

1 **The assembled Banana dihaploid mitochondrial genome is compact** 2 **with a high number of gene copies**

3 Shruthy Priya Prakash¹, Vaidheki Chandrasekar¹, Selvi Subramanian^{1*}, Rahamatthunnisha Ummar¹

4 ¹PSG College of Technology, Department of Biotechnology, Coimbatore, Tamil Nadu, India 641004.

5 Email ID: selvi.bio@psgtech.ac.in ORCID:Orcid.org/0000-0003-4061-0973

6 *Corresponding author

7 **ABSTRACT**

8 Banana being a major food crop all around the world, attracts various research interests in crop improvement. In
9 banana, complete genome sequences of *Musa accuminata* and *Musa balbisiana* are available. However, the
10 mitochondrial genome is not sequenced or assembled. Mitochondrial (mt) genes play an important role in flower and
11 seed development and in Cytoplasmic Male Sterility. Unraveling banana mt genome architecture will be a
12 foundation for understanding inheritance of traits and their evolution. In this study, the complete banana mt genome
13 is assembled from the whole genome sequence data of the *Musa acuminata subsp. malaccensis* DH-Pahang. The mt
14 genome sequence acquired by this approach was 409574 bp and it contains, 54 genes coding for 25 respiratory
15 complex proteins 15 ribosomal proteins, 12 tRNA genes and two ribosomal RNA gene. Except atpB, rps11 and
16 rps19 other genes are in multiple copies. The copy number is 12 in tRNA genes. In addition, nearly 25% tandem
17 repeats are also present in it. These mt proteins are identical to the mt proteins present in the other members of AA
18 genome and share 98% sequence similarity with *M. balbisiana*. The C to U RNA editing is profoundly higher (87 vs
19 13%) in transcripts of *M. balbisiana* (BB) compared to *M. accuminata* (AA). The banana AA mitochondrial genome
20 is tightly packed with 233 genes, with less rearrangements and just 5.3% chloroplast DNA in it. The maintenance of
21 high copy number of functional mt genes suggest that they have a crucial role in the evolution of banana.

22 **Key words:** Mitochondrial genome, *Musa acuminata*, DH Pahang, genome assembly, RNA editing, mitochondrial
23 genes

24 **INTRODUCTION**

25 The genomes of present day cultivated bananas consist of either diploids, triploids or tetraploids of AA (*Musa*
26 *acuminata*), BB (*M. balbisiana*) and combination of both (AB). The mitochondrial (mt) genome of banana is not
27 sequenced/assembled to date. The available AA genomic (D'Hont et al 2012) and BB genomic (Wang et al. 2019)
28 sequence resources help to reconstitute the mt genomes. NOVOPlasty is a *denovo* organellar genome assembly tool
29 which uses Whole Genome Shotgun (WGS) sequences to assemble circular organellar genomes (Dierckxsens et al.
30 2017). This program was tested to assemble the mt genomes of model plants rice and Arabidopsis and gave 99.9%
31 accuracy. Norgal (de Novo ORGAnellar DNA extractor) is one another pipeline available to extract the organellar DNA
32 from the Whole genome Shotgun (WGS) sequences (Al-Nakheeb et al. 2017). Using this, full circular mt genome of

33 Panda, a sea weed, butterfly and fungal genomes were assembled with 99.5 % sequence similarity with reference
34 sequences. Wang et al., (2018) used Newbler, Amos and Minimus software to assemble mt and chloroplast genomes
35 of an ornamental plant *Salix suchowensis*, and a fruit tree, *Ziziphus jujuba*. In the above methods there is no need to
36 separately isolate the organellar DNA of high quality. *Brassica oleracea* var. capitata mt genome assembly from
37 WGS is found to be 219,975 bp in size with no large repeats (Yang et al.2018). Recently, mt genomes of *Sinapis*
38 *arvensis* var. ‘Yeyou 18’, a cytoplasmic male sterile line (Nsa CMS) and its corresponding maintainer line
39 ‘Zhongshuang 4’ were assembled using the mt sequences present in the total DNA (Sang et al. 2020). Present work
40 deals with *Insilico* approaches to assemble and reconstitute the mt genome of banana species *Musa accuminata*
41 using the WGS data in comparison with other model plants such as maize, rice, sugar cane, sugar beet and
42 Arabidopsis.

43 The mt genome size of land plants range between 66 kb in *Viscum scurruloidem* to 11.3Mbp in *Silene*
44 *conica* (Omelchenko et al. 2020). Among them the herbaceous monocots have a range of 400-500 kb (Cuenca et al.
45 2013). Plant mt genomes evolve in a dramatic burst due to the presence of several repeats (Wynn and Christensen
46 2019). The reasons for the larger mt genome size in plants and lesser number of assembled mt genomes compared to
47 animals are also due to the rearrangements and recombinations caused by these repeats (Kovar et al. 2018).
48 Comparison of mt genome sizes of a diploid (A_2) and an allotetraploid (AD_2) cotton species indicated only a slight
49 variation in size and found to be 644 and 677 kb respectively (Chen et al. 2017). The cotton A, D and AD mt
50 genomes differed drastically and had four or six large repeats leading to lot of inversions and translocations. Mt
51 genomes of rice wide crosses and backcross inbred lines have shown radical change in the gene order and copy
52 number (Yang et al. 2020). In Fabales, a different kind of genome expansion has occurred in mt genomes.
53 Horizontal gene transfer events between intercellular and interspecific level have led to the genome size variation
54 (Choi et al. 2019). Considerable amount of mt DNA in plants are horizontally transferred to either nuclear or
55 chloroplast genome. DNA transfer events from mitochondria to nucleus are positively correlated to the size of
56 nuclear genomes in several plants (Zang et al. 2020). Often, these horizontally transferred genes are not activated
57 (Pinard et al.2019).

58 The mt gene content in plants do not vary much. Barley wild and cultivated varieties were found to have similar mt
59 genome size and gene content with only three SNPs. These genomes contain 33 protein coding genes, three rRNA
60 and 16 tRNAs (Hisano et al. 2016). Similarly, chiltepin pepper (*Capsicum annuum* var *glabriusculum*) has 31
61 known protein coding genes, three rRNA genes and 25 tRNA genes (Magdy and Ouyong 2020). In *Raphanus*
62 *sativus*, L. 40 protein coding genes, three rRNA genes and 23 tRNA genes were found in the mt genome (Peng and
63 Gao 2020). Very recently, six mitogenomes of *Damnacanthus indicus* was sequenced and found to have 32 protein
64 coding genes after several losses (Han et al. 2021). Plant organellar genes undergo post transcriptional modifications
65 such as splicing and editing, among them RNA editing specifically C to U change is found to be evolved during
66 early land plant development (Liu et al. 2011). Most of the editing sites create nonsynonymous changes leading to
67 protein change however they lead to a conservative change to maintain the function (Omelchenko et al. 2020). RNA
68 editing in protein coding genes is found to increase the protein function and codon bias. Higher frequency of editing

69 was observed immediately after exposure to salt stress in Barley mt nad3 gene (Ramadan 2020). However, in
70 Arabidopsis reduced RNA editing rate is found in heat stressed plants and suggested to have a regulatory role in
71 abiotic stress tolerance (Chu and Wei 2020). Differential RNA editing pattern was observed between Cytoplasmic
72 Male Sterile (CMS) and Fertile plants in Pigeon Pea (Kaila et al. 2019). Nodulation process which fixes atmospheric
73 Nitrogen is an energy demanding development involving higher mt activity. Higher splicing and RNA editing
74 efficiency is seen in nodulating roots (Sun et al. 2020). Edera and Sanchez-Puerta (2021) has recently developed a
75 computational frame work to identify editing sites in *Nicotiana tabacum* mt genome.

76 Assembly of AA mt genome will provide the genome size, number of respiratory and ribosomal proteins, tRNA and
77 rRNA genes in it. Assembling the organellar genome from the whole genome sequence data can be possibly done de
78 novo by using programs like Norgal (Al-Nakeeb et al. 2017) or by using reference sequence with CONTIGuator
79 (Galardini et al.2015, Halim et al.2016). This information will be useful for researchers who are interested in
80 analyzing characters which are controlled by nuclear and mitochondrial interaction (Hanson and Bentolila 2004). In
81 Maize (Weiwei et al. 2017) and Arabidopsis (Lee et al. 2017) embryo, seed development is modulated by genes
82 located in mt genome which is regulated by nuclear genes. Mitochondrial genes undergo post transcriptional
83 modifications such as splicing and editing which are regulated tissue and stage specifically by nuclear genes
84 (Hanson and Bentolila 2004). The knowledge on gene content, gene structure cis or trans spliced would enable a
85 researcher to correlate a molecular factor to a phenotype. This study focuses on assembling the mt genome of
86 banana for the above-mentioned applications. Besides, paternal vs maternal transmission (Faure et al.,1994) of mt
87 genes to the hybrids can be ascertained if the full genome is known in completion.

