



Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

Lilian Oliveira Machado, Suziane Alves Barcelos, Deise Sarzi Shröder,

*Valdir Marcos Stefenon

*Universidade Federal do Pampa - UNIPAMPA, Nucleus of Genomics and Molecular Ecology, Interdisciplinary Center of Biotechnological Research, Av. Antonio Trilha 1847, 97300-000, São Gabriel, RS, Brazil

valdirstefenon@unipampa.edu.br

Received Date: May 17, 2017

Accepted Date: May 31, 2017

Published Date: June 02, 2017

Abstract: This study investigated the phylogenetic signal of the nuclear gene GA20ox1 in seed plants focusing in the relationship between Monocots and Eudicots. Sequences were obtained from GenBank and analyzed using the maximum likelihood and the maximum parsimony approaches. A maximum likelihood tree was built using sequences of the *rbcl* plastid gene in order to enable comparison of the results. The GA20ox1 gene presents neutral evolution, levels of homoplasy equivalent to that observed in chloroplast sequences and generated well-resolved phylogenetic relationships. The relationship between Monocots and Eudicots based on the GA20ox1 gene was clear resolved, revealing the evolution of both groups. All these characteristics taken together make the GA20ox1 gene a promissory marker to corroborate as well as to complement and resolve phylogenetic relationships among species within one to several genera.

Keywords: Nuclear gene, flowering plants, systematics, gibberellin, phylogeny

INTRODUCTION

The large amount of DNA sequences generated in the last decades for an increasing number of different species has enabled to refine the phylogenetic relationships among flowering plants and enabled the generation of better-resolved classifications for this group (APG 2009, Babineau et al. 2013). Despite this progress, some undefined or weakly supported clades still remain and additional molecular data are needed to increase support for these relationships. Soltis et al. (2000) reported a well resolved and highly supported topology of the angiosperm phylogeny by combining chloroplast and nuclear sequence data sets (*rbcl*, *matK* and 18S rDNA) and suggested that most of the remaining phylogenetic questions could best be addressed by sequencing additional genes, without the need of adding more taxa in the analysis.

In general, plant molecular phylogenetics has been very dependent on nucleotide sequences of chloroplast (cpDNA) and nuclear ribosomal (rDNA) genes, but the importance of using additional new nuclear sequences has been proposed as an alternative to improve the resolution of phylogenetic relationships (Qiu et al. 1999, Soltis et al. 1999, Smal et al. 2004). Phylogenetic markers must be at least 500 bp in length and should neither be too conserved nor too variable (Belinky et al. 2012). Since introns and intergenic regions are usually too variable to be used as informative markers above the intraspecific level, the evaluation of exon regions may be a valuable alternative for phylogenetic reconstructions above the species level (e.g. Belinky et al. 2012). In addition, attributes as the elevated rate of sequence evolution, existence of multiple independent loci and the

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

biparental inheritance make nuclear genes a very attractive alternative for estimating species trees (Smal et al. 2004). Following this rationality, a phylogenetic investigation based on nuclear genes should start by selecting candidate genes for a preliminary study (Smal et al. 2004), in which the utility of such DNA sequences can be evaluated. Considering the very large amount of gene sequences deposited in the public databanks of nucleic acids (e.g. GenBank, EMBL, DDBJ), such repositories may be an attractive source of DNA sequences for initial screening of promissory exonic regions with phylogenetic signal. However, an evaluation of the reliability of the sequences deposited in public repositories for phylogenetic analysis is crucial for testing the usefulness of these databanks for such a purpose.

The aim of the present study was to investigate the phylogenetic signal of the nuclear gene Giberellin 20-oxidase1 (GA20ox1) in seed plants focusing in the relationship between Monocots and Eudicots. Using nucleotide sequences deposited in the GenBank, we intended to test the hypotheses that the GA20ox1 is a useful gene for phylogenetic inferences for seed plants.

MATERIAL AND METHODS

Initially an exhaustive search was performed in the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>) in order to find all sequences deposited for the gene GA20ox1, as suggested by Small et al. (2004). The GA20ox is a low-copy nuclear gene (four copies characterized in flowering plants) related to the activation of the plant growth hormone gibberellin (Hedden and Phillipos 2000) and seems to be relatively conserved across plant species (Huerta et al. 2009). Although there are no grounds to expect that any particular gene will be universally useful at any given phylogenetic depth (Small et al. 2004), the GA20ox1 gene may match the main features of a phylogenetic useful nuclear gene, an elevated rate of neutral sequence evolution.

