

UNIVERSITY OF ESWATINI
FACULTY OF SCIENCE AND ENGINEERING
DEPARTMENT OF BIOLOGICAL SCIENCES
SUPPLEMENTARY EXAMINATION PAPER 2019/2020

COURSE CODE: BIO451
TITLE OF PAPER: BIOINFORMATICS
TIME ALLOWED: THREE (3) HOURS
INSTRUCTIONS: NUMBERS IN BRACKETS DENOTE THE NUMBER OF MARKS
THIS PAPER COMPRISES OF TWO SECTIONS.
SECTION A COMPRISES OF SIX (6) QUESTIONS.
SECTION B COMPRISES OF SIX (6) QUESTIONS.
ANSWER ALL QUESTIONS IN SECTIONS A AND B.

TOTAL MARKS AVAILABLE: 70

DO NOT OPEN THIS PAPER UNTIL PERMISSION HAS BEEN GRANTED BY THE CHIEF INVIGILATOR

Section A

Total marks available: 40

Eswatini National Trust Commission has been doing a botanical survey of their national parks. They have collected a number of plants and have requested your help in identifying them. You have extracted the genomic DNA from one of the plants, in triplicate, (Sample 1) and sent each for sequencing.

The data is back, and the laboratory assistant has handed you a USB key. You load the USB key into your computer and see the following:

ENTC_Sample1a.abi

ENTC_Sample1b.abi

ENTC_Sample1c.abi

AQ1: Describe the steps you will take to help identify the plant. In your answer, also include the format of the data, the name(s) of the database(s) you would use and how you would select the appropriate result. [10]

You have performed the steps as described in AQ1 for Sample 1 (a, b and c). The laboratory assistant has handed you the results (Annex 1). Carefully study the three documents in Annex 1.

AQ2: Identify the scientific and common name of the plant collected by the team from ENTC. [3]

AQ3: Despite the DNA extraction occurring in triplicate, the results were different. What are the most likely reasons for the different results observed? [4]

AQ4: Using the data in Annex 1. Identify the following:

- a) The size and type of molecule. [2]
- b) The full name of the GenBank Division. [1]
- c) The number and name of the proteins encoded by this section of DNA. [4]
- d) The name and number of genes in this section of DNA. [4]
- e) The coding sequence for one of the genes in this section of DNA [2]

You have sent the identifications of all the plants to ENTC. A botanist from ENTC has contacted you and wants to do further analysis on the plant (Sample 1). She would like to carry out some morphological (physical) analysis against the plant found in the national park and the plant data your analysis identified.

AQ5: What piece(s) of information from Annex 1 will you need to provide to the botanist? [5]

The ENTC botanist has contact you again, requesting your assistance in determining the structure and function of one of the proteins.

AQ6: Describe how you would use bioinformatics to determine the structure and function of this protein, including which database(s) you would use. [5]

Section B

Total marks available: 30

BQ1: You are supervising an intern in your laboratory. Unfortunately he has got his results all mixed up and has asked you for help. Based on the sequences below, identify the **types of molecules** these sequences are from.

- a) AGAAGTTCCAGATCATTCCAAGA [1]
- b) AUCCAGGGUUCACCAGAGCCAGGAGA [1]
- c) LKAGTAKCLATVAPPACQUALEALOCCAGAPLAGNQSS [1]
- d) AGATCAGCCATGACTAG [1]
- e) SPALVGUPPSCQURPALAGUCWNMU [1]
- f) GGGCAUGGCUGAAACUACCAGUCAAGUAG [1]

BQ2: What causes a single strand of RNA to fold back on itself? [2]

BQ3: Describe three key differences between CDS and ORF in a nucleotide sequence. [3]

BQ4: Describe the difference between homology and similarity. [3]

BQ5: Describe when and why you would use the:

- a) Smith-Waterman algorithm [3]
- b) Needleman-Wunsch algorithms [3]

BQ6: Assuming a scoring method: match = +2, mismatch = -3, and gap = -2, using the Needleman-Wunsch algorithm:

- a) Initialize the scoring matrix below. [2]

| | | A | C | T | G | A | T | T | C | A |
|---|--|---|---|---|---|---|---|---|---|---|
| A | | | | | | | | | | |
| C | | | | | | | | | | |
| G | | | | | | | | | | |
| C | | | | | | | | | | |
| A | | | | | | | | | | |
| T | | | | | | | | | | |
| C | | | | | | | | | | |
| A | | | | | | | | | | |

- b) Calculate the score of the matrix and fill the traceback matrix. Use arrows to indicate how you determined the path forward. Use circles and arrows to determine your trace back. [5]
- c) Deduce the best alignment based on the traceback matrix. [3]

END OF EXAMINATION

Nucleotide

GenBank

Homo sapiens keratin 15 (KRT15), RefSeqGene on chromosome 17

NCBI Reference Sequence: NG_012284.1

FASTA [Graphics](#)

Go to:

LOCUS NG_012284 12274 bp DNA linear PRI 30-DEC-2018
 DEFINITION Homo sapiens keratin 15 (KRT15), RefSeqGene on chromosome 17.
 ACCESSION NG_012284
 VERSION NG_012284.1
 KEYWORDS RefSeq; RefSeqGene.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
 Catarrhini; Hominidae; Homo.
 COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The
 reference sequence was derived from [AC019349.28](#).
 This sequence is a reference standard in the [RefSeqGene](#) project.
 Summary: The protein encoded by this gene is a member of the
 keratin gene family. The keratins are intermediate filament
 proteins responsible for the structural integrity of epithelial
 cells and are subdivided into cytokeratins and hair keratins. Most
 of the type I cytokeratins consist of acidic proteins which are
 arranged in pairs of heterotypic keratin chains and are clustered
 in a region on chromosome 17q21.2. [provided by RefSeq, Jul 2008].
 PRIMARY REFSEQ_SPAN PRIMARY_IDENTIFIER PRIMARY_SPAN CONP
 1-9358 AC019349.28 74117-83474
 9359-9359 "C" 1-1
 9360-12274 AC019349.28 83476-86390
 FEATURES Location/Qualifiers

source 1..12274
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17q21.2"

gene <1..402
 /gene="KRT19"
 /gene_synonym="CK19; K19; K1CS"
 /note="keratin 19"
 /db_xref="GeneID:3880"
 /db_xref="HGNC:HGNC:6436"
 /db_xref="MIM:148020"

mRNA <22..402
 /gene="KRT19"
 /gene_synonym="CK19; K19; K1CS"
 /product="keratin 19"
 /inference="similar to RNA sequence, mRNA (same
 species):RefSeq:NM_002276.5"
 /exception="annotated by transcript or proteomic data"
 /transcript_id="NM_002276.5"
 /db_xref="GeneID:3880"
 /db_xref="HGNC:HGNC:6436"
 /db_xref="MIM:148020"

