Profile HMM Report

Xinyuan Min. 950829573070

1. Read in the supplied small file test.fasta (example from "Biological Sequence Analysis"). What is the number of match states needed?

```
['V', 'V', 'V', 'V', 'V', 'F', 'I', '*']
['G', '-', 'E', 'K', 'Y', 'N', 'A', '*']
['A', '-', 'A', 'G', 'S', 'A', 'G', '*']
['-', '-', '-', '-', '-', '-', 'A', '.']
['-', '-', '-', '-', '-', '-', 'D', '.']
['H', 'N', 'D', '-', 'T', 'N', 'N', '*']
['A', 'V', 'V', '-', 'Y', 'I', 'G', '*']
['G', 'D', 'A', '-', 'E', 'P', 'A', '*']
['E', 'E', 'G', '-', 'T', 'K', 'G', '*']
['Y', 'V', 'H', 'D', 'S', 'H', 'V', '*']
number of match states: 8
```

2. Train the HMM, i.e. the emission and transition probabilities. For emissions in the insert states, use the values supplied in pa. Print the estimated probabilities. Explain whether they make sense

Emission probabilities:

```
0 : {'V': 0.222, 'F': 0.074, 'I': 0.074, 'A': 0.037, 'C': 0.037, 'D': 0.037, 'E': 0.037, 'G': 0.037, 'H': 0.037, 'L': 0.037,
'K': 0.037, 'M': 0.037, 'N': 0.037, 'P': 0.037, 'Q': 0.037, 'R': 0.037, 'S': 0.037, 'T': 0.037, 'W': 0.037, 'Y': 0.037}
1 : {'G': 0.077, 'E': 0.077, 'K': 0.077, 'Y': 0.077, 'N': 0.077, 'A': 0.077, 'C': 0.038, 'D': 0.038, 'F': 0.038, 'H': 0.038,
'I': 0.038, 'L': 0.038, 'M': 0.038, 'P': 0.038, 'Q': 0.038, 'R': 0.038, 'S': 0.038, 'T': 0.038, 'V': 0.038, 'W': 0.038}
2 : {'A': 0.154, 'G': 0.115, 'S': 0.077, 'C': 0.038, 'D': 0.038, 'E': 0.038, 'F': 0.038, 'H': 0.038, 'I': 0.038, 'L': 0.038,
'K': 0.038, 'M': 0.038, 'N': 0.038, 'P': 0.038, 'Q': 0.038, 'R': 0.038, 'T': 0.038, 'V': 0.038, 'W': 0.038, 'Y': 0.038}
3 : {'A': 0.074, 'C': 0.025, 'D': 0.054, 'E': 0.054, 'F': 0.047, 'G': 0.074, 'H': 0.026, 'I': 0.068, 'L': 0.099, 'K': 0.058,
'M': 0.025, 'N': 0.045, 'P': 0.039, 'Q': 0.034, 'R': 0.052, 'S': 0.057, 'T': 0.051, 'V': 0.073, 'W': 0.013, 'Y': 0.034}
4 : {'A': 0.074, 'C': 0.025, 'D': 0.054, 'E': 0.054, 'F': 0.047, 'G': 0.074, 'H': 0.026, 'I': 0.068, 'L': 0.099, 'K': 0.058,
'M': 0.025, 'N': 0.045, 'P': 0.039, 'Q': 0.034, 'R': 0.052, 'S': 0.057, 'T': 0.051, 'V': 0.073, 'W': 0.013, 'Y': 0.034}
5 : {'H': 0.077, 'N': 0.154, 'D': 0.077, 'T': 0.077, 'A': 0.038, 'C': 0.038, 'E': 0.038, 'F': 0.038, 'G': 0.038, 'I': 0.038,
'L': 0.038, 'K': 0.038, 'M': 0.038, 'P': 0.038, 'Q': 0.038, 'R': 0.038, 'S': 0.038, 'V': 0.038, 'W': 0.038, 'Y': 0.038}
6 : {'A': 0.077, 'V': 0.115, 'Y': 0.077, 'I': 0.077, 'G': 0.077, 'C': 0.038, 'D': 0.038, 'E': 0.038, 'F': 0.038, 'H': 0.038,
'L': 0.038, 'K': 0.038, 'M': 0.038, 'N': 0.038, 'P': 0.038, 'Q': 0.038, 'R': 0.038, 'S': 0.038, 'T': 0.038, 'W': 0.038}
7 : {'G': 0.077, 'D': 0.077, 'A': 0.115, 'E': 0.077, 'P': 0.077, 'C': 0.038, 'F': 0.038, 'H': 0.038, 'I': 0.038, 'L': 0.038,
'K': 0.038, 'M': 0.038, 'N': 0.038, 'Q': 0.038, 'R': 0.038, 'S': 0.038, 'T': 0.038, 'V': 0.038, 'W': 0.038, 'Y': 0.038}
8 : {'E': 0.115, 'G': 0.115, 'T': 0.077, 'K': 0.077, 'A': 0.038, 'C': 0.038, 'D': 0.038, 'F': 0.038, 'H': 0.038, 'I': 0.038,
'L': 0.038, 'M': 0.038, 'N': 0.038, 'P': 0.038, 'Q': 0.038, 'R': 0.038, 'S': 0.038, 'V': 0.038, 'W': 0.038, 'Y': 0.038}
9 : {'Y': 0.074, 'V': 0.111, 'H': 0.111, 'D': 0.074, 'S': 0.074, 'A': 0.037, 'C': 0.037, 'E': 0.037, 'F': 0.037, 'G': 0.037,
'I': 0.037, 'L': 0.037, 'K': 0.037, 'M': 0.037, 'N': 0.037, 'P': 0.037, 'Q': 0.037, 'R': 0.037, 'T': 0.037, 'W': 0.037}
```

