



REDIdb 3.0: A Comprehensive Collection of RNA Editing Events in Plant Organellar Genomes

Claudio Lo Giudice¹, Graziano Pesole^{1,2} and Ernesto Picardi^{1,2*}

¹ Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Consiglio Nazionale delle Ricerche, Bari, Italy, ² Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari A. Moro, Bari, Italy

RNA editing is an important epigenetic mechanism by which genome-encoded transcripts are modified by substitutions, insertions and/or deletions. It was first discovered in kinetoplastid protozoa followed by its reporting in a wide range of organisms. In plants, RNA editing occurs mostly by cytidine (C) to uridine (U) conversion in translated regions of organelle mRNAs and tends to modify affected codons restoring evolutionary conserved aminoacid residues. RNA editing has also been described in non-protein coding regions such as group II introns and structural RNAs. Despite its impact on organellar transcriptome and proteome complexity, current primary databases still do not provide a specific field for RNA editing events. To overcome these limitations, we developed REDIdb a specialized database for RNA editing modifications in plant organelles. Hereafter we describe its third release containing more than 26,000 events in a completely novel web interface to accommodate RNA editing in its genomics, biological and evolutionary context through whole genome maps and multiple sequence alignments. REDIdb is freely available at http://srv00.recas.ba.infn.it/redidb/index.html

OPEN ACCESS

Edited by:

Giovanni Nigita, The Ohio State University, United States

Reviewed by:

Giorgio Giurato, Università degli Studi di Salerno, Italy Shihao Shen, University of California, Los Angeles, United States

> *Correspondence: Ernesto Picardi

ernesto.picardi@uniba.it

Specialty section:

This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Plant Science

> Received: 01 February 2018 Accepted: 29 March 2018 Published: 11 April 2018

Citation:

Lo Giudice C, Pesole G and Picardi E (2018) REDIdb 3.0: A Comprehensive Collection of RNA Editing Events in Plant Organellar Genomes. Front. Plant Sci. 9:482. doi: 10.3389/fpls.2018.00482 Keywords: organellar genomes, RNA editing, plant database, mitochondria, chloroplasts

INTRODUCTION

RNA editing is an essential co/post transcriptional process able to expand transcriptome and proteome diversity in addition to alternative splicing. The term RNA editing was first introduced in 1986 to describe the addition and deletion of uridine nucleotides to and from mRNAs in trypanosome mitochondria (Benne et al., 1986). Since then, RNA editing events have been found in a wide range of organisms and can occur in the nucleus and cytoplasm as well as in organelles (Bowe and depamphilis, 1996). Modifications due to RNA editing comprise nucleotide substitutions and insertions or deletions that can affect both protein coding and Non-protein coding RNAs (Maier et al., 1996; Steinhauser et al., 1999).

In humans, the most prevalent type of RNA editing event is the deamination of adenosine (A) in inosine (I) in double RNA strands (dsRNAs) through the catalytic activity of the adenosine deaminase (ADAR) family of enzymes. To date, more than 4 million events have been collected and annotated in dedicated resources such as DARNED, RADAR, and REDIportal (Kiran et al., 2013; Ramaswami and Li, 2014; Picardi et al., 2017).

1

In plants, RNA editing occurs mostly in organelles in the form of cytidine (C) to uridine (U) conversion particularly in translated regions of mRNAs, albeit the opposite event (U-to-C substitutions) has been observed in some taxa, especially in chloroplasts RNAs (Takenaka et al., 2013). Plant RNA editing sites are recognized by specific pentatricopeptide repeat (PPR) proteins that are encoded in the nuclear genome. In flowering plants, the editosome machinery requires several additional Non-PPR protein factors, even though its molecular assembly has yet to be clarified (Sun et al., 2016).

Most of the C-to-U changes in the protein coding regions tends to modify affected codons restoring evolutionary conserved aminoacid residues (Gray, 2003). Therefore, plant RNA editing is believed to act as an additional proofreading mechanism to generate fully functional proteins. Occasionally, C-to-U modifications occur in untranslated regions, structural RNAs and intervening sequencing, affecting splicing and translation efficiency. Indeed, RNA editing changes in the domain V of plant group II introns is mandatory for the splicing process (Castandet et al., 2010).

With the advent of high-throughput sequencing technologies, many complete plant organellar genomes have been released and numerous novel RNA editing events uncovered. Nevertheless, RNA editing changes are not always correctly or completely annotated in primary databases (GenBank, ENA and DDBJ) and an appropriate field to unambiguously describe them is not provided. RNA editing modifications are often reported as misc_feature or even as simple exception notes. With the aim to overcome these limitations and create a cured catalog of plant RNA editing events, we developed the specialized REDIdb database. Its first release stored 9,964 modifications distributed over 706 different nucleotide sequences, increased to 11,897 in the following update.

After 10 years of massively parallel sequencing, we present here REDIdb 3.0, an upgraded release that annotates 26,618 RNA editing events distributed among 281 organisms and 85 complete organellar genomes.

All changes have been recovered from Genbank and literature using a semi-automated bioinformatics procedure in which each annotation has been manually checked to avoid redundancy or inconsistencies due to errors in flatfiles.

The web-interface was totally restyled and developed using the latest computational technologies in the field of database querying and managing.

Furthermore, many computational facilities have been integrated to improve the user experience and ensure continuous and future updates of the database. Indeed, REDIdb 3.0 accommodates RNA editing in its genomics, biological and evolutionary context through whole genome maps and multiple sequence alignments.

Although a variety of RNA editing databases have been released such as DARNED (Kiran et al., 2013), RADAR (Ramaswami and Li, 2014), and REDIportal (Picardi et al., 2017), REDIdb is the only one devoted to editing changes in plant organelles. Indeed, similar resources such as dbRES (He et al., 2007), RESOPS (Yura et al., 2009), ChloroplastDB (Cui et al.,

2006), or GOBASE (O'Brien et al., 2009) have been dismissed or not updated.

MATERIALS AND METHODS

All editing events stored in REDIdb derive from GenBank flatfiles through a semi-automated parsing algorithm implemented in custom python (2.7.13) scripts. Each flatfile is screened for RNA editing features using the SeqIO parser included in the Biopython (1.68) module (Cock et al., 2009).

All annotations have been manually checked to identify and correct potential errors, taking into account other related flatfile fields or literature. REDIdb database is organized in MySQL tables and queries are in python employing the MySQL-python (1.2.5) module, a data access library to MySQL engine. The web interface, instead, is built in BootStrap (3.3.7), while data presentation is based on DataTables, an *ad hoc* Javascript library (1.10.13) to efficiently show large tables in html documents. Genome rendering, available for complete organellar genomes, has been developed in pure python, mimicking OGDraw graphics (Lohse et al., 2013).

Query results are dynamically generated using the CGI (common gateway interface) technology. Multiple sequence alignments of edited cDNAs and proteins have been generated by ClustalOmega (Sievers et al., 2011) and displayed in html pages through the MSAViewer (Yachdav et al., 2016), a JavaScript component of the BioJS collection (https://biojs.net/).

The distribution of RNA editing events along functional domains and predicted protein secondary structures are shown by the feature-viewer JavaScript library (https://github.com/ calipho-sib/feature-viewer) based on the powerful D3 JavaScript library for visualizing data using web standards (https://d3js. org/). Functional domains have been detected using InterPro engine (Jones et al., 2014), while protein secondary structures have been predicted using the stand-alone version of Spider2 program (Yang et al., 2017).

All the scripts to parse multiple alignments, InterPro html files and Spider2 outputs have been created in Python. Scripts used to extract RNA editing positions from Genbank flatfiles are freely available at the REDIdb help page. Additional details and supplementary scripts are available upon request.

RESULTS

Database Content

Previous REDIdb release contained 11,897 editing events distributed over 198 organisms and 929 different nucleotide sequences. This upgraded version, instead, collects more than 26,000 editing events from 281 organisms, 85 complete organellar genomes and 3,467 sequences. REDIdb 3.0 includes 26,545 events in protein coding sequences and 73 in untranslated regions, structural RNAs and introns. The vast majority of editing changes occur in the mitochondrion, accounting for a total of 23,553 events over 2,300 sequences.

The most recurrent RNA editing modification is the C-to-U substitution, that accounts for more than 92% of all

TABLE 1 | Number of RNA Editing events in complete genomes stored in REDIdb.

