

Dataset S1. List of predicted EIEs

Word	Ns	Na	z
AAGAAA	4155	1304	45.23
ACCTAC	886	105	40.13
GAAGAA	4724	1828	39.88
CCTACC	854	100	39.67
AGAAAA	3567	1216	39.02
AGAAGA	4540	1769	38.82
TACCTG	1438	289	36.98
AAAGAA	3593	1315	36.70
AAGAAG	4052	1624	35.65
CTTACC	604	69	33.77
CTACCT	1154	227	33.59
GAAAAA	2763	978	33.19
CAAGAA	2884	1067	32.54
AAAAGA	2629	964	31.33
GCCTAC	815	141	30.64
CCTACA	1155	274	29.55
ACCCTG	1861	608	29.25
GGAAAA	2653	1058	28.99
GAAAGA	2452	956	28.51
GGAGAA	3284	1474	28.37
AAGGAA	2635	1099	27.57
GAGAAA	2512	1027	27.49
CCTGGA	3928	1947	27.45
GACCTG	1980	727	27.15
TGGAGA	3671	1794	27.03
TTACCT	747	147	26.99
ACAAGA	1893	685	26.92
GAGGAG	3278	1569	26.23
AACTAC	883	208	25.96
CTACAA	1362	427	25.90
ACTACA	1192	346	25.80
CCCTAC	671	131	25.71
AACCTG	1671	605	25.27
GAAAAG	2296	974	25.27
GAGAAG	3239	1592	25.20
AGAGAA	2745	1264	25.16
GACTAC	775	176	24.95
TGGAAA	3004	1446	24.93
CTACCG	389	52	24.67
AACAAG	1726	654	24.60
CTGGAG	4159	2275	24.56
TGTGGA	2573	1177	24.55
ACCTGG	2471	1112	24.53
TGAAGA	4197	2310	24.44
AAAAAA	2093	878	24.42
AAGGAG	2679	1258	24.28

Word	Ns	Na	z
GAAAAT	2405	1084	24.15
ACCTGA	1455	515	24.08
AAGAGA	2307	1025	24.05
AGAAAG	2573	1210	23.75
CAAGGA	2395	1099	23.60
ATGGAA	2303	1042	23.53
GACCTA	554	107	23.51
TGGACC	1751	703	23.38
CTACAG	1424	517	23.27
ACCTGC	1966	838	23.26
AAAAAG	2143	952	23.18
GGAGGA	3127	1621	23.04
CTACAC	903	254	23.01
CGCTAC	388	59	22.82
GCTACA	1189	400	22.78
CCTATG	875	246	22.66
AAAGGA	2311	1085	22.55
GGAAGA	3115	1644	22.42
GAAGGA	2724	1373	22.35
CCTACT	757	197	22.35
TGAAAA	2645	1320	22.32
ACTTAC	573	123	22.28
CTACGA	537	110	22.26
ATGAAA	2271	1072	22.21
ACTATG	973	300	22.19
CTATGA	1264	457	22.00
CCTACG	444	80	21.99
GAAAAC	1765	754	21.98
CACCTA	731	191	21.89
GAAGAT	2950	1562	21.71
CAACCT	1281	477	21.54
GAACAA	1678	716	21.46
AATGGA	1973	904	21.46
AAGAAC	2005	926	21.43
GGCTAC	968	315	21.15
GACCCT	1283	492	20.95
CTGGAC	2384	1207	20.78
ACTACC	723	201	20.77
AGAAAT	2465	1265	20.74
GCTACC	791	236	20.55
AGGAGA	2855	1554	20.50
ACCCTA	425	83	20.41
TCAAGA	1885	885	20.36
TATGAA	1417	590	20.25
AGGAAA	2553	1350	20.24
TACCTA	389	72	20.21
CAGAAA	2886	1594	20.15
AACCTA	501	113	20.13

Word	Ns	Na	z
TGGAGC	2620	1408	20.02
AGTGGA	1964	951	20.00
GCCCTG	2585	1384	20.00
CCCTGG	2808	1547	19.96
AGACCT	1408	593	19.94
GTGGAC	1665	756	19.92
GCCCTA	510	119	19.83
AGAGGA	2510	1339	19.81
TCCTAC	850	278	19.73
GTGGAG	2437	1291	19.72
AGAAGC	2548	1377	19.58
CAACAA	1780	852	19.32
GAACTA	753	237	19.18
CTCAAC	1136	448	19.17
ACCTGT	1279	535	19.14
CAACCC	995	369	19.06
GGACCT	1495	674	19.03
GAGACC	1425	631	18.97
AACAGA	1816	890	18.94
AACAAC	1359	591	18.91
GCTGGA	3234	1921	18.91
CCCTGC	2163	1135	18.83
AGAACA	2097	1089	18.82
GAGAAC	1629	769	18.80
AGAAAC	1935	980	18.71
TACCTC	846	294	18.66
CTACCC	607	174	18.57
AAACAA	1736	849	18.56
CTATGG	987	375	18.54
GACAAC	1127	459	18.48
AATGAA	2180	1165	18.41
TACAAG	1196	504	18.36
CAGAAG	3142	1883	18.34
CTACAT	1102	448	18.31
GTGGAA	1996	1039	18.30
ACCTAT	540	147	18.23
ACAGAA	2211	1196	18.21
ATGAAG	2966	1758	18.18
GCAAGA	1542	734	18.11
CAAGAC	1300	577	18.07
ACAACA	1602	776	18.05
TATGGA	1248	545	18.03
CAACTA	730	243	18.00
GACAAG	1433	666	17.97
ACCTCA	1577	762	17.97
ATGGAG	2388	1337	17.94
CCTGAA	2155	1176	17.74
AAACCT	1207	527	17.74

