Convergent consequences of parthenogenesis on stick insect genomes

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

[1] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland;

[2] Swiss Institute of Bioinformatics, Lausanne, Switzerland;

[3] Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL

[4] ISEM - Institut des Sciences de l'Evolution, Montpellier, France;

[5] Current address: CoBIUS lab, Department of Computer Science, University of Sherbrooke, Sherbrooke, Canada;

[6] Current address: Institute for Zoology, University of Cologne, Köln, Germany;

[7] Current address: Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France;

[8] Current address: Université de Namur, LEGE, URBE, Namur, 5000, Belgium

[a] shared first authorship

[b] equal contribution

Correspondence should be addressed to DarrenJames.Parker@unil.ch, Kamil.Jaron@ed.ac.uk, or Tanja.Schwander@unil.ch

Abstract

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

Introduction

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

Parthenogenesis is thought to be favored in the short term because it generates a transmission advantage (*4*, *5*), as well as the advantage of assured reproduction when mates are scarce (*6*, *7*). The short-term benefits of parthenogenesis, however, are

62 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 natural
64 selection (e.g. (*8–10*), reviewed in (*11*)). This is expected to translate into reduced
65 rates of adaptation and increased accumulation of mildly deleterious mutations, which
66 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 structural
76 variants, which in sexuals are counter-selected due to the constraints of properly
77 pairing homologous chromosomes during meiosis (*17*).
78
79 We tested these predictions by comparing the genomes of five independently derived
80 parthenogenetic stick insect species in the genus *Timema* with their close sexual
81 relatives (Fig. 1). These replicate comparisons allowed us to solve the key problem in
82 understanding the consequences of parthenogenesis for genome evolution:
83 separating the consequences of parthenogenesis from lineage specific effects (*17*).
84 *Timema* are wingless, plant-feeding insects endemic to western North America.
85 Parthenogenetic species in this genus are diploid and of non-hybrid origin (*18*) and
86 ecologically similar to their sexual relatives. Previous research, based on a small
87 number of microsatellite markers, has suggested that oogenesis in parthenogenetic
88 *Timema* is functionally mitotic, as no loss of heterozygosity between females and their
89 offspring was detected (*18*).
90

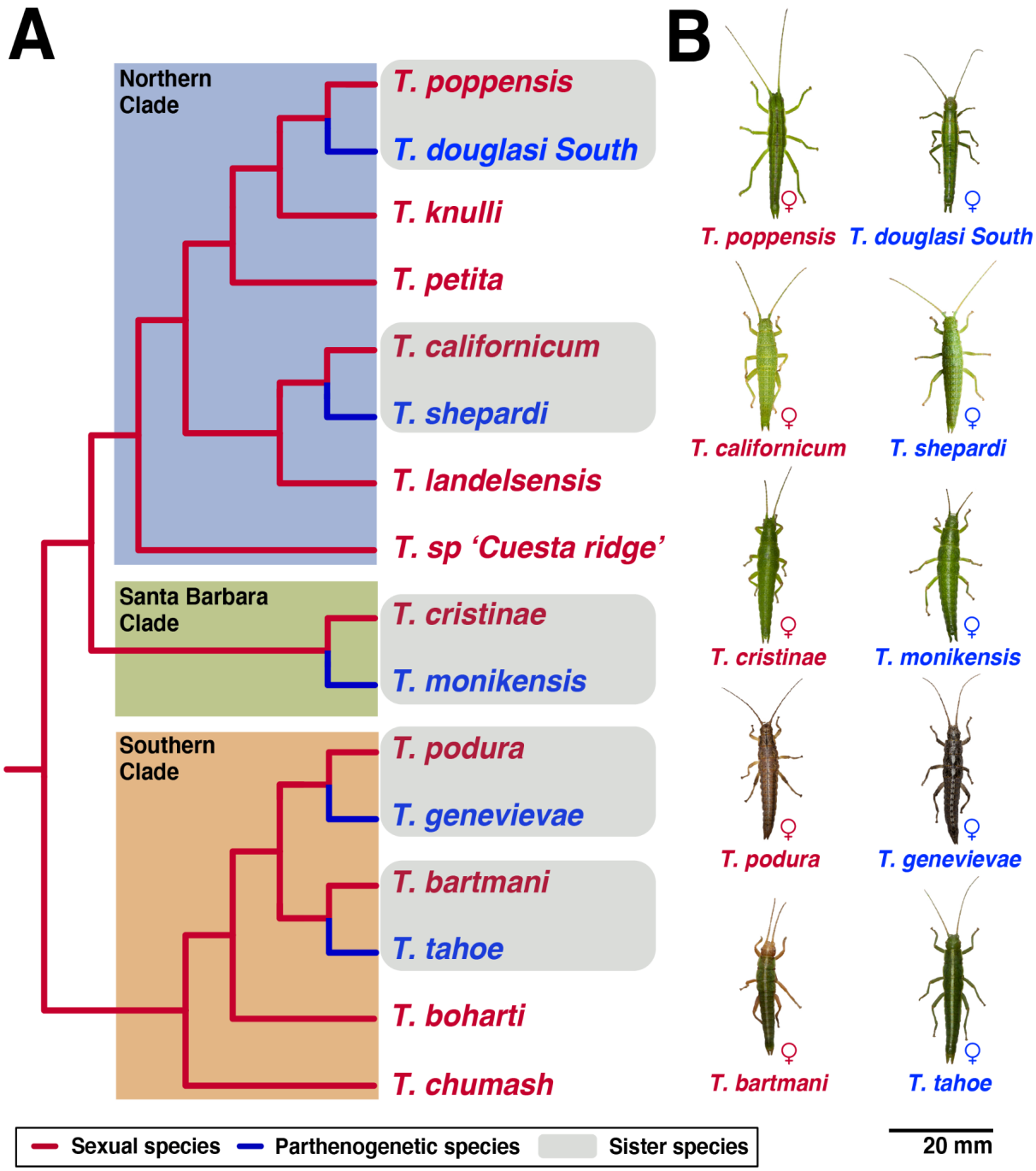


Fig. 1. Multiple, independent transitions from sexual to parthenogenetic reproduction are known in the genus *Timema (19)*, each representing a biological replicate of parthenogenesis, and with a close sexual relative at hand for comparison **A.** Phylogenetic relationships of *Timema* species (adapted from (*19*, *20*)). **B.** Species sequenced in this study. Photos taken by © Bart Zijlstra - www.bartzijlstra.com.

97 # Results

98 ***De novo* genome assemblies 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 (Fig. 1, Tables S1, S2). Genomes were
102 subjected to quality control, screened for contamination, and annotated (see Methods,
103 Supplementary Text). The final reference genomes were largely haploid, spanned 75-
104 95% of the estimated genome size (1.38 Gbp (*21*)), and were sufficiently complete for
105 downstream analyses, as shown by the count of single copy orthologs conserved
106 across insects (96% of BUSCO genes (*22*) detected on average; Table S3). A
107 phylogeny based on a conservative set of 3975 1:1 orthologous genes (Data S1)
108 corroborated published phylogenies and molecular divergence estimates in the
109 *Timema* genus (Fig. S1). Finally, we identified 55 candidate horizontal gene transfers
110 (HGTs) from non-metazoans. Using long reads, we were able to corroborate six out
111 of the eight HGT candidates present in one species (*T. douglasi*, see Supplementary
112 Text), indicating that the majority of our HGT candidates are likely true HGTs rather
113 than false positives. All 55 candidate HGTs occurred early on in *Timema* evolutionary
114 history, well before the evolution of parthenogenesis in the genus (see Supplementary
115 Text).
116
117 We estimated genome-wide nucleotide heterozygosity in each reference genome
118 directly from sequencing reads, using a reference-free technique (genome profiling
119 analysis (*23*)). These analyses revealed extreme heterozygosity differences between
120 the sexual and parthenogenetic species. The five sexual *Timema* featured nucleotide
121 heterozygosities within the range previously observed in other sexual species (Fig. 2;
122 (*24*, *25*)). The heterozygosities in the parthenogenetic species were substantially
123 lower, and in fact so low that reference-free analyses could not distinguish
124 heterozygosity from sequencing error (see Supplementary Text). We therefore
125 compared heterozygosity between sexuals and parthenogens by calling SNPs in five
126 re-sequenced individuals per species. This analysis corroborated the finding that
127 parthenogens have extremely low ($<10^{-5}$) heterozygosity, being at least 140 times
128 lower than that found in their sexual sister species (permutation ANOVA, reproductive

