1 **Interspecific crossing and genetic mapping reveal intrinsic genomic**

2 **incompatibility between two *Senecio* species that form a hybrid zone on**

3 **Mount Etna, Sicily**

4

5 Adrian C. Brennan[1,2,3*], Simon J. Hiscock[4] and Richard J. Abbott[1]

6

7 [1]*School of Biology, Harold Mitchell Building, University of St Andrews, St*

8 *Andrews, Fife, KY16 9TH, UK*

9 [2]*Estación Biológica de Doñana (EBD-CSIC), Avenida Américo Vespucio*

10 *s/n, 41092 Sevilla, Spain*

11 [3]*School of Biological and Biomedical Sciences, University of Durham,*

12 *South Road, Durham, DH1 3LE, UK*

13 [4]*School of Biological Sciences, University of Bristol, Woodland Road,*

14 *Bristol, BS8 1UG, UK*

15

16 *Corresponding author: Adrian C. Brennan, School of Biological and

17 Biomedical Sciences, University of Durham, South Road, Durham, DH1

18 3LE, UK, Tel: 0044 191 3341251, Fax: 0044 191 3341201, email:

19 a.c.brennan@durham.ac.uk

20

21 Running title: Interspecific incompatibility in *Senecio*

22

23 Main text word count: 6892

24

25

ABSTRACT

Studies of hybridizing species can reveal much about the genetic basis and maintenance of species divergence in the face of gene flow. Here we report a genetic segregation and linkage analysis conducted on F_2 progeny of a reciprocal cross between *Senecio aethnensis* and *S. chrysanthemifolius* that form a hybrid zone on Mount Etna, Sicily, aimed at determining the genetic basis of intrinsic hybrid barriers between them. Significant transmission ratio distortion (TRD) was detected at 34 (~27%) of 127 marker loci located in nine distinct clusters across seven of the ten linkage groups detected, indicating genomic incompatibility between the species. TRD at these loci could not be attributed entirely to post-zygotic selective loss of F_2 individuals that failed to germinate or flower (16.7 %). At four loci tests indicated that pre-zygotic events, such as meiotic drive in F_1 parents or gametophytic selection, contributed to TRD. Additional tests revealed that cytonuclear incompatibility contributed to TRD at five loci, Bateson-Dobzhansky-Muller (BDM) incompatibilities involving epistatic interactions between loci contributed to TRD at four loci, and underdominance (heterozygote disadvantage) was a possible cause of TRD at one locus. Major chromosomal rearrangements were probably not a cause of interspecific incompatibility at the scale that could be examined with current map marker density. Intrinsic genomic incompatibility between *S. aethnensis* and *S. chrysanthemifolius* revealed by TRD across multiple genomic regions in early generation hybrids is likely to impact the genetic structure of the natural hybrid zone on Mount Etna by limiting introgression and promoting divergence across the genome.

51

52 **Keywords:** genetic divergence; genetic maps; genomic incompatibility;

53 hybrid zone; intrinsic hybrid barrier; transmission ratio distortion

54

55 **INTRODUCTION**

56

57 Although strong, divergent natural selection can maintain population

58 divergence in the face of gene flow (Nosil, 2012), intrinsic genetic barriers

59 can also evolve under such conditions (Rundle and Nosil, 2005; Agrawal *et*

60 *al.*, 2011) or may already be in place to varying degrees between

61 hybridizing populations that diverged in allopatry (Coyne and Orr, 2004;

62 Bierne *et al.*, 2011; Feder *et al.*, 2012; Abbott *et al.*, 2013). Intrinsic

63 incompatibility between species is generally detected by crossing studies,

64 which can also reveal the nature of such incompatibility at both genetic and

65 genomic levels. For example, transmission ratio distortion (TRD), which is

66 often observed at segregating loci among progeny of interspecific crosses or

67 crosses between divergent lineages within species, can result from selection

68 against particular hybrid genotype combinations at these loci (TRDLs;

69 Fishman *et al.*, 2001; Moyle and Graham, 2006). TRD may result from

70 Bateson-Dobzhansky-Muller (BDM) incompatibilities at haploid and/or

71 diploid stages of the life cycle caused by negative epistatic interactions

72 between nuclear loci showing polymorphisms within and between

73 populations of the same or different species (Fishman *et al.*, 2008; Cutter,

74 2012; Bomblies, 2013; Corbett-Detig *et al.,* 2013; Ouyang and Zhang,

75 2013). It may also arise from cytonuclear incompatibility (Levin, 2003;

76   Fishman and Willis, 2006; Turelli and Moyle, 2007) or from "selfish"

77   meiotic drive of alleles in a new genomic background (Fishman and Willis,

78   2005). Incompatibility caused by cytonuclear or haploid-diploid

79   incompatibilities may depend on cross direction leading to asymmetric

80   incompatibility, which can influence patterns of introgression following

81   hybridization (Fishman *et al.*, 2001; Turelli and Moyle, 2007; Tang *et al.*,

82   2010). Genetic incompatibilities of all kinds accumulate with increasing

83   phylogenetic distance until complete reproductive isolation is evident

84   (Matute *et al.*, 2010; Moyle and Nakazato, 2010; Levin, 2012; Corbett-

85   Detig *et al.*, 2013).

86   Interspecific crossing, when used in linkage analysis and mapping,

87   may also reveal chromosomal rearrangements between species that can

88   reduce the fitness of hybrids and suppress recombination that, in turn, will

89   affect rates of interspecific gene flow (Rieseberg, 2001; Ortiz-Barrientos *et*

90   *al.*, 2002; Kirkpatrick and Barton, 2006). Studies of hybridizing annual

91   sunflowers have revealed that differences in genetic architecture due to

92   chromosomal inversions and translocations are an important cause of hybrid

93   sterility (Lai *et al.*, 2005; Yatabe *et al.*, 2007), whereas in other groups of

94   hybridizing species such as irises (Taylor *et al.*, 2012) and sculpin fish

95   (Stemshorn *et al.*, 2011) such rearrangements appear to be absent or minor.

96   In contrast, the presence of TRDLs in genetic maps of interspecific crosses

97   appears to be common, if not the rule (Fishman *et al.*, 2001; Lu *et al.*, 2002;

98   Tang *et al.*, 2010). Either way, the cumulative action of multiple intrinsic

99   incompatibilities with diverse modes of action distributed across the

100 genome will act as a potent barrier to gene flow between species and seems

101   to be a common, if not universal, intermediate stage in the process of

102   speciation (Bierne *et al.*, 2011; Feder *et al.*, 2012; Abbott *et al*., 2013;

103   Bomblies, 2013).

104   In the present study, we utilise interspecific crossing and linkage

105   analysis to investigate the genetic nature of intrinsic incompatibility

106   between two diploid, ragwort species, *Senecio aethnensis* (2n = 20) and *S.*

107   *chrysanthemifolius* (2n = 20) (Asteraceae), that form a hybrid zone on

108   Mount Etna, Sicily (James and Abbott, 2005; Brennan *et al*., 2009). These

109   two species are self-incompatible, short-lived, herbaceous perennials, which

110   grow at high and low altitudes, respectively, on Mount Etna. They are

111   connected by a series of hybrid populations, which potentially provides a

112   corridor for high levels of interspecific gene flow to occur. Interestingly,

113   material collected from this hybrid zone and introduced to Britain in the late

114   17th century subsequently gave rise to a highly invasive homoploid hybrid

115   species, *S. squalidus*, which spread through much of Britain in the 19th and

116   20th centuries (James and Abbott, 2005; Abbott *et al.*, 2009).

117   Previous analyses of the hybrid zone on Mount Etna showed that

118   while extrinsic environmental selection is important in determining the

119   ecological differences and relative distributions of the two *Senecio* species

120   (Brennan *et al.*, 2009; Ross *et al.*, 2012; Muir *et al.*, 2013; Osborne *et al.*,

121   2013; Chapman *et al*., 2013), intrinsic selection against hybrids was

122   predominant in determining changes in allele frequencies and quantitative

123   trait expression in the hybrid zone (Brennan *et al.*, 2009). Indeed, based on

124   indirect measures of assessment, Brennan *et al.* (2009) showed that the

125   hybrid zone was characterized by strong selection against hybrids, high

126 dispersal rates, and few loci differentiating quantitative traits. The evidence

127 for strong selection against hybrids was somewhat surprising as both

128 Hegarty *et al.* (2009) and Brennan *et al.* (2013) have reported that fertile

129 hybrids are easily produced from crosses between the two species, while

130 Chapman *et al.* (2005) showed that *S. aethnensis* exhibits no conspecific

131 pollen advantage when pollinated with mixtures of pollen from both

132 species, whereas *S. chrysanthemifolius* shows only a small conspecific

133 pollen advantage when treated similarly. However, Hegarty *et al.* (2009)

134 noted that a marked decline in germination rate and survival occurred in the

135 F_2 generation of crosses they examined, indicating that post-mating

136 incompatibility between the two species becomes apparent after the F_1

137 generation, i.e. as a consequence of hybrid breakdown.

