Evolutionary forces on A-to-I RNA editing revealed by sequencing individual honeybee drones

Yuange Duan[1,2], Shengqian Dou[1], Jiaxing Huang[3], Eli Eisenberg[4*], and Jian Lu[1*]

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. 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
4. Raymond and Beverly Sackler School of Physics and Astronomy and Sagol School Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel

*Correspondence:
Jian Lu
Email: luj@pku.edu.cn

Eli Eisenberg
Email: elieis@post.tau.ac.il

Abstract

Social insects exhibit extensive phenotypic diversities among the genetically similar individuals, suggesting a role for the epigenetic regulations beyond the genome level. Recent studies propose that the ADAR-mediated adenosine-to-inosine (A-to-I) RNA editing, an evolutionarily conserved mechanism, facilitates adaptive evolution by expanding proteomic diversities temporally and spatially. Here, we characterize the A-to-I RNA editome honeybees (*Apis mellifera*). We collected heads, thoraxes, and abdomens of four haploid drone individuals, and deep sequenced the genomes and transcriptomes. We systematically identified 464 A-to-I editing sites and found that editing events in honeybees, especially in heads, are positively selected. There are four recoding and one synonymous editing site highly conserved between honeybee, bumblebee, and *Drosophila*. We discovered a diverged auto-editing site in *Adar* mRNA in bees and flies, which might play an auto-regulatory role in the two clades. We also found a recoding site in honeybee gene *tipE* that was reported to be task-related in bumblebee. Besides, although only a few editing sites are conserved between bees and flies, the species-specific editing sites exhibit convergent adaptation. We also detected thousands of hyper-editing sites in the lowly expressed regions, which possibly represents the non-specific targeting by ADAR.

Keywords: A-to-I RNA editing; Honeybee; Adaptation

Introduction

Adenosine (A) to inosine (A-to-I) RNA editing is an evolutionarily conserved mechanism that expands RNA diversity at the co-transcriptional or post-transcriptional level in metazoans (*1, 2*). 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 (*3-6*), microRNA (miRNA) biogenesis or target recognition (*7-9*), and mRNA translation (*10-13*). 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 (*14-18*), and dysregulation of A-to-I editing might be associated with cancer, autoimmune disorders, or other human diseases (*1*).

ADAR (adenosine deaminase acting on RNA) family are the enzymes that catalyze A-to-I editing (*19-22*). The substrates of ADAR are usually double-stranded RNAs (*20, 22-25*). Mammalian genomes encode three *Adar* genes: ADAR1 and ADAR2, which are expressed ubiquitously, and ADAR3 which is primarily expressed in the brains but catalytically inactive(*1, 2*). Human ADAR1 primarily recognizes the double-stranded RNAs (dsRNAs) that are formed by repetitive sequences such as *Alu* repeats. More than 100 million adenosine sites were edited in humans, and the majority of them were located in *Alu* repeats but lowly edited (*26*). In mammalian cells, the binding of MDA5 to dsRNAs will activate the interferon response, and this cytosolic innate immune response can be suppressed by the ADAR1-mediated editing of dsRNAs (*27-30*). Mammalian ADAR2 primarily convert A-to-I editing in the coding (CDS) or untranslated regions (UTRs) of mRNAs. ADAR2-catalyzed editing can cause a Q-to-R change in *GluA2*, which is a subunit of the AMPA receptor that controls Ca^{2+} permeability (*31*). ADAR2-catalyzed editing in the serotonin 5-hydroxytryptamine$_{2c}$ (5-HT$_{2c}$) receptor can change the protein sequences and reduce the affinity of the receptor for its G protein (*32*). There is only one *Adar* locus in *Drosophila* (*33, 34*), and it is predominately expressed in the nervous system (*35*). In *Drosophila*, A-to-I editing mainly occurs in mRNAs, suggesting *Drosophila Adar* is orthologous to the mammalian *Adar2* gene (*1, 2*). The abolition of *Adar* in *D. melanogaster* severely affects its viability and behavior (*19, 20, 36*). Overall, although A-to-I editing is an evolutionarily conserved mechanism, the editing systems have diversified in mammals after its divergence

with invertebrates.

The A-to-I RNA editing sites have been systematically characterized in various metazoan species during the past decade (*37, 38*). The majority of RNA editing sites are located in clusters in non-coding regions of humans (*39-48*), monkeys (*49, 50*), mice (*51*), worms (*52, 53*), and coral (*54*). Nevertheless, a considerable number of editing sites are clustered in mRNA regions in *Drosophila* (*55-62*), ants (*63*), and cephalopod species (*64, 65*). Only a very few editing targets are conserved between virtually all recoding sites in mammals, *Drosophila*, and cephalopods (*1*). Therefore, despite the deep conservation of the A-to-I editing mechanism, the target landscapes of editing have considerably evolved during metazoan evolution. The editing level is mainly regulated by the expression level of ADAR, and the sequence context flanking the focal editing sites as previous studies found cis-elements affect editing in humans (*66*) and *Drosophila* (*60, 67, 68*). Overall, the current results suggest that RNA editing modulates the diversity of the transcriptomes and proteomes in a lineage-specific manner, although the evolutionary driving forces behind the evolution of A-to-I editing across species at the macro-evolutional scale requires further investigation.

RNA editing was hypothesized to facilitate adaptive evolution by increasing proteomic diversities temporally or spatially, in a more flexible manner (*13, 69-72*). However, only a few examples of the advantageous effects conferred by RNA editing have been demonstrated (*12, 13, 64, 70*). For example, the A-to-I editing events in *Kv1* mRNA provide numerous adaptive amino acid changes that allow the octopus to adapt to extremely cold temperatures (*73*). In corals, ADAR and editing levels peak at the time of gamete release during spawning, resulting in over a thousand recoding events, which increases the survival probability for gametes in the dynamic and varied environmental conditions they encounter in the early developmental stages (*74*). Another example of how RNA editing contributes to adaptation occurs in *Drosophila melanogaster* that lives in the two opposing slopes of 'Evolution Canyon' in Israel that have divergent microclimates. Researchers found dozens of A-to-I edits might participate in environmental adaption to the microclimates (*75*).

A-to-I editing has the advantage of quick response to environmental stress and adjusts the activity of final protein products (*76, 77*). Accordingly, several previous studies, including ours, have demonstrated that in *Drosophila*, RNA editing level changes in response to

temperature(*60, 77, 78*). It is possible that the editing level change might recode the amino acid to compensate for the protein-folding caused by altered temperature(*79*).

At the editome levels, the evidence of adaption of A-to-I editing has been found in *Drosophila* and cephalopods. *Drosophila* is a powerful model system that is under the intense study of A-to-I RNA editing (*1, 2*). Previous studies have revealed hundreds of editing sites that are conserved in the coding regions across *Drosophila* species that span 60 million years of evolution (*55-62*). In particular, we have previously identified over 2,000 A-to-I editing sites in brains of *D. melanogaster* (*60*). By comparing the observed nonsynonymous *(N)* to synonymous (*S*) editing sites versus the expected *N/S* ratio under randomness (neutrality), we found abundant nonsynonymous editing events in *Drosophila* brain are adaptive and maintained by natural selection. Furthermore, we found the signal of adaptive evolution is more pronounced in the sites that have conserved editing events between *D. melanogaster* and its sibling species. Beyond that, we later found that many of the nonsynonymous editing events are linked in the same mRNA molecules, highlighting the importance of considering the combined effect of the linked editing events in the functional annotation (*80*). Similarly, in cephalopods, it has been shown that A-to-I RNA editing is particularly common in behaviorally sophisticated cephalopods, with tens of thousands of evolutionarily conserved sites (*64, 65, 73*). Editing is enriched in the nervous system, affecting the molecules pertinent to excitability and neuronal morphology.

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 genome. The social insects provide us a system to study how phenotypic diversity is regulated among these genetically similar individuals(*81, 82*). Honeybee (*Apis mellifera*), as an important pollinator in the ecological environment, also presents complicated social behaviors, such as the division of labor for reproduction or foraging(*83*). Genetic mapping demonstrated that phenotypic plasticity was associated with complex epistatic and pleiotropic genetic networks that influence reproductive regulation and foraging behaviors (*83*). Compared to DNA substitution, RNA modifications such as A-to-I editing contributes to proteomic diversity beyond the genomic level with more flexibility (*13, 69-72*).

Here, we investigate the A-to-I RNA editome in different tissues of four honeybee drone individuals. Notably, the drone of the honeybee is haploid, enabling us to detect A-to-I editing events with high accuracy. In this study, we systematically identified 464 A-to-I sites in multiple tissues of the honeybee by deep sequencing of the genome and transcriptome. We found that editing events in honeybees, especially in heads, are positively selected. There are four recoding and one synonymous editing site highly conserved between honeybee, bumblebee, and *Drosophila*. We also discovered a diverged auto-editing site in *Adar* mRNA in bees and flies, which might play a similar auto-regulatory role in the two clades. Although only a few editing sites are conserved between bees and flies, the species-specific editing sites exhibit convergent adaptation. We also detected thousands of hyper-editing sites in the lowly expressed regions, which possibly represents the promiscuous targeting by ADAR.