mode effect p = 0.0049; Fig. 2). Screening for structural variants (indels, tandem duplications, and inversions) in sexual and parthenogenetic individuals revealed the same pattern: extensive and variable heterozygosity in sexual species and homozygosity in the parthenogens (Fig. 2, Supplementary Text). Some heterozygosity in *Timema* parthenogens could be present in genomic regions not represented in our assemblies, such as centromeric and telomeric regions. These regions however represent a relatively small fraction of the total genome, meaning that for most of the genome at least, *Timema* parthenogens are either largely or completely homozygous for all types of variants (see Supplementary Text).

We are able to demonstrate that extremely low heterozygosity is a convergent consequence of parthenogenesis in *Timema*. However, while extremely low heterozygosity is the rule for *Timema* parthenogens this is not the case for all parthenogenetic species. For example, we have previously shown that some parthenogenetic animal species, particularly those of hybrid origin, are characterized by relatively high heterozygosity levels (*17*). This indicates that the consequences of parthenogenesis for heterozygosity are likely to be lineage-specific.

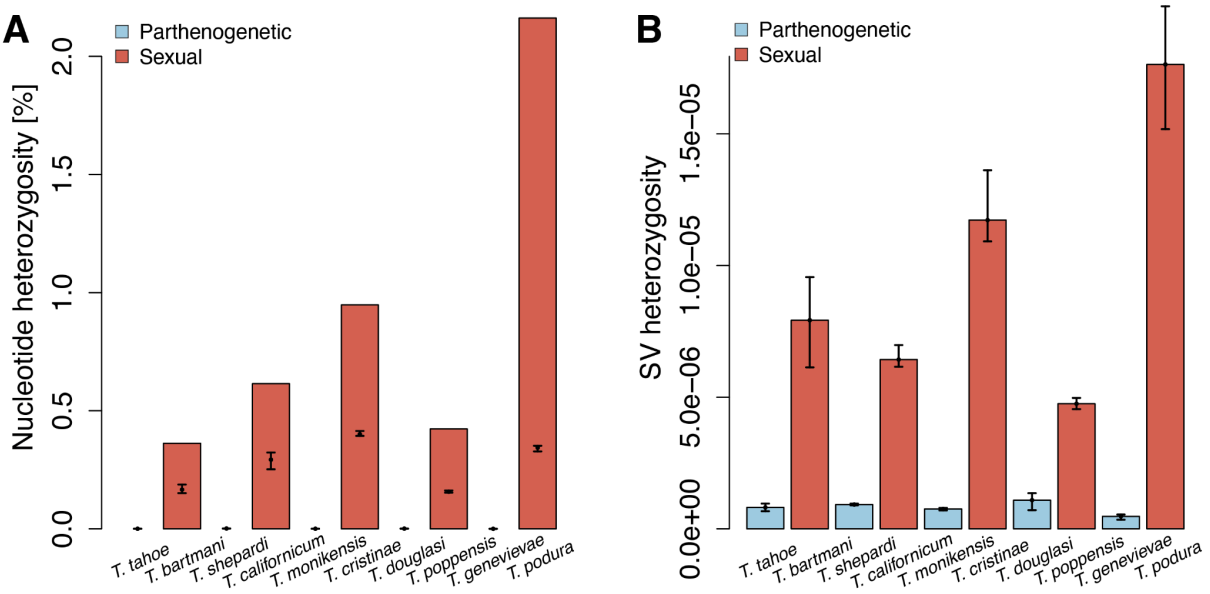


Fig. 2. Extremely low heterozygosity in parthenogenetic *Timema* species for different types of variants. **A.** Nucleotide heterozygosity represented by bars indicates genome-wide estimates for the reference individuals (based on raw reads, see Methods), heterozygosity based on SNP calls in re-sequenced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

The unexpected finding of extremely low heterozygosity in *Timema* parthenogens raises the question of when and how heterozygosity was lost. For example, the bulk of heterozygosity could have been lost during the transition from sexual reproduction to obligate parthenogenesis (*26*). This would be the case if functionally mitotic parthenogenesis was derived from automictic parthenogenesis, meaning that recombination and meiosis take place, and diploidy is restored secondarily, for example via fusion of two of the four meiotic products. Similar to inbreeding, automictic parthenogenesis thus results in the rapid loss of heterozygosity over time (*27*). Automixis can then be co-opted into functionally mitotic parthenogenesis via

169 recombination suppression, and by solely fusing meiotic products separated during
170 meiosis I and not meiosis II (*18*). Alternatively, heterozygosity loss in *Timema*
171 parthenogens could be a continuous and ongoing process. To distinguish these
172 options, we investigated the origin of the genetic variation present among different
173 homozygous genotypes in each parthenogenetic species. We found that only 6-19%
174 of the SNPs called in a parthenogen are at positions that are also polymorphic in the
175 sexual relative (Table S4). This means that most of the variation in parthenogens likely
176 results from mutations that appeared after the split from the sexual lineage. Although
177 it is possible that some of the parthenogen-specific polymorphisms may in fact
178 represent ancestral polymorphisms that were purged in the sexual species, it is
179 unlikely that this would be the case for the majority of SNPs in parthenogens. This
180 implies that heterozygosity generated through new mutations is most likely lost
181 continuously in parthenogens, and was not solely lost at the inception of
182 parthenogenesis. The most likely explanation for these findings is that parthenogenetic
183 *Timema* are, in fact, not functionally mitotic but automictic. Formally distinguishing
184 between automixis and functionally mitotic parthenogenesis with gene conversion will
185 require cell biological data which is currently not available for *Timema*. It is important
186 to note that the key theoretical predictions regarding the consequences of sex do not
187 change between automixis or functionally mitotic parthenogenesis. Thus, although
188 automixis can allow for the purging of heterozygous deleterious mutations (*28*), the
189 classical predictions for the long-term costs of asexuality extend to automictic
190 parthenogens because, as for obligate selfers, linkage among genes is still much
191 stronger than in classical sexual species (*29*). This is especially the case in largely
192 homozygous parthenogens, where recombination and segregation, even if
193 mechanistically present, have no effect on genotype diversities.
194
195 Functional mitosis in *Timema* parthenogens was previously inferred from the
196 inheritance of heterozygous microsatellite genotypes between females and their
197 offspring (*18*), a technique widely used in non-model organisms with no cytological
198 data available (e.g., (*30*, *31*)). The most likely reconciliation of these results with our
199 finding of extreme homozygosity is that heterozygosity is maintained in only a small
200 portion of the genome, for example the centromeres or telomeres, or between
201 paralogs. Consistent with this idea, we were unable to locate several of the
202 microsatellite-containing regions in even the best *Timema* genome assemblies (see

203 Supplementary Text), suggesting that these regions are not present in our assemblies
204 due to the inherent difficulty of assembling repetitive genome regions from short read
205 data (*32*).
206

207 **Extensive variation in genotype diversity between parthenogenetic populations**

