Chapter 4: Results

4.1. Major ginsenosides in rhizome of *Panax sokpayensis*

Major ginsenosides, viz., Rb1, Rb2, Rc, Rd, Re, Rf, Rg1 and Rg2 were quantified in the rhizome of 10 years old *P. sokpayensis* from Sikkim Himalaya. Among these major ginsenosides, all the ginsenosides were detected and quantified in the rhizomes of *P. sokpayensis* except ginsenoside Rc (Table 4.1). Ginsenoside Rg2 was found in highest amount, 7.97 ± 0.11 mg/g of dry weight (DW) followed by Rd (7.88 ± 0.78 mg/g of DW), Rb1 (5.19 ± 0.56 mg/g of DW), Re (4.40 ± 0.43 mg/g of DW), Rg1 (2.44 ± 0.24 mg/g of DW) and Rf (0.61 ± 0.23 mg/g of DW). Rb2 was present in least amount (0.04 ± 0.02 mg/g of DW) among the ginsenosides tested. Total content of ginsenosides was 28.53 ± 2.07 mg/g of DW. Among the dammarane type of ginsenosides, concentrations of protopanaxadiol and protopanaxatriol types were 13.11 ± 1.35 and 15.41 ± 0.75 mg/g of DW, respectively (Table 4.1). The Rb1:Rg1 ratio was 2.13 (Table 4.1).

Table 4.1 Quantification of major ginsenosides (mg/g of dry weight) in *P. sokpayensis*

Ginsenosides	*P. sokpayensis*
Rg2	7.97 ± 0.11
Rg1	2.44 ± 0.24
Rf	0.61 ± 0.23
Re	4.40 ± 0.43
Rd	7.88 ± 0.78
Rc	ND
Rb2	0.04 ± 0.02
Rb1	5.19 ± 0.56
Protopanaxadiol (Rb1,Rb2,Rc,Rd)	13.11 ± 1.35
Protopanaxatriol (Re,Rf,Rg1,Rg2)	15.41 ± 0.75
Rb1:Rg1	2.13
Total	28.53 ± 2.07

Mean value ± standard deviation (n = 3). ND – Not detected
Source: Adapted from Gurung et al.,2018

4.2. Cloning and analysis of genes involved in ginsenoside biosynthetic pathway

4.2.1. Genes cloned through Suppression Subtractive Hybridization (SSH) cDNA libraries

SSH cDNA libraries were prepared using PCR-select^{TM} cDNA SSH kit (Clontech, USA) as per the manufacturer's instructions (section 3.4). The total RNA from leaf and rhizome (Figure 4.1A) was used for the purification of respective mRNAs (Figure 4.1B). These mRNAs were used for the construction of leaf and rhizome specific SSH libraries (Figures 4.2, 4.3). Briefly, after the synthesis of double stranded leaf and rhizome cDNAs (Figure 4.2A), they were restriction digested with *Rsa* I to create shorter, blunt ended ds cDNA fragments (Figure 4.2A). The experimental tester and driver cDNAs were then ligated with adaptors (section 3.4.5.4). After checking the ligation efficiency (Figure 4.2B), first and second hybridizations of tester and driver cDNAs were performed (section 3.4.5.4). Then, two rounds of primary and secondary PCRs were performed to analyze differentially expressed cDNAs (Figure 4.2C). These enriched cDNAs were cloned as mentioned under section 3.4.6. The insert size was confirmed using colony PCR before sequencing (Figure 4.3).

4.2.1.1. SSH cDNA libraries, sequencing and assembly of ESTs

From the total of 658 clones that were sequenced from the leaf SSH cDNA library, 513 high quality ESTs were used for sequence assembly (Table 4.2). In the case of rhizome SSH cDNA library, 444 clones were sequenced out of which 374 high quality ESTs were used for sequence assembly. The above ESTs clustered into 59 singletons and 21 contigs accounting for 80 unigenes in the leaf SSH library whereas the ESTs of rhizome SSH library clustered into 119 singletons and 41 contigs adding up to 160 unigenes (Table 4.2).

4.2.1.2. Functional annotation of the ESTs

The functional annotation of high quality ESTs from both leaf and rhizome SSH cDNA libraries were performed using Blast2GO gene annotation tool (Conesa et al., 2005). 64.91 % ESTs (333 out of 513) from the leaf and 69.25 % (259 out of 374) from the rhizome SSH cDNA library showed significant homology to the proteins in the NCBI database (Table 4.2). Thus remaining 179 ESTs (34.89 %) from the leaf SSH library and 115 ESTs (30.75 %) from the rhizome library did not show significant homology to any

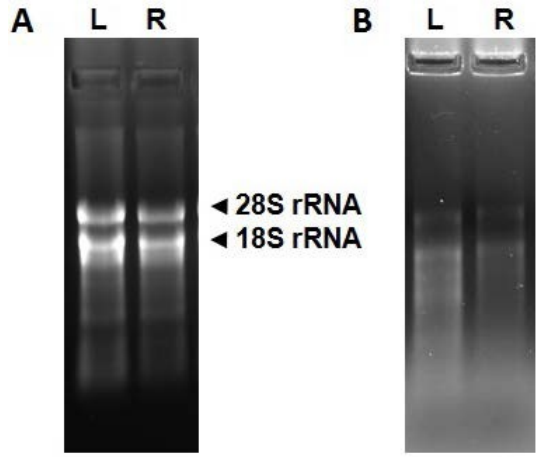


Figure 4.1 Total RNA and mRNA used for SSH cDNA library construction (A) Leaf and rhizome total RNA (B) Leaf and rhizome mRNA. L: leaf, R: rhizome.

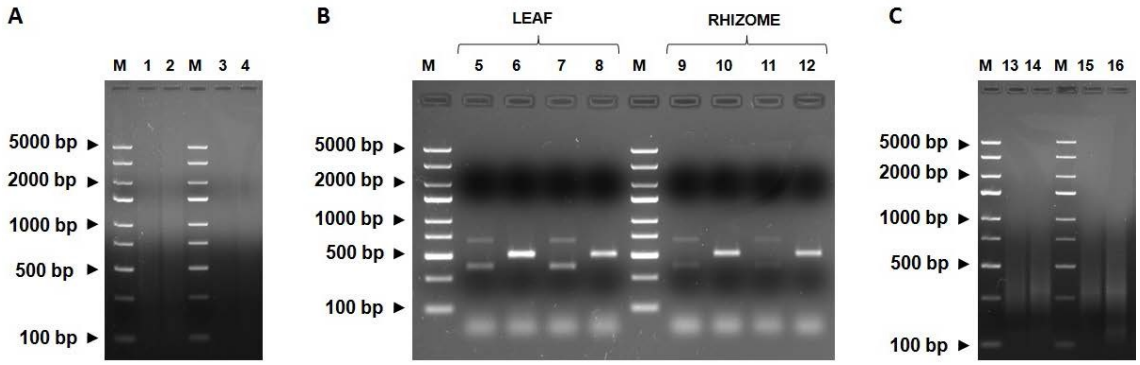


Figure 4.2 Construction of SSH cDNA libraries (A) cDNA synthesized from mRNA of leaf and rhizome (lanes 1 and 2), leaf and rhizome cDNAs after *Rsa*I digestion (lanes 3 and 4) (B) Adapter ligation efficiency test using PCR primers (supplied by the kit) and 26SrRNA primers (Singh et al., 2004). Lane 5: PCR products using tester 1-1 (adapter 1 ligated leaf cDNA) as template and *26SrRNA* reverse primer and PCR primer 1. Lane 6: – PCR products using tester 1-1 as template and 26SrRNA forward and reverse primers. Lane 7: PCR products using tester 1-2 (adapter 2 ligated leaf cDNA) as template and *26SrRNA* reverse primer and PCR primer 1. Lane 8: PCR products using tester 1-2 as template and *26SrRNA* forward and reverse primers. Lanes 9 – 12 represent PCR products amplified using tester 2-1 (adapter 1 ligated rhizome cDNA) and tester 2-2 (adapter 2 ligated rhizome cDNA) with PCR primer 1 and *26SrRNA* primers in similar pattern to lanes 5 – 8. (C) PCR amplification of subtracted cDNAs. Lanes 13 and 14: primary and secondary PCR with leaf cDNA; lanes 15 and 16: primary and secondary PCR with rhizome cDNA. M: DNA marker.

protein in the NCBI database (Table 4.2) suggesting that these ESTs are novel. The ESTs showed homology to genes from several species. In the leaf SSH library, the highest number of homology corresponded to the genes from *P. trichocarpa* (78 ESTs) followed by *Theobroma cacao* (32 ESTs), *P. ginseng* (31 ESTs), and *Solanum tuberosum* (31 ESTs) (Figure 4.4A). In the rhizome SSH library, the highest number of homologies corresponded to genes from *Vitis vinifera* (80 ESTs) followed by *Sesamum indicum* (48 ESTs), *P. ginseng* (30 ESTs) and *Citrus clementina* (19 ESTs) (Figure 4.4B).

The most abundant EST in the leaf SSH library was that of *galactinol synthase 2* (82 ESTs). It was then followed by ESTs of genes encoding novel protein (40 ESTs), ribosomal RNA processing Brix domain protein isoform 2 (33 ESTs), cell division cycle 20.1 (31 ESTs), plastocyanin (29 ESTs), ribonuclease t2 family protein (21 ESTs), metallothionein-like protein type 3 (17 ESTs), ginsenoside biosynthesis related-5 (GBR-5) protein (16 ESTs), glycolate oxidase family protein (15 ESTs), ribulose 1,5 bisphosphate carboxylase oxygenase (15 ESTs),

Table 4.2 Summary of ESTs from SSH leaf and rhizome cDNA libraries.

Descriptive Category	**Leaf SSH library**	**Rhizome SSH library**
Number of clones sequenced	658	444
Number of high quality ESTs	513	374
Longest EST (bp)	581	756
Mean EST length (bp)	231.20	216.10
ESTs with known function	333	259
ESTs with unknown function	179	115
Number of singletons	59	119
Number of contigs	21	41
Number of unigenes	80	160

Source: Reproduced from Gurung et al., 2016

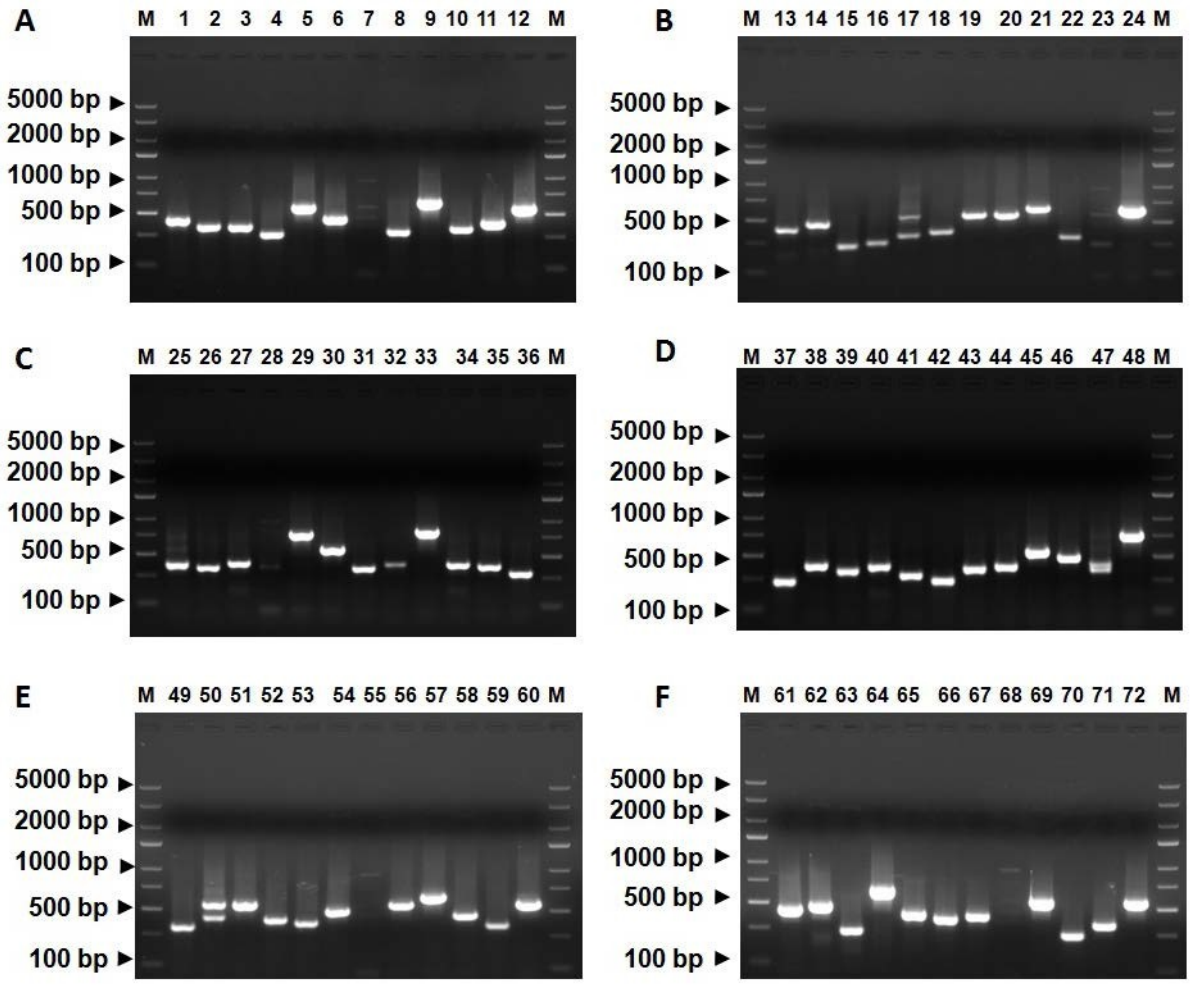


Figure 4.3 Colony PCR analysis (A – C) putative recombinant clones of leaf SSH library (D – F) putative recombinant clones of rhizome SSH library. M: DNA marker.

carbonic anhydrase (14 ESTs), PsbA (12 ESTs) and chloroplast chlorophyll a/b binding protein (10 ESTs) (Table 4.3). In the rhizome SSH library, ESTs of genes encoding protein KIAA0664 homologue (60 ESTs) were the most abundant (Table 4.4). It was followed by ESTs of genes encoding ubiquitin-activating enzyme e1 1 (33 ESTs), 40s ribosomal protein s26-3 (20 ESTs), major latex-like protein (12 ESTs), novel protein (8 ESTs), ribonuclease-like storage protein (7 ESTs), alpha-1,4 glucan phosphorylase L isozyme (7 ESTs), actin (6 ESTs), glyceraldehyde-3-phosphate (5 ESTs), ankyrin repeat-containing protein (5 ESTs), protochlorophyllide reductase (4 ESTs), probable polygalacturonase (4 ESTs), β-amylase (4 ESTs) and polyadenylate-binding protein RBP47 isoform (3 ESTs) (Table 4.4). Besides, some ESTs of genes encoding FPS, SS and DS were also reported in the rhizome library (Table 4.4).

Table 4.3 ESTs redundancy in leaf SSH library.

GenBank accession number	Seq. Name	Seq. Description	Length (bp)	E-Value	Mean Similarity (%)	Redundancy	Clone ID
JZ822892	PsF1	---NA---	178			29	1, 27, 34, 40, 49, 114, 189, 198, 249, 343, 376, 387, 411, 425, 433, 435, 457, 458, 480, 495, 522, 527, 538, 582, 600, 602, 612, 629, 645
JZ822893	PsF2	Glycolate oxidase family protein	250	1.67E-15	86.05%	15	2, 17, 84, 101, 107, 203, 324, 336, 361, 391, 409, 525, 547, 564, 570
JZ822894	PsF3	---NA---	245			28	3, 7, 13, 50, 86, 123, 160, 170, 187, 245, 248, 251, 256, 260, 302, 345, 346, 402, 406, 412, 456, 462, 512, 531, 546, 585, 639, 651
JZ822895	PsF4	Carbonic anhydrase	574	1.96E-27	83.35%	14	4, 29, 48, 235, 269, 448, 471, 472, 483, 543, 561, 586, 648, 657
JZ822896	PsF5	PsbA	575	1.70E-115	98.70%	12	5, 21, 33, 130, 151, 229, 278, 314, 400, 440, 503, 558
JZ822897	PsF6	Ribonuclease t2 family protein	414	2.97E-20	52.20%	21	6, 57, 60, 69, 79, 88, 97, 155, 188, 193, 214, 261, 276, 320, 368, 369, 427, 554, 595, 637, 640
JZ822898	PsF8	---NA---	157			21	8, 20, 90, 105, 115, 133, 150, 175, 177, 196, 197, 224, 240, 308, 319, 337, 403, 446, 485, 557, 567
JZ822899	PsF9	Ribosomal RNA processing Brix domain protein isoform 2	164	5.51E-24	97.80%	33	9, 17, 19, 22, 24, 32, 35, 43, 54, 65, 110, 131, 159, 171, 259, 258, 280, 304, 380, 396, 407, 422, 473, 497, 505, 511, 516, 539, 588, 596, 611, 618, 636
JZ822900	PsF10	GBR5-like protein	112	1.68E-04	95.00%	16	10, 52, 74, 100, 180, 208, 333, 338, 341, 398, 404, 507, 520, 535, 574, 587
JZ822901	PsF12	Plastocyanin	176	2.99E-28	93.80%	29	12, 26, 31, 39, 53, 72, 73, 80, 82, 95, 98, 117, 136, 182, 212, 215, 226, 242, 271, 277, 303, 321, 328, 363, 537, 555, 560, 566, 577
JZ822902	PsF14	Galactinol synthase 2	332	2.09E-60	90.05%	82	14, 16, 25, 30, 55, 56, 63, 75, 83, 85, 93, 96, 102, 104, 113, 119, 129, 137, 148, 161, 186, 192, 194, 200, 201, 204, 205, 207, 211, 213, 218, 221, 223, 225, 232, 234, 241, 244, 247, 253, 262, 263, 268, 273, 274, 310, 315, 322, 330, 335, 353, 366, 374, 390, 397, 415, 424, 426, 431, 434, 450, 455, 465, 466, 474, 475, 476, 477, 482, 498, 499, 502, 524, 544, 545, 569, 592, 617, 626, 628, 635, 638

JZ822903	PsF15	Cell division cycle 20.1, cofactor of apc complex-like	186	3.70E-19	75.40%	31	15, 23, 38, 44, 81, 89, 91, 149, 162, 178, 222, 228, 239, 243, 250, 272, 306, 350, 370, 372, 375, 377, 393, 444, 451, 517, 523, 532, 552, 573, 631
JZ822904	PsF18	Galactinol synthase 2	332	3.65E-63	91.40%	13	18, 45, 68, 165, 168, 195, 378, 447, 449, 504, 590, 598, 599
JZ822905	PsF36	Ribulose 1,5 bisphosphate carboxylase oxygenase small subunit	99	6.84E-15	99.80%	15	36, 41, 58, 78, 230, 360, 357, 381, 421, 501, 515, 529, 572, 594, 615
JZ822906	PsF37	---NA---	48			40	37, 51, 87, 124, 132, 158, 172, 173, 179, 206, 216, 237, 252, 254, 264, 301, 309, 334, 340, 342, 344, 349, 379, 392, 395, 420, 423, 452, 486, 496, 526, 542, 550, 553, 579, 581, 589, 597, 609, 627
JZ822907	PsF42	---NA---	56			11	42, 94, 103, 106, 227, 317, 329, 410, 441, 519, 541
JZ822908	PsF46	---NA---	273			10	46, 181, 191, 209, 220, 231, 416, 437, 478, 494
JZ822909	PsF47	---NA---	287			1	47
JZ822910	PsF59	---NA---	319			1	59
JZ822911	PsF61	Chloroplast chlorophyll a b binding protein	119	1.61E-17	100.00%	10	61, 219, 326, 365, 384, 419, 454, 493, 559, 563
JZ822912	PsF62	Galactinol synthase	224	7.86E-09	73.00%	1	62
JZ822913	PsF64	---NA---	196			1	64
JZ822914	PsF66	---NA---	273			1	66
JZ822915	PsF67	---NA---	236			1	67
JZ822916	PsF71	---NA---	196			1	71
JZ822917	PsF92	Galactinol synthase	190	2.84E-06	75.78%	1	92
JZ822918	PsF116	---NA---	96			1	116
JZ822919	PsF157	---NA---	273			1	157
JZ822920	PsF163	---NA---	245			1	163
JZ822921	PsF164	Glycolate oxidase-like protein	138	2.02E-20	97.75%	4	164, 399, 443, 653
JZ822922	PsF166	---NA---	273			1	166
JZ822923	PsF184	Metallothionein-like protein type 3	204	1.02E-08	72.10%	17	11, 70, 121, 184, 318, 364, 408, 438, 461, 484, 488, 506, 514, 551, 605, 632, 655
JZ822924	PsF210	Glycolate oxidase-like protein	138	6.66E-20	95.50%	1	210

JZ822925	PsF233	Plastocyanin	177	2.81E-25	93.55%	1	233
JZ822926	PsF246	---NA---	416			1	246
JZ822927	PsF255	---NA---	245			1	255
JZ822928	PsF270	---NA---	196			1	270
JZ822929	PsF305	Photosystem I reaction center subunit psaK, chloroplastic	254	1.83E-34	98.00%	3	305, 323, 489
JZ822930	PsF312	Glycolate oxidase-like protein	139	2.37E-17	97.75%	1	312
JZ822931	PsF313	---NA---	112			1	313
JZ822932	PsF325	Oxygen-evolving enhancer protein-2	239	7.68E-41	90.75%	1	325
JZ822933	PsF332	---NA---	178			1	332
JZ822934	PsF351	---NA---	245			1	351
JZ822935	PsF355	Plastocyanin	176	8.28E-28	92.60%	1	355
JZ822936	PsF358	Photosystem I reaction center subunit psaK, chloroplastic	254	7.00E-34	95.00%	1	358
JZ822937	PsF383	Plastocyanin	176	5.97E-24	85.70%	1	383
JZ822938	PsF413	---NA---	213			1	413
JZ822939	PsF414	PsbA	577	4.87E-101	99.25%	1	414
JZ822940	PsF428	---NA---	245			1	428
JZ822941	PsF430	Carbonic anhydrase	301	4.57E-58	92.50%	1	430
JZ822942	PsF432	---NA---	56			1	432
JZ822943	PsF436	---NA---	114			1	436
JZ822944	PsF442	Metallothionein 3-like protein	222	6.04E-19	80.25%	1	442
JZ822945	PsF445	---NA---	244			1	445
JZ822946	PsF453	---NA---	270			1	453
JZ822947	PsF459	---NA---	178			1	459
JZ822948	PsF460	---NA---	90			1	460
JZ822949	PsF463	---NA---	178			1	463
JZ822950	PsF479	---NA---	196			1	479
JZ822951	PsF487	Chlorophyll a b binding protein 13,chloroplastic like	334	1.93E-71	98.30%	1	487

JZ822952	PsF500	Oxygen-evolving enhancer protein 2	239	7.68E-41	90.75%	1	500
JZ822953	PsF508	---NA---	196			1	508
JZ822954	PsF513	---NA---	198			1	513
JZ822955	PsF521	---NA---	245			1	521
JZ822956	PsF528	---NA---	197			1	528
JZ822957	PsF534	---NA---	170			1	534
JZ822958	PsF540	Plastocyanin A	331	1.16E-45	89.00%	1	540
JZ822959	PsF548	---NA---	207			1	548
JZ822960	PsF549	Galactinol synthase 1	259	6.11E-42	88.85%	1	549
JZ822961	PsF556	---NA---	146			1	556
JZ822962	PsF565	PsbA	578	7.47E-92	88.05%	1	565
JZ822963	PsF571	---NA---	176			1	571
JZ822964	PsF575	---NA---	217			1	575
JZ822965	PsF593	---NA---	175			1	593
JZ822966	PsF601	---NA---	50			1	601
JZ822967	PsF606	PsbA	581	8.91E-83	99.75%	1	606
JZ822968	PsF608	---NA---	414			1	608
JZ822969	PsF644	Plastocyanin	177	1.97E-27	92.85%	1	644
JZ822970	PsF647	---NA---	181			1	647
JZ822971	PsF650	---NA---	64			1	650
					Total	513	

Source: Reproduced from Gurung et al., 2016

Table 4.4 ESTs redundancy in rhizome SSH library.

GenBank accession number	Seq. Name	Seq. Description	Length (bp)	E-Value	Mean Similarity (%)	Redundancy	Clone ID
JZ822732	PsR2	---NA---	193			1	2
JZ822733	PsR3	Polyadenylate-binding protein	192	1.30E-25	95.15%	1	3
JZ822734	PsR5	Neutral invertase	312	2.60E-55	95.30%	1	5
JZ822735	PsR6	Tetratricopeptide repeat-like	169	5.34E-26	88.45%	1	6

		superfamily protein					
JZ822736	PsR7	---NA---	453			5	7, 35, 40, 45, 52
JZ822737	PsR8	---NA---	308			1	8
JZ822738	PsR9	Transcription factor bhlh96 like	215	9.54E-14	72.45%	1	9
JZ822739	PsR11	Major latex-like protein	439	1.18E-75	75.95%	12	11, 27, 291, 305, 355, 357, 358, 388, 392, 401, 416, 422
JZ822740	PsR12	---NA---	167			1	12
JZ822741	PsR13	---NA---	145			1	13
JZ822742	PsR14	ARF guanine-nucleotide exchange factor gnom-like	280	2.29E-13	76.50%	1	14
JZ822743	PsR15	Phosphatidylinositol -trisphosphate 3-phosphatase and dual-specificity protein phosphatase pten like isoform x1	475	5.14E-27	68.75%	1	15
JZ822744	PsR16	---NA---	149			1	16
JZ822745	PsR17	40s ribosomal protein s18	462	1.09E-85	95.55%	2	17, 31
JZ822746	PsR19	Tetratricopeptide repeat-like superfamily protein	169	5.34E-26	88.45%	1	19
JZ822747	PsR20	---NA---	106			2	20, 53
JZ822748	PsR21	---NA---	106			1	21
JZ822749	PsR22	60s ribosomal protein l18	312	2.33E-35	94.00%	1	22
JZ822750	PsR23	---NA---	88			1	23
JZ822751	PsR25	ADP-ribosylation factor	213	1.07E-35	98.00%	1	25
JZ822752	PsR26	Ribonuclease like storage protein	217	3.77E-12	64.00%	6	26, 49, 269, 275, 279, 314
JZ822753	PsR30	Aldehyde dehydrogenase	260	6.93E-27	81.55%	2	30, 39
JZ822754	PsR32	RING U-box superfamily protein	192	6.32E-06	56.00%	1	32
JZ822755	PsR33	60s ribosomal protein l18	324	1.86E-18	91.45%	1	33
JZ822756	PsR34	Poly A binding protein	192	1.30E-25	95.15%	1	34
JZ822757	PsR36	---NA---	308			1	36

JZ822758	PsR38	---NA---	108			2	38, 50
JZ822759	PsR41	Ribonuclease like storage protein	131	1.39E-17	100.00%	1	41
JZ822760	PsR43	Zinc finger cch domain-containing protein 56	251	6.48E-15	88.45%	1	43
JZ822761	PsR44	---NA---	126			1	44
JZ822762	PsR46	---NA---	93			1	46
JZ822763	PsR47	Phosphatidylinositol -trisphosphate 3-phosphatase and dual-specificity protein phosphatase pten like isoform x1	474	6.03E-27	68.75%	1	47
JZ822764	PsR48	---NA---	420			1	48
JZ822765	PsR54	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	33	54, 73, 112, 113, 114, 123, 126, 153, 158, 160, 163, 165, 168, 169, 184, 185, 186, 187, 188, 189, 190, 193, 204, 205, 206, 213, 214, 216, 220 , 223, 225, 226, 227
JZ822766	PsR55	Protochlorophyllide reductase, chloroplastic	144	8.37E-07	76.95%	4	55, 68, 120, 159
JZ822767	PsR56	Protein KIAA0664 homolog	304	8.39E-15	61.35%	60	56, 58, 60, 66, 75, 79, 80, 81, 83, 84, 85, 86, 89, 93, 94, 95, 98, 109, 111, 115, 116, 118, 119, 121, 124, 129, 130, 132, 134, 136, 139, 141, 144, 146, 147, 148, 149, 150, 152, 154, 155, 157, 162, 170, 171, 177, 181, 182, 191, 197, 201, 203, 210, 212, 215, 217, 218, 219, 221, 224
JZ822768	PsR57	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	57
JZ822769	PsR59	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	59
JZ822770	PsR61	Ubiquitin-activating enzyme e1 1	166	1.94E-22	91.30%	1	61
JZ822771	PsR62	---NA---	148			1	62
JZ822772	PsR65	---NA---	139			1	65
JZ822773	PsR67	Ubiquitin-activating enzyme e1 1-like isoform x1	163	3.11E-25	94.50%	1	67
JZ822774	PsR70	Ubiquitin-activating enzyme e1 1-like isoform x1	128	1.73E-17	90.60%	1	70
JZ822775	PsR72	---NA---	151			5	72, 76, 77, 99, 100

JZ822776	PsR74	Ubiquitin-activating enzyme e1 1	179	2.49E-28	94.50%	1	74
JZ822777	PsR78	Ubiquitin-activating enzyme e1 1	179	7.77E-26	89.50%	1	78
JZ822778	PsR82	Ubiquitin-activating enzyme e1 1	174	1.36E-22	90.10%	1	82
JZ822779	PsR87	Ubiquitin-activating enzyme e1 1-like isoform x2	171	7.98E-23	85.50%	1	87
JZ822780	PsR88	Ubiquitin-activating enzyme e1 1-like isoform x1	171	7.56E-25	89.40%	1	88
JZ822781	PsR90	Ubiquitin-activating enzyme e1 1	179	4.46E-28	93.45%	1	90
JZ822782	PsR91	Ubiquitin-activating enzyme e1 1	179	9.70E-28	93.10%	1	91
JZ822783	PsR92	---NA---	133			1	92
JZ822784	PsR96	Ubiquitin-activating enzyme e1 1	174	5.93E-21	91.15%	1	96
JZ822785	PsR97	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	97
JZ822786	PsR117	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	117
JZ822787	PsR122	Ubiquitin-activating enzyme e1 1	179	2.80E-28	93.05%	1	122
JZ822788	PsR127	Protochlorophyllide reductase	144	1.30E-06	74.70%	1	127
JZ822789	PsR131	Ubiquitin-activating enzyme e1 1	179	4.59E-28	94.55%	1	131
JZ822790	PsR137	Protein KIAA0664 homolog	304	4.01E-12	56.50%	1	137
JZ822791	PsR142	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	142
JZ822792	PsR156	Ubiquitin-activating enzyme e1 1	179	1.38E-25	89.50%	1	156
JZ822793	PsR164	Protochlorophyllide reductase, chloroplastic	144	1.39E-06	76.15%	1	164
JZ822794	PsR166	Ubiquitin-activating enzyme e1 1	179	1.44E-26	93.25%	1	166
JZ822795	PsR167	Ubiquitin-activating enzyme e1 1	179	8.87E-28	91.30%	1	167
JZ822796	PsR172	Protein KIAA0664 homolog	304	6.20E-14	60.25%	1	172

