

Supplementary Materials: Whole-Genome Sequencing of Chinese Yellow Catfish Provides a Valuable Genetic Resource for High-Throughput Identification of Toxin Genes

Shiyong Zhang, Jia Li, Qin Qin, Wei Liu, Chao Bian, Yunhai Yi, Minghua Wang, Liqiang Zhong, Xinxin You, Shengkai Tang, Yanshan Liu, Yu Huang, Ruobo Gu, Junmin Xu, Wenji Bian, Qiong Shi, and Xiaohui Chen

Data Availability: The genome assembly of Chinese yellow catfish has been deposited at the NCBI Genbank under the project ID of PRJNA494039.

Table S1. Summary of the next-generation sequencing data from an Illumina X-Ten platform.

Insert Size (bp)	Sequencing Length (bp)	Raw Data (Gb)	Clean Data (Gb)
250	150	51.34	45.14
500	150	56.23	48.16
800	120	18.40	17.01
2000	150	46.08	35.62
5000	150	46.90	30.55
10,000	150	48.71	32.99
20,000	150	46.71	22.13
Total		314.37	231.60

Table S2. Summary of the third-generation sequencing data from a PacBio Bioscience Sequel platform.

Library ID	Total Bases (Gb)	Total Reads	Average Length (bp)	Max Length (bp)	Min Length (bp)	N50 (bp)
r54040_20180213_070316-2_B01	4.18	511,615	8178.57	77,479	50	13,730
r54040_20180213_070316-3_C01	4.17	534,723	7799.99	75,635	50	13,282
r54040_20180213_070316-4_D01	4.05	480,249	8435.54	66,100	50	13,944
r54040_20180213_070316-5_E01	2.08	383,891	5419.65	97,683	50	9150
r54266_20180412_092831-1_C01	6.83	1,077,808	6337.74	67,030	50	10,989
r54272_20180413_101835-2_D01	4.49	695,485	6453.19	61,481	50	11,228
Total	25.47	3,683,771	7104.11	97,683	50	11,805

Table S3. Genome-size estimation based on the 17-mer frequencies.

K-mer	K-mer Number	K-mer Depth	Genome Size	Used Base	Used Read	Sequencing Depth
17	41,049,532,138	57	720,167,230	46,390,459,050	333,807,932	64.4

Table S4. The detailed repetitive elements in the yellow catfish genome.

Type	Length	Proportion of Genome (%)
LTR	200327135	28.06
LINE	7387444	1.03
SINE	237822	0.03
DNA	33850171	4.74
Other	2584755	0.36
Unknown	5977250	0.83
Total	242600066	33.99

Table S5. Statistics of gene annotation from the genome assembly of Chinese yellow catfish.

Method	Software, Fish Species, or Data Source	Gene Number	Average Transcript Length (bp)	Average CDS Length (bp)	Average Exons Per Gene	Average Exons Length (bp)	Average Intron Length (bp)
<i>De novo</i>	AUGUSTUS	21,392	14,898.52	1,434.06	8.17	175.43	1,876.73
	GeneScan	28,229	16,786.98	1,478.17	7.72	191.46	2,277.95
Homolog	<i>Ictalurus punctatus</i> BGI	25,553	10,219.12	1,459.43	7.29	200.24	1,392.98
	<i>Danio rerio</i>	25,390	12,070.41	1,569.74	8.05	194.89	1,488.49
	<i>Gadus morhua</i>	22,759	8,385.83	1,197.41	6.64	180.22	1,273.59
	<i>Gasterosteus aculeatus</i>	26,322	7,804.13	1,171.03	6.31	185.53	1,248.77
	<i>Latimeria chalumnae</i>	19,518	8,582.87	1,377.42	6.99	196.92	1,201.98
	<i>Oreochromis niloticus</i>	24,516	8,883.31	1,324.56	7.05	187.86	1,249.24
	<i>Oryzias latipes</i>	26,464	7,215.32	1,148.88	6.08	189.06	1,194.92
	<i>Takifugu rubripes</i>	18,808	10,998.62	1,484.95	8.19	181.28	1,322.94
	<i>Tetraodon nigroviridis</i>	17,438	10,978.66	1,486.42	8.41	176.66	1,280.30
	<i>Xiphophorus maculatus</i>	23,040	9,053.33	1,353.53	7.20	188.07	1,242.49
Transcript	Dataset 1 *	20,132	9,395.39	2,147.20	8.37	256.55	983.55
	Dataset 2 **	33,361	6,890.79	1,692.43	6.05	279.97	1,030.39
Consensus		21,562	16,730.36	1,697.60	9.46	179.37	1,776.09

* Transcriptomic Dataset 1 were generated in this present work. ** Transcriptomic Dataset 2 were adopted from our previous report (Xie et al., 2016).

Table S6. Functional assignments from the genome assembly of Chinese yellow catfish.

Catagory	Gene Number	% of Gene
Total	21,562	100
Nr	20,941	97.12
Swissprot	19,647	91.12
KEGG	18,787	87.13
TrEMBL	20,936	97.10
Interpro	20,004	92.77
Annotated	21,042	97.59
Unannotated	520	2.41

Table S7. Information of the fish species used for phylogenetic analyses.

Species	Genome Version	Gene Number	Database
<i>Astyanax mexicanus</i>	AstMex102	23,041	ensembl
<i>Cynoglossus semilaevis</i>	Cse_v1.0	22,144	NCBI
<i>Danio rerio</i>	GRCz11	25,778	ensembl
<i>Gasterosteus aculeatus</i>	BROAD S1	20,785	ensembl
<i>Ictalurus punctatus</i>	-	21,556	BGI
<i>Latimeria chalumnae</i>	LatCha1	19,568	ensembl
<i>Lepisosteus oculatus</i>	LepOcu1	18,341	ensembl
<i>Oreochromis niloticus</i>	Orenil1.0	21,437	ensembl
<i>Oryzias latipes</i>	HdrR	19,682	ensembl
<i>Paralichthys olivaceus</i>	ParOli_1.1	21,787	NCBI
<i>Scleropages formosus</i>	-	22,274	BGI
<i>Takifugu rubripes</i>	FUGU 4.0	18,518	ensembl
<i>Tetraodon nigroviridis</i>	TETRAODON 8.0	19,595	ensembl
<i>Xiphophorus maculatus</i>	Xipmac4.4.2	20,379	ensembl

* The protein datasets of channel catfish (*Ictalurus punctatus*) and Asian arowana (*Scleropages formosus*) were obtained from our laboratory. All the protein datasets had been filtered as follows: 1) Only the longest transcript was retained if one gene has multiple transcripts. 2) Those protein sequences were discarded if the amino acid number was less than 20.

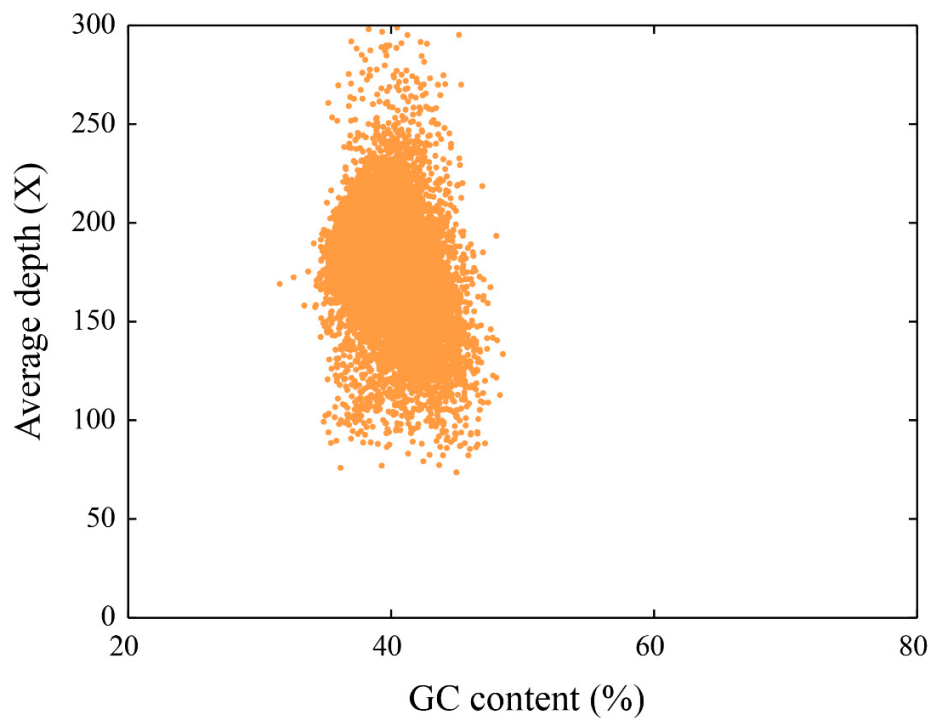


Figure S1. GC content and sequencing depth of the yellow catfish genome. This scatter plot was drawn by sliding 50-kb non-overlapping windows against the assembled yellow catfish genome. The x-axis means GC content (%), and the y-axis represents the average depth.

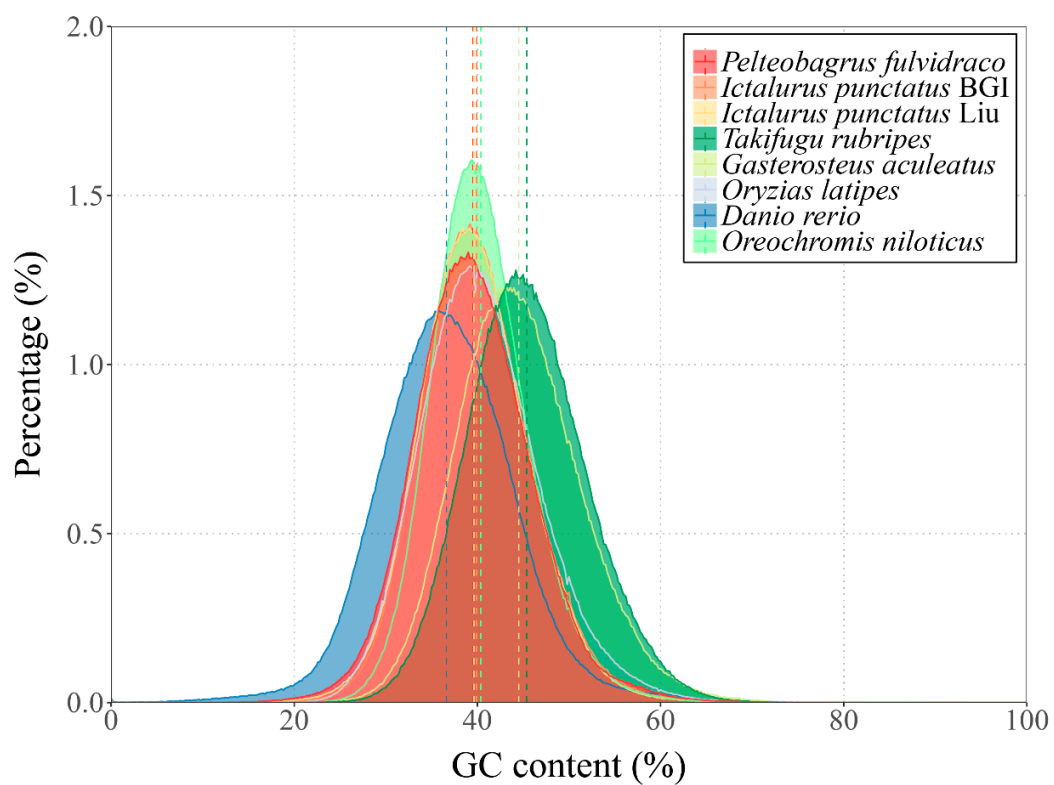


Figure S2. Comparisons of GC content between the yellow catfish and other seven fish species. The values in this plot was calculated by sliding 50-kb non-overlapping windows against these eight genome sequences. The x-axis means GC content (%), and the y-axis represents the percentage of each GC content.

