**Current Biology, Volume *28***

## Supplemental Information

# The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird

**Mathieu Gautier, Junichi Yamaguchi, Julien Foucaud, Anne Loiseau, Aurélien Ausset, Benoit Facon, Bernhard Gschloessl, Jacques Lagnel, Etienne Loire, Hugues Parrinello, Dany Severac, Celine Lopez-Roques, Cecile Donnadieu, Maxime Manno, Helene Berges, Karim Gharbi, Lori Lawson-Handley, Lian-Sheng Zang, Heiko Vogel, Arnaud Estoup, and Benjamin Prud'homme**

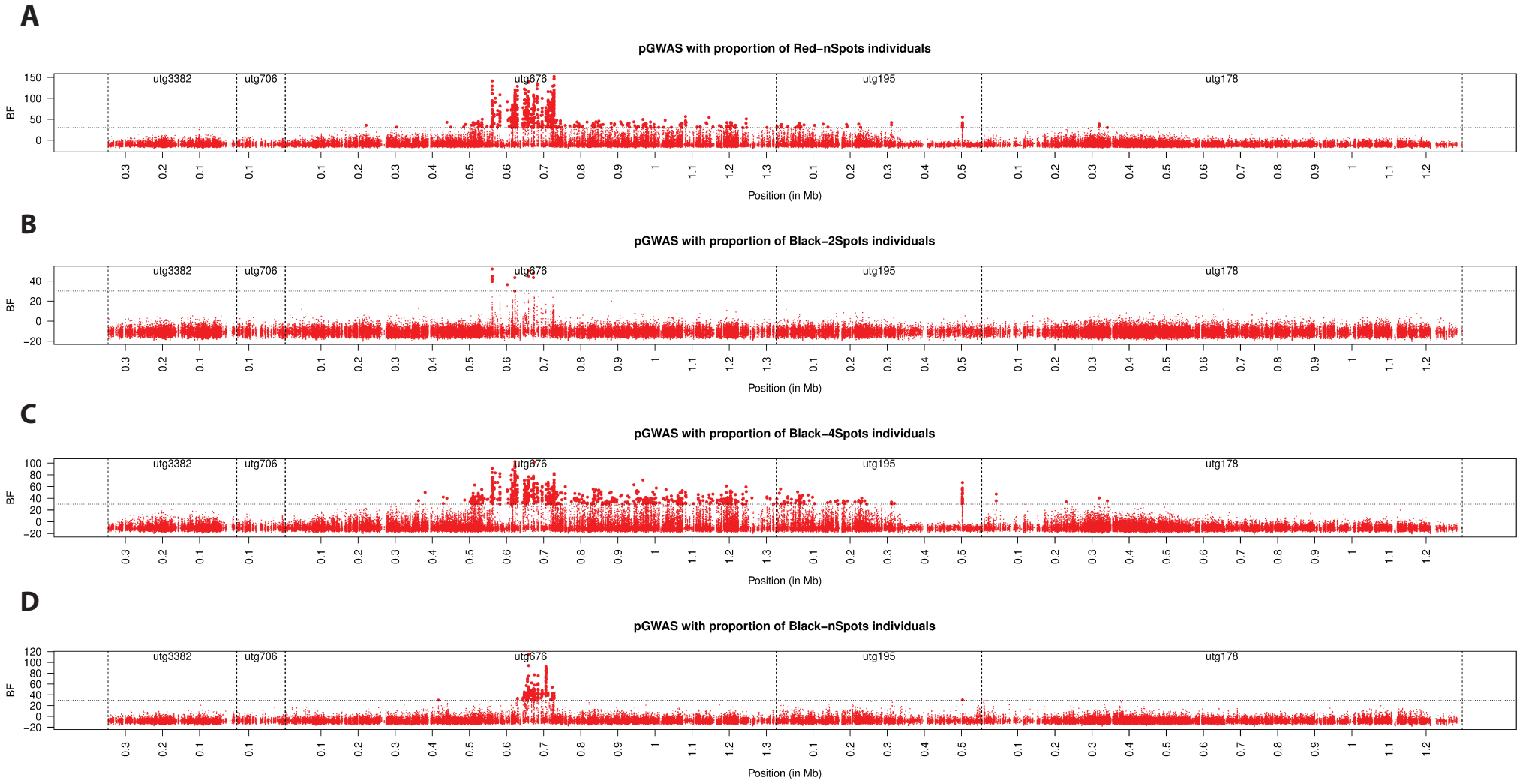


**Figure S1. Genome-wide association study focusing on the color pattern genomic region of the *HaxR* assembly. Related to Figure 1.**
(A) Association with the proportion of Red-nSpots individuals. (B) Association with the proportion of Black-2Spots individuals.
(C) Association with the proportion of Black-4Spots individuals. (D) Association with the proportion of Black-nSpots individuals.

