

The Science Circle

"The future of education"



"The Dragon's Smile"

By
Dr. Mary Anne Clark

Ph.D. Biology, teaching in the Biology
Department at Texas Wesleyan
University. Creator of the Genome
Island in Second Life.

OCT 17
10:00 am PDT



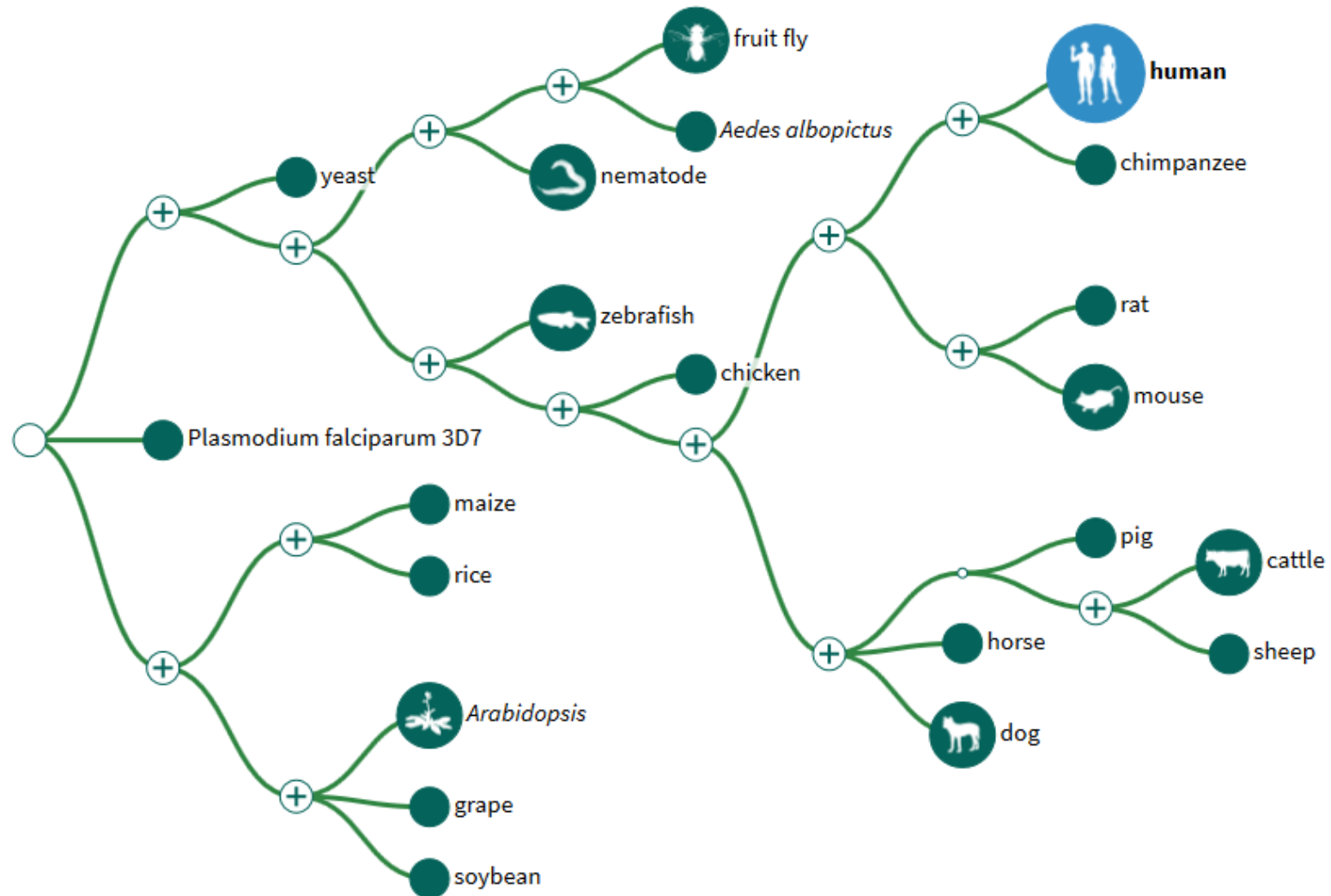
sciencecircle.org

Snooping in The Genetics Databases

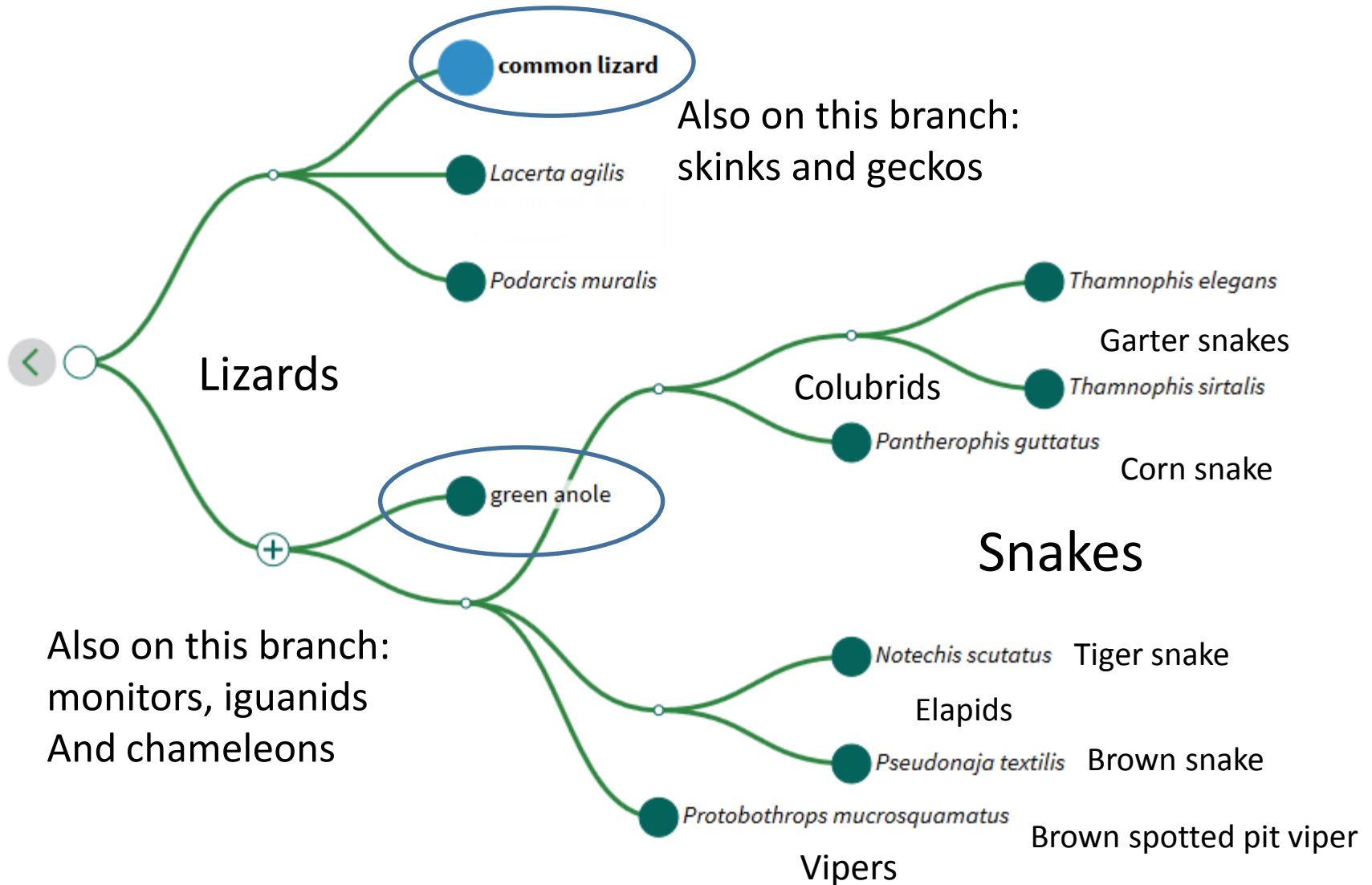
- New genomes are being sequenced all the time.
- Sequencing occurs in stages:
 - Shotgun sequencing of genome fragments
 - Assembly of fragments into contigs
 - Assembly of contigs into scaffolds.
 - Assembly and identification of chromosomes
- Sequences are usually published at the scaffold stage.
- Annotation takes years.