Materials and Methods

Honeybee collection

The sequencing experiment was performed using drones from a colony of *Apis mellifera*, which were raised by professional beekeepers in Jie Wu's Lab, Chinese Academy of Agricultural Sciences. For each drone adult (haploid), the head and thorax were separated with surgical scissors, and the testis was dissected, with residual abdomen tissues 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, and total RNAs were extracted from head, thorax and testis from each individual with TRIzol reagent (Thermo Fisher), separately. For total RNAs from two individuals (individuals 1 and 2), Poly(A)+ mRNAs were selected on oligo-dT25 DynaBeads (Thermo Fisher). For total RNAs from the other two individuals (individuals 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, and 40-80 nt RNA fragments were purified from 15% TBE-Urea gels for deep sequencing. Then were sequentially 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 subjected to amplification by PCR and size-selected in 20% TBE gels for fragments within 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 the SNPs from RNA editing sites, we extracted genomic DNA from abdomen tissues from each drone individual mentioned above with the Genomic DNA Extraction Kit (TIANGEN) separately, following the manufacturer's instructions. The library preparation and sequencing were performed in Biomedical Pioneering Innovation Center, Peking University (Illumina HiSeq-2500 sequencer; run type: paired-end; read length: 100 nt).

Identification of SNPs in four drones

The sequence of the honeybee reference genome (*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 (*84*). PCR duplicates were removed using Picard v1.119, SNPs were called using SAMtools mpileup (*85*).

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) (*86*) and BWA (*84*) 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 (*85*). The mismatches between RNA-Seq reads and the reference genome were extracted using Sam2Tsv (*87*). Soft clipping bases in the alignments were not considered, and only the mismatch sites supported by both STAR and BWA were retained. Mismatches in repeat regions were removed.

Next, we masked the reference genome by substituting the SNP sites in each individual, and mapped the RNA-Seq reads to the masked genomes again with STAR and BWA. The variation sites supported by mapping to both the reference genome and the masked genomes were kept. This excludes the false-positive sites confounded by SNPs. Among the variations in twelve samples, including the head, thorax, and abdomen of four honeybee individuals, we

required a variation site to appear in at least three libraries. Under these criteria, we obtained 1,742 candidate variation sites.

Defining A-to-I RNA editing sites

For each candidate site, we first pooled the reads of four drones and identified editing sites in heads, thoraxes, and abdomens, respectively. Based on the pipeline we developed in our previous work, we calculated the probability $P_k(E_0)$ that the A-to-G mismatch in tissue k (k = head, thorax, or abdomen) is solely caused by sequencing error (the A-to-G error rate ε = 0.00167) (*60*). The probability for each candidate site was adjusted for multiple testing correction (*88*). Under adjusted P value (FDR) < 0.05, we obtained 499, 205, and 245 variant sites in heads, thoraxes, and abdomens, among which 432 (86.6%), 146 (71.2%), and 179 (73.1%) of the variations were A-to-G alterations. If we pooled these variations in three tissues, we obtained in total 580 unique variant sites, and 464 (80.0%) of them were A-to-G variation. These 464 A-to-G sites were regarded as putative A-to-I RNA editing sites. The editing level is defined as G/(G+A), in which G is the alternative allele count, and A is the reference allele count.

Annotation of A-to-I RNA editing sites

The functional annotations of the editing sites were inferred with the software SnpEff version 4.3 (*89*). The canonical transcript of each gene defined by the software was chosen for annotation.

Conservation analysis

We parsed the genomic coordinates with the pairwise CDS alignments between *D. melanogaster* and honeybee. The reciprocal best orthologous genes were obtained based on pairwise BLASTP (*90*) between *D. melanogaster* and honeybee. The protein sequences of the orthologous genes were aligned with the clustalw program (*91*), and the CDS alignments were achieved by the tranalign program (*92*) based on the corresponding protein alignments.

mRNA secondary structure prediction

We folded the mRNAs of honeybee with RNALfold (*93*). The local structures with Z score < -3 were defined as stable hairpins.

Data availability

All deep-sequencing data generated in this study were deposited in the Genome Sequencing

Archive (GSA) under accession number CRA002262.

Results and discussion

1. Editome of honeybee

We deep sequenced the genomic DNA of four individual drones as well as their transcriptomes (Fig. 1A). For each individual, we dissected the head, thorax, and abdomen, and therefore, we totally sequenced the transcriptomes of 12 samples. 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 Ribo-Zero kit to deplete ribosomal RNAs (Materials and Methods). The male drone is haploid, which gives us the power to accurately identify RNA editing sites by excluding the SNPs from the mismatches between RNA-Seq reads and the reference genome (Fig. 1B).

We mapped the DNA-Seq reads to the reference genome and obtained 826,890~858,949 SNPs in each of the four drones (Table S1). For the RNA-Seq data, we obtained 9.2~21.5 million reads per sample that could be uniquely mapped to the genome (Table S2). We used a two-step strategy to identify RNA editing sites by masking the SNPs in the genome (Materials and Methods). We pooled the reads from different individuals and identified editing sites in three tissues, respectively. With FDR < 0.05 (Materials and Methods), we identified 432 (86.6%) A-to-G sites among the 499 variations in heads, 146 (71.2%) A-to-G sites among the 205 variations in thoraxes, and 179 (73.1%) A-to-G sites among the 245 variations in abdomens (Fig. S1). In total, we obtained 464 (80.0%) A-to-G sites among the 580 variations (Fig. 1C). These 464 A-to-G changes are regarded as A-to-I RNA editing sites. The nucleotide context around these putative editing sites is consistent with the known ADAR binding motif, while the context around unedited adenosines or SNP sites do not show any preferences (Fig. S2).

Among the 464 editing sites we identified, 227 sites are located in gene regions, including 131 nonsynonymous (*N*) sites, 10 synonymous (*S*) sites, and several sites in UTRs and introns (Fig. 1D and Table 1). The other 237 sites are annotated as intergenic. Note that editing sites in heads have a higher fraction in coding regions compared to the other two tissues (Fig. S3). Besides, due to the different library construction strategies, we found that

sites in drones 1 and 2 have higher fractions in coding regions while drones 3 and 4 retrieved more sites in intergenic regions (22.0 ± 2.9% in CDS in drones 1 and 2 versus 14.3 ± 5.3% in CDS in drones 3 and 4; 57.8 ± 2.5% in intergenic regions in drones 1 and 2 versus 71.6 ± 5.7% in intergenic regions in drones 3 and 4).

In each sample, we identified 50~314 editing sites (Table 1). Compared to thorax and abdomen, the number of editing sites is remarkably greater in heads (Fig. 2A). The editing sites among four drones show moderate proportions of overlapping (Fig. S4). The editing levels are generally higher in heads than other tissues (Fig. 2B). Consistent with the observed higher editing efficiency in heads, the ADAR expression is higher in heads than other tissues (Fig. S5).

2. Signals of adaptation

Among the 464 editing sites, the nonsynonymous to synonymous ratio (*N*/*S*) is 131/10 = 13.1 (Table 1). If we change all the adenosines in edited genes to guanosines, the expected *N*/*S* ratio is 2.26. This suggests that the nonsynonymous editing events in honeybees are overall adaptive, which is similar to the pattern we previously found in brains of *Drosophila* (*60*). In each sample, the observed *N*/*S* ratio is generally higher than neutral expectation, especially in heads (Table 1). Analogous results are obtained if we pool the sites from different individuals together and look at the *N*/*S* ratio in the three tissues (Fig. 2C, "All sites") or if we only focus on the sites that appear in at least two individuals (Fig. 2C, "In ≥ 2 individuals"). Furthermore, in heads of honeybees, the *N* sites have significantly higher editing levels than the *S* sites when we combined the sites in four drones to increase the statistical power (Fig. 3D).

Interestingly, in our previous study in fly brains, we have discovered higher *N*/*S* ratio for the X-linked genes than autosomal genes (*60*). However, honeybees and flies have different sex-determination systems, and accordingly, we observed even lower *N*/*S* ratios in the honeybee genes which are orthologous to the X-linked genes in flies (*N*/*S* = 8.0 for all sites and 7.0 for sites in heads), suggesting that nonsynonymous editing in these honeybee genes are not subjected to additional selection as their counterparts in flies.

In our previous study in *Drosophila*, we have defined a set of PSEB (Positively Selected Editing in Brain) genes with *N*/*S* ratio > 5 (*60*). With the same criteria (*N*/*S* ratio > 5), we found 62 genes with abundant nonsynonymous editing in honeybees (if we only focus on

editing sites in heads, the PSEB gene number is 60). The overall *N*/*S* ratio among PSEB genes is 125/1=125 for all editing sites and 123/1=123 for editing sites in heads, suggesting that these genes are likely the targets favored by natural selection. Due to the limited number of genes, the gene ontology enrichment could not be performed. However, we indeed found nerve-system-related genes like *Ca-alpha1D*, *DopEcR*, *eag*, *qvr*, *Shab*, *Syt1* and *tutl*. The gene expression profile reveals a significantly higher expression level for PSEB genes in heads than thoraxes or abdomens (Fig. S6).

3. Evolutionarily conserved and non-conserved editing sites

We found only four nonsynonymous and one synonymous sites that are edited in heads of both *Drosophila* and honeybees (Fig. 3A and 3B). These sites are in gene *Shab* (*shaker cognate b*, two nonsynonymous and one synonymous sites) and *qvr* (*quiver*, two nonsynonymous sites). These sites are also highly edited in *Drosophila* and honeybee heads. Interestingly, the recent study in bumblebee also found four editing sites in gene *Shab* and two sites in *qvr* that are highly conserved between fly and bumblebee (*82*), including the five editing sites we mentioned above in honeybee (Fig. 3A and 3B). These findings suggest a potential functional role of the conserved editing sites.

The majority of the editing sites are not shared between *Drosophila* and honeybees. For example, in *Drosophila*, the auto-editing site that changes Ser to Gly in *Adar* mRNA (S430G) is moderately edited and conserved among different fly species, but the adenosine site in genome is not conserved in honeybee nor bumblebee (Fig. 4A), which prevents the editing at this position. Note that the un-editable codon in bees also encodes Ser (Fig. 4A). In *Drosophila*, the auto-edited *Adar* isoform would result in a less active ADAR protein, which forms a negative feedback loop of editing activity (*21, 94*). Since the orthologous sequence in bee species is un-editable, so that the bees do not use this particular site to regulate the global editing activity.

