Values for GA Variables

We derived values for each of the variables utilized in our GA by estimating frequencies from the training data sets. These data were derived from the *A. thaliana* mitochondrial genome data. Each variable listed here corresponds to one of the six variables discussed in the text and shown in Figure 1.

Hydrophobicity values were based on the frequency with which an edited codon yielded a known hydrophobic amino acid. For example, 34% of all edited codons yielded a serine, so the value for serine would be 0.34 as shown below.

Hydrophobicity Values

Amino acid	Value
S	0.3411
P	0.2670
R	0.1505
L	0.0729
H	0.0565
I	0.0305
A	0.0117
V	0.0071
C	0.0024
Y	0.0024
Q	0.0024
D	0.0024

Amino acid transition probabilities were calculated based on the frequency of finding an amino acid in a transcript post-editing compared with pre-editing. The values below represent the frequency seen in the training data pre- and post-editing. Some amino acids were never edited in our data sets, and these have no entries in the post-editing portion of the table. Other amino acids were edited to more than one amino acid as shown by double entries below.

Amino Acid Transition Probabilities

Pre-edit Amino acid	Post-edit Amino acid	Transition Probability
A	A	0.00234741784037559
	V	0.0140845070422535
F	F	0.0234741784037559
S	S	0.0117370892018779
	F	0.115023474178404
	L	0.211267605633803
T	M	0.00469483568075117
	T	0.00704225352112676
	I	0.0164319248826291
P	F	0.0187793427230047
	S	0.0633802816901408
	P	0.00704225352112676
	L	0.199530516431925
Y	Y	0.00234741784037559
V	V	0.00704225352112676
H	Y	0.0539906103286385
Q	*	0.00469483568075117
D	D	0.00234741784037559
I	I	0.0211267605633803
R	W	0.0610328638497653
	C	0.0751173708920188
	R	0.00234741784037559
	*	0.00234741784037559
G	G	0.00234741784037559
L	F	0.0422535211267606
	L	0.028169014084507

Codon transition probabilities were estimated from the training data in a similar manner as the amino acid transition probabilities. These are listed below for each pre-edit codon and its corresponding post-edit versions.

Codon Transition Probabilities

Pre-edit codon	Post-edit codon	Transition Probability
GCC	GTT	0.00234741784037559
CTT	TTT	0.028169014084507
GGA	GGA	0.00469483568075117
GTC	GTC	0.00234741784037559
TCA	TTA	0.117370892018779
CGA	TGA	0.00234741784037559
ATT	ATT	0.00704225352112676
TAT	TAT	0.00234741784037559
TAC	TAC	0.00234741784037559
ACA	ATA	0.00469483568075117,
	ACA	0.00234741784037559
TCG	TTG	0.0985915492957746
ACT	ATT	0.00938967136150235
CAA	TAA	0.00234741784037559
CCG	TTG	0.0140845070422535,
	TCG	0.0117370892018779,
	CTG	0.0352112676056338
CTG	TTG	0.00469483568075117,
	CTG	0.00234741784037559
GGT	GGT	0.00704225352112676
GCA	GTA	0.00234741784037559
GTG	GTG	0.00234741784037559
TCC	TTT	0.00469483568075117,
	TTC	0.0328638497652582
TTT	TTT	0.00234741784037559
CAC	TAC	0.0140845070422535
CGT	CGT	0.00234741784037559,
	TGT	0.0586854460093897
CAT	CAT	0.00234741784037559,

	TAT	0.039906103286385
CGG	TGG	0.0610328638497653
ATA	ATA	0.00469483568075117
ACC	ACC	0.00234741784037559,
	ATC	0.00234741784037559
CCC	TCC	0.00704225352112676,
	CCC	0.00234741784037559,
	TCT	0.00234741784037559,
	TTC	0.00469483568075117,
	CTC	0.00938967136150235
TTA	TTA	0.00234741784037559
CCA	TCA	0.0117370892018779,
	CTA	0.068075117370892,
	TTA	0.0187793427230047
CTA	TTA	0.0140845070422535
GAT	GAT	0.00234741784037559
TCT	TTT	0.0751173708920188,
	TCT	0.00704225352112676
CTC	TTT	0.00469483568075117,
	CTT	0.00234741784037559,
	TTC	0.00938967136150235
CGC	TGC	0.0234741784037559
TTG	TTG	0.00234741784037559
GCG	GTG	0.00469483568075117
GGC	GGC	0.00234741784037559
GAA	GAA	0.00234741784037559
CAG	CAG	0.00234741784037559,
	TAG	0.00234741784037559
GCT	GCT	0.00234741784037559,
	GTT	0.00234741784037559
CCT	TTT	0.0140845070422535,
	CTT	0.0539906103286385,
	TCT	0.0328638497652582
ACG	ATG	0.00469483568075117
AAA	AAA	0.00469483568075117
ATG	ATG	0.00704225352112676
GTA	GTA	0.00234741784037559

The frequency of an edit occurring in each of the three codon positions is listed below.

