



Title	Cis-regulatory evolution underlying the changes in wingless expression pattern associated with wing pigmentation of <i>Drosophila</i>
Author(s)	Karasawa, Takumi; Saito, Namiho; Koshikawa, Shigeyuki
Citation	FEBS Letters, 597(14), 1837-1847 https://doi.org/10.1002/1873-3468.14637
Issue Date	2023-07
Doc URL	https://hdl.handle.net/2115/92895
Rights	This is the peer reviewed version of the following article: Takumi Karasawa, Namiho Saito, Shigeyuki Koshikawa. (2023) "Cis-regulatory evolution underlying the changes in wingless expression pattern associated with wing pigmentation of <i>Drosophila</i> ", FEBS Letters 597(14) 1837-1847, which has been published in final form at https://doi.org/10.1002/1873-3468.14637 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.
Type	journal article
File Information	20230510CVT_Karasawa_all.pdf



Cis*-regulatory evolution underlying the changes in *wingless* expression pattern associated with wing pigmentation of *Drosophila

Takumi Karasawa¹, Namiho Saito², Shigeyuki Koshikawa^{1,3}

¹ Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan

² School of Science, Hokkaido University, Sapporo, Japan

³ Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan

Corresponding author:

Shigeyuki Koshikawa

Address: Faculty of Environmental Earth Science, Hokkaido University, N10W5, Kita-ku, Sapporo, Hokkaido 060-0810, Japan

E-mail: koshi@ees.hokudai.ac.jp

Tel: +81-11-706-2236

Running title:

Cis-regulatory evolution of pigmentation pattern

Abbreviations:

EGFP: Enhanced Green Fluorescent Protein

Mad: Mothers against Decapentaplegic

BMP-AE: Bone Morphogenetic Protein-Activating Element

Keywords:

fruit fly, *cis*-regulatory evolution, gene regulatory network, CRE, Evo-Devo

Conflict of interest:

The authors declare no conflict of interest.

Abstract

The co-option of regulatory genes has the potential to play an important role in the evolutionary gain of new traits. However, the changes at the sequence level that underlie such a co-option event are still elusive. We identified the changes in the *cis*-regulatory sequence of *wingless* that caused co-option of *wingless* and led to its expression in new places in *Drosophila guttifer*, which has unique pigmentation patterns on its wings. The newly gained function of gene expression activation was acquired evolutionarily via a combination of pre-existing sequences containing a putative binding site for SMAD transcription factors that exhibit an ancestral function in driving expression at crossveins, and a sequence that is specific to the lineage leading to *D. guttifer*.

Introduction

Evolution has produced various traits throughout the history of life. In evolutionary developmental biology, genetic mechanisms underlying the gain of new traits have been discussed [1-3], and in particular, the importance of gene network co-option has been examined [4, 5]. In gene network co-option, a regulatory factor is redeployed in a novel cellular context, and genes in the network controlled by this regulatory factor are also made to function in this context. Thus, changes to a few genes can potentially alter the expression of many downstream genes. Some examples in which gene network co-option may have played an important role in the gain of new morphological traits have been reported [6-10]. However, we do not have a good understanding of how the co-option that leads to the acquisition of traits occurs at the sequence level.

The wing pigmentation pattern of *Drosophila* species is one of the traits that have been studied to reveal the genetic changes underlying the acquisition of traits [11-14]. For example, in *Drosophila biarmipes*, the co-option of *Distal-less (Dll)*, a gene that also functions in the patterning of the wing margin, contributes to the gain of the male-specific spot on the wing [15]. *Drosophila guttifer* has a unique pigmentation pattern on its wing, and the pigmentation is regulated by *wingless (wg)*, a gene which activates the expression of pigmentation genes such as *yellow* (Fig. 1A) [16, 17]. In *D. guttifer*, *wg* is expressed in the crossveins, longitudinal vein tips, and campaniform sensilla of the pupal wing, prefiguring the location of pigmentation [18, 19]. In *Drosophila melanogaster*, on the other hand, *wg* is expressed in the crossveins, but not in the longitudinal vein tips or campaniform sensilla. Furthermore, in fruit flies of the quinnia group, to which *D. guttifer* is closely related [20], the expression pattern of *wg* is similar to that of *D. melanogaster*, suggesting that the acquisition of ectopic expression occurred in the lineage after the divergence of *D. guttifer* from its closely related species [18]. Thus, the co-option of *wg* that leads to the expression at the future pigmentation sites was one of the key events in the gain of the polka-dotted pigmentation pattern in *D. guttifer*.

A 760 bp *cis*-regulatory region in the vicinity of *wg* that drives expression in the crossveins

and longitudinal vein tips (gutCVT-core) was identified in *D. guttifera* (Fig. 1C, E). The corresponding region of *D. melanogaster* (melCV-core and its larger version melCV) drives expression only in the crossveins (Fig. 1D, E). gutCVT-core maintains its function of regulating gene expression when introduced into *D. melanogaster*, indicating that the new expression pattern of *wg* should have evolved via the changes in this *cis*-regulatory region rather than via changes in the *trans*-regulatory factors (Fig. 1C) [18, 21]. Then, what specific changes in the sequence were responsible for gaining the new expression pattern?

In this research, we altered the sequences of the cloned *cis*-regulatory regions and examined their regulatory function by performing transgenic EGFP reporter assays. By deleting parts of gutCVT-core, we identified regions essential for the regulatory function and showed that they consist of regions needed to activate expression in the entire wing veins and a region required for repressing expression in certain parts of wing veins. In addition, by investigating the regulatory function of chimeric sequences with the homologous sequence of *D. melanogaster*, we revealed that some of the regions required for the regulation of the expression possessed their functions before the *cis*-regulatory evolution. Knockout of the potential binding sites of transcription factors revealed that a sequence resembling the SMAD binding site is necessary for activating *wg* expression and that this binding site also exists in the homologous region of *D. melanogaster*. These results suggest that the functions of pre-existing sequences and newly gained sequences were combined to acquire a new expression pattern of *wg*.

Results

Identification of the regions required for activating and repressing expression

We reasoned that the newly gained sequences that brought about the novel expression pattern of *wg* should be included among sequences required for regulating expression. Thus, we first narrowed down the location of the sequences necessary for the regulatory function. gutCVT-core and melCV-core have several perfectly matched stretches of sequence (anchor-points in GenePalette function, 14–43 bp), and we divided these *cis*-regulatory regions into five regions bordered by anchor-points (Fig. 1E). We confirmed the colinearity and homology of these anchor-points by comparison with the sequences of other closely related species. The function of each region flanked by anchor-points is not known, but each such region is thought to have evolved while maintaining its position in the spatial order (Fig. S1). However, the possibility that very small blocks of a few base pairs may have moved between the regions by rearrangement has not been completely ruled out. *Cis*-regulatory sequences that lack one or some of these regions were then cloned next to an EGFP reporter gene. By integrating these constructs into a specific site in the genome of *D. melanogaster*, we examined their regulatory function in developing pupal wings (Fig. 2A). While complete

gutCVT-core drove reporter EGFP expression in crossveins and longitudinal vein tips in pupal wings, gut2-5, which lacks region 1, activated expression in entire wing veins (Fig. 1C, 2B). gut3-5, lacking both regions 1 and 2, did not drive expression (Fig. 2C). These results indicate that gutCVT-core requires region 1 for repressing expression and region 2 for activating expression. gut1-4, which lacks region 5, drove EGFP expression in the crossveins and vein tips, which is very similar to the expression driven by complete gutCVT-core (Fig. 2D). gut1-3, lacking both regions 4 and 5, activated expression only in the crossveins, and this EGFP expression pattern was similar to the one driven by melCV-core, although the fluorescence appeared weak (Fig. 2E). These results suggest that region 4 has an important role in activating expression, while region 5 is not necessary. Moreover, gut2-3, which lacks regions 1, 4, and 5, did not drive expression (Fig. 2F). This result indicates that region 1 has a role to activate expression in the crossveins.

Similarly, to examine the regulatory function of the corresponding regions in melCV-core, we tested mel1-4, which lacks region 5, and mel2-4, which lacks regions 1 and 5. Similar to the complete melCV-core, mel1-4 activated expression in the crossveins (Fig. 2G). Thus, region 5 was also unnecessary for the function of melCV-core. In contrast, mel2-4 failed to activate the expression (Fig. 2H). Thus, it was shown that region 1 has an important role in driving expression in melCV-core while region 2-4 alone cannot activate expression in veins.

