 implies that heterozygosity generated through new mutations is lost continuously in  
 159 parthenogens, and was not suddenly lost at the inception of parthenogenesis. The  
 160 most likely explanation for these findings is that parthenogenetic *Timema* are, in fact,  
 161 not functionally mitotic but automictic. Automictic parthenogenesis frequently  
 162 involves recombination and segregation, and can lead to homozygosity in most or all  
 163 of the genome (27, 28). Although automixis can allow for the purging of  
 164 heterozygous deleterious mutations (29), the classical predictions for the long-term  
 165 costs of asexuality extend to automictic parthenogens because, as for obligate  
 166 selfers, linkage among genes is still much stronger than in classical sexual species  
 167 (30). This is especially the case in largely homozygous parthenogens, where  
 168 recombination and segregation, even if mechanistically present, have no effect on  
 169 genotype diversities.

170

171 Functional mitosis in *Timema* was previously inferred from the inheritance of  
 172 heterozygous microsatellite genotypes between females and their offspring (18), a  
 173 technique widely used in non-model organisms with no cytological data available  
 174 (e.g., (31, 32)). The most likely reconciliation of these contrasting results is that  
 175 heterozygosity is maintained in only a small portion of the genome, for example the  
 176 centromeres or telomeres, or between paralogs. Consistent with this idea, we were  
 177 unable to locate several of the microsatellite-containing regions in even the best  
 178 *Timema* genome assemblies (SM text 4), suggesting that these regions are not  
 179 present in our assemblies due to the inherent difficulty of assembling repetitive  
 180 genome regions from short read data (33).

181



# Extensive variation in genotype diversity between parthenogenetic populations

Parthenogenesis and sexual reproduction are expected to drive strikingly different distributions of polymorphisms in genomes and populations. Different regions within genomes experience different types of selection with sometimes opposite effects on the levels of polymorphisms within populations, such as purifying versus balancing selection (34). The increased linkage among genes in parthenogenetic as compared to sexual species is expected to homogenize diversity levels across different genome regions. Furthermore, recurrent sweeps of specific genotypes in parthenogenetic populations can lead to extremely low genetic diversity and even to the fixation of a single genotype, while sweeps in sexual populations typically reduce diversity only in specific genome regions.

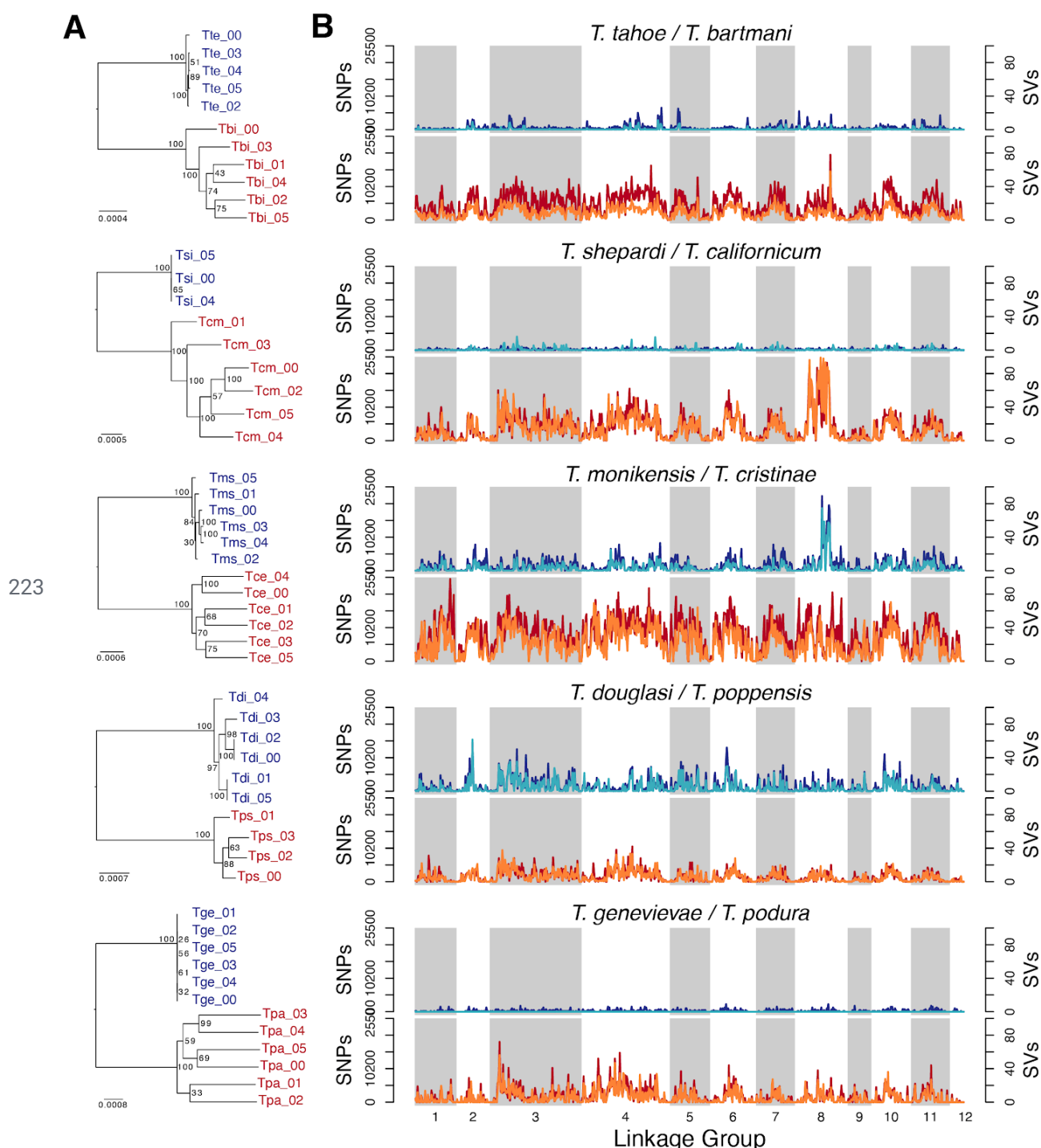
To address these aspects in the genomes of sexual and parthenogenetic *Timema* species, we mapped population-level variation for the SNPs and SVs inferred above to our species-specific reference genomes. We then anchored our reference genome scaffolds to the 12 autosomal linkage groups of a previously published assembly of the sexual species *T. cristinae* (v1.3 from (35), SM text 5). This revealed that different types of polymorphisms (SNPs and SVs) tended to co-occur across the genomes in all species, independently of reproductive mode (Figure 3).

