Supplementary:

Plastic-hydrolytic enzyme classification using explainable deep learning

Woo-Haeng Lee[1+], Louis Dumontet[2+], KyungMin Jung[3], Hyun Lee[3,4], Gobinda Thapa[1], Tae-Jin Oh[1,4,5*], and Mingon Kang[2*]

[1] Department of Life Science and Biochemical Engineering, SunMoon University, Asan, Republic of Korea 31460

[2] Department of Computer Science at the University of Nevada, Las Vegas, NV, USA 89154

[3] Department of Computer Science and Engineering, SunMoon University, Asan, Republic of Korea 31460

[4] Genome-based BioIT Convergence Institute, Asan, Republic of Korea 31460

[5] Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan, Republic of Korea 31460

[+]The first authors: Woo-Haeng Lee and Louis Dumontet
*Corresponding authors: Tae-Jin Oh and Mingon Kang

Supplementary Note 1. Transformer architecture.

We considered an encoder-only transformer for the benchmark. It is composed of an embedding module followed by four encoders. The embedding module relies on two layers: the amino acid matrix embedding and the positional encoding. The amino acid matrix embedding is a $23 \times d_{model}$ ($d_{model} = 256$) matrix learned during training such that the amino acid (assigned index a) has its embedding in the a-th row of this matrix. Note that we need an embedding for the 20 different amino acids: the undetermined amino acid often referred to X, the classification token, as well as the padding token which account for the 23 rows of our matrix. The positional encoding module takes the position p of the amino acid as input and encodes it as:

$$PE(p, 2i) = sin(\frac{p}{10000^{2i/d_{model}}}), \tag{1}$$

$$PE(p, 2i + 1) = cos(\frac{p}{10000^{2i/d_{model}}}), \tag{2}$$

where 2i and 2i+1 represent even and odd indices, respectively. The outputs of the two layers are then added to form the output of our embedding module, which is fed to the first encoder.

Each encoder comprises a multi-headed attention module followed by two fully connected layers activated by the ReLU function, a dropout layer, and a layer normalization. The multi-headed attention layer consists of projecting the input into 3 vectors Q (query), K (key), and V (value) of size $d_k = 256$. The output of the attention module is computed as:

$$Attention\ Scores = Softmax(\frac{Q \cdot K^T}{\sqrt{d_k}}), \tag{3}$$

$$Attention\ output = Attention\ Scores \cdot K. \tag{4}$$

Then, the output is fed to a dropout layer with a dropout rate of *0.1*. The multi-headed attention module incorporates a skip connection, which adds the input of this module to its output. The given result is normalized and fed to two successive position-wise feed-forward layers of size $4 * d_k = 1024$ in which the first one uses ReLU as an activation function, and both layers apply a dropout of 0.1 to their output. These two layers are skipped through a skip connection that adds the module's input to its output.

The final output of the encoder is passed through a dropout layer with a dropout rate of 0.1. After going through the embedding module and the four encoders, the final prediction is made by a linear layer, which takes as input the representation of the classification token by the last encoder of the transformer and uses the sigmoid function as activation to produce a score between 0 and 1 for each of the 11 classes.

Supplementary Note 2. Predictive performance comparison of CNNs across various encoders.
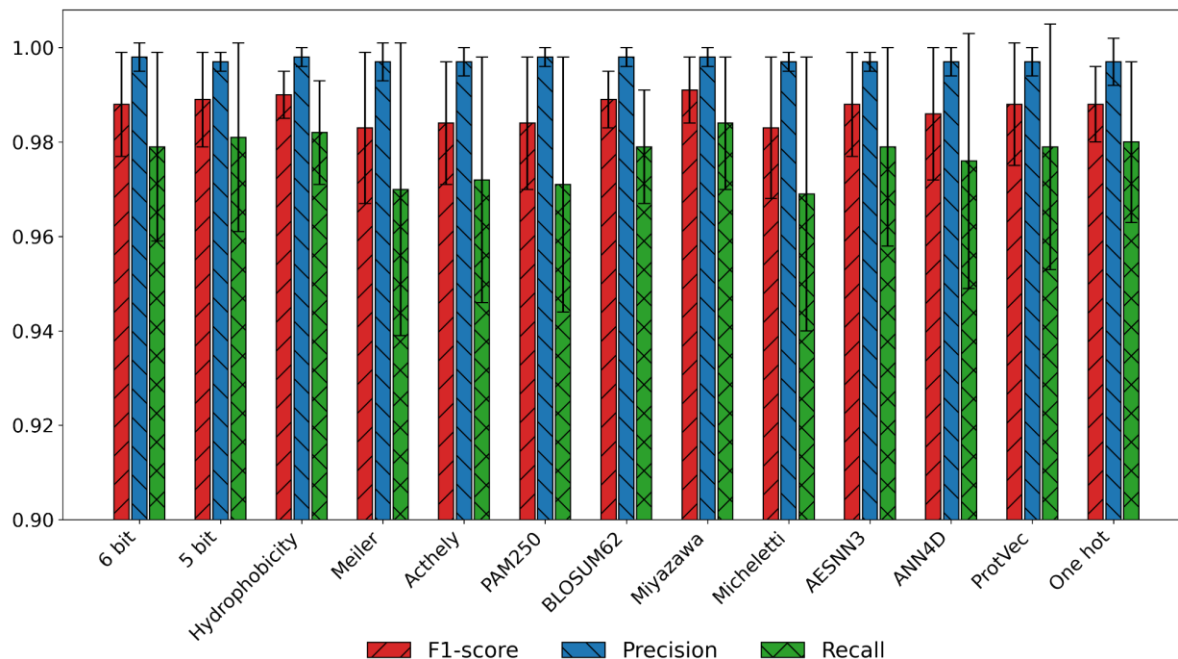

Fig. S1. Predictive performance of convolutional neural networks on the cross-validation dataset across multiple encoding methods

We evaluated the predictive performance of thirteen amino acid encoding methods using PEPIC on the cross-validation dataset (Fig. S1). Performance was assessed on the validation sets using micro-averaged F1-score across twenty repetitions to ensure statistical robustness. Among the tested encodings, Miyazawa energies achieved the highest micro-averaged F1-score (0.991 ± 0.007), followed closely by hydrophobicity matrix (0.990 ± 0.005), binary 5-bit encoding (0.989 ± 0.010), and BLOSUM62 (0.989 ± 0.006). Other encoding methods, such as one-hot (0.988 ± 0.008), AESNN3 (0.988 ± 0.011), and PAM250 (0.984 ± 0.014), also demonstrated competitive performance. Despite slight variations in the results, none of the differences were statistically significant. Given its strong performance, we adopted the Miyazawa energies encoding for the remainder of this study.

Supplementary Table 1. Sequences for interpretation of plastics-hydrolyzing enzymes.

