Mblk-1/E93, an ecdysone related-transcription factor, targets synaptic plasticity-related genes in the honey bee mushroom bodies

Yasuhiro Matsumura
The University of Tokyo
Taiko Kim To
The University of Tokyo
Hiroki Kohno
The University of Tokyo
Tetsuji Kakutani
The University of Tokyo
Takeo Kubo (stkubo@bs.s.u-tokyo.ac.jp)
The University of Tokyo

Article

Keywords:

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1612011/v1

Abstract

Among hymenopteran insects, aculeate species such as bees, ants, and wasps have enlarged and morphologically elaborate mushroom bodies (MBs), a higher-order brain center in the insect, implying their relationship with the advanced behavioral traits of aculeate species. The molecular bases leading to the acquisition of complicated MB functions, however, remains unclear. We previously reported the constitutive and MB-preferential expression of an ecdysone-signaling related transcription factor, Mblk-1/E93, in the honey bee brain. Here, we searched for target genes of Mblk-1 in the worker honey bee MBs using chromatin immunoprecipitation sequence analyses and found that Mblk-1 targets several genes involved in synaptic plasticity, learning, and memory abilities. We also demonstrated that Mblk-1 expression is self-regulated *via* Mblk-1-binding sites, which are located upstream of *Mblk-1*. Furthermore, we showed that the number of the Mblk-1-binding motif located upstream of *Mblk-1* homologs increased associated with evolution of hymenopteran insects. Our findings suggest that *Mblk-1*, which has been focused on as a developmental gene transiently induced by ecdysone, has acquired a novel expression pattern to play a role in synaptic plasticity in honey bee MBs, raising a possibility that molecular evolution of Mblk-1 may have partly contributed to the elaboration of MB function in insects.

Introduction

In insects, mushroom bodies (MBs), a higher-order center in insect brains, play important roles in olfactory associative learning [1], spatial learning [2], sensory integration [3], and social behavior [4]. In Hymenoptera, aculeata species such as bees, ants, and wasps have enlarged and morphologically elaborated MBs compared to more primitive hymenopteran insects like parasitoid wasps and sawflies, suggesting a relationship between MB elaboration and the advanced behavioral traits of these insects [5, 6]. The molecular and neural bases underlying how aculeate MBs acquired elaborated functions, however, remain elusive. Aculeate hymenopteran species show advanced social behaviors [7, 8]. For example, European honey bee (*Apis mellifera* L.) workers perform the 'waggle dance' to inform their nestmates of the location of food sources that they memorize during foraging flights [9]. Therefore, honey bees are useful model insects for understanding neural mechanisms underlying sociality, learning ability [10, 11] and brain evolution in insects.

The honey bee MBs comprise 4 types of interneurons termed Kenyon Cells (KCs): class I large-, middle-, small-type KCs, and class II KCs, which are distinguished on the basis of their location, the size of their somata, and their gene expression profiles [12]. Among these KC subtypes, large-type KCs (lKCs) are characteristic in that they preferentially express genes involved in Ca^{2+}-signaling, including *Ca^{2+}/calmodulin-dependent protein kinase II* (*CaMKII*) [13, 14]. Since Ca^{2+}-signaling pathways play an important role in synaptic plasticity, the primary mechanism of learning and memory functions in many animal species [15], it is plausible that lKCs are involved in the learning and memory abilities that underlie honey bee social behaviors.

In addition to the genes involved in Ca^{2+}-signaling, a sequence-specific transcription factor, termed *mushroom body/large-type Kenyon-cell specific protein-1* (*Mblk-1*) */ ecdysone-induced protein 93F* (*E93*) (hereafter termed as *Mblk-1*) is preferentially expressed in lKCs [16–18]. *Mblk-1* is widely conserved in animal species from invertebrates to vertebrates [19]. Especially in insects, *E93*, the *Mblk-1* homolog, is upregulated by the molting hormone ecdysone and ecdysone receptor and triggers programmed cell death of larval tissues during metamorphosis [20–24]. In contrast to E93's transient roles reported so far, Mblk-1 (*Am*E93) is expressed constitutively in lKCs through the pupal to adult stages [18]. Thus, in addition to its common roles in insect metamorphosis, Mblk-1 appears to have novel functions in honey bee lKCs. The function of Mblk-1 as a transcription factor and its target genes in honey bee lKCs, however, are unknown.

We previously used an affinity-purified antibody raised against a partial recombinant Mblk-1 protein to examine Mblk-1 protein expression in the honey bee brain [18]. In the present study, we performed chromatin immunoprecipitation sequencing (ChIP-seq) using this antibody to identify Mblk-1 target genes in honey bee brains. We found that Mblk-1 regulates genes related to synaptic plasticity, such as *CaMKII* [25–27] and *Mblk-1* itself, suggesting that Mblk-1 has acquired a new expression mechanism and roles related to synaptic plasticity in the honey bee MBs. We propose that acquisition of unconventional functions of Mblk-1 in lKCs might have partly contributed to the evolution of learning and memory abilities and social behaviors in aculeate Hymenoptera.

Methods

Insects

European honey bee colonies maintained at The University of Tokyo (Hongo Campus) were used. Some colonies were also purchased from a local dealer (Kumagaya Honeybee Farm, Saitama, Japan). For ChIP-seq analysis using adult worker samples, we randomly collected workers from the hives irrespective of age and/or role. For ChIP-seq analysis using pupal worker samples, we selected and used worker pupae with brown or purple eyes (stages: P3-P5). Sawflies (*Athalia rosae*) were a gift from Mr. Takayoshi Kuwabara. Adult sawflies within 10 days after emergence were used for the experiment.

Analysis of genes expressed in adult worker MBs using RNA-seq data

The RNA-seq data (low-value reward & no-dance) was obtained from GEO accession: GSM3747967 [28]. The value of 2012_309_TTAGGC_L007_Aligned.sortedByCoord.out.bam in GSM3747967_2012_309_TTAGGC_L007_Aligned.sortedByCoord.out.bam.counts.txt was divided by each gene length, and the top 40 genes with the highest divided values were displayed.

Chromatin preparation

80 MBs dissected from adult workers or 60 whole brains dissected from worker pupae were used per single chromatin immunoprecipitation sequence (ChIP-seq) experiment. The ChIP-seq experiments were

conducted using 2 biological replicates for the adult and pupal samples. The MBs or whole brains were dissected under binocular microscopy from cooled and anesthetized honey bees and homogenized with plastic pestles in 200 µl of buffered insect saline (20 mM Tris-HCl [pH 7.4], 130 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$, 10 mM HEPES buffer, protease inhibitor cocktail [Roche], and PhosSTOP [Roche]). The cells were fixed with formaldehyde (1% final conc.) for 12 min at room temperature. The fixing reaction was quenched with 2.5 M glycine (125 mM final conc.) for 10 min on ice. The cells were washed 3 times with 150 µl of buffered insect saline and centrifuged (600 x g, 5 min, 4℃). The supernatant was removed, and the cells were suspended in 300 µl of buffered insect saline. The cells were crushed for 25 s using zirconium beads (TOMY) and Micro Smash MS-100 (TOMY). After centrifugation, the supernatant was discarded and 200 µl of nuclei lysis solution (10% glycerol, 0.5% NP-40, 2 mM EDTA, 50 mM Tris-HCl [pH 7.4], and protease inhibitor cocktail [Roche]) were added. The sample was incubated on ice for 30 min. After centrifugation, the supernatant was discarded and 500 µl of Micrococcal Nuclease buffer (50 mM Tris-HCl [pH 7.4] and 5 mM $CaCl_2$) was added. The sample was kept on ice for 10 min and centrifuged. After the supernatant was discarded, the pellet was resuspended in 140 µl of Micrococcal Nuclease buffer containing protease inhibitor cocktail and 0.2 µl of Micrococcal Nuclease (2000 gel units/µl, New England BioLabs). The sample was incubated at 37°C for 12 min and digestion was stopped by 500 mM EDTA (total volume: 150 µl). After adding 300 µl of 1.5x ChIP dilution buffer (1x ChIP dilution buffer: 50 mM Tris-HCl [pH 7.4], 167 mM NaCl, 1.1% Triton X-100, 0.11% SDS), the sample was sonicated (output 10%, 12s at 60s intervals, 3 times) using a SONIFIER250 (Branson). After sonication, the sample was centrifuged, and the supernatant was collected for immunoprecipitation.

