

ARTICLE

<https://doi.org/10.1038/s41467-019-13275-2>

OPEN

The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive

Daohan Jiang ¹ & Jianzhi Zhang ^{1*}

A-to-I editing enzymatically converts the base adenosine (A) in RNA molecules to inosine (I), which is recognized as guanine (G) in translation. Exceptionally abundant A-to-I editing was recently discovered in the neural tissues of coleoids (octopuses, squids, and cuttlefishes), with a greater fraction of nonsynonymous sites than synonymous sites subject to high levels of editing. Although this phenomenon is thought to indicate widespread adaptive editing, its potential advantage is unknown. Here we propose an alternative, nonadaptive explanation. Specifically, increasing the cellular editing activity permits some otherwise harmful G-to-A nonsynonymous substitutions, because the As are edited to Is at sufficiently high levels. These high editing levels are constrained upon substitutions, resulting in the predominance of nonsynonymous editing at highly edited sites. Our evidence for this explanation suggests that the prevalent nonsynonymous editing in coleoids is generally nonadaptive, as in species with much lower editing activities.

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA. *email: jianzhi@umich.edu

RNA editing refers to a variety of posttranscriptional alterations of RNA molecules, including chemical modifications as well as insertions and deletions of nucleotides, but excluding RNA processing events such as splicing, capping, and polyadenylation^{1,2}. Transcriptome-wide profiling of each type of RNA editing and understanding its biochemical and physiological functions are a major task of molecular and genome biology, and have seen a rapid progress in the last decade^{3–6}. Among over 100 different types of RNA editing, adenosine (A)-to-inosine (I) editing of RNAs transcribed from animal nuclear genomes is arguably best studied^{7–9}. The A-to-I conversion is catalyzed by a family of adenosine deaminase acting on RNA (ADAR) and the resultant I is recognized as guanine (G) in translation. For simplicity, we refer to A-to-I editing as A-to-G editing hereafter. If the editing takes place in protein-coding regions, it could be either nonsynonymous (also known as recoding) or synonymous, depending on whether the encoded amino acid is altered or not. A-to-G editing has been reported in multiple animal phyla^{10,11}, such as many vertebrates^{10,12–17}, as well as fruit flies^{18–23}, cephalopods^{24–27}, nematodes^{28,29}, and cnidarians³⁰. Although an editing mechanism could emerge by chance and become fixed by genetic drift³¹, studies of functional consequences of a handful of A-to-G recoding events led to the initial belief that recoding offers an “extreme advantage³²,” because disrupting recoding could be lethal³³. This view has been challenged in the last few years by transcriptome-wide analysis of RNA editing. Specifically, there is a long tradition in molecular evolutionary genetics to compare the rate of synonymous nucleotide substitution (d_S) with that of nonsynonymous substitution (d_N) in protein-coding DNA sequence evolution. As synonymous changes are presumably neutral, while nonsynonymous changes may or may not be neutral, an observation of $d_N > d_S$ indicates overall positive selection promoting beneficial nonsynonymous substitutions, whereas $d_N < d_S$ indicates overall purifying selection hindering deleterious nonsynonymous substitutions. Although RNA editing is a molecular phenotype, similar comparisons between synonymous and nonsynonymous editing can be made³⁴. For instance, in humans, the fraction of sites subject to nonsynonymous editing is lower than that subject to synonymous editing and the editing level (i.e., the proportion of RNA molecules edited at a site) is also lower for nonsynonymous than synonymous editing³⁴. These patterns suggest that nonsynonymous editing is generally deleterious and is selectively removed and/or suppressed when compared with synonymous editing, which is presumably inconsequential to protein function. Therefore, most A-to-G coding RNA-editing events appear to be nonadaptive and are probably attributable to cellular errors resulting from ADARs’ limited specificity³⁴. This conclusion is compatible with the fact that only a handful of editing events have known functions³³, and that only 1.8% of ~2000 human coding RNA-editing events are shared with mouse^{35,36}.

The trend, however, is drastically different in coleoid cephalopods, which include octopuses, squids, and cuttlefishes. Tens of thousands of coding A-to-G editing events, including a considerable proportion of recoding, have been identified in the neural tissues of coleoids^{25,27}. In particular, the frequency of nonsynonymous sites subject to high levels of editing exceeds that of synonymous sites, leading to the inference that nonsynonymous editing has been promoted by positive selection and is generally advantageous in coleoids^{25,27}. We will refer to this hypothesis as the adaptive hypothesis. Furthermore, because the high editing activity appears to be limited to their neural tissues, it was speculated that the extraordinary abundance of RNA editing in coleoids is related to their complex nervous system and behavior^{24,25,27,37}. Nonetheless, with the exception of recoding of an octopus potassium channel that is associated with cold

adaptation²⁶, no benefit of the widespread editing is known in coleoids. Here we propose and provide evidence for an alternative, nonadaptive explanation of the preponderance of highly edited nonsynonymous sites in coleoids.

Results

A nonadaptive hypothesis and its predictions. Let us consider a genomic position in a coding region that is currently occupied by G and does not accept A (see top row in Fig. 1a). As the editing activity in the species rises, a G-to-A mutation at the site may become neutral and fixed if the resultant A is edited back to G in a sufficiently large proportion of mRNA molecules (see middle row in Fig. 1a). Upon the G-to-A substitution, the high editing level at the site will be selectively maintained, because it is G rather than A that is permissible at the mRNA level. As the above situation applies only to nonsynonymous G-to-A substitutions and the coupled nonsynonymous A-to-G editing, it inflates the number of nonsynonymous editing sites and nonsynonymous editing levels relative to the corresponding synonymous values. Although here the nonsynonymous editing has permitted the fixation of the otherwise deleterious G-to-A mutation, the derived genotype with a genomic A that is highly edited is no fitter than the original genotype with a genomic G. Thus, the editing is nonadaptive. We assumed in the above scenario that the editing level is so high that the otherwise deleterious G-to-A mutation becomes neutral. It is also possible that the editing level is not high enough, rendering the G-to-A mutation slightly deleterious (see bottom row in Fig. 1a). A slightly deleterious mutation may nevertheless get fixed and the editing level may be selectively increased in subsequent evolution. Even under this scenario, there is no net fitness gain from the original genotype with a genomic G to the derived genotype with a genomic A that is highly edited. We refer to the above nonadaptive model including both of the described scenarios as the harm-permitting model, because RNA editing permits the fixation of otherwise harmful mutations. Although the possibility of harm-permitting by RNA editing has been proposed multiple times^{31,38–40}, especially regarding the editing of organelle transcriptomes, empirical evidence that it is entirely or primarily responsible for creating “adaptive signals” of RNA editing is lacking.

