BIOLOGICAL SCIENCES: Agricultural Sciences

# RNA editing fine-tunes transcriptional plasticity and enhances the adaptability of *Myzus persicae* to diverse host plants

Wenyuan Yu [a,b], Jun Wu [a,b], Rong Hu [a,b], Zhimou Lin [a,b], Peiyu Yang [a,b], Zhichao Hu [a,b], Pengshuai Peng [a,b], Shuai Zhan [c], Jean-Christophe Simon[d], Saskia A. Hogenhout [e], Gang Wu [b, #], Shuangxia Jin [a,f, #], Yazhou Chen [a,b, #]

[a] Hubei Hongshan Laboratory, Wuhan, 430070, China;

[b] Hubei Insect Resources Utilization and Sustainable Pest Management Key Laboratory, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China;

[c] Key Laboratory of Plant Design, Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China;

[d] INRAE (National Institute of Agriculture, Food and Environment), UMR IGEPP, Le Rheu, France;

[e] Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK.

[f] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei 430070, P. R. China

[#] Correspondence: Yazhou Chen, Shuangxia Jin, Gang Wu

Email: yzchen@mail.hzau.edu.cn; jsx@mail.hzau.edu.cn; wugang@mail.hzau.edu.cn

**Author Contributions:**

Y.C., S.J., and G.W. conceived and designed the study, supervised the project, and finalized the manuscript. W.Y. performed the majority of the experimental work, and Y.C. and W.Y. jointly conducted the primary data analyses. J.W., Z.L., and P.P. contributed to the experimental setup and bioassays. Bioinformatic analyses were collaboratively carried out by Y.C., S.Z., W.Y., and R.H. P.Y. and Z.H. contributed to the execution of the research. The manuscript was drafted by Y.C. and W.Y., with W.Y. also managing the next-generation sequencing data. S.H., J.-C.S., G.W., S.J.,

and S.Z. provided academic consultation and contributed to manuscript revision. All authors reviewed the manuscript, provided feedback, and approved the final version.

**Competing Interest Statement:** The authors declare no competing interests.

**Data deposition**: All data reported in this paper were deposited in the NCBI BioProject under project numbers PRJNA1211840 and PRJNA1344291. All data are available in the main text or the supplementary materials.

**This PDF file includes:**

Main Text

Figures 1 to 5

**Abstract**

Parthenogenetic organisms can adapt to diverse and rapidly changing environments despite limited standing genetic variation, but the molecular mechanisms enabling such adaptability remain unclear. Here, we investigate this question in the green peach aphid (*Myzus persicae*, GPA), an asexual generalist herbivore capable of colonizing a wide range of host plants without underlying genetic change. By integrating RNA-seq and DNA-seq from genetically identical clones reared on different hosts, we identified 1,368 high-confidence RNA editing sites (RES), many of which exhibited strong host-dependent differences in editing levels. These RES encompassed all 12 possible nucleotide substitutions, with A-to-I, C-to-U, and U-to-C conversions being particularly prevalent. Editing at several sites shifted dynamically within 48 hours of host transfer and often coincided with changes in transcript abundance. Functional knockdown of the editing genes *ADAR1* and *ADAR2* reduced editing at specific sites, diminished fecundity, and abolished aphid survival on new hosts. Comparative analyses across aphid species reveal that although the biochemical signatures of A-to-I editing are deeply conserved, the genomic locations and functional roles of edited transcripts are species specific, reflecting lineage-specific adaptive strategies. Together, our findings demonstrate that RNA editing provides a rapid, reversible mechanism for environmental acclimation in GPA, enabling adaptive phenotypic plasticity even in the absence of genetic variation. This work establishes RNA editing as a key molecular basis for the remarkable ecological generalism of clonal organisms.

**Keywords:** herbivorous insects, RNA editing, host adaptation, transcriptional plasticity, generalism

## Significance statement

A major question in evolutionary biology is how clonal organisms with little genetic variation rapidly adapt to diverse, changing environments. Here we show that RNA editing provides a mechanism enabling the parthenogenetic green peach aphid (*Myzus persicae*) to colonize a wide range of host plants. RNA editing generates transcriptomic diversity, responds quickly to host shifts, and is essential for adaptation, as silencing ADAR genes disrupts A-to-I editing, reduces fitness, and prevents survival on new hosts. Comparative analyses across aphid species reveal that although the biochemical features of A-to-I editing are conserved, its genomic deployment and functional outcomes differ among species, reflecting lineage-specific strategies. These findings demonstrate that RNA editing creates a dynamic regulatory layer that offsets constraints of asexual reproduction.

## Main Text

## Introduction

Standing genetic variation fuels adaptation, as preexisting mutations that change phenotypes can be selected for and spread through the population when advantageous (1). However, whether genetic mutations alone are sufficient for an organism to adapt to rapidly changing environments remains unclear. In particular, this concept is challenged by parthenogenetic or clonal organisms that are unlikely to accumulate favorable mutations in a short period but can plastically produce alternative phenotypes in response to environmental change.

Aphids are small sap-feeding insects of the order Hemiptera and are highly specialized plant feeders that often transmit plant pathogens, especially viruses (2). Aphid life cycles typically encompass alternation of sexual and parthenogenetic stages. Sexual reproduction promotes recombination and gene shuffling, ultimately enriching genetic variation (3). However, many aphid lineages have lost the sexual phase in their life cycle and reproduce parthenogenically (4). Under conditions such as the absence of winter hosts on which sexual reproduction occurs and/or living in warm regions, parthenogenesis becomes the sole reproductive strategy year-round. Aphids produced parthenogenetically are genetically identical across the vast majority of the genomes (4, 5), and consequently, genetic variation in asexual populations is extremely low (6). Nevertheless, asexual aphids can rapidly adapt to harsh environments in both laboratory and field settings (7–10), for example, by acquiring resistance to insecticides (11) or adapting to resistant host plants (12). The mechanisms underlying such adaptability, likely independent of preexisting genetic variation, remain largely unknown.

In addition to genetic variations, nucleotide variations in RNA molecules generated by RNA editing have increasingly been recognized to promote insect adaptation (13, 14). RNA editing in insects contributes to caste determination (15), task performance (16), wing dimorphism (17), temperature adjustment (18, 19), and insecticide resistance (20–23). RNA editing modifies nucleotides in RNA molecules, leading to a significant diversity of transcripts in an environment-responsive manner. The conversion of one type of nucleotide to any of the other three types theoretically produces 12 different editing types in primary RNA transcripts (24). Consequently, editing results in amino acid changes in the encoded proteins (25), altered

secondary structure of RNAs (26), replaced splice consensus sites (22, 27), and modified RNA fate (28). The most common editing event in insects is the deamination of adenosine into inosine (A>I) by Adenosine Deaminase Acting on RNA (ADAR) (26, 29). Inosines (Is) are recognized as guanosines (Gs) by the translational machinery and the reverse transcriptase used in protocols of RT-PCR and RNA-sequencing, resulting in A-to-G (A>G) substitution (30). C-to-U (C>U) editing mediated by cytidine deaminase Apolipoprotein B mRNA Editing Enzyme, Catalytic Polypeptide 1 (APOBEC-1), U-to-C (U>C) (31, 32), and G-to-A (G>A) by *trans*-amination (33) are less frequently observed in insects so far examined (24).

To investigate the RNA editing in aphids, we focused on the green peach aphid (GPA), *Myzus persicae*, a highly polyphagous generalist capable of colonizing a broad range of host plants and transmitting more than 100 plant viruses (34). GPA has gained resistance to most classes of insecticide, making it one of the most widely and strongly resistant insect species worldwide (11, 35, 36). GPA has both sexual and parthenogenetic stages in the life cycle. Sexual reproduction occurs in autumn on peach trees, and the parthenogenetic generations feed on a wide number of secondary host species, including many economically important crops. In many areas where peach trees are absent or the climate is warm, the life cycle is simplified to continual parthenogenesis throughout the year, leading to rapidly increasing population and escalating field damage. Unlike some herbivorous insects that are generalists at the species level but show plant-based population differentiation, GPA is a true generalist even at the level of a single clone, capable of shifting between highly divergent plant hosts (37–40). Previously, we reported that GPA Clone O, which descended from a single asexual female, successfully colonized nine hosts from five different plant families (41). Individuals that feed on one type of host can also survive and reproduce on distantly related host plants (41). These findings demonstrate that GPA can colonize diverse host plant species in the absence of genetic variation.

In this study, we investigated the RNA editing events in the transcriptome of GPA and found that GPA was capable of diverse editing events. Many RNA editing sites were responsive to host change. The GPA genome encoded two copies of *ADAR*s, which were responsible for A-to-I editing, while lacking AID/APOBEC deaminases. RNAi of each of the *ADAR*s resulted in a significant reduction in A-to-I editing levels and led to a reduced survival rate of GPA on host plants. These

findings demonstrate that RNA editing fine-tunes transcriptomic plasticity and enables rapid, reversible adaptation to environmental change, even in the absence of genetic variation.

## Results

### RNA single-nucleotide variants in *M. persicae* vary according to host plants

Previously, we established stable colonies on nine plant species using colonies of *M. persicae* Clone O that derived from a single female. RNA-seq analysis of *M. persicae* on these hosts revealed host-specific patterns of coordinated gene expression (37). Remarkably, numerous single-nucleotide variants (SNVs) were detected in the RNA-seq reads aligned to the chromosome-level genome of Clone O (Fig. 1A).

To investigate the origin of these SNVs, we compared the SNVs to those identified in DNA-seq data from aphids maintained on *Brassica rapa* (Br) before (2016) and after (2018) the nine-host experiment conducted in 2017 (37, 42). 78.7%–88.4% of SNVs in the RNA-seq corresponded to genomic mutations (Figure 1A). After subtracting the genomic SNVs, a significant number of SNVs remained in RNA-seq data from aphids on Br and on the other eight hosts (Figure 1A), indicating the presence of nucleotide mismatches in RNA relative to genomic DNA of *M. persicae*.

To elaborate on the occurrence of the RNA-specific SNVs in this species, we established stable colonies of *M. persicae* Clone Xibei (descended from a single female) on four different hosts, Bn (*B. napus*), At (*Arabidopsis thaliana*), Ta (*Triticum aestivum*), and Zm (*Zea mays*). Those aphid lines were maintained on corresponding plants for more than a year. Three aphid samples on each host were harvested for RNA-seq, and two additional samples for DNA-seq. Reads were aligned to the reference genome of Clone O. PCA of SNVs revealed that lines of Clone Xibei maintained on different host plants clustered closely (Fig. 1B), indicating that their genomes remained largely identical. Similarly, lines of Clone O, which were maintained on *B. napus* for over two years, exhibited few differences (Fig. 1B). After excluding DNA-based SNVs, thousands of RNA-specific SNVs were detected across the lines (Fig. 1C). This pattern was consistent across both Clone O and Clone Xibei (Fig. 1A and 1C), highlighting the presence of RNA-specific nucleotide variations in aphid lines of the same clone maintained on distinct host plants.

## RNA editing in *M. persicae* is diverse

SNVs in RNA that are independent of genomic variation may arise from RNA editing (43–45). To explore RNA editing in *M. persicae*, we employed four bioinformatic tools—REDItools2 (46), JACUSA2 (47), BCFtools (48), and SPRINT (49)—to identify RNA-editing sites (RES) by comparing RNA-seq data from Xibei clones with corresponding DNA-seq data. Strand-specific RNA-seq reads were aligned to the reference genome, and single-nucleotide mismatches between RNA and DNA were detected. SNVs present in the DNA samples were excluded to eliminate potential genomic mutations that could mimic RNA editing.

To ensure high-confidence identification of RNA editing events, we applied stringent filtering criteria (Fig. 1D). Candidate sites were required to (i) contain at least 10 uniquely mapped reads, including ≥5 edited reads, (ii) show an editing level between 5% and 95%, (iii) be consistently detected across all three biological replicates, and (iv) be called by all four prediction tools. Using these criteria, we identified 1,368 high-confidence RNA editing sites (Fig. 1E; Dataset S1). Of these, 951 sites (69.5%) were located in single-copy genes, 21.5% mapped to multi-copy gene regions, and the remaining 9.0% were found in intergenic regions (Fig. S1).

The identified RNA editing events were then categorized into transitions and transversions based on the type of nucleotide change (Fig. 1F). Transition events accounted for 71.6% of all editing sites and included A-to-I (18.4%), C-to-U (20.1%), G-to-A (18.4%), and U-to-C (14.6%) (Fig. 1F). Transversions comprised the remaining 28.4%, with A-to-U (8.6%) and U-to-A (6.7%) being the most frequent, while other transversion types occurred at lower frequencies, each representing less than 11% of the total (Fig. 1F). Independent analyses using the four software tools yielded highly consistent substitution patterns and relative frequencies (Fig. S2). While A-to-I editing is a common editing event in many organisms, the prevalence of other types of RNA editing in *M. persicae* is notable and suggests a distinct editing repertoire compared to other species (14, 17, 19, 24).

Analysis of the genomic distribution of RES revealed significant differences between editing types. Most RNA editing sites were located in coding sequences (CDS), and the impact of editing varied between transitions and transversions (Fig. 1G). Within the CDS, both transition and transversion edits generated a substantial number of nonsynonymous substitutions (Fig. 1H). The RNA editing profiles

observed in *M. persicae* Clone O were further supported by high-depth RNA-seq data (37), confirming the robustness of the identified editing patterns (Fig. S3). Functional enrichment analysis revealed that CDS-resident RNA editing events in *M. persicae* were associated with cell adhesion and tissue morphogenesis, phosphorylation-dependent signal transduction, synaptic and neuromuscular structural organization, and extracellular matrix function (Fig. S4), suggesting that RNA editing may act as a key regulatory mechanism modulating developmental plasticity and host adaptation in *M. persicae*.

**RES are differentially edited in GPA on different host plants**

To investigate the potential role of RNA editing in host adaptation by *M. persicae*, we identified RES that were differentially edited (DE) in response to different host plants. Editing levels for RES were obtained by Reditools and compared between lines of Clone Xibei maintained on four host plants, At, Zm, Ta, Bn. A total of 154 RESs were found to be DE among these hosts (Turkey test, log2FC ≥ 2, $P < 0.05$; Fig. S5A, Dataset S2). Of these, 53 RES were DE between aphids on At and Bn, 54 between those on Ta and Bn, and 99 between Zm and Bn (Fig. S5A, Dataset S2). A heatmap of editing levels indicated that RNA editing events were host-responsive, suggesting a potential role for RNA editing in enabling *M. persicae* to colonize different plant species (Fig. 2A).

To validate these RES, PCR amplification and sequencing were performed using specific primers targeting the identified RES (Fig. 2B). Among the 154 RESs (Dataset S2), 71 sites were amenable for primer design (Table S1). Primers covered RES sites were consistently and specifically amplified from cDNA and gDNA samples of aphids on the four host plants (Fig. S5B). For verification, we used biological samples different from those employed in the RNA-seq analysis to ensure the reliability of RES across lines. Sequence analysis revealed that nine RES were present in PCR products derived from cDNA but absent from gDNA (Fig. 2C). The nine RES included five A-to-I sites, three G-to-A sites, and one U-to-C site (Fig. 2C), which were repeatedly validated across multiple independent samples. TPM of editing levels at these sites varied significantly across lines reared on different host plants (Fig. 2D), further supporting the hypothesis that RNA editing contributes to host adaptation in GPA.

**Rapid response of RES editing levels correlates with gene expression to host change**

To determine whether RNA editing levels in GPA respond to different host plants, we developed a RES-specific qPCR approach to quantify the editing levels of specific RNA editing sites modified from allelic-specific PCR (50) and mismatch primer design (51-53) (Fig. 3A). This method was designed to distinguish between edited and unedited transcripts by placing the edited or unedited nucleotide at the 3′ end of the forward primer and introducing an additional mismatch 1–2 bases upstream. These primers, combined with a common reverse primer, allowed for specific amplification of edited versus unedited cDNA templates (Fig. S6A).

We first validated primer specificity by showing that they amplified edited and unedited cDNA, but not gDNA (Fig. 3B). Sanger sequencing confirmed that the PCR products contained the expected edits and primer-introduced mismatches (Fig. S6B). Edited and unedited primers showed similar amplification efficiencies on cDNA, but amplification of gDNA with edited primers was strongly suppressed due to the 3′-end mismatch (ΔCt = 5.36–9.05) (Fig. S6C), which is consistent with previous reports (53). Using this RES-qPCR method, we quantified three A-to-I sites (RES1, RES29, RES46-2) and one U-to-C site (RES14) (Fig. 3B). Editing at three of these sites—RES29, RES46-2, and RES14—varied with host plant, showing higher levels in GPA feeding on Zm than on Bn (Fig. 3C).

