Supplementary Materials

Ancient origin and conserved gene function in terpene pheromone and defense evolution of stink bugs and hemipteran insects

Zarley Rebholz[1†], Jason Lancaster[1†], Hailey Larose[1†], Ashot Khrimian[2], Katrin Luck[3], Michael E. Sparks[2], Kerry L. Gendreau[1], Leena Shewade[4], Tobias G. Köllner[3], Donald C. Weber[2], Dawn E. Gundersen-Rindal[2], Paul O'Maille[4], Alexandre V. Morozov[5], and Dorothea Tholl[1*]

[1] Department of Biological Sciences, Virginia Tech, Latham Hall, 220 Ag Quad Lane, Blacksburg, VA, 24061, USA

[2] Invasive Insect Biocontrol and Behavior Laboratory, USDA Agricultural Research Service, Beltsville, MD, 20705, USA

[3] Department of Biochemistry, Max Planck Institute for Chemical Ecology, Hans-Knöll-Strasse 8, D-07745 Jena, Germany

[4] SRI International, Biosciences Division, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493, USA

[5] Department of Physics & Astronomy and Center for Quantitative Biology, Rutgers University, 136 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA

*Corresponding author: Dorothea Tholl, phone: +1-540-231-4567; e-mail: tholl@vt.edu

Tables S1, S4, and S5 are in xlsx format and can be downloaded in this format.

Table S2. Placement of *H. halys* IDS-type genes in genome scaffolds

H. halys Genome Wide Shotgun Assembly Hhal_1.0	
NW_014466702.1	
*Hh*IDS3	Location 607289..620830
*Hh*IDS4	Location 625390..639949
*Hh*IDS7/TPS1 - MG917093	Location 660896.. 647588
NW_014466461.1	
*Hh*IDS1/TPS2 - MG870388	Location (71539) 66498..56095
*Hh*IDS5	Location 90118..75481
*Hh*IDS6	Location 107902..93802
NW_014466714.1	
*Hh*IDS2/FPPS - MG870389	Location 76720..43812

Table S3. Primer sequences used for amplification, cloning, and qRT-PCR analysis of IDS-type genes from *H. halys* and *N. viridula.*

Gene	Primers (5'-3')	Amplicon size (bp)	Purpose
HhIDS1/ *HhTPS2*	HhIDS1_1F ATGATACCGAAGACGCTTGG HhIDS1_1134R TTATGGAGCTTTTAGGATCTCCAATTC	1134	Amplification and cloning from cDNA
	HhIDS1_QF GCCAAGAAGCAGCCATCTATG HhIDS1_QR CACATCTTGGTGAAACCTGGATC	100	qRT-PCR
	HhFPPS1_860F CAATGAGTCACTTTAATCCGGCC HhFPPS1_958R CGTCAAACAGTTCGCGTACTC	100	qRT-PCR
HhIDS2/ *HhFPPS*	HhFPPS_1F ATGCCTTTTGCAAAACTGTG HhFPPS_1212R CTACTGCTTTCTACCATACATCTTATG	1212	Amplification and cloning from cDNA
	HhFPPS 240 F TGTTCGGGATCTGACACAGC HhFPPS 382 R GGTCAGGAGCAAGCATACGA	100	qRT-PCR
HhIDS3	HhIDS3_1F ATGGCGTTCGTGTCTGC HhIDS3_1125R TTAATCTAAATTTTCATCAGGAGTTTCTC	1125	Amplification from cDNA
HhIDS4	HhIDS4_1F ATGGCGAACATGGCTGG HhIDS4_1143R TCAAACATTCGTAACTTTAGGGTC	1143	Amplification from cDNA
HhIDS5	HhIDS5_1F ATGGCGTCAAAGGTGTCG HhIDS5_1131R TCAGAATGATTCTAATCTTTCAAGTTGAA	1131	Amplification from cDNA
HhIDS6	HhIDS6_1F ATGGCAGCGAAGGCATC HhIDS6_1131R TCAGAATGATTTTAATCGTTCAAGTTG	1131	Amplification from cDNA
HhIDS7/ *HhTPS1*	HhIDS7_1F ATGGCGTCCGTGGCTAC HhIDS7_1107R TCACTCTTCTCGAATCACGAGC	1107	Amplification and cloning from cDNA
	HhFPPS7_265F AGGATTGTAGCCGATGAGAGC HhFPPS7_364R GTACTGCTGACATCGTGAACAAC	100	qRT-PCR
NvTPS2	NvTPSp_F ATGGTGTCCGTTGCCG NvTPSp_R2 TCAAACATTCGCAACTTCAAG	1143	Amplification from cDNA
	NvTPSp_TEV_F GAAAACCTGTATTTTCAGGGCATGGTGTCCGTTGCCG	1164	Addition of TEV cleavage site to NvTPS2

NvTPSp_R1_attb2 GGGGACCACTTTGTACAAGAAAGCTGGGTATTCTCGAATTACAAGCCACTC	1205	Gateway cloning into pDONR/Zeo
attB1_tev GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGC		

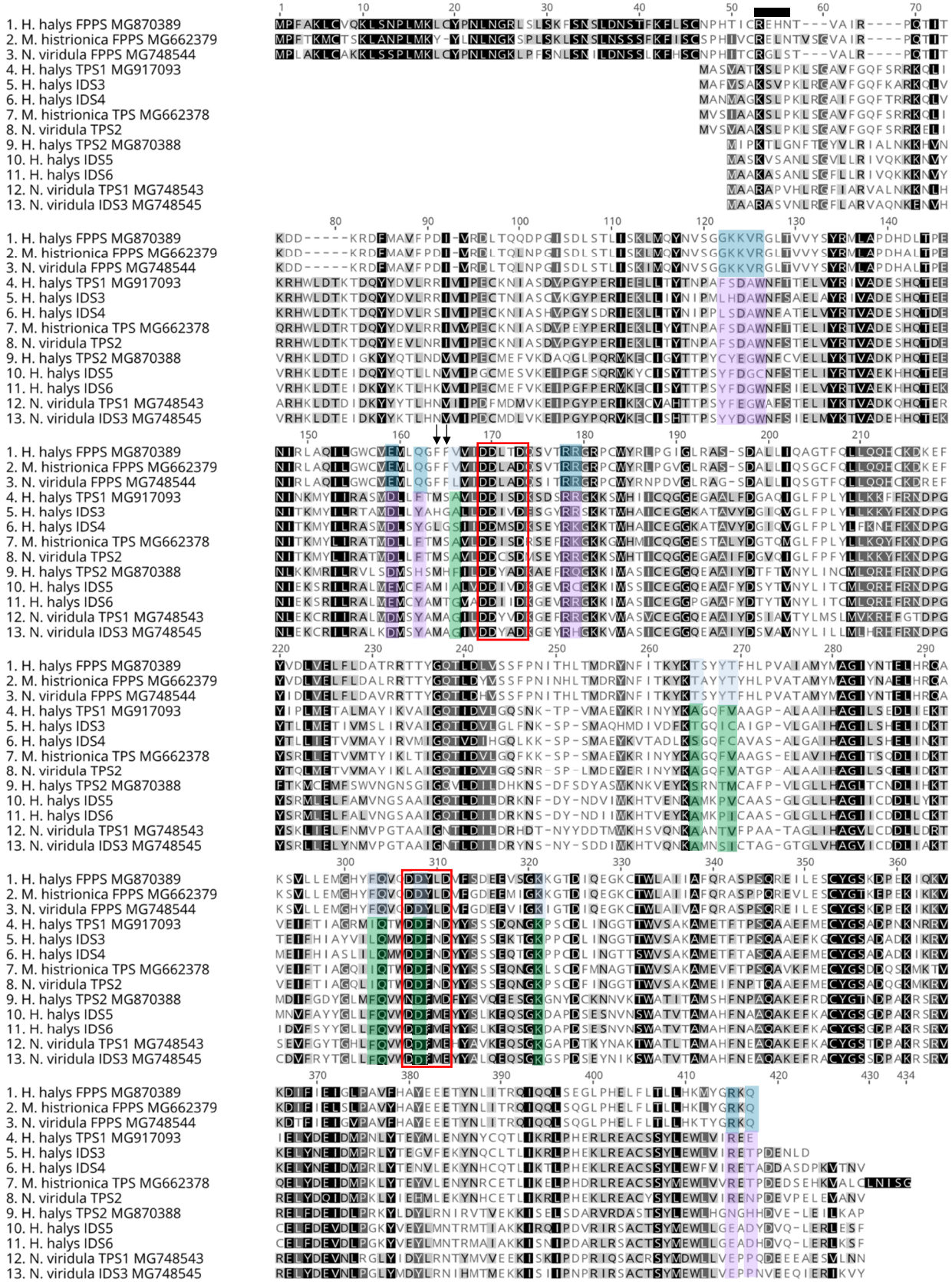


Fig. S1. Amino acid sequence alignment of IDS (FPPS) and IDS-type enzymes of *H. halys*, *M. histrionic*a, and *N. viridula*. *H. halys* FPPS = *Hh*IDS2; *H. halys* TPS1 = *Hh*IDS7; *H. halys* TPS2 = *Hh*IDS1. Red boxes indicate the position of the first and second aspartate rich motifs. Dark blue color marks residues in FPPS proteins that bind the diphosphate moiety of IPP and magenta color marks corresponding residue substitutions in IDS-type proteins. Light blue color marks residues in FPPS proteins that bind the prenyl tail of IPP and green color marks corresponding residue substitutions in IDS-type proteins. Arrows mark residues at positions 4 and 5 upstream of the first aspartate rich motif. The black bar indicates a putative mitochondrial targeting sequence cleavage site in FPPS proteins.