JZ822797	PsR173	---NA---	139			1	173
JZ822798	PsR176	---NA---	151			1	176
JZ822799	PsR178	---NA---	140			1	178
JZ822800	PsR179	Ubiquitin-activating enzyme e1 1	179	2.31E-28	94.50%	1	179
JZ822801	PsR183	Protein KIAA0664 homolog	324	1.80E-15	60.50%	1	183
JZ822802	PsR192	Ubiquitin-activating enzyme e1 1	179	2.49E-28	94.50%	1	192
JZ822803	PsR194	---NA---	151			1	194
JZ822804	PsR195	Ubiquitin-activating enzyme e1 1	183	1.61E-28	94.55%	1	195
JZ822805	PsR196	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	196
JZ822806	PsR198	---NA---	141			1	198
JZ822807	PsR199	Protein KIAA0664 homolog	329	2.29E-14	61.95%	1	199
JZ822808	PsR202	Ubiquitin-activating enzyme e1 1-like isoform x1	180	2.57E-25	93.70%	1	202
JZ822809	PsR209	Protein KIAA0664 homolog	306	2.54E-07	62.83%	1	209
JZ822810	PsR222	Ubiquitin-activating enzyme e1 1	179	8.09E-29	94.45%	1	222
JZ822811	PsR229	Ubiquitin-conjugating enzyme E2	171	2.37E-31	100.00%	1	229
JZ822812	PsR232	---NA---	76			4	232, 271, 277, 320
JZ822813	PsR235	---NA---	276			1	235
JZ822814	PsR236	Major latex-like protein	246	7.45E-27	77.00%	1	236
JZ822815	PsR238	---NA---	276			4	233, 238, 295, 331
JZ822816	PsR239	Ribonuclease-like storage protein	524	9.55E-83	62.65%	1	239
JZ822817	PsR241	Polyadenylate-binding protein rbp47 isoform x1	167	1.36E-20	87.60%	3	241, 281, 312
JZ822818	PsR242	CBL-interacting serine threonine-protein kinase 7	132	1.10E-11	77.05%	2	242, 260
JZ822819	PsR243	---NA---	96			1	243
JZ822820	PsR244	---NA---	229			5	244, 257, 294, 297, 316

JZ822821	PsR245	Probable chromatin-remodeling complex ATPase chain	382	2.72E-06	94.75%	1	245
JZ822822	PsR247	---NA---	111			1	247
JZ822823	PsR248	---NA---	93			5	248, 256, 278, 290, 299
JZ822824	PsR249	---NA---	115			1	249
JZ822825	PsR250	---NA---	136			1	250
JZ822826	PsR251	Ubiquitin-conjugating enzyme	164	7.82E-29	98.00%	1	251
JZ822827	PsR252	Ribonuclease like storage protein	93	4.25E-08	96.00%	1	252
JZ822828	PsR253	---NA---	127			1	253
JZ822829	PsR254	---NA---	110			1	254
JZ822830	PsR255	Ribonuclease-like storage protein	150	3.11E-20	98.50%	1	255
JZ822831	PsR258	---NA---	85			2	258, 303
JZ822832	PsR259	---NA---	40			3	259, 286, 317
JZ822833	PsR261	Glyceraldehyde-3-phosphate dehydrogenase	114	8.12E-16	99.70%	5	261, 273, 318, 322, 369
JZ822834	PsR262	Rhodanese-like domain-containing protein6	138	1.09E-05	90.40%	2	262, 268
JZ822835	PsR264	---NA---	90			1	264
JZ822836	PsR267	Rnase-like major storage protein	267	5.29E-51	68.00%	1	267
JZ822837	PsR270	Ribonuclease like storage protein	245	1.59E-44	97.00%	1	270
JZ822838	PsR272	Ribonuclease like storage protein	179	4.44E-29	98.00%	1	272
JZ822839	PsR276	---NA---	131			3	276, 327, 332
JZ822840	PsR280	Probable polygalacturonase	164	3.20E-24	83.95%	4	280, 298, 304, 315
JZ822841	PsR282	---NA---	83			1	282
JZ822842	PsR284	---NA---	141			1	284
JZ822843	PsR289	---NA---	95			2	289, 329
JZ822844	PsR292	---NA---	54			2	292, 300
JZ822845	PsR293	Ribosomal protein l13a	197	4.74E-18	94.20%	2	293, 310
JZ822846	PsR296	---NA---	250			1	296

JZ822847	PsR302	60s ribosomal protein l21	189	2.56E-22	93.70%	3	302, 309, 231
JZ822848	PsR306	Type II ribosome-inactivating protein cinnamomin	329	2.09E-28	67.90%	1	306
JZ822849	PsR307	Polyubiquitin like protein	258	2.49E-43	99.25%	1	307
JZ822850	PsR328	---NA---	149			1	328
JZ822851	PsR330	Riboflavin synthase alpha chain	405	4.32E-22	76.30%	1	330
JZ822852	PsR334	---NA---	47			1	334
JZ822853	PsR337	Polyadenylate-binding protein	192	6.18E-25	95.15%	1	337
JZ822854	PsR338	---NA---	185			8	338, 346, 347, 349, 363, 371, 380, 385
JZ822855	PsR339	40s ribosomal protein s26	130	1.18E-06	89.45%	20	339, 348, 351, 352, 362, 366, 373, 389, 393, 399, 402, 408, 410, 411, 413, 415, 421, 424, 427, 429
JZ822856	PsR340	Ankyrin repeat-containing protein	185	6.21E-30	95.75%	5	340, 367, 381, 405, 417
JZ822857	PsR341	---NA---	108			1	341
JZ822858	PsR343	---NA---	95			1	343
JZ822859	PsR344	---NA---	296	1.08E-10	75.15%	4	344, 370, 409, 419
JZ822860	PsR345	Alpha 1,4 glucan phosphorylase L isozyme, chloroplastic	544	1.46E-61	69.70%	7	345, 397, 407, 361, 365, 382, 404
JZ822861	PsR350	---NA---	185			2	350, 430
JZ822862	PsR353	Major latex-like protein	119	4.69E-08	90.88%	1	353
JZ822863	PsR354	Ankyrin repeat-containing protein	185	5.88E-29	94.50%	1	354
JZ822864	PsR356	Beta-amylase	357	2.81E-64	88.10%	4	356, 420, 434, 436
JZ822865	PsR359	---NA---	103			1	359
JZ822866	PsR364	---NA---	81			2	364, 387
JZ822867	PsR368	Actin	410	4.51E-34	99.40%	6	368, 376, 390, 391, 394, 432
JZ822868	PsR372	---NA---	180			1	372
JZ822869	PsR374	60s ribosomal protein l9	167	4.29E-26	92.90%	1	374
JZ822870	PsR375	---NA---	198			2	375, 428
JZ822871	PsR377	60s ribosomal	241	1.48E-36	94.65%	1	377

		protein l9					
JZ822872	PsR378	---NA---	122			3	378, 379, 431
JZ822873	PsR383	---NA---	76			1	383
JZ822874	PsR384	Type 2 ribosome-inactivating protein cinnamomin I precursor	378	7.37E-40	70.80%	1	384
JZ822875	PsR396	---NA---	80			2	396, 414
JZ822876	PsR398	---NA---	76			1	398
JZ822877	PsR400	Ribonuclease like storage protein	90	1.73E-07	96.00%	1	400
JZ822878	PsR403	---NA---	108			1	403
JZ822879	PsR406	Heat shock protein 70	253	1.52E-51	100.00%	1	406
JZ822880	PsR412	Maternal effect embryo arrest 18 protein	406	2.91E-32	73.30%	2	412, 423
JZ822881	PsR418	---NA---	184			1	418
JZ822882	PsR425	---NA---	241			3	425, 435, 342
JZ822883	PsR433	---NA---	118			1	433
JZ822884	PsR437	Squalene synthase	734	1.34E-180	97.20%	1	437
JZ822885	PsR438	Dammarenediol synthase	165	4.00E-32	97.20%	1	438
JZ822886	PsR439	---NA---	64			1	439
JZ822887	PsR440	Farnesyl diphosphate synthase	756	1.55E-150	96.75%	1	440
JZ822888	PsR441	---NA---	198	3.12E-12	72.25%	1	441
JZ822889	PsR442	---NA---	60			1	442
JZ822890	PsR443	---NA---	60			1	443
JZ822891	PsR444	Squalene synthase	351	6.80E-80	96.85%	1	444
					Total	374	

Source: Reproduced from Gurung et al., 2016

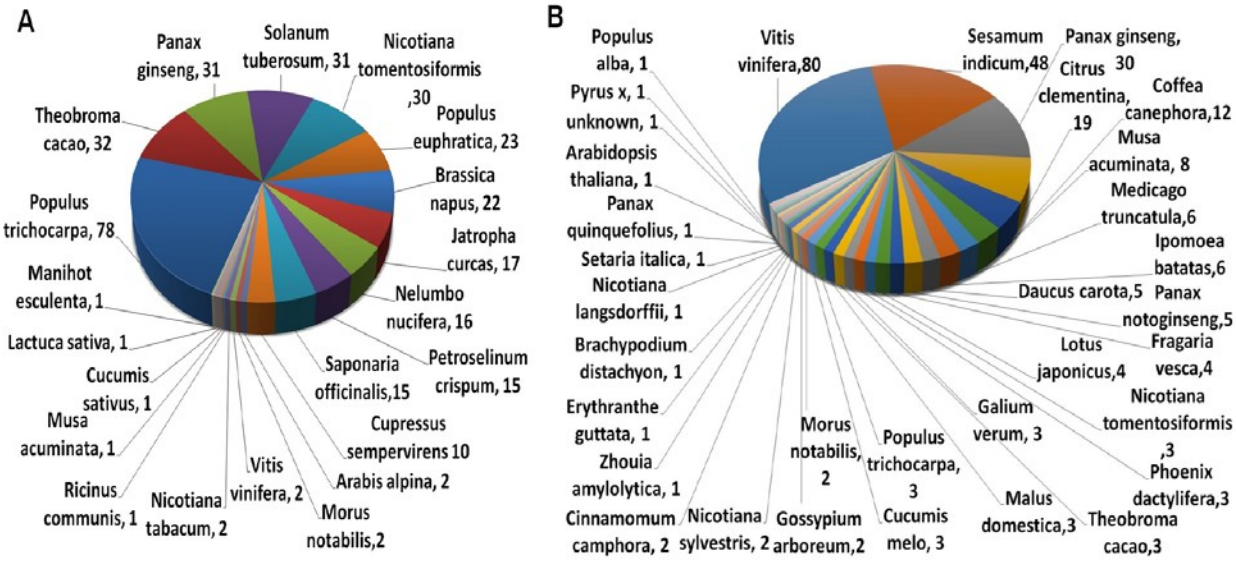


Figure 4.4 Species with BLAST top-hits. (A) Species against which leaf SSH library ESTs recorded hits (B) Species against which rhizome SSH library recorded hits.

4.2.1.3. Gene ontology (GO) annotation of ESTs

The ESTs of each SSH library were grouped into three gene ontology (GO) categories: biological process, molecular function and cellular component (Figure 4.5). For comparison, functional annotation was performed separately for leaf and rhizome SSH libraries. In the leaf SSH library, the biological processes that were represented by majority of ESTs were photosynthesis, small molecule metabolic process, precursors for metabolites and energy, response to radiation and cellular biosynthetic process (Figure 4.5A). The EST of rhizome SSH library represented macromolecule metabolic process, organic substance biosynthetic process, cellular biosynthetic process, ribonucleoprotein complex biogenesis, cellular aromatic compound metabolic process and oxidation-reduction process (Figure 4.5B). Among the molecular functions, many ESTs from the leaf SSH library belonged to transferases (transferring glycosyl groups) and enzymes related to cation binding, nucleic acid binding, hydrolase activity and action on ester bonds (Figure 4.5C). In the rhizome SSH library, the ESTs pertaining to nucleoside phosphate binding, nucleoside binding, nucleic acid binding enzymes, transferases (transferring glycosyl, alkyl, or allyl groups) and enzymes with hydrolase activity were present (Figure 4.5D). Intracellular region, organelle lumen, viral capsid, catalytic complex, envelope and intrinsic component of membrane were the cellular components to which most of the ESTs from the leaf SSH library belonged (Figure 4.5E). In the rhizome SSH library, majority of ESTs belonged to the intracellular region followed by cell periphery, organelle lumen and plasmodesma (Figure 4.5F).

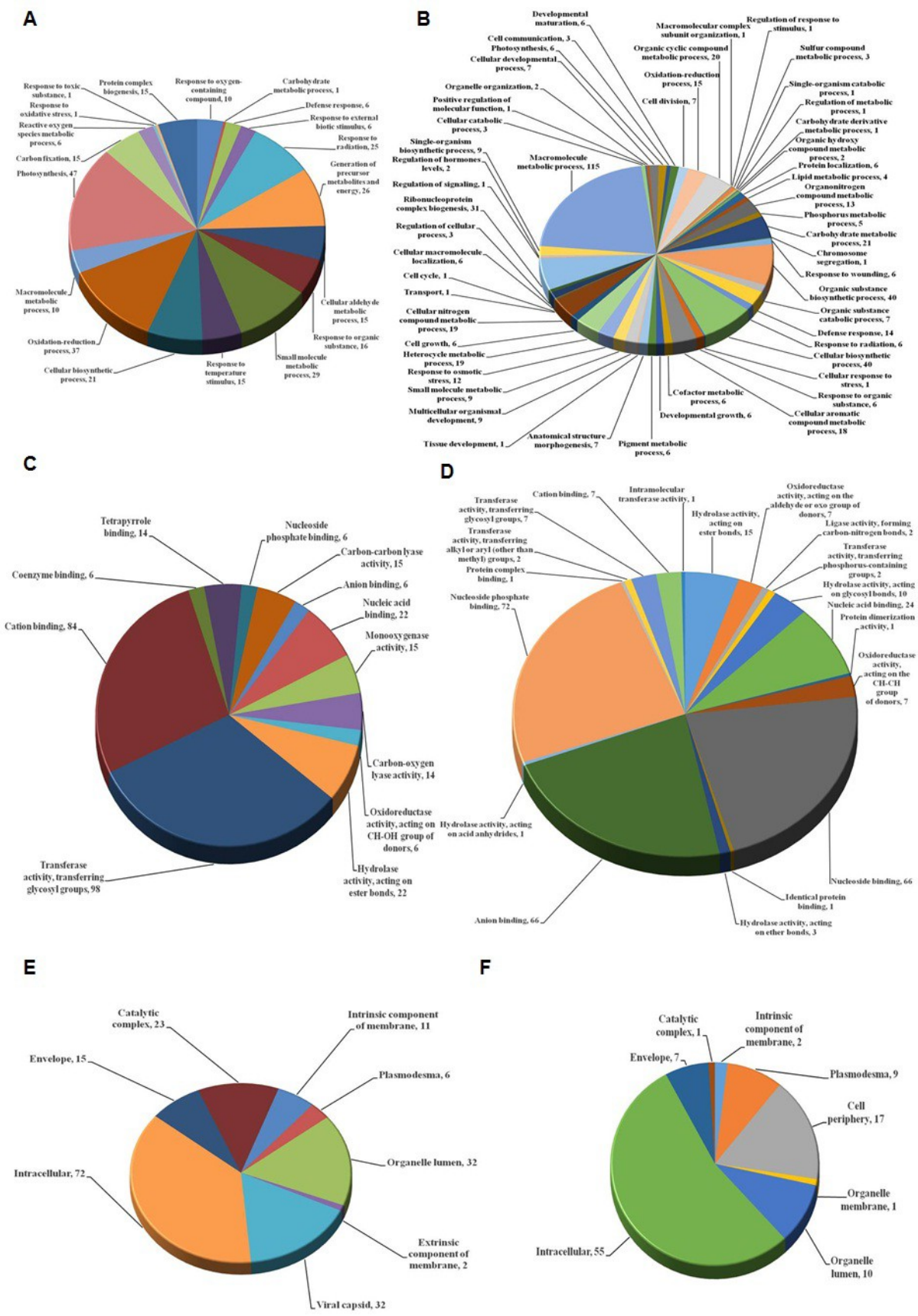


Figure 4.5 Gene ontology (GO) based functional characterization of leaf and rhizome ESTs obtained through SSH. (A, B) leaf and rhizome biological process (C, D) leaf and rhizome

molecular function (E, F) leaf and rhizome cellular component. Source: Reproduced from Gurung et al., 2016.

4.2.1.4. InterPro Scan analysis of ESTs

Searching protein domains using InterProScan tool (Zdobnov and Apweiler, 2001) embedded in Blast2GO software (Conesa et al., 2005) found 337 ESTs with InterPro protein domains and 175 ESTs without InterPro protein domains for the ESTs of leaf SSH library (Figure 4.6A). A total of 233 ESTs had GO annotations (Figure 4.6A). In the rhizome SSH library, the number of ESTs with and without InterPro protein domains were 160 and 214, respectively (Figure 4.6B). 105 ESTs from this library had GO annotations (Figure 4.6B). Nine protein domains viz., nucleotide-diphospho-sugar transferases domain, WD40/YVTN repeat- like-containing domain, cupredoxin/blue (type1) copper domain, FMN-dependent dehydrogenase domain, photosynthetic reaction center, photosystem I PsaG/PsaK domain, Mog1/PsbP, alpha/beta/alpha sandwich, myotoxin/anenome neurotoxin domain and chlorophyll a/b binding protein domain were detected (Figure 4.7A) among the ESTs of leaf SSH library.

Similarly, rhizome SSH library ESTs had protein domains like ubiquitin-activating enzyme repeat, glycosyl transferase, family 35, glyceraldehyde 3-phosphate dehydrogenase, catalytic domain, pectin lyase fold, glycoside hydrolase, catalytic domain, nucleotide-binding alpha-beta plait domain, polyadenylate-binding protein/hyperplastic disc protein, isoprenoid synthase domain, 30s ribosomal protein S13, C-terminal, tetratricopeptide-like helical domain, ricin B lectin domain, Myc-type, basic helix-loop-helix (bHLH) domain, P-loop containing nucleoside triphosphate hydrolase domain and terpenoid cyclases/protein prenyltransferase alpha-alpha toroid (Figure 4.7B).

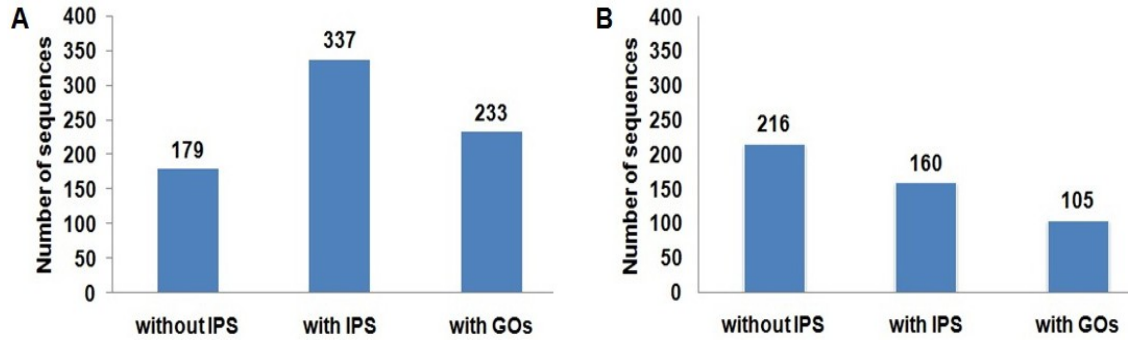


Figure 4.6 InterProScan (IPS) analysis showing ESTs without IPS hits, with IPS hits and with gene ontology (GO) annotations. (A) IPS analysis of leaf cDNA SSH library (B) IPS analysis of rhizome SSH cDNA library. Source: Adapted from Gurung et al., 2016.

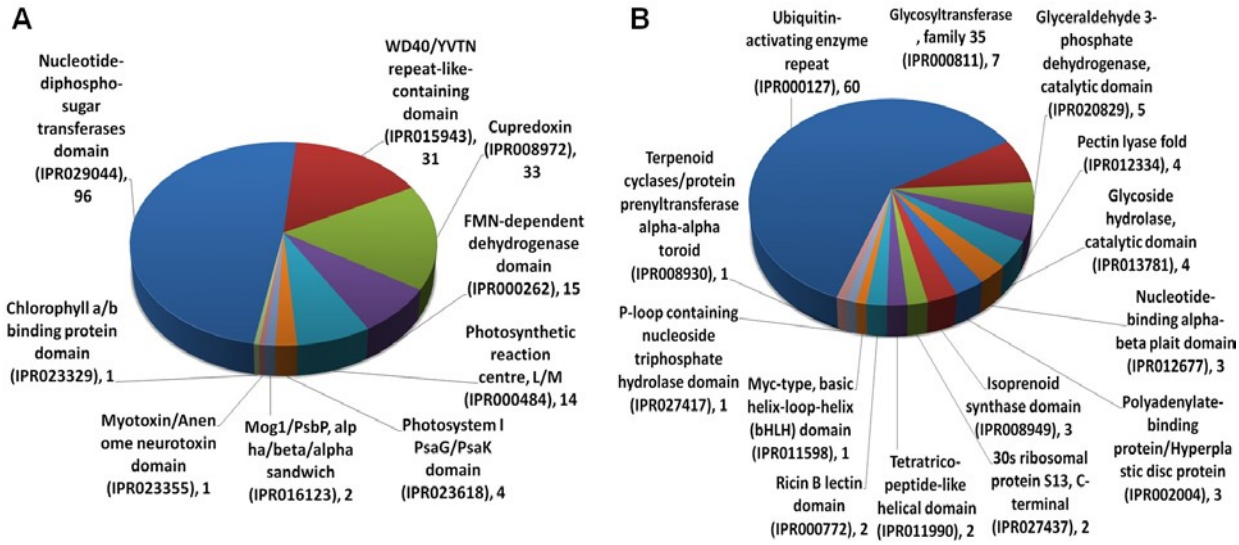


Figure 4.7 Most abundant protein domains detected in ESTs from (A) leaf SSH cDNA library (B) rhizome SSH cDNA library.

4.2.1.5. KEGG analysis of ESTs

In order to find the metabolic pathways represented by the ESTs of the two libraries, KEGG pathway analysis was performed using inbuilt KEGG analysis tool (Kanehisa and Goto, 2000) in the Blast2GO software (Conesa et al., 2005). In the leaf SSH library, ESTs belonging to four KEGG pathways namely nitrogen metabolism, glyoxylate and dicarboxylase metabolism, carbon fixation and galactose metabolism (Figure 4.8A) were present. In the rhizome SSH library, ESTs were annotated KEGG pathways involved in sesquiterpine and triterpenoid biosynthesis, pentose and glucuronase interconversion, porphyrin and chlorophyll metabolism, retinol metabolism, terpenoid backbone biosynthesis, cysteine and methionine metabolism, starch metabolism, riboflavin biosynthesis, steroid biosynthesis, aminobenzoate degradation, purine metabolism and thiamine metabolism (Figure 4.8B).

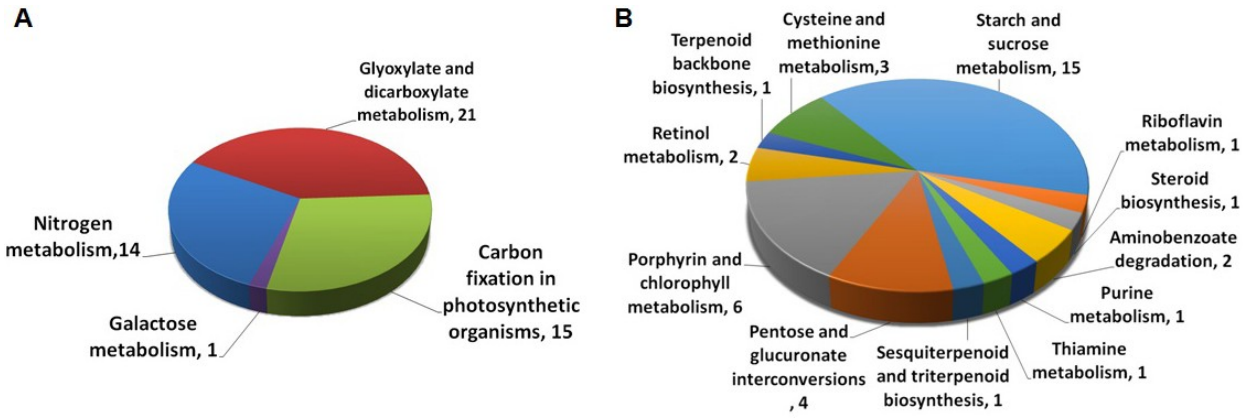


Figure 4.8 KEGG pathway analysis of functionally annotated ESTs from (A) leaf cDNA SSH library (B) rhizome cDNA SSH library.

4.2.1.6. Validation of subtractive cDNA libraries by semiquantitative RT-PCR

The expression studies were performed through semiquantitative RT-PCR to validate the quality of two libraries. The majority of the genes were randomly selected. However, three genes, viz., *PsFPS*, *PsSS* and *PsDS* were chosen as they were from the ginsenoside biosynthetic pathway. Other genes selected were *galactinol synthase 2*, *cell division cycle 20.1*, *metallothionein 3-like protein*, *GBR-5*, *PsbA*, *protein KIAA0664 homologue*, *major latex-like protein*, *RNase-like major storage protein*, *glyceraldehyde-3-phosphate dehydrogenase*, *ankyrin repeat-containing protein*, *β-amylase*, *transcription factor bHLH96*, *ubiquitin-conjugating enzyme E2*, *heat shock protein 70* and *polyubiquitin*. The results showed that the genes selected from the leaf SSH library were overexpressed in leaf tissue, whereas those belonging to the rhizome SSH library showed upregulation in rhizome tissue (Figure 4.9).

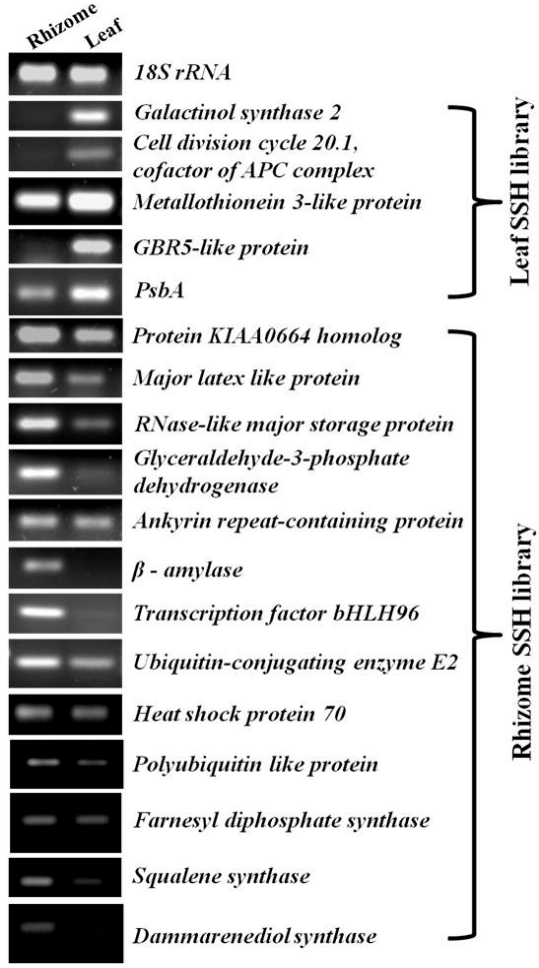


Figure 4.9 Semi-quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) analyses of genes from leaf and rhizome SSH libraries. RNA isolated from the leaf and rhizome tissues were used for this study. 18S rRNA was used as internal control for equal loading. Names of the genes are shown on right side of the panel. Primer details and PCR cycles are given in table 3 (appendix C). Source: Reproduced from Gurung et al., 2016.

4.2.1.7. Identification of rare and novel transcripts by comparing all the unigenes with the available *Panax* transcriptomes

The unigenes from both libraries were compared for sequence homology against SRA available for *Panax* transcriptomes at NCBI to identify novel sequences in our dataset (Leinonen et al., 2011). BLASTN analysis revealed that around 13.25 % of unigenes from the leaf subtractive library were not represented in leaf transcriptome available for *P. ginseng* (Table 4.5). Similarly, 18.12, 23.75, 25 and 6.25 % of the unigenes from the subtractive rhizome library were not represented in the root transcriptomes available for *P. ginseng*, *P. notoginseng*, *P. quinquefolius*, and *P. vietnamensis*, respectively (Table 4.6), indicating a significant fraction of rare and novel transcripts.

Table 4.5 *In silico* comparative analysis of unigenes from leaf SSH library of *P. sokpayensis* with available leaf transcriptome of *Panax* species. Novel unigenes are highlighted with yellow color.

Clone ID	Gene Name	*P. sokpayensis*	*P. ginseng*
PsF1	---NA---	29	100
PsF2	Glycolate oxidase family protein	15	100
PsF3	---NA---	28	200
PsF4	carbonic anhydrase	14	163
PsF5	PsbA	12	71
PsF6	Ribonuclease t2 family protein	21	100
PsF8	---NA---	21	0
PsF9	Ribosomal RNA processing Brix domain protein isoform 2	33	4
PsF10	GBR5-like protein	16	100
PsF12	plastocyanin	29	86
PsF14	galactinol synthase 2	82	100
PsF15	cell division cycle 20.1, cofactor of apc complex-like	31	0
PsF18	galactinol synthase 2	13	100
PsF36	ribulose 1,5 bisphosphate carboxylase oxygenase small subunit	15	100
PsF37	---NA---	40	0
PsF42	---NA---	11	0

PsF46	---NA---	10	100
PsF47	---NA---	1	0
PsF59	---NA---	1	0
PsF61	chloroplast chlorophyll a b binding protein	10	100
PsF62	galactinol synthase	1	100
PsF64	---NA---	1	11
PsF66	---NA---	1	100
PsF67	---NA---	1	0
PsF71	---NA---	1	11
PsF92	galactinol synthase	1	0
PsF116	---NA---	1	100
PsF157	---NA---	1	100
PsF163	---NA---	1	200
PsF164	Glycolate oxidase-like protein	4	100
PsF166	---NA---	1	100
PsF184	Metallothionein-like protein type 3	17	100
PsF210	Glycolate oxidase-like protein	1	100
PsF233	plastocyanin	1	86
PsF246	---NA---	1	2
PsF255	---NA---	1	200
PsF270	---NA---	1	11
PsF305	Photosystem I reaction center subunit psaK, chloroplastic	3	100
PsF312	Glycolate oxidase-like protein	1	100
PsF313	---NA---	1	100
PsF325	Oxygen-evolving enhancer protein-2	1	100
PsF332	---NA---	1	144
PsF351	---NA---	1	200
PsF355	plastocyanin	1	86
PsF358	Photosystem I reaction center subunit psaK, chloroplastic	1	100
PsF383	plastocyanin	1	86
PsF413	---NA---	1	173

PsF414	PsbA	1	71
PsF428	---NA---	1	200
PsF430	carbonic anhydrase	1	100
PsF432	---NA---	1	0
PsF436	---NA---	1	100
PsF442	Metallothionein 3-like protein	1	100
PsF445	---NA---	1	17
PsF453	---NA---	1	100
PsF459	---NA---	1	144
PsF460	---NA---	1	2
PsF463	---NA---	1	144
PsF479	---NA---	1	11
PsF487	Chlorophyll a b binding protein 13,chloroplastic like	1	100
PsF500	oxygen-evolving enhancer protein 2	1	100
PsF508	---NA---	1	11
PsF513	---NA---	1	149
PsF521	---NA---	1	200
PsF528	---NA---	1	114
PsF534	---NA---	1	114
PsF540	plastocyanin A	1	100
PsF548	---NA---	1	9
PsF549	galactinol synthase 1	1	100
PsF556	---NA---	1	4
PsF565	PsbA	1	71
PsF571	---NA---	1	137
PsF575	---NA---	1	200
PsF593	---NA---	1	200
PsF601	---NA---	1	0
PsF606	PsbA	1	68
PsF608	---NA---	1	100
PsF644	plastocyanin	1	100
PsF647	---NA---	1	100

PsF650	---NA---	1	0

Source: Reproduced from Gurung et al., 2016

Table 4.6 *In silico* comparative analysis of unigenes from rhizome SSH library of *P. sokpayensis* with available root/rhizome transcriptome of *Panax* species. Novel unigenes are highlighted with yellow color.