To further test whether RNA editing responds dynamically to different host plants, we performed reciprocal host-switch experiments (Fig. 3D, 3G). Third-instar GPAs reared on Bn were transferred to either Bn (control) or Zm. After 48 hours, editing at RES29, RES46-2, and RES14 was significantly increased in aphids switched from Bn to Zm (Fig. 3E). RES14 (a U-to-C event) is located in the 5′ UTR of a lncRNA (MYZPE13164_O_EIv2.1_0099850, hereafter lncRNA850), whereas RES46-2 (an A-to-I event) resides in the 3′ UTR of the *ORC3* gene. Consistent with the induction of editing, lncRNA850 expression decreased, and expressions of *ZBED4* and *ORC3* were increased in Bn-to-Zm aphids (Fig. 3F). The opposite trend was observed in the reciprocal switch. When GPAs were transferred from Zm to Bn (Fig. 3G), editing levels at RES29, RES46-2, and RES14 significantly decreased (Fig. 3H). Correspondingly, expressions of *ZBED4* and *ORC3* were downregulated, and lncRNA850 expression was upregulated in Zm-to-Bn aphids (Fig. 3I).

Together, these results suggest that RNA editing in GPA is responsive to host plants and may play a role in host adaptation, likely through site-specific RNA editing events affecting both coding and non-coding transcript features.

**Identification and domain architecture of RNA-editing deaminases in GPA**

In eukaryotes, the RNA editing mechanisms remain largely unknown. Among the best-characterized events are transition edits such as A-to-I and C-to-U, mediated by deamination reactions. Few U-to-C and G-to-A events resulting from transamination reactions (33, 48) have been reported, but the enzymes responsible for these are still unknown. Therefore, we focused on identifying deaminases in *M. persicae*.

Using the Pfam database, we surveyed the deaminase superfamily (Pfam clan CL0109), which includes 33 domain families (Dataset S3), along with two additional domains—dsRNA-binding (dsRB) and Z-DNA/RNA-binding—known from mammalian ADARs. We searched for all 35 domains in the GPA Pfam annotations (Dataset S4) and identified the presence of A_deamin and dsRB domains, but not the Z-DNA/RNA-binding domain (Dataset S3), suggesting that GPA ADARs contain A_deamin and dsRB domains. Three proteins with A_deamin domains were found: two also contained dsRB domains and were annotated as ADAR-like based on NCBI homology, while the third was annotated as an ADAT (adenosine deaminase acting on tRNA) (Dataset S4). No domains related to AID/APOBEC deaminases were detected in GPA or other arthropod genomes examined.

ADARs and ADATs both catalyze A-to-I RNA editing, but act on distinct substrates: ADARs on double-stranded RNAs and ADATs on tRNAs. Comparative analysis of 22 arthropod genomes revealed that the ADAT gene is present as a single copy in most species (Dataset S3), with multiple copies detected only in a few cases. In contrast, ADAR genes occur as multiple copies in several species, and some copies lack canonical dsRB domains (Fig. 4A). These findings indicate that ADAR genes have undergone lineage-specific diversification in arthropods, which may lead to differences among species in RNA substrate recognition or post-transcriptional regulatory functions.

***ADARs* are essential for GPA host adaptation**

In GPA, expression of *ADAR* genes varied significantly depending on the host plant (Figure 4B), and their expression was increased in response to changed host

plants (SI Figure 7), implicating their potential roles in RNA editing-mediated adaptation to different hosts. To directly assess the functional contribution of ADAR genes to host adaptation, we silenced *ADAR1* and *ADAR2* in aphids originally maintained on Bn via dsRNA feeding (Fig. 4C). Knockdown of either gene significantly reduced aphid fecundity compared with aphids fed a blank treatment or *dsLTA* control (Fig. 4D). These findings, together with the host-dependent expression patterns of *ADAR*s, indicate that *ADAR*s are required for GPA fitness and likely play an essential role in host plant adaptation.

To test the impact of *ADAR*s in host adaptation, we performed host-change experiments by transferring RNAi-treated and control aphids from Bn to Zm. In *ADAR1*- and *ADAR2*-silenced aphids on Bn, the editing level at RES46-2—a site of A-to-I editing—was decreased relative to control. After 24 hours on Zm, control aphids exhibited upregulation of both RES46-2 editing (Figure 4E), which was abolished in RNAi aphids (Figure 4E). As a result of this impaired response, *ADAR1*- and *ADAR2*-RNAi aphids experienced significantly higher mortality following transfer to Zm compared to non-silenced aphids (Figure 4F).

These results indicate that RES46-2 editing is regulated by ADARs in response to host plant changes. Disruption of the *ADAR*-mediated pathway is associated with lower GPA survival on a novel host, underscoring the role of RNA editing in host adaptation.

**Conserved A-to-I RNA editing with species-specific functional divergence across aphid lineages**

To investigate RNA editing in aphids, we analyzed editing events in two specialist species, *Rhopalosiphum padi* (BOA) and *Brevicoryne brassicae* (CA), together with the generalist GPA. Previously, we generated RNA-seq datasets for GPA, CA, and BOA subjected to host plant shifts (38) and corresponding genomic DNA sequences (PRJNA1344291). After filtering out genomic SNPs using Reditools, we identified 1270 RNA editing sites in GPA, 368 in CA, and 500 in BOA (Fig. 5A, Dataset S5).

The genomic distribution of all identified editing sites also showed clear lineage specificity. The three species differed in editing preferences between untranslated regions (UTRs) and coding sequences (CDSs), and CDS edits contributed differently to synonymous and nonsynonymous substitutions (Fig. 5B). Across all three species,

A-to-I was the predominant editing type, accounting for 33.8% in GPA, 72.0% in CA, and 25.4% in BOA. Notably, GPA and BOA exhibited a broader diversity of non–A-to-I editing events than CA, a pattern consistent with their wider host ranges (Fig. 5A).

Sequence context analysis revealed conserved motifs and characteristic RNA secondary structures surrounding A-to-I sites, which were preferentially located within stem–loop regions (Fig. 5C–D), supporting that A-to-I editing is a conserved and structurally guided feature of aphid transcriptomes. Despite this conservation, the functional targets of RNA editing differed among species (Fig. 5D, Fig. S8). In GPA, edited genes were enriched in neural and apoptotic pathways, including cell fate determination, differentiation, and programmed cell death. In BOA, editing events were concentrated in genes related to V-type ATPase assembly and chromatin regulation. In CA, edits were enriched in developmental and histone modification pathways (Fig. 5E, Table S4).

Overall, A-to-I RNA editing represents a conserved molecular mechanism in aphids, but its genomic distribution and functional targets are species-specific, reflecting potential differences in ecological adaptation and host utilization strategies.

## Discussion

In this study, we investigated parthenogenetic lines of GPA, a highly generalist aphid capable of colonizing over 1,000 plant species across more than 50 plant families (34). We show that extensive RNA editing in GPA generates substantial transcriptomic variation, potentially compensating for its limited genetic diversity. These editing events are diverse and, in many cases, strongly influenced by host plants, with editing levels shifting in response to different hosts. Notably, several RNA editing sites respond rapidly to host changes, correlating with altered expression of the corresponding transcripts.

RNA editing in GPA is highly dynamic. More than 150 editing sites differed among aphids feeding on four host plants, and several showed reversible shifts within 48 hours of host transfer. Editing occurred in both protein-coding and non-coding regions and was associated with host-responsive changes in transcript abundance. Of particular interest, we identified an A-to-I site (RES46-2) in the 3′UTR of *ORC3* and a U-to-C site (RES14) in the 5′UTR of a long noncoding RNA (lncRNA850). *ORC3* is a subunit of the origin recognition complex, a conserved

multi-protein assembly that initiates DNA replication, and in insects it is implicated in replication licensing, cell cycle regulation, and developmental plasticity (54-58). These observations suggest that RNA editing may modulate transcript stability or translation in a host-specific manner, providing a mechanism for fine-tuning fundamental cellular processes in response to environmental change.

RNA editing likely underlies the transcriptional plasticity that allows GPA to thrive on diverse hosts. It confers several adaptive advantages: it enables rapid phenotypic plasticity without heritable genetic changes; allows environment-specific regulation of gene function through post-transcriptional modifications; and can modulate key physiological activities via nonsynonymous edits that alter protein function. For example, RES1 represents an A-to-I edit in the ion channel transcript *KCNB1* (Table S5), which encodes a voltage-gated potassium channel, suggesting that RNA editing may influence neural or physiological processes critical for host recognition, feeding, or detoxification.

RNA editing events in GPA can be classified into transitions and transversions, with transitions being the most common. GPA possesses two *ADAR* genes, each containing 1-2 double-stranded RNA-binding motifs, but lacks genes encoded for AID/APOBEC deaminases, which mediate other RNA editing types in many organisms (31). Knockdown of *ADAR* genes significantly reduced editing, decreased aphid fecundity, and impaired adaptation to novel hosts, demonstrating that RNA editing is essential for environmental acclimation in this parthenogenetic species. The enzymes responsible for less common editing types in GPA remain unknown, and understanding their origins may provide insights into the molecular basis of ecological generalism and host adaptation in insects.

A-to-I editing is the most common RNA editing type across metazoans, with other types observed less frequently. In GPA, however, RNA editing is exceptionally diverse, encompassing 12 types, many of which are rarely reported in other organisms (24). Among 22 metazoan species studied, A-to-I dominates, with other types largely absent except in the marine invertebrate *Trichoplax adhaerens*, which exhibits a broader editing repertoire (17, 24). Even in *Metopolophium dirhodum*, a close relative of GPA and a specialist on cereal plants, A-to-I remains predominant (17). In GPA, A-to-I edits account for only a fraction of the total, raising the intriguing question of whether the broader diversity of RNA editing types contributes to its polyphagy. Indeed, we observed a correlation between RNA-editing diversity and

host range among aphids: GPA and BOA, which colonize multiple plant families, exhibit more non-A-to-I editing events than CA, a Brassicaceae specialist (59). Although A-to-I editing retains conserved biochemical features across aphids, its genomic distribution and functional consequences are species-specific, reflecting lineage-specific adaptation strategies.

In summary, GPA employs a diverse, dynamic, and host-responsive RNA editing system that overcomes the constraints of clonal reproduction and facilitates adaptation to a wide range of host plants. More broadly, these findings highlight RNA editing as a flexible and underappreciated driver of environmental adaptation in metazoans, particularly in species with limited genetic variation.

## Acknowledgments

This project is funded by the National Natural Science Foundation of China (project No. 32172392 to Y.C.), the National Key Research and Development Program of China (project No. 2023YFF1000703 to Y.C.), and supported by Hubei Hongshan Laboratory (project No. 2022hszd026 to Y.C.), the Startup Foundation for Advanced Talents at HZAU to Y.C., the Fundamental Research Funds for the Central Universities (Program No. 2022ZKPY003 to Y.C.), the Wuhan Yingcai Talent Program to Y.C., and grants from the National Science Fund for Distinguished Young Scholars (32325039 to S.J.).

## References

1. R. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol*. **23**, 38–44 (2008).
2. N. I. Katis, J. A. Tsitsipis, M. Stevens, G. Powell, “Transmission of plant viruses” in *Aphids as Crop Pests*, R. Harrington, H. F. Van Emden, Eds. (CABI, 2007), pp. 353–390.
3. J.-C. Simon, F. Delmotte, C. Rispe, T. Crease, Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biol. J. Linn. Soc.* **79**, 151–163 (2003).
4. J.-C. Simon, S. Stoeckel, D. Tagu, Evolutionary and functional insights into reproductive strategies of aphids. *C. R. Biol.* **333**, 488–496 (2010).

5. F. D. Hales *et al*., Lack of detectable genetic recombination on the X chromosome during the parthenogenetic production of female and male aphids. *Genet. Res.* **79**, 203–209 (2002).
6. H. D. Loxdale, A. Balog, D. G. Biron, Aphids in focus: Unravelling their complex ecology and evolution using genetic and molecular approaches. *Biol. J. Linn. Soc.* **129**, 507–531 (2020).
7. J. S. Harrison, E. B. Mondor, Evidence for an invasive aphid “superclone”: Extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern united states. *PLoS One* **6**, e17524 (2011).
8. S. I. Martel, C. G. Ossa, J. Simon, C. C. Figueroa, F. Bozinovic, Latitudinal trend in the reproductive mode of the pea aphid *Acyrthosiphon pisum* invading a wide climatic range. *Ecol. Evol*. **10**, 8289–8298 (2020).
9. M. E. Rubio-Meléndez, J. Barrios-SanMartin, F. E. Pina-Castro, C. C. Figueroa, C. C. Ramirez, Asexual reproduction of a few genotypes favored the invasion of the cereal aphid *Rhopalosiphum padi* in chile. *PeerJ* **7**, e7366 (2019).
10. D. Liu *et al*., Life-history responses of insects to water-deficit stress: A case study with the aphid *Sitobion avenae*. *BMC Ecol.* **18**, 17 (2018).
11. C. Bass, R. Nauen, The molecular mechanisms of insecticide resistance in aphid crop pests. *Insect Biochem. Mol. Biol.* **156**, 103937 (2023).
12. A. D. Yates, A. Michel, Mechanisms of aphid adaptation to host plant resistance. *Curr. Opin. Insect Sci.* **26**, 41–49 (2018).
13. W. M. Gommans, S. P. Mullen, S. Maas, RNA editing: A driving force for adaptive evolution? *Bioessays* **31**, 1137–1145 (2009).
14. X. Grau-Bové, D. Weetman, RNA editing: An overlooked source of fine-scale adaptation in insect vectors? *Curr. Opin. Insect Sci.* **40**, 48–55 (2020).
15. Y. Duan *et al*., A-to-I RNA editing in honeybees shows signals of adaptation and convergent evolution. *iScience* **24**, 101983 (2021).
16. H. T. Porath *et al*., RNA editing is abundant and correlates with task performance in a social bumblebee. *Nat. Commun.* **10**, 1605 (2019).
17. B. Zhu *et al*., A-to-I RNA editing of CYP18A1 mediates transgenerational wing dimorphism in aphids. *eLife* **13**, RP96540 (2025).
18. L. E. Rieder *et al*., Dynamic response of RNA editing to temperature in drosophila. *BMC Biol.* **13**, 1 (2015).

19. Y. Duan *et al*., 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.* **81**, 136 (2024).
20. M. Grauso, R. A. Reenan, E. Culetto, D. B. Sattelle, Novel putative nicotinic acetylcholine receptor subunit genes, *Dα5*, *Dα6* and *Dα7* in *Drosophila melanogaster* identify a new and highly conserved target of adenosine deaminase acting on RNA-mediated a-to-I pre-mRNA editing. *Genetics* **160**, 1519–1533 (2002).
21. Z. Liu, W. Song, K. Dong, Persistent tetrodotoxin-sensitive sodium current resulting from U-to-C RNA editing of an insect sodium channel. *Proc. Natl. Acad. Sci.* **101**, 11862–11867 (2004).
22. A. K. Jones *et al*., Splice-variant- and stage-specific RNA editing of the drosophila GABA receptor modulates agonist potency. *J. Neurosci.* **29**, 4287–4292 (2009).
23. X. Yao *et al*., Nicotinic acetylcholine receptor β1 subunit from the brown planthopper, *Nilaparvata lugens*: A-to-I RNA editing and its possible roles in neonicotinoid sensitivity. *Insect Biochem. Mol. Biol.* **39**, 348–354 (2009).
24. P. Zhang *et al*., On the origin and evolution of RNA editing in metazoans. *Cell Rep.* **42**, 112112 (2023).
25. E. Eisenberg, “Proteome diversification by RNA editing” in RNA Editing: Methods and Protocols, Methods in Molecular Biology., E. Picardi, G. Pesole, Eds. (Humana Press Inc, 2021), pp. 229–251.
26. E. Eisenberg, E. Y. Levanon, A-to-I RNA editing — immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
27. Y.-H. E. Hsiao *et al*., RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res.* **28**, 812–823 (2018).
28. A. Brümmer, Y. Yang, T. W. Chan, X. Xiao, Structure-mediated modulation of mRNA abundance by A-to-I editing. *Nat. Commun.* **8**, 1255 (2017).
29. Y. A. Savva, L. E. Rieder, R. A. Reenan, The ADAR protein family. *Genome Biol.* **13**, 252 (2012).
30. B.-E. Wulff, M. Sakurai, K. Nishikura, Elucidating the inosinome: Global approaches to adenosine-to-inosine RNA editing. *Nat. Rev. Genet.* **12**, 81–85 (2011).

