BMC Genomics

RESEARCH

Open Access

Adaptive evolution of A-to-I auto-editing site in *Adar* of eusocial insects

Caiqing Zheng[1†], Jiyao Liu[1†] and Yuange Duan[1*]

Abstract

Background Adenosine-to-inosine (A-to-I) RNA editing is a co-/post-transcriptional modification introducing A-to-G variations in RNAs. There is extensive discussion on whether the flexibility of RNA editing exerts a proteomic diversification role, or it just acts like hardwired mutations to correct the genomic allele. Eusocial insects evolved the ability to generate phenotypically differentiated individuals with the same genome, indicating the involvement of epigenetic/transcriptomic regulation.

Methods We obtained the genomes of 104 Hymenoptera insects and the transcriptomes of representative species. Comparative genomic analysis was performed to parse the evolutionary trajectory of a regulatory Ile > Met auto-recoding site in *Adar* gene.

Results At genome level, the pre-editing Ile codon is conserved across a node containing all eusocial hymenopterans. At RNA level, the editing events are confirmed in representative species and shows considerable condition-specificity. Compared to random expectation, the editable Ile codon avoids genomic substitutions to Met or to uneditable Ile codons, but does not avoid mutations to other unrelated amino acids.

Conclusions The flexibility of *Adar* auto-recoding site in Hymenoptera is selectively maintained, supporting the flexible RNA editing hypothesis. We proposed a new angle to view the adaptation of RNA editing, providing another layer to explain the great phenotypical plasticity of eusocial insects.

Keywords A-to-I RNA editing, *Adar*, Auto-recoding, Hymenoptera, Adaptive

Background

A-to-I RNA editing in metazoans

The ADAR (adenosine deaminase acting on RNA) enzymes in metazoans mediate the prevalent adenosine-to-inosine (A-to-I) RNA editing [1]. Numerous adenosines in the transcripts are potential targets of ADARs which preferentially bind to a particular sequence context in the double-stranded RNA (dsRNA) structures [2, 3]. Thousands to millions of editing sites were identified in different animal species but the genomic distributions of such sites are lineage-specific [4, 5]. While the mammalian RNA editing sites mainly occur in non-coding regions and repetitive elements [6–9], insects have a large fraction of editing sites in coding sequence (CDS) [10–15]. Since inosines will be recognized as guanosines by cellular machineries [16, 17], A-to-I RNA editing resembles A-to-G mutations and can change the amino acid (AA) encoded by the DNA. Therefore, nonsynonymous RNA editing events are also termed "recoding" sites and largely diversify the proteome [18, 19]. There has been a

[†]Caiqing Zheng and Jiyao Liu co-first authors.

*Correspondence:
Yuange Duan
duanyuange@cau.edu.cn
[1]Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing 100193, China

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

long-term discussion on whether the RNA recoding sites are designed to facilitate adaptive evolution [20, 21], or alternatively, the recoding events are just byproducts of promiscuous RNA editing [22, 23].

Signals of adaptation of A-to-I RNA recoding

The original evidence for adaptive RNA recoding was found by observing overrepresented numbers of recoding sites compared to the random expectation. Also, the editing levels of recoding sites were higher than the neutral synonymous sites [12, 24, 25]. These suggest the positive selection on recoding sites, leaving us an impression that RNA recoding drives adaptive evolution due to the flexible regulation because recoding sites can increase proteomic diversity in a controllable manner. However, researchers realized that positive selection on recoding sites does not fully stand for the advantage of flexibility. If G-allele is constantly better than A-allele, then A-to-I RNA editing will also be maintained, but it just makes an incomplete compensation to rescue the genomic A-allele rather that to diversify the proteome. Under the "G-better hypothesis", one could envision that the edited adenosines should prefer to be directly replaced with a genomic G during evolution (Fig. 1A). This scenario is different from our common understanding of the post-transcriptional regulatory role of RNA editing.

In contrast, if the advantage of RNA editing is the flexibility of the temporal-spatial regulation, then the editing sites should avoid genomic mutations because the alteration of genome sequence would permanently abolish the flexibility of RNA editing (Fig. 1B). Therefore, the stringent definition of "adaptive recoding" is the narrow-sense one, which only refers to the advantage of "flexible RNA editing" over the hardwired DNA mutations. Conceivably, the essence of flexibility is to adjust the relative proportions of A- and G-alleles according to which version is fitter under a particular condition [26–28]. In other words, the flexible RNA editing hypothesis predicts that editing levels should be variable and condition-dependent.

Different from the comparative genomics analysis to prove the overall avoidance of mutations on editing sites, the condition-specificity of RNA editing can be tested on individual sites, but this is not the most challenging part

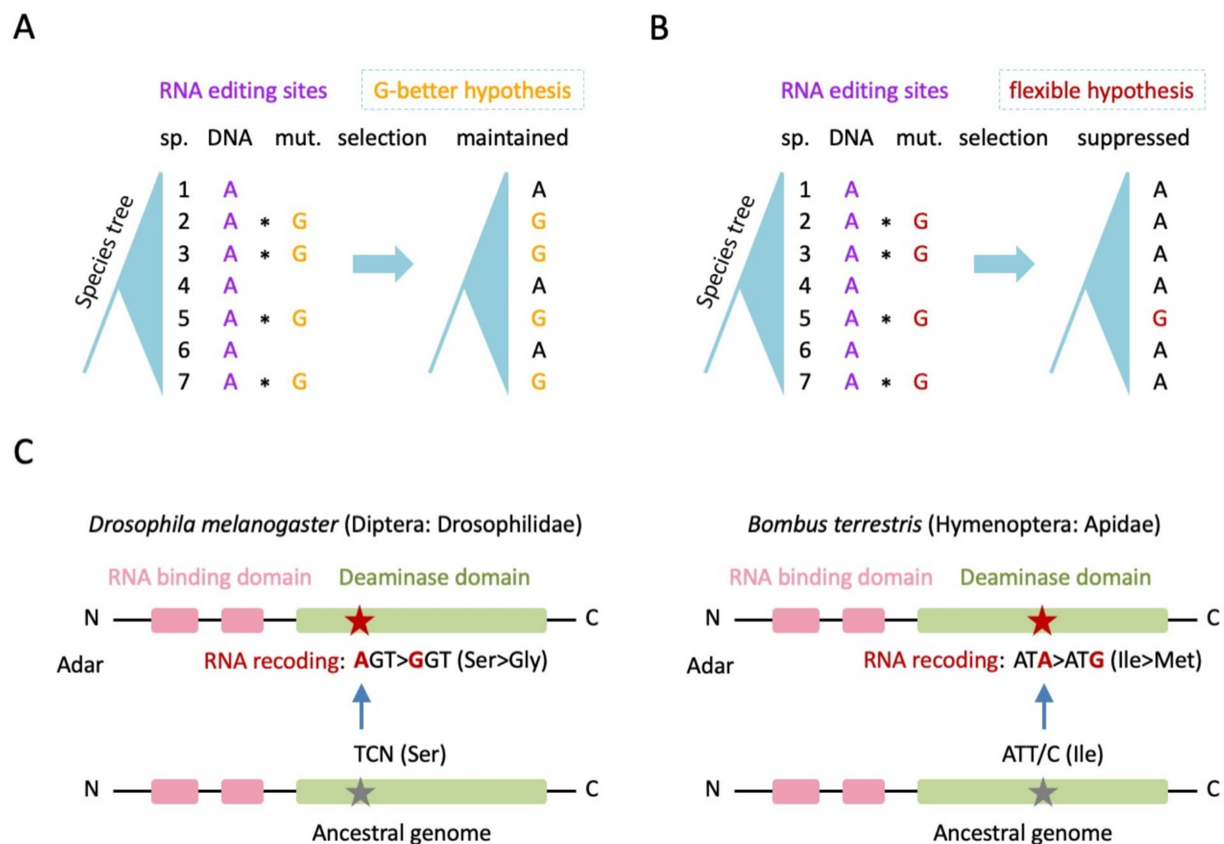

Fig. 1 Comparative genomic evidence for adaptive A-to-I RNA recoding. (**A**) The G-better hypothesis predicts that the genomic A-to-G substitution at RNA editing sites should be beneficial and be maintained by natural selection. (**B**) The flexible RNA editing hypothesis predicts that genomic mutations at RNA editing sites should be suppressed by natural selection. (**C**) Auto-recoding sites in *Adar* gene. The domains of Adar protein are displayed. The current recoding sites and the ancestral genome sequences are shown for two different editing sites in *D. melanogaster* and *B. terrestris*. Abbreviations: sp., species; mut., mutation

in proving the flexible RNA editing hypothesis. The best approach to demonstrate the advantage of editing is to construct the fully-edited and uneditable mutant organisms and measure their fitness changes. This idea has already been accomplished in fungi [27]. However, since the genetic manipulation and fitness assays are unfeasible for most animal species, in this article, we only discuss the potential bioinformatic and comparative genomic methodologies to judge the adaptation of RNA editing events. Moreover, to compare the advantage of RNA editing over an uneditable state, question comes that how to define an uneditable allele given that an adenosine always has the potential to be edited? The solution will require the degeneracy of synonymous codons.

