An Evolutionary Landscape of A-to-I RNA Editome across Metazoan Species

Li-Yuan Hung[1,†], Yen-Ju Chen[1,2,†], Te-Lun Mai[1,†], Chia-Ying Chen[1], Min-Yu Yang[1], Tai-Wei Chiang[1], Yi-Da Wang[1], and Trees-Juen Chuang[1,2,*]

[1]Genomics Research Center, Academia Sinica, Taipei, Taiwan

[2]Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: trees@gate.sinica.edu.tw.

Accepted: December 23, 2017

Abstract

Adenosine-to-inosine (A-to-I) editing is widespread across the kingdom Metazoa. However, for the lack of comprehensive analysis in nonmodel animals, the evolutionary history of A-to-I editing remains largely unexplored. Here, we detect high-confidence editing sites using clustering and conservation strategies based on RNA sequencing data alone, without using single-nucleotide polymorphism information or genome sequencing data from the same sample. We thereby unveil the first evolutionary landscape of A-to-I editing maps across 20 metazoan species (from worm to human), providing unprecedented evidence on how the editing mechanism gradually expands its territory and increases its influence along the history of evolution. Our result revealed that highly clustered and conserved editing sites tended to have a higher editing level and a higher magnitude of the ADAR motif. The ratio of the frequencies of nonsynonymous editing to that of synonymous editing remarkably increased with increasing the conservation level of A-to-I editing. These results thus suggest potentially functional benefit of highly clustered and conserved editing sites. In addition, spatiotemporal dynamics analyses reveal a conserved enrichment of editing and ADAR expression in the central nervous system throughout more than 300 Myr of divergent evolution in complex animals and the comparability of editing patterns between invertebrates and between vertebrates during development. This study provides evolutionary and dynamic aspects of A-to-I editome across metazoan species, expanding this important but understudied class of nongenomically encoded events for comprehensive characterization.

Key words: A-to-I RNA editing, ADAR, ADAR motif, dynamic editome, evolution.

Introduction

Adenosine-to-inosine (A-to-I) RNA editing is a common co- or post-transcriptional mechanism in metazoans, which is mediated by the ADAR (adenosine deaminases that act on RNA) family of proteins and converts adenosine (A) into inosine (I) (Bass 2002; Nishikura 2006). Since inosine is then recognized as guanine (G) in living cells, A-to-I editing is also known as A-to-G editing (Bass 2002; Nishikura 2006). Although an A-to-I editing event only alters one base at the RNA level, which seems to slightly affect the corresponding RNA molecule, emerging evidence shows that it can affect both transcription and translation in different aspects such as microRNA regulation (Rueter et al. 1999; Kawahara et al. 2007), alternative splicing (Rueter et al. 1999; Lev-Maor et al. 2007), structural alteration (Kimelman and Kirschner 1989; Wagner et al. 1989), and coding potential (Burns et al. 1997). RNA editing events have been reported to be highly associated with modification of regulatory RNAs (Kawahara et al. 2007; Tomaselli et al. 2015), cancer mechanisms (Maas et al. 2006; Han et al. 2015), embryogenesis (Wang et al. 2000; Osenberg et al. 2010), and brain development or neural differentiation (Wahlstedt et al. 2009; Hwang et al. 2016). These observations indicate the functional importance of A-to-I RNA editing not only for the immediate biological relevance but also for the potential molecular complexity.

The past decade has seen a rapid growth in efforts to identify A-to-I sites on a transcriptome-wide scale; a tremendous number of editing events have been detected (Levanon et al. 2004; Danecek et al. 2012; Peng et al. 2012; Ramaswami et al. 2013; Bazak et al. 2014; Ramaswami and

Li 2014; Picardi et al. 2015; Zhang and Xiao 2015). However, identification of RNA editing often severely hampers by false positives arising from RNA/DNA sequencing errors, single nucleotide polymorphisms (SNPs) or somatic mutations between the individual cells, incompleteness of genomic sequences and gene annotation, and mapping errors of the cDNA sequences to the reference genome (Bass et al. 2012; Ulbricht and Emeson 2014). Several RNA editing-detecting methods minimize false positives from SNPs by comparing the DNA and RNA sequences from a single individual (Bahn et al. 2012; Peng et al. 2012; Ramaswami et al. 2012). However, this approach simultaneously requires genome and transcriptome sequencing data from the same sample, limiting its practicability. To address the lack of genome sequencing data, a complete (or a partial) SNP database is often used to filter out SNPs (Bahn et al. 2012; Ramaswami et al. 2012, 2013; Bazak et al. 2014; Zhang and Xiao 2015; John et al. 2017; Sun et al. 2016). These approaches require a priori knowledge from known SNPs such as dbSNP, also decreasing their capability to identify editing sites in diverse organs/animals, especially in nonmodel species due to the lack of SNP information. Moreover, the phenomenon that most detected RNA editing sites on coding regions tend to be nonadaptive (Chen et al. 2014; Xu and Zhang 2014) reveal another challenge in extracting functionally beneficial editing events from the sea of sites with a very low level of editing (Pinto et al. 2014; Savva and Reenan 2014; Xu and Zhang 2015). These challenges thus limit our understanding of the evolutionary landscape and spatiotemporal dynamics of A-to-I editing in context.

Previously, it was observed an enrichment for clusters of detected editing sites, suggesting that the effect of sequencing errors, SNPs, and mutations on identification of A-to-I editing could remarkably decrease with increasing the number of consecutive variants (Levanon et al. 2004; Zaranek et al. 2010). The clustering strategy has been demonstrated to be effective to detect RNA editing sites without the need for SNP information or genome sequencing data from the same sample (Porath et al. 2014; Feng Zhang et al. 2017). To generate an evolutionary landscape of A-to-I RNA editome across metazoan species, most of which have no (or limited) SNP data, we employ the clustering strategy to identify A-to-I editing sites using RNA-seq data alone. For accuracy, we selected A-to-I editing sites by controlling for the fraction of A-to-G mismatches to all types of mismatches (designated as "%AG"; >95%) and false discovery rate (FDR; <1%) or cross-species conservation of A-to-I editing, and thereby identified 429,509 high-confidence A-to-I sites in diverse species from worm to human, including model and nonmodel animals. We thus constructed the first evolutionary landscape and spatiotemporal atlas of A-to-I editing across metazoan species. In contrast with previous studies that investigated editing dynamics in limited genes/species (Wahlstedt et al. 2009; Shtrichman et al. 2012; Veno et al. 2012; Picardi et al. 2015), our spatiotemporal atlas provided an unprecedented opportunity for studying dynamic editome in diverse species and for assessing the correlation between the global editing and expression of the two functional editors (ADAR1/ADAR2) in the context of evolution.

Materials and Methods

Data Retrieval and Access

The information of the RNA-seq data used in this study was listed in supplementary data set S1, Supplementary Material online. TEs and other repetitive elements were identified by RepeatMasker and were downloaded from the University of California Santa Cruz (UCSC) genome browser at https://genome.ucsc.edu/. The PhyloP scores were also downloaded from the UCSC genome browser. All supplementary data sets (S1–S7) are publicly downloadable at http://idv.sinica.edu.tw/trees/RNA_editing/RNA_editing.html. The in-house programs for identifying RNA editing sites are publicly accessible from GitHub at https://github.com/TreesLab/ICARES/tree/dev. A visualizable website was also provided (https://sites.google.com/site/recodedatabase/), which allows users to search for the identified RNA editing sites by providing the coordinates of genomic regions or gene symbols.

The Pipeline for Identifying A-to-I Sites

The RNA editing sites were identified by two strategies: the clustering strategy within the same species and the conservation strategy with cross-species comparison (fig. 1*A*). The clustering process was made up of three main phases to distinguish editing sites from technical and biological noises. First, we collected 309 RNA-seq data from 20 metazoan species (supplementary data set S1, Supplementary Material online), and used BWA (Li and Durbin 2009) (version 1.2.3) for short-read mapping. SNV calling was then performed using SAMtools pileup (Li et al. 2009) (version 1.0.2) with a pileup parser program downloaded from Galaxy (pileup_parser.pl with parameter settings: 3 9 10 8 20 1 "No" "Yes" 2 "Yes" "Yes"). The type of nucleotide substitution was determined on the basis of the reference genome and the strand information of the Ensembl-annotated genes (see supplementary data set S1, Supplementary Material online). Second, we selected dimorphic variants (i.e., sites with two distinct nucleotide types), and discarded singleton and multi-allelic ones. To eliminate potential strand-specific miscalls, we discarded variants with strand bias. For each species, we used MAQ (Li et al. 2008) to simulate all possible cDNA short-reads based on Ensembl-annotated transcripts, mapped these simulated reads to the reference genome, and compiled a mapping error set according to the mapping result. The called variants within this set were discarded since they were subject to mapping errors. Besides, sequencing errors were reported to occur in proximity (Nakamura et al. 2011). We then identified a