However, in bumblebees, another two nonsynonymous auto-editing sites K481E (level 0.07) and I482M (level 0.45) were found (*82*), and the auto-editing level is positively correlated with the global editing activity. Interestingly, we also found the I482M (ATA to ATG) auto-editing site in honeybee (Fig. 4A), and the editing level in the head of individual-1 is 0.58, similar to the level reported in brains of the bumblebee. More strikingly, this Ile

codon in bees (ATA) is changed to another Ile codon ATT in flies that could not be edited at the third position (Fig. 4A). We summarize the auto-editing patterns as follows, the auto-editing S430G in flies is abolished in bee by changing the editable Ser codon to an un-editable Ser codon, and the auto-editing I482M in bees is abolished in flies by changing the editable Ile codon to an un-editable Ile codon. While the auto-editing sites in flies and bees are supposed to have a function of the feed-back loop, this case might be an example of convergent evolution of the ADAR regulation strategy.

For some other editing sites in fly heads, we also found cases of non-conserved adenosines in honeybee, e.g., several editing events in fly gene *NaCP60E* and *eag*, the adenosines of which are mutated to other nucleotides (Fig. 4B), which canceled the editability. There are also cases conserved adenosines in honeybee where the editing events are not detected in RNA-Seq data (Fig. 4B).

Apart from the conservation patterns between flies and bees, we also noticed that in bumblebees, there are a set of putative task-related editing sites in CDS that show differential editing levels in brains of nurses versus foragers(*82*). These task-related sites are bumblebee-specific and non-conserved in flies. For example, a task-related recoding site in gene *XM_003393369* in bumblebee leads to a Ser>Gly change(*82*). In honeybees, gene *GB47375* is orthologous to bumblebee *XM_003393369* and *Drosophila* gene *tipE* (*temperature-induced paralytic E*), and we found two recoding editing sites leading to Ser>Gly in this gene (Fig. 4C) and site1 is orthologous to the task-related site in bumblebee. In our honeybee data, editing levels of site1 range from 0.63~0.76, and the level of site2 ranges from 0.24~0.42 in head samples, while in other tissues, this region is lowly expressed and poorly edited. In the genomes of *Drosophila* species, the orthologous site of site1 encodes the post-edit Gly codon GGC while site2 remains unedited (Fig. 4C). These observations again suggest that the *tipE* recoding sites in bees are possibly related to the bee-specific social behavior like task performance.

4. Convergent adaptation of A-to-I editing?

Although there are only a limited number of conserved editing sites between honeybee and *Drosophila*, we wonder whether the edited genes are conserved between the two species regardless of the editing position in the genes. If we focus on the genes that have editing sites

in the coding regions, we found over 19 genes that are shared between honeybee and *D. melanogaster* (Fig. 5A). For the genes that have editing sites specifically in honeybee or *Drosophila*, we still found their *N*/*S* ratios significantly higher than the neutral expectation (Fig. 5B). These results suggest the possible convergent adaptation of A-to-I editing in the two clades. Although the conserved editing sites in CDS are rare, the species-specific editing sites have evolved respectively and exhibit signals of adaptation. The case of two different *Adar* auto-editing site in flies and bees is strong evidence indicating the convergent evolution of recoding sites in the two clades.

5. The tradeoff between RNA diversity and genome evolution

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(*95*). This conclusion is based on the fact that RNA editing sites are majorly located in dsRNA structures that allow ADAR to bind and catalyze. Consistent with the established theory, we confirmed that the editing sites in honeybees are enriched in hairpin structures compared to randomness (Fig. S7A). Similar to the results in cephalopods(*95*), SNPs are depleted in the vicinity of editing sites rather than randomly picked unedited adenosines in honeybees (Fig. S7B). This indicates the tradeoff between transcriptome diversity and genome evolution. The editing sites in honeybees are selectively maintained so that the genomic mutations close to the editing sites are suppressed.

6. Hyper-editing pipeline detects numerous new sites in lowly expressed regions

We used the hyper editing script (*96*) with the default parameter and totally identified 5,706 unique sites that have hyper-editing events detected. Only 69 of these 5,706 sites overlapped with the 1,417 editing sites identified by traditional pipeline, and 61 of the overlapped sites are located in the intergenic region. Only the non-overlapped 5,637 hyper-editing sites are discussed in the following analyses.

The sequence context of these hyper-editing sites is similar to the pattern observed in bumblebees (*82*), where the upstream nucleotide shows a preference of A > T > C > G and the downstream nucleotide exhibits A > T > G > C (Fig. S8). In each sample, 89~1,307 hyper editing sites were detected in different tissues of four honeybee individuals (Table S3). Heads bear the greatest numbers of hyper editing sites than other tissues (Fig. 6A), although the sites

from heads of different individuals do not show much overlap (Fig. S9).

Compared to the editing sites identified by our normal pipeline, those hyper-editing sites have a remarkably higher fraction in intergenic and intron regions (Fig. 6B). This result agrees with the finding that hyper-edited regions are usually lowly expressed(*97*). This notion could also be inferred from the sequencing coverage spectrum that the coverage of hyper-editing sites is skewed to lower depth compared to normal editing sites (Fig. S10). Moreover, none of the hyper-editing sites overlap with the 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. Another feature of the hyper-editing sites is that they appear in clusters. We cluster the editing sites with a distance smaller than 100 bps, we found 53.9% (250/464) of normal editing sites, and all of the hyper-editing sites are located in clusters. The distribution of cluster length also demonstrates that hyper-editing sites tend to cluster together (Fig. S11), possibly due to the promiscuous targeting of the editing enzyme.

Acknowledgment

This work was supported by grants from the Ministry of Science and Technology of the People's Republic of China (2016YFA0500800), National Natural Science Foundation of China (No. 91431101 and 91731301). 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.

Competing interests

The authors declare that they have no competing interests.

References

1. E. Eisenberg, E. Y. Levanon, A-to-I RNA editing — immune protector and transcriptome diversifier. *Nature Reviews Genetics* **19**, 473-490 (2018).
2. A. L. Yablonovitch, P. Deng, D. Jacobson, J. B. Li, The evolution and adaptation of A-to-I RNA editing. *PLOS Genetics* **13**, e1007064 (2017).
3. S. M. Rueter, T. R. Dawson, R. B. Emeson, Regulation of alternative splicing by RNA editing. *Nature* **399**, 75-80 (1999).
4. R. Flomen, J. Knight, P. Sham, R. Kerwin, A. Makoff, Evidence that RNA editing

modulates splice site selection in the 5-HT2C receptor gene. *Nucleic Acids Research* **32**, 2113-2122 (2004).

5. Y. Jin *et al.*, RNA editing and alternative splicing of the insect nAChR subunit alpha6 transcript: evolutionary conservation, divergence and regulation. *BMC Evolutionary Biology* **7**, 1-12 (2007).
6. G. Lev-Maor *et al.*, RNA-editing-mediated exon evolution. *Genome biology* **8**, R29 (2007).
7. H. Liang, L. F. Landweber, Hypothesis: RNA editing of microRNA target sites in humans? *RNA* **13**, 463-467 (2007).
8. G. M. Borchert *et al.*, Adenosine deamination in human transcripts generates novel microRNA binding sites. *Human Molecular Genetics* **18**, 4801-4807 (2009).
9. S. Alon *et al.*, Systematic identification of edited microRNAs in the human brain. *Genome research* **22**, 1533-1540 (2012).
10. A. J. Wahba *et al.*, Synthetic polynucleotides and the amino acid code. VI. *Proc Natl Acad Sci U S A* **48**, 1683-1686 (1962).
11. B. Sommer, M. Köhler, R. Sprengel, P. H. Seeburg, RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11-19 (1991).
12. K. Nishikura, Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Biol* **7**, 919-931 (2006).
13. K. Nishikura, Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry* **79**, 321-349 (2010).
14. B. L. Bass, RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **71**, 817-846 (2002).
15. M. S. Paul, B. L. Bass, Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *The EMBO journal* **17**, 1120-1127 (1998).
16. I. H. Greger, P. Akamine, L. Khatri, E. B. Ziff, Developmentally regulated, combinatorial RNA processing modulates AMPA receptor biogenesis. *Neuron* **51**, 85-97 (2006).
17. H. Wahlstedt, C. Daniel, M. Enstero, M. Ohman, Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome research* **19**, 978-986 (2009).
18. J. Ohlson, J. S. Pedersen, D. Haussler, M. Ohman, Editing modifies the GABA(A) receptor subunit alpha3. *Rna* **13**, 698-703 (2007).
19. M. J. Palladino, L. P. Keegan, M. A. O'Connell, R. A. Reenan, A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. *Cell* **102**, 437-449 (2000).
20. L. P. Keegan, A. Gallo, M. A. O'Connell, The many roles of an RNA editor. *Nature reviews. Genetics* **2**, 869-878 (2001).
21. Y. A. Savva *et al.*, Auto-regulatory RNA editing fine-tunes mRNA re-coding and complex behaviour in Drosophila. *Nat Commun* **3**, 790 (2012).
22. L. P. Keegan *et al.*, Tuning of RNA editing by ADAR is required in Drosophila. *The EMBO journal* **24**, 2183-2193 (2005).
23. M. Higuchi *et al.*, RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361-1370 (1993).
24. G. Ramaswami *et al.*, Genetic mapping uncovers cis-regulatory landscape of RNA

