Repeated inversions within a *pannier* intron drive diversification of intraspecific colour patterns of ladybird beetles

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 & Teruyuki Niimi

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 >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. axyridis*’ extraordinary intraspecific variation may have arisen from ancient modifications in conserved elytral colour-patterning mechanisms in ladybird beetles.
Here are approximately 6000 ladybird beetle species described worldwide\(^1\). Charismatic and popular, ladybird beetles are famous for the red and black spot patterns on their elytra (forewings), thought to be a warning signal to predators that they store bitter alkaloids in their body fluids\(^2,3\) and are unpalatable. This red/black warning signal is shared among many ladybird beetle species, and provides a model for colour pattern mimicry by other insect orders. While most ladybird beetle species have only a single spot pattern, a few display remarkable intraspecific diversities, such as the multicoloured Asian ladybird beetle, *Harmonia axyridis*, which exhibits >200 different elytral colour forms (Fig. 1a). This striking intraspecific variation prompted us to investigate its genetic and evolutionary basis.

The first predictions regarding the genetics underlying the highly diverse elytral colour patterns of *H. axyridis* and its relevance to speciation 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 supergene\(^8,9\)) (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 (>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 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.

**Results**

**Elytral pigmentation during *H. axyridis* pupal development.**

To identify the gene regulating elytral colour pattern formation of *H. axyridis*, we first investigated the pigmentation processes during development. In the developing pupal elytra, red pigment (carotenoids\(^11\)) was accumulated in the future red-pigmented regions (Fig. 2a, pharate adult elytron). Red pigmentation occurred only in the thick ventral epidermal cells of the two layers of the elytral epidermis (Fig. 2b, c, red), and started at 80 h after pupation (80 h AP). Black pigmentation (melanin accumulation\(^11\)) occurred only in the dorsal cuticle of black-pigmented regions (Fig. 2d, black), and started approximately 2 h after eclosion. Although pharate adult elytra are not black, we detected a strong upregulation of enzymatic activity related to melanin synthesis\(^12\) in the nascent dorsal cuticle in the future black regions from 80 h AP (Fig. 2a, lower panels; Fig. 2c, black; Supplementary Figure 1). Every black-pigmented region was deployed complementary to the red regions. Therefore, we concluded that the developmental programs for both red and black pigmentation started around 80 h AP.

*pannier* promotes melanin and represses carotenoids in elytra.

We hypothesised that some of the conserved genes essential for insect wing/body wall patterning\(^3–18\) are recruited to regulate these elytral pigmentation processes, and tested this possibility using larval RNAi\(^19\). We performed a small-scale candidate screening focusing on genes involved in wing/body wall patterning (Supplementary Table 1), and found that the *Harmonia* orthologue 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

---

**Fig. 1** Intraspecific genetic polymorphisms of elytral colour patterns in *H. axyridis*. 
**a** Highly diverse elytral colour patterns of *H. axyridis*. **b** Four major alleles of the elytral colour patterns, \(h^C\), conspicius; \(h^S\), spectabilis; \(h^A\), axyridis; \(h\), succinea. **c** An example of inheritance of elytral colour forms. When \(h^C/h^S\) 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 in the F1 progeny. When the F1 heterozygotes are sibcrossed, F2 progeny shows three phenotypes (\(h^C/h^S\), \(h^C/h\) 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. Images in Fig. 1a are used under license from Insect DNA Research Society Japan (newsletter, vol. 13, September 2010). All rights reserved.
for wing blade patterning in *Drosophila*, but rather essential for patterning of the dorsal body plate attached to the wings (notum) \(^{21-23}\). *pannier* mRNA was upregulated from 48 h AP to 96 h AP in elytra (Supplementary Figure 2a), and preferentially in black regions (*h*\(^C\)) (Supplementary Figure 2b, b'). Immediately before or after 80 h AP (start of the pigmentation program, 76–84 h AP), *pannier* seemingly showed higher expression in the future black regions in the dorsal elytral epidermis (Fig. 3b). These data suggest that region-specific upregulation of *pannier* during the pupal stage regulates black pigmentation in the ladybird beetle’s dorsal elytral cells, and that regions of expression differ among the major *h* alleles to form different black patterns in *H. axyris*.

**The genomic basis for the colour pattern polymorphism.** These data led us to test whether *pannier* is associated with the classically identified locus *h*, which regulates elytral colour patterns. To identify DNA sequences near the *h* locus, we assembled de novo genome sequences (assembly version 1: 423 Mb; contig N50, 63.5 kb; scaffold N50, 1.6 Mb), and performed a genetic association study using the strains with different *h* alleles. We obtained the scaffold containing *pannier* and two additional adjacent scaffolds based on the truncated gene structures at the scaffold ends (Fig. 4a, Bgb and *pnr*). Restriction-site Associated DNA Sequencing (RAD-seq) analysis of backcrossed progenies (BC1, *h*\(^A\) × *h*\(^C\) F0 cross, *n* = 183) revealed that these three scaffolds are included in the five scaffolds that showed complete association with colour patterns (Fig. 4a, the upper left panel). In addition, genotyping of F2 individuals from two other independent genetic sib-crosses (*h*\(^C\) × *h* F0 cross (*n* = 80) and *h*\(^A\) × *h*\(^B\) F0 cross (*n* = 273)) indicated that the *pannier* locus is included in the relevant regions of all of the major four *h* alleles (*h*\(^C\), 690 kb; *h*\(^B\), 750 kb; *h*\(^A\), 660 kb; *h*, >2.1 Mb) (Fig. 4a, Supplementary Data 1).

To test contiguity of these three scaffolds, we re-assembled the genome using a novel genome assembler (Platanus2), and performed additional de novo genomic assemblies of *h*\(^C\), *h*\(^A\) and *h* alleles using linked-read and long read sequencing platforms (10× Genomics Chromium system; PacBio system). We obtained contiguous longer genomic scaffolds including the three described above (*h*\(^C\), 3.13 Mb/2.74 Mb; *h*\(^A\), 1.42 + 1.61 Mb; *h*, 2.79 Mb) (Supplementary Figure 3; Supplementary Data 2, *H. axyris*) and the genotyping markers showing complete association with colour patterns (Fig. 4a, the upper left panel). These data support the result of our genetic association studies.

To further delimit the candidate genes associated with the elytral colour patterns, we performed RNA-seq analysis using epidermal tissues isolated from the developing red or black regions before pigmentation in the *h*\(^C\) genetic background (Fig. 4b, 24 and 72 h AP RNA-seq). We found that *pannier* was the only gene statistically significantly upregulated in the developing black region compared to the red region at 72 h AP within the *h* locus candidate region (Fig. 4b, red bars, false discovery rate (FDR) < 0.01; Supplementary Data 3). These data pinpoint *pannier* as the major gene regulating the elytral colour pattern variation in *H. axyris*.

