Predicting potential anti-codon shifts (alloaccepter shifts) in TryTryp tRNA models

Method:

1) Using the co-aligned 3542 genes from intersection of two gene-finders Aragorn and tRNAscan-SE (tRNA genes with the same predicted identity by both gene-finders), generated profiles for 22 models similar to tfam.

   Number of sequences used for making each profile:

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
211	64	105	160	104	228	80	171	183	336	97	125	200	162	347	228	218	241	52	76	78	76

2) Calculated the score of each sequence (from training set), against each profile using leave-one-out method.
3) Using the output of step 2, calculated the z score of each sequence against each model.
4) For each sequence compared <u>the model with the highest z score</u> or <u>the model with the highest score</u> with <u>the model reported by two gene-finders.</u>
5) Defined the potential anti-codon shifts for each sequence as <u>models with the highest z score or highest score which do not match the model predicted by gene-finders</u>.

```
geneid                                              sequence                                                                                           func   potential_shifts_z potential_shifts_s
LenriettiiLEM3045_LenLEM3045_36_95                  CCTTCTGTAGCTCAATTGGTGGAGCATGTGACTGTAGATCACAGGGCGCTGGTTCGATTCCGGCCGGAAGGA                            Y      I                  Y
PconfusumCUL13_Pcon_scaffold00007_2_43              AGCATCGTAGCGCAGT-GGTAGCGCGATGGGCTCATAACCCATAGGCGTTGGATCGAAACCAACCGATGCTA                            X      M                  X
PconfusumCUL13_Pcon_scaffold00007_3_45              GGGAAGATAGTTCAGT-GGCAGAAACACTAGATTGGGGCTTCAGG-ACCCGGGTTCGATTCCCGGTCTTCCCT                           P      H                  H
TbruceiTREU927_Tb927_11_bin_v5.1_73                 GCGCCCGTAGCTCAGCTGGAAGAGCGCTGCCCTCCGGAGGCAGAGGCTCAGGTTCGAATCCTGTCGGGCGCA                            R      I                  I
TbruceiTREU927_Tb927_11_bin_v5.1_74                 GTGGCTATAGCTCAGTTGGTAGAGCCCTGGATTGTGATTCCAGTTGCCGTGGGTTCGAATCCCATTAGCCACC                           H      H I                W
TbruceiTREU927_Tb927_11_bin_v5.1_75                 GCGAAGGTGGCGGAATTGGTGACGCGCTAGCTTCAGGTGTTAGT-GTGGGGGTTCAAGTCCCCCCCCTCGCCA                           L      V                  L
TbruceiTREU927_Tb927_11_bin_v5.1_76                 CGGCGAGTAGCGCAGCTGGTAGCGCAACTGGTTTGGGACCAGTGGGCGGAGGTTCGAATCCTCTCTCGCCGA                            P      H I M P R          R
TcruziCLBrener_Tcruzi_11019_2                       GCGAGCGTAGCTCAGCTGGTAGAGCACACGACTGATAATCGTGAGGCGATGGTTCAAGTCCATTCGTTCGCA                            I      M                  M W
TcruziCLBrener_Tcruzi_20084_17                      GCCACCATAGCCAAAC-GGCAAGGCATAGGTCTGCAACACCT-TGATACCGGTTCGAATCCGGTTGGTGGCT                            C      I                  S
TcruziCLBrener_Tcruzi_22215_21                      GGACCCATAGCTCAAT-GGTAGAGCAACCGGCTCATAACCGGTCGGTACAGGTTCGAGTCCTGTTGGGTCCA                            M      I                  C
TcruziCLBrener_Tcruzi_22215_22                      GGCATTGTAGCTCAGTTGGTAGAGCGTCCGGTTCATACCCGGAAGGCAAGAGTTCGACTCTCTTCCGGTGCTA                           M      I M                I M
TcruziCLBrener_Tcruzi_22215_23                      GGGGGGGTAGCTCACCTGGCAGAGCGCCTGCCTTGCACGCAGGAGGAGGGGTTCGACCCTTCCCTCTCCA                              A      I V                A