sequencing error set (SES) wherein variant calls of different mismatch types occurred in proximity (i.e., distance between each other <100 bp). Variants overlapped with the SES were subsequently discarded. In addition, for accuracy, only the variants with multi-sample evidence were retained. Third, the retained variants were compiled to identify high-confidence A-to-I sites using the clustering strategy. Of note, the editing type of each mismatch was determined on the basis of the Ensembl annotation. The clustering strategy compiled variants of the same mismatch type in close proximity (i.e., distance between each other <100 bp), because it was observed that the effect of sequencing errors, SNPs, and mutations on identification of A-to-I editing could remarkably decrease with increasing the number of consecutive variants (Levanon et al. 2004; Zaranek et al. 2010). The proximal distance was set as 100 bp because of the observation that the vast majority (>95%) of the previously identified A-to-I editing sites have at least one neighbor site within the proximal distance of 100 bp (supplementary table S1, Supplementary Material online). For each species, the qualified number of consecutive variants (qualified $N_{cluster}$) was determined by the FDR and %AG. The FDR of A-to-I sites was defined as FDR=the number of G-to-A mismatches/the number of A-to-G mismatches. The qualified $N_{cluster}$ was determined while the detected sites satisfied both of the two rules: %AG >95% and FDR <1%. For the conservation strategy, the A-to-I editing sites were identified from three clades (12 vertebrates, 4 *Drosophila* species, and 4 *Caenorhabditis* species), respectively. We considered the A-to-G mismatches that have passed the strand-bias, mapping error, and sequencing error filters described in Phase II (i.e., ad hoc filters; fig. 1*A*) for each species. Such A-to-G variants that were observed at the orthologous sites in more than one species of the examined clade were regarded as evolutionarily conserved A-to-I editing sites (fig. 1*A*). A full list of the identified A-to-I editing sites (including clustered and conserved editing sites) referred to the Ensembl annotation (releases 66 and 85) is given in supplementary data set S2, Supplementary Material online. Compared with previously identified A-to-I editing sites in human, mouse, and fly [collected in the well-known public databases: DARNED (Kiran and Baranov 2010), RADAR (Ramaswami and Li 2014), and REDIportal (Picardi et al. 2017)], we found that most the identified human editing sites (~90%) were also detected in the public databases (supplementary dat aset S2, Supplementary Material online). In contrast, more than one third of the identified mouse and fly A-to-I editing sites (92% for mouse and 33% for fly) were not found in these databases (supplementary data set S2, Supplementary Material online). To validate the newly identified A-to-I editing sites in mouse and fly, we selected 11 mouse and 12 fly editing sites and performed PCR amplification and Sanger sequencing of both DNA and RNA of mouse brain and wild-type fly from the same individual. Our result revealed that 10 mouse and 12 fly editing sites were experimentally confirmed (supplementary fig. S1*A* and *B*, Supplementary Material online; supplementary dat aset S3, Supplementary Material online). In addition to mouse and fly editing sites, which have been comprehensively detected and collected in public databases, we also selected nine editing sites from the identified chicken editing sites and successfully confirmed eight of them in chicken brain (supplementary fig. S1*C*, Supplementary Material online; supplementary data set S3, Supplementary Material online).

Samples and Validation

Genomic DNA and total RNA were extracted from frozen brain of mouse (C57BL/6J) at postnatal day 49, fresh brain of adult chicken, and wild-type fly (*Drosophila melanogaster*). Both genomic DNA and total RNA were obtained from the same individual/sample. Genomic DNA and total RNA were extracted by PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) and PureLink RNA Mini Kit (Thermo Fisher Scientific) with DNase I according to the manufacturer's instructions, respectively. Primers were designed in flanking sequences of the tested editing sites. 5 μg total RNA transforms to cDNA with SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) using Random Hexamer and Oligo-DT primers. 50 ng of genomic DNA and cDNA were used for the PCR analysis of the editing sites. The PCR was performed using DreamTaq Green PCR Master Mix (Thermo Fisher Scientific) on Veriti Thermal Cycler (Thermo Fisher Scientific). PCR products were validated by gel, and then treated with QIAquick Gel Extraction Kit (Qiagen). The editing sites were selected to be tested, if they were detected by the RNA-seq data (number of mapped reads at the site should be >10) to be editing events with editing level >10% in at least one sample. Sanger sequencing was performed to validate the corresponding editing positions of genomic DNA and cDNA sequences (see supplementary fig. S1, Supplementary Material online). The primer sequences used were listed in supplementary data set S3, Supplementary Material online.

Observed-to-Expected Ratio of "*G*"

The observed-to-expected (*O/E*) ratio of the presence of *G* was calculated to examine ADAR preference for A-to-I editing, which was defined as $P_{Obs(G)}/P_{Exp(G)}$, where $P_{Obs(G)}$ represented the frequency of observed presence of "*G*" immediately upstream or downstream to the A-to-I sites, and $P_{Exp(G)}$ represented that of expected presence of "*G*," which was calculated by the frequency of "*G*" in the examined genome. The statistical significance of difference between the observed and expected ratios of the presence of "*G*" was evaluated by the two-tailed Fisher's exact test using the R package (https://www.r-project.org/).

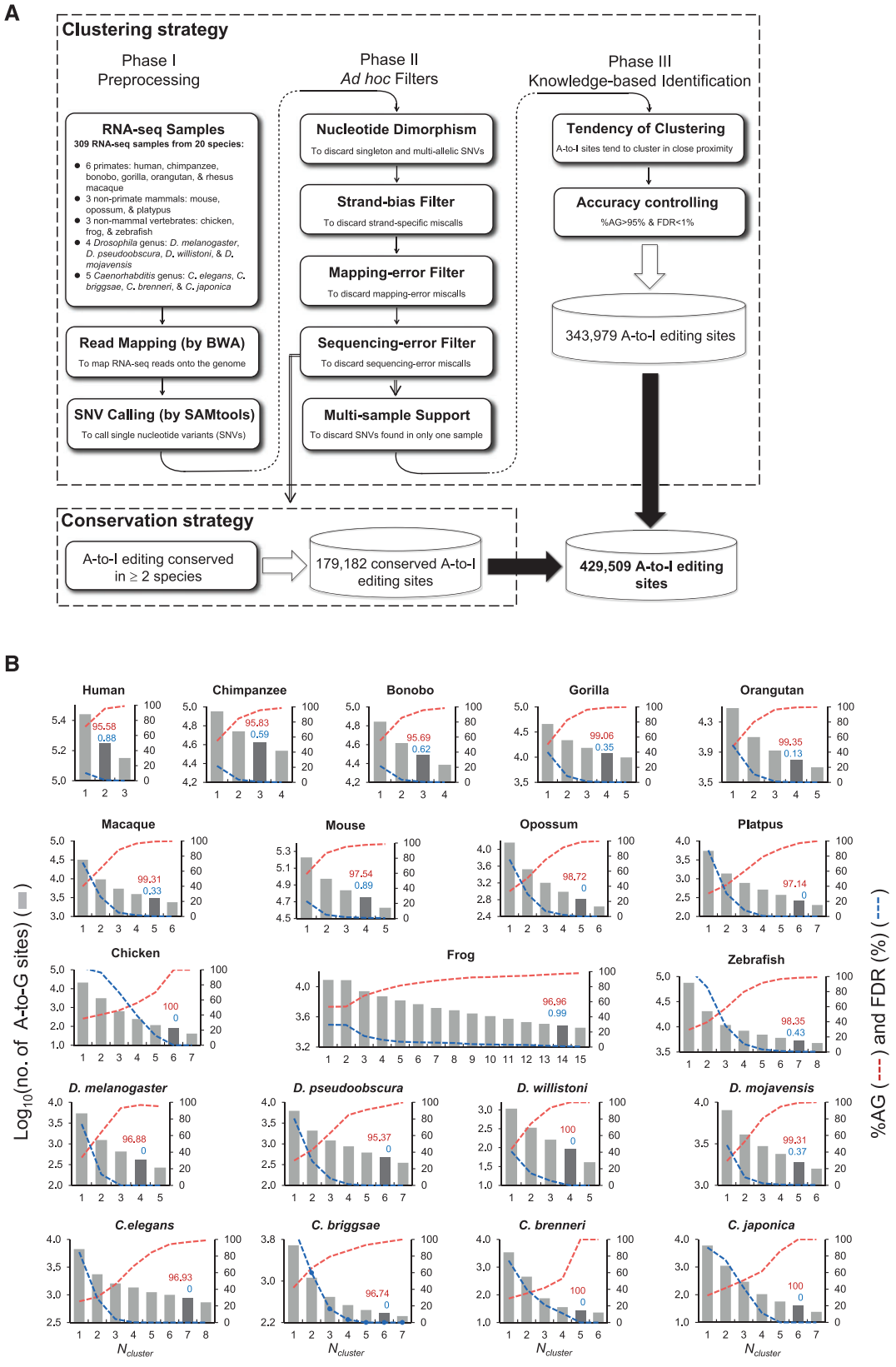

Fig. 1.—Identification of high-confidence A-to-I RNA editing sites, without a priori knowledge of SNP information. (*A*) Overview of the identification of RNA editing sites. The editing sites were identified by the clustering strategy within the same species or by cross-species comparison. The clustering process involved three main phases: Phase I: preprocessing, Phase II: ad hoc filters, and Phase III: ad hoc identification. (*B*) Correlation between the number of

Analysis of the Conservation Level of A-to-I Editing and Conservation Scores

To assess the correlation between the conservation level of nonsynonymous A-to-I editing and conservation scores, we first performed the UCSC LiftOver tool to convert genome coordinates of nonhuman species into those of human at the nonsynonymous sites of adenosine for the four categories of conservation (i.e., human–chimpanzee, human–chimpanzee–mouse, and human–chimpanzee–mouse–chicken conservation). For each category of conservation, the control sites with the same number of "A" of the corresponding category of conserved editing sites were selected randomly. The PhyloP score of each selected "A" was extracted from the UCSC genome browser. The simulation was performed 1,000 replicates for each category of conservation.

Comparative Analysis of A-to-I Editing Levels

We selected 16 animal individuals (2 for chimpanzee, 2 for bonobo, 2 for gorilla, 1 for orangutan, 2 for macaque, 3 for mouse, 1 for opossum, 1 for platypus, and 2 for chicken) for assessing editing dynamics in 5 organs (i.e., cerebellum, brain, liver, kidney, and heart). Of note, samples were selected for spatial profiling if they were eligible for measurement in five organs from a single individual. Each individual must contain all five types of organ samples with reliable RNA quality (i.e., RNA integrity number [Schroeder et al. 2006] >7.0). On the other hand, we used four species (i.e., *Caenorhabditis elegans*, *D. melanogaster*, zebrafish, and frog) in the profiling of developmental dynamics. Since A-to-I editing could influence its target sites in a local dsRNA region (Zaranek et al. 2010; Bazak Levanon, et al. 2014), the level of editing was determined in a clustering manner as

$$\text{A}-\text{to}-\text{I editing level} = \frac{\sum_i \text{Number of G in the clustered sites}}{\sum_i \text{Number of A and G in the clustered sites}}$$

with $i = 1, \ldots, N$, where N is the total number of editing sites in the designated cluster. For each individual in the spatial context or animal in the temporal context, the editing levels were accumulated and the highest value among the examined organs or development stages was used to normalize the A-to-I editing index. To ensure the statistical accuracy, we only considered the editing clusters with reasonable base coverage (i.e., the total number of *A* and *G* in the clustered sites $\geq$10) for the analysis of editing dynamics.