CDs <22..276
 /gene="KRT19"
 /gene_synonym="CK19; K19; K1CS"
 /inference="similar to AA sequence (same
 species):RefSeq:NP_002267.2"
 /exception="annotated by transcript or proteomic data"
 /note="keratin, type I, 40-kd; cytokeratin 19; 40-kDa
 keratin intermediate filament; CK-19; keratin 19, type 1"
 /codon_start=1
 /product="keratin, type I cytoskeletal 19"
 /protein_id="NP_002267.2"
 /db_xref="CCDS:CCDS1359.1"
 /db_xref="GeneID:3880"
 /db_xref="HGNC:HGNC:6436"
 /db_xref="MIM:148020"
 /translation="HTSVYRQSSATSSFGGLGGGSVAFGPGVAFRAPS IHGSGGGRR
 VSVSARFVSSSSGAYGGVGLTASDGLAGNKLTHQLNORLASYLDRVRALE
 ANGELVKIIRDAKQKQPGPSRDVSHYTTIQDIBDKILGATIEHSRIVLQIDMARL
 ADDDFIKFETEQLRMSVEADINGLRRVLDELTAARDLEHQIEGLKRELAVLKNKH
 EEFISTLRGQVGGQVSVEDVAPGDLAKLISDMASQVSVMAEQNRKDAEAMFTSRTE
 ELNREVASHTEQVNSRSEVTDLRRLTQLGLEIELQSLSKAALEDLAEAEAFRGAQ
 LAHTQALISGIEAQLGDWRADSERQNEQYQLPDIKRSLEQEIATYRSLLAEQEDHYN
 NLSASKVL"

STS 259..378
 /gene="KRT19"
 /gene_synonym="CK19; K19; K1CS"
 /standard_name="RH71340"
 /db_xref="UniSTS:89034"

gene 5001..10274
 /gene="KRT15"
 /gene_synonym="CK15; K15; K1CO"
 /note="keratin 15"
 /db_xref="GeneID:3866"
 /db_xref="HGNC:HGNC:6421"
 /db_xref="MIM:148030"

mRNA join(5001..5689,6054..6926,7055..7211,7754..7915,
 8000..8134,8327..8547,9345..9370,9899..10274)
 /gene="KRT15"
 /gene_synonym="CK15; K15; K1CO"
 /product="keratin 15"
 /transcript_id="NM_002275.3"
 /db_xref="GeneID:3866"
 /db_xref="HGNC:HGNC:6421"

```

/db_xref="MIM:148030"

exon      5001..5089
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=1

CDS       join(5192..5089,6054..6936,7055..7211,7754..7915,
8009..8134,8327..8547,9345..9370,9899..9996)
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/notes="keratin-15, beta; type I cytoskeletal 15;
cytokeratin 15; keratin-15, basic; CK-15; keratin 15, type
1"
/codon_start=1
/product="keratin, type I cytoskeletal 15"
/protein_id="IP_022465_2"
/db_xref="CCDS:CCDS11398.1"
/db_xref="GeneID:3565"
/db_xref="HGNC:HGNC:6421"
/db_xref="MIM:148030"
/translation="MTITFLQTSSTYFGGGSTRGSLLAGGGFGGGSLSGGGGSRSI
SASARFVSSGGGQGGHRVCGFGGAGSVFGGFGGGVGGFGGFGGGGGLLS
GNEIKTQHLNDRLASYLKVRALAEAHADLVKIHDAVQKQPTISPCDYSQFKTI
EELRDKIHATTIONSRVILEIDNARLAADDFRLKYENELARQVVEADINLRRVLDL
LTLARTDLEQIEGLNEELAVLKKHHEENKEFSSQLAGQVNVEMDAAPGVDLTRVLA
EMHQYEAQAEKRVROVEAHFVSKTEELNKEVASHTEHIQTSKTEITDLRRITQEELEI
ELQSLSHKAGLENSLAETECRYATLQQIQLIGGLEAQLSELRCENEAQHQYKHL
LDIXTRLEQETIATYRSLLEGQAKMAGIGIREASSGGGGSSSHFINVEVSDGQVVS
SHKREI"

miRNA     6802..6855
/gene="HNR6510"
/gene_synonym="hsa-mir-6510; mir-6510"
/notes="microRNA 6510"
/db_xref="GeneID:102466658"
/db_xref="HGNC:HGNC:50101"
/db_xref="miRBase:MI0022222"

miRNA     6802..6855
/gene="HNR6510"
/gene_synonym="hsa-mir-6510; mir-6510"
/product="microRNA 6510"
/transcript_id="HS_189765_1"
/db_xref="GeneID:102466658"
/db_xref="HGNC:HGNC:50101"
/db_xref="miRBase:MI0022222"

ncRNA     6804..6825
/ncRNA_class="miRNA"
/gene="HNR6510"
/gene_synonym="hsa-mir-6510; mir-6510"
/product="hsa-mir-6510-5p"
/db_xref="miRBase:MI00225476"
/db_xref="GeneID:102466658"
/db_xref="HGNC:HGNC:50101"
/db_xref="miRBase:MI0022222"

ncRNA     6833..6853
/ncRNA_class="miRNA"
/gene="HNR6510"
/gene_synonym="hsa-mir-6510; mir-6510"
/product="hsa-mir-6510-3p"
/db_xref="miRBase:MI00225477"
/db_xref="GeneID:102466658"
/db_xref="HGNC:HGNC:50101"
/db_xref="miRBase:MI0022222"

exon      6854..6936
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=2

exon      7055..7211
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=3

exon      7754..7915
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=4

STS       7837..8464
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/standard_name="Krt15"
/db_xref="UniSTS:474754"

exon      8009..8134
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=5

exon      8327..8547
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"
/inference="alignment:Splign:2.1.0"
/number=6

exon      9345..9370
/gene="KRT15"
/gene_synonym="CK15; K15; K1CO"