NCBI Genome Data Viewer

Some Annotated Genomes



Lizards and Snakes



Podarcis muralis (Common wall lizard) genome

Search in genome



Examples: [ATP6](#), [chrMT:6000-10000](#), [DNA repair](#)

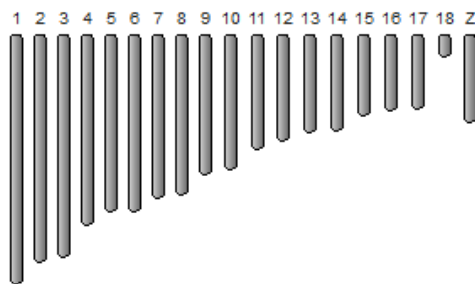
[Browse genome](#)[BLAST genome](#)

Assembly details

Name	PodMur_1.0
RefSeq accession	GCF_004329235.1
GenBank accession	GCA_004329235.1
Download via FTP	RefSeq , GenBank
Submitter	CIBIO-InBIO
Level	Chromosome
Category	Representative genome

Annotation details

Annotation Release	100
Release date	2019-04-21



Anolis carolinensis (green anole) genome

Search in genome



Examples: [rho](#), [DNA repair](#)

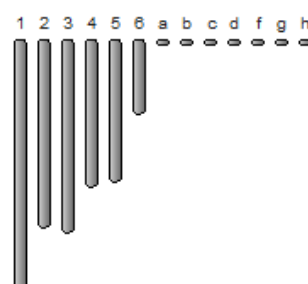
[Browse genome](#)[BLAST genome](#)

Assembly details

Name	AnoCar2.0
RefSeq accession	GCF_000090745.1
GenBank accession	GCA_000090745.2
Download via FTP	RefSeq , GenBank
Submitter	Broad Institute
Level	Chromosome
Category	Representative genome

Annotation details

Annotation Release	102
Release date	2016-05-31



My snoopee: *Varanus komodoensis*



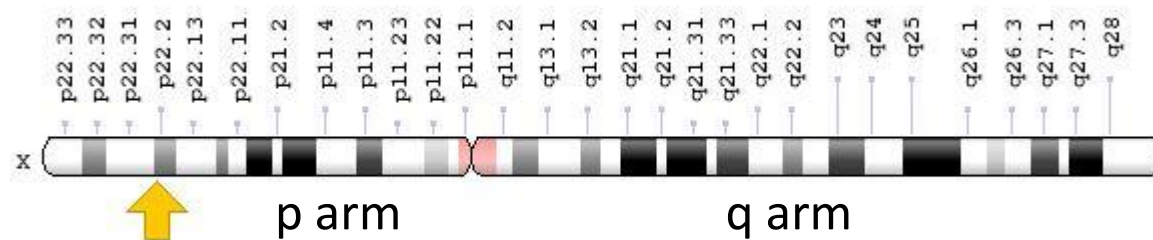
The genome of the komodo monitor was reported in 2019 (Lind, et al. Cardiovascular Research Institute, UCSF). It is still being annotated. The genome contains about 1.5 Gb of DNA.

The target gene: Amelogenin

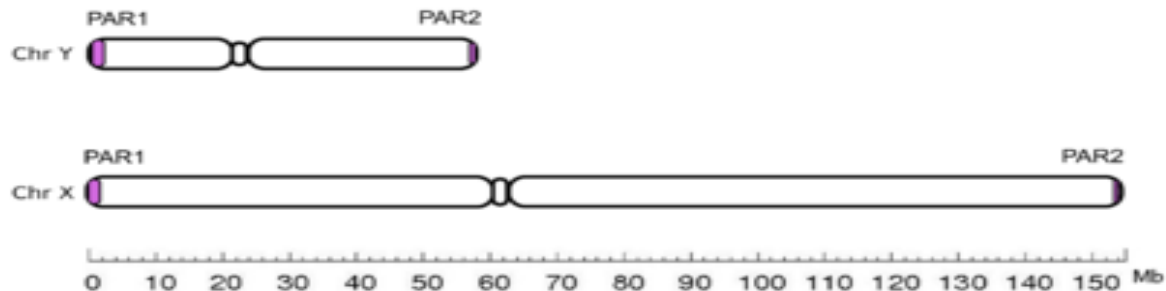
- Vertebrate teeth have two layers: an inner bony layer (dentine), and a hard outer enamel layer.
- Amelogenin is a protein involved with the mineralization of tooth enamel.
- Amelogenin is synthesized by ameloblasts, a layer of cells that lies between the dentine and enamel layers.
- The amino acid sequence of amelogenin is highly conserved and high in the amino acids proline, glutamine, leucine and histidine.
- In mammals, the amelogenin gene is located on the X chromosome, within the first intron of another unrelated gene, ARHGAP6 (Rho GTPase activating protein-6)

Human Amelogenin

Chromosome location of human amelogenin gene: Xp22.2



AmelX is located in the pseudoautosomal region (PAR1) of the X chromosome, with its homolog AmelY located in the PAR1 region of the Y chromosome. The gene has 7 exons.



