

The Evolution Force of Genome Reduction in Carnivorous Plants

¹Ming-Yue Ma, ²Tao Tan, ¹Xin-Xin Yin, ¹Jia-Ling Zou and ¹Kun-Xian Shu

¹Chongqing Key Laboratory of Big Data for Bio Intelligence,

Chongqing University of Posts and Telecommunications, Nan'an District, 400065, Chongqing, China

²Chongqing Medical University, No.1 Yixueyuan Road, Yuzhong District, 400016, Chongqing, China

Article history

Received: 25-04-2018

Revised: 11-05-2018

Accepted: 24-05-2018

Corresponding Author:

Kun-Xian Shu

Chongqing Key Laboratory of

Big Data for Bio Intelligence,

Chongqing University of Posts

and Telecommunications,

Nan'an District, 400065,

Chongqing, China

Email: sukx@cqupt.edu.cn

Abstract: The introns are widely present in the genome of eukaryotes and the distribution of intron varies greatly among different organisms or different genes. Generally, introns loss is an important way for uneven distribution of intron during genome evolution. In this study, two closely related carnivorous plants (*Genlisea aurea* and *Utricularia gibba*) were chosen, their genome were relatively integrity and high quality, especially, the large difference in genome size between them. We detected intron loss events, then investigated the relationship between the genome size, intron density, intron loss and the mutation rate in the carnivorous plants. Finally, a total of 752 and 124 intron loss positions were identified in *G. aurea* and *U. gibba*, respectively. In carnivorous plants, we found that the region around lost site had high mutation rate, the genes of intron loss had high mutation rate. Besides, for the species with more intron losses, the genome size was smaller and the mutation rate was higher. Thus, we propose that the mutation rate was positively correlated with intron losses, but negatively correlated with intron number and genome size. These could be explained by the selection to minimize mutational hazards.

Keywords: Genome Size, Intron Loss, Mutational-Hazard Hypothesis, *Genlisea aurea*, *Utricularia gibba*

Introduction

Generally, the introns are ubiquitous in the eukaryotic genomes. Nonetheless the distribution of introns among different species and different genes are quite different (Ma *et al.*, 2015b), studying the evolutionary dynamics of introns may provide clues to the evolution of eukaryotic genome. However, there are still controversies on the evolution forces that cause the present distribution of introns, which can be divided into neutral evolution (also known as genetic drift) and selective evolution. In addition, there is another way also could change the frequency of alleles in the population, which is the mutation bias (Duret, 2008; Weber *et al.*, 2014; Behringer and Hall, 2016). It means that, in the species, the probability of the beneficial mutation being fixed is higher than that of the neutral mutation, on the contrary, the probability of the disadvantage mutation being fixed is lower than that of the neutral mutation. In recent years, a popular view about intron evolution is the mutational-hazard hypothesis proposed by Lynch. He believes that non-coding sequences such as introns are

harmful to organisms, but the power of these deleterious variations is very weak and is therefore considered weakly harmful. For the species with small population size (N_e), the effective selection efficiency of weakly deleterious mutations is significantly reduced. The hypothesis also gives a reasonable explanation towards the evolutionary model between the genome size and the complexity of the gene structure (Lynch, 2006; Lan and Pritchard, 2016). Lynch emphasizes the role of mutation rate (μ) and N_e during the evolution of non-coding sequences, the genome with high $N_e\mu$ is structurally more compact than the genome with low $N_e\mu$. The species of large N_e are more capable of eliminating harmful mutations and effectively deleting introns that are harmful to the organism. On the contrary, in the species with small N_e , the genetic drift plays a major role, such as eukaryotes, which cannot efficiently eliminate deleterious introns sequences and thus retain relatively high number of introns. In addition, multicellular organisms not only have higher mutation rate and produce more mutations than unicellular organisms, but also their genetic structure

are more complex and are more likely to be mutated target sites during evolutionary processes. By comparing the nuclear genomes and organellar genomes, Lynch found a large number of introns accumulation in the nuclear genome and he believes that it is due to differences in the mutation rates between the two genomic groups. Meanwhile, he also found that the complexity of the organelle genome was significantly reduced, supporting the mutational-hazard hypothesis (Lynch *et al.*, 2006; Smith, 2016).