TcruziCLBrener_Tcruzi_22215_24                      GGAGAAGTAGCTCAGCTGGTAGAGCACTTGACTTGGGCTCAAGCGGCGCAGGTTCGAATCCTGTCTTCTTCA                            P      H I                W
TcruziCLBrener_Tcruzi_22215_25                      GGGCAATTACCCAAGA-GGCAAGGGAGCAGTCTTGAAAACTGCAGGCGGGGGTTCAAATCCCTCATTGCCCG                            S      H P                T
TcruziCLBrenerNon-Esmeraldo-like_TcChr21-P_8        GCGCAGTTGGTCCAGT-GGGAGAATTCCCGCATCACAAGCGGGAGGCCCGGGTTCGATTCCCGGACTGCGCA                            V      G                  G
TcruziEsmeraldo_ANOX01006378.1_11                   GCGGCAATAGCTCAGTTGGTAGAGCAACGACCTTGCCAAGGTCGGGGCGCGAGTTCGAGTCTCGTTGCCGCT                            G      H I                H
TcruziEsmeraldo_KB205213.1_31                       GCGCAGTTGGTCCAGT-GGGAGAATTCCCGCATCACAAGCGGGAGGCCCGGGTTCGATTCCCGGACTGCGCA                            V      G                  G
TcruziEsmeraldo_KB205435.1_59                       GCGCAGTTGGTCCAGT-GGGAGAATTCCCGCATCACAAGCGGGAGGCCCGGGTTCGATTCCCGGACTGCGCA                            V      G                  G
TcruziEsmeraldo_KB205928.1_89                       GCGCAGTTGGTCCAGT-GGGAGAATTCCCGCATCACAAGCGGGAGGCCCGGGTTCGATTCCCGGACTGCGCA                            V      G                  G
TcruziJRcl4_AODP01003975.1_2                        GCCTCGGTGGTGTAAT-GGTAGCACGGCGACCGGATATCTGTA-ATGCGGGTTCGAGTCCCGTCTGAGGTT                             S      H R V              R
TcruziJRcl4_AODP01004654.1_4                        GCTCTCATAGCTCAGTCGGTAGAGCGTGGGTCTAACAAGCCCAAGGCACAGGTTCGACCCCTGTTGGGAGCA                            V      I                  I
TcruziSylvioX10-1_TcX10_chr15_18                    TTCTCGGTAGTATAGT-GGTAGTATACCCGCCTCACGCGGGT-CACGCGGGTTCAATTCC-GGCCGGGAAG                             D      G V                D
TcruziTulacl2_KB851446.1_107                        GCGCAGTTGGTCCAGT-GGGAGAATTCCCGCATCACAAGCGGGAGGCCCGGGTTCGATTCCCGGACTGCGCA                            V      G                  G
TvivaxY486_TvY486_bin_archived_contig_1183_36       GCCCGGATAGCTCAGTTGGTAGAGCATCAGACATTTTAATCTAAGGGTGAGGGTTCAAGTCCTTGTTTGGGTG                           K      H I R              R
TvivaxY486_TvY486_bin_archived_contig_3413_48       GCCGTGATCGTATAGT-GGTAGTACTCTGCGTTGTGGCCGCAGC-ACCCCGGTCCGAGTCCGGGTCATGGCA                            H      H V                V
TvivaxY486_TvY486_bin_archived_contig_3413_51       GCCCGGATAGCTCAGTTGGTAGAGCATCAGACTTTTAATCTAAGGGTGAGGGTTCAAGTCCTTGTTTGGGTG                           K      H I R              R
TvivaxY486_TvY486_bin_archived_contig_4539_66       GTCTCTGTGGCTCAATAGGTAGCGCGTTCGGCTGTTAACTGAAAGATGGTGGTTCAAGCCCACCCAGGGACG                            N      I                  R
TvivaxY486_TvY486_bin_archived_contig_4539_67       TTCCATATGGTCTAGC-AGTAGGATTCCTGATTTTCACCCAGGT-GCCCGGGTTCGACTCCCGGTATGGGAA                            E      G                  E
TvivaxY486_TvY486_bin_archived_contig_7832_92       GTCGAAATAGCTCAGTTGGGAGAGCGTTAGACTAAAGATCTAAAGGCCCTGGTTCGATCCCGGGTTTAGGCA                            F      H I                F
TvivaxY486_TvY486_bin_archived_contig_7832_95       GCCGAAATAACTCAGTTGGGAGAACGTTAGACTGAAGATCTAAAGGCCCTGGTTCGATCCCGGGTTTCGGCA                            F      H                  F
                                                    >>>>>>..>>>>........<<<<.>>>>.......<<<<<....>>>>.......<<<<<<<<<<<<<.