Measurement of ADAR mRNA Expression

ADAR mRNA expression was represented by BPKM (Bases Per Kilo-base of gene model per Million mapped bases) (Mortazavi et al. 2008). To improve the precision of mRNA expression measurement in poly(A)-selected RNA-seq samples, we only considered 3′UTR exons in the gene model of calculation.

Results

Identification of Editing Sites across 20 Metazoan Species

To investigate the evolutionary landscape and dynamic changes of A-to-I editing across metazoan species, we first collected RNA-seq data from various cell types of 20 species, including 6 primates, 3 nonprimate mammals, 3 nonmammalian vertebrates, 4 *Drosophila* species, and 4 *Caenorhabditis* species (309 samples in total; supplementary data set S1, Supplementary Material online). We then used the clustering strategy to detect RNA editing sites based on RNA-seq data alone, without the need for a priori knowledge from known SNPs or genome sequencing data from the same sample (Materials and Methods). Two useful measures are often applied to evaluating the specificity of the identified A-to-I editing sites: 1) the fraction of A-to-G mismatches to all types of mismatches (designated as "%AG") because of the extreme infrequency of non-A-to-G editing events (Kleinman and Majewski 2012; Lin et al. 2012; Pickrell et al. 2012); and 2) the ratio of the number of G-to-A mismatches to the number of A-to-G mismatches (i.e., FDR, see Materials and Methods) because of the assumption that G-to-A mismatches often reflected sequencing errors (Bahn et al. 2012; Liscovitch-Brauer et al. 2017). We can find that the percentage of A-to-G/T-to-C variants grew (or the FDR values decreased) rapidly with the increase of the number of consecutive single-nucleotide variants (SNVs) of the same type (designated as "$N_{cluster}$") in all examined species (fig. 1*B*). Of note, the T-to-C variants were considered, because the used RNA-seq data were not strand-specific and these variants could possibly be A-to-G editing sites in an antisense transcript. This revealed that a priori knowledge of the clustering tendency for A-to-I editing held well across species, including model and nonmodel animals. For accuracy, the qualified $N_{cluster}$ was determined while the detected sites satisfied both %AG >95% and FDR <1% (fig. 1*B*). Upon completion of the screening process, 343,979 A-to-G editing sites were identified in diverse species from worm to human. Since the more species the variants were supported by, the higher level

Fig. 1.—Continued

consecutive SNVs of the same type ($N_{cluster}$) and the two measures of specificity: %AG (the percentage of A-to-G & T-to-C among 12 SNV types) and FDR (the ratio of the number of G-to-A mismatches to the number of A-to-G mismatches). Histograms for each species represent A-to-G & T-to-C percentage in the subset of $\geq N_{cluster}$ consecutive SNVs of the same type. The dark gray histogram for each species represents the qualified $N_{cluster}$, which satisfies both %AG > 95% and FDR < 1%.

of functional potentiality (and thus accuracy) they contained (Bahn et al. 2012), we recovered the A-to-G variants at the orthologous sites in multiple species (editing sites supported by more than one species; the conservation strategy). By doing so, 179,182 evolutionarily conserved A-to-G editing sites were identified in the 20 species. Integrating the editing sites identified by these two strategies, we totally identified 429,509 editing sites (table 1 and supplementary data set S2, Supplementary Material online). Of note, we only considered the sites that were located within genic regions. We emphasize that the identified editing sites are highly confident with controlling for FDR $< 1\%$ and %AG $> 95\%$ (the clustering strategy) or evolutionary conservation of editing in multiple species (the conservation strategy).

The Correlation between $N_{cluster}$, Conservation Level of Editing, Editing Level, and the Magnitude of the ADAR Motif

The A-to-I editors (i.e., ADAR proteins) are known to be highly conserved across species (Bass 2002). Previous studies have observed that in some animals ADARs have a sequence preference (or targets) for "G" depletion and "G" enrichment at the 5′ and 3′ neighbor nucleotides next to A-to-I editing sites, respectively (Kiran and Baranov 2010; Lehmann and Bass 2000; Eggington et al. 2011; Ramaswami et al. 2013; Chen et al. 2014; Pinto et al. 2014; Porath et al. 2014; Alon et al. 2015; Liscovitch-Brauer et al. 2017). We evaluated the observed-to-expected (*O*/*E*) ratios of the presence of "*G*" immediately upstream or downstream to the A-to-I editing sites (see Materials and Methods) and showed that the trend of the known ADAR motif generally held true across metazoan species, regardless of the detection strategies of A-to-G editing (i.e., the clustered and conserved sites; fig. 2*A* and *B*).

We processed to examine the correlations among the magnitude of clustering of editing sites ($N_{cluster}$), the conservation level of editing, the magnitude of the ADAR motif, and editing level. First, we found a positive correlation between the magnitude of the ADAR motif and $N_{cluster}$. Such a trend that the magnitude of the ADAR motif increased with increasing $N_{cluster}$ became flat after the qualified $N_{cluster}$ (fig. 2*C*). Second, we classified the identified human editing sites into three groups according to the conservation level (supplementary data set S4, Supplementary Material online) and found that the magnitude of the ADAR motif significantly increased with increasing the conservation level of A-to-I editing (fig. 2*D*). This also reflected the sequence and structural conservation in the double-stranded RNA binding domains (dsRBDs) of ADARs across 14 vertebrates (supplementary fig. S2*A*, Supplementary Material online), in which the protein residues involved in RNA binding (Stefl et al. 2010) were even highly conserved from vertebrates to *Drosophila* (supplementary fig. S2*B* and *C*, Supplementary Material online). These results respond to a recent notion that *cis* sequence changes are highly associated with the evolution of RNA editing between *Drosophila* species (Sapiro et al. 2015). Third, we divided all examined human editing sites into 20 equal-size bins according to their editing levels and found a remarkably positive correlation between the magnitude of the ADAR motif and editing level (fig. 2*E*). Of note, if a detected site appeared in multiple samples, the highest level observed was considered. Finally, we also observed that both $N_{cluster}$ (fig. 2*F*) and the conservation level of editing (fig. 2*G*) were positively correlated with editing level. The correlations among $N_{cluster}$, the conservation level of editing, the magnitude of the ADAR motif, and editing level were summarized in figure 2*H*. Since a relatively higher editing level at a site is expected to be functionally important (Xu and Zhang 2015), our results suggested the biological significance of highly clustered and conserved editing sites.

Evolutionary Analysis of A-to-I Editomes

Since transposable elements (TEs) are pervasive in the genomes of higher eukaryotes, the double-stranded RNA structure formed by TE-complementary pairs often provides an ideal target for ADAR binding (Levanon et al. 2004). To understand the extent of ADAR substrates among different lineages and species, we constructed an evolutionary landscape of A-to-I editing maps (i.e., A-to-I editomes). We showed that the scaffolding of A-to-I editomes was highly associated with the expansion of TEs (fig. 3*A*). We found that the distribution of clustered A-to-I sites pertaining to TEs (fig. 3*B*) generally reflected the TE density in the genome (fig. 3*A*), echoing the observation that A-to-I editing tended to be clustered within TEs (Kim et al. 2004). In addition, the landscape was likely to encompass all the A-to-I editing events that were highly conserved across species, whilst contained lineage- or species-specific events that might have been recruited as a result of divergent evolution. Importantly, short interspersed elements (SINEs), which were apparently enriched in the mammalian genomes (fig. 3*A*, top), constituted a major landmark in mammalian editomes (fig. 3*A*, bottom). This co-evolution of SINEs and editomes might accelerate the divergence of species, not only in the genomic content, but also in the transcriptomic complexity. For example, in primates, genes associated with neurological processes or disorders were more likely to develop a host of SINE-mediated substrates of A-to-I editing (Paz-Yaacov et al. 2010), possibly resulting in the development of more complex transcriptome in the primate brain.

To further examine the biological significance of conserved editing sites, we retrieved 66,689 primate-only, 39 mammal-only, and 16 vertebrate-conserved editing events according to the conservation level of the human A-to-I editing events identified by our conservation strategy (supplementary data set S4, Supplementary Material online). Of note, primate-only events were human editing events observed at the

Table 1
Summary of A-to-I Editing Sites Identified in This Study

Organis	Number of Samples	Number of A-to-I Editing Sites[a]				Host Genes
		Clustering Strategy (A)	Conservation Strategy (B)	Total Sites (A∪B)	TE Sites[b]	
Homo sapiens (human)	45	177,734	66,800	222,300	213,772	11,035
Pan troglodytes (chimpanzee)	15	42,082	45,117	54,891	52,939	5,292
Pan paniscus (bonobo)	12	30,807	37,021	42,420	40,911	4,929
Gorilla gorilla (gorilla)	11	12,169	15,936	19,956	18,561	3,442
Pongo abelii (orangutan)	9	6,230	8,828	11,125	10,820	2,357
Macaca mulatta (macaque)	13	3,034	4,593	6,206	5,997	1,691
Mus musculus (mouse)	32	57,125	115	57,143	47,859	3,738
Monodelphis domestica (opossum)	12	661	27	682	588	89
Ornithorhynchus anatinus (platypus)	12	267	7	274	253	36
Gallus gallus (chicken)	12	77	17	94	30	20
Xenopus tropicalis (frog)	40	4,495	4	4,499	1,812	146
Danio rerio (zebrafish)	8	5,331	5	5,336	4,571	340
Fly species						
D. melanogaster	34	317	229	534	208	114
D. pseudoobscura	8	464	206	636	N/A[c]	112
D. willistoni	8	93	63	149	N/A[c]	124
D. mojavensis	4	1,894	214	2,065	208	410
Nematode species						
Caenorhabditis elegans	16	888	0	888	584	57
C. briggsae	6	241	0	241	164	21
C. brenneri	6	28	0	28	N/A[c]	3
C. japonica	6	42	0	42	N/A[c]	2

[a]The detailed information of the identified sites (e.g., genomic location and host genes) is listed in supplementary data set S2, Supplementary Material online.
[b]TE sites: A-to-I editing sites that are located in TEs.
[c]TE information of the species is not available.

