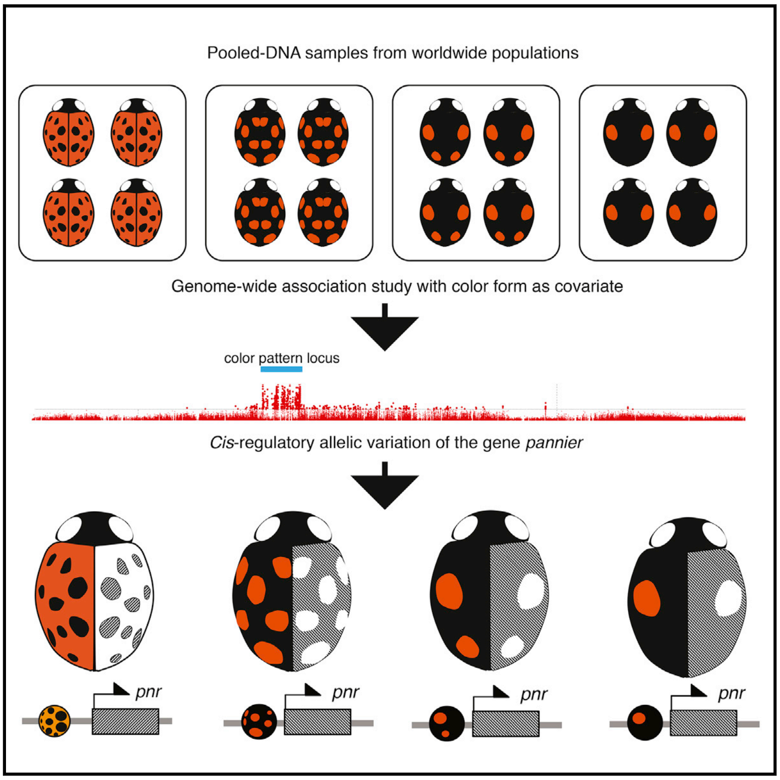
Report

Current Biology

The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird

Graphical Abstract


Pooled-DNA samples from worldwide populations
Genome-wide association study with color form as covariate
color pattern locus
Cis-regulatory allelic variation of the gene *pannier*
pnr
pnr
pnr
pnr

Authors

Mathieu Gautier, Junichi Yamaguchi, Julien Foucaud, ..., Heiko Vogel, Arnaud Estoup, Benjamin Prud'homme

Correspondence

arnaud.estoup@inra.fr (A.E.), benjamin.prudhomme@univ-amu.fr (B.P.)

In Brief

More than 200 distinct color forms have been described in natural populations of the harlequin ladybird, *Harmonia axyridis*. Gautier et al. show that this variation is controlled by the transcription factor Pannier. Pannier is necessary to produce black pigment, and its expression pattern prefigures the coloration pattern in each color form.

Highlights

- The harlequin ladybird displays various color pattern forms in natural populations
- The transcription factor Pannier controls this color polymorphism
- Pannier is essential for the formation of melanic elements on the elytra
- The *cis*-regulatory regions of *pannier* have diverged extensively among color forms

Gautier et al., 2018, Current Biology *28*, 3296–3302
October 22, 2018 © 2018 The Authors. Published by Elsevier Ltd.
https://doi.org/10.1016/j.cub.2018.08.023

The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird

Mathieu Gautier,[1,15] Junichi Yamaguchi,[2,15] Julien Foucaud,[1] Anne Loiseau,[1] Aurélien Ausset,[1] Benoit Facon,[1,10] Bernhard Gschloessl,[1] Jacques Lagnel,[1,11] Etienne Loire,[1,12,13] Hugues Parrinello,[3] Dany Severac,[3] Celine Lopez-Roques,[4] Cecile Donnadieu,[4] Maxime Manno,[4] Helene Berges,[5] Karim Gharbi,[6,14] Lori Lawson-Handley,[7] Lian-Sheng Zang,[8] Heiko Vogel,[9] Arnaud Estoup,[1,16,*] and Benjamin Prud'homme[2,16,17,*]

[1]CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier, Montpellier, France
[2]Aix Marseille Université, CNRS, IBDM, Marseille, France
[3]MGX, Biocampus Montpellier, CNRS, INSERM, Université de Montpellier, Montpellier, France
[4]INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France
[5]INRA, Centre National de Ressources Génomiques Végétales, 31326 Castanet-Tolosan, France
[6]Edinburgh Genomics, University of Edinburgh, Edinburgh, UK
[7]Evolutionary and Environmental Genomics Group, School of Environmental Sciences, University of Hull, Hull HU6 7RX, UK
[8]Institute of Biological Control, Jilin Agricultural University, Changchun, China
[9]Department of Entomology, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany
[10]Present address: UMR Peuplements Végétaux et Bioagresseurs en Milieu Tropical, INRA, Saint-Pierre, Réunion, France
[11]Present address: INRA, UR1052, Génétique et Amélioration des Fruits et Légumes (GAFL), Domaine Saint Maurice, 67 Allée des Chênes, CS 60094, 84143 Montfavet Cedex, France
[12]Present address: CIRAD, UMR ASTRE, 34398 Montpellier, France
[13]Present address: ASTRE, Université de Montpellier, CIRAD, INRA, Montpellier, France
[14]Present address: The Earlham Institute, Norwich Research Park, Norwich, UK
[15]These authors contributed equally
[16]These authors contributed equally
[17]Lead Contact
*Correspondence: arnaud.estoup@inra.fr (A.E.), benjamin.prudhomme@univ-amu.fr (B.P.)
https://doi.org/10.1016/j.cub.2018.08.023

SUMMARY

Many animal species comprise discrete phenotypic forms. A common example in natural populations of insects is the occurrence of different color patterns, which has motivated a rich body of ecological and genetic research [1–6]. The occurrence of dark, i.e., melanic, forms displaying discrete color patterns is found across multiple taxa, but the underlying genomic basis remains poorly characterized. In numerous ladybird species (Coccinellidae), the spatial arrangement of black and red patches on adult elytra varies wildly within species, forming strikingly different complex color patterns [7, 8]. In the harlequin ladybird, *Harmonia axyridis*, more than 200 distinct color forms have been described, which classic genetic studies suggest result from allelic variation at a single, unknown, locus [9, 10]. Here, we combined whole-genome sequencing, population-based genome-wide association studies, gene expression, and functional analyses to establish that the transcription factor Pannier controls melanic pattern polymorphism in *H. axyridis*. We show that *pannier* is necessary for the formation of melanic elements on the elytra. Allelic variation in *pannier* leads to protein expression in distinct domains on the elytra and thus determines the distinct color patterns in *H. axyridis*. Recombination between *pannier* alleles may be reduced by a highly divergent sequence of ∼170 kb in the *cis*-regulatory regions of *pannier*, with a 50 kb inversion between color forms. This most likely helps maintain the distinct alleles found in natural populations. Thus, we propose that highly variable discrete color forms can arise in natural populations through *cis*-regulatory allelic variation of a single gene.

RESULTS AND DISCUSSION

Ladybird species have long been studied by geneticists and evolutionary biologists to investigate the origin and maintenance of discrete color pattern forms in natural populations. In particular, the harlequin ladybird, *Harmonia axyridis*, is an emblematic species of elytral color pattern polymorphism, with more than 200 color pattern forms described from different localities [11, 12]. However, four forms dominate natural populations with high frequencies (Figure 1A) [14]: three distinct melanic forms harboring different patterns (from darkest to lightest, form [f.] *conspicua*, f. *spectabilis*, and f. *axyridis*, hereafter called Black-2Spots, Black-4Spots, and Black-nSpots, respectively) and a non-melanic form (f. *succinea*, called Red-nSpots). The striking array of color patterns documented in *H. axyridis* in the wild has been attributed to a combination of allelic diversity, interactions between allelic forms, and plastic response to environmental factors [11, 14]. Genetic crosses have demonstrated

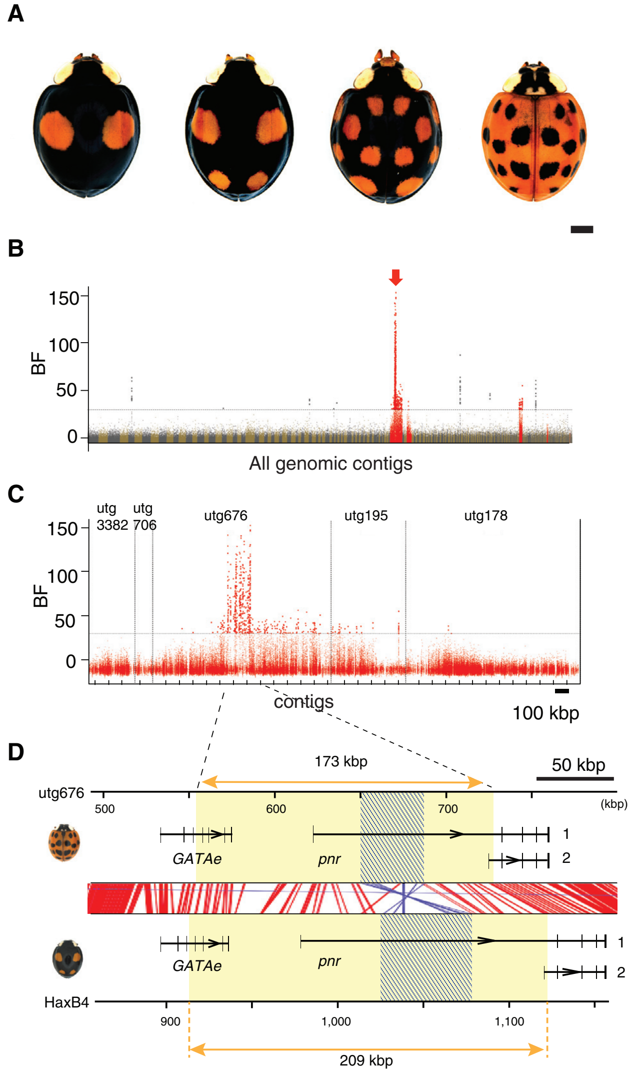


Figure 1. Genome-wide Association Study Identifies the Main Color Pattern Locus in *H. axyridis*

(A) The four most frequent color pattern forms of *H. axyridis*. From left to right: the form Black-2Spots (f. *conspicua*), Black-4Spots (f. *spectabilis*), Black-nSpots (f. *axyridis*), and Red-nSpots (f. *succinea*).

