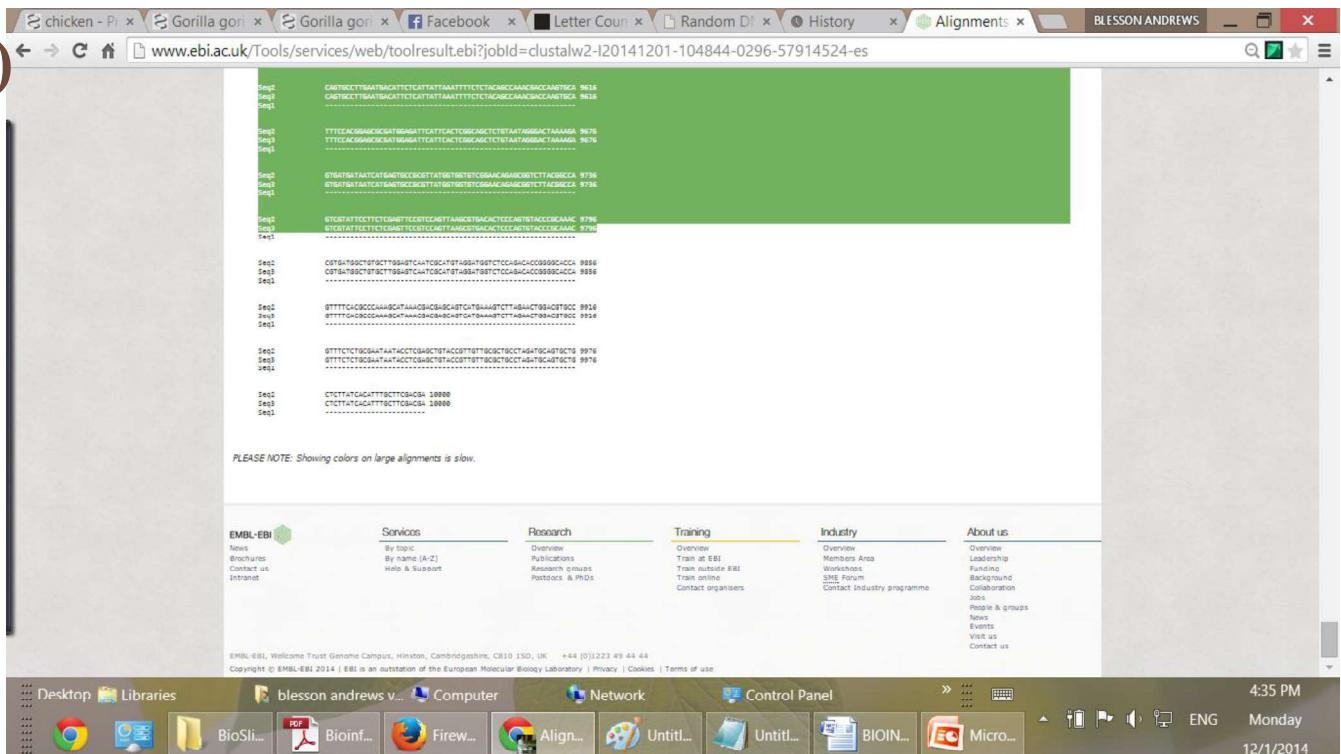


(xiv)

(xv)

(xvi)



1.5 Home Address Vs Protein Sequence

Home Address: THUNDIYIL HOUSE POST OFFICE LANE NALANCHIRA P O TRIVANDRUM KERALA INDIA

Seq1: Mutant Beta-Glucosidase In Wheat Complexed With Dimbooa-Glc

Organism: Triticum aestivum (bread wheat) **PDB:** 3AIS_A

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDKAMAGTPSKPAEPIGPVF
TKLKPWQIPK

Input

chicken x Gorill... x wheat x Chain x Chain x Chain x hare - x Letter x Random x History x Clustal x BLESSON ANDREWS

www.ebi.ac.uk/Tools/msa/clustalw2/

ClustalW2

[Input form](#) | [Web services](#) | [Help & Documentation](#)

Tools > Multiple Sequence Alignment > ClustalW2

[Share](#) | [Feedback](#)

Multiple Sequence Alignment

ClustalW2 is a general purpose DNA or protein multiple sequence alignment program for **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Note: ClustalW2 is no longer being maintained. Please consider using the new version instead: [Clustal Omega](#)

STEP 1 - Enter your input sequences

Enter or paste a set of Protein sequences in any supported format:

```
>Seq1
SGLVPRGSGMKTAAKFERQHMDSPDGLTDDDKAMAGTPSKPAEPIGPVFTKLKPWQPK
>Seq2
MHHHHHHSSGLVPRGSGMKTAAKFERQHMDSPDGLTDDDKAMAGTPSKPAEPIGPVFTKLKPWQPK
```

Or, upload a file: [Choose File](#) No file chosen

STEP 2 - Set your Pairwise Alignment Options

Alignment Type: Slow Fast

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 3 - Set your Multiple Sequence Alignment Options

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 4 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

Output:

chicken x Gorill x wheat x Chain x Chain x hare - x Letter x Random x History x Align x BLESSON ANDREWS

www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2-l20141201-112624-0080-22902243-oy

EMBL-EBI Services Research Training About us

ClustalW2

Input form Web services Help & Documentation Share Feedback

Tools > Multiple Sequence Alignment > ClustalW2

Results for job clustalw2-l20141201-112624-0080-22902243-oy

Alignments Result Summary Guide Tree Phylogenetic Tree Submission Details

Download Alignment File Show Colors Send to ClustalW2_Phylogeny

CLUSTAL 2.1 multiple sequence alignment

	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10	Seq11	Seq12	Seq13	Seq14	Seq15	Seq16	Seq17	Seq18	Seq19	Seq20	Seq21	Seq22	Seq23	Seq24	Seq25	Seq26	Seq27	Seq28	Seq29	Seq30	Seq31	Seq32	Seq33	Seq34	Seq35	Seq36	Seq37	Seq38	Seq39	Seq40	Seq41	Seq42	Seq43	Seq44	Seq45	Seq46	Seq47	Seq48	Seq49	Seq50	Seq51	Seq52	Seq53	Seq54	Seq55	Seq56	Seq57	Seq58	Seq59	Seq60	Seq61	Seq62	Seq63	Seq64	Seq65	Seq66	Seq67	Seq68	Seq69	Seq70	Seq71	Seq72	Seq73	Seq74	Seq75	Seq76	Seq77	Seq78	Seq79	Seq80	Seq81	Seq82	Seq83	Seq84	Seq85	Seq86	Seq87	Seq88	Seq89	Seq90	Seq91	Seq92	Seq93	Seq94	Seq95	Seq96	Seq97	Seq98	Seq99	Seq100	Seq101	Seq102	Seq103	Seq104	Seq105	Seq106	Seq107	Seq108	Seq109	Seq110	Seq111	Seq112	Seq113	Seq114	Seq115	Seq116	Seq117	Seq118	Seq119	Seq120	Seq121	Seq122	Seq123	Seq124	Seq125	Seq126	Seq127	Seq128	Seq129	Seq130	Seq131	Seq132	Seq133	Seq134	Seq135	Seq136	Seq137	Seq138	Seq139	Seq140	Seq141	Seq142	Seq143	Seq144	Seq145	Seq146	Seq147	Seq148	Seq149	Seq150	Seq151	Seq152	Seq153	Seq154	Seq155	Seq156	Seq157	Seq158	Seq159	Seq160	Seq161	Seq162	Seq163	Seq164	Seq165	Seq166	Seq167	Seq168	Seq169	Seq170	Seq171	Seq172	Seq173	Seq174	Seq175	Seq176	Seq177	Seq178	Seq179	Seq180	Seq181	Seq182	Seq183	Seq184	Seq185	Seq186	Seq187	Seq188	Seq189	Seq190	Seq191	Seq192	Seq193	Seq194	Seq195	Seq196	Seq197	Seq198	Seq199	Seq200	Seq201	Seq202	Seq203	Seq204	Seq205	Seq206	Seq207	Seq208	Seq209	Seq210	Seq211	Seq212	Seq213	Seq214	Seq215	Seq216	Seq217	Seq218	Seq219	Seq220	Seq221	Seq222	Seq223	Seq224	Seq225	Seq226	Seq227	Seq228	Seq229	Seq230	Seq231	Seq232	Seq233	Seq234	Seq235	Seq236	Seq237	Seq238	Seq239	Seq240	Seq241	Seq242	Seq243	Seq244	Seq245	Seq246	Seq247	Seq248	Seq249	Seq250	Seq251	Seq252	Seq253	Seq254	Seq255	Seq256	Seq257	Seq258	Seq259	Seq260	Seq261	Seq262	Seq263	Seq264	Seq265	Seq266	Seq267	Seq268	Seq269	Seq270	Seq271	Seq272	Seq273	Seq274	Seq275	Seq276	Seq277	Seq278	Seq279	Seq280	Seq281	Seq282	Seq283	Seq284	Seq285	Seq286	Seq287	Seq288	Seq289	Seq290	Seq291	Seq292	Seq293	Seq294	Seq295	Seq296	Seq297	Seq298	Seq299	Seq300	Seq301	Seq302	Seq303	Seq304	Seq305	Seq306	Seq307	Seq308	Seq309	Seq310	Seq311	Seq312	Seq313	Seq314	Seq315	Seq316	Seq317	Seq318	Seq319	Seq320	Seq321	Seq322	Seq323	Seq324	Seq325	Seq326	Seq327	Seq328	Seq329	Seq330	Seq331	Seq332	Seq333	Seq334	Seq335	Seq336	Seq337	Seq338	Seq339	Seq340	Seq341	Seq342	Seq343	Seq344	Seq345	Seq346	Seq347	Seq348	Seq349	Seq350	Seq351	Seq352	Seq353	Seq354	Seq355	Seq356	Seq357	Seq358	Seq359	Seq360	Seq361	Seq362	Seq363	Seq364	Seq365	Seq366	Seq367	Seq368	Seq369	Seq370	Seq371	Seq372	Seq373	Seq374	Seq375	Seq376	Seq377	Seq378	Seq379	Seq380	Seq381	Seq382	Seq383	Seq384	Seq385	Seq386	Seq387	Seq388	Seq389	Seq390	Seq391	Seq392	Seq393	Seq394	Seq395	Seq396	Seq397	Seq398	Seq399	Seq400	Seq401	Seq402	Seq403	Seq404	Seq405	Seq406	Seq407	Seq408	Seq409	Seq410	Seq411	Seq412	Seq413	Seq414	Seq415	Seq416	Seq417	Seq418	Seq419	Seq420	Seq421	Seq422	Seq423	Seq424	Seq425	Seq426	Seq427	Seq428	Seq429	Seq430	Seq431	Seq432	Seq433	Seq434	Seq435	Seq436	Seq437	Seq438	Seq439	Seq440	Seq441	Seq442	Seq443	Seq444	Seq445	Seq446	Seq447	Seq448	Seq449	Seq450	Seq451	Seq452	Seq453	Seq454	Seq455	Seq456	Seq457	Seq458	Seq459	Seq460	Seq461	Seq462	Seq463	Seq464	Seq465	Seq466	Seq467	Seq468	Seq469	Seq470	Seq471	Seq472	Seq473	Seq474	Seq475	Seq476	Seq477	Seq478	Seq479	Seq480	Seq481	Seq482	Seq483	Seq484	Seq485	Seq486	Seq487	Seq488	Seq489	Seq490	Seq491	Seq492	Seq493	Seq494	Seq495	Seq496	Seq497	Seq498	Seq499	Seq500	Seq501	Seq502	Seq503	Seq504	Seq505	Seq506	Seq507	Seq508	Seq509	Seq510	Seq511	Seq512	Seq513	Seq514	Seq515	Seq516	Seq517	Seq518	Seq519	Seq520	Seq521	Seq522	Seq523	Seq524	Seq525	Seq526	Seq527	Seq528	Seq529	Seq530	Seq531	Seq532	Seq533	Seq534	Seq535	Seq536	Seq537	Seq538	Seq539	Seq540	Seq541	Seq542	Seq543	Seq544	Seq545	Seq546	Seq547	Seq548	Seq549	Seq550	Seq551	Seq552	Seq553	Seq554	Seq555	Seq556	Seq557	Seq558	Seq559	Seq560	Seq561	Seq562	Seq563	Seq564	Seq565	Seq566	Seq567	Seq568	Seq569	Seq570	Seq571	Seq572	Seq573	Seq574	Seq575	Seq576	Seq577	Seq578	Seq579	Seq580	Seq581	Seq582	Seq583	Seq584	Seq585	Seq586	Seq587	Seq588	Seq589	Seq590	Seq591	Seq592	Seq593	Seq594	Seq595	Seq596	Seq597	Seq598	Seq599	Seq600	Seq601	Seq602	Seq603	Seq604	Seq605	Seq606	Seq607	Seq608	Seq609	Seq610	Seq611	Seq612	Seq613	Seq614	Seq615	Seq616	Seq617	Seq618	Seq619	Seq620	Seq621	Seq622	Seq623	Seq624	Seq625	Seq626	Seq627	Seq628	Seq629	Seq630	Seq631	Seq632	Seq633	Seq634	Seq635	Seq636	Seq637	Seq638	Seq639	Seq640	Seq641	Seq642	Seq643	Seq644	Seq645	Seq646	Seq647	Seq648	Seq649	Seq650	Seq651	Seq652	Seq653	Seq654	Seq655	Seq656	Seq657	Seq658	Seq659	Seq660	Seq661	Seq662	Seq663	Seq664	Seq665	Seq666	Seq667	Seq668	Seq669	Seq670	Seq671	Seq672	Seq673	Seq674	Seq675	Seq676	Seq677	Seq678	Seq679	Seq680	Seq681	Seq682	Seq683	Seq684	Seq685	Seq686	Seq687	Seq688	Seq689	Seq690	Seq691	Seq692	Seq693	Seq694	Seq695	Seq696	Seq697	Seq698	Seq699	Seq700	Seq701	Seq702	Seq703	Seq704	Seq705	Seq706	Seq707	Seq708	Seq709	Seq710	Seq711	Seq712	Seq713	Seq714	Seq715	Seq716	Seq717	Seq718	Seq719	Seq720	Seq721	Seq722	Seq723	Seq724	Seq725	Seq726	Seq727	Seq728	Seq729	Seq730	Seq731	Seq732	Seq733	Seq734	Seq735	Seq736	Seq737	Seq738	Seq739	Seq740	Seq741	Seq742	Seq743	Seq744	Seq745	Seq746	Seq747	Seq748	Seq749	Seq750	Seq751	Seq752	Seq753	Seq754	Seq755	Seq756	Seq757	Seq758	Seq759	Seq760	Seq761	Seq762	Seq763	Seq764	Seq765	Seq766	Seq767	Seq768	Seq769	Seq770	Seq771	Seq772	Seq773	Seq774	Seq775	Seq776	Seq777	Seq778	Seq779	Seq780	Seq781	Seq782	Seq783	Seq784	Seq785	Seq786	Seq787	Seq788	Seq789	Seq790	Seq791	Seq792	Seq793	Seq794	Seq795	Seq796	Seq797	Seq798	Seq799	Seq800	Seq801	Seq802	Seq803	Seq804	Seq805	Seq806	Seq807	Seq808	Seq809	Seq810	Seq811	Seq812	Seq813	Seq814	Seq815	Seq816	Seq817	Seq818	Seq819	Seq820	Seq821	Seq822	Seq823	Seq824	Seq825	Seq826	Seq827	Seq828	Seq829	Seq830	Seq831	Seq832	Seq833	Seq834	Seq835	Seq836	Seq837	Seq838	Seq839	Seq840	Seq841	Seq842	Seq843	Seq844	Seq845	Seq846	Seq847	Seq848	Seq849	Seq850	Seq851	Seq852	Seq853	Seq854	Seq855	Seq856	Seq857	Seq858	Seq859	Seq860	Seq861	Seq862	Seq863	Seq864	Seq865	Seq866	Seq867	Seq868	Seq869	Seq870	Seq871	Seq872	Seq873	Seq874	Seq875	Seq876	Seq877	Seq878	Seq879	Seq880	Seq881	Seq882	Seq883	Seq884	Seq885	Seq886	Seq887	Seq888	Seq889	Seq890	Seq891	Seq892	Seq893	Seq894	Seq895	Seq896	Seq897	Seq898	Seq899	Seq900	Seq901	Seq902	Seq903	Seq904	Seq905	Seq906	Seq907	Seq908	Seq909	Seq910	Seq911	Seq912	Seq913	Seq914	Seq915	Seq916	Seq917	Seq918	Seq919	Seq920	Seq921	Seq922	Seq923	Seq924	Seq925	Seq926	Seq927	Seq928	Seq929	Seq930	Seq931	Seq932	Seq933	Seq934	Seq935	Seq936	Seq937	Seq938	Seq939	Seq940	Seq941	Seq942	Seq943	Seq944	Seq945	Seq946	Seq947	Seq948	Seq949	Seq950	Seq951	Seq952	Seq953	Seq954	Seq955	Seq956	Seq957	Seq958	Seq959	Seq960	Seq961	Seq962	Seq963	Seq964	Seq965	Seq966	Seq967	Seq968	Seq969	Seq970	Seq971	Seq972	Seq973	Seq974	Seq975	Seq976	Seq977	Seq978	Seq979	Seq980	Seq981	Seq982	Seq983	Seq984	Seq985	Seq986	Seq987	Seq988	Seq989	Seq990	Seq991	Seq992	Seq993	Seq994	Seq995	Seq996	Seq997	Seq998	Seq999	Seq1000	Seq1001	Seq1002	Seq1003	Seq1004	Seq1005	Seq1006	Seq1007	Seq1008	Seq1009	Seq1010	Seq1011	Seq1012	Seq1013	Seq1014	Seq1015	Seq1016	Seq1017	Seq1018	Seq1019	Seq1020	Seq1021	Seq1022	Seq1023	Seq1024	Seq1025	Seq1026	Seq1027	Seq1028	Seq1029	Seq1030	Seq1031	Seq1032	Seq1033	Seq1034	Seq1035	Seq1036	Seq1037	Seq1038	Seq1039	Seq1040	Seq1041	Seq1042	Seq1043	Seq1044	Seq1045	Seq1046	Seq1047	Seq1048	Seq1049	Seq1050	Seq1051	Seq1052	Seq1053	Seq1054	Seq1055	Seq1056	Seq1057	Seq1058	Seq1059	Seq1060	Seq1061	Seq1062	Seq1063	Seq1064	Seq1065	Seq1066	Seq1067	Seq1068	Seq1069	Seq1070	Seq1071	Seq1072	Seq1073	Seq1074	Seq1075	Seq1076	Seq1077	Seq1078	Seq1079	Seq1080	Seq1081	Seq1082	Seq1083	Seq1084	Seq1085	Seq1086	Seq1087	Seq1088	Seq1089	Seq1090	Seq1091	Seq1092	Seq1093	Seq1094	Seq1095	Seq1096	Seq1097	Seq1098	Seq1099	Seq1100	Seq1101	Seq1102	Seq1103	Seq1104	Seq1105	Seq1106	Seq1107	Seq1108	Seq1109	Seq1110	Seq1111	Seq1112	Seq1113	Seq1114	Seq1115	Seq1116	Seq1117	Seq1118	Seq1119	Seq1120	Seq1121	Seq1122	Seq1123	Seq1124	Seq1125	Seq1126	Seq1127	Seq1128	Seq1129	Seq1130	Seq1131	Seq1132	Seq1133	Seq1134	Seq1135	Seq1136	Seq1137	Seq1138	Seq1139	Seq1140	Seq1141	Seq1142	Seq1143	Seq1144	Seq1145	Seq1146	Seq1147	Seq1148	Seq1149	Seq1150	Seq1151	Seq1152	Seq1153	Seq1154	Seq1155	Seq1156	Seq1157	Seq1158	Seq1159	Seq1160	Seq1161	Seq1162	Seq1163	Seq1164	Seq1165	Seq1166	Seq1167	Seq1168	Seq1169	Seq1170	Seq1171	Seq1172	Seq1173	Seq1174	Seq1175	Seq1176	Seq1177	Seq1178	Seq1179	Seq1180	Seq1181	Seq1182	Seq1183	Seq1184	Seq1185	Seq1186	Seq1187	Seq1188	Seq1189	Seq1190	Seq1191	Seq1192	Seq1193	Seq1194	Seq1195	Seq1196	Seq1197	Seq1198	Seq1199	Seq1200	Seq1201	Seq1202	Seq1203	Seq1204	Seq1205	Seq1206	Seq1207	Seq1208	Seq1209	Seq1210	Seq1211	Seq1212	Seq1213	Seq1214	Seq1215	Seq1216	Seq1217	Seq1218	Seq1219	Seq1220	Seq1221	Seq1222	Seq1223	Seq1224	Seq1225	Seq1226	Seq1227	Seq1228	Seq1229	Seq1230	Seq1231	Seq1232	Seq1233	Seq1234	Seq1235	Seq1236	Seq1237	Seq1238	Seq1239	Seq1240	Seq1241	Seq1242	Seq1243	Seq1244	Seq1245	Seq1246	Seq1247	Seq1248	Seq1249	Seq1250	Seq1251	Seq1252	Seq1253	Seq1254	Seq1255	Seq1256	Seq1257	Seq1258	Seq1259	Seq1260	Seq1261	Seq1262	Seq1263	Seq1264	Seq1265	Seq1266	Seq1267	Seq1268	Seq1269	Seq1270	Seq1271	Seq1272	Seq1273	Seq1274	Seq1275	Seq1276	Seq1277	Seq1278	Seq1279	Seq1280	Seq1281	Seq1282	Seq1283	Seq1284	Seq1285	Seq1286	Seq1287	Seq1288	Seq1289	Seq1290	Seq1291	Seq1292	Seq1293	Seq1294	Seq1295	Seq1296	Seq1297	Seq1298	Seq1299	Seq1300	Seq1301	Seq1302	Seq1303	Seq1304	Seq1305	Seq1306	Seq1307	Seq1308	Seq1309	Seq1310	Seq1311	Seq1312	Seq1313	Seq1314	Seq1315	Seq1316	Seq1317	Seq1318	Seq1319	Seq132
--	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	--------

1.6 Random DNA Vs Home Address

Random DNA :

CTGCCGCCTCGCTGTTCCCTAGACACTCAACAATAAGCGCTTTGTAGGCAGGGC
ACC

Home Address: THUNDIYIL HOUSE POST OFFICE LANE NALANCHIRA P O
TRIVANDRUM KERALA INDIA

Input

chicken - P x Gorilla gorilla - Gorilla gorilla - Facebook Letter Count Random DI History ClustalW2 BLESSON ANDREWS

www.ebi.ac.uk/Tools/msa/clustalw2/

ClustalW2

Input form | Web services | Help & Documentation

Tools > Multiple Sequence Alignment > ClustalW2

Multiple Sequence Alignment

ClustalW2 is a general purpose DNA or protein multiple sequence alignment program for **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Note: ClustalW2 is no longer being maintained. Please consider using the new version instead: [Clustal Omega](#)

STEP 1 - Enter your input sequences

Enter or paste a set of Protein sequences in any supported format:

Sequences:
Sseq1: EUNQVILHOUSEPQSTOFIEELANENALANCHIRAPOTRIVANDRUNKERALAINDA
Sseq2: CTGCGCGCTTCGGTGTTCCTAGACACTAAACATAAGCCCTTTGTAGGCAGGGGACC

Or, upload a file: [Choose File](#) No file chosen

STEP 2 - Set your Pairwise Alignment Options

Alignment Type: Slow / Fast

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 3 - Set your Multiple Sequence Alignment Options

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 4 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

Desktop Libraries blesson andrews v... Computer Network Control Panel 4:40 PM ENG Monday 12/1/2014

Output

Waiting for cdn.spotflux.com...

EMBL-EBI

ClustalW2

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Multiple Sequence Alignment > ClustalW2

Results for job clustalw2-l20141201-111243-0261-47427062-es

Alignments Result Summary Guide Tree Phylogenetic Tree Submission Details

Download Alignment File Show Colors Send to ClustalW2_Phylogeny

CLUSTAL 2.1 multiple sequence alignment

Seq1	-THUDIVTLIHKVSPDSTOPPTCEIEN-ENLIAHCTTGTGTTWANDRIRIPFALATH 58
Seq2	CTGCCGCCTTG-CCTG-TT-TCCCTAGACACTCAACATAAGGCGCTTTTGTAGCAGGG 57
Seq1	DAA-- 61
Seq2	GCAAC 62

PLEASE NOTE: Showing colors on large alignments is slow.

EMBL-EBI

News
Brochures
Contact us
Intranet

Services

By topic
By name (A-Z)
Help & Support

Research

Overview
Publications
Research groups
Postdocs & PhDs

Training

Overview
Train at EBI
Train outside EBI
Train online
Contact organisers

Industry

Overview
Members Area
Workshops
SME Forum
Contact Industry programme

About us

Overview
Leadership
Funding
Background
Collaboration
Jobs
People & groups
Press
Events
Visit us
Contact us

4:43 PM 12/1/2014

2. BLAST
Basic Local Alignment Search Tool

2.1 Human DNA Sequence

Seq1: Homo sapiens cDNA

Organism: Human ; GenBank: BC047343.2

CCCACGCGTCCGTTCTCTCCAAGGAACCTTTGTTGTTTATAAAGCAATGGGAAATCC
AGAAAAACAT

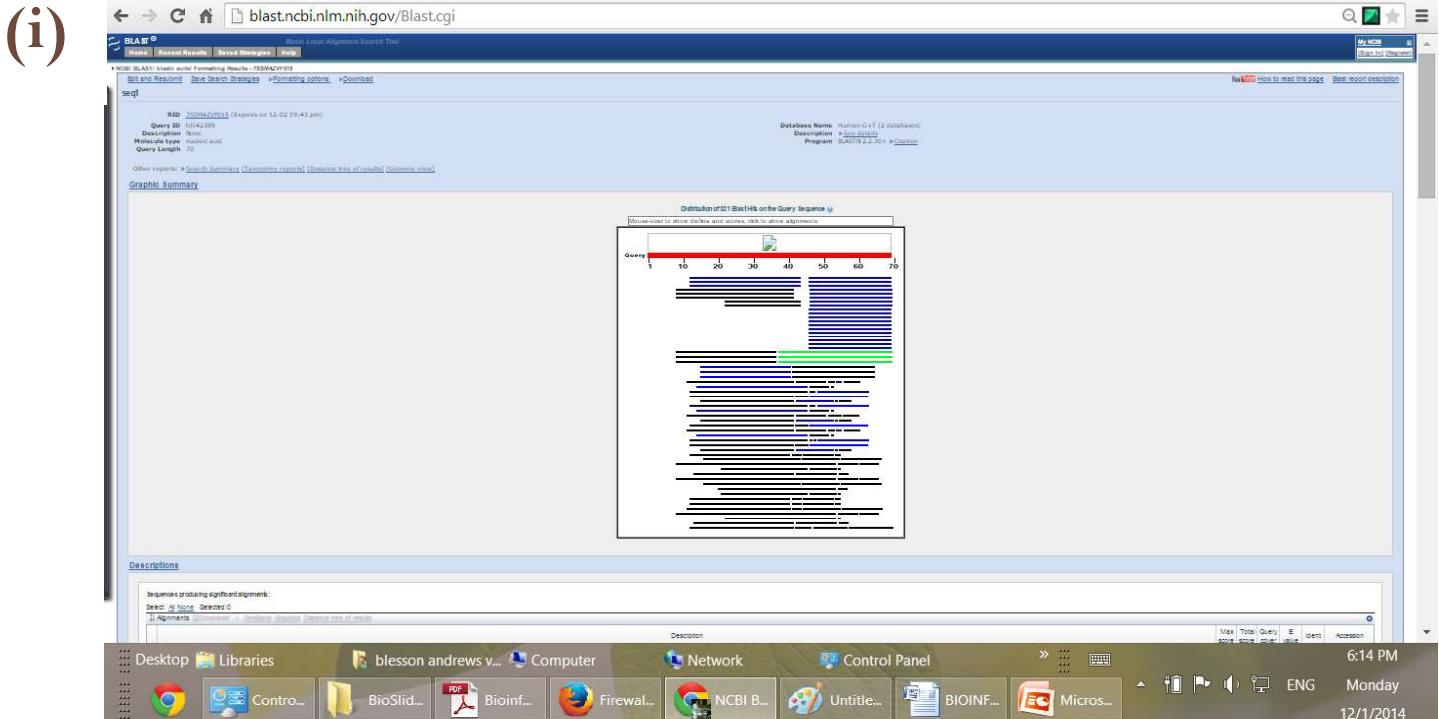
Input

The screenshot shows the NCBI BLAST search interface on a Windows desktop. The URL in the address bar is `blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome`. The search form contains the following input fields:

- Enter accession number(s), gi(s), or Fasta sequence(s):** CCCACGGCTCCGTCTTCCAAGGAACTTTTTGTTTATAAAGCAATGGAAATTCCAGAAAAACAT
- From:** (empty)
- To:** (empty)
- Or, upload file:** Choose File (No file chosen)
- Job Title:** seq1
- Enter a descriptive title for your BLAST search:** (empty)
- Align two or more sequences:** (unchecked)
- Choose Search Set:**
 - Database:** Human genomic + transcript (radio button selected)
 - Exclude:** Models (XM/XP) (checkbox)
 - Limit to:** Uncultured/environmental sample sequences (checkbox)
- Entrez Query:** (empty)
- Program Selection:**
 - Optimize for:** Somewhat similar sequences (blastn) (radio button selected)
 - Choose a BLAST algorithm:** (empty)

The status bar at the bottom shows the search parameters: **BLAST** | Search database Human G+T using Blastn (Optimize for somewhat similar sequences).

