

# Structure and Mechanism of the Diterpene Cyclase *ent*- Copalyl Diphosphate Synthase

## --- SUPPLEMENTARY INFORMATION ---

Mustafa Köksal<sup>1</sup>, Huayou Hu<sup>2,4</sup>, Robert M. Coates<sup>2</sup>, Reuben J. Peters<sup>3</sup>, and David W.  
Christianson<sup>1\*</sup>

<sup>1</sup>Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania,  
231 South 34th Street, Philadelphia, PA 19104-6323 USA. <sup>2</sup>Department of Chemistry,  
University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. <sup>3</sup>Department of  
Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50011 USA.  
<sup>4</sup>Current address: Department of Chemistry, Huaiyin Normal University, Huaian, Jiangsu  
223300 P. R. China.

\*e-mail: [chris@sas.upenn.edu](mailto:chris@sas.upenn.edu).

## Supplementary Methods

**Isoprenoid diphosphate ligands.** The preparation and characterization of (*S*)-15-aza-14,15-dihydrogeranylgeranyl thiolodiphosphate (**1**) and 13-aza-13,14-dihydrocopalyl diphosphate (**2**) were described previously<sup>26,27</sup>.

**Cloning, expression and purification of copalyl diphosphate synthase.** The pseudomature CPS gene from *Arabidopsis thaliana* lacking the N-terminal segment M1-Q84 in pDEST14 vector (AtCPSd84) was described previously<sup>23</sup>. For more efficient purification, this gene was cloned into a T7 promoter-based expression vector with a C-terminal hexa-histidine tag. Accordingly, the M84-AtCPS gene was amplified by PCR with the following forward and reverse primers with flanking *Eag*I and *Bam*HI sites, respectively:

GCACGGCCGATGATTAGTGTGGAAGTAATAG and

GCAGGATCCGACTTTTTGAAACAAGACTTTG. A variant of the pET22b vector (Novagen, USA) (pET22bCV) was created by PCR with the following forward and reverse primers with complementary flanking restriction sites:

GCACGGCCGATGTATATCTCCTTCTTAAAGTTAAACAAA and

GCAGGATCCCACCACCACCACCACCTGAGATCCGGCT. The M84-AtCPS gene and the pET22bCV vector were ligated and the intervening *Eag*I site was removed using the Quickchange kit (Qiagen Inc., USA) to generate a plasmid encoding the M84-AtCPS polypeptide with the C-terminal hexa-histidine tag GSHHHHHH (M84-AtCPS-CHT). The resulting clones were confirmed by DNA sequencing.

The M84-AtCPS-CHT protein (henceforth designated "CPS") was expressed using *E. coli* OverExpress C41(DE3) cells (Lucigen, USA). CPS was expressed as previously described

for taxadiene synthase<sup>16</sup>. The cell pellet (19 g) was suspended in 30 mL of buffer E (50 mM K<sub>2</sub>HPO<sub>4</sub> (pH 7.5), 300 mM NaCl, 10% (v/v) glycerol, 3 mM β-mercaptoethanol) containing 1 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride, then incubated at 4°C for 2 h with shaking. Cells were disrupted by sonication on ice with a large probe at medium power, 10 × (30 s on + 30 s off). Cell debris was cleared by centrifugation twice at 30,000g for 1 h. The clear supernatant was applied to a pre-equilibrated Ni-NTA column (QIAGEN, USA) at a flow rate of 1 mL/min using an ÄKTAprime plus FPLC system (GE Healthcare Bio-Sciences AB, Sweden). The loaded column was washed 3 times with 3 column volumes of first buffer E, then buffer E plus 2 mM imidazole, then buffer E plus 4 mM imidazole. The CPS protein was eluted with a gradient of 8-500 mM imidazole in buffer E at a flow rate of 2.5 mL/min. Selected fractions were combined, concentrated to a volume of 5 mL, and purified by gel-filtration chromatography as previously described for taxadiene synthase<sup>16</sup>. CPS was concentrated to a maximum of 5 mg/mL and the sample purity was 99% based on SDS-PAGE analysis.

**Crystallization of CPS.** Screening for CPS crystallization conditions was performed by the batch crystallization method at the HTS facilities of the Hauptman-Woodward Institute<sup>28</sup> and optimized at the University of Pennsylvania by the sitting-drop vapor diffusion method at 15-21 °C. Briefly, a 1-μL drop of protein solution [5 mg/mL CPS, 25 mM MOPSO (pH 6.8), 10% glycerol, 300 mM NaCl, 1 mM DTT] was added to a 1-μL drop of precipitant solution [100 mM sodium citrate (pH 5.4), 30 % polyethylene glycol 400, 200 mM KH<sub>2</sub>PO<sub>4</sub>] and equilibrated against a 500-μL reservoir of precipitant solution. Rectangular prism-shaped plates appeared in 2-3 weeks and grew to maximal dimensions of 50 μm × 100 μm × 200 μm. These crystals diffracted to only 4 Å resolution. Addition of 2.5 mM MgCl<sub>2</sub> and 2.5 mM 13-aza-13,14-dihydrocopalyl diphosphate (**2**)<sup>27</sup>, or 2.5 mM MnCl<sub>2</sub> and 2.5 mM (*S*)-15-aza-14,15-

dihydrogeranylgeranyl thiolodiphosphate (**1**)<sup>26</sup>, to the protein solution 4-8 hrs prior to crystallization improved the diffraction quality. Additionally, 0.1 M strontium chloride hexahydrate and 40% v/v 1,4-butanediol from the Additive Screen (Hampton Research, USA) were used as described in the supplier's guidelines for co-crystallization of the CPS-1 and CPS-2 complexes, respectively. Crystals were flash-cooled after transfer to a cryoprotectant solution consisting of the mother liquor augmented with 10% glycerol.

**X-ray diffraction data collection and processing.** Crystals of the CPS-1 complex and the CPS-2 complex diffracted X-rays to 2.25 Å and 2.75 Å resolution, respectively, at the Advanced Photon Source (APS), Argonne National Laboratory, beamline 24-ID-C, using incident radiation with  $\lambda = 0.9795$  Å. All diffraction data were processed with HKL2000<sup>29</sup>. Crystals of the CPS-1 complex belonged to space group  $P2_12_12_1$  with unit cell parameters  $a = 51.32$  Å,  $b = 114.31$  Å,  $c = 129.41$  Å; with one molecule in the asymmetric unit, the Matthews coefficient  $V_M = 2.32$  Å<sup>3</sup>/Dalton (solvent content = 47%). Crystals of the CPS-2 complex belonged to space group  $P2_12_12_1$  with unit cell parameters  $a = 63.74$  Å,  $b = 188.55$  Å,  $c = 229.07$  Å; with three molecules in the asymmetric unit, the Matthews coefficient  $V_M = 2.80$  Å<sup>3</sup>/Dalton (solvent content = 56%). Data collection and reduction statistics are recorded in **Supplementary Table 1**.

**Phasing and structure refinement.** The recently-described crystal structure of taxadiene synthase from *Taxus brevifolia*<sup>16</sup> (PDB ID: 3P5P) allowed solution of the phase problem by the molecular replacement method. Molecular replacement, rigid body refinement, positional refinement, grouped and individual atomic B-factor refinement, automatic addition of waters (using  $|F_o| - |F_c|$  map at  $3.0\sigma$  with 3.5 Å distance cut-off) was performed with PHENIX<sup>30</sup>. Manual model rebuilding was performed with COOT<sup>31</sup>. In the final model of the CPS-1 complex, 688 of

727 residues are present; disordered segments excluded from the final model include N-terminal residues M84-N90 (M84 is the N-terminus of the construct), the C-terminal hexa-histidine tag and one of its associated linker residues (S804-H810), and surface loops A618-R636 and Y732-R737. Some portions of **1** are modeled in two alternative conformations; each conformation is assigned 40% occupancy.

The polypeptide model of the CPS-**1** complex was used in molecular replacement calculations to determine the structure of the CPS-**2** complex at 2.75 Å resolution as described above. In the final model of the CPS-**2** complex, 685, 689 and 685 of 727 residues are present in monomers A, B, and C, respectively. Disordered segments excluded from the final models of monomers A, B, and C include N-terminal residues M84-S89, the C-terminal hexa-histidine tag and one of its associated linker residues (S804-H810), and surface loops R621-P637 and R730-K741 (R730-R737 in monomer B).

For both structures, refinement statistics are recorded in **Supplementary Table 1**. Ramachandran plot statistics, calculated with PROCHECK<sup>32</sup>, were as follows. CPS-**1** complex: allowed 91.9%; additionally allowed 7.8%; generously allowed 0.3%. CPS-**2** complex: allowed 88.8%; additionally allowed 10.6%; generously allowed 0.6%. Simulated-annealing omit maps were calculated with CNS<sup>33</sup>. Protein structure figures were prepared with the graphics program PyMol (<http://www.pymol.org>).

**Sequence alignments.** The sequence alignments discussed in the text were performed with ClustalX using default parameters<sup>34</sup>. Annotated alignments and descriptions of aligned sequences are included as **Supplementary Figures 4 and 5**.

## Supplementary Results

**Supplementary Table 1.** Data collection and refinement statistics

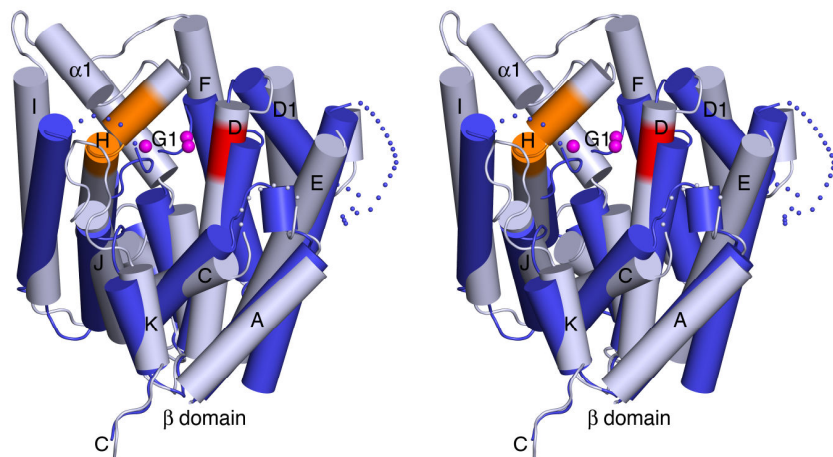
	CPS-1 complex	CPS-2 complex
<b>Data collection</b>		
Space group	<i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i>	<i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i>
Cell dimensions		
<i>a, b, c</i> (Å)	51.32, 114.31, 129.41	63.74, 188.55, 229.07
$\alpha, \beta, \gamma$ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	40.36-2.25 (2.33-2.25)*	44.52-2.75 (2.85-2.75)*
$R_{\text{sym}}$	0.120 (0.780)	0.139 (0.768)
$I / \sigma I$	11.0 (2.0)	12.0 (2.2)
Completeness (%)	99.4 (99.1)	94.8 (86.1)
Redundancy	5.1 (4.5)	4.8 (4.5)
<b>Refinement</b>		
Resolution (Å)	40.36-2.25	44.52-2.75
No. reflections	36117	62150
$R_{\text{work}} / R_{\text{free}}$	0.172 / 0.233	0.177 / 0.256
No. atoms		
Protein	5660	16889
Ligand/ion	67	173
Water	263	201
B-factors		
Protein	46	60
Ligand/ion	50	73
Water	48	50
R.m.s. deviations		
Bond lengths (Å)	0.005	0.004
Bond angles (°)	0.9	0.8

\*Both data sets were collected from single crystals. Highest-resolution shell is shown in parentheses.

