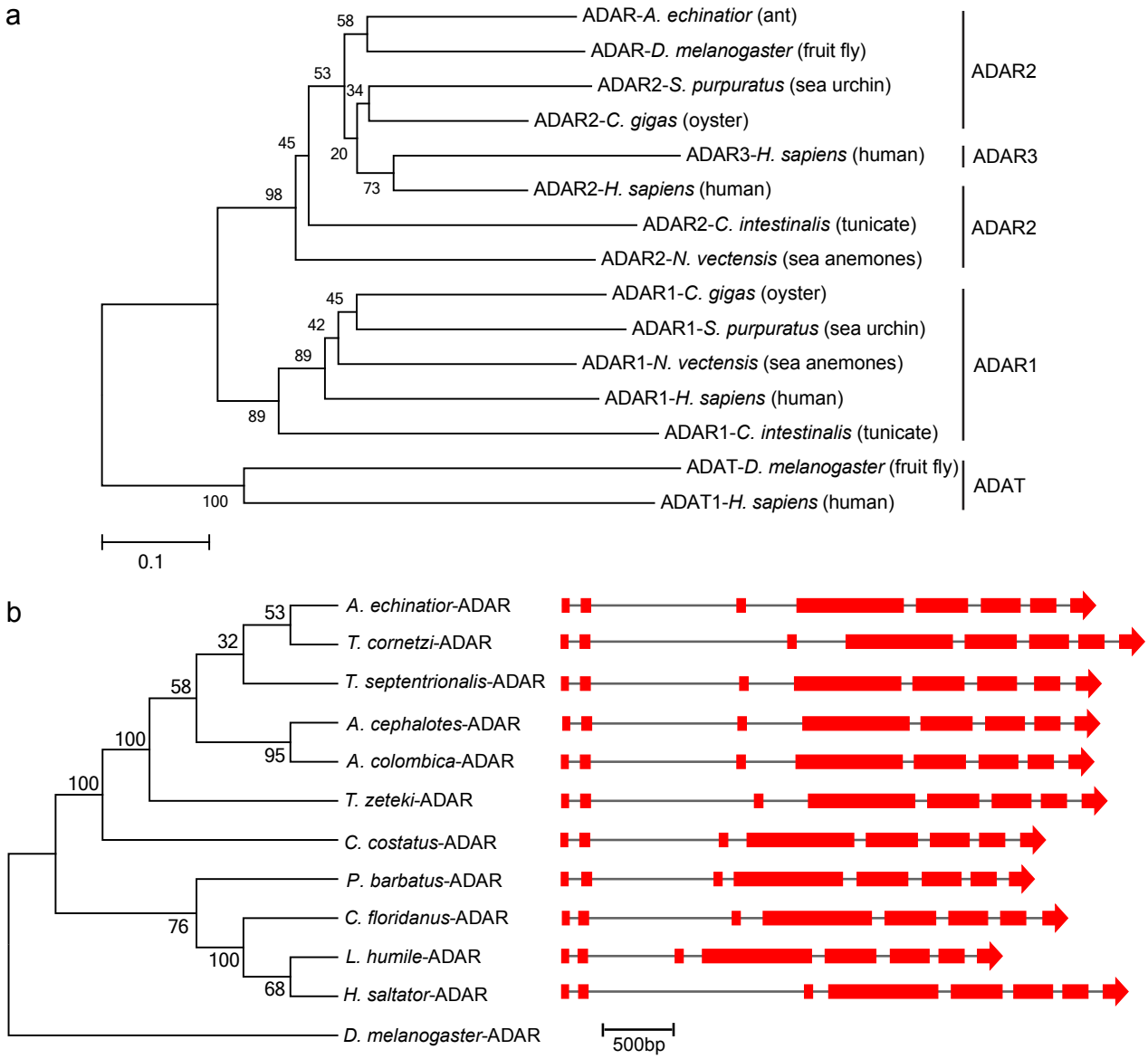
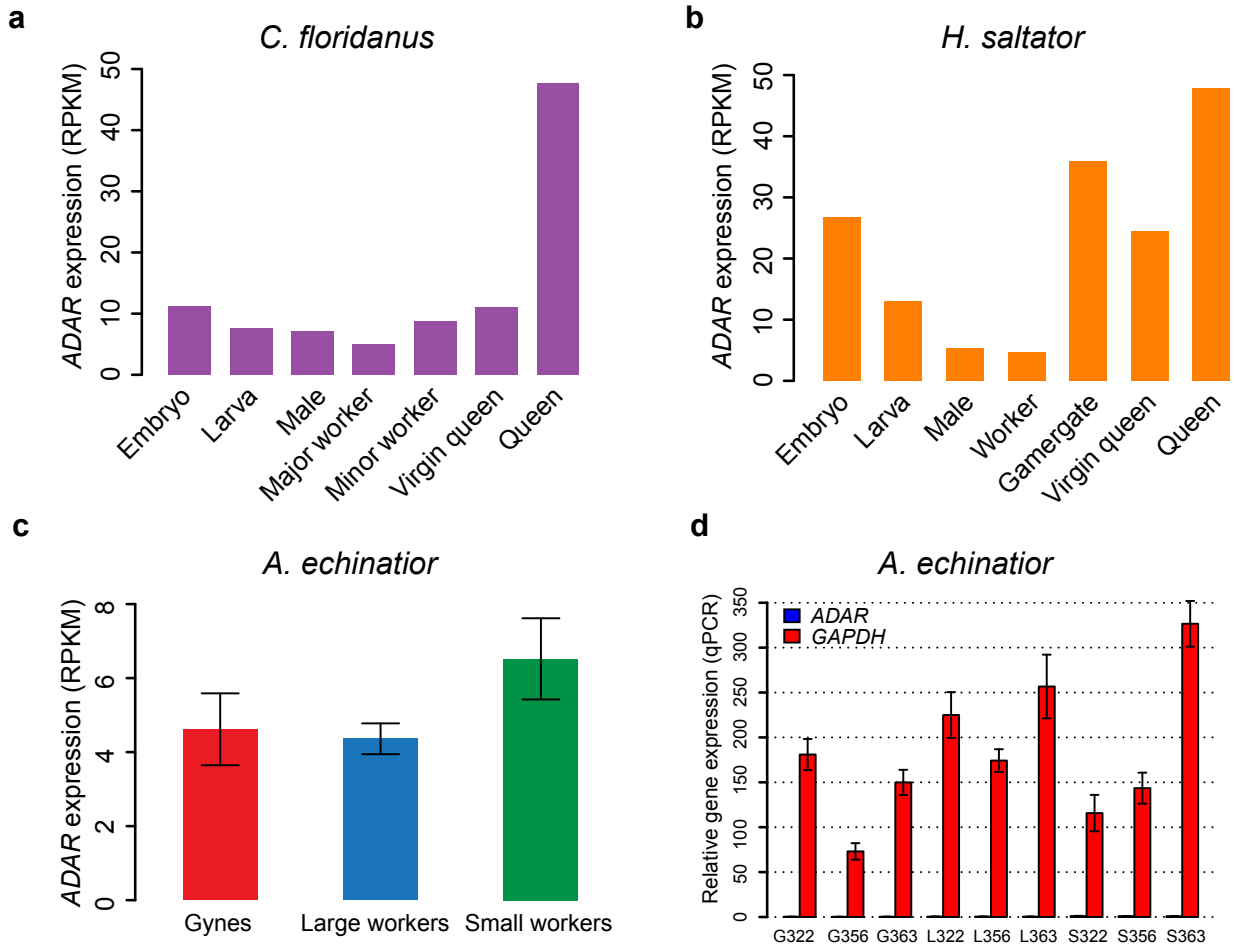
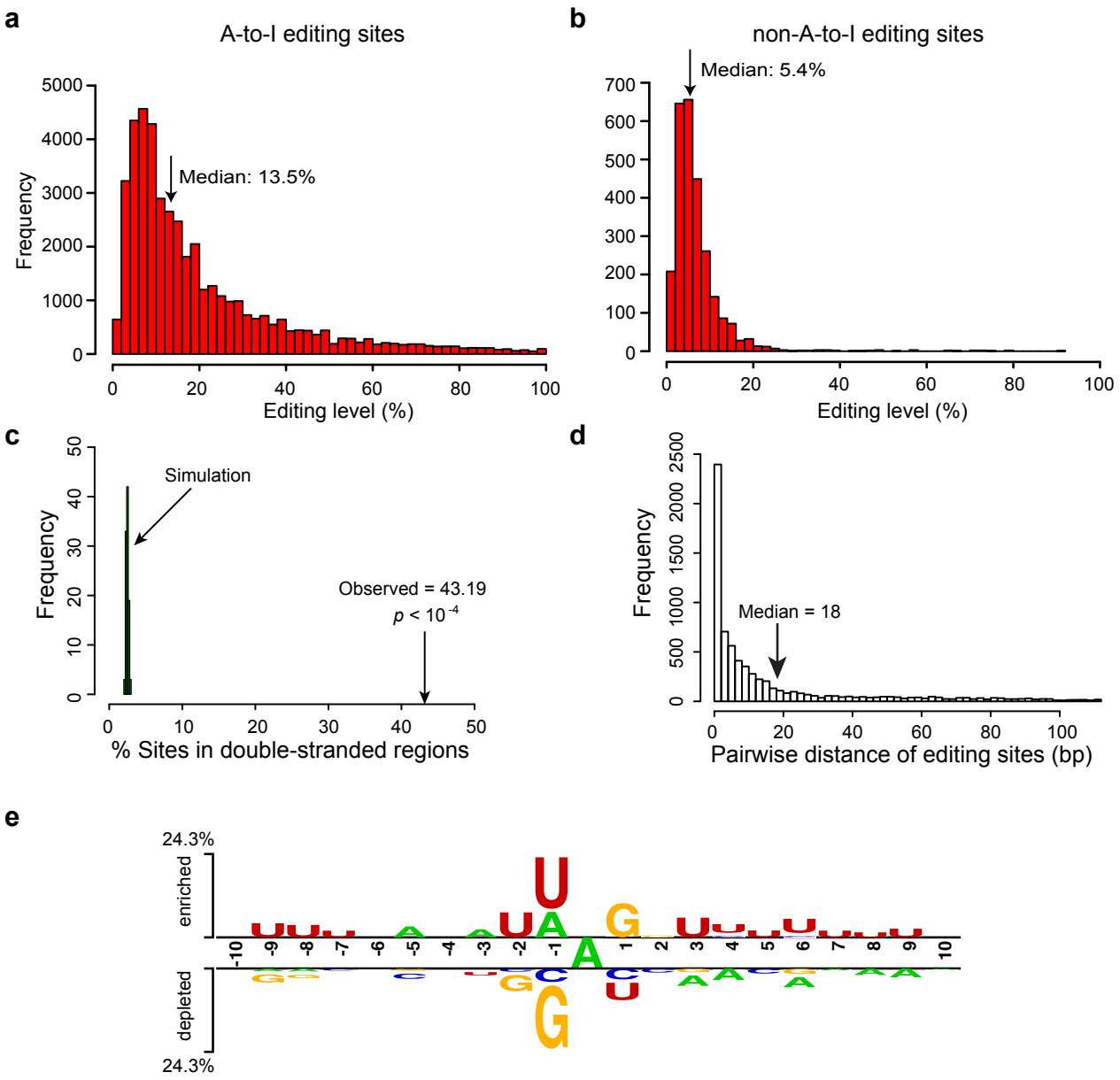
Supplementary Figures



