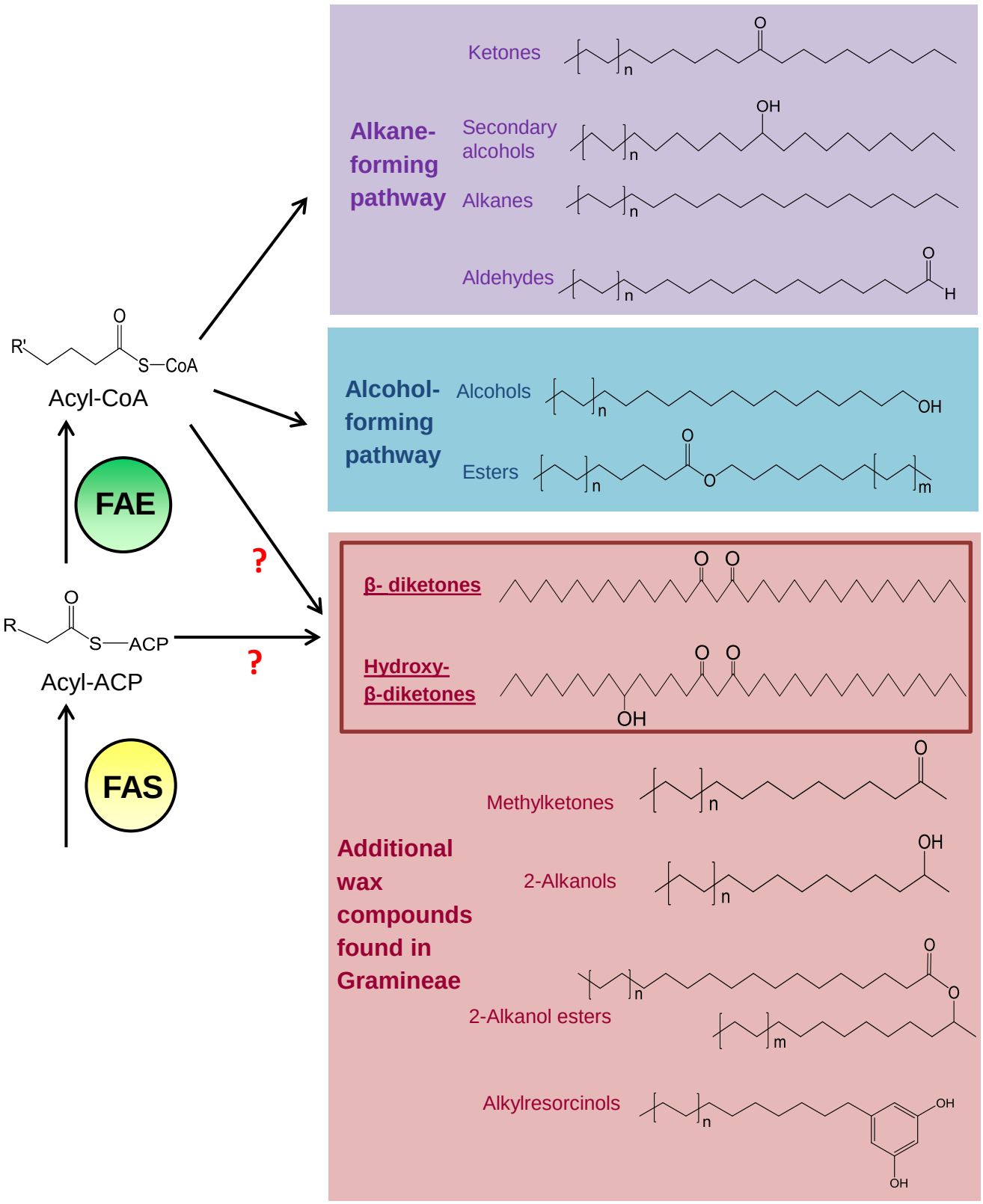
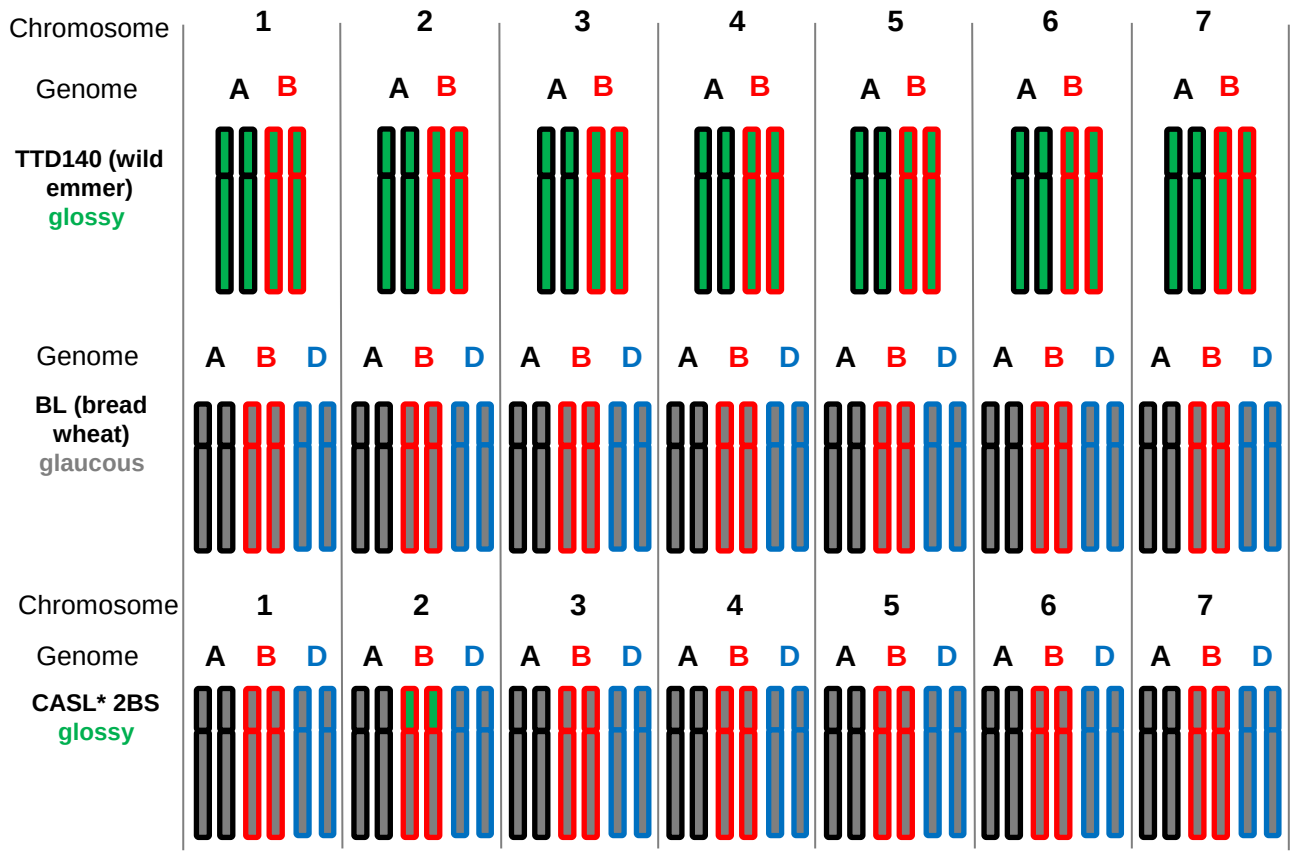
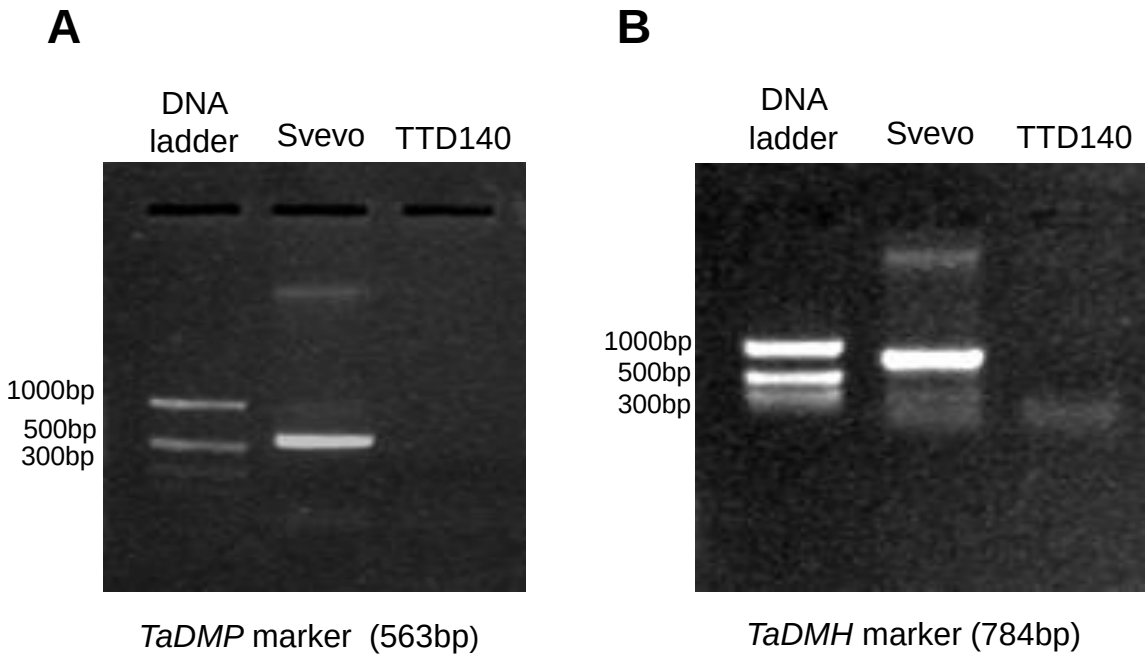


Supplemental Fig. 1

Supplemental Figure 1. General scheme of biosynthetic pathways generating different cuticular wax components. The alkane-forming (in purple) and alcohol-forming (in blue) pathways of cuticular wax biosynthesis. Many Gramineae species have additional wax compounds (in red). β-diketones and hydroxy-β-diketones, synthesized through an independent pathway, are shown in a red rectangle. Methylketones, 2-alkanols and 2-alkanol esters are side products of the β-diketone biosynthesis pathway. The β-diketones may be derived from FAS products or FAE products. Alkylresorsinols are polyketides also found in many Gramineae species. FAS- fatty acid synthase system, FAE- fatty acid elongase system. n, m indicate varying chain lengths.

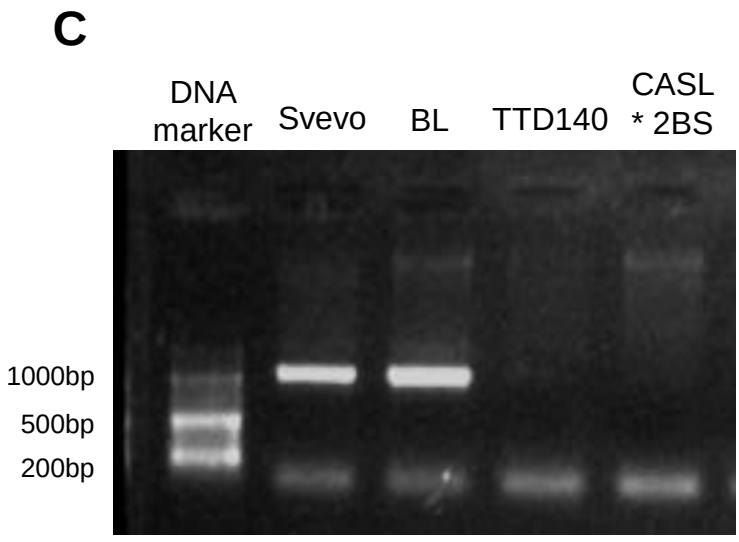


Supplemental Figure 2. Illustration of wheat parental and chromosome arm substitution lines used in this study. Wheat substitution lines used in this study were produced by Millet et al. (2013). A series of wild emmer [*Triticum turgidum* ssp. *dicoccoides* (line TTD140)] chromosome arm substitution lines (CASLs) were produced in the background of the modern common wheat [*Triticum aestivum* ssp. *aestivum* cv. Bethlehem (BL)]. In CASL* 2BS, the whole 2BS arm from TTD140 was inserted to the BL genome background. The wax phenotype is indicated for each line. Grey and green chromosome colors indicates BL and TTD140-derived chromosomes, correspondingly.



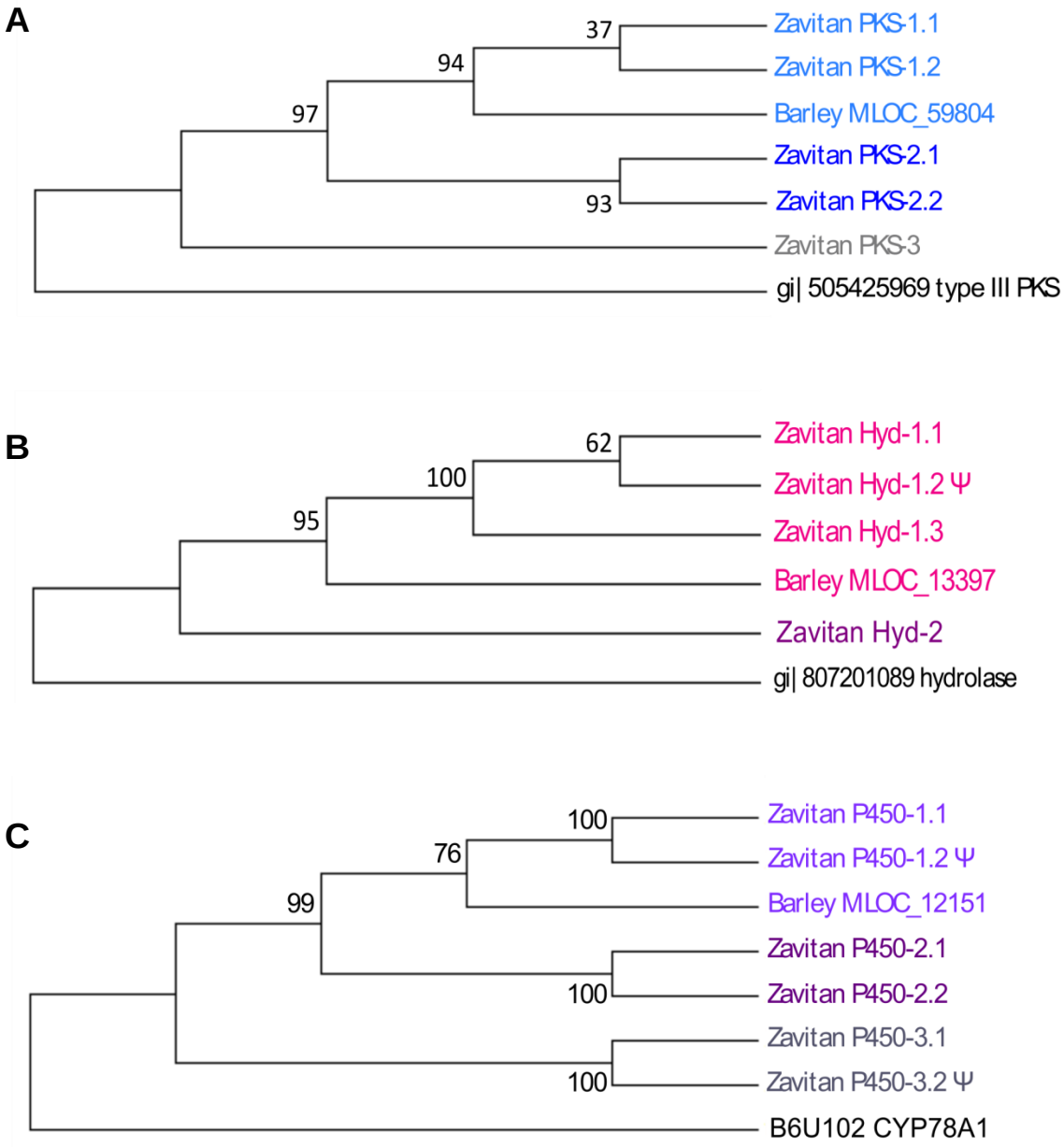
TaDMP marker (563bp)

TaDMH marker (784bp)

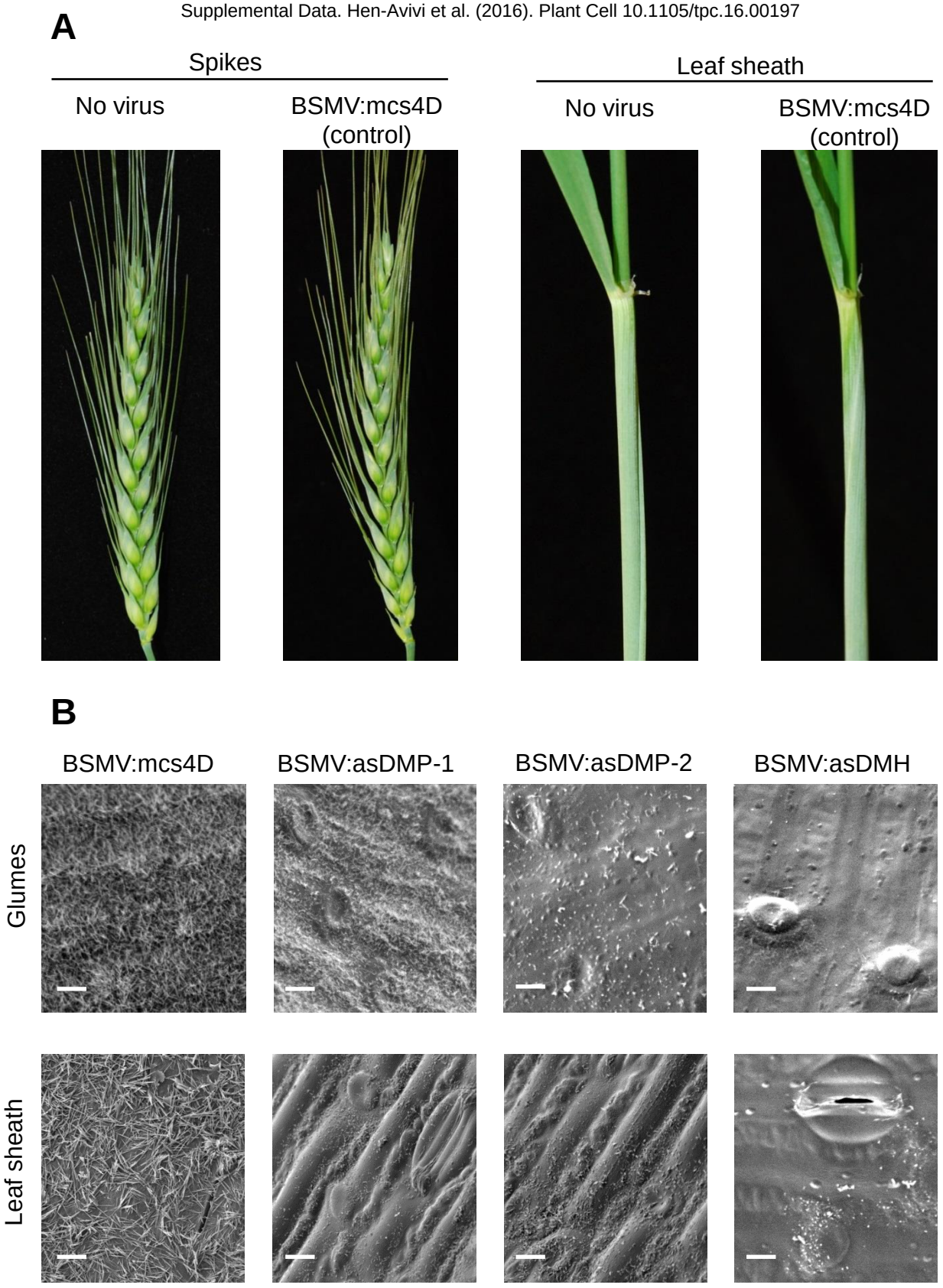


TaDMC marker *(*1057bp)

Supplemental Figure 3. Amplification of markers developed for the three candidate genes in glaucous and glossy wheat. The β-diketone biosynthesis candidate genes are present in the glaucous parental line 'Svevo‘ however absent in glossy parental line 'TTD140‘. Specific markers were developed for *TaDMP* **(A)**, *TaDMH* **(B)** and *TaDMC* **(C)**. The marker developed for the *TaDMC* gene was also polymorphic in the hexaploid population while the other two were not, apparently due to homologues in the D genome.

