

Supplementary Data

Identification and comparative analysis of the epidermal differentiation complex in snakes

Karin Brigit Holthaus, Veronika Mlitz, Bettina Strasser, Erwin Tschachler, Lorenzo Alibardi, Leopold Eckhart

Content

Supplementary Figures S1-S10

Supplementary Tables S1-S6

MSQQQQCKQIPCTPPTSSCKCTSSCQKPCEPLACQGSAGPQCPGSTGNSSQKPTCPPGKNECFCHCHQQQQC

MTHQCKLPPELPPACLKLATKTSLOTKSPPPPQTVTVCAVFTPAFSPPPPQSVTCGGVFLQVPQSPPPPQRTVTCAVFAPAFSPPPPQSVT
 CCGVFLQVPQSPPPPQRTVTCAVFAPAFSPPPPQSVTCGGVFLQVPQSPPPPQRTVTCAVFAPAFSPPPPQSVTCGGVFLQVPQSPPPPQRTV
 TVCAVFPAPAFSPPPPQSVTCGGVFLQVPQSPPPPQRTVTCAVFAPAFSPPPPQSVTCGGVFLQVPQSPPPPQRTVTCAVFAPAFSPPPPQSV
 VTCCGVFLQVPQSPPPPQRTVTCIAPIFALAPPPAQSITCGGFLQVLENNPPTVTICASFAPASAAFTLVSVPCSSACTSCVFTLSLV
 QFTCVAVPCCSSCKTCHGS

MAHCSKCGSPCCCKKHGHSHCRKKCRNKGCSPPCCCKKHGHSHCHKKCRCKKCGCSPPCCCHLCSGSPSCCQQCGGSPCCCGGCGGSP
SCSQQCGLSSSCCKKCGCSPPCCCGGCGSPSCCKKCGCSPPCCCGGCSGCSQCCCEKSCCCTPCSCCSPCCGSGKGCASAKGGSSTGCFPQQ
CEC

M T C C S K G C S F P C C C K Q S C C S F P C C C N Q S C C S F P C C C K Q S C C S F P C C C K Q S C C S P C C C K Q S C C S P C C G Q Q S C C S P C C G Q Q S C C S Q C S C C K S C C
 C T P C S C C C S K G C S S K D C S K G C S S K G C S S K G C C S K G S C S K G C C S N Q C K C

MASSGTTKTPKCSKCGSPCCCFKASSSTKCSKCGSSPCCCSKASSSSKTCKSKCGCSFCCCSKSLCSKRCRSPCCCFQSRCSKCGSPC
CCFQSQCSKCGSPCCCFQSSXXXSPCCGAQPRCSKCGCSFCCCFQSRCSKCGCSFRCSSSRCSKCGSPCCCFQSRCSKCGCSFRC
SQSRCSKCGSPCCCFQSPCSKCGSPCCCSQSRCSKCGCSFRCRCSQSRCSKCGSPCCCFQSPCSKCGSPCCCSQSRCSRCGSPCC
RSQSPCSRRCRSPCCCSQSRCSRCGRSPCCSQSCSRCGCSPCCRSQSPCSRRCRSPCCCSQSRCSRCGRSPCCSQSCSRCGSPCCR
SQSPCSRRCRSPCCCSQSRCSRCGRSPCCSQSCSRCGCSPCCRSQSPCSRRCRSPCCCSQSRCSRCGRSPCCSQSCSRWGRSPCCP
QSPCSKCGSPCCCSQSQCSKCGSPCCCFQSPCSKCGCSFCCCSQSRCSRCGCSFRCRCSQSRCSKRCRSPCCSQSCSRCGSPCCCS
QSRCSRCSFCCCSQSRCSGRSPCCYSQSRCSKCGSPCCCSQSRCSKCGSPCCCSQSRCSKCGSPCCCSQSRCSKCGCFPCRCSQSRCCRCGCCP
CCTPSCCPRCNCPYCCQNSCCCPCCRYSCCCPCCYSCCNSSPLKQPK

MTCCPICGSSPCGAPCYSCSPCCGCGSGSPCCSSCGSPCCSSSCAPCCSCSPSGGSGSGSSSCPQCSSSPCCACGSCCNQSS
GCGCGCGSCPCSSCPCCCGSGCCSPSCSSCCGSPCCSSCGSSCCGSGCCFPKQCF

MKELPPCGWFCFCNVFLQRLGLHDPCKKVVDPFQLHLSLFLILLPSLCTMEFEYVFEILLDETDPFTFLETYRQRWVKWVDPCKERYCPPK
 YFPFPCPPPCPPPCPPPKPQRIPCCPLYELFQCPKYFPCCPPYPCCYFPQWCYPRFLPCVPQKFWYFCKVPCPFGWVPKYPRFYDLLE
 LY

MA**SS**HNQH**QCKQ**AP**TL****S**EL**CK**SAP**CP**D**QCKQ**PPVV**VP**TP**CP**EP**PK**CP**PP**EP**CK**EPPIV**IV**EP**EP**PK**CP**EP**EL**CK**EP**EA**TP**CP
EP**KE**EP**EP**CP**PK**Q**EP**TP**CS**Q**QCK**S**QCK**EP**VP**VP**VP**TP**CP**EP**PK**CP**QEP**EP**CK**EP**VP**VP**VP**TP**CP**PK**PK**CP**QD**PP**CK**EP**EP**AP**TP**CP**PK**EP**EP**
CP**PT**TP**CS**Q**QCK**S**QCK**Q**PP**VP**VP**TP**CP**EP**PK**CP**QEP**EP**CK**EP**VP**VP**VP**TP**CP**PK**PK**CP**QEP**PP**CK**EP**VP**VP**VP**TP**CP**EP**PK**CP**QEP**PP**CK**
Q**PP**CP**DQ**Q**QCK**Q**CP**SW**PP**Q**PN**

MYTKGYGDGGGSYRNYGRSACCEPSCHSVRRRSSEVKCCDTS PQSCCPEVQKCCLEQKYCPEVQKCCPFPRKYCPEVQDYCPEVQFCCS
EVQKYCPEVEFCYPEVQKYCPEVQCCPEVQKYCPEVQFCCPFIKKYCPEIQCCPEAQKYCPEVQRCYPPFQRYCPEVQFCCPAEQRY
CPEVQCCPPEQRYCPEVQCGSPPEQRYCPEVQCCPPEQRYCPPIQCCPPOQRYCPEVEPSCPELEISQVQVQCRVPPHLLK

MSYHOOOC~~K~~OPCO~~P~~PPPOCOKT~~S~~CPPACPPOKCVPKPCPPOE~~C~~VPKPCPPOE~~C~~VPKPCPPOE~~C~~PPKPCPPPOOCCONPKRC

MSYLNNOCKOPCPIPPPOCKKTIPPPEOCCPPPPOOYCPPPPOOCCPOOODODPKCC

MFHCCVPSCSMGPTAPLCFQFCQSSRMETISIPCRVHPHSMPPPTSIINLCPVPSMPPPSSSPSICIPLCPSPSSSQRPSPILCLLS
PQFSGSSSGNSRISISLSPSSSSQSLSIPLCLCPQFSSSSSGSSSPSISFSFNQSSSSQELYCIPLSPSQPPPSISISFSQS
QSSSSHSICIFCPSPLSSSSQPSFCILRPSPTSSSSSSSPRISISLSPSPSSSGQSCLILCFSPQPTSGSGSGSSPRISIS
LSFSPFSSSQSPSYCIPLCFSQPPPSMSERPAPPPTCCIPIIIISNCFCGC

XXXXSSSSGSSQSPISIPLCFSPQPSGSGSGSGSGSGSSPSISISLGLPSHSSSSGSSQSPISIPLCFSPQPSGSGSGSGSGSGSSPSISISLGLPSSSS
SSSQSPFYCIPLCPTQSPSSSMSPPSSSSPFTCCIPISISNPPCFVC

MSSCCVPSRCVGIAPLRVQVRDQSSSLPSGISCRVQVPSMPPTSIISLSVPSMSQPGPSPLSSCIPLCPMQSPFSMSLSPPPPPT
 SSOPPSYCIPLCPSPOPPPSMSETPSPPPSCCIPITIPSTPCFIVF

MDCCSQQCKQFCLPPPIQQQKIGIKRTLWQQRGSSQCTKPCSPSWPQSYDECPPEAYFPQGRAVWAQFVQQCPSPWRQPCSPFCCKSVS
FNSKGLRAPHGMQQSASKKRRQCLLQCSQDQPHQCLKVCQVKGCKKQANKGSPQCCTKESRGGGLPCAQYPHTVKGDWTFWIKGNQQCVS
QPCVTKVNPVLYSSKGSQQYRRKGGYCKTQLKVSSNGQKYCSASNNWFW

>Pb_EDQL_partial

MCSSHHDRGCHGVSHDHGCHSGSSCHEDHGSCHNRSSCHDGGNSFRVMTIEVAQDKNEMPIVQQQSEVVYPAPKLQQTQQLKQPTTYPE
QQQLQQPQQQQKQKSKTQRKSSXXX

>Pb_EDQM

MCSSREDKDQCYKQEKDESPCCYSHRSSGCGNKSSRACSGSYGWGESSPYCQEKSDQPPQQQQKKQGCQLPPQK

>Pb_EDQSG

MQQCKQPGFPPVGQMSTAAQCKASAVSQSQGATVDQCHGSSGNQTQGSTGIQCGGSAVSQGKSGTGNQCGSMESQTOSSGVSQWRGS
TGDQCGGSAVSQSKSGTGNQCGSMFNQSSGVSQWRGSTGDQCGVSVVSQGEQESVAFQDQESMVFQGGGSPRNQCGGPFVSDQGST
GCQCCHESAGTKY

>Pb_EDSC1

MSQQIQQSSCCCCGSSGGCCSRGSSGGGCCSSGGVVYVQSSQQSSRGCCINRTGACCGRCCGRRCGGGCCRRGSSGGCCSRGTGGGC
CGSSGGAIYVQSSSRCCCINRTGFCGGRCCGRRCGGGCCSSGSSQKLLK

>Pb_EDSC2_partial

MSQQQQRGGCCCCRCGGGGGGGGCGCHSGGGSVSRSSGGGCCGRSSGGSSQSSXXXSSGGGCCGRSSGGSSQSSHSSGGGCCGRSSGG
SSQQTHSSGRCCGGGGSSQKMK

>Pb_EDSCP

MVFSCHRPTTSSSGPIFFISGGSHSCVEVCQPTCSIPSGPTTTRSSGPIFFISEGSHSCVEVCPPHCCIPCGPPTVQVQPPPPCVFV
CFSPCCI PCGPPTVQVQRPQSCIFVCPSPCSPNCGPPTVQVQRRSSPCCI SCTTFSRPVITVSPCCSVPSCSIAHLCPQPGSSSCVV
FSGFIFLVSSGQSCIFISVHPCSPCLTFF

>Pb_EDSPR1

MACFYQQCKQFCLPPPICGKGGVKQCAVCPFACFCQSVCAVPSVKVVDGCTFIQIFPCAPKCAVDNICVQSCPFSGQSCICVQPPQVCF
CDSKCEDECCCKQSGC

>Pb_EDSPR2

MSQCKQACKAPFCFECCEFTCKCFPKTCCSSPGTKAVSKKGGCCPFPSADVCKPPPKVKEPSKDKCC

>Pb_EDSQ

MSYQCKQFCLPPFFMKGTTVCAGFGTTVCVSTQGGQSIIEVCASFTGAICVTOGQNAQCANVCQDCCGSSVVVSFVQNGQFSPCATICTK
QNGAVCVTKAPFQSSAVCVQQQACQCSVCQGNISVAEIQSQSGACATICASSPSSVCVTPIATLGLAATICTSFGACATVCQDPCGS
VCVTSKGATKCATKCATKCAFGATVCVDPCGGVKSVKSCCTKCTSPCFAVSNEQVCVKACPSISMNQCNVKKC

>Pb_EDSRWM

MNFYMPQEYWDLNSWETNYENNYDFPSEYYYAYNERSYILDYQSRWGSFSSYSSSDCCDPCYSAICGSPFRYRSRSSYCFSCAPSSNSSW
GSPDCSSNSSCCPFCWTGGSNFRYSPICRFQCHRSRVSFFRSNRSSSCCPCCRCGRWAPFSSSSSQSPDGSNSSYCFPCWTGSGC
NHYSPICRFRCHRSRVSFFRSNRSSSYCFPCRCGRWAPFSSSTQSFPDGSNSSYCFPCWTGGSNFRYSPICRFRCHRSRVSFF
RSNRSSSCCPCCRCGRWAPFSSSSSQSPDGSNSSYCFPCRCGRWAPFSSSQSPDGSNSSYCFPCWTGGSNFRYSPICRFRCH
RSRVSFFRSNRSSSYCFPCRCGRWAPFSSSQSPDGSNSSYCFPCWTGGSNFRYSPICRFRCHRSRVSFFRSNRSSSYCFPC
RCGRWAPFSSSQSPDGSNSSYCFPCWTGGSNFRYSPICRFRCHRSRVSFFRSNRSSSCCPCCRCGRWAPFSDSCQSPDCSCH
SSCCPFCCTCGSNFCYCFICGPOPSNFRSSSRRCFCCRCGRCAFPCCSSQCCRSSSCCRPCWTCSSNFCYCFSCRPQPPNSSFS
SCCCPFCPCGCGYGPYDPS

>Pb_EDWM

MSEERIYSSGREHYFNLNSTWYDEAGSWLDNRRKFFCYVENTACVTCCNERTNVERRGGHNYRCYCYRRSTCRPGGNERVRCCVHNHSG
GFRDYWGRFIDGACNGCTGGYYSHGDCGSCCGSSGECGTGGGLVACAQPCVSSGGVCAEPGCRFAGRVCAPFCVMSGGWGSGRGVCA
EEECSSASGCGSRRRGVCSSEFCARASGGC

>Pb_EDY1

MIGITILSSHSEGTAKIYCPPTIYKPSGSDSSAYPPYTLAGTAARCYIERKSRYLEGPTYYLSGWDCSRPMPTHYRSTYGPCSYEYL
LDNSQGGYDPCVYKYLKGVQRSYDPLYLEYLYSYRRRYDPSDYRYPSYGGRCDFCVYNNPSYGRSSYSRGYTYRCGFSNDPCGY
SCAFGYRQPRYADRRRYSSRPISSDPYGSCPESFYASECRRPSSDSSYCGPC

>Pb_EDYM2

MHYAYHCKEIGFPLELFVKKRLFRYGSQYFELYGSRVYSSRSIFHRRGIWMTYSLASVAGYQFWIIPRVSEACSSKSPPKLRTRNE
VRHYVDYQYFLQTTVPYDGMTKGFHGFINKGSLGCVIFERVLVTKGFVLVSSKDLHFSSAQILRRACKGPFVYASQDPELLCSSKEE
LVRVSKGFDQYLLKGAQRRAADSYLMTEDPQSGAKVTQERATKGERLIATKVERFSALTVSRSFLLRGVHSLTKTSQTNLCKGSR
SSLSKTSWLSLKGSGPFLVRTATTAPKTLKSNVKISSGKKYCSSTTKWPF