**Supplementary Table 2.** Domain structural alignments among class I and II cyclases\*

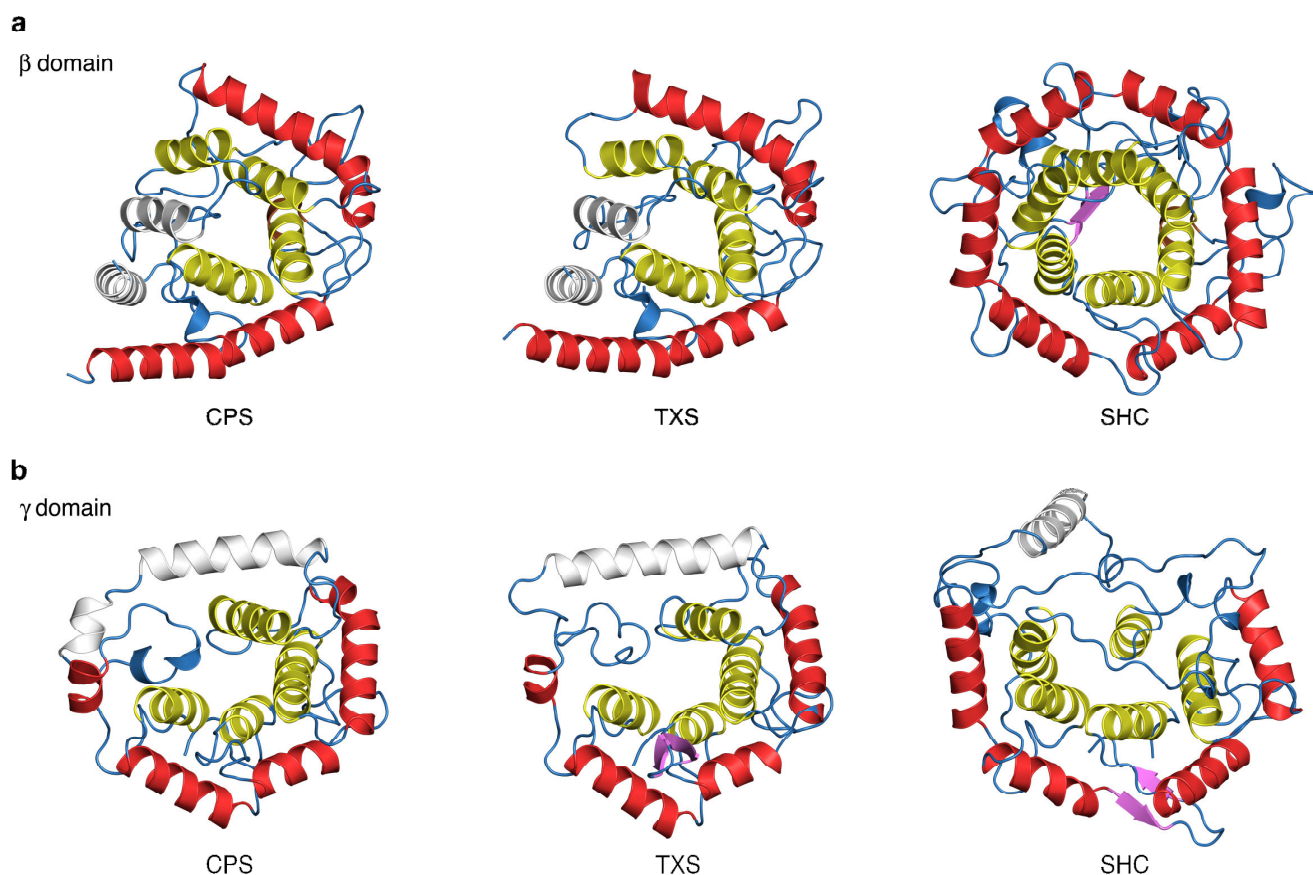
CPS domain	Sequence identity (%)			No. of C <sub>α</sub> atoms aligned			r. m. s. deviation (Å)		
	TXS	SHC	OSC	TXS	SHC	OSC	TXS	SHC	OSC
α (269)	15	-	-	207 (309)	-	-	2.5	-	-
β (232)	45	10	9	219 (229)	178 (348)	180 (442)	1.1	3.0	2.7
γ (211)	42	16	12	196 (213)	139 (271)	140 (284)	1.5	3.3	3.3

\*Number of residues in aligned domains is given in parentheses. TXS = taxadiene synthase, PDB 3P5R; SHC = squalene-hopene cyclase, PDB 1SQC; OSC = oxidosqualene cyclase, PDB 1WK6.

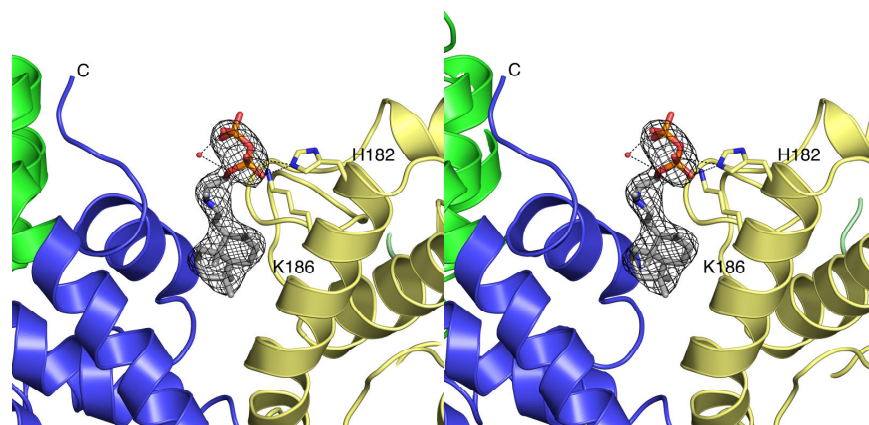


**Supplementary Figure 1. Comparison of the  $\alpha$  domain of class I and class II diterpene cyclases.** Vestigial  $\alpha$  domain of CPS (dark blue) and the functional  $\alpha$  domain of taxadiene synthase (light blue) are compared. The  $D^{613}$ DMAD and  $N^{757}$ DTKTYQAE metal binding motifs of taxadiene synthase are red and orange, respectively, and  $Mg^{2+}$  ions are magenta spheres.





**Supplementary Figure 2. Comparison of the  $\beta$  and  $\gamma$  domains of class I and class II diterpene cyclases and a class II triterpene cyclase.** The functional  $\beta\gamma$  domains of CPS, the vestigial  $\beta\gamma$  domains of taxadiene synthase (TXS), and the functional  $\beta\gamma$  domains of squalene-hopene cyclase (SHC) are compared. To highlight the double  $\alpha$ -barrel fold, which is complete in the  $\beta$  domain of squalene-hopene cyclase as  $(\alpha\alpha)_6$  and incomplete in all others, outer  $\alpha$  helices are red and inner  $\alpha$  helices are yellow. Additional helices (such as the membrane anchoring helix 8 of squalene-hopene cyclase) are white,  $\beta$  strands are violet, and loops are blue. The color scheme is identical to that utilized in a previous structural comparison<sup>4</sup>.



**Supplementary Figure 3. Binding of ACP at a crystal packing interface.** Simulated annealing  $|F_o|-|F_c|$  omit map in which 13-aza-13,14-dihydrocopalyl diphosphate (**2**) is omitted from the structure factor calculation (contoured at  $3.0\sigma$ ); the side chains of the residues surrounding the diphosphate group are indicated. While  $\alpha$ ,  $\beta$ , and  $\gamma$  domains are blue, green, and yellow, respectively, the  $\gamma$  domain (light yellow) is from a different monomer. The diphosphate group of **2** accepts hydrogen bonds from H182, K186, and a water molecule (red sphere).