Chromatin immunoprecipitation (ChIP)

The sample containing the fragmented chromatin and anti-Mblk-1 antibody-attached Dynabeads Protein A (Invitrogen) were incubated with rotation for 45 min at 4℃. The Dynabeads complex was washed twice with RIPA buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% TritonX-100), and 3 times with Phosphate-Buffered Saline (PBS). The washed Dynabeads complex was resuspended in PBS. The solution containing the beads was transferred to a new clean tube. The tube was placed on magnet and the supernatant was removed. The anti-Mblk-1 antibody and protein complex was eluted by elution buffer (1% SDS, 50 mM Tris-HCl [pH 7.4], 8.5 mM EDTA). The tube was placed on magnet and the supernatant was transferred to a new tube. After adding TE buffer and 10 mg/ml RNase A, the sample was incubated for 1 h at 50℃, and then 20 mg/ml proteinase K was added, and the crosslink was reversed by incubation for 12 h at 50℃. The DNA was purified and processed to sequencing preparation. The sequencing library was prepared according to the Illumina protocol and sequenced on an Illumina Hiseq2500.

ChIP-seq data processing

Sequence reads were aligned to the reference genome (Amel_HAv3.1) [29] using Bowtie2 with default parameters [30]. Peak calling was performed using MACS2 [31] and the input DNA sample as a control. After peak calling, we visualized the ChIP-seq data using the Integrative Genomics Viewer (IGV) [32]. The

ChIP-seq experiments were performed twice using 2 biological replicates. The peak signals detected in the 2 replicates were considered reproducible if they were located within 1 kb in the genome. Subsequent analysis was conducted using the signals that were reproducible.

Motif analysis

Sequences of ± 150 bp (300 bp) from the centers of the reproducible signals with a peak calling q-value threshold of 0.001 in the analysis using adult MBs were obtained. We submitted these as input sequences into MEME-ChIP (Version 5.2.0). As the motif width options, -minw was set to 6, and -maxw was set to 17.

Luciferase assay

We obtained the coding sequence of *Mblk-1* from Mblk-1/pPac-PL [33] and inserted it into a multiple cloning region of an expression vector, pHEK293 Ultra Expression Vector II (TaKaRa). Reporter vectors were constructed by inserting 6xGAGA, 6xGAGA with a 4-bp spacer sequence, 6xATTTTG, 6xGAAATTTT, 6xATTTTGG, and an Mblk-1 enhancer candidate sequence (NC_037652.1:7063408–7063595, 188 bp), respectively, into the upstream region of minimal promoter of pGL4.23 [luc2/minp]. For the Mblk-1 enhancer candidate sequence, we used the overlap region of the 2 Mblk-1-binding sequences (NC_037652.1:7063147–7063709 and NC_037652.1:7063400–7063629) obtained from the 2 independent ChIP-seq experiments. pRL-TK was used as an internal control vector to normalize transfection efficiency. The Mblk-1 expression vector or pHEK 293 Ultra Expression Vector II without the Mblk-1 coding sequence as a negative control, pHEK293 Enhancer Vector, one of the reporter vectors (pGL4.23 with the motif sequences), and the internal control vector (pRL-TK) were transfected to HEK293T cells (RCB2202) from RIKEN BioResource Center using Lipofectamine LTX Reagent with PLUS Reagent (Invitrogen). The pHEK293 Enhancer Vector was used to enhance the expression of *Mblk-1* in pHEK293 Ultra Expression Vector II. We measured *Photinus pyralis* (firefly) and *Renilla reniformis* luciferase activities 40–48 h after transfection with a Dual-Luciferase Reporter Assay System (Promega) and 2030 ARVO X5 (PerkinElmer). *Photinus pyralis* (firefly) luciferase activity was normalized with the co-transfected internal control luciferase activity and then relative luciferase activity of the control assay using the pHEK 293 Ultra Expression Vector II without the Mblk-1 coding sequence was subtracted to exclude effects other than those of Mblk-1.

Single-cell RNA-seq data analysis

We obtained the raw single-cell RNA-seq data from the previously reported GEO accession ID GSE130785 [34]. Using SeuratV3 Wizard [35], we performed MB cell clustering and found marker genes for each cluster (cluster number: 10, granularity: 1.0). We selected cluster marker genes that were more highly expressed than in other regions; that is, cluster marker genes with positive avg_logFC values.

RNA probe synthesis and *in situ* hybridization

Total RNA was extracted from 3 honey bee adult worker brains using TRIzol Reagent (Invitrogen). After removing the genomic DNA with the gDNA Eraser from the PrimeScript RT reagent Kit with gDNA Eraser (Perfect Real Time), the extracted RNA was reverse-transcribed into cDNA using the same kit. The cDNA was partially amplified by PCR using Ex *Taq* (TaKaRa) with primers specific for each gene. The PCR products were cloned using pGEM-T Easy Vector Systems (Promega). The primer sequences used to synthesize the *in situ* hybridization RNA probes are listed in Supplementary Table 1. The DNA templates for *in vitro* transcription were amplified by PCR with M13 forward and reverse primers. The digoxigenin (DIG)-labeled sense and antisense RNA probes were synthesized using the amplified DNA template, a DIG RNA Labeling Mix (Roche), SP6 RNA polymerase (Roche), and T7 RNA polymerase (Roche). For *in situ* hybridization, dissected whole brains were embedded in Tissue-Tek OCT compound (Sakura Finetek Japan) without fixation and immediately frozen and stored at -80℃ until use. The frozen whole brains were sliced into 10-μm thick sections. *In situ* hybridization was performed essentially as described previously [36]. Images were captured with a light microscope BX-50 (Olympus).

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR)

Total RNA was extracted from each tissue (brain, head without brain, thorax, and abdomen) using TRIzol Reagent (Invitrogen). cDNA synthesis was performed using a PrimeScript RT reagent Kit with a gDNA Eraser (Perfect Real Time; TaKaRa). qRT-PCR analysis was performed using TB Green *Premix Ex Taq II* (Tli RNaseH plus; Takara) and Light Cycler 480 Instrument II (Roche Life Science, Indianapolis, IN, USA). The following primers were used for gene amplification: *Mblk-1* homolog (LOC105691346), 5'-GCAACAGCAACAGAGAGAAAGG-3' and 5'-CGTTCCCGTACCGTCATAC-3'; *60S ribosomal protein L32* (LOC105692622), 5'-ACAGAGTTCGTAGGCGCTTC-3' and 5'-GCATCATCAAGACCTCCAAC-3'. The PCR conditions were as follows: denaturation at 95℃ for 30 s; 45 cycles of PCR (95℃ for 5 s, 61℃ for 15 s, 72℃ for 20 s), melting, 95℃ for 5 s, 65℃ for 60 s, then 95℃ (Ramp Rate: 0.11℃/s) and cooling, 50℃ for 30 s.

Analysis of the GAGA motifs upstream of *Mblk-1* homologs

Upstream sequences within 2 kb or 10 kb of the transcription start site of each *Mblk-1* homolog were obtained from NCBI. The numbers of GAGA or GAG sequences in the upstream region of each *Mblk-1* homolog were counted, allowing for overlap. For example, a 'GAGA' sequence was counted as one GAGA and one GAG, and a 'GAGAG' sequence was counted as one GAGA and two GAGs. The assemblies used were as follows: *Apis mellifera*, Amel_HAv3.1 [29]; *Harpegnathos saltator*, Hsal_v8.5 [37]; *Solenopsis invicta*, UNIL_Sinv_3.0; *Vespa mandarinia*, V.mandarinia_Nanaimo_p1.0; *Fopius arisanus*, ASM80636v1 [38]; *Nasonia vitripennis*, Nvit_psr_1.1 [39]; *Athalia rosae*, Aros_2.0 [40]; and *Drosophila melanogaster*, Release 6 plus ISO1 MT [41].

