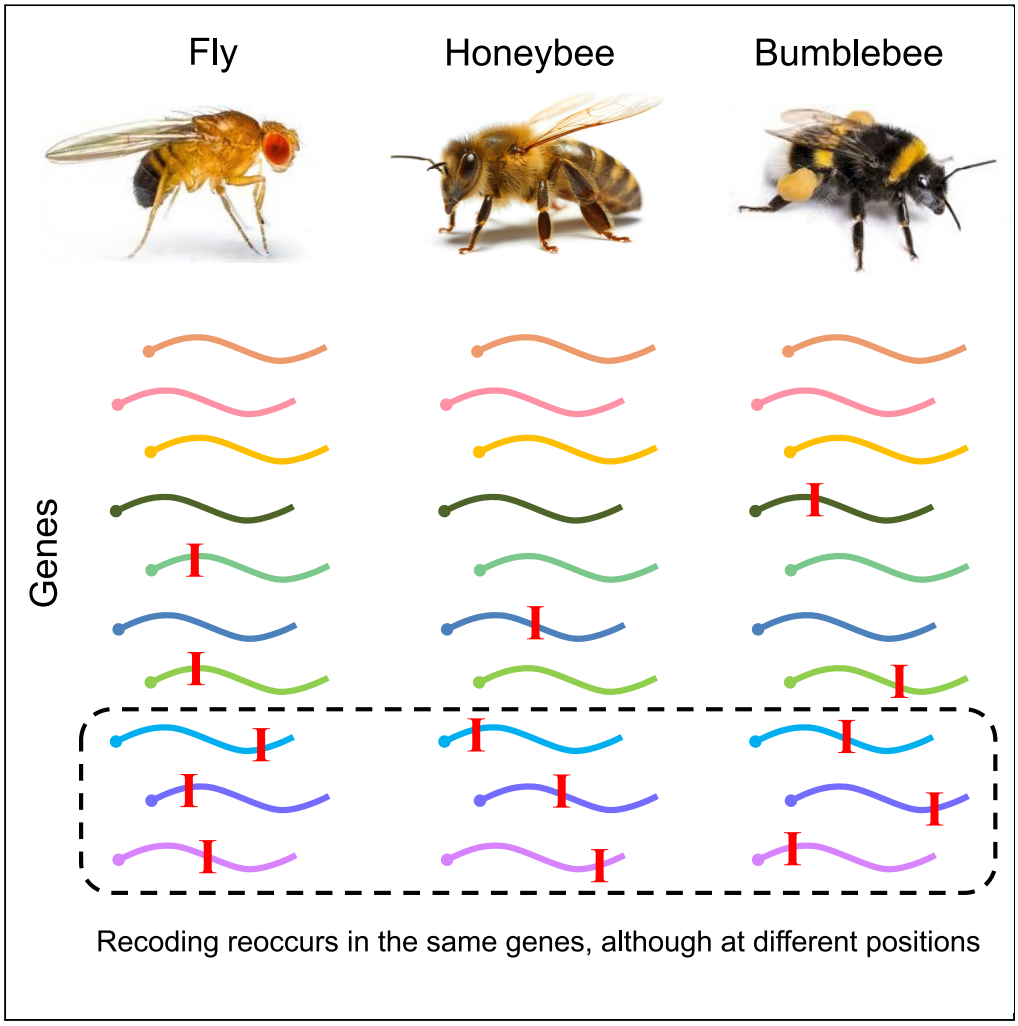
iScience

CellPress
OPEN ACCESS

Article

A-to-I RNA editing in honeybees shows signals of adaptation and convergent evolution



Yuange Duan, Shengqian Dou, Hagit T. Porath, Jiaxing Huang, Eli Eisenberg, Jian Lu

elieis@post.tau.ac.il (E.E.)
luj@pku.edu.cn (J.L.)

Highlights

Nonsynonymous editing sites in honeybees were under positive selection

Differential editing may contribute to the phenotypic diversity between sub-castes

Target genes acquire editing in different clades, suggesting convergent evolution

Duan et al., iScience *24*, 101983
January 22, 2021
https://doi.org/10.1016/j.isci.2020.101983

iScience

CellPress
OPEN ACCESS

Article

A-to-I RNA editing in honeybees shows signals of adaptation and convergent evolution

Yuange Duan,[1,2] Shengqian Dou,[1] Hagit T. Porath,[3] Jiaxing Huang,[4] Eli Eisenberg,[5,*] and Jian Lu[1,6,*]

Summary

Social insects exhibit extensive phenotypic diversities among the genetically similar individuals, suggesting a role for the epigenetic regulations beyond the genome level. The ADAR-mediated adenosine-to-inosine (A-to-I) RNA editing, an evolutionarily conserved mechanism, facilitates adaptive evolution by expanding proteomic diversities. Here, we characterize the A-to-I RNA editome of honeybees (*Apis mellifera*), identifying 407 high-confidence A-to-I editing sites. Editing is most abundant in the heads and shows signatures for positive selection. Editing behavior differs between foragers and nurses, suggesting a role for editing in caste differentiation. Although only five sites are conserved between bees and flies, an unexpectedly large number of genes exhibit editing in both species, albeit at different locations, including the nonsynonymous auto-editing of *Adar*. This convergent evolution, where the same target genes independently acquire recoding events in distant diverged clades, together with the signals of adaptation observed in honeybees alone, further supports the notion of recoding being adaptive.

Introduction

Adenosine (A)-to-inosine (A-to-I) RNA editing, catalyzed by enzymes of the ADAR (adenosine deaminase acting on RNA) family (Savva et al., 2012b), is an evolutionarily conserved mechanism that expands RNA diversity at the co-transcriptional or post-transcriptional level in metazoans (Bass 2002; Nishikura 2010; Eisenberg and Levanon 2018). Two catalytically active *Adar* genes are encoded in most metazoans. However, insects have lost ADAR1 and encode only a single *Adar* gene (Keegan et al., 2011). Due to the structural similarity between inosine (I) and guanosine (G), I is generally believed to be recognized as G in many cellular processes such as mRNA splicing (Rueter et al., 1999; Flomen et al., 2004; Jin et al., 2007; Lev-Maor et al., 2007), microRNA (miRNA) biogenesis or target recognition (Liang and Landweber 2007; Borchert et al., 2009; Alon et al., 2012), and mRNA translation (Basilio et al., 1962; Licht et al., 2019). Therefore, an A-to-I RNA editing usually has a similar effect as an A-to-G DNA substitution. A-to-I editing plays essential roles in many biological and physiological processes (Keegan et al., 2001; Nishikura 2010), and dysregulation of A-to-I editing might be associated with cancer, autoimmune disorders, or other human diseases (Gallo et al., 2017).

During the past two decades, A-to-I RNA editing sites have been systematically characterized in various metazoan species (Ramaswami and Li 2016). The majority of RNA editing sites are located in clusters in non-coding regions of humans (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004; Picardi et al., 2017), monkeys (Chen et al., 2014; Yang et al., 2015), mice (Neeman et al., 2006; Danecek et al., 2012), worms (Morse and Bass 1999; Zhao et al., 2015; Goldstein et al., 2017), corals (Porath et al., 2017b), and many other species (Porath et al., 2017a). In most species, only a minute fraction of the edits resides within the coding sequence. Notable exceptions are *Drosophila* (Graveley et al., 2011; Rodriguez et al., 2012; St Laurent et al., 2013; Mazloomian and Meyer 2015; Yu et al., 2016; Buchumenski et al., 2017; Duan et al., 2017; Zhang et al., 2017) and cephalopod species (Alon et al., 2015; Liscovitch-Brauer et al., 2017). Despite the deep conservation of A-to-I editing mechanism, the target landscapes of editing have considerably evolved during metazoan evolution. Only one editing site is known to be conserved across virtually all mammals, *Drosophila*, and cephalopods (Porath et al., 2019), and it thus seems that RNA editing modulates the diversity of the transcriptomes and proteomes in a lineage-specific manner.

[1]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, China

[2]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China

[3]The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Ramat-Gan 52900, Israel

[4]Key Laboratory for Insect-Pollinator Biology of the Ministry of Agriculture and Rural Affairs, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, 100093, China

[5]Raymond and Beverly Sackler School of Physics and Astronomy and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel

[6]Lead contact

*Correspondence: elieis@post.tau.ac.il (E.E.), luj@pku.edu.cn (J.L.)

https://doi.org/10.1016/j.isci.2020.101983

The forces driving the evolution of A-to-I editing across species at the macro-evolutional scale are not well understood. RNA editing was hypothesized to facilitate adaptive evolution by increasing proteomic diversities temporally or spatially, in a manner more flexible than genomic mutations (Gommans et al., 2009; Nishikura 2010, 2016; Klironomos et al., 2013; Rosenthal 2015). However, in most species, the studied ratio of nonsynonymous *(N)* to synonymous (*S*) editing sites (*N/S*) is lower than expected for random sites, suggesting that recoding events may be overall non-adaptive (Xu and Zhang 2014). Here too *Drosophila* and cephalopods stand out as exceptions, exhibiting high *N/S* ratios indicating positive selection of recoding. Furthermore, hundreds of recoding sites were shown to be conserved across the *Drosophila* lineage (Graveley et al., 2011; Rodriguez et al., 2012; St Laurent et al., 2013; Mazloomian and Meyer 2015; Yu et al., 2016; Buchumenski et al., 2017; Duan et al., 2017; Zhang et al., 2017) and thousands in the behaviorally sophisticated cephalopods (Garrett and Rosenthal 2012; Alon et al., 2015; Liscovitch-Brauer et al., 2017), supporting the functional importance of these editing events. However, in only a few examples was the advantageous effect conferred by RNA editing explicitly demonstrated (see for example [Garrett and Rosenthal 2012]), and even it is not yet clear to what extent editing is indeed utilized for proteome diversification.

Social insects, including bees and ants, show extensive phenotypic plasticity. The morphologically and behaviorally differentiated social castes such as queens, workers, and drones have the same set of diploid or haploid genomes. The social insects provide us with model systems to study how phenotypic diversity is regulated (Page et al., 2012; Yan et al., 2014). RNA editing is well suited to contribute to the behavioral variation among genetically similar individuals. Indeed, differential editing was demonstrated for a few sites in leaf-cutting ants and in worker bumblebees (Li et al., 2014; Porath et al., 2019).

Here, we wish to study the contribution of RNA editing to caste differentiation in honeybee (*Apis mellifera*), an important pollinator and a model for complicated social behaviors in insects (Page et al., 2012). Unlike bumblebees, honeybees have a sharp task specialization, with distinct worker castes. Genetic mapping demonstrated that the phenotypic plasticity in honeybees is associated with complex epistatic and pleiotropic genetic networks that influence reproductive regulation and foraging behaviors (Page et al., 2012). We wish to check whether A-to-I RNA editing may contribute to the proteomic diversity underlying this phenotypic plasticity. We investigate the A-to-I RNA editome in different tissues of four honeybee drone individuals and detect over four hundred A-to-I sites in multiple tissues. Editing is enriched in the head and exhibits signs for positive selection and a particularly high *N/S* ratio. We show editing is elevated in foragers compared to nurses. Five editing sites are conserved between honeybee, bumblebee, and *Drosophila*. Interestingly, we find a significantly high number of cases where the same gene is edited in both bees and flies, even if not at the same position. One example is the auto-editing of *Adar* mRNA in bees and flies, which might play an auto-regulatory role in the two clades. This finding may further support possible convergent evolution and adaptation.

Results

Editome of honeybee

We deep sequenced genomic DNA and RNA from the head, thorax, and abdomen for each of four individual drones (12 RNA-seq samples) (Figure 1A). The transcriptome libraries of drones 1 and 2 were constructed by selecting polyA tailed mRNAs while the libraries of drones 3 and 4 were constructed by using the Ribo-Zero kit to deplete ribosomal RNAs (Transparent methods). The male drone is haploid, simplifying identification of its genomic single nucleotide polymorphisms (SNPs). DNA-Seq reads were mapped to the reference genome to identify 826,890-858,949 SNPs in each of the four drones and 1,303,225 unique sites combined (Transparent methods; Table S1). Median sequencing coverage at the SNP sites was 39, and in 96.6% of the sites, all DNA reads supported the SNP (Figure S1). The identified SNPs were then used to produce a masked genome version where the reference allele is substituted by the individual-specific alleles identified for each drone to facilitate accurate detection of RNA editing sites (Figure 1B; Transparent methods).