(B) Manhattan plots of genome-wide association for the proportion of Red-nSpots individuals in 14 DNA pooled samples of wild *H. axyridis* populations, with Bayes factor (BF) for individual SNPs. The horizontal dashed line indicates the 30 deciban (db) threshold. SNPs above this threshold are highlighted, and those assigned to contig utg676 (red arrow, containing the color pattern locus) in the *HaxR* assembly and four neighboring contigs are shown in red. Contigs are ordered by length, and only autosomal contigs are shown.

(C) Same as (B), with a focus on SNPs belonging to the five neighboring contigs including and surrounding the color pattern locus of the *HaxR* assembly (in red in Figure 1B). The relative ordering of these contigs was derived from the *de novo* sequencing of the Black-4Spots allele extended region (see STAR Methods).

(D) The gene content at the identified color pattern locus. Fifty-six SNPs with the strongest association signal delimit a candidate color pattern locus region of ~170 kb (yellow boxes) that extends from the first coding exon of *pannier* (*pnr*) to the 5′ upstream gene *GATAe*. Red and blue lines show conserved sequence blocks in forward or reverse direction, respectively, detected by [13]. The first intron of *pannier* contains the footprint of an ~50 kb inversion (shaded boxes). Two alternative splice variants of *pnr* are produced (named 1 and 2).

See also Figures S1, S3, and S4.

that the majority of *H. axyridis* melanic forms result from variation of multiple alleles segregating at a single, uncharacterized, autosomal locus [9, 10], hereafter referred to as the color pattern locus.

To identify this color pattern locus and the mechanisms underlying discrete color pattern variation, we used a population genomics approach, taking advantage of the co-occurrence of multiple color pattern forms in natural populations. To that end, we first performed a *de novo* genome assembly of the *H. axyridis* Red-nSpots form (*HaxR*) using long reads produced by a MinION sequencer (Oxford Nanopore) (Table S1). Then, to fine map the color pattern locus on this assembly, we sequenced, on a HiSeq 2500 (Illumina), DNA from 14 pools of individuals that were representative of the world-wide genetic diversity (i.e., eight different geographic origins) and the four main color pattern forms of *H. axyridis.* Our aim was to characterize genetic variation associated with phenotypic differences across pool samples, using the proportion of individuals of a given color form in each pool as a covariate. Table S2 shows how individuals were split both by color form and by geographic location for most pools in order to maximize the proportion of alleles of a given color form in the pools (n = 40 to n = 100 individuals per pool). Because the trait is monogenic (and autosomal), the highest mapping power would have been achieved if all of the color pattern alleles were co-dominant (i.e., pool frequencies for each allele would then be directly derived from the color pattern of the pooled individuals). However, previous work demonstrated that the observed color pattern of an individual imperfectly predicts its genotype at the underlying color pattern locus due to the hierarchical dominance between all of the color pattern alleles, with Black-2Spots > Black-4Spots > Black-nSpots > Red-nSpots [9, 11, 14]. Since the Red-nSpots allele is the most recessive one (with seven pools including 100% of Red-nSpots individuals; Table S2), we performed a genome-wide association study, following Gautier [15], to identify SNPs associated with the proportion of Red-nSpots individuals in each sequenced pool as a covariate to achieve the highest mapping power. Among 18,425,210 SNPs we called on the 457 autosomal contigs (totaling 377.5 Mb), we found 710 SNPs strongly associated with the proportion of the Red-nSpots form (Bayes factor > 30 deciban [db]), the vast majority (86%) of which are located within a single 1.3 Mb contig, *utg676* (Figures 1B and 1C). The 56 SNPs with the strongest association signals (Bayes factor > 100 db) delineate an ~170 kb region on *HaxR*, representing the strongest candidate region for the color pattern locus. Importantly, additional genome-wide association studies using the proportions of Black-4Spots, Black-2Spots, or Black-nSpots individuals in the pools as covariates pointed exclusively to the same region (Figure S1), although these analyses were less powerful.

The candidate color pattern locus extends from the first coding exon of the ortholog of the *Drosophila* gene *pannier*, including its first intron and first non-coding exon, to the end of the neighboring 5′ gene, the *GATAe* ortholog (Figure 1D). To test a possible role of *pannier* or *GATAe* in adult color pattern formation, we used RNA interference (RNAi) [16]. Because adult pigmentation patterns are specified during pupal development in insects, we injected larvae of the different *H. axyridis* forms, just before pupation, with double-stranded RNA (dsRNA) targeting the coding sequences of *pannier* or *GATAe*. We also targeted

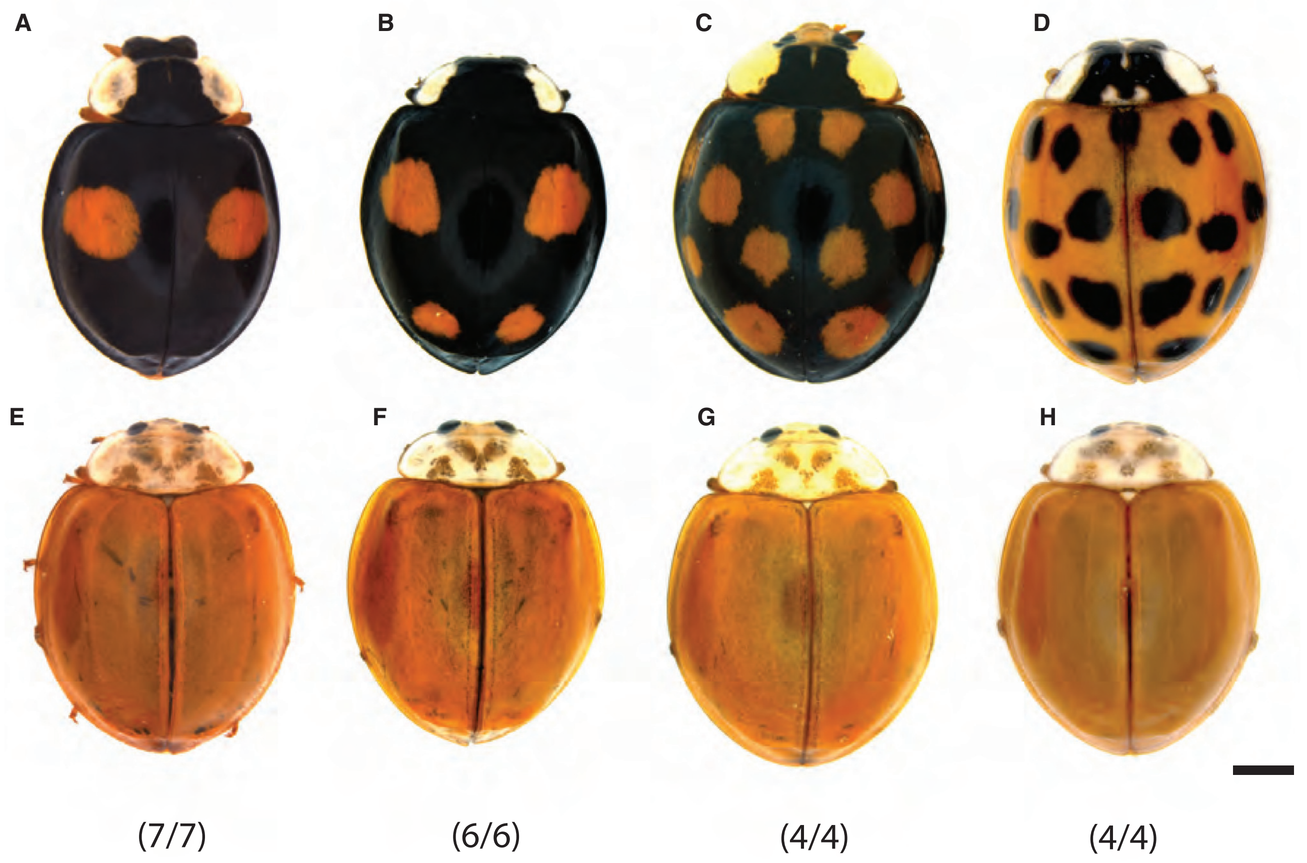


Figure 2. ***pannier*** **Is Necessary for Black Pigment Production in** ***H. axyridis***
Wild-type color pattern forms (top) and representative phenotypes for each form when knocking down *pannier* by larval RNAi (bottom). Black-2Spots (A and E), Black-4Spots (B and F), Black-nSpots (C and G), and Red-nSpots (D and H) are shown. Numbers indicate the fraction of eclosed adults with the representative phenotype for each form. Scale bars, 1 mm. See also Figures S2 and S3.