Output



(ii)

blast.ncbi.nlm.nih.gov/Blast.cgi

Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Transcripts						
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X15_mRNA	44.6	44.6	34%	0.014	100%	XM_006716164.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X14_mRNA	44.6	44.6	34%	0.014	100%	XM_006716163.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X13_mRNA	44.6	44.6	34%	0.014	100%	XM_006716162.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X12_mRNA	44.6	44.6	34%	0.014	100%	XM_006716161.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X11_mRNA	44.6	44.6	34%	0.014	100%	XM_006716160.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X10_mRNA	44.6	44.6	34%	0.014	100%	XM_005250669.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X9_mRNA	44.6	44.6	34%	0.014	100%	XM_005250668.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X8_mRNA	44.6	44.6	34%	0.014	100%	XM_005250667.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X7_mRNA	44.6	44.6	34%	0.014	100%	XM_005250666.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X6_mRNA	44.6	44.6	34%	0.014	100%	XM_005250665.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X5_mRNA	44.6	44.6	34%	0.014	100%	XM_005250664.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X4_mRNA	44.6	44.6	34%	0.014	100%	XM_005250663.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X3_mRNA	44.6	44.6	34%	0.014	100%	XM_005250662.1
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X1_mRNA	44.6	44.6	34%	0.014	100%	XM_005250660.1

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:15 PM Monday 12/1/2014

(iii)

blast.ncbi.nlm.nih.gov/Blast.cgi

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X1_mRNA	44.6	44.6	34%	0.014	100%	XM_005250660.1
Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant 5_mRNA	44.6	44.6	34%	0.014	100%	NM_001013406.1
Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant 3_mRNA	44.6	44.6	34%	0.014	100%	NM_194454.1
Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant 2_mRNA	44.6	44.6	34%	0.014	100%	NM_004912.3
Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant 1_mRNA	44.6	44.6	34%	0.014	100%	NM_194456.1
Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant 4_mRNA	44.6	44.6	34%	0.014	100%	NM_194455.1
Genomic sequences [show first]						
Homo sapiens chromosome 7_alternate assembly CHM1_1.1	73.4	292	88%	3e-11	91%	NC_018918.2
Homo sapiens chromosome 7_alternate assembly HuRef	73.4	292	88%	3e-11	91%	AC_000139.1
Homo sapiens chromosome 7_GRCh38 Primary Assembly	73.4	292	88%	3e-11	91%	NC_000007.14
Homo sapiens chromosome 16_alternate assembly CHM1_1.1	42.8	78.3	71%	0.050	96%	NC_018927.2
Homo sapiens chromosome 16_alternate assembly HuRef	42.8	78.3	71%	0.050	96%	AC_000148.1
Homo sapiens chromosome 16_GRCh38 Primary Assembly	42.8	78.3	71%	0.050	96%	NC_000016.10
Homo sapiens chromosome 2_alternate assembly CHM1_1.1	41.0	407	70%	0.18	93%	NC_018913.2
Homo sapiens chromosome 6_alternate assembly CHM1_1.1	41.0	332	55%	0.18	88%	NC_018917.2
Homo sapiens chromosome 9_alternate assembly CHM1_1.1	41.0	147	72%	0.18	84%	NC_018920.2
Homo sapiens chromosome 11_alternate assembly CHM1_1.1	11.0	151	72%	0.18	81%	NC_018922.2
Homo sapiens chromosome 12_alternate assembly CHM1_1.1	41.0	256	61%	0.18	90%	NC_018923.2
Homo sapiens chromosome 20_alternate assembly CHM1_1.1	41.0	41.0	45%	0.18	88%	NC_018931.2
Homo sapiens chromosome 9_alternate assembly HuRef	41.0	147	72%	0.18	84%	AC_000141.1
Homo sapiens chromosome 6_alternate assembly HuRef	41.0	332	55%	0.18	88%	AC_000138.1

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:16 PM Monday 12/1/2014

(iv)

(v)

1	Homo sapiens chromosome 3_GRCv38 Primary Assembly	39.2	291	80%	0.61	90%	NC_000003.12				
2	Homo sapiens chromosome 4_GRCv38 Primary Assembly	39.2	259	82%	0.61	84%	NC_000004.12				
3	Homo sapiens chromosome 14_GRCv38 Primary Assembly	39.2	112	47%	0.61	88%	NC_000014.9				
4	Homo sapiens chromosome 18_GRCv38 Primary Assembly	39.2	216	62%	0.61	93%	NC_000018.10				
5	Homo sapiens chromosome 5_alternate assembly CHM1_1.1	37.4	289	82%	2.1	90%	NC_018916.2				
6	Homo sapiens chromosome 17_alternate assembly CHM1_1.1	37.4	108	70%	2.1	89%	NC_018928.2				
7	Homo sapiens chromosome X_alternate assembly CHM1_1.1	37.4	72.9	35%	2.1	92%	NC_018934.2				
8	Homo sapiens chromosome X_alternate assembly HuRef	37.4	72.9	35%	2.1	92%	AC_000155.1				
9	Homo sapiens chromosome 5_alternate assembly HuRef	37.4	252	82%	2.1	90%	AC_000137.1				
10	Homo sapiens chromosome 17_alternate assembly HuRef	37.4	108	70%	2.1	89%	AC_000149.1				
11	Homo sapiens chromosome 5_GRCv38 Primary Assembly	37.4	252	82%	2.1	90%	NC_000005.10				
12	Homo sapiens chromosome 17_GRCv38 Primary Assembly	37.4	108	70%	2.1	89%	NC_000017.11				
13	Homo sapiens chromosome X_GRCv38 Primary Assembly	37.4	72.9	35%	2.1	92%	NC_000023.11				
14	Homo sapiens chromosome 8_alternate assembly CHM1_1.1	35.6	35.6	31%	7.5	95%	NC_018919.2				
15	Homo sapiens chromosome 10_alternate assembly CHM1_1.1	35.6	71.1	31%	7.5	100%	NC_018921.2				
16	Homo sapiens chromosome 13_alternate assembly CHM1_1.1	35.6	177	45%	7.5	92%	NC_018924.2				
17	Homo sapiens chromosome 19_alternate assembly CHM1_1.1	35.6	35.6	48%	7.5	82%	NC_018930.2				
18	Homo sapiens chromosome 22_alternate assembly CHM1_1.1	35.6	35.6	27%	7.5	100%	NC_018933.2				
19	Homo sapiens chromosome 8_alternate assembly HuRef	35.6	35.6	31%	7.5	95%	AC_000140.1				
20	Homo sapiens chromosome 22_alternate assembly HuRef	35.6	35.6	27%	7.5	100%	AC_000154.1				
21	Homo sapiens chromosome 19_alternate assembly HuRef	35.6	35.6	48%	7.5	82%	AC_000151.1				

(vi)

blast.ncbi.nlm.nih.gov/Blast.cgi

BLESSON ANDREWS

Homologous sequences found in the database:

	Score	Length	Identity (%)	Gaps	Strand	Accession
1	35.6	177	45%	7.5	92%	AC_000145.1
2	35.6	71.1	31%	7.5	100%	AC_000142.1
3	35.6	35.6	31%	7.5	95%	NC_000008.11
4	35.6	71.1	31%	7.5	100%	NC_000010.11
5	35.6	177	45%	7.5	92%	NC_000013.11
6	35.6	35.6	48%	7.5	82%	NC_000019.10
7	35.6	35.6	27%	7.5	100%	NC_000022.11
8	35.6	35.6	31%	7.5	95%	NT_187647.1
9	35.6	35.6	31%	7.5	95%	NT_187523.1
10	35.6	35.6	35%	7.5	92%	NW_003315953.2

Alignments

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X15, mRNA
Sequence ID: ref|XM_006716164.1| Length: 3217 Number of Matches: 1

Range 1: 417 to 440 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

Query 47 AGCAATGGAAATCCAGAAAAACAT 70
Sbjct 417 AGCAATGGAAATCCAGAAAAACAT 440

Related Information
[Gene](#) - associated gene details
[Map Viewer](#) - aligned genomic context

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:17 PM Monday 12/1/2014

(vii)

blast.ncbi.nlm.nih.gov/Blast.cgi

BLESSON ANDREWS

PRFDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X15, mRNA
Sequence ID: ref|XM_006716164.1| Length: 3217 Number of Matches: 1

Range 1: 417 to 440 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

Query 47 AGCAATGGAAATCCAGAAAAACAT 70
Sbjct 417 AGCAATGGAAATCCAGAAAAACAT 440

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X14, mRNA
Sequence ID: ref|XM_006716163.1| Length: 3251 Number of Matches: 1

Range 1: 307 to 330 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

Query 47 AGCAATGGAAATCCAGAAAAACAT 70
Sbjct 307 AGCAATGGAAATCCAGAAAAACAT 330

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X13, mRNA
Sequence ID: ref|XM_006716162.1| Length: 3401 Number of Matches: 1

Related Information
[Gene](#) - associated gene details
[Map Viewer](#) - aligned genomic context

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:17 PM Monday 12/1/2014

(viii)

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X13, mRNA
Sequence ID: ref|XM_006716162.1| Length: 3401 Number of Matches: 1

Range 1: 457 to 480 GenBank Graphics				
Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

```

Query 47 AGCAATGGGAAATCCAGAAAAACAT 70
Sbjct 457 AGCAATGGGAAATCCAGAAAAACAT 480

```

[Download](#) [GenBank](#) [Graphics](#)

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X12, mRNA
Sequence ID: ref|XM_006716161.1| Length: 3584 Number of Matches: 1

Range 1: 640 to 663 GenBank Graphics				
Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

```

Query 47 AGCAATGGGAAATCCAGAAAAACAT 70
Sbjct 640 AGCAATGGGAAATCCAGAAAAACAT 663

```

[Download](#) [GenBank](#) [Graphics](#)

PREDICTED: Homo sapiens KRIT1, ankyrin repeat containing (KRIT1), transcript variant X11, mRNA
Sequence ID: ref|XM_006716160.1| Length: 3573 Number of Matches: 1

Range 1: 629 to 652 GenBank Graphics				
Score	Expect	Identities	Gaps	Strand
44.6 bits(48)	0.014	24/24(100%)	0/24(0%)	Plus/Plus

```

Query 47 AGCAATGGGAAATCCAGAAAAACAT 70
Sbjct 629 AGCAATGGGAAATCCAGAAAAACAT 652

```

[Download](#) [GenBank](#) [Graphics](#)

Related Information

[Gene - associated gene details](#)
[Map Viewer - aligned genomic context](#)

Next ▲ Previous ▲ Descriptions

Related Information

[Gene - associated gene details](#)
[Map Viewer - aligned genomic context](#)

Next ▲ Previous ▲ Descriptions

Related Information

[Gene - associated gene details](#)
[Map Viewer - aligned genomic context](#)

Next ▲ Previous ▲ Descriptions

Desktop Libraries blesson andrews v... Computer Network Control Panel

6:18 PM Monday 12/1/2014

2.2 Yeast Protein Sequence

Seq1: Yeast Iso-1-Cytochrome C

Organism:**Saccharomyces cerevisiae** (baker's yeast) ; PDB: **1YCC_A**

TEFKAGSAKKGATLFKTRCLQCHTVEKGGPHKVGPNLHGIFGRHSGQAEGYSYTDANIKKN
VLWDENNMS

Input

The screenshot shows the NCBI BLAST suite interface. In the 'Enter Query Sequence' field, the sequence `TEPIASDEAKKXGATLPITRCLQCHTVIISPHKIVDPIILH42T8RHSQ4831STDAEIKKVLIOBIRWIS` is entered. Below it, there are fields for 'From' and 'To'. Under 'Or, upload file', there is a 'Choose File' button. A 'Job Title' field contains the text 'Enter a descriptive title for your BLAST search'. There is also an option to 'Align two or more sequences'. The 'Choose Search Set' section includes a dropdown for 'Database' set to 'Non-redundant protein sequences (nr)', an 'Organism optional' field, and an 'Exclude optional' field with 'Models (XMP)' checked. The 'Entrez Query optional' field has a link to 'Create custom database'. The 'Program Selection' section shows 'blastp (protein-protein BLAST)' selected. The 'BLAST' button is highlighted. The taskbar at the bottom shows various application icons and the date/time: 12/1/2014, 6:34 PM.

Output

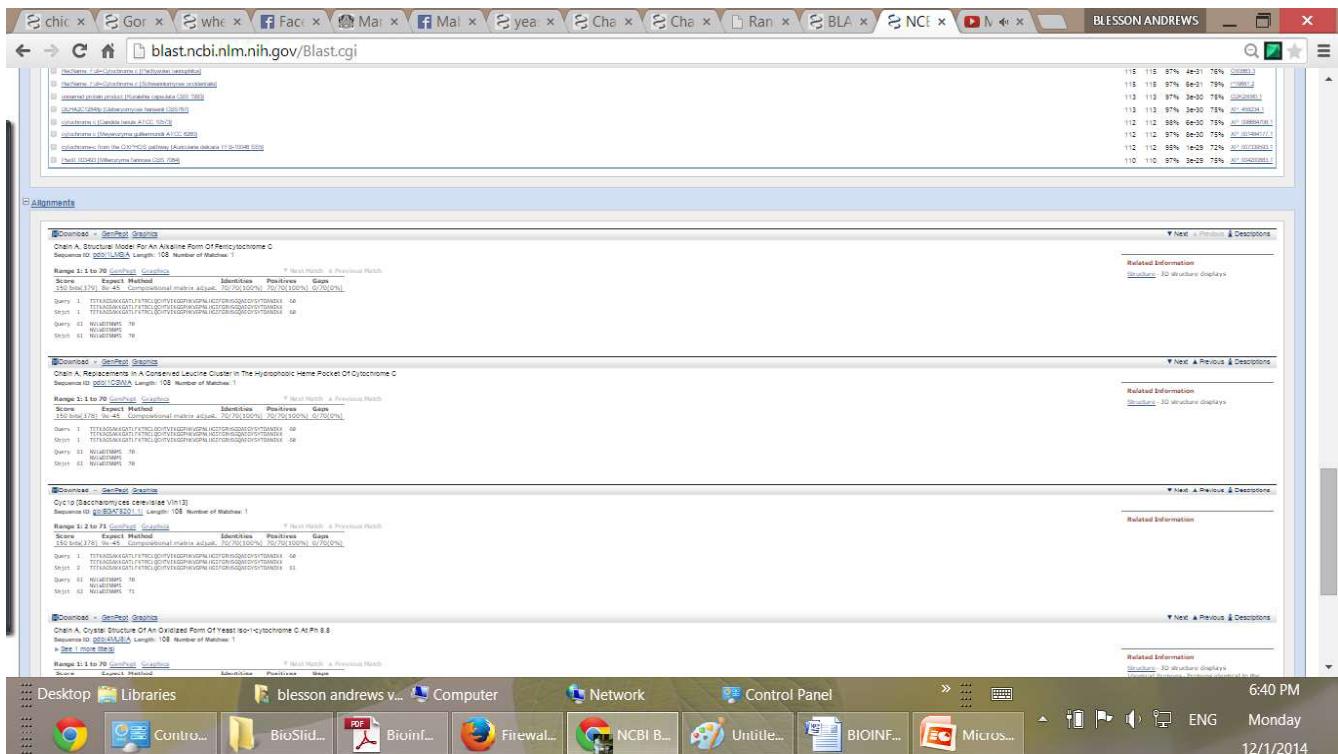
(i)

The screenshot shows the NCBI BLAST results page. The search ID is 2027190024, and the search date is 12/02/2014 at 20:09 pm. The query sequence length is 70. The database used is All non-redundant GenBank CDS translations+PDB + SwissProt + P535+PRF excluding environmental samples from WGS projects. The program used is BLASTP 2.2.30+. The results show a 'Graphic Summary' where putative conserved domains have been detected. A color key for alignment scores ranges from <40 (blue) to >2000 (red). The results window also includes sections for 'Descriptions' and 'Statistics'.

(ii)

(iii)

The screenshot shows a Microsoft Windows desktop environment. At the top, there is a taskbar with several open browser tabs, each displaying a different URL related to BLAST or NCBI services. The tabs include: chic, Gor, whe, Fac, Mar, Mai, Cha, Cha, Ran, BLA, NCE, and BLESSON ANDREWS. Below the taskbar is a vertical scroll bar. The main workspace contains a file explorer window titled 'blast.ncbi.nlm.nih.gov/Blast.cgi' showing a list of BLAST results. The results are organized into two columns: 'Query' and 'Database'. The 'Query' column lists various protein sequences from different organisms, and the 'Database' column lists the corresponding BLAST databases used for comparison. The file explorer window has a dark theme with white text. At the bottom of the screen is the Windows Start menu with categories like Desktop, Libraries, Control Panel, Network, and Control Panel.

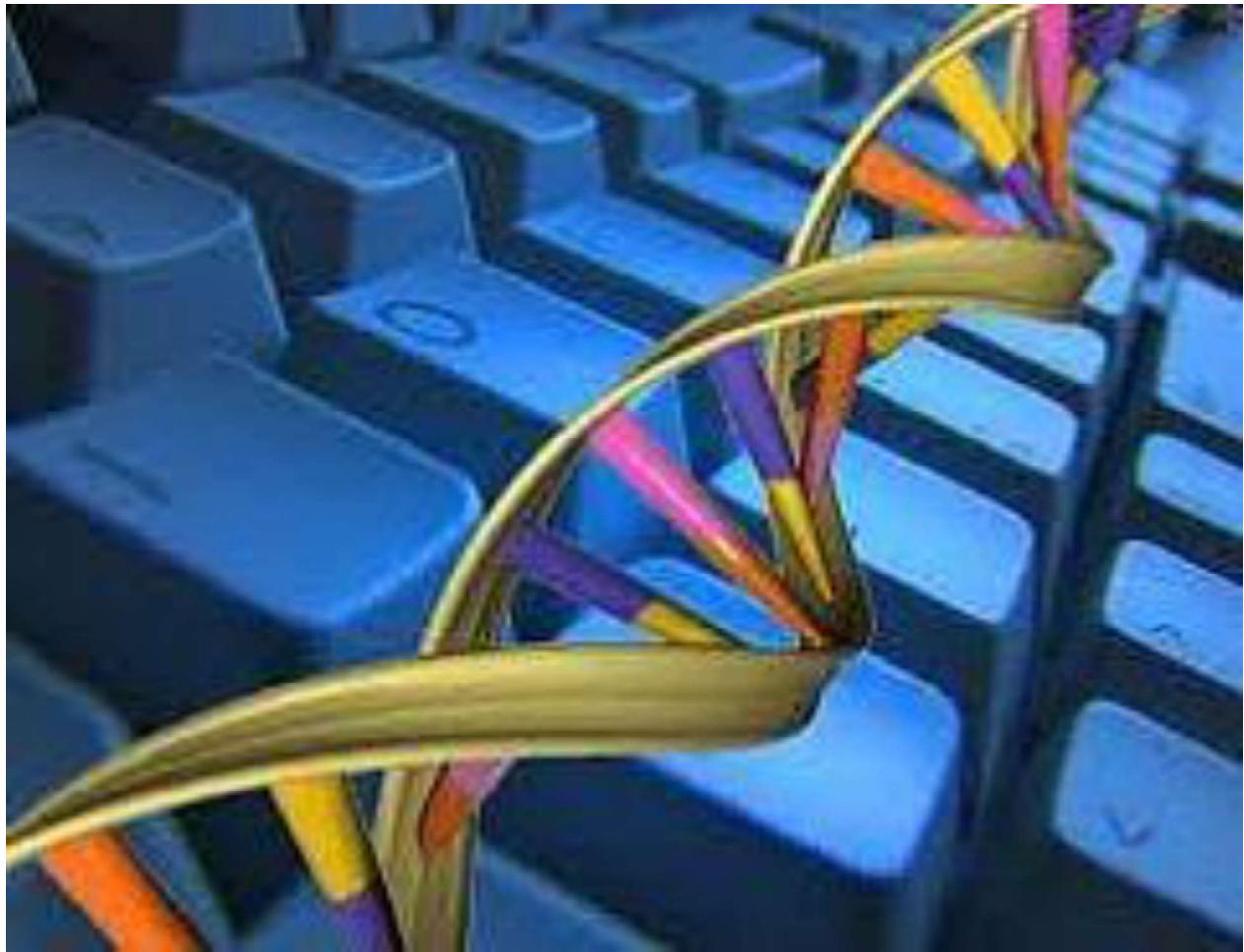


2.3 Random DNA Sequence of length 5000

Seq1:

CCCTATCAGTGGCTGCACAAACATCTCGGATCCCCTGTCCAATCAAATTGATCGAA
 TTCTTCATTTAAGACCCATAATATGACATCATTAGTGATTAAATGCCACTCCAAAATTCTG
 CCTAGAAATGTTAAGTCGCTCCACTAAAGTTAAAACGACTACTAAATCCGCGTGA
 TAGGGGATTCATATTAATCTTTATCGTAAGGAACAGCCGATCTTAATGGATGGCCGC
 AGGTGGTATGGAAGCTATAAGCGCGGGTGAGAGGGTAATTAGGCGTGTACCTACACT
 ACGCTAACGGCGATTCTATAAGATTGCACATTGCGTCTACTTATAAGATGTCTAACCG
 CATGCGCAACTTGTGAAGTGCCTACTATCCTAAACGCATATCTGCACAGTAACCCCC
 AATATGTGAGCATCTGATGTTGCCCGGGCCGAGTTAGTCTTGCTCACGGAACCTATTG
 TATGAGTAGTGATTGAAAGAGTTGTCAGTTAGCTCGTCAGGTAATGGTCCCTCACACT
 ACGTCAAAATAAGAGAGCGGTCGTGACATTATCCGTATTTCCTACTACTATCAGTACT
 CACGACTCGATTCTGCCGCAGCCACGTATGCCAGAAAGCCAGTCAGCATTAGGAGTG
 CTCTGGGCAGGACAACTCGCATAGTGAGAGTTACATGTTGTTGGGCTCTCCGACACG
 AACCTCAGTTGGCCTACATCCTACCTGAGGTCTGTGCCCGGGTGGTGAGAAGTGCAC
 TTCGTTCTTGCAGCTCGTCAGTACTTCAGAATCATGGCCTGCACGGTAGAATGACGCTT
 ATAATGGACTTCGACATGGCAATAACCCCCCGTTTCTACCTCAAGAGGGAGAAAAGTATT
 ACATGACTGCTGTCGGACAAGGGCAAAGAAGTCTCCAATTCTATTCCGAATAACA
 TCCGTCTCCCTGCAGGGAAAATCACCGACCGCATTTCATAGAAGCCTGGGGAACAGATA
 GGTCTAATTAGCTTAAGAGAGTAAATCCTGGGATCATTAGTAGTAACCACAAACCTACG
 CTGGGGCTTCTTGGCGGATTTCAGATACTAACCGAGGTGATTGAGTAAATTAGTT
 GAGGATTAGCCCGCTATCCGTAATCTCAAATTAAAACATACCGTTCCATGAGGGCT
 AGAATTACTACCGGCCTTCACCATGCCTGCCTACCGGCCACTCTCCGTTATCC
 GTCCAAGCGGATGCAATGCGATCCTCCGCTAAGATATTCTACGTGTAACGTAGCTATGT
 ATTTCAGAGCTGGCGTACCGTTGAACACTCACAGATGATAGGGATTGGTAAAGA
 GCGTGTATTGGGACTTACACAGCGTAGACTACAATGGGCCACTCAATCACAGCT
 CGAGCGCCTTGAATAACGTACTCATCTCTACATTCTCGACAATCTATCGAGCGACTCG
 ATTATCAACGGGTCTTGCAGTTCAATCTCTGCCAGCATCGTAATAGCCTCCAAGAG
 ATTGATGATAGTCATGGGCACAGAGCTGAGACGGCGCCGATGGATAGCGGACTTCGGT
 CAACCCACAATTCCCCACGAGACAGGTCTGCCGTGCGCATCACTCTGAATGTACAAGCA
 ACCCAAGAGGGCTGAGCCTGGACTCAGCTGGTTCTGGGTGAGCTGAGACTCGGGGT
 GACAGCTCTTCATACATAGAGCGGGCGTCGAACGGTCGTAAAGTCATAGTACCCCG
 GGTACCAACTTACTGAGGATATTGCTGAAGCTGTACCGTTAGGGGGGAAACGCTGA
 AGATCTCTTCTCATGACTGAACCTCGCAGGGCGTGTATGTCGGTCTCAAAGGTT
 AAAGAACAAAGGCTTACTGTGCGCAGAGGAACGCCATTAGCGGCTGGCGTCTTGAAT
 CCTCGGCCCCCTGTCTTCCAGATTAATCCATTCCCTCATTACGAGCTTACCAAGTC
 AACATTGGTATATGAATGCGACCTTGAAAGAGGCCGCTAAAAATGGCAGTGGTTATGCT
 CTAAACTCCATTGGTTAACTCGTGTATCACCGCGATAGGCTGATAGAGGTTAATATTGT
 ATAGCAAGGTACTCCGGTCTCAATGAATGGCCGGAAAGGTACGCGCGCGTATGG
 AGGGTCAAGGGGCCAATAGAGAGGGCTCCTCTCACTCGCTAGGAGGCAATTGTATAAC
 AATGCTTACTGCATCGATACATAAAACGTGTCCATCGGTTGCCAAACTGTGAAGTGTCT
 ATCACCCCTAGGCCGTTCCCGCATATAAACGCCAGGTTGTATCCGCATTGATGCTAC
 CGTGGATGAGTCAGCGTCGAGCACGCCACTTATTGCATGAGTAGGGTTGACTAAGAG
 CCGTTAGATGCCTCGCTGACTAATAGTTGTCGACAGATCGTCAAGATTAGAAAACGGTA
 CCAGCATTTCGGAGGTTCTCTAAACTAGTATGGATAGCCGTGTCTTCACTGTGCTGCC
 TACCCATCGCCTGAAATCCAGTTGGTGTCAAGCCATCCCCTGTCCAGGACGCCGATGT
 AGTAAACATACAGCTGCTCGGGTTCACCCCGGTCCGTTCTGAGTCGACCAAGGACAC
 AATCGAGCTCCGATCCGTACTGTCGAGAAACTTGTATCCGACCCCCGCCAGCTGCCAGC

BIOINFORMATICS ASSIGNMENT



TEAM MEMBERS

BLESSON ANDREWS VARGHESE - B110087CS	: Chapter 1 , 2
NIRANJAN G S - B110773CS	: Chapter 3
RONY JOSHY - B110419CS	: Chapter 4

CONTENTS

1. CLUSTAL.....	3
1.1 DNA Vs DNA	4
1.2 Protein Vs Protein.....	6
1.3 DNA Vs Random DNA.....	9
1.4 Random DNA Vs Random DNA	11
1.5 Protein Vs Home Address.....	37
1.6 Random DNA Vs Home Address.....	41
2. BLAST	43
2.1 Nucleotide and protein BLASTs	44
2.2 Human DNA sequence	50
2.3 random DNA sequence of length 5000.....	54
2.4 random protein sequence of length 10000	58
2.5 Concatenated string of 10 names.....	61
3. ORF Finder.....	63
3.1 DNA Sequence of 3 Organisms.....	64
3.2 random DNA sequence of length 100.....	70
3.3 random DNA sequence of length 500.....	72
3.4 random DNA sequence of length 1000.....	74
4. GenScan.....	76
4.1 human DNA sequence of length 10000	77
4.2 random DNA sequence of length 10000	83

1. CLUSTAL Multiple Sequence Alignment

1.1 DNA Vs DNA

Seq1 : Panthera tigris

Organism: Tiger ; GenBank: KC879292.1

TCTATAATCACCAATAACACCAATAAGCAAAGACCAACCAGTGACAACC
ACTAGCCAGGTTCCATAACTAT

Seq2 : Psittacus erythracus

Organism: African Gray Parrot ; GenBank: AE014296.5

GGACTTTGAGGTGGAAAACGAGGAGCGAGAGCAATGCAGGTGGCATTCGAAG
GGGACTATGTGGG

Seq3 : Papilio xuthus

Organism: Swallowtail Butterfly ; GenBank: DI249007.1

TTCGCAGATTTTCAGTGTCTTGTATTACAAACACAAAATGAA
ATACGTACAATCATCCTGTTCGTT

Input

The screenshot shows the ClustalW2 input interface. The user has pasted a DNA sequence into the input field:

```
>seq2
GGACTTGGTGGG-GAAAAACGAGGAGCG-AGAGCAATGGTGCGA-TCATT 51
>seq3
TTCCGAGATTTCAGTGTCTTAACTAACAAACCAAAATGAAATACGTCAAATCTC 60
>seq1
---TCTATATAATGAACTAC-----AGCAATAAGCAAAATGAAATGAAATGAAATCT 51
```

Below the sequence, there are several configuration steps:

- STEP 1 - Enter your input sequences**: A dropdown menu shows "DNA". The sequence is pasted into the text area.
- STEP 2 - Set your Pairwise Alignment Options**: Includes "Alignment Type" (radio buttons for "Slow" or "Fast"), "More options..." (link), and "Default settings" (link).
- STEP 3 - Set your Multiple Sequence Alignment Options**: Includes "More options..." (link) and "Default settings" (link).
- STEP 4 - Submit your job**: Includes "Be notified by email" (checkbox checked, value "blessonandrews0@gmail.com"), "TITLE" (text input "driv vs driv"), and "Submit" button.