orthologous sites in nonhuman primate(s) but not observed in nonprimate vertebrates. Mammal-only ones were human editing events observed at the orthologous sites in both nonhuman primate(s) and nonprimate mammal(s) but not observed in nonmammal vertebrates. Vertebrate-conserved ones were human editing events simultaneously observed at the orthologous sites in nonhuman primate(s), nonprimate mammal(s), and nonmammal vertebrate(s). We found that the great majority of the conserved editing events (99%) were primate-only, of which 98% (65,109 out of 66,689 events) were located in TEs, especially in the primate-specific *Alu* sequences (63,745 events; 96%), whereas no mammal-only or vertebrate-conserved events occurred in *Alu* repeats (fig. 4*A*). This also reflected our abovementioned notion that the scaffolding of A-to-I editomes could be raised by the introduction of TEs in a lineage-specific manner (fig. 3). We examined the effect of the conservation level on the distribution of editing events. We found that the only 0.23% of primate-only events caused amino acid changes (nonsynonymous changes), whereas as high as 26% and 88% of mammal-only and vertebrate-conserved events leaded to nonsynonymous changes, respectively (fig. 4*B*). This reveals that the percentage of nonsynonymous editing sites increase with increasing the conservation level of A-to-I editing. Furthermore, considering the total "*A*" sites within the human coding sequences that would cause nonsynonymous and synonymous changes if edited to "*G*," respectively, we calculated the frequencies of nonsynonymous (f_n) and synonymous (f_s) editing (Xu and Zhang 2014). If nonsynonymous editing events are generally deleterious and are destined to selective elimination, f_n should exhibit remarkably smaller than f_s (Chen 2013; Xu and Zhang 2014). To examine the relationship between the conservation level of A-to-I editing and the f_n-to-f_s ratio, we retrieved human nonconserved, human–chimpanzee shared, human–chimpanzee–mouse shared, and human–chimpanzee–mouse–chicken shared synonymous (or nonsynonymous) editing sites and the corresponding shared "*A*" sites that would have synonymous (or nonsynonymous) changes if edited to "*G*," respectively (supplementary table S2, Supplementary Material online). Intriguingly, we found that the f_n-to-f_s ratio remarkably increased with increasing the conservation level of A-to-I editing, and the ratio even exhibited ~100% >1 for the human–chimpanzee–mouse–chicken shared events (fig. 4*C*). A previous study reported that nonsynonymous editing events were observed to occur less frequently than expected by

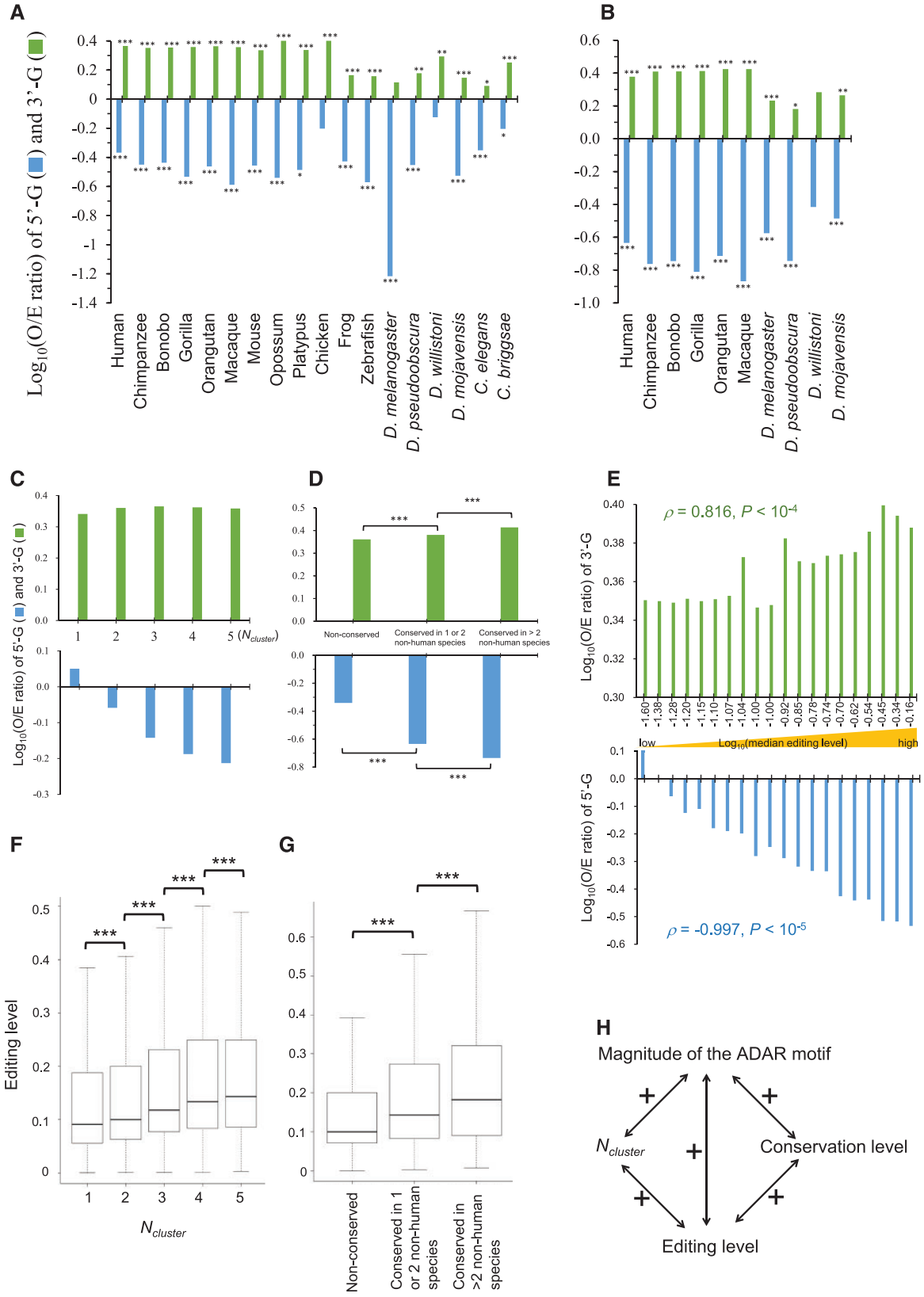

Fig. 2.—The *cis*-preference of ADARs and editing level for the detected A-to-G editing sites. (*A, B*) The *cis*-preference of ADARs (or the ADAR motif, which was measured by the observed-to-expected (*O/E*) ratio of the presence of "*G*" immediately upstream and downstream to the A-to-G editing sites) for

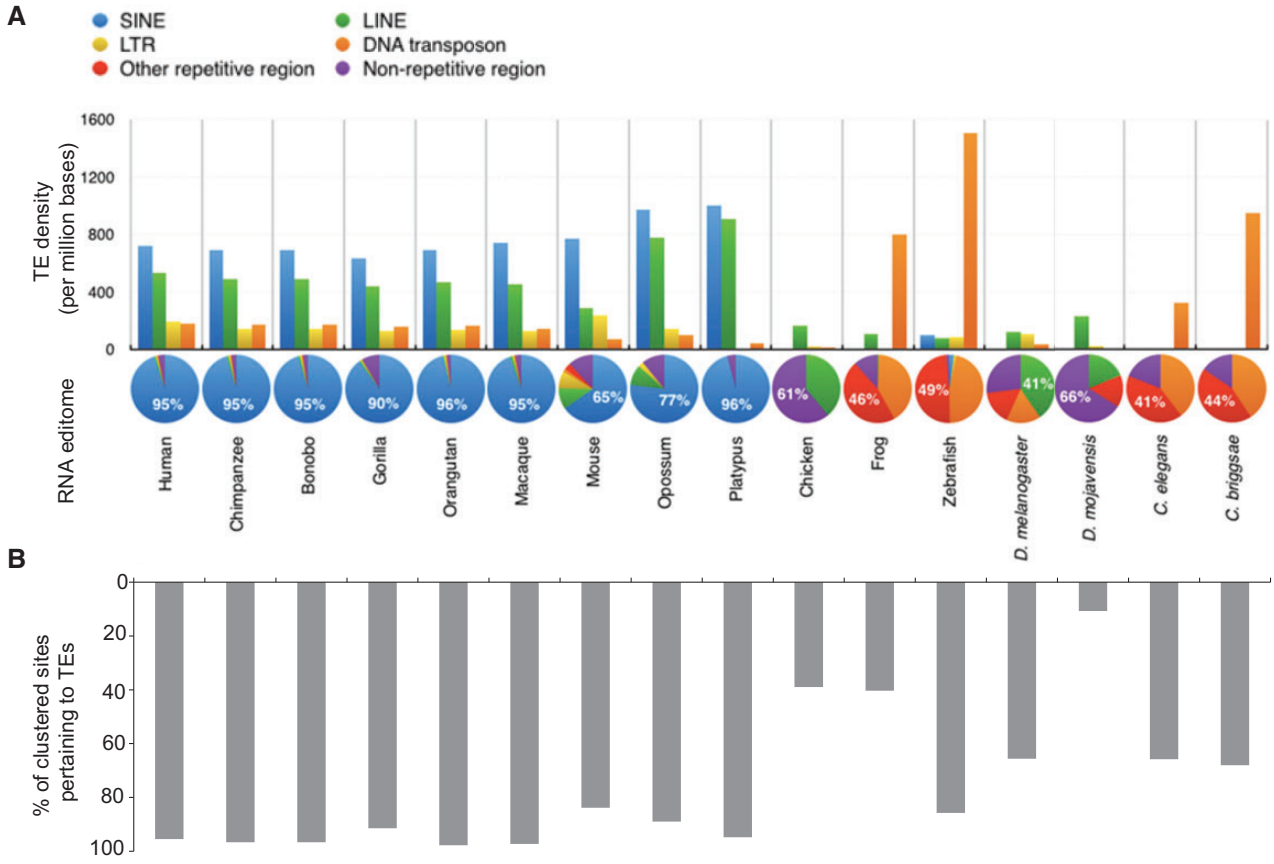

Fig. 3.—The relationship between the expansion of TEs and the increase of A-to-I editing sites. (*A*) The average numbers of TEs (i.e., SINEs, LINEs, LTRs, and DNA transposons) per million bases (top) and the compositions of A-to-I editing sites in the four types of TEs (i.e., SINE, LINE, LTR, and DNA transposon), other repetitive region, and nonrepetitive region (bottom). (*B*) The distribution of clustered A-to-I sites pertaining to TEs across species. TE: transposable element. SINE: short interspersed nuclear element. LINE: long interspersed nuclear element. LTR: long terminal repeat element.