**Inversions and high diversification within a *pannier* intron.** We next investigated allele-specific polymorphisms at the *pannier* locus. We found that alleles of the first intronic region of *pannier* are more diverse than the surrounding genomic regions (Fig. 5a, asterisk, the middle whitish regions in *h*\(^C\) vs. *h*\(^A\), *h*\(^B\) vs. *h*, and *h* vs. *h*\(^C\) comparisons), whereas the same allele in different strains shows conserved fragments distributed throughout the region (Fig. 5a, blue bars, *h*\(^C\) (F2-3) vs. *h*\(^C\) (NT6) comparison).
comparisons between the alleles, we consistently found traces of large inversions in the upstream half of the first intron (Fig. 5a, reddish lines, $h^C$ vs. $h^A$, $h^A$ vs. $h$, $h$ vs. $h^C$; 56 kb–76 kb in size) (Supplementary Figure 5, dot-plot). However, we found that the coding sequences of pannier only showed a single nonsynonymous substitution in the region not conserved among organisms (G235V, $h^{Sp}$) (Supplementary Figure 6, 7), suggesting that cis-regulatory differences in the first intronic region of pannier are the major cause of intraspecific colour variation.

Moreover, we found that in H. axyridis, the size of the upstream noncoding sequences of the pannier locus (including the first intron of pannier, and the upstream intergenic region between the 5′ end of pannier 5′ UTR and 3′ end of the GATAe 3′ UTR) are 46–65 kb larger than the currently available corresponding genomic sequences of the other holometabolous insects (Fig. 5b, H. axyridis, 153–172 kb; other holometabolous insects, 13–107 kb). The expanded region in H. axyridis included at least four transcription initiation sites of pannier transcripts (Fig. 4c, pnr-1A–4B). In addition, in this region, several known DNA-binding motifs of transcription factors involved in Drosophila wing formation were more enriched allele-specifically than those in the other genomic regions (Table 1, allele-specifically enriched motifs; Supplementary Data 4). 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 the upstream and the downstream regions of the first intron of pannier specifically in the $h^C$ allele (Table 1, allele-specifically enriched motifs, $h^C$, Sd in the upstream and the downstream regions of the first intron). Furthermore, the RNA-seq data for the $h^C$ background also revealed that the sd coactivator gene vestigial was the only transcription factor gene that was significantly upregulated in the future black region from early pupal stages (Supplementary Figure 8), implicating Vestigial as one of the upstream trans- regulatory factors acting together with Sd to form the two-spotted elytral colour pattern of $h^C$. It is noteworthy that the noncoding region of pannier in each allele possesses putative DNA-binding motifs that can respond to variety of developmental contexts such as anterior–posterior patterning$^{26–28}$ (En, Inv), wing fate specification$^{17,24,29}$ (Sd), hinge-wing blade patterning$^{17,18,30,31}$ and wing vein patterning$^{29,32–34}$ (Ab, Al, B-H1, B-H2, Brk, Exd, H, Hth, Kn, Mad, Med, Nub, Rn, Rs, Vvl), hormonal cues$^{35,36}$ (EcR, Tai, Usp), and auto-regulation (Pnr) (Table 1, allele-specifically enriched motifs). These results suggest that allele-specific elytral colour patterns of H. axyridis may be formed by integrating appropriate combinations of developmental contexts of wing formation shared among insects.

**Colour-patterning function of pannier conserved in ladybirds.** We further tested whether the regulatory function of the red/black colour pattern in elytra is a conserved or a derived aspect of pannier function in ladybird beetles using the seven-spotted ladybird beetle, *Coccinella septempunctata*, which shows a monomorphic seven-spotted elytral colour pattern. The pannier mRNA was detected in the larval elytral primordium, was upregulated from 24 h AP to 96 h AP (Supplementary Figure 9a), and preferentially expressed in the black spots of elytra in *C. septempunctata* (Supplementary Figure 9b, b’) similarly to that in H.

**Fig. 3** pannier expression foreshadowing the adult colour pattern switches red/black pigmentation processes. a The adult phenotypes of RNAi treatments targeting GFP (negative controls, GFP RNAi) and pannier (panrier 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 (panrier) 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 $h$) adapted from Fig. 2a. White arrowhead, the region with a weak signal. Black arrowheads, the regions with intense signals. Scale bars, 1 mm
The black-to-red switching phenotype was also observed in *C. septempunctata* adults treated with larval RNAi targeting *pannier* (Fig. 3a, *C. septempunctata*). These data suggest that the elytral colour-patterning function of *pannier* may be conserved at the inter-genus level in ladybird beetles. To investigate the putative regulatory sequences at the *pannier* locus, we performed de novo assembly of the *C. septempunctata* genome using a linked-read sequencing platform (10× Genomics Chromium system), and obtained a contiguous genomic scaffold including the *pannier* locus (2.41 Mb; Supplementary Fig. 4d; Supplementary Data 2, *C. septempunctata*). Whereas the noncoding sequences of *C. septempunctata pannier* are enriched with several species-specific DNA-binding motifs (Table 1, *C. septempunctata*), we found DNA-binding motifs commonly enriched between *H. axyridis* and *C. septempunctata* at the *pannier* locus (Fig. 4).
Fig. 5 Traces of inversions and high sequence diversification within a *pannier* intron 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 program (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 BLAST72 hit blocks (bluish, forward hit; reddish, reverse hit) in the comparison between the two surveyed insects is adapted from ref.107 (Coleoptera), and TIMETREE108 (Diptera, Hymenoptera, Lepidoptera). The sizes of the noncoding regions at the *pannier* locus were extracted from the genome assembly of each allele in *C. septempunctata* (haplotype-1) and *C. septempunctata* (haplotype-2). 

**b** Overview of the size of the upper noncoding regions (the first intron + the upstream region) of *H. axyridis* (*h^C^*, *h^A^*, *h*) and *C. septempunctata* (*C. sep*). 

**c** *pannier* coding region

**d** *pannier* 1st intron

Bootstrap values were calculated from 1000 resampling of the alignment data. Bars, 0.01 substitutions/site.
Harmonia and Coccinella, which are associated with wing vein formation and wing/body wall patterning (Exd, Hh and Mad)\textsuperscript{29,31,32} (Table 1, commonly enriched motifs). Therefore, co-option of such wing developmental modules in the regulatory region may have facilitated acquisition of a novel expression domain of \textit{pannier} in pupal elytral blades in ladybird beetles.

In order to explore the history of the emergence of elytral colour patterns in \textit{H. axyridis}, we also performed a molecular phylogenetic analysis focusing on the highly conserved \textit{pannier} intronic sequences shared among \textit{Harmonia} and \textit{Coccinella} (three blocks, totalling 1.1 kb in length, Supplementary Data 5). The maximum likelihood (ML) phylogenetic tree inferred from nucleotide sequences of the \textit{pannier} coding region did not resolve the phylogenetic relationship among the alleles in \textit{H. axyridis} to a satisfactory level (Fig. 5c, bootstrap values <75). However, the ML tree inferred from the conserved intronic sequence suggested that in \textit{H. axyridis} the contrasting colour patterns of the \textit{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. 5d, bootstrap values >90).

