

1 **Widespread Recombination Suppression Facilitates Plant Sex**

2 **Chromosome Evolution**

3 *Joanna L. Rifkin**, *Felix E.G. Beaudry**, *Zoë Humphries*, *Baharul I. Choudhury*, *Spencer C.H. Barrett*
4 *and Stephen I. Wright.*

5 *These authors contributed equally to this work.

6 **1 Summary**

7 Classical models suggest recombination rates on sex chromosomes evolve in a stepwise manner to
8 localize the inheritance of sexually antagonistic variation in the sex where it is beneficial, thereby
9 lowering rates of recombination between X and Y chromosomes. However, it is also possible that sex
10 chromosome formation occurs in regions with pre-existing recombination suppression. To evaluate
11 these possibilities, we constructed linkage maps and a chromosome-scale genome assembly for the
12 dioecious plant *Rumex hastatulus*, a species with a young neo-sex chromosome found in part of its
13 geographical range. We found that the ancestral sex-linked region is located in a large region
14 characterized by low recombination. Furthermore, comparison between the recombination landscape
15 of the neo-sex chromosome and its autosomal homologue indicates that low recombination rates
16 preceded sex linkage. Our findings suggest that ancestrally low rates of recombination have facilitated
17 the formation and evolution of heteromorphic sex chromosomes.

18 **2 Introduction**

19 Plant and animal genomes vary widely in recombination rate, both between and along chromosomes
20 (1), but how this variation contributes to genome evolution remains unclear. Recombination rate
21 evolution is likely to be governed by natural selection because it is predicted to affect local adaptation
22 and speciation (2), the efficacy of selection (3, 4), and the maintenance of genetic polymorphism (5,
23 6). However, deterministic models suggest that the conditions under which natural selection favors the

24 invasion of variants that change the rate of recombination (recombination modifiers) are quite
25 restrictive (7). Many other factors, including chromosomal position (8), chromatin structure (9) and
26 transposable elements (10) influence rates of recombination, and evolutionary changes in these
27 properties may indirectly drive recombination rate evolution. Nonetheless, compelling evidence
28 supporting the evolution of recombination rate by natural selection is growing (11, 12). Disentangling
29 the relative contributions of direct selection and other factors in shaping the rate of recombination is
30 essential for detailed understanding of the forces shaping genome evolution.

31 Sex chromosomes are particularly valuable for the study of recombination evolution because
32 they represent an example of convergent recombination suppression, and evolutionary theory predicts
33 an important role for natural selection in this process. Classical models of sex chromosome evolution
34 predict that sex chromosomes evolve from autosomes to alleviate the cost of sexually antagonistic
35 alleles in the sex to which those alleles are deleterious (13, 14). Because of differences between the
36 sexes in their optimal reproductive strategies (15), some alleles beneficial in one sex can be
37 detrimental in the other sex and create a selective load in the population (14, 16). The cost of this
38 genetic load can be resolved by the evolution of sex-specific gene expression (17), or by the invasion
39 of recombination modifiers that link the sexually antagonistic variant with the genomic region
40 responsible for determination of the sex to which that variant is beneficial (18). Thus, sexually
41 antagonistic selection can promote the spread of structural rearrangements including inversions and
42 autosome-sex chromosome fusions (13, 19), or cause the recurrent spread of recombination loss
43 further along the X and Y (20). Theory suggests that this selection has to be very strong, and almost
44 equal and opposite between the sexes (21, 22), for recombination modifiers to invade. Over time,
45 according to these models, recombination suppression spreads in a stepwise fashion along an incipient
46 sex chromosome, leaving a pattern of “strata” with distinct levels of divergence between the X and the
47 Y. Evidence for evolutionary strata between sex chromosomes has been found in several unrelated
48 organisms, including humans (23), chickens (24), and the plant *Silene latifolia* (25). Plant sex
49 chromosomes are generally in earlier stages of divergence than the vertebrate sex chromosomes that

50 led to the development of this “strata” model (26) and are therefore especially useful for identifying
51 the earliest stages of recombination evolution across sex chromosomes.

52 *Rumex hastatulus* is one of a relatively small number of dioecious plants with heteromorphic sex
53 chromosomes (27). It offers a unique model for the study of sex chromosome evolution because of a
54 polymorphism in sex chromosome karyotype (28). Males to the west of the Mississippi River have
55 four autosomes and an XY pair (XY cytotype), while males to the east of the Mississippi River have
56 three autosomes, a single larger X chromosome and two Y chromosomes (XYY cytotype).
57 Cytological studies (29, 30), as well as patterns of chromosome number and size (28, 31), suggest that
58 this sex chromosome polymorphism arose from a Robertsonian fusion between the ancestral X
59 chromosome and an autosome. The species’ current polymorphic karyotype includes both the
60 ancestral and the derived state of this chromosome. *Rumex hastatulus* therefore provides an unusual
61 opportunity to investigate recombination rates before and after linkage to the sex-determining region.
62 The principal goal of our study was to understand the changes in recombination rate associated with
63 sex chromosome turnover in *R. hastatulus* and to relate our results to the classic model of sex
64 chromosome evolution.

65 Here, we present a chromosome-scale genome assembly of *R. hastatulus* and describe the patterns
66 of recombination and genome content on the sex chromosomes and autosomes. With these data, we
67 investigate whether the pattern of recombination rate heterogeneity and polymorphism is consistent
68 with the classic stepwise model of sex chromosome evolution.

69 **3 Results**

70 **3.1 The sex-linked region in the XY cytotype exhibits low sex-averaged recombination rates**

71 For the genome assembly, we sequenced PACBio reads for one male individual from the XY cytotype
72 (see supplementary materials). Following *de novo* assembly of PACBio SMRTcell reads, contigs
73 were assembled into longer scaffolds using *in vitro* reconstituted chromatin Chicago libraries (32),
74 followed by Hi-C chromosome conformation capture libraries (33, 34). This primary assembly

75 comprised 1.647 Gb of the *R. hastatulus* genome, with half of the genome assembled into 25 scaffolds
76 larger than 11.886Mb (N50). Our assembly size is consistent with C-values for *R. hastatulus* based on
77 flow cytometry (30; Table S1), representing approximately 92% of the estimated total genome size.

78 Linkage mapping allowed for further scaffolding and correction of misjoins to assemble the
79 genome into the expected five major scaffolds, representing the five chromosome pairs of the XY
80 cytotype of *R. hastatulus* (28). Using RNAseq from 96 individuals representing both sexes (see Table
81 S2 for sex information and supplementary methods for linkage mapping information), we constructed
82 the final genetic map from 988 independent markers on five linkage groups (Table S3). These five
83 chromosomal scaffolds comprised a total of 1.08Gb, or 65% of the primary assembly. Based on the
84 patterns of recombination suppression, our linkage mapping indicated three sub-metacentric and two
85 metacentric chromosomes (green in Figure 1A), which is consistent with the microscopic karyotype
86 (28).

