Title

Repeated inversions at the *pannier* intron drive diversification of intraspecific colour patterns of ladybird beetles

Abstract

How genetic information is modified to generate phenotypic variation within a species is one of the central questions in evolutionary biology. Here we focus on the striking intraspecific diversity of more than 200 aposematic elytral (forewing) colour patterns of the multicoloured Asian ladybird beetle, *Harmonia axyridis*, which is regulated by a tightly linked genetic locus *h*. Our loss-of-function analyses, genetic association studies, *de novo* genome assemblies, and gene expression data reveal that the GATA transcription factor gene *pannier* is the major regulatory gene located at the *h* locus, and suggest that repeated inversions and cis-regulatory modifications at *pannier* led to the expansion of colour pattern variation in *H. axyridis*. Moreover, we show that the colour patterning function of *pannier* is conserved in the seven spotted ladybird beetle, *Coccinella septempunctata*, suggesting that *H.*

12 *axyridis'* extraordinary intra-specific variation may have arisen from ancient modifications in

13 a conserved elytral colour patterning mechanisms in ladybird beetles.

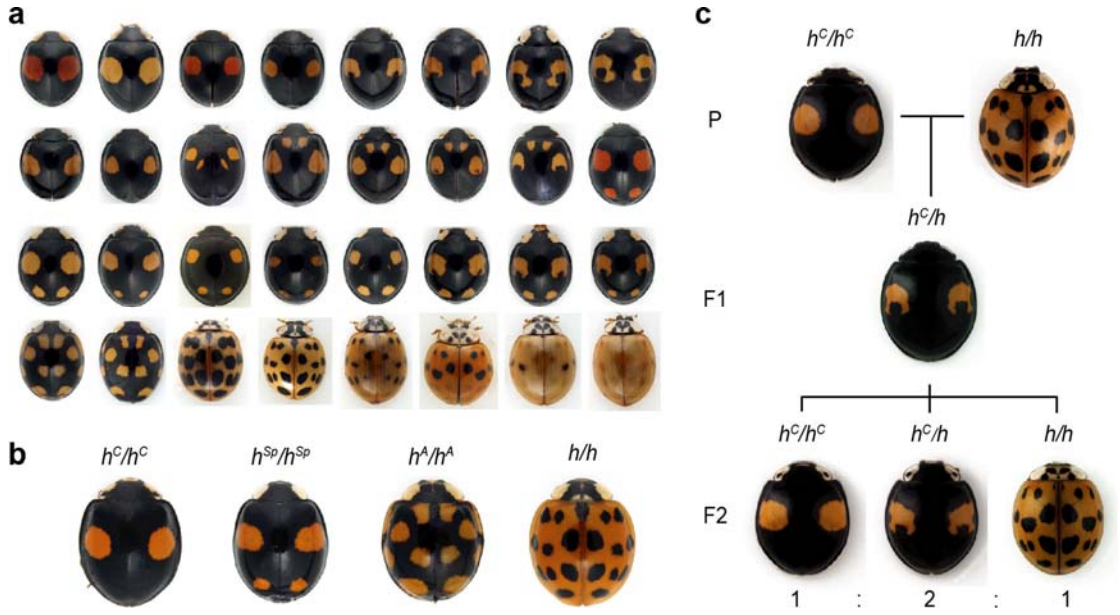
14 **Main text**

15 There are approximately 6,000 ladybird beetle species described worldwide[1].
16 Charismatic and popular, ladybird beetles are famous for the red and black spot patterns on
17 their elytra (forewings), thought to be a warning signal to predators that they store bitter
18 alkaloids in their body fluids[2, 3] and are unpalatable. This red/black warning signal is shared
19 among many ladybird beetle species, and provides a model for colour pattern mimicry by
20 other insect orders. While most ladybird beetle species have only a single spot pattern, a few
21 display remarkable intraspecific diversities, such as the multicoloured Asian ladybird beetle,
22 *Harmonia axyridis*, which exhibits more than 200 different elytral colour forms (Fig.1a).
23 This striking intraspecific variation prompted us to investigate its genetic and evolutionary
24 basis.

The first predictions regarding the genetics underlying the highly diverse elytral colour patterns of *H. axyridis* were made by the evolutionary biologist, Theodosius Dobzhansky based on his comprehensive classification of specimens collected from various regions in Asia[4]. Successive genetic analyses[5–7] revealed that many of these colour patterns are actually regulated by a tightly linked genetic locus, *h*, which segregates either as a single gene, or as strongly linked pseudoallelic genes (a super-gene[8, 9]) (Fig.1b, c). The elytral colour patterns are assumed to be formed by the superposition of combinations of two of the four major allelic patterns and dozens of minor allelic colour patterns (more than 20 different allelic patterns in total). The major allelic patterns cover more than 95% of colour patterns in the natural population[4]. In the elytral regions where the different colour elements are overlapped in heterozygotes, black colour elements are invariably dominant against red

colour elements (mosaic dominance[10]). Whether all of the supposed alleles linked to the *h* locus correspond to a single gene or multiple genes is yet unknown. Elucidating the DNA structure and the mechanisms underlying the evolution of this tightly linked genetic locus that encodes such a strikingly diverse intraspecific colour pattern polymorphism would provide a case-study that bears upon a major evolutionary-developmental biology question; how does morphology evolve? Here we show that the gene *pannier* is responsible for controlling the major four elytral colour patterns of *H. axyridis*. Moreover, we illustrate how modification to this ancient colour-patterning gene likely contributed to an explosive diversification of colour forms.

Figure 1. Intraspecific genetic polymorphisms of elytral colour patterns in *Harmonia axyridis*. a, Highly diverse elytral colour patterns of *H. ayridis*. **b**, Four major alleles of the elytral colour patterns. h^C, *conspicua*; h^{Sp}, *spectabilis*; h^A, *axyridis*; *h*, *succinea*. **c**, An example of inheritance of elytral colour forms. When h^C/h^C and *h*/*h* are crossed (P), all F1 progenies show the colour pattern of h^C/h. Note the small black spots within the red spots. When the F1 heterozygotes are sib-crossed, F2 progeny show three phenotypes (h^C/h^C, h^C/h and *h*/*h*) at the 1:2:1 ratio predicted for Mendelian segregation of a single locus. Inheritance of any combination of colour patterns follows this segregation pattern.



55 **Results**

56 To identify the gene regulating elytral colour pattern formation of *H. axyridis*, we

57 first investigated the pigmentation processes during development. In the developing pupal

58 elytra, red pigment (carotenoids[11]) was accumulated in the future red-pigmented regions (Fig.

59 2a pharate adult elytron). Red pigmentation occurred only in the thick ventral epidermal cells

60 of the two layers of the elytral epidermis (Fig. 2b; 2c, Red), and started at 80 hours after

61 pupation (80 h AP). Black pigmentation (melanin accumulation[11]) occurred only in the dorsal

62 cuticle of black-pigmented regions (Fig. 2d), and started approximately 2 h after eclosion.

63 Although pharate adult elytra are not black, we detected a strong upregulation of enzymatic

64 activity related to melanin synthesis[12] in the nascent dorsal cuticle in the future black regions

65 from 80 h AP (Fig. 2a, lower panels; Fig. 2c, black; Supplementary Figure 1). Every

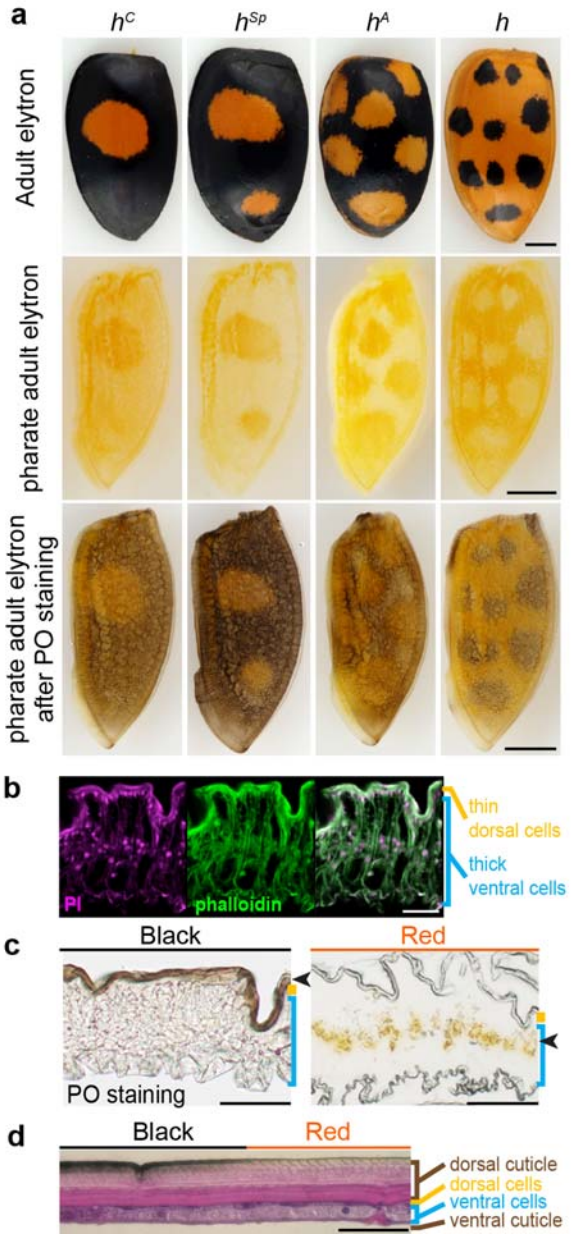
66 black-pigmented region was deployed complementary to the red regions. Therefore, we

67 concluded that the developmental programmes for both red and black pigmentation started

68 around 80 h AP.

69

Figure 2 | Developmental programmes for elytral pigmentation initiate at the late pupal stage in *H. axyridis*. **a**, Adult elytral colour patterns (upper panels), and localisation of carotenoid (middle panels, orange) and phenol oxidase (PO) activity (lower panels, black) in pharate adult elytra at 96 h AP. Proximal is up. Outer rims are to the left. **b**–**d,** Cross sections of elytra at 96 h AP (**b**, **c**) and adult (**d**). Dorsal is up. **b**, magenta, nuclei (propidium iodide); green, F-actin (phalloidin). **c**, left, PO staining; right, No staining. Arrowheads indicate pigmented areas. **d**, Haematoxylin and eosin stain. Scale bars, 1 mm in (**a**), 50 μm in (**b**, **c**, **d**).

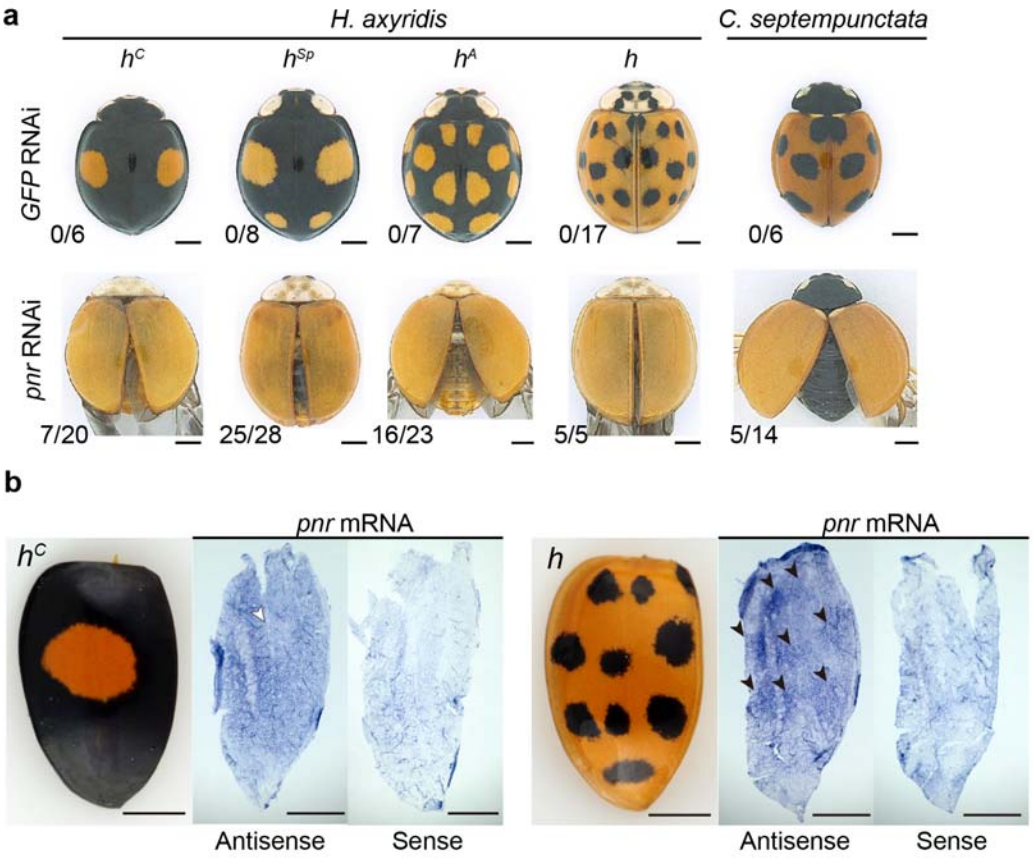


We hypothesised that some of the conserved genes essential for insect wing/body wall patterning[13–18] are recruited to regulate these elytral pigmentation processes, and tested this possibility using larval RNAi[19]. We performed small-scale candidate screening focusing on genes involved in wing/body wall patterning (Supplementary Table 1), and found that the *Harmonia* ortholog of *Drosophila pannier*, which encodes a GATA transcription factor[20], is essential for formation of all of the black-pigmented regions in the elytra. For all four major *h* allele backgrounds, larval RNAi targeting *pannier* resulted in complete loss of black colour elements and alternative emergence of red colour elements in the elytra (Fig. 3a, *H. axyridis*), indicating that *pannier* is essential for inducing black pigmentation in dorsal elytral cells and supressing red pigmentation in ventral elytral cells. This result was unexpected because *pannier* is not essential for wing blade patterning in *Drosophila*, but rather essential for

89 patterning of the dorsal body plate attached to the wings (notum) [21–23].

90

Figure 3 | The pattern of *pannier* expression foreshadowing the adult elytral colour pattern is necessary for switching red/black pigmentation processes in ladybird elytra. **a.** The adult phenotypes of RNAi treatments targeting *GFP* (negative controls, *GFP* RNAi) and *pannier* (*pnr* RNAi) in *H. axyridis* (h^C, h^{Sp}, h^A, *h*) and *C. septempunctata*. Scores in the lower left corners indicate penetrance of the loss-of-pattern phenotype in surviving animals. **b**, The pattern of *pannier* expression (*pnr*) in the dorsal elytral epidermal cells immediately before or after pigmentation (76−84 h AP). Left panels indicate the corresponding adult elytral phenotypes (h^C and *h*). White arrowhead, the region with a weak signal. Black arrowheads, the regions with intense signals. Scale bars, 1 mm.