Sequences of 60 species representing Monocots (12 species), Eudicots (57 species) and Pinaceae (1 species) were recorded in the GenBank (Table 1). Sequence of the GA20ox1 gene from the lycophyte *Selaginella moellendorffii* Hieron. (Lycopodiidae) was included in the analysis as outgroup. Sequences of the GA20ox1 gene were aligned with the software Muscle (Edgar 2004), using the Neighbor Joining clustering method for the interactions and all default parameters of the program.

In order to characterize evolutionary patterns of the GA20ox1 sequences, the transition/transversion matrix (using the maximum likelihood method) and a codon-based Z-test of selection (using the Pamilo-Bianchi-Li method, complete deletion of gaps and 1000 bootstraps for significance determination) were estimated using Mega 5.05 (Tamura et al. 2011).

The phylogenetic relationship among species was analyzed using the maximum parsimony (MP) and the maximum likelihood (ML) approaches. The best-fit model of nucleotide substitutions for the sequences was determined through the Akaike Information Criterion (AIC) in the software jModelTest 0.1.1 (Posada 2008). The MP analysis was performed using the Tree-Bisection-Regrafting (TBR) model, all characters unordered and gaps treated as missing data. The ML tree was built using the GTR+G+I mutation model with rate variation among sites modeled with a discrete gamma distribution with five categories. A bootstrap analysis with 500 replicates was employed to assess the internal support of the clades for ML and MP trees. Phylogenetic trees were built in Mega 5.05 (Tamura et al. 2011).

Aiming to compare the phylogenetic signal of the GA20ox1 gene with the phylogenetic relationship retrieved with plastid genes traditionally employed in such studies, sequences of the chloroplast region rbcL deposited in the GenBank were recorded for the same species or, when not available, a species belonging to the same genera recorded for the former gene. The rbcL sequences were aligned and the phylogenetic relationship among species was analyzed using the ML approach as described above. The number of parsimony informative sites, the CI and the RI were computed as for the GA20ox1 sequences.

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

Table 1. Summary of the species included in this study, base composition (%), sequence size (in base pairs) and main taxonomic group of the analysed sequences.