Statistical analysis

Statistical analysis was conducted as indicated in the text and figure legends. Where required P values were adjusted using Bonferroni or Benjamini-Hochberg correction. In Fig. 3D, Tukey-Kramer's tests after

one-way ANOVA were performed (n = 3). F values, P values, and degrees of freedom were 8.09, 0.00831 and 3 (right panel) and 99.3, 2.52×10^{-5} and 2 (left panel), respectively. Student's t test used in Fig. 6B was two sided (n = 4). In Fig. 6D, one-way ANOVA was performed. F value was 2.09, P value was 0.180 and degrees of freedom was 3 (n = 3).

Results

Identification of Mblk-1 target gene candidates in the worker MBs by ChIP-seq analysis

Mblk-1 is expressed in a lKC-preferential manner in adult worker MBs (Fig. 1A) [16, 18]. When we searched for highly expressed genes in the honey bee MBs using previously reported RNA-seq data [28], we found that Mblk-1 was the 38th most highly expressed gene and the most highly expressed among transcription factors in MBs (Fig. 1B), implying its important role in the adult honey bee MBs.

We previously prepared an antibody that binds specifically to Mblk-1 protein (hereafter termed anti-Mblk-1 antibody) (Fig. S1A) [18]. We first tested whether the anti-Mblk-1 antibody could be used for ChIP-seq analysis. Full-length Mblk-1 protein was mainly detected as 250-kDa bands in the immunoprecipitated sample as well as in the input and supernatant samples after immunoprecipitation using worker MB lysate (Fig. S1B), indicating that the anti-Mblk-1 antibody can be used for immunoprecipitation.

We performed ChIP-seq analysis using homogenates of adult worker MBs and the anti-Mblk-1 antibody. A total of 593 genes were identified as target gene candidates of Mblk-1, in which Mblk-1-binding signals were located within 10 kb upstream or downstream of the corresponding genes (Table 1 and Supplementary Data 1&3) when using a peak calling q-value threshold of 0.005 and the input DNA sample as a control. Schematic diagrams of some peak signals and their surrounding Mblk-1 target gene candidates are shown in Fig. 2A-F. The signal with the lowest q-value was in the third intron of an uncharacterized non-coding RNA gene (Gene 1 in Table 1 and Fig. 2A). Genes related to neural function, such as *suppressor of lurcher protein 1* (Gene 2 − 1 in Table 1 and Fig. 2B), *fasciclin-3* (Gene 4 in Table 1 and Fig. 2C), *potassium voltage-gated channel subfamily KQT member 1* (Gene 9 in Table 1 and Fig. 2D), and *CaMKII* (Gene 17 − 1 in Table 1 and Fig. 2F) were also identified as Mblk-1 target gene candidates. Moreover, among the Mblk-1 target gene candidates, some ecdysone-signaling pathway-related genes such as *ecdysone-induced protein 75* (*E75*) and *Ultraspiracle* (*Usp*) were identified. This is consistent with the fact that *E75* is a direct regulatory target of *Drosophila melanogaster* E93 (*Dm*E93) [21]. *Usp* was listed in Table 1, and the peak signal was detected in the 5' untranslated region of this gene (Fig. 2E).

Three distinct DNA binding motifs of Mblk-1

To search for Mblk-1-binding motifs, we next performed motif analysis using sequences corresponding to the Mblk-1-binding regions and MEME-ChIP [42]. Three consensus sequences were identified by DREME: GA-rich sequence (GAGA or GAG, hereafter termed a GAGA motif), GAAATTTT, and AT(C)TTTGTA (Fig. 3A). In addition, a combination of these motifs was found by MEME (Fig. 3B). In a previous study, we identified an Mblk-1 preferred binding sequence named MBE (Mblk-1-binding element, 5'-

AGGTAGAGATCGATCGATAGGG-3') by using an *in vitro* binding site selection method [17]. The motif we found in this study is consistent with MBE, because MBE has a GAGA sequence. In addition, the ATTTTC(T)GG motif is reported as a binding motif of *Dm*E93 [43], and it is also reported that *Bombyx mori* E93 (*Bm*E93) binds to GAGA-containing motifs [44]. These previous findings support the quality of our data and suggest that Mblk-1-binding motifs are conserved in insects.

We next performed a luciferase assay to examine whether Mblk-1 recognizes these 3 binding motifs and affects the transcription of a reporter gene. Human HEK293T cells were transfected with an Mblk-1 expression vector; a *Renilla* luciferase reporter vector; and 5 types of pGL4.23 firefly luciferase reporter vectors, each of which contains 6xGAGA, 6xGAGA with 4-bp spacer sequences, 6xATTTTG, 6xGAAATTTT, or 6xATTTTGG upstream of the minimal promoter, respectively (Fig. 3C). Normalized firefly luciferase activity increased about 3.0-fold and 6.6-fold when co-transfected with pGL4.23 vectors containing 6xGAGA or 6xGAGA with 4-bp spacer sequences, respectively, compared with that when co-transfected with a pGL4.23 vector containing no binding motifs (Fig. 3D). This indicates that Mblk-1 recognizes GAGA motifs and upregulates the transcription of genes that have these motifs. By contrast, normalized firefly luciferase activity significantly decreased by approximately 60% when co-transfected with pGL4.23 vectors containing 6xGAAATTTT, compared with that when co-transfected with a pGL4.23 vector containing no binding motifs (Fig. 3D). These results strongly suggested that Mblk-1 recognizes GAAATTTT motif to affect the transcription of genes with these motifs. On the other hand, no significant decrease in firefly luciferase activity was observed when co-transfected with a pGL4.23 vector containing 6xATTTTG or 6xATTTTGG (Fig. 3D), although these results do not exclude the possibility that Mblk-1 binds to these motifs. Taken together, our results indicated that Mblk-1 affects transcription by recognizing the identified Mblk-1-binding motifs.

Identification of genes upregulated in lKCs as candidates for Mblk-1 target genes

Given that Mblk-1 upregulates its target genes in a lKC-preferential manner in the honey bee MBs, Mblk-1 target gene candidates are expected to be expressed preferentially in lKCs. To find genes expressed preferentially in lKCs among the identified Mblk-1 target gene candidates, we used recently reported honey bee MB single-cell RNA-seq data [34]. When the MB cells were classified into 10 clusters (Fig. 4A), preferential Mblk-1 expression was observed in clusters 1, 3, and 5 (Fig. 4B). In addition to *Mblk-1*, genes shown to be preferentially expressed in lKCs by *in situ* hybridization, such as *Ryr* (cluster 1) [45], *IP3R* (clusters 1, 3, and 5) [14], *disks large homolog 5* (clusters 1 and 3) [36], and *broad-complex* (cluster 3) [46] were also more highly expressed in clusters 1, 3, and/or 5 than in other clusters (Fig. 4B). Therefore, we regarded these 3 clusters as lKCs.

We newly identified 12 Mblk-1 target gene candidates with higher expression in at least one of clusters 1, 3, and 5 than in the other clusters (Fig. 4B). *In situ* hybridization analysis of genes with the top 6 peak signals was performed to confirm that these genes are expressed preferentially in lKCs. As expected, *CaMKII*, *Baiap3* and *pumilio homolog 2* were expressed preferentially in lKCs (Fig. 4D, E and H), and *CUGBP Elav-like family member 4* was expressed in the whole MBs, but not in the other brain regions

(Fig. 4F). For the other 2 genes, specific expression in lKCs or MBs was not observed (Fig. 4C and G). Schematic diagrams of the peak signals and their surrounding Mblk-1 target gene candidates are shown in Fig. S2.

Comparison of Mblk-1 target gene candidate profiles between pupal and adult worker honey bee brains

Mblk-1 expression level is higher in pupal brains than adult brains [18]. In addition, *Dm*E93 is expressed transiently in pupal MB neuroblasts to activate autophagy during metamorphosis [24]. Thus, we next examined whether Mblk-1 has distinct target genes between adult and pupal worker brains in the honey bee. To this end, we performed ChIP-seq analysis using pupal worker brain homogenates and the anti-Mblk-1 antibody. A total of 263 genes were identified as pupal Mblk-1 target gene candidates, for which Mblk-1-binding signals are located within 10 kb upstream or downstream of the corresponding genes when using a peak calling q-value threshold of 0.005 and the input DNA sample as a control. Approximately half of the pupal Mblk-1 target gene candidates (138/263, 52%) were detected specifically in the ChIP-seq analysis using pupal brains (Fig. 5A and Supplementary Data 2&3). The Mblk-1-binding signals with the 20 lowest q-values (top 20 signals) and genes that have these signals within 10 kb upstream or downstream are shown in Fig. 5B. Among the 20 pupal target gene candidates shown in Fig. 5B, 8 genes were pupal-specific (8/20, 40%). For almost half of the 20 signals (9/20, 45%), no genes were localized within 10 kb upstream or downstream of the signals. For example, a neural developmental gene, *Down syndrome cell adhesion molecule* (*Dscam*) shown in Fig. 5B was not present among the adult target gene candidates, suggesting that this gene is a pupal-specific target. The peak signal was found in the first intron of *Dscam*, and only the GAGA motif was found in the Mblk-1 binding region. In addition to *Dscam*, several developmental genes, including *Ultrabithorax* and *Tyrosine-protein kinase transmembrane receptor Ror*, were specifically detected in the pupal analysis (Supplementary Data 2).