Class	Accession number (Nr)	Sequence
PET	MAA58622.1	MNTYLLRTLSICLFAGLFSMQVQAITPDPEPDPDPDPPSTCSNCYQRGPNPT VSALEADSGPYSVRTINVSSWVSGFGGGTIHYPVGTEGTMGAIAVIPGYVS YERSIKWWGPRLASWGFVVITTDTNTIYDQPDSRADQLSAALDYVISQSN SRSPIYGMVDANRLGAMGWSMGGGGTLKLSTERELKAAIPQAPYYAGFNP FDEITTPTLIIACELDVVAPVAQHASPFYREIPGSTAKAFLEINGGDHFCANS GYPDEDILGKYGIAWMKRFIDEDRRYDQFLCGPNHEADRSISEYRDTCNY
	RLP53020.1	MLEGRYMKTVRFNTAAAVFTSALLSSQVFAITDDPVDPVDPVDPPSSGTVR GPDPTLSALESTRSGPYSVRTENVSNLSASGFGGGTIHYPTNAGENMGAIAV IPGYVSYESSIEWWGPRLASWGFVVITIDTNTIYDQPDSRADQLSAALDHLI DESGSSSSPISGLVDASRLGVIGWSMGGGGTLKLATERNLKAIIPQAPWYSG FNSFDRITTPTMIIACESDAIAPVGQHASPFYNDIPNSTAKAFLEINGGSHYC ANSGYSDEDILGKYGISWMKRFMDNDTRYSQFLCGPNHESDRSISEYRDTC NY
	WP_0210188 94.1	MNVLTKCKLALGIIAIFFSLPSFAVPCSDCSNGFERGQVPRVDQLESSRGPYS VKTINVSRLARGFGGGTIHYSTESGGQQGIIAVVPGYVSLEGSIKWWGPRL ASWGFTVITIDTNTIYDQPDSRASQLSAAIDYVIDKGNDRSSPIYGLVDPNR VGVIGWSMGGGGSLKLATDRKIDAVIPQAPWYLGLSRFSSITSPTMIIACQA DVVAPVSVHASRFYNQIPGTTPKAYFEIALGSHFCANTGYPSEDILGRNGVA WMKRFIDKDERYTQFLCGQNFDSSLRVSEYRDNCSYY
	WP_0773153 88.1	MKFLIKVNFILSIFAVFISPQIFAATVACSDCSNGFQRGTLPRVDQLESSRGP YSVKTSNVSVFARGFGGGTIHYSTDSGGQQGIIAVIPGYVSYESSIKWWGPR LASWGFTVITINTNTIYDQPNSRANQLSAAIDYVIDKGNDRSSPIYGLVDPER VGVIGWSMGGGGTLKLATDRDIDAIIPQAPWYSGLNSFSRITTPTMVIACQA DAVAPVALHASIFYNQIPRSTPKAFFEIAAGSHFCGNSGYPNEDILGRNGVA WMKRFIDNDTRYNQFLCGQNFDRSLRVSDYKDTCNTY
	RMH89651.1	MKLKAYLARITSLVTVSLLASSLAYAAPGPSAPCADCSRGPNPTVASLQSRS GPFTVSTFSVSGYLRGFGNSTVHYPTNATGKMGAIAVIPGYLSYEDSIRWW GPRLASHGFVVITMNTNTIYDQPDSRATQLSRALDYVIEQSNSRSSPISGKV DSTRLGAIGWSMGGGGSLKLSTERSLNAIIPQAPYYAGLNRFNTINTPTMIL ACSADVVAPVGSHASPFYNRIPEATPKAFLEIYGGSHFCANSGYPNEDLLG MYGISWMKRFIDFDSRYSQFLCGPNHAADLRISEYRENCNY
	OUS39971.1	MPNPNPAPCQEDCDFTRGPDPTISSLEASAGPYSVANQGVSRSVDGFGGGTI FYPMNTTGTMGAIAIAPGFLAGESSIEWWGPRFASHGFVIITIATNSVFDQPN SRETQLSSALDYVISQSNSGNSPISGMVDSTRVGAMGWSMGGGGALRLAS GDRLSAVIPLAPWHQGRNSFDQLETPTLIIACENDTVAPVNRHASSFYNSIPS STDKALLEISNGAHSCANGGGANGGLLGKYGVSWMKRFIDNDLRYDQFLC GPNHAANSAVSEYRGTCNY
	WP_0859886 67.1	MPLSMNKLQKLPLTLTSAALLCAGGLTVNTAVAETRGPDPTEAYVEAEGP YNVDTINVSSLALGFGGGTIHYPTNTTGQMGGIVVIPGYLSYESSIEWWGER LASHGFVVMTIDTNTIYDQPGSRRDQIDAALDYLVDRSDSSFSAISGMVDG DRLGAVGWSMGGGGTLQLASGDRLSAAIPLAPWNSSFNDFDDIETPTLIFA CENDTVAPVGVHASPFYYDIPASTDKAFFEINNGNHFCANGDNANDAVLS KYGVSWMKLHIDQDARYGQFLCGPNHESQYRISEYRGTCPY
	MBF78136.1	MPFNKKGILAACGAGALLFSMSALANNPPPTDPDPGDGGSSPYQRGPDPSV SFLEADRGNYSVSTSRVSGLVSGFGGGTIHYPSGTTGTMAAIVVIPGFVSAE SSIEWWGPKLASYGFVVMTIDTNSGFDQPGSRATQINNALDYLVSQNTSSS