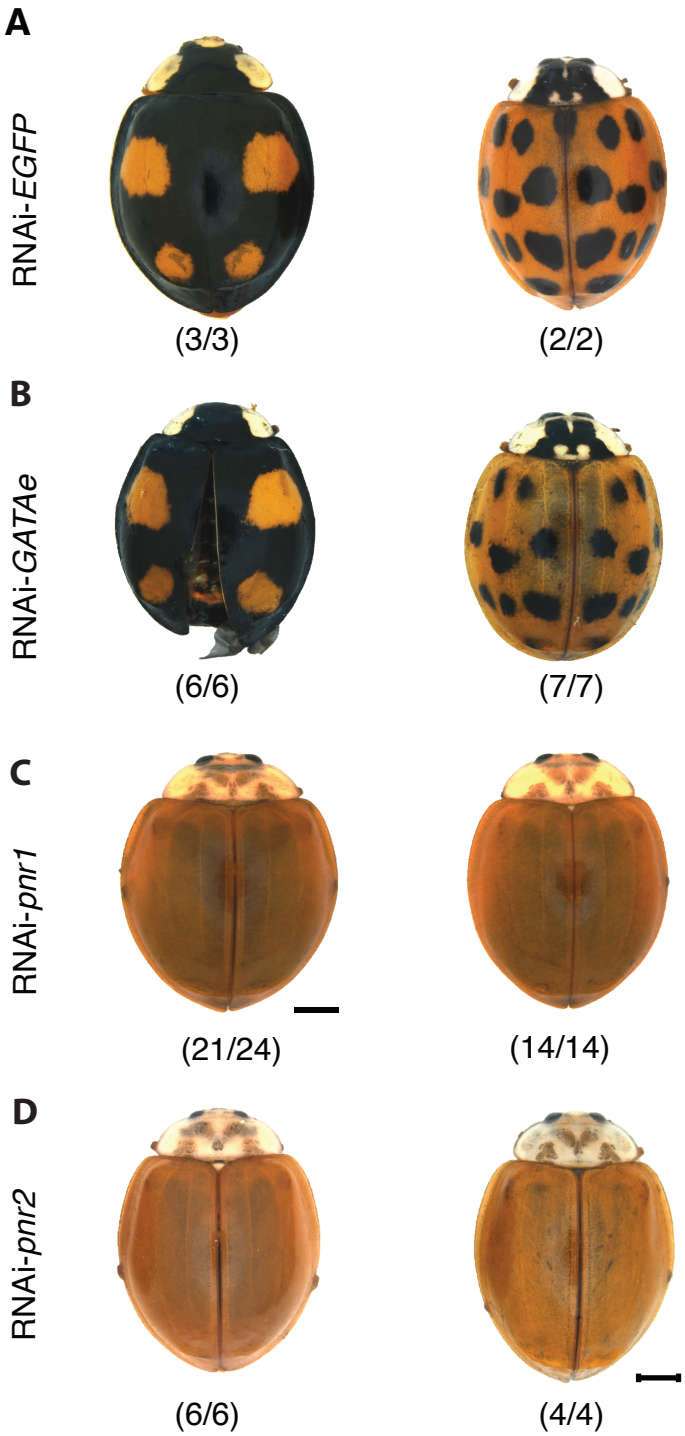


**Figure S2. *pannier*, but not *GATAe*, is necessary for adult pigmentation patterns.**
**Related to Figure 2; Figure S3.**
(A-D) RNAi phenotypes after larval injection of dsRNA in Black-4Spots (left column) or Red-nSpots (right column). (A) dsRNA targeting *eGFP* (negative control). (B) dsRNA targeting *GATAe*. (C) dsRNA targeting *pnr_1* region. (D) dsRNA targeting *pnr_2* region. *pnr_1* and *pnr_2* are non-overlapping regions of the *pannier* cDNA. Numbers in parentheses indicate the proportion of eclosed adults showing the pigmentation pattern of the representative individual. Scale bar; 1mm.