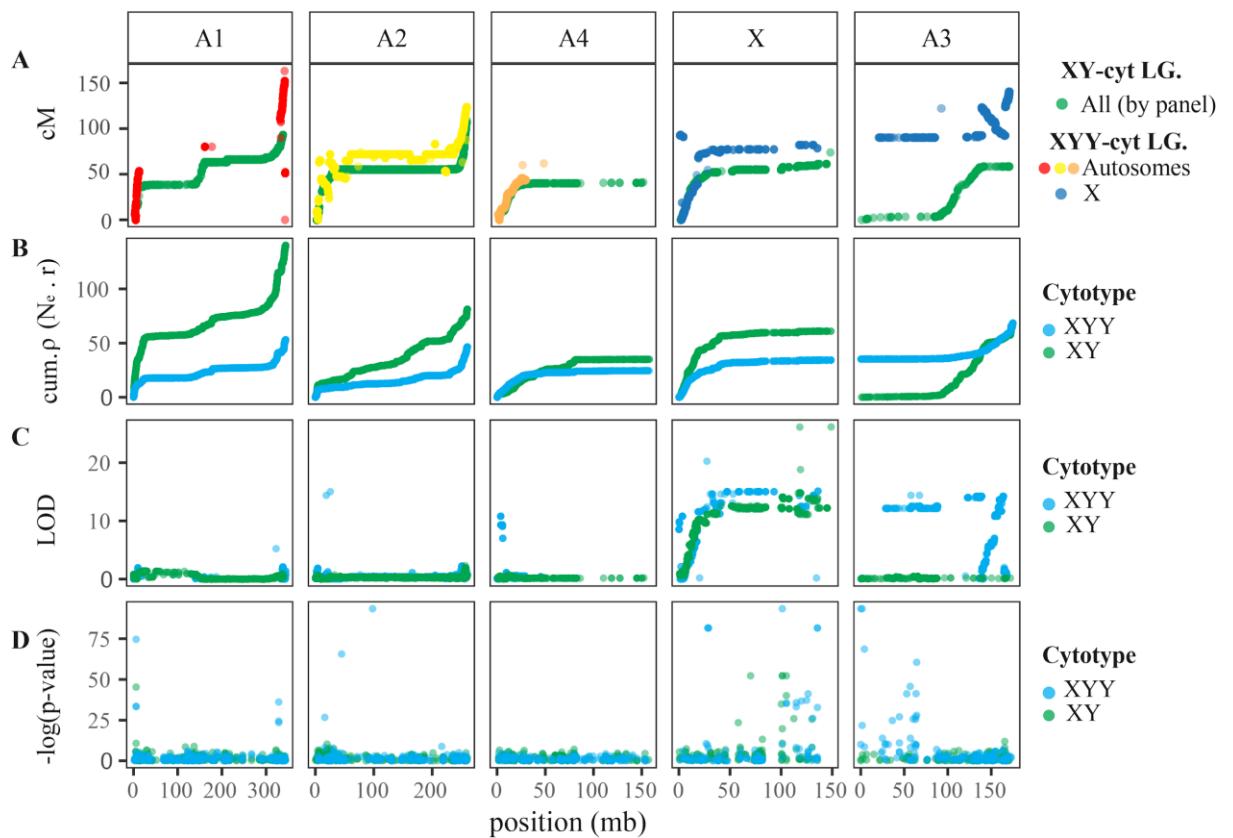
87 Using our chromosome-level assembly, we first used existing RNAseq data from a within-
88 population cross (35) to identify SNPs showing X-linked, Y-linked, X-hemizygous, and autosomal
89 segregation. From this, one linkage group contained the majority of X- and Y-linked SNPs, as well as
90 loci following hemizygous inheritance patterns (Table S4). In particular, of the sex-linked SNPs
91 identified on our five major chromosomal scaffolds, 97% of Y-linked SNPs, 96% of X-linked SNPs
92 and 70% of hemizygous SNPs were mapped onto this scaffold, hereafter referred to as the X. Overall,
93 52% of sex-linked SNPs were mapped to the major chromosomal scaffolds, indicating that a
94 significant proportion of the sex chromosome sequence could not be positioned in our linkage map,
95 but was present in our assembly as smaller scaffolds. In contrast, 84% of autosomal SNPs were
96 mapped to the major scaffolds; this suggests that the X chromosome assembly is less complete, likely
97 due to the high effective heterozygosity in joint assembly of X and Y chromosomes, and to high
98 repeat density on the Y (30). As expected, windowed analyses of sex-associated SNPs from the
99 experimental cross used for the linkage map showed largely similar patterns. With a quantitative trait
100 locus (QTL) mapping approach, all markers above 9.78Mb on the X chromosome were significantly
101 associated with sex phenotype (LOD, $P < 0.01$, Figure 1C, 2D). Thus, overall, we have identified a

102 very large (>140Mb) sex-linked region comprising the vast majority of the X chromosome. The SNP
103 segregation patterns and QTL analyses also suggested the presence of a pseudoautosomal region of X-
104 Y recombination of about 9-10Mb at the beginning of the chromosome (Figures 1 and 2).

105 Since cross-based analysis may not capture rare recombination events between the X and Y
106 chromosomes, we also used population-level RNA sequence data (35, 36) to assess the boundaries of
107 the sex-linked region. We identified sites as fixed differences between the X and Y chromosome
108 when all males were heterozygous and all females were homozygous, allowing for no genotyping
109 errors. The vast majority of X-Y fixed differences in the population were on the X (Table S5). At the
110 population level, the region of X-Y fixed differences was slightly narrower than in the crossing data,
111 suggesting that this approach did indeed capture more recombination events (Figure 2B). We also
112 conducted a genome-wide association study (GWAS) for sex phenotype using population-level
113 genotyping-by-sequencing (GBS) data (36). We found that regions with significant association with
114 sex, after correcting for multiple hits, were located on the X (Figure 1D, 2E). As these multiple,
115 largely independent approaches converged on one region of one chromosome and showed a
116 significant association with sex, we conclude that we have effectively identified the X-chromosome.

117 112Mb of the 151Mb X-chromosome were within 9cM of each other. This suggests that the
118 sex-averaged recombination rate in the sex-linked region of the X chromosome was very low (<0.1
119 cM/Mb). This does not simply reflect suppressed recombination between the X and Y, because our
120 estimate of recombination rate was sex-averaged. Even if X-Y recombination in males was zero, we
121 would estimate a female recombination rate of 0.16 cM/Mb in this region, which is much lower than
122 the average of 1.87 cM/Mb in the pseudoautosomal region. To determine whether this pattern of low
123 recombination was reflected at the population-level scale, we estimated ρ , the effective recombination
124 rate ($N_e \cdot r$), using LD_hat (37, 38) from our population-level RNAseq sampling data (35, 36). At the
125 population level, which is also sex-averaged, we observed very low effective rates of recombination
126 in the sex-linked region (Figure 1B). Since the low rate of recombination is likely to be the result of
127 X-X as well as X-Y recombination rates, recombination may have been ancestrally low in this region.
128 Low rates of recombination are common in regions surrounding the centromere (39), and it therefore