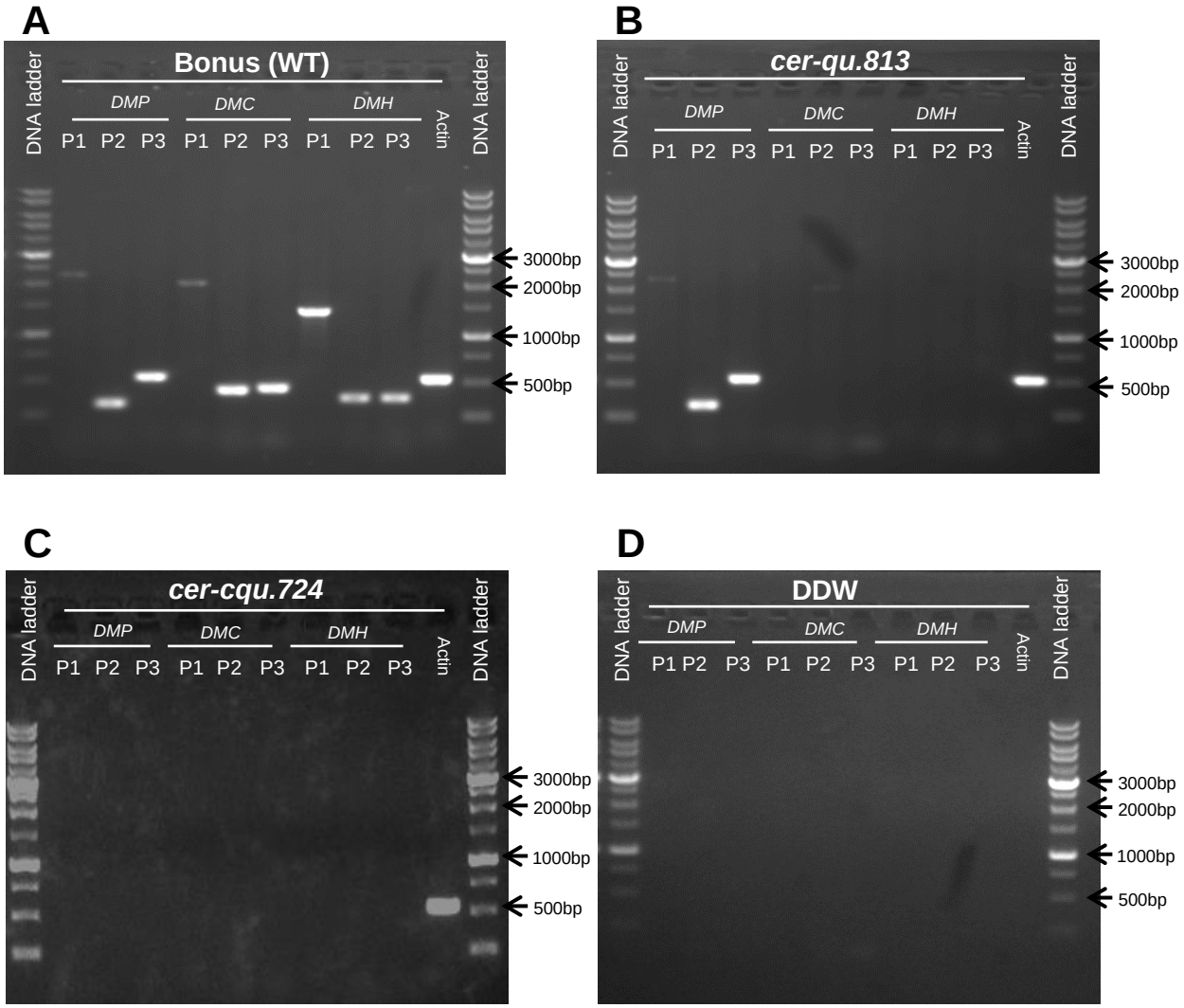


Supplemental Figure 4. Phylogenetic trees of Zavitan proteins. Phylogenetic trees describe the similarity between Zavitan and barley β-diketone proteins. **(A)** PKS, **(B)** Hydrolase (HYD) and **(C)** P450 phylogenetic trees. Sequences for the Zavitan proteins can be found in Supplemental File 1. Outgroup protein accessions are: For PKS tree: gi|505425969 (type III PKS from *Streptomyces fulvissimus*), for HYD: gi|807201089 (adenosylhomocysteinase from *Solanum lycopersicum*), for P450: B6U102 (CYP78A1 from *Zea mays*). Ψ- Premature termination codon (PTC). Closely related proteins are denoted with the same color.



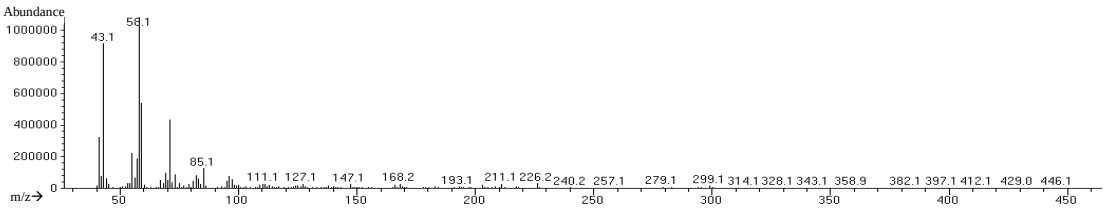
Supplemental Fig. 5

Supplemental Figure 5. Virus-induced gene silencing (VIGS) of the β-diketone genes reduces wax accumulation and deposition of epicuticular wax crystals. (A) Control infection with BSMV:mcs4D does not reduce glaucousness. Non-infected plants (no virus) and BSMV:mcs4D-infected (control) plants both have glaucous phenotype in glumes and leaf sheath **(B)** Epicuticular wax on surfaces of flag leaf sheath and glumes tissues was examined using scanning electron microscopy (SEM). Plants infected with constructs targeting the β-diketone biosynthesis gene transcripts (BSMV:*asDMP-1*, BSMV:*asDMP-2* and BSMV:*asDMH*) displayed reduced wax amount as well as reduced deposition of tubular wax crystals, typical for β-diketone-rich wax. Magnification: X500. Scale bar indicates 2 µm.

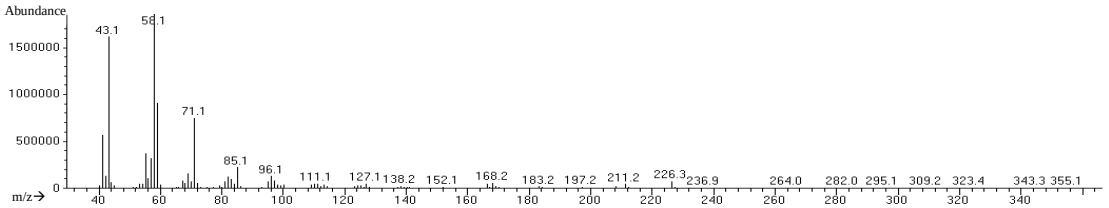


Supplemental Figure 6. The barley *cer-qu.813* double and *cer.cqu-724* triple mutants have deletions that include the metabolic gene cluster genes. Amplifications of the three candidate genes, *DMP*, *DMH* and *DMC,* were performed using three different primer pairs (P1- P3) per gene and genomic DNA template. Actin oligonucleotides were used as positive control, DDW as negative control. Agarose gels are shown for reactions using the following templates: **(A)** bonus (WT), **(B)** *cer-qu.813,* **(C)** *cer-cqu.724,* **(D)** DDW control. An unspecific band appears in the reaction mixture of *DMC* P2 primers and *cer-qu.813* gDNA, however has different product size from the corresponding reaction with WT template.

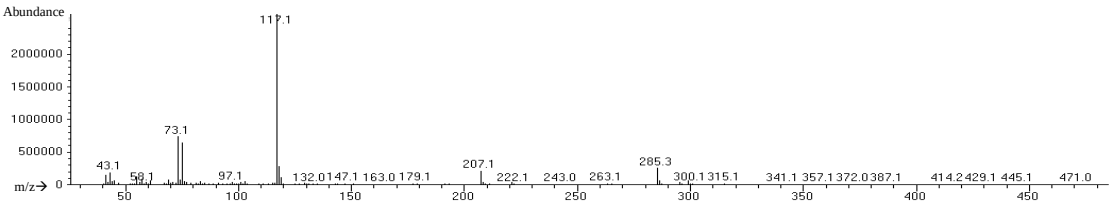
A Peak at 7.5 min: pentadecan-2-one (C_{15} 2-ketone)



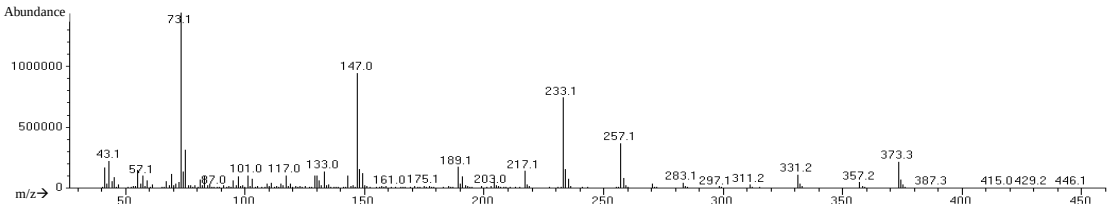
B Peak at 7.5 min: pentadecan-2-one (C_{15} 2-ketone) standard



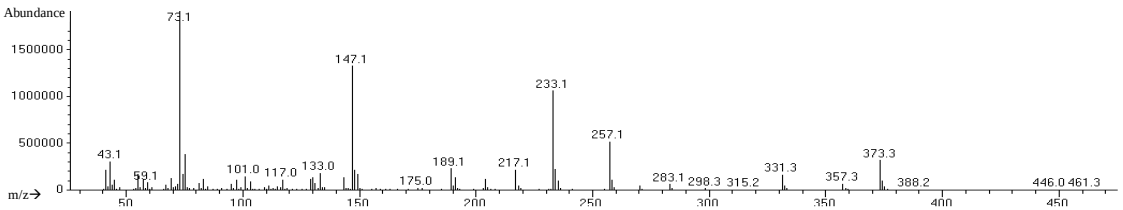
C Peak at 8.4 min: pentadecan-2-ol (C_{15} 2-alkanol)



D Peak at 10.8 min: 3-OH tetradecanoic acid (C_{14} 3-hydroxy-acid)



E Peak at 10.8 min: 3-OH tetradecanoic acid (C_{14} 3-hydroxy-acid) standard



Supplemental Figure 7. Compound identification of *E-coli* expressing *DMH* metabolites. Gas chromatography mass spectrometry (GC-MS) spectra of increased compounds in lipids extracted from *E-coli* expressing DMH from barley **(A, C-D)** and commercial standards **(B, E)**. **(A)** C_{15} 2-ketone, RT 7.5 min **(B)** C_{15} 2-ketone standard (Sigma, cat. W372404), RT 7.5 min. **(C)** C_{15} 2-alkanol, RT 8.4 min. **(D)** C_{14} 3-hydroxy-acid, RT 10.8 min. **(E)** C_{14} 3-hydroxy-acid standard (Sigma, cat. H4148), RT 10.8 min.

Supplemental Text 1: Fine genetic map of 2BS *W1/ Iw1* loci in different segregating populations

Kofa+*Lr19* (glaucous)x AUS2499 (glossy):

Out of 352 F_2 progeny screened we observed 271 individuals with a glaucous phenotype and 81 individuals with a glossy phenotype. This correlates well with the expected 3:1 segregation ratio for a single locus ($\chi2=0.74$, df=1, P=0.3889). Using genetic markers *W1_1*, *W1_2*, *W1_3* and *W1_4* (corresponding to markers *JIC005*, *JIC007*, *JIC009* and *JIC010*, respectively) we have mapped this single locus to the short arm of chromosome 2B between markers *W1_3* and *W1_4*; two recombinants were identified between markers *W1_3* and *W1_4*, giving a genetic distance of 0.57 cM. We then proceeded to screen another 4,423 F_2 progeny using markers *W1_3* and *W1_4*, giving a total of 4,775 F_2 plants screened (9,550 gametes) out of which 15 plants were shown to be recombinant between markers *W1_3* and *W1_4*. This represents a genetic distance of 0.16 cM. By selfing the 15 recombinants we obtained homozygous recombinants and verified both the mapping data as well as the phenotype. These results suggest that the dominant locus for β-diketone production designated *W1* is confined in a 0.16 cM interval flanked by markers *W1_3* and *W1_4*.

Svevo (glaucous) x TTD140 (glossy)

In this population *Iw1* derived from TTD140 segregates. Out of 299 F_2 progeny screened we observed 77 individuals with a glaucous phenotype and 222 individuals with a glossy phenotype, fitting a single locus 1:3 Mendelian segregation ratio ($\chi2=0.09$, df=1, P=0.7674). Using genetic markers *Xwmc661*, *W1_3* and *W1_4* we have mapped this single locus to the short arm of chromosome 2B completely linked to *W1_3*; seven recombinants were identified between markers *Xgwm614* and *W1_4*, giving a genetic distance of 1.17 cM. By selfing the 7 recombinants we obtained homozygous recombinants and verified both the mapping data as well as the phenotype.

Supplemental Text 2: Analysis of the Zavitan *W1/lw1* interval

Using the wild emmer Zavitan genome sequence (unpublished) we analysed the genes within the interval between genetic markers *JIC036* and *JIC032* which include the *W1* locus and the three β-diketones candidate genes (see Figure 2C). We predicted 22 genes in this region, 15 out of the 22 were annotated similarly to the β-diketones candidate genes; six *CYP450* family *(P450)* genes, five *Type III polyketide synthase (PKS)* family members and four members of the *hydrolase/lipase (Hydrolase; HYD)* family.

Among the five *PKS* genes there are three distinct genes, two of which have been duplicated in tandem. One of the tandem duplicated genes has not been fully captured in the Zavitan genome sequence and only the first exon is present, but the gene is likely intact. The similarity between the tandemly duplicated genes is ~99.5%, indicating a relatively recent duplication event. Comparing the coding sequence (CDS) of these five genes against that of barley *MLOC_59804* shows that one of the duplicated gene pairs (*PKS-1.1* and *PKS-1.2*) is more closely related than the other three genes, indicating that these are the likely orthologues. However, *MLOC_59804* is also the best hit for *PKS-2.1* and *PKS-2.2*, suggesting that either their barley orthologue is missing from the database or that these are older duplicated versions of *PKS-1.1* and *PKS-1.2* (Supplemental Table 3 and Supplemental Figure 4A).

Among the four *HYD* genes there are two distinct genes, one of which has been duplicated twice (*HYD-1*). The similarity between these three genes is >99.6% indicating relatively recent duplication events. Comparing the CDS of these four genes against that of the barley *MLOC_13397* showed that the triplicated gene pair (*HYD-1*) is more closely related than the other gene (*HYD-2*), indicating that the triplicated genes are the likely orthologues (Supplemental Table 3 and Supplemental Figure 4B). Interestingly, *HYD-1.2* has a single base pair deletion within its CDS, causing a frameshift and premature termination codon (PTC), suggesting that *HYD-1.1* and *HYD-1.3* are the functional orthologues of *MLOC_13397*.

Among the six P450 genes, three distinct genes are represented, all of which have been duplicated in tandem. The similarity between the tandemly duplicated genes

is >99.3% indicating relatively recent duplication events. Comparison of the CDS of these six genes against that of *MLOC_12151* showed that one of the duplicated gene pairs (*P450-1.1* and *P450-1.2*) is more closely related than the other four genes, indicating that genes in this pair are the likely the orthologues (Supplemental Table 3 and Supplemental Figure 4C). Similar to the HYD example, *P450-1.2* has a single base pair deletion within its CDS, causing a frameshift leading to a PTC. This suggests that *P450-1.1* is the only functional orthologue of *MLOC_12151*. The orthologous loci of the other P450 genes appear to be *MLOC_13649* and *MLOC_71974*. Notably, *P450-3.2* also has a PTC.

The pattern of gene duplications within the Zavitan interval has occurred after the divergence from barley based on the high sequence similarity of the duplicated genes (>99.3 % in all cases). Moreover, the position of the genes within the interval provides clues as to how those duplications occurred. The arrangement of the genes within the interval precludes a large scale duplication event and is most consistent with a series of independent small scale local duplication events.

One duplication consists of *PKS-1.1* and *PKS-1.2,* the two genes most closely related to *MLOC_59804*. Another block includes the duplication of *P450-1.1*, *HYD-1.1* and *Hly-1.1* into a second block (*P450-1.2*, *HYD-1.3* and *Hly-1.2*) which was followed by the separate duplication of *HYD-1.1* into *HYD-1.2*. A third duplication block within this interval encompasses *PKS-2*, *P450-2* and *P450-3* (see Figure 2C).

In addition to β-diketones candidate genes, two more relevant genes were found to be located in the *W1/ Iw1* Zavitan interval: wax ester synthase (*WES*) and *CYP96B30 (P450-3.1)*. The genes annotation as wax related and their genomic localization suggests that they might also be involved in β-diketones or related metabolites biosynthesis.

