Genetic Basis of a Violation of Dollo's Law: Re-Evolution of Rotating Sex Combs in *Drosophila bipectinata*

Thaddeus D. Seher, Chen Siang Ng,[1] Sarah A. Signor, Ondrej Podlaha,[2] Olga Barmina, and Artyom Kopp[3]
Department of Evolution and Ecology, University of California, Davis, California 95616

ABSTRACT Phylogenetic analyses suggest that violations of "Dollo's law"—that is, re-evolution of lost complex structures—do occur, albeit infrequently. However, the genetic basis of such reversals has not been examined. Here, we address this question using the *Drosophila* sex comb, a recently evolved, male-specific morphological structure composed of modified bristles. In some species, sex comb development involves only the modification of individual bristles, while other species have more complex "rotated" sex combs that are shaped by coordinated migration of epithelial tissues. Rotated sex combs were lost in the *ananassae* species subgroup and subsequently re-evolved, ~12 million years later, in *Drosophila bipectinata* and its sibling species. We examine the genetic basis of the differences in sex comb morphology between *D. bipectinata* and *D. malerkotliana*, a closely related species with a much simpler sex comb representing the ancestral condition. QTL mapping reveals that >50% of this difference is controlled by one chromosomal inversion that covers ~5% of the genome. Several other, larger inversions do not contribute appreciably to the phenotype. This genetic architecture suggests that rotating sex combs may have re-evolved through changes in relatively few genes. We discuss potential developmental mechanisms that may allow lost complex structures to be regained.

DOLLO'S "law of irreversibility" posits that complex morphological structures, once lost during evolution, cannot be regained in the same form. This principle makes intuitive sense: resurrecting an extinct developmental pathway in close to its ancestral condition seems biologically as well as statistically implausible. And yet, phylogenetic analyses have revealed several cases where Dollo's law is apparently violated. Examples include re-evolution of lost digits in lizards (Kohlsdorf and Wagner 2006; Brandley *et al.* 2008; Kohlsdorf *et al.* 2010), eggshells and oviparity in boas (Lynch and Wagner 2010), wings in stick insects (Whiting *et al.* 2003), shell coiling in limpets (Collin and Cipriani 2003), mandibular teeth in frogs (Wiens 2011), molars in lynx (Kurten 1963), and others. There is a growing consensus that lost structures can sometimes be regained, especially if that happens soon after the initial loss (Wiens 2011) (but see Goldberg and Igic 2008 and Galis *et al.* 2010 for counterarguments). This shifts the question from the realm of phylogenies to developmental genetics: how can complex structures re-evolve? What is the genetic basis of such reversals?

In this report, we examine the genetic basis of a likely violation of Dollo's law that occurred during the evolution of *Drosophila* sex combs. Sex combs are male-specific arrays of modified mechanosensory bristles ("teeth") that evolved within the genus *Drosophila* and are used by males during courtship and mating (Kopp 2011). These structures develop from either transverse or longitudinal bristle rows that are present on the front legs of both sexes and show extensive morphological diversity but essentially fall into three distinct types. "Rotating" sex combs develop from one or several transverse bristle rows (TBRs) that undergo a 90° rotation. This rotation is driven by a precisely coordinated rearrangement of several hundred epithelial cells that, assisted by strong homophilic adhesion between adjacent bristle cells, moves the embedded bristle rows from a transverse to

doi: 10.1534/genetics.112.145524
Manuscript received August 30, 2012; accepted for publication September 28, 2012
Supporting information is available online at http://www.genetics.org/content/suppl/2012/10/11/genetics.112.145524.DC1.
Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession NM_079524.
[1]Present address: Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan.
[2]Present address: Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115.
[3]Corresponding author: Department of Evolution and Ecology, University of California, Davis, CA 95616. E-mail: akopp@ucdavis.edu

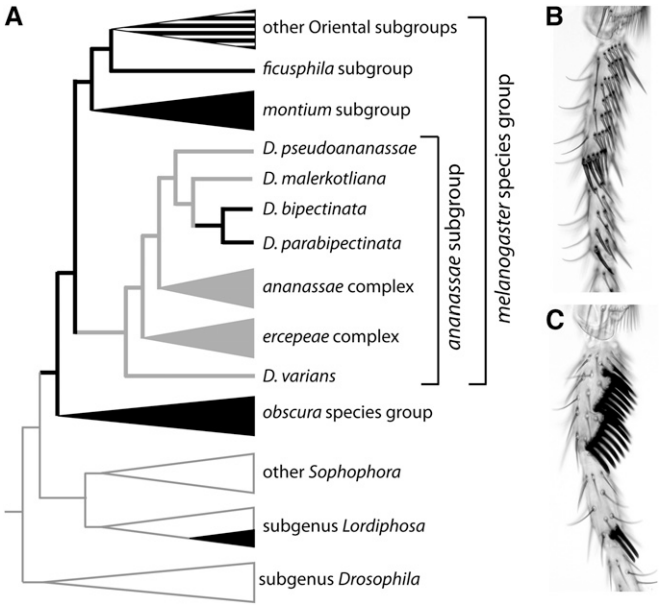

Figure 1 Re-evolution of rotating sex combs in *D. bipectinata*. (A) A simplified phylogenetic tree showing the position of *D. bipectinata* and its relatives. Black indicates a rotated (longitudinal) sex comb, gray a transverse sex comb, and white a primitively absent sex comb. The striped triangle shows a lineage where species have either rotated or transverse sex comb. In the *ficusphila* and *montium* subgroups, sex combs develop from longitudinal bristle rows and do not undergo active rotations. The *obscura* species group, the Oriental subgroups, and *D. bipectinata* and *D. parabipectinata* have actively rotating sex combs (Kopp 2011). Some species in the subgenus *Lordiphosa* also have actively rotating sex combs (Atallah *et al.* 2012). The most likely evolutionary scenario is that actively rotating sex combs were present in the last common ancestor of the *melanogaster* and *obscura* species group, were lost at the base of the *ananassae* subgroup, and re-evolved in the last common ancestor of *D. bipectinata* and *D. parabipectinata* (Matsuda *et al.* 2009). (B) The sex comb of *D. malerkotliana*. (C) The sex comb of *D. bipectinata*.

a more longitudinal orientation (Atallah *et al.* 2009a,b; Tanaka *et al.* 2009). In contrast, "transverse" sex combs are simply TBRs composed of modified bristles. In this case, sex comb development is limited to the modification of individual bristle shafts and does not involve any morphogenetic movements (Kopp 2011). Finally, "longitudinal" sex combs resemble rotating sex combs in adult flies but actually develop from longitudinal bristle rows and are not homologous to the rotating sex combs on a cell-by-cell basis (Atallah *et al.* 2009b; Tanaka *et al.* 2009). Sex comb evolution presents many examples of divergence and convergence, and each developmental mechanism has evolved more than once (Atallah *et al.* 2009b, 2012; Tanaka *et al.* 2009) (Figure 1A).

An apparent violation of Dollo's law is observed in the *ananassae* subgroup of the *melanogaster* species group. This lineage consists of over 20 species, most of which have simple transverse sex combs (Matsuda *et al.* 2009) (Figure 1, A and B). The only exception is found in one pair of sibling species, *D. bipectinata* and *D. parabipectinata*, which have much more dramatic sex combs that develop by active rotation (Figure 1, A and C). Detailed analysis of cell behavior during the pupal stage shows that this rotation occurs by the same mechanism as in distantly related groups such as the *obscura* species group and *D. melanogaster* and its relatives (Atallah *et al.* 2009b; Tanaka *et al.* 2009). *D. bipectinata* and *D. parabipectinata* branch deeply within the *ananassae* subgroup, while all basal lineages have transverse sex combs. Phylogenetic analysis strongly suggests that the common ancestor of these species has re-evolved a rotating sex comb following a previous loss at the base of the *ananassae* subgroup (Kopp and Barmina 2005; Barmina and Kopp 2007; Matsuda *et al.* 2009) (Figure 1A).

D. bipectinata can be hybridized with its close relative *D. malerkotliana* (Bock 1978; Kopp and Barmina 2005), which has simple transverse sex combs, opening the way for a direct genetic analysis of sex comb re-evolution. We used a QTL mapping approach to identify the genomic regions responsible for the differences in sex comb morphology between *D. bipectinata* and *D. malerkotliana*. Surprisingly, we find that much of the species difference maps to a single chromosomal inversion that covers ~5% of the genome.

Materials and Methods

Drosophila strains, crosses, and phenotypic analysis

Drosophila strains were obtained from the *Drosophila* Species Stock Center or provided by Y. Fuyama and M. Matsuda and maintained on standard cornmeal media. Polytene chromosome spreads were prepared from salivary glands of female larvae using acetic orcein staining. Strains *D. malerkotliana* 14024-0391.00 and *D. bipectinata* 14024-0381.03 were inbred by single-pair, full-sib crosses for 12 and 18 generations, respectively, to generate the derivative strains *mal0-sc2* and *bip3-isoA*. Interspecific hybridizations and subsequent crosses were performed in mass cultures with at least 20 males and 20 females per generation. For phenotypic analysis, male legs were removed just above the tibia–tarsus joint, mounted in Hoyers media between two 60- × 22-mm coverslips, and examined under a high-power compound microscope with brightfield illumination. Right and left legs from each male were kept together on an individual slide, and the average number of sex comb teeth per leg was recorded.

Transcriptome sequencing and identification of fixed differences between parental strains

To identify single nucleotide polymorphisms that distinguish *mal0-sc2* and *bip3-isoA*, normalized cDNA libraries were synthesized from whole-body adult RNA samples extracted from each strain. Paired-end libraries were prepared from the sheared, normalized cDNA using the standard Illumina protocol and sequenced with 85-base paired-end reads on an Illumina Genome Analyzer II at the University of California Davis Genome Center (Supporting Information, Table S1). The partial transcriptomes of *mal0-sc2* and *bip3-isoA* were assembled *de novo* using ABySS (Birol *et al.* 2009; Simpson

et al. 2009; Miller *et al.* 2010) followed by Trans-ABySS (Robertson *et al.* 2010) and CAP3 (Huang and Madan 1999) (Table S2; File S1a and File S1b; and File S2a and File S2b). For genetic analysis, we required fixed differences (FD) between parental strains, *i.e.*, nucleotide positions where *bip3-isoA* is fixed for one allele and *mal0-sc2* for a different allele. To identify FDs for the first round of genotyping, *mal0-sc2* and *bip3-isoA* read libraries were each mapped separately to both *bip3-isoA* and *mal0-sc2 de novo* transcriptome assemblies using SOAP2 (Li *et al.* 2009). Same-species mapping (*mal0-sc2* reads to *mal0-sc2* transcriptome and *bip3-isoA* reads to *bip3-isoA* transcriptome) serves to correct assembly errors, gauge coverage at each position, and identify nucleotide positions where multiple alleles segregate within each parental strain despite inbreeding. Cross-species mapping allows us to identify polymorphic positions and select FDs that meet coverage cutoffs. Transcriptome alignments yielded >40,000 FDs between *bip3-isoA* and *mal0-sc2* that are located in pairs of orthologous transcripts (File S3). For the second round of genotyping, *bip3-isoA* and *mal0-sc2* cDNA reads were mapped to the *D. bipectinata* reference genome (GenBank AFFE00000000.1) (File S4). To identify the genomic locations of FDs, we BLASTed the sequence flanking each FD against the *D. ananassae* FlyBase 1.3 (July 2011) reference genome and transcriptome. The transcriptome assemblies and SNP data sets for *D. malerkotliana*, *D. bipectinata*, and several related species are described in detail elsewhere (Signor *et al.* 2013).

