Supplementary material for

Molecular convergence of infrared vision in snakes

Shozo Yokoyama[1] Ahmet Altun[1,2] and Dale F. DeNardo[3]

[1] Department of Biology, Emory University

[2] Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University

[3] School of Life Sciences, Arizona State University

Materials and Methods

Statistical analyses of DNA and amino acid sequences

To establish the phylogenetic relationship of the TRPA1 proteins of not only the infrared-sensitive boa *(Corallus hortulanus*), python (*Python regius*), and pit viper (*Crotalus atrox*) but also infrared-nonsensitive rat snake (*Elaphe obsoleta lindheimeri*), we aligned the nucleotide sequences of TRPA1 genes of rattlesnake (GenBank accession number: GU562967), python (GU562965), boa (GU562969), rat snake (GU562966), zebrafish a (NM_001007065), zebrafish b (NM_001007066), frog (NM_001127962), zebra finch (XM_002197822), chicken (XM_418294), human (NM_007332), macaque (XM_001083172), mouse (AY231177), rat (NM_207608 XM_232586), dog (XM_544123), XM_001493464 (XM_001493464), pig (XM_001926115), cow (XM_581588), and opposum (XM_001378390).

The alignment of these sequences was done first by applying Clustal W to the amino acid sequences deduced from these genes and then adjusted visually. Then, we applied the NJ method (Saitou and Nei 1987) to the p idstances for a total of 929 common codons (positions 62-88, 91-609, 612-622, 626-685, 687-746, 756-823, and 825-1010; numbering follows the western diamondback rattlesnake sequence) and orthologous mosquito (EU624401) and fruit fly (NM_140006) sequences, and constructed a rooted phylogenetic tree of vertebrate TRPA1 genes (and TRPA1 proteins). To infer the amino acid sequences at all ancestral nodes in Fig. 1A, the codeml of PAML (Yang 2007) was applied to the corresponding amino acid sequences. In the PAML analyses, Dayhoff, Wagner, and JTT models were used.

Molecular structure of ankyrin repeats

To study the effect of three amino acid changes (L330M, Q391H, and S434T) on the structure of the 17 ankyrin repeats of TRPA1 proteins that are located at the amino acid sites 68-655 (fig. S2), we evaluated their structures of the ancestral snake TRPA1 with and without the three mutations.

The initial coordinates of the ancestral TRPA1 were obtained by applying homology modeling (CPH models 3.0 server, www.cbs.dtu.dk/services/CPHmodels) to human ankyrin repeat (pdb code: 1N11). Since the number of amino acids between the ancestral TRPA1 (588) and the human orthologue (404 resolved out of 437) differ, we considered the three overlapping ankyrin segments of TRPA1 separately: 1) the segment without the first and last 70 amino acids (M1), 2) the first 150 amino acids (M2), and 3) the last 150 amino acids (M3). Then, the last/first 80 amino acids of M2/M3 structure were superimposed with the corresponding amino acids in the M1 structure. Good agreement was observed for the overlapped parts, indicating a high quality for the structures. Then, M1-3 structures were combined by deleting the coordinates of the common amino acids in M1 with M2 and M3.

Based on PROPKA (http://propka.ki.ku.dk/) pKa calculations and visual inspection, we applied standard protonation states for all amino acids of the resulting structure except histidines: 12 histidines at the protein surface were doubly protonated while the remaining 18 histidines were singly protonated at either δ or ε nitrogen depending on their local environment. Standard N and C terminals were used for L68 and K655. The net charge of the resulting model is zero.

After protonating the modeled TRPA1, a series of geometry optimizations were performed with AMBER96 (http://ambermd.org) force field. First, only hydrogens were relaxed. Next, only C_α atoms were fixed while the rest was relaxed. Then, all coordinates were fully relaxed to obtain final structure of the ancestral TRPA1.