Human Amelogenin Protein Sequence

MGTWILFACL LGAAFAMPLP PHPGHPGYIN FSYEVLTPK WYQSIRPPYP
SYGYEPMGGW LHHQIIPVLS QQHPPTHTLQ PHHHIPVVPA QQPVIPQQPM
MPVPGQHSMT PIQHHQPNLP PPAQQPYQPQ PVQPQPHQPM QPQPPVHPMQ
PLPPQPPLPP MFPMQPLPPM LPDLTLEAWP STDKTKREEV D

P = proline 49

L = leucine 16

H = histidine 14

Q = glutamine 24

191 amino acids, with 4
amino acids at high frequency.

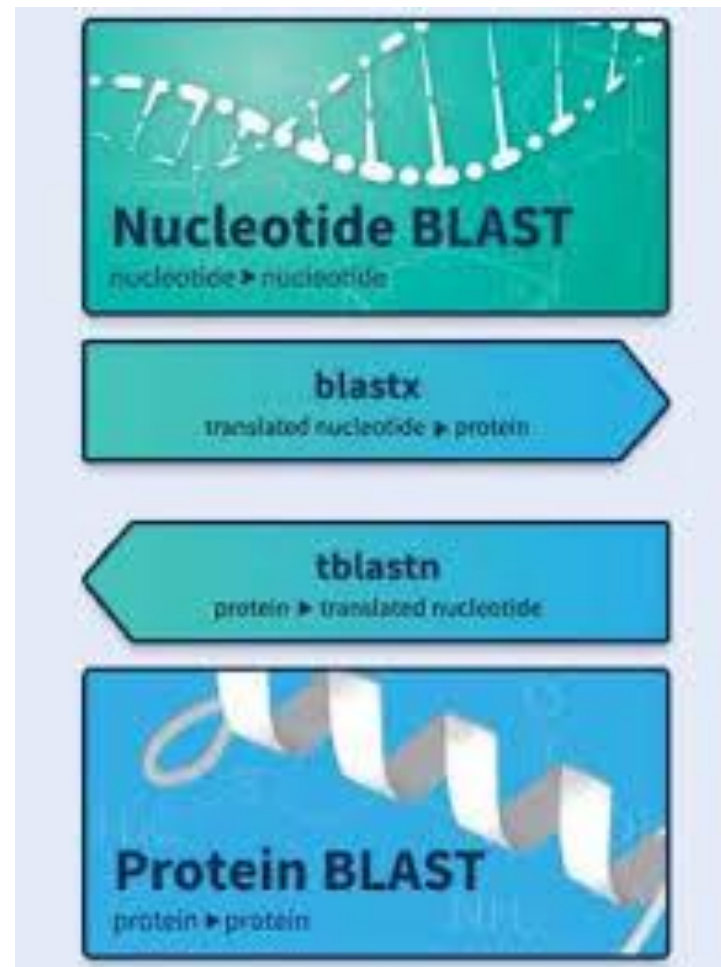
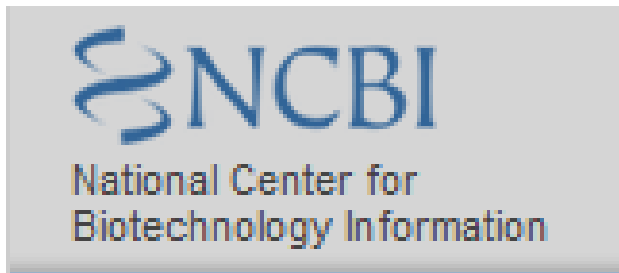
How to find a gene

- How do you find a gene in an unannotated species?
- Genes are older than species.
- Both coding sequences and exon breaks are conserved across species.
- To find a gene in a newly sequenced genome, start with genes that you know about in the genomes of related organisms.
- A former student (Tuan Nguyen) did a similar search for amelogenin in the Anole genome, which was the first reptilian genome to be sequenced (2011). He started with the human sequence.

Amelogenin in the Anole

- The genome of the Carolina anole (*Anolis carolinensis*) was the first reptilian genome to be fully sequenced: 2011 at the Broad Institute. The genome has 1.8 Gb.
- Anoles are similar to mammals in having X and Y chromosomes, with the male as the heterogametic sex.
- The amelogenin gene in the anole is on anole chromosome 3.
- Anole chromosome 3 also contains other markers found in the pseudoautosomal region of the human X chromosome.
- The anole X chromosome is therefore derived from a different ancestral chromosome than the human X.

Databases and Search Tools



Basic Local Alignment Search Tool
(NCBI)

Locating the Komodo Amelogenin

- The anole mRNA was used as the query sequence to search for the Komodo amelogenin DNA sequence.
- The database searched was the *Varanus komodoensis* whole-genome shotgun scaffolds from Lind et al. 2019.
- Scaffolds are segments composed of overlapping sequence not identified with specific chromosomes.
- Matches to anole exons 3-7 were identified in a single scaffold: SLA01.
- The Komodo amelogenin coding sequence was identified in sequences corresponding to anole exons 3-7.
- The mRNA was translated using Ensembl Transeq.

