

Association of *orthodenticle* with Natural Variation for Early Embryonic Patterning in *Drosophila melanogaster*

LISA M. GOERING^{1,2*}, PRISCILLA K. HUNT¹, CASSANDRA HEIGHINGTON^{1,3}, CHRISTOPHER BUSICK⁴, PLEUNI S. PENNING⁵, JOACHIM HERMISSON⁶, SUDHIR KUMAR⁴, AND GREG GIBSON^{1,7}

¹Department of Genetics, North Carolina State University, Raleigh, North Carolina

²Department of Biological Sciences, St. Edward's University, Austin, Texas

³Department of Genetics, University of Georgia, Athens, Georgia

⁴Center for Evolutionary Functional Genomics, The BioDesign Institute, Arizona State University, Tempe, Arizona

⁵Department of Biology II, Ludwig-Maximilian University, Munich, Germany

⁶Department of Mathematics and Max F. Perutz Laboratories, University of Vienna, Vienna, Austria

⁷School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

ABSTRACT Although it is well established that *cis*-acting regulatory variation contributes to morphological evolution between species, few concrete examples of polymorphism affecting developmental patterning within species have been demonstrated. Early embryogenesis in *Drosophila* is initiated by a gradient of Bicoid morphogen activity that results in differential expression of multiple target genes. In a screen for genetic variation affecting this process, we surveyed 96 wild-type lines of *Drosophila melanogaster* for polymorphisms in binding sites within 16 Bicoid *cis*-regulatory response elements. One common polymorphism in the *orthodenticle* (*otd*) early head enhancer is associated with a complex series of indels/substitutions that define two distinct haplotypes. The middle region of this enhancer exhibits an unusual pattern of nucleotide diversity that does not easily fit into standard models of selection and demography. Population Gene Expression Maps, generated by extracting binary expression profiles from normalized embryo images, revealed a ventral reduction of *otd* transcript abundance in one of the haplotypes that was recapitulated in expression of transgenic constructs containing the two alleles. We thus demonstrate that even a process as robust as early developmental patterning is affected by standing genetic variation, intriguingly involving *otd*, whose morphogenetic function *bicoid* is thought to have displaced during dipteran evolution. *J. Exp. Zool. (Mol. Dev. Evol.)* 312B:841–854, 2009. © 2009 Wiley-Liss, Inc.

How to cite this article: Goering LM, Hunt PK, Heighington C, Busick C, Pennings P, Hermisson J, Kumar S, Gibson G. 2009. Association of *orthodenticle* with natural variation for early embryonic patterning in *Drosophila melanogaster*. *J. Exp. Zool. (Mol. Dev. Evol.)* 312B:841–854.

The basic patterning processes of early embryogenesis have been conserved for tens of millions of years in the genus *Drosophila*. There is nevertheless a constant tension between robustness and flexibility during ontogeny, and this gives rise to a surprising evolutionary lability of detailed features of the patterning process that is observed between genera and species (Peel et al., 2005;

Grant sponsor: US National Institutes of Health; Grant number: 2R01GM61600; Grant sponsor: Australian Research Council; Grant number: DP0880204.

*Correspondence to: Lisa M. Goering, Department of Biological Sciences, St Edward's University, Austin, TX 78704. E-mail: lisago@stedwards.edu

Received 14 January 2009; Revised 6 April 2009; Accepted 4 May 2009

Published online 1 June 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21299

Damen, 2007). The fundamental morphogen Bicoid (Bcd), that sets up the gradient of anterior-posterior (A-P) positional information (Ephrussi and St Johnston, 2004) is actually an invention of higher diptera that seems to have taken over a role largely performed by another transcription factor, Orthodenticle (Otd), in more primitive insects (Schroder, 2003; Lynch et al., 2006). One of the target genes of Bicoid, the pair-rule gene *even-skipped* (*eve*), initiates segmentation, and expression in the second stripe of *eve* is controlled by an enhancer that has evolved extensive substitutions despite retaining its overall function between divergent *Drosophila* species (Ludwig et al., 2005). Even closely related species of *Drosophila* show alterations in the timing of early pattern formation (Kim et al., 2000). These observations suggest that there is also likely to be functional polymorphism for early embryonic patterning within a species, but the extent and nature of such variation has not been documented.

Bicoid protein is distributed in a gradient, and activates transcription of target genes at different positions along the A-P axis (Driever and Nüsslein-Volhard, '88a,b; Driever et al., '89; Ephrussi and St Johnston, 2004; Spirov et al., 2009). Any differences in the amount of *bcd* mRNA deposited at the anterior pole of the embryo as a result of maternal genotype or nutrition, or any physiological variation in the diffusion of Bcd protein and mRNA after fertilization, will alter the gradient and hence the position and timing of transcriptional responses. Yet early embryonic patterning mediated by Bcd takes place robustly despite variation for embryonic size and shape, and across a range of temperatures, and segmentation defects are presumably as rare in nature as in the laboratory (Houchmandzadeh et al., 2002). Manu et al. (2009) demonstrate that cross-regulation among the gap genes is essential for robustness during development, but it remains to be established whether this homeostatic mechanism actually permits cryptic variation to be maintained in populations, or whether evolutionary canalization simply reflects the purging of any variation that would tend to disrupt embryogenesis.

In order to identify potential sources of genetic variation in the response to Bcd, we surveyed 16 well-characterized short *cis*-regulatory regions (Ochoa-Espinoza et al., 2005) associated with a dozen genes, in a total of 96 highly inbred lines from two populations of *Drosophila melanogaster*. Among several polymorphisms identified, one in the promoter of the aforementioned *orthodenticle*

gene stands out as unusual in structure and diversity, and we subsequently show that it is part of a functionally distinct common haplotype dimorphism. The results build a bridge between studies of evolution and development between species, and quantitative genetic variation within (Stern, 2000), and demonstrate that even the earliest and most highly conserved developmental patterning processes are genetically variable.

MATERIALS AND METHODS

Fly culture

Individual, female *D. melanogaster* were collected in the summer of 2004 and established as isofemale lines. The North Carolina population (NC2) was caught by us in a peach orchard near West End, North Carolina; the Maine (ME) population was collected by Martin Kreitman in a blueberry field close to Cherryfield, Maine. Lines were bred to isogenicity through 15–20 generations of full sib-mating. Five African *D. melanogaster* lines (MW6, MW12, MW29, MW46 and MW51), and one *D. sechellia* line (Robertson 3C) were provided by Corbin Jones (UNC). The *D. simulans* line NC112T was provided by Richard Lyman and Trudy Mackay (NCSU). All lines were kept on 12 hr light–dark cycles in vials with 10 mL standard cornmeal medium supplemented with yeast.

Sequencing

All genotype data for 48 NC2 and 48 ME lines was obtained by direct sequencing of PCR products amplified from genomic DNA extracted from a single female fly. PCR amplification and sequencing was performed by Genaisance Pharmaceuticals (now Cogenics Division of Clinical Data, New Haven, CT); primer sequences will be provided upon request. All fragments were unidirectionally sequenced using the forward PCR primer except for *Dichaete* that was bidirectionally sequenced. Base calling and initial contig assembly was performed with Phred and Phrap (Codon-Code, Deadham, MA). Alignment of multiple contigs, manual sequence editing, and single nucleotide polymorphism (SNP) calling was performed with BioLign software (Tom Hall and Ed Buckler; <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). *D. simulans* (Release 1.0, April 2005) and *D. yakuba* (Release 2.0, November 2005) outgroup sequences were produced by the

Washington University School of Medicine Genome Sequencing Center, St. Louis, MO.

Additional sequencing in the *otd* Bicoid cis-regulatory response element (BCRE) region was done in six African *D. melanogaster* lines, one *D. simulans* line, and one *D. sechellia* line. PCR primers (*otd*SSM F: 5'-gggaggetaggttaggcgtag-3'; *otd*SSM R: 5'-ggacaacaatgaaggcgatt-3') were designed that would amplify in all three species; amplification was performed with Taq polymerase (Promega, Madison, WI) at an annealing temperature of 54°C for 35 cycles. For the *D. simulans* and *D. sechellia* lines, PCR products generated from genomic DNA were directly sequenced with the forward and reverse PCR primers, using BigDye Terminator v. 3.1 (Applied Biosystems, Foster City, CA) on ABI 3700 sequencers at the NCSU Genome Research Laboratory. Because the African lines are not isogenic, PCR products were first cloned using the Promega pGEM-T Easy TA cloning kit, and four clones per line were sequenced using plasmid DNA as the template.