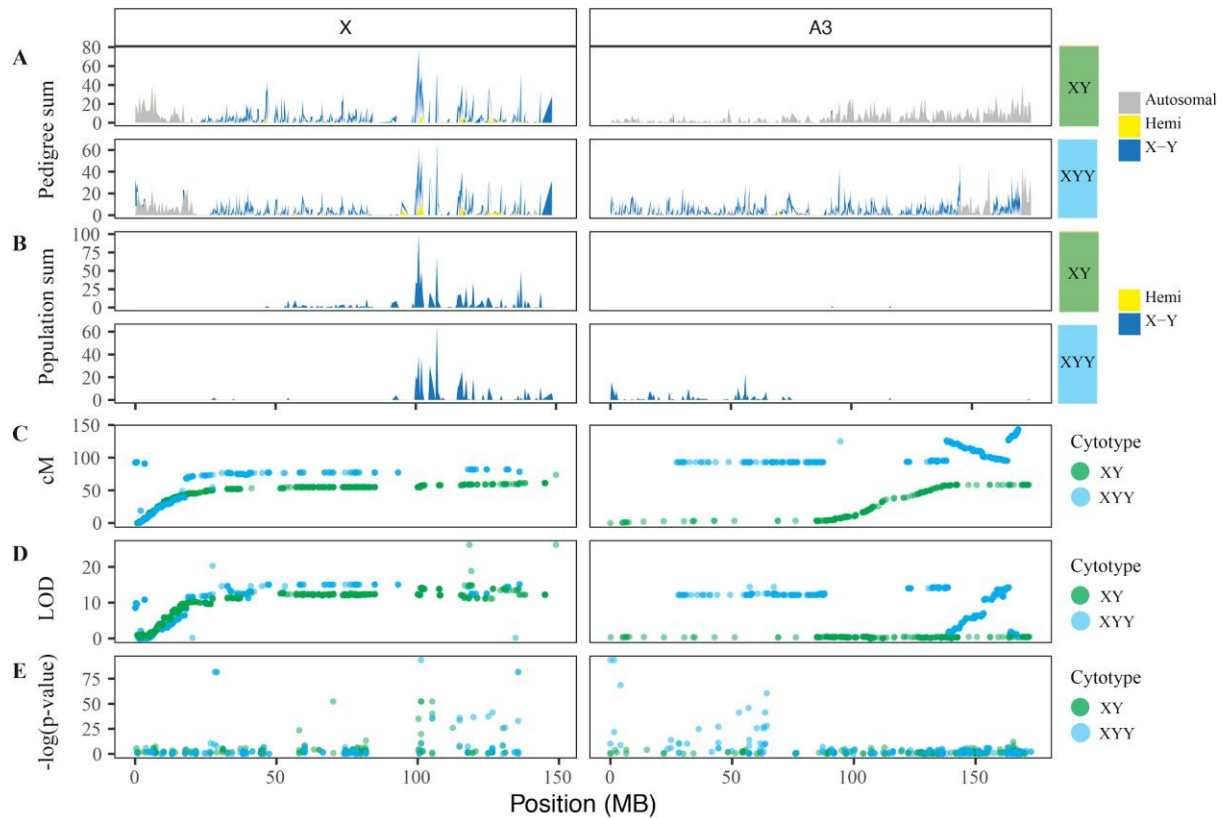
129 seems likely that the sex-linked region is within a very large pericentromeric region. We observed
 130 considerable heterogeneity in the extent of X-Y differentiation along the X chromosome (Figure 2B).
 131 Although this is consistent with our previous results, which also identified varying levels of
 132 divergence possibly consistent with strata (35), a clear stepwise pattern is not apparent from the
 133 ordered data. However, the suppression of recombination limited our ability to order scaffolds based
 134 on recombination position in the non-recombining region, and our results could suggest the presence
 135 of 2-3 strata on the X chromosome.



136

137 **Figure 1.** Recombination rate (cM), cumulative recombination rate (ρ), QTL LOD for sex, and
 138 GWAS for sex across the five chromosomes of the XY-cytotype *Rumex hastatulus*. **A.** Marey map
 139 relating linkage map recombination position (cM) to physical genetic position in the XY-cytotype *R.*
 140 *hastatulus* genome assembly (Mb - panels are XY-cytotype chromosomes) for the XY cytotype (all
 141 green) and XYY cytotype (LG split by color). **B.** Cumulative effective recombination rate estimated
 142 from population-level genetic sequencing for the XY cytotype (green) and XYY cytotype (blue). **C.**
 143 QTL analysis for sex as a binary trait, for the XY cytotype (green) and XYY cytotype (blue). **D.**

144 GWAS for association with sex in population GBS data from (36) for the XY cytotype (green) and
145 XYY cytotype (blue).



146

147 **Figure 2.** Sex-specific associations across the ancestral X chromosome (panel X)
148 chromosome (panel A3). **A-B.** Sum of SNPs identified as autosomal (gray – A only), hemizygous
149 (yellow), and X- or Y-linked (dark blue) for the XY cytotype (top panel) and XYY cytotype (bottom-
150 panel) in **A.**, cross data from (35), and **B.**, population-wide data from (35, 36). **C.** Marey map relating
151 linkage map recombination position (cM) to physical genetic position in the XY-cytotype *Rumex*
152 *hastatus* genome assembly (Mb - panels are XY-cytotype chromosomes) for the XY cytotype
153 (green) and the XYY cytotype (blue). **D.** QTL analysis of sex as a binary trait, for the XY cytotype
154 (green) and XYY cytotype (blue). **E.** GWAS for association with sex in population GBS data from
155 (36) for the XY cytotype (green) and XYY cytotype (blue).

156 **3.2 Neo-XY shows low recombination rates**

157 To determine whether low rates of recombination are a widespread feature of sex chromosomes in *R.*
158 *hastatulus*, we also considered the recombination rate on the neo-X in the XYY cytotype. We
159 constructed a linkage map using 877 independent markers for the XYY cytotype, complementary to
160 the XY-cytotype map (Figure 1A). Recombination rate estimates indicated that the linkage groups
161 represent three metacentric chromosomes and one submetacentric chromosome, suggesting a
162 reduction in chromosome count, as expected given the X-autosome fusion and consistent with the
163 karyotype of this cytotype (28). A single large metacentric linkage group in the XYY-cytotype
164 linkage map joined the XY-cytotype X chromosome with chromosome A3, suggesting A3 in the XY
165 cytotype is the autosomal homologue to the neo-X chromosome. Our linkage map also identified a
166 large inversion on the recombining end of the neo-X. Inversions have been shown to be important in
167 the evolution of sex chromosomes, as well as in ecological differentiation and reproductive isolation
168 (11, 12). Using cross data from the XYY cytotype from (35), we identified X- and Y-linked SNPs on
169 both the ancestral (panel X) and neo-X (panel A3) sections of the large fused X chromosome (Figure
170 1B, 2C; Table S4).

171 Our windowed analyses of sex-linked SNPs also confirmed that sex-linkage extends across
172 both the ancestral X and the neo-X sex chromosomes (Figure 2A, 2B), consistent with an X-autosome
173 fusion event. At the population level, we observed fixed X-Y differences on both the ancestral and
174 neo-X, but the region of fixed X-Y differences on the ancestral X was less extensive than in the XY
175 cytotype (Figure 2B). Both QTL analysis (Figure 2D, LOD $P < 0.01$) and GWAS (Figure 2E, lnL,
176 $P < 0.01$) of the XYY cytotype identified large low-recombination regions on both the X and A3
177 chromosomes to be significantly associated with sex phenotype.