Word	Ns	Na	z
ACTACG	382	85	17.72
CCCTGA	1801	926	17.68
AACCCC	932	362	17.63
ACCCCT	952	375	17.57
GCTACG	427	105	17.47
AAAACA	1979	1062	17.43
ACCCCA	1463	703	17.43
CCTGAC	1308	601	17.41
AAAATG	2239	1255	17.34
AGCTAC	920	361	17.34
CATGGA	2213	1241	17.23
GGAGCT	2452	1424	17.12
CCTGCC	2144	1196	17.10
GAAATG	1830	969	17.10
TCAACC	1033	435	17.07
CTATGC	730	258	17.07
GACCCC	1213	552	16.96
GGACCC	1298	608	16.94
TACTAC	676	231	16.93
TACAGA	1394	674	16.88
ACCCCC	960	395	16.87
GATGAA	2547	1509	16.86
ACCAAG	1632	839	16.84
CTCTAC	991	415	16.82
CCAAGA	2173	1230	16.82
GAGGAA	2509	1486	16.74
GCCAAG	1832	985	16.73
AACCCT	850	333	16.69
CTCAAG	1435	709	16.65
AAAGAG	2041	1141	16.63
AAGATG	2704	1641	16.63
GGACAA	1435	713	16.53
CTACGC	324	71	16.48
CTACCA	1104	495	16.46
TACGAG	445	122	16.45
CTACGT	418	110	16.44
ACGAGA	612	205	16.39
GAGGAC	1770	955	16.35
GAAGAC	2035	1149	16.34
CCTGTG	2089	1190	16.32
AAGACC	1399	696	16.29
ATTACC	596	198	16.29
CAGACC	1390	691	16.26
AAAAAC	1540	796	16.23
ACAAGC	1246	595	16.21
AGACAA	1507	775	16.17
ACAACC	1041	462	16.17
AAAACC	1291	627	16.15

Word	Ns	Na	z
AAGACA	1834	1010	16.14
AAGAAT	1968	1109	16.12
TACAAC	951	407	16.09
CTACGG	424	116	16.08
GCCTAT	590	198	16.06
ATCGAG	657	235	16.01
CCACCC	1511	784	15.99
TGAAAG	1915	1076	15.98
ATCAAC	1252	607	15.94
GGCCTA	544	176	15.92
CCCCTG	1714	933	15.88
GCTATG	879	368	15.85
GAAATT	1734	949	15.84
AAATGA	1897	1069	15.83
AAAAAT	2078	1204	15.82
ATACCT	672	247	15.79
TCTACA	1272	626	15.76
TAAAGA	1465	760	15.75
CAAGCA	1615	867	15.74
AAACTA	760	299	15.71
CAAGAT	1719	944	15.69
CACCCT	1297	646	15.66
AAATTA	1166	560	15.57
ACCCAG	1743	968	15.53
GGAAAT	1785	1000	15.50
TCTACG	456	137	15.49
GCAGAA	2379	1450	15.47
TGGACA	2275	1370	15.47
ATCTAC	906	395	15.39
TATGTG	924	407	15.36
GAACCT	1178	574	15.36
GAAACC	1160	563	15.32
AGACCC	1139	549	15.32
AGGAGC	2360	1444	15.30
ACCTAA	418	121	15.28
ATCAAG	1489	795	15.23
CTGACC	1420	746	15.23
AGAATG	1748	986	15.17
CCCTAT	482	154	15.14
AACGAG	559	195	15.10
GGAGCA	2144	1289	15.06
CGAGAA	806	340	15.05
TGACCC	1243	628	15.04
GTGAAA	1526	830	15.00
GAGCTA	707	281	15.00
AACAAA	1719	973	14.96
CAAGCC	1234	625	14.94
CACCCC	1232	626	14.86

Word	Ns	Na	z
TGCCTA	655	253	14.85
ACCAGA	1675	945	14.84
TGCTAC	916	416	14.77
GGAGAC	1699	966	14.76
CTACTC	660	258	14.73
ATTGAA	1445	784	14.65
GAAGAG	2749	1787	14.61
AGAAGG	2486	1578	14.61
GAAACA	1723	991	14.59
ACCTTA	484	162	14.58
GGAAAG	1770	1027	14.57
AGATGA	2516	1604	14.57
CGAGGA	1063	519	14.55
AACTAT	721	298	14.55
CAGCTA	1011	485	14.51
CCAACC	1119	560	14.46
TACGAC	330	88	14.45
TACCAG	1251	654	14.39
CCTGCG	1018	493	14.39
AAGCAA	1522	850	14.38
TCTACC	894	411	14.38
TACGTG	456	150	14.36
TGACCT	1468	812	14.34
GGGAAA	1416	775	14.31
TACAAA	1188	613	14.29
GAAGCA	2113	1305	14.22
GAGCTG	2940	1966	14.19
TCAACA	1706	996	14.15
ATCCTG	2057	1265	14.14
CCCAAC	1036	513	14.11
ATGGAC	1334	723	14.10
GACTAT	626	249	14.10
TACCCT	557	209	14.09
CTACTG	1050	524	14.06
TGTGAA	1863	1120	14.04
TCAAGC	1100	559	14.04
GAGAGA	1661	967	14.03
TCAACG	520	189	14.03
CTCACC	1199	630	13.99
AGCCTA	490	173	13.98
AAGCTG	2527	1645	13.97
GGACTA	551	208	13.93
TGGGAA	1941	1186	13.91
AAAGAC	1419	793	13.87
CCTTAC	533	199	13.84
TGGAGG	2908	1961	13.83
ATTATG	905	433	13.78
TTACCG	233	52	13.77