Supplemental Text 3: Comparison of the *W1*/*Iw1* interval between the wild emmer accessions TTD140 and Zavitan

Genetic markers *JIC007*/*IWB7407* (distal) and *JIC010*/*CD927782* (proximal) (see Figure 2C) which are flanking both the *W1* and *Iw1* interval were used to anchor genomic sequence of Zavitan (glaucous) and TTD140 (glossy). The distal TTD140 and Zavitan sequences are 97.6% similar over ~8 kb of intergenic, exon and intron sequence. Proximal to the *W1/Iw1* interval the similarity is only 93% over ~18 kb of intergenic, exon and intron sequence, but this rises to 96% if corrected for gaps in the alignment. (Figure 2C, Supplemental Table 3). However, the interval between these highly similar regions is completely different between TTD140 and Zavitan, both for the intergenic space and also for the gene content, with the exception of two NB-Arc-like genes (NB-Arc-like_1 and _2; Figure 2C). These two genes are located at the distal border of the conserved interval in Zavitan and TTD140. However, in the latter the NB-Arc like genes have undergone a series of tandem duplication events; the available sequence of TTD140 suggests at least 13 copies, but there are likely more. The duplications in TTD140 are not restricted to the open reading frames themselves, but also include intergenic sequence and transposable elements.

None of the other genes identified in Zavitan can be found in the TTD140 physical map covering the *W1*/*Iw1* interval. This could be attributed to the incomplete state of the TTD140 physical map. But the reverse is also true, i.e. none of the genes identified in TTD140 can be found in the Zavitan whole genome assembly. Interestingly, the TTD140 physical map contains two P450 and two PKS genes, but these have low similarity (~85 % in the nucleotide level) to the *P450* and *PKS* genes identified in the Zavitan interval (see Supplemental Table 3). In summary, at least 808,361 base pairs of TTD140 sequence (excluding the region of tandem duplicated NB-Arc-like genes) are completely different in the Zavitan whole genome assembly, suggesting a divergence in haplotype between these two wild emmer accessions across the *Iw1*/*W1* interval.

The haplotype divergence is supported by the observation that no recombinant plants between markers *JIC036* and *JIC032*, the two markers closest to the interval, were observed in F_2 progeny of a cross between three different mapping populations in

which TT140 is a parent or the 2BS is derived from TTD140 (CASL*2BS). The studied populations were TTD140 and *T. durum* cultivar Langdon (glaucous; *iw1*/*iw1*; *W1*/*W1*), TTD140 and *T. durum* cultivar Svevo and CASL*2BS and *T. aestivum* cultivar BL. Based on the disparity in sequence content between TTD140 and Zavitan we hypothesise that Langdon and TTD140 also have equivalent sequence disparity, i.e. different haplotypes between markers *JIC036* and *JIC032*. Recombination numbers from Langdon X TTD140 population are given in purple in Figure 2C.

Supplemental Table 1: Linked deleted markers in the TTD140 accession detected by the 90K chip analysis

Number	SNP name	Reside inside β-diketone gene	SNP ID	Marker sequence	Index on iSelect 90K SNP bead chip
1	wsnp_Ra_rep_c106727_90434958	*DMP*	IWA8128	AGGTTCCATGGTGTCTGCTACCAGTGTGGGCAGTTTGGTGTAAACTTTCCCATGAGTGC CATTGCTCCCCAAGTTCAGGGTGAGCACATTCTCAGTGCTC[A/G]GCACGAGGGTCTGA GAGACGGACACCATCTCAAACAGAGGATTCTCGATGGGGTGCATGGGGTTGGCGCCG ACGATTACCGCGCCTGCCCCATCACCAAA	81290
2	RAC875_rep_c109471_154	*DMP*	IWB61884	CCCACCTTATCGTCAGCACCAACTCTGACACTGGCGCCCCAAGCGCTGAC[A/G]TACGC TTGGTTTCACTCCTTGGCCTTCGCGCCGATGTCTGTCATACCATG	61884
3	RAC875_rep_c115433_378	*DMP*	IWB62322	CAAACTGCCCACACTGGTAGCAGACACCATGGAACCTTGTCTTCTGGAAG[T/C]GTTTG GTCCACTTGAGATGGACTTCCAAT	62322
4	wsnp_Ex_c851_1654297	*DMC*	IWA4808	GCAAACGGAKACGACTGGAAGCGGCGSCGCAAATTCATCCACCCGGCTTTCAGCCAAG AGAAGATCAAGTCCATGTCAGCAATAACGTTGGAGTGCACAC[A/G]GCAGATGATGGAA CAGTGGCGCACTCAAATGCAGGAGAGCAACATGCAGCAAGCCGAGATCGACATGAGG TACGACTCCGATGACATAGCAATGCGTGTC	78868
5	wsnp_Ra_c1660_3275687	*DMC*	IWA7656	TTGATGATGGTCGCTATCACAATCTGCACCTCAACCATGGCGAAGTTTTGCCCGGCACA AACCCTCGGCCCGAATGAGAAGGCCAGCAGCGCGTGTGAAT[A/G]CTTGGCGGCTCTC GATAAGCCATTCTGGAACCTCATCGGATTAAACTCGTCGGCGTCGGGTCCCCAAATCTC CTTGTCCCGGTGCAACAAAACTAATGGT	80948
6	Ra_c2544_1566	*DMC*	IWB51601	CTCTTGGCTGAAAGCCGGGTGGATGAATTTGCGCCGCCGCTTCCAGTCGT[A/C]TCCGT TTGCAAATATTACCCCGTTCCCTAGAATTGCTTCCAAACTGGGAT	51601
7	RAC875_c9013_185	*DMC*	IWB60992	AAGACTGCCTCAGACACTATCCTAACGCACGTGAAGGTGCCGAAAGGAAC[A/G]ATGAT AACAATACCATTAGTTTTGTTGCACCGGGACAAGGAGATTTGGGG	60992
8	BS00084667_51		IWB11378	GTCACCACTCTGCCGCCAGCTGAGGCCACTAAACCAACATTCATTGGTGT[T/C]GAAGG ATTTGAGGGCGACTATGGTGGAATTGAGTTTGGTAACGAGGAAGC	11378
9	Ra_c491_902		IWB52168	ATACCATTAGTTTTGTTGCACCGGGACAAGGAGATTTGGGG	52168
10	wsnp_Ex_c1996_3754394		IWA2482	CGTGCGATGATCATTGGCTACATGTTCGCTGCTAGGGATACAGTTGGAACAACCCTGA CATGGATCTTCTACAAACTCGCCCAGAACCCTAACATTGTTT[T/C]GAATATCCGAAAAG AACTCTCACCCATTGCATCACGCAAAGCAGCGGTTGGTGTGGATGCCATGTTGATCTTT GAGCCAAAAGAGACCAGATCTCTAGTA	77079
11	Ra_c11464_294		IWB50940	GCATCAACGCTATGGAACTTCGATGTGGAGGTGATGGACGGGCAAACAAT[T/C]CAGCC CAAGCAAGCTTGTATACAGCAGATGAAAAATGGGCTCATAGTTAA	50940
12	BS00023068_51		IWB7407	TACTTGCTTTGCACTGGCATTCTTAGTAGTAGTTCCCCGAACGTTGACTG[T/C]AAATATG CGAGGATTTCCAACTTGCTGGTTGCATCACCAGGACAGTAGTT	7407

Supplemental Table 2: Markers and oligonucleotides used for linkage analysis

Kofa+*Lr19* x AUS2499 population

	Marker	Kofa+*Lr19* allele	AUS2499 allele	Common oligonucleotide	corresponds to
1	W1_1	TAACGCCGCTCCTCTCGT	TAACGCCGCTCCTCTCCTT	TCAGTTGGGTATTTTCTCTTCCAG	JIC005
2	W1_2	ACTTCAAATGTGATGTGCTTGCTT	ACTTCAAATGTGATGTGCTTGCTC	CGACAGTCGGATTCTTTCTTACC	JIC007
3	W1_3	CAGAAGACTGCACACACATTACTTG	CAGAAGACTGCACACACATTACTTT	CCAGTTAAGTGAAAACAGTGATCT	JIC009
4	W1_4	CGTGAGCTGCTAAACAAAGGTG	CGTGAGCTGCTAAACAAAGGTA	AGAGAAGTGCCTTGAGACATATTTAC	JIC010

Svevo x TTD140:

	Marker	Forward oligonucleotide (5'-3')	Reverse oligonucleotide (5'-3')	Tm (°C)	Type	Amplicon length (bp)
1	*Xwmc764*	CCTCGAACCTGAAGCTCTGA	TTCGCAAGGACTCCGTAACA	61	SSR	
2	*RAC875_c38665_247 (IWB57438)*	TGCTGAGTGCAAAGCATTGC	TGGGTCATAAGTTGAGCATGCAGTT	59	SNP (HRM)	
3	*Xwmc661*	CCACCATGGTGCTAATAGTGTC	AGCTCGTAACGTAATGCAACTG	60.8	SSR	
4	*Xgwm614*	GATCACATGCATGCGTCATG	TTTTACCGTTCCGGCCTT	60	SSR	
5	*BS00027693_51 (IWB7661)*	GCTGTGTAACCTGAAGACCATC	TCATCATGCACCATCAGTCA	59	SNP (HRM)	
6	*CD893659 (JIC009)*	CAAATTAAGGCAAACAATAACACGA	GCTAGGGTTATTACTGTACACC	58	PCR	723bp
7	*TaDMC gene*[A]	CTGACCCATCAAAGGCAGTT	CTTCAGTTGCTCCCTTCCTG	68	PCR	1057bp
8	*TaDMH gene*[A]	CCCGTGCCGACGTCCTACGAT	TGCGAGCTTGCTTCAAAGGTA	65	PCR	784bp
9	*TaDMP gene*[A]	GGTTACTGCACCGAGACCAT	TGAAACCAAGCGTACGTCAG	67.5	PCR	563bp
10	*CD927782 (JIC0010,IWB43910)*	CTGTGTAACCTGAAGACCATCACC	GCACCATCAGTCACATGCTTA	60	SNP assay	
11	*CD927782 (JIC0010,IWB43910)* probes	allele one- CAACGATGCCCGTCT	allele two-CAACGATGCTCGTCTTG	60	SNP assay	

[A] Domimnat in Svevo, null in TTD140

Supplemental table 3: Homology between Zavitan, TTD140 and the corresponding barley genes

TTD140 vs Zavitan flanking sequence similarity

	Intergenic/exon/intron	
	Identities	**gaps**
Distal side	7827/8019 (97.6%)	72/8019 (0%)
Proximal side	17008/18287 (93%)	475/18287 (3%)

TTD140 vs Zavitan gene similarity

	PKS_1.1		PKS_1.2		PKS_2.1		PKS_2.2		PKS_3	
	Identities	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**
TTD140_CHS-1	no alignment		no alignment		no alignment		no alignment		no alignment	
TTD140_CHS-2	969/1192 (81%)	26/1192 (2%)	966/1187 (81%)	26/1187 (2%)	966/1192 (81%)	26/1192 (2%)	116/135 (86%)	0/135 (0%)	768/1167 (66%)	59/1167 (5%)

	P450_1.1		P450_1.2		P450_2.1		P450_2.2		P450_3.1		P450_3.2	
	Identities	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**
TTD140_P450-1	1326/1556 (85%)	9/1556 (0.58%)	1322/1556 (85%)	10/1556 (0.64%)	957/1117 (86%)	30/1117 (3%)	955/1117 (86%)	30/1117 (3%)	no alignment		no alignment	
TTD140_P450-2	no alignment		no alignment		no alignment		no alignment		no alignment		no alignment	

Barley vs Zavitan gene similarity

	PKS_1.1		PKS_1.2		PKS_2.1		PKS_2.2		PKS_3	
	Identities	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**
MLOC_59804/AK375690	1107/1181(94%)	0/1181(0%)	1107/1182(94%)	0/1181(0%)	1084/1181(92%)	0/1181(0%)	131/147(89%)	0/147(0%)	785/1154(68%)	33/1154(2%)
1182 bp										

	P450_1.1		P450_1.2		P450_2.1		P450_2.2		P450_3.1		P450_3.2	
	Identities	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**
MLOC_12151/AK373499	1444/1545(93%)	0/1545(0%)	1440/1545(93%)	1/1545(0.006%)	947/1116(85%)	21/1116(1%)	945/1116(85%)	21/1116(1%)	no alignment		no alignment	
1548 bp												

	HYD_1.1		HYD_1.2		HYD_1.3		HYD_2	
	Identities	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**	**Identities**	**gaps**
MLOC_13397	1110/1235(90%)	7/1235(0.056%)	1108/1235(90%)	8/1235(0.064%)	1112/1235(90%)	7/1235(0.056%)	717/993(72%)	36/993(3%)
1230 bp								

Supplemental Table 4: Oligonucleotides used in this study

no.	primer_Fwd name	primer Fwd seq 5'-3'	primer_Rev name	primer Rev seq 5'-3'	Figure
1	TaPKS_RT_F1	GTTGGAGCCCACGAAGCTAG	TaPKS_RT_R1	GACGCCGTAGCTCATCAAGC	Figure 3
2	TaHyd_RT_F1	GACGAATTCTACCCATTAATCC	TaHyd_RT_R1	GGACACACCGGTGCCCTG	
3	TaP450_RT_F1	GGCAAATTGAAGCTGCTCAAC	TaP450_RT_R1	TTCGGCACCTTCACATGTG	
4	ACT2_RT_F	CAAATCATGTTTGAGACCTTCAATG[1]	ACT2_RT_R	ACCAGAATCCAACACGATACCTG[1]	
5	PKS_VIGS_F1	AACCACCACCACCGTGGTTACTGCACCGAGACC[2]	PKS_VIGS_R1	AAGGAAGTTTAACTTCTTTGCTGCTGCTG[2]	Figure 4
6	PKS_VIGS_F2	AACCACCACCACCGTCACACTGGTAGCAGACACC[2]	PKS_VIGS_R2	AAGGAAGTTTAAGCTGCGTACATAGCTTAACG[2]	
7	Hyd_VIGS_F1	AACCACCACCACCGTGTGAGGACCATGGCTTC[2]	Hyd_VIGS_R1	AAGGAAGTTTAACCTCATCACGAATCTTACC[2]	
8	TaPKS_RT_F2	CCATCTTAACGGCTGCTTC	TaPKS_RT_R2	CTGCAGAAGCCAGAAATG	
9	TaHyd_RT_F2	ATGACAGCATCATGGACATC	TaHyd_RT_R2	CACTGGTGGGAACATGG	
10	CDC48_RT_F	GTCCTCCTGGCTGTGGTAAAAC[3]	CDC48_RT_R	AGCAGCTCAGGTCCCTTGATAC[3]	
11	MLOC_59804_F1	CCATAATTGTTTAGTCTGCTGC	MLOC_59804_R1	AATCAAAATAACATTTGGGTGAG	Figure 5, Supp. Figure 6
12	MLOC_59804_F2	CAGTCCATAATTGTTTAGTCTGC	MLOC_59804_R2	CTACAAAGTGTAAGGTGTGTGC	
13	MLOC_59804_F3	CGTTGACCCACATTTTCAG	MLOC_59804_R3	GGAGTTGAGCAGTTCATCC	
14	MLOC_13397_F1	TCGTCAACATATCCCGTAG	MLOC_13397_R1	ACTCCAGACGTGATCATAACC	
15	MLOC_13397_F2	GGTTGGTCATGGTGCATC	MLOC_13397_R2	GCTCTTCATAAGCTTCTTGCTG	
16	MLOC_13397_F3	CCATGGTGATCCCAGATG	MLOC_13397_R3	CAATGTTCCATTGCAATACC	
17	MLOC_12151_F1	GAAAAACATTCCTGTACTGGTTG	MLOC_12151_R1	CAGTATTACTCAAGTCTCTTCTGAAC	
18	MLOC_12151_F2	AACATTCCTGTACTGGTTGG	MLOC_12151_R2	GAGTCGTGCATCATGTCG	
19	MLOC_12151_F3	GGAAGGATGTTAAAGCTGCTC	MLOC_12151_R3	GAACATGGCCCAGGTAATAAG	
20	MLOC_59804_RT_F	TGTCCCAAGAAGAGTACCCTGAC	MLOC_59804_RT_R	GCCCGTGAGACCACATATTATCT	Figure 5
21	MLOC_13397_RT_F	CCAAACATAATATAGATCTGACAGC	MLOC_13397_RT_R	GGTGACGGGTGACCTACG	
22	MLOC_12151_RT_F	TCCTGTACTGGTTGGGACC	MLOC_12151_RT_R	TTGGAACATGTCTGTCCTG	
23	HvActin RT_F	GCCGTGCTTTCCCTCTATG[4]	HvActin RT_R	GCTTCTCCTTGATGTCCCTTA[5]	
24	Hyd_RF_F	TCCGCGGGTGAAAACCTGTACTTCCAGGGT CCTGCAAACAAGACTTACCCCTC[5]	Hyd_RF_F_R	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTA GAAACAGTTGTTCATCATGGATC[5]	Figure 6

[1]Wheat actin gene [TC234027] (Tenea, G.N., et al., Reference genes for gene expression studies in wheat flag leaves grown under different farming conditions. BMC Res Notes, 2011. 4: p. 373.)

[2]Sequences complementary to the target gene cDNA sequence are underlined. Ligation independent cloning (LIC) adaptor sequences are shown in bold.

[3]Wheat cell division cycle 48 [Unigene Ta.46201]

[4]Barley actin gene [GenBank: AY145451] (Kapazoglou, A., et al., Epigenetic chromatin modifiers in barley: IV. The study of barley polycomb group (PcG) genes during seed development and in response to external ABA. BMC Plant Biol, 2010. 10: p. 73.)

[5]Sequences complementary to the target gene cDNA sequence are underlined. Restriction free (RF) sequences adaptors from pET28.TEV are shown in bold.

Supplemental File 1: Gene Sequences of β-diketone biosynthesis and those annotated in the genome of the Zavitan accession.