>Pb_LOR1_partial

MTSWQQTSGCCSSGSI CCYFCQGESSCCC GGGRRDRGRSENQSSQGSICCGSSGSSGSGQTIIIVFGSSGGDGWCCCCGGGGSSD
GSSGGSGQIIISGSGVGSRSGDRQSSGCCIGGSGWGGSGSAMSQOKDGGEI CCGGVSGGGSPGASGGVTINMGSSGGGSGGQTII
VFGRRGSGSGGLSSMGGSGGGSGGQTIIIVFGRRGSGSGGLSSMGGSGGGSGGQTIIIVFGRRGSGSGGLSSMGGSGGQTIIIVGGR
GSSGGQTIIIVFGRRGSGSGGRSXXXSGHSSSMGGSGGGSGGQTIIIVFGRRGSGSGGLSSMGGSGGGSGGQTIIIVGGR
RGSGSGQSSGSSMGGSGGGSGGQTIIIVFGRRGSGSGGLSSSMGGSGGGSGGQTIIIVFGRRGSGSGGLSSGSTGGSGSGSG
GGQTIIIVQCGGASGGSSRAQSSGCCIRGGMGAGSMQTKQFISIPFCIGFTK

>Pb_LOR2_partial

MSKQIQSSVCCGGGSSAGCCVRSSRSGGSSCCCACCCCRGGSGHVIASGSGSSCCGGIQISSSSCCGGSSSTGVVVVPGGQSSGCC
IGGGHGGGIQQKIPLIGVGGGICCGGSGGGSGGQTVIVPGGSGCCGGSGGVRVVGGSGSGWCCGGSGGGQTVIVPGSGSGCCGG
SGGVKVI GGSGSGWCCGGSGGGQTVIVPGGSGSGWGGSGGGGGSGGVRDVGGSG

>Pb_PGLYRP3

MVTLQILFLIICSLSQATGCFQLITPSKWAKFANCSQPLRDVPEYVVIIHTAGNFCRTHADCRKEVKNIQDFHMLKGCWDIAYSFL
IGEDGYVYEGRCWRNEGSHTYGYNDLSLGIAFIGTFVERSPODVAWALRCFLNFSVKIGYLAPEYILLAHSDVSDLVSGGEFVRAEIS
KWNENYKHSFYVLDREGGQ

>Pb_SCFN1

MTYFLDSVCTIVGIFHKYAQWGRNLTNLNRREMKMLIQTEFAEVLNFCDFQORVEFTFQLLDVNGDSLVDNFNEYLIFFIQIAKACYSYL
QREYLLQEDGSRALHEGEGGSKRDHWQLQDEEREEDYVHERRGSDRTPLRMEDSRRGELGKCYLSEPEEEEEEEKEEERNFQGRDR
ELRDGDRRIHSWEHQERETERRRLEPKQOEDIETFEHHQLRQREPALEEGSPRRRRRNDNDRQADEHLDRDEHREGRETEHREAEG
RLFRCCSKARVDQGNHHAEYNSERSQASCEPRELEDGRRDLETHDELRTAFEHVTDERRDQIRNRREVDDYEEHRLREGERRSS
HETEHRESERRRRDHSAFQSRDDVRIANRAEEEEVRGCERRRDEDEWSSRQCEALRRPEVEELEHRERERRSYSELENMDRRNQHE
SEGLRRSRHREDEERRRQFSTWEETQATESRRRRFSHEPEPRDFKYGRSQRYEEEFKRADQRRQHNVEVQSPERDVARRRRLQTR
AEREDNQERQQYERESRSGSGDRRRFQLCYSEPREGEKRRHQLTDSRVVERERLQSRRESVTRFTQDRFSFSDSSGEYDWSRQQIYES
RDIEQREGRRRSYDRRSCGQRQIQGIDIDPREGEGQRRTNAAEGDVDQRTQTRGVDHGESEQQRRTSRRDSDGDKDIDQRRQTDEASS
RDGEDQRRTSSSHDSNTRQVNOQMOTSDPDFRDEIPIQHNKGNAPQREVPSLSNQTGDREFAEDLLRDQRPPESGESQLSRKPRQHK
EQSSSIEPTLEQKQATRGSQCLSGFELRGFORRWLQPHQTEAQEGEAGVEQTQKVGSFESKRAASQFGEQOTTEQESRVQRHHQKFA
TODEGESVAGKGKAQEPATVQKQQLQAGASYEGEGEQEVEGFKKVEEDDENGQRAQESQPELEDDQOAVTEKPCDSLGSKSSVVC
NELYEYLLAQKQEQP

>Pb_SCFN2_partial

MACLVDVSVCTIIGVFHKYAERNACSTMKREMKRLIQKEFGEILENPRDPIQVLTFLQLLDVNRDSLVDNFNEFLLLIFEVATACYSSG
YFKECLSPKEERSRVVRNGEPRGNENNHRQFQDEEREEDYVHERRGSDRTPLRMEDSRRGELGKCYLSEPEEEEEEEKEEERNFQGRD
RELRDGDRRVHLSREQHKKETERRRLEPKQOEDIETFERHQLRQREPALEERSRRRSREPMRRNDSNRRAGDFLDRDEQDLYSEVVT
REDGRQRRGRETEHREAEGRLFRCCSKARVEEGNHRAECDSERSRASREPRELEDGRRDLETHDELRTAFEPRTDEXXATAEARRE
RHHEDDEQRRQSNREETRATESSRRRFSHEPEPRDFKYGRSQGYEERRRDDRRQNYKVESPERDVARRRRLQTRSAEREDNQDR
QDYEHESRDSGDRRRFMILEPETRDEQRRQETGSRYGERERRQERDGRIRRTGRELGSESDGGQRELDSSRQISRCRTAELSRSE
PRSRKFPARRG

B

>Pb_Beta1

MSGSGVKCVTTPCITSCFDAKVVVHPPPLILTLGLSLRTSNQCLLETQTSCLTNGSEVGCNEGSTAIVTKSSGVCALSGGDTCCCTT
TCHDSQVVIQPPPVCTITPGAVLTSPNECLIKSTPCVSEGPQPHALTRSTSVDCLTPRRLSRASVPSGLDSLRCVTRGPPTKV
VIHPPPIEVTIPGFVLEIAAQCAVEVYNPCSSNNALTSSDQKAITSSDEGEVKALTTQAKSCCTVCGLMDTSSCISQGPCHKIIPPF
HIEVELEGFIELVFBEACKIETLNSLELEPAITGSETKALSSKMTSTALATTSNVKRELENLRRARREWAEMYSRSMTEPTIALSQQ
TRLAKYRNTLHSMNFQPSF

>Pb_Beta2

MERRPLCYASCPESTVTIQPPPFRTLTIIPGPAIFCFNQPFHIMQYNPCARDGMGIGGRMMFSDFISEDLPDLKLESSSGHGSALVTLYRS
RMALLAMTGSN

>Pb_Beta3

MSECYASCPEASTVTIQPPPFVLTIPGPALYCPDQVFGIEQYNPCAGFGGMVTRGGGGIIVEGRGGTLGALGAGRGWGGSWSGSGIGS
RIGGGGLIGGVGGPRGGGLIGGVGGRRGGGLIGGVGGPRGGGLIGGVGGPRGGGLIGDVGGTGGGLDIDGRGTTGSWSGGSGV
GGTVGGRRGSIGAGSITGGDLGTGGVGGTRGGLSTGGTRVTTGGWSGGSRGGSGVGGRRSIGAGGTVAGWSGGGRRGSGVGTVSVTG
GDYGLSGAGGGWSSSHRISGSHRWSGSGYGGSGGYSGGGYGGGYGSSGYGGGFSSGRYGTSGYGGGYGHTGGYRRIKYSVGGGSGHRSG
GFSRSSYGGSYGGTLAVLPGPSDISCY

>Pb_Beta4

MSSTSYNPADVCAFSCPPTMTVTIQPPPFRTLTIIPGPTIHCPDQPIQFQQHNPCVFNKEHSPLLASRNSSFYSRALPSGSYTSSIYRY

>Pb_Beta5

MSANRSFWPNQYACCPASTVTIQPPPFVLTIPGPALYCPDQPLGIDQYNPCYEGAYPLPLGRGGSNLTNFYSRGLPSGSFSSSTTYRSSN

>Pb_Beta6

MASGWNPCASCPEPLTVTIQPPPYTVNIIPGPSLYCADQPLCIEQCNPVPIQPIHGGYSLLESSAVSDFSSSSQSLRSQRF

>Pb_Beta7

MASGWDQCYSNCPPTMTVTIQPPPFRTLNLPGPALHCPEQPCIEQCNPVPTPHCDNSLFSSTASGFSSSSVASQKSVSGRQSSGCFPCR
RC

>Pb_Beta8

MSFGSNSCYASCPPVTVTVPQPPFFTLNIPGFSLYCFDQPLCIDQYNPCVPPPICPPRRSYTNSISSSSSLPQKSLERSQSEWSSVSQRF

>Pb_Beta9
MFCGWNPCSSSRMTAQPPFFSLNIPGISICCFEWFCCPEQYNPCAFVQPCGRMNYGRYGSYWSGRASDCNSLKFSAPKHLHQSQTKSLHY

>Pb_Beta10
MASRWDLCSFNCPPFTSVTIQPPFFFTVSIIEGFSLYCFDQPLCIEQCNEPCVPCYGGSSYGRREVILASISSQKTVSAYQKGSSEFFKRF

>Pb_Beta11
MSFGWNQCYENCTMTVTTLQPPFFFTMNIPGTSHCCFDOLLCVDSCSPCSFTVCAFMVCSFTVCSFTLCAPTHCGPTHCAFSQCAPVVCTSGLNRSRYSSCFSSCSLESRKCLERVSVCDECTKY

>Pb_Beta12
MASRWDLCSFNCPATTVTIQPPFFFTINIIEGFSLYCFDQPLCIEQCNEPCVPCYGGSNYGRFRSSAISAAISIYSQKSVSQOQSGCDPCQRF

>Pb_Beta13
MNAARVYSGWDRFNSCSMTVTVHPPLTSLIVSPAHSFNHSEPCIEPCNPCAFCQWDGRNRYGYHALTNQVAPCLTSTNSSQKCLFRSKSGCVPQRSY

>Pb_Beta14
MTSYKLCCTCVSCSSKLPYSTAKESCLCMFVSEINIPSHQELVFASETLSEAQNIRRFSRQKKSNTLCLETSQLLGHLGRGPTRRISYNLQKPPPLRSYRIKC

>Pb_Beta15
MAFWDYPSFNSGLCGVSVSCVQIFASEVIIQPPFFVALTIIEGFI LSASTEFAVATQNSPCALGALGACAPFGYGYSGGRRYHYGICARESYAQKSS

>Pb_Beta16
MSYCGFVCHVSCASSGGSCGLVFAYGSSSLSGAVSSYYGGGFI TCSSQLTGSEVILQPPASVVTIIEGFI LSATTEIVTVGQVTPCSYSHFLSFVGGSSYGGYGRLGYSGYGTWRRYSRRCLFNPCG

>Pb_Beta17
MAFCGPPCCLPPCAIPSCAIPSCASSPSVGFPGGLGGLASRSFGLPSQCPASSLGTLGVTSCINQIPAAEVVIQPPPVIVTIIEGFI LSASCDPVAVGGNTPCAAGGYQGLPTGLLGGGSRQGSRYRFVGNRGSICYIPC

>Pb_Beta18
MAFCGPPCCLPPCAIPSCASSPSVGFPGGLGGLASRSFGLPSQCPASSLGTLGVTSCINQIPAAAVVSSXXEVVIQPPSVIVTLIEGFI LSASCDPVTVGGNTPCAAGGYQGLPTGLLGGGSSGPRPGRRFTFVGRRGSIQYSPF

>Pb_Beta19
MHFCGPPSCAIPSCASSPVVGFSGAGLGYRGLGLGCGYGLGYGYGLSGYGLSGYGLSGYGLGALATSSGSLGTLAGVIPSSINQIPFAEVVIQPPASIVTIIEGFI LAASCEPVCVGGNTPCAAGGFGRYGGYLGGRWGRLGRRGSVCELPCLLPC

>Pb_Beta20_partial
MHFCGPPSCAIPSCASSPVVGLGSTGCGPSGLGYRGLGLGYGYGGWGESASNLGTLAGVNPSCISQIPFSEVVIQPPFVVVTVPFGFI LSASCEPVSVGGYTXXX

>Pb_Beta21
MHFCGPPSCAIPSCASSPVVGFSGAGLGLSGYGLD SGYGLSGYGLGYGYGSGALATSSGSLGTLAGVNPSCINQIPFAEVVIQPPASVVTIIEGFI LSASCEPVCVGGNTPCA VSDSGLRGSWGYGDWGYGGLGLRNRGLLGRFELSRGSIQYSRRGSIQCF

>Pb_Beta22_partial
XXXFGFI LSASCDPVSVGGYTAYATGGFGRYGGS LGRLGRFGRRGSVCELPCLLPCPLPCTLPC

>Pb_Beta23_partial
XXXVFGFI LSASCDPVSVGGYTACATGGFGRYGGS LGRLGRFGRRGSVCTLPCLLPCPLPCTLPC

>Pb_Beta24
MSCCVPSCTVPTCVPSGSPPICTYFVGGLSLNPCAISTAGGVSASSLGIVPGASVGCINQTPSELVIQPPFITVVIIEGFI VLSASCEPVRVGGFTACSGGSSNGGSSRVRQYPCNECKV

>Pb_Beta25
MHFCGPPSCAIPSCASSPVVGFSGAGLGYRGLGLGSGYGLSGYGSGALATSSGSLGTLAGVIPSCINQIPFAEVVIQPPASIVTIIEGFI LSASCEPVVYGGNTPCAAGGFGRYGGYLGGRWGRLGRRGSVCTLPCLLPC

>Pb_Beta26_partial
MAFCAPVCGLPTCAIPSCASAPQYGLSSGSGGLGLGLGGRGGLGGDSLGGSTTSGQLGTLAGVNPQINQIPSAEVVIRPFPVVVTIIEGFI LSXXX

>Pb_Beta27_partial
 XXXGRGGLGGSLGGSTTSGQLGTLAGVNPQINQIPSAEVVIQPPPVVVVTIPGFIILSASCPEVAIGGNTPCALSGSGPLGRPLALGTG
 PFGRCLLGGRGNCLFPFCGF

>Pb_Beta28
 MSYCGPACAIPLSLASNPIVGFSGAGLGGPGYGLLYSSGSSALAESSGGGLTLAGITPSCVNQIPPAEVVIQPPASIMTIPIGFIILSASC
 EPVAVGGNTPCAVSGSGSDLFGNCLPGNPLGLLRRGTLLGRRSLLGNHGDVCL

>Pb_Beta29_partial
 XXXCAISTAGGVSASSLGIVPGASVGCINQTPSEVVIQPSFPLITIPGELLSASCPEVVRVGGFTAGGGGSSNGGSSRGRCYPCICNP
 CRS

>Pb_Beta30
 MSCYVETCTITCTCVETCPPASSSICYPVGGVGSLSSSSSISSGGGTAASSLGIAPGATVSCINQTPASEIVIQAASVILTVPGFIILSA
 SCEPVVRVGGYTACAGSSGRSSRSMMSSGRFYICRS

>Pb_Beta31
 MSCCIPACLPPCPPPCPPTCAVPSCAAAPTIGLSCGSGSGILGYGGGGAASASSLGILPGASVGLSQTIPPSEVVIQPPPFVITIPGA
 IILSASCDEPVAVGGYSFPCASGSGGYLGGGLGGICRRKFSICKYPC

>Pb_Beta32
 MSCCRPCCPPTCAIPSCASRLIGLGGCGGSGALGGFGSGGGGAASASSLGMLAGVNPSSISQIPSEVVIQPPPVVVVTIPGFIILSASC
 DEPVAVGGYSACSSGSYGSYSRGLLGASSRGIAPRRYSICSSPC

>Pb_Beta33
 MYCGSPCPPTCAIPSCRETLGLGPGGLSGSGGIGGGSGIIGYGGSGLAASASSLGMIPGVVSCVSIQPPSEVVIQPPPYVLTIPG
 FIILSSSCPEPLAVGGYSFPCATGGYGASGGYGYSRYLGSAGGYLGSRCYGSKRRQSVCGYPC

>Pb_Beta34
 MACFIPTYTTFAAASNVLGSGCRIGPNISDCGGYGYGGSGGVITGGPSSILGLTSGANIARTSRLPPSEIVIQPPLCVLTVPPEVVS
 AITPELILGGSSSYIYGFPGYGYGGSVGSCLRGDIMGGFQNIQSRGRCNIIQYPS

>Pb_Beta35
 MSLCNYCCNPVYRVDTFCILQITLSEVAVQALFLITVPGFVMSASDELVAVRGNTFRAAADCYEYSGSGWVLEGGHGAFLAFIHT
 GFTKGYF

Suppl. Fig. S1. Amino acid sequences of proteins encoded by EDC genes of the python (*P. bivittatus*).
(A) Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X's indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Pb, *P. bivittatus*.