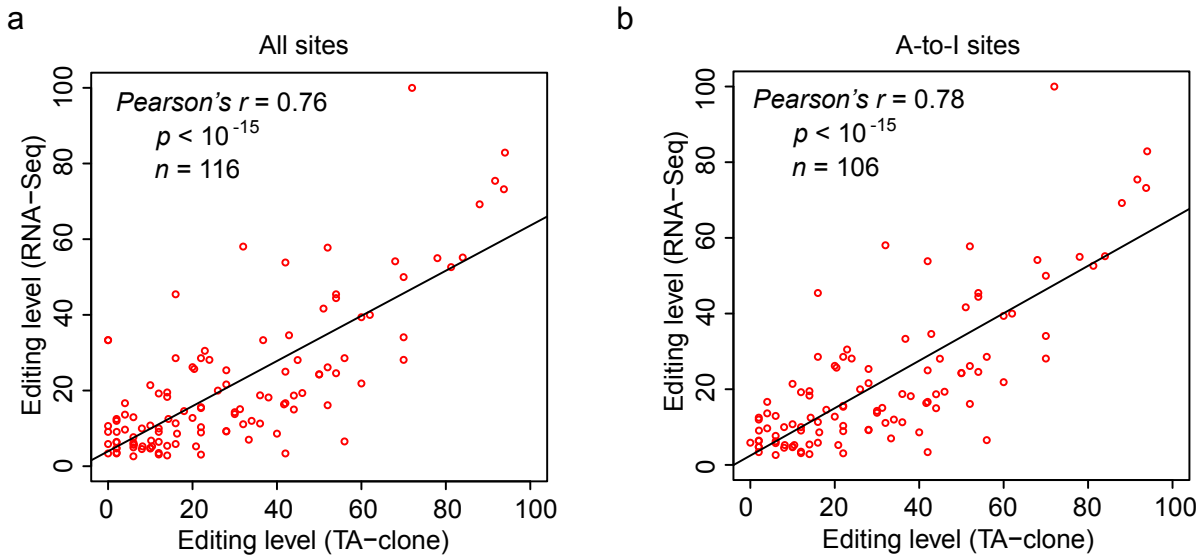
Supplementary Figure 1. **The *ADAR* gene in ants.** (**a**) Phylogenetic tree of ADARs in different animals with adenosine deaminases that acts on tRNA (ADATs) of *Drosophila* and humans as outgroups. (**b**) Phylogenetic tree and gene structure of ADARs in 11 ant species with *Drosophila* as outgroup. The number and size of all coding exons (red blocks) and introns (gray lines) of ADAR are conserved, except for limited length variation in the 2nd and 3rd introns. Protein sequences were aligned with MUSCLE in MEGA5[1] after which trees were generated using the neighbor-joining (NJ) method. Numbers along branches are bootstrap values.