208
209 Parthenogenesis and sexual reproduction are expected to drive strikingly different
210 distributions of polymorphisms in genomes and populations. Within genomes, different
211 regions experience different types of selection with sometimes opposite effects on the
212 levels of polymorphisms within populations, such as purifying versus balancing
213 selection (*33*). The increased linkage among genes in parthenogenetic as compared
214 to sexual species is expected to homogenize diversity levels across different genome
215 regions. At the species level, parthenogens are expected to have reduced genetic
216 diversities relative to sexual species because of likely bottlenecks occurring during the
217 origin of parthenogenesis. At the population level, recurrent sweeps of specific
218 genotypes in parthenogenetic populations can lead to extremely low genetic diversity
219 and even to the fixation of a single genotype, while sweeps in sexual populations
220 typically reduce diversity only in specific genome regions.
221
222 To address these aspects in the genomes of sexual and parthenogenetic *Timema*
223 species, we mapped population-level variation for the SNPs and SVs inferred above
224 to our species-specific reference genomes. We then anchored our reference genome
225 scaffolds to the 12 autosomal linkage groups of a previously published assembly of
226 the sexual species *T. cristinae* (v1.3 from (*34*), see Supplementary Text). This
227 revealed that different types of polymorphisms (SNPs and SVs) tended to co-occur
228 across the genomes in all species, independently of reproductive mode (Fig. 3).
229
230 The focal population for three of the five parthenogenetic species (*T. genevievae*, *T.*
231 *tahoe* and *T. shepardi*) consisted largely of a single genotype with only minor variation
232 among individuals. By contrast, genotype diversity was considerable in *T. monikensis*
233 and *T. douglasi* (Fig. 3A). In the former species, there was further a conspicuous
234 diversity peak on LG8, supporting the idea that parthenogenesis is automictic in
235 *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 diversity levels in *T. monikensis* and *T. douglasi* were comparable to the diversities in populations of some of the sexual *Timema* species (Fig. 3A). Different mechanisms could contribute to such unexpected diversities in parthenogenetic *Timema*, including the presence of lineages that derived independently from their sexual ancestor, or rare sex. While a single transition to parthenogenesis is believed to have occurred in *T. monikensis*, the nominal species *T. douglasi* is polyphyletic and known to consist of independently derived clonal lineages. These lineages have broadly different geographic distributions but can overlap locally (*19*). Identifying the causes of genotypic variation in these species, including the possibility of rare sex, requires further investigation and is a challenge for future studies.

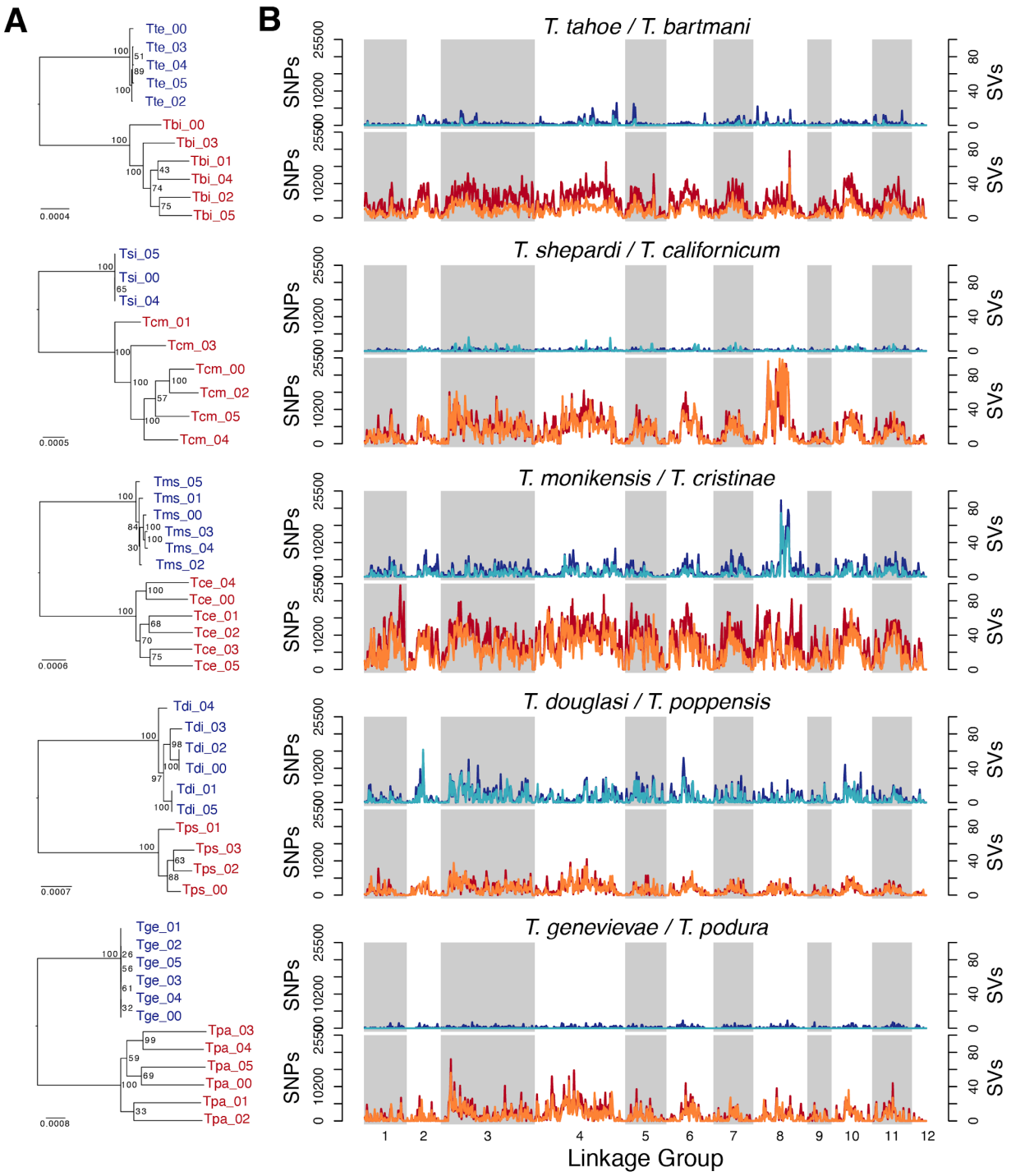


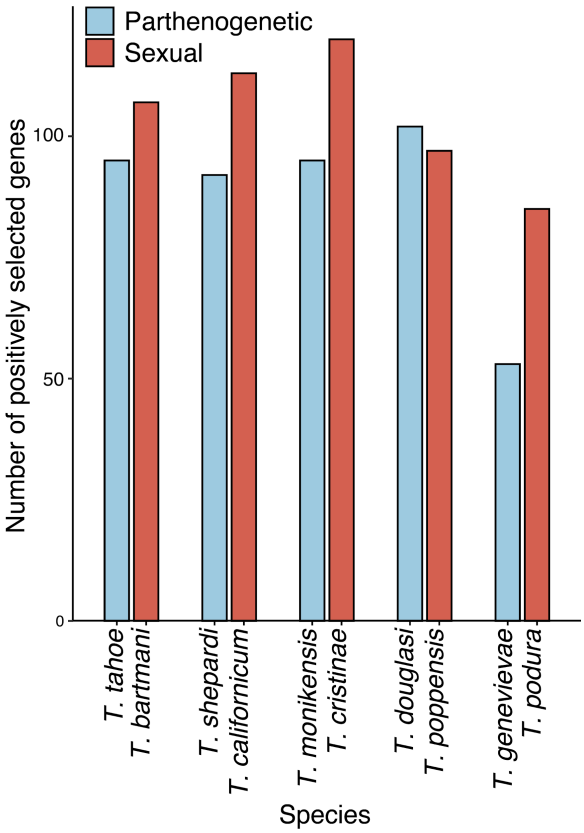
Fig. 3. Population polymorphism levels in parthenogenetic (blue) and sexual (red) *Timema* species. **A.** Phylogenies based on 1:1 orthologous genes reflect the different levels of genotype diversities in parthenogenetic *Timema* species **B.** Distribution of structural variants (SVs; dark blue and 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* (see Supplementary Text).