Comparison of the regulatory function with that of a homologous sequence

The function of activating expression was not harbored in the same way in gutCVT-core and melCV-core. The activation function of region 2-4 was only possessed by gutCVT-core, which raised the question, is the repressive function of region 1 only possessed by gutCVT-core? To test whether the repressive function of region 1 was newly gained in the lineage leading to *D. guttifera*, we examined the regulatory function of mel1gut2-5, a chimeric sequence in which the region 1 of gutCVT-core is replaced with the corresponding sequence of melCV-core. mel1gut2-5 showed a fluorescence pattern similar to that shown by gutCVT-core, with expression repressed in longitudinal veins except for the vein tips (Fig. 3A). Thus, the homologous sequence of *D. melanogaster* also has the repressive function of region 1.

In addition, since we could not confirm the necessity for region 3 in the previous experiment, we also tested the function of mel3gut1-2+4, a chimeric sequence in which region 3 is replaced with the corresponding sequence of melCV-core. This *cis*-regulatory sequence also has a regulatory function similar to that of gutCVT-core, whereby EGFP was expressed in the crossveins and longitudinal vein tips (Fig. 3B). This result indicates that changes in region 3 did not play a critical role in the *cis*-regulatory evolution that brought about the new expression pattern.

Knockout of SMAD binding sites

Since the most important sequences for the regulatory function of a *cis*-regulatory region are binding sites of transcription factors, we focused on transcription factors expressed in the pupal wing. One of them, Mothers against Decapentaplegic (Mad), is a transcription factor required for Decapentaplegic signaling readout, and phosphorylated Mad (pMad) forms a complex with Medea, also a transcription factor, and binds to specific sequences to activate target gene expression [22-26]. The localization of pMad in the pupal wings is similar to the pattern driven by gutCVT-core [27, 28].

We hypothesized that Mad is the transcription factor that binds to gutCVT-core. We searched for candidate binding sites of Mad-Medea complex, BMP-AE (GGCGCCANNNGNCV) [29], and candidate binding sites of Mad, SMAD consensus (GCCGNCGC) [22], in the *cis*-regulatory regions. In gutCVT-core, these binding sequences or very similar sequences are present, and one sequence similar to BMP-AE (GGGGCCAATGTGTCG) (hereinafter called BMP-AE for simplicity) is present in region 5, two sequences similar to SMAD consensus (GCCGNCGC or 1 bp change of it) (hereinafter called SMAD consensus for simplicity) are present in region 1 (CCCGCCGC and GCCGCCAC), one in region 3 (GCAGACGC), and four in region 4 (GCAGCCGC, GCCGCCGC, GCCGCCTC and GCCGCCTC) with overlap (Fig. 1E, 4A). To test the necessity of these putative binding sites for regulating expression, we knocked out these sequences and examined their regulatory function (Fig. 4A). As predicted from our result with gut1-4, BMP-AE KO possessed regulatory function very similar to that of gutCVT-core (Fig. 4B). On the other hand, gutSMAD bs KO, lacking all SMAD consensus, failed to activate expression in the crossveins and vein tips, and EGFP was weakly expressed only around the crossveins and a part of the longitudinal vein between two crossveins (a region where *wg* expression was not shown by *in situ* hybridization) (Fig. 4C). This demonstrates that SMAD consensus is essential for activating expression. To identify which of the multiple SMAD consensus is necessary for activating expression, we recovered SMAD consensus in regions 3 and 4, which can be involved in promoting expression, respectively, and tested their regulatory function. gutSMAD bs3 Rec, in which only SMAD consensus in region 3 was recovered, regained its function of promoting expression and drove EGFP expression in the crossveins and longitudinal vein tips (Fig. 4D). On the other hand, gutSMAD bs4 Rec, in which only SMAD consensus in region 4 were recovered, failed to activate the expression and showed a non-specific EGFP expression pattern similar to gutSMAD bs KO (Fig. 4E). Thus, the only SMAD consensus required for regulating expression in gutCVT-core is the one in region 3.

While no BMP-AE was found in melCV-core, two sequences very similar to SMAD consensus were present. One was in region 2 (GACGACGC), and the other was in region 3, and was completely identical to the sequence in gutCVT-core (GCAGACGC, Fig. 1E, 4A). To determine whether SMAD consensus is essential in melCV-core, we knocked out these sequences (melSMAD bs KO) (Fig. 4A). Deletion of SMAD consensus resulted in the loss of the EGFP expression in the

crossveins, indicating that the SMAD consensus is required for melCV-core to activate expression (Fig. 4F).

Discussion

New expression pattern brought about by the gain of enhancer activity in the entire wing veins

Although gutCVT-core expresses *wg* in the crossveins and longitudinal vein tips, this expression pattern did not simply result from the activated expression in these parts of wing veins, but from a combination of activation in the entire wing veins by region 2–4 and repression in excess parts of longitudinal veins by region 1 (Fig. 5).

In addition, the function of region 1 of gutCVT-core to repress expression throughout the longitudinal veins except the vein tips was also present in the homologous region of melCV-core, suggesting that a silencer was present there in both species (Fig. 2B, 3A). The function of activating expression in the entire wing veins possessed by region 2–4 was not possessed by melCV-core, and was shown to be newly gained in gutCVT-core (Fig. 2B, H). Thus, it was revealed that the new expression of *wg* in the longitudinal vein tips in *D. guttifera* was not simply brought about by the gain of activation in these parts, but rather the activation in the entire wing veins combined with the repression in excess parts of the veins.

One interesting point is that melCV-core possesses the repressive function in region 1, even though the *D. melanogaster* region 2–4 has no function to activate gene expression in the entire wing veins. The role of the sequence involved in repressing expression in melCV-core is still unknown, but one possibility is that this sequence functions in developmental stages other than pupae or in other tissues in *D. melanogaster*. *Cis*-regulatory sequences sometimes change the place where or the developmental timing when they work over the course of evolution [30]. We can consider several other possibilities. For example, it is possible that *D. melanogaster* has a function of activation in longitudinal veins in regions outside the melCV-core, and region 1 represses it. It is also possible that region 1 has no function in *D. melanogaster*, but only in combination with gut2-4 it acquires a function of repression. Alternatively, an ancestor of *D. melanogaster* might have had activating function in region 2–4 and even pigmentation on veins, like *D. guttifera*, and lost them secondarily. These possibilities are not mutually exclusive, but the actual evolutionary pathway may be a composite of them.

The role of the pre-existing SMAD consensus in the gain of the new regulatory function of *wg* expression

SMAD consensus in region 3 was required for activating expression in the entire wing veins (Fig. 4). However, this does not mean that the gain of the SMAD consensus brought about the

ectopic expression of *wg*. The SMAD consensus required for promoting expression is also conserved in *melCV*-core, and the chimeric sequence in which region 3 was replaced with a homologous sequence of *melCV*-core drove the same expression pattern as full *gutCVT*-core (Fig. 3B). Furthermore, also in *melCV*-core, the deletion of all SMAD consensus causes the loss of gene expression in the crossveins, suggesting that the SMAD consensus in region 3, which is conserved between the two species, is necessary for activating expression by *melCV*-core (Fig. 4F). Thus, we propose that the SMAD consensus in ancestral *cis*-regulatory regions coordinates with some newly gained sequences to cause acquisition of *wg* expression in the new parts of the wing veins (Fig. 5).

The causal sequence change in *gutCVT*-core responsible for the evolution of the *wg* expression pattern has not yet been identified at nucleotide level resolution. That sequence is presumed to coordinate with SMAD consensus to activate expression in the entire wing veins. Therefore, newly gained elements may interact with SMAD proteins. For example, when pMad works to activate transcription, it makes a complex with Medea and the complex binds to BMP-AE, but the SMAD consensus is shorter than BMP-AE and does not contain the binding sites of both [22-26]. There was one Medea consensus on the SMAD consensus in region 3, but they are unlikely to work together because the binding consensus were overlapped. The Medea consensus is short (GTCTG, [31]) and similar sequences are abundant within *gutCVT*-core; thus, Medea may bind to some of them. Some factors are known to interact with pMad, for example, the protein CBP/p300, which works as a transcriptional co-activator [32, 33]. Such Mad-related factors may provide clues for identifying newly gained regulatory sequences in *gutCVT*-core.

