

## Supplementary Data Figure.

The sequences amplified around each gene selected for reverse engineering are shown. The primer positions for the left and right flanking sequences are given together with the length of the flanking region and each flanking region is shaded. The position numbers at each end of the amplified region are indicated and the positions of start and stop codons for the genes within the amplified region are shown. For *M005L/R* both copies of the gene are shown.

### Flanking sequences for *M005L*

LH flanking sequence 308 bp shown in green: Forward: 4595-4611; Reverse: 4883-4902. RH flanking sequence 299 bp shown in yellow: Forward: 6383-6401; Reverse: 6667-6681.

Start codon for *M004L* at 4634-4636 is underlined; stop codon for *M004.1* at 4638-4640 is underlined; start codon for *M004.1L* at 4908-4910 is underlined; stop codon for *M005L* at 4932-4934 is shown in red; start codon for *M005L* at 6380-6382 is shown in red; stop codon for *M006L* at 6418-6420 is underlined. Total length of amplified region 2087 bp.

```
4595 : CCATGGTTCCTACCCCTCGATAATCGCGAGTACATTTTCATTAATTAACCCGTCACCCCAAACCAGGAACT
CTTCTTAGGCTCTTCCTTTTAAACGGATTAAAGGAAGCGATGAATCCGTCGTCCTTCTTCTCCTCTTTCTTAA
CGGGTTGAAAGAGGGCAGATAAGAGTACGGGTTGAGAGAAGGCAGATAAGAGTACGGACTTCCCACGACGAACCA
TATCATTTTCACGACCAGGTATTGGACTAATCCGGTCAATCGTCCTATAACGAAGTAGCAAATCGATAGGATAGG
ACCTACCAAATAGGATAGCATTATAATAACGAGAAAGTTCCTACGCGTGTATCTTTACAAAAAGATCGTGTATTT
CGTCATCCGTCAGATAATCCACGACGCTGTATCGGACTTCTACGGGAAGACGGGTCCATAGACCGTCGGGTCCGC
ACAGGTCGGCGATACGGCGAACGAGACGATCCGTTTTGAACCGTCGCTTGTGTAGGGATCGGAGAATCTTCTTTA
CTCGATCTCCGTAGACGCCTCGTACGTACTTCGAGAGAATCCGAACGTGTTTCCAACGAGGCCCTGGCGACTAGAC
CGAACGCTGTATCGAACACGCTGACCGCGTCCAGAACGTCCGCCCTCATGGCGTCTCGTTCCCTTATAACATTCGT
CCACAACTCTCTAAAGTAGTCGTAGATGTGAACGTATGTAAGATCTCCCTATCCGAGTCTTTTACGAAGCTTC
GAATGAGGGCTTTGTACAAGGTTCCGTGTGCTTTACGAACTCCGACATGCCGTACGACTTGATCGAGTAAATGT
TAGAGTTTACCACGTTCAACGCCCTCGTGTAGGCGTCCGGCGTGATCGTCTTTCGTAGCAGACGATGGACCATGT
ACTTGTTCGGTTTAAACACGGCCAGGTCTAGACAAGAGTTCCCCAGAAAGTTCACCACGTTGGGATCGGCTCCTC
TCTCCAGCAGATAGTCGAACGTTCCGCCGTCGTTGTGTATCACGCAATAGTTAATGGGTGTAAAGTCGAGAGAGT
TTTGTTCGTTTATGCTTATTTGTGTAAAGATAAAAATCCAACACCTTTCGATTAAAGACGCCGTGGCTGATCTTAC
GTTCCAACGTGTTAAACATGACGGTTCGCGTGTAAATCCTCGGGAGGAATCGACCCCGTGACGCAGAAGGAATCGTA
GAATCTGCACGTTTCGACGAACGCACGTGGGCGATCATGTAACATTGAAGGACGTTAAACCCGTACGTCCTTGTCTG
GGTTCGTGTACGGTGGCTCCCGCGCGGACCAACTCTCGTAACGTATCCAATACCACGTTCTTGAATGTATGTAGG
TTTGTACGGGCCCCGATCCGTCCGTTCCCCGAACGTACACGGACGTGCCTCGGTCGATCATGTAGCGTACAAACC
TCGCCGTATTAATACGGTCGTTTTCCAGCAGACAGAGCAAGGGATATCGTCCATCTGGCGCCATCGCGTTGACGT
TGGCTCCCGCATCGATGAGCTGTTTCGCCAACGCCTCCGCCCTCATTTCCCAAATCTTTATTGGATAGGACCGTAC
ACAGAGGAGACGTACGAGAACCACGAACCTCCGTTGACGTCCGCGCCTCGTGTAAAAACAAACGTAAACCCCTCCG
GGGTACAGTCGTATCTCGTTAGGTAAAGGCGAAAGGGGTATCCCGCGAACATAGTTCGTCCGGGTCGTATACGT
TCAAGAGACCATCGAGTACATCGTGTGCGGATACGGGGCGTACACGACACATACCCGTATAGATCCATCCTCGTC
CGGGGTTATATCAACTACACTTATTTTTTTCACCGTTGTATAAAAAATGGACACGCCGTGTTTCGTACACCCCGTGT
ATAGAGAGACGTCGTTGATGCGGTGCCATTTCCACCCGTTATACGTCTCCGCCGAGTTCGTCCGTCCCTCCGAATA
CGTACAGCGCGTTGTTGTAGACGCAGCAGATGGGATTCCTGCGCGACTCCGTCATATTCTCGACACGTTCCACG
CGCCCCGGTCATCGTTCGTATCGGTAGACGTGGTTCGTGTGCGGTGTGTTTCGTTTCGTACGTCCTCCGAC : 6681
```