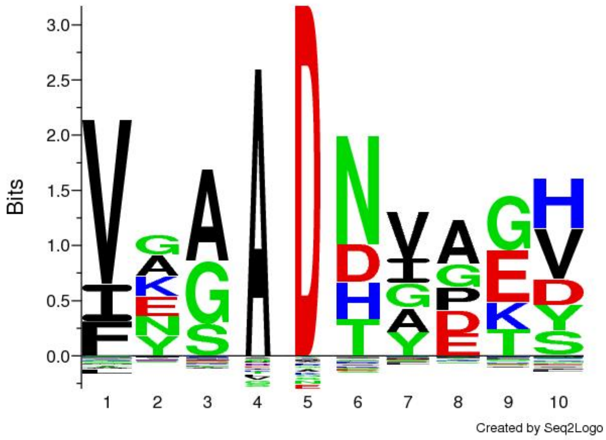
Transition probabilities:

```
0 : {'m': {'mm': 0.7, 'mi': 0.1, 'md': 0.2}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
1 : {'m': {'mm': 0.667, 'mi': 0.167, 'md': 0.167}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.286, 'dd': 0.429, 'di': 0.286}}
2 : {'m': {'mm': 0.467, 'mi': 0.267, 'md': 0.267}, 'i': {'im': 0.364, 'id': 0.273, 'ii': 0.364}, 'd': {'dm': 0.4, 'dd': 0.3, 'di': 0.3}}
5 : {'m': {'mm': 0.556, 'mi': 0.222, 'md': 0.222}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.308, 'dd': 0.385, 'di': 0.308}}
6 : {'m': {'mm': 0.524, 'mi': 0.238, 'md': 0.238}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.312, 'dd': 0.375, 'di': 0.312}}
7 : {'m': {'mm': 0.5, 'mi': 0.25, 'md': 0.25}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.316, 'dd': 0.368, 'di': 0.316}}
8 : {'m': {'mm': 0.481, 'mi': 0.259, 'md': 0.259}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.364, 'dd': 0.318, 'di': 0.318}}
```