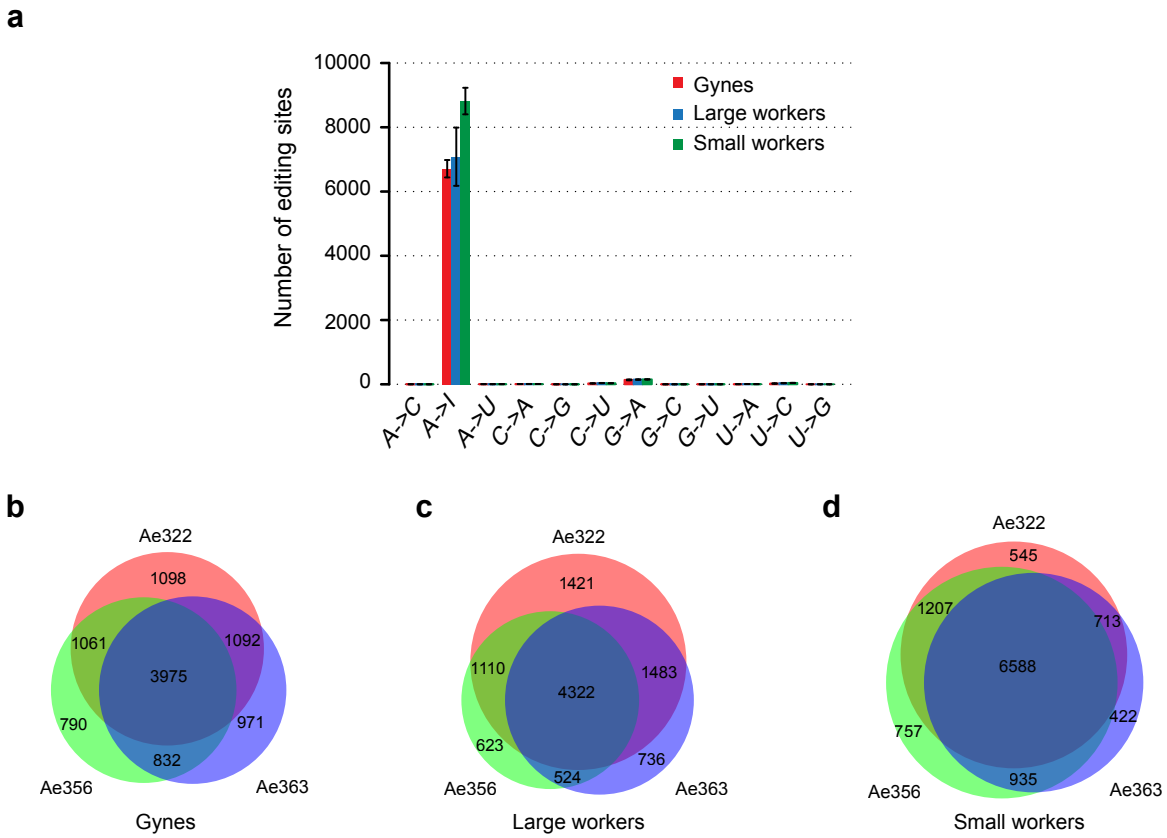
Supplementary Figure 2. Expression of *ADAR* genes. Expression levels of the *ADAR* genes (RPKM) in different castes and developmental stages of the ants *Camponotus floridanus* (**a**) and *Harpegnathos saltator* (**b**). RNA-Seq data for *C. floridanus* and *H. saltator* were obtained from Bonasio *et al*[2,3]. Each sample of these two ant species was the pool of multiple individuals and RNA was isolated from whole organisms. (**c**) Expression levels of *ADAR* genes (RPKM) in the heads of different castes of *A. echinatior*. Error bars are standard deviations (SD) across the three sampled colonies. (**d**) qPCR analysis of *ADAR* and *GAPDH* expression using *Ef1-beta* as internal reference gene, showing that *ADAR* expression levels in the head samples of *A. echinatior* are quite low (as indicated by RPKM in panel c). Error bars are standard deviations (SD) across three replicates. These low *ADAR* expression levels were somewhat surprising, but may be due to *ADAR* expression being restricted to the brains, i.e. in a small fraction of the total head samples that we used. RPKM: reads per kilobase of exon model per million mapped reads. qPCR: real-time quantitative PCR.



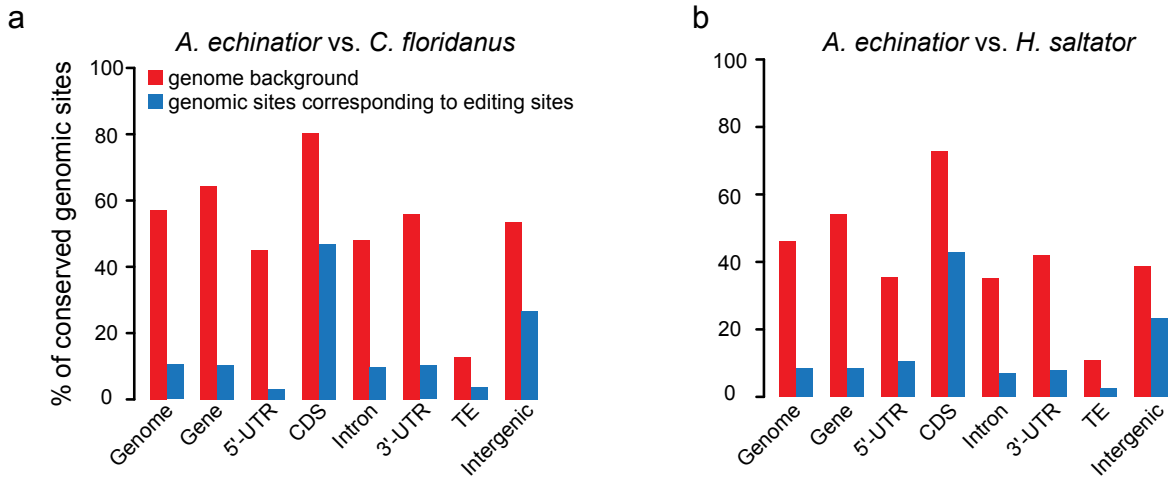
Supplementary Figure 3. Characteristics of A-to-I editing sites in *A. echinatior*. Frequency distribution of editing levels for A-to-I editing sites (**a**) and non-A-to-I editing sites (**b**) with RNA coverage ≥ 20X after combining editing sites from all 9 samples (arrows are the respective median editing levels of 13.5% and 5.4%). (**c**) Proportion of A-to-I editing sites in regions with the potential of forming double-stranded RNA structures relative to random 'A's in regions with RNA-Seq read coverage ≥ 1X; the *p*-value was estimated by the Monte Carlo simulation with 10,000 iterations. (**d**) Frequency distribution of pairwise distances among A-to-I editing sites estimated in the gyne sample from colony 322 (estimations from the other samples gave similar results and figures are therefore not shown). (**e**) Sequence preferences for base positions flanking predicted A-to-I editing sites as displayed by the Two Sample Logo program[4] (http://www.twosamplelogo.org).



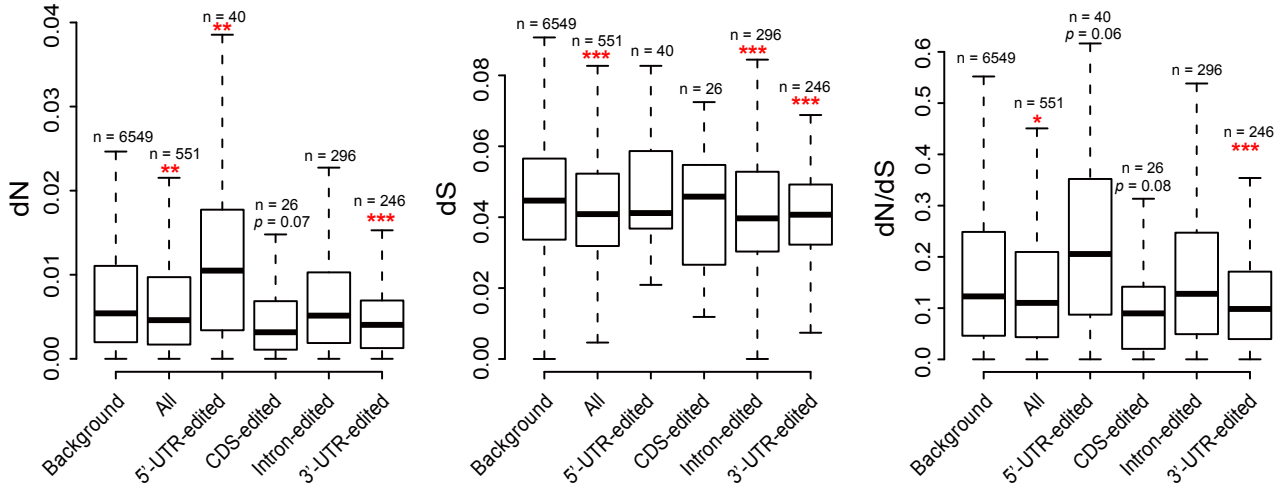
Supplementary Figure 4. Correlations between editing levels estimated by TA-clonal sequencing and RNA-Seq for all editing sites (a) and A-to-I editing sites (b). For each site ca. 50 clones were randomly picked for Sanger sequencing. Fitted lines are regressions predicting RNA-Seq estimates from TA-cloning estimates.