The focal population for three of the five parthenogenetic species (*T. genevieveae*, *T. tahoe* and *T. shepardii*) consisted largely of a single genotype with only minor variation among individuals. By contrast, genotype diversity was considerable in *T. monikensis* and *T. douglasi* (Figure 3A). In the former species, there was further a conspicuous diversity peak on LG8, supporting the idea that parthenogenesis is automictic in *Timema*. Indeed, under complete linkage (functionally mitotic parthenogenesis), putative effects of selection on this LG would be expected to propagate to the whole genome. Independently of local diversity peaks, overall



212 diversity levels in *T. monikensis* and *T. douglasi* were comparable to the diversities in  
 213 populations of some of the sexual *Timema* species (Figure 3A). Different  
 214 mechanisms could contribute to such unexpected diversities in parthenogenetic  
 215 *Timema*, including the presence of lineages that derived independently from their  
 216 sexual ancestor, or rare sex. While a single transition to parthenogenesis is believed  
 217 to have occurred in *T. monikensis*, the nominal species *T. douglasi* is polyphyletic  
 218 and known to consist of independently derived clonal lineages. These lineages have  
 219 broadly different geographic distributions but can overlap locally (19). Identifying the  
 220 causes of genotypic variation in these species, including the possibility of rare sex,  
 221 requires further investigation and is a challenge for future studies.

222



**Figure 3.** Population polymorphism levels in parthenogenetic (blue) and sexual (red) *Timema* species. **A.** Phylogenies based on 1:1 orthologous genes reflect the different levels of genotype diversities in parthenogenetic *Timema* species **B.** Distribution of structural variants (SVs; dark blue and red) and SNPs (light blue and orange) along the genome. Scaffolds from the ten *de novo* genomes are anchored on autosomal linkage groups from the sexual species *T. cristinae* (SM text 5).

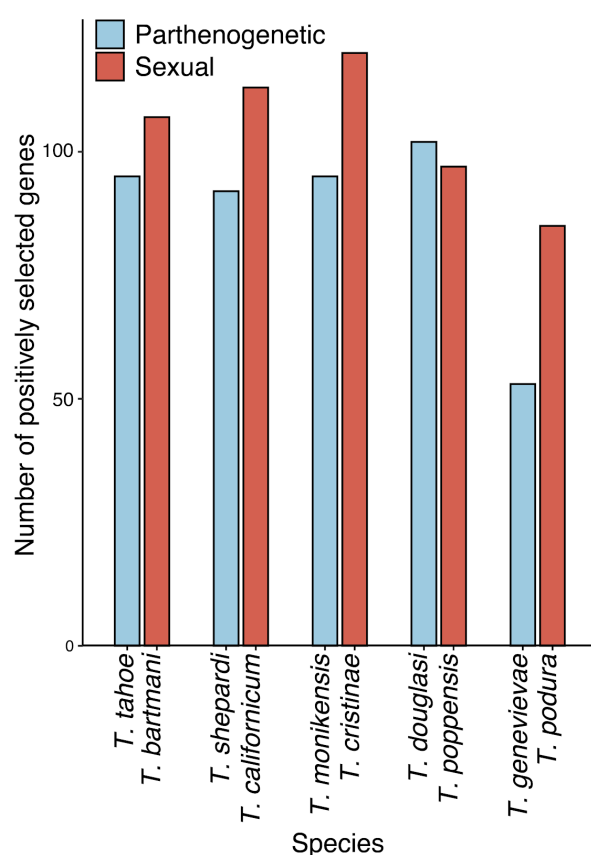
Independently of the mechanisms underlying polymorphism in the parthenogenetic species *T. monikensis*, the polymorphism peak on LG8 is striking (Figure 3B). This peak occurs in a region previously shown to determine color morph (green, green-striped, or brownish (“melanistic”)) in the sexual sister species of *T. monikensis*, *T. cristinae* (35). Our focal *T. monikensis* population features four discrete color morphs (green, dark brown, yellow, and beige), suggesting that additional color morphs may be regulated by the region identified in *T. cristinae*. We also found a peak in polymorphism on LG8, spanning over approximately two-thirds of LG8, in the sexual species *T. californicum*, which features a different panel of color morphs than *T. cristinae* (36). Interestingly, this diversity peak in *T. californicum* was generated by the presence of two divergent haplotypes (approximately 24Mbp long), with grey individuals homozygous for one haplotype and green individuals heterozygous or homozygous for the alternative haplotype (SM text 6). Note that the grey color morph is not known in the monomorphic green parthenogenetic sister of *T. californicum* (*T. shepardi*), and we therefore do not expect the same pattern of polymorphism on LG8 in this species.

## Faster rate of adaptive evolution in sexual than parthenogenetic species

We have shown previously that parthenogenetic *Timema* species accumulate deleterious mutations faster than sexual species (37, 38), a pattern also reported in other parthenogenetic taxa (reviewed in (17, 39)). This is expected given that linkage among loci in parthenogens prevents selection from acting individually on each locus, which generates different forms of selective interference (9, 10, 40). In addition to facilitating the accumulation of deleterious mutations, selective interference among loci in parthenogens should also constrain the efficiency of positive selection. While there is accumulating evidence for this process in experimental evolution studies (e.g., (41–43)), its impact on natural populations remains unclear (17, 39). To compare the efficiency of positive selection in sexual and parthenogenetic *Timema*, we used a branch-site model on the gene trees ((44), Methods). We compared the terminal branches leading to sexual or parthenogenetic

species in one-to-one orthologous genes identified in at least three species pairs (SM Table 4), using a threshold of  $q < 0.05$  to classify which terminal branches show evidence of positive selection.

We found a greater number of positively selected genes in sexual than parthenogenetic species (Figure 4, binomial GLMM  $p = 0.005$ ). In addition, we also examined if there was more evidence for positive selection in sexual species in a threshold-free way by comparing the likelihood ratio test statistic between parthenogenetic and sexual species (as in (45, 46)). This confirmed that the evidence for positive selection was stronger for sexual species (permutation glm  $p = 0.011$ ).