138   Clearly, the two *Senecio* species and the hybrid zone they form on

139 Mount Etna comprise a very useful system for investigating adaptive

140 divergence and mechanisms of reproductive isolation between hybridizing,

141 diploid species. In the present study we investigate further the nature of

142 intrinsic reproductive isolation between *S. aethnensis* and *S.*

143 *chrysanthemifolius* by examining genetic segregation in an F_2 population

144 derived from a reciprocal cross between the two species. This enabled us to

145 (i) construct a genetic linkage map based on segregation in this F_2

146 population, (ii) identify if large-scale linkage group (chromosomal)

147 rearrangements exist between the two species, (iii) determine the occurrence

148 and extent of transmission ratio distortion across the linkage groups

149 identified, and (iv) examine some of the possible causes of transmission

150 ratio distortion at particular loci. Overall, the aim of our study was to obtain

151   a better understanding of the occurrence and genetics of intrinsic

152   incompatibility between these two species.

153

154   **MATERIALS AND METHODS**

155   **F_2 population**

156   Parent plants were raised from seed collected from wild populations located

157   at approximately 2600 m (*S. aethnensis* population VB) and 600 m (*S.*

158   *chrysanthemifolius* population C1) altitude, respectively, on Mount Etna,

159   Sicily (see James and Abbott, 2005, for further details of populations).

160   Reciprocal crosses (after Brennan *et al.*, 2013) were made between one

161   representative of each of *S. aethnensis* (A) and *S. chrysanthemifolius* (C) to

162   produce 16 F_1 individuals. The F_1s were grown to flowering and inter-

163   crossed in a partial diallel design. The most compatible reciprocal cross

164   between a pair of F_1s, i.e., producing highest seed-set in both directions, was

165   chosen to found an F_2 reciprocal cross family (hereafter referred to as

166   F_2AC). The maternal and paternal origin of each F_2 individual and its F_1

167   progenitors were recorded to permit testing of cytoplasmic effects. A total

168   of 120 F_2AC seed was sown, half coming from each direction of a cross

169   between the selected pair of F_1 parents. Following germination, F_2AC

170   individuals were grown, one per pot, in a glasshouse. Pots were randomized

171   on a bench and re-randomized weekly until commencement of flowering.

172   Twenty F_2AC individuals (16.7%) failed to flower because they either failed

173   to germinate, experienced early mortality, or exhibited stunted growth and

174   remained vegetative during the course of the study. As plants comprising

175   this F_2 family were also to be used in another study aimed at examining the

176 quantitative genetics of morphological and life history differences between

177 *S. aethnensis* and *S. chrysanthemifolius*, non-flowering individuals were

178 excluded from further analysis. This left 100 F_2 plants for genotyping and

179 use in the construction of linkage maps.

180

181 **DNA isolation and genotyping**

182 DNA was extracted from young fresh leaves of all plants, i.e. the two

183 parents, 16 F_1 progeny, 100 F_2AC progeny (after Brennan *et al.*, 2009). For

184 a subsample of 9.4% of randomly chosen plants (41 out of 436 plants from a

185 larger genetic study), two independent DNA extracts were made to test for

186 genotyping reliability.

187

188 **AFLP genotyping**

189 A protocol modified from that described by Wolf, PG

190 (http://archive.is/LCe6) was used to generate amplified fragment length

191 polymorphisms (AFLPs; supplementary methods). Fluorescently labelled

192 forward primers were multiplexed in the final selective PCR step by adding

193 0.05 µl of each of two labelled primers per sample and adjusting the total 20

194 µl PCR volume accordingly.

195   The AFLP protocol was initially tested on a panel of individuals

196 comprising the two parents and their F_1 progeny. AFLPs were detected by

197 running samples on a Beckman Coulter CEQ 8000 capillary sequencer (see

198 Brennan *et al.*, 2009) and analyzing output with CEQ v9.0 (Beckman

199 Coulter Inc. Fullerton, CA). In total, 28 paired combinations of five

200 selective forward and eight selective reverse AFLP primers with different

201   combinations of three final 5' bases were tested on these individuals.

202   Samples were genotyped first using the automatic AFLP binning options of

203   CEQ v9.0 with a minimum relative fluorescence unit (rfu) cut-off of 100 rfu

204   and bin widths of 1 bp. Manual checks of the AFLP genotypes of parents,

205   repeat extracts, and F_1 plants were then conducted to assess what subset of

206   AFLP bands were suitable for genotyping F_2AC individuals based on

207   reliable amplification and inheritance pattern. Eight of the 28 primer

208   combinations (E1M3, E1M5, E1M7, E4M7, E5M3, E5M6, E8M5, E8M7)

209   were chosen for genotyping F_2AC plants because they produced many

210   reliably scored amplified polymorphic bands showing high levels of

211   repeatability across replicate extracts (>95% across all scored loci) and the

212   expected inheritance pattern from parents to F_1s, and also because they

213   could be used in multiplexed combinations (Table S2a). AFLP bands were

214   scored in F_2AC individuals from the automatic AFLP CEQ v9.0 binning

215   output with reference to parent controls and then checked manually across

216   all samples to correct for genotyping errors.

217

218   **SSR, EST and INDEL genotypes**

219   A screen of 340 existing and newly developed single locus molecular

220   genetic markers was initially carried out on the two parent individuals.

221   These loci comprised 37 simple sequence repeat loci (SSRs) developed

222   previously by repeat motif enrichment of genomic DNA of *S. aethnensis, S.*

223   *chrysanthemifolius, S. squalidus*, and *S. vulgaris* (Liu *et al.*, 2004), 25 newly

224   developed SSRs (A Brennan and G-Q Liu unpublished), 10 SSRs developed

225   for *S. madagascariensis* (Le Roux and Wieczorek, 2006), 216 newly

226 developed expressed sequence tag (EST)-SSRs and 45 EST-indels (SSR

227 markers labelled as ES and NES, indels labelled as EC; Hegarty *et al.*, 2008;

228 www.seneciodb.org/), and 7 additional indel markers comprising five

229 derived from non-duplicated Asteraceae genes (Chapman *et al.*, 2007) and

230 two from published polymorphic *Senecio* gene sequences (*SSP*, which

231 encodes a stigma-specific peroxidase involved in pollination (McInnis *et al.*,

232 2005), and *Ray2a*, which encodes a cycloidea-like transcription factor

233 involved in the control of ray floret development in *Senecio* (Kim *et al.*,

234 2008)). A single PCR protocol was used to amplify all markers (see

235 Brennan *et al.*, 2009). This PCR protocol consisted of a three-primer system

236 with universal fluorescently labelled M13 primers. Mulitplexed PCR

237 products of all F_2 plants, wild sampled plants, control repeats, and both

238 parents were run on a Beckman Coulter CEQ 8000 capillary sequencer and

239 genotypes were assessed and scored using CEQ v9.0. Fifty four of these

240 marker loci were appropriate for genetic mapping in the current study based

241 on reliable PCR amplification, fragment length scoring and the presence of

242 fragment length polymorphisms in the F_2AC family (Tables 1 and S1).

243

244 **Genetic mapping**

245 A genetic linkage map was constructed from the segregation of alleles at

246 marker loci in the F_2AC family using the demonstration version of Joinmap

247 v4.0 (Van Ooijen, 2001). In the analysis, the F_2AC was treated as an

248 outcrossed mapping family (CP type) because many AFLP and codominant

249 marker genotypes were heterozygous in the F_0 parents (*S. aethenensis* and *S.*

250 *chrysanthemifolius*) and therefore did not exhibit the type of F_2 inheritance

251   that assumes parents are homozygous for alternative alleles at polymorphic

252   loci. The outcross mapping family option (CP type) in Joinmap v4.0 allows

253   mapping of loci with a variety of parental genotypes showing different

254   segregation patterns. The software automatically assigns the most likely

255   linkage phase of heterozygous parental alleles to each of the two parental

256   chromosomes for each group of linked loci, thus enabling more loci to be

257   mapped. A genetic map was assembled using Joinmap's default regression

258   mapping algorithm parameters. Linkage groups were identified at greater

259   than four logarithm of odds (LOD) score with less than 20 Kosambi

260   centiMorgan (cM) map distance units between loci. Map quality was

261   assessed by examining goodness of fit G^2 statistics and markers responsible

262   for incompatible linkage interactions were removed to generate linkage

263   groups with high map support. In some cases, removal of suspect loci led to

264   the splitting of large linkage groups. Diagrams of linkage groups were

265   constructed using MapChart v2.2 (Voorrips, 2002).

266   Summary statistics describing map characteristics were calculated as

267   follows. Genome length was estimated by adding twice the mean marker

268   distance to the length of each linkage group to account for ends beyond the

269   terminal markers (Fishman *et al.*, 2001) and also by multiplying the length

270   of each linkage group by the correction factor (marker number +1)/(marker

271   number -1) (Chakravarti *et al.*, 1991). Map coverage in terms of the

272   percentage of the genome that is within 5 or 10 cM of a mapped marker was

273   assessed according to the formula: 1-exponent((-2*distance*marker

274   number)/map length)) (Fishman *et al.*, 2001). The extent of marker

275   clustering was tested using a chi^2 dispersion test against a null Poisson

276   distribution of evenly distributed markers separated by mean marker

277   distance.