```

/inference=alignment:SpIign:2.1.0"
/number=7

nise_difference 9359
/ gene="KRT15"
/ gene_synonym="CK15; K15; K1CO"
/ note="This sequence differs from the reference genome assembly (NCBI Build 36) at this position. C was replaced by G to represent the standard allele identified by alignment with human cDNAs."
/ replace="c"

STS 9793..9983
/ gene="KRT15"
/ gene_synonym="CK15; K15; K1CO"
/ standard_name="SHGC-79685"
/ db_xref="UniSTS:184592"

STS 9870..10463
/ standard_name="KRT15_3016"
/ db_xref="UniSTS:467281"

EXON 9899..10274
/ gene="KRT15"
/ gene_synonym="CK15; K15; K1CO"
/ inference=alignment:SpIign:2.1.0"
/ number=8

STS 9999..10273
/ gene="KRT15"
/ gene_synonym="CK15; K15; K1CO"
/ standard_name="SHGC-12347"
/ db_xref="UniSTS:57368"

STS 10015..10140
/ gene="KRT15"
/ gene_synonym="CK15; K15; K1CO"
/ standard_name="RH17477"
/ db_xref="UniSTS:48736"

ORIGIN

1 accccctctt cctatctcta gaaagctgcc ttggaaagaca cactggcaga accggaggcc
51 ccctctggag ccagctggc gcatatccag gcctggtca cgggtatga agccagctg
121 ggcctctctt ggcctctctt gaaagctgcc ttggaaagaca cactggcaga accggaggcc
181 aagctggcgc tggagcaga gattgcacc taccagacc tctcaggg accggaagat
241 cactacaaca attgtctgc ctccaagtc ctctgagca gcaggctctg gggcttctc
301 tctctcttgg aggtgtctt ctggctgag ggtggagag gaaggagacc taaccaccg
361 ctctctctct gactctcaca taagaatta tggctcagc gaggagacc gtttggttta
421 ttatttcagc ccagagagc aatagctctg tggtttctt atttgagcc atggagagat
481 gggctctctt ttggctctt ttgatttca ttgaaacca agtctactct gttccaggaa
541 agggagggga ggtactcac atacttgcg tgaatgata attctgtt acagaagctt
601 tcatattatt accctctcc ttaanaaac ctatgaat aagcttctt tcatattac
661 aagtaaac taagctcag agaggtgag ttgttttct tatatgaa tctctaatg
721 catggcaag gtggatttg aattgtttt tctctctaga tttagttt tctctactct
781 actcccagg gcttgaggt atctgtgta aattaatgt tattgaac ctactatgt
841 ccatgaatt aaaaacagc ctaaaagta gtagagctg catgtatag atcagaagat
901 agctcagag aggtcagta actgtccaa gccacacag caactacat gtagagcag
961 gactagaacc taagtctct gattcttag ctaggtgac agccctctc aggtgtctc
1021 ctgaggtga cgcactctg ccgctctg ctgacatcc agtctgtct ccatgaagc
1081 ctgagcagg gactcctag ggcagcttc tcccagaca ttctctacc tcaatgaac
1141 tggaggtttt gggcaggca tccccttgg acaagctg gggagctgt acctgcaat
1201 gttgtttgt tgtgtctag caatacaca tccaccatt ttggagctg acctgcaat
1261 gcccttaac cccaccaga caatacctt gccagctgc cagatggga cccagcagt
1321 accagacct cccacagga ccaaacctt gctcagctg ggcctctcc tggcagaca
1381 aagaaagca gtcagactt ctgacaca ctctgacct ggttagtat agccctgctc
1441 ctgggttta cactcagct gccagaact pagcccaat caagtagg agagagccc
1501 aaaaagtc caactctc ctgtggaga ggtgagct gaggccca cctatgagt
1561 gatggagca agtagagc actagttca cccctcag ctgcaaatg tactcagac
1621 cctctagag aggtggaga cccctcag gccatgcaa tcccgaag gaggctgtt
1681 caagttgt tatgaatg taaccagga gccctctgc tgtgctgg actggagcc
1741 cctctgccc cctcctagc tctctctc tggctctcc caccacaa caaccaccc
1801 caactaac tctcagag cttgagct gggagagag gcttctgc cccccaagt
1861 gggactgcc agggcctct tctctctc atgaacct ccttctac agctgaacc
1921 tctctctgc ctgagggcc cctgaatg gctcaggg cgggtctg gccacacag
1981 gtagggcag tgcctctc gctcagcc tgaatgagc gaatgaacc tgcctttgt
2041 tctcagct gcaggctgc gctgactg ggaacttt ctgtctcag ctggagact
2101 gggataggc agagagggg gtagagact tgaagagct gcttctac cctctgtc
2161 atctgtgca cttcagct ttgagact ctgtgagc atctgtctc aactgagat
2221 caggctgag tgaattat tcaattctc tgaatgag actaagctc agagagagg
2281 ggcagcatta cccaaagcc cacagctgt aatgggaa agaggggtt aaggtatg
2341 ctctggctc cagagacac cccctctcc tggacttcc ctgggagca agagaggtg
2401 gggaggttc taagttgtt ggaactca tctctctg gctcaaat gctcaatg
2461 gccctggag gactctgt taagctca acttaaat aggaaccag ctagaggg
2521 ctgactct tctcactg ctgactgc tcaattca gagaagcc cagctccca
2581 gatggccca tctgctga gtagcaag gtagaact ggcagact cctggagcc
2641 ctgctggag caggccccc gctcagacc tggctggg ccagagag cttatctc
2701 gcaagccat ccaactct gtaatgag agctcggg ccccaacc atctggact
2761 caagtagc gtttgaact aaccagga ttgtctct ggcagtag ggggagct
2821 gccctctg gcttggct gagatgag agaaagcc ctagagct tctgagag
2881 gactcagtg cagaagggc tctgctgg gtagcact tctcactc tctgagag
2941 gctctgag gctctgag taatgctg gtagcact tctcactc tctgagag
3001 gctctgag atctgagt gattattt ctctggag aagaagct accctctc
3061 cctgagca ataccact gctctgccc agcttggag gccagctca ctggaagat
3121 gtagagag gaaccaccc cagctctg tctctgag tctctctc cagccagct
3181 cagatcagc aattgggt ctgagagag gactctcc ctgtctct tttgttgg
3241 gctgagga cctgttgg tctgcttt gccctaac agtctgca cctgcaac
3301 gttgctcacc ctctcaag tcaattca tctctctg atctgggt ggggtcacc
3361 tctgttga gtagagag aggtatga tcaaatgt ctttgaat caaagtg
3421 aggttaga taagpata tctattat gcaaatag aactaaga caactcgt
3481 gcccctag gactgctt tgaagctg atccttga gcttaaac atctcagg
3541 tcaagggca tgtctgat ccaagaca agccctcc tggtaggg actcaggc
3601 agtgccgt catcaact gcccctga tctctgac ctgagggg atccaggg
3661 ggcagagtg gggctctt gcccaggg ctctctga agtagagt gagagccc
3721 ctgcccag gaaacagag actaaagag gtaaatgccc acagcttcc catgtaatt
3781 gtagggta gatttgaac ccagcctca cagcttacc atgataatg ataaagcc
3841 gattcaaac cggctcagc ggcctcga gccctgca ctcccaatg ctgcaactc
3901 gctaggaat cctctctg agttctct agccctggg ccaggtccag aggtctgag
3961 cacactcga agtgcact ccaagagta tcaagctcc ctggaaggt gctctgagc
4021 tgtgtaga gaaagggct cctctgag tcaactccc cgtctgtg agcaatgct
4081 ttgtctca agatctccc tctctcag cctgctgc agtctctt cctctctt
4141 ccaagcatt cagcacaata gactctct cagggccc tcccagca cacttga
4201 gcaagatg ggcctaac ggcctgca agcagaca gctctctt gggagggca
4261 gggctctc gaaactgc ccagctat cagggagt tggcccaac acatccaga
4321 aagcagaa ggtcagcc gtagagtg gcaagctg gtagatcag tcatccctt
4381 gtagcctca agctctg gctttgtt ggtctccc aacaagct gggagatg
4441 aggggggg taggtctg aggggggg tcaactca gccacagta agggatga