		SPVRGMIDTSRLGVVGWSMGGGGTLRVAREGRIKAAIPLAPWDTSTYYSS RSQAPTLIFACESDVIAPVYQHASPFYNALPSSIDKAFVEINNGSHYCGNGG SIYNDVLSRFGVSWMKLHLDEDARYKQFLCGPNHTSDSQISDYRGNCPY
	WP_1181318 88.1	MSALANNPPPTDPDPGNGGSSPYQRGPDPSVNFLEADRGQYNVDDERVSSF VSGFGGGTIHYPTGTTGTMAAIVVIPGFVSAESSIEWWGPKLASYGFVVMTI DTNSGFDQPGSRATQINNALDYLVDQNTSVGSPVRGMIDTDRLGVIGWSM GGGGTLRVGREGRIKAAIPLAPWDTSSYYASRAQAPTLIFACESDVIAPVYQ HASPFYNALPSNIDKAFVEINNGSHYCGNGGSIYNDVLSRFGVSWMKLHLD EDARYKQFLCGPNHTSDSQISDYRGNCPY
	WP_0918498 96.1	MSDHYASNPLRSVVAAASLLFSASVFAAGGGGSDGGDDGCTSNCGYERGP APTESFLEASSGPYSVRTDRVSSLVGGFGGGTIHYPTGTSGTMGAVVVIPGF VSAESSIEWWGPKLASHGFVVMTIDTNSGFDQPPSRATQINNALDYLIEQN GSSSSPVSGMIDTNRLGVIGWSMGGGGTLRVASEGRIQAAIPLAPWDTSSFR FRNIETPTLIIACESDIIAPVRSHADPFYEAIPSSTDKAFVELNNGSHYCGNGG NSYNDVLSRFGVSWMKLHLDNDQRYNQFLCGPDHERDWDISEYRGTCPY
	CAH17554.1	MAVMTPRRERSSLLSRALRFTAAAATALVTAVSLAAPAHAANPYERGPNP TDALLEARSGPFSVSEERASRFGADGFGGGTIYYPRENNTYGAVAISPGYTG TQASVAWLGKRIASHGFVVITIDTNTTLDQPDSRARQLNAALDYMINDASS AVRSRIDSSRLAVMGHSMGGGGSLRLASQRPDLKAAIPLTPWHLNKNWSS VRVPTLIIGADLDTIAPVLTHARPFYNSLPTSISKAYLELDGATHFAPNIPNKI IGKYSVAWLKRFVDNDTRYTQFLCGPRDGLFGEVEEYRSTCPF
	G8GER6.1	MPPHAARPGPAQNRRGRAMAVITPRRERSSLLSRALRFTAAAATALVTAVS LAAPAHAANPYERGPNPTDALLEARSGPFSVSEERASRFGADGFGGGTIYY PRENNTYGAVAISPGYTGTQASVAWLGERIASHGFVVITIDTNTTLDQPDSR ARQLNAALDYMINDASSAVRSRIDSSRLAVMGHSMGGGGTLRLASQRPDL KAAIPLTPWHLNKNWSSVRVPTLIIGADLDTIAPVLTHARPFYNSLPTSISKA YLELDGATHFAPNIPNKIIGKYSVAWLKRFVDNDTRYTQFLCGPRDGLFG EVEEYRSTCPF
	WP_1247733 20.1	MSSPTTTRPRSVVARLALAAVLAAGGVLAGPAGVAQAASPYERGPAPTTAI LEASRGPFATASQSVSSLSVVGFGGGVIYYPTSTSEGTFGAIAISPGFTASWS SISWLGPRIASHGFVVIGIETNTRLDQPDSRGRQLLAALDYLTERSSVRSRID SSRLAVAGHSMGGGGSLEAASSRPSLQAAVPLAPWNTDKSWSELRVPTLII GGESDSVAPVATHSVPFYNSIPASAEKAYLELNGASHFFPQTTNTPTARQM VAWLKRFVDDDTRYEQFLCPGPSGSQIQEYRNTCPSA
	TKK88911.1	MSRIATAALAFTLATGTAVTLAPSAQAAGFERGPNPTSAILEASRGPFSVAT TSVSSLVSGFGGGTIYYPTDTSQGTFGAIAISPGYTARWSSLEWLGPRIASHG FVVIGIETNSTLDQPASRGNQLLAALDYLVNSSSTTVRSRIDRNRLAVAGHS MGGGGTLHAAEDRPSLKAAVPIAPWNTDKTWGSVRVPTLIVAGESDSVAS PTTHASPFYNSITQTEKAYLELNSASHFFPQTTNTPFAKQFVAWLKRWVDE DTRYSQFICPGPSGLAIEEYRSTCPV
	WP_1608756 56.1	MTTTTWRTRIASLALAAAAATGLTSGVGLGAPISAVAATANPYERGPAPTR ASIEATRGAYATAQTSVSSLVSGFGGGTIYYPTTTADGTFGAVIIAPGYTAT SSSLAWLGPRLASQGFVVFTIDTDSRYDQPASRGDQLLEAADYLTRTSVVA SRVDARRVALMGHSMGGGGTLEAIKDRPSIKAAIPLTPWNLDKTWPEVTT PTLIIGADNDSVAPVASHAEPFYGSLPSTLDKAYLELRNASHFAPNSANTTI ASYSIAWLKRFVDDDTRYSQFLCPTPASSLAIAEYRSTCPY
	WP_1425690 56.1	MVTRLALVVLLALAGGLLTAPAAHAAVHGPDPTDALLESSRGPYATAQTD VSSLSVSGFGGGTIYYPTTTSEGTFGGVAIAPGYTADKSSLAWLAARLASH GFVVFNIDTLTRLDQPDSRGRQLLAALDYLTQRSSVRGRVDATRLGVMGH SMGGGGTLEAVDDRPSVRAAVPLTPWNLDKTWSGVRTPTLIIGAEADTVA PVASHAVPFYTSLSSSLDKAYLELNGATHFAPNTTNTTIGKYAVAWMKRF VDDDTRYDQFLCPGPGRSLTVEEYRSTCPF