31. V. Blanc, N. O. Davidson, APOBEC-1-mediated RNA editing. WIREs Syst. *Biol. Med.* **2**, 594–602 (2010).
32. Ruchika, C. Okudaira, M. Sakari, T. Tsukahara, Genome-wide identification of U-to-C RNA editing events for nuclear genes in *Arabidopsis thaliana*. *Cells* **10**, 635.
33. J. Tao *et al*., Transcriptome-wide identification of G-to-A RNA editing in chronic social defeat stress mouse models. *Front. Genet.* **12**, 680548 (2021).
34. Myzus persicae (green peach aphid). CABI Compend. Available at: https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.35642 [Accessed 12 October 2025].
35. C. Bass *et al*., The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect Biochem. Mol. Biol.* **51**, 41–51 (2014).
36. K. S. Singh *et al*., Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Commun. Biol.* **4**, 847 (2021).
37. T. C. Mathers *et al*., Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* **18**, 27 (2017).
38. J. Wu *et al*., Plant-induced transcriptional plasticity diverges between generalist and specialist herbivores. bioRxiv, Preprint (2024). https://doi.org/10.1101/2024.12.07.627354
39. J. Wu, Z. feng Zhang, H. H. Cao, T.-X. Liu, Transcriptional and physiological plasticity of the green peach aphid (hemiptera: Aphididae) to cabbage and pepper plants. *J. Econ. Entomol.* **118**, 416–429 (2025).
40. Y. Nio *et al*., Host specialisation or generalism? Population genetics of the aphid *Myzus persicae* reveals dominance of superclones across diverse host plants. *Front. Ecol. Evol.* **13**, 1635527 (2025).
41. Y. Chen *et al*., An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proc. Natl. Acad. Sci.* **117**, 12763–12771 (2020).
42. T. C. Mathers *et al*., Sex-specific changes in the aphid DNA methylation landscape. *Mol. Ecol.* **28**, 4228–4241 (2019).
43. J. B. Li *et al*., Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).

44. G. Ramaswami *et al*., Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **10**, 128–132 (2013).
45. R. Zhang *et al*., Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nat. Methods* **11**, 51–54 (2014).
46. E. Picardi, A. M. D'Erchia, A. Montalvo, G. Pesole, Using REDItools to detect RNA editing events in NGS datasets. *Curr. Protoc. Bioinforma.* **49** (2015).
47. M. Piechotta, I. S. Naarmann-de Vries, Q. Wang, J. Altmüller, C. Dieterich, RNA modification mapping with JACUSA2. *Genome Biol.* **23**, 115 (2022).
48. M. Lefouili, K. Nam, The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci. Rep.* **12**, 11331 (2022).
49. F. Zhang, Y. Lu, S. Yan, Q. Xing, W. Tian, SPRINT: An SNP-free toolkit for identifying RNA editing sites. *Bioinformatics* **33**, 3538–3548 (2017).
50. Y. Chen *et al*., Allelic-specific expression in relation to bombyx mori resistance to bt toxin. *Insect Biochem. Mol. Biol.* **54**, 53–60 (2014).
51. S. Sauer, A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res.* **28**, 13e–113 (2000).
52. F. M. You *et al*., BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253 (2008).
53. S. Lefever, F. Pattyn, J. Hellemans, J. Vandesompele, Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clin. Chem.* **59**, 1470–1480 (2013).
54. D. Kopytova *et al*., ORC interacts with THSC/TREX-2 and its subunits promote Nxf1 association with mRNP and mRNA export in drosophila. *Nucleic Acids Res.* **44**, 4920–4933 (2016).
55. B. L. Hua, T. L. Orr-Weaver, DNA replication control during drosophila development: Insights into the onset of S phase, replication initiation, and fork progression. *Genetics* **207**, 29–47 (2017).
56. D. Remus, M. Blanchette, D. C. Rio, M. R. Botchan, CDK phosphorylation inhibits the DNA-binding and ATP-hydrolysis activities of the drosophila origin recognition complex. *J. Biol. Chem.* **280**, 39740–39751 (2005).
57. V. V. Popova, A. V. Brechalov, S. G. Georgieva, D. V. Kopytova, Nonreplicative functions of the origin recognition complex. *Nucleus* **9**, 460–473 (2018).

58. J. Liu, K. McConnell, M. Dixon, B. R. Calvi, Analysis of model replication origins in drosophila reveals new aspects of the chromatin landscape and its relationship to origin activity and the prereplicative complex. Mol. Biol. *Cell* **23**, 200–212 (2012).

59. E. Kazana *et al*., The cabbage aphid: A walking mustard oil bomb. *Proc. R. Soc. B Biol. Sci.* **274**, 2271–2277 (2007).

## Main figures

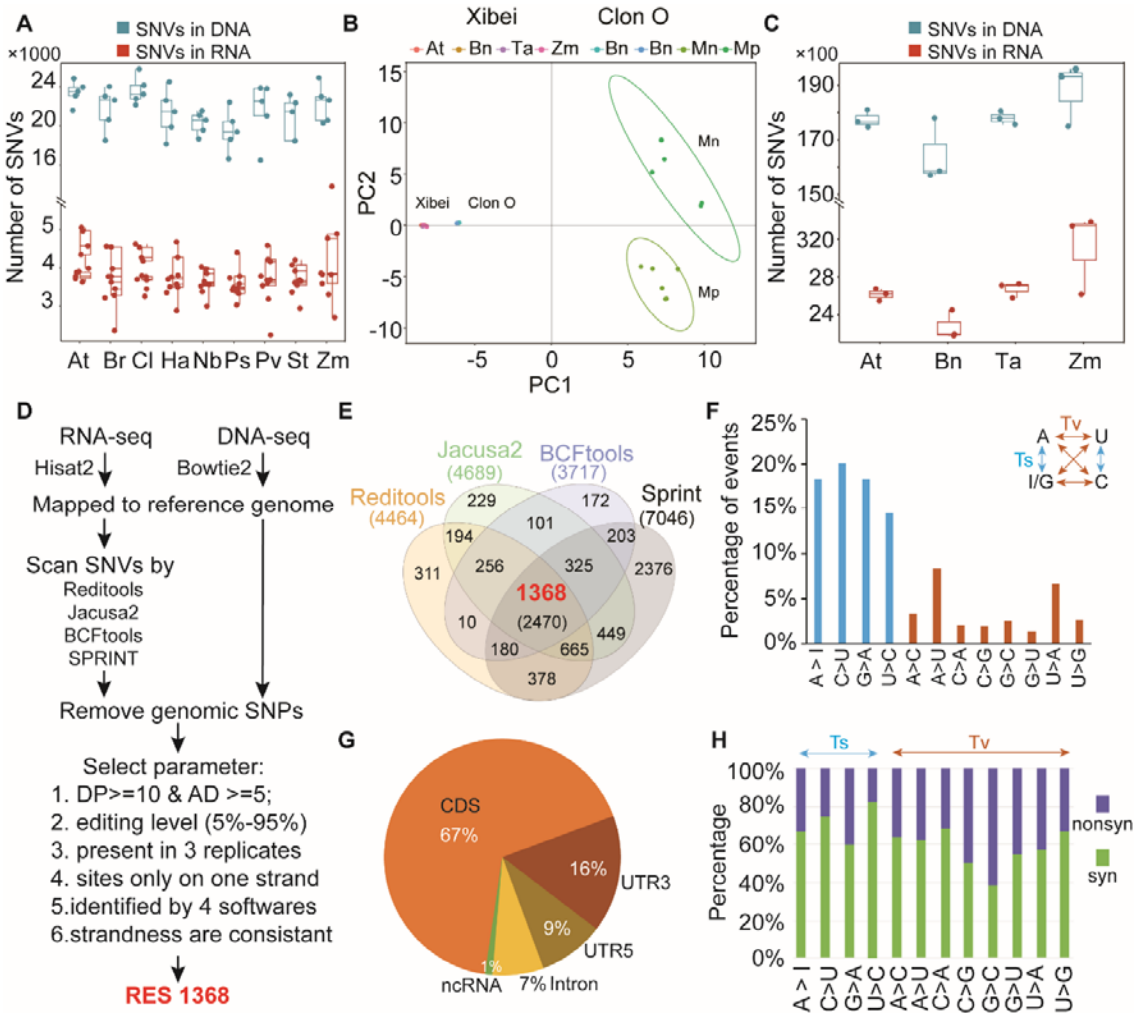


**Fig. 1. Mismatches between RNA and DNA are derived from RNA editing**

(A) Number of SNVs detected in RNA and DNA of *M. persicae* Clone O across nine different host plants. RNA-seq data were collected from 2017 to 2018 and published in Chen *et al*. (2020) (41). DNA-seq data from 2016 and 2018 were published in Mathers *et al*. (2017, 2019) (37,42). Host plants include *Brassica rapa* (Br), *Arabidopsis thaliana* (At), *Nicotiana benthamiana* (Nb), *Solanum tuberosum* (St), *Chrysanthemum indicum* (Ci), *Helianthus annuus* (Ha), *Pisum sativum* (Ps), *Phaseolus vulgaris* (Pv), and *Zea mays* (Zm). (B) Single-female–derived GPA lines on different hosts (Clone O and Xibei) clustered tightly together, whereas the Mp and Mn geographic lineages were more divergent. Principal component analysis (PCA) of *M. persicae* colonies based on DNA-seq–derived SNVs. DNA-seq data for Clone O were obtained from Mathers *et al*. (2017, 2019), for the Xibei clone from Wu *et al*. (2024), and for Mp and Mn from Singh *et al*. (2021). (C) Number of SNVs detected in

RNA and DNA of the Xibei clone across four different host plants: Bn, At, Ta, and Zm. (D) Workflow for identifying RNA editing sites. Strand-specific RNA-seq data were aligned to the *M. persicae* reference genome using HISAT2, and DNA-seq data were aligned using Bowtie2. SNVs in RNA were independently identified using Reditools, JACUSA2, BCFtools, and SPRINT. SNVs also present in the DNA were excluded. After applying six filtering criteria, 1,368 high-confidence RES were identified. DP (depth) indicates the total number of reads (edited and unedited), and AD (allele depth) refers to the number of edited reads. (E) Overlap of RES detected by the four tools: Reditools (46), JACUSA2 (47), BCFtools (48), and SPRINT (49). (F) Distribution of the 12 possible nucleotide substitution types among total RES. Transition events were more frequent than transversions. (G) Genomic features associated with RES. (H) Proportion of nonsynonymous (nonysn) and synonymous (syn) changes among RES located in coding sequences (CDS).

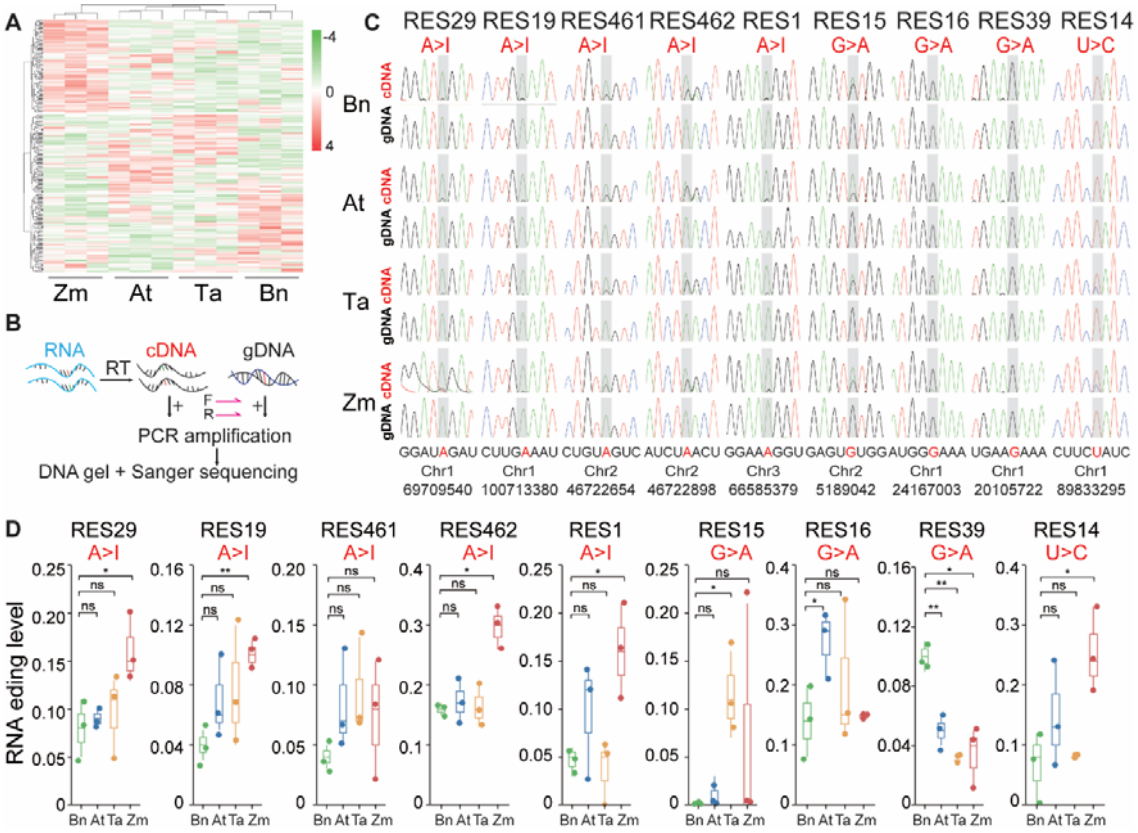


**Fig. 2. RES differentially edited in *M. persicae* on divergent plant hosts.** (A) Heatmap of differentially edited RES across host plant treatments. (B) Workflow for RES validation. Forward and reverse primers were designed to amplify each target RES with 150–200 bp flanking sequences. PCR was performed using cDNA and gDNA as templates, followed by Sanger sequencing of the products. (C) Representative nucleotide sequences of RES in cDNA and gDNA from aphids reared on four host plants. Each panel is labeled with the corresponding RES identifier and editing type. Edited sites are highlighted in gray. In cDNA chromatograms, black peaks with gray highlights represent inosines (I) that were detected as guanosines (G) in Sanger sequencing. Genomic coordinates are indicated below the panels. (D) Editing levels of selected RES in aphids on different host plants. Colored dots represent RNA editing levels of individual samples for each host, and boxplots show the data distribution within each host. Statistical significance is indicated as *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and n.s. denotes not significant ($P > 0.05$), based on Wilcoxon rank-sum test.

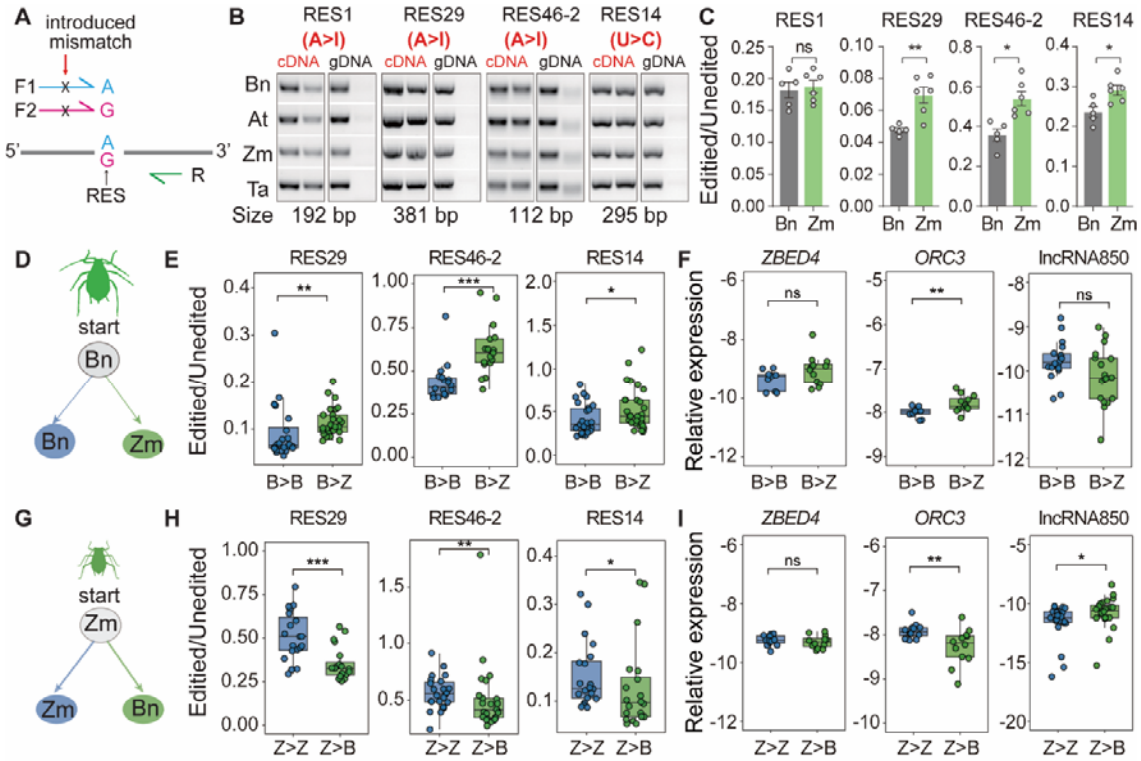


**Fig. 3. Rapid response of RES editing levels and gene expression to host change.**

(A) Design of RES-qPCR. The 3′ end of each forward primer targets the edited or unedited nucleotide, with an additional mismatch introduced 1–2 nucleotides upstream. Each forward primer is paired with a common reverse primer to selectively amplify edited or unedited transcripts. (B) RES-qPCR selectively amplifies both edited and unedited transcripts from cDNA, but only unedited transcripts from gDNA. (C) Ratio of edited to unedited transcripts in *M. persicae* reared on Bn and Zm. (D, G) Schematics of host-switch experiments. Aphids reared on Bn were transferred either to Zm (B>Z) or to Bn (B>B), and aphids reared on Zm were transferred either to Bn (Z>B) or to Zm (Z>Z). (E, H) Ratio of edited to unedited transcripts following host transfer from Bn to Zm and from Zm to Bn, respectively. (F, I) Expression levels of genes containing RES after host transfer from Bn to Zm and from Zm to Bn, respectively. In Figures C, E, and H, RNA editing levels were quantified by qPCR. For each gene, the total transcript abundance was used as an internal reference. ΔCt values for the edited and unedited transcripts were calculated relative to this reference and subsequently converted to relative expression levels ($2^{-\Delta Ct}$). The RNA editing level was then expressed as the ratio of $2^{-\Delta Ct(edited)}$ to $2^{-\Delta Ct(unedited)}$, yielding results consistent with those normalized using conventional housekeeping genes. In

Figure F and I, relative expression of genes containing candidate editing sites was calculated using the $2^{-\Delta Ct}$ method and normalized to the *EF1α* gene. RES29 corresponds to *ZBED4*, RES46-2 to *ORC3*, and RES14 to long non-coding RNA 850 (lncRNA850). In Figure C, E, F, I, and J, each data point (white dot with a black outline) represents a ratio or expression level, and bar graphs show the distribution across treatments. Each treatment includes n = 15 aphids. Statistical significance is indicated as *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and n.s. denotes not significant ($P > 0.05$), determined by Wilcoxon rank-sum test.