278

279   **Transmission ratio distortion**

280   Genotype frequencies at each mapped marker locus in the entire

281   F_2AC mapping family were tested for Mendelian segregation of genotypes

282   using chi^2 with Microsoft Excel 2003 (Microsoft corp, 2003). Associations

283   between chi^2 statistics for genotype segregation and marker characteristics,

284   including marker type (classified as dominant or codominant) and AFLP

285   fragment length, were tested with linear models after natural log

286   transformation using R v2.13 (R Development Core Team 2011).

287   The distribution of loci showing TRD, suggestive of the number of

288   independent TRDLs, was examined by plotting per locus chi^2 values onto

289   their genetic map positions. The number of distinct TRDLs across the

290   genetic map was assessed as the number of clusters of distorted markers at a

291   single-locus 95% confidence level. Clustered distorted markers within each

292   TRDL were located at a map distance of less than 10 cM from each other.

293   The most likely map position of a TRDL was interpreted as the map

294   position of the locus exhibiting the greatest TRD within each cluster of

295   distorted markers. Some markers showing TRD were isolated by more than

296   10 cM from the nearest marker also exhibiting TRD. These markers were

297   considered as possibly representing different TRDLs, albeit with weaker

298   supporting evidence. Clustering of distorted markers across the entire

299   genetic map was tested with a binomial test of the hypothesis that the

300   observed number of neighbouring pairs of distorted loci per chromosome

301   was greater than the expected number of pairs of such loci given their

302   observed frequency. A Poisson test of TRDL clustering was not conducted,

303   since the underlying marker distribution was significantly clustered

304   according to a dispersion test. A possible bias of TRDLs being located at the

305   ends of linkage groups, due possibly to weak linkage to all other mapped

306   markers, was investigated using a binomial test of whether loci at linkage

307   group ends more often exhibited segregation distortion relative to the

308   overall likelihood of loci showing TRD.

309

310   Causes of transmission ratio distortion

311   *Post-zygotic mortality or failure to flower.* Chi^2 tests were performed on the

312   F_2AC family of segregating offspring to test the extent to which selective

313   post-zygotic mortality or failure to flower (exhibited by 16.7% of F_2

314   individuals examined) could be responsible for generating the observed

315   TRD. For these tests, expected genotype and allele segregation patterns

316   were based on the minimum frequencies expected assuming that all non-

317   genotyped F_2 individuals possessed the under-represented allele or

318   genotype.

319

320   *Pre-zygotic events.* TRD caused by pre-zygotic events, such as biased

321   gamete production (meiotic drive), gametophyte selection or other maternal

322   effects (Turelli and Moyle, 2007; Fishman *et al.*, 2008), was tested at 36

323   codominant loci where the parental origin of alleles could be identified

324   unambiguously. These early reproductive events occur at the haploid stage

325   of the life cycle and require tests on allelic rather than genotype segregation.

326   Because such modes of incompatibility may show unilateral effects and

327   therefore be visible in only one cross direction, allelic segregation tests were

328   also performed on subsets of F_2 individuals containing different parental

329   cytoplasms.

330

331   *Cytonuclear incompatibility.* Because the F_2AC family was generated from

332   a reciprocal cross, it was possible to examine whether TRD of genotypes at

333   particular loci resulted from cytonuclear incompatibility by testing if TRD

334   was dependent on the direction of the cross (Fishman and Willis, 2006;

335   Turelli and Moyle, 2007). Thus, we tested TRD of genotypes in separate

336   subsets of 49 and 51 individuals in the F_2AC mapping family that had

337   inherited either the *S. aethnensis* or *S. chrysanthemifolius* cytoplasm,

338   respectively (Figures S2 and S4).

339

340   *BDM incompatibilities.* Hybrid incompatibilities are often thought to be due

341   to epistatic BDM incompatibilities between parental alleles at two or more

342   interacting genetic loci. To test for this, interactions between all pairs of

343   identified TRDLs were examined using Fisher's exact tests of contingency

344   tables of paired genotype counts using R v2.13 (R Development Core Team

345   2011).

346

347   *Deficiency or excess of heterozygotes.* Tests of whether TRD was caused by

348   either a deficiency or excess of heterozygous genotypes were conducted at

349   36 codominant loci where the heterozygous state of progeny in terms of

350   parental alleles could be identified. A deficiency of heterozygous genotypes

351 might reflect negative allelic interactions at a locus (underdominance). In

352 contrast, an excess of heterozygous genotypes could reflect inbreeding

353 depression (due to the expression of deleterious recessives) as a

354 consequence of the full-sib cross used to generate the F_2 family (Remington

355 and O´Malley, 2000; Schwarz-Sommer *et al.*, 2003), heterosis exhibited by

356 heterozygous genotypes generated from parents homozygous for different

357 alleles (Latta *et al.*, 2007), or the occurrence of recessive BDM

358 incompatibilities across two or more loci (Fishman *et al.*, 2008).

359

360 *Locally reduced recombination.* Genomic regions exhibiting limited local

361 recombination might show enhanced associations with TRD because of the

362 effects of chromosomal rearrangements on chromosomal segregation during

363 meiosis, or because they are more likely by chance to show linkage to

364 nearby loci under selection in hybrids (Fishman *et al.*, 2013). We

365 investigated this by testing for significant associations between local

366 recombination rate and TRD using logistic regression of markers scored for

367 various categories of TRD against marker distance. Different TRD

368 categories were chosen to reflect different stringencies in defining TRDLs

369 and TRDL map locations. These included categories where markers

370 exhibited TRD detected by chi^2 tests at (i) $p = 0.05$ or (ii) $p = 0.001$, (iii)

371 where markers exhibited greatest TRD within each of the nine identified

372 TRDLs, and (iv) where markers showed greatest TRD within each of the

373 four multi-locus TRDLs.

374

375 **RESULTS**

376 **Molecular genetic markers**

377 Details of the AFLP, and codominant markers selected to genotype *S.*

378 *aethnensis* and *S. chrysanthemifolius* parents, and their F_1 and F_2 offspring

379 for construction of linkage maps are presented in Tables 1 and S1.

380

381 **Genetic linkage map**

382 A genetic linkage map constructed from the segregation of markers in the

383 F_2AC mapping family comprised 127 marker loci distributed across 14

384 linkage groups of average 22.4 cM length (st. dev. 15.6 cM) giving a total

385 genetic map length of 313.8 cM (Figure 1, Table 2). Markers showing

386 transmission ratio distortion (TRD) were included in the construction of this

387 map because such distortion has been shown not to bias recombination

388 statistics nor the resulting linkage maps (Xu, 2008). Nine markers were not

389 mapped because they were unlinked at the <4 LOD and >20 cM thresholds

390 for inclusion in linkage groups (Table 1). Another nine markers were

391 excluded from the map because they caused problematic linkage

392 interactions within linkage groups (Table 1) or, in the case of two of these

393 markers, showed linkage to multiple different linkage groups (EC77 linked

394 AC1 and AC5A, and EC1687 linked AC5A, AC5B, and AC6). Weak

395 linkage (<4 LOD or >20 cM distances) between markers located on four

396 pairs of linkage groups was attributed to them belonging to distant ends of

397 the same chromosome. Thus, ten independent linkage groups were inferred,

398 which matches the haploid chromosome number ($n = 10$) of both species

399 (Alexander, 1977). Distance between mapped markers averaged 2.8 cM, but

400 was highly variable (st. dev. = 3.6 cM). The dispersion index for markers

401   was large (4.66) and highly significant (dispersion test, $p = 5.85e-46$), which

402   indicated a clumped distribution of markers across the map. The total

403   predicted map length was estimated to be 391.6 cM (after Fishman *et al.*,

404   2001) or 407.1 cM (using the method of Chakravarti *et al.*, 1991) and

405   approximately 96% and >99% map coverage of the genome was predicted

406   to be within 5 and 10 cM of a mapped marker, respectively.

407

408   **Transmission ratio distortion**

409   Transmission ratio distortion (TRD) was common for markers included in

410   the F_2AC genetic map with 34.0% of codominant markers and 22.1% of

411   dominant markers showing significant deviations from Mendelian

412   expectations of genotype segregation according to a per-locus 95%

413   confidence level (26.8% overall percentage frequency; Figure 1 and Table

414   S2). Mapped codominant markers exhibited TRD more frequently than

415   mapped dominant markers ($F_{1,125} = 10.41$, $p = 0.0016$), probably reflecting

416   the fact that AFLPs, but not codominant markers, were included on the basis

417   of inheritance patterns in the F_1 progeny panel. No association was found

418   between AFLP fragment length and TRD ($F_{1,75} = 0.67$ $p = 0.4173$).

419   The map locations of markers were investigated to better understand

420   the genetic architecture of incompatibilities between the two parent species,

421   *S. aethnensis* and *S. chrysanthemifolius*. Markers showing genotypic

422   segregation distortion were clustered in the genetic map (one-way binomial

423   tests $p = 0.0007$ and $3.5e-6$, markers showing TRD at individual chi^2 test 95

424   % and 99.9 % confidence levels, respectively) to form four distinct groups

425   located in linkage groups AC1, AC3 (distal position), AC7A, and AC10A,

426   and also occurred individually in linkage groups AC3 (proximal and central

427   widely separated positions), AC4, AC6, and AC9 (Figure 1 and Table 3).