β-diketone genes from *T. aestivum* cultivar Bethlehem (BL):

>TaDMP

ATGGCAGGCAGCTCACCGAAGGTTAGTGAGATCCGGTGTGCGCAGCGTGCGGAAGGCTCCGCGGCAATGC
TGGCTATCGGAACAGCAAATCCGGCGAACAAGGTGTCCCAAGAAGAGTACCCTGACTATTATTTCCGCGT
TACCAAGAGCGAGCACCTTACTGACCATAAAGACACATTCAAGATAATATGTGGTCTAACGGGCACGGAG
AATCGTTTCTTCTACCACACGGATGAACTGCTCAACTCCCACCCTGTCTTGCTGGACAACACGTCACCGT
CCCGTGAGGCTCGGCATGATATCGTGGCCAAGGCTGCTCCAGAGCTTGCGGCAGCAGCAGCAAAGAAGGC
CATCGCAAAGTGGGGCCGTCCGGCCAGTGACATCACCCACCTTATCGTCAGCACCAACTCTGACGCTGGC
GCCCCAAGCGCTGACGTACGCTTGGTTTCACTCCTTGGCCTTCGCGCCGATGTCTGTCGTACCATGCTCC
ATCTTAACGGCTGCTTCGCCGGCTGCTCCGCCTTGCGCCTAGCCAAGGACCTTGCTGAGAATAACTGTGG
GGCACGCGTCCTCGTAGCTTGCGTGGAGCTCGCCATTTCTGGCTTCTGCAGCCCCGGCGAAGGGGACTGC
TTAGACACCCTCATCACTCATGCGTTGTTTGGTGATGGGGCAGGCGCGGTAATCGTCGGCGCCGACCCCA
TGCACCCCATCGAGAATCCTCTGTTTGAGATGGTGTCCGTCTCTCAGACCCTCGTGCCGAGCACTGAGAA
TGTGCTCACCCTGAACTTGGGGAGCAATGGCACTCATGGGAAAGTTTACACCAAACTGCCCACACTGGTA
GCAGACACCATGGAACCTTGTCTTCTGGAAGCGTTTGGTCCACTTGAGATGGACTTCCAATGGAATGACC
TCTTCTGGGCAGTGCACCCTGGCAGCCGTGGGATCTTGGACCAACTTGACAAGACACTCCAGTTGGAGCC
CACGAAGCTAGCGGCGAGCAGAACTGTCGTACAGAAGTTTGGGAACATGTTTAGCGCCACCGTGATCTTC
GTGCTTGATGAGCTACAGCGTCGAATGGAGGAGGAAGGAGAGCAAGCTGAGTGGGGGGCCATGGTGGGAT
TTGGACCAGGCTTCACTATCGAGACCATGGTGCTCCATGCAACCGGCGCTCTCAAGAAAAAT**TAG**

> TaDMP_protein

MAGSSPKVSEIRCAQRAEGSAAMLAIGTANPANKVSQEEYPDYYFRVTKSEHLTDHKDTFKIICGLTGTE
NRFFYHTDELLNSHPVLLDNTSPSREARHDIVAKAAPELAAAAAKKAIAKWGRPASDITHLIVSTNSDAG
APSADVRLVSLLGLRADVCRTMLHLNGCFAGCSALRLAKDLAENNCGARVLVACVELAISGFCSPGEGDC
LDTLITHALFGDGAGAVIVGADPMHPIENPLFEMVSVSQTLVPSTENVLTLNLGSNGTHGKVYTKLPTLV
ADTMEPCLLEAFGPLEMDFQWNDLFWAVHPGSRGILDQLDKTLQLEPTKLAASRTVVQKFGNMFSATVIF
VLDELQRRMEEEGEQAEWGAMVGFGPGFTIETMVLHATGALKKK*

>TaDMH

ATGCCTGCAAACAAGACTTACCCCTCCCATAAGAATGCCAACGGTGAGGTGGACGATGAATTCTACCCAT
TAATCCGCAAGTACAAGGATGGCCGGATCGAGCGGTTCATGAGCTCATTCGTGCCGGCGTCGGAGGACCC
GGCCGCCAGTCGTGGTGTGGTGACGAGGGACGTCGTCATCGACCAGGGCACCGGCGTGTCCGTGCGCCTG
TTCCTTCCTGTCCAGGCTGCCCAGGCTGGCACGAGGCTCCCCCTTGTTGTGTACGTCCATGGTGGTTCCT
TTTGCACGGAGAGTGCCTTCTCCCGGACGTACCACCGTTACGCCACTTCCCTCGCCGCCAGCGCTGGGGC
GCTCATCGTGTCCGTGGAGTACCGTCTGGCGCCGGAATATCCCGTGCCAACGTCCTACGATGACACATGG
GCCGCGCTGCGGTGGGTGGCGTCCTTGTCCGACCCTTGGCTCGCCAAGTACGCAGACCCTAGCCGCACAT
TCCTCGCCGGCGACAGCGCCGGCGGCAACATCGTGTACCACACGGCCGTGCGCGCCACACGTGATGACAG
CATCATGGACATCCAGGGGTTGGTCATGGTGCATCCATTCTTCTGGGGCCCCGAGCGTCTCCCGGCGGAG
AAGGTTTTGGACGGCGACGCCATGTTCCCGCCAGTGTGGGTGGATAAGCTGTGGCCCTTCGTGACGGCGG
GCGGGGCTGGCAACGATGATCCTCGGATCAATCCTCCGGACGAGGAGATCGCGTTGCTAACTGGCAGGCG

GGTGCTTGTGGCCATTGCAGAGAAGGACACCCTGCGCGACCGGGGGCGCCAGTTTGTGTGCAGCATGCGC
GGGTGTGGGTGGGTTGATGGCAGCCTCACCGTGGTGGAGTCGGAGGGTGAGGACCATGGTTTCCATCTGT
ATGCTCCCCTACGTGCGACCAGCAAGAAGCTTATGAAGAGCATCGTGCAGTTCATAAACCATCGCGCCAC
CTTGCCGTCACCGGCCATGGTGATCCCAGGAGGCTCGGCCGAAACTATGCTAGGCGTCCCTAGTAGGCCA
TTTAAGGACATATTTGGCTACGGGATGCGCATGAAACGTTGGAGTGGCACGAGTTTTGGGCTCAAAGTTG
GTCGTGCAAAAGCATCGACGACGAGCTATGGGTTACGTTTGAAGCAAGCTCGTACCTTCGGAGACCCTGT
TTCAGCACCAACTTCGGTAAGATTCGCGATGAGGAACTGTTTCTAG

>TaDMH_protein
MPANKTYPSHKNANGEVDDEFYPLIRKYKDGRIERFMSSFVPASEDPAASRGVVTRDVVIDQGTGVSVRL
FLPVQAAQAGTRLPLVVYVHGGSFCTESAFSRTYHRYATSLAASAGALIVSVEYRLAPEYPVPTSYDDTW
AALRWVASLSDPWLAKYADPSRTFLAGDSAGGNIVYHTAVRATRDDSIMDIQGLVMVHPFFWGPERLPAE
KVLDGDAMFPPVWVDKLWPFVTAGGAGNDDPRINPPDEEIALLTGRRVLVAIAEKDTLRDRGRQFVCSMR
GCGWVDGSLTVVESEGEDHGFHLYAPLRATSKKLMKSIVQFINHRATLPSPAMVIPGGSAETMLGVPSRP
FKDIFGYGMRMKRWSGTSFGLKVGRAKASTTSYGLRLKQARTFGDPVSAPTSVRFAMRNCF*

>TaDMC
ATGGCGGCTAACGTAGTGCAAGCCCTGGCCGCTGTCCTCACACTGCTGGTGATCACAAGAGCGCTATGGT
ATCTGCTTTGGAGGCCATACGCTGTGGCCAGGTGGTTCGAGCAGCAGGGCATCAGAGGTCCGCCTTACAA
ATTCCTGGTTGGGTCCCTGCCCGACTGCCAAAGGATGCTCGTTGCCGGGAGAGTCAAGGATCTGGACACG
AGCTCCCACGACTGCATCACCACCGTGCAACCCTTCTTCCGGAAATGGGCCTCACTGAAAAACATTCCTG
TACTGGTTGGGACCGACACCAGCACTGTGTACTACAGACATAGAAATAGTGAAGAAAGTGTTAAGCGACA
GGACAGACATGTTCCAAAAAGATTACTTGAATCCCAGTTTGGAAGCAATTCTAGGGAACGGGGTAATATT
TGCAAACGGAGACGACTGGAAGCGGCGCCGCAAATTCATCCACCCGGCTTTCAGCCAAGAGAAGATCAAG
TCCATGTCAGCAATAACGTTGGAGTGCACACAGCAGATGATGGAACAGTGGCGCACTCAAATGCAGGAGA
GAACATGCAGCAAGCCGAGATCGACATGAGGTACGACTCCGATGACATAGCAATGCGTGTCATAGCACGA
GTGATGCTGGGCAAGAACTACAGGGAGGCCTGGGAAGTGTTCATGGCAGGAAGGGAGCAACTGAAGCTCG
CCGCGTATGCATTTGCGGATCCTCCAGTACCTGGATTCAGGTACCTGCCGACACGTCGCAACCGTCGGAC
CTGGCAACTCGACAAGCTTGTGAGAAGCAAGATTACAGAGATCATAAAGGCGCGGCTTGCTAGCAGTGTC
TATGGAGACGACCTGCTCGGGCAGATGTTGTGGCTGCAGAGGTCGGGTGCTGGCGCCAACGCCGAGAACC
TGAGCACGGAGGAGATGGTCGGCGAGTGCAGGACCTTCTTCATGGCTGGGTACGAAACCAGCGCCAACCT
CATTACCTGGGCCATGTTCCTGCTCGCCAGCCACCCACGGTGGCAGGAGATGGTCAGGGACGAGGTCGTC
CAGGAGTTTCCTGCTCACAAGCCACCCTTAGGTGACGGCCTCGGCAAATTGAAGCTGCTCAACATGTTAC
TCTGGGAGACATTGAGGCTCTATGGCCCCATATCATTCCTGCAGAGGAAGACAGCCTCAGACACAATCCT
CACACATGTGAAGGTGCCGAAAGGAACGATGATAACGATACCTCTGGTGATGTTGCACCGGGACAAAGAG
GTCTGGGGACCCGACGCCGACAAGTTTAACCCAATGAGGTTCCAGAATGGCTTCGCGAGAGCCGCCAAGC
ATTCACATGCACTGCTGGCCTTCTCGTATGGGCCTAGGGTCTGTGTCGGGCAGAACTTTGCCATGGTGGA
GGTGCAGATCGTCATAGCGACGATGCTCAAAAGTTTCTCCTTCTCCCTGTCCCCCACTTATGTGCACAAG
CCGAGCAATTTCGTCACATTGACGCCCAAGTACGGGCTCCCTCTCATCGTGAGGAACCTGCAGCTGACTA
GG**TGA**

>TaDMC_protein
MAANVVQALAAVLTLLVITRALWYLLWRPYAVARWFEQQGIRGPPYKFLVGSLPDCQRMLVAGRVKDLDT
SSHDCITTVQPFFRKWASLKNIPVLVGTDTSTVYYRHRNSEESVKRQDRHVPKRLLESQFGSNSRERGNI
CKRRRLEAAPQIHPPGFQPREDQVHVSNNVGVHTADDGTVAHSNAGENMQQAEIDMRYDSDDIAMRVIAR
VMLGKNYREAWEVFMAGREQLKLAAYAFADPPVPGFRYLPTRRNRRTWQLDKLVRSKITEIIKARLASSV
YGDDLLGQMLWLQRSGAGANAENLSTEEMVGECRTFFMAGYETSANLITWAMFLLASHPRWQEMVRDEVV
QEFPAHKPPLGDGLGKLKLLNMLLWETLRLYGPISFLQRKTASDTILTHVKVPKGTMITIPLVMLHRDKE
VWGPDADKFNPMRFQNGFARAAKHSHALLAFSYGPRVCVGQNFAMVEVQIVIATMLKSFSFSLSPTYVHK
PSNFVTLTPKYGLPLIVRNLQLTR*

Zavitan genes:

Nucleotide and protein sequence of 22 annotated genes in the Zavitan *W1* physical map interval. Gene/protein nomenclature is similar to Figure 2C.

>NB-Arc-like_1

ATGGCGACGTTCGTGATCCAGCACAACACAGCACTAGCTAAGCAAGAGTCCGTGGTTAGCACTAATAAGC
AAGTCATATCACACGGGCCCTCCCCTGCATCCCTTCCCTGTATAAACGATTCTCCACGGCACAAACAGAA
TCATCAAACCAACACCAGGTACGCACTACGCAGCACACACGTTAACCTCAGAGTCCCCATTACATTGATG
GCTTCCCTGGCCCGAAATACCGTGCGGCTCCTTGTCCTGGTCGGCGTCGTTCTGCAGCTGAGCTCCATCA
TGCCGGCCGCGGCGGCTGGCAGGATGCTTGCAGGGGACGTACTGAAGTGTGAAGGCGGGTTGGTATTCAA
GAGAGATCCCTCCATTAGCGGCGGCGGCATGTGCGTCGTAGCCGACATACTCAAGTGCGGCGGCGGGTTG
GTATTCAAGGAAGATCCCTCCGTTAGCGGCGGCGGCGAGTGCGTCGTAGCCAACGTACTCAAGTGCGGCG
CCGGGTTCGAATTCAAGGAAGATCCATCCATTAACGGCGGTGGCGGGTGCGTCGTAGCCGCCGTACTCAA
GTGCGGCGCCGGGTTCGAATTCAAGAAAGATCCCTCCATGAGCGGCGGCGGCGGATGCGTCGTAGCCGCC
TTACTCAAGTGCGGTGCTGGGTTCGAATTCAAGGAAGATCCCTCCATTCGCGGCGGCGGCGAGTGTGTTG
TAACCAACGTAGTCAAGTGCGACGCCGGGTTCGAATTCAAGAAAGACAACCCCCCTACCGGCGGCGGCAA
GTTCCCATTAATGGTGCATGCGTCTGTCGTCGTCGTCGTTGATTGGCTATTTCCTTTGTTCCTTCAATCC
TCTGTGTATGGTTTGTCGATAACAGTGACTATTTTATTCAAGGCTGGGGCGCCCCCCTCCCGTCGATCGC
TCAGCTATGCCATTGTTTCGTCTGAGCAGCTTGCAGTTGCACCCCAAACACCTTGCAAAAGTGTAGATCA
CCTCATGGCAATCGAGCCAATGATGGGCATGCAAAGGCCCAATGTACTCCGTTTCATGCTAGCCCAGGAG
GAAGCGGCAATGAAACAACGCTTTGCCGAGTACAAAGTAGATGTGCCCACGGTTGTAGTCCGTGCTGTGA
TGAAGATTGGTTCGATTTTAAATGTTGAAGCTGCTGGGGCTAGCATATATAAACTGTCTGATGAAGTTAC
TAATAGGAAAGAATTGCCCGTGAAGGTCTATCAAATAAGTGAGCCACTGATAATGTCAGGAAGTGTTATT
GGGAAGATAGGCACGGTGTACGGCACGGACAAAGTTGCCAAGAGTTGTATTGGACAGGCCCGCAAACTGG
CCTATCACGTCGAAGATTGGATCACTGTGTCACAAACTTATACTGTTGAAGTGTTGATGCCATTGGCAGG
AGGCACTGACAAAACGGACATGGAGCAAGAGGTATACTTCCATATACATGATGCTCTCCAGAATTTCCAA
GGAAGTTGCATCGTCAGCACGACATGGAACGAACATGTTGCAATGATTTCTTCTCCCACCGGTCGCCTTC
GGCTCGAGCAGTTGAGTCAGCCTGATGCATTTGACCTTCTCTGCAGACGGGATTTTACAGAAACAACGGC
TACATGTGCCCCGAGGAGCTTGAGGCAGTTGCGACCTCAAAGGCTGCTTTTGTACTCCAGTCTGTTCCCT
AAAGACCATCCCATGTCACGTGAGAGCCTGGTACGACTTTGGGTTTCAGAAGGCCTTGTGGTGAGTAGAG
GCATGGACACAACAGAGATGCTAGCCGAGGAAAATATCAGGGAATTGGTCTGCCGCAATATGCTTGAAGT
TATGAAGAATGATGGGCTTGGCAGGGTCAAAACCTGTGAGATGTGA

>NB-Arc-like_1_protein

MATFVIQHNTALAKQESVVSTNKQVISHGPSPASLPCINDSPRHKQNHQTNTRYALRSTHVNLRVPITLMASLARNT
VRLLVLVGVVLQLSSIMPAAAAGRMLAGDVLKCEGGLVFKRDPSISGGGMCVVADILKCGGGLVFKEDPSVSGGGEC
VVANVLKCGAGFEFKEDPSINGGGGCVVAAVLKCGAGFEFKKDPSMSGGGGCVVAALLKCGAGFEFKEDPSIRGGGE
CVVTNVVKCDAGFEFKKDNPPTGGGKFPLMVHASVVVVVDWLFPLFLQSSVYGLSITVTILFKAGAPPSRRSLSYAI
VSSEQLAVAPQTPCKSVDHLMAIEPMMGMQRPNVLRFMLAQEEAAMKQRFAEYKVDVPTVVVRAVMKIGSILNVEAA
GASIYKLSDEVTNRKELPVKVYQISEPLIMSGSVIGKIGTVYGTDKVAKSCIGQARKLAYHVEDWITVSQTYTVEVL
MPLAGGTDKTDMEQEVYFHIHDALQNFQGSCIVSTTWNEHVAMISSPTGRLRLEQLSQPDAFDLLCRRDFTETTATC
APRSLRQLRPQRLLLYSSLFPKDHPMSRESLVRLWVSEGLVVSRGMDTTEMLAEENIRELVCRNMLEVMKNDGLGRV
KTCEM*

>NB-Arc-like_2

ATGGCTTCCCTGGCCCGAAATACCGTGCTGCTCCTTGTCCTGGTCGGCGTCGTTCTGCAGCTGAGCTCCATCATGCC
GGCCGCGGCGGCTGGCAGGATGCTTGCAGAGGACGTACTGAAGTGTGAAGGCGGGTTGGTATTCAAGAGAGATCCCT
CCATTAGCGGCGGCGGCATGTGCGTCGTAGCCGACATACTCAAGTGCGGTGGCGGGTTGGTATTCAAGGAAGATCCC
TCCGTTAGCGGCGGCGGCGAGTGCGTCGTAGCCAACGTACTCAAGTGCGGCGCCGGGTTCGAATTCAAGGAAGATCC

ATCCATTAATGGCGGTGGCGGGTGCGTCGTAGCCGCCATAGTCAAGTGCGGCGCCGGGTTCGAATTCAAGAAAGATC
CCTCCATAAGCGGCGGCGGCGGATGCGTCGTAGCCGCCTTACTCAAGTGCGGTGCTGGGTTCGAATTCAAGGAAGAT
CCCTCCATTCGCGGCGGCGGCGAGTGTGTTGTAACCAACGTAGTCAAGTGCGACGCCGGGTTCGAATTCAAGAAAGA
CAACCCCCCTACCGGCGGCAGCAAGATAACAGTGACTATTTTATTCAAGGCTGGGGCGCCCCCCTCCCGTCGATCGC
TCAGCTATGCCAATGTTTCGTCTGAGCAGCTTGCAGTTGCACCCCAAACACCTTGCAAAAGTGTAGATCACCCTCGT
GGCAATCCAGCCAATGATGGGCATGCAAAGGCCCAATATCATCCGTTTCATGCTAGCCCAGGAGGAAGCCGGATATC
ATGA