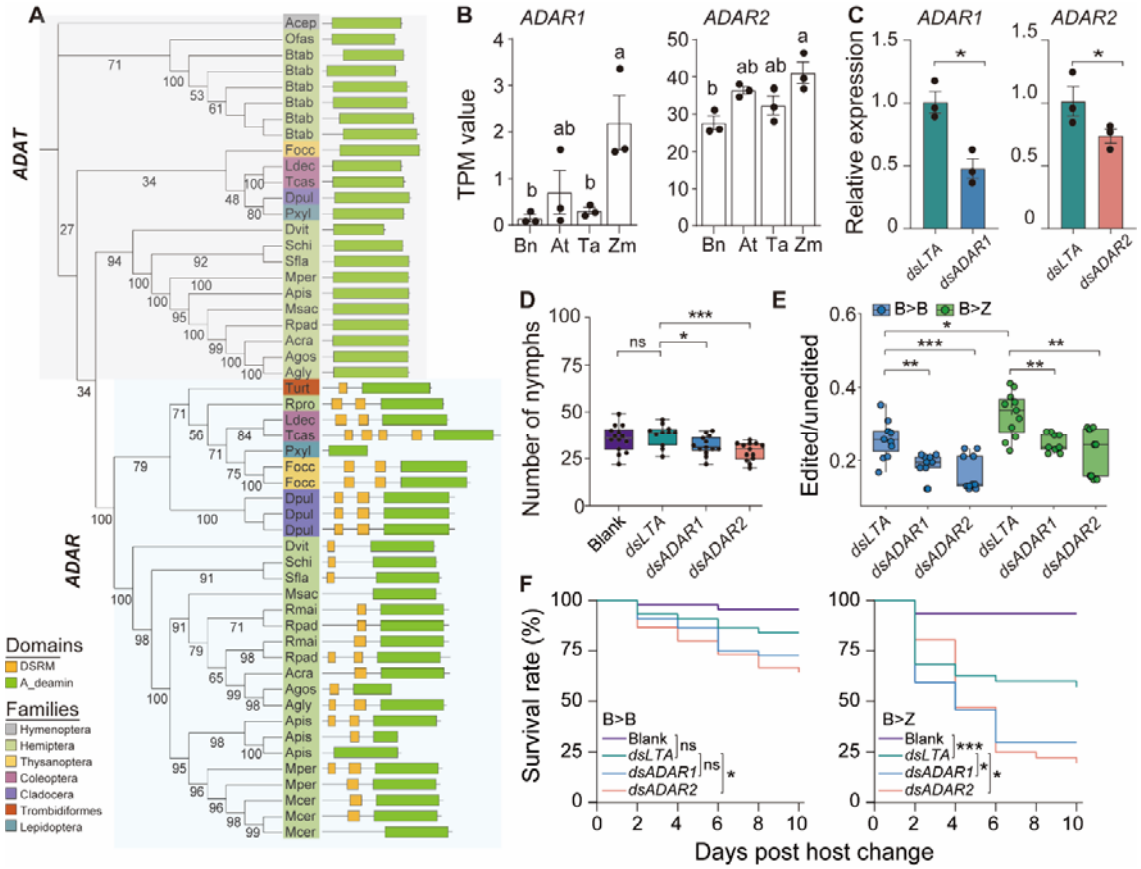


**Fig. 4 ADARs regulate RNA editing and contribute to *M. persicae* host adaptation.**

(A) Phylogenetic tree of ADARs and ADAT. *Tetranychus urticae* was used as the outgroup. Bootstrap values (%) are indicated at the nodes. Shaded colors indicate insect orders. Branch labels show species abbreviations (e.g., Mper = *M. persicae*; full list in Table S3). Functional domains are shown to the right: yellow boxes = dsrm (double-stranded RNA-binding) domain, green bars = adenosine deaminase domain. The large gray and blue shaded regions highlight the ADAT and ADAR clades, respectively. (B) TPM of *ADAR1*, *ADAR2* in *M. persicae* reared on different host plants. Different lowercase letters indicate significant differences (one-way ANOVA with Tukey's test). (C–D) RNAi knockdown of *ADAR1* and *ADAR2* by dsRNA feeding reduced transcript levels (normalized to *MpEF1α*, mean ± SEM, n = 3), decreased fecundity (each dot = 1 aphid, n = 12–15; blank: 30% sucrose, *dsLTA*: unrelated dsRNA). (E) Silencing *ADAR1* or *ADAR2* reduced RES46-2 editing levels on Bn and blocked the rapid upregulation observed after transfer to Zm (mean ± SEM, n = 10). (F) Silencing *ADAR1* or *ADAR2* reduced survival on the new host (surviving aphids

counted every two days, n = 10–12). Statistical significance: Wilcoxon rank-sum test for transcript levels, fecundity and editing levels, log-rank test for survival; ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, n.s = not significant.

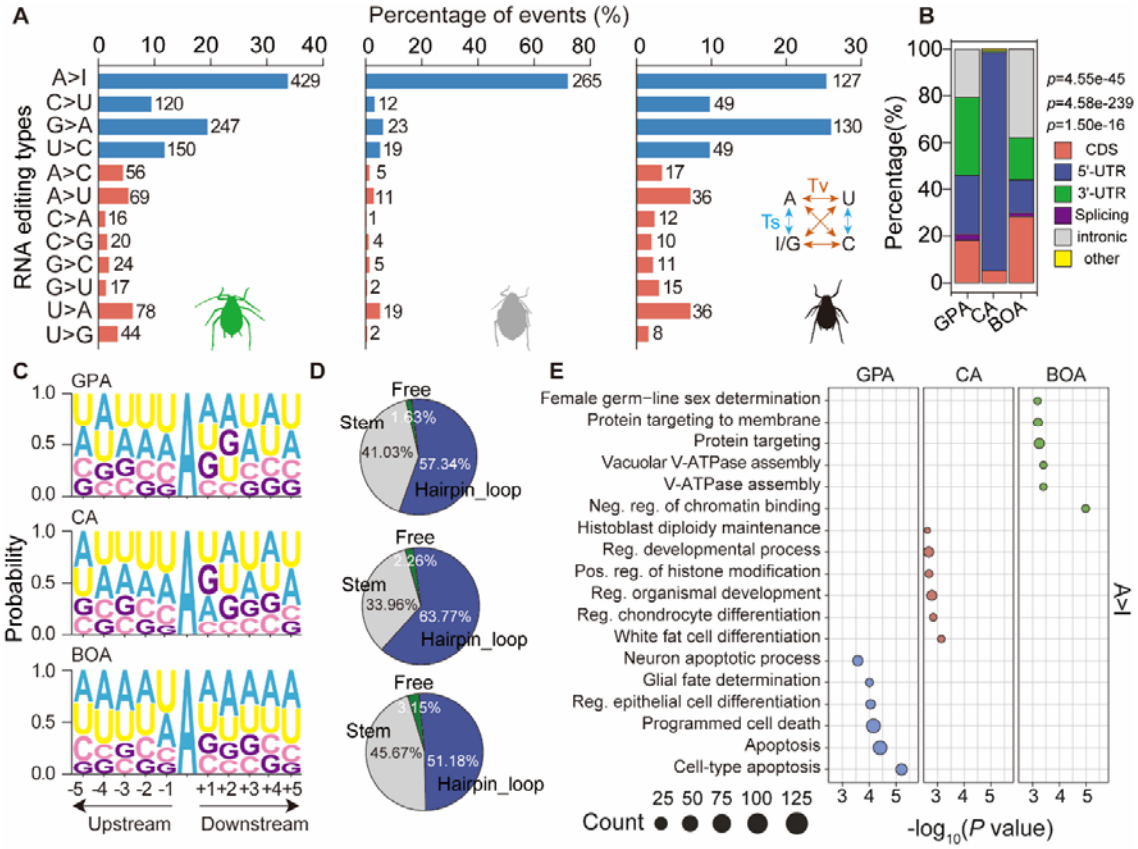


**Fig. 5 Comparative analysis reveals conserved and species-specific RNA editing patterns in aphids.**

(A) Distribution of the 12 types of RES in *M. persicae* (GPA, green), *B. brassicae* (CA, gray), and *R. padi* (BOA, black). The number of RNA variants for each nucleotide substitution type is shown to the right of each bar. A-to-I is the most prevalent type across all species. (B) Genomic distribution of A-to-I RES in the three aphid species. The stacked bar chart shows the percentage of editing events located in coding sequences (CDS), 5′-UTRs, 3′-UTRs, splice sites, intronic regions, and other regions. Significant differences in regional preference among species are indicated (*P* values from $\chi^2$ tests). (C) Sequence conservation of A-to-I editing sites and nucleotide preferences in the flanking sequences (5 bp upstream and downstream). (D) Percentage of A-to-I RES located in different types of RNA secondary structure elements. Prediction of RNA secondary structures is based on 30-nt upstream and 30-nt downstream sequences surrounding the edited sites. (E) Gene Ontology (GO) enrichment analysis of genes harboring A-to-I RES. GO terms were detailed in Table S4. Enriched biological processes are plotted based on gene count (dot size) and significance ($-\log_{10}$ (*P value*)).

**Supporting Information for**

# RNA editing fine-tunes transcriptional plasticity and enhances the adaptability of *Myzus persicae* to diverse host plants

Wenyuan Yu [a,b], Jun Wu [a,b], Rong Hu [a,b], Zhimou Lin [a,b], Peiyu Yang [a,b], Zhichao Hu [a,b], Pengshuai Peng [a,b], Shuai Zhan [c], Jean-Christophe Simon[d], Saskia A. Hogenhout [e], Gang Wu [b, #], Shuangxia Jin [a,f, #], Yazhou Chen [a,b, #]

[a] Hubei Hongshan Laboratory, Wuhan, 430070, China;
[b] Hubei Insect Resources Utilization and Sustainable Pest Management Key Laboratory, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China;
[c] Key Laboratory of Plant Design, Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China;
[d] INRAE (National Institute of Agriculture, Food and Environment), UMR IGEPP, Le Rheu, France;
[e] Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK.
[f] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei 430070, P. R. China

[#] Correspondence: Yazhou Chen, Shuangxia Jin, Gang Wu
Email: yzchen@mail.hzau.edu.cn; jsx@mail.hzau.edu.cn; wugang@mail.hzau.edu.cn

**This PDF file includes:**

Supporting text
Figures S1 to S8
Tables S1 to S5
Legends for Datasets S1 to S6
SI References

**Other supporting materials for this manuscript include the following:**

Datasets S1 to S5

## Supplementary text

## Materials and methods

### Aphid rearing

*M. persicae* (GPA) Clone Xibei was obtained from a stock culture maintained on cabbage (*B. oleracea*) in Prof. Tongxian Liu's laboratory at Northwest A&F University (Yangling, Shaanxi). This clone was subsequently propagated into lines maintained on four host species: *Z. mays* (Zm) B73, *B. napus* (Bn) Zhongshuang 11, *A. thaliana* (At), and *T. aestivum* (Ta). The GPA was initially reared on Bn for 2–3 weeks, after which a single parthenogenesis female was transferred to Zm, At, and Ta to establish stable parthenogenetic lines. The four aphid lines were maintained at 24 ± 1°C, 60 ± 10% relative humidity, and a 16 h light/8 h dark photoperiod for over 2 years.

For each GPA line, five samples were collected: three for RNA sequencing and the remaining two for DNA sequencing. Each sample comprised 30 age-synchronized adults.

### RNA extraction and RNA sequencing

Aphid samples for RNA-seq were ground to a fine powder with sterilized stainless-steel beads in a TissueLyser II (Jingxin, Shanghai, China). Total RNA was isolated and purified from the samples using TRIzol® Reagent (Invitrogen,15596026CN, USA) in conjunction with VAHTS® mRNA Capture Beads (Vazyme, N403, Nanjing, China). RNA concentration and purity were quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and Qubit 2.0 Fluorometer (Thermo Scientific, USA). RNA integrity was subsequently evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA).

Strand-specific libraries were constructed following the protocol provided with the VAHTS® Universal V6 RNA-seq Library Prep Kit (Vazyme, NR604, Nanjing, China). Sequencing was then performed on the DNBSEQ-T7 platform (OEbiotech, Shanghai, China), generating paired-end reads of 150 bp.

### DNA extraction and DNA sequencing

Genomic DNA of each GPA line, was extracted using the cetyltrimethylammonium bromide (CTAB) method and further purified with the Blood and Cell Culture DNA Midi Kit (QIAGEN, 13343, Germany).

Library construction and genome sequencing were performed by Novogene (Beijing, China) according to the manufacturer's protocol. Sequencing libraries were generated using the NEBNext Ultra DNA Library Prep Kit (Illumina, E7645S, USA). Sequencing was then carried out on the Illumina NovaSeq 6000 platform (Illumina, USA), with paired-end reads of 2 × 150 bp for each sample.

**SNP Calling by GATK**

To minimize false-positive SNVs introduced by sequencing errors or mapping bias, we implemented stringent multi-step filtering and cross-validation. In addition to standard GATK filtering (QUAL, QD, FS, SOR, and cluster thresholds), we verified variant calling accuracy by benchmarking with simulated datasets of known variants, ensuring tool performance under our sequencing depth and read length conditions. Variants supported only by a single aligner or by strand-biased reads were discarded. This conservative pipeline reduces spurious SNVs and increases confidence in downstream RNA editing detection.

We used the published genome sequencing data of the *M. persicae* Clone O strain (PRJEB11304, PRJNA613055) and previous transcriptome sequencing data of GPA colonized on nine host plants (PRJEB24317, PRJEB11304, PRJNA532419) (1-3), for subsequent analysis. DNA-seq and RNA-seq data were aligned to the *M. persicae* Clone O genome (4) using BWA (version 0.7.17-r1188) (5) and HISATS2 (6) (version 2.2.1), respectively. The alignment results were converted to BAM format using SAMtools (version 1.18) (7). The BAM files were then sorted and indexed using SortSam, and potential PCR duplicates were marked using MarkDuplicates both from the Picard Toolkit (version 3.0) (broadinstitute.github.io/picard/). Additional indexing was performed with SAMtools to facilitate downstream analyses. Single nucleotide polymorphisms (SNPs) were identified using HaplotypeCaller from the GATK (version 4.1.8.1) tool (8), and multi-dimensional filtering was applied to retain high-quality variants: (a) quality score (QUAL) below 30 were discarded; (b) quality-by-depth ratio (QD) below 2 were excluded to avoid variants influenced by abnormal sequencing depth; (c) significant strand bias, as indicated by Fisher Strand Bias (FS) and Symmetric Odds Ratio (SOR) values exceeding 60 and 4, respectively, were

removed; (d) clusters of more than three variants within a 10-base window were identified and excluded as potential false positives; and (e) missing values were considered non-passing variants. Then, VCF files from all samples were merged using BCFtools (version 1.8) (9). To eliminate genomic SNPs from the transcriptomic data, DNA variant sites were excluded using VCFtools (version 0.1.16) (10), resulting in a final set of transcriptome-specific SNPs.