	EFL43114.1	MHRPAGGSPRQRGPLVVQQHPHTGGRRTGRFAALAAAVAAVVGLTTLGG PGAHAADNPYERGPAPTESSIEALRGPYAVSETSVSSLSVTGFGGGTIYYPT STADGTFGAIAVSPGFTAYQSSIAWLGPRLASQGFVVFTIDTNTTLDQPASR GDQLLAALDYLTQRSAVRGRIDSSRLGVMGHSMGGGGTLEAAKDRPSLQ AAIPLTPWNLDKTWPEVRTPTLLFGADGDTVAPVGTHAEPLYTGLPSSLDR AYLELNGATHFTPNSSNTTIAKYSISWLKRFIDNDTRYEQFLCPLPRPSLTIE ESRGNCPHTS
	WP_0855761 51.1	MQQHPHTSRRGTGRFAALTAAVAAVVGLTTLNGPGAQAADNPYERGPAP TESSIEALRGPYAVSDVSVSSLSVTGFGGGTIYYPTSTADGTFGAIAISPGFT AYQSSIAWLGPRLASQGFVVFTIDTNTTLDQPASRGDQLLAALDYLTQRSA VRGRVDSSRLGVMGHSMGGGGTLEAAKDRPSLQAAIPLTPWNLDKTWPE VRTPTLLFGADGDTVAPVSSHAEPLYSGLPSSLDRAYIELNGATHFTPNSSN TTIAKYSVSWLKRFIDNDTRYEQFLCPLPRPSLTVEESRGNCPHTS
	WP_1504740 04.1	MQQQARTGAAHRVPTGSPHRRSARFAGLATAIAAVVGLTTLNGAGAQAA DNPYERGPAPTTSSIEAARGSYSVSQTSVSSLAVTGFGGGTIYYPTSTADGT FGAVAISPGYTGTQSTMAWLGPRLASQGFVVFTIDTNTTLDQPDSRGRQLL AALDYLTQRSSVRGRVDSTRLGVMGHSMGGGGSLEAAKTRPSLQAAIPLT PWNTDKSWPEISTPTLIFGADGDTIAPVASHAEPFYSSLPSSLDRAYLELNGT SHLTPISSNTTIAKYSVSWLKRFIDNDTRYEQFLCPLPRPSLTIEEYRGNCPHT S
	WP_0537570 25.1	MQQHPRSTTASAAPGPARGAGRRGTRRFAGAAAAIAAAVALSTLTGPGAR AADNPYERGPAPTTASIEASRGPYSVSETSVSSLAVSGFGGGTIYYPTSTAD GTFGAVAVSPGYTGTQSSIAWLGPRLASQGFVVFTIDTLTTLDQPDSRGRQ LLAALDYLTTRSSVRGRVDSTRLGVMGHSMGGGGSLEAAKSRPSLQAAIP LTPWNLDKSWPEVTTPTLIVGADGDSIAPVSSHAEPFYGSLRSSLDRAYLEL NGASHFTPNSSNTTIAKYSVSWLKRFIDNDTRYEQFLCPLPSPSLTIEEYRGN CPHTS
	BCL25765.1	MQQHSQNSSLTAGTVPPESAGRRPRRRDGRGGWAAKRITGALAALTTVVG LSSLASPGAHAADNPYERGPAPTTSSIEAARGSYSVSQTTVSSLAVTGFGGG TVYYPTSTADGTFGAVVISPGYTGTQSSISWLGARLASQGFAVFTIDTLTTL DQPDSRGRQLLAALDYLTERSSVRTRVDGSRLAVMGHSMGGGGSLEAAK SRPSLQAAIPLTPWNTDKSWPEVSTPTLIVGADGDTIAPVASHAEPFYGSLPS STDKAYLELNNATHFTPNSSDTTIAKYSISWLKRFVDNDTRYEQFLCPLPRP SLTIEEYRGNCPHTS
	WP_0699337 76.1	MRNAPAHRRRGSGRLRSLVAGLAALLAVGGLSSVATPAAQAADNPYERG PAPTTASIEAPNGPYAVSQTSVSSLSVTGFGGGTVYYPTTTGDGTFGAVAIS PGFTAGESSIAWLGPRLASQGFVVFTIGTLTRYDQPDSRGSQLLAALDYLTQ RSTVRARIDSGRLGVMGHSMGGGGSLEAAKSRPSLQAAIPLTGWNTDKTW PEIKTPTLVVGADGDTVASVGSHSEPFYESLPSSLDKAYLELNNATHFTPNT SNTTIAKYSISWLKRFIDNDTRYEQFLCPLPRPSLTIEEYRGNCPHTS
	WP_2373253 67.1	MQQHLLARRQTPHPSRSRTLTGLLTAAAATAGLLLTALAPGAQAVAANPY ERGPAPTNASIEASRGSYATSQTSVSSLAVSGFGGGTIYYPTSTADGTFGAV VISPGFTAYQSSIAWLGPRLASQGFVVFTIDTNTTLDQPDSRGRQLLSALDY LTQRSSVRTRVDASRLGVMGHSMGGGGSLEAAKSRTSLKAAIPLTGWNTD KTWPELRTPTLVVGADGDTVAPVGTHSKPFYESLPGSLDKAYLELRGASHF TPNSSDTTIAKYSLSWLKRFIDNDTRYEQFLCPIPRPSLTIAEYRGTCPHTS
	WP_1612676 37.1	MQQHLLARRQAPRPSRSRTLTGLLTAAAATAGLLLTGLAPGAQAVAANPY ERGPAPTNASIEASRGSYATSQTSVSSLAVSGFGGGTIYYPTSTADGTFGAV VISPGFTAYQSSIAWLGPRLASQGFVVFTIDTNTTLDQPDSRGRQLLSALDY LTQRSSVRTRVDASRLGVMGHSMGGGGSLEAAKSRTSLKAAIPLTGWNTD KTWPELRTPTLVVGADGDTVAPVATHSKPFYESLPGSLDKAYLELRGASHF TPNSSDTTIAKYSLSWLKRFIDNDTRYEQFLCPIPRPSLTIAEYRGTCPHSS

	AOS64284.1	MQSSSIASRRARVRSAGRPRTRLAGLVLALTMVATGLAAAPAATAQENPY ERGPAPTERSIEALRGPFAVAEDRVSSLVIGFGGGTIYYPTDTSEGTFGAVA VSPGYTGTQSSMAWLGPRLASQGFVVFTIDTNTTVDQPDSRGRQLLAALD YLVEDSDVRNRIDPNRLGVMGHSMGGGGSLSAAESRPALQAAIPLTGWHL SKNWSRVTVPTLVVGAENDLIAPVRSHSIPFYESLSSSLDKAYLELDGASHF APNISNTTIAKYSISWLKRFIDDDLRYEQFLCPPPDDREISEYRNTCPHS
	KOV83888.1	MPNEVYSAVQLRTLIPLALTLVLVTGTAAQAADNPYERGPAPTVSSIEALR GPFAVSETSVSSLVGGFGGGTIYYPTSTTSGTFGAVAVSPGYTGTQSSISWL GPRLASQGFVVFTIDTNTIYDQPDSRASQLLAALDYLTQQSSVRSRIDATRL GVMGHSMGGGGTLRAASQRPTLQAAIPLTAWHTTKNWSSVRVPTLVVGA EDDSIAPVATHSEPFYTTLPSTLDKAYLELNNATHFAPNSNNTTIAKYSISW LKRFIDNDTRYEQFLCPAPGRSTLIEEYRDTCPHS
	WP_373873195.1	MDKVIPKLFGIAAAVALGGAGITLIPDADAATASFAKGPAPSNASIEAVRGP FAVAQSNVSRASVSGFGGGDIYAPTDTGAGTFGAVVIAPGFTARKSSMAW LAPRLASQGFVVFNIDTLSTSDQPASRGRQLLAAADFLTQRSTVRARIDAGR VAVAIGHSMGGGGALEAAGSRPALAAAIPLTPWNLTKSFSRNAVPTLVIGAE ADSIAPVRSHAQPFFQSLPAVPGKAFLNLNGASHFAPNTPNTTIAKFSISWL KLFVDDDTRYQQFVCPGPGAGAAVQEYRSTCDQFTP
	WP_328594753.1	MPRTTLRTLAAAVLAAGAVGVLPTAPAHAAGFERGPAPTEASVTAAKGPF AIDRIEVPAGSGTGFNSGTIYYPTSTAEGTFGAVAISPGFVSPKSWIDWYGPR LASQGFVVMTLETFSYFDAPDGRADQLLAALDYLTAKSKVKDRIDPNRLA AMGHSMGGGGALSAAVKRPSLKAVVPLAPWYVGGGLEQSTVPTMIFGAD NDFIAPVASNARPFYQSLTKVPEKAYLELENAGHVGSFNSPNTTIAKYAISW LKRFVDDDTRYSQFLCPAPKFPSSTIQEYRDTCPS
	G9BY57.1	MDGVLWRVRTAALMAALLALAAWALVWASPSVEAQSNPYQRGPNPTRS ALTADGPFSVATYTVSRLSVSGFGGGVIYYPTGTSLTFGGIAMSPGYTADAS SLAWLGRRLASHGFVVLVINTNSRFDYPDSRASQLSAALNYLRTSSPSAVR ARLDANRLAVAGHSMGGGGTLRIAEQNPSLKAAVPLTPWHTDKTFNTSVP VLIVGAEADTVAPVSQHAIPFYQNLPSTTPKVYVELDNASHFAPNSNNAAIS VYTISWMKLWVDNDTRYRQFLCNVNDPALSDFRTNNRHCQ
	P19833.1	MFIMIKKSELAKAIIVTGALVFSIPTLAEVTLSETTVSSIKSEATVSSTKKALP ATPSDCIADSKITAVALSDTRDNGPFSIRTKRISRQSAKGFGGGTIHYPTNAS GCGLLGAIAVVPGYVSYENSIKWWGPRLASWGFVVITINTNSIYDDPDSRA AQLNAALDNMIADDTVGSMIDPKRLGAIGWSMGGGGALKLATERSTVRAI MPLAPYHDKSYGEVKTPTLVIACEDDRIAETKKYANAFYKNAIGPKMKVE VNNGSHFCPSYRFNEILLSKPGIAWMQRYINNDTRFDKFLCANENYSKSPRI SAYDYKDCP
PCL	WP_025804184.1	MSTLSWVRGVNGTLGWVAPKLVASKMRLAFMTPREHLPRDWELPLLARS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLIDALVGAGYSVVAL DGPAHGRSPGHEANVMLFARAMLEAAAELPPLRAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVRLPPKARSVFIRQVEQDVGMRA AAMDVAHYQLDMPGLIVHAEDDNFVPVKESELIHDAWFDSRLLRLKEGG HQRVLADPRVIEGVLTLLAGRSLQARQSA
	WP_059181212.1	MSTLKWVRGVNGTLGWFAPQLVASRMRLAFMSPRALPPRDWELPLLAKS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASIITALVDAGYSVVALD GPAHGRSPGEEANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQ LGLRTETLVTIAAPARILGVLRGFAKYVRLPPKARSAFIRQVEKDVGMRAA ALDVAHYQLDMPGLIVHAEDDRFVSAKESQLIHEAWFDSRLLRLEEGGHQ RVLADPRVIDGVLSLLAGRSLHSRQSA
	WP_052145136.1	MERLYRGMGRAALDRAYNNTRAIANFPAVLADFRTRSAALYERVRGRRD LRYGDRPRERFDWLPGGRANAPTFVFIHGGYWQNCAKEDFAFVAHGPLAR