257 Independently of the mechanisms underlying polymorphism in the parthenogenetic
258 species *T. monikensis,* the polymorphism peak on LG8 is striking (Fig. 3B). This peak
259 occurs in a region previously shown to determine color morph (green, green-striped,
260 or brownish ("melanistic")) in the sexual sister species of *T. monikensis*, *T. cristinae*
261 *(34)*. Our focal *T. monikensis* population features four discrete color morphs (green,
262 dark brown, yellow, and beige), suggesting that additional color morphs may be
263 regulated by the region identified in *T. cristinae*. We also found a peak in polymorphism
264 on LG8, spanning over approximately two-thirds of LG8, in the sexual species *T.*
265 *californicum*, which features a different panel of color morphs than *T. cristinae (35)*.
266 Interestingly, this diversity peak in *T. californicum* was generated by the presence of
267 two divergent haplotypes (approximately 24Mbp long), with grey individuals
268 homozygous for one haplotype and green individuals heterozygous or homozygous
269 for the alternative haplotype (see Supplementary Text). Note that the grey color morph
270 is not known in the monomorphic green parthenogenetic sister of *T. californicum* (*T.*
271 *shepardi*), and we therefore do not expect the same pattern of polymorphism on LG8
272 in this species.

273 **Faster rate of adaptive evolution in sexual than parthenogenetic species**

274 We have shown previously that parthenogenetic *Timema* species accumulate
275 deleterious mutations faster than sexual species (*36*, *37*), a pattern also reported in
276 other parthenogenetic taxa (reviewed in (*17*, *38*)). This is expected given that
277 increased linkage among loci in parthenogens reduces the ability of selection to act
278 individually on each locus, which generates different forms of selective interference (*9*,
279 *10*, *39*). In addition to facilitating the accumulation of deleterious mutations, selective
280 interference among loci in parthenogens should also constrain the efficiency of
281 positive selection. While there is accumulating evidence for this process in
282 experimental evolution studies (e.g., (*40–42*)), its impact on natural populations
283 remains unclear (*17*, *38*). To compare the efficiency of positive selection in sexual and
284 parthenogenetic *Timema,* we used a branch-site model on the gene trees ((*43*),
285 Methods). We compared the terminal branches leading to sexual or parthenogenetic
286 species in one-to-one orthologous genes identified in at least three species pairs (Data
287 S1), using a threshold of $q < 0.05$ to classify which terminal branches show evidence
288 of positive selection.
289

290 We found a greater number of positively selected genes in sexual than
291 parthenogenetic species (Fig. 4, binomial GLMM p = 0.005). In addition, we also
292 examined if there was more evidence for positive selection in sexual species in a
293 threshold-free way by comparing the likelihood ratio test statistic between
294 parthenogenetic and sexual species (as in (*44*, *45*)). This confirmed that the evidence
295 for positive selection was stronger for sexual species (permutation glm p = 0.011).
296



297
298 **Fig. 4.** Number of genes showing evidence for positive selection in each species (total
299 number of genes = 7155) . In addition to reproductive mode, species pair also had a
300 significant influence on the number of positively selected branches (binomial GLMM p
301 = 0.015). There was no significant interaction between species pair and reproductive
302 mode (p = 0.197). Note, the difference between reproductive modes is robust to a
303 more stringent cutoff (q < 0.01 instead of 0.05, Fig. S2A), and if genes with polymorphic
304 positively-selected sites were excluded (Fig. S2B).
305
306 The positively selected genes we identified are most likely associated with species-
307 specific adaptations. Few of them were shared between species, with overlap between
308 species not greater than expected by chance (Fig. S3, FDR < 0.4), and there was little

309 enrichment of functional processes in positively selected genes (0-19 GO terms per
310 species, Table S5). Interestingly, most of the significant GO terms were associated
311 with positively selected genes in parthenogenetic *Timema* (Table S5), likely because
312 a much smaller proportion of positively selected genes in sexual species had
313 annotations (Fig. S4). We speculate that positively selected genes in sexuals could
314 often be involved in sexual selection and species recognition. Indeed, genes
315 associated with processes such as pheromone production and reception often evolve
316 very fast, which makes them difficult to annotate through homology-based inference
317 (*46*). For the parthenogenetic species, although some terms could be associated with
318 their mode of reproduction (e.g. GO:0033206 meiotic cytokinesis in *T. douglasi*), most
319 are not clearly linked to a parthenogenetic life cycle.

320 **Transposable element loads are similar between species with sexual and**
321 **parthenogenetic reproduction**

322

323 Upon the loss of sexual reproduction, transposable element (TE) dynamics are
324 expected to change (*14*, *16*, *47*). How these changes affect genome-wide TE loads is
325 however unclear as sex can facilitate both the spread and the elimination of TEs (*17*).
326 In parthenogens, TE load might initially increase as a result of weaker purifying
327 selection, a pattern well illustrated by the accumulation of TEs in non-recombining
328 parts of sex chromosomes and other supergenes (*48*, *49*). However, TE loads in
329 parthenogens are expected to decrease over time via at least two non-mutually
330 exclusive mechanisms. First, TEs are expected to evolve lower activity over time as
331 their evolutionary interests are aligned with their hosts (*14*, *47*). Second, TE copies
332 that were purged via excision can re-colonize a sexual but not a parthenogenetic
333 genomic background (*15*, *16*). Finally, it is important to note that the predicted effects
334 of reproductive mode on TE loads require some amount of TE activity (transposition
335 or excision). Without such activity, TE content does not vary among individuals and is
336 therefore not expected to change over time.

337

338 We generated a *Timema* genus-level TE library by merging *de novo* TE libraries
339 generated separately for each of the ten *Timema* species. We then quantified TE loads
340 in each *Timema* genome by mapping reads to this merged library (see Methods). The
341 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; Fig. 5; Fig. S5).

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 or that transposition and excision rates are in balance. We find that the TE divergence patterns in *Timema* are best explained by the absence of recent TE activity (see Supplementary Text). Inactive TEs might facilitate the persistence of incipient parthenogenetic strains (*17*) and thus help to explain the high frequency of established parthenogenetic species in *Timema*.