Given the exceptionally high editing activity in coleoid neural tissues^{25,27}, we hypothesize that the reported preponderance of nonsynonymous editing is explained by the harm-permitting model and is nonadaptive. To test this hypothesis, we divide nonsynonymous editing into two categories: restorative and diversifying⁴¹. Restorative editing converts the amino acid state back to an ancestral state (Fig. 1b), whereas diversifying editing converts the amino acid state to a non-ancestral state (Fig. 1c). As restorative editing but not diversifying editing can confer a harm-permitting effect, our hypothesis predicts that the reported preponderance of nonsynonymous editing in coleoids is attributable to restorative but not diversifying editing. In particular, we predict that (i) the frequency of sites edited is greater for restorative (F_R) than synonymous (F_S) editing, and that (ii) the median editing level is higher for restorative (L_R) than synonymous (L_S) editing. It further predicts that (iii) the frequency of sites edited is no greater for diversifying (F_D) than synonymous (F_S) editing, and that (iv) the median editing level is no higher for diversifying (L_D) than synonymous (L_S) editing. By contrast, the adaptive hypothesis does not have specific predictions about F_R and L_R , but predicts that F_D and L_D are respectively greater than F_S and L_S . It is noteworthy that although only restorative editing can be harm-permitting, not all restorative editing is necessarily harm-permitting. For instance, the restorative editing would be neutral if it restores a neutral G-to-A substitution.

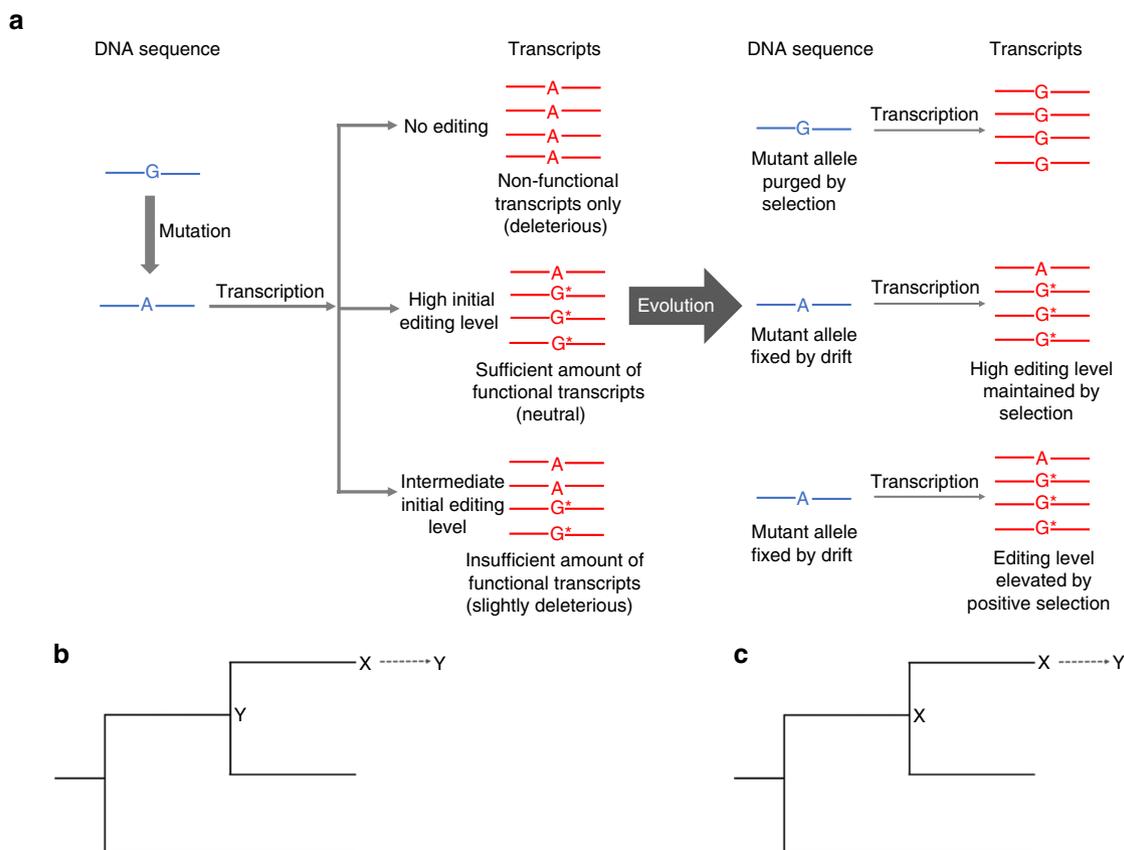


Fig. 1 The harm-permitting model and a strategy to detect the harm-permitting effect. **a** The harm-permitting effect of nonsynonymous editing. The top row shows that, when a nonsynonymous A site is not edited (or is subject to a low level of editing), a G-to-A mutation at the site is too deleterious to get fixed. The middle row shows that, when the site is highly edited, the G-to-A mutation becomes neutral and is fixed by genetic drift. The high editing level is then selectively constrained. The bottom row shows that, when the editing level of the site is intermediate, the G-to-A mutation is slightly deleterious and fixed by genetic drift. The editing level may be further elevated by positive selection (or maintained by negative selection). Despite the relatively high nonsynonymous editing levels in the middle and bottom rows, no adaptation (i.e., no net increase in fitness) occurred when the final genotype is compared with the original genotype. DNA is shown in blue, whereas RNA is in red. Post-edited nucleotides are marked with stars. **b** Restorative editing restores an ancestral amino acid state lost upon an amino acid substitution, which may have occurred in the exterior branch as shown here or in an earlier branch. In other words, the post-editing state is identical to an ancestral pre-editing state. **c** Diversifying editing creates an amino acid state that differs from pre-editing states in a set of ancestors considered. Although only the state of one ancestor is shown here, the states of multiple ancestors may be considered. In **b** and **c**, X and Y represent different amino acid states, whereas the arrow shows the effect of editing. Restorative but not diversifying editing can confer a harm-permitting effect.

Patterns of restorative and diversifying editing. To test the nonadaptive hypothesis, we analyzed the published neural transcriptomes of six mollusk species²⁷, whose phylogenetic relationships are depicted in Fig. 2a. Among them, the four coleoids have widespread coding A-to-G editing in neural tissues, whereas the two outgroups have substantially fewer editing sites²⁷.

We identified 3979 one-to-one orthologous genes in these six species and inferred ancestral coding sequences at all interior nodes of the species tree (Fig. 2a). We regarded a nonsynonymous editing event in an exterior node of the tree that modifies the amino acid state from X to Y as restorative if the inferred genomic sequence-based amino acid state is Y at any node of the tree that is ancestral to the focal exterior node (Fig. 1b; also see Methods), or diversifying if Y is not present at any node of the tree that is ancestral to the focal exterior node (Fig. 1c). It is worth noting that these definitions are based on amino acid states and are applied to nonsynonymous editing only. Synonymous editing is presumably neutral, so need not be separated into restorative and diversifying editing. Furthermore, separating synonymous editing into the two categories would be less accurate because of lower reliabilities in inferring ancestral sequences at synonymous sites. Of the two categories of nonsynonymous editing sites, the

number of diversifying editing sites is 15.7–20.4 times that of restorative editing sites in the four coleoids (Supplementary Table 1).