88 **METHODOLOGY**

89 *Data collection*

90 The contigs using WGS of the *Musa acuminata* subsp. *malaccensis* was collected from NCBI GenBank
91 CAJGYN000000000.1 (Bioproject: PRJEA82777) (D'Hont et al. 2012). Among the genomic sequences,12 mt
92 fragments have been separated from the nuclear genome. The contigs were quality checked removed from the other
93 contigs and stored in a separate file. These contigs were then subjected to nucleotide search using NCBI BLASTN
94 using 5 different reference mt genomes of Maize (NC_007982.1), rice (NC_011033), Arabidopsis (NC_037304.1),
95 sugar-beet (NC_002511.2) and sugarcane (NC_031164.1).

96 *Sequence assembly and circularization*

97 The sequence comparison of the mt contigs of *Musa acuminata* subsp. *malaccensis* DH Pahang resulted in several
98 fragments of sequence from each reference. The sequences obtained as a result of NCBI BLASTN are pooled
99 together and assembled using CONTIGuator (Galardini et al.2015) (<http://contiguator.sourceforge.net/>) that resulted
100 in a single scaffold. This program uses BLAST to align the draft sequence against the reference sequence and
101 provides a single scaffold. CGView (Grant and Stothard 2008) (<http://cgview.ca/>) is a Circular Genome
102 Visualization server was used to depict a circular genome of mitochondria. The scaffold sequence (1.2 Mb) as well
103 as assembled sequence in single fasta format were analyzed through CG viewer. Gene labeling in the map was done

104 in the same tool. Individual track of protein coding genes, ORFs, tRNAs and rRNAs were created as separate text
105 file and fed to the server for visualization. GC skew and GC content can be obtained from the server directly. The
106 location of the genes whether in sense or antisense strand were also manually recorded while labeling. Legend and
107 captions were also included in the figure through the software.

108 ***Repeats Identification***

109 Repeats in the assembled mitochondrial genome was identified using Tandem Repeat Finder which is a public
110 repository tool to identify repeats present in genomic DNA. (<https://tandem.bu.edu/cgi-bin/trdb/trdb.exe>). The
111 assembled mt genome in fasta file format was fed to the server to find repeats. Result was immediately obtained
112 with coordinates, sequences and copy number. It can be downloaded in any formats for further analyses. The copy
113 number and percentage of repeats were calculated using the obtained results.

114 ***Gene Annotation***

115 MFANNOT is a mt genome annotation (<https://megasun.bch.umontreal.ca/RNAweasel/>) (Beck and Lang 2010)
116 server that utilizes various tools and programs to provide detailed annotation on the introns, exon sequences using
117 comparative analysis. The input given is the scaffold sequence obtained from the CONTIGuator, as well as the
118 merged fasta file and the parameters are set to default with standard mitochondrion annotation settings. Time taken
119 for annotation depends on the size of the mt genome given as input, as the plant mt genomes are bigger in size, total
120 time needed can be up to 1- 2 hours. The results are directly sent to the email entered to the server. Results consists
121 of 3 files, one with the sequences translated, one with the annotations and other with the genes present. The GC
122 content of the scaffold was calculated using GC calculator. The translated sequences are separated and stored for
123 future use. Annotated information is manually recorded in MS excel sheet for easy handling of information. The
124 results were also compared with another annotation tool MITOFY (<https://dogma.cccb.utexas.edu/mitofy/>).

125 ***Comparative Analyses and Editing site identification***

126 The annotated list of mt genes identified in *Musa accuminata subsp. malaccensis* DH-Pahang are compared against
127 the available genomic data of other varieties, subspecies, species and genus by doing a BLASTP search. The results
128 were analyzed for variation in gene content, split genes and length variation in nucleotide and protein sequences
129 with reference to the *M. balbisiana* mt genes. Chloroplast DNA content in banana mt genome was analyzed by
130 BLASTN search of both mt and chloroplast genomes of banana. Editing sites for all the mt genes of *M. accuminata*
131 and *M. balbisiana* were identified using their DNA compared against their respective transcriptome data
132 (SRX10460839; SRX767394) available in NCBI. Pairwise comparison of mt DNA and transcript sequences
133 (cDNA) were carried using BLASTN program for all annotated genes. Variations corresponding to C in mt DNA
134 and T in cDNA were considered as C to U edited sites.

135 **RESULTS**

136 ***Mitochondrial genome assembly***

137 As a result of banana mt sequences assembly using the CONTIGuator, with five model organisms as
138 references, a single scaffold was obtained. The scaffold sequence was 409,574 bp in length with a GC content of
139 45.3% (Table 1). The percentage coverage and identity obtained in comparison with related model organisms,
140 Maize, rice, sugarcane, sugar beet and Arabidopsis are given in Fig 1. Annotation of this assembled genome by
141 MFANNOT resulted in the identification 97 genes including ORFs (Table 1; Fig.2). There are 45 ORFs present in
142 the genome which are 300bp and above reported by the program. The number of respiratory genes in banana mt
143 genome is 25, ribosomal protein genes 15, tRNA genes 12 and rRNA gene one. The coding regions of the genome
144 accounts to 22.8% of the total genome. There are 169 tandem repeats present in the banana mt genome
145 (Supplementary Table1). The length of the repeats ranges from 15 bp to 1500 bp. These tandem repeats accounts to
146 about 25.34% of the total mt genome size. The two copy repeats contribute to 21.22%. The three and four copy
147 repeats contribute to 2.61% and 1.51% of the mt genome size respectively.

148 ***Mitochondrial gene Analyses***

149 The location, copy number, presence and number of introns of these genes are listed in the Table2. The
150 number of genes with known unique function excluding the additional copies is 41 (Table 2). Out of these 41 genes
151 only three of them are single copy genes. Other genes have maximum of seven copies. The number of copies of
152 tRNA is found to be even more and one of them have fourteen copies. (Supplementary table2). There are ten split
153 genes present in the mt genome. These genes also contain groupII introns. Among them, atpB and ccmF are cis-
154 spliced and the rest eight of them are trans-spliced genes. Among the cis-spliced genes, atpB has two introns, and
155 ccmF is with single intron in four copies and one copy has two introns. Among the 233 genes including all copies
156 present in the mt genome (Fig.2) 141 of them are present in sense strand and the rest in antisense strand. A
157 comparative analyses of gene content with the four model organisms, maize, rice, Arabidopsis and sugarbeet are
158 presented in Table3. Among the respiratory complex genes, ccmFC and ccmB genes are not present in banana with
159 reference to the four model organisms. Ribosomal protein genes, rps1, rps7 and rpl2 are pseudogenes. Other
160 category genes, matr, mttB are absent compared to the other organism whereas ftsH is a pseudogene in banana.
161 There are seven pseudogenes identified in the annotation results which are all truncated in nature. However, there
162 are few genes which are intact with unconventional start and stop codons (Table 2). The amino acid length and the
163 percentage similarity of the mt proteins of other *M. accuminata* subspecies or varieties show 100% similarity except
164 a few (supplementary Table2). The average percentage similarity shared between *M. accuminata* and *M. balbisiana*
165 mt proteins is 98% (Table 2). There are four protein coding genes that are missing in *M. balbisiana* genome
166 compared with *M.accuminata*.

167 ***Mitochondrial RNA editing***

168 The mt RNA editing status was analyzed for all protein coding genes of *M. accuminata* and compared
169 against the *M. balbisiana*. The number of editing sites, their position, change of amino acid if any and intron
170 position are given in table 4. There is a drastic difference (41 vs 6) observed in the editing status between *M.*
171 *balbisiana* and *M. accuminata* mt genomes. The graph (Fig. 3) represents the genes in which editing differences are
172 found between *M. accuminata* and *M. balbisiana*. The bar length corresponds to the number of editing sites

173 observed. There were 18 genes which showed editing of mRNA in *M. balbisiana* species compared to only five
174 genes in *M. accuminata*. Among them nad6 and rps1 are unique which were not edited in *M. balbisiana*. The other
175 three genes, ccmC, rps12 and petB genes had just one editing site. The nad6 gene is the only gene that had two
176 editing sites in *Accuminata* however one of them is a synonymous change (Table 4). Twenty seven percent of the
177 editing observed in *M. balbisiana* are synonymous. atp 9 gene had the maximum of seven editing sites and a stop
178 codon is created by editing. Introns are also found to be edited in nad4L and nad5 genes of *Balbisiana*.