Anole Amelogenin mRNA

The anole amelogenin gene has 7 exons. Exons are numbered and marked by different colors in the sequence. The start codon in exon 3 and the stop codon in exon 7 are circled.

```

1      1  ttgcatttca cctaaatatt catattcaaa gttttgatct tttcctggat tcagccactg
      61  ccgtgtatac ccaggtttta gcaatggttg gaacgtatct gctaagctat ggggtccatgg
     121  accataataa agttatatat taatagcttg gaggcacaaa atagttgcat cgtggtctcc
2      181  tctccacagt ctgggaacag cctccagggg caggcactga gaaggtaaca agaataccag
3      241  aaagcttcta agaggacctg caatgaggaa ATGgaaggc tggactttgg ttatgtgcct
4      301  tttaagtacc acatttgcta ttccattgcc acaacatcct ggttatatca atttcagtta
5      361  tgaggtaaatg acacctttaa aatggtacca gagcctaata ggacaccagt atccacgtta
6      421  tggctttgaa ccaatgggag gatggatgca ccatgctgca ggaccaacaa tgcaccaaac
     481  taccttccaa agccatcctt cagtgcactc aactctccat caaatgcagc caccacatcc
     541  agcactgaac ccacacatgc agccacctgg gcacaaccca tttgggccaa tgcctgggtca
     601  aaatacattg atgccccagt ttcaaccagc tcatggagga ccaattcatc atccattcca
     661  accacatgca ggagaacacc caatgcaccc tcagcaaaca gggaatccag tccacccaat
     721  gcatcctcag caaccagcaa atccaaattc accaatttac ccagttcagc aattgccacc
     781  actgatatca gatacacctc ttgaatcatg gccaccagct gacaagacca agcaagaaga
     841  agtggacTAA aagacccaag aaaccagaat aatgcaaaga ataatcaggc cttgtatttt
7      901  gctttcaaga acccatactt ttacttcagt gagtgatata tggttccatt aaccacaggt
     961  tttagtaaat gttgtaacta gatggcaaac taatgtgaag tagcaaataa taaaatatgt
    1021  tttgatcaat a (polyadenylation)
  
```

Note the location of the stop codon

BLAST Search for Komodo Amelogenin

Varanus komodoensis isolate **SLA01 scaffold1**, whole genome shotgun sequence

Sequence ID: SJPD01000001.1 Length: 138,280,312 (= Hs Chr9) (Hs X = 153 Mb)

Number of Matches: 10 Match #1 (= Anole Exon 6) Identities 340/438(78%)

Ac Query	408	AGTATCCACGTTATGGCTTTGAACCAATGGGAGGATGGATGCACCATGCTGCAGGACCAA	467
Vk sbjct	79060653	AGTATCCACGCTATGGCTATGAGCCCATGGGAGGATGGATGCACCACACTTCAGGGCCAA	79060594
Query	468	CAATGCACCAAACCTACCTTCCAAAGCCATCCTTCAGTGCACCTCAACTCTCCATCAAATGC	527
sbjct	79060593	TGCTCCACCAAACCTCATTTTCAAGGGCTCCATCCAATGCATCCACCTCTCCACCAAATGC	79060534
Query	528	AGCCACCACATCCAGCACTGAACCCACACATGCAGCCACCTGGGCACAACCCATTTGGGC	587
sbjct	79060533	AGCCACAGCAGCCATCTCTGAATCCACAGATGCAGCAACCTGGGCATAGCCCATTCGTGC	79060474
Query	588	CAATGCCTGGTCAAATAACATTGATGCCCCAGTTTCAACCAGCTCATGGAGGACCAATTC	647
sbjct	79060473	CAATGACTGGGCAAAGCAACGTAATACCACACTATCAACCGGCTCATGCAGGACCAGCCC	79060414
Query	648	ATCATCCATTCCAACCACATGCAGGAGAACACCCAATGCACCCTCAGCAAACAGGGAATC	707
sbjct	79060413	AACATCCCCTCCCACCACATGCAGGAGAGCACCCAATGCAGCCCCCGCAACCAGGGAACC	79060354
Query	708	CAGTCCACCCAATGCATCCTCAGCAACCAGCAAATCCAAATTCACCAATTTACCCAGTTC	767
sbjct	79060353	CAAATCATCCCATTATATCCCAGCAACCAGGGAATCCCAGCCAGCCAATGTATCCAATTC	79060294
Query	768	AGCAATTGCCACCACTGATATCAGATACACCTCTTGAATCATGGCCACCAGCTGACAAGA	827
sbjct	79060293	AGCCACTGCCTCCATTAATCCCAGATACACCTCTTGAATCGTGG---CCAGTTGACAAGA	79060237
Query	828	CCAAGCAAGAAGAAGTGG	845
sbjct	79060236	CTAAGCAAGAAGAAGTGG	79060219

Identifying the exons

- Exons are conserved across species
- All 5 coding exons were matched in the scaffold
- Introns begin with GT and end with AG

Range 4: 79062004 to 79062050 [GenBankGraphics](#) Next Match Previous Match
Alignment statistics for match #4: **ANOLE Exon 5**
Score Expect Identities Gaps Strand
72.5 bits(79) 9e-10 44/47(94%) 0/47(0%) Plus/Minus
Query 363 AGGTAATGACACCTTTAAATGGTACCAGAGCCTAATAGGACACCAG 409
|||||
Sbjct 79062050 AGGTAATGACACCTTTAAATGGTACCAGAGCCTAATTGGGCATCAG 79062004

Range 5: 79063568 to 79063609 [GenBankGraphics](#) Next Match Previous Match
Alignment statistics for match #5: **ANOLE Exon 4**
Score Expect Identities Gaps Strand
72.5 bits(79) 9e-10 41/42(98%) 0/42(0%) Plus/Minus
Query 326 TTGCCACAACATCCTGGTTATATCAATTTCAGTTATGAGGTAA 368
|||||
Sbjct 79063609 NTGCCACAGCATCCTGGTTATATCAATTTCAGTTATGAGGTAA 79063568

364
↓

Komodo Amelogenin Coding Sequence: mRNA exons 3-7.