Word	Ns	Na	z
TGAGAA	2005	1245	13.71
TGGAAG	2798	1880	13.68
GATGGA	2053	1284	13.68
ACAAAA	1609	943	13.65
AAGCAG	2291	1474	13.64
GAAAGC	1483	849	13.64
CCCAGA	2139	1357	13.57
AAATGG	1654	981	13.55
TGGACG	856	406	13.54
AAAGAT	1816	1106	13.54
ATCCTA	564	223	13.46
AGCCTG	2127	1353	13.45
AAGGAC	1476	851	13.44
ATGGGA	1373	775	13.43
CCCCTA	489	181	13.37
AGAACC	1265	699	13.33
ACCAAC	1106	585	13.31
CAGAGA	2438	1610	13.29
AGTACC	797	373	13.29
TTACCA	909	449	13.25
CAAGAG	1566	928	13.21
GTACCT	710	318	13.21
TCAGAA	2163	1396	13.16
TATGAG	911	454	13.11
AATTAC	756	350	13.11
CCCACC	1545	916	13.11
TTACCC	480	180	13.08
ATGAGA	1492	877	13.08
GGTGGA	1990	1263	13.07
GTGACC	1233	685	13.05
GACCAA	1127	608	13.05
ACTACT	734	337	13.05
AGAACT	1751	1077	13.05
CCTGGC	2516	1689	13.00
TATGAC	791	376	12.98
GATTAC	539	216	12.98
AAGATT	1431	836	12.94
ATCGTG	604	257	12.90
TACCTT	637	278	12.88
GAAGCT	2244	1476	12.87
GACGAG	692	314	12.84
CAACGA	567	237	12.74
ACCCAA	903	459	12.71
CAAGTG	1320	762	12.69
CCAAGC	1446	858	12.66
AGATCA	1467	874	12.66
GCGCTA	227	56	12.66
GGAGAT	1828	1155	12.64

Word	Ns	Na	z
ACCCTC	1019	543	12.64
TTACAA	1012	539	12.60
ATGGAT	1452	865	12.60
AGCAGA	2312	1545	12.60
CGACCT	493	195	12.58
GAGCAA	1333	776	12.57
AAGACT	1267	727	12.56
CCCTAA	405	145	12.55
GTCTAC	595	258	12.54
AGTGAA	1547	941	12.52
TACCGG	313	97	12.50
CAGAAC	1615	996	12.46
TACCGC	334	108	12.46
CCTAGA	511	208	12.44
TGCAGA	2598	1790	12.41
ATTACA	956	505	12.39
AGATTG	1057	578	12.38
AGACCA	1553	952	12.36
CGCCTA	243	65	12.35
GGGACC	1189	676	12.35
TGAAAC	1326	781	12.28
CGAGAT	573	249	12.28
CCAGAA	2581	1784	12.27
AAGCTA	729	351	12.26
GAACCC	883	457	12.26
AAGTGG	1547	953	12.22
TTATGA	1021	557	12.21
CCTAAA	604	270	12.21
TGTGTG	1570	972	12.20
CGAGAC	516	215	12.19
CCGAGA	827	420	12.18
ACAACT	1185	679	12.17
AAACAG	1754	1118	12.17
GCCAAC	1155	657	12.16
AGACTA	510	212	12.15
CGGGAA	707	339	12.14
TTGGAA	2009	1323	12.14
TCTATG	993	539	12.14
GACCCA	1135	644	12.10
AGCAAG	1630	1023	12.10
TTTACC	716	346	12.10
AGTATG	855	442	12.09
AGATTA	706	340	12.07
AGCAAC	1318	783	12.06
ATTGGA	1229	716	12.05
AAAAGG	1532	949	12.04
CACCTG	2084	1389	12.03
CCTATT	539	232	12.03

Word	Ns	Na	z
CAAAAA	1810	1170	12.00
CAAGCT	1456	892	11.98
CCTGAT	1329	795	11.96
TTATGG	823	423	11.95
GCAACC	830	428	11.95
AGAGAC	1303	776	11.94
CCTAAT	407	154	11.94
ATCGAC	405	153	11.92
TGGAAC	1639	1040	11.87
AACTTA	636	298	11.85
GACAGA	1467	906	11.85
GTATGA	750	375	11.85
ATGTGG	1687	1080	11.82
GAGTAC	776	394	11.80
TCGAGA	581	263	11.80
AATTGA	1031	577	11.79
CTTACT	594	272	11.77
GAAACT	1469	911	11.76
CACTAC	707	348	11.74
GGAACC	1116	642	11.73
GTGGAT	1447	895	11.73
AGCTAT	753	380	11.72
GAAAGT	1393	854	11.71
GAATTA	843	444	11.69
GACCTC	1275	766	11.63
CTGAAA	1976	1322	11.62
ACACCC	962	532	11.61
TGGACT	1484	929	11.60
CCGGGA	984	549	11.58
CCCCCC	704	350	11.56
CGCCTG	947	523	11.54
CATCGA	822	433	11.54
GAACGA	474	200	11.53
AAACCC	847	452	11.49
CTATGT	727	368	11.47
AGCCCT	1510	955	11.47
ATGCTA	711	357	11.47
CTTACA	747	382	11.47
TACCCA	719	363	11.45
GGAAAC	1292	786	11.44
ACCTAG	315	109	11.40
AAGCCC	1298	792	11.40
ACAACG	496	216	11.39
ACAAGG	1338	824	11.37
GCTACT	792	416	11.37
TATGCC	699	351	11.37
GACCGG	471	201	11.35
CGGGAG	959	538	11.33