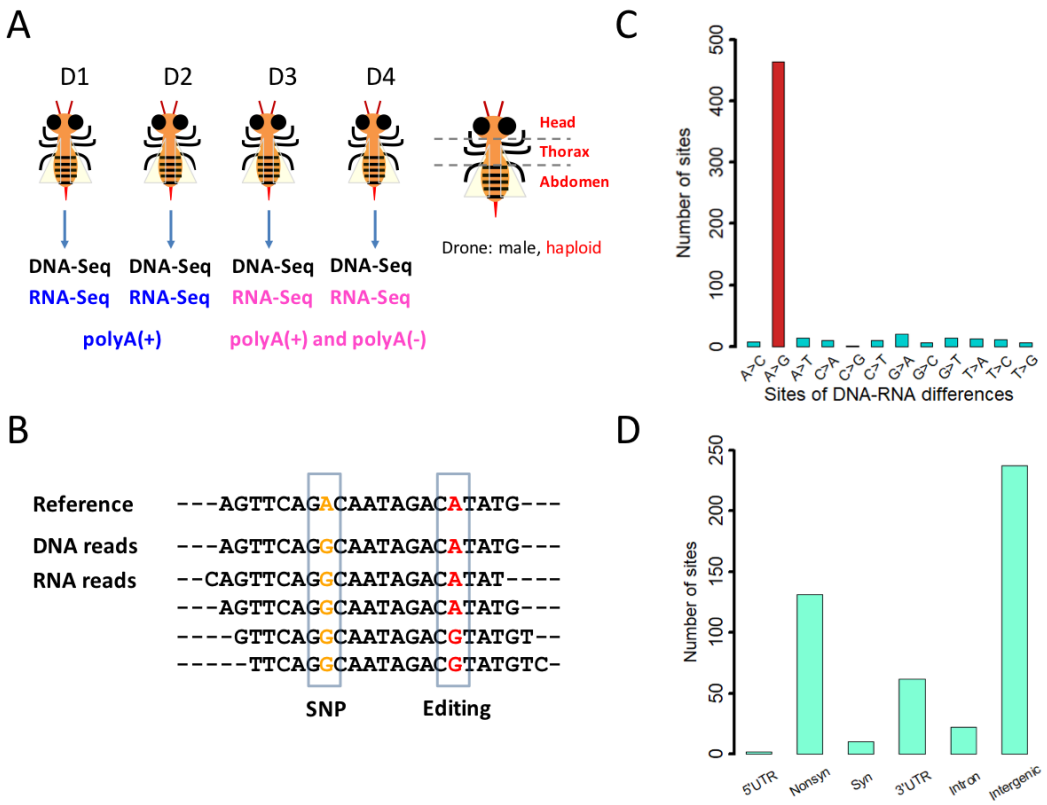
editing. *Nat Commun* **6**, (2015).
25. E. Y. Levanon *et al.*, Evolutionarily conserved human targets of adenosine to inosine RNA editing. *Nucleic Acids Research* **33**, 1162-1168 (2005).
26. L. Bazak, A-To-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, (2014).
27. B. J. Liddicoat, RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, (2015).
28. K. Pestal, Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **43**, (2015).
29. N. M. Mannion, The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, (2014).
30. Q. Feng, MDA5 detects the double-stranded RNA replicative form in picornavirus-infected cells. *Cell Rep.* **2**, (2012).
31. B. Sommer, M. Kohler, R. Sprengel, P. H. Seeburg, RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, (1991).
32. C. M. Burns, Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, (1997).
33. Y. A. Savva, L. E. Rieder, R. A. Reenan, The ADAR protein family. *Genome Biology* **13**, 1-10 (2012).
34. M. J. Palladino, L. P. Keegan, M. A. O'Connell, R. A. Reenan, dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* **6**, (2000).
35. J. E. C. Jepson *et al.*, Engineered Alterations in RNA Editing Modulate Complex Behavior in Drosophila: REGULATORY DIVERSITY OF ADENOSINE DEAMINASE ACTING ON RNA (ADAR) TARGETS. *Journal of Biological Chemistry* **286**, 8325-8337 (2011).
36. C. Maldonado, D. Alicea, M. Gonzalez, M. Bykhovskaia, B. Marie, Adar is essential for optimal presynaptic function. *Molecular and cellular neurosciences* **52**, 173-180 (2013).
37. G. Ramaswami, J. B. Li, Identification of human RNA editing sites: A historical perspective. *Methods* **107**, 42-47 (2016).
38. Y. A. Savva, G. S. Laurent, R. A. Reenan, in *Post-Transcriptional Gene Regulation,* E. Dassi, Ed. (Springer New York, New York, NY, 2016), pp. 255-268.
39. Z. Peng *et al.*, Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol* **30**, 253-260 (2012).
40. L. Bazak *et al.*, A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Research* **24**, 365-376 (2014).
41. J. B. Li *et al.*, Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science (New York, N.Y.)* **324**, 1210-1213 (2009).
42. M. Sakurai *et al.*, A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome research* **24**, 522-534 (2014).
43. G. Ramaswami *et al.*, Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods* **10**, 128-132 (2013).
44. E. Y. Levanon *et al.*, Systematic identification of abundant A-to-I editing sites in the

human transcriptome. *Nat Biotech* **22**, 1001-1005 (2004).
45. M. Blow, P. A. Futreal, R. Wooster, M. R. Stratton, A survey of RNA editing in human brain. *Genome Research* **14**, 2379-2387 (2004).
46. E. Picardi *et al.*, Profiling RNA editing in human tissues: towards the inosinome Atlas. *Sci Rep* **5**, 14941 (2015).
47. E. Picardi, A. M. D'Erchia, C. Lo Giudice, G. Pesole, REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* **45**, D750-D757 (2017).
48. P. Mazin *et al.*, Widespread splicing changes in human brain development and aging. *Molecular Systems Biology* **9**, (2013).
49. X.-Z. Yang *et al.*, Selectively Constrained RNA Editing Regulation Crosstalks with piRNA Biogenesis in Primates. *Molecular Biology and Evolution* **32**, 3143-3157 (2015).
50. J.-Y. Chen *et al.*, RNA Editome in Rhesus Macaque Shaped by Purifying Selection. *PLoS Genet* **10**, e1004274 (2014).
51. P. Danecek *et al.*, High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol* **13**, 26 (2012).
52. H. Q. Zhao *et al.*, Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. *Genome Res* **25**, 66-75 (2015).
53. B. Goldstein *et al.*, A-to-I RNA editing promotes developmental stage–specific gene and lncRNA expression. *Genome Research* **27**, 462-470 (2017).
54. H. T. Porath *et al.*, A-to-I RNA Editing in the Earliest-Diverging Eumetazoan Phyla. *Molecular Biology and Evolution* **34**, 1890-1901 (2017).
55. G. St Laurent *et al.*, Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. *Nat Struct Mol Biol* **20**, 1333-1339 (2013).
56. J. Rodriguez, J. S. Menet, M. Rosbash, Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. *Mol Cell* **47**, 27-37 (2012).
57. B. R. Graveley *et al.*, The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473-479 (2011).
58. Y. Yu *et al.*, The Landscape of A-to-I RNA Editome Is Shaped by Both Positive and Purifying Selection. *PLoS Genet* **12**, e1006191 (2016).
59. R. Zhang, P. Deng, D. Jacobson, J. B. Li, Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. *PLoS genetics* **13**, e1006563 (2017).
60. Y. Duan, S. Dou, S. Luo, H. Zhang, J. Lu, Adaptation of A-to-I RNA editing in Drosophila. *PLoS Genet* **13**, e1006648 (2017).
61. A. Mazloomian, I. M. Meyer, Genome-wide identification and characterization of tissue-specific RNA editing events in D. melanogaster and their potential role in regulating alternative splicing. *RNA Biology* **12**, 1391-1401 (2015).
62. I. Buchumenski *et al.*, Dynamic hyper-editing underlies temperature adaptation in Drosophila. *PLOS Genetics* **13**, e1006931 (2017).
63. Q. Li *et al.*, Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior. **5**, 4943 (2014).
64. S. Alon *et al.*, The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *eLife* **4**, e05198 (2015).

65. N. Liscovitch-Brauer *et al.*, Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* **169**, 191-202.e111 (2017).
66. E. Park *et al.*, Population and allelic variation of A-to-I RNA editing in human transcriptomes. *Genome Biol* **18**, 143 (2017).
67. A. L. Sapiro, P. Deng, R. Zhang, J. B. Li, Cis regulatory effects on A-to-I RNA editing in related Drosophila species. *Cell Rep* **11**, 697-703 (2015).
68. G. Ramaswami *et al.*, Genetic mapping uncovers cis-regulatory landscape of RNA editing. *Nature Communications* **6**, 8194 (2015).
69. F. D. Klironomos, J. Berg, S. Collins, How epigenetic mutations can affect genetic evolution: Model and mechanism. *BioEssays* **35**, 571-578 (2013).
70. W. M. Gommans, S. P. Mullen, S. Maas, RNA editing: a driving force for adaptive evolution? *BioEssays* **31**, 1137-1145 (2009).
71. K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**, 83-96 (2016).
72. J. J. C. Rosenthal, The emerging role of RNA editing in plasticity. *The Journal of Experimental Biology* **218**, 1812 (2015).
73. S. Garrett, J. J. Rosenthal, RNA editing underlies temperature adaptation in K+ channels from polar octopuses. *Science* **335**, 848-851 (2012).
74. H. T. Porath, A-To-I RNA editing in the earliest-diverging eumetazoan phyla. *Mol. Biol. Evol* **34**, (2017).
75. A. L. Yablonovitch *et al.*, Regulation of gene expression and RNA editing in Drosophila adapting to divergent microclimates. *Nature Communications* **8**, 1570 (2017).
76. K. Licht, M. F. Jantsch, Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *The Journal of Cell Biology* **213**, 15-22 (2016).
77. L. E. Rieder *et al.*, Dynamic response of RNA editing to temperature in Drosophila. *BMC Biol* **13**, 1 (2015).
78. I. Buchumenski *et al.*, Dynamic hyper-editing underlies temperature adaptation in Drosophila. *PLoS Genet* **13**, e1006931 (2017).
79. S. C. Garrett, J. J. C. A. Rosenthal, Role for A-to-I RNA editing in temperature adaptation. *Physiology* **27**, (2012).
80. Y. Duan *et al.*, Linkage of A-to-I RNA Editing in Metazoans and the Impact on Genome Evolution. *Molecular Biology and Evolution* **35**, 132-148 (2018).
81. Q. Li *et al.*, Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior. *Nat Commun* **5**, 4943 (2014).
82. H. T. Porath *et al.*, RNA editing is abundant and correlates with task performance in a social bumblebee. *Nature Communications* **10**, (2019).
83. R. E. Page, Jr., O. Rueppell, G. V. Amdam, Genetics of reproduction and regulation of honeybee (Apis mellifera L.) social behavior. *Annu Rev Genet* **46**, 97-119 (2012).
84. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
85. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
86. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21