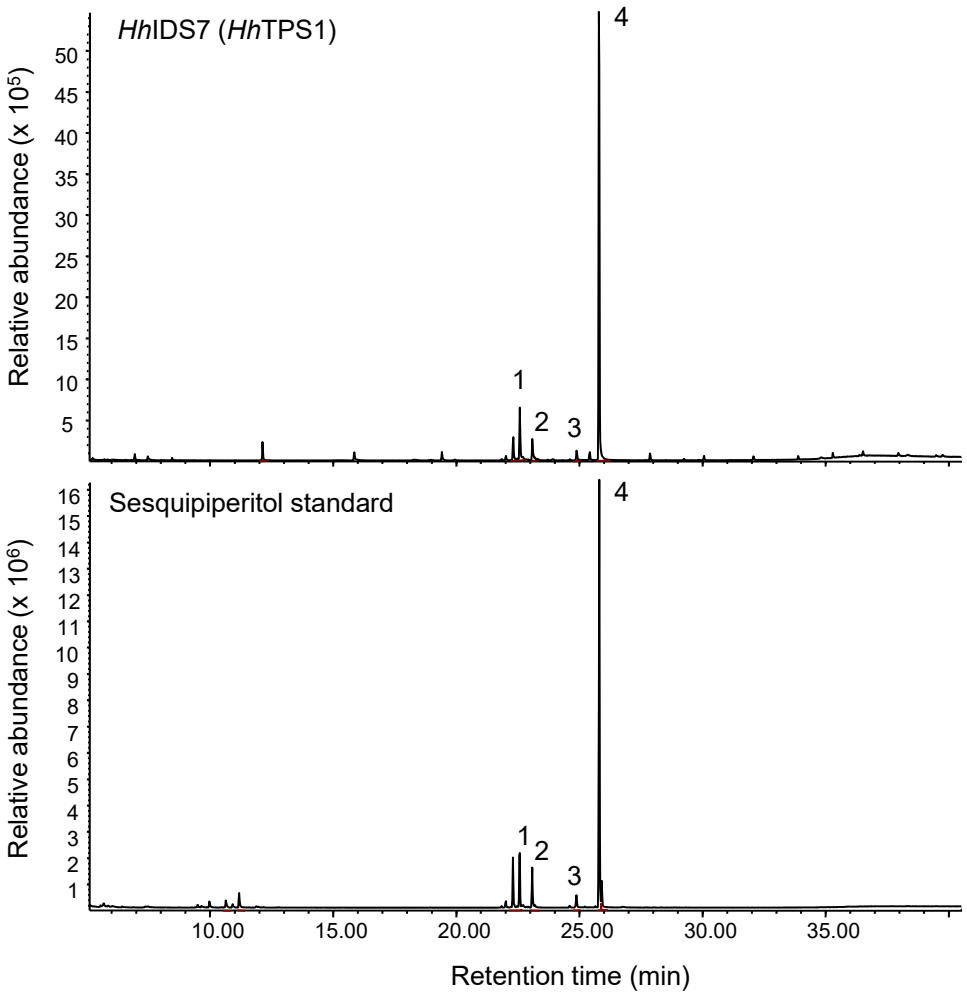


Fig. S2. GC-MS analysis of a hexane extract from an *Hh*IDS7 (*Hh*TPS1) activity assay in comparison to a sesquipiperitol standard. The sample was injected without split at 260°C. 1, zingiberene; 2, sesquiphellandrene; 3, *RSR*-zingiberenol; 4, sesquipiperitol

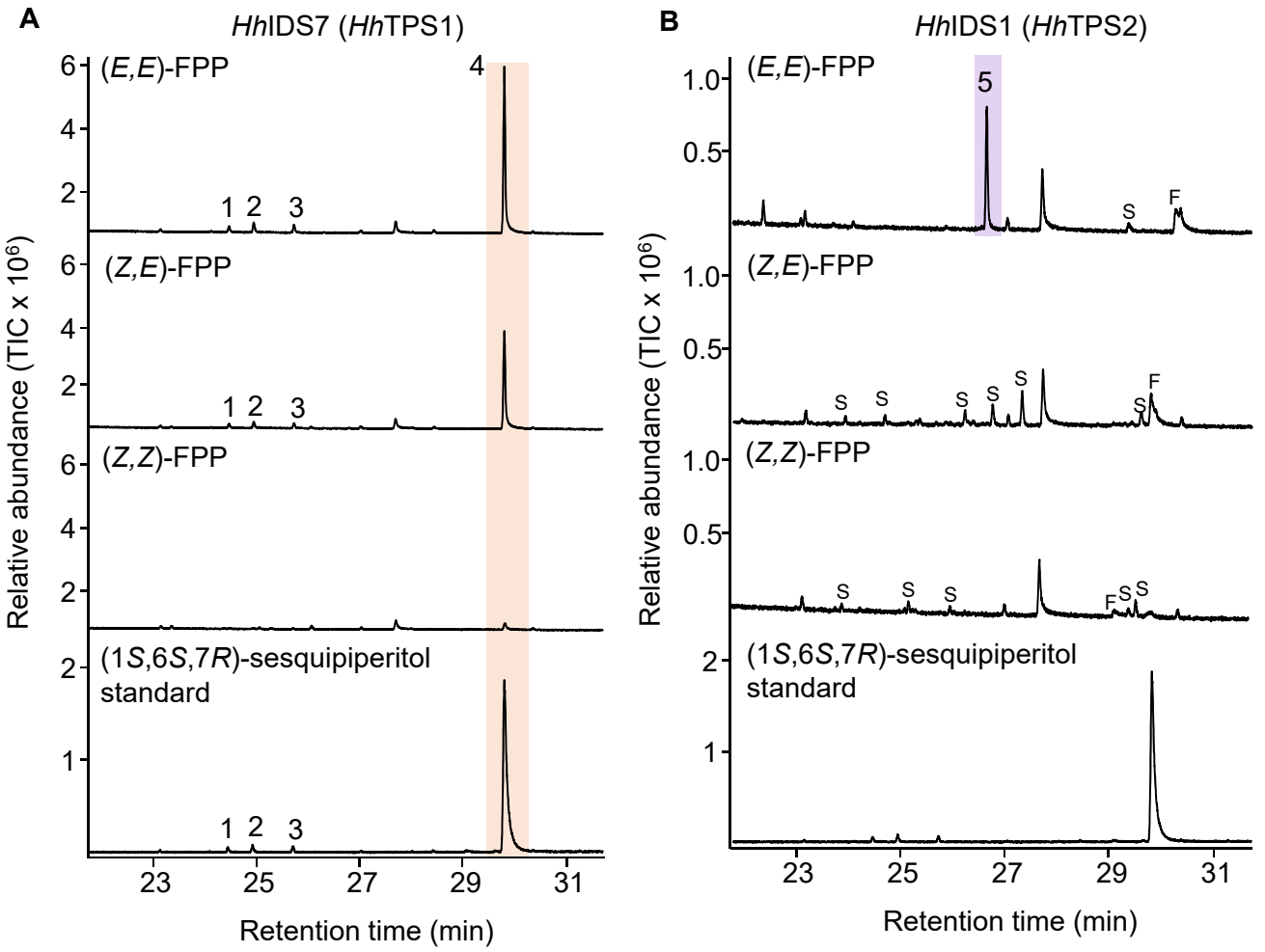


Fig. S3. Terpene synthase assay of *Hh*IDS7 *(Hh*TPS1) and *Hh*IDS1 *(Hh*TPS2) with various FPP isomeric substrates. Partially purified enzymes were assayed with 50 µM FPP substrates in assay buffer. (*A*) *Hh*TPS1 produces an isomer of the pheromone precursor sesquipiperitol (4) along with thermal rearrangement products (1-3). (*B*) *Hh*TPS2 produces putative sesquiterpene products (S) with the most abundant compound being identified as elemol (5). 1, γ-curcumene; 2, α-zingiberene; 3, β-sesquiphellandrene; 4, sesquipiperitol isomer, 5, elemol; F, farnesol isomer. Elemol was identified only by mass-spectral library comparison.