428   The widespread distribution of clusters and individual markers showing

429   TRD across several linkage groups indicates that multiple loci contributed

430   to the frequent TRD observed in the F_2 of this cross (Figures 1 and S1,

431   Tables 3 and S2). Within each cluster, all markers showed the same TRD

432   direction favouring either alleles from the same parent or heterozygous or

433   homozygous parental genotypes (Figures S1 to S4). Markers located at the

434   ends of six linkage groups exhibited significant TRD, which raised the

435   possibility that such markers that are weakly linked to other mapped

436   markers might be erroneously interpreted as distinct TRDLs. However, a

437   binomial test for overrepresentation of linkage group ends exhibiting TRD

438   (6 out of 28) relative to the overall frequency of markers showing

439   segregation distortion (26.4%) was not significant ($p = 0.788$). We conclude

440   that at least four TRDLs influencing multiple markers were present in the

441   mapping family or as many as nine TRDLs if additional unlinked markers

442   showing TRD were also included.

443

444   **Possible causes of transmission ratio distortion**

445   *Post-zygotic mortality or failure to flower*. The observed TRD could have

446   been due to selective mortality or selective failure to flower of some

447   individuals in the mapping family (16.7% of F_2 progeny). Tests for TRD

448   greater than could be accounted for by these causes showed that seven of the

449   nine TRDLs (excluding the proximal and distal TRDLs on AC3) required

450 additional mechanisms to explain the observed TRD in these regions

451 (Tables 3 and S2, Figure S1).

452

453 *Pre-zygotic events*. Tests of TRD caused by events at early stages of

454 reproduction, such as meiotic drive and gametophytic selection, were

455 conducted on codominant loci where the parental origin of alleles could be

456 identified. These tests showed that for TRDLs in linkage groups AC1 and

457 AC10A, and also for a new TRDL in AC10B, F_2 progeny lacked *S.*

458 *aethnensis* alleles, while for the central TRDL in linkage group AC3, F_2

459 individuals lacked *S. chrysanthemifolius* alleles (Tables 3 and S2, Figure

460 S3). When cross direction was also tested for, TRDLs in AC5B and

461 AC10A were shown to express TRD in the *S. aethnensis* maternal

462 background only (Figure S4 and Table S2).

463

464 *Cytonuclear incompatibility*. Tests of TRD of genotypes in subsets of F_2

465 individuals that reflected differences in cross direction showed an overall

466 similar genomic distribution of TRD irrespective of parental cytoplasm

467 (Figures S2 and S4). However, the TRD of several TRDLs was cytoplasm

468 dependent. This was the case for TRDLs located in linkage groups AC3

469 (central and distal positions) and AC9 (all dependent on *S.*

470 *chrysanthemifolius* cytoplasm), and also the TRDL located in AC10A

471 (dependent on *S. aethnensis* cytoplasm) (Tables 3 and S2, Figure S2). A

472 further instance of TRD in linkage group AC5B was detected only in

473 progeny possessing *S. aethnensis* cytoplasm (Table 3, Figure S2). Taken

474 overall, these tests found asymmetric TRD occurring at five TRDLs

475   dependent on individuals possessing either *S. chrysanthemifolius* or *S.*

476   *aethnensis* cytoplasm.

477

478   *Epistatic BDM incompatibilities*. Multi-locus BDM incompatibilities were

479   investigated by looking for negative interactions between particular

480   genotype combinations of pairs of markers close to the nine TRDLs

481   identified. Of the 36 paired TRDL combinations tested, five significantly

482   non-independent pairs were found that affected four of the nine TRDLs (all

483   three TRDLs on AC3 and one on AC6, Table 3). However, two of the five

484   interacting TRDLs were present in the same linkage group indicating that

485   physical proximity might contribute to their non-independence. The

486   underrepresented genotype combinations typically included one of the loci

487   homozygous for *S. aethnensis* alleles and the other locus homozygous for *S.*

488   *chrysanthemifolius* alleles.

489

490   *Deficiency or excess of heterozygotes.* Tests of a deficiency or excess of

491   heterozygotes in terms of parental alleles at codominant loci where the

492   parental state of alleles could be identified, detected a significant deficiency

493   of heterozygous genotypes at the TRDL of large effect located in linkage

494   group AC1 (Tables 3 and S2, Figure S1). However, this deficiency is more

495   likely caused by selection against *S. aethnensis* alleles, than

496   underdominance (heterozygote disadvantage), because there was an even

497   stronger bias against genotypes homozygous for *S. aethnensis* alleles across

498   this particular region of AC1 (Figure S3). No TRDLs showing an excess of

499   heterozygotes relative to homozygotes were identified, thus discounting

500   inbreeding depression, heterosis, or recessive BDM incompatibilities, as

501   important causes of TRD in this F_2 family.

502

503   *Locally reduced recombination.* In no instance was an association detected

504   between locally reduced recombination and TRD for any of the four

505   categories of markers tested. Thus, for markers that showed TRD at 95%

506   and 99.9% confidence levels, the probability of a stronger association with

507   recombination rate than the null hypothesis of no association was $p = 0.300$,

508   and $p = 0.291$, respectively, while the probability was $p = 0.261$ for the most

509   strongly distorted marker within each of the nine identified TRDLs, and $p =$

510   0.178 for the most strongly distorted marker within each of the four multi-

511   locus TRDLs treated separately. Thus, these results do not support the

512   hypothesis that chromosomal regions with reduced recombination are

513   associated with hybrid incompatibility in this system.

514

515   **DISCUSSION**

516   Studies of closely related species that form hybrid zones can reveal much

517   about the genetic basis and maintenance of species divergence in the face of

518   frequent interspecific hybridization and gene flow. Our previous analysis of

519   the hybrid zone between *S. aethnensis* (occurring at high altitude) and *S.*

520   *chrysanthemifolius* (occurring at low altitude) on Mount Etna, Sicily,

521   compared clines for molecular variation with those for phenotypic trait

522   variation, and indicated that both extrinsic and intrinsic selection against

523   hybrids act to maintain the hybrid zone despite high levels of gene flow

524   (Brennan *et al.*, 2009). The study reported here has expanded on these

525   previous results and provided insights into the genetics of intrinsic

526   reproductive isolation between *S. aethnensis* and *S. chrysanthemifolius* on

527   Mount Etna, Sicily. We found intrinsic genomic incompatibility between

528   these two species caused by a variety of genetic mechanisms at multiple

529   genetic loci. However, large-scale genomic rearrangements or translocations

530   between the species did not appear to contribute greatly to this

531   incompatibility.

532

533   **F_2 linkage map structure**

534   To investigate the genomic architecture of these hybridizing *Senecio*

535   species, we constructed a genetic linkage map from the segregation of

536   dominant and codominant molecular markers in the F_2 of a reciprocal cross

537   between the two species. The resulting F_2AC linkage map comprised ten

538   distinct linkage groups (taking account of four weakly linked linkage group

539   pairs), which corresponds to the haploid chromosome number of the two

540   species (Alexander, 1979). If the parental species were distinguished by

541   chromosomal translocations, the affected regions would link different

542   linkage groups in genetic maps of hybrids and reduce the number of

543   independent linkage groups to below the haploid chromosome number of

544   the species investigated (Fishman *et al.*, 2013). Therefore, the ability to

545   distinguish ten distinct linkage groups in the F_2AC map and the removal

546   from maps of only two markers with suspect linkage to multiple different

547   linkage groups indicates that large-scale genomic translocations between

548   chromosomes probably do not distinguish the two species in contrast to

549   what has been found in hybridizing annual sunflowers (Lai *et al.,* 2005;

550   Yatabe *et al.*, 2007).

551   Overall, the map showed good coverage with an average locus

552   distance of just 2.8 cM and >99% of the genome predicted to be within 10

553   cM of a mapped marker. However, markers exhibited a highly clumped

554   distribution within the map, for which there are several possible

555   explanations. First, there may be technical reasons for marker clustering,

556   such as marker position uncertainty due to a limited number of

557   recombination events observed in the relatively small F_2 family examined,

558   and/or to genotyping error, which though estimated to be reasonably low

559   (2.2% based on duplicated samples tested) translates to a 2.2 cM uncertainty

560   in the position of loci within maps. Secondly, some marker clusters could be

561   explained by sequence heterogeneity across the genome. For example,

562   clusters of AFLP loci could signal repetitive genomic regions containing

563   many closely spaced repeated restriction enzyme cut sites, while clusters of

564   EST and gene loci could signal highly expressed, gene-rich genomic

565   regions. Thirdly, the occurrence of clusters could reflect variable

566   recombination rates across the genome caused by intrinsic features such as

567   low recombination near centromeres or in regions where local chromosome

568   rearrangements, such as inversions, exist between species.

569

570   **Transmission ratio distortion**

571   A notable feature to emerge from the current study was the large number of

572   molecular marker loci that showed transmission ratio distortion (TRD) in

573   the F_2AC mapping population, i.e. 34 (26.8%) of 127 markers tested. These

574 loci were non-randomly distributed across the genetic map and generally

575 formed clusters in which all loci showed the same bias against alleles from

576 one parent or against heterozygous combinations of parental alleles. We

577 found strong multi-locus evidence for the occurrence of four TRDLs and

578 weaker single locus evidence for five additional TRDLs located in seven of

579 the ten linkage groups identified. Individual markers showing TRD were

580 sometimes located at the ends of linkage groups raising the possibility that

581 they were technical artifacts; however, such markers were not

582 overrepresented relative to the overall observed numbers of markers

583 exhibiting TRD.