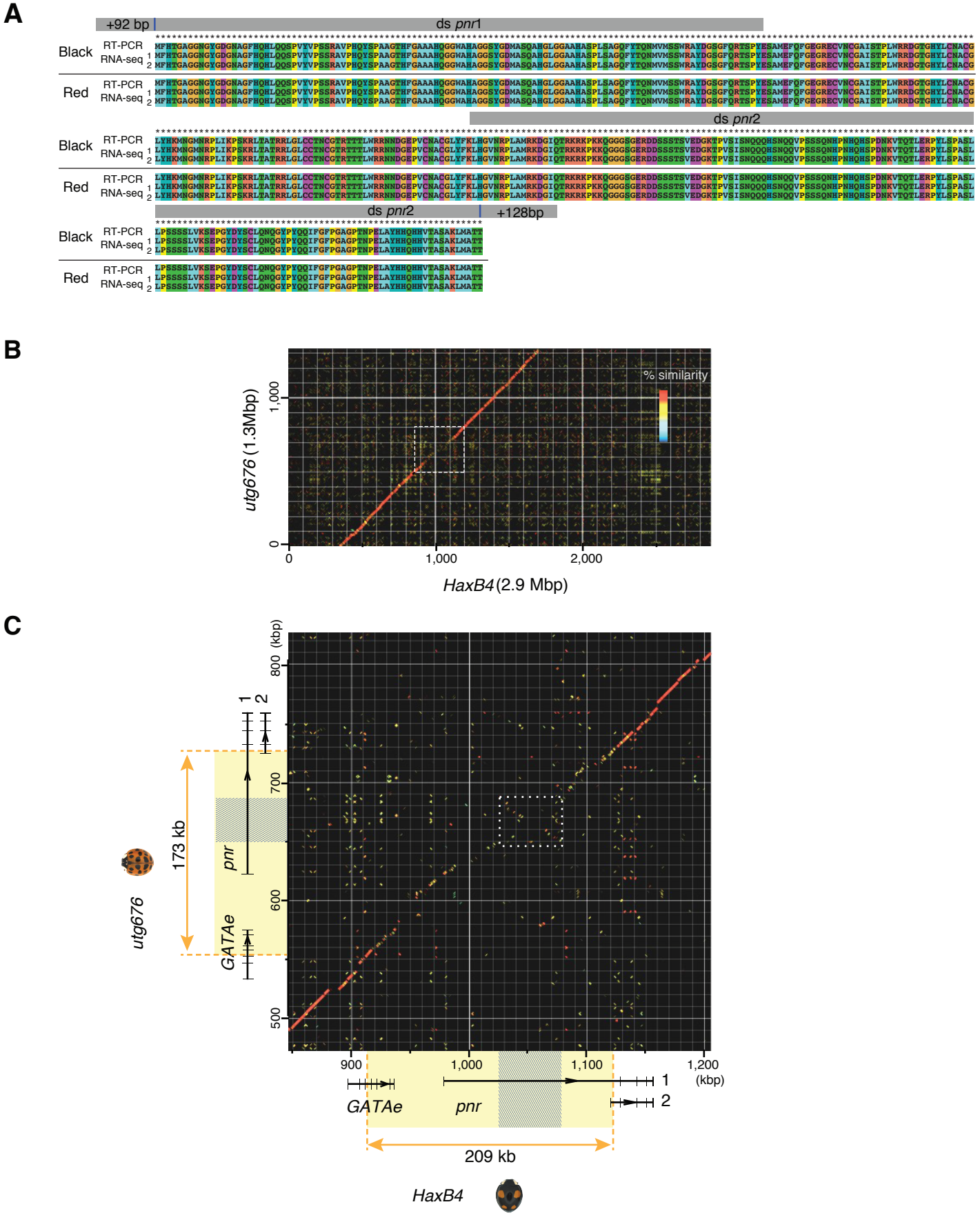
A
+92 bp
ds pnr1
Black
Red
RT-PCR
RNA-seq
ds pnr2
+128bp
B
% similarity
utg676 (1.3Mbp)
HaxB4 (2.9 Mbp)
C
(kbp)
173 kb
pnr
GATAe
utg676
209 kb
HaxB4