However, the number of introns and the size of genome were accumulated during the long-term evolution process, the mutation rate and the size of effective population were the current condition yet. Therefore, there was still some controversy about the mutational-hazard hypothesis (Whitney and Garland 2010; Cohen *et al.*, 2012). For example, Mohlhenrich analyzed the non-coding genes of salamanders and frogs, he concluded that the mutational hazard did not play a mainly role in the genomic gigantism of salamanders (Mohlhenrich and Mueller, 2016). But there are still some studies that support the mutational-hazard hypothesis. The first study, Smith found that, in *Chlamydomonas reinhardtii* plastid DNA (ptDNA), the nucleotide polymorphisms of the silent sites were lower than nuclear DNA but higher than mitochondrial DNA (mtDNA). By comparing the ptDNA of *C. reinhardtii* and land plants, the number of silent sites in the former genome was found to be about 4 times more than land plant's genomes and the genomic structure of the former was also more compact. The silent-site ptDNA diversity was conformed to the mutational-hazard hypothesis (Smith and Lee, 2009). The second study, Yang have systematically analyzed the genome of *Arabidopsis thaliana* from intron, gene and genome. It was found that the introns with high mutation rate were more likely to be lost, the genes with high mutation rate had higher frequency of intron deletions and the species that lost more introns have a higher genomic mutation rate. These research results strongly support the hypothesis that mutation hazard is the choice driving force of intron loss (Yang *et al.*, 2013).

In this study, we selected the carnivorous plants to investigate whether the driving force in the evolution of the genome was mutation hazard. The only two species of carnivorous plants (*Genlisea aurea* (Leushkin *et al.*,

2013) and *Utricularia gibba* (Ibarra-Laclette *et al.*, 2013) that genome sequencing and annotation results were relatively integrity. The *Genlisea* and *Utricularia* plants have similar habits that usually in a relatively stable environment, like underground or under water. They often feed on microscopic and perform digestive functions in closed traps. Thus, they are less vulnerable to the environmental constraints in exposed environments (Butts *et al.*, 2016). The genome sequencing results showed that the genome size difference between *G. aurea* and *U. gibba* was more than one time, 43 Mb vs 101 Mb. This provides an opportunity to assess the role of the mutation hazard hypothesis in the genome evolution of the carnivorous plants.

Materials and Methods

Data Sets

We downloaded the genomes sequences and annotation files of *G. aurea*, *U. gibba*, *Mimulus guttatus*, *Solanum tuberosum*, *Arabidopsis thaliana* and *Vitis vinifera* from CoGe (<https://genomeevolution.org/coge/>). The genome version and other information of the six species were shown on the Table1.

Identification of Conserved Introns, Conserved Genes and Lost Introns

Using BlastN alignment program to identify the orthologous genes among *G. aurea*, *U. gibba*, *M. guttatus* and the outgroups, by comparing the CDS of the genes. And the orthologous sequences were aligned using ClustalW (version 2.1) and MUSCLE (version 3.8.31) multiple sequence alignment programs, we retained the alignment results of higher quality (like similarity). Then the intron positions of *G. aurea* and *U. gibba* were located to the sequence alignment results, determining the state of the intron (absence or presence) on the orthologous sequences. In this step, we required that the similarity of the 30 bp coding bases upstream and downstream of the target intron site was no less than 0.5. Finally, the conserved or discordant sites of introns were detected. If a pair of orthologous genes only had conserved introns and no one discordant intron, these genes were called conservative genes in this article.

Table 1: The number of proteins belonging to four structural classes in the datasets

Species	Order	Family	Genome ID	Version
<i>Genlisea aurea</i>	Lamiales	Lentibulariaceae	25393	v1
<i>Utricularia gibba</i>	Lamiales	Lentibulariaceae	29027	vPNAS_May15_2017
<i>Mimulus guttatus</i>	Lamiales	Phrymaceae	22665	v2
<i>Solanum tuberosum</i>	Solanales	Solanaceae	24771	PGSC_DM_v4.03
<i>Arabidopsis thaliana</i>	Brassicales	Brassicaceae	25869	vTAIR10.26
<i>Vitis vinifera</i>	Vitales	Vitaceae	19990	v12x

For the discordant intron positions, intron deletion events were determined based on the presence or absence of the orthologous sequence of the closely related species and the Dollo parsimony (Dollo and Polymorphism Parsimony Program, phylip version 3.6) (Felsenstein, 2004).

Calculation of the Substitution Rates of Coding Sequences of Gene

Take the intron losses of *G. aurea* introns as an example. First, found out the gene with intron loss (or conservative gene) and the orthologous gene in *M. guttatus*. Second, using ClustalW (version 2.1) and MUSCLE (version 3.8.31) to align protein sequences of the pair of orthologous genes. Third, converting the protein alignments into codon sequences one by one and constructing the new alignments of the coding sequences. Fourth, using Gblocks program (version 0.91b) (Talavera and Castresana, 2007) to filter the new coding sequence alignments (-t = c -p = t) and convert the alignment results to phylip format. Lastly, the codeml.ctl program (PAML, version 4.7) (Yang, 2007) was used to calculate the synonymous and nonsynonymous substitution rate of the group of orthologous genes.

Calculation of the Substitution Rates Around Lost/Conserved Intron Positions

Take the intron losses of *U. gibba* introns as an example. At first, we found out the orthologous positions of the target introns (the loss or conservative positions of *U. gibba*) in *M. guttatus*. Then, cutting 50 complete codons (150 bp coding sequences) upstream and downstream of the target intron positions of *U. gibba* and *M. guttatus*, ClustalW (version 2.1) and MUSCLE (version 3.8.31) were used to align the coding sequences. Next, similar to the 2.3 of "Material and methods" part, we used Gblocks (Talavera and Castresana, 2007) to filter alignments and

the codeml.ctl program (Yang, 2007) to calculate mutation rate of the target coding sequences.