The taskbar at the bottom shows various open windows and the date/time: 12/1/2014, 1:08 PM.

Output

The screenshot shows the ClustalW2 output page for the same job ID. It displays the aligned sequences:

```
CLUSTAL 2.1 multiple sequence alignment

seq2      ----GGACTTGGTGGG-GAAAAACGAGGAGCG-AGAGCAATGGTGCGA-TCATT 51
seq3      TTCCGAGATTTCAGTGTCTTAACTAACAAACCAAAATGAAATACGTCAAATCTC 60
seq1      ---TCTATATAATGAACTAC-----AGCAATAAGCAAAATGAAATGAAATGAAATCT 51

seq2      GCGGAAGGGGATCTATGCGG- 70
seq3      CTG-----TCGTTT---- 70
seq1      A-GCCAGTTCCATAACAT 70
          *
```

Below the sequences, a note says: "PLEASE NOTE: Showing colors on large alignments is slow."

The taskbar at the bottom shows the same date/time: 12/1/2014, 1:10 PM.

1.2 Protien VS Protien

Seq1 : Serine protease inhibitor 7

Organism: Solanum tuberosum (potato) ; UniProtKB/Swiss-Prot: P30941.2
MKCLFLLCLCLVPIVFSSTFTSKNPINLPSDATPVLDVAGKELDSRLSYRIISTFWGALGGD
VYLGKSP

Seq2 : Chicken Liver Dihydrofolate Reductase

Organism: Gallus gallus (chicken) ; PDB: 1DR3_A
VRSLNSIVAVCQNMIGKDGNLPWPPLRNEYKYFQRMTSTSHVEGKQNAVIMGKKTWFSIP
EKNRPLKDR

Seq3 : Buffalo (Bubalus Bubalis) Hemoglobin At 2

Organism: Buffalo (Bubalus Bubalis) ; PDB: 3CY5_D
MLTAEEKAAVTAFWGKVHVDEVGGEALGRLLVVPWTQRFFESFGDLSTADAVMNNPKVK
AHGKKVLDSF

BLOSUM

Input

The screenshot shows the ClustalW2 input interface on a Windows desktop. The window title is "ClustalW2 < Multiple". The interface includes several tabs: "Input form", "Web services", and "Help & Documentation". Under "Input form", the "Multiple Sequence Alignment" tab is selected. The main area displays two protein sequences:

```

>seq1
VRSLNSIVAVCQNMIGHIGGNLPWPPLRNEYKKYFQRTSTSHVEKGKQNAVMGKKTWFSEPKNRPLKDR
>seq2
MLTAEIKAATTAFWGKVMDEVGEALGRLLVYPWTQRFESFGDLSTADAVINNNPKVIAHGKKVLDSE

```

Below the sequences, there are three steps for configuration:

- STEP 1 - Enter your input sequences:** A text area for pasting sequences.
- STEP 2 - Set your Pairwise Alignment Options:** Includes "Alignment Type" (radio buttons for "Slow" or "Fast"), "GAP OPEN" (set to 10), "GAP EXTENSION" (set to 0.20), "GAP DISTANCES" (set to 5), and "NO END GAPS" (checkbox checked).
- STEP 3 - Set your Multiple Sequence Alignment Options:** Includes "Protein Weight Matrix" (set to "BLOSUM"), "ITERATION" (set to "none"), "NUMBER" (set to 1), and "CLUSTERING" (set to "NU").

The "Output Options" section shows "FORMAT" set to "Clustal w numbers" and "ORDER" set to "aligned".

STEP 4 - Submit your job: A checkbox for "Be notified by email" is checked, and the email address "takemotu94@gmail.com" is entered. The status bar at the bottom right shows the date and time: 12/1/2014, 1:28 PM.

Output

The screenshot shows the ClustalW2 results interface on the same Windows desktop. The window title is "Alignments < ClustalW2". The main content area displays the aligned sequences from the previous input screen:

```

seq1  ---VRSLNSIVAVCQNMIGHIGGNLPWPPLRNEYKKYFQRTSTSHVEKGKQNAVMGKKTWFSEPKNRPLKDR
seq2  MLTAEIKAATTAFWGKVMDEVGEALGRLLVYPWTQRFESFGDLSTADAVINNNPKVIAHGKKVLDSE

```

The interface includes tabs for "Alignments", "Result Summary", "Guide Tree", "Phylogenetic Tree", and "Submission Details". Below the sequences, a note says "PLEASE NOTE: Shading colors on large alignments is slow." At the bottom, there are links for "EMBL-EBI" (News, Bioresources, Contact us, Intranet) and "Services" (By topic, By name (A-Z), Help & Support). The status bar at the bottom right shows the date and time: 12/1/2014, 1:56 PM.

PAM

Input

The screenshot shows the ClustalW2 web interface. In the 'Input form' section, there is a text area containing four protein sequences:

```

>29821
MKCFPLCLCLVPMVFSTFTSKNPINLPSDATPVLVDAGKELDSRLSYRISTPVGALGGDVVLGKSP
>29822
VRSLNSAVACQNMGIGIQLNLVPLNVEKKYFQRMSTSHWE6KQNAVMIGKTKWFSPIKEKNRPLKDR
>29823
MLTAEEKAAVTAFWGKVHDEVGEALGRLLVYPINTQRIFESFGDLSADAVMINNPKVKAHKKVLDSD

```

Below the sequences, there are buttons for 'Upload a file' and 'Choose File'. The 'Output Options' section includes dropdown menus for 'FORMAT' (set to 'Clustal w numbers') and 'ORDER' (set to 'aligned'). At the bottom, there is a 'Submit' button.

Output

The screenshot shows the ClustalW2 web interface displaying the results of a multiple sequence alignment. The 'Alignments' tab is selected, showing the alignment details for job clustalw2-120141201-101013-0287-36681443-oy. The alignment output is shown as a text block:

```

CLUSTAL 2.3 multiple sequence alignment
seq1 VRSLSIVIAVCFISGSGQILPPPLRIEVVYFQQTHTSPWEGIQUWZHSVTFI
seq2 NLTAEFLCLCPLPQVYVSEELDNLVLYVPTQDFPPEFADLTLQDNTVPRV
seq3 -HICUFLCLCPLPQVYVSEELDNLVLYVPTQDFPPEFADLTLQDNTVPRV
seq4 PKEKUPLD-> 70
seq5 AHGSVVLQDF-> 70
seq6 LAGSVVLLDF-> 70

```

A note at the bottom says 'PLEASE NOTE: Showing colors on large alignments is slow.' The interface includes a navigation bar with links like 'Services', 'Research', 'Training', 'Industry', and 'About us'. The status bar at the bottom shows the date and time as 12/1/2014 3:40 PM.

1.3 DNA Vs Random DNA

Seq1: Gorilla gorilla isolate Snowflake 5' upstream regulatory sequence

Organism: Gorilla gorilla (Western Gorilla) GenBank: AY738219.1

TGTGTGTGTTGTTCCCTGATATTCTATGTATCATGATTGACCTAGATAAAAGGTTGTA
TTTTGGC

Random DNA Sequence

TTCATTCTCTGCCAGCACTGTAATAGGCAC TAAAAGAGTGATGATAATCATGAGTGCCG
CGCTAAGGTGGTGT CGGAAC

(Source: <http://www.faculty.ucr.edu/~mmaduro/random.htm>)

Input

chicken - Pi x Gorilla gor x Gorilla gor x Facebook x random DN x Random DI x History x ClustalW2 x BLESSON ANDREWS

www.ebi.ac.uk/Tools/msa/clustalw2/

ClustalW2

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Multiple Sequence Alignment > ClustalW2

Multiple Sequence Alignment

ClustalW2 is a general purpose DNA or protein multiple sequence alignment program for **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment](#) tools.

Note: ClustalW2 is no longer being maintained. Please consider using the new version instead: Clustal Omega

STEP 1 - Enter your input sequences

Enter or paste a set of Protein sequences in any supported format:

```
TGTGTTGTTGGTCTTGATTTCTATGATCATGATTGACCCTAGATAAAAAGGTGTATTTTGGC>3852TTCACTCTCGGAGGACTGTAATAGGCACTAAAGAGTGATGATAATCATGAGTGCCTGGCTAAGGTGGTGGAAAC
```

Or, upload a file | Choose File | No file chosen

STEP 2 - Set your Pairwise Alignment Options

Alignment Type: Slow Fast
The default settings will fulfill the needs of most users and, for that reason, are not visible.
[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 3 - Set your Multiple Sequence Alignment Options

The default settings will fulfill the needs of most users and, for that reason, are not visible.
[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 4 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

If you plan to use these services during a course please [contact us](#).

Support staff, BioINF...

Output

The screenshot shows a web browser window with the URL www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2-l20141201-102431-0751-11675971-oy. The page title is "ClustalW2". The main content displays a multiple sequence alignment between two sequences, seq1 and seq2, with 54 and 60 characters respectively. The alignment shows high conservation across most positions, with some variations indicated by asterisks (*). A note at the bottom states: "PLEASE NOTE: Showing colors on large alignments is slow."

	seq1	seq2
seq1	TGTGTGTGTTTGTTT-CCTTGATATT--TCTAT -GTATCATGTTGACCTAGATAAA	54
seq2	TTTATCTCTCGCCAGACTTATAGGCACCTAAAGAGTGTGATGATATACTATGAGTCGCG	60
seq1	AG-GTTTATTTTTGG-c	70
seq2	CGCTAAGGTGTTGCGAAC	80

1.4 Random DNA Vs Random DNA

Sequence Length: 10

Seq1: AAAGCGGTCT

Seq2: TACGGTCAGT

Seq3: CGTATTCCCTT

Input

The screenshot shows the ClustalW2 web interface for multiple sequence alignment. The sequences entered are:

```

>Seq1
AAAGCGGTCT
>Seq2
--TACG--GTCAGT
>Seq3
---CGTATTCCTT
  **  **
  
```

Output

The screenshot shows the ClustalW2 output page. The alignment results are displayed as follows:

```

CLUSTAL 2.1 multiple sequence alignment

Seq1      AAAGCG--GTCT-- 10
Seq2      --TACG--GTCAGT 10
Seq3      ---CGTATTCCTT 10
  **  **
  
```

PLEASE NOTE: Showing colors on large alignments is slow.

Sequence Length: 100

Seq1:

CTCGAGTTCCGTCCAGTTGAGCGTGTCACTCCCAGTGTACCTGCAAGCCGAGATGGCTG
TGCTTGGAGTCAATCGCATGTAGGATGGTCTCCAGACACCG

Seq2:

GGGCACCAGTTTCACGCCCTAAAGCATAAACGACGAGCAGTCATGAAAGTCTTAGTACTG
GACGTGCCGTTCACTGCGAATAATACTGGAGCTGTACC

Seq3:

GTTATTGCGCTGCATAGATGCAGTGCTGCTTATCACATTGTTCGACGACAGCCGCC
TTCGCAGTTCCCTCAGACACTTAAGAATAAGCGCTTATTG

Input

ClustalW2

[Input form](#) | [Web services](#) | [Help & Documentation](#)

Tools > Multiple Sequence Alignment > ClustalW2

Multiple Sequence Alignment

ClustalW2 is a general purpose DNA or protein multiple sequence alignment program for **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Note: ClustalW2 is no longer being maintained. Please consider using the new version instead: [Clustal Omega](#)

STEP 1 - Enter your input sequences

Enter or paste a set of Protein sequences in any supported format.

```
>Seq1
CTGAGTCCGTCAAGTGGACGCTGACTCCAGTGACCTGCAAAGCGAGATGGCTGTGCTTGGACTAATGCATGAGATGGCTCCAGACACCG
>Seq2
GGGACACAGTTTCAACGCCAACATAACGACGAGCAGTGAAGCTTAACTGAGCTGCGTCCACTGCGAATAACCTGGAGCTGTAC
>Seq3
GTTATTGCGCTGCATAGATGCAAGTGTGCTCTTACATTTGCGACGACAGCCGCTTCGAGTCTTCAGACACTTAAGATAAGCGCTTATTG
```

Or, upload a file: [Choose File](#) | No file chosen

STEP 2 - Set your Pairwise Alignment Options

Alignment Type: Slow Fast

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 3 - Set your Multiple Sequence Alignment Options

The default settings will fulfill the needs of most users and, for that reason, are not visible.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 4 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

Output

chicken - P Gorilla gorilla x Gorilla gorilla x Facebook x random DN x Random DN x History x Alignment x LESSON ANDREWS

www.ebi.ac.uk/Tools/services/web/toolresult.cgi?jobId=clustalw2-l20141201-103610-0614-41043774-oy

EMBL-EBI Services Research Training About us

ClustalW2

Input form Web services Help & Documentation Share Feedback

Tools > Multiple Sequence Alignment > ClustalW2

Results for job clustalw2-l20141201-103610-0614-41043774-oy

Alignments Result Summary Guide Tree Phylogenetic Tree Submission Details

Download Alignment File Show Colors Send to ClustalW2_Phylogeny

CLUSTAL 2.1 multiple sequence alignment

	Seq2	Seq3	Seq1							
GGG	CACCA	TTT	CGCTAAAGCAT	AACGAGCAG	TCA	TGAAGCTT	AGT-AC	58		
-----	-----	-GT	ATTGCGCT	-----	GCT	ATGCACTCT	TATCACATTTGTT	TC	57	
---	CTG	BAG	TTCCGTCG	CAAGT	GAGT	CTCCCA	GTGACCTGCAA	GGAGATGG	56	
***	***	*	***	***	*	***	***	***	*	
Seq2	TG-GACG-----	TGCC	TTT-CAGT	-CGA	TAA	TACCTG	GAAGCT	TACC	100	
Seq3	GACG	ACGCCG	CTTGGC	CA	GACT	TAAG	AA	AGCGCTT	ATTG	100
Seq1	TGT	GCTT	-----	-GAGT	CA	TG	TAG	GA	TGCTTCCAGAGACCGG	---

PLEASE NOTE: Showing colors on large alignments is slow.

Sequence Length: 1000

Seq1:

TAGGCAGAGGCACGCCCTATTAGTGGCTGCGGCAAAATATCTCGGATCCCTTGCCA
 ACCAAATTGATCGAATTCTTCATTAAAGACCTAACATGTCTCATTAGTGATTAATGCC
 ACTCCGAAAATACCGCCTAGAAATGTCTAACAGATCGGCCACTAAAGTTAAAACGAC
 TGCTAAATCCCGGTGATAGGGGATTGAAGTTAACATCTCTATCGCAAGGAACGCCGAT
 CTTAATGGATGGCGGAGGTGGTATGGAAGCTATAAGCGCGGGTGAGAGGGTAATTAG
 GCGTGTTCACCTACGCTAACGGCGATTCTATAAGATTGCACATTGCGTCAACTC
 ATAAGATGTCTAACGGCATGCGCAACTGTGAAGTGTCTACTATCCTAAACGCATATC
 TCGCACAGTAACCTCCGAATATGTCGGCATCTGATGTTGCCCGGCCAGTTAGTGTG
 AGCTCACGGAACCTATTGTATGAGTAGTGATTGTAAGAGTTGTCAGTTAGCTCGTTAG
 GTAATAGTTGCCACACAACGTCAAATAAGAGAACGTCGTAACATTATCCGTGATTT
 CTCACTACTATCAGTACTCACGACTCGACTCTGCCGCAGCACGTATGCCGTGAAAGCC
 AGTCAGCGTTAAGGAGTGCTCTGAGCAGGACAACTCGCGTAGTGAGAGTTACATGTTG
 TTGGGCTTCCGACACGGACCTGAGTTGGCCAACGTCACCTGAGGTCTGTGCCCG
 GGTGATGAGAAGTGTGCATCTCGTTGCAGCTCGTAGTACTTCAGAATCATGGCGT
 GCATGGTAGAATGACCTATAACGGACTCGACATGGCAATAACCCCCCGTTCTACTT
 CTAGAGGAGAAAAGTATTGACATGAGCGCTCCCGACAAGGGCAAAGAAGTCTCAA
 TTTCTTATTCCGAATGACATGCGTCTCCTGCGGGTAAATCACCAGA

Seq2:

CCGCAATTCATAGAAGCCTGGGGAACAGATAGGTCTAATTAGCTTAAGAGAGTAAATCC
 TGGGATCATTCACTAGTAGTAACCATAAACTTACGCTGGGCTTCTCGGCCGATTTTACAG
 TTACCAACCAGGAGATTGAAGTAAATCAGTTGAGGATTAGCCGCGCTATCCGTAATC
 TCCAAATTAAACATACCGTCCATGAAGGCTAGAATTACTTACCGGCCCTTCCATGCC
 GCGCTATACCCCCCCCACCTCTCCGCTTATCCGTCCGAGCGGGAGGCAGTGCGATCCTCC
 GTTAAGATATTCTTACGTGTACGTAGCTATGTATTTGCAGAGCTGGGAACGCGTTGA
 ACACCTCACAGATGGTAGGGATTCGGGTAAAGGGCGTATAATTGGGACTAACATAGGC
 GTAGACTACGATGGCGCCAACCTCAATCGCAGCTCGAGCGCCCTGAATAACGTACTCATC
 TCAACTCATTCTCGGCAATCTACCGAGCGACTCGATTATCACCGCTGTCTAGCAGTTCT
 AATCTTTGCCAGCATCGTAATAGCCTCCAAGAGATTGATGATAGCTATCGGCACAGAAC
 TGAGACGGCGCCGATGGATAGCGGACTTCGGTCAACCACAATTCCCCACGGGACAGG
 TCCTGCCGTGCCCATCACTCTGAATGTACAAGCAACCAAGTGGGCCAGCCTGGACTC
 AGCTGGTCTCGGTGAGACTCGGAGACTCGGGATGACAGCTCTTAAACATAGAGCGGGG
 GCGTCGAACGGTCGAGAAAGTCATAGTACCTCGGGTACCAACTACTCAGGTTATTGCT
 GAAGCTGTACTATTTAGGGGGAGCGCTGAAGGTCTCTTCTCATGACTGAACCTC
 GCGAGGGTCGTGAAGTCGGTCTCAATGGTAAAAAACAAAGGCTACTGTGCGCAG
 AGGAACGCCATCTAGCGGCTGGCGTCTGAATGCTCGGTCCCTTGT

Seq3:

CATTCCGGATTAATCCATTCCCTCATTACGAGCTGCGAAGTCTACATTGGTATATGAA
 TGCGACCTAGAAGAGGGCGCTTAAATTGGCAGTGGTGTATGCTCTAAACTCCATTGGT
 TTACTCGTGCATCACCGCGATAGGCTGACAAAGGTTAACATTGAATAGCAAGGCACCTC
 CGGTCTCAATGAACGGCCGGAAAGGTACGCGCGGTATGGGAGGATCAAGGGGCC
 AATAGAGAGGGCTCCTCTCACTCGCTAGGAGGCAAATGTAAAACAATGGTTACTGCATC
 GATACATAAAACATGTCCATCGGTTGCCAAAGTGTAAAGTGCTATCACCCCTAGGGCC
 GTTCCCGCATATAACGCCAGGTTGTATCCGCATTGATGCTACCGTGGATGAGTCTGC

GTCGAGCGCGCCGCATTATTGCATGAGTAGGGTTGACTAAGAGCCGTTAGATGCGTCG
CTGTACTAATAGTTGTCGACAGACCCTCGAGATTAGAAAATGGTACCAAGCATTTCGGAG
GTTCTCTAACTAGTATGGATTGCGGTGCTTCACTGTGCTGCGGCTACCCATGCCGTGAA
ATCCAGCTGGTGTCAAGCCATCCCCCTCTCCGGGACGCCGCATGTAGTGAAACATATAACG
TTGCACGGGTTACCGCGGTCCGTTCTGAGTCGACCAAGGACACAATCGAGCTCCGATC
CGTACCCCTCGACAAACTTGTACCCGACCCCCGGAGCTGCCAGCTCCTCGGGTATCATG
GAGCCTGTGGTTCATCGCGTCCGATATCAAACCTCGTCATGATAAAAGTCCCCCCCCTCGG
GAGTACCAAGAGAAGATGACTACTGAGTTGTGCGATCCCTGCACTTCAGCTAAGGAAGCT
ACCAATATTAGTTCCGAGTCTCACGTCTGACCTCGCGGGTAGATTGCCATGCGTAGAG
CTTACGAGCCAGCGGAAAGTGTGAGGCGCTTTAAGTATGGCGAGTAA

Input

The screenshot shows the ClustalW2 web interface. In the main area, there are two sequence inputs:

```

Seq1: TAGGCAGAGGCGACGGCTTATTAGTGGCTGGGAAATAATCTTCGGATCCCTTGTGGAA
Seq2: -----

```

Below the sequences, there are four steps for alignment options:

- STEP 1: Enter your input sequences**: Shows the pasted sequences.
- STEP 2: Set your pairwise alignment options**: Contains "Alignments" and "Show Options".
- STEP 3: Set your multiple sequence alignment options**: Contains "Alignments" and "Show Options".
- STEP 4: Submit your job**: Includes a checkbox for notifications and a "Submit" button.

The status bar at the bottom shows the date and time: 12/1/2014 4:07 PM.

Output

(i)

The screenshot shows the ClustalW2 results page for job ID clustalw2-l20141201-104017-0891-71632106-oy. The results section includes:

- Alignments**: Shows the aligned sequences for Seq1, Seq2, and Seq3.
- Result Summary**: Provides a summary of the alignment.
- Guide Tree**: Displays a phylogenetic tree.
- Phylogenetic Tree**: Shows the tree structure.
- Submission Details**: Includes download links for the alignment file and phylogeny.

The aligned sequences are:

```

Seq1: TAGGCAGAGGCGACGGCTTATTAGTGGCTGGGAAATAATCTTCGGATCCCTTGTGGAA 68
Seq3: -----
Seq2: -----

```

... (continues for several more lines of sequence alignment).

The status bar at the bottom shows the date and time: 12/1/2014 4:09 PM.

(ii)

chicken - Pi Gorilla gori Gorilla gori Facebook random DN Random DI History Alignments BLESSON ANDREWS

Seq1 GCTGTTTCACTACGCTAACCTAACGGCGGATTTCTATAAGATTGCACTGCGTCACA 359
Seq3 AG4GCGCTCTCTCACTCGCTAGGGCGAAATGTAAAACATGGTTACTGCATGATC 363
Seq2 CTGGCTATACCCCCCGACTCTCCGCTTATCGTCGCGAGCGAGCTGGATCTCC 299

Seq1 ATAAAGATGTCGACGGCATGG-----CAAGATG-----TGAAAGTGTCACTACCTCCAAAC 412
Seq3 ATAAAACATGTCGATCGTTGCC-----CAAAAGTG-----TTAAAGTGTCTATCACCCCTAGGG 356
Seq2 GTTAAGATATTCACGTGTGAGCTAGTGTATTTGCAAGCTCGCAAGCGCTGTGA 359

Seq1 GCTATCTGCGACAGTAACCTCCGGAATATGCGCGACTGATGCG-----TTGCGCGGGCG 467
Seq3 CGGTTTCCCACATAAACCGGGTTGATCGCGATTG-----CTACCGTGGATG 411
Seq2 ACACITCCACAGATGGTAGGGATTGCGGTAAGGGCGTATAATTGGGGACTAACATAGCG 419

Seq1 -AGTTAGTGTGAGCTCGGACACTTATTTGATGTGAGTGTGATTTGTAAGAGTTGTCG 526
Seq3 -AGTCCTGGCTGGCGCGCCGCGATTATTGATGTGAGGTGACTAAAGGCCCTTGA 479
Seq2 TAGACATCGATG-----CTGCAACTCACTCGACG-----CTCGAGCCGSCCTGATAAACG-TACT 473

Seq1 TAGCTGTTGAGTGTCTCACTACTACGACTGAGCTCGAGCTCGCGCAACAGTAC 586
Seq3 TCGCTGCTGATG-----TATGTTGTCGAGACCGTCTGAGTTTAAAGTACACAGCG 530
Seq2 TATTGATGACTTATGCTGGATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTG 533

Seq1 ATCCGTTGATTCTCACTACTACGACTGAGCTCGAGCTCGCGCAACAGTAC 646
Seq3 TTGGGAGTTCTTCACTACTGTTGATTGCGGTGCTTCACGCTGCGGCTACCCATC 590
Seq2 TTCT-----AAATCTTCCGACGATCC-----TAATAGGCTCCAGAGATTGATGATGCTCATC 507

Seq1 GCTGAAAGCC-ACTCGECA-TAGGAGTGTCTGGCGAGCA-CGCTGAGCTGCGCGACACGATC 704
Seq3 GCTGAAAGCTC-ACTGCTGCTGCA-BGCGCTTCCCTCTCGGGGAGCGCCATGCTGAGA 645
Seq2 GG-LC-RG-AV-GAGL-LG-LX-LCLRLI-LG-LI-LG-LG-LC-LA-LI-LCL-LC-LA- 64/

Seq1 GTTACATGTTGTTGGCTCTTCCGACACGGAGCTGTGGGCGAACAGTCCCACCTGAGG 764
Seq3 CATATAGTTGACGGGGTTCACCGCGCTGCTCTGAGTCGACCAAGAACACAACTGAGG 708
Seq2 GGGACAGCTCTCGCG-----TGGCCTACCTCTGAGTATCAAGCGAACCCAAGTGAGG 701

Seq1 T-----CTGTCGGGGTGTAGAGAAGTGTGCACTGCTGTCAGCTGCTGAGCTTC 821
Seq3 T-----CCGATCGTACCTCTGACAAACCTGTGACCCACCCCGAGCTTGCGACCTCTG 765
Seq2 CGAGCTGGACTCAGCTGTTCTCGTGTGAGCTGAGCTGAGCTGAGCTGTTA 759

Seq1 AGAATCATG-GCTGCTATGGTAG-AATGACCCCTTATAACGGAGCTCGACATGCAAAAC 879
Seq3 GGTATTCATG-GNGCTGTTTCATCGCTGCTGATGATGATGATGATGATGATGATG 823

4:10 PM blesson andrews v... Computer Network Control Panel 12/1/2014 ENG Monday

(iii)

chicken - Pi Gorilla gori Gorilla gori Facebook random DN Random DI History Alignments BLESSON ANDREWS

Seq1 AGAATCATG-GCTGCTATGGTAG-AATGACCCCTTATAACGGAGCTCGACATGCAAAAC 879
Seq3 GGTATCATG-GNGCTGTTTCATCGCTGCTGATGATGATGATGATGATGATGATG 823
Seq2 AACATAGAGCGGGGGCGTCGAGCGTGAAGAAAGTCAATGACCTCTGAGCTTCACTTC 818

Seq1 CCCCCGGTTTCACTCTTCACTGAGGAGAAAGTATGACATGAGCTGGCTCCGG-CACAGGGC 938
Seq3 CCCCGCTGGGGGATGACAGAGAGTACTGAGCTGCTGGGGATCTCTGAGCTGG 882
Seq2 CTCAAGTTATGCTTGAAGCTGACTATTTTGGGGGGGGGGCTGAGGCTCTCTCTC 878

Seq1 CAAA-----GAAGTCTCCAAATCTTATTTCCGAATGACATGGCTCTGGGGGTTAA 993
Seq3 TAAG-----GAAGTCTCCAAATTTAGTTGGCTGGAGTCTCAGCTGCTGACTCTGGGGTTAA 937
Seq2 TCATGAGCTGAGCTGGCGAGGGTGTGAGCTGGCTCTTCAATGGTTAAAACAAAGGC 938

Seq1 TCAAGG----- 1000
Seq3 TTGGCATGCTGAGG-----CTTACGAGCCAGCGGAAATGAGGCGCTTAAATGAGG 995
Seq2 TTACTGTGCGCGAGGAGGACCGGCAATCTGGGGCTGGCTGCTGAGCTGGGGCTT 997

Seq1 -----
Seq3 AGTAA 1000
Seq2 TGT- 1000

PLEASE NOTE: Showing colors on large alignments is slow.