179 **DISCUSSION**

180 **Banana mt genome assembly from Whole Genome Shotgun sequence**

181 The contigs of mt sequences of DH Pahang were assembled from the genomic shotgun sequences. This approach of
182 extracting mt sequences from total genomic sequence was followed in several algal mt genome assembly.
183 Mitochondrial genomes of diatoms *Phaeodaetylum tricornutum* and *Talassiosira pseudonana* were obtained by
184 sequencing the total genome including plastid, mt and nuclear genomes (Secq and Green 2011). The shotgun
185 sequences were assembled by JGI/JAZZ assembler. The mt genomes were assembled by sequence similarity to
186 other algal mt genomes. Similarly, sequences of mt genomes of 10 algae were assembled from the genomic
187 sequences downloaded either from genbank or related published data (Guillory et al. 2018). These algal mt sequences
188 were selected based on GC content, size, BLASTN sequence similarity to other mt genomes. Falcon (Ver.4) was
189 used to assemble the genome and circular mt topology was predicted. CONTIGuator tool was used to assemble the
190 mt sequences listed in the DH Pahang CAJGYN000000000.1. This tool can address the sequence gaps and more
191 than one circular molecule (Galardini et al. 2011). Besides, this can also align contigs from a draft genome by
192 comparing it to several reference genomes based on their alignment and orientation (Galardini et al. 2015).

193 **Compact mt genome**

194 The mt genome size of banana is 409 kb. This is smaller than two other monocots, maize, 539 kb (Clifton et al.
195 2004) and rice, 490 kb (Notsu et al. 2002) mt genomes. The smaller size implicates compactly arranged genes with
196 lesser extent of repeated sequences. The gene number is found to be comparatively higher (97 vs 58) than maize mt
197 genome. Similarly, when the number of tandem repeat sequences were analyzed most of them are two copies and
198 the three and four copy number repeats contribute to only 4% of the genome (SupplementaryTable1). In sugar beet,
199 the repeat sequence TR1, which contains an array of 32 bp repeat sequences was found to be repeated to a maximum
200 of 13 times (Nishizawa et al. 2000). In solanaceous plants, a common origin of these short repeats was observed.
201 However, only in members of the tribe Hyoscyameae it has expanded to eight copies (Gandhini et al. 2019). Melon
202 mt genome which is 2740 kb has nearly 42% of repeat sequences (Rodrigues – Moreno et al. 2011). These repeat
203 sequences beside contributing to the mt genome size, also involve in recombination events that can further increase
204 the genome size. Sullivan et al. (2020) have performed comparative analysis on the repeat abundance and
205 recombination frequency in plant mt genomes. They reported that the recombination dynamics was heterogenous
206 among gymnosperms and short repeats of 200bp and below are actively involved in recombination in one third of
207 plant species they analyzed. There is a possibility of at least one recombination event in banana mt genome also.

208 Nearly 40% of the genes are in the antisense strand and the rest in the sense strand, there are only three shifts in
209 orientation in the full genome (Table2 and Fig 2.). Presence of promiscuous DNA from other compartments by
210 horizontal transfer might also contribute to the size of the mt genomes in plants. In apple approximately 20% of the
211 mt DNA is transferred from other compartments (Goremykin et al. 2012). In banana, just 5.3% of chloroplast DNA
212 was found in the mt genome (Table 1). The above observations, small genome size, possibility of one or two
213 recombination events, high gene copies and low chloroplast DNA content could be the reasons for the compact
214 nature of the genome.

215 ***High gene copy number***

216 Most of the banana mitochondrial genes are found duplicated and the maximum copy number is six in many of these
217 genes. This suggests genome duplication rather than gene duplication event. Segmental duplication of genes can be
218 identified when looking at the gene order repeated (Fig.2). For example, *rps12-nad3-nad1* order is repeated in the
219 banana mt genome. *M. accuminata* genome is a dihaploid with AA genome. The process of dihaploid development
220 by tissue culture may have increased the copy number of genes in this banana mt genome. Furthermore, paternal
221 transmission and the possibility of maintaining several subcircular forms of the intact mt genome may contribute to
222 this high copy number (Nagata et al., 1999; Tsujimura et al. 2019). In yeast, a linear scale increase in copy number
223 with ploidy number is observed in both genomes (De Chiara et al. 2020). Similarly, in Arabidopsis autopolyploids,
224 nuclear mitochondrial coordination is observed in genome duplication followed by tuned gene expression patterns in
225 mt genes (Coate et al. 2020). The copy number of tRNA genes are higher than the protein coding genes. Out of the
226 12 tRNA genes five of them match to the chloroplast genes. Chloroplast tRNA genes are found functional in
227 mitochondria of angiosperms (Richardson et al. 2013). In Magnolia six tRNA genes are transferred from plastid and
228 they are functional (Dong et al. 2020). This suggests that some of the tRNA could have chloroplast origin in banana
229 also. RNA mediated gene duplication events are found in several taxa of plants (Cuenca et al. 2012). They have
230 observed existence of processed paralogues along with precursors of *nad1* genes. In the case of banana DH Pahang
231 mt genome, DNA duplication is most likely as introns are retained in all copies. However, the possibility of
232 precursor mRNA/cDNA insertion is also likely since additional introns are gained in *ccmF* gene. Intact nature of
233 these additional copies of these genes in banana suggests that they are under functional constraint.

234 ***Gene content***

235 The total number of unique protein coding genes present in the banana mitochondrial genome is slightly
236 higher than other model plants. The core protein coding genes observed in Angiosperms is 24 (reviewed by Zardoya
237 2020). Mangnolia has the largest set of 64 unique mitochondrial genes among flowering plant (Dong et al. 2020).
238 However, the largest mt (11.7 Mb) genome assembled to date is from a gymnosperm which has 77 unique genes
239 (Putintseva et al. 2020). The total number of unique genes is 52 in banana which goes up to 233 when all copies are
240 considered. Hence the genome is compact with very little intergenic space. The number of tRNA and rRNA genes
241 are slightly lower than other mt genomes. The average number of tRNA genes found in mt genomes across
242 angiosperms is 17-29 (Michaud et al. 2011). In banana, the copy number of tRNA genes is up to twelve. The
243 number of pseudogenes is seven in banana mt genome; many of them could be remnant genes that might have been

244 recently transferred to nucleus (Subramanian et al. 2001). The *rpl2* gene is one example, which is a pseudogene in
245 wheat, functional in rice whereas many other plants do not have a mitochondrial copy. In banana, also a truncated
246 *rpl2* is present, the rest of the gene could not be matched, but the annotation result suggests both possibility of
247 pseudogene and a trans-spliced gene. A complete analysis of these truncated pseudogenes in closely related
248 organisms in both nuclear and mt compartment might provide evidence for the recent horizontal transfer events.
249 Recently, O’conner and Li (2020) also have proposed ‘Mitochondrial Fostering’ theory in which mitochondria play
250 an important role in arrival and development of orphan genes which are not present in any other plants. In banana,
251 there are several genes with either or both start and stop codon missing or non-classical are present. The
252 unconventional start and stop codons may not make the gene to pseudogene. In *Arabidopsis ccmF* (N2) gene is
253 functional despite the absence of classical start codon (Rayapuram et al. 2008). The *tatC* is a membrane transport
254 protein which is present other than respiratory complex and ribosomal proteins in banana. This is also found in few
255 other species including moss *Lycopodium cernuum* (Kanagara et al. 2021).

256 ***Banana A and B mitochondrial genome comparison***

257 Comparative analyses of *M. acuminata* sub sp *malacensis* DH-Pahang mt proteins with other AA genome mt
258 proteins show 100 percentage similarity (Supplementary Table 3). In some multicopy genes, one copy of these
259 genes shares a slightly lesser similarity however that is higher than the similarity percentage found between the two
260 species *M. Accuminata* and *M. balbisiana*. The average sequence similarity between these two *Musa* species is
261 98%. Barley and wheat mitochondrial respiratory proteins were almost similar with only few differences in few
262 genes (Hisano et al. 2016). In cotton allopolyploids, among A and D genomes, A is found to be the donor of
263 mitochondrial genomes of the progenitors (Chen et al. 2017). In banana also A and B genomes are available,
264 comparing B mitochondrial genome and autopolyploid progenitors would resolve the origin of banana mt genome.
265 In fishes, the mitochondrial gene similarity between allopolyploids were found to be higher when transmitted
266 maternally than paternal transmission (You et al. 2014). The paternal/biparental transmission of mitochondrial DNA
267 is documented in banana (Nagata et al.1999). When the editing sites were compared between two *Musa* genomes,
268 *Balbisiana* is having a drastic high number. However, the editing sites were not leading to conservative changes in B
269 genome. Similarly, A genome editing sites were different to that of B genome editing event. The A and B genome
270 sequences used for comparison are from wild genotypes that are under natural selection. Hence, they may be
271 evolving independently.

272 **Data availability**

273 The data underlying this article are available in NCBI (National Center for Biotechnology Information) and can be
274 accessed with accession numbers -CAJGYN000000000.1- *Musa acuminata* subsp. *malaccensis*, NC_007982.1-
275 Maize, NC_011033- rice, NC_037304.1- *Arabidopsis*, NC_002511.2- sugar-beet and NC_031164.1- sugarcane. The
276 assembled AA genome data will be submitted to NCBI after annotation.