A

>Oh_CRNN

MSHLLRNIRSIIGIFEXYAKDDGCDTSLKGEIKQLIQTEFADVIVNPHDPKTIETVLHLLDTDCDGVGFEEFTVLVFKVVKACYKKV
QECQGTAGGFSKRGITSTQQQEI PPQOISRDGQSORDITSIQQQEI PPQOISRDEFSORDISRTOQQDI PPQOLSRDGF SKRDITSTQQ
QDI PPQOISRDF SQRDTSRTQORNI PPQOLSRDGF SKRDITRTOQQDI PPQVSRDEFSQRDTSSTQQQEI PPQOISXDGF SQRDTS
TQORDI PPQVSRDGF SQRDTSSTQQRETQTSQVSPRRCQKLVPPKKILDAEELFQEQRERTRLET SYDEKEQSQQEEASIRPRTQTS
RVL SVORTIEDMIRGNLTQMSLAEMRTEQAFAQGHPSALPERRRCLOEPLSPDQEASQDVLQSVFPQQSSQORQVEERSSTCQEEVTSQ
TSGQDRRSQQSAVEMEIQSSGESLRETRQRSSGGRNQQVQVEERSSAGHLRI PQVVQICDQESRSVTLRQHNDQQRCCSQGVEQSSVV
QEDQEVQA PGQVISRROEPQMFVYRPHQEKERYETTLECRGFSEQRAVSQEVTRQVAERYPRGEFGELTIVQFSQDSSTRRPFQFSQ
FQLKQRELFQFPFWSTKQ

>Oh_EDCATM

MFFHKKLGLWTNTVSSFTDSCLSSTPSCSGGSFCLGSRTPFVHKVFRMAGLKPPASVTSSEPLACEGLSSSI PKASGTFIVLMSCSFR
DAGFNPNWRNESAIWKERNIFTNELNLSVFLSLPASGGFVLTLCQPCDITFLCLTSFFLGHGFCGFLCVVFCATQBPVNLFP CGTQ
VQLAPNCANMVECTVKIAALGFCAMKGASL NTEFPPLASKCASLSCTMTCPYAAKDANSLPCTMEIILASNCASTVSCATKGAAEMQY
DAKGAALTSPCGSKGADIHPCDVEILPAPNYASTSCATKGATIIFYATKGATLTSFGGTKGANFHPGVEILPAPKCTSPVSCATKGM
TTIFYATKGATMSCFCATKCVSAITLFCACAFVFCATKGAISALTB SWEVLTLC TPN SCHNAFAFTVFPQVQK

>Oh_EDCG

MGMWDCCSCCGNSSGRSTIVLCMFSGGYQSCYFGCMFSCGCAPMCCMFIYCTNMRRNSCCGCGCGC

>Oh_EDCM

MSCCRCCGGRCGCCCNRYRFRVYYAAGRSATAGCCCRFRGCCGRFGCCCTSYYSYFYRVFVQTYGTQVKKC

>Oh_EDCP

MSYQCKQPCSCFTVCVKSAFACTFCSPFSSCICVSEFVKVVKCFSTCQKSC TDTKAF AASCTSPCSSFPCPVVQVSSCESKPACSCQTAT
GVVTTQSCDCTKFCODTKSCAFABEFCOTQCTCIEVOLSCCCGRIVLVEVSA CAAPSAPEAPCCTCTCCCKCCN

>Oh_EDCS1

MFCNCCNYCSFCGGSITNKIICAFGGSYHVSFSCCMCGGGGCCSGCGSCSSCCCFPPQMMCFSPMKCCCFFAMKCPFFVOLTQSCCLF
KKCC

>Oh_EDCS2

MYGGCGCGGGGGSGDTTIICIFCGNSCGSMGCCMPCCGGMRGCCSPCCYMPCSFGGYGCGCGGIMAYRI PSSRRGSSSSCCGGY

>Oh_EDCS3

MTCCGDCCDCGATCGNCGAVLLCVFCCGCQQCCSFCCGQSSSSCGGQSSCSFCGGTKSSCGQSSCGGQSSCSSCGGSKSSCXGSKSS
CGQSSCSGGSKSGCGQSSCSGGSKSSCGQSSCSGGSKSSCGQSSCSGGSKSSCGQSSCSGGSKSSCGQSSCGGQSSCGQSSCGGQSSC
GQSCCGGQSSCGQSSCGGCGQSSCSFPCSCALFCITQTCFISCPFCCKK

>Oh_EDEPK

MSTEQQQQKQTFVFP SGP SKLMAELVPEIASQKHGDESLPSAEQGDVVPEDLKESHQKKEWTSQSAIEIFVSEDTSGAKHGSLTRKE
KEHLTSTDQHCREFPTLSQVFNKETFKSSSEPFVREAEGHFATQFSVLGEQKQQQKKQTNOWLHK

>Oh_EDEPT

MAYQCKEFCLPPFIDVAQCSGDAQDSCTLTDMAQCVQTCDCVLEACSAFDSLSYQTKQPCVSPSFKVALCPKAISIFCTSPCLSLRTE
MCAGSPSVFCGSGVALQFSGPCFEDSQPCSTGTADACAPSPPTCGSRCSGHCSGFGDGGMSYVFPGSTVCLEPSGQVSVKPGVGSFEM
EGTTMFLASYPHF

>Oh_EDETM

MACLGLSSYNFAYSASLPSKSFLLRPFPFVAETRLTKLEILVFEHIPWDVPDVCSHFYSPLAFYHEPPIVVGFETSLHEV

>Oh_EDGPC

MGNMESNCGGGGCGCGSGCGGGDYGGSDAHVICMFTGGCGCGSCCAFPCCAPCCSPCCAPCCSPCCCSFCCCSFGCGGSSQCCSQ
CCKQCK

>Oh_EDGY1

MNFYTQPESIDSSYLEFYLSCYAFQYPOYGYGTTGTSYCDP GNGCNYGYQPYSGGGLNTCSTSCSYGYRYTDRGSCYE PWNRYRMGYG
ARRNFPRRRYNCFNF

>Oh_EDGY2

MSCYPNPSCYDFCRAKSFSSGSSGLKRRTRWSSGS CYGSGYGRGVRIYFCGSGYGGYQSPAHGGICYQPIRYDCGYGYGNKYCPFVC
YGGYSGGGYDYGCKYGGLYGRRQSI CYDECYGERIYQYSGRRGSFCWFC

>Oh_EDHEM

MECTTFNKSTEEITAMHFVSIQELDEISLGSHFHNHHDGIFLEDENVMECIHFNDAHFHGEFYRRIHIVEMIELFSFDAKESDCFL
HKHVHPSGSEHSQEDPEKATS SSSSEPKSKEKLETANDGGGENIASNGDANNLS SKDPEDAVGLMYTCKLEYQWSMLKYCTMEELGYK
FLHLL

>Oh_EDKM

MSRLLRAFTEMIEGNSKVAERKSTDAELLKKSEFKTLIQKELTFSKTRKHKNAILLHESDGEELMNEKELMGQWR

>Oh_EDP3

MSQQQCKQTFCVFFVSHRCTSSCKPCEESTCYASSGFHCPGSTGSHCFGSTGSQCFGMTGIFCSGSTGGSCQNFCTCRSGNNGCFQRHFQQQC

>Oh_EDPAM

MTHQCKLLEVELGGLKIGTKSSQTKSSSAFSTACCCGIFVSAFPFSPFPQTVTVCALPAPAPATQSMTCGGIFVOSTSSFPFQTVSICAVPAPAPAQSSITCCGIFLQTKSATFPFQTLTICVPAQAVASQCQSVSCCGIFVQAAPAPAQTLTICAVPAPAPAPAFSCVVTLHLVQPASAPTSSCMCCSGCKT

>Oh_EDPCCC1

MTKCSKCGCSFCCCKCGCSQCKTCCCPKCGCCFCCCCCCPKCGCCFCCCCSPCCCQQCNCAFPCCCKQCVCAPCCCVGCGSPCCCQSSCSQSCCGQSKGCESKGCCSKCCFFKQGKS

>Oh_EDPCCC2

MTCCSKCGCSFCCCQCCCCSPCCCQOTCCQSCCEKCGCGCGCGCKGGCGCGCXGCGCGCKGGCGCGCGCGCGCGCKGSCGCGCGCGKGGCGCGCKGGCGCGCGCKGGCGCGCGCKKQGGK

>Oh_EDPCCC3_partial

MSSGTKSKPKCAKGGSPCCCSKAASKKKCAKCGSSPCCCSKAASSLKSKCSKCGCSFCCCSKTLQSKCGCWFPCCCFQSPCAKCGCXXXSRCRCGCSFCCCSQSRCSRCCGCSFCCCSQSRCSRGGCCPCCCFQSRCCRCGCSFCCCSQSRRCSSKCGCFPCRCSFPCRCGCCPCCFPQACCPRCGFPCQQIQPYCYFCHR

>Oh_EDPCCC4_partial

MACCPICGCFCCCVPCCSCANCCGCGPCCSSCGAFCCCSFPCSSCCPCCACGGCGSFPCSCPQCSSCCVCGFPCCSRQSSGGGCGCGCSFCGCGSCGGS CGGSCGXXXC GGS CGGSCGSGCSGCFNQCFS

>Oh_EDPCS

MNYEYLFPEFLEEAFYFFEANEDIVMPTKGGTYCFRFPFPYKCVTPRFRYCRPKYFCPCPPPFPFPFPFPFPCLFPFPFIRPQCFSPCYFLYFLPQWPKCYFSYPCRYFPQWICFKPLFCVFRFPYFSCQLFCFPYWGRCFYFKCLPQYPCDLLEIY

>Oh_EDPKC

MASSHNOHCKQTLTLPFEICKTALCFPEEQCKQFPVVEVPKPKFCPPQEPCKEFPVVVIFTPEFEPKPCPPKEFPCKEFPVVVITPCEFPKPCPPKEFPCKEFPAPITCPPEEEQCCQKQSQCKKPPVITPCEFPKKQEPFPCKEFPPIVVIPTPTPCEFPKPCPPKEFPCKEFPAPTFCRPKEFPCLPKEPTPDQCKTQCKKPPVVVIPSPEFPFCPPQERPCKEFPVVVTPTPCEFPKPCPPQEPFPCKQPPPCDQQQKKQFCSWFPQNK

>Oh_EDPQ1

MYTTENGDEGDSYSNYMRVSCEFPFSYSVRWSAEMEYCDTPPPFVRSYCPFRFIYCFPPAPKGCFTKQKYCFFVQRYCFFVQKYRPFVKYCCFPGRKSPFPFAQWCSFPFVQKCHFPVQSYCFPVQKYCRFVQRYCFPQKYHFPFAQFCYFPQKYCFPAQSYRFPFAQFCYFPQKYCPFAQSYRFPFAQCYFPQKYCPFAQSYCPFAQFCYFPQKYCPFAQSYCPFAQFCYFPQKYCPFAQSYCPFAQFCYFPQKYCPFAQCCSFQFKICQIKEVCKAPPHLLK

>Oh_EDPQ2

MSYQQQQCKQFCQPPPCQKTCFTFVCTPQVCIFPKPCPPQDCIFDFCPPKVCIFRRCSPQECLPKFCFTQECLPRFCSPQECLPKFCSEQECLPKFCFPQVCLFEFCFQDQQSQRFSQRKRC

>Oh_EDPQ3_partial

MSYQSQCKQFCFI PPQCKIIPPFQCCPPFKCCFPFQXXX

>Oh_EDPS1

MFHCCMPSCMQFIAPLQVQFLQCFSSISIPCSPLFPMPFSVIFVLSPPFSCIFVCRROPSSSSPTISIFLSSMQPPPSISLAFSSSSGSSPTISIFLSALIFSPSCISLCPQSSSGPSISIFLAFSSSSRPPCFLLCPGQSPCISLVPSPSSGSPPTISIFLSALQLRSSISLAFSPSSRSPPTISIFLSALIFSPSRPCCIFGSPSSGPTISIFMNFSPRNSHICCFPIPMSSPFCFIC

>Oh_EDPS2

MFGYCLFSYYMGFIAPLDVQFYQOTSSMFAIGIFCRSVQSSMPPFCIFIRFSFGFSTSSGFGSFFSCIPICMQSPFSFLPPYSFSFSRQHTYCIFFDNFSPSSSQTSIFLFSFCPPFSSSQTPYCIIFLCPSAQSSSGQTTSAQPSYCIFFVSTCQPSFSMSFCFAPQSSSRREMSSISLSLSLSSGSQSYCIFFCSYFSPSSSNQFSYVFCPSQTFSDANVSIFVSFSPSPSPNPFFTCRIFCSFCFLVF

>Oh_EDPS3_partial

XXXIFLSSLI PPF SKPSPFCYIFLCPSPSSNSSSGISIFLSSSSSSNPPCYIFICPGQSPFPSSSGSPPTISIFLSSLI PPFSSSSSFCYIFIVSFCQSSSGPTISIFMNLSPSSSNPPSCMIFIFMSSSFCFIY

>Oh_EDQK

MSSKDQQQKKFCGTFSTFQEPQKQCLPKNCCVVNFPETRSECCLSFPQFCTFPQKFGCLSCHFCFQNPFSYQQ

>Oh_EDQL

MC SRNNEGCSHDNTCHSGRSCCHEDRGYCGDRSNE SFTVTALFVQDKNPLTQTQKQCPIVYSTFQLQHTQQLKQFIHYFPQQQLKQFQ

QLKK

>Oh_EDQSG

MQQSKQPGFPMGQTTNCKGSTGGQTQGFVAVQQSQCCDSAFSQCCSSGVTHVHSSGSVSQMGGCVSQMGSSGVTQIQSSGLGSSQSS
DLDDCQGSIGISQCCGSGGNQGGKGSMDCCGCSFVSQKGSSSGGQCKGSPVDCCGCSFVSQKGSSSGGQCKGSPVDCCGSSDVQCKGSS
GGQVKGCFVDQCKGSKGKY