The sequence was annotated for Bcd binding sites in two ways. Approximately one-third of the BCREs had been previously characterized, with binding sites identified and functionally verified in some way (see Gao and Finkelstein, '98; Driever et al., '89; Howard and Struhl, '90; Small et al., '92; Rivera-Pomar et al., '95; Kuhnlein et al., '97; La Rosee et al., '97; Fujioka et al., '99). The other BCREs had been identified as potential Bcd targets by computational means (see Ochoa-Espinoza et al., 2005); these BCREs were searched for potential Bcd binding sites using TFSEARCH (Yutaka Akiyama: <http://www.rwcp.or.jp/papia>) using a threshold score of 80. TFSEARCH makes use of the TRANSFAC transcription factor database (Heinemeyer et al., '98).

All sequences have been deposited in the NCBI PopSet database as alignments and can be accessed using the following accession numbers: *CG9571* (FJ713898); *Dichaete* (FJ713992); *eve1* (FJ714081); *eve2* (FJ714176); *gt1* (FJ14270); *gt23* (FJ14362); *hairy0* (FJ714453); *hairy2* (FJ714543); *hairy7* (FJ714632); *hbP2* (FJ71425); *knirps* (FJ14810); *otd* early (FJ714902); *salmBE* (FJ715017); *slpA* (FJ715200); *slpB* (FJ715200); *tll* (FJ715292); *otd* early MW lines (FJ714921); *otd* early—*D. simulans* and *D. sechellia* (FJ714923).

Analysis of nucleotide diversity

DNA polymorphism data was analyzed in DnaSP version 4.0 (Rozas et al., 2003). Nucleotide

diversity was estimated as the average number of pairwise differences per site (π) and as Watterson's estimate of Theta (θ_w) based on the number of segregating sites. Tests of neutrality (Tajima's *D*, Fu and Li's *D*) were also performed in DnaSP; *D. simulans* sequence was used as the outgroup for Fu and Li's *D* test. Divergence between *D. melanogaster* and *D. simulans* was calculated in DnaSP using the Jukes and Cantor method. Linkage disequilibrium (LD) was assessed in Tassel version 2.0.1 (Ed Buckler, <http://www.maizegenetics.net>) using Fisher's exact test to assess the significance of the squared correlation of the allele frequencies (r^2). Only sites with a minor allele frequency of 0.05, and that were present in at least 30 alleles were used in the analysis. Haplotype networks were constructed using TCS version 1.21 (Clement et al. 2000), which estimates phylogenetic networks using the statistical parsimony algorithm (Templeton et al., '92). Recombination rates were estimated using the Drosophila Recombination Rate Calculator (http://petrov.stanford.edu/cgi-bin/recombination-rates_updateR5.pl; Singh et al., 2005). A summary of haplotype sequences is shown in Table 1.

Generation of transgenic *otd*-BCRE lines

The 925 bp *otd* early head enhancer was PCR amplified from D and I haplotype lines with the primers F: 5'-AACCGAATTCCTAGGCCCGA-GATTAAGTACCG-3' and R: 5'-AACCGGCG-CGCCGATACAGATCTCGTGGATTGC-3'. The resulting PCR fragments contained an EcoRI and AscI site on the 5' and 3' ends respectively. Fragments were cloned into pAOE1 (a gift of A. Ochoa-Espinoza), which contains the *eve* basal promoter fused to the *lacZ* gene and the α -*tubulin* 3' UTR (Ochoa-Espinoza et al., 2005). Transgenic lines were generated using standard injection techniques by the Duke University Model System Genomics facility (Durham, NC). Five and six independent lines were obtained for the D and I haplotype reporter constructs, respectively.

In situ hybridization

Expression of endogenous *otd* mRNA was analyzed by in situ hybridization using an antisense *otd* probe generated from DNA clone RE10280 (GenBank ID: BI169182) obtained from the Berkley *Drosophila* Genome Project (BDGP) EST project (www.fruitfly.org/EST). Reporter gene expression was analyzed using an antisense *lacZ* probe. Females were allowed to lay eggs on apple juice-agar

TABLE 1. *otd* enhancer haplotypes

HapI consensus (25 ME; 15 NC2)	C C C C G C A A A T T A C C A G T G G G
HapI-2 (1 ME; 5 NC2)	A G
HapI-3 (3 ME; 0 NC2) C A
HapI-4 (0 ME; 3 NC2)	A G T
NC2-24	A G C A
NC2-01	A G A A A
NC2-27	. . A A
HapD consensus (14 ME; 12 NC2)	A G A A A T C - G A C T T T T T G C . .
HapD-2 (0 ME; 2 NC2)	. . A A A T C - G A C T T T T T G
HapD-3 (3 ME; 1 NC2)	A G A A A T C - G A C T T T T T G
HapD-4 (0 ME; 2 NC2)	A G C A A T C - G A C T T T T T G C
ME48	A G - A A T C - G A C T T T T T G C - -
NC2-37	A G - A A T C - G A C T T T T T G C
NC2-31	A G A A A T C - G A C T T T T T G C - -
NC2-40	A G . . A T C - G A C T T T T T G C
NC2-16	A G A A A T C - G A C T T T T T G . C A

HapI consensus: ME03, ME04, ME05, ME06, ME07, ME09, ME11, ME12, ME16, ME17, ME19, ME21, ME25, ME26, ME29, ME30, ME31, ME32, ME33, ME35, ME36, ME38, ME42, ME44, ME47, NC2-04, NC2-05, NC2-06, NC2-07, NC2-09, NC2-11, NC2-13, NC2-14, NC2-18, NC2-19, NC2-30, NC2-36, NC2-38, NC2-42, NC2-44. *HapI-2*: ME02, NC2-02, NC2-12, NC2-22, NC2-46, NC2-47. *HapI-3*: ME37, ME41, ME46. *HapI-4*: NC2-21, NC2-23, NC2-28. *HapD* consensus: ME08, ME10, ME15, ME18, ME20, ME22, ME23, ME27, ME28, ME34, ME39, ME40, ME43, ME45, NC2-03, NC2-08, NC2-17, NC2-20, NC2-25, NC2-32, NC2-33, NC2-34, NC2-35, NC2-41, NC2-43, NC2-45. *HapD-2*: NC2-15, NC2-29. *HapD-3*: ME01, ME14, ME24, NC2-39. *HapD-4*: NC2-10, NC2-26.

plates for 3 hr, after which embryos were dechorionated and fixed. Fixation and hybridization was performed as described by Tautz and Pfeifle ('89).

Phenotypic analysis

After in situ hybridization, embryos were mounted on microscope slides in 70% glycerol/1X PBS. Lateral images of blastoderm-stage embryos were captured at 20 \times magnification, using a Nikon Eclipse E800 microscope with camera attachment. Because we were interested in differences between haplotypes, and not within lines, single embryos from 99 wild type (WT) lines (38 Hap D, 61 Hap I) and 11 transgenic lines (6 *otd*ED, 5 *otd*EI) were analyzed. To analyze differences in the ventral domain of *otd* expression, a line was drawn from dorsal to ventral near the anterior border of the expression domain. Pixel intensity along the line was calculated using NIH ImageJ software (<http://rsb.info.nih.gov/ij/>) and plotted against the percent egg height (dorsal margin = 0%, ventral margin = 100%). Note that dark colors have a lower pixel intensity than light colors; therefore, as staining decreases the relative pixel intensity increases (black would be 0 pixel intensity, a complete absence of color). The inflection point was determined as the percent egg height at which at least four points in a row increased in pixel intensity above background levels for that image. The difference in distribution of inflection points among D and I haplotypes

within wild type or transgenic lines was assessed using a nonparametric Mann–Whitney/Wilcoxon sum-rank test implemented in JMP Genomics (SAS Institute, Cary, NC). A second replicate of single images from all lines was performed, with similar results (data not shown). To analyze differences in the anterior boundary of the *otd* expression domain, a line was drawn from the anterior tip to posterior (center of the pole cells). Pixel intensity along the line was calculated using NIH ImageJ software (<http://rsb.info.nih.gov/ij/>) and plotted against the percent egg length (posterior = 0%, anterior = 100%). The border of the expression domain was determined as the percent egg length at which at least four points in a row increased in pixel intensity at the expression boundary. The difference in the position of anterior expression boundaries among D and I haplotypes within wild type or transgenic lines was also assessed using a Mann–Whitney/Wilcoxon test.