Acquisition of the new expression pattern by the combination of pre-existing and newly gained *cis*-elements

In the evolution of the *wg* expression pattern associated with the wing pigmentation pattern formation of *D. guttifera*, at least two pre-existing *cis*-regulatory sequences were co-opted. One is the sequence in region 1 to repress expression throughout the longitudinal veins except for the vein tips, and the other is the SMAD consensus in region 3 required for activating expression in the entire wing veins. These pre-existing *cis*-regulatory sequences function together with newly gained *cis*-regulatory sequences in *gutCVT*-core, resulting in the new expression pattern of *wg* (Fig. 5). This means that not all of the sequences required for the new expression pattern were newly gained in the co-option of *wg*. Similar to this case, there are several examples of new expression patterns evolved by the modification of pre-existing enhancers [34-36].

Combinations of multiple control functions could possibly produce diverse expression patterns. Such *cis*-regulatory regions would require many regulatory sequences, and therefore the probability of acquiring a new one from a non-functional sequence by mutation seems intuitively low. The frequency of *de novo* generation of enhancers is still unclear and being discussed [37].

Alternatively, by reusing preexisting regulatory sequences, fewer sequences are required to gain new expression patterns. Co-option of pre-existing sequences is one of the main mechanisms that cause the gain of ectopic expression of a gene, and if it occurs on important regulatory genes, as shown in this study, it may lead to gene network co-option [38].

Materials and Methods

Prediction of binding sites in *cis*-regulatory regions

To identify putative binding sites, we used GenePalette 2.1 [39]. To find candidates for BMP-AE, we searched for GGCGCCANNNGNCV [29], allowing a 1 bp mismatch, in gutCVT-core and melCV-core. In the same way, we found the putative SMAD consensus by searching for GCCGNCGC [22], allowing a 1bp mismatch, in the two *cis*-regulatory regions.

Examination of the function of modified *cis*-regulatory sequences by EGFP reporter assay

Within the scope of this study, it is assumed that the expression patterns driven by the *cis*-regulatory regions tested in *D. melanogaster* and *D. guttifera* are almost the same, and therefore the composition of the *trans*-regulatory factors involved is also almost the same. This assumption is based on a previous study showing the same expression pattern (crossveins and longitudinal vein tips) of gutCVT in *D. melanogaster* as in *D. guttifera* [18]. Some of the expression of the reporter in locations where *wingless* expression would not normally be observed, such as ectopic expression at the fourth longitudinal vein, may be due to this difference in host species.

We first amplified *cis*-regulatory sequences to test the regulatory function by PCR, using gutCVT-core or melCV-core [18] as templates. To synthesize sequences that lack some of the regions, PCR was conducted using primers based on sequences from the borders of the divided regions. To obtain chimeric sequences, we made DNA fragments of the respective regions of gutCVT-core or melCV-core and joined them together by overlap extension PCR. To obtain sequences in which putative binding sites were knocked out, we first performed PCR using primers that contain sequences of about 10 bases before and after the sequence to be knocked out, with the putative binding sites changed to random sequences (in the case of BMP-AE KO, all bases were changed to A). These were then joined together by overlap extension PCR. By using PROMO [40, 41], we confirmed that no important binding site was newly created. All primers used in this research are shown in Supporting Information (Table S1, S2).

Amplified sequences were then inserted into S3aG [42] with the 5' end oriented closer to the EGFP sequence and the constructs were then introduced into *D. melanogaster*, VK00006 (attP on cytogenetic location 19E7) [43] using PhiC31 integrase-mediated transgenesis systems (BestGene, CA). Transgenic flies were then crossed to be homozygous for the introduced sequence and

maintained at 25 °C with standard cornmeal/sugar/yeast/agar food [44].

To observe EGFP reporter expression, wandering larvae were collected and then incubated at 25 °C for 40–44 hours, and P6-P7 pupae were obtained [45]. In these stages, pupal wings have not yet folded, and wing veins are visible. Obtained pupae were taken out from the puparium, had their fat and organs removed in 1 x PBS, and were fixed in 1% paraformaldehyde at 4 °C overnight. Fixed wings were taken out from pupal cuticles in 1 x PBS and observed in the fluorescence field using a BX60 microscope (Olympus) [46]. To examine the EGFP expression pattern in wings, we took photographs under the same conditions of light intensity and exposure. The brightness and contrast of images were adjusted with ImageJ [47].

Fluorescence patterns were confirmed in one line each for BMP-AE KO, gut1-3, and gutSMAD bs3 Rec, and in multiple lines for each of the other constructs.

Author contributions

TK, NS, and SK conceived and designed the study. TK and NS collected data and conducted the analysis. TK and SK wrote the manuscript. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Acknowledgements

We thank Sean B. Carroll for providing template DNA; Elizabeth Nakajima for English editing; Masato Koseki for technical advice; and Haruka Takahashi, Masato Koseki, Satoshi Hayakawa, Shunsuke Otake, Sota Sakai, Takuma Niida, Yusuke Yoshida and Yuto Terashima for their discussions and comments. This work was funded by KAKENHI (18H02486) to SK.

Data accessibility

The data that supports the findings of this study are available in [Fig. 1, Fig. 2, Fig. 3, Fig. 4 and Fig. 5] and the supplementary material of this article.

References

1. Wagner GP & Lynch VJ (2010) Evolutionary novelties. *Current Biology* 20, R48-R52.
2. True JR & Carroll SB (2002) Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18, 53-80.
3. Shubin N, Tabin C & Carroll S (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.

4. Monteiro A, & Podlaha O (2009) Wings, horns, and butterfly eyespots: how do complex traits evolve?. *PLoS Biol* 7, e1000037.
5. McQueen E & Rebeiz M (2020) On the specificity of gene regulatory networks: How does network co-option affect subsequent evolution? *Curr Top Dev Biol* 139, 375-405.
6. Keys DN, Lewis DL, Selegue JE, Pearson BJ, Goodrich LV, Johnson RL, Gates J, Scott MP & Carroll SB (1999) Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. *Science* 283, 532-534.
7. Gao F & Davidson EH (2008) Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proc Natl Acad Sci USA* 105, 6091-6096.
8. Lemons D, Fritzenwanker H, Gerhart J, Lowe CJ & McGinnis W (2010) Co-option of an anteroposterior head axis patterning system for proximodistal patterning of appendages in early bilaterian evolution. *Dev Biol.* 344, 358-362.
9. Nakayama H, Yamaguchi T & Tsukaya H (2012) Acquisition and diversification of cladodes: leaf-like organs in the genus *Asparagus*. *Plant Cell.* 24, 929-940. (doi: 10.1105/tpc.111.092924)
10. Murugesan SN, Connahs H, Matsuoka Y, Das Gupta, M, Tiong GJ, Huq M, Gowri V, Monroe S, Deem KD, Werner T, Tomoyasu Y & Monteiro A (2022). Butterfly eyespots evolved via cooption of an ancestral gene-regulatory network that also patterns antennae, legs, and wings. *Proc Natl Acad Sci USA* 119, e2108661119.
11. Gompel N, Prud'homme B, Wittkopp PJ & Carroll SB (2005) Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481-487.
12. Le Poul Y, Xin Y, Ling L, Mühling B, Jaenichen R, Hörl D, Bunk D, Harz H, Leonhardt H, Wang Y, Osipova E, Museridze M, Dharmadhikari D, Murphy E, Rohs R, Preibisch S, Prud'homme B & Gompel N (2020) Regulatory encoding of quantitative variation in spatial activity of a *Drosophila* enhancer. *Sci Adv* 6, eabe2955.
13. Raja KK, Bachman EA, Fernholz CE, Trine DS, Hobmeier RE, Maki NJ, Massoglia TJ, Werner T (2023) The genetic mechanisms underlying the concerted expression of the *yellow* and *tan* genes in complex patterns on the abdomen and wings of *Drosophila guttifera*. *Genes* 14, 304.