Location	Protein_ coding_events	No_protein_ coding_events	Genbank_id	Organism_name	Cultivar	Strain
mito	6		NC_021931	Anomodon attenuatus		
mito	6		NC_016121	Anomodon rugelii		
mito	493		NC_001284	Arabidopsis thaliana		
mito	2		NC_024520	Atrichum angustatum		
mito	1		NC_024519	Bartramia pomiformis		
mito	344		NC_015994	Beta macrocarpa		
mito	340		NC_015099	<i>Beta vulgaris</i> subsp. maritima		
mito	4		NC_031212	Brachythecium rivulare		
mito	417		AP006444	Brassica napus	Westar	
mito	467		KJ820683	Brassica oleracea var. botrytis		
mito		5	KJ820683	Brassica oleracea var. botrytis		
mito	2		NC_024518	Buxbaumia aphylla		
mito	2		NC_012116	Carica papaya	SunUp	
mito	547		NC_014043	Citrullus lanatus	Florida	
					giant	
mito	3		NC_024515	Climacium americanum		
mito	475		NC_014050	Cucurbita pepo	Dark green zucchini	
mito	719		NC_027976	Ginkgo biloba		
mito		8	AY182006	Harpochytrium sp. JEL105		JEL105
mito		5	AY182005	Harpochytrium sp. JEL94		JEL94
mito	1		AP014526	Hevea brasiliensis	BPM 24	
mito	1		AP017300	<i>Hordeum vulgare</i> subsp. spontaneum		H602
mito	1		AP017301	Hordeum vulgare subsp. vulgare		HarunaNijo
mito	3		NC_026515	Hyoscyamus niger		
mito	5		NC_024516	Hypnum imponens		
mito	888		KC821969	Liriodendron tulipifera		
mito	488		NC_016743	Lotus japonicus		
mito		1	NC_016743	Lotus japonicus		
mito	488		JN872551	Lotus japonicus		MG-20
mito		1	JN872551	Lotus japonicus		MG-20
mito	485		NC_016742	Millettia pinnata		
mito		1	NC_016742	Millettia pinnata		
mito		10	AY182007	Monoblepharella sp. JEL15		JEL15
mito	847		NC_030753	Nelumbo nucifera		
mito	3		NC_029805	Nicotiana sylvestris	TW 137	
mito	1		NC_006581	Nicotiana tabacum	Bright Yellow 4	
mito	60		NC_012651	Nothoceros aenigmaticus		
mito		1	NC_012651	Nothoceros aenigmaticus		
mito	1		NC_029356	Orthotrichum diaphanum		
mito	1		NC_029355	Orthotrichum macrocephalum		
mito	2		NC_024522	Orthotrichum stellatum		
mito	3		AP017386	Oryza sativa Indica Group		BT-CMS
mito	97		NC_013765	Phaeoceros laevis		
mito	16		NC_017755	Phlegmariurus squarrosus		
mito	8		NC_013444	Pleurozia purpurea		
mito	8		NC_024514	Ptychomnion cygnisetum		
mito	60		JQ083668	Raphanus sativus		

(Continued)

TABLE 1 | Continued

Location	Protein_ coding_events	No_protein_ coding_events	Genbank_id	Organism_name	Cultivar	Strain
mito	7		NC_027974	Sanionia uncinata		
mito	326		NC_014487	Silene latifolia		
mito	3		NC_024521	Sphagnum palustre		
mito	1		NC_017840	Spirodela polyrhiza		7498
mito	1		NC_027515	Syntrichia filaris		
mito	5		KC784953	Tetraphis pellucida		
mito	2		NC_028191	Tetraplodon fuegianus		
mito	1		NC_016122	Treubia lacunosa		
mito	2		NC_024517	Ulota hutchinsiae		
mito	7		NC_015121	Vigna radiata		
mito	109		AY506529	Zea mays		NB
chloro	79		KU764518	Actinostachys pennula		
chloro	343		AY178864	Adiantum capillus-veneris		
chloro		1	AY178864	Adiantum capillus-veneris		
chloro	564		NC_004543	Anthoceros formosae		
chloro	56		NC_019628	Apopellia endiviifolia		
chloro	25		LC154068	Arabidopsis lyrata subsp. lyrata		MN47
chloro	37		NC_000932	Arabidopsis thaliana		
chloro	36		NC_004561	Atropa belladonna		Ab5p(kan)
chloro	1		NC_031894	Citrus depressa		
chloro	1		NC_009618	Cycas taitungensis		
chloro	65		NC_028542	Cyrtomium devexiscapulae		
chloro	55		NC_028705	Cyrtomium falcatum		
chloro	2		NC_031159	lpomoea nil	Tokyo- kokei standard	
chloro	1		NC 016058	Larix decidua		
chloro	3		NC 017006	Mankvua cheiuensis		
chloro	38		NC 016708	Millettia pinnata		
chloro	00	1	NC 016708	Millettia pinnata		
chloro	58	·	NC 014592	Myriopteris lindheimeri		
chloro	40		NC_001879	Nicotiana tabacum	Bright Yellow 4	
chloro	5		NC_007602	Nicotiana tomentosiformis		
chloro	44		AY916449	Phalaenopsis aphrodite subsp. formosana	Taisugar TS-97	
chloro	1		NC_005087	Physcomitrella patens subsp. patens		
chloro	23		NC_001631	Pinus thunbergii		
chloro	66		NC_014348	<i>Pteridium aquilinum</i> subsp. aquilinum		
chloro	16		NC_005878	Saccharum hybrid cultivar SP80-3280		
chloro	1		NC_020098	Tectona grandis		
chloro	72		NC_028543	Woodwardia unigemmata		
chloro	32		NC_001666	Zea mays		

Events are divided by sequence (coding/Non-coding) and according to their intracellular location. In presence of multiple accession numbers for the same organism, only the RefSeq record (if present in Genbank) has been considered.

annotated events and, when located in protein coding regions, tends to modify the aminoacid coded by the edited codon. Indeed, the majority of RNA editing events affects the first and second codon position leading to aminoacid changes resulting the most conserved in the comparison with related orthologs.



Differently from the previous releases, the novel REDIdb database annotates 85 complete organellar genomes. Of these 57 are mitochondrial genomes and include 7791 events. As reported in **Table 1**, the most edited mitochondrial genomes are those from *Liriodendron tulipifera*, *Nelumbo nucifera* and *Ginkgo biloba* with 888, 847, and 717 events, respectively. Of 27 annotated chloroplast genomes, instead, the one from *Anthoceros formosae* comprising 564 modifications results the richest in editing events.

All REDIdb sequences including RNA editing events are identified by unique accession numbers (e.g., EDI0000.). To preserve the full compatibility with previous database versions, accession numbers linked to old entries have been maintained unchanged.

Query Form and Output Tables

REDIdb implements a modular query form (**Figure 1A**) allowing users to make flexible searches by selecting the organism or the intracellular location or the gene name. Regarding nucleotide sequences, users can retrieve the original sequence submitted to the primary database or the RefSeq version or both. In addition, the search can be limited to full open reading frames and include individual exons in case of interrupted genes.

Query results are shown in a sortable and exportable summary table (**Figure 1B**) comprising several info such as the GenBank accession number, the organism and the link to the related taxonomy, the organelle type and the link to the complete genome (if available), the gene name and a flag indicating its partial or full nature, the editing types and details and the total number of events. Column can be selectively included in the final table and results are downloadable in pdf or csv format. The "Taxonomy" column includes a link to an interactive taxonomy chart, while the "Genome" column contains a link to the complete genome (if available in primary databases) chart in which RNA editing events are displayed in their genomics context.

Using the link in the "Gene_name" column, users can browse individual RNA editing events organized in flatfiles.

Entry Organization

RNA editing events stored in REDIdb are organized in specific flat-files comprising four main sections. The first section (**Figure 2A**) contains a general description of the entry including the organism name, the taxonomy (according with the NCBI Taxonomy database), the GenBank and PubMed accession numbers, the intracellular location (mitochondrion or chloroplast) and the official gene name.

The second section (**Figure 2B**) is devoted to Gene Ontologies (GO), obtained by matching each protein sequence contained in REDIdb against the InterPro database (Finn et al., 2017). In the case of protein coding genes, it contains information regarding the molecular functions, the biological processes and the cellular localization of the protein product. The third section (**Figure 2C**) shows all the editing features that characterize the record. Here, for each editing event the position on the transcript is reported and, if the complete reference genome is available, also the genomic location. In case of editing within protein coding genes, the genomic codon, edited codon and aminoacidic change are determined and reported. Finally, the fourth

Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Control Contro Contro Contro Contr	ORGANISM Nelumbo nucifera DIVISION Eukarvota Viridiplantae	Streptophyta Embryophyta Tr	acheophyta, Spermatophyta, Mag	noliophyta eudicotyledose	Proteales, Nelumbonaceae	Nelumbo	
A A	LOCATION mitochondrion	on optopriyta, embryopriyta, fr	aonoophyta, opennatophyta, Mag	nonopriyta, eudicotyled005,	rotoaico, neiumbonaceae,	reambo	
Control of the second of	STATUS complete genome						
Concerning Concern	SECHENCE Full						A
Transmission The form of t							
Number delighted by a possible delightedlightedlightedlighted by a possible delighted by a possible del	SOURCE gene						
Submary biology Submary bi							
Columbic billion Columbic billion<	GENBANK NC_030753						
Of the control of control or cutable address is a control of control or control	PUBMED 27444405						
<text></text>	GO Gene Ontology						
Control Automatical basis Aut			-				
BUILDER AMPLICASS GUIDER Developmentered CENTRAL FAMILIER Sevenic position CDNA position Generalic codin Effect codin Generalic codin Effect codin Sevenic position Effect codin Tenneric codin Effect codin Tenneric codin A Effect codin A Effect codin A<	MOLECULAR FUNCTION GO:000	4129 cytochrome-c oxidase act	wity				≻ B
Circle Control (Control (Contro) (Contro) (Control (Control (Control (Control (Control (Control (BIOLOGICAL PROCESS GO:0022	900 electron transport chain					
Conside position COMA position Genomic coden Edited coden Genomic AA Edited AA 100716 161 Ca Ta S L 100716 161 Ca Ta S L 100716 161 Ca Ta S L 100716 273 CdQ CT L L 100716 273 CdQ CT L L 100727 373 CdQ TT I I 100728 354 atC atT I I 10089 443 cdQ atT P L 10089 443 cdQ Ta P UL 10089 544 CC Ta P S 10089 544 CC Ta P S 10089 544 CC Ta P S 10089 544 CC Ta P S <	CELLULAR COMPONENT GO:001	6020 membrane					
Genomic position Oth A position Genomic coden Edited coden Genomic AA Edited AA 1 105713 1 61 1 Ca 1 fr S L L 1 105713 1 63 0 gg 7 gg R W W 1 105713 1 63 0 gg 7 gg R W W 1 10588 2 72 0 Gg 1 fr L L L 1 10588 2 73 0 Gg 1 fr L L L 1 10582 3 3 GG atr 1 1 1 L L L 1 10589 43 0 Gg 1 fr P L	EDITING FEATURES						
OUTPOTED International and the second se	Genomic position	cDNA position	Genomic codon	Edited codon	Genomic AA	Edited AA	
100010 100 000 <t< td=""><td>105715</td><td>161</td><td>tCa</td><td>tTa</td><td>S</td><td>L</td><td></td></t<>	105715	161	tCa	tTa	S	L	
100000 127 cb cb r	105713	163	Caa	Taa	B	W/	
100000 273 CC C1 P L 100000 33 atC atT 1 1 100000 33 atC atT 1 1 100000 379 Cgg Tgg R W 100000 379 Cgg Tgg R W 100000 443 aCg aTg T M 100000 443 aCg aTg P L 100000 443 CG CG TG P L 100000 557 CG TG F L L 100000 652 CG TG F L L 100001 652 CG TG F S L 100001 70 TG CG S F	105940	100	099	ryy ctT	1	1	
100000 2/3 Gog Gig F L 100000 33 adC afT I I 100000 33 adC afT I I 100000 33 adC afT I I 100000 443 adg adg afg T M 100000 443 adg afg T M L 100000 443 Ca Ca Ta S L L 100000 557 cb Ca Ta S L L L L L L L L L L L L L L L L L L L </td <td>105049</td> <td>2/</td> <td>00</td> <td>oTo</td> <td>L.</td> <td></td> <td>(</td>	105049	2/	00	oTo	L.		(
the set of the se	105390	210	200	cig		L	
100042 379 Cg0 Tg0 R W 100083 379 Cg0 Tg0 R W 100083 379 Cg0 Tg0 R W L 100083 379 Cg0 Tg0 F L M 100082 460 CcaCCa Ta/TTa P L L 100081 461 CcaCCa Ta/TTa P L L 100086 470 Ga Ta S L L 100086 470 Ga Ta S L L 100086 470 Ga Ta S L L 100086 670 Cd Ta S L L 100086 670 Cd Ta S L<	105843	33	alC	dll			í l
100000 33 COUNT 144 P V 1000000 443 acg afg T M 1000000 443 acg afg T M 1000000 443 acg afg T M 1000000 4461 CCarCCa TeaTTa P LL C 1000000 4461 CCarCCa TeaTTa P LL C 1000000 4461 CCarCCa TeaT P L C 1000000 561 Itaa Ita S L 1000000 562 CC CT P L C 1000000 562 CC TT P L C 1000000 652 CC TT N C C 1000000 71 CO Tea Ced S F 1000000 71 CC Tea Ced S F 1000000 71 CC Tea Ced S F	105322	304	alu	Tag		W/	
100000 30 CA CH F L 100000 30 CA CH F L 100000 443 acg afg T M 100000 443 CaCaCCa TaTTa P SL 100000 443 tba tfa S L 100000 557 tba tfa S L 100000 652 tba tfa S L 100446 676 Cgt tfg R C 100447 695 tbt <tdttfit< td=""> Tft S F 100447 695 tbt<tdttt<tdttt< td=""> Tft S F 1000071 700 Tdt Cdt Tft P S 100005 771 Cdt Tft S F S <!--</td--><td>105497</td><td>3/3</td><td>cgg</td><td>rgg cTt</td><td>P</td><td>v.</td><td>r I</td></tdttt<tdttt<></tdttfit<>	105497	3/3	cgg	rgg cTt	P	v.	r I
100000 1 400 CaCCa TaTTa P SL 100000 461 CCaCCa cTaTTa P UL C 100000 461 CCaCCa cTaTTa P UL C 100000 461 CC Ta S L 100000 557 CC C1 T P L 100000 557 CC C1 T P L 100000 557 CC C1 Ta S L 100000 552 C0 Ta S L L 100000 652 C0 Ta S F L 100001 652 C0 Ta C S P 100001 700 Ta Cd T M M 100001 700 Ta Cd T M M M 100002 711 Cd Ta P S M M M 1000002 711 Cd <td< td=""><td>102600</td><td>38</td><td>207</td><td>aTa</td><td>T</td><td>M</td><td></td></td<>	102600	38	207	aTa	T	M	
1000000000000000000000000000000000000	103699	443	aug Coa/CCa	alg		IVI S/I	
1000000 901 Cubicol Unit Till P Cubicol 1000000 476 ICa Till P S 1000000 557 CCI CII P L 1000000 652 ICG Til S L 100000 652 ICG Til S F 100000 652 ICG Til S F 1000001 700 Til Cci S P 1000001 700 Til Cci Til P S 1000001 721 Cci Til P S S 1000001 721 Cci Til P S S 1000001 721 Cci Til P S S 1000001 721 Cci Til <td>103682</td> <td>460</td> <td></td> <td>ica/iia</td> <td>P</td> <td>5/L</td> <td>– C</td>	103682	460		ica/iia	P	5/L	– C
100000 1/0 Lua Lia S L 100000 1/0 Lua Lia S L 100000 5951 CC CT P L 100000 5951 ICa ITa S L 100000 5951 ICa ITa S L 100000 652 ICq ITa S L 100000 652 ICq ITa S L 100000 652 ICq ITa S F 100000 652 ICq ITa S F 100000 700 Tet Cct S P 100000 721 Cct Tet P S 100000 721 Cct Tet P S 100000 721 Cct Tet P S 100000 721 Cct Tet P S S 100000 721 Cct Tet P S S S S<	103681	461	tCo	tTo	P		