### Table 1: Known DNA-binding motifs enriched in the noncoding regions of the \textit{pannier} locus

| Category | Allele/species-specifically enriched motifs | Upstream intergenic region | Upstream region of the first intron | Downstream region of the first intron |
|----------|------------------------------------------|-----------------------------|-------------------------------------|---------------------------------------|
| Allele/species-specifically enriched motifs | \textit{h}^C | EcR, Foxo | En, Exd, H, HLH106, Mad, Mhc, Mnt, Ss, Pad, Pan, Poxn, Sd, Tgo, Tai, Vvl | Ab, B-H1, Crc, Crol, Dr, Nn, Sd |
| | \textit{h}^A | Exd, Pan, Vvl | Ab, Ato, Ets21C, H, Mad, Sd, Tgo, Tgo | Usp |
| Commonly enriched motifs | \textit{h} | Kni, Rn, Sqz | B-H2, Eg, Pnr, Tgo, Tgo | Al, B-H1, Lms, Nub, Sd, Sth, Tgo |
| | \textit{C. sep} | Ato, EcR, H, Ss, Tgo | Al, Brk, E(spl)m, H, Ss, Tgo | Dr, En, Inv, Mad, Sens, Slou, Umpg |
| Region size (bp) | \textit{H. axy} & \textit{C. sep} | 38,192 (F2-3)/51,566 (NT6) | 71,107 (F2-3)/76,264 (NT6) | 47,450 (F2-3)/44,236 (NT6) |
| | \textit{h}^A | 42,826 | 75,417 | 52,210 |
| | \textit{h} | 40,778 | 67,340 | 67,340 |
| | \textit{C. sep} | 20,370 | 56,564 | 18,432 |

Enriched DNA-binding motifs of \textit{Drosophila} transcription factors involved in wing formation are listed (\textit{p} < 0.05). The Scalloped binding motif (Sd) discussed in the text is underlined.

Discussion

The \textit{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 \textit{h} locus in ladybird beetles\textsuperscript{12}. Also, it is worth noting that a concurrent study by Prudhomme, Estoup and their colleagues independently identified the same locus in \textit{H. axyridis} by whole-genome sequencing, population genomics, gene expression and functional genetics approaches\textsuperscript{37}. 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 \textit{H. axyridis}. In addition, we also discuss the underlying evolutionary developmental backgrounds specific to ladybird beetles.

The common ancestor of \textit{Harmonia} and \textit{Coccinella} (\textit{Coccinellinae}) diverged more than 33.9 million years ago, according to molecular phylogenetic analyses and fossil records\textsuperscript{38,39}. Therefore, the elytral colour-pattern function of \textit{pannier} shared between \textit{H. axyridis} and \textit{C. septempunctata} was most likely acquired before this divergence event. The 1.1 kb sequence blocks in the first intron of \textit{pannier} conserved between \textit{H. axyridis} and \textit{C. septempunctata} are a likely candidate for a regulatory element associated with the ladybird beetle-specific elytral expression of \textit{pannier} in the pupal elytra. The effects of enhancer activities of these sequence blocks have not yet been experimentally addressed. However, the acquisition of such regulatory sequences during evolution would have coincided with the acquisition of the elytral-colour-pattern function of \textit{pannier} (Fig. 6, blue diamond). These conserved sequence blocks are located in the expanded intronic region specific to \textit{H. axyridis} (Fig. 5a, black arrow). Therefore, the expansion of the first intron in the ancestral lineage of \textit{H. axyridis} (Fig. 6, intronic expansion) might be one of the events that facilitated diversification of the intraspecific elytral colour patterns.

In the genus \textit{Harmonia}, colour patterns similar to those encoded by the \textit{h} allele and those of \textit{C. septempunctata} (black spots in red background) are commonly observed. Also, the position of the spots is similar across species (e.g. \textit{H. quadripunctata}, \textit{H. octomaculata}, and \textit{H. dimidiata}). Therefore, we speculate that the intronic sequence of \textit{pannier} in the \textit{h} allele of \textit{H. axyridis} might retain a repertoire of regulatory sequences acquired in a common ancestor of the genus \textit{Harmonia} (Fig. 6, green arrowhead). However, in the ancestral lineage of \textit{H. axyridis}, the regulatory region of \textit{pannier} appears to have been modified to generate novel colour patterns of the recently diverged alleles (\textit{h}^C, \textit{h}^9 and \textit{h}^A; red spots in black background; Fig. 6 magenta, red and purple arrowheads). The 70 kb-scale noncoding sequences located at the upstream region of the first intron of \textit{pannier} that is specifically expanded in \textit{H. axyridis} (Fig. 6, Intrinsic expansion, yellow box) might have facilitated accommodation of the allele-specific regulatory motifs responsible for the diversified colour pattern of elytra. In addition, traces of inversions in this region consistently found in allele comparisons suggest that repeated inversions in this region (Fig. 6, white arrowheads) created opportunities to diverge the noncoding sequence of \textit{pannier} to successively generate novel diverse alleles within a species by suppressing recombination within this region. Such inversion events would have occurred in the common ancestor of \textit{H. axyridis} and its reproductively isolated sister species, \textit{H. yedoensis} because the major elytral colour patterns are shared between the two species\textsuperscript{40}. Large-scale chromosomal inversion is believed to be one of the major driving forces generating and maintaining intraspecific morphological variation within a species\textsuperscript{41–44}. Our study exemplifies that not only a single inversion event but also repeated inversion events at an expanded intron can lead to the acquisition of novel morphological traits within a species.

From the viewpoint of evolutionary developmental biology, it is noteworthy that in \textit{H. axyridis}, all of the developmental genes known to regulate colour pattern and pigmentation, a
single gene, *pannier*, is responsible for the major classes of intraspecific entire wing colour pattern diversification. This evolutionary pattern contrasts with that of the intensely studied warning signals of *Heliconius* butterflies. In the case of *Heliconius erato* and *Heliconius melpomene*, five major loci and several minor loci located on different chromosomes regulate multiple intraspecific wing colour patterns prevailing in the population\(^{(45)}\). This difference in evolutionary mechanisms may stem from a paucity of available options of evolvable genes in the gene regulatory network of elytral colour patterning. Ladybird beetles diverged from ancestral species of Cucujoidea\(^{(38)}\) (Fig. 5b, Cucujoidea), leaf-litter or rotten-tree dwelling insects. Thus, the ancestor of ladybird beetles would have had far less colourful and more simply patterned forewings (elytra) than the ancestors of butterflies, moths. Therefore, these ancestors presumably would have possessed far fewer colour pattern regulatory genes. In *H. axyridis*, this developmental constraint may have led to the selection of *pannier* as the major evolvable gene to a signal-integrating “input—output” regulatory gene\(^{(46,47)}\). This might have generated >200 colour patterns genetically tightly linked to the h locus by utilising the expanded regulatory DNA sequence. Future research aiming to identify specific regulatory inputs to *pannier* will help clarify the regulatory mechanisms underlying the generation of highly diverse intraspecific polymorphism at the interspecific level. Another important issue to clarify whether *pannier* is indeed the hotspot of morphological evolution in ladybird beetles is whether *pannier* is responsible for the remaining >20 minor colour patterns in *H. axyridis*.

