MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 5829–5841 doi: 10.1111/mec.13432

Complex patterns of *cis*-regulatory polymorphisms in *ebony* underlie standing pigmentation variation in *Drosophila melanogaster*

RYUTARO MIYAGI,* NORIYOSHI AKIYAMA,* NAOKI OSADA†‡[1] and AYA TAKAHASHI*§

*Department of Biological Sciences, Tokyo Metropolitan University, Minamiosawa 1-1, Hachioji 192-0397, Japan, †Department of Population Genetics, National Institute of Genetics, 1111 Yata, Mishima 411-8540, Japan, ‡Department of Genetics, SOKENDAI (The Graduate University for Advanced Studies), 1111 Yata, Mishima 411-8540, Japan, §Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Minamiosawa 1-1, Hachioji 192-0397, Japan

Abstract

Pigmentation traits in adult *Drosophila melanogaster* were used in this study to investigate how phenotypic variations in continuous ecological traits can be maintained in a natural population. First, pigmentation variation in the adult female was measured at seven different body positions in 20 strains from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) originating from a natural population in North Carolina. Next, to assess the contributions of *cis*-regulatory polymorphisms of the genes involved in the melanin biosynthesis pathway, allele-specific expression levels of four genes were quantified by amplicon sequencing using a 454 GS Junior. Among those genes, *ebony* was significantly associated with pigmentation intensity of the thoracic segment. Detailed sequence analysis of the gene regulatory regions of this gene indicated that many different functional *cis*-regulatory alleles are segregating in the population and that variations outside the core enhancer element could potentially play important roles in the regulation of gene expression. In addition, a slight enrichment of distantly associated SNP pairs was observed in the ~10 kb *cis*-regulatory region of *ebony*, which suggested the presence of interacting elements scattered across the region. In contrast, sequence analysis in the core *cis*-regulatory region of *tan* indicated that SNPs within the region are significantly associated with allele-specific expression level of this gene. Collectively, the data suggest that the underlying genetic differences in the *cis*-regulatory regions that control intraspecific pigmentation variation can be more complex than those of interspecific pigmentation trait differences, where causal genetic changes are typically confined to modular enhancer elements.

Keywords: allele-specific expression level, amplicon sequencing, *cis*-regulatory variation, linkage disequilibrium, melanin biosynthesis

Received 28 March 2015; revision received 16 October 2015; accepted 21 October 2015

Introduction

It is a fundamental question to ask how phenotypic variation is maintained in a population. In particular, it is not well understood how variations in continuous ecological traits persist within a population. One approach to address this question is to conduct a detailed investigation of DNA sequence polymorphisms underlying the particular variation. In *Drosophila melanogaster*, many morphological traits such as bristle numbers (Shrimpton & Robertson 1988a,b; Long *et al*. 1995; Gurganus *et al*. 1999; Nuzhdin *et al*. 1999; Mackay & Lyman 2005), body size (Turner *et al*. 2011) and wing shape (Weber *et al*. 1999, 2001; Zimmerman *et al*. 2000; Mezey *et al*. 2005) exhibit continuous variation and these traits are caused by numerous QTLs. While these

Correspondence: Aya Takahashi, Fax: +81-42-677-2559; E-mail: ayat@tmu.ac.jp

[1]Present address: Graduate School of Information Science and Technology, Hokkaido University, Kita 14-jo, Nishi 9-chome, Kita-ku, Sapporo 060-0814, Japan

traits are important for future study, as they likely represent characteristic complex variation within a wild-type population, simple systems can be more experimentally tractable. In this sense, pigmentation traits are ideal for focused analyses, as they tend to show continuous variations even with small numbers of major responsive genes (Kopp *et al.* 2003; Takahashi *et al.* 2007; Bastide *et al.* 2013; Dembeck *et al.* 2015).

Pigmentation variation in *D. melanogaster*, which is exhibited as variable pigmentation intensity in the thoracic trident and variable darkness of abdominal segments, was first noticed by early geneticists (Morgan & Bridges 1919) and was subsequently reported in many population samples from different continents (Jacobs 1960; David *et al.* 1985; Munjal *et al.* 1997; Pool & Aquadro 2007; Takahashi *et al.* 2007; Parkash *et al.* 2008, 2009; Rebeiz *et al.* 2009; Takahashi & Takano-Shimizu 2011; Telonis-Scott *et al.* 2011; Bastide *et al.* 2013, 2014; Dembeck *et al.* 2015). Clinal distribution of pigmentation intensity according to latitude or altitude has been observed repeatedly (David *et al.* 1985; Munjal *et al.* 1997; Pool & Aquadro 2007; Parkash *et al.* 2008, 2009; Telonis-Scott *et al.* 2011) and has suggested that this trait is under some type of natural selection. Thus, a considerable amount of pigmentation variation in abdominal and thoracic segments appears to exist universally in natural populations of this species.

Genetic studies using strains from some variable populations indicated that the key causal gene for pigmentation intensity variation in thoracic and abdominal segments is *ebony*, which encodes an enzyme within the melanin biosynthesis pathway (Pool & Aquadro 2007; Takahashi *et al.* 2007; Rebeiz *et al.* 2009; Telonis-Scott *et al.* 2011). The extent of dark melanin on abdominal segments A5–7 is sexually dimorphic in this species and is regulated by a transcription factor gene, *bric-a-brac* (Kopp *et al.* 2000). This sexual dimorphism is distinct from the nonsexually dimorphic dark stripes on the abdominal segments A1–4 (Kopp *et al.* 2000), which are regulated by the transcription factor *optomotor blind* (Kopp & Duncan 1997). Previous studies showed that stripe width in A5–7 was associated with *bric-a-brac* and another gene *tan*, which codes for an enzyme in the melanin biosynthesis pathway (Kopp *et al.* 2003; True *et al.* 2005; Bastide *et al.* 2013; Dembeck *et al.* 2015). The causal variations in these genes are thought to be in the *cis*-regulatory control regions that regulate gene expression levels during the late pupal and the early post-eclosion periods (Pool & Aquadro 2007; Takahashi *et al.* 2007; Rebeiz *et al.* 2009; Bastide *et al.* 2013). Interestingly, the two enzymes, Ebony and Tan, play reciprocal roles in the chemical reaction between dopamine and beta-alanyl-dopamine in the melanin biosynthesis pathway (Wright 1987; Wittkopp *et al.* 2003). In addition to these genes with major effects, a genome-wide association study (GWAS) on abdominal pigmentation using the *Drosophila melanogaster* Genetic Reference Panel (DGRP), which originated from a natural population in North Carolina (Mackay *et al.* 2012), showed contributions of multiple genes with relatively minor effects involved in various developmental and regulatory pathways (Dembeck *et al.* 2015).

An extensive transgenic assay of flies from African populations conducted by Rebeiz *et al.* (2009) identified five causal *cis*-regulatory SNPs in *ebony* that were associated with gene expression levels and abdominal pigmentation intensity. However, these SNPs were not associated with thoracic pigmentation intensity in the Iriomote population from southern Japan (Takahashi & Takano-Shimizu 2011) or with extent of abdominal pigmentation in the populations from Europe (Bastide *et al.* 2013) and North America (Dembeck *et al.* 2015). Distinct haplotypes within the core *cis*-regulatory element for the light and dark strains were observed in the Iriomote population; this haplotype association was not observed in the African populations (Takahashi & Takano-Shimizu 2011). These complex pictures are in marked contrast to observed interspecies *cis*-regulatory changes that are typically simple changes within the modular enhancer elements (Gompel *et al.* 2005; Jeong *et al.* 2006, 2008; Prud'homme *et al.* 2006). Further investigation of various natural populations will broaden our understanding of the genetic mechanisms underlying this universal trait variation.

