

1 **Supplementary materials: Convergent consequences of**
 2 **parthenogenesis on stick insect genomes**

3 **SM Table 1. Origin of biological material**

4 All six females per species were taken from a single location at the indicated
 5 coordinates. Red species reproduce sexually (s), blue species via parthenogenesis
 6 (p).

7

Species	Host plant	Coordinates	
		longitude	latitude
<i>T. tahoe</i> (p)	<i>Abies concolor</i>	38.7610110	-120.1600530
<i>T. bartmani</i> (s)	<i>Abies concolor</i>	34.1700000	-117.0020167
<i>T. shepardii</i> (p)	<i>Arctostaphylos</i> sp.	39.1926500	-123.2617833
<i>T. californicum</i> (s)	<i>Quercus</i> sp.	37.3431667	-121.6364667
<i>T. douglasi</i> (p)	<i>Pseudotsuga menziesii</i>	38.9825500	-123.4697500
<i>T. poppensis</i> (s)	<i>Sequoia sempervirens</i>	37.1655167	-122.0155500
<i>T. monikensis</i> (p)	<i>Cercocarpus betuloides</i>	34.1148833	-118.8531333
<i>T. cristinae</i> (s)	<i>Cercocarpus betuloides</i>	34.5362700	-119.2444300
<i>T. genevieveae</i> (p)	<i>Adenostoma fasciculatum</i>	38.9957833	-122.9257667
<i>T. podura</i> (s)	<i>Adenostoma fasciculatum</i>	33.7976020	-116.7769850

8

9 **SM Table 2. Sequencing coverage**

10 Read coverage for the reference assemblies of individual *Timema* species was
 11 estimated using the haploid genome size of *Timema cristinae* of 1.381Gbp (21). Red
 12 species are reproducing sexually, while blue species are parthenogenetic. Is: insert
 13 size [bp].

14

species	paired-end			mate-pair		orphans	Total
	Is 350	Is 550	Is 700	Is 3000	Is 5000		
<i>T. tahoe</i>	15	12.2	5.7	4.1	3	3	43.1
<i>T. bartmani</i>	12.3	13.5	3.7	2.7	2.5	2.4	37.0
<i>T. shepardi</i>	12.4	11.6	8.3	3.8	3.6	2.8	42.7
<i>T. californicum</i>	16.4	13.2	8.1	4.4	2.8	3.1	48.2
<i>T. douglasi</i>	13.2	11	8.8	4.3	2.8	2.9	43.1
<i>T. poppensis</i>	12.5	12.1	7.1	2.9	2.8	2.7	40.2
<i>T. monikensis</i>	13.8	12.6	9.6	3.4	4.2	3	46.6
<i>T. cristinae</i>	13.7	10.9	10	4	3.6	3	45.3
<i>T. genevieveae</i>	15	13.4	4.3	2.5	5.4	2.8	43.5
<i>T. podura</i>	15.7	10.8	3.1	3.1	2.6	2.3	37.7

15

16 **SM Table 3. Genome assembly statistics**

17 Genome assembly statistics of sequenced *Timema* species. Haploid genome size
 18 represents the estimate from genome profiling of raw reads using Genomescope
 19 (23). Total sum represents the sum of all scaffolds. The BUSCO score (22) is the
 20 percentage of conserved single copy orthologs among insects. N is the percentage
 21 of unknown nucleotides (gaps) in the assembly. Genes are the number of annotated
 22 protein coding genes. Red species reproduce sexually, blue species through
 23 parthenogenesis. Although the sequencing coverage was similar across the ten
 24 sequenced species (approximately 40x, SM Table 2), all five parthenogenetic
 25 species had both higher continuity (NG50 61.9 - 147.4 kbp for parthenogens, vs. 2.1
 26 - 76.1 kbp for sexuals) and higher completeness (97.2 - 98.3% BUSCO genes in
 27 parthenogens, vs. 86.4 - 97.2% in sexuals), likely because of systematic differences
 28 in heterozygosity between species with different reproductive modes (see main text).
 29

species	Haploid genome size [Gpb]	Σ [Gpb]	BUSCO [%]	Ns [%]	Genes
<i>T. tahoe</i>	1.13	1.093	97.5	2.4	12771
<i>T. bartmani</i>	1.15	1.109	97.2	2.6	14066
<i>T. shepardi</i>	1.23	1.153	97.2	1.7	14033
<i>T. californicum</i>	1.3	1.220	94.4	1.8	14563
<i>T. douglasi</i>	1.26	1.124	97.2	1.6	13824
<i>T. poppensis</i>	1.31	1.137	93.9	1.6	15605
<i>T. monikensis</i>	1.12	1.099	98.3	1.7	12837
<i>T. cristinae</i>	1.11	1.178	96.9	2.3	13882
<i>T. genevieveae</i>	1.07	1.049	97.9	1.6	12009
<i>T. podura</i>	1.04	1.105	86.4	0.4	16529

30

31 **SM Table 4. Numbers of 1:1 orthologs in different sets of *Timema* species**

32 See external file “SM_Table_4.tsv”.

33

34

35 **SM Table 5. Origin of the genetic variation among genotypes in**
 36 **parthenogenetic populations**

37

38 To distinguish between putative ancestral polymorphisms (shared between sexual
 39 and parthenogenetic species) and polymorphisms that appeared in the
 40 parthenogenetic lineage after the split from the sexual lineage, we used the SNPs
 41 generated for heterozygosity estimates via GATK best practices pipeline (67) (see
 42 Methods) but with less stringent downstream filtering (min 10x coverage).
 43 Homologous SNPs within a species pair were identified with MUMmer v4.0.0beta2
 44 (nucmer and dnadiff with default parameters to keep only unique alignments of
 45 genome segments, and custom scripts to discard overlapping alignments), using the
 46 genome of the parthenogenetic species as the reference and the one from its sexual
 47 relative as the query.

48

Species pair	Number of positions analyzed	Variable (within and/or between species)	Same variants in both species	Different variants	Variable only in sexual species	Variable only in partheno-genetic species
<i>T. bartmani</i> <i>T. tahoe</i>	852224058	8945655	26137	3683201	5052559	183758
<i>T. californicum</i> <i>T. shepardi</i>	725333178	12631427	51243	4752391	7604702	223091
<i>T. cristinae</i> <i>T. monikensis</i>	816642553	19793873	87370	7078329	11677904	950270
<i>T. poppensis</i> <i>T. douglasi</i>	781906596	14109700	206188	8296511	3989068	1617933
<i>T. podura</i> <i>T. genevieveae</i>	636577084	27365408	325	5947476	21410167	7440

50 **SM Table 6. Number of RNA-seq libraries used for genome annotation**51 Species are abbreviated as follows: Tbi = *T. bartmani*, Tce = *T. cristinae*, Tps = *T.*52 *poppensis*, Tcm = *T. californicum*, Tpa = *T. podura*, Tte = *T. tahoe*, Tms = *T.*53 *monikensis*, Tdi = *T. douglasi*, Tsi = *T. shepardii*, and Tge = *T. genevieveae*

54

Tissue	Library type	Tbi	Tte	Tce	Tms	Tcm	Tsi	Tpa	Tge	Tps	Tdi
Whole-Body (Female)	Single-end	6	3	6	3	6	3	6	3	6	3
Whole-Body (Male)	Single-end	3		3		3		3		3	
Rep. tract (Female)	Single-end	3	3	3	3	3	3	3	3	3	3
Rep. tract (Male)	Single-end	3		3		3		3		3	
Heads (Female)	Single-end	3	3	3	3	3	3	3	3	3	3
Heads (Male)	Single-end	3		3		3		3		3	
Legs (Female)	Single-end	3	3	3	3	3	3	3	3	3	3
Legs (Male)	Single-end	3		3		3		3		3	
Juvenile (Female)	Paired-end					3	3				
Juvenile (Male)	Paired-end					3					
Hatchling (Unknown)	Paired-end			7	3	6	3	5			3

55

56

57 **SM Table 7A. Accession numbers for raw reads of reference individuals**

58 Species are abbreviated as follows: Tbi = *T. bartmani*, Tce = *T. cristinae*, Tps = *T.*

59 *poppensis*, Tcm = *T. californicum*, Tpa = *T. podura*, Tte = *T. tahoe*, Tms = *T.*

60 *monikensis*, Tdi = *T. douglasi*, Tsi = *T. shepardii*, and Tge = *T. genevieveae*

61

Library Name	Sp ID	Sample ID	Insert size	SRA sample accession	SRA run accession	Assembly	Genome profiling	Variants
HYI-7_125	4_Tte	Tte_00	350	SRS1972401	SRR5248900	*	*	
HYI-7_150	4_Tte	Tte_00	350	SRS1972401	SRR5248899		*	
HYI-17	4_Tte	Tte_00	550	SRS1972401	SRR5248898	*	*	*
HYI-51	4_Tte	Tte_00	700	SRS1972401	SRR5248897	*	*	
HYI-37	4_Tte	Tte_00	3000	SRS1972401	SRR5248896	*		
HYI-47	4_Tte	Tte_00	5000	SRS1972401	SRR5248895	*		
HYI-18_125	4_Tbi	Tbi_00	350	SRS1972400	SRR5248892	*	*	
HYI-8_125	4_Tbi	Tbi_00	350	SRS1972400	SRR5248894		*	
HYI-8_150	4_Tbi	Tbi_00	350	SRS1972400	SRR5248893	*	*	*
HYI-28	4_Tbi	Tbi_00	700	SRS1972400	SRR5248891	*	*	
HYI-38	4_Tbi	Tbi_00	3000	SRS1972400	SRR5248890	*		
HYI-48	4_Tbi	Tbi_00	5000	SRS1972400	SRR5248889	*		
HYI-4_125	2_Tsi	Tsi_00	350	SRS1972405	SRR5248924	*	*	
HYI-4_150	2_Tsi	Tsi_00	350	SRS1972405	SRR5248923		*	
HYI-14	2_Tsi	Tsi_00	550	SRS1972405	SRR5248922	*	*	*
HYI-24	2_Tsi	Tsi_00	700	SRS1972405	SRR5248921	*	*	
HYI-34	2_Tsi	Tsi_00	3000	SRS1972405	SRR5248920	*		
HYI-44	2_Tsi	Tsi_00	5000	SRS1972405	SRR5248919	*		
HYI-3_125	2_Tcm	Tcm_00	350	SRS1972404	SRR5248918	*	*	
HYI-3_150	2_Tcm	Tcm_00	350	SRS1972404	SRR5248917		*	
HYI-13	2_Tcm	Tcm_00	550	SRS1972404	SRR5248916	*	*	*
HYI-23	2_Tcm	Tcm_00	700	SRS1972404	SRR5248915	*	*	
HYI-33	2_Tcm	Tcm_00	3000	SRS1972404	SRR5248914	*		
HYI-43	2_Tcm	Tcm_00	5000	SRS1972404	SRR5248913	*		
HYI-5_125	3_Tms	Tms_00	350	SRS1972403	SRR5248912	*	*	
HYI-5_150	3_Tms	Tms_00	350	SRS1972403	SRR5248911		*	
HYI-15	3_Tms	Tms_00	550	SRS1972403	SRR5248910	*	*	*
HYI-25	3_Tms	Tms_00	700	SRS1972403	SRR5248909	*	*	
HYI-35	3_Tms	Tms_00	3000	SRS1972403	SRR5248908	*		
HYI-45	3_Tms	Tms_00	5000	SRS1972403	SRR5248907	*		
HYI-6_125	3_Tce	Tce_00	350	SRS1972402	SRR5248906	*	*	
HYI-6_150	3_Tce	Tce_00	350	SRS1972402	SRR5248905		*	
HYI-16	3_Tce	Tce_00	550	SRS1972402	SRR5248904	*	*	*
HYI-26	3_Tce	Tce_00	700	SRS1972402	SRR5248903	*	*	
HYI-36	3_Tce	Tce_00	3000	SRS1972402	SRR5248902	*		
HYI-46	3_Tce	Tce_00	5000	SRS1972402	SRR5248901	*		
HYI-1_125	1_Tdi	Tdi_00	350	SRS1972407	SRR5248936	*	*	
HYI-1_150	1_Tdi	Tdi_00	350	SRS1972407	SRR5248935		*	
HYI-11	1_Tdi	Tdi_00	550	SRS1972407	SRR5248934	*	*	*
HYI-21	1_Tdi	Tdi_00	700	SRS1972407	SRR5248933	*	*	

HYI-31	1_Tdi	Tdi_00	3000	SRS1972407	SRR5248932	*		
HYI-41	1_Tdi	Tdi_00	5000	SRS1972407	SRR5248931	*		
HYI-2_125	1_Tps	Tps_00	350	SRS1972406	SRR5248930	*	*	
HYI-2_150	1_Tps	Tps_00	350	SRS1972406	SRR5248929		*	
HYI-12	1_Tps	Tps_00	550	SRS1972406	SRR5248928	*	*	*
HYI-22	1_Tps	Tps_00	700	SRS1972406	SRR5248927	*	*	
HYI-32	1_Tps	Tps_00	3000	SRS1972406	SRR5248926	*		
HYI-42	1_Tps	Tps_00	5000	SRS1972406	SRR5248925	*		
HYI-10_125	5_Tge	Tge_00	350	SRS1972399	SRR5248888	*	*	
HYI-10_150	5_Tge	Tge_00	350	SRS1972399	SRR5248887		*	
HYI-20	5_Tge	Tge_00	550	SRS1972399	SRR5248886	*	*	*
HYI-53	5_Tge	Tge_00	700	SRS1972399	SRR5248885	*	*	
HYI-40	5_Tge	Tge_00	3000	SRS1972399	SRR5248884	*		
HYI-50	5_Tge	Tge_00	5000	SRS1972399	SRR5248883	*		
HYI-9_125	5_Tpa	Tpa_00	350	SRS1972398	SRR5248882	*	*	
HYI-9_150	5_Tpa	Tpa_00	350	SRS1972398	SRR5248881		*	
HYI-19	5_Tpa	Tpa_00	550	SRS1972398	SRR5248880	*	*	*
HYI-52	5_Tpa	Tpa_00	700	SRS1972398	SRR5248879	*	*	
HYI-39	5_Tpa	Tpa_00	3000	SRS1972398	SRR5248878	*		
HYI-49	5_Tpa	Tpa_00	5000	SRS1972398	SRR5248877	*		