For each of the 12 RNA-Seq samples, 9.2-21.5 million reads were uniquely mapped to the reference genome, and similar numbers of reads were mapped to the masked genome sequences (Transparent methods; Table S2). Analyzing pooled data for each of the three tissues separately, we identified (false discovery rate [FDR] = 0.05) 376 (84.3%) A-to-G sites among 446 variations in heads, 106 (66.7%) A-to-G sites among 159 variations in thoraxes, and 137 (67.5%) A-to-G sites among 203 variations in abdomens (Figure S2). Combined, we obtained 407 (80.1%) unique A-to-G sites among 508 variations (Figure 1C). These

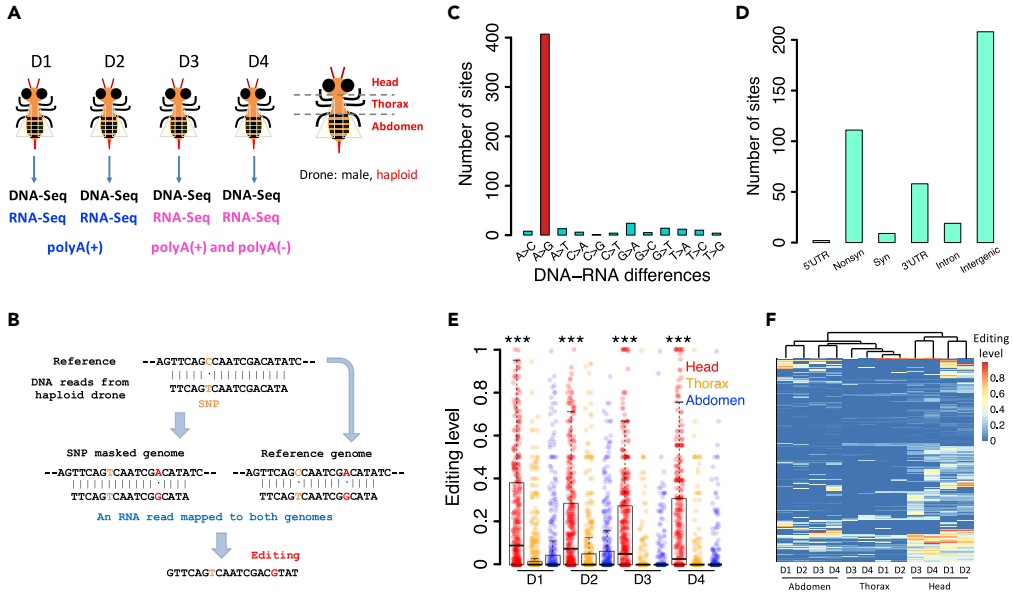


Figure 1. Identification and annotation of the A-to-I RNA editing sites in honeybee

(A) Workflow of sample collection, dissection, and library construction.

(B) Identification of A-to-I RNA editing sites is facilitated by mapping to both the reference genome and the masked genome.

(C) Distribution of DNA-RNA mismatch types detected in honeybee.

(D) Distribution of the detected A-to-I editing sites over different genic regions.

(E) Editing levels at the detected sites as measured in three tissues of each of four individuals (D1-D4). Editing levels are higher in the head (head versus pooled thorax and abdomen, Wilcoxon rank-sum test; ***: $P < 0.001$. Exact p values, left to right: 2.1×10^{-35}, 5.7×10^{-20}, 4.0×10^{-29}, and 3.9×10^{-38}).

(F) The editing profiles for the twelve tissues cluster according to their tissue of origin.

407 sites are regarded as A-to-I RNA editing sites (Table S3). The nucleotide context around these putative editing sites is consistent with the known ADAR binding motif (Figure S3). The median DNA coverage (over the 407 A-to-I editing sites) was 25 reads, sufficient to exclude possible SNPs (Figure S4).

To improve mapping accuracy, we discarded variants in repeat regions and required a variation site to be found by mapping to both the reference genome and the masked genome (Transparent methods). To test whether these criteria are too stringent, we checked the variation sites discarded by the two filtering steps. First, we looked at the variation sites located in the repeat regions. Without any filter, we found 19,573 unique variation sites that overlapped with repeat regions, only 3,694 (19%) of which were A-to-G variants. Following a binomial test and multiple testing correction (to remove random sequencing errors), 1,277 variation sites were maintained, only 324 (25%) of them A-to-G variations. Thus, filtering repetitive regions does contribute appreciably to the precision of our detection. Consistently, hyper-editing analysis did not reveal multiple sites in repetitive regions (Transparent methods).

As for the masked genome filter, only 38% (98/256) of the variation sites found by mapping to the masked genome but not to the reference genome are A-to-G mismatches. Following binomial test for sequencing error and multiple testing correction for p values, only 41 variation sites were maintained, none of them was an A-to-G variation. Conversely, only 31% (82/268) variation sites found by mapping to the reference genome but not to the masked genome are A-to-G mismatches. Of these 268 variation sites, 82 were maintained after multiple testing correction, 20 of them were A-to-G variations. Thus, most of the sites supported by only one method of mapping are likely not due to A-to-I RNA editing.

Among the 407 editing sites we identified, 199 sites are located in gene regions, including 111 nonsynonymous (*N*) sites, 9 synonymous (*S*) sites, and several sites in untranslated regions (UTRs) and introns (Figure 1D). The other 208 sites are annotated as intergenic. As expected from the different library construction strategies, the fraction of sites in coding regions is higher and the intergenic fraction is lower for sites

observed in drones 1 and 2 compared with drones 3 and 4 (23.2 ± 2.6% in coding sequence (CDS) in drones 1 and 2 versus 14.5 ± 5.1% in CDS in drones 3 and 4; 54.9 ± 2.2% in intergenic regions in drones 1 and 2 versus 70.5 ± 5.4% in intergenic regions in drones 3 and 4). The number of sites for which editing is observed in each sample (per-sample editing level >0) varies considerably (40–278 sites per sample; Table S4). The partial overlap between these sets of sites is mostly due to sites being undetected at a given sample due to low per-sample coverage (Figure S5).

Interestingly, editing seems to be enriched in heads. The number of detected sites (Figure S6), the editing levels (Figure 1E), and the fraction of coding sites (Figure S7) are all higher in heads. Clustering the samples by their editing profile, the three tissue types form distinct clusters (Figure 1F). Consistently, *Adar* expression is also higher in heads than in other tissues (Figure S8). Looking at specific sites, 78 sites (out of 378 sites with sufficient coverage, see Transparent methods) are differentially edited in the head, compared to non-head tissues. In all of these 78 sites, editing is higher in the head (Table S3. Information of the candidate editing sites identified in this study, Figure 1).

Signals of adaptation

The nonsynonymous to synonymous (*N/S*) ratio for sites in the coding region is 111/9 = 12.3 (Table S4), compared to a ratio of 2.26 observed for random adenosine to guanosine substitutions (see Transparent methods). This strongly suggests that the nonsynonymous editing events in honeybees are overall adaptive. Notably, the *N/S* is exceptionally high, even compared to the ratios we previously found in the brains of *Drosophila* (Duan et al., 2017) or in cephalopods (Alon et al., 2015; Liscovitch-Brauer et al., 2017). Looking at each sample separately, at pooled tissue data, or focusing on sites that appear in at least two individuals per tissue all show the *N/S* ratio to be especially high in heads (Figure 2A and Table S4). Furthermore, the editing level at nonsynonymous sites is significantly higher than synonymous sites editing levels in heads of honeybees (Figure 2B).

A previous study in cephalopods proposed that the coleoids massively edit their RNAs to diversify the transcriptome at the cost of constraining the evolution of genomic sequences (Liscovitch-Brauer et al., 2017). Maintaining beneficial editing requires genomic conservation of the genomic sequence encoding for the dsRNA structures that allow ADAR to bind and deaminate the adenosine. To test this, we first confirmed that the editing sites in honeybees are enriched in hairpin structures (Figure 2C). Then, we demonstrated that SNPs are depleted in the vicinity of nonsynonymous sites (Figure 2D), further supporting the notion of selective advantage for the nonsynonymous editing events that justifies the trade-off between transcriptome diversity and genome evolution.

Evolutionarily conserved and non-conserved editing sites

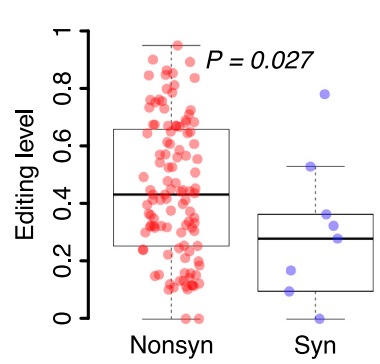
Editing at four nonsynonymous and one synonymous site is conserved between *Drosophila* and honeybees (Figure 3). Three of these sites are in the *Shab* transcript (*shaker cognate b*, two nonsynonymous and one synonymous) and the other two are in *qvr* (*quiver*). Editing at these sites is also observed in the bumblebee *Bombus terrestris* (Porath et al., 2019), and editing levels are high in all of these species (Figure 3). These findings suggest a potential functional role for these widely conserved editing sites. However, the majority of the editing sites are not shared between *Drosophila* and honeybees. For some sites, the editable adenosine is not conserved in the honeybee genome (Figure S9) and are thus clearly uneditable, while other adenosines are conserved but editing was not detected in RNA-Seq data (Figure S9). Overall, of the 639 *Drosophila* editing sites in CDS for which an orthologous site was identified in honeybees, 349 were conserved as A (55%), and 85, 127, and 78 (13%, 20%, and 12%) were mutated to C, G, and T, respectively.

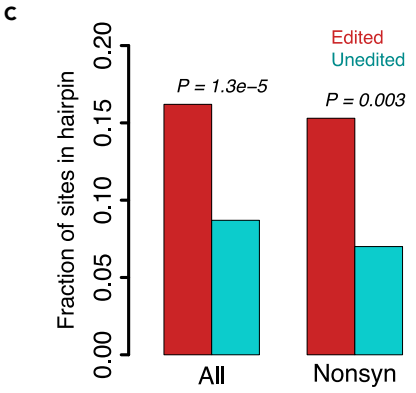
Similarly, most sites are not even conserved across the two bee species. The bumblebee study has reported 219 editing sites in CDS, 164 of which are nonsynonymous (Porath et al., 2019). Of these, we found 9 editing sites in coding regions conserved between honeybee and bumblebee (Table S5).

One of these conserved sites is a Ser > Gly site within the *tipE* (temperature-induced paralytic E; *GB47375*) transcript, editing levels of which correlate with task performance in bumblebees (nurses versus foragers) (Porath et al., 2019). Editing levels at this site in honeybee heads are ~0.7, similar to those observed in bumblebee brains. In the two other tissues, this gene is lowly expressed and poorly edited. The orthologous site in *Drosophila* genomes encodes the post-edit Gly codon GGC (Figure S10). Thus, *tipE* recoding seems to be bee specific, possibly related to social behavior and task performance.

A

	N	S	N/S	P value
All sites				
Head	108	8	13.5	8.5E-10
Thorax	14	5	2.8	0.81
Abdomen	16	1	16	0.032
In ≥ 2 individuals				
Head	80	8	10	2.2E-6
Thorax	11	4	2.75	1
Abdomen	13	1	13	0.078





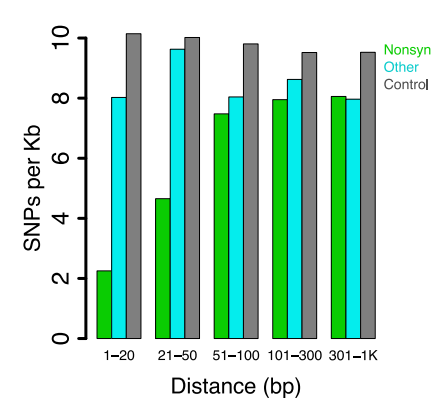


Figure 2. Signals of adaptation of A-to-I editing in honeybee

(A) The observed numbers of nonsynonymous (*N*) and synonymous (*S*) editing sites in each tissue. p values are obtained from Fisher's exact test, comparing the observed *N* and *S* counts to all adenosines in the coding sequence of in *A. mel* (*N* = 4,492,737 sites where A-to-G substitution is nonsynonymous and *S* = 1,986,130 for synonymous). "All sites" represents the numbers of sites that appear in at least one individual. Comparisons for sites appearing in at least two individuals are also shown.

(B) Editing levels of the nonsynonymous and synonymous sites in the head. Data from four drones were pooled to increase the statistical power. p values are calculated using Wilcoxon rank-sum test.

(C) The fraction of sites in predicted stable hairpin structures. The observed editing sites were compared to the unedited adenosines (p values using Fisher's exact test).

(D) The SNP density in the vicinity of nonsynonymous editing sites (green), other editing sites (cyan), or unedited adenosines (gray).

Sex-dependent and caste-dependent editing

We further looked for potential sex-dependence and caste-dependent editing, comparing our honeybee drone head data with previously published brain RNA-seq data studying two sub-castes of female workers, foragers and reverted nurses (Herb et al., 2012) (Table S6). First, we compared the pooled editing levels between drones and workers for each well-covered (>10 reads in each pool) editing site separately. Twenty six sites exhibit differential editing (Wilcoxon rank-sum test; FDR = 0.05) between workers (females) and drones (males), including 4 recoding sites, 2 sites in 3′UTRs, 3 intronic sites, and 17 intergenic sites (Figure 4A and Table S7). In most cases, editing in drones is higher, but this may result from looking at sites identified in drone data to begin with. The difference in *Adar* expression was not statistically significant (Figure S11).