Results and Discussion

Intron Consensus Sequence in the Carnivorous Plants

We removed the introns with sequence length less than 30 bp and detected 62,377 and 88,915 satisfied introns in the annotation results of *G. aurea* and *U. gibba*, respectively. Then, using WebLogo (version 3.6.0) (Crooks *et al.*, 2004) to analyze the conservatism of sequences, including 10 bp intronic sequences of 5' end of introns, 20 bp intronic sequences of 3' end of introns and 2 bp exonic sequences of the upstream and downstream of introns. Figure 1 shown that 5' splicing signal, the polypyrimidine tract and 3' splicing signal of introns from left to right. We found that the consensus signals of the introns of the carnivorous plants were similar to those of eukaryotes, like land plants, metazoan and mammals (Irimia and Roy, 2008; Ma *et al.*, 2015a; 2016).

Conserved and Discordant Intron Sites

We identified 14,776 pairs of orthologous genes between the *G. aurea* and *U. gibba*, covering more than 80% of the genes in the *G. aurea* genome. By comparing the alignment results of orthologous genes, we obtained 4,187 discordant intron positions. Using *Mimulus guttatus*, *Solanum tuberosum*, *Arabidopsis thaliana* and *Vitis vinifera* as outgroup species to identify the intron loss cases, 937 and 175 intron loss sites were found in *G. aurea* and *U. gibba*, respectively. Since *M. guttatus* was used as a control species for comparative analysis in this study, only retained the discordant positions when *M. guttatus* existed the introns at the target sites (Fig. 2 (1)-(3)). In addition, there is at least one or more conserved introns in the genes of lost intron in *G. aurea* or *U. gibba*.

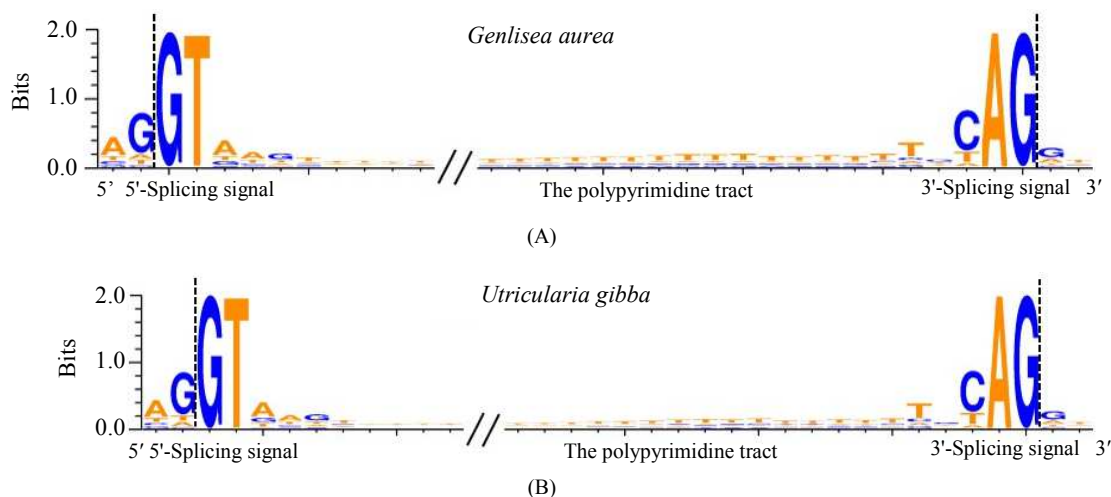


Fig. 1: Consensus sequence of introns in *Genlisea aurea* and *Utricularia gibba*. From left to right, there were 5' splicing signal, the polypyrimidine tract and 3' splicing signal of introns

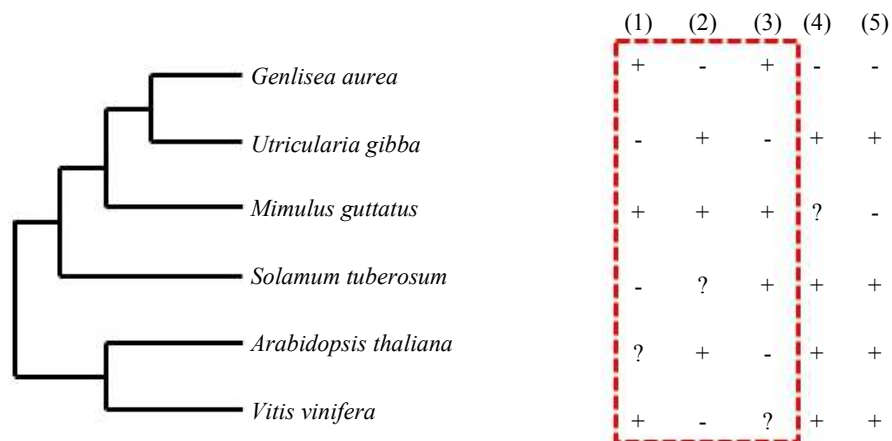


Fig. 2: Phylogenetic tree used to identify intron loss in *Genlisea aurea* and *Utricularia gibba*. In this figure, the (1) and (3) displayed as intron losses in *U. gibba*; (2), (4) and (5) displayed as intron losses in *G. aurea*. In this study, we only kept the case of intron loss like (1) ~ (3)