Adar auto-recoding sites in insects enable the comparison between editable and uneditable states

Insects experienced an ancestral loss of *ADAR1* and only had one copy of *Adar* gene which is orthologous to mammalian *ADAR2* [29]. Insect Adar has two dsRNA-binding domains at the N-terminal and one deaminase domain at the C-terminal [30]. In *Drosophila melanogaster* (Diptera: Drosophilidae), *Adar* mRNA has an auto-recoding site in the deaminase domain that changes a serine residue (*A*GC/T) to a glycine residue (*G*GC/T) [31]. This recoding site is named the Ser>Gly or S>G site (Fig. 1C). The post-edited $Adar^G$ isoform has lower catalytic activity than the unedited $Adar^S$ isoform, making a negative feedback loop that stabilize the global RNA editing efficiency [31]. A few years after the discovery of *Drosophila Adar* Ser>Gly site, another Ile>Met (AT*A*>AT*G*) auto-recoding site in *Adar* was found in eusocial insect bumblebee *Bombus terrestris* (Hymenoptera: Apidae) (Fig. 1C), and this site again seemed to affect the overall editing efficiency [32]. The two auto-recoding sites at different positions in *Adar* CDS were independently gained in fly and bee. Given these observations, it is very likely that the functional auto-recoding site in *Adar* exerts a flexible regulatory role.

Interestingly, comparative genomics analyses revealed that the *Adar* Ser>Gly (*A*GC/T>*G*GC/T) auto-recoding site in Diptera had an uneditable Ser codon (*T*CN) in the ancestral genome of all insects; accordingly, the Ile>Met (AT*A*>AT*G*) auto-recoding site in Hymenoptera had an uneditable Ile codon (AT*C*/*T*, denoted as "unIle") in the ancestral node [33] (Fig. 1C). This provides an exciting evidence supporting the flexible regulatory role of the auto-recoding sites in *Adar* [33]. The uneditable-to-editable codon transition indicates the evolutionary advantage of having the editing potential over the hardwired pre-editing allele. But for the majority of recoding sites in the genome, an uneditable codon encoding the pre-editing AA is not available, preventing the manifestation of the advantage of RNA editing using evolutionary genomic approach. Thus, the *Adar* auto-recoding sites in insects is a very precious case for studying adaptive RNA editing.

Eusocial insects, RNA editing, and transcriptomic plasticity

While the Ser>Gly recoding site in *Adar* was sufficiently studied in Diptera [30, 31, 33], the Ile>Met recoding site in Hymenoptera is still poorly understood [32]. Given that (1) many eusocial insects in Hymenoptera have evolved the ability to generate high phenotypic plasticity with the same set of genome [34, 35], and that (2) the transcriptomic plasticity introduced by A-to-I RNA editing is well-suited to shape the proteome, phenome, and fitness of organisms, there comes a tempting possibility that RNA editing facilitates the phenotypic regulation of a broad range of eusocial hymenopterans. And the auto-regulatory site in Adar might act as the key to the accurate control of the overall cellular RNA editing events.

Regarding this Ile>Met auto-recoding site in Hymenoptera *Adar* gene, several questions remain to be answered: (1) What is the evolutionary trajectory of the genome sequence of this Ile codon in eusocial hymenopterans? (2) Can this Ile>Met RNA editing event be detected in other eusocial hymenopterans apart from bumblebee? (3) Can we observe condition-specificity of editing levels for this Ile>Met recoding site? (4) If the editable Ile codon has evolved from an ancestral uneditable Ile codon, how unexpected was this transition compared to random control? Is there a suppression on genomic mutations on this current recoding site? (5) If the uneditable-to-editable transition was adaptive, then what overall signals should we observe in the current genome and editome?

In this study, we utilized the high-quality genomes and transcriptomes of Hymenoptera species to answer the above-mentioned questions. The Ile>Met auto-recoding event in *Adar* was identified in multiple eusocial hymenopterans like bumblebee, honeybee, and leaf-cutting ant. The Ile>Met editing levels showed remarkable condition-specificity in tested species like bees and ants. The ancestral state of this auto-recoding site appeared to be an uneditable Ile codon, suggesting the evolutionary advantage of having the editing potential. In the Hymenoptera phylogeny, the A-to-G and editable-to-uneditable transitions were significantly suppressed at the Ile>Met site, suggesting the strong maintenance on the "editability", but the mutations to other AAs were not suppressed. Interestingly, termites (Blattodea) do not have this editable Ile codon, suggesting that distantly-related eusocial insects might utilize different molecular strategies to achieve the flexibility of eusociality. In the light of comparative and evolutionary genomics, our work developed a new angle to view and test the adaptation of global or individual RNA recoding

sites. We revealed adaptive RNA editing sites in eusocial hymenopterans, providing another layer to account for the great phenotypical plasticity of eusocial insects.

Results

Conservation of *Adar* Ile > Met auto-recoding site in eusocial insects of Hymenoptera

To obtain a clear evolutionary trajectory of Ile>Met auto-recoding site in Hymenoptera, we searched the public database NCBI for the high-quality genome sequences of Hymenoptera species. We selected 104 well-annotated Hymenoptera genomes and constructed the species tree using orthologous genes (Materials and Methods). By aligning the *Adar* CDS, we found that most of the species had an ATA codon identical to the editable Ile codon reported in bumblebee [32]. Particularly, there is a monophyletic branch with all species having ATA codon, and we defined its ancestral node as node1 (Fig. 2). This node includes 57 species and contains all the best-known eusocial insects in Hymenoptera such as bumblebee, honeybee *Apis mellifera* (Hymenoptera: Apidae), and leaf-cutting ant *Acromyrmex echinatior* (Hymenoptera: Formicidae). Totally, 42 eusocial insects are included in node1 (Fig. 2).

To examine whether the RNA editing event in bumblebee is conserved across node1, we downloaded publicly available transcriptomes and confirmed that this AT*A* codon is indeed edited to AT*G* in other eusocial hymenopterans like honeybee and leaf-cutting ant (Fig. 2). Note that bees (Apidae) and leaf-cutting ants (Formicidae), which have diverged for 160 Mya [36], belonged to the two earliest diverging clades in node1. This suggests that the Ile>Met recoding event was likely to already exist at node1 and might be conserved across all the 57 extant species in node1 (Fig. 2). However, we reserve the possibility that the honeybee and ant editing events might be independently gained during evolution, and thus a wider range of transcriptome data are needed to confirm that the Ile>Met recoding event is indeed conserved across species in node1.

Remarkable condition-specificity of Ile > Met editing site in representative eusocial hymenopterans

Next, we try to find evidences to support the flexible RNA editing hypothesis of this Ile>Met recoding site in *Adar* gene. Before we resort to the comparative genomic approaches, we first dig into this Ile>Met site itself. As we have introduced, flexible RNA editing hypothesis requires that the editing levels are condition-dependent because the relative proportions of A- and G-allele should be adjusted according to which allele is fitter under a specific condition [27]. The condition-specificity, which generally includes tissue-specificity, developmental stage-specificity, environment-sensitivity, treatment-sensitivity, and caste-specificity for eusocial insects, is a necessary but insufficient requirement for adaptive RNA editing.

We investigated the transcriptomes of different tissues and castes/sub-castes of honeybee and leaf-cutting ant [38]. If the Ile>Met RNA recoding exerts a flexible role, then there might be differentiated editing levels across different conditions as reported in bumblebee [32]. Not surprisingly, in honeybees, the Ile>Met editing level was highest in heads and much lower in thoraces and abdomens, exhibiting high tissue-specificity (Fig. 2 and Additional file1: Supplementary Figure S1). Moreover, the editing levels in drone (male), forager (female, worker), and nurse (female, worker) were also remarkably different (Fig. 2). Drone heads had an editing level 2.7 times higher than the average level in worker brains and then the editing levels in foragers were also significantly higher than those in nurses (Fig. 2). In leaf-cutting ants, the recoding levels also showed remarkable differences among gynes, large workers, and small workers (Fig. 2). Apart from the six nurse and six forager samples of honeybees, the comparison between single samples without biological replicates is not applicable. However, Wilcoxon rank sum tests could be used to calculate the differential editing levels between 12 honeybee worker samples *versus* any of the remaining bee or ant samples, respectively. All of the remaining single sample showed a marginally significant editing level when compared to the 12 honeybee workers (1 *versus* 12 comparison, $P=0.07$). The pattern will become significant ($P=0.0057$) if we treat the three ant brain samples as biological replicates and perform a 3 *versus* 12 comparison between ant brains and honeybee brains. All these observations increased the possibility that the *Adar* Ile>Met auto-recoding levels in eusocial hymenopterans largely depend on different conditions like the species identity, castes/sub-castes, and tissues, supporting the notion that this site exerts its function with a flexible manner. Moreover, apart from the drone (male, haploid) sample in *A. mellifera* with high editing level in heads, the other samples displayed in Fig. 2 were all female (diploid) bee/ant samples that showed considerably lower editing levels in brains. This again raises the possible sex-specific or even allele-specific regulation of *Adar* Ile>Met recoding site. Nevertheless, the condition-specificity is only tested in two species and more data are needed to obtain a robust conclusion across the phylogenetic tree. Next, as we have stated, the condition-specific editing level is only a necessary but insufficient requirement for the flexible RNA editing hypothesis. To make a robust *in silico* inference, we need comparative genomic evidence to show the evolutionary constraint on this site that maintains its "editability".