277 **Acknowledgement:**

278 We thank all the authors and fellows who supported this idea and their encouragement towards completing this
279 article.

280 **Funding:** This study was initiated as part of a “North East Banana Consortium funded project, by Department of
281 Biotechnology, Government of India. It is a full in silico work, only the second Authors salary was from the fund.

282 **Conflicts of interest/Competing interests:** No conflicting interest

283 Literature cited

- 284 1. Al-Nakeeb K, Petersen TN, Sicheritz-Pontén T (2017) Norgal: extraction and de novo assembly of
285 mitochondrial DNA from whole-genome sequencing data. *BMC bioinform* 18(1):1-7
286 <https://doi.org/10.1186/s12859-017-1927-y>
- 287 2. Beck N, Lang B (2010) MFannot, organelle genome annotation webserver.
288 <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>
- 289 3. Chen Z, Nie H, Grover CE, Wang Y, Li P, Wang M, Pei H, Zhao Y, Li S, Wendel JF, Hua J (2017) Entire
290 nucleotide sequences of *Gossypium raimondii* and *G. arboreum* mitochondrial genomes revealed A-
291 genome species as cytoplasmic donor of the allotetraploid species. *Plant Biology* 19(3):484-493
292 <https://doi.org/10.1111/plb.12536>
- 293 4. Chen Z, Nie H, Wang Y, Pei H, Li S, Zhang L, Hua J (2017) Rapid evolutionary divergence of diploid and
294 allotetraploid *Gossypium* mitochondrial genomes. *BMC genomics* 18(1):1-15
295 <https://doi.org/10.1186/s12864-017-4282-5>
- 296 5. Choi IS, Schwarz EN, Ruhlman TA, Khiyami MA, Sabir JS, Hajarrah NH, Sabir MJ, Rabah SO, Jansen RK
297 (2019) Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. *BMC*
298 *plant biology* 19(1): 1-15 <https://doi.org/10.1186/s12870-019-2064-8>
- 299 6. Chu D, Wei L (2020) Reduced C-to-U RNA editing rates might play a regulatory role in stress response of
300 *Arabidopsis*. *J of plant physiol* 244:153081 <https://doi.org/10.1016/j.jplph.2019.153081>
- 301 7. Ck N, Lang B (2010) MFannot, organelle genome annotation webserver.
302 <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>
- 303 8. Clifton SW, Minx P, Fauron CMR, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti
304 S, Tayloe C, Meyer L (2004) Sequence and comparative analysis of the maize NB mitochondrial
305 genome. *Plant physiology* 136(3):3486-3503 <https://doi.org/10.1104/pp.104.044602>
- 306 9. Coate JE, Schreyer WM, Kum D, Doyle JJ (2020) Robust cytonuclear coordination of transcription in
307 nascent *Arabidopsis thaliana* autopolyploids. *Genes* 11(2):134 <https://doi.org/10.3390/genes11020134>
- 308 10. Cuenca A, Petersen G, Seberg O, Jahren AH (2012) Genes and processed paralogs co-exist in plant
309 mitochondria. *J of mol evol* 74(3):158-169 <https://doi.org/10.1007/s00239-012-9496-1>
- 310 11. D’hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M,
311 Da Silva C (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants.
312 *Nature* 488(7410):213-217 <https://doi.org/10.1038/nature11241>

- 313 12. De Chiara M, Barré B, Persson K, Chioma AO, Irizar A, Schacherer J, Warringer J, Liti G (2020)
314 Domestication reprogrammed the budding yeast life cycle. bioRxiv
315 <https://doi.org/10.1101/2020.02.08.939314>
- 316 13. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from
317 whole genome data. *Nucleic acids res* 45(4): e18-e18. <https://doi.org/10.1093/nar/gkw955>.
- 318 14. Dong S, Chen L, Liu Y, Wang Y, Zhang S, Yang L, Lang X, Zhang,S (2020) The draft mitochondrial
319 genome of *Magnolia biondii* and mitochondrial phylogenomics of angiosperms. *PloS one* 15(4):0231020
320 <https://doi.org/10.1371/journal.pone.0231020>
- 321 15. Fauré S, Noyer JL, Carreel F, Horry JP, Bakry F, Lanaud C (1994) Maternal inheritance of chloroplast
322 genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Current*
323 *Genetics* 25(3): 265-269 <https://doi.org/10.1007/BF00357172>
- 324 16. Galardini M, Biondi EG, Bazzicalupo M and Mengoni A (2011) CONTIGuator: a bacterial genomes
325 finishing tool for structural insights on draft genomes. *Source code for biology and medicine* 6 (1) :1-5
326 <https://doi.org/10.1186/1751-0473-6-11>
- 327 17. Galardini, M, Mengoni A and Bazzicalupo M (2015) Mapping contigs using CONTIGuator. In *Bacterial*
328 *Pangenomics* 163-176 Humana Press, New York, NY https://doi.org/10.1007/978-1-4939-1720-4_11
- 329
- 330 18. Gandini CL, Garcia LE, Abbona CC, Sanchez-Puerta MV (2019) The complete organelle genomes of
331 *Physochlaina orientalis*: Insights into short sequence repeats across seed plant mitochondrial genomes. *Mol*
332 *phylogenetics and evol* 137:274-284 <https://doi.org/10.1016/j.ympev.2019.05.012>
- 333 19. Garcia LE, Edera A, Palmer JD, Sato H, Sanchez-Puerta MV (2021) Horizontal gene transfers dominate the
334 functional mitochondrial gene space of a holoparasitic plant. *New Phytologist* 229(3): 1701-1714
335 <https://doi.org/10.1111/nph.16926>.
- 336 20. Goremykin VV, Lockhart PJ, Viola R, Velasco R (2012) The mitochondrial genome of *Malus domestica*
337 and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *The Plant J* 71(4)
338 :615-626 <https://doi.org/10.1111/j.1365-313X.2012.05014.x>
- 339 21. Grant JR, Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes.
340 *Nucleic acids res* 36(suppl_2): W181-W184 <https://doi.org/10.1093/nar/gkn179>
- 341 22. Guillory WX, Onyshchenko A, Ruck EC, Parks M, Nakov T, Wickett NJ and Alverson AJ (2018)
342 Recurrent loss, horizontal transfer, and the obscure origins of mitochondrial introns in diatoms
343 (*Bacillariophyta*). *Genome biology and evolution*, 10(6):1504-1515 <https://doi.org/10.1093/gbe/evy103>
- 344 23. Halim, MA, Rahman AY, Sim KS, Yam HC, Rahim AA, Ghazali AHA and Najimudin N (2016) Genome
345 sequence of a Gram-positive diazotroph, *Paenibacillus durus* type strain ATCC 35681. *Genome*
346 *announcements* 4(1): e00005-16. <https://doi.org/10.1128/genomeA.00005-16>
- 347 24. Han EK, Cho WB, Tamaki I, Choi IS, Lee JH (2021) Comparative Mitogenomic Analysis Reveals Gene
348 and Intron Dynamics in Rubiaceae and Intra-Specific Diversification in *Damnacanthus indicus*. *Int j of mol*
349 *sciences* 22(13): 7237 <https://doi.org/10.3390/ijms22137237>