In each of the four coleoids, F_R is significantly lower than F_S when all editing sites are considered (Fig. 2b). However, because the harm-permitting hypothesis concerns relatively highly edited sites, we analyzed sites with editing levels exceeding 10% and found F_R to be significantly higher than F_S (Fig. 2b inset). We observed that L_R is significantly greater than L_S in all four coleoids (Fig. 2c). By contrast, F_D is significantly smaller than F_S when we considered all editing sites (Fig. 2b) or sites with >10% editing levels (Fig. 2b inset). L_D is not significantly different from L_S except in the squid (Fig. 2c). These results largely confirm all four predictions of the nonadaptive hypothesis and are at odds with the predictions of the adaptive hypothesis, strongly suggesting that the preponderance of nonsynonymous editing in coleoids is explained by the harm-permitting model and is nonadaptive. Figure 2c shows that, although L_R is significantly higher than L_S in each coleoid, it is lower than 2.5%. One might ask whether such low median levels of restorative editing can be harm-permitting. As mentioned, not all restorative editing is necessarily harm-permitting, which could explain why L_R is not particularly

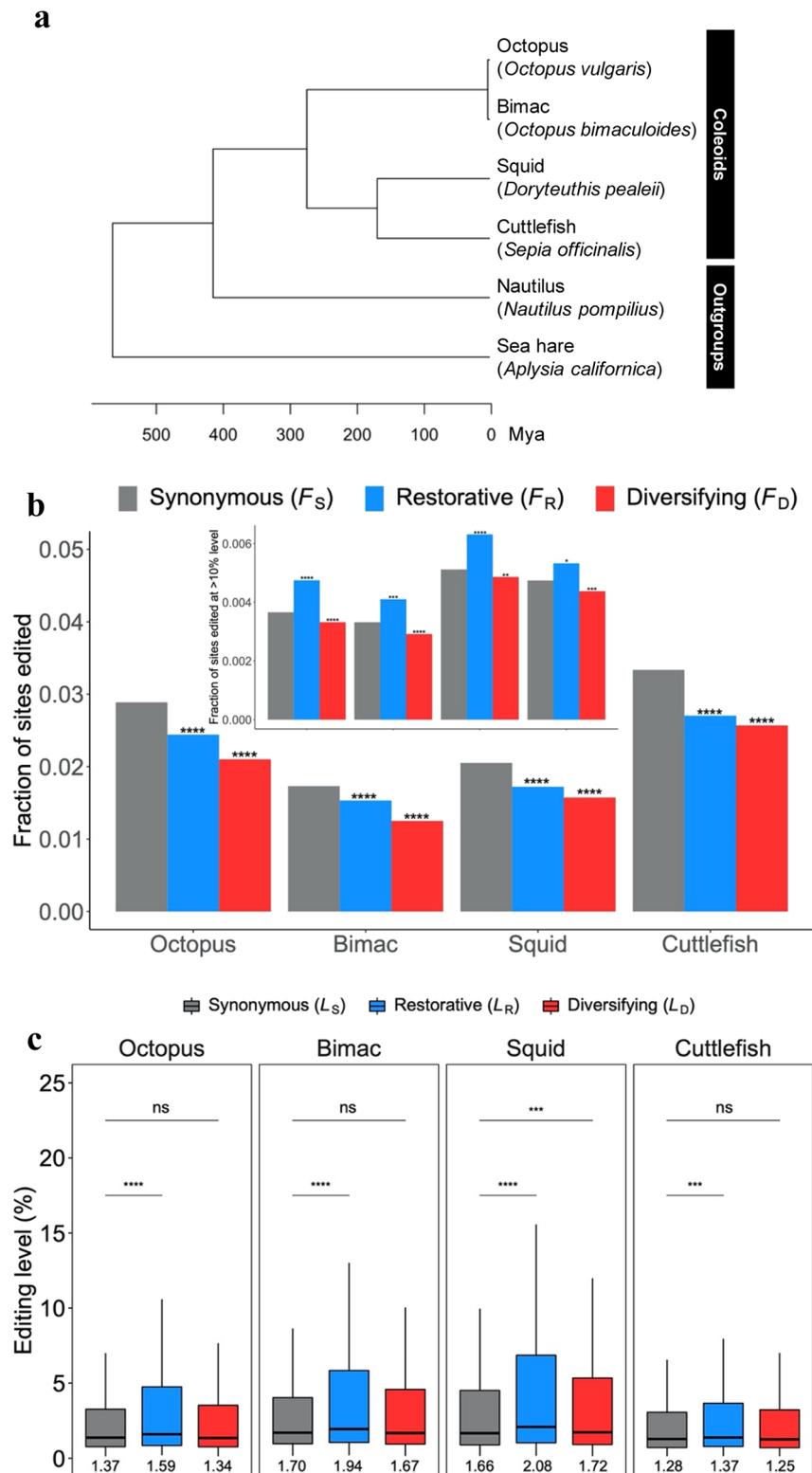


Fig. 2 Comparison of restorative and diversifying editing with synonymous editing in coleoids. **a** The phylogenetic relationship of the six mollusks studied here. Branch lengths represent divergence times based on the mid-points of the divergence time ranges in a previous study²⁷. **b** Frequencies of sites with synonymous (F_S), restorative (F_R), and diversifying (F_D) editing, respectively, in each of the four coleoids. A significant difference between F_S and F_R (or F_D) is indicated by stars above the bin of F_R (or F_D) ($*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$; ns, not significant; χ^2 -test). **c** Synonymous (L_S), restorative (L_R), and diversifying (L_D) editing levels in each of the four coleoids. The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), and the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5 (qu_3 - qu_1)$. The median editing levels are also given below the corresponding boxes. A significant difference between L_S and L_R (or L_D) is indicated by stars ($*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$; ns, not significant; Mann-Whitney U -test). Source data are provided as a Source Data file.

high. Nevertheless, Fig. 2c reveals a larger fraction of restorative editing than synonymous editing with appreciable editing levels. For example, in the squid, 29.14% and 15.65% of restorative editing sites but only 23.03% and 6.82% of synonymous editing sites have editing levels >5% and >20%, respectively. Depending on the harm of the G-to-A mutation and the relative dominance of the A and G isoforms, these appreciable levels of A-to-G editing could substantially increase the fixation probability of the G-to-A mutation. It should also be noted that the harm-permitting hypothesis is proposed as an alternative to the adaptive hypothesis. If moderate levels of nonsynonymous editing could be beneficial as asserted by the adaptive hypothesis, there is no reason why they could not be harm-permitting. Furthermore, the general trend of $L_R > L_S$ and $L_D \sim L_S$ supports the harm-permitting hypothesis relative to the adaptive hypothesis.