Word	Ns	Na	z
ATGTGA	1239	750	11.31
CCCAAG	1454	918	11.29
TACCGA	275	89	11.29
GTCAAC	852	462	11.26
GGACGA	679	340	11.25
ACCGGC	534	243	11.24
GATCGA	424	174	11.23
GAGAAT	1448	917	11.20
GGAGCG	833	450	11.19
AGGACC	1415	892	11.17
TACCCC	524	238	11.16
GAGATC	1279	787	11.14
AACCTC	1045	608	11.14
CAGATG	2326	1637	11.11
GCGGGA	732	381	11.07
GTGAAG	2020	1385	11.07
CTCCTA	747	393	11.02
GTCAAG	1009	585	11.01
ATCGCC	509	231	11.01
GCTTAC	425	178	11.00
TGCCCT	1598	1046	10.97
CGAGTG	573	274	10.96
CGTGGA	1027	600	10.96
CCCTGT	1222	750	10.94
ATATGA	988	572	10.92
CTACTT	892	501	10.91
CAACGG	487	219	10.88
AATGTG	1397	891	10.84
ATGAAC	1345	850	10.84
CTAAAG	770	415	10.80
GGACCG	553	264	10.79
AGAACG	601	297	10.77
TGGATG	2367	1691	10.76
ACCGAG	651	332	10.74
TATGGG	672	347	10.73
GCAAGC	991	581	10.70
TACCAA	856	481	10.68
CCGCTA	231	72	10.67
AAGATC	1515	993	10.65
AACGAC	372	151	10.63
CGACTA	224	69	10.61
GGAGAG	1942	1344	10.61
GATTAT	807	447	10.60
AAGATA	1164	718	10.58
TGGAAT	1794	1224	10.56
CCCCAA	1316	838	10.56
TATGAT	931	541	10.53
TACACC	786	434	10.52

Word	Ns	Na	z
CTGGTG	2497	1818	10.46
TATTAC	564	278	10.46
CTCAGA	1714	1164	10.44
CTGCGG	1095	670	10.42
TACGCC	281	101	10.40
CGAGAG	672	355	10.39
CAAAGA	2343	1692	10.38
GCTCAA	1111	684	10.37
CAACAG	1666	1128	10.37
CCCTCA	1335	861	10.35
TGTGAC	1208	761	10.34
ATGCAA	1067	651	10.34
AACCAG	1561	1044	10.33
TGCAAG	1381	899	10.32
ACAAGT	1056	643	10.32
TCTTAC	600	306	10.31
AAAAGC	1461	965	10.28
CTGGCC	2519	1848	10.27
GCGAGA	561	280	10.27
TTTATG	1073	659	10.24
GAGACA	1400	918	10.23
AATATG	1151	721	10.21
CTGTGG	2473	1814	10.18
AGAAGT	1907	1336	10.18
GAATGG	1325	860	10.18
AACCAA	1145	718	10.16
AACTAG	390	169	10.15
AGGAGG	2573	1902	10.14
CAAGCG	455	211	10.14
AGTGTG	1295	838	10.13
CAACCA	1316	855	10.12
TGGCTA	934	556	10.12
ATGAAT	1330	867	10.10
CCTCAA	1449	963	10.10
GACCGA	430	196	10.06
AAAGTG	1447	963	10.06
TGAAGC	1960	1387	10.05
ACTGTG	1769	1228	10.04
TGATGA	2705	2022	10.04
AAATTG	1268	820	10.03
AAACCA	1499	1007	10.02
CGACAA	393	173	10.01
GGCCCT	1618	1105	10.01
CTGCGA	651	349	10.01
TGCCTG	1957	1387	10.00
AAGGAT	1375	908	9.98
CTACTA	548	277	9.97
CAACCG	412	186	9.96

Word	Ns	Na	z
CGCTAT	240	83	9.95
ACCAA	1371	907	9.93
AGAGAG	1653	1138	9.91
TGATCG	452	213	9.91
AAATCA	1514	1024	9.91
CCTGCT	2460	1819	9.91
ACCGGA	447	210	9.89
GAATGA	1313	862	9.88
CCCCCA	1473	992	9.87
GATCAA	1085	682	9.84
CCTATC	459	219	9.84
ACGAGT	395	177	9.84
AGATGG	2021	1451	9.80
CAGTGG	2208	1610	9.80
AAGTAC	927	561	9.78
AAGAGG	1872	1327	9.78
CTGCCC	2217	1619	9.77
TTCGAG	569	296	9.77
CTGTGA	2007	1442	9.75
CAACAT	1532	1046	9.74
TGTATG	921	558	9.73
TGACTA	555	287	9.73
TACATC	1038	649	9.73
GCCTGA	1330	882	9.72
AGATCG	463	224	9.71
GAGCGA	482	237	9.70
AACGTG	613	329	9.69
AAGTGA	1218	793	9.69
GTTACC	449	215	9.68
ACAGAT	1457	987	9.68
ACGAGG	730	415	9.67
CACCCA	1455	986	9.67
ACTGGA	1970	1415	9.67
TGCCAA	1679	1171	9.67
GAGAGC	1467	996	9.66
GAGCCT	1395	938	9.64
GCCGAG	894	540	9.64
GGACCA	1427	966	9.60
TGAACC	1068	677	9.60
TATGGC	782	456	9.60
CACCAA	1761	1244	9.57
CAACGC	402	186	9.56
GACCTT	1080	688	9.55
GATTGA	829	493	9.55
AAATAC	1116	717	9.54
AGTGCC	1288	856	9.52
GCCAGA	1674	1174	9.51
ACGTGG	885	537	9.51