In this study, we utilized strains from the DGRP as in Dembeck *et al.* (2015), but took a different approach from their GWAS. First, pigmentation variations at seven different body positions and allele-specific expression levels of four genes in the melanin biosynthesis pathway were quantified during the early post-eclosion period when melanin synthesis occurs. Next, the contributions of *cis*-regulatory polymorphisms to the variations in pigmentation patterns were analysed. Finally, a detailed sequence analysis was performed on *ebony*, which was associated with the pigmentation traits, and *tan*. Our data demonstrate that the complex features of *cis*-regulatory polymorphisms underlie pigmentation variation in natural populations of *D. melanogaster*.

Materials and methods

Fly strains

The following inbred *Drosophila melanogaster* strains from the DGRP (Ayroles *et al.* 2009; Mackay *et al.* 2012; Huang *et al.* 2014) were used in this study: RAL-208, RAL-303, RAL-324, RAL-335, RAL-357, RAL-358,

RAL-360, RAL-365, RAL-380, RAL-399, RAL-517, RAL-555, RAL-705, RAL-707, RAL-732, RAL-774, RAL-786, RAL-799, RAL-820 and RAL-852. These strains were chosen from the set of 40 strains with available microarray expression data (Ayroles *et al.* 2009). Mel6 (G59), an inbred strain of *D. melanogaster* originally from Benin, West Africa (Takahashi & Takano-Shimizu 2005), was used as a reference strain. All flies were grown at 20 °C with a 12 h light–dark cycle on standard corn-meal fly medium.

Imaging and scoring pigmentation traits

Pigmentation intensities were measured in females at 3–6 days post eclosion. Note that the initial assessment using four sampled DGRP strains showed no detectable difference in pigmentation intensity measurements between 3-day-old and 6-day-old individuals, with only one exceptional case (Fig. S1, Supporting information). Flies were placed in 10% glycerol in ethanol overnight at 4 °C. After removal of the head, legs and wings, the head-less bodies (thorax and abdomen) were stored in 10% glycerol in PBS for 1–3 days at 4 °C. Images of dorsal bodies were captured from the flies embedded in 10% glycerol/PBS using a digital camera (DP73, OLYMPUS) connected to a stereoscopic microscope (SZX16, OLYMPUS). All the images were captured under identical parameters (exposure time, zoom width and illumination) with the reference greyscale (Brightness = 128; ColorChecker, X-rite). White balance was corrected using the white scale (Brightness = 255; ColorChecker, X-rite) with CELLSENS STANDARD 1.6 software (OLYMPUS). Images were captured using the RGB mode. Pigmentation was measured using IMAGEJ software (Schneider *et al.* 2012) in four manually selected areas of the body indicated in Fig. 1B. The mode greyscale brightness value from each area (0–255) was corrected using the reference grey background area (b in Fig. 1B). Thus, % darkness was calculated as follows:

$$\left(1-\frac{\text{brightness}}{\text{background brightness}}\times\frac{128}{255}\right)\times 100\,(\%).$$

The percent of dark-pigmented area was measured in females at 6–7 days post eclosion. Flies were placed in 10% glycerol in ethanol overnight at 4 °C. After the head and thorax were removed, the abdomen was incubated in 10% KOH for 2 h at 65 °C and was subsequently washed twice with 10% glycerol in PBS. The abdominal cuticle was flattened, cut and mounted in Hoyer's solution on a slide. The cuticle was cut at two positions indicated in Fig. 1C to flatten the whole segment and avoid folding or overlapping when mounted on a slide. Images were captured as described above. The dark-pigmented area of A4, A5 and A6 (Fig. 1C)

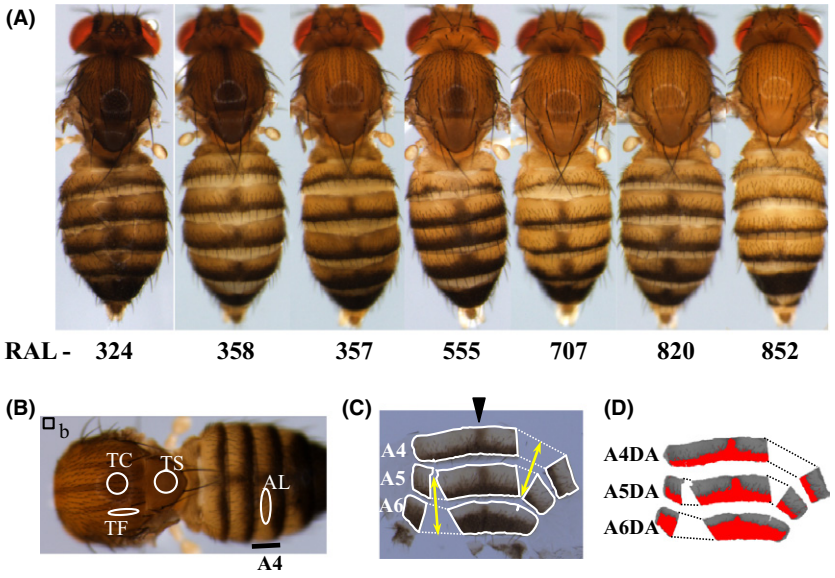

Fig. 1 Pigmentation variation in the DGRP. (A) Dorsal images of characteristic DGRP strains reared at 20 °C. (B) Pigmentation intensity measurement positions: % darkness of grey background (b); centre of thoracic trident, Trident Centre (TC); outside flanking area of thoracic trident, Trident Flank (TF); centre of thoracic scutellum, Thoracic Scutellum (TS); and lightly pigmented nonstripe area of the fourth abdominal (A4) segment, Abdominal Light (AL). (C) Flattened abdominal cuticle of A4, A5, and A6. Dorsal midline is indicated by an arrowhead. The cuticle was cut at two positions indicated by yellow arrows to flatten the whole segment. (D) Pigmented area measurement positions: % dark-pigmented area of A4 (A4DA), A5 (A5DA) and A6 (A6DA). Extracted dark-pigmented area is marked in red.

was defined as the dark-pigmented stripe region extracted by setting the threshold to the midpoint brightness value (midpoint between peaks of bimodal pixel distribution) within each segment that distinguishes stripe and nonstripe regions (Fig. 1D). The extraction of the dark-pigmented area and the quantification of % dark-pigmented area within each segment were performed using IMAGEJ software (Schneider *et al.* 2012).