Three mutations (L330M/Q392H/S435T) were then introduced into the ancestral TRPA1 structure by first relaxing only the coordinates of the mutated residues and then all coordinates of the mutant TRPA1.

```
                                                             .
                      10        20        30        40        50        60
Rattlesnake   MQRSILKLFQSRVQKQQSVD--SSDGVVSEGDACSVASQDIFKVVSDGNTSRLRSLIKKNRR
Rat snake     MKRSILKWFQSRDQKQH---PSSYEGVVCEADACSVASQDVFKVISDGSTCRLRSFIKKNRE
Python        MMRSILKWFQSHVQKQEQQDALSYEGVVCEPDACSVASQDVFKVISGGSTSRLRSFIKKNRN
Boa           MKRWILKWFQSHGQK---EQDIYYDGVVCEPDACSVASQDIFKVMSDGNTSRLRSFIKKNSK
                                   ANK1
                     _70________80________90_______100___ ___110_______120
ancestor       LKKLDELNATPLHHAAGNGQLELMQMIMDDSSPEALNVADSSGNTPLHWATKKHQIESV
Rattlesnake   RLKKLDRLNATPLHHAAGNGQLELMQMIMDGSSPEALNVADVSGNTPLHWATKKHQTESI
Rat snake     GLKKLDKLNATPLHHAAGKGQLELMQMIMDDSSFEALNVTDSSGNTPLHWATKKQQTESV
Python        GLKKLDEMKATPLHHAAGGGQLELMLMIMDGSSAEALNIGDAYGNTPLHWATKKHQIESV
Boa           GLKKLDDMNASPLHHAAANGQLEFMQMIMDDSPPEALNVADNAGNTPLHWATKKQQIESV
                ANK2                          ANK 3                    .
              _______130_____ _140_______150_______160_______170_______180
ancestor      KLLLSRGANPNILNSNMMAPLHWAVQYLFNDLVKIFLENSITDVNLEGEGGNTPILLACY
Rattlesnake   KLLLSRGANPNILNVNMIAPLHWALHYLLDDLVTIFLECSNTNINLEGEGGNTPILLACY
Rat snake     KLLLSRGANPNILNSNMVSPLHWAVQYLCNDLVKIFLECSITDVNLEGEGGNTPILVACY
Python        KLLLSRGASPNSLNKNNMAPLHWAVQYFFDDLLKIFIENDMTEINLEGESGNTAILLACY
Boa           KLLLRKGANPNVANANMMTPFHWVVQYHFNDLVKIFLENSLTDVNLGGEGGNTPILLACY
                 .       ANK 4 .                       ANK 5 .
              _______190_______200_ _____210_______220_______230_______240
ancestor      KDNPEALKLLIEKGGDICKANNMGCMPVHAAAFSGSKLCMEIIIKRGEELGYSPESHINF
Rattlesnake   KDNPTALKILIEKGGDICKVNNMGCMPVHAAAFSGSKLCLEMIIKRGEQLGYSPKNHINF
Rat snake     KDNSEALKLLIENGGDIAKANNMGCMPVHAAAFSGSKLCLEIIIKRGVELGYSPENHINF
Python        KDNPEALKLLIERGADICKANHMGTMPIHAAAFSGSKICMEMVIKQAEKLGHSPESHINS
Boa           KDNPEALKLLIEKGGDICKANNTGCMAVHAAATSGAKLCMEIVIRRGEDLGYTRESHINF
                                ANK 6                            ANK 7
              ___ ___250_______260_______270_____ _280_______290_______300
ancestor      TNNGKSSPLHLAVQSRDLEMIKMCIEYGAQIDLKQNDNCTALHFAATQGATEILKLMMSS
Rattlesnake   INNEKSSPLHLAVQSRDVEMIKMCIEYGAQIDLKQSDNCTALHIAAIQGATEIIELLMSA
Rat snake     TNNGKCSPLHLAVQSRDLEMIKMCIEYGAQIDLKQNDNCTALHFAATQGATEILKLMMSS
Python        LDNARNSPLHLAVQSRDLEMIRMCIEYGAQVDLKQNDKCTALHFAATQGATEILKLMISS
Boa           TTTGKSSPLHLAVQSRDLEMIKMCIEYGAQIDLKQNDNCTALHFAATQGATDILKLMMST
                                     ANK 8 *           .
              _______310__ ____320_______330_______340_____ _350_______360
ancestor      YTGEESIINAVDGNKETLLHRAALFDHHELAEYLISKGANIDSVDTEGRTPLLLATSCAS
Rattlesnake   YSGEECLINASDENKETLLHRAALFDHDEMTDYLISKGANIDSVDIEGRTPLILATSRAS
Rat snake     YTGEESIINALDGNKETLLHRAALFDHHELAEYLISKGANINSVDIEGRTPLLLATSCAS
Python        YTGEVSLINVVDGKKETLLHRAALFDHCEMAEYLISMGAHIDSVDTEGRTPLLLATSCGS
Boa           YEGDESIVNAADGNKETMLHRAALFDHCEMAEYLISKGAKIDCVDHEGRTPLLLGTACGS
```