- 350 25. Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male
351 gametophyte development. *The Plant Cell*, 16(suppl_1): S154-S169 <https://doi.org/10.1105/tpc.015966>
- 352 26. Hisano H, Tsujimura M, Yoshida H, Terachi T, Sato K (2016) Mitochondrial genome sequences from wild
353 and cultivated barley (*Hordeum vulgare*). *BMC genomics* 17(1): 1-12 <https://doi.org/10.1186/s12864-016-3159-3>
- 354
355 27. Kaila T, Saxena S, Ramakrishna G, Tyagi A, Tribhuvan KU, Srivastava H, Chaudhury A, Singh NK,
356 Gaikwad K (2019) Comparative RNA editing profile of mitochondrial transcripts in cytoplasmic male
357 sterile and fertile pigeonpea reveal significant changes at the protein level. *Mol biology rep* 46(2): 2067-
358 2084 <https://doi.org/10.1007/s11033-019-04657-2>
- 359 28. Kanagara k, kanzariya r, bhatt p, monpara j, chudasama k, thaker v (2021) Mitochondrial genome
360 characterization of lycopodium cernuum l. *J of Cell and Tissue Res* 21(1):7069-7076
- 361 29. Kovar L, Nageswara-Rao M, Ortega-Rodriguez S, Dugas DV, Straub S, Cronn R, Strickler SR, Hughes
362 CE, Hanley KA, Rodriguez DN, Langhorst BW (2018) PacBio-based mitochondrial genome assembly of
363 *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing.
364 *Genome biology and evol* 10(9):2501-2517 <https://doi.org/10.1093/gbe/evy179>
- 365 30. Lee K, Han JH, Park YI, Colas des Francs-Small C, Small I, Kang H (2017) The mitochondrial
366 pentatricopeptide repeat protein PPR 19 is involved in the stabilization of NADH dehydrogenase 1
367 transcripts and is crucial for mitochondrial function and *Arabidopsis thaliana* development. *New*
368 *Phytologist* 215(1):202-216 <https://doi.org/10.1111/nph.14528>
- 369 31. Liu Y, Xue JY, Wang B, Li L, Qiu YL (2011) The mitochondrial genomes of the early land plants *Treubia*
370 *lacunosa* and *Anomodon rugelii*: dynamic and conservative evolution. *PLoS One* 6(10):25836
371 <https://doi.org/10.1371/journal.pone.0025836>
- 372 32. Magdy M, Ouyang B (2020) The complete mitochondrial genome of the chiltepin pepper (*Capsicum*
373 *annuum* var. *glabriusculum*), the wild progenitor of *Capsicum annum* L. *Mitochondrial DNA Part*
374 *B* 5(1):683-684. <https://doi.org/10.1080/23802359.2020.1714496>.
- 375 33. Martínez-Cuenca MR, Iglesias DJ, Talón M, Abadía J, López-Millán AF, Primo-Millo E, Legaz F (2013)
376 Metabolic responses to iron deficiency in roots of Carrizo citrange [*Citrus sinensis* (L.) Osbeck.× *Poncirus*
377 *trifoliata* (L.) Raf. *Tree physiol* 33(3):320-329 <https://doi.org/10.1093/treephys/tpt011>
- 378 34. Michaud M, Cognat V, Duchêne AM, Maréchal-Drouard L (2011) A global picture of tRNA genes in plant
379 genomes. *The Plant J* 66(1):80-93 <https://doi.org/10.1111/j.1365-313X.2011.04490.x>
- 380 35. Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T (1999) The selective increase or decrease of organellar
381 DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta* 209(1):53-65
382 <https://doi.org/10.1007/s004250050606>
- 383 36. Nishizawa S, Kubo T, Mikami T (2000) Variable number of tandem repeat loci in the mitochondrial
384 genomes of beets. *Curr Genetics* 37(1):34-38 <https://doi.org/10.1007/s002940050005>
- 385 37. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The
386 complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition

- 387 and loss during the evolution of flowering plants. *Mol Genetics and Genom* 268(4):434-445
388 <https://doi.org/10.1007/s00438-002-0767-1>
- 389 38. O’Conner S, Li, L (2020) Mitochondrial fostering: the mitochondrial genome may play a role in plant
390 orphan gene evolution. *Front in plant science* 11:1855 <https://doi.org/10.3389/fpls.2020.600117>
- 391 39. Omelchenko DO, Makarenko MS, Kasianov AS, Schelkunov MI, Logacheva MD, Penin AA (2020)
392 Assembly and Analysis of the Complete Mitochondrial Genome of *Capsella bursa-pastoris*. *Plants*
393 9(4):469. <https://doi.org/10.3390/plants9040469>
- 394 40. Oudot-Le Secq MP and Green BR. (2011) Complex repeat structures and novel features in the
395 mitochondrial genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Gene*
396 476(1-2):20-26 <https://doi.org/10.1016/j.gene.2011.02.001>
- 397 41. Peng H, Gao J (2020) The whole genome assembly and evolution analyze of carmine radish (*Raphanus*
398 *sativus* L.) Mitochondrion. *Mitochondrial DNA Part B* 5(3):2252-2253
399 <https://doi.org/10.1080/23802359.2020.1772136>
- 400 42. Pinard D, Myburg AA, Mizrachi E (2019) The plastid and mitochondrial genomes of *Eucalyptus grandis*.
401 *BMC genomics* 20(1):1-14 <https://doi.org/10.1186/s12864-019-5444-4>
- 402 43. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, Konstantinov YM,
403 Shmakov VN, Belkov VI, Sadovsky MG, Keech O (2020) Siberian larch (*Larix sibirica* Ledeb.)
404 mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the
405 largest known mitogenome. *BMC genomics* 21(1):1-12 <https://doi.org/10.1186/s12864-020-07061-4>
- 406 44. Ramadan AM (2020) Salinity effects on nad3 gene RNA editing of wild barley mitochondria. *Mol biology*
407 *rep* 47(5): 3857-3865 <https://doi.org/10.1007/s11033-020-05475-7>
- 408 45. Rayapuram N, Hagenmulle J, Grienberger JM, Bonnard G, Giegé P (2008) The three mitochondrial
409 encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in
410 *Arabidopsis*. *J of Biol Chemistry* 283(37):25200-25208 <https://doi.org/10.1074/jbc.M802621200>
- 411 46. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD (2013) The “fossilized” mitochondrial
412 genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and
413 extraordinarily low mutation rate. *BMC biology* 11(1):1-17 <https://doi.org/10.1186/1741-7007-11-29>
- 414 47. Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, Garcia-Mas J
415 (2011) Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest
416 reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear
417 origin. *BMC genomics* 12(1):1-14 <https://doi.org/10.1186/1471-2164-12-424>
- 418 48. Sang S, Mei D, Zaman QU, Liu J, Cheng H, Fu L, Wang W, Wang H, Hu Q (2020) An efficient approach
419 for obtaining plant organelle genomes. *Oil Crop Science* 5(3):129-135
420 <https://doi.org/10.1016/j.ocsci.2020.07.003>
- 421 49. Subramanian S, Fallahi M, Bonen L (2001) Truncated and dispersed rpl2 and rps19 pseudogenes are co-
422 transcribed with neighbouring downstream genes in wheat mitochondria. *Curr genetics* 39(4):264-272
423 <https://doi.org/10.1007/s002940100204>

- 424 50. Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, Keech O, Öberg L, Møller
425 IM, Arvestad L, Street NR (2020) The mitogenome of Norway spruce and a reappraisal of mitochondrial
426 recombination in plants. *Genome biology and evol* 12(1):3586-3598 <https://doi.org/10.1093/gbe/evz263>
- 427 51. Sun Y, Xie M, Xu Z, Chan KC, Zhong JY, Fan K, Wong-Bajracharya J, Lam HM, Lim BL (2020)
428 Differential RNA Editing and Intron Splicing in Soybean Mitochondria during Nodulation. *Int j of mol*
429 *sciences* 21(24): 9378 <https://doi.org/10.3390/ijms21249378>
- 430 52. Tsujimura M, Kaneko T, Sakamoto T, Kimura S, Shigyo M, Yamagishi H, Terachi T (2019)
431 Multichromosomal structure of the onion mitochondrial genome and a transcript
432 analysis. *Mitochondrion* 46:179-186 <https://doi.org/10.1016/j.mito.2018.05.001>
- 433 53. Wang X, Cheng F, Rohlsen D, Bi C, Wang C, Xu Y, Wei S, Ye Q, Yin T, Ye N (2018) Organellar genome
434 assembly methods and comparative analysis of horticultural plants. *Horticulture research* 5(1):1-13
435 <https://doi.org/10.1038/s41438-017-0002-1>
- 436 54. Wang Z, Miao H, Liu J, Xu B, Yao X, Xu C, Zhao S, Fang X, Jia C, Wang J, Zhang J (2019) *Musa*
437 *balbisiana* genome reveals subgenome evolution and functional divergence. *Nature plants* 5(8):810-821
438 <https://doi.org/10.1038/s41477-019-0452-6>
- 439 55. Weiwei Qi, Zhongrui Tian, Lei Lu, Xiuzu Chen, Xinze Chen, Wei Zhang, Rentao Song (2017) Editing of
440 Mitochondrial Transcripts nad3 and cox2 by Dek10 Is Essential for Mitochondrial Function and Maize
441 Plant Development, *Genetics* 205(4):1489–1501 <https://doi.org/10.1534/genetics.116.199331>
- 442 56. Wynn EL, Christensen AC (2019) Repeats of unusual size in plant mitochondrial genomes: identification,
443 incidence and evolution. *G3: Genes, Genomes, Genetics* 9(2):549-559
444 <https://doi.org/10.1534/g3.118.200948>
- 445 57. Yang K, Nath UK, Biswas MK, Kayum MA, Yi GE, Lee J, Yang TJ, Nou IS (2018) Whole-genome
446 sequencing of *Brassica oleracea* var. *capitata* reveals new diversity of the mitogenome. *PloS one*
447 13(3):0194356 <https://doi.org/10.1371/journal.pone.0194356>
- 448 58. Yang W, Zou J, Wang J, Li N, Luo X, Jiang X, Li S (2020) Wide crossing diversify mitogenomes of rice.
449 *BMC plant biology* 20(1):1-12 <https://doi.org/10.1186/s12870-020-02380-w>
- 450 59. Yu, Jeong-Nam, Soonok Kim, Myounghai Kwak (2014) Complete mitochondrial genome sequence of a
451 Korean *Pungtungia herzi* (Cypriniformes, Gobioninae). *The J of DNA Mapp Sequencing and Analysis* 25
452 96):414-415 <https://doi.org/10.3109/19401736.2013.809435>
- 453 60. Zardoya R (2020) Recent advances in understanding mitochondrial genome diversity. *F1000Research* 9
454 <https://doi.org/10.12688/f1000research.21490.1>
- 455 61. Zhang GJ, Dong R, Lan LN, Li SF, Gao WJ, Niu HX (2020) Nuclear integrants of organellar DNA
456 contribute to genome structure and evolution in plants. *Int J of mol sciences* 21(3):707
457 <https://doi.org/10.3390/ijms21030707>