RNA and DNA extraction

Males from each of the 20 DGRP strains were crossed to females of the Mel6 reference strain, and the F_1 progenies were subjected to cDNA Amplicon sequencing (Amp-seq; Fig. S2, Supporting information) as in Bickel *et al.* (2011). Expression differences between parental strains within each parental strain pair that consisted of Mel6 and each one of the 10 selected DGRP strains (RAL-335, RAL-358, RAL-365, RAL-555, RAL-705, RAL-774, RAL-786, RAL-799, RAL-820 and RAL-852) were quantified using cDNA Amp-seq. F_1 virgin females (n = 10–15) from the above crosses and virgin females from the selected parental strains (Mel6: n = 26; DGRP: n = 5–8) were collected within 30 min of eclosion during a 1–2 h period after the light was turned on. The collected flies were immediately placed on ice in 50% RNAlater (Ambion) in PBS with 0.1% Tween20 (MP Biomedicals) and their heads were removed. The remaining headless body samples were transferred to RNAlater and kept at −80 °C until further use. Total RNA was extracted using a TRIzol Plus RNA Purification Kit (Life Technologies) and was quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific). Two biological replicate samples were obtained for each cross or each strain. To compare expression differences between the parental strains, parental RNA mix was prepared by mixing 5 μg total RNA from the Mel6 strain and 5 μg total RNA from each DGRP strain. Genomic DNA was extracted from five females according to the method described by Boom *et al.* (1990) with minor modifications.

Plasmid construct for real-time quantitative PCR (qPCR)

All target regions for qPCR were amplified from RAL-820 and Mel6 cDNA. Products were ligated into the pGEM-T vector using a pGEM-T Easy kit (Promega), and transformed into DH5α *E. coli* competent cells (Promega). Plasmid DNA containing the target insert was extracted using a QIAGEN Midi kit (QIAGEN) and was then linearized using Not I. Digested RAL-820 and Mel6 constructs were purified using a standard phenol-chloroform method, quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific), and then used to obtain standard curves for qPCR.

Amp-seq

First strand cDNA was reverse-transcribed from 1 μg of total RNA using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa). Oligo dT Primer included in the kit was used for reverse transcription. PCR amplification of target genes (*ebony*, *tan*, *black* and *pale*) was performed using a 48.48 Access Array system (Fluidigm) in accordance with the manufacturer's instructions. Briefly, to equalize the amplicons from each gene, PCR was performed using the following conditions: 50 °C for 2 min, 70 °C for 20 min and 95 °C for 10 min, followed by 10 cycles of the first step (95 °C for 15 s, 60 °C for 30 s and 72 °C for 1 min), 2 cycles of the second step (95 °C for 15 s, 80 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min), 8 cycles of the first step, 2 cycles of the second step, 8 cycles of the first step and 5 cycles of the second step. PCR was performed using a FastStart High Fidelity PCR system dNTPack (Roche) and four primer pairs for each sample: forward and reverse primers comprising 3′ target-specific sequences (19–23 bp) and 5′ universal tag sequences (22 bp), as well as forward and reverse sample barcoded primers comprising 3′ universal tag sequences (22 bp), sample barcode sequences (10 bp) and 5′ adaptor sequences (25 bp) for 454 GS Junior library construction (Table S1, Supporting information). The amplified regions were 150–200 bp in length and contained SNPs known to differ between the DGRP and Mel6 strains. Target-specific primer pairs were designed within the same exon. The nucleotide positions of the target SNPs are described in Table S2 (Supporting information). Read counts from primer pairs targeting the same SNP were pooled for subsequent analyses. Amplification was conducted in 35 nL reactions from cDNA (synthesized from 1.05 ng of total RNA) or genomic DNA (0.35 ng). Amplicons were quantified using a Bioanalyzer 2100 or a TapeStation (Agilent Technologies). Amp-seq was performed using a 454 GS Junior (Roche) sequencer. DNA libraries were constructed using a GS Junior Titanium emPCR kit (Lib-A, Roche). Sample barcodes (12–29) were pooled in each run.

Real-time qPCR

First strand cDNA was reverse-transcribed from 500 ng of total RNA extracted from the parental strains above (Mel6 and 10 selected DGRP strains) using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa). The PCR reaction solution (in a 25 μL total volume) contained

12.5 μL of SYBR *Premix Ex Taq* II (Tli RNaseH Plus, TaKaRa), 6 μL of 50-fold dilution of the cDNA samples above and 10 pmol each of the forward and reverse primers (α*Tub84B*: 5′−CATAGCCGGCAGTTCGAACG−3′ and 5′−CGCTGAAGAAGGTGTTGAACGA−3′; *RP49*: 5′−TCGGATCGATATGCTAAGCTG−3′ and 5′−TCGATCCGTAACCGATGTTG−3′; *Act57B*: 5′−CGTGTCATCCTTGGTTCGAGA−3′ and 5′−ACCGCGAGCGATTAACAAGTG−3′). PCR reactions were performed using the Thermal Cycler Dice Real Time System II (TaKaRa) with the following conditions: 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Correction of PCR efficiency of each primer set was performed using a standard curve drawn from the 10-fold dilution series of the plasmid DNA described above. Each sample was measured in triplicate.

Quantification of allelic expression ratio (AER)

Nucleotide sequences with quality scores were obtained from 454 GS Junior sequencing. After low-quality sequence reads were eliminated, the GS Reference Mapper (Roche) was used to map remaining reads to the reference sequences of the *D. melanogaster* genome (Dmel 5.13 reference genome). Sequenced reads were assigned to DGRP or Mel6 allele-derived reads according to SNP types. The information of the allele-specific read counts was extracted by SAMTOOLS (Li *et al.* 2009). Read counts from primer pairs targeting the same SNP were pooled (Table S2, Supporting information).

For F_1 hybrids, the fold-difference between alleles (D_{F1}/M_{F1}), where D_{F1} and M_{F1} indicate the number of sequenced reads from DGRP- and Mel6-derived amplicons from the F_1 cDNA, respectively, was calculated. The fold-difference (D_{F1}/M_{F1}) was averaged across SNPs within each gene and across replicates. The fold-difference between alleles in F_1 genomic DNA (D_{G_F1}/M_{G_F1}) was used to assess the amplification bias, where D_{G_F1} and M_{G_F1} indicate the number of sequenced reads from DGRP- and Mel6-derived amplicons from the F_1 genomic DNA respectively.

For parental comparisons, fold-difference between alleles (D_{C_par}/M_{C_par}) was calculated, where D_{C_par} and M_{C_par} indicate the number of sequenced reads from DGRP- and Mel6-derived amplicons from parental cDNA mix respectively. Then, the fold difference between parental alleles (D_{par}/M_{par}) was scaled by the fold-difference in amount of parental cDNA in the parental cDNA mix (Table S3, Supporting information). Expression levels of *aTub84B*, *RP49* and *Act57B* quantified by qPCR from parental cDNA samples (geometric means) were used for scaling. The fold-difference (D_{par}/M_{par}) was averaged across SNPs within each gene and across replicates.

The allelic expression ratio of the DGRP-derived allele ($AER_{DGRP} = D_{F1}/(D_{F1} + M_{F1})$) was used for the comparisons with pigmentation trait scores and was calculated using the average D_{F1}/M_{F1} across SNPs within each gene ($AER_{DGRP} = 1/\{1 + (M_{F1}/D_{F1})\}$) for each replicate and the ratio was subsequently averaged across replicates.

Regulatory effect categories

Regulatory effects were assessed for each strain pair by comparing allele counts in the parental and the hybrid cDNA amplicons. Differential expression of alleles in the parents and F_1s were determined by the deviations from $D_{par}/M_{par} = 1$ and $D_{F1}/M_{F1} = 1$, respectively, and the *trans*-regulatory effect was determined by the deviation from $D_{par}/M_{par} = D_{F1}/M_{F1}$. According to the significance of these deviations, the regulatory status of each strain pair was classified into seven categories (*cis* only, *trans* only, *cis* + *trans*, *cis*-by-*trans*, compensatory, conserved and ambiguous), following the criteria used in previous studies (Landry *et al.* 2005; McManus *et al.* 2010; Suvorov *et al.* 2013).