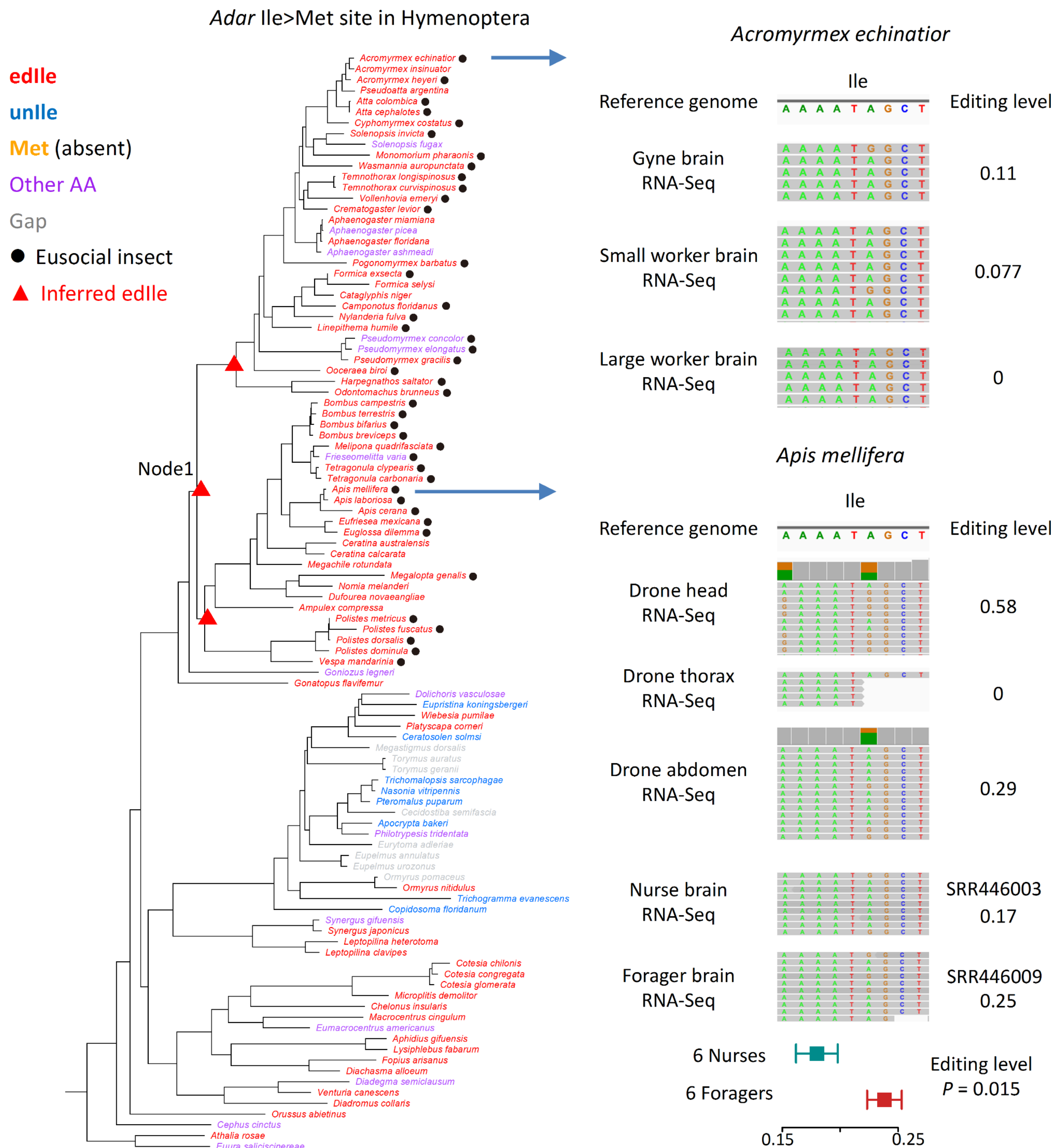

Fig. 2 Conservation of *Adar* Ile > Met auto-recoding site in eusocial insects of Hymenoptera. Left panel: The species tree constructed from orthologous genes of 104 hymenopterans is displayed. The eusocial insects are labeled with a dot after the Latin name. The genomic codon at Ile > Met site in each species is indicated by different colors: editable isoleucine (edIle, red), uneditable isoleucine (unIle, blue), methionine (orange), other AAs (purple), and gap (grey). Note that we aim to show that none of the species mutated to Met and thus we keep Met in the color caption and indicate that it is absent in the tree. The inferred codons in crucial nodes were labeled by colored triangles. Right panel: The editing levels of Ile > Met site in representative samples of eusocial species *Acromyrmex echinatior* and *Apis mellifera* are visualized by IGV. Screenshots were shown. Non-representative samples of nurses and foragers were displayed in Additional file1: Supplementary Figure S1. The differential editing levels between six nurses and six foragers (mean ± S.E.) was measured by Wilcoxon rank sum test. Note that in drone heads of *Apis mellifera*, an AAA > GAA (Lys > Glu) RNA editing site shows strong linkage disequilibrium [37] with the nearby Ile > Met recoding site. The linkage will not be discussed in this study

The general conservation level of RNA editing sites does not directly imply the advantage of flexible editing

As previously reported in other non-social insects like *Drosophila*, the edited adenosines are more conserved than unedited adenosines at the nonsynonymous positions, meaning that RNA recoding sites tend to avoid genomic mutations during evolution [12]. Similar patterns were seen in cephalopods [21, 24]. These observations seem to support the advantage of flexible RNA editing. However, the general conservation level of recoding sites does not tell us which kind of mutation was avoided. In fact, the flexible RNA editing hypothesis only states that an editable A-allele is better than G-allele or an uneditable A-allele (and thus the genomic A-to-G mutations should be avoided), but does not predict that A-allele is better than other alleles/AAs (Fig. 3A). The fitness of the pre-editing AA and post-edited AA is condition-specific, and the flexible RNA recoding can switch between A- and G-alleles so that having editing potential is conceivably better than having hardwired A- or G-allele alone [27]. But the relative fitness of other alleles, which largely relies on the function of the proteins with the particular AA, is unpredictable (Fig. 3A). Therefore, merely measuring the conservation level of editing sites is not informative enough. Under the flexible RNA editing hypothesis, one may expect that: compared to the many unedited adenosines in the genome (as a control), the edited adenosines should avoid genomic A-to-G mutations or the editable-to-uneditable transition, but should not avoid the mutations leading to other AAs (Fig. 3A).

Adar Ile > Met site in eusocial hymenopterans is evolutionarily constrained

Next, we will test the flexible RNA editing hypothesis on the *Adar* Ile>Met (AT*A*>AT*G*) auto-recoding site in Hymenoptera. Most of the hymenopterans, especially the species in node1, have an ATA codon at the orthologous

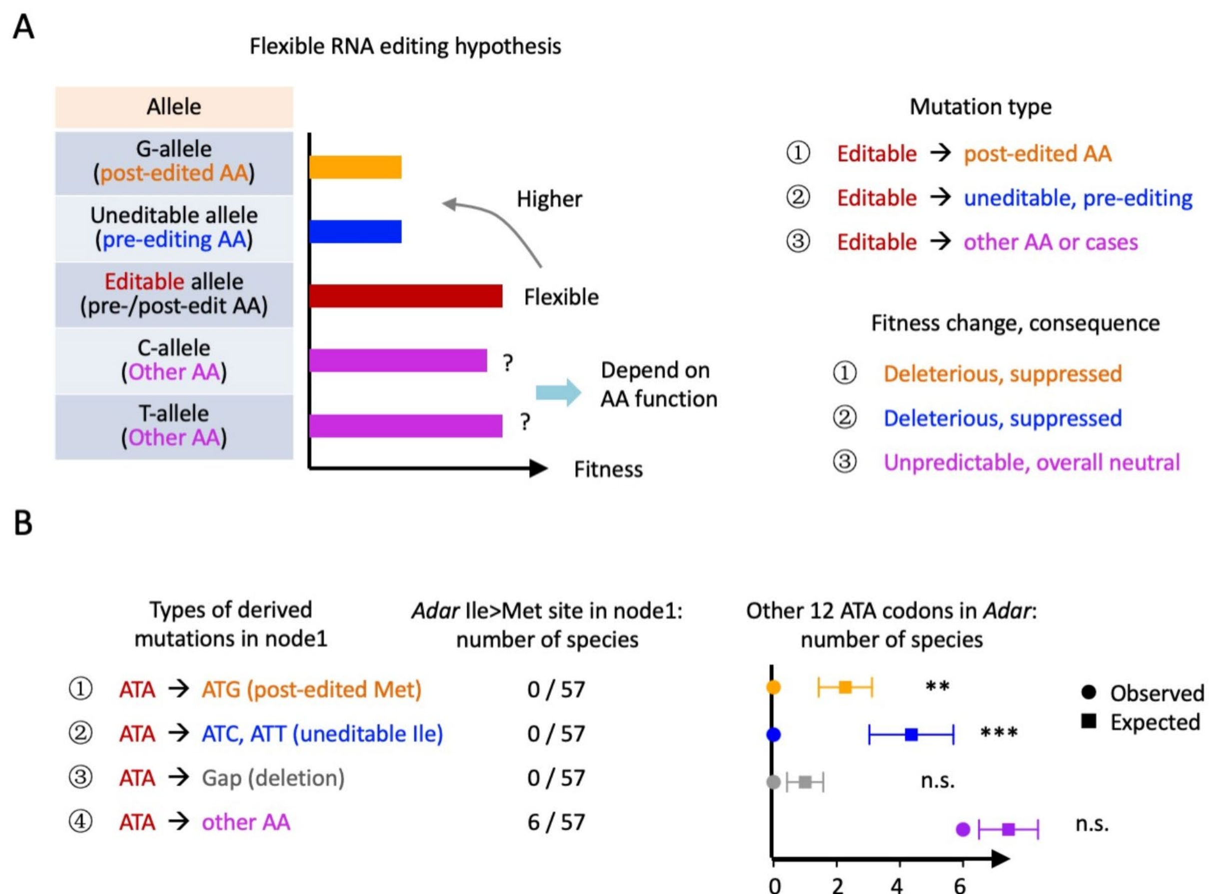

Fig. 3 Evolutionary constraint on *Adar* Ile > Met recoding site in node1 of hymenopterans. (**A**) Schematic diagram showing the fitness of different alleles, and the effects of several types of mutations under the flexible RNA editing hypothesis. The editable allele is conceivably fitter than the pre-editing AA or post-edited AA alone, but the fitness of other AA is unpredictable. (**B**) The observed and expected occurrences of different types of derived mutations on Ile > Met site in the species of node1. The other 12 unedited ATA codons in *Adar* CDS were used as control (neutral expectation). Only the ATA > ATG and the ATA > unIle mutations rather than other mutations were suppressed. The mean ± standard error (S.E.) of each group was shown. **, observed < mean – 2×S.E.; ***, observed < mean – 3×S.E

position (Fig. 2). The other non-ATA codons in the phylogeny include the post-edited codon ATG (encoding Met), the uneditable Ile codons (ATC and ATT), codons encoding other AAs, or gaps (Fig. 2). According to the different types of mutations classified in Fig. 3A, here we defined the following types of mutations on the ATA codon: (1) ATA > ATG; (2) ATA > ATC or ATA > ATT; (3) ATA to other codons not encoding Ile, namely ATA > other AAs; (4) deletion of ATA, namely ATA > gap (Fig. 3B).