Genotyping and genetic mapping

To construct linkage maps and identify QTL intervals, we genotyped the progeny of two separate F_2 backcrosses and an F_{38} introgression line (see *Results*). For the first round of genotyping, we selected 32 FDs that were evenly distributed among the major chromosome arms (Muller elements A–E) and were located at least 2.4 Mb from each other in the *D. ananassae* genome (Table S3). Because the initial analysis suggested the presence of one or more strong QTL on Muller E or proximal Muller D, we performed a second round of genotyping with 32 additional FDs concentrated on these chromosome arms. Genotyping primers were designed using Typer (Sequenom) based on the sequences of at least 70 bp upstream and 70 bp downstream from each candidate FD, after accounting for within-strain polymorphisms (Table S4). Individual flies were genotyped using MASSARRAY (Sequenom) single-base extension in a 32-plex format using standard protocols (Table S5).

Linkage maps were constructed using R-QTL (Broman *et al.* 2003; Broman and Sen 2009). To map QTL responsible for the differences in sex comb morphology, we applied the Haley–Knott, multiple imputation, and expectation-maximization models (Dempster *et al.* 1977; Haley and Knott 1992; Sen and Churchill 2001) to our data using the R-QTL package. All three methods gave nearly identical peak locations, LOD scores, and significance levels, indicating that the data are robust to overparameterization. We first performed single-QTL scans to identify likely regions of genotype–phenotype association. For each detected QTL, we performed composite interval mapping and determined that genotypes at neighboring markers did not significantly affect the peak LOD score or width. We calculated the additive and epistatic interactions between all markers with scans utilizing two-QTL models. To determine the statistical significance of QTL peaks, we used a genome scan-adjusted *P*-value corresponding to the observed LOD score. The null distribution was derived through a standard permutation test. To test for the presence of a QTL on the nonrecombining fourth chromosome (Muller F), we genotyped two FDs in eight of the lightest and eight of the darkest individuals in each F_2 backcross using cleaved amplified polymorphism sequences (Darvasi and Soller 1992; Konieczny and Ausubel 1993). QTL association power analysis was performed using R-QTLDesign (Sen *et al.* 2006). A more detailed description of sequencing and genotyping methods is provided in the Expanded Methods Section online (File S5).

Mapping candidate genes to the D. bipectinata genome

To determine the locations of *Scr* and *dsx* on our linkage maps, we BLASTed the full-length sequences of the *D. ananassae* genomic regions encompassing each gene, as well as the mature transcript of each gene, against the *D. bipectinata* genome assembly. Both the genomic and the transcript sequences mapped unambiguously to genome scaffolds that contained several of our genotyping markers. *Scr* mapped to scaffold scf7180000396708, which also contained markers E-In(2L)D-u7 through E-In(2L)D-u13, while *dsx* mapped to scaffold scf7180000395971, which also contained marker E-In(2L)D-u16. On the linkage map, both of these scaffolds are located in the distal-most, nonrecombining segment of Muller E (2L) corresponding to the inversion in In(2L)D. A similar BLAST analysis shows that genotyping markers linked to the major Muller E QTL are located in a different, more proximal nonrecombining region corresponding to the inversion In(2L)M (Table S6).

Allele-specific expression analysis

Allele-specific pyrosequencing was performed in male F_1 hybrids between *D. bipectinata* bip3-isoA and *D. malerkotliana* mal0-sc2. First and second pupal legs between 16 and 20 hr after pupariation were dissected and stored in TRIzol (Invitrogen) at −70°. Three replicates of 24–43 individuals each were collected. RNA samples were extracted and reverse-transcribed using oligo(dT) primer and Superscript II (Invitrogen) following the manufacturer's protocols. Second-strand synthesis was performed using DNA polymerase I and RNase-H. Nucleotide substitutions in the *Scr* sequence were identified by amplifying and sequencing an ~500-bp fragment of the second coding *Scr* exon from each parental strain. A 163-bp region flanking the chosen SNP was amplified using primers Fwd CATGTGGTACGGCACGATGTTCA and Rev biotin-GAGTTCCACTTCAACCGCTACCTG. Extension primer CTTGTGCTCCTTCTTCCACTTCA, which anneals

Table 1 Sex comb size in three species of the *bipectinata* species complex

Species	Strain	No. of sex-comb teeth (mean ± SD)	*n*
D. malerkotliana	14024-0391.00	6.85 ± 1.03	53
D. malerkotliana	SWB17	5.24 ± 0.72	37
D. bipectinata	14024-0381.00	14.10 ± 1.14	59
D. bipectinata	14024-0381.02	13.55 ± 1.2	43
D. bipectinata	14024-0381.03	18.69 ± 1.37	46
D. bipectinata	14024-0381.04	16.81 ± 1.49	31
D. parabipectinata	14024-0401.00	12.38 ± 1.08	40
D. parabipectinata	14024-0401.02	12.98 ± 1.38	52

upstream from the targeted SNP site, was used to measure allele-specific gene expression as described (Wittkopp *et al.* 2004). The sequence immediately downstream of this primer is TAC in *D. bipectinata* and TGC in *D. malerkotliana*. The polymorphic A/T site (underlined) corresponds to position 1871 in the *D. melanogaster Scr-A* transcript.

Results

Phenotypic and chromosomal variation in D. bipectinata and D. malerkotliana

To examine the genetics of re-evolution of large rotated sex combs, we carried out a series of crosses between *D. bipectinata* and *D. malerkotliana*. Prior to that analysis, we analyzed several strains of each species to estimate the degree of intraspecific variation. The number of sex comb teeth per leg varied from 5.2 ± 0.72 to 6.9 ± 1.03 in *D. malerkotliana* and from 13.6 ± 1.2 to 18.7 ± 1.37 in *D. bipectinata* (Table 1). There was no detectable variation in other aspects of sex comb morphology such as their position and orientation or the shape and color of teeth. Two strains of *D. bipectinata* and two of *D. malerkotliana* were crossed in all possible combinations and in both directions. F_1 hybrid males showed sex comb morphology that was intermediate in all respects (number of teeth, orientation, and tooth morphology and color). Males from reciprocal crosses showed only slight differences in sex comb size (Table S7), indicating that this phenotype is controlled predominantly by autosomal genes.

D. bipectinata and *D. malerkotliana* differ by several fixed chromosomal inversions, and each species is also polymorphic for many inversions (Tomimura *et al.* 2005). By examining the polytene chromosomes of multiple strains of each species and their F_1 hybrids, we determined that the strains *D. malerkotliana* 14024-0391.00 and *D. bipectinata* 14024-0381.03 differed by the smallest number of inversions, all but one of which are completely fixed between the two species. The remaining inversion In(2L)D, which occupies the distal part of the chromosome arm 2L from 18A to 28D (Table S8), is polymorphic within *D. bipectinata*. However, repeated attempts to cross the only available strain that lacked this inversion to different strains of *D. malerkotliana* did not succeed.

We inbred *D. malerkotliana* 14024-0391.00 and *D. bipectinata* 14024-0381.03 for 12 and 18 generations by single-pair full-sib crosses, respectively. The resulting strains, called *mal0-sc2* and *bip3-isoA*, were used for all subsequent experiments. These strains differ by one inversion each on XL and XR (Muller element A), two adjacent inversions on 2L (Muller E), none on 2R (Muller D), one on 3L (Muller C), and several overlapping inversions that cover almost the entire 3R (Muller B). The chromosome order of each strain is given in Table S8, and the autosomal inversions are shown in Figure S1. In crosses between these two strains, ~50% of the euchromatic genome is locked inside chromosomal inversions.

QTL mapping in F_2 hybrids identifies a major QTL on Muller-E

To estimate the number of genomic regions contributing to the dramatic difference in sex comb morphology between *D. bipectinata* and *D. malerkotliana*, we first performed two F_2 backcrosses between *mal0-sc2* and *bip3-isoA*. F_1 females from the cross between *D. malerkotliana* females and *D. bipectinata* males were crossed separately to *mal0-sc2* or *bip3-isoA* males. In the former cross, F_2 recombinant males are either heterozygous for the *D. malerkotliana* and *D. bipectinata* alleles or homozygous for the *D. malerkotliana* allele at each locus; in the latter, F_2 males are either heterozygous or homozygous for the *D. bipectinata* alleles. We examined the sex combs of 427 and 528 F_2 males from these two crosses, respectively. In both panels, all aspects of sex comb morphology were correlated: males with the largest number of sex comb teeth had fully rotated sex combs with curved dark teeth, males with the smallest number of teeth had unrotated sex combs with straight light teeth, and those with an intermediate number of teeth were also intermediate in sex comb orientation and tooth morphology. The number of teeth could be quantified unambiguously, while the angle of rotation proved difficult to quantify due to the variation in the orientation of legs mounted on slides. We therefore used the number of teeth (average between the left and right forelegs) as proxy for species-specific sex comb morphology in subsequent analyses. The distribution of sex comb size in each F_2 backcross is shown in Figure 2, A and B.

A total of 188 F_2 males from the *mal0-sc2* backcross and 163 males from the *bip3-isoA* backcross were genotyped for 28 SNP markers distributed among all major chromosome

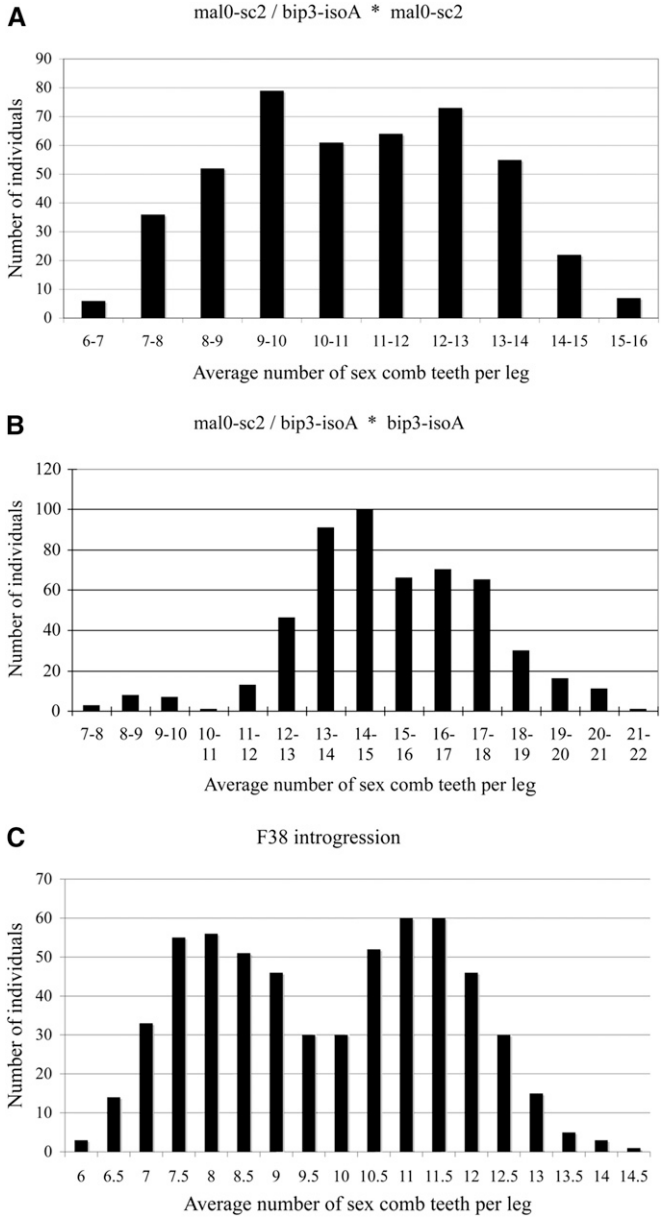

Figure 2 Distribution of sex comb size in the genotyping panels. (A) Progeny of *D. malerkotliana* mal0-sc2/*D. bipectinata bip3-isoA* F_1 hybrid females and *bip3-isoA* males. (B) Progeny of *D. malerkotliana mal0-sc2*/*D. bipectinata bip3-isoA* F_1 females and *mal0-sc2* males. (C) F_{38} introgression (see text for details).

arms. Since recombination was low and uneven due to inversion heterozygosity in F_1 females, linkage maps were inferred using a combination of empirical genetic distances and physical locations of the markers in the genomes of *D. bipectinata* and *D. ananassae* (Figure S2). QTL analysis showed that, in both crosses, a large fraction of the difference between species was explained by a single genomic region that spanned most of Muller E (2L) (Figure 3). In the *mal0-sc2* backcross, mean sex comb size was 9.1 ± 1.12 for *malerkotliana* homozygotes and 12.3 ± 1.17 for *bipectinata/malerkotliana* heterozygotes for all markers on distal Muller E. In the *bip3-isoA* backcross, mean sex comb size was 14.2 ± 1.31 for *bipectinata/malerkotliana* heterozygotes and 17.2 ± 1.79 for *bipectinata* homozygotes. Thus, this QTL interval accounts for ~6–6.4 teeth, or slightly >50% of the total difference between parental strains.