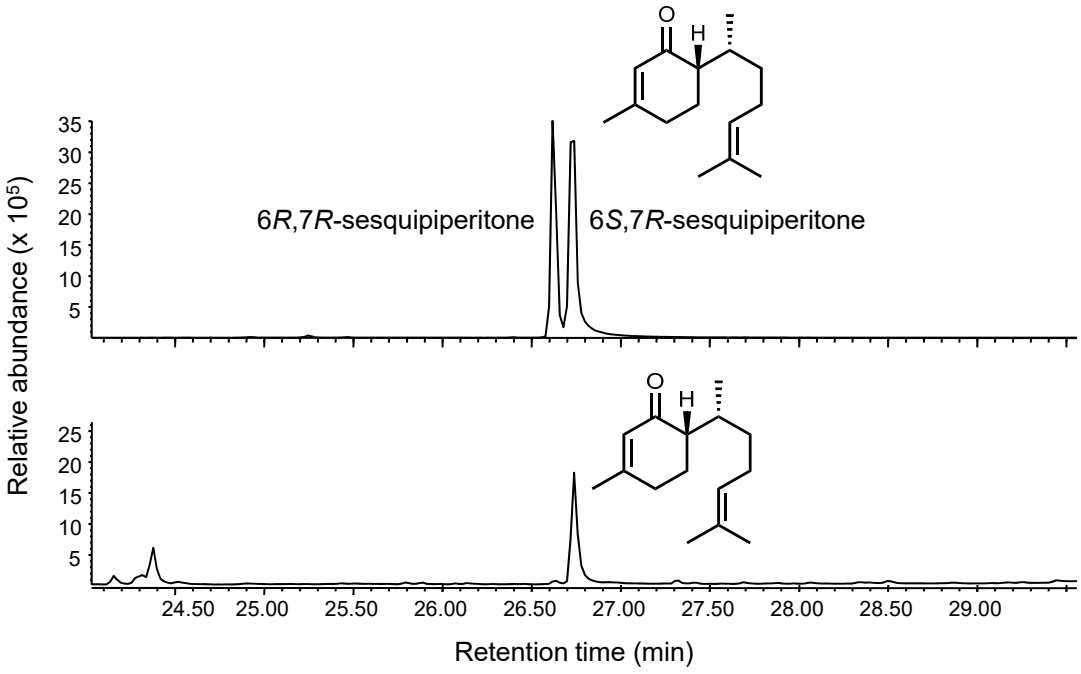


Fig. S4. Gas-chromatographic determination of the relative configuration at C-6 and C-7 of the *Hh*IDS7 (*Hh*TPS1) sesquipiperitol product. Gas chromatograms of the oxidized enzymatic product (sesquipiperitone, lower panel) and that of 6*R*,7*R*/6*S*,7*R* sesquipiperitone standards (upper panel). Comparison of retention times defined a relative 6*S*,7*R* or 6*R*,7*S* configuration of the *Hh*IDS7 product.

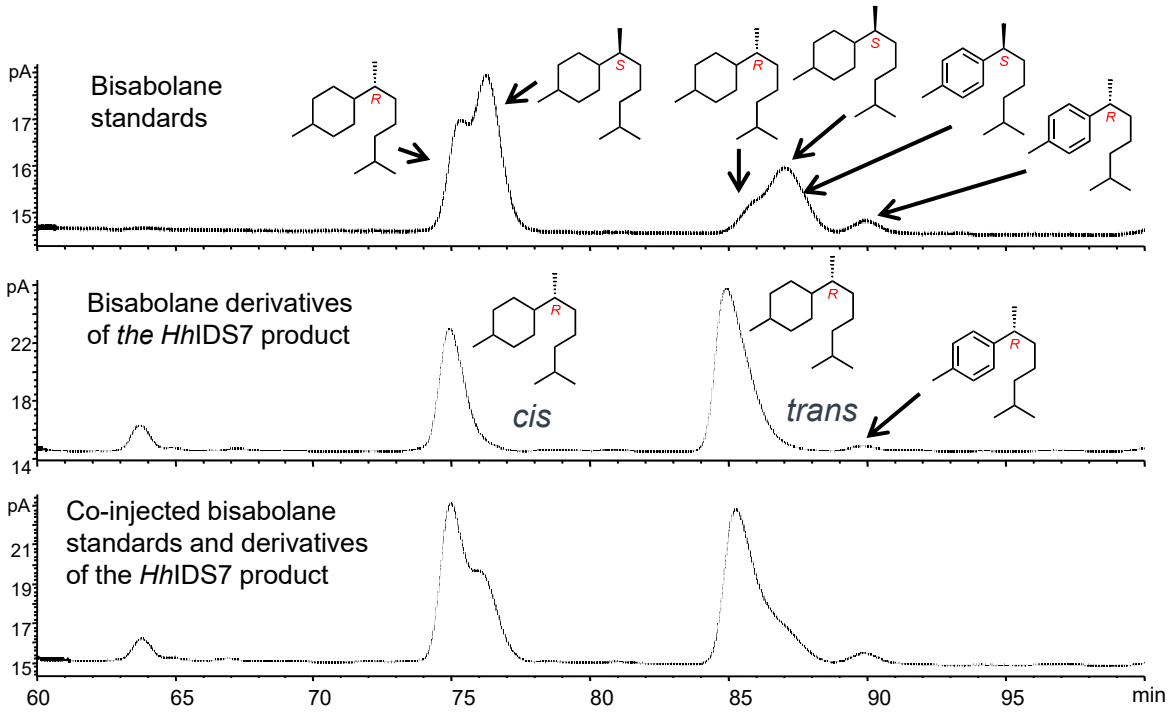


Fig. S5. Gas-chromatographic determination of the absolute configuration at C-6 and C-7 of the *Hh*IDS7 (*Hh*TPS1) sesquipiperitol product. The enzymatic product was converted to bisabolanes (middle panel) and their retention times were compared to those of bisabolane standards of 7*S* and 7*R* configuration by separate injection (upper panel) and by co-injection with the standards (lower panel). The stereochemistry at C-7 was determined to be *R*, which concluded an absolute configuration of 6*S*,7*R* of the *Hh*IDS7 product.