chance, suggesting that coding RNA editing is generally not beneficial (Xu and Zhang 2014). However, if the editing events were highly conserved across species, we observed a different trend (fig. 4*C*). We further examined the correlation between the conservation level of nonsynonymous A-to-I editing and the conservation level of the corresponding individual nucleotides (measured by the PhyloP score) (Pollard et al. 2010). We found that the PhyloP scores of the conserved editing sites (including human–chimpanzee, human–chimpanzee–mouse, and human–chimpanzee–mouse–chicken shared editing sites) were generally higher than control (by simulating the sequence data to infer the PhyloP scores; Materials and Methods) and the conservation level of nonsynonymous A-to-I editing was indeed positively correlated with the PhyloP scores (fig. 4*D*). This also reflected the above observations that both the magnitude of the ADAR motif and editing level were positively correlated with the conservation level of editing (fig. 2*D*, *G*, and *H*). These results thus support a previous assumption that cross-species shared nonsynonymous editing is potentially beneficial and unlikely due to nonediting-related processes that may cause the inevitable consequence of sequence conservation (Xu and Zhang 2015). For example, considering the 87 human-editing events highly conserved in 5 nonhuman vertebrates

Fig. 2.—Continued
the editing sites identified by the clustering (*A*) and conservation (*B*) strategies across metazoan species. Only the species with more than 50 detected editing sites were considered. (*C*) The correlation between the magnitude of the ADAR motif and the magnitude of clustering of editing sites ($N_{cluster}$) in human. (*D*) The correlation between the magnitude of the ADAR motif and the conservation level of editing. (*E*) The correlation between the magnitude of the ADAR motif and editing level (measured by the Spearman's rank correlation). (*F, G*) The correlation between editing level and $N_{cluster}$ (*F*) and the conservation level of editing (*G*). (*H*) The correlations among $N_{cluster}$, the conservation level of editing, the magnitude of the ADAR motif, and editing level. "+" represents a positive correlation. The statistical significance was evaluated using the two-tailed Fisher's exact test (*A, B, D*), the Spearman's rank correlation (*E*), and the two-tailed Wilcoxon rank-sum test (*F, G*) using the R package: **P* value < 0.05, ***P* value < 0.01, and ****P* value < 0.001.

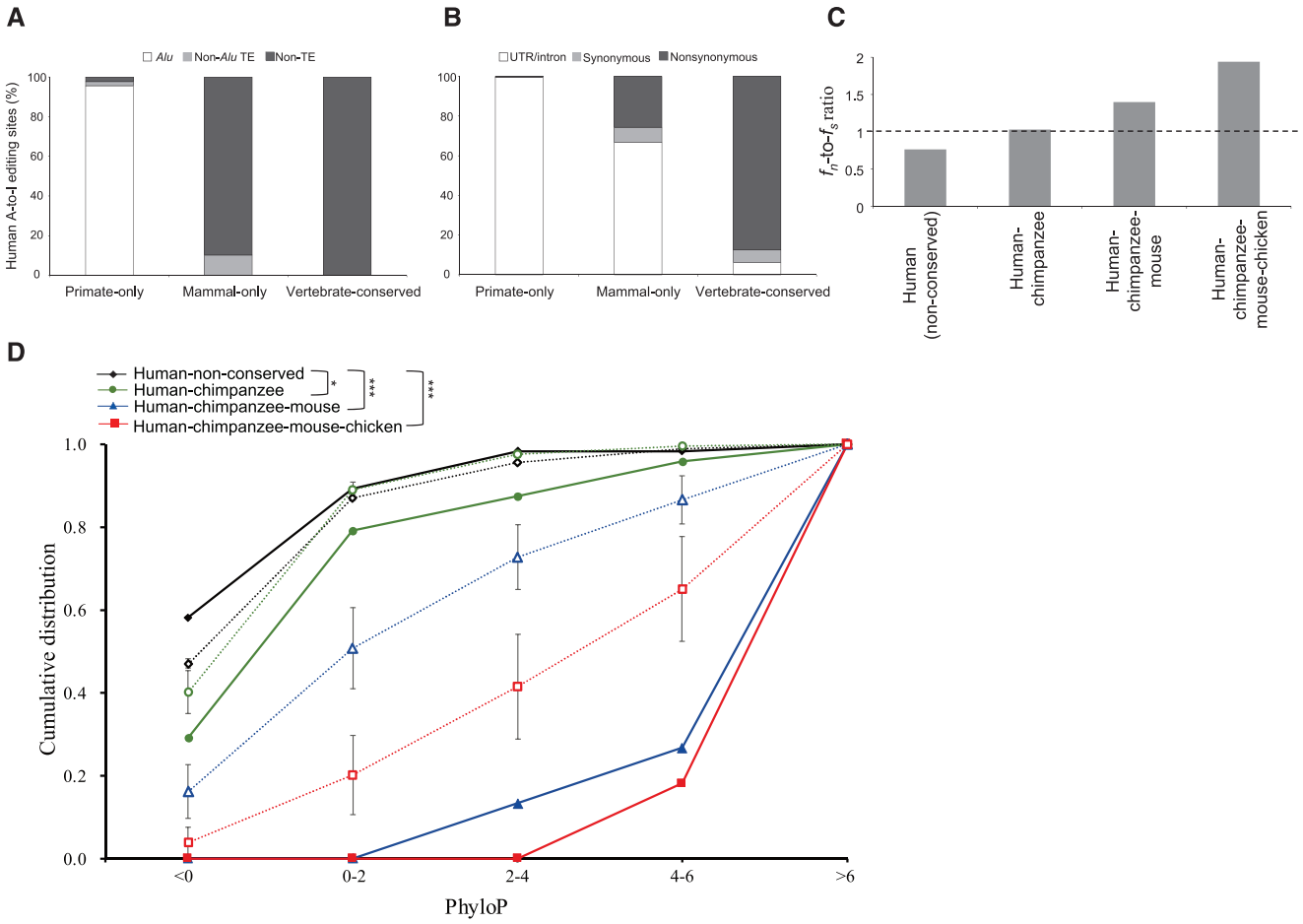

Fig. 4.—Analysis of the identified human editing events according to different conservation levels of A-to-I editing. (*A*, *B*) The distribution of human A-to-I editing sites (*A*) located in *Alu*, non-*Alu* TE, and non-TE regions, and (*B*) located in UTR/intron and editing sites leading to synonymous and nonsynonymous changes for primate-only, mammal-only, and vertebrate-conserved editing events. UTR: untranslated region. (*C*) The f_n-to-f_s ratios for human (all identified human A-to-I editing sites), human–chimpanzee shared, human–chimpanzee-mouse shared, and human–chimpanzee–mouse–chicken shared A-to-I editing sites. (*D*) Comparison of the conservation level of nonsynonymous A-to-I editing and conservation scores (measured by the PhyloP score). The empty diamond, circle, rectangle, and triangle represents the control (the average values of PhyloP scores of the simulation; see Materials and Methods) for each bin of the four categories of conservation, respectively. The error bar represents the standard error of the mean. The *P* values were estimated by the Kolmogorov–Smirnov test. **P* value < 0.05 and ****P* value < 0.001.

(supplementary data set S4, Supplementary Material online), all the nonsynonymous editing events (18 events located in 15 genes) were conserved across primates and nonprimate vertebrates. In fact, alteration of A-to-I editing has been shown to affect the function of all these 15 genes in diverse species (supplementary table S3, Supplementary Material online), further supporting the functional importance of these evolutionarily conserved nonsynonymous editing events.

Spatiotemporal Dynamics of A-to-I Editing across Species

To survey the transcriptome-wide dynamics of A-to-I editing, we constructed the spatiotemporal atlas in diverse species, including nine species with spatial profiles in different organs and four species with temporal stages during development. Samples were selected for spatial profiling if they were eligible for measurement in five organs (cerebellum, brain, liver, kidney, and heart) from a single individual, and for temporal profiling if they were based on a time-course experiment using a single animal strain. The constructed spatiotemporal atlas thus enabled us to assess the correlation between the global editing (supplementary data set S5, Supplementary Material online) and expression of the two functional editors (i.e., ADAR1 and ADAR2 expression; supplementary data set S6, Supplementary Material online) in the context of evolution.