For the 6 genes analyzed by *in situ* hybridization, *potassium voltage-gated channel subfamily KQT member 1*, *CaMKII*, *CUGBP Elav-like family member 4*, *neural-cadherin*, and *pumilio homolog 2* were listed as the common target gene candidates. *Baiap3* was the only adult-specific gene. Although *CaMKII* is reported to regulate learning and memory abilities in adult workers [26, 27], it is likely that *CaMKII* is regulated by Mblk-1 in both pupal and adult brains and is involved in neural development through regulating synaptic plasticity in pupae. All the Mblk-1 target genes specific for adult MBs or pupal brains and the common target gene candidates are shown in Supplementary Data 1, 2, and 3, respectively.

Self-regulation of Mblk-1 expression in honey bee MBs

Interestingly, two distinct peak signals were found 2 kb and 25 kb upstream of the *Mblk-1* gene, respectively (Fig. 6A). All these signals contained GAGA motifs, which induced the expression of the downstream *luciferase* gene in the reporter assay (Fig. 3D). From this result, we expected that Mblk-1 induces its own expression *via* these GAGA motifs. To confirm this, we performed a luciferase assay with a pGL4.23 firefly luciferase reporter vector that contains the sequence corresponding to the peak signal located 25 kb upstream of *Mblk-1* gene (Fig. 6A). We observed an approximately 4.9-fold increase in firefly luciferase activity when using the pGL4.23 vector containing the peak signal sequence compared

with that when using the pGL4.23 vector containing no binding motifs (Fig. 6B). Therefore, Mblk-1 is suggested to recognize this GAGA-containing region as an enhancer to upregulate its own expression. It is possible that once *Mblk-1* is expressed, it can upregulate its own expression, which leads to its constitutive expression in adult honey bee lKCs.

Expression analysis of *Mblk-1* homolog in the sawfly

In the honey bee, *Mblk-1* is almost selectively expressed in the brain among various adult tissues [18]. In addition, in the ant *Harpegnathos saltator* workers and gamergates, the expression of the *Mblk-1* homolog in the brain is observed [47]. We next tested whether the constitutive expression of *Mblk-1* homologs in the adult brain is conserved even in other hymenopteran insect species (Fig. 6C). We performed qRT-PCR analysis to examine the expression of the *Mblk-1* homolog (LOC105691346) in various adult body parts of the sawfly (*Athalia rosae*), which is a solitary, phytophagous, and primitive hymenopteran insect species [40]. Although significant expression of the sawfly *Mblk-1* homolog was detected in the head without the brain, thorax, and abdomen, the expression in the brain was below the detection threshold (crossing point Cp: 40) (Table S2). The relative expression levels of the *Mblk-1* homolog did not differ significantly among body parts when the brain Cp values were set to 40 (Fig. 6D). These results indicate that the expression level of *Mblk-1* homolog is very low in the sawfly brain, and that brain-specific expression, which is observed in the honey bee, is not conserved in the sawfly.

Comparative analysis of the upstream sequences of *Mblk-1* homologs

Finally, we investigated the relationship between the accumulation of the GAGA motif (GAGA and GAG) and the expression of *Mblk-1* homologs in the brains of *Drosophila melanogaster* and hymenopteran insects. We counted the number of GAGA motifs located within upstream of 2 kb and 10 kb from the transcription start site of each *Mblk-1* homolog (Fig. 6E and Fig. S3). In all the four species of Aculeata, the numbers of GAG sequences within upstream of 2 kb were over 100, and the numbers of GAGA sequences were over 30. On the other hand, in *Drosophila melanogaster* and *Athalia rosae*, either of which no *Mblk-1* homolog expression was observed in adult brains, the number of GAG sequences was 76 and 41, respectively. In addition, to determine the evolutional stage in hymenopteran insects at which the accumulation of GAGA motifs was acquired, we investigated the numbers of GAGA motifs in the 2 parasitoid wasps, *Nasonia vitripennis* and *Fopius arisanus*. The numbers of GAGA motifs in *Nasonia vitripennis* were comparable to those in Aculeata, while the numbers in *Fopius arisanus* were closer to those in *Drosophila melanogaster* and *Athalia rosae*. The same trend was observed in the numbers of GAGA motifs within upstream of 10 kb (Fig. S3).

Discussion

Mblk-1 targets genes related to synaptic plasticity and ecdysone-signaling in adult honey bee MBs

Genes related to neural function, such as *CaMKII*, *Baiap3* and *pumilio homolog 2*, were newly identified as Mblk-1 target gene candidates in the worker honey bee MBs by ChIP-seq analyses. *CaMKII* and *Baiap3*

are involved in synaptic functions [25–27, 48]. In addition, *D. melanogaster pumilio* has a role in synaptic plasticity and learning ability [49]. Therefore, it is likely that Mblk-1 is involved in neural function, such as learning and memory abilities through transcriptional regulation of these synaptic plasticity-related genes (Fig. 4).

Several ecdysone-related genes such as *Usp* were also identified as Mblk-1 target gene candidates. In forager bees (> 21 days), the expression of *Usp* is stronger in lKCs than in the inner compact KCs, which correspond to mKCs and sKCs [50]. It is likely that *Usp* is regulated differently between lKCs and inner compact KCs, and its expression in lKCs is controlled by *Mblk-1*. It is plausible that *Usp* has some neural functions that differ from its developmental functions like *Mblk-1*. This is consistent with reports that ecdysone and ecdysone-related genes are involved in neural functions, including learning and memory abilities, in *Drosophila melanogaster* [51].

The honeybee Mblk-1 has three DNA binding motifs that are shared with *Dm*E93 or *Bm*E93

We identified three Mblk-1-binding motifs (GAGA, GAAATTTT, and AT(C)TTTGTA; Fig. 3A and B). The luciferase assay suggested that the GAGA and GAAATTTT motifs have different functions in transcriptional regulation by Mblk-1. GAGA and ATTTTGG motifs are suggested to be binding motifs for *Bm*E93 and *Dm*E93, respectively [43, 44], indicating that these two Mblk-1-binding motifs are conserved in insects. To the best of our knowledge, however, only AWTTTYGG and HTTTBGG, both of which are like ATTTTGG, are reported as *Dm*E93-binding motifs in Fly Factor Survey and previous research using ChIP-seq analysis [43]. It is thus likely that Mblk-1 possesses a binding motif (the GAGA motif) that is not observed in *Dm*E93. In the luciferase assay experiments, the expression of a reporter gene was upregulated when using the reporter vector containing the 6xGAGA sequence (Fig. 3D). Therefore, it is probable that Mblk-1 recognizes the GAGA motifs and upregulates the expression of genes around them in lKCs.

Distinct functions of Mblk-1 between pupal and adult brains

We identified pupal-specific Mblk-1 target gene candidates (Supplementary Data 2), which include some neural developmental genes. Approximately half of the target gene candidates for pupal brains were not identified as Mblk-1 target gene candidates for adult MBs (Fig. 5), suggesting that Mblk-1 alters its role by changing its target genes as honey bees develop from pupae to adults. There are two possible explanations for this change. One is that target genes of Mblk-1 that functions in regions other than MBs, such as OLs and ALs where Mblk-1 is expressed in early pupal stage [36], were included since we used pupal whole brains for the ChIP-seq analysis. The other possible explanation is that Mblk-1 regulates different genes when it is phosphorylated. Mblk-1 is phosphorylated by mitogen-activated protein kinase [33] in a pupae-specific manner [18]. The anti-Mblk-1 antibody was designed to recognize both non-phosphorylated and phosphorylated Mblk-1 [18]. It is thus possible that Mblk-1 changes its binding regions and target genes depending on phosphorylation.