Clone ID	Gene Name	*P. sokpayensis*	*P. ginseng*	*P. notoginseng*	*P. quinquefolius*	*P. vietnamensis* var. *fuscidicus*
PsR2	---NA---	1	11	0	0	10
PsR3	polyadenylate-binding protein	1	40	3	17	100
PsR5	neutral invertase	1	0	1	2	7
PsR6	tetratricopeptide repeat-like superfamily protein	1	62	35	97	100
PsR7	---NA---	5	0	28	0	200
PsR8	---NA---	1	30	0	3	100
PsR9	transcription factor bhlh96 like	1	0	0	1	70
PsR11	major latex-like protein	12	100	100	100	100
PsR12	---NA---	1	0	0	0	0
PsR13	---NA---	1	53	1	0	100
PsR14	arf guanine-nucleotide exchange factor gnom-like	1	7	0	0	100
PsR15	phosphatidylinositol - trisphosphate 3-phosphatase and dual-specificity protein phosphatase pten like isoform x1	1	38	9	21	100
PsR16	---NA---	1	0	1	0	100
PsR17	40s ribosomal protein s18	2	100	51	79	100
PsR19	tetratricopeptide repeat-like superfamily protein	1	62	35	97	100
PsR20	---NA---	2	45	7	14	100
PsR21	---NA---	1	45	7	14	100

PsR22	60s ribosomal protein l18	1	18	4	3	100
PsR23	---NA---	1	100	100	100	100
PsR25	adp-ribosylation factor	1	47	2	13	100
PsR26	ribonuclease like storage protein	7	3	2	0	0
PsR30	aldehyde dehydrogenase	2	11	1	4	100
PsR32	RING U-box superfamily protein	1	13	11	5	100
PsR33	60s ribosomal protein l18	1	15	16	9	100
PsR34	Poly A binding protein	1	40	3	17	100
PsR36	---NA---	1	30	0	3	100
PsR38	---NA---	2	44	0	0	76
PsR41	Ribonuclease like storage protein	1	100	100	100	100
PsR43	zinc finger ccch domain-containing protein 56	1	1	5	0	100
PsR44	---NA---	1	33	4	9	45
PsR46	---NA---	1	100	100	100	100
PsR47	phosphatidylinositol - trisphosphate 3-phosphatase and dual-specificity protein phosphatase pten like isoform x1	1	38	9	21	100
PsR48	---NA---	1	100	100	100	100
PsR54	ubiquitin-activating enzyme e1 1	33	7	1	7	100
PsR55	protochlorophyllide reductase, chloroplastic	4	0	4	0	101
PsR56	Protein KIAA0664 homolog	60	90	2	11	100
PsR57	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR59	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR61	ubiquitin-activating enzyme e1 1	1	7	1	6	100

PsR62	---NA---	1	1	0	0	6
PsR65	---NA---	1	1	0	0	6
PsR67	ubiquitin-activating enzyme e1 1-like isoform x1	1	7	1	7	100
PsR70	ubiquitin-activating enzyme e1 1-like isoform x1	1	7	1	6	100
PsR72	---NA---	5	1	0	0	100
PsR74	ubiquitin-activating enzyme e1 1	1	7	1	6	100
PsR78	ubiquitin-activating enzyme e1 1	1	7	1	6	100
PsR82	ubiquitin-activating enzyme e1 1	1	7	1	6	100
PsR87	ubiquitin-activating enzyme e1 1-like isoform x2	1	6	1	6	100
PsR88	ubiquitin-activating enzyme e1 1-like isoform x1	1	7	1	7	100
PsR90	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR91	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR92	---NA---	1	1	0	0	6
PsR96	ubiquitin-activating enzyme e1 1	1	7	1	6	100
PsR97	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR117	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR122	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR127	protochlorophyllide reductase, chloroplastic	1	0	4	0	101
PsR131	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR137	Protein KIAA0664 homolog	1	89	2	11	100
PsR142	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR156	ubiquitin-activating	1	5	1	7	100

	enzyme e1 1					
PsR164	protochlorophyllide reductase,chloroplastic	1	0	4	6	101
PsR166	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR167	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR172	Protein KIAA0664 homolog	1	90	2	11	100
PsR173	---NA---	1	0	0	0	5
PsR176	---NA---	1	0	0	0	5
PsR178	---NA---	1	0	0	0	5
PsR179	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR183	Protein KIAA0664 homolog	1	90	2	12	100
PsR192	ubiquitin-activating enzyme e1 1	1	7	1	6	100
PsR194	---NA---	1	0	0	0	5
PsR195	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR196	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR198	---NA---	1	0	0	0	5
PsR199	Protein KIAA0664 homolog	1	90	2	11	100
PsR202	ubiquitin-activating enzyme e1 1-like isoform x1	1	7	1	7	100
PsR209	Protein KIAA0664 homolog	1	90	2	11	100
PsR222	ubiquitin-activating enzyme e1 1	1	7	1	7	100
PsR229	ubiquitin-conjugating enzyme E2	1	100	24	100	100
PsR232	---NA---	4	0	0	0	6
PsR235	---NA---	1	100	84	100	100
PsR236	major latex-like protein	1	100	100	100	100
PsR238	---NA---	4	100	85	100	100
PsR239	Ribonuclease-like	1	100	100	100	100

	storage protein					
PsR241	polyadenylate-binding protein rbp47 isoform x1	3	37	20	7	100
PsR242	cbl-interacting serine threonine-protein kinase 7	2	0	0	0	61
PsR243	---NA---	1	3	0	0	11
PsR244	---NA---	5	1	0	1	57
PsR245	Probable chromatin-remodeling complex ATPase chain	1	6	0	1	100
PsR247	---NA---	1	8	4	10	93
PsR248	---NA---	5	100	100	100	100
PsR249	---NA---	1	4	3	16	67
PsR250	---NA---	1	8	4	10	96
PsR251	ubiquitin-conjugating enzyme	1	100	22	100	100
PsR252	Ribonuclease like storage protein	1	100	100	100	100
PsR253	---NA---	1	0	0	0	14
PsR254	---NA---	1	8	4	10	93
PsR255	ribonuclease-like storage protein	1	100	100	100	100
PsR258	---NA---	2	1	100	93	100
PsR259	---NA---	3	42	0	38	0
PsR261	glyceraldehyde-3-phosphate dehydrogenase	5	100	41	100	100
PsR262	rhodanese-like domain-containing protein6	2	0	0	0	14
PsR264	---NA---	1	0	0	0	0
PsR267	rnase-like major storage protein	1	100	100	100	100
PsR270	ribonuclease like storage protein	1	100	100	100	100
PsR272	ribonuclease like storage protein	1	100	100	100	100
PsR276	---NA---	3	71	4	2	100
PsR280	probable	4	0	5	24	26

	polygalacturonase					
PsR282	---NA---	1	100	84	100	100
PsR284	---NA---	1	0	0	0	0
PsR289	---NA---	2	100	100	100	100
PsR292	---NA---	2	12	8	4	100
PsR293	ribosomal protein l13a	2	68	8	22	100
PsR296	---NA---	1	70	60	31	100
PsR302	60s ribosomal protein l21	3	100	17	58	100
PsR306	Type II ribosome-inactivating protein cinnamomin	1	4	7	0	100
PsR307	polyubiquitin like protein	1	100	100	100	100
PsR328	---NA---	1	0	0	0	0
PsR330	riboflavin synthase alpha chain	1	9	17	27	100
PsR334	---NA---	1	0	1	0	100
PsR337	polyadenylate-binding protein	1	40	3	17	100
PsR338	---NA---	8	100	45	11	100
PsR339	40s ribosomal protein s26	20	37	3	33	100
PsR340	ankyrin repeat-containing protein	5	3	6	2	100
PsR341	---NA---	1	37	3	33	100
PsR343	---NA---	1	100	100	100	100
PsR344	---NA---	4	39	6	12	100
PsR345	alpha 1,4 glucan phosphorylase L isozyme, chloroplastic	7	196	153	132	100
PsR350	---NA---	2	2	0	2	100
PsR353	major latex-like protein	1	100	100	100	100
PsR354	ankyrin repeat-containing protein	1	2	5	2	100
PsR356	beta-amylase	4	101	14	100	100
PsR359	---NA---	1	13	9	7	67
PsR364	---NA---	2	0	0	1	20

PsR368	actin	6	106	101	113	100
PsR372	---NA---	1	2	0	2	100
PsR374	60s ribosomal protein l9	1	34	1	3	100
PsR375	---NA---	2	4	103	0	100
PsR377	60s ribosomal protein l9	1	36	1	3	100
PsR378	---NA---	3	100	1	100	100
PsR383	---NA---	1	0	0	0	5
PsR384	type 2 ribosome-inactivating protein cinnamomin I precursor	1	4	10	1	100
PsR396	---NA---	2	100	100	100	100
PsR398	---NA---	1	0	0	0	6
PsR400	Ribonuclease like storage protein	1	100	100	100	100
PsR403	---NA---	1	44	0	0	76
PsR406	Heat shock protein 70	1	100	100	100	100
PsR412	maternal effect embryo arrest 18 protein	2	101	100	100	100
PsR418	---NA---	1	3	0	0	2
PsR425	---NA---	3	91	33	100	100
PsR433	---NA---	1	10	1	0	100
PsR437	squalene synthase	1	7	40	16	100
PsR438	dammarenediol synthase	1	2	100	10	100
PsR439	---NA---	1	0	0	0	0
PsR440	farnesyl diphosphate synthase	1	9	101	43	100
PsR441	---NA---	1	0	0	0	0
PsR442	---NA---	1	0	0	0	0
PsR443	---NA---	1	0	0	0	0
PsR444	squalene synthase	1	8	31	13	100

Source: Reproduced from Gurung et al., 2016

4.2.2. Genes cloned through degenerate primer approach

Partial fragments of genes belonging to mevalonate, non – mevalonate, ginsenoside biosynthetic pathway and that of *CS*, which belongs to phytosterol biosynthetic pathway, were amplified using degenerate primers (Table 4.7, Figure 4.10). The degenerate primer set and PCR parameters used for the respective gene are given in table 1 (Appendix C).

Table 4.7 Partial genes pertaining to ginsenoside biosynthesis amplified using degenerate primers.

Gene name	GenBank accession number	Length (bp)	Total score	Query cover	min. e value	Identity (%)
Acetyl-CoA C-acetyltransferase (AACT)	KY513109	653	444	99	9e-158	100
3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS)	MF682466	265	182	98	1e-58	98
Mevalonate kinase (MVK)	MF682465	488	305	99	3e-101	99
Phosphomevalonate kinase (PMVK)	KY513110	402	222	100	7e-68	88
1-deoxy-D-xylulose-5-phosphate synthase (DXS)	KY513105	265	172	99	3e-53	92
1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)	KY513106	332	204	99	2e-65	92
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (CMS)	KY513101	481	327	99	5e-111	98
4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK)	MF682467	467	307	99	5e-102	95
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS)	KY513108	337	221	98	3e-71	97
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS)	KY513103	343	233	98	4e-76	98

4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR)	KY513104	500	323	99	8e-108	94
Isopentenyl diphosphate isomerase 2 (IDI2)	MF682461	333	220	99	5e-71	97
Squalene epoxidase (SE)	-	270	188	98	2e-55	99
β-amyrin synthase (β-AS)	MF682464	513	347	98	5e-113	97
Protopanaxadiol synthase (PPDS)	MF682462	307	217	99	1e-66	100
Protopanaxatriol synthase (PPTS)	MF682463	423	285	99	1e-92	99
Cycloartenol synthase (CS)	-	314	217	99	1e-64	98

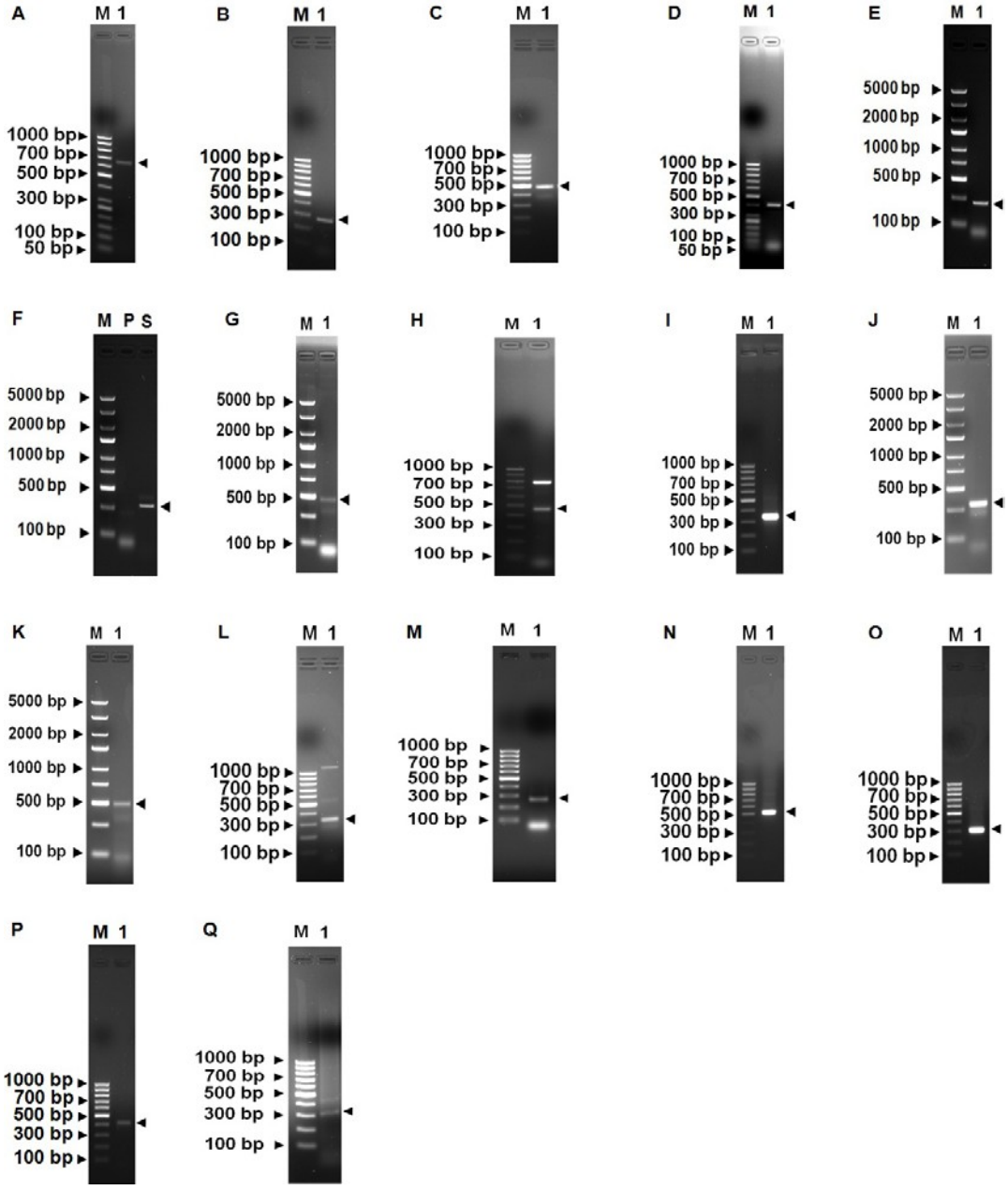


Figure 4.10 Partial fragments of genes belonging to mevalonate, non – mevalonate, ginsenoside biosynthetic and phytosterol biosynthetic pathway amplified using degenerate primers (A) *Acetyl-CoA C-acetyltransferase* (B) *3-hydroxy-3-methylglutaryl coenzyme A synthase* (C) *Mevalonate kinase* (D) *Phosphomevalonate kinase* (E) *1-deoxy-D-xylulose-5-phosphate synthase* (F) *1-deoxy-D-xylulose 5-phosphate reductoisomerase* (G) *2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase* (H) *4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase* (I) *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (J) *4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (K) *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* (L) *Isopentenyl diphosphate isomerase* (M) *Squalene epoxidase* (N) *β-amyrin synthase* (O) *Protopanaxadiol synthase* (P) *Protopanaxatriol synthase* (Q) *Cycloartenol synthase*. M: Marker, lane 1: amplicon amplified using degenerate primers, P: Primary PCR, S: secondary PCR.

4.3. Cloning and characterization of full length genes through RACE

Seven genes were cloned to full length following RACE. They were *PsFPS*, *PsSS*, *PsDS*, *PsCMK*, *PsPMVK*, *PsSE* and *PsCS*. These cDNAs were further characterized and the results are presented below:

4.3.1. *PsFPS*

4.3.1.1. Construction of full length *PsFPS* using RACE

5′ and 3′ RACE fragments were 545 and 626 bp long, respectively (Figure 4.11). A full-length *P. sokpayensis FPS* (*PsFPS*) of 1437 bp constructed through the alignment of these RACE fragments were submitted to NCBI with accession number KT936527. This full-length cDNA comprised of 1026 bp long ORF between 143 bp and 1168 bp, 5′ untranslated region (UTR) of 142 bp and 3′ UTR of 234 bp followed by a poly A tail (Figure 4.12).

4.3.1.2. *In silico* analysis of full length cDNA of *PsFPS*

Three identical polyadenylation signals (PASs) were detected in 3′ UTR at positions 1334 – 1339, 1338 – 1343 and 1383 – 1388 (Figure 4.12). *In silico* analysis predicted a molecular weight of 39.6 kDa for deduced 342 amino acids of PsFPS and theoretical pI of 5. Seven conserved features, viz., substrate binding pocket (22/22 residues), substrate – Mg^{2+} binding site (13/13 residues), active site lid residues (27/28 residues), chain length determination region (10/10 residues), catalytic residues (8/8 residues), first aspartate rich motif (FARM) (7/7 residues) and second aspartate rich motif (SARM) (6/6 residues) were mapped to PsFPS (Figure 4.13). DDXXD, a highly conserved motif of FARM and SARM was also found on PsFPS. Multiple sequence alignment of FPSs from across the genera and families from the plant kingdom revealed that all the domains present in this protein were highly conserved (Figure 4.13). Phylogenetic analysis revealed that FPSs of *Panax* species were highly related to each other and thus clustered close to one another in the cladogram (Figure 4.14). The conserved domain search annotated PsFPS with trans – isoprenyl diphosphate synthase, head to tail domain containing protein belonging to Isoprenoid_Biosyn_C1 superfamily (Figure 4.15A). The predicted secondary structure contained 57.89 % alpha helix, 10.82 % extended strand, 7.60 % beta turn and 23.68 % random coil (Figure 4.15B). Kyte and Doolittle hydropathicity plot and TMHMM program suggested PsFPS to be a cytosolic protein (Figures 4.15C, D).

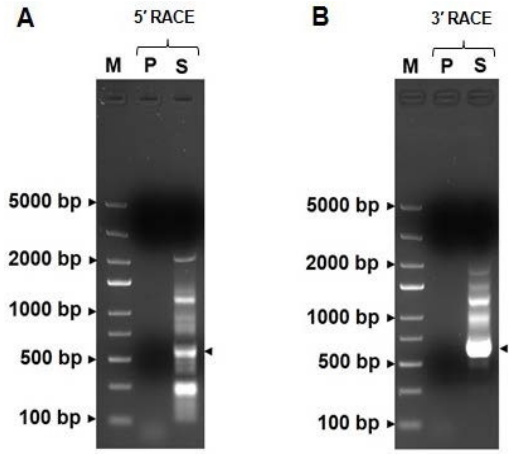


Figure 4.11 Agarose gel picture showing PCR products of *P. Sokpayensis farnesyl pyrophosphate synthase* (A) 5′ RACE (B) 3′ RACE. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

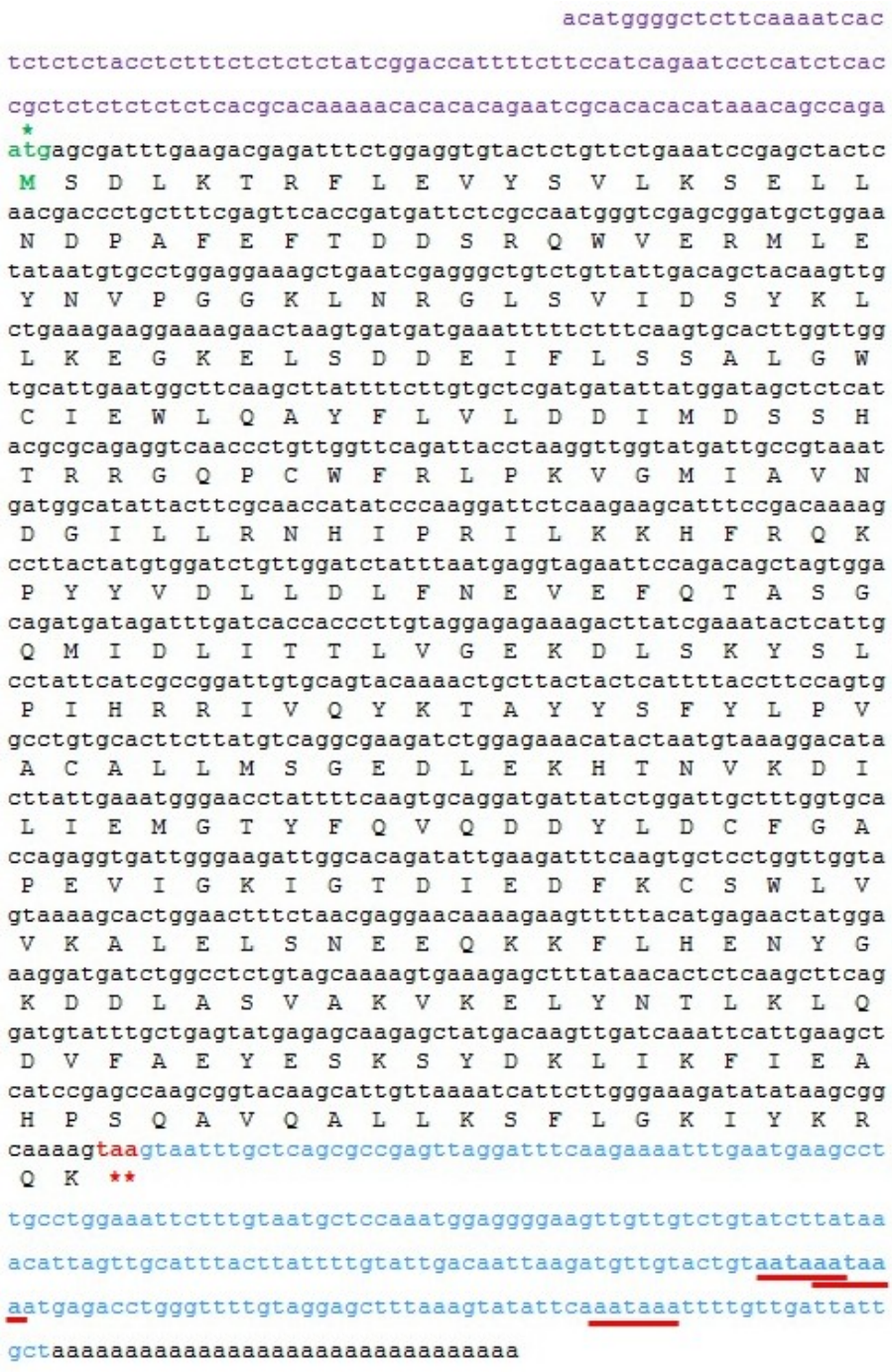


Figure 4.12 Nucleotide and deduced amino acid sequence of *P. Sokpayensis farnesyl pyrophosphate synthase* (*PsFPS*, KT936527). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**" symbol. 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end. Putative poly A signals are bold and underlined by red bars.

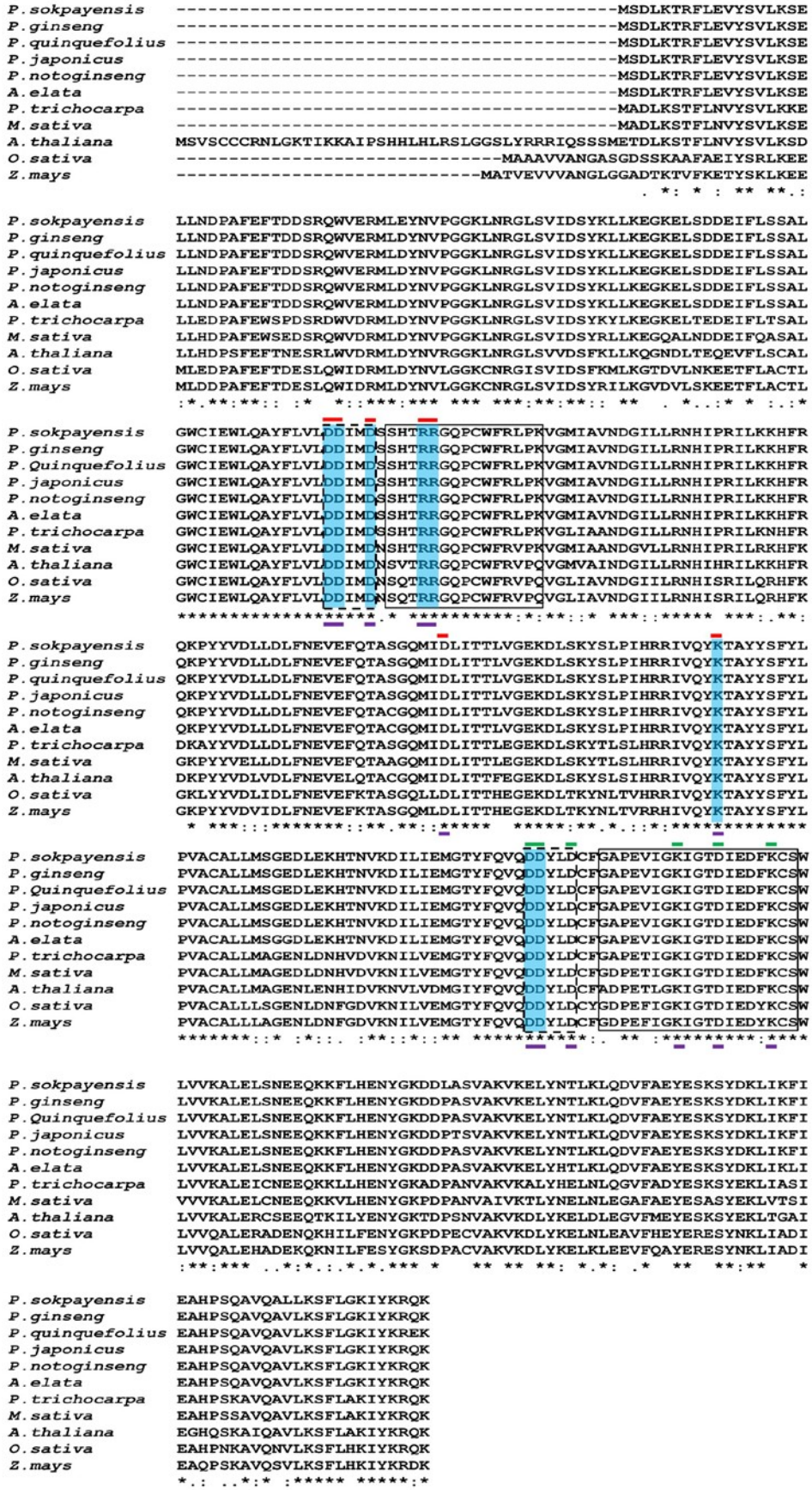


Figure 4.13 Multiple sequence alignment of deduced amino acid sequences of FPSs from the *Panax* species and other plants. Active site lid residues are shown in boxes,

catalytic residues are highlighted in blue, purple lines indicate residues that form substrate Mg2+ binding region. First aspartate rich region (FARM) and second aspartate rich region (SARM) are indicated with red and green lines respectively. FARM and SARM contain highly conserved motif DDXXD which are shown in dashed boxes. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75532), *P. ginseng* (AAY87903), *P. japonicus* (AKN52395), *P. quinquefolius* (ADJ68004), *P. notoginseng* (AGS79228), *A. thaliana* (OAO94430), *Aralia elata* (ADK12004), *P. trichocarpa* (XP_002308751), *O. sativa* (BAA19856), *Zea mays* (AFW83683), *Medicago sativa* (ADC32809).

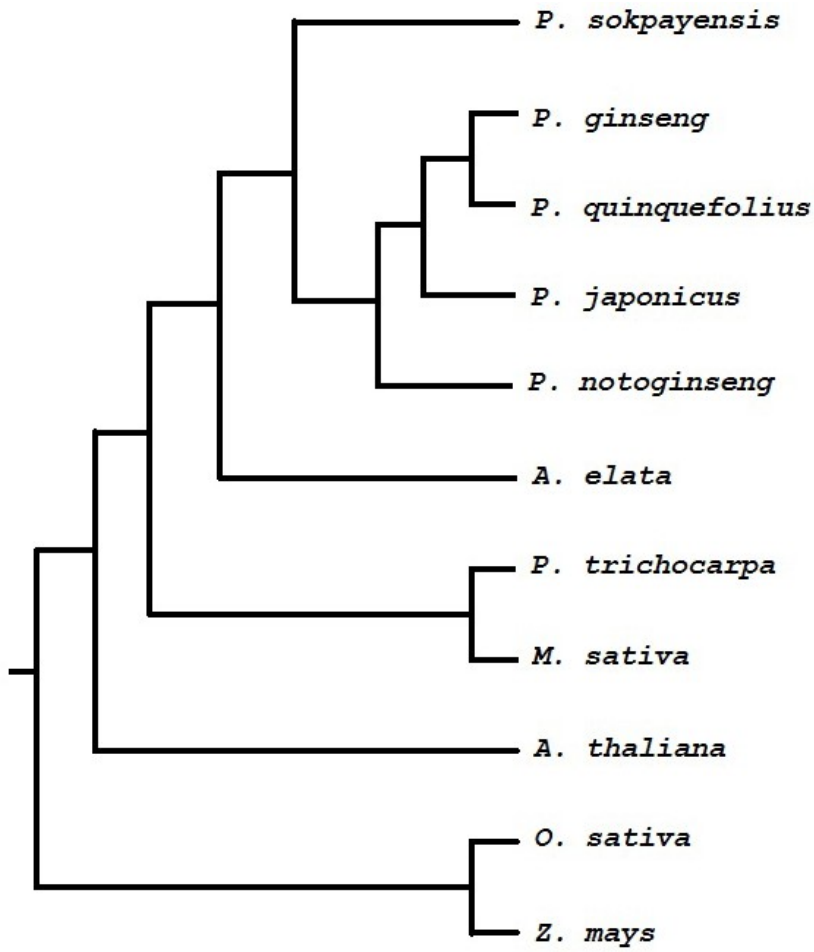


Figure 4.14 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of FPSs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75532), *P. ginseng* (AAY87903), *P. japonicus* (AKN52395), *P. quinquefolius* (ADJ68004), *P. notoginseng* (AGS79228), *A. thaliana* (OAO94430), *A. elata* (ADK12004), *P. trichocarpa* (XP_002308751), *O. sativa* (BAA19856), *Z. mays* (AFW83683), *M. sativa* (ADC32809).