63 **SM Table 7B. Accession numbers for raw reads of resequenced individuals**

64 Species are abbreviated as follows: Tbi = *T. bartmani*, Tce = *T. cristinae*, Tps = *T.*
 65 *poppensis*, Tcm = *T. californicum*, Tpa = *T. podura*, Tte = *T. tahoe*, Tms = *T.*
 66 *monikensis*, Tdi = *T. douglasi*, Tsi = *T. shepardi*, and Tge = *T. genevieveae*
 67

Library Name	Species ID	Sample ID	SRA sample accession	SRA run accession
ReSeq_Te07	4_Tte	Tte_01	SRS7638306	SRR12928425, SRR12928426, SRR12928429-SRR12928438, SRR12928440-SRR12928449
ReSeq_Te08	4_Tte	Tte_02	SRS7638305	SRR12928399-SRR12928404, SRR12928406-SRR12928415, SRR12928417-SRR12928424
ReSeq_Te09	4_Tte	Tte_03	SRS7638326	SRR12928367-SRR12928371, SRR12928373-SRR12928382, SRR12928384-SRR12928393, SRR12928395-SRR12928398
ReSeq_Te10	4_Tte	Tte_04	SRS7638327	SRR12928340-SRR12928349, SRR12928351-SRR12928360, SRR12928362-SRR12928366
ReSeq_Te11	4_Tte	Tte_05	SRS7638328	SRR12928311-SRR12928315, SRR12928318-SRR12928327, SRR12928329-SRR12928338
CC86B	4_Tbi	Tbi_01	SRS7637496	SRR12928843-SRR12928847, SRR12928849-SRR12928858, SRR12928860
CC86C	4_Tbi	Tbi_02	SRS7637495	SRR12928821-SRR12928824, SRR12928826-SRR12928835, SRR12928838-SRR12928842
CC87B	4_Tbi	Tbi_03	SRS7637498	SRR12928490-SRR12928493, SRR12928495-SRR12928504, SRR12928506-SRR12928515,

				SRR12928517-SRR12928520, SRR12928818, SRR12928820
CC87C	4_Tbi	Tbi_04	SRS7638307	SRR12928468-SRR12928471, SRR12928473-SRR12928482, SRR12928484-SRR12928489
CC88B	4_Tbi	Tbi_05	SRS7638309	SRR12928451-SRR12928460, SRR12928462-SRR12928467
ReSeq_Si01	2_Tsi	Tsi_01	SRS7638289	SRR12928651-SRR12928659, SRR12928661-SRR12928663
ReSeq_S14	2_Tsi	Tsi_02	SRS7638288	SRR12928664-SRR12928670, SRR12928672-SRR12928676
ReSeq_Si03	2_Tsi	Tsi_03	SRS7638287	SRR12928635-SRR12928637, SRR12928639-SRR12928648, SRR12928650
ReSeq_Si16	2_Tsi	Tsi_04	SRS7638284	SRR12928621-SRR12928626, SRR12928628-SRR12928634
ReSeq_Si18	2_Tsi	Tsi_05	SRS7638285	SRR12928604, SRR12928606-SRR12928615, SRR12928617-SRR12928620
HM217	2_Tcm	Tcm_01	SRS7638279	SRR12928757-SRR12928759, SRR12928761-SRR12928770, SRR12928772-SRR12928778
HM218	2_Tcm	Tcm_02	SRS7638277	SRR12928735-SRR12928737, SRR12928739-SRR12928748, SRR12928750-SRR12928756
HM219	2_Tcm	Tcm_03	SRS7638281	SRR12928713-SRR12928715, SRR12928717-SRR12928726, SRR12928728-SRR12928734
HM220	2_Tcm	Tcm_04	SRS7638282	SRR12928695-SRR12928703, SRR12928706-SRR12928712

HM221	2_Tcm	Tcm_05	SRS7638286	SRR12928677-SRR12928681, SRR12928683-SRR12928692, SRR12928694
ReSeq_Ms01	3_Tms	Tms_01	SRS7637486	SRR12928998-SRR12929002, SRR12929004-SRR12929013, SRR12929015
ReSeq_Ms02	3_Tms	Tms_02	SRS7637485	SRR12928916, SRR12928918-SRR12928924, SRR12928988, SRR12928990, SRR12928991, SRR12928993-SRR12928997
ReSeq_Ms03	3_Tms	Tms_03	SRS7637493	SRR12928896, SRR12928898-SRR12928905, SRR12928907-SRR12928915
MS_Alpo3b	3_Tms	Tms_04	SRS7637467	SRR12929069-SRR12929077, SRR12929080-SRR12929089, SRR12929091, SRR12929092
MS_Alpo4b	3_Tms	Tms_05	SRS7637463	SRR12929016, SRR12929048-SRR12929055, SRR12929057-SRR12929066, SRR12929068
CC22B	3_Tce	Tce_01	SRS7638290	SRR12928577-SRR12928581, SRR12928583-SRR12928592, SRR12928595-SRR12928603
CC22C	3_Tce	Tce_02	SRS7638291	SRR12928555-SRR12928559, SRR12928561-SRR12928570, SRR12928572-SRR12928576
CC24B	3_Tce	Tce_03	SRS7638292	SRR12928533-SRR12928537, SRR12928539-SRR12928548, SRR12928550-SRR12928554

CC24C	3_Tce	Tce_04	SRS7637466	SRR12928521-SRR12928526, SRR12928528-SRR12928532, SRR12928819, SRR12929111, SRR12929113-SRR12929115
CC25B	3_Tce	Tce_05	SRS7637461	SRR12929093-SRR12929100, SRR12929102-SRR12929110
ReSeq_Di02	1_Tdi	Tdi_01	SRS7637469	SRR12928239, SRR12928250, SRR12928261, SRR12928272, SRR12928283, SRR12928294, SRR12928305, SRR12928961, SRR12928972, SRR12928983, SRR12929022, SRR12929034, SRR12929045
ReSeq_Di04	1_Tdi	Tdi_02	SRS7637489	SRR12928865-SRR12928870, SRR12928872, SRR12928878, SRR12928889, SRR12928928, SRR12928939, SRR12928950
ReSeq_Di06	1_Tdi	Tdi_03	SRS7637497	SRR12928806-SRR12928814, SRR12928861-SRR12928864
ReSeq_Di08	1_Tdi	Tdi_04	SRS7638280	SRR12928792, SRR12928794-SRR12928803, SRR12928805
ReSeq_Di10	1_Tdi	Tdi_05	SRS7638278	SRR12928779-SRR12928781, SRR12928783-SRR12928791
ReSeq_Ps14	1_Tps	Tps_01	SRS7637462	SRR12928527, SRR12928538, SRR12928549, SRR12928560, SRR12928571, SRR12928582, SRR12928593, SRR12928605, SRR12929014, SRR12929056, SRR12929067, SRR12929078, SRR12929090, SRR12929101, SRR12929112

ReSeq_Ps16	1_Tps	Tps_02	SRS7637490	SRR12928483, SRR12928494, SRR12928505, SRR12928516, SRR12928825, SRR12928836, SRR12928848, SRR12928859, SRR12928906, SRR12928917, SRR12928992, SRR12929003
ReSeq_Ps18	1_Tps	Tps_03	SRS7638308	SRR12928316, SRR12928328, SRR12928339, SRR12928350, SRR12928361, SRR12928372, SRR12928383, SRR12928394, SRR12928405, SRR12928416, SRR12928427, SRR12928439, SRR12928450, SRR12928461, SRR12928472
ReSeq_Ps08	1_Tps	Tps_04	SRS7637470	SRR12928317, SRR12928428, SRR12928594, SRR12928705, SRR12928782, SRR12928793, SRR12928804, SRR12928815-SRR12928817, SRR12928837, SRR12928871, SRR12929023, SRR12929079
ReSeq_Ps12	1_Tps	Tps_05	SRS7638283	SRR12928616, SRR12928627, SRR12928638, SRR12928649, SRR12928660, SRR12928671, SRR12928682, SRR12928693, SRR12928704, SRR12928716, SRR12928727, SRR12928738, SRR12928749, SRR12928760, SRR12928771
CC59_A	5_Tge	Tge_01	SRS7637468	SRR12928980-SRR12928982, SRR12928984-SRR12928987, SRR12928989, SRR12929017-SRR12929021, SRR12929024-SRR12929026

CC59_C	5_Tge	Tge_02	SRS7637484	SRR12928958-SRR12928960, SRR12928962-SRR12928971, SRR12928973-SRR12928979
CC65_B	5_Tge	Tge_03	SRS7637488	SRR12928937, SRR12928938, SRR12928940-SRR12928949, SRR12928951-SRR12928957
CC66_A	5_Tge	Tge_04	SRS7637487	SRR12928892-SRR12928895, SRR12928897, SRR12928925-SRR12928927, SRR12928929-SRR12928936
CC67_A	5_Tge	Tge_05	SRS7637494	SRR12928873-SRR12928877, SRR12928879-SRR12928888, SRR12928890, SRR12928891
Pa_AB	5_Tpa	Tpa_01	SRS7637465	SRR12929027-SRR12929033, SRR12929035-SRR12929043
PA_CD	5_Tpa	Tpa_02	SRS7638329	SRR12928245-SRR12928249, SRR12928251-SRR12928260, SRR12928262, SRR12928263
PA_E	5_Tpa	Tpa_03	SRS7637464	SRR12928231-SRR12928238, SRR12928240-SRR12928244, SRR12929044, SRR12929046, SRR12929047
H54	5_Tpa	Tpa_04	SRS7638331	SRR12928293, SRR12928295-SRR12928304, SRR12928306-SRR12928310
H56	5_Tpa	Tpa_05	SRS7638330	SRR12928264-SRR12928271, SRR12928273-SRR12928282, SRR12928284-SRR12928292

68

69

70 **SM Table 8. GO terms enriched in positively selected genes**

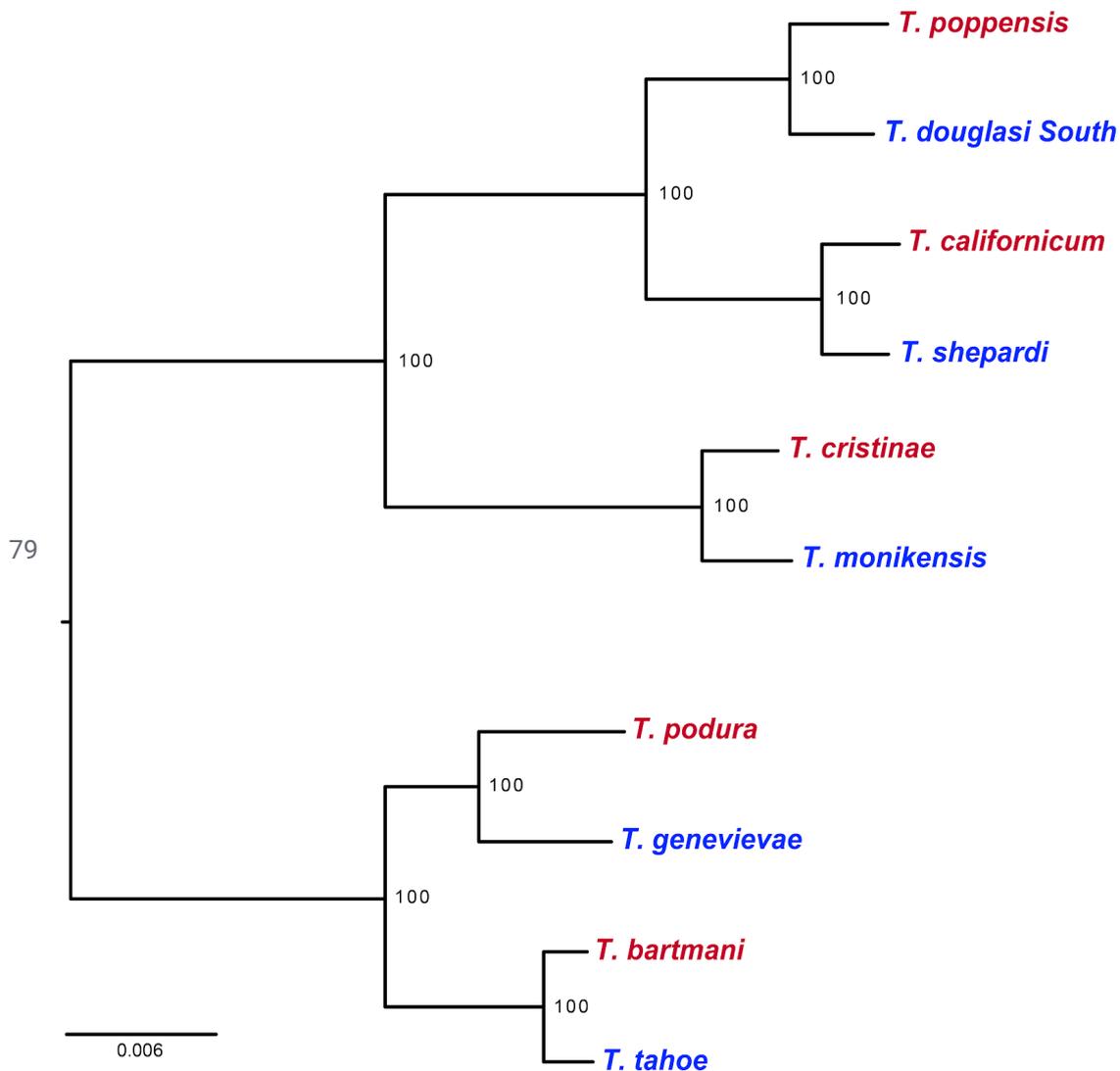
71 Few GO terms are enriched in positively selected genes in each species. This may
 72 be partly due to the difficulty in obtaining functional annotations in *Timema*, due to
 73 their evolutionary distance from a well characterised insect model system. Species
 74 are abbreviated as follows: Tbi = *T. bartmani*, Tce = *T. cristinae*, Tps = *T. poppensis*,
 75 Tcm = *T. californicum*, Tpa = *T. podura*, Tte = *T. tahoe*, Tms = *T. monikensis*, Tdi = *T.*
 76 *douglasi*, Tsi = *T. shepardi*, and Tge = *T. genevieveae*