Data S1. The short-length toxin proteins in the yellow catfish.

```
>Q4JCS0-D3
SAPYCGNKLVDVGEECDGCGSEEECEKDPCCEPKTCKLRAGAECAGVCCCKYCRVTHHT
>P0DMW6-D3
CGLPSLQGPCKFYKPRWAYSSTFNQCQPFYGGCGGNDNNFETKEACEDLC
>P17497-D4
SMPMCGNQILEGTEECDGPPETCNKCCDAATCTFTKGSVCAAGVCCCKDCQVLVSGTPCRVSVNECDLPEYCTGQS
GFCP
>P22028-D7
LLDPPECGNGYVEQGEECDGCGS
>C0HJW4-D1
DPRSSDMNECLKKSHYYIKCFSDIC
>P82475-D1
DVYERSSCQPRETLVEVWQEFPWETHHLFLPSCVSVRRRCGGCCGDEALECVPSHTDMVTMEVTPNT
>Q7Z091-D1
PCWYHSDCLCKLHKISRPSLKWLL
>P04362-D2
AWIFPGTLWCGHGNRAGDYEQLGMFERVDRCCREHDHC
>P0CI21-D24
GSARGRCEAGNGPRLAGARTSRSRVWECSPKRVVNSI
>P0CI21-D7
GSARGRCEAGNGPRLAGARTSRSRVWECSPKRVVNSI
>P0CI21-D8
GSARGRCEAGNGPRLAGARTSRSRVWECSPKRVVNSI
>P81995-D2
ILDHNLKLRGQVYPPASNM
>P86468-D24
CLNGGVCHDGVKTFHCSCP KGYTGNRCQ
>P0DMW7-D4
SEVCWANAETGPCRAMLPWYFVKEEGRCAPFIYGGCGGNRNNFESEYCL SVC
>P86468-D5
NPCQHGGICIEKRGYICKCTD MYTGHNCEK
>P81990-D1
QEEIVNIHNTFRREVPTASNMLKM
>P0DN17-D8
CTQPPVTGPCRGSN TMWHYNPYEQKCSRFNYGGCQGNDNQFDTEEKCMKFC
>P86402-D1
LCVSCFTTKYDIQSTNCCGGNTCVCFG
>Q90W96-D1
CDLPHTAGLCDEWTARYYYDPASSRCVHFWYGGCPGNSNNFATMEECQQT C
>P0CV91-D2
GLFVIDDKGILRQITINDLPVGRSVDETLRLVQAFQHTDK
>P86468-D10
CLNGGTCSNTGPDKYHCSPDGYSGQN CER
>P86468-D23
CFNGGRCADNPDGGYFCQCPTGYAGFNCEK
```

>P0CV91-D1
GLFIIDDKGTLRQITMNDLPVGRSVDETLRLVQAFQYTDK
>P0C8W6-D2
ASEPASEPASEPASEPASEPASEPARQAARSCP
>P0DMD6-D6
CFGFKIDRISDLSGMGC
>P22028-D11
PICGNGMVEAGEQCDCG
>P0DMW6-D2
CSLPSVQGPCKSYEPRWAYNHLLKKCQSFVYGGCGGNENNFSKEACEEMC
>P00985-D1
LCELVLDDQGSCRDYSIRWYYDRTANACAQFWYGGCHGNRNRFDTEDECKKTCV
>P80163-D2
CSCSNWMDKECIYFCHLDIIW
>P80163-D1
CSCATFLDKECVYFCHLDIIW
>P0C8W6-D4
LAVEPAVEPAVEPAVEPTVEPAVELAVEPA
>P0CI21-D1
GSARGRCEAGNPRLAGARTSRSRVWECSPKRVVNSI
>P0CI21-D23
GSARGRCEAGNPRLAGARTSRSRVWECSPKRVVNSI
>P0CI21-D3
GSARGRCEAGNPRLAGARTSRSRVWECSPKRVVNSI
>P0CI21-D4
GSARGRCEAGNPRLAGARTSRSRVWECSPKRVVNSI
>B3EWP2-D1
PFFTVWNAPTKKCASQYGVDLDLSMFDIIHN
>Q2XXR8-D1
MRSSFCLFFLLLVSYGSGAVITGCTKDSQCGGSMCCAVSLWIRSLRMCVPMGMEGEECHPM SHKVPFGGKRLH
HTCPCLPNLACITTEEGNSRCL
>P0C6S3-D1
YCISGVIGHRRTECQ
>P0CH88-D4
IVGGNAASEGSWPWQVSL
>P0DMD6-D3
SGCFGHKMDRIGTISGMGC
>P0DMW6-D6
CTMPVVGPKGVFPRWYYDPTAGECKHFLYSGCKGNHNNFLQQADCANEC
>P22028-D8
VSSQPICGNRITEKGEECDVGLNAN
>Q90W97-D1
CQLAQDMGTCINFMLKWHYEASRKECSRFWYGGCGGNRNRFDTEEECEARC
>Q9TWF9-D4
CSLDMDMGRVCSSYESRWYYDRNAHECKHFWYGGCDGNGNRFMTECEETCG
>P86468-D16
CLNGGTCQDGYGTYKCTCPHGYTSTNCQ
>P0DP55-D1
SWTAIFTLFLCVLLYLSTVVDA YPPKPENPGENAAPEELAKYYTALRHYINLITRQRYGKRSTQEGILTDLLFGDSND
>P43318-D1
GFTYPTLWCGAGNIADNYDQLGEF

>P0CI21-D10
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D25
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D9
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CH17-D1
LRKDTLCSSSFLSRCCDSVCILI
>P86821-D1
ANTPEHAPEHAPEHAPEHAPEHAPE
>C0HJB4-D1
PCRSVEPMMRSGSCIRGW
>P0C8X6-D1
QCPSNCNKNGECVNGKCVCH
>P0CV87-D1
GCFGLKLDRIGSMSGLGC
>P22028-D10
MLGGPRCGNLYVEKGEECDG
>P80163-D3
CSCINQLDSECYYFCHLDIVW
>P86468-D1
HHPCVNGGTCINSEPNEYNCACPEGYSGKNCE
>P86468-D6
CRNGGTCLAHSTKSYQCRCPEGFRGQWCE
>P0DMZ2-D1
FMSSYTDGMPWLRMTHHSPC
>P86468-D8
CKNGGTCQDGINDYICTCPPGYTGKNC
>P86468-D9
CLNGGRCIDRINNFHCLCPAGYTGRTE
>P84038-D1
EGCEHGWKKFHGHGHCYKLFSSRRHTWEDA EKDCRELSGHL
>P86468-D19
CLNGGLCMSPGVCICPPGYGVSC
>P0DMB7-D2
PHVYSDDVTRCHVMSRENKKH
>P0DJ94-D1
HADGLFTSGYSKLLGQLSAKEYLESLL
>P0DMJ1-D1
VDICKIPKDEGTCAKFVLKWHFDSIEKSCKRFWYGGCGGNQNRFDTQKECEIACGKA
>P0CV91-D3
GLFIIDDKGILRQITINDLPVGRSIDETLRLVQAFQFTDK
>C0HJR6-D1
QERLPEYNITGRMMCSGYPEGGIDTC
>C0HJF4-D3
CALKKDEGPCKALKDRFYFDMEMFRCEPFYGGCQGNENNFETI
>P0DN17-D3
CALKKDEGPCKALKDRFYFDMEMFRCEPFYGGCQGNENNFETIEECEEMCLVSK
>P86468-D4
CLNGGRCVDGIGHYTCVCPPGFTGERCE
>B5KL36-D5

CKGVPSDGPFCGMLHRYHYNSSIMTCQRFDYGGCMGNQNNFLTEKECLQTC
>P0DN12-D1
CRLPVDVGSKMALQFWAFDSKIGKCVSFIYGGCDGNGNRFYQTQKECEEYCGVSRD
>P0C6S2-D1
GQGRCIYKCMNHNLSLQLHTLC
>A6MGY1-D1
CTLVLSQGTCDYIIRWYYDKQANACAQFWYGGCEGNENRFDTEGDCKKTC
>P0DKR0-D9
LFEPTECGNGYVEVGEECDG
>P0DMW7-D3
SAVCWAPARKGPCRAKLSRWYFVAEKGRCAPFTFGCGGNRNNFESEYCMAVCSS
>P85843-D1
HDPHTKHGRTSIVHLFEWRWQDIAAECERYLAPNGY
>P0DL27-D3
IVGGYECKPNSQPWQVSLNVGYHFCGGSLINQNWVVSAAHC
>P0DN42-D2
CMACGPRDSGRFCFNPICCAAGLGCSVGSPEALSCTEEDYIIPCENGGRACGSKRGRCAAPGVCCNSGLCYFNTLVK
>P82972-D3
KSKNFSGCFGGRLDRIGSSSTLGCNAMK
>P83231-D3
SSSKYSGCFGRRLDRIGSMSTLGCNTV
>Q2XXR8-D2
ACDRDVQCGIGMCCA VSLWLRGLRMCTPQGFEGDECHPFSHKVPFPGKRH
HTCPCLPHLVCTRYTDSRYRCKDYKSI
>P0C8X6-D7
RSCPGNCNNKGRCVNGKCVC
>P86468-D3
NICLNGGTCKFDQKGQVNCLCPLGTSGLYCE
>P0DMH1-D1
CPANWVPFSGHCY
>P86468-D11
CLNGGTCVEGTDISYTCLCPKGFTGNNCEE
>Q9PS06-D1
CDHNWRKFHGHGCRYRYFTHRLNWEDAEDKDC
>P0DKM8-D1
VCGTDGKTYTNECELRTSCQEKNIEVARPGSCDE
>P0DKM8-D2
VCGSDGLDYPSECALNMKACSTNKNIRLQHVGSCE
>P0DKM9-D1
VCGTDGVTYADYCQLRTIACRQDKEITVKHLGQC
>P0DKM9-D2
CLDTCLQASDPVCGSDGQTYTSQCQMNAISCTLQKHIQIKHKGPC
>P86468-D12
CHNGGTCHNLVGGFSCSCPEGFTGMACER
>P00983-D1
ACHLPKAPGDCYGHYLRYYYDAAHGKCTTFAWTGCVGNRFLDLNHCNATC
>P86468-D17
CLHDGTCILDSSHSYHCACLAGYTGKRCE
>P0CH88-D3
IIGGIDATLGRWPWQVSL
>C0HJR6-D2