Population Gene Expression Maps (popGEMs) are based on carefully standardized and aligned images (Kumar et al., 2002; see also www.flyexpress.net). In this standardization process, the embryo is first enclosed in a rectangle with sides tangent to the perimeter of the embryo. This is followed by rectangle rotation to ensure that anterior of the embryo is on the left and that the A-P axis is parallel to the horizontal axis. After cropping the image, all embryos are size-standardized to a 320 \times 128 pixel dimension, such that each row of pixels in the image as well as each

image ends on a byte, word and long word boundaries, which obviates the need for pixel padding. The aspect ratio chosen is based on the average aspect ratio found in the collection of original in situ images in FlyExpress (www.flyexpress.net). The size-standardized images are then subjected to expression pattern extraction from the foreground, which eliminates the background (including noise; see methods described in Kumar et al. (2002) and Gurunathan et al. (2004). This process produces a binary expression extraction (BEE) for each in situ image. All binary patterns from each class of image (Hap D, Hap I, *otdED*, and *otdEI*) are summarized in a two-dimensional grid, with each spatial coordinate containing the number of images (from different individuals) exhibiting a binary expression. In a popGEM, the color depth reflects the relative fraction of individuals that show expression at the given location. popGEMs were generated from the same images as above, for a total of 99 WT images (38 Hap D, 61 Hap I) and 11 transgenic images (6 *otdED*, 5 *otdEI*).

RESULTS

Polymorphism in Bicoid cis-regulatory response elements

The sequenced BCRES averaged 726 bp in length and have been shown to drive reporter gene

expression in transgenic embryos with boundaries of expression at specific positions along the embryo (Hoch et al., '90; Howard and Struhl, '90; Small et al., '92; Rivera-Pomar et al., '95; Kuhnlein et al., '97; La Rosee et al., '97; Gao and Finkelstein, '98; Fujioka et al., '99; Ochoa-Espinoza et al., 2005). Previous studies of these BCRES suggest that gene expression along the A-P axis is owing to combinatorial inputs of Bcd and other transcription factors in the regulatory network (Ochoa-Espinoza et al., 2005). Levels of sequence diversity were estimated on the basis of average nucleotide heterozygosity (π) and Watterson's estimate of theta (θ_w) (Moriyama and Powell, '96; Hutter et al., 2007). Overall levels of polymorphism in the BCRES are low (Table 2; avg $\pi = 0.0047$, avg $\theta_w = 0.0046$), but within the range of other noncoding regions in *D. melanogaster* (Andolfatto, 2005), and the expected correlation between diversity and recombination rate first described by Begun and Aquadro ('92) was observed (not shown). We had postulated that there may be a relationship between sequence diversity and level of expression along the A-P axis, but sequence diversity was too low to detect such a trend.

Our data set contained 119 potential Bcd binding sites that cover approximately 10% of the sequence. 34 of these sites had been functionally characterized previously, and 85 sites were

TABLE 2. Nucleotide diversity and tests of neutrality in BCRES

BCRE ^a	Length (bp)	π	θ_w	Tajima's D	Fu and Li's D	Total sequences ^b (ME/NC2)
<i>CG9571</i>	734	0.003	0.002	0.35	1.15	(38/44)
<i>Dichaete</i>	850	0.004	0.008	-1.60	-3.23 [†]	(47/47)
<i>eve1</i>	741	0.002	0.002	-0.04	-1.16	(46/43)
<i>eve2</i>	485	0.003	0.002	-0.03	-0.03	(47/48)
<i>gt1</i>	772	0.002	0.002	0.78	0.33	(47/47)
<i>gt23</i>	1,076	0.003	0.004	-0.73	-0.62	(45/47)
<i>hairy0</i>	541	0.006	0.007	-0.49	0.69	(46/45)
<i>hairy2</i>	982	0.009	0.007	0.52	-0.51	(44/46)
<i>hairy7</i>	875	0.006	0.006	0.17	1.06	(45/44)
<i>hb P2</i>	280	0.001	0.003	-1.28	-1.48	(48/45)
<i>kni</i>	768	0.002	0.002	0.05	0.35	(42/44)
<i>otd early</i>	874	0.01	0.006	2.64 [*]	0.17	(46/46)
<i>salBE</i>	569	0.006	0.005	-0.04	-0.42	(48/46)
<i>slpA</i>	304	0.005	0.006	-0.44	-2.20	(48/48)
<i>slpB</i>	824	0.004	0.007	-1.00	-1.30	(44/43)
<i>tll</i>	946	0.01	0.009	0.38	-0.56	(47/45)
Total (all)	11,621					
Average (all)		0.0047	0.0046			

[†] $P < 0.02$; ^{*} $P < 0.05$.

^aBCRE identifiers reflect the names used in the primary report of each enhancer element. See Materials and Methods for references.

^bIndicates the total number of sequenced lines for each population (Maine/North Carolina) used in these analyses.

TABLE 3. Polymorphisms in Bicoid binding sites

BCRE	Bicoid binding site	Minor allele frequency
<i>eve1</i>	T(A/G)CAAATGCG	0.01
<i>eve1</i>	(C/T)CCTAATCCTT	0.01
<i>eve2</i> ^{a,b}	CTAATAATCT(G/C)	0.14
<i>eve2</i> ^a	CC(A/C)TAATCCCT	0.01
<i>gt23</i>	AGGTAAT(G/C)C(T/C)	0.09, 0.09
<i>hairy0</i>	T(T/C)CAAATCCAA	0.09
<i>hairy2</i>	C(C/T)GTAAGCCC	0.10
<i>hairy2</i>	GATTAAGC(C/G)G	0.10
<i>hairy2</i>	CG(--/ TA)ACCCT	0.08
<i>hairy7</i> ^a	TTT(A/T)(A/G)(A/T)CTT	0.11, 0.23, 0.26
<i>otd</i> early	TTTAACC(G/A)T	0.41
<i>slpA</i>	TCCTAA(T/A)CCC	0.01
<i>slpB</i>	GCTAGTCC(C/-)	0.08

SNPs are bracketed in bold.

^aPreviously defined and functionally characterized binding site: Ludwig and Kreitman ('95), La Rosée et al. ('97).

^bPreviously reported SNP: Small et al. ('92).

predicted based on sequence similarity to known Bcd binding sites (see Materials and Methods). Thirteen sites, three of which have been functionally characterized, were polymorphic in our sample (Table 3). Only one of the SNPs, the G/C polymorphism in the *eve* stripe 2 CRE, has been previously reported by Ludwig and Kreitman ('95). In general, polymorphisms tend to be situated outside the core 5'-TAATCC-3' domain where most deviations from the consensus sequence have been found to reside in naturally occurring binding sites (Lifanov et al., 2003). However, SNPs do occur in the core region: the site in the *hairy* stripe 7 enhancer is a good example of this. Whether these polymorphisms affect Bcd binding or target gene expression awaits further study.

