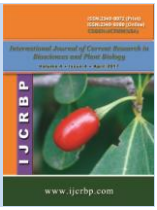
International Journal of Current Research in
Biosciences and Plant Biology
Volume 4 • Number 4 (April-2017) • ISSN: 2349-8080 (Online)
Journal homepage: www.ijcrbp.com



Original Research Article **doi:** https://doi.org/10.20546/ijcrbp.2017.404.005

Bioinformatics Analysis of Squalene Epoxidase (*SE*) Gene from Several Plant Species

Li Zhu[1], Daoju Jiang[2], Qiangwen Chen[1], Weiwei Zhang[1]* and Feng Xu[1]

[1]*College of Horticulture and Gardening, Yangtze University, Jingzhou, 434025, China*
[2]*Jingzhou Forestry Bureau, Jingzhou, Hubei, 434020, China*

Corresponding author.

Abstract

Squalene epoxidase (SE) is one of the important enzymes of the triterpenoid saponin biosynthesis pathway in plants. In this study, SE protein sequences from 16 species of plant were analyzed using bioinformatics methods. The secondary structure, three dimensional structures for these enzymes were predicted. The physico-chemical properties including pI, molecular weight (MW), GRAVY, subcellular localization and conserved domains were also studied. The results of sequence comparison showed that there is high identity between SE proteins from different plants. Phylogenetic analysis showed that SE proteins from plants, mammals, fungus can be clustered into different branches, the relationship was closer within plant species. These results indicated that SE proteins from various species have conserved region, and were derived from a common ancestor.

Article Info

Accepted: 20 March 2017
Available Online: 06 April 2017

Keywords

Bioinformatics
Species
Squalene epoxidase
Triterpenoid saponin

Introduction

Squalene epoxidase (SE) belong to NADB Rossmann superfamily, is a key enzyme, as a bio-catalyst, in the process of squalene turning epoxidation squalene, which exists in the microsome of the endoplasmic reticulum (Jin et al., 2016; Xu et al., 2016). The biosynthesis efficiency of synthesize sterol, steroid saponin, triterpenoid saponin is determined by the activity of SE that also affect the activity and quality of them. Therefore squalene epoxidase has been taken for an important rate-limiting enzyme in the biosynthesis of cycloartenol. *SE* genes have been cloned from several plants and their functions were studied (Liu et al., 2013; Hu et al., 2012). In recent years, the molecular mechanism research of secondary metabolism has become a hot topic; some reports have focused on *SE* genes involved in triterpenoid saponin biosynthetic pathway.

Using bioinformatics to analyze the protein sequence, thus infer and predict its structure and function, which also have become a shortcut for initial identification gene structure and function (Ran et al., 2015). So far, bioinformatics analysis of squalene epoxidase gene was rarely reported (Ding et al., 2014). In this study, we analyzed physicochemical properties, structural characteristic and functional domain of SE proteins from16 plant species *via* bioinformatics method, which may provide reference for the cloning and genetic manipulation.

Materials and methods

Sequence retrieval

A total of 16 amino acid sequences of SE from different plants used in this paper were obtained from NCBI GenBank (www.ncbi.nlm.nih.gov) (Table 1). The protein sequences were retrieved in FASTA format and used for further analysis.

Bioinformatics analyses

The composition and physicochemical properties of the amino acid sequences was analyzed by ProtParam (http://web.expasy.org/protparam/). Using Compute pI/Mw tool (http://web.expasy.org/comute_pi/) to estimate the molecular weight and theoretical isoelectric point. Hydrophilicity or hydrophobicity, subcellular localization were analyzed using ProtScale (http://web.expasy.org/protscale/), Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) (Chou and Shen, 2010), respectively. The secondary structure of the proteins was analysed by using NPS (https://npsa-prabi.ibcp.fr/cgi-bin). The three dimensional structure of the protein was predicted by using SWISS-MODEL (http://swissmodel.expasy. org/) (Biasini et al., 2014). Multiple sequence alignment was performed by DNAMAN 8. Using WebLogo (http://weblogo.berkeley. edu/logo.cgi) to produce sequence logo maps and SMART to predict functional domain of protein. With Clustal X 2.0 and MEGA 6.0 software, the phylogenetic tree of SE protein sequences was constructed using the Neighbor-Joining (NJ) method.

