Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior

Li, Qiye; Wang, Zongji; Lian, Jinmin; Schiøtt, Morten; Jin, Lijun; Zhang, Pei; Zhang, Yanyan; Nygaard, Sanne; Peng, Zhiyu; Zhou, Yang; Deng, Yuan; Zhang, Wenwei; Boomsma, Jacobus Jan; Zhang, Guojie

Published in:
Nature Communications

DOI:
10.1038/ncomms5943

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):

Download date: 27. jan. 2019
Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior

Qiye Li1,2,*, Zongji Wang1,2,*, Jinmin Lian2,*, Morten Schiøtt3, Lijun Jin2, Pei Zhang2, Yanyan Zhang4, Sanne Nygaard3, Zhiyu Peng4, Yang Zhou1,2, Yuan Deng2, Wenwei Zhang4, Jacobus J. Boomsma3 & Guojie Zhang2,3

Eusocial insects have evolved the capacity to generate adults with distinct morphological, reproductive and behavioural phenotypes from the same genome. Recent studies suggest that RNA editing might enhance the diversity of gene products at the post-transcriptional level, particularly to induce functional changes in the nervous system. Using head samples from the leaf-cutting ant Acromyrmex echinatior, we compare RNA editomes across eusocial castes, identifying ca. 11,000 RNA editing sites in gynes, large workers and small workers. Those editing sites map to 800 genes functionally enriched for neurotransmission, circadian rhythm, temperature response, RNA splicing and carboxylic acid biosynthesis. Most A. echinatior editing sites are species specific, but 8–23% are conserved across ant sub-families and likely to have been important for the evolution of eusociality in ants. The level of editing varies for the same site between castes, suggesting that RNA editing might be a general mechanism that shapes caste behaviour in ants.

1 School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China. 2 China National GeneBank, BGI-Shenzhen, Building No. 11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China. 3 Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. 4 BGI-Shenzhen, Shenzhen 518083, China. * These authors contributed equally to this work. † Present address: Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen 1350, Denmark. Correspondence and requests for materials should be addressed to G.Z. (email: zhanggj@genomics.cn) or to J.J.B. (email: jjboomsma@bio.ku.dk).
nts have highly organized eusocial colonies consisting of morphologically differentiated castes that specialize in egg-laying (queens) or non-reproductive functions such as nursing, nest-building, foraging and defense (workers and occasionally soldiers). Caste phenotypes represent distinct developmental pathways that are plastic during early larval development and produce adults that vary in size, physiology, life span and behaviour. Previous studies have suggested that epigenetic factors may be more important for caste differentiation than originally appreciated. Recent studies in honeybees and ants have also uncovered details of differential gene expression, DNA methylation and histone modification associated with profound behavioural differences between eusocial castes.

However, post-transcriptional gene regulation via RNA editing has remained unexplored in social insects, in spite of its known importance for nervous system functions that have been documented in model species such as nematodes, fruit flies, mice and humans. RNA editing alters RNA sequences through individual base substitutions, insertions or deletions, and thus enhances gene product diversity. The most prevalent type of RNA editing in the animal kingdom is A-to-I editing, where adenosine (A) residues are converted to inosine (I) catalysed by adenosine deaminase (ADAR) enzymes that use double-stranded RNAs (dsRNAs) as substrate. As inosine is recognized as guanosine by the ribosome during translation, A-to-I editing in protein-coding RNAs results in amino acid changes that alter the functional properties of proteins. A-to-I editing can also play an important role in regulating gene expression by affecting alternative splicing, editing microRNA (miRNA) sequences or changing miRNA target sites in messenger RNA. Drosophila with an ADAR-null mutation preventing A-to-I editing exhibit severe behavioural and locomotory abnormalities, while Caenorhabditis elegans with homozygous deletions of both adr-1 and adr-2 show chemosensory defects. The effects of preventing RNA editing in mammals are even more striking, Adar1−/− mice die in the embryonic stage and Adar2−/− mice suffer profound epileptic seizures and die shortly after birth. The recent application of next-generation sequencing technology has greatly expanded the genome-wide detection of RNA editing events and demonstrated that RNA editing of genomic information is pervasive.

RNA editing appears to frequently target transcripts that encode proteins involved in the propagation of fast electrical and chemical signals in animal nervous systems, such as voltage-gated ion channels, ligand-gated receptors and key components of the pre-synaptic release machinery. This was recently confirmed by a study in octopuses, which showed that levels of RNA editing were associated with the function of K+ channels mediating adaptation to temperature change. As the nervous systems coordinate animal behaviours in response to both internal and external stimuli, there are multiple reasons to expect that RNA editing of genes expressed in the nervous systems of eusocial insects might provide an important regulatory mechanism for behavioural differences among eusocial castes.

We used the leaf-cutting ant Acromyrmex echinatior to investigate the characteristics and putative roles of RNA editing in eusocial insects. This ant species has three female castes: small workers, large workers and queens (referred to as gynes when still winged and unmated). A. echinatior belongs to the crown group of the attine fungus-growing ants and has the peculiar characteristics associated with this group, such as a diet almost exclusively derived from a specialized fungus-garden symbiont, foraging on fresh leaves that are processed in a conveyor-belt manner and additional symbioses with antibiotic-producing bacteria. We recently sequenced the genome of A. echinatior and produced a high-quality assembled and annotated draft genome, which provided us with the opportunity to further investigate possible genome-wide mechanisms of caste differentiation. Here we use strand-specific RNA-Seq on head tissue samples of the female castes of A. echinatior and data from other genome-sequenced ants to evaluate the possible role of RNA editing in modulating functional gene differentiation among castes, and show that RNA editing may be an important but underappreciated mechanism that shapes caste behaviour in ants.
To identify candidate RNA-editing sites, we designed a statistical framework to detect sites that were homozygous for genomic DNA (gDNA), but heterozygous for transcripts (see Methods). The orientation information provided by strand-specific RNA-Seq then allowed us to determine the RNA–DNA base differences independent of existing genomic annotations. We identified an average of 10,715 editing sites (range 8,146–13,823) per sample (Supplementary Table 3), with an average of 969 genes (range 812–1,125) being edited in one or more sites (Supplementary Table 4). As observed in humans17–20, mice21 and fruit flies2,4,15, most (up to 97%) of the editing sites represented A-to-I editing (Fig. 2a and Supplementary Table 3), indicating that non-A-to-I RNA editing is rare in A. echinatior.