Two possible start codons in Exon 3

GACGTGTG**ATG**AGAAAA**ATG**GAGATCTGGACTTTGGTTATGTGCCTCTTA
AGTATAACATGTGCTATTCCACTGCCACAGCATCCTGGTTATATCAATTT
CAGTTATGAGGTAATGACACCTTTAAAATGGTACCAGAGCCTAATTGGGC
ATCAGTATCCACGCTATGGCTATGAGCCCATGGGAGGATGGATGCACCAC
ACTTCAGGGCCAATGCTCCACCAAACTCATTTTCAAGGGCTCCATCCAAT
GCATCCACCTCTCCACCAAATGCAGCCACAGCAGCCATCTCTGAATCCAC
AGATGCAGCAACCTGGGCATAGCCCATTCGTGCCAATGACTGGGCAAAGC
AACGTAATACCACACTATCAACCGGCTCATGCAGGACCAGCCCAACATCC
CCTCCCACCACATGCAGGAGAGCACCCAATGCAGCCCCCGCAACCAGGGA
ACCCAAATCATCCCATTTATCCCCAGCAACCAGGGAATCCCAGCCAGCCA
ATGTATCCAATTCAGCCACTGCCTCCATTAATCCCAGATACACCTCTTGA
ATCGTGGCCAGTTGACAAGACTAAGCAAGAAGAAGTGGAT**TAA**AGGAGTC
AAGAAACCAGAAAAAGGAGGAAAAAATAGGCACTGCTTCTTGCTTTCAAG
AACTCATCCTCCTACTGCAGTCAGAGATGCATGTTTTTCATTAAACCATGT
TCTTAGAGAAAGCTATAGCTAAAAGGCAAACAAATACAATGTAGCAAACA
ATAAATATGTTTTAATCAATA

Translation of Komodo Amelogenin mRNA



Probably the real start position

MRKMEIWTLV MCLLSITCAI PLPQHPGYIN FSYEVMTPLK
WYQSLIGHQY PRYGYEPMGG WMHHTSGPML HQTHFQGLHP
MHPPLHQMQP QQPSLNPQMQ QPGHSPFVPM TGQSNVIPHY
QPAHAGPAQH PLPPHAGEHP MQPPQPGNPN HPIYPQQPGN
PSQPMYPIQP LPPLIPDTPL ESWPVDKTKQ EEVD

P = proline 39

L = leucine 14

H = histidine 16

Q = glutamine 22

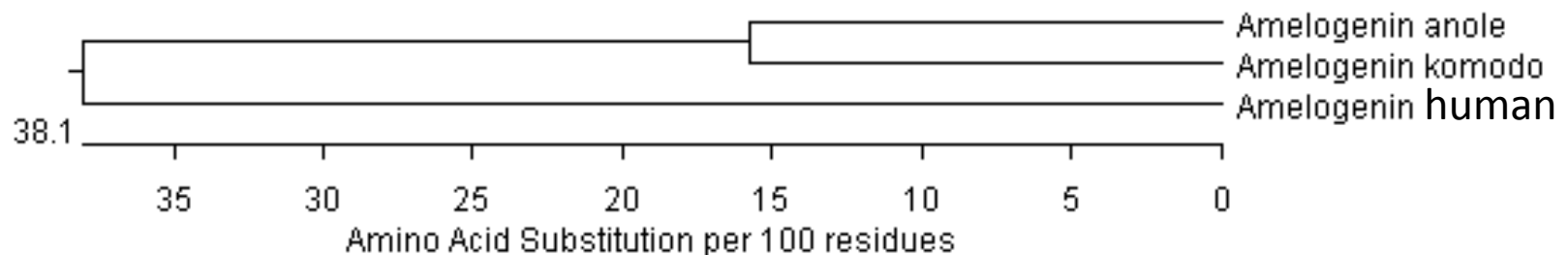
194 amino acids, with 4
amino acids at high frequency.

Sequence comparisons

- Sequences of the amelogenin proteins were compared using DNASTAR Lasergene MegAlign14.
- Sequences of anole, monitor and human were aligned and compared
- The alignment and associated phylogeny are seen on the following slide.
- In the next slide the phylogeny also includes the horse sequence as a second reference mammal.

Comparison of Anole, Human and Komodo Amelogenins

Majority	- - - MEXWLVLMCLLSXTFAI PLPQHP- - - GYI NF SYEVM TPLKWY QSLI GHQYPRYGYEPMGGWMMHHXXGPXLHQT HF QX	
	10 20 30 40 50 60 70 80	
Amelogenin komodo	MRKMEI WTLVLMCLLSI TCAI PLPQHP- - - GYI NF SYEVM TPLKWY QSLI GHQYPRYGYEPMGGWMMHHTSGPMLHQT HF QG	77
Amelogenin anole	- - - MEGWTLVLMCLLSITFAI PLPQHP- - - GYI NF SYEVM TPLKWY QSLI GHQYPRYGYEPMGGWMMHHAAGPTMHQTTF QS	74
Amelogenin	- - - MGTWMLFACLLGAFAFAMPLPPHPGHPGYI NF SYEVL TPLKWY QSI R- PPYP SYGYEPMGGWMLHHQI I PVLSQQQHP-	75
Majority	XHXXHPXLHQMQPXXPALNPXMQQPGHXPFXPMPGQXXXXP- - HXQPAHXGPAQHPXQPHAG- EHPMQPXQPGNPVHPMX	
	90 100 110 120 130 140 150 160	
Amelogenin komodo	LHPMHPLHLHQMQPQQPSLNPQMQQPGHSPFVPMTGQSNVI P- - HYQPAHAGGPAQHPLPPHAG- EHPMQPPQPGNPNHPI Y	154
Amelogenin anole	HPSVHSTLHQMQPPHPALNPQMPPGHNPFGPMPPGQNTLMP- - QFQPAHGGPI HHPFQPHAG- EHPMHPQQTGNPVHPMH	151
Amelogenin	THTLQP- - HHHI PVVPAQQPVI PQ- - - QPMMPVPGQHSMTPI QHHQPNLPPPAQQPYQPQPVQPQPHQPMQPQPPVHPMQ	150
Majority	PQQPXNPXXPMYPXQPLPPLI PDTPLESWPXXDKTKQEEVD	
	170 180 190 200	
Amelogenin komodo	PQQPGNPSQPMYPI QPLPPLI PDTPLESWP- VDKTKQEEVD	194
Amelogenin anole	PQQPANPNSPI YPVQQLPPLI SDTPLESWPADTKTKQEEVD	192
Amelogenin	PLPPQPLPMPFPMQPLPMLPDLTLEAWPSTDKTKEEVD	191