From the phylogeny of node1, we discovered that none of the 57 species had this editing ATA codon (Ile) mutated to hardwired ATG codon (Met) or uneditable Ile codon (ATC and ATT) (Figs. 2 and 3B). The observed transition rates were 0 / 57 among the species. In contrast, there are 12 additional unedited ATA codons in the *Adar* CDS serving as control groups to calculate the random expectation (Fig. 3B). By focusing on ATA > ATG and ATA > unIle mutations, we found that the expected numbers of species with such codon substitutions in node1 were significantly higher than the observed zero occurrence (Fig. 3B). This means that the Ile > Met editing site strongly avoids the genomic mutations to the post-edited AA or the hardwired pre-editing status, and thus indicating the need for being editable at this Ile > Met codon position. However, for the mutations leading to other unrelated AAs or gaps, the differences between observed and expected numbers of species were insignificant (Fig. 3B). This comparison cancels the bias caused by the background conservation level of different positions and therefore the results suggest that the avoidance of ATA > ATG or ATA > uneditable mutations at Ile > Met recoding site is not simply due to the background noises.

Here we give two examples of high substitution rates for unedited ATA codons in *Adar* CDS (Fig. 4). Codon #106 (according to the CDS coordinate annotated in *A. mellifera*) is an ATA codon and it has considerable instances of ATA > ATG and ATA > unIle transitions among the species in node1 (including a few eusocial insects). Another unedited ATA codon #336 has as many as 12 ATA > unIle transitions among the species in node1, including multiple eusocial insects (Fig. 4). These cases demonstrate a considerable evolvability of the *Adar* CDS in Hymenoptera, but for the Ile > Met recoding sites, the ATA > ATG and ATA > unIle substitutions were particularly suppressed in the genomes. The results support the flexible RNA editing hypothesis which stresses that an editable allele is fitter than hardwired G-allele or uneditable allele.

Baseline A > G substitution rate in *adar* CDS in Hymenoptera

Since the ATA > ATG mutation takes place at the 3rd codon position, we then looked at the nonsynonymous A > G mutations at the 1st and 2nd codon positions as well as the synonymous A > G mutations at the 3rd codon position to get a clearer profile of the baseline substitution rate in the phylogeny (Fig. 5). The best annotated species in Hymenoptera is *A. mellifera*, which has an *Adar* CDS of 664 codons and thus 663 AAs. Among them, 431 codons contain adenosines and totally 631 adenosines are there (one codon might contain more than one adenosine). Apart from the adenosine at the 3rd position of the Ile > Met auto-editing codon (codon #482), there are 630 unedited adenosines in *Adar* CDS (Fig. 5A). Presume that an adenosine is mutated to a guanosine, the 630 adenosines will have 212 nonsynonymous mutations at the 1st codon position, 215 nonsynonymous mutations at the 2nd codon position, 12 nonsynonymous mutations at the 3rd codon position, and 191 synonymous mutations at the 3rd codon position (Fig. 5A). The 12 nonsynonymous sites at the 3rd codon positions are all AT*A* > AT*G* mutations which have been discussed above.

Interestingly, compared to the mutations in the 1st and 2nd codon positions, the 3rd codon position had a remarkably higher A > G substitution rate in the node1 species (Fig. 5B). Given this high baseline A > G substitution rate at the 3rd codon position, the Ile > Met RNA recoding site has robustly maintained 0 occurrence of genomic A > G mutations among the 57 species in node1, suggesting the strong suppression on mutations at this RNA editing sites. The results again highlight the flexible RNA editing hypothesis where an editable status is more advantageous than hardwired alleles.

Adar is differentially expressed in different honeybee samples and correlates with its auto-recoding level

In addition to the potential impact of Ile > Met recoding on the catalytic activity of Adar enzyme, the expression level of *Adar* itself also affects the global editing efficiency. We calculated the *Adar* expression level in different honeybee samples (Materials and Methods). It turns out that *Adar* expression varies a lot across samples (Fig. 6A). Foragers have significantly higher *Adar* expression than nurses; while in drones, head has much higher *Adar* expression than thorax and abdomen. Interestingly, *Adar* expression is significantly positively correlated with the Ile > Met auto-recoding level (Fig. 6A). These observations might suggest that this conserved Ile > Met editing in *Adar* has a potential causal correlation with differential RNA editing being observed either in *Adar* itself or transcriptome-wide. Although only one editing site in a few species is studied in this work, the correlation between Ile > Met auto-recoding level and the transcriptome-wide editing efficiency has been observed in bumblebee [32].

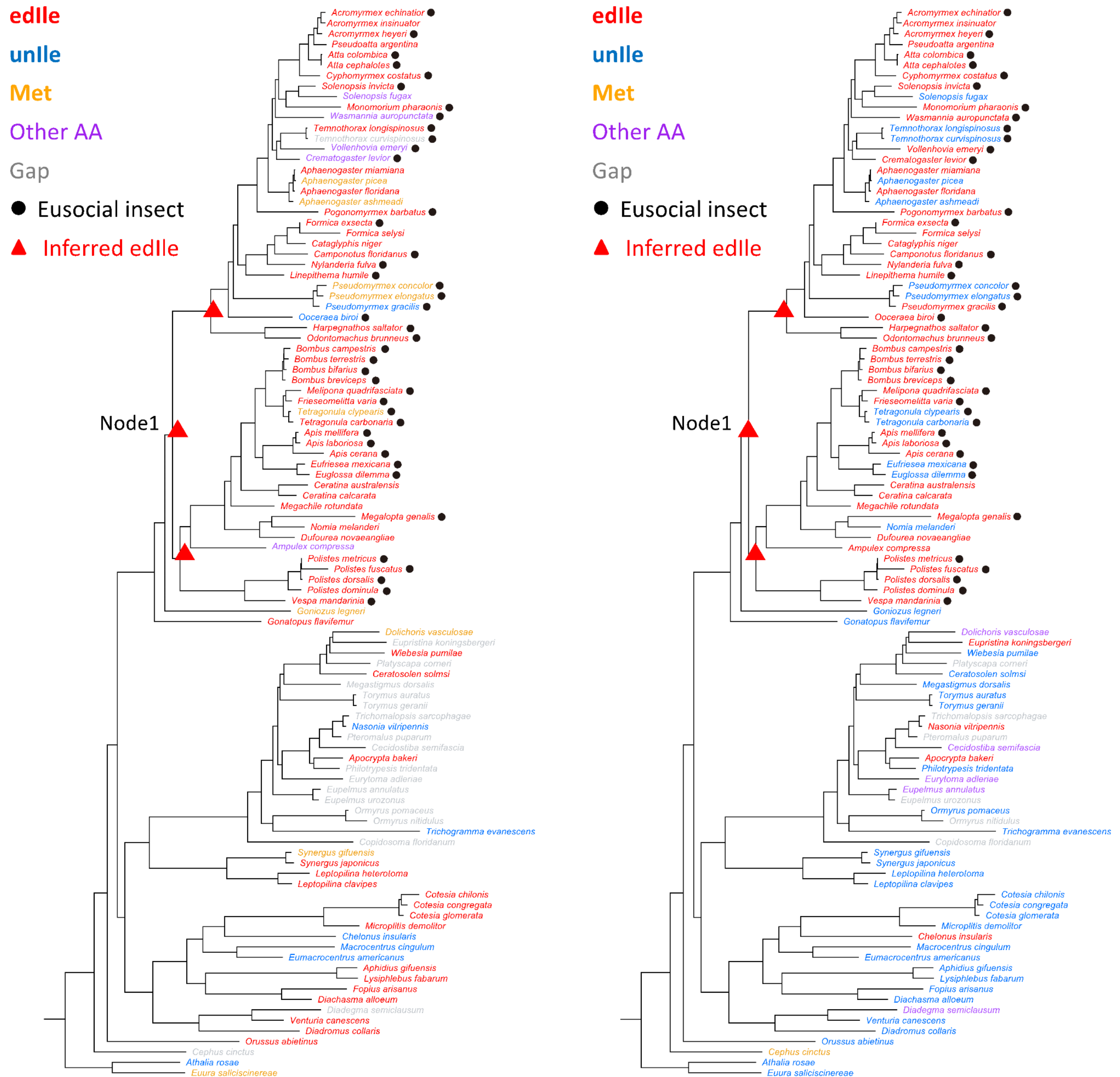

Fig. 4 Two representative unedited ATA codons in *Adar* CDS and the evolutionary trajectories in Hymenoptera. Eusocial insects are labeled with black circle. In the hymenopteran genomes, these two codons have a remarkable number of substitutions from ATA to other codons. ATA was defined as editable Ile codons (edIle, red). The other codons can be an uneditable Ile codon (unIle, ATC or ATT, blue), a codon encoding Met (orange), a codon encoding other AAs (purple), or be deleted in a few species (gap, grey). In both cases, the non-ATA codons appeared in eusocial insects that were labeled with dot after the Latin name. The inferred codons in crucial nodes were labeled by colored triangles. The most recent common ancestor of all eusocial hymenopterans converged to node1, which was inferred to have an ATA codon

Why is Ile > Met recoding site special along *adar* CDS?