To examine the robustness of our results, we conducted four additional analyses. First, we respectively examined editing sites that are specific to each of the four coleoids, because species-specific editing events have similar evolutionary ages, allowing fairer comparisons. The results obtained are highly similar to those in Fig. 2 and are robust to potential misidentifications of species-specific editing (Supplementary Fig. 1). Second, we probed editing events identified from individual tissues in bimac. $F_D < F_S$ holds across tissues, but editing level comparisons as well as comparisons between F_R and F_S are mostly non-significant, likely due to the reduced statistical power as a result of decreased sample sizes (Supplementary Table 2). Third, because editing levels of neighboring editing sites may be co-affected by a mutation, which would reduce the statistical power in comparing synonymous with nonsynonymous editing sites, we compared synonymous editing sites in one half of the gene set with nonsynonymous editing sites in the other half. Specifically, we ranked all genes by the d_N/d_S ratio between octopus and squid orthologs, and respectively grouped genes with odd ranks into bin 1 and those with even ranks into bin 2. We then compared synonymous editing in bin 1 with nonsynonymous editing in bin 2, as well as synonymous editing in bin 2 with nonsynonymous editing in bin 1. The results (Supplementary Fig. 2) are similar to those obtained from all editing sites (Fig. 2). Fourth, we respectively investigated F_R/F_S and F_D/F_S in five editing level ranges (0–20%, 20–40%, 40–60%, 60–80%, and 80–100%) in each coleoid (Supplementary Fig. 3). While both F_R/F_S and F_D/F_S generally increase with the editing level, F_D/F_S is smaller than 1 except when the editing level exceeds 40%. It is important to stress that only a few percent of diversifying editing sites in a coleoid fall in this editing level range (Supplementary Table 3), suggesting that the vast majority of diversifying editing is nonadaptive (see below for quantitative estimates).

Accelerated nonsynonymous G-to-A substitutions. The harm-permitting model further predicts that the rate of nonsynonymous G-to-A substitution relative to that of synonymous G-to-A substitution (d_N/d_S for G-to-A) should be elevated, because the high editing activity renders some otherwise deleterious nonsynonymous G-to-A mutations acceptable. Furthermore, this elevation should be particularly pronounced in genes exclusively expressed in neural tissues but not in genes unexpressed in neural tissues, because the high editing activity is so far observed only in neural tissues^{25,27}. However, because only bimac and squid have available RNA-sequencing data from several non-neural tissues and because genes unexpressed in neural tissues are not in the transcript sequence data of the octopus and cuttlefish, and hence are excluded from our alignments, we had to define two groups of genes with relatively high and relatively low specificities in neural expression, respectively. The genes with high neural expression

specificities are expressed exclusively in neural tissues in the bimac or squid, whereas those with low neural expression specificities are expressed in both neural and non-neural tissues in both the bimac and squid. The harm-permitting model predicts that d_N/d_S for G-to-A is greater for genes with relatively high neural expression specificities than for those of relatively low neural expression specificities. As the harm-permitting effect is present only when a G-to-A mutation at a site is deleterious without editing, we focused on nonsynonymous sites that are conserved in the two outgroup species (i.e., nautilus, sea hare, and the immediately ancestral node of the focal species share the same pre-editing state) to increase the sensitivity of our test. Furthermore, the elevation in d_N/d_S should be specific to G-to-A changes, because the potential harms of other changes such as C/T-to-A and G-to-C/T cannot be alleviated by A-to-G editing.

To this end, we considered all six branches descendent from the common ancestor of the four coleoids. We computed d_N and d_S of each of these branches using the extant and inferred ancestral sequences, and then calculated d_N/d_S by dividing the total d_N by the total d_S of these branches. In support of our prediction, d_N/d_S for G-to-A changes is greater for genes of relatively high neural expression specificities than those of relatively low specificities (Fig. 3). By respectively bootstrapping the two groups of genes 200 times, we found that the above difference is statistically significant ($P < 0.005$). By contrast, no significant difference in d_N/d_S exists between the two groups of genes when C/T-to-A changes or G-to-C/T changes are considered (Fig. 3). It is noteworthy that $d_N/d_S < 1$ in all cases in Fig. 3, consistent with the harm-permitting model that does not involve positive selection.

The potential benefit of shared editing among species. It has been suggested that shared editing among multiple species is likely beneficial, because otherwise the editing status is unlikely to be evolutionarily conserved³⁶. In support of this suggestion was the finding that, even in mammals, where most nonsynonymous editing appears neutral or deleterious, the frequency of conserved sites subject to nonsynonymous editing in both human and

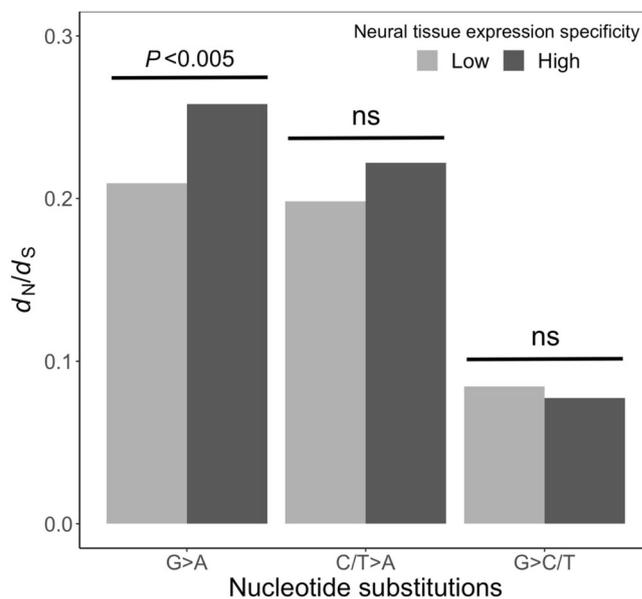


Fig. 3 Coleoid nonsynonymous to synonymous substitution rate ratios (d_N/d_S) for various nucleotide changes. The P -value is based on 200 bootstrap samples; ns, not significant. Source data are provided as a Source Data file.

