Supplementary Text

Results

Phenotypes of intermediate constructs

Homozygous replacement of only the 5' portion of the \textit{D.mel pb} locus with sequences from \textit{D.mim} (fragment 1) resulted in flies with a completely normal labellum, but with notable alterations in maxillary palp morphology and bristle patterning (fig. S4). Maxillary palps of fragment 1 replaced flies were bulbous with increased numbers of tufted bristles in the distal region of the maxillary palps (fig. S4A). Correspondingly, we observed greater numbers of underlying ELAV positive staining neurons with a broader distribution in developing pupal maxillary palps (fig. S4D), a phenotype reminiscent of that observed for the full \textit{pb-mim} replacement (Fig. 2G). Next, we replaced the 3'-most portion of the pb locus (fragment 2) alone (fig. S5, A and B) or in combination with replacement of fragment 1 (fig. S5, C and D). Flies homozygous for fragment 2 only displayed a partial loss of the \textit{pb} homeotic phenotype which consists of a transformation of labellum into arista (fig. S5, A and B). The hypomorphic fragment 2 phenotype dominated when combined with the fragment 1 replacement. In light of the fact that a replacement of the entire \textit{D.mel pb} locus with that of \textit{D.mim} provides nearly full rescue of \textit{pb} function, we speculate that long-range interactions between multiple species-specific cis-acting regulatory sequences are required to act in concert to provide full activity of the locus.

It is noteworthy that the phenotype of the full \textit{D.mim pb} replacement differs from the sum of its component parts in that replacement of only 5' portion of the locus leads to alteration in bristle numbers and underlying neurons (also observed in the full replacement) but does not result in altered orientation of the maxillary palps nor heart-shape palps in males, while replacement of only the 3' portion of the locus results in a partial loss-of-function phenotype. These observations underscore the importance of replacing the entire locus to accommodate potential species-specific evolution of interactions among distant cis-regulatory elements.
Fig. S1: Expression of the *pb*, *Dfd*, and *Scr* loci in third instar larvae. Expression of the *melPb-G4* (A), *Dfd-GAL4* (B) and *Scr-Gal4* (C) genes was visualized by the UAS-nlsGFP reporter gene.
(green) driven by T2A-GAL4 insertions into these loci respectively (Fig. 3A) in labial discs, salivary glands, leg disc T1, leg disc T2, eye-antennal discs and the brain/central nervous system.
Fig. S2: Expression of *pb*, *Dfd*, and *Scr* during metamorphosis. The expression of *melPb-G4* (A), *Dfd-GAL4* (B) and *Scr-Gal4* (C) in developing mouthparts using *UAS-nlsGFP* as the reporter gene (green). Each image represents individual frames from live imaging shown in Supplemental Data Movie 1, 2 and 3. The time the image was recorded is indicated (hr) and the arrowhead indicate the developing labial segment, the arrow indicates the maxillary primordia, and asterisk indicates T1 leg primordia expression. Scale bar, 250 μm.
**Fig. S3: Deletion of the pb gene locus in D.mel** (A) Scheme depicting whole pb locus deletion in D.mel pb, with pb and Zen2 gene introns and exons depicted on top. Also shown are the two gRNAs (gRNA1 and 2) and Cas9 used to delete the 38 kb pb locus and replace it with plasmid carrying an eGFP fluorescent eye marker cassette. (B) Fly heads from a pb wild type (Wt, w1118 strain) and a pb gene locus deletion (pb-/-) fly head. False color indicates maxillary palps (green), and arrowheads indicate labellar lobes. Scale bar, 50 μm.
Fig. S4: Replacement of pb fragment 1 from D.mim to D.mel (A) Scheme depicting the entire pb genomic locus from D.mel (38 kb), along with the replacement DNA fragment 1 (D.mim F1) from D.mim pb with 3xP3-GFP (GFP) marker cassette flanked by lox sites. (B) Bright field images of D.mel and D.mim F1 replaced head. False color indicates maxillary palps (green), and arrowheads indicate labellar lobes. Scale bar, 50 μm. (C) Female and male maxillary palps of D.mel and D.mim F1 replaced. Scale bar, 25 μm. Graph represents number of bristles on maxillary palp in D.mel, and D.mim F1 replaced genotypes. Error bars indicate standard deviation; No. of maxillary palp scored (n)=16; ****p < 0.0001. (D) Female maxillary palp neurons in the D.mel and D.mim F1 replaced animals, as visualized by staining with anti-Elav antibodies. Scale bar, 25 μm. Graph represents number of neurons in maxillary palp in D.mel and D.mim F1 genotypes. Error bars indicate standard deviation; No. of maxillary palp scored (n)=6; ****p < 0.0001.
Fig. S5: Replacement of pb fragment 2 from D.mim to D.mel (A) Scheme depicting whole pb genomic locus from D.mel (38 kb) and a replacement of D.mim fragments 2 (F2) along with 3xP3-RFP (RFP) marker cassette flanked by lox sites. (B) Bright field image of D.mel and D.mim F2
replaced heads. False color indicates maxillary palps (green), and arrowheads indicate labellar lobes. Scale bar, 50 μm. (C) Scheme depicting whole pb genomic locus from D.mel (38kb) and a replacement of both fragments-1 and 2 (D.mim F1-F2) replacement. (D) Bright field image of wild type D.mel and D.mim F1-F2 replaced heads. False color indicates maxillary palps (green), and arrowheads indicate labellar lobes. Scale bar, 50 μm.
Fig. S6: Variation of maxillary palp phenotypes in pb-mim replacement. Array of maxillary palps from *D. melanogaster* and *pb-mim* replaced females (A) and males (B). Asterisks indicate notch on inner surface. Scale bar, 25 μm.
A. Pb protein sequence conservation between *D. melanogaster* and *D. mimica*