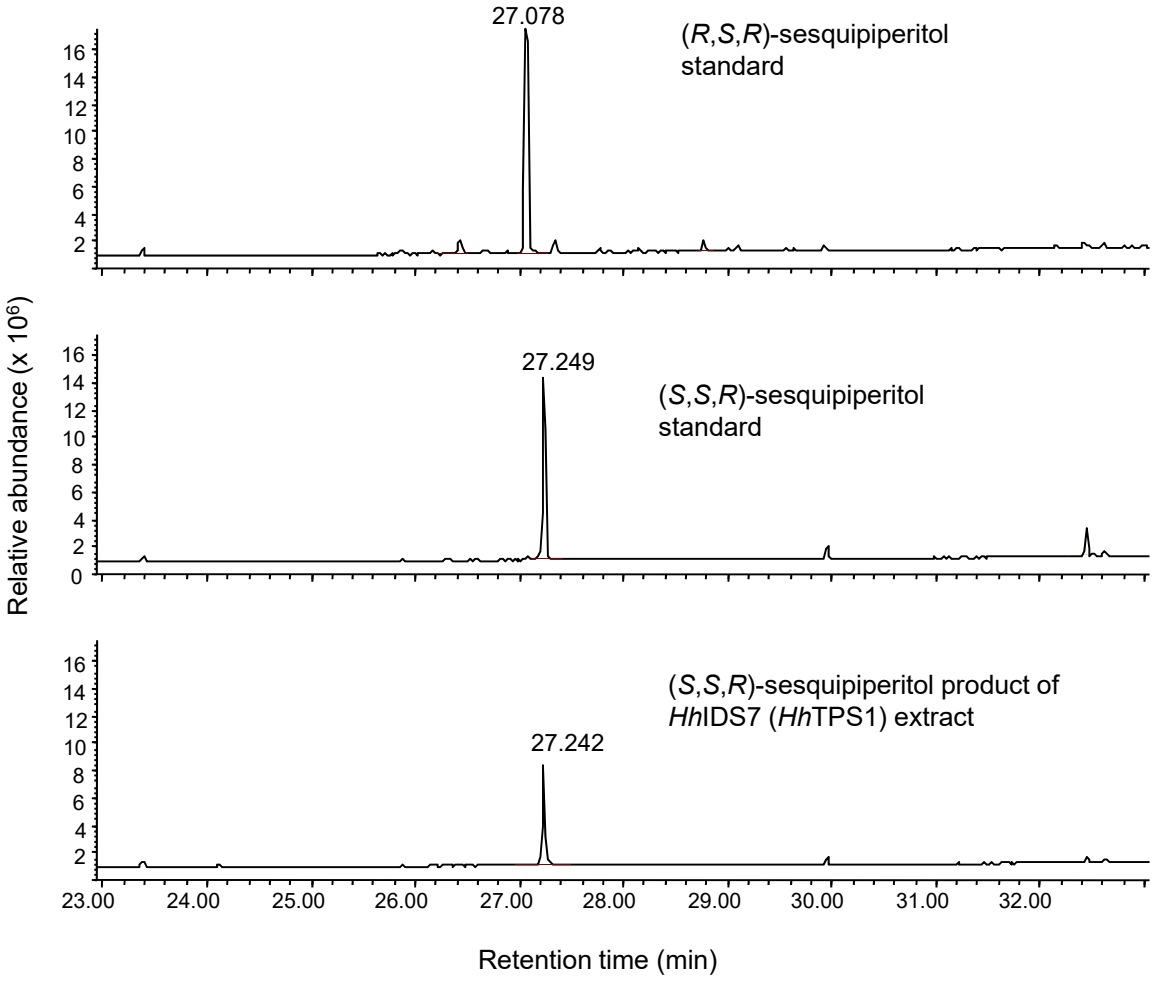
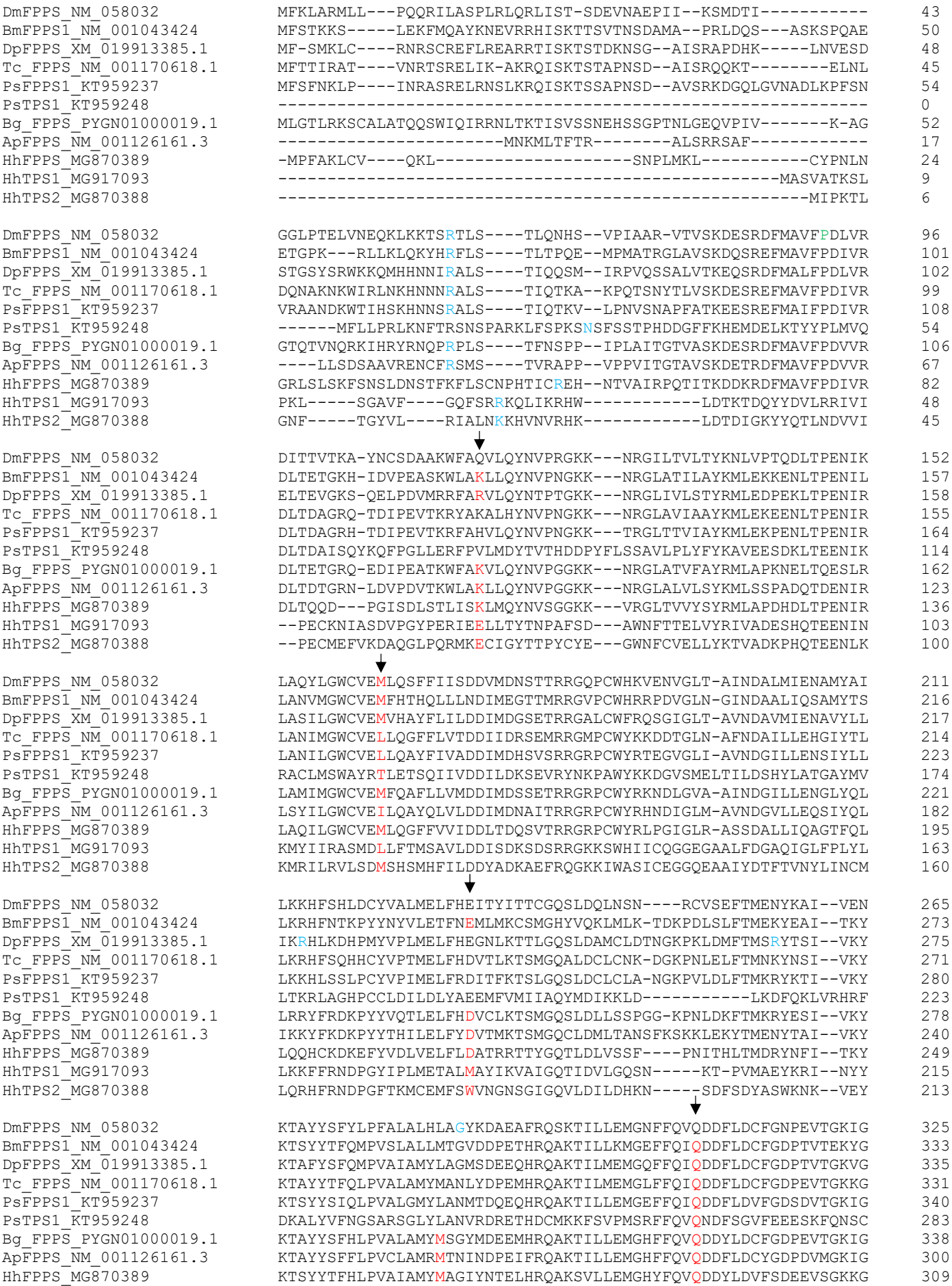


Fig. S6. Gas-chromatographic comparison of sesquipiperitol produced by *Hh*TPS7 (*Hh*TPS1) with (1*S*,6*S*,7*R*)- and (1*R*,6*S*,7*R*)-sesquipiperitols on a chiral column. Retention time comparisons determined an (*S*) configuration at C-1 and thus an overall (1*S*,6*S*,7*R*) stereochemistry of the enzymatic product.

Fig. S7. Transcript abundance of *HhFPPS, HhTPS1, HhTPS2* as determined by qRT-PCR in mature *H. halys* female and male whole bug tissue (n=3, ±SD). Significance was determined using student's t-test and means grouped by Tukey's HSD; *P < 0.05. *H. G*ene expression was normalized against the *RpS4* housekeeping gene and transcript abundance is shown relative to that in female as determined by qRT-PCR**.**