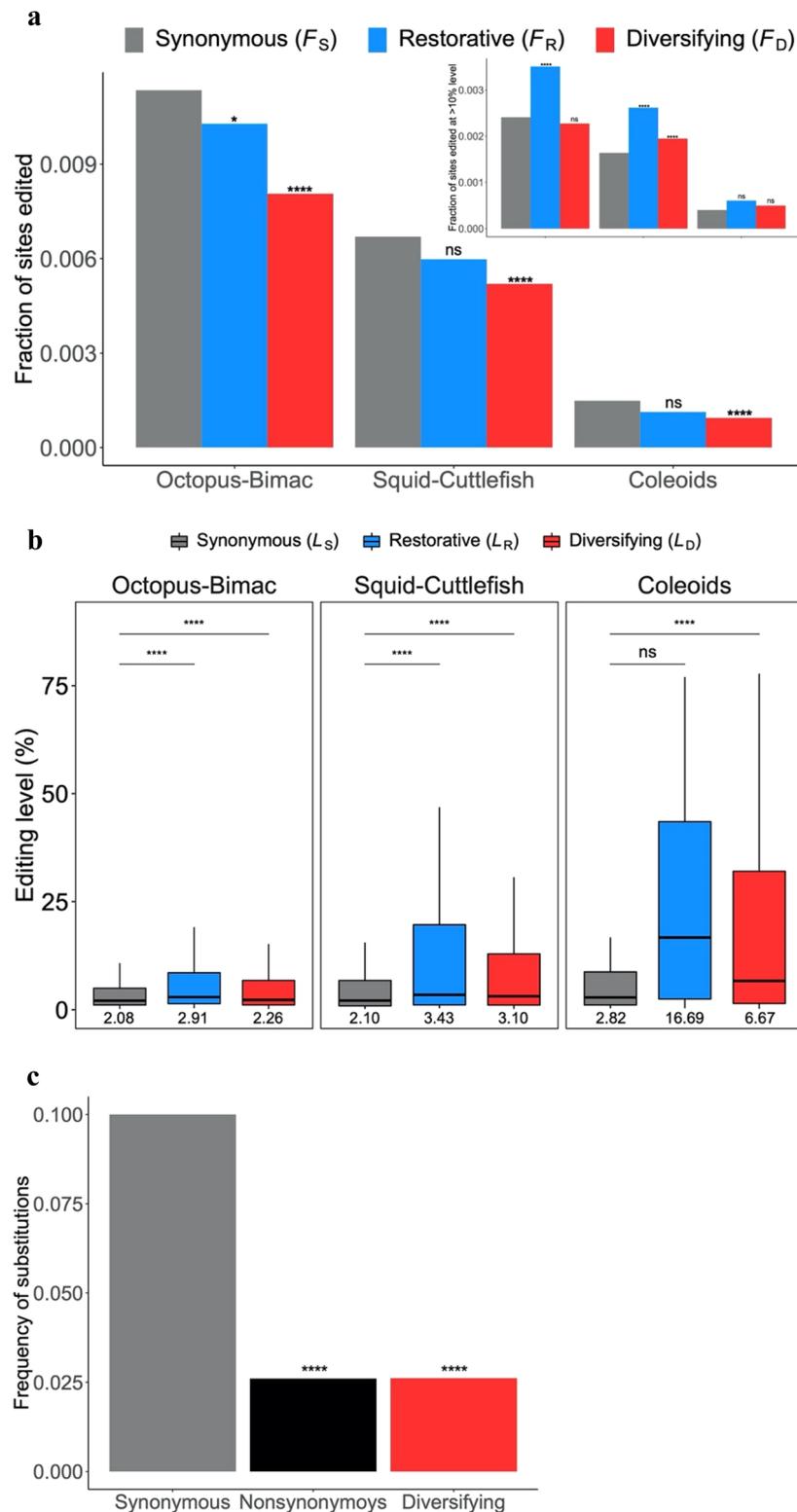


Fig. 4 Patterns of shared editing in coleoids. **a** Frequencies of sites with shared synonymous (F_S), restorative (F_R), and diversifying (F_D) editing, respectively, among sites shared between the octopus and bimac, between the squid and cuttlefish, and among all four coleoids, respectively. A significant difference between F_S and F_R (or F_D) is indicated by stars above the bin of F_R (or F_D) (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001; ns, not significant; χ^2 -test). The inset shows the corresponding fractions of sites with editing levels >10%. **b** Synonymous (L_S), restorative (L_R), and diversifying (L_D) editing levels among the three sets of shared editing sites. The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), and the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3 - qu_1)$. The median editing levels are also given below the corresponding boxes. A significant difference between L_S and L_R (or L_D) is indicated by stars (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001; ns, not significant; Mann-Whitney U -test). **c** Fraction of sites edited in the common ancestor of the four coleoids that have a genomic G in a coleoid. (****, P < 0.0001; chi-squared test). Source data are provided as a Source Data file.

mouse significantly exceeds the frequency of conserved sites subject to synonymous editing in both species³⁶. A similar phenomenon was reported in fruit flies²³. In coleoids, a sizable fraction of nonsynonymous editing is shared by at least two species and highly edited sites tend to be shared²⁷. To understand the potential evolutionary forces maintaining RNA editing at specific sites across multiple coleoids, we analyzed editing shared by a clade of two or more species.

A nonsynonymous editing event shared by a clade of species that modifies the amino acid state from X to Y is considered restorative if the inferred genomic sequence-based amino acid state is Y at any node of the tree that is ancestral to the most recent common ancestor of the clade, or diversifying if Y is not present at any of these ancestral nodes. In the study of shared editing, we considered the average editing level in the clade where the editing is shared. For each of the three groups of editing respectively shared between octopus and bimac, between squid and cuttlefish, and among the four coleoids, both F_R and F_D are lower than F_S (Fig. 4a). However, when the editing level exceeds 10%, F_R is significantly higher than F_S for shared editing between octopus and bimac and that between squid and cuttlefish, and F_D is significantly higher than F_S for shared editing between squid and cuttlefish (Fig. 4a inset). L_R and L_D are both higher than L_S for each group of shared editing (Fig. 4b). A significantly greater F_D than F_S for shared editing could be caused by (i) positive selection promoting the initial fixation of mutations that lead to nonsynonymous editing and/or (ii) purifying selection preventing the loss of presumably beneficial nonsynonymous editing; therefore, it is a clear indicator of adaptive nonsynonymous editing. A significantly greater L_D than L_S for shared editing could be caused by (i) positive selection promoting the increase of editing levels of presumably beneficial nonsynonymous editing, (ii) purifying selection preventing the decrease of editing levels of presumably beneficial nonsynonymous editing, (iii) purifying selection preferentially preventing the loss of high-level nonsynonymous editing presumably because high editing levels are associated with larger benefits than low editing levels, and/or (iv) positive selection preferentially promoting the loss of low-level nonsynonymous editing, probably because an A-to-G substitution is favored at an edited site, especially when the editing level is low. Regardless, a significantly greater L_D over L_S also indicates adaptive nonsynonymous editing. Hence, diversifying editing shared by different coleoids shows adaptive signals, suggesting that a fraction is adaptive.

As most nonsynonymous editing is species-specific (Supplementary Table 1), the above finding is not inconsistent with the analysis of individual species revealing the nonadaptive nature of most editing events. We estimated that, of species-specific diversifying editing sites, 0.15%, 0%, 1.20%, and 0.16% are adaptive in the octopus, bimac, squid, and cuttlefish, respectively (see Methods). Similarly, 3.50%, 3.00%, 14.98%, and 12.26% of shared diversifying editing sites are adaptive in octopus, bimac, squid, and cuttlefish, respectively. Taken together, 1.37%, 1.72%, 4.91%, and 2.18% of diversifying editing sites are adaptive in the four coleoids, respectively.

What is the general benefit of the shared editing that shows adaptive signals? Two hypotheses exist. First, editing may be beneficial because of the intra-organism protein diversity created^{25,27,32,42}. That is, editing allows the existence of two protein isoforms per edited site in an organism, which may confer a higher fitness, analogous to heterozygote advantage at polymorphic sites. Alternatively, editing offers a new isoform that may be simply fitter than the unedited isoform. In this latter hypothesis, the benefit of editing is comparable to that of a nucleotide substitution. To distinguish between these two hypotheses, we focused on sites that are edited in at least three

of the four coleoids, because editing should have existed at these sites in the common ancestor of the four species according to the parsimony principle (Fig. 2a). We then estimated the frequency of replacement of editing with an A-to-G substitution in any of the four species. Such replacements are expected to be more or less neutral for synonymous editing. For nonsynonymous editing, such replacements are deleterious under the first hypothesis due to the loss of protein diversity but are neutral under the second hypothesis. Hence, the first hypothesis predicts a lower frequency of such replacements for nonsynonymous editing than synonymous editing, whereas the second hypothesis predicts equal frequencies of such replacements for synonymous and nonsynonymous editing.