**Figure 4.** Number of genes showing evidence for positive selection in each species. In addition to reproductive mode, species pair also had a significant influence on the number of positively selected branches (binomial GLMM  $p = 0.015$ ). There was no significant interaction between species pair and reproductive mode ( $p = 0.197$ ). Note, the difference between reproductive modes is robust to a more stringent cutoff ( $q < 0.01$  instead of 0.05, SM Figure 2).

278 The positively selected genes we identified are most likely associated with  
 279 species-specific adaptations. Few of them were shared between species, with  
 280 overlap between species not greater than expected by chance (SM Figure 3, FDR <  
 281 0.4), and there was little enrichment of functional processes in positively selected  
 282 genes (0-19 GO terms per species, SM Table 8). Interestingly, most of the significant  
 283 GO terms were associated with positively selected genes in parthenogenetic  
 284 *Timema* (SM Table 8), likely because a much smaller proportion of positively  
 285 selected genes in sexual species had annotations (SM Figure 4). We speculate that  
 286 positively selected genes in sexuals could often be involved in sexual selection and  
 287 species recognition. Indeed, genes associated with processes such as pheromone  
 288 production and reception often evolve very fast, which makes them difficult to  
 289 annotate through homology-based inference (47). For the parthenogenetic species,  
 290 although some terms could be associated with their mode of reproduction (e.g.  
 291 GO:0033206 meiotic cytokinesis in *T. douglasi*), most are not clearly linked to a  
 292 parthenogenetic life cycle.

## 293 Transposable element loads are similar between species with 294 sexual and parthenogenetic reproduction

295

296 Upon the loss of sexual reproduction, transposable element (TE) dynamics are  
 297 expected to change (14, 16, 48). How these changes affect genome-wide TE loads  
 298 is however unclear as sex can facilitate both the spread and the elimination of TEs  
 299 (17). In parthenogens, TE load might initially increase as a result of weaker purifying  
 300 selection, a pattern well illustrated by the accumulation of TEs in non-recombining  
 301 parts of sex chromosomes and other supergenes (49, 50). However, TE loads in  
 302 parthenogens are expected to decrease over time via at least two non-mutually  
 303 exclusive mechanisms. First, TEs are expected to evolve lower activity over time as  
 304 their evolutionary interests are aligned with their hosts (14, 48). Second, TE copies  
 305 that were purged via excision can re-colonize a sexual but not a parthenogenetic  
 306 genomic background (15, 16). Finally, it is important to note that the predicted effects  
 307 of reproductive mode on TE loads require some amount of TE activity (active

transposition or excision) to occur. Without such activity, TE content does not vary among individuals and can therefore not change over time.

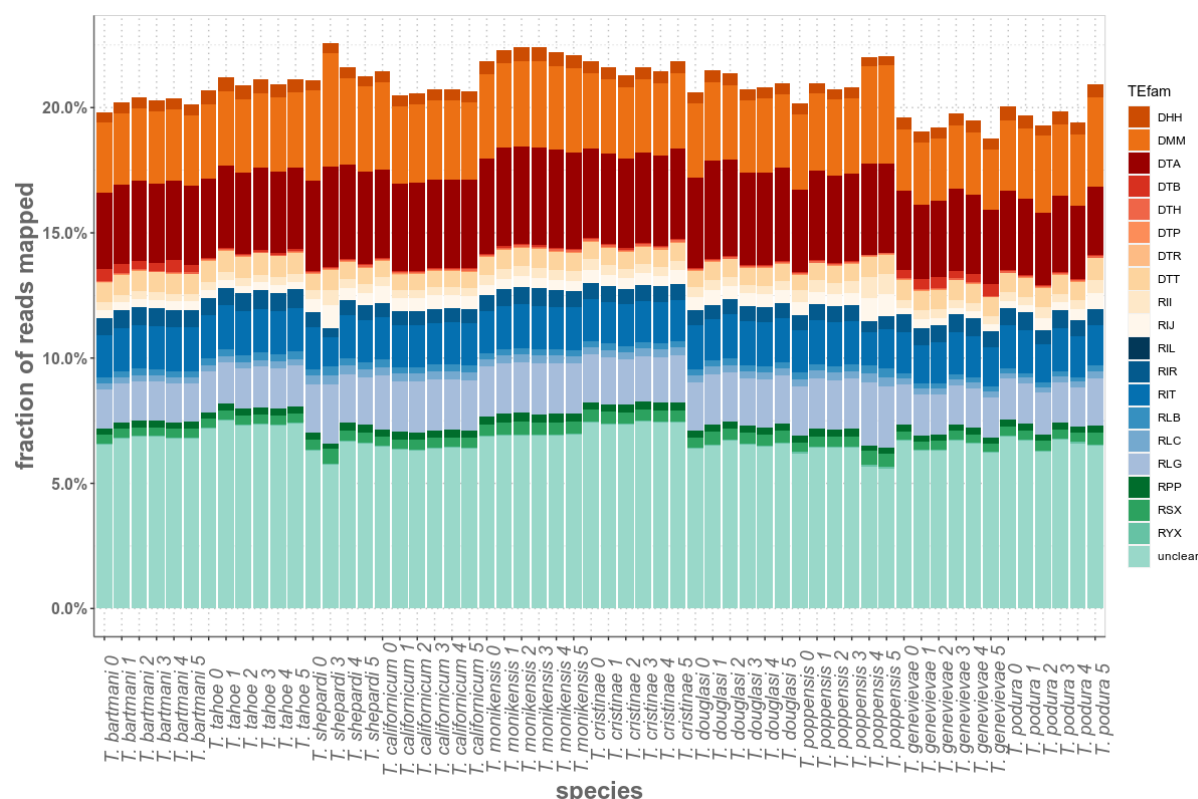
310

We generated a *Timema* genus-level TE library by merging *de novo* TE libraries generated separately for each of the ten *Timema* species. We then quantified TE loads in each *Timema* genome by mapping reads to this merged library (see **Methods**). The overall TE content was very similar in all ten species (20 - 23.6%), with significant differences in abundance of TE superfamilies between species groups but no significant effect of reproductive mode ( $p=0.43$ ; Figure 5; SM Figure 5).

318

No difference in TE load between sexual and parthenogenetic *Timema* would be expected if TEs were already well controlled in their ancestor, without any subsequent TE activity. Consistent with this idea, we find very little evidence for ongoing TE activity in the genus. The oldest node in our *Timema* phylogeny has an age estimate of 30 Mya (20) but the TE contents of the two clades separating at this node have only diverged by 1.3%, suggesting that TEs remained largely silent during the evolution of the genus. Inactive TEs may facilitate the persistence of incipient parthenogenetic strains (17) and thus help to explain the high frequency of established parthenogenetic species in *Timema*.