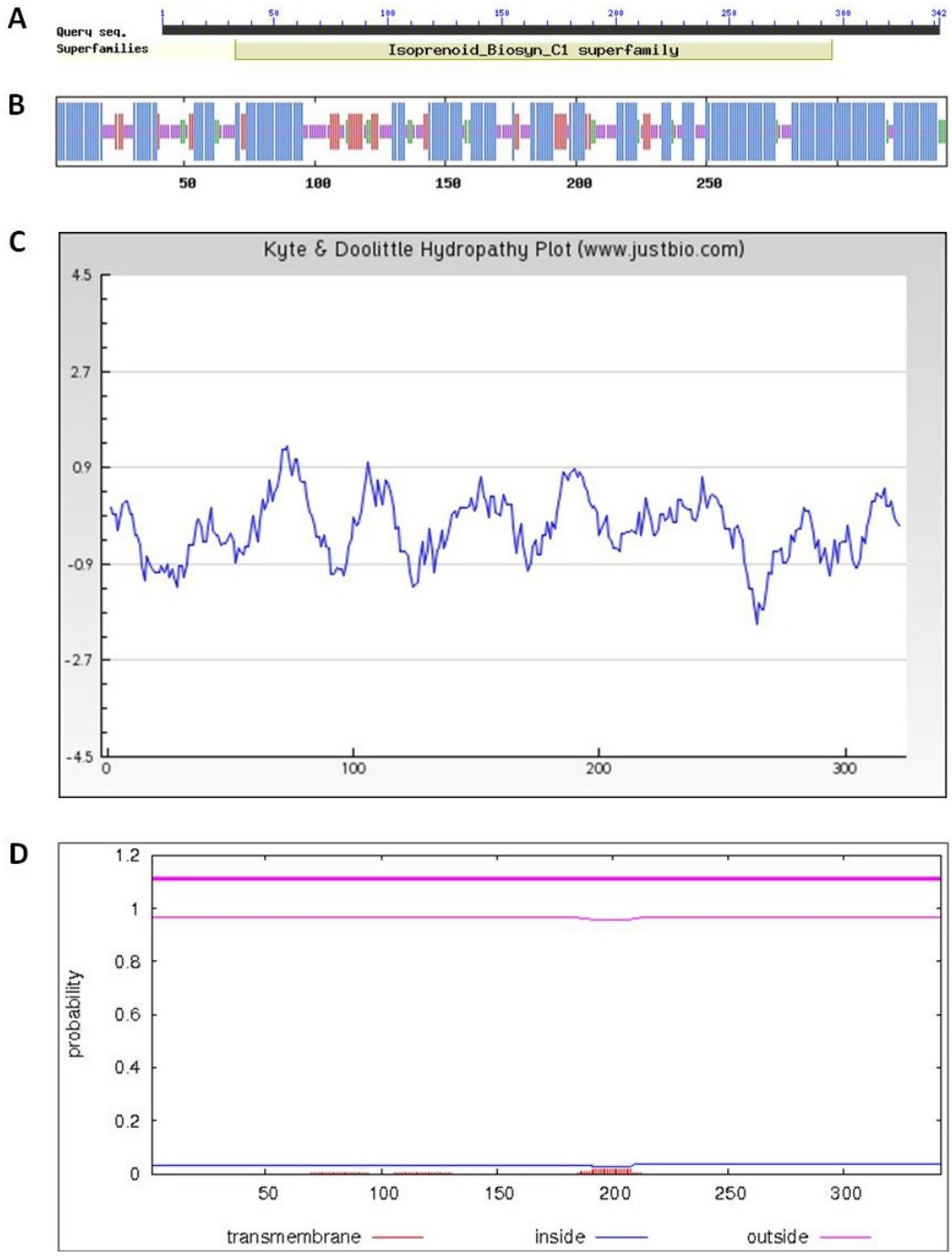


Figure 4.15: (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of PsFPS using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsFPS (D) Prediction of transmembrane region using TMHMM program.

4.3.2. *PsSS*

4.3.2.1. Construction of full length *PsSS* using RACE

Partial fragment of *PsSS* (accession number JZ822884) was used to design RACE primers. 5′ and 3′ RACE fragments were 331 and 664 bp long, respectively (Figure 4.16). A full-length *PsSS* of 1507 bp constructed through the alignment of these RACE fragments were submitted to NCBI with accession number KT936528 (Figure 4.17). This full-length cDNA comprised of 1245 bp long ORF between 96 bp and 1340 bp, 5′ UTR of 95 bp and 3′ UTR of 143 bp containing putative poly A signal, AATAAA followed by a poly A tail (Figure 4.17).

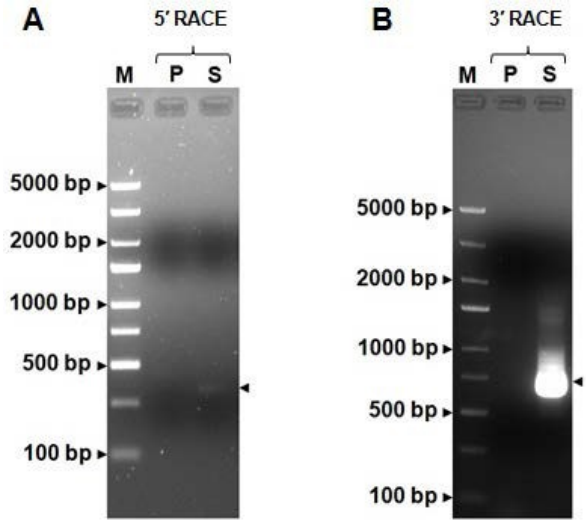


Figure 4.16 Agarose gel picture showing RACE products of *PsSS* (A) 5′ RACE (B) 3′ RACE. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

4.3.2.2. *In silico* analysis of full length cDNA of *PsSS*

Putative PAS was detected in 3′ UTR at positions 1395 – 1400 bp (Figure 4.17). *In silico* analysis predicted a molecular weight of 47.13 kDa for a deduced 415 amino acids of PsSS and theoretical pI of 6.13. Six conserved features, viz., substrate binding pocket (18/18 residues), substrate – Mg^{2+} binding site (10/10 residues), active site lid residues (9/9 residues), catalytic residues (15/15 residues), first aspartate rich motif (FARM) (5/5 residues) and second aspartate rich motif (SARM) (5/5 residues) were mapped to PsSS (Figure 4.18). DXXXD, a highly conserved motif of FARM and SARM was also found on PsSS. Multiple sequence alignment of SSs from across the genera and families from the plant kingdom revealed that all the domains present in this protein were highly conserved (Figure 4.18). Phylogenetic analysis revealed that SSs of *Panax* species clustered close to

one another in the cladogram (Figure 4.19). The conserved domain search annotated PsSS with trans – isoprenyl diphosphate synthase, head to head domain containing protein belonging to Isoprenoid_Biosyn_C1 superfamily (Figure 4.20A). The predicted secondary structure contained 66.75 % alpha helix, 7.71 % extended strand, 6.51 % beta turn and 19.04 % random coil (Figure 4.20B). Both Kyte and Doolittle hydropathicity plot and TMHMM program predicted transmembrane helix at the C-terminal of PsSS (Figures 4.20C, D).

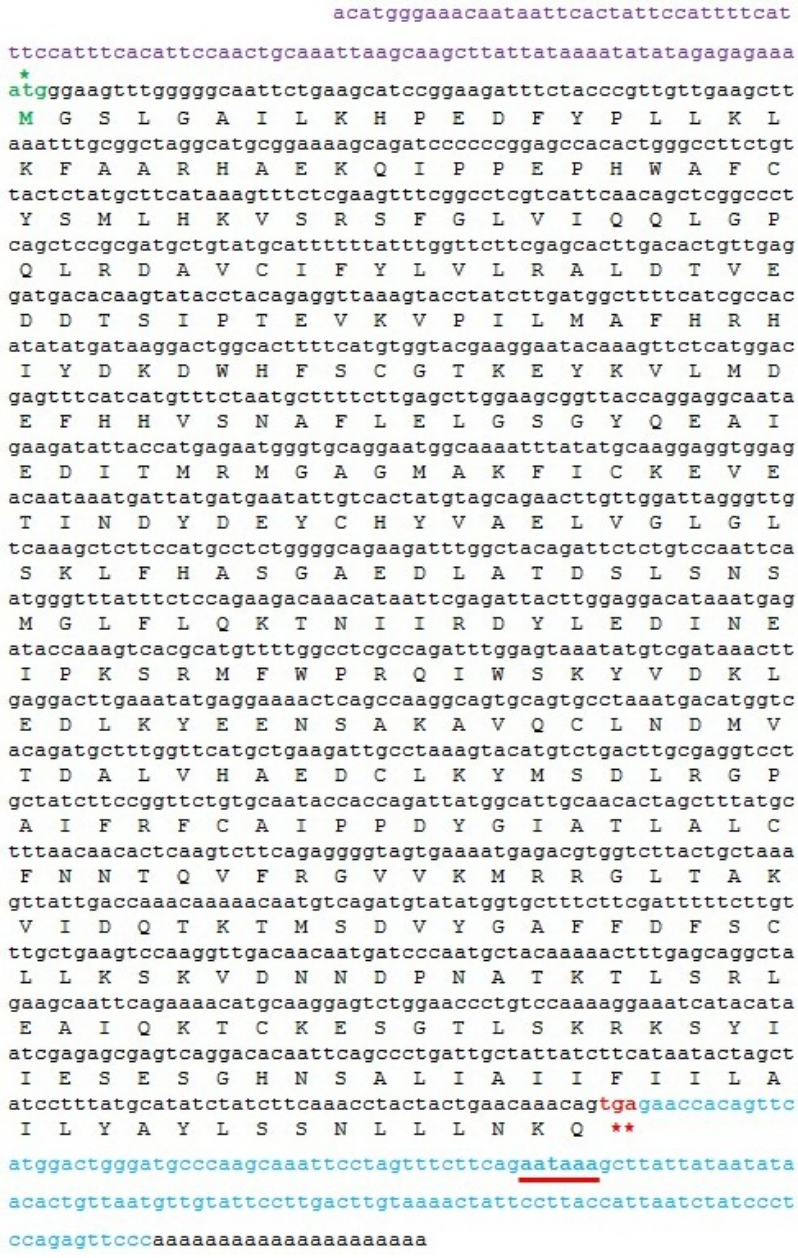


Figure 4.17 Nucleotide and deduced amino acid sequence of *P. sokpayensis squalene synthase* (*PsSS*, KT936528). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**" symbol. 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end. Putative poly A signal is bold and underlined by a red bar.

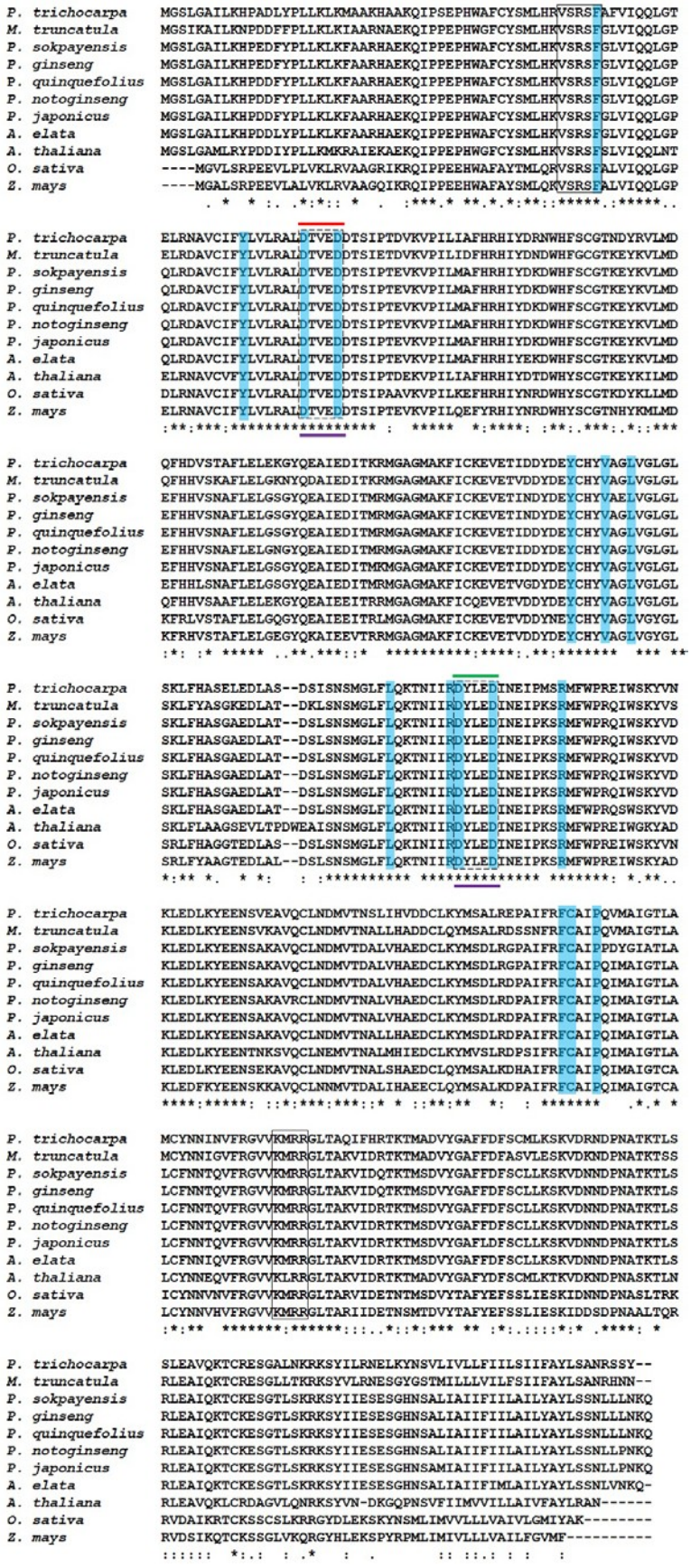


Figure 4.18: Multiple sequence alignment of deduced amino acid sequences of SSs from the *Panax* species and other plants. Active site lid residues are shown in boxes, catalytic

residues are highlighted in blue, purple lines indicate residues that form substrate Mg2+ binding region. First aspartate rich region (FARM) and second aspartate rich region (SARM) are indicated with red and green lines respectively. FARM and SARM contain highly conserved motif DXXXD which are shown in dashed boxes. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75533), *P. ginseng* (BAA24289), *P. japonicus* (ALB38664), *P. quinquefolius* (AED99863), *P. notoginseng* (ABA29019), *A. thaliana* (OAO97129), *A. elata* (ADC32654), *P. trichocarpa* (XP_002313765), *O. sativa* (XP_015630544), *Z. mays* (NP_001104839), *M. truncatula* (XP_003607040).

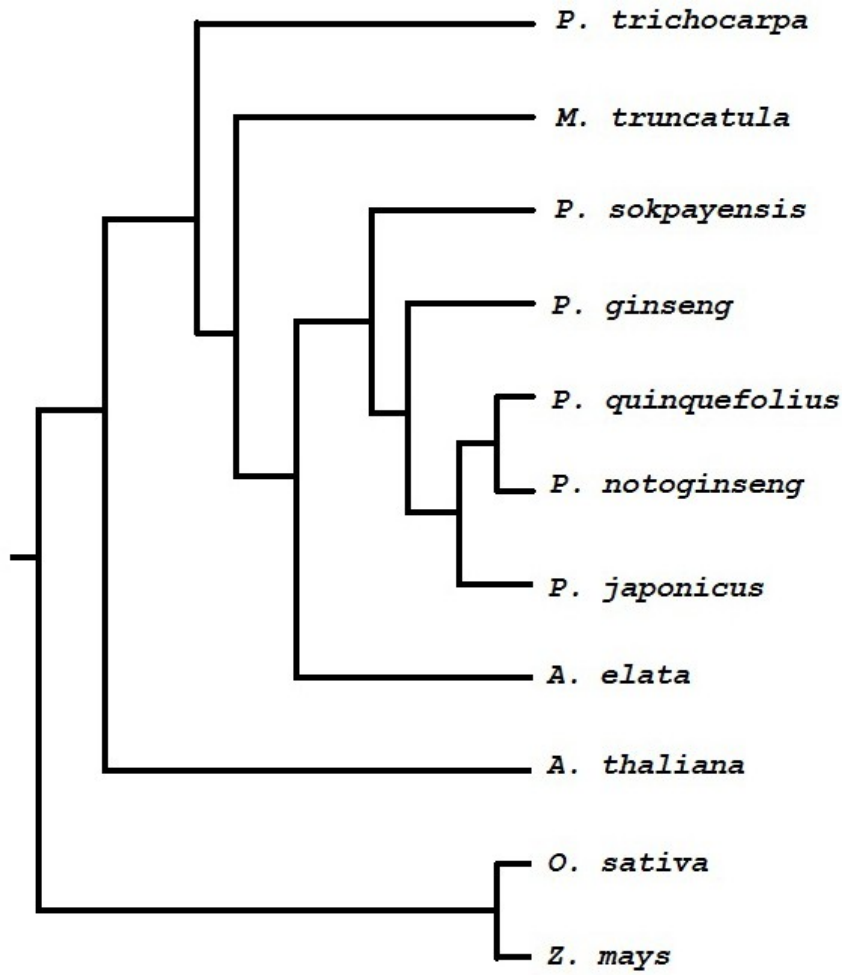


Figure 4.19 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of SSs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75533), *P. ginseng* (BAA24289), *P. japonicus* (ALB38664), *P. quinquefolius* (AED99863), *P. notoginseng* (ABA29019), *A. thaliana* (OAO97129), *A. elata* (ADC32654), *P. trichocarpa* (XP_002313765), *O. sativa* (XP_015630544), *Z. mays* (NP_001104839), *M. truncatula* (XP_003607040).

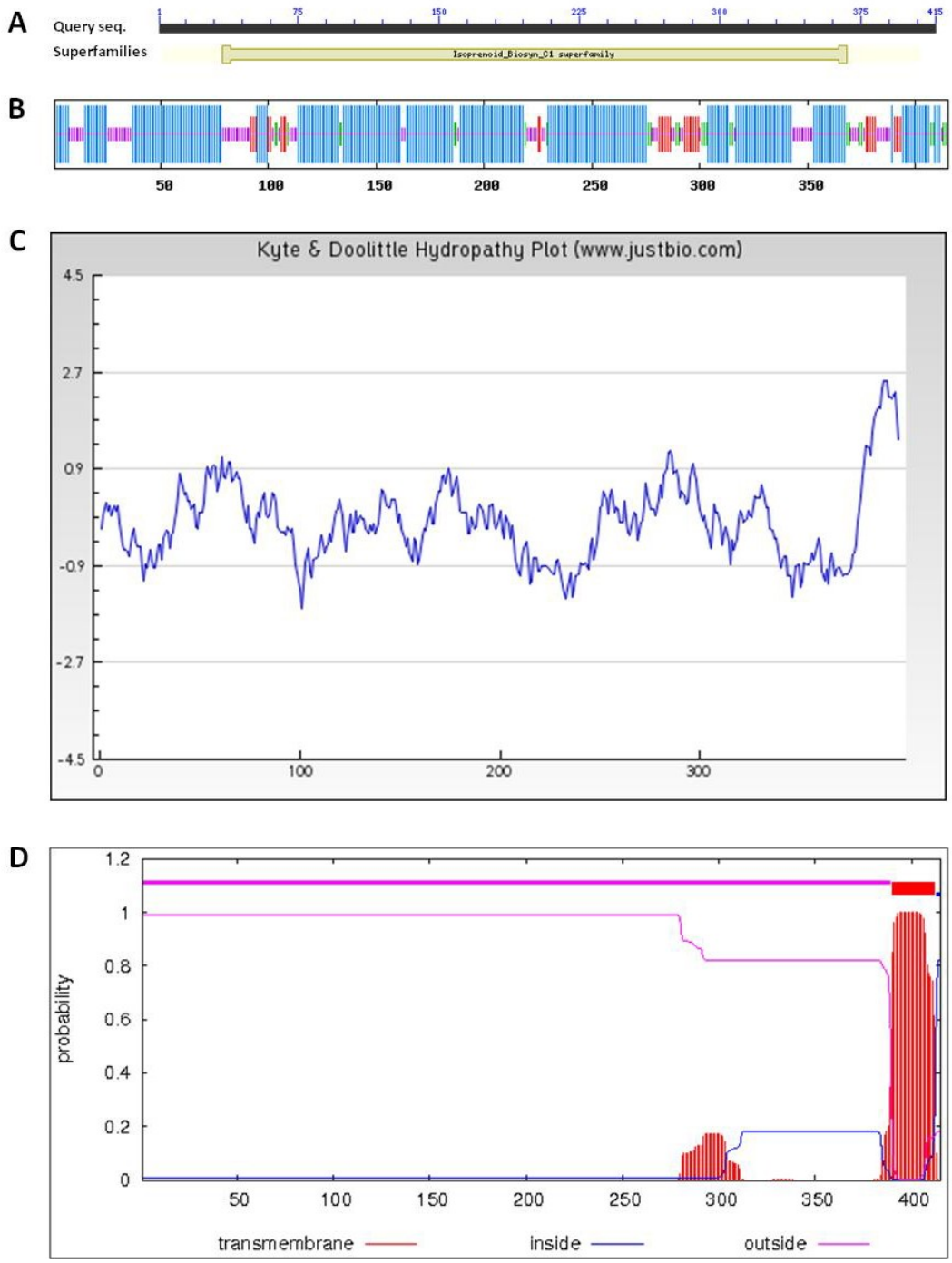


Figure 4.20 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of Squalene Synthase (SS) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsSS. (D) Prediction of transmembrane region using TMHMM program.

4.3.3. *PsDS*

4.3.3.1. Construction of full length *PsDS* using RACE

5′ and 3′ RACE fragments were 810 and 1822 bp long, respectively (Figure 4.21). A full-length *P. sokpayensis DS* (*PsDS*) of 2663 bp constructed through the alignment of these RACE fragments were submitted to NCBI with accession number KU196775 (Figure 4.22). This full-length cDNA comprised of 2310 bp long ORF between 48 bp and 2359 bp, 5′ UTR of 48 bp and 3′ UTR of 277 bp followed by a poly A tail (Figure 4.22).

4.3.3.2. *In silico* analysis of full length cDNA of *PsDS*

Putative PASs were detected in 3′ UTR at positions 2505 – 2510, 2600 – 2605 and 2607 – 2612 bp (Figure 4.22). *In silico* analysis predicted a molecular weight of 88.37 kDa for a deduced 769 amino acids of PsDS and theoretical pI of 6.47. Two conserved features, viz., catalytic acid (1/1 residue) and active site cavity (34/34 residues) were mapped to PsDS (Figure 4.23). Multiple sequence alignment of DSs from across the genera and families from the plant kingdom revealed that all the domains present in this protein were highly conserved (Figure 4.23). Phylogenetic analysis found DSs from the *Panax* species forming distinct clade (Figure 4.24). The conserved domain search annotated PsDS with squalene cyclase (SQCY) domain subgroup 1 containing protein belonging to SQHop_cyclase_C superfamily (Figure 4.25A). The predicted secondary structure contained 35.50 % alpha helix, 17.43 % extended strand, 12.35 % beta turn and 34.72 % random coil (Figure 4.25B). Kyte and Doolittle hydropathicity plot did not predict a distinct transmembrane region in PsDS (Figure 4.25C). However, TMHMM analysis shows a single transmembrane helix near the C terminal of PsDS (Figure 4.25D).

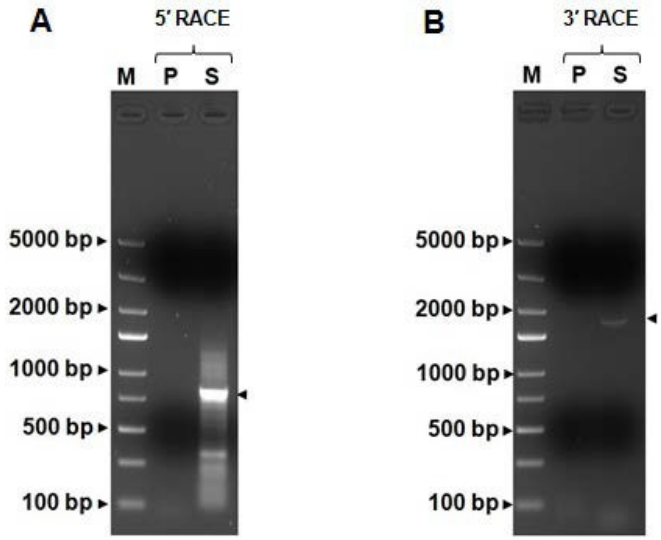


Figure 4.21 Agarose gel picture showing RACE products of *PsDS* (A) 5′ RACE (B) 3′ RACE. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

```
                                                acatggggagccatctaggtgcaccacataccctcaagtagtagaaga
                                                *
atgtggaagctgaaggttgctcaaggaaatgatccatatttgtatagcactaacaacttt
 M  W  K  L  K  V  A  Q  G  N  D  P  Y  L  Y  S  T  N  N  F
gttggcaggcaatattgggagtttcagcccgatgctggttctccagaagagagggaagag
 V  G  R  Q  Y  W  E  F  Q  P  D  A  G  S  P  E  E  R  E  E
gttgaaaaagcacgcaaggattatgtaaacaataagaagctacatggaattcatccatgc
 V  E  K  A  R  K  D  Y  V  N  N  K  K  L  H  G  I  H  P  C
agtgatatgctgatgcgcaggcagcttattaaagaaagtggaatcgatctcctaagcatg
 S  D  M  L  M  R  R  Q  L  I  K  E  S  G  I  D  L  L  S  M
ccgccggtgagattagatgaaaacgaacaagtgaactacgatgcagttacaaccgctgtg
 P  P  V  R  L  D  E  N  E  Q  V  N  Y  D  A  V  T  T  A  V
aagaaagctcttcgattgaaccgggcaattcaagcacacgatggtcactggccagctgaa
 K  K  A  L  R  L  N  R  A  I  Q  A  H  D  G  H  W  P  A  E
aatgcaggctctttactttatacacctccccttatcattgccctatatatcagcggaacg
 N  A  G  S  L  L  Y  T  P  P  L  I  I  A  L  Y  I  S  G  T
attgacactattctgacaaaacaacacaagaaggaactgattcgcttcgtttacaaccat
 I  D  T  I  L  T  K  Q  H  K  K  E  L  I  R  F  V  Y  N  H
caaaatgaggatggtggatggggatcctatattgaggggcacagcacgatgattgggtca
 Q  N  E  D  G  G  W  G  S  Y  I  E  G  H  S  T  M  I  G  S
gtacttagctacgtgatgttacgtttgctaggagaaggattagctgagtctgatggtgga
 V  L  S  Y  V  M  L  R  L  L  G  E  G  L  A  E  S  D  G  G
aatggtgcagttgagagaggccggaagtggatacttgatcatggaggtgcagccagcata
 N  G  A  V  E  R  G  R  K  W  I  L  D  H  G  G  A  A  S  I
ccctcttggggaaagacttatctagcggtgcttggagtatatgagtgggaaggggtgcaac
 P  S  W  G  K  T  Y  L  A  V  L  G  V  Y  E  W  E  G  C  N
ccgctgcccccagaattctggcttttcccttcaagttttccttttcatccagcaaaaatg
 P  L  P  P  E  F  W  L  F  P  S  S  F  P  F  H  P  A  K  M
tggatctactgccggtgcacttacatgccaatgtcgtatttgtatgggaagagatatcat
 W  I  Y  C  R  C  T  Y  M  P  M  S  Y  L  Y  G  K  R  Y  H
ggaccaataaccgatcttgttttatctttgagacaagaaatttacaacattccttatgag
 G  P  I  T  D  L  V  L  S  L  R  Q  E  I  Y  N  I  P  Y  E
cagataaagtggaatcaacagcgccataactgttgcaaggaggatctctactaccctcat
 Q  I  K  W  N  Q  Q  R  H  N  C  C  K  E  D  L  Y  Y  P  H
acccttgtacaagacctggtttgggatggtcttcactactttagtgaaccattcctcaaa
 T  L  V  Q  D  L  V  W  D  G  L  H  Y  F  S  E  P  F  L  K
cgttggcccttcaacaaactgcgaaaaagaggtctaaaaagagtggttgaaccaatgcgc
 R  W  P  F  N  K  L  R  K  R  G  L  K  R  V  V  E  P  M  R
tatggtgccaccgagaccagattcataaccacaggaaatggggaaaaagctttacaaata
 Y  G  A  T  E  T  R  F  I  T  T  G  N  G  E  K  A  L  Q  I
```

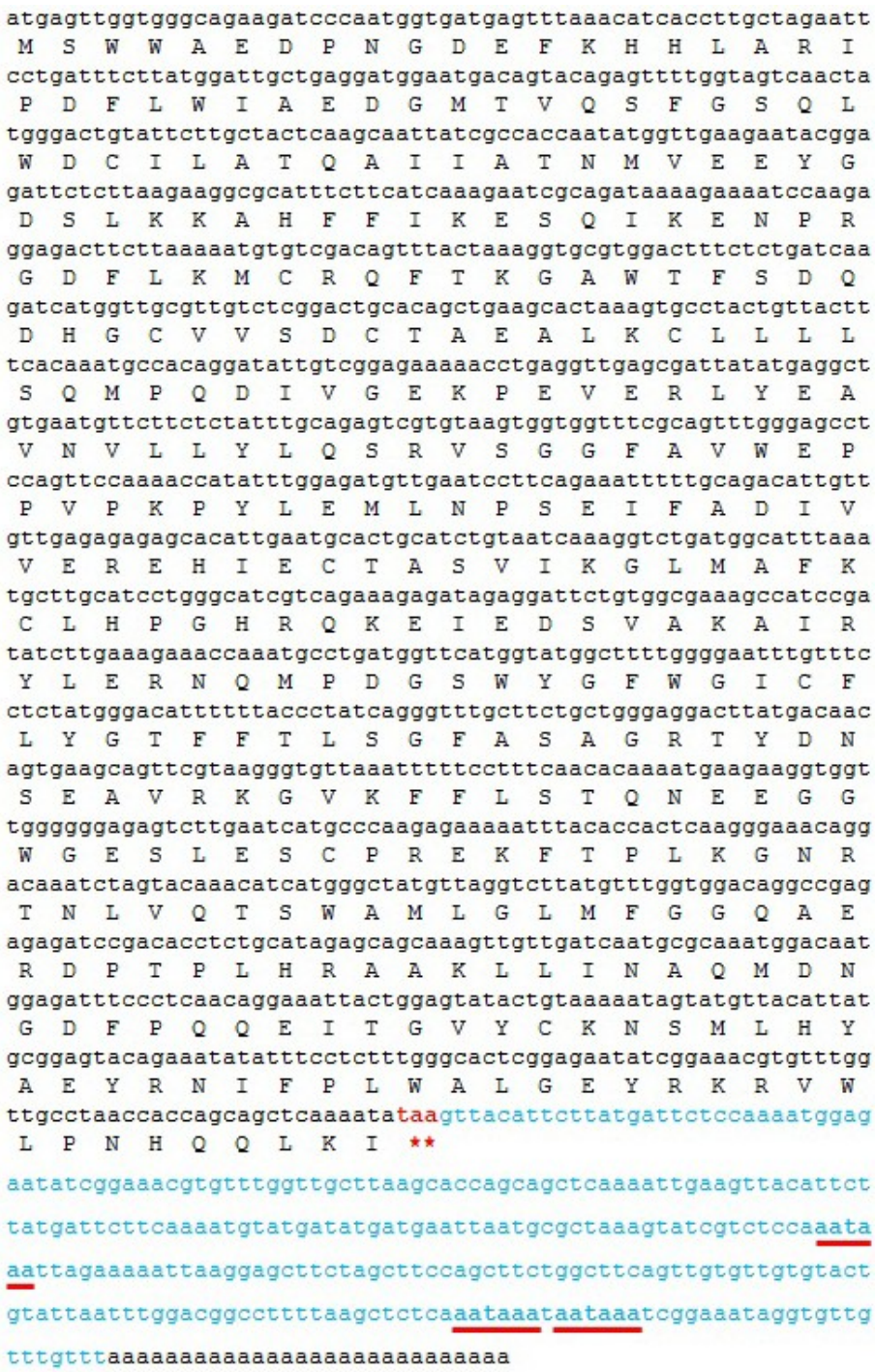
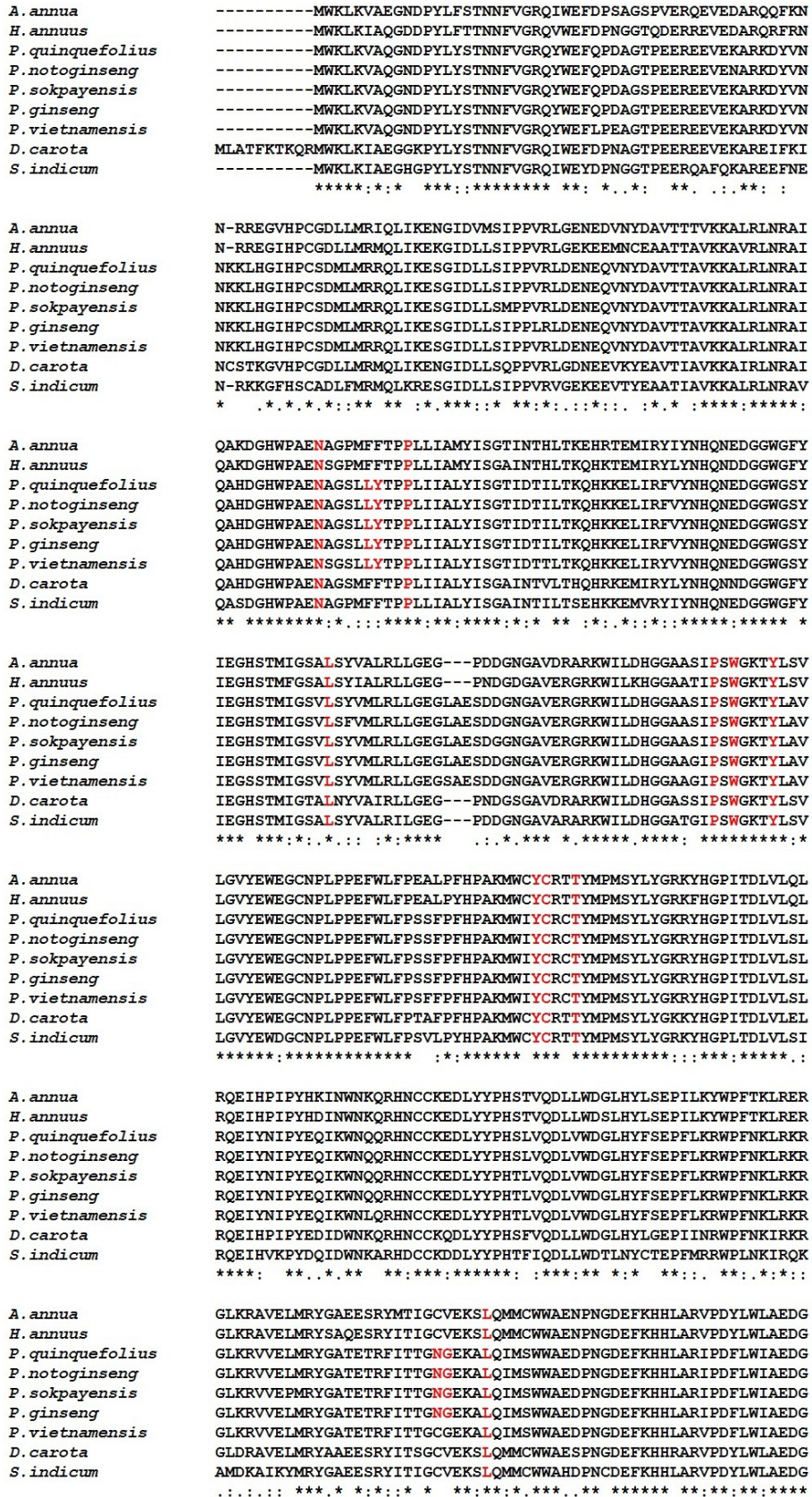


Figure 4.22 Nucleotide and deduced amino acid sequence of *P. sokpayensis dammarenediol synthase* (*PsDS*, KU196775). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and is indicated by "**". 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end. Putative poly A signals are bold and underlined by red bars.