eGFP as a negative control. Targeting of *GATAe* or *eGFP* had no effect on pigmentation, both in the Red-nSpots and in the Black-4Spots forms, suggesting that *GATAe* does not play any role in elytral pigmentation (Figures S2A and S2B). In contrast, knockdown of *pannier* dramatically reduced the formation of black pigment in all different forms, resulting in adults with almost homogeneous red elytra (Figure 2). Dark pigment formation is strongly reduced not only in the elytra, but also in the head and the rest of the body. Use of a different, non-overlapping dsRNA fragment of *pannier* (Figure S3) produced similar results, ruling out RNAi off-target effects (Figures S2C and S2D). These results show that *pannier* is necessary for the formation of black pigment in *H. axyridis* adults. Furthermore, combined with our genome-wide association study, our data indicate that *pannier* is the main gene responsible for color pattern polymorphism in *H. axyridis* and that different *pannier* alleles determine the color pattern in the different forms.

To understand how *pannier* contributes to the formation of different color patterns, we compared its coding sequences between the Red-nSpots and Black-4Spots forms. We did not find any non-synonymous mutation, thus ruling out changes in Pannier protein composition (Figure S3A). We next hypothesized that *pannier* might have evolved divergent expression patterns during the development of the elytra, resulting in different color pattern forms. We therefore compared *pannier* expression level by qRT-PCR in late-developing pupal elytra between Red-nSpots and Black-4Spots forms. We found that *pannier* is expressed at a higher level in the elytra of the Black-4Spots form compared to the Red-nSpots form (Figure 3A). In order to determine how this difference reflected in Pannier spatial expression pattern, we compared the relative *pannier* expression levels in different parts of a Black-4Spots elytron. We found that *pannier* is expressed at a higher level in a presumptive black area, in the middle of the elytron, compared to the presumptive red areas (Figure 3B). In order to map these differences onto spatial expression patterns, we stained late pupal elytra with an antibody raised against *H. axyridis* Pannier. We found that Pannier spatial distribution on the elytra is different between color pattern forms (Figures 3C–3F). Strikingly, in all forms, areas with the strongest Pannier expression levels prefigure the adult elytral pattern of melanic elements. This tight spatial correlation, coupled with our genomic association study and the essential role of *pannier* in governing melanic patterns, provides strong evidence that *cis*-regulatory changes at the *pannier* locus drive divergent *pannier* expression patterns and, in turn, the polymorphic melanic patterns of *H. axyridis*.

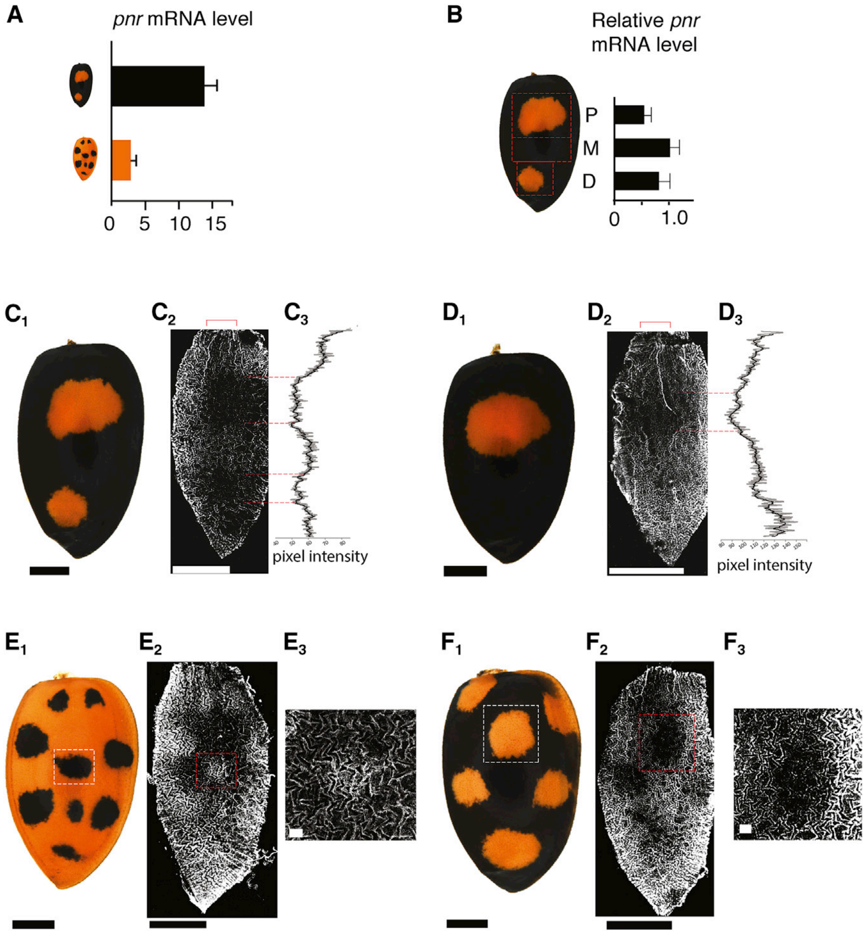


Figure 3. *pannier* Expression Pattern Determines Melanic Color Pattern in Each *H. axyridis* Form

(A) qPCR of *pannier* mRNA from whole elytra reveals the difference between the Black-4Spots and Red-nSpots forms (two-tailed t test; $n \geq 6$; $p = 0.0005$). Error bars indicate the SEM.

(B) *pannier* is expressed at different levels in the presumptive red or black elytral areas in the Black-4Spots form (P, proximal; M, medial; D, distal) ($n = 5$; two-tailed paired t test, $p = 0.02$ [P versus M], $p = 0.005$ [M versus D]). Error bars indicate the SEM.

(C–F) Immunodetection of Pannier protein in each color form (~96 hr after pupation). (C_1)–(F_1) show adult elytron; (C_2)–(F_2) show anti-Pannier staining; (C_3) and (D_3) show pixel intensity along the PD axis of the elytron in the region delineated with a bracket in (C_2) and (D_2), respectively; and (E_3) and (F_3) show higher magnification of the selected areas in (E_2) and (F_2), respectively. Note the reduction of signal intensity in the presumptive red areas in each form. Scale bars, 1 mm, 100 μm for insets.

In addition to allelic variation, color pattern diversity in *H. axyridis* is shaped by the dominance relationships among color form alleles [11]. Indeed, similarly to other species (e.g., [17]), heterozygous individuals resulting from the cross of distinct homozygous *H. axyridis* forms produce black pigmentation in any part of the elytra that is black in either parental form (Figure 4) [11]. Our results explain this phenomenon, known as mosaic dominance, at the molecular level. Since elytral Pannier expression patterns mimic adult black pigmentation patterns, and since each *pannier* allele carries its own *cis*-regulatory determinant to drive specific expression pattern, the expression pattern of *pannier* in heterozygotes reflect the sum of individual form patterns (Figure 4). In other words, the mosaic expression of *pannier* during development, driven by the heterozygous alleles, produces a mosaic pigmentation pattern in adults. This phenomenon compounds the effect of *pannier* allelic variation, increasing the complexity of color patterns polymorphism in *H. axyridis*. We suggest that this explanation can account for other cases of mosaic dominance, providing a simple and general mechanism to expand the pigmentation pattern repertoire of any species.

Finally, in order to precisely compare the sequences of *pannier* alleles between the Red-nSpots and another form, we sequenced the Black-4Spots form *de novo*. We chose the Black-4Spots form because it seemed the most divergent when compared to the Red-nSpots, based on genome-wide association results (Figure S1) and careful inspection of read coverage of the pools representative of color forms (Figure S4). We generated the Black-4Spots draft assembly (*HaxB4*) from Illumina sequencing reads (Table S1) and supplemented it by targeted bacterial artificial chromosome (BAC) sequencing to derive a high-quality 2.87 Mb sequence (% N = 0.96) spanning *pannier* and adjacent genomic regions. Strikingly, we found that the Red-nSpots and Black-4Spots sequences of the 5′ non-coding DNA and the first intron of *pannier* align poorly, in contrast to adjacent regions (Figures S3B and S3C). Furthermore, we detected the footprint of an ~50-kb-long inversion within the first intron of *pannier* (Figure 1D, Figure S3C). This comparison indicates that the sequences of the non-coding DNA of *pannier* have diverged extensively between these forms.