100398 944 Cci 1ci P S 100395 557 Cci cfi P L 100395 551 103 1fa S L 100395 614 103 1fa S L 100395 614 103 1fa S L 103466 676 Cgt Tgt R C 103444 685 acgg aTg T M 1004971 700 Tct Cct S P 100695 711 101 ITt S F 1006971 700 Tct Cct S P 100695 711 101 ITt S F 1006971 700 Tct Cct Tt P S 100697 71 101 Tt S F S 100697 721 Cct Tt P S S 100697 1721 Cct Tt S S S	103666	4/6	iCa Cat	tia	S	L	
100583 957 CCL CTL F L 100584 957 CCL CTL F L 100584 957 CCL CTL F L 100584 957 CCL CTL S L 100584 653 CCg Tgl R C 100447 6555 ICL TTL M P 100447 6555 ICL TTL S F 100447 6555 ICL TTL S F 100471 700 TtL CCL S F 100550 721 CcL TTL S F 100550 721 CcL TTL S K 20050 721 CcL TTL S K 210050 721 CcL TTL S K 210050 721 CcL TTL K K K 211 ACCHNC SEQUENCE 730 b K K K K K <tr< td=""><td>103598</td><td>544</td><td>CCI</td><td>T</td><td>P</td><td>5</td><td></td></tr<>	103598	544	CCI	T	P	5	
100301 381 LLa 114 S L 100302 381 LLa 114 S L 100310 6322 LGg 117g S L 1003466 676 Ggt Tgt R C 103466 676 Ggt Tgt R C 103447 695 LGt Tt M P 100671 700 Td Cd T M 100620 721 Cd Td P S 100620 721 Cd Td Cd M 100620 721 Cd	103585	557	CUT	CIT	P		
103528 614 CCC ALTOROUTE ATTECTACE ATTECTOR TATECTOR COARCEATE CANTAGENT TATECTACE ATTECTOR TATECTOR ACCORCEATE ATTACTOR COARCEATE ATTACTOR ACCORCEATE ATTACTORA COARCEATE CARATECE TATEGEATE TATEGEATE TATEGEATE TATEGEATE TATEGEATE TATEGEATE TATEGEATE TATEGEATE TATEGEATE ATTACTOR ACCORCEATE ATT	103561	581	tCa	tia	5	L	1
103310 6522 LCg 11g R C 103466 675 Cg1 Tit R C 103447 695 LC1 Tit S F 103444 695 LC2 Tit M 100571 700 Td Cd T M 1005805 71 IC1 Td P S 100829 721 Cd Td P S 100829 742 Cgg Tgg R W	103528	614	tCa	tla	S		
103465 676 Cgt Igt P C 103447 655 CL ITI S F 103444 698 aCg aTg T M 100671 700 Tu CcL S P 1006805 71 ICL ITI S F 1006805 721 CcL Td P S 1006805 721 CcL Td P S 100829 742 Cgg Tgg R W	103510	632	tCg	tig	s	L	
103447 695 tCl TI M 103444 6956 aCg aTg T M 100671 700 Tcl Ccl S P 1005805 71 tCl Ttl S F 100650 721 Cct Tcl P S 100829 742 Cgg Tgg R W	103466	676	Cgt	Igt	н	C	
103444 698 aCg alg I M 100971 700 Tct Cct S P 103605 71 1Ct It S F 100650 721 Cct Tct P S 100829 742 Cgg Tgg R W SECUENCES SECUENCE 783 bg FST BASE COUNT 214a 169: 159g 2411 1 AtGATTGTTC TAGAATGGCT ATTCCTCACA ACGCTCCTT GTGATGCTGC GGAACCATGG CAATAGGAT TCCACATACA ACGCAACGCA	103447	695	tCt	tīt	s	F	
100871 700 Ict Cct S P 1008805 71 ICi ITi S F 100850 721 Cct Tct P S 100829 742 Ggg Tgg R W SEOUENCES GENOMIC SEQUENCE 783 bp FACTA BASE COUNT 214a 1896 1599 2411 1 ATGATTGTTC TAGAATGGCT ATTCCTCACA ATCGCTCCTT GTGATGCTGC GGAACCATGG CAATAGGAT CTCAAGAGCG AGCAACACCT ATGATGCAGG GAATAAGGAT CTCGAACAGCT ATGAGTGCT CAGAACTGAGT TAGCATTC CAGAACTGCT ATGAGTGCTAGC CAGAACTGGT TAGCATCA CAGCAACGCT ATGAGTGCT CAGAACTGGT CAGAACTGCT ATGAGTGCT CAGACTGAGT CAGAACTGCAT ACCACACT ATGAGTGCT CAGCAGGAGGA GTATGTGGT CAGACTGAGT CAGAACTGCT ATGAGTGCT ACCACCTA TTCCGTGAT ATACCATCA TTTCCGTCGTGT ATGACTGCG CAGCAAGGTGT ATACCATCA TTGCCTGCA ATGGTGTGGG CAGTTAGGAGT TAGCATTCAGGTGCT CAGGTGTGGG CAGTTAGGAGT TAGCATTCAG CAGAACTGCT ATGGTGTCGGGATTCC CAGAATGGGT CAGATGGGT CAGATGGTGG CAGTTAGGAGT TAGCATTCA ATGCGTGT CAGGGGAGGGT TAGCATCG CAGAACTGGT AGAACGCGA ATGGGGT CTAACGTGCT TAGGTGCGGGAACTGT TAGGTGTGGG CAGTTAGGGT TTGCGGGG AGCTAACGGT AGAAGGGGT TAGGTGTT CGGGGAAGGCT TAGGTGTGGGGCT TTGGGGGAT TGCAGGGG AGTTAGGGTT TGCGGGGAAGCT ATGGTGTGGG CAGTTAGGGAT TGCGGGGAAGCT TAGGTGGGG CAGTTAGGGAT TGCGGGGAAGCT TAGGTGTGGG CAGTTAGGGAT TGCGGGGAAGCT TAGGTGTGGGG CAGTTAGGGAT TGCGGGGAAGCT ATGGTGTGGG CGTTGGGGGCT TAGGTGGGGGCT TAG 1 ATGATGTGT CAGAAGGGA ATTGCGT ATGGTGCGGC GAGAAGGGG TAGGTGGGGGC TTAGGTGGG CAGTTAGGAGT TGCGGGGAAGCT ATGGTGGG CGTTGGGGCT TAGGGGGATGT TGCGGGGAAGCT TAGGTGGGGGGA GTAGTGGGGGG GGAAGGGGGA GTAGGGGGGA GTAGGGGT GGAAGGGGG GGAAGGGGGA GTAGGGGGGA GTAGGGGGGGA GTAGGGGGGA GTAGGAGGGGGGAGGGA	103444	698	aCg	alg	T	м	
105805 71 ICI III S F 100850 721 Cct Tct P S 100829 742 Cgg Tgg R W SECUENCES GENOMIC SECUENCE 789 bp FASTA BASE COUNT 214a 169c 1599 2411 1 ATGATISTIC TAGAATGGCT ATTECTACA ATCECTCCT GIGATECTIGE GBAACCATGE CAATTAGGAT CCAAGGACGC AGCAACCA ATGCCGCAAG GAATTAGGA CTTACATCAC SECUENCE TOTIGE GAATTAGGACT THE GITTICETAT CACGGATETIE GATTACCACTA THACCAGTC CAGGATACCA TG GAAGGATACA TACCACTACT TAGACATCA CAGGATACCA TG GAAGGAGAT ATACCACTCACT TAGACAGCA AGCAACCACCA ATCCCGCGAA GAATTAGGA CTTACACTCACT SECUENCE 783 bp FASTA SECUENCE TAGGATGGC ATTECTACA ATCECCCA TATTEC TATACCATC TATECTACT TH TGACGATT ACCGATCA TAGGATAT CAGGATCACTG GAATACCACTG TATACACTCACT TAGCATCACT TAGATGCACACTG CAGTAGAGATACA AATCACACTCACTG TATACCACTACT TAGCATCACT TAGCATCATCA TAGCATCAT TAGCATCACT T	100871	700	Ict	Cct	5	P	
100890 721 Cut Int P S 100829 742 Ggg Tgg R W SEQUENCES SECUENCE 783 bp [ASIN] BASE COUNT 214a 1896 159g 2411 1 A TGATTGTTE TAGAATGEGT ATTCCTACA ATGEGTCCTT GTATGECTEG GGAACCATGE CATTAGGAT TCAAGCGC AGCAACCT ATGATGCAAG GAATAATGGA CTAACATCA 21 ATGATTGTTE TAGAATGEGT ATTCCTACA ATGEGTCCTT GTATGECTEG GGAACCATGE CATTAGGAT TCAAGCGC AGCAACACCT ATGATGCAAG GAATAATGGA CTAACATCA 21 ATACTTTT CTCTACAT ATTGCTATG CACGA TGEGTCCTT GTATGECTEG GGAACCATGE CATTAGGAT TCAAGAGCC ATCCCCCAAG GAATAATGGA CTAACAACCT 21 ATGATTGTT CAGAATGEGT ATTCCTACA ATGEGTCCTT GTATGECTEG CGAACCATGE CATTAGGAT TCAAGCACCA ATCCCCCAAG GAATAATGGA CTAACAACCT 21 ATGATGTT TAGGATTATGE GATTGGACATA ATTCCCCCAA ATGCGTCAAT TATGGTCGT TATACTCACA AGCAGCACATA TATGCATGCT ATGATGAT TATGCATGTT ATGCAGGAGAGT TAACATCAT TATGGACGT ATGATAGCT ATGCAGCAA ATGCATAG TATGATGA CAGCCTCAAT TATGGATGTA TAGGATTATGGA TATGATGAT CAGGACGAGT TATACATGAC GAGAAGGAGT TTACATGGA GAATAGTGGA GAATAATGGA CTAACAGTGT CAGGGAGGAGT TTACATGGA CGGGGAAGCT TA 21 ATGATTGTT TAGAATGGA ATTGCTTACA ATGCTCTT TGGATGATA TAGGAT TTCAAGGCGA AGCACCCA ATGCGCCCCCAA ATGGATGAT ATGGATGATGA GAATAATGGA CTATACATCAC GGGGGGAGCT TATGCTAATGGAG GAACCAGTGC CAATAGGAT TTCCAAGGACAA GAACGAACCAA ATGCGGGAAGGATTATAGGAT TTGCAAGAAGGA AGCAACCAA ATGCGGGAAGGATTATGGA TATGGATGATTGGA GAATAATGGA GAATAATGGA CTTAACAATGAA TTGGATGATA TAGGATGATTGGA GAATAATGGA CTATACATGGA TATGCTAATGCTA TATGGTCGTT TAGGACGAGAGCAA TAGGAGAGGAA ATTAGGAT TTCCAAGAACGAA AGCAACCAA ATGCGGGAAGGATTATACACGGGGAAGGAGT TTACATGGA GAGGAAGGACAA ATGCGGGAAGGAA	105805	71	tCt	tit	s	F	