The median editing level was rather low (12.6%) and only ca. 8% of the editing sites had editing levels (the ratio of RNA-Seq reads with editing versus total RNA reads) over 50% (Fig. 2b, also see Supplementary Fig. 3 for separating A-to-I and non-A-to-I editing sites). Consistent with the known properties of ADAR enzymes22, A-to-I editing sites were concentrated in potentially dsRNA regions, relative to random controls in A. echinatior (Supplementary Methods). Similar to observations in humans18,19 and fruit flies15, the A. echinatior bases in upstream and downstream positions adjacent to A-to-I editing sites (−1 and +1 positions) had a deficiency and excess of guanines, respectively (Supplementary Fig. 3e and Supplementary Methods). In addition, A-to-I editing sites clearly occurred in clusters (Fig. 2d,e, Supplementary Table 5 and Supplementary Methods). In addition, A-to-I editing sites were less likely to be clustered compared with editing sites in other genomic elements (Fig. 2d), which is unlikely to be simply due to the low number of CDS-editing sites, as 5’-UTRs harbour similarly low numbers (Fig. 2c). We also found that simultaneous editing of genes in UTRs, CDSs and introns was rare, and that only a few genes were in fact edited in more than one of these genomic regions (Fig. 2f).

Up to 32% of A-to-I editing sites occurred in transposable elements (TEs), a figure that was four times higher than the background expectation of 8% (Fig. 2c). The fact that TE transcripts are another major target of RNA editing in A. echinatior is interesting, because only one case of RNA editing in TE transcripts has been reported in insects: A-to-I, U-to-C and C-to-U editing in read-through transcripts of the KP elements of Drosophila melanogaster35. In humans, A-to-I editing is pervasive in Alu repeats19, a primate-specific short interspersed element that represents about 10% of the entire human genome37. Our data thus provide the first evidence for substantial RNA editing in
insect TEs, and may imply that the activity of a significant fraction of A. echinatior TEs is regulated by RNA editing.

Functional editing sites shared across colonies. Comparing the A-to-I editing sites among samples of the same caste from different colonies, we found that ca. 60% (range 43–76%) of the caste-specific editing sites were shared by the three sampled colonies (Fig. 3a–c). This indicates a marked disparity of RNA-editing events among sympatric colonies of the same species, consistent with the limited reproducibility of editing events among different human individuals17,20. This variability may be due to colony-specific habitat conditions, colony size, colony age or even random expression noise. However, as our focus was on the editing sites most likely to contribute to caste-specific functions, we continued our analyses using only those sites where editing was confirmed in samples of a caste across all three colonies. This allowed us to identify 4,682, 5,049 and 9,120 caste-specific A-to-I editing sites for gynes, large workers and small workers, respectively (Fig. 3a–c), representing a total of 10,282 numbers of small worker-specific editing sites could not be explained by variation in sequencing depth among samples (Supplementary Fig. 5), but corresponded with the higher expression level of ADAR in small workers (Fig. 1c).

We identified 533, 538 and 779 genes that harboured at least one consistently edited site in gynes, large workers and small workers, respectively, corresponding to a total of 835 genes (Supplementary Data 2). The most highly edited gene was Ach_13462, with >70 consistently edited sites in its 3′-UTR in all three castes (Fig. 2e and Supplementary Data 2). This gene is an orthologue of Mdh1 (malate dehydrogenase 1), which has a conserved role in aerobic energy production28. The high expression level of this gene in the heads of Acromyrmex leaf-cutting ants (mean RPKM (reads per kilobase of transcript per million reads mapped) across 9 samples: 264; Supplementary Methods) probably implies that the aerobic energy requirements for brain activity are extremely high, and extensive RNA editing of its 3′-UTR may reflect a continuous need for fine-tuning gene expression related to energy metabolism. Another gene with a considerable number of consistently edited sites was Ach_00116, an orthologue of Hsc70-4 (heat shock 70-kDa protein cognate 4) in Drosophila, which is involved in nerve-evoked neurotransmitter exocytosis, a process critical for signal transmission across synapses39. It had >50 consistently edited sites in its 3′-UTR in all 3 castes of A. echinatior (Supplementary Data 2), which may offer a special mechanism for post-transcriptional regulation of this gene.