Statistical tests were conducted by generating distributions of D_{par}/M_{par}, D_{F1}/M_{F1}, or $(D_{par}/M_{par})/(D_{F1}/M_{F1})$ using the expected value = 1 by computer simulations. These values were simulated 10^6 times using the observed read count for each SNP and were averaged across SNPs within each gene and across replicates, as was performed for the observed values. The two-tailed *P*-value was calculated from the generated distribution and FDR <0.05 was regarded as statistically significant. Simulation and *P*-value calculation were conducted using custom Perl scripts. FDR was calculated using R (R Core Team 2014).

DNA sequencing of ebony and tan cis-regulatory regions

Putative *cis*-regulatory control regions of *ebony* (*e_cis*, Fig. 2A) and the core *cis*-element of *tan* (*t*_core_*cis*, Fig. S6A, Supporting information) were amplified from genomic DNA samples using Ex Taq (TaKaRa). PCR products were purified and were directly sequenced using BigDye Terminator Cycle sequencing kit version 3 and an ABI 3130xl Genetic Analyzer (Applied Biosystems).

Testing association between polymorphisms and AER_{DGRP}

Spearman's rank correlation test was applied to each SNP and indel within the *e*_core_*cis* and the *t*_core_*cis* regions to test for an association with AER_{DGRP} of *ebony*

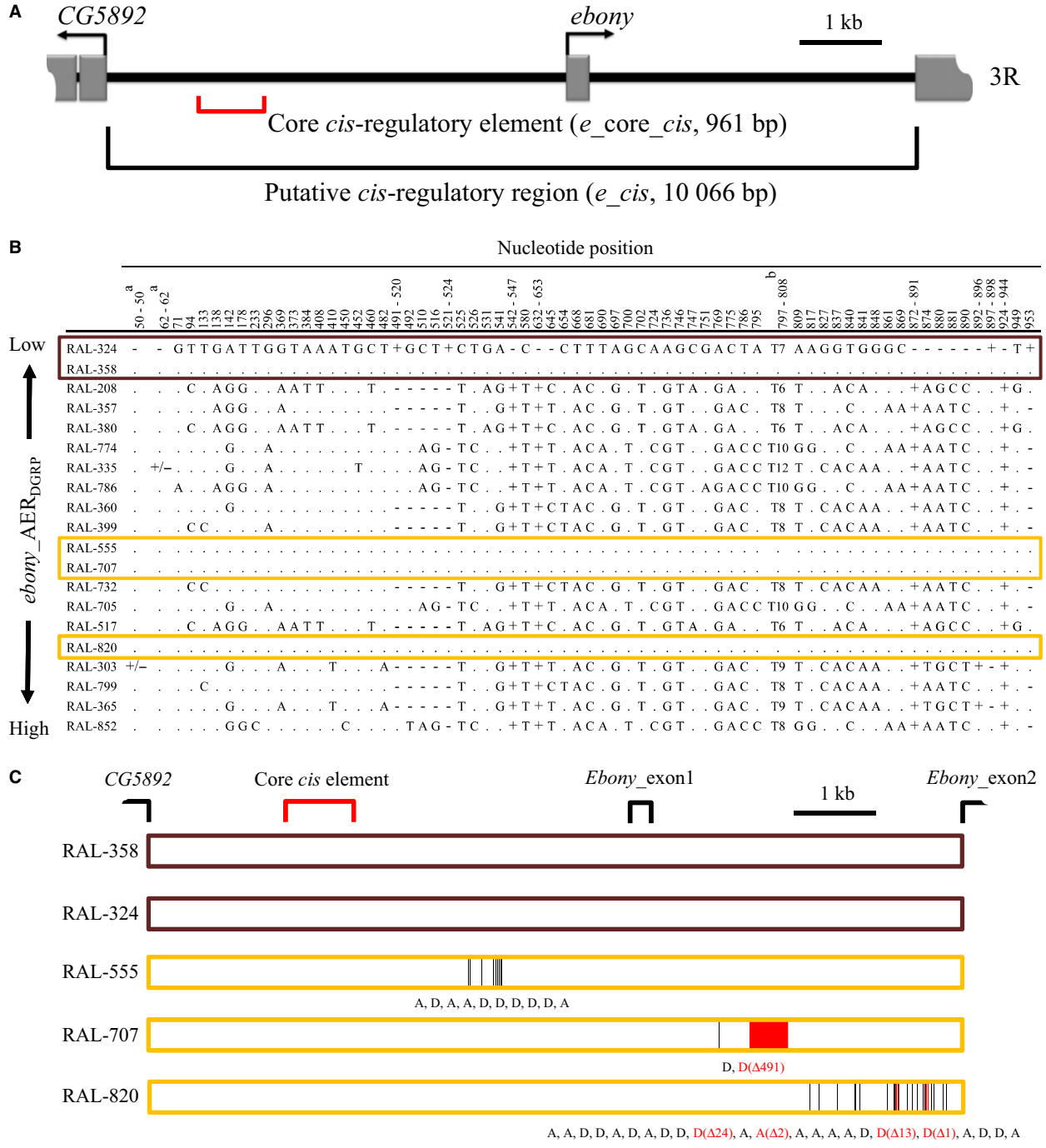

Fig. 2 Sequence polymorphisms in the putative *cis*-regulatory region of *ebony* in the DGRP strains used in this study. (A) Schematic representation of the locations of annotated genes near *ebony*. The putative *cis*-regulatory region controlling *ebony* expression (*e_cis*:17 058 756..17 068 350 in Dmel 5.13) is indicated by the black bracket. Grey boxes represent exons. The red bracket indicates the core *cis*-regulatory element, which regulates *ebony* expression in the thoracic and abdominal epidermis (Rebeiz *et al.* 2009; Takahashi & Takano-Shimizu 2011). (B) Nucleotide alignment of the core *cis*-regulatory element. The 20 DGRP strains are listed in ascending order of *ebony*_AER$_{DGRP}$. Note that the core sequences of the two strains with the lowest *ebony*_AER$_{DGRP}$ values (boxed in brown) were identical to the sequences of the three strains (boxed in yellow) with higher AER$_{DGRP}$ values. "." indicates nucleotide status identical to that of RAL-358, [a]indicates indel of unknown length and [b]indicates nucleotide status as indicated by the number of T repeats. (C) Polymorphic sites in the putative *cis*-regulatory region of the five DGRP strains indicated by brown and yellow boxes in (B). SNPs and indels are indicated with black and red vertical bars respectively. Ancestry (A: ancestral and D: derived) was inferred using the *D. sechellia* sequence (droSec1).

and *tan* respectively. Indels of unknown length and variation in poly-T repeats were excluded from the analysis. FDR was calculated using R (R Core Team 2014).

Linkage disequilibrium (LD) analysis

Nucleotide polymorphism data of 205 DGRP strains were obtained from the DGRP freeze 2 data set (Mackay *et al.* 2012; Huang *et al.* 2014; http://dgrp.gnets.ncsu.edu/). SNP sites were used for LD calculation and gap sites were excluded from the analysis. SNPs with minor allele frequencies <0.05 and sites with >10% missing data (including heterozygotes) were eliminated. R^2 values were calculated using HAPLOVIEW 4.2 (Barrett *et al.* 2005).