DmFPPS_NM_058032 MFKLARMLL---PQQRILASPLRLQRLIST-SDEVNAEPII--KSMDTI---------- 43
BmFPPS1_NM_001043424 MFSTKKS-----LEKFMQAYKNEVRRHISKTTSVTNSDAMA--PRLDQS---ASKSPQAE 50
DpFPPS_XM_019913385.1 MF-SMKLC----RNRSCREFLREARRTISKTSTDKNSG--AISRAPDHK-----LNVESD 48
Tc_FPPS_NM_001170618.1 MFTTIRAT----VNRTSRELIK-AKRQISKTSTAPNSD--AISRQQKT--------ELNL 45
PsFPPS1_KT959237 MFSFNKLP----INRASRELRNSLKRQISKTSSAPNSD--AVSRKDGQLGVNADLKPFSN 54
PsTPS1_KT959248 ---------- 0
Bg_FPPS_PYGN01000019.1 MLGTLRKSCALATQQSWIQIRRNLTKTISVSSNEHSSGPTNLGEQVPIV-------K-AG 52
ApFPPS_NM_001126161.3 ----------MNKMLTFTR--------ALSRRSAF---------- 17
HhFPPS_MG870389 -MPFAKLCV----QKL----------SNPLMKL----------CYPNLN 24
HhTPS1_MG917093 ----------MASVATKSL 9
HhTPS2_MG870388 ----------MIPKTL 6
DmFPPS_NM_058032 GGLPTELVNEQKLKKTSRTLS----TLQNHS--VPIAAR-VTVSKDESRDFMAVFPDLVR 96
BmFPPS1_NM_001043424 ETGPK---RLLKLQKYHRFLS----TLTPQE--MPMATRGLAVSKDQSREFMAVFPDIVR 101
DpFPPS_XM_019913385.1 STGSYSRWKKQMHHNNIRALS----TIQQSM--IRPVQSSALVTKEQSRDFMALFPDLVR 102
Tc_FPPS_NM_001170618.1 DQNAKNKWIRLNKHNNNRALS----TIQTKA--KPQTSNYTLVSKDESREFMAVFPDIVR 99
PsFPPS1_KT959237 VRAANDKWTIHSKHNNSRALS----TIQTKV--LPNVSNAPFATKEESREFMAIFPDIVR 108
PsTPS1_KT959248 ------MFLLPRLKNFTRSNSPARKLFSPKSNSFSSTPHDDGFFKHEMDELKTYYPLMVQ 54
Bg_FPPS_PYGN01000019.1 GTQTVNQRKIHRYRNQPRPLS----TFNSPP--IPLAITGTVASKDESRDFMAVFPDVVR 106
ApFPPS_NM_001126161.3 ----LLSDSAAVRENCFRSMS----TVRAPP--VPPVITGTAVSKDETRDFMAVFPDVVR 67
HhFPPS_MG870389 GRLSLSKFSNSLDNSTFKFLSCNPHTICREH--NTVAIRPQTITKDDKRDFMAVFPDIVR 82
HhTPS1_MG917093 PKL-----SGAVF----GQFSRRKQLIKRHW----------LDTKTDQYYDVLRRIVI 48
HhTPS2_MG870388 GNF-----TGYVL----RIALNKKHVNVRHK----------LDTDIGKYYQTLNDVVI 45
DmFPPS_NM_058032 DITTVTKA-YNCSDAAKWFAQVLQYNVPRGKK---NRGILTVLTYKNLVPTQDLTPENIK 152
BmFPPS1_NM_001043424 DLTETGKH-IDVPEASKWLAKLLQYNVPNGKK---NRGLATILAYKMLEKKENLTPENIL 157
DpFPPS_XM_019913385.1 ELTEVGKS-QELPDVMRRFARVLQYNTPTGKK---NRGLIVLSTYRMLEDPEKLTPENIR 158
Tc_FPPS_NM_001170618.1 DLTDAGRQ-TDIPEVTKRYAKALHYNVPNGKK---NRGLAVIAAYKMLEKEENLTPENIR 155
PsFPPS1_KT959237 DLTDAGRH-TDIPEVTKRFAHVLQYNVPNGKK---TRGLTTVIAYKMLEKPENLTPENIR 164
PsTPS1_KT959248 DLTDAISQYKQFPGLLERFPVLMDYTVTHDDPYFLSSAVLPLYFYKAVEESDKLTEENIK 114
Bg_FPPS_PYGN01000019.1 DLTETGRQ-EDIPEATKWFAKVLQYNVPGGKK---NRGLATVFAYRMLAPKNELTQESLR 162
ApFPPS_NM_001126161.3 DLTDTGRN-LDVPDVTKWLAKLLQYNVPGGKK---NRGLALVLSYKMLSSPADQTDENIR 123
HhFPPS_MG870389 DLTQQD---PGISDLSTLISKLMQYNVSGGKK---VRGLTVVYSYRMLAPDHDLTPENIR 136
HhTPS1_MG917093 --PECKNIASDVPGYPERIEELLTYTNPAFSD---AWNFTTELVYRIVADESHQTEENIN 103
HhTPS2_MG870388 --PECMEFVKDAQGLPQRMKECIGYTTPYCYE---GWNFCVELLYKTVADKPHQTEENLK 100
DmFPPS_NM_058032 LAQYLGWCVEMLQSFFIISDDVMDNSTTRRGQPCWHKVENVGLT-AINDALMIENAMYAI 211
BmFPPS1_NM_001043424 LANVMGWCVEMFHTHQLLLNDIMEGTTMRRGVPCWHRRPDVGLN-GINDAALIQSAMYTS 216
DpFPPS_XM_019913385.1 LASILGWCVEMVHAYFLILDDIMDGSETRRGALCWFRQSGIGLT-AVNDAVMIENAVYLL 217
Tc_FPPS_NM_001170618.1 LANIMGWCVELLQGFFLVTDDIIDRSEMRRGMPCWYKKDDTGLN-AFNDAILLEHGIYTL 214
PsFPPS1_KT959237 LANILGWCVELLQAYFIVADDIMDHSVSRRGRPCWYRTEGVGLI-AVNDGILLENSIYLL 223
PsTPS1_KT959248 RACLMSWAYRTLETSQIIVDDILDKSEVRYNKPAWYKKDGVSMELTILDSHYLATGAYMV 174
Bg_FPPS_PYGN01000019.1 LAMIMGWCVEMFQAFLLVMDDIMDSSETRRGRPCWYRKNDLGVA-AINDGILLENGLYQL 221
ApFPPS_NM_001126161.3 LSYILGWCVEILQAYQLVLDDIMDNAITRRGRPCWYRHNDIGLM-AVNDGVLLEQSIYQL 182
HhFPPS_MG870389 LAQILGWCVEMLQGFFVVIDDLTDQSVTRRGRPCWYRLPGIGLR-ASSDALLIQAGTFQL 195
HhTPS1_MG917093 KMYIIRASMDLLFTMSAVLDDISDKSDSRRGKKSWHIICQGGEGAALFDGAQIGLFPLYL 163
HhTPS2_MG870388 KMRILRVLSDMSHSMHFILDDYADKAEFRQGKKIWASICEGGQEAAIYDTFTVNYLINCM 160
DmFPPS_NM_058032 LKKHFSHLDCYVALMELFHEITYITTCGQSLDQLNSN----RCVSEFTMENYKAI--VEN 265
BmFPPS1_NM_001043424 LKRHFNTKPYYNYVLETFNEMLMKCSMGHYVQKLMLK-TDKPDLSLFTMEKYEAI--TKY 273
DpFPPS_XM_019913385.1 IKRHLKDHPMYVPLMELFHEGNLKTTLGQSLDAMCLDTNGKPKLDMFTMSRYTSI--VKY 275
Tc_FPPS_NM_001170618.1 LKRHFSQHHCYVPTMELFHDVTLKTSMGQALDCLCNK-DGKPNLELFTMNKYNSI--VKY 271
PsFPPS1_KT959237 LKKHLSSLPCYVPIMELFRDITFKTSLGQSLDCLCLA-NGKPVLDLFTMKRYKTI--VKY 280
PsTPS1_KT959248 LTKRLAGHPCCLDILDLYAEEMFVMIIAQYMDIKKLD----------LKDFQKLVRHRF 223
Bg_FPPS_PYGN01000019.1 LRRYFRDKPYYVQTLELFHDVCLKTSMGQSLDLLSSPGG-KPNLDKFTMKRYESI--VKY 278
ApFPPS_NM_001126161.3 IKKYFKDKPYYTHILELFYDVTMKTSMGQCLDMLTANSFKSKKLEKYTMENYTAI--VKY 240
HhFPPS_MG870389 LQQHCKDKEFYVDLVELFLDATRRTTYGQTLDLVSSF----PNITHLTMDRYNFI--TKY 249
HhTPS1_MG917093 LKKFFRNDPGYIPLMETALMAYIKVAIGQTIDVLGQSN-----KT-PVMAEYKRI--NYY 215
HhTPS2_MG870388 LQRHFRNDPGFTKMCEMFSWVNGNSGIGQVLDILDHKN-----SDFSDYASWKNK--VEY 213
DmFPPS_NM_058032 KTAYYSFYLPFALALHLAGYKDAEAFRQSKTILLEMGNFFQVQDDFLDCFGNPEVTGKIG 325
BmFPPS1_NM_001043424 KTSYYTFQMPVSLALLMTGVDDPETHRQAKTILLKMGEFFQIQDDFLDCFGDPTVTEKYG 333
DpFPPS_XM_019913385.1 KTAFYSFQMPVAIAMYLAGMSDEEQHRQAKTILMEMGQFFQIQDDFLDCFGDPTVTGKVG 335
Tc_FPPS_NM_001170618.1 KTAYYTFQLPVALAMYMANLYDPEMHRQAKTILMEMGLFFQIQDDFLDCFGDPEVTGKKG 331
PsFPPS1_KT959237 KTSYYSIQLPVALGMYLANMTDQEQHRQAKTILLEMGEFFQIQDDFLDVFGDSDVTGKIG 340
PsTPS1_KT959248 DKALYVFNGSARSGLYLANVRDRETHDCMKKFSVPMSRFFQVQNDFSGVFEEESKFQNSC 283
Bg_FPPS_PYGN01000019.1 KTAYYSFHLPVALAMYMSGYMDEEMHRQAKTILLEMGHFFQVQDDYLDCFGDPEVTGKIG 338
ApFPPS_NM_001126161.3 KTAYYSFFLPVCLAMRMTNINDPEIFRQAKTILLEMGHFFQVQDDFLDCYGDPDVMGKIG 300
HhFPPS_MG870389 KTSYYTFHLPVAIAMYMAGIYNTELHRQAKSVLLEMGHYFQVQDDYLDVFSDEEVSGKKG 309