	T	C	A	G	Size	Main group
<i>Elaeis guineensis</i>	21.1	31.9	19.9	27.1	1634.0	Commelinids
<i>Eragrostis tef</i>	15.6	32.7	17.7	34.0	468.0	Commelinids
<i>Hordeum vulgare</i>	19.3	34.0	16.0	30.8	1459.0	Commelinids
<i>Lolium perenne</i>	16.8	35.7	16.5	31.0	1167.0	Commelinids
<i>Musa acuminata</i>	18.8	31.6	19.5	30.1	1507.0	Commelinids
<i>Oryza sativa</i>	13.0	38.9	14.6	33.5	1170.0	Commelinids
<i>Phoenix dactylifera</i>	20.9	30.7	21.1	27.3	1574.0	Commelinids
<i>Setaria italica</i>	18.3	34.5	17.4	29.8	1372.0	Commelinids
<i>Sorghum bicolor</i>	16.1	37.4	15.3	31.2	465.0	Commelinids
<i>Triticum aestivum</i>	17.4	34.5	17.2	30.9	1324.0	Commelinids
<i>Brachypodium</i>	15.6	35.4	15.8	33.2	1192.0	Commelinids
<i>Zea mays</i>	15.8	36.6	15.6	32.0	1169.0	Commelinids
<i>Camellia sinensis</i>	27.4	25.2	28.2	19.3	1374.0	Asterids
<i>Capsicum annuum</i>	30.5	20.4	30.3	18.8	1662.0	Asterids
<i>Chrysanthemum morifolium</i>	27.4	21.4	31.0	20.2	1173.0	Asterids
<i>Datura ferox</i>	27.8	19.6	31.0	21.6	496.0	Asterids
<i>Daucus carota</i>	29.8	21.5	30.3	18.4	1456.0	Asterids
<i>Helianthus annuus</i>	30.9	19.2	30.4	19.5	1608.0	Asterids
<i>Ipomoea nil</i>	23.3	27.9	27.9	21.0	1382.0	Asterids
<i>Lactuca sativa</i>	27.6	21.2	31.6	19.5	1511.0	Asterids
<i>Nerium oleander</i>	26.5	21.8	30.8	20.9	1542.0	Asterids
<i>Nicotiana glauca</i>	28.2	20.7	32.2	18.9	1729.0	Asterids
<i>Panax quinquefolius</i>	25.5	23.0	29.3	22.3	1092.0	Asterids
<i>Solanum lycopersicum</i>	29.0	20.8	31.7	18.5	1137.0	Asterids
<i>Torenia fournieri</i>	26.6	19.2	28.3	25.9	286.0	Asterids
<i>Beta vulgaris</i>	29.4	19.0	32.7	19.0	1528.0	Caryophyllales
<i>Dianthus caryophyllus</i>	28.0	19.0	32.5	20.5	1510.0	Caryophyllales
<i>Rumex palustris</i>	24.0	24.6	28.3	23.1	1200.0	Caryophyllales
<i>Spinacia oleracea</i>	26.0	21.6	32.0	20.5	1364.0	Caryophyllales
<i>Pinus tabulaeformis</i>	26.8	21.7	27.1	24.5	1127.0	Pinales
<i>Nelumbo nucifera</i>	25.3	26.1	26.2	22.3	1593.0	Proteales
<i>Acacia mangium</i>	26.8	20.4	31.1	21.6	1356.0	Rosids
<i>Arabidopsis thaliana</i>	27.9	23.4	29.3	19.5	1429.0	Rosids
<i>Betula luminifera</i>	24.3	23.6	28.8	23.3	1321.0	Rosids
<i>Castanea mollissima</i>	26.7	23.0	30.2	20.1	1305.0	Rosids

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

<i>Cicer arietinum</i>	33.0	16.4	33.3	17.3	1368.0	Rosids
<i>Citrullus lanatus</i>	29.7	18.1	29.0	23.2	1420.0	Rosids
<i>Citrus sinensis</i>	28.3	20.1	30.8	20.9	1715.0	Rosids
<i>Cucumis sativus</i>	28.3	19.6	30.3	21.7	1122.0	Rosids
<i>Cucurbita maxima</i>	26.6	20.7	28.8	23.8	1398.0	Rosids
<i>Eucalyptus grandis</i>	22.4	26.1	26.8	24.7	1722.0	Rosids
<i>Fagus sylvatica</i>	27.6	21.4	31.9	19.2	1629.0	Rosids
<i>Fragaria vesca</i>	26.3	23.2	29.2	21.4	1131.0	Rosids
<i>Glycine max</i>	27.9	21.0	28.2	22.8	1038.0	Rosids
<i>Gossypium hirsutum</i>	25.5	21.5	29.4	23.6	1379.0	Rosids
<i>Medicago truncatula</i>	26.5	22.4	27.4	23.6	1138.0	Rosids
<i>Phaseolus vulgaris</i>	28.8	20.8	30.0	20.4	1601.0	Rosids
<i>Pisum sativum</i>	31.6	18.6	32.4	17.4	1625.0	Rosids
<i>Populus tremula</i>	28.8	22.4	29.0	19.8	1689.0	Rosids
<i>Prunus dulcis</i>	25.7	23.1	28.9	22.2	1209.0	Rosids
<i>Pyrus communis</i>	23.2	24.1	29.2	23.5	1179.0	Rosids
<i>Raphanus sativus</i>	26.2	25.2	29.4	19.3	1311.0	Rosids
<i>Ricinus communis</i>	29.1	19.7	29.0	22.2	1143.0	Rosids
<i>Rorippa aquatica</i>	26.6	24.5	27.6	21.2	1149.0	Rosids
<i>Rosa wichurana</i>	23.2	25.1	26.9	24.8	973.0	Rosids
<i>Salix viminalis</i>	31.1	20.6	28.2	20.1	897.0	Rosids
<i>Sesbania rostrata</i>	29.0	21.2	28.6	21.2	1355.0	Rosids
<i>Theobroma cacao</i>	32.5	19.4	28.3	19.7	1666.0	Rosids
<i>Vitis vinifera</i>	25.0	24.0	27.0	24.0	1149.0	Rosids
<i>Kalanchoe blossfeldiana</i>	27.5	23.3	26.9	22.2	1452.0	Saxifragales
<i>Selaginella moellendorffii</i>	20.2	26.7	23.8	29.3	1050.0	Lycopodiales