Other QTL intervals had much weaker effects. A QTL located on Muller C + B (3L + 3R) was significant at the 95% level and had a total effect of ~1.43 teeth in the *D. malerkotliana* backcross, but was not significant with an effect of ~0.93 teeth in the *D. bipectinata* backcross (Figure 3). A possible QTL on Muller A (XL + XR) had an effect of ~0.98 teeth in the *D. bipectinata* backcross and ~0.65 teeth in the *D. malerkotliana* backcross and did not reach significance in either panel. No QTL were detected on Muller D (2R) or Muller F (the dot chromosome). Two-QTL scans did not reveal any epistatic interactions between the Muller E, Muller C + B, and Muller A QTL, suggesting that the loci controlling variation in sex comb size act in an additive manner. Under the additive model, and assuming that both Muller C + B and Muller A contain true QTL, all detected QTL together explain only 7.9–8.4 teeth, or ~67–72% of the difference in sex comb size between *bip3-isoA* and *mal0-sc2*. The rest of this difference is likely to be controlled by even weaker QTL that are below our power of detection. Assuming that all QTL are fully additive, the number of genotyped males provides 95% power to detect a QTL with an effect size of at least 1.01 teeth in the *D. malerkotliana* backcross, or at least 1.41 teeth in the *D. bipectinata* backcross, for QTL located on Muller A–E (Lynch and Walsh 1998; Sen *et al.* 2007). These estimates are in agreement with the effect size of the weak QTL detected on Muller C+B.

Refined mapping localizes the major QTL to a single chromosomal inversion

The 2L/Muller E chromosome arm carries two adjacent interspecific inversions, In(2L)D (18A; 28D) and In(2L)M (28D; 34A) (Table S8 and Figure S1). The proximal 2L (34A–45D) and all of 2R/Muller D are free of inversions. To examine the genetic basis of interspecific differences more closely, we sought to increase the amount of recombination between the *D. bipectinata* and *D. malerkotliana* genomes using an introgression approach. F_2 males were sterile in both backcrosses. We crossed F_2 females from the *mal0-sc2* backcross to *mal0-sc2* males. Some fraction of F_3 males was fertile when crossed to *mal0-sc2* females. In the F_4, we selected males with the largest sex combs, which were used to found an introgression strain. There is some recombination in *D. bipectinata* males, but it is low compared to females (Singh and Banerjee 1996). Since the sex comb phenotype can be scored only in males, we used the following crossing scheme: in even-numbered generations, hybrid males with the largest sex combs were selected and crossed to *mal0-sc2* females, while in odd-numbered generations randomly chosen hybrid females were crossed to *mal0-sc2* males (Figure 4A). These crosses should eventually make the introgression strain homozygous for *D. malerkotliana*

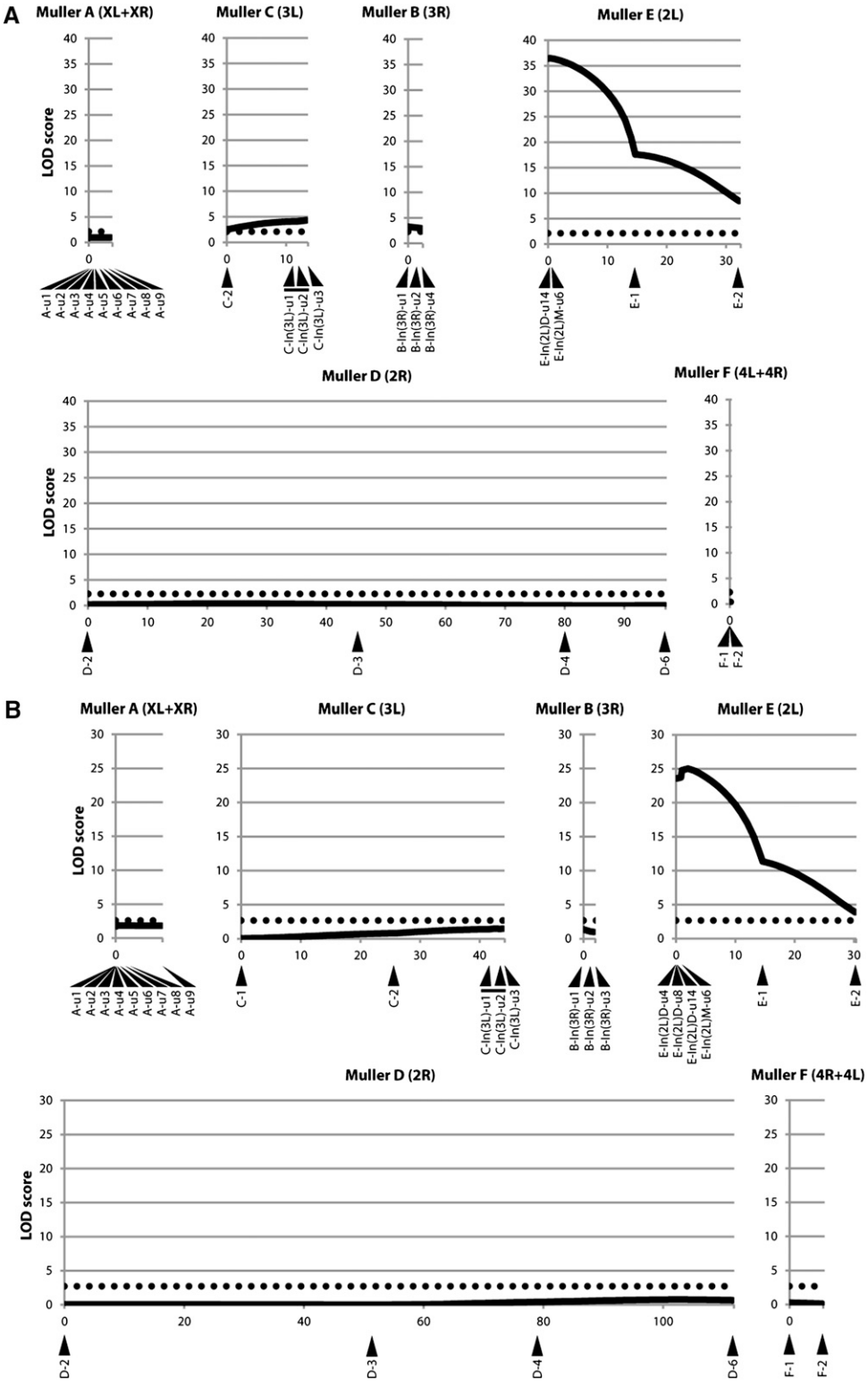

Figure 3 QTL mapping of sex comb size in the F_2 backcrosses. (A) The progeny of *D. malerkotliana mal0-sc2/D. bipectinata bip3-isoA* F_1 females and *mal0-sc2* males. (B) Progeny of *D. malerkotliana mal0-sc2/D. bipectinata bip3-isoA* F_1 hybrid females and *bip3-isoA* males. The genome-wide significance threshold is 2.27 in the former cross and 2.72 in the latter. The *x*-axis is in centimorgans. Marker names and locations are indicated as follows. Each marker name begins with the Muller element on which it is located and ends with the arbitrary marker number. For markers inferred to be inside a chromosomal inversion, the name of that inversion is added in the middle. Since these and adjacent markers cosegregated as nonrecombining blocks, their relative positions could not be determined by meiotic mapping. The numbers of such markers are preceded by "u" for "unmapped," and their order on the map is arbitrary. For example, E-In(2L)D-u4 is marker #4 located on Muller E (chromosome arm 2L) in the inversion In (2L)D and could not be mapped by recombination, while D-3 is marker #3 located on Muller D (chromosome arm 2R) outside of any inversions.

alleles at all loci, with the exception of genomic regions that are strongly linked to genes responsible for the interspecific differences in sex comb morphology.

We examined polytene chromosomes in the introgression strain after 20 generations. We found that it was polymorphic for In(2L)D and In(2L)M, indicating that it was heterozygous for the *D. bipectinata* and *D. malerkotliana* alleles on Muller E. On all other chromosome arms, the introgression was homozygous for the *D. malerkotliana* arrangement. This result suggested that only the 2L (and potentially

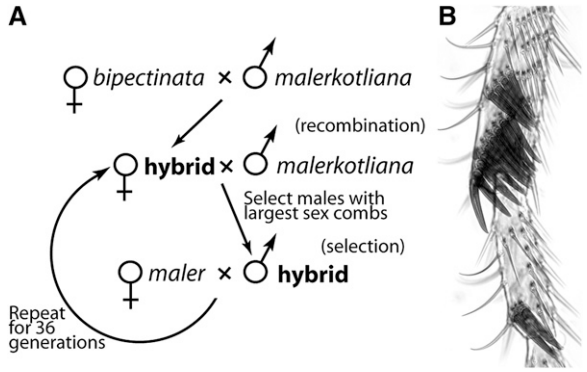

Figure 4 Phenotypic introgression between *D. bipectinata* and *D. malerkotliana*. (A) Crossing scheme. (B) Sex comb of a male from the introgression strain after >20 generations of backcrossing.

2R) made a major contribution to species differences, confirming the results of F_2 QTL mapping. Throughout the introgression process, different aspects of sex comb morphology (orientation and the number, shape, and color of teeth) continued to be correlated (Figure 4B).

After 36 total generations of introgression (corresponding to 19 recombining female generations), males with the largest sex combs were crossed to *mal0-sc2* females, and the resulting F_{37} males and females were crossed to each other en masse. This resulted in F_{38} males that could in principle be homozygous for the *D. bipectinata* allele, heterozygous, or homozygous for the *D. malerkotliana* allele at any locus. We examined the sex combs of 590 F_{38} males. The distribution of sex comb sizes was more clearly bimodal than in the F_2 (Figure 2C), suggesting that this phenotype was largely controlled by a single genomic region and that some of the weaker QTL have been removed by repeated backcrossing. A total of 185 F_{38} males were first genotyped for the same 28 SNP markers as the F_2 panels; preliminary QTL mapping was consistent with the F_2 results. We therefore genotyped the F_{38} panel for 25 additional markers on Muller E and 6 additional markers on Muller D.