B. Pb protein sequence conservation between *D. mimica* and *D. grimshawi*

C. Bar graph showing % protein conservation between different species.

D. Phylogenetic tree showing the relationship between different *Drosophila* species with time scale.

E. Diagram of the *pb* gene with exons, regulatory elements, and rescue experiments.

F. Fluorescent images showing expression patterns of genes in *Drosophila*.
Fig. S7: Comparison of Pb protein coding sequences and function in *D.mel* versus *D.mim* (A)
The Pb protein sequence alignment of *D.mim* compared to *D.mel*. The bar graph represents the conservation percentage of amino acid sequences. (B) The Pb protein sequence alignment of *D.mim* compared to *D.grim*. The bar graph represents the conservation percentage between single amino acid. (C) Pb (green) and zen2 (gray) paralog protein sequence comparison plotting percentage of amino acid conserved between *D.mim* and *D.mel* (*mim-mel*), *D.grim* and *D.mel* (*grim-mel*), and *D.mim* and *D.grim* (*mim-grim*). (D) Phylogenetic tree of Pb protein sequences of six *Drosophila* species. The tree was made using the maximum likelihood method with the CLC Main Workbench program. Bar indicates distance. (E) Schematic of the *pb* genomic regions engineered to incorporate Gal4-T2A fused to HoxA2 or HoxB2 cDNAs followed by 2 stop codons. (F) Images of control (*melPb-G4*) and Gal4-T2A:HoxA2 cDNAs driving expression of *UAS-nlsGFP* in primordia of the maxillary palps (Mx) and proboscis (Lb) in late pupae. The HoxA2 proboscis shape is altered because it has been partially transformed to legs.
Fig. S8: Comparison of genomic pb sequences in *Drosophilids* (A) Genomic sequence conservation between *pb* loci in *D.mim* and *D.mel*. Blue regions indicate exons and red regions indicate conserved non-coding regions. (B) Genomic sequence conservation between *pb* loci in *D.mim* and *D.grim*. Blue regions indicate exons and red regions indicate conserved non-coding regions. (C) Approximately equal mean GFP intensities were measured in the labellum of late pupae for the *melPb-G4* and *mimPb-G4* reporters driving expression of *UAS-nlsGFP*. Error bars indicate standard deviation; No. of pupae scored (n)=20; ns >0.05.
Fig. S9: Ectopic expression of Dfd, dsxM and dsxF in the D.mel pattern (A) Bright field image of melPb-G4, and melPb-G4>UAS-Dfd head. False color indicates maxillary palps (green) and arrowheads indicate labellar lobes. (B) Maxillary palp of melPb-G4>UAS-Dfd. (C) Representative maxillary palp of melPb-G4>UAS-dsxM male and female. (D) Bright field image of melPb-G4>UAS-dsxF fly head. Scale bar, 50 μm.
| Source     | Genotype                                                                 |
|------------|---------------------------------------------------------------------------|
| BDSC:4775  | w[1118]; P{w[+mC]=UAS-GFP.nls}14                                           |
| BDSC:7299  | w[1118]; P{w[+mC]=UAS-Dfd.B}W4                                             |
| BDSC:7302  | w[1118]; P{w[+mC]=UAS-Scr.M}EE2/TM6B, Tb[+]                                  |
| BDSC:44223 | y[1] w*; P{w[+mC]=UAS-dsx.F}24-3                                           |
| BDSC:44224 | y[1] w*; P{w[+mC]=UAS-dsx.M}2/CyO                                           |
| BDSC:7108  | w*; P{w[+mC]=tubP-GAL80[ts]}10; TM2/TM6B, Tb[1]                             |
| In this study | In this study                                                                 |
| In this study | w[1118]; melPb-Gal4                                                        |
| In this study | w[1118]; melPb-Gal4_T2A_cDNA:Pb                                            |