Supplementary Figure 1: The TRPA1 proteins of the rattlesnake, rat snake, python, boa, and their common ancestor. Amino acid sites that are conserved among the rattlesnake, python, and boa are shown by dots (.) and stars (*). The dots indicate that amino acids are highly variable among the 18 orthologous receptors and/or not unique to the three snake proteins, whereas the stars indicate not only are amino acids unique to the three infrared-sensitive proteins but also all others are identical (see also Fig. 1). A partial sequence of the ancestral snake TRPA1 protein is also shown, where amino acids that have a posterior probability of < 0.95 are underlined. TM I-VI is indicated by black boxes.

```
                     ANK 9                  *    ANK 10
                     370       380       390       400       410       420
ancestor      WKIVNLLLSKGANVELKDHLGRNFLHLTVLQPGGLQHLNEEFLKMKHIRDLVTEEDHEGC
Rattlesnake   WKIVNLLISKGANIQLKDHLGRSFLHLTVLHPGGLQHLNEEFLKMKHIRDLLTDEDHEGC
Rat snake     WKIVNLLLSKGANVELKDLLGHNFLHLTVLQPGGLQHLNEDFLKMKHIRDLITEEDQEGC
Python        WKLVNLLLSKGANLELKDHLGRNFLHLTVLHPGGLQHLSEQFLKMKVIEDLVGEEDREGC
Boa           WKIVNLLLSKGANLELKDHLGRNVLHLAVLHPGGLQRLNEEILKLTCIRNLATEEDHEGC
                           *  ANK 11                        ANK 12
                     430       440       450       460       470       480
ancestor      TPLHYACKQGVPLSVNILLGMNVSVYAKSRDKKSPLHFAASYGRINTCHRLLEAMKDTRL
Rattlesnake   TPLHYACKQGMPLTVNILLGMNVSVYAKSRDKKSPLHFAASHGRLNTCLRLLESMDDTRL
Rat snake     TPLHYASKQGVPLSVNILLEMNVSVYAKSRDKKSPLHFAASYGRINTCLRLLEAMEDTRL
Python        TPLHYACKQGVPLTVNILLKMNVSVYAKSREKKSPLHYAASYGRIHTCHRLLESMPDTRL
Boa           TPLHYASRQGVPLTINTLLSMNVSVYCKSRDKKSPLHFAASNGRLNTCCLLLEAMKDTRL
                                 ANK 13                       ANK 14
                     490       500       510       520       530       540
ancestor      LNEGDKKGMTPLHLAAQNGHEKVVQFLLKKGALFLCDYKGWTALHHAAFGGYTRTMQIIL
Rattlesnake   LNEGDRKGMTPLHLAAQYGHEKVTQLLLKKGALFNSDYKGWTPLHHAALGGYSRTMEIIL
Rat snake     LNEGDKKGMTPLHLAAQNGHEKVVQFLLKKGALFLCDYKGWTALHHAAFGGYTRTMQIIL
Python        LNEGDKKGLTPLHLASQNGHEKVVQLLLKRGALFGCDNKGWTALHHAAFGGYTRTMQIIL
Boa           LNEGDKKGMTPLHLAAENGHEKIAQFLLKKGALFLSDNKGWTALHHAAFGGYSRTMQAVL
                                  ANK 15
                     550       560       570       580       590       600
ancestor      NTNMKATDKVDEEGNTALHLAAREGHAKAVKLLLDDNAKILLNKAEASFLHEAIHNGRKD
Rattlesnake   NTNMKSTDKVNDKGDTALHLAAREGHARAVKLLLDANAKILLNESEASFLHEAIHNERKD
Rat snake     NTNMKATDKVNDEGNTALHLAAREGHAKAVKLLLDDNAKILLNSAEASFLHEAIHNGRKD
Python        NTNMIATDKEDEDGNTGLHLAAREGHAKAVKLLLDGNAKIVLNKAEASFLHEAIRNGRKN
Boa           DTNVKATDNVDEDGNTALHLAAREGHAKAVKLLLDEGAKILLNKVEASFLHEAIHNGQRD
                        ANK 16               ANK 17
                     610       620       630       640       650       660
ancestor      VVNVVILHKRWEESITTFSHNSSINKCAILEMVEYLPDCLKLVLDNCIIESPEEKTSRDF
Rattlesnake   VVKIVILHKRWEESISTFSHFSSINKCAILEMVEYLPDCLKLVLDNCIVKSPEDKGSKDF
Rat snake     VVNAVILHKRWEESITTFSHHSSINKCAILEMVEYLPECLKLVLDNCIIESPDEKGSKDF
Python        VVNVVILHKRWEESMTTFSHNSSINKCAILEMVEYLPDCLKLVLDNCIVRSPDEKTSRDF
Boa           VVTVVILHKRWEEAISTFSHQSSVNKCAILDMVEYLPGCLKLVLDNCIIRSPEEKTSADF

                     670       680       690       700       710       720
ancestor      FIEYNFRYLQCPLKLK  VKENEGV YEPLLALNAMVRHNRMELLSHPVCTEYLLMKWMA
Rattlesnake   FVEYNFTYLQCPLQFKKKVKENEGVVYEPLLALNGMVRHNRVELLSHPVCTQYLLMKWMA
Rat snake     CIEYNFRYLQCPLKLKKKFKENEGIIYEPLLALNGMVRHNRVELLSHPVCTEYLLMKWMA
Python        SIQYNFRYLQCPLTLTKKSKAQDGVVYEPLLALNAMVRYNRMELLSHPVCTEYLLMKWMA
Boa           FIEYNFRYLQCPLKLTKKSKGNGGALYEPLLALNAMVLHNRMELLSHPVCTEYLLMKWMA
                          TM I          .                        TM II
                     730       740       750       760       770       780
ancestor      YGFRAHIMNLAVYSLGLIPLTLLVTS         TLKHGPFDNKDSNFIKVCMCLVFIM
Rattlesnake   YGLRAHILNLAVYSLGLIPLTLLVTSLEPDISFDEPLQYGPFDNKKSYFIKVCMSLVFIM
Rat snake     YGFRAHILNLAVYSLGLIPLTLLVTHLEPDVCFNATLKYGPFDNKDSNFIKVCMSLVFIM
Python        YGFRAHIMNLAVYSLGLIPLTLLITSIEPEVSFNTTFKHGPFDKKDSNFIKVCMCLVFIM
Boa           YGFRAHIMNLAVYSLGLIPLTLLITSLEPKVTFNATLKHGPFDKKNSSFIKVCMCLVFIM
```