>Oh_EDSC1

MSQQQQQGSCHGSSSGLGGGCCRRGSSGGCCSRGTTGGGGCCGRRGSSGGCCSRGTGFVYVQQSQQTARCCCNINRSGFCCGRCCGRR
CGGGCCGGGSTQLKQK

>Oh_EDSC2

MSQQQQRGGCCCCRCCGGGGSSQVSRSSGSQGRSGGGSSQQAQSSGGGCCGRSGGGSSQQTSSSGSCCGGSSQCKMK

>Oh_EDSCP

MVLSCHRFTCYSSGPIFFVGGSSCCVEVCHAFGPPPTVCVQFVQSSCVFVCASFPGMSGPPPTVCVQFLOSFCVFFVCASFPGISG
PPTVCVQFLOSFCIFVCGSPCGPFAMCIQPKSSFFSCVSPSGPIFFVSSSSSIIVAPLCVQPSKLSQVSPSGPIFFVSGGHSSCMTC
IPVSVHPCSPCHLWF

>Oh_EDSPR1

MSYFYQQCKQFCLPPPVCCKGCATQCKVCFACFCDFPICATFGVKVVGCACTVCFPPPCAPKAGFAVDICVFKQGVCFPCDCKCFDFC
CK

>Oh_EDSPR2

MSQQKQCKKIPFCEACCPFKGFKQKTCSSHENKDVLFQKGFCCPFKDKVKCPFKAKQCKPKHC

>Oh_EDSQ

MSYQCKQFCLPPFCLKGATVCGFGTTVCISFTQSGGIEVCASFPGAICVTQGGSSCCAKISHCCGFVVVTVQTSFSPCATIAVS
NQGNSTICVPSQGCQCVTCRGSVATIOSSQGCATVCASFVCISFVAAPLAAAVVCFECGNVCVASKGTAKCATKCAAFGTTVQV
DFCECVKSVKNQDDAVCAKGCQSVNQCNVKKC

>Oh_EDSRWM

MIFYTFQQYLDLDSWDINYGNDDNSSEYFYADDETSFVLDPKSRWDSPPYSOSSDLGPRVRASSGSRSSCHPSCKCHCWSPSCDFE
SCPPWSSGCSNFCYSPLCKRFRGRSRAPFSMGNWGPSSSSSCPPWRCGFCNFCAPCCRWGPPSMRRRGSQDSTSSCRYFPPLS
RCGFCNFCAPCCHWGPPSMRRRGSQDSTSSCRYFPPLSRGFCNFCAPCCRWGPPSMRRRGSQDRSSCYFPRSTCGSKPCYSFICKER
CHSLRCPPFSMDSREPQSSSSSCPPWRCVFCNFCSECCRWGPPSMRRRGSQDSSSCRYFPPLSRGFCNFCSESCCCWNFSRIRWE
SQDRSSCYPPGSTGSCNFCYSFICKPRCHSLRCPPPSMDSSCCPPFCWSPPFMNRWGPPVCRSCFTIYYYPCYYC

>Oh_EDWM

MEERIYSSGREAYFNLNSTWYDEAGSWLDTRRKFERYVDNTACVTCCNFRSNVRRGGHNYRCYCYROCTCTPGGNERVTCVHNHSG
GPRDYWGRFIGDACDGGCTGGHYSHAGSDGSSCCGSLGGCGTGGRTMACAOPCATSGGVCAEPGCRFAGRGVCAEFCITSSGGCSSGRGV
CAEPGCRFAGRGVCAEFCITSSGGCRTGVCAEFTCTESGGRRRRGVCFEPCSGTSNGC

>Oh_EDY1

MIGITILNNDSGDTKAIYCPFTTYGFSASNFSICFPYSSLAGTAARCYIFRRYRHLFGLSFNICYLSDLSMACGYSSSLSDWRSACQFC
NYGHLFKSNQGYDDECTFGFLSRGVQRSSDEYGLELYLSYRRRYCDEFCGYSSSYRRRYCYDECVYNYPNYSNGRSSYSRGTYRCGY
NDFCGYSACATWSQPRYAERRRYSCQPVSCDYGFSESPYSSGSWRRSSDFCNDGCPY

>Oh_EDYM2

MYYAYHCKEIGFPLPMFVKRLEFKYGSQYILYGSKLISSRNILYRRGIWVTYSLAHVAGYQRTIEVEFDAQLNKFSKLRTRTNL
VRSDVEDYFYQIQSTVGYDGTTKGFHGFHKGSLFCVVEPERVFRMTKGFVLVSSRDLHQGLTQVFRVSRRTTFYASEDELLCSSKEELS
QRTKRLTTQYFSKVFQGNRVRSDLEGGIDDELQBEAKISSQPTATKGSRVITTKVFRANFLTLSHSFIKRGARSFLTKTQAHFYRGSR
ASLVKPSRSTSLVKGSQPSLPHTAIFPAPKNLSKTVKISSGKXYSSAVNWPF

>Oh_LOR1

MASCCQSKSCYSSGGRCVCIQSGGGSLCCCKGRGRGSVRCVSIQQTQTSVCCGGGRRGSSGQTIIVLPSSGGGDGCCCCRGRRGSSG
GRSGQIIISPGGSGGGQSSCCIGGGICGVFGIGMNQKCGGGIISGGWRQSGSAGGSCGVKVISGGSSGGRRGSGSRSSGSPIG
RSSGGGQTIIVFCGCKGSGSRGSGFIVGSCGDQTIIVFCERRGSGSGSRGSGFIKGVCSGGQTIIVFCERRGSGSGSRGSGFIKGVSA
GSQTIIVFCERRGSGSGSRGSGFIIGGFGGGSIDGGKTIIVFCDDGSGGGGSSSVQVSGGGITVGVGAASMQTKQSIIILPPCIGQTKQA
FQCPPLQKL

>Oh_LOR2

MSQQIQSSGCCSTGGSSGHCHRRRGSGGTGCCACCCCRGSRHVMVVSQSQSSCCGGIQQISYGCCGSPTEGVVVVFGGSSGCC
IGGGLGGGIQCKVFLIGDSSGGICCGGGSGGLGGQTTIVAGGSGSGCCGGSGVVKVIIGGSGSGSGWCGTGGSGGGQTTIVITGGS
GGCCGGSGGVKVIIGGSGSGSGWCGTSGSGSGQTTVVTSGGGCGGSGVKVIIGGSGSVQVDGSGSGVKVIIGGSSGGGSGFDVKV
IGGSGSGVCGTGGSGGAKAVVVTGSSSEASACCSTGSSSGTVAVVGTGGSGSGGKTIVVSGGGGSDCCSGGSGGASVVIAGGGGSS
QTKCFIVVFCIGQTKQVCFLEPHIK

>Oh_PGLYRP3

MVILQISFLILSLSQDLASSISFATGCFRLITFSKWEAKAANSQPLKDVFAEYVVIHTAGNFCRTHRDCHNEVRMIQNYHMNLKGW
CDIAYSFLIGEDGYVYEGRGWRNEGSHTYGYNDLSLGIAFIGTFVERSFPEDKAWKALRCFLDFSVKIGYLSPEYIMLAHSDVSDIVSG

EFIRAEIAKWNKHSPLYILNRGGQ

>Oh_SCFN1

MSYFLDSVCTIVGIFHKYARCDGNLALNRREMKALIQKEFAEVLNPCDPQTIELTFKLLDVNGDSLVDNFNEYLIFVFQIAKGCYRYL
QREYLLRDESSRALHEGEAGGSKRGDHQLDGERRGDYVHERGGLDGTFLHSTEEGSRGELVGRYLIEEVEEEDSNFEGHEPKLRDG
DRRDHESQEHQEREDEQQWREPKEWEDIETEPQOWRQOESTLEEDSBRQSSRELVRNRNDNRREARGLLDRDEEDLYSIVESSREEMR
QHRQGETEGLRFRYGSQARVDHTDYDVERSWSHWEQEVEDGRRELEMHNERSSHAFEHVADDQRDRIHRSVEEYEEHHPLSTEGE
RRSQFHEDDHRESERRRQGRSVFSRKDVRRAFRAEEEEVRRSERRHDEDEWNRQPREASRRSEVQDLERRERETRREYSSSELKNMDS
SSSHESESREEVRRSQHFEDIVORRREFSREDTRVTEVRRRRSSQSLFVARDIRYVRSQRYEEFRNDQRRFHNYEGRSLERDVARRR
RLQLRESFKREDNQERQWYYEHETWDSERDGRRCOLRYSEPREREMRRYQVVDSRVDEREQSSREAVTRPTQWDRFSSSDSPEGYDQSGQ
QIYERHTLQREGRGQRYDRLSNGQRQIWAIDIAQRESEGQRSSNVEVRIVDQFRGVHDHRESEHQRTSIHSDDDENTDQROTQTYEAS
SREDEDQRRSSHDSGTROIQRRTETYEVGKDDQRRRRMSRESVDQRWRHTQTDLAEDGOKSNFYETDPRDGEQORTVQSSDEN
FRDTEATQHNMGDSTQPREATSLSNQATADREFDLRRDQRPESGEGSQSRKEGQSKAEQSEVSYEQHQPSTRESQRLSGBELRGQRRW
LFHHQTEAPEEAGVDQVQRKVGSPVSKRAISQFSERQITEGQESRIQGYHRTVVSQAEGESELSLDGEGDKAQEPAALNQQQVHREGEV
QFVAEEFKVEEDDGGSSWAQESQPFLEDDDDQAITEKSSSLSVESKSCVICNLYEYLLAQKKQEQE

>Oh_SCFN2

MAGLVDSICTIIAVFHKYADRKSESSSMKRRQMKRLIQKEFGDVLENPRDQIVKLTFLQLLDVNGDNRVDFNEFLFLIFEMATACYSVW
HIRECFSYNEERRRAVEDEEPRGDESNRREFLGEDRRRDDVRERRGADRTPLSMEEGRRGELSSSELREEVDRRFEGHDRELRDGDRRD
RESRELQEREPEQRLEPKWEDVDIPEPRQRQRESAVEEDSBRQSSRELVRNRNDNRREARGLLDRDEEDLYSIVTISRREDERQROG
RETEHQEAEGLRFRHGSQARVDTRNHRPEYDVERGQPSRERREGEDGRRELEMHNERSSRAFEPRVADDRDRRIHRSVEEYEEHHPL
SREGRRSSQYEDDHRESESRQGRSVFSREDVRRRAIRAEIEEEVRRSERRHVEDEWSRPREASRRSEVQDLERRERETRREYSSSEL
ENMDRPSRSHSEGREAVRRSQRFEDVRRRQELSWEESTRATESRGRPEQSEPRDPKYDRSSQSYEPRRDDQRRFRNYEVRSSQERD
AERRRRLQPRELVEREDNQERQHYERESRDRERDGRRAMSLEAETREDEVRRRQESASRYGERERLEEGDARMRRTGRDMASPSDAGRR
ELDESSPQOISRRRAPELRRESEHSMFPAATKG

B

>Oh_Beta1

MSGSGVKCVTTFCVTSCEPAKVVVHPPFLVLTLGLSLKTSFNQCLVESQTSCLTNGCEVNCDEGSKAIVTKTSGLVALSGGDTFCCTT
TCFDSQOVVIPPFPVCITIPGAVLTSYPNECLISSSSTNCITSGVQRALERSTSVSDCSLTPSRLTRSASVPSGLNSSTRCLTQGSNKV
VIYPPPIEITIPGVLEIAAEECTVEVYNSCTDNNAITGSDQAATSGDESEVTLIPRAKSCTTVSGLMDTSTCISQGPPEMKIIIPF
PIEVELPGFIFLEVFEACKVETLTCEPFKPEAITGNEIKALCDSKIPSTALATTTSTKRELPDVRPPPRFPAEMYSSSMTFRSLALSQ
QSRLLAKYRSALYSMSHSQSSY

>Oh_Beta2

MDRSECYASCPFSTTVIRPPTFTLTIPGEAIFCFNQLQIAQHNPCAHGGMGVEERMMLSDFINEDLENLELELSSRYGSALATLYRS
RMAFPSESWL

>Oh_Beta3

MSECYASCEPASTVTIQPPPFVLNIPGPALFCFQAFGIEQHNPCAGFGMGARGSGVGIEGTVGGSMGALPGRGGWGGSSWGGSGIG
SRMGGGLGVAGVGGVGGGLGGGLGIGGPTGLGIGGLGGERSGGLGIGGLGGERSGGLGIGGLGGERSGGLGIGGVGFRGSLGLGGER
GTTAGWSSGSRLLGGXSVGMGSMGTGGDLGSGEGGDMSGRRGSIAGGSTVGVWSSGSRLLGGTVGGRRGSGVMGSMGTGGDLGSGEGGVMSG
RRGSGIAGGSTVGVWSGVGDVTVRRGSGVTGSTITEGDLTGTVGVDGIHGGSSGSTRITGGRRGGSGTGVSVGGRRGSGITGGTIGG
GVGSSMSGRRGSGITVSVTGGDYGLSGTRGSTVGGWTSGRRMSSGSHRYGGSIGYGGGGYGGGYSSGGYSSSGYGDGYGHTGHGSGGYG
GGYGHTGHGSGGYRRIGYSVGGGDGYRSGGFSSRSYGGSGGNLAVMPGFSIDIYSY

>Oh_Beta4

MSDHCFSFSCPFLKVTIQPPPFIIITVPGPSICCFDQPLCIEQHNPCAINHNSPPQAIRASSFYSRALPSTPYSPSIYRF

>Oh_Beta5

MWNECYAQCFPASTVTIQPPPFVLTIIPGPALYCFDQPLGIEQHNPCALPMRGGSNLTSFYSRALPSTNFSSTYRSSY

>Oh_Beta6

MASGWNICYTSCPFMMVTIQPFVYTLNIPGPSLYCADQSVCEQCNPCAAIPSSGGINHGGYGLLPANVASDFSSSSKSLENAYYF

>Oh_Beta7

MTSGWNYCYTSCPSMIVTIQPFVFAMNLTGPALHCPQQSLCIEQHNPCVPTPHCKDSLCSSTASNFSSTSSVASQKSISSRSQSGCFPCR
WR

>Oh_Beta8

MSCGRMPCYSSCPFVTVTIQPFVFTLSIPGPSLYCFDQPLCIDQCNPCVPPPIPCPFITRRSYASSVVSSPSLPKSLPSSQSEWSSVS
QRF

>Oh_Beta9

MASRWDLCSFSCPFSTVTIQPPPFVTVTVGGSLSLFCFDQPLCIDQYNPCVPCQLSYNRRRELILASTSSQKTVSKF

>Oh_Beta10

MSYGWNNCYFNCTFVTVTIQPPPFMSMNISSSICCFDQYACIDPCRECSFTVCSTMVCSFTVCGFSMCAFKACGFAACGFAHCGFVSCA

FVSCGPFVSCAFVSCGPFVSCGPFVSCAFVSCAFVVGCPAFSSRRYNSYYNSCSLPSRKCLPRVQTICDFCTKY

>Oh_Beta11

MASKWDLCSFNCPPFTTVTIQPPFFCLSI PGPSLHCPDQPLCIEQCNCVPLPYRGSLPLPSSAASIYSQKSQQQQQQQYQORF

>Oh_Beta12

MHSGWDHCYNCTFMTMMVQPPFLTLSIMSPAHSFGHPACIKETHNFCAPVHCWDRSHYGHGGLTGAAAPCFTSTSQKCLPKSKSGYV
FCRTY

>Oh_Beta13

MTSFNFCTCASCSSTRPCFHAKFSCLCMYFSQINMPLHQELVSTREMLSTTQRIQGFLKQKTREYGCLSTSRPLKYLRGPTSGISYSL
QRNFPLOSYRYIKC

>Oh_Beta14

MCRENKLVMAIHFYFLDGRNPPFAVAVTI PGFILSACTEPIAITQHSFCAPSGGECAPFGYGNRSGRRYHNRI GRSGGTQKSSSGNGG
PYLKNCKLEKIIRKSCKVFRNSENS

>Oh_Beta15

MSYCGSICHVYASNSGSCALVSSYSGTTLTGLAANCYTGGHVNCSQLTGSELIIOPPASVVSI PGFIISSTHDVSVGQVTPCSYTH
PLNSYGAYGRLGYWGYGNYGNYGGWRRYSRKCLTYN