178 Like the ancestral X chromosome, the sex-linked region of the neo-X exhibited a very large
179 non-recombining region: 107Mb of the neo-X is characterized by a total of 0.56cM (0.0052cM/Mb).
180 However, the severity of recombination suppression is even greater in the context of the large fused
181 chromosome. Interestingly, recombination rates on these chromosomes were similar between the

182 cytotypes, including a 92.16Mb region of 5.682cM (0.074cM/Mb) on A3 (the homologue of the neo-
183 X), despite A3 segregating independently from the sex chromosomes and showing no signal of sex
184 linkage in the XY cytotype (Figure 1A, 2A). This pattern accords with our population-level estimates
185 of recombination rate (Figure 1B). Taken together, this evidence implies that strong recombination
186 suppression in the genomic region that formed the neo-sex chromosome of the XYY-cytotype in *R.*
187 *hastatulus* preceded its status as a sex chromosome.

188 **3.3 Low recombination is a genome-wide phenomenon**

189 The sex chromosomes of *R. hastatulus* were not unusual in exhibiting suppressed recombination. Our
190 analyses revealed that all autosomes also had massive (>100Mb) regions of suppressed
191 recombination, with evidence for recombination restricted primarily to the very tips of the
192 chromosomes (Figure 1A, B). This striking finding is consistent with patterns observed from
193 comparative data, which suggest that species with large chromosomes, such as *R. hastatulus*, often
194 have highly peripheral recombination (8). However, *R. hastatulus* appears to represent an extreme
195 case, with all chromosomes exhibiting over one hundred megabases with recombination rates near
196 zero. Remarkably, approximately 80% of the genome, including the sex chromosomes, exhibits
197 highly suppressed recombination, and given that our linkage map contains only 65% of the complete
198 genome this is likely an underestimate.

199 **4 Discussion**

200 The major finding of this study is that the sex-linked regions of both cytotypes of *R. hastatulus* are
201 embedded in a region of highly suppressed recombination which cannot be explained simply by a lack
202 of X-Y recombination. The classic model of sex chromosome evolution assumes that the invasion of
203 recombination modifiers is subsequent to the appearance and maintenance of sexually antagonistic
204 variants (13, 18). However, this may not be necessary if the region already has low rates of
205 recombination (40, 41). In regions with low rates of recombination, the cost of the invasion and
206 maintenance of sexually antagonistic variation can be avoided from the start. Indeed, low rates of

207 recombination are predicted to increase the likelihood of the maintenance of sexually antagonistic
208 polymorphisms (18, 22, 40, 42). Thus, regions of the genome with low rates of recombination may be
209 generally predisposed to evolve sex-linked regions (40). Indeed, evidence for a role for ancestrally
210 low rates of recombination in the evolution of sex chromosomes has started to emerge in plants, as
211 reported in papaya, *Carica papaya*, (43) and kiwifruit, *Actinidia chinensis*, (45). Analogously, self-
212 incompatibility alleles in *Petunia* have been identified in a region of low recombination close to a
213 centromere (45). By comparing a neo-sex chromosome to its ancestral autosome, we provide direct
214 evidence suggesting that suppressed recombination was indeed the ancestral state prior to the
215 evolution of sex-linkage.

216 This ancestral state is part of a genome-wide global pattern of large, recombination-
217 suppressed chromosomes that produces extremely high linkage disequilibrium even in a dioecious
218 obligate outcrosser. The recombination landscape we observed, of high recombination near
219 chromosome ends and suppressed recombination away from the tips, is consistent with a “telomere-
220 initiate” model of recombination interference (8, 46). Simulations suggest this pattern can lead to
221 heterogeneity in divergence between incipient species at chromosome centers (47), a pattern that may
222 be similar to sex chromosome evolution. Although recombination rates have not yet been quantified
223 in other *Rumex* species, it is noteworthy that sex chromosomes have arisen only in sections of the
224 genus with reduced chromosome numbers (48).

225 Our results, indicating that low recombination is likely ancestral to the evolution of sex and
226 neo-sex chromosomes in *R. hastatulus*, may seem at odds with two earlier findings from our work on
227 this species: 1) Here we have shown that recombination rate is low in both males and females in this
228 region, yet previous polymorphism analyses reported dramatically lower amounts of genetic diversity
229 caused by linked selection on the Y but not on the X chromosome (49); 2) Given that ancestrally low
230 recombination removes the requirement for the recruitment of recombination modifiers for sex
231 chromosome expansion, how can we also account for our earlier results demonstrating that genes with
232 pollen-biased expression are disproportionately found on the Y chromosome (50)?

233 The resolution to these apparent contradictions may involve the nonlinear dynamics that
234 likely occur in the transition from low rates of recombination to an absence of recombination. When
235 recombination is low, even very small changes to recombination rate can have important effects on
236 the fixation of adaptive or deleterious mutations (51). For example, simulations suggest that a very
237 low rate of recombination in the European common frog, *Rana temporaria*, can account for the
238 maintenance of high genetic diversity on the Y (52) and, similarly, of high genetic diversity across the
239 genome of the largely asexual apomictic goldilocks buttercup, *Ranunculus auricomus* (53). In our
240 case, a low rate of recombination on the X in females of *R. hastatulus* is not the same genomic
241 environment as the complete absence of recombination on the Y, and the low rates of recombination
242 on the X may be sufficient to maintain genetic diversity in this region. This would imply that the Y
243 and possibly the neo-Y chromosomes have experienced additional recombination suppression.

244 If this interpretation is correct and recombination suppression has progressed further on the Y,
245 what is the relevance of an initially low-recombination origin for the sex chromosomes? Low but non-
246 zero rates of recombination are in fact likely to be the optimal parameter space for the invasion of
247 recombination modifiers. Indeed, very tight linkage is required for the maintenance of sexually
248 antagonistic polymorphism and, subsequently, for the invasion of recombination modifiers on the sex
249 chromosomes (13). The very low, but non-zero, rates of recombination on the sex chromosomes
250 demonstrated in our study are thus suitable for the maintenance of polymorphism along the proto-sex
251 chromosomes, for example, see Figure 1 in (22). Low rates of recombination may allow for the
252 invasion of recombination modifiers completely linking haploid-expressed (pollen) genes to the Y
253 (50, 54). Therefore, there are theoretical reasons to expect that pre-existing low-recombination plays a
254 crucial role in sex-chromosome formation (40), and in both kiwifruit and papaya, sex-determining
255 regions originated in centromeric regions that were likely ancestrally recombination-suppressed (43,
256 44). In both of these systems, however, the sex chromosomes are small and homomorphic or micro-
257 heteromorphic. In *R. hastatulus*, the massive genome-wide scale of recombination suppression may
258 have contributed to the ongoing formation of large, heteromorphic sex chromosomes. The ancestral

259 recombination landscape may thus be a major determinant of the genomic structure of sex
260 chromosomes.

