

Supplementary Figure S3. The prediction of *DUXA* derived from *Bos taurus* isolate L1 Dominette 01449 registration number 42190680 breed Hereford chromosome 18, ARS-UCD1.2, whole genome shotgun sequence. Three possible ORFs for exons, but not introns, are depicted. Putative protein sequence is highlighted in yellow. Sequences from StringTie merge prediction and confirmed cDNA are drawn as lines below the corresponding sequences. Cloning primers are drawn as line arrows. Splice sites are underlined and codons split by two exons are coloured red. The homeodomains are highlighted in green.

[illegible]

Cloning primer 1 _____

StringTie merge

DUXA_16c_01

S K F
 V S S
 Q * V P

10101 CAGTAAAGTTCCTCACGCTTCAGACCGCTGTGCTCTTTTCCTTGGTTGGGGGGAGAGGGAGGGGAATTGGACTGAAAACATACTGAGCTTATGAGAAGGTA 10200
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10201 TGGGAGTCCCATTTTTCAAGCGTTTAAATAACTGTGCTAGGTAAGAGAAAAGTATCCCTCCTGTAATATGAAAGCATGCTGTTTTCTCTGCATCTGAGC 10300
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10301 ACTGGGAATTTCAAAAAGATGGAGGCACGCAAGAACCAGACAATACAGGAGGAGGGTGAAAACAACCTGAACAATACAAAACGTTGGTTCAAGAAGGGAGA 10400
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10401 TGGAAGGGAAAAATTGATGGAAGAAGGAAGAGGAGCTATGTGCTGTGCTAAGTGGCTTCAGTCGTGTCCAACCTATTTGCAGCCACCTGGGCTGTAGCTCG 10500
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10501 CCAGGCTCCTCTGTCCCTGGGATTCTCCAGGCAAGAACACTGGAGCGGGTTGCCGTTTTCTTCTCCAGGGGATCTCCCCGCCAGGGATGGAACCCACAT 10600
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10601 CTCTTATCGGCAGCCAGGTTCTTTACCGTTAGTGCCACCGGGGAAACCCAAGGAAATGATGGGGGTATCAAAAAGCCCTGAGACCAGTAAGCTTTTGA 10700
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10701 ATCGAAGTAGCTTTTCAAGATTTCTCTCTCCAGACTGAACTTGTCTCCTGAACTTCAGCTTTGTGCATCTAACTGCCTGTAATAGGTGTCTCAAATCAACG 10800
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|
 10801 TGTCCAAAATGAAACTCCTTATGTTTCAGACCCAAATTCTCCTCCTGTTCTGCGTGCTACCAAATATCAACTGTTTGTGCCCCAACCTAAGCGTTCTT 10900
 ---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|---:---|

F1
F2
F3

10901 GGCCTTGTCTTTCTCTCATAGGGCTTCCCTGGTGGCTCAGATGGTAAGGAATTTGCCTGCAATGCAGGAGACCTGGGTTTGATCCCTGGGTAGGGAAGAT 11000
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11001 CCCCTGGAGGAAGGAATGGCTACCCACTCCAGTGTCTTGTACTGGGGAATTCATAAGCTACACCTAACCAGGCTGAGAATCCTGTTGGCTTTAACAAAAC 11100
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11101 ACTTAAACCATCATCTGCTGGATTATGGCAAGAATTGATCTCTCTGCCCCATCCTGGCCCTGATTCAATCTATTTTCAACTCTGTAGCCTGTGTGATCTT 11200
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11201 GTCAAACCTGTCTGATCAACTCTTCTCACTATTGGCCAGCCAAGATTGCTCAGTCGTGTCCCACTCTGAGACCCCATGACTGTAGGCCTGCCAGCTCCTCT 11300
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11301 GTCCATGGGATTTCCCAGGCAAGAATACTGGAGTGGGTGCGCCATTTCCTTCTCCATGGGATCCTCCTGACCCAGGGATTGCACCTGGGTCTACCTGGGAAG 11400
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11401 CCCCACTACTAGAGTCAAAGGCACATGGCCTCAAGAGGCCCATAGAATCTGCCCACTTGAGGGTCTAAGACCTCATCTGCTTCTGGTTCATTTTCTC 11500
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11501 TTCTCCAACCCTCCTGCTCTACCGAGTTTATTTCTCTGTCAATTTTTCCCTCATTACTTTGTCTTGTGTGTAAAATTCCATTACGGGAATATACCAGGTT 11600
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11601 TTACTTATCCATTCTCCTAGCAATGGGTGTTTCAGTTGCTTCAGCTTACTACAAAATACCATTCTGAAAACCTGTCATGGAGATGTATTTGGGGTTAGAATT 11700
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11701 ACAAAATTATTCTTTGTAATCTGTTACAAAGAAATGGAGTTGGGGTTAAACTACCTACTGTTGTTGTTTAGCCACTCAATTGTGTCCGACTCTTTCGCAC 11800
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11801 CCTATGGACCACCAGGCTCCTCTGCCTGTGGAATTTTTCAGGCAAGAAATACTGGAGAGGGTTGCCATTTCTCTCCAGGGGATCTTCTTGACCCTGGG 11900
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