In the spatial context, consistent with a previous observation (Picardi et al. 2015), we found that A-to-I editing tended to be tissue-specific (fig. 5*A*), and that most of the editing events observed in only one tissue were brain-/cerebellum-specific (fig. 5*B*). The clustered heatmap analysis

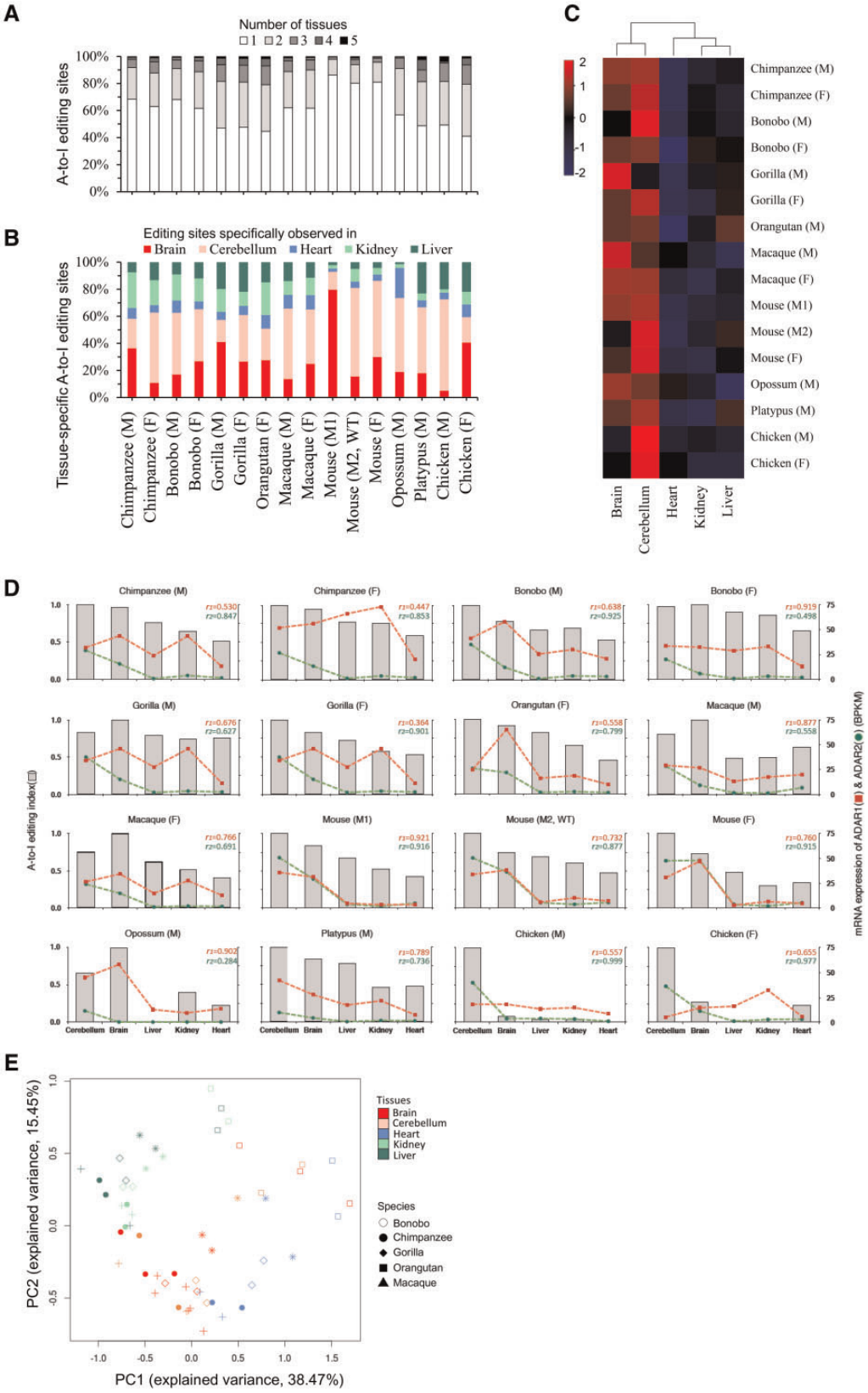

Fig. 5.—Spatial profiling of A-to-I editing in five organs (cerebellum, brain, liver, kidney, and heart) among metazoans. (*A*) Analysis of tissue-specificity of A-to-I editing. (*B*) Distribution of tissue-specific A-to-I editing sites in varied individuals across species. (*C*) Clustered heatmap for A-to-I editing across five types

on 16 individuals across species further revealed that A-to-I editing events grouped tissue samples into 2 distinct groups (fig. 5*C*). This reflected that A-to-I editing activity was more abundant in the central nervous system (CNS, e.g., cerebellum and brain) than in the rest (liver, kidney, and heart), regardless of species, sex (fig. 5*C* and *D*, and supplementary fig. S3, Supplementary Material online), or where the editing sites were located (TE vs. non-TE and coding vs. noncoding; supplementary fig. S4, Supplementary Material online). These results indicate a highly conserved, enriched pattern of editing activity in the CNS. We further showed that the expression levels of ADAR1/ADAR2 were generally higher in the CNS as compared with other organs (fig. 5*D*). Such an expression pattern holds well among amniotes from primates to birds, suggesting both editors are subject to strong evolutionary constraint in the perspective of gene expression. In addition, a positive correlation between A-to-I editing levels in tissues and ADAR (ADAR1 and ADAR2) expression was generally observed across species (fig. 5*D* and supplementary fig. S3, Supplementary Material online). We asked whether our result may bias toward the editing sites identified by our stringent criteria (i.e., the clustering and conservation strategies). To address this, we extracted the A-to-G variants that failed to pass our strategies in the three mouse individuals examined in this study (for which the five examined organs were obtained from the same individual) but were previously identified to be editing sites (collected in DARNED or RADAR; supplementary data set S7, Supplementary Material online). We performed the similar analysis and observed a consistent result (supplementary fig. S5, Supplementary Material online). These results thus suggest that global A-to-I editing activity could be largely attributed to the two editors in tissues, although other confounding factors may also affect the regulation of A-to-I editing activity. This result also reflected the enrichment of RNA editing activity in the CNS. Furthermore, considering editing levels of 589 one-to-one orthologous sites in 5 organs from 5 primates, the principal component analysis (PCA) showed that the data tended to be clustered by organ (fig. 5*E*). This result also reflects the tendency of tissue-specificity of A-to-I editing (fig. 5*A* and *B*).

In the temporal context, we examined the dynamics of A-to-I editing in two invertebrates (*C. elegans* and *D. melanogaster*) during the development from embryo to adult (fig. 6*A*) and two vertebrates (zebrafish and frog) during the embryogenesis from cleavage to organogenesis (fig. 6*B*), respectively. We observed that the dynamics of global editing exhibited one-stage lagging when compared with the fluctuation of ADAR expression (supplementary fig. S6, Supplementary Material online), as we used mRNA expression to represent the actual abundance of ADARs proteins. When taking this into account, ADAR expression was precisely synchronized with global editing in all examined species (fig. 6*A* and *B*). In *C. elegans*, consistent with a previous study (Tonkin et al. 2002), the first larval stage (L1) exhibited a strong signal of editing during embryonic and larval (L1–L4) development (fig. 6*A*, left). Intriguingly, another strong signal was present during dauer stages (i.e., dauer entry, dauer, and dauer exit) (fig. 6*A*, left), which represented an alternative development pathway in response to environmental stress. In *D. melanogaster*, similarly, an elevated trend of editing was also observed in L1; and the patterns of global editing were generally comparable with the patterns observed in *C. elegans* during development (fig. 6*A*, right). Such an elevated trend of editing in L1 generally held during worm and fly development. In vertebrates, frog and zebrafish shared a conserved pattern of editing during embryogenesis: A-to-I editing was highly active in early embryonic stages, and rapidly fell off during gastrulation (from shield to bud for zebrafish, and from 6 to 12 h for frog), followed by a stable trajectory that remained at low levels in later stages (fig. 6*B*). These results thus suggest that A-to-I editing may play a stage-dependent role during development. The similar patterns of editing between invertebrates (nematode and fly) and between vertebrates (zebrafish and frog) during development were generally observed for TE, non-TE, coding, and noncoding editing events (fig. 6*C* and supplementary fig. S7, Supplementary Material online). To avoid a possible bias toward our identified editing sites, we also analyzed the A-to-G variants that were previously identified to be editing sites (collected in DARNED or RADAR) but excluded by our stringent strategies in the fly samples examined in this study (supplementary data set S7, Supplementary Material online), and showed the similar results as above (supplementary fig. S8, Supplementary Material online). The comparability of global editing patterns between species implied that editing activity principally reflects ADAR expression inherited from their common ancestors. Intriguingly, we found 22 nonsynonymous editing sites (in 13 genes; supplementary table S4, Supplementary Material online), all of which were not

Fig. 5.—Continued

of tissues, with rows representing accumulated editing levels of detected editing events and columns representing tissues. (*D*) Correlation between A-to-I editing activity and ADAR expression in the spatial context. Transcriptome-wide activities of editing were examined in the five organs from the same individual. For each individual, the highest editing level among the five organs was used to normalize the A-to-I editing index (left *Y*-axis; Material and Methods). ADAR expression levels were estimated in terms of BPKM (right *Y*-axis; Material and Methods). Pearson's coefficient of correlation (*r*) (performed by the R package) was used to evaluate the correlation between A-to-I editing index and ADAR (ADAR1 (r_1) and ADAR2 (r_2)) expression levels. (*E*) PCA based on the editing levels of 589 orthologous sites in 5 organs from 5 primates. PCA was performed by the "princomp" function in the "stats" package of the R package. The distance metric between samples was calculated by $1 - |\rho|$. ρ represents pairwise Spearman's correlation coefficient of RNA editing level between samples. M, male; F, female; WT, wild type (mouse).