### Flanking sequence for *M005R*

LH flanking sequence 299 bp: Forward: 155083-155100; Rev: 155363-155381. RH flanking sequence 308 bp: Forward: 156861-156881; Reverse: 157153-157169. TGA stop codon for *M006R* at 155344-155346 is underlined; ATG start codon for *M005R* at 155382-155384 is in red; TAG stop codon for *M005R* at 156832-156834 is in red; ATG start codon for *M004.1R* at 156854-156856 is underlined; *M004.1R* stop codon at 157124-157126 is underlined; *M004R* ATG start codon at 157130-157132 is underlined.

155083 : GTCGGAGGACGTACGAACGAACACACCCGACACGAACCACGTCTACCGATACGACGATGACCGGGGCGC  
GTGGGAACGTGTTCGAGAATATGACGGAGTCGCGCAGGAATCCCATCTGCTGCGTCTACAACAACGCGCTGTACGT  
ATTCGGAGGACGGACGAACCTCGGCGGAGACGTATAACGGGTGAAATGGCACCGCATCAACGACGTCTCTCTATA  
CACGGGGTGTACGAACACGGCGTGTCCATTTTTTATACAACGGTGAAAAATAAGTGTAGTTGATATAACCCCGG  
ACGAGGATGGATCTATAACGGGTATGTGTCGTGTACGCCCCGTATCCGACACGATGTACTCGATGGTCTCTTGAA  
CGTATACGACCCGGACGAACCTATGTTTCGCGGGATAACCCCTTTCGCTTTACCTAACGAGATACGACTGTACCCC  
GGAGGGTTTACGTTTGTAAAAACACGAGGCGGGACGTCAACGGAGTTCGTGGTTCTCGTACGTCTCCTCTGTG  
TACGGTCCATCCAATAAAGATTTGGGAAATGAGGCGGAGGCGTTGGCGAAACAGCTCATCGATGCGGGAGCCAA  
CGTCAACGCGATGGCGCCAGATGGACGATATCCCTTGCTCTGTCTGCTGAAAAACACCGTATTAATACGGCGAG  
GTTTGTACGCTACATGATCGACCGAGGCACGTCCGTGTACGTTCGGGAAACGGACGGATACGGGCCCGTACAAAC  
CTACATAATTCCAAGAACGTGGTATTGGATACGTTACGAGAGTTGGTCCGCGGGGAGCCACCGTACACGACCC  
AGACAAGACGTACGGGTTTAAACGTCTTCAATGTTACATGATCGCCACGTGCGTTCGTCGAACGTGCAGATTCT  
ACGATTCCTTCTGCGTCACGGGGTTCGATTCCCTCCGAGGATTACACGCGACCGTCATGTTTAAACAGTTGGAACG  
TAAGATCAGCCACGGCGTCTTTAATCGAAAGGTGTTGGATTTTATCTTTACACAAAATAAGCATAAACGAACAAA  
CTCTCTCGACTTTACACCCATTAACCTATTGCGTGATACACAACGACCGGCGAACGTTTCGACTATCTGCTGGAGAG  
AGGAGCCGATCCCAACGTGGTGAACCTTCTGGGAACTCTTGTCTAGACCTGGCCGTGTTAAACGGAAACAAGTA  
CATGGTCCATCGTCTGCTACGAAAGACGATCACGCCGACGCCTACACGAGGGCGTTGAACGTGGTAAACTCTAA  
CATTTACTCGATCAAGTCGTACGGCATGTCCGAGTTCGTAAAAGCGACACGGAACTTGTACAAAGCCCTCATTCG  
AAGCTTCGTAAAAGACTCGGATAGGGAGATCTTTACATACGTTTACATCTACGACTACTTTAGAGAGTTTGTGGA  
CGAATGTATAAGGGAACGAGACGCCATGAAGGCGGACGTTCTGGACGCGGTCAGCGTTCGATACAGCGTTCCG  
TCTAGTCGCCAGGCCTCGTTGGAAACACGTTCCGATTCTCTCGAAGTACGTACGAGGCGTCTACGGAGATCGAGT  
AAAGAAGATTCTCCGATCCCTACACAAGCGACGGTTCAAAAACGGATCGTCTCGTTCGCCGTATCGCCGACCTGTG  
CGGACCCGACGGTCTATGGACCCGTTCCCGTAGAAGTCCGATACAGCGTCGTGGATTATCTGACGGATGACGA  
AATACACGATCTTTTTGTAAAGATACACGCGTAGGAACTTTCTCGTATTATAAATGCTATCCTATTTGGTAGGTC  
CTATCCTATCGATTGCTACTTCGTTATAGGACGATTGACCGGATTAGTCCAATACCTGGTCTGAAAAATGATAT  
GGTTCGTCGTGGGAAGTCCGTAATCTTATCTGCCTTCTCTCAACCCGTACTCTTATCTGCCCTTTTCAACCCGT  
TTAAGAAAGAGGAGAAGAAGGACGACGGATTTCATCGCTTCCCTTAAATCCGTTTAAAAAGGAAGAGCCTAAGAAGA  
GTTTCTGGTTTGGGTGGACGGGTTAATTAATGAAAAATGACTCGCGATTATCGAGGGTAGGAACCATGG :

157169

### Flanking sequences for *M012L*

LH flanking sequence 330 bp: Forward: 13800-13824; Reverse: 14104-14129. RH flanking sequence 288 bp: Forward: 14620-14644; Reverse: 14883-14907.

Stop codon of *M012L* at position 14137-14139 is shown in red; start codon of *M012L* is shown in red at position 14581-14583. Start codon of *M011L* is underlined at position 14122-14124 and stop codon of *M013L* is underlined at 14603-14605. Total length of amplified region 1108 bp.