Word	Ns	Na	z
TTACAG	1037	656	9.50
AACTAA	511	261	9.49
TTTGGA	1967	1422	9.48
GAACAG	1619	1131	9.45
AGTACA	1009	636	9.44
AACCCA	1001	630	9.43
GACAAA	1325	891	9.40
CGCAAG	411	195	9.37
CTGGAA	2900	2235	9.36
GAATAT	1231	817	9.34
CCTGAG	1814	1301	9.32
ATTTAC	769	454	9.31
TCCCTA	452	224	9.30
CTTATG	672	382	9.28
TGCCCC	1346	913	9.28
CTATAC	410	196	9.27
TCAAGG	1394	953	9.26
GCGCAA	293	121	9.26
GACGAC	468	236	9.24
TGCGAG	498	257	9.24
AGAGAT	1475	1021	9.24
ACAGAG	1751	1252	9.23
AGCTGG	2875	2222	9.22
TGACAA	1534	1072	9.19
CTTCGA	713	415	9.19
GGATGA	1959	1431	9.18
AACGGC	440	218	9.18
AGTTAC	602	333	9.17
GTACAA	829	504	9.17
GTCCTA	460	232	9.16
GTGTGA	1069	694	9.13
TACTAT	556	301	9.11
CGCCAA	500	261	9.11
TATACC	472	242	9.08
GAGATT	1110	729	9.08
AATCAA	1085	709	9.07
GAGATG	1855	1348	9.07
GCCTGG	2469	1877	9.06
GCCTTA	550	298	9.05
TGTACC	828	507	9.04
GAAGTG	1540	1084	9.04
GCCTGC	1854	1349	9.03
ACCTCT	1171	780	9.03
CGGGAC	616	348	8.98
GCCCAA	1047	682	8.97
ACGACG	279	116	8.97
ACCGGG	547	298	8.95
CCTGTC	1303	891	8.95

Word	Ns	Na	z
CAGTAC	954	609	8.94
AAGTTA	724	431	8.90
GAGTGG	1214	821	8.87
CTGAAG	2849	2222	8.87
GACCGC	436	221	8.86
GCACCC	1003	651	8.85
GACGTG	628	360	8.85
TTGAAA	1760	1279	8.83
ATGCTG	2250	1701	8.82
CCCCCT	1000	650	8.81
AGACCG	423	213	8.80
GAAATC	1367	950	8.80
GCCCCT	1204	816	8.79
CGACGA	262	108	8.77
AAAATT	1692	1225	8.76
GAAAGG	1309	904	8.75
GGAACA	1585	1135	8.75
AAAGCA	1716	1246	8.74
TATTGA	872	551	8.73
CGACCC	374	181	8.72
AATTAT	893	568	8.71
GAAGCC	1818	1335	8.70
AAGCCT	1179	800	8.68
TTCAAG	1639	1184	8.67
TGGGAC	1253	861	8.67
GAAATA	1266	873	8.64
AATACC	637	372	8.64
ACGACC	389	193	8.61
AAAATA	1492	1063	8.61
GGAGTG	1206	825	8.60
ACGCCT	470	250	8.60
TATGCA	780	483	8.59
GAGCCC	1427	1010	8.57
AGACAG	1614	1168	8.56
AATTAA	840	531	8.56
GAATAC	735	449	8.56
AGAATT	1534	1101	8.55
ACCGAC	343	163	8.54
ATGACA	1593	1152	8.52
GGCCAA	1421	1007	8.52
TACATG	929	603	8.52
ACATCG	613	357	8.51
GGAAGC	1819	1346	8.50
GTACGA	293	131	8.49
AGATGC	1569	1134	8.47
AAAACT	1431	1018	8.46
GCTAGA	473	255	8.45
AACGCC	415	214	8.44

Word	Ns	Na	z
GACGGC	631	373	8.42
CAGCCC	2140	1629	8.40
ATCAGA	1452	1039	8.39
ACAGAC	1211	838	8.38
TACGGG	267	116	8.37
GAATGC	992	659	8.36
AAGCCA	1821	1355	8.36
CCTATA	374	187	8.36
CAGACA	1751	1295	8.35
GCTGTG	2305	1776	8.35
CCAACG	509	284	8.32
AGCGAC	427	225	8.30
ATTGCC	950	627	8.30
ACGGGA	474	259	8.29
CATGAA	1753	1300	8.29
ATGACC	1239	865	8.28
GGCAAC	883	574	8.28
CCAGAC	1340	950	8.27
CCTGCA	2485	1940	8.25
ACTAAA	598	352	8.25
CAACGT	485	268	8.25
AGCCCC	1494	1081	8.24
TAATGA	1066	724	8.23
TACGGC	315	149	8.23
GAGGCC	1852	1389	8.21
TAAAAA	1202	837	8.21
TTATGC	570	332	8.21
GTGCCC	1257	884	8.18
TCCTAT	631	379	8.18
GCAACG	415	219	8.17
ACGGAA	459	251	8.15
TGACCG	541	312	8.13
ACTCTA	476	264	8.12
TGAGGA	2228	1722	8.12
ACCCCG	535	308	8.11
GACTTA	482	269	8.09
CCAGAG	2274	1765	8.07
GAGTGC	913	605	8.07
AGTGAC	1073	736	8.06
AGACAT	1277	906	8.05
TGAACG	497	281	8.05
CGGAAA	553	323	8.04
GAGCGG	705	440	8.04
TTCTAC	1015	689	8.04
CTAGAA	720	452	8.03
CAACTG	1401	1012	8.02
TGGTGG	2571	2031	8.01
CCAACA	1528	1121	8.00