14. Koshikawa S (2020) Evolution of wing pigmentation in *Drosophila*: Diversity, physiological regulation, and *cis*-regulatory evolution. *Dev Growth Diff* 92, 269-278.
15. Arnoult L, Su KFY, Manoel D, Minervino C, Magriña J, Gompel N & Prud'homme B (2013) Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* 339, 1423-1426.
16. Werner T, Koshikawa S, Williams TM & Carroll SB (2010) Generation of a novel wing colour pattern by the Wingless morphogen. *Nature* 464, 1143-1148.
17. Fukutomi Y, Kondo S, Toyoda A, Shigenobu S & Koshikawa S (2021) Transcriptome analysis reveals *wingless* regulates neural development and signaling genes in the region of wing pigmentation of a polka-dotted fruit fly. *FEBS J* 288, 115-126.
18. Koshikawa S, Giorgianni MW, Vaccaro K, Kassner VA, Yoder JH, Werner T & Carroll SB (2015) Gain of *cis*-regulatory activities underlies novel domains of *wingless* gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 112, 7524–7529.
19. Koseki M, Tanaka NK & Koshikawa S (2021) The color pattern inducing gene *wingless* is expressed in specific cell types of campaniform sensilla of a polka-dotted fruit fly, *Drosophila guttifera*. *Dev Genes Evol* 231, 85-93.
20. Chialvo CHS, White BE, Reed LK & Dyer KA (2019) A phylogenetic examination of host use evolution in the *quinaria* and *testacea* groups of *Drosophila*. *Mol Phylogenet Evol* 130, 233-243.
21. Koshikawa S (2015) Enhancer modularity and the evolution of new traits. *Fly* 9, 155-159.
22. Kim J, Johnson K, Chen HJ, Carroll S & Laughon A (1997) *Drosophila* Mad binds to DNA and directly mediates activation of *vestigial* by Decapentaplegic. *Nature* 388, 304-308.
23. Hudson JB, Podos SD, Keith K, Simpson SL & Ferguson EL (1998) The *Drosophila* *Medea* gene is required downstream of *dpp* and encodes a functional homolog of human Smad4. *Development* 125, 1407-1420.
24. Affolter M, Marty T, Vigano MA & Jazwińska A (2001) Nuclear interpretation of Dpp signaling

in *Drosophila*. *EMBO J* 20, 3298-3305.

25. Raftery LA & Sutherland DJ (1999) TGF-beta family signal transduction in *Drosophila* development: from *Mad* to Smads. *Dev Biol* 210, 251-268.

26. Hamaratoglu F, Affolter M & Pyrowolakis G (2014) Dpp/BMP signaling in flies: from molecules to biology. *Semin Cell Dev Biol* 32, 128-136.

27. Matsuda S & Shimmi O (2012) Directional transport and active retention of Dpp/BMP create wing vein patterns in *Drosophila*. *Dev Biol* 366, 153-162.

28. Conley CA, Silburn R, Singer MA, Ralston A, Rohwer-Nutter D, Olson DJ, Gelbart W & Blair SS (2000) Crossveinless 2 contains cysteine-rich domains and is required for high levels of BMP-like activity during the formation of the cross veins in *Drosophila*. *Development* 127, 3947-3959.

29. Weiss A, Charbonnier E, Ellertsdóttir E, Tsigos A, Wolf C, Schuh R, Pyrowolakis G & Affolter M (2010) A conserved activation element in BMP signaling during *Drosophila* development. *Nat Struct Mol Biol* 17, 69-76.

30. Pham T, Day SM, Glassford WJ, Williams TM & Rebeiz M (2017) The evolutionary origination of a novel expression pattern through an extreme heterochronic shift. *Evol Dev* 19, 43-55.

31. Pyrowolakis G, Hartmann B, Müller B, Basler K, Affolter M (2004) A simple molecular complex mediates widespread BMP-induced repression during *Drosophila* development. *Dev Cell* 7, 229-40.

32. Waltzer L & Bienz M (1999) A function of CBP as a transcriptional co-activator during Dpp signalling. *EMBO J* 18, 1630-1641.

33. Morikawa M, Koinuma D, Miyazono K & Heldin CH (2013) Genome-wide mechanisms of Smad binding. *Oncogene* 32, 1609-1615.

34. Rebeiz M, Jikomes N, Kassner VA, & Carroll SB (2011) Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci USA* 108, 10036-10043.

35. Glassford WJ, Johnson WC, Dall NR, Smith SJ, Liu Y, Boll W, Noll M & Rebeiz M (2015) Co-

option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty.
Dev Cell 34, 520-531.

36. Xin Y, Le Poul Y, Ling L, Museridze M, Mühling B, Jaenichen R, Osipova E & Gompel N
 (2020) Ancestral and derived transcriptional enhancers share regulatory sequence and a pleiotropic
 site affecting chromatin accessibility. *Proc Natl Acad Sci USA* 117, 20636-20644.

37. Eichenlaub MP, & Ettwiller L (2011) *De novo* genesis of enhancers in vertebrates. *PLoS Biol* 9,
 e1001188.

38. Rebeiz M & Tsiantis M. (2017) Enhancer evolution and the origins of morphological novelty.
Curr Opin Genet Dev, 45, 115-123.

39. Smith AF, Posakony JW & Rebeiz M. (2017) Automated tools for comparative sequence analysis
 of genic regions using the GenePalette application. *Dev Biol* 429, 158-164.

40. Messeguer X, Escudero R, Farré D, Núñez O, Martínez J & Albà MM (2002) PROMO: detection
 of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18, 333-
 334.

41. Farré D, Roset R, Huerta M, Adsua JE, Roselló L, Albà MM & Messeguer X (2003)
 Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN.
Nucleic Acids Res 31, 3651-3653.

42. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A & Carroll SB (2008) The regulation and
 evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* 134, 610-623.

43. Venken KJ, He Y, Hoskins RA & Bellen HJ (2006) P[acman]: a BAC transgenic platform for
 targeted insertion of large DNA fragments in *D. melanogaster*. *Science* 314, 1747-1451.

44. Fukutomi Y, Matsumoto K, Funayama N & Koshikawa S (2018) Methods for staging pupal
 periods and measurement of wing pigmentation of *Drosophila guttifera*. *J Vis Exp* 131, e56935.

45. Bainbridge SP & Bownes M (1981) Staging the metamorphosis of *Drosophila melanogaster*. *J*
Embryol Exp Morph 66, 57-80.

463 46. Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image
464 analysis. *Nature Methods* 9, 671-675.
465

Figure Legends

Fig. 1. Adult wing of *D. guttifer* and *D. melanogaster* and *cis*-regulation of *wingless*. (A) Adult wing of *D. guttifer*. (B) Adult wing of *D. melanogaster*. (C) *wingless* expression pattern driven by gutCVT-core. Reporter EGFP is expressed in the crossveins (arrows), and the longitudinal vein tips (asterisks). (D) *wingless* expression pattern driven by melCV. Reporter EGFP is expressed in the crossvein only. (E) Comparison of gutCVT-core and melCV-core. Gray boxes indicate sequences that match in at least 14 bp, visualized as anchor-points in GenePalette [39] and the two *cis*-regulatory sequences are divided into five regions by some of these anchor-points. The divided regions are called region 1, 2, 3, 4, or 5. Red squares indicate the sequences similar to SMAD consensus (GCCGnCGC or 1 bp change of it), blue pentagon indicates the sequence similar to BMP-AE (GCTGTGTAACCGGG), and green triangles indicate the Medea binding motif (GTCTG). A–D were reproduced from [18]. Expression on the middle of longitudinal vein 4 and at the base of all longitudinal veins was not observed (or only weakly present) in the reporter expression of gutCVT (a larger version of gutCVT-core) or in *in situ* hybridization of *wingless*; therefore, repressing elements were present in gutCVT outside of gutCVT-core [18].

Fig. 2. EGFP reporter assay for *cis*-regulatory sequences lacking some of the regions. (A) Enhancer activity of partial sequences of gutCVT-core (blue) and melCV-core (orange) was examined. (B) gut2-5 drove the reporter expression in the entire wing veins. (C) gut3-5 did not show strong expression in the wing veins. (D) gut1-4 activated expression in the crossveins (arrows) and longitudinal vein tips (asterisks). (E) gut1-3 drove expression in the crossveins (arrows) only. (F) gut2-3 did not show strong expression in the wing veins. (G) mel1-4 drove strong expression in the crossveins (arrow). (H) mel2-4 did not drive expression in the wing veins. White bars indicate 100 μ m (B–H).

Fig. 3. EGFP reporter assay for chimeric sequences. (A) Chimeric sequence in which region 1 of gutCVT-core is replaced with that of melCV-core. mel1gut2-5 drove reporter EGFP expression in the crossveins (arrows) and longitudinal vein tips (asterisks). (B) Chimeric sequence in which sequences from gutCVT-core are used for regions 1, 2, and 4, and a sequence from melCV-core is used for region 3. mel3gut1-2+4 activated expression in the crossveins (arrows) and longitudinal vein tips (asterisks). White bars indicate 100 μ m (A, B).