13800 : TGACTTCCGAGAACATAGCCCGTCCGTATGTTGTATCTTGTTTAACTTTTTAATCACGCTCGCCAACA  
GAGTTAATGTGGCTAGTTTACGGACGGTGACCGTCTCGTAACCGAACGTC AACGTAATTTTTAATGTTACGGA  
TATTATACGACAAAATGTCCCGTTCATGATATCATAACAACGTGTCGTAGTTTTTTGCGTACGTTTCGTTGATAA  
AATCGCAACAATACTCAACTGACACAGTAGATCCATTTCCGTACACTCCGTTATATCCACGTCGTTACGATAAT  
CGTATACGGCCGTTCTTTAAACGAGACATCATCGACGATAAAAAATAAACCCTTGAACCGAATCCGTTATGGC  
CGCGTGCGGTTCGTGTTCAACGTGCCCACGCTTTCTATATTTACACTTGTACTTTTTCAAAAATTATTTGTGCTA  
CCTTATCTCCTTTTTTTACATAGTACGTATCATTTCATTTGATGAGGATGATTCCCTATATTGCCCTGTAAT  
CTCTGTGATAAATCCACCGCCACGTGCGATGGAGAATTTCAACGACATACCGGATCGTGGCGGATCCTGCCGT  
AACACCCCTTCCGGAATACCCAACCTATATCCGTTTTAACCACACACCGTTTGTGTGGAGGAATGATATACGCAT  
AGGCGCTGTATAGGTCGTATCCTGCCGATCGGCCGTGGCCTTTGTAGGAAGACTAGCGTTCGGAGAACTTGA  
CGCACTTGACCGTTCATGACTTCTTTGTAGAAAAACATACTTTCTTTTAACTTTTGTAAAAATAACAATTTG  
CGACATACACCTCGTTGCTTCTTCGGCGTTTTTGTAGTACACCAATGGCCATTTTTTCTTTTCGTGACTGCTTTTGT  
ACCGCTTCTGTACCGCTTCCATGGTGGCTTCCGCGTCTGCTTCCGCGCGCCAAATAGGGAATCCACGTACACT  
TTATTAGGAATGAAGTCGGTAACCGTTTTCAAACTCAATATAGTCGGTTCCAGATATTGTTTCAGATATCTTT  
GGCCAGGTCTATTCTGTCGATCTTTTCTTTTCCGCCGTTTCGATGTATAGGTCCTCCATCG : 14907

### **M014L flanking sequences**

LH flanking sequence 284 bp: Forward 14847-14867; Reverse: 15109-15130. RH Flanking sequence for 265 bp: Forward 16676-16698; Reverse: 16919-16940.

Start codon of *M013L* is underlined at position 14981-14983. *M014L* stop codon is shown in red at position 15013-15011. *M014L* start codon TAC is shown in red at 16562-16564. *M015L* stop codon is underlined at position 16617-16619. Total length of amplified region 2094 bp.