Self-regulation: a possible model for constitutive expression of *Mblk-1* in adult honey bee MBs

Two distinct peak signals were found upstream of the *Mblk-1* gene, and the GAGA motif (GAGA or GAG) is enriched in those signal regions. The results of the luciferase assay indicated that the DNA sequence corresponding to one of the peak signals was recognized by Mblk-1 and the expression of its downstream reporter gene was upregulated (Fig. 6B). Generally, in insect metamorphosis, *E93* is transiently induced by ecdysone and ecdysone receptors [20]. On the other hand, Mblk-1 is constitutively expressed in honey bee lKCs [18], whereas *ecdysone receptor* (*EcR*) is mainly expressed in sKCs in the honey bee MBs [52]. Therefore, together with the result of the luciferase assay, it is likely that the constitutive expression of Mblk-1 in lKCs is, at least partially, accounted for by self-regulation instead of ecdysone-dependent regulation. To our knowledge, this is the first report demonstrating the self-induction of *Mblk-1*/*E93* in insects.

This finding also suggests that the accumulation of the GAGA motif sequences upstream of *Mblk-1* is important in the self-induction of *Mblk-1*. The numbers of the GAGA motif sequences in species in which the constitutive expression of *Mblk-1* homologs in adult brains are observed (*Apis mellifera* and *Harpegnathos saltator* [47]) were much greater than those in *Athalia rosae*. Consistent with this, qRT-PCR analysis using sawfly body parts indicated that the *Mblk-1* homolog is not selectively expressed in the brains of adult sawflies (Fig. 6D). For *Nasonia vitripennis*, the GAGA motifs are accumulated upstream of the *Mblk-1* homolog, although we have not analyzed whether their *Mblk-1* homologs are preferentially expressed in the adult brain. If the GAGA motif accumulation and the constitutive expression of *Mblk-1* homologs in brains has been acquired in parasitoid wasps, the role of *Mblk-1* homologs related to neural functions, including learning and memory abilities, might have been a platform for the regulation of parasitic behaviors and prerequisite for further advanced behaviors of aculeate species.

In summary, although Mblk-1/E93 is known as a transcription factor required for metamorphosis in insects [21–24], the findings of the present study indicate that the roles of Mblk-1 have expanded to regulate synaptic plasticity. By acquiring new expression mechanisms, binding motifs, and target genes during the evolution of hymenopteran insects, Mblk-1 may contribute to neural development in pupae and the advanced learning and memory abilities in adults. Our study provides an example of molecular evolution that could contribute to the acquisition of novel MB function in insects. Similar mechanisms may be at work in the brains of animals in general, and our findings could contribute to a better understanding of the evolution of brains and behaviors.

Declarations

Data Availability

ChIP-seq data in this study were deposited in the GEO with the accession number GSE173409. The link for data access is as follows: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173409.

We declare that all data supporting the findings of this study are available within the article and its supplementary information files.

Acknowledgements

We are grateful to Mr. Takayoshi Kuwabara for his help in using sawflies.

Funding

This work was supported by Grant-in-Aid for Scientific Research(B) (JSPS KAKENHI Grant Number 20H03300) from Japan Society for the Promotion of Science.

Author Contributions

Y. Matsumura designed and performed all experiments except for NGS preparation and sequencing. T. K. To performed NGS preparation and sequencing. All analyzed and discussed the data. Y.M prepared the figures. Y.M, H.Kohno and T. Kubo drafted the manuscript and all wrote and approved the paper. T. Kakutani and T. Kubo supervised the study.

References

1. McGuire SE, Le PT, Davis RL. The role of *Drosophila* mushroom body signaling in olfactory memory. Science **293**: 1330–1333 (2001).
2. Mizunami M, Weibrecht JM, Strausfeld NJ. Mushroom Bodies of the Cockroach: Their Participation in Place Memory. J Comp Neurol. **402**: 520–37 (1998).
3. Schildberger K. Multimodal interneurons in the cricket brain: properties of identified extrinsic mushroom body cells. J Comp Physiol. **154**: 71–79 (1984).
4. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE. Influence of gene action across different time scales on behavior. Science **296**: 741–744 (2002).
5. Farris SM, Schulmeister S. Parasitoidism, not sociality, is associated with the evolution of elaborate mushroom bodies in the brains of hymenopteran insects. Proc Biol Sci. **278**: 940–951 (2011).
6. Oya S, Kohno H, Kainoh Y, Ono M, Kubo T. Increased complexity of mushroom body Kenyon cell subtypes in the brain is associated with behavioral evolution in hymenopteran insects. Sci Rep. **7**: 13785 (2017).
7. Winston ML. *The Biology of the Honeybee*; Cambridge: Harvard University Press (1987).
8. Seeley TD. *The Wisdom of the Hive: The Social Physiology of the Honey Bee Colonies*; Cambridge: Harvard University Press (1995).
9. Von Frisch K. *The Dance Language and Orientation of Bees*; Cambridge: Harvard University Press (1993).
10. Menzel R, Leboulle G, Eisenhardt D. Small brains, bright minds. Cell **124**: 237–239 (2006).
11. Giurfa M. Behavioral and neural analysis of associative learning in the honeybee: A taste from the magic well. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. **193**: 801–824 (2007).

12. Kaneko K, Suenami S, Kubo T. Gene expression profiles and neural activities of Kenyon cell subtypes in the honeybee brain: identification of novel 'middle-type' Kenyon cells. Zool Lett. **2**: 14 (2016).
13. Kamikouchi A, Takeuchi H, Sawata M, Natori S, Kubo T. Concentrated expression of Ca^{2+}/calmodulin-dependent protein kinase II and protein kinase C in the mushroom bodies of the brain of the honeybee *Apis mellifera* L. J Comp Neurol. **417**: 501–510 (2000).
14. Kamikouchi A, Takeuchi H, Sawata M, Ohashi K, Natori S, Kubo T. Preferential expression of the gene for a putative inositol 1,4,5-trisphosphate receptor homologue in the mushroom bodies of the brain of the worker honeybee *Apis mellifera* L. Biochem Biophys Res Commun. **242**: 181–186 (1998).
15. Neves G, Cooke SF, Bliss TV. Synaptic plasticity, memory and the hippocampus: A neural network approach to causality. Nat Rev Neurosci. **9**: 65–75 (2008).
16. Takeuchi H, Kage E, Sawata M, Kamikouchi A, Ohashi K, Ohara M, et al. Identification of a novel gene, *Mblk-1*, that encodes a putative transcription factor expressed preferentially in the large-type Kenyon cells of the honeybee brain. Insect Mol Biol. **10**: 487–494 (2001).
17. Park JM, Kunieda T, Takeuchi H, Kubo T. DNA-binding properties of Mblk-1, a putative transcription factor from the honeybee. Biochem Biophys Res Commun. **291**: 23–8 (2002).
18. Kumagai H, Kunieda T, Nakamura K, Matsumura Y, Namiki M, Kohno H, et al. Developmental stage-specific distribution and phosphorylation of Mblk-1, a transcription factor involved in ecdysteroid-signaling in the honey bee brain. Sci Rep. **10**: 8735 (2020).
19. Takayanagi-Kiya S, Kiya T, Kunieda T, Kubo T. Mblk-1 Transcription Factor Family: Its Roles in Various Animals and Regulation by NOL4 Splice Variants in Mammals. Int J Mol Sci. **18**: 246 (2017).
20. Baehrecke EH, Thummel CS. The *Drosophila* E93 gene from the 93F early puff displays stage- and tissue-specific regulation by 20-hydroxyecdysone. Dev Biol. **171**: 85–97 (1995).
21. Lee CY, Wendel DP, Reid P, Lam G, Thummel CS, Baehrecke EH. E93 directs steroid-triggered programmed cell death in *Drosophila*. Mol Cell. **6**: 433–443 (2000).
22. Lee CY, Baehrecke EH. Steroid regulation of autophagic programmed cell death during development. Development. **128**: 1443–1445 (2001).
23. Ureña E, Manjón C, Franch-Marro X, Martín D. Transcription factor E93 specifies adult metamorphosis in hemimetabolous and holometabolous insects. Proc Natl Acad Sci USA. **111**: 7024–7029 (2014).
24. Pahl MC, Doyle SE, Siegrist SE. E93 Integrates Neuroblast Intrinsic State with Developmental Time to Terminate MB Neurogenesis via Autophagy. Curr Biol. **29**: 750–762.e3 (2019).
25. Shonesy BC, Jalan-Sakrikar N, Cavener VS, Colbran RJ. CaMKII: A molecular substrate for synaptic plasticity and memory. Prog Mol Biol Transl Sci. **122**: 61–87 (2014).
26. Matsumoto Y, Sandoz JC, Devaud JM, Lormant F, Mizunami M, Giurfa M. cGMP-gated channels, calmodulin, AC, and CamKII required for late, but not early, long-term memory formation in the honeybee. Learn Mem. **21**: 272–286 (2014).