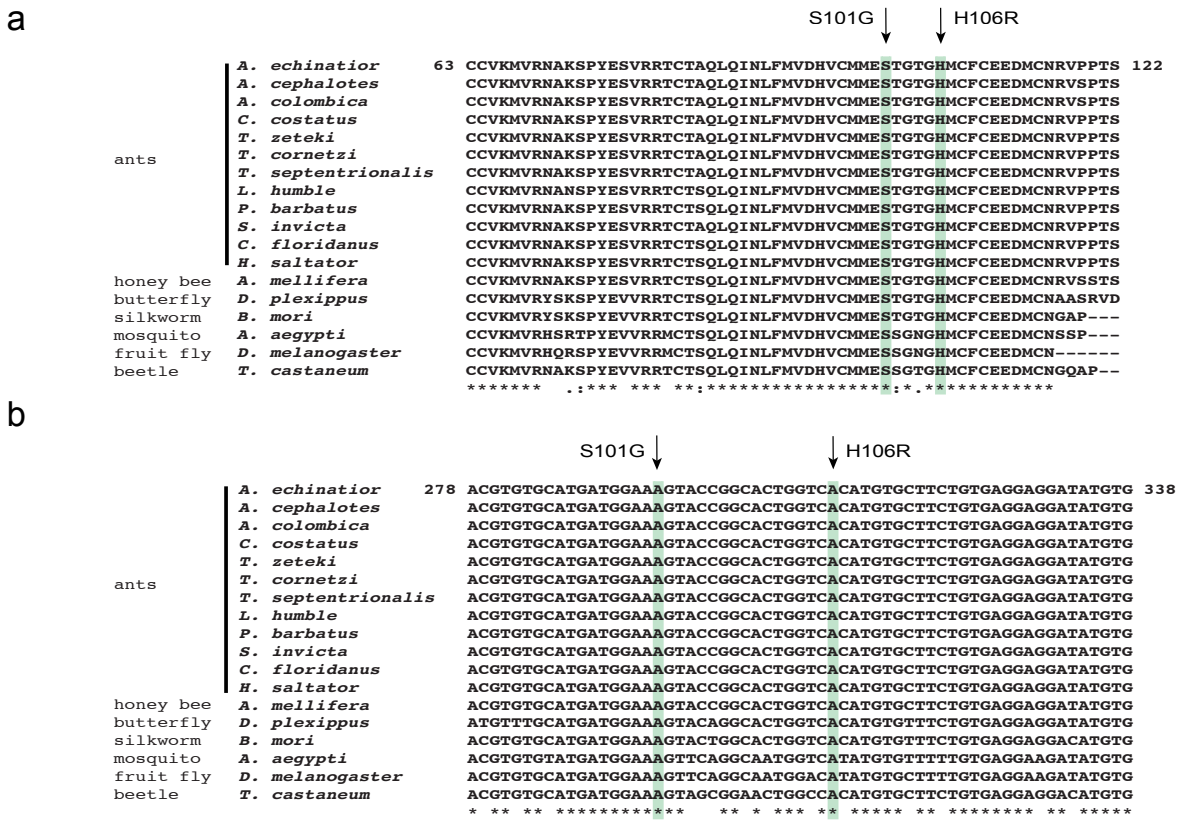
Supplementary Figure 5. Number of editing sites in each sample identified from equal amount of sequencing data. (**a**) Distribution of the mean number of editing sites across the 12 possible types of RNA editing in each caste. (**b-d**) Venn diagrams of caste-specific editing sites across the three different colonies. For each of the nine samples, we randomly picked 8X of usable data for editing site identification, and we could still observe a consistently higher number of editing sites in small workers than in the other two castes.



Supplementary Figure 6. Percentage of conserved genomic sites that display the same DNA bases between *A. echinatior* and *C. floridanus* (a) and between *A. echinatior* and *H. saltator* (b). The "Gene" category is the sum of 5'-UTR, 3'-UTR, CDS and intron, 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 (i.e. 'A's covered by ≥ 1 RNA reads) in each category, and the blue bars were calculated from genomic sites corresponding to A-to-I RNA editing sites in each category.



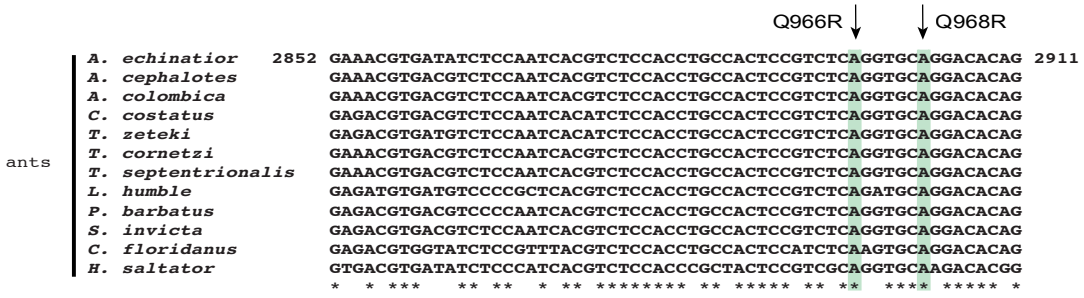
Supplementary Figure 7. Synonymous (*dn*) and non-synonymous (*ds*) mutation rates and *dn/ds* ratios of 835 consistently edited genes. Sequences of one-to-one orthologs of seven attine ants (*Acromyrmex echinatior*, *Atta cephalotes*, *Atta colombica*, *Trachymyrmex septentrionalis*, *Trachymyrmex cornetzi*, *Trachymyrmex zeteki* and *Cyphomyrmex costatus*) and two outgroup species (*Solenopsis invicta*, *Pogonomyrmex barbatus*) were aligned by PRANK[5,6], after which *dn* and *ds* and the *dn/ds* ratios were estimated for each gene using the codeml program with the free-ratio model in the PAML package[7]. All expressed genes (RPKM > 1 in at least one of the nine sequenced samples) were used as background. "*" represents a *p* value < 0.05 when comparing a given group of edited genes with background genes using a Wilcoxon test, "**" represents a *p* value < 0.01 and "***" represents a *p* value < 0.001. *P* values between 0.05 and 0.1 are given as precise values.