```

Figure 1: the list of sequences along with their predicted model by gene finders as column "func". The last two columns shows potential anti-codon shifts based on z-score and score, in order. The last row shows the final structure of the aligned sequences.

```
A    C    D    E    F    G    H    I    K    L    M    N    P    Q    R    S    T    V    W    X    Y    Z    geneid
-28  -8   -381 -19  -19  -11  -8   -5   -10  -17  -7   -13  -10  -11  -7   -10  -27  -8   -33  -77  |-9   -38  LenriettiiLEM3045_LenLEM3045_36_95
-23  -8   -432 -22  -29  -11  -9   -6   -9   -14  -5   -10  -10  -12  -9   -9   -22  -7   -33  -6   -311 -37  PconfusumCUL13_Pcon_scaffold00007_2_43
-25  -8   -329 -18  -25  -7   -2   -8   -14  -13  -9   -20  -8   -10  -12  -11  -21  -7   -32  -90  -329 -42  PconfusumCUL13_Pcon_scaffold00007_3_45
-23  -8   -401 -23  -24  -6   -8   -4   -11  -17  -7   -15  -7   -10  -6   -9   -17  -5   -31  -62  -266 -36  TbruceiTREU927_Tb927_11_bin_v5.1_73
-20  -8   -360 -22  -25  -10  -6   -6   -10  -12  -7   -15  -9   -10  -8   -9   -26  -8   -29  -72  -284 -37  TbruceiTREU927_Tb927_11_bin_v5.1_74
-26  -11  -411 -24  -33  -8   -8   -9   -14  -9   -8   -17  -9   -15  -10  -10  -30  -7   -43  -84  -354 -45  TbruceiTREU927_Tb927_11_bin_v5.1_75
-18  -9   -432 -21  -33  -12  -8   -8   -14  -17  -8   -16  -8   -13  -8   -11  -25  -9   -37  -66  -266 -35  TbruceiTREU927_Tb927_11_bin_v5.1_76
-20  -8   -381 -20  -24  -8   -7   -6   -8   -17  -5   -14  -9   -12  -10  -9   -24  -7   -29  -63  -240 -36  TcruziCLBrener_Tcruzi_11019_2
-34  -8   -411 -23  -27  -10  -8   -7   -13  -17  -8   -16  -10  -11  -9   -8   -25  -8   -42  -77  -302 -29  TcruziCLBrener_Tcruzi_20084_17
-20  -5   -370 -18  -25  -8   -7   -4   -9   -19  -8   -15  -8   -9   -7   -9   -18  -8   -33  -59  -275 -34  TcruziCLBrener_Tcruzi_22215_21
-16  -8   -370 -18  -25  -9   -7   -5   -9   -16  -5   -15  -8   -13  -8   -11  -19  -6   -40  -54  -293 -32  TcruziCLBrener_Tcruzi_22215_22
-10  -10  -360 -18  -27  -7   -7   -6   -11  -16  -7   -18  -8   -14  -9   -10  -19  -6   -39  -79  -311 -36  TcruziCLBrener_Tcruzi_22215_23
-18  -7   -401 -19  -24  -9   -6   -6   -12  -13  -9   -18  -8   -12  -7   -7   -24  -8   -27  -77  -240 -39  TcruziCLBrener_Tcruzi_22215_24
-21  -10  -411 -25  -33  -9   -7   -8   -14  -15  -9   -17  -7   -14  -11  -9   -19  -8   -39  -82  -382 -38  TcruziCLBrener_Tcruzi_22215_25