>NB-Arc-like_2_protein

MASLARNTVLLLVLVGVVLQLSSIMPAAAAGRMLAEDVLKCEGGLVFKRDPSISGGGMCVVADILKCGGGLVFKEDP
SVSGGGECVVANVLKCGAGFEFKEDPSINGGGGCVVAAIVKCGAGFEFKKDPSISGGGGCVVAALLKCGAGFEFKED
PSIRGGGECVVTNVVKCDAGFEFKKDNPPTGGSKITVTILFKAGAPPSRRSLSYANVSSEQLAVAPQTPCKSVDHPR
GNPANDGHAKAQYHPFHASPGGSRIS*

>DUF4220

ATGGCTGGAGTGGTGCGCCTATGGAACGAGTGGGAGGTCCAAATCCTAATGCTTGCAAGCTTCATGCTGCAAGTGTT
CCTGCTCCTATTCGCAGGGAGGCGTCGGCAGAACATCTCCGTCGTCCCACGGACCCTCCTATGGCTAGCATACCTCC
TGGCTGACTCCACAGTGATATACATACTCGGCCACATGTCAATTTGCAGCAAGTCCCACGAGAAGCAACAGCTCATG
GCGTTCTGGGCGCCGTTCTTGCTGGTGCATCTCGGTGGCCAGGATACCATCACCGCCTACTCCATAGAAGACAACCA
GCTGTGGCCGCGCCACCTGCTCTCTCTGGTGGTGCAGGCATTAGGAGTAGGTTATGTCTTCTACAAGTATGTCGCTG
GTAGCTGGACCTTGTTCACAGCCGCAGCTCTGGTATTTGTGGCCGGTGTTCTTAAGTATGGCGAGAGAGTATGGGCG
CTCAGGAACTCCAAGCTGGACAGCATCAGGAAGTTCCTTGATGAGGTTGAGAAGCCGAGGGTGGTGAAGGAGGCTGA
ACGAGAAGAAGAACGGCCCTACCCTCTGCCTGCACAGTGGGGCGAAGAGGATAAGTTAGATTCTGAAGAAGTACTGC
AGGGAGCACATGATCTATTACCCATTTGCATGGGTCAGTTTGTCGATTATAAGTTCTTGCCGTCCCGGTTGCAGATT
CAAGCCAACTGCCGATTCAAGCGCAAGGGGTGCTTGTATGAGCTGATCGAGATGCAGCTCTCCTTGATGCACGACGT
CCTGTACACCAAGGCGGCGGTGGTCTTCACATGGTATGGATGCTCCATCCGTGCGATATCATTGGTTGCCACCATCA
GTGCTTTCTTCCTATTTCAATCAAGCATTGGCAAGAATGACTTCAGCAGAGGTGATATCATTGTCACATATATCTTA
ATAGCTGGGGCTGTCCTCCTAGAGATGACATCATTGTTGAAGGCAATGGGTTCAACATGGGCATGCACATTGTTGCG
TGCTAGAAGATGGGTTCGGCTTCATAGCATGGTCGTATCTGTTCGACAACATGTCAGGGCTGCGGAAAGAAATAGAA
GATGGTCAGGTTTTATTGGACAGCCCAGGCCGTCCACTGGCATCGGACATGGACAGGACAGGGGTGGCTGGACCCGG
GAAGTTGCCAAGTGGTTTGAATTGAAGCACTGCAAAGATTTGTGGAAGAAGCTGCACCGCTCCTTACCTGTCATTTC
ATATGGTACTAAGGAGTTGGTGATGAGACATGTAACAAGAATGGTGGAGGAATGCCATGGCCAAGAGGATGTCATGA
GGAGTTACAGTGGTCAGTGCGCGCTGAGTCGATGGCAGGTAACCTTTGAGGATCCCACCTCGCGCGCTGGCATTGAT
TTTGATGACAAAATCCTTGCCTGGTACTTTGCCACCAAAATGTTCTTCCAATTTCACGTGGAACCAAGGCAGCTAGA
GCCAGTTGTGGAGGCAATCGATACGCTGTCCGAATACATGATATTCCTCCTTTTGGAACGTCCGTACATGCTGCCTA
GCCCTGTTCGCCCCGTACCATATGCCAATGCTGAAGCAGCATATCAAAAGCTTAATTTGCATTGTGTGGTGGGTTCC
TTTATTAAGCGCATCACAAGAATGGCGGAGCTGGACAAAAGAAGAGAGGAGCTGGAGAAAACAAAGAGGGATCGATG
GCATGCTAGAGAGCTGAAAAAAATAAGGGAGGAGTTGGAAACAATAAGGGAGGAGGTGCAGGTGCTGGAGAGAGGAA
AGGAGAACCTGCTTTCGGTACGTTTACATCCACCCGCCTTAGTCCGTGGAGCCGAGATTGTTGAAAGGCTGATATAT
AATGATGCGCGGGACCTGGACAACATGGCGGTCGGGACGGTACTGCTTGGGGTGTGGGTGGAGATGCTGTGCTACGC
GGCGCACCACTGCAGCAGGGATTCACACGCCAGGCAGCTCAACAATGGCGGTGAGTTCATTACCATTGTCTGGCTTT
TGTCAACTGCCATTTTCAATGGAAACTACATCAGGTGA

>DUF4220_protein

MAGVVRLWNEWEVQILMLASFMLQVFLLLFAGRRRQNISVVPRTLLWLAYLLADSTVIYILGHMSICSKSHEKQQLM
AFWAPFLLVHLGGQDTITAYSIEDNQLWPRHLLSLVVQALGVGYVFYKYVAGSWTLFTAAALVFVAGVLKYGERVWA
LRNSKLDSIRKFLDEVEKPRVVKEAEREEERPYPLPAQWGEEDKLDSEEVLQGAHDLLPICMGQFVDYKFLPSRLQI
QANCRFKRKGCLYELIEMQLSLMHDVLYTKAAVVFTWYGCSIRAISLVATISAFFLFQSSIGKNDFSRGDIIVTYIL
IAGAVLLEMTSLLKAMGSTWACTLLRARRWVRLHSMVVSVRQHVRAAERNRRWSGFIGQPRPSTGIGHGQDRGGWTR
EVAKWFELKHCKDLWKKLHRSLPVISYGTKELVMRHVTRMVEECHGQEDVMRSYSGQCALSRWQVTFEDPTSRAGID
FDDKILAWYFATKMFFQFHVEPRQLEPVVEAIDTLSEYMIFLLLERPYMLPSPVRPVPYANAEAAYQKLNLHCVVGS
FIKRITRMAELDKRREELEKTKRDRWHARELKKIREELETIREEVQVLERGKENLLSVRLHPPALVRGAEIVERLIY
NDARDLDNMAVGTVLLGVWVEMLCYAAHHCSRDSHARQLNNGGEFITIVWLLSTAIFNGNYIR*

>PKS-1.1

ATGGCAGGCAGCTCACCGAAGGTTAGTGAGATCCGGTGTGCGCAGCGTGCGGAAGGCTCCGCGGCAATGCTGGCTAT
CGGAACAGCAAATCCGGCGAACAAGGTGTCCCAAGAAGAGTACCCTGACTATTATTTCCGCGTTACCAAGAGCGAGC
ACCTTACTGACCATAAAGACACATTCAAGATAATATGTGGTCTAACGGGCACGGAGAATCGTTTCTTCTACCACACG
GATGAACTGCTCAACTCCCACCCTGTCTTGCTGGACAACACGTCACCGTCCCGTGAGGCTCGGCATGATATCGTGGC
CAAGGCTGCTCCAGAGCTTGCGGCAGCAGCAGCAAAGAAGGCCATCGCAAAGTGGGGCCGTCCGGCCAGTGACATCA
CCCACCTTATCGTCAGCACCAACTCTGACGCTGGCGCCCCAAGCGCTGACGTACGCTTGGTTTCACTCCTTGGCCTT
CGCGCCGATGTCTGTCGTACCATGCTCCATCTTAACGGCTGCTTCGCCGGCTGCTCCGCCTTGCGCCTAGCCAAGGA
CCTTGCTGAGAATAACTGTGGGGCACGCGTCCTCGTAGCTTGCGTGGAGCTCGCCATTTCTGGCTTCTGCAGCCCCG
GCGAAGGGGACTGCTTAGACACCCTCATCACTCATGCGTTGTTTGGTGATGGGGCAGGCGCGGTAATCGTCGGCGCC
GACCCCATGCACCCCATCGAGAATCCTCTGTTTGAGATGGTGTCCGTCTCTCAGACCCTCGTGCCGAGCACTGAGAA
TGTGCTCACCCTGAACTTGGGGAGCAATGGCACTCATGGGAAAGTTTACACCAAACTGCCCACACTGGTAGCAGACA
CCATGGAACCTTGTCTTCTGGAAGCGTTTGGTCCACTTGAGATGGACTTCCAATGGAATGACCTCTTCTGGGCAGTG
CACCCTGGCAGCCGTGGGATCTTGGACCAACTTGACAAGACACTCCAGTTGGAGCCCACGAAGCTAGCGGCGAGCAG
AACTGTCGTACAGAAGTTTGGGAACATGTTTAGCGCCACCGTGATCTTCGTGCTTGATGAGCTACAGCGTCGAATGG
AGGAGGAAGGAGAGCAAGCTGAGTGGGGGGCCATGGTGGGATTTGGACCAGGCTTCACTATCGAGACCATGGTGCTC
CATGCAACCGGCGCTCTCAAGAAAAATTAG

>PKS-1.1_protein

MAGSSPKVSEIRCAQRAEGSAAMLAIGTANPANKVSQEEYPDYYFRVTKSEHLTDHKDTFKIICGLTGTENRFFYHT
DELLNSHPVLLDNTSPSREARHDIVAKAAPELAAAAAKKAIAKWGRPASDITHLIVSTNSDAGAPSADVRLVSLLGL
RADVCRTMLHLNGCFAGCSALRLAKDLAENNCGARVLVACVELAISGFCSPGEGDCLDTLITHALFGDGAGAVIVGA
DPMHPIENPLFEMVSVSQTLVPSTENVLTLNLGSNGTHGKVYTKLPTLVADTMEPCLLEAFGPLEMDFQWNDLFWAV
HPGSRGILDQLDKTLQLEPTKLAASRTVVQKFGNMFSATVIFVLDELQRRMEEEGEQAEWGAMVGFGPGFTIETMVL
HATGALKKN*

>PKS-1.2

ATGGCAGGCAGCTCACCGAAGGTTAGTGAGATCCGGTGTGCGCAGCGTGCGGAAGGCTCCGCGGCAATGCTGGCTAT
CGGAACAGCAAATCCGGCGAACAAGGTGTCCCAAGAAGAGTACCCTGACTATTATTTCCGCGTTACCAAGAGCGAGC
ACCTTACTGACCATAAAGACACATTCAAGATAATATGTGGTCTAACGGGCACGGAGAATCGTTTCTTCTACCACACG
GATGAACTGCTCAACTCCCACCCTGTCTTGCTGGACAACACGTCACCGTCCCGTGAGGCTCGGCATGATATCGTGGC
CAAGGCTGCTCCAGAGCTTGCGGCAGCAGCAGCAAAGAAGGCCATCGCAAAGTGGGGCCGTCCGGCCAGTGACATCA
CCCACCTTATCGTCAGCACCAACTCTGACGCTGGCGCCCCAAGCGCTGACGTACGCTTGGTTTCACTCCTTGGCCTT
CGCGCCGATGTCTGTCGTACCATGCTCCATCTTAACGGCTGCTTCGCCGGCTGCTCCGCCTTGCGCCTAGCCAAGGA
CCTTGCTGAGAATAACTGTGGGGCACGCGTCCTCGTAGCTTGCGTGGAGCTCGCCATTTCTGGCTTCTGCAGCCCCG
GCGAAGGGGACTGCTTAGACACCCTCATCACTCATGCATTGTTTGGTGATGGGGCAGGCGCGGTAATCGTCGGCGCC
GACCCCATGCACCCCATCGAGAATCCTCTGTTTGAGATGGTGTCCGTCTCTCAGACCCTCGTGCCGAGCACTGAGAA
TGTGCTCACCCTGAACTTGGGGAGCAATGGCACTCATGGGAAAGTTTACACCAAACTGCCCACACTGGTAGCAGACA
CCATGGAACCTTGTCTTCTGGAAGCGTTTGGTCCACTTGAGATGGACTTCCAATGGAATGACCTCTTCTGGGCAGTG
CACCCTGGCAGCCGTGGGATCTTGGACCAACTTGACAAGACACTCCAGTTGGAGCCCACGAAGCTAGCGGCGAGCAG
AACTGTCGTACAGAAGTTTGGGAACATGTTTAGCGCCACCGTGATCTTCGTGCTTGATGAGCTACAGCGTCGAATGG
AGGAGGAAGGAGAGCAAGCTGAGTGGGGGGCCATGGTGGGATTTGGACCAGGCTTCACTATCGAGACCATGGTGCTC
CATGCAACCGGCGCTCTCAAGAAAAAATAG

>PKS-1.2_protein

MAGSSPKVSEIRCAQRAEGSAAMLAIGTANPANKVSQEEYPDYYFRVTKSEHLTDHKDTFKIICGLTGTENRFFYHT
DELLNSHPVLLDNTSPSREARHDIVAKAAPELAAAAAKKAIAKWGRPASDITHLIVSTNSDAGAPSADVRLVSLLGL
RADVCRTMLHLNGCFAGCSALRLAKDLAENNCGARVLVACVELAISGFCSPGEGDCLDTLITHALFGDGAGAVIVGA
DPMHPIENPLFEMVSVSQTLVPSTENVLTLNLGSNGTHGKVYTKLPTLVADTMEPCLLEAFGPLEMDFQWNDLFWAV
HPGSRGILDQLDKTLQLEPTKLAASRTVVQKFGNMFSATVIFVLDELQRRMEEEGEQAEWGAMVGFGPGFTIETMVL
HATGALKKK*

>P450-1.1

ATGGCGGCTAACGTAGTGCAAGCCCTGGCCGCTGTCCTCACACTGCTGGTGATCACAAGAGCGCTATGGTATCTGCT
TTGGAGGCCATACGCTGTGGCCAGGTGGTTCGAGCAGCAGGGCATCAGAGGTCCGCCTTACAAATTCCTGGTTGGGT
CCCTGCCCGACTGCCAAAGGATGCTCGTTGCCGGGAGAGTCAAGGATCTGGACACGAGCTCCCACGACTGCATCACC
ACCGTGCAACCCTTCTTCCGGAAATGGGCCTCACTGTATGGAAAAACATTCCTGTACTGGTTGGGACCGACACCAGC
ACTGTGTACTACAGACATAGAAATAGTGAAGAAAGTGTTAAGCGACAGGACAGACATGTTCCAAAAAGATTACTTGA
ATCCCAGTTTGGAAGCAATTCTAGGGAACGGGGTAATATTTGCAAACGGAGACGACTGGAAGCGGCGCCGCAAATTC
ATCCACCCGGCTTTCAGCCAAGAGAAGATCAAGTCCATGTCAGCAATAACGTTGGAGTGCACACAGCAGATGATGGA
ACAGTGGCGCACTCAAATGCAGGAGAGCAACATGCAGCAAGCCGAGATCGACATGAGGTACGACTCCGATGACATAG
CAATGCGTGTCATAGCACGAGTGATGCTGGGCAAGAACTACAGGGAGGCCTGGGAAGTGTTCATGGCAGGAAGGGAG
CAACTGAAGCTCGCCGCGTATGCATTTGCGGATCCTCCAGTACCTGGATTCAGGTACCTGCCGACACGTCGCAACCG
TCGGACCTGGCAACTCGACAAGCTTGTGAGAAGCAAGATTACAGAGATCATAAAGGCGCGGCTTGCTAGCAGTGTCT
ATGGAGACGACCTGCTCGGGCAGATGTTGTGGCTGCAGAGGTCGGGTGCTGGCGCCAACGCCGAGACCCTGAGCACG
GAGGAGATGGTCGGCGAGTGCAGGACCTTCTTCATGGCTGGGTACGAAACCAGCGCCAACCTCATTACCTGGGCCAT
GTTCCTGCTCGCCAGCCACCCACGGTGGCAGGAGATGGTCAGGGACGAGGTCGTCCAGGAGTTTCCTGCTCACAAGC
CACCCTTAGGTGACGGTCTCGGCAAATTGAAGCTGCTCAACATGTTACTCTGGGAGACATTGAGGCTCTATGGCCCC
CTATCATTCCTGCAGAGGAAGACAGCCTCAGACACAATCCTCACACATGTGAAGGTGCCGAAAGGAACGATGATAAC
GATACCTCTGGTGATGTTGCACCGGGACAAAGAGGTCTGGGGACCCGACGCCGACGAGTTTAACCCAATGAGGTTCC
AGAATGGCTTCGCGAGAGCCGCCAAGCATTCACATGCACTGCTGGCCTTCTCGTATGGGCCTAGGGTCTGTGTCGGG
CAGAACTTTGCCATGGTGGAGGTGCAGATCGTCATAGCGACGATGCTCAAAAGTTTCTCCTTCTCCCTGTCCCCCAC
TTATGTGCACAAGCCGAGCAATTTCGTCACATTGACGCCCAAGTACGGGCTCCCTCTCATCGTGAGGAACCTGCAGC
TGACTAGGTGA

>P450-1.1_protein

MAANVVQALAAVLTLLVITRALWYLLWRPYAVARWFEQQGIRGPPYKFLVGSLPDCQRMLVAGRVKDLDTSSHDCIT
TVQPFFRKWASLYGKTFLYWLGPTPALCTTDIEIVKKVLSDRTDMFQKDYLNPSLEAILGNGVIFANGDDWKRRRKF
IHPAFSQEKIKSMSAITLECTQQMMEQWRTQMQESNMQQAEIDMRYDSDDIAMRVIARVMLGKNYREAWEVFMAGRE
QLKLAAYAFADPPVPGFRYLPTRRNRRTWQLDKLVRSKITEIIKARLASSVYGDDLLGQMLWLQRSGAGANAETLST
EEMVGECRTFFMAGYETSANLITWAMFLLASHPRWQEMVRDEVVQEFPAHKPPLGDGLGKLKLLNMLLWETLRLYGP
LSFLQRKTASDTILTHVKVPKGTMITIPLVMLHRDKEVWGPDADEFNPMRFQNGFARAAKHSHALLAFSYGPRVCVG
QNFAMVEVQIVIATMLKSFSFSLSPTYVHKPSNFVTLTPKYGLPLIVRNLQLTR*