101 *pannier* mRNA was upregulated from 48 h AP to 96 h AP in elytra, and preferentially in

102 black regions (h^C, Supplementary Figure 2). Immediately before or after 80 h AP (start of the

103 pigmentation programme, 76−84h AP), *pannier* seemingly showed higher expression in the

104 future black regions in the dorsal elytral epidermis (Fig. 3b). These data suggest that

105 region-specific upregulation of *pannier* during the pupal stage regulates black pigmentation

106 in the ladybird beetle's dorsal elytral cells, and that regions of expression differ among the

107 major *h* alleles to form different black patterns in *H. axyridis*.

108 These data led us to test whether *pannier* is associated with the classically

109 identified locus *h*, which regulates elytral colour patterns. To identify DNA sequences near

110 the *h* locus, we assembled *de novo* genome sequences (assembly version 1: 423 Mb; contig

111 N50, 63.5 kb; scaffold N50, 1.6 Mb), and performed a genetic association study using the

112 strains with different *h* alleles. We obtained the scaffold containing *pannier* and two

113 additional adjacent scaffolds based on the truncated gene structures at the scaffold ends (Fig.

114 4a, *Bgb* and *pnr*). Restriction-site Associated DNA Sequencing (RAD-seq) analysis of

115 backcrossed progenies (BC1, $h^A \times h^C$ F0 cross, n = 183) revealed that these three scaffolds are

116 included in the five scaffolds that showed complete association with colour patterns (Fig. 4a,

117 the upper left panel). In addition, genotyping of F2 individuals from two other independent

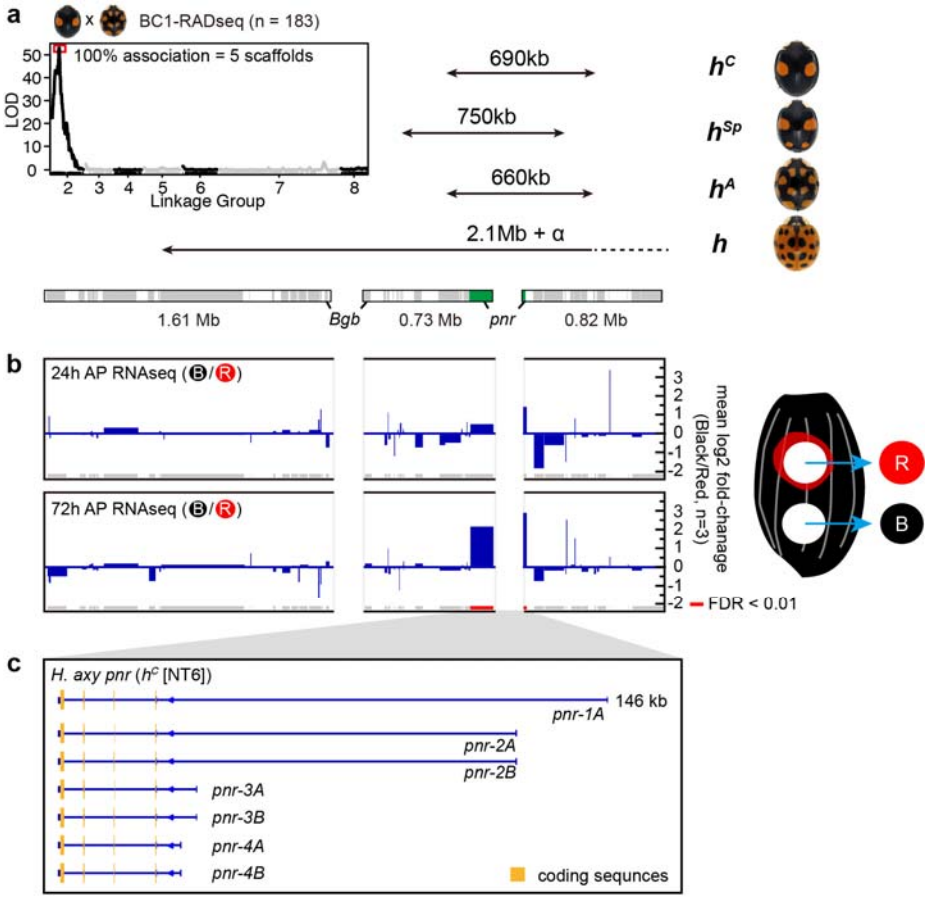
118 genetic sib-crosses ($h^C \times h$ F0 cross [n = 80] and $h^A \times h^{Sp}$ F0 cross [n = 273]) indicated that the

119 *pannier* locus is included in the relevant regions of all of the major four *h* alleles (h^C, 690 kb;

120 h^{Sp}, 750 kb; h^A, 660 kb; *h*, >2.1 Mb) (Fig. 4a, Supplementary Table 2).

121

Figure 4 | *pannier* is the major elytral colour pattering gene located at the *h* locus. a, Genetic association study of the *h* locus. Upper left panel, LOD plot for 4,419 RAD tags deduced from genotyping BC1 progeny from the h^C–h^A F0 cross. RAD tags showing segregation patterns of markers located on the X chromosome were excluded from the analysis. The LOD score peaked at RAD tags in Linkage Group 2. RAD tags with complete association with elytral colour patterns corresponded to 5 genomic scaffolds including the three scaffolds adjacent to the *pannier* locus (lower bars). Upper right panel, the candidate genomic regions responsible for the major four *h* alleles. Lower bars, the three genomic scaffolds adjacent to the *pannier* locus. Grey, genes predicted from RNA-seq. Green, the *pannier* locus. Contiguity of the scaffolds was predicted based on the truncated genes at the end of scaffolds (*Bgb* and *pnr* [*pannier*]). Arrows indicate the respective candidate genomic regions responsible for each allele. *pannier* is included in all four relevant regions. **b**, Fold changes of gene expression in the presumptive red and black elytral epidermis before pigmentation around the candidate genomic region responsible for *h*. Grey and red bars on the bottom indicate predicted genes. Red, FDR < 0.01. Only *pannier* was significantly upregulated in this region. The samples for RNA-seq were collected as depicted in the right panel in h^C background. **c**, Gene structures of *pannier* in *H. axyridis*. 1A isoform cDNA was cloned by rapid amplification of cDNA ends (RACE). 2A−4B isoforms were predicted from RNA-seq analysis. *pannier* has at least 4 transcription start sites. Coding sequences are located from exon 2 to exon 5 (yellow). There are two alterative exons at exon 3, one encoding one of the two zinc finger domains of Pannier (A isoforms), and the other skipping that zinc finger domain (B isoforms).



146 To test contiguity of these three scaffolds, we re-assembled the genome using a novel

147 genome assembler (Platanus2), and additional *de novo* genomic assemblies of h^C, h^A and h

148 alleles using linked-read and long read sequencing platforms (10x Genomics Chromium

149 system; PacBio system). We obtained longer genomic scaffolds including the three described

150 above (h^C, 3.13 Mb/2.74 Mb; h^A, 1.42 + 1.61 Mb; h, 2.79 Mb, Supplementary Table 3) and

151 the genotyping markers showing complete association with colour patterns and incomplete

152 association at both ends (h^C and h) or one end (h^A) of each scaffold (Supplementary Figures 3,

153 4a−c). These data support the result of our genetic association studies.

154 To further delimit the candidate genes associated with the elytral colour patterns,

155 we performed RNA-seq analysis using epidermal tissues isolated from the developing red or

156 black regions before pigmentation in the h^C genetic background (Fig. 4b, 24 h & 72 h AP

157 RNA-seq). We found that *pannier* was the only gene statistically significantly upregulated in

158 the developing black region compared to the red region at 72 h AP within the *h* locus

159 candidate region (Fig. 4b, red bars, False discovery rate [FDR] < 0.01; Supplementary Table

160 4). These data pinpoint *pannier* as the major gene regulating the elytral colour pattern

161 variation in *H. axyridis*.

162 We next investigated allele-specific polymorphisms at the *pannier* locus. We found

163 that alleles of the first intronic region of *pannier* are more diverse than the surrounding

164 genomic regions (Fig. 5a, the whitish regions in the middle), whereas the same allele in

165 different strains shows conserved fragments distributed throughout the region (Fig. 5a, blue

166 bars, h^C[F2-3]–h^C[NT6]). In comparisons between the alleles, we consistently found traces of

167 large inversions in the upstream half of the first intron (reddish lines in Fig 5a, h^C–h^A, h^A–h,

168 *h*–h^{C}; 56 kb–76 kb in size). However, we found only a single non-synonymous substitution in

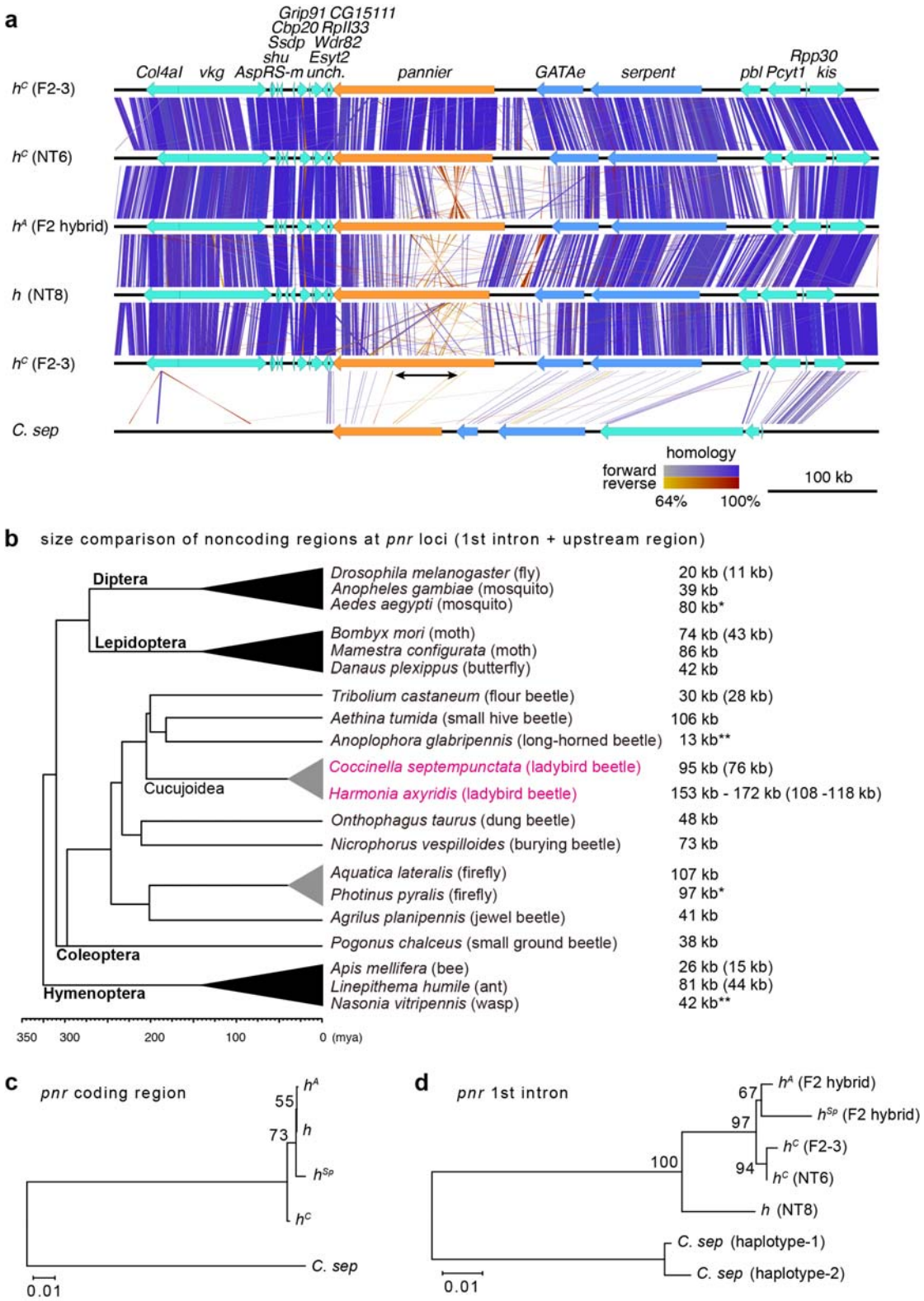
169 the region not conserved among organisms (G235V, h^{Sp}), (Supplementary Figure 5, 6),

170 suggesting that cis-regulatory differences in the 1st intronic region of *pannier* are the major

171 cause of intraspecific colour variation.

172

Figure 5. Intraspecific diversification and interspecific conservation of the 1st intronic region of *pannier* in ladybird beetles. a, Sequence comparison of the genomic region surrounding the *pannier* locus. 700 kb genomic sequences surrounding the *pannier* locus were extracted from the genome assembly of each allele in *H. axyridis* (h^C, h^A, *h*) and *C. septempunctata* (*C. sep*). Strain names are given in parentheses. Arrows indicate genes predicted by the exonerate programme (Orange, *pannier*; Blue, GATA transcription factor genes paralogous to *pannier*; Green, other genes). Gene names are listed at the top. Vertical or diagonal bars connecting adjacent genomic structures indicate BLAST[47] hit blocks (bluish, forward hit; reddish, reverse hit). The colour code for colouring the bars is at the bottom. The upper half (1st intron) of the *pannier* locus is diversified (whitish) between different *h* alleles in *H. axyridis*, and shows traces of inversions (crossed reddish bars). Several intronic sequences are conserved between *H. axyridis* and *C. septempunctata* (bars located in the upper half of *pannier* in *C. sep*). The black arrow indicates the region specifically expanded in *H. axyridis*. **b,** Overview of the size of the upper noncoding regions (the 1st intron + the upper intergenic region) at the *pannier* locus in holometabolous insects. The topology of the phylogenetic tree of surveyed insects is adapted from ref. [48] (Coleoptera), and TIMETREE[49] (Diptera, Hymenoptera, Lepidoptera). The sizes of the 1st intron are given in parenthesis if cDNA information was available. *H. axyridis* has the largest noncoding sequence at the *pannier* locus. In some species, synteny of the three paralogous GATA genes was broken up by translocation (*) or insertion (**). **c, d,** ML phylogenetic trees constructed with nucleotide sequences of *pannier* coding region (**c),** and those of the conserved region in the 1st intron (**d)**. The trees were drawn to scale with branch lengths measured in the number of substitutions per site. Bootstrap values were calculated from 1000 resampling of the alignment data.