Altogether, our results conclusively show that *pannier* plays a key role in the specification and the diversification of the main color pattern forms in *H. axyridis*. *pannier* has never been reported to play a role in pigmentation in insects. It has therefore been co-opted for this function in the lineage leading to *H. axyridis*, presumably through the evolution of new regulatory connections with downstream effector genes directly involved in black pigment production [18]. This contrasts with other insect groups, including butterflies and fruit flies, in which different regulatory genes have been co-opted to generate wing color patterns [1, 2, 18, 19]. Furthermore, we show that polymorphic color patterns in *H. axyridis* arise from differential regulation of *pannier* spatial expression. The divergent non-coding sequences we have identified in the *H. axyridis* color pattern forms are quite large (~170 kb) and may host multiple discrete *cis*-regulatory elements. We propose that these *cis*-regulatory elements produce different *pannier* expression patterns and, ultimately, discrete melanic patterns, by interpreting differently a *trans*-regulatory

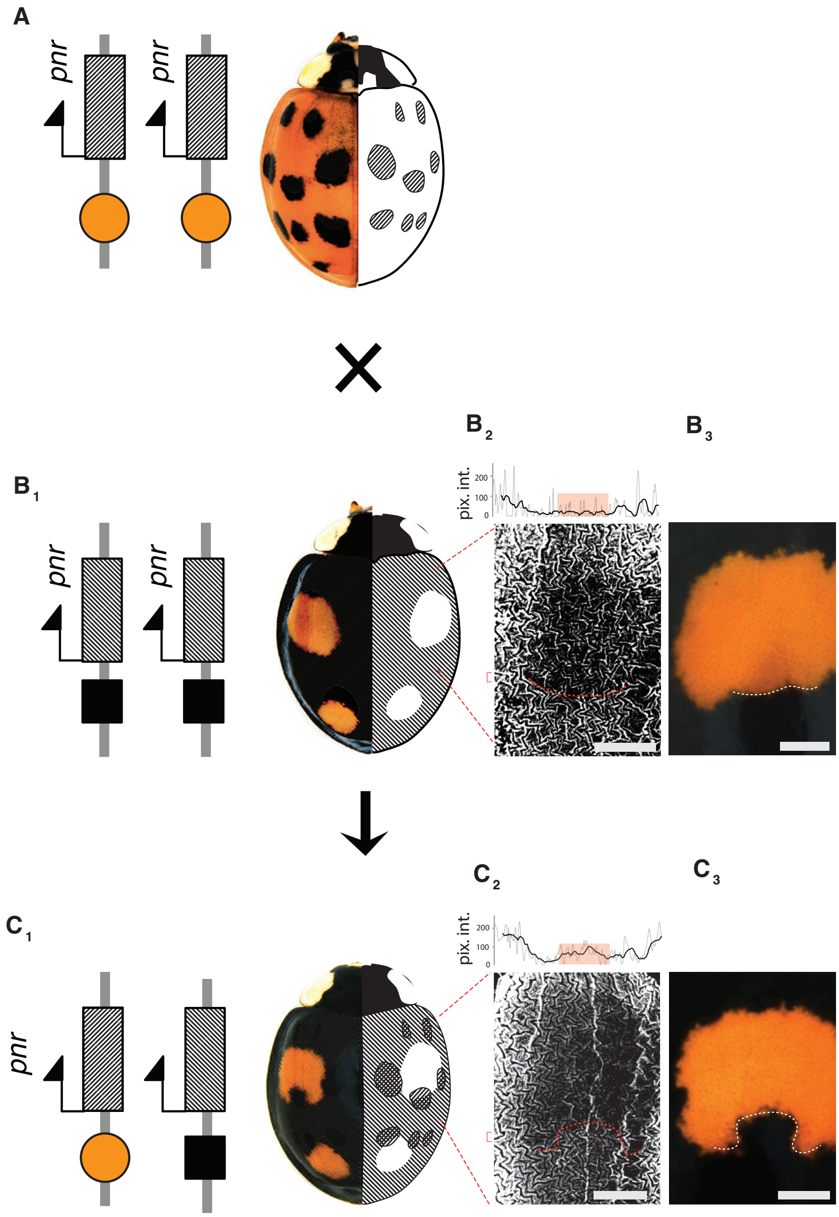


Figure 4. Genetic Basis of the Mosaic Dominance

Each color pattern form is determined by a particular *cis*-regulatory allele of *pannier* (symbolized with orange circles and black squares). In heterozygous individual carrying a *pnr* allele from the Red-nSpots form (A) and a *pnr* allele from the Black-4Spots form (B$_1$), the mosaic color pattern reflects *pannier* expression pattern, which is the sum of the two allelic forms. For each individual in (A), (B$_1$), and (C$_1$), the left elytron shows Pannier expression pattern, and the right elytron shows the corresponding adult pigmentation pattern. (B$_2$) and (C$_2$) show Pannier expression pattern in the presumptive area shown in (B$_3$) and (C$_3$), respectively. Note that in the heterozygote (C$_1$), Pannier expression pattern prefigures the ectopic formation of black pigment in the anterior red patch (red dotted lines in B$_2$ and C$_2$). (B$_2$) is a higher magnification image of Figure 3C$_2$. Graphs above (B$_2$) and (C$_2$) represent pixel intensity along the medial-lateral axis in the region delineated with a red bracket. Scale bars, 0.5 mm.

landscape that is common to all *H. axyridis* color pattern forms. This model is reminiscent of the mechanisms underlying thoracic bristles or wing pigmentation pattern diversity in *Drosophila* species [18, 20, 21] and that have been proposed to explain wing color pattern evolution in butterflies [1].

We have demonstrated that the *cis*-regulatory region of *pannier* is highly divergent between the Red-nSpots and Black-4Spots alleles and that it includes an ~50 kb inversion. Furthermore, our data provide evidence of large-scale sequence variation among all four alleles of the main color pattern forms in natural populations (Figure S4). These results are in agreement with a recent, independent study [22]. We hypothesize that the numerous, rare color pattern forms that have been described in *H. axyridis* [11] are also determined by *pannier cis*-regulatory variation and that they might result from rare mutational events, including rare recombination between the alleles of the main forms.

The striking sequence divergence among the *pannier* alleles of the main color forms brings into question their evolutionary origin. Possibilities include ancient mutational events or even across species introgression events [10, 23]. Thorough characterization of the *pannier* genomic region from different color pattern forms both within *H. axyridis* and across coccinelid species (especially those harboring color pattern polymorphisms [7]) will illuminate the evolutionary origin of the genomic determinants of color pattern forms in *H. axyridis* and other ladybird species.

Finally, if sequence divergence helps preserve distinct *pannier* alleles by reducing recombination among them, selective mechanisms are also suspected to maintain different color forms in natural populations of *H. axyridis* and to affect their frequencies. Both local adaptation to climatic factors and seasonal variations in temperature have been suggested to affect color forms proportions in space and time, possibly mediated by mate choice [10, 12, 24]. The identification of the genomic basis of color pattern polymorphism will help to better characterize the evolutionary mechanisms that shape the striking color pattern diversity in natural populations of *H. axyridis* and to reveal potential pleiotropic effects of *pannier* alleles on traits involved in survival and reproduction [24, 25].

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, two tables, and two data files and can be found with this article online at https://doi.org/10.1016/j.cub.2018.08.023.

ACKNOWLEDGMENTS

We thank G. Beydon for constructing and screening the BAC library; L. Sauné and M. Galan for assistance in MiSeq sequencing; E. Lombaert, I. Goryacheva, R. Koch, J. Lundgren, O. Nedved, and C.E. Thomas for help in sampling *H. axyridis* populations; A. Tayeh for assistance in lab-rearing of *H. axyridis*; and A. Bretaudeau for making the *H. axyridis* genome assemblies publicly accessible. We are grateful to N. Gompel, N. Rode, Ruth Hufbauer, and R. Vitalis for comments on the manuscript. A.E., B.F., J.F., and M.G. acknowledge financial support from the national funding agency ANR (France) through the European Union program ERA-Net BiodivERsA (project EXOTIC) and the GenoToul bioinformatics platform Toulouse Midi-Pyrenees for providing computing resources. B.P. and J.Y. acknowledge financial support from the CNRS, the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC grant agreement no. 615789, and the France-BioImaging infrastructure supported by the Agence Nationale de la Recherche (ANR-10-INSB-04-01, call "Investissements d'Avenir"), as well as the support of the NIG supercomputer at ROIS National Institute of Genetics (Japan). C.L.P., C.D., and M.M. acknowledge financial support from the France Génomique National infrastructure, funded as part of "Investissement d'Avenir" program managed by Agence Nationale de la Recherche (contract ANR-10-INBS-09).

AUTHOR CONTRIBUTIONS

M.G. conceived the project, designed the study, carried out bioinformatics and statistical treatments for genome-wide association studies and the identification of *HaxR* autosomal contigs, performed or supervised bioinformatic analyses and BAC contig construction, and wrote the manuscript. J.Y. carried out larval RNAi, cDNA sequencing, qPCR, immunohistochemistry, and imaging studies; processed various bioinformatics treatments; and wrote the manuscript. J.F. helped design the study and contributed to obtaining and maintaining various *H. axyridis* populations in the lab. A.L. processed molecular work for the *de novo* assembly, genome-wide association studies, and BAC library PCR screening. A.A. contributed to obtaining and maintaining various *H. axyridis* populations in the lab. B.F. helped design the study and obtained funding. B.G. analyzed the quality of the *de novo* genome assemblies and annotated the coding genes. J.L. designed the MinION study and processed associated bioinformatics treatments leading to the *HaxR de novo* assembly. E.L. processed bioinformatics treatments for the *HaxB4 de novo* assembly. H.P. and D.S. produced pool-sequencing (pool-seq) and individual next-generation sequencing (NGS) data. C.L.R., C.D., and M.M. helped designed the MinION study, produced the MinION data, and processed upstream bioinformatics treatments. H.B. developed and helped screening the Black-4Spots BAC library. K.G. helped design the study and produced some NGS data to construct the *HaxB4 de novo* assembly. L.L.H. provided *H. axyridis* individuals from Japan and contributed to drafting the manuscript. L.S.Z. provided *H. axyridis* individuals from China. H.V. helped design the study, obtained funding, and produced RNA-seq data and genomic resources for the *HaxB4 de novo* assembly. B.P. and A.E. designed and directed the project, obtained funding, interpreted results, and wrote the manuscript. All authors commented on the manuscript.