Our analysis showed that the frequency of A-to-G substitutions at nonsynonymous editing sites is significantly lower than that at synonymous editing sites (Fig. 4c; Supplementary Table 4). Because it is the shared diversifying editing for which the nature of the benefit is in question, we restricted the analysis to diversifying editing only, but obtained a similar result (Fig. 4c; Supplementary Table 4). Furthermore, A sites edited in at least three coleoid species are more likely to be edited in all four coleoids when the editing is nonsynonymous than when it is synonymous (36.6% vs. 27.6%, $P = 0.001$, chi-squared test; Supplementary Table 4), suggesting that shared nonsynonymous editing is less likely to be lost than shared synonymous editing. Together, these observations suggest that the benefit of adaptive shared editing is the provision of two protein isoforms per edited site in an organism.

Discussion

The recent discovery of the preponderance of nonsynonymous A-to-G RNA editing among highly edited sites in coleoid neural tissues led to the assertion of widespread adaptive editing in these organisms, but the potential benefits of the editing are unknown. In this work, we proposed an alternative, nonadaptive explanation. Our reanalysis of published transcriptome data from four coleoids and two outgroup species lends strong support to the nonadaptive hypothesis. Combined with previous findings from other species, the new finding suggests a generally nonadaptive nature of coding A-to-G editing among animals. As explained in the harm-permitting model, nonadaptive editing such as some restorative editing, may, however, be selectively protected (middle row in Fig. 1a) or even promoted (bottom row in Fig. 1a). Although such editing events likely originated as molecular errors due to ADARs' limited target specificity, they are no longer errors today. The fact that a nonadaptive feature can nevertheless be under purifying selection or even be positively selected is well known in evolutionary biology^{40,43}.

In the harm-permitting model, A-to-G editing permits the fixation of otherwise deleterious G-to-A mutations and hence the editing is nonadaptive. In theory, it is also possible that A-to-G editing emerged in evolution after a G-to-A substitution at the same site. If the substitution is slightly deleterious, the editing would be slightly beneficial (i.e., compensatory). However, such sites have minimal contributions to F_R and L_R , so this possibility does not alter our interpretation of the nonadaptive nature of restorative editing (see Methods).

The principle of our test of the nonadaptive hypothesis of RNA editing is similar to that of the test of the adaptive hypothesis, except that the new test requires a distinction between restorative and diversifying editing, which in turn depends on ancestral coding sequences inferred for the interior nodes of a phylogeny (Fig. 1b, c). Although ancestral sequence inference is generally reliable, it is not expected to be 100% correct⁴⁴. Will errors and potential biases in this inference bias our test? The answer is no.

F_R is the number of edited sites with an ancestral nonsynonymous G-to-A substitution divided by the total number of sites with an ancestral nonsynonymous G-to-A substitution. As our ancestral sequence inference is based on genomic sequences and is blind to RNA editing, any potential bias in estimating the number of sites with an ancestral G-to-A substitution is cancelled out in computing F_R . The same applies to F_D . Errors and potential biases in ancestral sequence inference only increase the stochastic errors of F_R and F_D estimates, reducing the statistical power in testing our hypothesis. Notwithstanding, the vast majority of our key statistical tests yielded significant results, suggesting that sufficient statistical power remains in these tests.

Although our study explains the preponderance of nonsynonymous editing in coleoids, we have not addressed a related question—why the editing activity was drastically elevated in neural tissues during coleoid evolution. A substantial rise in editing activity is expected to be harmful, because its effect is similar to inducing A-to-G mutations. Indeed, expression of the human *ADAR2* gene in the budding yeast *Saccharomyces cerevisiae*, which does not naturally possess any ADAR gene, inhibits yeast growth because of ADAR2's RNA editing activity⁴⁵. Our observation of a significantly lower F_D than F_S in every coleoid examined (Fig. 2b) strongly suggests that diversifying editing is generally deleterious and has been selectively purged. Hence, it is almost certain that the pervasive coding RNA editing was not the reason for the elevation of the editing activity in coleoids but its byproduct. Whatever the reason was, the relevant benefit must at least offset the harm from pervasive nonsynonymous editing, under the assumption that the evolutionary elevation of the editing activity was not due to genetic drift alone, because the population size of ancestral coleoids was probably not small. It is worth mentioning that a number of physiological functions have been proposed for A-to-G editing, including suppressing the proliferation of transposons⁴⁶, inhibiting viral replication⁴⁷, marking RNAs for degradation³², marking RNAs to prevent innate immunity against self-RNAs^{48,49}, regulating alternative splicing³², and modulating nuclear retention of RNAs³². As the primary physiological function of A-to-G editing is unknown, it is difficult to discern why the editing activity rose drastically in coleoids.

Similar to previous findings in mammals and flies^{23,36}, we observed some adaptive signals from nonsynonymous editing shared between species. Our additional analysis suggests that the benefit of these adaptive editing events lies in the protein diversity brought by editing. While this finding supports the prevailing view on why coding RNA editing may be adaptive, it is important to stress that, based on our estimation, only about 2.5% of nonsynonymous editing appears adaptive in an average coleoid.

Liscovitch-Brauer and colleagues²⁷ noted that flanking regions of sites edited in multiple species tend to be evolutionarily conserved and asserted that coleoids “use extensive RNA editing to diversify their neural proteome at the cost of limiting genomic sequence flexibility and evolution.” However, we believe that the observation prompting such a conclusion is caused by an ascertainment bias. Specifically, because of the various requirements for a site to be edited, such as specific flanking sequences²⁷ and secondary structures⁵⁰, a shared editing site by definition satisfies these requirements in its neighborhood in multiple species. Thus, the site is expected to show a higher interspecific similarity in flanking sequences than a randomly picked site, regardless of whether the editing is shared because of selective constraints or not. The same ascertainment bias occurs in the comparison of intraspecific polymorphisms of flanking sequences between shared editing sites and random sites. In particular, given the flanking sequence requirement for editing, an edited site with a lower flanking sequence polymorphism is expected to be edited in

a greater percentage of individuals in the species. Hence, provided that a site is found to be edited in multiple species when only one individual is examined per species, the polymorphism is expected to be low irrespective of the presence/absence of selective constraints on the editing.

The nonadaptive hypothesis we proposed is based on the harm-permitting effect of high levels of editing, which inflates the frequency and level of restorative editing, relative to those of synonymous editing. As previous comparisons of synonymous and nonsynonymous editing in non-coleoid species never considered this effect, one wonders whether their conclusions are still valid. Ignoring the harm-permitting effect renders conclusions of nonadaptive editing more conservative. Hence, such conclusions should still hold. For claims of adaptive editing that are based on comparisons between synonymous and nonsynonymous editing frequencies and levels, a reanalysis taking into account the harm-permitting effect is warranted. In other words, a significantly greater F_D than F_S and/or a significantly greater L_D than L_S are required to demonstrate positive selection promoting nonsynonymous editing. This is especially true to the group of fungi that show pervasive A-to-G editing as in coleoids^{51–53}.