| In this study | w[1118]; melPb-Gal4_T2A_cDNA:Pbmim                                         |
| In this study | w[1118]; melPb-Gal4_T2A_cDNA:HoxA2                                         |
| In this study | w[1118]; melPb-Gal4_T2A_cDNA:HoxB2                                         |
| In this study | w[1118]; mimPb-gal4                                                        |
| In this study | w[1118]; Dfd-Gal4                                                          |
| In this study | w[1118]; Scr-Gal4                                                          |
| In this study | w[1118]; Pbdel                                                            |
| In this study | w[1118]; pb-mim Frag1                                                       |
| In this study | w[1118]; pb-mim Frag2                                                       |
| In this study | w[1118]; pb-mim Frag1-2                                                     |
| In this study | w[1118]; pb-mim Replaced                                                    |
| gRNA       | Sequences                                      |
|------------|------------------------------------------------|
| Del1_gRNA  | GCACCGCGCATGTGCCGACC                           |
| Del2_gRNA  | GACGACTTGGGGTTGTATGA                           |
| F1a_gRNA   | GGAAGGCAGTGCGATTAGCG                           |
| F1b_gRNA   | GATCATCTGATAGGCTACTC                           |
| F2a_gRNA   | GGGGTGGAGAGTGCTCCGGG                           |
| F2b_gRNA   | GACGACTTGGGGTTGTATGA                           |
| F3a_gRNA   | GCAGTGCGATTAGCGTCTG                           |
| F3b_gRNA   | GCAGACTTCTTGCATAGCTT                           |
| pbGal4_gRNA| GCATGAGAAAAGAGCTCATGA                          |
| DfdGal4_gRNA| GCACCGTTCGTCACCAGA                            |
| ScrGal4_gRNA| GTACCAGTTTGCACCTCGC                           |
| Oligos_mimica | Fragment 1 Amplification Oligos | Fragment 2 Amplification Oligos | Fragment 3 Amplification Oligos |
|--------------|--------------------------------|--------------------------------|--------------------------------|
| AA83         | GTGGAAATCACTTTATGATGTGATGACACC | AA181                          | GAGAATAAGCATAATCCAGATCCACCCC   |
| AA319        | GGTGAGACGGGTTGCGGGGTGGAATTTGCTGCC | AA182R                         | CGAAACGAAACTTATCAGCAACTACGTTC |
| AA320        | GGCAAGAATCCTCCACCCCGGCAACCGTCTCAACC | AA246                          | GCTGCTGCACAAAAAGTATCCGACGTAGTGTGCTGATAAGTTTCGGTTTCG |
| AA281        | CTACACATCCGCGCCTAAGTGCTGTG     | AA247                          | CGGCAGATTTTGAAACGAGCAGCTTACCTGGAAGTGGTAGCAAAATGGTAG |
| AA282        | CCAAGAAAGGGACAGCTACACACAGCAC  | AA186                          | CTACCAGTTGCTACCATTCCAGGTTA    |
| AA237        | GATTCCGTACGAGTGTCGGCGCTTGGAATGCTTCACGCATTTATTTGTCG |
| AA238        | CCAAGTCCGAGTGAAAGTGACCAATAAATAGCGTAAGCATTCCAAGCG | AA187                          | CCAAAAGTTAAAAATTITSACGGAGAGCCC |
| AA239        | GTATCTTGGCGCACTTATCTCCATAAATAGCTGAACAAAAATTGCGGC   | AA217                          | CAGACAGACATCTAGTTAGGGCTCTCCTG |
| AA240        | AA241                          | AA120                          | GAGTGGGGTGGGAAAACGTGCAAGATGAATG |
| AA241        | GGCCTGAGCCACTGAAAAGGTCGCGCAATTAATTGTACGCCTTTATATGAG | AA178                          | GTCGGCTACGCACTTATGTAATAAATAGC |
| AA120        | CAGTAGGTTGGACACGTCTACTCTACTCTAGTGACAGTTTCGACACCCCAC | AA191                          | CACTCCGACACTCTGACTTTATGCGCG    |
| AA178        | AATGGGTGGGAAAACGTGCAAGATGAATG | AA192                          | TCCTCAATAGCGCACTTCTGTTCGGCC   |