*Ci/TbTXS/1-862* 1 - MAQLSFNAALKMNALGNKAIHDP...TNCRAKSERQMMVWCSSRSRGRTRVKMSRSGSGGPGVVMMS... 64  
*Ci/HsKS/1-774* 1 .....MNITITQSMT.....GPLSITKPYRSWALSAIHATP..... 31  
*Ci/OsES/1-812* 1 .....MMLLLPSSSSCCRCPCGGQFHGAPPR...VMAPRRGVTR... 38  
*Ci/OsES/1-820* 1 .....MMLSSSYS...GGQFPVPLGTRPKRSTTVVPRPVVTR... 37  
*Ci/OsOS/1-816* 1 .....MMLSSSYS...GGQFPVPLGTRPKRSTTVVPLPVVTRA... 38  
*Ci/SiKS/1-784* 1 .....MNLSLCIA.....SPL.LTKSSRPTALSAIHATST... 29  
*Ci/AgABS/1-853* 1 .....AHLHTAN.....TQSIPIHFSTTLNAGSSASKRRSLYLRWGKGSNKI.IACVGEGGATSV... 53  
*Ci/MgABS/1-873* 1 MAGVLFANLPQSLQLSPKVPFR...QSTNILIPHFHRSFGFNAQHCVRSHLRLRWKCVGIHASAET... 65  
*Ci/MPLAS/1-850* 1 .....TOCIPHFHGLNAGTSAGKRRSLYLRWGKGSNKI.IACVAGQ... 57  
*Ci/MPLAS/1-859* 1 .....MALLSSLSQIPTGAHLLTNAYANTQOCIPHFHSTLNAGTSAGKRRSLYLRWGKGSNKI.IACVGE... 66  
*Ci/MFDTS/1-850* 1 .....TOCIPHFHGLNAGTSAGKRRSLYLRWGKGSNKI.IACVAGQ... 57  
*Ci/MFAPS/1-867* 1 .....MALLSSLSQIPTGSHPLTH.....TOCIPHFSTTLNAGISAGKPRSFYLRWGKGSNKI.IACVGE.G.TTSL... 66  
*Ci/AtCPPS/1-802* 1 .....MSLOYHVLNIPSTTF.....LSSTKTTISSSFLTISGSPLNVARDKSR... 44  
*Ci/MpCPPS/1-801* 1 .....MFTHTFSTHFLP.....SSSSLLFLHFPFYKS.SSLGAVSFVAKDK... E 40  
*Ci/CmCPPS/1-823* 1 .....MKALSLSRFPF.....CSSDATKLSSRPPPPVGVSCSFKVES...IRSSR...I 44  
*Ci/CmCPPS/2/1-827* 1 .....MSSSSLSLSRHCL.....SSSF.SFLPNLFPFPAAGG.CSLRVKDKG...AVLSEIRCI 52  
*Ci/LcCPPS/1-800* 1 .....MSISASFL.....RFLSHTAHYQSPSSSPNQPFKFLKSNR...EHVEFNRI 44  
*Ci/SrCPPS/1-787* 1 .....MKTGFI.SPATVHFHRI...SPATTFRHHS.PATTNSTG.IVALRDIN... 43  
*Ci/SmCPPS/1-793* 1 .....MASLSTIL.....SRSPAARRNITPASAKLHRPECFATSAWM...G... 39  
*Ci/ZnCPPS/1/1-823* 1 .....MPYPHPYVW...SRRRRRRRRGGDARGPQARR...VVER...AAA 39  
*Ci/ZnCPPS/2/1-827* 1 .....MVLSSSTVVP...HLSSLLAVVQLGKPSRSRIKKTDTVAVP...AAA 41  
*Ci/OsCPPS/1/1-867* 1 .....MIHLHSPPTAPAAFGGAGSADWRRRRRWSWSSSRAPVAKGHLRPCVWRRGGDDGGEDHHADGGGGGGGAA 74  
*Ci/OsCPPS/2/1-800* 1 .....MQMQVLTAA...SLPRATLLPAAAEPWRS...FLLQLQ 34  
*Ci/OsCPPS/4/1-767* 1 .....MPVFTASFQC...VTLFGQPSADAADQLLQGGQP...FLHLH 37  
*Ci/AaSHC/1-631* 1 .....MTGTC.LRRRGGPKYTE.PATDLGRWRLNCE... 30  
*Ci/HsOSC/1-732*



*Ci/TbTXS/1-862* 65 TGTSKVSVETSSTIVDDIPRLSANYHGD.LWHNNVIQTLETPFRSSTY...QERADELVVKIKDMFNALGD...GDISPSAYDTAWVAVRATISSDGS 158  
*Ci/HsKS/1-774* 32 .....SHVG...QANPHTLAIDTTKERIR...KLFNN.VDLSVPSAYDTAWVAVRATISSDGS 158  
*Ci/OsES/1-812* 39 .....VYIEKRLOGVGNASSLRDHRKELQARTR...DQLQT.LELSTLYDTAWVAVRATISSDGS 95  
*Ci/OsES/1-820* 38 .....AGVRRNNLEVVG-NAGTLQMGIDELRVLIVR...KQLQG.VELSPSAYDTAWVAVRATISSDGS 94  
*Ci/OsOS/1-816* 39 .....TAGVRRNNLEVVG-NAGTLQMGIDELRVLIVR...KQLQG.VELSPSAYDTAWVAVRATISSDGS 94  
*Ci/SiKS/1-784* 30 .....SHGG...QNTPTNLIIDTTKERIQ...KLFKN.VELSPSAYDTAWVAVRATISSDGS 97  
*Ci/AgABS/1-853* 54 PYQSAEKNDLSLSS...TLVKREFPPGFWKDDLDLSLSSHVAASD...EKRIETLISEKNMFRSMGY.GETNP.SAYDTAWVAVRATISSDGS 139  
*Ci/MgABS/1-873* 66 .....RPDQLPQEERFVSRNLADYHVAWKKDFIDSLTSPNSHATSKSSVDEITNKRIQTLVKEIQCMFQSMGD.GETNP.SAYDTAWVAVRATISSDGS 156  
*Ci/MPLAS/1-850* 58 .....DPFSVP...TLVKREFPPGFWKDHWIE.SLMPYKVPASD...EKRIETLISEKNMFRSMGY.GETNP.SAYDTAWVAVRATISSDGS 135  
*Ci/MPLAS/1-859* 67 .....DLSLAP...TLVKREFPPGFWKDHWIE.SLSSHVAASD...EKRIETLISEKNMFRSMGY.GETNP.SAYDTAWVAVRATISSDGS 144  
*Ci/MFDTS/1-850* 58 .....DPFSVP...TLVKREFPPGFWKDHWIE.SLMPYKVPASD...EKRIETLISEKNMFRSMGY.GETNP.SAYDTAWVAVRATISSDGS 135  
*Ci/MFAPS/1-867* 67 PYQSAEKNDLSLAP...TLVKREFPPGFWKDHWIE.SLSSHVAASAE...EKRMETLISEKNMFRSMGY.GETNP.SAYDTAWVAVRATISSDGS 152  
*Ci/AtCPPS/1-802* 45 SGIHCSKLRTEQYINSQEVQHDLP.LIHEWQQLQDDEAPQIS...VGSNSN...AFKEAVKSVKTLRNLTD.GEITI.SAYDTAWVAVRATISSDGS 132  
*Ci/MpCPPS/1-801* 41 KRCRAIKSRTQEQYGVF...QTN.VATLKLSEINVEDVIVIDDEEQDIR...VGLVKNKIKSLLSLED.GEITI.SAYDTAWVAVRATISSDGS 125  
*Ci/CmCPPS/1-823* 46 GRRRALARAQHTSESAIVAKSSLTPIVRTDAESRRTPWTD...N.EHDIDAIRSMLGSMGD.GEISV.SAYDTAWVAVRATISSDGS 141  
*Ci/CmCPPS/2/1-827* 53 IKCNAIKSPPQDYSV...QSG.VPLKQQWF.VEEGIESETAAQVSWVE...EIEERVKIKSMLSSMD.DGIS.SAYDTAWVAVRATISSDGS 131  
*Ci/LcCPPS/1-800* 44 FRCKAVS...KEYSLL...QKDEASFTKDDDDHDDHDT...KNLYPN...EIEKVESKAMFGSMND.GEINVS.SAYDTAWVAVRATISSDGS 126  
*Ci/SrCPPS/1-787* 40 SSSKNLLSYQLNHKIKS...VATVDAQVHDHGTTVHGHDAVK...N.IEDPIEYIRTLRTTGD.GRISV.PYDTAWVAVRATISSDGS 122  
*Ci/ZnCPPS/1/1-823* 40 GPGHATTTQQPDNVSSAKVFQTSRVETES...KLRNGKRPQDDDEHQAEEA...ELQPLIDQVRAMLRSMND.GDTS.SAYDTAWVAVRATISSDGS 130  
*Ci/ZnCPPS/2/1-827* 42 GRWRRALARAQHTSESAIVAKSSLTPIVRTDAESRRTPWTD...DAEPLVDEIRAMLTSMGD.DGIS.SAYDTAWVAVRATISSDGS 128  
*Ci/OsCPPS/1/1-867* 75 WRARATTAAGVSSSSSTAKGLQANIIEHETPRIKWPNEKRDLDHQQNEADEADDELQPLVEQVRSMLSSMED.GAIT.SAYDTAWVAVRATISSDGS 173  
*Ci/OsCPPS/2/1-800* 35 ARP IQRPEIMLHCKAQLQ...GQETREERLQDDEHARPPQGGDDVAASTS...ELPYMIESIKSKLRAARNLQDGETT.SAYDTAWVAVRATISSDGS 127  
*Ci/OsCPPS/4/1-767* 38 ARRRRRCPPMLIKSPPYPASEETREWEAEGQHEHT...DELRETT...TTMIDGIRTAALRIGEGEIS.SAYDTAWVAVRATISSDGS 119  
*Ci/AaSHC/1-631* 1 .....MAEQLVEAPAYARTLDRAVEYLLSCOKD...EGYWGPLLS...NVTMAEYVLLCHLDRVD... 57  
*Ci/HsOSC/1-732* 31 .....RGRQWTWYLDQDERAGREQTGLEAYALGLDTKNYFKDLPKAHTAFEG...ALNGMTFYVGLQAEDGHWITGYDGPLLPLGLLITCHVARIP 119