Fig. 4. Changes in regulatory function by the knockout of candidate binding sites. (A) Sequences modified from gutCVT-core (blue) and melCV-core (orange) were examined. Red indicates the sections where the binding sites were knocked out and green indicates the sections where the

knocked-out sequences were recovered (same as wild type). (B) gutCVT-core whose BMP-AE was knocked out (BMP-AE KO) drove EGFP expression in crossveins (arrows) and longitudinal vein tips (asterisks). (C) gutCVT-core whose SMAD consensus were all knocked out (gutSMAD bs KO) lost the enhancer activity in the wing veins. (D) When SMAD consensus in region 3 was recovered (gutSMAD bs3 Rec), the sequence regained the enhancer activity in the crossveins (arrows) and longitudinal vein tips (asterisks). (E) Even when all SMAD consensus in region 4 were recovered (gutSMAD bs4 Rec), the sequence did not activate expression in the wing veins. (F) When SMAD consensus in melCV-core were all knocked out (melSMAD bs KO), the sequence lost the enhancer activity in the crossveins. White bars indicate 100 μ m (B-F).

Fig. 5. Predicted evolution of the sequence and the function of gutCVT-core. Repression of the gene expression in the longitudinal veins except the vein tips was exerted by both melCV-core and gutCVT-core (left, orange). With melCV-core, expression was activated only in the crossveins by sequences in region 1 and SMAD consensus in region 3 (red square) (upper center, green). In gutCVT-core, sequences that coordinate with the SMAD consensus in region 3 were acquired, producing a set of sequences which has activating function in the entire wing veins (lower center, green). With gutCVT-core, gene expression was driven in crossveins and longitudinal vein tips by the combination of the ancestral repression and the newly gained activation in the entire wing veins (lower right).