(2013).
87. L. Pierre, *JVarkit: java-based utilities for Bioinformatics*. (2015).
88. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).
89. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80-92 (2012).
90. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
91. J. D. Thompson, D. G. Higgins, T. J. Gibson, 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 (1994).
92. P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277 (2000).
93. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6**, 26 (2011).
94. M. J. Palladino, L. P. Keegan, M. A. O'Connell, R. A. Reenan, 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 (2000).
95. N. Liscovitch-Brauer *et al.*, Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* **169**, 191-202 e111 (2017).
96. H. T. Porath, S. Carmi, E. Y. Levanon, A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nature Communications* **5**, (2014).
97. H. T. Porath, B. A. Knisbacher, E. Eisenberg, E. Y. Levanon, Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biol* **18**, 185 (2017).

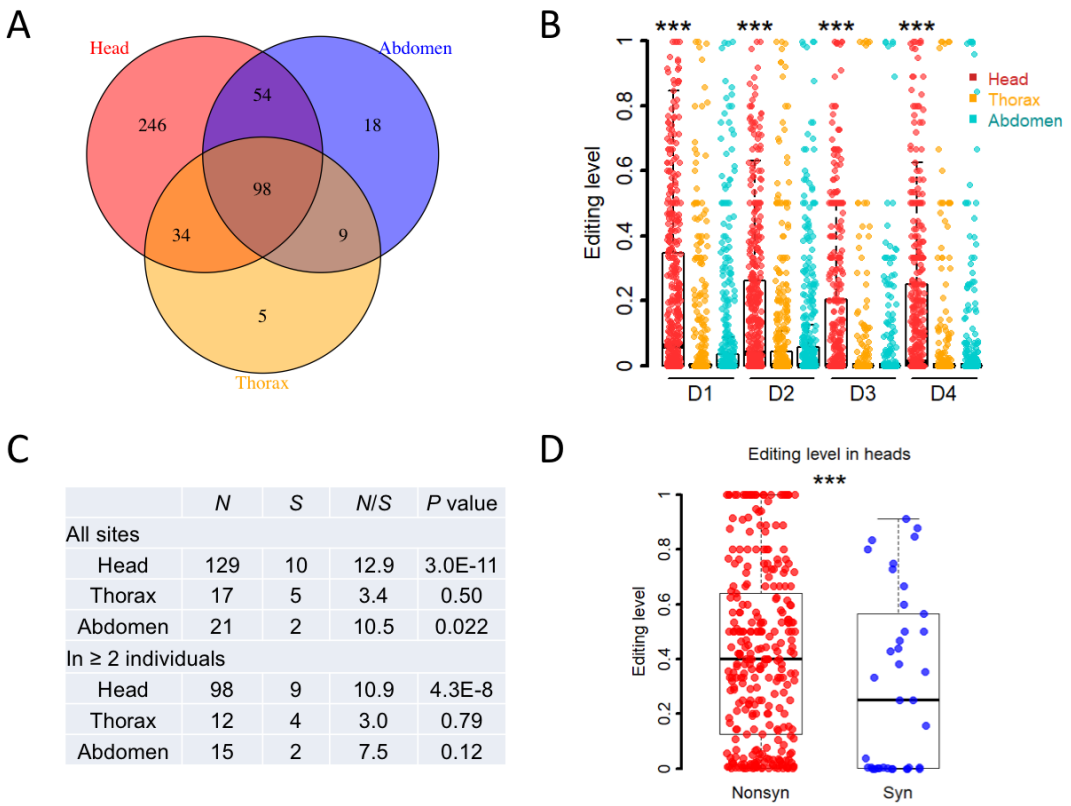
1 **Figure legends**



2

3 **Figure 1. Identification and annotation of the A-to-I RNA editing sites in the honeybee.**
4 (A) Workflow of sample collection, dissection and library construction. (B) Schematic
5 diagram illustrating the identification of A-to-I RNA editing sites. (C) Numbers of different
6 types of DNA-RNA differences detected in honeybee. (D) A summary of A-to-I editing sites
7 according to their functional annotations.

8



	N	*S*	*N*/*S*	*P* value
All sites				
Head	129	10	12.9	3.0E-11
Thorax	17	5	3.4	0.50
Abdomen	21	2	10.5	0.022
In ≥ 2 individuals				
Head	98	9	10.9	4.3E-8
Thorax	12	4	3.0	0.79
Abdomen	15	2	7.5	0.12

Figure 2. Signals of adaptation of A-to-I editing in the honeybee. (A) Venn diagram demonstrating the overlaps of editing sites in three honeybee tissues. (B) Editing levels of detected editing sites in 12 samples. The Wilcoxon rank sum tests were performed to detect the statistical significance between levels in heads and other tissues (***, $P < 0.001$). (C) The observed numbers of nonsynonymous (*N*) and synonymous (*S*) editing sites in each tissue. The *P* values of the observed *N*/*S* ratios compared to the expected *N*/*S* ratio under neutral evolution were calculated from Fisher's exact tests. "All sites" represents the number of sites that appear in at least one individual. The numbers of sites appearing in at least two individuals are also shown. (D) Editing levels of the nonsynonymous and synonymous sites in the head. Sites in four drones were combined to increase the statistical power. The Wilcoxon rank sum test was performed to detect statistical significance (***, $P < 0.001$).

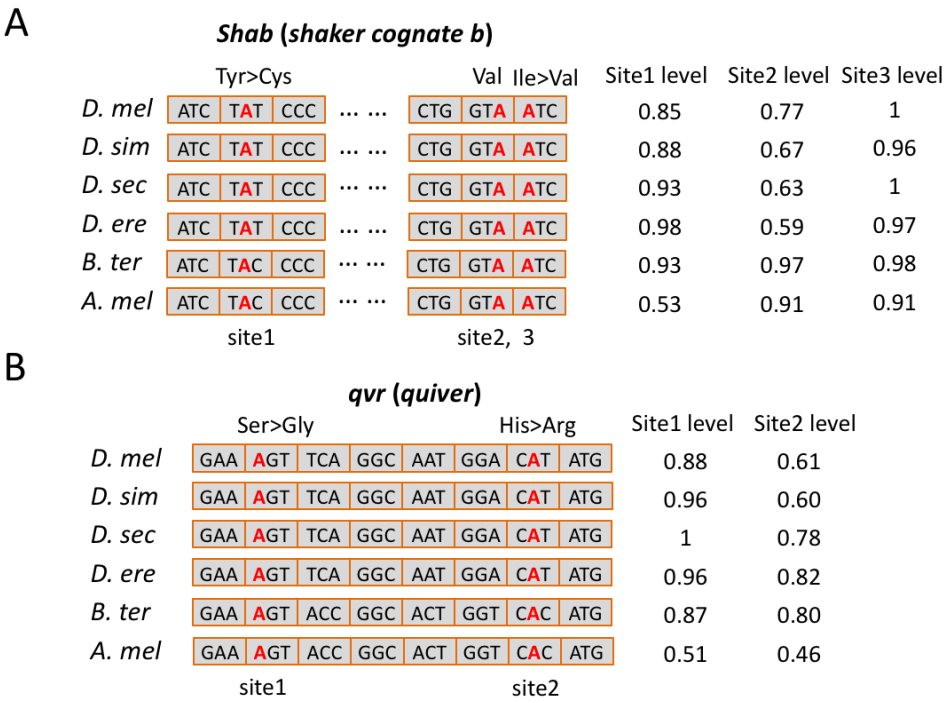


Figure 3. Conserved editing sites between bees and flies. (A) In gene *Shab* (*shaker cognate b*), one synonymous and two nonsynonymous editing sites are conserved in heads of bees and four *Drosophila* species. (B) In gene *qvr* (*quiver*), two nonsynonymous editing sites are conserved in heads of bees and four *Drosophila* species. The editing sites are colored in red. The editing levels from male heads of flies (unpublished), head of honeybee individual-1 and brains of bumblebee(*82*) 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*.