TNKMICAGLLQGGKDTC
>P0CH88-D2
IVGGQSASAGAWPWQVSL
>P86991-D1
FPDGGCGKRLPCAIDSN
>P0CV91-D4
GLFIIDPNGIIKHMSINDLPVGRSVEETLRLVRAFQFVE
>Q8T0W4-D5
CQAEPQVGLCRASIPRYYYTSGTCKRFRFGGCGGNSNNYNTTEECMKTC
>Q9TWF8-D4
CRFEKVVGHCRA SFPRYYDVTDQTCKTFVYGGCGGNNNNFKTKEECENAC
>P0CI21-D22
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>Q2ES47-D3
PSICRLPVEKGLCFAISTRYFFNMASMQCQAFYGGCGGNNNNFPDHISC
MEYC
>Q6T6S5-D1
DVCFLQVDEGPCLDDVPRFYNTLTQSCEEFSGGCEGNANNFKSYEACY
KTCFS
>E2E4L2-D1
LVPIAVVLCVLVCMAEAYPPKPEPPAGDAGPEELAKYHTALRHYINLITRQRYGKRSSPEVEMAELLFG
>P86468-D2
CFNGGTCEERFTGGYVCRCPPAYTGSNCEK
>P86468-D21
CLNGGRCIAPYECECPKGWTGKRC
>P0DMW7-D1
ALAVCSLESETGPCRASMPRWRFDIHLGKCVRFIYGGCAGNRNNFESEEDCMAVCKS
>P59068-D1
LWQLRGMILCVKPHSWPALDYADYG CYGYGSGTPLDDLDRCCQVHDKCYSDA
>P0DN43-D1
CYISNCPIGGKRAVQDLPT RQCMGCGPGDKGRCFGNICCGEEIGCMVGTLEAMRCLEEDYLPSPCETGGKPCGSV
TGRCAAPGVCC
>P86468-D20
NPCLNGGVCARREGGYTCICRENYTG
>P0CI21-D19
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D20
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0DM77-D1
CSQLKQAGTMCRSAAGSCDLPEYCTGGSPYCPSNVY
>P0CI21-D13
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D14
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D16
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>P0CI21-D17
GSARGRCEAGNGPRLAGARTSRSRVVWECSPKRVVNSI
>Q6T6T5-D1
LCNLPAVQGPCRNWEPRWAYNAVTRLCQAFVYGGCRGNSNNFRSKAECQASCPRQSSR
>P13211-D4

CSCENLKDRECVYFCHIGIVW
 >P86468-D15
 CMNGGTCQDSYGYKCTCPHGYHGFNCQ
 >P0DN20-D1
 CNLKMEPGSCFARQAMYYYDSQEACRMFLYGGCQGNRFDTKEDCEKMC
 >Q90WA0-D3
 CVDPPVTGPCRASMPHWFYDPLKQTCFRFTYGGCAGNANRFGKKDDCMNVCA

Data S2. The medium-length toxin proteins in the yellow catfish.

>Za2G-D17
 TAVTPGINFPEFTIVGQVDGGHIEYYDSNIMKPIPKAEWKKKVTADDPDYWNKQTQSTQETLKDNIILKVNTHHTHT
 QGCVLHRGVHTMQVMYGCELDGDDGTVRGYTQFGYDGEDFLSLDLKPQAVISKKNWDNDRGAIVGEKNYLENTCI
 EWLKKYVSYGKETLERKKGK
 >Za2G-D3
 MSLCSTVMKTLIFFTSLHLSSAVTHSLQYFYTGVTGPGINFPEFTNVGLLDGQQISYYDSNIRKKIPKTEWIKMTADD
 DYWNSGTQNLQDTQESFKVNVATAMQRFNQTTGVHVSQVMYGCELDGDDGTVRGYRQYGYDGEDFISLDLKTLTW
 TAAKPQAVITKNKWDHNPVSVGRKNYLENICIDWLKKYVSYGRETLERNKG
 >Za2G-D22
 VTHSLQYLYTAITPQINFPEFSAVGLVDGGQFVYYDSNIRKMIPKTEWIRSDKISAYDSYWKRETERSRSDHEDLHHL
 LHTVMKSFNHTQGVHTFQRMYGCELDGDDGTVGGYNQFGYDGEDFISLDLKTGSWTAAKPQAVIKNQWDSTGAQ
 GQYWKSYLETECIEWVKLVSYGRETLERKVRPEMSLFQK
 >Za2G-D38
 TYNKVLIFLTYCVHLSLADTHALQYLYTALTSGINVPEFIAVVLVDDEEQSVYYDSNIRKMIPKTEWIKFSAADDPKYW
 NRETERMQNDQEDYKVDKTTLMHRFNQTTGVHTLQRMCGCELDGSGTTGGHNQFGYDGVDFISLDLNTETWTVN
 NNKAEIFIKEDWPEGEKAKYWTTQLTYECIDQLKKFVSYGRETLER
 >Za2G-D40
 TILKVLPLFTLTHVSSADTHSLQFFYTAGTPGIKVTAVGLLDGEQFVYYDSNIRKMIPKTEWIKISTDDKDYWNRET
 EHVEDNQDSVATVMKNLNAEGDHTLQWMFGCGLDNGTIRGYSQYRYDGEDFISLDLNCQKPGCARTWTAANE
 KAEIFIKKWDPRRDQARYWMDYLQFNCIDQLKKFVPHSRETLEKDPPTASVIQK
 >Za2G-D44
 IPKTEWIKKVDADNPYWNSTETEVWQDEEELKHDDVTAMNHFNQTEGVHTVQLMFGCEIDDDGTTRVFSHFY
 DGEDFISLDLKTKTWTAASIKTLITKKKWNPTGHKAKGCTLYLENECIDWLKKFVSYRRDTERKVCPTASVFQK
 >Adamalysin-D5
 FSQKYIERFYSFQKGPPGHKKRSHPSFSSKLKDMQSFGLDQGTGNLDTETLDEMKRPRCGVPDIEEYVYNRGNRWKK
 NVITYNIGKYTSDLPVSTVDSLIGAALDVWANASPLRFFRSSSQADIMVEFGSKYHGDNYFPDGPKGTLAHAFDPG
 EGVGGDVHFDDDELWTADSRGFSLYLVAHEFGHSLGLKHSQNPQSLMYPTYRKRQPNMLSSDILNINALYGIN
 >Za2G-D43
 FFHIFVLSSADTHSLQFLYTAVTPGINVTAVGLLNGEQFVYYDSNIRKMIPKTEWIKISTDDKEMTGTEKPSVVKD
 HQYQLQDIVTTVTKRLNAEGDHTLQWMFGCGLDNGTIRGYSQYRYDGEDFMSLDLNNQEHGTWIAANEKAE
 LFIKEWNYKEHAIYWMNYLKTECIDRLKKFVPHGRETLERKKGK
 >Za2G-D27
 SVTHSLQYFYTAVTPGINFPEFTAVGLVDGGQFVYYDSNIRRMIPKTEWIKMNYDDPDYWNRETQNLQGSEETFKV
 NVDILMMRFNQTTGVHVSQVMYGCELDGDDGTVRGYRHYGFDGEDFISFDLKTVTWIAPKPQAVITKNKWDNYPG
 MTVARKNYLDDICIEWLKKYVSYGKETLERKVRPEASVFQ
 >Za2G-D31
 VTHSLQYLYTAITPQINLPEFTAVGLVDGGQIVYYDSNIRKMIPKTEWIKISADDPDYWKRETEHSRSDHEDLHHL
 HTVMKSFNHTQGVYTLQRMYGCELDGDDGTVGGYNQFGYDGEDFISLDLKTGRWTAAPQAVIKNQWESTGAQG
 QYWKSYLETECIEWVKLVSYGRETLERK
 >Za2G-D32

MDLSCSLVKVLLFLMNGFHQTSVTHSLQYFYTAVTPGINFPEFTAVGLVDGGQFVYYDSNIRRMIPKTEWIQKMNY
DDPDYWNRETQNLQGSQETFKVVIKIIAVYTNVYVYAMQRFNQTTGVHVSQVQMYGCELDGDDGTVRGYRQYGYDGED
FISFDLKTWTWIAAKPQAVITKNKWDNDPGMSVARKNYLENICIEWLKKYVSYGR

>Za2G-D46

MSPRSTVMKTLIFFTFSLHLSSAVTHSLQYFYTAVTPGINFPEFTAVGQVDGQFGCYDSERREVIKTEWIHSDHPDH
WNRITRIAKEHQEAYKAHVFTQMVFNQTTGVHTVQRMVGCELDGGGTVRGYRQYGYDGEDFLSFDPKTLTWTA
AKPQAVITKNRRNTEPGNNNEMKNFLEIDCIDWLKQYVSYGKETLERKVRPTVSMFQK

>Za2G-D5

SNHLKSALYVSVTHSLQYFYTAVTPGINFPEFTAVGLVDGGQFVYYDSNIRRMIPKTEWIQKMNYDDPDYWNRETQ
NLLGSEETFKVNVYVYAMQRFNQTTGVHVSQVQMYGCELDGDDGTVRGYRQYGYDGEDFISDLKTWTWIAKPQAVIT
KNKWDNDPGMTVARKNYLENICIEWLKKYVSYGKETLERKVRPETSFLQE

>Za2G-D8

SVTHSLQYFYTAVTTGINFSEFTAVGLVDGGQFVYYDSNIRRMIPKTEWIQKMNYDDPDYWNRETQNLQGSQETFK
VNVYVYAMQHFNQTTGVHVSQVQMYGCELDGDDGTIRGYRQYGYDGEDFISDLKTLTWIAKPQAVITKNKWDNDP
GMTVARKNYLENICIEWLKKYVSYGKETLERK

>Za2G-D9

MSLCTVMKTLIFFTFSLHLSSAVTHSMQYFITA VTPGTNFPEFTIVGLVDGGQFVYYDSNIMKMIPKTEWIDVMGEDY
WNREAQKQQQSQETFKADVSTVMQRFNQTTGVHTVQVIYGCELDGGGTVRGYRQDGYDGEDFISFDLKTWTWTA
AKPQAVITKNKWDNDSGYNVARKNYLENICIEWLKQYMSYGRETLEKVRPTASVFQK

>Za2G-D2

MSLCSTVMKTLIFFTFSLHLSSAVTHSLQYSYTA VTPGINFPEFAE VVGQVDGGQISYYDSNIRKKIPKTEWIQKVTADDP
DYWNSGTQILQGAQDTFKVSVSSLMQRFNQTTGVHVSQVQMYGCELDGDDGTVRGYEQYGYDGEDFISDLKTWTWT
AVKPQAVITKNKWDNDHGATVGEKNYLEKDCNEWLKKYVSYGRETLEKKGK