328



**Figure 5.** Total TE abundance in the ten *Timema* species. TE abundance is expressed as the fraction of reads that map to a genus-level TE library. TE families are named following the Wicker classification (51). The first character corresponds to the TE class (Class I are retrotransposons (R), Class II are DNA transposons (D)), the second character corresponds to the Order (e.g. LTR) and the third to the Superfamily (e.g. *Gypsy*); for example, RLG is a *Gypsy* retroelement. The character X indicates unknown classification at the superfamily level (because of fragmentation or lack of detectable homology).



## Conclusion

We present genomes of five independently derived parthenogenetic lineages of *Timema* stick insects, together with their five sexual sister species. This design with replicated species pairs allows us, for the first time, to disentangle consequences of parthenogenesis from species-specific effects. All parthenogenetic *Timema* species are largely or completely homozygous for both SNPs and SVs, and frequently feature lower levels of population polymorphism than their close sexual relatives. Low population polymorphism can exacerbate the effects of linkage for reducing the efficacy of selection, resulting in reduced rates of positive selection in parthenogenetic *Timema*, in addition to the accumulation of deleterious mutations previously documented (37). In spite of these negative genomic consequences, parthenogenesis is an unusually successful strategy in *Timema*. It evolved and persisted repeatedly in the genus, and parthenogenetic species often occur across large geographic areas. Because *Timema* are wingless and their populations subjected to frequent extinction-recolonization dynamics in their fire-prone Californian shrubland habitats, the genomic costs of parthenogenesis are likely offset by one of the most classical benefits of parthenogenesis: the ability to reproduce without a mate.

## Methods

### Sample collection and sequencing

For each of the ten species, the DNA for Illumina shotgun sequencing was derived from virgin adult females collected in 2015 from natural populations in California (SM Table 1). Extractions were done using the Qiagen Mag Attract de HMW DNA kit, following manufacturer indications. Five PCR-free libraries were generated for each reference genome (three 2x125bp paired end libraries with average insert sizes of respectively 350, 550 and 700bp, and two mate-pair libraries with 3000 and 5000bp insert sizes), one library (550bp insert size) was generated for each re-sequenced

individual. Libraries were prepared using the illumina TruSeq DNA PCR-Free or  
 Nextera Mate Pair Library Prep Kits, following manufacturer instructions, and  
 sequenced on the Illumina HiSeq 2500 system, using v4 chemistry and 2x 125 bp  
 reads at FASTERIS SA, Plan-les-Ouates, Switzerland.

## Genome assembly and annotation

The total coverage for the reference genomes (all libraries combined) ranged  
 between 37-45x (SM Table 2). Trimmed paired-end reads were assembled into  
 contigs using ABySS (52) and further scaffolded using paired-end and mate pairs  
 using BESST (53). Scaffolds identified as contaminants were filtered using Blobtools  
 (54). The assembly details can be found in supplementary materials (SM text 1).

Publically available RNA-seq libraries for *Timema* (37, 55, 56) were used as  
 expression evidence for annotation. Trimmed reads were assembled using Trinity  
 v2.5.1 (57) to produce reference-guided transcriptomes. The transcriptomes and  
 protein evidence were combined with *ab initio* gene finders to predict protein coding  
 genes using MAKER v2.31.8 (58). The annotation details can be found in the  
 supplementary materials (SM text 1).

## Orthologs

*Timema* orthologous groups (OGs) were inferred with the OrthoDB standalone  
 pipeline (v. 2.4.4) using default parameters (59). In short, genes are clustered with a  
 graph-based approach based on all best reciprocal hits between each pair of  
 genomes. The high level of fragmentation typical for Illumina-based genomes  
 constrains the ability to identify 1:1 orthologs across all ten *Timema* species. To  
 maximize the number of single copy OGs covering all ten *Timema* species,  
 transcriptomes were included during orthology inference. Thus, transcripts were  
 used to complete OGs in absence of a gene from the corresponding species. Using  
 this approach, 7157 single copy OGs covering at least three sexual-parthenogenetic  
 sister species pairs were obtained (SM Table 4).

## 396 Horizontal gene transfers (HGT)

397 To detect HGT from non-metazoan species, we first used the pipeline of foreign  
 398 sequence detection developed by Francois et al. (60). We used the set of CDS  
 399 identified in publicly available transcriptomes (37) and the genome assemblies prior  
 400 to the decontamination procedure with Blobtools (54). The rationale is that some  
 401 genuine HGT could have been wrongly considered as contaminant sequences  
 402 during this decontamination step and thus been removed from the assembly.  
 403 Scaffolds filtered during decontamination are available from our github repository  
 404 ([https://github.com/AsexGenomeEvol/Timema\\_asex\\_genomes/tree/main/4\\_Horizontal](https://github.com/AsexGenomeEvol/Timema_asex_genomes/tree/main/4_Horizontal_Gene_Transfers/contamination_sequences)  
 405 [al\\_Gene\\_Transfers/contamination\\_sequences](https://github.com/AsexGenomeEvol/Timema_asex_genomes/tree/main/4_Horizontal_Gene_Transfers/contamination_sequences)), and will be archived upon  
 406 acceptance.

407 Briefly, a DIAMOND BlastP (v0.8.33) (61) allows to detect candidate non-metazoan  
 408 genes in the set of CDS of each species. Taxonomic assignment is based on the 10  
 409 best blast hits to account for potential contaminations and other sources of  
 410 taxonomic misassignment in the reference database. Candidate non-metazoan  
 411 sequences are then subjected to a synteny-based screen with Gmap (v2016-11-07)  
 412 (62) to discriminate between contaminant sequences and potential HGT-derived  
 413 sequences. A sequence is considered as a HGT candidate if it is physically linked to  
 414 (i.e., mapped to the same scaffold as) at least one “confident-arthropod” CDS  
 415 (previously identified in the DIAMOND blast).