| AA191        | GTCGGCTACGCACTTATGTAATAAATAGCGC | AA242                          | TAGGCTGACAAAAAGCTTTAAAGCGGAAACAGAGAGTCGGCGCTATGAG |
| AA192        | TCCTCAATAGCGCACTTCTGTTCGGCC   | AA243                          | ATGGCGGACCTTCTCTGCGGGGTGGAATCTGGAATTATGCTTTACTCT   |
| AA244        | GCTTTGCAATCTGGCGAATACGATAAATGCTCAATCCAGATCCAC      |
| Sequencing oligos |  |
|------------------|--|
| AA104            | CAAATGTGTCAGAGCTGAAACGTCATGATTAATG |
| AA106            | CATCATTAATCATGACGTTTCAGCTCTGACAC  |
| AA107            | TCGTCGGGGCAGGCGGTGGTTTCTG          |
| AA107            | GGAAATGTACATCGCATGGATGAAATGCCG    |
| AA108            | GCAGCTGAATTCTGTCCAGAAACCGCCTGCG  |
| AA119            | CATTCAATCTGACAGTTTTACCACCACACTC  |
| AA121            | CCATCAGGGATGCTAAGAGCTGTCATC       |
| AA122            | GATTGACAGCTCTTAGCTAGCAATCCCTGATG  |
| AA123            | CGAATTGGAGTCTGACATTTCACGGG        |
| AA124            | CCCGTGAAATGCAAGACTCCAATTCCG       |
| AA146            | CACGCACATCGAAGAAAATACAG           |
| AA147            | CGCCTTGGTGCGGCGTTGTTC             |
| AA148            | GATTGACAGCTCTTAGCTAGCAATCCCTGATG  |
| AA149            | CCTCATGGGTTACACGACAC              |
| AA150            | GCCACTGCGAGTCAAGTGCC              |
| AA151            | GGCTGCAATTGGTCTCGCAATTGG          |
| AA152            | GCCGTGGAAGTTGAACCTCG              |
| AA152            | GCCGTGGAAGTTGAACCTCG              |
| AA153            | CAACGGGAGTGGAACCTCAACTCC          |
| AA154            | GGAGTTGAAGTCTACACTCCGTTGGG        |
| AA155            | GGATGATACAGGGAGCGGG               |
| AA156            | GTGTTGCCCGGTGGTACGGC              |
| AA157            | GCCAGGCATCAGCATGCAAATCG           |
| AA158            | CTTGGTGCGGCGGCGGCTG               |
| AA159            | GGGATCTCCGGACTCAAAGTG             |
| AA160            | CGACTCGGATGTTGGTGCTG              |
| AA160            | CGACTCGGATGTTGGTGCTG              |
| AA161            | CGTGCTGAAATGGAAGTCAATGC           |
| AA176            | CGAGTTGGATCGACATCCCACGG          |
| AA177            | CGGCAGACAAAGTGCTGTC              |
| AA180            | GGGGTGAGATCGGATGATGCTTATTCCT     |
| AA180            | GGGGTGAGATCGGATGATGCTTATTCCT     |
| AA181            | GAGAATAAGCATAATCCAGATCCACCCC     |
|   |   |
|---|---|
| AA184 | GACGTA-GTTGCTGATAAGTTTCGTTTGC |
| AA185 | TAACCTGGGAGTGTAGCAAAAATGGTAG |
| AA189 | GCATGGGCTTGGTGAAGATCGCAACAA |
| AA190 | TGCAAAATCTTCTAGTCCAGGCAACCAGGC |
| AA193 | GCCGAACAGAAATGCTCCTCTTACG |
| AA194 | TAAGGCCCTCTGGTCCATAAGACCTCTCCCC |
| AA197 | AAGACTCGTATACCCCTGTGG |
| AA198 | GCAGCTGTTGGCCGGTGTGCTG |
| AA199 | GCACAGCAGCCAATTCAACA |
| AA200 | CACATGCACCTGGACATGGGC |
| AA201 | ATTGATGATGCGGCTTTGGG |
| AA202 | CCCATAAGCTGCCTAATGGC |
| AA202 | CCCATAAGCTGCCTAATGGC |
| AA203 | GGAGGGTCCTCCGAACAGAG |
| AA203 | GGAGGGTCCTCCGAACAGAG |
| AA204 | CATCTCTGGCGCATTCCTAGG |
| AA205 | GCGAGCTTTGCTGGTCCTCACG |
| AA206 | CAGTGCTCTATGTCTGGCTG |
| AA208 | CGGATGATGCAACGATCCAGCAAG |
| AA209 | CCCGTCTACTTCAGCGCAGACTG |
| AA209 | CCCGTCTACTTCAGCGCAGACTG |
| AA211 | CCGCTCTTTGGACTCGGTCTG |
| AA212 | GCCAAGTTGGAAAGGCATTGG |
| AA213 | CATTTTGCTGGCGCAGACTG |
| AA214 | GCTGCCACGGTCACAGGACT |
| AA215 | GCTGCCACCGTCACAGGACT |
| AA216 | GCCTCTTCCATCTCTGGCTG |
| AA216 | GCCTCTTCCATCTCTGGCTG |
| AA230 | GTACACACTTTAGTAAGTGGTCCACCTAAAAACTCTACGCTTGC |
| AA231 | CTGGTGGCCTGCAACTAAATGTCGACAAACTTTAGT |
| AA264 | CATGGGACATTGAGCAGTTG |
| AA264 | CATGGGACATTGAGCAGTTG |
| AA276 | ACTCGGTGGCCAGTATCCCTGGATG |
| AA281 | GCAGCGGTGTGGCTCAAATAATGGC |
| AA323 | AGACCGAGCTGTCCATCAAGGAGGCCAAC |
| AA348 | GCAGCTCGACATGATGGAATGGAAATGGGACTG |
| AA434 | TGGGTTCAGCAACAGCACC |
Table S5: Maxillary palp Normalized GFP intensity