Word	Ns	Na	z
CAGCGA	861	566	7.98
ATACAA	837	547	7.97
GGGAAG	1659	1235	7.96
TTAAAG	1054	724	7.96
CAACAC	1067	735	7.95
AAGAGC	1586	1173	7.95
AGCGAG	681	424	7.94
ATGCCA	1662	1239	7.94
TTACAC	573	341	7.92
CTAGAG	544	319	7.92
GACATC	1441	1051	7.91
ACCGCA	502	288	7.90
TGTGCC	1373	994	7.89
ATCTTA	590	355	7.88
AGAATA	1143	801	7.88
AATGAG	1244	886	7.87
GGAATG	1283	919	7.86
GGACAT	1339	967	7.85
TCACCC	1034	712	7.84
GATTTA	753	484	7.83
GCACTA	481	274	7.82
GCGACC	332	167	7.81
AATTTA	885	591	7.80
TTCGAC	369	194	7.74
CTAACC	291	140	7.74
GACCAT	1061	738	7.74
GGAATT	1270	914	7.72
GTGTGC	1054	733	7.71
CGGACC	421	232	7.71
GGCTAT	650	406	7.71
GCAACA	1243	892	7.70
ACTCAA	836	555	7.69
CCCCAG	2203	1725	7.68
TCCTGA	2003	1549	7.68
GACGGA	474	272	7.67
ATTGTG	1046	728	7.67
CTATTC	575	349	7.66
GAGCGC	571	346	7.66
CTGCTA	940	641	7.65
AATGCA	1254	904	7.63
AGCTGA	2094	1632	7.63
GAGTAT	691	441	7.61
AACGGA	346	180	7.61
CCAAGG	1733	1317	7.60
ACGACA	445	252	7.60
GCCTGT	1319	961	7.59
CCCAGC	2397	1903	7.58
CGGCTA	277	133	7.57

Word	Ns	Na	z
ACATTA	716	462	7.57
CAGTGA	1928	1489	7.57
AGCTAA	567	345	7.57
ATTACT	722	467	7.56
GTTACA	645	406	7.56
ACAGTG	1718	1306	7.56
TGAACA	1597	1201	7.56
CCGCCT	853	573	7.56
GAATTG	1008	701	7.54
GGGAGA	1517	1133	7.54
GCGGAG	663	421	7.53
CCGAAA	423	237	7.53
CGAAAA	394	216	7.52
ACCGAA	376	203	7.52
GTGCTA	497	293	7.50
ACCGCC	497	293	7.50
GCAACT	925	634	7.50
CGGAGA	727	473	7.50
GACCAG	1524	1141	7.49
ACGACT	371	200	7.48
CCCGAG	767	506	7.47
AGTACG	289	143	7.44
TTAGAA	892	609	7.43
CGTGTG	640	406	7.42
CCGACC	394	218	7.41
GACATG	1200	868	7.39
CGAGCG	261	125	7.37
ATACTA	386	213	7.37
GATACC	549	336	7.37
TACAGC	1168	842	7.37
GTGAAC	975	680	7.36
CCAGTG	2022	1583	7.36
ACCCTT	691	448	7.36
TCTACT	877	599	7.36
CCTTCG	623	394	7.36
GCTAAA	607	382	7.34
AATGCC	1175	849	7.34
ACCCAC	931	646	7.29
TGAAGG	2177	1725	7.28
CGAGCT	629	401	7.28
AACATG	1297	956	7.26
AAGCAC	1111	798	7.26
ACAAAT	1235	904	7.24
ATGCAG	1779	1376	7.23
AACGGG	381	212	7.23
GATATG	717	473	7.22
GTGGCC	1902	1485	7.22
GCCCCC	1324	983	7.18

Word	Ns	Na	z
ACACCT	1119	808	7.17
TACGGA	276	138	7.17
ACGCTA	160	63	7.17
ACTAGA	449	264	7.16
TACCAC	771	519	7.15
GCGAGC	463	275	7.15
AGATAT	944	662	7.14
TTACTG	934	654	7.13
CGCAAC	265	131	7.13
CCCTTA	392	222	7.13
ACGTGA	461	274	7.12
ACCACC	1537	1170	7.12
CTTACG	187	80	7.11
ATCACC	1281	949	7.11
GATGTG	1488	1128	7.10
GCAAAA	1332	993	7.10
AGGACA	1709	1322	7.08
GGGAAC	960	678	7.07
TCAACT	1120	813	7.07
GATCCT	1231	908	7.06
AACCTT	826	567	7.06
TCAAGT	1193	876	7.05
AACAGC	1642	1266	7.03
AGATCC	1308	976	7.02
TACTTA	385	219	7.01
TGCAAC	955	676	7.01
CAGTGC	1671	1293	7.00
CAAACC	927	653	6.99
CTATTG	593	380	6.99
ACGAGC	452	270	6.99
AAATGC	1260	936	6.99
AAACGA	404	234	6.97
ACAATG	1329	997	6.95
ACATGA	1221	904	6.95
GACTCA	860	599	6.94
GAGCAG	2388	1935	6.93
GGCCGA	732	494	6.92
CCATCG	723	487	6.91
AGAGTG	1136	833	6.91
AGATAC	768	524	6.91
AGAGCA	1710	1332	6.90
GAACGG	424	251	6.88
ACGTGC	564	360	6.87
ATGCCT	1085	791	6.87
GCTCTA	611	398	6.85
AACCGC	339	188	6.85
ACAAAG	1579	1220	6.84
ACGGAG	681	455	6.84