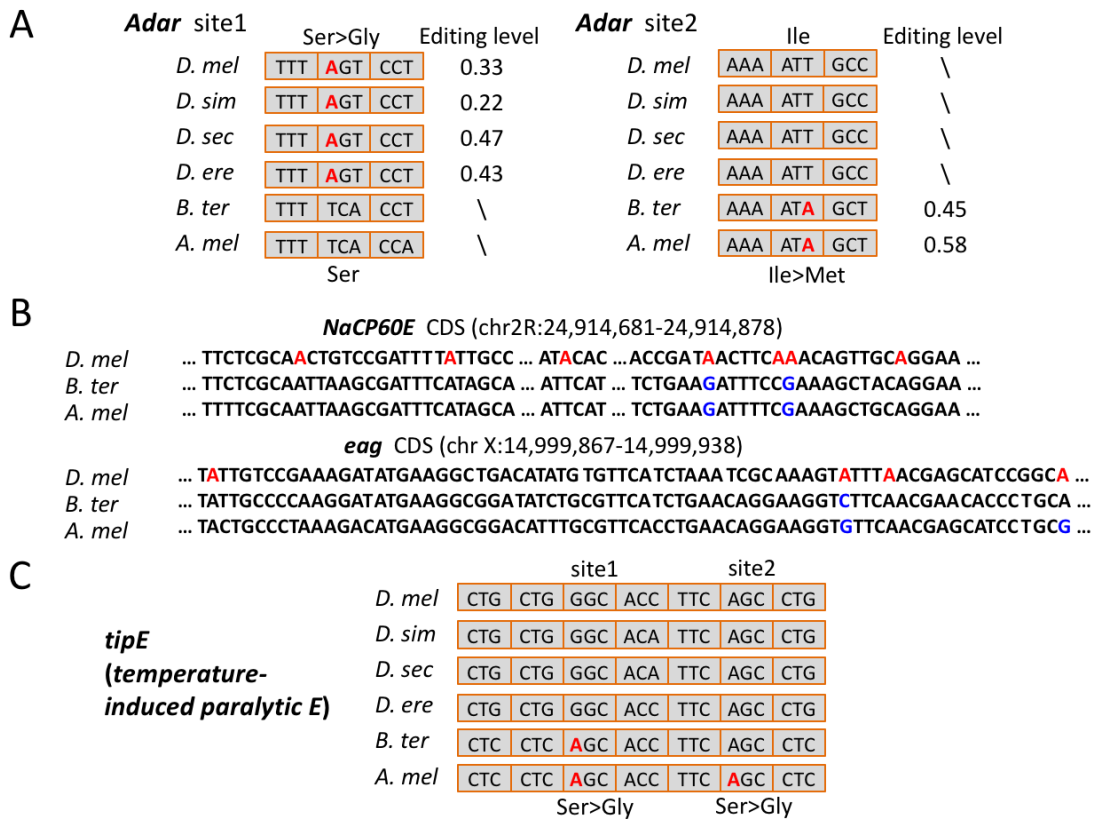
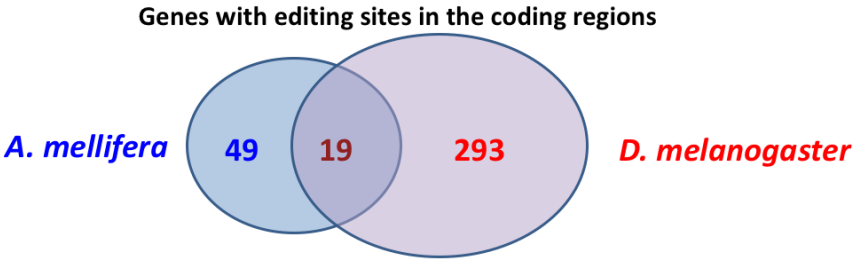


Figure 4. Non-conserved editing sites between bees and flies. (A) In gene *Adar*, the auto-editing site in *Drosophila* leads to a Ser>Gly change, which is highly conserved across fly species. The editable Ser codon AGT in flies is changed to an un-editable Ser codon TCA in bees. Another auto-editing site in bees leads to a Ile>Met change and it is conserved in honeybee and bumblebee. The editable Ile codon ATA (the third position) in bees is changed to an un-editable Ile codon ATT in flies. The editing sites are colored in red. The editing levels from male heads of flies (unpublished), head of honeybee individual-1 and brains of bumblebee(*82*) 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) Non-conserved editing sites in gene *NaCP60E* and *eag*. The *Drosophila* specific editing sites are colored in red. (C) Non-conserved editing sites in fly gene *tipE* (*temperature-induced paralytic E*). The editing sites in honeybee (orthologous gene *GB47375*) are colored in red. A task-related editing site (site1) in bumblebee orthologous gene *XM_003393369* was reported(*82*).

A



B

	N	*S*	*N*/*S*	*P* value
A. mellifera (expected *N*/*S* = 2.26)				
Specific	104	8	13.0	1.9E-9
Shared	25	2	12.5	0.006
D. melanogaster (expected *N*/*S* = 3.80)				
Specific	589	125	4.71	0.030
Shared	89	19	4.68	0.003

Figure 5. Evidence for convergent adaptation in fly and honeybee. (A) Venn diagram demonstrating the overlap between orthologous genes with editing sites in coding regions. (B) The numbers of nonsynonymous (*N*) and synonymous (*S*) editing sites that are shared between fly and honeybee or species-specific. The *P* values of the observed *N*/*S* ratios compared to the expected *N*/*S* ratio under neutral evolution were calculated from Fisher's exact tests.

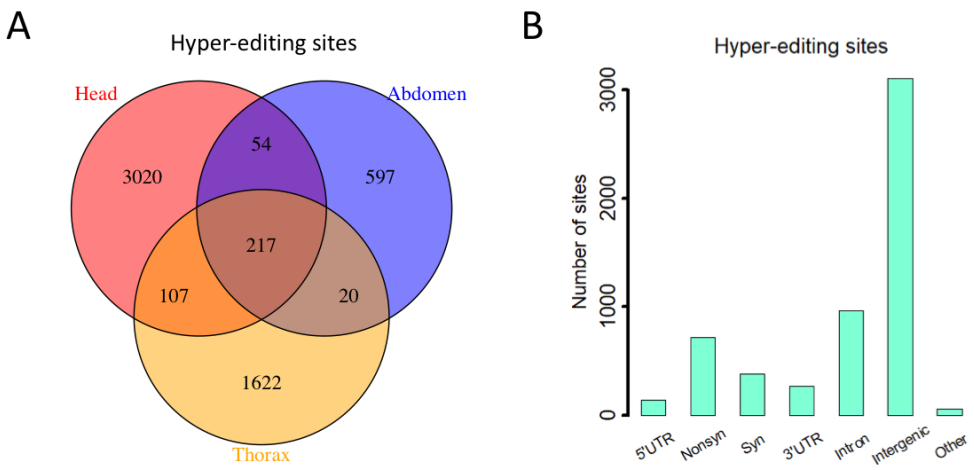


Figure 6. Identification and annotation of the hyper-editing sites in honeybee. (A) Venn diagram demonstrating the overlaps of hyper-editing sites in three honeybee tissues. (B) Functional annotation of hyper-editing sites.

Table 1. Number of editing sites in each sample.

Drones	Tissue	Total sites	Nonsyn (*N*)	Syn (*S*)	*N*/*S*	*P* value
	Head	314	85	9	9.4	1.6E-06
D1	Thorax	93	15	4	3.8	0.46
	Abdomen	114	17	1	17.0	0.020
	Head	294	86	7	12.3	1.0E-07
D2	Thorax	93	13	5	2.6	1.0
	Abdomen	132	18	2	9.0	0.051
	Head	202	53	6	8.8	3.2E-4
D3	Thorax	50	3	0	Inf	0.56
	Abdomen	69	0	1	0	0.31
	Head	239	69	8	8.6	3.5E-05
D4	Thorax	58	4	1	4.0	1.0
	Abdomen	73	5	1	5.0	0.67
All head		432	129	10	12.9	3.0E-11
All thorax		146	17	5	3.4	0.50
All abdomen		179	21	2	10.5	0.022
All		464	131	10	13.1	1.2E-11

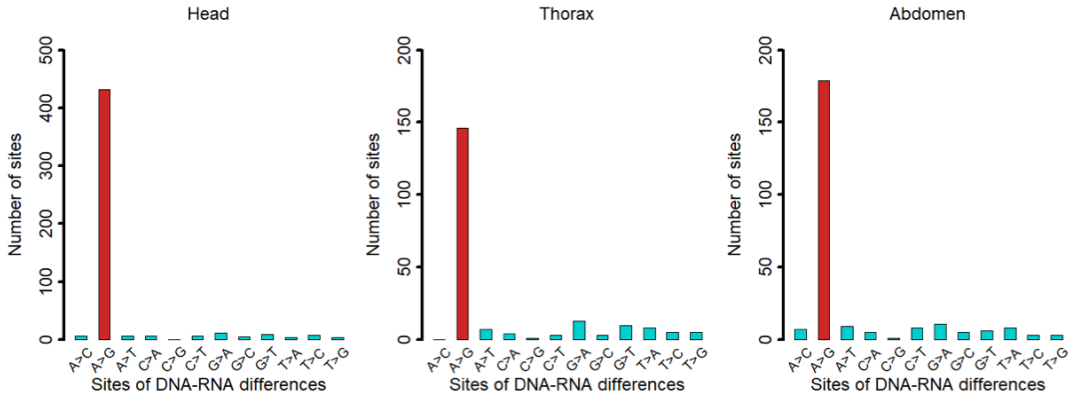
Totally 464 unique editing sites were identified. The numbers of editing sites in each sample are listed, respectively. "All head" is the number of sites that appear in at least one head sample among the four individuals. The same goes for "All thorax" and "All abdomen".
P value: The *P* value of the observed *N*/*S* ratios compared to the expected *N*/*S* ratio under neutral evolution calculated from Fisher's exact tests.