4501 tgetgctec telgttacc teagttgaa caactactaa ggtccctggc tectgtctc
4561 caectgecc cagggtccct gnttrecat ccccagggec teactctcc ctgagccctg
4621 cagtctctca ctactctact caetctacta tteactactc caetctacta cagggtttg
4681 agttgataa cacagataa tgatggcca gttgactccc aaactaaga cacggagctg
4741 tctcttggg gggagctca gggctgggg aagtaagtc cagctgagg cacagtggc
4801 ctgagctct ctactgacc gcagagccct tggtagctgt cgtgtgatg gogtctctc
4861 cctgtctct tctctggca tggagagctg aacctgtat ccaagtgtt aaacctgccc
4921 ctggggaaa acactataa tctgataca tctcttttg tgcattaaa aaaaataga
4981 gaaaaaact tcaatgacc caetcaagt gtgagcagc ctgtgtttg caggaaagca
5041 gaagagctg ptttctgtt agggagggg acgagctccc acaacacct ctgaaaggtg
5101 tataagagc cccagctgc agctggctt ggtacctct gccagatct ctgggtttg
5161 ctgagaaact acggctcca gctactggc catgaccac acatctctg aaactcttc
5221 ctccacctt ggggtggct caacctggg ggttccctc ctggctggg gogtggctt
5281 tgggggggg agtctctct gggaggggt aagccgaagt atctagctt cttctctag
5341 gtttctctc tcaagctcg gggagagata tgggggggc atgaggtct gttgcttgg
5401 tggagggct agtgggttt tgggtgggg cttctgggg gggctgggt ggggtttgg
5461 tggggcttt gggggggg atgggtctt cctctgggc aatgaaaaa ttaccatga
5521 gaacctcaat gacgctggg cctctctct ggcagagta cgtctctgg agggactca
5581 tctgactcg gaggtaaga tcaagactg gtacagagag cagaccaca ccagccaga
5641 atgctctac agcaaatct tcaagactt tgaaggtct cggagcaag tgggtctctg
5701 gattctaaag aggggggac agtggctgg taccctgtac cggactctaa tttcagttt
5761 actgtcaat caagaaact cttctctgc caggcaatg actgtctca tttcagttt
5821 caaatctgg aatgactct ctgctcaggt gcagagcca gtgcactct tctcaactg
5881 ctgtgtctg gaaagggag gggaggtctt gggatctgg ttaggacca caggggaca
5941 gttggggca attataaga gccagagca gtactcaca cctgtaacc cagtgtttg
6001 gggggcaag gttatagat tctctgttc caggattag agctgcatg gactatgoc
6061 tttgcaact cactcggcc taggtgacag agtgaacct gtatcaatt aataaact
6121 nttaaaaaa aagagcaac tgcctctcc tactctatg acctgagg cttctgata
6181 aagtattct ttactcttt tcaagaagt ggtttgggt gtgctctat gattctaga
6241 tctgagacc tctgtcttt tctataaag aggtctttg tgcctaaat actgtgggt
6301 cttaggagc ctgggaact attacactt ctgcaacca ctaactcacc atttactcc
6361 ctaagactg gttctctat cgtaaatgc cttcttaact gccctcagg ttgactagg
6421 ctggaatgt ctgtagact atgctgtgt ccaaggtgt caggagcag tgactagag
6481 ttgtccacag tgggctgtt ggggtgtct tgaagagct atgtcagag ctgctgac
6541 ccaactgcc cactctctg ggaacaact tcccacagg aatactgct cttgtctt
6601 tctgctctg gataaaga agcagctcc tcaaggtct caagggctt ggtgggaca
6661 aaacaagaa tgaactgca aagtggctt gactttgct caccacaga gcttcaaga
6721 ttttgggg tgggtcaga ggggtgatt tcaatgggg gtaatttca atgagggct
6781 gcttgggt cagagcaga ggcagcagg ggaagagag gogtctctca gacacgact
6841 ctgtctctg cagatctgg ccaacacat cgaactccc gggctctcc tggagatga
6901 caatgcagg ctgctctgg agactctag gttcaagtc gttctctctt cttctctcc
6961 tctctctat gccaactta cttgctctc tctgtctct gactgggca cccaagacc
7021 ctctctctc cttctctgc cctctctcc tcaagtatg gaatggctg gctcaagcc
7081 aggggttga ggtgactc aagctctgc cctgactct ggtgactg cctctgccc
7141 gactgactt ggaatgctg atcgagctg tgaatgaga gttactctc ctgagaga
7201 acaagagat gggagact gttctctcc acctgtaac cagagagag tgggactct
7261 ctcccaaaa ctacttaaa ttcccagag gaaagacag tggctcagg aagctctcc
7321 agatcttca gaaatagat cagtgaggg cagatgggg cttgtgttc tctctatg
7381 ttgctctgg gcttagtgt tcaaaaac acctaccaca caaacaccg tttctctag
7441 gagaagagg ggcagctgt gttgctcag gcccaagcc agccaggtg gctctatg
7501 atagttact tctgattgg cttgtgagc tcaatcagg gctctcag cctgggtcc
7561 tctcttcta ggcactaga cagcagaga cttcttagg ctgagggct cagtgaaca
7621 aactctca tgtcnaag ggaacaggg ttgggtgtg ttacggcag agcaggtt
7681 gctgtgtg agccagact ctgactctc tctctctct cctgtctca agtggctcc
7741 tctctctg cagagatga aggattcag cagcagctg gccgctagg tcaatgtga
7801 gatgacgca gcaacggat tgaactcag cctgtgtct gcagagatg gggagcagt
7861 cagagcctg gggagaga acgcccagg tctcagggc tggttctca gcaagttgg
7921 cttatctc tgcctctcc tctctctcc aggtcctga gttctgag agccctggg
7981 gctactct gttgtctg actacagac tgaagagct acaaaagg tggctctca