Codon Position

First Position	**Second Position**	**Third Position**
0.349	0.535	0.116

The frequency with which each of the four nucleotides occur in the -1 position with respect to the edited C is listed below.

-1 Nucleotide

A	**C**	**G**	**U**
0.040	0.305	0.024	0.631

Nucleotide frequencies in the +1 position, immediately downstream of the edited C are listed here.

+1 Nucleotide

A	**C**	**G**	**U**
0.29	0.162	0.325	0.223

Estimating Posterior Probabilities for REGAL Predictions

We used the false positive and false negative rates to estimate the posterior probability that a given prediction by REGAL is true. Below we report the false positive and false negative rates for the range of GA scores from 0 to 60,000.

GA Score	False positive rate	False negative rate
0.000	1	0
500.000	1	0
1000.000	1	0
1500.000	1	0
2000.000	1	0
2500.000	1	0
3000.000	1	0
3500.000	1	0
4000.000	1	0
4500.000	1	0
5000.000	1	0
5500.000	1	0
6000.000	1	0
6500.000	1	0
7000.000	0.981	0
7500.000	0.977	0
8000.000	0.974	0
8500.000	0.971	0
9000.000	0.948	0
9500.000	0.919	0.003
10000.000	0.91	0.003
10500.000	0.884	0.00600000000000001
11000.000	0.874	0.00600000000000001
11500.000	0.813	0.01
12000.000	0.806	0.01
12500.000	0.729	0.019
13000.000	0.674	0.019
13500.000	0.668	0.019
14000.000	0.603	0.023

14500.000	0.587	0.026
15000.000	0.558	0.035
15500.000	0.542	0.035
16000.000	0.487	0.045
16500.000	0.487	0.048
17000.000	0.477	0.048
17500.000	0.465	0.052
18000.000	0.448	0.0580000000000001
18500.000	0.445	0.0610000000000001
19000.000	0.439	0.0649999999999999
19500.000	0.429	0.0679999999999999
20000.000	0.426	0.1
20500.000	0.394	0.1
21000.000	0.381	0.1
21500.000	0.377	0.1
22000.000	0.377	0.1
22500.000	0.374	0.1
23000.000	0.374	0.1
23500.000	0.355	0.106
24000.000	0.339	0.123
24500.000	0.335	0.129
25000.000	0.335	0.129
25500.000	0.329	0.139
26000.000	0.313	0.161
26500.000	0.313	0.174
27000.000	0.297	0.174
27500.000	0.297	0.174
28000.000	0.287	0.187
28500.000	0.287	0.187
29000.000	0.268	0.197
29500.000	0.268	0.197
30000.000	0.194	0.261
30500.000	0.194	0.261
31000.000	0.181	0.274
31500.000	0.177	0.284
32000.000	0.174	0.284
32500.000	0.152	0.352
33000.000	0.1	0.429

33500.000	0.1	0.484
34000.000	0.1	0.484
34500.000	0.1	0.484
35000.000	0.074	0.626
35500.000	0.074	0.639
36000.000	0.026	0.774
36500.000	0.026	0.774
37000.000	0.026	0.774
37500.000	0.026	0.774
38000.000	0.026	0.774
38500.000	0.026	0.774
39000.000	0.026	0.774
39500.000	0.026	0.774
40000.000	0.026	0.774
40500.000	0.026	0.774
41000.000	0.026	0.774
41500.000	0.026	0.774
42000.000	0.026	0.774
42500.000	0.026	0.774
43000.000	0.026	0.774
43500.000	0.026	0.774
44000.000	0.026	0.774
44500.000	0.026	0.774
45000.000	0.026	0.774
45500.000	0.019	0.9
46000.000	0.019	0.9
46500.000	0	1
47000.000	0	1
47500.000	0	1
48000.000	0	1
48500.000	0	1
49000.000	0	1
49500.000	0	1
50000.000	0	1
50500.000	0	1
51000.000	0	1
51500.000	0	1
52000.000	0	1

```
52500.000         0          1
53000.000         0          1
53500.000         0          1
54000.000         0          1
54500.000         0          1
55000.000         0          1
55500.000         0          1
56000.000         0          1
56500.000         0          1
57000.000         0          1
57500.000         0          1
58000.000         0          1
58500.000         0          1
59000.000         0          1
59500.000         0          1
60000.000         0          1
```