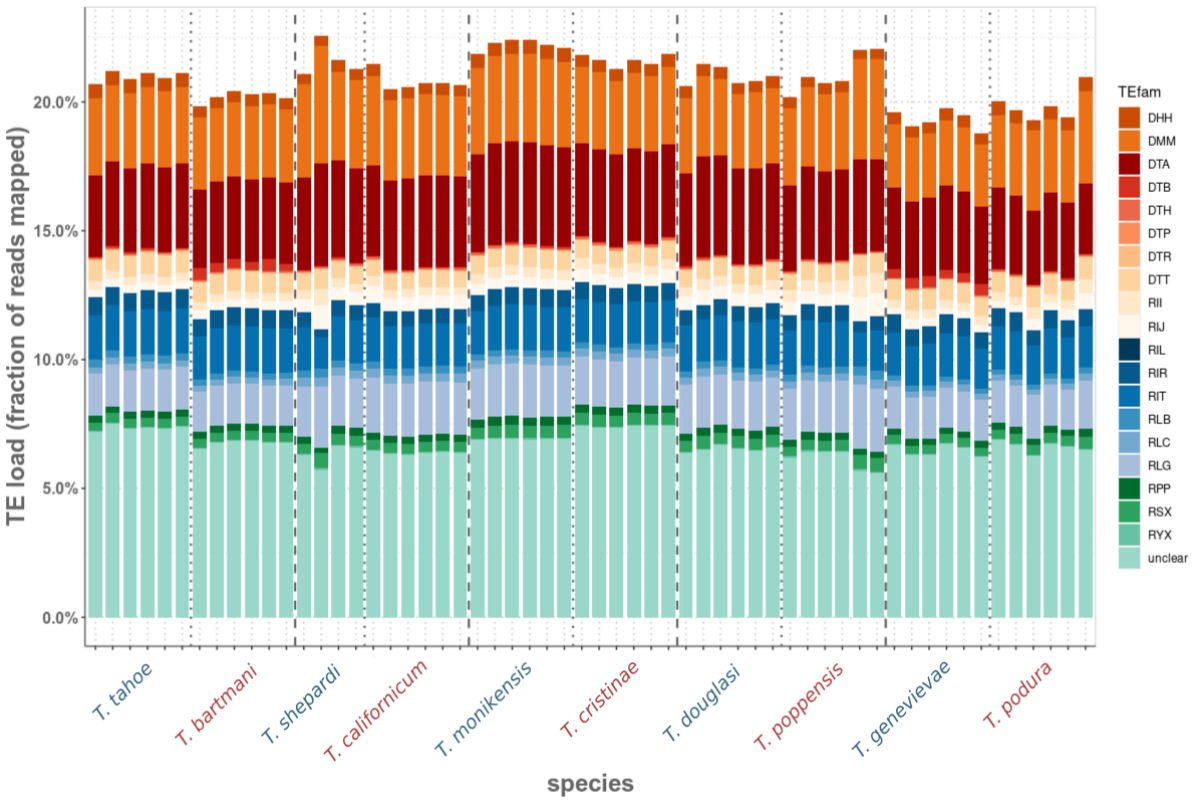


Fig. 5. Total TE abundance in the ten *Timema* species. TE abundance in each genome is expressed as the fraction of reads that map to a genus-level TE library (replicates within species correspond to the reference genome and three to five re-sequenced individuals). TE families are named following the Wicker classification (*50*). 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

362 the third to the Superfamily (e.g. *Gypsy*); for example, RLG is a *Gypsy* retroelement.
363 The character X indicates unknown classification at the superfamily level (because of
364 fragmentation or lack of detectable homology).
365

366 # Conclusion

367

368 We present genomes of five independently derived parthenogenetic lineages of
369 *Timema* stick insects, together with their five sexual sister species. This design with
370 replicated species pairs allows us, for the first time, to disentangle consequences of
371 parthenogenesis from species-specific effects. All parthenogenetic *Timema* species
372 are largely or completely homozygous for both SNPs and SVs, and frequently feature
373 lower levels of population polymorphism than their close sexual relatives. Low
374 population polymorphism can exacerbate the effects of linkage for reducing the
375 efficacy of selection, resulting in reduced rates of positive selection in parthenogenetic
376 *Timema,* in addition to the accumulation of deleterious mutations previously
377 documented (*36*). In spite of these negative genomic consequences, parthenogenesis
378 is an unusually successful strategy in *Timema*. It evolved and persisted repeatedly in
379 the genus, and parthenogenetic species often occur across large geographic areas.
380 Because *Timema* are wingless and their populations subjected to frequent extinction-
381 recolonization dynamics in their fire-prone Californian shrubland habitats, the genomic
382 costs of parthenogenesis are likely offset by one of the most classical benefits of
383 parthenogenesis: the ability to reproduce without a mate.

384

385

386 # Methods

387 **Sample collection and sequencing**

388 For each of the ten species, the DNA for Illumina shotgun sequencing was derived
389 from virgin adult females collected in 2015 from natural populations in California (Table
390 S1). Extractions were done using the Qiagen Mag Attract de HMW DNA kit, following
391 manufacturer indications. Five PCR-free libraries were generated for each reference
392 genome (three 2x125bp paired end libraries with average insert sizes of respectively
393 350, 550 and 700bp, and two mate-pair libraries with 3000 and 5000bp insert sizes),
394 one library (550bp insert size) was generated for each re-sequenced individual.
395 Libraries were prepared using the illumina TruSeq DNA PCR-Free or Nextera Mate
396 Pair Library Prep Kits, following manufacturer instructions, and sequenced on the
397 Illumina HiSeq 2500 system, using v4 chemistry and 2x 125 bp reads at FASTERIS
398 SA, Plan-les-Ouates, Switzerland and the Lausanne Genomic Technologies Facility,
399 Switzerland.

400 **Genome assembly and annotation**

401 The total coverage for the reference genomes (all libraries combined) ranged between
402 37-45x (Table S2). Trimmed paired-end reads were assembled into contigs using
403 ABySS (*51*) and further scaffolded using paired-end and mate pairs using BESST (*52*).
404 Scaffolds identified as contaminants were filtered using Blobtools (*53*). The assembly
405 details can be found in supplementary materials (see Supplementary Text).
406
407 Publically available RNA-seq libraries for *Timema (36, 54, 55)* were used as
408 expression evidence for annotation. Trimmed reads were assembled using Trinity
409 v2.5.1 (*56*) to produce reference-guided transcriptomes. The transcriptomes and
410 protein evidence were combined with *ab initio* gene finders to predict protein coding
411 genes using MAKER v2.31.8 (*57*). The annotation details can be found in the
412 supplementary materials (see Supplementary Text).

413 **Orthologs**

414 *Timema* orthologous groups (OGs) were inferred with the OrthoDB standalone
415 pipeline (v. 2.4.4) using default parameters (*58*). In short, genes are clustered with a

416 graph-based approach based on all best reciprocal hits between each pair of
417 genomes. The high level of fragmentation typical for Illumina-based genomes
418 constrains the ability to identify 1:1 orthologs across all ten *Timema* species. To
419 maximize the number of single copy OGs covering all ten *Timema* species,
420 transcriptomes were included during orthology inference. Thus, transcripts were used
421 to complete OGs in absence of a gene from the corresponding species. Using this
422 approach, 7157 single copy OGs covering at least three sexual-parthenogenetic sister
423 species pairs were obtained (Data S1).

424 **Horizontal gene transfers (HGT)**

425 To detect HGT from non-metazoan species, we first used the pipeline of foreign
426 sequence detection developed by Francois et al. (*59*). We used the set of CDS
427 identified in publicly available transcriptomes (*36*) and the genome assemblies prior to
428 the decontamination procedure with Blobtools (*53*). The rationale is that some genuine
429 HGT could have been wrongly considered as contaminant sequences during this
430 decontamination step and thus been removed from the assembly. Scaffolds filtered
431 during decontamination are available from our github repository
432 (https://github.com/AsexGenomeEvol/Timema_asex_genomes/tree/main/4_Horizont
433 al_Gene_Transfers/contamination_sequences), and will be archived upon
434 acceptance.

435 Briefly, a DIAMOND BlastP (v0.8.33) (*60*) allows us to detect candidate non-metazoan
436 genes in the set of CDS of each species. Taxonomic assignment is based on the 10
437 best blast hits to account for potential contaminations and other sources of taxonomic
438 misassignment in the reference database. Candidate non-metazoan sequences are
439 then subjected to a synteny-based screen with Gmap (v2016-11-07) (*61*) to
440 discriminate between contaminant sequences and potential HGT-derived sequences.
441 A sequence is considered as a HGT candidate if it is physically linked to (i.e., mapped
442 to the same scaffold as) at least one “confident-arthropod” CDS (previously identified
443 in the DIAMOND blast).

444 We then clustered all HGT candidates identified in each of the 10 *Timema* species into
445 HGT families using Silix (v1.2.10) (*62*), requiring a minimum of 85% identity (default
446 parameters otherwise). These HGT families were then “completed” as much as

447 possible by adding homologs from the genome assemblies not identified as HGT
448 candidates (this could occur if the corresponding sequences are fragmented or on
449 short scaffolds for example). To this end, the longest sequence of each HGT family
450 was mapped (using Gmap) on the genomic scaffolds of all species, requiring a
451 minimum of 85% identity.