EMBL-EBI Services Research Training Industry About us

News Brochures Contact us Intranet By topic By name (A-Z) Help & Support Overview Publications Research groups Postdocs & PhDs Overview Train at EBI Train outside EBI Train online Contact organisers Overview Members Area Workshops SME Forum Contact Industry programme Overview Leadership Funding Background Collaboration Jobs People & groups News Events

4:10 PM blesson andrews v... Computer Network Control Panel 12/1/2014 ENG Monday

Sequence Length: 10000

Seq1:

GTGATCCAACGCTTCGGATGTGACTATATACTTAGGTCGATCTCGTCCCGAGAATTAA
 AGCCTCAGCATCTACGAGTTAGGAGCTAGCCAAAAAAGCACGTGGTGGCGCCCACCGA
 CTGTTCCCAAACGTAGCTCTCGTCCGTCAAGGCCGACTTCATCGCGGCCATTCC
 ATGCGCGGACCATAACCGTCTAATTCTCGGTTATGTTCCGATGTAGGAGTGAGCCTAC
 CTGCCTTGCCTGATACCAATGAAAAACCTATGCACCTTGTACAGGGTGCCATCGGG
 TTTCTGAACCTCTCAGATAGTGGGGATCCCAGGTAAGACCTATATCTCGGGTCAAACCTA
 GGCATAAACCTCCATGCTACCTAGTCAGACCCACCCCGCACGGGTAATATGGCACGC
 GTCCGACCTGGTCTGGCGTACGCTGCCACGTGTTCTTAACGTGTTGGTAG
 ACAAAAGTATTACCATGGTCTAGAAGTTCCGACAGTTAGTCAGGCTAATGTCACAA
 ATGACGCAGAACGCCAATGAGTGCCAGACATTAGGTGGAGTTCAAGTCAGGTAACGGAGA
 GACTCTGCGGCGTACTTAATTATGCATTGAAACGCCAAGTGACGCTAGGCAAGTC
 AGAGCAGGTTCCCCTGGTAGCTGAGGGTAAACATACAAGCCGATTGAAGATGGTAGG
 GGGCTCAAATCGTCCAGCACTCCACAGTACCTCCGAGAGCAAGTAGGGCACCCCTGTAG
 TTCGAAGCGGAACATTGAGGGCGAGCCACATCGTCTCTCGGGATGACTTAA
 CACGCTAGGGAGGTGGAGTCGATTCCATCGATGGTTATAAATCAAAAAAATCGGAACGCT
 GTCTGGAGGATGAATCTAACGGTGCCTCGATCGCTCAGTCGCTTTCTGACTGCG
 CGAAAGTCGCACCGCTCATACACTTGGTCCGAAGCCTGTCTGATATATGAATCCAAA
 CTAGAGCGGGGCTTGCAGCTTGGAGTTAAATATCTAATATTCCAATCGGCTTTAC
 GTGCACCAACCGCGGGCGTACAGGAGACTCACACCGAGAAACTAGACAGTTGCGCG
 CTGGAAGTAGGCCGGCTAAGAAAGACGCCCTGGTACAGCAGGACTATGAAACCCGTACA
 AAGGCAACATCCTCACTCGGTGAATCGAAACGCCATCAAGGTTACTTTGGATACC
 TGAAACAAAACCCATCGTAGTCCTAGACTTGGACACTTCACCCCTAGGGCCATATC
 TGGAAATAGACGCCAAGTTCAATCCGTACTCCGACGTACGATGGAACAGTGTGGATGT
 ACGAGCTCATTATACCCCTCGCGCGCCGGACCGGGTCCGAAGGCCGCGCGTGC
 ACAAGCAATTGACAACTAACCACCGTGTATTGTTATGGCACCAGGGAGTTAAGCCGAG
 TCAATGGAGCTCGCAATACAGAGTTACCGCATCTGCCCTAAGTGACAAACTGTGATCG
 ACCACAAGCCAAGCCATTGCCCTTAGACACGCCGTTACAGTGATTATGAAAACCTTGCG
 GGGCATGGCTACGACTTGTTCAGCCACGTCCGAGGGCAGAAACCTATCCCCATTGTAT
 GTTCAGCTATCTTACCCATCCCCGGAGGTTAAGTAGGTTGTGAGATGCGGAAGAGGC
 TCTCGATCATCCGTGGGACATCAACCTTCCCTGATAAAGCACCCCGCTGGGTATG
 GCAGAGAGAACGCCCTCTGAATTGTGCTATCCTCGACCTATCAAAGCTTGTACCAAT
 AATTAGGATTATTGCCCTGCGACAGACCTCCTACTCACACTGCCCTCACATTGAGCTAGTC
 AGTGAGCGATTAGCTTGACCCGCTCTAGGGCGAGTACGTGAGCTAGGGCTCCG
 GACTGGCTATATAGTCGAGTCTGATCTCGCCCCGACAACCTGCAAACCTCAACTTTTA
 GATAACATGGTAGCCGAAGTTGCACGAGGTGCCGACCGCGACTGCTCCCCGGGTGT
 GGCTCCTCATCTGACAACGTGCAACCGCTATGCCATCGATTGAGTCAGCGGACGGTG
 TTGTCGTAGTTGGCATGTTCCCTGAGGTGTGAAATCACTTAGCTCGCGCCG
 TAGTCTAAAGGAAAACCTATGAGACTATGTTGGGTAGCACCAGGAATCTGAAGCACGT
 GTATGTGGACGTGGCGTACACCTTAATCTCGCTCATGCTAAGGATCTGGCTGC
 ATGCTATGTTGATAACACCTACACTGCTCGAAGAAAATATCGAAGCGGGCGGCTGGCC
 GGAGCCCTACCGCATCGACACGTGTTACTGTTAATTGGTGGCACATAAGCAATAT
 TGTAGTCGTCAATTGAGCCCTTATCCTCGCGTTATGTTGTCAGGCTAGATCT
 GGATTGACTGTATGACGGTACCTGCTGATCGGTAGGGAGACCGAGAATCTATCGGGCTA
 TGTCACTAAAACCTTCCAAACGCCCGTGTGACTGAACTGAGACATGCGTACGCCCTTAATACATC
 TCCTGAAAACGCATAAACATACAAGTGGCAGATGATGCGTACGCCCTTAATACATC
 CAACACTCTACGCCCTTCAAGAGCTGGAAGGGCACCCCTGCACTGGATAGGGATTAA

TCTCGTAAGGCAAGCTCGTACCGTCATCGGGAAAGAGTTAACACGATTGGAAGTAG
GGATAGTTCGAACCTCGGTTACTAGTCCTAATAAGGGAACGCTGTGAAGGGATGAGT
GTCAGCCAGTGTAAACCGATGAGCTACCCAGTAGTCGAACCTGGGCCAGACAACCCGGC
GCTAATGCACTCAATCCCGAGGCCTGACGCGACATATCAGCTAACAGAGTAGGCCGGGG
TGTAGACCTTGGGGTTGAATAAAATCTATTGTACTAACCGGTTCAACGTCGCCGACCGGG
TGGCACCTCAGGAGGGCCCACAGGGAGGAAGTAAACTGCTATTGTCGCCGATGGTG
GTAACATAATTATGTTCTGCCACTACAATTGTATCTAACGCCGTGTAATGAGAACATCCAC
ACCTTAGTGAATCGATGCGCCGCTCGGAATACCGTTGGCTACCTGTTACTAACGCCA
TCGCGATTTCAGGTATCGTGCACGTAGGGTTGCACCGCACGCATGTCGAACGGTGGC
GAAGTACGATTCCACGACC GGTTGACGATTCAACTACCGGGACGTGACGAGCTTAT
ATATGCTTCGCCCGCCGGACCGGCCTCGTATGGGGTAGCTGCGCATGAGCTTATGAC
AAATAACGAGTGTGTACTTGTAAATCATCTCACAGTGAAGTCGGGAGAACATGGAGCCG
CTACACACATTACCGCAACTACACCTAACTGAGATACTACCATAGATGACTACCCATCC
CTCTAGGCCTAGATAACCGGATACAGTGACTTGATAGGTTGTGGGGTACAGCTATGA
CTTGCATAGCTCGTATGGAGGAAGGAACCTTGCGTGTAGTATGTTGACCCGTGATT
ACGCATGCGGGTAGATCATGTGGGTAGAGACATCCAGGTCAAGTTCTGACCTTCTCGT
GGGAGGTGATCCAGTTCACTATAGGACCATTCCATTGAGCATGGCACTAAGTACGCCG
TCCCCATTCTGGTAATCATCATCCCTATCAGGGTTGGAGTGAGTGGTACGGTTATTCC
CCAGGAATGGACTTCTTATTCACAGTCGGTCACATTGGCTACTCCTGGTCTTCCGCT
TGGCCCGGTCTGTTGGCCGCCATTGCGTGAGTTCGGCCCTGTGCTGCGCTGTGCG
GCGATTCTCATTGGGGCCTCACATCTGGAAACCCCGACCTATTGACGGCACCATGG
CGGAAGTTGCTGGCTTGTGCACCGTGAAAGTCCTCCGGGGCTCCCCCTACAAAAG
ACGATAAGCTCCGGCAAGCAGCATTGATCAACGCAAGGATCGGCGATAAACAGAGAT
ACGGCTGATTACTCTTGTGGTGGTATCGCTAAACTGCGCCGGAGCCTATGGCA
AAGTCGTCGCGAACCATCCGGTAGCGCTTAAGGTCCATAGCACATTATCGCATCCC
GGCGTGCCTCAATTGACGACCCCTGGCGAAAGGTGCTGGCACCGTGCTAAATTAA
AGCGGCTGCACTGCTGAAGTCCGTTACGGAGAGGGCGGCCTGGGAGAGCACTAGC
CCATCGACCTGTACAGGAACACTCTATATTGTTCTCGGACGGACAGATTACTAGAGTGCC
GCTTCAAGCCCCCTGTCGCCGACGTCTGTAATATGGCGTTGTGATTCACTCT
ATTGAGGCATTAACTGATGCGTAAGGAGATCTGGAATGAACTGGCTATGTCAGTAAAC
TGTGCAAACACCCAATGTCGTTAGTGTAGGTTCTGACCGATACTGCTTGAGAACT
CACAATTTCACAACGGGACATAATCCCTACGCCCATCATCTACTCGCGTCCCTGTGGC
TCCAGTTCATGTAATGGGAGAGTATCCTCCACAAGATCTAGTGCTATGGGGTATAGTAA
GCTCGTACTGTGGTAGACCGCACCGGGTAGGACCATCAGTAATAAGGATAGTGCAGAA
GCTCACAGACCACTGCCCTAGGGGGCTTACTTTACAAAAAGCGACTGTCAGTATAA
CCCCACGAGGATTGAAAAGGTGAACCGACCCAGACGATCCGGAGGCACGGGCTCAA
AGCCGCGACACGACGGCTGTCGGCCGGTAACAGTAACCCGGAGTGAACTCCTATGGG
GCTGGATAGAACAGCCCTGGTGGGCCCATCAGCAACCCGAATACGTGGCTTTCGGG
AGGCAGGGCGAGGGCGATGTCCTTCACTATTGAGGCCGTTGTTAACACTTGTGCG
TTCCTAGCCGCTATATTGTCCTTGCCGACTAATGTGGACAAGCACACCATAGCCATT
ATCGGAGCGCCTCGGAATACGGTATGAGCAGGCGCCTCGTGAGACCATTGCGAATACC
AGGTGTCCTGTAAGCAGCGAAGGCCCATACCGAGATAAAACTGCTAGGAAACCGCGTGT
CTACGACCGGTGGTGAATTAAATCTCGCTGACGTGTAGACATTCCAGGCAGTGCCTG
CTGTCGGGCTCCCTCTGGTACTGGTAGTTGAGACTTGCCTTGGAAAGACATAGCAAGAC
CCTGCCCTCTATTGATGTCACGGCGAATGTGGGGAGACAGCAGCGGCTGCAGACATC
AGATCGGAGTAATACTAACGTGGATAACTCCGTAACTGACTACGCCCTCTAGACTT
TACTTGACCAGATACGCTGTCTTGGCACGTGGATGGTTAGAGGAATCACATCCAAGAC
TGGCTAACGACGAAGCAACTCTTGAGTGTAAATTGTTCTCCTGTATTGGGGATGCGG
GTACTAGATGACTGCAGGGACTCCGACGTTAAGTACATTACCCGTATAGGCGCCGTT

CAGGATCACGTTACGCCATAAGATGGGAGCATGACTTCTCCGCTGCGCCCACGCC
AGTAGTGATTACTCCTATAACCCTCTGAGAGTCGGAGGCAGAAATCCGCCACGAATG
AGAATGTATTCCCCGACAATCATTATGGGGCGCTCTAACGCTTCACTCGGTTGAGC
CGGCTAGGCCTCTGCCGGAGTTGACGGACTGCTGCCACACCCGGCATTGTT
TTAGGGGGTTATCGAGGGCACTCGCAGCCAACCTGTCGGGACCAGCCGGCTGGTC
ATCGGGCTTATATAGCGAAATGCCGAGGACCCGGCCCCACGCTATGGAACGTCTTAGC
TCCGGCAGGCAATTAGGACAACGTAAGTATGGCGGATATAAACAGAGAAACGGGCGAA
TATACCTATTGATCGTACGGTAAATAGCCTCGCGAGGATGTCGACCGACACTAGA
CGGAGCACTCTAGTTATGCATATGGTCCACAGGACACTCGTCGCTTCGGATTGCCCTC
TATGTGACGGTTTGTAGCGCACTTATGCTCAGCACCGTTAAACCAGACCGACACTAGA
TCTATAAGGTCCGCCACGCAGACGAGAGCGCACGGAGATCACCGAGCGATCTATGAT
CGGCGACCATTGTGTGGTACTGGGGCCAGAGGTAACACTACGGTGCCGTAACAAACCT
CTCGGTCGCTGACGTTGTAGTCTAGTCTCATTATGATCGTACGCTATTAGGGATT
GACTGATACCGGAAGACATCTCAGTTGAAGTGGTGTATACGACAGAGACCGTGCACCTA
CCAAACCTCCTAGTCTAAGTTCAGACCAATTGGTAGTTGTCCAGAAACTCAGATTATC
ACCAGAGGACGCACGCTCACCTTATGATCCATTGATGTCCTGAGGGCTGCAATACATG
CAACCAGGCAGTCTCGCGTAAGTCCTAGTGCAATGGGCTTTTCCTGGTCCTCG
AGAAGAGGAGACGTCAGTCCAGATATCTTAATGTGGTAATTGGAAGGACTCTGGCCCT
CCACCCCTAGGCAGTGTACTCTTCATAAACGGCTATTAGTTATGGGTCCGAAGAT
TGAAAAAGGTGAGCGAACTCGGCCAACCGGAAAGACGGGCTCAAAGCAACCTGACC
ACGGTTGCGCGTCCGTATCAAGATCCTCTTAATAAGCCCCGTCACTGTTGGTTGTAGAG
CCCAGGACGGGTTGGCCAGATGTGCGACTATATCGCTTAGTGGCTCTGGGCCGCGT
GCGTTACCTTGCAGGAATTGAGGCCGTCGTTAATTCCCTGCATACATATTGCGTTTT
TTGTCCTTTATCCGCTACTTAGAAAAGAGACAGATAGCTTCTACCGGTGCGCCTCCG
TACGCAGTACGATCGCACGCCCATGAGAACGATAGGTAACCTGGTCTGTGAGCG
ACAAAAGCTAAATGGGAAATACGCCGATAAGTCGGTGCAGAACCTGGCTCGTAGCAA
TGTTGGTCTGACTATGATCTACATATTACAGGCCGTCGTTCTGCTCTGGTCAGCCTCTAA
TGGCTCGTAAGATAGTGCAGGCCGTCGTTCTGCTACTCGATGACCTCGGCTCCCCATTGCT
ACTACGGCGATTCTGGAGAGGCCAGCTGCCTGCTAATGTGAGGACAGTGTAGTATTA
GCAAACGATAAGTCCCCACTGGTTGTGGCTATCGAAAAGTGAACCTCATAACACATGC
TGTCCCACGCACATGGATGATTGGACAAATTGATTGAGTCTGAGTCTGATCACCTCACACA
GATCTAGAATCGAAAGCAGTGATCTCCGGTGCAGAACAAAAACTAGGTAACTAGAG
GGTCTCGACGTTCTAAACGTTGGTCCGTCTGAAACCGCCATCCAGGATCACGTCGCCCT
GAAAAAAAGATATCAGGAACTCTCCTCCTCAGCAGTCTGGTCTATGGAAACTACAGGACT
AACCTCCTGGCAACCGGGGCTGGGAATCTGTCACATGAGTCAAGGTATTGCTCGAT
AATCTATACTCCAGGCATCTAACCTTCCACTGCCTTAAGCGGGCCTGCCCTTCTGCC
TGTCGATCCATAGGACTCGTGCACCGCGCAGGCTTAGTCAGGGAGAAATATCCGGGG
CCAAAGACAACCAGCATCTCGCGTCTGGCTAACCCCCCTACATGCTGTTATAGAGAAC
AGTGGAAACCCGGTGCAGGCGATGGAATGTCCTAACCTGGCAGGGAAATCAAAGGGA
ACGTATATAACGCAAAGAAGCTGGAAATTGGCGAGGGAAATCCTGTTCTGTCTATCC
AAGAATGGGCATGAGGTGGCAACCGTCGTCAGCGTACAGGGTGCACTTGTAACCAT
TTGGGACACCGGACACTCGCTTTGAAATTACCTTAAGCGCGGGTATTGAACCGAG
GCTTATGCCCAAGATCGTAGCAAGCAGACTCAAACAGATATATTGCCCCCTTACAG
ACGAAACTAGTTGGAGGTTATGGAGCATACTACGTCGGGCGGCCACTGGTGA
TACACCCCAAGGGCAACGTTGATGCTCTAAACCTGGCAGGGCTGATCTACACTGGATTGCCATTCT
CTCAAAGTATTATGCAGGACGGCGTGCCTCATGTAACACTGTCATAACTTACCTGA
GACTAGTTGGAAGTGTGGCTAGATCTTGCTCACGCATCTAGTCGGTCCACGTTGGTT
TTAAGATCCAATGATCTCAAAACGCTGCAAGATTCTCAACCTGCTTACTAAGCGCTGG

GTCCTACTCCAGCGGGATTTTATCTAAAGACGATGAGAGGGAGTATTCGTAGACCACA
 TAGCTTCATGTCCTGATCGGAAGGATCGTGGCGCCCGACCCTCAGACTCTGTAGTGA
 GTTCTATGTCGAGCCATTGCATGCGAGATCGTAGATTGATAGGGGATACAGAATATCC
 CTGGATGCAATAGACGGACAGCTTGGTATCCTAACCGTAGTCGCGCGTCCGAACCCAGC
 TCTACTTTAGAGGCCCTCGGATTCTGGTGCCTGCAGGCCGCAGAACCGATTAGGGGCATG
 TACAACAATATTATTAGTCACCTTGAGACACGATCTCCCACCTCACTGGAATTAGTTC
 CTGCTATAATTAGCCTCCTCATAAGTTGCATTACTTCAGCGTCCCAACTGCACCCCTTACC
 ACGAAGACAGGTTGTCATTCCACTGCAGCGTTGGCAGGGGGTCGATGTC
 CGCGAAACGTTGCTGAAGGCTCAGGTTCTGAGCGACAAAAGCTTAAACGCGAGTTCC
 CGCTCATAACTTGGTCCGAATGCGGTTCTGCATCGTCCACTGAGTTGTTCATGTA
 GGACGGCGCAAAGTATACTTAGTTCAATCTCAATACCTGTATTATTGTACACCTACCG
 GTCACCAGCCAACAATGTGCGGACGGCGTTGCAACTCCAGGGCTAATCTGACCGTCC
 TAGATACGGCACTGTGGCAATACGAGGTAATGGCAGACACCCAGTGTGAAACAACACC
 TGACCTAACGGTAAGAGAGTCACATAATGCCTCCGGCCGCGTGCCTAGGGTATATTGG
 TCAGTATCGAATGGACTGAGATGAATCTTACACCGAAGCGGAAACGGGTGCGTGGACT
 AGCCAGGAGCAAACGAAAAATCCTGGCCTGCTGATGTCGTAACGTTCTAGAGATG
 GACGAAATGTTCACGACCTAGGAAAAGGTCGCCCTACAAAATAGATTGTGCTACTCTC
 CTCATAAGGAGTCCGGTGTACCGAAAGAACAGGCGAGCCTAGGTAGCAACCGCCGGC
 TACGGGGTAAGGTATCACTCAAGAACGAGGCTCGTAACACACGGCTAGCTGACTGT
 CTATCGCCTAGGTATAGGGACCTTGATATCTGCATGTCCAGCCTAGAATTACCT
 CAGCGCGCAGGTTGGTCGAGATAAAATCACCAGTACCCAGGACAGGGGGCTCGG
 CGCGTTGGCTAATCCTGGTACATCTGTTATGAATATTCACTAGTAAAGGATCTGTGTTAGAG
 GGACGAGTCACCATGTACCAAAAGCGATATTAAATCGGTGGAGTATTGTCGGTGA
 GACGCTGGTTACGTGGAAAGGTGCTGTGCTAACAGGCTAGGATATAATGCTGA
 AACCGTCCCCCAAGCGTTCAAG

Seq2:

CTATGACGGTATCTGCTGATGGGTAGGGAGATCCGGAATCTATCGGCCTATGTCACTGA
 AACTATCAAACACCCCATGTCGATACTGAAACGTATCGACGCATACTCCCTCCTGAAA
 ACGCACAATCATACAACGGGCACATAATGCGTACGCCATCTAGTACACCCATCTCTGT
 AGGTCCAGTTCAAGAGCTGGAAAGAGCACCCCTCCACTTGGTCAAGTGTATCTGGTAAG
 GTAAGCTCGTACCGTGATTGATCGGGCAGGGGTAAGACCATTAGAAGTAGGGATAGTCC
 CAAACCTCACTTACCACTGCCAATAAGGGCTCTATCTGAAGGATGAGTGTAGCCAGT
 GTAACCCGATGAGGAACCCAGAACGCCAACTGGGCCAGACAAACCCGGCGCTAACGCAC
 TCAAAGCCGGGACGCGACGCGACATAACGGCTAAGAGTAGCCCCGGAGTGTAGACCTT
 TGGGTTGGATAAAATCTGCGTGGTAACCGGCTCAACGACCCGTACACGTGGCACTTC
 AGGAGGCGCCCGCAGGGGGAAAGTTTCTGCTATTGAGGGCGTTGTGGTAAGTGT
 GCGTCCCTAGCCGCTATAATTGTTCTATGCCGAGTAATGAGAACACCACACCATAGCG
 ATTTGACGCGGGCGCTCGGAATACCGTTTGGCAGGGCGCTGCTAAGGCCATCGCGAAT
 TCCAGGTATCGTCATGCGTAGCGTAGGGCCGCACGCAAGTTAAACTGCTGGGAACCGC
 GTTTCCACGACCGGTGACGATTAATTGCGACGTGACGACATTCTGCTAATGCC
 CACCCGCCGGACCCCTCTCGTGTAGGGTAGCTGGACATGTCCTGTGAGATATAACAA
 GAGCCTGCCTGTTAATGATCTCACGGCGAAAGTCGGGGGACAGCAGCGGCTGCAGA
 CATTATAACCGCAACAAACACTAAGGTGAGATAACTCCGTAGTTGACTACGCATTCTCTAG
 ACCTTACTTGACCGGATACAGTGACTTGTACACGTTGTGGTTACAGCAATCACATCCA
 AGACTGCCTATGGAGGAAGCAACTCTTGAGTGTTAATATGTTGACCCCTGTATTAGGGAT
 GCGGTAGTAGATGAGCGCAGGGACACCGAGGTCAAGTACATTACCCCTCTCATAGGAG
 GTGTTCTAGATCACCACCATATCGAGCATGACACTATCTGCGCTGTCCCCA

TCCTGGTAGTCATTATTCTATCACGCTTCGAGTGACTGGTACGGATATCCCCCACGA
ATGAAAATCTTTCACTGACAGTCATATTGGGTGCTCCTAACGCTTCCACTGGCTGG
GTCTGCTAGGCCTCCGTCCCCGAGTTCGGCCTGCTGCCAGAGCCGGCCATTG
TCATTGGGCCTCACTTGAGGATACCCGACCTATTTGTCGGGACCCTCGGGTAGT
CGTTGGCTTATGCACCGTAAAGTCCTCCGCCCTCCCCCTACAGAAGATGATAAG
CTCCGGCAAGCAATTATGAACAACGCAAGGATCGGCATATAAACAGAGAAACGGCTGA
TTACACTTGTCGTGGTATCGCTAAATAGCCTCGGGAGCCTATGCCATACTCGTCC
GCGGAGCACTCTGTAACGCTTATGGCCATAGGACATTATCGCTCCGGTATGCC
TCTATTGACGATCCTTGGCGCACAGATGCTGGCACGAGCTAAATTAGAGCGACTGC
ACAACGTAAAGGTCGTCACGCAGACGGCCCAGGGAGACCACTGACCCATCAACC
TGTACGGGAACCTCTGTATCGTCTCGGACGGAGAGATAACTACAGTGCCGCTACAG
CCCCTCTGTCGCGCCACGCTGTAAATGTAGCCTATTGTGATTCCACCCATTGAGGC
ATTGACTGATGCGGGAAAGAGATCTGAAATGAACGGTCTATGCGACAGAAACTGTGCAG
CTACCTAATCTCCTAGTGTAGGTTCTGACCGATTGCTCGTTGAGAACTCACAATT
AACAAACAGAGGACATAAGCCCTACGCCATGATCTACTGACGTCCTGAGGCTGCAATT
CATGTAATGGGACAGTATCCGCGGCAAGTCCTAGTGCAATGGCGTATTCTACCCCTCGT
ACTGTAGTAGAGGCGACGCCGGTGCCTGTCATCACTAATAAGGATATTGGGAAAGACTCAC
AGGCCTCCGCCTTAGGCGGTGCTTACTCTTACATAAAAGGGCTGTTAGTATTACCCCG
CGAGGATTGAAAAGGTGAGCCAACCCGGCGATCCGGAGAGACGGGCTCAAAGCCG
CGTGACGACGGCTGTGGCCCGTAACAAATCCCCGAATAAGCTCCGTGAGCGTCG
GTTGAACAGCCCTGGTCGGCCCCATCAGTAGGCCAATATGTCGCTTACGGGCTCTGG
GCCGGGGTGCATAACCTTGCAGAAATCGAGGCCGTTGTTAATTCTGTGATTGCAATT
CCGCCTATATTGTCTTTGCCGGCTTATGGACAAGCATAGCATAGCCATTATCGG
AGCGCCTCCGTACACGGTATGATCGGACGCCGTGAGATCAAACGTATACCAGGTG
CCTGTGAGCGAAAGCCTATACGCGAGATAACTGCCAAAAATCCGCGTATTACGA
GTCGTGGCAAATTGGCTGGCTGTGGCTAGACATTCCAGGCCGTCGCTCTCG
GGTGCCTCTAGTGGCTGGCTAGATAGACTAGCCGCTGGTAAACACACCAGGCC
TCTCCATTGATGCCACGGCGATTGTTGAGAGACGCCAGCAGCAGTGC
GAGTAATACTAGCATGCGATAAGTCCCTAACTGACTATGCCCTCTGTAGAGTC
ACCACATATGCTCTGGCACGTGGATGGTTAGAGGAATCAGATTCAAGTCTGGT
ACCATCAAACAGGTCTTGAGTCTAAATTGTCGTCCTGCGTACGAGATGGAAACTA
GGTAACTACAGGGACTCCGACGTTATGTCGTTGCTCCGTAGAGGCCATTAGGAT
CACGTTACCGCGAAAAAAAGGGACCAGGAGCTCTCTCCCTGCGGTACGTC
AATTACACCATTAAACCCCTGAGAACCGGGAGGCAGGAAATCCGTACGTATGAGAAGG
TATTGCCGATAATCAATACCCAGGCTTAACCTTTCCACTCGCTGAGCCGGCTA
GGCCTTCTGCCGAAGTTCGATGGACTGGTGCACGCCAGGCATAGTTAGGAG
AATTATTGGGGCAGTGACAACCAACATCTGGGTCTGCCAACCGGTCTACACGCT
AATATAGCGAATCACCAGAACCCGGGCCACGCAATGGAACGTCTTAAC
GCAATTAAAGGGAACGTATGTATAACGCAAAAAACAGAAAAATAGGCGAATGAATCTT
TCTCTGTGATCGAAGAATGCCCTCGCGGAGGCATCGTCATGCTAGCGTGC
CTCTTGCTATCCATATGGTCCACAGGACACTCGTTGTTCGGATTACCG
CGGTTTCAGCCACGCTTATGCCAGCATCGTTACAACCAAGACCG
AGTCCGCCATGCAGACGAGACCAGTCGGAGATTACCGAGCATTCTACAGG
CACTAGTGAGCTACTGGAGCCAGGGTAACACGATGCCGCTAAGAAC
GACGCAAGCGATTACACTCCTGTCACATCATAATCGTTGCTATT
ACCGGAAACTTTCACTTGAGTATTGTATACGACAGGGTGC
GTTAAACTAAGTTCAGACTAGTTGGAAGTGTGCTAGATCTAG
GCCCACGCTTATTGATCCATTGATCTCCAGACGCTGCAAGATTGCA
CAGACTGGCGGTAGGTCCTAGTGAGCGGGACTTTCTATAGTC
CTTGAGAGGAG