196

197

198 Moreover, we found that in *H. axyridis*, the size of upstream noncoding

199 sequences of the *pannier* locus (including the 1st intron and the upstream intergenic region)

200 are 46−65 kb larger than the currently available corresponding genomic sequences of the

201 other holometabolous insects (Fig. 4b, *H. axyridis*, 153−172 kb; other holometabolous

202 insects, 13−107 kb). Comparison of the exon-intron structures of *H. axyridis* to those of some

203 of the holometabolous insects also suggested that especially the 1st intron of *pannier* is

204 expanded in *H. axyridis* (*H. axyridis*, 108−118 kb; the other holometabolous insects, 11−44

205 kb). The expanded region in *H. axyridis* included at least four transcription initiation sites of

206 *pannier* transcripts (Fig. 4c, *pnr-1A*−*4B*). In addition, in this region, several known

207 DNA-binding motifs of transcription factors involved in *Drosophila* wing formation were

208 more enriched allele-specifically than those in the other genomic regions (Table 1a). For

example, the highly conserved Scalloped (SD) DNA binding motif of the insect wing selector transcription factor complex Vestigial/Scalloped[24, 25] occurred frequently in upstream and downstream regions of the first intron of *pannier* specifically in the h^C allele (Table 1a, the upstream and donwstream regions of the 1st intron, *sd*; Supplementary Table 5).

214

Table 1 | Known DNA binding motifs enriched in the non-coding regions of the *pannier* loci.

Enriched DNA binding motifs of *Drosophila* transcription factors involved in wing formation are listed ($p < 0.05$).

a. Allele- or Specie-specifically enriched motifs

strain	upstream intergenic region	upstream region of 1st intron	downstream region of 1st intron
h^C	*EcR, foxo*	*Myc, en, exd, HLH106, h, ss, pan Mad, Mnt, pad, Poxn, sd, tgo, tai, vvl*	*B-H1, ab, crc, crol, Dr, rn, sd*
h^A	*exd, pan, vvl*	*ab, Spps, brk, usp, pnr*	*usp*
h	*kni, rn, sqz*	*ato, ab, Ets21C, rn, sqz*	*al, B-H1, lms, nub, sqz, tup*
C. sep	*ato, EcR, h, ss, tgo*	*B-H2, eg, tai, pnr, tgo*	*E(spl)mβ, ken, brk, tgo, h, tai*

b. Commonly enriched motifs

group	upstream intergenic region	upstream region of 1st intron	downstream region of 1st intron
Harmonia	*Mad, brk, ab, taxi, tgo*	*lms, exd, hth, tgo, tai, Mad al, brk, tx, E(spl)mβ, tgo*	*Dr, en, inv, Med, sens, slou, unpg*
Harmonia & Coccinella	*Mad*	*lms, exd, hth, tgo, tai, Mad*	–

c. Region size (bp)

strain	upstream intergenic region	upstream region of 1st intron	downstream region of 1st intron
h^C	38,192 (F2-3) / 51,566 (NT6)	71,107 (F2-3) / 76,264 (NT6)	47,450 (F2-3) / 44,236 (NT6)
h^A	42,826	75,417	52,210
h	40,778	67,340	67,340
C. sep	20,370	56,564	18,432

215

216

217 Furthermore, the RNA-seq data for the h^C background also revealed that the *sd* co-activator

218 gene *vestigial* was the only transcription factor gene that was significantly upregulated in the

219 future black region from early pupal stages (Supplementary Figure 7), implicating Vestigial

220 as one of the upstream trans-regulatory factors acting together with Sd to form the

221 two-spotted elytral colour pattern of h^C. It is noteworthy that the non-coding region of

222 *pannier* in each allele possesses putative DNA-binding motifs that can respond to variety of

223 developmental contexts such as anterior-posterior patterning[26–28] (*en*, *inv*), wing fate

224 specification[17,24,29] (*sd*), hinge-wing blade patterning[17,18,30,31] & wing vein patterning[29, 32–34]

225 (*nub*, *rn*, *ab*, *al*, *B-H1*, *B-H2*, *ss*, *hth*, *exd*, *kni*, *brk*, *vvl*, *Mad*, *Med*, *h*), hormonal cues[35, 36] (*EcR*,

226 *usp*, *tai*), and auto-regulation (*pnr*) (Table 1a). These results suggest that allele-specific

227 elytral colour patterns of *H. axyridis* may be formed by integrating appropriate combinations

228 of developmental contexts of wing formation shared among insects.

229 We further tested whether the regulatory function of the red/black colour pattern in

230 elytra is a conserved or a derived aspect of *pannier* function in ladybird beetles using the

231 seven-spotted ladybird beetle, *Coccinella septempunctata*, which shows a monomorphic

232 seven-spotted elytral colour pattern. The *pannie*r mRNA was detected in the larval elytral

233 primordium, was upregulated from 24h AP to 96h AP, and preferentially expressed in the

234 black spots of elytra in *C. septempunctata* similar to that in *H. axyridis* (Supplementary

235 Figure 8). The black-to-red switching phenotype was also observed in *C. septempunctata*

236 adults treated with larval RNAi targeting *pannier* (Fig. 3a, *C. septempunctata*). These data

237 suggest that the elytral colour-patterning function of *pannier* may be conserved at the

238 inter-genus level in ladybird beetles. To investigate the putative regulatory sequences at the

239 *pannier* locus, we performed *de novo* assembly of the *C. septempunctata* genome using a

240 linked-read sequencing platform (10x Genomics Chromium system), and obtained a

241 contiguous genomic scaffold including the *pannier* locus (2.41 Mb, Supplementary Figure 4d,

242 Supplementary Table 3). Whereas the noncoding sequences of *C. septempunctata pannier*

243 are enriched with several species-specific DNA binding motifs (Table 1a, *C. sep*), we found

244 DNA-binding motifs commonly enriched between *Harmonia* and *Coccinella*, which are

245 associated with wing vein formation and wing/body wall patterning (*Mad*, *hth* and *exd*) [29,31,32]

246 (Table 1b). Therefore, co-option of such wing developmental modules in the regulatory

247 region may have facilitated acquisition of a novel expression domain of *pannier* in pupal

248 elytral blades in ladybird beetles.

In order to explore the history of the emergence of elytral colour patterns in *H. axyridis*, we also performed a molecular phylogenetic analysis focusing on the highly conserved *pannier* intronic sequences shared among *H. axyridis* alleles and *C. septempunctata* (3 blocks, totalling 1.1 kb in length, Supplementary Data 1). The maximum likelihood (ML) phylogenetic tree inferred from nucleotide sequences of the *pannier* coding region did not resolve the phylogenetic relationship among the alleles in *H. axyridis* to a satisfactory level (Fig. 5d, bootstrap values < 75). However, the ML tree inferred from the conserved intronic sequence suggested that in *H. axyridis* the contrasting colour patterns of the *h* allele (black spots in red background) and the other three alleles (red spots in black background) diverged first. The latter three alleles diverged more recently (Fig. 5e, bootstrap values > 90).

Discussion

The *pannier* locus identified in this study appears to be the key genetic locus responsible for the origin of large-scale intraspecific variation genetically linked to the *h* locus in ladybird beetles[1,2]. Based on the results presented in this study, we propose an evolutionary model that might underlie the high level of diversification of the intraspecific elytral colour patterns of *H. axyridis*. In addition, we also discuss the underlying evolutionary developmental backgrounds specific to ladybird beetles.

The common ancestor of *Harmonia* and *Coccinella* (Coccinellinae) diverged more than 33.9 million years ago, according to molecular phylogenetic analyses and fossil records[37, 38]. Therefore, the elytral colour-patterning function of *pannier* shared between *H. axyridis* and *C. septempunctata* was most likely acquired before this divergence event. The

272 1.1 kb sequence blocks in the 1st intron of *pannier* conserved between *H. axyridis* and *C.*

273 *septempunctata* are a likely candidate for a regulatory element associated with the ladybird

274 beetle-specific elytral expression of *pannier* in the pupal elytra. The effects of enhancer

275 activities of these sequence blocks have not yet been experimentally addressed. However, the

276 acquisition of such regulatory sequences during evolution would have coincided with the

277 acquisition of the elytral colour patterning function of *pannier* (Fig. 6, blue diamond). These

278 conserved sequence blocks are located in the expanded intronic region specific to *H. axyridis*

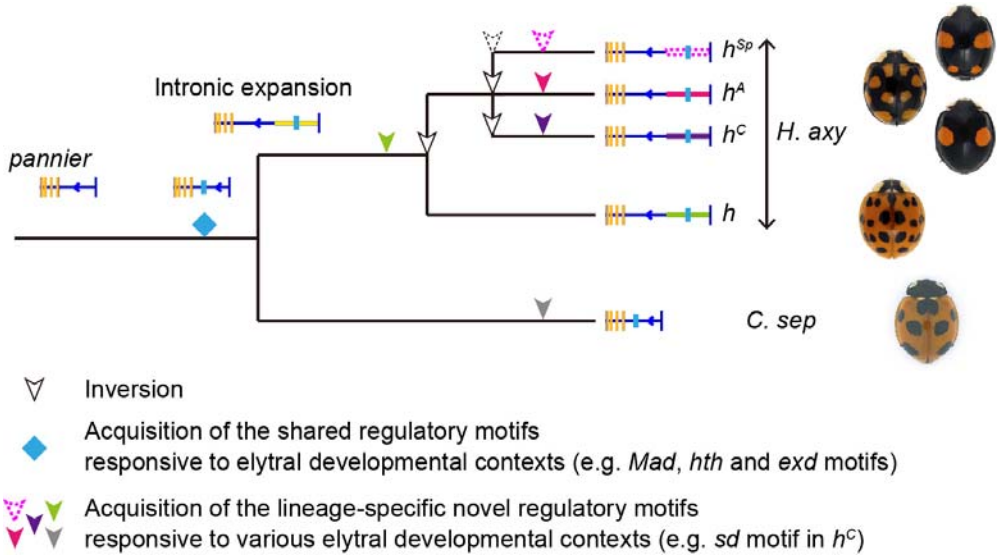
279 (Fig. 5a, black arrow). Therefore, the expansion of the 1st intron in the ancestral lineage of *H.*

280 *axyridis* (Fig. 6, intronic expansion) might be one of the events that facilitated diversification

281 of the intraspecific elytral colour patterns.

282

Figure 6. Evolutionary model for the acquisition of the highly diversified intraspecific elytral colour patterns of *H. axyridis*. See details in the Discussion.



In the genus *Harmonia*, colour patterns similar to those encoded by the *h* allele and those of *C. septempunctata* (black spots in red background) are commonly observed. Also the position of the spots is similar across species (e.g. *H. quadripunctata*, *H. octomaculata*, and *H. dimidiata*). Therefore, we speculate that the intronic sequence of *pannier* in the *h* allele of *H. axyridis* might retain a repertoire of regulatory sequences acquired in a common ancestor of the genus *Harmonia* (Fig 6, green arrowhead). However, in the ancestral lineage of *H. axyridis*, the regulatory region of *pannier* appears to have been modified to generate novel colour patterns of the recently diverged alleles (h^{C}, h^{Sp} and h^{A}; red spots in black background; Fig. 6 magenta, red and purple arrowheads). The 46−65 kb-scale intronic region at the *pannier* locus that is specifically expanded in *H. axyridis* (Fig. 6, intronic expansion, yellow box) might have facilitated accommodation of the allele-specific regulatory motifs

298 responsible for the diversified colour pattern of elytra. In addition, traces of inversions in this

299 region consistently found in allele comparisons suggest that repeated inversions in this region

300 (Fig. 6, white arrowheads) created opportunities to diverge the noncoding sequence of

301 *pannier* to successively generate novel diverse alleles within a species by suppressing

302 recombination within this region. Such inversion events would have occurred in the common

303 ancestor of *H. axyridis* and its reproductively isolated sister species, *H. yedoensis* because the

304 major elytral colour patterns are shared between the two species[39]. Large-scale chromosomal

305 inversion is believed to be one of the major driving forces generating and maintaining

306 intraspecific morphological variation within a species [40–43]. Our study exemplifies that not

307 only a single inversion event but also repeated inversion events at an expanded intron can

308 lead to the acquisition of novel morphological traits within a species.

309 From the viewpoint of evolutionary developmental biology it is noteworthy that in

310 *H. axyridis*, of all of the developmental genes known to regulate colour pattern and

311 pigmentation, a single gene, *pannier*, is responsible for the major classes of intraspecific

312 entire wing colour pattern diversification. This evolutionary pattern contrasts with that of the

313 intensely studied warning signals of *Heliconius* butterflies. In this case, more than 10 loci

314 regulate multiple intraspecific wing colour patterns prevailing in the population[44]. This

315 difference in evolutionary mechanisms may stem from a paucity of available options of

316 evolvable genes in the gene regulatory network of elytral colour patterning. Ladybird beetles

317 diverged from ancestral species of Cucujoidea[37] (Fig. 5a, Cucujoidea), leaf-litter or

318 rotten-tree dwelling insects. Thus, the ancestor of ladybird beetles would have had far less

319 colourful and more simply patterned forewings (elytra) than the ancestors of butterflies,

320 moths. Therefore, these ancestors presumably would have possessed far fewer colour pattern

321 regulatory genes. In *H. axyridis*, this developmental constraint may have led to the selection

322 of *pannier* as the major evolvable gene to a signal-integrating 'input-output' regulatory

323 gene[45, 46]. This might have generated more than 200 colour patterns genetically tightly linked

324 to the *h* locus by utilising the expanded regulatory DNA sequence. Future research aiming to

325 identify specific regulatory inputs to *pannier* will help clarify the regulatory mechanisms

326 underlying the generation of highly diverse intraspecific polymorphism at the interspecific

327 level. Another important issue to clarify whether *pannier* is indeed the hotspot of

328 morphological evolution in ladybird beetles is whether *pannier* is responsible for the

329 remaining >20 minor colour patterns in *H. axyridis*.

330 **Methods**

331 **Insects**

332 Laboratory stocks of *H. axyridis* and *C. septempunctata* were derived from field

333 collections in Japan. They were reared at 25°C and usually fed on artificial diet[50], or fed on

334 the pea aphid *Acyrthosiphon pisum* (kindly provided by Dr. T. Miura) for egg collection.