261 **5 Methods**

262 **5.1 Preliminary genome assembly**

263 We sent three grams of leaf tissue from an *R. hastatulus* male F₁ from two parents from Wesley
264 Chapel, TX (TX-WES in (55)) to Dovetail Genomics LLC, Santa Cruz, California 95060, USA. At
265 Dovetail, high molecular weight (HMW) DNA was extracted and sequenced on 15 PacBio
266 SMRTcells (Single Molecule, Real-Time; (56)). The sample was sequenced to 35x coverage for a
267 total of 6.7M reads. After error correction, 5.7M reads were retained (24x coverage) with an N50 of
268 9.5kb. The error-corrected reads were assembled by Dovetail Genomics into a primary assembly
269 using Falcon (57, 58) and the assembly was polished with Arrow from the PacBio GenomicConsensus
270 toolkit (<https://github.com/PacificBiosciences/GenomicConsensus>). This assembly yielded 43,461
271 contigs with an N50 of 74.7kb.

272 We sent an additional three grams of leaf tissue from a full-sib male for two orders of further
273 scaffolding by Dovetail Genomics to improve the primary PacBio-Falcon assembly. The assembly
274 was first scaffolded using the Chicago technique (32), which uses *in vitro* reconstituted chromatin for
275 positioning. Two Chicago libraries were sequenced from ~500ng of HMW gDNA reconstituted into
276 chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with DpnII, 5'
277 overhangs were filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation,
278 crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove
279 biotin that was not internal to ligated fragments. The DNA was sheared to ~350bp mean fragment size
280 and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible
281 adapters. Dovetail isolated biotin-containing fragments using streptavidin beads before PCR
282 enrichment of each library. Libraries were sequenced on an Illumina HiSeq X. The number and length
283 of read pairs produced was: 189 million, 2x150bp for library 1; 170 million, 2x150bp for library 2.
284 Together, these Chicago library reads provided 33.42x physical coverage of the genome (1-100kb).

285 These reads were used to scaffold the PacBio-Falcon assembly using the HiRise pipeline (32) which
286 is designed specifically for using proximity ligation data to scaffold genome assemblies. Dovetail
287 conducted an iterative analysis. They aligned Shotgun and Chicago library sequences to the draft
288 input assembly using a modified SNAP read mapper (<http://snap.cs.berkeley.edu>). The separations of
289 Chicago read pairs mapped within draft scaffolds were then analyzed by HiRise to produce a
290 likelihood model for genomic distance between read pairs, and the model was then used to identify
291 and break putative misjoins, to score prospective joins, and to make joins above a threshold. The
292 longest scaffold increased from 600kb to 1,977kb, and the L50/N50 increased from 0.075Mb in 6,213
293 scaffolds to 0.248Mb in 1,887 scaffolds.

294 We further improved the Chicago-scaffolded assembly using 2 Hi-C chromosome
295 conformation capture libraries from Dovetail (33, 34). Briefly, for each library, chromatin was fixed
296 in place with formaldehyde in the nucleus and extracted. Fixed chromatin was digested with DpnII, 5'
297 overhangs were filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation,
298 crosslinks were reversed and the DNA was purified from protein. Purified DNA was treated to
299 remove biotin not internal to ligated fragments. The DNA was then sheared to ~350bp mean fragment
300 size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible
301 adapters. Streptavidin beads were used to isolate biotin-containing fragments before PCR enrichment
302 of each library. Dovetail sequenced these libraries on an Illumina HiSeq X. The number and length of
303 read pairs produced for each library was: 158 million, 2x150bp for library 1; 194 million, 2x150bp for
304 library 2. Together, these Dovetail HiC library reads provided 87.47x physical coverage of the
305 genome (10-10,000kb). After assembly with HiRise, the N50/L50 of the assembly increased to
306 11.89Mb in 25 scaffolds and a longest scaffold of 146,334kb.

307 We obtained C-values for genome size estimation from Plant Cytometry Services of Didam,
308 The Netherlands. We shipped fresh leaf tissue, and DNA content was estimated with flow cytometry
309 relative to *Vinca minor* with both DAPI and PI staining.

310 **5.2 Linkage mapping**

311 We generated F₂ linkage-mapping populations for both the XY and XYY cytotypes of 96 offspring
312 each. The original parents were collected from Wesley Chapel, TX and Marion, SC, respectively (see
313 (55)). Seeds from F₁ plants were sterilized in 5% (V/V) bleach for one minute and then washed in
314 running tap water and distilled water. Sterilized seeds were spread on wet filter paper in Petri dishes
315 and incubated in the dark at 4°C to germinate. After germination (usually within 2-3 weeks), we
316 transplanted seedlings into 6-inch plastic pots filled with Promix® soil and sand (3:1 ratio) with
317 300ml of Nutricote® (14:13:13, slow releasing fertilizer) for each 60lbs of Promix soil. We grew
318 seedlings in a glasshouse set for 22°C daytime, 18°C nighttime temperature and 16-hour day length at
319 the University of Toronto St. George campus. We watered on alternate days and randomized pots
320 twice weekly for uniform growing conditions and to avoid edge effects. We phenotyped plants for sex
321 at onset of flowering. After plants were sexed, we collected approximately 30mg of young and
322 healthy leaf tissue from individual plants and flash froze it in liquid nitrogen for RNA extraction.
323 Total RNA was extracted using Spectrum™ plant total RNA Kit (Sigma-Aldrich) according to
324 manufacturer's instructions. We sent RNA samples to Genome Quebec Innovation Centre (McGill
325 University), Montreal, QC for library preparation and sequencing. Libraries were prepared using
326 NEB® mRNA stranded library preparation method and sequenced on two lanes of Illumina NovaSeq
327 S2 PE100 (2x100) sequencing platform using 96 barcodes. A total of ~6.25 billion reads
328 (6,244,145,277) were generated, ranging from ~13 to ~104 million reads per sample with an average
329 of ~32.5 (32,521,589) (median ~28 million (28,041,358)).

330 We aligned samples to the *R. hastatulus* TX draft assembly using STAR 2-pass
331 (STAR_2.5.3a (59)). We sorted alignments, assigned read groups, and marked PCR duplicates using
332 PicardTools 2.18.21-SNAPSHOT [<https://broadinstitute.github.io/picard/>]. We called variants with
333 bcftools 1.9-67-g626e46b mpileup and call, and filtered using bcftools view for minimum sample
334 depth 10, minimum sample quality 10, minimum site quality 50, allele frequency between 0.1 and 0.9,
335 and minimum 25 individuals called. For use in linkage mapping, we further filtered our variants using
336 bcftools filter and vcftools (60) to remove mislabeled individuals and sites with more than 5% missing

337 data. For linkage mapping, we converted data from vcf to 012 format using vcftools and transformed
338 to csvr format using custom scripts.

339 We generated a linkage map using the R package ASMAP (61), which implements the
340 minimum spanning tree algorithm in an R/qtl-compatible interface (62, 63). We removed two and
341 three individuals from the XY and XYY mapping populations, respectively, because of apparent
342 contamination. Individuals with high missing data and genetic clones were removed. We generated
343 our XY linkage map using only markers on the 100 biggest scaffolds of our draft assembly and 322
344 smaller scaffolds that we identified as sex-linked, which contain 1.2Gb of sequence data. Our XYY
345 linkage map used only markers on the 100 biggest scaffolds of our draft assembly. We filtered our
346 initial set of variants (25010 XY / 25544 XYY) to remove colocalized markers (leaving 2255 XY /
347 2103 XYY) and distorted markers. We constructed the final maps from 988 and 877 markers for XY
348 and XYY, respectively (see Table S1 for progeny sample sizes). We used custom scripts to fuse
349 colocalized markers into the map to maximize the number of scaffolds in the chromosome-scale
350 assembly.