27. Scholl C, Kübert N, Muenz TS, Rössler W. CaMKII knockdown affects both early and late phases of olfactory long-term memory in the honeybee. J Exp Biol. **218**: 3788–3796 (2015).
28. McNeill MS, Kapheim KM, Brockmann A, McGill TA, Robinson GE. Brain regions and molecular pathways responding to food reward type and value in honey bees. Genes Brain Behav. **15**: 305–317 (2016).
29. Wallberg A, Bunikis I, Pettersson OV, Mosbech MB, Childers AK, Evans JD, et al. A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genomics. **20**: 275 (2019).
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. **9**: 357–359 (2012).
31. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9**: R137 (2008).
32. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. **29**: 24–26 (2011).
33. Park JM, Kunieda T, Kubo T. The activity of Mblk-1, a mushroom body-selective transcription factor from the honeybee, is modulated by the ras/MAPK pathway. J Biol Chem. **278**: 18689–94 (2003).
34. Traniello IM, Bukhari SA, Kevill J, Ahmed AC, Hamilton AR, Naeger NL, et al. Meta-analysis of honey bee neurogenomic response links Deformed wing virus type A to precocious behavioral maturation. Sci Rep. **10**: 3101 (2020).
35. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell **177**: 1888–1902.e21 (2019).
36. Suenami S, Paul RK, Takeuchi H, Okude G, Fujiyuki T, Shirai K, et al. Analysis of the differentiation of Kenyon cell subtypes using three mushroom body-preferential genes during metamorphosis in the honeybee (*Apis mellifera* L.). PLoS ONE. **11**: e0157841 (2016).
37. Shields EJ, Sheng L, Weiner AK, Garcia BA, Bonasio R. High-Quality Genome Assemblies Reveal Long Non-coding RNAs Expressed in Ant Brains. Cell Rep. **23**: 3078–3090 (2018).
38. Geib SM, Liang GH, Murphy TD, Sim SB. Whole Genome Sequencing of the Braconid Parasitoid Wasp *Fopius arisanus*, an Important Biocontrol Agent of Pest Tepritid Fruit Flies. G3 (Bethesda). **7**: 2407–2411 (2017).
39. Dalla Benetta E, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS. Genome elimination mediated by gene expression from a selfish chromosome. Sci Adv. **6**: eaaz9808 (2020).
40. Oeyen JP, Baa-Puyoulet P, Benoit JB, Beukeboom LW, Bornberg-Bauer E, Buttstedt A, et al. Sawfly genomes reveal evolutionary acquisitions that fostered the mega-radiation of parasitoid and eusocial hymenoptera. Genome Biol Evol. **12**: 1099–1188 (2020).
41. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. **25**: 445–458 (2015).

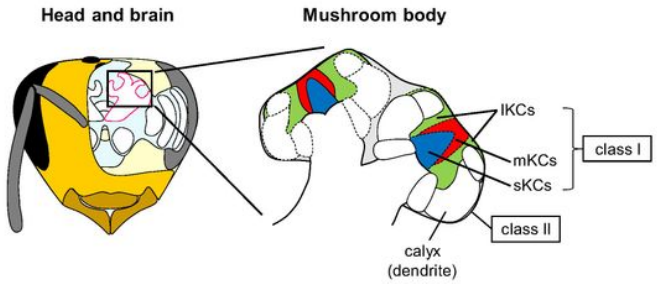
42. Machanick P, Bailey TL. MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics. **27**: 1696–1697 (2011).
43. Uyehara CM, Nystrom SL, Niederhuber MJ, Leatham-Jensen M, Ma Y, Buttitta LA, et al. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. **31**: 862–875 (2017).
44. Liu X, Dai F, Guo E, Li K, Ma L, Tian L, et al. 20-Hydroxyecdysone (20E) primary response gene E93 modulates 20E signaling to promote bombyx larval-pupal metamorphosis. J Biol Chem. **290**: 27370–27383 (2015).
45. Uno Y, Fujiyuki T, Morioka M, Kubo T. Mushroom body-preferential expression of proteins/genes involved in endoplasmic reticulum Ca^{2+}-transport in the worker honeybee (*Apis mellifera* L.) brain. Insect Mol Biol. **22**: 52–61 (2013).
46. Paul RK, Takeuchi H, Kubo T. Expression of two ecdysteroid-regulated genes, Broad-Complex and E75, in the brain and ovary of the honeybee (*Apis mellifera* L.). Zoolog Sci. **23**: 1085–1092 (2006).
47. Sheng L, Shields EJ, Gospocic J, Glastad KM, Ratchasanmuang P, Berger SL, et al. Social reprogramming in ants induces longevity-associated glia remodeling. Sci Adv. **6**: eaba9869 (2020).
48. Aravamudan B, Fergestad T, Davis WS, Rodesch CK, Broadie K. Drosophila UNC-13 is essential for synaptic transmission. Nat Neurosci. **2**: 965–971 (1999).
49. Dubnau J, Chiang AS, Grady L, Barditch J, Gossweiler S, McNeil J, et al. The staufen/pumilio pathway is involved in *drosophila* long-term memory. Curr Biol. **13**: 286–296 (2003).
50. Velarde RA, Robinson GE, Fahrbach SE. Nuclear receptors of the honey bee: Annotation and expression in the adult brain. Insect Mol Biol. **15**: 583–595 (2006).
51. Ishimoto H, Sakai T, Kitamoto T. Ecdysone signaling regulates the formation of long-term courtship memory in adult *Drosophila melanogaster*. Proc Natl Acad Sci U S A. **106**: 6381–6386 (2009).
52. Takeuchi H, Paul RK, Matsuzaka E, Kubo T. *EcR-A* expression in the brain and ovary of the honeybee (*Apis mellifera* L.). Zoolog Sci. **24**: 596–603 (2007).

Tables

Table 1 is available in the Supplemental Files section.

Figures

A



B

Gene ID	Gene description	Gene ID	Gene description
1. LOC102656354	uncharacterized protein	21. LOC724211	cytochrome P450 9e2
2. Tsf1	transferrin 1	22. LOC408951	14-3-3 protein epsilon
3. LOC102653599	uncharacterized ncRNA	23. Ant	ADP/ATP translocase
4. LOC408695	uncharacterized protein	24. LOC726472	prohormone-3
5. Mir6058	microRNA 6058	25. LOC551859	calmodulin
6. LOC408696	uncharacterized protein	26. LOC724405	protein lethal(2)essential for life
7. Mir6037	microRNA 6037	27. LOC409326	40S ribosomal protein S2
8. LOC409077	aminopeptidase N	28. LOC102655329	14-3-3 protein zeta
9. LOC725960	neurofilament heavy polypeptide	29. LOC409791	cAMP-dependent protein kinase catalytic subunit
10. LOC408388	tubulin alpha-1 chain	30. LOC409675	polyubiquitin-A
11. Mir6054	microRNA 6054	31. LOC408782	tubulin beta-1 chain
12. Hsc70-4	heat shock protein cognate 4	32. Fabp	FABP-like protein
13. LOC552326	uncharacterized protein	33. LOC726321	uncharacterized protein
14. LOC552734	poly(U)-specific endoribonuclease homolog	34. LOC409536	arginine-glutamic acid dipeptide repeats protein
15. Mir9883	microRNA 9883	35. Ef-1a-f1	translation elongation factor eEF-1 alpha chain
16. Ancr-1	AncR-1 non-coding nuclear RNA	36. LOC410467	uncharacterized protein
17. LOC100577092	uncharacterized protein	37. LOC724274	protein lethal(2)essential for life
18. Mir6049	microRNA 6049	38. Mblk-1	transcription factor mblk-1-like
19. LOC413324	chitinase-like protein Idgf4	39. LOC725970	protein still life, isoform SIF type 1
20. LOC408452	cytochrome P450 9e2	40. LOC409634	uncharacterized

Figure 1

Analysis of genes highly expressed in adult worker MBs. (A) Schematic drawings of the head, the brain, and the MBs. lKCs, large-type KCs shown in green; mKCs, middle-type KCs shown in red; sKCs, small-type KCs shown in blue. (B) Top 40 genes highly expressed in adult worker MBs.