>Za2G-D47

MSLCSTVMKTLIFFTFSLHLSSAVTHSLQYFYTA VTPGLNFPEFAE VVGQVDGGQISYYDSNIRKKIPKTEWIQKVTADD
PDYWNSGTQNLQGAQETFKVSVASLMQRFNQNGVHVSQVQMYGCELDGDDGTVRGYEQYGYDGEDFISDLKTWT
WTA AKRQAVITKNKWDNIPGYNMDRKNYLEKECIEWKYVSYGRETLEKVRPTASLFQE

>Adamalysin-D3

MKTYYLQCLVALVFRVRSNPVPQNNDLTDEEFAKNYLKRLYNMKEVNKPSFGRTTSEMSLKSQMQQFFGLKVTG
ILDDETIAMMKKPRCGVPDVA AFKSNALPIKWNSTNSLYRIENYTPDMSVAEVDETIERALQVWARVTPLRFTRINSG
VADIMISFGRGSHGDAYPFDGPSGLAHAFAPSSGIGGDAHFDEDENFTFSSTNGFILFLVAAHEFGHSMGLSHSSDP
GALMYPTYSYRDPKTFVLPADDVKGIQSLYGPNDKPVDPSPKNPPPV

>Za2G-D41

TLIYLLWNSFSVHLSYAVTHTLQYIHTAVTPGINLPEYTDVGLVDGEPFVYYDSNIRKYIPRTEWINKITDYPDYWS
SGTQNQNYVQEIYKDDVITLMRRFNQTEGVHILQRIYGCELHDDGTVRGFEQLAYDGEDFMSLDLEHVTWTAVKP
QALNTKNKLNTEGA AHFEKFYLENICIEWLQKYMTYGRDTERKGR

>Za2G-D48

IIFFFTVTVTHTLQYIRTA VTPGLNLPEYTDVGLVDGEPFVYYDSNIKKYIPKTEWINKITDDDPDYWSSGTQNQNYE
QEYKENVITLKRFRNHTEGLHTLQRIYGCELHDDGTVRGFEQLAYDGEDFMSLDLERVTWTAVKPQALNTKNKLN
TEGAATFEKFYLENICIEWLQKYVIYGRETLERKARPEVSFHK

>Za2G-D1

MSLCSTVMKTLIFFTFSLHLSSAVTHSLQYFYTA VTPGINFPEFTDVGKVDGQFQSYDSNIRKKIPKTEWIQKVTADD
PDYWNSGTQNLQGGQEIFKVNVASLMQRFNQTTGVHTVQRMVGCELDGGGTVRGYTQYGYDGEDFLIFDLKTVT
YTAPKPQAVISKNKWDNDRGDTVQKNYLENTCIEWLKKYVSYGRETLEKDRPTASVFRK

>Za2G-D6

TLCSTVMKTLIFFTFSLHLSSAVTHSLQNFYTG VTPGINFPEFTAVGQVDGQFVYYDSNIRRETPKTEWIQKVTADDP
DYWNRETQILQGNQENFKANVATLMQRFNQTTGVHTVQVMVGCELDGGGTVRGYTQYGYDGEDFLIFDLKTSTW
IAPKPQAVITKNKLDNPNPGVSVGRKNYLENICIEWLKKYVSYGRETLEKDRPTASVFRK

>Za2G-D18

VTHTLQYIHTAITPGINLPEYTDVGLVDGEPFVYYDSKIRKYIPKTEWIKRITDDDPDYWSSGTQIQNNAQEIYKVNVI
TLMRRFNQTEGVHTLQRIYGCELHDDGTVRGYEQLAYDGEDFISLDLEHVTWTAVKPQALITKNKLNTGAAIFEK
VYLKNMCIEWLQKYVITYGRETLEKDRPEVSVFHK

>Za2G-D25

SVTVTHTLQYIHTAVTPGINLPEYIDVGLVDGEPFVYYDSNIKKYIPKTEWIKKITDYDPDYWSSGTQNNQNDQEIKD
DVITLMRRFNQTEGVHTLQRIYGCELHDDGTVRGYEQLAYDGEDFISLDLEHVTWTAAPQALITKNKLNTAGAAT
FEKLYLENICIEWLQKYVSYGRETLEKARPEVSVFHK

>Za2G-D49

VTHSLQYFYTGITPGIHFPEFTAVGLLDGQHFGCYDSKRREIKTEWISFDDPDHWNRITGANDHQVAFKAHVITQM
QRFNQTTGVTQVQMYGCELDDDGTVRGYRQYGYDGEDFISLDLKTLTWTAAPQAVITKNKWDNNPGATVGEK
SYLEDTCIDWLKKYVSYGRETLEKVRPETSFLQK

>P28891-D3

SLSQMYIALNIRVVLVGLEIWSVVNFNIDGSAGEVLGRFTQWREKELVHRRRHDSAQLILNKYGSTAGMAFVGTA
CSRSHGGGINAVLSTFASIVAHELGHNLMNHDDGRVCRCDTGNCIMNSGATGSRNPFSSCSADDFEKLILNTGGT
CLLNVP

>P31989-D2

VAATMAHEMGNHFGMSHDSEGCCQALPKDGGCIMAATGYPFPSVFNQCNQAEKRYLNSGGGKCLFNLPNTR
VMYGGQRCNGYLEEGEECDCEVEECSSPCCNANNCTLKAGAECAEGVCCENCR

>Q1PHZ4-D1

VFQFYRVLNIRVALVGLVWSDSDKCAVTQDPFTTLHEFLDWRKLKLLPLRPHDNAQLISGVYFQGTIGMAPIMSM
CTAEQSGGIVMDHSDNPLGAAVTLAHELGHNFNMHDTPERGCGCGMTVERGGCIMPSTGYPFPTVFSTCSKKD
LAASLDKGVGMCLYNIPEVKVLYGGQKCGNGYVEEGEECDCEPEECVNPCCNATTCTLKEDAVCAHGQCCEDC
KLKPAGTPCRELSNSCDLPEFCTGSNPHCPANVYLHDGHACHTVDGYCYNGICQTHEQQCITLWG

>B5U6Y3-D1

SLQYNSQHYCGGTLIHPQWVVSAAHCWKPNYLIKVVNLNEYDLFKKEGVEQVFNVSKTLVYYLYNYRTFDNDIMLL
KLEQPADLNSNVQPVKMPSTDTPTFGGMLCVVSGWGVTVQVSYSPMLRAVDVQIIPNCQNYLYRITDNMVC
AGSPLGGKDSCQGDGGPLVCNGYLEGIVSWGIGCASPYFPGVYTKVRNYRWINLTIDAN

>T1DKS4-D1

MIIFKDIISGDEMFSDIYKLKESENGMMIEVEGKMISRSEGDIDDSLIGGNASAEVQDDGCDSTTVSGVDIVLNHKLQE
TSYDKKSYMVIKDYMKAVKTKLETCPDRVEPFMANAPAEVKKIIGNIKNFQFFTGESMNPDGSGILLDFREDGVTP
YMLFFKDGLESEKC

>P30894-D4

RGELSVCDSSISQWVTAVDKKTAIMSGQTVTVLEKVPVANGQLKQFYETKCNPLGYTKEGCRGIDKRHYNSQCRT
TQSYVRALTMDSKRKIGWRFIRIDTSCVCTLTIK

>Q91516-D1

DENGFCGGTLINQRWVVTAAHCLQETPDHVTLGDFDKFRPDAGEQKIKVEKVVVHPHFHEYTFDSVDVALLYLAEP
VVFSSVSPVCLPNTHLAKRLRPGENGLVSGWGATHFLGRSSRFLMKVSLPVVDQKECMDSTDQVITDNMFCAGF
LRAEKDACSGDSGGPFVNYRGTWFLTGVVSWGEQCAADGKYGVYTRIRNFWLHIE

>Q9I8W9-D2

CGGSLINEIWLVSAAHCFQSSSTGITINLGMESELTNSNQQRSSASSIINQNYDSTTKDNDIALVQLSSSVTFNNYIQ
PVCLAASSSSFPAGTEVWVTGWGTIASGVSLPSPQTLQEVQLPIVSNSDCTTSYGNGSITGNMCMCAGLAQGGKDSCQ
GDSGGPLVVQENGWVWQAGIVSFGYGCALPNIPGVYTRVSYQDQWISSWI

>E0Y421-D1

CGGILIHQQWVVSAAHCWRPSNIIQVVLSAHNLAEDGSEQVFNVSRIITNPTYNLKTYNGDIMLLKLSQPAVLNAY
VQPAPLPDDSTPPLDAGTTCTVSGWGVTRVYSFFLSPVLRAVDVDYIPNCYFYFRVNNMICAGSRYGGKDSCQG
DSGGPLMCNGILEGVSWGIGCANPYYPGVYTKVRNYSKWISGISSDS

>Q8JH85-D6

CGGTLINKYWVLTAACHCNVGVGNMRIVAGDYSVSVYEGNEQYRSPKLLVPHPLYNKTTNNADIMLIKMQTPVTVN
QFVFPAPLPRQGFDMPEGRVCRVSGWGFTSPTGGIPTSLHTATVPIVSSTHCNSTDSYNGNITENMICAGYAEGGTDA
CRFITLPSLQGDGGPLVCEGRVYGIVSWGNSCADPKYPGVYTAVSKFRTWIDSTVFGN