**Methods**

**Insects.** Laboratory stocks of *H. axyridis* and *C. septempunctata* were derived from field collections in Japan. They were reared at 25 °C and usually fed on artificial diet\(^{(46)}\), or fed on the pea aphid *Acyrthosiphon pisum* (kindly provided by Dr. T. Miura) for egg collection. Larvae and pupae analysed in this study were not sexed.

**Phenoloxidase (PO) activity staining.** Pupa elytral discs were dissected in a potassium phosphate buffer (K-PO\(_4\)) buffer; 100 mM KH\(_2\)PO\(_4-K\)PO\(_4\), 150 mM NaCl, pH 6.3) on ice. PO staining was performed using 0.4 mg/ml dopamine as a substrate in 40% K-PO\(_4\) buffer/60% isopropl alcohol for 2 h at room temperature as previously described\(^{(42)}\). 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\(_2\)HPO\(_4\), 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 dehydation, the elytra were embedded in 4% carboxymethyl cellulose (FINETEC), and were frozen in hexane cooled with dry ice. The freeze-embedded elytra were stored at −80 °C until use. The 6 μm frozen sections were prepared using an adhesive film (Cryofilm Type 1; FINETEC)\(^{(49)}\). Sections of the PO activity-stained elytra were dried at least 1 h at room temperature, mounted in PBS, and photographed under an inverted microscope (IX70, Olympus). For nuclear and F-actin staining, sections were treated with 2.5 μg/ml propidium iodide and 1 μg/ml 4′,6-diamidino-2-phenylindol (Molecular Probes) for 1 h at 37 °C under a dark condition. After washing three times in PBS, the sections were mounted in an antifade reagent (FluoGuard\(^{TM}\), Bio-Rad), and images were captured with a confocal laser-scanning microscope (LSM 510; Carl Zeiss).

For localisation of carotenoids, elytra at 96 h AP were embedded and sectioned as described above. All procedures were rapidly performed to prevent diffusion of carotenoids. The sections were dried for 1 min, mounted in PBS and immediately photographed under an inverted microscope (IX70, Olympus).

**cDNA cloning.** Larval and pupal elytral discs and pharate adult elytra of *H. axyridis* (h\(^{Xy}\)) and *C. septempunctata* were dissected in PBS on ice. Soon after dissection, the tissues were frozen in liquid nitrogen and stored at −80 °C until use. Total RNA was extracted from each sample using TRIzol Reagent (Invitrogen) or RNaseasy Micro Kit (Qiagen) according to the manufacturer’s instructions, and treated with 2 U DNase I (Ambion) for 30 min at 37 °C. The first-strand cDNA was synthesised with SMARTer PCR cDNA Amplification Kit (Clontech) using 1 μg of total RNA according to the manufacturer’s instructions. *H. axyridis* and *C. septempunctata* cDNA fragments were amplified by reverse transcription-polymerase chain reaction (RT-PCR) and rapid amplification of cDNA ends (RACE) with the primers listed in Supplementary Tables 2, 3. The PCR product was cloned into the EcoR V site of the pBluescript KS + vector (Stratagene) or pCR4-TOPO vector (TOPO TA Cloning Kit; Invitrogen). The nucleotide sequences of the PCR products were determined using a DNA sequencer 3130 genetic analyser (Applied Biosystems). The SNPs in open reading frame (ORF) of *pannier* were determined through direct sequencing of the PCR products treated with EcoSAP-IT (Affymetrix). Sequencing was performed by DNA sequencing service (FASMAC) using the primers listed in Supplementary Table 4. Sequence analysis was carried out using DNASK (Hitachi Software Engineering) or ApE\(^{(50)}\) (version 2.0.45) software.

The nucleotide sequences and deduced amino acid sequences were aligned with ClustalW in MEGA\(^{(51)}\) software (version 7.0.18). The alignment figures were generated using Boxshade\(^{(52)}\) (version 3.21).

**Gene expression analysis by RT-PCR.** For the gene expression analysis in each developmental stage, elytral tissues of three individuals of *H. axyridis* (h\(^{Xy}\)) and *C. septempunctata* were dissected as described above. Six elytral tissues from each sampling stage were pooled in one test tube. Total RNA extractions and the subsequent first-strand cDNA syntheses (using 425 and 267 ng of total RNA for *H. axyridis* and *C. septempunctata* samples, respectively) were performed as described above. Three microtubes of 100 and 62.8 times diluted *H. axyridis* and *C. septempunctata* first-strand cDNA was used as a template for each PCR, respectively. The PCR cycle number was 35 for all genes. A set of primers #1 and #2 for each gene was used for this analysis (Supplementary Table 5).

For the gene expression analysis in the future red and black regions, the red and black regions of pharate adult elytra at 84 h AP were collected by boring with

---

**Fig. 6** An evolutionary model for the colour pattern diversification in *H. axyridis*. See details in the Discussion. The images of ladybird beetles are adapted from Figs. 1b and 3c.

---

H. axyridis C. septempunctata

[Diagram showing evolutionary relationships and genetic mechanisms.]
From 72 h AP (described previously). In brief, the cloned cDNA fragments in DNA vectors were samples for the RT-PCR experiments. No band was detected in these reactions for reverse transcriptase were performed with cDNA synthesis as negative control.

After post-equilibrated to hybridisation solution with 0.4 ng/cRNA probes overnight. After reverse transcriptase were performed with cDNA synthesis as negative control.