A.annua ----------MWKLKVAEGNDPYLFSTNNFVGRQIWEFDPSAGSPVERQEVEDARQQFKN
H.annuus ----------MWKLKIAQGDDPYLFTTNNFVGRQVWEFDPNGGTQDERREVEDARQRFRN
P.quinquefolius ----------MWKLKVAQGNDPYLYSTNNFVGRQYWEFQPDAGTPEEREEVEKARKDYVN
P.notoginseng ----------MWKLKVAQGNDPYLYSTNNFVGRQYWEFQPDAGTPEEREEVENARKDYVN
P.sokpayensis ----------MWKLKVAQGNDPYLYSTNNFVGRQYWEFQPDAGSPEEREEVEKARKDYVN
P.ginseng ----------MWKLKVAQGNDPYLYSTNNFVGRQYWEFQPDAGTPEEREEVEKARKDYVN
P.vietnamensis ----------MWKLKVAQGNDPYLYSTNNFVGRQYWEFLPEAGTPEEREEVEKARKDYVN
D.carota MLATFKTKQRMWKLKIAEGGKPYLYSTNNFVGRQIWEFDPNAGTPEEREEVEKAREIFKI
S.indicum ----------MWKLKIAEGHGPYLYSTNNFVGRQIWEYDPNGGTPEERQAFQKAREEFNE
*****:*:* ***::******** **: *..*: **. .:.**: :
A.annua N-RREGVHPCGDLLMRIQLIKENGIDVMSIPPVRLGENEDVNYDAVTTTVKKALRLNRAI
H.annuus N-RREGIHPCGDLLMRMQLIKEKGIDLLSIPPVRLGEKEEMNCEAATTAVKKAVRLNRAI
P.quinquefolius NKKLHGIHPCSDMLMRRQLIKESGIDLLSIPPVRLDENEQVNYDAVTTAVKKALRLNRAI
P.notoginseng NKKLHGIHPCSDMLMRRQLIKESGIDLLSIPPVRLDENEQVNYDAVTTAVKKALRLNRAI
P.sokpayensis NKKLHGIHPCSDMLMRRQLIKESGIDLLSMPPVRLDENEQVNYDAVTTAVKKALRLNRAI
P.ginseng NKKLHGIHPCSDMLMRRQLIKESGIDLLSIPPLRLDENEQVNYDAVTTAVKKALRLNRAI
P.vietnamensis NKKLHGIHPCSDMLMRRQLIKESGIDLLSIPPVRLDENEQVNYDAVTTAVKKALRLNRAI
D.carota NCSTKGVHPCGDLLMRMQLIKENGIDLLSQPPVRLGDNEEVKYEAVTIAVKKAIRLNRAI
S.indicum N-RKKGFHSCADLFMRMQLKRESGIDLLSIPPVRVGEKEEVTYEAATIAVKKALRLNRAV
* .*.*.*.*::** ** :*.***::* **:*:.::*::. :*.* :****:*****:
A.annua QAKDGHWPAENAGPMFFTPPLLIAMYISGTINTHLTKEHRTEMIRYIYNHQNEDGGWGFY
H.annuus QAKDGHWPAENSGPMFFTPPLLIAMYISGAINTHLTKQHKTEMIRYLYNHQNDDGGWGFY
P.quinquefolius QAHDGHWPAENAGSLLYTPPLIIALYISGTIDTILTKQHKKELIRFVYNHQNEDGGWGSY
P.notoginseng QAHDGHWPAENAGSLLYTPPLIIALYISGTIDTILTKQHKKELIRFVYNHQNEDGGWGSY
P.sokpayensis QAHDGHWPAENAGSLLYTPPLIIALYISGTIDTILTKQHKKELIRFVYNHQNEDGGWGSY
P.ginseng QAHDGHWPAENAGSLLYTPPLIIALYISGTIDTILTKQHKKELIRFVYNHQNEDGGWGSY
P.vietnamensis QAHDGHWPAENSGSLLYTPPLIIALYISGTIDTTLTKQHKKELIRYVYNHQNEDGGWGSY
D.carota QAHDGHWPAENAGSMFFTPPLIIALYISGAINTVLTHQHRKEMIRYLYNHQNNDGGWGFY
S.indicum QASDGHWPAENAGPMFFTPPLLIALYISGAINTILTSEHKKEMVRYIYNHQNEDGGWGFY
** ********:*.:::****:**:****:*:* ** :*:.*::*::*****:***** *
A.annua IEGHSTMIGSALSYVALRLLGEG---PDDGNGAVDRARKWILDHGGAASIPSWGKTYLSV
H.annuus IEGHSTMFGSALSYIALRLLGEG---PNDGDGAVERGRKWILKHGGAATIPSWGKTYLSV
P.quinquefolius IEGHSTMIGSVLSYVMLRLLGEGLAESDDGNGAVERGRKWILDHGGAASIPSWGKTYLAV
P.notoginseng IEGHSTMIGSVLSFVMLRLLGEGLAESDDGNGAVERGRKWILDHGGAASIPSWGKTYLAV
P.sokpayensis IEGHSTMIGSVLSYVMLRLLGEGLAESDGGNGAVERGRKWILDHGGAASIPSWGKTYLAV
P.ginseng IEGHSTMIGSVLSYVMLRLLGEGLAESDDGNGAVERGRKWILDHGGAAGIPSWGKTYLAV
P.vietnamensis IEGSSTMIGSVLSYVMLRLLGEGSAESDDGNGAVERGRKWILDHGGAAGIPSWGKTYLAV
D.carota IEGHSTMIGTALNYVAIRLLGEG---PNDGSGAVDRARKWILDHGGASSIPSWGKTYLSV
S.indicum IEGHSTMIGSALSYVALRILGEG---PDDGNGAVARARKWILDHGGATGIPSWGKTYLSV
*** ***:*:.*.:: :*:**** .:.*.*** *.*****.****: *********:*
A.annua LGVYEWEGCNPLPPEFWLFPEALPFHPAKMWCYCRTTYMPMSYLYGRKYHGPITDLVLQL
H.annuus LGVYEWEGCNPLPPEFWLFPEALPYHPAKMWCYCRTTYMPMSYLYGRKFHGPITDLVLQL
P.quinquefolius LGVYEWEGCNPLPPEFWLFPSSFPFHPAKMWIYCRCTYMPMSYLYGKRYHGPITDLVLSL
P.notoginseng LGVYEWEGCNPLPPEFWLFPSSFPFHPAKMWIYCRCTYMPMSYLYGKRYHGPITDLVLSL
P.sokpayensis LGVYEWEGCNPLPPEFWLFPSSFPFHPAKMWIYCRCTYMPMSYLYGKRYHGPITDLVLSL
P.ginseng LGVYEWEGCNPLPPEFWLFPSSFPFHPAKMWIYCRCTYMPMSYLYGKRYHGPITDLVLSL
P.vietnamensis LGVYEWEGCNPLPPEFWLFPSFFPFHPAKMWIYCRCTYMPMSYLYGKRYHGPITDLVLSL
D.carota LGVYEWEGCNPLPPEFWLFPTAFPFHPAKMWCYCRTTYMPMSYLYGKKYHGPITDLVLEL
S.indicum LGVYEWDGCNPLPPEFWLFPSVLPYHPAKMWCYCRTTYMPMSYLYGRKYHGPLTDLVLSI
******:************** :*:****** *** **********:::***:*****.:
A.annua RQEIHPIPYHKINWNKQRHNCCKEDLYYPHSTVQDLLWDGLHYLSEPILKYWPFTKLRER
H.annuus RQEIHPIPYHDINWNKQRHNCCKEDLYYPHSTVQDLLWDSLHYLSEPILKYWPFTKLRER
P.quinquefolius RQEIYNIPYEQIKWNQQRHNCCKEDLYYPHSLVQDLVWDGLHYFSEPFLKRWPFNKLRKR
P.notoginseng RQEIYNIPYEQIKWNQQRHNCCKEDLYYPHSLVQDLVWDGLHYFSEPFLKRWPFNKLRKR
P.sokpayensis RQEIYNIPYEQIKWNQQRHNCCKEDLYYPHTLVQDLVWDGLHYFSEPFLKRWPFNKLRKR
P.ginseng RQEIYNIPYEQIKWNQQRHNCCKEDLYYPHTLVQDLVWDGLHYFSEPFLKRWPFNKLRKR
P.vietnamensis RQEIYNIPYEQIKWNLQRHNCCKEDLYYPHTLVQDLVWDGLHYFSEPFLKRWPFNKLRKR
D.carota RQEIHPIPYEDIDWNKQRHNCCKQDLYYPHSFVQDLLWDGLHYLGEPIINRWPFNKIRKR
S.indicum RQEIHVKPYDQIDWNKARHDCCKDDLYYPHTFIQDLLWDTLNYCTEPFMRRWPLNKIRQK
****: **..*.** **:***:******: :***:** *:* **:: **:.*:*::
A.annua GLKRAVELMRYGAEESRYMTIGCVEKSLQMMCWWAENPNGDEFKHHLARVPDYLWLAEDG
H.annuus GLKRAVELMRYSAQESRYITIGCVEKSLQMMCWWAENPNGDEFKHHLARVPDYLWLAEDG
P.quinquefolius GLKRVVELMRYGATETRFITTGNGEKALQIMSWWAEDPNGDEFKHHLARIPDFLWIAEDG
P.notoginseng GLKRVVELMRYGATETRFITTGNGEKALQIMSWWAEDPNGDEFKHHLARIPDFLWIAEDG
P.sokpayensis GLKRVVEPMRYGATETRFITTGNGEKALQIMSWWAEDPNGDEFKHHLARIPDFLWIAEDG
P.ginseng GLKRVVELMRYGATETRFITTGNGEKALQIMSWWAEDPNGDEFKHHLARIPDFLWIAEDG
P.vietnamensis GLKRVVELMRYGATETRFITTGCGEKALQIMSWWAEDPNGDEFKHHLARVPDFLWIAEDG
D.carota GLDRAVELMRYAAEESRYITSGCVEKSLQMMCWWAESPNGDEFKHHRARVPDYLWLAEDG
S.indicum AMDKAIKYMRYGAEESRYITIGCVEKSLQMMCWWAHDPNCDEFKHHLARVPDYLWLAEDG
.:.:.:: ***.* *:*::* * **:**:*.***..** ****** **:**:**:****

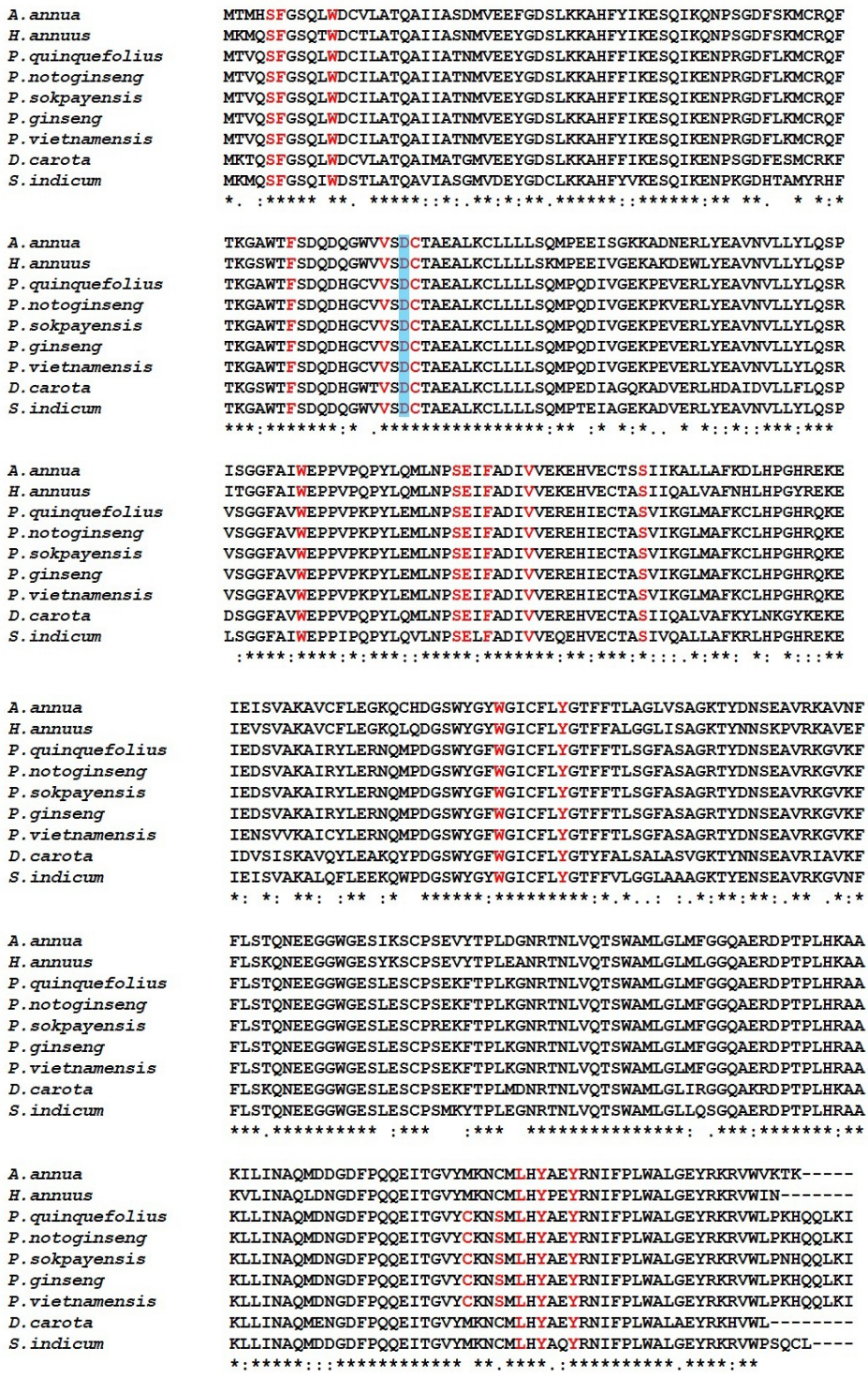


Figure 4.23 Multiple sequence alignment of deduced amino acid sequences of DSs from the *Panax* species and other plants. Catalytic acid residue is highlighted in blue and residues forming active site cavity are shown in red. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (ANB82450), *P. ginseng* (ACZ71036), *P. quinquefolius* (AGI15962), *P. notoginseng* (AGS79229), *P. vietnamensis* var *fuscidiscus* (AGS16975), *Daucus carota* subsp. *Sativus* (XP_017229739), *A. annua* (AHF22084), *H. annuus* (OTG28709), *S. indicum* (XP_011096562).

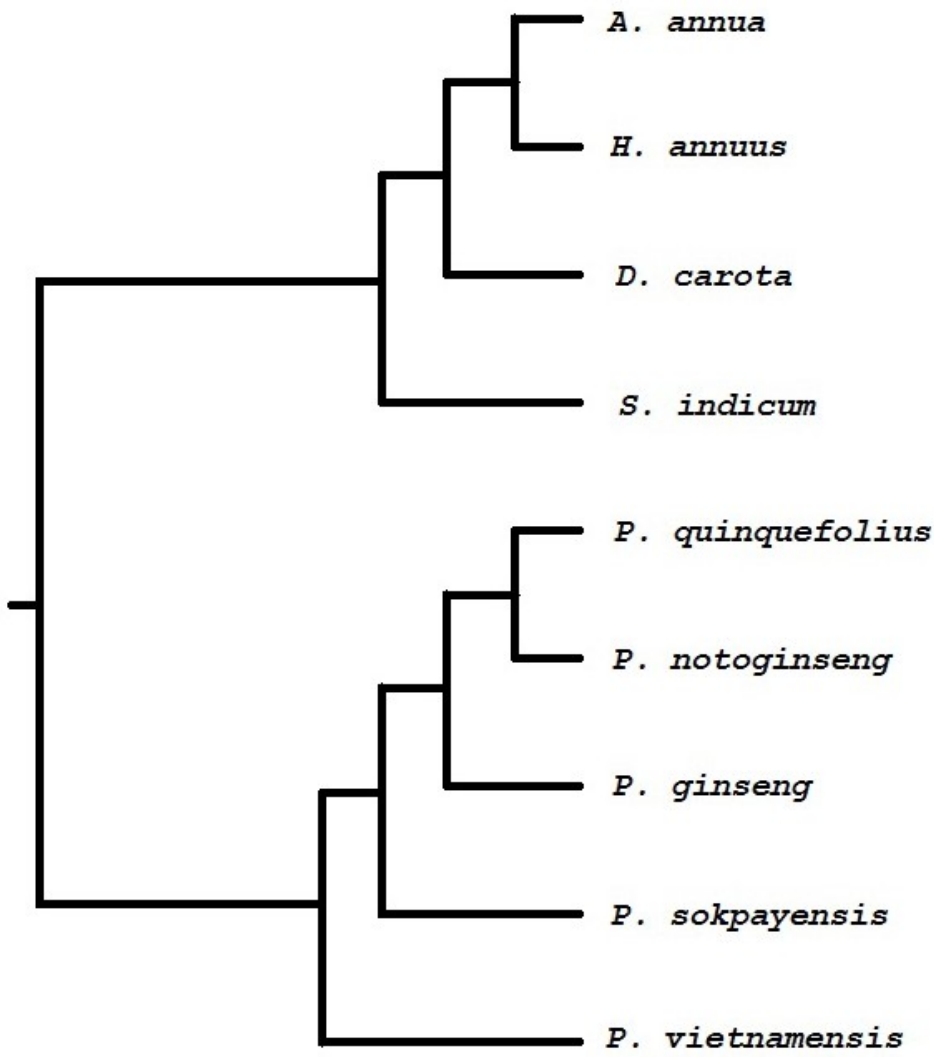


Figure 4.24 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of DSs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (ANB82450), *P. ginseng* (ACZ71036), *P. quinquefolius* (AGI15962), *P. notoginseng* (AGS79229), *P. vietnamensis* var *fuscidiscus* (AGS16975), *D. carota* subsp. *Sativus* (XP_017229739), *A. annua* (AHF22084), *H. annuus* (OTG28709), *S. indicum* (XP_011096562).

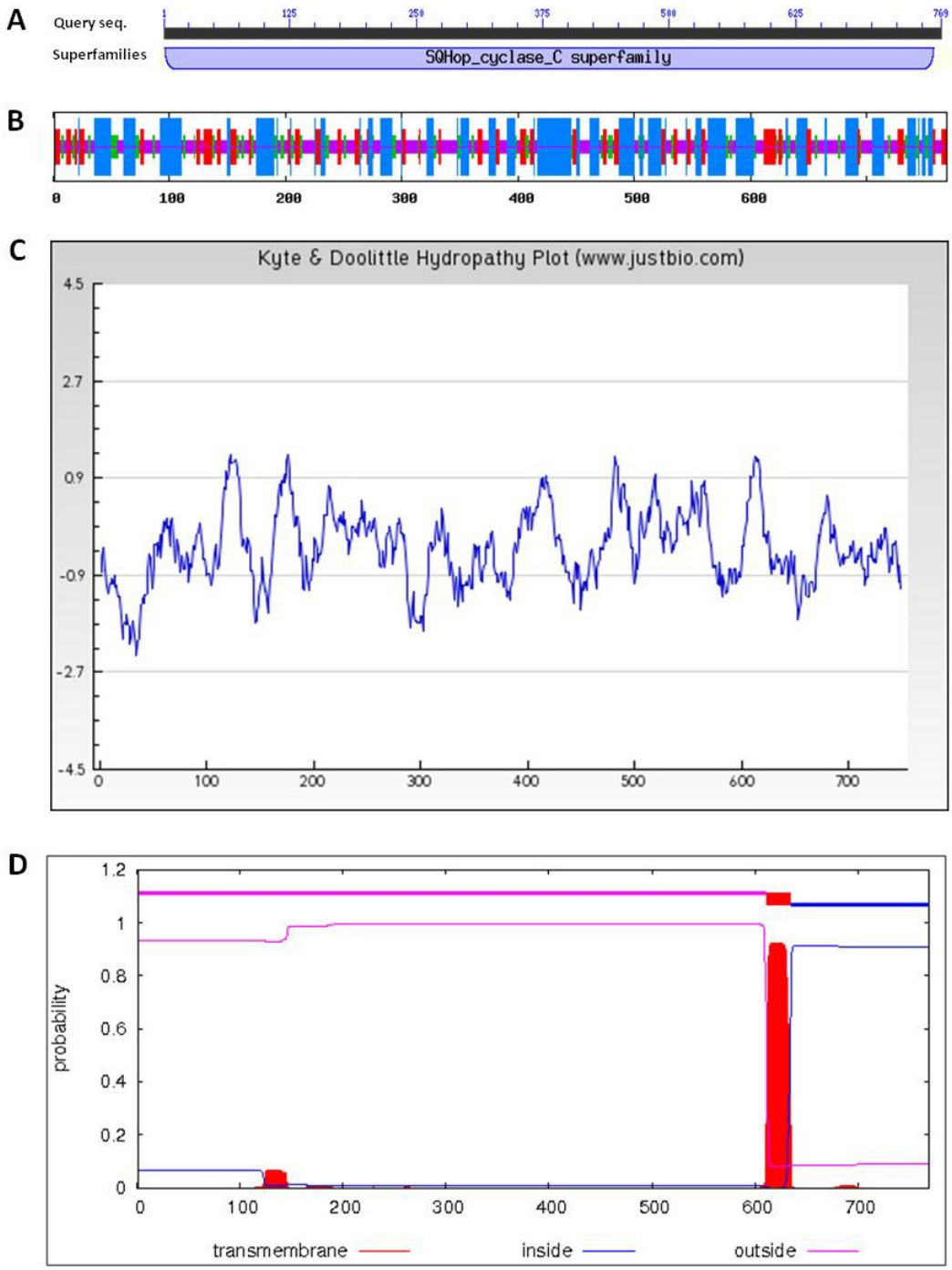


Figure 4.25 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of Dammarenediol Synthase (DS) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsDS. (D) Prediction of transmembrane region using TMHMM program.

4.3.4. *PsSE*

4.3.4.1. Construction of full length *PsSE* using RACE

The partial sequence of *PsSE* was used to design RACE primers for cloning of full-length cDNA using rapid amplification of cDNA ends (RACE). 5′ RACE and 3′ RACE amplicons were approximately 1000 and 1100 bp, respectively (Figure 4.26). A full-length *PsSE* of 1917 bp constructed through the alignment of these RACE fragments were submitted to NCBI with accession number KT936529. This full-length cDNA comprised of 1617 bp long ORF, 5′ UTR of 102 bp and 3′ UTR of 195 bp which included a 24 bp long adenine repeats representing a section of poly A tail (Figure 4.27).

4.3.4.2. *In silico* analysis of full length cDNA of *PsSE*

Putative PASs were detected in 3′ UTR at positions 1740 – 1745 (ATTAAA) and 1865 – 1870 (AATAAA) (Figure 4.27). *In silico* analysis predicted a molecular weight of 59.42 kDa for a deduced 539 amino acids of PsSE and theoretical pI of 8.78. Two conserved features, viz., squalene epoxidase and NAD (P) domains were found in PsSE (Figure 4.28). Multiple sequence alignment of SEs from across the genera and families from the plant kingdom revealed that these two domains were highly conserved (Figure 4.28). On phylogenetic analysis, SEs from different *Panax* species formed a distinct clade (Figure 4.29). The conserved domain search annotated PsSE with squalene monooxygenase (PLN02985 superfamily) (Figure 4.30A). The predicted secondary structure contained 32.84 % alpha helix, 22.26 % extended strand, 10.95 % beta turn and 33.95 % random coil (Figure 4.30B). Kyte and Doolittle hydropathicity plot and TMHMM program suggested the presence of four transmembrane helices (Figures 4.30C, D).

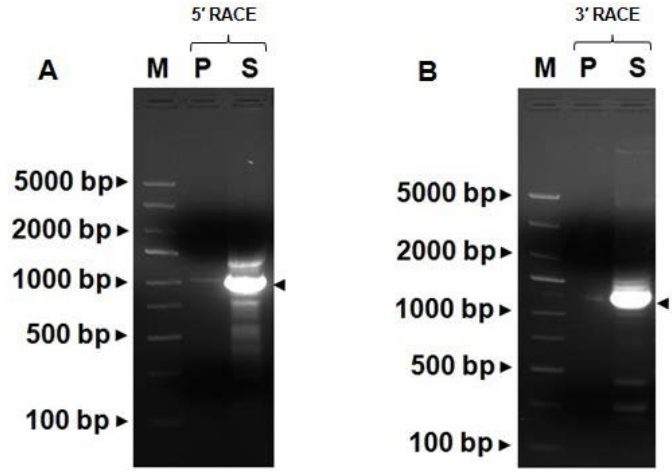


Figure 4.26 Agarose gel showing (A) 5′ RACE products, and (B) 3′ RACE products, of *PsSE*. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

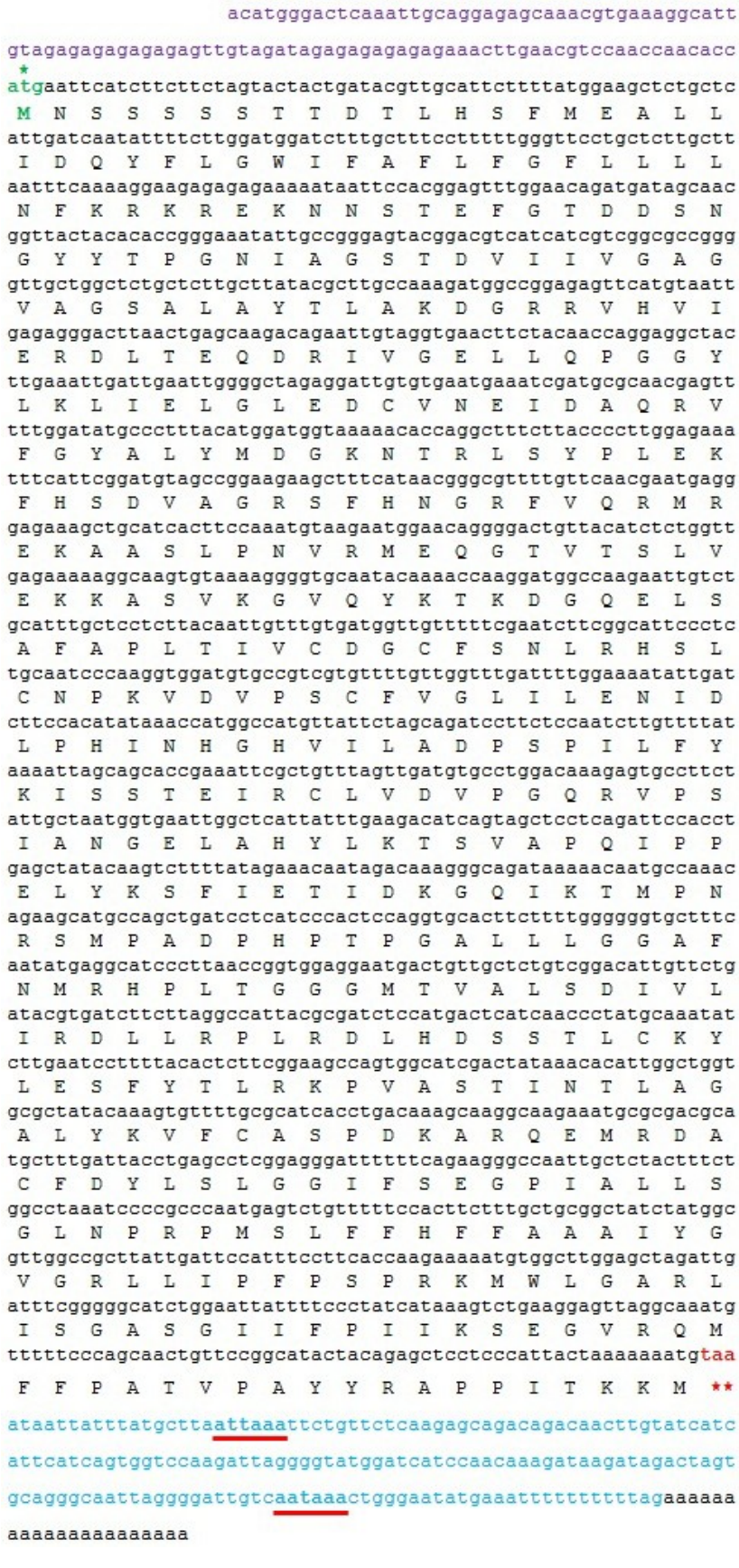


Figure 4.27 Nucleotide and deduced amino acid sequence of *P. sokpayensis squalene*

epoxidase (*PsSE*, KT936529). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**". 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end. Putative poly A signals are bold and underlined by red bars.