Finally, we obtained 752 lost intron positions in 601 genes in *G. aurea* and 124 lost intron positions in 118 genes in *U. gibba*. By comparing the orthologous gene sequences of *G. aurea*, *U. gibba* and the control species (*M. guttatus*), we determined 21,343 conservative intron positions and 4,582 conserved genes for next comparative analysis.

High Mutation Rates around Intron Lost Sites

A comparison of the ratio of the number of synonymous substitutions per synonymous site (synonymous substitution ratio, d_S) and the nonsynonymous substitution ratio (d_N) between the flanking coding sequences of intron loss sites and intron conserved sites. In this study, we analyzed the 150 bp coding sequence based on 50 amino acids of the both sides of the target intron positions.

In *G. aurea*, there are 706 sites of lost intron and 20,008 sites of conserved intron that met the analysis conditions. The PAML software was used to calculate the mutation ratio of the target intron sites in *G. aurea* and the corresponding region sequence in *M. guttatus*. Then the Mann-Whitney U test was used for statistical analysis. We discovered that the d_S values of the flanking coding sequences of the lost intron positions were significantly higher than that of conserved intron positions (P -value = 6.96×10^{-52}) in *G. aurea*. Fig. 3A shown that comparison of d_S values of the target coding sequences of all lost introns sites (shown as red line) and that of all conserved introns sites (show as blue line) in each percentage point. In addition, the values of d_N/d_S of the flanking coding sequences of the lost intron positions were significantly lower than that of conserved intron positions, the P -value was 2.23×10^{-22} (Mann-Whitney U test, Fig. 3B).

Using the same method, we investigated the substitution ratios around lost and conserved of intron positions in *U. gibba*. After filtering the length of target sequences, there were 120 intron loss sites and 20,499 conservative intron positions could be capable for statistics.

Through calculating and comparing the lost and conserved introns of *U. gibba* and the corresponding regions in *M. guttatus*, we found that: (1) the d_S values of the flanking coding sequences of the lost intron sites were significantly higher than that of conserved intron sites, the median values were 2.572 and 1.564 respectively and the P -value was 7.78×10^{-6} of Mann-Whitney U test. (2) But, there was no significantly difference between the d_N/d_S values of the flanking coding sequences of the lost intron sites and that of conserved intron sites, the median values were 0.058 and 0.069 respectively and the P -value was 0.526 (Mann-Whitney U test).

High Mutation Rates of the Intron Lost Genes

By comparing the orthologous gene alignment sequences and the conserved positions of *G. aurea*, *U. gibba* and *M. guttatus*, a total of 4,582 conserved genes were obtained in this study.

In *G. aurea*, we identified 752 cases of intron losses and distributed on 601 genes. Compared the mutation ratio of the coding sequences of the conserved genes ($N = 4,582$) and the genes of lost introns ($N = 601$). Using PAML software to calculate the synonymous substitution ratios and the non-synonymous substitution ratios of the coding sequences of the target genes in *G. aurea* and that of the orthologous genes in *M. guttatus*. The statistical analysis results shown that the d_S values of coding sequences of the genes with intron loss were significantly higher than the d_S values of coding sequences of conserved genes (Fig. 4A), the median values were 1.850 and 1.732 respectively, the P -value was 0.015 (Mann-Whitney U test). Moreover, the Mann-Whitney U test revealed d_N/d_S values of coding sequences of the genes with intron loss were significantly lower than that of conserved genes, (Fig. 4B) the median values were 0.084 and 0.093 respectively, the P -value was 5.47×10^{-6} . Conforming to the mutation-hazard supposition.