Table 1. Amino acid composition and physicochemical of SE proteins in 16 plants.

Species	GenBank	Amino acids	MW (kDa)	pI	Instability index (%)	Subcellular localization	GRAVY
Panax notoginseng	ABE73759.1	537	59.1	8.63	45.50	Endoplasmic reticulum	0.046
Withania somnifera	ADD17678.1	531	57.7	8.48	38.65	Endoplasmic reticulum	0.046
Gynostemma pentaphyllum	ACQ90301.1	525	57.8	9.04	42.99	Endoplasmic reticulum	0.021
Euphorbia tirucalli	BAF79915.1	531	57.6	8.75	38.15	Endoplasmic reticulum	0.111
Medicago sativa	ABC94943.1	524	57.2	8.54	41.23	Endoplasmic reticulum	0.025
Chlorophytum borivilianum	AFN61200.1	532	57.5	8.42	43.91	Endoplasmic reticulum	0.077
Betula platyphylla	AKR76254.1	526	57.1	8.83	42.78	Endoplasmic reticulum	0.021
Eleutherococcus senticosus	AEJ79819.1	554	60.8	8.98	42.18	Endoplasmic reticulum	0.018
Aralia elata	ADC32655.1	547	60.3	8.85	44.03	Endoplasmic reticulum	0.012
Panax ginseng	ACJ24907.2	545	60.2	9.06	48.68	Endoplasmic reticulum	-0.038
Morus alba	ALD84329.1	524	57.2	8.56	42.55	Endoplasmic reticulum	0.050
Crocus sativus	KGN62420.1	528	57.7	8.99	42.94	Endoplasmic reticulum	0.043
Astraagalus membranaceus	AHY94896.1	528	57.4	8.92	41.39	Endoplasmic reticulum	0.045
Jatropha curcas	AGW81843.1	531	58.3	8.61	49.68	Endoplasmic reticulum	0.063
Camellia oleifera	AGH32908.1	534	58.3	8.74	40.17	Endoplasmic reticulum	0.062
Huperzia serrata	AFO53557.1	561	62.8	9.19	57.75	Endoplasmic reticulum	0.007

Results

Composition and physicochemical properties

Sequence analysis showed that SE proteins sequences from different plants have similar composition and physicochemical properties (Table 1). The length of amino acid sequences was from 524 to 561, and the predicted pI was between 8.48 and 9.19, which indicating the SE is alkaline. According to instability index all sequences are unstable protein exception *Withania somnifera* and *Euphorbia tirucalli*. Subcellular localization analysis showed that SE mainly targeted to endoplasmic reticulum. GRAVY found that all the amino acids were positive values except *Panax ginseng*, which indicates these SE proteins were hydrophobicity, and *Panax ginseng* was hydrophilicity protein.

Structural analysis of SE proteins

The secondary structure of proteins was analysed for considering relative availability of alpha helix, beta turn, extended strand and random coils. The percentage distribution of predicted secondary structure of SE proteins from 16 plant species is represented in Table 2. All the SE proteins were composed of alpha helices, beta turn, extended strand, random coil, with the smallest proportion of beta turn, then of extended strand, the most proportion of the alpha helices or shift random coil. Differences of SE proteins secondary structure indicated that the four structural elements may have unique functions in different species. For prediction of 3D structure of SE proteins, the homology modeling (SWISS-MODEL) analysis was carried out. The 3D structure of SE proteins was mainly consisted of secondary structure elements alpha helices, beta turn, extended strand and random coil. The result of molecular modeling showed that SE proteins from different species are under a similar spatial structure in the conserved regions (Fig. 1). However, the whole space structure of SE proteins from various species was different. Such as SE proteins from *Chlorophytum borivilianum*, *Huperzia serrata* and *Panax notoginseng*, especially in C-terminal.