458

459 Fig 1: Graph showing query coverage of *M.accuminata* mitochondrial genome with Rice, Maize,
460 Sugarcane, Arabidopsis and Beet mitochondrial genomes

461 Fig 2: Circular view of assembled *M.accuminata* subsp. *malaccensis* DH Pahang with individual tracks
462 which depicts respiratory complex, open reading frame, tRNA and rRNA genes, along with GC skew and
463 GC content

464 Fig 3: Bar graph showing number of C to U editing sites identified in mitochondrial transcripts between
465 *M.accuminata* (blue) and *M.balbisiana* (orange). Length of the bar represents the number of editing sites
466 in respective genes

Fig 1

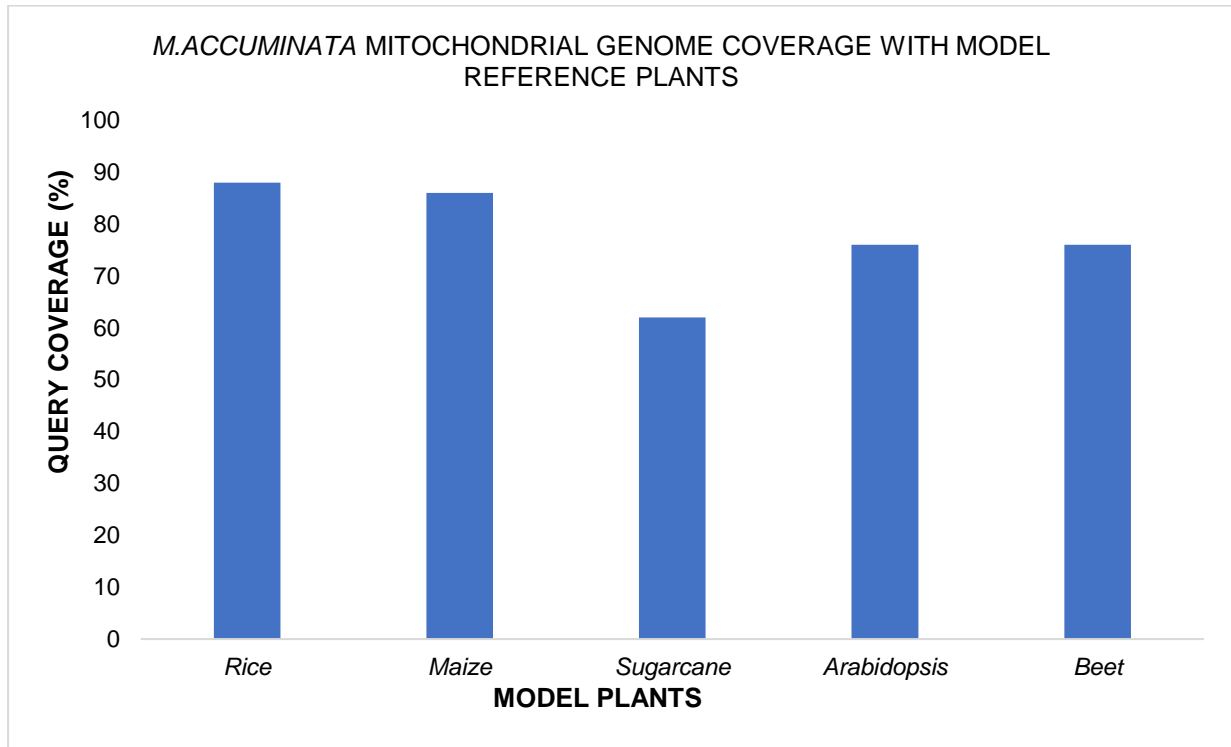


Fig 2

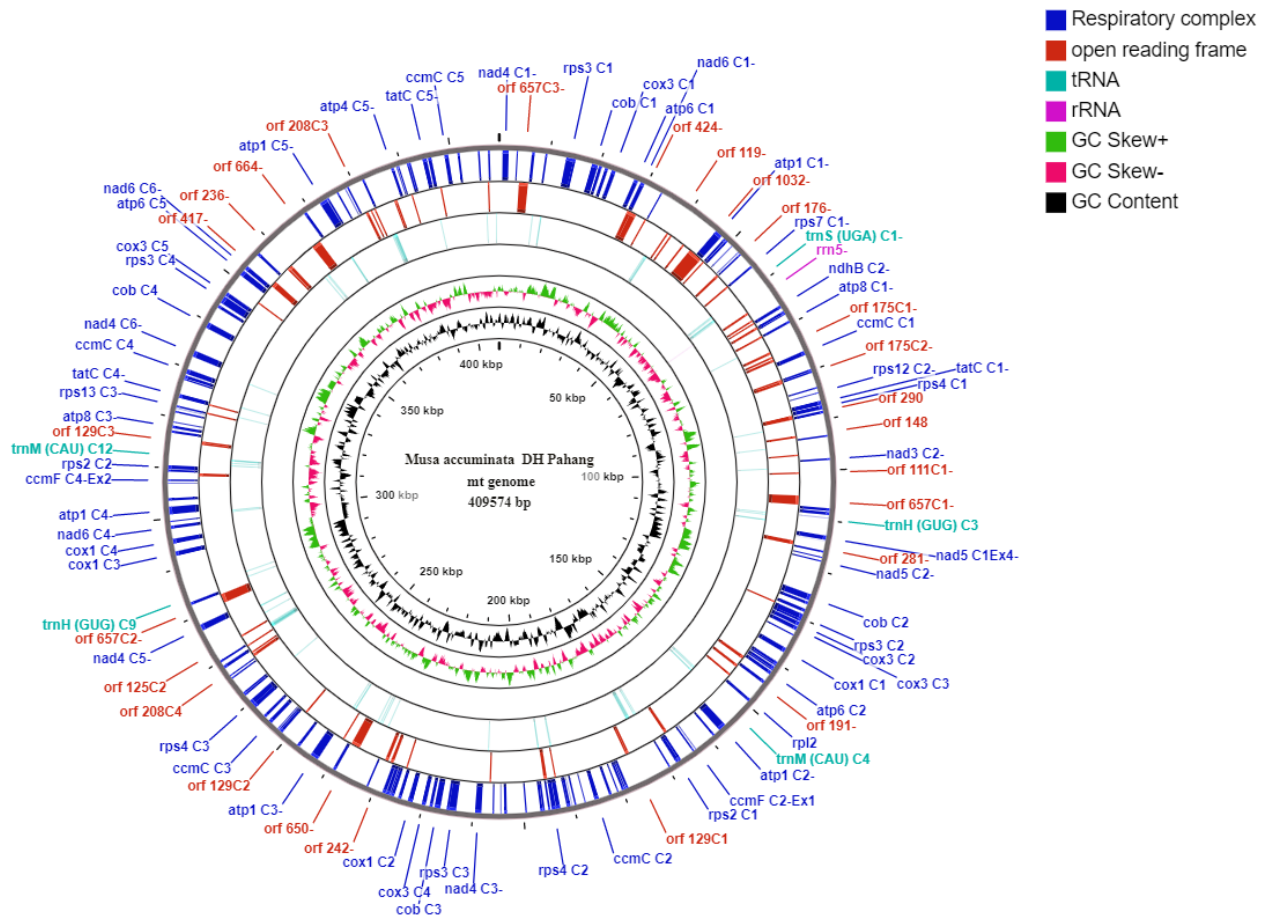


Fig 3

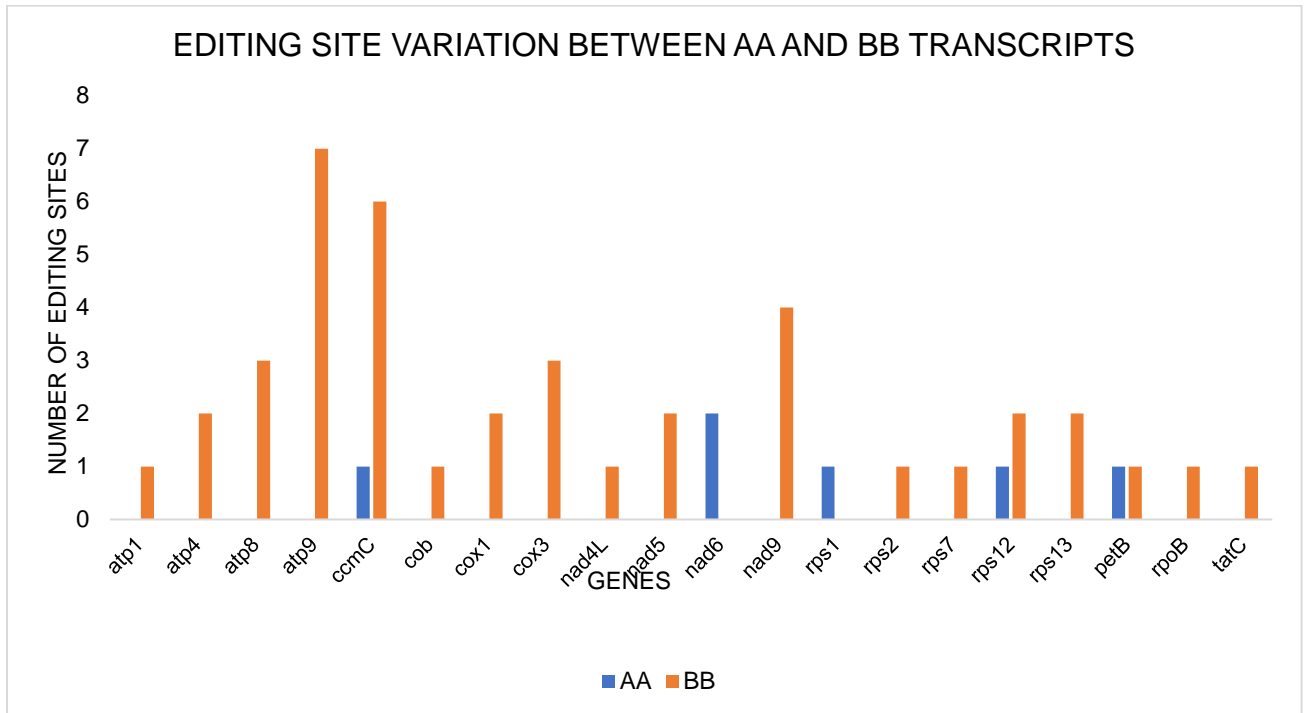


Table 1: General characters of the assembled *M. accuminata* mitochondrial genome

Features	Content
Genome size	409754 bp
GC content	45.3%
Chloroplast content	5.3%
Tandem repeats	25.34%
Respiratory complex	22
Ribosomal protein	10
Other protein	1
Pseudo genes	7
tRNAs	12
rRNA	2
ORF	43
Total no of genes	97

Table 2: List of mitochondrial genes annotated in *M.accuminata* DH Pahang subspecies malaccensis and their comparison with *M.balbisiana*

	GENE	FUNCTION DESCRIPTION	COPY NUMBER	START REGION	STOP REGION	INTRON	S/A	LENGTH OF AMINO ACID	PERCENTAGE IDENTITY WITH <i>M. Balbisiana</i>		
Complex I	nad1	NADH dehydrogenase subunit 1	1		83721	83425	83424-3705	A	182	100	
					3704	3453			190283-189807	139	97.85
			2 ^a	Ex-2	190364	190284	190283-189807			A	99
				Ex-1	189806	189468					
			3 ^b		199621	199325	327587-321855, 321773-321297		A	238	99.48
			4	Ex-3	327884	327588					
				Ex-2	321854	321774					
			4	Ex-1	321296	320958	327587-321855, 321773-321297		A	238	99.48
				1		21211					
			nad2	NADH dehydrogenase subunit 2	2		50853		50701	50700-49080	A
	Ex-1	49079			48612						
	nad3	NADH dehydrogenase subunit 3	1		80485	80129		A	118	97.46	
			2		98031	97666			121		
			3 ^b		196654	196199			151	96.61	
			4 ^b		262385	262029			118	97.03	
			5 ^b		391885	391529			118	100	
			6		407705	407340			121		
	nad4	NADH dehydrogenase subunit 4	1 ^a		1865	642		A	407	99.49	
			2		182740	182213			175	97.39	
			3 ^c		209572	208349			407	99.49	
			4		256304	255777			175	97.39	
			5 ^a		278161	276938			407	99.49	
			6 ^a		333740	332517			407	99.5	
			7		401734	401294			146	97.69	
	nad4L	NADH dehydrogenase subunit 4L	1		67863	67591		A	90	98.89	
			2		179515	179243			90	97.5	
			3		252249	251977			90	98.8	
4			389546	389274	90	98.86					
nad5	NADH dehydrogenase subunit 5	1		Ex-5	117453	117223	117222-116379, 116003-114085, 114061-113809, 112982-109546	A	506	97.97	
				Ex-4	116378	116004					
				Ex-3	114084	114062					
				Ex-2	113808	112983					
				Ex-1	109545	109480					
		2		118226	117960	88	96.2				
nad6	NADH dehydrogenase subunit 6	1		29584	28793		A	263	99.48		
		2		144886	144263			207	99.49		
		3		225041	224391			216	99.5		
		4		298572	297940			210	99.5		
		5		299518	299291			75	98.31		
		6		353359	352547			270	99.5		

	nad7	NADH dehydrogenase subunit 7	1	181036	180773	A	87	
			2	252574	252386		62	
			3	319150	318962		62	
			4	389157	388969		62	
	nad9	NADH dehydrogenase subunit 9	1	6445	7017	S	190	100
			2	107541	108113		190	100