Results

Pigmentation variation within the DGRP population

Four pigmentation intensity measurements from thoracic and abdominal segments were obtained from 10 to 13 adult females from each of the 20 DGRP strains (with the exception of RAL-786: only four females were measured due to sample limitations). Dark-pigmented stripe area measurements were obtained from three segments (A5, A6 and A7) for five adult females from each of the 20 DGRP strains. Pigmentation variations within the DGRP population were apparent (Fig. 1A) and significant differences were detected among the DGRP strains for all seven pigmentation traits (Welch's ANOVA *F*-test, Fig. S3, Supporting information).

Quantification of allelic expression ratio of the DGRP-derived allele (AER_{DGRP})

Some of the enzymes involved in the melanin biosynthesis pathway have pleiotropic functions in photoreceptor terminals and in the central nervous system (Newby & Jackson 1991; Borycz *et al.* 2002; True *et al.* 2005; Gavin *et al.* 2007; Suh & Jackson 2007; Wagner *et al.* 2007). Therefore, to focus on the functions of these enzymes in the epidermis, expression was analysed in headless body samples.

Amp-seq was conducted to quantify allele-specific expression levels in F_1 heterozygotes. Compared to a regular RNA-seq method, which requires sufficiently deep sequencing to obtain accurate quantities for all the target genes, this method proved cost effective for the quantification of allele-specific expression levels in the limited number of target genes. AER_{DGRP} values of the four melanin biosynthesis genes (*ebony*, *tan*, *black* and *pale*) were obtained using the Amp-seq method. SNPs used for allele identification and the read counts obtained for each SNP are summarized in Table S2 (Supporting information). There were three cases (SNP 1583 of *tan* in RAL-820 and SNP 2757 of *pale* in RAL-786 and RAL-208) where parental alleles in F1 genomic DNA amplicons, D_{G_F1} and M_{G_F1}, showed more than two-fold differences. In the cases of *pale*, SNPs within the target-specific primer sites were found, which may have biased the amplification efficiency. Therefore, these samples were removed from the subsequent analyses. Significant differences in AER_{DGRP} were observed among DGRP strains for all four melanin biosynthesis genes (Welch's ANOVA *F*-test, Fig. S4, Supporting information). Two strains (RAL-324 and RAL-358) showed notably low AER_{DGRP} values for the *ebony* gene compared to the other strains (Fig. S4A, Supporting information).

AER_{DGRP} was used as an index for the *cis*-regulatory effect on gene expression. Here, allele-specific expression levels of DGRP alleles were compared to those of the reference allele (Mel6 allele) in F_1 heterozygotes. Regarding F_1s from a single cross, both of the parental alleles share an identical *trans* environment, and thus, comparative analysis using AER determines the effect of *cis*-regulatory factors on the expression differences between alleles, assuming that the *cis–trans* interaction is negligible. To estimate the extent of such interactions, parent-hybrid comparisons were conducted using 10 selected strains (Fig. S5, Supporting information). Error can be introduced when there are prominent *cis*-by-*trans* and compensatory interaction effects (C × T and Comp) between the reference Mel6 and the sampled DGRP alleles. In all genes, the proportions of strain pairs that showed significant *cis*-by-*trans* and compensatory interaction effects were small (between 1 and 3 strain pairs, Fig. S5, Supporting information). In addition, the relative magnitudes of their effects (deviation from 0 in the *y*-axis) were small in *ebony*, *black* and *pale*, indicating that the interaction between *cis* and *trans* factors are negligible in these genes. In *tan*, the effect size of compensatory interaction was relatively large (Fig. S5B, Supporting information), suggesting that the power for detecting the *cis*-regulatory effect using AER_{DGRP} of *tan* may be slightly reduced.

Among the crosses, the Mel6 and the DGRP alleles in F_1s do not share identical *trans*-environments due to variability in *trans*-effects. However, when the *cis–trans* interaction is negligible, *trans*-regulatory variation affects the absolute transcript quantity and not the relative expression levels of the two alleles. We therefore reasoned that comparison of AERs (as described in Materials and Methods) in F_1 flies obtained from crosses between the Mel6 reference strain and each of the 20 DGRP strains (Fig. S2, Supporting information)

Table 1 Pearson's correlation coefficient *r* between AER_{DGRP} and pigmentation scores

	Pigmentation traits						
AER_{DGRP}	TC	TF	TS	AL	A4DA	A5DA	A6DA
ebony	−0.68*	−0.84***	−0.76**	−0.44	−0.42	−0.49	−0.17
tan	−0.19	−0.10	0.07	−0.31	0.12	−0.03	−0.16
black	0.22	0.24	0.29	−0.17	0.19	0.21	0.20
pale	−0.35	−0.34	−0.10	−0.40	−0.22	−0.27	−0.32

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (after Bonferroni correction).

allowed us to extract the *cis*-regulatory variation at each locus.

Correlation between pigmentation traits and AER_{DGRP}

To investigate the relationship between adult pigmentation and *cis*-regulatory gene expression, correlation tests were performed between scores of pigmentation intensity and AER_{DGRP} values. AER_{DGRP} of *ebony* showed strong correlations with pigmentation intensities of thoracic segment, whereas, the other genes showed no significant relationship with the pigmentation scores (Table 1).

DNA sequence variation in the cis-regulatory regions of ebony and tan

Next, DNA sequences in the *cis*-regulatory region of *ebony* were analysed. Previous reporter assays by Rebeiz *et al.* (2009) showed that an ~10 kb region that included an upstream intergenic region and intron 1 of *ebony* (Fig. 2A) was sufficient to drive most of the endogenous expression pattern of this gene. Therefore, we obtained sequences for this putative ~10 kb *cis*-regulatory region (*e_cis*) by resequencing the region in our DGRP strains obtained from the stock centre to verify heterozygous sites and to correct minor mapping errors in the published genome sequences of the strains (Mackay *et al.* 2012; Huang *et al.* 2014).

Previous studies identified a *cis*-regulatory element (<1 kb core *cis*-element) that controlled *ebony* expression in the epidermis of young adults (Fig. 2A; Rebeiz *et al.* 2009; Takahashi & Takano-Shimizu 2011). Analysis of the sequence variation in this core *cis*-element (*e*_core_*cis*) among the 20 DGRP strains revealed that there were no SNPs or indels associated with the AER_{DGRP} variation in *ebony* (Spearman's rank correlation test, FDR >0.05; Fig. 2B). Moreover, the *e*_core_*cis* sequences of the two strains with the lowest AER_{DGRP} of *ebony* (RAL-324 and RAL-358) were identical to one another and to the core sequences of three other strains with higher AER_{DGRP} values (RAL-555, RAL-707 and RAL-820). Interestingly, RAL-324 and RAL-358 *e_cis* sequences were identical across their entire ~10 kb length. The other three strains (RAL-555, RAL-707 and RAL-820) had nearly identical haplotypes throughout *e_cis*, with each having distinct single nucleotide or indel differences compared to the RAL-324/358 sequence (Fig. 2C). The minimal sequence variation in the *e_cis* region of these five strains provided a unique opportunity to investigate the *cis*-regulatory structure of this gene. Assuming that these nucleotide differences in *e_cis* were responsible for the differences in *ebony* expression between the two dark strains (RAL-324 and RAL-358) and the three light strains (RAL-555, RAL-707 and RAL-820), the sequence variation pattern indicated that different regions outside the *e*_core_*cis* region could potentially affect expression of this gene. This further suggests that there are likely to be multiple functional *cis*-regulatory alleles that have similar effects on expression and are segregating within a population. A SNP (A/G) located 220 bp upstream of the *ebony* transcription start site that was previously associated with abdominal pigmentation in Dembeck *et al.* (2015) exhibited an identical genotype (A) in all of the strains used in this study except RAL-517, which had a 17 bp deletion surrounding the SNP site.