GAGTCGTCACTTGTGATGTCCTGATTGGAAGGACCGTTGGCCCCCACCC
TTAGGCAGTGTACTCAGTCCATAAACGAGCTATTAGATATGAGGTCCGTAGATTGAAAAA
GGGTGACGGAATTGCCCCAACGGAAAGACGGACAACTAGGTATCCTGAGCACGGTT
GCGCGTCCGTATCAAGCTCCTCTTATAGGCCCGGTTCTGTTGGTCGTAGAGCGCAG
AACGGGTGGGGGATGTACGACAATATCGCTTAGTCACCTTGGGCCACGGTCCGCTA
CCTTACAGGAATTGAGACCGTCCTTAATTCCCTGCATATATGTTGCGTTCTCGACC
TTTAACCGCTCCCTAGGAGAAAGACAGATAGCTTACCCGTACTCCACCGTTGGCA
GCACGATCGCATGTCCCACGTGAACCATTGGTAAACCCGTGGCCTGTGAGCGACAAA
GCTTTAATGGGAAATTGCGCCATAACTTGGTCCGAATACGGGTCTAGCAACGTTCGT
CTGAGTTGATCTATATAATACGGGCGGTATGTCCTTGATCAACCTCCAATAGCTCG
TATGATAGTGCACCCGCTGGTGTACTCAATGATCTGGGCTCCCCGTTGCAACTACGG
GGATTTTCGAGACCGACCTGCGTTGGCATTGTGGCAGCTGAAGTATTAGCAAACG
TTAAGTCCCGAACTAGATGTGACCTAACGGTAAGAGAATTCTATAATACGTCCGCCGCA
CGCGCAAGGTACATTGGACAGTATTGAATGGACTCTGATCAACCTCACACCGATCTAG
AATCGAATGCGTAGATCAGCCAGGTGCAAACCAAAAAATTCTAGGTTACTAGAAGTTTGC
GACGTTCTAAGTGGACGAAATGATTGCGACCCAGGATGAGGTGCGCCCTAAAAAAT
AGATTCTGCAACTCTCCTCGTGAGCAGTCTGGTGTATCGAAAGTACAGGACTAGCCTTC
CTAGCAACCGCGGGCTGGAGTCTGAGACATCACTCAAGATATATGCTCGGTAAAGTAT
GCTCTAGCCATCTAACTATTCCCTATGTCCTATAGGGGCTACGTTATGCGCTGCGAA
CCATAGGATTGCGTCAGCGCGCAGGCTGGATCGAGATGAAATCTCCGGAGCCTAAGA
CCACGAGCGTCTGGCGTCTGGCTAATCCCCATACATGTTGTTATAAACAAATCAGTGGAA
ACTCAGTGCTAGAGGGTGGAGTGACCTTAAATCAAGGACGATATTAATCGGAAGGAGTA
TTCAACGCAATGAAGTGCAGGGTTGACGTGGGAATGGTGTCTGTCCAAACAGGTTAA
GGGTATGAGGCCGAAACCGTCCCCAAGCGTACAGGGTGCACTTGCAACGATTTCGG
AGTCAAAGACTCGCTTTGAAATTGCGCTCAAGGGCGAGTATTGAACCAGGCTTA
CGCCCAAGAACGTTAGCAAGGTGACTCAAACAAAGTACATCTGCCCGTTCATATGAA
TCAAGTTAGAAGTTATGGAGCATAATAACATGTGGATGGCCAGTGGTCGGTTGCTACACC
CCTGCCGCAACGTTGAAGGTCCCAGGATTAGACTGGCTGGATCTATGCCGTGACACCCGT
TATACTCCATTACCGTCTGTGGTCACAGCTTGTGGACTGGATTGCCATTCTCTCAG
TGTATTACGCAGGCCGGCGCACGGTCCCATATAAACCTGTCTAGCTTACCTGACTCTA
CTTGGAAATGTGGCTAGGCCTTGCCCACGCACCTGATCGTCCTCGTTGCTTTAGG
ACCGGATGAACTACAGAGCATTGCAAGAATCTTACCTGCTTACAAAGTGTGGATCCT
ATTCCAGCGGGATGTTTATCTAAACACGATGAGAGGAGTATTGTCAGGCCACATGGCT
TTCTTGTCTGGTCGGATCCATCGTTGGCGCCGACCCCCCATTCCATAGTGTGAGTTCT
CGTCCGAGCCATTGTATGCCAGATCGACAGACAGATAGCGGATCCAGTATATCCCTGGA
AACTATAGACGCACAGGGTGGAAATCTTAAAGTGAAGTGCAGCGTCCAAACCCAGCTCTATT
TTAGTGGTCATGGGTTCTGGTCCCCCGAGCCCGGAAACCGATTAGGACCATGTACAAC
AATACTTATTAGTCATCTTGTGACTACTTCTGCGACCCAAATGCACCCCTACCGAAG
AATTAGCCACCCCTCATAAGTGCACACTTCTGCGACCCAAATGCACCCCTACCGAAG
ACAGGATTGTCGATCCTATATTACGACTTGGCAGGGGTTGCAAGTCCCACCCCAA
ACGATGCTGAAGGCTCAGGTTACACAGGCACAAGTACTATATACGAGTTCCCCTCTT
AACCTGGATCGAATGCAAGTACATGCATCGTACCACTGTGTTGTCATCTAGGACGG
GCGCAAAGGATATATAATTCAATTAGAATACCTTATATTGTACACCTACCGGTACC
AGCCAACAATGTGCGGATGGCGTTACGACTTACTGGCCTGATCTCACCGCTTAGATA
CCGCACACTGGCAATACGAGGTAAGCCAGTCACCCAGTGTGATCAACAGCTAACGT
AACGGTAAGAGGCTCACAAATCGCACTGTGCGTCCCTGGTATTTACGTTAGCAT
CAGGTGGACTAGCATGAATCTTACTCCCAGGGCAAAACGGGTGCGTGGACAAGCGAG
CAGCAAACGAAAATTCTGGCCTGCTGGTGTCTCGTATTCTCTTGGAGATCGAGGAAA
TGTTCATGACCAAGGGAAAGGTCGCCCTACGAAATAGATTGCGTTACTGTCTGCATAA

GGAGTCGGTAGCGAAGGATGAAGGCGACCCCTAGGTAGCAACC GCCGGCTCGCG
GTAAGGTATCACTCAGGAAGCAGACACAGAAAGACACGGCTAGCAGATCGTTATCGG
CTAGGTCAAATAGAGTGCTTGTATCAGCATGTCTAGCTTACAATTAGTTAGTGC
TGATCTGAGTCGAGATAAAATCACCACTACCCAAAACCAGGCGGGCTGCCACGTTGGC
TAATCCTGGTACATTGTAAATCAATGTTAGAAGAAAATTGTTAGAAGGACGAGTCA
CCACGTACCAATAGCAACAAACGATCGGACTATTGTTAGTGCAGCTCGGA
TTACACGGAAAGGTGCTGTCCCACAGGCTAGGATATAATGCTGAGGCGTTACCC
CAATCGTCAGCGTGGGATTGCTACAACCTGAGTGTACATGTACGAGACCAGTTA
TGTATGCACAAGGCCACAATAGGACGTAGCCTCGAGTTAGTACGTAGCGTGGTC
TAAGCACAGTAGATCCTCCCCCGCATCCTATTAAAGTTAATTCTATAGCAATACGAT
CATATGCGGATGGGCGAGTGGCCGGTAGTCACACGTCTACCGCGGTGCTCAATGACCGG
GACTAAAGAGGCGAAGATTATGGTGTGACCCGTTATGCTCGAGTTAGTGCAGAGCGT
CATTGCGAGTAGTCGATTGCTTCTCAATCTCCAGCGATTAGCGTACAGCCCCAGG
GAACCCACAAAATGTGATCGCAGTCCATCCGATCGTACACAGAAAGGAAGGTCCCCATA
CACCGACGCACCTGTTACACGTGATGCATAAACGAGCCGCACGAACCAGAGAGCAT
AAAGAGGACCTCTAGTCCCTTACAAAGTACAGGTTCGCCTGTCGCCAAGATGCCTTACC
TAGATGCAATGACGGACGTATTCCCTGCGCTCAACGTTCTGCTTGCCTGGGATCC
AAGATTGGCAGCTGAAACCGCCTTCAAAGTGAAGTCCTCGTGTGACTAACTGTGCC
AAATCGTCTTGCAAAACTCCTGATCCAGTTAACTCACCAAATTATAGCCATACAGACCAA
ATTTAAATCATATCACCGACTAGCCTCTGCTTAATTCTGTGCTCAAGGGTTGGTCC
GCCCGAGCGGTGCGAGCGATTAGGACCATCTAATGCACTGTTACAAGACTTCTTAAA
TACTTCTCCTGCCCAGTAGCGGATGATAATGGTGTGCCAGCGGTGTTAGGTA
ACAGCACCGGTGCGAGCCTAATGTGCCGTCTCCACCAACACAAGGCTATCCGGTC
AATAGGATTCCGCAATGGGGTTAGCAAATGGCAGCCTAAACGATATGGGGACTTGC
TGTACATGCTTGGTCAATACATACGTGACCCAGTAGTTATCCTGTATCGGAACATCA
TGTGCATGGGCCAGCATAATCATGTCTGGGAAGTGGCCGTAGGATAAAATAATTCAA
TAAAGATGTCGTTTGCTAGTATACGTCTAGGCGTCACCCGCATCTGTGCAAGGTGG
GCCGACGAGACATTGTCCTGATTCTCCACTACTAATAGCACACACGGGGCAATACCA
GCACAAGCTAGTCGGGGACGCTCGTCAAGCATAACGAAAGAGCTAACGGCACGCC
TTCGCACTGTCAGGGTCACTGGGTGTTGCACTACCGTCAGGTACGCTAGTATGCG
CTTCCTCCAGAGGTATGTGGCTGCGTGGTCAAAGTGCAGGCATTGTATTGCTC
GTGTTACTCTACAAACTGACCTGGAGATCAAGGAGATGCTTCTGTGGAACGGACA
ACGCATCAACGCAACGGATCTACGTTACAGCGTCAGTGAATGTGAAACGGAGTTG
ACGAAAGCGACATTGGGATCTGTCGTTGCAATTGCGGGAAACATCCGTTACGAGGC
GGACACTGATTGACACGGTTTGCAAGAGTTAGGGAAATAGGTTAAATTGAGTGGCT
AAAATGCTATGTCGGATTAAAGTGTAGTAAACTGTGATTAACGGAGACGGTTAAAGA
CAGGAGTCGCAAAATCAAGCGGGCTATTACAACGGTTATTCTGTGGTTAGGCG
ACAATGTCCTGAAGAATATTAAGAAAAAGCACCCCTCGCCTAGAATTACCTACCG
CGGTCGACCATACTCGATTATCGCGCCACTCTCCATTAGTCGGCAGAGGTGGTT
TGTTGCGATAGCCCAGTATGATATTCTAAGGCGTTACGCTGATGAATATTCTACGG
GCCATAGGCGTTGAACGCTACACGGACGATCGAATTATGTATAGAGCGGGCATCG
AAGGTTAACTCTGTAGTTAACATGTAGCCCAGCTATTAGTACAGCAGTGC
TGACATTCTCATTAAATTCTACAGCCAAACGACCAAGTGCATTCCACGGAGCG
CGATGGAGATTCACTCGGAGCTCTGTAATAGGGACTAAAGAGTGTATGATAATCA
TGAGTGCCCGTTATGGTGGTGGAAACAGAGCGGTCTACGGCCAGTCGATTCT
CTCGAGTTCCGTCAGTTAACGCGTACACTCCCAGTGTACCCGCAAACCGTGATGG
TGCTTGGAGTCATCGCATGTAGGATGGTCTCCAGACACCAGGGCACCAGTTTC
CCAAAGCATAACGACGAGCAGTCATGAAAGTCTAGAAACTGGACGTGCC
TTCTCG

CGAATAATACCTCGAGCTGTACCGTTGCGCTGCCTAGATGCAGTGCTGCTTATCA
CATTGCTTCGACGA

Seq3:

CTATGACGGTATCTGCTGATGGGTAGGGAGATCCGGAATCTATCGGCCTATGTCACTGA
AACTATCAAACACCCCATGTCGATACTGAACGTATCGACGCATACCTCCTCCTGAAA
ACGCACAATCATACAACGGCACATAATCGTACGCCATCTAGTACACCCATCTCTGT
AGGTCCAGTTCAAGAGCTGGAAGAGCACCCTCCACTTGGTCAAGTGTATCCTGGTAAG
GTAAGCTCGTACCGTATTGTCGGCAGGGTAAGACCATTAGAAGTAGGGATAGTCC
CAAACCTCACTTACCACTGCCAATAAGGGTCCTTATCTGAAGGATGAGTGTAGGCCAGT
GTAACCCGATGAGGAACCCAGAACGCCAAGGGCTAAGAGTAGCCCCGGAGTGTAGACCTT
TCAAAGCCGGGACGCGACGCGACATAACGGCTAAGAGTAGCCCCGGAGTGTAGACCTT
TGGGGTTGGATAATCTGTCGTGGTAACCGGCTCAACGACCCGTACACGTGGCACTTC
AGGAGGCGCCCGCAGGGGGAGTTTCTGCTATTGAGGCCGTTCTGGTAACTAGTT
GCGTCCCTAGCCGCTATAATTGTTCTATGCCGAGTAATGAGAACACACCACAGCG
ATTGACGCGGCCCTCGGAATACCGTTTGGCAGGGCTGCTAAGGCCATCGCGAAT
TCCAGGTATCGTCATGAGCGTAGGGCCGACGCAAGTTAACTGCTGGGAACCGC
GTTTCCACGACCGGTGCACGATTAATTGCCCACGTGACGACATTCTGCTAATGCC
CACCCGCCGGACCCCTCGTGTGGTAGCTGGACATGTCCTGTGAGATAACAA
GAGCCTGCCTTTAATGATCTCACGGCAAAGTCGGGGGACAGCAGCGCTGCAGA
CATTATACCGCAACAAACACTAAGGTGAGATAACTCCGTAGTTGACTACGCATTCTCTAG
ACCTTACTTGACCGGATACAGTGACTTTGACACGTTGTGGTTACAGCAATCACATCCA
AGACTGCCTATGGAGGAAGCAACTTTGAGTGTAAATATGTTGACCCCTGTATTAGGGAT
GCGGGTAGTGTAGATGAGCGCAGGGACACCGAGGTCAAGTACATTACCCCTCATAGGAG
GTGTTCTAGATCACCATAACCACCATATCATTGAGCATGACACTATCTGCGCTGTCCCCA
TCCTGGTAGTCATTATTCTATCACGCTTCGAGTGACTGGTACGGATATCCCCCACGA
ATGAAAATCTTTCACTGACAGTCATATTGGGGTCTCTAACGCTTCCACTTGGCTGG
GTCTGCTAGGCCTCCGTGCCCCGGAGTTCGGCGCTGTGCTGCCGAGAGCCGGCCATTG
TCATTGGGCCTCACTTGAGGATACCCGACCTATTGTCGGGACCACTCGGGTAGT
CGTTGGCTTATGCACCGTAAAGTCCTCCGCCCTCCCCGCTACAGAACGATGATAAG
CTCCGGCAAGCAATTATGAACAAACGCAAGGATCGGCGATATAAACAGAGAACGGCTGA
TTACACTTGGTGTGGTATCGCTAAATAGCCTCGGGAGCCTATGCCATACTCGTCC
GCGGAGCACTCTGGTAACGCTTATGGCCATAGGACATTGACGCTTCCGGGTATGCGC
TCTATTGACGATCCTTGGCGCACAGATGCTGGCCACGAGCTAAATTAGAGCGACTGC
ACAACGTAAAGGTCCGTACGCAGACGACGGCCCAGGGAGACCACTGACCCATCAACC
TGTACGGGAACCTCTGTATCGTTCTCGGACGGAGAGATAACTACAGTGCCGCTTACAG
CCCCTCTGTCGCCGACGCTGTAAATGTAGCCTCATTGTGATTCCACCCATTGAGGC
ATTGACTGATGCGGGAAAGAGATCTGAAATGAACGTTCTATGCGACAGAAACTGTGCAG
CTACCTAATCTCCTTAGTGTAGGTTCTGACCGATTGCTGCTTGTGAGAAACTACAATT
AACAAACAGAGGACATAAGCCCTACGCCATGATCTACTGACGCTCCCTGAGGCTGCAATT
CATGTAATGGGACAGTATCCGCGGCAAGTCCTAGTGAATGGCGGTATTCTACCCCTG
ACTGTAGTAGAGGCGACGCCGGTGCCTACTCTTACATAAAAGGGCTGTTAGTATTACCC
AGGCCTCCGCCCTTAGGCCGGTGCCTACTCTTACATAAAAGGGCTGTTAGTATTACCC
CGAGGATTGAAAAGGTGAGCCAACCCGGCCGATCCGGAGAGACGGGCCTAAAGGCCG
CGTGACGACGGCTGTGGCCCGTAACAAATCCCGCAATAAGCTCCGTGAGCGTCG
GTTGAACAGCCCTGGTCGGCCCCATCAGTAGCCGAATATGTCGCTTACGGGTCTGG
GCCGGGGTGCAGATACCTTGCAGAAATCGAGGCCGTTGTTAATTCTGTGATTGCAATT
CCGCCTATATTGTCTTTGCCGGCTTATATGGACAAGCATAGCAGGCCATTATCGG
AGCGCCTCCGTACACGGTATGATCGGACGCCGTGAGATCAATACGTATACCGAGGTGT

CCTGTGAGCAGCGAAAGCCTATACCGAGATAACTGCCAAAATCCGCGTATTACGA
GTCGTGGCAAATTGGTCTGGCTGGTAGACATTCCAGGCAGTGCCTGCTCTCG
GGTGCCTCTAGGGCTGGTAGATAGACTAGCCGCTGGAAACACACCATGACCCGGC
TCTCCATTGATGCCACGGCGATTGGAGAGAGCCAGCAGCAGTCAACATCAGATCA
GAGTAATACTAGCATGCGATAAGTCCCTAACTGACTATGGCCTCTGTAGAGTCAACTTC
ACCACATATGCTCTGGCACGTGGATGGTTAGAGGAATCAGATTCAAGTCTGGTTA
ACCATCAAACAGGTCTTGAGTCTAAATTGTCGTCTCGTACAGAGATGGAAATACTA
GGTAACTACAGGGACTCCGACGTTATGTACGTTGCTCCGTAGAGGGGCCATTAGGAT
CACGTTACCGCGAAAAAAAGGGACCAGGAGCTCTCTCCCCTGCGGTACGTCTATAGA
AATTACACCATTAAACCCCTGAGAACCGGGAGGCGGGAAATCCGTACGTATGAGAAGG
TATTGCCCCATAATCAATAACCCAGGCTTAACCTTCCACTCGCTTGAGCCGGCTA
GGCCTTCTGCCGAAGTTGATGGACTGGTCCAACGCGCAGGCATAGTTAGGAG
AATTATTGGGGGCAGTGACAACCATCTCGGTCTGCCAACCGGTCTACACGCT
AATATAGCGAACATACCGAGAACCCGGGCCACGCAATGGAACGTCCTAACTCCGGCAG
GCAATTAAAGGGAACGTATGTATAACGCAAAAAACAGAAAAATAGGCGAATGAATCTT
TCTCTGTATCGAAGAATGGCCTCGCGGAGGCATCGTCATGCTAGCGTGCAGGGTA
CTCTGCTATCCATATGGTCCACAGGACACTCGTTGTTGGATTACCCCTTATGCGC
CGGTTTCAGCCACGCTTATGCCAGCATCGTTACAACCAGACCGATACTAGATGTATAA
AGTCCGCCATGCAGACGAGACCAGTGGAGATTACCGAGCATTCTATCAGGTCGGCAGC
CACTAGTGAGCTACTGGAGCCGAGGGTAACCACGATGCCGTAAGAACCTCTGGTC
GACGCAAGCGATTACACTCCTGTCACATCATAATCGTTGCTATTAGGGTTGACCAAC
ACCGGAAAATTTCACTTGAAGTATTGTATAACGACAGGGTGCCTGACCTACCAAACCT
GTTTAAACTAAGTCAGACTAGTTGAAGTGTGTCTAGATCTAGTTGTCCTAGAGG
GCCACGCTTATTTTATGATCCATTGATCTCCAGACGCTGCAAGATTGCAACCAGG
CAGACTTGGCGGTAGGTCTAGTGCAGCGGACTTTTCTATAGTCTTGAGAGGAG
GAGTCGTAGTCCAGATACTTGTCTGATTGGAAGGACCGTGGCCCCCACC
TTAGGCAGTGTACTCAGTCCATAACGAGCTATTAGATATGAGGTCCGTAGATTGAAA
GGGTGACGGAATTGCCGAACGGAAAGACGGACAACACTAGGTATCCTGAGCACGGTT
GCGCGTCCGTATCAAGCTCTTATAGGCCCGTTCTGTTGGCTGTAGAGCGCAG
AACGGTTGGGGGATGTACGACAATATCGCTTAGTCACCTTGGCCACGGTCCGCTA
CCTACAGGAATTGAGACCGTCTTAATTCCCTGCAATATGTTGCGTTCTCGACC
TTTAACCGCTCCCTAGGAGAAAGACAGATAGCTTACCCGTACTCCACCGTTGGCA
GCACGATCGCATGCCCACGTGAACCATTGGTAAACCCCTGTGGCCTGTGAGCGACAAA
GCTTAATGGAAATTGCGCCATAACTGGTCCGAATACGGTCCTAGCAACGTTCTG
CTGAGTTGATCTATATAATACGGCGGTATGTCGCTTGATCAACCTCAATAGCTCG
TATGATAGTGCACCCGCTGGTGATCACTCAATGATCTGGCTCCCGTTGCAACTACGG
GGATTTGAGACCGACCTGCGTTGGCATTGTGGCACAGTGAAGTATTAGCAAACG
TTAAGTCCGAACTAGATGTGACCTAACGGTAAGAGAATTCTATAATACGTCCTGCCGCA
CGCGCAAGGTACATTGGACAGTATTGAATGGACTCTGATCAACCTCACACCGATCTAG
AATCGAATGCGTAGATCAGCCAGGTGCAAACCAAAAATTCTAGGTTACTAGAAGTTGC
GACGTTCTAAGTGTGGACGAAATGATTGCGACCCAGGATGAGGTGCCCTAAAAAAAT
AGATTCTGCAACTCTCTCGTGAGCAGTCTGGTATCGAAAGTACAGGACTAGCCTTC
CTAGCAACCGCGGGCTGGAGTCTGAGACATCACTCAAGATATGCTCGTAACGTAT
GCTCTAGCCATCTAACTATTCCCTATGCTTATAGGGCCTACGTTATCTGCCGTGCAA
CCATAGGATTGCGTCAGCGCCAGGCTGGATCGAGATGAAATCTCCGGAGCCTAAGA
CCACGAGCGTCTGGCGTCTGGCTAATCCCCCTACATGTTGTTATAAACAAATCAGTGGAA
ACTCAGTGCTAGAGGGTGGAGTGACCTAAATCAAGGACGATATTAAATCGGAAGGAGTA
TTCAACGCAATGAAGTGCAGGGTTGACGTGGAAATGGTCTGTCCAAACAGGTAA
GGGTATGAGGCCGCAACCGTCCCCAAGCGTACAGGGTGCACTTGCAACGATTTCGG

AGTCCAAAGACTCGCTTTGCAAATTGCGCTCAAGGGCGAGTATTGAACCAGGCTTA
CGCCCAAGAACGTAGCAAGGTGACTCAAACAAAGTACATCTGCCCGCTTCATATGAA
TCAAGTTAGAAGTTATGGAGCATAATAACATGTGGATGGCCAGTGGTCGGTGCTACACC
CCTGCCGCAACGTTGAAGGTCCCAGGATTAGACTGGCTGGATCTATGCCGTGACACCCGT
TATACTCCATTACCGTCTGTGGTCACAGCTTGTGGACTGGATTGCCATTCTCTCAG
TGTATTACGCAGGCCGGCGCACGGTCCCATAAAACCTGTATAGCTTACCTGACTCTA
CTTGGAAATGTGGCTAGGCCTTGCCCACGCACCTGATCGGTCCTCGTTGCTTTAGG
ACCGGATGAACTACAGAGCATTGCAAGAATCTCTACCTGCTTACAAAGTGCTGGATCCT
ATTCCAGCGGGATGTTTATCTAAACACGATGAGAGGAGTATTGTCAGGCCACATGGCT
TTCTTGTCTGGTCGGATCCATCGTGGCGCCGACCCCCCATTCCATAGTGAGTTCTT
CGTCCGAGCCATTGTATGCCAGATCGACAGACAGATAGCGGATCCAGTATATCCCTGGA
AACTATAGACGCACAGGTTGGAATCTTAAGTGAAGTCGCGCGTCCAAACCCAGCTCTATT
TTAGTGGTCATGGTTCTGGTCCCCCGAGCCGGAACCGATTAGGACCATGTACAAC
AATACTTATTAGTCATCTTAGACACAACTCCCTGCTCAGTGGTATATGGTTTGCTAT
AATTAGCCACCCCTCATAAGTTGCACTACTTCTGCGACCCAAATGCACCCCTTACACGAAG
ACAGGATTGTCGATCCTATATTACGACTTGGCAGGGGTTGCAAGTCCCACCCAA
ACGATGCTGAAGGCTCAGGTTACACAGGCACAAGTACTATATATACGAGTTCCCCTCTT
AACCTGGATCGAATGCAGAATCATGCATCGTACCACTGTGTTGTCATCTAGGACGG
GCGCAAAGGATATATAATTCAATTAGAATACCTTATATTGTACACCTACCGGTCAACC
AGCCAACAATGTGCGGATGGCGTTACGACTTACTGGCCTGATCTCACCGCTTAGATA
CCGCACACTGGCAATACGAGGTTAAAGCCAGTCACCCAGTGTGATCACAGCTAACGT
AACGGTAAGAGGCTCACAAAATCGCACTGTCGGCGTCCCTGGTATTTACGTTAGCAT
CAGGTGGACTAGCATGAATCTTACTCCCAGGCAAAACGGGTGCGTGGACAAGCGAG
CAGCAAACGAAAATTCTGGCCTGTTGGTGTCTCGTATTCTCTTGGAGATCGAGGAAA
TGTTTCATGACCAAGGGAAAGGTGCGCCCTACGAAATAGATTGCGTTACTGTCTGCATAA
GGAGTCCGGTAGCGAAGGATGAAGGCGACCCCTAGGTAGCAACCGCCGGCTCGCG
GTAAGGTATCACTCAGGAAGCAGACACAGAAAAGACACGGCTAGCAGATCGTTATCGG
CTAGGTCAAATAGAGTGTGTTGATATCAGCATGTCTAGCTTAGAATTGAGTTAGTGC
TGATCTGAGTCGAGATAAAATCACCAGTACCCAAAACAGGCGGGCTGCCACGTTGGC
TAATCCTGGTACATTTGTAATCAATGTTAGAAGAAAATTGTGTTAGAAGGACGAGTCA
CCACGTACCAATAGCAACAAACGATCGGCGGACTATTCAATTGTTGGTGGTACGCTCGGA
TTACACGGAAAGGTGCTTGTGCCCAGGCTAGGATAATGCTGAGGCGTTACCC
CAATCGTCAGCGTGGATTGCTACAACCTCTGAGTGCTACATGTACGAGACCATGTTA
TGTATGCACAAGGCCGACAATAGGACGTAGCCTCGAGTTAGTACGTAGCGTGGTCGCA
TAAGCACAGTAGATCCTCCCCCGCATCCTATTAAAGTTAATTCTATAGCAATACGAT
CATATGCGGATGGCAGTGGCGGTAGTCACACGTCTACCGCGGTGCTCAATGACCGG
GACTAAAGAGGCGAAGATTATGGTGTGACCCGTTATGCTCGAGTTGGTCAAGAGCGT
CATTGCGAGTAGTCGATTGCTTCTCAATCTCCAGCGATTTAGCGTACAGGCGTACAG
GAACCCACAAAATGTGATCGCAGTCCATCCGATCGTACACAGAAAGGAAGGTCCCCATA
CACCGACGCACCTGTTACACGTCGTATGCATAAACGAGCCGCACGAACCAGAGAGCAT
AAAGAGGACCTCTAGTTCTTACAAAGTACAGGTTGCCCTGTCGCCAAGATGCCTTAC
TAGATGCAATGACGGACGTATTCTCTGGCCTCAACGGTTCTGCTTGCCTGGGATCC
AAGATTGGCAGCTGAAACCGCCTTCCAAAGTGAGTCCTCGTGTGACTAACTGTGCC
AAATCGCTTGCAAACCTCTGATCCAGTTAACTCACCAAATTAGCCATACAGACCCAA
ATTTAAATCATATCACCGCACTAGCCTCTGCTTAATTCTGTGCTCAAGGGTTGGTCC
GCCCGAGCGGTGCAGCCGATTAGGACCATCTAATGCACCTGTTACAAGACTTCTTTAAA
TACTTCTTCTGCCAGTAGCGGATGATAATGGTTGCCAGCCGGTGTGAAAGGTA
ACAGCACCGGTGCGAGCCTAATGTGCCGTCTCCACCAACACAAGGCTATCCGGTCGTAT
AATAGGATTCCGCAATGGGTTAGCAAATGGCAGCCTAAACGATATGGGGACTTGCAG

TGTACATGCTTGGTTCAATAACATACGTGACCCAGTAGTTATCCTGTATCGGAACATCAAT
TGTGCATGGGCCAGCATAATCATGTCATCTGGGAAGTGGCCGTAGGATAAAATAATTCAA
TAAAGATGTCGTTTGCTAGTATACGTCTAGGCCTACCCGCCATCTGTGCAGGTGG
GCCGACGAGACATTGTCCCTGATTCTCCACTACTAATAGCACACACGGGGCAATACCA
GCACAAGCTAGTCTCGCGGGAACGCTCGTCAGCATAACGAAAGAGCTTAAGGCACGCCAA
TTCGCACTGTCAGGGTCACTGGGTGTTGCACCTACCGTCAGGTACGCTAGTATGCGTT
CTTCCTTCCAGAGGTATGTGGCTCGTGGTCAAAAGTGCAGGCATTGTATTGCTCCTC
GTGTTTACTCTCACAAACTTGACCTGGAGATCAAGGAGATGCTTCTTGGAACGGACA
ACGCATCAACGCAACGGATCTACGTTACAGCGTGCATAGTAAAACGGAGTTGCTGACG
ACGAAAGCGACATTGGGATCTGTCTGTCATTGCAGGAAAACATCCGTTACGAGGC
GGACACTGATTGACACGGTTTGCAGAAGGTTAGGGAAATAGGTTAAATTGAGTGGCTTA
AAAATGCTATGTCGGATTAAAGTGTAGTAAACTGTGATTAACGGAGACGGTTAAAGA
CAGGAGTTCGCAAAATCAAGCGGGTCATTACAACGGTTATTCTGGTGGTTAGGCGT
ACAATGTCCTGAAGAATATTAAGAAAAAGCACCCCTCGTCGCCTAGAATTACCTACCG
CGGTCGACCACCTCGATTATCGCGCCCCTCTCCATTAGTCGGCAGAGGTGGTTG
TGTTGCGATAGCCCAGTATGATATTCTAAGGCCTACGCTGATGAATATTCTACGGAATT
GCCATAGGCGTTGAACGCTACACGGACGATACGAATTATGTATAGAGCGGGTCATCGA
AAGGTTATACTCTGTAGTTAACATGTAGCCCAGCTTACAGCCAAACGACCAAGTGCATTCCACGGAGCG
CGATGGAGATTCACTCACTCGGAGCTCTGTAATAGGACTAAAAGAGTGTATGATAATCA
TGAGTGCCCGCTTATGGTGGTGCAGAACAGAGCGGTCTACGGCCAGTCGTATTCTT
CTCGAGTTCCCGTCCAGTTAACGCGTACACTCCCAGTGTACCCGCAAACCGTGATGGCTG
TGCTTGGAGTCAATCGCATGTAGGATGGTCTCCAGACACCGGGGACCCAGTTTCACGC
CCAAAGCATAAACGACGAGCAGTCATGAAAGTCTTAGAAACTGGACGTGCCGTTCTG
CGAATAATACCTCGAGCTGTACCGTTGCGCTGCCTAGATGCAGTGCTGCTTATCA
CATTGCTTCGACGA

Input

Output

The screenshot shows a web browser window with the URL www.ebi.ac.uk/Tools/services/web/toolresult.cgi?jobId=clustalw2-l20141201-104844-0296-57914524-es. The title bar includes tabs for 'chicken - P', 'Gorilla gorilla - P', 'Gorilla gorilla - P', 'Facebook', 'Letter Counter', 'Random DI', 'History', 'Alignments', and 'LESSON ANDREWS'. The main content area is titled 'ClustalW2' and displays a multiple sequence alignment for 'CLUSTAL 1.1 multiple sequence alignment'. The alignment involves three sequences: Seq1, Seq2, and Seq3. Seq1 contains the sequence: GTGATCCAAAGCTTGGATGTGAATCTAATATCCTTAAGTTGGATTCCTGGCCGAAATTTC. Seq2 contains: AGCTTCAGCATCTACGAAATTGATGAAATTGCCAAGAAAAGCACTGTTGGCGCCGCC. Seq3 contains: TTTTGCGAAACCTGTGAGCTTCTTCGCTCAAAAGCCGAGCTTCTGATGCGGGCCGATCTGG. The bottom of the page shows the desktop taskbar with icons for File Explorer, Libraries, blesson andrews v..., Computer, Network, Control Panel, and several pinned applications: BioSli..., Bioinf..., Firefox, Align..., Untit..., Untit..., BIOIN..., and Micro... The system tray shows the date as Monday, 12/1/2014 and the time as 4:26 PM.