In situ hybridisation. Essentially the same protocol for whole mount pupal antennal primordia of the silk moth (was used. For sclerotized pupal elytra of ladybird beetles, several procedures were modified as follows: to increase RNA probe penetration in elytral epidermis covered with sclerotised cuticle at 76–84 h AP, the peripheral edge of an elytron was cut off, and then, ventral and dorsal elytral epidermis layers appressed together were carefully separated with fine forceps after fixation; to reduce nonspecific probe hybridisation, fixed, separated and detergent-permeabilised elytral epidermal samples were stored in 100% methanol for more than 12 h at −30 °C, and prehybridisation treatment was extended to two overnight incubation; the concentration of cRNA probes were reduced to 0.4 ng/µl, the ventral epidermis samples were not used for analysis because of high nonspecific background signals. Sample washing was performed for 10 min three times unless otherwise noted.

pannier antisense probes were designed at 5′ and 3′ regions of ORF excluding the two conserved GATA zinc finger coding regions in the middle to prevent cross-hybridisation with other GATA family genes. The PCR primers used to amplify the template DNA for in vitro RNA probe synthesis were listed in Supplementary Table 6. Essentially, panner ORF fragment was amplified by RT-PCR using cDNA from 72 h AP (h), and cloned into pcR-TOPO vector (Invitrogen). Sense and antisense probe templates were amplified from the cloned cDNA. Sense and antisense DIG-labelled riboprobes were transcribed using the flanking T7, T3 or SP6 promoter sequences, and DIG RNA labelling kit (Roche). Mixture of 5′ and 3′ probes was used for hybridisation. The concentrations of RNA probes were quantified by agarose gel electrophoresis.

First, pupal elytra were dissected in PBS, and fixed with 4% PFA in PBS for 2 h, and washed with PBS including 0.1% Tween20 (PTw). Fixed elytra were dorsoventrally separated using fine forceps, and permeabilized with 0.5% Triton X-100. After permeabilized with PTw, samples were treated with 20 µg/ml proteinase K in PTW at 37 °C for 30 min. Proteinase K was immediately washed out by quick washes with 2 mg/ml glycine in PBS and following washes with PTW. After post-fixation with 4% PFA and 0.1% glutaraldehyde for 30 min, samples were equilibrated to hybridisation solution with five steps of hybridisation solution wash series (10×), 0.5×, 0.1× and 0.5× hybridisation buffer (57°C over two nages for ca. 40 h), and hybridised with 0.4 ng/µl cRNA probes overnight. After reverse hybridisation wash series wash series (50, 25, 12.5, 0% (PTw) 0% (PTw)) and a wash with RNase A reaction buffer (10 mM Tris-HCl, 500 mM NaCl, pH 8.0), single stranded probes not hybridised to mRNA were degraded with 20 µg/ml RNase A at 37°C for 1 h. After hybridisation to hybridisation solution again, non-specifically bound degraded probes were washed out with 4 times of 20 min washes with hybridisation solution at 57°C and following reverse hybridisation solution wash series (50, 25, 12.5%; 20 min for each). After washing with PBTW and PBSW (PBS with 0.01% Saponin), and blocking for 30 min with BSS-BSA (PBS with 0.2% BSA and 0.1% PEG), samples were incubated with alkaline phosphatase-conjugated anti-hC antibody (1:2000, Roche) in BSS-BSA. After washing with PBSW and NTMT buffer (0.1 M NaCl, 0.1 M Tris-HCl (pH 9.5), 50 mM MgCl2, 0.1% Tween20), Colour development was conducted using 170 µg/ml BCIP and 340 µg/ml NBT (Roche) diluted in NTMT buffer. After washing with NTMT buffer and PBS, the samples were stained with PO activity were used for boring because carotenoid localisation was observed as follows: to increase RNA probe penetration in elytral epidermis covered with sclerotised cuticle at 76–84 h AP, the peripheral edge of an elytron was cut off, and then, ventral and dorsal elytral epidermis layers appressed together were carefully separated with fine forceps after fixation; to reduce nonspecific probe hybridisation, fixed, separated and detergent-permeabilised elytral epidermal samples were stored in 100% methanol for more than 12 h at −30 °C, and prehybridisation treatment was extended to two overnight incubation; the concentration of cRNA probes were reduced to 0.4 ng/µl, the ventral epidermis samples were not used for analysis because of high nonspecific background signals. Sample washing was performed for 10 min three times unless otherwise noted.

Reassembly of the genomic scaffold at hC pannier. Adaptor sequences and low-quality regions in paired-end and mate-pair reads were trimmed using Platanus trim (version 1.0.7) with default parameters. Trimmed reads were assembled by Platanus2 (version 2.0.0), which was derived from Platanus56 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 nonbubble scaffolds are collectively called “phased-scaffolds”. (7) cDNA library, Platanus2 can construct primary-bubbles and nonbubble scaffolds to construct long “consensus scaffolds”, which consists of mosaic structure of haplotypes (i.e. paternal and maternal haplotypes are mixed). Employing the strategy of Platanus2, certain highly heterozygous regions were expected to be assembled contiguously compared to Platanus.

Using the markers of the responsible region for elytral colour patterns (the h locus), we found that two long bubbles and one short nonbubble scaffold corresponded to the locus. Consequently, one consensus scaffold covering the breakpoint markers at the h locus was constructed from these phased scaffolds. We used that consensus scaffold (3.13 Mb for the downstream in silhouette sequence analysis). We assessed the completeness of the genome assembly using BUSCO58 (version 3.0.2, Insecta dataset (1658 orthologues)).

Genome sequencing by long reads and linked-reads. High molecular weight (HMW) genomic DNA was extracted using QIAGEN Genomic-tip 100/G (QIA-GEN) 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 10× Genomics Chromium system, one pupa (hF1 (NT6 strain) and h (NT8 strain) or one adult (hF2 adult progenies in genetic cross hF1 × hF2) was used. Size selection by BluePippin (range: 50 kb–80 kb, Sage Science) was performed only for hF1 genomic DNA used in 10x linked-read library preparation.

Preparation of gel bead-in-emulsions (GBEs) for each 10× Genomics Chromium library was performed using 0.5–0.6 ng of HMW genomic DNA according to the manufacturer’s instructions. The prepared GBEs were quality-checked using Qubit dsDNA HS kit (Thermo Fisher) and Bioanalyzer (Agilent), and sequenced together with Chromium CAR Chromium PIC (10× Genomics). The constructed DNA libraries were quality-checked again in the same way. Sequencing of the libraries was performed in the HiSeq X ten (Illumina) platform (1 library/lane) at Macrogen. In total, we generated 66.9, 64.6, 64.9 and 60.3 Gb of raw reads for linked-read library preparation.

For library preparation for 10× Genomics Chromium system, one pupa (hF1 (NT6 strain) and h (NT8 strain) or one adult (hF2 adult progenies in genetic cross hF1 × hF2)) was used. The libraries were prepared according to the 20-kb Template Preparation Using BluePippin™ Size-Selection System (Sage Science). Sequencing of the libraries was performed in the PacBio RS II (Pacific Biosciences) platform. In total, 4.3, 4.1, 4.4 and 4.0 Gb of insert sequences (approximately 10× coverage of the genome, assuming a genome size of 423 Mb) were obtained from 4 to 5 SMRT cells for each library.