Fig.1

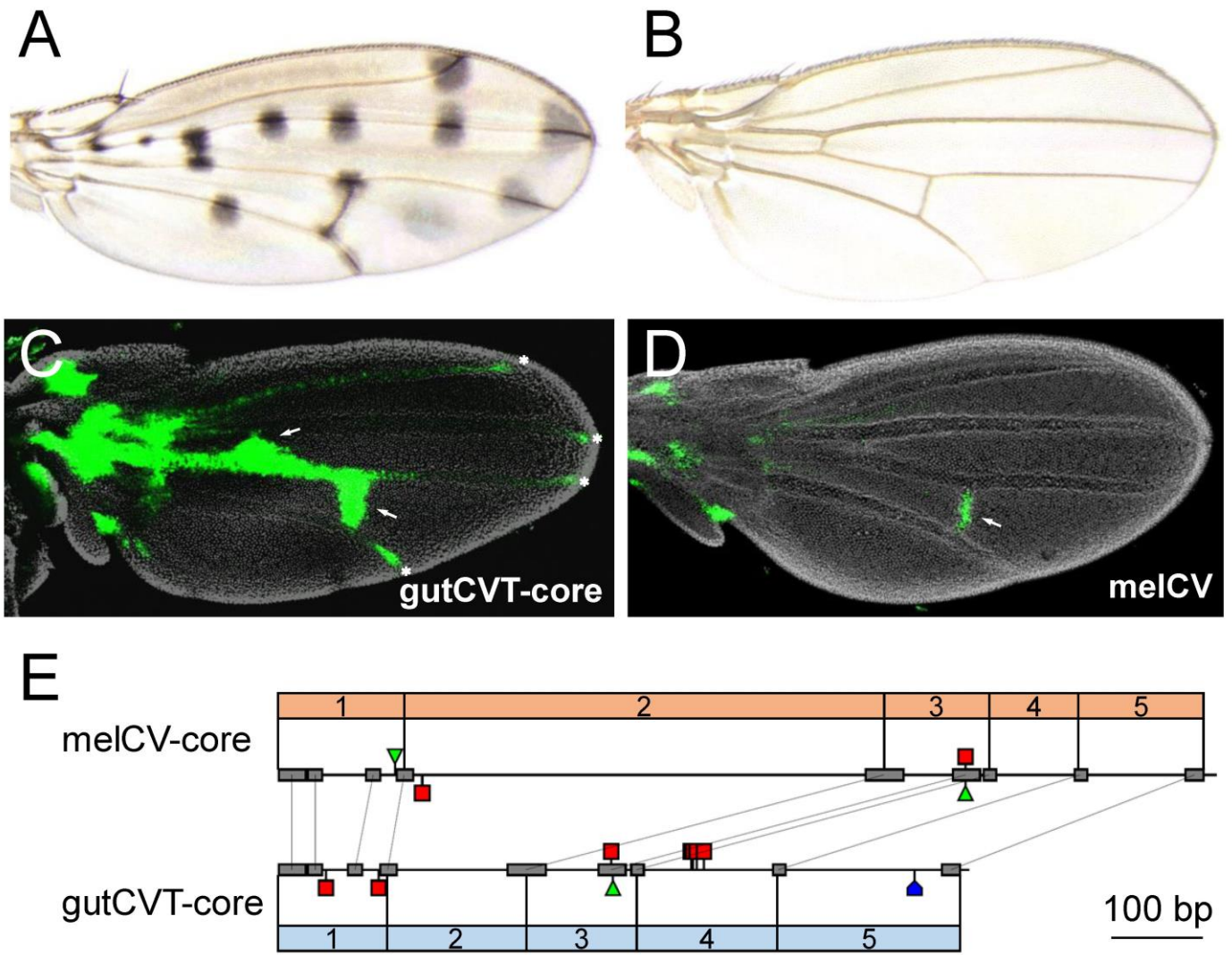


Fig.2

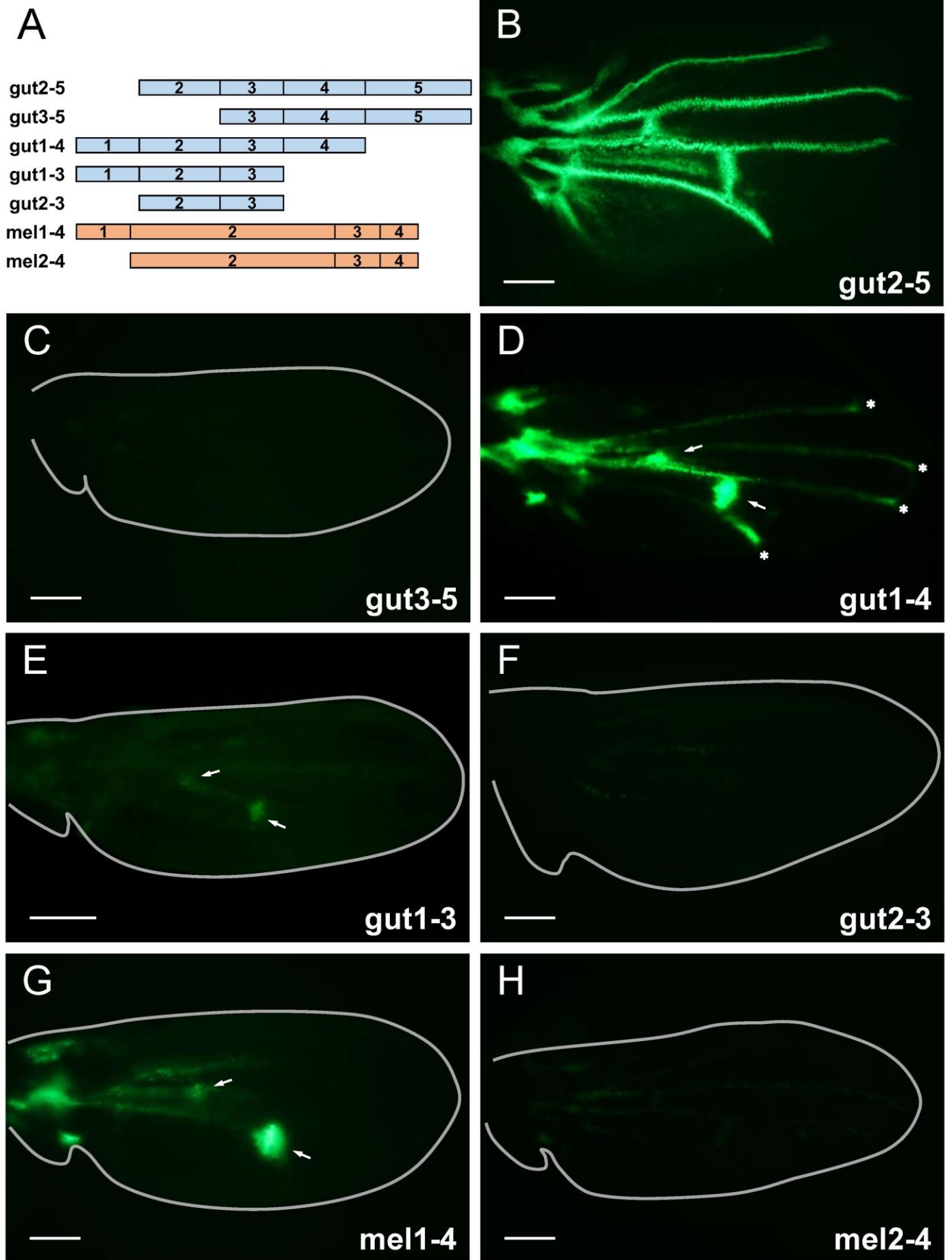


Fig.3

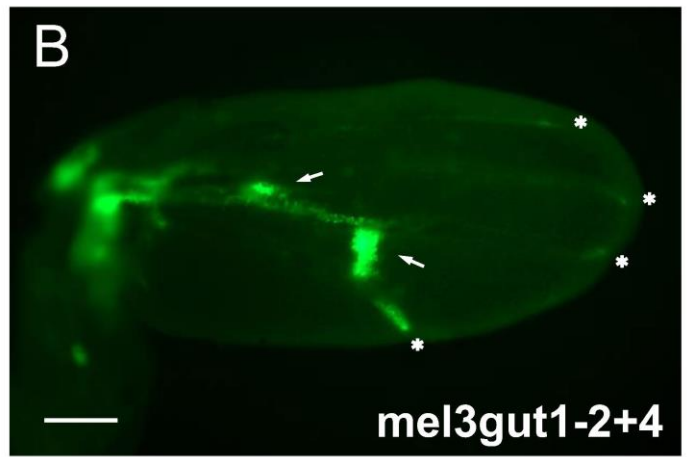
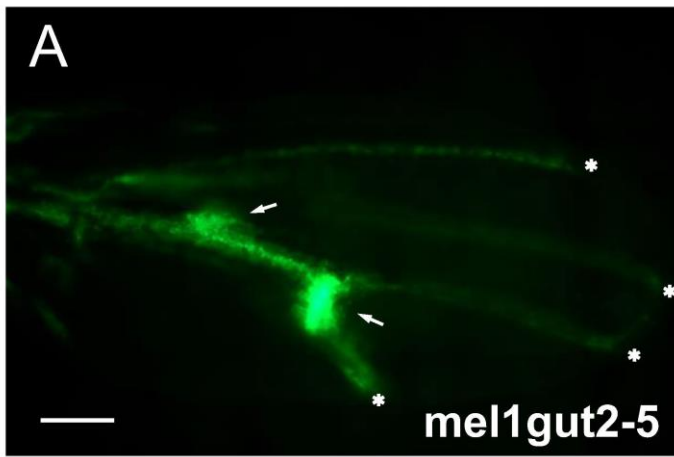


Fig.4

A

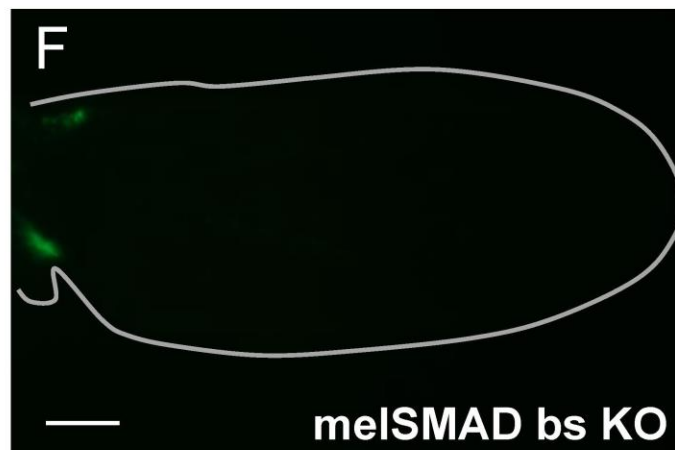
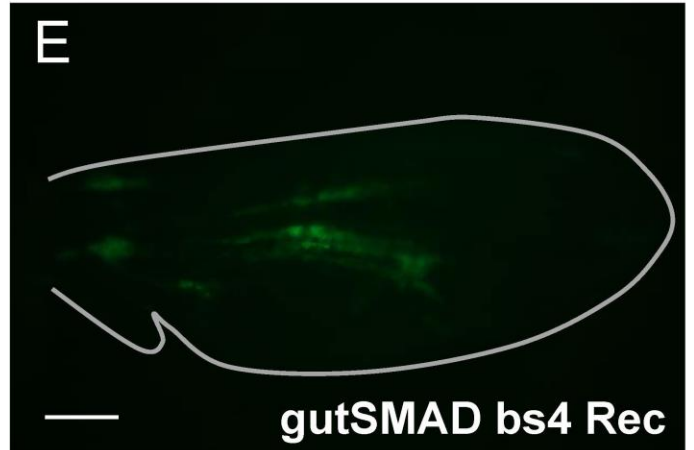
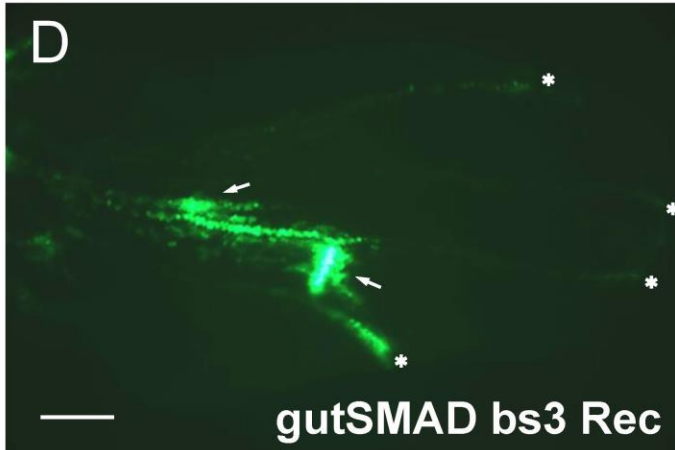
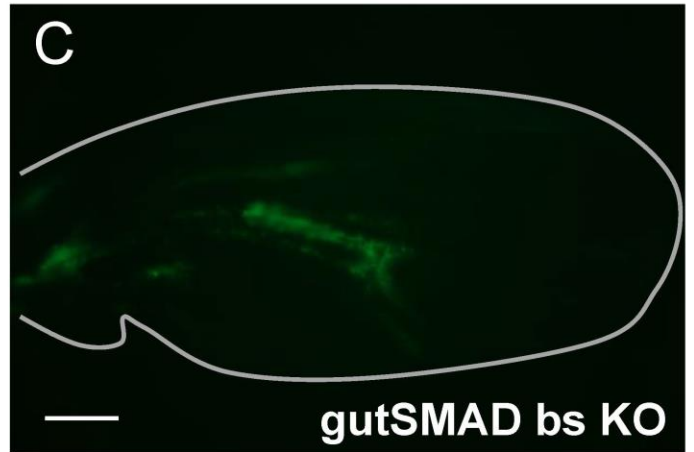
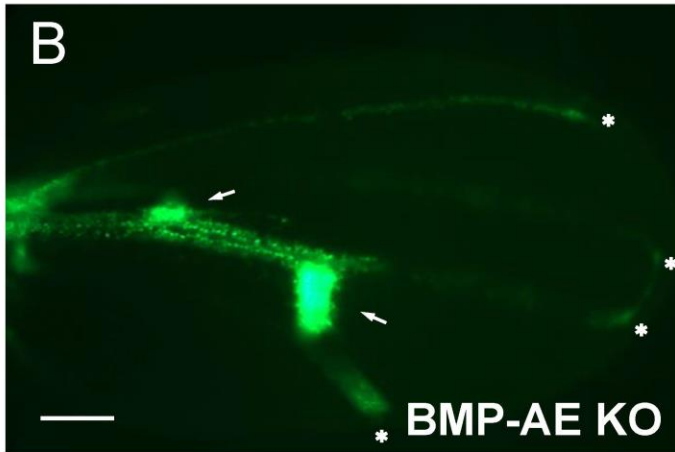
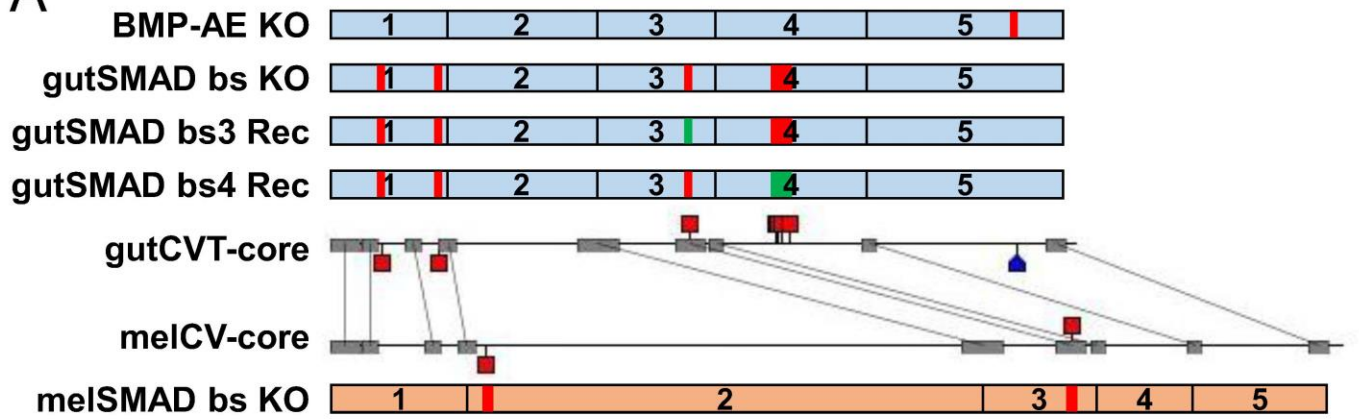
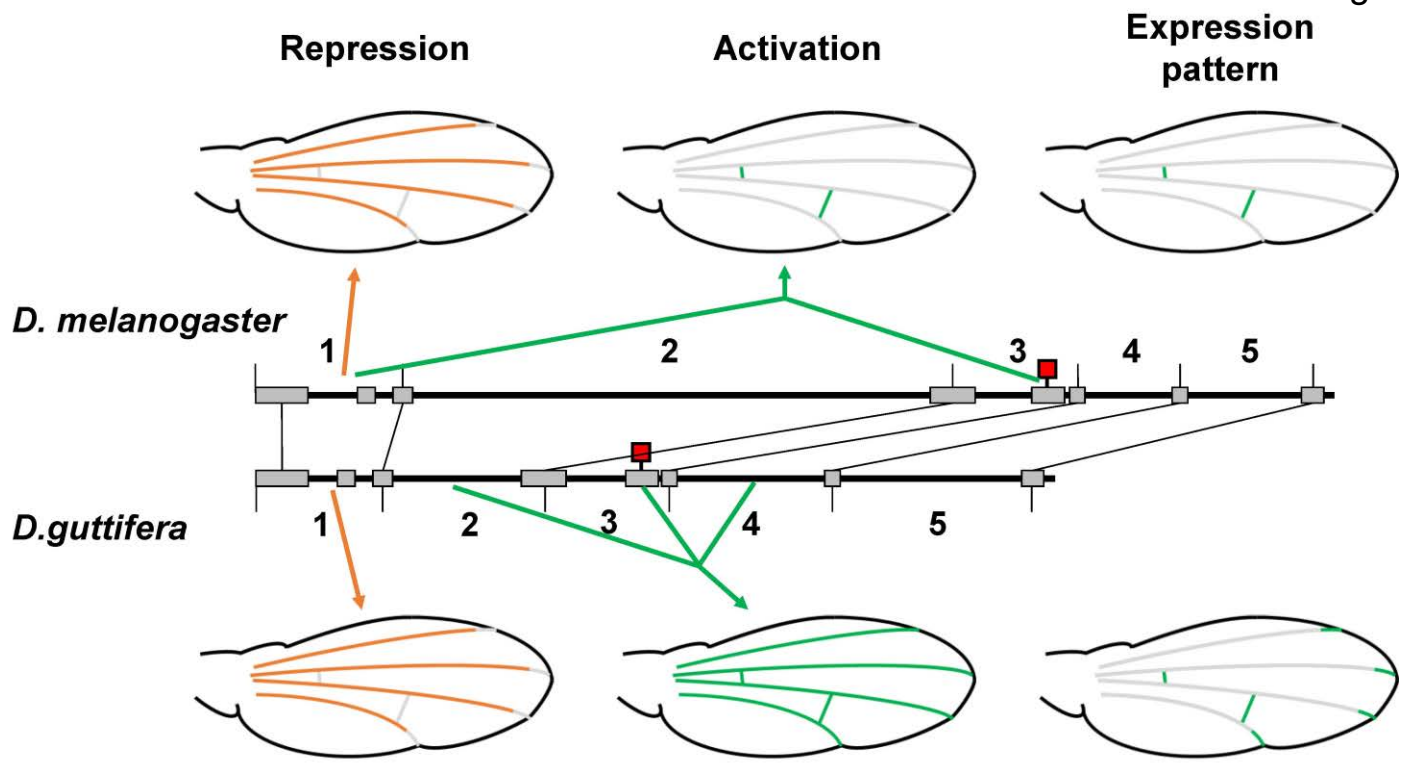