-19  -10  -257 -14  -24  -1   -6   -6   -12  -16  -8   -18  -5   -10  -10  -11  -20  -2   -37  -88  -346 -40  TcruziCLBrenerNon-Esmeraldo-like_TcChr21-P_8
-23  -9   -381 -21  -24  -8   -6   -6   -11  -17  -7   -17  -9   -14  -9   -11  -20  -7   -34  -92  -266 -34  TcruziEsmeraldo_ANOX01006378.1_11
-19  -10  -257 -14  -24  -1   -6   -6   -12  -16  -8   -18  -5   -10  -10  -11  -20  -2   -37  -88  -346 -40  TcruziEsmeraldo_KB205213.1_31
-19  -10  -257 -14  -24  -1   -6   -6   -12  -16  -8   -18  -5   -10  -10  -11  -20  -2   -37  -88  -346 -40  TcruziEsmeraldo_KB205435.1_59
-19  -10  -257 -14  -24  -1   -6   -6   -12  -16  -8   -18  -5   -10  -10  -11  -20  -2   -37  -88  -346 -40  TcruziEsmeraldo_KB205928.1_89
-32  -11  -319 -18  -25  -8   -7   -8   -14  -14  -9   -14  -9   -9   -7   -10  -27  -7   -51  -77  -284 -35  TcruziJRcl4_AODP01003975.1_2
-21  -7   -411 -20  -20  -8   -8   0    -9   -16  -4   -14  -9   -9   -6   -9   -18  -5   -30  -68  -231 -34  TcruziJRcl4_AODP01004654.1_4
-26  -9   -10  -10  -29  -5   -7   -8   -13  -18  -11  -16  -6   -11  -12  -10  -23  -5   -47  -90  -320 -41  TcruziSylvioX10-1_TcX10_chr15_18
-19  -10  -257 -14  -24  -1   -6   -6   -12  -16  -8   -18  -5   -10  -10  -11  -20  -2   -37  -88  -346 -40  TcruziTulacl2_KB851446.1_107
-21  -8   -360 -18  -20  -10  -6   -6   -8   -15  -7   -13  -10  -10  -6   -8   -20  -7   -33  -81  -257 -31  TvivaxY486_TvY486_bin_archived_contig_1183_36
-36  -9   -329 -17  -28  -8   -6   -8   -14  -13  -10  -19  -9   -10  -14  -13  -31  -6   -45  -91  -373 -38  TvivaxY486_TvY486_bin_archived_contig_3413_48
-21  -8   -360 -18  -20  -10  -6   -6   -8   -15  -7   -13  -10  -10  -6   -8   -20  -7   -33  -81  -257 -31  TvivaxY486_TvY486_bin_archived_contig_3413_51
-27  -8   -360 -21  -27  -11  -9   -6   -10  -18  -8   -11  -10  -12  -7   -10  -24  -8   -42  -62  -275 -37  TvivaxY486_TvY486_bin_archived_contig_4539_66
-26  -11  -236 -9   -32  -5   -6   -8   -13  -14  -11  -18  -6   -9   -11  -9   -31  -6   -49  -99  -364 -41  TvivaxY486_TvY486_bin_archived_contig_4539_67
-22  -6   -411 -19  -6   -9   -5   -5   -11  -11  -7   -13  -9   -12  -8   -9   -21  -7   -30  -73  -222 -33  TvivaxY486_TvY486_bin_archived_contig_7832_92
-24  -7   -411 -17  -5   -8   -4   -5   -11  -11  -7   -14  -10  -11  -9   -9   -24  -7   -32  -77  -231 -34  TvivaxY486_TvY486_bin_archived_contig_7832_95
```

figure 2: z-score of each tRNA sequence against each profile.