**Figure S3. Comparison of Pannier sequences and *utg676* and *HaxB4* sequences. Related to Figure 1; Figure 2; Figure S2.**

(A) Alignment of Pannier sequences, resulting from conceptual translation of cDNA obtained by RT-PCR or by mapping two independent RNA-seq reads from Black-4Spots or Red-nSpots individuals. Asterisks indicate conserved amino acids. The positions of the two fragments amplified for dsRNAs (*ds_pnr1* and *ds_pnr2*) are shown in grey. (B) Dot-plot between genomic scaffolds from the Black-4Spots (assembly *HaxB4*) and Red-nSpots (assembly *HaxR*, contig *utg676*) forms. The most divergent sequence (white dashed box in B) is shown at a higher magnification in C). (C) Traces of a sequence inversion are visible in the divergent region (white dashed box).

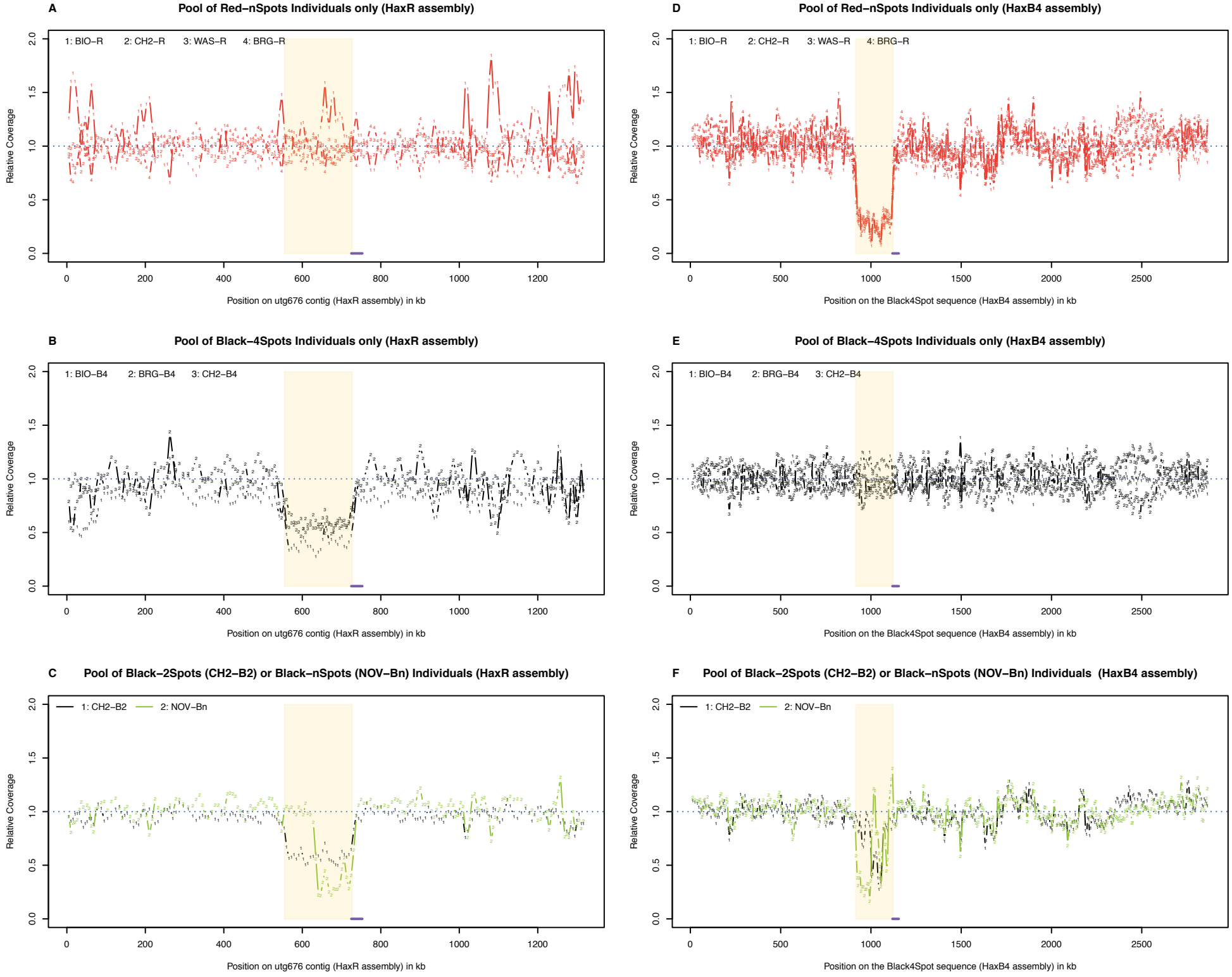
A
Pool of Red−nSpots Individuals only (HaxR assembly)
1: BIO−R
2: CH2−R
3: WAS−R
4: BRG−R
Relative Coverage
Position on utg676 contig (HaxR assembly) in kb
D
Pool of Red−nSpots Individuals only (HaxB4 assembly)
1: BIO−R
2: CH2−R
3: WAS−R
4: BRG−R
Relative Coverage
Position on the Black4Spot sequence (HaxB4 assembly) in kb
B
Pool of Black−4Spots Individuals only (HaxR assembly)
1: BIO−B4
2: BRG−B4
3: CH2−B4
Relative Coverage
Position on utg676 contig (HaxR assembly) in kb
E
Pool of Black−4Spots Individuals only (HaxB4 assembly)
1: BIO−B4
2: BRG−B4
3: CH2−B4
Relative Coverage
Position on the Black4Spot sequence (HaxB4 assembly) in kb
C
Pool of Black−2Spots (CH2−B2) or Black−nSpots (NOV−Bn) Individuals (HaxR assembly)
1: CH2−B2
2: NOV−Bn
Relative Coverage
Position on utg676 contig (HaxR assembly) in kb
F
Pool of Black−2Spots (CH2−B2) or Black−nSpots (NOV−Bn) Individuals (HaxB4 assembly)
1: CH2−B2
2: NOV−Bn
Relative Coverage
Position on the Black4Spot sequence (HaxB4 assembly) in kb