In the F_{38} recombinant panel, all marker loci located on the proximal Muller E, most of Muller D, and all other chromosome arms were homozygous for *D. malerkotliana* alleles, while the distal portion of Muller E that carries the inversions In(2L)D and In(2L)M was polymorphic for the *D. bipectinata* and *D. malerkotliana* alleles and showed a strong association with the sex comb phenotype (Figure 5). We did not observe any males homozygous for the *D. bipectinata* alleles on Muller E, suggesting that some interaction between one or more *D. bipectinata* genes in this region with homozygous *D. malerkotliana* alleles elsewhere in the genome results in hybrid lethality. With a single exception (see below), the entire distal region of Muller E segregated as a single block, as expected from the inversion heterozygosity. Males homozygous for the *D. malerkotliana* alleles throughout this region had small sex combs (8.3 ± 1.13 teeth), whereas the *malerkotliana/bipectinata* heterozygotes had sex combs that were intermediate in size between the parental species (11.3 ± 1.04) (Figure 5C). Thus, in the F_{38} as well as in the F_2, this single QTL region accounts for ~6 teeth, slightly >50% of the total difference between the *D. bipectinata* and *D. malerkotliana* parents. Sex comb size in the F_{38} males homozygous for the *D. malerkotliana* Muller E (8.3 ± 1.13 teeth) is larger than in the *mal0-sc2* parent (7.13 ± 1.13 teeth) but smaller than in the F_2 males homozygous for the *D. malerkotliana* Muller E (9.1 ± 1.12 teeth). This suggests that some of the minor QTL persist in the introgression strain but are below our power of detection.

We observed a single recombination event between In(2L)D and In(2L)M. This rarity is not surprising, given the close proximity between the inversion breakpoints. This fortuitous event allowed us to localize the region responsible for the differences in sex comb morphology more precisely. The recombinant male was homozygous for *D. malerkotliana* alleles in the more proximal block of markers, presumably corresponding to In(2L)M, but heterozygous for the *D. bipectinata* and *D. malerkotliana* alleles in the more distal block that presumably corresponds to In(2L)D (Figure 5C). The number of sex comb teeth in this male (8) almost exactly matches the mean sex comb size of the 93 males that are homozygous for *D. malerkotliana* alleles over the entire Muller E (8.30 teeth), but is clearly different from the mean phenotype of the 91 males that are heterozygous for the *D. bipectinata* and *D. malerkotliana* alleles in both In(2L)D and In(2L)M (11.32 teeth). This indicates that the major Muller E QTL interval corresponds to the In(2L)M inversion.

The markers that cosegregate with In(2L)M map to four scaffolds in the genome of *D. bipectinata*. Together, these scaffolds cover 6049 kb and contain several hundred genes including transcription factors, Polycomb and Trithorax group genes, and other regulatory genes that could potentially affect sex comb development (Table S9). Since recombination mapping within the In(2L)M inversion is not feasible, we cannot determine whether this QTL corresponds to a single locus or reflects the cumulative effect of several weaker QTL.

Scr and dsx are not directly responsible for the interspecific differences

The HOX gene *Sex combs reduced* (*Scr*) and the sex determination gene *doublesex* (*dsx*) play central roles in sex comb development. The evolutionary origin of sex combs coincides with the origin of a new *dsx* expression domain and novel regulatory interactions between *Scr* and *dsx* (Barmina and Kopp 2007; Tanaka *et al.* 2011). A combination of experimental and comparative evidence suggests that changes in *dsx* and *Scr* expression were responsible for sex comb evolution (Kopp 2011; Tanaka *et al.* 2011). Moreover, *dsx* and *Scr* expression differs between *D. bipectinata* and *D. malerkotliana* in a way consistent with their morphological differences (Barmina and Kopp 2007; Tanaka *et al.* 2011). Since both genes are located on Muller E (2L

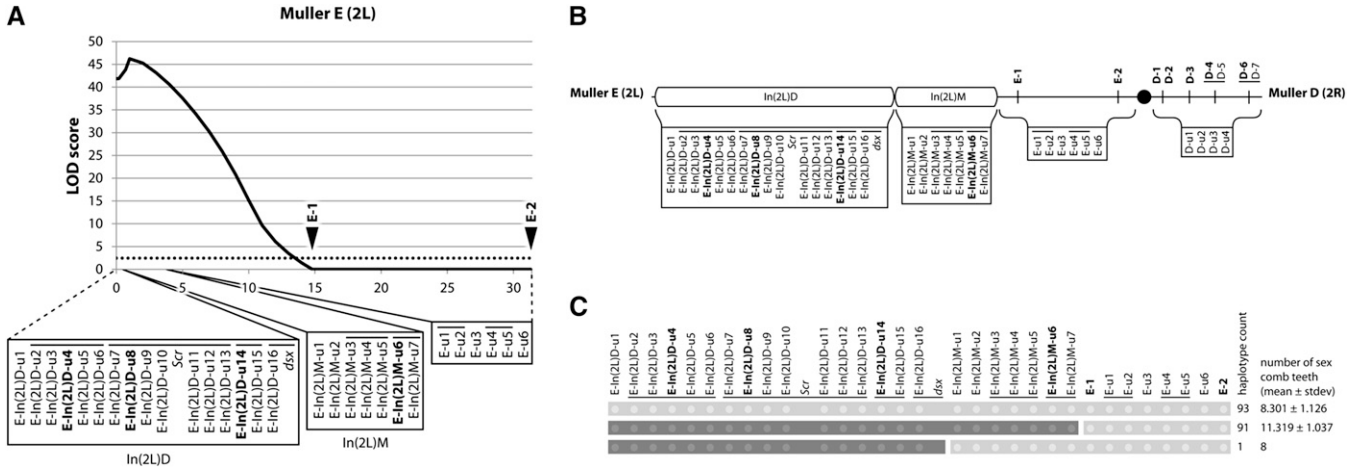

Figure 5 QTL mapping of sex comb size in the introgression line. (A) QTL plot. Markers genotyped in both the F_2 and the introgression are highlighted by boldface type; markers that were genotyped only in the introgression are in roman. Markers and candidate genes located on the same scaffold in the genome sequence of *D. bipectinata* are connected by horizontal lines at the top. See legends to Figure 3 and Figure S2 for marker nomenclature. (B) Locations of genotyping markers and candidate genes relative to chromosomal inversions. Note that most markers could not be mapped by recombination because they either are located inside inversions [In(2L)D and In(2L)M] or were homozygous for *D. malerkotliana* alleles. Markers that were mapped by recombination in the F_2 are shown above the line. (C) Phenotypes associated with different haplotype blocks. Light gray shows markers that were homozygous for the *D. malerkotliana* allele, and dark gray indicates markers that were heterozygous for the *D. malerkotliana* and *D. bipectinata* alleles.

of *D. bipectinata*), we tested whether they could be responsible for the differences in sex comb morphology between these species.

We used allele-specific pyrosequencing (Cowles *et al.* 2002; Wittkopp *et al.* 2004) in F_1 hybrids between *bip3-isoA* and *mal0-sc2* to test whether the interspecific differences in *Scr* expression had a *cis*-regulatory component. We found that the *D. bipectinata* allele of *Scr* was expressed at a significantly higher level than the *D. malerkotliana* allele in the prothoracic, but not in the mesothoracic, pupal legs of F_1 hybrid males (*t*-test $P = 0.0003$; Figure 6). Thus, *Scr* expression has diverged between *D. malerkotliana* and *D. bipectinata* due at least in part to changes at the *Scr* locus.

To localize *Scr* and *dsx* relative to the linkage map and inversion boundaries, we BLASTed the coding sequences of these genes and the transcriptome contigs that contained our genotyping markers against the *D. bipectinata* genome scaffolds. We found that both genes were located on genomic scaffolds that also included SNP markers that were part of the In(2L)D linkage block (Figure 5, Table S6). In contrast, phenotypic differences between *D. malerkotliana* and *D. bipectinata* are associated with the In(2L)M block (Figure 5C). Thus, despite the evidence for *cis*-regulatory divergence at the *Scr* locus, neither that gene nor *dsx* are directly responsible for species divergence.

Discussion

A single inversion, In(2L)M, is responsible for slightly more than half of the difference in sex comb morphology between *D. bipectinata* and *D. malerkotliana*. Minor QTL located elsewhere in the genome have much weaker effects. In(2L)M spans chromosomal bands 28D–34A, out of the total of 100 chromosome divisions. In other words, much of the interspecific difference maps to ~5% of the genome. Other chromosomal inversions that are also completely fixed between *D. malerkotliana* and *D. bipectinata*, including several inversions that are much larger than In(2L)M, make little or no contribution to the differences in sex comb morphology. Thus, although we cannot determine the number of genes in the In(2L)M inversion that contribute to its total effect, it is possible that this number is relatively small.

Re-evolution of large rotating sex combs in the last common ancestor of *D. bipectinata* and *D. parabipectinata* represents a major developmental change. The number of bristles recruited into the sex comb, rotation of the surrounding epithelium, the shape of bristle shafts, and their pigmentation have very different molecular underpinnings (Kopp 2011). Epithelial rotation in particular is a complex morphogenetic process involving localized convergent extension and must require numerous genes involved in cell polarity, cell adhesion, and cytoskeletal dynamics (Atallah *et al.* 2009a; Tanaka *et al.* 2009). What types of genetic changes could allow this entire suite of processes to re-evolve following an earlier loss (Barmina and Kopp 2007; Matsuda *et al.* 2009)?

We suggest that violations of Dollo's law can be made more likely by the modular organization of developmental pathways. Many pathways have "nexus" regulatory genes that activate multiple downstream targets responsible for different cellular processes. The loss of a complex trait can happen easily if the expression of the nexus gene in the progenitor tissue is disrupted, since the entire downstream

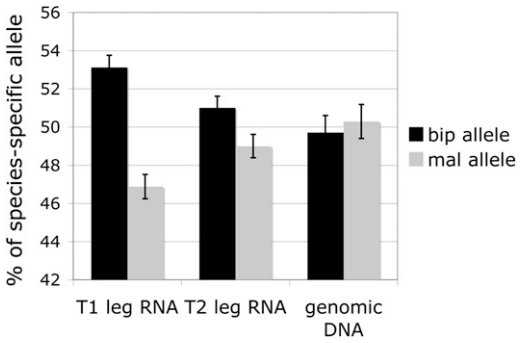

Figure 6 Allele-specific expression of *Scr* in the F_1 hybrids between *D. bipectinata* bip3-isoA and *D. malerkotliana* mal0-sc2. The *y*-axis shows the proportion of overall expression represented by each species-specific allele. Error bars are based on three biological replicates.

developmental program will be lost automatically. In fact, rapid loss of morphological structures by loss-of-function mutations in a single regulatory gene has been documented in several cases (McGregor *et al.* 2007; Chan *et al.* 2010). The same property of developmental pathways could also explain the re-evolution of lost traits: as long as the downstream pathway remains intact, regaining the expression of the nexus gene or genes in the progenitor tissue will be sufficient to restore much of the original structure. For example, several lineages of swordtail fish (*Xiphophorus*) have secondarily lost the male swords (extended tail fins) (Meyer *et al.* 1994, 2006). In at least some of these species, small vestigial "swordlets" can be restored by exposure to abnormally high levels of testosterone (Gordon *et al.* 1943). In *X. maculatus*, a swordless species, a single mutation is sufficient to form a similar (although not identical) fin extension, presumably by enhancing cell response to endogenous testosterone (Offen *et al.* 2008). These observations suggest that the pathway responsible for sword development has decayed only partially following the loss of the sword and can be brought back by changes in a relatively small number of genes. On a much deeper timescale, teeth are absent in all modern birds; however, a single mutation in the chicken *talpid2* gene can partially restore tooth development, inducing integumentary outgrowths that resemble crocodilian teeth (Harris *et al.* 2006). Of course, this is only possible because some of the regulatory landscape that controls odontogenesis has been retained in avian oral tissues (Chen *et al.* 2000; Mitsiadis *et al.* 2006).

The question, then, is, why would the downstream pathway (*i.e.*, a large set of regulatory interactions among genes) stay intact and not succumb to mutation accumulation in the absence of selection on the defunct structure? One possible explanation is that the vast majority of genes have pleiotropic functions. In the simplest case, the coding sequence of a gene whose expression has been lost in one tissue will remain under purifying selection as long as it continues to be expressed in other tissues (Marshall *et al.* 1994; Collin and Miglietta 2008). More generally, the modular organization of development suggests that pleiotropy can protect entire pathways: selection pressure on most regulatory interactions may still be present if the pathway acts in other tissues. This mechanism may be particularly important for structures that have serial homologs, which share largely the same developmental programs. For example, re-evolution of lost digits in lizards and mandibular teeth in frogs is likely to be enabled by the fact that other digits, and other teeth, have always been retained (Brandley *et al.* 2008; Kohlsdorf *et al.* 2010; Wiens 2011).