11901 ATGGAGCCCTTGTCTCTTGCCTCTCCTGCATTGGCAGGCGGGTTCTTTACCACTAGCGCCACCTGGGAGAGCAGTGTGGAGGCTCCTTAAAAAAGAAAAA 12000
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

12001 TAGAACCACGATACGATCCAGCGTGTCTTTTCTGTGAGGTGTCAACCAAAGCAAATCCCTAAGTTTCTCTTTGTATGAGTCCAGCTCCAAACGAACCT 12100
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

12101 CTAGCATGCACAGTAACTTTTCTTTTCATCTTATTTAGTTTACCAAGCCTAATTCATTCATGTTCCAAAATGAACGCCAGGTTTTCAGATTGGCTAT 12200
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

12201 TTTAATGAGCTGTATCATATCTAAATCATGCAACATGGCATGTATTGCTTATTTGAAGATTCTACTGAATGTACGAATTGTGTTACCCCATGGTGTTT 12300
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

12301 AAAAAAATATTTGTAGGGGGATTTCCCTGATGGTCCGGCAGGTAAGGATCGGTGCTTCCGTTGCAGGGGGCACAGGTTCAAGCACTCGTTGGGGAACATA 12400

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
12401 GATCCTGCAGGCTATGCAGTCAGGCAAAAAAAAAATGGGAAGCTGATCCTATTTGTTGAGGGCAATGAACTTCAAAATGAGTCAATCTGTACTAGGAATTTA 12500
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
12501 ACCAGAAAACCTGAAAGGAGCTCAACAATACCTGTAACAGGTGCTCTTTGTGAGTTCAAATGCCCATCATTGTCATGACATCACCTGCGTCATCCTGACT 12600
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
12601 TGCATTGTTCTTATCAAGAACAATGTCCCATAGGACTGGGACCCATCAGTTTTTCATGTTTTTTTTTTTTTCTTTTTTCTTTTTTAAACAGAAGACA 12700
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
12701 TATGTCCCTAGTCATGACCTTCAGGTCAACTAAGGAAGCTTCAAGAATCACCACCCTTGTGCACTTAAGGGCTTTATAATCGAGGGTTCTTTCCATCTGG 12800
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
12801 GTTCTGGTGGATTGAGAATGAAAAGAGGAAATGAAAGACAAGCTCTAAGCTGAAGCGTGAATACATACGATGGCCCATTTAATGCTTCTCTCCTCCGATT 12900
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

F1
F2
L F3
Cloning primer 2 —
DUXA_16c_02 —

Y W S L S P Q R P W R H S V G A A V P R S Q K S S * R S S S R H S A F1
I G L S L P R D H G D T V * A Q P Y L V H R A V E D P R P G I Q F2
L V S L S P E T M A T Q C R R S R T S F T E E Q L K I L V Q A F S F3
12901 TATTGGTCTCTCTCTCCCCAGAGACCATGGCGACACAGTGTAGGCGCAGCCGTACCTCGTTACAGAAGAGCAGTTGAAGATCCTCGTCCAGGCATTCAG 13000
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