**Ortholog analysis**

Protein sequences of GPA were analyzed using OrthoFinder (v2.5.4) (11) with the parameters "-M msa -t 30 -a 30", which identified 23,912 genes in *M. persicae*, including 8,109 single-copy genes (33.91%) and 15,803 multi-copy genes (66.09%).

**GO term enrichment analysis**

GO term enrichment analyses using the clusterProfiler package (version 3.18.1) (12). The input for these analyses was a series of defined DEG IDs.

**Identification of RNA-editing sites**

To enhance the robustness of RNA editing site (RES) detection, we employed four independent tools (REDItools2, JACUSA2, BCFtools, and SPRINT) and retained only consensus sites consistently identified by all tools. To further minimize potential mapping artifacts, candidates located in paralogous or repetitive genomic regions annotated by RepeatMasker were excluded. To ensure reproducibility, only sites present in all three biological replicates were considered. RES identification was based on paired transcriptome and genome resequencing data generated in this study, following the principle that target sites must be homozygous in genomic DNA (gDNA) and exhibit consistent nucleotide mismatches between RNA and DNA alignments (13). Based on this principle, two sequential rounds of RES identification and filtering were performed. In the first round, SNP detection was conducted on both RNA-seq and DNA-seq data to identify transcription-derived variants. Raw RNA-seq and DNA-seq reads were quality-controlled using fastp (v0.23.4) (14) and aligned to the reference genome as previously described. RNA-seq editing sites were independently identified using four bioinformatics tools: REDItools2 (v2.0) (15) with the REDItoolDnaRna.py script (-S -mrl 30 -q 20 -bq 30 -men 1 -T 1 -Tv 0.7 -N -B); JACUSA2 (v2.04) (16) in call-1 -a D -p 20 -f V mode; BCFtools (v1.10) (17) using

mpileup and call functions (--max-depth 10000 -q 20 -Q 20); and SPRINT (v0.1.8) (18) (-cd 200 -csrg 5). SNP detection for DNA-seq data was performed as described above. RES identified from RNA-seq data were then filtered against gDNA SNPs to exclude sites originating from genomic variation. In the second round of filtering, high-confidence RES were defined using the following criteria: (a) Coverage depth (DP) ≥ 10 and supporting reads for the edited base (AD) ≥ 5; (b) Editing level between 5% and 95%; (c) Presence in all three biological replicates; (d) Occurrence on a single strand only; (e) Identification by all four bioinformatics tools; (f) Consistency of editing results between forward and reverse strands.

**Genomic feature of RNA editing sites**

RNA editing sites (RES) were annotated using SnpEff (v5.2) with the parameters *-canon -no-upstream -no-downstream -hgvs* (19), restricting annotation to canonical transcripts and avoiding ambiguous intergenic assignments. Annotation incorporated aphid-specific gene models from Liu *et al*. (4). Because SnpEff assumes eukaryotic transcript structures, putative prokaryotic-like RNA modification features were excluded. For each RES, genomic context—including UTRs, introns, exons, and lncRNAs—was determined based on the *M. persicae* genome annotation (4). Additionally, the numbers of nonsynonymous and synonymous mutations were quantified to assess their potential impact on protein-coding sequences. To further explore sequence features and RNA structures, 30 nucleotides upstream and downstream of each A-to-I RNA editing site were extracted. WebLogo 3 (20) was used to evaluate and visualize sequence preferences surrounding editing sites, while RNAFold (21) was employed to predict the secondary structures of RNA sequences containing edited A sites.

**Differential editing (DE) RES analysis**

To assess whether RNA editing levels were significantly influenced by host plant species, we performed linear regression analyses using R. Editing levels (dependent variable: value) were modeled as a function of host identity (independent variable: hosts) for each editing site individually.

For each RNA editing site, we applied a linear model: *lm* (*value* ~ *hosts*). Here, *value* represents the editing level, and hosts is a categorical variable representing

the host plant. The regression was conducted using the dplyr package in R by grouping the Dataset by editing site and applying the linear model across each group.

To evaluate the effect of the host on editing level, we extracted the *p*-value associated with the *hosts* coefficient from the model summary. This corresponds to an ANOVA F-test assessing whether host species explains a significant proportion of the variance in editing level. The analysis was performed using basic R functions, and the resulting *p*-values were used to determine host-dependent RNA editing events.

To define DE RES, we applied two criteria: a statistically significant difference in editing levels across hosts ($P < 0.05$), and a fold-change in editing level of ≥2 between at least two host conditions.

**Verification of the candidate RES by Sanger sequencing**

To validate whether the candidate sites are edited in *M. persicae*, primers were designed with 150-200 bp flanking sequences upstream and downstream of the candidate sites (Figure 2B). PCR amplification was performed using cDNA and genomic DNA (gDNA) as templates, followed by Sanger sequencing of the amplified products. For cDNA synthesis, 1 µg of total RNA was reverse transcribed using the PrimeScript™ RT Reagent Kit with gDNA Eraser Kit (TaKaRa, RR047A, Japan) according to the manufacturer's instructions. The PCR reaction was typically conducted in a 25 µL volume, containing Phanta® Max Super-Fidelity DNA Polymerase (Vazyme, P505, Nanjing, China), 100 ng of cDNA (or 10 ng of gDNA) template, and 10 µM each of forward and reverse primers. The PCR program was set as follows: 95°C for 3 minutes for initial denaturation, followed by 38 cycles of 95°C for 15 seconds, 57°C for 15 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 5 minutes. Primers were synthesized by Shanghai Sangon Biotech Co., Ltd., and Sanger sequencing was performed by Wuhan Quintarabio Biotech Co., Ltd. The presence of RNA editing was visualized using SnapGene.

**RES-specific qPCR**

To assess the variation in editing levels of candidate sites across different hosts, we propose an efficient and convenient quantification method—RES-specific qPCR (RES-qPCR). This method exploits the mismatch effect between the template and primers to influence the DNA polymerase amplification efficiency, leading to changes

in fluorescence signals, thereby enabling the quantification of different transcript variants. The primer design strategy is adapted from genome-wide allele-specific SNP detection methods (22, 23). For each candidate editing site, a set of primers is designed, including two specific editing primers and one universal primer (Figure 3A). When the specific editing primer serves as the forward primer, the universal primer is designed as the reverse primer, and vice versa. The two specific editing primers differ only at the first base of the 3' end, allowing effective discrimination between edited and unedited sites. To further enhance primer specificity, a second mismatch is introduced at the 3' end of the specific editing primers, either at the first or second position upstream. The rules for introducing mismatched bases are as follows: a “strong” mismatch at the 3' end of the specific editing primer requires a “weak” mismatch at the second position, and vice versa; while a “medium” mismatch at the 3' end should be paired with a “medium” second mismatch (22, 24) (Table S2). Such intentional mismatches can increase the quantification cycle (Ct) by approximately 3–9 cycles, thereby enabling effective discrimination between edited and unedited templates while maintaining measurable amplification for quantitative analysis. The reaction system and protocol are performed on the Bio-Rad CFX Connect™ Real-Time System (Bio-Rad, USA). The reaction mixture is 20 µL, containing Hieff® qPCR SYBR® Green Master Mix (YEASEN, 11201ES05, Shanghai, China), 100 ng cDNA, and 10 µM of primers. The PCR conditions (three-step protocol) are as follows: 95°C for 5 minutes for pre-denaturation, followed by 41 cycles of 95°C for 10 seconds, 62°C for 20 seconds, and 72°C for 20 seconds.

For the RES-qPCR results, after the reaction is complete, the amplification products are analyzed by 1.5% agarose gel electrophoresis and Sanger sequencing. The primers that meet the expected criteria are selected for subsequent data quantification. The standards are as follows: (1) When cDNA is used as the template, both the unedited- and edited-stie primers should amplify normally, producing a single band, while no amplification product should be observed when gDNA is used as the template; (2) Sanger sequencing results should confirm the presence of both the edited and unedited transcript forms, along with the introduced mismatch bases.

The relative expression levels of RNA-editing transcripts are calculated using the $2^{-\triangle Ct}$ method, where the relative expression of RNA-editing transcripts is normalized to the gene expression level of the RES in the sample, while other data

are normalized to the expression level of the nuclear elongation factor 1-alpha (*EF1α*) gene.

**Gene expression quantification**

Transcript abundance was also quantified as TPM (transcripts per million) using TPMCalculator (v0.0.3) (25). BAM files generated by RMTA and the corresponding GTF annotation files were provided to TPMCalculator (-b for BAM, -g for annotation).

**Analysis of RNA editing enzymes**

Protein annotation files of 21 arthropod species were retrieved from the NCBI public database (Table S3). The Pfam-A database (downloaded on 2025-08-03) was used to identify conserved domains. In total, 35 relevant domains were iok selected, including members of the deaminase superfamily clan CL0109 (33 families), the double-stranded RNA-binding domain (PF00035), and Z-DNA/RNA-binding domains (PF02295 and PF03178) (Dataset S4). Domain annotation was performed with hmmscan in HMMER (version 3.2) (parameters: --cut_ga, domain E-value ≤ 1e-5, coverage ≥ 0.5) (26), retaining only the highest-scoring annotation in cases of overlapping hits. Candidate sequences were further validated with InterProScan to refine domain boundaries and eliminate false positives (27). For genes with multiple isoforms, only the longest transcript was retained, and truncated sequences shorter than 100 amino acids were discarded. Multiple sequence alignments of candidate proteins were generated with MUSCLE (version 5.3) (28), and alignments were trimmed with trimAl (version 1.5.rev0) (-automated1) (29) to remove highly gapped columns and low-quality sequences. Phylogenetic trees were constructed with IQ-TREE (version 2.4.0) using ModelFinder (-m MFP) and branch supports were assessed with 1,000 ultrafast bootstrap replicates (-bb 1000) and 1,000 SH-aLRT replicates (-alrt 1000) (69). Finally, the resulting trees were visualized and refined with iTOL (https://itol.embl.de/) (30).

**Double-stranded RNA synthesis**

Double-stranded RNA (dsRNA) of *ADAR1* and *ADAR2* was synthesized in vitro using the T7 High Yield RNA Transcription Kit (Vazyme, DD4201, Nanjing, China) according to the manufacturer's instructions. Gene-specific primers containing the T7

polymerase promoter sequence were designed using the DSIR website (http://biodev.extra.cea.fr/DSIR/DSIR.html). All synthesized dsRNAs were dissolved in nuclease-free water and quantified by a K5600C micro-volume spectrophotometer (KAIAO, Beijing, China)

**RNAi experiments**

The feeding treatment was carried out according to the method described in (63), with minor modifications. Double-layered stretched parafilm (Parafilm, PM996, USA) was used to cover acrylic rings (25 mm in diameter × 15 mm in height), constructing the feeding apparatus. Each device was filled with 50 μL of artificial feed mixed with dsRNA, which included a final concentration of 30% sucrose solution, 0.02% neutral red dye, and 1 μg/μL of dsRNA.

To assess the impact of ADARs on *M. persicae* host colonization, 20 synchronized third-instar GPA reared on Bn were fed with *dsADARs* (*dsADAR1* and *dsADAR2*) for 48 hours. The surviving aphids were then transferred to Bn plants for a 10-day reproductive capacity assessment. *dsLta* (Lymphotoxin A gene of *Mus musculus*) was used as a control (33). The interference efficiency was evaluated via qPCR. Each biological replicate was independently tested three times (n = 3), and for each biological replicate, two technical replicates were performed.

**Host plant change experiments**

Approximately 150 adult aphids from GPA-Bn and GPA-Zm were collected and introduced onto 4-week-old host plants. After 24 hours, adults were removed, while the newly hatched nymphs remained on the plants for three days. Thirty synchronized 3-day-old aphids were then transferred simultaneously to both their original and novel host plants. This transfer procedure was repeated at least three times, with each transfer serving as an independent biological replicate. To prevent escape, aphids were confined to the plant leaves within foam clip-cages (2 cm height × 3.5 cm diameter), securely sealed at the base of the petiole. Two days after the transfer, live nymphs from both original and novel plants were collected for RNA extraction. Each biological replicate was subjected to two technical replicates of quantitative reverse transcription PCR (qRT-PCR) analysis.

To investigate the effects of gene silencing of *ADARs* on RNA editing and rapid host adaptation of *M. persicae*, GPA-Bn 3-day-old nymphs were fed *dsADAR1* and

*dsADAR2* (1 μg/μL) for 48 hours. Control groups included nymphs fed with *dsLta*. After the feeding period, the same transfer procedure was performed, and subsequent experiments included: (a) quantification of candidate RNA editing site (RES) levels and related gene expression; (b) measurement of fecundity and survival rate over 10 days. All experiments were performed independently three times, and for each biological replicate, two technical replicates were included for quantitative measurements.

**Statistical analysis**

All analyses were performed in R. Two-group comparisons (e.g., dsRNA-treated vs. control) for RNA editing levels, transcript ratios, gene expression, and fecundity were conducted using Wilcoxon rank-sum tests. Differences in ADAR TPM across hosts were analyzed by one-way ANOVA, with normality and homogeneity of variances assessed using the Shapiro–Wilk and Levene's tests, respectively; post-hoc pairwise comparisons were performed using Tukey's HSD or Games–Howell as appropriate. Survival curves were compared using the log-rank test. Host-dependent RNA editing per site was assessed by linear regression (lm(value ~ hosts)), with ANOVA F-tests determining significance ($P < 0.05$ and fold-change $\geq 2$). The distribution of A-to-I RNA editing sites across genomic regions and the proportions of synonymous and nonsynonymous sites within CDS regions were evaluated using Pearson's chi-square tests, with post-hoc pairwise comparisons where appropriate

**Supplementary Figures**

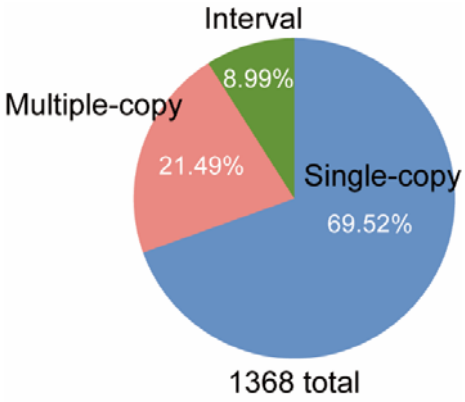


**Fig. S1 Proportional distribution of candidate RES across single-copy, multi-copy, and intergenic regions.**

The pie chart shows the proportion of 1,368 sites with percentages indicated. Subsequent analyses focused only on single-copy genes to avoid potential artifacts from unassembled duplicates or tandem repeats.

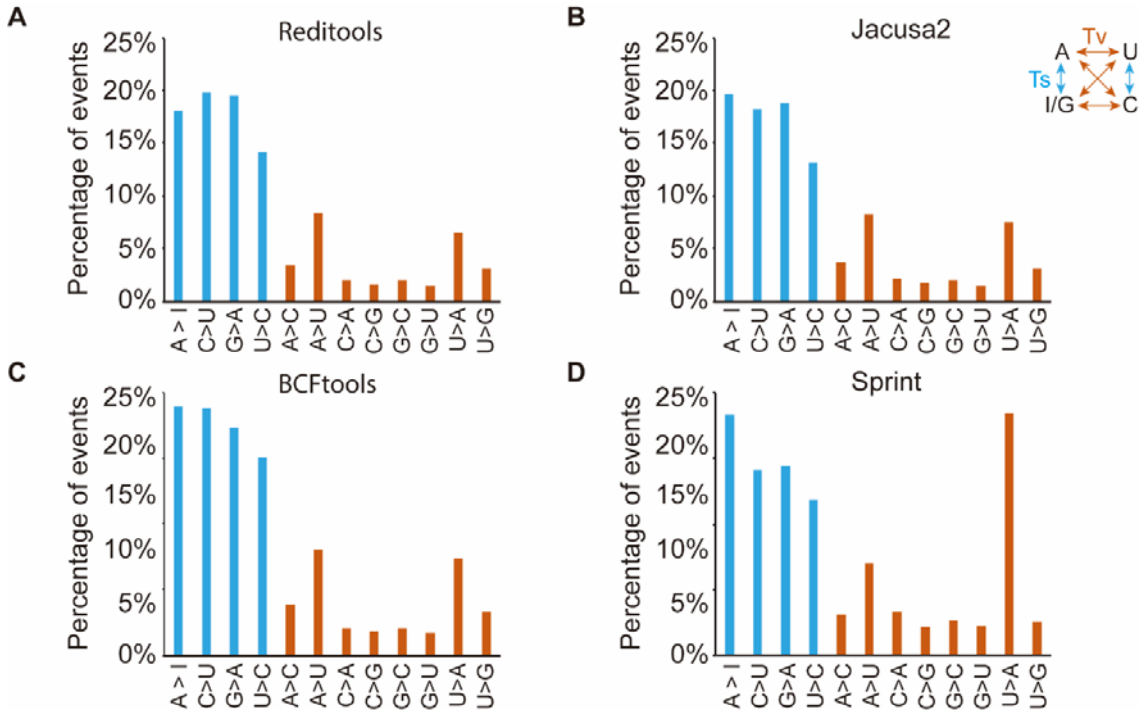


**Fig. S2 Distribution of the 12 possible nucleotide substitution types among all RES identified by four bioinformatic software tools.**

The 12 substitution types were counted based on RES detected by the four tools following the workflow in Fig. 1D, showing that the results are largely consistent with those in Fig. 1F.