As we have mentioned, apart from this Ile>Met auto-recoding site, the other 12 ATA codons in *Adar* CDS are not edited. Here comes a question that why is this auto-recoding site so special along *Adar* CDS? It is well-established that animal ADAR (Adar) enzymes preferentially target dsRNA structure [2, 3, 39]. We folded the flanking sequences of Ile>Met recoding site within *Adar* pre-mRNA (Materials and Methods) and found that the auto-recoding site is located in a highly stable hairpin structure (Fig. 6B). This potentially explains the editability of this site in eusocial insects.

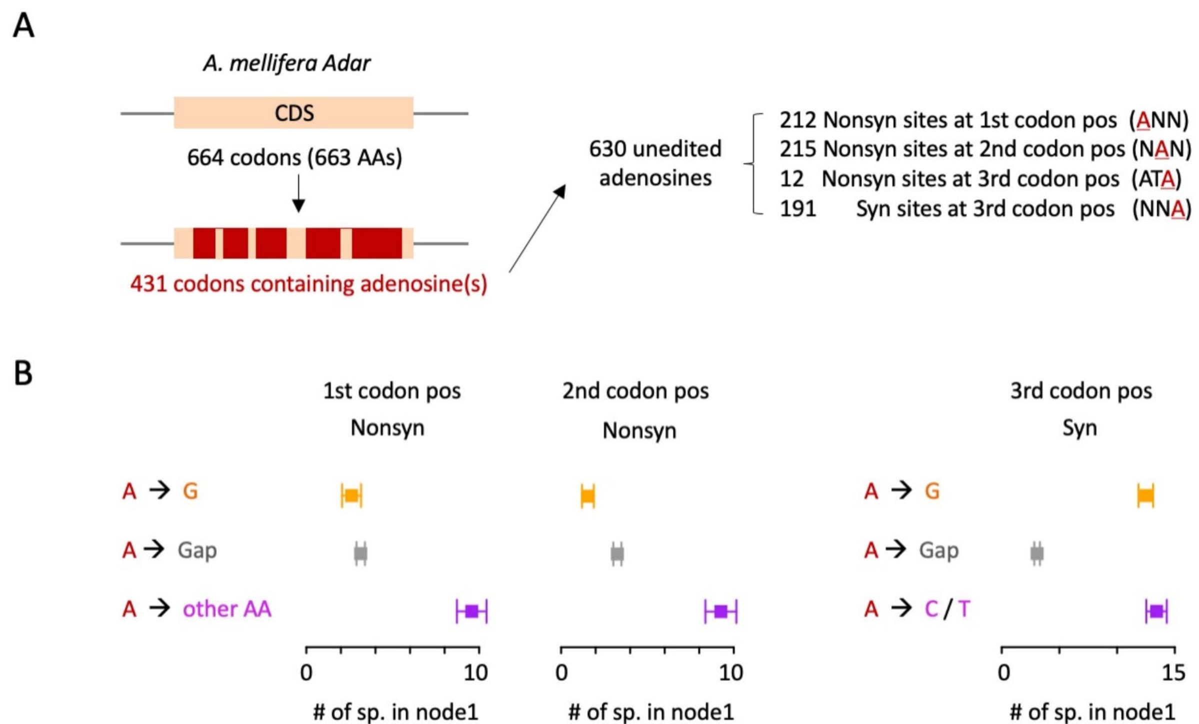

Fig. 5 Baseline substitution rates in node1 of hymenopterans. All adenosine sites in the *Adar* CDS were used. (**A**) *A. mellifera Adar* sequence was used as an example. All adenosines were divided into different groups according to their codon positions and consequences. (**B**) The numbers of occurrences of different types of derived mutations in the species of node1. Mean ± standard error (S.E.) were shown. The result of the 3rd codon position was compared to the results of the 1st and 2nd codon positions. Abbreviations: Nonsyn, nonsynonymous; Syn, synonymous; sp., species

In addition, we looked at the protein sequence similarity of different species. In the phylogenetic tree of Hymenoptera, we defined two sister clades: node2 and node3 (Fig. 7A). Node2 contains all the eusocial hymenopterans. We did not use node1 because it does not have a sister clade with sufficient number of species. We calculated the pairwise identity of protein sequences for the two RNA binding domains and one deaminase domain, respectively. The within-clade identity is slightly lower in node2 than node3, but a clear trend is that within-clade identity is higher than the inter-clade identity between node2 and node3 (Fig. 7A), suggesting that the editability of Ile>Met recoding site is generally maintained in node2 eusocial insects while in node3 the sequence context might not be favorable for RNA editing. However, we acknowledge that if we observe a higher within-group identity among node2 species, then it will be a more convincing evidence supporting the uniqueness and editability of Ile>Met recoding in eusocial insects. Nevertheless, the moderate conservation level of Adar protein sequence among eusocial Hymenoptera makes a sharp contrast to the highly constrained Ile>Met recoding site in this clade, which highlights a putative function and essentiality of this RNA editing site.

Adar is expressed in eusocial termites (Blattodea) but the orthologous site is an uneditable ile codon

Hymenoptera contributes most of the well-known eusocial insects, but eusociality did independently emerged in insects of other orders. Termites, originally classified as Isoptera but has now been grouped into Blattodea, are eusocial insects of an earlier-diverging clade [40]. We are curious whether termites exhibit a similar adaptive evolution of RNA editing. We obtained the *Adar* sequence of three Blattodea species including two termites (*Coptotermes formosanus* and *Zootermopsis nevadensis*) and a cockroach (*Blattella germanica*), and searched for their head transcriptomes. Genome sequence comparison shows that the position corresponding to honeybee Ile>Met editing site is an uneditable Ile codon ATT in three Blattodea species (Fig. 7B). This precludes the possibility of RNA editing at this position. Transcriptome mapping shows that *Adar* is expressed in heads but no A-to-I RNA editing events were observed in the entire CDS (Fig. 7B).

These results seem contradictory to the notion that the Ile>Met recoding contributes to the flexibility of eusociality in Hymenoptera. However, the "evolutionary tinkering" theory [41], originally proposed by Jacob in 1977 [42], tells us that the evolutionary process is imperfect and that the genomes of organisms

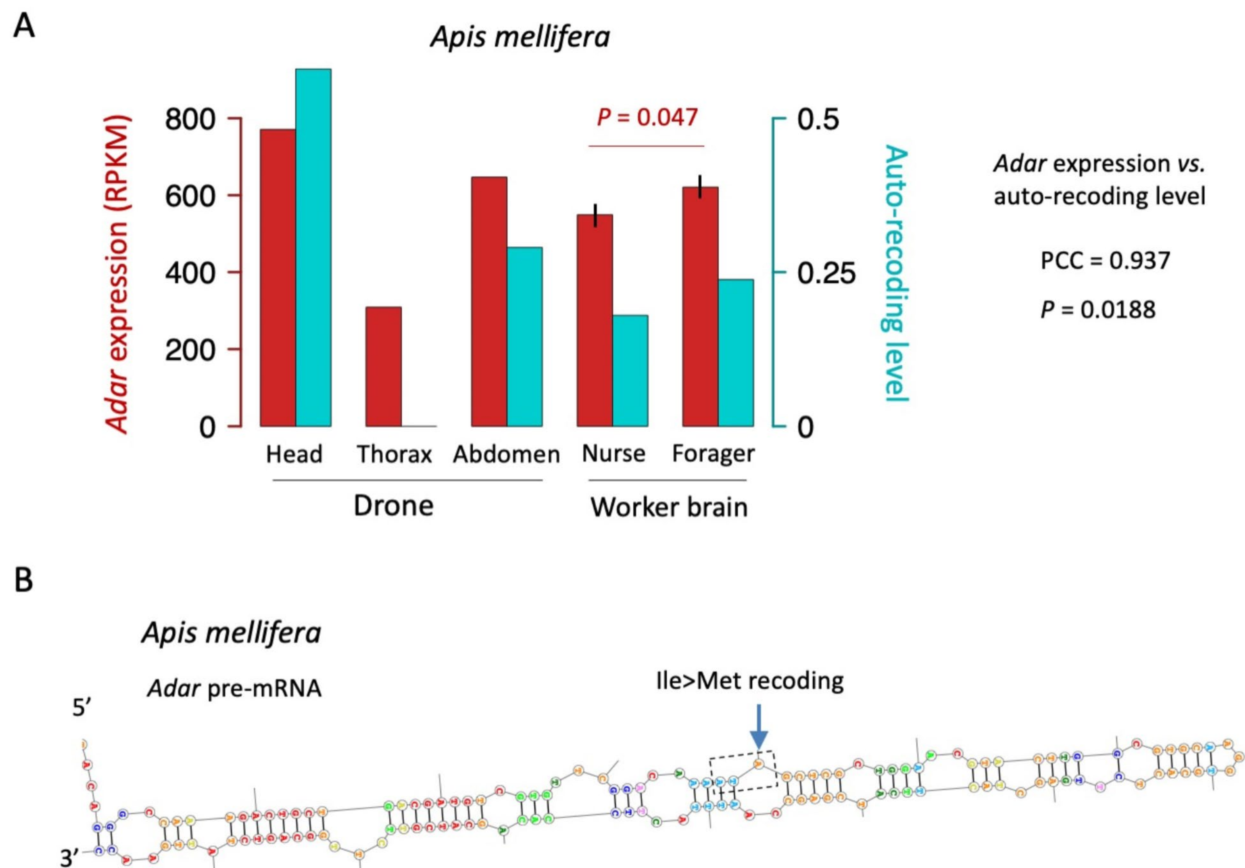

Fig. 6 *Adar* expression and RNA structure in honeybee. (**A**) *A. mellifera Adar* gene expression and Ile > Met recoding levels in different examples. Mean ± standard error (S.E.) were shown for the expression in six nurses and six foragers. *P* value was measured by Wilcoxon rank sum test. Pearson correlation between *Adar* expression and Ile > Met recoding level was calculated. (**B**) The secondary structure around Ile > Met recoding site in *Adar* pre-mRNA of honeybee

are making corrections and improvements based on what they already have. It is possible that eusociality of distantly related clades was achieved through different molecular mechanisms. While eusocial Hymenoptera take advantage of auto-recoding in *Adar* gene, termites might benefit from RNA editing at other positions, or from other mechanisms like gene expression regulation.