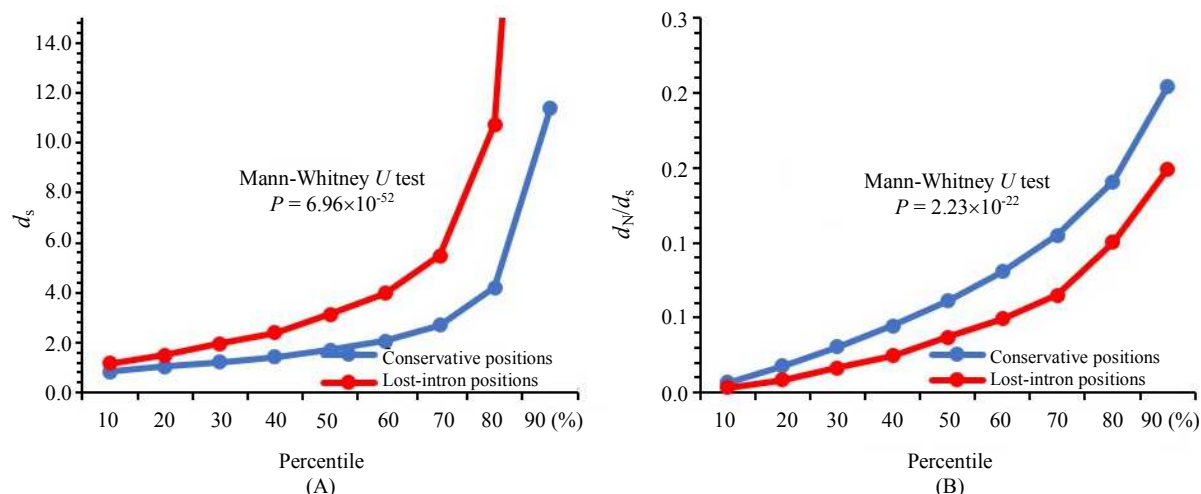


Fig. 3: The mutation rate of coding sequences around lost introns and conserved introns in *Genlisea aurea*. The X-axis represents the percentile of the d_S or d_N/d_S and the Y-axis represents the values of the d_S or d_N/d_S . The blue line represents as the d_S of coding sequences around conserved intron positions ($N = 20,008$). The red line represents as the d_N/d_S of coding sequences around lost-intron positions ($N = 706$)

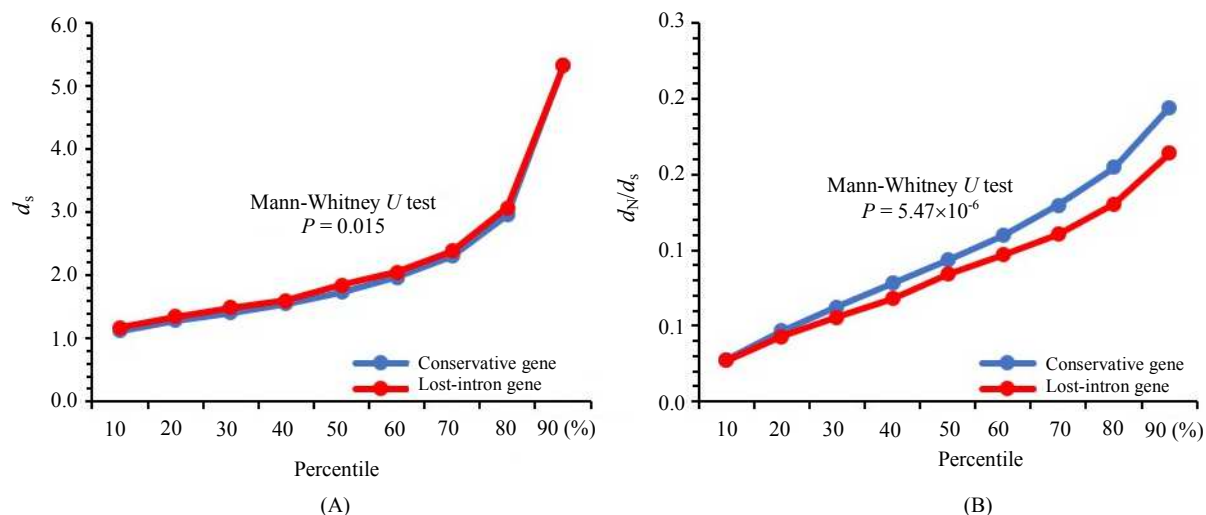


Fig. 4: The mutation rate of lost genes and conserved genes in *Genlisea aurea*. The X-axis represents the percentile of the d_S or d_N/d_S and the Y-axis represents the values of the d_S or d_N/d_S . The blue line represents as the d_S of coding sequences of conservative genes ($N = 4,582$). The red line represents as the d_N/d_S of coding sequences of lost-intron genes ($N = 601$)

In a similar way, the mutation rate of the coding sequence of the genes with the intron deletions ($N = 118$) and that of the conserved genes ($N = 4,582$) in *U. gibba* was investigated. At first, we calculated the mutation rate between the conserved genes or variational genes in the *U. gibba* and the corresponding orthologous genes in the control species *M. guttatus*. Next, the Mann-Whitney *U* test was used to analyze the difference in the mutation rate between the genes with intron loss and the conserved genes. The results shown that there were no significant differences of the d_S nor of the d_N/d_S between the coding sequences of conserved genes and

the coding sequence of variant genes. The *P*-values were 0.791 and 0.479, respectively.