77

GO ID	Term	Annotated	Significant	Expected	p	sp
GO:0007399	nervous system development	315	12	5.08	0.0028	Tte
GO:0006338	chromatin remodeling	20	3	0.32	0.0035	Tte
GO:0007476	imaginal disc-derived wing morphogenesis	45	4	0.73	0.0055	Tte
GO:0050775	positive regulation of dendrite morphogenesis	30	3	0.48	0.0118	Tte
GO:0030178	negative regulation of Wnt signaling pathway	31	3	0.5	0.0129	Tte
GO:0043039	tRNA aminoacylation	12	2	0.19	0.0152	Tte
GO:0031935	regulation of chromatin silencing	14	2	0.23	0.0205	Tte
GO:0008593	regulation of Notch signaling pathway	14	2	0.23	0.0205	Tte
GO:0045931	positive regulation of mitotic cell cycle	15	2	0.24	0.0234	Tte
GO:0006030	chitin metabolic process	17	2	0.27	0.0297	Tte
GO:0009058	biosynthetic process	454	10	7.33	0.0306	Tte
GO:0060966	regulation of gene silencing by RNA	16	2	0.26	0.0315	Tte
GO:0046331	lateral inhibition	44	3	0.71	0.0329	Tte
GO:0007155	cell adhesion	47	3	0.76	0.0389	Tte
GO:0007286	spermatid development	20	2	0.32	0.0403	Tte
GO:0006997	nucleus organization	20	2	0.32	0.0403	Tte
GO:0032990	cell part morphogenesis	127	7	2.05	0.0435	Tte
GO:0048814	regulation of dendrite morphogenesis	33	4	0.53	0.0451	Tte
GO:0002064	epithelial cell development	84	4	1.36	0.0459	Tte
GO:0032259	methylation	30	3	0.33	0.0011	Tbi
GO:0007631	feeding behavior	12	2	0.13	0.007	Tms
GO:0007450	dorsal/ventral pattern formation, imaginal disc	12	2	0.13	0.007	Tms
GO:0016485	protein processing	20	2	0.22	0.019	Tms
GO:0007601	visual perception	22	2	0.24	0.023	Tms
GO:0007088	regulation of mitotic nuclear division	22	2	0.24	0.023	Tms
GO:0031667	response to nutrient levels	29	3	0.31	0.03	Tms
GO:0007623	circadian rhythm	29	2	0.31	0.039	Tms
GO:0110116	regulation of compound eye photoreceptor cell differentiation	31	3	0.44	0.0019	Tsi
GO:0045732	positive regulation of protein catabolic process	10	2	0.14	0.0084	Tsi
GO:0043269	regulation of ion transport	12	2	0.17	0.0121	Tsi
GO:0031331	positive regulation of cellular catabolic process	15	2	0.21	0.0187	Tsi
GO:0045466	R7 cell differentiation	17	2	0.24	0.0237	Tsi
GO:0016197	endosomal transport	17	2	0.24	0.0237	Tsi
GO:0044773	mitotic DNA damage checkpoint	19	2	0.27	0.0279	Tsi

GO:0035088	establishment or maintenance of apical/basal cell polarity	21	2	0.3	0.0354	Tsi
GO:0006520	cellular amino acid metabolic process	32	2	0.46	0.0418	Tsi
GO:0006261	DNA-dependent DNA replication	13	2	0.18	0.013	Tcm
GO:0098869	cellular oxidant detoxification	14	2	0.2	0.016	Tcm
GO:1903008	organelle disassembly	14	2	0.2	0.016	Tcm
GO:0007052	mitotic spindle organization	14	2	0.2	0.016	Tcm
GO:0003007	heart morphogenesis	10	2	0.2	0.016	Tdi
GO:0071985	multivesicular body sorting pathway	10	2	0.2	0.016	Tdi
GO:0044262	cellular carbohydrate metabolic process	42	3	0.84	0.03	Tdi
GO:0033206	meiotic cytokinesis	15	2	0.3	0.039	Tdi
GO:0051046	regulation of secretion	21	2	0.42	0.039	Tdi
GO:0072657	protein localization to membrane	22	2	0.44	0.039	Tdi
GO:0098656	anion transmembrane transport	16	2	0.32	0.039	Tdi
GO:0030001	metal ion transport	17	2	0.34	0.044	Tdi
GO:0016197	endosomal transport	34	3	0.68	0.048	Tdi
GO:0030855	epithelial cell differentiation	81	2	0.57	0.027	Tge
GO:0007030	Golgi organization	50	2	0.35	0.046	Tge
GO:0010499	proteasomal ubiquitin-independent protein catabolic process	13	2	0.18	0.013	Tpa
GO:0001510	RNA methylation	14	2	0.19	0.015	Tpa

78



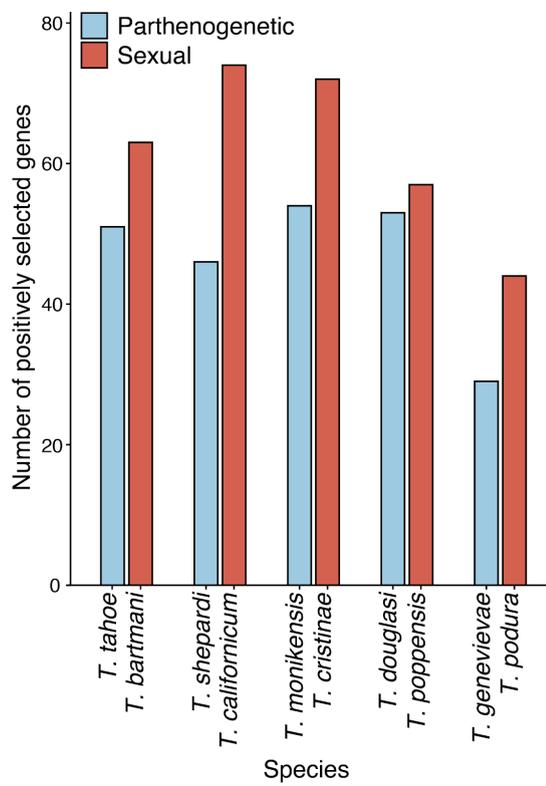
80 **SM Figure 1** | *Timema* phylogeny. Maximum likelihood tree based on 2377398
 81 orthologous coding DNA positions (from 3975 orthologs), rooted at the midpoint.
 82 Branch lengths represent the mean number of substitutions per site. Node labels
 83 indicate branch support (%) from 1000 bootstrap replicates. Othologs were aligned
 84 using MCoffee (v11.00.8cbe486) (79) which was run with the following aligners:
 85 mafft_msa, muscle_msa, clustalo_msa (80), and t_coffee_msa (81). Alignments
 86 were concatenated together, and filtered with GBlocks (v. 0.91b, type = codons,
 87 minimum block length = 12) to remove large alignment gaps and blocks of Ns (86).
 88 The tree was then generated with RAxML (66), with a GTR+gamma model with 40
 89 rate categories for each codon position.

90

91

92

93

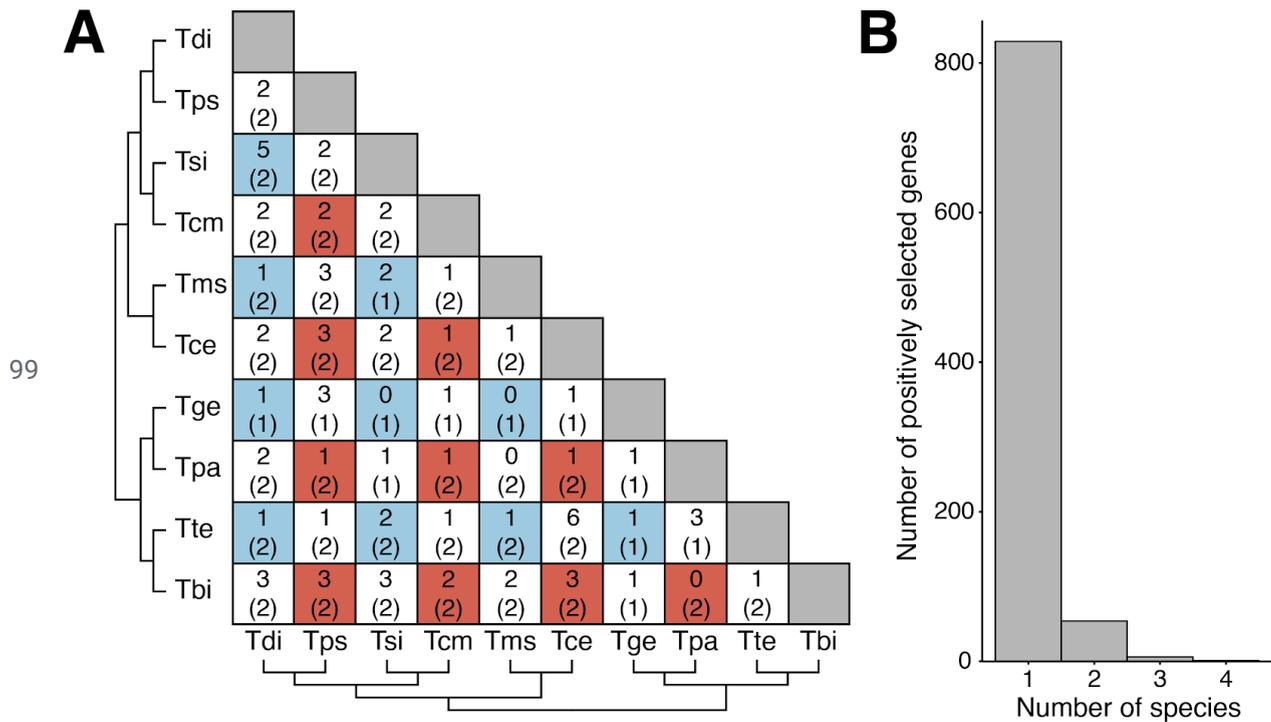


94 **SM Figure 2** | Number of genes showing evidence for branch-site positive selection
95 on terminal branches with a q-value threshold of 0.01.