>Hyd-1.1

ATGCCTGCAAACAAGACTTACCCCTCCCATAAAAATGCCAACGGTGAGGTGGACGACGAATTCTACCCATTAATCCG
CAAGTACAAGGACGGCCGGATCGAGCGGTTCATGAGCTCATTCGTGCCGGCGTCGGAGGACCCGGCCGCCAGCCGTG
GTGTGGCGACGAGGGACGTCGTCGTCGACCAGGGCACCGGTGTGTCCGTGCGCCTGTTCCTTCCTGCCCAGGCTGCC
GAGGCCGGCACGAGGCTCCCCCTTGTTGTGTACGTCCATGGTGGTTCCTTCTGCACGGAGAGTGCCTTCTCCCGGAC
GTACCACCGTTACGCCACTTCCCTCGCCGCCAGCGCAGGGGCGCTCATCGTGTCCGTGGAGTACCGTCTGGCGCCGG
AATATCCCGTCCCGACGTCCTACGATGACACATGGGCCGCGCTGCGGTGGGTGGCGTCCTTGTCCGACCCTTGGCTC
GCCAAATACACAGACCCTAGCCGCACGTTCCTCGCCGGCGACAGCGCTGGCGGCAACATCGTGTACCACACGGCCGT
GCGCGCCACACATGATGACAGCATCATGGACATCCAGGGGTTGGTCATGGTGCATCCATTCTTCTGGGGGCCCGAGC
GTCTCCCGGCGGAGAAGGTCTTGGACGGCGACGCCATGTTCCCACCAGTGTGGGTGGATAAGCTGTGGCCGTTCGTG
ACGGCGGGCGGGGCTGGCAACGATGATCCTCGGATCAATCCTCCGGACGAGGAGATCGCGTTGCTAACTGGCAGGCG
GGTGCTTGTGGCCGTTGCAGAGAAGGACACCCTGCGCGACCGGGGGCGCCAGTTTGTGTGCAGCATGCGCAGGTGTG
GGTGGGTTGATGGCAGCCTCACCGTGGTGGAGTCGGAGGGTGAGGACCATGGCTTCCACTTGTACGCCCCCCTACGT
GCGACCAGCAAGAAGCTTATGAAGAGCATGGTGCAGTTCATAAACCATCGCGCCACCTTGCCGTCACCGGCCATGGT
GATCCCAGAAGGCTCGGCCGAAACTATGCTAGGCGTCCCTAGTAGGCCATTTAAGGACATATTTGGCTACGGGATGC
GCATGAAACGTTGGAGTGGCACGAGTTTTGGGCTCAAAGTTGGTCGTGCAAAAGCATCGACGACGAGCTATGGGTTA
CCTTTGAAGCAAGCTCGCACCTTCGGAGACCCTGTTTCAGCACCAACTTCGGTAAGATTCGTGATGAGGAACTGTTT
CTAG

>Hyd-1.1_protein

MPANKTYPSHKNANGEVDDEFYPLIRKYKDGRIERFMSSFVPASEDPAASRGVATRDVVVDQGTGVSVRLFLPAQAA
EAGTRLPLVVYVHGGSFCTESAFSRTYHRYATSLAASAGALIVSVEYRLAPEYPVPTSYDDTWAALRWVASLSDPWL
AKYTDPSRTFLAGDSAGGNIVYHTAVRATHDDSIMDIQGLVMVHPFFWGPERLPAEKVLDGDAMFPPVWVDKLWPFV
TAGGAGNDDPRINPPDEEIALLTGRRVLVAVAEKDTLRDRGRQFVCSMRRCGWVDGSLTVVESEGEDHGFHLYAPLR

ATSKKLMKSMVQFINHRATLPSPAMVIPEGSAETMLGVPSRPFKDIFGYGMRMKRWSGTSFGLKVGRAKASTTSYGL
PLKQARTFGDPVSAPTSVRFVMRNCF*

>Hyd-1.2 (pseudogene)

ATGCCTGCAAACAAGACTTACCCCTCCCATAAAAATGCCAACGGTGAGGTGGACGACGAATTCTACCCATTAATCCG
CAAGTACAAGGACGGCCGGATCGAGCGGTTCATGAGCTCATTCGTGCCGGCGTCGGAGGACCCGGCCGCCAGCCGTG
GTGTGGCGACGAGGGACGTCGTCGTCGACCAGGGCACCGGTGTGTCCGTGCGCCTGTTCCTTCCTGCCCAGGCTGCC
GAGGCCGGCACGAGGCTCCCCCTTGTTGTGTACGTCCATGGTGGTTCCTTCTGCACGGAGAGTGCCTTCTCCCGGAC
GTACCACCGTTACGCCACTTCCCTCGCCGCCAGCGCAGGGGCGCTCATCGTGTCCGTGGAGTACCGTCTGGCGCCGG
AATATCCGGTCCCAACGTCCTACGATGACACATGGGCCGCGCTGCGGTGGGTGGCGTCCTTGTCCGACCCTTGGCTC
GCCAAATACACAGACCCTAGCCGCACGTTCCTCGCCGGCGACAGCGCTGGCGGCAACATCGTGTACCACACGGCCGT
GCGCGCCACACATGATGACAGCATCATGGACATCCAGGGGTTGGTCATGGTGCATCCATTCTTCTGGGGGCCCGAGC
GTCTCCCGGCGGAGAAGGTCTTGGACGGCGACGCCATGTTCCCACCAGTGTGGGTGGATAAGCTGTGGCCGTTCGTG
ACGGCGGGCGGGGCTGGCAACGATGATCCTCGGATCAATCCTCCGGACGAGGAGATCGCGTTGCTAACTGGCAGGCG
GGTGCTTGTGGCCGTTGCAGAGAAGGACACCCTGCGCGACCGGGGGCGCCAGTTTGTGTGCAGCATGCGCAGGTGTG
GGTGGGTTGATGGCAGCCTCACCGTGGTGGAGTCGGAGGGTGAGGACCATGGCTTCCACTTGTACGCCCCCTACGTG
CGACCAGCAAGAAGCTTATGAAGAGCATCGTGCAGTTCATAAACCATCGCGCCACCTTGCCGTCACCGGCCATGGTG
ATCCCAGAAGGCTCGGCCGAAACTATGCTAGGCGTCCCTAGTAGGCCATTTAAGGACATATTTGGCTACGGGATGCG
CATGAAACGTTGGAGTGGCACGAGTTTTGGGCTCAAAGTTGGTCGTGCAAAAGCATCGACGACGAGCTATGTGTTAC
CTTTGAAGCAAGCTCGCACCTTCGGAGACCCTGTTTCAGCACCAACTTCGGTAAGATTCGTGATGAGGAACTGTTTC
TAG

>Hyd-1.2 (pseudogene)_protein

MPANKTYPSHKNANGEVDDEFYPLIRKYKDGRIERFMSSFVPASEDPAASRGVATRDVVVDQGTGVSVRLFLPAQAA
EAGTRLPLVVYVHGGSFCTESAFSRTYHRYATSLAASAGALIVSVEYRLAPEYPVPTSYDDTWAALRWVASLSDPWL
AKYTDPSRTFLAGDSAGGNIVYHTAVRATHDDSIMDIQGLVMVHPFFWGPERLPAEKVLDGDAMFPPVWVDKLWPFV
TAGGAGNDDPRINPPDEEIALLTGRRVLVAVAEKDTLRDRGRQFVCSMRRCGWVDGSLTVVESEGEDHGFHLYAPYV
RPARSL*

>HlyIII-1.1

ATGGGGAACGGCGGCGGCGTCCATGCCGATCAAGCTGTTGCATTGAAAAACAAAGGCAGCCCCACCTGCAGTGACGT
GGCGAAGCCGAGCTGCAAGCATAGGTACGAGCTCGTCTGCTACGATGCGCTCCCGGCCTTCTTGAAGCACGACGAAT
TCATCCTCGATCACTACCGCAGCGAATGGCCCGTCAAGCAGGCGATCCTCAGCGCCTTCGCCCTCCACAACGAGACG
ATCAACGTCTGGACGCATTTGATCGGCTTCTTCATCTTCCTCGCGCTCACCATGTCCGTCTGGTCGGGGGAGCACGG
CCCGGTCACGCGGTGGCCGTTCTACACGTACCTGTGCGGCGCCATGTTCTGCCTGCTGATGAGCAGCGGTTGCCATC
TGCTGGCGTGCCACTCGGAGCATGCCACCTACGTGCTGCTCCGGCTCGACTACGCCGGCATCACCGGCCTTATCGTC
ACCTCCTTCTACCCGGTCGTCTACTACACCTTCCTCTGTGACCCCTTCTCCCGTACACTCTACCTTGGCTCCATCAC
CATCTGCGGCGCGGCCGCTGTGGCTGTCTCGCTGCTGCCGGTCTTCCAGGCACCTGATCTGCGGTGGGCGCGCGCTG
CGCTATTCGCGTGCATGGGCGCGTCCGGCATCGTGCCCATCGTGCACAAGATGCTCATGTTCCATGCACGACCCGAG
GCGGTGCTCACGACTGGTTACGAGGTGGTCATGGGGGCGTTCTACCTTGCCGGCGTGTTGGTTTACGCCACAAGGAT
GCCGGAGAGGTTTATGCCAGGAATGTTTGACCTCGTCGGGCATAGCCACCAGCTGTTCCACGTGCTAGTCATTGCCG
GCGCTTACGCGCACTACCATGCTGGCCTGGTTTACTTGAGTTGGCGGGACAGGGACCAATGCTGA

>HlyIII-1.1_protein

MGNGGGVHADQAVALKNKGSPTCSDVAKPSCKHRYELVCYDALPAFLKHDEFILDHYRSEWPVKQAILSAFALHNET
INVWTHLIGFFIFLALTMSVWSGEHGPVTRWPFYTYLCGAMFCLLMSSGCHLLACHSEHATYVLLRLDYAGITGLIV
TSFYPVVYYTFLCDPFSRTLYLGSITICGAAAVAVSLLPVFQAPDLRWARAALFACMGASGIVPIVHKMLMFHARPE
AVLTTGYEVVMGAFYLAGVLVYATRMPERFMPGMFDLVGHSHQLFHVLVIAGAYAHYHAGLVYLSWRDRDQC*

>P450-1.2 (pseudogene)

ATGGCGGCTAACGTAGTGCAAGCCCTGGCCGCTGTCCTCACACTGCTGGTGATCACAAGAGCGCTATGGTATCTGCT
TTGGAGGCCATACGCTGTGGCCAGGTGGTTCGAGCAGCAGGGCATCAGAGGTCCGCCTTACAAATTCCTGGTTGGGT
CCCTGCCCGACTGCCAAAGGATGCTCGTTGCCGGGAGAGTCAAGGATCTGGACACGAGCTCCCACGACTGCATCACC

ACCGTGCAACCCTTCTTCCGGAAATGGGCCTCACTGTATGGAAAAACATTCCTGTACTGGTTGGGACCGACACCAGC
ACTGTGTACTACAGACATAGAAATAGTGAAGAAAGTGTTAAGCGACAGGACAGACATGTTCCAAAAAGATTACTTGA
ATCCCAGTTTGGAAGCAATTCTAGGGAACGGGGTAATATTTGCAAACGGAGACGACTGGAAGCGGCGCCGCAAATTC
ATCCACCCGGCTTTCAGCCAAGAGAAGATCAAGTCCATGTCAGCAATAACGTTGGAGTGCACACAGCAGATGATGGA
ACAGTGGCGCACTCAAATGCAGGAGAGCAACATGCAGCAAGCCGAGATCGACATGAGGTACGACTCCGATGACATAG
CAATGCGTGTCATAGCACGAGTGATGCTGGCAAGAACTACAGGGAGGCCTGGGAAGTGTTCATGGCAGGAAGGGAGC
GACTGAAGCTCGCCGCGTATGCATTTGCGGATCCTCCAGTACCTGGATTCAGGTACCTGCCGACACGTCGCAACCGT
CGGACCTGGCAACTCGACAAGCTTGTGAGAAGCAAGATTACAGAGATCATAAAGGCGCGGCTTGCTAGCAGTGTCTA
TGGAGACGACCTGCTCGGGCAGATGTTGTGGCTGCAGAGGTCGGGTGCTGGCGCCAACGCCGAGACCCTGAGCACGG
AGGAGATGGTCGGCGAGTGCAGGACCTTCTTCATGGCTGGGTACGAAACCAGCGCCAACCTCATTACCTGGGCCATG
TTCCTGCTCGCCAGCCACCCACGGTGGCAGGAGATGGTCAGGGACGAGGTCGTCCAGGAGTTTCCTGCTCACAAGCC
ACCCTTAGGTGACGGTCTCGGCAAATTGAAGCTGCTCAACATGTTACTCTGGGAGACATTGAGGCTCTATGGCCCCC
TATCATTCCTGCAGAGGAAGACAGCCTCAGACACAATCCTCACACATGTGAAGGTGCCGAAAGGAACGATGATAACG
ATACCTCTGGTGATGTTGCACCGGGACAAAGAGGTCTGGGGACCCGACGCCGACGAGTTTAACCCAATGAGGTTCCA
GAATGGCTTCGCGAGAGCCGCCAAGCATTCACATGCACTGCTGGCCTTCTCGTATGGGCCTAGGGTCTGTGTCGGGC
AGAACTTTGCCATGGTGGAGGTGCAGATCGTCATAGCGACGATGCTCAAAAGTTTCTCCTTCTCCCTGTCCCCCAGT
TATGTGCACAAGCCGAGGAATTTCGTCACATTGACGCCCAAGTACGGGCTCCCTCTCATCGTGAGGAACCTGCAGCT
GACTAGGTGA

>P450-1.2 (pseudogene)_protein

MAANVVQALAAVLTLLVITRALWYLLWRPYAVARWFEQQGIRGPPYKFLVGSLPDCQRMLVAGRVKDLDTSSHDCIT
TVQPFFRKWASLYGKTFLYWLGPTPALCTTDIEIVKKVLSDRTDMFQKDYLNPSLEAILGNGVIFANGDDWKRRRKF
IHPAFSQEKIKSMSAITLECTQQMMEQWRTQMQESNMQQAEIDMRYDSDDIAMRVIARVMLARTTGRPGKCSWQEGS
D*

>Hyd-1.3

ATGCCTGCAAACAAGACTTACCCCTCCCATAAAAATGCCAACGGTGAGGTGGACGACGAATTCTACCCATTAATCCG
CAAGTACAAGGACGGCCGGATCGAGCGGTTCATGAGCTCATTCGTGCCGGCGTCAGAGGACCCGGCCGCCAGCCGTG
GTGTGGCGACGAGGGACGTCGTCGTCGACCAGGGCACCGGTGTGTCCGTGCGCCTGTTCCTTCCTGCCCAGGCTGCC
GAGGCCGGCACGAGGCTCCCCCTTGTTGTGTACGTCCATGGTGGTTCCTTCTGCACGGAGAGTGCCTTCTCCCGGAC
GTACCACCGTTACGCCACTTCCCTCGCCGCCAGCGCAGGGGCGCTCATCGTGTCCGTGGAGTACCGTCTGGCGCCGG
AATATCCCGTGCCGACGTCCTACGATGACACATGGGCCGCGCTGCGGTGGGTGGCGTCCTTGTCCGACCCTTGGCTC
GCCAAATACGCAGACCCTAGCCGCACGTTCCTCGCCGGCGACAGCGCTGGCGGCAACATCGTGTACCACACGGCCGT
GCGCGCCACACATGATGACAGCATCATGGACATCCAGGGGTTGGTCATGGTGCATCCATTCTTCTGGGGGCCCGAGC
GTCTCCCGGCGGAGAAGGTCTTGGACGGCGACGCCATGTTCCCACCAGTGTGGGTGGATAAGCTGTGGCCGTTCGTG
ACGGCGGGCGGGGCTGGCAACGATGATCCTCGGATCAATCCTCCGGACGAGGAGATCGCGTTGCTAACTGGCAGGCG
GGTGCTTGTGGCCGTTGCAGAGAAGGACACCCTGCGCGACCGGGGGCGCCAGTTTGTGTGCAGCATGCGCAGGTGTG
GGTGGGTTGATGGCAGCCTCACCGTGGTGGAGTCGGAGGGTGAGGACCATGGCTTCCACTTGTACGCCCCCCTACGT
GCGACCAGCAAGAAGCTTATGAAGAGCATCGTGCAGTTCATAAACCATCGCGCCACCTTGCCGTCACCGGCCATGGT
GATCCCAGAAGGCTCGGCCGAAACTATGCTAGGCGTCCCTAGTAGGCCATTTAAGGACATATTTGGCTACGGGATGC
GCATGAAACGTTGGAGTGGCACGAGTTTTGGGCTCAAAGTTGGTCGTGCAAAAGCATCGACGACGAGCTATGGGTTA
CCTTTGAAGCAAGCTCGCACCTTCGGAGACCCTGTTTCAGCACCAACTTCGGTAAGATTCGTGATGAGGAACTGTTT
CTAG

>Hyd-1.3_protein

MPANKTYPSHKNANGEVDDEFYPLIRKYKDGRIERFMSSFVPASEDPAASRGVATRDVVVDQGTGVSVRLFLPAQAA
EAGTRLPLVVYVHGGSFCTESAFSRTYHRYATSLAASAGALIVSVEYRLAPEYPVPTSYDDTWAALRWVASLSDPWL
AKYADPSRTFLAGDSAGGNIVYHTAVRATHDDSIMDIQGLVMVHPFFWGPERLPAEKVLDGDAMFPPVWVDKLWPFV
TAGGAGNDDPRINPPDEEIALLTGRRVLVAVAEKDTLRDRGRQFVCSMRRCGWVDGSLTVVESEGEDHGFHLYAPLR
ATSKKLMKSIVQFINHRATLPSPAMVIPEGSAETMLGVPSRPFKDIFGYGMRMKRWSGTSFGLKVGRAKASTTSYGL
PLKQARTFGDPVSAPTSVRFVMRNCF*

>HlyIII-1.2 (partial)

GCGGCGTCCATGCCGATCAAGCTGTTGCATTGAAAAACAAAGGCAGCCCCACCTGCAGTGACGTGGCGAAGCCGAGC
TGCAAGCATAGGTACGAGCTCGTCTGCTACGACGCGCTCCCGGCCTTCTTGAAGCACGACGAATTCATCCTCGATCA
CTACCGCAGCGAATGGCCCGTCAAGCAGGCGATCCTCAGCGCCTTCGCCCTCCACAACGAGACGATCAACGTCTGGA