Even if selection does not preserve the entire pathway, preservation of its component modules may be enough to retain capacity for re-evolution. In *Drosophila*, the cellular module responsible for making thickened, rounded, darkly pigmented sex comb teeth may also be deployed to make similar bristles in other body parts such as male genitalia; the module responsible for sex comb rotation may act in other epithelial sheets that undergo convergent extension, and so on. If selection on pleiotropically linked traits preserves most regulatory interactions from mutational decay, the pathway as a whole could remain largely intact and require changes in only a few genes to restore it to a modestly functional state that can then be refined by additional mutations. The genetic architecture of sex comb morphology in *D. bipectinata vs. D. malerkotliana* could conceivably be the result of such a process.

Understanding the molecular mechanisms that enable re-evolution of complex traits will require a developmental genetics perspective; QTL mapping is only the first step in this direction. Although the limitations imposed by fixed chromosomal inversions preclude us from identifying the major gene or genes responsible for the re-evolution of large rotating sex combs in *D. bipectinata* by linkage mapping, our growing understanding of sex comb development may ultimately allow us to overcome these limitations.

Acknowledgments

We are grateful to Muneo Matsuda for his help in analyzing chromosome order in the *bipectinata* species complex; to Y. Fuyama and the San Diego *Drosophila* Species Stock Center for providing fly strains; and to Rachael Curtis, Andrew Hamilton, and Mary Magsombol for help with phenotypic analysis and DNA extractions. Illumina sequencing was performed at the University of California Davis Genome Center. Genotyping was performed at the University of California Davis Veterinary Genetics Laboratory. This work was supported by National Institutes of Health grant 5-R01GM082843-02.

Literature Cited

Atallah, J., N. H. Liu, P. Dennis, A. Hon, D. Godt *et al.*, 2009a Cell dynamics and developmental bias in the ontogeny of a complex sexually dimorphic trait in Drosophila melanogaster. Evol. Dev. 11: 191–204.

Atallah, J., N. H. Liu, P. Dennis, A. Hon, and E. W. Larsen, 2009b Developmental constraints and convergent evolution in Drosophila sex comb formation. Evol. Dev. 11: 205–218.

Atallah, J., H. Watabe, and A. Kopp, 2012 Many ways to make a novel structure: different mechanisms of sex comb development in Lordiphosa and Sophophora. Evol. Dev. (in press).

Barmina, O., and A. Kopp, 2007 Sex-specific expression of a HOX gene associated with rapid morphological evolution. Dev. Biol. 311: 277–286.

Birol, I., S. D. Jackman, C. B. Nielsen, J. Q. Qian, R. Varhol *et al.*, 2009 De novo transcriptome assembly with ABySS. Bioinformatics 25: 2872–2877.

Bock, I. R., 1978 The bipectinata complex: a study of interspecific hybridization in the genus Drosophila. Aust. J. Biol. Sci. 31: 197–208.

Brandley, M. C., J. P. Huelsenbeck, and J. J. Wiens, 2008 Rates and patterns in the evolution of snake-like body form in squamate reptiles: evidence for repeated re-evolution of lost digits and long-term persistence of intermediate body forms. Evolution 62: 2042–2064.

Broman, K. W., and Ś. Sen, 2009 *A Guide to QTL Mapping with R/qtl*. Springer, New York.

Broman, K. W., H. Wu, Ś. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal Jr, M. D. Shapiro *et al.*, 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327: 302–305.

Chen, Y., Y. Zhang, T. X. Jiang, A. J. Barlow, T. R. St. Amand *et al.*, 2000 Conservation of early odontogenic signaling pathways in Aves. Proc. Natl. Acad. Sci. USA 97: 10044–10049.

Collin, R., and R. Cipriani, 2003 Dollo's law and the re-evolution of shell coiling. Proc. Biol. Sci. 270: 2551–2555.

Collin, R., and M. P. Miglietta, 2008 Reversing opinions on Dollo's Law. Trends Ecol. Evol. 23: 602–609.

Cowles, C. R., J. N. Hirschhorn, D. Altshuler, and E. S. Lander, 2002 Detection of regulatory variation in mouse genes. Nat. Genet. 32: 432–437.

Darvasi, A., and M. Soller, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor. Appl. Genetics 85: 353–359.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977 Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B 39: 1–38.

Galis, F., J. W. Arntzen, and R. Lande, 2010 Dollo's law and the irreversibility of digit loss in Bachia. Evolution 64: 2466–2476; discussion 2477–2485.

Goldberg, E. E., and B. Igic, 2008 On phylogenetic tests of irreversible evolution. Evolution 62: 2727–2741.

Gordon, M., H. Cohen, and R. Nigrelli, 1943 A hormone-produced taxonomic character in Platypoecilus maculatus diagnostic of wild P. xiphidium. Am. Nat. 77: 569–572.

Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315–324.

Harris, M. P., S. M. Hasso, M. W. Ferguson, and J. F. Fallon, 2006 The development of archosaurian first-generation teeth in a chicken mutant. Curr. Biol. 16: 371–377.

Huang, X., and A. Madan, 1999 CAP3: a DNA sequence assembly program. Genome Res. 9: 868–877.

Kohlsdorf, T., and G. P. Wagner, 2006 Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in Bachia (Gymnophthalmidae: Squamata). Evolution 60: 1896–1912.

Kohlsdorf, T., V. J. Lynch, M. T. Rodrigues, M. C. Brandley, and G. P. Wagner, 2010 Data and data-interpretation in the study of limb evolution: a reply to Galis *et al.*, on the re-evolution of digits in the lizard genus Bachia. Evolution 64: 2477–2485.

Konieczny, A., and F. M. Ausubel, 1993 A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4: 403–410.

Kopp, A., 2011 Drosophila sex combs as a model of evolutionary innovations. Evol. Dev. 13: 504–522.

Kopp, A., and O. Barmina, 2005 Evolutionary history of the Drosophila bipectinata species complex. Genet. Res. 85: 23–46.

Kurten, B., 1963 Return of a lost structure in the evolution of the felid dentition. Biol. 26: 1–12.

Li, R., C. Yu, Y. Li, T.-W. Lam, S.-M. Yiu *et al.*, 2009 SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966–1967.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.

Lynch, V. J., and G. P. Wagner, 2010 Did egg-laying boas break Dollo's law? Phylogenetic evidence for reversal to oviparity in sand boas (Eryx: Boidae). Evolution 64: 207–216.

Marshall, C. R., E. C. Raff, and R. A. Raff, 1994 Dollo's law and the death and resurrection of genes. Proc. Natl. Acad. Sci. USA 91: 12283–12287.

Matsuda, M., C.-S. Ng, M. Doi, A. Kopp, and Y. N. Tobari, 2009 Evolution in the Drosophila ananassae species subgroup. Fly (Austin) 3: 1–13.

McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan *et al.*, 2007 Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448: 587–590.

Meyer, A., J. M. Morrissey, and M. Schartl, 1994 Recurrent origin of a sexually selected trait in Xiphophorus fishes inferred from a molecular phylogeny. Nature 368: 539–542.

Meyer, A., W. Salzburger, and M. Schartl, 2006 Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. Mol. Ecol. 15: 721–730.

Miller, J. R., S. Koren, and G. Sutton, 2010 Assembly algorithms for next-generation sequencing data. Genomics 95: 315–327.

Mitsiadis, T. A., J. Caton, and M. Cobourne, 2006 Waking-up the sleeping beauty: recovery of the ancestral bird odontogenic program. J. Exp. Zool. B Mol. Dev. Evol. 306: 227–233.

Offen, N., N. Blum, A. Meyer, and G. Begemann, 2008 Fgfr1 signalling in the development of a sexually selected trait in vertebrates, the sword of swordtail fish. BMC Dev. Biol. 8: 98.

Robertson, G., J. Schein, R. Chiu, R. Corbett, M. Field *et al.*, 2010 De novo assembly and analysis of RNA-seq data. Nat. Methods 7: 909–912.

Sen, S., and G. A. Churchill, 2001 A statistical framework for quantitative trait mapping. Genetics 159: 371–387.

Sen, S., J. M. Satagopan, K. W. Broman, and G. A. Churchill, 2007 R/qtlDesign: inbred line cross experimental design. Mamm. Genome 18: 87–93.

Signor, S. A., T. D. Seher, and A. Kopp, 2013 Genomic resources for multiple species in the *Drosophila ananassae* species group. Fly 7(1) (in press).

Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones *et al.*, 2009 ABySS: a parallel assembler for short read sequence data. Genome Res. 19: 1117–1123.

Singh, B. N., and R. Banerjee, 1996 Spontaneous recombination in males of Drosophila bipectinata. J. Biosci. 21: 775–779.

Tanaka, K., O. Barmina, and A. Kopp, 2009 Distinct developmental mechanisms underlie the evolutionary diversification of Drosophila sex combs. Proc. Natl. Acad. Sci. USA 106: 4764–4769.

Tanaka, K., O. Barmina, L. E. Sanders, M. N. Arbeitman, and A. Kopp, 2011 Evolution of sex-specific traits through

changes in HOX-dependent doublesex expression. PLoS Biol. 9: e1001131.
Tomimura, Y., M. Matsuda, and Y. N. Tobari, 2005 Chromosomal phylogeny and geographical divergence in the Drosophila bipectinata complex. Genome 48: 487–502.
Whiting, M. F., S. Bradler, and T. Maxwell, 2003 Loss and recovery of wings in stick insects. Nature 421: 264–267.
Wiens, J. J., 2011 Re-evolution of lost mandibular teeth in frogs after more than 200 million years, and re-evaluating Dollo's law. Evolution 65: 1283–1296.
Wittkopp, P. J., B. K. Haerum, and A. G. Clark, 2004 Evolutionary changes in cis and trans gene regulation. Nature 430: 85–88.

Communicating editor: C. D. Jones

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2012/10/11/genetics.112.145524.DC1

Genetic Basis of a Violation of Dollo's Law: Re-Evolution of Rotating Sex Combs in *Drosophila bipectinata*

Thaddeus D. Seher, Chen Siang Ng, Sarah A. Signor, Ondrej Podlaha, Olga Barmina, and Artyom Kopp

DOI: 10.1534/genetics.112.145524

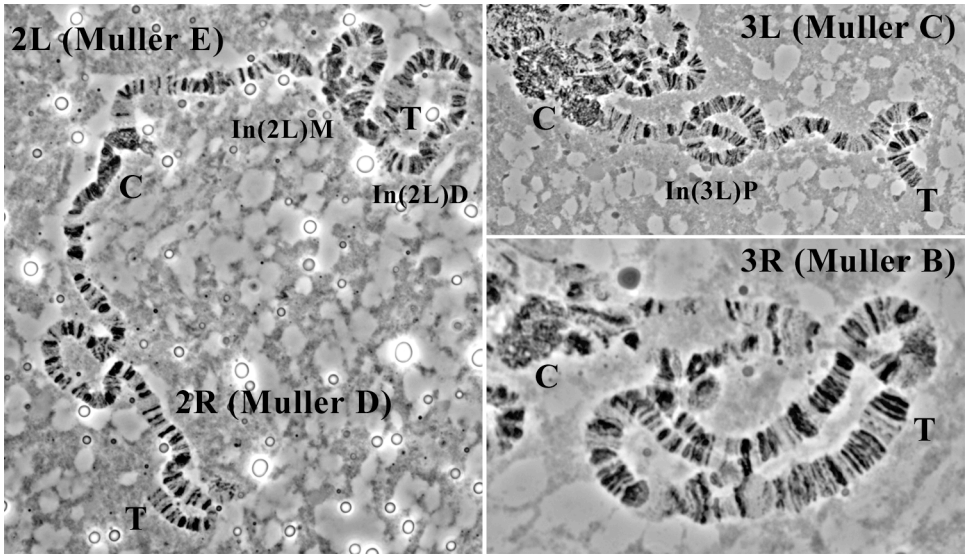

Figure S1 Polytene chromosomes of F_1 hybrids between *D. malerkotliana* mal0-sc2 and *D. bipectinata* bip3-isoA. Each inversion is indicated on 2L and 3L; 2R is free of inversions; and 3R carries several overlapping inversions. C, chromocenter; T, telomere.