```
A    C    D    E    F    G    H    I    K    L    M    N    P    Q    R    S    T    V    W    X    Y    Z    geneid
-176 -124 -175 -153 -104 -169 -148 -106 -135 -173 -131 -143 -177 -151 -110 -145 -159 -139 -118 -157 -8       -168 LenriettiiLEM3045_LenLEM3045_36_95
-150 -129 -199 -175 -160 -175 -174 -123 -121 -145 -95  -106 -180 -155 -130 -127 -134 -131 -118 -18   -156 -160 PconfusumCUL13_Pcon_scaffold00007_2_43
-161 -123 -152 -147 -134 -113 -40  -153 -181 -136 -156 -207 -144 -139 -174 -160 -128 -125 -114 -182 -165 -184 PconfusumCUL13_Pcon_scaffold00007_3_45
-148 -123 -185 -182 -133 -110 -155 -81  -142 -174 -126 -156 -132 -137 -94  -131 -110 -105 -109 -126 -135 -156 TbruceiTREU927_Tb927_11_bin_v5.1_73
-133 -128 -166 -171 -137 -155 -122 -120 -130 -129 -128 -156 -153 -136 -126 -127 -155 -138 -105 -147 -143 -162 TbruceiTREU927_Tb927_11_bin_v5.1_74
-169 -176 -189 -191 -177 -138 -157 -167 -175 -104 -142 -183 -156 -191 -150 -148 -173 -123 -150 -170 -178 -197 TbruceiTREU927_Tb927_11_bin_v5.1_75
-121 -148 -199 -166 -176 -186 -151 -147 -173 -173 -148 -174 -139 -168 -118 -159 -148 -151 -129 -134 -135 -154 TbruceiTREU927_Tb927_11_bin_v5.1_76
-132 -129 -175 -162 -133 -133 -145 -116 -108 -172 -102 -149 -167 -160 -142 -132 -142 -123 -102 -130 -122 -157 TcruziCLBrener_Tcruzi_11019_2
-212 -122 -189 -179 -147 -158 -152 -129 -170 -169 -146 -167 -176 -147 -140 -116 -150 -149 -146 -156 -152 -127 TcruziCLBrener_Tcruzi_20084_17
-131 -86  -171 -146 -135 -134 -145 -93  -115 -193 -143 -163 -144 -122 -107 -126 -114 -142 -116 -121 -139 -148 TcruziCLBrener_Tcruzi_22215_21
-106 -124 -171 -146 -138 -146 -137 -103 -118 -159 -103 -161 -146 -170 -124 -151 -120 -115 -138 -112 -148 -143 TcruziCLBrener_Tcruzi_22215_22
-73  -159 -166 -144 -146 -127 -130 -127 -147 -166 -137 -190 -144 -187 -129 -140 -118 -119 -136 -161 -156 -158 TcruziCLBrener_Tcruzi_22215_23
-122 -119 -185 -151 -133 -149 -117 -124 -149 -133 -162 -186 -144 -161 -105 -113 -145 -147 -97  -156 -122 -170 TcruziCLBrener_Tcruzi_22215_24
-135 -156 -189 -195 -177 -143 -142 -160 -174 -154 -162 -177 -132 -181 -155 -136 -116 -147 -136 -179 -191 -165 TcruziCLBrener_Tcruzi_22215_25
-124 -153 -119 -114 -129 -31  -114 -118 -149 -168 -143 -191 -97  -140 -145 -162 -125 -58  -129 -179 -174 -174 TcruziCLBrenerNon-Esmeraldo-like_TcChr21-P_8