Although, contribution of *tan* to the pigmentation scores was not detected in this study, previous studies have found several linked SNPs associated with the extent of abdominal pigmentation (Bastide *et al.* 2013; Dembeck *et al.* 2015). The putative core *cis*-regulatory region of this gene controlling the abdominal stripe pattern in segments A5–7 (*t*_core_*cis*) resides in the intergenic region between *CG15370* and *Gr8a* (Fig. S6A, Supporting information; Jeong *et al.* 2008). DNA sequences of this ~450 bp region were obtained by resequencing from the 20 DGRP strains (but RAL-820 was excluded from the analysis because of the possible amplification bias), as with the *ebony cis*-regulatory region. Seven SNPs including four that were previously identified were associated with the AER_{DGRP} of *tan* (Spearman's rank correlation test, FDR <0.05; Fig. S6B, Supporting information). Four SNPs associated with the extent of dark pigmentation on the A7 abdominal seg-

ment in European populations (Bastide *et al.* 2013) were present in the DGRP. Three of them were shown to have association with the pigmentation of A5 and A6 segments in the DGRP (Dembeck *et al.* 2015). Those SNPs (dark and light allele types) as well as another SNP near *t_core_cis* detected by Dembeck *et al.* (2015) were all significantly associated with the AER_{DGRP} of *tan* (Fig. S6B, Supporting information).

LD analyses

Our data showed that there were many potential *cis*-regulatory elements outside the core *cis*-element that could affect the expression level of *ebony*. These elements are likely to be interacting with one another to control the expression of the gene. Therefore, if multiple functional *cis*-regulatory alleles with different combinations of those elements are segregating in a population, the interactions between those scattered elements may be detected as long distance LD. The intensity of LD is highly dependent on local recombination rate, and decay of LD (R^2) with distance in the *ebony cis*-regulatory region (*e_cis*, ~10 kb) was, therefore, compared to that of the surrounding ~350 kb region (Fig. S7A, Supporting information). Plots and nonlinear regression lines showed that the overall LD level of the *ebony cis*-regulatory region was similar to that of the surrounding region, with a slightly faster decay of LD with distance, indicating that there was no reduction in the population recombination rate in this region. However, a noticeable number of distant SNP pairs with high LD were observed. In particular, the number of SNP pairs that were ≥3 kb apart with $R^2 \geq 0.5$ was enriched in the *e_cis* region (449 of 19 306 SNP pairs in *e_cis*, and 5562 of 833 841 in the ~350 kb region; Fisher's exact test $P < 10^{-16}$). To investigate this further, the percent of SNP pairs that were ≥3 kb apart with $R^2 \geq 0.5$ were plotted using a sliding 10 kb window (Fig. S8, Supporting information). The sliding window plots indicated that the *e_cis* region exhibited one of several peaks in the ~350 kb region. These results indicated a slight excess of distantly associated SNPs in the *e_cis* region compared to the surrounding regions. Similar analysis of the *tan* gene region identified showed a very small number of distantly associated SNPs (Fig. S7B, Supporting information) and demonstrated faster decay of R^2 with nucleotide distance in both the 10 kb *t_cis* region and the surrounding ~300 kb region.

Discussion

Our precise quantification of allele-specific expression levels and pigmentation properties in a *D. melanogaster* population revealed that *cis*-regulatory variation in *ebony* accounted for pigmentation intensities of thoracic segment, but at least not as strongly for the extent of abdominal pigmentation (Table 1). There was no detectable contribution of *tan* and *black*, which is involved in the same reactions between two metabolites, dopamine and beta-alanyl-dopamine, in the melanin biosynthesis pathway (Wright 1987; Wittkopp *et al.* 2003). There was also no contribution of *pale* (a hydrogenase gene), a gene that catalyses the first step of the metabolic pathway.

Our findings in this study, albeit from a limited number of genes, correspond with data from previous studies that suggest most of the pigmentation trait variations are influenced by a small number of genes with large effects (Kopp *et al.* 2003; Takahashi *et al.* 2007; Bastide *et al.* 2013; Dembeck *et al.* 2015). Our analyses focused on effector enzymes in the melanin biosynthesis pathway and did not include their *trans*-acting upstream regulators. Bickel *et al.* (2011) have conducted extensive analyses on the *cis*-regulatory polymorphisms of one of the regulators, *bric-a-brac*, and they found that SNPs in the three different regulatory regions affect allele-specific expression of *tan*. Using a similar approach to conduct more extensive analyses on the *cis*-regulatory effects of other *trans*-acting upstream regulators, such as *Abdominal-B*, *optomotor blind*, *hedgehog*, *wingless* and *decapentaplegic* (reviewed in Wittkopp *et al.* 2003) may allow the confirmation and/or identification of other genetic factors that contribute to natural pigmentation variation.

The ~450 bp core *cis*-element of *tan* (*t_core_cis*) was identified as harbouring variants responsible for pigmentation differences between *D. yakuba* and *D. santomea* in sexually dimorphic abdominal segments (Jeong *et al.* 2008). Pigmentation-associated SNPs in *t_core_cis* from European *D. melanogaster* populations were associated with the extent of pigmentation in the sexually dimorphic A7 region (Bastide *et al.* 2013), and those SNPs in the DGRP were also associated with pigmentation in sexually dimorphic A5 and A6 segments (Dembeck *et al.* 2015). All these SNPs were significantly associated with the AER_{DGRP} of *tan* in our study (Fig. S6B, Supporting information), despite the lack of associations detected between AER_{DGRP} of *tan* and pigmentation traits from our samples (Table 1). It should be noted that because AER was determined from the whole headless body rather than from particular body segments, and because of the effect of *cis*–*trans* interaction on AER_{DGRP} of *tan* (Fig. S5B, Supporting information), the resolution for determining *cis*-regulatory effects on *tan* gene expression may be limited. However, our data indicated that these SNPs in the core *cis*-regulatory region are likely to be actually determining the intraspecific variation in expression level of this gene.

The sequence polymorphisms in the DGRP samples provided a rare opportunity to investigate the potential *cis*-regulatory mutational targets in the *ebony* gene region. Within-haplotype comparisons between dark and light strains revealed that there were at least three different regions outside the core *cis*-regulatory element that could affect gene expression (Fig. 2C). A reporter assay conducted by Rebeiz *et al.* (2009) showed that the core enhancer located upstream of the *ebony* transcription start site activated expression in the abdominal and thoracic epidermis, and regulatory sequences located in the intron repressed expression in abdominal stripes. Evolutionarily conserved male-specific silencers also reside in the upstream sequence outside the core enhancer of this gene (Rebeiz *et al.* 2009; Ordway *et al.* 2014; Johnson *et al.* 2015). These results suggested that the regulatory architecture of *ebony* may not resemble a classic example with a discrete modular enhancer, but that expression might be affected profoundly by sequences outside the core enhancer region. However, it has also been shown by a precise sequence analysis on the *cis*-regulatory SNPs of the *bric-a-brac* locus that most causative nucleotide polymorphisms for transcription reside around several core functional elements, but not inside the core enhancer (Bickel *et al.* 2011). Therefore, diffused targets for *cis*-regulatory mutations outside core elements may be a common feature of regulatory variation within species.