Cloning primer 1 →

Cloning primer 2 →

K T L T L A I L P N R D L L W K S T L K S L E F F1
P K P L P W L Y C Q T E T C C G N Q R * R V * N S F2
Q N P Y P G Y T A K Q R L A V E I N A E E S R I Q F3
13001 CCAAAACCCCTTACCTGGCTATACTGCCAAACAGAGACTTGCTGTGGAATCAACGCTGAAGAGTCTAGAATTCAGGTACATTCCAAGTTCAGATCTCC 13100
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
13101 AGGACGAGAAGCAGAAACTCAGGGTGGGAATGGGGGGTGGGGACTTATCTGGTGGCCAGTAGCTAAGACTCCGAGCTAGATGCAGCGAGACAGAGCTA 13200
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
13201 CAGGAGTTTGTGTCGACTTAAAAACAGTCACAACCGAAAAGCTGAGAGTTAGGTTTTTACTCAGTGGAATTTTATAGGACTCCAAGCCCGGGAGGCAGC 13300
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
13301 ATCTCAAGTAACCCTGAGAGAATTGCTCCAAAGAGGCTATAGGGGAAGAGGGAGGAGTCTGGTTATATTTGTGCTCAGTCCCGTCCGACTCTTTGTGAGA 13400

[illegible]

14801 TGCCAGAGGAATGACGGATGGTGTGAGGGCTAACCTTCTCCTTTCAATGAATTCTGGCAGACCAGGGAATAGCAGACTGGTGTGTGCGCTCAGTTGTG 14900
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

14901 TGACTCTTTGCAACCCCATGGACTGTAGCCCTCCAGGCTCCTCTCTCTATGGGATTTCTCAGGCAAGAATACTGGAGTGGGTTGTCATTCCCTTCTCTAG 15000
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15001 GGGATCTTGCCAACCTAGAGATGGAACCCACGTCTCCTGGATTGGCAGGTAGATTCTTCACCACTGTACCACCTGGGAAGCCCAGGTGTAATGTTGCAGA 15100
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15101 TTCAGGCTGGACGGAGATCTTACATGCAGAATAATTACAGCATCCATTTATTAACCTTAAAGAGTATTACAGGGGACTTCCCTGCTGGTGCAGTGGGTAAG 15200
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15201 GCATTGCCTTCCAATGAAGGAATGCAGGTTTAACTGCTCACTGGGGAGCTAGGATCCACGTGCCTCATGGCCAAAAACCAGAAGGGAAACAACAGATGC 15300
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15301 AATATTGTAACAAATTCAATGCAGACTTTAAAAATGGTCCATATAAAAAAATCTTTTAAAAAGAGATTATTATAGCAAATTGTGTTAGGACCAGGTCAT 15400
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15401 GAAGATGCTGAATGCCAAGCCTTTATCTAAGGCTTAGTTGAATGTAACCAGGATCAGTGATAACCTGAGTTTAGGGTTCATAAGAATCACTGCCCACCC 15500
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15501 CCTTTTGGGGGCCATGTCATGGCACGTGGGATTTCACTTCCCTGACCAGGGATTGAGCTCATGCTACCTGCATGTACTAACCCTGGACCACCAGGGCAG 15600
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15601 TCCCCGAACCACTCCCTTTGGAGAAGAGCAGCAACGGAGCCGAAACGACATAAACATGTATAATTCTCTCGTTTTCTTTTCCCTCTCTTAGATATGGTTTC 15700
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15701 AGAATCGAAGAGCGAGGCATGGATTCCCTGAAGAAACCCAAGAAGAACGTAGACCCAAGCAAAGACCACGCCTCCCCATAGAGAAGATTCAAGTAAACA 15800
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15801 GCCTGCTCCAGCCCTGCTCCCCACCAGAGCCTCCTGGGATGCAAGTGCCTCTTGAGATTGGGGCCACCTAAAAGAAACCACCTCTTCTTTCTCCGCCTT 15900
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

15901 GTAAGAGGAGAGTGGGAAAGGAATGTTTCCTCCTCCGGGTTTCTGGTCAAGGCATCCAGGAGCAAAGGATTCAATGAGGCCCTTGAAAGTCTCCCC 16000
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16001 ACAAGGTCCAGATATTCTGAGTCTTAGCTTATTTGCGCTGGAGTCTAGCAATTTAGCCAAATACTACTGCTTCTCACAAGAGATGCGGGTTCAAGGCCTG 16100
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