Amelogenins from

Monitor, Anole, Human, Horse

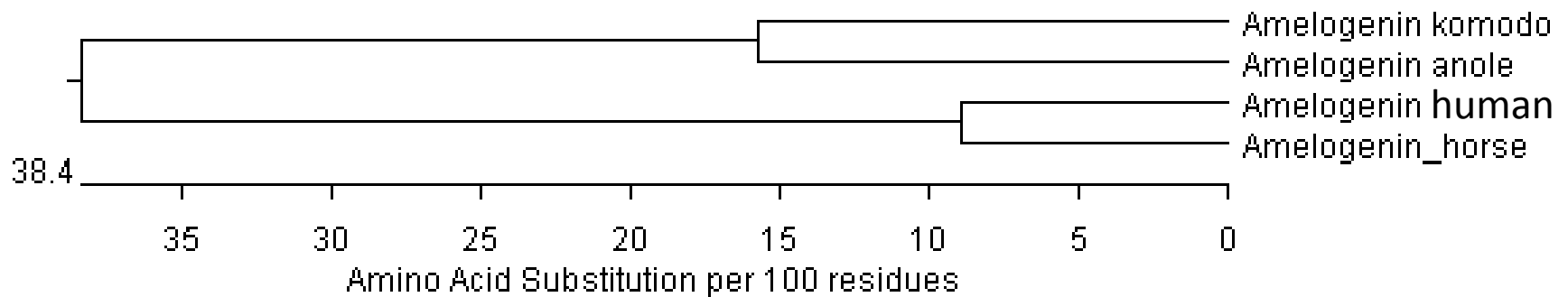
MEIWTLMCLLSITCAIPLPQHFGYINFSYEVMTPWKYQSLIGHQYPRY
GYEPMGGWMHHTSGPMLHQTHFQGLHPMHPPLHQMPPQPSLNPQMPPG
HSPFVPMTGQSNVIPHYPAHAGPAQHPLPPHAGEHPMQPPQGNPNHPI
YPQQPGNPSQPMYPIQPLPPLIPDTPLESWPVDKTKQEEVD

MEGWTLMCLLSTTFAIPLPQHFGYINFSYEVMTPWKYQSLIGHQYPRY
GFEPMGGMHHAAGPTMHQTTFQSHPSVHSTLHQMPPHPALNPHMQPPG
HNPFGPMPGQNTLMPQFQPAHGGPIHHPFQPHAGEHPMHPQQTGNPVHPM
HPQQPANPNSPIYPVQQLPPLISDTPLESWPPADKTKQEEVD

MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYEVLTPLWKYQSIRPPYP
SYGYEPMGGWLHHQIIPVLSQQHPPTHTLQPHHHIPVVPAQQPVIPQQPM
MPVPGQHSMTPIQHHQPNLPPPAQQPYQPQPVPQPHQPMQPQPPVHPMQ
PLPPQPPLPPMFPMQPLPPMLPDLTLEAWPSTDKTKREEVD

MGTWILFASLLGAAFAMPLPPHPGHPGYINFSYEVLTPLWKYQSLIRQPYT
SYGYEPMGGWLHHQIIPVLSQQNPSNHALQPHHHIPMVSAQHPVVPQQPM
MPLPGQHSMVPTQHHQPNLPPPVQQPFHPQPVPQPHQPIQPQPPLHPIQ
PLPPQPPLPPIFPLQPLPPMLPDLPLEAWPATDKTKREEVD

Amelogenins in 4 species



Note that komodo and anole are more distantly related to each other than are human and horse.

Sequence comparisons

- The komodo sequence differs from the anole sequence by about 15 amino acids/100.
- The human and horse sequences differ by about 8 amino acids/100.
- The lizard and mammalian sequences differ by about 38 amino acids/100.
- The Komodo monitor scaffold annotation has been recently updated. How does it compare?

AmelX from *V. komodoensis* annotation

- /translation="MEFTPQRGPPGTVITGFSAPVTRKEIAFIKKSEDMAKFSPKALP
TKPSLHYTMERCVIERKNTERERGGGAISHFAVTRSGVSNRPAGQLPGQAGHGGGALQ
PSKCWWRLGQLQDEWPTTTAVVEYLMRPSLTQTLLPAAPILLIIFKTVNQTPVMTVL
EDSSEEKEWEVPKMAANEMSLLGIGEACDDLAQHFKEKIVQIRRDLDSTVDTVVPVRE
VSRPPSGPSLMDEFQLLQPDDVDKVLGMLRPYLDEHSLATVIHALVTSCLDYCNALYV
GLLLKTVWILQLVQNRAARLLTGTSHYSQIMPVLYQLHWLPVEVQAQFKVLVLTYSKAL
NGLGPGYLKERRVMRK
- MEIWTLMVCL LSITCAIPLP QHPGYINFSY EVMTPLKWYQ SLIGHQYPRY
GYEPMGGWMH HTSGPMLHQT HFQGLHPMHP PLHQMPPQP SLNPQMPPQP
HSPFVPMTGQ SNVIPHYQPA HAGPAQHPLP PHAGEHPMQP PQPGNPNHPI
YPQQPGNPSQ PMYPIQPLPP LIPDTPLESW PVDKTKQEEV (D)
- VREHSCNLILLFYNSLLVHDMQIAME
QFSIEQCILIVKTHYKNGESYAETVCKLGMIFGCDNAPTTTTVSRLVQKSEEPGSIAT
RKLPSRNCTGKLKTVFINCGGKSRLKESARQQRTVITAAGTDDDDDDDDQDIKQTKH
IMGKAGLDESPVGIKIAGRNINNLRYADDTTLMAESEQELKSLLMRVKEESTKVGLKL
NIKKTKIMASGPLTSWQIDGEEMEVVTDIFLGSKITTDGDCSQEIKRHLLGRNAMA
NLDSIIKSRDITLPTKVCIVKAMVFPVAMYGCESWTIRNAECRRNEAFELWCWQRLLR
VPWTTRRSNRSVLEEINPDCSLEGQILKMMLKYFGHLM"
- I think the database annotation for exon 7 may be incorrect

Is Amelogenin Embedded in Monitors?