100829 742 Cgg 1gg H W SEQUENCES GENOMIC SEQUENCE 783 bp [k51] 11 1 ATGATTGTTC TAGAATGGCT ATTCCTCACA ATCCCTCCTT GTGATGCTGC GGAACCATGG CAATTAGGAT CTCAAGAGCG AGCAACACCT ATGATGCAG GAATTAATGGA CTTACATCAC 121 GATACTTTT ICTTCCTCAT ICTGATTIG GTTTTCGTAT CACGGATGTT GGTTGCGGCT TATAGGAT ICTCAACTACA GAACGAACCA ATCCCTGCACA GACTAATGGA CTTACAACACT 122 GATACTTTT ICTTCCTCAT ICTGATTIG GTTTTCGTAT CACGGATGTT GGTTGCGGCT TATAGGAT ITCCAACTACA GACGAACGCA ATCCCGGCAATGA GAGTGATC CAGGATGATC GAGCATACAATGA TATCGAACT TATGCACTT ITGCACTTACA GACGAACGCA ATCCCTGCACT ATCCAACGACT 241 ATCGAACTCA ATTGGTTGATGA ATACTGGAC TATACCAGTT ITGCATTGTA GACCGAGACGAACGA TACCAGCACT ATGATGGAC 351 ATTGGACAT ATGGTGTGTGA CCAGCCAAAA CTAATCCAGC TATATTGTA CACCCTGCTG ATGTACCTCA ATGGATGTGGA GTACCTCCA CAGGGGTGAA ATTGGATGTGA 351 ATTGGACAT AGGGGACATAG AGTGGTGTGA CCAGCCAAAA CTAATCCAGC TATATTGTA CACCCTGCTG ATGTACCTCA ATGTGGGCT GTACCTCCC AGGGGTGCAA ATGGAGTGTGA 351 ATTGGACAT AGGGGACATAG GACGGTGAAC CAACAGACGCA ATTACCTGGA TATATTGTA CACCCTCACT ATGTGGACT GTACCTCCA TATGTGGGCT TAGGGGCT GACCTTCCC AGGGGTGCAA TATGGAGTGT 351 ATTGGACAT AGGGCACTATTCCTTACA ATTGCTCAAC CAGAAGAGGACT TAA CDNA SEQUENCE 783 bp [kst] MGA BASE COUNT 214a 147c 159g 2631 MGA 1 ATGATGTTT CTGGATATGA ATTGCTCTTT GGATGCTGC GGAACCATGG CAATTAGGAT TTGCAGCACA AGCAGCCA ATCCCTCCT CAGGGGGAACTATC 353 ATTGGACAT AATGCGATATCCTGAA ATTGCGACT ATTGCCTGTT GGAAGGACGC AGCAACGCC AGCAACCCC ATGGGCAATATAGGA CTTACAACGT 354 ATTGGACATA AGCGAATGAA ACTGACTGAC ATTGCGTGCA TATACATCAT TIGCCTGTT TAGGACGTA TAGGACGA ATTGCGGCAT ATTACAACGT 354 ATTGGACATA AGCGAATGAAGA CTAACAGTCC CAGGGGAGGACT TAAC 355 ATTGGACATA AGGGAATGAA AGCAGGTA ATACGGACGA ATTACCACGAC ATTGCTGCA ATGAAGGACGA ATACCAGGACGA ATCCCCCACT ATGGACGA TAGG	100850	721	Cct	Tct	Р	S	
SEQUENCES GENOMIC SEQUENCE 783 bp FASTA BASE COUNT 214a 1690 159g 2411 1 ATGATIGTIC TAGAATGGCT ATTCCTCACA ATCCCCCTT GTGATGCTGC GGAACCATGG CAATTAGGAT CTCAAGGACC ATGCCCCAAA GGAATGTGA GGAATATGGA CTTAACACCA 212 GATACTTIT TCTCTCTAT TCTGATTTTG GTTTCCTAT CACGGAAGTT GGTTCGCCGCT TATGGCAT TCCACTACA ACCGACACCA ATCCCGCAAA GGAATGTGA GGAATGTGA GGAATGTGA 236 ATTGGAACTA ATTGCTAGGA TATTCCCAGA TATACCCTCA TTGCTCTT TATGCCATT TCCACTGATA ACCGAACCCA ATCCCGCGAAA GGAATGTGA GGAATGATGGA CTTAACACT 241 ATCGAAGTC TTGGACAT ATTTCCTAGA TATTCCCAGA GGAAGGAT ATACGATCA CGAGAGGAGA TTACCAGTT TTGCACATT TCCACTGAT TCCACAGGAAGAAC CAAAGCACACC TATGGACGA TATGCAGATGAAAAGAAAA	100829	742	Cgg	Igg	R	W	
GENOMIC SEQUENCE 783 bp FASTA BASE COUNT 214a 1690 159g 2411 1 ATGATTGTTC TAGAATGGCT ATTCCTCACA ATGCGTCCTT GTGATGCTGC GGAACCATGG CAATTAGGAT CTCAAGACGC AGCAACACCT ATGATGCAAG GAATAATGGA CTTACATCAC 21 ATGATTGTTCT TCTCCTCAT TCTCGATTTG GTTTCGTAT CACGGATGTT GGTTCGCGCT TTATGCCAAT GAACGAACCCA ATGCCGCAAA GGATTGTTCA TGGAACTACT 24 ATGCGAATCA TCTCCTCAT TCTCGATTTG GTTTCGTAT CACGGATGTT GGTTCGCGCT TTATGCCAAT GGACGAGGTA GTAGTACATC CAGCCATTAC TATCCAAGCT 35 A ATTGGAACTC ATTGCTACTAGAG TATTCCCCAA ATGCACTGG CTATATGGAT CCAAGCAG CCACTCGT TATGCCAAT GGACGAGGTA GTAGTACATC CAGCGATTAC TATCCAAGCT 481 TTATTAGAAG TGGACATAGA GATGGTTGTA CCACGCAAAA CTAATCTACG TATTTTGTA ACACCTGGT GATACTCCA TAGTTGGGT CAATCACGT 481 TATTAGAAG TGGACATAGA GATGGTTGTA CCAGCCAAAA CTAATCTACG TATTATTGTA ACACCTGGTG AGTACTCCA TAGTTGGGT CAATCGACGT 481 TATTAGAAG TGGACATAGA GATGGTTGTA CCAGCCAAAA CTAATCTACG TATTATTGTA ACACCTGGTG AGTATCGGTG AACTAATCAT GCCTCACGT CTATCGTGGT CAATGGAGT 481 TATTAGAAG TGGACATGA ATTGCTCTTAT TGGTAGTGCTG GGAACCAGGG CAATTAGGAT TTCCAGTAGA TAGTGGAT CTAAGGATGT 482 TATTGTTT TCTCCTCAT TCGGATGTAC AATTGCATCTT GTGTGGTGGC GGAACCAGGG CAATTAGGAT TACATGAG 483 TATGGAAGTC ATTGCCTTATC GGAGTGTC AATGCATTT TGGTGGTGGC GGAACCAGGG CAATTAGGAT TTCCAGGAGGA GAATAATGGA CTTACATCAC 484 TTATTAGAAG TGGACATAGA ATTGCCTCTTTG GTATGCTGCT GGGAACCAGGC CAATTAGGAT TTAGGAAGT TACGGAAGGAT ATTGGGA CTTACATCAC 214 AGCAGATGCT ATTGCCTCAA TCTGCTGAT TATGGAGTT TAGGAGTT TAGGAGTT TAGGAGTAT TAGGAGATATTGGG GAATTAGGA CTTACATCAC 214 ATGCGAATTC TTGGACAA TATCGGACT ATACCTCGGA TATCCAGGC TATAGGAT TTAGGAAGTATT TAGGAAGTATT TAGGATGATT TAGGACT ATTGGGACTATAGGA TATCGAGGAT TATACGGAT ATACGAGCAGA TATGGAGATTAGGA TATCAGGACT ATACCTCGA TATGGAGACT TAATCGAGAGAGATATGGA CTTACATGGA TATAGGACT TAATGGAAGGACT TATCGAGGAT TAGGACTAGGA TATGGAGGAT TAGGACTATAGGA TATGGAGACTATAGGA TATGGAGACTAGGA TATGGAGACTAGGA TATGCAGGAAGGACT TAACGAAGCAGA TATGGAAGGAGGAT TACCTCAATGGA ATGTGGACT TAGGACGAAGGACT TACCTCAATGGA ATGTGGACTAGGA CTAATGGAA TAGGAGGAGGAT TACCTCAATGGA ATGTGGACGAGGAGGAAGGAAGGAGGAT TACCTGACGAGGAAGGACT TAA 41 TATAGAAG TGGACAATAG AGG	SEQUENCES						
BASE COUNT 214a 169c 159g 2411 1 ATGATTGTTC TAGAATGGCT ATTCCTCACA ATCGCTCCTT GTGATGCTGC GGAACCATGG CAATTAGGAT CTCAAGACGC AGCAACACCT ATGATGCAAG GAATTAATGGA CTTACATCAC 12 GATATCTTTT TCTTCCTCAT TATCATGTTTC GTGATGCTGG GGAACCATGG CAATTAGGAT CTCAAGACGC AGCAACACCT ATGATGCAAG GGATTGTTCA TGGAACACCT 241 ATCGACATC AATGGTATCG GACTTATGAG TATTCCCGA TGTCATGCA TTTGCTCGT TATATCCAAT GGACGAGGAT GATTAGATC CAGCACTAC ATGCGTCAA ATGCGACCAATG AGTGGTTCCCAA TTGCTCAGT TATACAGTC 345 ATGGACATC AATGGTTCT GGACTATGAG TATACAGTTC CAGTGACACG TATATGGAACCCAA TTACTGATGGACTACT TGGACATACG 445 TTATAGAAG TGGACATAAG AGTGGTTGTA CACCCCAAGACGAAG CAATTAGGA TTAACAGTC CAGTGACAGG TGATTACCTAT TAGAATGGAC CTATCGTCG AGTGGTCGAA ATGGAGTGTC 721 CCTAGGAAAG ATTATGGTTC TGGGTGACAC GAGAAGGAGT TTACTATGGT CAGTGCCG GAATGGGG GACTATCAT GCCCTACGT CTATCGTCGA ATGGAGTGTC 721 CCTAGGAAGA ATTATGGTTC TGGGTGTCC CAGTCACAG GGAGAGGAT TTACTATGGT CAGTGCGG GAATTAGGGT CTACCGTCG AGAAGGCTGT 721 CCTAGGAAGA ATTATGGTC TTGGATGTCT TGGATGCTC GGAAGCGCT TAA DONA SEQUENCE 783 bp [75576] MSA BASE COUNT 214a 147c 159g 2631 D D D D D D D D D D	GENOMIC SEQUENCE 783 bp	STA					