Previous studies on populations from Africa and southern Japan (Rebeiz *et al.* 2009; Takahashi & Takano-Shimizu 2011), and this study of a North American population, showed that different *cis*-regulatory variants were associated with expression variability in different populations (Fig. S9, Supporting information). These results indicated that the multiple combinations of mutations can give rise to similar expression phenotypes. Corroborating this, our data showed that even within the same natural population, many different functional mutations with similar effects were segregating in the *cis*-regulatory region of *ebony*. This feature resembles the *cis*-regulatory polymorphisms at the *tan* locus that underlie intraspecific pigmentation variation in *D. americana* (Wittkopp *et al.* 2009).

Because variations in pigmentation traits are observed universally, including within ancient populations in Africa (David & Capy 1988; Lachaise *et al.* 1988; Pool & Aquadro 2007; Rebeiz *et al.* 2009), the standing variation is likely to have persisted for a long time. A close investigation of the sequence polymorphisms in the responsible genes, *ebony* and *tan*, indicated that unlike genes under strong balancing selection that show long-term persistence of a similar set of alleles (*e.g.* Clark & Kao 1991; Klein *et al.* 1993), different populations possessed different segregating mutations that were associated with similar phenotypic variations (Fig. S9, Supporting information). However, unlike neutrally evolving loci, some signatures of the action of natural selection targeting the *ebony* gene were observed. An identical haplotype block in the *ebony cis*-regulatory region was present in the two darkest strains (Fig. 2C), and this block differed from the haplotype blocks reported by Pool & Aquadro (2007) in the darkest African strains (Fig. S9A, Supporting information). Similarly, the two darkest strains in our study did not harbour the pigmentation-associated SNP types of either the African or Japanese populations (Fig. S9B,C, Supporting information). These long-range haplotypes suggest that there may have been multiple selective events in the past that favoured dark-pigmented phenotypes and targeted different alleles. These alleles may have arisen independently in different populations from new mutations, from new combinations of segregating SNPs, or both. In the Iriomote population from southern Japan, a strong association of the Light-type allele with the inversion *In(3R)Payne*, which is prevalent in tropical areas, was indicative of selection favouring a light pigmentation phenotype. Possible selective pressures might include environmental factors such as temperature, humidity and UV radiation, which were previously shown to correlate with body colour in populations sampled from different continents (reviewed in True 2003; Takahashi 2013; Bastide *et al.* 2014).

Because those environmental factors tend to fluctuate seasonally and annually, one possible explanation for the maintenance of functional polymorphisms at the *cis*-regulatory region of *ebony* would be to assume that selection is temporal and/or fluctuating. Occasional strong positive selection may produce long-ranged haplotype structures, but a relaxation of the selective pressure in subsequent generations may cause an eventual admixture of haplotypes (Charlesworth 2006). Moreover, if the selected allele is subjected to balancing selection as a result of a fluctuating environment, it may not rise to a high frequency and may segregate along with other alleles in the population. This type of short-term balancing selection scenario is consistent with the occasionally observed long-range haplotype structure at a low frequency surrounding the *ebony* gene (Fig. 2; Pool & Aquadro 2007). Therefore, fluctuating selective forces on the standing genetic variation, perhaps together with a high migration rate, might explain the maintenance of variable functional alleles within different populations.

Our LD analyses of the *ebony* gene region indicated that although there was no reduction in the population recombination rate in the *cis*-regulatory region of this gene, a slight enrichment of distantly associated SNP pairs was observed (Figs S7A and S8, Supporting information). This is suggestive of interacting *cis*-regulatory

elements scattered along the ~10 kb region. However, identification of these elements may not be straightforward because the interactions may be complex and the effects may be weak. In the case of *tan*, the causative SNPs are likely to reside inside the known core *cis*-regulatory element (Fig. S6, Supporting information). Unlike in the case of *ebony*, the *cis*-regulatory region of *tan* did not exhibit an excess of distantly located SNP pairs in high LD (Fig. S7B, Supporting information). This could be due to the higher population recombination rate compared to the *ebony* gene region. Nevertheless, different structural properties and/or selective histories of the *cis*-regulatory regions of *ebony* and *tan* may also have contributed to the differences between these genes. Collectively, together with the case of the *bric-a-brac* locus (Bickel *et al.* 2011), our data suggest that the underlying genetic differences in the *cis*-regulatory regions controlling intraspecific pigmentation variation are more complex than those typically observed for interspecific pigmentation trait differences (Gompel *et al.* 2005; Jeong *et al.* 2006, 2008; Prud'homme *et al.* 2006), where causal genetic changes are confined to modular enhancer elements.

Acknowledgements

We thank the Bloomington Stock Center for fly stocks and Koichiro Tamura for valuable discussions. This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Ayroles JF, Carbone MA, Stone EA *et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, **41**, 299–307.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Bastide H, Betancourt A, Nolte V *et al.* (2013) A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genetics*, **9**, e1003534.

Bastide H, Yassin A, Johanning EJ, Pool JE (2014) Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan Africa. *BMC Evolutionary Biology*, **14**, 179.

Bickel RD, Kopp A, Nuzhdin SV (2011) Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. *PLoS Genetics*, **7**, e1001275.

Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, **28**, 495–503.

Borycz J, Borycz JA, Loubani M, Meinertzhagen IA (2002) *tan* and *ebony* genes regulate a novel pathway for transmitter metabolism at fly photoreceptor terminals. *The Journal of Neuroscience*, **22**, 10549–10557.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, e64.

Clark AG, Kao TH (1991) Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Sciences, USA*, **88**, 9823–9827.

David JR, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics: TIG*, **4**, 106–111.

David JR, Capy P, Payant V, Tsakas S (1985) Thoracic trident pigmentation in *Drosophila melanogaster*: differentiation of geographical populations. *Genetique, Selection, Evolution*, **17**, 211–224.

Dembeck LM, Huang W, Magwire MM, Lawrence F, Lyman RF, Mackay TF (2015) Genetic architecture of abdominal pigmentation in *Drosophila melanogaster*. *PLoS Genetics*, **11**, e1005163.

Gavin BA, Arruda SE, Dolph PJ (2007) The role of carcinine in signaling at the *Drosophila* photoreceptor synapse. *PLoS Genetics*, **3**, e206.

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

Gurganus MC, Nuzhdin SV, Leips JW, Mackay TF (1999) High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. *Genetics*, **152**, 1585–1604.

Huang W, Massouras A, Inoue Y *et al.* (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research*, **24**, 1193–1208.

Jacobs ME (1960) Influence of light on mating of *Drosophila melanogaster*. *Ecology*, **41**, 182–188.

Jeong S, Rokas A, Carroll SB (2006) Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell*, **125**, 1387–1399.

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB (2008) The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell*, **132**, 783–793.

Johnson WC, Ordway AJ, Watada M, Pruitt JN, Williams TM, Rebeiz M (2015) Genetic changes to a transcriptional silencer element confers phenotypic diversity within and between *Drosophila* species. *PLoS Genetics*, **11**, e1005279.

Klein J, Satta Y, O'HUigin C, Takahata N (1993) The molecular descent of the major histocompatibility complex. *Annual Review of Immunology*, **11**, 269–295.