416 We then clustered all HGT candidates identified in each of the 10 *Timema* species  
 417 into HGT families using Silix (v1.2.10) (63), requiring a minimum of 85% identity  
 418 (default parameters otherwise). These HGT families were then “completed” as much  
 419 as possible by adding homologs from the genome assemblies not identified as HGT  
 420 candidates (this could occur if the corresponding sequences are fragmented or on  
 421 short scaffolds for example). To this end, the longest sequence of each HGT family  
 422 was mapped (using Gmap) on the genomic scaffolds of all species, requiring a  
 423 minimum of 85% identity.

424 For each completed HGT family, a protein alignment of the candidate HGT  
425 sequence(s) and its (their) 50 best DIAMOND blastP hits in the reference database  
426 (1<sup>st</sup> step of the pipeline) was generated with MAFFT (v7) (64). The alignments were  
427 cleaned using HMMcleaner (stringency parameter = 12) (65) and sites with more  
428 than 50% missing data were removed. Phylogenetic trees were inferred using  
429 RAXML (v8.2) (66) with the model 'PROTGAMMALGX' of amino-acid substitution  
430 and 100 bootstrap replicates. Phylogenetic trees were inspected by eye to confirm or  
431 not an evolutionary history consistent with the hypothesis of HGT.

### 432 Heterozygosity

433  
434 Genome-wide nucleotide heterozygosity was estimated using genome profiling  
435 analysis of raw reads from the reference genomes using GenomeScope (v2) (23). A  
436 second, SNP-based heterozygosity estimate was generated using re-sequenced  
437 individuals. We re-sequenced five individuals per species, but 3 individuals of *T.*  
438 *shepardi*, 2 individuals of *T. poppensis* and one *T. tahoe* individual did not pass  
439 quality control and were discarded from all downstream analyses. SNP calling was  
440 based on the GATK best practices pipeline (67). We used a conservative set of  
441 SNPs with quality scores  $\geq 300$ , and supported by 15x coverage in at least one of the  
442 individuals. SNP heterozygosity was then estimated as the number of heterozygous  
443 SNPs divided by the number of callable sites in each genome. Due to stringent  
444 filtering criteria, our SNP based heterozygosity is an underestimation of  
445 genome-wide heterozygosity.

### 446 Structural variants

447 We used Manta (v1.5.0) (68), a diploid-aware pipeline for structural variant (SV)  
448 calling, in the same set of re-sequenced individuals used for SNP heterozygosity  
449 estimates. We found a high frequency of heterozygous SVs with approximately twice  
450 the expected coverage (SM Figure 7), which likely represent false positives. To  
451 reduce the number of false positives, we filtered very short SVs (30 bases or less)  
452 and kept only variant calls that had either split read or paired-end read support  
453 within the expected coverage range, where the coverage range was defined

454 individually for each sample by manual inspection of coverage distributions. The  
455 filtered SV calls were subsequently merged into population SV calls using  
456 SURVIVOR (v1.0.2) (69). The merging criteria were: SV calls of the same type on  
457 the same strand with breakpoints distances shorter than 100 bp.

#### 458 Genome alignment

459 We anchored our genome assemblies to the reference of *T. cristinae* (BioProject  
460 Accession PRJNA417530) (35) using MUMmer (version 4.0.0beta2) (70) with  
461 parameter --mum. The alignments were processed by other tools within the package:  
462 show-coords with parameters -THrc1 to generate tab-delimited alignment files and  
463 dnadiff to generate 1-to-1 alignments. We used only uniquely anchored scaffolds for  
464 which we were able to map at least 10k nucleotides to the *T. cristinae* reference  
465 genome.

#### 466 Transposable elements

467 For each species, specific repeat libraries were constructed and annotated to the TE  
468 superfamily level (51) wherever possible. For collecting repetitive sequences, we  
469 used a raw read based approach DNAPipeTE v1.2 (71) with parameters  
470 -genome\_coverage 0.5 -sample\_number 4 and respective species genome size, as  
471 well as an assembly based approach (RepeatModeler v1.0.8 available at  
472 <http://www.repeatmasker.org/RepeatModeler/>), such that repeats not present in the  
473 assembly can still be represented in the repeat library. The two raw libraries were  
474 merged and clustered by 95% identity (the TE family threshold) using usearch  
475 v10.0.240 (72) with the centroid option. To annotate TEs larger than 500 bp in the  
476 repeat library, we used an approach that combines homology and structural  
477 evidence (PASTECClassifier (73)). Because PASTECClassifier did not annotate to TE  
478 superfamily levels, we additionally compared by BlastN (v. 2.7.1+) (74) the repeat  
479 libraries to the well curated *T. cristinae* TE library from Soria-Carrasco et al. (21).  
480 Blast hits were filtered according to TE classification standards: identity percentage  
481 >80%, alignment length >80 bp, and the best hit per contig was kept. The two  
482 classification outputs were compared and in case of conflict the classification level of  
483 PASTECClassifier was preferred. All non-annotated repeats were labelled 'unknown'.

484 Repeat library header naming was done according to RepeatMasker standard, but  
 485 keeping the Wicker naming for elements (i.e., Wicker#Repeatmasker, e.g.,  
 486 DTA#DNA/hAT). TE libraries were sorted by header and TE annotations to similar  
 487 families numbered consecutively. Species-specific TE libraries were merged into a  
 488 genus-level *Timema* TE library to account for any TE families that might have not  
 489 been detected in the single species assemblies.

490

491 To estimate the TE load of reference genomes and resequenced individuals, we first  
 492 repeat masked the assemblies with the genus-level TE library using RepeatMasker  
 493 v4.1.0 with parameters set as -gccalc -gff -u -a -xsmall -no\_is -div 30 -engine rmbblast  
 494 (75). Second, we mapped the 350 bp insert paired-end reads back to the reference  
 495 genome assemblies using BWA-MEM v0.7.17 (76) with standard parameters. We  
 496 then counted the fraction of reads mapping to TEs out of total mappable reads by  
 497 counting the number of reads that mapped to each genomic location annotated as  
 498 TE using htseq-counts (v0.6.1.1p1) (77) with parameters set to -r name -s no -t  
 499 similarity -i Target --nonunique using the mapped read alignments and the gff output  
 500 of RepeatMasker (filtered for TE length of >80 bp). TE loads were compared among  
 501 species using a permutation ANOVA with 5000 bootstrap replicates.