High Mutation Rate in *Genlisea Aurea*

In this study, the 752 and 124 cases of intron loss were found in *G. aurea* and *U. gibba*, respectively. We found a significant difference in the number of introns lost in the two species, the *P*-value was much less than 0.05 (Table 2). The Pearson's chi-squared test indicated that the number of introns lost in the *G. aurea* genome was significantly higher than that in the *U. gibba*, that is, there are more introns lost in *G. aurea*.

Using *M. guttatus* as the reference species, we explored the differences in the mutation rates between *G. aurea* and *U. gibba* in two aspects.

In the first aspect, we examined the flanking coding sequences of the conserved intron sites among *G. aurea*, *U. gibba* and *M. guttatus*. Computing the synonymous substitution ratios and the nonsynonymous substitution ratios of the upstream and downstream coding sequences

(150 bp) of conserved intron positions of two groups (*G. aurea* vs *M. guttatus* and *U. gibba* vs *M. guttatus*). Then, comparing the differences between the two groups. The number of conserved intron sites satisfying the comparison conditions was 19,847. Using the PAML software to calculate the substitution ratios of the coding sequences of the conserved orthologous genes and using the Wilcoxon signed-rank test for statistical analysis.

Table 2: *Genlisea aurea* has lost more introns than *Utricularia gibba*

Intron type	<i>Genlisea aurea</i>	<i>Utricularia gibba</i>	χ^2 Test
Loss	752	124	5.79×10^{-98}
Conserved	21,343	21,343	

Table 3: The size of coding sequences and introns sequences in genomes

Species	Genome size	Coding sequence	Intron sequence	Intron/Coding ¹
<i>Genlisea aurea</i>	43 Mb	18 Mb	8.3 Mb	0.46
<i>Utricularia gibba</i>	101 Mb	28 Mb	16.7 Mb	0.60
<i>Mimulus guttatus</i>	313 Mb	67 Mb	97.0 Mb	1.45
<i>Solanum tuberosum</i>	773 Mb	38 Mb	60.1 Mb	1.58
<i>Arabidopsis thaliana</i>	120 Mb	48 Mb	25.4 Mb	0.53
<i>Vitis vinifera</i>	486 Mb	31 Mb	126.4 Mb	4.08