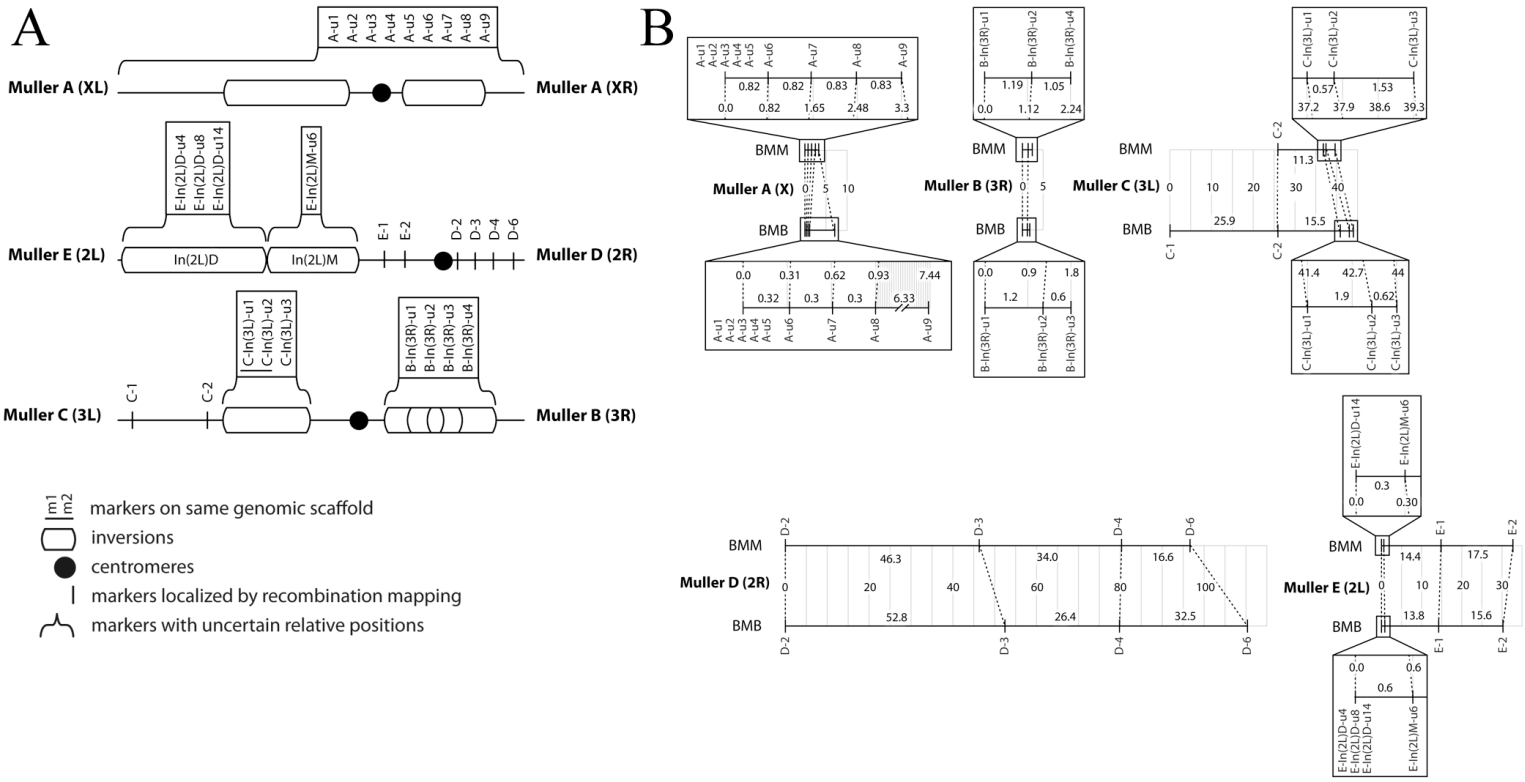

Figure S2 Genetic and chromosome maps of the *D. malerkotliana* / *D. bipectinata* hybrids. **A**. A schematic of the major chromosomes showing the locations of inversions and genotypic markers. **B**. Linkage maps reconstructed from the progeny of *D. malerkotliana* mal0-sc2 / *D. bipectinata* bip3-isoA F_1 females and mal0-sc2 males (BMM, top row) and the progeny of *D. malerkotliana* mal0-sc2 / *D. bipectinata* bip3-isoA F_1 females and bip3-isoA males (BMB, bottom row). The X axis is in centimorgans. Positions of markers on the linkage maps are shown under each marker, and the distance between each pair of adjacent markers is indicated between them. Dotted lines connect markers that were successfully genotyped in both crosses. Marker names and locations are indicated as follows. Each marker name begins with the Muller element on which it is located and ends with the arbitrary marker number. For markers inferred to be inside a chromosomal inversion, the name of that inversion is added in the middle. Groups of markers that co-segregate as a single block are in rectangular boxes. Since the relative positions of such markers cannot be determined by mapping, their numbers are preceded by "u" for "unmapped" and their order on the map is arbitrary. For example, E-In(2L)D-u4 is marker #4 located on Muller E (chromosome arm 2L) in the inversion In(2L)D and could not be mapped by recombination, while D-3 is marker #3 located on Muller D (chromosome arm 2R) outside of any inversions.

File S1 – S4

Supporting Data

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.145524/-/DC1.

File S1A Compressed FASTA file containing basecall results of *de novo* transcriptome assembly of *D. bipectinata bip*3-isoA.

File S1B Compressed FASTA file containing quality results of *de novo* transcriptome assembly of *D. bipectinata bip*3-isoA.

File S1C Compressed FASTA file containing IUPAC ambiguity basecalls of *de novo* transcriptome assembly of *D. bipectinata bip*3-isoA. The polymorphism sensitivity is relatively high in order to design robust primers.

File S2A Compressed FASTA file containing basecall results of *de novo* transcriptome assembly of *D. malerkotliana mal*0-sc2.

File S2B Compressed FASTA file containing quality results of *de novo* transcriptome assembly of *D. malerkotliana mal*0-sc2.

File S2C Compressed FASTA file containing IUPAC ambiguity basecalls of *de novo* transcriptome assembly of *D. malerkotliana mal*0-sc2. The polymorphism sensitivity is relatively high in order to design robust primers.

File S3 A list of all fixed differences between *D. bipectinata* bip3-isoA and *D. malerkotliana* mal0-sc2. FDs were identified as described in the text except the following filters have NOT been applied to this table: FDs cannot be within X base pairs of one another; FDs cannot be within Y bases from the edge of the contig; FDs must be within a homologous *D. ananassae* exon. There are three worksheets for each transcriptome cross-mapping. The worksheets designated with "all" contain the FD position, its depth, the reference basecall, and the allele frequencies. The sheets with the "homologs" label contain the position of the exact same base in several other references, as described in Supplementary Table 3. The "shared" worksheets list only the positions that appear in both the *bipectinata*-to-*malerkotliana* and the *malerkotliana*-to-*bipectinata* cross-mappings.

File S4 This file holds the locations of FDs based on the mapping of *D. malerkotliana* mal0-sc2 reads to the modified genome assembly of *D. bipectinata*. Markers for the second round of genotyping were drawn from this list.

File S5

Expanded Methods Section

cDNA library construction and sequencing

For each parental strain (*mal0-sc2* and *bip3-isoA*), we extracted total RNA from pools of 20-30 randomly picked flies using the standard TRIzol (Invitrogen) protocol. RNA quality and concentration were measured on a Bioanalyzer (Agilent) and NanoDrop (Thermo Scientific). cDNA libraries were synthesized from 5 µg of total RNA using the MINT cDNA synthesis kit (Evrogen) and normalized using the TRIMMER kit (Evrogen) following the manufacturer's protocols. AMPure beads (Agencourt) were used for purification steps. We then used a Bioruptor (Diagenode) to shear the cDNA, and agarose gel electrophoresis to select a narrow range of ~300 bp fragments. Paired-end libraries were prepared from the sheared, normalized cDNA using the standard Illumina protocol, and sequenced with 85-base paired-end reads on an Illumina Genome Analyzer II at the UC Davis Genome Center (Table S1).

Data filtering and transcriptome assembly

Illumina and Evrogen adapter sequences used during library construction were trimmed from the Illumina reads using custom scripts. We then filtered out any reads shorter than 36 bases to minimize repetitive read mapping. Read quality appeared poor because many cDNA fragments contained Evrogen adapters, creating repetitive sequences that influenced basecalling probabilities. We therefore relaxed the standard Illumina 3' trimming of contiguous bases with Phred scores (COCK *et al.* 2010; EWING and GREEN 1998; EWING *et al.* 1998) of 15 or less, and fed the remaining reads into the assembly pipeline without quality filtering.

The partial transcriptomes of *mal0-sc2* and *bip3-isoA* were assembled *de novo* using ABySS (BIROL *et al.* 2009; MILLER *et al.* 2010; SIMPSON *et al.* 2009) followed by Trans-ABySS (ROBERTSON *et al.* 2010) to merge multiple unscaffolded assemblies with different K-mer sizes. The assembly procedure was as follows:

1. We ran paired-end ABySS assemblies with kmer sizes increasing from 29 to 63 in increments of two (n=10, q=0, s=160, c=2, ABYSS_OPTIONS= "--no-chastity --no-trim-masked --illumina-quality").

2. Next we ran Trans-ABySS phase 1, which leverages the BLAT (KENT 2002) algorithm, to exclude repetitive and inferior contigs generated by the individual assemblies.

3. We selected contigs 100 bp or longer from the combined assembly, and ran them through CAP3 (HUANG and MADAN 1999) in order to merge overlapping contigs.

4. We selected contigs 200 bp or longer, and removed contigs that were identical except for one mismatch, keeping the copy with the fewest ambiguous bases.

5. Finally, we grouped all the contigs by homology using NCBI BLASTn (ALTSCHUL *et al.* 1990; ALTSCHUL *et al.* 1997) (word_size= 14, gapopen=12, gapextend=4, penalty=-6, reward=4, minimal 110 bp overlap, minimal bitscore of 130, mismatch to identity ratio ≤ 0.05, allowing for large gaps), uniformly oriented all contigs within each group using a directed graph, and assigned unique identifiers to these groups (Table S2 and Files S1A-B and S2A-B).

This procedure produces a unique set of genes, some of which have multiple isoforms. Subsequent read mapping assigns most reads unambiguously to contigs, but can map the same read redundantly to a group of transcripts corresponding to a single alternatively spliced locus.

Identification of fixed SNP differences

For our genetic analysis, we required fixed differences (FD) between parental strains, i. e. nucleotide positions where *bip3-isoA* is fixed for one allele and *mal0-sc2* for a different allele. To identify FDs for the first round of genotyping, *mal0-sc2* and *bip3-isoA* read libraries were each mapped separately to both *bip3-isoA* and *mal0-sc2 de novo* transcriptome assemblies. Mapping was performed using SOAP2 (LI *et al.* 2009) (m=1, x=1200, v=3, l=40, g=300, s=50, r=2). In order to eliminate ambiguously mapping poly(A+) tails, we first aligned all reads to the *de novo* transcriptomes, then removed any reads that mapped to multiple contig groups and had GC content below 10%. Same-species mapping (*mal0-sc2* reads to *mal0-sc2* transcriptome and *bip3-isoA* reads to *bip3-isoA* transcriptome) serves to correct assembly errors, gauge coverage at each position, and identify nucleotide positions where multiple alleles segregate within each parental strain despite inbreeding. Cross-species mapping allows us to identify polymorphic positions and select FDs that meet coverage cut-offs (File S3).