>Oh_Beta16

MALCGPSCAIFSCASAPSVGFGSAGLGGLAPGSLGFS PFPLSESSGSLGTLAGIVESCINQIPPAEVVLQPPSCVVTIPGPILSASCEP
VAVGGNTPCAVAGGFGQLPLSLGGPSRLGARYGFVGNRGSICYTFC

>Oh_Beta17

MAFCGPSCTVFSCASAPSVGFGSAGLGGLAPGSLGLSPFPLSESSGSLGTLGIVESCINQIPPAEVVLQPPPCVVTIPGPILSASCEP
VAVGGNTPCALAGGFGQLPTGLLGAGLLGSRLGRRYNFVGRKRSICYSPC

>Oh_Beta18_partial

MAYCGPACAVFSCASAPVVGFGSAGGKGLGWGLGYGGLGYGLGYGRGLGXXXGFI LSASCEPVAVGGNTPCAPGGIGGIGGFHYGGFY
GGRLGRLGRRGSICNLPC

>Oh_Beta19

MAYCGPACAVFSCASAPVVGFGSGGSKGLGWGHGYGLGYGLGYGRGLGYGAGALAETSGSLGTLAGVIPSCINQIPASEVTIQ
PFSSVVTIPGFILSASCEPVAVGGNTPCAPGGLRGGWGYGGWYGGWGHGGLGFRGYGGLLKRYWNRRGSI CLSRRGSVCL

>Oh_Beta20

MAYCGPSCAVFSCASAPVIGLSSGCGPFGYGYRGLGLGYGAGALAETSGSLGTLAGVIPSCINQIPASEVTVQPFSSVVTIPGFILSA
SCEPVAVGGHTPCAAGGYGRYGGYYGRLGRFGRRSV CALPCNFC

>Oh_Beta21

MACCPPSCAVFSCASTFVVGGLGSTGCGPCGYGSWGYGGGLGLGYGYGSARFGESARNLGTLAGVVFSCISQIPASEVTIQPAPVVLTI
PGFILSASCDPVAVGGYTCAFGGFGGYGGYYGRLGRFGRRSICSVGRRGSICTLPC

>Oh_Beta22_partial

MAYCGPACAVFSCASAXXXRLGRLGRRGSICALGSRGSI CNLPC

>Oh_Beta23

MAYCGPACAVFSCASSPVVGFGSAGARGLGWGLGYGGLGYAGLGYGAGALAETSGSLGTLAGVVFQFINQIPASEVTIQPPSFVITVVG
FILSASCEPVAVGGNTPCAPGGIGRLGASYLGRLGRLGRRGSI ICNFCNLPC

>Oh_Beta24

MSCCAPACAVFTCI PACSSPVGYVGGGLSLTSCGVGSYGMSSAGGSSSAASLALAPGASVSCVNQIPPSSEIMVQPTPIAVIIPGAIL
AATCEPVRVGGYTACASGSSSGGSSKLRYVPCNFCGFCCK

>Oh_Beta25

MAYCGPSCAVFSCASAPAVGFGSAGARGLGWGLGYGGLGYGLGFLGFRGLGYGAGALAETSGSLGTLAGVIPQFINQIPASEVTI
QPFSSFVITVPGFILSASCEPVAVGGYTCAFGGIGRLGASYLGRLGRLGRRGSI ICNFCNLPC

>Oh_Beta26

MALCPPTYVIFSCASTPQFGLGSAGASAGLGLGGRGLGGGLQGGSIGLQGGSMGGLGGGLGGGVSSGELGTLGIVPQFINQIPPAE
IVIQPPSFVITVPGFILSASSDPVAIGGNTPCAAPGSRILGRSAPSRGLLGGRNLOEGRGNITIVFCGY

>Oh_Beta27

MAFCPPSCITFSCASAPQFGLGSAGASAGLGLGGRGLGGGLQGGSMGGLGGGLGGGVSSGELGTLGSTITPQAINQIPPAEIVIQPPSF
IVTIPGPILSASCDPVAI GGNTCAAPGSGILGRSALSRLGGLGGRNLOGGRGNVTLI PCGY

>Oh_Beta28_partial

MAFCPPSCIIIFSCCYTSQGLGSAGVSGGASAGGLXXX

>Oh_Beta29

MAYCGPACAVESLASSFTVGFSSAGAGGLGYGILYNFAASALAESSGSLGTLAGINSCINQIPFAEVLVLPSSVLVTIEGFIILSASC
ETAVGGNTCAISGSGIVGSDLYGNLLSGNLGLGLRRGTLLGRKSLLSRGNICL

>Oh_Beta30

MSCCAPACAVETIIPACSSVVCYFVGGLGSLTSCGMGSYGISSAGGSMASATLALAFGASVSCVNQTPPSSELVIQPPFVAVVIEGAILA
STSEFVRVGGYTACASGSSSGSSKLRYYPCTPCNFCRN

>Oh_Beta31

MSCCAPACAVETIIPACSSVVCYFVGGLGSLTSCGMGSYGISSAGGSMASALALAFGASVSCVNQTPPSSELVIQPPFVAVVIEGAILA
STSEFVRVGGYTACASGSSSGSSKLRYFPCNFCGFC

>Oh_Beta32

MSCYVETCSVETVFTCAPPFSSFIICYFVGGLGSLSSGGGMMSSGGSMMSSGGGGATLAASLGMAFGASVSCINQIPSSSEVVIQPAFLM
LTIEGAILSASCEFVRVGGYTACATGFSGSSGRASRSMYSCTFYICRS

>Oh_Beta33

MSCCPACLPCCPFSCAIPSCASVFRIGLGSCTGIGSGLLGYGGGAASASSLGILFGANVGCINQIPSSSEVVVQPFYFTVTIEGFIILSA
SCFVAVGGYSACAAGYGGYLGGSSGLGAICTRRRRASICKYEC

>Oh_Beta34

MSCCPPCVPCCPFSCVIFSCAARETVGLGGCTFSALGGFGGIGGYGGFGGFGGYGFSGSGFASASSLTLAGVTFSPISQIPSSSEV
VVQPPPVVLTIEGFIILASCEFVAVGGYSACAAGSFGGSYSRGLLGAGSRGICGERRKFSICGSC

>Oh_Beta35

MQSQCAPSCAIPSCREVOALGSGIGGGFGSGGFSGGGYGMGYGGGSGLATASSLGLIEGVVSCISQIPSSSEVVIQPPPFSLTIE
GFVLASSCEFVAVGGYSPCSSGGYGVSGVYGSSRYGLGSSCYLGGSSGYLGGRSYGSKRROSICGYPC

>Oh_Beta36

MACFNPSTYTMFCATSNVGLGSCRTGFTVSGWGGYGYSFGGSIIGGFSGMLGLTSGANFSRTSOLPFSSEIVIQPFQVLTIEPEPVVTHI
TPFMVLSSSSPCSYGSLGYGYSGGYGSGSVGGCFRGDIMRGFQSIGPPRGRONIYRCPC

Suppl. Fig. S2. Amino acid sequences of proteins encoded by EDC genes of the cobra (*O. hannah*). (A) Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X's indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Oh, *O. hannah*.

A***EDPCCC4***

P. bivittatus gen TATAAAA GGTCCCCCATTTCCCCAAACCTCCATTTCAGCTGGATTAACTCTCCTGGCCAC
P. regius RNA-seq -----CTCTCCTGGACAC

P. bivittatus gen TTCCACTGCTTACAAAGGTAAAG//TTCAGCTAGACCTTCGCTACTCATTGCAAAGATGAC
P. regius RNA-seq TTCCACTGCTTACAAAG-----CTAGACCTTCGCAACTCATTGCAAAGATGAC

P. bivittatus gen CTGCTGCCCCATCTGTGGCTCTTCACCCTGCGGTTGTGCCCTTGCTACTCTTGTTCCCC
P. regius RNA-seq CTGCTGCCCCATCTGTGGCTCTTCACCCTGCGGTTGTGC-----

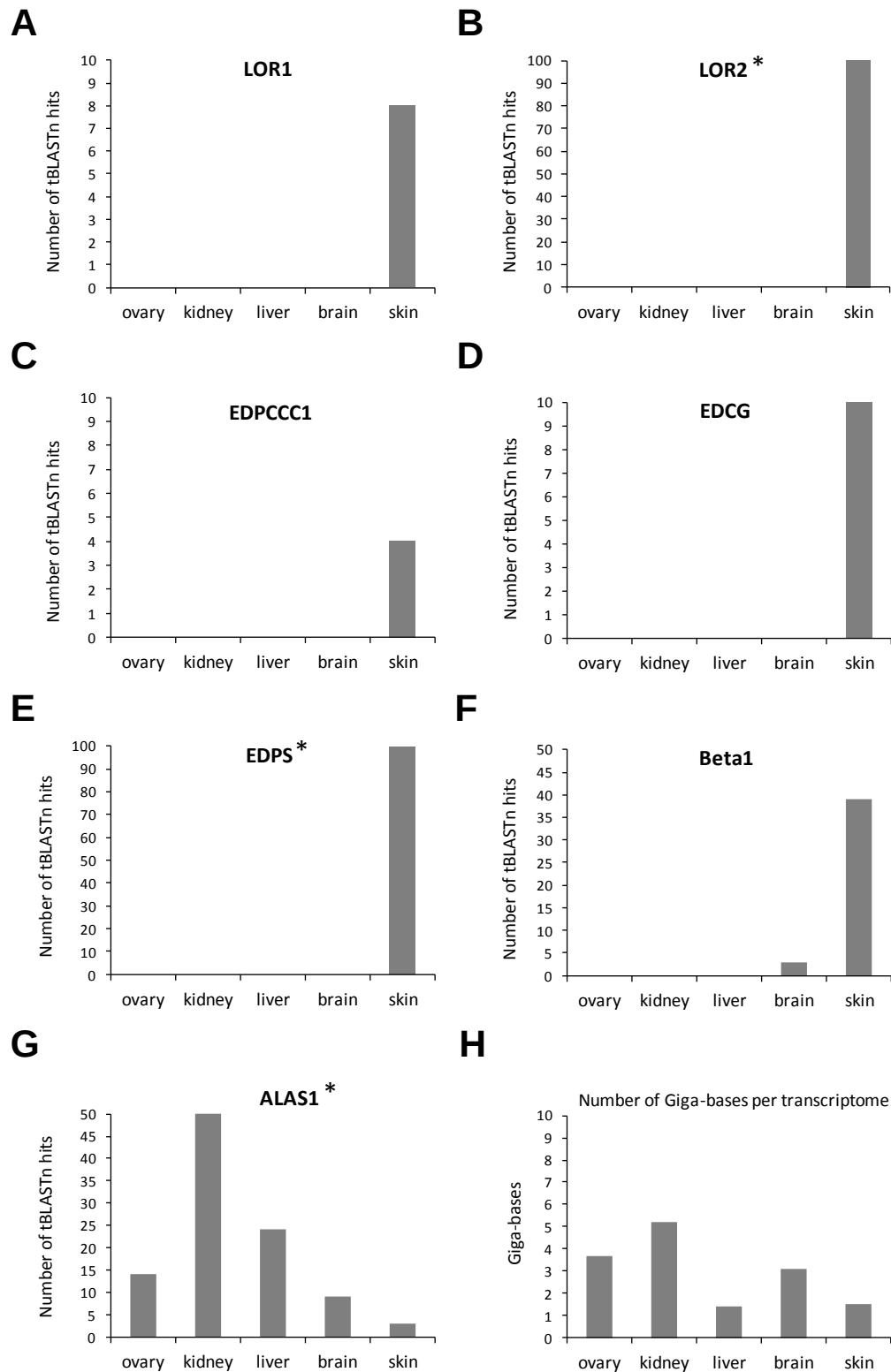
B***EDSRWM***

P. bivittatus gen TATAAAA GTCAGTTGAGTCCTGAGAAGTTCTAGAGATCCCTCCATACTGTCTATTGTCT
P. regius RNA-seq -----CTATTGTCT

P. bivittatus gen GTGGTCCAGGTAAAG//TTCAGACCCACCTTCATCCTCTCAAGATGAACCTCTACATGCCT
P. regius RNA-seq GTGGTCCAG-----ATCCACCTTCATCCTCTCAAGATGAACATCTACATACCT

P. bivittatus gen CAAGAATACTGGGACCTAAACAGCTGGGAGACCAACTATGAAAATAACTATGATTTCCCC
P. regius RNA-seq CAAGAATACTGGGACCTAAACGGCTGGGAGACCGACTATGAAA-----

Suppl. Fig. S3. Alignment of RNA sequence reads versus genome sequences confirms the expression and the presence of an intron in the 5'-non-coding region of *EDPCCC4* and *EDSRWM* genes of pythons. (A) The nucleotide sequence of *EDPCCC4* in the *P. bivittatus* genome (gen) was aligned to the sequence of a *P. regius* RNA-seq read (GenBank sequence read archive, SRA, accession number ERR216300.5250820.1). **(B)** The nucleotide sequence of *EDSRWM* in the *P. bivittatus* genome (gen) was aligned to the sequence of a *P. regius* RNA-seq read (ERR216300.7157883.2). The complete sequences of the RNA-seq reads (100 nucleotides) were aligned to python gene segments corresponding to exon 1 and the start of exon 2 including the flanking sequences. TATA boxes are highlighted by green shading, splicing signals (GT and AG) at the ends of the intron are highlighted by blue shading, and start codons are highlighted by yellow shading. Hyphens were introduced to maximize the alignment. Red letters indicate identical nucleotides in both sequences.



Suppl. Fig. S4. Identification of EDC gene transcripts in tissues of the painted saw-scaled viper (*Echis coloratus*). (A-F) Sequence segments of proteins (Suppl. Fig. S2) encoded by genes within the EDC were used as queries in tBLASTn searches in the tissue transcriptomes of *E. coloratus* (Hargreaves et al., Genome Biol Evol 2014;6:2088-2095). Accession numbers: SRX566773 (brain), SRX566774 (kidney), SRX566777 (liver), SRX566778 (ovary), ERX190966 (skin). (G) RNA-seq reads of the house-keeping gene ALAS1 (located outside the EDC) were investigated as a control. Default settings of tBLASTn at the GenBank server were used. The results help to determine the tissue in which each gene is predominantly expressed. These data do not allow quantitative comparisons of expression levels of different genes. *, only the top 100 hits of the tBLASTn search were counted for LOR2, EDPS, and ALAS1. (H) Number of Giga-bases per transcriptome.