8041 caagaaatg atccagaca gcaagagga gatcacagc ctgagacga cagtgcagg
8101 gctggagatc gactgcat ccaactcag ctatgacgg ccaactgccc ccaactcag
8161 cctctgggt ggtgaagct gcaactctc ctgactttt aagcttga gttctcagg
8221 cctctcaag gttgttccc tgaactcat ctactctcc cctctgctt gctctcagt
8281 gctctctc tagctgat tttctgca tctgtctcc cccaagag ctgggttga
8341 gaactcact ggcagacag agtgcctga gttcagcag ctgagcaga cctgactca
8401 cttgtggc ctgagggccc agtgaatga gttcagcag gaaagggg ctgactcag
8461 ggggtcaag gttgtctg acataaagc agcagggag cagagctgg cttactcag
8521 caactctc ggggtccc agcaaaagc agcagggag cagagctgg cttactcag
8581 gctgtctg ggtctgag cagactcat ctaaaagga tggcgggtg tctctcaga
8641 gctgtgag actgtgaaa gaaactgag caaaaggg ggtctgag tctctctc
8701 gttctgtg gttactgag gaaactcag tctctctc tgaactctg tttctctg
8761 aatgataga gctgggcta gttgattga agatctagt ctactctc tctctcag
8821 agatagatg gttctgtg agttgaat ctgaccagc cagtggctc atgctgaa
8881 tcccagact ttggggctc agcgggtgt gatacttga ggtcaggt tcaagcag
8941 calgccaat atgtgaaac cctctctca ctaaaatac aaaaatagc tgggctggt
9001 ggcgggtgc tgaatccca gctactcgg agctgaggg caaagaatca cttgaaccg
9061 ggggggggg gttcagtg gctgagatg tgcactgca cctcagctg ggcagagag
9121 tgagattcc tctcaaaaa caaacaaaa caaacaaaa tctgtctct tgggtagc
9181 agcctgact gggctgagc gatgctccc cctcagaga gttcagatg tgggtgtg
9241 tgggtgagg tctgaaga acgctagac tactctcat agtctctc acgacactt
9301 gactgtgac ctttctgta aatcagctc cctctctct cagagtggc tggactgag
9361 atcaggaag gtaggaagc ttactcttc tgggtgagg tgggggggg agatgtga
9421 gccacccc cttcgatg agggagagg gagaagagc caaacatcc cttaggtg
9481 gaacatccc atctataga tgcctctcc caagtggaa cttctgtag cccaagtg
9541 gttgagct gttgattat tgaagaaact caagtggag taggtctat ctagggact
9601 tctctccc agcccctac tctactcact tgccttaga tgcacaaat tctctccc
9661 tcatgtgag ctgaaaaac cctgcaatt tttctctct cgggctaac ctgactg
9721 cactgccc cctctgtag aggaatcaa caccagatct ctggcagct tgaaggt
9781 ggcaggggc cactctagg gttgtgctt agtttccca gacccctcc agtctgag
9841 cccctggg agaggggg gttctccc cttcaagcc tactctct cttctgag
9901 tctcaggg gttgtgtag cagcagcaat ctccacata atgtagaag cttaggtg
10021 ctggcactc cctctctca caggcaag atcagagctg ctatggag ggaagctcc
10081 aactcagct cctgctctc agagagca aagggctcc tgggtcttt attcaggg
10141 ttgactgct cctctccc cctctgctc tcccactct cttggagca agpatgca
10201 gctgtatgt gtaaacact cttgttaca gttgtctc atgtaataa gaattactt
10261 tcttttga aatctctt gacagtagt cagtttct tpaatgggt cttgttgg
10321 ctctgactc gaaagagc caggtccac atgagatct cttcccac tttgctct
10381 tctctgctc gcaacaatt tttgggtac ccaagggc agtggctgg acaaatag
10441 tctactcag gtaattgct cttgtgctt acctgggtg gaccaggaa gtaaggtg
10501 gaaagttcc taaaagagt ttgtgtacc cagcagaga ctaacacat ctataatta
10561 tgaatttta ccaagact gatgctgtt atgtctctg tgcctatcc cttgctgag
10621 atgatgag gcaactcuc ccaagggct ctttaacc tcaactgt agtggctg
10681 gattattat tcccacta gactgaggt cactggggc agagagaaga ataactgct
10741 cccggcttg tggatgat atagtagca gggctgggc tggcctgtg tctgtagc
10801 tcaagacca gggctgctc tggcccaca gggcctttt cacagctgc tgcaggacc
10861 ctgctctg ctctgacc tcaacatg tctctctca atctccatc cttctgagc
10921 agcataaca cctcagat gggcagcca gttcccaat gactcctct gcaagctgg
10981 cccatccc cctgaagct atgctctcc caaaagca cttatagc ccaagggaa
11041 agacaccag cctgagagc cctgagatt tcccagat taagtgaaa gcatatca
11101 cctctctac cctcttant caaatctg cctctgctc ccaatccca gcatctcc
11161 cagactct ctgaggggt agactcagc caaatatg tttggctc tttctcagg
11221 ctgtttctc tgaactctg cttatctg ccaaacaga ggcagaaaa tggagagca
11281 caaactct gttcagag ctggctggc tcaagactg ggtgagat tggggtgg
11341 ggggagaa cgtttctg gttcactg gaagagaga tgggggtgc cttctctt
11401 ggaatccc atctttgg ggaaggtca agcctctga ggggtccc atgtctctg
11461 ttgtgaaag cctctctcc caaacctt gttggaggt aaggtgtg cttctctt
11521 tggggagat gaaataga atgtctcc tgaagcaga caaggtgag cttctctt
11581 cttggagaa gaaatgaa tttagggg gttataga tttcaggcc tttctctt
11641 atccggat ggcactcat taactctc aatgacttt taagttaga tcaactca
11701 gctcagtt cagaagga aactgagc tggagagt tagtaactg aatgagcca