Comparing the pooled editing levels between the two sub-castes (foragers and nurses), one finds 230 out of 407 sites with lower levels in nurses, compared to only 91 exhibiting higher levels in nurses (equal level, mostly zero, was observed in 84 sites), suggesting a globally higher editing activity (at the detected sites) in foragers ($P = 1.3 \times 10^{-14}$, proportion test). The surcharge of forager higher sites is maintained for various coverage and editing difference cutoffs (Figure S12). Furthermore, *Adar* expression (Figure 4B) and the editing index (Figure 4C) are both higher in foragers, and the editing index correlates with *Adar* expression. However, reliable detection of specific differentially edited sites with the available sample size (6 foragers

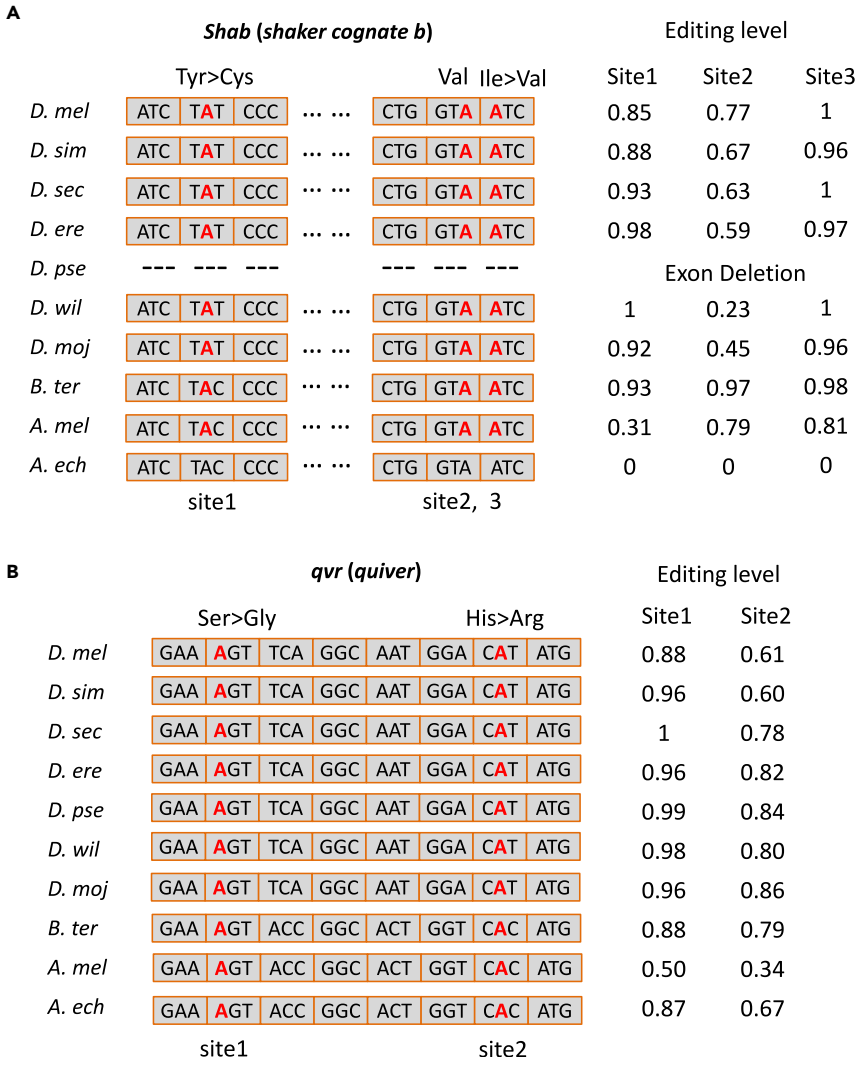


Figure 3. Editing sites conserved between bees and flies

(A) One synonymous and two nonsynonymous editing sites in *Shab* transcripts are edited in heads of two bee and four *Drosophila* species.

(B) Two nonsynonymous editing sites in *qvr* transcripts are edited in heads of two bee and four *Drosophila* species. The editing sites are colored red. Editing levels measured in male heads of flies (unpublished), pooled heads of honeybee drones, and brains of bumblebees (Porath et al., 2019) are shown. *D. mel*, *Drosophila melanogaster*; *D. sim*, *Drosophila simulans*; *D. sec*, *Drosophila sechellia*; *D. ere*, *Drosophila erecta*; *D. pse*, *Drosophila pseudoobscura*; *D. wil*, *Drosophila willistoni*; *D. moj*, *Drosophila mojavensis*; *A. mel*, *Apis mellifera*; *B. ter*, *Bombus terrestris*.

and 6 nurses) is challenging, and not even a single site was identified with FDR = 0.05 (Wilcoxon rank-sum test). We have thus relaxed the statistical test, allowing for FDR = 0.3, and found 61 candidate sites, most of which are differentially edited (Figure 4D and Table S8). These include 10 nonsynonymous, 16 3′UTR, and 35 intergenic sites. Consistently, 55 of these sites exhibit higher levels of editing in foragers. Finally, PCA analysis of the editing profile across sites results in distinct clusters of drones, nurses, and foragers, a classification that is not achieved by the expression profile (Figures 4E and 4F). Taken together, these results show a global increase in editing in foragers compared to nurses.

Convergent adaptation of A-to-I editing?

One of the recoding sites conserved between the two bee species resides within *Adar* transcript (Table S5 and Figure 5A). In bumblebee, the recoding level of the conserved site, I482M (ATA to ATG), positively

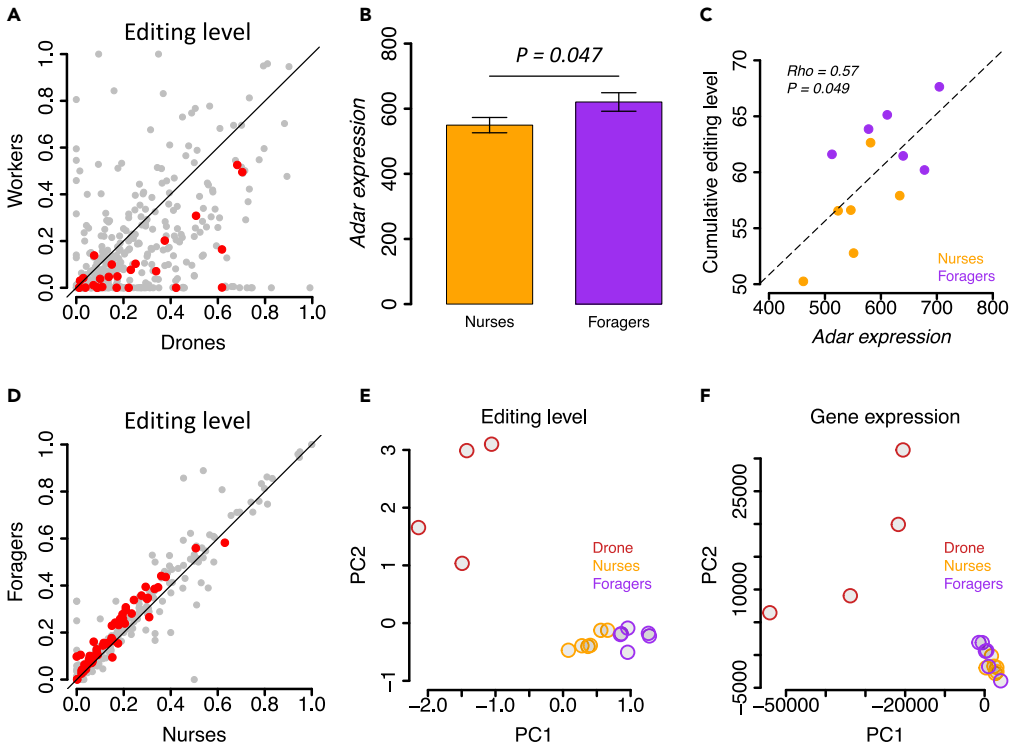


Figure 4. Sex-dependent and caste-dependent editing

(A) Comparison of editing levels between drones (males) and workers (females). Statistically significant sites (FDR <0.05) are colored red.

(B) *Adar* expression level (reads count normalized by DESeq2) in nurses and foragers. Data are presented as mean ± SEM (standard error of mean). p value was calculated using Wilcoxon rank-sum test.

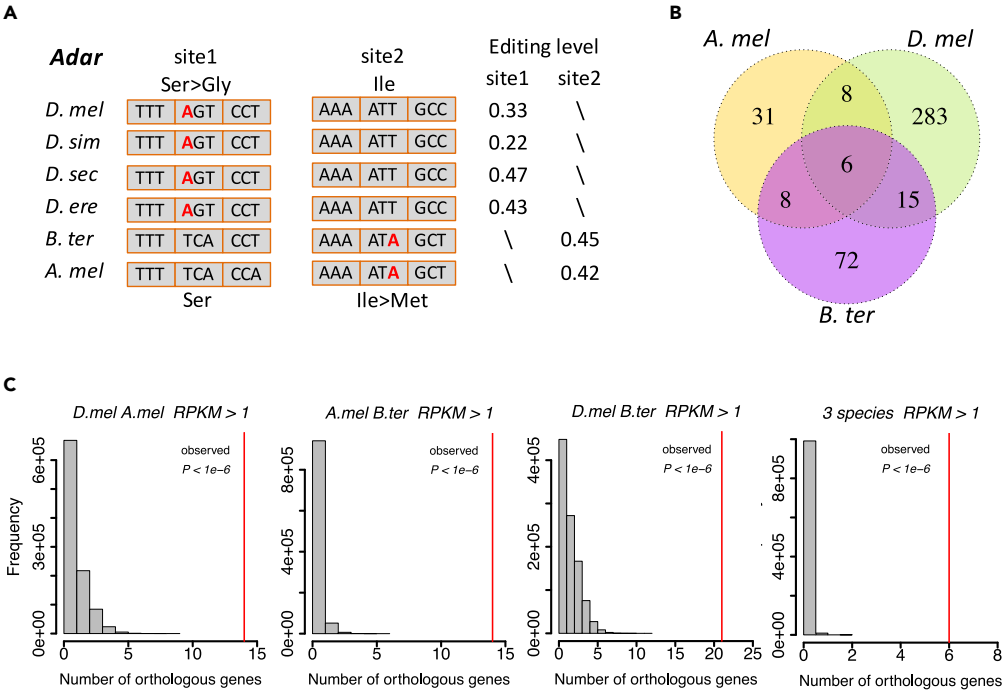
(C) Spearman's correlation between *Adar* expression and editing index (sum over all G alleles in all sites divided by sum of all the coverages in all sites) for all the nurse (orange) and forager (purple) samples.

(D) Comparison of editing levels between two sub-castes of workers: nurses and foragers. Statistically significant sites (FDR <0.3; see Results) are colored red.

(E and F) (E) PCA analysis of the editing profile shows a different behavior for nurses and foragers, while (F) PCA of the expression profile (genes with RPKM>1) does not lead to a clear separation.

correlates with the global editing activity (Porath et al., 2019), suggesting a possible auto-regulation mechanism. Interestingly, *Drosophila Adar* is also auto-edited at a different position, where a Ser (AGT) to Gly (GGT) substitution leads to a less active ADAR protein, resulting in a negative feedback loop of editing activity (Palladino et al., 2000; Savva et al., 2012a). Intriguingly, fly-edited Ser amino acid is conserved in honeybee and bumblebee, but a different codon (TCA, uneditable at the first codon position) is used. Similarly, the bee-edited Ile is conserved in flies, but the edited adenosines (at the third position of ATA) is synonymously mutated to T, and the editable ATA bee codon is substituted by ATT codon in flies, which could not be edited at the third position (Figure 5A). Thus, while auto-editing S430G in flies is abolished in bees due to an uneditable Ser codon, and the auto-editing I482M in bees is abolished in flies due to an uneditable Ile codon. However, the mechanism of auto-editing, possibly used for global ADAR regulation, is shared by the two lineages, albeit at different positions and possibly with different effects on the protein. This might hint at the possibility of convergent evolution of ADAR auto-regulation strategy.

Following this example, we wondered whether there are more genes for which editing is observed in both honeybees and *Drosophila*, even if the exact location of the editing site is not conserved. Excluding conserved sites, there are 53 genes exhibiting CDS editing in honeybee heads, 101 genes exhibiting CDS editing in bumblebee heads, and 312 genes edited in CDS in male brains of *D. melanogaster*. Of these, 21 genes exhibit editing in both bee species, 14 are edited in honeybee and *Drosophila*, 14 are edited in bumblebee and *Drosophila*, and six genes are found to be edited in all three species (Figure 5B and Table S9). These numbers are significantly higher than expected by random sampling, ($p < 1 \times 10^{-6}$,



D

	N	S	Observed N/S	Expected N/S	P value
Sites in genes shared by at least two species					
D. melanogaster	127	22	5.77	3.80	0.038
A. mellifera	37	3	12.33	2.26	4.5E-4
B. terrestris	76	21	3.62	2.29	0.036
Sites in species-specific edited genes					
D. melanogaster	539	117	4.61	3.80	0.031
A. mellifera	74	6	12.33	2.26	5.2E-7
B. terrestris	87	35	2.49	2.29	0.38

Figure 5. Convergent evolution of editing

(A) *Adar* transcripts are auto-edited in *Drosophila* and bees. The Ser > Gly site is highly conserved across *Drosophila* species. The *Drosophila*-editable Ser codon AGT is changed to an uneditable Ser codon TCA in bees. On the other hand, the Ile > Met auto-editing site is conserved between honeybee and bumblebee. The bee-editable Ile codon ATA (edited at the third position) appears as an uneditable Ile codon ATT in flies. Editing sites are colored red. The editing levels from male heads of flies (unpublished), pooled heads of honeybee drones, and brains of bumblebee (Porath et al., 2019) are shown. *D. mel*, *Drosophila melanogaster*; *D. sim*, *Drosophila simulans*; *D. sec*, *Drosophila sechellia*; *D. ere*, *Drosophila erecta*; *A. mel*, *Apis mellifera*; *B. ter*, *Bombus terrestris*.

(B) Venn diagram demonstrating the overlap between orthologous genes with editing sites in coding regions.

(C) The observed number of orthologous genes with editing in coding regions, compared to the expected distribution. Genes with RPKM >1 in both species were chosen to calculate the expected numbers (results are essentially the same for other expression cutoffs, Figure S13). p values calculated by randomization test.

(D) The nonsynonymous to synonymous (*N*/*S*) ratios are significantly higher than expected, for both shared and species-specific sites (except for species-specific sites in *B. terrestris*). Here, the shared and species-specific sites refer to the sites in shared genes and species-specific genes as defined in the above Venn diagram in (B). The expected ratio is evaluated by calculating the effect of putative editing on all adenosines in the coding region. p values are obtained from Fisher's exact test.

randomization test), even if one accounts for the expression profile (Transparent methods, Figures 5C and S13) (Table S9). We have repeated the analysis using stricter criteria to characterize sites, considering as edited only sites with an observed editing level >10% (all samples pooled), and a binomial P < 0.05

(excluding the null hypothesis of no editing), and considering a site to be not edited if the binomial $P < 0.05$ for the null hypothesis that the site is edited at 10%. Reassuringly, the results are robust to this change of definitions (Figures S14 and S15). These results suggest an interesting convergent evolution of A-to-I editing. Recoding sites are rarely conserved across clades, but the same genes in different lineages tend to acquire a recoding event, apparently independently, indicating a functional importance for recoding of this gene.