Discussion

Measuring the adaption of individual editing sites using phylogenetic methodology

In this article, we found a novel angle to measure the adaptation of RNA recoding sites. Previous methodologies mainly focus on the genome-wide comparison of the occurrence and editing levels between nonsynonymous and synonymous sites [12], making an attempt to discover the selection force acting on the overall recoding sites. This methodology has two potential limitations: (1) we can only know the selection on the global RNA editing profile. But for individual editing sites, we cannot tell whether it is beneficial or not; (2) the analysis can show the signal of adaptation to indicate that recoding events are beneficial, but it is unclear what exactly the advantage is. Flexible diversification of the proteome is one of the possible advantages, but mimicking a DNA mutation (if G-allele is constantly better) can also confer an advantage to the organism compared to no editing. Either flexible RNA editing or mimicking DNA mutations can lead to the overrepresentation of recoding sites compared to random expectation [12, 43].

Our current works propose a new methodology for judging the adaptation of individual RNA editing sites and can even exactly tell what hypothesis does this site conform to. In the context of phylogeny, the G-better hypothesis predicts a preference on genomic substitutions from A to G; but the flexible RNA editing hypothesis predicts that such A-to-G mutations should be avoided, so do the mutations to an uneditable allele. The question is, how to measure the preference or avoidance? What is the negative control group to be compared with? The solution is to look at the same type of codons located in the same gene (Figs. 3B and 4), or the adenosines located at the same codon position (1st, 2nd, or 3rd) within the same gene (Fig. 5).

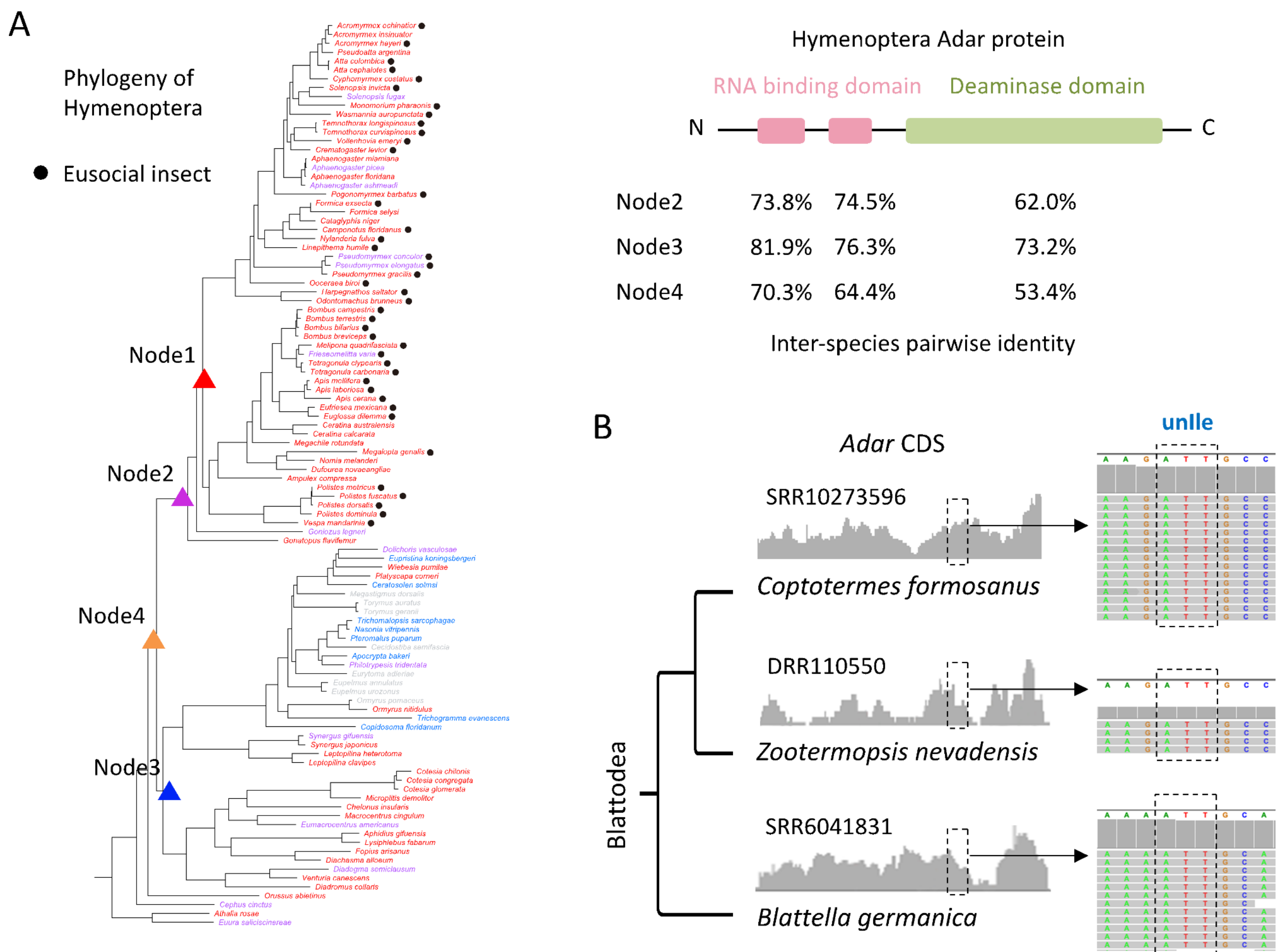

Fig. 7 Protein identity of hymenopteran Adar and the *Adar* expression in three Blattodea species. (**A**) Pairwise protein identity of different groups of species. The nodes of interest are labeled in the tree. Node4 includes the species of node2 and node3. Three domains are compared separately. (**B**) Expression of *Adar* in heads of three Blattodea species. The orthologous position of the bee Ile > Met recoding site is enlarged and visualized. No RNA editing is seen in the corresponding site as the codon itself is an uneditable Ile (ATT). No other editing events are observed in the entire CDS either

Using the new methodology, our observations on the *Adar* Ile>Met auto-recoding sites in Hymenoptera support the flexible RNA editing hypothesis. This methodology fully takes advantage of the large number of publicly available genomes, it circumvents the searching for appropriate transcriptomes to identify RNA editing sites. As established in this field, the accurate identification of RNA editing sites requires transcriptomes from the head or nerve systems and the matched DNA re-sequencing from the matched individual [10, 24, 44]. These requirements are extremely unfeasible in small-sized insects. Therefore, our comparative genomic approach in the light of phylogeny will facilitate the inference of adaptation of RNA recoding.

Vertical and horizontal approaches to investigate the flexible RNA recoding sites

Our analyses on substitution rates in the phylogeny of Hymenoptera were performed in a vertical dimension where the sequences in many different species were used. By showing that the Ile>Met edited codon was rarely replaced with hardwired genomic alleles, the flexible RNA editing hypothesis was supported. In contrast, there is another horizontal dimension that can putatively judge the advantage of flexible RNA editing. Take the *Adar* auto-recoding site for instance, Hymenoptera has an Ile>Met site evolved from an ancestral uneditable Ile codon, and Diptera has a Ser>Gly site evolved from an ancestral uneditable Ser codon (Fig. 1C). If the uneditable-to-editable transitions are beneficial at genome-wide level and are likely to be maintained by natural selection, then in the current genomes of extant species, we should observe an enrichment of such nonsynonymous editing sites whose pre-editing codon has an uneditable counterpart(s). The AT*A*>AT*G* (Ile>Met) and *A*GC/T>*G*GC/T (Ser>Gly) recoding events are the perfect cases.

We interrogate the previously reported RNA editome of bumblebee [32]. Totally 164 recoding sites were identified, among which 27 sites were Ser>Gly (16.5%) and 8 sites were Ile>Met (4.9%). In contrast, presume that all the adenosines in CDS were replaced with guanosines, there will be 6,020,316 nonsynonymous mutations, including 176,475 Ser>Gly mutations (2.9%) and 133,351 Ile>Met mutations (2.2%). The observed occurrences of Ser>Gly and Ile>Met recoding were significantly higher than random expectation estimated from genome-wide investigation ($P<0.05$ in both cases under Fisher's exact tests). This enrichment of Ser>Gly and Ile>Met RNA recoding sites supports the advantage of RNA editing (being editable), where the uneditable-to-editable transitions might be driven by the need for editing and be selectively maintained during evolution. However, we do not claim that all the observed Ser>Gly and Ile>Met editing sites in bumblebee have an uneditable codon in the ancestral genome of Hymenoptera or insects. The enrichment is just a tendency: many unedited codons (e.g. ATA) in the genome might also evolve from an ancestral uneditable codon (e.g. ATC or ATT), but editing has not yet occurred in the current species; the current recoding sites (e.g. edited AT*A*) might also be inherited from a deeply conserved editing site that appeared in the common ancestor.

Flexible RNA editing facilitates the transcriptomic plasticity of eusocial insects?