CBPs with 4 beta-sheets

This study, Fig. 3

Python Beta1 sheet 1	PCITS-CPDAKVVVHPPPLILTLPGLSLRTPNQCLLETQTSC
Cobra Beta1 sheet 1	PCVTS-CPDAKVVVHPPPLVLTLPGLSLKTPNQCLVESQTSC
Lizard Beta1 sheet 1	PCVTK-CPDAKMVIHPPPLAITLPGLRLTTNPGKCLVETQTSC
Gecko Beta1 sheet 1	PCVTS-IPDAKVIIHPPSLKLTLPGLILATSPGKTLVETETAC
Python Beta1 sheet 2	PCCTTTCPPDSQVVIQPPPVCTITPGAVLTSYPNECLIKTSTPC
Cobra Beta1 sheet 2	PCCTTTCPPDSQVVIQPPPVCTITPGAVLTSYPNECLISSSTNC
Lizard Beta1 sheet 2	PCSTTACSDSQLVIYPPPVCMIPGAMLSSHPNECLIESSIPC
Gecko Beta1 sheet 2	PCCISTTPDSQMVIKPPPVCTITPGAVLTSFPNECLIATSQPC
Python Beta1 sheet 3	RCVTR-GPTTKVVIHPPPIEVTIPGFVLEIAAQECAVEVYNPC
Cobra Beta1 sheet 3	RCLTQ-GPSNKVVIYPPPIEITIPGFVLEIAAEECTVEVYNPC
Lizard Beta1 sheet 3	PCVTQ-GPGSKVVIHPPPIEIEICLPGFIVEIMAAECAVEVYNPC
Gecko Beta1 sheet 3	ACVAQ-GPSSKVVIYPPPIELIIPGFILLEIEAEECAVEVHNPC
Python Beta1 sheet 4	SCISQ-GPEMKIIIQPPPIEVELPGFILEVFPKCIETLNPS
Cobra Beta1 sheet 4	TCISQ-GPEMKIIIQPPPIEVELPGFILEVFPKCKVETLTPC
Lizard Beta1 sheet 4	SCVSE-GPELRIVVKPPPIEVDMPGFILQVFPETCKVEVLNSG
Gecko Beta1 sheet 4	PCVSQ-GPEMKITIQPPPIEVDLPGFILRVFPKCKVETFSPC

CBPs with 1 beta-sheet

Fraser & Parry 2014

Beta-sheet consensus PCVRQ-CPDSEVVIQPPPVVVTLPGFILSSFPQNTAVGSST

CBPs with 1 beta-sheet

Calvaresi et al. 2016

Chick feather CBP (Fe)	PCVRQ-CQDSRVVIQPSPPVVVTLPGFILSSFPQNTAAGSST
Turtle shell CBP (Tu)	PCVRRQCQDSEVVIRPSPPVVVTLPGFILSNFQHSVGAVG
Snake scale CBP (Sn)	SCINQ-IPASEVTIQPPAVVVTIPGFILSASCDPVAVGGYN
Lizard CBP (HC10)	SCITQ-LPASEMVIQPPAVVVTIGFIMSASCEPIAVGGTT
Gecko scale CBP (Ge)	SCINQ-IPPSEVTIQPPCTVVVPGFVLAASCEPLRVGGYT

Suppl. Fig. S5. Comparison of amino acid sequences in the beta-sheet regions of corneous beta-proteins (CBPs)/beta-keratins with 4 beta-sheets and CBPs with 1 beta-sheet. Proteins encoded by the *Beta1* gene (first gene of the CBP gene cluster within the EDC) of squamates are predicted to form 4 beta-sheets. The sequences shown here correspond to the core sequences indicated in Fig. 3. All other CBPs are predicted to form only 1 beta-sheet. For comparison, the consensus sequence of the beta-sheet region of CBPs, as reported by Fraser and Parry (Fraser, R. D. & Parry, D. A. Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. *J. Struct. Biol.* 188, 213-224. 2014) and core sequences of selected CBPs, as reported by Calvaresi *et al.* (Calvaresi, M., Eckhart, L. & Alibardi, L. The molecular organization of the beta-sheet region in Corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. *J. Struct. Biol.* 194, 282-291. 2016) are shown. Each sheet is composed of anti-parallel beta-strands that are separated by tight turns. Predictions of beta-strand-forming residues are indicated by yellow shading. Proline residues (green fonts) are critical for the formation of tight turns. There are uncertainties in the number of residues in individual beta-strands, and the number of beta-strands per beta-sheet differs in the various models (5 beta-strands per sheet in the model of Fraser & Parry and in Fig. 3; 4 beta-strands per sheet in the model of Calvaresi *et al.*). Cysteine residues (red fonts) are involved in the formation of disulfide bonds which have not been integrated in the models discussed here.

>Ac EDQK

MSSKDQQKKQQSQKPP EEQKASKNECPEKSPKTKQKKQEEKEEPTKKKQN

>Ac EDQM

MCSRQDKDKAYKQEKKEKPSCGRQNSGNESRGCSGRQNSGSSCGKPKPKPSQDQQQQQQQQQVCKVPEQKQK

>Ac EDQSG

MSYQREQQYKEMYPFAPCCPKFYSPPEYSRFPSPSSSSSGSTGIQCQSSSGTHCRWSTRIQSSSGTQCQESSGNRRQWFTGNPCQGSTWNS
CQEPFSGTQYQGFTGSLWQCSSGTDGGTHSQYHYHYRQGSQDPKQC

>Ac EDSC2

MSQQQRGSSCCCSGGSSGGGGGGGGGGCCGRSGGSQQSQSSGGCCRRRSSGGCCSRGGSSQQSQGSSGGCCGRSSGGSSGGCCGGSS
GGGSSQKKMTIQQMK

>Ac EDSRWM

MTYYGCLGYCDFCNYSLEPKGSASNGYQRLYIQILPSEYYACSPCWDNYGHQISCVFHWYRMRFSSYCSFPCYSYNCCSFSYYCDFCHTL
KSSDDYIFRYKSTSRYLCPHGGSYNYSFSYCFKTRCYSFYYSSKTKKCSHSSSNHQFTHRYGSFSSYCSFCYRTETIRGTFCYSSNECCSSC
YSSPLSSCCSYDAKTKAWSFCYSSNESSSSCYGSFSSSCSSCYAFKTKGCTPCYNNHQHSFCHRCSPFYTSRRCSPGPTYRRCSPC
CSSKGYSPHYNSGSKCRTFSYRYSQCLSCYRSRCHSFYYHFGSRGAGFYRNSHYFPCRGADFGCHSFSHSSSTCSSSHSTSCITYSMKSN
SYSSGCRRCSCEW

>Ac EDY1

MAGFILFGYFFDLQKIYGSQGLYHPWYEFMYEFPWSYTDPSDVFANCYVPGYYNGLFNYNNTSFLQFNNSYGAFEGREPNGPDLPSNNG
 GSGYKPRSYSHL SGLSRNGYQFGDSGNRGSNSKPHHFGSFTSDGDRYLNLYDSGCYRFRAYLSNSSRRCYDPCGYNFNSGLSNPCGYPY
 LSGCGGSCYELFPYSYGYFYSYGGSCYNPCGYNYPTSYGGRYCRPRGYTYISGYGGSCYDPPYSAMCFYFYTGRRNYFS

>Aca EDYM1

MTNCNYQCWFQNFPPQGGFQFGFGPQCGLQCFFGPGQCGFGPQCGLQCFFGPGQYGMMPKPSASFMNMNQPPFMNMNQPANFPEFYKMN
FCNIQGGMFCACGPTPCDSKYAESSDAKSTEKHTSSCDTSGHESCTMKNSSQPTDRCSSQRPPCFSPMGISLCSFSCGFSMGRFHGF
SSSCCFPPRNMYTTSKTFKSCYAK

>Ac EDYM2

MYYRYRYSKELGPPLSPFVKRLPKYGSQYIPFYGRYFSSRTLPLEYKGLFSSKTLFLYQRLGIPIFYHQSGAGYQPRGLNPRFESH
GTLPLQORNLRNPQPHVNIYQRHLLQSFYSGMTKDRHGTITKGSVFGAPRALPLRMTKGFPVLYSFKDLEPHLPAKALPRVTKGFPVLYS
PRDLEPHLPKSPPPHIPKSTKTHSAKRIPFVVPDTSQNLNVTAKQQHHTSKSMLSRATKGSRTIASKVFERNLLNFSRSLRKKASRRNRA
KTSQPSLVKGSQPSLVRDSRGSGLAKSSQENLVHTATCFDLKKLSSNVKVSKTGKKYCSAAKWPF

Suppl. Fig. S6. Update of amino acid sequences of proteins encoded by EDC genes of *Anolis carolinensis* (Ac). (A) Newly identified amino acid sequences of EDC proteins in this study. Other EDC genes have already been reported by Strasser et al. (2014) Mol Biol Evol 31:3194-3205. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X's indicate unknown numbers of amino acid residues that could not be predicted because of gaps in the corresponding gene sequences.

A

Pb	EDSPR1	MA	Y	--	Q	C	K	O	F	C	L	P	P	F			
Pb	EDPSQ	MD	C	C	S	--	Q	C	K	O	F	C	L	P	P	F	
Pb	EDSPR2	MS	-----	Q	C	K	O	A	C	K	A	P	F				
Pb	EDCP	MSF	-----	Q	C	K	O	A	C	F	C	S					
Pb	EDEPK	MSTE	--	Q	Q	R	K	O	T	S	V	L	P	F			
Pb	EDEPT	MAY	-----	Q	Y	K	O	F	C	P	P	P	F				
Pb	EDQSG	M	-----	Q	C	K	O	F	O	G	F	P	F				
Pb	EDP3	MS	---	Q	Q	Q	C	K	O	I	F	C	T	P	F		
Pb	EDPQ2	MSYH	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Pb	EDPQ3	MSYLN	--	Q	Q	C	K	O	F	C	F	I	P	F			
Pb	EDSQ	MSY	-----	Q	C	K	O	F	C	L	P	P	F				
Pb	EDPAM	MTH	-----	Q	C	K	L	P	P	E	L	P	F				
Ac	EDCP	MSY	-----	Q	C	K	O	R	C	L	P	P	F				
Ac	EDPQ2	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Ac	EDPQ3	MSSDSF	--	Q	C	T	O	F	C	K	A	P	F				
Ac	EDSPR2	MS	-----	Q	C	K	O	G	C	K	A	P	F				
Ac	EDSQ	MSY	-----	Q	V	K	O	A	S	L	P	P	F				
Ac	EDEPT	MSY	-----	Q	A	R	O	F	C	T	A	P	F				
Ac	EDP3	MSH	-	H	Q	Q	Q	C	V	O	F	S	F	P	F		
Ac	EDQSG	MSY	-	Q	R	E	Q	Q	Y	K	-	E	-	M	Y	P	F
Ac	EDSPR1	MA	C	H	--	Q	C	K	O	F	C	L	P	P	F		
Ac	EDPSQ	MY	C	--	T	D	Q	Q	C	K	O	A	C	L	P	P	F
Hs	Lor	MSY	-----	Q	K	K	O	F	T	P	O	P	F				
Hs	PRR9	MSFS	--	E	Q	Q	C	K	O	F	C	V	P	P	F		
Hs	SPRR1A	MNS	---	Q	Q	-	K	O	F	C	T	P	P	F			
Hs	SPRR1B	MSS	---	Q	Q	-	K	O	F	C	T	P	P	F			
Hs	SPRR2A	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR2B	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR2D	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR2E	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR2F	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR2G	MSY	--	Q	Q	Q	C	K	O	F	C	P	P	F			
Hs	SPRR4	MSS	--	Q	Q	Q	R	Q	Q	Q	C	P	P	Q			
Gg	EDQCM	MSYY	--	E	Q	C	K	O	F	C	L	P	P	F			
Gg	EDPE	M	-----	Q	C	K	O	E	V	T	L	P	F				
Gg	EDYM1	MSYWY	---	Q	Y	K	Q	C	F	I	S						
Gg	EDP3	MSSH	---	Q	-	K	Q	Q	Q	I	A						

B

Pb	EDPKC	Q	Q	K	K	O	F	C	S	---	W	P	F	---	Q	N	K	*					
Pb	EDQL	Q	Q	K	K	O	G	C	Q	---	L	P	F	---	Q	N	K	*					
Pb	Lor1	Q	-	T	K	O	F	I	S	I	P	E	C	I	G	-	T	-	K	*			
Ac	EDCQ2	Q	Q	V	K	O	F	---	T	Q	W	P	F	---	Q	N	A	K	*				
Ac	EDQL	Q	Q	V	K	O	F	---	T	Q	W	P	S	---	Q	N	Q	K	*				
Ac	EDEPK	Q	Q	R	K	O	F	---	S	T	W	P	L	---	-	-	-	-	K	*			
Ac	EDPKC	H	Q	K	K	O	F	C	---	Y	W	E	H	---	-	-	-	-	H	K	*		
Ac	Lor1	Q	-	T	K	O	---	-	-	M	N	T	W	P	S	G	---	Q	N	K	*		
Gg	Lor1	Q	Q	T	-	Q	F	I	S	---	W	P	P	Q	T	-	K	H	K	*			
Gg	EDGH	Q	Q	I	K	O	---	-	-	S	S	Q	W	P	P	S	---	Q	N	K	*		
Gg	EDPE	Q	Q	V	K	O	F	---	-	-	S	W	P	L	T	---	Q	N	K	*			
Gg	EDQL	Q	Q	I	K	O	F	---	-	-	V	Q	W	P	T	---	Q	Q	N	K	*		
Hs	Iv1	Q	Q	-	K	O	---	-	-	E	V	Q	W	P	E	---	-	-	-	K	H	K	*
Hs	Lor	Q	Q	-	K	O	---	-	-	A	T	W	P	S	---	-	-	-	-	K	*		

C

Pb	EDPSQ	G	Q	K	---	Y	C	S	A	S	N	N	W	P	W	*	
Pb	EDYM2	T	G	K	K	-	Y	C	S	T	T	-	K	W	P	F	*
Ac	EDYM2	T	G	K	K	-	Y	C	S	A	A	-	K	W	P	F	*
Gg	EDQrep	H	A	K	K	-	Y	C	S	A	S	-	K	W	P	W	*
Gg	EDYM2	H	S	K	K	S	R	C	-	A	S	-	K	W	L	W	*