Word	Ns	Na	z
TGTACG	340	189	6.84
AGCCAA	1521	1170	6.82
ACATGC	953	681	6.82
AACAGT	1105	810	6.82
GGACGG	576	371	6.82
GTACTA	313	170	6.80
CTAGAC	306	165	6.79
ATTAAA	1014	734	6.78
AATGAC	1121	826	6.76
TACAAT	721	490	6.76
CGCCCT	609	399	6.76
TCCCTG	1803	1422	6.75
AGCGGC	827	578	6.74
CCTAAC	284	150	6.74
GCCGAC	476	294	6.74
AACGAA	362	207	6.74
ACCAGT	1191	887	6.74
TCCTAA	592	386	6.73
TCACCG	567	366	6.73
CATCGC	567	366	6.73
AAAGCC	1365	1038	6.73
TGCGGC	719	490	6.70
TTACGT	191	87	6.69
TGTTAC	629	417	6.69
TTACTA	519	329	6.69
ATACCC	456	280	6.67
ACCCGC	421	253	6.67
GAAACG	430	260	6.67
CCTAAG	392	231	6.66
TGCTAA	714	487	6.66
TCAGAT	1350	1028	6.66
TTACGA	238	119	6.66
ACTGCG	471	292	6.66
GACAGC	1694	1331	6.65
TGGATT	1451	1117	6.64
GGGAGC	1428	1097	6.64
CCAGAT	1609	1257	6.62
ACCTTT	923	663	6.61
ACCGTG	660	445	6.59
ACCTTC	1456	1125	6.57
CGACCG	167	73	6.56
GATGCT	1646	1294	6.54
GTGGTG	1845	1471	6.53
ATTCAG	1429	1103	6.53
CTATCG	179	81	6.53
ACCGCT	421	256	6.53
AGGAAG	2601	2153	6.53
ATGACT	1109	825	6.53

Word	Ns	Na	z
TGGAGT	1642	1292	6.51
CTAGTG	432	265	6.50
CCGCAA	420	256	6.49
TGAAAT	1633	1285	6.49
ATTTAT	1001	734	6.48
GTGTGG	1383	1065	6.48
GCTATT	640	432	6.47
AGCTAG	405	245	6.46
AGCAAA	1664	1314	6.46
TACGAA	275	148	6.46
ACATGG	1409	1089	6.45
ATTAGA	600	400	6.45
CCTGGT	1823	1456	6.45
CGTGAA	544	355	6.44
CGCGAG	212	104	6.44
CCCTAG	306	171	6.43
TATTGG	760	533	6.41
GACACC	1107	828	6.41
CCTTAT	624	421	6.40
TGTGCG	472	299	6.38
TTCGTG	531	346	6.38
CAAAAC	1107	829	6.38
TCGACA	395	239	6.38
ATGATG	2261	1854	6.37
CTATCC	478	304	6.37
GGATTA	569	377	6.37
GGAGCC	1714	1363	6.37
CTAAAC	406	248	6.36
AGGAAC	1466	1144	6.35
AAGGGA	1027	762	6.33
CAGTTA	735	515	6.32
CTGAAC	1313	1011	6.32
TGAATG	1416	1102	6.30
AATACA	978	721	6.30
CAGAAT	1638	1299	6.30
CTGTGC	1800	1444	6.29
AGGCTA	488	314	6.29
AACCGA	312	178	6.28
GACAAT	965	711	6.27
TTACAT	762	539	6.27
CGAAAG	393	240	6.26
CTAAAA	772	548	6.25
TCGACT	347	205	6.24
GAACCG	400	246	6.23
GCAGTG	1812	1458	6.23
CAGATC	1355	1052	6.22
AGACGG	568	380	6.22
ACCATG	1366	1062	6.22

Word	Ns	Na	z
ACTTTG	1451	1137	6.22
TTTGAA	1878	1518	6.21
CCGAGG	933	686	6.21
ACGAAA	350	208	6.20
GTGCCT	1156	879	6.19
AACCGG	360	216	6.18
TCATCG	709	498	6.16
TACGTC	241	128	6.16
TTAAAA	1169	892	6.15
CATCTA	905	664	6.15
GCCTAA	274	152	6.15
TCAGAC	1041	781	6.15
TATATG	592	402	6.13
CCGGAA	697	489	6.13
GACAGT	1128	858	6.11
ACTCAC	624	429	6.11
AATGGG	981	731	6.10
CAAAAG	1510	1195	6.10
GTCAGA	1105	839	6.09
ATATGG	827	600	6.08
ACCCGA	310	180	6.08
GATGAC	1565	1246	6.05
CTGCCT	1997	1635	6.04
GCAGAC	1186	912	6.03
GAACAT	1285	999	6.03
TAGAAA	1039	784	6.03
AATCTA	515	342	6.03
AACAAT	1010	759	6.03
GTATGG	586	400	6.03
TGCGAC	296	171	5.99
GGAGGC	1912	1561	5.98
GGAACG	429	274	5.98
TTGTGG	1416	1117	5.98
TATGCT	696	493	5.97
TGTGGC	1712	1382	5.96
TCGACC	302	176	5.96
CCTCTA	737	528	5.95
CGTGAC	376	233	5.95
AAAAGT	1289	1006	5.95
GATTGC	598	412	5.95
AAAGCT	1466	1163	5.95
CCTCAC	1202	930	5.94
GCTGCG	983	739	5.94
TTGGAC	984	740	5.93
CAGCAA	2278	1897	5.92
CCCCAC	1312	1028	5.91
ACGGCA	587	404	5.91
ACTTCG	512	343	5.89