This sequence overlaps the *M014L* 3' end but doesn't include the mutated sequence in Ur.

```
14847 : GCCAGGTCTATTCTGTTCGATCTTTTCTTTTTCCGCCCGTTCGATGTATAGGTCCTCCATCGCTAGATAG
ATGAACATCTTAAACTGGTAATCGGATAGATTTTCCAACACGGTTATAATGACGCCCTCGGTGCTCCATTTTGATA
AAAAGACGTCGTTTTTCACTTTAAAAAATATCACATTCGTGTATATTTCTACGATATCCCACGTTTTTGATATA
TCGTTATATACCTCTATCTCGTTCGATGTATCTGTACATGAAACCTCCCAGAACGACAAGAGACTCCTCAAGGAGG
CACGCCGATGCATTGACCCGGGGGAGATTCAACGAAGATTCCATATGCCAGGTGCAGGTCACGTGTATTATACCCC
TCTACCATGGTATACACCTTTATATCGTCTATGTACGAGACTCCTCCCAAGACGTAATGACGTCTCCGCGTGAT
ACACAACATCCCCCGCATGAGAGTATTTTCATGGGATTACACGGTCTCCATTGTTTGGAGAACGGAGAGTAGACT
TCCACCGTTTTTATCAAACCTCGTCAATGCCTCCCATGATGTACAACAGGTCGTTAACGTGACGGAACATGCGCCG
TATCTGGGCTGTAGCAGAGTCGATTTCGAGTGTCCAGGCGTCGTCCTCCCGGGGCCAACTCTCTACGGAGATGAGG
CTTCCGTTTCGTCCTTATTCCTCCTATTGCGTAAATTTCTATCACGTAAGGCGGATACAGTGGGTAACTTTCTTGCC
ACGTTTTAACTCGGGCACGAGTACGCGCTGCGACGTACGAGTGTTTATACCCACCGCGTACGGCACGGGTTTCGTCC
AATTCGTTTCATGCCCCCTATCAGATACACCACGTCGTTTAAACGTGACGCATCCACACCCCGCTATAATGGTCGTCT
ATAAGTCTATTAATTCGCTCCACGTCCTCCTTGGGAGATGTTGGCGGATCGTCGAGGGGTTTTGGAGGAGGAACG
TCTTTCGCCGTGTAATATTTACGACGACGTGTTGTAGACCGTCTGTCTTGTGTTGATGAACTTGAACACCCCGTGC
GATCTCCACCGTCTTAACACTTTACTTCCACGCTCCGTCAAAATACGTGTCTCGTAACACGTCGTTTCATAAGACCT
AACGTTTTTCATTTTTATTTTTGTTTTTATACCTGGCCCATTTGATAATGAACTCTAATACCACGTCTTCCCTTGGA
ACGTTTACCTCGTCGCCTTTGAGTATACTACTCAGTTTCGAAATATGACAAGGGCATCAACGTTTTATCCGCCGCT
AAGACTTCGAACCGGCTCCGATGTACTCCATGGCGCGCTGTACACTTTGTAACATTTGTTGCTCAATCCGGTT
TTTAGAACCCCCACACACACGTCGTCGGTCATGTGGTTTATTATATATTTCTGTACACGAGGCTGTACGAATTCC
AAAAACGTTTTTACAAGCGATGGTAAATAGTTTTTACTCGTTCGAGGAGTTACGTCGAATGATCCTGTGTGATG