(ii)

(iii)

(iv)

The screenshot displays a sequence alignment between two genomes. The top part of the page shows several browser tabs: 'chicken - P' (highlighted), 'Gorilla gori', 'Facebook', 'Letter Counter', 'Random DI', 'History', 'Alignments', and 'LESSON ANDREWS'. Below the tabs, the URL is 'www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2-i20141201-104844-0296-57914524-es'. The main content area shows a sequence alignment with three columns: 'Seq1', 'Seq2', and 'Seq3'. The sequences are aligned with various mutations and annotations. For example, Seq1 has a mutation 'A>T' at position 1000, and Seq3 has a mutation 'C>G' at position 1000. There are also several 'N' characters representing gaps or undefined positions. The alignment is color-coded, with red and green boxes highlighting specific regions of interest.

(v)

The screenshot shows a browser window with multiple tabs open, including 'chicken - Pi', 'Gorilla gori', 'Facebook', 'Letter Cour', 'Random D', 'History', 'Alignments', and 'LESSON ANDREWS'. The main content area displays a sequence alignment between the Gorilla gori genome (left) and the chicken genome (right). The alignment is color-coded to highlight matches (green), mismatches (red), and gaps (blue). The sequences are presented in two columns, with the Gorilla gori sequence on top and the chicken sequence on the bottom. The alignment covers several thousand base pairs, with labels indicating the start and end of each sequence. The bottom of the screen shows the Windows taskbar with icons for the desktop, libraries, file explorer, bioinformatics tools (BioSli... Bioinf... Align...), network, control panel, and system status.

(vi)

(vii)

The screenshot shows a web browser window with multiple tabs open. The active tab displays a sequence alignment between the Gorilla gori genome (hggori) and the chicken genome (Gallus gallus). The alignment is presented in a grid format where rows represent the Gorilla gori sequence and columns represent the chicken sequence. The alignment highlights identical nucleotides in green and mismatched nucleotides in red. A vertical scale bar on the left indicates positions 1 through 4800. The sequence starts with ATGGGGAGCTCGACGGCTTCCGGCTTAAAGGGCTTCCTTCAAGGGCTGCTG 2296.

(viii)

(ix)

A screenshot of a web browser window titled "www.ebi.ac.uk/Tools/services/web/toolresult.cgi?jobId=clustalw2-l20141201-104844-0296-57914524-es". The main content area displays a sequence alignment between two DNA strands. The top strand is labeled "Seq1" and the bottom strand is labeled "Seq2". The alignment shows numerous matches with vertical bars, mismatches with diagonal bars, and gaps with dashes. The sequences consist of long stretches of A, T, C, and G nucleotides. The browser interface includes a navigation bar at the top with tabs for "chicken - P", "Gorilla gori", "Facebook", "Letter Cour", "Random D", "History", "Alignments", and "BLESSON ANDREWS". Below the main content are standard Windows taskbar icons for "Desktop", "Libraries", "blesson andrews v...", "Computer", "Network", "Control Panel", and system status indicators.

(x)

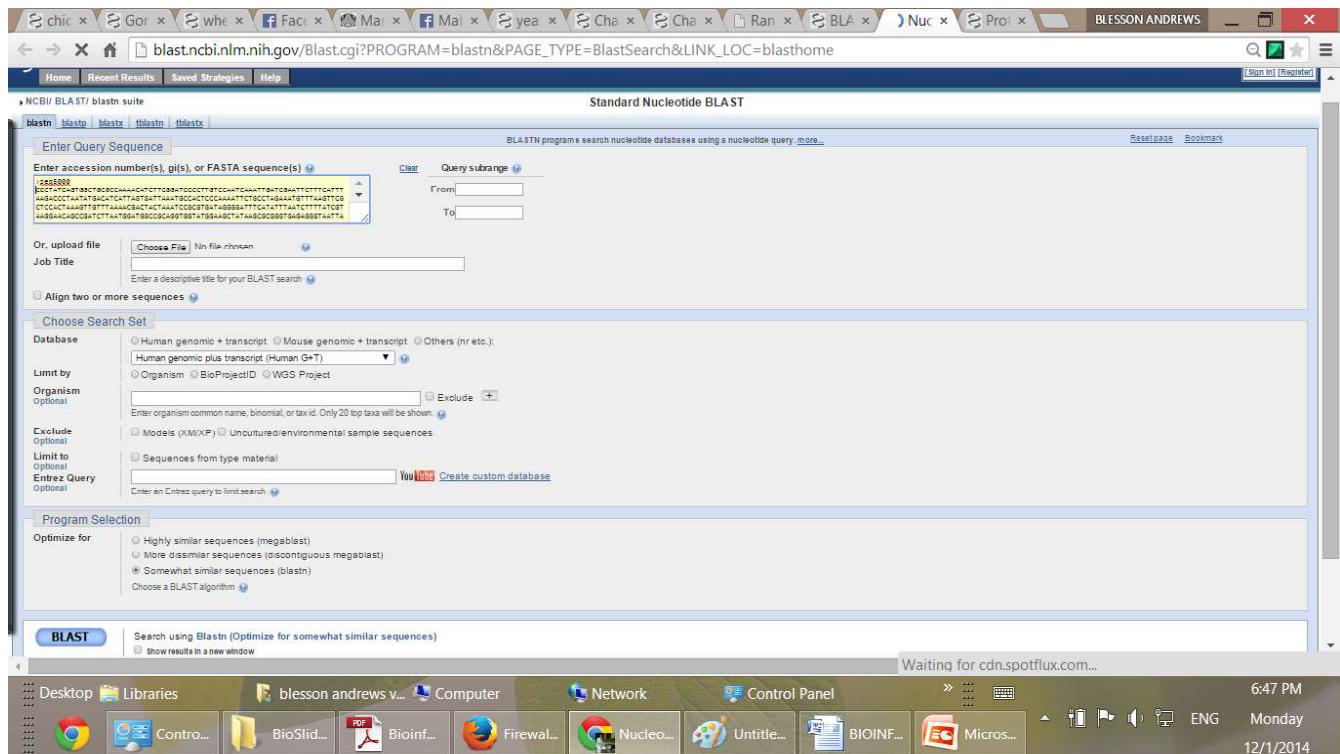
(xi)

(xii)

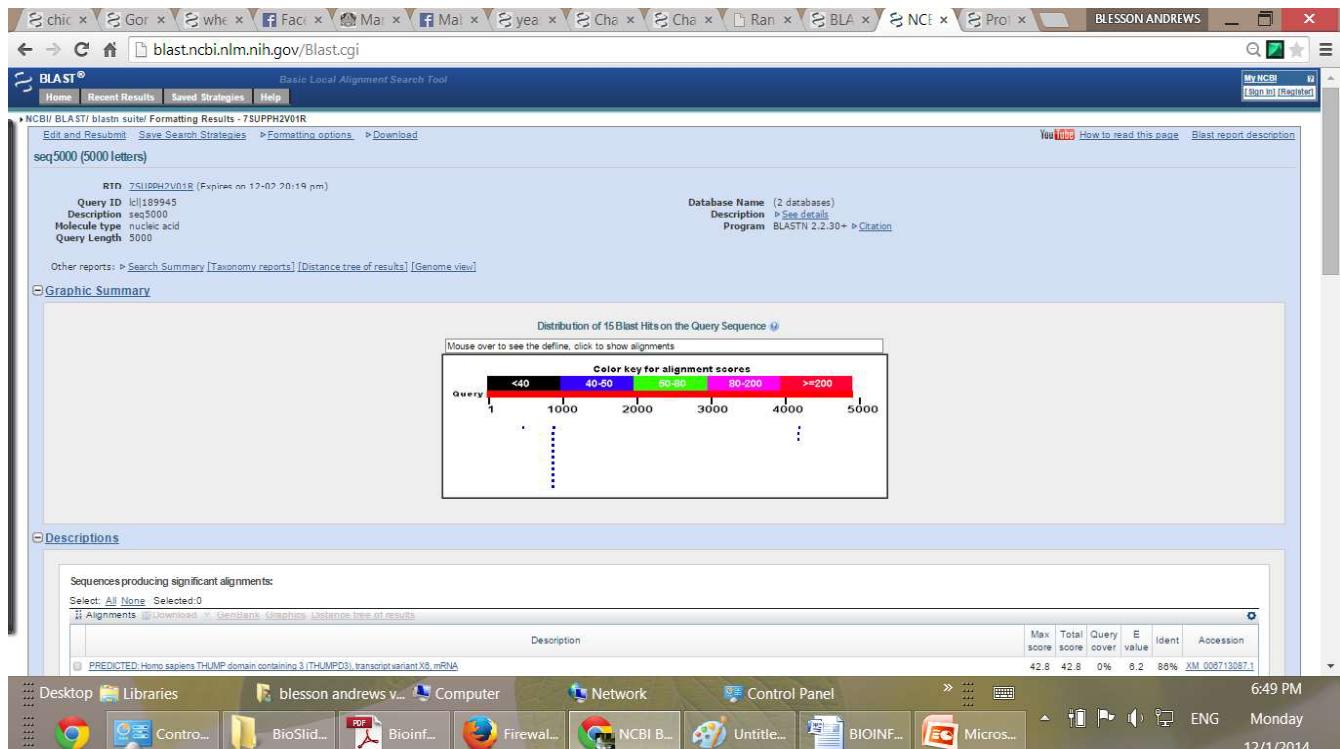
(xiii)

TCTTCGGGTATCATGGAGCCTATGGTTAACGTGTCGATAACGAACCTCGACATGATAA
AGTCCCCCCCCTCGCGACTACCAGAGAAGAAGACTACTGAGTTGAGCGTCCCAGCACTT
CAGCCAAGGAAGCTACCAATTAGTTAGTTCGAGTGTACGTCACGCTGACCTCGCGGGTAGAT
TGCCGAGCGTAGAGCTACGAGCCAGCGAAACAGTAAGGCCTTTAAGTATGGGGAG
TAAGTGATCGAACGCTTCAGATGTGACCATACTTAGGCTGGATCTCGTCCCCTGAATT
TTAACCCCTACCAACTACGAGATATGAGGTAAGCCAAAAAGCACGTGGTGGCGCTCAC
CGACTGTTCCCAAACGTAACTCATCGTCCGTCAAGGCCTGACTTACTTCCCAGGCCCT
TCCATGCGCGGACCATAACCGTCTAGTTCTCGGTTATGTTCCGATGTAGGAGTGAGC
CTACCTCCGTTGCGTCTTACCAATGAAAAAGCTATGCACTTGTACAGGGTGCCT
CGGGTTCTGAACCTCTCAGATAGTGGGATCCCAGGAAAGGGCCTATATTGCGGTCCA
ACTTAGGCCTAACCTCGATGCTACCTACTCAGACCCACCCCGCGCGGGTAAATATGG
CACTCATCCCAGCTGGTCTTGGCGTCTACGCAGCCACATGTTCTTAACAGTTGTCTG
GTAGCACAAAAGTATTACCATGGTCTAGAAGCCCAGAGTTAGTCGAACCTAATGC
CACAAATGAGACAGGACGCCAATGGGTACCGGACATTAGGTCGAGCTCAGTCGGTAAC
GGAGAGACCCTGCGCGTACTTAATTATGCATATGAAACCGGCCAAGTGACGCCAGGC
AAGTCTCAGCAGGTTCCCGTGTAGCTGAGGGTAAACATACAAGCCGATTGAACATGG
GTTGGGGCTTCAAATCGTCGAGGACCCCACAGTACCTCGGAGACCAAGTAGGGCACC
CTATAGTTCGAACGAGAACTATTGAGGGCGAGCCCTCATCGTCTCTGCGGATG
ACTTAACACGCTAGGGACGTGGAGTCGATTCCATCGATGGTTATAATCAAAGATTGGA
ATGCTGTCGGAGGGTGAATCTAACGGTGCCTACTCGATTGCTCAGTCGCTTCGTA
TGCAGCAGGTTCCGCTATTCACTTGGTCCGAAGCCTGCTGATATATGAATC
CAAACATAGAGCGGGCTCTTGACATTGGAGTTGAAATATCTAATAACTCCAATCGGCTT
TTACGTGCACCACCGCGGGCGTGAAGGAGGGTCTCACACCGAGAAACAAGACAGTC
CGGGCTGGAAGTAGCGCCGGCTAACGGAAAGACGCCCTGGTACAGCAGGACTATGAAACCG
GTACAAAGGCAACATCCTCACTTGGTGAACCGAACCGCGGTATCAAGGTTACTTTTGG
ATACCTGAAACAAATCCATGGTAGTCCTAGACTTGGAGTCTATCACCCCTAGGGCCC
ATATCTGAAATAGACGCCAACGTTCAATCCGTATTCCGACGTACGATGGAACAGTGTGG
GTGAGACGTGCTTCATTATACCCCTACGCAGGCTGGACCGAGGTCCGCAAGGCGCGGC
GGTGCACAAGCAATTGACAACCAACCGTGTATTCAATTGGTACCGAGGGACTTAAG
CCGAGTCAATGGAGCTGCAATACAGAGTTACCGCATCTGCCGTAACGTACAAACTGT
GATCCACCACAAGTCAAGCCATTGCCCTTAGACACGCCGTTAGAGTAATTATGTAAC
TTGCGCGGCTTGACTACGACTCGTCAGTCAGTCAGTCCGAGGGCACAATCCTATTCCCATT
GTATGTTCAGCTATCTTCTACCCATCCCCGGAAGTTAAGTAGGTCGTGAGATGCCGAGG
AGGCTCTCGTTATCCCGTGGACATCAAGCTCCCTTGATAAAAGCACCCCGCTCGGG
TATGGCAGAGAGGGACGCCCTCTGAATTGTGCTATCCCTGACCTTATCAAAGCTGCTAC
CAATAATTAGGATTATTGCCCTGCGACAGACCTCCTACTCAGACTGCCACATTGAGCT
AGTCAGTGAGCGATAAGCTGACCCGCTTCTAGGGTCGAGTACGTGAACTAGGGCT
CCGGACAGGGCTATACTCGAGTTGATCTGCCCGACAACACTGCAAACCTCAACT

Input



Output



(ii)

blast.ncbi.nlm.nih.gov/Blast.cgi

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenBank Graphics Distance list of results

Description	Max score	Total cover	E value	Ident	Accession
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X6, mRNA	42.8	42.8	0%	6.2	86% XM_008713087_1
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X5, mRNA	42.8	42.8	0%	6.2	86% XM_008713088_1
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X4, mRNA	42.8	42.8	0%	6.2	86% XM_005265024_2
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X3, mRNA	42.8	42.8	0%	6.2	86% XM_005265023_2
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X2, mRNA	42.8	42.8	0%	6.2	86% XM_005265022_2
PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X1, mRNA	42.8	42.8	0%	6.2	86% XM_005265021_2
Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant 2, mRNA	42.8	42.8	0%	6.2	86% NM_011140382_1
Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant 1, mRNA	42.8	42.8	0%	6.2	86% NM_015453_2
Homo sapiens chromosome 3, alternate assembly CHM1_1	42.8	42.8	0%	6.2	86% NC_018014_2
Homo sapiens chromosome 13, alternate assembly CHM1_1	42.8	42.8	0%	6.2	86% NC_018024_2
Homo sapiens chromosome 3, alternate assembly HuRef	42.8	42.8	0%	6.2	86% AC_000135_1
Homo sapiens chromosome 13, alternate assembly HuRef	42.8	42.8	0%	6.2	86% AC_000145_1
Homo sapiens chromosome 3, GRCh38 Primary Assembly	42.8	42.8	0%	6.2	86% NC_00003_12
Homo sapiens chromosome 13, GRCh38 Primary Assembly	42.8	42.8	0%	6.2	86% NC_00003_11
Homo sapiens chromosome 6 genomic scaffold, GRCh38 alternate locus group ALT_REF_LOCI_1 HSCHB6_1 CTG8	42.8	42.8	0%	6.2	86% NT_07554_1

Alignments

Download GenBank Graphics

PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X6, mRNA

Sequence ID: ref|XM_008713087_1 Length: 4148 Number of Matches: 1

Range 1: 2108 to 2143

Score	Expect	Identities	Gaps	Strand
42.8 bits(46)	6.2	31/36 (86%)	0/36 (0%)	Plus/Plus

Query 883 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 918

Subject 2108 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 2143

Related Information

Gene - associated gene details
Map Viewer - aligned genomic context

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:50 PM ENG Monday 12/1/2014

(iii)

blast.ncbi.nlm.nih.gov/Blast.cgi

Alignments

Download GenBank Graphics

PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X6, mRNA

Sequence ID: ref|XM_008713087_1 Length: 4148 Number of Matches: 1

Range 1: 2108 to 2143

Score	Expect	Identities	Gaps	Strand
42.8 bits(46)	6.2	31/36 (86%)	0/36 (0%)	Plus/Plus

Query 883 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 918

Subject 2108 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 2143

Related Information

Gene - associated gene details
Map Viewer - aligned genomic context

Download GenBank Graphics

PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X5, mRNA

Sequence ID: ref|XM_008713088_1 Length: 3888 Number of Matches: 1

Range 1: 1852 to 1887

Score	Expect	Identities	Gaps	Strand
42.8 bits(46)	6.2	31/36 (86%)	0/36 (0%)	Plus/Plus

Query 883 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 918

Subject 1852 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 1887

Related Information

Gene - associated gene details
Map Viewer - aligned genomic context

Download GenBank Graphics

PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X4, mRNA

Sequence ID: ref|XM_005265024_2 Length: 3942 Number of Matches: 1

Range 1: 1906 to 1941

Score	Expect	Identities	Gaps	Strand
42.8 bits(46)	6.2	31/36 (86%)	0/36 (0%)	Plus/Plus

Query 886 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 918

Subject 1906 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 1941

Related Information

Gene - associated gene details
Map Viewer - aligned genomic context

Download GenBank Graphics

PREDICTED: Homo sapiens THUMP domain containing 3 (THUMPD3), transcript variant X3, mRNA

Sequence ID: ref|XM_005265023_2 Length: 3916 Number of Matches: 1

Range 1: 1880 to 1915

Score	Expect	Identities	Gaps	Strand
42.8 bits(46)	6.2	31/36 (86%)	0/36 (0%)	Plus/Plus

Query 886 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 918

Subject 1880 AGAGGAGAAGAAATTTAACCTGCTCTGTGCGAC 1915

Related Information

Gene - associated gene details
Map Viewer - aligned genomic context

Desktop Libraries blesson andrews v... Computer Network Control Panel 6:50 PM ENG Monday 12/1/2014

2.4 Random Protein Sequence of length 1000

DGGLGSFILTRLFDKKAMGMFMQNPSNATSIYTSDEKVYLYGSLASRDILESLCLSVAEIPKAR
PDAVGDVETGTAGQPIGWMDADEDMSDQTIEACALGFPM DLLNEEHLRC DMLILSFELKDEL
KMPITVKEVTVPRLIVGALVRAAAIKMDFLIDKRHAEAWGIKIEKPLFENNSVLDSSINAERSLG
CVAKDIGNVFEDQWPVPKAQWLTEEDSREP GGVKIVL LDDAHGKSSAKQLERLPGQPLRL
DGSIYVVAPKQNNFLGRPTIQSRQSGEEEADTSDQLEVYDIAKRYGLPDLETIHIPQIACVIV
VKSQHRLMEEDESKFVRPYDSTALLADTQFGEDGSYVQLCCREVIEKPVFILD RNAAVWLNI
KSVQTYRVFPRRIIASNHTLEGFKA VYLPLVHRGQKT DASKRANNTLSHDTGGRV PRAGHE
WSLHFGLIIKAFEPYFSYHMGIRSM MVSQYKV KQPELKQARKEKRPCDFSRIVDATVKFAENL
NGEINRKPKLMKGFGWAMGYRLVAYETRNAPNSFQQVVMLRNAKGLLTSLSRDPINLFNKV
LAFCGQRIGMGHVIKYQESPTFGFAYNTRDDVRSEGMFH DGS LRNL LPQGNPAITNNIL
TDTEKGHEVVSNPLVPVVGKGQAIHQELHETLIK YHALGWYT MGAECTPGYGAIGTYEGHKS
DDHGTARYFVASDTQKP KLVFESGSIDNGITINVPPAPYNYMTMDVRLKFSKG VYRLLVDA
HSLEPATTLGAEVPSLAAQEYESKFPLGLEFSELQLEIKNAFTNPIAM LGAYVGSTQMFLV SLSA
LTIVEATKPEEHTGATSDEKQFILRRAARRSSANA IKYAATGHSKAEPGVTRADDQLANDFV
KQLFADKVSWPSHLECR LGDLQYIAGTKV TSLKLSLSIIAGLRVNSLEEVVILSVNGWPARTID
LQASEMLPLRVNLKRN RQCTNFGIGLCTENAEIRTFIDACVLAITALEFATQKQRAIW

(Generated using web.expasy.org/randseq)

Input

The screenshot shows the NCBI BLAST search interface. In the 'Enter Query Sequence' field, a protein sequence is pasted:

```
SSDIDETTINVPAPVYNTIVLFLFQGVLLVDDHLEPTTLGQWPSLAAQESEPPGLER
SLLCQHLSKSIAPDPTTCCCGPQCPQKAPGVSHPHCKLSDCQVQzdtYTVTSLSLZ224LAVII
SLEEVVZLSVHSIPARTIDQ428ILPLVILKRIKQCTHIF028LCTB14E2NTFD40VL2TALEFTQ
HQKQDQ
```

Below the sequence, there are fields for 'Job Title' and 'Job Description'. Under 'Choose Search Set', the 'Database' is set to 'Non-redundant protein sequences (nr)'. The 'Program Selection' section shows 'blastp (protein-protein BLAST)' selected. At the bottom, the search button is labeled 'BLAST'.

Output

(i)

The screenshot shows the NCBI BLAST search results page. The query details are as follows:

- RID: 7SVDFPTN015 (Expires on 12-02 20:31 pm)
- Query ID: Icl42587
- Description: seq1
- Molecule type: amino acid
- Query Length: 1000

The database used is 'nr' (All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects). The program used is 'BLASTP 2.2.30+'.

The main content area displays a 'Graphic Summary' showing the distribution of blast hits on the query sequence. A color key for alignment scores is provided, ranging from <40 (black) to >200 (red).

(ii)

2.5 String formed by concatenating names of 10 friends

RONYNIRANJANMITHUNSANDEEPNITHINAMALASHBINARAJMANUGOUTHAM

Input

The screenshot shows the NCBI BLAST search interface. In the 'Enter Query Sequence' field, the string 'RONYNIRANJANMITHUNSANDEEPNITHINAMALASHBINARAJMANUGOUTHAM' is entered. The 'Database' dropdown is set to 'UniProtKB/Swiss-Prot(swissprot)'. Under 'Organism', 'human (taxid:9606)' is selected. The 'Algorithm' section has 'blastp (protein-protein BLAST)' selected. The search button at the bottom left is labeled 'BLAST'.

Output

(i)

The screenshot shows the NCBI BLAST search results for the query sequence. The 'seq1' section displays the sequence information: RID 7SW3DPM1E015, Query ID lcl|88697, Description seq1, Molecule type amino acid, and Query Length 56. The 'Database Name' is swissprot, Description Non-redundant UniProtKB/SwissProt sequences, and Program BLASTP 2.2.30+. The 'Graphic Summary' section shows a color key for alignment scores from <40 (black) to >=200 (red). A distribution plot indicates no putative conserved domains have been detected. The 'Descriptions' section lists sequences producing significant alignments.

(ii)

blast.ncbi.nlm.nih.gov/Blast.cgi

Descriptions

Sequences producing significant alignments:

Select All None Selected 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
RecName: Full=Gamma-glutamyltransferase 5; Short=GGT 5; AltName: Full=Gamma-glutamyl transpeptidase-related enzyme; Short=GGT-rel; AltName: Full=Gamma-glutamyltransferase-like activity 1; AltName: Full=Gamma-glutamyltranspeptidase 5; AltName: Full=Glutathione hydrolase 5; AltName: Full=Leukotriene-C4 hydrolase; Contains: RecName: Full=Gamma-glutamyltransferase 5 heavy chain; Contains: RecName: Full=Gamma-glutamyltransferase 5 light chain; Flags: Precursor [Homo sapiens]	23.5	23.5	48%	9.3	34%	P36269.2	

Alignments

Download GenPept Graphics Next Previous Descriptions

RecName: Full=Gamma-glutamyltransferase 5; Short=GGT 5; AltName: Full=Gamma-glutamyl transpeptidase-related enzyme; Short=GGT-rel; AltName: Full=Gamma-glutamyltransferase-like activity 1; AltName: Full=Gamma-glutamyltranspeptidase 5; AltName: Full=Glutathione hydrolase 5; AltName: Full=Leukotriene-C4 hydrolase; Contains: RecName: Full=Gamma-glutamyltransferase 5 heavy chain; Contains: RecName: Full=Gamma-glutamyltransferase 5 light chain; Flags: Precursor [Homo sapiens]

Sequence ID: sp|P36269.2|GGT5_HUMAN Length: 586 Number of Matches: 1

Range 1: 510 to 538 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
23.5 bits(49)	9.3	Compositional matrix adjust.	10/29(34%)	15/29(51%)	2/29(6%)

Query 3 NYNIRANANWTHNSAN- DEEPNITH 29
 +--+RA A I H NS + EPN +
 Sbjct 510 GFDLRAIAAIPILHVNSKGCVYEPNFSQ 538

Related Information Gene-associated gene details

Desktop Libraries blesson andrews v... Computer Network Control Panel 7:12 PM Monday 12/1/2014

Control... BioSlid... Bioinf... sads - ... NCBI B... Untile... BIOINF... Micros...