335 Larvae and pupae analysed in this study were not sexed.

336

337 **Phenoloxidase (PO) activity staining**

338 Pupa elytral discs were dissected in a potassium phosphate buffer (100 mM

339 KH_2PO_4/K_2PO_4, 150 mM NaCl, pH 6.3) on ice. PO staining was performed using 0.4 mg/ml

340 dopamine as a substrate for 2 hours at room temperature as previously described[13]. After

washing several times in the potassium phosphate buffer containing 0.3% Triton-X100 and mounted in this solution. Images were captured with a stereoscopic microscope (MZ FLIII, Leica) equipped with a digital camera (DP70, Olympus).

Histological analysis

To visualise tissue morphology and PO active tissues, pharate adult elytra dissected in ice-cold PBS (137 mM NaCl, 2.68 mM KCl, 10.14 mM Na_2HPO_4, pH 7.2) at 96 h AP or those after PO activity staining were fixed with 4% paraformaldehyde (PFA) in PBS for 15 min on ice and for 75 min at room temperature. After fixation, the elytra were washed several times in 100% methanol and stored in 100% methanol at –20°C until use. After dehydration, the elytra were embedded in 4% carboxymethyl cellulose (CMC; FINETEC), and were

352 frozen in hexane cooled with dry ice. The freeze-embedded elytra were stored at −80°C until

353 use. The 6 μm frozen sections were prepared as described previously[51] using an adhesive film

354 (Cryofilm Type 1; FINETEC). Sections of the PO activity stained elytra were mounted in

355 PBS and photographed using IX70. For nuclear and F-actin staining, sections were treated

356 with 2.5 μg/ml propidium iodide, 1 mg/ml RNase A and 5 U/ml AlexaFluor 488 phalloidin

357 (Molecular Probes) for 1 hour at 37°C under a dark condition. After washing three times in

358 PBS, the sections were mounted in an antifade reagent (FluoroGuard™; Bio-Rad), and

359 images were captured with a confocal laser-scanning microscope (LSM 510; Carl Zeiss).

360 For localisation of carotenoid, elytra at 96 h AP were embedded and sectioned as

361 described above. All procedures were rapidly performed to prevent diffusion of carotenoids.

362 The sections were dried for 1 min, mounted in PBS and immediately photographed using

363 IX70.

364

365 **cDNA cloning**

366 Larval and pupal elytral discs and pharate adult elytra of *H. axyridis* (h^C) and *C.*

367 *septempunctata* were dissected in PBS on ice. Soon after dissection, the tissues were frozen

368 in liquid nitrogen and stored at −80°C until use. Total RNA was extracted from each sample

369 using TRIzol Reagent (Invitrogen) or RNeasy Micro Kit (Qiagen) according to the

370 manufacture's instructions, and treated with 2 U DNase I (Ambion) for 30 min at 37 °C. The

371 first-strand cDNA was synthesised with SMARTer PCR cDNA Amplification Kit (Clontech)

372 using 1 μg of total RNA according to the manufacture's instructions. *H. axyridis* and *C.*

373 *septempunctata* cDNA fragments were amplified by reverse transcription-polymerase chain

374 reaction (RT-PCR) and rapid amplification of cDNA ends (RACE) with the primers listed in

375 Supplementary Table 6−8. The PCR product was cloned into the *Eco*R V site of the

376 pBluescript KS+ vector (Stratagene) or pCR4-TOPO vector (TOPO TA Cloning Kit;

377 Invitrogen). The nucleotide sequences of the PCR products were determined using a DNA

378 sequencer 3130 genetic analyser (Applied Biosystems). The SNPs in open reading frame

379 (ORF) of *pannier* were determined through direct sequencing of the PCR products treated

380 with ExoSAP-IT (Affymetrix). Sequencing was performed by DNA sequencing service

381 (FASMAC) using the primers listed in Supplementary Table 6. Sequence analysis was

382 carried out using DNASIS (Hitachi Software Engineering) or ApE[52] (version 2.0.45)

383 software. Nucleotide sequences and deduced amino acid sequences were aligned with

384 ClustalW in MEGA[53] software (version 7.0.18) and the alignment figure was generated using

385 Boxshade[54] (version 3.21).

386

387 **Gene expression analysis by RT-PCR**

388 For the gene expression analysis in each developmental stage, elytral tissues of

389 three individuals of *H. axyridis* (h^C) and *C. septempunctata* were dissected as described

390 above. Six elytral tissues from each sampling stage were pooled in one tube. Total RNA

391 extractions and the subsequent first-strand cDNA syntheses (using 425 ng and 267 ng of total

392 RNA for *H. axyridis* and *C. septempunctata* samples, respectively) were performed as

393 described above. Three μl of 100 and 62.8 times diluted *H. axyridis* and *C. septempunctata*

394 first-strand cDNA was used as a template for PCR, respectively. The PCR cycle numbers are

395 35 for all genes. A set of primers #1 and #2 for each gene was used for this analysis

396 (Supplementary Table 7).

397 For the gene expression analysis in the future red and black regions, red and black

398 regions of pharate adult elytra at 84 h AP were bored with a needle. Internal diameters of 0.7

399 and 0.6 mm were used for *H. axyridis* and *C. septempunctata*, respectively. In the case of *C.*

400 *septempunctata,* elytra stained with PO activity were used for boring since carotenoid

401 localisation was not observable unlike *H. axyridis*. cDNA synthesis was performed as

402 described above, using as much total RNA as we could extract. 20 μl of 10 times diluted

403 first-strand cDNA was used as a template for PCR. The PCR cycle numbers are 45 cycles for

404 *Ha-pnr*, 38 cycles for *Ha-rp49*, 47 cycles for *Cs-pnr* and 40 cycles for *Cs-rp49*. A set of

405 primers #3 and #4 for each gene was used for this analysis other than *rp49*. *Ha-rp49* and

406 *Cs-rp49* were used as internal controls. Reactions without reverse transcriptase were

407 performed with cDNA synthesis as negative control samples for the RT-PCR experiments.

408 No band was detected in these reactions for all genes. The primers used for this analysis are

409 described in Supplementary Table 7.

410

411 **Larval RNAi**

412 dsRNA synthesis and microinjection into larvae were performed as previously

413 described[19]. The cloned cDNA fragments amplified by PCR were used as templates for

414 dsRNA synthesis. Approximately 1.4 to 2.7 μg and 1.4 to 2.0 μg of the dsRNAs of *Ha-pnr*

415 and *Cs-pnr* were injected into two-day-old forth (final) instar larvae, respectively.

416 Approximately 2.0 to 2.7 μg and 1.4 to 2.7 μg of the *EGFP* dsRNA were injected into *H.*

417 *axyridis* and *C. septempunctata* larvae as negative controls, respectively. Different amount of

418 dsRNA for each gene in this range gave no difference in phenotypic effects (data not shown).

419 In order to give enough time for the completion of pigmentation, images of adults were

420 captured at more than 2 days after eclosion using a digital microscope (VHX-900, Keyence).

421

422 ***In situ* hybridisation**

423 Essentially the same protocol for whole mount pupal antennal primordia of the silk

424 moth[55] was used. To increase RNA probe penetrance in elytral epidermis covered with

425 sclerotised cuticle at 76−84 h AP, the peripheral edge of an elytron was cut off, and then,

426 ventral and dorsal elytral epidermis layers appressed together were carefully separated with

427 fine forceps after fixation. To reduce non-specific probe hybridisation, fixed, separated and

428 detergent-permeabilised elytra epidermal samples were stored in 100% methanol for more

429 than 12 hours at −30°C, and prehybridisation treatment was extended to two overnights. The

430 ventral epidermis samples were not used for analysis because of high non-specific

431 background signals. *pannier* antisense probes were designed at 5' and 3' regions of ORF

432 excluding the two conserved GATA zinc finger coding regions in the middle to prevent

433 cross-hybridisation with other GATA family genes. The PCR primers used to amplify the

434 template DNA for *in vitro* RNA probe synthesis were listed in Supplementary Table 9. Briefly,

435 *pannier* ORF fragment was amplified by RT-PCR using cDNA from 72 h AP (h^C), and cloned

436 into pCR4-TOPO vector (Invitrogen). Sense and antisense probe templates were amplified

437 from the cloned cDNA. Sense and antisense riboprobes were transcribed using the flanking

438 T7, T3 or SP6 promoter sequences. Mixture of 5' and 3' probes was used for hybridisation.

439

440 **_De novo_ genome assembly of _H. axyridis_**

441 A single female adult from F2-3 strain sibcrossed for 3 generations (h^C) was used

442 for the first version of *de novo* genome assembly of *H. axyridis*. Genomic DNA was extracted

443 using DNeasy Blood and Tissue Kit (Qiagen). Paired end (300 bp and 500 bp) and mate pair

444 (3 kb, 5 kb, 8 kb, 10 kb, 12 kb and 15 kb) libraries were constructed using TruSeq DNA

445 PCR-Free LT Sample Prep Kit and Nextera Mate Pair Sample Prep Kit (Illumina) following

446 the manufacturer's protocols. Sequencing libraries were run on Illumina HiSeq2500

447 sequencers. In total, we generated 133.6 Gb of raw sequence data for de novo genome

448 assembly. Genome assembly was performed using the Platanus v1.2.1.1 assembler[56] after

449 removal of adapter sequences and error correction (SOAPec v2.01)[57].

Re-assembly of the genomic sequence at the *pannier*-locus in the *H. axyridis* F2-3 sample

Adaptor sequences and low-quality regions in paired-end and mate-pair reads were trimmed using Platanus_trim[58] (version 1.0.7) with default parameters. Trimmed reads were assembled by Platanus2[59] (version 2.0.0), which was derived from Platanus[56] to assemble haplotype sequences (i.e. haplotype phasing) instead of consensus sequences. Procedures of Platanus2 are briefly described as follows:

(1) De Bruijn graphs and scaffold graphs are constructed without removal of bubble structures caused from heterozygosity. Paths that do not contain junctions correspond to assembly results (scaffolds). Scaffold pairs in bubbles represent heterozygous haplotypes.

(2) Paired-ends or mate-pairs are mapped to the graphs to detect links between bubbles, and linked bubbles are fused to extend haplotype sequences.

(3) Each haplotype (contig or scaffold) is independently extended by modules of *de novo* assembly derived from Platanus.

(4) Homologous pairs of haplotype scaffolds are detected using bubble information in the initial de Bruijn graph.

(5) Steps 1–4 are iterated using various libraries (paired-ends or mate-pairs).

(6) Homologous pairs of scaffolds are formatted into bubble structures as output. For each pair, longer and shorter scaffold were called "primary-bubble" and "secondary-bubble", respectively. Primary-bubbles, secondary-bubbles and non-bubble scaffolds are collectively called "phased-scaffolds".

472 In addition, Platanus2 can connect primary-bubbles and non-bubble scaffolds to construct

473 long "consensus scaffolds", which consists of mosaic structure of haplotypes (i.e. paternal

474 and maternal haplotypes are mixed). Employing the strategy of Platanus2, certain highly

475 heterozygous regions were expected to be assembled contiguously compared to Platanus.

476 Using the markers of the responsible region of wing pattern (*pannier*-locus), we found that

477 two long bubbles and one short non-bubble scaffold corresponded to the locus. Consequently,

478 one consensus scaffold was constructed from these phased scaffolds, which covering the

479 breakpoint markers at the *pannier* locus. We used that consensus scaffold (3.13 Mb) for the

480 downstream *in silico* sequence analyses. We assessed the completeness of the genome

481 assemblies using BUSCO[60] (version 3.0.2, Insecta dataset [1,658 orthologs]).

482

DNA sequencing for *de novo* genome assembly by long read and linked-read platforms

High molecular weight (HMW) genomic DNA was extracted using QIAGEN Genomic-tip 100/G (QIAGEN) according to the manufacturer's instructions. The concentrations and qualities of the extracted HMW genomic DNA were evaluated using Qubit dsDNA, and RNA HS kits (Thermo Fisher).

For library preparation for 10x Genomics Chromium system, one pupa (h^C [NT6 strain] and *h* [NT8 strain]) or one adult (h^A [F2 adult progenies in genetic cross $h^A \times h^{Sp}$] and *C. septempunctata* [MD-4 strain]) was used. Size selection by BluePippin (range: 50kb−80kb, Sage Science) was performed only for h^A genomic DNA used in 10x linked-read library preparation.

493 Preparation of Gel Bead-in-Emulsions (GEMs) for each 10x Genomics Chromium

494 library was performed using 0.5−0.6 ng of HMW genomic DNA according to the

495 manufacturer's instructions. The prepared GEMs were quality checked using Qubit dsDNA

496 HS kit (Thermo Fisher) and Bioanalyzer (Agilent), and processed with Chromium Controller

497 (10x Genomics). The constructed DNA libraries were quality checked again in the same way.

498 Sequencing of the libraries was performed in the Hiseq X ten (Illumina) platform (1

499 library/lane) at Macrogen. In total, we generated 66.9 Gb, 64.6Gb, 64.9 Gb and 60.3 Gb of

500 raw reads for linked-read sequencing (h^C, h^A, *h*, and *C. septempunctata*, respectively).

501 For library preparation for PacBio system, ten to eleven pupae (h^C [NT6 strain] and

502 *h* [NT8 strain]) or adults (h^A [F2 adult progenies in genetic cross $h^A \times h^{Sp}$]) were used.

503 The libraries were prepared according to the 20-kb Template Preparation Using

504 BluePippin™ Size-Selection System (Sage Science, MA, USA). In total, 4.31 Gb, 4.92 Gb

505 and 4.44 Gb of insert sequences (approximately10 × coverage of the genome, assuming a

506 genome size of 423 Mb) were obtained from 4 to 5 SMRT cells for h^A, h^C and *h*, respectively.

507

508 ***De novo* assembly of 10x linked-reads**

509 For 10x linked-reads libraries of four samples (three *H. axyridis* and one *C.*

510 *septempunctata*), Supernova (version 2.0.0)[61] was executed with default parameters except

511 for the maximum number of used reads (the --maxreads option) to obtain the optimum

512 coverage depth for Supernova (56×). For each sample, the value for --maxreads was

513 determined as follows:

514 (1) Barcode sequences in raw linked-reads were excluded using "longranger basic" command

515 of Long Ranger[62] (version 2.1.2), resulting in "barcoded.fastq" file.

516 (2) Adaptor sequences and low-quality regions in "barcoded.fastq" were trimmed using

517 Platanus_trim (version 1.0.7) with default parameters.

518 (3) 32-mers in the trimmed reads were counted by Jellyfish[63] (version 2.2.3) using the

519 following two commands and options:

520 $ jellyfish count -m 32 -s 20M -C -o out.jf barcoded_1.trimmed barcoded_2.trimmed

521 $ jellyfish histo -h 1000000000 -o out.histo out.jf

522 In summary, all 32-mers in both strands (-C) were counted and distribution of the number of

523 occurrences without upper limit of occurrences (-h 1000000000).