		GFNVVLAEYTLAPDASMTQIVDEIGRLIDHLRADRDGLGTAGRPLCLSGHS AGGHLSALHRGHAFVTSALAISPLVDLEPISLSWLNEKLQLSEREIAAYSPL WHVGKGAPTVVAVGADELPELVRQADDYTAACAAAGEPVWGAHVPGCT HFSVLDDLAQPNGTLMRLLDAAIAGSGHGGDERADEERR
	WP_1243200 71.1	MSTLSWVRGVNGTLGWVAPKLVASKMRLAFMTPRERLPRDWELPLLARS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLIDALVGAGYSVVAL DGPAHGRSPGQEANVMLFARAMLEAAAELPPLRAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVRLPPKARSAFIRQVEQDVGMRA AAMDVAHYQLDMPGLIVHAEDDNFVPVKESDLIHEAWFDSRLLRLKEGG HQRVLADPRVIEGVLTLLAGRSLQARQSA
	WP_1341743 23.1	MNTLKWIRGVNGTLGWIAPKRVASKMRLAFMTPRSLPLRDWELPLLASSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVDAGYTVVALD GPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQ LGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRAA TLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGHQ RVLADPRVIDGVLSLLAGRSLQARQSA
	WP_0967961 49.1	MNTLKWVRGVNGTLGWIAPQRVASKMRQAFMTPRTLPLRDWELPLLASA ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLLRLESGGHQ RVLADPRVVDGVLSLLAGRSLQARQSA
	WP_0545966 12.1	MSTLSWVRGVNGTLGWVAPQWVASKMRSVFMTPRELPPRDWEMPLLAK SERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLITALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILAVLRGFARHVRMPPKVRSAFIRKVERDVGIQAS RLDVAHYQLDMPGLIVHAEDDVFVSVNESQLIHDAWFDSRLLRLEEGGHQ RVLADPRVIEGVLSLLSGRSLQARQSA
	WP_1235928 31.1	MNTLKWVRGVNGTLGWIAPQRVASKMRLAFMTPRSLPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVDAGYTVVAL DGPAHGRSLGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGH QRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_0597283 96.1	MTILYRGMDRAALDAAYLNTKVVPDFPALLASMQARSAALYDTAHGRRD LRYGAQPAQRFDWLSCGRPDAPLFVFIHGGYWQHCAKEDFAYAASGPLAR GFDVILAEYTLAPVATMTGIVGEIGALLDYLANDPDAIGTAGRPIHLSGHSA GGHLTAVYRAHPAVVSALAISPLVDLEPISLCVLNDKLQLGAREVDAYSPL RHVGPGAPTVVAVGDAELPELVRQARDYATACEAAGERVVHVGLPGMRH FDVLDDLAKPDGAMLAALQSIAPR
	WP_0386312 69.1	MSTLSWVRGVNGTLGWVAPTLVASKMRLAFMTPRERLPRDWELPLLARS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLIDALVGAGYSVVAL DGPAHGRSPGHEANVMLFARAMLEAAAELPPLRAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVRLPPKARSAFIRQVEQDVGMRA AAMDVAHYQLDMPGLIVHAEDDNFVPVKESDLIHEAWFDSRLLRLKEGG HQRVLADPRVIEGVLTLLAGRSLQARQSA
	AIC18917.1	MTMTLLYRDMNQAQLDAAYNNTQAVPDFPGIYAALQARSASFYASAAGR LNLPYGTAPRQRYDWLPCGKADAPTLIFIHGGYWQNCSKEDFAFIAAGPLA AGFNIVLAEYTLAPQASMTQIVSEIGSLLEHLQADADQLGIAGHKVVLSGH SAGGHLALQFRSHPWVTDVLAISALVDLEPISLSWLNEKLSLSEAEIDAYSP LYHIDKGANTWVAVGADELSELVRQSDEYAKQALARGESVQLIHVPGCTH FSVLDEMAKPQGALLQALSSIR

	WP_1390555 57.1	MNTLKWVRGVNGTLGWIAPQRVASKMRQAFMTPRTLPLRDWELPLLASS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLLRLESGGHQ RVLADPRVVDGVLSLLAGRSLQARQSA
	WP_1033048 52.1	MNTLKWIRGVNGTLGWIAPKQVASKMRLAFMTPRALPLRDWELPLLANS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLESGGH QRVLADPRVIDGVLSLLAGRSLQARQSA
	WP_0079692 26.1	MNTLKWVRGVNGTLGWIAPKQVASKMRTAFMTPRALPLRDWELPLLASS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLMRLESGGH QRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_0476001 95.1	MNTLKWVRGVNGTLGWIAPQRVANKMRQAFMTPRTLPLRDWELPLLASS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLESGGH QRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_1309278 96.1	MNALKWVRGVNGTLGWFAPKLVASKMRLAFMTPRALPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLINALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEASAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPSRILGVLRGFARMVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHEAWFDSRLLRLESGGH QRVLADPRVIDGVLSLLAGRSLQARQSA
	WP_1234634 37.1	MNTLKWIRGVNGTLGWIAPKRVASKMRLAFMTPRTLPLRDWELPLLANSE RITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVEAGYTVVALD GPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQ LGLRTETLVSIAAPARILGVLRGFARYVGLPPRARSAFIRQVEQDVGMRAA TLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLLRLESGGHQR VLADPRVVDGVLSLLAGRSLQARQSA
	WP_1515516 51.1	MNALKWVRGVNGTLGWFAPKLVASKMRLAFMTPRALPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLINALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARMVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHEAWFDSRLLRLESGGH QRVLADPRVIDGVLSLLAGRSLQARKSA
	WP_1234486 44.1	MNTLKWVRGVNGTLGWIAPKRVASKMRLAFMTPRSLPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDSFVSVKESQLIHEAWFDSRLLRLEGGGH QRVLADPRVIDGVLSLLAGRSLQARQSA
	WP_1276503 88.1	MNTLKWVRGVNGTLGWIAPKRVASKMRLAFMTPRALPLRDWELPLLASS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA

		ATLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLLRLESGGHQRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_160057137.1	MNTLKWVRGVNGTLGWIAPQRVASKMRLAFMTPRALPLRDWELPLLASSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVDAGYTVVALDGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRAATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGHQRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_108591690.1	MNTLKWVRGVNGTLGWIAPKQVASKMRTAFMTPRSLPLRDWELPLLASSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVALDGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRAATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLESGGHQRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_332373221.1	MTILYRGMDRTALDAAYLNTKAVPDFPALLASCQTRSAALYAATPGRRDLRYGAQPAQRFDWLPCGQPDAPLFVFIHGGYWQHCTKEDFAYAASGPLARGYDVVLAEYTLAPVATMTDIVGEIGALLDHLAADRDGLGTAGRPIHLSGHSAGGHLTAMHRAHPAVVSALAISPLVDLEPIALCCLNDKLQLTAREVDAYSPLHHIGPGAPTVIAVGDAELPELIRQADVYATACEAAGERIARTWLRGMQHFAVLDDLATPDGAMLDALHAIAPR
	WP_034186871.1	MTILYRGMDRAALDAAYLNTKAVPDFPALLASCQARSAALYDETPGRRDLRYGAQPAQRFDWLSCGQAGAPLFVFIHGGYWQHCTKADFAYAASGPLACGFDVILAEYTLAPVATMTGIVAEIGMLLDHLAADPDRLGTARRPIHLSGHSAGGHLTAMHRAHPAVVSALAISPLVDLEPISLCCLNDKLQLTAHEVDAYSPLRHVGPGAPTVVAVGDAELPELIRQADEYATACEAAGERIARVWLPGMQHFAVLDDLARPDGAMLAALRSITPR
	WP_130207001.1	MSALKWVRGVNGTLGWFAPKLVARKMRLAFMTPRDLPPRDWELPLLAKSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDAGYTVVALDGPAHGRSPGTEANVALFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQLGLRTETLVTIAAPARILGVLRGFARYVGLPPKARSAFIRQVEKDVGMRAATLDVAHYQLDMPGLIVHAEDDKLVSVKESQAIHEAWFDSRLLRLQEGGHQRVLADPQVIDGVLSLLAGRSLQSRQSA
	WP_150773989.1	MNTLKWIRGVNGTLGWIAPKRVASKMRLAFMTPRALPLRDWELPLLASSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVDAGYTVVALDGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRAATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGHQRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_069745339.1	MTILYRGMDRAALDAAYLNTKAVPDFPALLASCQSRSAALYDAIAGRRELRYGALPAQRYDWLPCGQPGAPLFVFIHGGYWQHCAKEDFAYAASGPLARGYDVVLAEYTLAPTASMTDIVAEIGALLDHLAADRDGLGIAGRPIHLSGHSAGGHLTAMYRAHPAVAAALSISPLVDLEPISLCCLNDKLQLTAQEIEACSPLRHIGPGAPTVVAVGDAELPELIRQAHDYAAACDAAGERIAHVQLPGMKHFAVLDDLANPDGKMLAALRAIAPR
	WP_003226446.1	MNTLKWVRGVNGTLGWIAPQRVASKMRQAFMTPRSLPLRDWELPLLASAERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVALDGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAVQLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRAATLDVAHYQLDMPGLIVHAEDDTFVSVKESQLIHESWFDSRLLRLESGGHQRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_1235329	MNTLKWVRGVNGTLGWIAPQRVASKMRLAFMTPRVLPLRDWELPLLANS

	98.1	ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPRRARSEFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEAGGH QRVLADPRVIDGVLSLLAGRSLQARQSA
	WP_102358940.1	MSTLKWVRGVNGTLGWFAPKLVASKMRLAFMTPRSLPLRDWELPLLASS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVERDVGMRA ATLDVAHYQLDMPGLIVHAEDDSFVSVKESQLIHESWFDSRLLRLESGGHQ RVLADPRVVDGVLSLLAGRSLQARQSA
	WP_007913423.1	MNTLKWVRGVNGTLGWIAPQRVASKMRLAFMTPRSLPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVDAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGH QRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_150794302.1	MNTLKWVRGVNGTLGWIAPQRVASKMRLAFMTPRSLPLRDWELPLLASS ERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAALITALVEAGYTVVAL DGPAHGRSPGREANVVLFARAMLEAAAELPPLQAVIGHSMGGASAMLAV QLGLRTETLVSIAAPARILGVLRGFARYVGMPPRARSAFIRQVEQDVGMRA ATLDVAHYQLDMPGLIVHAEDDNFVSVKESQLIHESWFDSRLLRLEGGGH QRVLADPRVVDGVLSLLAGRSLQARQSA
	WP_133210951.1	MNTLSWIRSVNGTLGRLAPEHIAGKMRHAFMTPRNLPPRDWELPLLASGE RITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAHLITTLVQAGYTAVALE GPAHGHSPGNQAHVALFARSLLEAAAELPPLRAVIGHSMGGASVMLALQ WGLRAEMAVSVAAPAQLLGVLRNFAHRLGMPSRARAAFVRQVERDVGIPI SRLDVSRYQLEIPALIAHAEDDRIVPASEALTIHQSWFDSRLLLLLPEGGHQR VLSDPQLIEGVMALLLRRHSTARQSA
	WP_038994175.1	MNTLSWIRSVNGTLGHLAPEHVARKMRRAFMTPRNRPPRDWELPLLARAE RITLRFGLSALRWGQGPTVLLMHGWEGRPTQFAHLIDSLVDAGYTAVALE GPAHGHSPGNEANVVLFARALLEAAAELPPLKAVVGHSMGGASMLLALQ WGLRAEVAVSIAAPAQLLGVIRGFARHLGMPARARAAFIRQIERDVGVQIS RLDVSGYQLELPGLIVHAEDDQLVPVDESDAIHRAWFDSRLLRLPDGGHLR VLADPQLREGVLALLQRSSSPARQSA
	WP_160087882.1	MNSMSWVRGFNASIGLLAPHALASKLRREFMTPHTLPPRDWELPLLAQAE RITLRFGLSALRWGSGPAVLLMHGWEGRPTQFAELIKALVNAGYGVVALD APAHGRSPGREANVVLFARALLEAASELPPLKAVIGHSMGGASALLATQLG LRTEALVSIAAPSRILTMLRRFSHYMGLPRQARAHFVQLVEEQAGIPAGQL DSAHYQLDFPGLVVHAVDDPMVPFSEAEAIHQRWFDSRLLRLERGGHQRV LADPQVVQAVLTLLASLNQAPSNALAS
	NMY13981.1	MNQMAWVRGVNATLGRVAPQLIASRLRERFMTPRTTSPRDWELPLLASSE RITLRFGLSALRWGSGPTVLLMHGWEGRPTQFALLIRGLVDAGYGVVALD APAHGRSPGREANVVLFARALLEAASELPPLRAVIGHSMGGASALLATQM GLRSETLVTIAAPSRILGLLRGFARFMGLPAEARAHFVRQVEKTAGIPAAHL DVQRYRLELPGLIVHAADDQVVPVSEADLIHKAWFDSQLLRLSAGGHQRL LSDPLLLQAVLELLEQVPQASLKALAS
PLA	WP_034126230.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQS ERITLRFGLSALRWHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALD GPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQ LGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLK

		HLDVAHYHLNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	WP_032896134.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQLGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLKHLDVAHYQMNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	WP_044274707.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQLGLRTQALVSIAAPSRFLDVLRGFTHMVGLPPRARSAFIQEVELSMGMPLKHLDVAHYQMNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	ABY53108.1	MGLAIAAAAVFSLPGVATATEPTGGVQPNIVGGGNATQVYSFMVSQQSSSGGHQCGGSLISSTWVVTAKHCGTPYQVRVGTTNRTSGGTVARVAQRIAHPSADLALLRLSTAVPQAPVTIADASGAVGTATRIIGWGQTCAPQGGCGAPITLQELNTSIVSDSRCLGISGASEICTNNPNGNSGACYGDSGGPQIKQVNGVWQLIGATSRAGNNSSTCATGPSIYVDVPYFRSWIRTNTGV
	WP_057005688.1	MGALTWIRGVNGTLGRLAPHTVANSMRRVFMTPRDLPPRDWELPLLAHAERVTLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVQNGYSVFALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLHAVVGHSMGGASAMLAVQLGLRTEALVSIAAPSRFLDVLGGFAGMVGLPSRARSAFIQEVELTLGMPLKHLDVAHYQMDLPGLIVHAEDDTFVPVSASQVIHDAWFDSRLLRLEQGGHQRVLADPRVVEAVLALLAGSCLQARQTA
	WP_043206304.1	MGTLTWIRRFNGTLGRLAPQTVANRMRRAFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLHAVIGHSMGGASAMLAVQLGLRTEALVSIASPSRFLDVLRGFTQMVGLPARARSAFIQEVELTFGMPLKHLDVAHYQMNIPGLIVHAEDDTFVPVKASQAIHEAWFDSRLLRLEQGGHQKVLADPRVIDAVLALLAGRRLQALQSA
	WP_116666539.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVGDGYSVIALDGPAHGRSPGREAHVLLFARAMLGAAAELPPLYAVVGHSMGGASAMLAVQLGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLKHLDVAHYQLNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	WP_064055161.1	MMGTLTWIRRFNGTLGRLAPQTVANRMRRAFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLHAVIGHSMGGASAMLAVQLGLRTEVLVSIASPSRFLDVLRGFTQMVGLPARARSAFIQEVELTFGMPLKHLDVAHYQMNIPGLIVHAEDDTFVPVKASQAIHEAWFDSRLLRLEQGGHQKVLADPRLIDAVLALLAGRRLQALQSA
	WP_076953514.1	MGTLTWIRRFNGTLGHLAPQTVANKMRRVFMTPRTLPPRDWELPLLAQSERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLITALVDKGYSVIALDGPAHGRSPGREAHVVLFARAMLEAAAELPPLRAVIGHSMGGASAMLAVQLGLRTEALVSIAAPSRCLDALRGFTTMVGLPSRARSAFIREVEMTFAMPLKHLDVAHYQMNIPGLIVHAEDDTFVPVKASQAIHEAWFDSRLLRLEQGGHQKVLADPRVIDATLSLLAGCGLQALQTA
	WP_019820855.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALD

		GPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQ LGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMALK HLDVAHYQLNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQ KVLADPQVIDAVLALLAGCRLQERQIA
	WP_0429370 80.1	MSTLSWIRGINGTLGRVAPRVVASRMRQMFMTPRARLPRDWELPLLATAE RITLRFGLSALRWGKGPTVLLMHGWEGRPTQFANLINALVAAGYTAVALD GPAHGRSPGREANVVVFARALLEAAAELPPLKAVVGHSMGGASAMLATQ LGLRTEALVSIAAPARVLGVLRGFARYVGLPPRARSAFIREVERDVGMRAA HLDIEHYQMDMPGLIVHAEDDRMVRVDESRRIHEAWFDSRLLRLESGGHL QVLADQRLIDGVLALLAGRSLAQRQSA
	WP_0995840 16.1	MGALTWIRRFNGTLGHVAPHTVANKMRRAFMTPRKLPPRDWELPLLAQS ERITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLIKALVDNGYCVIAL DGPAHGRSPGREAHVLLFARAMLEAAAELPPLHAVVGHSMGGASAMLAV QLGLRTQALVSIAAPSRFLDALRGFTRMVGLPARARSAFIQEVEMTFGMPL KYLDVAHYQMNIPGLIVHAEDDTFVPVKASQAIHDAWFDSRLLRLEQGGH QKVLADPRVIEAVLTLLAGCCLQERQSA
	WP_0567858 24.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQS ERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALD GPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQ LGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLK HLDVAHYQLNIPGLIVHAEDDTFVPVRAAQAIHQAWFDSRLLRLEQGGHQ KVLADPQVIDAVLALLAGCRLQERQSA
	WP_1976274 15.1	MGALTWIRGFNGTVGRLAPHMVASKLRRTFMTPRNLAPRDWELPLLAQSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVADGYSVIALDG PAHGRSPGREAHVLLFARAMLEAAAELPPLHAVVGHSMGGASAMLAVQL GLRTEALVSIAAPSRFLDVLRGFAGMVGLPPRARSAFIHEVELTFGMPLKHL DVAHYQMNIPGLIVHAEDDTFVPVNASQAIHDAWFDSRLLRLEQGGHQKV LADPRVIDAVLSLLAGRRLQERQTA
	WP_1106820 84.1	MSSMSWIRGFNATVGRLAPDLVASKMHRAFLTPRDLPPRDWELPLLAESE RITLRFGLSALRWGQGPAVLLLHGWEGRPTQFAELIRALVRAGYGVVALD APTHGRSPGHEANVVLFARALLEAAGELPPLKAVIGHSLGGASALLATQLG LRTEALVTIAAPARILGALRRFAHFVGLPKQARARFVRMVEQSAGMPAAQ LDVARYQLDFPGLVVHAEDDPMVPYGEAQSIHAAWPGSRLLPLERGGHSK PLGDPRVVEAVLELLGSADLHSAVSRRVLAATVLAS
	XOQ14903.1	AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGISTHPDLNIRGGASFVP GEPSTQDGNGHGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASGSGS VSSIAQGLEWAGNNGMHVANLSLGSPSPSATLEQAVNSATSRGVLVVAAS GNSGAGSISYPARYANAMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQ STYPGSTYASLNGTSMATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATS LGSTNLYGSGLVNAEAATR
	SPU21234.1	AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGISTHPDLNIRGGASFVP GEPSTQDGNGHGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGADGRGA ISSIAQGLEWAGNNGMHVANLSLGSPSPSATLEQAVNSATSRGVLVVAASG NSGASSISYPARYANAMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQST YPGSTYASLNGTSMATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATSLG STNLYGSGLVNAEAATR
	WP_0532580 10.1	MGALTWIRGFNGTVGRLAPRTVASKLRRTFMTPRNLPPRDWELPLLAQSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDG PAHGRSPGREAHVLLFARAMLEAAAELPPLQAVVGHSMGGASALLAVQL GLRTEALVSIAAPSRFLDVLRGFAGMVGLPARARAAFIREVELTFGMPLKH LDVAHYQMNLPGLIVHAEDDTFVPVKASQAIHDAWFDSRLLRLEQGGHQK VLADPRVIDGVLTLLAGCRLQARQTA