584   One TRDL, located at the proximal end of linkage group AC1,

585 showed particularly strong segregation distortion, both in terms of the extent

586 to which allele and genotypes frequencies were distorted away from

587 Mendelian expectations for *S. aethnensis* alleles and genotypes, and the

588 length of the genome affected (all of AC1, i.e. 44.5 cM). The fact that this

589 TRDL, and also other multilocus TRDLs, affected large genomic regions

590 was likely due to both their strong effects on segregation distortion and the

591 limited post-hybridization recombination that had occurred in the F_2AC

592 mapping family. Strong TRD in early-generation hybrids is likely to bias

593 patterns of introgression across large genomic regions in later generation

594 hybrids because hybrid genotypes at linked loci are eliminated before they

595 have an opportunity to recombine away from the TRDL (Bierne *et al.*,

596 2011).

597

598 **Causes of transmission ratio distortion**

599   Through further analyses of the patterns of segregation in the F_2AC family

600   we obtained some insights into the causes of TRD at different TRDLs.

601   Twenty of the original 120 progeny that comprised the F_2AC mapping

602   family failed to flower either because of a failure to germinate, the

603   occurrence of early mortality, or an inability to develop to the flowering

604   stage. These individuals were not genotyped and their absence may have

605   contributed to the TRD observed at marker loci. The low intrinsic fitness of

606   these individuals might stem from several causes (see below), including the

607   disruption of coadapted gene complexes (hybrid breakdown) following

608   recombination. We did not examine survival and flowering in the F_1

609   generation in our study, however a previous study (Hegarty et al., 2009)

610   recorded a marked drop in intrinsic fitness (measured in terms of seed

611   germination and seedling survival under glasshouse conditions) between the

612   F_1 and F_3 generations of a reciprocal cross between the same $S.$ $aethnensis$

613   and $S.$ $chrysanthemifolius$ parental individuals as this study, indicating the

614   occurrence of hybrid breakdown.

615   Although post-zygotic selective mortality or inability to flower of

616   some F_2 individuals would contribute to TRD at certain marker loci, we

617   found that the level of TRD exhibited by seven of the nine TRDLs

618   identified could not be entirely explained in this way. However, it remains

619   possible that earlier acting post-zygotic incompatibility in the form of a

620   failure of seed development (e.g. due to negative interactions between

621   zygote and endosperm) could have contributed to the observed TRD.

622   Alternatively, there may be pre-zygotic causes of TRD either in the

623   production of gametes containing particular alleles (meiotic drive) or

624 selection against gametes containing particular alleles (gametophytic

625 selection) (Fishman and Willis, 2005; Fishman *et al.*, 2008). We tested for

626 these possibilities at loci where the parental origin of alleles could be

627 identified unambiguously allowing TRD of alleles rather than genotypes to

628 be examined and found three TRDLs where there was a bias against *S.*

629 *aethnensis* alleles and one TRDL where there was a bias against *S.*

630 *chrysanthemifolius* alleles. Thus, pre-zygotic factors of the type mentioned

631 above could have contributed to the TRD recorded at these loci.

632 Incompatibility between diverging genomes is often asymmetric

633 when the underlying causes of incompatibility include either pre-zygotic

634 haploid stages of the life-cycle, such as meiotic drive or pollen fitness, or

635 cytonuclear incompatibilities between nuclear and organelle genomes

636 (Levin, 2003; Fishman and Willis, 2006; Turelli and Moyle, 2007). Strong

637 crossing asymmetry or unilateral incompatibility is most likely observed

638 when few asymmetric incompatibilities of large effect are involved (Turelli

639 and Moyle, 2007). The extent to which TRD was asymmetric and dependent

640 on cross direction was tested by comparing TRD across loci in F_2 progeny

641 possessing cytoplasm inherited from either the *S. aethnensis* or *S.*

642 *chrysanthemifolius* parent. These tests revealed five TRDLs exhibiting

643 asymmetric differences in TRD dependent on parental cytoplasmic

644 background. We may conclude that the asymmetric TRD at these loci

645 reflects cytonuclear incompatibility and/or the effects of meiotic drive

646 and/or gametophytic selection.

647 There is growing evidence for the widespread occurrence of BDM

648 incompatibilities between allopatric or parapatric populations and their role

649   in limiting subsequent hybridization and possibly promoting further

650   divergence during speciation (Coyne and Orr, 2004; Corbett-Detig *et al.*,

651   2013). These BDM incompatibilities are frequently caused by deleterious

652   interactions between different parental alleles occurring at two or more loci.

653   Significantly non-independent, paired-locus genotype frequencies were

654   observed for five pairs of TRDLs affecting four of the nine TRDLs (Table

655   3). However, three of the five interacting TRDL pairs were present in the

656   same linkage group indicating that physical proximity might contribute to

657   their non-independence. The underrepresented genotype combination

658   typically included one of the loci homozygous for *S. aethnensis* alleles and

659   the other locus homozygous for *S. chrysanthemifolius* alleles.

660   We also tested if TRD could be due to (i) underdominance (reflected

661   by a deficiency of heterozygotes at certain loci) or (ii) inbreeding

662   depression, heterosis or recessive BDM incompatibilities (reflected by an

663   excess of heterozygotes). Inbreeding depression could be caused by the

664   expression of deleterious recessive alleles as a result of using a full-sib F_2

665   mapping family from two outcrossed self-incompatible parents. Tests of

666   heterozygosity revealed that underdominance could have contributed to a

667   significant deficiency of heterozygous genotypes at only one TRDL, in the

668   AC1 linkage group. However, the heterozygote deficiency at this locus was

669   more likely caused by selection against *S. aethnensis* alleles, rather than

670   underdominance, because there was an even stronger bias against genotypes

671   homozygous for *S. aethnensis* alleles across the AC1 TRDL (Figure S1). An

672   excess of heterozygotes was not evident at any TRDL and thus it is

673   concluded that inbreeding depression, heterosis or recessive BDM

674   incompatibilities have not contributed to the TRD recorded in the F_2AC

675   mapping family.

676   As mentioned earlier in the discussion, reduced recombination near

677   centromeres or due to local inversions may be causes of marker clustering

678   observed in the genetic map. These causes of reduced recombination may

679   also contribute to TRD at marker loci. For example, selfish drive elements

680   are typically found near centromeres where they directly influence

681   chromosome segregation patterns in their favour during meiosis (Henikoff

682   *et al.,* 2001), while chromosomal rearrangements may also influence

683   inheritance patterns because recombination within the rearranged region is

684   typically selected against (Ortiz-Barrientos *et al.,* 2002, Kirkpatrick and

685   Barton, 2006; Lowry and Willis, 2010). To examine whether genomic

686   regions showing limited recombination could be a cause of TRD in the

687   F_2AC family, we tested for, but did not find associations between, marker

688   clustering and TRDLs. Therefore, genomic regions showing limited

689   recombination do not seem to play a strong role in reinforcing genomic

690   divergence in *Senecio*. However, our current F_2 genetic map lacks sufficient

691   marker density to adequately test this association since the clumped marker

692   map distribution could be due to a variety of reasons other than variation in

693   recombination rates and the genomic scale of TRDL-low recombination

694   associations could be highly localized (Yatabe *et al.*, 2007; Jones *et al.*,

695   2012; Renaut *et al.*, 2013). Future studies aimed at determining whether

696   differences in chromosomal rearrangement may be a cause of genomic

697   incompatibility between *S. aethnensis* and *S. chrysanthemifolius* should

698   involve cytogenetic comparisons of karyotypes and/or detailed comparisons

699   of high marker density genetic maps of the two species.

700   Overall, it is clear from our analyses that a variety of genetic

701   mechanisms at multiple genetic loci across the genome are likely to

702   contribute to the intrinsic incompatibility existing between *S. aethnensis* and

703   *S. chrysanthemifolius*. Further fine-scale genetic mapping studies of *S.*

704   *aethnensis* and *S. chrysanthemifolius* involving quantitative trait locus

705   mapping of traits of adaptive relevance, larger families to better estimate

706   recombination, more markers for better genome coverage, and families from

707   intraspecific crosses, will be necessary to further investigate the various

708   mechanisms contributing to hybrid incompatibility and ecological

709   divergence in this species pair.