It is worth mentioning that transcriptome-wide analyses of several other types of RNA editing such as C-to-U editing⁵⁴ and m⁶A modification (methylation of A at the nitrogen-6 position)⁵⁵ also suggest that most editing events are nonadaptive. In addition, variations in several steps of RNA production and processing such as alternative transcriptional initiation⁵⁶, alternative splicing⁵⁷, and alternative polyadenylation⁵⁸ have been shown to be largely molecular errors. Similarly, it is plausible that variations in the translational process such as stop-codon read-through⁵⁹ and events of posttranslational modifications such as phosphorylation⁶⁰ and glycosylation⁶¹ are primarily manifestations of molecular errors. Whether it is generally true that phenotypic variations at the molecular level are less likely to be adaptive than those at the cellular, tissue, organ, and organismal levels is worth exploration⁶².

Methods

Transcriptomes, editing sites, and ancestral sequences. The transcriptomes of six mollusk species and the list of A-to-I editing sites in the four coleoid species were previously published²⁷. We extracted coding sequences from the previously assembled transcriptomes²⁷ on the basis of the annotations in the data set and converted Gs in the sequences to As at edited sites. In some genes, we observed stop codons occurring upstream of the last three nucleotides of the annotated coding sequence, possibly due to erroneous inclusions of 3'-untranslated regions. We therefore removed nucleotides downstream of the first stop codon in these sequences. All but one A-to-G editing site in the data are upstream of the first stop codons, suggesting that these annotation errors barely influenced the previous analysis of RNA editing. If a gene appeared more than once in the original dataset for a species, only the longest sequence was retained in our analyses.

Orthologous genes among the six mollusks were previously identified²⁷ and a total of 3979 genes have orthologs in all 6 species and contain at least 1 A-to-G editing site in at least 1 coleoid. We first made a protein sequence alignment of orthologous sequences using MAFFT⁶³ and then generated a coding sequence alignment of these genes using PAL2NAL⁶⁴. Ancestral sequences were inferred using the codeml program in PAML⁶⁵ under default parameters without excluding gap sites and the best joint inferences of all interior nodes were used in subsequent analyses. The unrooted topology of the tree in Fig. 2a was used in ancestral sequence inference. Subsequent analyses used in-house Perl scripts.

All reported editing sites in the 3979 genes²⁷ were included in our analyses, unless otherwise noted. Although some editing sites may be sequencing errors, the probability of error is expected to be low given the tiny amount of other types of DNA–RNA mismatches observed²⁷.

Restorative and diversifying editing. The tree in Fig. 2a shows three interior nodes ancestral to each coleoid species. A coding A site in a coleoid is considered a potential site for restorative editing if changing the A to G is nonsynonymous and if the corresponding amino acid after the change becomes identical to the amino acid state at any one of the three ancestral nodes. A potential site for restorative editing becomes a restorative editing site if it is edited in the focal species. By definition, F_R is the number of sites with restorative editing divided by the number

of potential sites for restorative editing, whereas L_R is the median editing level at restorative editing sites. A coding A site in a focal species is considered a potential site for diversifying editing if changing the A to G is nonsynonymous and if the corresponding amino acid after the change differs from all amino acid states of the three ancestral nodes. A potential site for diversifying editing becomes a diversifying editing site if it is edited in the focal species. By definition, F_D is the number of sites with diversifying editing divided by the number of potential sites for diversifying editing, whereas L_D is the median editing level at diversifying editing sites. F_S is the number of sites with synonymous editing divided by the number of A sites where A-to-G editing would be synonymous, whereas L_S is the median editing level at synonymous editing sites³⁴. Although the comparison between F_R (or F_D) and F_S , and that between L_R (or L_D) and L_S are not entirely independent from each other, each comparison is fair.

An editing event is considered to be shared by a clade of two or more species if the event occurs in all species of the clade in the tree of Fig. 2a and all of these species have the same pre- and post-editing amino acid states. In studying shared editing by a clade, we followed the above procedure in distinguishing restorative from diversifying editing, except that we considered all interior nodes ancestral to the most recent common ancestor of the clade instead of all interior nodes ancestral to one species.

Comparing median editing levels. When the mRNA concentration is low, RNA editing cannot be detected unless the editing level is sufficiently high. This bias would make the median editing level appear higher in weakly expressed genes than strongly expressed genes even when no such difference actually exists. To alleviate this bias, we considered only those sites that are covered by at least 400 RNA-sequencing reads when comparing median editing levels. Nevertheless, the bias does not affect the comparison between synonymous and nonsynonymous editing, because their detectabilities are equally influenced by the gene expression level. For a shared editing site, the average editing level and average read number of all species in the focal clade are used to represent the site. We did not apply editing level cutoffs in the comparison of editing levels of different sites due to potential biases that may arise.

Proportion of diversifying editing that is adaptive. Under the presumption that the excess of F_D over F_S represents adaptive editing, we calculated F_D and F_S in each of 10 editing level intervals (0–10%, 10–20%, till 90–100%). For each interval exhibiting $F_D > F_S$, the number of adaptive diversifying editing sites equals $ADP = N_D(1 - F_S/F_D)$, where N_D is the number of diversifying editing sites in the interval. Summing up these ADP numbers yields the total number of diversifying editing sites that are adaptive.

Contributions of compensatory editing to F_R and L_R . In the harm-permitting model, A-to-G editing permits the fixation of otherwise deleterious G-to-A mutations and hence the editing is nonadaptive. In theory, it is also possible that A-to-G editing emerged in evolution after a G-to-A substitution at the same site. If the substitution is slightly deleterious, the editing would be slightly beneficial (i.e., compensatory). For several reasons, such sites should contribute minimally to F_R and L_R . First, the probability that the G-to-A substitution occurred in the most recent common ancestor of cephalopods (the top five species in Fig. 2a) or more recently is small, because it could occur at any time prior to the emergence of the editing at the site, which most likely took place when the cellular editing activity rose substantially in the branch immediately preceding the common ancestor of coleoids. Hence, the probability that the editing is classified as restorative is small and such compensatory events are unlikely to affect our analysis of restorative editing sites. Although such compensatory events are potentially included in diversifying editing sites we analyzed, diversifying editing still show lower editing frequencies and editing levels when compared with synonymous editing. Thus, our interpretation that diversifying editing is overall under purifying selection remains valid. Furthermore, even for the minority of compensatory events that are classified as restorative, the impact is small. This is because deleterious G-to-A mutations that could get fixed without editing are presumably only slightly deleterious. Hence, the benefit of A-to-G editing at such sites is also presumably small such that their editing level may not be selectively raised or selectively maintained at high levels. More importantly, there will be a comparable number of slightly beneficial G-to-A substitutions followed by slightly deleterious A-to-G editing that are included in the category of restorative editing. The effects of these two groups of events are likely cancelled out.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The original data were previously published. All new results are in the manuscript including Supplementary Materials. The source data underlying Figs. 2–4 and Supplementary Figs. 1–3 are provided as a Source Data file. Intermediate results are available from the authors.

Code availability

Custom code is available from the authors.