**3. ORF Finder
Open Reading Frames**

3.1 DNA Sequence Of any 3 organisms

Seq1 : Panthera leo voucher ZMA710 haplotype 9 control region

Organism: mitochondrion *Panthera leo* (lion) ; GenBank: EF517785.1

CTTATTCCCCACGAAAAGCAAGTGAAAATCCCCAACCTCCACAGCACAAACGCACAATGT
AAAATAACCAGTCAACTTCTTTCCCCACATACACTGTATCATCGACTGCCCTCCCATG
AATATTAA

Input

Output

Seq2 : Drosophila melanogaster chromosome X

Organism: Drosophila melanogaster (fruit fly) ; GenBank: AE014298.5

GAATTCGTAGAAATGAGCTAACAAATTAAATCATTAAATGCGAGCGGCGAATCCGGAA
ACAGCAACTTCAAACCAGTCACTCTGGCTGAACTAAATGGCCTGATAAAACTCACTGGAATT
AAAGAAAGCCCCAGGAAGTGACAATCTAACAAACAAGACCATAATAAACTTACCTACAAAG
GCCAGAATATTTAATACTTATTATAACAAACATCCTGAGAACTGGACATTCCCAGACAAA
TGGAAAGCACGCTAGCATCTCAATGATTCCAAACCAGGGAAATCACCATTGCTCTAAATT
CATACCGCCCAATCAGCTTACTCTGGTCTTCCAAACTACT

Input

Sequence Manipulation Suite:

ORF Finder

ORF Finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments. ORF Finder supports the entire IUPAC alphabet and several genetic codes.

Paste the text into the text area below. Input limit is 100000 characters.

```
>>>seq2
GAATTCTGTCAGAAATGAGCTAAACAAATTAAATCATTAAATGCAGGGCGAATCCGGA...
ACAGCAACTTCAAACCGACTCTGGCTGAACTAAATGGCCTGATAAAACTCACTGGAAATT...
AAAGAAAAGCCCCAGGAACTGAAACATTAAACAAACAGAACATAATAAACTTACCTACAAG...
GCCGAAATATTTAAATACCTATTATAAACAAACATCTGAGAACCTGGACATTTCCCGAAC...

```

Please check the browser compatibility page before using this program.

Submit | Clear | Reset

- ORFs can begin with: any codon
- Search for ORFs in reading frame 1 on the direct strand
- Only return ORFs that are at least 30 codons long
- Use the standard (1) genetic code

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Wed Jul 30 10:02:54 2014
Valid XHTML 1.0; Valid CSS

Output

Sequence Manipulation Suite - Google Chrome

ORF Finder results

Results for 350 residue sequence "seq2" starting "GAATTCTGTC..."

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 348.

GAATTCTGTCAGAAATGAGCTAAACAAATTAAATCATTAAATGCAGGGCGAATCCGGA...
ACAGCAACTTCAAACCGACTCTGGCTGAACTAAATGGCCTGATAAAACTCACTGGAA...
TTAAAGAAAAGCCCCAGGAACTGACAACTTTAACACAAACGACCATATAAACTTACCTAC...
AAAGGCCAGAAATATTTAAATACTTATTATAAACAAACATCTGAGAACCTGGACATTTCCCG...
ARCAATGGAGGACGACGCTAGCATCTCAATGATTCCRAACCAGGGAAATCACCATTTGCT...
CTAAATTCAACCGCCCAATCAGCTTACTCTCTGGCTTTCCAACAACTA

>Translation of ORF number 1 in reading frame 1 on the direct strand.

E F V R N L K F K S L N A S G E S G N S N F K P V T L A E L N G L I N S I L E L K K A P G T D N I N N K T I I N L P T K A R I Y L I L I Y N N I L R T G H F P N K W K H A S I S M I P K P G K S P F A L N S Y R P I S L S G L S K L

7:55 PM
12/1/2014

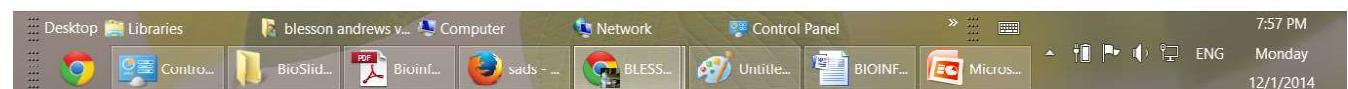
Seq3 : GBE1 Equine BACs from the INRA Horse BAC library

Organism: Equus caballus (horse) ; GenBank: BV725494.1

GATCTTATTCACCTTTAACACATATTTGTCATCAAAATGAGTTTGACCATGCAGATT
AACATTTCTTACAAACACATTATGCTTTGAACCCAAAAGAACTTGTACTTAGGCTATAAAG
ACACACCCCTCCCCAAACTTGCTCTAACACAGCCTCAATGGAGTCCAACACAATATAGAT
AAGTGGAAAGATACTCGGAAACTTCAAAGTCCTGAAAAATAATGACAATTATAAGTAGTA
AGCAAGCATCATCTGCCATTTCAGGCAGACAGGCAAAACAGGAGCAAGTATCCTAAA
ATAGTAAAATATGGTGGGTGCAGCCAGGGCTTGGGTCCAACAATTACGTTCTAATCCC
AGTCCTGCCAATTACCAACTTGTGTCTGCGAGCAAGTTAGTGAATTCCCTCGAGGCTCAGT
TTCTCGTTGGTAAAACAGAGGTACTGTATCAGAGGCTGCTGTGAGGACGGACTGAAGCAA
TGCACAGAGGGCTTGACACCGTGCAAACACTGGGACACGTCTCAAATTATTAGCTGCTG
TTTATTGTCATTATAGACTTGGACCATTAAACTACTGTTTAAGTTATGAAAAATAAAA
GTGGGTCAATTATTCTCTTACGTTTACACCCAACTTGGACTCTTCTTAATGACCTTATT
TCCATTCCCTCTCCCCACTGTCTTGTCTCATTAGAACATTTCTGTTATATA
AAAACTTAACATTAAAGCTTAATTAGCACA
CCCAACTCTTTGAGGAC

Input

Output



3.2 Random DNA Sequence of Length 100

CGCCTCGGAATACGGTATGAGCAGGCGCCTCGTGAGACCATTGCGAATACCAGGTATCG
TGTAAGTAGCGAAGGCCGTACGCGAGATAAACTGCTAGGA

Input

Sequence Manipulation Suite: ORF Finder

ORF Finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments. ORF Finder supports the entire IUPAC alphabet and several genetic codes.

Paste the text into the text area below. Input limit is 100000 characters.

```
>seq1
CGCCTCGGAATACGGTATGAGCAGGCCCTCGTGAGACCATTGCGAATACCAAGGTATCGTGTAA
GTAAGTAGCGAAGGCCCGTACGCGAGATAAACTGCTAGGA
```

Please check the browser compatibility page before using this program.

- ORFs can begin with:
- Search for ORFs in reading frame on the strand.
- Only return ORFs that are at least codons long.
- Use the genetic code.

*This page requires JavaScript. See [browser compatibility](#).
*You can mirror this page or use it off-line.

new window | home | citation

Wed Jul 30 16:02:54 2014
Valid XHTML 1.0; Valid CSS

8:02 PM
12/1/2014

Output

Sequence Manipulation Suite - Google Chrome

ORF Finder results

Results for 100 residue sequence "seq1" starting "CGCCTCGGA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 90.
CGCCTCGGAATACGGTATGAGCAGGCCCTCGTGAGACCATTGCGAATACCAAGGTATCGTGTAA GTAAGTAGCGAAGGCCCGTACGCGAGATAAA

>Translation of ORF number 1 in reading frame 1 on the direct strand.
RLGIRYEQAPRETIANTRYRVSSEGPyAR*

8:04 PM
12/1/2014

3.3 Random DNA Sequence of length 500

AACCGCGTCTCTACGACC GG TGCTCGATTTAATT CGCTGACGTGATGACATTCCAGGCA
GTGCGTCTGCTGCCGGGTC CCTCTCGTGATTGGGTAGTTGGACATGCCCTGAAAGACA
TAGCAAGAGCCTGCCTCTATTGATGTCACGGCGAATGTCGGGGAGACAGCAGCGGCT
GCAGACATTACATCGGAGTAACACTAAGGTGGGATAACTCCGTAACTGACTACGCCTTC
TCTAGACCTTACTTGACCA GAGATA CACTGTCTTGACACGTTGATGGATTAGAGCAATCAC
ATCCAAGACTGGCTATGCACGAAGCAACTCTTGAGTGTAAAATGTTGTCTCCTGTATT
GGGATGCGGGTACTAGATGACTGCAGGGACTCCGACGTTAAGTACATTACCCGTCA
GGCGGCGTT CAGGATCACGTTACCGCCATATGATGCGAGCATGACATCATCTCCGCTGT
GCCCACCCCAGTAGTGATTATTCC

3.3 Random DNA Sequence of length 500

AACCGCGTCTCTACGACC GG TGCTCGATTTAATT CGCTGACGTGATGACATTCCAGGCA
GTGCGTCTGCTGCCGGGTC CCTCTCGTGATTGGTAGTTGGACATGCCCTGAAAGACA
TAGCAAGAGCCTGCCTCTATTGATGTCACGGCGAATGTCGGGGAGACAGCAGCGGCT
GCAGACATTACATCGGAGTAACACTAAGGTGGGATAACTCCGTAACTGACTACGCCTTC
TCTAGACCTTACTTGACCA GAGATA CACTGTCTTGACACGTTGATGGATTAGAGCAATCAC
ATCCAAGACTGGCTATGCACGAAGCAACTCTTGAGTGTAAAATGTTGTCTCCTGTATT
GGGATGCGGGTACTAGATGACTGCAGGGACTCCGACGTTAAGTACATTACCCGTCA
GGCGGCGTTCA GGATCACGTTACCGCCATATGATGCGAGCATGACATCATCTCCGCTGT
GCCCA CCCCAGTAGTGATTATTCC

Input

SMS

Sequence Manipulation Suite:

ORF Finder

ORF Finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments. ORF Finder supports the entire IUPAC alphabet and several genetic codes.

Paste the text into the text area below. Input limit is 100000 characters.

```
>seq1
AACC CGCG TCTAC GACCG GTGCTCGATTAA TTTCGCTGACGTGATGACATTCCAGGCAG
TGCG GTCTGCTGCGGGTCCCCTCGTGAATTGGTAGTTGGACATGCCCTGAAAGACATAG
CAAG AGCGCTGCCCTCTATTGATGTCACGGCGAATGTCGGGGAGACAGCAGCGCGTCAGA
CATTACATCGGGAGTAAACACTAAAGGTTGGATAACCTCGTAACGTGACTACGCCCTTTCCTAGA
CTT TACATTGACCAAGATACACTGCTTTGACAGTTGATGATTAGAGCAATCACATCCAAG
```

Please check the browser compatibility page before using this program.

Submit | Clear | Reset

- ORFs can begin with: any codon
- Search for ORFs in reading frame 1 on the direct strand.
- Only return ORFs that are at least 30 codons long.
- Use the standard (1) genetic code.

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Wed Jul 30 18:02:54 2014
Valid XHTML 1.0; Valid CSS

8:05 PM
Monday
12/1/2014

Output

Sequence Manipulation Suite - Google Chrome

ORF Finder results

Results for 500 residue sequence "seq1" starting "AACCGCGTCT..."

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 144.
AACCGCGTCTACGACCGGTGCTCGATTAA TTTCGCTGACGTGATGACATTCCAGGCAG
TGCG GTCTGCTGCGGGTCCCCTCGTGAATTGGTAGTTGGACATGCCCTGAAAGACATAG
AGCAAGAGCGCTGCCCTCTATTGATGTCACGGCGAATGTCGGGGAGACAGCAGCGCGTCAGA
CTT TACATTGACCAAGATACACTGCTTTGACAGTTGATGATTAGAGCAATCACATCCAAG

>Translation of ORF number 1 in reading frame 1 on the direct strand.
NRVTTGAREFNADVMTFQAVRLLPGPSRDWVVGHALERHSKSLPLY*

>ORF number 2 in reading frame 1 on the direct strand extends from base 145 to base 243.
TGTACCGGCATGTGGGGAGACAGCAGCGGCTGCAGACATTACATCGGAGTAAACACTA
AGGTGGGATAACTCCGTAACTGACTACGCCCTTCTCTAG

>Translation of ORF number 2 in reading frame 1 on the direct strand.
CHGECRGDSSGCRHYIGVTLRWDNSVTDYAFL*

8:06 PM
Monday
12/1/2014

3.2.3 Random Sequence Of length 1000

TATAACCCTCTGAGTGTCCGGAGGC GGAAATCCGCCACGAATGAGAATGTATTCCCC
GACAATCATAATGGGGCGCTCCTAACAGCTTTCACTGGTGGCCGGCTAGGCCTCTC
TGCCC GGAGTT CGGCGCACTGCTGCCACAGCCGGCATTGTTAGGGCGTTATTC
GAGGGCACTCGGAGCTAACTTGTCGGGACCAGCCGGGTAGTCATCGGGCTTACAG
CGAAAAGCCCAGGACC CGGCTCCACGCTATGGAACGTCTTAGCTCCGGCAAGCAATTA
AGAACAA CGCAAGCATCGCGGATATAAACAGAGAAACGGCCGAATACACCTGTTCGTAT
CGTATCGGTAATAGCCTCGGGAGCCATGTGCCATACTGGTCTCGGGAGCACTCTGGT
TATGCATATGGTCCACAGGACACTCGTCGCTCCGGTATCGCCTCTATGTGACGGTCT
TTAGGCGCACTAATGCTCAGCACCAATTAAACCAAGACCGACACCAGATCTGTAAGGTCC
GCCACGCAGACGACAGCCCACGGAGATCACCGACCGATCTATCTGATCGGCGACCATT
GTGTGGTACTGGGGCGGAGAGGTA ACTACGGTGCCGCTAACAAACCCCTGT CGTCGC
TGACGTTGTAGTCTAGTCATTATGATTGTACGCTATT CAGGGATTGACTGATACCGG
AAGACATCTCAGTTGAAGTGGTCTATACGACAGAGACCGTGACCTACCAAATCTCCTTA
GTGTAAGTTCAGACCAATTGGTAGTTGTCCAGAACTCAGATTAAACAGCAGAGGACGC
ATGCTCTACCTT CATGATCCACTGACGTCCCTGAGGCTGCAATACATGCAACGAGGCAG
TCTCCGGTAAGTCCTAGTGCAATGGCGCTTTTACCCCTCGCCTCGAGAAGAGGGG
ACGCCAGTGCAGATATTTAATGTGGTAATTGGGAGGACTCTGGCCCTCGCC

Input

Sequence Manipulation Suite:
ORF Finder

ORF Finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments. ORF Finder supports the entire IUPAC alphabet and several genetic codes.

Paste the text into the text area below. Input limit is 100000 characters.

```
>seq1
TATAACCCCTTGTGGAGCGGGAAATCCGCCACGAATTGAGAATGTATTTCCCCGA
CAATCATAATGGGGCGCTCTTAAGCTTTCACCTGGTGGCCGGCTAGGGCCCTCTGGCC
CGGAGTTTCGGCGACTCTGGCGACAGCGGGCATGTTTAAAGGGCGTTATTGGAGGGC
ACTCGGAGCTAACATTGGGGACCCAGCGGGTAGTCATCGGGCTTATACGGGAAAGGCC
CAGGAGCCCAGCTTACAGCTATGGCTCTGGCTCCGGCAAGCAAGCAATTAAAGACACAGCA
```

Please check the browser compatibility page before using this program.

Submit | Clear | Reset

- ORFs can begin with: any codon
- Search for ORFs in reading frame 1 on the direct strand.
- Only return ORFs that are at least 30 codons long.
- Use the standard (1) genetic code.

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Wed Jul 30 18:02:54 2014
Valid XHTML 1.0; Valid CSS

8:07 PM ENG Monday 12/1/2014

Output

Sequence Manipulation Suite - Google Chrome

ORF Finder results

Results for 1000 residue sequence "seq1" starting "TATAACCCCTT"

>ORF number 1 in reading frame 1 on the direct strand extends from base 196 to base 321.
CTTGTGGGACCAGCGGGGTAGTCATGGGGCTTACAGCGAAAAGGCCAGGACCGC
TCCACGCTATGGAAACGCTTTAGCTCCGGCAAGCAATTAAAGAACACGCAAGCATCGGG
ATATAA

>Translation of ORF number 1 in reading frame 1 on the direct strand.
LVGTSRGSHRAYTAKSPGPGSTLWNVFSKGQLRTTQASRI*

>ORF number 2 in reading frame 1 on the direct strand extends from base 322 to base 465.
ACAGAGAAAACGGCCGAAATACACCTGTCTGTATCGTAAAGCTCGCTCGGAGCCA
TGAGCCATACTGGCTGGGGAGCACTGGTTATGCATATGGTCCACAGGACACTCGTCG
CTTCGGGTATGGCTCTATGTGA

>Translation of ORF number 2 in reading frame 1 on the direct strand.
TEKRPNTPVIRIVSVNLSAEPCAILVCGALWLCIWSSTGHSSLPGMRSM*

>ORF number 3 in reading frame 1 on the direct strand extends from base 577 to base 675.
TCGGCGACCATTGGTGGTACTGGGGGGAGAGGTAAACTACGGTGCCTGCTAACACCCCC
TCGGTGGTGCCTGACGTTGTAGTCTAGTCTCATTTATGGA

>Translation of ORF number 3 in reading frame 1 on the direct strand.
SATTCVVLGRGRNYGAANNPSVVALVCSLVSL*

>ORF number 4 in reading frame 1 on the direct strand extends from base 862 to base 975.
GGCTGCAATACATGCAACCGAGGCCACTCTCCGGCTAAAGTCTTAGTGCATGGCGTTTTT
TACCCCTCGTCTCGAGAAGAGGGGACGCCAGTCAGATCTTAAATGTGGTAA

>Translation of ORF number 4 in reading frame 1 on the direct strand.
GCNTCNEAVSAVSPSAMALFYPRPREEGTPVQISLMW*

8:07 PM ENG Monday 12/1/2014

4. GenScan

4.1 Human DNA

Seq2 : Homo sapiens 12 BAC CTD-2325H11 (Cal Tech Human BAC Library D)

Organism: Homo Sapien ; GenBank: AC139763.4

AACCCTTCCATCCGAGGTAAAAAGACACATGCTCCAGTATGCTTTCTGGTAATAAA
 GGCAATATTGACCCCTCCATCTGCTCATTCTTGCTTATTGGCTCAGAACCT
 ATGCAAGGAAAAGATGTTATGTATTGACCCCTCAGATTCCAAGAGCTGTGTTGCAGAA
 AGAGGATTGCTTCTGTGTCTGGCTGGGATGGGGATGGAACAAAGGAAGTGAAGTCCCAG
 TGGAATCTGTTCTTGATGAAC TGCAACTCACTTCTCCAAGACTGAATTGCAGAGG
 CCTGCTCAGAGGCTGGAGGAAGGGATGGGATGACTTCTGGGAGTGGGAAGGTCTCC
 CCTAAACAACATACTCCTGGACAGACCTATCACGGGAACCTAACCAAATAGGTGGAGAA
 GCCAAGACTATTCA GTGGCCTTATACCCACTTTAAATTCTGGAAGACC ATCACCCCC
 ACCTGATTCTCATCACTTTGGGCGTGTGATCTATTCTGTGACTCCTGGAATAAAAGACCA
 ACACCAAGAATGTGAAGTTCAGTGCAATTATTGAGAAAAGAAAGTGCAACATACTCATCA
 CCAGAGGATCTACAGCTCAGGGATTAGGGACAGAAGGGAGTCAGGATCTGAAGCCC
 AGATGTAGGGAGGGTCCCTGGGAACCAGCAGTCTGGAGATCTGGCCGTGCACCCCTC
 CAGCACAGAACCCATGGCTCTGCCCTCCTCAGACAGTGCTGAGATGACTGGCTTCA
 AATCTCCA ACTCCCAGTAAGAAAGGGAGATGGGCCAGAGTGGAACAGGCCAGGCTGT
 GAAGGCCAAACTGGCTGGCTCAGGGTTGCTGCCAAAGTAAGGGGCCATTGCAG
 GAAGGTTCCAGGGGCATCATCATCCCAGACCAGCAGCAGGACCTCCATGGCCCTGG
 AAGGTATGGGATGGAGAAACCAGGAGTTCA GCTTCTCCAGGGAACTTGAGGAAA
 AATGTGGTTGACCC TTATGCATCTATTGATGCCAAGCAGAGAGTGGAACACTATTGCA
 GGCTGAGTGGGAAGCAGGTGGTTAGAGATTGGAACAGTAGATCACTGGTGGAGCT
 ATGCTGGCTTGAGCTCGTGGCTGGGAGAACGGGATACTGGTGCTGCCCTCACCACCA
 GACTTGAGCCACTCCTCCAGAACAGTCTGGATGCTGCCAGAGCTGGAGCCCATTCCACT
 GTGGCTCACAGAGATGCTACCTGCACCGCCGAGCCTGCTCCACTACTGCTGCTCTGGT
 AGCTCCGCTGCTACCCCTGCCCTCAGTGCTGCCACTGCTGACCCCTCCATGCCACTG
 CCGCTGCCGCCACTGACTCCATAGCCGCTGCTGCTGCTGCTGCCGCCCTTGT
 GCCACCACTGCCACTGCCACTGACCCCTCCAAACTCCACGGCTACTGCCAGAGCTGC
 CACTTGCTGGT GACATTGCTGACACTGCTACAAGACAACAGGCACACAATTAGGC
 AAAGGGCTGGAAACTGGAAAGCCAAGGACTGTATTACATCCCCACCCCTGCTGGCCAT
 GTGACTTCGAGAAAGATGCTAACTTATCTAATGCCCTCAGTTGCAAACCATGAGAGTGAG
 CCTGGCCTCTAGTGCCTCTCCAAGAACAGGAACCAAGGGCTTAGCTTGGCCTGAGAGGTTGTC
 TTCCAACCTAGGCCACTCCCCAAACCAGAGGGCTTAGCTTGGCCTGAGAGGTTGTC
 AGCTTACACTAGGGCAGAGCATGTGCCATGGCCTGAGATCCTAGACGCTGTCAGGT
 GGCCTGACTCCTGGATGGAACAGGCCTTCTGCCTGAAGGTCAGTCAGTGAGGTTGT
 TTCTAGCTGCCCTTTACAACCCAGGGAGAGGGTTTTGAATGGCTACTTACAAATGC
 ACACGGCACTCTGACATTCTCCAGACATCCTGAAAAGGAAGAGGAGAGGTGAATTCTTC
 TGTTATGAGTCAATAACCAAAGAGTTGAGAGTTCAACACTTGGGAGGCCACGCCCTCCCA
 TTGCCCCACTAGGGTCAAGCTCCTGCAAGAAGAAAGCCTATTCCAAGTCTGTTCCAT
 GACCATATCTCCCCAGAAAAGGGGCTGCTCAAGAGTTGAGAATTCCCACACTTTCCC
 CTTAGCCCTAGAGATTATCTCTAGGCAAACACTTGATGCCAGCTGCTGCCCTGATT
 GGCCTGCCCCACTTCTCATCCCCCTGCTAAGACCATCTTCCCACCTTAGCATGCCCTTC
 ATTCTCTCTACCCAGCCCTGGATTTTATTCCAATTCTGCCCTTCCAAGAAGGCTTCC
 ACAATGGGCCCTCTGGGTGGT GACATGATTGGGTTGTCAGGGTGGAGAAAAT

ATGGGAGGGGCTGAAACTCAATGCAGGATGGATTGAAATGTACCACCTACACTGTCCA
TCTCCTGTGTCATTCTGTTAATGGAGCAGGCACATCCACATATCTGAGAAGATATGG
ATTGGGCTGCATATCTGCCAATGTGCCCCACTGGTGTGACCATCCCTCCTCATCC
CATGACTCCTAGTACACTGACATCACCTCACTCTCACTTCCTCTCATACTGGGTT
TGCTTCCTCCATCAGATTGGGAGCTACCCAAGGACAGAGGGTAGCTACTCCCTGAGCA
CATGATGCATTGGCAGAAACTCAGCAGAAAGATCTCAATTCTGACACTCCTTAGC
ATTGATAAACACTGCTAAGACTTATCTGCTGCTTAATGCTGTATTATTTCCAAGAGCA
TGGATGTATCAACCGTCTCCCCATATGGACTCAGCTTCAGGGTGGAGCTGTAGGACT
GTCTATATTGAGCTTCCAGTATGGTAGCCAGTAGCCACATGATGCTATTAAAGTTAATT
AAATTAAATTAAATGTAAAAGTGTGTTCTCGGTTGCAGTAGCCACATTCCAGGGCTCAA
GTGCCAGTGGACTGTATTGGACAGCACAAGAATAGAAAATTTCATCATCTCAAAAA
GTTCCAGTGGACAGCACTGGTCTATCAGTGTCTAATGCCAAGAACCTTCCTCATACA
TTGCACATAGTGAAGATCAATAGAGATATGTAACATGATATTGATGTCCATTAAACATTG
AGAGTTGGATGATCTTGATGGACATAAAAGTGGGGCTGAACCCTGACTCTGTCTT
CAAGACACTCCTGCCCTTGTCTTAGTGTCCCTCTCCTTATCATCTTGCCCCCTTCCAC
ATTATCAAAGGTCACTCTAAAACCATATCTGGAGAATGGTGTCCCATTGGCCCTCACCT
GCACTCCTCTCCCTCAGCAGCTGCGGTAGGTGGCAATCTCCACATCCAGGGCCAGCT
TGACGTTCATCAGCTCTGGTAGTCACGCAGGAGCCAGGTGATCCTTAGCCTTC
TGTAGGGCAGTCTGCAAGTCTGGAGCTGGCATTGGCGTCTTGAGGGCCATCTCTCC
ACGCTGCTCAGCCTCTGCAATTGCCGCTGCAAGGTTGGCATTCTGAGGTAGAAAATCAAT
TAGCATAGTGAACCGGAAACAAACACTTGATCCTACTAGGAGTAGATCTCTCCCCAC
ATCCTGCATCCCTCCTCCCTTCTAGAAATCCAATGGAGCATGCTCAGAGAAAAGA
AGTCAGTTTGCAAGTTGTTAATTCATCTTGCTCCATTGAAATCAACTGCGGCTCATCA
AGAAAGGGAGCACTAATTGCTTCTAGTACAGGTAGTAGTGGTCTGGTTAGTC
CCTAATCTTTTGCAAGACTGTTAAAAGTCTCTGCTTAATCACTTCAATTCCAATGATAC
AGAGCAGCCATCAAGGTTATGTGAAAGGTGTGAGAAACAGCTTCACTAAGGGCAGATCC
TGCTGACATCTGAGTGTAAAGTCCAGAAGTGCAGTTGCAAGATTGTTAGGCAAAGTAG
AATAAACAAACATCTTCTTGAGTGTGAGCTTCCAAGAAGTTCAACCTCCAACAAAACATC
TGGGAGAACAGAGTTCCCAGACACTGCGCTATCCCACATGTGACACCAGTAAAGGATG
GAGAGAGCATGTAGCAGAGGAGCAGAGCTTATGCTGCTCTCCGGGGCCAGAAGA
CCTCTGGATCTCCGTTAAACCCAGCCCCACCTGCTTCTGACATTTCATCTCAG
CCGTAGCCTCTGGATCATCCTGTTGAGCTCATGATCTCACTCTGGTGTCTCAGGTC
ATCCCCATGCCTGCCAGCTGTGGCTGCAGCTCCCCAAGCTGACAGAGGAAAACCAAT
CAACGTGGCTTTGCAAAGTGTGGCACATTCTCAATGAAGGTGATCAGGATCCAAGGAC
ACAGGGAGGGTCCCACCTTGGTCTGGTACAGGGCCTCAGCTCAGACTGCTCCTCTG
GGCAATCTCCTCATACTGGCGCGACCTCGCAATGATGCTGCCAGGTCCAGGCAG
CGGTTGTTGTCATGGACAGAACACAGACGTGTCAGTGGCATGGCTTGACATTGG
CAGCTCCTGCAGGAAACATGGATAGTTACTCTTACCATCGCCTGGACAGGGATGCTG
GGAATCCTCTACTAGCGGAGATATCCAAGGATGGGGTAGAGGTGGATCAGTTGGAAA
ATAAGCTCACATTAGGCCTGGCTACAGCCTAAAGGATGTTAGACAACAAGGCTAAAG
GAAATATGCAAAATATGCAATATAATTAGGCTATCAGCCTGAAAGTATCAGCCAATGAGTC
TTATTGTGTAAGTTCCATGGGCCCCCAGCAGGGAGGAGTATTATGAGAAGCCAAT
GACATGGGTATAGAAGCCAGAGAGAAGGTGAGAGCAAATGGCCTGAGGACCTGTGG
GCCAGGTCTATGCAGACTGAAGAAGCTCAGTTAACATTCCCAGGTGGGGTGGAGCTGA
GTCCTGAGCCCCATTGCTGGTTGTGAGTGTGCTGTTGGCTGCTCCCTAAAGTTA
TGCTTGGGAAGAATCAGACAGGAAGGTTCTGCAGCTGCAAACCAAGGAATACCCATCT
CATAGAGGGCTCAGGAAGCTGACTTCATCTGTCAGGCTGCCACTTGGCCTGCAGC
TCCACCTGTTCATGAAAGCCGCATCCACATCCTGCAGAGGGAGGTAGAAACCCAGAGA
ATGACCCCTCTGTTCCCACCCAGTGGCTTTCTAAGCCCTCATGTCTCGTCTT