TACTTGATAATTTGACATACGACGTCATGGTTCGACTGTATCATGACCTTTACGTTATTCCTGATAAACATCTTA
AAATACACGGATGCCGAAAGAAGTAAATCCTTTTCGTACCGTAACAGCCCCATGGTGGTAACGACGGAGACCACT
TCGGGCGTACACGGAGGTAGCATATCAAGTCGTTTCTTGTACTCATGATCGACGTACGTTACCGTCTCCGCATT
TTGTGTATTTTTATATACTTATTTGCTCAACTTTTTTGTAAAAAATACTTAAAAATCTTCATCCAATGAAAA
TGTGTTATCTTGTGTTATGAACAACACCCATTCTTTGGTAATCGCTAACGCGTTTTTTCGAAAAAGTTAGTTTTGCC
CTCTAGAGATATATTCTCCATGAAATCAAAGGATTAGTAGCTGAAAAATTTTAGAGTACCCAGTTTCGGCGAT
TAATCTATCCGCTACAAATTCATATATTGGCACATTAATTCGCAGTTCAATTCCTATAAGACGTACCGGAAGAGC
TTCCGTCAAAAATCGTTGCTCTATAACAACCGCTTCCTTAATAATACTAGCAACTCGTTTCTCGGACGGTGTATG
:16940
```

### **M153R flanking sequences**

LH flanking sequence 258 bp: Forward: 148269-148291; Reverse: 148505-148526. RH flanking sequence 263 bp: Forward: 149238-149266; Reverse: 149479-149500. *M153R* ATG start codon at 148157 and TAG stop codon at 149137 are shown in red. TAA stop codon for *M152R* at 148479 is shown underlined. *M154L* stop codon at 149230 is underlined. Total length of amplified region 1232 bp.

```
148269 : CGGAAACGGGACACAACGTAACGTGTCTAACTCCCGCACACGTCAATCTATGTCATCGTTGTCGTATA
ACGATTACCAAAACGGGCGTGGACGCGACCGCGTTCTCATGCGTCGACGGCGATACATGCACCGAACACGACACG
ACCGCGTCAACGTGTACGATTATTATAAAAAACGACGGGTCTAGACTTTTGTTTTATGGGAAAACCTTAAAAAATA
TGTCAATTAAGTAAATAGGTTGTGTAAACATGGCTACTGTTGTAAACATGGACACTGTTGTAAACCTGGACGAT
GTAAGCCTGGCGGATAAATGTTGTTGGATATGCAAAGAAGCCTGTGACATCGTACCCAATTACTGCAAATGCCGA
GGGGATAATAAGATCGTGCATAAGGAATGTTTGAAGAGTGGATCAACACAGATGTCGTGAAAAACAAATCGTGT
GCCATATGCGAGAGTCCGTACAACCTAAAGCGGCGGTACAAAAAATAACCAAATGGAGGTGTTATAAACGCGAC
TGCCACGATAGTTTATTAGTAAACATGTCGCTTGTGTTGATCGTGGGTGGAATGGGGGGCTATTTACTCATTTC
```