| Pbmel_F   | Pbmim_F   | Pbmel_M   | Pbmim_M   |
|-----------|-----------|-----------|-----------|
| 46.19951  | 51.31782  | 59.16485  | 48.7396   |
| 42.48317  | 44.69703  | 67.46287  | 52.63787  |
| 69.6953   | 60.86609  | 58.07277  | 38.97624  |
| 67.52748  | 46.04604  | 44.36089  | 33.70371  |
| 63.08936  | 53.90941  | 58.88911  | 32.82252  |
| 77.16411  | 55        | 54.98317  | 42.98094  |
| 66.83144  | 36.79876  | 45.97104  | 41.58614  |
| 65.5349   | 34.76807  | 54.54604  | 43.39059  |
| 51.55545  | 58.85272  | 59.16485  | 46.3896   |
| 54.95545  | 30.77401  | 65.46287  | 40.86262  |
| 44.06436  | 40.80743  | 58.07277  | 38.77327  |
| 51.66609  | 36.10767  | 44.36089  | 42.00965  |
| 60.19233  | 79.36386  | 58.88911  | 28.69183  |
| 68.6255   | 61.24233  | 54.98317  | 38.41287  |
| 50.40866  | 45.79802  | 45.97104  | 54.75718  |
| 42.56436  | 53.11238  | 54.54604  | 31.35272  |
| 43.57302  | 46.04134  | 47.55495  | 35.81188  |
| 48.44505  | 60.31757  | 49.48812  | 38.11634  |
| 52.53342  | 51.31782  | 55.81436  | 34.64653  |
| 68.99381  | 44.69703  | 42.56436  | 32.52401  |
| 47.55495  | 60.86609  | 43.57302  | 37.31114  |
| 49.48812  | 46.04604  | 48.44505  | 32.71708  |
| 55.81436  | 53.90941  | 54.98317  |          |
| 62.02946  | 54.98317  |          |          |
| 58.61906  | 36.79876  |          |          |
| 57.33861  | 34.76807  |          |          |
| 49.64183  | 58.85272  |          |          |
| 56.0255   | 40.80743  |          |          |
| 53.32129  | 36.10767  |          |          |
| 54.34158  | 79.36386  |          |          |
Table S6: Maxillary palp raw GFP intensity