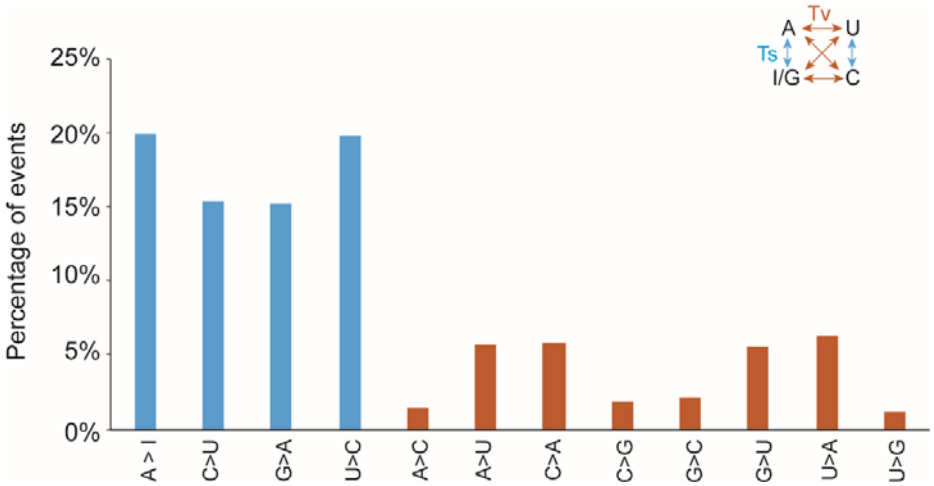


**Fig. S3 Proportion of 12 types of RES.**

RNA-seq data (accession: ERR1661483) (1) from *M. persicae* Clone O reared on *B. rapa* was used for the analysis. RES were identified using Reditools.

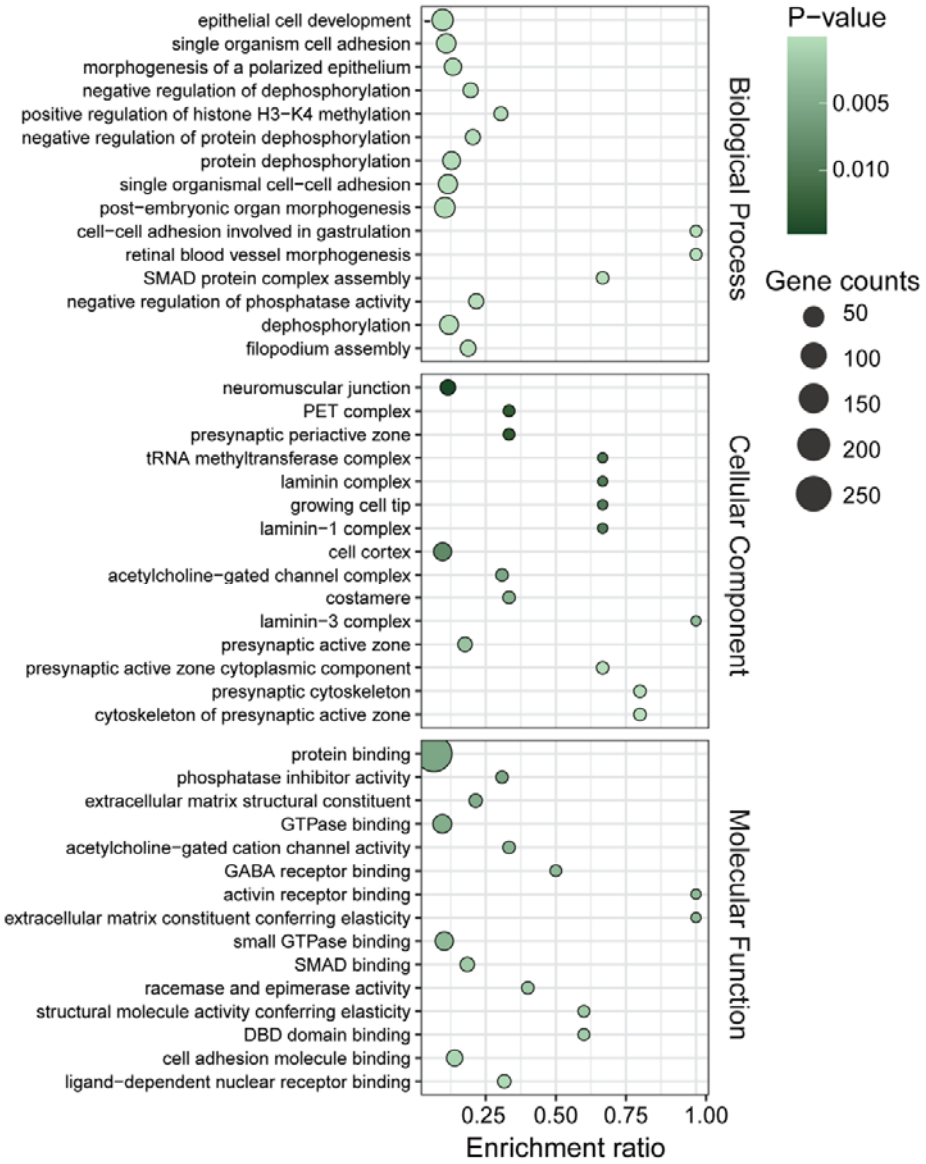


**Fig. S4 GO term enrichment analysis of genes within RESs located coding region.**

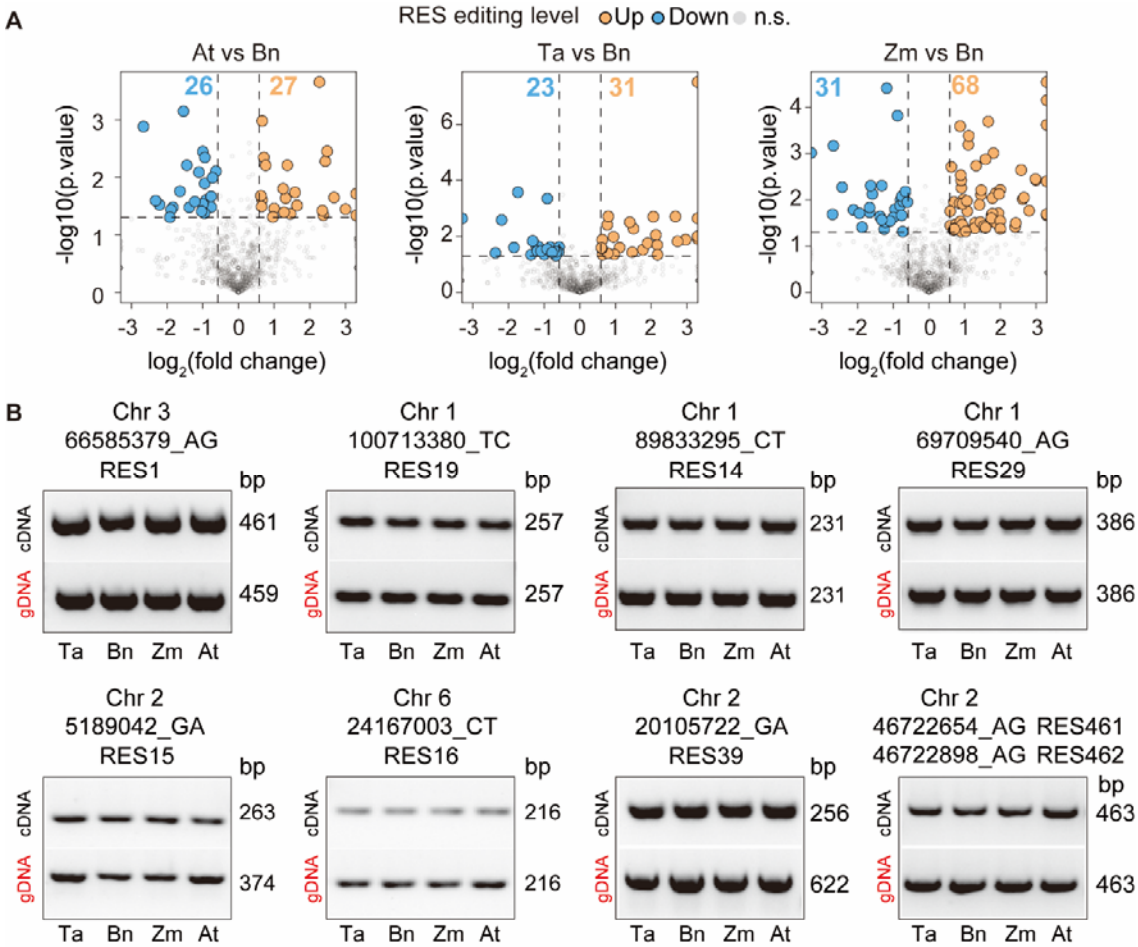


**Fig. S5 RES are differentially edited in *M. persicae* on different plant hosts.**

(A) Volcano plot showing differential RNA editing between aphids reared on various host plants. Editing levels in aphids on Bn were used as the control. DE RES were defined as those with a ≥2-fold change in editing level and $P < 0.05$ (ANOVA F-test). (B) Gel electrophoresis image of PCR amplicons containing selected RES. Forward and reverse primers were designed to flank the target RES. PCR was performed using both cDNA and genomic DNA (gDNA) templates, followed by Sanger sequencing.

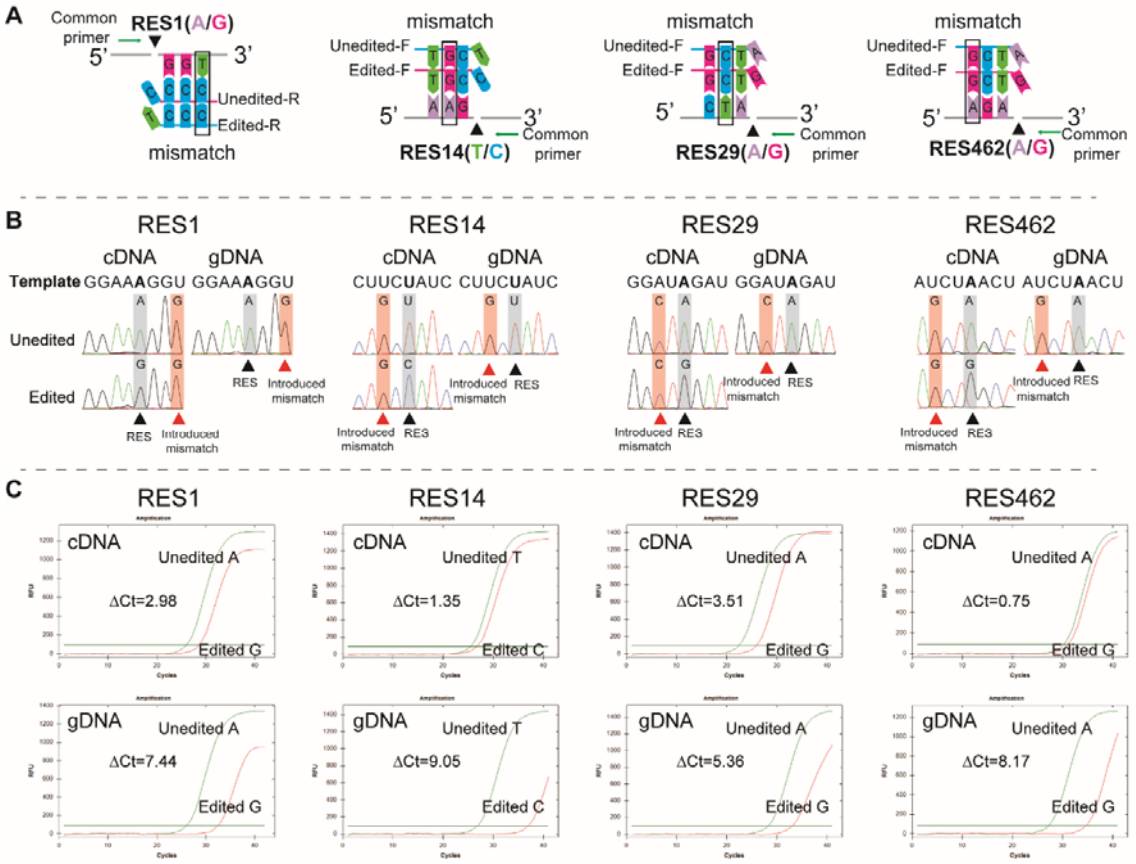


**Fig. S6 Validation of RES-qPCR for distinguishing edited and unedited transcripts.**

(A) Schematic illustration of RES-qPCR primer design. The forward primers contain the edited or unedited nucleotide at the 3′ end and include an additional mismatch 1–2 bases upstream (black boxes), paired with a common reverse primer.

(B) Representative Sanger sequencing chromatograms of PCR products amplified from cDNA and gDNA templates for four selected RES (RES1, RES14, RES29, and RES462). Arrows indicate RES (black) and the introduced mismatch nucleotide (red).

(C) Amplification curves of edited and unedited primers using cDNA and gDNA templates. Green curves correspond to amplification with unedited primers, and red curves correspond to amplification with edited primers. ΔCt values indicate differences in amplification efficiency between the two primer sets. In cDNA templates containing edited transcripts, the assay clearly distinguishes between edited and unedited transcripts. In contrast, amplification from gDNA templates is strongly suppressed due to 3′-end mismatches and the presence of artificially introduced sequence alterations.

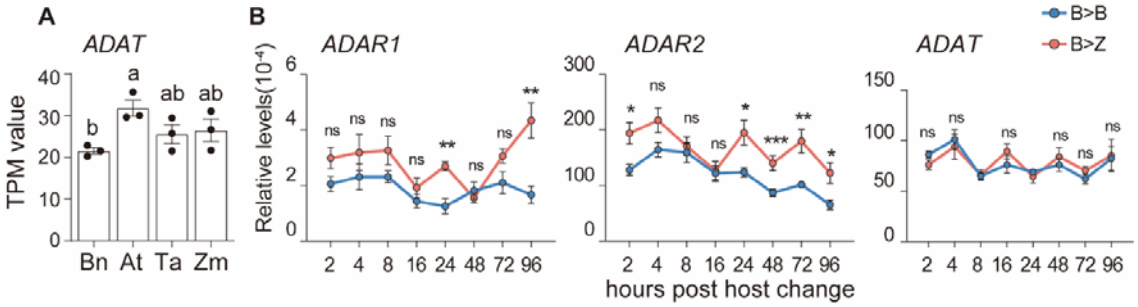


**Fig. S7. Dynamic gene expression and RNA editing in aphids after host transfer.**

(A) TPM of *ADAT* in *M. persicae* reared on different host plants. Different lowercase letters indicate significant differences (one-way ANOVA with Tukey's test). (B) Expression levels of *ADAR1*, *ADAR2* and *ADAT* at different time points after transfer from Bn to Zm**.** Data are presented as mean ± SEM (n = 6; 15 aphids per biological replicate). Symbols indicate statistical significance: $P < 0.05$, $P < 0.01$, $P < 0.001$; ns = not significant. Blue lines: B > B; red lines: B >Z.

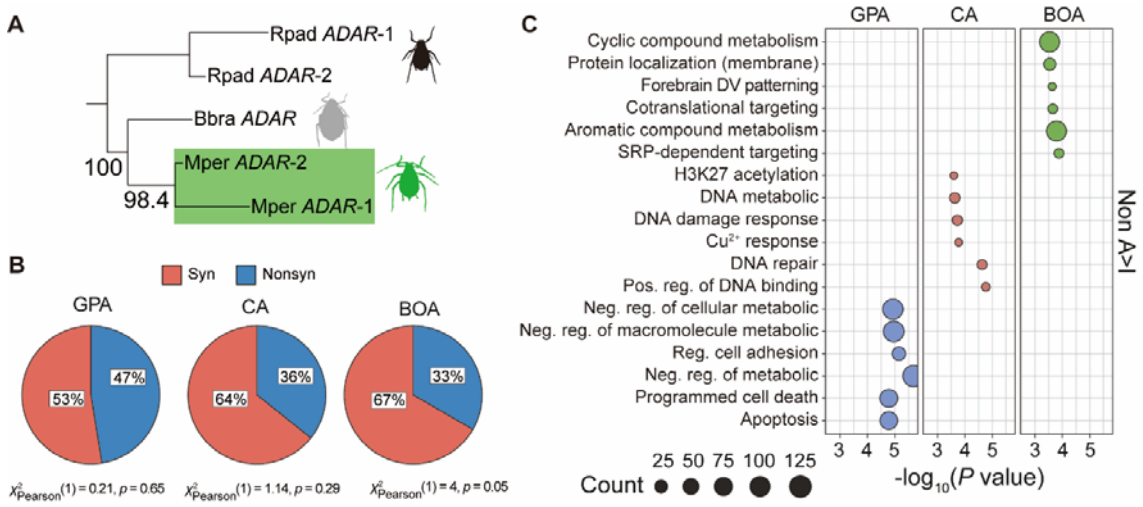


**Fig. S8. Dynamic gene expression and RNA editing in aphids after host transfer.**

(A) Phylogeny of ADAR homologs from *Myzus persicae* (Mper), *Rhopalosiphum padi* (Rpad), and *Brevicoryne brassicae* (Bbra) (B) Proportions of synonymous (Syn) and nonsynonymous (NonSyn) A-to-I editing sites within CDS regions in the three aphid species. Pearson's chi-square test results were used to evaluate differences in proportions among the aphid species. (C) Gene Ontology (GO) enrichment analysis of genes harboring non-A-to-I RES. GO terms were detailed in Table S4. Enriched biological processes are plotted based on gene count (dot size) and significance ($-\log_{10}$ (*P value*))

## Supplementary Tables

**Table S1 Primers used for validation of candidate RNA editing sites (RESs).**

Primers were used to amplify regions containing candidate RESs using cDNA and genomic DNA (as a control) as templates, in order to validate the presence of RNA editing.