(The probabilities were rounded to 3 decimals for readability)

3. Print the emission probability matrix in a format demanded by the SeqLogo server (http://www.cbs.dtu.dk/biotools/Seq2Logo-2.0/), e.g. the PSSM format (http://www.cbs.dtu.dk/biotools/Seq2Logo-2.0/bin/weight.txt), and use SeqLogo to generate a logo. Put both the matrix and logo in your answers.

```
        A      C      D      E      F      G      H      I      L      K      M      N      P      Q      R      S      T      V      W      Y
0 [0.037, 0.037, 0.037, 0.037, 0.074, 0.037, 0.037, 0.074, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.222, 0.037, 0.037]
1 [0.077, 0.038, 0.038, 0.077, 0.038, 0.077, 0.038, 0.038, 0.038, 0.077, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.077]
2 [0.154, 0.038, 0.038, 0.038, 0.038, 0.115, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038]
3 [0.074, 0.025, 0.054, 0.054, 0.047, 0.074, 0.026, 0.068, 0.099, 0.058, 0.025, 0.045, 0.039, 0.034, 0.052, 0.057, 0.051, 0.073, 0.013, 0.034]
4 [0.074, 0.025, 0.054, 0.054, 0.047, 0.074, 0.026, 0.068, 0.099, 0.058, 0.025, 0.045, 0.039, 0.034, 0.052, 0.057, 0.051, 0.073, 0.013, 0.034]
5 [0.038, 0.038, 0.077, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.154, 0.038, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038]
6 [0.077, 0.038, 0.038, 0.038, 0.038, 0.077, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.115, 0.038, 0.077]
7 [0.115, 0.038, 0.077, 0.077, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038]
8 [0.038, 0.038, 0.038, 0.115, 0.038, 0.115, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038, 0.038, 0.038, 0.038, 0.077, 0.038, 0.038, 0.038]
9 [0.037, 0.037, 0.074, 0.037, 0.037, 0.037, 0.111, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.037, 0.074, 0.037, 0.111, 0.037, 0.074]
```



4. List at least 10 random sequences generated using your HMM

1 ITGAL
2 WVLRWDA
3 VLRKIDM
4 QYPAL
5 EYVGRTW
6 GWKYTV
7 ILRN
8 HCCVIFTE
9 VDHKLL
10 VFTNSA

5. Change from using the estimated emission probabilities (EQ6.25) in your code to the background frequencies (EQ6.26) and repeat steps 2-4. How do the results change?