## 502 Positive selection analysis

503 Only one-to-one orthologs in at least three pairs of species (sister-species sex-asex)  
 504 were used. The species phylogeny was imposed on every gene as the "gene tree".  
 505 We used a customized version of the Selectome pipeline (78). All alignment building  
 506 and filtering was performed on predicted amino acid sequences, and the final amino  
 507 acid MSAs (multiple sequence alignments) were used to infer the nucleotide MSAs  
 508 used for positive selection inference. MSAs were obtained by MAFFT (v. 7.310) (64)  
 509 with the allowshift option, which avoids over-aligning non homologous regions (e.g.  
 510 gene prediction errors, or alternative transcripts). All the next steps "mask" rather  
 511 than remove sites, by replacing the amino acid with a 'X' and the corresponding  
 512 codon with 'NNN'. MCoffee (v11.00.8cbe486) (79) was run with the following  
 513 aligners: mafft\_msa, muscle\_msa, clustalo\_msa (80), and t\_coffee\_msa (81).  
 514 MCoffee provides a consistency score per amino acid, indicating how robust the

alignment is at that position for that sequence. Residues with a consistency score less than 5 were masked. TrimAl (v. 1.4.1) (82) was used to mask columns with less than 4 residues (neither gap nor 'X').

518

The branch-site model with rate variation at the DNA level (44) was run using the Godon software (<https://bitbucket.org/Davydov/godon/>, version 2020-02-17, option BSG --ncat 4). Each branch was tested iteratively, in one run per gene tree. For each branch, we obtain a  $\Delta\ln L$  which measures the evidence for positive selection, a corresponding p-value and associated q-value (estimated from the distribution of p-values over all branches of all genes), and an estimate of the proportion of sites under positive selection if any. All positive selection results, and detailed methods, will be available at <https://selectome.org/timema>. To determine if there the number of positively selected genes differed between sexual than parthenogenetic species we used a binomial GLMM approach (lme4 (83)) with q-value threshold of 0.05 or 0.01. Significance of model terms was determined with a Wald statistic. In addition, we also examined if there was more evidence for positive selection in sexual species in a threshold-free way by comparing  $\Delta\ln L$  values between parthenogenetic and sexual species (as in (45, 46)). To do this we used a permutation glm approach where reproductive mode (sexual or parthenogenetic) was randomly switched within a species-pair. To determine if the overlap of positively selected genes was greater than expected by chance we used the SuperExactTest package (v. 0.99.4) (84) in R. The resulting p-values were multiple test corrected using Benjamini and Hochberg's algorithm implemented in R. Functional enrichment analyses were performed using TopGO (v. 2.28.0) (85) using the *D. melanogaster* functional annotation (see SM text 1). To determine if a GO term was enriched we used a Fisher's exact test with the 'weight01' algorithm to account for the GO topology. GO terms were considered to be significantly enriched when  $p < 0.05$ .



542

## 543 Data and code availability

544 Raw sequence reads have been deposited in NCBI's sequence read archive under  
 545 the following bioprojects: PRJNA371785 (reference genomes, SM Table 7A),  
 546 PRJNA670663 (resequenced individuals, SM Table 7B), and PRJNA673001 (PacBio  
 547 reads for *T. douglasi*). Genome assemblies and annotations PRJEB31411. Scripts  
 548 for the analyses in this paper are available at: [https://github.com/AsexGenomeEvol/](https://github.com/AsexGenomeEvol/Timema_asex_genomes)  
 549 Timema\_asex\_genomes. Data were processed to generate plots and statistics using  
 550 R v3.4.4.

## 551 Acknowledgements

552 We thank Bart Zijlstra, Chloé Larose, and Kirsten Jalvingh for help in the field, and  
 553 Deborah Charlesworth, Daniel Wegmann, and current and previous members of the  
 554 Schwander and Robinson-Rechavi labs for discussions. This project was supported  
 555 by several grants from the Swiss Science Foundation (CRSII3\_160723 as well as  
 556 PP00P3\_170627, 31003A\_182495, IZLRZ3\_163872, and 407540\_167276) and  
 557 funding from the University of Lausanne.

## 558 References

- 559 1. M. Neiman, C. M. Lively, S. Meirmans, Why sex? A pluralist approach revisited. *Trends*  
 560 *Ecol. Evol.* **32**, 589–600 (2017).
- 561 2. N. P. Sharp, S. P. Otto, Evolution of sex: Using experimental genomics to select among  
 562 competing theories. *Bioessays*. **38**, 751–757 (2016).
- 563 3. G. Bell, *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ of  
 564 California Press, Los Angeles, 1982).
- 565 4. G. C. Williams, *Sex and Evolution* (Princeton University Press, 1975).
- 566 5. J. Maynard Smith, *The Evolution of Sex* (CUP, 1978).
- 567 6. J. Gerritsen, Sex and Parthenogenesis in Sparse Populations. *Am. Nat.* **115**, 718–742  
 568 (1980).
- 569 7. S. K. Jain, The Evolution of Inbreeding in Plants. *Annu. Rev. Ecol. Syst.* **7**, 469–495