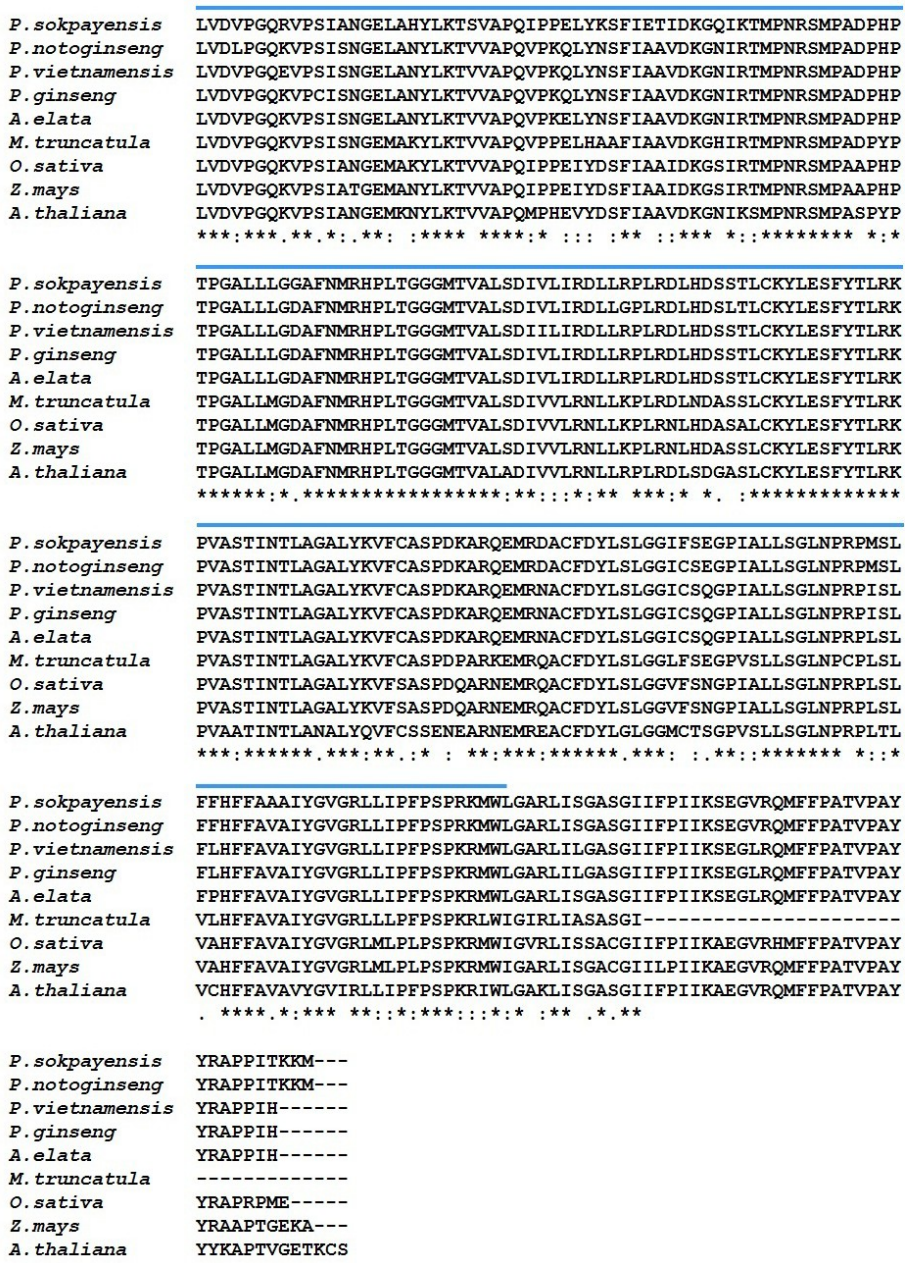


Figure 4.28 Multiple sequence alignment of deduced amino acid sequences of SEs from the *Panax* species and other plants. Conserved Squalene Epoxidase domain is shown by a blue line and NAD(P)-binding Rossmann-like domain is shown in red color. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75534), *P. ginseng* (BAD15330), *P. notoginseng* (AIK21785), *P. vietnamensis* var *fuscidiscus* (AIK23030), *A. elata* (ADC32655), *M. truncatula* (XP_013461821), *O. sativa* (BAF11377), *Z. mays* (ONL95392), *A. thaliana* (NP_564734).

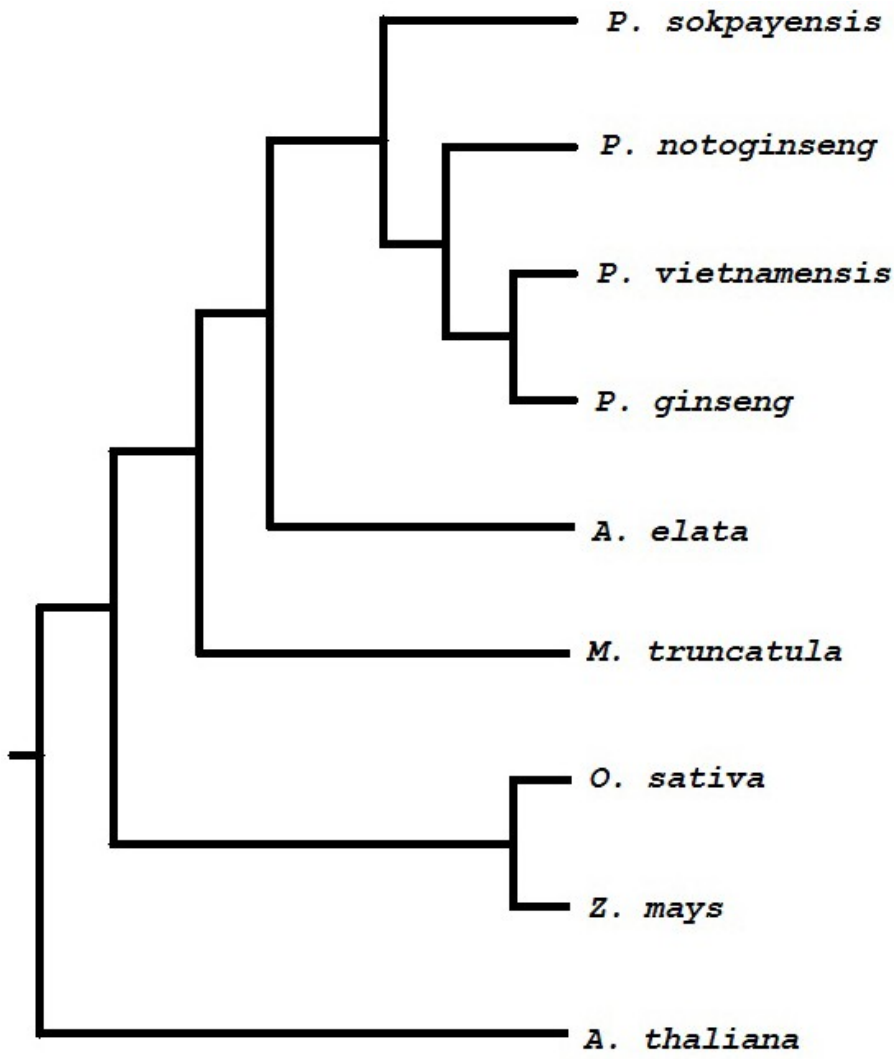


Figure 4.29 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of SEs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75534), *P. ginseng* (BAD15330), *P. notoginseng* (AIK21785), *P. vietnamensis* var *fuscidiscus* (AIK23030), *A. elata* (ADC32655), *M. truncatula* (XP_013461821), *O. sativa* (BAF11377), *Z. mays* (ONL95392), *A. thaliana* (NP_564734).

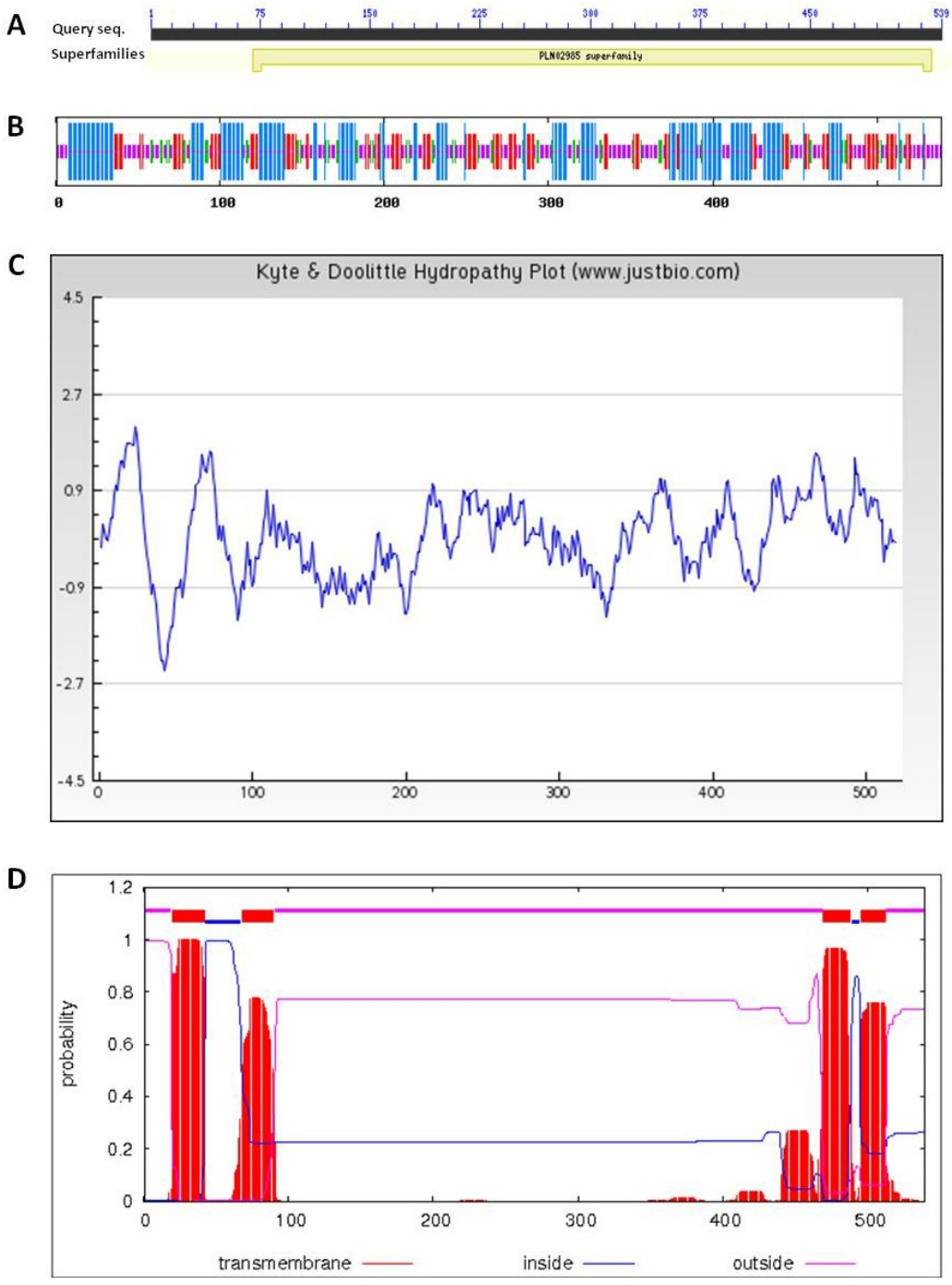


Figure 4.30 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of Squalene Epoxidase (SE) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsSE. (D) Prediction of transmembrane region using TMHMM program.

4.3.5. *PsCS*

4.3.5.1. Construction of full length *PsCS* using RACE

The partial sequence of *PsCS* was used to design RACE primers for cloning of full-length cDNA using rapid amplification of cDNA ends (RACE). 5′ RACE and 3′ RACE amplicons were approximately 1100 and 1500 bp, respectively (Figure 4.31). A full-length *PsSE* of 2520 bp constructed through the alignment of these RACE fragments were submitted to NCBI with accession number KT936530. This full-length cDNA comprised of 2277 bp long ORF, 5′ UTR of 43 bp and 3′ UTR of 200 bp which included a 22 bp long adenine repeats representing a section of poly A tail (Figure 4.32).

4.3.5.2. *In silico* analysis of full length cDNA of *PsCS*

In silico analysis predicted a molecular weight of 86.01 kDa for a deduced 758 amino acids of PsCS and theoretical pI of 6.48. Two conserved features, viz., catalytic acid (1/1) and catalytic site cavity (34/34) were found in PsCS (Figure 4.33). Multiple sequence alignment of CSs from across the genera and families from the plant kingdom revealed that these two features were highly conserved (Figure 4.33). A cladogram of CSs from different plant species contained a distinct clade formed by CSs of *Panax* species (Figure 4.34). The conserved domain search annotated PsCS with squalene cyclase (SQCY) domain subgroup 1 (SQCY_1) proteins belonging to SQHop_cyclase_C superfamily (Figure 4.35A). The predicted secondary structure contained 35.88 % alpha helix, 21.64 % extended strand, 8.71 % beta turn and 33.77 % random coil (Figure 4.35B). Kyte and Doolittle hydropathy plot analysis indicated the presence of potential transmembrane region at the amino terminal of PsCS (Figure 4.35C). However, TMHMM program suggests that the probability of presence of such region is low (Figure 4.35D).

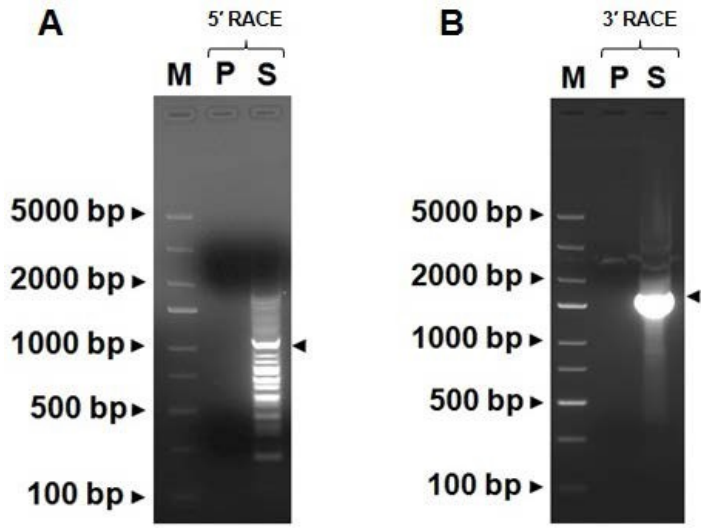


Figure 4.31 Agarose gel showing (A) 5′ RACE products, and (B) 3′ RACE products, of *PsCS*. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

```
*                   acatgggacatgaatgtggagttgaattttgtgcagattgaaa
atgtggaagcttaaaattgcagagggaggaaatccatggcttcggagtctgaacgatcac
 M  W  K  L  K  I  A  E  G  G  N  P  W  L  R  S  L  N  D  H
gtcggccggcaaacttgggagttcgatcccaaactcggatctccggaagagcttgcggag
 V  G  R  Q  T  W  E  F  D  P  K  L  G  S  P  E  E  L  A  E
atcgagacagctcgcgaaacttttcgcaaacatagtttcgagaaaaagcatagttccgat
 I  E  T  A  R  E  T  F  R  K  H  S  F  E  K  K  H  S  S  D
ctcctcatgcgcattcagttttccaatgaaaatcgaggcagtatactcttaccacaagtt
 L  L  M  R  I  Q  F  S  N  E  N  R  G  S  I  L  L  P  Q  V
aaagtaaaagatacagaagatatttcagatgacaaagtaacagttacgttaaaaagagct
 K  V  K  D  T  E  D  I  S  D  D  K  V  T  V  T  L  K  R  A
attaatttccattcaactcttcaggcccatgatggacattggccaggagattatggtggt
 I  N  F  H  S  T  L  Q  A  H  D  G  H  W  P  G  D  Y  G  G
cctatgtttctaatgcctggtttggttattacgctatcgataactggggcactgaatgca
 P  M  F  L  M  P  G  L  V  I  T  L  S  I  T  G  A  L  N  A
gtcttatccaaagaacataaacgtgagatgtgccgttatctttacaatcatcagaacaca
 V  L  S  K  E  H  K  R  E  M  C  R  Y  L  Y  N  H  Q  N  T
gatggtgggtggggtttgcacattgagggtccaagtactatgtttggtactgctttgaac
 D  G  G  W  G  L  H  I  E  G  P  S  T  M  F  G  T  A  L  N
tatgttactttgaggttgcttggcgagggagctaatgatggacaaggggcaatggaaaaa
 Y  V  T  L  R  L  L  G  E  G  A  N  D  G  Q  G  A  M  E  K
gggcgtcaatggattctggatcatggtggtgctactgcaatatcatcatggggaaaaatg
 G  R  Q  W  I  L  D  H  G  G  A  T  A  I  S  S  W  G  K  M
tggctttcagttcttggtgtatttgaatggtccggaaataatcctctgccccccagagata
 W  L  S  V  L  G  V  F  E  W  S  G  N  N  P  L  P  P  E  I
tggctatttccatatatccttccgttccacccaggaaggatgtggtgtcactgtcggatg
 W  L  F  P  Y  I  L  P  F  H  P  G  R  M  W  C  H  C  R  M
gtttatctgcctatgtcatacttgtatgggaagaggtttgttggtccaatcacatcgact
 V  Y  L  P  M  S  Y  L  Y  G  K  R  F  V  G  P  I  T  S  T
gtgctctctttgagaaaggagctcttcactgtcccttatcacgaaatagattggaatcaa
 V  L  S  L  R  K  E  L  F  T  V  P  Y  H  E  I  D  W  N  Q
```

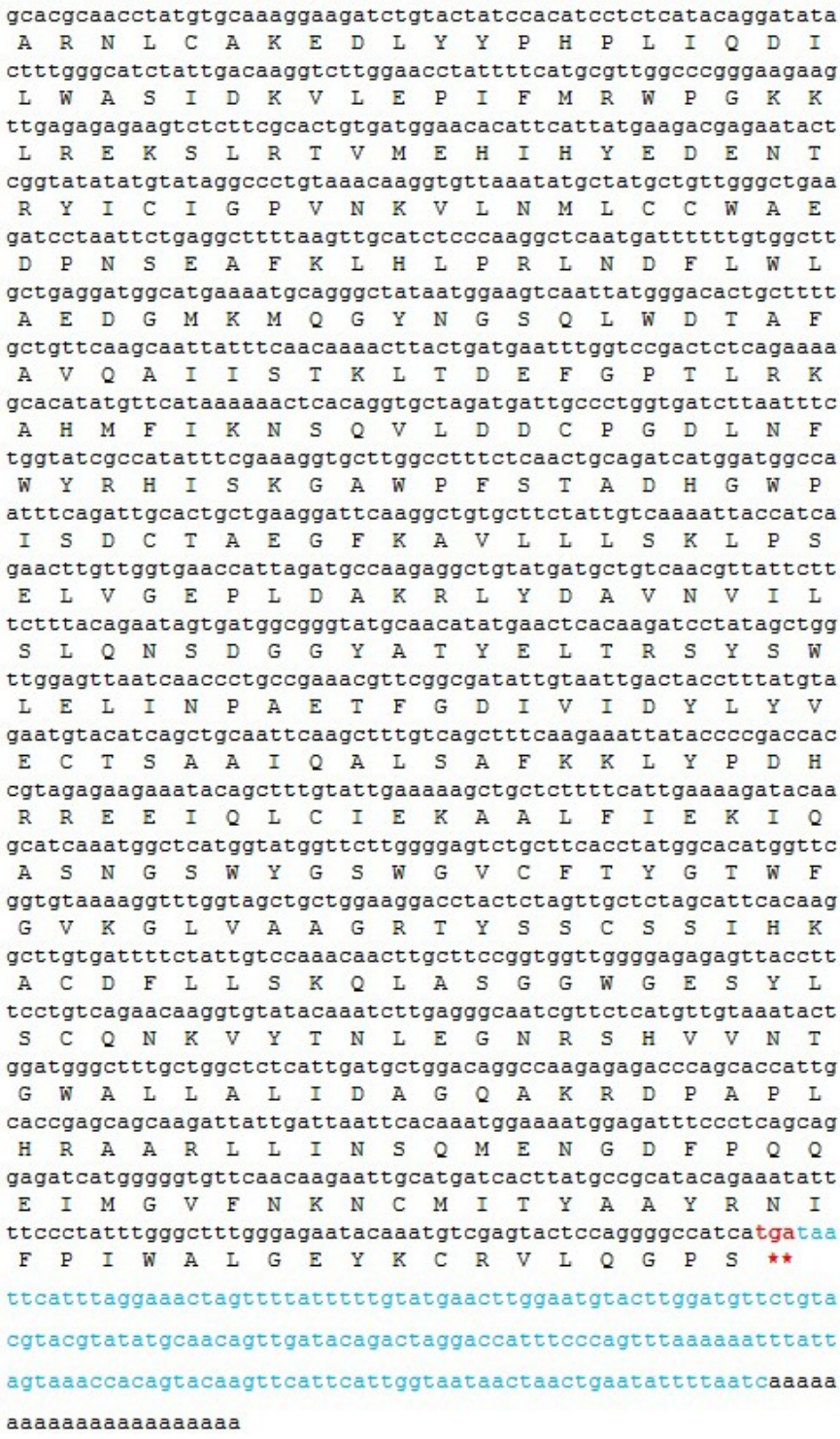


Figure 4.32 Nucleotide and deduced amino acid sequence of *P. sokpayensis cycloartenol synthase* (*PsCS*, KT936530). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**". 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end.

```
P.sokpayensis     MWKLKIAEGG-NPWLRSLNDHVGRQTWEFDPKLGS-PEELAEIETARETFRKHSFEKKHS
P.japonicus       MWKLKIAEGG-NPWLRSLNDHVGRQTWEFDPKLGS-PEELAEIEKARETFRKHRFEKKHS
P.notoginseng     MWKLKIAEGG-NPWLRSLNDHVGRQTWEFDPKLGS-PEELADIEKARETFRKHRFEKKHS
P.ginseng         MWKLKIAEGG-NPWLRTLNDHVGRQIWEFDPNIGS-PEELAEVEKVRENFRNHRFEKKHS
P.quinquefolius   MWKLKIAEGG-NPWLRTLNDHVGRQIWEFDPNIGS-PEELAEIEKVRENFRDHRFEKKHS
M.truncatula      MWKLKVAEGG-SPWLRTLNNHVGRQLWEFDPNSGS-PQDLADIESVRQNFRNNRFSHKHS
P.trichocarpa     MWRLKIAAGG-NPWLRTNNDHIGRQIWEFDPKLTLSPEEISEIENARQNFTKNRFQSKHS
A.thaliana        MWKLKIAEGG-SPWLRTTNNHVGRQFWEFDPNLGT-PEDLAAVEEARKSFSDNRFVQKHS
O.sativa          MWRLRVAEGGGDPWLRTKNGHVGRQVWEFDP-AAGDPDELAAVEAARRGFAARRHELKHS
Z.mays            MWRLRIAEGGGDPWLRTKNAHVGRQVWEFDA-SADPD---PAVDAARRAFAGSRGHLKHS
                  **:*::* ** .****: * *:*** ****.          . :: .*. *       ***

P.sokpayensis     SDLLMRIQFSNENRGSILLPQVKVKDTEDISDDKVTVTLKRAINFHSTLQAHDGHWPGDY
P.japonicus       SDLLMRIQFSNENRGSILLPQVKVKDTEDISDDKVTVTLKRAINFHSTLQAHDGHWPGDY
P.notoginseng     SDLLMRIQFSNENRGSILLPQVKVKDTEDISDAKVTVTLKRAINFHSTLQAHDGHWPGDY
P.ginseng         ADLLMRIQFANENPGSVVLPQVKVNDGEDISEDKVTVTLKRAMSFYSTLQAHDGHWPGDY
P.quinquefolius   ADLLMRIQFANENPGSVVLPQVKVNDGEDILEDKVTVTLKRAMSFYSTLQAHDGHWPGDY
M.truncatula      DDLLMRIQFAKENPMKEVIPKVRVKDVEDVNEESVTTTLRRALNFYSTLQSHDGHWPGDY
P.trichocarpa     ADLIMRMQFEKENPVSEVLPQVKLKESEKVTEEAVTTTLKRGLDYYSSIQAHDGHWPGDY
A.thaliana        ADLLMRLQFSRENLISPVLPQVKIEDTDDVTEEMVETTLKRGLDFYSTIQAHDGHWPGDY
O.sativa          SDLLMRMQFAKANPLKLDIPAIKLEEHEAVTGEAVLSSLKRAIARYSTFQAHDGHWPGDY
Z.mays            ADLLMRIQFAKENPLELDLPAIKLGEHEDVTEEAVSTTLRRAISRLSTLQAHDGHWPGDY
                   **:**:** . *  .  :* ::: : : :    *  :*:*.:   *::*:*********

P.sokpayensis     GGPMFLMPGLVITLSITGALNAVLSKEHKREMCRYLYNHQNTDGGWGLHIEGPSTMFGTA
P.japonicus       GGPMFLMPGLVITLSITGALNAVLSKEHKREMCRYLYNHQNRDGGWGLHIEGPGTMFGTA
P.notoginseng     GGPMFLMPGLVITLSITGALNAVLSKEHKHEMCRYLYNHQNRDGGWGLHIEGPSTMFGTA
P.ginseng         GGPMFLMPGLVITLSITGVLNVVLSKEHKREICRYLYNHQNRDGGWGLHIEGPSTMFGTV
P.quinquefolius   GGPMFLMPGLVITLSITGALNAVLSKEHKREICRYLYNHQNRDGGWGLHIEGPSTMFGTV
M.truncatula      AGPMFLLPGLVITLSITGALNAVLTDEHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSV
P.trichocarpa     GGPMFLMPGLVITLSITGALHAVLSDEHKKEIIRYLYNHQNRDGGWGLHIEGPSTMFGSV
A.thaliana        GGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQNEDGGWGLHIEGPSTMFGSV
O.sativa          GGPMFLMPGLIITLYVSGALNTALSSEHQKEIRRYLYNHQNEDGGWGLHIEGHSTMFGSA
Z.mays            GGPMFLMPGLIITLYVTGALNTVLSSEHQKEIRRYLYNHQNEDGGWGLHIEGPSTMFGSA
                  .*****:***:*** ::*.*:..*:.:*::*: ******** ********** .****:.

P.sokpayensis     LNYVTLRLLGEGANDGQGAMEKGRQWILDHGGATAISSWGKMWLSVLGVFEWSGNNPLPP
P.japonicus       LNYVTLRLLGEGANDGQGAMEKGRQWILDHGGATAISSWGKMWLSVLGVFEWSGNNPLPP
P.notoginseng     LNYVTLRLLGEGTNDGQGAMEKGRQWILDHGGATAISSWGKMWLSVLGVFEWSGNNPLPP
P.ginseng         LNYVTLRLLGEGANDGQGAMEKGRQWILDHGSATAITSWGKMWLSVLGVFEWSGNNPLPP
P.quinquefolius   LNYVTLRLLGEGANDGQGAMEKGRQWILDHGSATAITSWGKMWLSVLGVFEWSGNNPLPP
M.truncatula      LSYVTLRLLGEGPNDGEGDMEKARDWILEHGGATHITSWGKMWLSVLGVFEWSGNNPLPP
P.trichocarpa     LNYVTLRLLGEGPNDGEGAMEKARDWILNHGGATMITSWGKMWLSVLGVFEWSGNNPMPP
A.thaliana        LNYVTLRLLGEGPNDGDGDMEKGRDWILNHGGATNITSWGKMWLSVLGAFEWSGNNPLPP
O.sativa          LTYVSLRLLGEGPDSGDGAMEKGRKWILDHGGATYITSWGKFWLSVLGVFDWSGNNPVPP
Z.mays            LTYVTLRLLGEGPDSGDGAMEKGRNWILDHGGATYITSWGKFWLSVLGVFEWSGNNPVPP
                  *.**:*******.:.*:* ***.*.***:**.** *:****:******.*:******:**

P.sokpayensis     EIWLFPYILPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITSTVLSLRKELFTVPYHEIDW
P.japonicus       EIWLFPYILPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITSTVLSLRKELFTVPYHEIDW
P.notoginseng     EIWLFPYILPFHPGRMWCHCRMVYLPMSYLYGKRFVGPITSTVLSLRKELFTVPYHEIDW
P.ginseng         ETWLLPYILPIHPGRMWCHRRMVYLPMSYLYGKRFVGPITPTVLSLRKEVFSVPYHEIDW
P.quinquefolius   ETWLLPYILPIHPGRMWCHCRMVYLPMSYLYGKRFVGPITPTVLSLR-RVFSVPYHEIDW
M.truncatula      ELWLLPYILPFHPGRMWCHCRMVYLPMCYLYGKRFVGPITPTILSLRKELFTVPYHDVDW
P.trichocarpa     EMWLLPYLLPVHPGRMWCHCRMVYLPMSYLYGKRFVGPITPTILSLRKELFTVPYHEIEW
A.thaliana        EIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGPITSTVLSLRKELFTVPYHEVNW
O.sativa          EIWLLPYFLPIHPGRMWCHCRMVYLPMCYIYGKRFVGPVTPIILELRKELYEVPYNEVDW
Z.mays            EVWLLPYLLPFHPGRMWCHCRMVYLPMCYIYGKRFVGRITPLVLELRKELFKDPYSKIDW
                  * **:**:**.******** *******.*:******* :*. :*.** .::  ** .::*

P.sokpayensis     NQARNLCAKEDLYYPHPLIQDILWASIDKVLEPIFMRWPGKKLREKSLRTVMEHIHYEDE
P.japonicus       NQARNLCAKEDLYYPHPLIQDILWASIDKVLEPIFMRWPGKKLREKSLRTVMEHIHYEDE
P.notoginseng     NQARNLCAKEDLYYPHPLIQDILWASIDKVLEPIFMCWPGKKLREKSLRTVMEHIHYEDE
P.ginseng         NQARNLCAKEDLYYPHPLIQDILWASLDKVWEPIFMHWPAKKLREKSLRTVMEHIHYEDE
P.quinquefolius   NQARNLCAKEDLYYPHPLIQDILWASLDKVWEPIFMHWPAKKLREKSLRTVMEHIHYEDE
M.truncatula      NQARNLCAKEDLYYPHPLVQDILWATLHKVVEPIFMNWPCKKLREKAVETVMEHIHYEDE
P.trichocarpa     NQARTLCAKEDLYYPHPLVQDVLWALLDKAAEPVLMHWPGKKLREKALCTAIEHIHYEDE
A.thaliana        NEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLMRWPGANLREKAIRTAIEHIHYEDE
O.sativa          DKARNLCAKEDLYYPHPFVQDVLWATLHKFVEPAMLRWPGNKLREKALDTVMQHIHYEDE
Z.mays            DKARNLCAKEDLYYPHPFVQDVLWATLHKFVEPVMMSWPGSKLREKALETAMQHVHYEDE
                  ::**.************::**:*** :.*  ** :: **  :****:: *.::*:*****

P.sokpayensis     NTRYICIGPVNKVLNMLCCWAEDPNSEAFKLHLPRLNDFLWLAEDGMKMQGYNGSQLWDT
P.japonicus       NTRYICIGPVNKVLNMLCCWAEDPNSEAFKLHLPRLNDFLWLAEDGMKMQGYNGSQLWDT
P.notoginseng     NTRYICIGPVNKVLNMLCCWAEDPNSEAFKLHLPRLNDFLWLAEDGMKMQGYNGSQLWDT
P.ginseng         NTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLPRLHDFLWLAEDGMKMQGYNGSQLWDT
P.quinquefolius   NTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLPRLHDFLWLAEDGMKMQGYNGSQLWDT
M.truncatula      NTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLPRIHDYLWIAEDGMKMQGYNGSQLWDT
P.trichocarpa     NTRYLCIGPVNKVLNMLCCWVEDPNSEAFKLHIPRIQDYLWLAEDGMKMQGYNGSQLWDT
A.thaliana        NTRYICIGPVNKVLNMLCCWVEDPNSEAFKLHLPRIHDFLWLAEDGMKMQGYNGSQLWDT
O.sativa          NTRYICIGPVNKVLNMLACWIEDPNSEAFKLHIPRVHDYLWIAEDGMKMQGYNGSQLWDT
Z.mays            NTRYICIGPVNKVLNMLACWIEDPNSEAFKLHIPRVYDYLWLAEDGMKMQGYNGSQLWDT
                  ****:************.** ***********:**: *:**:******************
```