710

711   **Conclusions**

712   While a number of studies have indicated that *S. aethnensis* and *S.*

713   *chrysanthemifolius* are highly interfertile (Chapman *et al.*, 2005; Brennan *et*

714   *al.*, 2013), the present study has revealed evidence of intrinsic

715   incompatibility between these two species in the form of multiple genomic

716   regions showing TRD. Our results, therefore, support the findings of our

717   previous clinal analysis of the natural hybrid zone on Mount Etna, which

718   indicated that intrinsic selection against hybrids was an important factor

719   maintaining species differences in the face of gene flow (Brennan *et al.*,

720   2009). The present study shows that hybrid incompatibility between these

721   two diverging plant lineages is more cryptic than previously considered, and

722   is manifested in the F_2 generation rather than in the F_1 (see also Hegarty *et*

723 *al.*, 2009). Some studies of other diverging plant lineages have yielded

724 similar findings (e.g. Fishman *et al.*, 2001; Moyle and Graham, 2006;

725 Fishman and Willis, 2006)

726   It follows that the effects of TRDLs with multiple modes of action at

727 multiple, relatively large, genomic regions in early generation hybrids

728 between *S. aethnensis* and *S. chrysanthemifolius* are likely to impact the

729 genetic structure of the natural hybrid zone on Mount Etna by limiting

730 introgression and promoting divergence across the genome (Feder *et al.*,

731 2012; Abbott *et al.*, 2013). However, large-scale genomic translocations or

732 other rearrangements between the species do not seem to contribute in any

733 major way to this process. It has been estimated that *S. aethnensis* and *S.*

734 *chrysanthemifolius* are of relative recent origin (<1 million years ago,

735 Comes and Abbott, 2001; ~108,000 to 150,000 years ago, Osborne *et al.*,

736 2013, Chapman *et al.*, 2013), which indicates that the various forms of

737 intrinsic hybrid incompatibility that clearly exist between these two species

738 must have evolved relatively rapidly.

739

740 **DATA ARCHIVING**

741 Mapping family genotype data will be deposited with the DRYAD data

742 repository. Other results can be found in supplementary information.

743 Table S1. Summary of codominant molecular markers used for mapping.

744 Table S2. Frequencies and segregation tests of mapped markers.

745 Figure S1. Genetic map distribution of genotype TRD.

746 Figure S2. Genetic map distribution of genotype TRD for separate

747 cytotypes.

748   Figure S3. Genetic map distribution of allelic TRD.

749   Figure S4. Genetic map distribution of allelic TRD for separate cytotypes.

750

762

763   **REFERENCES**

764   Abbott RJ, Brennan AC, James JK, Forbes DF, Hegarty MJ, Hiscock SJ

765   (2009). Recent hybrid origin and invasion of the British Isles by a

766   self-incompatible species, Oxford ragwort (*Senecio squalidus* L.,

767   Asteraceae). *Biol. Invasions* **11**: 1145–1158.

768   Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N *et al.*

769   (2013). Hybridization and speciation. *J. Evol. Biol.* **26**: 229-246.

770   Agrawal AF, Feder JL, Nosil P (2011). Ecological divergence and the origin

771   of intrinsic postmating isolation with gene flow. *Int. J. Ecol.* 2011 ID

772   435357, doi:10.1155/2011/435357.

773   Alexander JCM (1979). Mediterranean species of *Senecio* sections *Senecio*

774   and *Delphinifolius*. *Notes Roy. Bot. Gard. Edin.* **37**: 387-428.

775   Bierne N, Welch J, Loire E, Bonhomme F, David P (2011). The coupling

776   hypothesis: why genome scans may fail to map local adaptation

777   genes. *Mol. Ecol.* **20**: 2044-2072.

778   Bomblies K (2013). Genes causing postzygotic hybrid incompatibility in

779   plants: a window into co-evolution. In: Chen ZJ, Birchler JA (eds)

780   *Polyploid and Hybrid Genomics,* John Wiley & Sons, Inc.: Oxford,

781   UK. pp 225-240.

782   Brennan AC, Harris SA, Hiscock SJ (2013). The population genetics of

783   sporophytic self-incompatibility in three hybridizing *Senecio*

784   (Asteraceae) species with contrasting population histories. *Evolution*

785   **67**: 1347-1367.

786   Brennan AC, Bridle JR, Wang A-L, Hiscock SJ, Abbott RJ (2009).

787   Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on

788   Mount Etna, Sicily. *New Phytol.* **183**: 702-717.

789   Chakravarti A, Lasher LK, Reefer JE (1991). A maximum likelihood

790   method for estimating genome length using genetic linkage data.

791   *Genetics* **128**: 175–182.

792   Chapman MA, Forbes DG, Abbott RJ (2005). Pollen competition among two

793   species of *Senecio* (Asteraceae) that form a hybrid zone on Mt. Etna,

794   Sicily. *Amer. J. Bot.* **92**: 730-735.

795   Chapman MA, Chang J, Weisman D, Kesseli RV, Burke JM (2007).

796   Universal markers for comparative mapping and phylogenetic

797   analysis in the Asteraceae (Compositae). *Theor. Appl. Genet.* **115**:

798   747-755.

799   Chapman MA, Hiscock SJ, Filatov DA (2013). Genomic divergence during

800   speciation driven by adaptation to altitude. *Mol. Biol. Evol.* **30**:

801   2553-2567.

802   Comes HP, Abbott RJ (2001). Molecular phylogeography, reticulation and

803   lineage sorting in the Mediterranean species complex of *Senecio* sect.

804   *Senecio* (Asteraceae). *Evolution* **55**: 1943-1962.

805   Corbett-Detig R, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic

806   incompatibilities are widespread within species. *Nature* **504**: 135-137.

807   Coyne JA, Orr HA (2004). *Speciation*. Sinauer Associates, Sunderland,

808   Mass, USA.

809   Cutter AD (2012). The polymorphic prelude to Bateson-Dobzhansky-Muller

810   incompatibilities. *Trends Ecol. Evol.* **27**: 209-218.

811   Feder JL, Egan SP, Nosil P (2012). The genomics of speciation with gene

812   flow. *Trends Genet.* **28**: 342–350.

813   Fishman L, Willis JH (2005). A novel meiotic drive locus almost

814   completely distorts segregation in *Mimulus* (monkeyflower) hybrids.

815   *Genetics* **169**: 347–353.

816   Fishman L, Willis JH (2006). A cytonuclear incompatibility causes anther

817   sterility in *Mimulus* hybrids. *Evolution* **60**: 1372–1381.

818   Fishman L, Aagaard J, Tuthill JC (2008). Toward the evolutionary

819   genomics of gametophytic divergence: patterns of transmission ratio

820   distortion in monkeyflower (*Mimulus*) hybrids reveal a complex

821   genetic basis for conspecific pollen precedence. *Evolution* **62**: 2958–

822   2970.

823   Fishman L, Kelly AJ, Morgan E, Willis JH (2001). A genetic map in the

824   *Mimulus guttatus* species complex reveals transmission ratio

825   distortion due to heterospecific interactions. *Genetics* **159**: 1701-

826   1716.

827   Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP (2013).

828   Chromosomal rearrangements and the genetics of reproductive

829   barriers in *Mimulus* (monkeyflowers). *Evolution* **69**: 2547-2560.

830   Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ

831   (2008). Changes to gene expression associated with hybrid

832   speciation in plants: Further insights from transcriptomic studies in

833   *Senecio. Phil. Trans. Royal Soc. B.* **363**: 3055-3069.

834   Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ

835   (2009). Extreme changes to gene expression associated with

836   homoploid hybrid speciation. *Mol. Ecol.* **18**: 877-889.

837   Henikoff S, Ahmad K, Malik HS (2001). Speciation and centromere

838   evolution. *Science* **294**: 2478-2480.

839   James JK, Abbott RJ (2005). Recent, allopatric, homoploid hybrid

840   speciation: the origin of Oxford ragwort, *Senecio squalidus*

841   (Asteraceae), in the British Isles from a hybrid zone on Mount Etna,

842   Sicily. *Evolution* **59**: 2533-2547.

843   Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J *et al.*

844   (2012). The genomic basis of adaptive evolution in threespine

845   sticklebacks. *Nature* **484**: 55–61.

846   Kim M, Cui ML, Cubas P, Gillies A, Lee K, Chapman MA *et al.* (2008).

847   Regulatory genes control a key morphological and ecological trait

848   transferred between species. *Science* **322**: 1116-1119.

849   Kirkpatrick M, Barton NH (2006). Chromosome inversions, local

850   adaptation and speciation. *Genetics* **173**: 419–434.

851   Lai Z, Nakazato T, Salmaso M, Burke JM, Tang S, Knapp SJ, Rieseberg LH

852   (2005). Extensive chromosomal repatterning and the evolution of

853   sterility barriers in hybrid sunflower species. *Genetics* **171**: 291–303.

854   Latta RG, Gardner KM, Johansen-Morris AD (2007). Hybridization,

855   recombination, and the genetic basis of fitness variation across

856   environments in *Avena barbata*. *Genetica* **129**: 167-177.

857   Le Roux JJ, Wieczorek AM (2006). Isolation and characterization of

858   polymorphic microsatellite markers from fireweed, *Senecio*

859   *madagascariensis* Poir. (Asteraceae). *Mol. Ecol. Notes* **7**: 327-329.

860   Levin DA (2012). The long wait for hybrid sterility in flowering plants. *New*

861   *Phytol.* **196**: 666-670.

862   Levin DA (2003). The cytoplasmic factor in plant speciation. *Syst. Bot.* **28**:

863   5–11.

864   Liu G-Q, Hegarty MJ, Edwards KJ, Hiscock SJ, Abbott RJ (2004). Isolation

865   and characterization of microsatellite loci in *Senecio. Mol. Ecol.*

866   *Notes* **4**: 611–614.

867   Lowry DB, Willis JH (2010). A widespread chromosomal inversion

868   polymorphism contributes to a major life-history transition, local

869   adaptation, and reproductive isolation. *PLoS Biol*. **8**: 1–14.