96

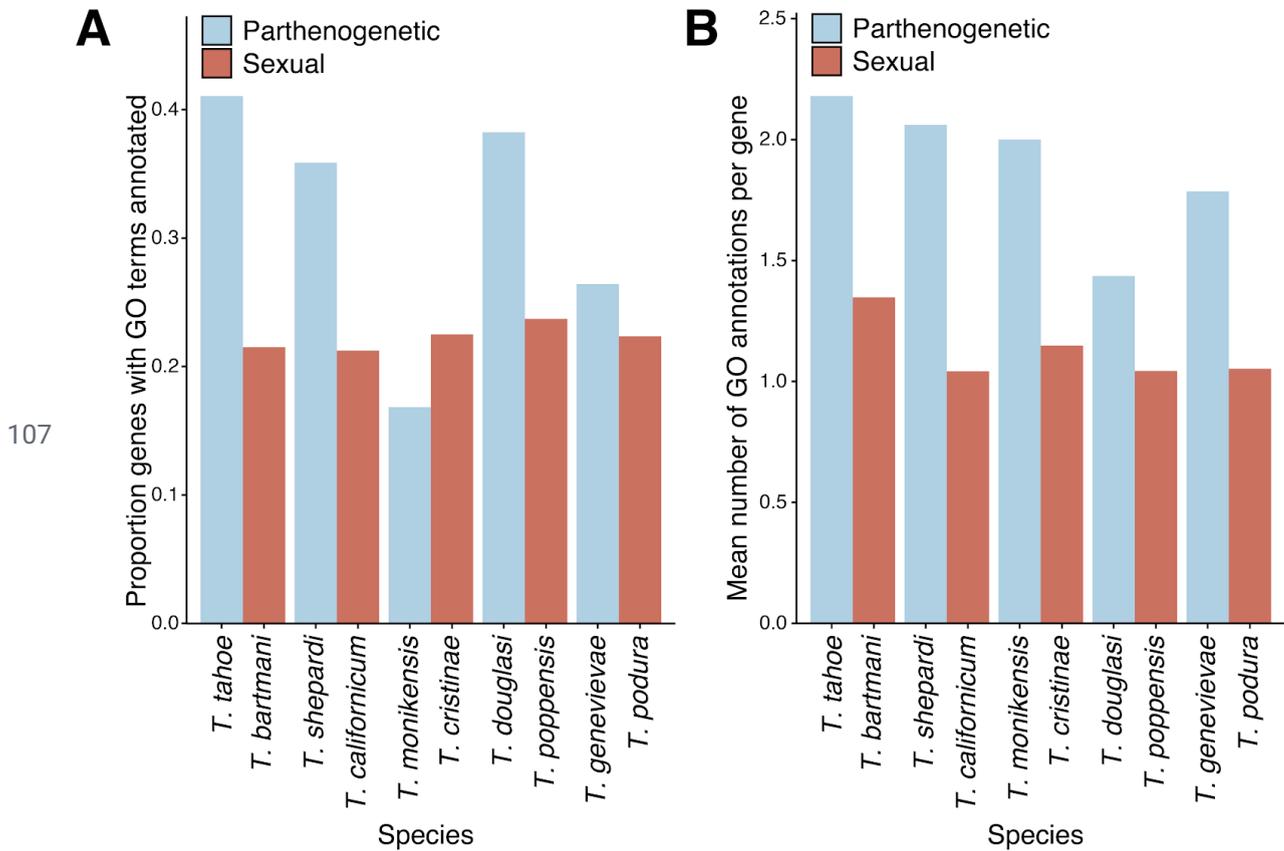
97

98



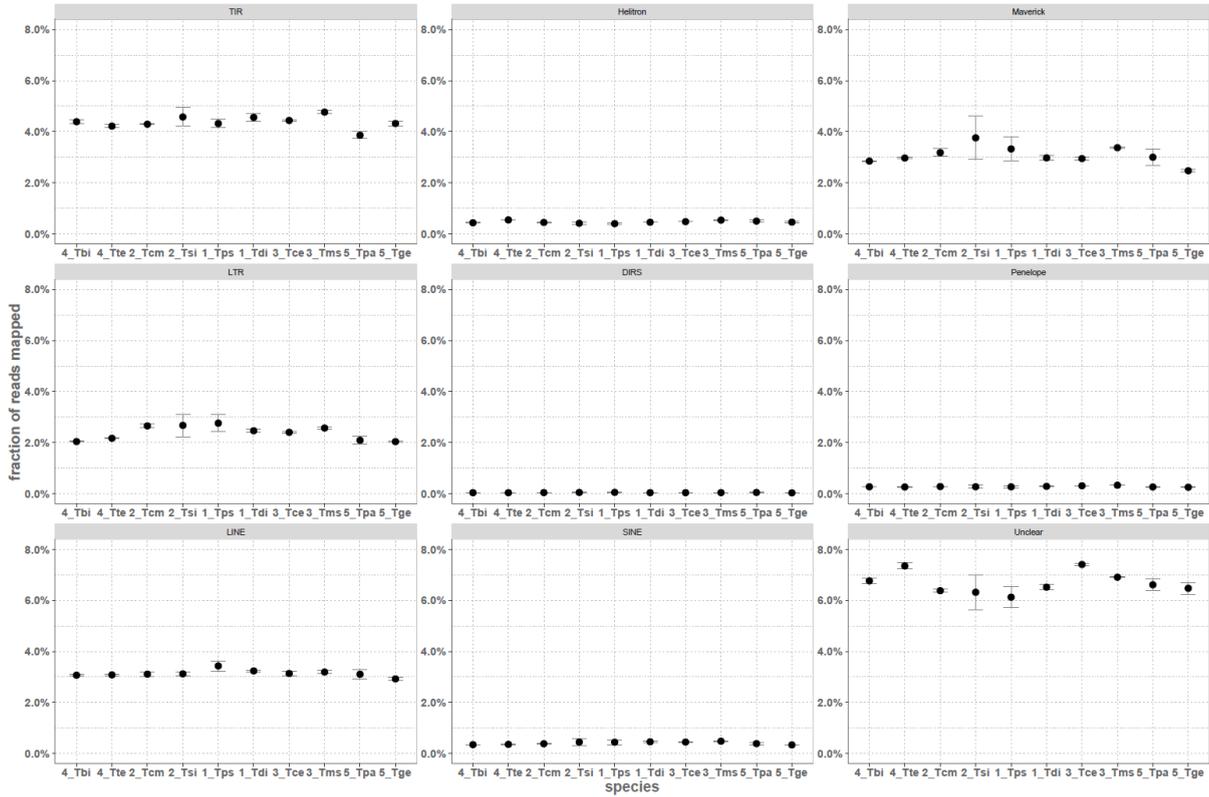
100 **SM Figure 3** | Positively selected genes are shared between few species. **A.** Matrix
 101 showing pairwise overlap of positively selected genes with the number of genes
 102 expected by chance given in parentheses. Red cells indicate the overlap between
 103 two sexual species, blue between two parthenogenetic species, and white between
 104 one sexual and one parthenogenetic species. **B.** Number of species positively
 105 selected genes are found in.

106



108 **SM Figure 4** | Positively selected genes in sexual species have fewer annotations
 109 than in parthenogenetic species. **A.** Proportion of positively selected genes with at
 110 least 1 GO term (biological processes) annotated. **B.** Mean number of GO terms
 111 annotated in positively selected genes with at least 1 GO term (biological processes)
 112 annotated

113



114 **SM Figure 5 | Genomic transposable element loads, separated by TE orders.**

115 Error bars represent standard deviation across the six (three for Tsi) sequenced
116 genomes in each species. Species are abbreviated as follows: Tbi = *T. bartmani*, Tce
117 = *T. cristinae*, Tps = *T. poppensis*, Tcm = *T. californicum*, Tpa = *T. podura*, Tte = *T.*
118 *tahoe*, Tms = *T. monikensis*, Tdi = *T. douglasi*, Tsi = *T. shepardi*, and Tge = *T.*
119 *genevievae*

120

121 **SM text 1: Assembly and annotation pipelines**

122 Paired-end raw reads were trimmed according to sequencing quality and matched to
123 known Illumina sequencing adapters using Trimmomatic (v.0.36) (87). Leading and
124 trailing bases below quality 9 were removed. Reads were scanned using a 4-base
125 sliding window, trimmed when the average quality dropped below 15, and discarded
126 if read length dropped below 96bp (Parameters: PE ILLUMINACLIP:
127 all-adapters.fa:3:25:6 LEADING:9 TRAILING:9 SLIDINGWINDOW:4:15 MINLEN:96).
128 The raw mate-pair reads were de-linked and reverse complemented using NxTrim (v.
129 0.4.1) (88) with the parameter "--preserve-mp". Unlinked pairs without identified
130 adapter sequence, called unknown pairs, were also considered as valid mate pairs
131 as they had a similar distribution of insert sizes as mate pairs with identified linker
132 sequence.

133

134 Filtered paired-end reads were *de novo* assembled using ABySS (v. 1.9.0) (52, 89)
135 with default parameters and k-mer sizes predicted to be optimal using kmergenie
136 (90). The k-mer sizes were 83, 87, 83, 87, 83, 89, 81, 81, 65 and 87 for *Timema*
137 *poppensis*, *T. douglasi*, *T. californicum*, *T. shepardii*, *T. cristinae*, *T. monikensis*, *T.*
138 *barmani*, *T. tahoe*, *T. podura* and *T. genevieve* respectively. Assembled contigs
139 longer than 250 bases were scaffolded using BESST (v. 2.2.5) (53) with default
140 parameters and gap-filled with GapCloser (v. 1.12-r6), a module of the SOAP
141 package (91).

142

143 Genome assemblies were decontaminated using BlobTools (v. 0.9.19.5) (54). Hit
144 files were generated after a BlastN (v. 2.6.0) (74) against the NCBI nt database (v
145 2016-06) (92), searching for hits with sequence identity above 85% and an e-value
146 below 1e-25 (Parameters: -task megablast -culling_limit 5 -evalue 1e-25
147 -perc_identity 85). Scaffolds without hits to metazoans were removed from the
148 assemblies. The genome assembly completeness was assessed with BUSCO (v.
149 3.0.2) (22) against the insecta_odb9 lineage and the -long option. For genome
150 annotation, we took a total of 231 publically available RNA-seq libraries for *Timema*
151 from different tissues, sexes and developmental stages as expression evidence (min

152 per species = 12, see SM Table 6) (37, 55, 56). Before mapping reads to the
153 genomes, adapter sequences were trimmed from raw reads with CutAdapt (v. 1.15)
154 (93). Reads were then quality trimmed using Trimmomatic (v. 0.36) (87), clipping
155 leading or trailing bases with a phred score of <10 from the read, before using a
156 sliding window from the 5' end to clip the read if 4 consecutive bases had an average
157 phred score of <20. Any reads with a sequence length of <80 after trimming were
158 discarded. All trimmed RNA-seq reads were then mapped against the genomes as
159 single end reads using STAR (v. 2.5.3a) (94) under the “2-pass mapping” mode and
160 default parameters. The STAR outputs were then used to produce transcriptome
161 assemblies using Trinity (v. 2.4.0) (57) “genome guided” mode (Parameters:
162 --genome_guided_max_intron 100000 --SS_lib_type R). Finally, the transcriptome
163 assemblies were filtered following Trinity developers recommendations
164 (<https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-FAQ>): Briefly, filtered
165 RNA-seq reads were mapped back against the transcriptomes using Kallisto (v.
166 0.43.0) (95) with options --bias and --single, then transcripts with at least 1 TPM in
167 any sample were retained.

168

169 Genome scaffolds >1000 bp were annotated, protein coding genes were predicted
170 using MAKER (v. 2.31.8) (58) in a 2-step iterative way as described in Campbell *et*
171 *al.* (96) with minor modifications following author recommendations. For the first
172 iteration, genes were predicted using Augustus (v. 3.2.3) (97) trained with the
173 BUSCO results. A combination of UniProtKB/Swiss-Prot (release 2018_01) (98) and
174 the BUSCO insecta_odb9 proteome were used as protein evidence. The Trinity
175 assembled RNA-seq reads (described above) were used as transcript evidence. The
176 resulting gene models were then used to retrain Augustus as well as SNAP (v.
177 2013.11.29) (99) and a second iteration was performed. Predicted protein coding
178 genes were then functionally annotated using Blast2GO v5.5.1 (100, 101) with
179 default parameters against both the NCBI non-redundant arthropods protein
180 database, and the *Drosophila melanogaster* (drosoph) database, to produce two sets
181 of functional annotations, one derived from all arthropods and one specifically from
182 *Drosophila melanogaster*.

183

184 **SM text 2: Horizontal Gene Transfers (HGTs) are not facilitated by**
185 **parthenogenesis**

186 Genomic analyses of bdelloid rotifers, a group that likely persisted and diversified in
187 the absence of canonical sex for over 40 million years (102), revealed that bdelloids
188 carry an unusually large amount (6.2% - 9.1%) of horizontally acquired genes
189 compared to sexual lophotrochozoans (0.08% - 0.7%) (17, 103–105). Unusually high
190 proportions of HGT-derived genes were also identified in parthenogenetic root-knot
191 nematodes (106, 107) and springtails (108). These findings led to the suggestion
192 that parthenogenesis might favor the retention of horizontally acquired genes, and
193 may perhaps confer adaptive benefits that could compensate for the absence of
194 recombination and outcrossing (106), although such patterns are not shared by most
195 other parthenogenetic animal genomes (17). Analyzing HGT events in *Timema*
196 provided no evidence for parthenogenesis facilitating the retention of HGTs. We
197 identified 55 putative HGT events in the 10 *Timema* species, with up to 50
198 sequences each, for a total of 704 HGT-derived sequences (351 in the five sexual
199 species vs. 353 in the five parthenogenetic species). The genome of each *Timema*
200 species included approximately 70 HGT-derived sequences, comparable to values
201 from metazoa in general (109). All putative HGT families were shared by at least six
202 *Timema* species, and only one putative HGT event occurred in a specific clade (HGT
203 family shared between two sexual and two parthenogenetic species of the Northern
204 clade) while all other HGT events were shared between at least two clades of
205 *Timema*.

206 Of note, out of the 55 HGT families, 34 featured significant similarities with
207 sequences from two plant pathogens (*Phytophthora infestans* and *Pythium ultimum*),
208 and displayed a high-glycine content, due to many 'GGG' repeats. This repeated
209 motif is very similar to the lorocrin-like protein described in *Phytophthora infestans* by
210 Guo et al. (110), which is suggested to be involved in plant infection.