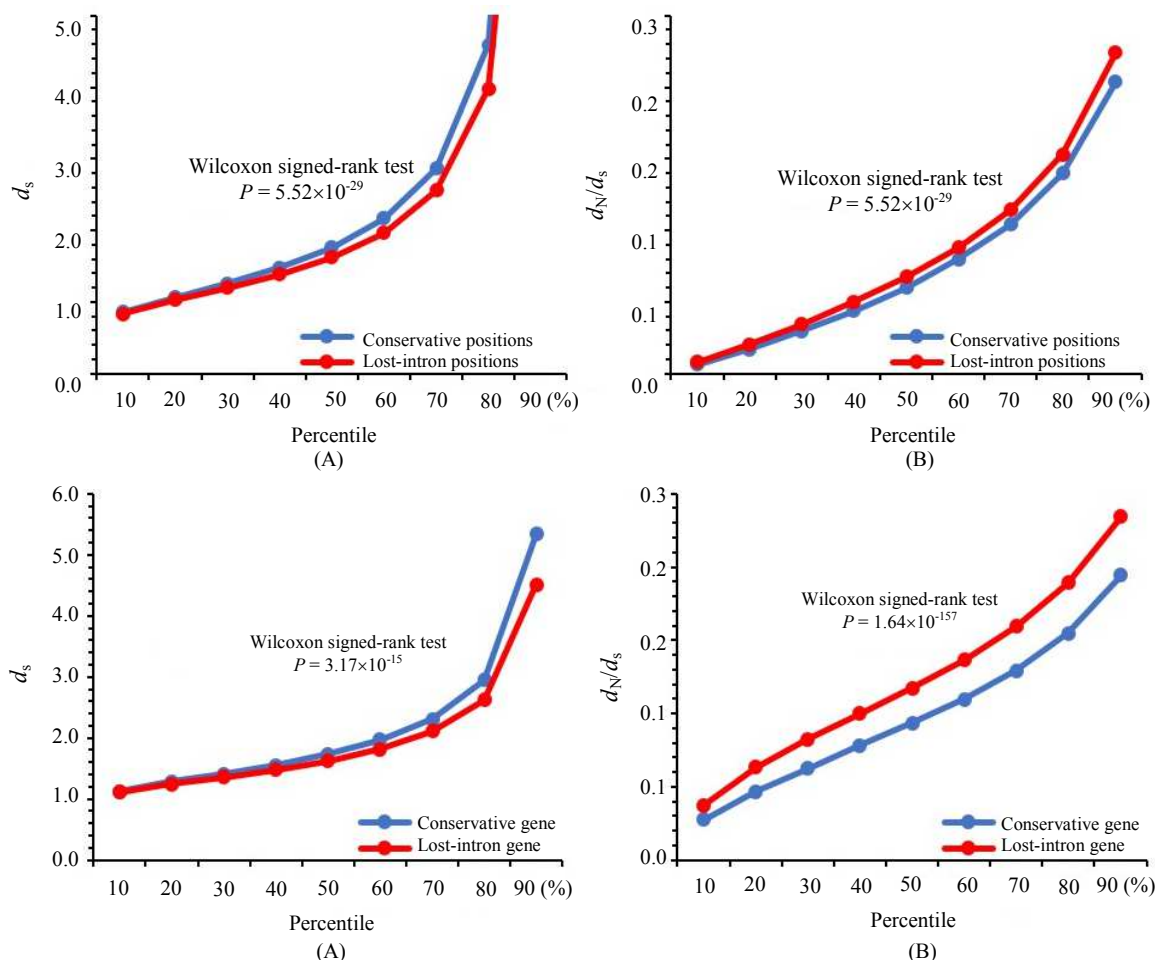


Fig. 5: The mutation rate in *Genlisea aurea* and *Utricularia gibba*. The X-axis represents the percentile of the d_s or d_N/d_S and the Y-axis represents the values of the d_s or d_N/d_S . (A) and (B) shown the mutation rate of coding sequences around conservative introns ($N = 19,847$) in *G. aurea* and *U. gibba*. (C) and (D) shown the conservative genes ($N = 4,582$)

The Wilcoxon signed-rank test results shown that the d_s values of flanking coding sequences of conserved intron positions in *G. aurea* were significantly higher than that of orthologous sequences in *U. gibba* (Fig. 5A), the median values were 1.732 and 1.626 respectively. In contrast, the d_N/d_S values of flanking coding sequences of conserved intron positions in *G. aurea* were significantly lower than that of orthologous sequences in *U. gibba* (Fig. 5B), the median values were 0.093 and 0.117 respectively.

In the second aspect, we examined the coding sequences of the conserved orthologous genes of the three species, the definition of conservative genes refers to the second part of "Methods" and identified 4,582 pairs of conserved genes. Also using the Wilcoxon signed-rank test to analysis, the results shown that the d_s values of coding sequences of conserved genes in *G. aurea* were significantly higher than that in *U. gibba* (Fig. 5C), the median values were 1.763 and 1.624, respectively. In contrast, the d_N/d_S values of coding sequences of conserved genes in *G. aurea* were significantly lower than that in *U. gibba* (Fig. 5D), the median values were 0.060 and 0.068, respectively.

Conclusion

The genome sizes of carnivorous plants (*G. aurea* and *U. gibba*) and *A. thaliana* were smaller than those of the other closely related species, like *M. guttatus*, *S. tuberosum* and *V. vinifera*. But, their coding sequences account for the genome sequences were higher than the other closely related species. That is, the ratios of the non-coding sequences to the whole genome size had reduced. In addition, it was found that the ratios of intron sequences to coding sequences were also lower (Table 3). In this study, we investigated the relationship between the number of intron loss, the upstream and downstream coding sequences of target positions, the coding sequence of the orthologous genes and the mutation rate, for the two carnivorous plants with relatively small genome size. Here, the d_s values represented the mutation rate and the d_N/d_S values represented the selected pressure of the coding sequences.

According to the phylogenetic relationship and genome sequences of the control species *M. guttatus* and the outgroups (*S. tuberosum*, *A. thaliana* and *V. vinifera*), we found out 752 and 124 cases of intron losses in *G. aurea* and *U. gibba*, respectively. And the *G. aurea* had lost more introns than *U. gibba* (Table 2).

In this study, we analyzed the mutation rate intron loss of two carnivorous plants and detected that: (1) The high mutation rate and under strong selection force around intron loss positions. The d_s around lost sites were higher than conservative sites in both two species. And the d_N/d_S around lost positions were lower than conservative positions in *G. aurea*, but there was no

significant difference in *U. gibba*. (2) The high mutation rate and under strong selection of genes with intron losses. The d_s of variant genes were higher than conservative genes and the d_N/d_S were lower in *G. aurea*. There had similar trends of the two types of genes in *U. gibba*, but the difference was not significant. (3) The high mutation rate and under strong selection of the species (*G. aurea*) that lost more introns. The d_s of orthologous genes of *G. aurea* were higher than that of *U. gibba* and the d_N/d_S were lower.

Compared with *U. gibba*, *G. aurea* had smaller genome size and fewer introns, but it lost more introns, had higher mutation rate and under relatively stronger purify selective. The mutation rate was positively correlated with intron losses and negatively correlated with intron number and genome size. These results were supported mutational-hazard hypothesis, the selection to minimize mutational risks might be the selective force on intron loss and possibly genome reduction.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant No. 31701093); the Chongqing Research Program of Basic Research and Frontier Technology (Grant No. cstc2017jcyjAX0200); the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJ1704094); and the Scientific Research Foundation of Chongqing University of Posts and Telecommunications (Grant No. A2016-87).

Author's Contributions

Ming-Yue Ma and Tao Tan: Designed the analyses and wrote the manuscript.