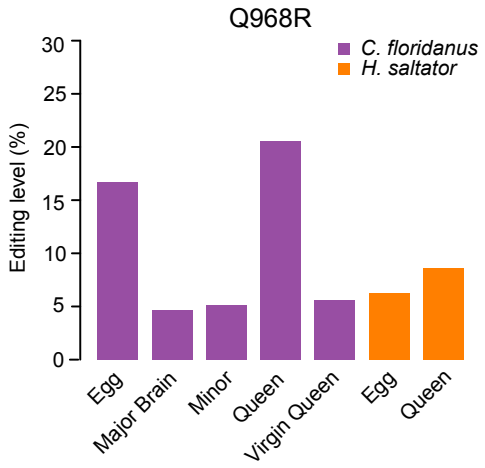
Supplementary Figure 8. Multiple sequence alignments for the protein (a) and mRNA (b) sequences of *qvr* in different insect species. The arrows indicate the positions of the two A-to-I RNA editing sites.

a

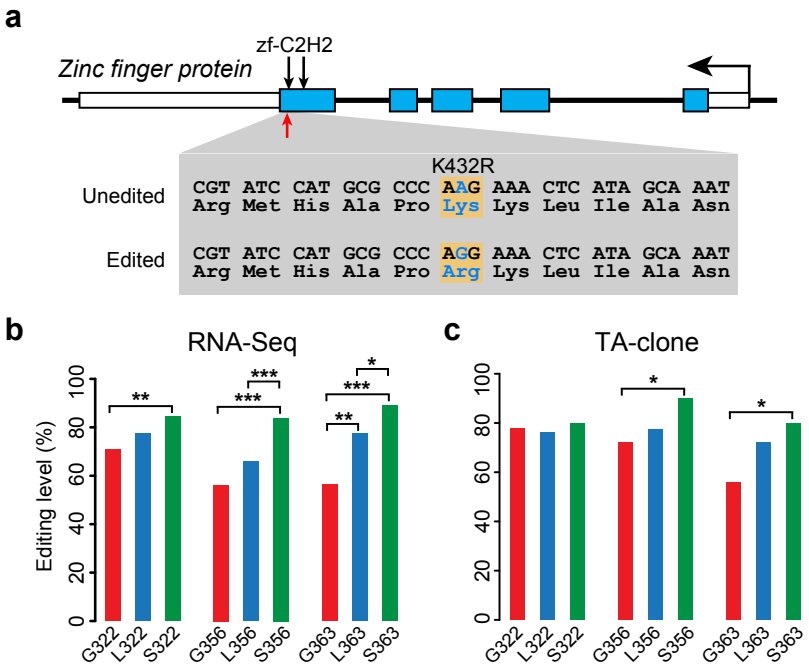
b



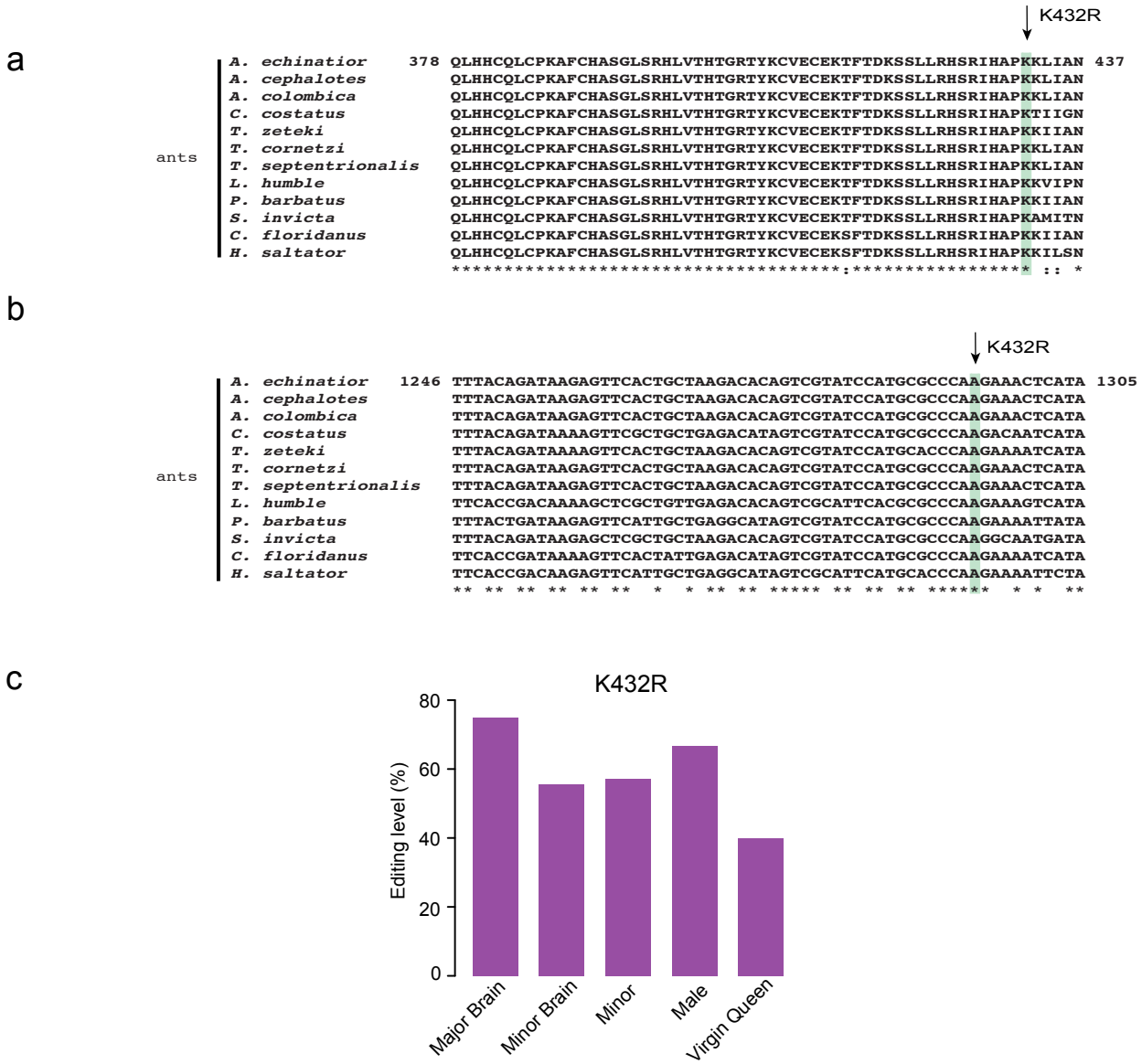
c



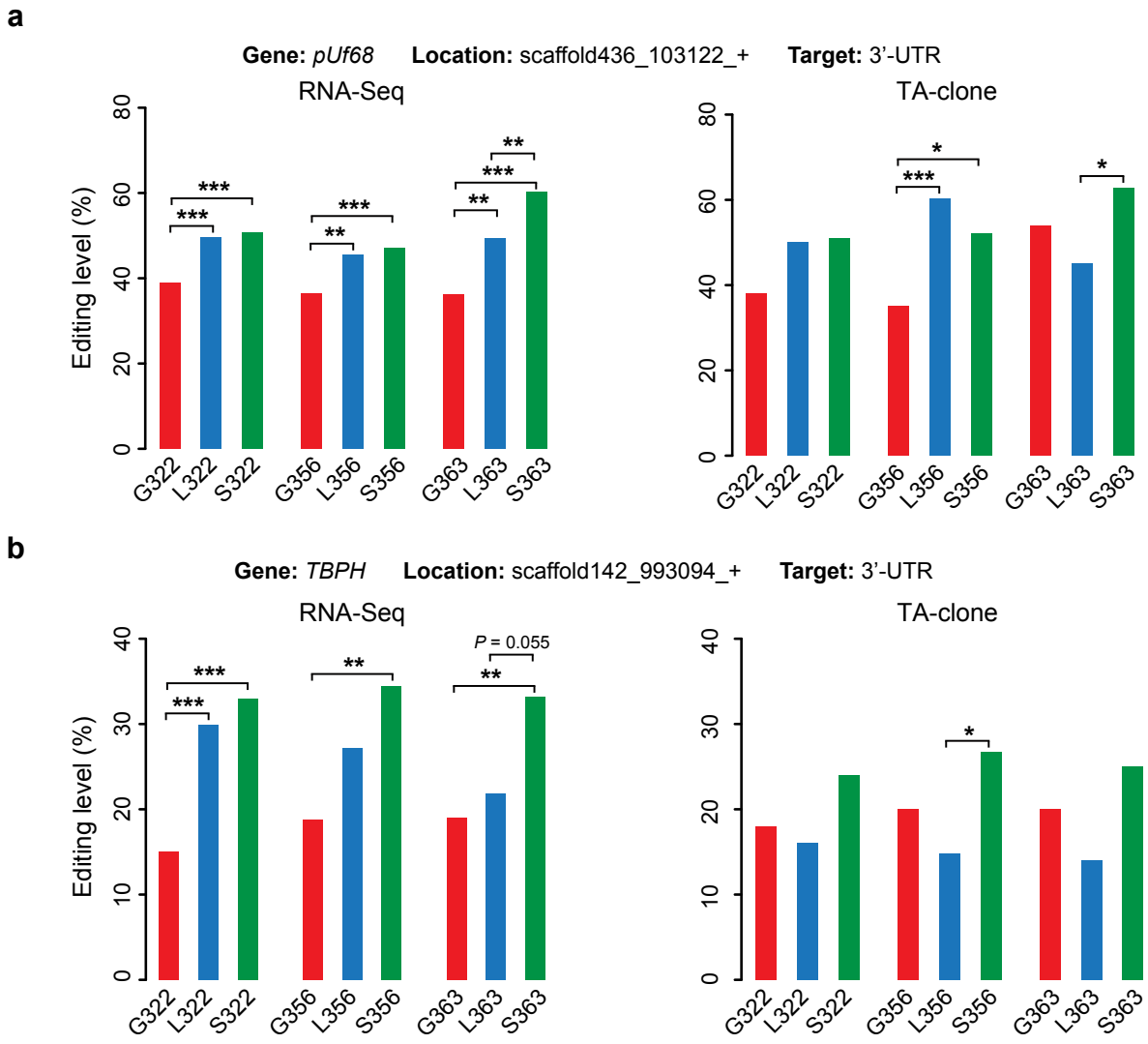
Supplementary Figure 9. Conservation of *wah* (*Aech_02313*) in ants. Multiple sequence alignments for the protein (**a**) and mRNA sequences (**b**) of the *wah* gene adjacent to two recoding RNA editing sites (arrows) across different ant species. (**c**) Editing levels of the Q968 recoding sites of *wah* in *C. floridanus* and *H. saltator*.



Supplementary Figure 10. RNA editing of *Aech_08485* (a Zinc finger protein) in *A. echinatior*. (**a**) The gene model of *Aech_08485* with blue boxes representing CDS and white boxes UTRs, the red arrow below the 5th exon indicating the editing site position that causes K423R recoding in the protein sequence, and the two black vertical arrows above the gene model indicating the exon regions that code for the zf-C2H2 domain. (**b**) Editing levels of the K432R editing site based on the RNA-Seq data and (**c**) the TA-clonal sequencing data (ca. 50 randomly selected clones for each sample used for Sanger sequencing). *P*-values of Fisher's exact test between any two of the three castes within the same colony are given above the bars (* = $p < 0.05$, ** = $p < 0.01$ and *** = $p < 0.001$).



Supplementary Figure 11. Conservation of *Aech_08485* (a Zinc finger protein) in ants. Multiple sequence alignments for the protein (**a**) and mRNA sequences (**b**) of the *Aech_08485* gene and its adjacent recoding RNA editing site among closely and distantly related ant species, with arrows indicating the positions of the A-to-I RNA editing sites. (**c**) Editing levels of the K432R recoding site of *Aech_08485* in *C. floridanus*.



Supplementary Figure 12. Consistent differential editing of *pUf68* and *TBPH* as confirmed by RNA-Seq and TA-clonal sequencing data. Editing levels of the differentially edited site in 3'-UTR of *pUf68* (**a**) and *TBPH* (**b**). For TA-clonal sequencing 50-100 clones were randomly selected for Sanger sequencing for each sample. *P* values of Fisher's exact test between any two of the three castes within the same colony are given above the bars (* = $p < 0.05$, ** = $p < 0.01$ and *** = $p < 0.001$. *P* values between 0.05 and 0.1 are given in precise figures).