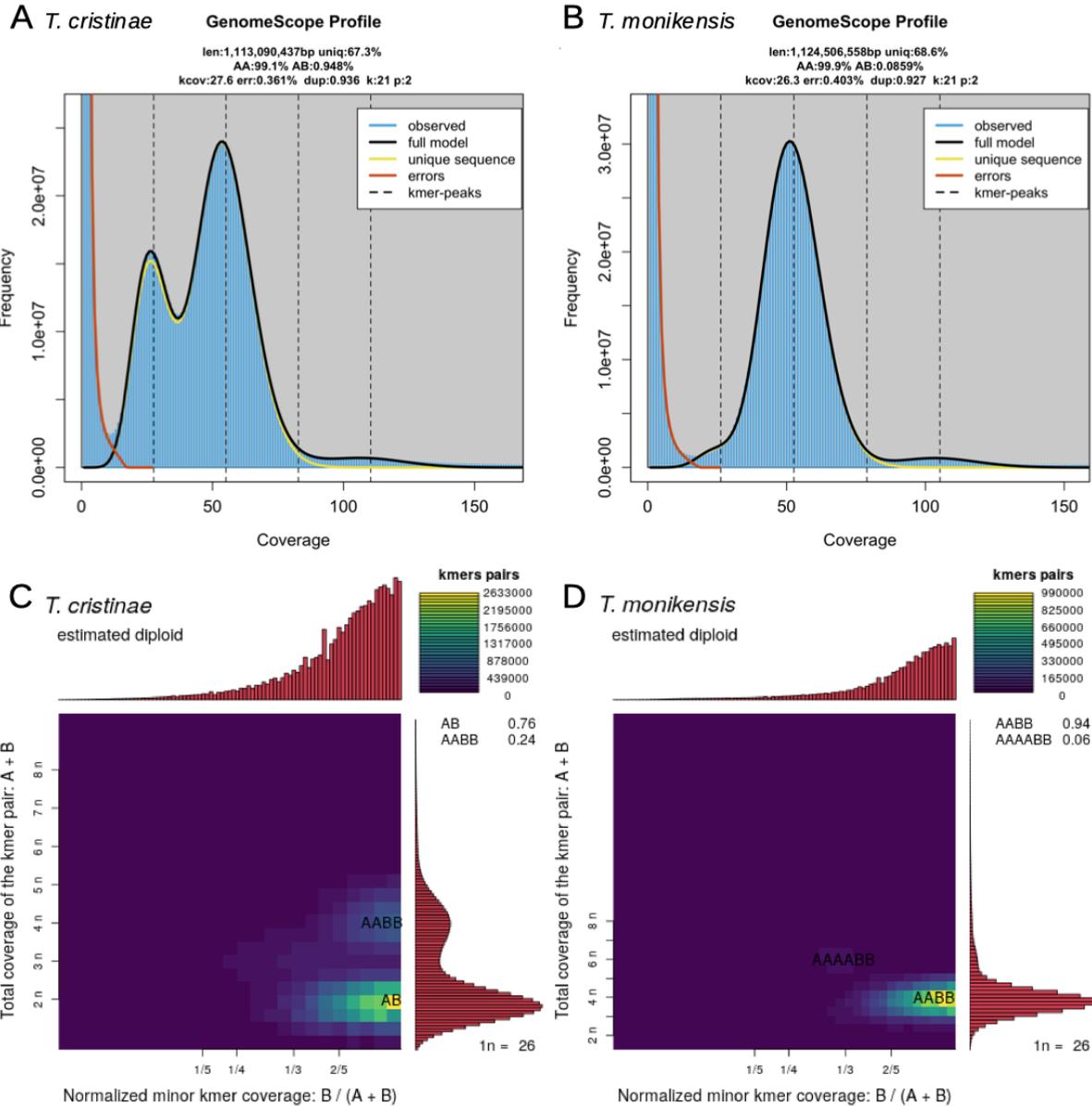
211 Out of the 21 remaining families, only 4 showed a phylogenetic pattern consistent
212 with an old HGT event, sometimes shared with *Zootermopsis nevadensis* (the

213 closest species in our reference database). However, the terminal branches leading
214 to the putatively-transferred sequences were too long for a reliable identification of
215 the donor species. The phylogenetic evaluation of the other families was not
216 conclusive, as commonly observed in HGT detection studies (e.g. see (111)).

217 **SM text 3: Analysis of heterozygosity for SNPs and SVs**

218 We present two estimates of heterozygosity, one based on a reference-free
219 technique (kmer spectra analysis using Genomescope (v. 2) (23), the other using
220 sequencing reads mapped to reference genomes to call SNPs with GATK ((67), see
221 Methods).

222 The kmer spectra of all sexual species displayed distinct haploid coverage peaks
223 representing heterozygous kmers (SM Figure 6A), contrasting with the kmer spectra
224 of parthenogenetic species, where no distinct peaks were visible (SM Figure 6B). To
225 confirm that no heterozygous kmers were present at the expected haploid coverage
226 in parthenogens, we used Smudgeplot (23), a technique to extract closely related
227 kmer pairs representing heterozygous and paralogous kmer pairs. While in the
228 sexual species, kmers from the $1n$ peak paired together in heterozygous kmer pairs
229 (AB smudge on SM Figure 6C), no diploid kmer pairs were detected in the
230 parthenogenetic species (SM Figure 6D). We conclude that heterozygosity estimates
231 for the parthenogenetic species cannot be based on k-mer spectra analyses
232 because the heterozygosity levels are too low to reliably fit the distribution estimating
233 haploid kmers in the kmer spectra. Unreliable heterozygosity estimates based on
234 k-mer spectra analyses for species with very low heterozygosity was already
235 reported in Jaron et al. (17), suggesting that with the current quality of sequencing
236 data, kmer methods do not have resolution for very small heterozygosity levels.



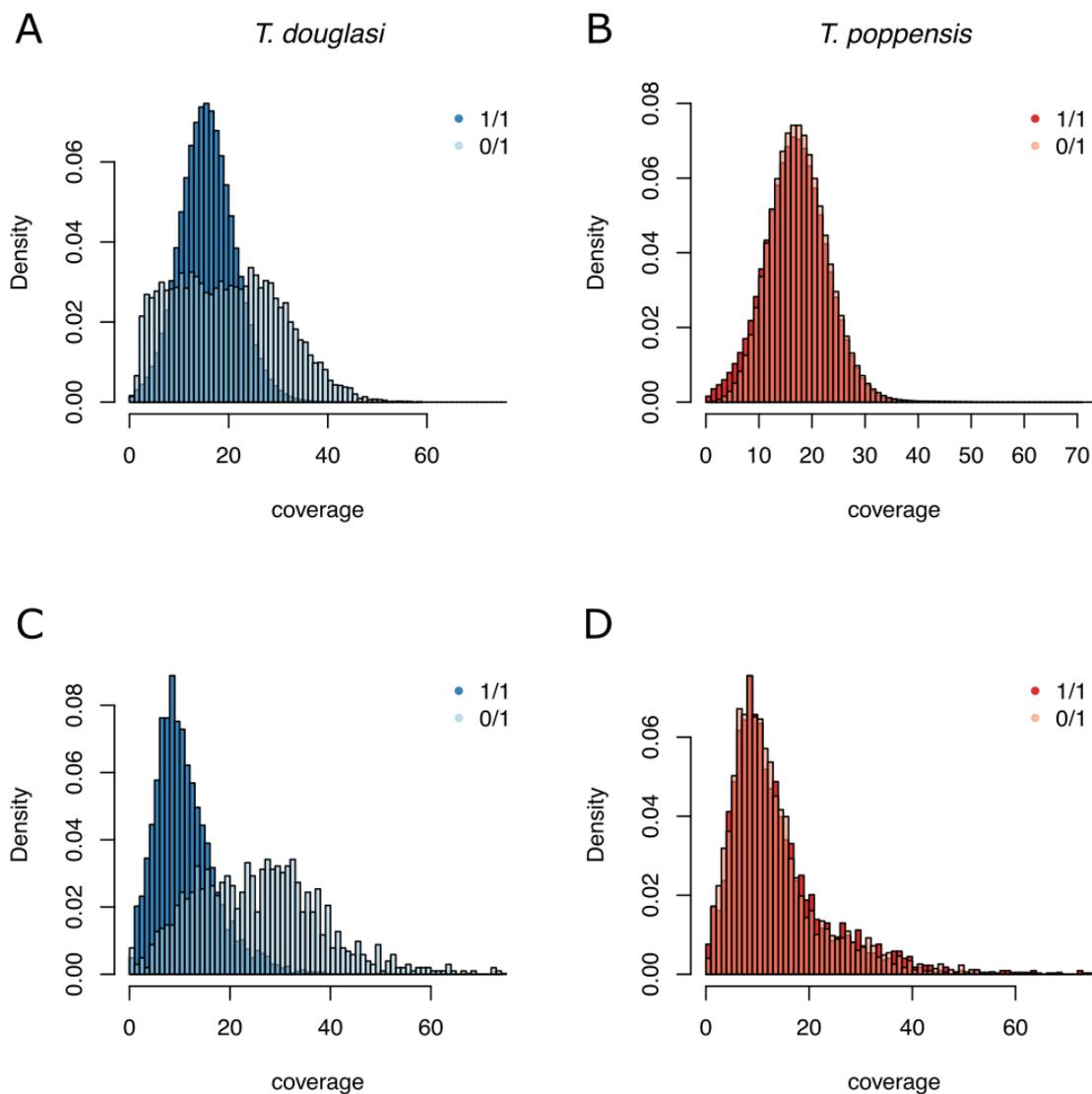
237

238 **SM Figure 6:** Genome profiling examples for a sexual (*T. cristinae*, panels A and C)
 239 and a parthenogenetic (*T. monikensis*, panels B and D) *Timema* species.

240

241 Because we could not estimate heterozygosity of parthenogens using kmer-spectra
 242 analyses, we estimated nucleotide heterozygosity using SNP calling. It is important
 243 to note, however, that this method generates an underestimation of heterozygosity
 244 given our fragmented reference genomes (SM Table 3) and relatively modest
 245 coverage (~14 - 21x) of re-sequenced samples. Therefore, our SNP heterozygosity
 246 estimates in *Timema* are useful for comparing sexual and parthenogenetic species,
 247 but are not accurate estimates of heterozygosity in *Timema* (which range from 0.36

248 to 2.16% for sexual species, i.e., 2-6 times higher than the SNP-based estimates,
249 Figure 2). In agreement with genome profiling, we find very low, nearly negligible
250 levels of heterozygosity in parthenogenetic species (Figure 2). Furthermore, a large
251 portion of the heterozygous SNP calls in parthenogens showed an unexpectedly
252 high coverage (SM Figure 7). This excess coverage of heterozygous positions in
253 parthenogens suggests that heterozygous sites in parthenogens largely stem from
254 merged paralogs, further supporting that a very large proportion (or maybe even all)
255 of the called heterozygous variants in parthenogens are just artifacts of the SNP
256 calling pipeline using whole genome data.
257



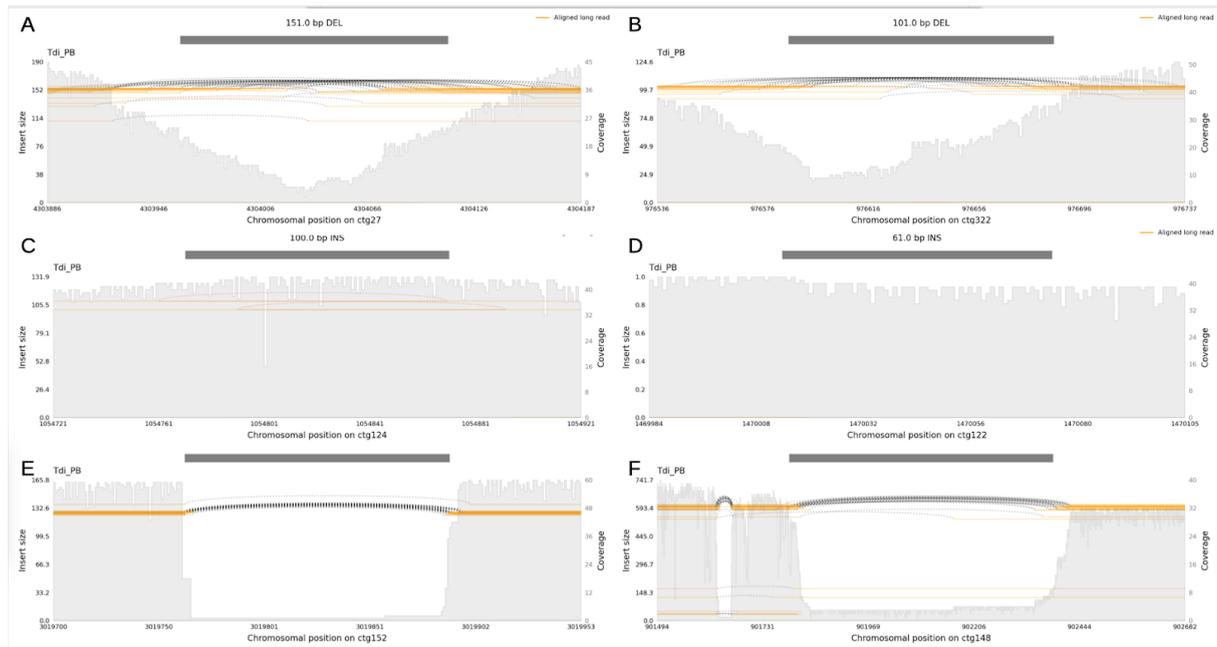
258

259 **SM Figure 7:** Densities of coverages supporting **A** SNPs found in the homozygous

260 state (1/1), or heterozygous (0/1) in *T. douglasi*. **B** In sexual *T. poppensis* **C** Densities
261 of split read coverage support of SVs in homozygous or heterozygous states in *T.*
262 *douglasi* **D.** *T. poppensis*. Both heterozygous SNPs and heterozygous SVs show
263 unexpected coverage distributions in parthenogenetic *T. douglasi* (blue), while
264 coverages supporting SNPs in sexual *T. poppensis* (red) are independent of the
265 genotype. There is a small difference in homozygous and heterozygous SV
266 coverages in sexuals, suggesting that at least some fraction of those heterozygous
267 SVs are also false positives. However, overlap of the two distributions is much
268 greater than in the case of parthenogenetic *T. douglasi* (panel C).

269

270 We further investigated if there were any heterozygous structural variations in
271 parthenogenetic *Timema*, as those could be potentially hidden to SNP analysis.
272 Consistent with the previous two analyses, the SV heterozygosity levels were
273 substantially lower in the parthenogens than in their sexual sister species (Figure 2).
274 However, we also detected a non-negligible amount of heterozygous structural
275 variants. We therefore manually curated all heterozygous structural variants found in
276 *T. monikensis* using samplot (v1.0.1) (112), but did not find a single variant clearly
277 supported by reads (results not shown). Since structural variant calling from short
278 read data has a high rate of false positives regardless of the method used (113), we
279 decided to verify variants using a PacBio long-read dataset (~32x coverage) for one
280 of the parthenogenetic species (*T. douglasi*; PRJNA673001). We assembled the
281 long-read data of *T. douglasi* using Redbean (formerly wtdbg; v2.5) assembler (114)
282 with parameters recommended for moderately sized genomes: -L 1000 -x preset3 -g
283 1300m. This genome assembly was used for SV calling using ngmlr (v0.2.7) and the
284 Sniffles (v1.0.11) pipeline (115) with default parameters for SV calling using long
285 read data. In total, we found only 6 heterozygous SVs: 4 deletions and 2 insertions.
286 We visualized the SVs alongside their read support using samplot and found that
287 none were well supported (SM Figure 8) suggesting that the heterozygous SVs
288 called using short read data represent noise in the absence of a signal from real
289 heterozygous SVs.