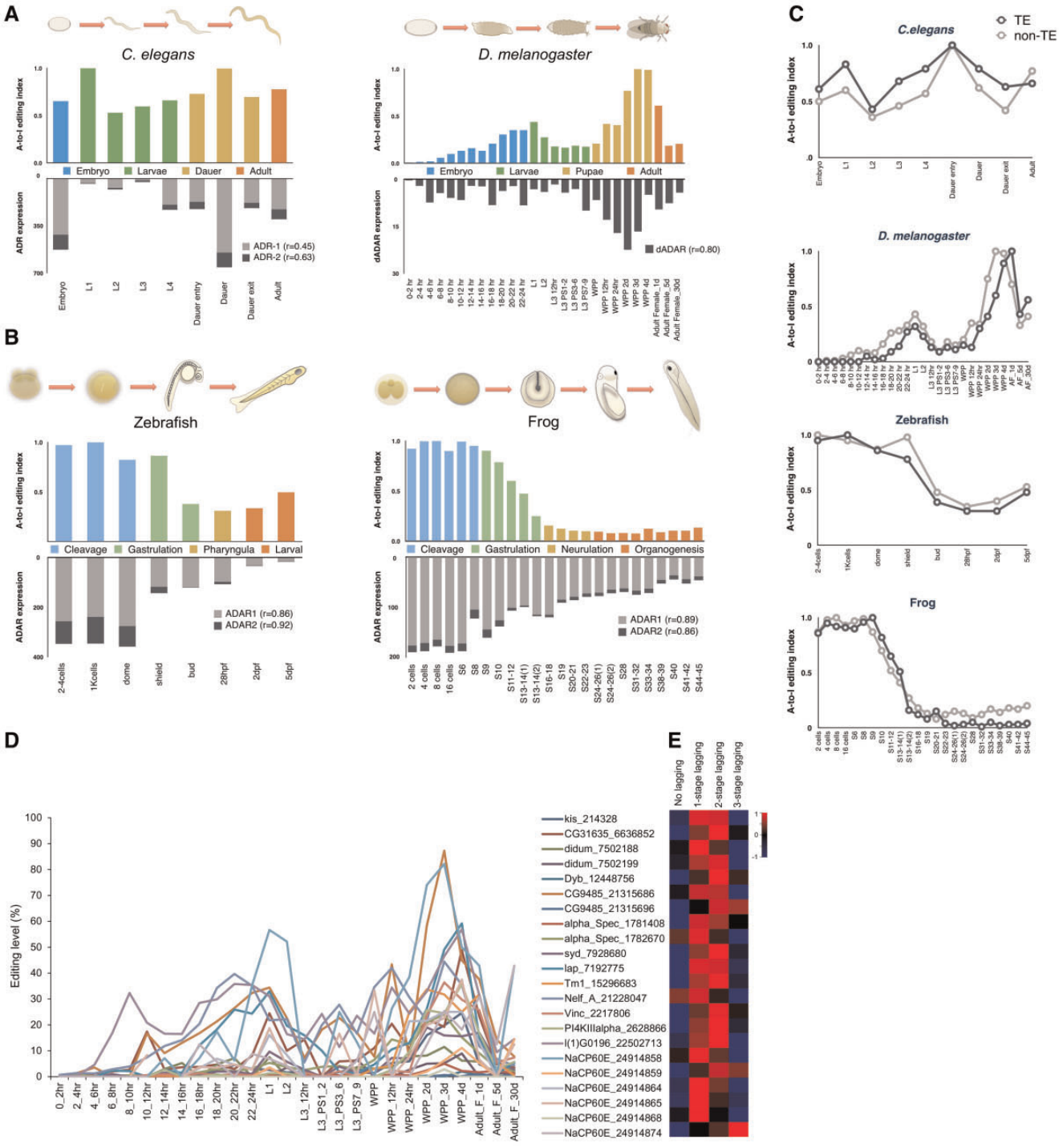

Fig. 6.—Temporal profiling of A-to-I editing among metazoans. (*A, B*) Temporal dynamics of A-to-I editing and ADAR expression during (*A*) invertebrate (*C. elegans* and *D. melanogaster*) and (*B*) vertebrate (zebrafish and frog) developments. For each animal, the highest editing level during development was used to normalize the A-to-I editing index (left *Y*-axis; see Material and Methods). ADAR expression levels were estimated in terms of BPKM (right *Y*-axis). Pearson's *r* between A-to-I editing index and ADAR expression level was calculated by considering the one-stage lagging of A-to-I editing as the fluctuation of ADAR expression during development (see the text). ADARs represent ADR-1 and ADR-2 for *C. elegant*, dADAR for *D. melanogaster*, and ADAR1 and ADAR2 for vertebrates (zebrafish and frog). (*C*) Temporal profiling of TE- and non-TE-associated A-to-I editing during *C. elegans*, *D. melanogaster*, zebrafish, and frog developments. For each animal, the editing levels of TE- and non-TE-associated sites were accumulated, respectively. The highest editing level during development was used to normalize the A-to-I editing index. (*D*) Changes in A-to-I editing levels for nonsynonymous editing sites during fly holometabolous development. (*E*) Heatmap representation of Pearson's correlations (*r*) between editing levels of individual nonsynonymous editing sites and different levels of lagging (no lagging and one-, two-, and three-stage lagging) in dADAR expression in the matching order of (*D*).

located in TEs and commonly present during fly holometabolous development from embryo to adult. Of the 22 nonsynonymous events, 16 were conserved editing sites, implying their importance of A-to-I editing for fly development. As shown in figure 6*D*, we further found different patterns in changes of RNA editing level on separate A-to-I editing sites during fly holometabolous development. We observed that the dynamics and active nature A-to-I editing on separate sites may reflect different levels of lagging as compared with the fluctuation of dADAR expression, even though editing sites located within the same gene loci (fig. 6*E* and supplementary fig. S9, Supplementary Material online). For example, NaCP60E (Na channel protein 60E) mediates the voltage-dependent sodium ion permeability of excitable membranes (Kulkarni et al. 2002), for which A-to-I editing may play a role in rapid electrical and chemical neurotransmission (Hoopengardner et al. 2003). Changes in editing levels for NaCP60E transcripts at the six nonsynonymous A-to-I editing sites exhibited distinct patterns throughout fruit fly life cycles, and were correlated with different levels of lagging in dADAR expression (fig. 6*E* and supplementary fig. S9, Supplementary Material online). Such diversity of NaCP60E proteins through A-to-I editing may contribute to fly nervous system. This is the first reported finding on the editing pattern of NaCP60E transcripts. These results reveal that A-to-I editing, which occurs at the transcriptome level, also provides a wide range of diversity for the proteome, and that individual editing sites may be developmentally regulated.

Discussion

In this study, we successfully identified high-confidence editing sites across diverse species by controlling for FDR and %AG (which is determined by $N_{cluster}$; the clustering strategy) or evolutionary conservation of editing in multiple species (the conservation strategy), without a priori knowledge of known SNPs or genome sequencing from the same sample. This enables us to identify high-confidence A-to-I sites in both model and nonmodel animals, and thus to construct the first A-to-I editomes across 20 metazoan species. Of note, previous studies have reported that the A/G pattern of human-chimpanzee coincident SNPs (the same *A/G* variants were present at human–chimpanzee orthologous positions) was over-represented (Hodgkinson et al. 2009; Chen et al. 2016). We suggest that the effect of coincident SNPs on the editing sites identified by the conservation strategy is slight because only 0.01% (5 out of 38,480) of the human–chimpanzee conserved A-to-G editing sites are also observed to be *A/G* polymorphic at human–chimpanzee orthologous positions on the basis of dbSNP (supplementary table S5, Supplementary Material online). Comparative analysis of the identified editing sites revealed that highly clustered and conserved editing sites tended to have a higher editing level and a higher magnitude of the ADAR motif (fig. 2*H*). We also found that both the ratio of the frequencies of nonsynonymous editing to that of synonymous editing (the f_n-to-f_s ratio; fig. 4*C*) and the conservation level of single nucleotides (the PhyloP score; fig. 4*D*) remarkably increased with increasing the conservation level of A-to-I editing. These results thus suggest potentially functional benefit of highly clustered and conserved editing sites.

While the scaffolding of A-to-I editomes could be raised by the introduction of TEs in a lineage-specific manner, a subset of highly conserved A-to-I editing events (87 human editing events observed in at least 5 nonhuman species, of which 34 events were conserved in nonprimate species; supplementary data set S4, Supplementary Material online) might exist long in the common ancestor of primates (or even the common ancestor of mammalian species for the 34 events). For example, *Gabra3* contains a highly conserved A-to-I editing site, at which amniotes encode an isoleucine codon (AUA) while zebrafish and frog retain that of methionine (AUG) as the ancestral state (Tian et al. 2011). This codon in amniotes can still revert to "AUG" by A-to-I editing during transcription, making the switch from isoleucine to methionine (I/M) possible without losing its original message in genome. In addition, we found that 18 of the 87 events (located in 15 genes) were nonsynonymous (supplementary data set S4,Supplementary Material online). Alteration of A-to-I editing has been shown to affect the function of these 15 genes in diverse species (supplementary table S3, Supplementary Material online). The gene ontology (GO) analysis further revealed that the host genes of these highly conserved nonsynonymous editing sites were enriched in "synaptic transmission," "synapse," "channel activity," "gated channel activity," and "ion channel activity" (supplementary table S6, Supplementary Material online), consistent with the functional effect of A-to-I editing in previous reports (supplementary table S3, Supplementary Material online). Of note, the host genes of *Drosophila* conserved editing sites (i.e., *D. melanogaster* editing sites observed in at least one other *Drosophila* species; supplementary data set S2, Supplementary Material online) were also enriched in similar GO terms (supplementary table S6, Supplementary Material online), indicating the evolutionary convergence of editing targets despite substantial divergence between vertebrate and invertebrate. These results also reflect that a conserved enrichment of editing in the CNS throughout more than 300 Myr of divergent evolution in complex animals from primates to chickens (fig. 5*D* and supplementary figs. S3 and S4, Supplementary Material online). Intriguingly, of the 66,800 human A-to-I editing events identified by our conservation strategy (table 1 and supplementary data set S4, Supplementary Material online), 6, 4, and 192 could cause stop codon losses, splice site alterations, and amino acid changes (i.e., nonsynonymous changes), respectively. It is worthwhile to further investigate the functional importance of these evolutionarily conserved editing events.

For the spatial context of A-to-I editing, we found that editing events tended to exhibit tissue-specificity (particularly for the CNS) (fig. 5*A* and *B*), and that global editing activity and ADAR (ADAR1 and ADAR2) expression were higher in the CNS than in the other tissues examined and positively correlated with each other (fig. 5*C* and *D* and supplementary fig. S3, Supplementary Material online). Importantly, these tendencies were consistently observed among amniotes from primates to birds, representing the first documented spatial profiling for A-to-I editing across amniotes. Using PCA, we further showed that global editing activity of primate-conserved sites exhibited an organ-preferential clustering (fig. 5*E*), reflecting that these editing events tended to be specific to a few organs (fig. 5*A* and *B*). On the other hand, our results exhibited the dynamic and active nature of A-to-I editome during the development of invertebrates (worm and fly) and during the embryogenesis of vertebrates (zebrafish and frog), respectively. We observed that shifting patterns of global A-to-I editing were precisely dependent on *ADAR* expression (fig. 6*A* and *B*), echoing the developmental effects of *ADAR* enzymes (Palladino et al. 2000; Jepson and Reenan 2008; Horsch et al. 2011). Particularly, this (fig. 6*D*) and other studies also observed that individual editing sites can exhibit distinct editing patterns during development in diverse species (Gurevich et al. 2002; Osenberg et al. 2010; Yu et al. 2016). In addition to a possible explanation that shift in expression levels of edited genes may accompany editing changes during development (Yu et al. 2016), we provided another possibility that individual editing sites may reflect different levels of lagging as compared with the fluctuation of ADAR expression (e.g., fig. 6*D* and *E*). The joint effects of both possibilities may be the cause for diverse editing patterns of individual sites during development. Further investigation of individual targeted editing sites will provide a fuller understanding of the developmental role of A-to-I RNA editing. Moreover, we observed that the editing patterns of *C. elegans* and fly were generally comparable with each other during development (fig. 6*A*). The comparability of editing pattern was also observed between vertebrates (zebrafish and frog) during embryogenesis (fig. 6*B*), regardless of where the editing sites were located (fig. 6*C* and supplementary fig. S7, Supplementary Material online). These results thus indicate that global editing activity principally reflects ADAR expression inherited from their common ancestors.