Annex 1 Sample 16.

Nucleotide [dropdown]

GenBank

Mentha spicata photosystem II cp47 protein (psbB) gene, partial cds; photosystem II subunit T (psbT) and photosystem II subunit N (psbN) genes, complete cds; and photosystem II subunit H (psbH) gene, partial cds; chloroplast

GenBank: HQ384581.1

FASTA Graphics PopSet

Go to:

LOCUS HQ384581 2207 bp DNA linear PLN 25-FEB-2014
DEFINITION Mentha spicata photosystem II cp47 protein (psbB) gene, partial cds; photosystem II subunit T (psbT) and photosystem II subunit N (psbN) genes, complete cds; and photosystem II subunit H (psbH) gene, partial cds; chloroplast.
ACCESSION HQ384581
VERSION HQ384581.1
KEYWORDS
SOURCE chloroplast Mentha spicata (spearmint)
ORGANISM Mentha spicata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetales; asterids; lamids; Lamiales; Lamiaceae; Nepetoideae; Mentheae; Mentha.
REFERENCE 1 (bases 1 to 2207)
AUTHORS Refugio-Rodriguez,N.F. and Olmstead,R.G.
TITLE Phylogeny of lamiidae
JOURNAL Am. J. Bot. 101 (2), 287-299 (2014)
PUBMED 24509797
REFERENCE 2 (bases 1 to 2207)
AUTHORS Refugio-Rodriguez,N.F. and Olmstead,R.G.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2010) Department of Biology, University of Washington, PO Box 355325, Seattle, WA 98195, USA
FEATURES Location/Qualifiers

source 1..2207
/organism="Mentha spicata"
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/specimen_voucher="R. G. Olmstead 92-191 (WTU)"
/db_xref="taxon:29719"
/tissue_type="leaf"
/note="authority: Mentha spicata L."

gene <1..1514
/gene="psbB"

CDS <1..1514
/gene="psbB"
/codon_start=3
/transl_table=11
/product="photosystem II cp47 protein"
/protein_id="AEN26622.1"
/translation="YRVHTVVLNDPGRLLSVHIMHTALVAGWAGSMALYELAVFDPSD
PILDPMWRQGHFVVPFMTRLGITNSWGGWSITGGTVTNPGIWSYEGVAGAHIVFSGLC
FLAAIHWVWDLIFSDERTGKPSLDLPKIFGTHLFLSGVACFGFGAFHVTGLYGP
INVSDPYGLTGRVQPVNPSWVEGDFPVPGGIASHHIAAGTLGILAGLFHLSVRPPQ
RLYKGLRMGNIETVLSSSIAAVFFAAFVAGTMMWGSATPIELFGPTRYQWDQGYFQ
QEIVRRVSGLAENKSLSEVWSKIPEKLAFYDYIGNPAKGGFLFRAGSMDNGDGIAGV
WLGHPIFRDKEGRELFVRRMPTFFETFPVVLVDGDIIVRADVPFRAESKYSVEQVGV
TVEFYGGELNGVSYSDPATVKKYARRAQLGEIFELDRATLKSQGVFRSSPRGWFTFGH
ASFALLFFGHIWNGARTLFRDVFAGIDPDLDAQVEFGAFQKLGOPTTRRQAV"

gene 1704..1805
/gene="psbT"

CDS 1704..1805
/gene="psbT"
/codon_start=1
/transl_table=11
/product="photosystem II subunit T"
/protein_id="AEN26623.1"
/translation="MEALVYTFLLVSTLGIIFFAIFFREPPKVPYTKK"

gene complement(1874..2005)

/gene="psbN"

CDS complement(1874..2005)
 /gene="psbN"
 /codon_start=1
 /transl_table=11
 /product="photosystem II subunit N"
 /protein_id="AEH26624.1"
 /translation="METATLVAIFISGLLVSFYGYALYAFGQP5QQLRDPFEEHGD"

gene 2111..>2207
 /gene="psbH"

CDS 2111..>2207
 /gene="psbH"
 /codon_start=1
 /transl_table=11
 /product="photosystem II subunit H"
 /protein_id="AEH26625.1"
 /translation="MATQIVENSRSRSGPRRTVGLLKLPLNSEYCK"