ERILVKG
ELAMMKA
TNGCGHL
GDPALTK
DEGYS
QRSLGIK
VKEFAGA
QKQT
GNVIEG
KSHALR

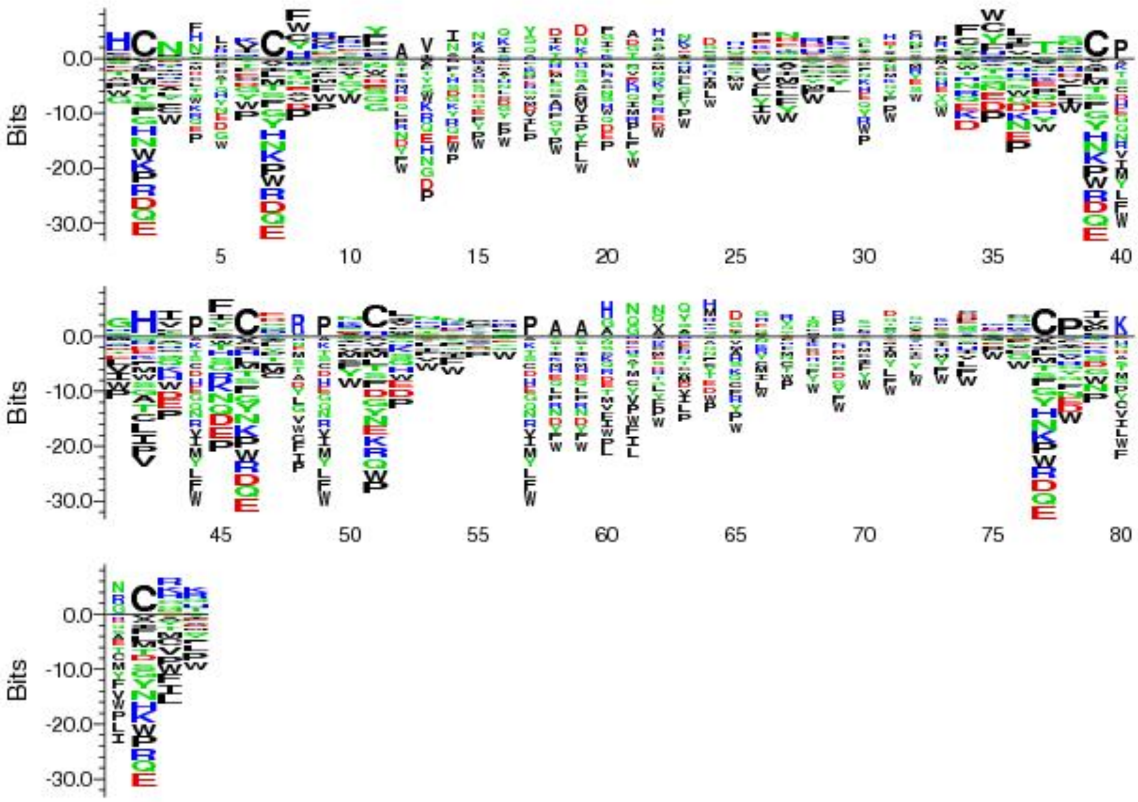
6. Read in the supplied large file test_large.fasta and repeat steps 2-5. Report the emission and transition probabilities, the logo (from SeqLogo), and the sampled sequences.