>Q3HXY9-D1

VDPKLFNKRRYRSPRVLFSELPPDSEPTEHQGSKDRTRKRRAGQPQSRGVYSVCESVSFWVGNKTKATDISGNEVTLP
 DVNINDVKKKQYFFETVCSGARTGGSGCLGIDVRHWSYCTDSHTFVRALTSFKNLVAWRLIRINVACVCVLSRKS
 >Q8AY75-D1
 PEQITEYKGVFEMFDEEGNGDVKTQELERLMSLMGINPTKRELSQMAKNVDKDGKGTFCNDRFLGLMALYHERAK
 NQDAELRAAFKVFDEAKGYIDWNTLKYVLMNAGEALSEEEAEQMMKEADKDGDTIDYEEFVAMMTGDSFKM
 T
 >P85031-D2
 VQVYCDMTTDGGGWIVFQRRQNGLTDFSRKWSEYRAGFGNLEDEFWLAQGRYELRIDMRNGQEAVYANYDRFSI
 GDSRSLYKLRIGEYNGTAGDLSYHHGRPFSTKDKDNDNAVTCALSYKGAWWYKNCHRVNLNGKY
 >P82942-D1
 YKPLNTFIALIGLEVWTDSDKIAVTPSSDTLGAFTKWRNEDLVKRIKHDNAHLITGIDFEGSTVGLAFIGTLCTGHST
 GVIQNHNPNAIAVGATLAHELGHNLGMNHDTSVCCTEDSCIMTAALSYTIPRLFSSCSINQFEQYLNRSPECLFNK
 PQPTTLILPPVCGNGFQESGEECDGSGVKIADATLMCRAKRDECDLPEYCTGNSPSPCEDVFAVNLQCKNGDGYC
 YNGQCPRLQDQCIKMWGPNAVVGDDFCYNQNTGRGTYYAYCVWPINGPYIACQKQDVKCGKLF
 >D8VNT0-D2
 RDCSDIYASGQREDGIYSVFIHFPSPGFQVYCDMSTDGGGWTVFQRRREDGSVNFYRDWAAYREGFGKITGEHWLGL
 KRIHALTIQANYELRIDLEDFENSTSFAYQGSFGVGLFSVDPDEDGYPLSIADYSGTAGDSSLKHNGMKFTTKDKDND
 HSENNCASFYHGAWWYRNCHMSNLNGQYLRGQHSSYADGIEWSSWTGWQYSLKFTTEMKIRP
 >P30894-D3
 RGEFSVCDMSHWVTDKTTAVDIHGYEVSVLTEVEIKRSTMKQYFYETTCQNSKPIKSGCRGIDDKHWNSSQCKTSQT
 YVRALTKYNNVMNWRWIRINTSCVCALSRKH
 >P67860-D1
 VVPFTEVYNKSMCRPREMLVEVHQEYPEDIEHIFIPSCVVLMMRCAGCCNDEMLECKPTATRNITMEVQKLKPMRIKR
 NYPMSFAEHTECECRVKKAMQENTDKKPRKGKGGKQKGRKKNRDKMRDFIHCPCSTCSERKKRLYVQDPET
 CQCTCKHSEADCKQKQLELNERTCRCDKPRR
 >F8S116-D2
 FCGGTLIDAQWVLTAACHCLERPSAYKVYMGHITERATEASKQIRDLQIIKGPPGTIDIALLLKDRPANLNDKVAKVC
 LPQKDYIVPSGTECYVTGWGETQGTGGEGILKETGFPVIENKVCNRPEYLNRRVKDFEMCAGNIEGGTDSCQGDGSG
 GPLVCYQNTFILQGVTSWGLGCANAMKPGVYARVSKFTDWIE
 >Q6T269-D1
 QWCALKKDEGPCKALKDRFYFDMEMFRCEPFYGGCQGNENNFETIEECEEMCLVKRDEPGPCRAVVPRYFFDSKV
 NECRRFFYGGCFGNANNFKTLKECKDRC
 >Q90W38-D2
 VDPKLFNKRRYLSRVHFSVVPFDGHTSPRTRRRKVRDFQNRGEYSACDSENHVWGNMTRATDLAGNEVTVL
 PDVRINNVVKKQMFYETTCRVTNRGSTQRGMKAGTTGCRGIDNKRWNSYCTNHTYVRLTSFKNQVTWRFIRIN
 AACVCVVSRSKS
 >P85031-D3
 PLQVFCDMTTDGGGWIVFVRRQSGKVDFFRNWRNYTAGFGDMNDEFWLVDLRDKGEHAYAQYDKFSVSEPRSY
 KVHVGGYSGTAGDSMTYHNGRPFSTYDHDNDIAVTNCALSYKGAFWYKNCHRVNIMGRY
 >Q8AY81-D3
 CGGFLINPSYVLTAACHCKSRHYHNFLNVVLGSHNINPERNDLKRYTVEKVHVHPFYKIKPDLGYDIMLLKVSKEISP
 NDVHVKTIEISSKHPDNNINCQVAGWGKTEDQVLSPLDLETDVTIINITVCEKEWTKADLFKLPDNLACAGGYETKS
 GACQADSGGPLVCNGVAVGIVSFNNNSNCKYPELPNVYTDISAYIDWINSVI
 >P0DP54-D2
 LELSGVVKCSTGRSTLAYIMYGCYCGVGEGWPRDPADWCCHKHDCCYAKAEDQGCYTKTHTYTPWSCDSQSLEC
 GSLTDRCEKMLCVDREAAKCLKKAPYNLKYVAVPDLFCPELPTCAYY
 >P0DP54-D1
 IHSLLVSMASLVDQRNVRSKRGLLELASIIKCTTGRSAFSYLMYGCYCGLGKGWPRDRADWCCHKHDCCYGDAE
 VAGCHTTTDDKYQWTCEDKEADCDSLKDRCEKILCRCDREAGRCLRKAPYNRKYAYWPDFLCGLYPTCNII
 >Q072L7-D6

VYNGHFCGGLINKDWVLTAAHCFSSMSSLTVYLKGQTLKGSNPNQIARSVKQMIHPNYSATHDNDIALLLLSSS
VTFNIRPVCLAGQGSSFPAGTNCWITGWGSIASGVQLPSPGVLQEAVVPTVNSFICDYLLGYGSITNMICAGYLQ
GGTDTCCQDSSGPMVAKKGAVWIQTGITSWGEGCARAFSPGVYTLVSQFQTWISSVINQN
>D8VNT0-D1
RDCSDIHASGQRENGVYSVFPTHYPAGFQVYCDMSTDGGGWTVIQRREDGSVNFFRDWDSYREGFGKITGEHWLG
LRQIHLSIQANYELRIDLEDFENSTAFAYDMFGVGLFSVDPEDDGYPLTIGDYTGAGDSLKHNGMKFTTKDRD
NDHSENNCASFYHGAWWYNSCHMSNLNGQYLHGQHTSYADGIEWSSWTGWQYSLKFSEMKIRPTREE
>T1DKS4-D5
VSGIDIVLNHKLQETSYDKKFYMYIKDYMKAVKAKLQETCPEVEPFVAIAPAEVKKTIGNIKNFQFFTGD SMNLD
GSIGLLDFCEDGLTPYMLFFKDSLELGKC
>C9E1S1-D2
TLLQFYRALNIRVALVGLEWSDSDKCPVSQDPFTTLHEFLDWRKLKLPQRPHDNAQLISGVYFQGTITIGMAPIMS
MCTAEQSGGIVMDHSDNPLGAAVTLAHELGHNFGMNHDTPERGCGCRVTAERGGCIMPSTGYPFPTVFSSCSKK
DLLVSLDKGVGMCLFNMPKMLVLYGGQKCGNGYVEEGEECDGDLCECMNPCCNASTCTLKLNVAHCAHGQCCE
DCQLKTAGTLCREPANSCLPEFCTGSDPHCPANVY
>T1DKS4-D4
GGGVSAEVQDDGCDLMTVSGIDIVLNHKLQETSYDKKFYMYIKDYMKAVKAKLQETCPEVEPFVAIAPAEVKKT
IGNIKNFQYFTGD SMNLDGSIGLLDFCEDGLTPYMLFFKDSLELGKC
>C0K3N5-D2
VVPFMEVYTKSRCNPRETLVDVQHEYPHDTHVYLPSCVVLQRCGGCCNDEALECVPHTNNTLELYRVKPGVG
EHTLLSFTEHTHCDRCVKPEVTKKEYRCEPCSERKKHWFVQDPLTCRCSTLTQLQCRSRKLELNERVCR
>T1DKS4-D3
EGDINDSLICGNVSAEVQDDGCDLMTVSGIDIVLNHKLQETSYDKKFYMYIKDYMKAVKAKLQETCPEVEPFVAI
APAEVKKTIGNIKNFQYFTGD SMNLDGSIGLLDFCEDGLTPYMLFFKDSLELGKC
>T1DKS4-D2
EGDINDSLIDGNVSAEVQDDGCDLMTVSGIDIVLNHKLQETSYDKKFYMYINDYMKAVKAKLQETCPEVEPFVA
IAPAEVKKTIGNIKNFQFFTGD SMNLDGSIGLLDFCEDGLTPYMLFFKDSLELGKC
>E0Y419-D1
CGGSLINANWVLSAAHCFQSSSTSGITIKLWKVWVSLANSNQQQSSASSIINQNYDSTTKDNDIALVQLSSSVTFNNYI
QPVCLAASSSSFPAGTEVWVTGWGTIASGVSLPSPQTLQEVQVPIVSNSDCATKYSSITDNMLCAGLAQGGKDSQCG
DSGGPLVVKRNGVWVQAGIVSFGHGCALPNIPGVYTRVSQYQDWINSNIGSN
>Q8UUJ2-D1
FCGGLINERWVLTAAHCFPSFSLGITANLGLSNLLGSNSNKQQRAIAALTIHPRYNSINNDNDIALQLTSPVFTN
YIRPTCLAASSTDFPHRTNVWVTGWGDIKSNVELPFPQTLQEVKPIVSNSDCAKSYGNDIITYNMMCAGLSEGGKD
SCQARNQHYYHISGHSTGLHGRKADIFGYGCALPNFPGVYTRVSQYQDWINSII
>P16354-D1
AWMIPGTLWCGSGNKASDFSDLGLFEDTDKCCREHDHCEQTISFQFGYGVFNSHFFTLSHCNCD SKFRRCLHNAN
DRMSDMVGYGFNVLMRCFEFSQRLECAERTWW
>P80966-D1
QPRALWQFGKMITCVQPNVNPFIYNNYGCYCGFSGSGSPKDQIDQCCLIHDKCYENARKHPDCPGVANLPYVKVY
NFSCSDKTITCSASNDKCKQAKVCECDQVAANCFQHNHTYNPNKNL

Data S3. The long-length toxin proteins in the yellow catfish.

>A7ISW1-D1
HTANTLSSDDLINVLAQTDIDRMWKNDLKPMLVVRYPGSPGSQAVQQRIKSTLSSLNAGWEVTEDRFQSWTPYQG
MPFNIIATLHPGSKRRIVLACHYDSKYFPPQWHGREFLGATDSAVPCSMLEMARALDNEKTLKVRSCGSDLTLQ
LIFFDGEEALYQWTSTDSL YGSRHLAAKMENTAHPVGATDTNQLDGLDFVLLDLIGGPMPHFGNQFSNTARWLSK
LQNIERRLHTLGQLEDHPNEVQYFWPGMPVGPVQDDHMPFLSKGVRVLHLIPTFPFSVWHTFDDNEENLDRATIQ
NLNKILQVVFVFEYLMN
>F8S0Z7-D1

PLLLLLCASCWLAAADWELTILHTNDVHARVEETNKDSGKCTKGCCYAGVARRFTKIQEIRSQKKPLLLLDAGDQF
 QGTVWFNYYKGSEAAHFMNRLGYDAMALGNHEFDNGVDGLKPFQAVNCTVLSANIKAVEPVASQISGYSSAST
 VLKVGTESVGIVGYTSRETPSLSMGPYMQFEEVAAALQLEVDKLITTSQVGNKIIALGHSGFETDKEIAKRVRGVDVVV
 GGHSNTFLYTGDHPSSEVPAGPYPFMVQSDDGRQVPVVQAYAFGKYLGYLKVTFDSGKVVKADGNPILLNSNIAE
 DPSIKAEVDAWKGLANYSAQFVGKTLVYLNNGTFEECFRECNLGNLICDAMVHHNIKYPDELQWNHVSACILNG
 GGIRGPIDERSRNGSITMEDVLSVLPFGGTFDLVMLKGSTLLQAFEHSVRRYGGNTGEFLQVSGFQIEYDLSKPSGERV
 KKVSVLCTECRVPHYELDTKKAYRVVMPSYLVDGGDGFMSMKKEEKLKHDSGDLDMSVFAGYITERQVRHPSVEGRI
 >F8S101-D1