- 570 (1976).
- 571 8. J. Felsenstein, The evolutionary advantage of recombination. *Genetics*. **78**, 737–756  
572 (1974).
- 573 9. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.*  
574 **8**, 269–294 (1966).
- 575 10. P. D. Keightley, S. P. Otto, Interference among deleterious mutations favours sex and  
576 recombination in finite populations. *Nature*. **443**, 89–92 (2006).
- 577 11. N. H. Barton, Why sex and recombination? *Cold Spring Harb. Symp. Quant. Biol.* **74**,  
578 187–195 (2009).
- 579 12. C. W. Birky Jr, Heterozygosity, heteromorphy, and phylogenetic trees in asexual  
580 eukaryotes. *Genetics*. **144**, 427–437 (1996).
- 581 13. D. Mark Welch, M. Meselson, Evidence for the evolution of bdelloid rotifers without  
582 sexual reproduction or genetic exchange. *Science*. **288**, 1211–1215 (2000).
- 583 14. D. A. Hickey, Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*. **101**,  
584 519–531 (1982).
- 585 15. E. S. Dolgin, B. Charlesworth, The effects of recombination rate on the distribution and  
586 abundance of transposable elements. *Genetics*. **178**, 2169–2177 (2008).
- 587 16. J. Bast, K. S. Jaron, D. Schuseil, D. Roze, T. Schwander, Asexual reproduction reduces  
588 transposable element load in experimental yeast populations. *Elife*. **8** (2019),  
589 doi:10.7554/eLife.48548.
- 590 17. K. S. Jaron, J. Bast, R. W. Nowell, T. R. Ranallo-Benavidez, M. Robinson-Rechavi, T.  
591 Schwander, Genomic Features of Parthenogenetic Animals. *J. Hered.* (2020),  
592 doi:10.1093/jhered/esaa031.
- 593 18. T. Schwander, B. J. Crespi, Multiple direct transitions from sexual reproduction to  
594 apomictic parthenogenesis in *Timema* stick insects. *Evolution*. **63**, 84–103 (2009).
- 595 19. T. Schwander, L. Henry, B. J. Crespi, Molecular evidence for ancient asexuality in  
596 *Timema* stick insects. *Curr. Biol.* **21**, 1129–1134 (2011).
- 597 20. R. Riesch, M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K.  
598 Lucek, E. Hellen, V. Soria-Carrasco, S. R. Dennis, C. F. de Carvalho, R. J. Safran, C. P.  
599 Sandoval, J. Feder, R. Gries, B. J. Crespi, G. Gries, Z. Gompert, P. Nosil, Transitions  
600 between phases of genomic differentiation during stick-insect speciation. *Nature*  
601 *Ecology & Evolution*. **1**, 0082 (2017).
- 602 21. V. Soria-Carrasco, Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.  
603 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, P.  
604 Nosil, Stick insect genomes reveal natural selection's role in parallel speciation.  
605 *Science*. **344**, 738–742 (2014).
- 606 22. R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E.  
607 V. Kriventseva, E. M. Zdobnov, BUSCO Applications from Quality Assessments to Gene  
608 Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

- 609 23. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot  
610 for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
- 611 24. J. Romiguier, P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R.  
612 Darnat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.  
613 Tsagkogeorga, A. A.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, N. Galtier,  
614 Comparative population genomics in animals uncovers the determinants of genetic  
615 diversity. *Nature*. **515**, 261–263 (2014).
- 616 25. A. Mackintosh, D. R. Laetsch, A. Hayward, B. Charlesworth, M. Waterfall, R. Vila, K.  
617 Lohse, The determinants of genetic diversity in butterflies. *Nat. Commun.* **10**, 3466  
618 (2019).
- 619 26. T. Schwander, S. Vuilleumier, J. Dubman, B. J. Crespi, Positive feedback in the  
620 transition from sexual reproduction to parthenogenesis. *Proc. Biol. Sci.* **277**, 1435–1442  
621 (2010).
- 622 27. E. Suomalainen, A. Saura, J. Lokki, *Cytology and evolution in parthenogenesis* (CRC  
623 Press, 1987).
- 624 28. J. Engelstädter, Asexual but Not Clonal: Evolutionary Processes in Automictic  
625 Populations. *Genetics*. **206**, 993–1009 (2017).
- 626 29. M. Neiman, T. Schwander, Using Parthenogenetic Lineages to Identify Advantages of  
627 Sex. *Evol. Biol.* **38**, 115–123 (2011).
- 628 30. S. Glémin, C. M. François, N. Galtier, Genome Evolution in Outcrossing vs. Selfing vs.  
629 Asexual Species. *Methods Mol. Biol.* **1910**, 331–369 (2019).
- 630 31. M. Percy, S. Aron, C. Doums, L. Keller, Conditional use of sex and parthenogenesis for  
631 worker and queen production in ants. *Science*. **306**, 1780–1783 (2004).
- 632 32. M. O. Lorenzo-Carballa, A. Cordero-Rivera, Thelytokous parthenogenesis in the  
633 damselfly *Ichnura hastata* (Odonata, Coenagrionidae): genetic mechanisms and lack  
634 of bacterial infection. *Heredity*. **103**, 377–384 (2009).
- 635 33. T. J. Treangen, S. L. Salzberg, Repetitive DNA and next-generation sequencing:  
636 computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
- 637 34. H. Ellegren, N. Galtier, Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433  
638 (2016).
- 639 35. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B.  
640 J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in *Timema*  
641 stick insects. *Science*. **359**, 765–770 (2018).
- 642 36. V. R. Vickery, Revision of *Timema scudder* (Phasmatoptera: Timematodea) including  
643 three new species. *Can. Entomol.* **125**, 657–692 (1993).
- 644 37. J. Bast, D. J. Parker, Z. Dumas, K. M. Jalvingh, P. Tran Van, K. S. Jaron, E. Figuet, A.  
645 Brandt, N. Galtier, T. Schwander, Consequences of asexuality in natural populations:  
646 insights from stick insects. *Mol. Biol. Evol.* **35**, 1668–1677 (2018).
- 647 38. L. Henry, T. Schwander, B. J. Crespi, Deleterious mutation accumulation in asexual

- 648 Timema stick insects. *Mol. Biol. Evol.* **29**, 401–408 (2012).
- 649 39. M. Neiman, P. G. Meirmans, T. Schwander, S. Meirmans, Sex in the wild: How and why  
650 field-based studies contribute to solving the problem of sex. *Evolution*. **72**, 1194–1203  
651 (2018).
- 652 40. S. P. Otto, Selective Interference and the Evolution of Sex. *J. Hered.* (2020),  
653 doi:10.1093/jhered/esaa026.
- 654 41. M. J. McDonald, D. P. Rice, M. M. Desai, Sex speeds adaptation by altering the  
655 dynamics of molecular evolution. *Nature*. **531**, 233–236 (2016).
- 656 42. O. Kaltz, G. Bell, The ecology and genetics of fitness in *Chlamydomonas*. XII. Repeated  
657 sexual episodes increase rates of adaptation to novel environments. *Evolution*. **56**,  
658 1743–1753 (2002).
- 659 43. M. R. Goddard, H. C. J. Godfray, A. Burt, Sex increases the efficacy of natural selection  
660 in experimental yeast populations. *Nature*. **434**, 636–640 (2005).
- 661 44. I. I. Davydov, N. Salamin, M. Robinson-Rechavi, Large-Scale Comparative Analysis of  
662 Codon Models Accounting for Protein and Nucleotide Selection. *Mol. Biol. Evol.* **36**,  
663 1316–1332 (2019).
- 664 45. J. T. Daub, S. Moretti, I. I. Davydov, L. Excoffier, M. Robinson-Rechavi, Detection of  
665 Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Mol.*  
666 *Biol. Evol.* **34**, 1391–1402 (2017).
- 667 46. J. Liu, M. Robinson-Rechavi, Adaptive Evolution of Animal Proteins over Development:  
668 Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo. *Mol. Biol. Evol.*  
669 **35**, 2862–2872 (2018).
- 670 47. W. Haerty, S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram, L. K. Sirot, L.  
671 Levesque, C. G. Artieri, M. F. Wolfner, A. Civetta, R. S. Singh, Evolution in the fast lane:  
672 rapidly evolving sex-related genes in *Drosophila*. *Genetics*. **177**, 1321–1335 (2007).
- 673 48. B. Charlesworth, C. H. Langley, The evolution of self-regulated transposition of  
674 transposable elements. *Genetics*. **112**, 359–383 (1986).
- 675 49. T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.*  
676 **24**, R288–94 (2014).
- 677 50. D. Bachtrog, Y-chromosome evolution: emerging insights into processes of  
678 Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
- 679 51. T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P.  
680 Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A. H. Schulman, A unified  
681 classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982  
682 (2007).
- 683 52. S. D. Jackman, B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G.  
684 Jahesh, H. Khan, L. Coombe, R. L. Warren, I. Birol, ABySS 2.0: resource-efficient  
685 assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).
- 686 53. K. Sahlin, R. Chikhi, L. Arvestad, Assembly scaffolding with PE-contaminated mate-pair