De novo assembly of 10×-linked reads. For 10× linked-reads libraries of four samples (three H. axyridis and one C. septempunctata), Supernova (version 2.0.0)59 was used.
was executed with default parameters except for the maximum number of used reads (the –max-reads option) to obtain the optimum coverage depth for Supernova (9).) For each sample, the value for –max-reads was determined as follows: (1) Barcode sequences in raw linked-reads were excluded using “longrange basis” command of LongRanger [version 2.1.2], resulting in “barcoded.fastq” file. (2) Adaptor sequences and low-quality regions in “barcoded.fastq” were trimmed using Platanus_trim (version 1.0.7) with default parameters. (3) 32-mers in the trimmed reads were counted using Jellyfish [version 2.2.3] using the following two commands and options: $ jellyfish count -m 32 -o 20M -C -o out.jf barcoded_1_trimmed barcoded_2_trimmed
$ jellyfish histo -h 1000000000 -o out.histo.out
In summary, all 32-mers in both strands (–C) were counted and distribution of the number of sequences according to the presence of repeats was counted. (4) The haploid genome size was estimated using the custom Perl script. For the determination of the number of 32-mers occurrences (“out.histo”), the number of occurrences corresponding to a homology peak was detected, and the total number of 32-mers whose occurrences were smaller than the number of occurrences corresponding to the bottom between zero and heterozygous peaks were excluded for the calculation to avoid the effect from sequencing errors. (5) The values for –max-reads 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-panner (hA) [NT6], 2.74 Mb; hH [F2 hybrid], 1.42 ± 1.61 Mb; hN [NT6] 2.79 Mb), and homologous regions in C. septempunctata (halytote, 1.10 ± 0.24 Mb; halytote 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 [version 3.0.2, Insecta database (1658 orthologues)].

Gap filling of the genomic scaffolds at the panner locus. Concerning the genome assemblies of H. axyridis, we used minimap2 [version 2.9.1] and PBjelly [version PBSuite_15.8.2.4] as a reference to fill gaps around the panner locus. In each genome of three strains of H. axyridis, we first mapped PacBio reads to the genome assemblies generated from the 10× linked-reads using minimap2. Then, we chose PacBio reads mapped to the scaffold containing panner gene. These PacBio reads were subjected to gap-filling of the scaffold with PBjelly. We obtained gap-free nucleotide sequences spanning the entire panner locus and the upstream intergenic regions.

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

Validation of the panner scaffold re-assembled by Platanus2. For the H. axyridis F2–3 sample, trimmed reads of the 15 kb-mate-pair library were mapped to the consensus scaffold set of Platanus2 using BWA-MEM [version 0.7.12-r1039] with default parameters. Next, a consensus scaffold corresponding to the panner locus was segmented into 2 kbp-windows, and links between windows (23 mate-pairs) were visualized by Circos [version 0.69-6] (version 0.69-6).

Preliminary resequencing of H. axyridis genome for RAD-seq. Genomic DNA was extracted from each of hH (F6 strain), hH (NT3 strain), and hH (CB-5 strain), and used to create Illumina libraries using TruSeq Nano DNA Sample Preparation Kit (Illumina). Paired-end libraries were sequenced on the Illumina HiSeq 1500 using a 2 × 106-nt paired-end sequencing protocol, yielding 84.7 M paired-end reads. SNP site identification was conducted basically according to the GATK Best Practice [version 3.7.3, August 7, 2015]. After trimming adaptor sequences with Cutadapt software (version 1.9.1), the sequence data were mapped to the de novo genome assembly data using bwa software [version 0.7.15, BWA-MEM algorithm]. SNPs were identified using low quality polymorphic sites that behaved as located on the X chromosome. In addition, polymorphic sites showing no recombination were identified individually using an automatic nucleic acid sequence program with a minimal read-depth to genome size ratio of 1:100. The resulting SNP data were filtered to obtain high-quality data using the following criteria: (1) SNP sites were filtered out when the reads for the minor allele were less than 5% of the total read count. (2) SNP sites were filtered out when the reads for both alleles were less than 5% of the total read count. (3) SNP sites were filtered out when the reads for the major allele were less than 20%. For each pair of the entire scaffolds and the extracted panner region, we constructed dot plots by performing pairwise-alignment using “nummer” program in the MUMmer package [version 3.1]. The options of nummer were as follows: (1) the entire scaffolds, H. axyridis vs. H. axyridis, Default parameters; (2) the entire scaffolds, H. axyridis vs. C. septempunctata each x 200; (3) the panner region, H. axyridis vs. H. axyridis, “-l 12”, Alignment results (delta files) were input into “mummerplot” program to generate dot plots. Note that resultant gnuplot scripts resulting from mummerplot were edited for visualisation.

We also visualised the homology and structural differences between the 700 kb-genome of H. axyridis using EasyGene [version 2.2.2]. Short BLAST [version 2.2.2] hit fragments less than 500 bp, and putative short repeat sequences less than 1250 bp, which showed more than two BLAST hit blocks within the 700 kb region, were filtered using a custom Perl script. Exon–intron structures of putative genes in the 700 kb regions were obtained using Exonerate [version 2.2.0] with the options “-m est2genome --showavglar -s -yoo” with the qvalue of 0.1. The cDNA sequences cloned by RT-PCR or predicted by RNA-seq were used as queries. If a single cDNA unit was split into multiple fragments, we merged the fragments by performing exonerate search again using the cDNA sequences 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 program “exonerate.gff3.pl” Perl script with the option “-e -EST”. The GFF3 file and a FASTA format file of each scaffold were converted to a GENBANK format file as a part of the EMBOSS ‘seqdef’ program (version 6.6.0.0) with the options “-format gff – sofset genbank”. The GENBANK format files corresponding to the 700 kb genomic sequences surrounding panner, which were used as input files of Easyfig, were extracted using the Genbank_slicer.py Python script (version 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 reactions to be completed sequentially without DNA purification in each step to make the procedures simple. In addition, we designed 96 sets of indexed and forked sequencing adapters compatible with Illumina pair-end sequencing (Supplementary Table S2).

Briefly, 100 ng of DNA was first double-digested with 15 U of EcoRI-HF and 15 U of HindIII-HF in 20 µl of NEB CutSmart Buffer (New England Biolabs) at 37 °C for 2 h. Fifteen microliters of the digested DNA, 4 µmol of adaptor DNA, 10 µmol of T4 DNA ligase were mixed in 20 µl, incubated at 22 °C for 2 h, and denatured at 65 °C for 10 min. Ligated library DNA fragments were purified with Agencourt AMPure XP (Beckman Coulter) according to the manufacturer’s instructions. Library DNA fragments ranging from 300 to 500 bp were size-selected with Pippin Prep (Sage Science). 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).
In the former cross, a single h male (D-5 strain) and a single virgin h' female (F2-3-B strain) were crossed, and the obtained F1 progenies were sibcrossed. Finally, 80 F2 adult progenies (h′ = 30, h′/h = 34, and h′ = 16) were collected for genotyping, and stored at −30 °C until use. In the latter cross, a single h′ male (CB-5 strain) and a single virgin h′ female (NT3 strain) were crossed, and the obtained F1 progenies were sibcrossed. Finally, 273 F2 adult progenies (h′ = 103, h′/h′ = 40, and h′′ = 90) were collected for genotyping, and stored at −30 °C until use.