Suppl. Fig. S7. Conserved amino acid sequence motifs of snake SEDC proteins. Amino acid sequence alignments of motifs present at the amino-terminus (**A**) and carboxy-terminus (**B, C**) of some but not all SEDC proteins of the lizard (*Anolis carolinensis*, Ac), chicken (*Gallus gallus*, Gg) and human (*Homo sapiens*, Hs). The amino acid sequence motifs shown in A and B were discussed in a previous paper (30). The present study shows that these motifs are also conserved in several SEDC proteins of snakes, represented here by the python (*Python bivittatus*, Pb). Panel **C** shows a newly identified sequence motif at the carboxy-terminus of proteins that are encoded by gene neighbors of the beta-protein gene cluster. *, end of the protein.

```

1
Pb_EDPS1 MFHCCVSCSMGTAIALCFQCCQSRSMETISIKRVVHSMETSIINLCIVSMET-----SSSSSTICILCLISSSSQRCSTICLSLQCSGSSSGNSRIS
Oh_EDPS1 MFHCCMSCQMCIATLGVQFLQQ-----SISITCSPELEHMSVSIIVSLSPFSCIVCRRCSSSSSTISILSSMQPFSISLASQ-----SSSGSSST
Ts_EDPS1 MFHCCMSSQMGRVSTICLQFLQH-----SISIT-----TMSITISILSLSTSISITL-----SSSSSTISILSSALQCSSTISLASQ-----SSSGSSST
Pb_EDPS2 MFGYCLSYVMCIATLDVQFYQOTSSMFAIGICRSVQSSMFFCIIIRFSIGSTSGS-----GGSFIS-CIICMQSPSTFSIIVYS-----
Ts_EDPS2 MFGYCLSYVMCIATLDVQFYQSSMTITIT-----VQSSMFFICTIVRFSIGSTSGS-----GGSFIS-CVICTMQSPSTFSIIVYS-----
Pb_EDPS3 MSSCCV-----SCRVGCIATLRVQCRDQSSSIF-----
Oh_EDPS3 -----XXXXXXXXXXISLIIETSKISPEYIILC

112
Pb_EDPS1 ISLSSTSSSSQSLSILCLCQSSSSSGSSSSSISISFSNOSSSQLYIILSSQPFISLSQSOSSSSHSIIIFCSLSSSSQSFILRLSQTST
Oh_EDPS1 ISILSALISSTISQCCI-----SLCQSSSGSISILASSTSSRPFCLFLCQCSFCISLVSSQSSSGSTIS-----ILSALQLRSSISLASQ
Ts_EDPS1 ISILSALISSSNIPCCI-----SCCQSSSGSRISISLASSTSSQPFCLFLCQCSFCVSLVSSQSSSGSTIS-----ILSALISSSQCCIILG
Pb_EDPS2 -----SGSGSGSSSISISLGHSSSSQSSSILCFSQSSG-----
Oh_EDPS2 -----NSSTSSSQTSILSPCFPSSSQTYIILCLCSAQSSSGQTTSAQGS-----YC-----IVSTQSSSMNFCIAQ
Ts_EDPS2 -----PVSROHTYCIIFG-----SSSSSYSQSVLSFCPPSSSQIYICIVCSAQSSSGQSSGGQSS-----YC-----IVSTQSSSLSFCSQ
Pb_EDPS3 -----SCI-----SCRVQSMETSIISLSIVISMSQSGSTLSSC-----
Oh_EDPS3 -----ESSNSSSGISILGSSSSSNQPFICYIICGQSFQSS-----

223
Pb_EDPS1 SSSSSSSRISISLSSTSSSGQSQCILCFSQCTGSGSGSGSSRISISLSSTSSSSQSYCIIILCSQPF-----SMSTRAPPTCCIIIIISNCFCCG
Oh_EDPS1 SSSRSSSTISILSALISSTSRCCIIFCQSS-----SSSGTISIMN-----FSRSNSTICCFI-----PM-----SSCFIC
Ts_EDPS1 SPSSSGTISILSALISSTSQCCIILGSS-----SSSGTISILSSLIPPPQSSSCYMLCSQSSSGIISIMSSHPPACCFM-----SSCFIC
Pb_EDPS2 -----SGSGSGSSSISISLGHSSSS-----SQSQFYCIIILCT-----CPSI-----SM-----SSSSSQTCCTIISISNCFCCV
Oh_EDPS2 -----SSSSRDEMSTISLSLS-SGQSQCIFCSYSS-----SSSNQSYSVCCSS-----QTSDANVSIIV-----SFSI-----SSSNPPTCRIC-----SCCLV
Ts_EDPS2 -----SSSGFVMSITLSSTSS-SGTQCYCIFCSA-----GSSNQSYCIIILCS-----CPSGASVSIIM-----GFT-----SSSNPPACCFM-----RCCFLDF
Pb_EDPS3 -----ILCPMQSPF-SMSTLSPFPF-----TSSQSYCIIILCS-----CPPI-----SM-----STISPPFSCCIITITISTCFIVF
Oh_EDPS3 -----SGSGTISILSSLIETSSS-----SQCY-IVSFC-----CPSGTISIVNLS-----SSSNPPSCMIIMSSSCCT-IY
331

```

Suppl. Fig. S8. Alignment of EDPS amino acid sequences of snakes. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X's indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Oh, *Ophiophagus hannah* (king cobra); Pb, *Python bivittatus* (Burmese python); Ts, *Thamnophis sirtalis* (common garter snake).

A**SCFN1**

O. hannah genome seq. TATAAAAAGGGTTTGGCTTCCTGGCGTTTCATACTTTTGTCTCTGAGGCTGCTCACCTG
E. coloratus RNA-seq1 -----CACACTTTTGTCTCTGAGGCTGCTCACCTA
E. coloratus RNA-seq2 -----

< intron >
O. hannah genome seq. ATTTTGCTGAGCTGGGTGAG//TGCAGAAAGGGTCTTCAGGATGAGCTACTTTCTGGACA
E. coloratus RNA-seq1 ATTTTGCTGAGCTGG-----AAAACACTTCAGGATGAGCTACTTTCTGGAAA
E. coloratus RNA-seq2 -----

O. hannah genome seq. GTGCTGTACCATTTGTTGGAATCTTTCACAAGTATGCCCGATGCCAAGATGGCAACCTCG
E. coloratus RNA-seq1 GTGCTGTACCATTTGTCGGAAT-----
E. coloratus RNA-seq2 -----TTCACAAGTATGCCCGATGTCAGGATGGCAACCTCG

O. hannah genome seq. CTCTCAACCGGAGAGAAATGAAGGCGCTTATCCAGAAAGAGTTTGCTGAAGTCTTGAGG
E. coloratus RNA-seq1 -----
E. coloratus RNA-seq2 CTCTGAACCGGAGAGAAATGAAGACGCTTATCCAGAAAGAGTTTGCTGAAGTCTTGAG-

intron >
O. hannah genome seq. TGAG//TCCAGAAATCCTTGCGACCCTCAGACAATTGAACTCACTTTCAAGCTGCTA
E. coloratus RNA-seq1 -----
E. coloratus RNA-seq2 -----AATCCTT-----

B**SCFN2**

O. hannah genome seq. TATAAAAAGGAATCGGATACCTGCTATTCTCCAACAGTTCCCTGAAGGCTTCCAACC
E. coloratus RNA-seq3 -----CTGAAGGCTCCCTGCT
E. coloratus RNA-seq4 -----

< intron >
O. hannah genome seq. GTTGAATGTACTGAGCTGGGTGAG//TAGAGACAAAGTCTCCAACATGGCTGGTCTC
E. coloratus RNA-seq3 GTTGAGCATACCGAGCTGG-----GCAAAGTCTCCAACATGGCTGGTCTC
E. coloratus RNA-seq4 -----

O. hannah genome seq. GTGGACAGTATCTGCACCATCATTGCTGTCTTTCAAGTATGCTGACAGGAAGAGT
E. coloratus RNA-seq3 GTGGACAGTATCTGCAACATCATCGTGGTCTTTCAAG-----
E. coloratus RNA-seq4 -----G

O. hannah genome seq. GAGAGTTCCTCCATGAAGCGAAGGCAGATGAAAAGACTCATCCAAAAGAGTTTGGT
E. coloratus RNA-seq3 -----
E. coloratus RNA-seq4 GAGTGTTCTCCTCCATGAAGCGGAGGCAGATGAAAAGACTCATCCAGAAGGAATTTGGT

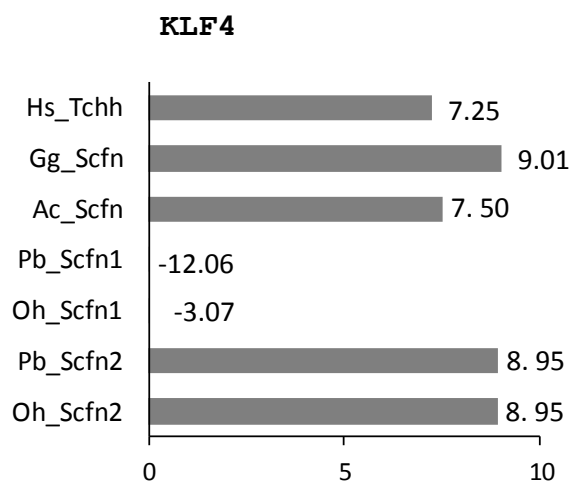
< intron >
O. hannah genome seq. GACGTTCTAGAGGTAAG//CCCAGAACCCCTCGTGATCCTCAGATTGTCAAGCTGACC
E. coloratus RNA-seq3 -----
E. coloratus RNA-seq4 GAAATTCTAGAG-----AACCCCTCGTGACCTCAATTGTCAAGCTG---

Suppl. Fig. S9. Alignment of RNA sequence reads versus genome sequences confirms the expression and the presence of 2 introns in the *SCFN1* and *SCFN2* genes of snakes. (A) The nucleotide sequence of the *scaffoldin 1* (*SCFN1*) gene of the king cobra (*O. hannah*) was aligned to RNA sequence (RNA-seq) reads of *E. coloratus*: RNA-seq1 (GenBank sequence read archive, SRA, accession number ERR216301.7454688.2) and RNA-seq4 (ERR216301.7454688.1). **(B)** The nucleotide sequence of the *scaffoldin 2* (*SCFN2*) gene of *O. hannah* was aligned to RNA-seq reads of *E. coloratus*: RNA-seq3 (ERR216319.8005522.2) and RNA-seq4 (ERR216319.8005522.1). Green shading highlights TATA boxes, blue shading splicing signals (GT and AG) at the ends of introns, and yellow shading start codons. Hyphens were introduced to maximize the alignment. Red letters indicate identical nucleotides in 2 sequences.

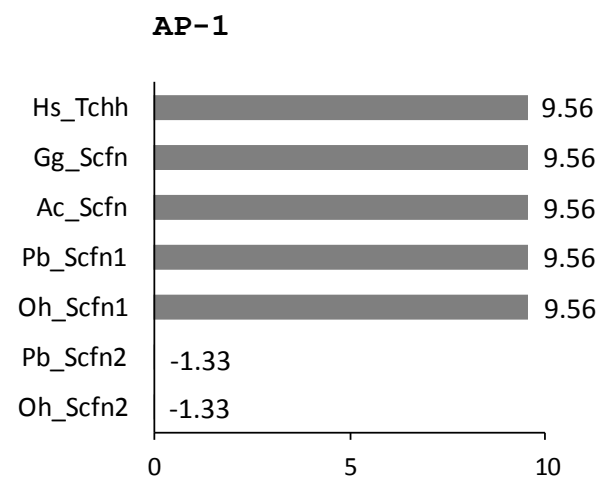
A

	KLF4 *****	AP-1 *****	TATA box *****
Hs_Tchh	GAGCTGGGCTTGGTTAGGAATGAATCAGGCC	//	CCCCATATAAAAGGCC
Gg_Scfn	CAGCTGGGTTTGGCCAAAGATGAATCAGGAC	//	ACCCATATAAAAGCGC
Ac_Scfn	GACTTGGGTTGGGTTGGAGATGAATCAGACC	//	ACTCATATAAAAGAG
Pb_Scfn1	GGAACAAGCCGGTCTAAGGATGAATCAGTTC	//	CAACTATAAAAGGGG
Oh_Scfn1	GGAACGAGCCGGGCTGAAGATGAATCAGATC	//	AACCTATAAAAGGGG
Pb_Scfn2	AAACTGGGCTGGGCTGCAATAAATTATCT	//	AATGTATAAAAGGA
Oh_Scfn2	AAACTGGGCTGGGCTGGACGTAATAATTATCT	//	AATATATAAAAGGA

B



C



Suppl. Fig. S10. Transcription factor binding sites in the promoters of SFTP genes. (A) Nucleotide sequences of the proximal promoters of the genes encoding S100 fused-type proteins (SFTPs) were aligned. As a previous investigation, which did not include sequences of snakes, suggested the presence of binding sites for the transcription factors KLF4 and AP-1 in SFTP promoters (see Figure S4 in Mlitz et al. 2014), the transcription factor binding scores were calculated using the JASPAR 2016 server (<http://jaspar.genereg.net>). The positions of putative binding sites are indicated. Nucleotides compatible with the consensus binding motifs are shaded grey whereas nucleotides incompatible with the transcription factor binding are highlighted with yellow and blue background. **(B)** Binding scores for KLF4 in the promoters of SFTP genes at the site indicated in panel **A**. **(C)** Binding scores for AP-1 in the promoters of SFTP genes at the site indicated in panel **A**. Human (*Homo sapiens*, Hs), chicken (*Gallus gallus*, Gg), green anole lizard (*Anolis carolinensis*, Ac), python (*Python bivittatus*, Pb), and cobra (*Ophiophagus hannah*, Oh) sequences were analyzed. Scfn, scaffoldin; Tchh, trichohyalin.

Suppl. Table S1**Tentative abbreviations and full names of EDC genes identified in this study**

Gene name abbreviation	Full gene name
Crnn	Cornulin
EDCATM	Epidermal Differentiation protein containing the CAT Motif
EDCG	Epidermal Differentiation protein rich in Cysteine and Glycine repeats
EDCM	Epidermal Differentiation protein containing a CCCC Motif
EDCP	Epidermal Differentiation protein rich in Cysteine and Proline
EDCRP	Epidermal Differentiation Cysteine-Rich Protein
EDCS1	Epidermal Differentiation protein, Cysteine-rich Short 1
EDCS2	Epidermal Differentiation protein, Cysteine-rich Short 2
EDCS3	Epidermal Differentiation protein, Cysteine-rich Short 3
EDEPK	Epidermal Differentiation protein rich in glutamic acid (E), Proline and lysine (K)
EDEPT	Epidermal Differentiation protein rich in glutamic acid (E), Proline and Threonine
EDETM	Epidermal Differentiation protein containing an ET Motif
EDETM2	Epidermal Differentiation protein containing an ET Motif 2
EDGPC	Epidermal Differentiation protein rich in Glycine, Proline and Cysteine
EDGPC2	Epidermal Differentiation protein rich in Glycine, Proline and Cysteine 2
EDGY1	Epidermal Differentiation protein rich in Glycine and tyrosine (Y) 1
EDGY2	Epidermal Differentiation protein rich in Glycine and tyrosine (Y) 2
EDHEM	Epidermal Differentiation protein containing a HEM Motif
EDKM	Epidermal Differentiation protein containing a KKLIQQ Motif
EDP3	Epidermal Differentiation protein rich in Proline 3
EDPAM	Epidermal Differentiation protein containing a PA Motif
EDPAML	Epidermal Differentiation protein containing a PA Motif Like
EDPCCC1	Epidermal Differentiation protein containing PCCC repeats 1
EDPCCC2	Epidermal Differentiation protein containing PCCC repeats 2
EDPCCC3	Epidermal Differentiation protein containing PCCC repeats 3
EDPCCC4	Epidermal Differentiation protein containing PCCC repeats 4
EDPCS	Epidermal Differentiation protein rich in Proline, Cysteine and Serine
EDPKC	Epidermal Differentiation protein rich in Proline, lysine (K) and Cysteine
EDPQ1	Epidermal Differentiation protein rich in Proline and glutamine (Q) 1
EDPQ2	Epidermal Differentiation protein rich in Proline and glutamine (Q) 2
EDPQ3	Epidermal Differentiation protein rich in Proline and glutamine (Q) 3
EDPS1	Epidermal Differentiation protein rich in Proline and Serine 1
EDPS2	Epidermal Differentiation protein rich in Proline and Serine 2
EDPS3	Epidermal Differentiation protein rich in Proline and Serine 3
EDPSQ	Epidermal Differentiation protein rich in Proline, Serine and glutamine (Q)
EDQK	Epidermal Differentiation protein containing glutamine (Q) and lysine (K) repeats
EDQL	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin
EDQL2	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin 2
EDQM	Epidermal Differentiation protein containing a glutamine (Q) Motif
EDQSG	Epidermal Differentiation protein rich in glutamine (Q), Serine and Glycine
EDSC1	Epidermal Differentiation protein rich in Serine and Cysteine 1
EDSC2	Epidermal Differentiation protein rich in Serine and Cysteine 2
EDSCP	Epidermal Differentiation protein rich in Serine, Cysteine and Proline
EDSPR1	Epidermal Differentiation protein Small Proline Rich 1
EDSPR2	Epidermal Differentiation protein Small Proline Rich 2
EDSQ	Epidermal Differentiation protein rich in Serine and glutamine (Q)
EDSRWM	Epidermal Differentiation protein containing a SRW Motif
EDWM	Epidermal Differentiation protein containing a WYDP Motif
EDY1	Epidermal Differentiation protein rich in tyrosine (Y)
EDYM1	Epidermal Differentiation protein containing Y Motif 1
EDYM2	Epidermal Differentiation protein containing Y Motif 2
Lor1	Loricrin 1
Lor2	Loricrin 2
Pglyrp3	Peptidoglycan recognition protein 3
Scfn1	Scaffoldin 1
Scfn2	Scaffoldin 2

Note - EDC genes encoding corneous beta-proteins (beta-keratins) and S100A proteins are not included here.