Fig.5



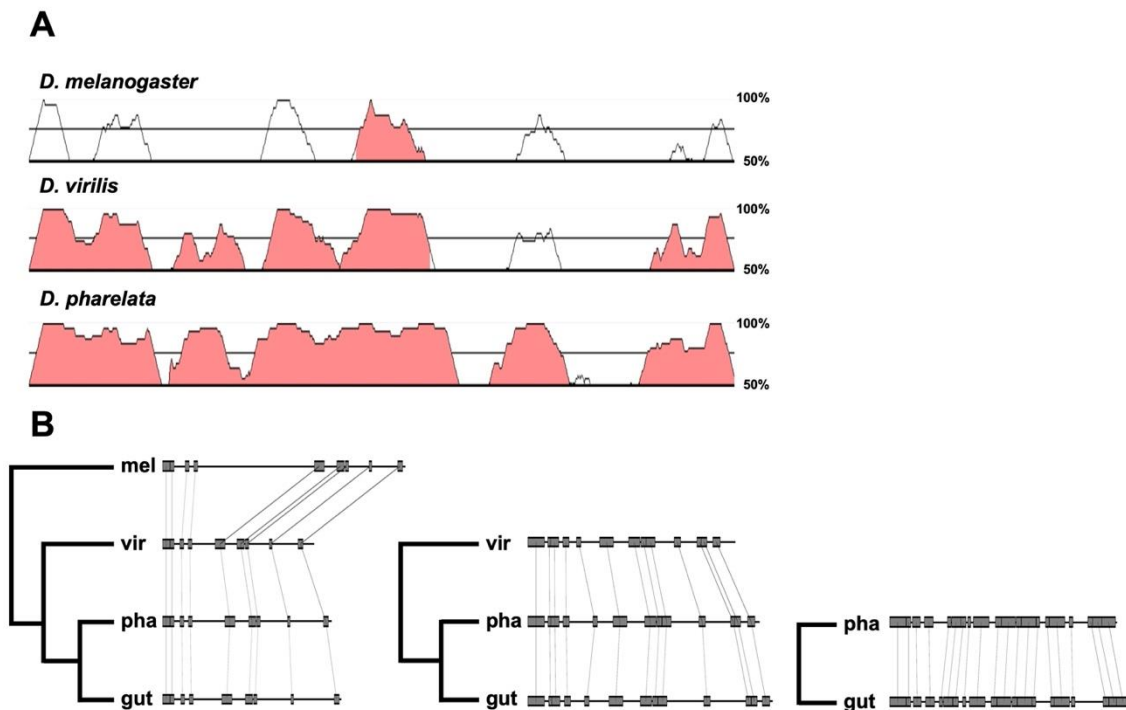


Figure. S1. Comparisons of crossvein enhancer regions in *D. melanogaster*, *D. virilis*, *D. phalerata*, and *D. guttifer*. (A) 30-bp sliding windows of crossvein enhancer regions using mVISTA [1]. Each sequence was compared with gutCVT-core and peaks show the degree of sequence conservation. (B) Crossvein enhancer regions of *D. melanogaster* (mel), *D. virilis* (vir), *D. phalerata* (pha), and *D. guttifer* (gut) were compared using GenePalette [2]. Gray boxes (anchor points in GenePalette function) indicate completely conserved sequences which are 10 bp or longer. Phylogenetic trees are based on [3]. Conserved sequences indicated as anchor points seem to remain colinear. The crossvein enhancer sequence of *D. phalerata* was obtained from the genome data of the Darwin Tree of Life Project (ERX9984265) [4].