- 687 libraries. *Bioinformatics*. **32**, 1925–1932 (2016).
- 688 54. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies. *F1000Res*.  
689 **6**, 1287 (2017).
- 690 55. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander,  
691 Sex-biased gene expression is repeatedly masculinized in asexual females. *Nat.*  
692 *Commun.* **10**, 4638 (2019).
- 693 56. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander,  
694 Repeated evolution of asexuality involves convergent gene expression changes. *Mol.*  
695 *Biol. Evol.* **36**, 350–364 (2019).
- 696 57. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B.  
697 Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F.  
698 Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc,  
699 N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq  
700 using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**,  
701 1494–1512 (2013).
- 702 58. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database  
703 management tool for second-generation genome projects. *BMC Bioinformatics*. **12**, 491  
704 (2011).
- 705 59. E. V. Kriventseva, F. Tegenfeldt, T. J. Petty, R. M. Waterhouse, F. A. Simão, I. A.  
706 Pozdnyakov, P. Ioannidis, E. M. Zdobnov, OrthoDB v8: update of the hierarchical  
707 catalog of orthologs and the underlying free software. *Nucleic Acids Res.* **43**, D250–6  
708 (2015).
- 709 60. C. M. Francois, F. Durand, E. Figuet, N. Galtier, Prevalence and Implications of  
710 Contamination in Public Genomic Resources: A Case Study of 43 Reference Arthropod  
711 Assemblies. *G3*. **10**, 721–730 (2020).
- 712 61. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND.  
713 *Nat. Methods*. **12**, 59–60 (2015).
- 714 62. T. D. Wu, C. K. Watanabe, GMAP: a genomic mapping and alignment program for  
715 mRNA and EST sequences. *Bioinformatics*. **21**, 1859–1875 (2005).
- 716 63. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with  
717 SiLiX. *BMC Bioinformatics*. **12**, 116 (2011).
- 718 64. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7:  
719 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- 720 65. A. Di Franco, R. Poujol, D. Baurain, H. Philippe, Evaluating the usefulness of alignment  
721 filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol.*  
722 *Biol.* **19**, 21 (2019).
- 723 66. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of  
724 large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
- 725 67. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A.  
726 Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella,



- 727 D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant  
728 calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*.  
729 **43**, 11.10.1–11.10.33 (2013).
- 730 68. X. Chen, O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, A. J. Cox,  
731 S. Kruglyak, C. T. Saunders, Manta: rapid detection of structural variants and indels for  
732 germline and cancer sequencing applications. *Bioinformatics*. **32**, 1220–1222 (2016).
- 733 69. D. C. Jeffares, C. Jolly, M. Hoti, D. Speed, L. Shaw, C. Rallis, F. Balloux, C. Dessimoz,  
734 J. Bähler, F. J. Sedlazeck, Transient structural variations have strong effects on  
735 quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061  
736 (2017).
- 737 70. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L.  
738 Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**,  
739 R12 (2004).
- 740 71. C. Goubert, L. Modolo, C. Vieira, C. ValienteMoro, P. Mavingui, M. Boulesteix, De novo  
741 assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with  
742 dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever  
743 mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **7**, 1192–1205 (2015).
- 744 72. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST.  
745 *Bioinformatics*. **26**, 2460–2461 (2010).
- 746 73. C. Hoede, S. Arnoux, M. Moisset, T. Chaumier, O. Inizan, V. Jamilloux, H. Quesneville,  
747 PASTEC: an automatic transposable element classification tool. *PLoS One*. **9**, e91929  
748 (2014).
- 749 74. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman,  
750 Gapped BLAST and PSI-BLAST: a new generation of protein database search  
751 programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- 752 75. A. Smit, R. Hubley, P. Green, *RepeatMasker Open-4.0* (2013-2015;  
753 <http://www.repeatmasker.org>).
- 754 76. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.  
755 *arXiv [q-bio.GN]* (2013), (available at <http://arxiv.org/abs/1303.3997>).
- 756 77. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with  
757 high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
- 758 78. S. Moretti, B. Laurency, W. H. Gharib, B. Castella, A. Kuzniar, H. Schabauer, R. A.  
759 Studer, M. Valle, N. Salamin, H. Stockinger, M. Robinson-Rechavi, Selectome update:  
760 quality control and computational improvements to a database of positive selection.  
761 *Nucleic Acids Res.* **42**, D917–21 (2014).
- 762 79. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: combining multiple  
763 sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
- 764 80. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam,  
765 M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of  
766 high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*

- 767        **7**, 539 (2011).
- 768 81. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and  
769        accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
- 770 82. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated  
771        alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**, 1972–1973  
772        (2009).
- 773 83. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using  
774        lme4. *Journal of Statistical Software, Articles.* **67**, 1–48 (2015).
- 775 84. M. Wang, Y. Zhao, B. Zhang, Efficient test and visualization of multi-set intersections.  
776        *Sci. Rep.* **5**, 16923 (2015).
- 777 85. A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from  
778        gene expression data by decorrelating GO graph structure. *Bioinformatics.* **22**,  
779        1600–1607 (2006).
- 780