I W F Q F1
Y G F F2
M V S F3

N R R A R H G F L K K P K K N V D P S K D H A S P I E K I Q F1
R I E E R G M D S * R N P R R T * T Q A K T T P P P * R R F K F2
E S K S E A W I P E E T Q E E R R P K Q R P R L P H R E D S R F3

16101 GGTCAAGGATGATCCCTGGAGGAGAAATGGCAACCCACCCAGTAATCCCATGGACAGAAGAGCCTGGCGTCTACAGTCCGTGGAGAGTAGGACGCGA 16200
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16201 CTGAGCACACCCCTTCTAGAACCTTCCACCTTGAATCACCTCAAGGAGTGGCAACAGATTCCAAGTTAGATGGTGAGAACATGTTTGGGTTCCTCATCTGA 16300
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

M * A D P K C F S T F N P T P T P D G T D R R C R T S F T S S Q L F1
 I C K L I P N V S L P S I L P L L Q M E Q T D G V V P A S L P L S Y F2
 Y V S * S Q M F L Y L Q S Y P Y S R W N R Q T V S Y Q L H F L S V F3
 16301 ATATGTAAGCTGATCCCAAATGTTTCTCTACCTTCAATCCTACCCCTACTCCAGATGGAACAGACAGACGGTGTCGTACCAGCTTCACTTCCTCTCAGTT 16400
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

Q T L K N A F T E N P Y P G I D S R E Q L A E E I G V P E S R V Q F1
 K L L R M H S Q K T L I L G L T P E S N L L K K L G F Q S Q E S R F2
 T N S * E C I H R K P L S W D * L Q R A T C * R N W G S R V K S P G F3
 16401 ACAAACTCTTAAGAATGCATTCACAGAAAACCCCTTATCCTGGGATTGACTCCAGAGCAACTTGCTGAAGAAATTGGGGTTCCAGAGTCAAGAGTCCAG 16500
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

GTAATCGAAAAGGGACTTGATATTGTGTGTTTCATAATGAGACATCTTTAAATGATTTCTCCATCAGGGACTTATTCTCATAGACATCATAAGCAGGACAG 16600
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16601 ATGAAGACTCCCTGGTGGTTTCAGTGGCTAAGACTTCATACTCCCAATACCTGGGGTGCAGTTCAATTCCTGGTCAAGGAAC TAGATCCACATGCCACAG 16700
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16701 GCATGCTACAAC TATAAAAATCCCAAGCAAGTGCTGCAACTAAGATCTGGTGCAGCCAAATAAATAAATATTAAGAAAAACAAATAGCATAATATGTTAT 16800
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16801 AAATGAAAAAAGATAAGCAAGACAGATGAAAAATGGGATTTCTGGCTTGAATAGTTCTTCCCCAAGGGCTCAGTAGGTGAAGAATGCAATGCAGGAGAC 16900
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

16901 ACTGAAGACTCATGTTCAATCCCTGGGTGGGGAGGATCCCGTGGAGTAGAAAAATGGCAACTCACTTCAGTATTCTTGCCTGGAAAAATCCCATGCACAGAG 17000
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

17001 AAGCCTGGTGGGCTACAGTCCATGGGGGTTCTGAAGAGTTGGATATGACTGAGTATGCATGCACCCATGGGATAACAATGTCTTCCTAGAAGACATTACC 17100
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

17101 CTATTCTGGGTCATAGAGTAAAAATCAGGCATTTTTGAGAGGAAGGAGAGAAGTGGGAAGAGAATTCCAGGCAGAGAGAAAAGCTGAAATACAGGTATGGC 17200
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

17201 CCTTTAGGAAAGACTAATACTTACAATGAGGGAATTGCGGGGACTTTTCAGGATATGAAACTTATCAGTAAATCATAAAGGTGTTTGTAAGATTTCAGAA 17300
 ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