A TREATTENT: TAGAATGECT ATTECTAGA ATGEGTECTT GTATGETE: GGAACCATGE CANTAGGAT CTCAMAGGE: AGCAACACCT ATGATGECAG GAATTATGGA ETTACATCA 21 ATGAGATTE TITEGACAT ATTECTAGA ATGGATCT TGATGETE: GGAACCATG GAACCATA TITGCTAGT TAGAATGGA CTTACAGCGA 21 ATGGAGATTE ATTECTAGA TAGGATTATGGAT TATACGGAT TAGAGTGE CANTAGGAT TITGCAGT TAGAGTGAT 31 ATGGAGATTE ATTGETAGT ATGGATTATGGAT TATACGGAT TITGCAGAT TITGCAGAT TITGCAGAT AGGAGGAGGTA GAAGAAGGAT 31 ATGGAGATTE ATGGATATGAG TATTGCGAGA TATGGACTAATGE ATTACAGTTE 324 ATGGAGATTE ATGGATGATE GAGCTAGAT GAAGAAGGAGAT TITACAGTTE ACACCTGGAT GATGACTAA TAGGATGATE 324 ATGGAGATTE AGGCCTATAT TGGATGACTE AGGAAGGAGGT TITACAGTTE ACACCTGGAT ATGGATGACTE CAGGATGGAT 324 ATGGAGATTE AGGCCTATAT TGGATGACTE AGGAAGGAGGT TTACAGTGAT GAGAAGGAGAT TACAGTTAA CAGCTGGAT GATGACTGA AGGACGAGATGAT CCAGATGGAT 324 ATGGAGATTA GAGATGAGT AGGCCTATAT TGGATGACTG AGGAAGGAGT TTACAGTGA GAGAAGGAGT TACA CDNA SEQUENCE 783 bp	BASE COUNT 214a 169c 159g 24	1t					
241 ATCGGGAATTC TITGGACCAT ATTICCTAGT ATCATCCCGA TGTTCATTG TATACCGTC TATACCGTC TATACCGAT TATACCGTC TATACCGTC TATACCGTC AGAGGAGGTA GAGGAGGTA GTCGAGAGGAGTA CAATCAAGGT 361 ATTGGGAATC AATGGTATCG GACTTATGGT TATACCGCCA TGTTCATTG TATACCGTC TATACCGTCT TGGACGAGGTA GAGGAGGTA GTCGAGGAGGTA GTCGAGGAGGTA 481 TATAGAAG TGGACATAG AGTGGTGTGA CCGCCGAAA CTAATCAGT CGATGACGAG TCACTCGCT GTGAGCCGA GAGGAGGTA GTCCCCAGTCGCCAA TGGATGGT 691 GTACCTGGTC GTTCAAATCA GACCTCTATT TGGATAGCCG GAATCAGTGA CCGCGGGAGGGT TTACAGGAGGAGGTA GTCGAGGAGGTA GTCGGCGT CAATCGGTG 692 GTACCTGGTC GTTCAAATCA GACCTCTATT TGGATGCCG AGAGGAGGT TTACTATGGT CAGTGCCG GAATCGGG GACTAATCAT GCCCTACGT CTATCGTCGA AGGAGGGTGT 721 CCTAGGAAAG ATTAGGTC TGGGGTAGCC AATCAATTAA TCCCCCCAAAC CGGGGAAGCT TAA CDNA SEQUENCE 783 bp MGA BASE COUNT 214a 147c 159g 2831 1 ATGATGGTCT TGGACGAT GTCGTGTA TGGGAGGTGT TGGATGGTGG GAATCAGGG CAATTAGGAT TTCCAGGAGGC AGCAACCCA ATGCGGCAAG GAATAATGGA CTTACAGT 224 ATCGAGATC ATTGCTGAT TTTCGTAG ATTGCTGTT TGGATGCTGC GAACCACGG CAATTAGGAT TTCCAGGACGCA AGCGAGCCAA AGCCCCCAATGGA GGATGGTGAT GGACTACGT 324 ATCGAGATC ATGGTATTG GACTTATGGAT GTTGGATGTC CAATCGAGT GGTCGGCGAGGTA GTACGGACCAATG CAGTGGACAATG ATTGCGAATGGAT GACTAATCGA TATCGAGAGGAGGTA TTACGATGGA CAATCGAGGAGGAGT TTACAGTCC AGGAGGAGGTA GATGAGGAC CCAATCGCATGCAATGGA GACTAATGGA 324 ATCGAGATC ATGGTATTG GACTTATGGA TATCGCAA TGTCGATGTC TATCGCT TATGGCAATGCACCAATGCAGGAAGGAT TATCGATGGA GACTAATGGA TTACAGTCC AATGCGAATGAA GACTGATGGAC TACCGGCAATGGA GACTAATGGA GACTAATGGA GACTAATGGA TATGGAATGCA GACCAATGGA GACTAATGGA TTACGATGGA GACTAATGGA TTACAGTGCT GAATGGAGT TTACAGTGC GAATGAGGAGGAT TTACAGTGGC GACTAACCG TATCGTCAA ATGGATGGAC GACTAATGGA TTACGATGGA GACTAATGGA GACTGATT TGGATAGGA GACTAATGGA GACTGATT TGGATGAT GACTGCG GAGAGGGAT TTACAGTGGC GACTGACCAATGGA GACTGATTAGGA GACCTCAATGGA GACTGATT TGGATGGAT GACTGACGA GAGAGGAGGT TTACAGTGA GACTGACGAGGAGGAT TTACAGTGGC GACTGACCAGGGGAAGGCT TACGGC GACAGCGA GACTATGGA GACTGAATGGA GACTGAATGGA GACTGAATGGA GACTGAATGGA GACTG	1 ATGATTGTTC TAGAATGG	CT ATTCCTCACA ATCGCTCCTT	GTGATGCTGC GGAACCATGG CAATT	AGGAT CTCAAGACGC AGCAA	CACCT ATGATGCAAG GAATA	ATGGA CTTACATCAC	
Statissaca Aligenatic Anigenatic Anigenatic Anices Aligenatic Anigenatic Anigenatics Castanacas Charlacas Castanacas Castanacas <td>241 ATCGAGATTC TTTGGACC</td> <td>AT ATTTCCTAGT ATCATCCCGA</td> <td>TGTTCATTGC TATACCATCA TTTGC</td> <td>TCTGT TATACTCAAT GGACG</td> <td>AGGTA GTAGTAGATC CAGCC</td> <td>ATTAC TATCAAAGCT</td> <td></td>	241 ATCGAGATTC TTTGGACC	AT ATTTCCTAGT ATCATCCCGA	TGTTCATTGC TATACCATCA TTTGC	TCTGT TATACTCAAT GGACG	AGGTA GTAGTAGATC CAGCC	ATTAC TATCAAAGCT	
691 GTACCTGGTC GTTCAAATCA GACCTCTATT TCGGTACAAC GAGAAGGAGT TTACTATGGT CAGTGCAGTG	361 ATTGGACATC AATGGTAT 481 TTATTAGAAG TGGACAAT	AG AGTGGTTGTA CCAGCCAAAA	ATAACAGIIC CGATGAACAG TCACT CTAATCTACG TATTATTGTA ACACC	TGCTG ATGTACCTCA TAGTA	GGGCT GTACCTTCCT CAGAA	GTCAA ATGTGATGCT	
P22 CONSISTING FRANCIng Property and the antipartic consistence in the antipart of the antipart	601 GTACCTGGTC GTTCAAAT	CA GACCTCTATT TCGGTACAAC	GAGAAGGAGT TTACTATGGT CAGTO	CAGTG AGATTCGTGG AACTA	ATCAT GCCTCTACGT CTATC	GTCGT AGAAGCTGTT	
CDNA SEQUENCE 783 bp [FASTA_MSA] BASE COUNT 214a 147c 159g 2631 D 1 ATGATTETIC TAGAATGGCT ATTCCTTACA ATTGCTCTTT GEATGCTCC GEAACCATGG CANTAGGCA TICAGAAGC AGCAACCCT ATGATGCAAG GATTATGGA CTACCT 21 ATGATTCTTT CTACTACTA CTACTATTGA TATGCTCTT TATGGATGT GETTGCGCCT TTATGGCATT TCCACTATCA ACGAACCCA ATGATGCAAG GATTATGGA CTACCT 214 ATGGACATC TTTGGACCAT CTGATTTGGA TATGCTGAT TATGGATGTC TATACCGCT TATAGGCATT TCCACTATCA ACGAACCCA ATGCCGCAAA GGATGTTCA TATGAAGCT 214 ATGGACATC ATGGTATTG GACTTATGAG TATCCGAT ATGCTGAT TATGGATGTC CAGTGACAG TCCCGCAAA GGATGATCA CAGTGATGCA 315 ATTGGACATC AATGGTATTG GACTTATGAG TATCCGAC TATACGGTT TAACAGTC CAGTGACAG TCCCTGCT TAGATGGCT TAGATGGCT CAATTAGGTGT 316 ATTAGAAG TGACAATGA AGTGGTTGTA CCACCGAAA CTAATCAGTCC CAGTGACAG TCACTCGCT TAGATGGCT TAGATTGGC TATGTGCAA TTGGAATGGCT 316 ATTAGAAG TGACAATGA GACCTCTATT TIGGTACAAC GAGAAGGAGT TTACTATGGT TAGATGGCT GACTGACTGC TAGATGGCT TAGATTGGC 317 TTAGAACTA GACCTCTATT TIGGTACAC GAGAAGGAGT TTACTATGGT CAGTGCCAT GCCTTTAGGCC TATCGTCAA TGGTGATGCT 318 TTATGAAGATTAGGACT CTTGGGTACACC GAGAAGGAGT TTACTATGGT CAGTGCCAGTGG AGATTGGGC GACTATCAT GCCTTTAGGCC TATGGTCAA ATGGATGCT 319 TTAGAAG TATAGGTTC TIGGGTACAC GAGAAGGAGT TTACTATGGT CAGTGCCAGT GAGATGGCT GAGAAGGCGT TAA PROTEIN SEQUENCE 261 AA [FASTA] MSA [OMANNS 1 mixlewifilt ilcldaapy glgfddaap mmggindihh difffillil vvVbmlra lwhfhygtnp lpqrivhgt leilwtifps illmfiaips fallysmey vvdpaitika	721 CUTAGGAAAG ATTATGGT	TO TOBOUTATOO AATCAATTAA	TUTUTAAAU UUUUUAAAUT TAA				