Table 2. Secondary structure elements of SE proteins.

Species	Alpha helices (%)	Beta turn (%)	Extended strand (%)	Random coil (%)
Panax notoginseng	34.82	10.06	21.97	33.15
Withania somnifera	30.13	9.42	24.29	36.16
Gynostemma pentaphyllum	32.92	11.05	21.71	34.29
Euphorbia tirucalli	30.89	9.79	23.92	35.40
Medicago sativa	32.25	9.73	23.28	34.73
Chlorophytum borivilianum	31.20	12.41	22.93	33.46
Betula platyphylla	29.85	12.17	24.90	33.08
Eleutherococcus senticosus	34.12	10.29	21.84	33.75
Aralia elata	31.99	10.60	23.77	10.60
Panax ginseng	33.94	11.19	23.49	31.38
Morus alba	32.06	12.21	23.85	31.87
Crocus sativus	28.98	10.80	22.16	38.07
Astraagalus membranaceus	28.79	11.17	24.43	35.61
Jatropha curcas	29.94	12.62	25.61	31.83
Camellia oleifera	34.27	11.80	22.10	31.84
Huperzia serrata	34.22	7.31	21.75	36.72

Homology of SE protein sequences

Homology of the SE proteins from different plants was analyzed by using BLASTP and DNAMAN. Multiple comparison results showed that SE protein sequences were high homology among different species (Fig. 2). The sequence identity of SE proteins was more than 88%. Conserved domains such as squalene epoxidase, FAD

binding domain were found in SE proteins. Squalene epoxidase domain contains a putative FAD binding site and it is a key enzyme in the sterol biosynthetic pathway. WebLogo was used to display the conserved domains. As shown in Fig. 2, motifs VGAGVAGxALA, VGELLQPGGYxKL, ELGLxDCV were included in the SE proteins. Alignment of deduced amino acid sequence of SE from 16 kinds of plants are shown in Fig. 3.

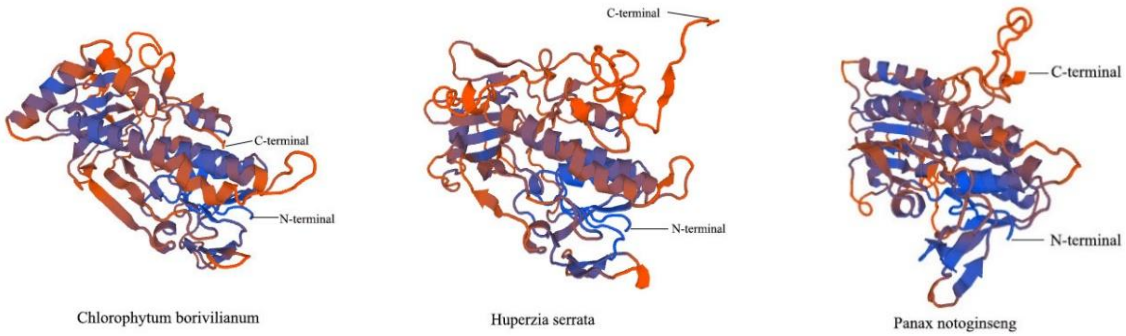


Fig. 1: Three-dimensional structure of SE proteins from *Panax notoginseng*, *Chlorophytum borivilianum* and *Huperzia serrata*.

Fig. 2: The sequence logos of SE proteins in 16 kinds of plants by WebLogo.

Phylogenetic analysis of SE proteins

Phylogenetic tree was constructed to analyze the relationship between the SE proteins from different species. The results showed that SE genes have a common ancestor and SE proteins from the different family were clustered into different branches (Fig. 4). As shown in the phylogenetic tree, SE proteins from monocotyledon and dicotyledon were clustered into different branches, however, they had a closer relationship with each other than with SE from mammals, fungi. SE proteins from the same family were clustered into the same branches, for example, SE proteins from gramineous plants, *Hordeum vulgare*, *Zea mays* and *Oryza sativa* were shared the same subclades. The analysis indicated that SE proteins shared a common evolutionary origin, these proteins had a high homology.