17301 TTTTGAGGGACATCTCTGGTGGCCAGTAGTTAGGACTCCACGCTTCCACTTCAGGGACACGGGTTTGATCCCTGATTGGGGAAC TAAGATCCCTGCAAG 17400

```

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
17401 CCACAGACACAGCAACAACCCCCCAAAAAAATTCAGAATTTTGAACTTTATCTTGAAATGTGAGACTTTCCCAGATGTCATATGGAATGTCCGTGGTT 17500
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

17501 TGAGAATTATGCCTAAAAGAAACCATCTCTTCCTCACCTTTGCCTTATAAGATGAGCGATGGTGAGACTTGCATGACATCAATTTTTTTTTCATTAAATC 17600
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

      L C F T Y L D L V P K P K S Q T P C P E K K R P * R S S I P K T K      F1
      F V L P T * I W F Q N R R A R L R V Q R K R G L E E A L Y Q R Q N      F2
S L F Y L L R F G S K T E E P D S V S R E K E A L K K L Y T K D K T      F3
17601 TCTTTGTTTTACCTACTTAGATTTGGTTCCAAAACCGAAGAGCCAGACTCCGTGTCCAGAGAAAAAGAGGCCTTGAAGAAGCTCTATACCAAAGACAAAA 17700
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

      P G T G S L I W E S *      F1
      Q G Q D L * Y G R V E      F2
      R D R I S D M G E L K      F3
17701 CCAGGGACAGGATCTCTGATATGGGAGAGTTGAAGGTACTAGCCTGCTGTTCCCAATGTGACTCACAATCAGAGTCTCCTGAGTGCCAGTTATTCGCAC 17800
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

      G G C G A C T C A A G T C C A G C T T C C T C T G G G T A G T G G A T A T G C T T C C A C T C A A G T G T G C C C C T G G G G T T G G G G T G G T C A C A G T C C C C A C G G C T C T C C C C A
17801 G G C G A C T C A A G T C C A G C T T C C T C T G G G T A G T G G A T A T G C T T C C A C T C A A G T G T G C C C C T G G G G T T G G G G T G G T C A C A G T C C C C A C G G C T C T C C C C A 17900
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

17901 AGTCTGTGGTCTCCCGAGGAATTCCACTGCCCCGCTTTTCAAGCTCTGAACTGGAACCCAGAAACCCTGTTAGTATGTGAGCTTTGTTTATTAATATTTT 18000
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

18001 CCCTAGTTCTGGCACATGGTAAACAATCAGATTTTGTGGAAGGTATGTCTGAGTTACTGGCAGGATGAATCCTCACGTGTGTTCTGCCAGGGGCTGGCTC 18100
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

18101 ATGCATGGCTTTCCCAAGGGGTAGGAGGTTCCATATTTGGCTGGAACGTTGGAAGAATCCACCTCCATTTTTGCATGTATAAACTGGTTTCTTCAATTT 18200
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

      V Y R T A L H S D C F I S L E P G Q S E K G D R F G      F1
      Y T E R H S I V T V L F L W S Q D K V K R E T D L V      F2
      I Q N G T P * * L F Y F S G A R T K * K G R Q I W C      F3
18201 TTCTTTCCCTTTTGTGCTTTAGGTATACAGAACGGCACTCCATAGTGACTGTTTTATTTCTCTGGAGCCAGGACAAAGTGAAAAGGGAGACAGATTTGGT 18300
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

      V F D V I S L D S T L I I F L N I S L * I Y L Q L Q L L E M V K R T      F1
      Y L M S S A W T P H * L S S L T Y P S K Y T F S F S F W R W * R G      F2
      I * C H Q P G L H I N Y L P * H I P L N I P S A S A S G D G K E D      F3
18301 GTATTGATGTCATCAGCCTGGACTCCACATTAATTATCTTCCTTAACATATCCCTCTAAATATACCTTCAGCTTCAGCTTCTGGAGATGGTAAAGAGGA 18400
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|

```

L G G R D R * Y K E P C S P S D Q I Q I Q S S T D L E D N K I I I F1
L * E A E T D D T K N H A A L V T K S R F R V Q Q I L R T I K * L * F2
F R R Q R Q M I Q R T M Q P * * P N P D S E F N R S * G Q * N N Y F3
18401 CTTTAGGAGGCAGAGACAGATGATACAAAGAACCATGCAGCCCTAGTGACCAAATCCAGATTCAGAGTTCAACAGATCTTGAGGACAATAAAATAATTAT 18500
----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|

Cloning primer 2 ←

Cloning primer 1 ←

N N I I F F G F1
I I S F F L V F2
K * Y H F F W S F3
18501 AAATAATATCATTTTTTTTGGTCA
----:----|----:----|----