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| RES1 | scaffold_3_66585379_AG_F | GCCTGCTCATGCTGTTTCTGG |
| | scaffold_3_66585379_AG_R | GCTAAGGATTGGCGGCGGAATGTAG |
| RES2 | scaffold_3_40961603_CT_F | GTTCTCGTGTGGCAGTTTCGT |
| | scaffold_3_40961603_CT_R | TTAGCAACGTGGCTTTTTTCCGATT |
| RES3 | scaffold_3_22994480_GA_F | CGAAGGCTTGGTACTCGGACAC |
| | scaffold_3_22994480_GA_R | GGTAAACTGCTGTTGGAGGTAT |
| RES4 | scaffold_2_5204493_GA_F | GGATACCATTTTGGCGAAGGC |
| | scaffold_2_5204493_GA_R | AGAACTTGCTCCAGCAGACAT |
| RES5 | scaffold_2_10293789_CT_F | ACGCCACAAACAACGTCATAG |
| | scaffold_2_10293789_CT_R | AGCCGATTTCCCATACACTCT |
| RES6 | scaffold_5_16876236_TC_F | TCTTCGGTGGTTGATGGTGTT |
| | scaffold_5_16876236_TC_R | CCATAATCGTGCAGAGGCAAT |
| RES7 | scaffold_3_19736043_AT_F | GATACCTACCCATCGAAAAGG |
| | scaffold_3_19736043_AT_R | AGCGCGTGATCGTAGCTAACTTTCG |
| RES8 | scaffold_1_14416567_AG_F | GACCCGACACTCACTGTCCAA |
| | scaffold_1_14416567_AG_R | AATCCCGATAAGAGAGCTTCCTGGA |
| RES9 | scaffold_5_11163586_AT_F | GTGGCTATCAACTACCAAAGG |
| | scaffold_5_11163586_AT_R | GTCTGGAATGTGAGGGCAAAG |
| RES10 | scaffold_4_18404091_AG_F | CAGTGCTGCTGCTTCACAAAT |
| | scaffold_4_18404091_AG_R | GTATGGCAGTCACCAGAGGTT |
| RES11 | scaffold_2_60575551_TC_F | GGGAAATGATCCACCCGTAAT |
| | scaffold_2_60575551_TC_cDNAR | TATTGTCCAGTCTGCGGAAGC |

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| | scaffold_2_60575551_TC_gDNAR | CTCTGATGCAGGCATAAGAGTT |
| RES12 | scaffold_4_21107550_AC_F | AGGAGTAGAAGGGTTCCGAGCAT |
| | scaffold_4_21107550_AC_R | TCATCTCCAGTGGCTTTTCAG |
| RES13 | scaffold_6_11706502_CT_F | TGACGGCAACAGACCAAAGTA |
| | scaffold_6_11706502_CT_R | TCAGTTCGGCCACACGATAAG |
| RES14 | scaffold_1_89833295_AG_F | TAAATGCCCAGCCTTGTCAAAC |
| | scaffold_1_89833295_AG_R | CTTCTGCTGTGGGGTTATGGT |
| RES15 | scaffold_2_5189042_CT_F | TTCTGGAGAAATCGCAGGTGGAT |
| | scaffold_2_5189042_CT_gDNAR | GCAAGGCTTGTGATTGGAGTGAA |
| | scaffold_2_5189042_CT_cDNAR | TATCCTGTGGCAGCTTAGGTGTA |
| RES16 | scaffold_6_24167003_CT_F | AGCCAAAGAAGGCAGTATTGAGG |
| | scaffold_6_24167003_CT_R | TACAGTTCGTACAGAGAGGGAGT |
| RES17 | scaffold_2_11864789_AG_F | CGACGATTACCGCACGTTATTTC |
| | scaffold_2_11864789_AG_R | CAGCTCCGCTTACCTGTTTATGTGT |
| RES18 | scaffold_4_15707182_TC_F | CTTACTCAACGGCGTCCACAGAT |
| | scaffold_4_15707182_TC_R | CCCTGGCACTACTTTTTCCGGTATC |
| RES19 | scaffold_1_100713398_TC_F | CACTGATCCTGTAGCAGCCTTAC |
| | scaffold_1_100713398_TC_R | CTTCAGATAATCTCATAAATTGTTC |
| RES20 | scaffold_2_80323514_AC_F | GCATCCAGTGACCGTTGTCC |
| | scaffold_2_80323514_AC_R | TTATCCCTCCTGGCAGCGTGACTCC |
| RES21 | scaffold_6_22030858_CT_F | ACGACCCACACACCTACGAT |
| | scaffold_6_22030858_CT_R | GTTGGTTCACACGGCTAAAT |
| RES22 | scaffold_6_22155452_CT_F | CGGCTTACCCAGCACACCAC |
| | scaffold_6_22155452_CT_R | CGGCTTACCCAGCACACCAC |
| RES23 | scaffold_6_28268790_AG_F | AGACCTCTGTAACCGTGATT |
| | scaffold_6_28268790_AG_R | TGAATAAATTAAAAAATAAAATTGA |
| RES24 | scaffold_2_45703273_AT_F | TCTGAGTTTTGTGGGTTGGC |
| | scaffold_2_45703273_AT_R | TCGTCACACTACGGTGCTTA |

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| RES25 | scaffold_4_20783659_CT_F | GAAAGTTCACGTCGTTAAACCG |
| | scaffold_4_20783659_CT_R | GTTGTGGTAATGTCCTCCTT |
| RES26 | scaffold_5_16511020_GA_F | AACTCACCAAGGCTGCTATT |
| | scaffold_5_16511020_GA_R | CATCAACCACCGAAGAAACT |
| RES27 | scaffold_6_10737055_AT_F | AGGCGTGCGTGATCTTGATT |
| | scaffold_6_10737055_AT_R | GATGGCCTTGGCAGATTGGACTTGT |
| RES28 | scaffold_3_60844485_AG_F | CTCGCGTCTCATAGTTCTTA |
| | scaffold_3_60844485_AG_R | TGGCTCGACTTTCTAGTTCC |
| RES29 | scaffold_1_69709540_AG_F | GGTTACGAAATCTCCCAGTA |
| | scaffold_1_69709540_AG_R | CCTCCAAAGTACAAACATCATCCAA |
| RES30 | scaffold_2_6909982_AG_F | GATGGCGATTGAAGGTAGTG |
| | scaffold_2_6909982_AG_R | ACCAACCCCAATTAGGTCTC |
| RES31 | scaffold_6_25972296_CA_F | GTGTACTCGCTGTTGAACCTAAT |
| | scaffold_6_25972296_CA_R | TTTTTCATTATTATCTTTTTGACCC |
| RES32 | scaffold_2_65920182_AT_F | GCTGATGACTGCTGGAAAGAATA |
| | scaffold_2_65920182_AT_R | TGATGATGACGTTGAAGATACGG |
| RES33 | scaffold_3_50433446_TC_F | TCTTAGAAAATCTAAAGGAAGGT |
| | scaffold_3_50433446_TC_R | AGCCAAGTATTATGAGTCAAATC |
| RES34 | scaffold_6_10507686_TA_F | CGATTCCGGCATAACATATCCTT |
| | scaffold_6_10507686_TA_R | GATTTGATGGCCTTGGCAGATTG |
| RES35 | scaffold_6_10534663_TA_F | CGGTCGGTGGCAGTGTATAAAAG |
| | scaffold_6_10534663_TA_R | ATGGCCTTGGCAGATTGGACTTG |
| RES36 | scaffold_3_38217599_AT_F | CGCCCGTCGCTCGAAGCGTTTAT |
| | scaffold_3_38217599_AT_R | TCCTCAGTCGCCGTCACTGTTTG |
| RES37 | scaffold_6_8887294_CT_gDNAF | CTAATGAACAATCCACAATACCCTTGA |
| | scaffold_6_8887294_CT_cDNAF | GCCGCTACAATGGTCATGTTCAG |
| | scaffold_6_8887294_CT_R | GCTCCTCCATCTTATCGCCGTCT |
| RES38 | scaffold_2_35319513_AG_F | TAAAATGCCCGAGGATGCAGCGTA |

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| | scaffold_2_35319513_AG_R | TGGTCAGCACTTGGTTTGACGAT |
| RES39 | scaffold_2_20105722_CT_F | GACTGCTTCTCAACAAGCCAAAC |
| | scaffold_2_20105722_CT_R | CCATTCTCCTTGTGTCAGTCTTTGG |
| RES40 | scaffold_2_80331586_TC_F | GACCAATCATTAACTTGGCTCAC |
| | scaffold_2_80331586_TC_R | CAAAACAATACTTTCACTTTCTCCT |
| RES41 | scaffold_6_22026693_TA_F | GTACTCTGCGCCCGCTTACTCGGCACCA |
| | scaffold_6_22026693_TA_R | CGACGAGGCTGTAGGATCCCTTGAC |
| RES42 | scaffold_6_10487336_AT_F | CGGTCGGTGGCAGTGTATAAAAG |
| | scaffold_6_10487336_AT_R | ATGGCCTTGGCAGATTGGACTTG |
| RES43 | scaffold_6_10474067_GA_F | ACATCATTTTGTGCAGATAAATTCTTTG |
| | scaffold_6_10474067_GA_R | AATTGGGTGGTGGCTGGGGTGGC |
| RES44 | scaffold_3_19307903_AG_F | TTGCCCAGTTATCGTACACTATT |
| | scaffold_3_19307903_AG_R | GTTTAGGAAGTGATGATGGAATAAA |
| RES45 | scaffold_4_18403885_AT_F | CAGTGCTGCTGCTTCACAAAT |
| | scaffold_4_18403885_AT_R | GTATGGCAGTCACCAGAGGTT |
| RES461 | scaffold_2_46722654_AG_F | GATCATGTGCAGCGCCTTAC |
| | scaffold_2_46722654_AG_R | CGTGGGTTGTCGTGGTCTTA |
| RES462 | scaffold_2_46722898_AG_F | GATCATGTGCAGCGCCTTAC |
| | scaffold_2_46722898_AG_R | CGTGGGTTGTCGTGGTCTTA |
| RES47 | scaffold_6_27881628_AG_F | TAATAGCAAGCCATAATCAG |
| | scaffold_6_27881628_AG_R | AACAAACCAAGACGGCAAAG |
| RES48 | scaffold_1_40764663_AG_F | CAAATGGTGAGAGGAGGTAGT |
| | scaffold_1_40764663_AG_R | TAGATGCTGTTGATAAATTTGAAGT |
| RES49 | scaffold_2_62693591_AG_F | TAAAACTTGATGGTAAGAATA |
| | scaffold_2_62693591_AG_R | ATAAAAAAGAAAATAGGAAAC |
| RES50 | scaffold_5_6912130_AG_F | TGACGTACTTCCGTTATTGGTT |
| | scaffold_5_6912130_AG_R | TACGCCACAATTCGCATACA |
| RES51 | scaffold_3_11529311_AG_F | TCTGTGGTCTCTCTTTGGGTT |

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| | scaffold_3_11529311_AG_R | ATCTGCGGTCGTTAAGTTGAA |
| RES52 | scaffold_2_79652819_AG_F | CTGAATTTGGGTGCTACTTCC |
| | scaffold_2_79652819_AG_R | GTCCATACATACTGTTGGGGT |
| RES53 | scaffold_1_84723037_GA_F | ATCACGTTCAAAAATATGTCAATG |
| | scaffold_1_84723037_GA_R | CAAACGAAGTCGAAGCTGAAGAAG |
| RES54 | scaffold_3_30030503_AG_F | TGGCAGTCCTTATGGCAGTTTGT |
| | scaffold_3_30030503_AG_R | GGTGGTTGTAGTATTGGTTGTTG |
| RES55 | scaffold_3_31396224_TA_F | ACTACAATGCCTGCGGTGAAATC |
| | scaffold_3_31396224_TA_R | GGTAAAGCCCTAACCTATAAAAA |
| RES56 | scaffold_3_31471977_CT_F | AGACACAAATAACTGCATTTTGG |
| | scaffold_3_31471977_CT_R | GCTGCTTCTGATCGTTGACGACC |
| RES57 | scaffold_3_43890171_GA_F | CAGTCCAGTAGCCGCATCTTGGT |
| | scaffold_3_43890171_GA_R | TCTTTTGTTTGAATGTCTCTTTC |
| RES58 | scaffold_3_55627006_TC_F | AGAAAATGTGTGATAAATGGTGGT |
| | scaffold_3_55627006_TC_R | AGTTGGAAATAGTGTTAAGGGAAA |
| RES59 | scaffold_4_32759230_TC_F | CCGTGCGCCTGGTCTTGAAAT |
| | scaffold_4_32759230_TC_R | ATAACGAGTCGAGGACATCATCAGT |
| RES60 | scaffold_4_33121905_CA_F | TGACAGTAATGATGTGGAAGGTA |
| | scaffold_4_33121905_CA_R | TTCTAAAGTCTTTGTTGGAGATG |
| RES61 | scaffold_4_48981178_AG_F | GATTGATTCATTTAATTTTGGGGA |
| | scaffold_4_48981178_AG_R | AACTGATATATGTAGTGGTACCCC |
| RES62 | scaffold_5_14251869_AT_F | TTTTTATTTCTTTCTTTTTATTGT |
| | scaffold_5_14251869_AT_R | GTTGTCGTATTTGTCTTTTACCGC |
| RES63 | scaffold_5_16531209_AG_F | TAAGAGGATCATAGGATCAAAGC |
| | scaffold_5_16531209_AG_R | ACCGGCAACAATTCAATACACAG |
| RES64 | scaffold_6_11459628_GA_F | GCTCTTAATTTACAAATAACTGAA |
| | scaffold_6_11459628_GA_R | GATAACCTTGGAACGCTGGGCTTC |
| RES65 | scaffold_6_22189385_GA_F | CGTCTTACTCCTCACCGTCTT |

| Candidate RES ID | Primers | Sequence (5'→3') |
|---|---|---|
| | scaffold_6_22189385_GA_R | CGTTCTTGCCGTCACTGTATT |
| RES66 | scaffold_6_22266688_CT_F | TGTTCGTGGTCAGCGGTGTAAGT |
| | scaffold_6_22266688_CT_R | ACCATCTCCGAGCACAGCAGTT |
| RES67 | scaffold_1_55074975_TC_F | ATGAAGAGGAAGCCACATCCACAT |
| | scaffold_1_55074975_TC_R | TTAAGAAGACGCCACAACCAGGT |
| RES68 | scaffold_1_58669866_TC_F | TGTTGTCTCTGTCTTACCTACTCAC |
| | scaffold_1_58669866_TC_R | TTCGAGGAAGATACGAGTATGTGAA |
| RES69 | scaffold_1_83194689_TC_F | GCCTTAGTCATCGTGATTGCCTATC |
| | scaffold_1_83194690_TC_R | CGGAGTCATCGTAATTGCCTATCTT |
| RES70 | scaffold_2_23269592_GA_F | GCAATGTTGTAACGCCTCCA |
| | scaffold_2_23269592_GA_R | GATTCAGTTGATTCAGTTGACTCTG |

**Table S2 List of primers used in this study and their experimental purposes.**

Primers were designed for validation of candidate RNA editing sites (RESs), quantification of edited and unedited transcript isoforms, dsRNA synthesis for RNA interference, and qRT-PCR assays of gene expression or RNAi efficiency.