Suppl. Table S2

Burmese python (*Python bivittatus*) EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by <i>P. regius</i> RNA-seq data
S100A1	NW_006533184.1	1588	555	yes	yes
S100A13	NW_006533184.1	14911	15919	yes	yes
S100A14	NW_006533184.1	19278	23485	yes	yes
S100A16	NW_006533184.1	37282	38809	yes	yes
S100A2	NW_006533184.1	46616	47229	yes	yes
S100A3	NW_006533184.1	51063	52413	yes	yes
S100A4	NW_006533184.1	56042	57171	yes	yes
S100A5	NW_006533184.1	61202	61812	yes	yes
S100A6	NW_006533184.1	66291	66748	yes	yes
S100A12	NW_006539396.1	10278	9093	yes	yes
PGLYRP3	NW_006539396.1	4288	1166	yes	yes
EDKM	NW_006540970.1	21714	20641	yes	yes
EDPQ3	NW_006540970.1	15196	15360	yes	yes
EDPQ2	NW_006540970.1	9469	9227	yes	yes
EDSC1	NW_006543838.1	4076	3660	yes	no
EDSC2	NW_006543838.1	12950	13385	no	no
EDQM	NW_006533945.1	464	240	yes	yes
EDWM	NW_006533945.1	10373	9753	yes	yes
EDPQ1	NW_006533945.1	15513	14725	yes	yes
EDCS1	NW_006533945.1	21723	22019	yes	yes
EDHEM	NW_006533945.1	24848	24321	yes	yes
EDCS2	NW_006533945.1	31505	31266	yes	yes
EDCM	NW_006533945.1	35942	36157	yes	yes
EDCS3	NW_006533945.1	40476	39946	yes	yes
EDPCCC1	NW_006533945.1	46880	46335	yes	yes
EDPCCC2	NW_006533945.1	51062	50646	yes	yes
EDPCCC3	NW_006533945.1	63364	61049	no	yes
EDCG	NW_006533945.1	77683	77486	yes	yes
EDPCCC4	NW_006533945.1	81785	81345	yes	yes
EDGPC	NW_006533945.1	85309	85572	yes	yes
EDQL	NW_006533945.1	92571	<92239	no	yes
LOR2	NW_006533945.1	102481	101773	no	yes
LOR1	NW_006533945.1	109065	107293	no	no
EDY1	NW_006538280.1	8900	8205	yes	yes
EDSRWM	NW_006538280.1	15145	17079	yes	yes
EDGY1	NW_006538280.1	23044	22733	yes	yes
EDGY2	NW_006538280.1	34934	35329	yes	yes
EDETM	NW_006540169.1	33959	34210	yes	yes
EDPS1	NW_006541849.1	19165	20127	yes	yes
EDPS2	NW_006547155.1	<1262	867	no	no
EDPS3	NW_006533133.1	7040	7441	yes	yes
EDSCP	NW_006533133.1	11620	10994	yes	yes
EDYM2	NW_006533133.1	51320	50361	yes	yes
EDPSQ	NW_006533133.1	59918	58921	yes	yes
EDEPK	NW_006533133.1	66212	65754	yes	no
EDPKC	NW_006533133.1	77786	78652	yes	yes
EDP3	NW_006533133.1	83736	83521	yes	yes
EDQSG	NW_006533133.1	87350	87931	yes	yes
EDSPR1	NW_006533133.1	90689	90375	yes	yes
EDPCS	NW_006533133.1	97800	98342	yes	yes
EDCP	NW_006533133.1	107373	106795	yes	yes
EDCATM	NW_006533133.1	113018	113974	yes	yes
EDPAM	NW_006533133.1	121454	122536	yes	yes
EDSQ	NW_006533133.1	130252	131010	yes	yes
EDEPT	NW_006533133.1	135106	135657	yes	yes
EDSPR2	NW_006533133.1	139240	139043	yes	yes
CRNN	NW_006533133.1	146971	149243	yes	yes
SCFN2	NW_006533133.1	154919	152258	no	no
SCFN1	NW_006533133.1	160207	164241	yes	no
S100A11	NW_006533133.1	167015	170608	yes	yes
S100A10	NW_006533133.1	183743	184878	yes	yes

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S3

Burmese python (*Python bivittatus*) corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by <i>P. regius</i> RNA-seq data
Beta1	NW_006533945.1	129573	128446	yes	yes
Beta2	NW_006533945.1	141471	141166	yes	yes
Beta3	NW_006533945.1	148257	149408	yes	yes
Beta4	NW_006533945.1	153207	152950	yes	yes
Beta5	NW_006533945.1	162249	161980	yes	yes
Beta6	NW_006533945.1	175889	175638	yes	no
Beta7	NW_006533945.1	178952	179227	yes	no
Beta8	NW_006533945.1	182958	182686	yes	yes
Beta9	NW_006533945.1	185558	185833	yes	no
Beta10	NW_006533945.1	187387	187124	yes	yes
Beta11	NW_006533945.1	193099	192722	yes	no
Beta12	NW_006533945.1	196728	197000	yes	no
Beta13	NW_006533945.1	204035	203727	yes	no
Beta14	NW_006533945.1	212727	213041	yes	yes
Beta15	NW_006538280.1	44064	44354	yes	no
Beta16	NW_006538280.1	49426	49821	yes	yes
Beta17	NW_006538280.1	52808	52377	yes	yes
Beta18	NW_006538280.1	57570	58020	yes	yes
Beta19	NW_006538280.1	62708	62235	yes	yes
Beta20	NW_006542926.1	15275	>15577	no	yes
Beta21	NW_006542926.1	2255	2746	yes	yes
Beta22	NW_006551345.1	>212	15	no	no
Beta23	NW_006555159.1	>522	334	no	no
Beta24	NW_006552039.1	197	556	yes	no
Beta25	NW_006540169.1	583	1002	yes	yes
Beta26	NW_006540169.1	5880	<5596	no	yes
Beta27	NW_006540169.1	<11505	11825	no	yes
Beta28	NW_006540169.1	17257	16829	yes	yes
Beta29	NW_006541849.1	>5559	5290	no	yes
Beta30	NW_006541849.1	13806	13432	yes	yes
Beta31	NW_006533133.1	21000	20548	yes	yes
Beta32	NW_006533133.1	27649	28053	yes	yes
Beta33	NW_006533133.1	31407	30952	yes	yes
Beta34	NW_006533133.1	39313	39747	yes	no
Beta35	NW_006533133.1	43812	43522	yes	no

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S4

King cobra (*Ophiophagus hannah*) EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete
S100A9	AZIM01042421.1	45	>188	no
PGLYRP3	AZIM01003248.1	1855	5074	yes
EDKM	AZIM01003248.1	12034	12171	yes
EDPQ3	AZIM01003248.1	19465	<19349	no
EDPQ2	AZIM01003248.1	33024	33395	yes
EDSC1	AZIM01003248.1	41385	41068	yes
EDSC2	AZIM01003248.1	52467	52712	yes
EDWM	AZIM01003248.1	68925	68212	yes
EDPQ1	AZIM01003248.1	73668	72802	yes
EDCS1	AZIM01003248.1	78687	78968	yes
EDHEM	AZIM01003248.1	82623	82117	yes
EDCS2	AZIM01003248.1	87881	87576	yes
EDCM	AZIM01003248.1	92670	92894	yes
EDCS3	AZIM01003248.1	97472	96797	yes
EDPCCC1	AZIM01003248.1	104405	104049	yes
EDPCCC2	AZIM01003248.1	109843	109472	yes
EDPCCC3	AZIM01003248.1	122470	118393	no
EDCG	AZIM01003248.1	133443	133243	yes
EDPCCC4	AZIM01003248.1	137976	137235	no
EDGPC	AZIM01003248.1	141722	142009	yes
EDQL	AZIM01003248.1	145280	144999	yes
LOR2	AZIM01003248.1	151769	150621	yes
LOR1	AZIM01003248.1	157066	155969	yes
EDY1	AZIM01004187.1	73541	74254	yes
EDSRWM	AZIM01004187.1	64099	62798	yes
EDGY1	AZIM01004187.1	58166	58483	yes
EDGY2	AZIM01004187.1	47891	47478	yes
EDETM	AZIM01000954.1	23704	23952	yes
EDPS1	AZIM01000954.1	41059	41844	yes
EDPS2	AZIM01000954.1	58107	58886	yes
EDPS3	AZIM01000954.1	<63568	63978	no
EDSCP	AZIM01000954.1	69647	69066	yes
EDYM2	AZIM01000954.1	107053	106094	yes
EDEPK	AZIM01000954.1	119831	119364	yes
EDPKC	AZIM01000954.1	131139	131972	yes
EDP3	AZIM01000954.1	135690	135409	yes
EDQSG	AZIM01000954.1	140286	140879	yes
EDSPR1	AZIM01000954.1	146227	145952	yes
EDPCS	AZIM01000954.1	153869	154345	yes
EDQK	AZIM01000954.1	156359	156141	yes
EDCP	AZIM01000954.1	162072	161572	yes
EDCATM	AZIM01000954.1	167242	168537	yes
EDPAM	AZIM01000954.1	174551	175153	yes
EDSQ	AZIM01000954.1	183997	184629	yes
EDEPT	AZIM01004605.1	3309	3890	yes
EDSPR2	AZIM01004605.1	8254	8066	yes
CRNN	AZIM01004605.1	14916	17816	yes
SCFN2	AZIM01004605.1	23685	21601	yes
SCFN1	AZIM01004605.1	28678	32401	yes
S100A11	AZIM01004605.1	<38940	40166	no

Notes - CDS, coding sequence. Further S100A genes are present close to border of the EDC.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S5

King cobra (*Ophiophagus hannah*) corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete
Beta1	AZIM01003248.1	174867	173737	yes
Beta2	AZIM01003044.1	87502	87801	yes
Beta3	AZIM01003044.1	81345	79850	yes
Beta4	AZIM01003044.1	76531	76767	yes
Beta5	AZIM01003044.1	68820	69074	yes
Beta6	AZIM01003044.1	55588	55848	yes
Beta7	AZIM01003044.1	52619	52344	yes
Beta8	AZIM01003044.1	49125	49403	yes
Beta9	AZIM01003044.1	43757	43984	yes
Beta10	AZIM01007131.1	26847	27329	yes
Beta11	AZIM01007131.1	17992	17738	yes
Beta12	AZIM01007131.1	13225	13509	yes
Beta13	AZIM01007131.1	3381	3070	yes
Beta14	AZIM01004187.1	42024	41680	yes
Beta15	AZIM01004187.1	36898	36512	yes
Beta16	AZIM01004187.1	34624	35034	yes
Beta17	AZIM01004187.1	30955	30533	yes
Beta18	AZIM01004187.1	27140	27531	no
Beta19	AZIM01004187.1	23240	22722	yes
Beta20	AZIM01004187.1	18585	18992	yes
Beta21	AZIM01004187.1	7539	7093	yes
Beta22	AZIM01004187.1	2467	2882	no
Beta23	AZIM01041946.1	772	344	yes
Beta24	AZIM01008786.1	16443	16057	yes
Beta25	AZIM01008286.1	17037	16576	yes
Beta26	AZIM01008286.1	11937	12416	yes
Beta27	AZIM01008286.1	8808	8353	yes
Beta28	AZIM01008286.1	1958	>2062	no
Beta29	AZIM01000954.1	3170	2733	yes
Beta30	AZIM01000954.1	29658	29272	yes
Beta31	AZIM01000954.1	38016	37636	yes
Beta32	AZIM01000954.1	47356	46940	yes
Beta33	AZIM01000954.1	76707	76306	yes
Beta34	AZIM01000954.1	85348	85815	yes
Beta35	AZIM01000954.1	89721	89242	yes
Beta36	AZIM01000954.1	100589	101041	yes

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S6

Green anole lizard (*Anolis carolinensis*) EDC genes newly (*) identified in the present study.

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq data (<i>A. carolinensis</i>)
EDSC2	NW_003338916.1	1040943	1040629	yes	yes
EDQM	NW_003338916.1	1036851	1037078	yes	yes
EDHEM	NW_003338916.1	995360	996070	yes	yes
EDCS1	NW_003338916.1	990209	990406	yes	yes
EDCS2	NW_003338916.1	982933	983241	yes	yes
EDCS3	NW_003338916.1	961905	962762	yes	yes
EDPCCC1	NW_003338916.1	951749	951994	yes	yes
EDPCCC2	NW_003338916.1	936661	936473	yes	yes
EDPCCC3	NW_003338916.1	<919157	919798	no	yes
EDGPC1	NW_003338916.1	893728	893540	yes	yes
EDGPC2	NW_003338916.1	876803	876600	yes	yes
EDYM1	NW_003338916.1	780003	780617	yes	yes
EDY1	NW_003338916.1	571522	572292	yes	yes
EDSRWM	NW_003338916.1	561996	560887	yes	yes
EDGY1	NW_003338916.1	556637	557008	yes	yes
EDGY2	NW_003338916.1	540103	539129	yes	yes
EDETM1	NW_003338916.1	305599	305339	yes	yes
EDETM2	NW_003338916.1	298118	297858	yes	yes
EDPCCC4	NW_003338916.1	238742	239044	yes	no
EDYM2	NW_003338916.1	184370	185365	yes	yes
EDP3	NW_003338916.1	129131	129376	yes	yes
EDQSG	NW_003338916.1	120578	120174	yes	yes
EDQK	NW_003338916.1	87721	87867	yes	yes
EDCATM	NW_003338916.1	57594	57070	yes	yes
EDPAML	NW_003338916.1	47259	46576	yes	yes

Notes - * other EDC genes have been reported in a previous paper of our laboratories (30).

CDS, coding sequence; the symbols < and > indicate that ends of the coding sequence were not present on the scaffold.¶