DECLARATION OF INTEREST

The authors declare no competing interests.

Received: July 12, 2018
Revised: July 25, 2018
Accepted: August 7, 2018
Published: August 23, 2018

REFERENCES

1. Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science *333*, 1137–1141.
2. Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D.H., Martin, A., Reed, R.D., Mullen, S.P., and Kronforst, M.R. (2014). doublesex is a mimicry supergene. Nature *507*, 229–232.
3. Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W.R., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., et al. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. Nature *534*, 106–110.
4. Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., and Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. Nature *534*, 102–105.
5. Yassin, A., Delaney, E.K., Reddiex, A.J., Seher, T.D., Bastide, H., Appleton, N.C., Lack, J.B., David, J.R., Chenoweth, S.F., Pool, J.E., and Kopp, A. (2016). The pdm3 Locus Is a Hotspot for Recurrent Evolution of Female-Limited Color Dimorphism in Drosophila. Curr. Biol. *26*, 2412–2422.
6. Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J., Dasmahapatra, K.K., and Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. Curr. Biol. *28*, 1839–1845.e3.
7. Majerus, M.E.N. (1994). Ladybirds (Collins New Naturalist).
8. Majerus, M.E.N. (1998). Melanism: Evolution in Action (Oxford University Press).
9. Tan, C., and Li, J. (1934). Inheritance of the elytral color patterns of the lady-bird beetle, Harmonia axyridis Pallas. Am. Nat. *68*, 252–265.
10. Komai, T. (1956). Genetics of ladybirds. Adv. Genet. *8*, 155–188.
11. Tan, C. (1946). Mosaic dominance in the inheritance of color patterns in the lady-bird beetle, Harmonia axyridi. Genetics *31*, 195–210.
12. Dobzhansky, T. (1933). Geographical variation in ladybeetles. The American Naturalist *67*, 97–126.

13. Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., and Parkhill, J. (2005). ACT: the Artemis Comparison Tool. Bioinformatics *21*, 3422–3423.

14. Michie, L.J., Mallard, F., Majerus, M.E.N., and Jiggins, F.M. (2010). Melanic through nature or nurture: genetic polymorphism and phenotypic plasticity in Harmonia axyridis. J. Evol. Biol. *23*, 1699–1707.

15. Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics *201*, 1555–1579.

16. Niimi, T., Kuwayama, H., and Yaginuma, T. (2005). Larval RNAi applied to the analysis of postembryonic development in the ladybird beetle, Harmonia axyridis. Sanshi Konchuu Baiotekku *74*, 95–102.

17. Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. Nat. Commun. *5*, 5644.

18. Arnoult, L., Su, K.F.Y., Manoel, D., Minervino, C., Magriña, J., Gompel, N., and Prud'homme, B. (2013). Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. Science *339*, 1423–1426.

19. Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481–487.

20. Marcellini, S., and Simpson, P. (2006). Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. PLoS Biol. *4*, e386.

21. Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution *62*, 2155–2177.

22. Ando, T., Matsuda, T., Goto, K., Hara, K., Ito, A., Hirata, J., Yatomi, J., Kajitani, R., Okuno, M., Yamaguchi, K., et al. (2018). Repeated inversions at the *pannier* intron drive diversification of intraspecific colour patterns of ladybird beetles. bioRxiv. https://doi.org/10.1101/347906.

23. Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature *487*, 94–98.

24. Wang, S., Michaud, J.P., and Zhang, R.Z. (2009). Seasonal cycles of assortative mating and reproductive behaviour in polymorphic populations of Harmonia axyridis in China. Ecol. Entomol. *34*, 483–494.

25. True, J.R. (2003). Insect melanism: the molecules matter. Trends Ecol. Evol. *18*, 640–647.

26. Havens, L.A., and MacManes, M.D. (2016). Characterizing the adult and larval transcriptome of the multicolored Asian lady beetle, Harmonia axyridis. PeerJ *4*, e2098.

27. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. *13*, e1005595.

28. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. Bioinformatics *34*, 2666–2669.

29. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

30. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

31. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

32. Benaglia, T., Chauveau, D., Hunter, D., and Young, D. (2009). mixtools: an r package for analyzing mixture models. J. Stat. Softw. *32*, 1–29.

33. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

34. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

35. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

36. Hivert, V., Leblois, R., Petit, E.J., Gautier, M., and Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. Genetics. Published online July 30, 2018. https://doi.org/10.1534/genetics.118.300900.

37. Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS ONE *6*, e17288.

38. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

39. Ribeiro, F.J., Przybylski, D., Yin, S., Sharpe, T., Gnerre, S., Abouelleil, A., Berlin, A.M., Montmayeur, A., Shea, T.P., Walker, B.J., et al. (2012). Finished bacterial genomes from shotgun sequence data. Genome Res. *22*, 2270–2277.

40. English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., and Gibbs, R.A. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS ONE *7*, e47768.

41. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.

42. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

43. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-seq. Bioinformatics *25*, 1105–1111.

44. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

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

46. Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y., and Tsuda, M. (2008). GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics *9*, 376.

47. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

48. Bidon, T., Schreck, N., Hailer, F., Nilsson, M.A., and Janke, A. (2015). Genome-wide search identifies 1.9 Mb from the polar bear Y chromosome for evolutionary analyses. Genome Biol. Evol. *7*, 2010–2022.

49. Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., Loomans, A., Malausa, T., Nedved, O., Rhule, E., et al. (2014). Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird Harmonia axyridis. Mol. Ecol. *23*, 5979–5997.

50. Good, I.J. (1985). Weight of evidence: a brief survey. 1985;2:249–270. Bayesian Stat. *2*, 249–270.

51. Gschloessl, B., Dorkeld, F., Berges, H., Beydon, G., Bouchez, O., Branco, M., Bretaudeau, A., Burban, C., Dubois, E., Gauthier, P., et al. (2018). Draft genome and reference transcriptomic resources for the urticating pine defoliator Thaumetopoea pityocampa (Lepidoptera: Notodontidae). Mol. Ecol. Resour. *18*, 602–619.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti_Pannier	This study	N/A
Alexa Fluor 568 goat anti-rabbit IgG (H+L)	ThermoFisher	CAT#A-11011; RRID: AB_143157
Bacterial and Virus Strains		
E. coli strain DH5A	Widely distributed	N/A
Biological Samples		
H. axyridis	Jilin, China	N/A
H. axyridis	Changchun, China	N/A
H. axyridis	Kyoto, Japan	N/A
H. axyridis	Novosibirsk, Russia	N/A
H. axyridis	Bourgogne, France	N/A
H. axyridis	Georgia, USA	N/A
H. axyridis	Washington, USA	N/A
H. axyridis	BIOTOP biocontrol population, France	N/A
H. axyridis, Red-nSpots	Mississippi, USA	N/A
H. axyridis, Black-4Spots	BIOTOP biocontrol population, France	N/A
Chemicals, Peptides, and Recombinant Proteins		
Vectashield	vector	CAT#H-1200
Tape	Tesa	CAT#Eco&Strong
BSA	Sigma	CAT#A9418
TRIzol	ThermoFisher	CAT#15596026
RQ1 DNase	Promega	CAT#M6101
Critical Commercial Assays		
Ligation Sequencing Kit 1D	Oxford Nanopore Technologies	CAT#SQK-LSK108
Genomic-tip 500/G	QIAGEN	CAT#10262
NEBNext FFPE Repair Mix	New England BioLabs	CAT#M6630S
NEBNext End repair / dA-tailing Module	New England BioLabs	CAT#E7546
TruSeq Nano DNA Library Preparation Kit	Illumina	CAT#FC-121-4001
Nextera Mate Pair Library Preparation Kit	Illumina	CAT#FC-132-1001
First Strand cDNA Synthesis Kit	NEB	CAT#E6300S
T7 RiboMax	Promega	CAT#P1700
Power SYBR Green PCR Master Mix	ThermoFisher	CAT#4367659
Deposited Data		
HaxR whole genome assembly	This paper	http://bipaa.genouest.org/sp/harmonia_axyridis
HaxB4 whole genome assembly	This paper	http://bipaa.genouest.org/sp/harmonia_axyridis
Pool-Seq data (14 pools)	This paper	SRA: PRJNA474099
RNA-Seq data	[26]	SRA: PRJEB13023
BAC library	This paper	https://cnrgv.toulouse.inra.fr/Library/Asian-ladybird
Experimental Models: Organisms/Strains		
Black-2Spots	Gard, France	N/A
Black-4Spots	Gard, France	N/A
Black-nSpots	Ol. Nedved’s Czech Republic	N/A
Red-nSpots	Gard, France	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
A list of oligonucleotides is given in Data S1	N/A	N/A
Recombinant DNA		
pIndigoBAC-5 HindIII-Cloning Ready vector	Epicenter (Madison, USA)	BACB085H
pGEM-T Easy	Promega	CAT#A1360
Software and Algorithms		
Minknow ONT software	Oxford Nanopore Technologies	v1.7.10 or v1.7.14
Albacore	Oxford Nanopore Technologies	v1.2.4 or v2.0.2
Porechop	[27]	v0.2.3
Nanofilt	[28]	v1.1.3
canu	[29]	v1.6
pilon	[30]	v1.22
BUSCO	[31]	v3
mixtools	[32]	v1.1.0
bwa	[33]	v0.7.12
samtools	[34]	v1.3.1
VarScan	[35]	v2.3.4
PoolFstat	[36]	v0.1
BayPass	[15]	v2.1
deconseq	[37]	v0.4.3
Trimmomatic	[38]	v0.22
AllPath-LG	[39]	N/A
PBJelly	[40]	v1.3.1
Mummer	[41]	v3.0
Blast+	[42]	V2.6.0
Tophat (2.1.0)	[43]	v2.1.0
Cufflinks (2.2.1)	[44]	v2.2.1
Integrative Genomics Viewer	[45]	N/A
GenomeMatcher	[46]	N/A
Artemis Comparison Tool	[13]	N/A
Photoshop	Adobe	N/A
Illustrator	Adobe	N/A
ImageJ 1.51	[47]	N/A
HeliconFocus	Helicon Soft	N/A
Other		
Statistics and information regarding genome sequencing, pool-seq librairies preparation & sequencing, and primers	This paper	Data S1
HaxR assembly contigs chromosome type assignation	This paper	Data S2