1

2 **Supplementary Material**

3 **Supplementary Figures**

4



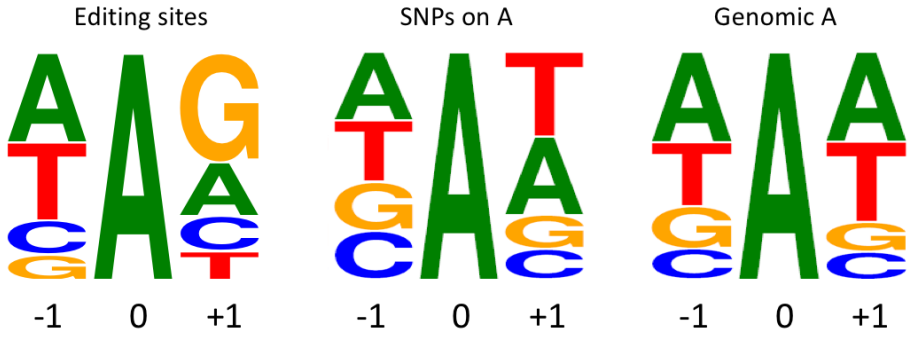
5

6 **Figure S1. Numbers of different types of DNA-RNA differences detected in heads,**

7 **thoraxes, or abdomens of honeybees.**

8

9



10

11 **Figure S2. The triplets centered by editing sites (left), SNPs (middle) or unedited**

12 **adenosines in genome (right).**

13

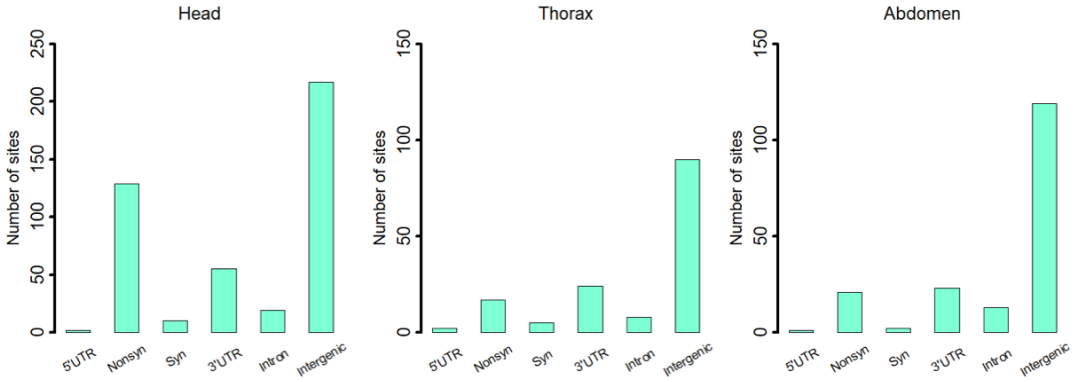


Figure S3. Functional annotation of editing sites detected in heads, thoraxes, or abdomens of honeybees.

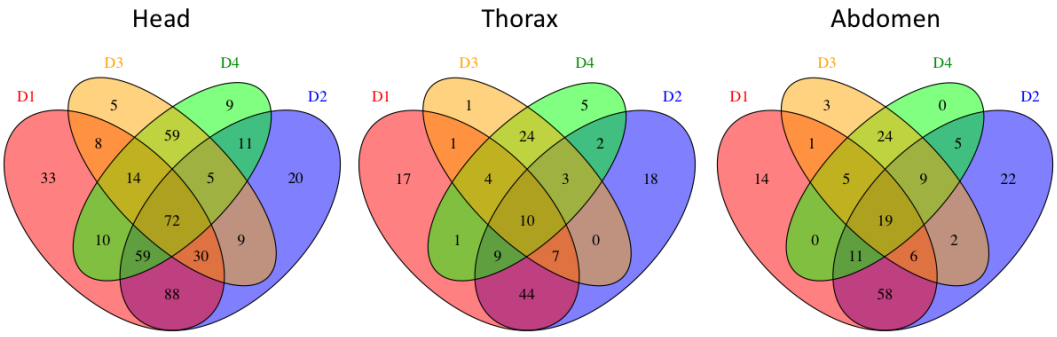


Figure S4. Venn diagram demonstrating the overlaps of editing sites in four honeybee drones.

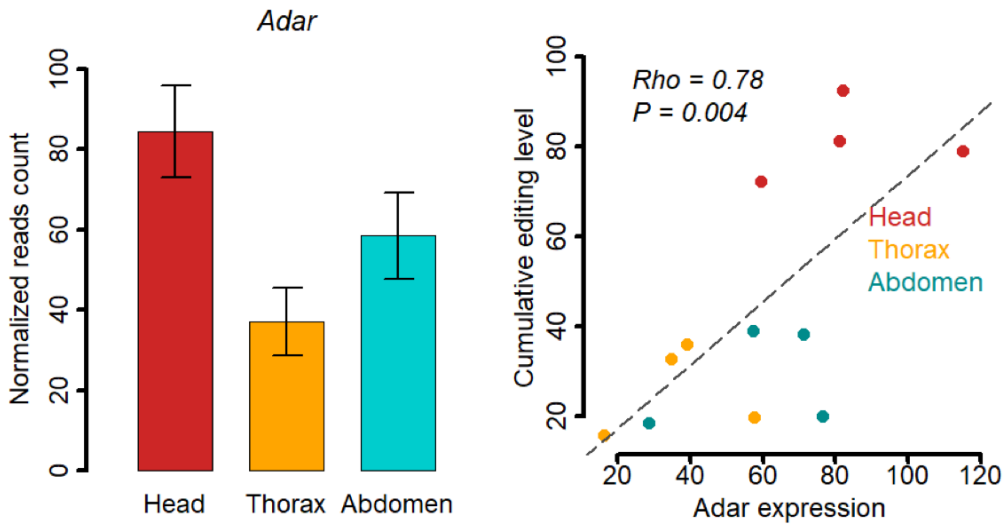


Figure S5. *Adar* expression is highest in heads (left) and is positively correlated to cumulative editing levels in each sample (right). Error bars represent standard error. Spearman's correlation is used in the correlation test.

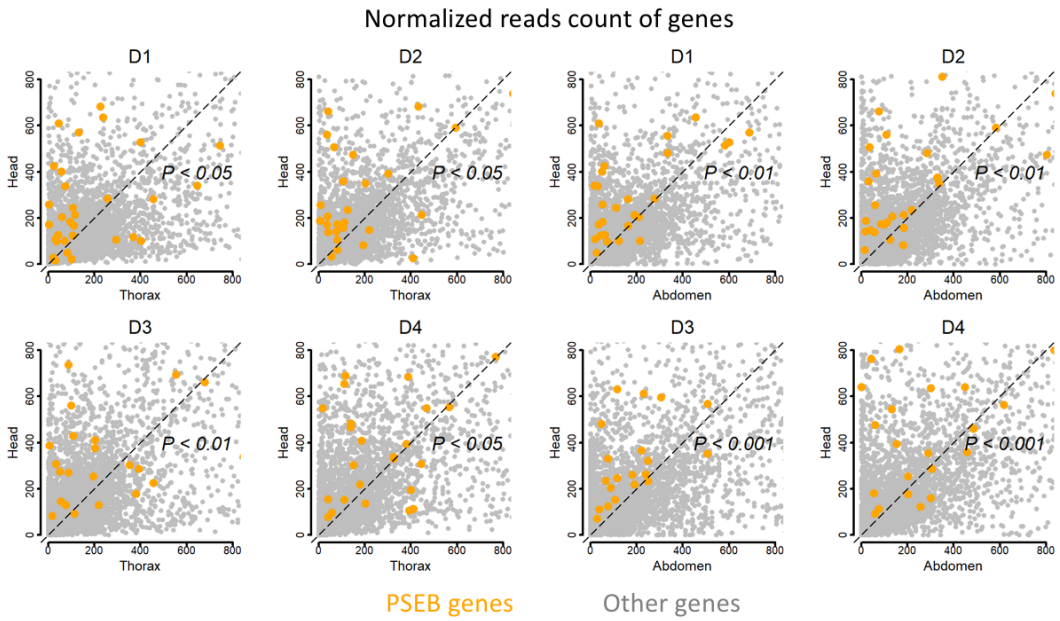


Figure S6. Comparison of the expression level of PSEB genes in the head versus thorax or abdomen. PSEB genes ($N/S > 5$) are colored in orange. Normalized reads count is used to represent the expression level. P values of the PSEB gene expression were calculated from paired Wilcoxon rank sum tests.

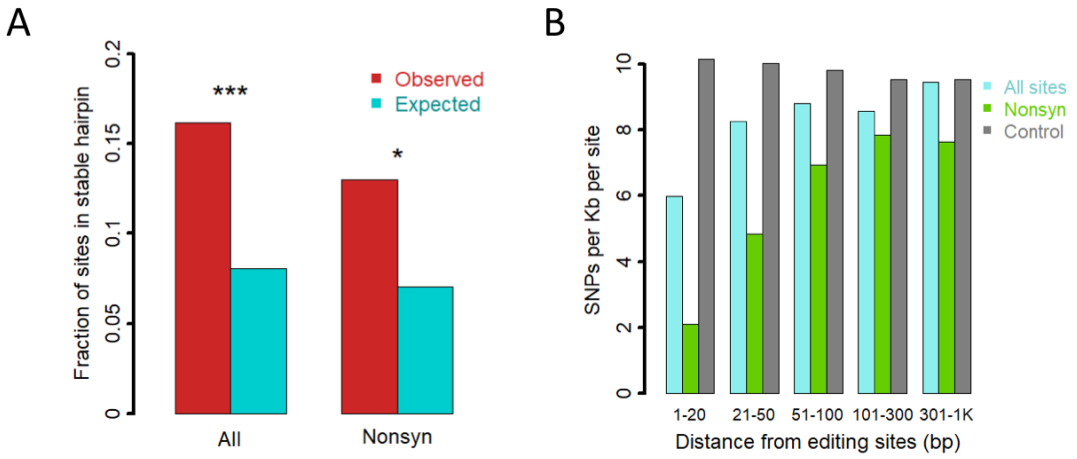


Figure S7. Tradeoff between RNA diversity and genome evolution in honeybee. (A) The fraction of sites in predicted stable hairpin structures. The observed editing sites were compared to the unedited adenosines. *P* values were calculated from Fisher's exact tests (*, $P < 0.05$; ***, $P < 0.001$). (B) The SNP density is depleted near editing sites (cyan) rather than the unedited adenosines (grey). The pattern is stronger for nonsynonymous editing sites (green).