RESULTS

Phylogenetic signal of the GA20ox1 gene

Concerning the patterns of nucleotide substitution, 40.3% of the mutations retained in the GA20ox1 sequences evaluated were transversions and 59.7% were transitional, with a ratio of transition to transversion $R = 1.49$. The codon-based Z-test of selection resulted in a statistically not significant value of $Z = -0.705$ ($p = 0.482$), suggesting neutral evolution of the studied sequences.

For the parsimony analysis of the GA20ox1 nucleotide sequences, 56.94% of the characters were parsimony informative, generating a consensus tree with 2808 steps (Fig. 1A), consistency index $CI = 0.23$ and retention index $RI = 0.51$. By comparison, the parsimony analysis of the rbcL sequences (tree not shown) resulted in 28.39% of the characters parsimony informative, generating a consensus tree with 396 steps, consistency index $CI = 0.44$ and retention index $RI = 0.66$.

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

The phylogenetic analysis of the *rbcl* sequences using the ML approach (Fig. 1C) grouped monocots in two separated clusters, while all species from asterids formed a single cluster, with weak bootstrap support (BP < 50%), and rosids revealed a main cluster (BP < 50%; Fig. 1C) mixed with *Kalanchoe* (Saxifragales) and *Rumex* (Charyophyllales). *Vitis* and *Rosa* (rosids) were located outside this main group and *Pinus* (Pinales) is basal to all species.

DISCUSSION

Considering that a gene tree is the phylogeny of a particular DNA sequence, viewing the alleles themselves as the operational taxonomic units (OTUs) and not the evolutionary pathway of a group of OTUs (Avice 1989), the present analysis of the GA20ox1 reflects the phylogenetic relationships of this gene across the recorded species. Even though the Z-test of selection suggested neutrality among the evaluated sequences, the observed phylogenetic relationships may be consequence of lineage sorting acting over the GA20ox1 gene.

The phylogenetic signal of the GA20ox1 gene was measured based on the nucleotide substitutions pattern and the Z-test of neutrality. By evolving slower than transitions, transversions are less susceptible to homoplasy and considered the more reliable type of mutations in constructing phylogenies (Quicke 1993, Yoder et al. 1996). So, a higher amount of transversional mutations are an important aspect in the informative capacity of DNA regions for phylogenetic analyses. The extent of transversional mutations observed in the GA20ox1 (40.3%) and its neutral evolution, confirmed through the Z-test, are desired characteristic for phylogenetic useful nucleotide sequences and assure a suitable phylogenetic signal for this gene.

The amount of parsimony informative characters and level of homoplasy of the GA20ox1 sequences is similar to that observed in useful cpDNA sequences as the *matK* and *trnK*. Evaluating the informative capacity of these two regions for phylogenetic studies of the early diverging eudicots, Hilu et al. (1996) found from 55% to 73% of parsimony informative characters, with CI of the trees ranging from 0.36 to 0.40 and RI ranging from 0.39 to 0.47. Although the proportion of parsimony informative characters of the GA20ox1 sequences is lower than those observed for *matK* and *trnK*, its length (>1200 bp) assures a number of informative characters larger than usually employed cpDNA sequences. In addition, the CI estimation of the GA20ox1 sequences is just slightly lower than the values observed for *matK* and *trnK* sequences, suggesting levels of homoplasy similar to that observe in widely used chloroplast sequences.

The GA20ox1 gene is responsible for the secondary growth in plants, an important trait differing monocots and Eudicots. While secondary growth is usually unexpressive in monocots, it is significant in Eudicots, mainly in tree species. The monocot-dicot divergence estimated through chloroplast gene sequences is about 200 ± 40 million years ago (Wolfe et al. 1989). Thus, the divergence of the GA20ox1 gene likely derives from this time, as clearly expressed in its phylogenetic analysis, which placed monocots as a monophyletic group with strong bootstrap support in the ML tree. Contrasting, the *rbcl* tree expressed the monocot group as paraphyletic. The paraphyletic origin of monocots has been reported in all studies based on the nuclear18S gene and was considered effect of differential lineage sorting (Duval and Erwin 2004).