Received: 1 October 2018; Accepted: 31 October 2019;

Published online: 27 November 2019

References

- Gott, J. M. & Emeson, R. B. Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* **34**, 499–531 (2000).
- Farajollahi, S. & Maas, S. Molecular diversity through RNA editing: a balancing act. *Trends Genet.* **26**, 221–230 (2010).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
- Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: form, distribution, and function. *Science* **352**, 1408–1412 (2016).
- Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* **14**, 23–31 (2016).
- Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339–346 (2017).
- Yablonovitch, A. L., Deng, P., Jacobson, D. & Li, J. B. The evolution and adaptation of A-to-I RNA editing. *PLoS Genet.* **13**, e1007064 (2017).
- Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing - immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
- Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
- Porath, H. T., Knisbacher, B. A., Eisenberg, E. & Levanon, E. Y. Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biol.* **18**, 185 (2017).
- Hung, L. Y. et al. An evolutionary landscape of A-to-I RNA editome across metazoan species. *Genome Biol. Evol.* **10**, 521–537 (2018).
- Bahn, J. H. et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
- Danecek, P. et al. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.* **13**, 26 (2012).
- Peng, Z. Y. et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **30**, 253–25 (2012).
- Bazak, L. et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).
- Chen, J. Y. et al. RNA editome in rhesus macaque shaped by purifying selection. *PLoS Genet.* **10**, e1004274 (2014).
- Li, J. B. et al. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
- Graveley, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
- Ramaswami, G. et al. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **10**, 128–132 (2013).
- Rodriguez, J., Menet, J. S. & Rosbash, M. Nascent-seq indicates widespread cotranscriptional RNA editing in *Drosophila*. *Mol. Cell* **47**, 27–37 (2012).
- Duan, Y., Dou, S., Luo, S., Zhang, H. & Lu, J. Adaptation of A-to-I RNA editing in *Drosophila*. *PLoS Genet.* **13**, e1006648 (2017).
- Zhang, R., Deng, P., Jacobson, D. & Li, J. B. Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. *PLoS Genet.* **13**, e1006563 (2017).
- Yu, Y. et al. The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. *PLoS Genet.* **12**, e1006191 (2016).
- Albertin, C. B. et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–224 (2015).
- Alon, S. et al. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *Elife* **4**, e05198 (2015).
- Garrett, S. & Rosenthal, J. J. C. RNA editing underlies temperature adaptation in K⁺ channels from polar octopuses. *Science* **335**, 848–851 (2012).
- Liscovitch-Brauer, N. et al. Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell* **169**, 191–202 (2017).
- Deffit, S. N. et al. The *C. elegans* neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife* **6**, e28625 (2017).
- Zhao, H. Q. et al. Profiling the RNA editomes of wild-type *C. elegans* and ADAR mutants. *Genome Res.* **25**, 66–75 (2015).
- Porath, H. T. et al. A-to-I RNA editing in the earliest-diverging eumetazoan phyla. *Mol. Biol. Evol.* **34**, 1890–1901 (2017).
- Covello, P. S. & Gray, M. W. On the evolution of RNA editing. *Trends Genet.* **9**, 265–268 (1993).
- Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).

33. Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. A-to-I RNA editing and human disease. *RNA Biol.* **3**, 1–9 (2006).
34. Xu, G. & Zhang, J. Human coding RNA editing is generally nonadaptive. *Proc. Natl Acad. Sci. USA* **111**, 3769–3774 (2014).
35. Pinto, Y., Cohen, H. Y. & Levanon, E. Y. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol.* **15**, R5 (2014).
36. Xu, G. & Zhang, J. In search of beneficial coding RNA editing. *Mol. Biol. Evol.* **32**, 536–541 (2015).
37. Rosenthal, J. J. C. & Seeburg, P. H. A-to-I RNA editing: Effects on proteins key to neural excitability. *Neuron* **74**, 432–439 (2012).
38. Klinger, C. M. et al. Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. *Genome Biol. Evol.* **10**, 1019–1038 (2018).
39. Tian, N., Wu, X. J., Zhang, Y. Z. & Jin, Y. F. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. *RNA* **14**, 211–216 (2008).
40. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
41. Sloan, D. B. Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. *Biol. Lett.* **13**, pii: 20170314 (2017).
42. Gommans, W. M., Mullen, S. P. & Maas, S. RNA editing: a driving force for adaptive evolution? *Bioessays* **31**, 1137–1145 (2009).
43. Hartl, D. L. & Taubes, C. H. Compensatory nearly neutral mutations: selection without adaptation. *J. Theor. Biol.* **182**, 303–309 (1996).
44. Zhang, J. & Nei, M. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **44** (Suppl 1), S139–S146 (1997).
45. Eifler, T., Pokharel, S. & Beal, P. A. RNA-Seq analysis identifies a novel set of editing substrates for human ADAR2 present in *Saccharomyces cerevisiae*. *Biochemistry* **52**, 7857–7869 (2013).
46. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
47. Taylor, D. R., Puig, M., Darnell, M. E., Mihalik, K. & Feinstone, S. M. New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *J. Virol.* **79**, 6291–6298 (2005).
48. Chung, H. et al. Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **172**, 811–824 (2018).
49. Liddicoat, B. J. et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120 (2015).
50. Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319 (2011).
51. Liu, H. Q. et al. A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **114**, E7756–E7765 (2017).
52. Liu, H. Q. et al. Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Res.* **26**, 499–509 (2016).
53. Teichert, I., Dahlmann, T. A., Kuck, U. & Nowrousian, M. RNA editing during sexual development occurs in distantly related filamentous ascomycetes. *Genome Biol. Evol.* **9**, 855–868 (2017).
54. Liu, Z. & Zhang, J. Human C-to-U coding RNA editing is largely nonadaptive. *Mol. Biol. Evol.* **35**, 963–969 (2018).
55. Liu, Z. & Zhang, J. Z. Most m⁶A RNA modifications in protein-coding regions are evolutionarily unconserved and likely nonfunctional. *Mol. Biol. Evol.* **35**, 666–675 (2018).
56. Xu, C., Park, J. K. & Zhang, J. Evidence that alternative transcriptional initiation is largely nonadaptive. *PLoS Biol.* **17**, e3000197 (2019).
57. Saudemont, B. et al. The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome Biol.* **18**, 208 (2017).
58. Xu, C. & Zhang, J. Alternative polyadenylation of mammalian transcripts is generally deleterious, not adaptive. *Cell Syst.* **6**, 734–742 (2018).
59. Li, C. & Zhang, J. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. *PLoS Genet.* **15**, e1008141 (2019).
60. Landry, C. R., Levy, E. D. & Michnick, S. W. Weak functional constraints on phosphoproteomes. *Trends Genet.* **25**, 193–197 (2009).
61. Park, C. & Zhang, J. Genome-wide evolutionary conservation of N-glycosylation sites. *Mol. Biol. Evol.* **28**, 2351–2357 (2011).
62. Zhang, J. Neutral theory and phenotypic evolution. *Mol. Biol. Evol.* **35**, 1327–1331 (2018).
63. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**, 772–780 (2013).
64. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
65. Yang, Z. H. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

Acknowledgements

We thank Zhen Liu and members of the Zhang lab for valuable comments. This work was supported by the U.S. National Institutes of Health research grant GM120093 to J.Z.

Author contributions

J.Z. conceived the project and secured funding. D.J. compiled and analyzed the data. D.J. and J.Z. designed the analyses and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13275-2>.

Correspondence and requests for materials should be addressed to J.Z.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019, corrected publication 2021