Emission probabilities

	A	C	D	E	F	G	H	I	L	K	M	N	P	Q	R	S	T	V	W	Y
0	0.014	0.022	0.022	0.072	0.058	0.007	0.275	0.051	0.043	0.072	0.022	0.036	0.014	0.036	0.094	0.065	0.029	0.036	0.007	0.022
1	0.007	0.855	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.014
2	0.022	0.007	0.007	0.022	0.014	0.065	0.007	0.036	0.036	0.051	0.022	0.442	0.072	0.029	0.014	0.072	0.022	0.036	0.007	0.014
3	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
4	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
5	0.043	0.007	0.014	0.014	0.014	0.014	0.036	0.130	0.051	0.225	0.022	0.043	0.007	0.043	0.065	0.065	0.014	0.174	0.007	0.007
6	0.014	0.855	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
7	0.015	0.066	0.015	0.022	0.234	0.051	0.066	0.015	0.058	0.036	0.015	0.015	0.007	0.022	0.102	0.073	0.029	0.007	0.051	0.102
8	0.044	0.044	0.051	0.095	0.007	0.036	0.022	0.036	0.036	0.146	0.007	0.066	0.007	0.051	0.139	0.080	0.051	0.051	0.007	0.022
9	0.029	0.015	0.058	0.102	0.022	0.015	0.029	0.036	0.124	0.102	0.036	0.029	0.124	0.058	0.088	0.080	0.022	0.015	0.007	0.007
10	0.022	0.029	0.022	0.022	0.139	0.007	0.036	0.058	0.153	0.044	0.022	0.022	0.124	0.007	0.051	0.022	0.029	0.036	0.007	0.146
11	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
12	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
13	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
14	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
15	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
16	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
17	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
18	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
19	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
20	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
21	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
22	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
23	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
24	0.065	0.022	0.043	0.087	0.022	0.033	0.054	0.054	0.065	0.054	0.011	0.076	0.043	0.022	0.098	0.130	0.054	0.033	0.011	0.022
25	0.038	0.010	0.086	0.152	0.019	0.067	0.019	0.010	0.019	0.133	0.019	0.076	0.114	0.029	0.019	0.114	0.038	0.019	0.010	0.010
26	0.016	0.008	0.102	0.133	0.016	0.047	0.039	0.023	0.016	0.086	0.008	0.117	0.031	0.062	0.055	0.078	0.125	0.023	0.008	0.008
27	0.023	0.015	0.115	0.130	0.015	0.046	0.046	0.046	0.046	0.061	0.008	0.015	0.038	0.046	0.153	0.053	0.046	0.076	0.008	0.015
28	0.076	0.023	0.031	0.038	0.031	0.076	0.031	0.031	0.008	0.145	0.031	0.031	0.137	0.076	0.084	0.046	0.046	0.038	0.015	0.008
29	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
30	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
31	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
32	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
33	0.094	0.065	0.007	0.014	0.261	0.043	0.007	0.022	0.072	0.007	0.036	0.014	0.109	0.014	0.022	0.043	0.022	0.065	0.007	0.072
34	0.036	0.094	0.007	0.007	0.123	0.022	0.043	0.065	0.080	0.051	0.022	0.022	0.007	0.029	0.043	0.029	0.022	0.080	0.072	0.145
35	0.036	0.029	0.014	0.007	0.087	0.014	0.014	0.167	0.312	0.007	0.029	0.007	0.007	0.022	0.022	0.029	0.029	0.145	0.007	0.014
36	0.022	0.007	0.007	0.014	0.022	0.080	0.007	0.029	0.130	0.014	0.029	0.014	0.014	0.007	0.022	0.065	0.449	0.051	0.007	0.007
37	0.094	0.022	0.007	0.036	0.014	0.051	0.007	0.007	0.014	0.036	0.014	0.080	0.080	0.065	0.065	0.261	0.101	0.014	0.007	0.022
38	0.007	0.862	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
39	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
40	0.051	0.014	0.022	0.022	0.043	0.370	0.043	0.007	0.029	0.043	0.022	0.036	0.007	0.065	0.029	0.138	0.022	0.007	0.007	0.022
41	0.007	0.007	0.014	0.007	0.007	0.007	0.848	0.007	0.007	0.007	0.007	0.014	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