In summary, our comparative analysis provides the first A-to-I editomes across 20 diverse species, from worm to human, and reconstructs an evolutionary landscape of editomes, which might serve as a valuable resource for understanding the co-evolution of the editing machinery (i.e., TEs, editomes, and editors). The spatiotemporal atlas provides unprecedented resolution of editing dynamics, and reveals conserved patterns of editing, which might be a key regulatory mechanism in living cells, especially in the CNS and during early development. This study thus sheds light on evolutionary and dynamic aspects of A-to-I editome across vertebrates and invertebrates, opening up this important but understudied class of nongenomically encoded events for comprehensive characterization.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Chan-Shuo Wu, Yen-Zho Chen, and Yu-Ting Hsiao for computational assistance, Prof. Hsien-Sung Huang for providing mouse brain tissue, Prof. Fu Huang for providing *D. melanogaster* sample, and Prof. Ying-Chih Chang for providing bench and related experimental facilities. This study was supported by Genomics Research Center, Academia Sinica, Taiwan; and the Ministry of Science and Technology, Taiwan (under the contracts NSC 102-2621-B-001-003 and MOST 103-2628-B-001-001-MY4).

Literature cited

Alon S, et al. 2015. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. Elife 4:e05198.
Bahn JH, et al. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. 22(1):142–150.
Bass B, et al. 2012. The difficult calls in RNA editing. Interviewed by H Craig Mak. Nat Biotechnol. 30(12):1207–1209.
Bass BL. 2002. RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem. 71:817–846.
Bazak L, et al. 2014. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. 24:365–376.
Bazak L, Levanon EY, Eisenberg E. 2014. Genome-wide analysis of Alu editability. Nucleic Acids Res. 42(11):6876–6884.
Burns CM, et al. 1997. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature 387(6630):303–308.
Chen CY, Hung LY, Wu CS, Chuang TJ. 2016. Purifying selection shapes the coincident SNP distribution of primate coding sequences. Sci Rep. 6:27272.
Chen JY, et al. 2014. RNA editome in rhesus macaque shaped by purifying selection. PLoS Genet. 10(4):e1004274.
Chen L. 2013. Characterization and comparison of human nuclear and cytosolic editomes. Proc Natl Acad Sci U S A. 110(29):E2741–E2747.
Danecek P, et al. 2012. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol. 13(4):26.
Eggington JM, Greene T, Bass BL. 2011. Predicting sites of ADAR editing in double-stranded RNA. Nat Commun. 2:319.
Feng Zhang YL, Yan S, Xing Q, Tian W. 2017. SPRINT: a SNP-free toolkit for identifying RNA editing sites. Bioinformatics 33:3538–3548.
Gurevich I, et al. 2002. Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. Neuron 34(3):349–356.
Han L, et al. 2015. The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. Cancer Cell 28(4):515–528.

Hodgkinson A, Ladoukakis E, Eyre-Walker A. 2009. Cryptic variation in the human mutation rate. PLoS Biol. 7(2):e1000027.
Hoopengardner B, Bhalla T, Staber C, Reenan R. 2003. Nervous system targets of RNA editing identified by comparative genomics. Science 301(5634):832–836.
Horsch M, et al. 2011. Requirement of the RNA-editing enzyme ADAR2 for normal physiology in mice. J Biol Chem. 286(21):18614–18622.
Hwang T, et al. 2016. Dynamic regulation of RNA editing in human brain development and disease. Nat Neurosci. 19(8):1093–1099.
Jepson JE, Reenan RA. 2008. RNA editing in regulating gene expression in the brain. Biochim Biophys Acta 1779(8):459–470.
John D, Weirick T, Dimmeler S, Uchida S. 2017. RNAEditor: easy detection of RNA editing events and the introduction of editing islands. Brief Bioinform. 18:993–1001.
Kawahara Y, et al. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 315(5815):1137–1140.
Kim DD, et al. 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. 14(9):1719–1725.
Kimelman D, Kirschner MW. 1989. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth-factor in *Xenopus* oocytes. Cell 59(4):687–696.
Kiran A, Baranov PV. 2010. DARNED: a DAtabase of RNa EDiting in humans. Bioinformatics 26(14):1772–1776.
Kleinman CL, Majewski J. 2012. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science 335(6074):1302; author reply 1302.
Kulkarni NH, Yamamoto AH, Robinson KO, Mackay TF, Anholt RR. 2002. The DSC1 channel, encoded by the smi60E locus, contributes to odor-guided behavior in *Drosophila melanogaster*. Genetics 161(4):1507–1516.
Lehmann KA, Bass BL. 2000. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39(42):12875–12884.
Lev-Maor G, et al. 2007. RNA-editing-mediated exon evolution. Genome Biol. 8(2):R29.
Levanon EY, et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol. 22(8):1001–1005.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18(11):1851–1858.
Lin W, Piskol R, Tan MH, Li JB. 2012. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science 335(6074):1302; author reply 1302.
Liscovitch-Brauer N, et al. 2017. Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell 169(2):191–202 e111.
Maas S, Kawahara Y, Tamburro KM, Nishikura K. 2006. A-to-I RNA editing and human disease. RNA Biol. 3(1):1–9.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628.
Nakamura K, et al. 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 39(13):e90.
Nishikura K. 2006. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol. 7(12):919–931.
Osenberg S, et al. 2010. Alu sequences in undifferentiated human embryonic stem cells display high levels of A-to-I RNA editing. PLoS ONE 5(6):e11173.
Palladino MJ, Keegan LP, O'Connell MA, Reenan RA. 2000. A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. Cell 102(4):437–449.
Paz-Yaacov N, et al. 2010. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc Natl Acad Sci U S A. 107(27):12174–12179.
Peng Z, et al. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol. 30(3):253–260.
Picardi E, D'Erchia AM, Lo Giudice C, Pesole G. 2017. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. Nucleic Acids Res. 45(D1):D750–D757.
Picardi E, et al. 2015. Profiling RNA editing in human tissues: towards the inosinome Atlas. Sci Rep. 5:14941.
Pickrell JK, Gilad Y, Pritchard JK. 2012. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science 335(6074):1302; author reply 1302.
Pinto Y, Cohen HY, Levanon EY. 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. Genome Biol. 15(1):R5.
Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of non-neutral substitution rates on mammalian phylogenies. Genome Res. 20:110–121.
Porath HT, Carmi S, Levanon EY. 2014. A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat Commun. 5:4726.
Ramaswami G, et al. 2012. Accurate identification of human Alu and non-Alu RNA editing sites. Nat Methods 9(6):579–581.
Ramaswami G, et al. 2013. Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 10(2):128–132.
Ramaswami G, Li JB. 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. 42(Database issue):D109–D113.
Rueter SM, Dawson TR, Emeson RB. 1999. Regulation of alternative splicing by RNA editing. Nature 399(6731):75–80.
Sapiro AL, Deng P, Zhang R, Li JB. 2015. Cis regulatory effects on A-to-I RNA editing in related *Drosophila* species. Cell Rep. 11(5):697–703.
Savva YA, Reenan RA. 2014. Identification of evolutionarily meaningful information within the mammalian RNA editing landscape. Genome Biol. 15(1):103.
Schroeder A, et al. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 7:3.
Shtrichman R, et al. 2012. Altered A-to-I RNA editing in human embryogenesis. PLoS ONE 7(7):e41576.
Stefl R, et al. 2010. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell 143(2):225–237.
Sun Y, et al. 2016. RED: a Java-MySQL software for identifying and visualizing RNA editing sites using rule-based and statistical filters. PLoS ONE 11(3):e0150465.
Tian N, et al. 2011. A structural determinant required for RNA editing. Nucleic Acids Res. 39(13):5669–5681.
Tomaselli S, et al. 2015. Modulation of microRNA editing, expression and processing by ADAR2 deaminase in glioblastoma. Genome Biol. 16:5.
Tonkin LA, et al. 2002. RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. Embo J. 21(22):6025–6035.
Ulbricht RJ, Emeson RB. 2014. One hundred million adenosine-to-inosine RNA editing sites: hearing through the noise. Bioessays 36(8):730–735.
Veno MT, et al. 2012. Spatio-temporal regulation of ADAR editing during development in porcine neural tissues. RNA Biol. 9:1054–1065.

Wagner RW, Smith JE, Cooperman BS, Nishikura K. 1989. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. Proc Natl Acad Sci U S A. 86(8):2647–2651.
Wahlstedt H, Daniel C, Enstero M, Ohman M. 2009. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res. 19(6):978–986.
Wang Q, Khillan J, Gadue P, Nishikura K. 2000. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. Science 290(5497):1765–1768.
Xu G, Zhang J. 2014. Human coding RNA editing is generally nonadaptive. Proc Natl Acad Sci U S A. 111(10):3769–3774.
Xu G, Zhang J. 2015. In search of beneficial coding RNA editing. Mol Biol Evol. 32(2):536–541.
Yu Y, et al. 2016. The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. PLoS Genet. 12(7):e1006191.
Zaranek AW, Levanon EY, Zecharia T, Clegg T, Church GM. 2010. A survey of genomic traces reveals a common sequencing error, RNA editing, and DNA editing. PLoS Genet. 6(5):e1000954.
Zhang Q, Xiao X. 2015. Genome sequence-independent identification of RNA editing sites. Nat Methods 12(4):347–350.

Associate editor: Wen-Hsiung Li