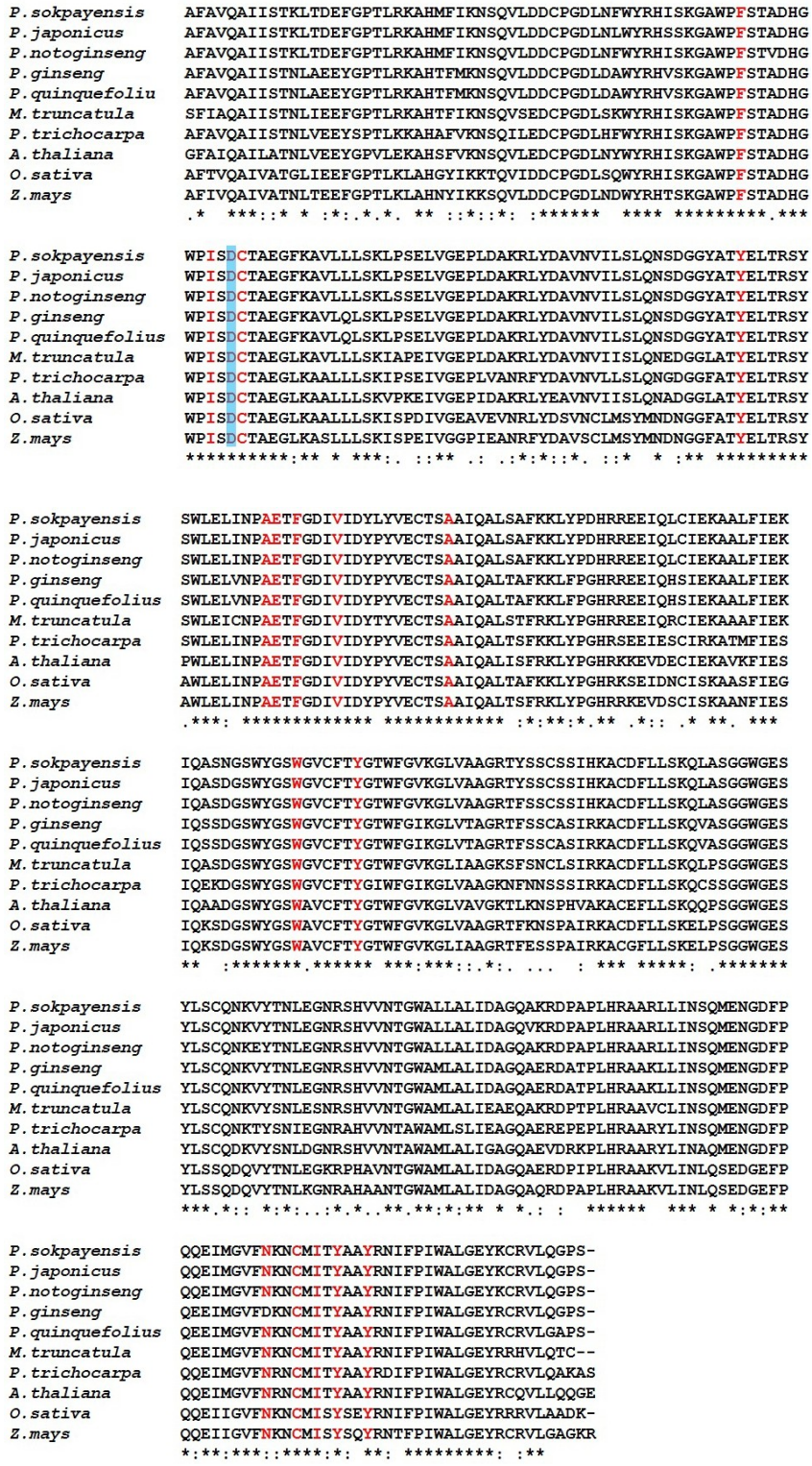


Figure 4.33 Multiple sequence alignment of deduced amino acid sequences of CSs from the *Panax* species and other plants. Catalytic acid residue is highlighted in blue and residues forming active site cavity are shown in red. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75535), *P. ginseng* (O82139), *P. quinquefolius* (AGK62447), *P. notoginseng* (ABY60426), *P. japonicus* (ALB38665), *M. truncatula* (XP_003610947), *P. trichocarpa* (XP_002308131), *A. thaliana* (AAC04931), *O. sativa* (XP_015625472), *Z. mays* (XP_008679954).

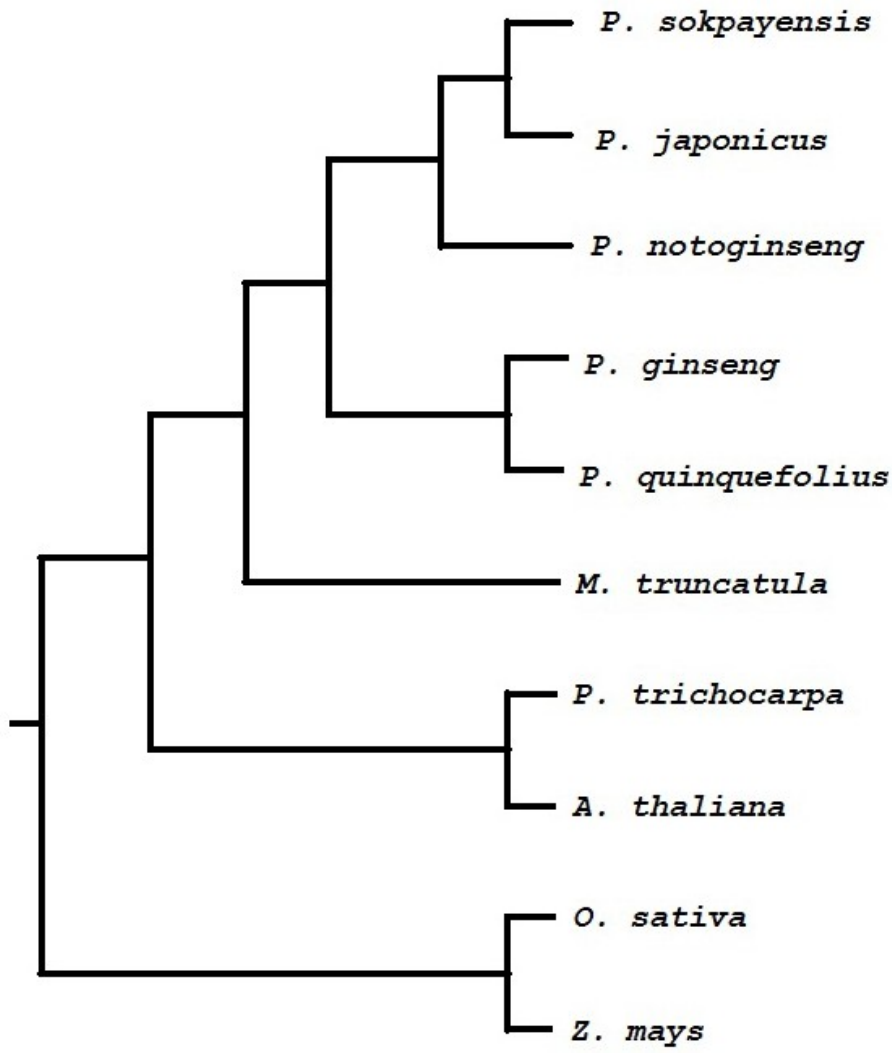


Figure 4.34 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of CSs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (AMT75535), *P. ginseng* (O82139), *P. quinquefolius* (AGK62447), *P. notoginseng* (ABY60426), *P. japonicus* (ALB38665), *M. truncatula* (XP_003610947), *P. trichocarpa* (XP_002308131), *A. thaliana* (AAC04931), *O. sativa* (XP_015625472), *Z. mays* (XP_008679954).

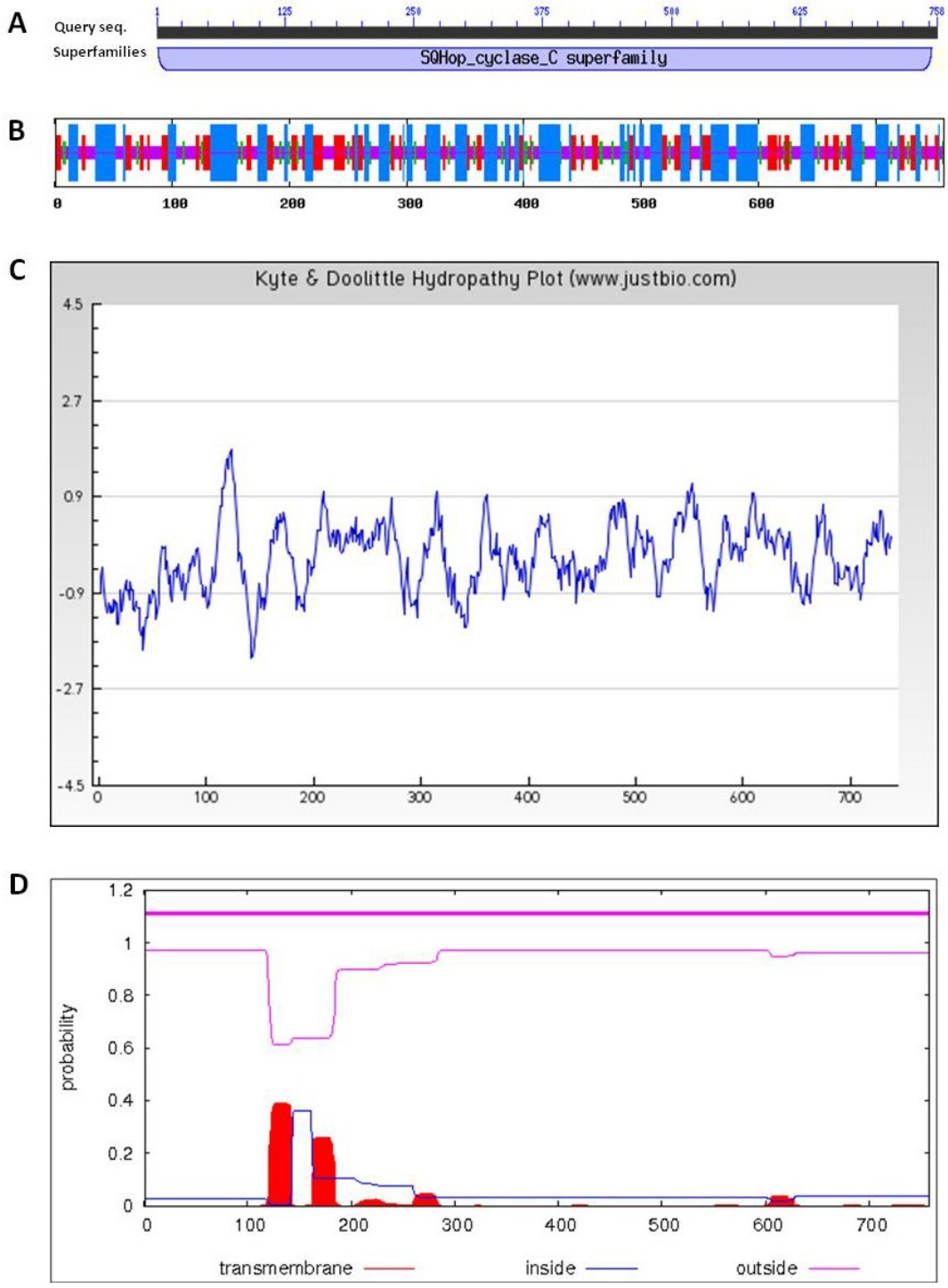


Figure 4.35 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of Cycloartenol Synthase (CS) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsCS. (D) Prediction of transmembrane region using TMHMM program.

4.3.6. *PsPMVK*

4.3.6.1. Construction of full length *PsPMVK* using RACE

The partial sequence of *PsPMVK* was used to design RACE primers for cloning of full-length cDNA using rapid amplification of cDNA ends (RACE). 3′ RACE fragment of ~1500 bp, was amplified (Figure 4.36). *PsPMVK* of 1806 bp was constructed through the alignment of partial fragment obtained using degenerate primers and 3′ RACE fragment. The sequence was submitted to NCBI with accession number (MF682468). This sequence comprised of 1518 bp long ORF and 288 bp long 3′ UTR which included a 26 bp long adenine repeats representing a section of poly A tail (Figure 4.37).

4.3.6.2. In silico analysis of full length cDNA of *PsPMVK*

Putative PAS was detected in 3′ UTR at position 1746 – 1751 (AATAAA) (Figure 4.37). *In silico* analysis predicted a molecular weight of 54.72 kDa for a deduced 505 amino acids of PsPMVK and theoretical pI of 5.39. Multiple sequence alignment of PMVKs from across the genera and families from the plant kingdom revealed that the major part of this protein is conserved (Figure 4.38). Cladogram formed by phylogenetic analysis of PMVKs of different plant species placed *Panax* PMVKs close to one another (Figure 4.39). The conserved domain search annotated PsPMVK with PMVK belonging to ERG8 superfamily (Figure 4.40A). The predicted secondary structure contained 42.57 % alpha helix, 18.61 % extended strand, 7.92 % beta turn and 30.89 % random coil (Figure 4.40B). Kyte and Doolittle hydropathy plot predicts two hydrophobic regions (Figure 4.40C). However, TMHMM analysis failed to detect any transmembrane helix in PsPMVK (Figure 4.40D).

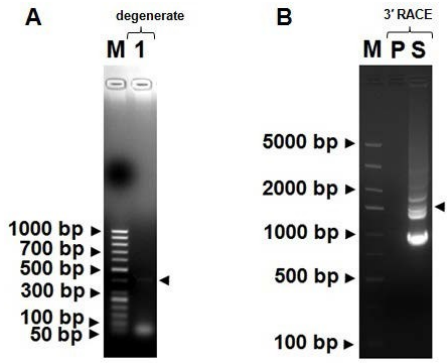


Figure 4.36 Agarose gel showing (A) amplicon obtained using degenerate primers and (B) 3′ RACE products of *PsPMVK*. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

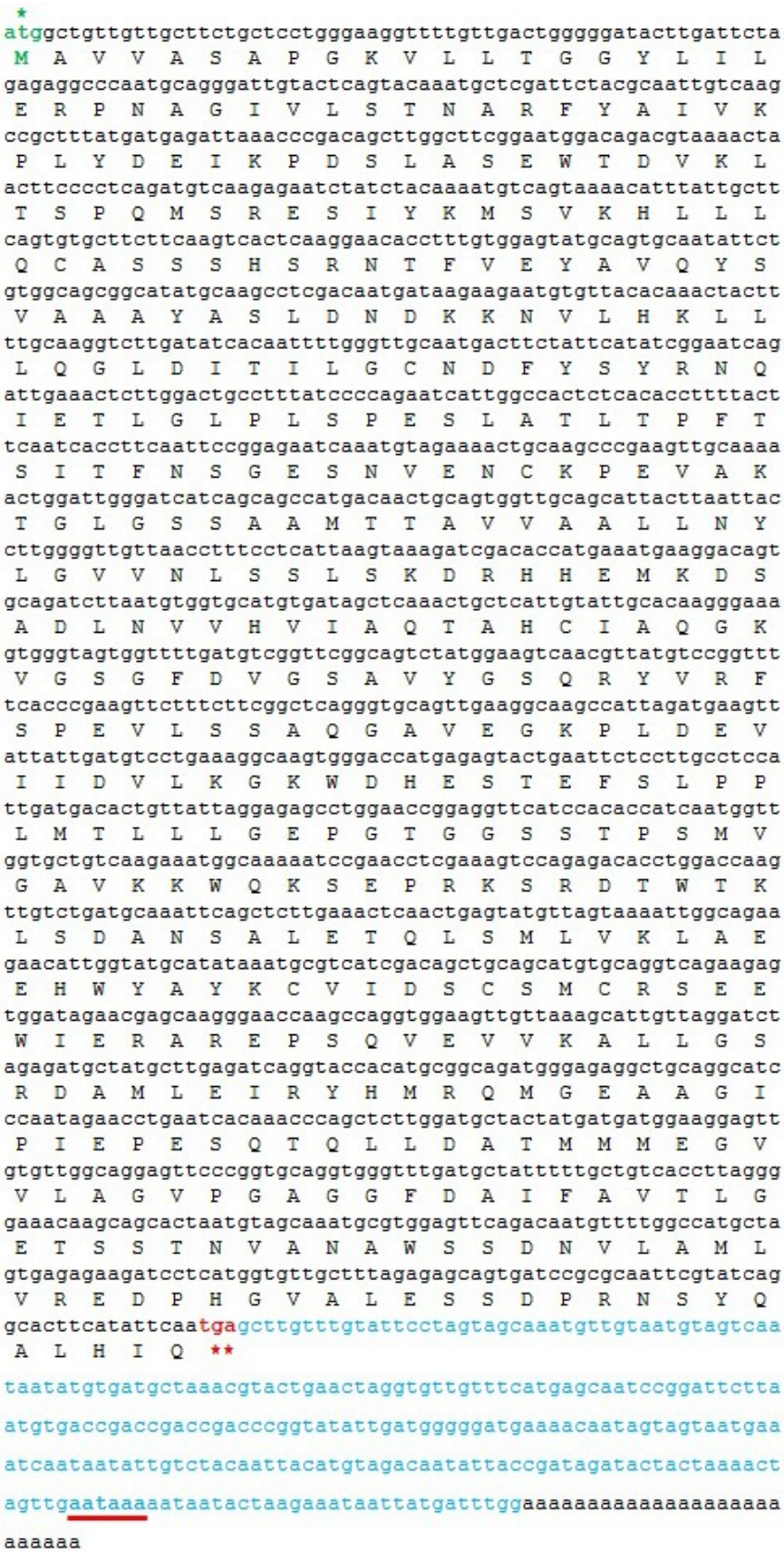


Figure 4.37 Nucleotide and deduced amino acid sequence of *P. sokpayensis phosphomevalonate kinase* (*PsPMVK*, MF682468). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and

its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**". 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end. Putative poly A signal is bold and underlined by a red bar.

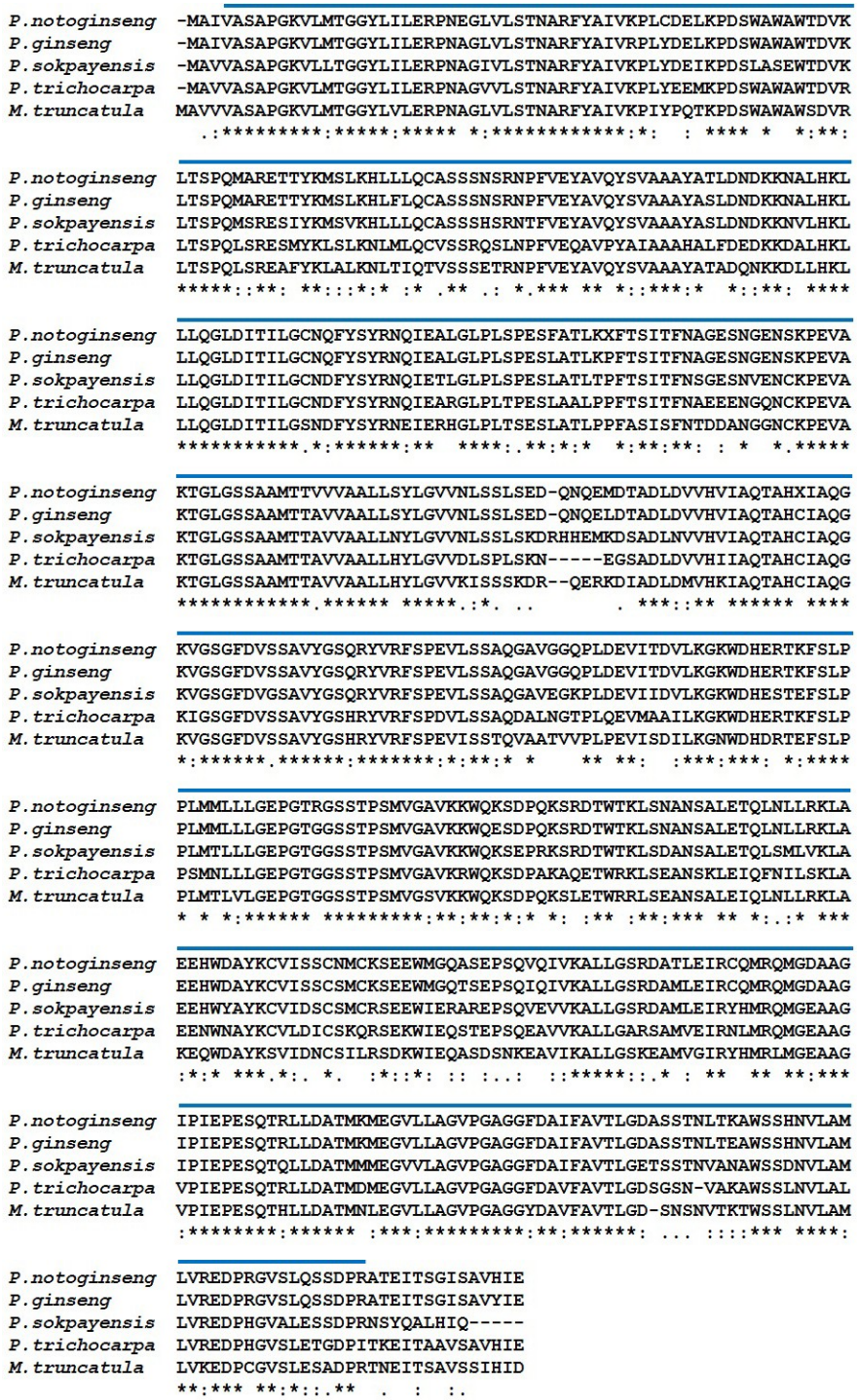


Figure 4.38 Multiple sequence alignment of deduced amino acid sequences of PMVKs from the *Panax* species and other plants. Phosphomevalonate Kinase domain is shown by a line in blue. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (MF682468), *P. ginseng* (AGZ15314), *P. notoginseng* (AIK21784), *M. truncatula* (XP_003602220), *P. trichocarpa* (XP_002303445).

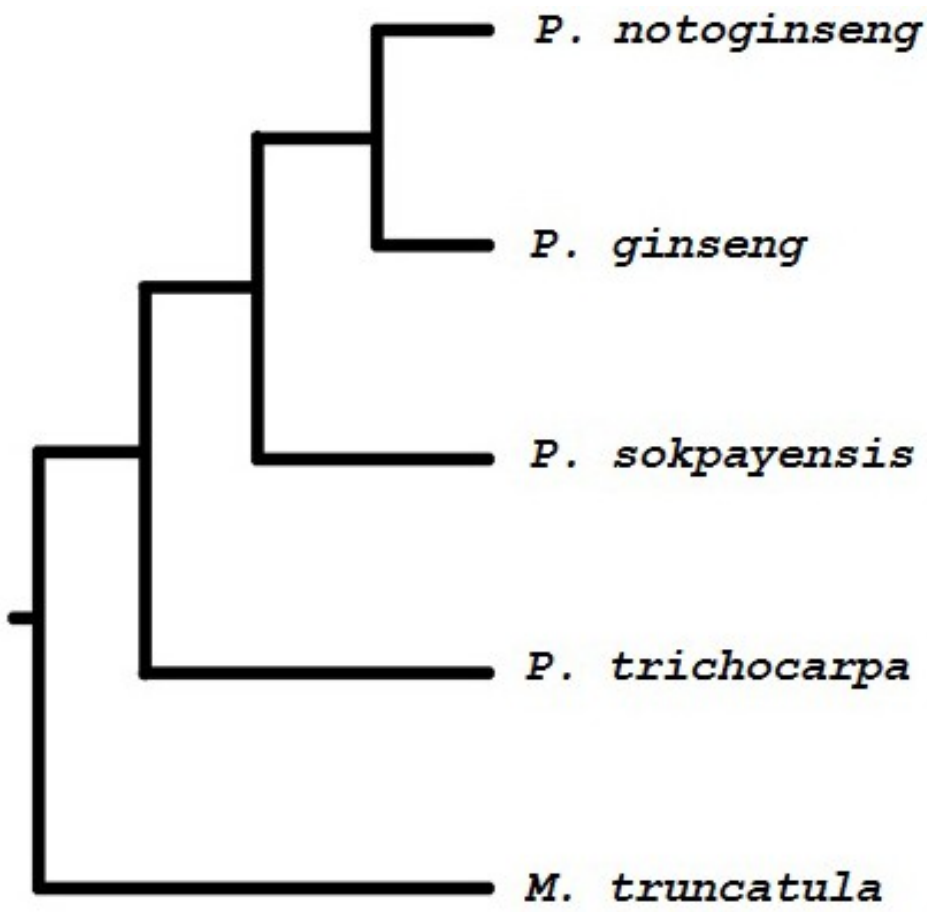


Figure 4.39 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of PMVKs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (MF682468), *P. ginseng* (AGZ15314), *P. notoginseng* (AIK21784), *M. truncatula* (XP_003602220), *P. trichocarpa* (XP_002303445).

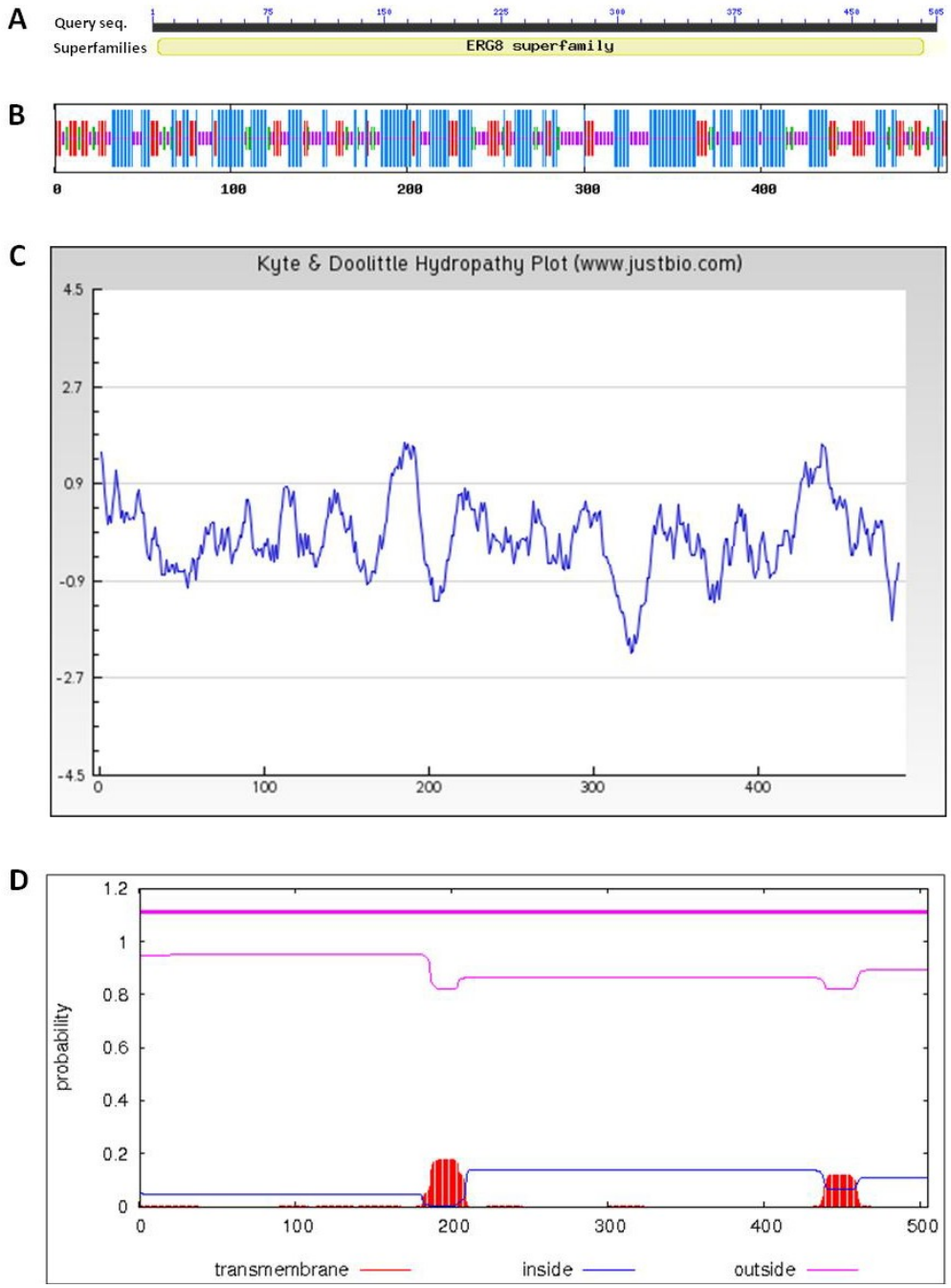


Figure 4.40 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of Phosphomevalonate Kinase (PMVK) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of PsPMVK. (D) Prediction of transmembrane region using TMHMM program.

4.3.7. *PsCMK*

4.3.7.1. Construction of full length *PsCMK* using RACE

The partial sequence of *PsCMK* was used to design RACE primers for cloning of full-length cDNA using rapid amplification of cDNA ends (RACE). 5′ and 3′ RACE fragments were ~800 and ~500 bp, respectively (Figure 4.41). *PsCMK* of 1604 bp was constructed through the alignment of amplicon obtained with degenerate primer set and RACE fragments. The sequence was submitted to NCBI with accession number (MF682467). This sequence comprised of 1212 bp long ORF, 149 bp long 5′ UTR and 243 bp long 3′ UTR which included a 26 bp long adenine repeats representing a section of poly A tail (Figure 4.42).

4.3.7.2. *In silico* analysis of *PsCMK*

In silico analysis predicted a molecular weight of 44.33 kDa for a deduced 403 amino acids of PsCMK and theoretical pI of 5.56. Multiple sequence alignments of PsCMK with CMKs from other plants revealed that both IspE superfamily domain and GHMP kinases N terminal domain were conserved across CMKs belonging to different genera (Figure 4.43). On phylogenetic analysis of CMKs from different plant species, PsCMK aligned itself close to CMK from *Hedera helix*, a plant belonging to same family, i.e., Araliaceae (Figure 4.44). The conserved domain search in PsCMK detected IspE superfamily domain along with GHMP kinases N terminal domain (Figure 4.45A). The predicted secondary structure contained 31.76 % alpha helix, 22.58 % extended strand, 7.69 % beta turn and 37.97 % random coil (Figure 4.45B). Kyte and Doolittle hydropathy plot and TMHMM analysis suggested PsCMK to be a non – membranous protein (Figures 4.45C, D).