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Benjamin Prud'homme (benjamin.prudhomme@univ-amu.fr).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Harmonia axyridis individuals were collected in the wild from eight geographic locations (Jilin-China, Changchun-China, Kyoto-Japan, Novosibirsk-Russia, Bourgogne-France, Georgia-USA, Washington-USA, and BIOTOP biocontrol population - France) to carry out genome-wide scans studies for association with the proportion of individuals of a given color pattern form in pool samples. Twelve Red-nSpots *H. axyridis* males from a laboratory reared population founded by Red-nSpots individuals originating from the biocontrol population BIOTOP (France) were used to build Oxford Nanopore Technologies (ONT) libraries. Three Red-nSpots males and three Red-nSpots females originating from a wild *H. axyridis* population from Mississippi (USA) were individually sequenced to identify the autosomal contigs of the *HaxR* assembly. We carried out four generations of full-sib crossings to obtain a Black-4Spots inbred line (origin: BIOTOP biocontrol population, France) in order to provide biological material for the *de novo* sequencing of the Black-4Spots allele and to build a Black-4Spots BAC library. Four *H. axyridis* strains homozygous for the color forms Red-nSpots, Black-2Spots (origin: Gard, France), Black-4Spots (origin: Gard, France) and Black-nSpots (origin: Oldrich Nedved's laboratory, Czech Republic) were produced and maintained in the laboratory to produce biological material for larval RNAi, qPCR and immuno-histochemistry experiments. Laboratory rearing conditions remained constant (24°C, 60% RH; L:D 14:10), individuals were fed *ad libitum* with irradiated eggs of *Ephestia kuehniella*.

METHOD DETAILS

Description and naming of the four color pattern forms that predominate in frequency in natural *Harmonia axyridis* populations

(i) form *conspicua* (hereafter named "Black-2Spots" for clarity) has black background color of elytra with two red spots on each elytron and the top one larger than the bottom one. (ii) f. *spectabilis* (hereafter "Black-4Spots") has black background of elytra with one large red spot in the top-center of each elytron, (iii) f. *axyridis* (hereafter "Black-nSpots") has black background color of elytra with many red spots, and (iv) f. *succinea* (hereafter "Red-nSpots") has red background color of elytra with the number of black spots ranging from 0 to 19. See Figure 1A of main text for illustrations.

De novo assembly of the *Harmonia axyridis* genome from individuals of the color pattern form Red-nSpots

Four Oxford Nanopore Technologies (ONT) libraries were prepared using the Ligation Sequencing Kit 1D (SQK-LSK108 ONT), according to the manufacturer's protocol 1D Genomic DNA by ligation (SQK-LSK108). Briefly, 60 μg of genomic DNA was extracted using the Genomic-tip 500/G kit (QIAGEN) from a pool of thorax tissue belonging to 12 Red-nSpots males from a lab-reared population founded by Red-nSpots individuals originating from the biocontrol population BIOTOP (France). The DNA sample was divided into four aliquots and sheared into 25 kb (n = 3) or 30 kb (n = 1) fragments using the Megaruptor system (Diagenode). One of the 25 kb sheared DNA sample was additionally size-selected prior to library preparation using the BluePippin system (Sage Science, Beverly, USA) to remove fragments smaller than 10 kb. A DNA Repair (NEBNext FFPE Repair Mix M6630) step as well as an End-repair and dA-tail step (NEBNext End repair / dA-tailing Module E7546) were then processed on 0.34 pmols of the sheared DNA sample, followed by ligation of sequencing adapters. Then 0.07 pmols of library were loaded onto an R9.5 flow cell containing at least 800 active pores and run for 48 hr, on a MinION sequencer (ONT) and sequenced by the Minknow ONT software (v1.7.10 or v1.7.14). Base calling was carried out using Albacore (from ONT, v1.2.4 or 2.0.2) with default settings yielding 2.46x10^6 reads (22.9 Gb) (Data S1) that were trimmed using Porechop 0.2.3 [27] with default options. Trimmed reads (Nanofilt v. 1.1.3 [2]) with a quality score Q > 9 and longer than 500 bp (1.34x10^6 reads; 17 Gb) were combined and further self-corrected using canu v.1.6 [29] with default options. Genome assembly was then performed based on the corrected long reads using SMARTdenovo v.1.0.0 run with default settings. A last polishing step with Pilon v1.22 [30] was carried out using paired-end (PE) Illumina sequencing reads (ca. 100X coverage) available for two pools of Red-nSpots individuals (see below). The resulting assembly, denoted *HaxR*, consisted of 1,071 contigs spanning 429 Mb (N50 = 1,434 Mb) encompassing 97.2% of the BUSCO (Benchmarking Universal Single-Copy Orthologs [31]) highly conserved arthropod gene set. See Table S1 for additional *HaxR* statistics.

Identification of *HaxR* autosomal contigs using a female to male read mapping coverage ratio

Barcoded DNA PE libraries with an insert size of ca. 550 bp were prepared using the Illumina TruSeq Nano DNA Library Preparation Kit following the manufacturer's protocols using six DNA samples extracted from three Red-nSpots males and three Red-nSpots females originating from a wild *H. axyridis* population from Mississippi (USA). Libraries were then validated on a DNA1000 chip on a Bioanalyzer (Agilent) to determine size and quantified by qPCR using the Kapa library quantification kit to determine concentration. The cluster generation process was performed on cBot (Illumina) using the Paired-End Clustering kit (Illumina). Each individual library was further paired-end sequenced on a HiSeq 2500 or 2000 (Illumina) using the Sequence by Synthesis technique (providing 2x125 or 2x100 bp reads, respectively) with base calling achieved by the RTA software (Illumina). After removal of sequencing adapters, reads were mapped onto the *HaxR* assembly using default options of the mem program from the bwa 0.7.12 software package [33]. Read alignments with a mapping quality Phred-score < 20 and PCR duplicates were further removed using the *view* (option *-q 20*) and rmdup programs from the samtools 1.3.1 software [34], respectively. Read coverage at each contig position for each individual sequences was then computed jointly using the default options of the samtools 1.3.1 depth program. To limit redundancy, only one count every 100 successive positions was retained for further analysis and highly covered positions (> 99.9th percentile of individual coverage) were discarded. The estimated individual median overall coverage ranged from 6 to 21 (see Data S1 for details).

To identify autosomal contigs, we used the ratio ρ of the relative (average) read coverage of contigs between all females and all males (weighted by the corresponding overall genome coverage) expected to equal 1 for autosomal contigs and 2 for X-linked contigs [48]. Contigs smaller than 100 kb were discarded from further analyses because they displayed a high coverage dispersion of their coverages (Data S2) together with 12 of the remaining contigs with extreme values of ρ ($\rho < 0.5$ or $\rho > 2.5$). We further fitted the ρ distribution of the 492 remaining contigs (398 Mb in total) as a Gaussian mixture model with two classes of unknown means and the same unknown variance. The latter parameters were estimated using an Expectation-Maximization algorithm as implemented in the mixtools R package [32]. As expected the estimated mean of the two classes $\mu_1 = 0.96$ and $\mu_2 = 1.90$ were only slightly lower (see Data S2 for further details) to that expected for autosomal and X-linked sequences allowing to classify the different contigs. We could hence therefore assign classify with a high confidence (p value < 0.01) 457 contigs as autosomal (377.5 Mb in total) and 18 contigs as X-linked (16.85 Mb in total).