Discussion

Despite over fifteen years of developments, reliable detection of CDS editing remains a challenge. Systematic searches often lead to high false-positive rates (reflected by the fraction of non-AG mismatches observed), especially for mammalian transcriptomes where the scope of recoding is rather low. Important exceptions are *Drosophila* and cephalopods, where the recoding signal is much more pronounced and easier to detect. Applying a novel strict alignment approach, we were able to achieve here a high accuracy in CDS editing detection (84.3% in heads) despite the overall modest scope of recoding sites. Another similarity between honeybees *Drosophila* and cephalopods is related to the important question regarding the extent to which recoding is adaptive. This is often estimated by the *N/S* ratio, where *Drosophila* and cephalopods exhibit a pattern markedly different from the one seen in other species studied (including human). In most species, the *N/S* ratio observed is lower or similar to the one expected under neutrality, and it is significantly higher in *Drosophila* and cephalopods. Again, we find here that honeybees show a strikingly high *N/S* ratio, similar to the above two clades. We suggest that the two last points are interconnected. The apparently low *N/S* ratio observed in many species (human included) may reflect the low accuracy of the lists of CDS editing sites and not the (lack of) adaptive potential for recoding in these species.

The importance of post-transcriptional and post-translational mechanisms in generating the proteomic complexity of higher organisms has been emphasized in recent decades. These epigenetic mechanisms allow for diversification of the proteome and functional heterogeneity across tissues, developmental stages, brain regions, or even among individual cells.

Recoding by A-to-I RNA editing is an epigenetic mechanism capable of diversifying the proteome, creating a range of proteins from a single genomically encoded gene in a temporally regulated, tissue-specific, condition-dependent way and providing the organism with a new means for acclimation and adaptation. Indeed, several studies have demonstrated how recoding levels at specific sites do change as a function of the organism's condition (Garrett and Rosenthal 2012; Robinson et al., 2016; Gallo et al., 2017; Porath 2017; Terajima et al., 2017; Yablonovitch et al., 2017). Importantly, many studies have demonstrated altered editing of individual recoding targets in various disease states (Gallo et al., 2017). However, these interesting examples notwithstanding, the extent to which the recoding phenomenon is actually used as a means for proteome diversification is still under debate. In fact, several recent studies have raised the possibility that nonsynonymous editing may be used to compensate for otherwise deleterious G-to-A mutations, rather than allowing for the two alleles (A and G) to co-exist (Jiang and Zhang 2019; Mai and Chuang 2019; Popitsch et al., 2020). Thus, a high *N/S* ratio is not sufficient to prove that editing serves for adaptation through proteome diversification.

The behavioral, physiological, and morphological plasticity exhibited by social insects suggests an important role for epigenetic mechanisms. Differential RNA editing was previously shown for different castes of the highly social leaf-cutting ant *Acromyrmex echinatior* (Li et al., 2014), as well as for bumblebee workers (Porath et al., 2019). Our results expand these finding to another species, further supporting the notion of RNA editing as a source of proteomic complexity, that may be recruited to provide phenotypic variation among genetically identical individuals.

ADAR enzymes and the recoding phenomenon are widely conserved across metazoan. However, the repertoire of recoding sites varies considerably and seems to have developed almost independently in different clades. Here, we point out that the same gene targets are being edited in distant species, even if the exact locations of the editing sites within the transcript vary. A particularly interesting example is that of nonsynonymous auto-editing of *Adar*, shown to affect ADAR activity in bumblebees, flies, and even mammals (Rueter et al., 1999; Savva et al., 2012a; Porath et al., 2019), possibly serving as global editing autoregulation. It thus seems possible that the selective advantage in having multiple versions of the protein product is shared by the target genes, while there is more than one specific way to achieve this diversity

through recoding. Clades developed along different evolutionary routes have converged on different implementations (in terms of the specific editing site) of the same diversifying solution. These ideas should be further tested in the future, using larger data sets and across additional species and clades. Moreover, as biochemical and functional understanding of the impact of the different edits are gained, it will be possible to look into a possible common effect for the different edits of the common targets.

Limitations of the study

The editomes of the bee species analyzed here are based on a limited number of samples. They probably represent a subset of the actual repertoire of editing sites. The results should be re-tested when more data are available. In particular, the honeybee editome was built using drone samples. Thus, some individual differential editing between the nurses and foragers sub-castes may have missed worker-specific sites. However, the global increased editing in foragers, supported by a higher Adar expression, is probably robust.

Resource availability

Lead contact

Further information and requests for resources should be directly and will be fulfilled by the Lead Contact, Jian Lu (luj@pku.edu.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

All deep-sequencing data generated in this study were deposited in the China National Genomics Data Center Genome Sequence Archive (GSA) under accession number CRA002262. Other deep-sequencing data analyzed were downloaded from SRA as follows. Brains of honeybee workers: accession numbers SRR445999 to SRR446004 (reverted nurse) and SRR446005 to SRR446010 (forager) (Herb et al., 2012). Bumblebee: SRP166322 (Porath et al., 2019). *D. mel* head: SRP067542 (Zhang et al., 2018). *D. sim* head: SRP074828 (Duan et al., 2017). *D. pse*: DRR055250 and DRR055251 (Nozawa et al., 2016). *D. wil*: SRR341127 and SRR341129 (Meisel et al., 2012). *D. moj*: SRR037508-SRR037518 (generated by the *Drosophila* modENCODE project). *D. sec*, and *D. ere* data are unpublished data generated in our own lab. The reference genomes versions used for *Drosophila* species are Dmel_r6.04 (http://flybase.org/) for *D. mel*, and versions droSim1, droSec1, droEre1, dp4, droWil1, and droMoj3 downloaded from UCSC Genome Browser (http://genome.ucsc.edu/) for the other fly species. For the *Bombus terrestris*, we used version Bter_1.0.

Methods

All methods can be found in the accompanying Transparent methods supplemental file.

Supplemental information

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101983.

Acknowledgments

We thank the National Center for Protein Sciences at Peking University for technical assistance. Part of the analysis was performed on the High Performance Computing Platform of the Center for Life Science at Peking University. This work was supported by the joint NSFC-ISF program (NSFC grant number 3201101147 to J.L. and ISF grant number 3371/20 to E.E.), as well as grants from the Ministry of Science and Technology of the People's Republic of China (2016YFA0500800) and National Natural Science Foundation of China (91731301) to J.L. and support from the Israel Science Foundation (1945/18) to E.E.

Author contributions

J.L. designed the research; J.X.H. contributed the honeybee materials; S.Q.D. performed the research; Y.G.D., H.T.P., E.E., and J.L. performed the bioinformatics analyses; E.E. and J.L. wrote the paper.

Declaration of interests

The authors declare no competing interests.

Received: September 4, 2020
Revised: November 18, 2020
Accepted: December 17, 2020
Published: January 22, 2021

References

Alon, S., Garrett, S.C., Levanon, E.Y., Olson, S., Graveley, B.R., Rosenthal, J.J., and Eisenberg, E. (2015). The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. Elife *4*, e05198.

Alon, S., Mor, E., Vigneault, F., Church, G.M., Locatelli, F., Galeano, F., Gallo, A., Shomron, N., and Eisenberg, E. (2012). Systematic identification of edited microRNAs in the human brain. Genome Res. *22*, 1533–1540.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. *2*, e391.

Basilio, C., Wahba, A.J., Lengyel, P., Speyer, J.F., and Ochoa, S. (1962). Synthetic polynucleotides and the amino acid code. Proc. Natl. Acad. Sci. U S A *48*, 613–616.

Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. *71*, 817–846.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res. *14*, 2379–2387.

Borchert, G.M., Gilmore, B.L., Spengler, R.M., Xing, Y., Lanier, W., Bhattacharya, D., and Davidson, B.L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. Hum. Mol. Genet. *18*, 4801–4807.

Buchumenski, I., Bartok, O., Ashwal-Fluss, R., Pandey, V., Porath, H.T., Levanon, E.Y., and Kadener, S. (2017). Dynamic hyper-editing underlies temperature adaptation in Drosophila. PLoS Genet. *13*, e1006931.

Chen, J.-Y., Peng, Z., Zhang, R., Yang, X.-Z., Tan, B.C.-M., Fang, H., Liu, C.-J., Shi, M., Ye, Z.-Q., Zhang, Y.E., et al. (2014). RNA editome in rhesus macaque shaped by purifying selection. PLoS Genet. *10*, e1004274.

Danecek, P., Nellaker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M., and Adams, D.J. (2012). High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol. *13*, 26.

Duan, Y., Dou, S., Luo, S., Zhang, H., and Lu, J. (2017). Adaptation of A-to-I RNA editing in Drosophila. PLoS Genet. *13*, e1006648.

Eisenberg, E., and Levanon, E.Y. (2018). A-to-I RNA editing — immune protector and transcriptome diversifier. Nat. Rev. Genet. *19*, 473–490.

Flomen, R., Knight, J., Sham, P., Kerwin, R., and Makoff, A. (2004). Evidence that RNA editing modulates splice site selection in the 5-HT2C receptor gene. Nucleic Acids Res. *32*, 2113–2122.

Gallo, A., Vukic, D., Michalik, D., O'Connell, M.A., and Keegan, L.P. (2017). ADAR RNA editing in human disease; more to it than meets the I. Hum. Genet. *136*, 1265–1278.

Garrett, S., and Rosenthal, J.J. (2012). RNA editing underlies temperature adaptation in K+ channels from polar octopuses. Science *335*, 848–851.

Goldstein, B., Agranat-Tamir, L., Light, D., Ben-Naim Zgayer, O., Fishman, A., and Lamm, A.T. (2017). A-to-I RNA editing promotes developmental stage–specific gene and lncRNA expression. Genome Res. *27*, 462–470.

Gommans, W.M., Mullen, S.P., and Maas, S. (2009). RNA editing: a driving force for adaptive evolution? BioEssays *31*, 1137–1145.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.

Herb, B.R., Wolschin, F., Hansen, K.D., Aryee, M.J., Langmead, B., Irizarry, R., Amdam, G.V., and Feinberg, A.P. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. Nat. Neurosci. *15*, 1371–1373.

Jiang, D., and Zhang, J. (2019). The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. Nat. Commun. *10*, 5411.

Jin, Y., Tian, N., Cao, J., Liang, J., Yang, Z., and Lv, J. (2007). RNA editing and alternative splicing of the insect nAChR subunit alpha6 transcript: evolutionary conservation, divergence and regulation. BMC Evol. Biol. *7*, 1–12.

Keegan, L.P., Gallo, A., and O'Connell, M.A. (2001). The many roles of an RNA editor. Nat. Rev. Genet. *2*, 869–878.

Keegan, L.P., McGurk, L., Palavicini, J.P., Brindle, J., Paro, S., Li, X., Rosenthal, J.J., and O'Connell, M.A. (2011). Functional conservation in human and Drosophila of Metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects. Nucleic Acids Res. *39*, 7249–7262.

Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. *14*, 1719–1725.

Klironomos, F.D., Berg, J., and Collins, S. (2013). How epigenetic mutations can affect genetic evolution: model and mechanism. BioEssays *35*, 571–578.

Lev-Maor, G., Sorek, R., Levanon, E.Y., Paz, N., Eisenberg, E., and Ast, G. (2007). RNA-editing-mediated exon evolution. Genome Biol. *8*, R29.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotech. *22*, 1001–1005.

Li, Q., Wang, Z., Lian, J., Schiott, M., Jin, L., Zhang, P., Zhang, Y., Nygaard, S., Peng, Z., Zhou, Y., et al. (2014). Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior. Nat. Commun. *5*, 4943.

Liang, H., and Landweber, L.F. (2007). Hypothesis: RNA editing of microRNA target sites in humans? RNA *13*, 463–467.

Licht, K., Hartl, M., Amman, F., Anrather, D., Janisiw, M.P., and Jantsch, M.F. (2019). Inosine induces context-dependent recoding and translational stalling. Nucleic Acids Res. *47*, 3–14.

Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J.C., and Eisenberg, E. (2017). Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell *169*, 191–202.e11.

Mai, T.L., and Chuang, T.J. (2019). A-to-I RNA editing contributes to the persistence of predicted damaging mutations in populations. Genome Res. *29*, 1766–1776.

Mazloomian, A., and Meyer, I.M. (2015). Genome-wide identification and characterization of tissue-specific RNA editing events in D. melanogaster and their potential role in regulating alternative splicing. RNA Biol. *12*, 1391–1401.

Meisel, R.P., Malone, J.H., and Clark, A.G. (2012). Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res. *22*, 1255–1265.

Morse, D.P., and Bass, B.L. (1999). Long RNA hairpins that contain inosine are present in Caenorhabditis elegans poly(A)+ RNA. Proc. Natl. Acad. Sci. U S A *96*, 6048–6053.

Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA *12*, 1802–1809.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. *79*, 321–349.

Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. Nat. Rev. Mol. Cell Biol. *17*, 83–96.

Nozawa, M., Onizuka, K., Fujimi, M., Ikeo, K., and Gojobori, T. (2016). Accelerated pseudogenization on the neo-X chromosome in Drosophila miranda. Nat. Commun. *7*, 13659.

Page, R.E., Jr., Rueppell, O., and Amdam, G.V. (2012). Genetics of reproduction and regulation of honeybee (Apis mellifera L.) social behavior. Annu. Rev. Genet. *46*, 97–119.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000). dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. RNA *6*, 1004–1018.

Picardi, E., D'Erchia, A.M., Lo Giudice, C., and Pesole, G. (2017). REDIportal: a comprehensive database of A-to-I RNA editing events in humans. Nucleic Acids Res. *45*, D750–D757.