-151 -140 -175 -169 -130 -140 -115 -120 -140 -170 -122 -178 -163 -179 -139 -155 -121 -127 -121 -186 -135 -151 TcruziEsmeraldo_ANOX01006378.1_11
-124 -153 -119 -114 -129 -31  -114 -118 -149 -168 -143 -191 -97  -140 -145 -162 -125 -58  -129 -179 -174 -174 TcruziEsmeraldo_KB205213.1_31
-124 -153 -119 -114 -129 -31  -114 -118 -149 -168 -143 -191 -97  -140 -145 -162 -125 -58  -129 -179 -174 -174 TcruziEsmeraldo_KB205435.1_59
-124 -153 -119 -114 -129 -31  -114 -118 -149 -168 -143 -191 -97  -140 -145 -162 -125 -58  -129 -179 -174 -174 TcruziEsmeraldo_KB205928.1_89
-200 -169 -147 -145 -138 -133 -133 -148 -176 -150 -157 -151 -159 -125 -110 -144 -158 -128 -176 -157 -143 -155 TcruziJRcl4_AODP01003975.1_2
-136 -105 -189 -160 -107 -137 -153 -15  -115 -160 -86  -151 -160 -119 -103 -131 -113 -107 -107 -138 -117 -152 TcruziJRcl4_AODP01004654.1_4
-167 -137 -7   -90  -159 -95  -131 -159 -162 -184 -193 -174 -118 -141 -170 -145 -140 -99  -161 -182 -161 -177 TcruziSylvioX10-1_TcX10_chr15_18
-124 -153 -119 -114 -129 -31  -114 -118 -149 -168 -143 -191 -97  -140 -145 -162 -125 -58  -129 -179 -174 -174 TcruziTulacl2_KB851446.1_107
-140 -129 -166 -148 -109 -158 -123 -115 -110 -151 -133 -141 -174 -132 -94  -125 -121 -130 -115 -164 -130 -139 TvivaxY486_TvY486_bin_archived_contig_1183_36
-226 -137 -152 -140 -152 -135 -123 -155 -183 -140 -184 -197 -157 -136 -192 -179 -181 -121 -155 -184 -187 -165 TvivaxY486_TvY486_bin_archived_contig_3413_48
-140 -129 -166 -148 -109 -158 -123 -115 -110 -151 -133 -141 -174 -132 -94  -125 -121 -130 -115 -164 -130 -139 TvivaxY486_TvY486_bin_archived_contig_3413_51
-171 -124 -166 -168 -149 -177 -183 -125 -125 -185 -138 -118 -175 -157 -106 -145 -144 -144 -145 -127 -139 -160 TvivaxY486_TvY486_bin_archived_contig_4539_66
-166 -167 -110 -77  -173 -93  -128 -146 -170 -142 -191 -194 -117 -119 -156 -138 -180 -109 -170 -200 -183 -179 TvivaxY486_TvY486_bin_archived_contig_4539_67
-143 -104 -189 -150 -36  -149 -98  -107 -141 -119 -132 -143 -166 -156 -119 -127 -129 -129 -107 -147 -113 -147 TvivaxY486_TvY486_bin_archived_contig_7832_92
-154 -107 -189 -140 -31  -138 -90  -112 -147 -121 -137 -149 -170 -151 -135 -134 -145 -130 -112 -156 -117 -151 TvivaxY486_TvY486_bin_archived_contig_7832_95
```
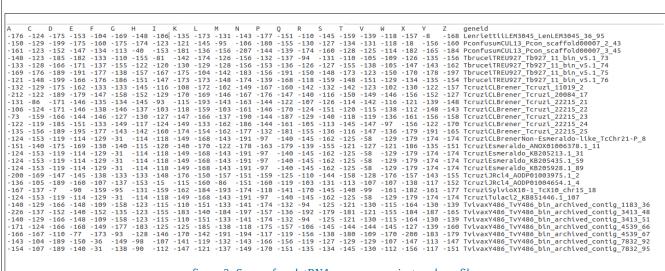
figure 2: Score of each tRNA sequence against each profile.