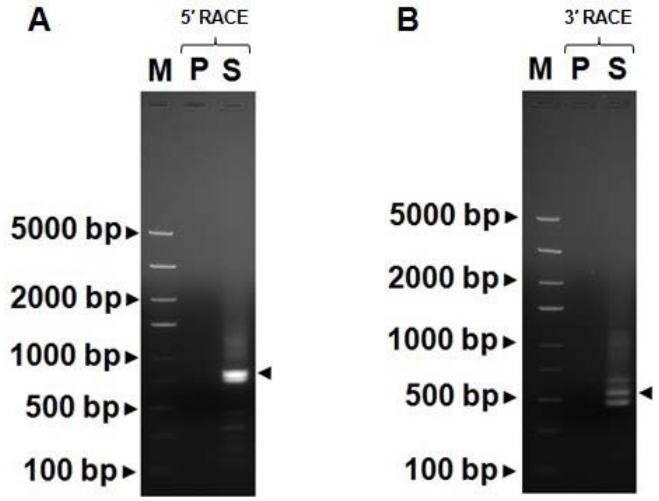


Figure 4.41 Agarose gel showing (A) 5′ RACE products, and (B) 3′ RACE products, of *PsCMK*. M: DNA marker; P: primary RACE PCR; S: secondary RACE PCR; bp – base pairs; arrow on the right hand side of the gel indicates the desired band.

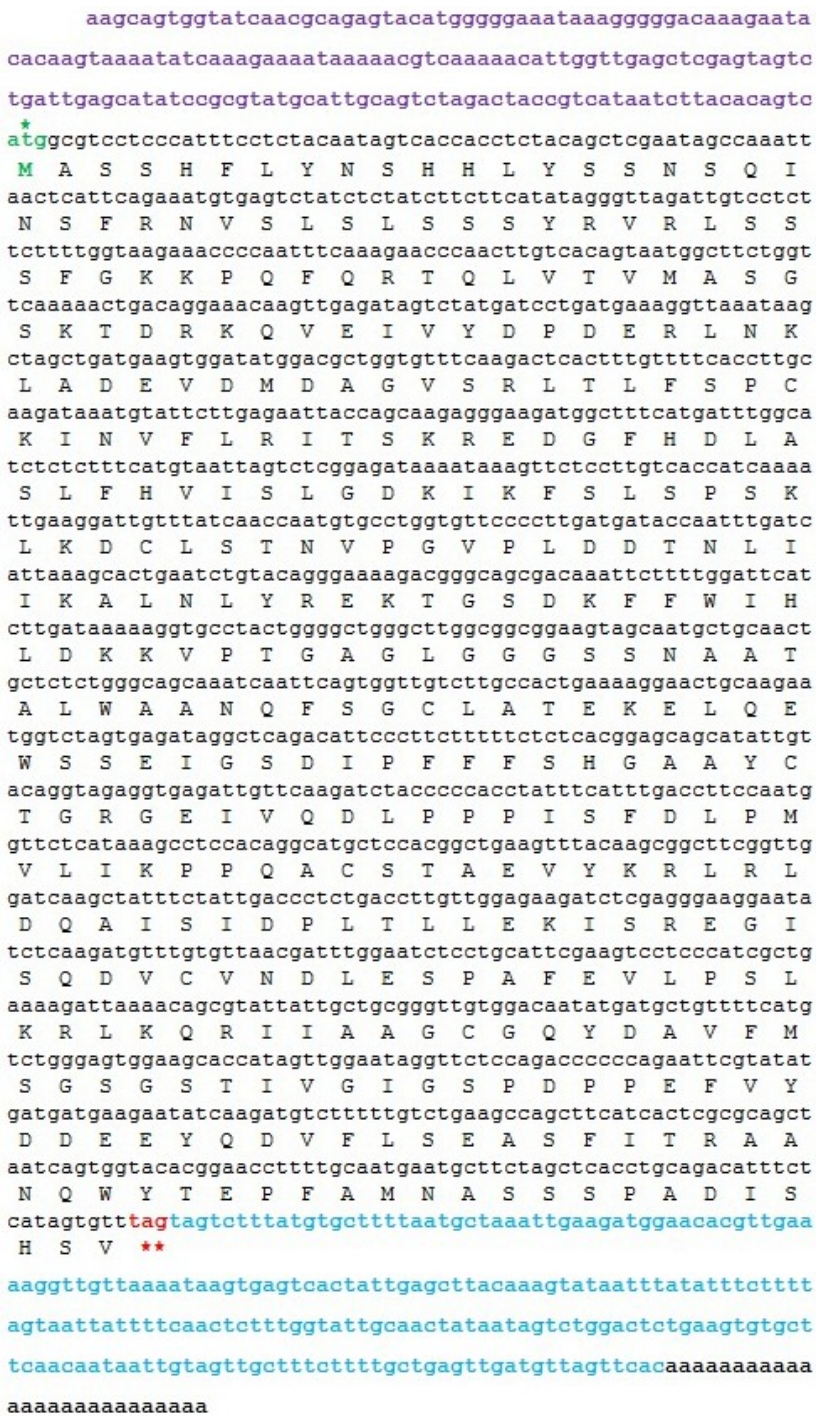


Figure 4.42 Nucleotide and deduced amino acid sequence of *P. sokpayensis 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase* (*PsCMK*, MF682467). The polypeptide sequence is represented by a single letter amino acid code under respective codon. The start codon and its corresponding amino acid are colored green and indicated by "*". The stop codon is colored red and indicated by "**". 5′ untranslated region (5′ UTR) is colored in purple and 3′ UTR is colored in blue. PolyA tail is represented by a stretch of adenine residues at the end.

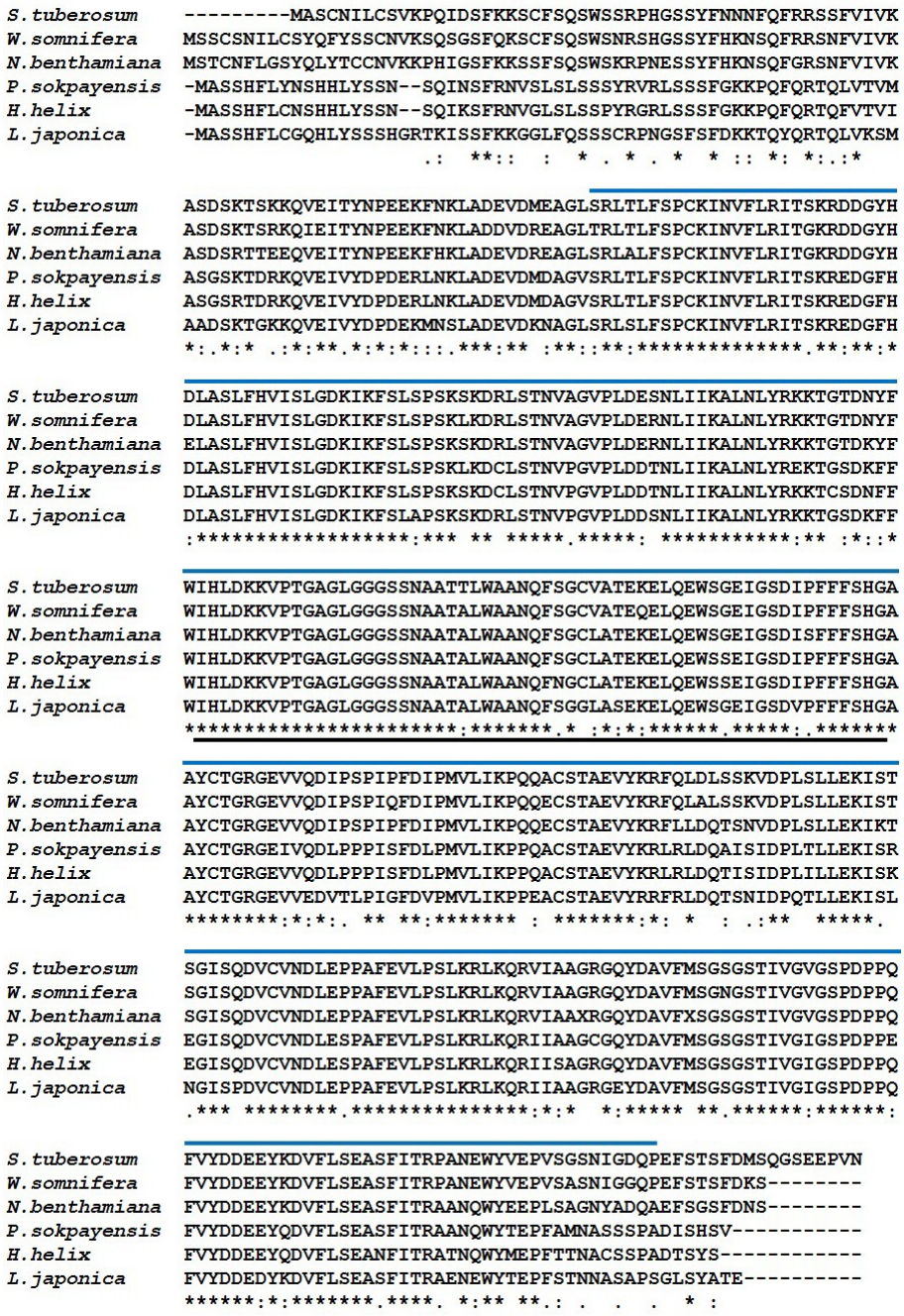


Figure 4.43 Multiple sequence alignment of deduced amino acid sequences of CMKs from the *Panax* species and other plants. 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK/IspE) domain is shown by a line in blue. GHMP kinases N terminal domain is shown by a black line. The sequences were retrieved from GenBank. Their accession numbers are as follows: *P. sokpayensis* (MF682467), *S. tuberosum* (XP_006362734), *Withania somnifera* (AOX15283), *Nicotiana benthamiana* (ABO87658), *H. helix* (APY22344), *Lonicera japonica* var. *chinensis* (AGE10581).

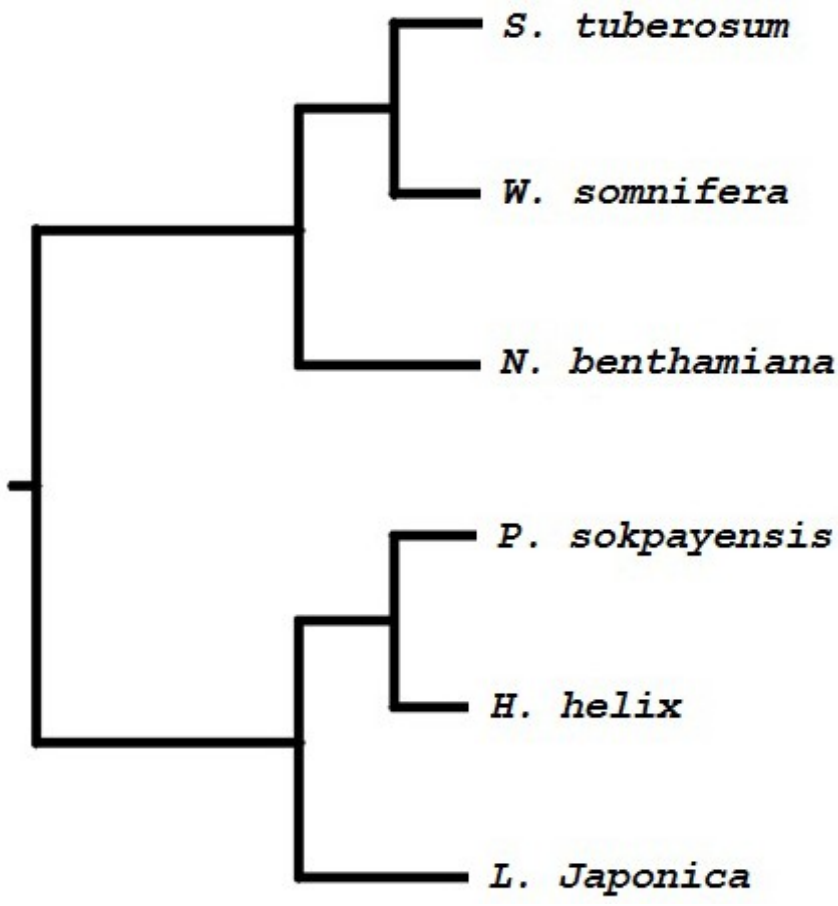


Figure 4.44 Phylogenetic tree constructed using UPGMA from the deduced amino acid sequences of CMKs from different plants retrieved from NCBI GenBank. Their accession numbers are as follows: *P. sokpayensis* (MF682467), *S. tuberosum* (XP_006362734), *W. somnifera* (AOX15283), *N. benthamiana* (ABO87658), *H. helix* (APY22344), *L. japonica* var. *chinensis* (AGE10581).

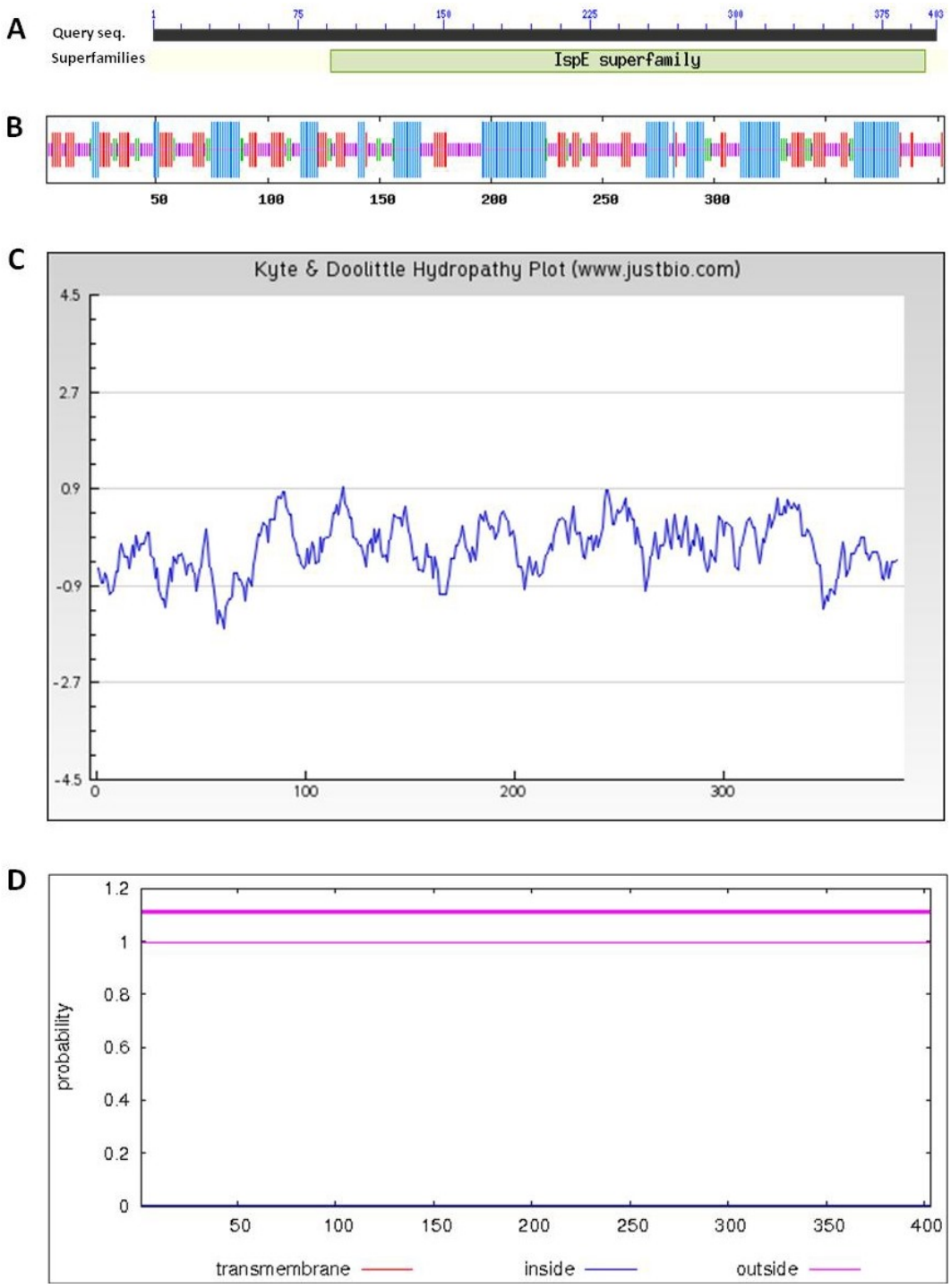


Figure 4.45 (A) Prediction of protein identity using NCBI conserved domain analysis (B) Secondary structure prediction of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK) using the deduced amino acid sequence. Blue lines indicate alpha helices, purple lines indicate random coils, red lines denote extended strands and green lines represent beta turns (C) Kyte and Doolittle hydropahty plot of Ps CMK. (D) Prediction of transmembrane region using TMHMM program.

4.4. Gene expression analysis of key genes of ginsenoside biosynthetic pathway

The expressions of the five ginsenoside biosynthetic pathway genes were performed in leaf, stem and rhizome tissues using qRT-PCR (Figure 4.46). The expression of four ginsenoside biosynthetic pathway genes, viz., *PsFPS*, *PsSS*, *PsDS* and *PsCS* were highest in rhizome when compared to other two tissues while the expression of *PsSE* was highest in leaf. The relative expression of *PsFPS* was ~1.5- and ~3.7-fold higher in rhizome than that of leaf and stem, respectively (Figure 4.46A). The expression of *PsSS* was high in the rhizome (~3-fold) as compared to that of leaf and stem (Figure 4.46B). On the other hand, the expression of *PsDS* was ~3.7- and ~19- fold higher in the rhizome than that of the leaf and stem, respectively (Figure 4.46C). The expression of *PsSE* was ~ 1.3 and ~1.5 fold higher in leaf than in rhizome and stem, respectively (Figure 4.46D). In case of *PsCS*, the expression was ~3.5 and ~2.4 fold higher in rhizome when compared to leaf and stem, respectively (Figure 4.46E).

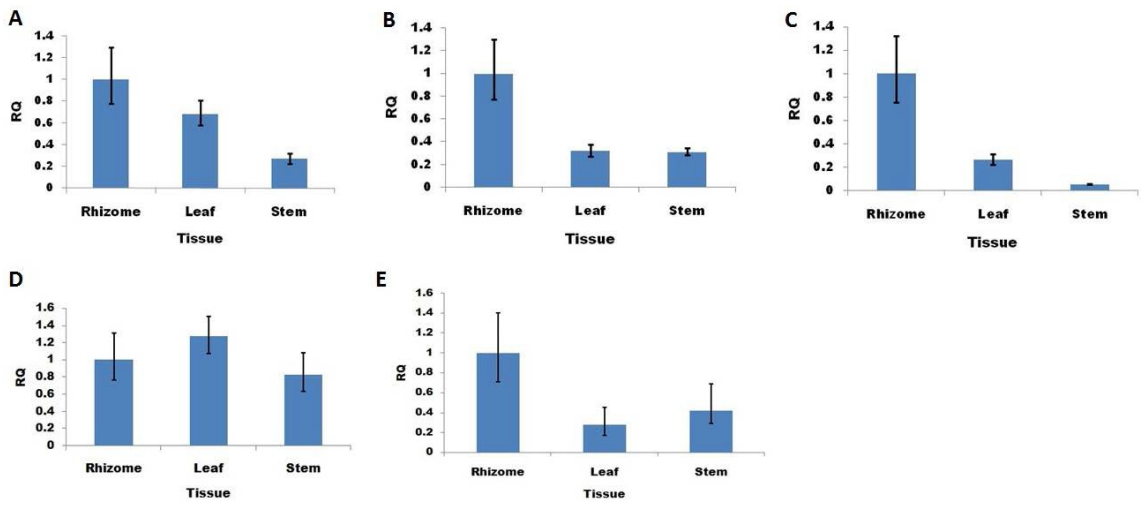


Figure 4.46 Relative expressions of (A) *farnesyl pyrophosphate synthase*, (B) *squalene synthase*, (C) *dammarenediol synthase*, (D) *squalene epoxidase* and (E) *cycloartenol synhtase* in rhizome, leaf and stem tissues. Data represented are relative quantitation (RQ) values of gene expression. Expression is shown after normalization to 18S rRNA. Values were calculated using ΔΔCt method and the error bars represent RQMIN and RQMAX. Data are representative of three biological replicate experiments. Source: Adapted from Gurung et al., 2016.

4.5. Genes expression analysis of other differentially expressed genes obtained through SSH

qRT-PCR analysis showed that the expression pattern of most of the genes from

both libraries largely corroborate with RT-PCR analysis (Figures 4.9 and 4.47). The expression of *galactinol synthase 2* was high in leaf but not detected in the stem and rhizome tissues (Figure 4.47A). The expression of *cell division cycle 20.1* was high (~4-fold) in the leaf than that of the stem and rhizome (Figure 4.47B). On the other hand, expression of *metallothionein 3-like protein* was ~2.2 and ~19-fold higher in the stem than that in the leaf and rhizome, respectively (Figure 4.47C). The expression of *GBR-5-like protein* was high (~7.0-fold) in the stem as compared to the leaf, whereas its expression was not detected in the rhizome (Figure 4.47D). The expression of *PsbA* was high in the leaf as compared to the stem and rhizome (Figure 4.47E). On the other hand, the expression of all the transcripts from the rhizome SSH library was high in the rhizome than that of the leaf and stem. The expression of *protein KIAA0664 homologue* was ~3.6 and ~2-fold higher in the rhizome than that of the leaf and stem, respectively (Figure 4.47F). Three genes, *major latex-like protein, RNase-like major storage protein* and *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) were found to be preferentially expressed in the rhizome when compared to the leaf and stem (Figures 4.47G-I). The expression of *ankyrin repeat-containing protein* was high as compared to the leaf and stem (Figure 4.47J). In case of *β-amylase*, *bHLH96*, and *ubiquitin-conjugating enzyme E2*, the expression was negligible in the leaf and stem as compared to the rhizome (Figures 4.47K-M). The expression of *polyubiquitin-like protein* was higher in the rhizome (~2.5- and ~5.6-fold) as compared to the leaf and stem, respectively (Figure 4.47N). A similar trend of expression was observed in *heat shock protein 70* with higher expression in the rhizome, ~3.4 and ~5.4-fold when compared to the leaf and stem (Figure 4.47O).

4.6. Heterologous expression of PsFPS in *E. coli*

Open reading frame of *PsFPS* was amplified using the primers designed from the start to stop codon (Table 4, Appendix C) which yielded 1029 bp fragment (Figure 4.48). The amplicon was eluted and cloned into pre-digested pQE30 expression vector, an expression vector with T5 promoter and His tag segment. The pQE30-*PsFPS* construct was transformed and the expression was induced. After confirmation of in-frame cloning of *PsFPS* in pQE30, the gene was induced for protein expression. The recombinant PsFPS fusion protein of size between 36 – 40 kDa was observed on 12 % denaturing SDS–PAGE gel (Figure 4.49). The time course analysis of expression of recombinant protein revealed high level of expression within 1 h after induction by 1 mM IPTG (Figure 4.49).

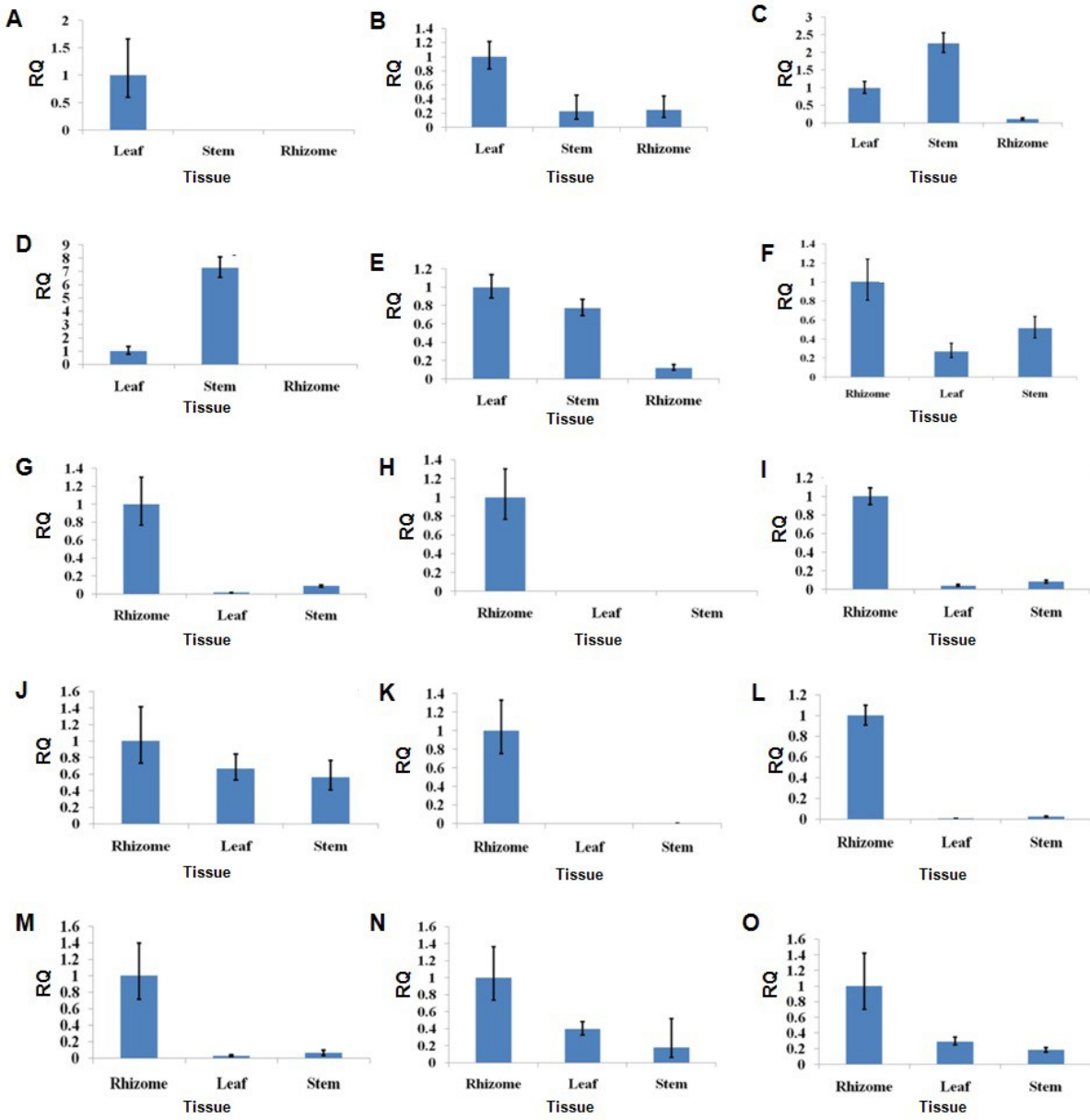


Figure 4.47 Relative expressions of (A) galactinol *synthase 2*, (B) *cell division cycle 20.1*, (C) *metallothionein 3-like protein*, (D) *GBR5-like protein*, (E) *PsbA*, (F) *protein KIAA0664 homologue*, (G) *major latex-like protein*, (H) *RNase-like major storage protein*, (I) *glyceraldehyde-3-phosphate dehydrogenase*, (J) *ankyrin repeat-containing protein* (K) *β-amylase*, (L) *bHLH*, (M) *ubiquitin-conjugating enzyme E2*, (N) *polyubiquitin like protein* and (O) *heat shock protein 70* in rhizome, leaf and stem tissues. Data represented here are relative quantitation (RQ) values of gene expression. Expression is shown after normalization to 18S rRNA. Values were calculated using ΔΔCT method, and the error bars represented RQMIN and RQMAX. Data are representative of three biological replicate experiments. Source: Adapted from Gurung et al., 2016.

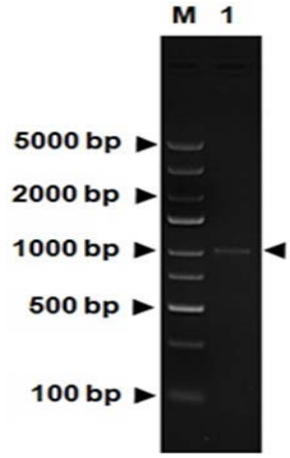


Figure 4.48 Agarose gel showing DNA band of open reading frame (ORF) of *PsFPS*.

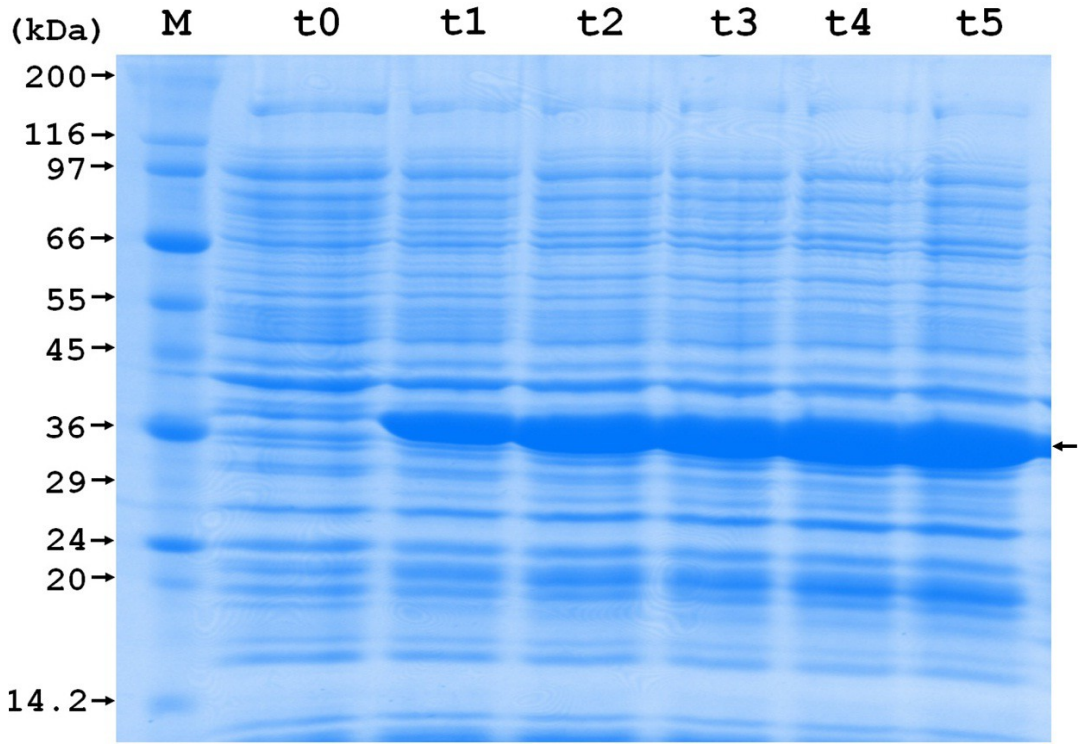


Figure 4.49 Induction analysis of recombinant PsFPS in *E. coli* M15 (pREP4). Crude protein extracts of approximately equal number of *E. coli* cells calculated using OD_{600} were used for detection by SDS PAGE stained with Coomassie Brilliant Blue. M – protein molecular weight marker; t0 – non induced control at 37 °C; t1, t2, t3, t4 and t5 – PsFPS produced after 1h, 2h, 3h, 4h and 5h of induction, respectively. The arrow indicates the position and size of recombinant PsFPS protein (~40 kDa).