CGCATTTGATCGGCTTCTTCATCTTCCTCGCGCTCACCATGTCCGTCTGGTCGGGGGAGCACGGCCCGGTCACGCGG
TGGCCGTTCTACACGTACCTGTGCGGCGCCATGTTCTGCCTGCTGATGAGCAGCGGTTGCCATCTGCTGGCGTGCCA
CTCGGAGCATGCCACCTACGTGCTGCTCCGGCTCGACTACGCCGGCATCACCGGCCTTATCGTCACCTCCTTCTACC
CGGTCGTCTACTACACCTTCCTCTGTGACCCCTTCTCCCGTACACTCTACCTTGGCTCCATCACCATCTGCGGCGCG
GCCGCTGTGGCTGTCTCGCTGCTGCCGGTCTTCCAGGCACCTGATCTGCGGTGGGCGCGCGCTGCGCTATTCGCGTG
CATGGGCGCGTCCGGCATCGTGCCCATCGTGCACAAGATGCTCATGTTCCATTCACGACCCGAGGCGGTGCTCACGA
CTGGTTACGAGGTGGTCATCGGGGCGTTCTACCTTGCCGGCGTGTTGGTTTACGCCACAAGGGTGCCGGAGAGGTTT
ATGCCAGGAATGTTTGACCTCGTCGGGCATAGCCACCAGCTGTTCCACGTGCTAGTCATTGCCGGCGCTTACGCGCA
CTACCATGCTGGCCTGGTTTACTTGAGTTGGCGGGACAGGGACCAGTGCTGA

>HlyIII-1.2 (partial_protein)

GVHADQAVALKNKGSPTCSDVAKPSCKHRYELVCYDALPAFLKHDEFILDHYRSEWPVKQAILSAFALHNETINVWT
HLIGFFIFLALTMSVWSGEHGPVTRWPFYTYLCGAMFCLLMSSGCHLLACHSEHATYVLLRLDYAGITGLIVTSFYP
VVYYTFLCDPFSRTLYLGSITICGAAAVAVSLLPVFQAPDLRWARAALFACMGASGIVPIVHKMLMFHSRPEAVLTT
GYEVVIGAFYLAGVLVYATRVPERFMPGMFDLVGHSHQLFHVLVIAGAYAHYHAGLVYLSWRDRDQC*

>PKS-2.1

ATGACAGGCAGCTCACCGAAAGTTAGTGAGATCCGGGGCGCGCAGCGTGCGGAAGGCTGCGCAGCCATGTTGGCTAT
TGGAACAGCAAATCCTGCGAACCAGGTGTCCCAAGAAGAGTACCCTTACTACTATTTCCGTGTCACAAAGTGCGAGC
ACCTTACTGACCATAAAGACACATTCAAGATAATATGTGGTCTCACGGGCACGGAGAACCGTTTCTTCTACCACACG
GATGAACTGCTCAACTCCCACCCTGTCTTGCTGGACCACACATCGCCGTCCCGCGAAGCTCGGCATGATATCGTCGC
CAAGGCTGCTCCAGGGCTTGCGGCATCAGCAGCAAAGAAGGCCATCGCAAAGTGGGGCCGTCCGGCCACTGACATCA
CCCACCTTATCGCCAGCACCAACTCTGACGCCGGCGCCCCAAGTGCGGATGTACGCTTGGTTTCACTCCTTGGCCTT
CGCGCCAACGTGTGCCGTACGATGCTCCATCTTAATGGCTGCTTCGCTGGCTGCTCCGCCTTGCGTCTAGCCAAGGA
CCTTGCTGAGAACAACCGTGGGGCACGCGTCCTCGTGGTTTGCGTGGAGCTCGCCATTTCTGGCTTCTGCAGCACCG
GCGAAGGGAACTGCTTAGACACCCTCATTACTCATGCGTTGTTTGGTGATGGGGCAGGCGCGGTTATCGTCGGCGCC
GATCCCATACACCCCGTTGAGAATCCTCTGTTTGAGATGGTGTCCGTCTCTCAGACCTTCGTGCCGAGCACCGAGCA
TGTGCTCACCCTGAACTTGGGGAGCCACGGCACTCATGGCAAAGTTTACACCAAACTGCCCACACTGGTAGCAGACA
CCATGGAACCTTGTCTTTCAGAAGCATTTGGTCGACTTGAGATGGACTTCAAATGGAATGACCTGTTCTGGGCTGTG
CACCCGGGCAGCCGTGGGATCTTGGACCAGCTTGACAAGACCCTCCAGTTGGATCCCACGAAGCTAGCGGCGAGCCG
AACTGTCGTACAAAAGTTTGGGAACATGTTTAGTGCCACCGTAATCTTCGTGCTTGACAAGCTGCGCCGGCGAATGG
AGGAGGAAGGAGAGCGAGCAGAGTGGGGGGCAATGGTAGCATTTGGACCAGGTTTCACTATCGAGACCATGGTGCTC
CACGCAACTAGTGCTCTGAAGGAAAATTAG

>PKS-2.1_protein

MTGSSPKVSEIRGAQRAEGCAAMLAIGTANPANQVSQEEYPYYYFRVTKCEHLTDHKDTFKIICGLTGTENRFFYHT
DELLNSHPVLLDHTSPSREARHDIVAKAAPGLAASAAKKAIAKWGRPATDITHLIASTNSDAGAPSADVRLVSLLGL
RANVCRTMLHLNGCFAGCSALRLAKDLAENNRGARVLVVCVELAISGFCSTGEGNCLDTLITHALFGDGAGAVIVGA
DPIHPVENPLFEMVSVSQTFVPSTEHVLTLNLGSHGTHGKVYTKLPTLVADTMEPCLSEAFGRLEMDFKWNDLFWAV
HPGSRGILDQLDKTLQLDPTKLAASRTVVQKFGNMFSATVIFVLDKLRRRMEEEGERAEWGAMVAFGPGFTIETMVL
HATSALKEN*

>P450-2.1

ATGGAGCTTAACGTAGTGCAAGCCATGGCAGCTGTCCTCATGGTGCTAGTAACCACGAGAGCGCTATGGTATCTGCT
TTGGAGGCCGTATGCCGTGGCCAGGTGGTTCCGACGGCAGGGCATCAGAGGTCCGCCTTACAATCTCCTGGTTGGGT
CCCTGCCGGAGTGCCAGAAGATGATTGTCGCCGGGAGAGCCAAGGATCTGGACACCATCTCCCACGACTGCATCACT
ACCGTGCAACCCTTCTTCCAGAAATGGGCCTCCATGTATGGGAAAACATTCCTATACTGGCTGGGGCCGACACCAGC
ACTGTGCTCTACGGACATGGAGCTGGTGAAGAAAGTGTTAACCGACAGGACGGACATGTTCCAAAAAGACTACTCAA
ATCCTAGCCTGGAAGCGGTTCTGGGGAATGGGGTTGTATTTGCAAACGGACACGACTGGAAGCGGCGGCGCAAATTC
ATCCACCCGGCTTTCAACCAAGAGAAGATCAAGTCCATGTCAGCAATAACGTTGGAGTGCACACAGCAGACAATGGA
ACGGTGGCGCAACAGGCAGCAAGCCGAGATAGACATGATGCACGACTCCGACGAGATAGCCATGAGTGTCATTGCGA
GAGTGATGTTGGGCAAGTGCTACAAGGAGGCCTGGGAGGTGTTCATCGCCGGAAAGGAGCAGCTGAAGCTAGCTACT
TATGCGTTTGCGGATCCTCCGGTACCTGGATTATGGTACCTGCCAACACGTCGCAACCGTCGGGCATGGTATCTTGA
CAAGCTTGTGAAACATAAGATCTCACAGATCATAGAGGCGCGACTCGCTAGCGGCGTCTACGAAGATGACCTGCTCG
GGCAGATGCTGCAGCTGCAGCTGCAAACATGTAGCAGCGGCAGCACCGAGACCCTGAGCACCGAGGAGATGGTTGGC
GAGTGCAGGACTTTCTTCGCGGCTGGGTACGAAACCAGTGCCAGCCTCATTACCTGGGCGATGTTCCTGCTCGCCAG

CTACCCACGGTGGCAGGAGATGGTCAGGGAGGAGGTCGTCCGGGAATATCCTGCTCACCGGTTACCCATAGGTGACG
CCCTCGGCAAGTTGAAGCTGGTATGCTACGCACGTACATACTCCCACCTTCTAAATTATATATGGTGCATATAA

>P450-2.1_protein

MELNVVQAMAAVLMVLVTTRALWYLLWRPYAVARWFRRQGIRGPPYNLLVGSLPECQKMIVAGRAKDLDTISHDCIT
TVQPFFQKWASMYGKTFLYWLGPTPALCSTDMELVKKVLTDRTDMFQKDYSNPSLEAVLGNGVVFANGHDWKRRRKF
IHPAFNQEKIKSMSAITLECTQQTMERWRNRQQAEIDMMHDSDEIAMSVIARVMLGKCYKEAWEVFIAGKEQLKLAT
YAFADPPVPGLWYLPTRRNRRAWYLDKLVKHKISQIIEARLASGVYEDDLLGQMLQLQLQTCSSGSTETLSTEEMVG
ECRTFFAAGYETSASLITWAMFLLASYPRWQEMVREEVVREYPAHRLPIGDALGKLKLVCYARTYSHLLNYIWCI*

>P450-3.1

ATGTCAATTCTGTTCTCCCAAGAGCTGCCCATCTACACAGCACTCGTGCTACTTGTTCCCCTGTACTGCCTGTACTC
AAGGGTTATATGTAGATCAGGGAACCCAGCACTCGTGCTCCCCACAAACTGGCCAATATTGCACATGTTTCCTTCCT
TCTTGGCCAACCTCCACAACTTGCACGAGTATTTAACCCTTGTCCTTGCCGGATCAGGGCACAACTTCAGGGCGCAT
GGCCCACCCGGGACCGGGTTGCGGTTCTTTGTCACATGCGACCCTACAAATGTCCGACACATTTTCACGACCAACTA
CACCAACTTCCCAAAGGGCGCGGAGTTTGCCGATATCTTTGACATCGTGGGTGGCAGCCTCTTTACCATCGACGGTG
AGCCGTCTCATCAACAGCGCGCAAAAGTCAAGGGCGCGCTCAGCAAACCGCGGATGGCTGCCACTATGGCGGCCTGC
TGCTGTGACAAGGTGGAGAACGGCCTCCTCCCGTTTTTTAGGCGAATGGCGAGCACTGGCACTCCATTTGACGTGCA
AGAACTGATGTCAAGGTTTATGTTTGACCTGGCTGCTATGCCTTTCTTTGGCGTGGATCCTGGCCTCCTGTCCTCGG
AAATGCTACAAGACAGGCTGCACATGGACGCTGCGGTTGCCCTTGACACGTTCATGGAGGTGGGATTTTCCCTGCTC
ATGACACCAAGTTCTTATTGGAAGTTGATGAGGTGGCTAAACATTGGCCCTGAGAGAAAGCTCAAAGTGGCGCGCAC
GGTGCTACGGGAGTTCGCCGCGGAGATGATGGAGAGGAGGAAGATGAATACGTGTTTTGTTGGTAATGAAGATGAAC
AAGAGGATGGGGATATCTTGTCTTGCTTCCTCAATGACCCAGACTACGCCGATGATGACTTGCTCCGTGCGATGATC
ATTGGCTACATGTTCGCTGCTAGGGATACAGTTGGAACAACCCTGACATGGATCTTCTACAAACTCGCCCAGAACCC
TAACATTGTTTCGAATATCCGAAAAGAACTCTCACCCATTGCATCACGCAAAGAAGCGGTTGGTGTGGATGCCATGT
TGATCTTTGAGCCAAAAGAGACCAGATCTCTAGTATATTTGAAAGCCGTCCTGTACGAGACTCTCAGGCTATACCCA
GCAGCGCCTCTTGAGTGCAAGACAGTAGTCGCCGATGATATCATGCCAAGTGGACATGAGGTTCATGCCGGCGACAC
CATTATTATTTCTATACACTCCATGGGGAGAATGGAGGGTGTGTGGGGTGAAGATTGCCTCCATTATAAACCAGATA
GGTGGCTCTTGGTGGGTGGCAATAACATGAGGTATGTACCATCTCACAAATTCTTGGCCTTCAACTCAGGCCCGAGG
ATGTGCCTCGGCAAGGACATCGCAATTATGCAGATGAAGACAGTCATCGCATCAACGCTATGGAACTTCGATGTGGA
GGTGATGGACGGGCAAACAATCCAGCCCAAGCAAGCTTGTATACAGCAGATGAAAAATGGGCTCATAGTTAAGCTAA
AGAAGCGGGAAATGTAA

>P450-3.1_protein

MSILFSQELPIYTALVLLVPLYCLYSRVICRSGNPALVLPTNWPILHMFPSFLANLHNLHEYLTLVLAGSGHNFRAH
GPPGTGLRFFVTCDPTNVRHIFTTNYTNFPKGAEFADIFDIVGGSLFTIDGEPSHQQRAKVKGALSKPRMAATMAAC
CCDKVENGLLPFFRRMASTGTPFDVQELMSRFMFDLAAMPFFGVDPGLLSSEMLQDRLHMDAAVALDTFMEVGFSLL
MTPSSYWKLMRWLNIGPERKLKVARTVLREFAAEMMERRKMNTCFVGNEDEQEDGDILSCFLNDPDYADDDLLRAMI
IGYMFAARDTVGTTLTWIFYKLAQNPNIVSNIRKELSPIASRKEAVGVDAMLIFEPKETRSLVYLKAVLYETLRLYP
AAPLECKTVVADDIMPSGHEVHAGDTIIISIHSMGRMEGVWGEDCLHYKPDRWLLVGGNNMRYVPSHKFLAFNSGPR
MCLGKDIAIMQMKTVIASTLWNFDVEVMDGQTIQPKQACIQQMKNGLIVKLKKREM*

>PKS-2.2 (partial)

ATGACAGGCAGCTCACCGAAAGTTAGTGAGATCCGGGGCGCGCAGCGTGCGGAAGGCTGCGCAGCCATGTTGGCTAT
TGGAACAGCAAATCCTGCGAACCAGGTGTCCCAAGAAGAGTACCCTTACTACTATTTCCGTGTCACAAAG

>PKS-2.2 (partial) _protein

MTGSSPKVSEIRGAQRAEGCAAMLAIGTANPANQVSQEEYPYYYFRVTK

>P450-2.2

ATGGAGCTTAACGTAGTGCAAGCCATGGCAGCTGTCCTCATGGTGCTAGTAACCACGAGAGCGCTATGGTATCTGCT
TTGGAGGCCGTATGCCGTGGCCAGGTGGTTCCGATGGCAGGGCATCAGAGGTCCGCCTTACAATCTCCTGGTTGGGT
CCCTGCCGGAGTGCCAGAAGATGATTGTCGCCGGGAGAGCCAAGGATCTGGACACCATCTCCCACGACTGCATCACT
ACCGTTCAACCCTTCTTCCAGAAATGGGCCTCCATGTATGGGAAAACATTCCTATACTGGCTGGGGCCGACACCAGC
ACTGTGCTCTACGGACATGGAGCTGGTGAAGAAAGTGTTAACCGACAGGACGGACATGTTCCAAAAAGACTACTCAA

ATCCTAGCCTGGAAGCGGTTCTGGGGAATGGGGTTGTATTTGCAAACGGACACGACTGGAAGCGGCGGCGCAAATTC
ATCCACCCGGCTTTCAACCAAGAGAAGATCAAGTCCATGTCAGCAATAACGTTGGAGTGCACACAGCAGACAATGGA
ACGGTGGCGCAACAGGCAGCAAGCCGAGATAGACATGATGCACGACTCCGACGAGATAGCCATGAGTGTCATTGCGA
GAGTGATGTTGGGCAAGTGCTACAAGGAGGCCTGGGAGGTGTTCATCGCCGGAAAGGAGCAGCTGAAGCTAGCTACT
TATGCGTTTGCGGATCCTCCGGTACCTGGATTATGGTACCTGCCAACACGTCGCAACCGTCGGGCATGGTATCTTGA
CAAGCTTGTGAAACATAAGATCTCACAGATCATAGAGGCGCGACTCGCTAGCGGCGTCTACGAAGATGACCTGCTCG
GGCAGATGCTGCAGCTGCAGCTGCAAACATGTAGCAGCGGCAGCACCGAGACCCTGAGCACCGAGGAGATGGTTGGC
GAGTGCAGGACTTTCTTCGCGGCTGGGTACGAAACCAGTGCCAGCCTCATTACCTGGGCGATGTTCCTGCTCGCCAG
CTACCCACGGTGGCAGGAGATGGTCAGGGAGGAGGTCGTCCGGGAATATCCTGCTCACCGGTTACCCATAGGTGACG
CCCTCGGCAAGTTGAAGCTGGTATGCTACGCACGTACATACTCCCACCTTCTAAATTATATATGGTGCATATAA

>P450-2.2_protein

MELNVVQAMAAVLMVLVTTRALWYLLWRPYAVARWFRWQGIRGPPYNLLVGSLPECQKMIVAGRAKDLDTISHDCIT
TVQPFFQKWASMYGKTFLYWLGPTPALCSTDMELVKKVLTDRTDMFQKDYSNPSLEAVLGNGVVFANGHDWKRRRKF
IHPAFNQEKIKSMSAITLECTQQTMERWRNRQQAEIDMMHDSDEIAMSVIARVMLGKCYKEAWEVFIAGKEQLKLAT
YAFADPPVPGLWYLPTRRNRRAWYLDKLVKHKISQIIEARLASGVYEDDLLGQMLQLQLQTCSSGSTETLSTEEMVG
ECRTFFAAGYETSASLITWAMFLLASYPRWQEMVREEVVREYPAHRLPIGDALGKLKLVCYARTYSHLLNYIWCI*

>P450-3.2 (pseudogene)