Eusocial insects have evolved the ability to generate phenotypically differentiated individuals from the same set of genomes [34, 35]. This phenomenon promotes the investigation on epigenetic or transcriptomic regulations, where RNA editing is possibly one of the regulatory mechanisms. In bumblebee, previous study showed the correlation between the *Adar* Ile>Met auto-recoding level and the global editing efficiency [32], suggesting a feedback regulatory role of the hymenopteran auto-editing site just like the *Adar* Ser>Gly site in *Drosophila* [31]. Our results show that although the Ile codon is mutated in a few Hymenoptera species, the latest branch containing all the eusocial hymenopterans consistently has the ATA codon across the phylogeny (Fig. 2). This observation raises a possibility that the Adar-mediated RNA editing contributes to the phenotypic plasticity in eusocial insects so that the global RNA editing activity should be precisely regulated by the Ile>Met auto-recoding site. This auto-recoding sites in *Adar* might act as a stabilizer through a feedback loop and this mechanism is indispensable only when RNA editing itself is important for the organism.

Future perspectives

Several questions are promisingly to be addressed in the future: (1) For the many other eusocial hymenopterans in node1 defined by us, can Ile>Met recoding be detected in the transcriptomes? This will be tested when appropriate data are generated. (2) Is there condition-specificity of the Ile>Met recoding site in other eusocial insects like what we have shown for leaf-cutting ant and honeybee? (3) For the non-social hymenopterans which might have lower needs for phenotypic plasticity, if Ile>Met editing is detected in their transcriptomes, then what will be the significance of RNA editing to these species? Answers to these questions will help us establish the connection between RNA editing and the sociality of insects, advancing our understanding on the molecular basis of phenotypic plasticity, and help bridge the gap between genotype and phenotype.

Conclusions

In the Hymenoptera phylogeny, the genomic A-to-G and editable-to-uneditable transitions were significantly suppressed at the Ile>Met site, but the mutations to other AAs were not suppressed. As a consequence of adaptive evolution from uneditable-to-editable status, the recoding sites with such editable codons are enriched in the current genome. In the light of comparative and evolutionary genomics, our work developed a new angle to view and test the adaptation of global or individual RNA recoding sites. We revealed adaptive RNA editing sites in eusocial hymenopterans, providing another layer to account for the great phenotypical plasticity of eusocial insects.

Materials and methods

Data Availability

The analysis of the *Adar* auto-recoding codon involves the reference sequences of Hymenoptera species, which were downloaded from NCBI https://www.ncbi.nlm.nih.gov/. The detailed links of the data were supplemented in our previous works [30, 33]. The phylogeny of Hymenoptera was also retrieved from our previous study [30]. The brain transcriptomes of leaf-cutting ant *Acromyrmex echinatior* were downloaded from NCBI under accession ID SRP031846. The caste- and tissue-specific transcriptomes of honeybee *Apis mellifera* were downloaded from NCBI under accession ID SRR445999-SRR446004 (nurse brain) and SRR446005-SRR446010 (forager brain), and from Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/) under accession ID CRA002262 (drone head/thorax/abdomen). The RNA editing sites in bumblebee *Bombus terrestris* was downloaded from the original study [32]. The *Adar* CDS sequences of 104 Hymenoptera species were provided as fasta format in Additional file2: Supplementary Data 1. The *Adar*

sequences of three Blattodea species were retrieved from a previous study involving gene alignment in insects [33]. Their transcriptomes were downloaded from NCBI with accession IDs SRR10273596 (*Coptotermes formosanus*), DRR110550 (*Zootermopsis nevadensis*), and SRR6041831 (*Blattella germanica*). The determination of eusocial hymenopteran insects was based on the following list of literatures [45–56].

Sequence alignment and codon extraction

The protein sequences of *Adar* genes were aligned using the G-INS-i strategy in MAFFT v7.310 [57] with default parameters. Then, with TranslatorX v1.1 [58], the nucleotide sequences (CDS sequence) of *Adar* genes were aligned based on the pre-aligned protein sequences. Default parameters were used. The alignment file was split with tri-nucleotide periodicity, extracting all the aligned codons.

Transcriptome mapping and the visualization of RNA editing sites

The transcriptomic reads were mapped to the *Adar* CDS using BWA v0.7.17 under default parameters [59]. The reference sequences of different species were used for mapping the reads of the corresponding species. The editing levels on the Ile>Met recoding site were visualized through IGV. Gene expression of *Adar* was measured by RPKM (reads per kilobase per million mapped reads). The number of mapped reads for each library refers to the reads mappable to the reference genome of *Apis mellifera* version Amel_HAv3.1.

Presumed A-to-G mutations in *adar* CDS as a random expectation

To obtain an expected occurrence of codon transitions in the Hymenoptera phylogeny, we utilized the *Apis mellifera Adar* CDS as the reference sequence due to its highest accuracy among hymenopteran species. The *Adar* CDS of 664 codons encoding 663 AAs except the last stop-codon. Among them, 431 codons contain at least one adenosine and totally 631 adenosines are there. Apart from the adenosine at the 3rd position of the Ile>Met auto-recoding site, there are 630 unedited adenosines in *Adar* CDS. Presume that an adenosine is mutated to a guanosine, according to the consequence in codon changes, these 630 adenosines will include 212 nonsynonymous mutations at the 1st codon position, 215 nonsynonymous mutations at the 2nd codon position, 12 nonsynonymous mutations at the 3rd codon position, and 191 synonymous mutations at the 3rd codon position. Then the substitution rate at each position was calculated by counting how many species have a different codon or AA in the Hymenoptera phylogeny.

RNA structure prediction

The dsRNA secondary structure was folded through online server RNAstructure (https://rna.urmc.rochester.edu/RNAstructureWeb/). The entire pre-mRNA of honeybee *Adar* gene was retrieved, and the 2000 bp region centered by the Ile>Met editing site was used as input sequence.

Statistics and graphic works

The statistics like mean and S.E. were performed in R studio (version 3.6.3). The graphic works were done in R studio or Adobe Illustrator version 2023.

Abbreviations

AA	Amino acid
A-to-I	Adenosine-to-inosine
ADAR	Adenosine deaminase acting on RNA
CDS	Coding sequence
dsRNA	Double-stranded RNA
mut.	Mutation
Nonsyn	Nonsynonymous
Syn	Synonymous
S.E.	Standard error
sp.	Species
edIle	Editable isoleucine codon
unIle	Uneditable isoleucine codon

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10709-0.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We thank the funders for the financial support. We thank Dr Jixiang Cui for providing the literatures and information of eusocial insects. We thank Dr Ling Ma for the help in aligning Adar protein sequences. We also thank the other members in Cai Lab for their support to this project.

Author contributions

Conceptualization & supervision: Y.D. Data analysis: Y.D., J.L., and C.Z. Writing – original draft: Y.D. Writing – review & editing: C.Z., J.L., and Y.D. All authors approved the submission of this manuscript.

Funding

This study is financially supported by the National Natural Science Foundation of China (no. 32300371) the Young Elite Scientist Sponsorship Program by CAST (no. 2023QNRC001), and the Young Elite Scientist Sponsorship Program by BAST (no. BYESS2023160).

Data availability

The analysis of the Adar auto-recoding codon involves the reference sequences of Hymenoptera species, which were downloaded from NCBI https://www.ncbi.nlm.nih.gov/. The detailed links of the data were supplemented in our previous works. The phylogeny of Hymenoptera was also retrieved from our previous study. The brain transcriptomes of leaf-cutting ant Acromyrmex echinatior were downloaded from NCBI under accession ID SRP031846. The caste- and tissue-specific transcriptomes of honeybee Apis mellifera were downloaded from NCBI under accession ID SRR445999-SRR446004 (nurse brain) and SRR446005-SRR446010 (forager brain), and from Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/) under accession ID CRA002262 (drone head/thorax/abdomen). The RNA editing sites in bumblebee Bombus terrestris was downloaded from the original study. The Adar CDS sequences of 104 Hymenoptera species were provided

as fasta format in Additional file2: Supplementary Data 1. The Adar sequences of three Blattodea species were retrieved from a previous study involving gene alignment in insects. Their transcriptomes were downloaded from NCBI with accession IDs SRR10273596 (Coptotermes formosanus), DRR110550 (Zootermopsis nevadensis), and SRR6041831 (Blattella germanica). The determination of eusocial hymenopteran insects was based on the following list of literatures.

Declarations

Ethics approval and consent to participate
This study does not use any sample of animals, plants, or humans.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 22 April 2024 / Accepted: 14 August 2024
Published online: 26 August 2024