42	0.029	0.007	0.007	0.007	0.058	0.022	0.007	0.297	0.080	0.014	0.014	0.029	0.007	0.014	0.029	0.065	0.101	0.181	0.007	0.022
43	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
44	0.043	0.014	0.007	0.007	0.333	0.014	0.007	0.174	0.152	0.007	0.022	0.007	0.007	0.007	0.007	0.029	0.014	0.072	0.007	0.065
45	0.007	0.855	0.007	0.007	0.007	0.007	0.014	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
46	0.065	0.007	0.109	0.152	0.014	0.058	0.036	0.036	0.058	0.094	0.007	0.065	0.022	0.094	0.036	0.058	0.014	0.036	0.022	0.014
47	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
48	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
49	0.094	0.007	0.058	0.087	0.014	0.051	0.029	0.029	0.043	0.130	0.007	0.101	0.036	0.036	0.080	0.116	0.043	0.022	0.007	0.007
50	0.007	0.797	0.014	0.014	0.007	0.007	0.058	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
51	0.145	0.014	0.007	0.014	0.058	0.043	0.007	0.130	0.217	0.014	0.036	0.029	0.007	0.014	0.036	0.014	0.058	0.116	0.007	0.029
52	0.036	0.014	0.072	0.051	0.014	0.051	0.036	0.022	0.051	0.116	0.029	0.094	0.036	0.087	0.036	0.116	0.087	0.014	0.007	0.029
53	0.072	0.014	0.043	0.087	0.007	0.058	0.029	0.014	0.051	0.116	0.014	0.101	0.036	0.065	0.094	0.043	0.072	0.051	0.007	0.022
54	0.101	0.051	0.029	0.058	0.043	0.058	0.029	0.029	0.152	0.080	0.036	0.022	0.014	0.058	0.029	0.072	0.051	0.043	0.014	0.029
55	0.065	0.014	0.043	0.036	0.065	0.043	0.036	0.051	0.087	0.116	0.043	0.029	0.058	0.072	0.036	0.058	0.080	0.043	0.007	0.014
56	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
57	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
58	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
59	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
60	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
61	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
62	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
63	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
64	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
65	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
66	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
67	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
68	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
69	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
70	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
71	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
72	0.070	0.023	0.047	0.093	0.012	0.070	0.012	0.023	0.070	0.116	0.012	0.047	0.058	0.070	0.070	0.105	0.047	0.035	0.012	0.012
73	0.053	0.044	0.096	0.105	0.009	0.079	0.009	0.035	0.018	0.061	0.018	0.061	0.088	0.061	0.061	0.114	0.035	0.018	0.009	0.026
74	0.073	0.016	0.073	0.024	0.032	0.065	0.024	0.040	0.065	0.113	0.024	0.040	0.016	0.048	0.105	0.065	0.056	0.056	0.008	0.056
75	0.030	0.037	0.052	0.052	0.015	0.015	0.037	0.059	0.052	0.096	0.022	0.022	0.037	0.126	0.074	0.067	0.081	0.089	0.007	0.030
76	0.007	0.862	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
77	0.029	0.014	0.007	0.007	0.014	0.022	0.029	0.029	0.058	0.022	0.022	0.007	0.565	0.014	0.022	0.036	0.051	0.022	0.007	0.022
78	0.101	0.014	0.014	0.036	0.065	0.014	0.029	0.174	0.116	0.014	0.043	0.007	0.007	0.051	0.029	0.014	0.065	0.181	0.007	0.014
79	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
80	0.074	0.025	0.054	0.054	0.047	0.074	0.026	0.068	0.099	0.058	0.025	0.045	0.039	0.034	0.052	0.057	0.051	0.073	0.013	0.034
81	0.007	0.841	0.022	0.007	0.014	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
82	0.014	0.007	0.036	0.072	0.007	0.152	0.065	0.007	0.007	0.181	0.007	0.065	0.007	0.109	0.167	0.051	0.014	0.007	0.007	0.014
83	0.094	0.014	0.022	0.022	0.014	0.022	0.043	0.036	0.014	0.174	0.014	0.029	0.007	0.087	0.123	0.094	0.130	0.029	0.007	0.022