Xin-Xin Yin, Jia-Ling Zou and Kun-Xian Shu: Proofed the data.

All authors read and approved the final manuscript.

Ethics

The authors declare their responsibility for any ethical issues that may arise after the publication of this manuscript.

Conflict of Interest

The authors declare that they have no competing interests. The corresponding author affirms that all of the authors have read and approved the manuscript.

References

Behringer, M.G. and D.W. Hall, 2016. Selection on position of nonsense codons in introns. *Genetics*, 204: 1239-1248. DOI: 10.1534/genetics.116.189894

- Butts, C.T., J.C. Bierma and R.W. Martin, 2016. Novel proteases from the genome of the carnivorous plant *Drosera capensis*: Structural prediction and comparative analysis. *Proteins*, 84: 1517-1533. DOI: 10.1002/prot.25095
- Cohen, N.E., R. Shen and L. Carmel, 2012. The role of reverse transcriptase in intron gain and loss mechanisms. *Molecular Biol. Evolut.*, 29: 179-186. DOI: 10.1093/molbev/msr192
- Crooks, G.E., G. Hon, J.M. Chandonia and S.E. Brenner, 2004. WebLogo: A sequence logo generator. *Genome Res.*, 14: 1188-1190. DOI: 10.1101/gr.849004
- Duret, L., 2008. Neutral theory: The null hypothesis of molecular evolution. *Nat. Educ.*, 1: 218-218.
- Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Seattle: Department of Genome Science University Washington.
- Ibarra-Laclette, E., E. Lyons, G. Hernandez-Guzman, C.A. Perez-Torres and L. Chang *et al.*, 2013. Architecture and evolution of a minute plant genome. *Nature*, 498: 94-98. DOI: 10.1038/nature12132
- Irimia, M. and S.W. Roy, 2008. Evolutionary convergence on highly-conserved 3' intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *Plos Genet.*, 4: 1-14. DOI: 10.1371/journal.pgen.1000148
- Lan, X. and J.K. Pritchard, 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*, 352: 1009-1013. DOI: 10.1101/019166
- Leushkin, E., R. Sutormin, E. Nabieva, A. Penin and A. Kondrashov *et al.*, 2013. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genom.*, 14: 1-11. DOI: 10.1186/1471-2164-14-476
- Lynch, M., 2006. The origins of eukaryotic gene structure. *Molecular Biol. Evolut.*, 23: 450-468. DOI: 10.1093/molbev/msj050
- Lynch, M., B. Koskella and S. Schaack, 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science*, 311: 1727-1730. DOI: 10.1126/science.1118884
- Ma, M.Y., X.R. Che, A. Porceddu and D.K. Niu, 2015a. Evaluation of the mechanisms of intron loss and gain in the social amoebae *Dictyostelium*. *BMC Evolut. Biol.*, 15: 286-286. DOI: 10.1186/s12862-015-0567-y
- Ma, M.Y., T. Zhu, X.N. Li, X.R. Lan and H.Y. Liu *et al.*, 2015b. Imprecise intron losses are less frequent than precise intron losses but are not rare in plants. *Biol. Direct*, 10: 24-24. DOI: 10.1186/s13062-015-0056-7
- Ma, M.Y., X.R. Lan and D.K. Niu, 2016. Intron gain by tandem genomic duplication: A novel case in a potato gene encoding RNA-dependent RNA polymerase. *Peer J.*, 4: 1-12. DOI: 10.7717/peerj.2272
- Mohlhenrich, E.R. and R.L. Mueller, 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution*, 70: 2865-2878. DOI: 10.1111/evo.13084
- Smith, D.R. and R.W. Lee, 2009. Nucleotide diversity of the *Chlamydomonas reinhardtii* plastid genome: Addressing the mutational-hazard hypothesis. *BMC Evolut. Biol.*, 9: 120-120. DOI: 10.1186/1471-2148-9-120
- Smith, D.R., 2016. The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Molecular Ecol.*, 25: 37-69. DOI.org/10.1111/mec.13742
- Talavera, G. and J. Castresana, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.*, 56: 564-577. DOI: 10.1080/10635150701472164
- Weber, C.C., B. Boussau, J. Romiguier, E.D. Jarvis and H. Ellegren, 2014. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biol.*, 15: 1-16. DOI: 10.1186/s13059-014-0549-1
- Whitney, K.D. and T. J. Garland, 2010. Did genetic drift drive increases in genome complexity? *PLoS Genet.*, 6: 1287-1292. DOI: 10.1371/journal.pgen.1001080
- Yang, Y.F., T. Zhu and D.K. Niu, 2013. Association of intron loss with high mutation rate in *Arabidopsis*: Implications for genome size evolution. *Genome Biol. Evolut.*, 5: 723-733. DOI: 10.1093/gbe/evt043
- Yang, Z.H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biol. Evolut.*, 24: 1586-1591. DOI: 10.1093/molbev/msm088