*Ci/TbTXS/1-862* 157 EKRFPPQALNIVFNNQLQDQSWGIESHFS...LQDRLLNTNSVIALSVKTKG...HSQVQGAEIFIAENLRLLNEDE-LSPDQII 237  
*Ci/HsKS/1-774* 80 KSPCFPECLNINLVDNQLDQSWGLVNHSS...THPLIKDLSLTLACIIVALKRWNVG...EDQINKGLRIFIESNLASATDKSQP-SFFGDII 165  
*Ci/OsES/1-812* 96 QHPCFPQCVWEI.LQNQLDQDQSWG-TRGEG...VAVTRDVLSTLACVLAALKRWNVG...QEHIRRGDIFGRNFSIAMDEQIA-APVGFNIT 179  
*Ci/OsES/1-820* 95 QSPCFPQCVWEI.LQNQLDQDQSWGHSAGPS...GEVNDKILLSTLACVLAALNIVNVG...QDHIRRGDIFGRNFSIAMDQCA-APVGFNIT 179  
*Ci/OsOS/1-816* 97 QSPCFPQCVWEI.LQNQLDQDQSWGHSAGPS...GEVNDKILLSTLACVLAALNIVNVG...QDHIRRGDIFGRNFSIAMDQCA-APVGFNIT 181  
*Ci/SiKS/1-784* 78 KSPCFPECLNINLINNQLNDQSWGLVNHHTNHNHPLKDLSSLTLACIIVALKRWNVG...EDQINKGLRIFIESNLASATDKSQP-SPIGDI 165  
*Ci/AgABS/1-853* 140 DNFHPFETVEI.LQNQLDQDQSWGEGFYFL...AYDRILATLACIITLTLWRTG...ETQVQKIEFRTQAGKMEDEADSHRSPGFEIV 222  
*Ci/MgABS/1-873* 157 GAPQFPQTLQWLINNQLPDQSWGEECIFL...AYDRVNLTLACLLTLKIWNKG...DIQVQKIEFVRKHMEEMKDEADNHRSPGFEIV 239  
*Ci/MPLAS/1-850* 136 EKQPFPETLEWI.LQNQLDQDQSWGEEFYFL...AYDRILATLACIITLTLWRTG...DQVQKIEFRTQAGKIEEADSHRSPGFEIV 218  
*Ci/MPLAS/1-859* 145 EOPFETLEWI.LQNQLDQDQSWGEGFYFL...AYDRILATLACIITLTLWRTG...EIQVQKIEFRTQAGKIEEADSHRSPGFEIV 227  
*Ci/MFDTS/1-850* 136 EKQPFPETLEWI.LQNQLDQDQSWGEEFYFL...AYDRILATLACIITLTLWRTG...DQVQKIEFRTQAGKIEEADSHRSPGFEIV 218  
*Ci/MFAPS/1-867* 153 EHPEFETLEWI.LQNQLDQDQSWGEGFYFL...AYDRILATLACIITLTLWRTG...ETQIRKIEFRTQAGKIEEADSHRSPGFEIV 235  
*Ci/AtCPPS/1-802* 133...PAPSSAVKVI.AENQLSDQSWGDALYLF...YHDLRINTLACVIALRSWNLF...PHQCNKIFTFRENLSKLENEDEHMPIGFEIV 213  
*Ci/MpCPPS/1-801* 126 STQFPSSLEWI.AKNNQLPDQSWGDSRLFS...AHDRIINTLACVIALRSWNLF...SEKCDKGMIFRENLSKLENEDEHMPIGFEIV 208  
*Ci/CmCPPS/1-823* 133 KTLPLFPSSLEWI.AQNNQLPDQSWGDSGIFS...AHDRIINTLACVIALRSWNLF...PKSEKGMVFNKNNIKLENEDEHMPIGFEIV 215  
*Ci/CmCPPS/2/1-827* 142 KSPFPSSLEWI.ANNQLPDQSWGDAQIFS...PHDRIINTLACVIALRSWNLF...PQNTKGVAFNQNWKLENEDEHMPIGFEIV 224  
*Ci/LcCPPS/1-800* 132 ETPFPSSLEWI.ANNQLPDQSWGDSIFL...VYDRVINTLACVIALRSWNLF...PDKILLGSMFRENLSKLENEDEHMPIGFEIV 214  
*Ci/SrCPPS/1-787* 127 GSPFPSSLEWI.ANNQLPDQSWGDHLLFS...AHDRIINTLACVIALRSWNLF...PSKCEKLNLENEDEHMPIGFEIV 209  
*Ci/SmCPPS/1-793* 123 DQFPFPSSLEWI.VQNNQLPDQSWGDKLFC...VYDRVINTLACVIALRSWNLF...AHKVKRQVYIKENVDKLENEDEHMPIGFEIV 205  
*Ci/ZnCPPS/1/1-823* 131 AQPFPATVRIVDHDLPDQSWGDSALFS...AYDRMINTLACVIALRSWNLF...PARCEAGLSLHENMWRLAEEAESMPIGFEIV 213  
*Ci/ZnCPPS/2/1-827* 129 PQ...FPAAVRIRNNQLPDQSWGDAALFS...AYDRILINTLACVIALRSWNLF...PEMRGRSLFLGRNNWKLATDEEESMPIGFEIV 209  
*Ci/OsCPPS/1/1-867* 174 TQ...FPAAVRIVGSQLDQSWGDEALFS...AYDRVINTLACVIALRSWNLF...HDQCKGLOLNLNLWRLAEEEDPTMRIGFEIV 254  
*Ci/OsCPPS/2/1-800* 128 RSPFPPEADWI.ANNQLPDQSWGDAGMFI...VQDRINTLACVIALRSWNLF...EEQRARGLAYIQDNLWRLAEEEDPTMRIGFEIV 201  
*Ci/OsCPPS/4/1-767* 120 -GQFPSTIDWI.VQNNQLPDQSWGDASFEM...MGRIMSTLACVIALRSWNLF...TDKCEKLNLENEDEHMPIGFEIV 210  
*Ci/AaSHC/1-631* 58 -RDRMEKIRRYLLHEQREDGTWALYPGGP...PDLDTIEAYVALKYIGMS...RDEEPMQALRFIQSOGGLESSRVEITRM 132  
*Ci/HsOSC/1-732* 120 PAGYREEIVRYLRSVLDQDQSWGLHIEDK...STVFGALNYSVLRILGVG...PDDPDLVRARNLHKKGGAVAIKSWGK 195



Supplementary Figure 4





Cj/TbTXS/1-862 497 VVEY PWHMSI PRLEARSY IDS YD -DNYVWQR--- KTL RMPSLNSKCLAEKLDNFNIVSLHQEELKLLTRWVWESGMADIN - FTRHRAVEVYFSSAT - 590  
Cj/HsKS/1-774 410 ALKVFNTGIERINVKRYIDHYHVDHTRIL--- KTT RSSNISNEDYLKFAVEDFNSSQSIYLEEMKGFKRWLAEATLQDLT - FAREKSAYCYFTAAT 504  
Cj/OsKS/1-812 435 ALKFFRYTTLRDLKHKRIERFD-AKDSQML--- KTE LLL-HANODILALAVEDFSSQSIYQDELNYLECVWKDEKLDLP - FAROKLTYCLLSAAAT 528  
Cj/OsSES/1-820 434 ALKRYFYATVEVLEHKGNIERFN-VNGFQRL--- KSG CGS-GADKEILALAVNKHYAQSIVYQELRYLESWVAEFLDELK - FARVILPQLSLLSAVVP 527  
Cj/OsSOS/1-816 436 ALKRYPLSTVDALHRRGNIERFN-VNGFQRP--- KSG CGS-GADKEILALAVDKFHYNQSVYQELRYLESWVAEFLDELK - FARVILPQLSLLSALVP 529  
Cj/SKs/1-784 410 ALKFIINTGLERINTRRNIQLYV-DNTRIL--- KTT HSSNISNTYYLRLAIVEDFYTCQSIYRELKGLERWVWQNLQDLK - FAROKTAYCYFSSAAT 504  
Cj/AgABS/1-853 488 ALKYPWHKSMRPRLEARSYIENYGV-PNDVWL--- KTY MMPYISNEKYLELAKLDFNKLQSIHQTELQDLRRWVKSSGFTELN - FTRRRTVEIYFSPASF 582  
Cj/MGABS/1-873 506 AIKYPWHRSMRPRLEARSYIEQFG-SNDVWL--- KTY KMLYSNEKYLELAKLDFNMVQALHQKTEQHVSVSWRRESGFDLNT - FTRORPVEMVYFSSAVP 600  
Cj/PALAS/1-850 484 AIKYPWHRSMRPRLEARSYIENYGV-PNDVWL--- KTY MMPNISNEKYLELAKLDFNRVQFFHROELQDIRRWNWSSGFSGLG - FTRRVAEYFSPASF 578  
Cj/PALAS/1-859 493 AIKYPWHRSMRPRLEARSYIENYGV-PNDVWL--- KTY MMPYISNEKYLELAKLDFNHVQSLHQKELRDLRRRWTSWSSGFTELK - FTRRRTVEIYFSPASF 587  
Cj/PDTS/1-860 484 AIKYPWHRSMRPRLEARSYIENYGV-PNDVWL--- KTY MMPNISNEKYLELAKLDFNRVQFFHROELQDIRRWNWSSGFSGLG - FTRRVAEYFSPASF 578  
Cj/PaIPS/1-867 501 AIKYPWHRSMRPRLEARSYIENYGV-PNDVWL--- KTY MMPYISNEKYLELAKLDFNHVQSLHQKELRDLRRRWTSWSSGFSELK - FTRRRTVEIYFSSAASP 595  
Cj/ACPPS/1-802 477 ALEIPWYASLPRVETRFYIDQYGGENDVWIG--- KTL RMPYVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - LS-KRDLLLAYFLATGS 568  
Cj/PCPPS/1-801 473 ALDMPWFASLPRVETRFYIDQYGAESDWWIG--- KTL RMYVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - LS-KRDLLLAYFLATGS 568  
Cj/CMPPS/1-823 483 ALDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 578  
Cj/CMPPS/2/1-827 480 ALDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 578  
Cj/LePPS/1-800 478 ALDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/SrPPS/1-787 473 ALDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/SmPPS/1-793 471 GLEMFWLATLPRVEAKYVIQYVAGSGDWWIG--- KTL RMPYISNIDYHDLAKDFNRVQFFHROELQDIRRWNWSSGFSELK - FTRRVAEYFSPASF 582  
Cj/ZnPPS/1/1-823 478 TLDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/ZnPPS/2/1-827 474 TLDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/OsCPPS/1/1-867 518 TLDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/OsCPPS/2/1-800 475 TLDVPWFASLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/OsCPPS/4/1-767 487 AMDYPWAKSLPRVETRFYIEQYGGENDVWIG--- KTL RMPVNNNYLLEAKLDYNNCQAQHLIEMWNIQTYWYLESRLGEGF - MR-RSLLSDFYGAACS 574  
Cj/AaSHC/1-631 396 QFDNVIYDQYVAVVWALNTLRLPDERRR--- KTL RMTKDFRWLVQKQSSSGGAYDNDSDLPNHPFCDQEVPT - DPPSEDVYAVMLEECP 457  
Cj/HsOSC/1-732 445 STLDGCVIWSQGTAAVVALKVLQEKCPHVT--- EHIPERELCDAAVYVLMNRNPDGQGFATYETKRGHLELLNPEVFDGIMDIYVETCSAVMQLK 542