Unusual sequence diversity in the orthodenticle BCRE

Theory predicts that the action of natural selection should leave a molecular signature in the region around selected loci that is different from the pattern of molecular variation seen under neutral conditions. To examine whether any of the BCREs have been subject to past natural selection, we used Tajima's *D* and Fu and Li's *D* statistics to test for a departure from neutrality (Haddrill et al., 2005; Begun et al., 2007). These tests indicate that the null hypothesis of neutrality cannot be rejected for most BCREs, with the exception of the *orthodenticle* (*otd*; also known as *ocelliless*, *oc*) early head

enhancer. The *otd* BCRE shows a positive and significant Tajima's *D* statistic, and is a clear outlier among the rest of the BCRE's (Table 2). Andolfatto (2005) demonstrated that Tajima's *D* values in *D. melanogaster* tend to be negatively skewed, making the positive value for the *otd* BCRE all the more interesting. In this case, a positive Tajima's *D* is the result of an excess of polymorphic alleles that segregate at intermediate frequency, as seen in a plot of the allele frequency spectrum (Fig. 1A). Furthermore, sliding window plots of π and Tajima's *D* reveal a high level of diversity within a 260 bp block in the center of the *otd* enhancer that corresponds to the region of elevated values of Tajima's *D* (Fig. 1B).

Unusually for *D. melanogaster*, this *otd* enhancer region also shows a large block of LD with nearly perfect association among alleles over the highly polymorphic 260 bp region (Fig. 1C). LD across the 16 other BCREs revealed no additional examples of such extensive haplotype block structure, including fragments with similar levels of polymorphism to the *otd* BCRE (data not shown), and this level of divergence exceeds that seen in our earlier demonstration of haplotype dimorphism in random genomic fragments of *D. melanogaster* (Teeter et al., 2000). The haplotype network diagram (Fig. 2A) shows two clear haplotype groups that are separated by a deep branch of 12 mutational steps. The haplotypes are further distinguished by several indel polymorphisms, and we use the presence of a major indel (16 bp) to define the main groups, denoted as D and I (see also Table 1). The polymorphisms that cause the strongly positive Tajima's *D* value, plus four indels, all appear in nearly perfect LD within the central stretch of 260 bp (Fig. 1C). There are additional mutational events distinguishing alleles within each haplotype group but these occur outside the 260 bp stretch. The more common haplotype (I) is present in approximately 60% of the samples, and allele frequencies do not differ substantially between two populations (Maine: 64% I, North Carolina: 54% I). We also sequenced a small sample of five African alleles, and observed both haplotypes (data not shown), implying that the dimorphism predates the bottleneck associated with the recent global spread of the species.

The same region where we find the strong haplotype structure appears to be part of a separate, complex insertion-deletion event leading to high divergence between *D. melanogaster* and its sibling species *D. simulans* and *D. sechellia*. The 163 base pair (bp) region between sites 363

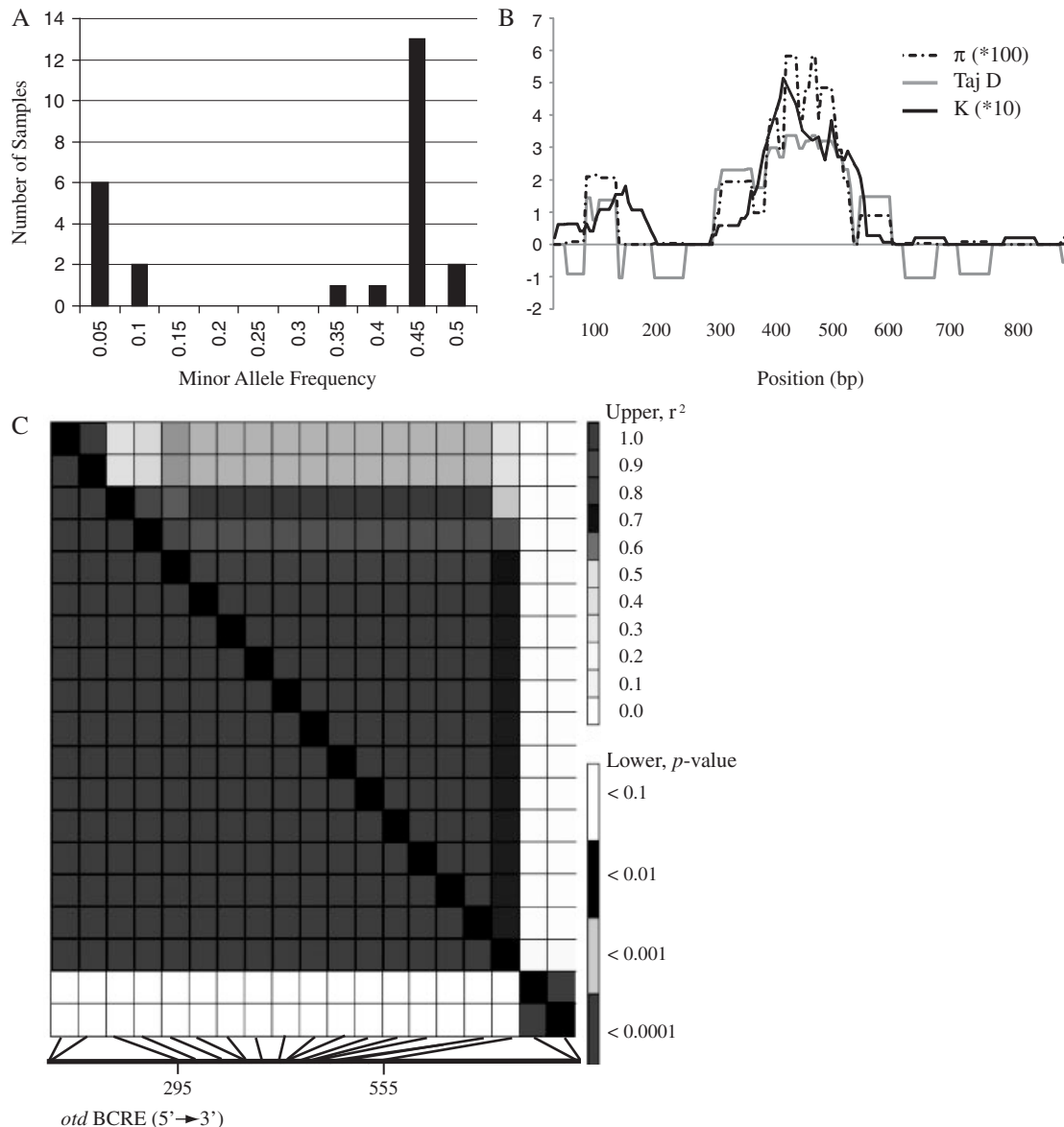


Fig. 1. Analysis of nucleotide diversity in the *otd* BCRE. (A) A plot of minor allele frequencies reveals an excess of common polymorphisms in the *otd* BCRE. (B) Sliding window plots of nucleotide diversity, Tajima's D , and divergence, indicate that most polymorphism within the *otd* BCRE is located in the central region of the enhancer, and corresponds to the region of the BCRE with elevated levels of Tajima's D and enhanced divergence between *D. melanogaster* and *D. simulans*. The plot was assembled in DnaSP using a window size of 50 and step size of 5. (C) The extent (r^2 , above the diagonal) and significance (P -value, below the diagonal) of linkage disequilibrium (LD) is plotted for each pairwise comparison of 19 polymorphic sites within the *otd* BCRE. The central 260 bp (sites 295–555) contains most of the variable sites, all of which are in nearly perfect LD.

and 516 in *D. melanogaster* is only 126 bp long in *D. simulans* and 119 bp in *D. sechellia*, of which, only short stretches can be aligned with the *D. melanogaster* sequence (Fig. 2B). Even without counting the indels, divergence between *D. melanogaster*, and the homologous region in *D. simulans*, *D. sechellia*, and *D. yakuba*, is increased (Fig. 2, and data not shown). In the 260 bp region, divergence between *D. melanogaster*

and *D. simulans* is about 14%, almost three times higher than the average for non coding regions reported by Begun et al. (2007).