| Primers | Sequence (5'→3') | Purpose |
|---|---|---|
| RES1_1R1 | CGCCGCCGATGCCCT | Primers used for detecting transcript expression levels of edited and unedited RES1 |
| RES1_1R2 | CGCCGCCGATGCCCC | |
| RES1_F | GCCTGCTCATGCTGTTTCTGG | |
| RES14_2F1 | CCCAGCCTTGTCAAACATACTGCT | Primers used for detecting transcript expression levels of edited and unedited RES14 |
| RES14_2F2 | CCCAGCCTTGTCAAACATACTGCC | |
| RES14_R | TCACCTTCTTCAAGATGGCAGCTAA | |
| RES29_2F1 | AAATCTCCCAGTAAATGTTAAAACTTGGCTA | Primers used for detecting transcript expression levels of edited and unedited RES29 |
| RES29_2F2 | AAATCTCCCAGTAAATGTTAAAACTTGGCTG | |
| RES29_R-new | ACCCTCCAAAGTACAAACATCATCC | |
| RES462_1F1 | CTTTCTAGTGTGGACATTTAGTTAGCTA | Primers used for detecting transcript expression levels of edited and unedited RES46 |
| RES462_1F2 | CTTTCTAGTGTGGACATTTAGTTAGCTG | |
| RES462_R | CGTGGGTTGTCGTGGTCTTA | |
| *dsMpADAR1*_Primer_F | TAATACGACTCACTATAGGGTTTGTTACGAAACATCAAGTCACTG | Primers used for dsRNA synthesis of *dsMpADAR1* |
| *dsMpADAR1*_Primer_R | TAATACGACTCACTATAGGGCTGCGGGTCAATATTTCTGCG | |
| *dsMpADAR2*_Primer_F | TAATACGACTCACTATAGGGTTGCCACGCAGAAATATTGA | Primers used for dsRNA synthesis of *dsMpADAR2* |
| *dsMpADAR2*_Primer_R | TAATACGACTCACTATAGGGGACGTTCACCCTGAGCAACT | |

| Primers | Sequence (5'→3') | Purpose |
|---|---|---|
| *dsLta_Primer*_F | TAATACGACTCACTATAGGGCACCCTCTCCACGAATTG | Primers used for dsRNA synthesis of *dsLTA* |
| *dsLta_Primer*_R | TAATACGACTCACTATAGGGTAGAAGATGCTGCTGTTTCA | |
| *qMpADAR1*_F | AAGACATTTACCACCTTCCCTT | Primers used for the detection of RNAi interference efficiency of *dsMpADAR1* |
| *qMpADAR1*_R | GGGTTACAGATAGAGGCAGGC | |
| *qMpADAR2*_F | AGCAAATAAGGCATCTATCCACCT | Primers used for the detection of RNAi interference efficiency of *dsMpADAR2* |
| *qMpADAR2*_R | AATGTCTTCAGTGATGTATTTGTGT | |
| *qMpTAD1*_F | CGGGCTGTGTCTGGAGAAAT | Primers used for the detection of expression level of *MpTAD1* |
| *qMpTAD1*_R | TGGCCAGTGAGACGACTTTC | |
| *qRES1*_F | CCAGCCTTGTCAAACATACT | Primers used for the detection of expression level of *gene_RES1* |
| *qRES1*_R | CTTGTGCGTACCTCAGAAAT | |
| *qRES14*_F | ACTTCGCCGAGAAGGAGGAGC | Primers used for the detection of expression level of *gene_RES14* |
| *qRES14*_R | CAGATACAACAGACGCCGCCGATGA | |
| *qRES29*_F | GCCTTCTCAGGGTGGAGTTGAC | Primers used for the |

| Primers | Sequence (5'→3') | Purpose |
|---|---|---|
| *qRES29*_R | TCTGCTCCTGCCTCAAGTGTCT | detection of expression level of *gene_RES29* |
| *qRES46*_F | CCCGGCCAATGACTGAGTTTGA | Primers used for the detection of expression level of *gene_RES46* |
| *qRES46*_R | AGGGCATGGTTGTTGGAGGTAC | |
| qEF1a__F | CTGATTGTGCCGTGCTTATTG | Primers used as reference genes for qRT-PCR |
| qEF1a__R | TATGGTGGTTCAGTAGAGTCC | |

1119

**Table S3 Genome data source of 22 species used in Phylogenetic analysis of ADARs and ADAT**.

| Species | Order | family | Genus | Abbreviation in tree | Source |
|---|---|---|---|---|---|
| *Tetranychus urticae* | Trombidiformes | Tetranychidae | *Tetranychus* | Turt | GRBI |
| *Atta cephalotes* | Hymenoptera | Formicidae | *Atta* | Acep | GCF_000143395.1 |
| *Tribolium castaneum* | Coleoptera | Tenebrionidae | *Tribolium* | Tcas | GCF_031307605.1 |
| *Daktulosphaira vitifoliae* | Hemiptera | Phylloxeridae | *Daktulosphaira* | Dvit | Aphid Base |
| *Myzus persicae* | Hemiptera | Aphididae | *Myzus* | Mper | (3) |
| *Acyrthosiphon pisum* | Hemiptera | Aphididae | *Acyrthosiphon* | Apis | (34) |
| *Rhopalosiphum padi* | Hemiptera | Aphididae | *Rhopalosiphum* | Rpad | (3) |
| *Rhopalosiphum maidis* | Hemiptera | Aphididae | *Rhopalosiphum* | Rmai | (35) |
| *Aphis gossypii* | Hemiptera | Aphididae | *Aphis* | Agos | (3) |
| *Aphis glycines* | Hemiptera | Aphididae | *Aphis* | Agly | (36) |
| *Melanaphis sacchari* | Hemiptera | Aphididae | *Aphis* | Msac | GCF_002803265.2 |
| *Bemisia tabaci* | Hemiptera | Aleyrodidae | *Bemisia* | Whitefly | GCF_001854935.1 |
| *Rhodnius prolixus* | Hemiptera | Reduviidae | *Rhodnius* | Rpro | (37) |
| *Oncopeltus fasciatus* | Hemiptera | Lygaeidae | *Oncopeltus* | Ofas | (38) |
| *Frankliniella occidentalis* | Thysanoptera | Thripidae | *Frankliniella* | Focc | (39) |
| *Daphnia pulex* | Cladocera | Daphniidae | *Daphnia* | Dpul | GCA_000187875.1 |
| *Leptinotarsa decemlineata* | Coleoptera | Chrysomelidae | *Leptinotarsa* | Ldec | NCBI genome |

| Species | Order | family | Genus | Abbreviation in tree | Source |
|---|---|---|---|---|---|
| *Myzus cerasi* | Hemiptera | Aphididae | *Myzus* | Mcer | Aphid Base |
| *Schlechtendalia chinensis* | Hemiptera | Aphididae | *Schlechtendalia* | Schi | GCA_019022885.1 |
| *Sipha flava* | Hemiptera | Aphididae | *Sipha* | Sfla | GCA_003268045.1 |
| Sitobion miscanthi | Hemiptera | Aphididae | *Sitobion* | Smis | GCA_008086715.1 |
| *Plutella xylostella* | Lepidoptera | Plutellidae | *Plutella* | Pxyl | GCF_000330985.1 |

1121

**Table S4 Detailed GO enrichment analysis of genes harboring RES identified in the three aphid species under host-shift conditions.**

| Description | Count | GeneRatio | Pvalue | Species | type |
|---|---|---|---|---|---|
| cell-type specific apoptotic process | 18 | 0.058 | 6.22E-06 | GPA | A>I |
| apoptotic process | 35 | 0.035 | 3.94E-05 | GPA | A>I |
| programmed cell death | 36 | 0.033 | 7.02E-05 | GPA | A>I |
| regulation of epithelial cell differentiation | 8 | 0.096 | 8.92E-05 | GPA | A>I |
| glial cell fate determination | 3 | 0.5 | 9.87E-05 | GPA | A>I |
| neuron apoptotic process | 11 | 0.061 | 0.000269 | GPA | A>I |
| white fat cell differentiation | 2 | 0.118 | 0.000726 | CA | A>I |
| regulation of chondrocyte differentiation | 2 | 0.083 | 0.001459 | CA | A>I |
| regulation of multicellular organismal development | 10 | 0.006 | 0.00165 | CA | A>I |
| positive regulation of histone modification | 3 | 0.028 | 0.002101 | CA | A>I |
| regulation of developmental process | 11 | 0.006 | 0.002162 | CA | A>I |
| maintenance of imaginal histoblast diploidy | 1 | 1 | 0.002396 | CA | A>I |
| negative regulation of chromatin binding | 3 | 0.5 | 1.04E-05 | BOA | A>I |
| proton-transporting V-type ATPase complex assembly | 2 | 0.5 | 0.000396 | BOA | A>I |
| vacuolar proton-transporting V-type ATPase complex assembly | 2 | 0.5 | 0.000396 | BOA | A>I |
| protein targeting | 11 | 0.026 | 0.000564 | BOA | A>I |
| protein targeting to membrane | 6 | 0.047 | 0.000632 | BOA | A>I |
| female germ-line sex determination | 2 | 0.4 | 0.000656 | BOA | A>I |
| negative regulation of metabolic process | 131 | 0.065 | 2.07E-06 | GPA | Non A>I |
| regulation of cell adhesion | 31 | 0.107 | 7.11E-06 | GPA | Non A>I |
| negative regulation of macromolecule metabolic process | 118 | 0.065 | 1.10E-05 | GPA | Non A>I |
| negative regulation of cellular metabolic process | 114 | 0.065 | 1.18E-05 | GPA | Non A>I |
| programmed cell death | 78 | 0.072 | 1.66E-05 | GPA | Non A>I |
| apoptotic process | 74 | 0.073 | 1.66E-05 | GPA | Non A>I |
| positive regulation of DNA binding | 4 | 0.105 | 1.73E-05 | CA | Non A>I |
| DNA repair | 9 | 0.023 | 2.36E-05 | CA | Non A>I |
| cellular response to copper ion | 2 | 0.4 | 0.00017 | CA | Non A>I |
| cellular response to DNA damage stimulus | 10 | 0.016 | 0.000189 | CA | Non A>I |

| | | | | | |
|---|---|---|---|---|---|
| DNA metabolic process | 11 | 0.014 | 0.000234 | CA | Non A>I |
| histone H3-K27 acetylation | 2 | 0.333 | 0.000254 | CA | Non A>I |
| SRP-dependent cotranslational protein targeting to membrane | 8 | 0.121 | 0.000133 | BOA | Non A>I |
| cellular aromatic compound metabolic process | 106 | 0.03 | 0.000164 | BOA | Non A>I |
| cotranslational protein targeting to membrane | 8 | 0.113 | 0.000224 | BOA | Non A>I |
| forebrain dorsal/ventral pattern formation | 3 | 0.5 | 0.000235 | BOA | Non A>I |
| protein localization to membrane | 21 | 0.054 | 0.000287 | BOA | Non A>I |
| organic cyclic compound metabolic process | 107 | 0.03 | 0.000294 | BOA | Non A>I |

1124

**Table S5 Detailed annotation of candidate RNA-editing sites (RESs).**

| Candidate RES ID | Site | Strand | gene_id | Genomic feture | Annotation |
|---|---|---|---|---|---|
| RES1 | scaffold_3_66585379_AG | + | MYZPE13164_O_EIv2.1_0259980 | CDS | PREDICTED: *Myzus persicae* potassium voltage-gated channel protein Shab (LOC111031521) |
| RES29 | scaffold_1_69709540_AG | + | MYZPE13164_O_EIv2.1_0085790 | 3'-UTR | Zinc finger BED domain-containing protein 4-like |
| RES462 | scaffold_2_46722898_AG | + | MYZPE13164_O_EIv2.1_0171060 | 3'-UTR | PREDICTED *Myzus persicae* origin recognition complex subunit 3 (LOC111040050) |
| RES14 | scaffold_1_89833295_AG | - | MYZPE13164_O_EIv2.1_0099850 | | PREDICTED: *Myzus persicae* uncharacterized LOC111038309, ncRNA |

**Legends for Datasets S1 to S6**

**Dataset S1: List of high-confidence RNA editing sites (RES) identified in *Myzus persicae***

Table 1: RES identified by REDItools2

Table 2: RES identified by JACUSA2

Table 3: RES identified by BCFtools

Table 4: RES identified by SPRINT

Table 5: RES identified by all four tools

**Dataset S2. Differentially edited RNA editing sites (DE-RES) in *Myzus persicae* in response to host plants**

Table 1: DE-RES between At and Bn

Table 2: DE-RES between Ta and Bn

Table 3: DE-RES between Zm and Bn

**Dataset S3: Distribution of deaminase-related Pfam domains across 21 arthropod species**

**Dataset S4: Pfam domain annotations of ADAR-related proteins in *Myzus persicae***

**Dataset S5: RES identified in GPA, CA, and BOA under host-shift treatments.**

Table 1: RES identified in *M. persicae* (GPA) across four host-shift conditions (B > B, B > Z, Z > Z, Z > B)

Table 2: RES identified in *B. brassicae* (CA) under two host-shift conditions (B > B, B > Z)

Table 3: RES identified in *R. padi* (BOA) under two host-shift conditions (Z > Z, Z > B)

## SI References


1. T. C. Mathers *et al*., Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* **18**, 27 (2017).
2. T. C. Mathers *et al*., Sex-specific changes in the aphid DNA methylation landscape. *Mol. Ecol.* **28**, 4228–4241 (2019).
3. T. C. Mathers *et al*., Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol. Biol. Evol.* **38**, 856–875 (2021).
4. Q. Liu *et al*., The salivary proteome of the green peach aphid/peach-potato aphid (*Myzus persicae*) (sulzer, 1776) (hemiptera, aphididae). [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13269257. Deposited 8 August 2024.
5. H. Li, R. Durbin, Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
6. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
7. H. Li *et al*., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
8. G. A. Van Der Auwera *et al*., From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43** 11.10.1–11.10.33 (2013).
9. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
10. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
11. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
12. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

13. B.-E. Wulff, M. Sakurai, K. Nishikura, Elucidating the inosinome: Global approaches to adenosine-to-inosine RNA editing. *Nat. Rev. Genet.* **12**, 81–85 (2011).
14. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
15. E. Picardi, A. M. D'Erchia, A. Montalvo, G. Pesole, Using REDItools to detect RNA editing events in NGS datasets. *Curr. Protoc. Bioinforma.* **49**, 12.12.1–12.12.15 (2015).
16. M. Piechotta, I. S. Naarmann-de Vries, Q. Wang, J. Altmüller, C. Dieterich, RNA modification mapping with JACUSA2. *Genome Biol.* **23**, 115 (2022).
17. M. Lefouili, K. Nam, The evaluation of bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci. Rep.* **12**, 11331 (2022).
18. F. Zhang, Y. Lu, S. Yan, Q. Xing, W. Tian, SPRINT: An SNP-free toolkit for identifying RNA editing sites. *Bioinformatics* **33**, 3538–3548 (2017).
19. P. Cingolani *et al*., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain $w^{1118}$; *iso-2*; *iso-3*. *Fly (Austin)* **6**, 80–92 (2012).
20. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
21. R. Lorenz *et al*., ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
22. S. Ye, An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* **29**, 88e–888 (2001).
23. F. M. You *et al*., BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253 (2008).
24. S. Lefever, F. Pattyn, J. Hellemans, J. Vandesompele, Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clin. Chem.* **59**, 1470–1480 (2013).
25. R. Vera Alvarez, L. S. Pongor, L. Mariño-Ramírez, D. Landsman, TPMCalculator: One-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960–1962 (2019).
26. J. Mistry *et al*., Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).

27. S. R. Eddy, Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).
28. M. Blum *et al*., InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Res*. **53**, D444–D456 (2025).
29. R. C. Edgar, Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* **13**, 6968 (2022).
30. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
31. B. Q. Minh *et al*., IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534 (2020).
32. V. Bilgi, J. Fosu-Nyarko, M. Jones, Using vital dyes to trace uptake of dsRNA by green peach aphid allows effective assessment of target gene knockdown. *Int. J. Mol. Sci.* **18**, 80 (2017).
33. X. Pei *et al*., Modulation of fatty acid elongation in cockroaches sustains sexually dimorphic hydrocarbons and female attractiveness. *PLOS Biol.* **19**, e3001330 (2021).
34. Y. Li, H. Park, T. E. Smith, N. A. Moran, Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol. Biol. Evol.* **36**, 2143–2156 (2019).
35. W. Chen *et al*., Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* fitch). *GigaScience* **8**, giz033 (2019).
36. J. A. Wenger *et al*., Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.* **123**, 102917 (2020).
37. R. D. Mesquita *et al*., Genome of *Rhodnius prolixus*, an insect vector of chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci.* **112**, 14936–14941 (2015).
38. K. A. Panfilio *et al.*, Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol.* **20**, 64 (2019).
39. D. Rotenberg *et al*., Genome-enabled insights into the biology of thrips as crop pests. *BMC Biol.* **18**, 142 (2020).