Reference list

1. Frazer KA, Pachter L, Poliakov A, Rubin EM & Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 1, 32 (Web Server issue): W273-9.
2. Smith AF, Posakony JW & Rebeiz M. (2017) Automated tools for comparative sequence analysis of genic regions using the GenePalette application. *Dev Biol* 429, 158-164.
3. Finet C, Kassner VA, Carvalho AB, Chung H, Day JP, Day S, Delaney EK, De Ré FC, Dufour HD, Dupim E, Izumitani HF, Gautério TB, Justen J, Katoh T, Kopp A, Koshikawa S, Longdon B, Loreto EL, Nunes MDS, Raja KKB, Rebeiz M, Ritchie MG, Saakyan G, Sneddon T, Teramoto M, Tyukmaeva V, Vanderlinde T, Wey EE, Werner T, Williams TM, Robe LJ, Toda MJ & Marlétaz F. (2021) Drosophila: Resources for Drosophilid Phylogeny and Systematics. *Genome Biol Evol* 3, 13(8): evab179.
4. Darwin Tree of Life Project Consortium. (2022) Sequence locally, think globally: The Darwin Tree of Life Project. *Proc Natl Acad Sci USA* 25, 119(4): e2115642118.

Table S1

Table. S1. All primers used in this research		
Primer name	Pimer Sequence	Resriction Site
gut BMP-AE knockout 1 AscI R1	TCTTGGCGCGCCCTCGAGCACAAACAATTGCCACTTAATTAACA CGAACTCCAACCTCTTTTTTTTTTTTTTTTAAAATACATATTTTCG	AscI
gut JY1A SMAD knockout1 SbfI F1	TAATCCTGCAGGGACATTGCTCCTAATCAATAAACTAACGAGC GGATTAAGATTATGTTGCACGATAACTCATCCACACAGAGGCG	SbfI
gut JY1L SMAD knockout1 F1	GTGTTGTAATTTGACGTCCCACACGTTTGC	
gut JY1L SMAD knockout1 R1	GCAAACGTGTGGGACGTCAAATTACAACAC	
gut JY1M SMAD knockout1 F1	CAAAAATGGTTTGTCTATCGCTAATGAATTGC	
gut JY1M SMAD knockout1 R1	GCAATTCATTAGCGATGACAAACCATTTTTTG	
gut multiSMAD knockout1 F1	AGACATTGAGAGCGCAACGCGTATTGTTACACGTCGCCTTCT	
gut multiSMAD knockout1 R1	AGAAGGCGACGTGTACAATACGCGTTGCGCTCTCAATGTCT	
gut wg JY1J AscI R1	TCTTGGCGCGCCCTCGAGCACAAACA	AscI
gut wg JY1L AscI R1	ATATGGCGCGCCCCAAATCAAACAGCGGCGCA	AscI
gut wg JY1L F1 Karasawa	CGCTGTTTGATTTGGCCAAA	
mel JY1L SMAD knockout1 F1	TGTTGTAATTTGCGAACCCACACGTTTTCG	
mel JY1L SMAD knockout1 R1	CGAAACGTGTGGGTTCGCAAATTACAACA	
mel JY1M SMAD knockout1 F1	CGAGTCGCCATCATGCATCACTATAAG	
mel JY1M SMAD knockout1 R1	CTTATAGTGATGCATGATGGCGACTCG	
mel wg JY1J AscI R1	TCTTGGCGCGCCCTCACGCCTCGAAAC	AscI
mel wg JY1L R1 Karasawa	CCAAATCAAACAGCGCGCGA	
wg JY1A SbfI F2	ATCGCCTGCAGGGACATTGCTCCTAA	SbfI
wg JY1M For Karasawa	TGGCGGCCTAATGAATTGC	
wg JY1M Rev Karasawa	GCAATTCATTAGGCCGCCA	
wg JY1M SbfI F1	TAATCCTGCAGGTGGCGGCCTAATGAATTGC	SbfI
wg JY1O AscI R1	ATATGGCGCGCCTTGTCTATGCCAAGAGTGGCG	AscI
wg JY1P F1 Karasawa	GATGTGACAATGACAGTTGTTGTTC	
wg JY1P R1 Karasawa	GAACAACAACGTGCATTGTCACATC	
wg JY1P SbfI F1	TAATCCTGCAGGGATGTGACAATGACAGTTGTTGTTC	SbfI

Table. S2. Primers used for each tested <i>cis</i> -regulatory sequences		
Tested sequence	Primers for producing dna fragments	Template sequence
gut2-5	wg JY1M SbfI F1	gutCVT-core
	gut wg JY1J AscI R1	gutCVT-core
gut3-5	wg JY1P SbfI F1	gutCVT-core
	gut wg JY1J AscI R1	gutCVT-core
gut1-4	wg JY1A SbfI F2	gutCVT-core
	wg JY1O AscI R1	gutCVT-core
gut1-3	wg JY1A SbfI F2	gutCVT-core
	gut wg JY1L AscI R1	gutCVT-core
gut2-3	wg JY1M SbfI F1	gutCVT-core
	gut wg JY1L AscI R1	gutCVT-core
mel1-4	wg JY1A SbfI F2	melCV-core
	wg JY1O AscI R1	melCV-core
mel2-4	wg JY1M SbfI F1	melCV-core
	wg JY1O AscI R1	melCV-core
mel1gut2-5	wg JY1A SbfI F2	melCV-core
	wg JY1M Rev Karasawa	melCV-core
	wg JY1M For Karasawa	gutCVT-core
	gut wg JY1J AscI R1	gutCVT-core
	wg JY1A SbfI F2	overlap extension PCR
	gut wg JY1J AscI R1	overlap extension PCR
mel3gut1-2+4	wg JY1A SbfI F2	gutCVT-core
	wg JY1P R1 Karasawa	gutCVT-core
	wg JY1P F1 Karasawa	melCV-core
	mel wg JY1L R1 Karasawa	melCV-core
	gut wg JY1L F1 Karasawa	gutCVT-core
	wg JY1O AscI R1	gutCVT-core
	wg JY1A SbfI F2	overlap extension PCR
	wg JY1O AscI R1	overlap extension PCR
BMP-AE KO	wg JY1A SbfI F2	gutCVT-core
	gut BMP-AE knockout 1 AscI R1	gutCVT-core

Table S2

gutSMAD bs KO	gut JY1A SMAD knockout1 Sbf1 F1	gutCVT-core
	gut JY1M SMAD knockout1 R1	gutCVT-core
	gut JY1M SMAD knockout1 F1	gutCVT-core
	gut JY1L SMAD knockout1 R1	gutCVT-core
	gut JY1L SMAD knockout1 F1	gutCVT-core
	gut multiSMAD knockout1 R1	gutCVT-core
	gut multiSMAD knockout1 F1	gutCVT-core
	gut wg JY1J Asc1 R1	gutCVT-core
	wg JY1A Sbf1 F2	overlap extension PCR
	gut wg JY1J Asc1 R1	overlap extension PCR
gutSMAD bs3 Rec	gut JY1A SMAD knockout1 Sbf1 F1	gutCVT-core
	gut JY1M SMAD knockout1 R1	gutCVT-core
	gut JY1M SMAD knockout1 F1	gutCVT-core
	gut multiSMAD knockout1 R1	gutCVT-core
	gut multiSMAD knockout1 F1	gutCVT-core
	gut wg JY1J Asc1 R1	gutCVT-core
	gut JY1A SMAD knockout1 Sbf1 F1	overlap extension PCR
	gut wg JY1J Asc1 R1	overlap extension PCR
gutSMAD bs4 Rec	gut JY1A SMAD knockout1 Sbf1 F1	gutCVT-core
	gut JY1M SMAD knockout1 R1	gutCVT-core
	gut JY1M SMAD knockout1 F1	gutCVT-core
	gut JY1L SMAD knockout1 R1	gutCVT-core
	gut JY1L SMAD knockout1 F1	gutCVT-core
	gut wg JY1J Asc1 R1	gutCVT-core
	gut JY1A SMAD knockout1 Sbf1 F1	overlap extension PCR
	gut wg JY1J Asc1 R1	overlap extension PCR
melSMAD KO	wg JY1A Sbf1 F2	melCV-core
	mel JY1M SMAD knockout1 R1	melCV-core
	mel JY1M SMAD knockout1 F1	melCV-core
	mel JY1L SMAD knockout1 R1	melCV-core
	mel JY1L SMAD knockout1 F1	melCV-core
	mel wg JY1J Asc1 R1	melCV-core
	wg JY1A Sbf1 F2	overlap extension PCR
	mel wg JY1J Asc1 R1	overlap extension PCR