Gene Ontology (GO) enrichment analysis for the 835 genes, which were consistently edited in at least 1 caste (using all expressed genes as background), showed that 5 functional categories were over-represented among the RNA-edited genes, including neurotransmission, circadian rhythm, response to...
temperature stimulus, RNA splicing and carboxylic acid biosynthesis (Fig. 3d; see Methods). Functions in neurotransmission, including GO terms of synaptic growth at neuromuscular junctions, synaptic transmission and cell signalling, are consistent with the known functions of RNA editing in animal nervous systems. For example, we found RNA editing in a series of genes encoding voltage-gated ion channel proteins, such as Shab (K⁺ channel), Shaw (K⁺ channel), Hk (K⁺ channel), NaCP60E (Na⁺ channel) and Ca-alpha1T (Ca²⁺ channel), as well as some neurotransmitter receptor-related proteins, such as Snap25 (Synapse protein 25), n-syb (n-synaptobrevin), Sytbeta (Synaptotagmin beta) and Syp12 (Synaptotagmin 12) (Supplementary Data 2). Circadian rhythms have rarely been reported to be affected by RNA editing, although the previous discovery of RNA editing of the qvr and Sh transcripts in Drosophila is consistent with this possibility. Yet, in A. echinatior we observed editing of many circadian rhythm genes: including qvr (quiver), Shaw (Shaker cognate v), lark, nocte (no circadian temperature entrainment), Af-2 (Activating transcription factor-2), cye (cycle), ctrip (circadian triph), ChIfz (casein kinase Ifz) and Rdl (Resistant to dieletrin) (Supplementary Data 2). Extensive RNA editing of circadian rhythm genes suggest that ants may possess a novel mechanism for adjusting circadian sleep/wake cycles with very short response times.

We only identified 27, 38 and 53 CDS-editing sites with amino acid recoding potential that were consistently edited in gynes, large workers and small workers, respectively, representing a total of 57 sites targeting 34 genes (Supplementary Data 3). Thirty-three per cent (19/57) of these recoding sites were located within or near (within 20 amino acids) known protein domains (Supplementary Data 3). The majority of these recoded genes were implicated in neural system functions such as synaptic transmission (for example, Caps and qvr), neurogenesis (for example, ctrip and Ars2), neuromuscular junction development (for example, TBPH and wath) and circadian rhythm (for example, qvr and ctrip). There were also some genes involved in mitotic processes (for example, Rpl7 and Mtor), RNA splicing (for example, CG7564 and CG2926) and organism development (for example, rhea and scrA) (Supplementary Data 3).

Conservation of RNA editing in ants. RNA editing may be conserved at two levels: the conservation of genomic sites corresponding to RNA-editing sites (DNA level) and the conservation of the RNA-editing events themselves (RNA level). We recently sequenced and assembled the genomes of another five attine fungus-growing ant species (Nygaard et al., in preparation) which, together with the published genomes of A. echinatior and Atta cephalotes, span the most recent ca. 25 MY of the attine ant lineage that originated ca. 50 MYA. Combining these data with the published genomes and transcriptomes of two non-fungus-growing ants, C. floridanus (Florida carpenter ant) and H. saltator (Jerdon’s jumping ant) that diverged from A. echinatior ca. 80 and 100 MYA, we had the opportunity to estimate the degree of conservation of RNA editing in ants over long evolutionary time scales.

Whole-genome alignment analysis (see Methods) showed that 67% of all transcribed ‘A’s (that is, ‘A’s covered by more than a single RNA read in at least one caste) in A. echinatior were highly conserved (that is, displaying the same DNA base) among the seven attine ant species, while 57% and 46% were conserved when comparing A. echinatior with C. floridanus and H. saltator, respectively. However, when we assessed the editing sites that appeared in at least one caste across all three colonies, we found that only 23% of these RNA-editing sites were conserved across the seven attine genomes and about 11% and 8% between A. echinatior and C. floridanus or H. saltator, respectively. The latter conservation percentages were all significantly lower than the former genome background percentages (Monte Carlo test $P < 10^{-4}$). Similar patterns were observed when we extended these comparisons to the separate genomic elements (Fig. 4a, Supplementary Fig. 6 and Supplementary Table 7).

We also estimated the conservation status of the genomic sites with RNA editing by using phastCon scores (base-by-base conservation scores ranging from 0 to 1) derived from the multiple genome alignments for the seven attine ant species (see Methods), and observed that the mean phastCon scores of genomic sites corresponding to RNA-editing sites were significantly lower than those of random genomic sites (Supplementary Table 8). Taken together, these results suggest that genomic sites with RNA editing are fast evolving, and imply that most

Figure 3 | Functional editing sites in the leaf-cutting ant A. echinatior. (a–c) Venn diagrams of A-to-I editing sites across the three different colonies Ae322, Ae356 and Ae363. (d) Over-represented Gene Ontology (GO) terms for the 835 genes that were consistently edited in at least one caste, showing the percentages of edited genes relative to all transcribed genes annotated to corresponding GO terms. The first figure after each bar indicates the number of edited genes annotated to this GO term, and the second number indicates the number of all transcribed genes (RPKM > 1 in at least one of the nine sequenced samples) annotated to this GO term.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms5943 ARTICLE

© 2014 Macmillan Publishers Limited. All rights reserved.
RNA-editing sites in *A. echinatior* are species-specific and thus likely to have emerged recently.

To further investigate why genomic sites corresponding to RNA-editing sites appear to evolve relatively fast, we checked the evolutionary rate of the 835 genes, which contained at least 1 caste-specific editing site that consistently appeared in 3 colonies by calculating the non-synonymous (dn) and synonymous (ds) mutation rate among the 7 attine ant species (Supplementary Methods). Except for the small proportion of 5′-UTR-edited genes, we generally found that the coding regions of RNA-edited genes evolved at equal or lower rates than the background genes (Supplementary Fig. 7). This suggests that the fast evolutionary rate of RNA-editing sites cannot be explained by hitchhiking effects connected to neighbouring genes. It thus seems more probably that the fast evolution of RNA-editing sites may relate to gene expression responses to changes in the specific social environments of different fungus-growing ant species.