Consensus

AL+YPWYASLPRLEARSYIEQYGGENDVWIG---KTLRMPYVNSDKYLELAKLDFNQCQALHQLQDLQRWYKESGLDELG-+FTRRVEIYF+AAAS

Cj/TbTXS/1-862 591 LPEPEYSA--- TRIAFTIGCQLVFLMAIIFATLDELQVTEGVRK--- WDTLSLLEIIP-ECMQTQFKVWFKLMEVEVNDVVKV- 666  
Cj/HsKS/1-774 505 LPAELSD--- ARLSWAKNISILTAITDPIIGGATIDESINLQYVEK--- WNVVDKDCGSEETLRFILFSAIAICWMDGADFVKW- 689  
Cj/OsKS/1-812 529 LPRELSD--- ARIAWAKNGVLTVVVDFFLDQGSKEELENLALVEK--- WDHQHEE-FYSEQVRIYFSAIYTVNQLGAKASAL- 608  
Cj/OsSES/1-820 528 LPRELSD--- ARIAWAKNGVLTVVVDFFLDQGSKEELENLALVEK--- WDHQHEE-FYSEQVRIYFSAIYTVNQLGAKASAL- 608  
Cj/OsSOS/1-816 530 LPAELSD--- ARIAWAKNGVLTVVVDFFLDQGSKEELENLALVEK--- WDHQHEE-FYSEQVRIYFSAIYTVNQLGAKASAL- 608  
Cj/SKs/1-784 505 LSPPELSD--- ARIAWAKNGVLTVVVDFFLDQGSKEELENLALVEK--- WDHQHEE-FYSEQVRIYFSAIYTVNQLGAKASAL- 608  
Cj/AgABS/1-853 593 LPEPEFSK--- CREVYTKTSNFTVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/MGABS/1-873 601 MPEPEFAA--- CRAIYAKTSCLAIVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/PALAS/1-850 579 LPEPEFAT--- CRAVYTKTSNFTVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/PALAS/1-859 588 MPEPEFAT--- CRAVYTKTSNFTVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/PDTS/1-860 579 LPEPEFAT--- CRAVYTKTSNFTVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/PaIPS/1-867 596 LPEPEFAT--- CRDVTYKISIFTVILGLYAHGTLDDLKLFSEAVRR--- WDLSLVDQMP-QMKKICFGFYNTFNDIAEKFRER- 682  
Cj/ACPPS/1-802 573 LPEPERSH--- ERMVWAKSSVLVKAISSSGESSDRRSDFQGEYIANARRS--- DHFFNDRNMRDRPQSVQASRLAGVLIQLNQMDFLFMS- 664  
Cj/PCPPS/1-801 569 LPEPERSH--- ERLAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/CMPPS/1/1-823 570 LPEPERAK--- ERLAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/CMPPS/2/1-827 585 LPEPERAK--- ERLAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/LePPS/1-800 574 LPEARSTK--- ERMWVKTAALMDCVRSQF- SPQVSA- AAFLECEFA--- HYSSTALN--- SRYNTEDRLVGLLTLNLSLISALLT- 651  
Cj/SrPPS/1-787 570 LPEPERSK--- ERIAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/SmPPS/1-793 567 LPELERTN--- ERIAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/ZnPPS/1/1-823 574 LPEPERSA--- ERLAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/ZnPPS/2/1-827 570 LPEPERSA--- ERLAWAKTITALLLETIKCYVR- NEDLR- KDFAKFN- DHDID- VRDYSIARR- MKRNKTEHELVESLFAITGIESWDLRSL- Y 611  
Cj/OsCPPS/1/1-867 614 LPEPNRAS--- ERLAWARVSVLANITSRHFYSMMSMKMERFWMSSL--- YEENGVNLGLEQYAKDGILARTLQQLIDLLSQEPFVRE- 698  
Cj/OsCPPS/2/1-800 571 LPEPNRAA--- ERLAWARVSVLANITSRHFYSMMSMKMERFWMSSL--- YEENGVNLGLEQYAKDGILARTLQQLIDLLSQEPFVRE- 698  
Cj/OsCPPS/4/1-767 565 LPEPNRAA--- ERLAWARVSVLANITSRHFYSMMSMKMERFWMSSL--- YEENGVNLGLEQYAKDGILARTLQQLIDLLSQEPFVRE- 698  
Cj/AaSHC/1-631 458 SFGYDADW--- KVIRRAVEYIKREQKPDGSWFGRWG--- VNYLYGTGAVVSLAKVAGIDTREP- 514  
Cj/HsOSC/1-732 543 YHKKRFPEHRAEIRETLTQGLFCRRQQRADGSWEGSWG--- VCFYGTGWFGLAFAACMGQTYRDG- 606

Consensus

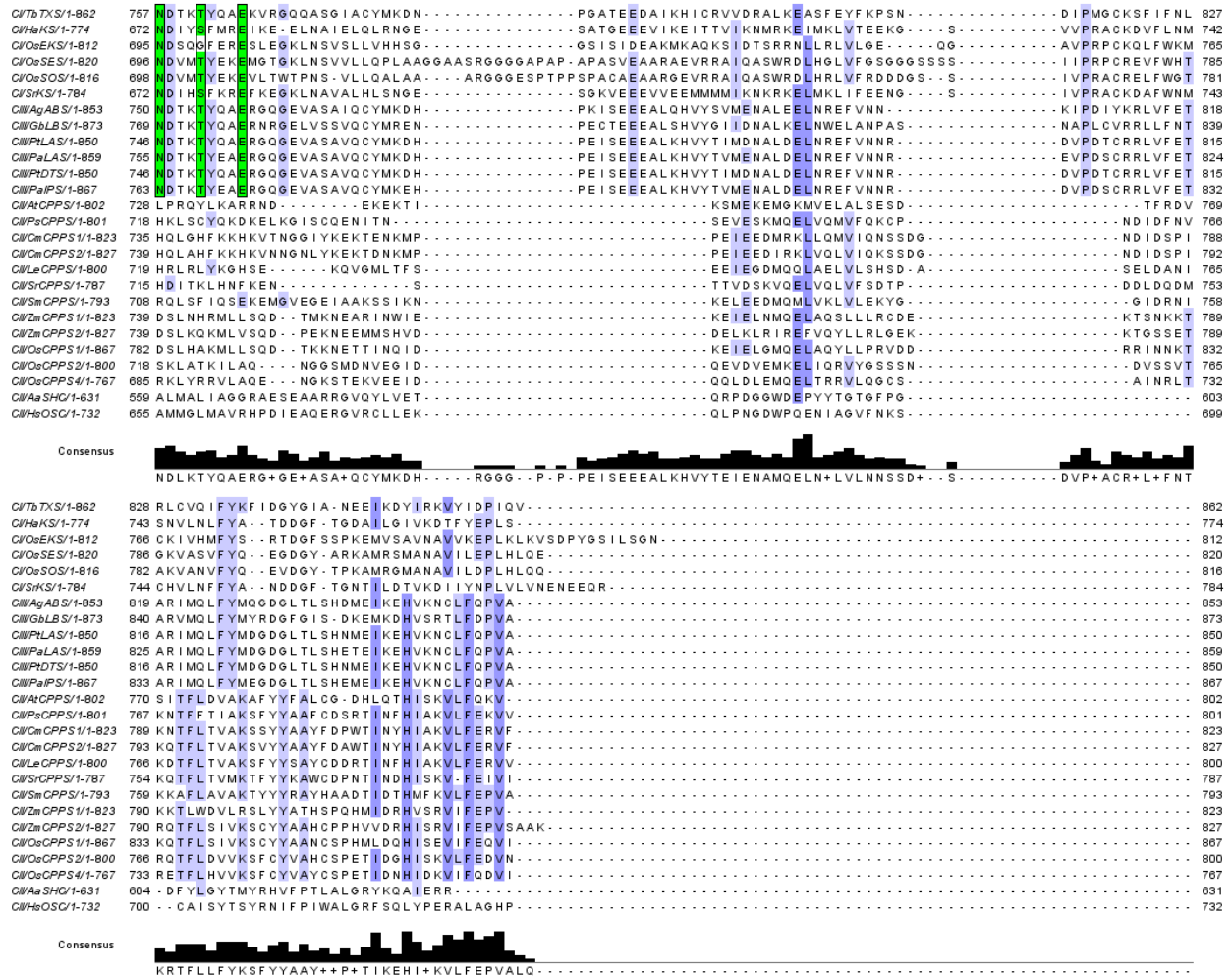
IFEPERAA---ERLAWAKTISILTVIIDD+FAHGSLDELKLFESVQRNFDIAD-+E++SWDLSLVD+MPSE+MKILFQALYNTLNEIALDGRK+-O