870   Lu H, Romero-Seveson J, Bernardo R (2002). Chromosomal regions

871   associated with segregation distortion in maize. *Theor. Appl. Genet.*

872   **105**: 622–628.

873   Matute DR, Butler IA, Turissini DA, Coyne JA (2010). A test of the

874   snowball theory for the rate of evolution of hybrid incompatibilities.

875   *Science* **329**:1518-1521.

876   McInnis SM, Costa LM, Gutiérrez-Marcos JF, Henderson CA, Hiscock SJ

877   (2005). Isolation and characterization of a polymorphic stigma-

878   specific class III peroxidase gene from *Senecio squalidus* L.

879   (Asteraceae). *Plant Mol. Biol.* **57**: 659–677.

880   Moyle LC, Graham EB (2006). Genome-wide associations between hybrid

881   sterility QTL and marker transmission ratio distortion. *Mol. Biol.*

882   *Evol.* **23**: 973-980.

883   Moyle LC, Nakazato T (2010). Hybrid incompatibility "snowballs" between

884   *Solanum* species. *Science* **329**: 1521-1523.

885   Muir G, Osborne OG, Sarasa J, Hiscock SJ, Filatov DA (2013). Recent

886   ecological selection on regulatory divergence is shaping clinal

887   variation in *Senecio* on Mount Etna. *Evolution* **67**: 3032-3042.

888   Nosil P (2012). *Ecological speciation*. Oxford Univ. Press, Oxford.

889   Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF (2002). Recombination

890   and the divergence of hybridizing species. *Genetica* **116**: 167-178.

891   Osborne OG, Batstone TE, Hiscock SJ, Filatov DA (2013). Rapid speciation

892   with gene flow following the formation of Mt. Etna. *Genome Biol.*

893   *Evol.* **5**: 1704-1715.

894 Ouyang Y, Zhang Q (2013). Understanding reproductive isolation based on

895   the rice model. *Ann. Rev. Plant Biol.* **64**: 11-35.

896 R Development Core Team (2011). *R: A language and environment for*

897   *statistical computing*. R Foundation for Statistical Computing:

898   Vienna.

899 Remington DL, O´Malley DM (2000). Whole genome characterization of

900   embryonic stage inbreeding depression in a selfed Loblolly pine

901   family. *Genetics* **155**: 337-348.

902 Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC *et al.* (2013).

903   Genomic islands of divergence are not affected by geography of

904   speciation in sunflowers. *Nature Comm.* **4**: 1827.

905   doi:10.1038/ncomms2833.

906 Rieseberg LH (2001). Chromosomal rearrangements and speciation. *Trends*

907   *Ecol. Evol.* **16**: 351–358.

908 Ross RIC, Ågren JA, Pannell JR (2012). Exogenous selection shapes

909   germination behaviour and seedling traits of populations at different

910   altitudes in a *Senecio* hybrid zone. *Ann. Bot.* **110**: 1439-1447.

911 Rundle HD, Nosil P (2005). Ecological speciation. *Ecol. Lett.* **8**: 336–352.

912 Schwarz-Sommer Z, de Andrade Silva E, Berndtgen R, Lönnig WE, Müller

913   A, Nindl I *et al.* (2003). A linkage map of an F_2 hybrid population of

914   *Antirrhinum majus* and *A. molle*. *Genetics* **163**: 699-710.

915 Stemshorn KC, Reed FA, Nolte AW, Tautz D (2011). Rapid formation of

916   distinct hybrid lineages after secondary contact of two fish species

917   (*Cottus* sp.). *Mol. Ecol.* **20**: 1475-1491.

918   Tang S, Okashah RA, Knapp SJ, Arnold ML, Martin NH (2010).

919   Transmission ratio distortion results in asymmetric introgression in

920   Louisiana *Iris*. *BMC Plant Biol.* 10: 48

921   Taylor SJ, Rojas LD, Ho SW, Martin NH (2012). Genomic collinearity and

922   the genetic architecture of floral differences between the homoploid

923   hybrid species *Iris nelsonii* and one of its progenitors, *Iris hexagona*.

924   *Heredity* **110**: 63-70.

925   Turelli M, Moyle LC (2007). Asymmetric postmating isolation: Darwin's

926   corollary to Haldane's rule. *Genetics* **176**: 1059–1088.

927   Van Ooijen JW (2001). JoinMap v3.0 Software for the calculation of

928   genetic linkage maps. Plant Res. Int., Wageningen, Netherlands.

929   http://www.joinmap.nl

930   Voorrips RE (2002). MapChart: Software for the graphical presentation of

931   linkage maps and QTLs. *J. Hered.* **93**: 77-78. www.biometris.nl

932   Xu S (2008). Quantitative trait locus mapping can benefit from segregation

933   distortion. *Genetics* **180**: 2201–2208.

934   Yatabe Y, Kane NC, Scotti-Saintaigne C, Reisberg LH (2007). Rampant

935   gene exchange across a strong reproductive barrier between the

936   annual sunflowers, *Helianthus annuus* and *H. petiolaris. Genetics*

937   **175**: 1883–1893.

938

Table 1. Summary of genetic markers screened and mapped in the *Senecio aethnensis* x *S. chrysanthemifolius* F$_2$ genetic mapping family.

Marker type	# screened	# developed	# genotyped	# mapped	unlinked	problematic
Codominant markers						
EST indel	45	15	10	8		EC77, EC1687
EST SSR	216	48	33	31	ES91	ES19
Indel	7	4	3	3		
SSR	72	9	8	8		
Total	340	76	54	50	1	3
Dominant markers						
E1M3, CAAC/ACAG	-	-	19	15	179, 219	
E1M5, CAAC/ACTA	-	-	15	10	168, 213	88, 160, 204
E1M7, CAAC/ACTG	-	-	13	12	275	
E4M7, CACT/ACTG	-	-	9	8	113	
E5M3, CACC/ACAG	-	-	8	7	302	
E5M6, CACC/ACTC	-	-	9	9		
E8M5, CAGG/ACTA	-	-	11	9	114	196
E8M7, CAGG/ACTG	-	-	7	7		

| Total | - | - | 91 | 77 | 8 | 6 |

Marker types are divided into codominant or dominant markers. Codominant markers are further divided into markers derived from expressed sequence tags (EST), simple sequence repeats (SSR) or insertion-deletions (indel). Dominant markers are amplified fragment length polymorphisms (AFLPs). Each AFLP primer combination is shown as the primer names used in this study followed by the three selective bases for the EcoRI and MseI primers, respectively (see supplementary methods). # screened in number of codominant marker primer pairs tested. # developed is number of codominant markers that could be scored in *Senecio*. # genotyped is number of markers showing polymorphism in the F2AC mapping family. # mapped is number of markers that were included in the final genetic map. Unlinked names markers that were unlinked at a >4 LOD or > 20 cM linkage threshold limits for mapping. Problematic refers to markers that caused problems with linkage group marker order or were present in multiple linkage groups. Dominant marker names are approximate fragment base pair lengths used to label AFLP loci.

Table 2. Summary genetic linkage map statistics for the *S. aethnensis* x *S. chrysanthemifolius* F$_2$ genetic map

Linkage group	Length	Total marker no.	Dominant marker no.	Codminant marker no.	Add2s length	Method4 length
1	44.5	18	4	14	50.05	49.74
2	29.1	7	4	3	34.65	38.80
3	42.6	10	6	4	48.15	52.07
4	41.3	10	6	4	46.85	50.48
5A	25.8	9	5	4	31.35	32.25
5B	9.5	4	0	4	15.05	15.83
6	41.7	9	7	2	47.25	52.13
7A	14.2	15	14	1	19.75	16.57
7B	3.2	2	1	1	8.75	9.60
8A	27.5	22	16	6	33.05	30.12
8B	5.2	2	1	1	10.75	15.60
9	15	8	7	1	20.55	19.29
10A	10	11	8	3	15.55	12.00
10B	4.2	2	0	2	9.75	12.60
Total	313.8	127	77	50	391.56	407.06
Mean	22.41	9.07	5.50	3.57	27.97	29.08
St. dev.	15.56	5.88	4.55	3.37	15.56	16.65

Notes. Map distance measures are in Kosambi centiMorgan units. Add2s length is an estimate of chromosome length calculated as linkage group length plus twice mean linkage group distance. Method4 length is another estimate of chromosome length calculated as linkage group length times (marker number + 1)/(marker number - 1).

Table 3. Summary of transmission ratio distortion loci and tests of different transmission ratio distortion mechanisms for the *S. aethnensis* x *S. chrysanthemifolius* F$_2$ genetic mapping family.