Word	Ns	Na	z
TAAGAA	1012	766	5.89
GACCCG	435	281	5.88
AAGACG	591	408	5.88
ATTACG	199	102	5.88
TGATTG	969	729	5.88
TGTGAG	1169	904	5.87
CTAATG	577	397	5.86
GAGTGT	858	634	5.86
TTTACA	1029	782	5.86
TTTGAC	1108	852	5.83
CGAAAT	333	202	5.83
CACAAC	902	673	5.83
CAGCCT	2266	1892	5.82
ACCATC	1457	1161	5.82
CCCGCC	726	523	5.82
GTACCC	450	295	5.80
GTGCGA	289	169	5.80
GGACTT	1152	893	5.77
CGACGG	271	156	5.76
CCCGAC	414	267	5.76
CTGCGC	738	535	5.76
GGAAT	848	629	5.76
ACTATT	583	405	5.75
AACCCG	284	166	5.75
TAACTA	268	154	5.75
TCGTGG	741	538	5.74
TACAGT	845	627	5.74
CGTCTA	223	121	5.74
TGTTCG	383	243	5.73
GAGACG	553	381	5.72
ATTGAC	802	591	5.72
CGCATC	488	328	5.70
ACGGGC	523	357	5.69
GAACGC	327	200	5.69
GCAGAG	2005	1662	5.69
ACTATA	428	280	5.68
TGGATA	1052	809	5.68
CGACAT	389	249	5.67
CTTTGA	1842	1515	5.67
TGCGGG	691	498	5.67
ATATTA	667	478	5.66
AACATC	1466	1177	5.65
TCGGGA	523	358	5.65
ATTTGA	1231	968	5.65
GCCCGA	524	359	5.65
GATTGG	733	535	5.62
TGGCCC	1703	1392	5.62
CATTAC	618	438	5.62

Word	Ns	Na	z
GCATCG	542	375	5.60
GACATT	1162	909	5.60
GCTGAA	2031	1691	5.59
GAATGT	1155	903	5.59
GGATAT	846	633	5.59
TACCAT	751	552	5.57
CGAGGC	725	530	5.57
CGTTAC	156	76	5.57
AGCTTA	508	348	5.56
GAACCA	1195	940	5.56
GAGACT	966	739	5.54
AATCAG	1126	881	5.51
GTGCTG	2161	1816	5.49
GCCCCA	1494	1210	5.49
GACTGC	1116	873	5.49
GTGGGA	1195	943	5.49
TTATGT	702	513	5.48
GGGCTA	473	322	5.45
TCACCT	1330	1065	5.45
AGTTAA	568	401	5.44
ACATTG	1200	950	5.43
CTCTAT	694	508	5.43
ACCCAT	724	534	5.41
TCAGTG	1625	1332	5.41
CTGATG	2028	1699	5.41
GGGGAA	1114	875	5.40
AATTGG	871	662	5.38
ACTTAA	512	356	5.38
AGTGGC	1347	1084	5.36
GTTTGA	1101	865	5.36
GGACGC	475	326	5.35
TGGGGA	1677	1383	5.34
CTCAGT	1178	935	5.32
ACGGAC	362	235	5.31
TACACA	809	611	5.30
CACAAG	1158	919	5.28
TGCTGG	3051	2654	5.27
ATATTG	917	707	5.25
AACTCA	1012	791	5.24
AGCAGT	1668	1380	5.24
CCGCCC	624	454	5.24
ATCTAT	631	460	5.24
GTTACG	158	81	5.23
CGACGT	220	126	5.23
TCGCTA	162	84	5.22
AGGACT	1229	985	5.21
GGGGAC	1021	800	5.21
ACATCC	1343	1088	5.20

Word	Ns	Na	z
AAGTAT	841	643	5.18
ACACTA	430	293	5.18
ACATCA	1760	1467	5.18
CGCCAT	564	405	5.17
CTATTA	458	317	5.14
TTATCG	194	108	5.14
ACACAA	920	714	5.13
GATGAG	1812	1517	5.13
TAACCC	331	214	5.12
CACCGA	569	411	5.11
GATCGG	280	174	5.10
AGGACG	688	513	5.10
GCACCT	1222	984	5.09
GGCAAG	1055	835	5.09
ACTTAT	503	356	5.08
CGACTG	461	321	5.08
TTACGG	201	114	5.08
GTTTAC	530	379	5.07
TGAGCG	518	369	5.07
ACGTGT	434	299	5.07
AAAACG	435	300	5.06
GCAAGT	894	694	5.06
CATCGT	685	512	5.05
TCGTGA	376	252	5.04
ATCCCT	868	672	5.03
CAGAGG	2006	1701	5.03
GCTTCG	545	393	5.02
TACGCG	92	39	5.01
TTGGAG	1844	1553	5.01