452 For each completed HGT family, a protein alignment of the candidate HGT
453 sequence(s) and its (their) 50 best DIAMOND blastP hits in the reference database
454 (1st step of the pipeline) was generated with MAFFT (v7) (*63*). The alignments were
455 cleaned using HMMcleaner (stringency parameter = 12) (*64*) and sites with more than
456 50% missing data were removed. Phylogenetic trees were inferred using RAxML
457 (v8.2) (*65*) with the model 'PROTGAMMALGX' of amino-acid substitution and 100
458 bootstrap replicates. Phylogenetic trees were inspected by eye to confirm whether or
459 not an evolutionary history was consistent with the hypothesis of HGT.

460 **Heterozygosity**

461

462 Genome-wide nucleotide heterozygosity was estimated using genome profiling
463 analysis of raw reads from the reference genomes using GenomeScope (v2) (*23*). A
464 second, SNP-based heterozygosity estimate was generated using re-sequenced
465 individuals. We re-sequenced five individuals per species, but 3 individuals of *T.*
466 *shepardi*, 2 individuals of *T. poppensis* and one *T. tahoe* individual did not pass quality
467 control and were discarded from all downstream analyses. SNP calling was based on
468 the GATK best practices pipeline (*66*). We used a conservative set of SNPs with
469 quality scores ≥300, and supported by 15x coverage in at least one of the individuals.
470 SNP heterozygosity was then estimated as the number of heterozygous SNPs divided
471 by the number of callable sites in each genome. Due to stringent filtering criteria, our
472 SNP based heterozygosity is an underestimation of genome-wide heterozygosity.

473 **Structural variants**

474 We used Manta (v1.5.0) (*67*), a diploid-aware pipeline for structural variant (SV)
475 calling, in the same set of re-sequenced individuals used for SNP heterozygosity
476 estimates. We found a high frequency of heterozygous SVs with approximately twice

477 the expected coverage (Fig. S6), which likely represent false positives. To reduce the
478 number of false positives, we filtered very short SVs (30 bases or less) and kept only
479 variant calls that had either split read or paired-end read support within the expected
480 coverage range, where the coverage range was defined individually for each sample
481 by manual inspection of coverage distributions. The filtered SV calls were
482 subsequently merged into population SV calls using SURVIVOR (v1.0.2) (*68*). The
483 merging criteria were: SV calls of the same type on the same strand with breakpoints
484 distances shorter than 100 bp.

485 **Genome alignment**

486 We anchored our genome assemblies to the reference of *T. cristinae* (BioProject
487 Accession PRJNA417530) (*34*) using MUMmer (version 4.0.0beta2) (*69*) with
488 parameter --mum. The alignments were processed by other tools within the package:
489 show-coords with parameters -THrcl to generate tab-delimited alignment files and
490 dnadiff to generate 1-to-1 alignments. We used only uniquely anchored scaffolds for
491 which we were able to map at least 10k nucleotides to the *T. cristinae* reference
492 genome.

493 **Transposable elements**

494 For each species, specific repeat libraries were constructed and annotated to the TE
495 superfamily level (*50*) wherever possible. For collecting repetitive sequences, we used
496 a raw read based approach DNAPipeTE v1.2 (*70*) with parameters -
497 genome_coverage 0.5 -sample_number 4 and respective species genome size, as
498 well as an assembly based approach (RepeatModeler v1.0.8 available at
499 http://www.repeatmasker.org/RepeatModeler/), such that repeats not present in the
500 assembly can still be represented in the repeat library. The two raw libraries were
501 merged and clustered by 95% identity (the TE family threshold) using usearch
502 v10.0.240 (*71*) with the centroid option. To annotate TEs larger than 500 bp in the
503 repeat library, we used an approach that combines homology and structural evidence
504 (PASTEClassifier (*72*)). Because PASTEClassifier did not annotate to TE superfamily
505 levels, we additionally compared by BlastN (v. 2.7.1+) (*73*) the repeat libraries to the
506 well curated *T. cristinae* TE library from Soria-Carrasco et al. (*21*). Blast hits were
507 filtered according to TE classification standards: identity percentage >80%, alignment
508 length >80 bp, and the best hit per contig was kept. The two classification outputs were

509 compared and in case of conflict the classification level of PASTEClassifier was
510 preferred. All non-annotated repeats were labelled ‘unknown’. Repeat library header
511 naming was done according to RepeatMasker standard, but keeping the Wicker
512 naming for elements (i.e., Wicker#Repeatmasker, e.g., DTA#DNA/hAT). TE libraries
513 were sorted by header and TE annotations to similar families numbered consecutively.
514 Species-specific TE libraries were merged into a genus-level *Timema* TE library to
515 account for any TE families that might have not been detected in the single species
516 assemblies.
517
518 To estimate the TE load of reference genomes and resequenced individuals, we first
519 repeat masked the assemblies with the genus-level TE library using RepeatMasker
520 v4.1.0 with parameters set as -gccalc -gff -u -a -xsmall -no_is -div 30 -engine rmblast
521 (*74*). Second, we mapped the 350 bp insert paired-end reads back to the reference
522 genome assemblies using BWA-MEM v0.7.17 (*75*) with standard parameters. We then
523 counted the fraction of reads mapping to TEs out of total mappable reads by counting
524 the number of reads that mapped to each genomic location annotated as TE using
525 htseq-counts (v0.6.1.1p1) (*76*) with parameters set to -r name -s no -t similarity -i
526 Target --nonunique using the mapped read alignments and the gff output of
527 RepeatMasker (filtered for TE length of >80 bp). TE loads were compared among
528 species using a permutation ANOVA with 5000 bootstrap replicates.
529
530 To generate the TE activity landscapes, utilities scripts calcDivergenceFromAlign.pl
531 and createRepeatLandscape_mod.pl were used on the outputs of Repeatmasker
532 v4.1.0. The createRepeatLandscape_mod.pl was modified to match the TE families
533 found in *Timema*.

534 **Positive selection analysis**

535 Only one-to-one orthologs in at least three pairs of species (sister-species sex-asex)
536 were used. The species phylogeny was imposed on every gene as the "gene tree".
537 We used a customized version of the Selectome pipeline (*77*). All alignment building
538 and filtering was performed on predicted amino acid sequences, and the final amino
539 acid MSAs (multiple sequence alignments) were used to infer the nucleotide MSAs
540 used for positive selection inference. MSAs were obtained by MAFFT (v. 7.310) (*63*)
541 with the allowshift option, which avoids over-aligning non homologous regions (e.g.

542 gene prediction errors, or alternative transcripts). All the next steps "mask" rather than
543 remove sites, by replacing the amino acid with a 'X' and the corresponding codon with
544 'NNN'. MCoffee (v11.00.8cbe486) (*78*) was run with the following aligners: mafft_msa,
545 muscle_msa, clustalo_msa (*79*), and t_coffee_msa (*80*). MCoffee provides a
546 consistency score per amino acid, indicating how robust the alignment is at that
547 position for that sequence. Residues with a consistency score less than 5 were
548 masked. TrimAl (v. 1.4.1) (*81*) was used to mask columns with less than 4 residues
549 (neither gap nor 'X'). Following this step 2 of the 7157 ortholog alignments consisted
550 only of gaps and were excluded from further analyses.
551
552 The branch-site model with rate variation at the DNA level (*43*) was run using the
553 Godon software (https://bitbucket.org/Davydov/godon/, version 2020-02-17, option
554 BSG --ncat 4). Each branch was tested iteratively, in one run per gene tree. For each
555 branch, we obtain a ΔlnL which measures the evidence for positive selection, a
556 corresponding p-value and associated q-value (estimated from the distribution of p-
557 values over all branches of all genes), and an estimate of the proportion of sites under
558 positive selection if any. All positive selection results, and detailed methods, will be
559 available at https://selectome.org/timema. To determine if the number of positively
560 selected genes differed between sexual and parthenogenetic species we used a
561 binomial GLMM approach (lme4 (*82*)) with q-value threshold of 0.05 or 0.01.
562 Significance of model terms was determined with a Wald statistic. In addition, we also
563 examined if there was more evidence for positive selection in sexual species in a
564 threshold-free way by comparing ΔlnL values between parthenogenetic and sexual
565 species (as in (*44*, *45*)). To do this we used a permutation glm approach where
566 reproductive mode (sexual or parthenogenetic) was randomly switched within a
567 species-pair. To determine if the overlap of positively selected genes was greater than
568 expected by chance we used the SuperExactTest package (v. 0.99.4) (*83*) in R. The
569 resulting p-values were multiple test corrected using Benjamini and Hochberg's
570 algorithm implemented in R. To assess the impact of polymorphism on our results we
571 repeated our analysis after excluding positively selected genes with non-synonymous
572 polymorphic variants at positively selected sites (those with a posterior probability of
573 >0.95 detected by detected by Bayes Empirical Bayes analysis in Godon). Functional
574 enrichment analyses were performed using TopGO (v. 2.28.0) (*84*) using the *D.*
575 *melanogaster* functional annotation (see Supplementary Text). 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 $p < 0.05$.