(4) The haploid genome size was estimated using the custom Perl script. For the distribution of the number of 32-mer occurrences ("out.histo"), the number of occurrences corresponding to a homozygous peak was detected, and the total number of 32-mers was divided by the homozygous-peak-occurrences. Here, 32-mers whose occurrences were small (< the number of occurrences corresponding to the bottom between zero and heterozygous peak) were excluded for the calculation to avoid the effect from sequencing errors.

(5) The values for --maxreads were calculated as follow:

estimated-haploid-genome-size / mean-read-length-of-barcoded.fastq × 56

As a result, we obtained the scaffolds including the genes surrounding *H. axyridis-pannier* (h^C [NT6], 2.74 Mb; h^A [F2 hybrid], 1.42 + 1.61 Mb; h^C [NT8] 2.79 Mb), and homologous regions in *C. septempunctata* (haplotype 1, 10.16 + 2.41 Mb; haplotype 2,

10.13 + 2.44 Mb). We used those sequences for the downstream *in silico* analyses. We assessed the completeness of the genome assembly using BUSCO[62] (version 3.0.2, Insecta dataset [1,658 orthologs]).

Gap filling of the genomic scaffolds at the *pannier* locus

Concerning the genome assemblies of *H. axyridis*, we used minimap2[64] (ver. 2.9) and PBjelly[65] (ver. PBSuite_15.8.24) software to fill gaps around the *pannier* locus. In each genome of three strains of *H. axyridis*, we first mapped PacBio reads to the assembly generated from the 10x linked-reads using minimap2. Then, we chose PacBio reads mapped to the scaffold containing *pannier* gene. These PacBio reads were subjected to gap-filling of

the scaffold with PBjelly. We obtained gap-free nucleotide sequences spanning the entire *pannier* locus and the upstream intergenic regions.

Concerning the genome assembly of *C. septempunctata*, there was a single gap estimated to be 15kb long by Supernova programme in the 1st intron of *pannier* locus. We handled this gap region as repeated N, and included it in the downstream *in silico* analyses.

Validation of the *pannier*-locus scaffold of Platanus2 for the *H. axyridis* F2-3 sample

For the *H. axyridis* F2-3 sample, trimmed reads of the 15 kbp-mate-pair library were mapped to the consensus scaffold set of Platanus2 using BWA-MEM[66] (version 0.7.12-r1039) with default parameters. Next, a consensus scaffold corresponding to *pannier*-locus was

segmented into 2 kbp-windows, and links between windows (the number of mate-pairs ≥ 3) were visualised by Circos[67] (version 0.69-6).

Additional comment for the editor and reviewers.

Although Platanus2 is not open currently, we plan to release the same version (2.0.0) of Platanus2 at the web site of Platanus (http://platanus.bio.titech.ac.jp/) by the publication of this study.

Resequencing of *Harmonia* genome for choosing appropriate strains for Genome Wide Association Study (GWAS)

565 Genomic DNA was extracted from each of h^C (F6 strain), h^A (NT3 strain) and h^{Sp}

566 (CB-5 strain), and used to create Illumina libraries using TruSeq Nano DNA Sample

567 Preparation Kit (Illumina) with insert size of approximately 400 bp. These libraries were

568 sequenced on the Illumina HiSeq 1500 using a 2 x 106-nt paired-end sequencing protocol,

569 yielding 84.7M paired-end reads. SNP site identification was conducted basically according

570 to GATK Best Practice[68] (ver. August 7 2015). After trimming adaptor sequences with

571 Cutadapt software (ver. 1.9.1) [69], the sequence data were mapped to the *de novo* genome

572 assembly data using bwa software (ver.0.7.15, BWA-MEM algorithm) [66]. Sequences and

573 alignments with low quality were filtered using Picard[70] tools (ver. 2.7.1) and GATK[68]

574 software (ver. 3.6 and 3.7), and 73,4443 SNP markers in the strains were identified. The most

distantly related strains (h^C [F6 strain] and h^A [NT3 strain]) were selected by performing phylogenetic analysis using SNPhylo[71] (Version: 20140701).

Comparison of the scaffolds at the *pannier* locus among ladybird beetles

For each pair of scaffolds, we constructed dot plots by performing pairwise-alignment using "nucmer" programme in the MUMmer package[72] (version 3.1). The options of nucmer were as follows: (1) *H. axyridis* vs. *H. axyridis*, Default parameters; (2) *H. axyridis* vs. *C. septempunctata*; '-l 8 -c 20'. Alignment results (delta file) were input into "mummerplot" programme to generate dot plots. Note that resultant gnuplot scripts resulted from mummerplot were edited for visualization.

585 We also visualised the homology and structural differences between the 700

586 kb-genomic region including *pannier* using Easyfig[73] (ver. 2s2.2). Short BLAST[47] hit

587 fragments less than 500 bp, and putative short repeat sequences less than 1250 bp, which

588 showed more than two BLAST hit blocks within the 700 kb region, were filtered using a

589 custom Perl script. Exon-intron structures of putative genes in the 700 kb regions were

590 obtained using Exonerate[74] (ver. 2.2.0) with the options '-m est2genome --showvulgar yes

591 --ryo ">%qi length=%ql alnlen=%qal\n>%ti length=%tl alnlen=%tal\n" --showtargetgff yes

592 --showalignment no --score 2000'. cDNA sequences cloned by RT-PCR or predicted by

593 RNA-seq were used as queries. If a single cDNA unit was split into multiple fragments, we

594 merged the fragments by performing exonerate search again using the cDNA sequences

595 whose subsequences were substituted by the genomic hit fragments in the first exonerate

search as a query. Exonerate output files were converted to the GFF3 format using our bug-fixed version of the process_exonerate_gff3.pl[75] with the option '–t EST'. The GFF3 file and a FASTA format file of each scaffold were converted to a GENBANK format file using EMBOSS Seqret[76] programme (ver. 6.6.0.0) with the options '-fformat gff -osformat genbank '. The GENBANK format files corresponding to the 700 kb genomic sequences surrounding *pannier*, which were used as input files of Easyfig, were extracted using the Genbank_slice.py[77] Python script (ver. 1.1.0).

Flexible ddRAD-seq

We newly constructed a flexible ddRAD-seq library preparation protocol to facilitate high-throughput ddRAD-seq analyses at low cost. We designed all enzymatic

607 reactions to be completed sequentially without DNA purification in each step to make the

608 procedures simple. In addition, we designed 96 sets of indexed and forked sequencing

609 adaptors compatible with Illumina platform sequencers (Supplementary Table 10).

610 Briefly, 100 ng of genomic DNA was first double-digested with 15 U of EcoRI-HF

611 and 15 U of HindIII-HF in 20 µl of NEB CutSmart Buffer (New England Biolabs) at 37°C for

612 2 hours. 15 µl of the digested DNA, 4 pmole of adaptor DNA, 10 µmole of ATP, 400 Units of

613 T4 DNA ligase were mixed in 20 µl, incubated at 22 °C for 2 hours, and denatured at 65°C for

614 10 minutes. Ligated library DNA fragments were purified with Agencourt AMPureXP

615 (Beckman Coulter) according to the manufacturer's instructions. Library DNA fragments

616 ranging form 300 bp to 500 bp were size-selected with Pippin Prep (Sage Science).

617 Concentration of each library DNA was quantified using KAPA Library Quantification Kits

(Roche) according to the manufacturer's instructions. Sequence data were obtained by applying 96 DNA libraries to a single lane of Hiseq 1500 (Illumina).

Linkage map construction and GWAS

A single h^C male (F6 strain) and a single virgin h^A female **(**NT3 strain) were crossed, and the obtained F1 progenies were backcrossed with the F0 male (h^C, F6 strain). Finally, 183 adult F2 progenies (h^C = 80, h^C/h^A = 103) and 2 F0 adults were collected for RAD-sequence analysis, and stored at −30°C until use. Genomic DNAs were extracted individually using an automatic nucleic acid extractor (PI-50α, KURABO). Briefly, each frozen ladybird beetle and a zirconia bead were transferred to 2 ml plastic tube (Eppendorf) on ice. Immediately, 250 µl of cold lysis buffer including Proteinase K and RNase, but not SDS was added to the

629 sample, and the tubes were vigorously shaken with a tissue grinder (Tissue Lyser LT, Qiagen)

630 at 3000 rpm for 1 min. Then, 250 µl of lysis buffer including SDS was added to each crushed

631 sample, and processed with the programme for DNA extraction from mouse tail, according to

632 the manufacturer's instructions. The extracted genomic DNA was diluted in 30 µl of TE

633 buffer. The DNA concentration of each sample was quantified using Qubit dsDNA BR Kit

634 according to the manufacture's instructions (Thermo Fisher Scientific). We performed

635 flexible ddRAD-seq using these genomic DNA samples. 0.6−6.0 millions (mean = 2.0

636 millions) of 106 bp paired-end reads per sample were generated using two lanes of Hiseq

637 1500 (Illumina) following the methods in the User Guide.

638 Mapping and polymorphic site calling were conducted as described in the

639 resequencing analysis above except that the procedure for filtering duplicated reads using

Picard was eliminated because we did not amplified DNA library by PCR. Count data at each SNP sites were extracted from the obtained vcf file using vcf_to_rqtl.py script in rtd software[78] with the options '5.0 80'. To avoid programme errors, we modified the script to skip the read depth data (DP) including characters in the GATK vcf file.

We constructed a linkage map using R/qtl[79] (version 1.42.8) and R/ASMap[80] (version 1.0.2) R[81] packages according to the QTL mapping workflow for BC1 population of *Jaltomata*[81]. Using the obtained csv file as an input, we eliminated the polymorph sites that behaved as located on the X chromosome. In addition, individuals with low mapping quality, and marker sites with low quality or highly distorted segregation patterns were eliminated as well. Finally, 4419 markers sites, and 177 F2 individuals were used. The linkage map was initially constructed with mstmap programme (R/ASMap) with the options 'dist.fun =

"kosambi", p.value = 1e-25', and highly linked linkage groups were merged manually. The markers consistently incongruent with neighboring markers were eliminated using correctGenotypes.py[82] script with the options '-i csvr -q 0.1 -t 4.0'.

Genome wide association study was conducted using calc.genoprob programme (R/qtl) with the options 'step=1, error.prob=0.001' and scanone programme (R/qtl) with the option 'model="binary"'. The result data were visualised with plot programme in R/ASMap.

Genetic association study focusing on the *pannier* locus

In addition to the genetic cross in GWAS ($h^C \times h^A$), two independent crosses ([$h \times h^C$] and [$h^A \times h^{Sp}$]) were performed.

In the former cross, a single *h* male (D-5 strain) and a single virgin h^C female (F2-3-B strain) were crossed, and the obtained F1 progenies were sibcrossed. Finally, 80 F2 adult progenies (h^C = 30, h^C/h = 34, and h = 16) were collected for genotyping, and stored at −30°C until use. In the latter cross, a single h^{Sp} male (CB-5 strain) and a single virgin h^A female (NT3 strain) were crossed, and the obtained F1 progenies were sibcrossed. Finally, 273 F2 adult progenies (h^{Sp} = 103, h^{Sp}/ h^A = 80, and h^A = 90) were collected for genotyping, and stored at −30°C until use.

Genomic DNA was extracted individually using the automatic nucleic acid extractor (PI-50α, KURABO) as described in the previous section, and diluted to approximately 100 ng/μl. We searched for genotyping markers by amplifying and sequencing the intronic region of the genes surrounding *pannier* with PCR. The individual

672 PCR was performed using approximately 100 ng of genomic DNA and Q5 DNA polymerase

673 (New England Biolabs) with 45 cycles. The primers used, the markers identified and the

674 typing results are summarised in Supplementary Table 2.

675

676 **RNA-seq analysis**

677 The total RNA extraction procedure for RNA-seq is essentially the same as that for

678 the gene expression analysis in the presumptive red and black regions by RT-PCR. The same

679 strain used for *de novo* genome sequencing (F2-5 strain, h^C) was used. In total, 12 samples (2

680 colours [Black/Red] × 2 developmental stages [24h AP/72h AP] × 3 biological replicates)

681 were prepared for RNA-seq analysis. Two fragments of bored epidermis from left and right

682 elytra were collected as a single sample in each condition. All total RNA extracted from each

683 sample (12 ng to 158 ng) using RNeasy Mini Kit (QIAGEN) and QIAcube (QIAGEN) was

684 used for each cDNA library preparation. RNA-seq library preparation was performed using

685 the SureSelect strand specific RNA library prep kit (Agilent) according to the manufacturer's

686 instructions. Briefly, mRNA was purified using Oligo-dT Microparticles. The strand-specific

687 RNA-seq libraries were prepared using dUTP and Uracil-DNA-Glycosylase. The libraries

688 and its intermediates were purified and size-fractionated by AMPure XP (Beckman Coulter).

689 For quality check and quantification of the RNA-seq libraries, we employed 2100

690 Bioanalyser and DNA 7500 kit (Agilent). 100 bp paired end read RNA-seq tags were

691 generated using the Hiseq 2500 (Illumina) following the methods in the User Guide.

692 In advance of reference mapping, adaptor and poly-A sequences were trimmed

693 from raw RNA-seq reads by using Cutadapt (ver. 1.9.1)[69]. Low-quality reads were also

694 filtered out by a custom Perl script as described previously[83]. The preprocessed RNA-seq

695 reads were mapped to the reference *H. axyridis* genome sequences using TopHat2[84] (ver.

696 2.1.0) with default parameters, and assembled by Cufflinks[85] (ver. 2.2.1) with the -u option in

697 each sample. All predicted transcript units and all loci from different samples were merged

698 by Cuffmerge in the Cufflinks suite. The RNA-seq read pairs (fragments) mapped to each

699 predicted transcript unit and locus were counted using HTSeq[86] (ver. 0.6.1) with the options

700 '-s no -t exon -i transcript' and '-s no -t exon -i locus', respectively. The downstream statistic

701 analyses were performed using edgeR[87, 88] package (ver. 3.16.5). The raw RNA-seq fragment

702 counts were normalised by the trimmed mean of M-values (TMM) method. Fold change

703 between black and red regions in each stage and its statistical significance (FDR) were

704 calculated. The mean fold changes of the genes in the scaffolds including the *h* locus

705 candidate region were visualised with IGV[89, 90] software (ver. 2.3.88).

706

707 **Comparison of the sizes of the upper noncoding regions of *pannier* among**

708 **holometabolous insects**

709 The holometabolous insects, whose genomic sequences are well assembled at the

710 *pannier* locus to the extent that at least the two paralogous GATA transcription factor genes,

711 *GATAe* and *serpent*, are included in the same scaffold, were selected for comparison.