We next examined whether RNA editing of the conserved *A. echinatior* genomic sites also occurred in the transcriptomes of other ant species by comparing the RNA editemes of *A. echinatior* with previously published transcriptome data from *C. floridanus* and *H. saltator* (Supplementary Table 9). We identified 113 and 84 *A. echinatior* sites that were also edited and displayed the same RNA base substitutions in at least one sample of *C. floridanus* and *H. saltator*, respectively. We identified 45 sites that displayed the same RNA base substitutions in all three ant species (Fig. 4b). The number of conserved RNA-editing events that this analysis identified is likely to be an underestimate because the transcriptome sequencing depth in *C. floridanus* and *H. saltator* was lower than in *A. echinatior*, and samples were also edited, as well as their protein recoding potential. The red arrow in the sixth exon indicates the positions of the two neighbouring editing sites. The lower part shows the editing levels of the two sites in the nine head samples (G, L, S stand for gynes, large workers and small workers, respectively, and are followed by colony number). (e,f) Editing levels of the corresponding sites of *qvr* in *C. floridanus* and *H. saltator*.

**Figure 4 | Conservation of RNA editing among ant species.** (a) Percentage of conserved genomic sites that display the same DNA bases among seven attine ant species. The ‘Gene’ category is the sum of 5′-UTRs, 3′-UTRs, CDSs and introns, and the ‘Intergenic’ category represents all genomic regions that are neither genes nor TEs. The red bars (genome background) were calculated from all transcribed ‘A’s (that is, ‘A’s covered by ≥1 RNA reads) in each category, and the blue bars from genomic sites corresponding to A-to-I RNA editing sites in each protein coding region. (b) The number of conserved editing events (sites displaying the same RNA base substitutions among species) between *A. echinatior* and *C. floridanus* and between *A. echinatior* and *H. saltator*. (c) Correlation of RNA-editing levels for the conserved editing events between heads of small workers of *A. echinatior* and brains of minor workers of *C. floridanus*. Only sites with RNA read coverage ≥10 × in both samples were used. Editing levels of *A. echinatior* small workers were calculated from the combined data of the three small worker samples. The blue line is a linear regression that is virtually identical to the diagonal. (d) Editing of the *qvr* gene in *A. echinatior*. The upper part shows the *qvr* gene model and the two editing sites, as well as their protein recoding potential. The red arrow in the sixth exon indicates the positions of the two neighbouring editing sites. The lower part shows the editing levels of the two sites in the nine head samples (G, L, S stand for gynes, large workers and small workers, respectively, and are followed by colony number).
all three castes of *A. echinatior* and in most samples of *C. floridanus* and *H. saltator*, causing S101G and H106R amino acid recoding in the protein sequence (Fig. 4d–f). Multiple sequence alignments indicated that genomic sites corresponding to these two editing sites are highly conserved in many insect species (Supplementary Fig. 8). Surprisingly, the *qvr* A-to-I RNA editing that we discovered occurs on exactly the same two sites in *Drosophila*40, indicating that both editing events have been conserved for > 300 MY of insect evolution. Other examples include the following: Rdfl and lark, two genes involved in circadian rhythm39,40, TBPH (also called TDP-43), a gene associated with neuromuscular junction development53; Caps (calcium-activated protein for secretion), a gene involved in synaptic transmission54, and tipE (temperature-induced paralytic E), a gene encoding voltage-gated sodium channel auxiliary subunits (Supplementary Data 4). These results once more confirm the importance of RNA editing in the regulation of nervous system function in *A. echinatior*.

Finally, the conserved editing events between head samples of *A. echinatior* and brain samples of *C. floridanus* allowed us to estimate the correlation of editing levels in similar tissues between the two ant species. We found that editing levels in *A. echinatior* were positively correlated with editing levels in *C. floridanus* at the same sites (Fig. 4c,e,f), implying that not only the editing events but also the degrees of editing have probably been under similar selection for many of these sites.

RNA editing and behavioural caste differentiation. In general, we observed that genome-wide editing patterns were quite similar across samples of the same caste from different colonies, with hierarchical clustering analysis separating samples by caste (Fig. 5a), which indicates that RNA-editing profiles differ more across castes than across colonies. Moreover, consistent with the ADAR expression pattern across the three castes (Fig. 1c), RNA-editing patterns of large workers were closer to those of gynes than small workers (Fig. 5a).

We used Fisher’s exact tests to identify sites that displayed significantly different levels of RNA editing between any two of the three female castes within each colony (see Methods) and detected an average of 276, 599 and 338 significantly differentially edited sites between gynes/large workers, gynes/small workers and large/small workers, respectively (Supplementary Table 10). To focus on sites where editing changes are likely to be consistently associated with caste status rather than colony age, colony size or habitat, we restricted our further analyses to target differentially edited sites that were shared across the three sampled colonies (see Methods) and identified 10, 124 and 15 sites with consistently different editing levels across the three colonies for gyne/large worker, gyne/small worker and large/small worker comparisons, respectively, which involved a total of 137 sites targeting 48 genes.

Of these 137 consistently caste-differentially edited sites, 43, 13 and 13 were conserved for their corresponding genomic sites among the 7 attine ants, between *A. echinatior* and *C. floridanus*, and between *A. echinatior* and *H. saltator*, respectively. This encompassed a total of 55 sites, representing 40% of the 137 differentially edited sites (Supplementary Data 5). Three of these sites were located in protein coding regions, with two of them targeting the ninth exon of the gene *Aech_02313*, which had significantly higher editing levels across colonies in small workers compared with gynes, and resulted in Q968R and Q968W amino acid recoding in the protein sequences (Fig. 5c). This gene is an orthologue of the *wah* gene (also known as *NSL1, CG4699*) in *D. melanogaster* and has been proposed to participate in neuromuscular junction development55. We found that editing of these two sites was always linked (Supplementary Table 11), and domain annotation revealed that they were close to the PEHE domain (histone acetyltransferase binding activity) of the *wah* gene (Fig. 5c), implying their potential to affect protein activity. Both of these two recoding sites were highly conserved in ants (Supplementary Fig. 9a,b) and Q968R was also edited in some samples of *C. floridanus* and *H. saltator* (Supplementary Fig. 9c).