The polymorphism and divergence pattern at the *otd* BCRE is thus characterized by three findings. First, there are two clear haplotypes with high divergence between them. Second, there is no polymorphism in the central 260 bp region within haplotype groups, even though both are at

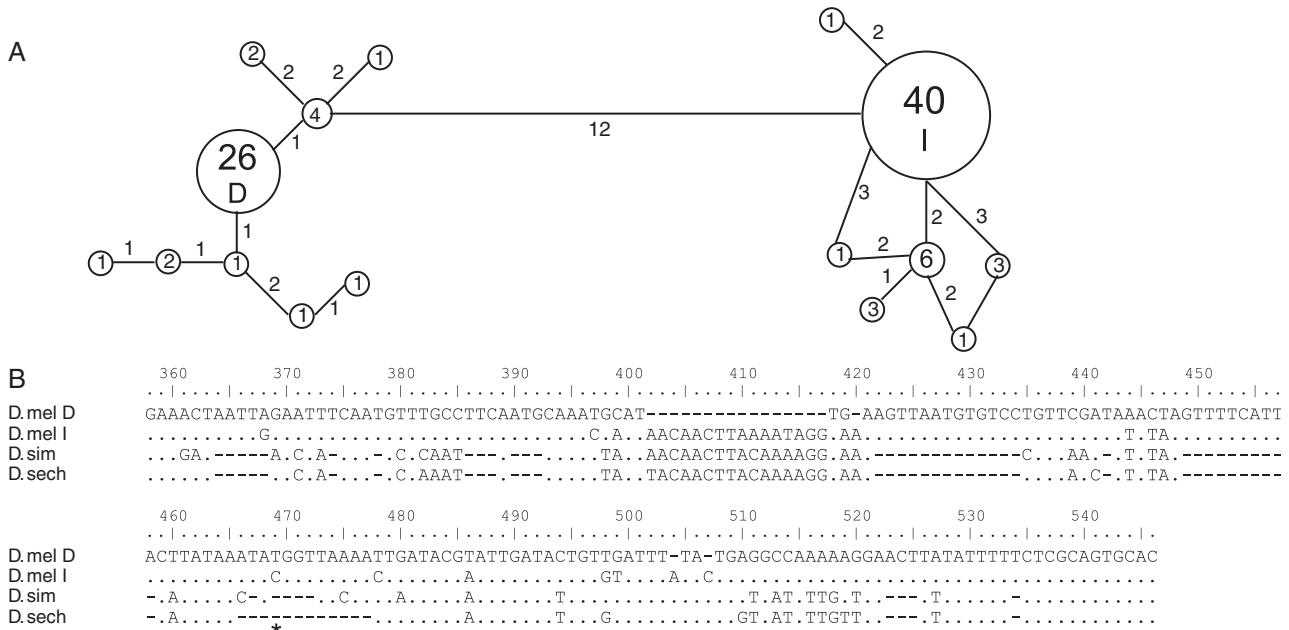


Fig. 2. Haplotype structure of the *otd* BCRE in *Drosophila*. **(A)** A haplotype network diagram of *otd* BCRE sequences from North American *D. melanogaster* reveals two distinct haplotype groups. All alleles can be unambiguously classified as harboring either the D (deletion) or I (insertion) haplotype. Numbers within circles refer to the total number of alleles at each node; numbers in italics refer to the number of mutational steps between each node. **(B)** A complex indel polymorphism distinguishes the *D. melanogaster* *otd* BCRE from its sibling species, *Drosophila simulans* and *Drosophila sechellia*. Sequence from the central region of the haplotype block is shown for representative D and I alleles of *D. melanogaster*, as well as homologous sequence from *D. simulans* and *D. sechellia*. The asterisk indicates the Bicoid binding site polymorphism in *D. melanogaster* haplotypes.

an intermediate frequency and there is a large sample size. And third, there is a high level of divergence relative to the sibling species *D. simulans* and *D. sechellia* in the same region. Each of these findings can result either from neutral processes such as enhanced mutation or a bottleneck, or from selective events. Haplotype structure with high divergence between the haplotypes could be caused by a population bottleneck, admixture, balancing selection or a soft selective sweep. Lack of polymorphism could be owing to a recent bottleneck or positive or negative selection. High divergence to the out-group species could be caused by an enhanced mutation rate or recurrent positive selection. Although the data do not allow a definitive conclusion as to which of these scenarios is most likely, they are suggestive of a complex history of both selective and demographic influences.

Association of the *otd* BCRE with expression variation

The *otd/oc* gene has pleiotropic roles in embryonic head, central nervous system, and adult eye development (Finkelstein et al., '90), and selection could affect any or all of these processes. To see

whether it could feasibly affect embryonic patterning, we examined the expression of *otd* at the cellular blastoderm stage of 99 wild-type inbred lines by in situ hybridization (Gao and Finkelstein, '98), 38 with the D haplotype and 61 with the I haplotype. Visual examination of staining in embryos suggested a tendency toward differential intensity of staining in the ventral portion of the expression domain as epitomized by the two embryos in Figure 3A. Specifically, embryos with the D haplotype appeared to have reduced staining in the ventral region, whereas I haplotype embryos showed a strong band of expression all the way to the ventral margin.

To quantify this difference, embryos were stained and photographed, and NIH ImageJ software was used to measure pixel intensity along a line drawn from the dorsal to the ventral edge of each embryo, near the anterior border of the expression domain (Fig. 3A). Plots of pixel intensity vs. percent egg height (dorsal = 0%, ventral = 100%) were generated and the inflection point was determined as the percent egg height at which at least four points in a row increased in pixel intensity above background levels (Fig. 3B). The distribution of inflection points for wild type animals is shown above the x-axis in Figure 3C.