290

291 **SM Figure 8.** All 6 heterozygous SVs called in the *T. douglasi* long read dataset.
 292 SVs on panels A and B are located in repetitive regions which is causing the uneven
 293 distribution of coverages and variable lengths of gaps. Variants on panels C - F are
 294 not supported by approximately half of the reads. Variant C is probably due to rare
 295 chimeric reads, and variant D does not seem to have any support at all. Conversely,
 296 SVs on panels E and F have very low support for the reference sequence. See
 297 examples provided in the manual of samplot for comparison to a well supported
 298 heterozygous SV.

299

300 In conclusion, we used four complementary approaches based on three different
 301 data sources: kmer spectra analysis on raw sequencing reads of the reference
 302 individuals, SNP and SV heterozygosity estimates using variant calling based on
 303 resequencing data, and finally a long read dataset of *T. douglasi*, which was
 304 independently assembled and is therefore free of any potential biases introduced in
 305 a short read assembly. Our analyses comprehensively show the absence of
 306 heterozygous loci in the parthenogenetic *Timema* genome assemblies. Residual
 307 heterozygosity could be potentially found in repetitive regions, such as centromeres
 308 and telomeres (see also SM text 5), as all our effort to detect heterozygosity focused
 309 on alleles with 1n coverage (half of the genome coverage). However, detecting
 310 heterozygosity in such regions requires chromosome-scale assemblies based on

311 long-read sequencing technologies, which are currently not available for
312 parthenogenetic *Timema*.

313

314 **SM text 4: Locating microsatellite markers in the genome assemblies**

315

316 Previous research, based on microsatellite markers, suggested that oogenesis in
317 parthenogenetic *Timema* was functionally mitotic, as there was no loss of
318 heterozygosity between females and their offspring (18). Yet our genome data reveal
319 complete or almost complete homozygosity in the genome assemblies of
320 parthenogens (see main text). The most likely reconciliation of these contrasting
321 results is that heterozygosity is maintained in only a small portion of the genome, for
322 example the centromeres or telomeres, or between paralogs.

323

324 To investigate these possibilities, we searched for the primer pairs used to amplify
325 the nine microsatellites in the genome assembly v1.3 of the sexual species *T.*
326 *cristinae* from Nosil et al (35). This assembly is currently the most complete and least
327 fragmented *Timema* assembly available, and the microsatellites used by (18) were
328 originally developed for *T. cristinae*. We used Blast to find primer pairs <500 bp
329 apart, on opposite strands, and retained significant hits with at least 80% of the
330 primer sequences covered. We then verified whether the retained hits comprised the
331 expected microsatellite repeat motif.

332

333 Using this approach, we were able to locate six of the nine microsatellites in the v1.3
334 assembly (SM Table 9). Two of the six microsatellites had multiple hits in the genome
335 (SM Table 9). In combination, these results support the idea that microsatellite
336 heterozygosity detected in *Timema* parthenogens may be a combination of
337 heterozygosity in centromere or telomere regions (microsatellites not detected in the
338 assembly) and heterozygosity between paralogs (microsatellites with multiple copies
339 in the *T. cristinae* assembly).

340

341

342 **SM Table 9** | Microsatellites located in the v1.3 genome of *T. cristinae*. Msat name:
 343 Microsatellite name from (18). Indicated are the scaffolds where a given
 344 microsatellite was found (Scaffold), the location of the microsatellite midpoint on the
 345 scaffold (Position), the linkage group (LG), the size of the microsatellite in the v1.3
 346 assembly and the expected size range given microsatellite genotypes in *T. cristinae*
 347 (Length (expected)), and whether the expected microsatellite repeat motif was
 348 present.

Msat name	Scaffold	LG	Position	Length (expected) [bp]	Motif
tim-3 (a)	CM009483.1	LG8	32962807	159 (82-109)	Yes
tim-3 (b)	CM009476.1	LG13	19984463	169 (82-109)	Yes
tim-4	CM009477.1	LG2	46600506	119 (83-125)	Yes
tim-5	CM009481.1	LG6	7421031	198 (124-238)	Yes
tim-6	CM009474.1	LG11	12218418	264 (253-283)	Yes
tim-7	CM009482.1	LG7	19035853	156 (120-195)	Yes
tim-8 (a)	CM009484.1	LG9	23487739	138 (133-148)	Yes
tim-8 (b)	CM009479.1	LG4	85533694	190 (133-148)	No
tim-8 (c)	CM009478.1	LG3	981498	498 (133-148)	No

349

350 **SM text 5: Polymorphism in parthenogenetic and sexual *Timema* populations**

351 To compare the distribution of polymorphism along different genomes, we mapped
 352 population-level variation for SNPs and SVs inferred from 2 to 5 re-sequenced
 353 individuals per population to our species-specific reference genomes (see main text).
 354 We then anchored our reference genome scaffolds to the 12 autosomal linkage
 355 groups of a previously published assembly of the sexual species *T. cristinae* (v1.3
 356 from Nosil et al. (35)) using MUMmer (v. 4.0.0beta2) (70) (see Methods for details).
 357 Note that we excluded LG13, classified as the X chromosome in Nosil et al (35),
 358 because our detailed analyses of X-chromosomes in *Timema* revealed that LG13 did

359 not correspond to the X (Parker et al in prep). We also removed X-linked scaffolds
360 assigned to autosomal LGs in the v1.3 *T. cristinae* assembly and used this “cleaned”
361 set of linkage groups (referred to as v1.4) in all our analyses with positional
362 information. Depending on the species, we were able to anchor between 59 and 558
363 Mbp of our genomes to the *T. cristinae* LGs.

364

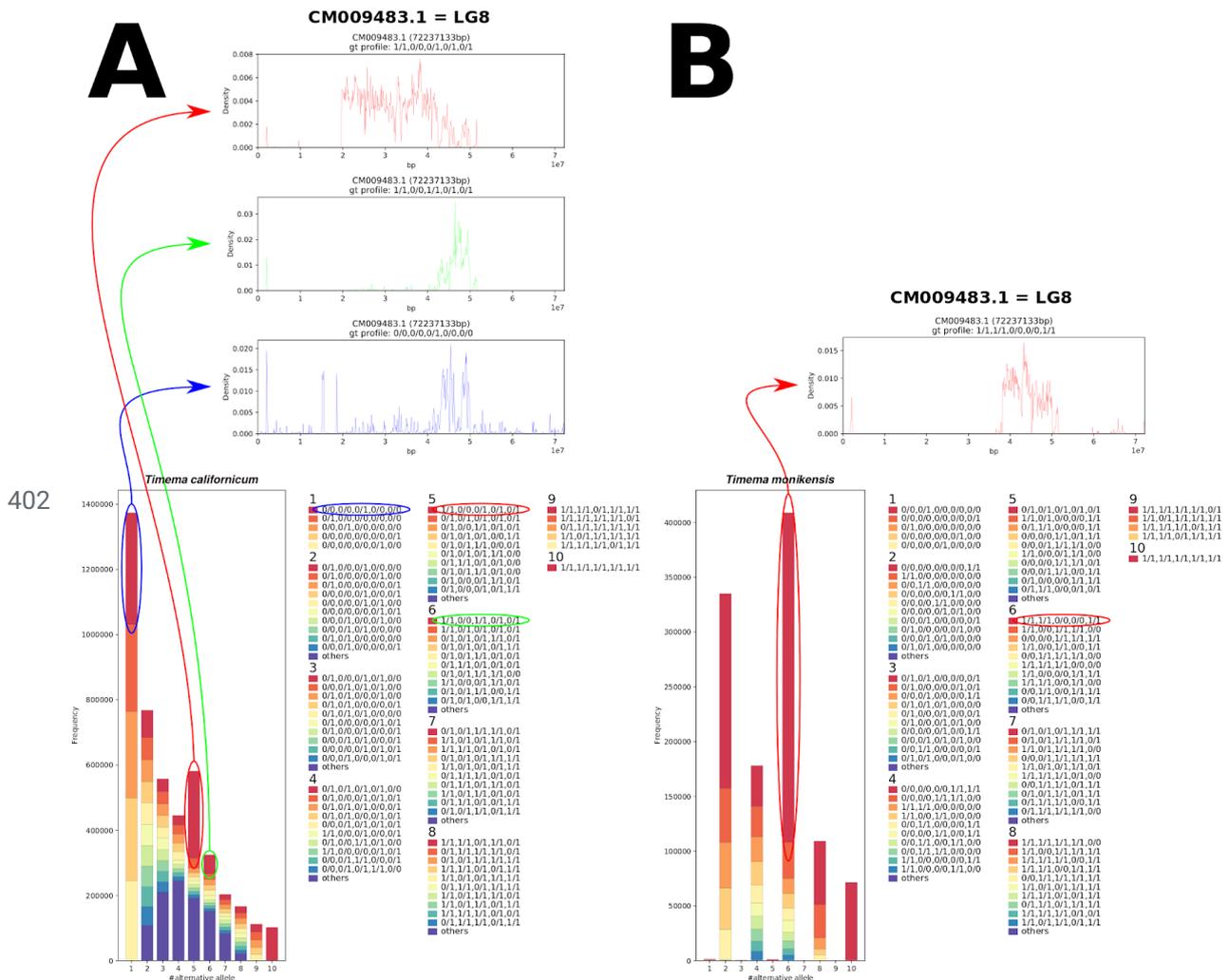
365 We also examined how genetic variation was distributed between individuals by
366 producing phylogenetic trees for the re-sequenced and reference individuals of each
367 sexual-parthenogenetic sister species pair. Sequences for re-sequenced individuals
368 were obtained by mapping reads of each re-sequenced individual to the reference
369 genomes with BWA-MEM (v0.7.15) (76). Multi-mapping and poor quality alignments
370 were filtered (removing reads with XA:Z or SA:Z tags or a mapq < 30). We removed
371 PCR duplicates with Picard (v. 2.9.0) (<http://broadinstitute.github.io/picard/>) and
372 performed indel realignment with GATK (v. 3.7) (67). Genomic sequences for each
373 re-sequenced individual were then generated using AngsD (v. 0.921) (-doCounts 1
374 -doFasta 2) (116) with a minimum depth of 5 and a maximum depth of twice the
375 median genome coverage. Coding sequences of 1-to-1 orthologs (2198) were
376 extracted using gffread from the Cufflinks (v. 2.2.1) package (117). These sequences
377 were codon-aligned using PRANK (v.100802) (118) concatenated together, and
378 filtered with GBlocks (v. 0.91b, type = codons, minimum block length = 12) to remove
379 large alignment gaps and blocks of Ns (86). Trees were generated with RAxML (66),
380 with a GTR+gamma model with 40 rate categories for each codon position (i.e. each
381 codon position (1st ,2nd, 3rd) was partitioned to allow a distinct model to be fitted to
382 it) to produce an ML tree with 1000 bootstraps.

383 **SM text 6: Polymorphism and color morphs on LG8 in the species *T.***

384 ***californicum* and *T. monikensis***

385 We found very high population polymorphism for both SNPs and SVs on LG8 in *T.*
386 *californicum* and *T. monikensis* (Figure 3B). A site frequency spectrum (SFS) in *T.*
387 *californicum* based on the SNPs called in the 5 resequenced individuals revealed
388 that the polymorphism in this species was likely generated by the presence of two
389 distinct haplotypes. The genotype structure for the 5 individuals was similar for a

390 large portion of SNPs on LG8 (1/1 0/0 0/1 0/1 0/1; SM Figure 9A), with three
 391 individuals heterozygous, and the two remaining ones homozygous for alternative
 392 alleles (SM Figure 9A). The LG8 SNP genotypes further matched the grey versus
 393 green color morphs, with the grey morph known to have recessive inheritance (119):
 394 the individual used to build the reference genome (SNP genotypes 0/0) was grey, as
 395 was resequenced individual 2 (Tcm_02) with 0/0 genotypes. The four remaining
 396 resequenced individuals (with 1/1 or 0/1 genotypes) were green. The size of the
 397 putative haplotypes associated with green or grey morphs is considerable, spanning
 398 approximately 24 Mbp on LG8. The genotypes at LG8 in *T. monikensis* were also
 399 correlated with color morphs. Brown or beige melanistic morphs featured 0/0
 400 genotypes in the high polymorphism region of LG8, while green individuals had 1/1
 401 genotypes (SM Figure 9B).