Discussion

Triterpene saponin, one of the important secondary metabolism, whichis the main active ingredients of many medicinal plants. Even though triterpene saponin are not essential for survival of the plant, it may be important for protecting against pest, pathogens and stress (Haralampidis et al., 2002). In vitro, it can be used in anti-cancer, anti-aging, sedation-analgesia and so on. The biosynthesis of triterpene saponin is regulated by squalene epoxidase (Augustin, 2011). Squalene epoxidase catalyzes squalene into oxidosqualene, which is the first oxygenation step and suggested it to be one of rate-limiting enzymes in sterol biosynthesis (Xing et al., 2005). Andoxidosqualene is the precursor substance of triterpene saponin synthesis, yet little study has been done on the squalene epoxidase and biochemical pathways involved in saponins biosynthesis.

In this study, SE proteins involved in secondary metabolism of 16 plant species were used to analyze structural information. Our results proved SE proteins sequences from different plants have similar composition, physicochemical properties, secondary structure element and 3D structure. Secondary structure analysis showed that all the SE proteins were with the smallest proportion of the beta turn, then of extended strand, and the most proportion of alpha helices or shift random coil. The results of phylogenetic showed that SE genes have a common ancestor and SE proteins from the different family were clustered into different branches. Bioinformatic analysis of SE genes can provide the references for the gene cloning and function identification in plants.