References
1. Keegan LP, Hajji K, O'Connell MA. Adenosine deaminase acting on RNA (ADAR) enzymes: a Journey from weird to wondrous. Acc Chem Res. 2023;56(22):3165–74.
2. Thuy-Boun AS, Thomas JM, Grajo HL, Palumbo CM, Park S, Nguyen LT, Fisher AJ, Beal PA. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. Nucleic Acids Res. 2020;48(14):7958–72.
3. Gallo A, Keegan LP, Ring GM, O'Connell MA. An ADAR that edits transcripts encoding ion channel subunits functions as a dimer. EMBO J. 2003;22(13):3421–30.
4. Duan Y, Li H, Cai W. Adaptation of A-to-I RNA editing in bacteria, fungi, and animals. Front Microbiol. 2023;14:1204080.
5. Zhang P, Zhu Y, Guo Q, Li J, Zhan X, Yu H, Xie N, Tan H, Lundholm N, Garcia-Cuetos L, et al. On the origin and evolution of RNA editing in metazoans. Cell Rep. 2023;42(2):112112.
6. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. 2014;24(3):365–76.
7. Bazak L, Levanon EY, Eisenberg E. Genome-wide analysis of Alu Editability. Nucleic Acids Res. 2014;42(11):6876–84.
8. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol. 2004;22(8):1001–5.
9. Rajendren S, Dhakal A, Vadlamani P, Townsend J, Deffit SN, Hundley HA. Profiling neural editomes reveals a molecular mechanism to regulate RNA editing during development. Genome Res. 2021;31(1):27–39.
10. Duan Y, Ma L, Liu J, Liu X, Song F, Tian L, Cai W, Li H. The first A-to-I RNA editome of hemipteran species *Coridius chinensis* reveals overrepresented recoding and prevalent intron editing in early-diverging insects. Cell Mol Life Sci. 2024;81(1):136.
11. Duan Y, Xu Y, Song F, Tian L, Cai W, Li H. Differential adaptive RNA editing signals between insects and plants revealed by a new measurement termed haplotype diversity. Biol Direct. 2023;18(1):47.
12. Yablonovitch AL, Deng P, Jacobson D, Li JB. The evolution and adaptation of A-to-I RNA editing. PLoS Genet. 2017;13(11):e1007064.
13. Yu Y, Zhou H, Kong Y, Pan B, Chen L, Wang H, Hao P, Li X. The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. PLoS Genet. 2016;12(7):e1006191.
14. Zhao T, Ma L, Xu S, Cai W, Li H, Duan Y. Narrowing down the candidates of beneficial A-to-I RNA editing by comparing the recoding sites with uneditable counterparts. Nucleus (Calcutta). 2024;15(1):2304503.
15. Zheng C, Ma L, Song F, Tian L, Cai W, Li H, Duan Y. Comparative genomic analyses reveal evidence for adaptive A-to-I RNA editing in insect *adar* gene. Epigenetics. 2024;19(1):2333665.
16. Basilio C, Wahba AJ, Lengyel P, Speyer JF, Ochoa S. Synthetic polynucleotides and the amino acid code. V. Proc Natl Acad Sci U S A. 1962;48:613–6.
17. Eisenberg E, Levanon EY. A-to-I RNA editing - immune protector and transcriptome diversifier. Nat Rev Genet. 2018;19(8):473–90.
18. Alon S, Garrett SC, Levanon EY, Olson S, Graveley BR, Rosenthal JJ, Eisenberg E. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. eLife. 2015;4:e05198.
19. Duan Y, Ma L, Zhao T, Liu J, Zheng C, Song F, Tian L, Cai W, Li H. Conserved A-to-I RNA editing with non-conserved recoding expands the candidates of functional editing sites. Fly (Austin). 2024;18(1):2367359.
20. Gommans WM, Mullen SP, Maas S. RNA editing: a driving force for adaptive evolution? BioEssays. 2009;31(10):1137–45.
21. Shoshan Y, Liscovitch-Brauer N, Rosenthal JJC, Eisenberg E. Adaptive proteome diversification by nonsynonymous A-to-I RNA editing in coleoid cephalopods. Mol Biol Evol. 2021;38(9):3775–88.
22. Xu G, Zhang J. Human coding RNA editing is generally nonadaptive. Proc Natl Acad Sci U S A. 2014;111(10):3769–74.
23. Zhang J, Xu C. Gene product diversity: adaptive or not? Trends Genet. 2022;38(11):1112–22.
24. Liscovitch-Brauer N, Alon S, Porath HT, Elstein B, Unger R, Ziv T, Admon A, Levanon EY, Rosenthal JJC, Eisenberg E. Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell. 2017;169(2):191–e202111.
25. Zhan D, Zheng C, Cai W, Li H, Duan Y. The many roles of A-to-I RNA editing in animals: functional or adaptive? Front Biosci Landmark. 2023;28(10):256.
26. Birk MA, Liscovitch-Brauer N, Dominguez MJ, McNeme S, Yue Y, Hoff JD, Twersky I, Verhey KJ, Sutton RB, Eisenberg E, et al. Temperature-dependent RNA editing in octopus extensively recodes the neural proteome. Cell. 2023;186(12):2544–e25552513.
27. Xin K, Zhang Y, Fan L, Qi Z, Feng C, Wang Q, Jiang C, Xu JR, Liu H. Experimental evidence for the functional importance and adaptive advantage of A-to-I RNA editing in fungi. Proc Natl Acad Sci U S A. 2023;120(12):e2219029120.
28. Rangan KJ, Reck-Peterson SL. RNA recoding in cephalopods tailors microtubule motor protein function. Cell. 2023;186(12):2531–e25432511.
29. Keegan LP, McGurk L, Palavicini JP, Brindle J, Paro S, Li X, Rosenthal JJ, O'Connell MA. Functional conservation in human and *Drosophila* of metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects. Nucleic Acids Res. 2011;39(16):7249–62.
30. Ma L, Zheng C, Xu S, Xu Y, Song F, Tian L, Cai W, Li H, Duan Y. A full repertoire of Hemiptera genomes reveals a multi-step evolutionary trajectory of auto-RNA editing site in insect *adar* gene. RNA Biol. 2023;20(1):703–14.
31. Savva YA, Jepson JE, Sahin A, Sugden AU, Dorsky JS, Alpert L, Lawrence C, Reenan RA. Auto-regulatory RNA editing fine-tunes mRNA re-coding and complex behaviour in *Drosophila*. Nat Commun. 2012;3:790.
32. Porath HT, Hazan E, Shpigler H, Cohen M, Band M, Ben-Shahar Y, Levanon EY, Eisenberg E, Bloch G. RNA editing is abundant and correlates with task performance in a social bumblebee. Nat Commun. 2019;10:1605.
33. Duan Y, Ma L, Song F, Tian L, Cai W, Li H. Autorecoding A-to-I RNA editing sites in the *Adar* gene underwent compensatory gains and losses in major insect clades. RNA. 2023;29(10):1509–19.
34. Rajakumar R, San Mauro D, Dijkstra MB, Huang MH, Wheeler DE, Hiou-Tim F, Khila A, Cournoyea M, Abouheif E. Ancestral developmental potential facilitates parallel evolution in ants. Science. 2012;335(6064):79–82.
35. Smith CR, Toth AL, Suarez AV, Robinson GE. Genetic and genomic analyses of the division of labour in insect societies. Nat Rev Genet. 2008;9(10):735–48.
36. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. Evolutionary history of the Hymenoptera. Curr Biol. 2017;27(7):1013–8.
37. Zhang Y, Duan Y. Genome-wide analysis on driver and passenger RNA editing sites suggests an underestimation of adaptive signals in insects. Genes (Basel). 2023;14(10):1951.
38. Li Q, Wang Z, Lian J, Schiott M, Jin L, Zhang P, Zhang Y, Nygaard S, Peng Z, Zhou Y, et al. Caste-specific RNA editomes in the leaf-cutting ant *Acromyrmex Echinatior*. Nat Commun. 2014;5:4943.
39. Savva YA, Rieder LE, Reenan RA. The ADAR protein family. Genome Biol. 2012;13(12):252.
40. Krishna K, Grimaldi DA, Krishna V, Engel MS. Treatise on the Isoptera of the world. Bull Am Museum Nat History. 2013;377(1):1–200.
41. Sercarz E. Redundant regulatory mechanisms as an outcome of evolutionary tinkering. Res Immunol. 1992;143(3):341–4.
42. Jacob F. Evolution and tinkering. Science. 1977;196(4295):1161–6.

43. Duan Y, Cai W, Li H. Chloroplast C-to-U RNA editing in vascular plants is adaptive due to its restorative effect: testing the restorative hypothesis. RNA. 2023;29(2):141–52.
44. Xu Y, Liu J, Zhao T, Song F, Tian L, Cai W, Li H, Duan Y. Identification and interpretation of A-to-I RNA editing events in insect transcriptomes. Int J Mol Sci. 2023;24(24):17126.
45. Field J, Leadbeater E. Cooperation between non-relatives in a primitively eusocial paper wasp, *Polistes Dominula*. Philos Trans R Soc Lond B Biol Sci. 2016;371(1687):20150093.
46. Huisken JL, Shell WA, Pare HK, Rehan SM. The influence of social environment on cooperation and conflict in an incipiently social bee, *Ceratina calcarata*. Behav Eco Sociobiol. 2021;75(4):74.
47. Hunt JH, Amdam GV. Bivoltinism as an antecedent to eusociality in the paper wasp genus *Polistes*. Science. 2005;308(5719):264–7.
48. Johnson RM, Harpur BA, Dogantzis KA, et al. Genomic footprint of evolution of eusociality in bees: floral food use and CYPome blooms. Insect Soc. 2018;65:445–54.
49. Koethe S, Reinartz L, Heard TA, Garcia JE, Dyer AG, Lunau K. Comparative psychophysics of western honey bee (*Apis mellifera*) and stingless bee (*Tetragonula carbonaria*) colour purity and intensity perception. J Comp Physiol Neuroethol Sens Neural Behav Physiol. 2022;208(5–6):641–52.
50. Shell WA. SM Rehan 2022 Social divergence: molecular pathways underlying castes and longevity in a facultatively eusocial small carpenter bee. Proc Biol Sci 289 1971 20212663.
51. Suenami S, Konishi Nobu M, Miyazaki R. Community analysis of gut microbiota in hornets, the largest eusocial wasps, *Vespa Mandarinia* and *V. Simillima*. Sci Rep. 2019;9(1):9830.
52. Tepedino VJ, Frohlich DR. Fratricide in *Megachile rotundata*, a non-social megachilid bee - impartial treatment of sibs and non-sibs. Behav Eco Sociobiol. 1984;15(1):19–23.
53. Libbrecht R, Keller L. The making of eusociality: insights from two bumblebee genomes. Genome Biol. 2015;16(1):75.
54. Mikhailova AA, Rinke S, Harrison MC. Genomic signatures of eusocial evolution in insects. Curr Opin Insect Sci. 2024;61:101136.
55. Rubin BER. Social insect colony size is correlated with rates of molecular evolution. Insect Soc. 2022;69(2–3):147–57.
56. Séguret A, Bernadou A, Paxton RJ. Facultative social insects can provide insights into the reversal of the longevity/fecundity trade-off across the eusocial insects. Curr Opin Insect Sci. 2016;16:95–103.
57. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
58. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38:W7–13.
59. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.