Transition probabilities:

0 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
1 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
2 : {'m': {'mm': 0.893, 'mi': 0.099, 'md': 0.008}, 'i': {'im': 0.48, 'id': 0.04, 'ii': 0.48}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
3 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
4 : {'m': {'mm': 0.975, 'mi': 0.008, 'md': 0.017}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
5 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.25, 'dd': 0.5, 'di': 0.25}}
6 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.25, 'dd': 0.5, 'di': 0.25}}
7 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.25, 'dd': 0.5, 'di': 0.25}}
8 : {'m': {'mm': 0.208, 'mi': 0.4, 'md': 0.392}, 'i': {'im': 0.41, 'id': 0.009, 'ii': 0.581}, 'd': {'dm': 0.5, 'dd': 0.25, 'di': 0.25}}
9 : {'m': {'mm': 0.973, 'mi': 0.013, 'md': 0.013}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.286, 'dd': 0.694, 'di': 0.02}}
10 : {'m': {'mm': 0.977, 'mi': 0.011, 'md': 0.011}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.667, 'dd': 0.306, 'di': 0.028}}
11 : {'m': {'mm': 0.982, 'mi': 0.009, 'md': 0.009}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.308, 'dd': 0.615, 'di': 0.077}}
12 : {'m': {'mm': 0.982, 'mi': 0.009, 'md': 0.009}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.1, 'dd': 0.8, 'di': 0.1}}
13 : {'m': {'mm': 0.868, 'mi': 0.123, 'md': 0.009}, 'i': {'im': 0.296, 'id': 0.014, 'ii': 0.69}, 'd': {'dm': 0.1, 'dd': 0.1, 'di': 0.8}}
14 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
15 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
16 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
17 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
18 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
19 : {'m': {'mm': 0.975, 'mi': 0.017, 'md': 0.008}, 'i': {'im': 0.5, 'id': 0.25, 'ii': 0.25}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
20 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
21 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
22 : {'m': {'mm': 0.975, 'mi': 0.017, 'md': 0.008}, 'i': {'im': 0.5, 'id': 0.25, 'ii': 0.25}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
23 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
24 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
25 : {'m': {'mm': 0.975, 'mi': 0.017, 'md': 0.008}, 'i': {'im': 0.4, 'id': 0.2, 'ii': 0.4}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
26 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
27 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
28 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
29 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
30 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
31 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
32 : {'m': {'mm': 0.157, 'mi': 0.405, 'md': 0.438}, 'i': {'im': 0.194, 'id': 0.004, 'ii': 0.802}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
33 : {'m': {'mm': 0.971, 'mi': 0.014, 'md': 0.014}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.527, 'dd': 0.455, 'di': 0.018}}
34 : {'m': {'mm': 0.979, 'mi': 0.01, 'md': 0.01}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.407, 'dd': 0.556, 'di': 0.037}}
35 : {'m': {'mm': 0.981, 'mi': 0.009, 'md': 0.009}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.706, 'dd': 0.235, 'di': 0.059}}
36 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.667, 'dd': 0.167, 'di': 0.167}}
37 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
38 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
39 : {'m': {'mm': 0.959, 'mi': 0.033, 'md': 0.008}, 'i': {'im': 0.571, 'id': 0.143, 'ii': 0.286}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
40 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}
41 : {'m': {'mm': 0.983, 'mi': 0.008, 'md': 0.008}, 'i': {'im': 0.333, 'id': 0.333, 'ii': 0.333}, 'd': {'dm': 0.333, 'dd': 0.333, 'di': 0.333}}



Created by Seq2Logo

Randomly generated sequence:

```
1 TCSLAKCCSELIADLFGKRDQIFNHPDQFFRIFLVNCPGHV
2 TCNSTVCFRPALGLWIVLTAADVTIGRFWASFFTACIGHV
3 MCKGPFCYMALSSEFIGVKLAAKRDEQQFLLLYEFMFSCLMHL
4 QCNKSLPTTDMRIIKCDTSERACAQNQTNVTRCHGHG
5 RCKLKVFGRSFDCLLFQAVVGCAEDDWVNHSAKFRRGKCFFHG
6 HCEDVACCELFAQRRSYTVVGMPLRTQFFDFMVIMQCAFHL
7 QCNNNAEHPLALGTKYFACKLKFLGKSIFVLNAVTMIDT
8 NCAVGKCFYKPELGVSGFFVLRFHFKKVKLRATSNCIGHA
9 ECITKKCYRDNTVKQDYDSVEKSRLQQVQDELLVILLPCASHV
10 HCNRYPCSNEVKLAFNQAATENVVKLVIGVQPPADRHT
```

7. Pick a randomly generated sequence and search for it in the PFAM database (http://pfam.xfam.org/). To what family does the sequence belong?
   I feel sorry that Pfam didn't find the belonging family of my randomly generated sequence …

8. What is the time complexity of estimating the emission and transition probabilities?
   The time complexity is O(n), as for each column, 2 count(amino acid + transition), (20+9) divisions are performed, and these operations are performed for n columns (n is the number of sequence)

9. Which part of your code is deterministic and which part is randomized?
   Both the emission and transition probabilities are deterministic, the calculation obeys certain mathematical expression.
   However, the sequence generation part is randomized, the transition between state and the amino acid at one position is randomly pick according certain probabilities distribution.