351 For the XY cytotype, we recovered five major linkage groups with over 100 markers each
352 (108-276), as well as six minor fragmentary linkage groups (5-36 markers). The sizes of the major
353 linkage groups ranged from 75.52 to 108.91 cM, with a total map length of 550.126. This is broadly
354 consistent with expectations for a linkage map with five chromosomes. Recombination frequency
355 (cM) roughly correlates with physical length (bp) (Figure S4). We recovered four major linkage
356 groups with over 100 markers each (103-258) for the XYY cytotype, as well as seven minor
357 fragmentary linkage groups (2-52 markers). The largest minor linkage group, LG11, was largely
358 collinear with the sex chromosome (LG10), containing the same scaffolds and overlapping positions.
359 The sizes of the major linkage groups ranged from 61.74 to 167.07 cM, with a total map length of
360 584.819 cM. We assessed recombination rates visually using Marey maps (64) relating physical
361 position of markers along scaffolds to recombination position along chromosomes.

362 We used Chromonomer (<http://catchenlab.life.illinois.edu/chromonomer/>) to relate our
363 linkage map to our genome assembly. We conducted manual edits to our linkage map before using
364 Chromonomer, to position scaffolds unique to minor linkage groups on major linkage groups based on
365 physically nearby markers and to remove minor linkage groups. We used a script from Nucleomics-
366 VIB's BioNano Tools (<https://github.com/Nucleomics-VIB/bionano-tools>) to convert Dovetail's
367 assembly table file into an .agp file. With Chromonomer, we were able to place 1.09GB (65%) of the
368 Dovetail assembly into five main pseudomolecules representing the major chromosomes. For
369 downstream analyses, we created custom Python scripts using the .agp file output from Chromonomer
370 to translate scaffold positions from our draft assembly to positions along the chromosomes.

371 **5.3 Sex-linked variant calling and filtration**

372 For RNA samples from (35, 36), we aligned samples to the *R. hastatulus* XY draft assembly using
373 STAR 2-pass STAR_2.5.3a (59). We aligned reads for GBS samples from (65) using NextGenMap
374 (66). For both alignments, we sorted the reads and assigned read groups using PicardTools 2.18.21-
375 SNAPSHOT [<https://broadinstitute.github.io/picard/>]. We marked PCR duplicates for RNASeq but
376 not for GBS data. For both datasets, variants were called using bcftools 1.9-67-g626e46b mpileup as
377 described above. We filtered the RNAseq datasets for minimum sample quality 20, minimum site
378 quality 20, minor allele frequency greater than 0.04, and no missing data. We filtered the GBS dataset
379 for minimum sample quality 20, minimum site quality 10, minimum mean depth of 6, minor allele
380 frequency greater than 0.05, and no more than 50% missing data. We converted data for windowed
381 analyses from vcf to 012 format using vcftools.

382 We identified SNPs as X-linked, Y-linked, hemizygous, autosomal, or male-only expressed in
383 the cross data, the F₂ data, and the population data using the 012 files described above and custom R
384 scripts based on the segregation patterns described in (35). We converted all sites to positions along
385 chromosomes using custom Python scripts and the Chromonomer .agp file. We summed the different
386 categories of sites across 500kb windows using custom R scripts.

387 **5.4 Quantitative trait locus (QTL) mapping and genome-wide association study (GWAS)**

388 We performed QTL mapping of sex as a binary phenotype using the scan1 function of R/qtl2
389 (67) using the F₂ mapping population. We adjusted eta_max (the maximum value for the linear
390 predictor in the model) downwards until the model was able to converge. We performed a
391 permutation analysis to identify significance thresholds. We performed GWAS analysis in Gemma
392 (68) using a likelihood ratio test for significance.
393

394 References

- 395 1. B. S. Gaut, S. I. Wright, C. Rizzon, J. Dvorak, L. K. Anderson, Recombination: an
396 underappreciated factor in the evolution of plant genomes. *Nat. Rev. Genet.* **8**, 77–84 (2007).
- 397 2. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics*
398 **173**, 419–434 (2006).
- 399 3. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**,
400 269–294 (1966).
- 401 4. J. Felsenstein, The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
- 402 5. J. M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35
403 (1974).
- 404 6. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on
405 neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
- 406 7. B. Charlesworth, D. Charlesworth, Selection of new inversions in multi-locus genetic systems.
407 *Genet. Res.* **21**, 167–183 (1973).
- 408 8. Q. Haenel, T. G. Laurentino, M. Roesti, D. Berner, Meta-analysis of chromosome-scale
409 crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol.*
410 *Ecol.*, 2477–2497 (2018).
- 411 9. K. Ohta, T. Shibata, A. Nicolas, Changes in chromatin structure at recombination initiation
412 sites during yeast meiosis. *EMBO J.* **13**, 5754–5763 (1994).
- 413 10. T. V. Kent, J. Uzunović, S. I. Wright, Coevolution between transposable elements and
414 recombination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372** (2017).
- 415 11. L. H. Rieseberg, C. Van Fossen, A. M. Desrochers, Hybrid speciation accompanied by
416 genomic reorganization in wild sunflowers. *Nature* **375**, 313 (1995).
- 417 12. M. A. F. Noor, K. L. Grams, L. A. Bertucci, J. Reiland, Chromosomal inversions and the
418 reproductive isolation of species. *Proc. Natl. Acad. Sci.* **98**, 12084–12088 (2001).
- 419 13. D. Charlesworth, B. Charlesworth, Sex differences in fitness and selection for centric fusions
420 between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980).
- 421 14. W. R. Rice, Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742
422 (1984).
- 423 15. R. L. Trivers, “Parental investment and sexual selection” in *Sexual Selection and the Descent*
424 *of Man, 1871-1971*, B. Campbell, Ed. (Aldine, 1972), pp. 136–179.
- 425 16. R. Lande, Sexual dimorphism, sexual selection, and adaptation in polygenic characters.
426 *Evolution* **34**, 292–305 (1980).
- 427 17. J. E. Mank, Sex chromosomes and the evolution of sexual dimorphism: lessons from the
428 genome. *Am. Nat.* **173**, 141–150 (2009).
- 429 18. W. R. Rice, The accumulation of sexually antagonistic genes as a selective agent promoting
430 the evolution of reduced recombination between the primitive sex chromosomes. *Evolution* **41**,
431 911–914 (1987).
- 432 19. L. H. Rieseberg, Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358
433 (2001).
- 434 20. J. J. Bull, Sex determining mechanisms: An evolutionary perspective. *Experientia* **41**, 1285–
435 1296 (1985).
- 436 21. J. F. Kidwell, M. T. Clegg, F. M. Stewart, T. Prout, Regions of stable equilibria for models of
437 differential selection in the two sexes under random mating. *Genetics* **85**, 171–183 (1977).
- 438 22. S. P. Otto, Evolutionary potential for genomic islands of sexual divergence on recombining
439 sex chromosomes. *New Phytol.* (2019).
- 440 23. B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science* **286**,
441 964–967 (1999).
- 442 24. L.-J. L. Handley, H. Ceplitis, H. Ellegren, Evolutionary strata on the chicken Z chromosome:
443 implications for sex chromosome evolution. *Genetics* **167**, 367–376 (2004).
- 444 25. R. Bergero, A. Forrest, E. Kamau, D. Charlesworth, Evolutionary strata on the X
445 chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes.
446 *Genetics* **175**, 1945–1954 (2007).
- 447 26. R. Ming, A. Bendahmane, S. S. Renner, Sex chromosomes in land plants. *Annu. Rev. Plant*