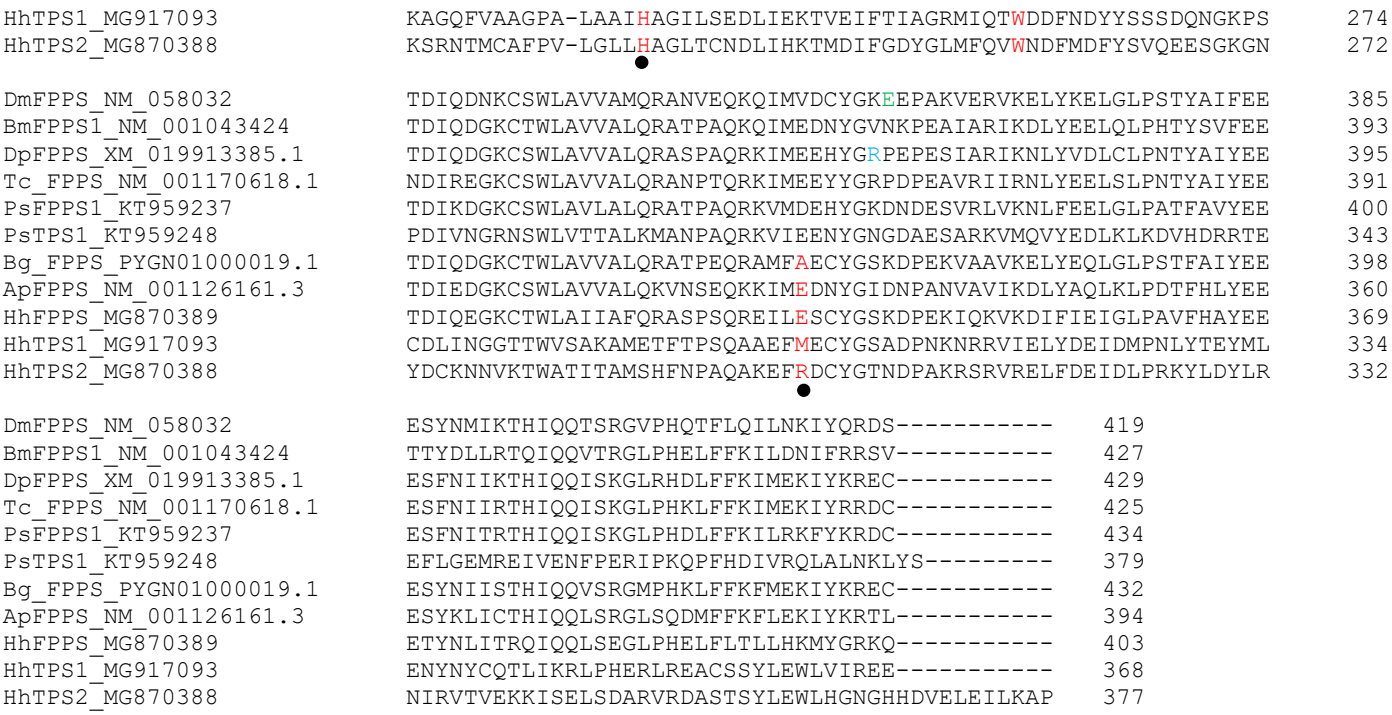


Fig. S8. Amino acid sequence alignment of *H. halys* and other insect FPPS and TPS enzymes with marked intron positions of the corresponding genes. Intron phases are highlighted in red (phase 0), green (phase 1), and blue (phase 2), respectively. *Drosophila melanogaster* (Dm), *Bombyx mori* (Bm), *Dendroctonus ponderosae* (Dp), and *Tribolium castaneum* (Tc), *Phyllotreta striolata* (Ps), *Blattella germanica* (BM), *Acyrthosiphon pisum* (Ap). Black arrowheads above the alignment indicate introns present in Hemiptera and other insect orders. Black dots below the alignment indicate intron positions conserved among the presented stink bug, aphid and cockroach sequences. The alignment was produced in Clustal Omega (Sievers and Higgins, 2018).

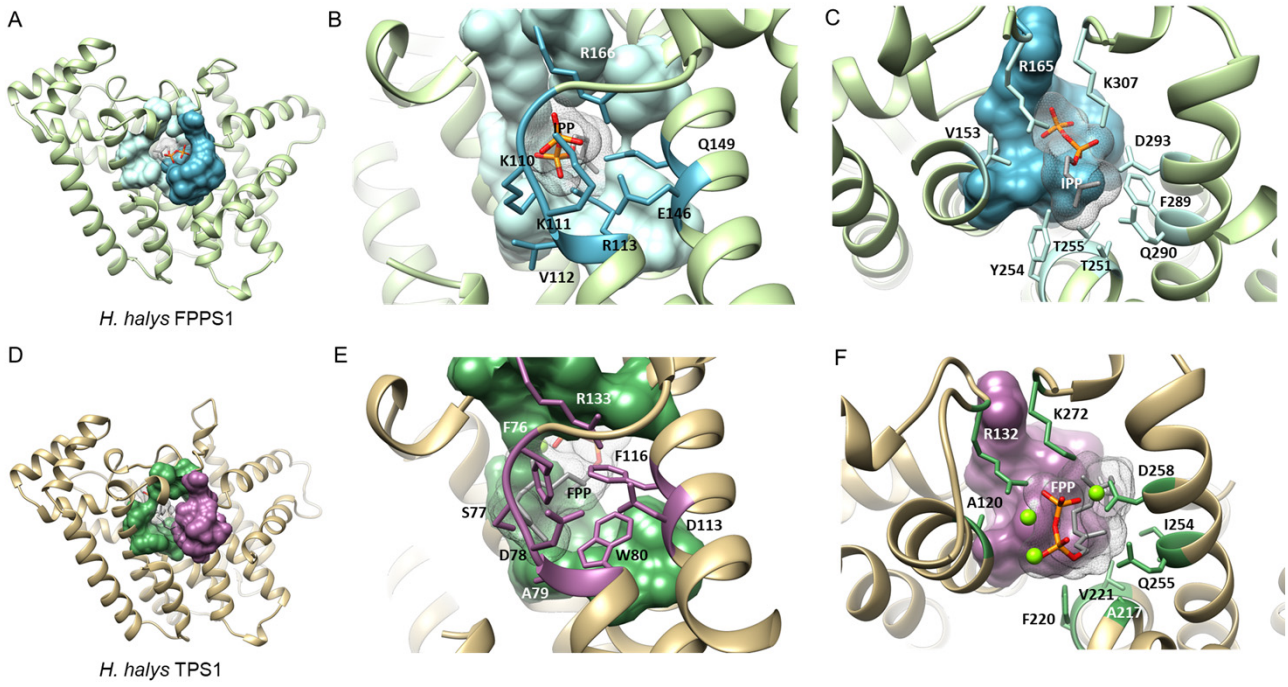


Fig. S9. Structural analysis of IPP binding residues in FPPS and IDS-type TPS homology models. (*A*) Structural model of *H. halys* FPPS with IPP binding pocket rendered as a colored surface; (*B*) *Hh*FPPS1 residues binding the diphosphate moiety of IPP are marked in dark blue and labeled; (C) *Hh*FPPS1 residues binding the prenyl tail of IPP are marked in light blue and labeled; (*D*) structural model of *H. halys* TPS1 with IPP binding pocket rendered as a colored surface; (*E*) *Hh*TPS1 residue substitutions of the diphosphate binding residues in (*B*) are marked in magenta and labeled. Aromatic substitutions in this region favor interactions with the isoprenyl tail of the docked FPP substrate; (*F*) *Hh*TPS1 residue substitutions of the prenyl tail binding residues in (C) are marked in green and labeled. Substitutions alter the substrate binding region to accommodate a larger isoprenyl diphosphate substrate.