712 Concerning Coleoptera, genomic sequences were collected from the Genome

713 database at NCBI[91] (GCA_000002335.3, Tcas5.2; GCA_001937115.1, Atum_1.0;

714 GCA_000390285.2, Agla_2.0; GCA_000648695.2, Otau_2.0; GCA_001412225.1,

715 Nicve_v1.0; GCF_000699045.1, Apla_1.0; GCA_002278615.1, Pchal_1.0) and

716 Fireflybase[92] (*Photinus pyralis* genome 1.3, *Aquatica lateralis* genome 1.3). Concerning

717 holometabolous insect other than Coleoptera, genomic information at Hymenoptera Genome

718 Database[93] (Hymenoptera) (GCF_000002195.4, Amel_4.5; GCF_000217595.1,

719 Lhum_UMD_V04; GCF_000002325.3, Nvit_2.1), Lepbase[94] (Lepidoptera, butterfly;

720 Danaus_plexippus_v3_scaffolds), SilkBase[95] (Lepidoptera, silk moth; Genome assembly

721 [Jan.2017]), Flybase[96] (Diptera, *Drosophila*, dmel_r6.12_FB2016_04), and the Genome

722 database at NCBI[91] (Diptera, mosquitos; GCA_000005575.1, AgamP3; GCA_002204515.1,

723 AaegL5.0) (Lepidoptera, moth; GCA_002192655.1, ASM219265v1) were utilised. We

724 performed BLAST[47] search (TBLASTN, ver. 2.2.26) using the amino acid sequence of *H.*

725 *axyridis* Pannier as a query, and identified *pannier* orthologs by focusing on the top hits, and

the conserved synteny of the three paralogous GATA transcription factor genes, in which *serpent*, *GATAe*, and *pannier* are tandemly aligned in this order from the 5' to 3' direction. The sizes of the upper noncoding region of *pannier* were estimated by calculating the difference between the coordinates of the 3' end of BLAST hit region of *GATAe* and the 5' hit region of *pannier*. Traces of translocation or insertions between the paralogous GATA genes were surveyed by looking into the annotations between the *GATAe* and *pannier* loci. If a noncoding exon 1 was annotated at the pannier locus, the size of the first intron was also calculated as well.

Motif enrichment analysis of the upper noncoding regions of *pannier*

736 To search for the *Drosophila* known DNA-binding sites at *pannier*, 1,139 DNA

737 motifs were retrieved from the JASPAR[97] database using the MotifDb[98] R package.

738 Concerning the SD-binding motif, the position weight matrix (PWM) scores were calculated

739 using the 2,557 ChIP-seq peaks in *Drosophila* genome obtained by Ikmi et al.[99] and the

740 RSAT peak-motifs programme[100]. The nucleotide sequences at the three upper noncoding

741 regions of *pannier* (the upper intergenic region, the upstream half of the 1st intron, and the

742 down stream half of the 1st intron) were collected by forging BSgenome[101] data packages

743 using our *H. axyridis* and *C. septempunctata* genome sequences, and by retrieving the

744 sequences using coordinate information obtained for annotation in Fig. 5a and the

745 GenomicFeatures[102] R package. We here defined the upstream half of the 1st intron as the

746 region including all of the traces of inversions or corresponding sequences shared among

different *h* alleles in *H. axyridis*. GRange objects were generated using the coordinate information obtained for annotation in Fig. 5a. Motif enrichment was quantified using the PWMEnrich[103] R package. As control background genomic regions for the upper intergenic regions of *pannier*, we used 2 kb promoter sequences of 11,279 genes to which RNA-seq reads were mapped (*H. axyridis* genome assembly version 1), and not located within the 10 kb from the end of the genomic scaffolds. As control background genomic regions for the upstream and downstream regions of the 1st intron of *pannier*, we used 2 kb sequences at the 5' and 3' end of the 1st introns, which are more than 2 kb long, and without gaps (2825 and 2810 sequences, respectively). Since there is no reliable gene annotation data for the *C. septempunctata* genome, we used the *H. axyridis* genomic background sequences. Each motif enrichment score, which is related to average time that transcription factors spend in

758 binding to a DNA sequence[104], was calculated using default parameter of PMWEnrichment.

759 *br* and *da* DNA-binding motifs were excluded from the analysis.

760

761 **Molecular phylogenetic analysis**

762 Concerning the conserved regions in the upper half of the 1st intron of *pannier*,

763 nucleotide sequences were collected from the BLAST hits obtained to construct Fig. 5a. The

764 collected BLAST hit sequences were arranged in the same directions using a custom Perl

765 script, and aligned using MAFFT[105] (ver. 7.222). We concatenated three alignment blocks,

766 and manually excluded GAP sites and seemingly nonhomologous sites in the alignment

767 (Supplementary Data 1). Concerning the coding region, nucleotide sequences of the cloned

768 cDNAs were aligned, and trimmed in the same way (Supplementary Data 1).

769 The maximum-likelihood (ML) phylogenetic trees were constructed using

770 RAxML[106] (ver. 8.0.0) with the options '--maxiterate 1000 --localpair --clustalout'. We

771 determined appropriate models of sequence evolution under the AICc4 criterion using

772 Kakusan4[107]. 100 replicates of shotgun search for the likelihood ratchet were performed.

773 Nodal support was calculated by bootstrap analyses with 1000 replications.

774

775

776 **Data availability**

777 The DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory

778 (EMBL)/GenBank accession numbers for the sequences reported in this article are

779 LC269047−LC269055 for *Ha-pnr*, LC269056 for *Cs-pnr* and LC269057 for *Cs-rp49*. The

780 genomic sequencing data, resequencing data and the RNA-seq data were deposited in the

781 DNA Data Bank of Japan (DDBJ) under the accession number DRA002559, X (under

782 curation) & DRA006068, and DRA005777, respectively.

References

1. Vandenberg, N. J. Coccinellidae Latreille, 1807 in *American Beetles*. *Vol.2. Polyphaga: Scarabaeoidea through Curculionoidea*. (ed. Arnett, R. H. & Thomas, M.C.) 371–389 (Boca Raton, CRC Press, 2002).
2. Daloze, D., Braekman, J. C. & Pasteels, J. M. Ladybird defence alkaloids: Structural, chemotaxonomic and biosynthetic aspects (Col.: Coccinellidae). *Chemoecology* **5–6,** 173–183 (1994).
3. Glisan King, A. & Meinwald, J. Review of the defensive chemistry of coccinellids. *Chem. Rev.* **96,** 1105–1122 (1996).
4. Dobzhansky, T. Die geographische und individuelle Variabilität von *Harmonia axyridis* Pallas in ihren Wechselbeziehungen. *Biol. Zentralbl.* **44,** 401–421 (1924).
5. Tan, C. C. & Li, J. C. Inheritance of the elytral color patterns of the lady-bird beetle, *Harmonia axyridis* Pallas. *Amer. Nat.* **68,** 252–265 (1934).
6. Hosino, Y. Genetical Studies of the lady-bird beetle, *Harmonia axyridis* PALLAS (Report II). *Japan. J. Genet.* **12,** 307–320 (1936).
7. Komai, T. Genetics of Ladybeetles. *Adv. Genet.* **8,** 155–188 (1956).
8. Thompson, M. J. & Jiggins, C. D. Supergenes and their role in evolution. *Heredity* **113,** 1–8 (2014).
9. Majerus, M. E. N. *Ladybirds*. (Harper Collins Publishers, London, 1994).
10. Tan, C. C. Mosaic dominance in the inheritance of color patterns in the lady-bird beetle, *Harmonia Axyridis*. *Genetics* **31,** 195–210 (1946).
11. Bezzerides, A. L., McGraw, K. J., Parker, R. S. & Husseini, J. Elytra color as a signal of chemical defense in the Asian ladybird beetle *Harmonia axyridis*. *Behav. Ecol. Sociobiol.* **61,** 1401–1408 (2007).
12. True, J. R., Kopp, A. & Carroll. S. B. Prophenoloxidase as a reporter of gene expression in *Drosophila*. *BioTechniques* **30,** 1008–1009 (2001).
13. Carroll, S. B. Developmental regulatory mechanisms in the evolution of insect diversity. *Development* **1994,** 217–223 (1994).

14. Abouheif, E. & Wray, G. A. Evolution of the gene network underlying wing polyphenism in ants. *Science* **297,** 249–252 (2002).
15. Tomoyasu, Y., Arakane, Y., Kramer, K. J. & Denell, R. E. Repeated co-options of exoskeleton formation during wing-to-elytron evolution in beetles. *Curr. Biol.* **19,** 2057–2065 (2009).
16. Strigini, M. & Cohen, S. M. Formation of morphogen gradients in the *Drosophila* wing. *Semin. Cell Dev. Biol.* **10,** 335–344 (1999).
17. Klein, T. Wing disc development in the fly: the early stages. *Curr. Opin. Genet. Dev.* **11,** 470–475 (2001).
18. Gómez-Skarmeta, J. L., Campuzano, S. & Modolell, J. Half a century of neural prepatterning: the story of a few bristles and many genes. *Nat. Rev. Neurosci.* **4,** 587–598 (2003).
19. Kuwayama, H., Yaginuma, T., Yamashita, O. & Niimi, T. Germ-line transformation and RNAi of the ladybird beetle, *Harmonia axyridis*. *Insect Mol. Biol.* **15,** 507–512 (2006).
20. Ramain, P., Heitzler, P., Haenlin, M. & Simpson, P. *pannier*, a negative regulator of *achaete* and *scute* in *Drosophila*, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. *Development* **119,** 1277–1291 (1993).
21. Garcia-Garcia, M. J., Ramain, P., Simpson, P. & Modolell, J. Different contributions of *pannier* and *wingless* to the patterning of the dorsal mesothorax of *Drosophila*. *Development* **126,** 3523–3532 (1999).
22. Tomoyasu, Y., Ueno, N. & Nakamura, M. The Decapentaplegic morphogen gradient regulates the notal *wingless* expression through induction of *pannier* and *u-shaped* in *Drosophila*. *Mech. Dev.* **96,** 37–49 (2000).
23. Sato, M. & Saigo, K. Involvement of *pannier* and *u-shaped* in regulation of Decapentaplegic-dependent wingless expression in developing *Drosophila* notum. *Mech. Dev.* **93,** 127–138 (2000).
24. Halder, G. *et al.* The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in *Drosophila*. *Genes Dev.* **12,** 3900–3909 (1998).

25. Landin-Malt, A., Benhaddou, A., Zider, A. & Flagiello, D. An evolutionary, structural and functional overview of the mammalian TEAD1 and TEAD2 transcription factors. *Gene* **591,** 292–303 (2016).

26. Blair, S. S. Compartments and appendage development in *Drosophila*. *BioEssays* **17,** 299–309 (1995).

27. Irvine, K. D. & Rauskolb, C. Boundaries in Development: Formation and Function. *Annu. Rev. Cell Dev. Biol.* **17,** 189–214 (2001).

28. Blair, S. S. Lineage compartments in *Drosophila*. *Curr. Biol.* **13,** R548–R551 (2003).

29. Mann, R. S. & Carroll, S. B. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* **12,** 592–600 (2002).

30. Duncan, D. M., Burgess, E. A. & Duncan, I. Control of distal antennal identity and tarsal development in *Drosophila* by *spineless–aristapedia*, a homolog of the mammalian dioxin receptor. *Genes Dev.* **12,** 1290–1303 (1998).

31. Rodríguez, D. del Á., Terriente, J., Galindo, M. I., Couso, J. P. & Díaz-Benjumea, F. J. Different mechanisms initiate and maintain *wingless* expression in the *Drosophila* wing hinge. *Development* **129,** 3995–4004 (2002).

32. de Celis, J. F. Pattern formation in the *Drosophila* wing: The development of the veins. *BioEssays* **25,** 443–451 (2003).

33. Blair, S. S. Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. *Annu. Rev. Cell Dev. Biol.* **23,** 293–319 (2007).

34. Martín, M., Ostalé, C. M. & de Celis, J. F. Patterning of the *Drosophila* L2 vein is driven by regulatory interactions between region-specific transcription factors expressed in response to Dpp signalling. *Development* **144,** 3168–3176 (2017).

35. Yao, T.P. *et al.* Functional ecdysone receptor is the product of *EcR* and *Ultraspiracle* genes. *Nature* **366,** 476–479 (1993).

54 36. Bai, J., Uehara, Y. & Montell, D. J. Regulation of invasive cell behavior by Taiman, a *Drosophila* protein
55 related to AIB1, a steroid receptor coactivator amplified in breast cancer. *Cell* **103,** 1047–1058 (2000).
56 37. Robertson, J. A. *et al.* Phylogeny and classification of Cucujoidea and the recognition of a new
57 superfamily Coccinelloidea (Coleoptera: Cucujiformia). *Syst. Entomol.* **40,** 745–778 (2015).
58 38. Encyclopedia of Life, Coccinellini, Available from http://eol.org/pages/8947605/overview, (Accessed 21
59 May 2018).
60 39. Sasaji, H. Biosystematics on *Harmonia axyridis*-complex, *Mem. Fac. Educ., Fukui Univ., Ser. II, Nat.*
61 *Sci.*, **30,** 61-79 (1980)
62 40. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly
63 mimicry. *Nature* **477,** 203–206 (2011).
64 41. Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*
65 **493,** 664–668 (2013).
66 42. Kunte, K. *et al. doublesex* is a mimicry supergene. *Nature* **507,** 229–232 (2014).
67 43. Nishikawa, H. *et al.* A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat.*
68 *Genet.* **47,** 405–409 (2015).
69 44. Merrill, R. M. *et al.* The diversification of *Heliconius* butterflies: what have we learned in 150 years? *J.*
70 *Evol. Biol.* **28,** 1417–1438 (2015).
71 45. Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295,** 1669–1678 (2002).
72 46. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323,** 746–751 (2009).
73 47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J.*
74 *Mol. Biol.* **215,** 403–410 (1990).
75 48. Zhang, S.Q. *et al.* Evolutionary history of Coleoptera revealed by extensive sampling of genes and
76 species. *Nat. Commun.* **9,** 205 (2018).
77 49. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and

78 divergence times. *Mol. Biol. Evol.* **34,** 1812–1819 (2017).
79 50. Niimi, T., Kuwayama, H. & Yaginuma, T. Larval RNAi applied to the analysis of postembryonic
80 development in the ladybird beetle, *Harmonia axyridis*. *J. Insect Biotech. Sericol.* **74,** 95–102 (2005).
81 51. Kawamoto, T. & Shimizu, M. A method for preparing 2- to 50-μm-thick fresh-frozen sections of large
82 samples and undecalcified hard tissues. *Histochem. Cell Biol.* **113,** 331–339 (2000).
83 52. Davis, M. W. *ApE*. Available from http://biologylabs.utah.edu/jorgensen/wayned/ape/, (Accessed: 21
84 May 2018)
85 53. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for
86 bigger datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).
87 54. Hofmann, K. & Baron, M. *Boxshade*. Available from
88 http://www.ch.embnet.org/software/BOX_form.html, (Accessed: 21 May 2018)
89 55. Ando, T., Kojima, T. & Fujiwara, H. Dramatic changes in patterning gene expression during
90 metamorphosis are associated with the formation of a feather-like antenna by the silk moth, *Bombyx mori*.
91 *Dev. Biol.* **357,** 53–63 (2011).
92 56. Kajitani, R. *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome
93 shotgun short reads. *Genome Res.* **24,** 1384–1395 (2014).
94 57. Yang, X., Chockalingam, S. P. & Aluru, S. A survey of error-correction methods for next-generation
95 sequencing. *Brief. Bioinform.* **14,** 56–66 (2013).
96 58. Platanus trim, Available from http://platanus.bio.titech.ac.jp/pltanus_trim, (Accessed 21 May 2018)
97 59. Platanus2, Available from http://platanus.bio.titech.ac.jp/, (in preparation)
98 60. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and
99 phylogenomics. *Mol. Biol. Evol.* **35,** 543–548 (2018).
00
01 61. Weisenfeld, N., Kumar, V., Shah, P., Church, D. & Jaffe, D. Direct determination of diploid genome

sequences. *Nat. Biotech.* **27,** 757–767 (2017).

62. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Nat. Commun.* **27,**764–70 (2011).

63. Long Ranger, Available at https://support.10xgenomics.com/genome-exome/software/downloads/latest, (Accessed: 21 May 2018)

64. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* doi:10.1093/bioinformatics/bty191 (2018)

65. English, A. C. *et al.* Mind the gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. *PLOS ONE* **7,** e47768 (2012).

66. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

67. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).

68. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., 2002).

69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

70. Picard, Available from http://broadinstitute.github.io/picard/, (Accessed: 21 May 2018)

71. Lam, H.M. *et al.* Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42,** 1053–1059 (2010).

72. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5,** R12 (2004).

73. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27,** 1009–1010 (2011).

74. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).

75. process_exonerate_gff3.pl, Available from https://github.com/hyphaltip/genome-scripts/blob/master/data_format/process_exonerate_gff3.pl (Accessed: 21 May 2018)

76. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).

77. Healey, J. Genbank_slice.py, Available from https://github.com/jrjhealey/bioinfo-tools/blob/master/Genbank_slicer.py, (Accessed: 21 May 2018)

78. Peterson, B. rtd, Available from https://github.com/brantp/rtd, (Accessed: 21 May 2018)

79. Arends, D., Prins, P., Jansen, R. C. & Broman, K. W. R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* **26,** 2990–2992 (2010).

80. Taylor, J. & Butler, D. R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. *J. Stat. Soft.* **79,** 1-29 (2017).

81. R Core Team. R: A language and environment for statistical computing. Available from https://www.R-project.org/ (Accesed: 21 May 2018)

82. Gibson, M., Jaltomata QTL mapping. Availabe from https://github.com/gibsonMatt/jaltomataQTL, (Accessed: 21 May 2018)

83. Ohyanagi, H. *et al.* Plant Omics Data Center: An Integrated Web Repository for Interspecies Gene Expression Networks with NLP-Based Curation. *Plant Cell Physiol.* **56,** e9–e9 (2015).

84. Kim, D. *et al*. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

85. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotech.* **28,** 511–515 (2010).

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

87. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

88. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).

89. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotech.* **29,** 24–26 (2011).

90. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).

91. NCBI Genome, Available from https://www.ncbi.nlm.nih.gov/genome/, (Accessed 21 May 2018)

92. Fireflybase, Available from http://www.fireflybase.org/, (Accessed 21 May 2018)

93. Hymenoptera Genome Database, Available from http://hymenopteragenome.org/, (Accessed 21 May 2018)

94. Lepbase, Available from http://lepbase.org/, (Accessed 21 May 2018)

95. SilkBase, Available from http://silkbase.ab.a.u-tokyo.ac.jp/cgi-bin/index.cgi, (Accessed 21 May 2018)

96. Flybase, Available from http://flybase.org/, (Accessed 21 May 2018)

97. JASPAR, Available from http://jaspar.genereg.net/, (Accessed 21 May 2018)

98. Shannon, P. MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs. Bioconductor. http://bioconductor.org/packages/MotifDb/ (2017).

99. Ikmi, A. *et al.* Molecular evolution of the Yap/Yorkie proto-oncogene and elucidation of its core transcriptional program. *Mol. Biol. Evol.* **31,** 1375–1390 (2014).

100. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40,** e31–e31 (2012).

101. Pagès, H. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs.

R package version 1.46.0. (2017). Available from: http://bioconductor.org/packages/BSgenome/. (Accessed: 7 April 2017)

102. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Comp. Biol.* **9,** e1003118 (2013).

103. Stojnic, R., & Diez, D. PWMEnrich: PWM enrichment analysis. R package version 4.14.0. (2015) Available from: http://bioconductor.org/packages/release/bioc/html/PWMEnrich.html. (Accessed: 21 May 2018)

104. Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16,** 16–23 (2000).

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

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

107. Tanabe, A. S. Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour*. **11,** 914–921 (2011).

989 **Supplementary Information**

990 Below are supplementary data related to our analyses. Included are one table on our

991 small-scale RNAi screen, one table on statistics of genome assemblies, five tables on

992 DNA primers designed for amplification and sequencing, and five figure legends for the

993 Supplementary data.

Supplementary Table 1 | The 1st smal-scale RNAi screening of wing/body wall patterning genes

gene name	phenotype in elytra	tested alleles
araucan	no phenotype	*h, hC, hSp, hA*
aristaless	no phenotype	*h*
apterous	wrinkled elytra	h^{Sp}, h^{A}
blistered	lethal	h^{Sp}, h^{A}
Cubitus interruptus	pupal lethal	h^{A}
Distal-less	no phenotype	*h*, h^{C}
decapentaplegic	smaller elytra	*h*, h^{A}
Epidermal growth factor receptor	lethal	h^{Sp}, h^{A}
pannier	transformation of black regions to red regions	*h*, h^{C}, h^{Sp}, h^{A}
wingless	no phenotype	*h*, h^{Sp}

997 **Supplementary Table 2 | Markers and Primers for gene association study on the *h* locus**

Marker/primer name	Target	Sequence (5'>3')	polymorphism (coordinate) between h-F0 and hC-F0, between hA-F0 and hSp-F0, and between hA-F0 and hC-F0	Recombinants' genotype	Cross	The *h* candidate regions	
mbl-F	scaffold898_cov129:657,001-657,297	AAAATGAGAGTGTAAGAAGA	A/G (657,138)	h/h-#27 (A/G heterozygote) hC/hC-#1 (A/G heterozygote)	h x h^C	< h	< h^C
mbl-R		CAAAGCAGAGAAACCTAACA			h x h^C		
DUS-F	scaffold898_cov129:1,560,979-1,561,277	CATTTCAACATCATCTTTCTG	A/T (1,561,052)	no recombinant	h x h^C	h	h^C
DUS-R		GAGACGGAGAGAGTGGAAC			h x h^C		
pnr-F	scaffold294_cov134:640,762-640,221	AAATTATTCTTACACAATTTTCTGG	600bp/500_or_480bp	no recombinant	h x h^C	h	h^C
pnr-R		AGGACACGTTTGGAAGAATAAGATAGT			h x h^C		
mus201-F	scaffold267_cov168:390,452-390,521	CCAAGAGGCTCAGGTAAATA	70bp/100bp	hC/hC-#14 (70bp/100bp heterozygote)	h x h^C	h	h^C >
mus201-R		ATTTCCATACAACACCAAAT			h x h^C		
Mink-F	scaffold99874_cov139:62,450-62,532	GAGTCTCCAAAATTACCAT	80bp/80_or_57bp	h/hC-#7 (57bp homozygote)	h x h^C	h >	
Mink-R		CAATGCTCTTCTGAAATAA			h x h^C		
Fam92_3'-F	scaffold294:205,919-205,999	GTCTGTCATTCTAGYGTATATGAAT	80bp/~60bp	hA/hSp-#41 (~60bp homozygote) hA/hSp-#49 (~60bp homozygote)	h^A x h^{Sp}	< h^A	< h^{Sp}
Fam92_3'-R		AGTTCGGTCAAGCAATTG			h^A x h^{Sp}		
sens-F	scaffold294:309,273_309,615	TCTTCATGGAGAGGTTTA	242bp/~250bp	no recombinant	h^A x h^{Sp}	h^A	h^{Sp}
sens-R		GMGTTTGTTCAAATTCATA			h^A x h^{Sp}		
srp_3'-F	scaffold267:650,101-650,598	ACTTTAACAACCCCGACR	~170bp/498bp	no recombinant	h^A x h^{Sp}	h^A	h^{Sp}
srp_3'-R		TACCATATTGAATCAACGTT			h^A x h^{Sp}		
srp_5'-F	scaffold267:579,180-579,337	AATAAGCATACAAACTTCCT	157bp/~160bp	hSp/hSp-#61 (157bp/~160bp heterozygote)	h^A x h^{Sp}	h^A	h^{Sp} >
srp_5'-R		ACTCACAATAGGTTTCTGAA			h^A x h^{Sp}		
kis_5'-F	scaffold267:465,373-465,708	AGCAATTTGTATGTATTCAG	~290bp/335bp	hSp/hSp-#61 (~290bp/335bp heterozygote)	h^A x h^{Sp}	h^A	
kis_5'-R		ATGTGAATTAAAACAAGGTG			h^A x h^{Sp}		
kis_3'-F	scaffold267:421,269-421,340	TTTTCTCGYTRGTATCAA	~60bp/71bp	hSp/hSp-#61 (~60bp/71bp heterozygote) hA/hSp-#9 (71bp homozygote)	h^A x h^{Sp}	h^A >	
kis_3'-R		CAAATTTTGAMACATGTTC			h^A x h^{Sp}		
RAD-tag_5'_breakpoint	scaffold294:466,936	- -	C/- (466,936)	hC/hC-#56 (C/- heterozygote) hA/hC-#1 (-/- homozygote)	h^A x h^C h^A x h^C	< h^A	< h^C

998

999

Supplementary Table 3 | Statistics of the genome assemblies

Columns labelled as “Complete”, “Duplicated & complete”, “Fragmented” and “Missing” are the results from the BUSCO programme.

Speceis	Sample	Assembler	Input data	Total (bp)	No. of sequences	N50 length (bp)	Gap rate (%)	Max length (bp)	Complete (%)	Duplicated & complete (%)	Fragmented (%)	Missing (%)
H. axyridis	h^C F2-3	Platanus2	pair end & mate pair reads	434,974,256	18,513	1,106,177	4.62	10,486,871	97.587	2.714	1.267	1.146
H. axyridis	h^C NT6	Supernova	10x linked-reads	581,452,270	50,767	100,132	10.24	8,301,291	95.476	9.288	2.654	1.87
H. axyridis	h^A F2 hybrid	Supernova	10x linked-reads	510,576,820	50,277	57,217	4.5	6,058,326	95.838	7.539	2.593	1.568
H. axyridis	*h* NT8	Supernova	10x linked-reads	557,776,217	44,316	157,105	11.07	15,325,500	95.356	7.358	2.774	1.87
C. septempunctata	MD-4	Supernova	10x linked-reads	514,983,385	55,574	67,752	9.12	12,752,393	96.019	5.79	2.292	1.689

1002 **Supplementary Table 6 | Primers for gene cloning**

Primer name	Target direction	Sequence (5'>3')
pannier-F	sense	ATIGAYTTYCARTTYGGIGA
pannier-R	antisense	GGYTTICKYTTICKIGTYTG
rp49-F	sense	ACIAARMAITTYATIMGICA
rp49-R	antisense	TGIGCIATYTCISCRCARTA
Ha-pannier-RACE-1	sense	GGGCAGGGAGTGCGTCAATTGTGGGGCC
Ha-pannier-RACE-2	sense	CCACCCCTCTGTGGAGGAGAGATGGTAC
Ha-pannier-RACE-3	antisense	GCCACAGGCGTTGCACACCGGTTCGCC
Ha-pannier-RACE-4	antisense	GATGCCGTCCTTGCGCATGGCCAGGGG
Cs-pannier-RACE-1	sense	AATTGCGGCACCAGGACGACGACGCTC
Cs-pannier-RACE-2	sense	AAGCTGCACGGCGTCAACAGGCCTCTG
Cs-pannier-RACE-3	antisense	ATCATCATGTTCTGGGCGTA
Cs-pannier-RACE-4	antisense	GTACGGGCTAAGTTCGGATC
Cs-pannier-5	sense	GATCCGAACTTAGCCCGTAC
Cs-pannier-6	antisense	GGGTCTGTTCATCCCATTCATCTTGTGG
Ha_pnr_intron_conserved1_F	sense	TCAGCRAATCTTCACATA
Ha_pnr_intron_conserved1_R	antisense	CTCCACTCGTTTATCTTAAT
Ha_pnr_intron_conserved2_F	sense	AGAGAAAAGAGACAASTTGA
Ha_pnr_intron_conserved2_R	antisense	AAAAGTRTTTSCTTCAGG
Ha_pnr_intron_conserved3_F	sense	AATGKATTCAAACCYCAGAC
Ha_pnr_intron_conserved3_R	antisense	MGRACGCTGAATGAAAGT

(I=inosine, K=G+T, M=A+C, R=A+G, S=C+G, Y=T+C)

1003 **Supplementary Table 7 | Primers for RT-PCR**

Primer name	Target direction	Sequence (5'>3')
Ha-pannier-1	sense	TCGAGCCTGGTGAAGAGCGAACCGGG
Ha-pannier-2	antisense	GGGTCTCCGGACGCAGTATTGATCTC
Ha-pannier-3	sense	GCTCCACCTCCGTAGAAGAC
Ha-pannier-4	antisense	AGCCATCAGTTTGGCAGAAG
Cs-pannier-1	sense	GGCGGTGAACGAGATGACAG
Cs-pannier-2	antisense	GCCATCAGTTTAGCGGACGC
Cs-pannier-3	sense	TGTGGAGGCGGGATGGTACT
Cs-pannier-4	antisense	CTGTTGACGCCGTGCAGCTT
Ha-rp49-1	sense	GCGATCGCTATGGAAAACTC
Ha-rp49-2	antisense	TACGATTTTGCATCAACAGT
Cs-rp49-1	sense	AGTGATCGTTATGGCAAGCT
Cs-rp49-2	antisense	TCTGTTTTGCATCAAAAGGAC