ACGGAAATTGTAAAACCTTATAGCGTCTGAGGAAGTTAGCAACATTGCTAAAGTATTTCTCGTCTCTGCGTCCATG  
GGTCCATTTTATGGTATCCGCATTAACCATGGTACGTGCGTGTATAGACTGTCGTACCTACTTCATAGCTACTCGT  
GAACGTAATACGATTACAGAGGTGGCAGAGATGGAAGATGTGGAGGAGGTGGAGGAGGTGAACGATGACGACGGC  
GATGAATACGTGACGCTGTGAGGAAATCGTGTGGAGTTACCCGCTTAGAAAAACGAATCGAATAAAAACCCG  
TGTACACACGGACGTTAATTTTTTTTTGTGGTTAAAAAATGACCACATTTACGCTTTTTTTTTAACGCGTTATATA  
AGGTATCTCGTTTGTCTATAACAAAGATCGTAACTGACCTTTTTTTATATCGAGAAAACATAACGTTTAGTTCATCC  
TCAAACGTAACACCGTAACTGCCTCGGACATCCTCCTTGTGTCGTACACAAACATACTAATCGGATGCGTGAAA  
TGAGGATTCACTTTAATCGGATTGGTTTCTAGGTTAACACATGTTACACAGGATCCTAAGATGGTTATGGACACA  
TCCTTGTGTGATGTAACGAGTCGGGAAGTTGATTGCCG:149500

### **M156R Flanking sequences**

LH flanking sequence 252 bp: Forward: 149948-149970; Reverse: 150178-150199. RH flanking sequence 252 bp: Forward: 150358-150381; Reverse: 150597-150618. The annotated ATG for *M156R* at 149988 is underlined however translation occurs from the second ATG shown in red at position 150069. The TGA stop codon at position 150296 (A) is shown in red. The ATG for *M008.1R* at position 150304 is shown in blue and underlined. The sequence 5' to the *M156R* annotated ATG is intergenic. Total length of amplified region 671 bp.

149948: GTAACAACACGTGTGTCGTAGCGTATACATAATGCCGTAAATGACAGTCATAAAAACCATCGAGTCGTC  
CCAGGCCGAGGAAAAACAAAAATATAAAAAGTAAATACATACAGAACGAGCGCCATGGATCTCTCTCCGGGAAGTG  
TCCACGAGGGTATCGTATATTTTAAAGACGGAATATTCAAAGTCCGCCTACTCGGATACGAGGGACACGAGTGTA  
TTCTTTTGGACTATCTGAACTACAGGCAAGACACGTTGGATCGGTTGAAGGAACGACTCGTGGGACGCGTGATTA  
AAACGCGAGTCGTTTCGCGCGGACGGTTTATACGTGGACCTGCGACGTTTTTTTTTGAAGGGTTAAATGAAGTATCTG  
GTCCTCGTCTTATGTTTAAACGTCGTGCGCGTGTGCGAGATATCGGACTATGGACGTTCCGATACGTCCTACAACGAA  
AGCGACAACGTCGTGTTCTCACCGTACGGCTTGACCTCCGCGTTGTCCGTGTTACGGATCGCGGCGGGCGGTAAC  
ACGAAACGAGAAATAGACGTCCCCGAATCCGTCGTGGAGGACTCCGACGCCTTTCTCGCGTTACGGGAGTTGTTTC  
GTAGACGCATCCGTTCCGTTACGTCCCGAGTTTACGGCGGAGTTCTCCTCGCGATTCAATACCTCCGTGCAACGC  
GTG:150618