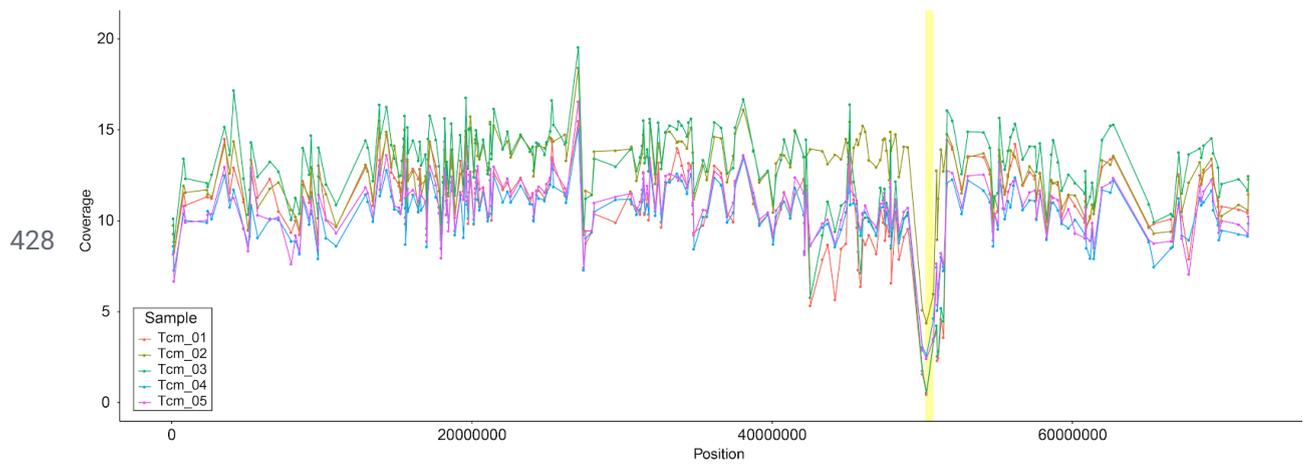


403 **SM Figure 9.** Site Frequency Spectrum for the five re-sequenced individuals of *T.*
404 *californicum* (A) and *Timema monikensis* (B), generated with Pop-Con
405 (<https://github.com/YoannAnselmetti/Pop-Con>), indicating the genotype distributions
406 for each count of alternative alleles. For *T. californicum*, the peak at count 5 is
407 generated by the overrepresented genotype structure 1/1 0/0 0/1 0/1 0/1, and almost
408 all SNPs with this structure (97.34%) map to LG8, suggesting the presence of two
409 divergent haplotypes on LG8. For *T. monikensis*, we observed a similar
410 overrepresentation at allele count 6, for the genotype structure 1/1 1/1 0/0 0/0 1/1,
411 and most of the SNPs with this structure (78.43%) map to LG8 .

412

413 We also investigated the presence of an approximately 0.5 Mb deletion on LG8,
414 suggested by Villoutreix et al. (120) to determine the green morphs in *T. californicum*,
415 in our ~24 Mbp long haplotypes. Because our reference genome was based on a
416 grey individual (which would be homozygous for the deletion-free haplotype), we
417 expected to observe normal coverage for this region in the re-sequenced grey
418 individual, zero coverage in the green re-sequenced individual homozygous for the
419 alternative haplotype, and half the coverage in the three heterozygous individuals.
420 We observed a coverage reduction in all individuals at the focal region, which could
421 be due to an enrichment in repetitive sequences (non-uniquely mapping reads are
422 not included for coverage estimations). Nevertheless, the grey individual featured
423 somewhat higher coverage, consistent with the deletion suggested by Villoutreix et
424 al. (120) (SM Figure 10). Further studies are required to characterize the contribution
425 of the two divergent, ~24 Mbp long haplotypes, and the putative 0.5 Mb deletion in
426 one of the haplotypes, to color polymorphism in *T. californicum*.

427



429 **SM Figure 10.** Coverage along LG8 for 5 resequenced *T. californicum*. Coverage
 430 was estimated by mapping reads to the *T. californicum* genome scaffolds, and
 431 scaffolds were anchored on *T. cristinae* linkage groups (see methods and SM text 5).
 432 The region with the expected deletion is highlighted in yellow. If there was a deletion
 433 on LG8 determining the green morph, grey individual Tcm_02 should feature higher
 434 coverage than the other individuals which are green.

435 References

- 436 1. M. Neiman, C. M. Lively, S. Meirmans, Why sex? A pluralist approach revisited. *Trends*
437 *Ecol. Evol.* **32**, 589–600 (2017).
- 438 2. N. P. Sharp, S. P. Otto, Evolution of sex: Using experimental genomics to select among
439 competing theories. *Bioessays*. **38**, 751–757 (2016).
- 440 3. G. Bell, *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ of
441 California Press, Los Angeles, 1982).
- 442 4. G. C. Williams, *Sex and Evolution* (Princeton University Press, 1975).
- 443 5. J. Maynard Smith, *The Evolution of Sex* (CUP, 1978).
- 444 6. J. Gerritsen, Sex and Parthenogenesis in Sparse Populations. *Am. Nat.* **115**, 718–742
445 (1980).
- 446 7. S. K. Jain, The Evolution of Inbreeding in Plants. *Annu. Rev. Ecol. Syst.* **7**, 469–495
447 (1976).
- 448 8. J. Felsenstein, The evolutionary advantage of recombination. *Genetics*. **78**, 737–756
449 (1974).
- 450 9. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.*
451 **8**, 269–294 (1966).
- 452 10. P. D. Keightley, S. P. Otto, Interference among deleterious mutations favours sex and
453 recombination in finite populations. *Nature*. **443**, 89–92 (2006).
- 454 11. N. H. Barton, Why sex and recombination? *Cold Spring Harb. Symp. Quant. Biol.* **74**,
455 187–195 (2009).
- 456 12. C. W. Birky Jr, Heterozygosity, heteromorphy, and phylogenetic trees in asexual
457 eukaryotes. *Genetics*. **144**, 427–437 (1996).
- 458 13. D. Mark Welch, M. Meselson, Evidence for the evolution of bdelloid rotifers without
459 sexual reproduction or genetic exchange. *Science*. **288**, 1211–1215 (2000).
- 460 14. D. A. Hickey, Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*. **101**,
461 519–531 (1982).
- 462 15. E. S. Dolgin, B. Charlesworth, The effects of recombination rate on the distribution and
463 abundance of transposable elements. *Genetics*. **178**, 2169–2177 (2008).
- 464 16. J. Bast, K. S. Jaron, D. Schuseil, D. Roze, T. Schwander, Asexual reproduction reduces
465 transposable element load in experimental yeast populations. *Elife*. **8** (2019),
466 doi:10.7554/eLife.48548.
- 467 17. K. S. Jaron, J. Bast, R. W. Nowell, T. R. Ranallo-Benavidez, M. Robinson-Rechavi, T.
468 Schwander, Genomic Features of Parthenogenetic Animals. *J. Hered.* (2020),
469 doi:10.1093/jhered/esaa031.

- 470 18. T. Schwander, B. J. Crespi, Multiple direct transitions from sexual reproduction to
471 apomictic parthenogenesis in *Timema* stick insects. *Evolution*. **63**, 84–103 (2009).
- 472 19. T. Schwander, L. Henry, B. J. Crespi, Molecular evidence for ancient asexuality in
473 *Timema* stick insects. *Curr. Biol.* **21**, 1129–1134 (2011).
- 474 20. R. Riesch, M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K.
475 Lucek, E. Hellen, V. Soria-Carrasco, S. R. Dennis, C. F. de Carvalho, R. J. Safran, C. P.
476 Sandoval, J. Feder, R. Gries, B. J. Crespi, G. Gries, Z. Gompert, P. Nosil, Transitions
477 between phases of genomic differentiation during stick-insect speciation. *Nature*
478 *Ecology & Evolution*. **1**, 0082 (2017).
- 479 21. V. Soria-Carrasco, Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S.
480 Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, P.
481 Nosil, Stick insect genomes reveal natural selection's role in parallel speciation.
482 *Science*. **344**, 738–742 (2014).
- 483 22. R. M. Waterhouse, M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E.
484 V. Kriventseva, E. M. Zdobnov, BUSCO Applications from Quality Assessments to Gene
485 Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
- 486 23. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot
487 for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
- 488 24. J. Romiguier, P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R.
489 Derrat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G.
490 Tsagkogeorga, A. A.-T. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glémin, N. Galtier,
491 Comparative population genomics in animals uncovers the determinants of genetic
492 diversity. *Nature*. **515**, 261–263 (2014).
- 493 25. A. Mackintosh, D. R. Laetsch, A. Hayward, B. Charlesworth, M. Waterfall, R. Vila, K.
494 Lohse, The determinants of genetic diversity in butterflies. *Nat. Commun.* **10**, 3466
495 (2019).
- 496 26. T. Schwander, S. Vuilleumier, J. Dubman, B. J. Crespi, Positive feedback in the
497 transition from sexual reproduction to parthenogenesis. *Proc. Biol. Sci.* **277**, 1435–1442
498 (2010).
- 499 27. E. Suomalainen, A. Saura, J. Lokki, *Cytology and evolution in parthenogenesis* (CRC
500 Press, 1987).
- 501 28. J. Engelstädter, Asexual but Not Clonal: Evolutionary Processes in Automictic
502 Populations. *Genetics*. **206**, 993–1009 (2017).
- 503 29. M. Neiman, T. Schwander, Using Parthenogenetic Lineages to Identify Advantages of
504 Sex. *Evol. Biol.* **38**, 115–123 (2011).
- 505 30. S. Glémin, C. M. François, N. Galtier, Genome Evolution in Outcrossing vs. Selfing vs.
506 Asexual Species. *Methods Mol. Biol.* **1910**, 331–369 (2019).
- 507 31. M. Percy, S. Aron, C. Doums, L. Keller, Conditional use of sex and parthenogenesis for
508 worker and queen production in ants. *Science*. **306**, 1780–1783 (2004).
- 509 32. M. O. Lorenzo-Carballea, A. Cordero-Rivera, Thelytokous parthenogenesis in the

- 510 damselfly *Ischnura hastata* (Odonata, Coenagrionidae): genetic mechanisms and lack
511 of bacterial infection. *Heredity* . **103**, 377–384 (2009).
- 512 33. T. J. Treangen, S. L. Salzberg, Repetitive DNA and next-generation sequencing:
513 computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
- 514 34. H. Ellegren, N. Galtier, Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433
515 (2016).
- 516 35. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B.
517 J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in *Timema*
518 stick insects. *Science*. **359**, 765–770 (2018).
- 519 36. V. R. Vickery, Revision of *Timema scudder* (Phasmatoptera: Timematodea) including
520 three new species. *Can. Entomol.* **125**, 657–692 (1993).
- 521 37. J. Bast, D. J. Parker, Z. Dumas, K. M. Jalvingh, P. Tran Van, K. S. Jaron, E. Figuet, A.
522 Brandt, N. Galtier, T. Schwander, Consequences of asexuality in natural populations:
523 insights from stick insects. *Mol. Biol. Evol.* **35**, 1668–1677 (2018).
- 524 38. L. Henry, T. Schwander, B. J. Crespi, Deleterious mutation accumulation in asexual
525 *Timema* stick insects. *Mol. Biol. Evol.* **29**, 401–408 (2012).
- 526 39. M. Neiman, P. G. Meirmans, T. Schwander, S. Meirmans, Sex in the wild: How and why
527 field-based studies contribute to solving the problem of sex. *Evolution*. **72**, 1194–1203
528 (2018).
- 529 40. S. P. Otto, Selective Interference and the Evolution of Sex. *J. Hered.* (2020),
530 doi:10.1093/jhered/esaa026.
- 531 41. M. J. McDonald, D. P. Rice, M. M. Desai, Sex speeds adaptation by altering the
532 dynamics of molecular evolution. *Nature*. **531**, 233–236 (2016).
- 533 42. O. Kaltz, G. Bell, The ecology and genetics of fitness in *Chlamydomonas*. XII. Repeated
534 sexual episodes increase rates of adaptation to novel environments. *Evolution*. **56**,
535 1743–1753 (2002).
- 536 43. M. R. Goddard, H. C. J. Godfray, A. Burt, Sex increases the efficacy of natural selection
537 in experimental yeast populations. *Nature*. **434**, 636–640 (2005).
- 538 44. I. I. Davydov, N. Salamin, M. Robinson-Rechavi, Large-Scale Comparative Analysis of
539 Codon Models Accounting for Protein and Nucleotide Selection. *Mol. Biol. Evol.* **36**,
540 1316–1332 (2019).
- 541 45. J. T. Daub, S. Moretti, I. I. Davydov, L. Excoffier, M. Robinson-Rechavi, Detection of
542 Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Mol.*
543 *Biol. Evol.* **34**, 1391–1402 (2017).
- 544 46. J. Liu, M. Robinson-Rechavi, Adaptive Evolution of Animal Proteins over Development:
545 Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo. *Mol. Biol. Evol.*
546 **35**, 2862–2872 (2018).
- 547 47. W. Haerty, S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram, L. K. Sirot, L.
548 Levesque, C. G. Artieri, M. F. Wolfner, A. Civetta, R. S. Singh, Evolution in the fast lane:

- 549 rapidly evolving sex-related genes in *Drosophila*. *Genetics*. **177**, 1321–1335 (2007).
- 550 48. B. Charlesworth, C. H. Langley, The evolution of self-regulated transposition of
551 transposable elements. *Genetics*. **112**, 359–383 (1986).
- 552 49. T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.*
553 **24**, R288–94 (2014).
- 554 50. D. Bachtrog, Y-chromosome evolution: emerging insights into processes of
555 Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
- 556 51. T. Wicker, F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P.
557 Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A. H. Schulman, A unified
558 classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982
559 (2007).
- 560 52. S. D. Jackman, B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G.
561 Jahesh, H. Khan, L. Coombe, R. L. Warren, I. Birol, ABySS 2.0: resource-efficient
562 assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).
- 563 53. K. Sahlin, R. Chikhi, L. Arvestad, Assembly scaffolding with PE-contaminated mate-pair
564 libraries. *Bioinformatics*. **32**, 1925–1932 (2016).
- 565 54. D. R. Laetsch, M. L. Blaxter, BlobTools: Interrogation of genome assemblies. *F1000Res.*
566 **6**, 1287 (2017).
- 567 55. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander,
568 Sex-biased gene expression is repeatedly masculinized in asexual females. *Nat.*
569 *Commun.* **10**, 4638 (2019).
- 570 56. D. J. Parker, J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, T. Schwander,
571 Repeated evolution of asexuality involves convergent gene expression changes. *Mol.*
572 *Biol. Evol.* **36**, 350–364 (2019).
- 573 57. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B.
574 Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F.
575 Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc,
576 N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq
577 using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**,
578 1494–1512 (2013).
- 579 58. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database
580 management tool for second-generation genome projects. *BMC Bioinformatics*. **12**, 491
581 (2011).
- 582 59. E. V. Kriventseva, F. Tegenfeldt, T. J. Petty, R. M. Waterhouse, F. A. Simão, I. A.
583 Pozdnyakov, P. Ioannidis, E. M. Zdobnov, OrthoDB v8: update of the hierarchical
584 catalog of orthologs and the underlying free software. *Nucleic Acids Res.* **43**, D250–6
585 (2015).
- 586 60. C. M. Francois, F. Durand, E. Figuet, N. Galtier, Prevalence and Implications of
587 Contamination in Public Genomic Resources: A Case Study of 43 Reference Arthropod
588 Assemblies. *G3*. **10**, 721–730 (2020).