	WP_0995254 09.1	MNQMTWVRGVNATLGRVAPQLIASRLRERFMTPRTQPPRDWELPLLASAE RITLRFGLSALRWGSGPTVLLMHGWEGRPTQFALLIRGLVDAGYGVIALDA PAHGRSPGREANVVLFARALLEAASELPPLRAVIGHSMGGASALLATQMG LRCETLVTVAAPSRILGLLRGFARFMGLPAEARAHFVRAVETTAGIPAAHL DVQRYQLDLPGLIVHAEDDQVVPVGEADLIHKAWFDSQLLRLPAGGHQRL LSDPLLLQAVLELLEQVPQASLKALAS
	WP_0540638 32.1	MNTLRWIRGINGTLGRVAPRVAASRMRQVFMTPRERSPRDWELPLLATAE RITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLIEALVAAGYTAVALD GPAHGQSPGHEANVVAFARALLEAAAELPPLKAVIGHSMGGASAMLATQL GLRTEALVSIAAPARVLGVLRGFAQHVGLPPRARSAFIREVERDVGMRAEH LDIGHYQMDMPGLIVHAEDDQLVAVDESRRIHEAWFDSRLLRLESGGHQR VLADPRLIDGVLALLAGRSMAQRQSA
	WP_0714881 09.1	MGTLTWVRRFNSTLGHLVPQTVANRMRRAFMTPRELPPRDWELPLLAQSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLISALVDNGYSVIALDG PAHGRSPGREAHVLLFARAMLEAAAELPPLHAVIGHSMGGASAMLAVQLG LRTEALVSIAAPSRFLDVLRGFTRMVGLPARARSAFIQEVELTFGMPLKHLD VAHYQMNIPGLIVHAEDDTFVPVKASQAIHEAWFDSRLLRLEQGGHQKVL ADPRVIDAVLALLAGCRLQERQSA
	WP_0714863 26.1	MGALTWIRGFNGTVGRLAPHTVASKMRRTFMTPRDLPPRDWELPLLAQSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALDG PAHGRSPGREAHVLLFARAMLEAAAELPPLHAVVGHSMGGASAMLAVQL GLRTQALVSIAAPSRFLDVLRGFAGMVGLPPRARSAFIQEVELTFGMPLKH LDVAHYQMNIPGLIVHAEDDTFVPVKASQQIHDTWFDSRLLRLEQGGHQK VLADPRVIDAVLALLAGRRLQERQTA
	WP_1258815 20.1	MNQMTWVRGVNATLGRVAPQLIASRLRERFMTPRTQPPRDWELPLLASAE RITLRFGLSALRWGSGPTVLLMHGWEGRPTQFALLIRGLVDAGYGVIALDA PAHGRSPGREANVVLFARALLEAASELPPLRAVIGHSMGGASALLATQMG LRCETLVTVAAPSRILGLLRGFARFMGLPAEARAHFVRAVETTAGIPAAHL DVQRYQLDLPGLIVHAEDDQVVPVGEADLIHKAWFDSQLLRLPAGGHQRL LSDPLLLQAVLELLEQVPQASLKAMAS
	WP_0549216 54.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQS ERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALD GPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQ LGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLK HLDVAHYQLNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQ KVLADPQVIDAVLALLAGCRLQERQSA
	WP_1698517 37.1	MSTLSWIRRVNGTVGRLAPQTIANQMRRAFMTPRDLPPRDWELPLLAQAE RVTLRFGLSALRWGQGPAVLMMHGWEGRPTQFASLIDALVGAGYSVIALD GPAHGRSPGREAHVVLFARAMLEAAAELPPLHAVVGHSMGGASAMLAIQ LGLRTNALVSIAAPSRLLDVLRGFAGVVGMPARARAAFIQEVEYSLGIPLK HLDVAHYQMNIPGLIVHAEDDTFVPVKASQMIHEAWFDSRLLRLEQGGHQ KVLADPRVVEGVLALLAGCREPTRQTA
	WP_0584268 95.1	MGTLTWIRRFNGTLGHLAPQTVANRMRRAFMTPRDLPLRDWELPLLAQSE RITLRFGLSALRWGQGPAVLLMHGWEGRPTQFASLISALVDNGYSVIALDG PAHGRSPGREAHVLLFARAMLEAAAELPPLHAVIGHSMGGASAMLAVQLG LRTEALVSIAAPSRFLDVLRGFTKMVGLPARARSAFIQEVELTFGMPLKHLD VAHYQMNIPGLIVHAEDDTFVPVKASQAIHEAWFDSRLLRLEQGGHQKVL ADPRVIDGVLALLAGARLQERQTA
	WP_0031925 69.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQS ERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLITALVDNGYSVIALD GPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQ LGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLK

		HLDVAHYQMNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	WP_060765121.1	MGALTWVRGFNGTVGRLAPHAVASKMRRTFMTPRDLPPRDWELPLLAQSERITLRFGLSALRWGHGPAVLLMHGWEGRPTQFASLIAALVDNGYSVIALDGPAHGRSPGREAHVLLFARAMLEAAAELPPLYAVVGHSMGGASAMLAVQLGLRTQALVSIAAPSRFLDVLRGFTRMVGLPPRARSAFIQEVELSMGMPLKHLDVAHYQLNIPGLIVHAEDDTFVPVRAAQAIHEAWFDSRLLRLEQGGHQKVLADPQVIDAVLALLAGCRLQERQSA
	WP_154743290.1	MSTFKWTRGVNGALGRLAPQIIASKMRRVFMTPRNFPPRDWELPLLAQSERITLRFGLSALRWGQGPTVLLMHGWEGRPTQFASLISALVGAGYSVIALEGPAHGRSPGREAHVLLFARAMLEAAAELPPLHAVIGHSMGGASAMLAVQLGLRTEALVSIAAPSRFLDVLRGFAGMVGLPARARSAFIQEVEHAFGMPLKYLDVAHYQMNMPGLIVHAEDDTFVSVRASQVIHEAWFDSRLMRLKQGGHQKVLADPHVIKGVLALLAGCRPAQRQTA