Popitsch, N., Huber, C.D., Buchumenski, I., Eisenberg, E., Jantsch, M., von Haeseler, A., and Gallach, M. (2020). A-to-I RNA editing uncovers hidden signals of adaptive genome evolution in animals. Genome Biol. Evol. *12*, 345–357.

Porath, H.T. (2017). A-To-I RNA editing in the earliest-diverging eumetazoan phyla. Mol. Biol. Evol. *34*, 1890–1901.

Porath, H.T., Hazan, E., Shpigler, H., Cohen, M., Band, M., Ben-Shahar, Y., Levanon, E.Y., Eisenberg, E., and Bloch, G. (2019). RNA editing is abundant and correlates with task performance in a social bumblebee. Nat. Commun. *10*, 1605.

Porath, H.T., Knisbacher, B.A., Eisenberg, E., and Levanon, E.Y. (2017a). Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. Genome Biol. *18*, 185.

Porath, H.T., Schaffer, A.A., Kaniewska, P., Alon, S., Eisenberg, E., Rosenthal, J., Levanon, E.Y., and Levy, O. (2017b). A-to-I RNA editing in the earliest-diverging eumetazoan phyla. Mol. Biol. Evol. *34*, 1890–1901.

Ramaswami, G., and Li, J.B. (2016). Identification of human RNA editing sites: a historical perspective. Methods *107*, 42–47.

Robinson, J.E., Paluch, J., Dickman, D.K., and Joiner, W.J. (2016). ADAR-mediated RNA editing suppresses sleep by acting as a brake on glutamatergic synaptic plasticity. Nat. Commun. *7*, 10512.

Rodriguez, J., Menet, J.S., and Rosbash, M. (2012). Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Mol. Cell *47*, 27–37.

Rosenthal, J.J.C. (2015). The emerging role of RNA editing in plasticity. J. Exp. Biol. *218*, 1812.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75–80.

Savva, Y.A., Jepson, J.E., Sahin, A., Sugden, A.U., Dorsky, J.S., Alpert, L., Lawrence, C., and Reenan, R.A. (2012a). Auto-regulatory RNA editing fine-tunes mRNA re-coding and complex behaviour in Drosophila. Nat. Commun. *3*, 790.

Savva, Y.A., Rieder, L.E., and Reenan, R.A. (2012b). The ADAR protein family. Genome Biol. *13*, 252.

St Laurent, G., Tackett, M.R., Nechkin, S., Shtokalo, D., Antonets, D., Savva, Y.A., Maloney, R., Kapranov, P., Lawrence, C.E., and Reenan, R.A. (2013). Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. Nat. Struct. Mol. Biol. *20*, 1333–1339.

Terajima, H., Yoshitane, H., Ozaki, H., Suzuki, Y., Shimba, S., Kuroda, S., Iwasaki, W., and Fukada, Y. (2017). ADARB1 catalyzes circadian A-to-I editing and regulates RNA rhythm. Nat. Genet. *49*, 146–151.

Xu, G., and Zhang, J. (2014). Human coding RNA editing is generally nonadaptive. Proc. Natl. Acad. Sci. U S A *111*, 3769–3774.

Yablonovitch, A.L., Fu, J., Li, K., Mahato, S., Kang, L., Rashkovetsky, E., Korol, A.B., Tang, H., Michalak, P., Zelhof, A.C., et al. (2017). Regulation of gene expression and RNA editing in Drosophila adapting to divergent microclimates. Nat. Commun. *8*, 1570.

Yan, H., Simola, D.F., Bonasio, R., Liebig, J., Berger, S.L., and Reinberg, D. (2014). Eusocial insects as emerging models for behavioural epigenetics. Nat. Rev. Genet. *15*, 677–688.

Yang, X.-Z., Chen, J.-Y., Liu, C.-J., Peng, J., Wee, Y.R., Han, X., Wang, C., Zhong, X., Shen, Q.S., Liu, H., et al. (2015). Selectively constrained RNA editing regulation crosstalks with piRNA biogenesis in primates. Mol. Biol. Evol. *32*, 3143–3157.

Yu, Y., Zhou, H., Kong, Y., Pan, B., Chen, L., Wang, H., Hao, P., and Li, X. (2016). The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. PLoS Genet. *12*, e1006191.

Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS Biol. *16*, e2003903.

Zhang, R., Deng, P., Jacobson, D., and Li, J.B. (2017). Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. PLoS Genet. *13*, e1006563.

Zhao, H.Q., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A.Y., Liu, X.M., Ye, A.Y., Dong, M.Q., and Wei, L. (2015). Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. Genome Res. *25*, 66–75.

iScience, Volume *24*

Supplemental Information

A-to-I RNA editing in honeybees shows signals of adaptation and convergent evolution

Yuange Duan, Shengqian Dou, Hagit T. Porath, Jiaxing Huang, Eli Eisenberg, and Jian Lu

Supplemental Figures

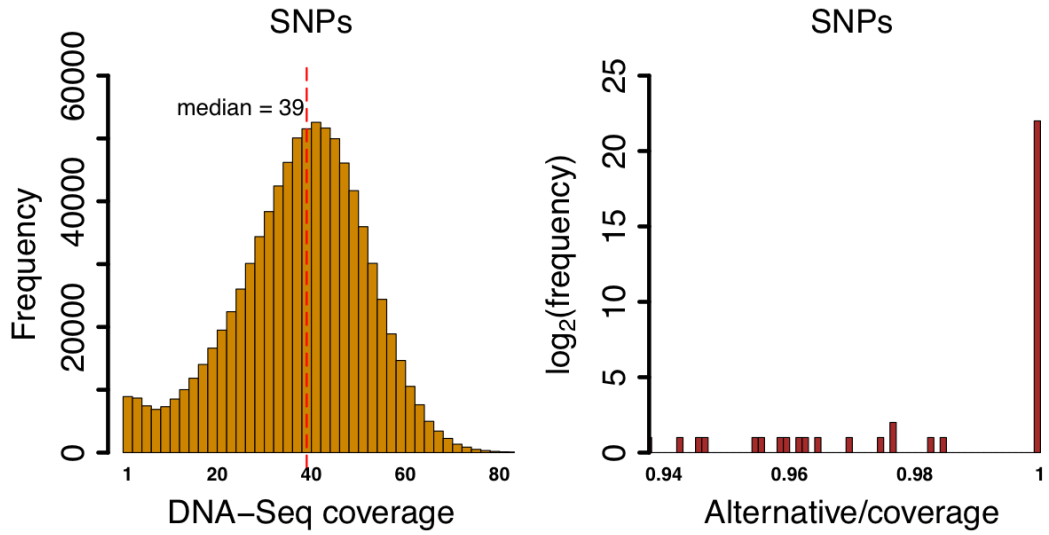


Figure S1. Per-individual DNA-Seq coverage at detected SNPs sites (left) and the variation level (alternative reads count / coverage) of these SNPs (right), Related to Figure1.

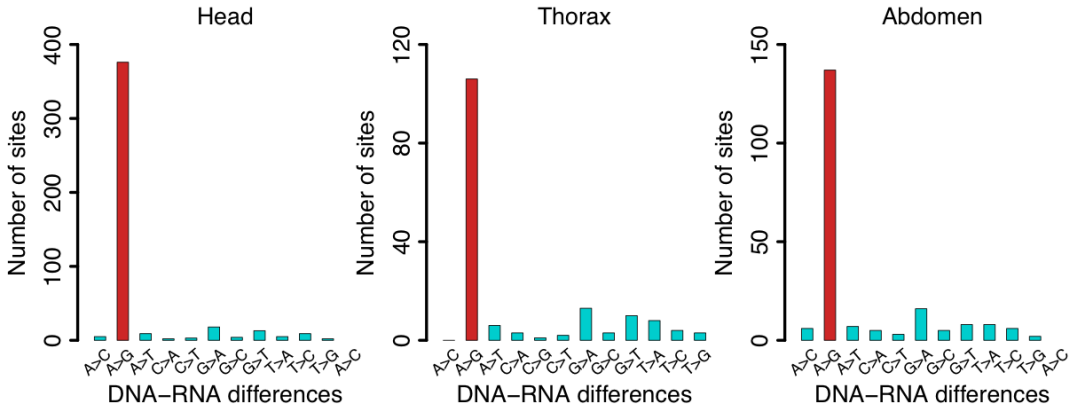


Figure S2. Distribution of DNA-RNA mismatch types for sites detected in heads, thoraxes, or abdomens of honeybees, Related to Figure 1.

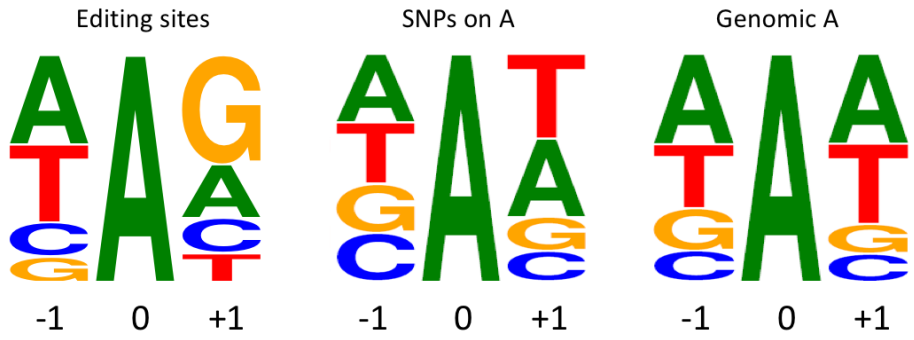


Figure S3. Local sequence motif around editing sites (left), SNPs (middle) or unedited adenosines in genome (right), Related to Figure 1.

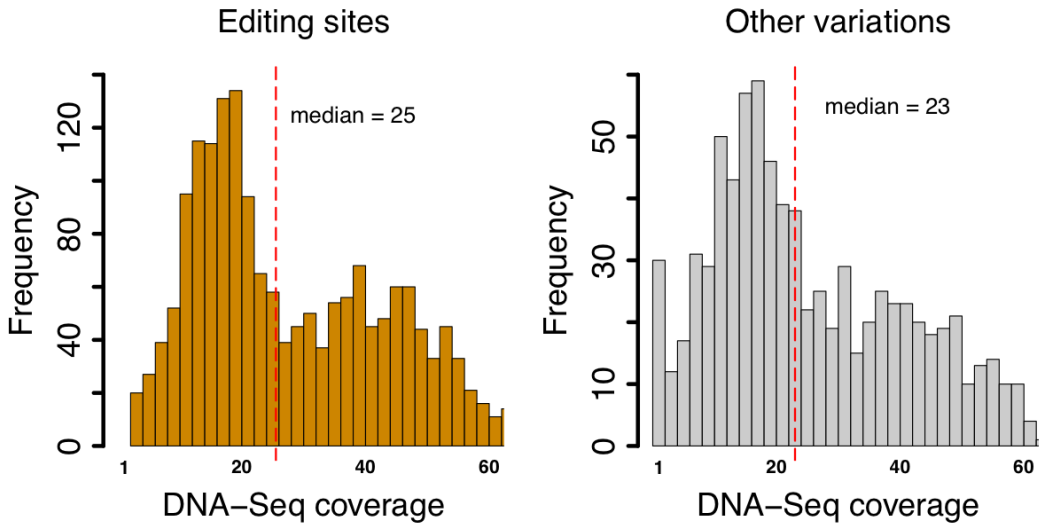


Figure S4. Histogram of per-individual DNA-Seq coverage at candidate RNA editing sites and other variations in each sample, Related to Figure 1.

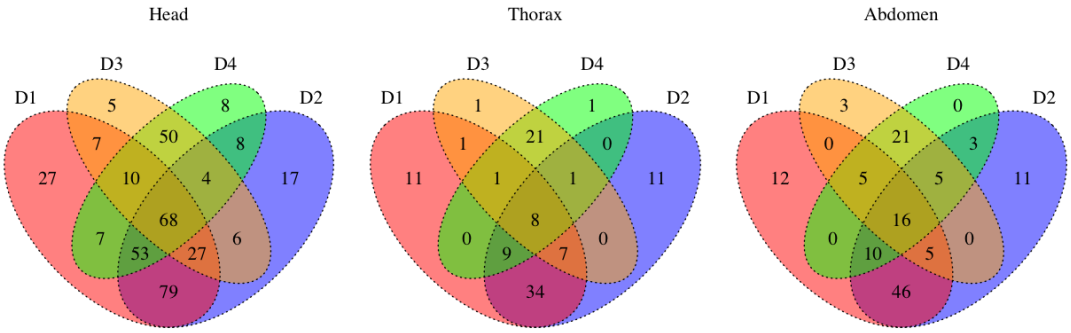


Figure S5. Venn diagram demonstrating the overlaps of editing sites in four honeybee drones, Related to Figure 1.

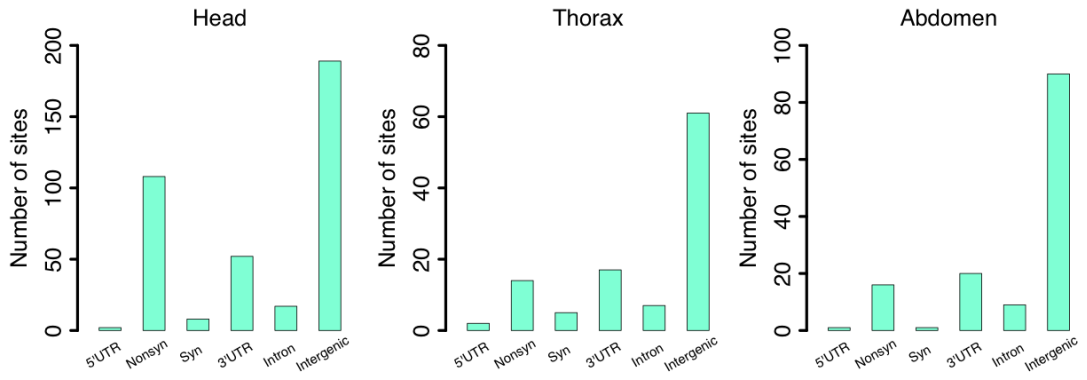


Figure S6. Functional annotation of editing sites detected in heads, thoraxes, or abdomens of honeybees, Related to Figure 1.