AAVHKATVYWDPDHKTIVLLKDGVLDDTGDAYGFYNDLSSETGWGVLEIRAGYGRTPRPDDKTFFLAGYLEGFLTA
 RQMVDHYTNMYPQLITKPNILQAVKTFIHKQSDWSRQQVKLNKTSDDLWAHTGFILAQLDGLQAGAAEWAKRSGG
 KALSFEVHFLNAVGDLLDLIQLMLVPEAKPSLRGYKQPPMGHCSALIKMLPGYENMLFAHSSWYVYAATMRIFKH
 WDFNVQEPHTATGKVSFSSYPGFLVSLDDFYLLGSGLMMTQTNNIFNQSLYESVTPFSLFAWQVRRLANALAHTGE
 EWAKIFSRFNSGTYNQYMMVVDVSKVSLGSRLDDGALTIVEQIPGLVEYSDQTEALRFGYWPSPYVFPFHQKIYTLSCY
 EEMWMDYGEDFSYELCPRAKIFRRDQASVTDLSSLKHIMRSNDYKDPYSKGDACKSICCRGDLRTSNPTPEGCYDT
 KVTDLRMAQQFRSEALNGPSTEGDLPPTWDAFNSTMHQGLPRFYNYTFITMHPVL

>J3RZ81-D1

DQQVKYNMRKSEDLDDDEGCFLQKGHNEVLQQCGFNNTAKTILIIHGWTIGGVFESWMYKLVSADVQREAEANVIV
 VDWLGLAHQLYPDAVNHTLRVGHSIATLLDWLKDEQQPLENVHLIGYSLGAHVAGYAGTFVNGKVGRITGLDP
 AGPMFEGAEPNKRSLPDDADFVDVLHTYTREALGVSIGIQQPIGHIDIYPNGGDVQPGCALGDVLSSAAAGDFMEV
 MKCEHERAVHLFVDSLMDKEHMSFAYQCTGPERFNKGICLSRKNRCNNVGYNTRSVRRTRNSKMYLKTRADTPF
 AGYHYQMKMHVFNKRHSDDADPTFYIKLFGAHNNDNDIYVDIADGVGLNLTNTFLIFTEEDIGELLKIRLTWESPTE
 SFSAVWKHIKSFWSTSSSTKVLQVRRIRIKCGESQRK

>J3RZ81-D2

DINDACTVKLFQAHTLQSCSFNSSHPLVIIIHGWSDMGMIESWVTRLAAALKSTQKDINILVSDWMTLAQQHYPIAV
 QNTRVIGQEITQLLMWLEDLTQFPVSKAHLIGYSLGAHAGFAGRNLATSGRTLGRITGLDPAGPLFEGMSSTDRLSP
 DDARFVDIAHTFTQQHGLSVGIKQPVAHYDFYPNGGSFQPGCHLQVKNLTHLSQYGLMGFEQTMKCAHERAV
 HLFIDSLNDRDKQIMAYKCRDSAAFNGKICLDCRKNRCNTLGYGINKVHTSTSKRLYLKTRSLMPYKMYHFQFRIQL
 FTQFENTDLSLTIKLTGTLESEALPITLVKVSNGKTYSLMTVDTDIGDLMMVHVSWEAESVWTNMWSKMKITLPW
 GSKEDEPQLTIGKIRVKAGETQQR

>J3SDX8-D1

SEIRHWGYPAAEFEEVVTEDGYILSINRIPHGIQSKFQEEPKPVVFLQHGLLAAGSNWVTNLPNTSLGFLADAGFDV
 WIGNSRGNTWSRKHVRFPKQKEYWQFSYDEMAKKDLPAVINFITKTGQEIQIFYVGHSSQGTIAFMAFSTMPELA
 SKIKMFFALAPVATVAFTESPMGKLSILPEFVIWKLFGNKDFLPQSALIKFFATQFCSKRPISILCGNIFLLCGFDERNL
 NMTRTPVYTTTCHPAGTSVQNMVHWSQAISKKLMAYDYGKAGNIVHYNQSTPPLYNVQDMKVPTALWSSGGHDT
 LADPKDMAVLLTQIPNLVFHRNIQHWEHLDFIWGLDAPQEMYEPMIKLM

>P81428-D1

VFIRRKVAHGVLSRAKRANSWFEELKMGDLERECIEEKCSYEEAREVFEHTEATSEFWNVYHDGDACISQPCLNKG
 VCKDGLGRYSYCPDQFQGYNCEIAIQQLCIVNNGGCEHFCTVEKKTGTVFCSCADGYKLADDRKACTSDMITSAN
 MTITIDTQSQAKSNSLFGMLDPSPDTKQIPELPESDGEKRVNGEDCPPGECWPQALLINEDKIGFCGGTILNEYFILSA
 AHCMKLSRSITVILGETYTRSWEGREAVHEVEQVLVHANYKPDTFHNDIALIKLVKPIQFSKYIIPACIPDHDFEAENVL
 MQQDEGLVSGFGRLQEGGIQSTTLQRLSVPYVDRSVCKESTKYNISPRMFCAGYGREEKDACCQDSSGGPHVTKYKN
 TWFTVTGVVSWGEGCARKDKYGIYTVQVSKYLKWIEGVM

>Q90YA8-D2

SHKPRKLPLAQVRNLAAQVDITRLWETHLRPILIERVPGTAGSKAVLQHIVSQLRSLSAGWTVEDSFQSSTPKGPVTF
 SNVLAVLDPSAPRRLLFACHHDSKILPRDPKDPQRFVIGASDSAVPCAMLLELASALDTELKSLKQQRSAITLQLVFF
 DGEEAFEEWTDTDLSYGRHLAELMAHKPHPPGSTHTTQLQAVDLFVLLDLLGAAEPLIVNHFDNTAHWFDRLIA
 AERRLRQQLSSQGSEKNYFRKDVYLGVPVQDDHVPFLLRGEVPVLHLLISTPFPFVHTLEDTEENMHRPTVENLTKI
 LAIFLAEYLHL

>Q92031-D1

KNPSAVEYALDIDGKTFTISLEKNRELLAKNYTLTYTEDGAKEITYPSNVDHCYHGGQIQNINDSSSVSGMCSGMR
 GFLRAENQVYLIEPLEDSVKGDAVYKQEHLRTRATHGYINDTVYDYGVAAPRLAGLQKSISSCYITCAWICNKTKN

VLGNVLLNMLFNNIYFQLYRPLNIRVMLVGLEVWSNKDQIDVSNVPDHTLDRFLKWRQTDLLPRKKHDNAQFVTG
MDFYGSTVGLAPVKSMCSPSSGAVNEDHSKNPLGIASIAHEMGMHNLGMSHDTSMVFGVIFLLSSSCRYVLPDRFSSC
SKMDLETFLQNYDVRCLLNSPNEDDLFGGPVCGNAIVE

>Q92035-D1

EFIVTTRLGRVQGLRLPMPDRSHVVAFLGIPFAEPPVGGKKRFPKPAEPKKPWNDVFEATDYSNACYQYVDTSYPGFPG
TEMWNPKNIMSEDCLYLNWVWPASPRPHNLTMVWVIYGGGFYSGSSSLDVYDGRYLAHSEKVVVVSMNYRVGAF
GFLALNGSSEAPGNVGLLDQRMALQWVQDNIQYFGGNPKQVTIFGESAGGASVGMHLLSPESRPKFTRAIMQSGVP
NSPWATVSFDEARRRAIKLAKFVSCPDDDDAELVDCLRNKQPQELIDHEFRVLPYSSLFRFSFVPVIDGVVLPDTPDA
MLSSGNFKDQTILLGVNQDEGSYFLIYGAPGFSKDNDLSISREDFLQGVMSVPHANDIGLEAVILQYTDWLDEDNP
LKNREAMDDIVGDHNVICPLQHFAFSYANSAQQTGTTQGNLVSYYLYLFDHRASNLAWPEWMGVIHGYEIEFVF
GLPLEKRLNYTAEKKLSRRIMRYWANFARTGNPNINIDGSVESRRRWPLFTATEQKHVGLNTDAMKVHKGLKTQF
CALWNRFLPRLLNITDNIDDAERQWKAEFHRWSSYMMHWKSQFDHYSKQERCTDL

>Q9IAB0-D1

KYPEKLLFQLLNGNNHTVHLQKNRLLIGRNYTEIHYEADGSTVITSPKLEDHCYYHGRIDGIDSSVSVGVCSGMR
GFVRAEEQMYLIEPLGNSTEGDHAIFYKREHLRRKRSAYGDSGITVYDTEPRTEALFKRSSMVVSKFTPHITGISQLMI
YISLQYRRFKSNVDTVRARMLEAINHVDKFYRRHNIRVLLVGLEVWNVEDQFLVSVNDNETLTRFIQWRQKRLSNIV
KHDNAQLVTGVDFLDKTVGLANKFAMCTKASGGINQDHSPLGLAATIAHEMGMHNMGMSHDVRGCTCGLSDCI
MTEDMNSAASVFPELFDCLDQLQIFLENVNPICLLDRPSSDKIYGGPVCGNAFLDPGEECDCGTEEECNPPCCNA
KTCKFTEGSFCAQGECCENCQVEYTGRCRASVNDNCDLDEYCTGMSEKCPQDSF