Cj/TbTXS/1-862 670 GRDMLAHIRKPVELYFNQVQEREWLEAGYIP-TFEEYLKTYAIVSGLGPCTLQ- PILLMDELVKKDDVVEKVVHYSNMFLVSLSWRLT 756  
Cj/HsKS/1-774 586 ERDVTSHVYQIWLDDLNSMLTEIWRDAYSPT-MDEYMKNASVFSALGPIVPLN- TLYFVGPKLSEEVIVESSEYH-NIFDQMSVMGRLL 671  
Cj/OsKS/1-812 609 GRDVTKHLTEIWLCLMRSMMTEAEWQRTKYP-TMEEYMANAVVFSALGPIVPLN- TLYFVGPKLSEEVIVESSEYH-NIFDQMSVMGRLL 694  
Cj/OsSES/1-820 609 KRVCVIDHIAEQVMVMRAMLTAEAWAAGKHIPATMEYMSVAFSALGPIVPLN- SAYLLGEELPEEAVRSPPEY-RLGLASAVGRLL 695  
Cj/OsSOS/1-816 611 NRCVMDHIAKQVMVMRAMLTAEAWAASRHIPATMEYMSVAFSALGPIVPLN- SAYLLGEELPEEAVRSPPEY-QLLRHASAVGRLL 697  
Cj/SKs/1-784 586 ARDVTSHVYQIWLLELMSMLREAIWTRDAYVPTLNNEYMENAVSFSALGPIVPLN- AIFYVGPKLSEEVIVESSEYH-NLFLMSTQGRLL 671  
Cj/AgABS/1-853 693 GRDVLGYIQNVVWQVLEAYTKAEAWSEAKYVPSFNEYIENASVSIALGTVVLI- SALFTGVELTDEVLSKIIRDRESRFLQMLGLTRVL 749  
Cj/MGABS/1-873 682 GRDVLGYIRNVWEIQLASYTEAEWSAAKYVPTFNEYIENASVSIALGTVVLI- SIFFTGEELPDYILQQVDLRSKFLHVLSTGLRL 768  
Cj/PALAS/1-850 659 GRDVLGYIQNVVWQVLEAYTKAEAWSAVRYVPSFDEYIENASVSIALGTVVLI- SALFTGELTDDILSKIGRDSRFLYMLGLTRVL 745  
Cj/PALAS/1-859 688 GRDVLGYIRNVWEIQLASYTEAEWSAAKYVPTFDEYIENASVSIALGTVVLI- SALFTGELTDDILSKIGRDSRFLYMLGLTRVL 745  
Cj/PDTS/1-860 659 GRDVLGYIQNVVWQVLEAYTKAEAWSAVRYVPSFDEYIENASVSIALGTVVLI- SALFTGELTDDILSKIGRDSRFLYMLGLTRVL 745  
Cj/PaIPS/1-867 676 GRDVLGYIRNVLEILLAAHTKAEAWSAAKYVPSFDEYIENASVSIALGTVVLI- SVLFTGELTDDILSKIGRDSRFLYMLGLTRVL 762  
Cj/ACPPS/1-802 665 GRDNNLLYLSWGDWMEKWKVLY- GDEGE- GELMVK- MIIILMKNNDLTN- FFFTHFVRLAEIINRI 727  
Cj/PCPPS/1-801 652 GHEIGYDMHQCKWKLLSSWQSE- GDCKCEGAELLIQ- GDLKNSHWI SEGP- SM-QSTIQLLQLTNSIC 717  
Cj/CMPPS/1/1-823 609 GFDITHHLHQAWEKWLFWKQWED- GDVHKEEAELLVQ- TILNSGCTLELDL- LSNPQYQKLSYLTNKKV 734  
Cj/CMPPS/2/1-827 673 GVDITHHLHQAWEKWLFWKQWED- GDVHKEEAELLVQ- TILNSGCTLELDL- LSNPQYQKLSYLTNKKV 738  
Cj/LePPS/1-800 652 GRDIPHQLRHAWENWLLTKLDGEG- GVEEGGAEIIR- TNLNCSVHWISKEI- LLSHPYQKLEIITNRV 718  
Cj/SrPPS/1-787 652 SQDIRHQLRHAWENWLLTKLDGEG- GVEEGGAEIIR- TNLNCSVHWISKEI- LLSHPYQKLEIITNRV 718  
Cj/SmPPS/1-793 644 DRYTRHQLKNAWSVWLTQLQHG- EDVET- AELMVO- MIMNTAGRWYSKEI- LTHPQYQKLEIITNRV 714  
Cj/ZnPPS/1/1-823 657 QRFIHSLLSLAWTEWMLQKANKEENKYHKCSG- IEP- GYMVHRDRTYLVLVQ- VJICAGRIGEAES- MINNKNDWFIQLTCTG 738  
Cj/ZnPPS/2/1-827 657 EDIHKLLRSAAEWREKADAAADVCSNGSSA- VHEQGSCTVHNKQCTCLLAR- MJEISAGRAAGEA- SEDG- DRRIQLTGSIC 738  
Cj/OsCPPS/1/1-867 699 QKCIHNLIRCAWIEWMMQINMKDGRYDKGRV- MHPG- SRTVHNKQCTCLLAR- IVEICAGRIGEAES- MINNTEGWSFIQLASSIC 781  
Cj/OsCPPS/2/1-800 640 GNA- SDSLRLEAWKQWLMSTWNE- SQSSTEGQDTALLVR- TIEICSRHSGAEQS- LKNSADYARLEQIASSMC 784  
Cj/OsCPPS/4/1-767 614 DDAYSGLRLEAWKQWLMMAWTAKES- SQSIEGQDTALLVR- ALEIFGGRHVLTGQ- RPPDWEYSQLEALTSIC 697  
Cj/AaSHC/1-631 515 ---YIQKALDQWVEHQHP- SQSIEGQDTALLVR- DGGWGEDCRSYEDPAYAGKASTPSQTAW 558  
Cj/HsOSC/1-732 607 ---TACAEVSRACDFLLSRQMA- DGGWGEDFESCEERRYVQSAQSIHNTCW 654

Consensus

GRDVL+HLRQAWVWLEA+TKEAWEWSAAKYVPAT++EY+SNASVSIALGELVQLQ-...SILFTGGLRLEEVLKIE+HSRFLRLM+LQGLR

Supplementary Figure 4 (continued)



**Supplementary Figure 4. Alignment of the amino acid sequences of 6 class I, 14 class II, and 6 bifunctional class I/II diterpene cyclases along with class II triterpene cyclases squalene-hopene cyclase and oxidosqualene cyclase. Annotations: “EDXXD-like” motif is orange; E211 of CPS is light green; H331 of CPS is cyan (arginines at the same position are magenta); T421 of CPS is violet; N425 of CPS is pink; Y511 of CPS is plum; DDXXD and NSE/DTE motif of class I cyclases are red and green; DXDD motif of class II cyclases is yellow. Full names of the sequences are given in the following table.**

Name	Full Name	Organism	Accession N.	Class
TbTXS	taxadiene synthase	<i>Taxus brevifolia</i>	AAC49310.1	I
HaKS	kaurene synthase	<i>Helianthus annuus</i>	CBM82408.1	I
OsEKS	<i>ent</i> -kaurene synthase	<i>Oryza sativa</i>	BAE72099.1	I
OsSES	stemer-13-ene synthase	<i>Oryza sativa</i>	BAD34478.1	I
OsSOS	stomodene synthase	<i>Oryza sativa</i>	AAZ76733.1	I
SrKS	copalyl pyrophosphate synthase	<i>Stevia rebaudiana</i>	AAD34294.1	I
AgABS	abietadiene synthase	<i>Abies grandis</i>	AAK83563.1	I / II
GbLBS	levopimaradiene synthase	<i>Ginkgo biloba</i>	AAS89668.1	I / II
PtLAS	abietadiene/levopimaradiene	<i>Pinus taeda</i>	Q50EK2.1	I / II
PaLAS	abietadiene/levopimaradiene	<i>Picea abies</i>	AAS47691.1	I / II
PtDTS	diterpene synthase	<i>Pinus taeda</i>	AAX07435.1	I / II
PaIPS	isopimaradiene synthase	<i>Picea abies</i>	Q675L5.2	I / II
AtCPPS	<i>ent</i> -copalyl diphosphate synthase	<i>Arabidopsis thaliana</i>	AAA53632.1	II
PscPPS	<i>ent</i> -copalyl diphosphate synthase	<i>Pisum sativum</i>	AAB58822.1	II
CmCPPS1	<i>ent</i> -copalyl diphosphate synthase 1	<i>Cucurbita maxima</i>	AAD04292.1	II
CmCPPS2	<i>ent</i> -copalyl diphosphate synthase 2	<i>Cucurbita maxima</i>	AAD04293.1	II
LeCPPS	<i>ent</i> -copalyl diphosphate synthase	<i>Lycopersicon esculentum</i>	BAA84918.1	II
SrCPPS	<i>ent</i> -copalyl pyrophosphate synthase	<i>Stevia rebaudiana</i>	AAB87091.1	II
SmCPPS	copalyl diphosphate synthase	<i>Salvia miltiorrhiza</i>	ABV57835.1	II
ZmCPPS1	<i>ent</i> -copalyl diphosphate synthase	<i>Zea mays</i>	AAA73960.1	II
ZmCPPS2	<i>ent</i> -copalyl diphosphate synthase	<i>Zea mays</i>	NP_001105257.1	II
OsCPPS1	<i>ent</i> -copalyl diphosphate synthase	<i>Oryza sativa</i>	Q6ET36.1	II
OsCPPS2	<i>ent</i> -copalyl diphosphate synthase	<i>Oryza sativa</i>	Q6Z510.1	II
OsCPPS4	<i>syn</i> -copalyl diphosphate synthase	<i>Oryza sativa</i>	Q0JF02.1	II
AaSHC	squalene-hopene cyclase	<i>Alicyclobacillus acidocaldarius</i>	BAA25185.1	II
HsOSC	oxidosqualene cyclase	<i>Homo sapiens</i>	P48449.1	II