A TGATTGITE TAGAATGGET ATTECTACA ATTGETETTI GTGATGETEG GGAACCATGG CAATTAGGAT TICAAGACGC AGCAACACCT ATGATGCAGA GAATAATGGA CTTACATCAC 12 GATATCTTAT TETTECTAAT TATGGATTITE GTTTTCGTAT TATGGATGET GGTGCGCGCT TTATGGCAT TICCACTACA AACGAACCCA ATGCCGGCAA GGATTGTGA TGTCACATCAC 241 ATCGAGATTC TTTGGACCAT ATTTCCTAGT ATCACCTGGA TGTTCGTGGC TATACCGTC ATTGCCTGT TATGGATGTT CCACTACAT AGGAAGGAGGTA GTAGTAGAC CAGCCATTAGC 341 ATGGACATC AATGGTATTG GACTTATGAG TTAACAGTC CAATGACGA TCACTCGCT ATTGCGATGTA ACATGGAC TACACTGCT 343 ATGGACATC AATGGTATTG GACTTATGAG TTAACAGTC CAATGACGA TTACTGCT TATGGACGTA ACATGTGGAC GTAGTAGAC 464 TTATAGAAG TGGACAATGA AGTGATGTGTA CACTCGCG ATTATGGAT TTACAATTGGAC GTACTAGAT 694 GTACCTGGTC GTTTAAATCA GACCTCTAT TTGGTACAAC GAGAAGGAGT TTACTATGGT CAATCGACTGCT 694 GTACCTGGTC GTTTAAATCA GACCTCTAT TTGGTACAAC GAGAAGGAGT TTACTATGGT CAATGCGC GACTAATCAT GCCTTTATGGCC GTACCATGCT 694 GTACCTGGTC TTGGGATGTGC CAATCAATTAA GACAACCAA GAAGAGGAGT TTACTATGGT CAATGGCG GACCTAATCAT GCCTTTATGGCC GTACCAGTGCAA CTGATGCT 694 GTACCTGGTC GTTTAAATCA GACCTCAAT TTGGTACAAC GAGAAGGAGT TTACTATGGT CAATGCAG GACTAATCAT GCCTTTATGC CTATCGTCGA AGGAGGGAGT 721 TCTAGGAAAG ATTATGGTTC TTGGGACGACCAACCAAC CGGGGAGGCT TAA PROTEIN SEQUENCE 261 AA FASTA MS DOMANNS 1 mixlwHIfIt ilcldaapy glgfddaap mmggindlhh diffflill vYubmilra lwhfhygtnp lpqrivhgtt ieilwtifps illmfiaps fallysmgev vydpaitika	CDNA SEQUENCE 783 bp FASTA	MSA					
1 ATGATTGTTC TAGAATGGCT ATTCCTTACA ATTGCTCTTT GTGATGCTGC GGAACCATGG CAATTAGGAT TICAAGAGCC AGCAACACCT ATGATGCAAG GAATAATGGA CTTACATCAC 121 GATATCTTTT CITCCTCAT CIGATATTIG GTTICGTAT TAGGATGT GGTGCGCGC TIATGGCAT TICAAGAGCC AGCAACACCT ATGATGCAAG 241 ATCGAGATCT TIFGGACCAT ATTTCCTGAT ATGGATGTTGAT TAGGATGT GGTGCACTACT AGCATGCA GTAGTAGATC CAGCCAATGCATTACTACT 361 ATTGGACATC ATGTGATTG GACTTATGGG TATTGGAGT TAACAGGTC CGATGAACAG TCACTCGCT ATGATGGATG GAAGAGGATT TAGAATGGG TCAATTACGT 481 TTATAGAAG TGGACATCA ATGTGGTTGTA CCAGCCAAAA CTAATCAGG TATACCGGT CGATGAACAG TCACTCGCT ATGATGGCC GTACCTCCT TAGGTGCAA ATGTGGGTCCAA TCAGTGGG 481 TTATTAGAAG TGGACATCA GACCTCTAT TIGGTACAC GTAACCAGTT CGATGACGAG TCACTCGCTG AGAAGGAGTTT TGGACGATGA GAAGGAGGAT 481 TTATTAGAAG TGGACATAG AGTGGTGTGA CCAGCCAAAA CTAATCACG TTACTAGGT TAACCTGCTG AGTGACTCA TAGTTGGCC GTACCTCCT TAGGTGCAA ATGTGGTGTCA 481 TTATTAGAAG TGGACATAG AGTGGTGTGA CCAGCCAAAA CTAATCACG TATACTGGT TAACTGGT GAGTAGATGAT TAGGATTGGG TCAATTAGGT 481 TTATTAGAAG TGGACATGA GACCTCTAT TIGGTACAAC GAGGAGGGAT TTACTATGGT CAGTGCCAGTG AGTTTGGCC GTACCTCCT TAGGTGCAA ATGTGGTGT 481 TTATTAGAAG TGGACATGA GACCTCTAT TIGGTACAAC GAGAAGGAGT TTACTATGGT CAGTGCCAGTG AGATTGGGC GTACCTCCT TAGGTGCAA ATGTGGTG 481 TTATTAGAAG TTAGGTCT CGATACACTACG AGGAAGGAGT TTACTATGGT CAGTGCCAGTG AGATTGGGG CGATGCT GTGCGCC GTACGCCGT GAGAAGCCGT 721 TCTAGGAAAG ATTAGGTTC CGATCAATCAATTAA TCCCCCAAAC CGGGGAGGCT TAA PROTEIN SEQUENCE 261 AA FASTA MSA DOMANS 1 mivlewlflt ilcldaapw glqfqdaap mmggindlhh diffflilil vrubmlra lwhfwynpn pqriwngtt ieilwtifps illmfiaips fallysmdev vvdpaitika	DAGE COUNT 2148 14/C 159g 26	51					U
241 ATCGGAGATIC TITGGACCAT ATTICIDAT INGUMENT GUIDANT GUIDANT GUIDANT AND	1 ATGATTGTTC TAGAATGG	CT ATTCCTTACA ATTGCTCTTT	GTGATGCTGC GGAACCATGG CAATT	AGGAT TTCAAGACGC AGCAA	CACCT ATGATGCAAG GAATA	ATGGA CTTACATCAC	
SEL ATTGEACATC AATGETATTG GACTTATGAG TATTCGGACT ATAACAGTIC GATGAACAG TCACTCACTT TAGACATTA TATGATTCA GAAGATGAT TAGATTGGG TCATTACGT 481 TTATAGAAG TGGACAATAG GATGGTGAT CCAGCCAAAA CTAATCTACG TATTATTGTA ACATCGCGT ATGACTTCA TAGTTGGGCT GTACTTCCT TAGGTGTCAA ATGGATGGT 681 GTACCTGGTC GTTAAATCA GACCTCATT TTGGTGACAAC GAGAAGGAGT TTACTATGTGA ACATCGCGTG AGATTTGTGG AACTAATCAT GCCTTTATGC CTATCGTCGT AGAAGCTGTT 721 TCTAGGAAAG ATTATGGTTC TTGGGTATCC AATCAACTAAC GAGAAGGAGT TTACTATGTC CAGTGCAGTG	241 ATCGAGATTC TTTGGACC	AT ATTTCCTAGT ATCATCCTGA	TGTTCATTGC TATACCATCA TTTGC	TCTGT TATACTCAAT GGACG	AGGTA GTAGTAGATC CAGCC	ATTAC TATTAAAGCT	
691 GTACCTGGTC GTTTAAATCA GACCTCTATT TTGGTACAAC GAGAAGGAGT TTACTATGGT CAGTGCAGTG	361 ATTGGACATC AATGGTAT 481 TTATTAGAAG TGGACAAT	TG GACTTATGAG TATTCGGACT	ATAACAGTTC CGATGAACAG TCACT	CACTT TTGACAGTTA TATGA	TTCCA GAAGATGATT TAGAA	TTGGG TCAATTACGT	
721 TCTAGGAAAG ATTATGGTTC TTGGGTATCC AATCAATTAA TCCCCCAAAC CGGGGAAGCT TAA PROTEIN SEQUENCE 261 AA FASTA MAA DOMANNS 1 mivlewlflt ialcdaeper glgfddatp mmggimdlhh diffflilil vfvlwmlyra lwhfhygtnp ipqrivhgtt ieilwtifps iilmfiaips fallysmdey vvdpaitika	601 GTACCTGGTC GTTTAAAT	CA GACCTCTATT TTGGTACAAC	GAGAAGGAGT TTACTATGGT CAGTO	CAGTG AGATTTGTGG AACTA	ATCAT GCCTTTATGC CTATC	GTCGT AGAAGCTGTT	
PROTEIN SEQUENCE 261 AA FASTA MSA DOMANS	721 TCTAGGAAAG ATTATGGT	TC TTGGGTATCC AATCAATTAA	TCCCCCAAAC CGGGGAAGCT TAA				
1 mivlewlflt ialcdaaepw qlgfqdaatp mmqgimdlhh diffflilil vfvlwmlvra lwhfhyqtnp ipqrivhgtt ieilwtifps iilmfiaips fallysmdev vvdpaitika	PROTEIN SEQUENCE 261 AA	STA MSA DOMAINS					
121 ignqwywtye ysuynssdeq sittasymip eddieigqir llevanrvvv paktnirliv tsadvihswa vpslgvkcda vpgrlnqtsi lvqregvyyg qcseicgtnh afmpivveav	1 mivlewlflt ialcdaae 121 ighqwywtye ysdynssd	pw qlgfqdaatp mmqgimdlhh eq sltfdsymip eddlelgqlr	diffflilil vfvlwmlvra lwhfh llevdnrvvv paktnlriiv tsadv	yqtnp ipqrivhgtt ieilw lhswa vpslgvkcda vpgrl	tifps iilmfiaips fally nqtsi lvqregvyyg qcsei	smdev vvdpaitika cgtnh afmpivveav	