Another site in which small workers showed significantly higher editing levels than gynes across colonies, was targeting the fifth exon of *Aech_08485*: a gene encoding a Zinc finger protein (Supplementary Fig. 10). This editing site can cause K432R recoding in the protein sequence downstream of the C2H2 zinc finger domain and is also conserved in all other ant species and edited in some samples of *C. floridanus* (Supplementary Fig. 11), but no function for this gene is known.

Most of the consistently differentially edited sites were targeting the non-coding regions of genes, mainly in 3’-UTRs (Fig. 5b), some of which are known to function in neural systems (Supplementary Data 5). For example, *Tap42*, *pUf68* and *Droj2* are known to participate in neurogenesis in *Drosophila*,56 and all of them harboured consistently differentially edited sites in their 3’-UTRs that displayed higher editing levels in worker castes than in gynes (Supplementary Fig. 12 and Supplementary Data 6). Another gene associated with neuromuscular junction development, *TBPH* (also called *TDP-43*)53, also harboured a consistently differentially edited site in its 3’-UTR with higher editing levels in small workers than in gynes (Supplementary Fig. 12 and Supplementary Data 6). The neuromuscular junction connects the nervous system to the muscular system via synapses between nerves and muscle fibres, and is critical in locomotory behaviour. Differential editing of *TBPH* and *wah* may thus imply that RNA editing is involved in modulating caste-specific adult locomotory behaviour. RNA editing in the 3’-UTR of genes cannot change protein sequences, but may affect the stability of mRNA or change the targets of miRNAs53 to regulate the final protein levels. Moreover, it has been proposed that A-to-I RNA editing may cross-talk with RNA-interference pathways, which involve dsRNAs; thus, RNA editing may play a role in gene silencing (reviewed in ref. 57). The differential patterns of RNA editing among *A. echinatior* castes thus appear to uncover a suite of novel potential mechanisms for differentiating caste phenotypes in eusocial insects.

**Discussion**

This is the first study focusing on the characteristics and functions of RNA editing in a eusocial insect. Unlike previous studies, we used a strand-specific RNA-Seq strategy, which can infer editing types directly from the sequencing reads, independent of whether genomic annotations are available. Our study also exploited the fact that *A. echinatior*, similar to most other eusocial Hymenoptera, has low degrees of genetic polymorphism within single colonies (our sampling unit) because all females are offspring of the same (diploid) mother queen and four to eight haploid males58. Our high-coverage sequencing for gDNA and RNA thus allowed us to capture essentially all genetic variation present in these colonies. We acknowledge that some of the low-level RNA editing events could represent false positives due to errors in the amplification or sequencing processes, but also note that scatter for low editing level estimates obtained with two independent methods was low; hence, this is unlikely to affect our main conclusions (Supplementary Fig. 4).

Consistent with observations in other animals, A-to-I editing appeared to be the dominant form in ants, with >96% of the identified editing sites belonging to this category, supporting the view that non-A-to-I RNA editing is very rare in animals.
The highly similar characteristics of A-to-I editing sites among ants, humans, mice and fruit flies, such as their frequent occurrence in clusters and in dsRNA regions, and their analogous depletion/enrichment of guanines in the upstream/downstream positions of edited adenosine, strongly suggest that the function of ADAR enzymes is highly conserved from insects to mammals, regardless of the variation in copy numbers of ADAR genes across lineages. In addition, two novel characteristics of RNA editing were reported in this study: (i) the low clustering tendency for editing in coding regions compared with editing in other regions and (ii) the mutually exclusive editing of different regions within gene bodies.

Our finding of RNA-edited genes that are functionally enriched for neurotransmission and circadian rhythm is consistent with the known role of RNA editing in the nervous system, but also offers extensions to the spectrum of possible functions of RNA editing. Our results suggest a new and potentially wide-ranging post-transcriptional gene regulation mechanism that could fine-tune gene expression in the eusocial insect brain. Eusocial insects, such as ants, combine individual and collective behaviours in a uniquely self-organized manner\(^59\), which might be particularly amenable to regulatory mechanisms such as RNA editing. However, much further work will be needed to detail the relative importance of alternative mechanisms that determine the expression of genes affecting ant behaviour.

Comparative analyses among both closely and distantly related ant species revealed mostly low levels of conservation in the genomic sites whose transcripts are subject to RNA editing, suggesting that most editing sites may have emerged in the recent history of *A. echinatior*. However, other genomic regions, representing 8–23% of the RNA-editing sites, appear not to have changed for hundreds of millions of years, indicating they have been under purifying selection during the very long evolutionary history of the ants since their single ancestor became eusocial\(^60\). The regulatory functions of RNA editing at the latter sites may hold important information about the fundamental origins of eusocial organization to be uncovered in future studies. In addition, the cross-species conservation of the relatively small number of RNA-editing sites that showed consistently different editing levels across adult caste phenotypes may offer valuable insights into the mechanisms that made eusociality in ants progress almost immediately towards morphological caste differentiation with distinct behavioural syndromes. In eusocial lineages of bees, wasps and termites, initial caste differentiation appears to have been based on conditionally expressed, reversible behaviours that did not become associated with permanent morphological castes until much later in their evolutionary history\(^61–84\).