ATGTCAATTCTGTTCTCCCAAGAGCTGCCCATCTACACAGCACTCGTGCTACTTGTTCCCCTGTACTGCCTGTACTC
AAGGGTTATATGTAGATCAGGGAACCCAGCACTCGTGCTCCCCACAAACTGGCCAATATTGCACATGTTTCCTTCCT
TCTTGGCCAACCTCCACAACTTGCACGAGTATTTAACCCTTGTCCTTGCCGGATCAGGGCACAACTTCAGGGCGCAT
GGCCCACCCGGGACCGGGTTGCGGTTCTTTGTCACATGCGACCCTACAAATGTCCGACACATTTTCACGACCAACTA
CACCAACTTCCCAAAGGGCGCGGAGTTTGCCGATATCTTTGACATCGTGGGTGGCAGCCTCTTTACCATCGACGGTG
AGCCGTCTCATCAACAGCGCGCAAAAGTCAAGGGCGCGCTCAGCAACCCGCGGATGGCTGCCACTATGGCGGCCTGC
TGCTGTGACAAGGTGGAGAACGGCCTCCTCCCGTTTTTAGGCGAATGGCGAGCACTGGCACTCCATTTGACGTGCAA
GAACTGATGTCAAGGTTTATGTTTGACCTGGCTGCTATGCCTTTCTTTGGCGTGGATCCTGGCCTCCTGTCCTCGGA
AATGCTACAAGACAGGCTGCACATGGATGCTGCGGTTGCCCTTGACACGTTCATGGAGGTGGGATTTTCCCTGCTCA
TGACACCAAGTTCTTATTGGAAGTTGATGAGGTGGCTAAACATTGGCCCTGAGAGAAAGCTCAAAGTGGCGCGCACG
GTGCTACGGGAGTTCGCCGCGGAGATGATGGAGAGGAAGATGAATACGTGTTTTGTTGGTAATGAAGATGAACAAGA
GGATGGGGATATCTTGTCTTGCTTCCTCAATGACCCAGACTACGCCGATGATGACTTGCTCCGTGCGATGATCATTG
GCTACATGTTCGCTGCTAGGGATACAGTTGGAACAACCCTGACATGGATCTTCTACAAACTCGCCCAGAACCCTAAC
ATTGTTTCGAATATCCGAAAAGAACTCTCACCCATTGCATCACGCAAAGCAGCGGTTGGTGTGGGTGCCATGTTGAT
CTTTGAGCCAAAAGAGACCAGATCTCTAGTATATTTGAAAGCCGTCATGTACGAGACTCTCAGGCTATACCCAGCAG
CGCCTCTTGAGTGCAAGACAGTAGTCGCCGATCATATCATGCCAAGTGGACATGAGGTTCATGCCGGCGACACCATT
ATTATTTCTATACACTCCATGGGGAGAATGGAGGGTGTGTGGGGTGAAGATTGCCTCCATTATAAACCAGATAGGTG
GCTCTCGGTGGGTGGCAATAACATGAGGTATGTACCATCTCACAAATTCTTGGCCTTCAACTCAGGCCCGAGGATGT
GCCTCGGCAAGGACATCGCAATTATGCAGATGAAGACAGTCATCGCATCAACGCTATGGAACTTCGATGTGGAGGTG
ATGGACGGGCAAACAATCCAGCCCAAGCAAGCTTGTATACAGCAGATGAAAAATGGGCTCATAGTTAAGCTAAAGAA
GCGGGAAATGTAA

>P450-3.2 (pseudogene) _protein

MSILFSQELPIYTALVLLVPLYCLYSRVICRSGNPALVLPTNWPILHMFPSFLANLHNLHEYLTLVLAGSGHNFRAH
GPPGTGLRFFVTCDPTNVRHIFTTNYTNFPKGAEFADIFDIVGGSLFTIDGEPSHQQRAKVKGALSNPRMAATMAAC
CCDKVENGLLPFLGEWRALALHLTCKN*

>PKS-3

ATGGGAAGCACCACCCTACCATCGGTGGGTGAGATCCGGCGTTTGCAGCGCGCGGAGGGGCCTGCAGCCATCCTCGC
CGTCGGCACGGCAAACCCGCCCAACTGCGTTTCCCAGGAGGAGTACCCCGACTACTACTTCCGCATCACCAAGAGCC
AACACCTCACAGACCTCAAACAAAAACTCAAGTCCTTCTGCCAGATGACCTCAACGGAGAAGCGCTACTTCCACCAC
ACGGAAGAGCTTCTGGAAGACCACCCCAATTTCTTCTGCCGCGGCATACCGTCCCTAGACGCCCGGCTGGAAATCGC
TGCTGCTGCTGCCCCGGAGCTGGCGGCGTCAGCAGCAGCCAAGGCCATAGCGAAGTGGGGCCGTCAGGCCACCGACA
TCACCCACCTCGTCGTCGGCACCAACTCTGGTGCGCAAGCCCCGGCCGTCGACCTCCGCCTGGCCTCTCTGCTCGGC
CTGCGCGCGTCCGTCTGCAGGACGATGCTCAACCTCAACGGCTGCTCCGCCGGTGCCGCCTCGCTCCGTCTCGCCAA
GGACCTGGCCGAGAACAACCGCGGCGCGCGCGTGCTGGTGGCCTGCGTCGAGCTCACCGTCGTCGCCTTCCACGGGC
CCGAGGAGGCGTACCCACACAAGCTCATCAGCCAGGCTGTCTTTGGCGACGGTGCGGGCGCGGTCATTGTCGGAGCT

GACGCCGTGCACCCCATCGAGCGCCCGCTCTTCGAGATGGTGTCGGCCTCACAGACCACGATACCAGCAACCGACGG
CGTGCTCACCATGCAGCTCACGGAAGCCGGCCTCGACGGCCACATCTTCACCAAGGAGCTAATTCCTCTAGCAGCGC
AGCACATCGAGCAGTGTCTCATGGACGCGTTCCAGCAGCTTGGAATAATGAACGGGGGCACGAACTGGAACGATCTC
TTCTTTGTGGTGCACCCTGGCACCCATGGAATAATGGACCACATCCAACGGGCTCTCCGGCTAGATCCTGGGAAGCT
GGCCGCTAGCCGGACTGTGCTGAGTGAGTACGGGAACATGCTTGGTGCCACACTGATCTTCGTGCTCGACGAGCAAC
GGCGGCAGATGGAGGAGGACGGAGAGACGGGTGAGTGGGGTGTGATGATGGGATTTGGACCAGGGTTCACAGTTGAG
ACCATGCTGCTACATGCGGTGGCCAGAAACCTGCACAACAAAAATTGA

>PKS-3_protein

MGSTTLPSVGEIRRLQRAEGPAAILAVGTANPPNCVSQEEYPDYYFRITKSQHLTDLKQKLKSFCQMTSTEKRYFHH
TEELLEDHPNFFCRGIPSLDARLEIAAAAAPELAASAAAKAIAKWGRQATDITHLVVGTNSGAQAPAVDLRLASLLG
LRASVCRTMLNLNGCSAGAASLRLAKDLAENNRGARVLVACVELTVVAFHGPEEAYPHKLISQAVFGDGAGAVIVGA
DAVHPIERPLFEMVSASQTTIPATDGVLTMQLTEAGLDGHIFTKELIPLAAQHIEQCLMDAFQQLGIMNGGTNWNDL
FFVVHPGTHGIMDHIQRALRLDPGKLAASRTVLSEYGNMLGATLIFVLDEQRRQMEEDGETGEWGVMMGFGPGFTVE
TMLLHAVARNLHNKN*

>Hyd-2

ATGCAGGAAAGCAAGAGTTCTCCGGTACTTAGCAAGAAGGATGACGGCGAGGAGGGCATCACCGTCGACCTGTATCC
ATTCATACGGGAATACAAGGTCGGCCGTGTGGAGCGCTTACTGCACAGCCCATTCGTGGCAGCGTCCAAGGACCCGG
CAGCAAACCGTGGAGTTGCGACCAGGGACGTGGTCGTCGACGAGAGCACCGGCGTGTTCGCACGCCTGTTCCTTCCC
TCCATTGCAGCCGCCGCTACGGGGATCGGCGGCGAGAGGCTTCCCGTCCTCATGTACGTCCACGGCGGATCCTTCTG
CACGGGGAGCGCGTTCTCTCGCACTTACCACAGCTACACCAGGTCTCTGGCGGCCAGCGCCGGGGCGCTCGTCGTGT
CCGTGGAGTACCGTCTCGCGCCGGAGCATCCTGTGCCTGCAGCCTATGAAGATGCATGGGCGGCGCTCCAGTGGGTG
GTGTCCCTCTCTGATCCGTGGCTGTCAGACTACGCAGACCTCGGGCGCACGTTCCTCGTTGGCGACAGTGCCGGCGG
GAATATTGTCTACAACACGGCGGTGCGTGCGGCCAGTGGCGGTGGCAGCCACATCCACATCGAAGGGCTGGTCATCG
TGCACCCATACTTCTGGGGGGTCGAGCGGCTGTCCAGCTCCGAGGCAGTCTGGGACGGCATCGCCATGTTTGCTCCC
GAGGACGTCGACAGGCTCTGGCCGTTCATTACGGCCGGCCGGCTGGGAAACGACGATCCCTTGGTCAACCCCGTCGA
CGAGGAGATCGCCTCGCTGACATGCCGGCGTGTGCTGGTGGCCGTGGCCGAGAAGGACACCTTGCGTGACCGCGGGC
GTCGGCTGGCAGCTCGCATGCGCGACTGCTGCTCATGGGCCGATGATGAAAACGCGGTGACATTGATAGAATCAAAG
GGAGAAGACCATGGCTTCCACCTGTACAACCCACTGCGTGCGACCAGCAAGATCCTCATGGAGAGCATCGTACGGTT
CATCAACCAGCGCACACCCTTGCCGCTGCCAAAACTTCACGAGCTGCATGCATGCCAAGGTAAGAGGAAGGCATCGT
CTCCGGGTGACATGGACACAGCCGTCCAGCATGTTCTAGGCGTGCCAACCAGGCCGTATATAGACGTACATGGCTAT
GGGACGTCCATGAGAGTCACAGATGTGACACGTAGTAGTTGCTTACGCATTGGAAATGGAAGAAAAGCATCCAAAAC
AAGTTCTCGTCATCAGGCCAAACATCACCAACATGAGGTTCTCGTCATCAGCAACAACAGCTCCTGA

>Hyd-2_protein

MQESKSSPVLSKKDDGEEGITVDLYPFIREYKVGRVERLLHSPFVAASKDPAANRGVATRDVVVDESTGVFARLFLP
SIAAAATGIGGERLPVLMYVHGGSFCTGSAFSRTYHSYTRSLAASAGALVVSVEYRLAPEHPVPAAYEDAWAALQWV
VSLSDPWLSDYADLGRTFLVGDSAGGNIVYNTAVRAASGGGSHIHIEGLVIVHPYFWGVERLSSSEAVWDGIAMFAP
EDVDRLWPFITAGRLGNDDPLVNPVDEEIASLTCRRVLVAVAEKDTLRDRGRRLAARMRDCCSWADDENAVTLIESK
GEDHGFHLYNPLRATSKILMESIVRFINQRTPLPLPKLHELHACQGKRKASSPGDMDTAVQHVLGVPTRPYIDVHGY
GTSMRVTDVTRSSCLRIGNGRKASKTSSRHQAKHHQHEVLVISNNSS*

>Wax ester synthase (WES)

ATGGACCCTGGCAGTGGTAGTCCTTGGGCATCAGCTCAAAGAAACCGACTGCTCCCGATCCGCTTGTCAAGAGAACC
TGCCAACACTGAGTGGACGCCGCCCGCCGGTTCTCCGGCGGAGGAGCCGGTGAGCCCTACAGCAAGAGCCATGGAAG
ATATAGGTATTTATATCGTCGTCACGGTCGGCCTCGACACGCCGATAAACCTGACCACCTTCCGAGCCGGCATCGAA
GCTATGCTTGCCCGATGCCCGCGCTATGGATGCATTCATGTGGCGGATGGGCTAAACAATGGCGAGGCGCGATGGGC
GCGCACTACTGTGAACGTCCAAGATCACATGATCGTGCCGAGGCTGGACCGCGCGACCGACCCAGACAAGGCCGTGG
AGGACTACGTCGCCTCGCTGTCCACGCTCCCCATGGTCCGCTCCCGCCCGCCATGGGAGTTCCACTTCCTCGACTTC
CCGACCTCCGAGGCGGCCTCCACGGTGGCGATCCGCGTGCACCACGCCTACGGCGATGGTATGTCCACCATGGCGCT
TCTCATGATGTCCACGCGCAGCGCCATCACCGACACCAAAAGTCATCCAGCCGCGCCTCCGCATCGTCGTCCTACGC
GCACGAGCGCCATATATGCCCCTCAACGCCCCCCACTTTCCGCGGGCGCCCTGGCGTTCGTCGCATGGGTCTGGTCG
TATCTTGTGCTGGCGTGGAACACCGCGGCAGACATTGCCTACTTCGCCGCCACGATTATGTTCTTGAGTGACCCGCG
CACTCTCTTCAAGCATGCCGACGACGGGTTCCACGCCAAGCGCTTTGTCCACCGGAGTCTTAGCCTCGATGACGTCA
AGTTCCTCAAGAACTCCATGAACTGTACTGTCAATGATGTGTTGGTTGCACTGACTTCTGCTGCTCTATCACGATAT

TACTTCCGCAAGTCCGAATCTGATAAGAGGAACGAGGTGACATGGGGAAATAAATTGGGCTACATCATTCTTCCATT
TTATCTAGCCAGACACGATGATCCACTTGCATACATTCACAAAGCAAAGAAGGTTCTGGATAGGAAAAAGAGGTCCT
TAGAAGTGATATTCACATATAAGATTGCTGAGATCTTCATGAATGTTTTTGGTGCCAAGTCATCATTGAACCAGGTA
GGAACTTCTATCTTCCGGTGTCTCTTTGCACGTACAACAATTGTTTTCTCAAACATGGTTGGACCAACCGAACAAGT
AGAATTATGTGGACACCCAGTGGCTTTCATTGCACCTAGTGTCTATGGGATTCCAGAGACTGTGAACTCTACTATGA
TGACCCTTCTTCAGTGA

>Wax ester synthase (WES) _protein

MDPGSGSPWASAQRNRLLPIRLSREPANTEWTPPAGSPAEEPVSPTARAMEDIGIYIVVTVGLDTPINLT
TFRAGIEAMLARCPRYGCIHVADGLNNGEARWARTTVNVQDHMIVPRLDRATDPDKAVEDYVASLSTLPM
VRSRPPWEFHFLDFPTSEAASTVAIRVHHAYGDGMSTMALLMMSTRSAITDTKSHPAAPPHRRPTRTSAI
YAPQRPPLSAGALAFVAWVWSYLVLAWNTAADIAYFAATIMFLSDPRTLFKHADDGFHAKRFVHRSLSLD
DVKFLKNSMNCTVNDVLVALTSAALSRYYFRKSESDKRNEVTWGNKLGYIILPFYLARHDDPLAYIHKAK
KVLDRKKRSLEVIFTYKIAEIFMNVFGAKSSLNQVGTSIFRCLFARTTIVFSNMVGPTEQVELCGHPVAF
IAPSVYGIPETVNSTMMTLLQ*

>Ank_PGG

ATGCAGATGGACAGACGGCTCCTGAAAGCAGCCACATCGGGTGACTCGGCGTCGATGCAGGCCATGGCTTCGCAGGA
TCTAAGCATTCTTCTCGGAACAACCCCGCAGGGGAACACTTGTCTCCACATATCCTCCATCCATGGCCACGAGGGGT
TCTGCGTGGACGTGATCGCACTAGAGGAGTCCCTGCTCGCCCAGACCAACCTGGACGGCGAGACGCCGCTTGTCGCC
GCCGTGAGAAGCGGCTCCGTGGCCGTGGCTTCCGCCTTGCTCGGACGCTACCGAGTGCGGCGGTTGAGCGAGGCCAT
CTTGGAGAAAGACAAGGACGGATGCAGTGCGCTCCACCATGCCATCTGCAGCGGCGACACGGAGCTCGCTCTGGAGC
TGATAGAAGCAGAGCCCGCCCTCTCCAAAGGTGTCAACAACTTCGGCGACCCGCCAATGTACGTTGCTGCCGCCAGA
GATCTTACTAGCGTTGCCCTGAGACTGTTGGAGATTGGCGATTCTGATCACGGGGGACAGTTCGGCTACAATGCTCT
CCATTCAGCAGTCAGAAATGGGAACTCAGTTATCGTTAAAAGAATCATGGCCAAACGTCCTGGGCTGGCCAAAGAAG
CTCACAAGCGTGGGTATACTCCAATGTGCACGGCCATGTACCGTGACAAGATCGACATGCTACGGCTGCTGCTGGAG
CATGATCCCGCTCTAGGGTATGACATGACGAGCGATGGTTATTCTCTCCTTCAGGTTGCTGCACACGCAGGTCATGT
TGCCACTGCACGCGAGCTGCTTAAGCACTGCCCAGATGCTCCTTGCCGCGGAGCACGCGTTGAGCGTTGGACGTGCC
TCCACACATCCGTATACTATGGTCATCTGGACTTTGTGAAATTTGTCCTCGCAACGCCACAGCTTGGCAAACTGGTC
AACATGCAGGACAACAAAGGAAGAACTGCTCTGCATCTCGCTGTGGAGAACTGTGATCCAAAAGTGGTTGCTGCTTT
ACTGTCTCATAAGGATACACACACACATGTGATGAACAACAACGGCACGTCACCAGCTTGGGTATTGTCAGGAATCA
TGCATCGTGCCACGACGTTAAACTGGAATGAAGTGATGATGCTTATGCTGAGAGCTGATCCCCAAAGTGCCCCTTCA
GCTTATAATCTTCATGTGCACACAACACGGCTGTTGACCGATGCATCACGGATGGATGCAAAGTCACTAACTAAAAC
ATACACAACCAACACCTCATTAGTGGCGATCCTCATTACAACAATCACCTTTGCCGCTGCTTTCACCCTCCCCGGAG
GATACAGCAGTGCTACTGGAAGCGAGGGACTTCCCATCATGTCTCAAAAGGCTGCATTCAAGGCATTCATAATCTCT
GACACCTTAGCAATGTGCTCTTCATTTGTTGTCGCCTTTATATGCATCGTAGCAAGGTGGAAGGATTATGAGTTCTT
GATCTATTACAGATCCTTCACTAGGAAGCTCATGTGGTTTGCATACGTGGCAACAACTACGGCCTTTTCAACTGGTT
TATACACAGTGCTGGCTCAACGTCTCCACTGGCTATCCATATCAATTTGTGTTTTGGTAGCTTTGTTGCCCATTCTT
ACTAAGTTGCTGGGCGAATGGCCTTACTTGAAGATCAGATTCGGCTTAGGTAGAACTTTCAGCTCCGATCTCCTTGA
CATGGTGTGA

>Ank_PGG_protein

MQMDRRLLKAATSGDSASMQAMASQDLSILLGTTPQGNTCLHISSIHGHEGFCVDVIALEESLLAQTNLDGETPLVA
AVRSGSVAVASALLGRYRVRRLSEAILEKDKDGCSALHHAICSGDTELALELIEAEPALSKGVNNFGDPPMYVAAAR
DLTSVALRLLEIGDSDHGGQFGYNALHSAVRNGNSVIVKRIMAKRPGLAKEAHKRGYTPMCTAMYRDKIDMLRLLLE
HDPALGYDMTSDGYSLLQVAAHAGHVATARELLKHCPDAPCRGARVERWTCLHTSVYYGHLDFVKFVLATPQLGKLV
NMQDNKGRTALHLAVENCDPKVVAALLSHKDTHTHVMNNNGTSPAWVLSGIMHRATTLNWNEVMMLMLRADPQSAPS
AYNLHVHTTRLLTDASRMDAKSLTKTYTTNTSLVAILITTITFAAAFTLPGGYSSATGSEGLPIMSQKAAFKAFIIS
DTLAMCSSFVVAFICIVARWKDYEFLIYYRSFTRKLMWFAYVATTTAFSTGLYTVLAQRLHWLSISICVLVALLPIL
TKLLGEWPYLKIRFGLGRTFSSDLLDMV*