Anole ARHGAP6 mRNA used as query sequence in BLAST:

Varanus komodoensis isolate [SLA01 scaffold1](#), whole genome shotgun sequence

Sequence ID: [SJPD01000001.1](#) Length: 138280312 Number of Matches: 18

Range 3: 79135088 to 79135347 [GenBankGraphics](#) Next Match Previous Match [First Match](#)

Alignment statistics for match #3 **Exon 4**

ARHGAP6 gene has 13 exons

	Score	Expect	Identities	Gaps	Strand	
	307 bits(340)	2e-80	224/260(86%)	0/260(0%)	Plus/Plus	
AC Query	437		AGAATTGTACCCAGGCATTTGGCATGCCTTTATCACAGGTAATAGCTAATGATCGGGC			496
Vk Sbjct	79135088		AGAATTGTTCCTCCAGGCTTTTGGAAATGCCCCTATCACAAAGTAATAGCGAATGACCGGGC			79135147
Query	497		ATACAAACTGAAGCAGGACATTCAGAAGGAAGAGCAGAGGGATGTTTCAGAATTTGTAGC			556
Sbjct	79135148		CTACAAACTGAAGCAGGATACGCAGAAGGAAGAACAGAGGGATGTTTCAGAATTTGTGGC			79135207
Query	557		CTCCCTTCTACCATTTGGGAGCAAAAGACAAAACAAGGAAGCTCTCAAGCAGTAACTCATC			616
Sbjct	79135208		TTCCCTTTTACCATTTGGGAATAAAAGGCAGAACAAAGAACTCTCAAGTAGTAACTCATC			79135267
Query	617		TCTTAGCTCCACCTCGGAAACACCAAATGAATCCACTTCACCAAATACTCCAGAACCAGC			676
Sbjct	79135268		TCTTAGCTCCACCTCAGAAACGCCCAACGAATCTACTACACCAAACACTCCTGAGGCTGC			79135327
Query	677		TCCACGGGCCCGGAGGAGAG			696
Sbjct	79135328		TCCACGGACAAGGAGGAGAG			79135347

ARHGAP 6

Exon 1: 79016994 – 79017092

Amelogenin

↑ Exon 1 + 2: unidentified
Exon 3: 79064248-79064318
Exon 4: 79063572-79063610
Exon 5: 79062004-79062048
Exon 6: 79060219-79060651
Exon 7: 79059063-79059246

Exon 2: 79099391 - 79099550

Exon 3: 79126580 - 79126651

Exon 4: 79135090 - 79135346

Exon 5: 79140542 - 79140735

Exon 6: 79142610 - 79142665

Exon 7: 79143647 - 79143797

Exon 8: 79144686 - 79144832

Exon 9: 79146640 - 79146819

Exon 10: 79151109 - 79151206

Exon 11: 79153968 - 79154230

Exon 12: 79156363 - 79156443

Exon 13: 79158693 - 79159984

ARHGAP6 and Amelogenin Exon Locations in Monitors

ARHGAP 6

Exon 1: 79016994 – 79017092

Amelogenin

Exon 1 + 2: unidentified

Exon 3: 79064248-79064318

Exon 4: 79063572-79063610

Exon 5: 79062004-79062048

Exon 6: 79060219-79060651

Exon 7: 79059063-79059246

Exon 2: 79099391 - 79099550

Exon 3: 79126580 - 79126651

Exon 4: 79135090 - 79135346

Exon 5: 79140542 - 79140735

Exon 6: 79142610 - 79142665

Exon 7: 79143647 - 79143797

Exon 8: 79144686 - 79144832

Exon 9: 79146640 - 79146819

Exon 10: 79151109 - 79151206

Exon 11: 79153968 - 79154230

Exon 12: 79156363 - 79156443

Exon 13: 79158693 - 79159984

ARHGAP6 and Amelogenin in Annotated Sequence

Annotated **Varanus amelogenin** sequence
identifies exons 1 and 2 here:

79071641..79071819

79070726..79071020

Exons 3-6 match mine, but Exon 7 is
79055871..79056083

Annotated **Varanus ARHGAP6** sequence
has 14 exons with first 3 exons here:

Exon 1: 79119403..79119411

Exon 2: 79120495..79120570

Exon 3: 79123378..79123431

Exon 4 = my Exon 3, etc.

ARHGAP6 Protein Sequence: Mine

Exon 1 = Green

Exon 2 = Blue

Exon 3+ = Black

MLQRRGDLGGCPPALGDQVPLNAKPSLYYAGDFTWNSMSGHSVRLKSVPV
QSLSELERARLQEVAFYQLQQECDLGCQITIPKDGQKRKKSLRKKLDSL
GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVSEFVASLL
PFGNKRQNKELSSSNSSLSTSETPNESTTPNTPEAAPRTRRRGAMSVDS
ITDLDDSQSRLLEALQLSLPAEAQNKKEKSRDKKLSLNPIYRQVPRLVDS
CCQHLEKHGLQTVGIFRVGSSKKRVRQLREEFDRGVDVILDEEHSIHDVA
ALLKEFLRDMPDPLLTRELYTPFINTLLLEPHEQLSTLQLLIYLLPPCNC
DTLHRLQLQFLATVASHAEDRIDQDSQEIPGNKMTSLNLATIFGPNLLHKQ
KTTDKEFTVQSTARAEESTAIIVVQKMIENYETLFMVPADLQNEVLISL
LETDPDVVDYLLRRKASQSSSPEMLQSEVSFSMEGRHSSIDSNRVSSGDI
SPYDNNSPVLSERSLLAMQEDTSLNSEKLFKVPEQCTLVSSFQPKSRENS
LGSWFGKDVSEDYFDIWGTWHSTLKSGSKDREMTGSYGDIYESSSLRPGQ
CSLSQGNLSLPSPQWQESNTELDNGKQIIRRSQTTAALPEYRPHLPLSRV
CSTPQIAGRNRSSQSSEQQLPKSDTEHLVLRD TDVTCWHSSSKPWHLKQ
AEGLRKERPPPPYPDKARPGSTSPLQAAEHALWRPRRPDLRPGAAGARVM
TQVPEQPALCEEQQDHVEGEGPTADSEQNSNNLAEPDWHDWQRDRWQIWE
ILSPDNPDALPETLV