Genome-wide scan for association with the proportion of Red-nSpots individuals using Pool-Seq data on 14 population samples

Barcoded DNA PE libraries with insert size of ca. 450 bp were prepared using either the Illumina Truseq DNA sample prep kit (n = 2) or the Nextera DNA Library Preparation kit (n = 12) following manufacturer protocols using 14 DNA pools (each pool including the head - or the leg for some pools - from n = 40 to n = 100 individuals) collected in eight populations representative of the world-wide genetic diversity [49] and the four main color pattern forms of the species (Table S2). Illumina sequencing, processing and mapping of reads to the *HaxR* assembly was performed as described above for individual data (see Data S1 for further details). The 14 Pool-Seq BAM files were processed using the mpileup program from the samtools v1.3.1 software [34] with default options and *-d 5000* and *-q 20*. Variant calling was then performed on the resulting mpileup file using VarScan mpileup2cns v2.3.4 [35] (options–min-coverage 50–min-avg-qual 20–min-var-freq 0.001–variants–output-v*cf.* 1). The resulting VCF file was processed with the vcf2pooldata function from the R package poolfstats v0.1 [36] retaining only bi-allelic SNPs covered by >4 reads, < 99.9th overall coverage percentile in each pool and with an overall MAF > 0.01 (computed from read counts). In total, 18,425,210 SNPs mapping to the 457 autosomal contigs were used for genome-wide association analysis with a median coverage ranging from 18 to 45X per pool (Data S1).

Genome-wide scans for association with the proportion of individuals of a given color pattern form in each pool were performed using the program BayPass 2.1 [15]. Capitalizing on the large number of available SNPs, we sub-sampled by taking one SNP every 200 SNPs along the genome, dividing the full dataset into 200 sub-datasets (each one including ca. 92,500 SNPs). These sub-datasets were further analyzed in parallel under the BayPass core model using default options for the Markov Chain Monte Carlo (MCMC) algorithm (except -npilot 15 -pilotlength 500 -burnin 2500). Three independent runs (using the option -seed) were performed for each dataset. The estimated model hyper-parameters were highly consistent across both runs and datasets. Support for association of each SNP with the corresponding prevalence covariate was then evaluated using the median Bayes Factor (BF) computed over the three independent runs. BFs were reported in deciban units (db) with 20 db corresponding to 100:1 odds, 30 db to 1000:1 odds, and so on [50].

De novo sequencing of the Black-4Spots color pattern allele

Starting from Black-4Spots individuals from the low diversity BIOTOP biocontrol population (France), we carried out four generations of full-sib crossings to produce a Black-4Spots inbred line, hereafter referred to as B4sIL. This aimed to improving further assembly steps by reducing the overall genetic variability. Four DNA PE libraries with insert sizes of ca. 250 bp (n = 2), 400 bp (n = 1) and 600 bp (n = 1) and two DNA Mate Pairs (MP) libraries with insert sizes of ca. 3 kb and 8 kb were constructed from B4sIL DNA (3-4 individuals per library) using standard Illumina kits; and two Long Jumping Distance libraries (Eurofins MWG Operon) with insert sizes of ca. 3 kb and 8 kb. All these libraries were sequenced on a HiSeq2500 sequencer with either 2x100 bp or 2x150 bp reads (Data S1). Raw reads were filtered for bacterial and human sequence contaminants using deconseq [37] and trimmed using Trimmomatic v0.22 [38]. Genome assembly was then performed using AllPath-LG [39] with default options except Haploidify = True to account for residual polymorphism in the sequenced individuals. This led to a first assembly consisting of 6,883 scaffolds (N50 = 921 kb) totaling 378 Mb (%N = 14.8). To further improve the assembly, we generated long reads using the Pacific Biosciences RS II platform. To that end seven SMRT bell libraries were prepared using size fractionated (shearing size of 25 kb and size selection cut-off of 10 kb) high molecular weight DNA prepared from B4sIL individuals and loaded into SMRT Cell 8Pac v3 for sequencing on Pacific Biosciences RS II system with P6C4 chemistry by Treecode (Malaysia). The seven resulting sequence movie files were processed and analyzed using Pacific Biosciences SMRT Analysis Server v2.3.0 with default settings. After the filtering step, a total of 422,222 reads (N50 = 21,521 kb) were used to carry out gap filling and scaffolding using PBJelly v1.3.1 [40]. The final Black-4Spots assembly, referred to as *HaxB4*, consisted of 6,586 scaffolds (N50 = 978.4 kb) totaling 393 Mb (%N = 5.84). Both genome-wide association studies conducted as described above but with *HaxB4* as a reference and sequence alignment of the utg686 contig of the *HaxR* assembly using Mummer [41] (and *BLAST*+ [42]) tools allowed unambiguous identification of a 5.96 Mb *HaxB4* scaffold that included the color pattern locus and adjacent genomic regions.

Because the *HaxB4* assembly remained less contiguous and accurate than the *HaxR* assembly (see Table S1 for a comparison), we performed a finishing step relying on a newly developed BAC (Bacterial Artificial Chromosomes) physical map covering the Black-4Spots allele of the color pattern locus. To that end, a BAC library of 13,824 BACs (141+/−41 kb insert size after sizing of 48 BACs) with ca. five genome equivalent coverage was constructed in the vector pIndigoBAC-5 using high molecular weight DNA extracted from 300 B4sIL larvae (of developmental stage L1), as described in [51]. The BAC library, deposited at the CNRGV (INRA, Toulouse, France; https://cnrgv.toulouse.inra.fr/Library/Asian-ladybird), was organized in 2-dimension pools for PCR screening. A total of 98 PCR primers pairs designed against the *HaxB4* assembly or newly generated BAC-end sequences were screened on the library. This allowed defining a Minimum Tiling Path of 14 BACs covering 1.9 Mb for local correction of scaffold mis-assembly. Shotgun sequencing was carried out using the Pacific Biosciences RS II system and Illumina MiSeq sequencer for 3 and 10 BACs, respectively [51]. Finally, we manually edited the *HaxB4* assembly using the BAC sequences to derive a high quality 2.87 Mb sequence (% N = 0.96) including the candidate region of the Black-4Spots allele of the color pattern locus as well as adjacent regions. Alignment of the latter sequence to the *HaxR* assembly (i.e., Red-nSpot allele) using nucmer from the package Mummer [41] allowed the scaffolding of the five neighboring *HaxR* contigs represented in Figure 1C.

Larval RNAi

We synthesized double-stranded RNAs (dsRNA) with T7 polymerase as described previously [16]. DNA fragments for the transcription were amplified by PCR using primers containing T7 polymerase promoter sequences at their 5′ ends (see Data S1for primer sequences). We used cDNA from Black-4Spots or Red-nSpots forms. Sense and antisense transcripts were simultaneously synthesized using RiboMax express RNAi System (Promega), annealed, treated with RQ1 DNase (Promega), and precipitated with ethanol. The quality of dsRNA was examined by agarose gel electrophoresis, and the concentration was roughly measured by spectrophotometer ND-1000 (NanoDrop Technologies), and 2 μg/μL in nuclease free water were used for injection. Larvae were anesthetized just before pupation on a CO_2 pad, and 400-600 nL of dsRNA was injected into hemolymph using Nanoject (Drummond Scientific).

cDNA sequencing

Fragments of *pannier* for Black-4Spots and Red-nSpots forms were amplified separately by PCR from cDNA of elytra of Black-4Spots or in Red-nSpots forms. Total RNA was extracted using TRI reagent (ThermoFisher), followed by DNase I treatment. cDNA was generated using First Strand cDNA Synthesis Kit (New England BioLabs). Resulting PCR fragments were approximately 1.8 kbp in length for both melanic forms when using the primer set (Ha_pnr-F1; CGGTACGAGATAAGCGAATAAGG, Ha_pnr-R1; TTACCATTTACAAATATATTTACATGGTTGTTG). Each PCR product was inserted into cloning vector pGEM-T Easy (Promega) for Sanger sequencing.

Quantitative PCR (qPCR)

Total RNA was extracted from whole elytron of homozygous Black-4Spots (n = 7) or Red-nSpots (n = 6), or dissected elytron of homozygous Black-4Spots (n = 5) at late pupal stage (96 hr after pupation) with TRI reagent (Invitrogen). RNA samples were reverse transcribed using First Strand cDNA Synthesis Kit (New England BioLabs). We omitted DNase I treatment, because all pairs of forward and reverse primer for qPCR were designed in different exons of each gene, which are separated by long introns. Furthermore, for accurate comparison, we confirmed the absence of nucleotide substitutions in the primer sequences in the different color pattern forms. qPCR and data analysis were performed on StepOne Real-Time PCR System (Applied Biosystems) with Power SYBR Green Master Mix (ThermoFisher). The data was normalized using eukaryotic initiation factor 4A (eIF4A) and 5A (eIF5A), and statistical significance of expression differences was established using two-tailed-t test. All primer sets and each R^2 value of standard curves are listed in Data S1.

Immuno-histochemistry

An antibody against *H. axyridis* Pannier (Ha_Pannier) was raised by Genscript, using as antigen the first 384 amino acids of the protein. To test that this antibody recognizes Ha_Pannier we ectopically expressed Ha_Pannier using the Gal4/UAS system in *Drosophila melanogaster*. Specifically, we stained *engrailed*-Gal4, UAS-GFP; UAS-Ha_Pannier larval imaginal disks with anti-Ha_Pannier and anti-GFP using standard procedures. We observed co-localization of the signal in the posterior compartement of the disk (data not shown), as expected, showing that the Ha_Pannier antibody recognizes Ha_Pannier *in vivo*.