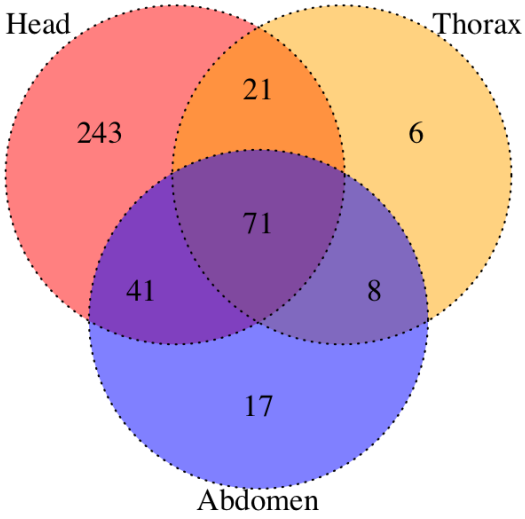


Figure S7. Venn diagram demonstrating the overlaps of editing sites in three honeybee tissues, Related to Figure 1.

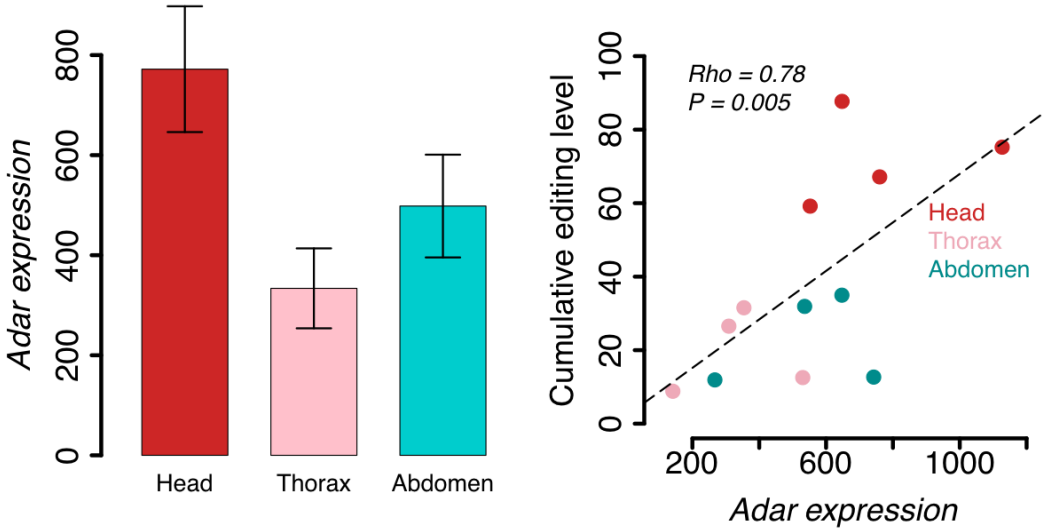


Figure S8. *Adar* expression is highest in heads (left) and is positively correlated (Spearman's correlation) to cumulative editing levels in each sample (right), Related to Figure 1. The cumulative editing level is the sum of the editing level over all sites. Error bars represent the standard error of mean. *Adar* expression values were calculated by the DESeq2 software.

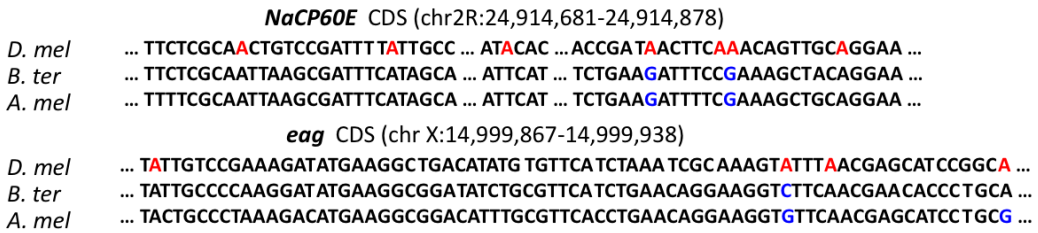


Figure S9. Non-conserved editing sites in gene *NaCP60E* and *eag*, Related to Figure 3.

Drosophila-specific editing sites are colored in red**.** Many of these sites are genomically encoded as G in the honeybee genome.

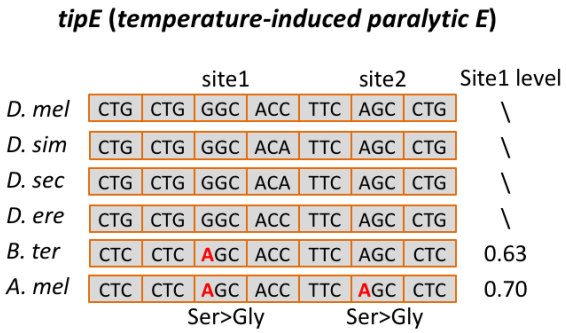


Figure S10. Bee-specific editing sites in *tipE*, Related to Figure 3. Two *tipE* recoding sites are observed in honeybee (orthologous gene *GB47375*), one of which is also seen in bumblebee (orthologous gene *XM_003393369*) (Porath et al. 2019), colored in red. Editing is not seen at this site in *Drosophila*, as the genome encodes edited form of the protein.

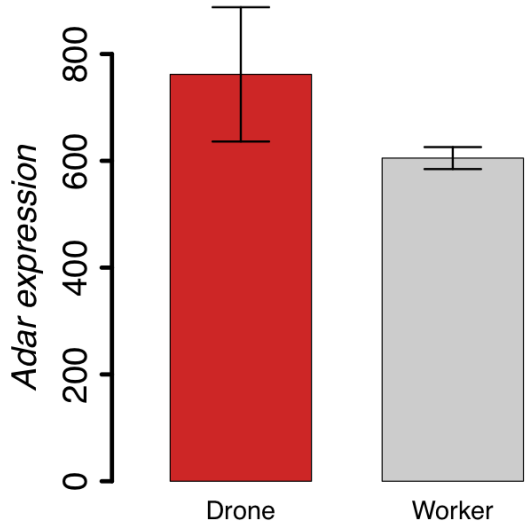


Figure S11. *Adar* expression in drones and workers, Related to Figure 4. Reads count was normalized by DESeq2. Error bars represent standard error of mean. T-test was used to calculate the statistical significance. No significant difference was obtained ($P = 0.23$).

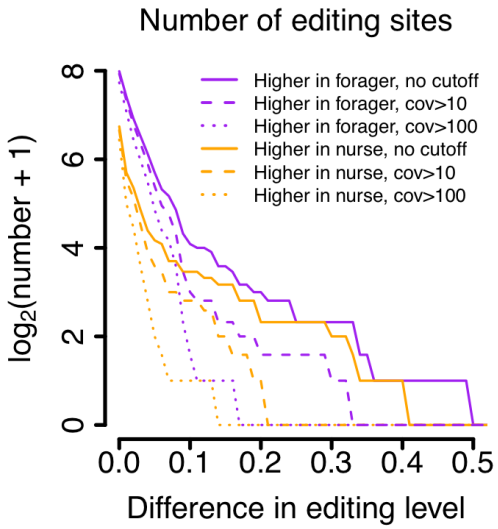


Figure S12. Comparison of editing levels in two sub-castes of workers, Related to Figure 4. The numbers of editing sites with levels higher in foragers (purple) and nurses (orange) are shown. Different cutoffs of sequencing coverages were used.

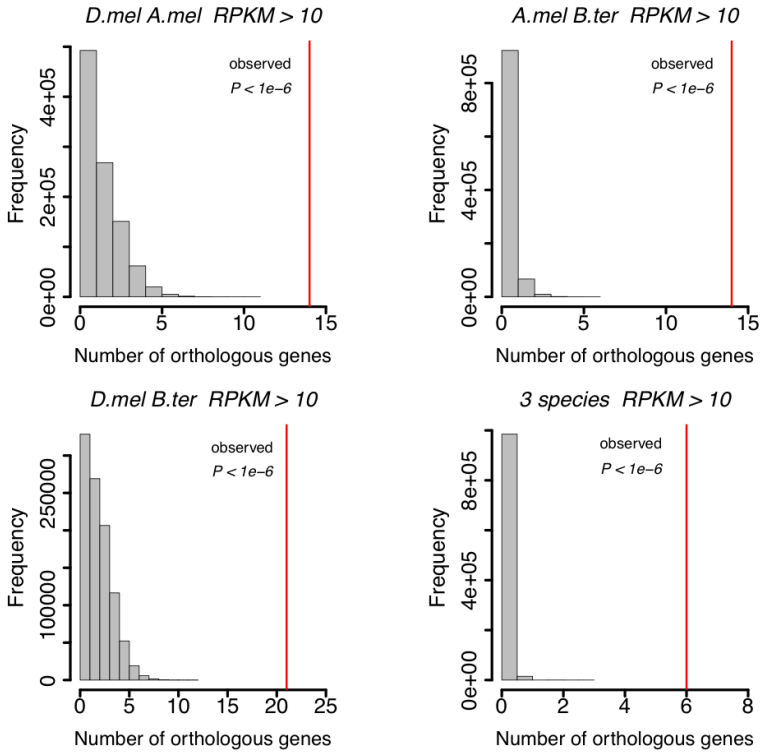


Figure S13. The observed and expected numbers of orthologous genes with editing in coding regions, Related to Figure 5. Genes with RPKM > 10 were chosen to calculate the expected numbers.

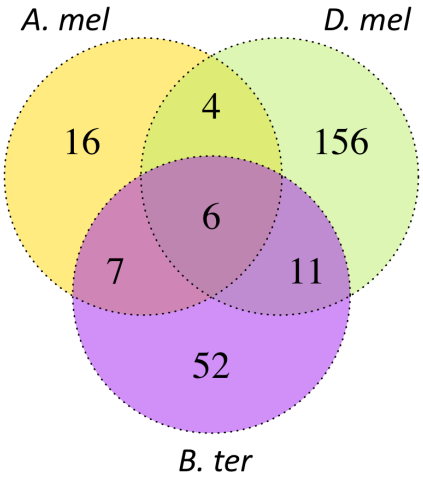


Figure S14. Venn diagram demonstrating the overlap between orthologous genes with editing sites in coding regions, Related to Figure 5. Strict criteria for determining sites as edited or not-edited were applied. The details are described in the main text.

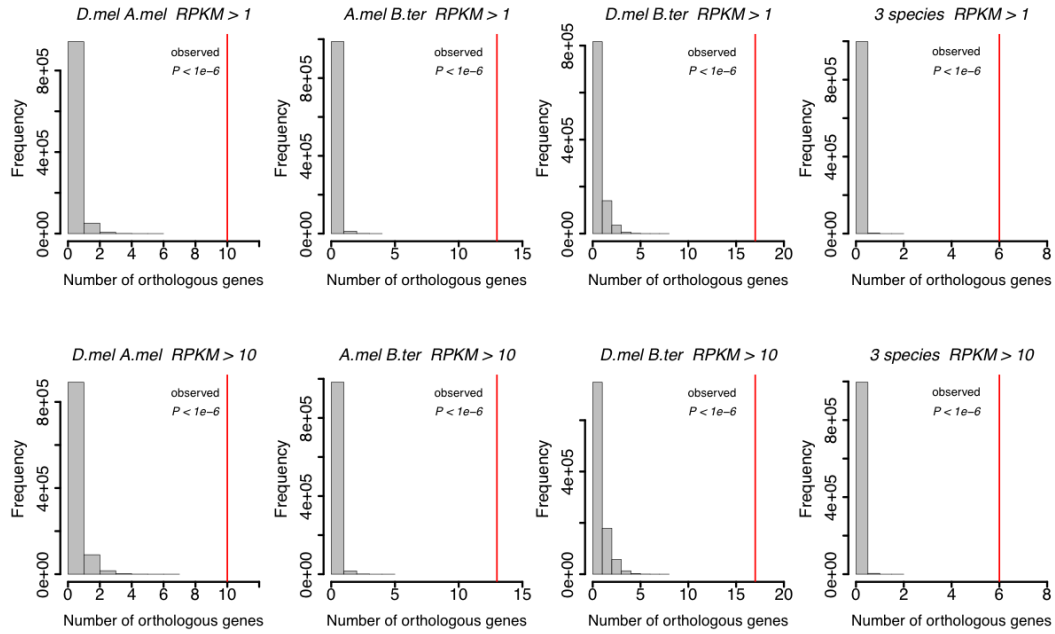


Figure S15. The observed and expected numbers of orthologous genes with editing in coding regions, Related to Figure 5. Strict criteria for determining sites as edited or not-edited were applied. The details are described in the main text. Genes with RPKM > 1 or RPKM > 10 were chosen to calculate the expected numbers.