1004 **Supplementary Table 8 | Primers for direct sequencing of *Ha-pannier* ORF**

Primer name	Target direction	Sequence (5'>3')	Use of primer
Ha-pannier-ORFa-F	sense	GCCACTGTCCGTAATTAGCCCGAACAGG	PCR-1
Ha-pannier-ORFa-R	antisense	TCCACCAGAAATAAGGAAATGAGG	PCR-1
Ha-pannier-seq1	sense	GGGCAGGGAGTGCGTCAATTGTGGGGCC	Sequencing-1
Ha-pannier-seq2	antisense	GGGTCTGTTCATCCCATTCATCTTGTGG	Sequencing-1
Ha-pannier-seq3	sense	AAAACAAGGTGGTGGTAGT	Sequencing-1
Ha-pannier-seq4	antisense	ATAAGGTGACGTCCGTTGGAATCCAGA	Sequencing-1
Ha-pannier-seq5	sense	CCACAAGATGAATGGGATGAACAGACCC	Sequencing-1
Ha-pannier-seq6	antisense	TCTTGTCTTGTTTATGTCGT	Sequencing-1
Ha-pannier-5'-F	sense	AGTTCTTCCAAGCCCTCTAAAGTTCAACGAC	PCR-2
Ha-pannier-5'-R	antisense	GGGTCTGTTCATCCCATTCATCTTGTGG	PCR-2
Ha-pannier-seq7	sense	CATCGTCTTCAGATTAGGTGTAACGACG	Sequencing-2
Ha-pannier-seq8	antisense	ATAAGGTGACGTCCGTTGGAATCCAGA	Sequencing-2
Ha-pannier-exon3B-F1	sense	GTTTCCACCAACACCTTC	PCR-3
Ha-pannier-exon3B-R1	antisense	GTCTGGTTGCAGTTAGTATT	PCR-3
Ha-pannier-seq9	antisense	GATGCCGTCCTTGCGCATGGCCAGGGG	Sequencing-3
Ha-pannier-exon3B-F2	sense	ACCTTATGAACATGGATAC	PCR-4
Ha-pannier-exon3B-R2	antisense	GATATGCTTACTGGTGTCT	PCR-4
Ha-pannier-seq10	sense	TCTCCCCTCTCCGCCGGTCAGTTCTAC	Sequencing-4

1005 **Supplementary Table 9 | Primers for riboprobe synthesis (*in vitro* transcription, IVT)**

Primer name	Target direction	Sequence (5'>3')	Use of primers
Ha-pannier-ORF-F	sense	GTCAGCATGTTCCACACC	ORF cloning
Ha-pannier-ORF-R	antisense	TCTTGTCTTGTTTATGTCGT	ORF cloning
T3	sense	ATTAACCCTCACTAAAGGGA	IVT template (5' ORF)
SP6-Ha-pnr-NR	antisense	atttaggtgacactatagaACTCCATGGCACTCTC	IVT template (5' ORF)
T7	antisense	TAATACGACTCACTATAGGG	IVT template (3' ORF)
SP6-Ha-pnr-CF	sense	atttaggtgacactatagaAAACAAGGTGGTGGTAGT	IVT template (3' ORF)

Supplementary Figure 1 | Snapshots of pharate adult elytral pigmentation in *H. axyridis*

Melanin synthesis activity (black) and carotenoids accumulation (orange) were simultaneously visualised using h^C elytra stained with PO activity. Strong black and orange signals appeared after 80 h AP.

Supplementary Figure 2 | Expression analysis of *H. axyridis* (*Ha*) *pannier* in elytral primordia by RT-PCR.

a, Developmental expression profiles of *pannier* from the final instar larvae to pupae. *Ha-pnr, Ha-pannier. Ha-rp49, Ha-ribosomal protein 49* (internal control). Adult emergence is at 4.5 days AP in our rearing condition. Days after the onset of the each stage are indicated above. **(b, b')** Spatial distribution of *pannier* at 84 h AP. Future black (B1-B3) and red (R) regions were isolated for RT-PCR. Pharate adult elytra of three individuals were analysed (#1-#3). *rp49*, internal control. **b',** Left panel, pharate adult elytra with 3 hour PO staining (84 h AP). Right panel, an example of an elytron after isolation of red and black regions. Scale bars, 1 mm.

Supplementary Figure 3 | Validation of the *pannier*-locus scaffold of Platanus2 for the *H. axyridis* F2-3 sample

The scaffold was segmented into 2 kbp-windows, and links of 15 kbp-mate-pairs between windows (the number of mate-pairs > 2) were visualised as arcs .The entire region of the scaffold was uniformly covered by the mate-pairs, and no links between a distal window-pair inferring mis-assembly was observed.

Supplementary Figure 4 | Dot plots of the genomic scaffolds including the *pannier* locus

a−d, The dot plots between the consensus scaffolds obtained by reassembly of F2-3 mate-pair reads with Platanus2, and each consensus scaffold obtained by *de novo* genome assembly of the 10x linked-reads. **a,** h^C (F2-3) vs. h^C (NT6). **b,** h^C (F2-3) vs. h^A (F2-hybrid). **c,** h^C (F2-3) vs. *h* (NT8). **d,** h^C (F2-3) vs. *C. sep*. The colour code for colouring the homologous segments is on the right side of each panel. Blue arrow, the *pannier* locus. The green and the magenta pins indicate the position of the breakpoint

genotyping markers located on the outermost side, of all genotyping markers found to show association with the *h* locus in the three crossing experiments (*mbl* and *Mink* in Supplementary Table 2, respectively). The subset regions of each scaffold between these two markers or the corresponding regions were extracted for dot plot analyses. In each pair of comparison, sequential linear homology between the two scaffolds was confirmed. The responsible region of each *h* allele (h^C, 690 kb; h^A, 660 kb; *h*, 2.1 Mb + α) was included in a single linked-read genomic scaffold (a−c). In *C. septempunctata*, the genomic region corresponding to the region just downstream of the *pannier* locus in *H. axyridis*, was located in the scaffold different from that including *pannier* (scaffold 47 and 92), implying at least one translocation event in either of the two ancestral lineages (**d**).

Supplementary Figure 5 | Polymorphism in ORF sequences of *pannier-A* isoform in the 4 allelic strains of the *h* locus.

a, Nucleotide polymorphisms in *pannier* ORF in 4 allelic strains of *h* (*h*, h^A, h^{Sp}, h^C). R = G or A, Y = T or C, M = A or C, S = G or C, W = A or T. **b**, Amino acid sequences deduced

from ORF sequences in (**a**). X = T or S. The alternative exon region (exon 3A) and GATA zinc finger domains are indicated with grey and yellow, respectively.

Supplementary Figure 6 | Polymorphism in ORF sequences of *pannier-B* isoform in the 4 allelic strains of the *h* locus.

a, Nucleotide polymorphisms in *pannier* ORF in 4 allelic strains of *h* (*h*, h^{A}, h^{Sp}, h^{C}). R = G or A, Y = T or C, M = A or C, S = G or C, W = A or T. **b**, Amino acid sequences deduced from ORF sequences in (**a**). X = T or S. The alternative exon region (exon 3B) and GATA zinc finger domains are indicated with grey and yellow, respectively.

Supplementary Figure 7 | *vestigial* mRNA is upregulated in the presumptive black regions from early pupal stages in the h^{C} background.

1B, black region at 24 h AP. 1R, red region at 24 h AP. 3B, black region at 72 h AP. 3R, red region at 72 h AP. Read count of each sample was extracted from RNA-seq analysis data. Bars, mean read counts. (n = 3). Error bars, standard error of means. *, FDR < 0.01.

Supplementary Figure 8 | Expression analysis of *C. septempunctata* (*Cs*) *pannier* in elytral primordia by RT-PCR.

a, Developmental expression profiles of *pannier* in pupal stages. *Cs-pnr, Cs-pannier. Cs-rp49, Cs-ribosomal protein 49* (internal control). Adult emergence is at 4.5 days AP in our rearing condition. Days after the onset of the each stage are indicated above. **b, b',** Spatial distribution of *pannier* at 84 h AP. Future black (B1−B2) and red (R1−R3) regions were isolated for RT-PCR. Pharate adult elytra of three individuals were analysed (#1−#3). *Cs-rp49*, internal control. **b',** Left panel, pharate adult elytra with 3 hour PO staining (84 h AP). Right panel, an example of an elytron after PO staining for 1 hour and isolation of red and black regions. Scale bars, 1 mm.

Supplementary Data 1 | DNA sequences used in the molecular phylogenetic analyses

a, The aligned nucleotide sequences of *pannier* ORF before trimming gapped and shifted regions. **b,** The aligned nucleotide sequences of *pannier* ORF after trimming. **c,** The aligned nucleotide sequences of the conserved intronic regions of *pannier* (concatenated three blocks) before trimming gapped and shifted regions. **d,** The aligned nucleotide

1086 sequences of the conserved intronic regions of pannier after trimming. A dash indicates a

1087 gap.

Acknowledgements

We thank G. Eguchi, T. Ohde, H. Gotoh, Y. Sato, T. Yaginuma, M. Kobayashi, M. Ikeda for discussions, D. J. Emlen for critical reading of the manuscript, J. Morita, T. Mizutani for experimental supports, H. Kawaguchi for rearing of ladybird beetles, H. Asao and A. Akita for library preparation and machine operation of the resequencing analyses, and Functional Genomics Facility, NIBB Core Research Facilities for technical support. Images in Fig. 1a are used under license from Insect DNA Research Society Japan (newsletter, vo.13, September, 2010). This study was supported by a Grant-in-Aid from Formation and Recognition, Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), MEXT KAKENHI Grant Numbers 18017012, 20017014, 26113708, 221S0002, 17H05848, 18H04828, JSPS KAKENHI Grant Number 22380035, and NIBB Collaborative Research Programs (18-433).

1101 **Author contributions**

1102 T.N. and T.A conceived this study. T.M. and T.N. analysed the elytral

1103 pigmentation processes. T.A., T.M., K.G., K.H., A.I. and J.Y. analysed the sequence data.

1104 T.A, T.M., K.G., A.I., K.H., and J.H. performed cloning of the *pannier* genes from

1105 different alleles and species of ladybirds. T.M., K.G., A.I., K.H., and J.H. performed the

1106 larval RNAi experiments. K.H., K.G. and J.H. performed the semi-quantitative RT-PCR.

1107 T.A. performed the *in situ* hybridisation. J.Y. collected the total RNA for the RNA-seq

1108 analysis, and the genomic DNA samples for the initial *de novo* genome assembly. T.A.

1109 collected the DNA samples for the resequencing analyses. M.S. and Y.S. collected the

1110 RNA-seq raw data. Y.M. and A.T. performed the initial *de novo* genome assembly. K.Y.

1111 and S.S. collected the raw data for the resequencing and the RAD-seq analyses. K.Y.

1112 constructed the flexible ddRAD-seq protocol. R.K, M.O, and T.I. performed reassembly

1113 of the genome, *de novo* assembly of the linked-read genomic data, and validation of the

1114 obtained genomic scaffolds. M.K., T.T., T.A. and K.Y. performed mapping and

1115 quantification of the RNA-seq data. T.A. performed the data analyses for the genetic

1116 association studies, the gene annotation, the motif enrichment analysis, and the molecular

phylogenetic analyses around the *pannier* locus. T.A and T.N. wrote, and all authors commented on the manuscript.

Author information

DNA Data Bank of Japan (DDBJ) accession numbers for the sequences reported in this article are LC269047−LC269055 for *Ha-pnr*, LC269056 for *Cs-pnr* and LC269057 for *Cs-rp49*, DRA002559 for the genomic sequencing data of *H. axyridis*, DRA006068 for the genomic resequencing data of *H. axyridis*, and DRA005777 for the RNA-seq data. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.N. (niimi@nibb.ac.jp).

Author list

Toshiya Ando, Takeshi Matsuda, Kumiko Goto, Kimiko Hara, Akinori Ito, Junya Hirata, Joichiro Yatomi, Rei Kajitani, Miki Okuno, Katsushi Yamaguchi, Masaaki Kobayashi, Tomoyuki Takano, Yohei Minakuchi, Masahide Seki, Yutaka Suzuki, Kentaro Yano, Takehiko Itoh, Shuji Shigenobu, Atsushi Toyoda, and Teruyuki Niimi

1136

1137 **Affiliations**

1138 Division of Evolutionary Developmental Biology, National Institute for Basic Biology,

1139 National Institutes of Natural Sciences, Okazaki, Aichi, 444-8585, Japan

1140 Toshiya Ando, and Teruyuki Niimi

1141

1142 Department of Basic Biology, School of Life Science, SOKENDAI (The Graduate

1143 University for Advanced Studies), Okazaki, Aichi, 444-8585, Japan

1144 Toshiya Ando, Shuji Shigenobu and Teruyuki Niimi

1145

1146 Laboratory of Sericulture and Entomoresources, Graduate School of Bioagricultural

1147 Sciences, Nagoya University, Nagoya, Aichi, 464-8601, Japan

1148 Takeshi Matsuda, Kumiko Goto, Kimiko Hara, Akinori Ito, Junya Hirata, Joichiro

1149 Yatomi, and Teruyuki Niimi

1150

1151

1152 Department of Biological Information, Tokyo Institute of Technology, Meguro-ku, Tokyo,

1153 152-8550, Japan.

1154 Rei Kajitani, Miki Okuno, and Takehiko Itoh

1155

1156 NIBB Core Research Facilities, National Institute for Basic Biology, National Institutes

1157 of Natural Sciences, Okazaki, Aichi, 444-8585, Japan

1158 Katsushi Yamaguchi, and Shuji Shigenobu

1159

1160 Laboratory of Systems Genomics, Department of Computational Biology and Medical

1161 Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa,

1162 Chiba, 277-8562, Japan

1163 Masahide Seki, and Yutaka Suzuki

1164

1165 Bioinformatics Laboratory, Department of Life Sciences, School of Agriculture, Meiji

1166 University, Kawasaki, Kanagawa, 214-8571, Japan

1167 Masaaki Kobayashi, Tomoyuki Takano, Kentaro Yano

Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540, Japan

Yohei Minakuchi, and Atsushi Toyoda

Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka, 411-8540, Japan

Atsushi Toyoda