In conclusion, our evaluation of the GA20ox1 nucleotide analysis revealed neutral evolution, levels of homoplasy equivalent to that observed in chloroplast sequences traditionally employed in phylogenetic analysis and phylogenetic relationships at genus and family levels well resolved with high support in both, ML and MP analyses. All these characteristics taken together make the GA20ox1 gene a promissory marker to corroborate

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

as well as to complement and resolve phylogenetic relationships among species within one to several genera, the level at which most systematists work. Considering that in this study we employed just DNA sequences deposited in the GenBank, further efforts are needed in developing universal primers to amplify this gene across different taxa, aiming to highlight relationships still unresolved in the flowering plants phylogeny. Since the admixture of species from different clades as rosids, asterids, Caryophyllales, Proteales and Saxifragales was observed in the phylogenetic analyses of both, GA20ox1 and rbcL sequences, this approach may solve these inconsistencies by enabling the selection of the more appropriated plant species for phylogenetic studies using the GA20ox1 gene.

ACKNOWLEDGEMENTS

This work was partially supported by financial support of FAPERGS (Process 1013351) and CNPq (Process 471812/2011-0). The authors thank CNPq, UNIPAMPA and CAPES for grant and scholarships.

REFERENCES

- APG. 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APGIII. *Bot. J. Linn. Soc.* 161: 105–121.
- Babineau M, Gagnon E, Bruneau A. 2013. Phylogenetic utility of 19 low copy nuclear genes in closely related genera and species of caesalpinoid legumes. *South Afr J. Bot.* 89: 94–105.
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoop SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress JW, Nixon KC, Farrise JS. 2000. Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn. Soc.* 133: 381-461.
- Qiu YL, Lee J, Quadroni FB, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chasek MW. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404-407.
- Soltis PS, Soltis DE, Chase MW. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402:402-404.
- Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* 17: 145-170.
- Belinky F, Szitenberg A, Goldfarb I, Feldstein T, Wörheide G, Ilan M, Huchon D. 2012. ALG11 – A new variable DNA marker for sponge phylogeny: Comparison of phylogenetic performances with the 18S rDNA and the COI gene. *Mol. Phyl. Evol.* 63: 702–713
- Hedden P, Phillips AL. 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.* 5: 523-530.
- Huerta L, Garcia-Lor A, Garcia-Martinez JL. 2009. Characterization of gibberellin 20-oxidases in the citrus hybrid Carrizo citrange. *Tree Physiol.* 29: 569-577.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.

Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots

- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.
- Posada D. 2008. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* 27: 1253-1256.
- Avise JC. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43: 1192-1208.
- Quicke DLJ. 1993. Principle and techniques of contemporary taxonomy. Chapman & Hall, Glasgow, London UK pp 311.
- Yoder AD, Vilgalys R, Ruvolo M. 1996. Molecular Evolutionary Dynamics of Cytochrome b in Strepsirrhine Primates: The Phylogenetic Significance of Third-Position Transversions. *Mol. Biol. Evol.* 13: 1339-1350.
- Hilu KW, Black C, Diouf D, Burleigh JG. 2000. Phylogenetic signal in matK vs. trnK: A case study in early diverging eudicots (angiosperms). *Mol. Phylogen. Evol.* 848: 1120–1130.
- Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6201-6205.
- Duvall MR, Erwin AB. 2004. 18S gene trees are positively misleading for monocot/dicot phylogenetics. *Mol. Phylogen. Evol.* 30: 97–106.

Citation: Lilian Oliveira Machado, Suziane Barcelos, Deise Sarzi Schröder, Valdir Marcos Stefenon "Phylogenetic Signal of the Nuclear Gene GA20ox1 in Seed Plants: The Relationship Between Monocots and Eudicots". *American Research Journal of Biosciences*; V3, I1; pp:1-8.

Copyright © 2017 Lilian Oliveira Machado, Suziane Barcelos, Deise Sarzi Schröder, Valdir Marcos Stefenon This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.