			3	108295	108867		190	100
			4	238525	239097		190	100
			5	282403	282975		190	100
			6 ^a	368548	369132		194	100
Complex III	cob	apocytochrome b	1	18542	19729	S	395	96.92
			2	124906	126168		420	97.58
			3	216483	217655		390	97.3
			4	338767	339939		390	97.3
	petB	cytochrome b6	1	169455	169787	S	110	
Complex IV	cox1	cytochrome c oxidase subunit 1	1	136592	137512	S	306	100
			2	222078	223007		309	99.53
			3	292559	293446		295	100
			4	295023	295952		309	99.53
	cox2	cytochrome c oxidase subunit 2	1 ^b	49153	49567	S	138	92.52
			2 ^b	158716	159130		138	92.52
			3 ^c	246972	247562		196	76.69
			4 ^c	303507	304005		166	94.29
			5	Ex-1 373903 Ex-2 376151	374290 376494		374291-376150	243
	cox3	cytochrome c oxidase subunit 3	1	22616	23413	S	265	98.71
			2	130801	131598		265	98.71
			3	131761	132558		265	98.71
			4	219422	220219		265	98.87
5			346092	346889	265		98.71	
Complex V	atp1	ATP synthase F1 subunit alpha	1 ^b	47970	46459	A	503	97.62
			2	157460	155865		531	97.81
			3	244265	242628		545	97.63
			4	302324	300807		505	97.77
			5	373812	372253		519	97.62
	atp4	ATP synthase F0 subunit b	1	67556	67077	A	159	93.59
			2	179158	178688		156	92.81
			3	251946	251365		193	93.1
			4 ^b	317401	316922		159	92.86
			5 ^b	388421	387879		180	94.22
	atp6	ATP synthase F0 subunit a	1	27408	28214	S	268	99.58
			2	141973	142650		225	100
			3	226885	227562		225	100
			4	227649	228326		225	100
			5	351147	352238		363	99.21

	atp8	ATP synthase F0 subunit 8	1	69721	69236		A	161	89.17	
			2	180048	179563			161	89.1	
			3	318394	317828			188	88.54	
	atp9	ATP synthase F0 subunit c	1	33539	33264		A	91	97.47	
			2	149428	149180			82	98.55	
			3	231683	231342			113	97.5	
	atpB	ATP synthase CF1 subunit beta	SINGLE COPY	Ex-1	203295	203804	203805-204403, 204551-205841	S	310	
				Ex-2	204404	204550				
				Ex-3	205842	206116				
	Cytochrome C biogenesis	ccmC	ABC transporter subunit C	1	75910	76716		S	268	97.67
2				184907	185713		268		97.67	
3				254150	254905		251		97.49	
4				328882	329838		318		97.22	
5				398681	399481		266		96.07	
ccmF		haem lyase	1 ^a	Ex-1	52477	52618	52619-52942	S	161	92.5
				Ex-2	52943	53284				
			2 ^a	Ex-1	163333	163474	163475-163798		161	92.5
				Ex-2	163799	164140				
			3	Ex-1	265781	266081	266082-266405, 266670-268309		391	92.64
				Ex-2	266406	266669				
				Ex-3	268310	268920				
			4 ^a	Ex-1	306904	307045	307046-307369		161	92.5
				Ex-2	307370	307711				
			5 ^a	Ex-1	377955	378096	378095-378420		161	92.5
Ex-2		378421		378762						
Ribosomal protein		rp15	ribosomal protein L5	1	139798	140352		S	184	97.83
				2	140881	141435			184	97.83
				3	225930	226484			184	97.83
	4			356278	356832		184		97.83	
	rp16	ribosomal protein L16	1 ^a	14772	15206		S	144		
			2	130192	130626			144		
			3 ^a	214756	215190			144		
	rps2	ribosomal protein S2	1	167418	168059		S	213	97.35	
			2	309768	310499			243	97	
	rps3	ribosomal protein S3	1	13337	14800		S	487	93	
			2 ^a	129117	130220			367	95.38	
			3	213321	214784			487	93.02	
			4	344457	345950			497	94.05	
	rps4	ribosomal protein S4	1	87140	87952		S	270	98.77	
			2	193587	194399			270	98.77	
			3	259282	260283			333	97.2	
			4	395410	396060			216		
rps11	ribosomal protein S11	SINGLE COPY	190637	191053		S	138	91.79		
rps12	ribosomal	1	80084	79707		A	125	98.4		