ORIGIN

```

1  ggtatcgtg  tcataccgtt  gtattgaatg  atcccggccg  gttgctttcc  gttcatataa
61  tgcatacagc  cctggttgct  ggttggccgg  gctcgaatgc  tctgtatgaa  ttagcagttt
121  ttgatccttc  tgaccctatt  ctgatccaa  tgtggagaca  gggtagtttc  gttataccct
181  tcatgactgc  tttaggaata  accaatcat  ggggggggtg  gagtatcaca  ggaggaactg
241  taacgaatcc  ggggatttgg  agttacgaag  gtgtagctgg  ggcgcatatt  gtgtttctg
301  gcttatgctt  tttggcagct  atctggcatt  gggctctatt  ggatctagaa  atattttctg
361  atgaacgtac  aggaaaacc  tctttggatt  tgcccaagat  cttggcatt  catttatttc
421  tctccggggg  ggcttgcttt  ggttttggtg  cattcatgt  aacaggctcg  tacggctcctg
481  gaatatgggt  gtccgatct  tatggactaa  cgggaagagt  acagcctga  aatccgtcgt
541  ggggcgtgga  aggtttgat  ccttttgtc  cgggaggaat  agcttctcat  catatgacag
601  caggtacact  ggcgatatta  gcgggtctat  tccatcttag  cgttcgaccg  ccacaacgtc
661  tatacaagg  attacgatg  ggaatatgt  aaaccgtact  ctccagtagt  atcgcggctg
721  tcttttttgc  cgcttttgtt  gtgctggaa  ctatgtgta  tgggtcagca  actactcca
781  tcgaattgtt  tggcccact  cgttatcaat  gggatcaggg  ttatttcag  caagagatat
841  atcgaagagt  tagcgcgggg  ctagccgaaa  ataaaagttt  atcagaagtc  tggctaaaa
901  ttccgtaaaa  attagctttt  tatgattata  tcggcaataa  tccggcaaaa  gggagattat
961  tcaggccagg  ctcaatggat  aacggagatg  gaatcgrgt  aggatggtta  ggacacccta
1021  tcttttagaga  taaagaagg  cgtgagcttt  ttgtacgtcg  tatgccacc  tttttgaaa
1081  catttccagt  gtttttggta  gacggcgcgc  ggattgttag  agctgatgtt  ctttttagaa
1141  gggcagaatc  caagtatagt  gttgaacaag  taggtgtaac  tgttgagttc  tatggtagcg
1201  aacttaattg  agtcagttat  agtgcctct  ctactgtgaa  aaaatatgct  agacgtgctc
1261  aattgggtga  aattttgaa  ttagatcgtg  cgactttgaa  atccgatggt  gttttcgtta
1321  gcagtcagag  gggttgggtt  acttttgggc  atgcttctgt  tgctttactc  tttttctctg
1381  gacacatttg  gcatggtgct  agaacctgt  tcagagatgt  ttttgccggc  attgaccagg
1441  atttggatgc  tcaagtagag  tttggagcat  tccaaaaact  tggggatcca  actacaagaa
1501  gacagcgagt  ctgatacaag  accactttgg  cttttttccc  ctctattttc  tttttggggg
1561  gattttacat  ttttacatag  agtaccggag  tgaatttgaa  tcaccgcttt  tttgatittt
1621  gctctttcaa  gatgattccc  aaaaagaaaa  tgaaaaaaaa  cgaacaagta  ggaagctat
1681  aattgtaaac  cagatcgaa  tttatggaag  cattggttta  tacattcctt  ttagtctcga
1741  ctctagggat  aatttttttc  gctatttttt  ttcgagaacc  gcctaaagt  ccaactaaaa
1801  aataaaaaaa  tgatttttca  ttatctcaat  tgaagtaatg  agcctcccaa  tgcctggagg
1861  cccattactt  caactagtcc  ccgtgttctc  cgaatggatc  tcttagttgt  tgagaagggt
1921  gcccaaaagc  ggtatataag  gcgtaccggg  taaaacttac  aagtaacca  gatataaaaa
1981  tggcgactag  gtttctgtt  tccattatga  ttatataatt  tcaagacccc  aactgtacta
2041  tcataaaatc  gtttatttac  aacggaatcg  tatacaaaag  caacagatct  caatgaatag
2101  aataggattt  atggctacac  aaactgttga  gaacagttct  agatcgggtc  caagcgaac
2161  tactgtaggg  agtttattaa  aaccattgaa  ttcggaatat  ggtaaag

```

//

Annex 1 SAMPLE 1c.

Nucleotide

GenBank

Mastomys natalensis voucher FMNH 166943 Brca1 (Brca1) gene, exon 12 and partial cds

GenBank: EU349660.1

FASTA Graphics PopSet

LOCUS EU349660 2540 bp DNA linear ROD 11-MAY-2017
DEFINITION Mastomys natalensis voucher FMNH 166943 Brca1 (Brca1) gene, exon 12 and partial cds.
ACCESSION EU349660
VERSION EU349660.1
KEYWORDS
SOURCE Mastomys natalensis (African soft-furred rat)
ORGANISM Mastomys natalensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mastomys.
REFERENCE 1 (bases 1 to 2540)
AUTHORS Rowe,K.C., Reno,M.L., Richmond,D.M., Adkins,R.M. and Steppan,S.J.
TITLE Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): multilocus systematics of the old endemic rodents (Muroidea: Murinae)
JOURNAL Mol. Phylogenet. Evol. 47 (1), 84-101 (2008)
PUBMED 18313945
REFERENCE 2 (bases 1 to 2540)
AUTHORS Rowe,K.C., Reno,M.L., Richmond,D.M., Adkins,R.M. and Steppan,S.J.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2007) Biological Sciences, The Florida State University, Tallahassee, FL 32306-1100, USA
FEATURES Location/Qualifiers

source 1..2540
/organism="Mastomys natalensis"
/mol_type="genomic DNA"
/specimen_voucher="FMNH 166943"
/db_xref="taxon:10112"

gene <1..>2540
/gene="Brca1"
/note="exon12"

mRNA <1..>2540
/gene="Brca1"
/product="Brca1"