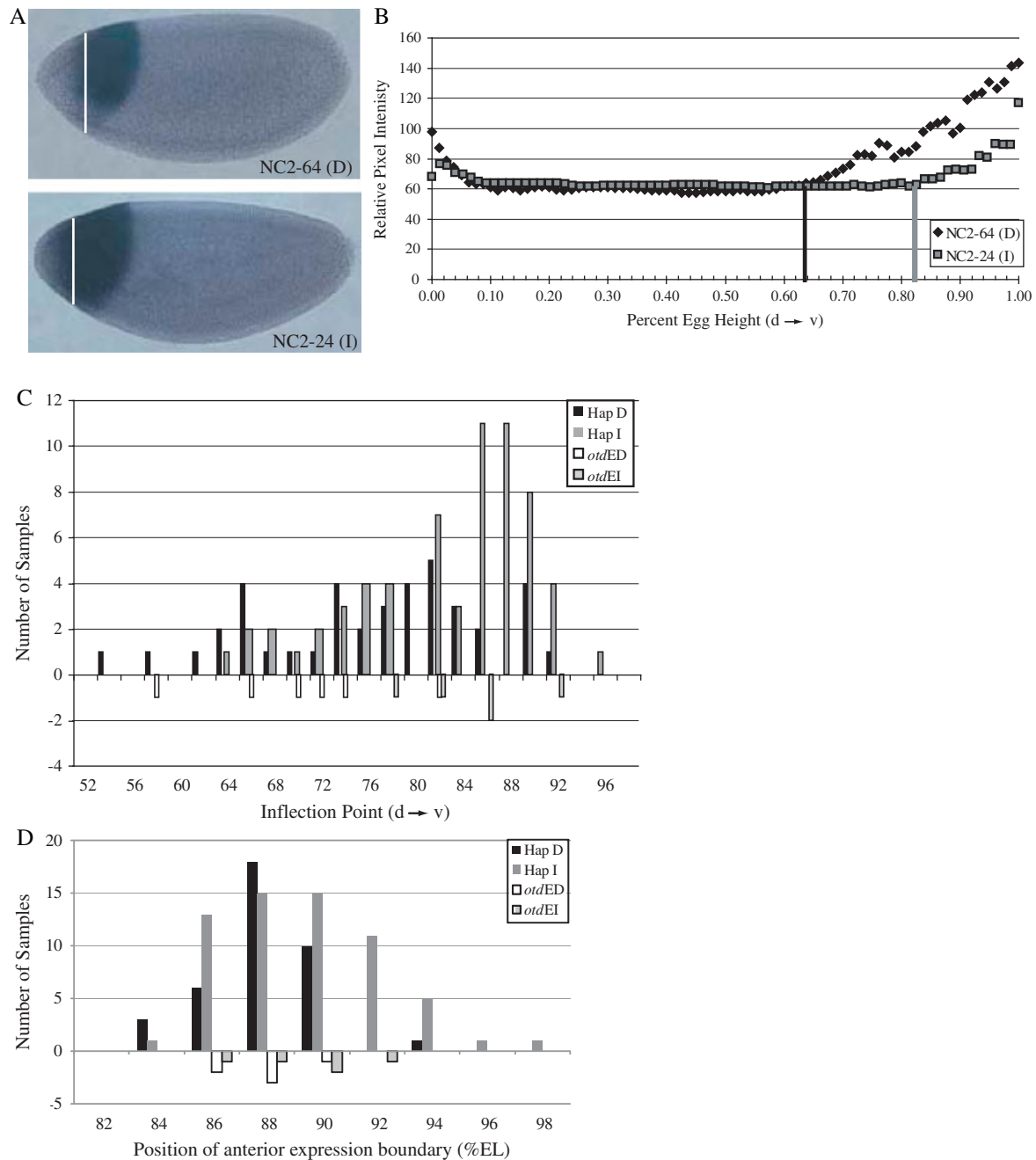


Fig. 3. Variation in *orthodenticle* expression between the two *otd* BCRE haplotypes. **(A)** *otd* expression in representative embryos harboring the D (line NC2-64) or I (line NC2-24) haplotype. Anterior is to the left and dorsal is up. Note the decrease in *otd* expression in the ventral and anterior regions of the D haplotype embryo. To quantify the difference in the ventral domain, pixel intensity was calculated along a line running from dorsal (d) to ventral (v) along the anterior border of the expression domain (white line in [A]). **(B)** A plot of relative pixel intensity vs. position along the dorsal-ventral (d-v) axis was generated to establish the inflection point (solid bar in B, see text for more details). **(C)** The distribution of inflection points for embryos from 99 wild type inbred lines (Hap D and Hap I, above the abscissa), and for 11 transgenic lines (*otdED* and *otdEI*, below the abscissa) indicates that the D haplotype is associated with a loss of expression ventrally, as seen by the shift in the distribution of inflection points towards the left. Transgenic embryos were stained for expression of the reporter gene, *lacZ*. **(D)** The distribution of the position of anterior expression boundaries for 99 wild type inbred lines (Hap D and Hap I, above the abscissa), and for 11 transgenic lines (*otdED* and *otdEI*, below the abscissa). Analysis indicates that there is a significant anterior shift of the anterior expression boundary in wild type I haplotype lines and a similar trend in transgenic embryos although the result is not significant. See Materials and Methods for details regarding anterior expression boundary measurements.

A Mann–Whitney U test indicates that the distribution of inflection point location along the dorsal–ventral axis was significantly different among embryos with different *otd* haplotypes (two-tailed t -test, $P = 0.0007$). Note that this procedure is independent of differences in staining intensity among embryos and that biases introduced by imperfect orientation of the embryos on the microscope slide would be randomized over the two classes and tend to reduce the true signal of differential ventral expression.

To further explore the expression differences between D and I haplotypes we used a novel mode of digital analysis, popGEMs. Each individual embryo image was standardized by forcing the embryo into a 320×128 pixel rectangle, as this aspect ratio corresponds to the average seen for thousands of embryos in the FlyExpress database (Gurunathan et al., 2004). A single best expression signature converts each image into a black/white BEE in which each pixel indicates either presence or absence of transcript spatial expression. Stacking of images from embryos with expression of the same gene from different lines results in popGEMs, wherein each pixel is now associated with a probability that the gene is expressed in that portion of the embryo in the sample. The *otd* popGEMs for the D and I haplotypes are shown in Figures 4A and 4B, and subtraction of D from I results in the difference plot in Figure 4C. The shades of red (negative values) indicate locations where *otd* is more commonly expressed in D than I lines, whereas green shades (positive values) indicate where expression is more common in I lines. If D and I are identical then this Δ popGEM should show a random scattering of red and green colors against a white background.

This is not the case, as the two haplotypes show substantial differential expression of two types. The extensive green band at the anterior tip of the embryo in Figure 4C indicates that I haplotypes tend to have expanded anterior expression, and this can be seen clearly in the representative embryos in Figure 3 as well as contrasting Figures 4A and 4B. We also generated a histogram showing the area, in pixels, of expression in each embryo of the two *otd* classes, and this too indicates a significant anterior expansion of the area of *otd* expression in I haplotype embryos (data not shown). We confirmed this difference using NIH ImageJ software to determine the position of the anterior expression boundary (see Materials and Methods). The distribution of anterior expression boundary position for wild

type animals is shown above the x -axis in Figure 3D. A Mann–Whitney U test indicates that the distribution of the position of the anterior boundary along the A–P axis was significantly different among embryos with different *otd* haplotypes (two-tailed t -test, $p = 0.04$); specifically, the boundary shows an anterior shift in I haplotype embryos.

By contrast, the red bands in Figure 4C are more difficult to explain as they suggest extensive dorsal and ventral expression in the D lines, which is not obvious from the popGEM images and is not what we would predict based on the results in Figure 3C. popGEMs give a visual representation of the probability that any embryo within the sample will exhibit expression at a particular location. Thus, the expression of *otd* at the ventral margin in the D haplotype popGEM in Figure 4A indicates that some embryos in the sample do have staining in this region; however, this spatial analysis method is not capturing quantitative differences in expression. The red bands in Figure 4C are more likely explained by an overall elevation of expression in D lines throughout the domain of expression common to both the D and I haplotypes. A substantial portion of the ventral domain is green, indicating more frequent ventral expression in I haplotype lines, consistent with the results in Figure 3C.

Differential reporter gene expression driven by the two enhancer haplotypes

To confirm that polymorphism in the BCRE is responsible for the differential expression of *otd*, transgenic constructs where either *otd* enhancer haplotype drives expression of the *lacZ* reporter gene were generated by P-element mediated transformation. Six independent insertions of the deletion construct, and five of the insertion construct were examined. The ventral expansion of expression by the I haplotype was confirmed, showing a considerably enhanced differentiation compared with the wild type alleles. The bars below the x -axis in Figure 3C indicate that the inflection point for ventral expression is between 78 and 92% egg height for the I lines, and between 58 and 74% egg height for 5 of the 6 D lines (two-sided Mann–Whitney U test, $P = 0.01$). Although there appears to be a slight anterior shift of the anterior expression boundary in transgenic I haplotype lines (Fig. 3D), the trend was not significant (two-sided Mann–Whitney U test, $P = 0.5$) suggesting either that the position of the