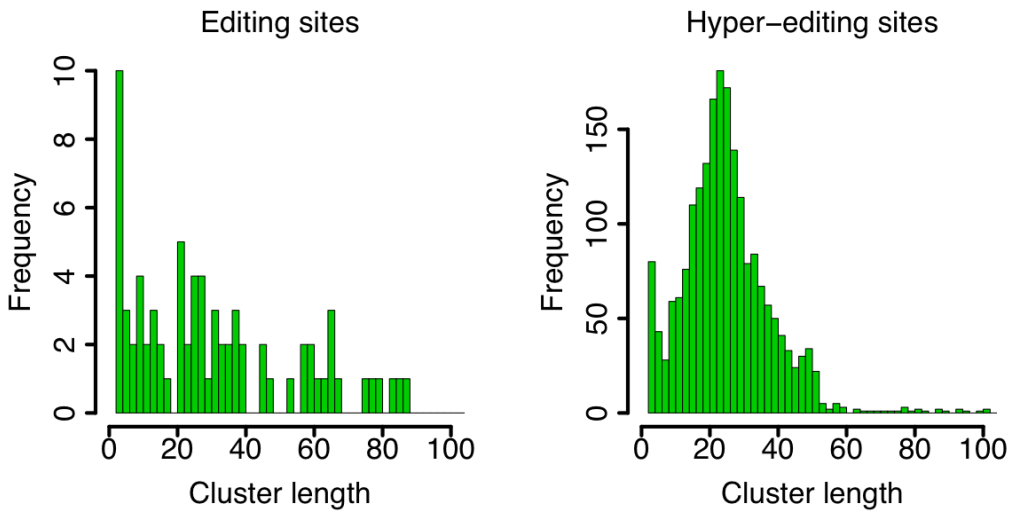


Figure S16. Histograms showing the cluster length of normal editing sites (left) and hyper-editing sites (right), Related to Figure 1. Editing sites within 100bps were clustered.

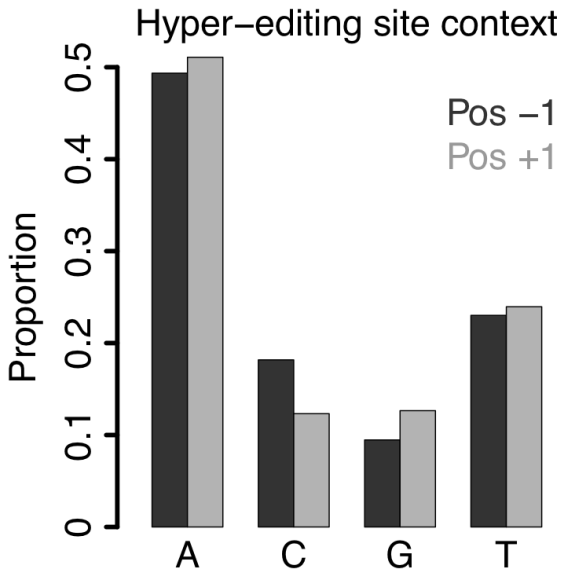


Figure S17. The nucleotide context of hyper-editing sites, Related to Figure 1. Pos -1, the upstream nucleotide. Pos +1, the downstream nucleotide.

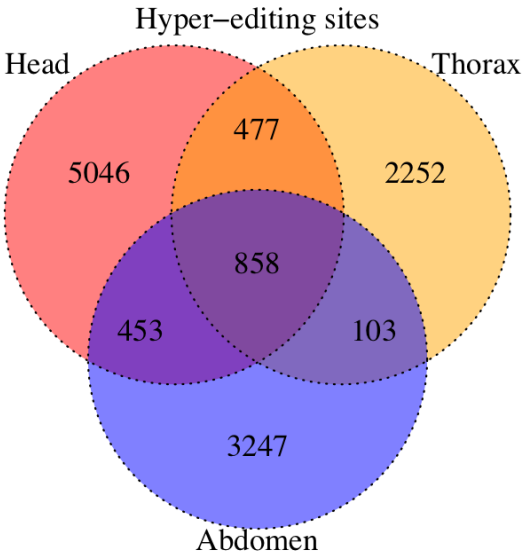


Figure S18. Venn diagram demonstrating the overlaps of hyper-editing sites in three tissues of honeybee drones, Related to Figure 1.

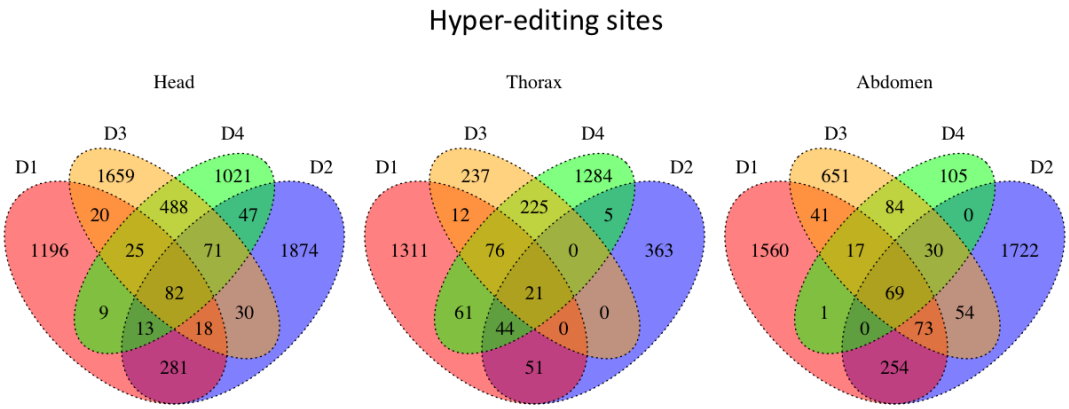


Figure S19. Venn diagram demonstrating the overlaps of hyper-editing sites in four honeybee drones, Related to Figure 1.

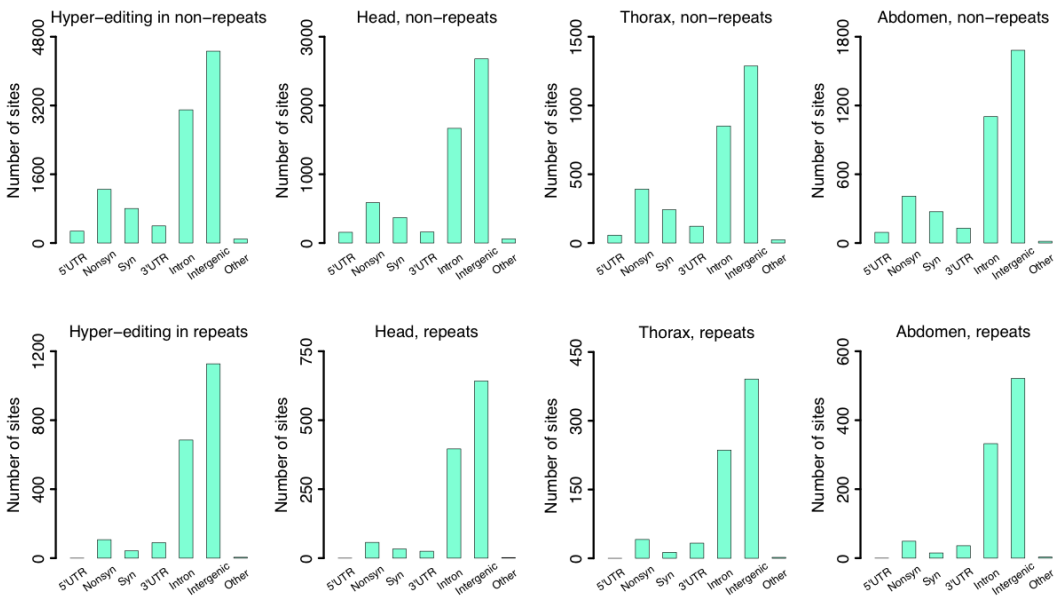


Figure S20. Annotation of hyper-editing sites for repetitive and non-repetitive regions, separately, Related to Figure 1. The hyper-editing sites in repeats have lower fractions of nonsynonymous and synonymous sites but have higher fractions of intronic and intergenic sites.

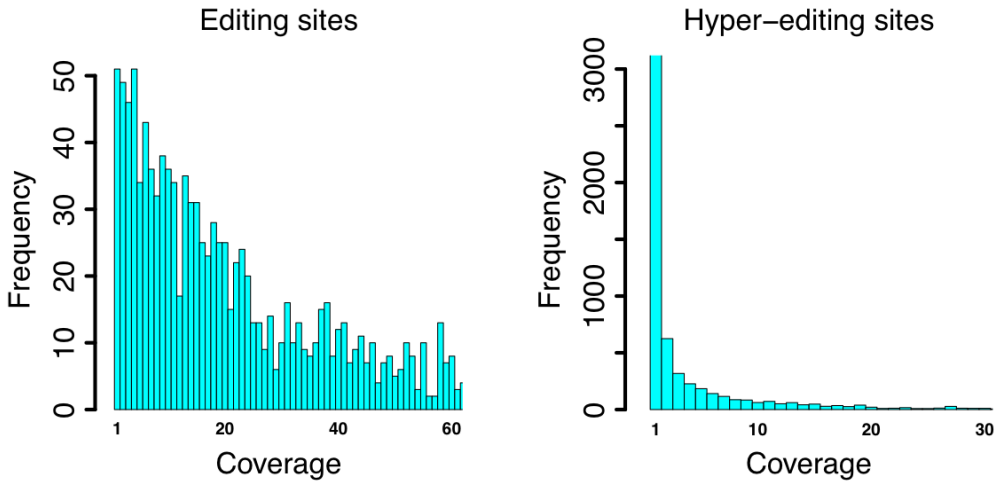


Figure S21. Histograms of sequencing coverage of normal editing sites (left) and hyper-editing sites (right), Related to Figure 1. Editing sites in all samples were used.

Supplemental Tables

Table S1. Number of SNPs identified in each individual honeybee, Related to Figure 1.

	D1	D2	D3	D4	All
5'UTR	2,573	2,495	2,480	2,445	3,769
Nonsyn	8,344	8,032	8,090	8,296	12,494
Syn	20,771	19,509	19,904	19,814	30,559
3'UTR	10,199	9,518	9,640	9,688	14,779
Intron	359,949	347,684	343,875	352,415	544,971
Intergenic	456,956	439,499	440,563	451,767	696,412
Other	157	153	154	159	241
Total	858,949	826,890	824,706	844,584	1,303,225

Table S2. Mapping summary of each drone sample, Related to Figure 1.

Drones	Tissue	Total reads (M)	Reference genome		Masked genome	
			Uniquely mapped reads (M)	Uniquely mapped reads (%)	Uniquely mapped reads (M)	Uniquely mapped reads (%)
D1	Head	25.0	17.9	71.6	17.8	71.1
	Thorax	25.8	16.8	65.1	16.1	62.4
	Abdomen	23.5	17.2	73.2	17.2	73.1
D2	Head	18.8	14.2	75.3	14.1	74.8
	Thorax	30.8	18.4	59.6	17.1	55.6
	Abdomen	28.5	21.5	75.3	21.4	75.2
D3	Head	37.2	12.4	33.4	13.0	34.9
	Thorax	32.9	10.2	31.1	10.6	32.3
	Abdomen	43.1	12.8	29.8	13.4	31.2
D4	Head	48.4	15.2	31.4	15.5	32.1
	Thorax	31.9	10.4	32.6	10.7	33.4
	Abdomen	32.9	9.2	28.1	9.5	28.8

Reference genome: the reads mapped to the reference honeybee genome.

Masked genome: the reads mapped to the genome sequences which have been replaced with the alternative alleles at SNP sites.

Table S4. Number of editing sites in each sample, Related to Figure 1 and Figure2.

Drones	Tissue	Total sites	Nonsyn (*N*)	Syn (*S*)	*N*/*S*	*P* value
	Head	278	72	8	9.0	1.59E-05
D1	Thorax	71	13	4	3.3	0.610
	Abdomen	94	14	1	14.0	0.0491
	Head	262	73	7	10.4	4.36E-06
D2	Thorax	70	12	5	2.4	1
	Abdomen	96	14	1	14.0	0.0491
	Head	177	44	6	7.3	0.00318
D3	Thorax	40	3	0	Inf	0.558
	Abdomen	55	0	0	0.0	1
	Head	208	59	6	9.8	6.67E-05
D4	Thorax	41	3	1	3.0	1
	Abdomen	60	5	1	5.0	0.674
All head		376	108	8	13.5	8.46E-10
All thorax		106	14	5	2.8	0.807
All abdomen		137	16	1	16.0	0.0318
All		407	111	9	12.3	1.02E-09

Totally 407 unique editing sites were identified. The numbers of editing sites in each sample are listed respectively. “All head” is the number of sites that appear in at least one head sample among the four individuals. The same goes for “All thorax” and “All abdomen”.

P value: The *P* value of the observed *N*/*S* ratios compared to the expected *N*/*S* ratio under neutral evolution calculated from Fisher’s exact tests.

Table S5. Conserved coding editing sites between honeybee and bumblebee, Related to Figure 3.

Apis mellifera			*Bombus terrestris*		
Site	Gene	Editing level	Site	Gene	Editing level
Group8.6:728981	GB40519	0.48	Group16.3:2744556	XM_003401834	0.87
Group8.6:728990	GB40519	0.42	Group16.3:2744565	XM_003401834	0.41
GroupUn131:18550	GB46982	0.31	Group10.1:6435961	XM_003398309	0.93
GroupUn131:18489	GB46982	0.79	Group10.1:6436022	XM_003398309	0.97
GroupUn131:18488	GB46982	0.81	Group10.1:6436023	XM_003398309	0.98
Group1.29:1345934	GB47508	0.50	Group1.8:1831958	XM_003393355	0.88
Group1.29:1345950	GB47508	0.34	Group1.8:1831942	XM_003393355	0.79
Group15.19:753243	GB50090	0.42	Group15.5:2877069	XM_003401089	0.45
Group1.29:1555391	GB47375	0.70	Group1.8:1578895	XM_003393369	0.63

Table S6. Mapping summary of each worker sample, Related to Figure 4.