FDs are positions in cross-species alignments where the consensus basecall is monomorphic and different from the reference base, subject to the following additional requirements: cross-species alignment depth of at least 20; a minimal mean Phred score of 20; mean depth of at least 5 at the 20 flanking bases; and contig coverage extending at least 70 bases upstream and downstream from the FD. We allowed for errors based on the depth of sequencing at the position of interest and on the expected number of individual chromosomes in the pooled samples (KOFLER *et al.* 2011). If depth (d) was lower than the expected number of chromosomes (c), the number of permissible errors for the position to be considered an FD was determined as ln(d), rounded down to the nearest whole number. If depth at the position was higher than the expected number of chromosomes, the number of permissible errors was ln(c)+(d/c), rounded down to the nearest whole number. For positions with more than two alleles, all minor alleles falling below this cut-off were considered sequencing errors if their frequencies were equal. If a majority of non-consensus basecalls supported one of the minor alleles, this position was considered polymorphic rather than FD and excluded from further analysis.

To identify FDs for the second round of genotyping, we first mapped *bip3-isoA* reads to the *D. bipectinata* reference genome (GenBank AFFE00000000.1) using SOAP2 with the same parameters as above, and examined positions

where the pileup consensus did not match the genome reference base. At these positions, we replaced the reference base with the new consensus base if it had a minimum depth of 5 and primary allele frequency above 0.5. This replacement was performed only for exonic positions, as determined by alignment to the annotated genome of *D. ananassae*. Finally, we mapped *mal0-sc2* reads to this modified reference genome and identified FDs as described above (File S4).

We used only FDs that had no other FDs within 70 bp on either side in order to avoid amplification biases during Sequenom genotyping. Since genotyping could also be hampered by intraspecific polymorphisms, we generated a reference sequence where IUPAC ambiguity codes were substituted at positions that were polymorphic in either one or both parental strains. Genotyping primers were designed based on this modified reference (Files S1C and S2C).

Marker design and genotyping

To identify the genomic locations of FDs, we BLASTed the sequence flanking each FD against the *D. ananassae* FlyBase 1.3 July 2011 reference genome and transcriptome. We assigned each FD to the chromosome arm of its BLAST match (SCHAEFFER *et al.* 2008), and used the *D. ananassae* transcriptome to make sure that target amplicons did not span splice junctions. Subsequently, marker flanking sequences were also BLASTed against the *D. bipectinata* genome. For the first round of genotyping, we selected 32 FDs that were evenly distributed among the major chromosome arms (Muller elements A-E) and were located at least 2.4Mb from each other in the *D. ananassae* genome (Table S3). Linkage mapping confirmed all homology-based chromosome assignments except for marker E-2, which was predicted to be located on Muller D (2R) but was linked to multiple markers on Muller E (2L). Because the initial analysis suggested the presence of one or more strong QTLs on Muller E or proximal Muller D, we performed a second round of genotyping with 32 additional FDs concentrated on these chromosome arms (Table S3). For these markers, we chose transcriptome contigs that were either located on different scaffolds in the modified *D. bipectinata* genome, or were separated by at least 0.75 Mb if located on the same scaffold.

Genotyping was performed by the UC Davis Veterinary Genetics Laboratory. Primers were designed using TYPER (Sequenom) based on the sequences of at least 70 bp upstream and 70 bp downstream from each candidate FD, after accounting for intraspecific polymorphisms with IUPAC codes as described above (Table S4). All primers were ordered from IDT. Individual flies were genotyped using MASSARRAY (Sequenom) single base extension in a 32-plex format using standard protocols. Genotype calls were made using default peak intensity thresholds (Table S5).

Construction of linkage maps

Genetic maps were constructed based on two separate F_2 backcrosses and a long-term introgression cross (see Results). Markers were assembled into linkage groups using R-QTL (BROMAN and SEN 2009; BROMAN *et al.* 2003); markers that deviated strongly from the expected Mendelian ratios were excluded. When marker density is low, recombination

frequency will underestimate the actual genetic distance due to the inability to detect double crossovers. We therefore used the standard Haldane (HALDANE 1919) and Kosambi (KOSAMBI 1944) mapping functions to estimate genetic distances between markers; the two formulas produced nearly identical results. We observed very high amounts of recombination on Muller D (chromosome arm 2R). This is probably caused by the interchromosomal effect (STEVISON *et al.* 2011), since all chromosome arms except 2R carry large inversions. To place minimal bounds on the large genetic distances on 2R, we used a simplified Haldane map function that does not include crossover interference or correction for multiple exchanges between adjacent markers.

Within each linkage group, we used the R-QTL "ripple" function to determine the marker order that minimized the number of crossovers. Due to the presence of chromosomal inversions, several potential marker orders had similar likelihoods. In such cases, we tested all possible orders on the chromosome arm and chose the one with the lowest total map length. On Muller A (X chromosome), where very little crossing-over was detected, we selected marker order at random from a set of possible orderings with the highest likelihood. Some pairs of markers showed no recombination and had identical map positions, suggesting they were located inside the same chromosomal inversion. We assigned such markers the same relative order in which their homologs appear in the *D. ananassae* genome. Finally, markers located on the same scaffold in the *D. bipectinata* genome assembly were placed adjacent to each other in the order in which they appear in that scaffold.

Linkage maps were constructed independently for each of the two backcrosses. To create a linkage map for the introgression panel, we merged the two backcross maps as follows. First, we determined the consensus marker order compatible with data from both backcrosses. Second, when data for a pair of adjacent markers was available in both datasets, we averaged the genetic distances between these markers from the two backcrosses. Finally, we used this low-resolution map to anchor additional, more densely spaced markers that were only genotyped in the introgression panel.

Genotyping on the dot chromosome (Muller F)

We used the FD alignments mapped to the *D. bipectinata* genome to identify FDs that were mapped by BLASTn to the fourth chromosome (Muller F) in both *D. melanogaster* and *D. ananassae*. Cleaved amplified polymorphism sequences (CAPSs) were selected by feeding the 500 bp upstream and downstream of the FD into SNP Cutter (ZHANG *et al.* 2005). Two potential CAPS sites were chosen, and gradient PCR was run to determine the optimal conditons. The sites were amplified and digested in single-plex format using restriction enzyme identified by SNP Cutter (PstI for markers F-1 and F-2, with PstI cutting the *D. bipectinata* but not the *D. malerkotliana* allele). Digested amplicons were examined on a 1.8% Agarose gel. CAPS genotyping was performed on eight of the lightest and eight of the darkest individuals in each of the two backcrosses. DNA extracted from pooled *bip3-isoA* flies, pooled *mal0-sc2* flies, and pooled *bip3-isoA* / *mal0-sc2* F_1

flies was used for controls. For every marker, post-restriction amplicon sizes matched the predicted values for each genotype.

QTL mapping

We applied the Haley-Knott, multiple imputation, and expectation-maximization models (DEMPSTER *et al.* 1977; HALEY and KNOTT 1992; SEN and CHURCHILL 2001) to our data using the R-QTL package (BROMAN and SEN 2009; BROMAN *et al.* 2003). All three methods gave nearly identical peak locations, LOD scores, and significance levels, indicating that the data are robust to over-parameterization. We first performed single-QTL scans to identify likely regions of genotype-phenotype association. For each detected QTL, we performed composite interval mapping and determined that genotypes at neighboring markers did not significantly affect the peak LOD score or width. To calculate the statistical significance of QTL peaks, we used a genome scan-adjusted P-value corresponding to an observed LOD score. The null distribution was derived through standard permutation test. The P-value represents the chance, under the null hypothesis of no QTL, of obtaining an LOD score that large or larger somewhere in the genome.

In the introgression cross, but not in either backcross, we detect a possible small-effect association between marker D-6 (most distal 2R) and sex combs. While the LOD score on the distal 2R is significant in both the single-QTL genome scan and the two-QTL genome scan that takes both the distal 2L and distal 2R peaks into account, the genotype ratios at marker D-6 diverge significantly from expectation ($P < 0.001$). We therefore discarded this marker. Marker D-4 (distal 2R) maintained the *D. bipectinata* allele through the introgression and is present in the introgression cross but is not associated with the phenotype. In both backcrosses, but not in the introgression cross, a two-QTL genome scan accounting for the 2L QTL reveals another significant QTL (P-value < 0.001) on proximal Muller C (3L), and to a lesser extent Muller B (3R). The relationship between the 2L and proximal 3L/3R QTL is additive. It appears that the *D. bipectinata* genes responsible for this weak QTL were lost from the introgression strain.

Power analysis was performed using R/qtlDesign (SEN *et al.* 2007) and confirmed using calculations set forth in (LYNCH and WALSH 1998). This analysis leverages the theories described in (DARVASI *et al.* 1993; HALEY and KNOTT 1992; REBAI *et al.* 1995; SIMPSON 1989; SIMPSON 1992). The `detectable()` function in R/qtlDesign was used to determine the minimum effect size of the QTL compared to the Muller E QTL. Environmental variance was estimated from the variance within each parental strain: $\sigma^2 = 1.967$ for *D. malerkotliana* homozygotes and $\sigma^2 = 2.857$ for *D. bipectinata* homozygotes. Genetic variance was calculated by taking the average of the means of the heterozygote and the homozygote, squaring it, and dividing by 4 (as mandated by the backcross model). We assumed a complete linkage between the QTL and a genotyping marker (recombination fraction of 0), and treated all QTLs as additive. The selection fraction was 1 for Muller A-E; for Muller F, it was set to 16/188 for the *D. malerkotliana* backcross and 16/163 for the *D. bipectinata* backcross. For QTLs located on Muller A-E, the minimal detectable effect size is nearly identical for a power of 0.95 and a power of 1. For

Muller F, selective genotyping of extreme individuals (DARVASI and SOLLER 1992) meant that requiring a power of 1 rather than 0.95 would dramatically reduce the probability of detecting QTLs of small effect. In the interest of consistency, we used the power of 0.95 for all chromosome arms.

Mapping candidate genes to the *D. bipectinata* genome

To determine the locations of *Scr* and *dsx* on our linkage maps, we BLASTed the full-length sequences of the *D. ananassae* genomic regions encompassing each gene, as well as the mature transcript of each gene, against the modified *D. bipectinata* genome assembly. Both the genomic and the transcript sequences mapped unambiguously to genome scaffolds that contained several of our genotyping markers. *Scr* mapped to scaffold scf7180000396708 (http://www.ncbi.nlm.nih.gov/nuccore/358402995), which also contained markers E-In(2L)D-u7 through E-In(2L)D-u13. This scaffold is built of 31 contigs that are stitched together with 30 short stretches (mean = 58.66 bp, mode = 20, min = 20, max = 523) of unknown bases (Ns). *Scr* almost wholly resides in a single contig, ctg7180000390941, which also includes marker E-In(2L)D-u10. *dsx* mapped to scaffold scf7180000395971 (http://www.ncbi.nlm.nih.gov/nuccore/358403364), which also contained marker E-In(2L)D-u16. This scaffold is composed of three contigs with a mean separation of only 20 bp. *dsx* and marker E-In(2L)D-u16 both lie on the same contig, ctg7180000389680. On the linkage map, both scf7180000396708 and scf7180000395971 are located in the distal-most, non-recombining segment of Muller E (2L) corresponding to the inversion in In(2L)D. A similar BLAST analysis shows that genotyping markers linked to the major Muller E QTL are located in a different, more proximal non-recombining region corresponding to the inversion In(2L)M (Table S6).