**Figure S4. Relative read mapping coverage for window of 10,000 positions. Related to STAR Methods**.

(A-C) relative read mapping coverage over the extended Red-nSpots (*HaxR* assembly), and (D-F) relative read mapping coverage over the Black-4Spots allele sequence (*HaxB4* assembly) for i) the four pools including Red-nSpots individuals only (A, D); ii) the three pools including Black-4Spots individuals only (B, E); and iii) the CH2-B2 and NOV-Bn pools including respectively Black-2Spots individuals only and Black-nSpots individuals only (C, F). The horizontal blue dotted line gives the value of 1 expected for the absence of departure in the relative coverage. The genomic region where the sequences of the Red-nSpots and Black-4Spots alleles diverge is represented as a shaded orange area. A purple line represents the position of the gene *pannier*. As expected, no obvious departure of the relative window coverage (from the expected value of 1) was observed over the entire Red-nSpots *utg676* contig for the four pools including Red-nSpots individuals only, although the BIO-R pools displayed several regions with high relative coverage values (A). In contrast, for the three pools including Black-4Spots individuals only, we found that the relative window coverage was almost halved in the ca. 170 kb region upstream *pannier*, with slight pattern differences likely explained by the pool frequency of heterozygous individuals carrying a Red-nSpots allele (B). When considering the extended Black-4Spots allele sequence as a reference (assembly *HaxB4*), the pattern was reverted: no obvious departure of the relative window coverage was observed for Black-4Spots pools (E) and the coverage was almost halved for Red-nSpots pools in the genomic region harbouring strong sequence divergence between the two color pattern form alleles (D). For the CH2-B2 pool including Black-2Spots individuals only, we observed a clear reduction in relative coverage over the entire genomic region where the sequences of the Red-nSpots and Black-4Spots alleles diverge on the HaxR assembly (C). Interestingly, the reduction in relative coverage concerned a more restricted region (located closer to *pannier*) on the *HaxB4* assembly (F). This suggests that the Black-2Spots allele might be more similar to the Black-4Spots allele, over at least a part of the upstream region of *pannier,* although it should be kept in mind that additional large scale variation (insertion or deletion) specific to the Black-2Spots allele are not identifiable with this approach. Finally, for the NOV-Bn pool including Black-nSpots individuals only, we observed a clear reduction in relative coverage over the second half of the region where the sequences of the Red-nSpots and Black-4Spots alleles diverge on the *HaxR* assembly (C). This pattern suggests a close similarity in sequence of the Black-nSpots allele with the Red-nSpots allele over the first half of its sequence. On the HaxB4 assembly we also observed a clear reduction of the NOV-Bn relative coverage over the first half of the region where the sequences of the Red-nSpots and Black-4Spots alleles diverge, whereas coverage reduction was less obvious in the second half of the region (F). This suggests some sequence similarity between the Black-nSpots and the Black-4Spots allele in the latter genomic region. It is worth noting, however, that for the Black-nSpots allele too, additional specific large-scale variations (insertion or deletion) are not identifiable with this approach.