Late pupal *H. axyridis* elytra are covered with a cuticle layer that is impenetrable for antibodies. Therefore, before staining, we split each elytron into two halves, separating the dorsal and ventral halves. For this we followed the protocol that has been developed for *Drosophila* wings [18] with some modifications. Elytra were dissected from pupae at late stage (around 96 hr after pupation) in PBS, and fixed in 4% paraformaldehyde (5-10min at room temperature). The edges of the elytra were trimmed off with a razor blade before transferring the elytra on a piece of adhesive tape (Tesa). Another piece of adhesive tape was positioned on top of the immobilized elytra, and then gently removed to separate the two faces of the elytra. The two pieces of tape with split elytra (one is dorsal, the other is ventral side of the elytra) were fixed 4% paraformaldehyde again (1-5min at room temperature) and stained (overnight at 4°C) with anti-Ha_Pannier antibody at 1:70 dilution in 1% bovine serum albumin (BSA), followed by visualization with Alexa-dye-conjugated secondary antibodies (ThermoFisher) at 1:100 dilution in 1% BSA (1 hr at room temperature). Cell nuclei were stained with DAPI. The pieces of tapes with stained elytra were mounted on microscope slides with VECTASHIELD (VECTOR).

Imaging

Anti-Pannier stainings were imaged under LSM510 confocal microscope (Zeiss) with identical settings (e.g., objective lens, pinhole size, laser power, number of stacks) for all samples. All raw confocal images were processed identically in ImageJ 1.51 [47], and then enhanced separately with (Adobe Photoshop). Mean intensities of anti-Pannier signal were measured in rectangular sections using the Plot Profile command of ImageJ 1.51.

Adult *H. axyridis* (2 days post-eclosion) were imaged on a Leica Z6Apo macroscope equipped with a ProgRes C5 ccd camera (Jenoptik). Several images were taken at different Z positions, and stacked together using HeliconFocus. Images were enhanced using Adobe Photoshop.

Genomic sequence divergence and gene structure at the color pattern locus

Genomic sequences of Red-nSpots (*utg676*) and Black-4Spots (*HaxB4*) contigs including *pannier* were visualized with dot plot using GenomeMatcher [46]. Conserved sequence blocks were further detected and visualized with Artemis Comparison Tool (ATC) [13] for Figure 1D. To identify reliable orthologous positions between the two contigs we first extracted long homologous blocks using blast2seq under high stringency parameters (blastn, e-value < 0.01, alignment length $\geq$ 2 kbp), and plotted them on the dot plot. The linear approximation was y = x+357764 (R^2 = 0.973, y axis; *utg676*, x axis; *HaxB4*). There was blank region (where there are no plots) in the middle of this linear approximation, which corresponds to a highly diverged region (202 kb and 234 kb in *utg676* and *HaxB4*, respectively). We subsequently checked shorter homology blocks ($\geq$1000 bp) around the breakpoints toward the center of the blank region to determine the borders more precisely. We considered continuous two or more conserved blocks ($\geq$1000 bp) within 20 × 20 kb sliding window along the line approximation. Thus, we defined two breakpoints for the boundary between continuous conservation and divergent regions, the latter one spanning 173,272 bp on *utg676* (554,399 - 727,671), in line with the ca. 170 kb region delimited by our genome-wide association study, and 209,085 bp on *HaxB4* (913,592 - 1,112,677).

To identify gene structures around the divergent genomic region (173 kb on *utg676*), we mapped RNA-seq reads PRJEB13023 [26] (100 bp paired-end, adult and larva *Harmonia* transcripts) deposited in Sequence Read Archive (SRA) to repeat-masked genome contigs using Tophat 2.1.0 [43], followed by assembling transcripts by Cufflinks v2.2.1 t [44] using default parameters on the NIG supercomputer at ROIS National Institute of Genetics. Resulting genes were named after sequence homology to protein database of *Drosophila melanogaster* (dmel-all-translation-r6.07) and *Tribolium_castaneum* (GCF_000002335.3_Tcas5.2). For the Black-4Spots sequence, since the number of mapped reads on exon-1 of *pannier* isoform-1 was low, we validated this gene structure using additional RNA-seq reads (H.V., unpublished data). The mapped reads were further confirmed by eye on Integrative Genomics Viewer (IGV) [45] to determine the gene structures.

Assessing large-scale sequence divergence among the four color form alleles of *Harmonia axyridis* based on read mapping coverage of Pool-Seq data on both the *HaxR* and *HaxB4* assemblies

Comparing *de novo* sequences surrounding *pannier* for the Red-nSpots and Black-4Spots alleles highlighted large-scale divergence in the upstream region covering ca. 170 kb in the *HaxR* and *HaxB4* assemblies (Figure 1D, Figure S3). This explained the clustering of SNPs harboring strong weights of evidence in favor of association with the proportion of Red-nSpot or Black-4Spots individuals in each sequenced pool (Figure S1). Interestingly, a similar clustering of strongly associated SNPs in the same genomic region was also observed in the genome scan for association with the Black-2Spots or Black-nSpots forms (Figure S1) suggesting extended sequence divergence in the upstream region of *pannier* for the Black-2Spots and Black-nSpots alleles too. To further assess the level of sequence divergence between the alleles of the four main color pattern forms, and especially for the Black-2Spots and Black-nSpots alleles that were not *de novo* sequenced, we examined read mapping coverage of the Pool-Seq data representative of these forms. The rationale of this approach was that extended divergence in the sequences of an allele represented at high frequency in a given pool, relatively to the reference assembly on which reads are mapped, is expected to translate into a local decrease in read coverage. We thus considered sequence data available for: (i) the four pools (CH2-R, BRG-R, WAS-R and BIO-R) including Red-nSpots individuals only (i.e., with a Red-nSpots allele frequency equal to 1); (ii) the three pools (CH2-B4, BIO-B4 and BRG-B4) consisting of Black-4Spots individuals only (i.e., with an expected Black-4Spots allele frequency $\geq$ 0.5 due to the possible presence of Red-nSpots alleles with an hidden expression in heterozygous Black-4Spots/Red-nSpots individuals); (iii) the CH2-B2 pool consisting of Black-2Spots individuals only (i.e., with an expected Black-2Spots allele frequency $\geq$ 0.5 due to the possible presence of Red-nSpots or Black-4Spots alleles with an hidden expression in heterozygous Black-2Spots/Red-nSpots or Black-2Spots/Black-4Spots individuals); and (iv) the NOV-Bn pool consisting of Black-nSpots individuals only (i.e., with an expected Black-nSpots allele frequency close to one due to the fixation of the Black-nSpots form in the corresponding wild population) (Figure S4). The five other remaining DNA pools that were used for the genome-wide association study were not considered here because they either consisted of mix of individuals from two melanic color pattern forms (e.g., CH2-B2 pool) or they displayed a genome-wide median coverage $\leq$ 25 (Data S1). Based on the mpileup file combining the mapping results of the pool-seq data onto the Red-nSpots assembly *HaxR*, we computed read coverages (using default options of the samtools 1.3.1 depth program) at each position of the *utg676* contig covering the Red-nSpots allele for the above nine DNA pools of interest. After discarding positions covered by <10 reads over all the pools or with a within pool coverage >95th percentile in at least one pool, the *utg676* contig was divided into consecutive windows of 10,000 positions with an overlap of 5,000 positions. Let $c_{i,p}$ represent the window coverage (computed as the average coverage over the 10,000 window positions) for window *i* in pool *p*; m_p the mean genome-wide coverage (computed over all *HaxR* autosomal contigs excluding *utg676* and over-covered position) in pool *p*; and $s_{i,p} = c_{i,p}/m_p$ the standardized window coverage for window *i* in pool *p*. To identify regions with lowered read mapping efficiency due to high sequence divergence with respect to the Red-nSpots allele (*HaxR* assembly), we computed a relative (standardized) window coverage as $rc_{i,p} = s_{ip}/s_i^R$ (for window *i* in pool *p*) where s_i^R is the standardized coverage for window *i* computed after merging reads from the four Red-nSpots representative pools. This correction allowed accounting for local variation in window coverage shared across experiments. Taking the *HaxB4* assembly as a reference, we similarly computed the relative (standardized) window coverage as $rc'_{i,p} = s'_{ip}/s'^{B4}_i$ (for window *i* in pool *p*) over the Black-4Spots allele sequence of the *HaxB4* assembly, where s'^{B4}_i the standardized coverage for window *i* computed after merging reads from the three Black-4Spots representative pools.

DATA AND SOFTWARE AVAILABILITY

The sequences of the genome assemblies *HaxR* and *HaxB4* are available from http://bipaa.genouest.org/sp/harmonia_axyridis. The data and sequences of the BAC library (color form Black-4Spots) are available from https://cnrgv.toulouse.inra.fr/Library/Asian-ladybird. The accession numbers for the Pool-Seq data are SRA: PRJNA474099 (see also sheet "c" in Data S1 for the SRA accession numbers of each 14 pools) and SRA: PRJEB13023 for RNA-Seq data. All software and algorithms are available from the references listed in the Key Resources Table (section "Software and Algorithms").