Supplementary Figure 1 (continued)

```
                                              TM III         ..
                     790       800       810       820       830       840
ancestor      SLFGICKEIIQLFQQKLNYLLDYSNLLDWTIYTTSIIFVSSLF KLPVHLQWDCGAIAIL
Rattlesnake   SLFGICKEIIQLFQQKLNYLMDYSNLLDWAIYTTSLIFVSSLLVTLPVHLQWHCGAIAIL
Rat snake     SLFGICKEIIQLFQQKLNYLLDYSNLLDWTIYTTSIIFVSSLFVMLPIRLQWDCGAIAIL
Python        SLFGICKEFMQLFQQKLRYLLDYTNLLDWTIYTTSIIFVSSLFVKLPVHLQWECGAIAIL
Boa           SLFGICKEIIQLFQQRLKYLLDYSNLLDWTIYTTSIIFVSSLFVKLPVHLQWDCGAIAIL
                TM IV                        TM V
                     850       860       870       880       890       900
ancestor      LSWTNFLLYLQRFENYGIYVVMFWEILRTLIRIVVVFFFLMLAFGLSFYVLLGSQQTYGT
Rattlesnake   LAWTNFLFYLQRFENYGIYIVMFWEILKTLIRIIVVFFFLMLAFGLSFHVLLGSQETYGT
Rat snake     LAWTNFLLYLQRFENYGIYIVMFWEILRTLIRIVIVFFFLMLAFGLSFYVLLGSQQTYGT
Python        LSWTNFLLYLQRFENYGIYVVMFWEILRTLIRIVVVFFFLMLAFGLSFHVLLGSQQTYGT
Boa           LSWTNFLLYLQRFESYGIYVVMFWEILRTLIRIVTVFFFLMLAFGLSFFVLLGSQQTYGT
                          .  .                             . TM VI
                     910       920       930       940       950       960
ancestor      PYLSVMQTFSMMLGDVNYRDAFLEPMLANKLPFPFLSFIILIIFSLLIPILLMNLLIGLA
Rattlesnake   PYLSVMQTFSMMLGDVNYREAFLEPMLVDKLPFPFLSFLILILFSLLIPILLMNLLIGLA
Rat snake     PYLSVMQTFSMMIGDNNYREAFLEPMLADKLPFPFLSFIILIIFSMLIPILLMNLLIGLA
Python        PYLSVMQTFSMMLGDVNYRDAFLEPMFTHKLPFPFLSFILLIIFSLLIPILLMNLLIGLA
Boa           PYLSVMQTFSMMLGDVNYRDAFLEPMLHNRLPRPFLSFIILIVFSLLIPILLMNLLIGLA
                                                                      .
                     970       980       990      1000      1010      1020
ancestor      VGDIAEVQKFAAMKRIAMQVNLHTNLEKKLPYWFLSRVDQESITVYPNRP
Rattlesnake   VGDIAEVQKFAAMKRIAMQINLHTNLEKKLPYWFLNRVDQQSMVIYPNRSRFGGVMSMFQ
Rat snake     VGDIAEVQKFAAMKRIAMQVNLHTNLEKKLPYWFLSRVDQESIVVYPNKPRYCGFMTVFQ
Python        VGDIAQVQKFAAMKRIAMQVNLHTNLEKKLPYWFLSRVDQESIILYPNRPRYCGFMSVFH
Boa           VGDIAQVQKFAAMKRIAMQVSLHTNLEKKLPYWFLTRVDQETLTLYPNKPRYCGFMSVLN
                     .                                 .           .
                    1030      1040      1050      1060      1070      1080
Rattlesnake   YCFGWDNIASDAKNADAAIELELLKQKYRLKDTANLVEKQHDLLKQIAKKMEVISVVEDE
Rat snake     YCFGWDNTAADTQSADTTLELEVLKQKYRLKDIAALVEKQHNLLKLVAQKMEIMSEAEDE
Python        YCFGWDSIAADAQGADTALELEVLKQKYRLKDVALLMEKQHDLLKLLTRKLEIISEVEDE
Boa           YCFGWDSIAADTQNADTTLELEILKLKYRLKDMATVIEKQHDLLKLVARKLEIISELEDE

                    1090      1100      1110
Rattlesnake   DLNDSLQSQFRKEHFENTNNKWDTVFKAVKS
Rat snake     DPNDLFQNKFRKEQLEHKNSKWDTVLKAVKSKCA
Python        DAADTFQQEFRKERSERINNKWDTVLKVEKSE
Boa           DLSDLSQANFRREQSDRNNSKWDTVLKVVKTK
```