CDS <1..>2540
/gene="Brca1"
/codon_start=3
/product="Brca1"
/protein_id="ABY60630.1"
/translation="SQQSRWAESKETCNERRPSTGEKVDANADSLCGGKKAHPKSLCPENSGATTDVPHITLNSSIKKVNWFSTRGEMLTSDGASDRDESNAEAAVLEVSNEVDECFSSSKKIDLAAPDSLNALMCKSGRDLKSPVENSTNDKIFGKTYQRKGSRPPLNHVTEIIGHFTKEPQITQEPPFTNKLKRKRSTCLOPEDFIKKADLAVVQRTPEMNNQGTNQHETNDQAMSI TSDYENKITGXNLQKEKNAYPIESLKKPEAFTAKVKSISNSISDLEVELNVHSSKAPKKNRLKRKSSTRCVLPLEPISRNPSPPTCAELQIDSCDSSEEKNNNSQTPARHIREPQLIKDTEPAADAKYEPNEHKKRSASDALPEEKLNNKAGLLTSCSSPRKPGQVNPSPQRKGIQLQETCQVPDNTKELGDLVGGEPGKPTQSEESTSVSLVPD TDYDTQNSVSVLEASTVRYTRTESAQCHTFVAVSNPKELVHGSNNAGSGTECFKHLRHELNHSQETTEMEDSELDYQLQNTFQVSKRQSFALFSKPRSPQKCAHSVPSREPSPKLTSKTEKEHQESEEISHVQVVTAVGLPVPCQEGKPDADIPCAGVSGLCPS SHYRSSENGLSTTGKAGISQNSHFQSVSPIRSSIKTDNRKPLTEERFEKHTSSTEKAHNGNETIVQSTVHTISQNNRSNACREASLDSIDEVCS TGEYFRGQLGRNRGPKVNTMLPFDSTIQPGVCKQSVASADKYLEIKKQEGE TVGADFSPCLFSDHLEQPMRSKGFQVQVCE TPDDLDDVEIRENTSFGEGDIMERSAVFNGSVLRRESSRSPSPVTHA"

exon <1..>2540
/gene="Brca1"
/number=12

ORIGIN

1 tgagccagca gagcagatgg gctgaaagta aagaaacgtg taatgagaag cgggttccca
61 gcactggggg aaggtagat gcaaacgctg attccctctg tgggggaaag aagtgggctc
121 atccgaaaag tctgtgccct gagaattctg gagctaccac cgaatgtcct tggataaac
181 taaatagcag cattaagaaa gttaatgagt ggtttccag aactggtaga atgtaacct
241 ctgagggcgc gctgacagg gatgagtaa atgctgaagc agctgtgtg ttggaagtt
301 caaatgaagt agatgaatgt ttcagttctt ccaagaaaat agacttagca gccctgatt
361 cccttaatgc tttaatgtgt aaaagtgaa gagacttacc caaacagta gagaatagta
421 ctaacgataa aatatttggg aaaacctatc agagaaaggg aagccgacct ccctgaacc
481 atgtgactga aattataggg atgttacta aagaaccaca gataacacaa gacgagcct

541 tcacaataa attaaaacga aaaagaagta catgccttca acctgaggac ttcatcaaga
601 aagcagactt agcagttgtt caaaggactc ctgaaaatat gaatcaggga actaaccaaa
661 tggagacaaa tgaccaagca atgagtatta ccagtgacta tcaggagaac aaaataacag
721 gtgrtaatct ccagaaagag aaaaacgctt atccaataga atcattgaaa aaggaaacctg
781 ctttcacagc taaagtcaaa tctataagca acagtataag tgatttggag gtagagttaa
841 atgtccacag ttcaaaaagca cctaagaaaa ataggctgaa gcggaggtct tctaccaggt
901 gtgtcttctc acttgaacca atcagttaga atccaagttt acctacttgt gctgaacttc
961 aaatcgatag ttgtgatagc agtgaagaaa agaaaaacaa ttccaaccaa acgccggcca
1021 ggcacattag agagcctcag ctcatcaag acactgaacc tgcagctgat gccaaagaat
1081 acgagcmeta tgaacacaaa aagaagagaa gtgccagcga tgctctcca gaagagaat
1141 taatgaacaa agctgggtta ttaactagct gttcaagctc tagaaaacct caeggacctg
1201 tcaatcctag ccctcagaga aaaggaatag agcaactga aacatgccaa gtgccgtgata
1261 ataccaaaga actcggggac ctggctctgg gggggagcc cagtgccaaa cctaccgagc
1321 aatctgagga gagcactagt gtgtccttgg tactgacac tgactacgac actcagaaca
1381 gtgtctcagt cctggaagct agcactgtca gatacaaaag aacagagctc gctcagtgta
1441 tgactcagtt tgcgcaagt gtaaacccca aggaacttgt ccatgttct aacaatgctg
1501 gaagtggcac agagtgtctc aagcaccgt tgagacacga gcttaacct agtcaggaga
1561 caacagaat ggaagacagt gaacttgata ctcagtatt gcagaataca tttcaagttt
1621 caaagcgtca gtcatttgc ttttttcaa aacctagaag tccccaaaag gactgtgctc
1681 actctgtgcc ctcaaggaaa ccgagtcгаа агттgactt тааггтаса гаааагаас
1741 atcaggaaca agaagagtct gaaatcagtc atgtacaggt agttacggcc acggtgggct
1801 tacttgtgcc ctgtcaggaa ggttaagccag atgctgatat accgtgtgct gtagtttctg
1861 ggctttgtcc atcatctcat tacagaagca gtgagaatgg actcagcacc acaggtaaag
1921 ctggaatttc acaaaaactca cattttaaac aatcagttt tccatcagg tcatctataa
1981 aaactgacaa taggaagcct ctgacagagg aacgatttga gaaacataca tcatcaactg
2041 aaaaggcaat gggaaatgag actattgttc aaagtacagt acacacaatt agccagaata
2101 acagaagcaa tgcttgtcga gaagccagct tggacagtat tgatgaagta tgttccactg
2161 gtgaatactt ccggggacaa ctaggtagaa acagagggcc caaggtaaac actatgcttc
2221 catttgatag tatacagcct ggtgtctgta agcaaatgt ttctgcaagt gataagtatc
2281 ttgagataaa aaagcaggag ggtgagactg ttggtgcaga cttctctca tgtctgttct
2341 cagatcacct tgaacaacct atgagaagt gcaagattt tcaggtttc tctgagacac
2401 ctgatgacct gttgatgat gttgaaatc gtgaaaatc tagcttgggt gaaggggaca
2461 taatggagag gtcctgtgtc tttaacggaa gtgcctgag aagagagctc agtaggagcc
2521 ctagccctgt aacctatgca

//