### Methods

**Analysis of RNA editing enzymes.** We obtained ADAR protein sequences of *Apis mellifera* (NP_001091684.1), *D. melanogaster* (NP_569940.2), *C. elegans* (NP_492153.2, NP_498594.1), *N. vectensis* (XP_001629062.1, XP_001629615.1), *C. gigas* (KE208855.1, KEC23699.1) and *C. intestinalis* (XP_002122509.2, XP_002128212.1) from NCBI, *Homo sapiens* (ENSP00000292205, ENSP00000435381, ENSP000002937013) and Xenopus tropicalis (ENSTETP0000008359, ENSTETP00000057325, ENSTETP0000039188) from Ensembl (release-68), and *S. purpuratus* (SPU_008095, SPU_012470) from EnsemblMetaZoa (release-18), and used these sequences as queries to search for...
ADAR genes in seven published ant genomes (http://hymenopteragenome.org/) and five unpublished ones (S. Nygaard et al., in preparation).

We first used DNA polymerase to amplify 3’-end (4,172 ± 223) and excised target-gene regions (2-kb flanking sequences) for gene structure and protein sequence determination using GeneWise (v2-2-0). Next, we aligned the predicted proteins to the nr database of NCBI to check whether the predicted proteins were ADARs. The gene model in Solenopsis invicta was only partially assembled, so for further analysis, protein organization was determined using GeneWise (http://www.megasoftware.net).

To search for APOBEC genes in the insect genomes, we collected the APOBEC protein sequences of H. sapiens (APOBEC1, APOBEC2, APOBEC3A-3H and APOBEC6), X. tropicalis (apo2bce2 and apoc4) and Danio rerio (apo2bca and apo2bc2b) from Ensembl (release 68) and performed the same analysis as described above. Here, size selection could have been performed in any of the 12 ant genomes and other insects with available genomes.

### Biological material and DNA/RNA extraction

Three colonies of A. echinatior (Ae322, Ae356 and Ae363) were collected in Gamboa, Panama, in 2006–2008, and maintained in the laboratory under standard conditions of ca. 25 °C and ca. 70% relative humidity where they were supplied with a diet of bramble leaves, rice and pieces of apple. RNA and DNA samples were collected from the female castes: gynes, large workers and small workers. Animals were flash frozen in liquid nitrogen and the heads separated from the bodies with a pair of forceps. Heads from 50 gynes, 50 large workers or 200 small workers originating from the same colony were pooled and used for RNA extraction as described previously.68 The remaining body parts were ground into powder and used for DNA extraction (Supplementary Table 1). DNA was extracted using a Blood & Cell Culture Midi kit (Qiagen) following the manufacturer’s instructions, with the modification that sample lysis was performed overnight.

### qPCR analysis of ADAR and GADPH expression

Total RNA of the nine samples was first converted to complementary DNA using SuperScript II reverse transcriptase (Invitrogen) and an oligo d(T) primer according to the manufacturer’s instructions. Forward and reverse primers for ADAR were 5’-GCT GAC CAA ATT GGC GAG AAT-3’ and 5’-CCG CTT CGA CTT AAA TGC TG-3’, respectively, and forward and reverse primers for GADPH were 5’-AGC CAC TCA GAA GAC CGT TG-3’ and 5’-CCA GTC TTC CGC TTG AG-3’, respectively. Primer pairs were spanning an intron to prevent amplification of gDNA. The final PCR reactions contained 0.8 µl of each primer (10 µM), 10 µl SYBR Green 1 (TaKaRa) and 2 µl template. Quantitative qPCR (qPCR) was performed on an ABI StepOnePlus system using Ex Taq HS polymerase (TaKaRa) in a final volume of 20 µl and the β-actin gene (primers: 5’-CAG CAG TTA TGG CAA AAT CG-3’ and 5’-ACC CAC ACC ATG TTC TCA-3’) as internal reference. All reactions were subjected to 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min, and melting curve analysis to ensure homogeneity of the reaction product. qPCR results were analysed using the StepOne Software.

### Library construction and sequencing

For each DNA sample, 5 µg gDNA was fragmented by sonication with a Covaris S2 system (Covaris, MA) to a mean size of 200–500 bp, followed by end repair, 3’-end overhangs ligation, double-stranded cDNA synthesis using MMLV-RT (Invitrogen) and second-strand cDNA synthesis using a DNA polymerase (Enzymatics). All libraries were size selected at 500 bp on an agarose gel and amplified by ten cycles of PCR to yield the DNA libraries. All libraries were subjected to 90 bp pair-end sequencing on the Illumina HiSeq 2000 platform.

### Strand-specific RNA-Seq

Strand-specific RNA-Seq libraries were constructed according to Parkhomchuk et al. Briefly, 2 µg total RNA per sample was treated with DNase I (New England Biolabs) to remove possible contamination with gDNA. Next, poly (A) mRNA was fragmented by sonication with a Covaris S2 system (Covaris, MA) to a mean size of 200–500 bp, followed by end repair, 3’-end overhangs ligation, double-stranded cDNA synthesis using a DNA polymerase (Enzymatics). All libraries were size selected at 500 bp on an agarose gel and amplified by ten cycles of PCR to yield the DNA libraries. All libraries were subjected to 90 bp pair-end sequencing on the Illumina HiSeq 2000 platform.

### Read alignment

RNA and DNA reads were aligned to the genome of A. echinatior (http://hymenopteragenome.org/) using Bowtie (v0.5.9-166), allowing five mismatches and zero gaps for the 90-bp reads. Only uniquely mapped reads (XT-A:U) with mapping quality ≥ 20 and ≥ 10× of the mean fragment length found by Bowtie were used for further analysis. Reads resulting from PCR duplicates (that is, read pairs that mapped to identical genomic locations) were removed except for the one with the highest sequencing quality. We then cut the first and last 6-bp of the aligned reads, as it is known that (i) the error rate of Illumina sequencing is higher towards both ends of a read, especially at the 3’-end; (ii) some 5’-end mismatches may be introduced by the random hexamers during the first- and second-strand syntheses of RNA library construction; and (iii) mapping errors around insertions/deletions relative to a reference genome can lead to mismatches occurring towards the beginning and ends of a read. In addition, to improve the accuracy of read alignment, we used BLAT (v. 34) to re-align all the uniquely mapped reads to the reference genome and only kept reads supported by both BLAT and BWA. As the fundamental algorithms of BLAT and short-read aligners such as BWA are different, the read-mapping results were often different and complementary, especially when handling reads derived from spliced junctions.69 Finally, ca. 63% of the raw RNA data and ca. 29% of the raw RNA data were retained for further analysis (Supplementary Table 2).