REFERENCES

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

BIROL, I., S. D. JACKMAN, C. B. NIELSEN, J. Q. QIAN, R. VARHOL *et al.*, 2009 De novo transcriptome assembly with ABySS. Bioinformatics **25:** 2872-2877.

BROMAN, K. W., and Ś. SEN, 2009 *A Guide to QTL Mapping with R/qtl.* Springer, New York.

BROMAN, K. W., H. WU, Ś. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

COCK, P. J. A., C. J. FIELDS, N. GOTO, M. L. HEUER and P. M. RICE, 2010 The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research **38:** 1767-1771.

DARVASI, A., and M. SOLLER, 1992 Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. TAG Theoretical and Applied Genetics **85:** 353-359.

DARVASI, A., A. WEINREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943-951.

DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN, 1977 Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). Journal of the Royal Statistical Society. Series B (Methodological) **39:** 1-38.

EWING, B., and P. GREEN, 1998 Base-Calling of Automated Sequencer Traces UsingPhred.II. Error Probabilities. Genome Research **8:** 186-194.

EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-Calling of Automated Sequencer Traces UsingPhred.I. Accuracy Assessment. Genome Research **8:** 175-185.

HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distances between the loci of linked factors. Journal of Genetics **8:** 299-309.

HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315-324.

HUANG, X., and A. MADAN, 1999 CAP3: A DNA Sequence Assembly Program. Genome Research **9:** 868-877.

KENT, W. J., 2002 BLAT—The BLAST-Like Alignment Tool. Genome Research **12:** 656-664.

KOFLER, R., P. OROZCO-TERWENGEL, N. DE MAIO, R. V. PANDEY, V. NOLTE *et al.*, 2011 PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLoS ONE **6:** e15925.

KOSAMBI, D. D., 1944 The estimation of map distance from recombination values. Ann. Eugen. **12:** 172–175.

LI, R., C. YU, Y. LI, T.-W. LAM, S.-M. YIU *et al.*, 2009 SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics **25:** 1966-1967.

LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantatative Traits*. Sinauer Associates, Inc., Sunderland, Massachusetts.

MILLER, J. R., S. KOREN and G. SUTTON, 2010 Assembly algorithms for next-generation sequencing data. Genomics **95:** 315-327.

REBAI, A., B. GOFFINET and B. MANGIN, 1995 Comparing Power of Different Methods for QTL Detection. Biometrics **51:** 87-99.

ROBERTSON, G., J. SCHEIN, R. CHIU, R. CORBETT, M. FIELD *et al.*, 2010 De novo assembly and analysis of RNA-seq data. Nat Meth **7:** 909-912.

SCHAEFFER, S. W., A. BHUTKAR, B. F. MCALLISTER, M. MATSUDA, L. M. MATZKIN *et al.*, 2008 Polytene Chromosomal Maps of 11 *Drosophila* Species: The Order of Genomic Scaffolds Inferred From Genetic and Physical Maps. Genetics **179:** 1601-1655.

SEN, S., and G. A. CHURCHILL, 2001 A Statistical Framework for Quantitative Trait Mapping. Genetics **159:** 371-387.

SEN, S., J. M. SATAGOPAN, K. W. BROMAN and G. A. CHURCHILL, 2007 R/qtlDesign: inbred line cross experimental design. Mamm Genome **18:** 87-93.

SIMPSON, J. T., K. WONG, S. D. JACKMAN, J. E. SCHEIN, S. J. JONES *et al.*, 2009 ABySS: a parallel assembler for short read sequence data. Genome Res **19:** 1117-1123.

SIMPSON, S. P., 1989 Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. TAG Theoretical and Applied Genetics **77:** 815-819.

SIMPSON, S. P., 1992 Correction: Detection of linkage between quantitative trait loci and restriction fragment length polymorphism using inbred lines. TAG Theoretical and Applied Genetics **85:** 110-111.

STEVISON, L. S., K. B. HOEHN and M. A. NOOR, 2011 Effects of inversions on within- and between-species recombination and divergence. Genome Biol Evol **3:** 830-841.

ZHANG, R., Z. ZHU, H. ZHU, T. NGUYEN, F. YAO *et al.*, 2005 SNP Cutter: a comprehensive tool for SNP PCR-RFLP assay design. Nucleic Acids Res **33:** W489-492.

Tables S1-S9

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.145524/-/DC1.

Table S1 Summary of Illumina sequence data.

Table S2 Summary statistics for the de novo assemblies of the *D. bipectinata* and *D. malerkotliana* transcriptomes.

Table S3 Positions of the 64 fixed differences used as genotyping markers. The columns show marker positions in the following references: *D. bipectinata de novo* transcriptome, *D. malerkotliana de novo* transcriptome, *D. ananassae* 1.3 July 2011 genome (chromosome), *D. ananassae* 1.3 July 2011 transcriptome (gene), *D. ananassae* 1.3 July 2011 coding region (transcript), modENCODE *D. bipectinata* genome (version 1), *D. melanogaster* r5.39 genome (chromosome), *D. melanogaster* r5.39 transcriptome (gene), and *D. melanogaster* r5.39 coding region (transcript). A blank pair of cells indicates that no BLASTn match was found. A vertical bar separating contig positions (for example, "1244|1245") indicates that the FD position in the *de novo* transcript reference query matches to a position inside a gap (in this case, a substring of hyphens "-") in the subject. The two numbers are the positions that flank the gap in the subject.

Table S4 Sequenom primer design output. The columns list the forward, reverse, and extension primers, their masses, and warning codes.

Table S5 The genotypes and phenotypes of all individuals. The top row lists marker names, the second row the Muller element where the marker is located, the third row the position of the marker on the linkage map, and the remaining rows show the genotype of each individual at each marker in the R-QTL CSV input format. "A" denotes homozygosity for the *D. bipectinata* allele, "B" homozygosity for the *D. malerkotliana* allele, and "H" denotes heterozygotes.

Table S6 Locations of the *Scr* and *dsx* genes and the Muller-E genotyping markers in the genome assembly of *D. bipectinata*.

Table S7 Sex comb size in F_1 interspecific hybrids.

Table S8 Chromosome arrangements in the parental strains of *D. malerkotliana* and *D. bipectinata*.

Table S9 List of genes located on the four *D. bipectinata* genome scaffolds that map to In(2L)M. This gene list is tentative because the *D. bipectinata* genome has not been annotated and no gene models are available for this species. To identify *D. bipectinata* genes that are likely to be located inside In(2L)M, we used BLASTp to compare the genome scaffolds that map to this inversion to the annotated transcriptomes of *D. ananassae* and *D. melanogaster*.

Table S7 Sex comb size in F_1 interspecific hybrids

Female parent	Male parent	Number of sex comb teeth (mean ± stdev)	n
D. malerkotliana 14024-0391.00	*D. bipectinata* 14024-0381.03	12.29 ± 1.74	30
D. bipectinata 14024-0381.03	*D. malerkotliana* 14024-0391.00	13.06 ± 1.22	35
D. malerkotliana 14024-0391.00	*D. bipectinata* 14024-0381.04	12.34 ± 1.17	37
D. bipectinata 14024-0381.04	*D. malerkotliana* 14024-0391.00	11.97 ± 1.22	72
D. malerkotliana SWB17	*D. bipectinata* 14024-0381.03	N/A	
D. bipectinata 14024-0381.03	*D. malerkotliana* SWB17	12.69 ± 1.51	28
D. malerkotliana SWB17	*D. bipectinata* 14024-0381.04	10.49 ± 1.35	38
D. bipectinata 14024-0381.04	*D. malerkotliana* SWB17	11.42 ± 1.14	42
D. malerkotliana 14024-0391.00	*D. parabipectinata* 14024-0401.00	10.2 ± 1.14	29
D. parabipectinata 14024-0401.00	*D. malerkotliana* 14024-0391.00	9.54 ± 1.04	34
D. malerkotliana 14024-0391.00	*D. parabipectinata* 14024-0401.02	10.4 ± 0.81	29
D. parabipectinata 14024-0401.02	*D. malerkotliana* 14024-0391.00	9.66 ± 0.92	49
D. malerkotliana SWB17	*D. parabipectinata* 14024-0401.00	10.03 ± 1.13	28
D. parabipectinata 14024-0401.00	*D. malerkotliana* SWB17	8.94 ± 1.09	32
D. malerkotliana SWB17	*D. parabipectinata* 14024-0401.02	N/A	
D. parabipectinata 14024-0401.02	*D. malerkotliana* SWB17	9.44 ± 0.89	33
D. bipectinata 14024-0381.03	*D. parabipectinata* 14024-0401.00	15.51 ± 1.31	78
D. parabipectinata 14024-0401.00	*D. bipectinata* 14024-0381.03	15.13 ± 1.13	38
D. bipectinata 14024-0381.03	*D. parabipectinata* 14024-0401.02	16.04 ± 1.28	71
D. parabipectinata 14024-0401.02	*D. bipectinata* 14024-0381.03	N/A	
D. bipectinata 14024-0381.04	*D. parabipectinata* 14024-0401.00	15.46 ± 1.04	29
D. parabipectinata 14024-0401.00	*D. bipectinata* 14024-0381.04	14.18 ± 1.16	32
D. bipectinata 14024-0381.04	*D. parabipectinata* 14024-0401.02	15.68 ± 1.16	27
D. parabipectinata 14024-0401.02	*D. bipectinata* 14024-0381.04	N/A	

Table S8 Chromosome arrangements in the parental strains of *D. malerkotliana* and *D. bipectinata*

Arm / Muller Element	Strain	Arrangement [a]	Chromosome Order [a]
XL (A)	bip3-isoA	A	**1A-5C**/10D-5C/10D-11D
XL (A)	mal0-sc2	B	1A-6D/11B-6D/**11B-11D**
XR (A)	bip3-isoA	A	**12A-12B**/14B-12B/14B-17D
XR (A)	mal0-sc2	B	12A-13C/16D-13C/**16D-17D**
2L (E)	bip3-isoA	AB+C	**18A**/24A-28D/22D-18A/24A-22D/28D-32D/39B-32D**/39B-45D**
2L (E)	mal0-sc2	A+CQS	18A-22D/28D-22D/28D-31B/36B-36D/34A-36B/31B-32D/39B-36D/**34A-32D/39B-45D**
2R (D)	bip3-isoA	AB	46A-47D/52D-49D/57B-52D/47D-49D/57B-61D
2R (D)	mal0-sc2	AB	**46A-47D/52D-49D/57B-52D/47D-49D/57B-61D**
3L (C)	bip3-isoA	A+CD	**62A-62C/66C-62C/66C-70B/**76A-77D/73C-70B/76A-73C/**77D-80D**
3L (C)	mal0-sc2	A	62A-62C/66C-62C/66C-80D
3R (B)	bip3-isoA	ABC	**81A-83B**/90C-96B/87A-90C/83B-87A/97A-96B/**97A-100D**
3R (B)	mal0-sc2	AL	81A-84A/88B-97A/87A-84A/88B-87A/97A-100D

[a] The names of chromosomal arrangements and chromosome orders refer to the standard polytene chromosome map of the *bipectinata* species complex (Tomimura et al 2005).

Homosequential genomic regions in *mal0-sc2* and *bip3-isoA*:

Chromosome X (Muller A)	1A – 5C; 11B – 12B; 16D – 17D
Chromosome 2 (Muller E + Muller D)	34A – 61D
Chromosome 3 (Muller C + Muller B)	62A – 70B; 77D – 83B; 97A – 100D

Muller E inversions distinguishing *mal0-sc2* and *bip3-isoA*:

In(2L)D: 18A – 28D
In(2L)M: 28D – 34A