- 448 *Biol.* **62**, 485–514 (2011).
- 449 27. D. Charlesworth, Plant sex chromosome evolution. *J. Exp. Bot.* **64**, 405–420 (2013).
- 450 28. B. W. Smith, The evolving karyotype of *Rumex hastatulus*. *Evolution* **18**, 93–104 (1964).
- 451 29. M. Kasjaniuk, A. Grabowska-Joachimciak, A. J. Joachimciak, Testing the translocation
452 hypothesis and Haldane’s rule in *Rumex hastatulus*. *Protoplasma* **256**, 237–247 (2019).
- 453 30. A. Grabowska-Joachimciak, *et al.*, Chromosome landmarks and autosome-sex chromosome
454 translocations in *Rumex hastatulus*, a plant with XX/X₁Y₁Y₂ sex chromosome system.
455 *Chromosom. Res.* **23**, 187–197 (2015).
- 456 31. B. W. Smith, Evolution of sex-determining mechanisms in *Rumex*. *Chromosom. Today* **2**, 172–
457 182 (1969).
- 458 32. N. H. Putnam, *et al.*, Chromosome-scale shotgun assembly using an in vitro method for long-
459 range linkage. *Genome Res.* **26**, 342–350 (2016).
- 460 33. E. Lieberman-Aiden, *et al.*, Comprehensive mapping of long-range interactions reveals folding
461 principles of the human genome. *Science* **326**, 289–293 (2009).
- 462 34. J.-M. Belton, *et al.*, Hi-C: a comprehensive technique to capture the conformation of genomes.
463 *Methods* **58**, 268–276 (2012).
- 464 35. J. Hough, J. D. Hollister, W. Wang, S. C. H. Barrett, S. I. Wright, Genetic degeneration of old
465 and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proc. Natl. Acad. Sci.*
466 **111**, 7713–7718 (2014).
- 467 36. F. E. G. Beaudry, S. C. H. Barrett, S. I. Wright, Ancestral and neo-sex chromosomes
468 contribute to population divergence in a dioecious plant. *Evolution* **54**, 180 (2019).
- 469 37. M. P. H. Stumpf, G. A. T. McVean, Estimating recombination rates from population-genetic
470 data. *Nat. Rev. Genet.* **4**, 959–968 (2003).
- 471 38. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots. *Genome*
472 *Res.* **17**, 1219–1227 (2007).
- 473 39. M. M. Mahtani, H. F. Willard, Physical and genetic mapping of the human X chromosome
474 centromere: repression of recombination. *Genome Res.* **8**, 100–110 (1998).
- 475 40. D. Charlesworth, Young sex chromosomes in plants and animals. *New Phytol.* **224**, 1095–1107
476 (2019).
- 477 41. D. Charlesworth, Does sexual dimorphism in plants promote sex chromosome evolution?
478 *Environ. Exp. Bot.* **146**, 5–12 (2018).
- 479 42. J. D. Fry, The genomic location of sexually antagonistic variation: some cautionary comments.
480 *Evolution* **64**, 1510–1516 (2010).
- 481 43. M. Iovene, Q. Yu, R. Ming, J. Jiang, Evidence for emergence of sex-determining gene(s) in a
482 centromeric region in *Vasconcellea parviflora*. *Genetics* **199**, 413–421 (2015).
- 483 44. S. M. Pilkington, *et al.*, Genetic and cytological analyses reveal the recombination landscape
484 of a partially differentiated plant sex chromosome in kiwifruit. *BMC Plant Biol.* **19**, 172
485 (2019).
- 486 45. R. ten Hoopen, R. M. Harbord, T. Maes, N. Nanninga, T. P. Robbins, The self-incompatibility
487 (*S*) locus in *Petunia hybrida* is located on chromosome III in a region, syntenic for the
488 Solanaceae. *Plant J.* **16**, 729–734 (1998).
- 489 46. S. P. Otto, B. A. Payseur, Crossover interference: shedding light on the evolution of
490 recombination. *Annu. Rev. Genet.* **53**, 19–44 (2019).
- 491 47. D. Berner, M. Roesti, Genomics of adaptive divergence with chromosome-scale heterogeneity
492 in crossover rate. *Mol. Ecol.* **26**, 6351–6369 (2017).
- 493 48. R. Navajas-Pérez, *et al.*, The evolution of reproductive systems and sex-determining
494 mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastial sequence
495 data. *Mol. Biol. Evol.* **22**, 1929–1939 (2005).
- 496 49. J. Hough, W. Wang, S. C. H. Barrett, S. I. Wright, Hill-Robertson interference reduces genetic
497 diversity. *Genetics* **207**, 685–695 (2017).
- 498 50. G. Sandler, F. E. G. Beaudry, S. C. H. Barrett, S. I. Wright, The effects of haploid selection on
499 Y chromosome evolution in two closely related dioecious plants. *Evol. Lett.* **197**, 368–377
500 (2018).
- 501 51. M. Hartfield, S. P. Otto, P. D. Keightley, The role of advantageous mutations in enhancing the
502 evolution of a recombination modifier. *Genetics* **184**, 1153–1164 (2010).

- 503 52. N. Rodrigues, T. Studer, C. Dufresnes, N. Perrin, Sex-chromosome recombination in common
504 frogs brings water to the fountain-of-youth. *Mol. Biol. Evol.* **35**, 942–948 (2018).
- 505 53. D. Hojsgaard, E. Hörandl, A little bit of sex matters for genome evolution in asexual plants.
506 *Front. Plant Sci.* **6**, 82 (2015).
- 507 54. M. F. Scott, S. P. Otto, Haploid selection favors suppressed recombination between sex
508 chromosomes despite causing biased sex ratios. *Genetics* **207**, 1631–1649 (2017).
- 509 55. M. Pickup, S. C. H. Barrett, The influence of demography and local mating environment on
510 sex ratios in a wind-pollinated dioecious plant. *Ecol. Evol.* **3**, 629–639 (2013).
- 511 56. J. Eid, *et al.*, Real-time DNA sequencing from single polymerase molecules. *Science* **323**,
512 132–138 (2009).
- 513 57. C.-S. Chin, *et al.*, Phased diploid genome assembly with single-molecule real-time sequencing.
514 *Nat. Methods* **13**, 1050–1054 (2016).
- 515 58. C.-S. Chin, *et al.*, Nonhybrid, finished microbial genome assemblies from long-read {SMRT}
516 sequencing data. *Nat. Methods* **10**, 563–569 (2013).
- 517 59. A. Dobin, T. R. Gingeras, Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics*
518 **51**, 11.14.1-11.14.9 (2015).
- 519 60. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
520 (2011).
- 521 61. J. Taylor, D. Butler, R package ASMap: efficient genetic linkage map construction and
522 diagnosis (2017).
- 523 62. Y. Wu, P. R. Bhat, T. J. Close, S. Lonardi, Efficient and accurate construction of genetic
524 linkage maps from the minimum spanning tree of a graph. *PLoS Genet.* **4** (2008).
- 525 63. K. W. Broman, S. Sen, A Guide to QTL Mapping with R/qtl (2009)
526 <https://doi.org/10.1007/978-0-387-92125-9>.
- 527 64. A. Chakravarti, A graphical representation of genetic and physical maps: the Marey map.
528 *Genomics* **11**, 219–222 (1991).
- 529 65. F. E. G. Beaudry, S. C. H. Barrett, S. I. Wright, Genomic loss and silencing on the Y
530 chromosomes of *Rumex*. *Genome Biol. Evol.* **9**, 3345–3355 (2017).
- 531 66. F. J. Sedlazeck, P. Rescheneder, A. Von Haeseler, NextGenMap: Fast and accurate read
532 mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).
- 533 67. K. W. Broman, *et al.*, R/qtl2: software for mapping quantitative trait loci with high-
534 dimensional data and multiparent populations. *Genetics* **211**, 495–502 (2019).
- 535 68. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies.
536 *Nat. Genet.* **44**, 821–824 (2012).
- 537