Acknowledgements

We thank Bart Zijlstra, Chloé Larose, and Kirsten Jalvingh for help in the field, and Deborah Charlesworth, Daniel Wegmann, and current and previous members of the Schwander and Robinson-Rechavi labs for discussions. **Funding:** This project was supported by several grants from the Swiss Science Foundation (CRSII3_160723 as well as PP00P3_170627, 31003A_182495, IZLRZ3_163872, and 407540_167276) and funding from the University of Lausanne. **Author contributions:** T.S., M.R.R., N.G, K.S.J., and D.J.P. designed the study. Z.D. performed molecular work. K.S.J., D.J.P., Y.A., P.T.V., J.B., E.F., C.M.F., K.H., V.R., and P.S. analysed the data with input from T.S., M.R.R., and N.G.. D.J.P., T.S and K.S.J., wrote the paper with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and code availability:** Raw sequence reads have been deposited in NCBI's sequence read archive under the following bioprojects: PRJNA371785 (reference genomes, Table S6), PRJNA670663 (resequenced individuals, Table S7), PRJNA679785 (RNAseq reads used for annotation), and PRJNA673001 (PacBio reads for *T. douglasi*). Genome assemblies and annotations PRJEB31411. Scripts for the analyses in this paper are available at: https://github.com/AsexGenomeEvol/Timema_asex_genomes. Data were processed to generate plots and statistics using R v3.4.4.

References

1. M. Neiman, C. M. Lively, S. Meirmans, Why sex? A pluralist approach revisited. *Trends Ecol. Evol.* **32**, 589–600 (2017).

2. N. P. Sharp, S. P. Otto, Evolution of sex: Using experimental genomics to select among competing theories. *Bioessays*. **38**, 751–757 (2016).

3. G. Bell, *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ of California Press, Los Angeles, 1982).

608 4. G. C. Williams, *Sex and Evolution* (Princeton University Press, Princeton, USA, 1975).

609 5. J. Maynard Smith, *The Evolution of Sex* (Cambridge University Press, Cambridge, UK,
610 1978).

611 6. J. Gerritsen, Sex and Parthenogenesis in Sparse Populations. *Am. Nat.* **115**, 718–742
612 (1980).

613 7. S. K. Jain, The Evolution of Inbreeding in Plants. *Annu. Rev. Ecol. Syst.* **7**, 469–495
614 (1976).

615 8. J. Felsenstein, The evolutionary advantage of recombination. *Genetics*. **78**, 737–756
616 (1974).

617 9. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.*
618 **8**, 269–294 (1966).

619 10. P. D. Keightley, S. P. Otto, Interference among deleterious mutations favours sex and
620 recombination in finite populations. *Nature*. **443**, 89–92 (2006).

621 11. N. H. Barton, Why sex and recombination? *Cold Spring Harb. Symp. Quant. Biol.* **74**,
622 187–195 (2009).

623 12. C. W. Birky Jr, Heterozygosity, heteromorphy, and phylogenetic trees in asexual
624 eukaryotes. *Genetics*. **144**, 427–437 (1996).

625 13. D. Mark Welch, M. Meselson, Evidence for the evolution of bdelloid rotifers without
626 sexual reproduction or genetic exchange. *Science*. **288**, 1211–1215 (2000).

627 14. D. A. Hickey, Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*. **101**, 519–
628 531 (1982).

629 15. E. S. Dolgin, B. Charlesworth, The effects of recombination rate on the distribution and
630 abundance of transposable elements. *Genetics*. **178**, 2169–2177 (2008).

631 16. J. Bast, K. S. Jaron, D. Schuseil, D. Roze, T. Schwander, Asexual reproduction reduces
632 transposable element load in experimental yeast populations. *Elife*. **8** (2019),
633 doi:10.7554/eLife.48548.

634 17. K. S. Jaron, J. Bast, R. W. Nowell, T. R. Ranallo-Benavidez, M. Robinson-Rechavi, T.
635 Schwander, Genomic Features of Parthenogenetic Animals. *J. Hered.* (2020),
636 doi:10.1093/jhered/esaa031.

637 18. T. Schwander, B. J. Crespi, Multiple direct transitions from sexual reproduction to
638 apomictic parthenogenesis in *Timema* stick insects. *Evolution*. **63**, 84–103 (2009).

639 19. T. Schwander, L. Henry, B. J. Crespi, Molecular evidence for ancient asexuality in
640 *Timema* stick insects. *Curr. Biol.* **21**, 1129–1134 (2011).

641 20. R. Riesch, M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K.
642 Lucek, E. Hellen, V. Soria-Carrasco, S. R. Dennis, C. F. de Carvalho, R. J. Safran, C. P.
643 Sandoval, J. Feder, R. Gries, B. J. Crespi, G. Gries, Z. Gompert, P. Nosil, Transitions
644 between phases of genomic differentiation during stick-insect speciation. *Nature*
645 *Ecology & Evolution*. **1**, 0082 (2017).

646 21. V. Soria-Carrasco, Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
647 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, P.

Nosil, Stick insect genomes reveal natural selection's role in parallel speciation. *Science*. **344**, 738–742 (2014).

22. R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, E. M. Zdobnov, BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

23. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).

24. J. Romiguier, P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. A.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, N. Galtier, Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*. **515**, 261–263 (2014).

25. A. Mackintosh, D. R. Laetsch, A. Hayward, B. Charlesworth, M. Waterfall, R. Vila, K. Lohse, The determinants of genetic diversity in butterflies. *Nat. Commun.* **10**, 3466 (2019).

26. T. Schwander, S. Vuilleumier, J. Dubman, B. J. Crespi, Positive feedback in the transition from sexual reproduction to parthenogenesis. *Proc. Biol. Sci.* **277**, 1435–1442 (2010).

27. J. Engelstädter, Asexual but Not Clonal: Evolutionary Processes in Automictic Populations. *Genetics*. **206**, 993–1009 (2017).

28. M. Neiman, T. Schwander, Using Parthenogenetic Lineages to Identify Advantages of Sex. *Evol. Biol.* **38**, 115–123 (2011).

29. S. Glémin, C. M. François, N. Galtier, Genome Evolution in Outcrossing vs. Selfing vs. Asexual Species. *Methods Mol. Biol.* **1910**, 331–369 (2019).

30. M. Pearcy, S. Aron, C. Doums, L. Keller, Conditional use of sex and parthenogenesis for worker and queen production in ants. *Science*. **306**, 1780–1783 (2004).

31. M. O. Lorenzo-Carballa, A. Cordero-Rivera, Thelytokous parthenogenesis in the damselfly Ischnura hastata (Odonata, Coenagrionidae): genetic mechanisms and lack of bacterial infection. *Heredity* . **103**, 377–384 (2009).

32. T. J. Treangen, S. L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).

33. H. Ellegren, N. Galtier, Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).

34. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in Timema stick insects. *Science*. **359**, 765–770 (2018).