Kopp A, Duncan I (1997) Control of cell fate and polarity in the adult abdominal segments of Drosophila by optomotor-blind. *Development*, **124**, 3715–3726.

Kopp A, Duncan I, Godt D, Carroll SB (2000) Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature*, **408**, 553–559.

Kopp A, Graze RM, Xu S, Carroll SB, Nuzhdin SV (2003) Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics*, **163**, 771–787.

Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) *Historical Biogeography of the Drosophila Melanogaster Species Subgroup*. Plenum Press, New York, New York.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL (2005) Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics*, **171**, 1813–1822.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF (1995) High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics*, **139**, 1273–1291.

Mackay TF, Lyman RF (2005) *Drosophila* bristles and the nature of quantitative genetic variation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1513–1527.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* genetic reference panel. *Nature*, **482**, 173–178.

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ (2010) Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Research*, **20**, 816–825.

Mezey JG, Houle D, Nuzhdin SV (2005) Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. *Genetics*, **169**, 2101–2113.

Morgan TH, Bridges CB (1919) The inheritance of a fluctuating character. *The Journal of General Physiology*, **1**, 639–643.

Munjal AK, Karan D, Gibert P, Moreteau B, Parkash R, David JR (1997) Thoracic trident pigmentation in *Drosophila melanogaster*: latitudinal and altitudinal clines in Indian populations. *Genetics Selection Evolution*, **29**, 601–610.

Newby LM, Jackson FR (1991) *Drosophila ebony* mutants have altered circadian activity rhythms but normal eclosion rhythms. *Journal of Neurogenetics*, **7**, 85–101.

Nuzhdin SV, Dilda CL, Mackay TF (1999) The genetic architecture of selection response. Inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. *Genetics*, **153**, 1317–1331.

Ordway AJ, Hancuch KN, Johnson W, Wiliams TM, Rebeiz M (2014) The expansion of body coloration involves coordinated evolution in *cis* and *trans* within the pigmentation regulatory network of *Drosophila prostipennis*. *Developmental Biology*, **392**, 431–440.

Parkash R, Sharma V, Kalra B (2008) Climatic adaptations of body melanisation in *Drosophila melanogaster* from Western Himalayas. *Fly*, **2**, 111–117.

Parkash R, Rajpurohit S, Ramniwas S (2009) Impact of darker, intermediate and lighter phenotypes of body melanization on desiccation resistance in *Drosophila melanogaster*. *Journal of Insect Science*, **9**, 1–10.

Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Molecular Ecology*, **16**, 2844–2851.

Prud'homme B, Gompel N, Rokas A *et al.* (2006) Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature*, **440**, 1050–1053.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science*, **326**, 1663–1667.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675.

Shrimpton AE, Robertson A (1988a) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics*, **118**, 437–443.

Shrimpton AE, Robertson A (1988b) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics*, **118**, 445–459.

Suh J, Jackson FR (2007) *Drosophila* Ebony activity is required in glia for the circadian regulation of locomotor activity. *Neuron*, **55**, 435–447.

Suvorov A, Nolte V, Pandey RV, Franssen SU, Futschik A, Schlötterer C (2013) Intra-specific regulatory variation in *Drosophila pseudoobscura*. *PLoS ONE*, **8**, e83547.

Takahashi A (2013) Pigmentation and behavior: potential association through pleiotropic genes in *Drosophila*. *Genes & Genetic Systems*, **88**, 165–174.

Takahashi A, Takano-Shimizu T (2005) A high-frequency null mutant of an odorant-binding protein gene, *Obp57e*, in *Drosophila melanogaster*. *Genetics*, **170**, 709–718.

Takahashi A, Takano-Shimizu T (2011) Divergent enhancer haplotype of *ebony* on inversion *In(3R)Payne* associated with pigmentation variation in a tropical population of *Drosophila melanogaster*. *Molecular Ecology*, **20**, 4277–4287.

Takahashi A, Takahashi K, Ueda R, Takano-Shimizu T (2007) Natural variation of *ebony* gene controlling thoracic pigmentation in *Drosophila melanogaster*. *Genetics*, **177**, 1233–1237.

Telonis-Scott M, Hoffmann AA, Sgrò CM (2011) The molecular genetics of clinal variation: a case study of *ebony* and thoracic trident pigmentation in *Drosophila melanogaster* from eastern Australia. *Molecular Ecology*, **20**, 2100–2110.

True JR (2003) Insect melanism: the molecules matter. *Trends in Ecology & Evolution*, **18**, 640–647.

True JR, Yeh SD, Hovemann BT *et al.* (2005) *Drosophila tan* encodes a novel hydrolase required in pigmentation and vision. *PLoS Genetics*, **1**, e63.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM (2011) Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genetics*, **7**, e1001336.

Wagner S, Heseding C, Szlachta K, True JR, Prinz H, Hovemann BT (2007) *Drosophila* photoreceptors express cysteine peptidase Tan. *The Journal of Comparative Neurology*, **500**, 601–611.

Weber K, Eisman R, Morey L *et al.* (1999) An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. *Genetics*, **153**, 773–786.

Weber K, Eisman R, Higgins S *et al.* (2001) An analysis of polygenes affecting wing shape on chromosome 2 in *Drosophila melanogaster*. *Genetics*, **159**, 1045–1057.

Wittkopp PJ, Carroll SB, Kopp A (2003) Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends in Genetics: TIG*, **19**, 495–504.

Wittkopp PJ, Stewart EE, Arnold LL *et al.* (2009) Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science*, **326**, 540–544.

Wright TR (1987) The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. *Advances in Genetics*, **24**, 127–222.

Zimmerman E, Palsson A, Gibson G (2000) Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics*, **155**, 671–683.

R.M. designed the study, conducted laboratory work, analysed data and wrote the manuscript; N.A. conducted laboratory work and analysed data; N.O. analysed data and commented on the manuscript; A.T. designed the study, analysed data and wrote the manuscript.

Data accessibility

Sequence data obtained by Sanger sequencing are deposited in DDBJ/EMBL/Genbank databases with accession numbers LC037995–LC038014 and LC038168–LC038187. Amp-seq fastq and bam files are submitted to DDBJ database under the accession numbers DRA004064 and DRZ007412 respectively. Pigmentation data (photographs and raw pigmentation scores) are deposited in Dryad Digital Repository (doi:10.5061/dryad.d4d19).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Pigmentation scores of the females at 3 and 6 days post eclosion.

Fig. S2 Crossing scheme performed to quantify allele-specific expression levels.

Fig. S3 Pigmentation scores of the 20 DGRP strains used in this study.

Fig. S4 Allelic expression ratio of the DGRP-derived allele (AER_{DGRP}) of the DGRP strains used in this study.

Fig. S5 *Cis*- and *trans*-regulatory effects estimated by parent-hybrid comparisons.

Fig. S6 Sequence polymorphisms in the core *cis*-regulatory element of *tan* in the DGRP strains used in this study.

Fig. S7 Decay of linkage disequilibrium (LD) with physical nucleotide distance.

Fig. S8 Sliding window analysis of LD (R^2) in ~350 kb region surrounding *ebony*.

Fig. S9 Comparison of RAL-358/324/707/555/820 haplotype in the *e_cis* region with haplotypes from African and Iriomote populations.

Table S1 Primer pairs used in 48.48 Access Array PCR

Table S2 Information on target SNP nucleotide positions, primer pairs and read counts from Amp-seq

Table S3 Quantification of expression levels of internal control genes by qRT-PCR