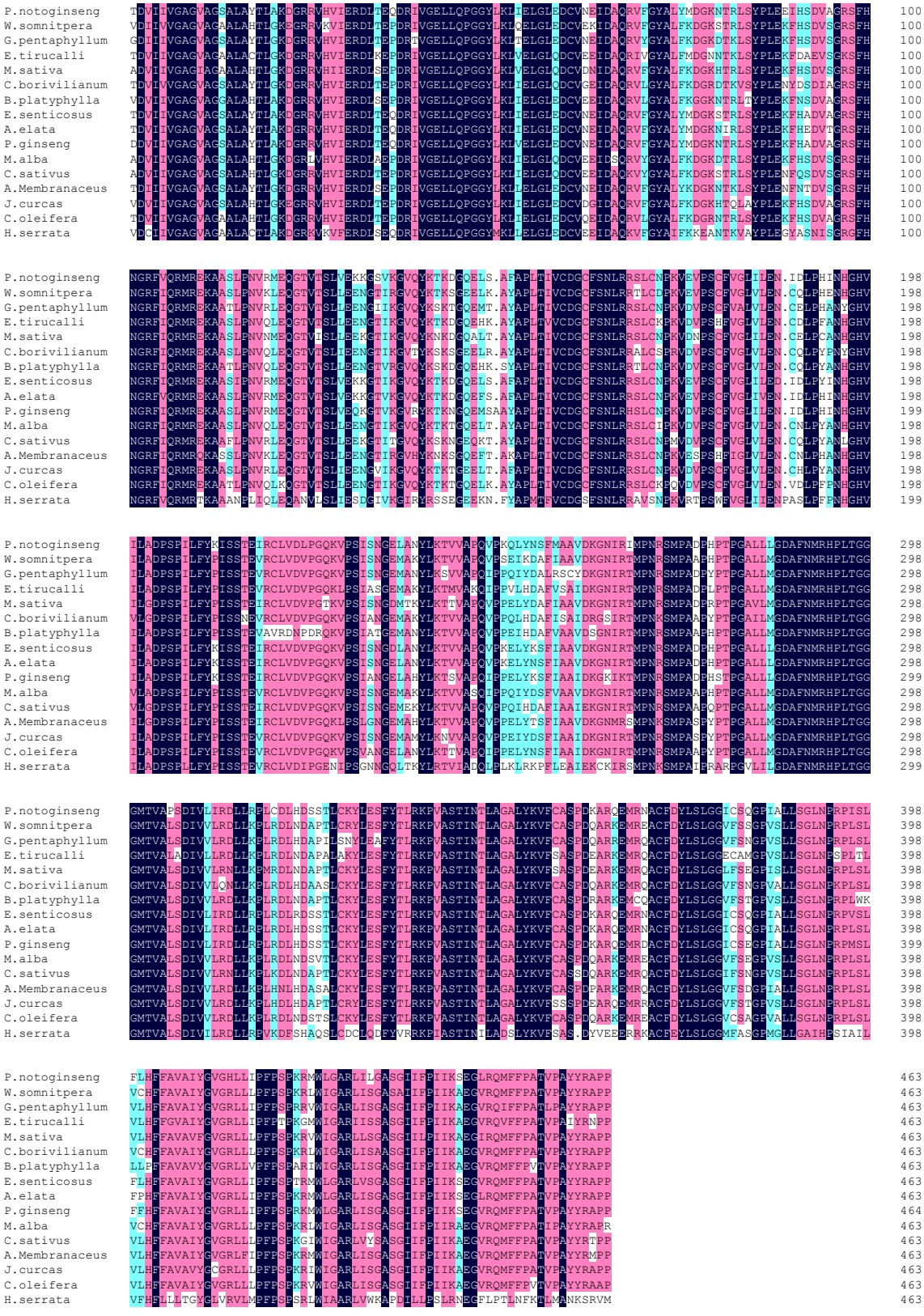


Fig. 3: Alignment of deduced amino acid sequence of SE from 16 kinds of plants. The completely identical amino acids are indicated with white foreground and dark blue background. The conserved amino acids are indicated with pink background. Weakly similar amino acids are indicated with blue background.