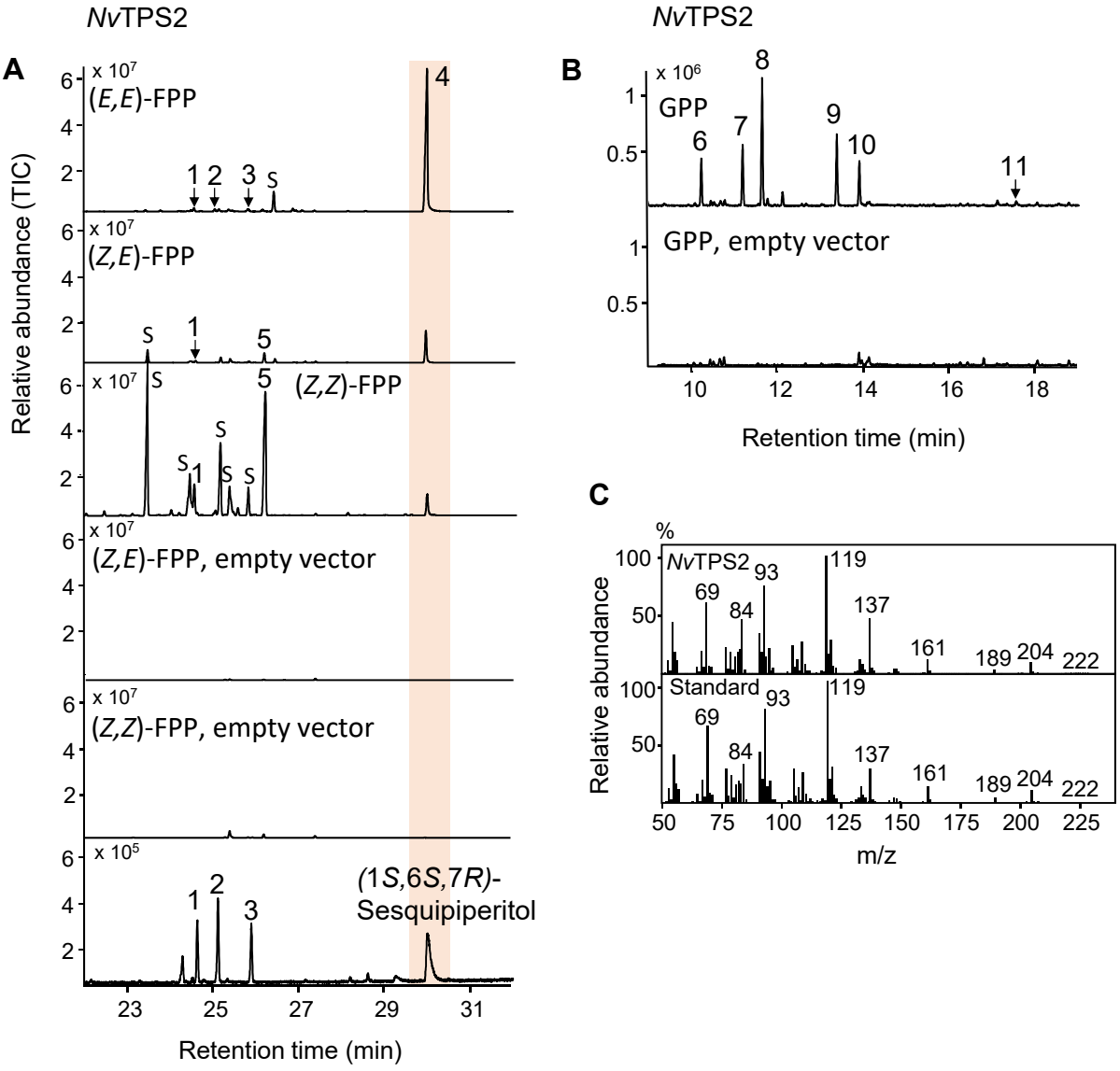


Fig. S10. Terpene synthase assay of *Nv*TPS2 with various FPP isomeric substrates and GPP. Partially purified enzyme was assayed with 50 µM substrates in assay buffer. (*A*) *Nv*TPS2 converts (*E*,*E*)-FPP to sesquipiperitol (4) along with thermal rearrangement products (1-3); chromatogram is the same as in Fig.4A. Formation of sesquipiperitol from (*Z*,*E*)-FPP and (*Z*,*Z*)-FPP is reduced relative to other sesquiterpene products. (*B*) *Nv*TPS2 converts GPP to monoterpene olefins and alcohols including a putative piperitol isomer (11). 1, γ-curcumene; 2, α-zingiberene; 3, β-sesquiphellandrene; 4, sesquipiperitol isomer; 5, (*Z*)-α-bisabolene; 6, myrcene; 7, α-terpinene; 8, limonene; 9, terpinolene; 10, linalool; 11, putative piperitol isomer. S, unidentified sesquiterpene olefins. (*C*) Mass spectra of enzyme product 4 and synthetic (1*S*,6*S*,7*R*)-sesquipiperitol.

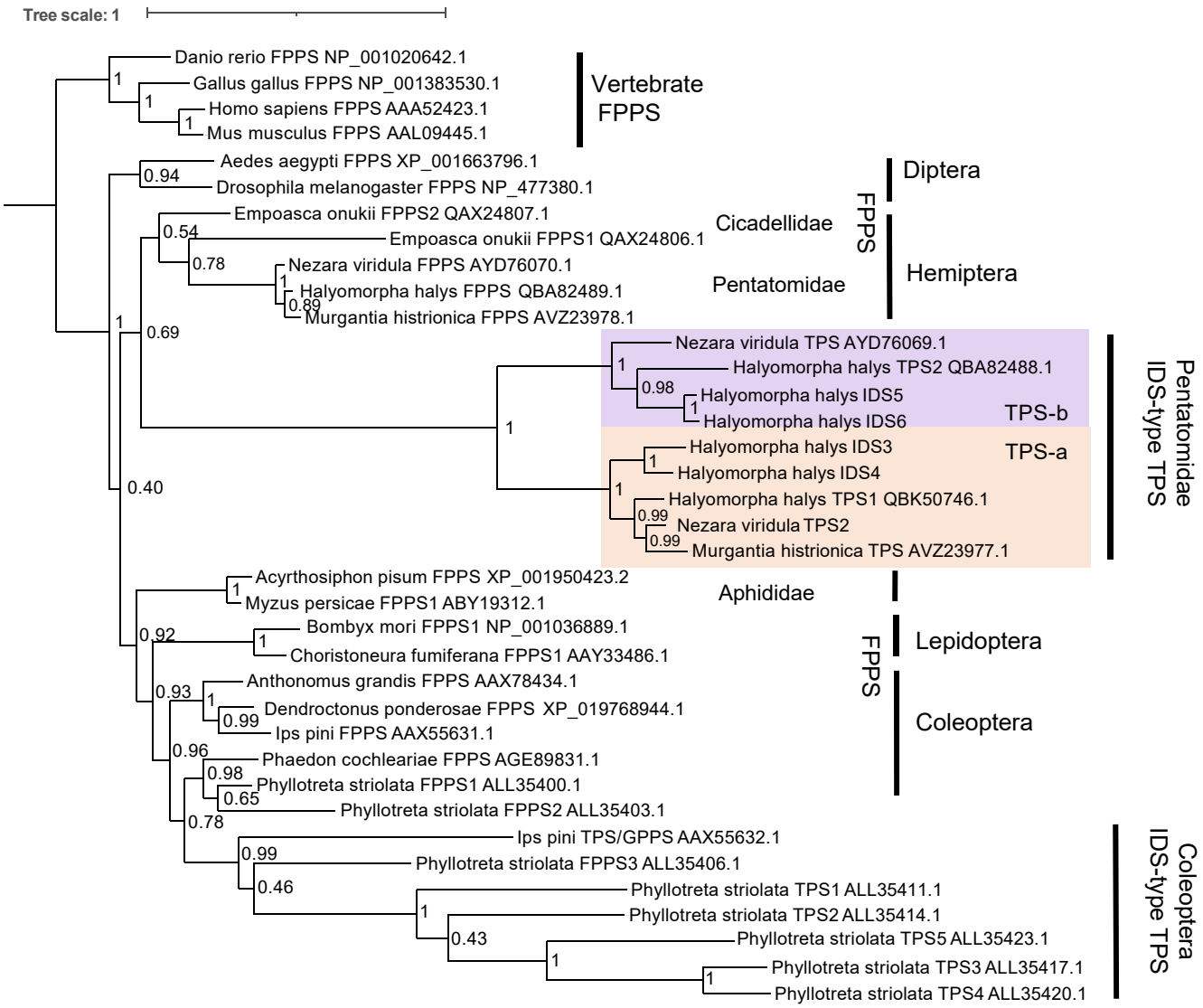


Fig. S11. Phylogram of characterized insect IDS (FPPS) and IDS-type proteins inferred from a Bayesian analysis (MRBAYES v3.2.5). Nodes are labeled with posterior probability support values and branch lengths are scaled to the number of amino substitutions per site. The phylogeny is rooted to a clade of vertebrate FPPS proteins. Pentatomid TPS-a and TPS-b clades are highlighted in orange and purple, respectively.