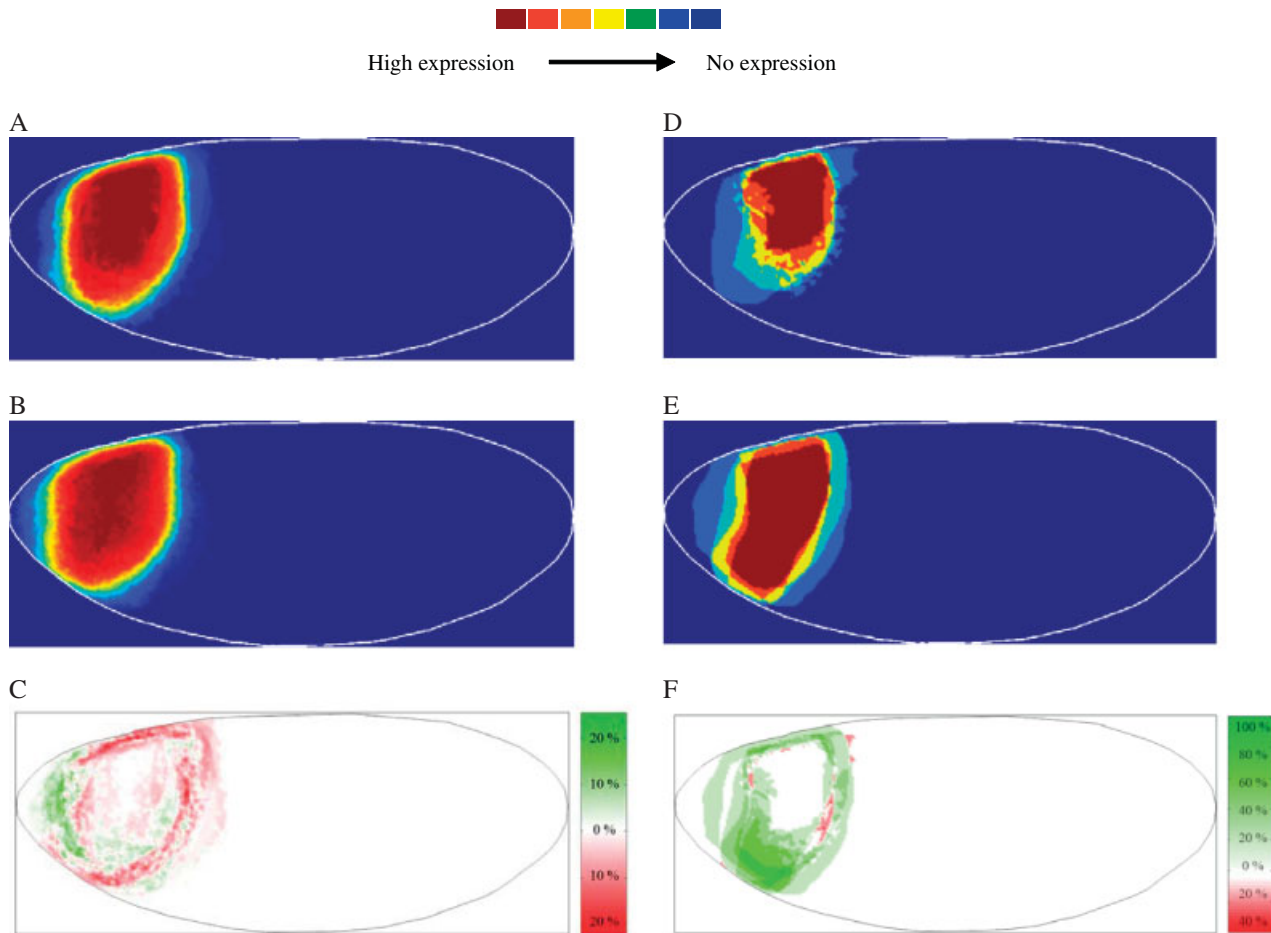


Fig. 4. Spatial analysis of *otd* expression. (A–B, D–E) Population Gene Expression Maps (popGEMs) for wild type embryos harboring the D and I haplotypes (A and B respectively) and transgenic embryos expressing *lacZ* under control of the D and I *otd* enhancer haplotypes (D and E respectively). Anterior is to the left and dorsal is up. Embryos were stained for *otd* (wild type) or *lacZ* (transgenic) by in situ hybridization, individual images were standardized into a 300×128 pixel rectangle, and a black/white binary expression extraction (BEE) was generated for each image. Each pixel in a BEE indicates the presence or absence of transcript expression. Images for all four sample groups were stacked to generate the popGEMs, where the depth of color represents the fraction of individuals with expression in that particular region of the embryo. See text for more details of analysis. (C, F) Hurricane-like difference plots were generated by subtracting the popGEM for Hap D embryos from Hap I embryos (C), and the popGEM for *otd*ED embryos from *otd*EI embryos (F). Shades of green (positive values) indicate regions where there is a higher probability of expression in embryos containing the I haplotype, whereas shades of red indicate regions where there is a higher probability of expression in D haplotype embryos. This analysis indicates that the I haplotype is associated with a higher probability of anterior expression, and confirms the decrease in ventral expression seen in *otd*ED embryos.

anterior boundary of *otd* expression is influenced by more than just this enhancer element or that this effect is subject to chromosomal position effects at the transgene integration site. We also suspect that variation in egg shape is not fully accounted for by the image transformation, and this may reduce the effectiveness of the popGEM method, which is most useful as a screening tool. The popGEMs and difference plot in Figures 4D–F nevertheless confirm the ventral expansion, despite the much reduced sample size relative to the wild-type comparisons.

One or more of the dozen sequence differences between the two haplotypes may be responsible for the expression differences. The Bcd binding site polymorphism (Table 3) is a clear candidate for mediating the anterior expansion in the I lines, but there are no obvious polymorphic binding sites for D–V transcription factors such as Snail that might explain the loss of ventral expression in the D lines. Attempts to demonstrate differential binding of Bcd protein to the two sequences were not successful, likely because Bcd is expected to bind the putative site with low affinity. More

comprehensive dissection of the element is needed to establish the molecular basis of the various aspects of differential expression.

DISCUSSION

An outstanding question in the field of evolutionary developmental biology regards the contribution of *cis*-regulatory mutations to phenotypic evolution (Hoekstra and Coyne, 2007; Wray, 2007). It has been proposed that certain types of phenotypic changes, notably those involving morphology, may be more likely to occur owing to adaptive mutation in *cis*-regulatory regions, rather than in the protein-coding regions of genes (Jacob and Monod, '61; Carroll, 2000; Stern, 2000; Gompel et al., 2005). This argument largely rests on the hypothesis that the modular nature of *cis*-regulatory regions reduces the likelihood of pleiotropic mutations by limiting the effect of a mutation to a small part of the overall transcriptional profile. In contrast, mutations in protein coding regions are expected to affect the function of the gene product in all expression domains, increasing the likelihood of a deleterious effect.

Until recently, a detailed analysis of the contribution of *cis*-regulatory mutations to phenotypic change has been difficult to come by. Unlike protein-coding regions of genes, regulatory sequences are notoriously difficult to identify, and predictions regarding the effects of mutations are hard to make. However, over the last several years the large volume of genomic data available to researchers, combined with improvements in technology, has made identification and analysis of *cis*-regulatory regions much more tractable. Several instances of polymorphisms in *cis*-regulatory elements have been shown to drive differential expression, and to be associated with complex disease susceptibility, and population genetic analyses have shown how a combination of balancing and diversifying selection as well as mutation pressures influence the distribution of allele frequencies in different human populations (Hahn et al., 2004; Rockman et al., 2003, 2004, 2005). At least three clear instances of multiple mutations affecting *cis*-regulatory modules responsible for divergence in expression between *Drosophila* species have been described (Ludwig and Kreitman, '95; Sucena et al., 2003; Jeong et al., 2008). To our knowledge, this study provides the first demonstration of intraspecific *cis*-regulatory polymorphisms affecting the

expression of a gene that is crucial for early embryonic patterning.

It is not clear what population genetic processes might explain the unusual nature of the sequence diversity in the *otd* enhancer. To explain all three observations (two divergent haplotypes, no polymorphism in the central 260 bp region of each haplotype, high divergence relative to the sibling species), at least two forces must be invoked. A purely neutral scenario would be a locally enhanced mutation rate combined with a strong recent bottleneck. The complicated history of North American *D. melanogaster* (which has certainly experienced bottlenecks and possibly admixture: Haddrill et al., 2005), accommodates a wide range of neutral processes. The difficulty here is that we need to assume that two entirely independent extreme processes have affected the same short sequence region, namely a large local mutation rate and an extreme shape of the local genealogy owing to a bottleneck.

Balancing selection maintains multiple alleles at intermediate frequency within a population by heterozygote advantage, frequency dependent selection, or selection on alternate alleles under different environmental conditions (environmental heterogeneity). Loci under old balancing selection will maintain common polymorphism in the region surrounding the selected polymorphism, and hence show an increased level of diversity that is organized in haplotype groups. However, in this case, we would expect neutral polymorphism to segregate within each of these haplotype groups. Surprisingly, the highly polymorphic 260 bp block shows no variation within either of the haplotype groups. To make the old balancing selection scenario compatible with the low polymorphism within the haplotypes, one needs to postulate additional recent positive selection events within both haplotype groups. Also, balancing selection does not explain the high divergence with *D. simulans*, for which we would still need to invoke an enhanced mutation rate or recurrent positive selection in the *D. melanogaster* or *D. simulans* lineage.