>CL3069.Contig2-D1

WAWHLSVIFLSLILIAHHDVSAANRCAASKATSCSACLQTGFCAAYCPDEVFGFDRCDLLENLKS HGC VQIVSVQSSM
SMLKNIEINKNLKHSQVAPQQMSMTLLPGEEREVEMEVFEPARGPLDLYILMDFSNSMSDDLNLKKMGDELARLV
GTLSDDYTIGFGKFVDKVTETPQDMRPSKLKEPWAKSDPPFSFQHVITLTSNISTFRQILQKERISGNLDAPEGGFDAI
QAAVCQSQIGWRQFHSTHLLVFSTESAFHYEGDGINVLGILPRNDEACHLDTKGLYTHATQQDYPSVPTLVRLVQ
NNIIPFAITNHSLSYEMLLHYPIAELGVLTDSSNILNIIKNAFDSIRSKISIQTENKPKAIISQILSASGVASDYGKFKI
QPGEIGKFKVLLSAKTSVNDKPVCSLNTNERTGTIRVKPTTFSSALEIQTTVLCETCNCEQFPITKASRCSGNGDLVCG
TCKCHNNWLGPFCNCSSQMSSDVSGCVAPGSTQPCNGRGDCLCGTCLCYNPNQYEGQFCQYDRSQCHRSEDDSY
RGRCYMGGCVCDSGWTGNACEPLSNATCLDNKGGLCNGHGVCKCGRCRVIRCDCTSKMWECLGVILCGGCEC
EYAGLELGTTCENPFAQLGMCESRRSCVQCQAWKTGEIKGDKCKECPFTIKMVDLKD RDKVIETCEYRDEEDDC
TYHYTVNYPNNPSDKHEHEVEVLKKKDCPPGGFLWLIPLIMFLMLLLGLLLCCWKYCACCKACLARLCVCSPPCCAM
GRIVGFKEDHYMLHQSMSSDHLDTMPMVRTGPTNSTDVVRWKVDNVHRGPNLPQNQIQPNPKEIIDVPLSLRLHR
AFSDELSRPEARNTMLKQEVEENLNNVFRQIPGAHRVQKTRFRTQRNAGKRQNTIVDTVLSAPRSSYHNITNLV
HKQVQSGNFSDLKVVPGYYTVATDREAMGVVELQEDVESVDVRVPLFIKDEDDDDKKQLLVEAIDVPMGIANICKRY
VNITVIKEHAKSILTFLOPSYTFSRQDGVANIPISRDIIEDGRTQFTYCTRDLTAKDKKDYISVDGDLMFQPGETQKMVP
VKLLELSEGDALLDKQPKQFVMDLSNPLQGA LKGYPRTTINADKPESVIMFKNSTQNFNISDALYTPVIRTQGL
ENPSTVNWRTISSRFNLGSPVKFAPGEMEKNIVIDTRTQQMVPKESFQLELFDPSANSLIRDRKTTLVNITDSRGDFVS
KTGSVPVKASSPTGRIFAPTNIETPTGPKKIHLNWNPPQGA KGYKVYWIYGDPEAEAQVVDVKNTHAELTNLYP
YCDYEMRVCSYNAQGDGDYSDITQCQTLEDVPSEPGRLAFNVIGQTVTQLSWAEPAPNGVITEYEVVYTPINEDSK
PIGPDKKVKIDNPKKRMILLIENLQKSQTYRYKVRASNKVWGWPYRDATINLATQPLRPMSPPIPDIPVDAEAGDDY
DSFLMYSNEVLRSPSSSTRPSVSEFSEEQIINGKWDQGFLFPGSGSMSRNISTSSSYNLSPRPGGANQTVETTMTYITN
KGGSMRPRKHDIHTEDVTLRKSENRYNDNDGIRDSIVMSELTGFGFSEVLSGSSFSQSTTTSYSMNSRVHNHSDDVNE
ALQNLDRVLQETRLQRGVPETPSRLVFSALGPTALKVSWQEPHCETPIRRYCVLYQLLSGGEIKSIDINNPTQNSVMV
QDLLPNQSYLFKVKAEHSGWGPREGVITIESAVDPKSPSPVPGSSFTLSTPSAPGPLVFTALSPETLQLSWDKPRKP
NGEIVGYVVTCEQLHGGGDQRSFQLSGNSATTLTVSDLSENVYPYKFKVQAQTTQGFPEREGIITIESQDGSVGQYGS
QSVMRREVFNMPQSSTTQTTHMTFTDPFITPEGIMMSGRQVTQHSEISGSITRQVEMVQRGVKSTVTKKQYEEA

>CL5628-D1

MFSFVDIRLGLLLAATVLVVRQGEGEDDSIYNSCQLEGQSYNDKDVWKPEPCRICVCDSGTVMCDEVICEDITDCANP
EIPDGECCPICIDGTESPQISGPNNGDKGDPGPKGDPGPVQPGNDGIPGENGLPGPPGPPGPPGLGGYNLPQLSYGSE
KSSCGQAFFPGPPGPMGPRGPPGSSGSPGPQGFTGPPGEPGEPGSPGPMGPRGSPGPPGKNGDDGEAGKSGRPGERG
AAGAQAARGFPPTPLPGIKGHRGFNGLDGAKGDSGPAGPKGEPGSPGENGVAGVMGPRGLPGERGRPGAAGPS
GARGNDGNSGPAAGSPGPTGPGPPGPGAAGAKGETGPAGGRGSEGPQGARGEPGNPGPAGAAGPPGPPGTDGS

AGNKGSPGAAGITGAPGFPGRGPAGPAGPLGAVGPKGLNGDAGAPGNKGEPGPKGEPGPVGPQGLSGPPGDEG
 KRGARGEPPGGPGLPPGARGAPGNRGFFPGAEGGP GKATPGELGPNGPAGAQQATGESGRPGEPGLPSILVFLFI
 LMNMSTFLVLSYSPVSPKGVGTGSPGAPGPDGKAGPAGAAGQDQASGAPGAGSRGLPGVMGFPGPKGADGEPGK
 PGERGVEGPTGSGVAPGKDGVDGAPGPSGPAGPSGEKGEQGPAGGPGFQGLPGPQGSTGETGKAGDQGLPGEAGV
 HGPPGPRGDRGLPGERGTPGAVGPTGSRGSPGSPGNDGARGDPGAAGAPGGVGGAGPQGMPPGERGSSGLPGARG
 DRGDGGPKGPDGAPGKDGVRGLTGAIGSPGSPGTSGEKGEPPVGPSTGPRGGPGDRGEPGSPGPAGFAGPPGA
 AGQPGAKEHGDSPGKGDAGAPPPGPVGPAGHQPPGATGPKGARGGAGPPWTGFGAAGRVGPPGPSGVAG
 PPGPTGPGGKEGPRGARGEKAPGRPGVEVGGVAPGPSGERGSAGADGPPGPPGAPGPQINGGGGIAGGPGQRG
 ERGMPPGPSGEPGKQGVGTGAGERGPPGPMGPPGLSGGPGEAGREGSVGHDGAPGRDGAPGPKGDRGESCHPG
 APGAPGLPGPPGPVGSCKPGDRGETGPAGPAGISGPAGARGALGPAGPRGDKGEAGETGERGMKGHRGFSGMSG
 LPGPAGHPGEPGPAGPTGPAGPRGPAGSTGAHGKDGMMNGIPGPVGPGRGTGEMGAAGAPGLPGPPGPPGA
 GIGEPFPIMPQPEKAADPYHGYRADEASVQDRDAEVDSSLKTISQKIENILSPEGTKTNPARMCRDLRMCHPEWKS
 SYWVDPNQCSPLDAIKVFCNMETGETCVNPTRASIPLKNWFISKNIKKHVWFGESEMPDGFQFYQYGESEADSESDVS
 IQMTFMRLMSNKAQNITYHCKNSIA

>CL5628-D2

MFSFVDIRLALLSATVLLARGQGEDDRTGSSCTLDGQVYNDRDVWKPEPCQICVCDSTVMCDEVICEDTSDCAN
 PVIIPPDECCPVCDDGKSYFLPKASVSHTLSHLLFCFQGLPGPPGNDGIPGQPLPGPPGPPGPPGLGGNFSPQLSGG
 YDEKSGGAAMAVPGPMGPMGPRGPPGPPGLSGPQGFTGPPGEPGEPGATGAMGPRGPAGPPGKNGEDGESGKPG
 RPDGRGPAGPQGARGFPPTPLPGIKGHRGFSGLDGAKGDSGPAGPKGEAGAPGENGTGAMGPRGLPGERGRA
 GANGAAGARGNDGAAGAAGPPGPTGPAGPPGFPGGPGSKGEVGPQGARGGEGPQGARGEPGSPGPAGAAGPAG
 NNGADGAPGAKGAPGAAGIAGAPGFPGRGPPGPNGAPGAVGPKGNTGEGGAPGAKGDAGAKGETGAPGIQGP
 PGAAGEEGKRGARGEPPGARGPPGERGAPGGRGFPADGAAGPKGGPGERGGPGVVGPKGASGEPGRNGEPG
 MPGSKGMTGSPGSPGPDGKMGPTGPAGQDGRPGPPGPVGGRGQPGVMGFPGPKGSAGEAGKPGERGVMGAIGAP
 GASGKDGVDGAPGAPGPAGPAGERGEQGPSGSPGFQGLPGPQGATGEPGKPGEQGLPGEAGAPGVSGARGDRGFP
 GERGAPGIAGPAGPRGSPGSAGNDGAKGEPGAPGAPGAQAGPLQGMPGDRGAAGLPGLKGDGRDQGAAGKGTG
 APGKDGIRGLTGPIGPPGPAGAQQDKGEPGAAGPLGPTGARGPPGERGETGAPGPAGFAGPPGTDGQAGAKGESG
 DTGAKGDAGAPGAGATGAPGPQGPVGSTGAKGARGPAGPPGATGFGAAGRLGPPGPAGNAGPPGPPGAPGKE
 GAKGLRGETGPSRGTGEVGAPGAPGAPGEGKPPGAEGPAGSAGIPGPQGLAGQRGIVGFPQKGERGFNGLPGPSG
 EPGKQGGPGPSGERGPPGPMGPPGLAGPPGEPGREGTPGNEGSPGRDGAAGPKGDRGESGAAGAPGAPGPPGAPG
 PIGPAGKTGDRGESGPAGPAGAAGPSGPRGSPGAPARGDKGETGEAGERGMKGHRGFTGMQGPPPGPPGPPGESG
 PAGSSGPAGPRGPVGASGTSCKDGMSGLPGPIGPPGPRGRNGEIGPAGPPGPPGPPGPPGPSGGGFDIGFIAAPQEKA
 PDPFRHYRADDANVMRDRDMEVDTTLSLSQIENIRSPDGTKKNPARTCRDLKMCHPDWKSGEYWIDPDQGCN
 QDAIKVYCNMETGETCVYPSEADIPKKS WYTSKNIKKHVWFGEAMTDGFQFEYGESESNPEDVNIQLTFLRLMST
 EASQNITYHCKNSIA

>CL5628-D3

MLRFVDSRTVLLL VATQFILQAVVRCQQEDDQEDLGGCIQDQGHFEDRAVWKPEACRVCVCDSGAVLCDEVICEEL
 RDCNNPIIPPGECCPICPADQDQTSEQGPGRDGRPKGEKGIPIPRGRDGEPTGNPGPPGPPGPPGNPPGLGGNF
 AAQMAVGDFDEKAGGATMGVMQGPMPGMPGPRGPPGPSGSPGQGFQGSPPGEAGEPGSPGPMGPRGPPGPSGKPGS
 DGENGKPGKPGDRGPTGSQARGFPPTPLPGIKGHRGHPGLDGAKEIGAAGAKGETGASGENGAPGPMGPRG
 LPGERGRPGPSGAAGARGNDGLPGPAGPPGPVGPSPGSPGFPSPGSKGEAGPTGHRGAEGAQGPGRGEAGTPGSPGP
 AGAGGNPGTDGIPGAKGSAGAPGIAGAPGFPGRGPPGPQCATGPLGPKGQSGDPGIPGFKGEAGPKGERGIVGPQ
 GAPGPIGEEGKRGRGEPGSAGPLGPPGERGAPGNRGFPQDGLAGAKGAPGDRGVSLTGPKGGNGDPGRPGEP
 GLPGARGLTGRPGDAGPQGVGSPGAPGEDGRPGPPGPQGTGQPGVMGFPGPKGASGEPGKPGKEGLIGSIGVRG
 LPGKDGGETGAAGPPGPAGPAGERGEQGPQPPGFQGLPGPSGPPGEGGKPGDQGVPEGGVPGVVGPGRGERGFP
 ERGSAGAQLQGPRGLPTGTDGPKGAIGPAGSAGAMGPPGLQGMPGERGTAGISGPKGDRGDTGEKGPAGP
 KDGSRGLTGPIGPPGPSGPNAGKGETGAIGPSGPAGTRGAPGDRGEVGPAGFAGPPGDDQPGIKGEQGESGQ
 KGDAGSPGPQGPSGAPGPVGPVGSVPKGARGAQGAPGATGFGAAGRVGPPGPNGNPGSAGPAGPAGKDGPKG
 VRGDAGPPGRPGDVGLRGAPGSPGEGKGEPPGEGHQHPLGPPGPSGLAGQRGIVGLPGQRGGERGFPGLPGSPGEPGK
 QGSPGSGGERGPPGPIGPPGITGPAGELGREGNPGSDGPPGRDGAPGVKGERGNTGPIGAPGAPGAPGSPGPVGPV
 KQGDRENGEPQGPAGAPGTAGARGMAGPQGPGRGDKGESGEVGERGQKGHRGFTGLQGLPGPPGPSGPDQASGP
 SGSPGPKGPPGPVGPAGKDGANGIPGPIGPPGPRGRSGESGTAGPPGNPGPPGPPGPPGPGIDMSAFAGLGHTTEKGP