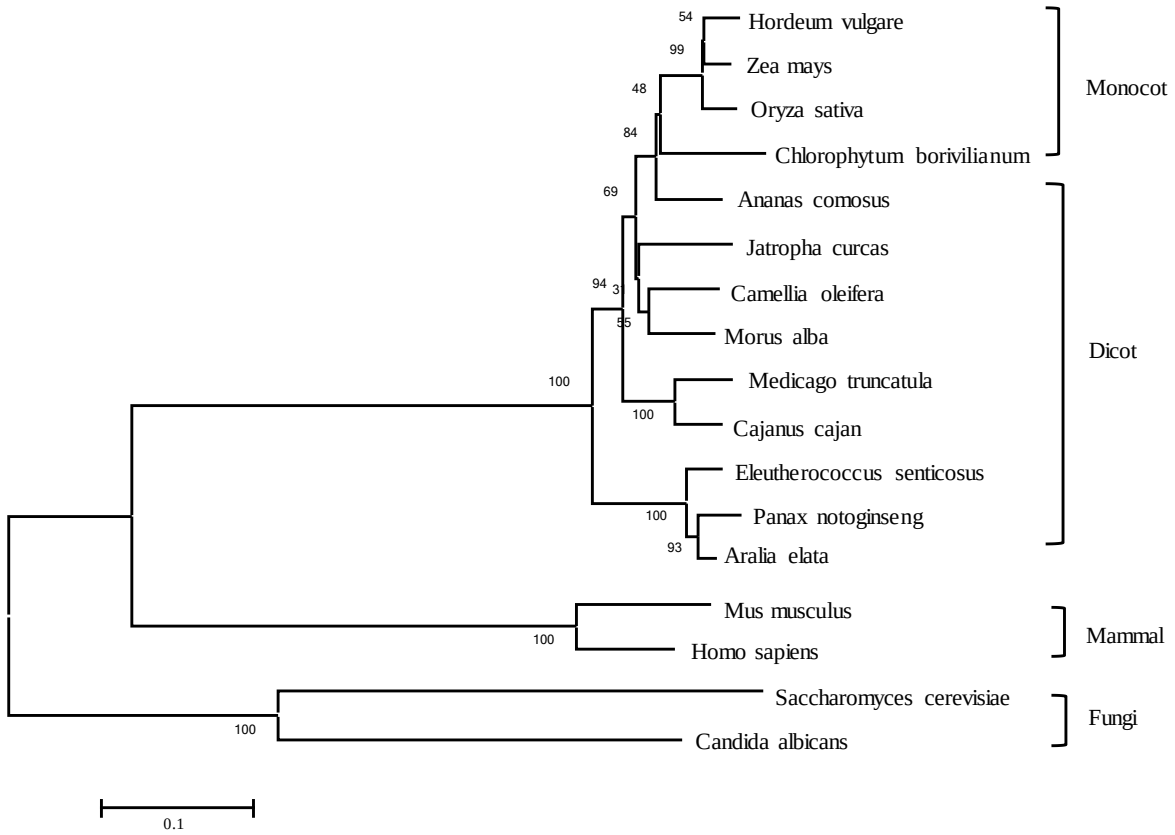


Fig. 4: Phylogenetic analysis of SE proteins. GenBank accession numbers of SS sequences are as follows: *Zea mays* (AFW89101.1), *Oryza sativa* (ABF94794.1), *Ananas comosus* (OAY63657.1), *Hordeum vulgare* (BAJ86899.1), *Jatropha curcas* (AGW81843.1), *Camellia oleifera* (AGH32908.1), *Morus alba* (ALD84329.1), *Panax notoginseng* (ABE73759.1), *Aralia elata* (ADC32655.1), *Eleutherococcus senticosus* (AEJ79819.1), *Chlorophytum borivilianum* (AFN61200.1), *Cajanus cajan* (KYP70371.1), *Mus musculus* (BAA07649.1), *Homo sapiens* (BAA22372.1), *Candida albicans* (BAA13565.1) and *Saccharomyces cerevisiae* (AAA34592.1).

Conflict of interest statement

Authors declare that they have no conflict of interest.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (31500546), the Doctor Foundation of Yangtze University (801190010127).

References

Augustin, J. M., Kuzina, V., Andersen, S. B., Bak, S., 2011. Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochem. 72(6), 435-57.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucl. Acids Res. 42(W1), W252-W258.

Chou, K. C., Shen, H. B., 2010. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. PloS One. 5(6), e11335.

Ding, C., Zhao, C. L., Chen, Z. J., Wei, F. G., Song, D. G., Miao, K. R., 2014. Bioinformatics analysis of the squalene synthase gene and the amino acid sequence in ginseng species. Int. J. Clin. Exp. Med. 8(8), 12818-12825.

Haralampidis, K., Trojanowska, M., Osbourn, A. E., 2002. Biosynthesis of triterpenoid saponins in plants. Adv. Biochem. Eng. 75(75), 31-49.

Jin, J., Ya-Yun, W.U., Yan, X., Huang, X., Zhang, S.P.,

Wu, H., 2016. Bioinformatics analysis of squalene epoxidase coding gene in *Panax* genus. J. Public Health Prev. Med. 27(3), 9-12.

Liu, N., Tian, Y. H., Meng, X. Y., Zhang, L. X., 2013. Cloning of Encoding Squalene Epoxidase Gene and Its Expression in Panax ginseng Root Tissue. J. Jilin Agric. Univ. 35(1), 40-45.

Ran, K., Wei, S. W., Wang, H. W., Zhang, Y., Wang, S. M., 2015. Bioinformatics analysis of physical and chemical properties and structure characteristics of aquaporins in 22 kinds of plants. Plant Physiol. J. 51(1), 97-104.

Xing, Z. B., Wang, Y. M., Chen, Z. C., Shen, H. L., 2005.The biosynthesis of triterpene saponin. Chem. Life. 25(5), 420-422.

Xu, Y., Zhao, S., Di, L., 2016. Research progress on squalene epoxidase gene of the medicinal plant. Anhui Med. Pharmaceut. J. 20(3), 417-420.

How to cite this article:
Zhu, L., Jiang, D., Chen, Q., Zhang, W., Xu, F., 2017. Bioinformatics analysis of squalene epoxidase (*SE*) gene from several plant species. Int. J. Curr. Res. Biosci. Plant Biol. 4(4), 25-31.
doi: https://doi.org/10.20546/ijcrbp.2017.404.005