- 589 61. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND.
590 *Nat. Methods.* **12**, 59–60 (2015).
- 591 62. T. D. Wu, C. K. Watanabe, GMAP: a genomic mapping and alignment program for
592 mRNA and EST sequences. *Bioinformatics.* **21**, 1859–1875 (2005).
- 593 63. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with
594 SiLiX. *BMC Bioinformatics.* **12**, 116 (2011).
- 595 64. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7:
596 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- 597 65. A. Di Franco, R. Poujol, D. Baurain, H. Philippe, Evaluating the usefulness of alignment
598 filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol.*
599 *Biol.* **19**, 21 (2019).
- 600 66. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
601 large phylogenies. *Bioinformatics.* **30**, 1312–1313 (2014).
- 602 67. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A.
603 Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella,
604 D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant
605 calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics.*
606 **43**, 11.10.1–11.10.33 (2013).
- 607 68. X. Chen, O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, A. J. Cox,
608 S. Kruglyak, C. T. Saunders, Manta: rapid detection of structural variants and indels for
609 germline and cancer sequencing applications. *Bioinformatics.* **32**, 1220–1222 (2016).
- 610 69. D. C. Jeffares, C. Jolly, M. Hoti, D. Speed, L. Shaw, C. Rallis, F. Balloux, C. Dessimoz,
611 J. Bähler, F. J. Sedlazeck, Transient structural variations have strong effects on
612 quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061
613 (2017).
- 614 70. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L.
615 Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**,
616 R12 (2004).
- 617 71. C. Goubert, L. Modolo, C. Vieira, C. ValienteMoro, P. Mavingui, M. Boulesteix, De novo
618 assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with
619 dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever
620 mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **7**, 1192–1205 (2015).
- 621 72. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST.
622 *Bioinformatics.* **26**, 2460–2461 (2010).
- 623 73. C. Hoede, S. Arnoux, M. Moisset, T. Chaumier, O. Inizan, V. Jamilloux, H. Quesneville,
624 PASTEC: an automatic transposable element classification tool. *PLoS One.* **9**, e91929
625 (2014).
- 626 74. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman,
627 Gapped BLAST and PSI-BLAST: a new generation of protein database search
628 programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

- 629 75. A. Smit, R. Hubley, P. Green, *RepeatMasker Open-4.0* (2013-2015;
630 <http://www.repeatmasker.org>).
- 631 76. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
632 *arXiv [q-bio.GN]* (2013), (available at <http://arxiv.org/abs/1303.3997>).
- 633 77. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with
634 high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
- 635 78. S. Moretti, B. Laurency, W. H. Gharib, B. Castella, A. Kuzniar, H. Schabauer, R. A.
636 Studer, M. Valle, N. Salamin, H. Stockinger, M. Robinson-Rechavi, Selectome update:
637 quality control and computational improvements to a database of positive selection.
638 *Nucleic Acids Res.* **42**, D917–21 (2014).
- 639 79. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: combining multiple
640 sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
- 641 80. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam,
642 M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of
643 high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*
644 **7**, 539 (2011).
- 645 81. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and
646 accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
- 647 82. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated
648 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973
649 (2009).
- 650 83. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using
651 lme4. *Journal of Statistical Software, Articles*. **67**, 1–48 (2015).
- 652 84. M. Wang, Y. Zhao, B. Zhang, Efficient test and visualization of multi-set intersections.
653 *Sci. Rep.* **5**, 16923 (2015).
- 654 85. A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from
655 gene expression data by decorrelating GO graph structure. *Bioinformatics*. **22**,
656 1600–1607 (2006).
- 657 86. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and
658 ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577
659 (2007).
- 660 87. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
661 sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).
- 662 88. J. O'Connell, O. Schulz-Trieglaff, E. Carlson, M. M. Hims, N. A. Gormley, A. J. Cox,
663 NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics*. **31**, 2035–2037
664 (2015).
- 665 89. J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, I. Birol, ABySS: a
666 parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
- 667 90. R. Chikhi, P. Medvedev, Informed and automated k-mer size selection for genome

- 668 assembly. *Bioinformatics*. **30**, 31–37 (2014).
- 669 91. R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang,
670 G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.-M. Yiu,
671 S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, J. Wang,
672 SOAPdenovo2: an empirically improved memory-efficient short-read de novo
673 assembler. *Gigascience*. **1**, 18 (2012).
- 674 92. E. W. Sayers, J. Beck, J. R. Brister, E. E. Bolton, K. Canese, D. C. Comeau, K. Funk, A.
675 Ketter, S. Kim, A. Kimchi, P. A. Kitts, A. Kuznetsov, S. Lathrop, Z. Lu, K. McGarvey, T. L.
676 Madden, T. D. Murphy, N. O’Leary, L. Phan, V. A. Schneider, F. Thibaud-Nissen, B. W.
677 Trawick, K. D. Pruitt, J. Ostell, Database resources of the National Center for
678 Biotechnology Information. *Nucleic Acids Res.* **48**, D9–D16 (2020).
- 679 93. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing
680 reads. *EMBnet J.* **17**, 10–12 (2011).
- 681 94. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M.
682 Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**,
683 15–21 (2013).
- 684 95. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
685 quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
- 686 96. M. S. Campbell, C. Holt, B. Moore, Genome annotation and curation using MAKER and
687 MAKER-P. *Current protocols in* (2014) (available at
688 <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471250953.bi0411s48>).
- 689 97. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: ab
690 initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).
- 691 98. UniProt Consortium, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*
692 **47**, D506–D515 (2019).
- 693 99. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics*. **5**, 59 (2004).
- 694 100. S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M.
695 Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and
696 data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
- 697 101. A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a
698 universal tool for annotation, visualization and analysis in functional genomics research.
699 *Bioinformatics*. **21**, 3674–3676 (2005).
- 700 102. D. Fontaneto, C. Q. Tang, U. Obertegger, F. Leasi, T. G. Barraclough, Different
701 diversification rates between sexual and asexual organisms. *Evol. Biol.* **39**, 262–270
702 (2012).
- 703 103. E. A. Gladyshev, M. Meselson, I. R. Arkhipova, Massive horizontal gene transfer in
704 bdelloid rotifers. *Science*. **320**, 1210–1213 (2008).
- 705 104. J.-F. Flot, B. Hespeels, X. Li, B. Noel, I. Arkhipova, E. G. J. Danchin, A. Hejnl, B.
706 Henrissat, R. Koszul, J.-M. Aury, V. Barbe, R.-M. Barthélémy, J. Bast, G. A. Bazykin, O.
707 Chabrol, A. Couloux, M. Da Rocha, C. Da Silva, E. Gladyshev, P. Gouret, O.

- 708 Hallatschek, B. Hecox-Lea, K. Labadie, B. Lejeune, O. Piskurek, J. Poulain, F.
709 Rodriguez, J. F. Ryan, O. A. Vakhrusheva, E. Wajnberg, B. Wirth, I. Yushenova, M.
710 Kellis, A. S. Kondrashov, D. B. Mark Welch, P. Pontarotti, J. Weissenbach, P. Wincker,
711 O. Jaillon, K. Van Doninck, Genomic evidence for ameiotic evolution in the bdelloid
712 rotifer *Adineta vaga*. *Nature*. **500**, 453–457 (2013).
- 713 105. R. W. Nowell, P. Almeida, C. G. Wilson, T. P. Smith, D. Fontaneto, A. Crisp, G.
714 Micklem, A. Tunnacliffe, C. Boschetti, T. G. Barraclough, Comparative genomics of
715 bdelloid rotifers: Insights from desiccating and nondesiccating species. *PLoS Biol.* **16**,
716 e2004830 (2018).
- 717 106. E. G. J. Danchin, M.-N. Rosso, P. Vieira, J. de Almeida-Engler, P. M. Coutinho, B.
718 Henrissat, P. Abad, Multiple lateral gene transfers and duplications have promoted plant
719 parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17651–17656
720 (2010).
- 721 107. P. Abad, J. Gouzy, J.-M. Aury, P. Castagnone-Sereno, E. G. J. Danchin, E. Deleury,
722 L. Perfus-Barbeoch, V. Anthouard, F. Artiguenave, V. C. Blok, M.-C. Caillaud, P. M.
723 Coutinho, C. Dasilva, F. De Luca, F. Deau, M. Esquibet, T. Flutre, J. V. Goldstone, N.
724 Hamamouch, T. Hewezi, O. Jaillon, C. Jubin, P. Leonetti, M. Magliano, T. R. Maier, G. V.
725 Markov, P. McVeigh, G. Pesole, J. Poulain, M. Robinson-Rechavi, E. Sallet, B.
726 Ségurens, D. Steinbach, T. Tytgat, E. Ugarte, C. van Ghelder, P. Veronico, T. J. Baum,
727 M. Blaxter, T. Bleve-Zacheo, E. L. Davis, J. J. Ewbank, B. Favery, E. Grenier, B.
728 Henrissat, J. T. Jones, V. Laudet, A. G. Maule, H. Quesneville, M.-N. Rosso, T. Schiex,
729 G. Smant, J. Weissenbach, P. Wincker, Genome sequence of the metazoan
730 plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **26**, 909–915 (2008).
- 731 108. A. Faddeeva-Vakhrusheva, K. Kraaijeveld, M. F. L. Derks, S. Y. Anvar, V.
732 Agamennone, W. Suring, A. A. Kampfraath, J. Ellers, G. Le Ngoc, C. A. M. van Gestel,
733 J. Mariën, S. Smit, N. M. van Straalen, D. Roelofs, Coping with living in the soil: the
734 genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics.* **18**, 493
735 (2017).
- 736 109. G. Schönknecht, A. P. M. Weber, M. J. Lercher, Horizontal gene acquisitions by
737 eukaryotes as drivers of adaptive evolution. *Bioessays.* **36**, 9–20 (2014).
- 738 110. T. Guo, X.-W. Wang, K. Shan, W. Sun, L.-Y. Guo, The Loricrin-Like Protein (LLP) of
739 *Phytophthora infestans* Is Required for Oospore Formation and Plant Infection. *Front.*
740 *Plant Sci.* **8**, 142 (2017).
- 741 111. Z. Yang, Y. Zhang, E. K. Wafula, L. A. Honaas, P. E. Ralph, S. Jones, C. R. Clarke,
742 S. Liu, C. Su, H. Zhang, N. S. Altman, S. C. Schuster, M. P. Timko, J. I. Yoder, J. H.
743 Westwood, C. W. dePamphilis, Horizontal gene transfer is more frequent with increased
744 heterotrophy and contributes to parasite adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **113**,
745 E7010–E7019 (2016).
- 746 112. J. R. Belyeu, M. Chowdhury, J. Brown, B. S. Pedersen, Samplot: A Platform for
747 Structural Variant Visual Validation and Automated Filtering. *bioRxiv* (2020) (available at
748 <https://www.biorxiv.org/content/10.1101/2020.09.23.310110v2.abstract>).
- 749 113. M. Mahmoud, N. Gobet, D. I. Cruz-Dávalos, N. Mounier, C. Dessimoz, F. J.
750 Sedlazeck, Structural variant calling: the long and the short of it. *Genome Biol.* **20**, 246

- 751 (2019).
- 752 114. J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2. *Nat. Methods.* **17**,
753 155–158 (2020).
- 754 115. F. J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von
755 Haeseler, M. C. Schatz, Accurate detection of complex structural variations using
756 single-molecule sequencing. *Nat. Methods.* **15**, 461–468 (2018).
- 757 116. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation
758 Sequencing Data. *BMC Bioinformatics.* **15**, 356 (2014).
- 759 117. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L.
760 Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of
761 RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
- 762 118. A. Löytynoja, N. Goldman, An algorithm for progressive multiple alignment of
763 sequences with insertions. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10557–10562 (2005).
- 764 119. A. A. Comeault, S. M. Flaxman, R. Riesch, E. Curran, V. Soria-Carrasco, Z.
765 Gompert, T. E. Farkas, M. Muschick, T. L. Parchman, T. Schwander, J. Slate, P. Nosil,
766 Selection on a genetic polymorphism counteracts ecological speciation in a stick insect.
767 *Curr. Biol.* **25**, 1975–1981 (2015).
- 768 120. R. Villoutreix, C. F. de Carvalho, V. Soria-Carrasco, D. Lindtke, M. De-la-Mora, M.
769 Muschick, J. L. Feder, T. L. Parchman, Z. Gompert, P. Nosil, Large-scale mutation in the
770 evolution of a gene complex for cryptic coloration. *Science.* **369**, 460–466 (2020).
- 771