Supplementary Tables

Supplementary Table 1. Samples for RNA-Seq and DNA-Seq used in this study.

Colony number	Caste name	# Pooled individuals	Abbr. used in this paper
Ae322	Gyne	50	G322
	Large worker	50	L322
	Small worker	200	S322
Ae356	Gyne	50	G356
	Large worker	50	L356
	Small worker	200	S356
Ae363	Gyne	50	G363
	Large worker	50	L363
	Small worker	200	S363

Supplementary Table 2. Statistics of the raw data, mapped data and final usable data retained for further analysis. Mapped data represent the subset of data that could be mapped to the reference genome of *A. echinatior* by BWA, and usable data represent the final data set after filtering (see Methods).

Sample	# Raw reads	Raw bases (Gb)	Raw depth (X)	Mapped bases (Gb)	Mapped ratio (%)	Usable bases (Gb)	Usable ratio (%)	Usable depth (X)
DNA-Seq data								
G322	118,570,444	10.67	36	8.99	84.23	6.71	62.84	22.37
G356	105,468,350	9.49	32	7.84	82.62	5.81	61.18	19.37
G363	136,691,374	12.30	41	10.53	85.60	7.76	63.06	25.87
L322	142,002,082	12.78	43	10.37	81.15	7.72	60.43	25.73
L356	130,751,450	11.77	39	9.93	84.36	7.35	62.49	24.50
L363	146,023,764	13.14	44	10.99	83.66	8.05	61.27	26.83
S322	123,049,284	11.07	37	9.51	85.89	7.12	64.29	23.73
S356	154,832,982	13.93	46	12.43	89.17	9.23	66.24	30.77
S363	100,832,530	9.07	30	7.80	85.91	5.76	63.46	19.20
Mean	128,691,362	11.58	39	9.82	84.73	7.28	62.81	24.26
RNA-Seq data								
G322	137,548,082	12.38	41	10.52	85.01	2.95	23.83	9.89
G356	121,487,936	10.93	36	9.44	86.40	2.84	25.98	9.54
G363	113,919,462	10.25	34	8.37	81.69	2.92	28.49	9.83
L322	119,183,424	10.73	36	9.53	88.81	3.54	32.99	11.89
L356	116,574,962	10.49	35	8.57	81.67	2.42	23.07	8.16
L363	113,808,324	10.24	34	8.85	86.47	2.87	28.03	9.63
S322	134,815,104	12.13	40	10.65	87.83	3.83	31.57	12.86
S356	122,556,330	11.03	37	9.94	90.08	3.81	34.54	12.81
S363	120,232,642	10.82	36	9.50	87.77	3.39	31.33	11.37
Mean	122,236,252	11.00	37	9.49	86.19	3.17	28.87	10.66

Supplementary Table 3. Numbers and percentages of editing sites identified in each sample.