Aa SHC/1-631 422 VDNTSDLPNHIPFCDFG...EVTDPSPSEDVTAHIVLECFGSGFYD...DAWKVIRRAVEYLKREQ...KPDGSWFGRWGVNLYLGTGAVVSALKAVGIDT 511  
 Bp SHC/1-657 438 PENTQYVYNNIPFSDHG...ALLDPPTADVTARVVSMLSQLQGETR...ATSRALDRDGVTYLLNDQ...EKDGSWYGRWGMNFIYGTWTALCSLNAAGLGH 528  
 Se SHC/1-636 423 ADNTRLVNELPFCDFG...AVIDPPSADVTARVVEALCVLGRG...DG-EAVRRGVRWLLDHO...ELDGSWFGWRGANHYVGTGAAPVALVRAGLRR 511  
 Gb SHC/1-651 431 ADNDKDYLNHIFPADHG...ALLDPPTADVTARISFLAQLNGR...EDKPVDIRAMAWLKERD...EADGSWFGWRGNTNYIYGTWSVLCALMNVAGMPH 520  
 Fa SHC/1-704 405 ADNVRTLTTKIPFCDFG...EVDVPPSADVTARHIVEMALDGRS...DH-PITQRAVQWLLDNO...EPGGSWFGWRGVNHLVGTGAVVPALIGAGVPT 583  
 Ct SHC/1-681 456 PENTHLYVNNIPFADHG...ALLDPPTADVTARVVSMLSQLQGETR...ANSEPARALRYLLAED...MPDGSWFGWRGNTNYIYGTWSVLCALNAAGLPH 546  
 Zn SHC/1-658 440 ANNSYTYVNNIPFADHG...ALLDPPTADVTARVVSMAQAQATP...ITDPKMKAAVDYLLKED...EEDGSWFGWRGVNFIYGTWSVLCALNVAALPH 529  
 Bc SHC/1-631 425 KGVSTHYLANLPIENAS...DMITDPSTPDI TGRVLEFFGTQYAGN...ELPEKQKQSAINWLMNVY...EENGSWYGRWGVNFIYGTWAVLTLGRSLIP 518  
 Mn SHC/1-663 444 ADNTHYVYNNIPFSDHG...ALLDPPTADVTARVVSMLSQLQGETR...ESCPPLDRGVAYLLAD...EADGSWYGRWGMNFIYGTWSVLCALNAAGVDP 534  
 Mc SHC/1-667 449 ADNNHHYVNNIPFSDHG...ALLDPPTADVTARVVSMLSQLQGETR...ATSRALDRDGVTYLLNDQ...EKDGSWYGRWGMNFIYGTWSVLCALNTAGVDP 539  
 Mp SHC/1-667 449 ADNNHHYVNNIPFSDHG...ALLDPPTADVTARVVSMLSQLQGETR...ETSRALDRDGVTYLLNDQ...EKDGSWYGRWGMNFIYGTWSVLCALNAAGVDP 539  
 Ssp SHC/1-666 449 ADNTSPFPNRLPFCDFG...EVIDPPSADVTARHIVEMAVEGLA...SH-PRTRGIEWLLAED...EACGAWFGRWGVNFIYGTGSVVPALITAGLPA 537  
 Te SHC/1-632 420 VDNNQDVLNATPYGDLK...AMIDPPNTADVTARVLEMGVCGGLA...MDSWRVVRGIDFLVRED...EEECSWFGWRGVNFIYGTSGVILALAVMARES 509  
 Rp SHC/1-685 447 KDNDKKIKAKIPFCDFG...EALDPPSADVTARHIVEMAVEGLA...RNNHPSIVRALDYLKRED...EPEEPWFGRWGVNFIYGTGAVLPAALAIIDEM 536  
 Ac SHC/1-633 428 ADNTRVLRLPFCDFG...AVIDPPSADVTARHIVEMAVEGLA...DH-PATVAVGVWLLAHO...EPDGSWFGWRGANHIYGTGAVVPALIGAGVSP 516  
 Su SHC/1-641 431 ADNNWFEVLSQVFFADHN...AMLDPDTCADITGRVLEALASQGLD...RNNKAVRRGAEWLIRHO...ENDGSWYGRWGVNFIYGTGCFALRGIAAGVEND 520  
 Hs OSC/1-732 505 TKRGGHLLELLNPSEVFGD...IMIDYTYVCTSAVMQALRHFHRRFPHRAAEIRETLTQGLFCRRQO...RADGSWEGSWGVCFITGTFWGLFAFCMGTY 603  
 Rn OSC/1-733 506 TKRGGHLLELLNPSEVFGD...IMIDYTYVCTSAVMQALRHFHRRFPHRAAEIRETLTQGLFCRRQO...RADGSWEGSWGVCFITGTFWGLFAFCMGHIY 604  
 At OSC/1-767 543 AARGRTWLEWLSVPEFMEDTIVEHEVYVCTGSAIVALARFLKEFPHRRREEVEKFIKNAVVKYIESFQ...MPDGSWYGNWGVCFMGTFAVRGVAAGKTY 641  
 Am OSC/1-762 536 LTRSYAWLEIINPAESFGD...IVIDYPTAESTSACIQALASFRLMYPGHRREIEKCIITKGVQFIEKTO...EHDGSWYGSWAVCYTNGTWYGVKGLISGKCY 634  
 La OSC/1-760 504 LTRSYAWLEIINPAEAFVGD...IVIDYQVYVCTSAAIQGLKAFMQLHPGHRKEIQKCIKAAANFIESIQ...QDGSWYGSWICVYTGTFWFKGLVACGRY 632  
 Ca OSC/1-728 506 GKAPLLLEKLNPAEAFVGD...IVIDYQVYVCTDSVSLGLTYFAKYYPDYKPEIKQKTISSAIQYILDSQDNIDGSWYGCWICVYASMFALAEALTVGLY 605  
 To OSC/1-758 532 LTRSYAWLEIINPAEAFVGD...IVIDYQVYVCTSAVVSQSRFTKLYPSHRIQIETCIEKAIKIAFIERSO...LGDGSWYGSWAVICVYTGTFWFKGLVACGRY 630  
 Pg OSC/1-780 547 LTRSYAWLEIINPAEAFVGD...IVIDYQVYVCTSAAIQGLKAFMQLHPGHRREIEKCIKAAANFIESIQ...LPGDGSWYGSWICVYTGTFWFKGLVACGRY 645  
 Lj OSC/1-761 535 LTRSYAWLEIINPAEAFVGD...IVIDYQVYVCTSAAIQGLKAFMQLHPGHRREIEKCIKAAANFIESIQ...LADGSWYGSWICVYTGTFWFKGLVACGRY 633  
 Ss OSC/1-728 506 KIRSTLLLEKLNPAEAFVGD...IVIDYQVYVCTDSVSLGLTYFAKYYPDYKPEIKQKTISSAIQYILDSQDNIDGSWYGCWICVYASMFALAEALTVGLY 605  
 Bt OSC/1-732 505 TKRGGHLLELLNPSEVFGD...IMIDYTYVCTSAVMQALRHFHRRFPHRAAEIRETLTQGLFCRRQO...RADGSWEGSWGVCFITGTFWGLFAFCMGHTY 603



Aa SHC/1-631 512 RE...FYIQKALDWVEQHNPDGGWGEDCR...SYEDP-AYAGK...GASTPQATWALLMALIAGG...RAESEARRGVQYLVETORPDGGWDEPYTGTG 600  
 Bp SHC/1-657 529 DD...PRVKRAAQLWLSIQNADGGWGEDCR...SYKLDYRGRYER...APSTSSQATWALLGLMAAG...EVDNPAVARGVDYLLGTORREHGLWDETRFTATG 618  
 Se SHC/1-636 512 DH...LALRRVAVRLEVHQNDDGGWGEDLRSYDDP-VVWGR...GRSTASQATWALLALLAVD...LHDTDAVRRGVRGFAETORPDGTWDEPFTGTG 610  
 Gb SHC/1-651 521 DD...PAITRAAVNVLVATQREDDGGWGEDDEETYPDASGAQPGRYKE...TPSQATWALLIGLMAAG...EAHEATRRGLAYLQATOKPDGGEDEAAYTAVG 612  
 Fa SHC/1-704 584 DH...PAITAAVRRVLLLEHDSPEGGWGEDLRSYDTP-AWIGR...GELTASQATWALLALLAVD...PHS-LAVKRRGVRWVLCETORPDGTWDEPYTGTG 671  
 Ct SHC/1-681 547 EA...PELCRAVAWALARIQNADGGWGEDDSSYRLDYSYGEPE...APVSASQATWALLALLMAAG...AAQHPAVARGIDYLLRTOQDGGWVWHPRYTAVG 636  
 Zn SHC/1-658 510 DH...LAVQKAVAWLKTIQNEDGGWGEDCR...SYALDYSYGEPE...MDSTASQATWALLGLMAVG...EANSEAVTKGINWLAQNDDEEGLWKEDEYSGGG 619  
 Bc SHC/1-631 517 SD...PSLKRALALWLEHIDHEDGGWGEDSCQSSVEKRFVTLF...FSTPQATWALLDALISY...DKETSIVIRKGINYLLSNPNYINETY...TGTG 601  
 Mn SHC/1-663 535 AS...EPVRRVAVNMLTTIQNPDGGWGEDAASVYKLERYGRYER...APSTASQATWALLGLMAAG...EADSPAVARGINYLTRSGDAGDLWDETRFTATG 624  
 Mc SHC/1-667 540 QS...PEIRKAVAWLIRIQNPDGGWGEDASVYKLN-PEFEP...GYSTASQATWALLALLMAAG...EVDPAVARGVNYLVRTGQDGLWSEERYATG 628  
 Mp SHC/1-667 540 QS...PEIRKAVAWLIRIQNPDGGWGEDASVYKLN-PEFEP...GYSTASQATWALLALLMAVG...EVDPAVARGVNYLVRTGQDGLWSEERYATG 628  
 Ssp SHC/1-666 538 GH...PAITRAVAVLESGVNDGGWGEDLRSYQDE-KWIGH...GESTASQATWALLALLAAG...RDRTPVARGVWVLEAQDQDGSWDEPYTGTG 626  
 Te SHC/1-632 510 HR...GIERGASVWLVGQNSDGGWGEDSCVSYNDP-SLKGK...GKSTASQATWALLIGLLAAGEGTGNFARADIDGGVGLVSTQNDGGWDEPYTGTG 602  
 Rp SHC/1-685 537 RT...PYIARACDVLARQANGGWGEDSCVSYMDA-QOAGE...GTASQATWALLMALJAADR...PODRDAIEGGLYLTETOR-DGTWQVEVHYTGTG 625  
 Ac SHC/1-633 517 DT...PPIRRAIRLWLEEHQNDGGWGEDLRSYTOPALWVGR...GVSTASQATWALLALLAAG...EASPAVDGVRVRLVTTQDGGWDEPYTGTG 606  
 Su SHC/1-641 521 RE...AHLLRGEWLSRIQNDGGWGEDSCVSYDNRIFTGPP...STPSTASQATWALLIGLLAAG...DANLSVQHGLEYLLETORSDGSWDEPFTGTG 608  
 Hs OSC/1-732 904 RDGTACAEVSRACDFLLSRQMDAGGGWGEDFECRERY...VQS-AQO...HNSQVWTCWALLGLMAVRHP...DIEAQRGVRWVLCETORPDGTWDEPYTGTG 697  
 Rn OSC/1-733 805 QNRACAEVQAACHFLSRQMDAGGGWGEDFECRERY...VQS-AQO...HNSQVWTCWALLGLMAVRHP...DIEAQRGVRWVLCETORPDGTWDEPYTGTG 697  
 At OSC/1-767 842 QN...LETDQNVESGGWGESYLSCPNKKYTLLEG-NRTNVNVSQALMLVIMGQME...RDLPLVHRAAKVILNSDLNDGPOEIMGVF 735  
 Am OSC/1-762 835 EN...SHSIRKACDFLLSKQLKSGGGWGESYLSQEQKVTYTLLEG-NRAHVNTSWMALLIDAGQAO...RDAEPLHRAAKVILNSDMENGEFPOEIMGVF 728  
 La OSC/1-760 833 DN...SKSIRKATEFLLSKQLKSGGGWGESYLSAHKHYVTLLEG-CKE...HIVNTSWMALLIKAGQAO...RDLPLHRAAKVILNSHDDGPOEIMGVF 728  
 Ca OSC/1-728 806 ES...SSAVKKGCDFLISKQLPDGGWSESMKOCETHSY-VNG-ENSLVVSAWALLIGLILQNY...DEEPIKRGIQFLMKRGLPTGSEWYKVIDGVF 695  
 To OSC/1-758 931 ET...SHSIRKACDFLLSKQLKSGGGWGESYLSQEQKVTYTLLEG-NKSHINTSWMALLIEAGQPR...RDRPLHRAAKVILNSDMENGEFPOEIMGVF 724  
 Pg OSC/1-780 846 RN...CYSIRKACDFLLSKQLKSGGGWGESYLSQEQKVTYTLLEG-NKSHVNTSWMALLIEAGQAO...RDPSPHRAAKVILNSDMENGEFPOEIMGVF 739  
 Lj OSC/1-761 834 QD...SHSIRKACDFLLSKQLKSGGGWGESYLSQEQKVTYTLLEG-NKSHVNTSWMALLIEAGQAO...RDPAPLHRAAKVILNSDMENGEFPOEIMGVF 727  
 Ss OSC/1-728 806 EN...SATVIRKGCDFLLISKQLPDGGWSESMKACETHSY-VNG-DKSYVVDWAVLIGLILGQYP...DQEPIRKGILQLMQRGLPTGSEWYKVIDGVF 695  
 Bt OSC/1-732 804 HNGVACAEISRACDFLLSRQMDAGGGWGEDFECRERY...VQS-AQO...HNSQVWTCWALLGLMAVRHP...DVAALERGVSYLLEKDLNSDGPQENISGVF 696