DPMRYMRADEASNNLRQHDVEVDATLKSINSQIEDIRSPDGSRKNPARSCHDLKICHPDWKSGEYWVDPNLGCTA
DAIKVFCNMETGETCVKPTTSNIPRKNWWSIKSKSQKHVWFGENMGGGFHFSYGEKIPTPNVASIQINFLRLSSEAS
QSITYHCKNSVA

>CL7398-D1

MWVSI GLIVALLFIALFLKYVFGNSGSPSPFDVDMREPLKPVQLDKKERNKVLKQGFLASRVPQDLDAIVIGSGIGGLSI
AVLLAKVGKKVLVLEQHDRAAGCCHSFSEKGFEDVGIHYIGNLEETGRFRCIVEQLTNGQLQWEPLENPFQDQVVL
GPPENRRVYPIYSGNTRFPEELKKCFPGEEKAIDEFMRLVKNCGRGVWMLLLKLLPSPVAKFLAYTGLANRMSYFFS
YGSRLSEVVNSLTENKELRAVL CYIFGTYNAPKEASFMSHLLVCHYLP GAWYPKGGASQIAYNMIPHIIEKAGGAV
LVRAPVSRILLNGANE AIGVSVMKQEEVHVRAPIVISDAGIFNTYQQLLPKEVQTHAAIQKQLSMLKHGEGGLSIFI
GLNGSKEELGLKAENYWIFCENNLDELFCNYIKGEREDSTKNIPMIFVASPSAKDPTWEQRQPKSTLSVVSFAPYSW
FEWKDGDGVKNRGGDDYEELKETIINSVLELLTQIYPEIKDKIEYIEAGTPITNQHYLAAPKGEIYGADHSTSRFTA EVCV
TLRPQTPIKNLFLTQDIFSCGFAGAIAGAMVCGSAILNRNL YRDIESLRKKLNTNSKKVQ

>Unigene1108-D1

MDQRVASILLFLVALVAGHGERYMVKKVMKAPYAVKGHVSVVAGEPGAPGEPGEPGPPGPPGHPGENAVGLPGP
QGPPGPPGPSYSAPGKPGTPGGPGKPGANGAPGPKGDTGSPGPQGPRGAPGPSGIPGPAGLSSPGKPGAAGMPGS
MGPRGETGPKGHPGIPGAPGQKGDRIQIGQPPGETGSTGPMGPV GAGQPGV GPKPGKPGYPGEPGKPGSSGLDGT
PGPMGPTGPKGHTGAPGIGMPGKPGENGAPGLPGPSGLKGPQGATGAPGAPGSPGYGKPGAPGFKGDTGPVGS
ATGQKGEAGARGATGYTGATGPMGPAGSQGPRGFPGEKGVTEKGETGPMGPQGFKGHKGQDQPGQPEGKSGY
PGAAGPQGPRGATGAPGSKGETGEAGATGAPGAPGVP GPKGHTGNSGPAGETGPAGAPGSRGSPSGSPGPPGAPGA
KGHPGLPGSPGAGLAAGKIPGPQGPPGLPGSDGAPGETGPAGPPGPPGPPGEVIAETKGT FVNEPFVKTPMSAFSA
LTTTPYPAGSPVKFEQVVYNAENHYDPESGIFT CQVPGVYFFSYSMHVNGANALVALYKNEEPI MFTYDEYNKGFL
DQMSGSTVLQ LNEQD TVYIQIPDDEANGVFAADNVHCSFSGLIAST

>Unigene1108-D2

MELHTVCVLLLLAVCVRATPDRIYIKKQSFVKGQELSGAPGIPGEPGPPGPPGPPGPPGMSIMGRQGP GPPGPA
GPAGYAAPGKHGSPGAPGKPGDNGMPGERGDSGPVGPQGARGPPGAPGSPGAGFSAPGKPGHGLPGAMGPRG
EPGSKGLSGIPGLPGQKGEPEGHGVQGHGPGGPGIPGSGPPGQPGQPGIGKPGATGYTGEPGKPLGRVGAPGPMG
APGPKGHQGAPGIGAPGNSGQNGAPGLPGMPGKGFQGPAGQPGAPGLQGVGKTGAPGIPGNRGAPGTPGTSGQ
KGEPGTTGFTGQPGASGPVGP TGPQGDRGFPGETGPLGPKGDAGIIGAPGSKGQKGD LGPQGFTGKPGIPGASGPQ
GPHGATGYPGPKGDRGYTGPSGSPGAMGPTGLKGHTGPQGMPGNRGENGLPGARGPMGPAGTSCSPGAKGNPG
LPGPPGPA GLVTKGLSGPQGPFGPGARGQDGRPGLAGPPGPPGPPGETVYYNEKSMPLKSEVYPVSHDLMKPRMS
AFTALLTTAYPNSGTPIVFNQIVYNGENHYDPSTGVFT CQVPGFYFV FHMHVNGANALVALYKNNEPV VFSYDEY
NKGFLDQMSASTV LMLHMGDTVYVQVPDDQSNGIYADNNVHCSFTGLIAST

>Un37646-D2

MRSILAFMVLGACTLSAWSKPINSRHVSFPGDILKNMTDIQMADSYLERFGYKRILDKSGRQGA VLTHKALRRLQTQ
LGLEETGELDQPTIDAMKKPRCGVPDVRNYQTFDGLKWDHNDVTYRILNYSPLDLDVSTIDDAFAFAFKVWSDVTP
LTFRRLYDGTADIMISFGKKDHGDPYFPDGKDGLLAHAYPPGEGVQGD A HFDDDEYWT LGKGPAIKTYFGNADGA
LCNFPFRFEGKSYSSCTTEGREDGLPWCATTANFDKDKKYGFCPSELLYTFDGNADGVPCVFPFVFEGKTYTSCTTEG
RDDGYRWCATTSSFDKDKKYGFCPNRDTAVIGGNSEGEPCQFPFIFLEKSFTSCTSEGRSDGKLWCATTSNYDKDQK
WGFCPDKGYSFLVAAHEFGHALGLDHSNIQDALMYPMYKYIADFSLHQDDIEGIQYLYGPKKGPKPTPPKPTTTA
STISTAKPTKKTPTTPSTASTTTPSVFTPVDPSPDCTADRFDAITEIQGELHFFKDGYYWTSSSSGNKERKGPFLVSE
WPGLPAKLDTAFEDPITKQMYFFAETQFWEFTGKNVRGPRSEIKLGLPANVDKVEGSLQRGKGKGVLLFSGENYWRM
DLKTQQMDKGFPQQTDMTFGGVPVDAHDVFLFKGNYYFCRNIIYWRMTSKRQVDRVGYVYELLNCPDY

>Un37646-D3

LKKMQKFFALPETGEIDQNTVEIMKKPRCGVPDVANYNFFPSKPKWQNN EITYRILGHSPDMDEETIDDAFFRAFKV
WSDVTP LKFTRIMSGDADIMINFRNEHGDGYPFDGKDGLLAHAFAPGPGIGGDSHFDDDEHWT LGDGQVVKVKF
GNAEGEFCKFPFLFMGKL YNSCTNQGRDDGFLWCSTTYDFDKDGKYGFCPHELLLT LGNGDGAPCKFPFTFQGEK
YDSC TTSGRDDGYRWCATTENYDQDKTYGFCPETALSTIGGNSEGAPCVFPFIFLGKSYDACTTSGRNDGKMWCSST
KSFDEDRKWGFCPDQGYSLFLVAAHEFGHALGLEHSQDPGALMAPIYTYTKHFRLSNDDIRGIQELYGAPTGKPLPP
TQGPVTPMDICSENIIFDAIAQFRGETFFFTDRFLRTTNAEKKPSGPLL VATFWSELPEKIDAAAYENPLEEKSVFFSGDE
MWVYSASTLERGYPKKISSLGLPSDLQHIDAAYAFHKSCKTYL FAGDKFWRYNEAKNKMDPGFPKLIADSWNGVP
DDLDAAFSTSGQGYSYFFKGSYHKMDENLRIVKVG EV

>Unigene40283-D1

MLSFVDTRILLLLAVTSYLASCQWPRGDKGPRGDRGPKGPDGKPGKHGLPGPPGPPGPPGLSGNFAAQYDGSKGPD
 PGP GPIGLMGP RGP SCPPGTPGPQGPQGHAGEPGEPQAGPGGPRGPPGPPGKSGEDGNNGRPGQPGDRGTPGSQ
 GARGFP GTPGLPGMKGHRGYNGLDGRKGEPGEMGTKGETGAHGSNGTPGQRGARGLPGERGRPGPPGPAGARG
 ADGNTGPSGPAGPLGAAGPPGFPGGPGPKGEMGPAGPSGPSGPQQQRGEPGTNGVSGPVGPPGNPGANGLNGAK
 GAAGTPGVAGTPGFPGPRGGPGPQGPAGPSGPRGLSGDPGPVGKGESGAKGEPGNIGAQQPTGAAGDEGKRST
 GEQGSAGPVGLRGARGAAGTRGLPGLAGRGGSMGMPGARGAAGAPGARGPPGDAGRAGEAGLVGARGLP GS PG
 SPGPQGKEGPAGPSGQDGRSGPPGPTGPRGQPGNIGFFPGPKGPSGEPGKPGEKGLAGAPGLRGPPGSDGNNGPAGP
 VLAGGPGKEKGEAGPAGAPGFQGLPGPAGPVGETGKPGDRGIPGDQGVAGPAGGKGERGNPGPAGASGAQQPTG
 PRGPSGTPGPDGNKGEPGSAGLAGAQQGPQGSVGMPPERGGAGTPGVKGEKGEFGYRGPEGNAGRDGSRGAPGPI
 GPP