538 **Supplementary Materials**

539 *Supplementary Tables*

540

Table S1. C-values were calculated from two XY-cytotype individuals [an F ₃ Male (ID: TxTx P5 MG) and F ₃ Female (ID: TxTx P5 MG)] of <i>Rumex hastatulus</i>	
	XY
Female	1.8
Male	1.99

541

Table S2. Counts of <i>Rumex hastatulus</i> individuals of each sex and cytotype sequenced for the linkage map.		
	TX (XY)	NC (XYY)
Male	53	40
Female	43	56

542

543

Table S3. Major linkage groups in both maps of <i>Rumex hastatulus</i>					
LG ID numbers after Smith 1964					
TX (XY)					
Description	LG ID	Number of independent markers (colocalized markers)	Length (cM)	Length (bp)	Length (μ m, Grabowska-Joachimiak 2015)
Metacentric autosome	A1	192 (1570)	101.022	344,497,533	3.8
Metacentric autosome	A2	276 (3241)	108.910	260,405,700	5.02
Neo-sex chromosome	A3	108 (338)	95.789	175,013,671	2.35
Sub-metacentric autosome	A4	119 (510)	48.148	158,234,227	2.81
Sex chromosome	XY	187 (1186)	75.521	150,617,407	3.4 (X) 5.89 (Y)
		882 (6845)	429.39	1088768538	

NC (XYY)					
Description	LG ID	Number of independent markers (colocalized markers)	Length (cM)	Length (bp)	Length (physical, as percentage of karyotype, Smith 1964)
Metacentric autosome B	A1	184 (645)	162.541		3.67
Metacentric autosome A	A2	258 (2576)	167.069		4.98
Sub-metacentric autosome	A4	103 (552)	60.524		2.2
Sex chromosome	XYY	250 (1270)	137.695		5.36 (X) 4.22 (Y1) 3.77 (Y2)
		537 (5043)	527.829		

545

Table S4. Sex-linked SNPs of *Rumex hastatulus* based on Hough et al. (2014) pedigree data. TX = XY, NC = XYY.

	Y-linked (TX)	Y-linked (NC)	X-linked (TX)	X-linked (NC)	Hemizygous (TX)	Hemizygous (NC)	Autosomal (TX)	Autosomal (NC)
A1	14	57	5	7	1	1	3894	1692
A2	3	137	1	18	13	5	1700	3950
A3	12	1172	11	763	6	7	1614	410
A4	4	59	0	33	2	0	1173	1691
X	1208	980	439	456	50	79	420	348
Not placed	1048	1110	143	173	317	345	1681	1714

546

547

548

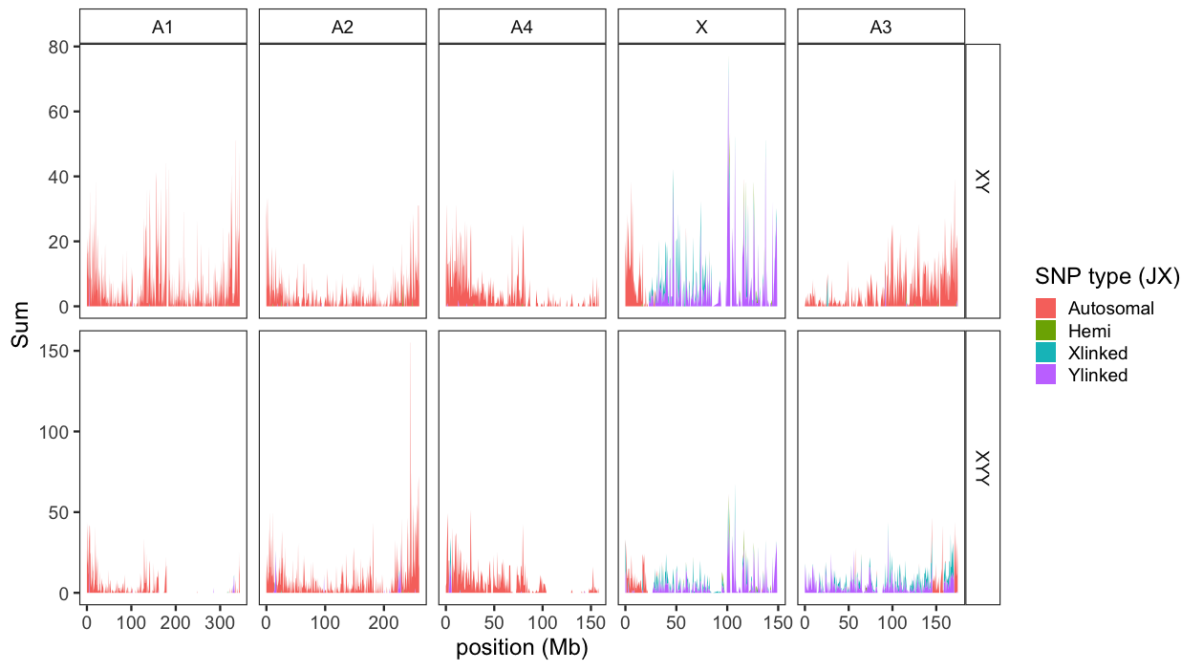
Table S5. Sex-linked SNPs of *Rumex hastatulus* based on Hough et al. (2014) and Beaudry et al. (2019) population data. TX = XY, NC = XYY.

	XY-fixed (TX)	XY-fixed (NC)
A1	17	16

A2	4	76
A3	7	390
A4	5	3
X	1048	583
Not placed	1403	1254

549

550 **Supplementary Figures**



551

552 **Figure S1.** Windowed counts of SNPs identified as autosomal (red), hemizygous (green), and X-

553 (blue) or Y-linked (purple) for the XY cytotype (top panel) and XYY cytotype (bottom-panel) of

554 *Rumex hastatulus* in cross data from (35).

555

556