35. V. R. Vickery, Revision of *Timema scudder* (Phasmatoptera: Timematodea) including three new species. *Can. Entomol.* **125**, 657–692 (1993).

36. J. Bast, D. J. Parker, Z. Dumas, K. M. Jalvingh, P. Tran Van, K. S. Jaron, E. Figuet, A. Brandt, N. Galtier, T. Schwander, Consequences of asexuality in natural populations: insights from stick insects. *Mol. Biol. Evol.* **35**, 1668–1677 (2018).

690 37. L. Henry, T. Schwander, B. J. Crespi, Deleterious mutation accumulation in asexual
691 Timema stick insects. *Mol. Biol. Evol.* **29**, 401–408 (2012).

692 38. M. Neiman, P. G. Meirmans, T. Schwander, S. Meirmans, Sex in the wild: How and why
693 field-based studies contribute to solving the problem of sex. *Evolution*. **72**, 1194–1203
694 (2018).

695 39. S. P. Otto, Selective Interference and the Evolution of Sex. *J. Hered.* (2020),
696 doi:10.1093/jhered/esaa026.

697 40. M. J. McDonald, D. P. Rice, M. M. Desai, Sex speeds adaptation by altering the
698 dynamics of molecular evolution. *Nature*. **531**, 233–236 (2016).

699 41. O. Kaltz, G. Bell, The ecology and genetics of fitness in Chlamydomonas. XII. Repeated
700 sexual episodes increase rates of adaptation to novel environments. *Evolution*. **56**,
701 1743–1753 (2002).

702 42. M. R. Goddard, H. C. J. Godfray, A. Burt, Sex increases the efficacy of natural selection
703 in experimental yeast populations. *Nature*. **434**, 636–640 (2005).

704 43. I. I. Davydov, N. Salamin, M. Robinson-Rechavi, Large-Scale Comparative Analysis of
705 Codon Models Accounting for Protein and Nucleotide Selection. *Mol. Biol. Evol.* **36**,
706 1316–1332 (2019).

707 44. J. T. Daub, S. Moretti, I. I. Davydov, L. Excoffier, M. Robinson-Rechavi, Detection of
708 Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Mol.*
709 *Biol. Evol.* **34**, 1391–1402 (2017).

710 45. J. Liu, M. Robinson-Rechavi, Adaptive Evolution of Animal Proteins over Development:
711 Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo. *Mol. Biol. Evol.*
712 **35**, 2862–2872 (2018).

713 46. W. Haerty, S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram, L. K. Sirot, L.
714 Levesque, C. G. Artieri, M. F. Wolfner, A. Civetta, R. S. Singh, Evolution in the fast lane:
715 rapidly evolving sex-related genes in Drosophila. *Genetics*. **177**, 1321–1335 (2007).

716 47. B. Charlesworth, C. H. Langley, The evolution of self-regulated transposition of
717 transposable elements. *Genetics*. **112**, 359–383 (1986).

718 48. T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.*
719 **24**, R288–94 (2014).

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

722 50. T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P.
723 Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A. H. Schulman, A unified
724 classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982
725 (2007).

726 51. S. D. Jackman, B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G.
727 Jahesh, H. Khan, L. Coombe, R. L. Warren, I. Birol, ABySS 2.0: resource-efficient
728 assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).

729 52. K. Sahlin, R. Chikhi, L. Arvestad, Assembly scaffolding with PE-contaminated mate-pair
730 libraries. *Bioinformatics*. **32**, 1925–1932 (2016).

731 53. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies.
732 *F1000Res.* **6**, 1287 (2017).

733 54. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander, Sex-
734 biased gene expression is repeatedly masculinized in asexual females. *Nat. Commun.*
735 **10**, 4638 (2019).

736 55. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander,
737 Repeated evolution of asexuality involves convergent gene expression changes. *Mol.*
738 *Biol. Evol.* **36** , 350–364 (2019).

739 56. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B.
740 Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F.
741 Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc,
742 N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq
743 using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–
744 1512 (2013).

745 57. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database
746 management tool for second-generation genome projects. *BMC Bioinformatics*. **12**, 491
747 (2011).

748 58. E. V. Kriventseva, F. Tegenfeldt, T. J. Petty, R. M. Waterhouse, F. A. Simão, I. A.
749 Pozdnyakov, P. Ioannidis, E. M. Zdobnov, OrthoDB v8: update of the hierarchical
750 catalog of orthologs and the underlying free software. *Nucleic Acids Res.* **43**, D250–6
751 (2015).

752 59. C. M. Francois, F. Durand, E. Figuet, N. Galtier, Prevalence and Implications of
753 Contamination in Public Genomic Resources: A Case Study of 43 Reference Arthropod
754 Assemblies. *G3* . **10**, 721–730 (2020).

755 60. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND.
756 *Nat. Methods*. **12**, 59–60 (2015).

757 61. T. D. Wu, C. K. Watanabe, GMAP: a genomic mapping and alignment program for
758 mRNA and EST sequences. *Bioinformatics*. **21**, 1859–1875 (2005).

759 62. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with
760 SiLiX. *BMC Bioinformatics*. **12**, 116 (2011).

761 63. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7:
762 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

763 64. A. Di Franco, R. Poujol, D. Baurain, H. Philippe, Evaluating the usefulness of alignment
764 filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol.*
765 *Biol.* **19**, 21 (2019).

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

768 66. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-
769 Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D.
770 Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant calls:
771 the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. **43**,
772 11.10.1–11.10.33 (2013).

773 67. X. Chen, O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, A. J.
774 Cox, S. Kruglyak, C. T. Saunders, Manta: rapid detection of structural variants and
775 indels for germline and cancer sequencing applications. *Bioinformatics*. **32**, 1220–1222
776 (2016).

777 68. D. C. Jeffares, C. Jolly, M. Hoti, D. Speed, L. Shaw, C. Rallis, F. Balloux, C. Dessimoz,
778 J. Bähler, F. J. Sedlazeck, Transient structural variations have strong effects on
779 quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061
780 (2017).

781 69. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L.
782 Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**,
783 R12 (2004).

784 70. C. Goubert, L. Modolo, C. Vieira, C. ValienteMoro, P. Mavingui, M. Boulesteix, De novo
785 assembly and annotation of the Asian tiger mosquito (Aedes albopictus) repeatome with
786 dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever
787 mosquito (Aedes aegypti). *Genome Biol. Evol.* **7**, 1192–1205 (2015).

788 71. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST.
789 *Bioinformatics*. **26**, 2460–2461 (2010).

790 72. C. Hoede, S. Arnoux, M. Moisset, T. Chaumier, O. Inizan, V. Jamilloux, H. Quesneville,
791 PASTEC: an automatic transposable element classification tool. *PLoS One*. **9**, e91929
792 (2014).

793 73. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman,
794 Gapped BLAST and PSI-BLAST: a new generation of protein database search
795 programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

796 74. A. Smit, R. Hubley, P. Green, *RepeatMasker Open-4.0* (2013-2015;
797 http://www.repeatmasker.org).

798 75. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-
799 MEM. *arXiv [q-bio.GN]* (2013), (available at http://arxiv.org/abs/1303.3997).

800 76. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-
801 throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

802 77. S. Moretti, B. Laurenczy, W. H. Gharib, B. Castella, A. Kuzniar, H. Schabauer, R. A.
803 Studer, M. Valle, N. Salamin, H. Stockinger, M. Robinson-Rechavi, Selectome update:
804 quality control and computational improvements to a database of positive selection.
805 *Nucleic Acids Res.* **42**, D917–21 (2014).

806 78. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: combining multiple
807 sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).

808 79. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam,
809 M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of
810 high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*
811 **7**, 539 (2011).

812 80. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and
813 accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).

814 81. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated

alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973 (2009).

82. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*. **67**, 1–48 (2015).

83. M. Wang, Y. Zhao, B. Zhang, Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).

84. A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. **22**, 1600–1607 (2006).