| | Assembly *HaxR* | Assembly *HaxB4* |
|---|---|---|
| Data | MinION long reads (65X)<br>Illumina PE reads (100X) | Illumina PE reads (65X)<br>Illumina MP reads (24X) |
| Assembler | SMARTdenovo | ALLPATH-LG |
| Nb. of sequences | 1,071 contigs | 6,586 scaffolds |
| Total length (Mbp) | 429 | 393.1 |
| Average length (Kbp) | 400.9 | 59.7 |
| Max size (Kbp) | 7,499 | 5,635 |
| Total Ns (bp) | 22 | 22,814,986 |
| N50 (Kbp) | 1,434 | 978.4 |
| BUSCO (complete) | 97.2 % | 86.0 % |
| BUSCO (fragmented) | 1.3 % | 8.7 % |
| BUSCO (missing) | 1.5 % | 5.3 % |

**Table S1 Statistics characterizing the *HaxR* assembly obtained from Red-nSpots individuals and the *HaxB4* assembly obtained from Black-4Spots individuals. Related to STAR Methods.**
More details can be found in DataS1.

| Pooled-sequencing sample code | Population sampling site | | Sampling year | Colour form in the pool | No. of sequenced individuals |
|---|---|---|---|---|---|
| | Country | Region or city | | | |
| CH1-R | China | Jilin | 2013 | Red-nSpots only | 100 |
| CH1-B | | | | 30 Black-4Spots<br>28 **Black-2Spots** | 58 |
| CH2-R | China | Changchun | 2015 | Red-nSpots only | 100 |
| CH2-B4 | | | | Black-4Spots only | 67 |
| CH2-B2 | | | | **Black-2Spots only** | 73 |
| JP-R | Japan | Kyoto and other cities | 2009 | Red-nSpots only | 57 |
| JP-B4 | | | | Black-4Spots only | 58 |
| NOV-Bn | Russia | Novosibirsk | 2007 | Black-nSpots only | 44 |
| BRG-R | France | Bourgogne | 2013 | Red-nSpots only | 50 |
| BRG-B4 | | | | Black-4Spots only | 50 |
| ENA-R | USA | Georgia | 2007 | Red-nSpots only | 45 |
| WAS-R | USA | Washington | 2007 | Red-nSpots only | 40 |
| BIO-R | France | Biological control population (BIOTOP) | 2012 | Red-nSpots only | 100 |
| BIO-B4 | | | | Black-4Spots only | 100 |

**Table S2 Sequenced pools of individuals representative of the world-wide diversity and the four main color pattern forms of *H. axyridis*. Related to STAR Methods.**

The colour pattern forms Red-nSpot, Black-nSpots, Black-4Spots and **Black-2Spots** correspond to the forms f. *succinea*, f. *axyridis*, f. *conspicua*, and f. *spectabilis*, respectively. Because of the hierarchical dominance between color form alleles (i.e. Black-2Spots > Black-4Spots > Black-nSpots > Red-nspot), a population pool sample including individuals of a single colour pattern form is characterized by a high proportion of the allele associated to that form but also contains alleles associated to other more recessive forms if present in the population. Only population pool samples with Red-nSpots individuals contain 100% of Red-nSpots alleles.