Another possibility is that a particular class of selective sweep, namely a soft selective sweep from recurrent beneficial mutation, has allowed two or more beneficial alleles at the same locus to quickly increase in frequency in the population as recently suggested at the *tan* locus (Jeong et al., 2008). This would be expected to create a strong haplotype structure, with very low polymorphism within the two haplotypes, and it is likely to result in a high

Tajima's D (Pennings and Hermisson, 2006a). However, unlike balancing selection, a soft sweep is not expected to lead to high divergence between the two haplotypes. This aspect of the data, as well as the high divergence to the outgroups, suggests a locally increased mutation rate, but as with the neutral model, two unusually strong phenomena must be invoked. Interestingly, a high mutation rate in itself increases the probability of soft sweeps (Pennings and Hermisson, 2006b), but a further conundrum is that the two alleles would be expected to have nearly identical fitness effects. This is not obviously reconciled with the clear functional difference between the haplotypes at the expression level, though it is possible that selection acts on a pleiotropic aspect of the enhancer function and that the embryonic expression phenotype is effectively neutral.

It is however, difficult to imagine that the expression differences described here would be without subtle consequences. Fowlkes et al. (2008) recently showed that precise standardization of embryonic shapes and staging relative to a control gene can be used to decipher spatial interactions that are important for the network of gene function in embryogenesis. Our results suggest that similar approaches applied to a diversity of wild type lines may provide insight into the variability of embryonic patterning even within a species. A fascinating outcome of such studies will be better understanding of the buffering mechanisms that ensure that development is robust, and of the canalization mechanisms that result in species-specific patterns of gene expression (Lott et al., 2007).

ACKNOWLEDGMENTS

The authors thank Erin Kennerly for help collecting flies; Martin Kreitman, Corbin Jones and Richard Lyman for generously providing fly lines; and Ian Dworkin for help with DNA collection and sequencing. We thank Ian Dworkin and Stefan Laurent for helpful discussions regarding the data. J. H. is supported by the Vienna Science and Technology Fund (WWTF), S. K. by the NHGRI, and G. G. by the US NIH (2R01-GM61600) and the Australian Research Council (ARC-DP0880204).

LITERATURE CITED

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–1152.

- Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:519–520.
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5:e310.
- Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101:577–580.
- Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659.
- Damen WG. 2007. Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev Dyn* 236:1379–1391.
- Driever W, Nüsslein-Volhard C. 1988a. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* 54:95–104.
- Driever W, Nüsslein-Volhard C. 1988b. A gradient of bicoid protein in *Drosophila* embryos. *Cell* 54:83–93.
- Driever W, Thoma G, Nüsslein-Volhard C. 1989. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* 340:363–367.
- Ephrussi A, St Johnston D. 2004. Seeing is believing: the bicoid morphogen gradient matures. *Cell* 116:143–152.
- Finkelstein R, Smouse D, Capaci TM, Spradling AC, Perrimon N. 1990. The orthodenticle gene encodes a novel homeo domain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. *Genes Dev* 4:1516–1527.
- Fowlkes CC, Hendriks CL, Keranen SV, Weber GH, Rubel O, et al. 2008. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* 133:364–374.
- Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB. 1999. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* 126:2527–2538.
- Gao Q, Finkelstein R. 1998. Targeting gene expression to the head: the *Drosophila* orthodenticle gene is a direct target of the Bicoid morphogen. *Development* 125:4185–4193.
- 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.
- Gurunathan R, Van Emden B, Panchanathan S, Kumar S. 2004. Identifying spatially similar gene expression patterns in early stage fruit fly embryo images: binary feature versus invariant moment digital representations. *BMC Bioinformatics* 5:202.
- Hadrill PRTK, Charlesworth B, Andolfatto P. 2005. Multi-locus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 16:790–799.
- Hahn MW, Rockman MV, Soranzo N, Goldstein DB, Wray GA. 2004. Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans. *Genetics* 167:867–877.
- Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, et al. 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* 26:362–367.
- Hoch M, Schroder C, Seifert E, Jäckle H. 1990. cis-acting control elements for Krüppel expression in the *Drosophila* embryo. *EMBO J* 9:2587–2595.

- Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution Int J Org Evolution* 61:995–1016.
- Houchmandzadeh B, Wieschaus E, Leibler S. 2002. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* 415:798–802.
- Howard KR, Struhl G. 1990. Decoding positional information: regulation of the pair-rule gene hairy. *Development* 110:1223–1231.
- Hutter SLH, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome-wide single nucleotide polymorphism data. *Genetics* 177:469–480.
- Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318–356.
- Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, et al. 2008. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* 132:783–793.
- Kim J, Kerr JQ, Min GS. 2000. Molecular heterochrony in the early development of *Drosophila*. *Proc Natl Acad Sci USA* 97:212–216.
- Kuhnlein RP, Bronner G, Taubert H, Schuh R. 1997. Regulation of *Drosophila* spalt gene expression. *Mech Dev* 66:107–118.
- Kumar S, Jayaraman K, Panchanathan S, Gurunathan R, Marti-Subirana A, et al. 2002. BEST: a novel computational approach for comparing gene expression patterns from early stages of *Drosophila melanogaster* development. *Genetics* 162:2037–2047.
- La Rosee A, Hader T, Taubert H, Rivera-Pomar R, Jäckle H. 1997. Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the *Drosophila* embryo. *EMBO J* 16:4403–4411.
- Lifanov AP, Makeev VJ, Nazina AG, Papatsenko DA. 2003. Homotypic regulatory clusters in *Drosophila*. *Genome Res* 13:579–588.
- Lott SE, Kreitman M, Palsson A, Alekseeva E, Ludwig MZ. 2007. Canalization of segmentation and its evolution in *Drosophila*. *Proc Natl Acad Sci USA* 104:10926–10931.
- Ludwig MZ, Kreitman M. 1995. Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*. *Mol Biol Evol* 12:1002–1011.
- Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, et al. 2005. Functional evolution of a cis-regulatory module. *PLoS Biol* 3:e93.
- Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C. 2006. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature* 439:728–732.
- Manu, Surkova S, Spirov AV, Gursky V, Janssens H, et al. 2009. The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proc Natl Acad Sci USA* 102:4960–4965.
- Moriyama EN, Powell JR. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol* 13:261–277.
- Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, et al. 2005. The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proc Natl Acad Sci USA* 102:4960–4965.
- Peel AD, Chipman AD, Akam M. 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet* 6:905–916.
- Pennings PS, Hermisson J. 2006a. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet* 2:e186.
- Pennings PS, Hermisson J. 2006b. Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol* 23:1076–1084.
- Rivera-Pomar R, Lu X, Perrimon N, Taubert H, Jäckle H. 1995. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* 376:253–256.
- Rockman MV, Hahn MW, Soranzo N, Goldstein DB, Wray GA. 2003. Positive selection on a human-specific transcription factor binding site regulating IL4 expression. *Curr Biol* 13:2118–2123.
- Rockman MV, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, et al. 2004. Positive selection on MMP3 regulation has shaped heart disease risk. *Curr Biol* 14:1531–1539.
- Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, et al. 2005. Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol* 3:e387.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Schroder R. 2003. The genes orthodenticle and hunchback substitute for bicoid in the beetle *Tribolium*. *Nature* 422:621–625.
- Singh ND, Davis JC, Petrov DA. 2005. Codon bias and noncoding GC content correlate negatively with recombination rate on the *Drosophila* X chromosome. *J Mol Evol* 61:315–324.
- Small S, Blair A, Levine M. 1992. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J* 11:4047–4057.
- Spirov A, Fahmy K, Schneider M, Frei E, Noll M, Baumgartner S. 2009. Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. *Development* 136:605–614.
- Stern DL. 2000. Evolutionary biology. The problem of variation. *Nature* 408:529–531.
- Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* 424:935–938.
- Tautz D, Pfeifle C. 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* 98:81–85.
- Teeter K, Naemuddin M, Gasperini R, Zimmerman E, White KP, et al. 2000. Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *J Exp Zool* 288:63–75.
- Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–633.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206–216.