### Identification of RNA-editing sites

The basic principle for identifying an RNA-editing site is that the site must be homozgyous for gDNA, while at the same time displaying a mismatch between RNA and DNA. We conducted two rounds of editing site identification in each sample. In the first round, we required that (i) a candidate RNA-editing site in a given sample must be supported by ≥3 RNA reads that were mapped to overlapping but not identical positions in the reference genome and this site had an editing level ≥5%. The editing level of a candidate editing site was calculated as the number of reads supporting editing divided by the total number of reads covering that site. RNA reads with quality score <30 for the first editing site (that is, the upper limit of sequencing error for a base was 0.1%). (ii) The candidate sites with multiple editing types were discarded. (iii) To avoid potential false positives resulting from mis-mapping of reads at splice junctions, we further required a candidate editing site to be supported by at least one RNA read in the middle of its length (that is, from positions 23–68 of TG3’ read). We discarded intronic candidate sites that were located within 10 bases of an splice site. (iv) To remove false positives resulting from heterozygous genomic SNPs, we only kept candidate editing sites in homozgyous genomic sites. Taking into account that female offspring are most likely to be half-sisters because queens of A. echinatior mate with multiple males67 and that offspring from separate pools of multiple individuals, some heterozygosity may have occurred in low frequencies in our data set, making commonly used SNP calling algorithms unsuitable as tools. Thus, to filter as many heterozygous SNPs as possible, we defined homozygous genomic sites with a strict criterion: sites with ≥10× RNA read coverage in each of the 9 samples and all the covering reads located on the same genotype. For each round of editing-site identification, we conducted a second round of editing-site identifications. We first combined all editing sites identified in the first round to obtain a comprehensive map of potentially editable positions in the genome of A. echinatior, which produced a total of 17,229 positions that were confirmed to be homozgyous for gDNA, to be located not be close to any intronic candidate sites, i.e. to be edited in at least one of the 9 samples. We then retrieved missing editing sites in each sample using the more liberal criterion of sites with ≥1 RNA read, supporting editing in already identified editable positions also being permissible.

To further distinguish between true and false positives due to sequencing error, we performed statistical tests for the candidate editing sites identified in each sample based on the binomial distribution B(n, p), with p = 0.1% (the upper limit of sequencing error for a base required in this study) and n equal to the total depth of a candidate editing site. For example, given a candidate editing site with k reads supporting editing and a total depth of n, we calculated the probability that all the k reads support the edited state (i.e. that k out of n true positives are edited at a given position, and then compared the probability of B(k, n, p) to 0.01 after adjusting the P-value by Benjamini–Hochberg False discovery rates (FDRs)67. Only candidate editing sites with adjusted P-values <0.01 were finally considered to be true positives and kept for subsequent analyses.

### PCR validation of identified RNA-editing sites

To confirm the reliability of our pipeline, we randomly picked 116 editing sites from large workers of colony L363 for PCR validation, including 106 A-to-I editing sites and 10 non-A-to-I editing sites identified in the first round to obtain a comprehensive map of potentially editable positions in the genome of A. echinatior, which produced a total of 17,229 positions that were confirmed to be homozgyous for gDNA, to be located not be close to any intronic candidate sites, i.e. to be edited in at least one of the 9 samples. We then retrieved missing editing sites in each sample using the more liberal criterion of sites with ≥1 RNA read, supporting editing in already identified editable positions also being permissible.

To further distinguish between true and false positives due to sequencing error, we performed statistical tests for the candidate editing sites identified in each sample based on the binomial distribution B(n, p), with p = 0.1% (the upper limit of sequencing error for a base required in this study) and n equal to the total depth of a candidate editing site. For example, given a candidate editing site with k reads supporting editing and a total depth of n, we calculated the probability that all the k reads support the edited state (i.e. that k out of n true positives are edited at a given position, and then compared the probability of B(k, n, p) to 0.01 after adjusting the P-value by Benjamini–Hochberg False discovery rates (FDRs)67. Only candidate editing sites with adjusted P-values <0.01 were finally considered to be true positives and kept for subsequent analyses.
subjected to Sanger sequencing on an ABI 3730 DNA sequencer, and PCR products from cDNA samples were used to construct TA clones of which we randomly selected 50 for Sanger sequencing on the same ABI 3730 sequencer to validate the existence of RNA editing. An editing site was considered to be true when its gDNA was homogenous based on the .ab1 chromatogram file and at least 1 out of the 50 sequenced clones supported editing. The editing level for each editing site was calculated as the number of clones with the edited nucleotide divided by the total number of sequenced clones.

Functional annotation and enrichment analysis. GO annotations for A. echinatior were based on orthologues in C. elegans, D. melanogaster, A. mellifera, Nasonia vitripennis and A. cephalotes from EnsemblMetazoa (release-18). Fisher’s exact test and \( z^{2} \)-test were employed to estimate whether a list of edited genes was enriched in a specific GO category when compared with background genes. All expressed genes, which we defined as genes with RPKM \( \geq 1 \) in at least one of the nine sequenced samples, were used as background genes in this GO enrichment analysis. \( P \)-values were adjusted for multiple testing by applying FDR67.