ARHGAP6 Sequence: Varanus annotation

Exon 1 = Green

Exon 2 = Blue

Exon 3 = Red

Exon 4+ = Black

MV**SR**AP**EE**GS**LP**Q**IS**R**GW**RTY**KEE**GAP**SG**NE**TA**VT**GY**L**PMS**FRA**EL**DGQK
RKKSLRKKLDSLKGKEKNKDKEFVPQAFGMPLSQVIANDRAYKLLKQDTQKE
EQRDVSEFVASLLPFGNKRQNKELSSSNSSLSTSTETPNESTTTPNTPEAA
PRTRRRGAMSVDSITDLDDSQSRLLEALQLSLPAEAQNKKEKSRDKKLSL
NPIYRQVPRLVDSCCQHLEKHGLQTVGIFRVGSSKKRVRQLREEFDRGVD
VILDEEHSHDVAALLKEFLRDMPDPLLTRELYTPFINTLLLEPHEQLST
LQLLIYLLPPCNCDTLHRLQLFLATVASHAEDRIDQDSQEIPGNKMTSLN
LATIFGPNLLHKQKTTDKEFTVQSTARAEESTAIIVVQKMIENYETLFM
VPADLQNEVLISLLETDPDVVDYLLRRKASQSSSPEMLQSEVSFSMEGRH
SSIDSNRVSSGDISPYDNNSPVLSERSLLAMQEDTSLNSEKLFKVPEQCT
LVSSFQPKSRENSLGSWFGKDVSEDYFDIWGTWHSTLKSGSKDREMTGSY
GDIYESSSLRPGQCSLSQGNLSLPSPQWQESNTELDNGKQIIRRSQTAA
LPEYRPHLPLSRVCSTPQIAGRRNSSQSSEQQLPKSDTEHLVLRD TDVT
CWHSSSKPWLKQAEGLRKERPPPPYPDKARPGSTSPLQAAEHALWRPRR
PDLRPGAAGARVMTQVPEQPALCEEQQDHVEGEGPTADSEQNSNNLAEPD
WHDWQRDRWQIWEILSPDNPDALPETLV

Matches with my ARHGAP6

Rho GTPase-activating protein 6 [Varanus komodoensis]

Sequence ID: [KAF7253702.1](#) Length: 778 Number of Matches: 1

Range 1: 47 to 778 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1499 bits(3882)	0.0	Compositional matrix adjust.	732/732(100%)	732/732(100%)	0/732(0%)

Query	84	DGQKRKSLRKKLDSLGEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVS	143
Sbjct	47	DGQKRKSLRKKLDSLGEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVS	106

Query	144	EFVASLLPFGNKRQ	202
Sbjct	107	EFVASLLPFGNKRQ	

rho GTPase-activating protein 6 isoform X2 [Zootoca vivipara]

Sequence ID: [XP_034972544.1](#) Length: 826 Number of Matches: 1

Range 1: 1 to 826 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1353 bits(3503)	0.0	Compositional matrix adjust.	711/826(86%)	741/826(89%)	11/826(1%)

Query	1	MLQRRGDLGGCPPALGDQVPLNAKPSLYYAGDFTWNSMSGHVSRLKSVFPVQSLSELERAR	60
Sbjct	1	MLQRRGDLGG PPAL LN KP LYYAGDFTWNSMSG SVRLKSVFPVQSLSELERAR	60

PREDICTED: rho GTPase-activating protein 6 [Anolis carolinensis]

Sequence ID: [XP_008105415.2](#) Length: 818 Number of Matches: 1

Range 1: 1 to 818 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1350 bits(3494)	0.0	Compositional matrix adjust.	703/818(86%)	746/818(91%)	3/818(0%)

Query	1	MLQRRGDLGGCPPALGDQVPLNAKPSLYYAGDFTWNSMSGHVSRLKSVFPVQSLSELERAR	60
Sbjct	1	MLQRRGDLG PP+ G+ L A+ SLYYAGDFTWNSMSG SVRLKSVFPVQSLSELERAR	60
Query	61	LQEVAFYQLQQECDLGCQITIPKDGQKRKSLRKKLDSLGEKNKDKEFVPQAFGMPLSQ	120
Sbjct	61	LQEVAFYQLQQ+CDLGCQITIPKDGQKRKSLRKKLDSLGEKNKDKEFVPQAFGMPLSQ	120



Exon 3



Exon 1



Exon 1

Matches with Annotated Varanus ARHGAP6

rho GTPase-activating protein 6 isoform X1 [Zootoca vivipara]

Sequence ID: [XP_034972543.1](#) Length: 983 Number of Matches: 1

Range 1: 245 to 983 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1213 bits(3138)	0.0	Compositional matrix adjust.	639/739(86%)	667/739(90%)	7/739(0%)
Query 47	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVS				106
Sbjct 245	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDT +EEQRDVS				304
Query 107	PREDICTED: rho GTPase-activating protein 6 [Anolis carolinensis]				
Sbjct 305	Sequence ID: XP_008105415.2 Length: 818 Number of Matches: 1				

Exon 3



Range 1: 84 to 818 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Exon 3

Score	Expect	Method	Identities	Positives	Gaps
1209 bits(3129)	0.0	Compositional matrix adjust.	633/735(86%)	672/735(91%)	3/735(0%)
Query 47	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVS				106
Sbjct 84	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQD QKEEQRDVS				143
Query 107	rho GTPase-activating protein 6 isoform X1 [Lacerta agilis]				
Sbjct 144	Sequence ID: XP_033003306.1 Length: 982 Number of Matches: 1				



Range 1: 244 to 982 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1214 bits(3141)	0.0	Compositional matrix adjust.	636/739(86%)	669/739(90%)	7/739(0%)
Query 47	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDTQKEEQRDVS				106
Sbjct 244	DGQKRKKS LRKKLDSL GKEKNKDKEFVPQAFGMPLSQVIANDRAYKLKQDT +EEQRDVS				303
Query 107	EFVASLLPFGNKRQNKELSSSNSSLSSTSETPNESTTTPNTPEAAPRTRRRGAMSVDSITD				166
Sbjct 304	EFV SLLPFG+KRQNKELSSSNSSLSSTSETPNEST+PNTPE APR RRRGAMSVDSITD				363

Exon 3



References and Software

Anole genome. <https://www.broadinstitute.org/news/first-lizard-genome-sequenced>

Komodo genome. Lind,et al. 2019.
<https://www.nature.com/articles/s41559-019-0945-8>

Fincham and Simmer, 1997. <https://www.ncbi.nlm.nih.gov/pubmed/9189621>

Gasse and Sire, 2015. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4587831/>

Anole amelogenin transcript: <https://www.ncbi.nlm.nih.gov/nuccore/GAFN01001597>

Human amelogenin NCBI: <https://www.ncbi.nlm.nih.gov/gene/265>

Human amelogenin OMIM: <https://www.omim.org/entry/30039>

Genetics Home Reference: <https://ghr.nlm.nih.gov/gene/AMELX>

DNA and protein searches: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Protein translation: https://www.ebi.ac.uk/Tools/st/emboss_transeq/

DNASTAR Lasergene 14 (<https://dnastar.com>)