	protein S12	2 ^a	84702	84520			60		
		3	196154	195777			125	98.39	
		4	261984	261607			125	98.15	
		5	391484	391107			125	98.25	
	rps13	ribosomal protein S13	1	93426	93076		A	116	100
			2	191393	191043			116	100
			3	322407	322057			116	100
			4	405701	405351			116	100
	rps14	ribosomal protein S14	1	140369	140671		S	100	96
			2	141452	141754			100	96
			3	226501	226803			100	96
	rps19	ribosomal protein S19	SINGLE COPY	133480	133758		S	92	92.39
Other protein	tatC	Sec-independent protein translocase component	1 ^a	87017	86226		A	263	96.88
			2 ^a	193508	192717			263	96.4
			3	259288	258521			255	96.39
			4 ^a	324749	323958			263	97
			5	395136	394279			285	94.78
Pseudo gene	ndhB	NADH-plastoquinone oxidoreductase subunit 2	1 ^a	19985	20590		S	201	95.81
			2 ^b	66365	65750		A	205	82.11
			3 ^a	126438	127043		S	201	95.81
			4 ^a	128009	128614			201	95.81
			5 ^a	218041	218646			201	95.81
			6 ^a	343310	343915			201	96.43
	rpoB	RNA polymerase subunit beta	SINGLE COPY ^a	187569	187775		S	68	75
	rpoC1	RNA polymerase b'-subunit	SINGLE COPY ^c	327981	328577		S	198	53.57
	rps1	ribosomal protein S1	1 ^a	88237	88773		S	178	95.33
			2 ^a	194684	195073			129	
	rpl2	ribosomal protein L2	SINGLE COPY ^b	148842	149288		S	148	95.8
	rps7	ribosomal protein S7	1 ^a	56555	56112		A	147	97.71
			2 ^a	167380	166937			147	97.44
3 ^a			270121	269678		147		97.58	
4 ^a			309660	309217		147		97.58	
5 ^a			382189	381746		147		97.5	
ftsH	cell division protein FTSH	SINGLE COPY ^c	9526	9149		A	125		

List of respiratory complex genes, ribosomal protein genes, other proteins and pseudo genes along with their co-ordinates, copy number, Exon and intron numbers and co-ordinates, presence in (+S/-A) strand, Protein length and percentage identity with *M.balbisiana* proteins are given. a-no start codon; b-no stop codon; c-no start and stop codon.

Table 3: *M.accuminata* mitochondrial gene content comparison with other model organisms used as reference in this study

	Genes	DH Pahang	Maize	Rice	Arabidopsis	Sugarbeet
Complex I	nad1	+	+, 2 exon 1	+	+	+
	nad2	+	+, 2 exons 4,5	+, 2 exons 3-5	+	+
	nad3	+	+	+	+	+
	nad4	+	+	+	+	+
	nad4L	+	+	+	+	+
	nad5	+	+	+, 2 exon 1	+	+
	nad6	+	+	+	+	+
	nad7	+	+	+	+	+
	nad9	+	+	+	+	+
	ndhB	Ψ	-	-	-	-
Complex II	sdh2	-	-	-	-	-
	sdh3	-	-	-	-	-
	sdh4	-	-	-	Ψ	-
Complex III	cob	+	+	+	+	+
	petB	+	-	-	-	-
Complex IV	cox1	+	+	+	+	+
	cox2	+	+	+	+	+
	cox3	+	+	+	+	+
Complex V	atp1	+	+	+	+	+
	atp4	+	+	+	+	+
	atp6	+	+	+	+	+
	atp8	+	+	+	+	+
	atp9	+	+	+	+	+
	atpB	+	-	-	-	-
Cytochrome C Biogenesis	ccmB	-	+	+	+	+
	ccmC	+	+	+	+	Ψ
	ccmF	+	+	+	-	+
	ccmFN1	-	-	-	+	-
	ccmFN2	-	-	-	+	-
	ccmFC	-	+	+	+	+
	ccmFC1	-	-	-	-	-
	ccmFC2	-	-	-	-	-
Ribosomal protein	rps1	Ψ	+	+	-	-
	rps2	+	+	+	-	-
	rps2B	-	+	-	-	-
	rps3	+	+	+	+	+
	rps4	+	+	+	+	+

	rps7	Ψ	+	+	+	+
	rps8	-	-	-	-	-
	rps10	-	-	-	-	-
	rps11	+	-	Ψ	-	-
	rps12	+	+	+	+	+
	rps13	+	+	+	-	+
	rps14	+	-	Ψ	Ψ	-
	rps15	-	-	-	-	-
	rps16	-	-	-	-	-
	rps19	+	-	+	Ψ	-
	rpl2	Ψ	-	+	+	+
	rpl5	+	-	+	-	-
	rpl6	-	-	-	+	-
	rpl8	-	-	-	-	-
	rpl16	+	+	-	+	-
Other proteins	ftsH	Ψ	-	-	-	-
	mat-r	-	+	+	+	+
	mttB	-	+	+	+	+
	tatC	+	-	-	+	-
rRNA	rrn5	+	+	+	+	+
	rrn18	-	+	+	+	+
	rrn26	+	+	+	+	+
RNA polymerase	rpoB	Ψ	-	-	-	-
	rpoC1	Ψ	-	-	-	-

List of core plant mitochondrial genes and their presence (+), absence (-) and pseudogenes (Ψ) of banana in comparison with other model plants used in this study are given.

Table 4: List of editing sites found in *M.accuminata* transcripts in comparison with *M.balbisiana*

Genes	Gene editing position	Codon change		Amino acid change		Intron/exon
		From	To	From	To	
AA editing sites						
ccmC	76	CGC	UGC	R	C	EXON
nad6	289	CTT	UTT	L	F	EXON
	306	TTC	TTU	F	F	EXON
rps1	179	CCA	CUA	P	L	EXON
rps12	115	CGG	UGG	R	W	EXON
petB	166	CTC	TTC	L	F	EXON
BB editing sites						
atp1	429	UCC	UCU	S	S	EXON
atp4	395	UCA	UUA	S	L	EXON
	416	ACC	AUC	T	I	EXON
atp8	47	UCA	UUA	S	L	EXON
	58	CUC	UUC	L	F	EXON
	123	CUC	CUU	L	L	EXON
atp9	504	UCC	UCU	S	S	EXON
	514	CGA	UGA	R	STOP	EXON

	518	UCA	UUA	S	L	EXON
	604	CGU	UGU	R	C	EXON
	627	UUC	UUU	F	F	EXON
	634	CAU	UAU	H	Y	EXON
	645	UCC	UCU	S	S	EXON
ccmC	497	UCU	UUU	S	F	EXON
	499	CCU	UCU	P	S	EXON
	568	CCU	UCU	P	S	EXON
	575	CCC	CUC	P	L	EXON
	608	CCC	CUC	P	L	EXON
	614	UCC	UUC	S	F	EXON
cob	923	GCC	GUC	A	V	EXON
cox1	1982	UCU	UUU	S	F	EXON
	2053	CCA	UCA	P	S	EXON
cox3	2227	CUU	UUU	L	F	EXON
	2366	UCG	UUG	S	L	EXON
	2376	CCC	CCU	P	P	EXON
nad4L	6315	AGC	AGU	S	S	INTRON
nad5	1363	CUC	UUC	L	F	EXON
	1441	CAA	UAA	Q	STOP	INTRON
nad9	356	UCU	UUU	S	F	EXON
	368	UCC	UUC	S	F	EXON
	398	UCA	UUA	S	L	EXON
	406	CAU	UAU	H	Y	EXON
rps 2	60	GUC	GUU	V	V	EXON
rps 7	215	CCA	CUA	P	L	EXON
rps 12	115	CGG	UGG	R	W	EXON
	156	UUC	UUU	F	F	EXON
rps 13	342	CUC	CUU	L	L	EXON
	373	CGC	UGC	R	C	EXON
petB	166	CTC	TTC	L	F	EXON
rpoB	991	CTA	TTA	L	L	EXON
tatC	235	CCG	TCG	P	S	EXON

List of genes and their position of editing sites with corresponding amino acid changes and their position in the gene (EXON/INTRON) are given.