Supplementary Figure 1 (continued)

	49	179	184	198	228	330	342	391	434	746	828	829	913	916	946	1017	1028	1062	1074
zebrafish a	Q	L	H	C	E	L	D	Q	S	N	S	S	V	M	L	F	R	N	S
zebrafish b	S	L	C	C	E	L	D	Q	S	N	A	T	V	L	L	V	R	N	T
frog	C	Q	H	C	E	L	D	Q	S	N	S	G	L	I	L	N	V	D	V
rattlesnake	S	L	P	C	E	M	D	H	T	S	V	H	L	V	L	S	I	D	I
rat snake	C	V	S	A	V	L	N	Q	S	H	I	R	I	N	M	T	T	N	M
python	S	L	P	C	E	M	D	H	T	S	V	H	L	V	L	S	I	D	I
Boa	S	L	P	C	E	M	D	H	T	S	V	H	L	V	L	S	I	D	I
chicken	C	V	P	C	E	L	D	Q	S	H	A	H	L	I	L				
zebra finch	C	V	P	C	E	L	D	Q	S	H	A	H	L	I	L	S	S	D	V
mouse	M	S	S	C	E	L	N	Q	S	K	A	Y	L	I	M	R	T	E	I
rat	M	S	S	C	E	L	N	Q	S	K	A	Y	L	I	M	R	T	E	I
macaque	S	I	S	C	E	L	N	Q	S	N	A	H	L	I	I	H	I	E	I
human	S	I	S	C	E	L	N	Q	S	N	A	H	L	I	I	H	I	E	I
cow	K	I	S	C	E	L	D	Q	S	S	A	Y	L	I	M	G	T	E	I
Pig	A	I	S	C	E	L	D	Q	S	S	A	Y	L	I	M	G	T	D	I
Horse	T	I	S	C	E	L	N	Q	S	N	A	H	L	I	M	R	A	E	I
dog	A	I	S	C	E	L	N	Q	S	S	A	Y	L	I	M	R	A	E	I
opposum	K	L	S	C	E	L	D	Q	S	N	N	D	L	I	L	N	T	E	I
mosquito	E	L	E	R	E	I	N	N	S	Q	F	S	L	M	I	N	D	Q	K
fruit fly	D	L	E	R	E	I	N	N	S	L	N	T	L	L	I	N	D	H	K

Supplementary Figure 2: Amino acids at the 19 critical residues that are common to the three infrared-sensitive snake TRPA1 receptors. The numerical column headings specify the amino acid positions of the rattlesnake receptor. The blank residues of the chicken receptor indicate missing amino acids.