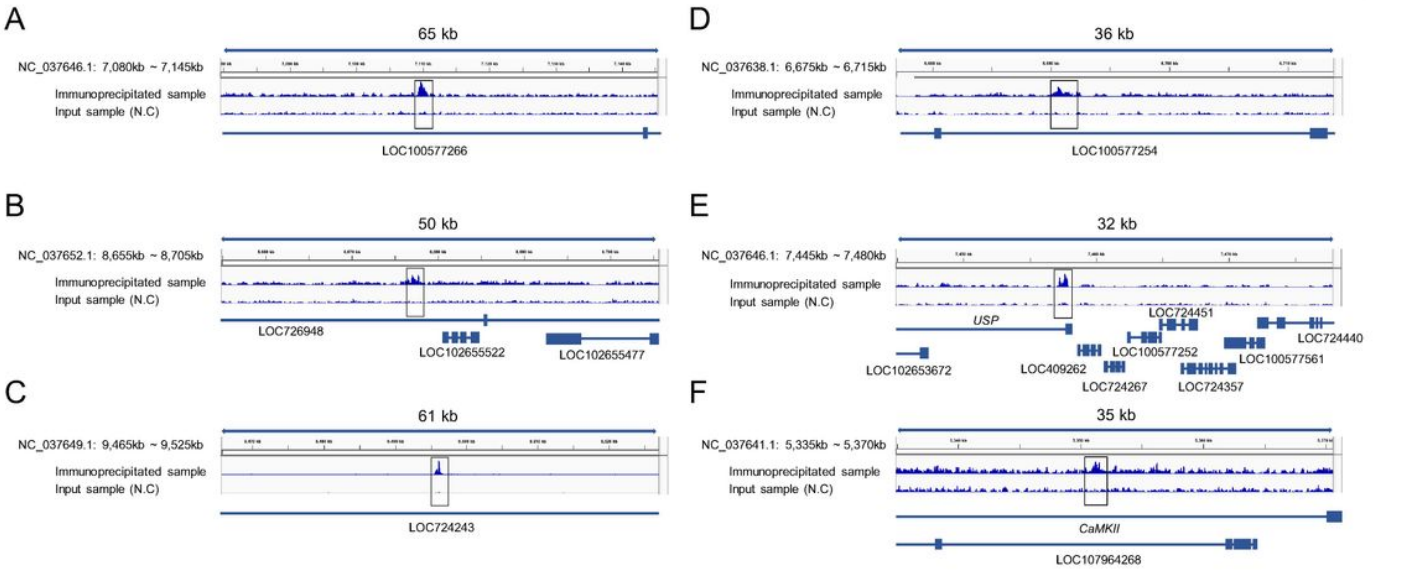


Figure 2

Mblk-1-binding signals and their surrounding genomic regions. The waveforms in each panel show the accumulation of reads in the corresponding genomic region when using samples immunoprecipitated by the anti-Mblk-1 antibody (upper lane), and input samples (lower lane, as negative control). The Mblk-1-binding signals are boxed with black lines in the waveform lanes. Bars and boxes under the waveforms indicate exons and introns, respectively, annotated within the corresponding genomic regions. The gene ID of each gene is shown below the gene structure. Panels A to F correspond to Gene Nos. 1, 2, 4, 9, 14 and 17 listed in Table 1, respectively.

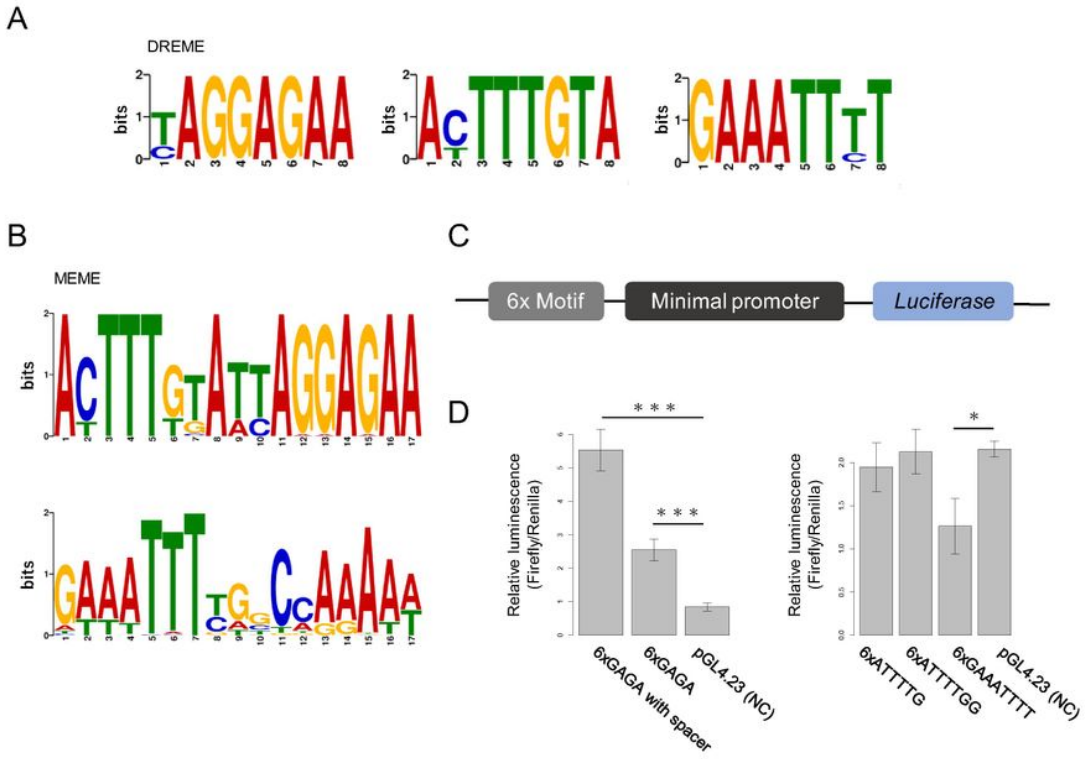


Figure 3

Mblk-1-binding motif analysis and luciferase assay. (A) Three Mblk-1-binding motifs identified with DREME. (B) Two Mblk-1-binding motifs identified with MEME. (C) pGL4.23 reporter vector construction used for the luciferase assay. (D) The results of the luciferase assay using reporter vectors containing motif 1: 6xGAGA or 6xGAGA with 4-bp spacer sequence (left panel) and motif 2: 6xGAAATTTT or motif 3: 6xATTTTGG sequence (right panel). Data represent means ± SD (n=3). Significant differences are indicated by asterisks (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.005$, Tukey-Kramer test).

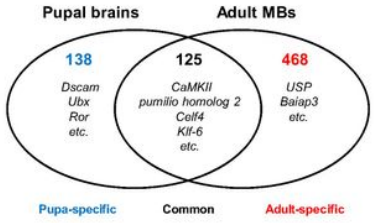
Figure 4

Mblk-1 target gene candidates preferentially expressed in lKCs identified by single-cell RNA-seq data and *in situ* hybridization analysis. (A) UMAP visualization of single-cell RNA-seq data obtained from adult honey bee MB in Traniello *et al.*, 2020 [34]. (B) The 13 Mblk-1 target gene candidates identified to be more highly expressed in clusters 1, 3, and/or 5 than in other clusters. Genes shown in red are already shown to be expressed preferentially in lKCs. (C-H) *in situ* hybridization results for the 6 genes. (C) *potassium voltage-gated channel subfamily KQT member 1* (D) *Ca^{2+}/calmodulin dependent protein kinase II* (E) *BAI1-associated protein 3* (F) *CUGBP Elav-like family member 4* (G) *neural-cadherin* (H) *pumilio homolog 2*. The upper panels show results using anti-sense RNA probes, whereas the lower panels show results using sense RNA probes. Schematic drawing of the expression pattern of each gene in the MB is shown

at the bottom. MB, mushroom bodies; OL, optic lobes; AL, antennal lobes; lKCs, large-type KCs. Scale bar: 250 μm. Note that genes that do not have a GB, which is the gene symbol in BEEBASE and was used in the single-cell RNA-seq analysis, are not included in this analysis, even if they are preferentially expressed in the three clusters.