| Pbmel_F | Pbmim_F | Pbmel_M | Pbmim_M |
|---------|---------|---------|---------|
| 186.646 | 207.324 | 239.026 | 196.908 |
| 171.632 | 180.576 | 272.55  | 212.657 |
| 281.569 | 245.899 | 234.614 | 157.464 |
| 272.811 | 186.026 | 179.218 | 136.163 |
| 254.881 | 217.794 | 237.912 | 132.603 |
| 311.743 | 222.2  | 222.132 | 173.643 |
| 269.999 | 148.667 | 185.723 | 168.008 |
| 264.761 | 140.463 | 220.366 | 175.298 |
| 208.284 | 237.765 | 239.026 | 187.414 |
| 222.02  | 124.327 | 272.55  | 165.085 |
| 178.02  | 164.862 | 234.614 | 156.644 |
| 208.731 | 145.875 | 179.218 | 169.719 |
| 243.177 | 320.63  | 237.912 | 115.915 |
| 277.247 | 247.419 | 222.132 | 155.188 |
| 203.651 | 185.024 | 185.723 | 221.219 |
| 171.96  | 214.574 | 220.366 | 126.665 |
| 176.035 | 186.007 | 192.122 | 144.68 |
| 195.718 | 243.683 | 199.932 | 153.99 |
| 212.235 | 207.324 | 225.49  | 139.972 |
| 278.735 | 180.576 | 171.96  | 131.397 |
| 192.122 | 245.899 | 176.035 | 150.737 |
| 199.932 | 186.026 | 195.718 | 132.177 |
| 225.49  | 217.794 |         |         |
| 250.599 | 222.2  |         |         |
| 236.821 | 148.667 |         |         |
| 231.648 | 140.463 |         |         |
| 200.553 | 237.765 |         |         |
| 226.343 | 124.327 |         |         |
| 252.978 | 164.862 |         |         |
| 215.418 | 145.875 |         |         |
| 250.038 | 320.63  |         |         |
Table S7: Labellum GFP intensity

| Pbmel  | Pbmim   |
|--------|---------|
| 397.567| 433.361 |
| 401.159| 384.79  |
| 392.197| 474.586 |
| 381.725| 289.157 |
| 404.983| 385.276 |
| 372.767| 426.509 |
| 370.844| 372.861 |
| 400.584| 474.345 |
| 384.702| 345.47  |
| 539.477| 324.968 |
| 397.567| 377.983 |
| 401.159| 417.626 |
| 392.197| 416.139 |
| 381.725| 404.574 |
| 404.983| 308.507 |
| 372.767| 349.938 |
| 370.844| 350.712 |
| 400.584| 464.522 |
| 384.702| 557.656 |
|        | 462.451 |
Movie S1: The expression of melPb-GAL4 during metamorphosis.

Movie S2: The expression of Dfd-GAL4 during metamorphosis.

Movie S3: The expression of Scr-GAL4 during metamorphosis.

Data S1: *D. mim* zen2 and Pb protein and DNA sequences