Genomic DNA was extracted individually using the automatic nucleic acid extractor (PI-50a, KURABO) as described in the previous section, and diluted to approximately 100 ng/µl. We searched for genotyping markers by amplifying and sequencing the region of the genes surrounding panner with PCR. The individual PCR was performed using approximately 100 ng of genomic DNA and Q5 DNA polymerase (New England Biolabs) with 45 cycles. The primers used, the markers identified and the typing results are summarised in Supplementary Data 1.

RNA-seq analysis. The total RNA extraction procedure for RNA-seq is essentially the same as that for the gene expression analysis in the presumptive red and black regions by RT-PCR. The same strain used for de novo genome sequencing (F2-3 strain, h′) was used. In total, 12 samples (2 colours [Black/Red] × 2 developmental stages [24 h AP/72 h AP] × 3 biological replicates) were prepared for RNA-seq analysis. Two fragments of bored epidermis from left and right elytra were collected as a single sample in each condition. All total RNA extracted from each sample (12–158 ng) using RNeasy Mini Kit (QIAGEN) and QIAcube (QIAGEN) was used for each DNA library preparation. RNA-seq library preparation was performed using the SureSelect strand-specific RNA library prep kit (Agilent) according to the manufacturer’s instructions. Briefly, mRNA was purified using oligo-dT MicroParticles. The strand-specific RNA-seq libraries were prepared using DUTP and Uracil-DNA Glycosylase. The libraries and its intermediates were purified and size-fractionated by AMPure XP (Beckman Coulter). For quality check and quantification of the RNA-seq libraries, we employed 2100 Bioanalyzer and DNA 7500 kit (Agilent). 100 bp paired-end read RNA-seq tags were generated using the Hiseq 2500 (Illumina) following the methods in the User Guide.

In advance of reference mapping, adaptor and poly-A sequences were trimmed from raw RNA-seq reads by using Cutadapt (ver. 1.9)118. Low-quality reads were also filtered out by a custom Perl script as described previously.90 The preprocessed RNA-seq reads were mapped to the reference *H. axyridis* genome (assembly version 1) using TopHat283 (ver. 2.1.0) with the -u option in each sample. All predicted transcript units and all loci from different samples were merged by Cuffmerge in the Cufflinks suite. The RNA-seq reads pairs (fragments) mapped to each predicted transcript unit and locus were counted using HTSeq85 (ver. 0.6.1) with the options “--s no-t exon -i transcript” and “--s no-t exon -l locus”, respectively. The downstream statistical analyses were performed using edgeR90 package (ver. 3.16.5). The raw RNA-seq fragment counts were normalised by the trimmed mean of M-values (TMM) method. Fold change between black and red regions in each stage and its statistical significance (FDR) were calculated. The mean fold changes of the genes in the scaffolds including the h locus candidate region were visualised with IGV90,99 software (ver. 2.3.88).

Comparison of the panner locus size. The holometabolous insects, whose genomic sequences are well assembled at the h genomic level, include the same scaffold, for comparison. Concerning Coleoptera, genomic sequences were collected from the Genome database at NCBI90 (GCA_001937115.1, Atum_1.0, GCA_000390285.2, Agla_2.0, GCA_000648969.2, Otua_2.0, GCA_001421225.1, Nicve_v1.0, GCF_006699045.1, Aplao_1.0, GCA_002278615.1, Pchal_1.0) and Fireflybase93 (*Photinus pyralis* genome 1.3.0, *Aquatilis latipes* genome 1.3.0). Concerning holometabolous insect other than Coleoptera, genomic information at Hymenoptera Genome Database92 (Hymenoptera) (GCF_000002195.4, Amel_4.5; GCF_000217595.1, Lhum_UMD_V04; GCF_000002325.3, Nvit_2.1), Lepbase93 (*Musca domestica* genome, GCA_000005575.1, AgamP3; GCA_002204515.1, AaegL5.0) was deposited in DDBJ under the accession numbers DRA002559, DRA006068, DRA007003, DRA007002, DRA007004, and DRA005777, respectively. The assembled genomic sequences were deposited in DDBJ under the accession numbers BHEG0200001 (H. axyridis genome assembly version 1), BHEG0200001 (H. axyridis genome assembly version 2), BHEF1000001 (H. axyridis linked-read genome assembly, h′, Nd7 strain), BHEF1000001 (H. axyridis linked-read genome assembly, h′, F2 hybrid), BHED1000001 (H. axyridis linked-read genome assembly, h′, F2 hybrid), BHEC0100001–BHEC0105573 (C. septempunctata linked-read genome assembly, MD8 strain), and AP018896–AP018898 (the H. axyridis gap-filled genomic scaffolds including the panner locus).

Received: 25 August 2017 Accepted: 15 August 2018

References

1. Vandenberg, N. J. Coccinellidae Latreille, 1807. In *American Beetles*. Vol. 2. Polyphaga: Scarabaeoidea through Curculionoidea (edsi Arnett, R. H. & Thomas, M. C.), 371–389. CRC Press, Boca Raton, FL.

2. Daleo, D., Braekman, J. C. & Pastels, 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, 1105–1112 (1996).

4. Dobzhansky, T. Die genetische und individuelle Variabilität von Harmonia axyridis Pallas in ihren Wechselbeziehungen. Biol. Zent. 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. Am. Nat. 68, 252–265 (1934).

6. Harlow, Y. G. Genetical studies of the lady-bird beetle, Harmonia axyridis PALLAS (Report II). Jpn. 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. Mosquitos in the inheritance of color patterns in the lady-bird, Harmonia Axyridis. Heredity 17, 401–421 (1996).

11. Bezzeraides, A. L., McGraw, K. J., Parker, R. S. & Hussein, 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., Kramer, K. J. & Denell, R. E. Repeated co-options of a mimicry supergene controlling butterfly mimicry. Nature 477, 203–206 (2011).

16. Wang, J. et al. A Y-like social chromosome causes alternative colony organization in ants. Nature 493, 664–668 (2013).

17. Kunte, K. et al. doubleshort causes extreme polyphenism in ladybirds. Nature 507, 229–232 (2014).

18. Nishikawa, H. et al. A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. Nature 47, 405–409 (2015).

19. Merrill, R. M. et al. The diversification of Heliconius butterflies: what have we learned in 150 years? J. Evol. Biol. 28, 1417–1435 (2015).

20. Davidson, E. H. & Guss, M. A. A genomic regulatory network for development. Science 295, 1669–1678 (2002).

21. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? Science 323, 746–751 (2009).