Conservation analysis of RNA-editing sites. To investigate the evolutionary conservation of genomic sites corresponding to RNA-editing sites, we performed cross-species whole-genome alignments between A. echinatior and six other attine ants (A. cephalotes, Atta colombica, Trachymyrmex septentrionalis, Trachymyrmex corneti, Trachymyrmex zeteki and Cyphomyrmmex costatus) plus two non-attine ants (C. floridanus and H. saltator). Whole-genome alignments were performed using LASTZ with default parameters and using the A. echinatior genome as the reference. Two methods were used to estimate the conservation degrees of the genomic sites corresponding to RNA-editing sites: (i) the percentage of genomic sites corresponding to RNA-editing sites showing the same DNA bases among seven attine ant species, between A. echinatior and C. floridanus and between A. echinatior and H. saltator, respectively; and (ii) mean phastCons scores for genomic sites corresponding to RNA-editing sites, which were calculated from the multiple genome alignments of the seven attine ant species. For both methods, Monte Carlo simulations were performed to estimate whether the observed conservation degrees were significantly different from that of background expectation. Briefly, the same number of transcribed sites (\( 1 \times 10^{6} \)) as that of A-to-I editing sites was randomly sampled in each genomic element, then the conservation degrees for each element were calculated. The \( p \)-value for each element was empirically calculated after the simulation of 10,000 iterations, and the expected conservation degree was the mean value of the 10,000 iterations (Supplementary Tables 7 and 8).

To identify conserved editing sites, we performed cross-species transcriptome comparisons between A. echinatior and C. floridanus (or H. saltator), RNA-Seq data of different developmental stages and castes of C. floridanus and H. saltator were collected from three studies6,7,31 (Supplementary Table 9) and whole-genome BS-Seq data from Bonasio et al. (GSE15176 in NCBI GEO database) were used for removing potential heterozygous SNPs. A conserved editing site between A. echinatior and C. floridanus (or H. saltator) was defined as a homogenous genomic site with consistent genotype and consistent editing type between A. echinatior and C. floridanus (or H. saltator). Considering the generally lower sequencing depth of the RNA-Seq samples of C. floridanus and H. saltator, we required a conserved editing site to be supported by \( \geq 2 \) RNA reads in at least one of the C. floridanus (or H. saltator) samples.

Hierarchical clustering analysis for RNA-editing sites. All editing sites that were confirmed in at least one of the nine sequenced samples by our pipeline and that had RNA read depth \( \geq 20 \times \) in all the nine samples were used for hierarchical clustering analysis (Fig. S5a). The method for distance measurement was ‘euclidean’ and the agglomeration method was ‘complete’. The heatmap as well as the hierarchical clustering tree were generated by using the heatmap function in R (www.r-project.org) based on the editing levels of the editing sites.

Identifying sites with caste-specific RNA editing. For the analysis of differential editing, we first identified differentially edited sites between any two castes within each individual colony (giving nine comparisons in total). To ensure that difference in editing levels between two castes were not due to random variation or insufficient coverage depth in one of the two samples, we performed Fisher’s exact test for each interrogated site. The four variables subjected to Fisher's exact test for each site were (i) the number of RNA reads supporting editing in caste one, (ii) the number of RNA reads supporting non-editing in caste one, (iii) the number of RNA reads supporting editing in caste two and (iv) the number of RNA reads supporting non-editing in caste two. Only sites with P-values \( < 0.01 \) were kept, and we also required that the differences in editing level were higher than 10% between any two samples compared (Supplementary Table 10).

We realized that the biological factors causing RNA-editing differences between castes may be complex; hence, we focused on identifying the sites that displayed consistent differences in editing levels between two castes across all three attine ant colonies. To identify such consistently differentially edited sites between two castes, we required that these sites satisfied the following two criteria: (i) the caste difference in editing level based on the criteria described in the above paragraph was significant in at least two out of the three colonies, while the third colony showed a difference in the same direction. (ii) The \( P \)-values of Fisher’s exact tests using the combined data of the three colonies (that is, pooling the number of reads supporting editing or non-editing across the three colonies) after adjusted by FDR was \( < 0.05 \).

PCR validation of differentially edited sites. Three consistently differentially edited sites, targeting CDSS of Arch, 06485 (a Zinc finger protein) and the 3’-UTRs of p3/l68 and TBPH, were randomly chosen for PCR validation in the nine samples. The nine gDNA and RNA samples for validations were independently collected from the same live lab colonies (Az2c2, Az356 and Az363) in early 2014 that provided the initial RNA-Seq samples in 2012. Primers are listed in Supplementary Data 6 and validation experiments were performed with the same protocol as described above.

References

Acknowledgements
This research was supported by the National Natural Science Foundation of China (31271392), an EU Marie Curie International Incoming Fellowship (300837) to G.Z., the China National GeneBank-Shenzhen, the Danish National Research Foundation (DNRF57) and an ERC Advanced grant to J.I.B. (320885). We thank Jack Howe for help with editing the final manuscript.

Author contributions
G.Z. and Q.L. conceived and designed the empirical work. J.J.B., G.Z., S.N. and M.S. provided the conceptual context and organized the genome materials/analysis tools. Q.L., G.Z. and J.J.B. wrote the paper, with contributions from M.S. and G.Z. coordinated and supervised the empirical work. G.Z. and Q.L. conceived and designed the empirical work. M.S. coordinated and supervised the empirical work. Q.L., G.Z. and J.J.B. wrote the paper, with contributions from M.S. and G.Z.

Additional information
Accession codes: DNA sequencing and strand-specific RNA sequencing data of A. echinatior produced in this study as well as the RNA-editing sites identified in the nine A. echinatior samples have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE51576. Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Li, Q. et al. Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior. Nat. Commun. 5:4943 doi: 10.1038/ncomms5943 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/