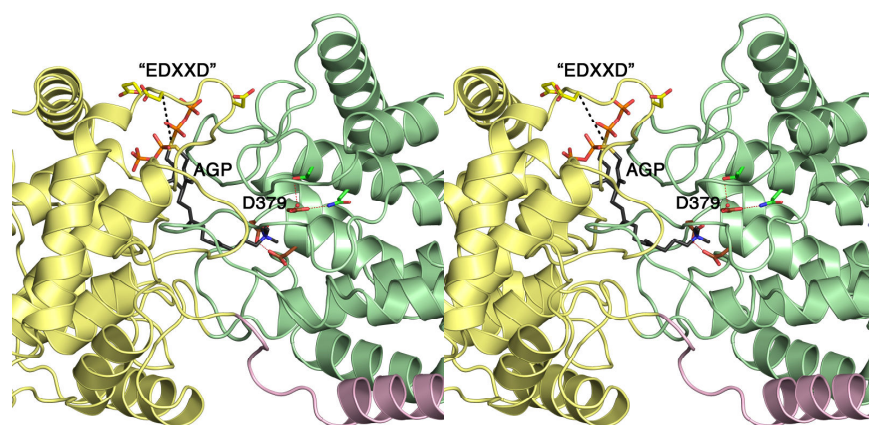
Aa SHC/1-631 801 FPG...DFYLG...TMYRHVFPPTALGRYKQAIER...R...ANATRVTVGM 631  
 Bp SHC/1-657 819 FPR...VFYLR...HGVRKFFPLWALARYRNLRK...ANATRVTVGM 657  
 Se SHC/1-636 801 FPG...DFYIN...HLVRLVFPVPTALGRYEQARRE...QSGGSG...636  
 Gb SHC/1-651 813 FPR...VFYLR...HGVRKFFPLMALSRYKLNRS...SNMKKVSFGF 651  
 Fa SHC/1-704 872 FPR...DFSLN...HLVRLVFPPTALGRYVSLTGV...ATP...704  
 Ct SHC/1-681 837 FPR...VFYLR...HGVRKFFPLWALARYRNLRK...ANQPIVHYGM 681  
 Zn SHC/1-658 820 FPR...VFYLR...HGVRKFFPLWALARYRNLRK...ANQPIVHYGM 658  
 Bc SHC/1-631 802 LPG...GFYIR...HSYAHYPLLTLAHYAKKYRK...631  
 Mn SHC/1-663 825 FPR...VFYLR...HGVRKFFPLWALARYRNLRK...SNRRVAVGM 663  
 Mc SHC/1-667 829 FPR...VFYLR...HGVRKFFPLWAMARFNLRK...GNSRQVQGM 667  
 Mp SHC/1-667 829 FPR...VFYLR...HGVRKFFPLWAMARFNLRK...GNSRQVQGM 667  
 Ssp SHC/1-666 827 FPR...DFSYN...HLVROVFPPTALGRYVHDDPF...ADRMAAEGA 666  
 Te SHC/1-632 803 FPG...HFYIK...HFSQYFPLMALGRYVSLLSG...632  
 Rp SHC/1-685 826 FPGVGGQTIKLNLDPLLSKRLMGPELRSFMLR...DLRHYFPMMAIGRVLQRGD...RSGH...685  
 Ac SHC/1-633 807 FPG...DFYIN...HLVRLVFPISALGRVNR...633  
 Su SHC/1-641 809 FPR...VFYLN...HMYKDYFPLLASFVKARAG...SNG...641  
 Hs OSC/1-732 897 NKS...CAIS...TSYRNIFPIWALGRFSQLYPERALAGHP...732  
 Rn OSC/1-733 898 NKS...CAIS...TSYRNIFPIWALGRFSQLYPNTLAGHI...733  
 At OSC/1-767 736 KMN...VMVH...ATYRNIFITLWALTYTKALRVPLO...767  
 Am OSC/1-762 729 NRN...CMIS...SAYRNIFPIWALGEYRTRVLS...AGH...762  
 La OSC/1-760 727 NKS...CMIS...SAYRNIFPIWALGEYRIRVLQLH...EKF...760  
 Ca OSC/1-728 996 NHS...CAIE...PSYRFLFPIKALGLYKKNYKQDKVLV...KSLG...728  
 To OSC/1-758 725 NKN...CMIS...SSYRNIFPIWALGEYLNHVITK...758  
 Pg OSC/1-780 740 NKN...CMIS...SAYRNIFPIWALGEYLNVRVLPSPRNILKTLNVV...780  
 Lj OSC/1-761 728 NKT...CTIS...SAYRNIFPIWALGVYRSHVLCGG...KE...761  
 Ss OSC/1-728 996 NHS...CAIE...PSYKFLFPIKALGLYKQKFNQEPVIV...728  
 Bt OSC/1-732 897 NKS...CAIS...TSYRNIFPIWALGRFSQLYPERALAGHP...732



Supplementary Figure 5 (continued)

**Supplementary Figure 5. Alignment of the amino acid sequences of 27 class II triterpene cyclases.** Annotations: D447 of squalene-hopene cyclase is **violet**; H451 of squalene-hopene cyclase is **pink**; Y495 of squalene-hopene cyclase is **orange**; DXDD motif of class II cyclases is **yellow**; Y609 of squalene-hopene cyclase is **plum**; T314 of squalene-hopene cyclase (corresponding to H331 of CPS) is **green**. Full names of the sequences are given in the following table.

Name	Full Name	Organism	Accession Number	Class
AaSHC	squalene-hopene cyclase	<i>Alicyclobacillus acidocaldarius</i>	BAA25185.1	II
BpSHC	squalene-hopene cyclase	<i>Burkholderia pseudomallei</i>	YP_112341.1	II
SeSHC	squalene-hopene cyclase	<i>Saccharopolyspora erythraea</i>	YP_001106521.1	II
GbSHC	squalene-hopene cyclase	<i>Granulibacter bethesdensis</i>	YP_745928.1	II
FaSHC	squalene-hopene cyclase	<i>Frankia alni</i>	YP_711680.1	II
CtSHC	squalene-hopene cyclase	<i>Cupriavidus taiwanensis</i>	YP_002008614.1	II
ZmSHC	squalene-hopene cyclase	<i>Zymomonas mobilis</i>	CAA51958.1	II
BcSHC	squalene-hopene cyclase	<i>Bacillus cereus</i>	ZP_00239607.1	II
MnSHC	squalene-hopene cyclase	<i>Methylobacterium nodulans</i>	YP_002502189.1	II
McSHC	squalene-hopene cyclase	<i>Methylobacterium chloromethanicum</i>	YP_002420995.1	II
MpSHC	squalene-hopene cyclase	<i>Methylobacterium populi</i>	ACB79998.1	II
SspSHC	squalene-hopene cyclase	<i>Streptomyces</i> sp. ACT-1	ZP_06282416.1	II
TeSHC	squalene-hopene cyclase	<i>Trichodesmium erythraeum</i>	YP_720632.1	II
RpSHC	squalene-hopene cyclase	<i>Rhodospseudomonas palustris</i>	ABJ08391.1	II
AcSHC	squalene-hopene cyclase	<i>Acidothermus cellulolyticus</i>	YP_873455.1	II
SuSHC	squalene-hopene cyclase	<i>Solibacter usitatus</i>	YP_822539.1	II
HsOSC	oxidosqualene cyclase	<i>Homo sapiens</i>	P48449.1	II
RnOSC	oxidosqualene cyclase	<i>Rattus norvegicus</i>	AAA91023.1	II
AtOSC	oxidosqualene cyclase	<i>Arabidopsis thaliana</i>	BAF80447.1	II
AmOSC	oxidosqualene cyclase	<i>Allium macrostemon</i>	BAA84603.1	II
LaOSC	oxidosqualene cyclase	<i>Luffa aegyptiaca</i>	BAA85267.1	II
CaOSC	oxidosqualene cyclase	<i>Candida albicans</i>	AAA34342.1	II
ToOSC	oxidosqualene cyclase	<i>Taraxacum officinale</i>	BAA86933.1	II
PgOSC	oxidosqualene cyclase	<i>Panax ginseng</i>	BAA33462.1	II
LjOSC	oxidosqualene cyclase	<i>Lotus japonicus</i>	BAE95409.1	II
SsOSC	oxidosqualene cyclase	<i>Scheffersomyces stipitis</i>	ABN66417.2	II
BtOSC	oxidosqualene cyclase	<i>Bos taurus</i>	ABD24094.1	II



**Supplementary Figure 6. Position of the “EDXXD like” motif with respect to the active site.**

The  $\beta$  and  $\gamma$  domains of CPS are green (with pink N-terminus) and yellow; side chains are shown with darker colors. Carbon atoms of (*S*)-15-aza-14,15-dihydrogeranylgeranyl thiolodiphosphate (**1**) are dark grey and aspartic acid side chains in the D<sup>377</sup>XDD motif are brown. The separation between the thiolodiphosphate group of **1** and the “EDXXD like” motif (18 Å) is shown with a black dashed line.



## References

26. Mann, F. M. *et al. J Biol Chem* **284**, 23574-23579 (2009).
27. Peters, R. J., Ravn, M. M., Coates, R. M. & Croteau, R. B. *J Am Chem Soc* **123**, 8974–8978 (2001).
28. Luft, J. R. *et al. J Struct Biol* **142**, 170-179 (2003).
29. Otwinowski, Z. & Minor, W. *Methods Enzymol* **276**, 307-326 (1997).
30. Adams, P. D. *et al. Acta Crystallogr* **D66**, 213-221 (2010).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. *Acta Crystallogr* **D66**, 486-501 (2010).
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J Appl Cryst* **26**, 283-291. (1993)
33. Brünger, A. T. *et al. Acta Crystallogr* **D54**, 905-921 (1998).
34. Larkin, M. A. *et al. Bioinformatics* **23**, 2947-2948 (2007).