Linkage group	Reference marker	Map position	Late-acting sufficient	Asymmetric	Pre-zygotic	Heterozygote	Epistasis (minority genotype)
AC1 cluster	EC296	0.0 cM	no	no	aeth	yes	no
AC3 singleton (proximal)	E1M3_147	0.0 cM	yes	no	-	-	AC3 central (CA), AC6 (CD)
AC3 singleton (central)	S20	24.3 cM	no	chrys	chrys	no	AC3 proximal (AC), AC3 distal (AC), AC6 (AC)
AC3 cluster (distal)	E5M3_405	36.5 cM	yes	chrys	-	-	AC3 central (CA), AC6 (DC)
AC4 singleton	E1M5_215	0.0 cM	no	no	-	-	no
AC6 singleton	E1M5_131	41.7 cM	no	no	-	-	AC3 proximal (DC), AC3 central (CA), AC3 distal (CD)
AC7 cluster (central)	E1M5_269	12.5 cM	no	no	-	-	no

AC9 singleton	E1M5_132	15 cM	no	chrys	-	-		no
AC10A cluster	EC290	3.4 cM	no	aeth	aeth	no		no
AC5B[1] singleton	ES25	0.0 cM	-	aeth	-	-		-
AC10B[1] singleton	S26	0.0 cM	-	no	aeth	no		-

Notes. Linkage group names correspond to Figure 1. Cluster or singleton refers to whether a cluster or a single marker within 10 cM was observed to have genotypic TRD. [1] indicates that the last two TRDLs were only observed when specific mechanisms were investigated so are not presented in Figure 1 or counted as primary TRLDs in the text. The 'reference marker' is the marker with strongest TRD where a cluster of distorted markers was observed. 'Late acting sufficient' tests if selective removal of the minority genotype according to observed post-mortality and reproductive failure would be sufficient to explain observed TRDLs. 'Asymmetric' tests if subsets of the mapping family divided according to parental cytotype showed asymmetric TRD in one cross direction only with "aeth" and "chrys" indicates if asymmetric TRD was observed in an *S. aethensis* or *S. chrysanthemifolius* cytoplasmic background, respectively. 'Pre-zygotic' tests for allelic TRD with "aeth" and "chrys" indicating the parental alleles showing significant deficiencies and "-" indicating that no suitable codominant markers were available for testing at this TRDL. 'Heterozygote' tests for significant excesses or deficiencies in heterozygosity of genotypes scored according to parental allelic state with "-" indicating that no suitable codominant markers were available for testing at this TRDL. 'Epistasis' tests for non-independence of

paired TRDL reference marker genotypes, with significantly interacting TRDLs listed. Letter codes in parentheses indicate minority genotype for the reference TRDL followed by the interacting TRDL. Genotype codes are: A = homozygous *S. aethnensis* alleles, C = not homozygous for *S. aethnensis* alleles, D = not homozygous for *S. chrysanthemifolius* alleles.

Figure 1. Genetic map of a reciprocal F₂ *S. aethnensis* and *S. chrysanthemifolius* mapping family showing mapped marker positions and associated transmission ratio distortion.
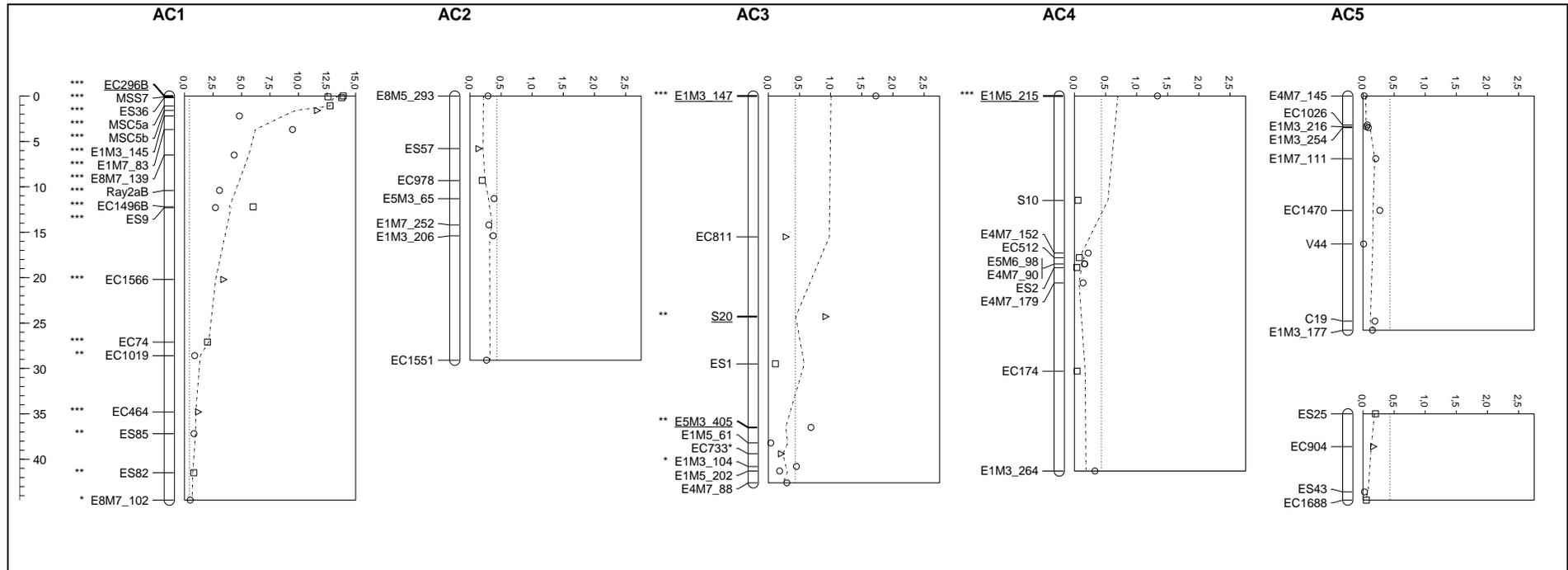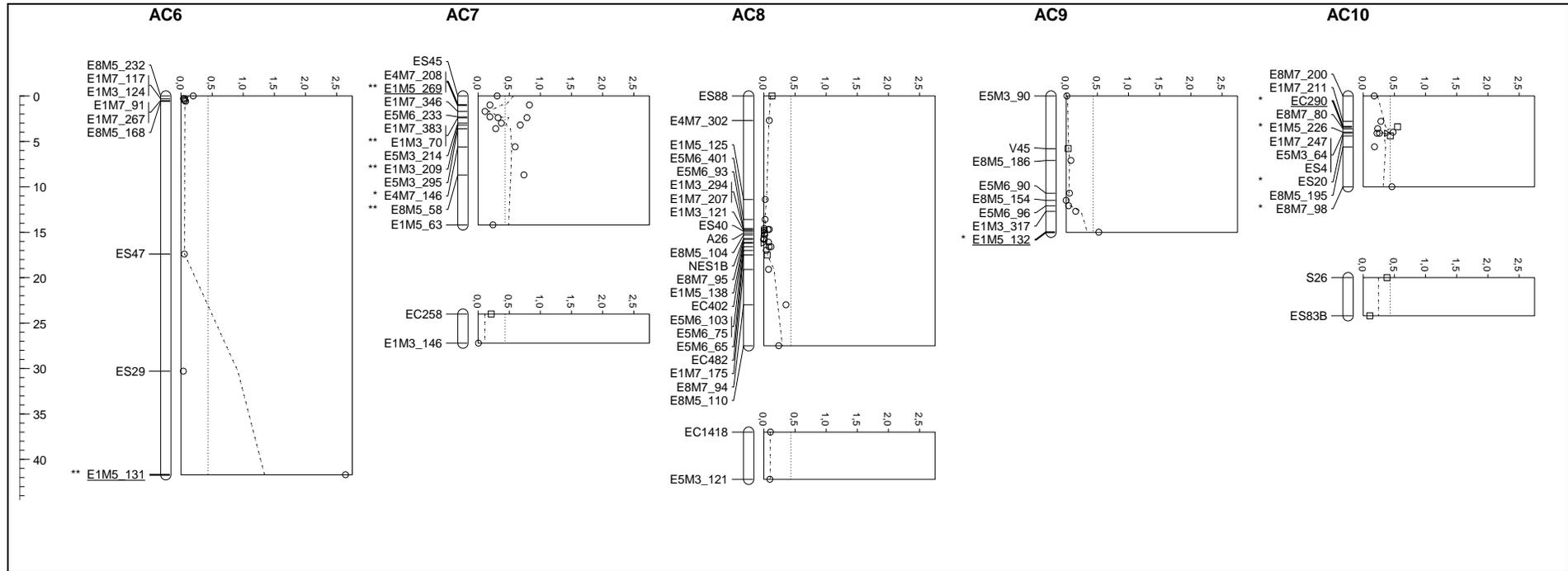
Map distances in Kosambi centiMorgans are shown in the scale to the left of linkage groups. Linkage groups are represented by vertical bars with mapped locus positions indicated with horizontal lines. Weakly linked linkage groups (< 4 LOD or > 20 cM) that are thought to belong to the same chromosome are aligned vertically. Locus names are listed to the left of linkage groups. Asterisks to the left of locus names indicate significant transmission ratio distortion of genotype frequencies at a single locus χ2 test 5 % confidence limit. Underlined locus names indicate

most strongly distorted locus within a particular cluster of distorted loci. Plots to the right of each linkage group show -log1000 p values for single locus $\chi2$ tests for transmission ratio distortion of genotype frequencies. Plot symbols indicate the number of degrees of freedom for the $\chi2$ tests; circles = 1 df, squares = 2 df, and triangles = 3 df. The dotted lines indicate the moving average of three neighbouring loci while dashed lines indicate the 5 % significance level, to the right of which, loci show significant transmission ratio distortion. The plot scale for linkage group AC1 is larger than the other linkage groups.