Fig. 5

A



B

Gene No	Gene ID	Gene description	Signal fold enrichment	-log$_{10}$ (q-value)	Signal position
1.	LOC724243	*fasciclin-3*	6.08	50.9	intron
2.	-	-	10.9	50.7	
3-1.	LOC726948	*suppressor of lurcher protein 1*	8.73	36.0	intron
3-2.	LOC102655522	uncharacterized ncRNA			upstream
4.	LOC100577266	uncharacterized ncRNA	7.75	30.7	intron
5.	LOC552397	*Krueppel-like factor 6*	4.92	30.2	intron
6.	-	-	6.21	26.7	
7.	LOC100577254	*potassium voltage-gated channel subfamily KQT member 1*	6.58	20.1	intron
8.	LOC552844	*complexin*	2.95	20.0	intron
9-1.	LOC413171	*myosin-I heavy chain*	5.61	19.6	intron
9-2.	LOC102656705	uncharacterized ncRNA			downstream
10.	-	-	2.68	17.8	
11.	-	-	6.16	16.8	
12.	-	-	5.76	15.6	
13.	-	-	2.50	15.2	
14.	-	-	2.63	14.8	
15-1.	*Dscam*	*Down syndrome cell adhesion molecule*	5.25	14.0	intron
15-2.	LOC409115	*probable dolichol-phosphate mannosyltransferase*			upstream
15-3.	LOC100578015	*centrosomal protein of 104 kDa*			upstream
15-4.	LOC411579	*tRNA (cytosine(34)-C(5))-methyltransferase*			downstream
16.	-	-	5.72	14.0	
17.	LOC551957	*headcase protein*	4.11	13.7	intron
18.	-	-	4.57	13.6	
19-1.	LOC102653663	uncharacterized	4.91	13.2	intron
19-2.	LOC726825	*homeobox protein HMX2*			intron
19-3.	LOC100577551	*eukaryotic translation initiation factor 5B*			downstream
19-4.	LOC409611	*disintegrin and metalloproteinase domain-containing protein 10*			upstream
20-1.	LOC411361	*protein tramtrack, beta isoform*	5.59	13.2	upstream
20-2.	*FAR1*	*fatty acyl-CoA reductase 1*			downstream

Figure 5

Identification of Mblk-1 target gene candidates in pupal brains. (A) Comparison of Mblk-1 target gene candidates between pupae and adults. (B) Top 20 ChIP-seq peak signals from pupal worker brain homogenates and Mblk-1 target gene candidates located within ±10 kb of the corresponding signals. Gene ID, gene description, signal fold enrichment, -log$_{10}$ (q-value), and signal position relative to the genes are shown for each ChIP-seq peak signal. - indicates that there are no genes located within ±10 kb of any peak signal.

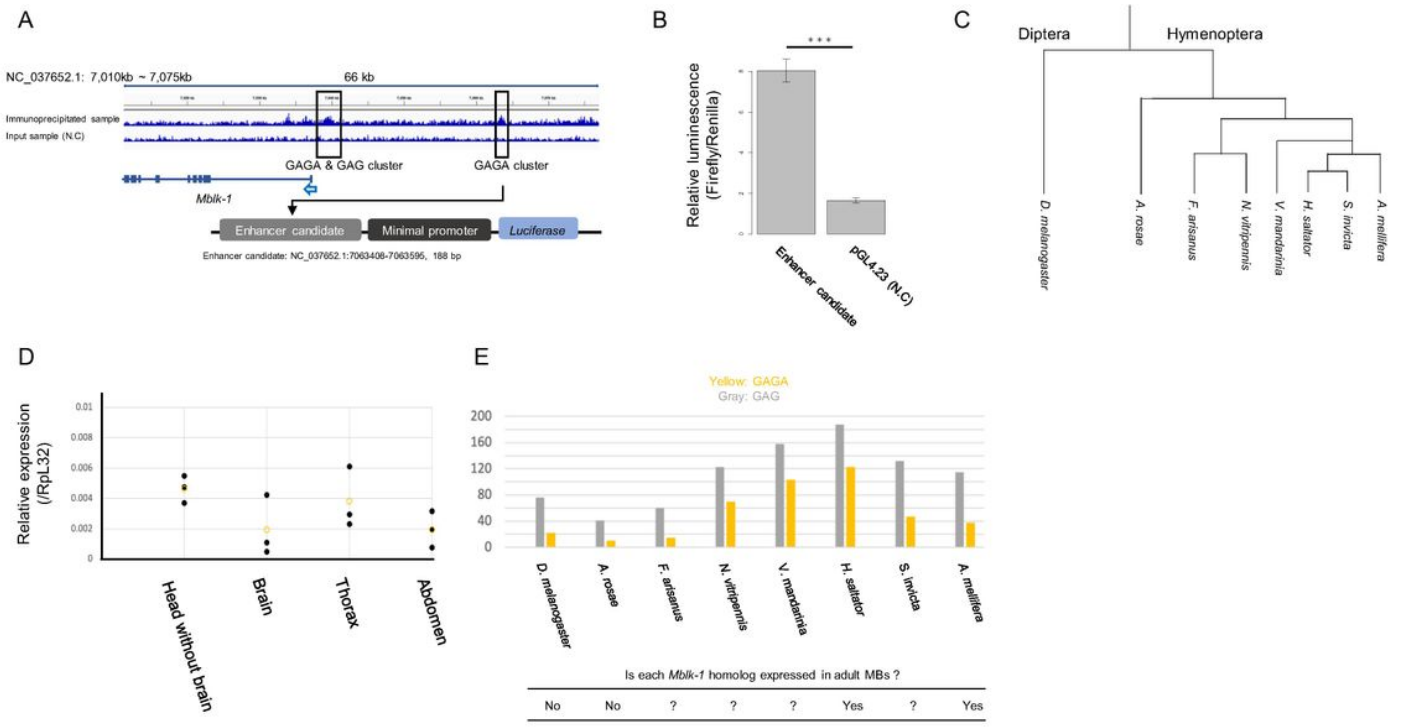


Figure 6

Mblk-1 upregulates its own expression *via* the GAGA motif-containing regions. (A) ChIP-seq peak signals were identified upstream of the *Mblk-1* gene (indicated by black outlines). The construction of the pGL4.23 reporter vector used in the luciferase assay is indicated below the *Mblk-1* gene structure. The start site of the *Mblk-1* gene is indicated by a blue arrow. (B) Functional interaction of Mblk-1 and the ChIP-seq peak signal region. Data represent means ± SD (n=4). Significant difference is indicated by 3 asterisks ($p < 0.005$, Student's t test). (C) Phylogenetic tree showing the relationship among Hymenoptera and *Drosophila melanogaster*. (D) Comparison of the relative expression of the *Athalia rosae Mblk-1* homolog with respect to *RpL32* among various adult body parts (head without brain, brain, thorax, and abdomen, n=3) of the *Athalia rosae*. Yellow circles represent average values. Because expression of the *Mblk-1* homolog in the brain samples was below the detection threshold, brain Cp values were set to 40 in a total of 45 PCR cycles and the relative expression levels were calculated. (E) The relationship between the number of GAGA motifs within 2 kb upstream of *Mblk-1* homologs and expression of *Mblk-1* homologs in the adult MBs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MatsumuraSupplFigs.SciRep220508.pptx
- MatsumuraSupplData1SciRep20220430final.xlsx

- MatsumuraSupplData2SciRep20220430final.xlsx
- MatsumuraSupplData3SciRep20220430final.xlsx
- Table1.jpg