Sample	Total reads (M)	Uniquely mapped reads (M)	Uniquely mapped reads (%)
SRR445999	213.8	118.2	55.3
SRR446000	295.4	203.8	69.0
SRR446001	152.0	95.2	62.6
SRR446002	154.5	87.9	56.9
SRR446003	138.6	89.2	64.4
SRR446004	226.0	149.9	66.3
Nurse pool	1180.3	744.2	63.0
SRR446005	178.9	97.5	54.5
SRR446006	193.1	130.3	67.5
SRR446007	121.2	72.8	60.1
SRR446008	164.6	100.2	60.9
SRR446009	179.7	119.4	66.5
SRR446010	139.4	84.5	60.6
Forager pool	976.8	604.6	61.9

Table S9. Shared genes with CDS editing across three species, Related to Figure 5.

D. melanogaster	*A. mellifera*	*B. terrestris*
FBgn0086372	GB43906	XM_003403010
FBgn0263354	GB48155	XM_003396523
FBgn0263111	GB51897	XM_003394356
FBgn0035538	GB54467	XM_003395885
FBgn0004242	GB54827	XM_003402654
FBgn0262483	GB55567	XM_003402632

The gene IDs of *B. terrestris* are retrieved from the bumblebee study (Porath et al. 2019) according to the editing site.

Table S10. Number of hyper-editing sites in each sample, Related to Figure 1.

Drones	Tissue	Total sites	5'UTR	Nonsyn	Syn	3'UTR	Intron	Intergenic	other
D1	Head	1,644	7	86	50	72	509	915	5
	Thorax	1,576	32	182	100	58	387	801	16
	Abdomen	2,015	29	215	114	65	607	977	8
D2	Head	2,416	110	290	170	37	660	1,110	39
	Thorax	484	19	54	44	46	143	172	6
	Abdomen	2,202	64	221	153	68	730	954	12
D3	Head	2,393	37	196	135	52	817	1,143	13
	Thorax	571	0	44	32	9	168	318	0
	Abdomen	1,019	6	57	49	17	291	598	1
D4	Head	1,756	11	145	92	29	495	977	7
	Thorax	1,716	18	189	102	42	602	755	8
	Abdomen	306	0	18	8	17	123	138	2
Pool		12,436	276	1,355	845	488	3,784	5,591	97

Transparent Methods

Honeybee collection

Drones were raised in a colony of *Apis mellifera* by professional beekeepers in Jie Wu's Lab, Chinese Academy of Agricultural Sciences. For each male drone adult (haploid), the head and thorax were separated with surgical scissors and the testis was dissected. Residual abdomen tissues were preserved for DNA extraction. All samples were flash-frozen in liquid nitrogen and stored at -80°C for further procedures.

RNA extraction and mRNA-Seq of four drone individuals

Four drone individuals were selected for mRNA-Seq. Total RNA was extracted from head, thorax, and testis from each individual, separately, using TRIzol reagent (Thermo Fisher). For two individuals (1 and 2), Poly(A)+ mRNAs were selected on oligo-dT25 DynaBeads (Thermo Fisher), while for the other two (3 and 4) RNAs were treated with Ribo-Zero Gold rRNA Removal Kit (Illumina) to remove the rRNA-derived fragments. All these purified RNAs were fragmented. The 40-80 nt fragments were purified from 15% TBE-Urea gels for deep sequencing, and subject to 3'-dephosphorylation with T4 Polynucleotide Kinase (NEB), 3'-ligation, 5'-phosphorylation with T4 Polynucleotide Kinase (NEB) and ATP, 5'-ligation, and reverse-transcription into cDNA with SuperScript™ III Reverse Transcriptase (Thermo Fisher). The cDNAs were PCR-amplified and size-selected in 20% TBE gels for fragments in correct ranges. Purified products were prepared for quality tests (Fragment Analyzer, Agilent Technologies) and sequencing (Illumina HiSeq-2500 sequencer; run type: single-end; read length: 50 nt).

Genomic sequencing of four drone individuals

To efficiently exclude SNPs from RNA editing sites, genomic DNA from abdomen tissue of each individual drone separately was extracted using the Genomic DNA Extraction Kit (TIANGEN) following manufacturer's instructions. The library preparation and sequencing were performed in Biomedical Pioneering Innovation Center, Peking University (Illumina HiSeq-2500 sequencer; 100bp paired-end reads).

Identification of SNPs in four drones

Honeybee reference genome sequence (*A. mel* 4.5) was downloaded from BeeBase (http://hymenopteragenome.org/beebase/). For each of the four individual drones, we mapped

the DNA-Seq reads to the reference genome with BWA v0.7.4 (Li and Durbin 2009). PCR duplicates were removed using Picard v1.119, and SNPs were called using SAMtools mpileup (Li 2011) with default parameters.

Identification of variation sites in four drones

To identify reliable A-to-I RNA editing events, we employed two different aligners STAR (2.4.2a) (Dobin et al. 2013) and BWA (Li and Durbin 2009) to map the RNA-Seq reads to the reference genome (*A. mel* 4.5). For each BAM sequence alignment file, we extracted all the alignments with mapping quality ≥ 10 using SAMtools 1.3.1 (Li 2011). The mismatches between RNA-Seq reads and reference genome were extracted using Sam2Tsv (Pierre 2015). Bases with mapping quality lower than 30, soft clipping bases, bases at the 10bp of reads' ends and mismatches in repeat regions were discarded. Only mismatch sites supported by both STAR and BWA were retained.

Next, we produced for each individual drone its own version of the genome, by substituting the SNP sites found in this individual to the reference genome. For example, in individual drone1, 858,949 SNPs were identified, and the reference genome (*A. mel* 4.5) was modified by replacing the reference alleles with the alternative alleles at all these SNP sites of drone1. The modified genome was regarded as the masked genome for drone1. Note that drone individuals are haploids and there are no allele-specific SNPs. Thus, the masked genome is the entire haploid genome sequence of each individual. We then mapped again the RNA-Seq reads of drone 1 to its masked genome, using both STAR and BWA as before. The same pipeline was applied to drones 2, 3, and 4.

Ideally, one would need to map the reads only to the masked genome (presumably, the true genome of each individual drone). However, our SNP detection is imperfect, and false positive SNPs identification could lead to RNA-DNA variations against the masked genome. Thus, to be conservative, we retained only variation sites supported by mapping to both the reference genome and the masked genomes. These RNA-DNA variations are unlikely to include many variations due to genomic polymorphisms, and should be enriched in RNA level alterations. We further discarded variations occurring in less than three out of twelve samples (head, thorax, and abdomen of four honeybee individuals), resulting in 1,742 candidate variation sites. Of these, 1417 (81.3%) were A-to-G.

Defining A-to-I RNA editing sites

Finally, we pooled the reads of the four drones for each of the tissues, and calculated the probability $P_k(E_0)$ that the mismatches at position E_0 (each of the 1,742 sites found above) observed in tissue k (k = head, thorax, or abdomen) can be explained by sequencing error (binomial test, with A-to-G error rate $\varepsilon = 0.00167$ (Duan et al. 2017), followed by Benjamini-Hochberg multiple testing correction with FDR = 0.05 (Benjamini and Hochberg 1995)). The multiple testing correction took into account all ~230 Mbp of the honeybee genome (excluding "N"s), at which the above variations may have been detected. Pooled data was analyzed separately for each tissue.

Editing level was estimated as G/(G+A), in which G is the alternative allele count and A is the reference allele count.

For the downloaded deep-sequencing data of brains of honeybee workers (Herb et al. 2012), we mapped the reads with STAR and called variants with SAMtools mpileup (Li 2011) with default parameters as we mentioned above. We directly retrieved the reads counts on the candidate editing sites and looked at the editing levels. Alignments with mapping quality ≥ 10 were maintained and no other filters were applied. Editing levels was compared between different castes or sub-castes using Fisher's exact test. Editing levels in other species (e.g. sites in genes *Shab* and *qvr*), were evaluated based on the datasets described below. We mapped the sequencing reads of each species to the reference genome using STAR (Dobin et al. 2013). The variants were called with SAMtools mpileup (Li 2011). The genomic coordinates were transferred from *D. melanogaster* to the other fly species by using liftOver chain downloaded from UCSC Genome Browser (http://genome.ucsc.edu/).

Annotation of A-to-I RNA editing sites

We used the software SnpEff version 4.3 (Cingolani et al. 2012b) for functional annotation of the editing sites. The canonical transcript of each gene, as defined by the software, was chosen for annotation.

Expected *N*/*S* ratio

To find the expected *N*/*S* ratio under neutrality, we replaced all adenosines in all coding regions with guanosines, one at a time, and used the software SnpEff (Cingolani et al. 2012a), (version 4.3, parameters "-ud -canon -v amel_OGSv3.2) to determine how many of these

substitutions would cause a nonsynonymous or synonymous change. In total, 4,492,737 nonsynonymous and 1,986,130 synonymous substitutions were obtained, and therefore the expected *N*/*S* ratio was 2.26. To verify that this ratio is not biased by editing being limited to a small set of genes, we have repeated the calculation focusing only on adenosines in genes harboring editing sites. Here we have found 68,220 nonsynonymous and 30,150 synonymous substitutions were obtained, and therefore the expected *N*/*S* ratio was very similar, 2.26. We therefore used the ratio 2.26 is all further analyses.

Conservation analysis

Orthologous genes (between *D. melanogaster* and honeybee, or between honeybee and bumblebee) were defined as reciprocal best hits of pairwise BLASTP (Camacho et al. 2009). The protein sequences of the orthologous genes were then aligned with the clustalw program (Thompson et al. 1994), and the alignments of RNA CDS were achieved by the tranalign program (Rice et al. 2000) based on the corresponding protein alignments. We used *D. melanogaster* reference genome version FlyBase_r6.04. For the bumblebee genome we used NCBI reference sequence version Bter_1.0. NCBI RefSeq annotations for genes and coding regions were downloaded from the BeeBase site (http://hymenopteragenome.org/beebase) on 7 May, 2014.

To calculate the expected numbers of genes sharing CDS editing at different positions within the same gene we first determined the set of expressed genes in in each of the three species as those genes with RPKM (reads per kilobase per million mapped reads) values exceeding a given cutoff. Then, we have randomly chosen 53 honeybee genes, 101 bumblebee genes and 312 fly genes (the numbers of genes harboring editing sites in CDS per species, excluding sites conserved in any two or three species), and looked for the number of orthologous gene groups reoccurring in any two of these sets or in all three of them. For each expression cutoff, the distribution of the expected overlap was calculated using 1,000,000 such random choices, and the p-value was calculated comparing the actual overlap to this distribution.

mRNA secondary structure prediction

We folded the mRNAs of honeybee with RNALfold (Lorenz et al. 2011). Local structures with Z score < -3 were defined as stable hairpins.

PCA Analysis

The PCA analysis of editing levels and gene expression levels was performed by function "princomp()" in R. The first two principal components (PC1 and PC2) were used to plot the graph.

Supplemental Note

The hyper-editing pipeline detects many sites in lowly expressed regions

We used the hyper editing script (Porath et al. 2014) with parameters suitable for 50bp reads "0.05 0.8 30 0.6 0.1 0.8 0.2" (Porath et al. 2014) to identify identified densely edited clusters, with a total of 12,436 unique sites. Of these, only 2,056 (17%) are located in repetitive regions. This fraction is higher than the one found by our main procedure, but still much lower than the one observed in most other species analyzed. A partial explanation is the rather short reads, preventing reliable alignment of heavily edited reads.

Hyper-editing sites are clustered. Clustering together editing sites (distance < 100 bps) we find that hyper-editing sites tend to cluster in big clusters (larger than the ones required for their detection) (Figure S16). Notably, half of regular editing sites are also clustered.

The sequence context of these hyper-editing sites is similar to the pattern observed in bumblebees (Porath et al. 2019), where the upstream nucleotide shows a preference of A > T > C > G and the downstream nucleotide exhibits A > T > G > C (Figure S17). In each sample, 306~2,416 hyper editing sites were detected in different tissues of four honeybee individuals (Table S10). Heads bear the greatest numbers of hyper editing sites than other tissues (Figure S18), although the sites from heads of different individuals do not show much overlap (Figure S19).

Compared to the editing sites identified by our regular pipeline, those hyper-editing sites have a remarkably higher fraction in intergenic and intron regions (Figure S20). In fact, only 52 of these 12,436 sites overlap with the 407 editing sites identified by traditional pipeline, and 45 of the overlapped sites are located in intergenic or intronic region. This result agrees with the finding that hyper-edited regions are often lowly expressed (Porath et al. 2017), and corroborated by the sequencing coverage spectrum which for hyper-editing sites is skewed to lower depths compared to normal editing sites (Figure S21). None of the hyper-editing sites overlap with conserved or task-related editing sites in the bumblebee, which further suggests

that hyper-edited regions are weakly expressed and likely to be non-conserved.

Supplemental References (Related to Supplemental Figures, Supplemental Tables, and Transparent Methods)

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289-300.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012a. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80-92.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**: 80-92.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

Duan Y, Dou S, Luo S, Zhang H, Lu J. 2017. Adaptation of A-to-I RNA editing in Drosophila. *PLoS Genet* **13**: e1006648.

Herb BR, Wolschin F, Hansen KD, Aryee MJ, Langmead B, Irizarry R, Amdam GV, Feinberg AP. 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat Neurosci* **15**: 1371-1373.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987-2993.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6**: 26.

Pierre L. 2015. *JVarkit: java-based utilities for Bioinformatics*.

Porath HT, Carmi S, Levanon EY. 2014. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nature Communications* **5**.

Porath HT, Hazan E, Shpigler H, Cohen M, Band M, Ben-Shahar Y, Levanon EY, Eisenberg E, Bloch G. 2019. RNA editing is abundant and correlates with task performance in a social bumblebee. *Nature Communications* **10**.

Porath HT, Knisbacher BA, Eisenberg E, Levanon EY. 2017. Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biol* **18**: 185.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.