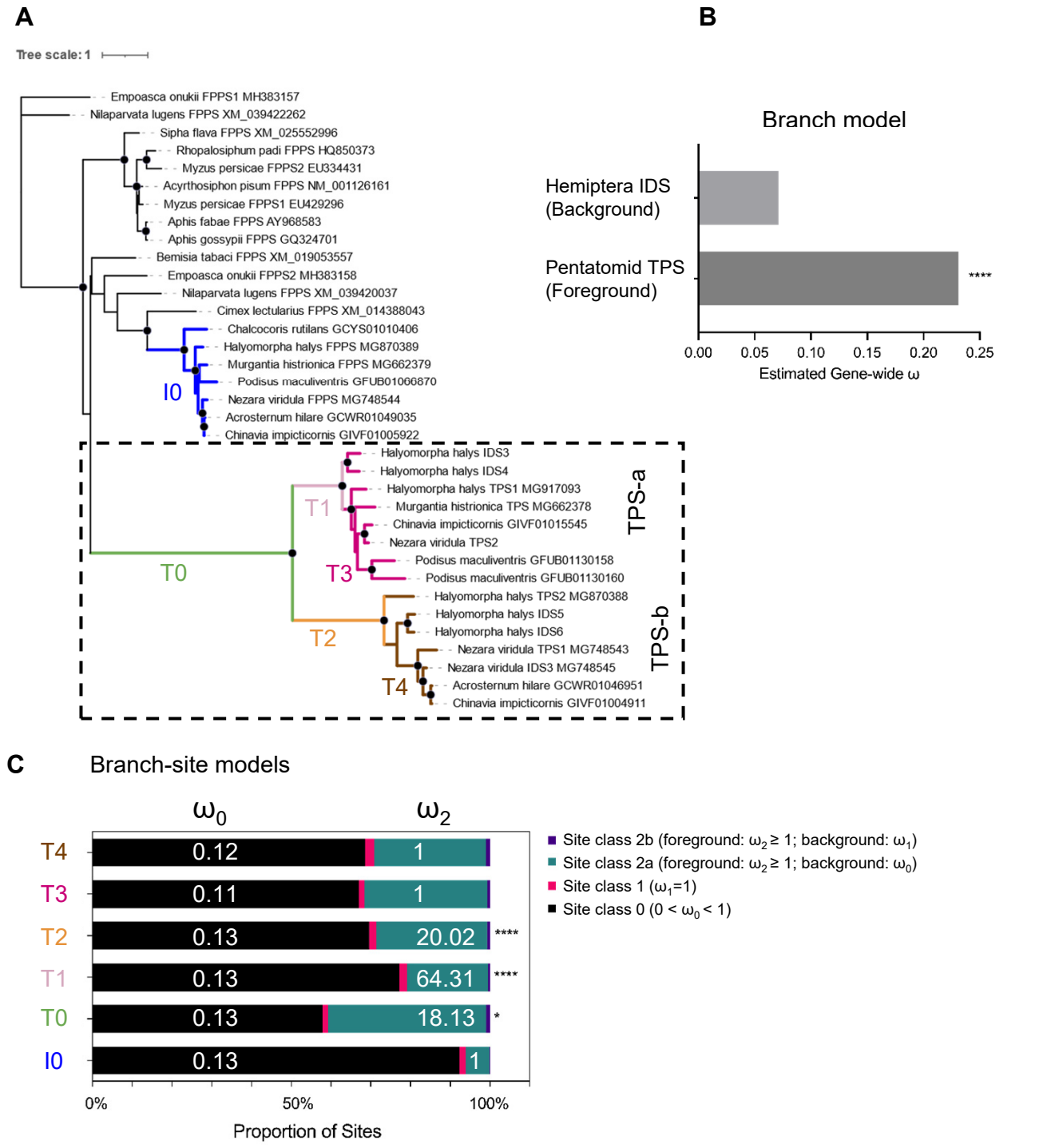


Fig. S12. Phylogeny and results from PAML selection analyses. (*A*) Phylogeny inferred from a codon alignment of characterized and putative hemipteran IDS (FPPS) and IDS-like/TPS proteins. Branches labeled as the TPS foreground in branch models are enclosed in a dashed box. Foreground branches used in branch-site models are color coded and labeled. Branch lengths are scaled proportionately to the number of amino acid substitutions per site. (*B*) Gene-wide ω estimates for IDS and TPS genes generated in a two-ratio branch model**.** Foreground and background ω values were significantly different based on a LRT comparing a two-ratio and one-ratio model. (*C*) Proportion of codons partitioned by site classes on select branches in branch-site models. Each bar represents the proportion of codons (colors) and corresponding estimated site class omega values (numbers) for codons on select branches. Estimated ω values for site class 0 and site class 2a/2b in branch-site models are displayed. Asterisks indicate significance in LRTs compared to the null models indicating positive selection. * p<0.05, **** p<0.0001.

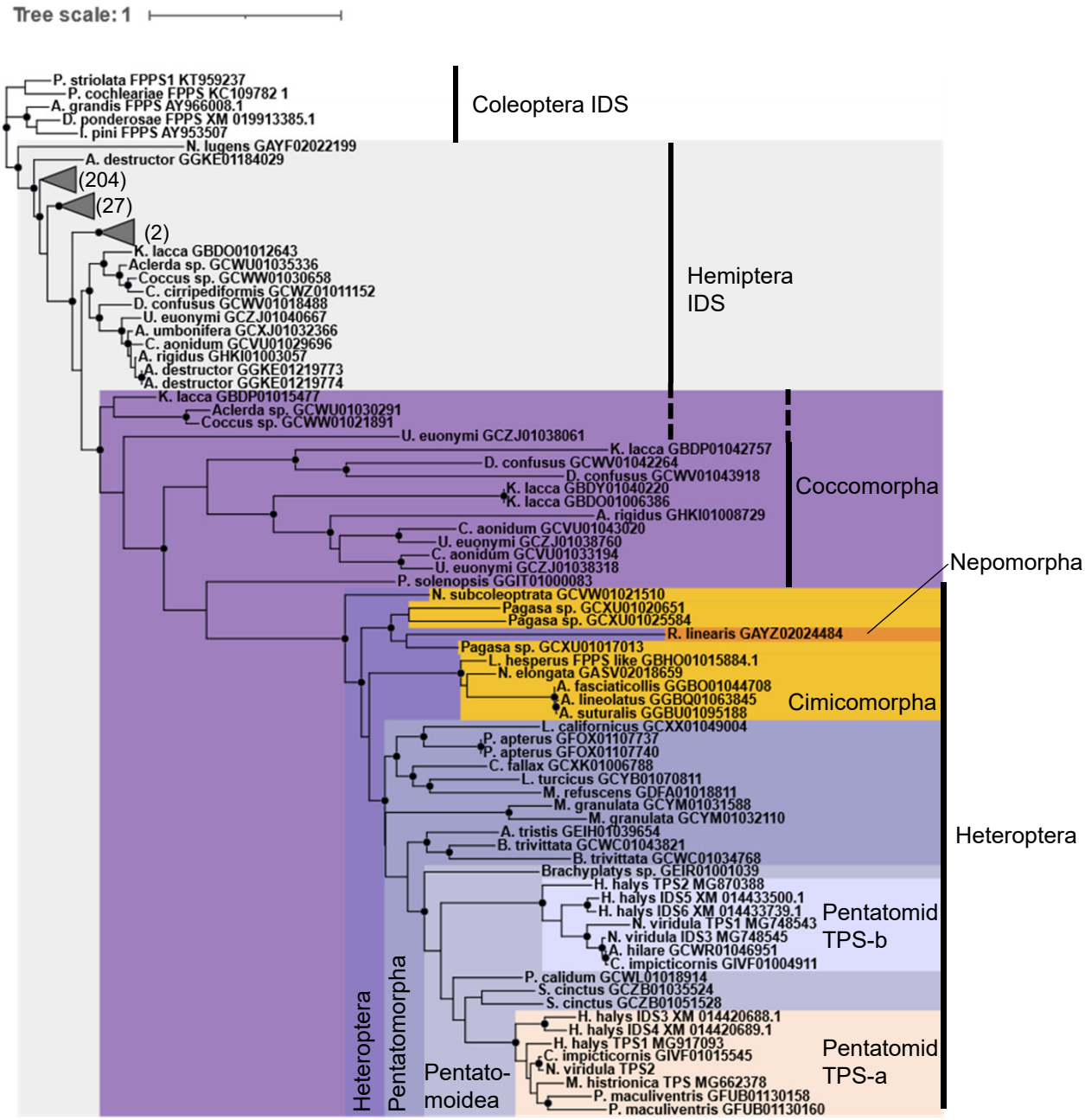


Fig. S13. Phylogeny of IDS (FPPS) and IDS-like proteins in the Hemiptera. A maximum likelihood consensus phylogram is depicted, rooted to a characterized Coleoptera FPPSs outgroup. Parenthetical values on collapsed clades represented as triangles indicate the number of collapsed leaves. Highly supported nodes (>80% or 0.80 support for all values) are labeled with a black circle. Branch lengths are scaled proportionally to the number of amino acid substitutions per site. Hemipteran sequences are listed in Table S5.