22. Niimi, T., Kuwayama, H. & Yaginuma, T. Larval RNAi applied to the analysis of postembryonic development in the ladybird beetle, Harmonia axyridis. J. Insect Biotech. Seriol. 74, 471–476 (2002).

23. Kawamoto, T. & Shimizu, M. A method for preparing 2- to 50-micrometer thick, fresh-frozen sections of large samples and undecalcified hard tissues. Histochem. Cell Biol. 113, 331–339 (2000).

24. Davis, M. W. Ap. http://biologylabs.utah.edu/jorgensen/wayned/ap/ (2018).

25. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

26. Hofmann, K. & Baron, M. Boxshade. http://www.ch.embnet.org/software/BOX_form.html (2018).

27. Ando, T., Kojima, T. & Fujiwara, H. Dramatic changes in pattern gene expression during metamorphosis are associated with the formation of a feather-like antenna by the silk moth, Bombyx mori. Dev. Genet. 35, 53–63 (2011).

28. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun reads. Genome Res. 24, 1384–1395 (2014).

29. Yang, X., Chockalingam, S. P. & Aluru, S. A survey of error-correction methods for next-generation sequencing. Nat. Rev. Genet. 13, 56–66 (2013).

30. Platanus trim. http://platanus.bio.titech.ac.jp/pltanus_trim (2018).

31. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35, 545–548 (2018).

32. Weisenfeld, N., Kumar, V., Shah, P., Church, D. & Jaife, D. Direct determination of diploid genome sequences. Nat. Biotechnol. 27, 757–767 (2017).

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

34. English, A. C. et al. Mind the gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. PLoS ONE 7, e47788 (2012).

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

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

37. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Nat. Protoc. Bioinformatics, 4, 11.10.1 (2013).

38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBet. J. 17, 10–12 (2011).

39. Picard. http://broadinstitute.github.io/picard/ (2018).

40. 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).

41. Kurtz, S. et al. Versatile and open software for large genome assemblies. Genome Biol. 5, R12 (2004).
71. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010 (2011).
72. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).
73. Slater, G. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6, 31 (2005).
74. Stajich, J. process_exonerate_gff3.pl. https://github.com/hyphaltip/genome-scraper/blob/master/data_format/process_exonerate_gff3.pl (2018).
75. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277 (2000).
76. Healey, J. Genbank_slicer.py. https://github.com/jrhealey/bioinfo-tools/blob/master/Genbank_slicer.py (2018).
77. Peterson, B. rtid. https://github.com/bbrant/rtid (2018).
78. Azends, D., Prins, P., Jansen, R. C. & Broman, K. W. R/qtl: high-throughput multiple QTL mapping. Bioinformatics 26, 2990–2992 (2010).
79. Taylor, J. & Butler, D. R. Package A5 Map: efficient genetic linkage map construction and diagnosis. J. Stat. Soft. 79, 1–29 (2017).
80. R. Core Team. R: a language and environment for statistical computing. https://www.R-project.org/ (2018).
81. Gibson, M. J. Jalmatata QTL mapping. https://github.com/gibsonMatt/jalmatataQTL (2018).
82. 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).
83. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
84. 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).
85. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
86. 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).
87. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25 (2010).
88. Robinson, T. J. et al. Integrative genomics viewer. Nat. Biotech. 29, 24–26 (2011).
89. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
90. NCBI Genome. https://www.ncbi.nlm.nih.gov/genome/ (2018).
91. Fireflybase. http://www.fireflybase.org/ (2018).
92. Hymenoptera Genome Database. http://hymenopteragenome.org/ (2018).
93. Lepbase. http://lepbase.org/ (2018).
94. SilkBase. http://silkbase.ab.a-u.tokyo.ac.jp/cgi-bin/index.cgi (2018).
95. Flybase. http://flybase.org/ (2018).
96. JASPAR. http://jaspar.genereg.net/ (2018).
97. Shannon, P. MotifDb: an annotated collection of protein-DNA binding sequence motifs. Bioconductor. https://doi.org/10.18129/B9.bioc.MotifDb (2017).
98. 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).
99. Thomas-Chollier, M. et al. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res. 40, e31 (2012).
100. Pagès, H. BSGenome: software infrastructure for efficient representation of full genomes and their SNPs. R package version 1.4.60. https://doi.org/10.18129/B9.bioc.BSGenome (2017).
101. Lawrence, M. et al. Software for computing and annotating genomic ranges. PLOS Comp. Biol. 9, e1003318 (2013).
102. Stojnic, R. & Diez, D. PWMEncipher: PWM enrichment analysis. R package version 4.14.0. https://doi.org/10.18129/B9.bioc.PWMEncipher (2015).
103. Stormo, G. D. DNA binding sites: representation and discovery. Bioinformatics 16, 16–23 (2000).
104. Katoh, K. & Standley, D. M. MAFFT: multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).
105. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).
106. 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).
107. Zhang, S. Q. et al. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nat. Commun. 9, 205 (2018).
108. Kumar, S., Stecher, G., Suleks, M. & Hedges, S. B. TimeTree: a resource for timelines, timescales, and divergence times. Mol. Biol. Evol. 34, 1812–1819 (2017).

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. Minatani for experimental support. 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. Computations were partially performed on the supercomputers at the Data Integration and Analysis Facility, National Institute for Basic Biology and at the ROIS National Institute of Genetics. 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, 20117014, 2613708, 22150002, 17H05848, 18H04828, JSPS KAKENHI Grant Number 22380035, and NIBB Collaborative Research Programs (18–433).

Author contributions

T.N. and T.A. conceived and designed this study. T.M. and T.N. analysed the eye larval pigmentation processes. T.A., T.M., K.G., K.H., A.I. and J.Y. analysed the sequence data. T.A., T.M., K.G., A.I. and J.H. performed cloning of the panner genes from different alleles and species of ladybirds. T.M., K.G., A.I. and J.H. performed the larval RNAi experiments. K.H., K.G. and J.H. performed the semi-quantitative RT-PCR. T.A. performed the in situ hybridisation. J.Y. collected the total RNA for the RNA-seq analysis, and the genomic DNA samples for the initial de novo genome assembly. T.A. collected the DNA samples for the ressequencing analyses. M.S. and Y.S. collected the RNA-seq raw data. Y.M. and A.T. performed the initial de novo genome assembly, K.Y. and S.S. collected the raw data for the ressequencing and the RAD-seq analyses. K.Y. constructed the flexible ddRAD-seq protocol. R.K., M.O. and T.I. performed reassembly of the genome, the de novo assembly of the linked-read genomic data, and validation of the obtained genomic scaffolds. M.K., T.T. and K.Y. performed mapping and quantification of the RNA-seq data. T.A. performed the data analyses for the genetic association studies, the gene annotation, the motif enrichment analysis, and the molecular phylogenetic analyses around the panner locus. T.A. and T.N. wrote, and all authors commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06116-1.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://www.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018