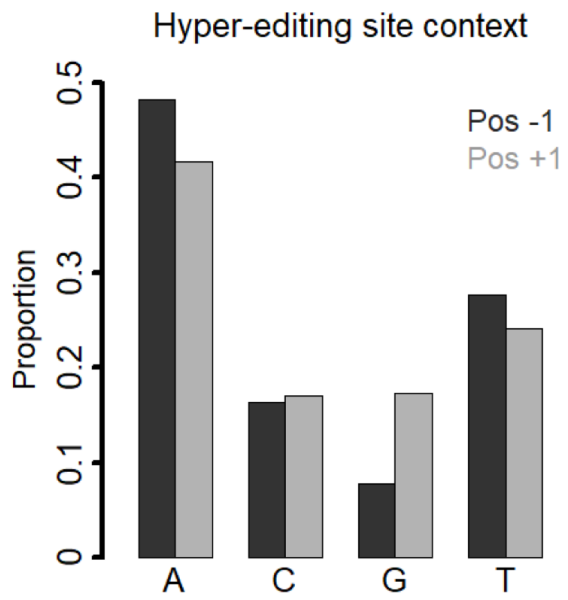


Figure S8. The nucleotide context of hyper-editing sites. Pos -1, the upstream nucleotide. Pos +1, the downstream nucleotide.

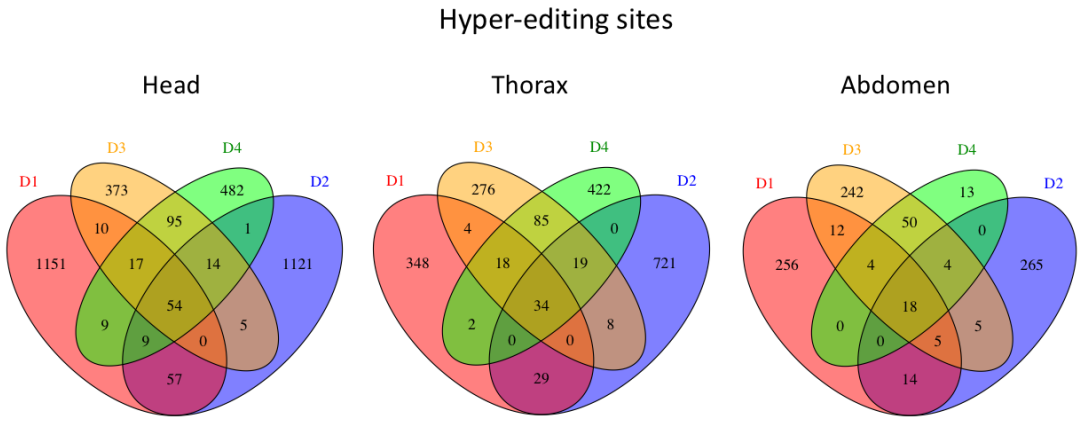


Figure S9. Venn diagram demonstrating the overlaps of hyper-editing sites in four honeybee drones.

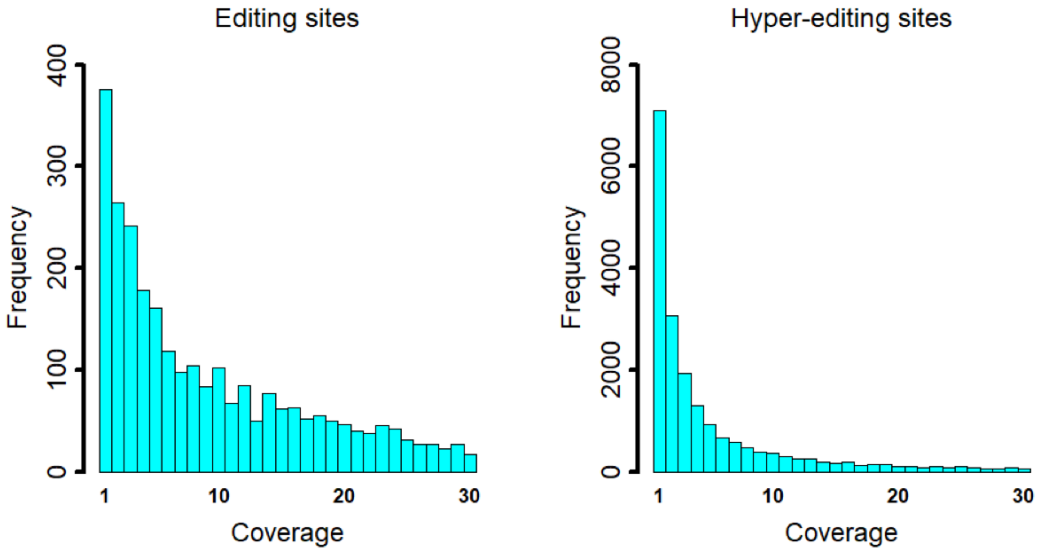


Figure S10. Histograms of sequencing coverage of normal editing sites (left) and hyper-editing sites (right). Editing sites in all samples were used.

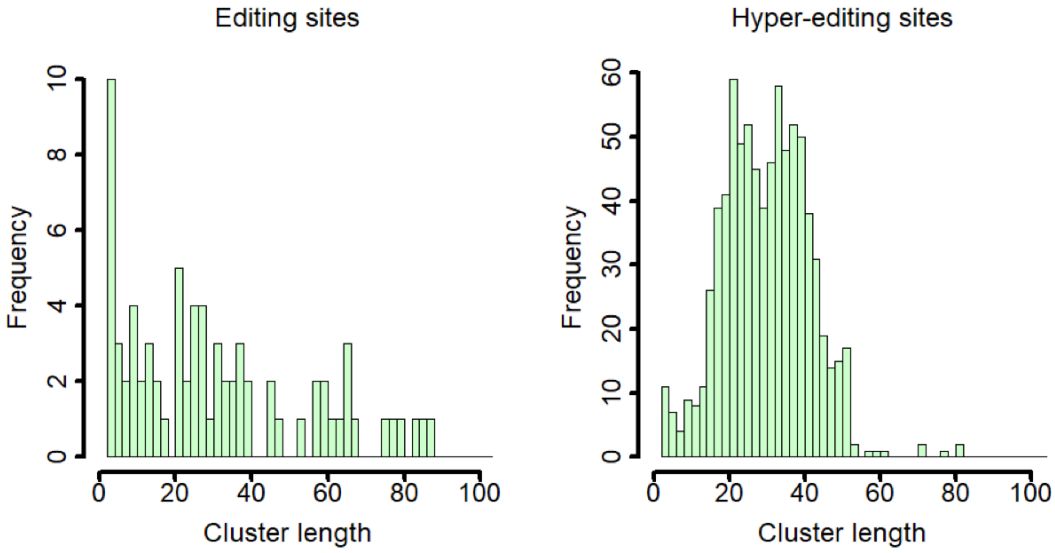


Figure S11. Histograms showing the cluster length of normal editing sites (left) and hyper-editing sites (right). Editing sites within 100bps were clustered.

Supplementary Tables

Table S1. Number of SNPs identified in each individual honeybee.

	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 sample.

Drones	Tissue	Total reads (M)	Uniquely mapped reads (M)	Uniquely mapped reads (%)
D1	Head	25.0	17.9	71.6
	Thorax	25.8	16.8	65.1
	Abdomen	23.5	17.2	73.2
D2	Head	18.8	14.2	75.3
	Thorax	30.8	18.4	59.6
	Abdomen	28.5	21.5	75.3
D3	Head	37.2	12.4	33.4
	Thorax	32.9	10.2	31.1
	Abdomen	43.1	12.8	29.8
D4	Head	48.4	15.2	31.4
	Thorax	31.9	10.4	32.6
	Abdomen	32.9	9.2	28.1

Table S3. Number of hyper-editing sites in each sample.

Drones	Tissue	Total sites	5'UTR	Nonsyn	Syn	3'UTR	Intron	Intergenic	other
D1	Head	1307	18	163	103	67	201	746	9
	Thorax	435	15	62	22	29	53	249	5
	Abdomen	309	3	28	10	54	63	149	2
D2	Head	1261	27	153	92	34	278	668	9
	Thorax	811	54	121	64	40	137	380	15
	Abdomen	311	1	58	27	3	53	167	2
D3	Head	568	14	41	22	22	112	350	7
	Thorax	444	1	38	16	0	96	291	2
	Abdomen	340	0	10	8	0	54	268	0
D4	Head	681	4	76	33	26	136	405	1
	Thorax	580	2	35	14	5	98	421	5
	Abdomen	89	0	1	0	0	29	59	0
Pool		5637	139	718	379	272	969	3103	57