Number of editing sites													
Sample	**Total**	**A->C**	**A->I**	**A->U**	**C->A**	**C->G**	**C->U**	**G->A**	**G->C**	**G->U**	**U->A**	**U->C**	**U->G**
G322	9,634	2	9,301	3	15	0	49	198	0	4	15	47	0
G356	8,878	0	8,586	5	9	0	43	190	0	4	8	32	1
G363	9,055	3	8,813	6	11	0	37	154	1	1	7	22	0
L322	12,044	2	11,664	3	14	0	54	234	0	6	20	46	1
L356	8,146	1	7,836	10	11	0	54	177	0	4	11	40	2
L363	9,197	1	8,873	2	9	0	61	183	2	4	10	52	0
S322	13,248	1	12,843	11	12	0	53	226	3	6	18	73	2
S356	13,823	2	13,409	8	12	1	59	245	0	3	17	65	2
S363	12,412	3	12,033	10	14	2	54	208	0	5	18	65	0
Mean	10,715	2	10,373	6	12	0	52	202	1	4	14	49	1
Percentage of editing sites													
G322	100.00	0.02	96.54	0.03	0.16	0.00	0.51	2.06	0.00	0.04	0.16	0.49	0.00
G356	100.00	0.00	96.71	0.06	0.10	0.00	0.48	2.14	0.00	0.05	0.09	0.36	0.01
G363	100.00	0.03	97.33	0.07	0.12	0.00	0.41	1.70	0.01	0.01	0.08	0.24	0.00
L322	100.00	0.02	96.84	0.02	0.12	0.00	0.45	1.94	0.00	0.05	0.17	0.38	0.01
L356	100.00	0.01	96.19	0.12	0.14	0.00	0.66	2.17	0.00	0.05	0.14	0.49	0.02
L363	100.00	0.01	96.48	0.02	0.10	0.00	0.66	1.99	0.02	0.04	0.11	0.57	0.00
S322	100.00	0.01	96.94	0.08	0.09	0.00	0.40	1.71	0.02	0.05	0.14	0.55	0.02
S356	100.00	0.01	97.00	0.06	0.09	0.01	0.43	1.77	0.00	0.02	0.12	0.47	0.01
S363	100.00	0.02	96.95	0.08	0.11	0.02	0.44	1.68	0.00	0.04	0.15	0.52	0.00
Mean	100.00	0.01	96.78	0.06	0.11	0.00	0.49	1.91	0.01	0.04	0.13	0.45	0.01

Supplementary Table 4. Numbers of genes with one or more editing sites in each sample. The numbers of 5'-UTR + CDS + Intron + 3'-UTR are slightly higher than the total number of edited genes, as some genes are edited in more than one genic region.

Sample	Total edited gene	Number of genes edited in different genic regions			
		5'-UTR	CDS	Intron	3'-UTR
G322	937	68	96	441	412
G356	865	67	79	411	389
G363	920	75	79	460	389
L322	1,027	74	98	520	438
L356	812	59	78	381	365
L363	898	66	93	441	386
S322	1,082	85	105	548	455
S356	1,125	88	95	584	477
S363	1,051	83	110	510	455
Mean	969	74	93	477	418

Supplementary Table 5. Numbers and percentages of A-to-I editing sites occurring in clusters.

Sample	Total	>=3 sites in 50-bp window		>=3 sites in 100-bp window	
		Number	Ratio (%)	Number	Ratio (%)
G322	9,301	5,244	56.38	6,203	66.69
G356	8,586	4,814	56.07	5,641	65.70
G363	8,813	4,838	54.90	5,756	65.31
L322	11,664	6,877	58.96	8,070	69.19
L356	7,836	4,358	55.62	5,127	65.43
L363	8,873	4,995	56.29	5,814	65.52
S322	12,843	7,731	60.20	8,972	69.86
S356	13,409	8,327	62.10	9,560	71.30
S363	12,033	7,231	60.09	8,420	69.97
Mean	10,373	6,046	57.85	7,063	67.66

Supplementary Table 6. Numbers of A-to-I editing sites within coding regions (CDS) that have the potential to change protein sequences.

Sample	Sites in CDS	Recode amino acid	Remove stop codon	Introduce stop codon	Proportion[a]
G322	64	47	0	0	0.73
G356	63	47	0	0	0.75
G363	60	47	0	0	0.78
L322	84	62	0	0	0.74
L356	58	47	0	0	0.81
L363	77	59	0	0	0.77
S322	90	68	1	0	0.77
S356	87	67	1	0	0.78
S363	88	69	1	0	0.80
Mean	74.56	57.00	0.33	0.00	0.77

a. Proportion is calculated as (*Recode amino acid* + *Remove stop codon* + *Introduce stop codon*) / *Sites in CDS.*

Supplementary Table 7. Monte Carlo simulations for the percentages of conserved genomic sites corresponding to A-to-I RNA editing sites. The same number of transcribed 'A's (i.e. 'A's covered by ≥ 1 RNA reads) as the identified A-to-I editing sites was randomly sampled in each element, after which the percentages of conserved 'A's were calculated. *P*-values for each element were empirically estimated after 10000 simulation iterations that produced random expectation values.

Among seven attine ants				
Element	**# Site**	**% of conserved genomic sites**		***P* value**
		Observation	**Expectation**	
Genome	10,282	22.82	66.80	$< 1\times10^{-04}$
Gene	5,052	24.18	73.36	$< 1\times10^{-04}$
5'-UTR	266	12.83	59.68	$< 1\times10^{-04}$
CDS	96	59.38	85.88	$< 1\times10^{-04}$
Intron	2,259	20.95	60.05	$< 1\times10^{-04}$
3'-UTR	2,431	26.88	66.85	$< 1\times10^{-04}$
TE	7,081	14.00	20.34	$< 1\times10^{-04}$
Intergenic	1,636	38.45	65.55	$< 1\times10^{-04}$
Between *A. echinatior* and *C. floridanus*				
Genome	10,282	10.54	57.11	$< 1\times10^{-04}$
Gene	5,052	10.33	64.49	$< 1\times10^{-04}$
5'-UTR	266	3.02	45.13	$< 1\times10^{-04}$
CDS	96	46.88	80.25	$< 1\times10^{-04}$
Intron	2,259	9.80	47.94	$< 1\times10^{-04}$
3'-UTR	2,431	10.21	55.89	$< 1\times10^{-04}$
TE	7,081	3.70	12.89	$< 1\times10^{-04}$
Intergenic	1,636	26.65	53.45	$< 1\times10^{-04}$
Between *A. echinatior* and *H. saltator*				
Genome	10,282	8.33	46.12	$< 1\times10^{-04}$
Gene	5,052	8.32	54.25	$< 1\times10^{-04}$
5'-UTR	266	10.57	35.59	$< 1\times10^{-04}$
CDS	96	42.71	73.01	$< 1\times10^{-04}$
Intron	2,259	7.06	35.19	$< 1\times10^{-04}$
3'-UTR	2,431	7.81	41.89	$< 1\times10^{-04}$
TE	7,081	2.38	10.82	$< 1\times10^{-04}$
Intergenic	1,636	23.17	38.82	$< 1\times10^{-04}$

Supplementary Table 8. Monte Carlo simulations for the mean phastCons scores of genomic sites corresponding to A-to-I RNA editing sites. The same number of transcribed 'A's (i.e. 'A's covered by ≥ 1 RNA reads) as the identified A-to-I editing sites was randomly sampled in each element, after which the mean phastCons scores of the sampled 'A's for each element were calculated. The *p*-values for each element were empirically estimated after 10000 simulation iterations that produced random expectation values.