FIGURE 2 | Editing informations stored in REDIdb are organized in specific flat-files in which it is possible to distinguish a header (A) containing the main features of the record (organism, Genbank accession, intracellular location, gene name, PubMed references, ecc.), a gene ontology box (B) describing the gene product properties, a feature table (C) with all the editing events and a sequence zone (D) with both the genomic sequence and the corresponding edited transcript/protein.

section (Figure 2D) contains the genomic sequence and the corresponding edited transcript. In coding protein genes, also the edited protein is displayed. Genomic sequences as well as edited transcripts and proteins can be retrieved in Fasta format.

Graphical Visualization

Edited cDNA and protein sequences can be explored in their evolutionary context through multiple alignments of available orthologs sequences. Since plant RNA editing tends to increase the sequence conservation along the evolution, annotated RNA editing changes are marked and visualized in the multiple alignment by the MSAViewer, to give rise to conservation levels and provide valuable comparative genomics information (**Figure 3A**).

In addition, RNA editing events are displayed along the edited sequence showing known functional domains and predicted secondary protein structures in order to better interpret the biological role of specific C-to-U or U-to-C changes (Figure 3B).

In case of complete organellar genomes, each genome is graphically rendered and edited genes can be selectively highlighted. Genome graphs are generated in SVG and include links to edited genes by mousing over. Further statistics such as the coding potential of the genome as well as the fraction of edited genes are also reported (**Figure 4**).

CONCLUSIONS AND PERSPECTIVES

As already mentioned, RNA editing plays an important role in transcriptome and proteome diversity. Since its first discovery in 1986 (Benne et al., 1986), a large number of events have been found in a wide range of eukaryotic organisms (Ichinose and Sugita, 2016). Only in humans more than 4 million events have been reported and dedicated resources such as DARNED, RADAR, and REDIportal have been developed to contain them into suitable specialized databases (Kiran et al., 2013; Ramaswami and Li, 2014; Picardi et al., 2017).

In the plant kingdom, RNA editing was first identified as C-to-U substitutions in mitochondrial transcripts (Hiesel et al., 1989), followed by its identification also in chloroplasts (Höch et al., 1991). In order to maintain a cured catalog of such events, we developed the specialized REDIdb database. Its third release, described here, contains three times more entries than the first version and two times more entries than the second version. To date, REDIdb is the unique bioinformatics resource collecting plant organellar RNA editing events. Indeed, similar databases such as dbRES





(He et al., 2007) or RESOPS (Yura et al., 2009) have been dismissed or are no more updated. Plant RNA editing events are also annotated in CloroplastDB (Cui et al., 2006), devoted to chloroplast genomes, and GOBASE (O'Brien et al., 2009), the organelle genome database. However, such resources are not specialized for RNA editing and include

potential not fixed errors due to the lack of manual curation (Picardi et al., 2011).

REDIdb 3.0 has been completely redrawn keeping in mind the simplicity as its working principle. RNA editing events are always shown in their biological context and novel graphical facilities have been added. Edited genes are now depicted in complete genome maps and RNA editing conservation can be investigated in pre-calculated multiple alignments of orthologous sequences. REDIdb 3.0 allows also the visualization of aminoacid changes induced by RNA editing in protein domains or secondary structures, providing insights into the potential functional consequences.

Next generation sequencing technologies, now arrived at their third generation, are expected to greatly increase the number of RNA editing candidates in the next future. Therefore, it will be indispensable to collect and annotate them in their biological context taking into account also the RNA editing levels.

Due to the unicity in its field, REDIdb is planned to be maintained and updated over time (as new editing sites or complete genomes are released), taking into account, as much as possible, eventual feedbacks from the users.

REFERENCES

- Benne, R., Van Den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986). Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819–826. doi: 10.1016/0092-8674(86)90063-2
- Bowe, L. M., and depamphilis, C. W. (1996). Effects of RNA editing and gene processing on phylogenetic reconstruction. *Mol. Biol. Evol.* 13, 1159–1166. doi: 10.1093/oxfordjournals.molbev.a025680
- Castandet, B., Choury, D., Bégu, D., Jordana, X., and Araya, A. (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.* 38, 7112–7121. doi: 10.1093/nar/gkq591
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163
- Cui, L., Veeraraghavan, N., Richter, A., Wall, K., Jansen, R. K., Leebens-Mack, J., et al. (2006). ChloroplastDB: the chloroplast genome database. *Nucleic Acids Res.* 34, D692–D696. doi: 10.1093/nar/gkj055
- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic. Acids. Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107
- Gray, M. W. (2003). Diversity and evolution of mitochondrial RNA editing systems. *IUBMB Life* 55, 227–233. doi: 10.1080/1521654031000119425
- He, T., Du, P., and Li, Y. (2007). dbRES: a web-oriented database for annotated RNA editing sites. *Nucleic Acids Res.* 35, D141–D144. doi: 10.1093/nar/gkl815
- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant mitochondria. *Science* 246, 1632–1634. doi: 10.1126/science.2480644
- Höch, B., Maier, R. M., Appel, K., Igloi, G. L., and Kossel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353, 178–180. doi: 10.1038/353178a0
- Ichinose, M., and Sugita, M. (2016). RNA editing and its molecular mechanism in plant organelles. *Genes* 8:5. doi: 10.3390/genes8010005
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., Mcanulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kiran, A. M., O'Mahony, J. J., Sanjeev, K., and Baranov, P. V. (2013). Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.* 41, D258–D261. doi: 10.1093/nar/gks961
- Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41, W575–W581. doi: 10.1093/nar/gkt289
- Maier, R. M., Zeltz, P., Kössel, H., Bonnard, G., Gualberto, J. M., and Grienenberger, J. M. (1996). RNA editing in plant mitochondria and chloroplasts. *Plant Mol. Biol.* 32, 343–365. doi: 10.1007/BF00039390

AUTHOR CONTRIBUTIONS

CL conducted the bioinformatics analyses and wrote the first manuscript draft; EP and GP conceived the study and contributed to writing and revising the manuscript.

FUNDING

This work was supported by ELIXIR IIB (CNR).

ACKNOWLEDGMENTS

We kindly thank TMR Regina and M. Takenaka for revising the database and fruitful suggestions, and L. Marra for technical and editorial assistance.

- O'Brien, E. A., Zhang, Y., Wang, E., Marie, V., Badejoko, W., Lang, B. F., et al. (2009). GOBASE: an organelle genome database. *Nucleic Acids Res.* 37, D946–D950. doi: 10.1093/nar/gkn819
- Picardi, E., D'erchia, A. M., Lo Giudice, C., and Pesole, G. (2017). REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* 45, D750–D757. doi: 10.1093/nar/gkw767
- Picardi, E., Regina, T. M., Verbitskiy, D., Brennicke, A., and Quagliariello, C. (2011). REDIdb: an upgraded bioinformatics resource for organellar RNA editing sites. *Mitochondrion* 11, 360–365. doi: 10.1016/j.mito.2010.10.005
- Ramaswami, G., and Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42, D109–D113. doi: 10.1093/nar/g kt996
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing. J. Mol. Evol. 48, 303–312. doi: 10.1007/PL00006473
- Sun, T., Bentolila, S., and Hanson, M. R. (2016). The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci.* 21, 962–973. doi: 10.1016/j.tplants.2016.07.005
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519
- Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., et al. (2016). MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics* 32, 3501–3503. doi: 10.1093/bioinformatics/btw474
- Yang, Y., Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., et al. (2017). SPIDER2: a package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. *Methods Mol. Biol.* 1484, 55–63. doi: 10.1007/978-1-4939-6406-2_6
- Yura, K., Sulaiman, S., Hatta, Y., Shionyu, M., and Go, M. (2009). RESOPS: a database for analyzing the correspondence of RNA editing sites to protein three-dimensional structures. *Plant Cell Physiol.* 50, 1865–1873. doi: 10.1093/pcp/pcp132

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Lo Giudice, Pesole and Picardi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.