Genomic sites corresponding to A-to-I RNA editing sites with phastCons score				
Element	**# Site**	**Mean phastCons score**		***P* value**
		Observation	**Expectation**	
Genome	4,004	0.0831	0.4938	$< 1\times10^{-04}$
Gene	2,042	0.0830	0.5280	$< 1\times10^{-04}$
5'-UTR	69	0.0334	0.3297	$< 1\times10^{-04}$
CDS	70	0.5489	0.6576	0.008
Intron	823	0.0662	0.3710	$< 1\times10^{-04}$
3'-UTR	1,080	0.0689	0.4549	$< 1\times10^{-04}$
TE	1,681	0.0435	0.1945	$< 1\times10^{-04}$
Intergenic	1,148	0.1092	0.4518	$< 1\times10^{-04}$

Supplementary Table 9. RNA-Seq samples of *Camponotus floridanus* and *Harpegnathos saltator* used for comparative analyses in this study. All RNA-Seq data were generated in an Illumina sequencing platform. "PE" refers to paired-end sequencing and "SE" to single-end sequencing. Except for the brain samples of major workers and minor workers from *C. floridanus*, all RNA samples were extracted from whole individuals (see ref. [2], ref. [3]and ref. [8] for details).

Camponotus floridanus				***Harpegnathos saltator***			
Sample	**Data type**	**Raw data (Gb)**	**SRA Accession**	**Sample**	**Data type**	**Raw data (Gb)**	**SRA Accession**
Embryo	75 PE	2.5	SRX023137	Embryo	75 PE	2.9	SRX023143
Larva	75 PE	2.4	SRX023138	Larva	75 PE	3.1	SRX023144
Male	75 PE	2.5	SRX023139	Male	75 PE	3.3	SRX023140
Major worker	75 PE	3.0	SRX023142	Worker	75 PE	3.0	SRX023145
Minor worker	75 PE	2.9	SRX023141	Gamergate	75 PE	3.0	SRX023146
Virgin queen	90 PE	3.0	SRX091809	Virgin queen	90 PE	3.2	SRX091811
Queen	90 PE	2.7	SRX091808	Queen	90 PE	5.3	SRX091810
Major worker brain	50 SE, 36 SE	3.5, 1.0	SRX144055, SRX144056				
Minor worker brain	50 SE, 36 SE	3.3, 1.0	SRX144057, SRX144058				

Supplementary Table 10. Number of sites with significantly differential editing levels between two castes within the same colony.

Caste1-vs-Caste2	Total	Up-regulated in Caste1	Up-regulated in Caste2
G322-vs-L322	295	70	225
G356-vs-L356	294	61	233
G363-vs-L363	238	61	177
G322-vs-S322	602	105	497
G356-vs-S356	569	74	495
G363-vs-S363	627	84	543
L322-vs-S322	292	94	198
L356-vs-S356	337	118	219
L363-vs-S363	384	107	277

Supplementary Table 11. Linkage editing of the two A-to-I recoding editing sites of the *wah* gene in *A. echinatior* with *p*-values calculated by Binomial Tests. Linkage editing represents that two editing events occur on the same transcripts, in other words, they are supported by the same RNA-Seq reads.

Sample	Reads spanning two sites	Reads supporting Q966R editing	Reads supporting Q968R editing	Reads supporting linkage editing		*P* value
				Observed	Expected	
G322	117	29	26	23	6.4	7.59×10^{-08}
G356	101	34	32	27	10.8	3.54×10^{-06}
G363	123	19	25	19	3.9	9.70×10^{-09}
L322	178	74	70	62	29.1	1.02×10^{-09}
L356	102	33	36	26	11.6	3.96×10^{-05}
L363	112	37	39	27	12.9	8.71×10^{-05}
S322	208	83	91	59	36.3	3.32×10^{-05}
S356	203	106	91	75	47.5	5.12×10^{-06}
S363	204	76	81	57	30.1	5.62×10^{-07}

Supplementary Methods

1. Analysis of RNA editing site positioning

To test whether the A-to-I editing sites had a tendency to be located in or near dsRNA regions, we followed the method of Bahn, *et al*[9]. Briefly, we extracted 4,001 bp genomic sequences centered around each A-to-I editing site and reverse-complemented and aligned these against a 401 bp sequence centered around an A-to-I editing site using BLASTN (v2.2.23). After excluding cases of self-alignment and provided the second-best alignment had an alignment length of ≥ 50 bp and an identity of ≥ 80%, we inferred that the A-to-I editing site was located in a dsRNA structure. We then calculated the fraction of A-to-I editing sites that was associated with dsRNA structures. To evaluate the statistical significance of dsRNA enrichment, we randomly selected the same number of transcribed 'A's ('A's covered by ≥ 1 RNA reads) as controls, and calculated *p* values based on 10,000 simulations (Supplementary Fig. 3c).

To test whether the A-to-I editing sites had a tendency to be located in clusters, we used 50 bp and 100 bp sliding windows to scan clustering sites. We defined a window with ≥ 3 A-to-I editing sites as a cluster and then calculated the fraction of A-to-I editing sites located in clusters (Supplementary Table 5). We also calculated pairwise distances for all A-to-I editing sites to obtain frequency distributions of these distances (Supplementary Fig. 3d).

Sequence preferences for base positions flanking A-to-I editing sites were analyzed using the Two Sample Logo program[4] (http://www.twosamplelogo.org). The same number of transcribed 'A's were randomly selected as controls (Supplementary Fig. 3e). The official Gene Set v3.8 of *Acromyrmex echinatior* (without HGD-IDs, built on Assembly v2.0: http://hymenopteragenome.org/acromyrmex/) and repeat annotations of the corresponding genome version[10] were used to investigate the genomic distribution of all editing sites. The 5'- and 3'-UTRs were identified using the RNA-Seq data from this study based on known gene models.

2. Gene expression analysis based on RNA-Seq data

The Tophat v2.0.7 package[11] was employed to map RNA-Seq reads to the reference genome with parameters of --bowtie-n, --min-anchor 8, --splice-mismatches 0, --min-intron-length 50, --max-intron-length 500000, --segment-mismatches 3, --segment-length 30, --read-mismatches 3, --read-edit-dist 3, --solexa1.3-quals and --library-type fr-firststrand. Gene expression levels were measured by RPKM (reads per kilobase per million mapped exonic reads)[12], and adjusted by scaling normalization[13].

3. Evolutionary analysis of edited genes

One-to-one orthologs of seven attine ants (*Acromyrmex echinatior*, *Atta cephalotes*, *Atta colombica*, *Trachymyrmex septentrionalis*, *Trachymyrmex cornetzi*, *Trachymyrmex zeteki* and *Cyphomyrmex costatus*) and two outgroup species (*Solenopsis invicta*, *Pogonomyrmex barbatus*) were determined by reciprocal best blast hits (RBH) and genome-syntenic relationships using *A. echinatior* as the reference species. Then orthologous sequences were aligned by PRANK[5,6], and the non-synonymous (*dn*) and synonymous (*ds*) mutation rates and the *dn*/*ds* ratio were estimated for each gene using the codeml program with the free-ratio model in the PAML package[7]. Finally, the *dn*, *ds* and *dn*/*ds* values of RNA edited genes were compared with that of all expressed genes (RPKM > 1 in at least one of the nine sequenced samples) by Wilcoxon tests to estimate whether the values of RNA edited genes differed significantly from a background expectation (Supplementary Figure 7).

Supplementary References

1. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731-2739 (2011).

2. Bonasio, R. *et al.* Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**, 1068-71 (2010).

3. Bonasio, R. *et al.* Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol* **22**, 1755-64 (2012).

4. Vacic, V., Iakoucheva, L.M. & Radivojac, P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* **22**, 1536-7 (2006).

5. Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* **102**, 10557-62 (2005).

6. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-5 (2008).

7. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-91 (2007).

8. Simola, D.F. *et al.* A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res* **23**, 486-96 (2013).

9. Bahn, J.H. *et al.* Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res* **22**, 142-50 (2012).

10. Nygaard, S. *et al.* The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res* **21**, 1339-48 (2011).

11. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).

12. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-8 (2008).

13. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).