GTTCCGGCCCCAGCTCCTCCTAATTGTGACTTCTCATCTTCCCTTAAAATAAGGAAG
ATCAAAATGGCTTCTCCTTCCCTGCTGACATATCAAGCCCCAATACTTAACAAAAT
ACAAACCATAAGGCACAGAGTAACCATCTGATTTAGGGTAGAAAAATCATATTACCTAA
TATGTATGTGACTTGGTTGTATAGTACCAAGAAATCCCACCCTCCTGAACAGAAATTAGG
GTATTCTCACTACTTCACCTTACGATGACTGGTTAAATATGCTACAAGTATTGTTGATT
GTTATTAAAGCATCTGGAGTACATTGCAGGCTGAAATCTGATTAATCTTACACATAA
AGGGCTCATTATTAGTCGCATTGCTAACAGCTCATTAGTCACAGTAACATTAAACATA
CCTTCCATGGTAATATACTTAGAGCAGCTCCTCATGAAACATTCACTTCACTTCAAAATG
TAAGGAAAAGTGTCAATCCCACACACAATATAAGGTCCCTGTGCCCGAGAACAGCTGGC
TGTGCACTCTCCATGGGGCAGCTCCCAGAAAGGAAACCTGGATATGCTGGTCCAGG
CTCACCTACCTCTTGAGCCCCACAAACTCATTCTCTGCAGTGCCTTGTGATTTC
ATCTTCATACCTGGAACAAGAAGAGGCACACATTGAACACTGGCATTTCAGAAGTAAA
AAGAGAGGTGAGGATTCCAGGCACTCTCCTCACCTGCGCAGAGACAGGACGAGCTCAT
GGGCACGTGCAGTTACACAGGACCCTGGCTCAGGGCCTCACACTTAGCTCAATGCTG
GCACCATATTGACATTCTTCTTTTTTTCTGAGACAAGGTTCACCCCTGTAC
CTAGGCTGGAGTGCAGTGGCACAAACATGGCTCACTGTGGCCTCAACCCCTCAGGCTCA
AGCAATCCTCCTGCCTCAGACTCCAAATAGCTGGACTGGAGGCATGCACCACACGT
CCAGCCAATTNTTAATTNTAGTAGAGACAAGGTCTCACTTGTGCCCAGGCTGGCTC
GAACGTGTTGGGCTCAAGTGATTCTCCCTCCTCAGCCTCCAAAGTACTGGGATTACAGG
CATGATTGAAATTCTTAATAATTGTAACTAGGTTCTGCATTATCATTGCACTGAGGC
CTGAAAATCACGTAGCTGGCCTGTCCAGAGACCTCTCAGAGCTCAGGGCACTGCCTGG
CCCTCTCCCTCAAATGCAGTGATGGGAGAGGTATGGGAGTCAGTGAGAGGGCAA
CTTCAAGGAGCCTGACCACTGCCTAGGCTCTTGATTGACTTGCACAAAGCACAATGTG
GCATATGGAACCCAGTGAAATGTGTGCTCTCATTGGCATTCTGCCTATTCCCTCAGTT
TCTGTTTCCTCATTACAGTTACTTCTTATAGGAATTTCAGTTAAGAAAGAATTCAACCT
TCTGAATGCATCAAAGTGTGGATTCTGCCTGTTAAGGACTAAAATCTACTACCAAGA
GTTTGTATTCATAAAGCTTGAGGAATTGGACTGATTCTAGGAGAAAAGAGGTCTTAGG
CAGGACCTCTTTCTCTAGGAGAAAAGAGGTCTCTGGGATCTTTGGGAATATG
AATCTGAGTTGCCACTGAATTCCCAGAGCCCAGCACAGAGCCTGCCACAGAGTGGTGC
CCACCACGTTGTGGTGAATTGGAATAGGGCCTCTGCAGAGCAGGTGTGCCGTCC
CTGTGGGTTCTGATCCCTGGGAAGTCGGAGCTCCCTCCACCTTGAGAGGCCAGT
TTGCTCCTGCAGACCTCAGCGGATATGGACTGCAACTTGGAACGTGGAGGAGGAGGCA
AACCAAAGCTGCAGAACAGGAGATTGGACGGACACACTGATAAAGGCCTTACCAAGT
GCAGCCCTGGACCTCTGCCTAGAATCCCCAGGAGCTAGGGGAACACTCACACTCCTGG
GGCCTGTCCACACCTATGATTGAGAACATACCTAACAGACTGGAAATACAGGAATACGCATTGTT
GACAAGCATCCAAGGTGATTCTGTGTCAGTCAGAACATGGAGTCTCAGTCCCCACCCCTTACCA
GGATTCTAAACTCTGGAGGACTCTTGTCCCTCTGGGCTGGAATAGGAGGGGCCA
TCTGGAGCCTGAAATCTGTGTCAAGCCACATGGAGTCTCAGTCCCCACCCCTTACCA
CCCTTCTGCTTGTCACTCAATAAATCCATCCCAGAGTATTGACTGAAGGCACA
GTTGGGAACAGTTGAAATCCAAATTAGGACCACACCTCCCTGCTAATTCCCCAGAAGG
GGAAACTCCCAGGAGGGAGCCCTACCTCTTTGAAGTCTTCCACCAAGGTCTGCATGC
TTTCAGCTCTCCCTCCAGGTTCCCCCTCCCTAGAAGTGAATCTAGCTGCTGCATA
GGAAGCTGATGTAGGATTCAAACAAAGGCTCCAGGCTGCTGGGCCCTGAGCCTGTGGT
CTGCTGCTGGAGCAGTCCCACCTGGCTCCAGGACCTGTTCTGCTGTTCCAGGAACC
GCACCTGCATGAAAGAGGGAGAAAAGGAGTCAGCCCTGAGTTGGCATGTGGGC
TGCCCATCCCAAATGTAGAACTGTGGAGAAGGAACCCAGTTCAATCTACAATTCCC
CTGTTTTAATTGGGAAGGATGAAGATTAGCAAAGGGAGAGCTAGAATTGTTGAGT
GATTATGCCTTACCTGTTATTTCATGTATTCCGTACATTAAATCCTCACAAACCACCTAT
AAAGGAGTTGCTGTGATCCCCATTGAGATAAGGAAACTGAGACTCAAAGAGTTAAG

TTTCCCCAAAGTGACAAGTCTACTGAGTGACAAAGCTAAGAACATCAAACCCAGGACTGTCT
GGGTCTATAGGGATCCTCAGAATAACCACTGAAGGAGGTAAAGCACAAATAAGTTGTCA
TTGTTTTTAGATGTAAGCTTTTCAGCTCCATAAAGGCAGTGAATACAGTCATGAGGT
CTAACATTCTTAGAAGTTCACATAGGAAGACCTCCTTCAGTTATTCAAGCTTGCAGCT
GAACTAGAATTCACTTCAGATCCTGGGCTCAGTCCCACACAGACCCAGAACACCCTCCT
CTGCCCGAGCATGGACAGCTCAGTTAATTATGGCTGAAGCACAGCATCATCTGGAAG
GTAAGGCAGCTGCCAATGCACCTCATTAGATGCAGCCTCAAGATCTGTGCTAACACAAG
GTTTGATTACCAAGCTACTTGGATTATCAAGCAATTGCTTTAGGAGAAATTAAAGTAA
CTGCAATAATTCATGACTAAAATGGCTATTCTCAAAAGCTAAGAGCATGTTGGACC
CCCACTTCAATTACCCCTCATCTCATTCTAGCAGGAAACCAGGACCAGGAATTCTCCCC
CCATTCACAGCTGGAAAAAGAAGGCCAAAAGTGAAGGGGCATGATCACTGTCTG
TAGGGACCAAGACAGCACCAGGACAAGAGCCAAGCGCCCCTGCCCTGTCTCACAAG
TCCCCATGGCATCTTCTACACCGACCCCTGGCACCCAAACCCCTCACAGCACCTCT
TGGCCTGCCCTCACCTGCGATGAAGGAGGCAAACCTGTTGAGGGTCTGATCT
GTTCCCGCTCCTGGCCTTACTGCCCCAATCTGGGGGTCGATCTCACATTGAGGGGC
TGCAGGAGACTCTGGTTACAATCACTCCTGAATTCCCCCAGGAAAGCCCCCAGGGCC
AAAGCCACCAAGGACCACCAAAGCCACCAGGCCCAAAACCAACTACCTACTCCTCTGCCACC
CCACCAAGCCACCACCATAGCTGCCCTCCAAAGCCACCTCCATAGCCACCTGCCAAAGCCAC
AGCTGCTCCGCCCTCCCCAAAGCCTCCAGGCCGGAGCTGCCAGCTGCCACGCTGAT
GGAGATGCTCTTGGCTGCCAGGTTGAGAGGCTGCGACTGCCAAAGCTGCCCTGCT
CGCTCCTGAAGCCACAGGCCCTCCACCAGCTCCCCCAGAGCGGGCCACACAGCTCAT
CCTGCTGCTGCCGGAGACCACAGCAGAGCGGCCAGAGAAACCCCTGGCTCTGCCACTG
AAGGATTCTTGCAAACCTGCTGTTAGTGAGAGAGCTGGAGGCAAGCTAGTGATC
ACTTAGCAAGAGCTGAGAGGGATAGGGACAAGAGAGGGAGACTGGCCTTATATAACAGG
GAGTTGGCTTGGCAAAAGGAGAGGGAGCAATTAGTTGAGAATCATCTGGTTGG
TTGCCTCGTAGGCATTAAGTTACTTCTGGTCCAAACACACCTCAGAGATAATCCTCTG
GGCCAGAGTGGCTTGGCTTCTAGCTTAATCACCTAAACTGCCAGTTGACAAATG
CTGGAGCTCCCTCCCTTTGTGCTGCCCTCAAG

Input

The figure shows a screenshot of a web browser window. The address bar contains the URL "genes.mit.edu/GENSCAN.html". The main content area displays the "The GENSCAN Web Server at MIT" page, which is titled "Identification of complete gene structures in genomic DNA". A red box highlights the "Organism" dropdown menu set to "Vitis vinifera". Below it, a message states: "Server update, November, 2009: We've been recently upgrading the GENSCAN webserver hardware, which resulted in some problems in the output of GENSCAN. We apologize for the inconvenience. These output errors were resolved." A red box highlights the "Sequence name (optional)" input field. Another red box highlights the "Upload your DNA sequence file (upper or lower case, spaces/numbers ignored):" section, showing a large DNA sequence text area. At the bottom left, there are "Run GENSCAN" and "Clear Input" buttons, and a link "Back to the top". The bottom status bar shows the system tray with icons for network, control panel, and other applications.

Output

genes.mit.edu/cgi-bin/genscanw_py.cgi

GENSCAN Output

View gene model output: [PS](#) | [PDF](#)

GENSCAN 1.0 Date run: 1-Dec-114 Time: 13:23:39

Sequence /tmp/12_01_14-13:23:39.fasta : 9994 bp : 48.76% C/G : T/A/C/G 9 (43 - 51 C/G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

1.10	PlyA	-	570	565	6		1.85					
1.09	Term	-	1585	1108	398	0	2	92	39	458	0.966	36.54
1.08	Intr	-	1978	1944	35	0	2	163	78	33	0.988	1.67
1.07	Intr	-	3614	3394	221	2	2	67	94	466	0.996	42.26
1.06	Intr	-	4391	4266	126	2	0	50	75	212	0.999	16.98
1.05	Intr	-	4653	4489	165	0	0	74	81	270	0.881	25.06
1.04	Intr	-	5273	5178	96	2	0	87	15	192	0.976	12.81
1.03	Intr	-	6837	5977	61	0	1	117	100	91	0.996	11.94
1.02	Intr	-	7907	7783	218	2	2	112	53	236	0.705	29.71
1.01	Init	-	9686	9087	680	2	0	63	100	599	0.669	53.89

Suboptimal exons with probability > 1.000

Exon Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr..

Suboptimal exons with probability > 1.000

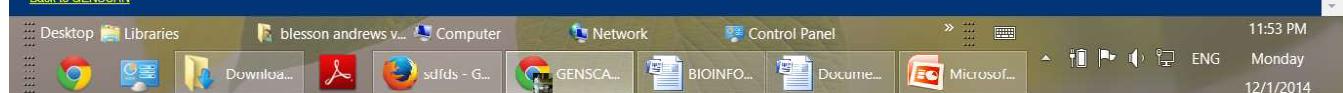
Exon Type S .Begin ...End .Len Fr Ph B/A/C D/G/T Cod/Rg P.... Tscr..

NO EXONS FOUND AT GIVEN PROBABILITY CUTOFF

Predicted peptide sequence(s):

```
>/tmp/12_01_14-13:23:39.fasta|GENSCAN_predicted_peptide_1|638_aa
INRQVCKXKSFSGRQGPSSRSAVWGSSEPRMPCVARS66AGG6ACGFRSGAGSFGRSLYN
LGSNKS1S1SVIAAGSSRAAGFPGGGSSCCFAGG6YGGF6GGSYGGF6GGGRGVGSQGFG6G
GPG6GAGGFGSPGVFGGPMSFGSPGPGFP6GFP6GQEVIVNQSLLQPUNVEIDPQTIGQW
AQERDQIKTUNKIFASPIDKVRLLEQQQKVLLETQMLLQQQTGSSPSSLEPCFESYSP
LCKQLDSLILGERQNLLEGELISHQQLVEDPKKKYEDIEINKR1AENEFVGULKKDVOAA/PN
KVELQAKVKSLSLDEVSFLRFLTYENELSGNQSHASDTSVLSNDNIRCLDLSITAEVRAQ
YEETAQRSKSEAEALYQTKLGELTQTTAGRHSQDQLINTKSEJHELNRMHQQLRAEIEVNKK
QNAHLQATAEAEQRGEMLAKDNALKQDQTAQKAKDOLARLLRDYQELHNVKLADW
ETATYRKLLGECECIMSGECQAVCSIVWSNTSTGGS6GSSRNGVFGGVSG6GSSGYYK66
SSSSSSSGYGVG6GSSGSGYGVG6GSSGSGY6GSSGSSGSGYQGSSSGSRL6GA6S1S1VSH6G6MS
SS651QTS6505WYS6066GTS1RF5QTTS5SQ55TK
```

[Back to GENSCAN](#)



Desktop Libraries blesson andrews v... Computer Network Control Panel 11:53 PM
Download... sdids - G... GENSCAN BIOINFO... Document Microsoft... ENG Monday 12/1/2014

4.2 Random DNA Sequence of length 10000

TCGAGAAAGTCATAGTACCTCGGGTACCAACTTACTCAGGTTATTGCTGAAGCTGTACT
ATTTAGGGGGGGAGCGCTGAAGGTCTTCTCATGACTGAACTCGCGAGGGTCGT
GAAGTCGGTCCCTCAATGGTAAAAAACAAAGGCTTACTGTGCGCAGAGGAACGCCA
TCTAGCGGCTGGCGTCTGAATGCTCGGTCCCCTTGTCATTCCGGATTAATCCATTCC
CTCATTACGAGCTTGCAGACTACATTGGTATATGAATGCGACCTAGAAGAGGGCGC
TTAAAATTGGCAGTGGTTATGCTCTAAACTCCATTGGTTACTCGTCATCACCGCGA
TAGGCTGACAAGGTTAACATTGAATAGCAAGGCACCTCCGGTCTCAATGAACGGCCG
GGAAAGGTACGCGCGGGTATGGGAGGATCAAGGGCCAATAGAGAGGCTCCTCTC
ACTCGCTAGGAGGCAAATGTAACAAATGGTTACTGCATCGATACATAAAACATGTCCAT
CGGTTGCCAAAGTGTAAAGTGTCTATCACCCCTAGGGCCGTTCCCGCATATAAACGC
CAGGTTGTATCCGATTGATGCTACCGTGGATGAGTCTCGTCGAGCGGCCGCATT
ATTGCATGAGTAGGGTTGACTAAGAGCCGTTAGATGCGTCGCTGTACTAATAGTTGTCGA
CAGACCGTCGAGATTAGAAAATGGTACCAGCATTTCGGAGGTTCTCTAACTAGTATGGA
TTGCGGTTCTTCACTGTGCTGCGGCTACCCATCGCCTGAAATCCAGCTGGTGTCAAGC
CATCCCCCTCTCCGGGACGCCGCATGTAGTGAAACATATACTGTTGCACGGGTTACCGCG
GTCCGTTCTGAGTCGACCAAGGACACAATCGAGCTCCGATCCGTACCCCTGACAAACTT
GTACCCGACCCCCGGAGCTTGCAGCTCCTCGGGTATCATGGAGCCTGTGGTTATCG
CGTCCGATATCAAACCTCGTATGATAAAGTCCCCCCCCTGGGAGTACCAAGAGAAGATG
ACTACTGAGTTGTGCGATCCCTGCACCCAGCTAAGGAAGCTACCAATATTAGTTCCG
AGTCTCACGTCTGACCTCGCGGGTAGATTGCCATGCGTAGAGCTTACGAGGCCAGCGGAA
AGTGTGAGGCCTTTAAGTATGGCGAGTAAGTGATCAACGCTCGGATGTGACTATAT
ACTTAGGTTGATCTCGTCCCAGAAATTAAAGCCTCAGCATCTACGAGTTATGAGGTTA
GCCAAAAAAAGCACGTGGTGGCGCCCACCGACTGTTCCAAACTGTAGCTCTCGTTCCG
TCAAGGCCCCGACTTCATCGCGGCCATTCCATGCGCGGACCATACCGTCCTAATTCTT
CGGTTATGTTCCGATGTAGGAGTGAGCCTACCTGCCTTGCCTTGTGATACCAATGAAA
AACCTATGCACTTGTACAGGGTGCATCGGGTTCTGAACCTCAGATAGTGGGATCC
CGGGTAAAGACCTATATCTCGGTCCAACCTAGGATAAACCTCCATGCTACCTAGTCAG
ACCCACCCCGCACGGGTAATATGGCACCGTCCGACCTGGTTCTGGCGTTCTACG
CTGCCACGTGTTCTTAACGTGTTGGTAGCACAAAGTATTACCATGGCCTAGAAG
TTCGGCACAGTTAGTCGAGCCTAATGTCACAAATGACGCAGAACGCCAATGAGTGC
GACATTAGGTGGAGTTCAAGTCAGTCGGTAACGGAGAGACTCTGCGCGTACTTAATTATGC
TTGAAACGCGCCAAGTGAAAGATGGTAGGGGGCTCAAATCGTCCAGCACTCCACA
GTACCTCCGAGAGCAAGTAGGGCACCCCTGTAGTCGAAGCGGAACATTTGAGGGGC
GAGCCCACATCGTCTTCTCGGGATGACTTAACACGCTAGGGAGGTGGAGTCGATTCC
ATCGATGGTTATAAATAAAAATCGGAACGCTGTCTGGAGGATGAATCTAACGGTGC
ATCTCGATCGCTCAGTCGCTTCTGACTGCGCAAAGTTCGCACCGCTCATACACTTGG
TTCCGAAGCCTGTCCTGATATGAATCCAAACTAGAGCGGGCTTGTGACGTTGGAGT
TGAAATATCTAATATTCCAATCGGCTTTACGTGCAACCGCAGGGCTGACGAGG
GACTCACACCGAGAAACTAGACAGTTGCGCGTGGAAAGTAGCGCCGGCTAAGAAAGAC
GCCTGGTACAGCAGGACTATGAAACCCGTACAAAGGCAACATCCTCACTCGGTGAATC
GAAACGCCGATCAAGGTTACTTTGGATACCTGAAACAAAACCCATCGTAGTC
ACTTGGGACACTTCACCCCTAGGGCCCATATCTGGAAATAGACGCCAAGTTCAATCC
ACTCCGACGTACGATGGAACAGTGTGGATGTGACGAGCTCATTATACCGCG
CCGGACCGGGTCCGCAAGGCGCGGGTGCACAAGCAATTGACAACTAACCACCGTG
TATTGTTATGGCACCAAGGGAGTTAACGCCAGTCAATGGAGCTCGCAATACAGAGTTA
CCGCATCTGCCCTAACTGACAAACTGTGATCGACCACAAGCCAAGCCATTGCCTTTAG

ACACGCCGTTACAGTATTGAAAACCTTGCGGGGCATGGCTACGACTTGTCAGCCA
CGTCCGAGGGCAGAAACCTATCCCCATTGTATGTCAGCTATCTTCTACCCATCCCCGG
AGGTTAAGTAGGTTGTGAGATGCGGAAGAGGCTCTCGATCATCCGTGGGACATCAACC
TTCCCTTGATAAAGCACCCCGCTCGGGTATGGCAGAGAGAACGCCTCTGAATTGTGC
TATCCTCGACCTTATCAAAGCTGCTACCAATAATTAGGATTATTGCCTTGCACAGAC
TCCTACTCACACTGCCTCACATTGAGCTAGTCAGTGAGCGATTAGCTTGACCCGCTCT
AGGGTCGCGAGTACGTGAGCTAGGGCTCCGGACTGGCTATATAGTCGAGTCTGATCTC
GCCCGACAACCTCAAAACCTCAACTTTTAGATAACATGGTAGCCGAAGTTGCACGAG
GTGCCGACCGCGGACTGCTCCCCGGGTGGCTCCTCATCTGACAACGTGCAACCGC
TATGCCATCGATTGATTGAGCTCGGGACGGTGTGTCGTAGTTGGCATGTTCCCTT
TAGGTGTGAAATCACTTAGCTCGCGCGTAGTCTAAAGGAAACCTATGGACTATGT
TTCGGGTAGCACCAGGAATCTGAAGCACGTGTATGTGGACGTGGCGTACACCTTA
ATCTCCGCTTATGCTAAGGATCTGGCTGCATGCTATGTTGATACACCTACACTGCTCGA
AGAAAATATACGAAGCGGGCGGCCTGGCCGGAGCCCTACCGCATCGACACGTGTTCGT
TTACTGTTAATTGGTGGCACATAAGCAATATTGAGTCCGTCAATTTCAGCCCTTTATCC
TCGGCGTTATGTGTCAAATGGCGTAGATCTGGATTGACTGTATGACGGTACCTGCTGATC
GGTAGGGAGACCGAGAATCTATCGGGCTATGTCACTAAAACCTTCCAAACGCCCGTGT
CGATACTGAACGAATCGATGCACGCTCCCTCCTGAAAACGCATAAACATACAAGTGGG
CAGATGATGCGTACGCCCTCTAACATCAACACTCTACGCCCTCTCAAGAGCTGGA
AGGGCACCCCTGCACTGGATAGGGATTATCTGTAAGGCAAGCTCGTACCGTCATTCA
TGCAGGAAAGAGTTAACACGATTGGAAGTAGGGATAGTTGAAACCTCGTTACTAGTCCTA
ATAAGGAAACGCTGTCTGAAGGATGAGTGTAGCCAGTGTAAACCGATGAGCTACCCAG
TAGTCGAACCTGGCCAGACAACCCGGCGCTAACACTCAATCCCAGGGCCTGACGCG
ACATATCAGCTAACAGTAGGCCGGGGTAGACCTTGGGTTGAATAAATCTATTGTA
CTAACGGCTTCAACGTGCCGCACGGGTGGCACCTCAGGAGGGGCCACAGGGAGGAA
GTAAACTGCTATTGTCGCCGATGGTGGTAACTAATTATGTTCTTGCCACTACAATTGTA
TCTAACGGCTGTAATGAGAACATCCACACCTTAGTGAATCGATGCGCCGCTCGGAATAC
CGTTTGGCTACCTGTTACTAACGCCATCGCGATTTCAGGTATCGTGCACGTAGGGTTG
CACCGCACGATGCGAACTGGTGGCGAAGTACGATTCCACGACCGGTGTACGATTCAA
CTACGCGGACGTGACGAGCTCTTATATGCTCGCCGCCGGACCGGCCTCGTGATG
GGTAGCTGCGCATGAGCTTATGACAAATAACGAGTGTGTACTTGTAAATCATCTACA
GTGAAAGTCGGAGAATAGGAGCCGCTACACACATTACCGCAACTACACCTAACTGA
GATACTACCATAGATGACTACCCATCCCTCTAGGCCTTAGATAACCGGATACAGTGA
TGATAGGTTGTGGGTTACAGCTATGACTTGCAAGCTGCTATGGAGGAAGGAACCTT
CCGTGTTAGTATGTTGACCCGTTGATTACGCATGCGGTAGATCATGTTGGTAGAGACA
TCCAGGTCAAGTTCTCGACCTTCTCGTGGGAGGTGATCCAGTTCACTATAGGACCATTCC
ATTGAGCATGGCACTAACGTACGCCGCCCCATTCTGGTAATCATCATCCCTATCAGGGT
TTGGAGTGAGTGGTGGCGATGGTTATTCCCCAGGAATGGACTTCTTATTACAGTCGGTCACA
TTGGGCTACTCCTGGGCTTCCGCTTGGCCCGTCTGTTGGGCCATTGCGTGAGT
TTCGGCCCTGTGCTGCGCTGTGCGGATTCTCATTGGGCCACATCTGGAAACCC
CGACCTATTTGACGGCACCATTGGCGGAAGTTGCTGGCTTGTGCACCGTGAAGTC
CCGCGGGCCTCCCCCTACAAAAGACGATAAGCTCCGGCAAGCAGCATTGATCAACCGA
AGGATCGGCGATATTAAACAGAGATACGGCTGATTACTCTTGTGTTGGTATCGCTAA
CTGCGCCGCGGAGCCTTATGGCAAAGTCGTCCGCAACCATTCCGGTAGCGCTTAAGG
TCCATAGCACATTGTCATCCGGCGTGCCTCAATTGACGACCCCTGGCGCAAA
GGTGTGGCCACGTGCTAAATTAAAGCGGCTGCACTGCTGTAAGGTCCGTTACGGAGAG
GGCGGCCCTGGGAGAGCACTAGCCCACGACCTGTACAGGAACACTCTATATTGTTCTC
GGACGGACAGATTACTAGAGTGCCGCTTCAGCCCCCTGTCGCGACGTCTGAA
TATGGCGTTGTGATTCCACTTATTGAGGCATTAACGATGCGTAAGGAGATCTGGA

ATGAACTGGCCTATGTCACTGAAACTGTGCAAACACCCAATGTCGTTAGTGTAGGTTCTG
ACCGATACTGCTTCGTTGAGAACTCACAACTTACAACACTGGGGACATAATCCCTACGCC
CATCATCTACTCGCGTCCCTGTGGCTCAGTCATGTAATGGGAGAGTATCCTCCACAAG
ATCTAGTGCTATGGTGGTATAGTAAAGCTCGTACTGTGGTAGACGCGACACGGTAGGAC
CATCAGTAATAAGGATAGTGCAGAACGCTCACAGACCACTGCCTATAGGGGTGCTTACTT
TTACAAAAAGCGACTGTCACTATAACCCCACGAGGATTGAAAAGGTGAACCGACCCAG
ACGATCCGGAGGCACGGGCCTCAAAGCCGCGACACGACGGCTGTCGGCCGGTAACAGT
AACCCCGGAGTGAACCTCTATGGGGCTGGATAGAACAGCCCTGGTGGGCCCATCAGC
AACCCGAATACGTGGCTTCGGAGGCAGGCGGAGGGCGATGTCTTCCACTATTG
AGGCCGTTCGTTAATACTTGTGCGTCTAGCCGCTATATTGTCTCTTGCCGACTAAT
GTGGACAAGCACACCAGGCCATTATCGGAGCGCCTCGGAATACGGTATGAGCAGGCG
CCTCGTGGAGACCATTGCGAATACCAAGGTGCTCTGTAAGCAGCGAAGGCCATACGCGAG
ATAAAACTGCTAGGAAACCGCGTGTCTACGACCGGTGGTAATTAACTCGCTGACGTGT
AGACATTCCAGGCAGTGCCTGCTGCTGGGCTCCCTGGTACTGGTAGTTGGACTT
GCCCTTGGAAAGACATAGCAAGACCCCTGCCTCTCTATTGATGTCACGGCGAATGTCGGGG
AGACAGCAGCGGCTGCAGACATCAGATCGGAGTAATACTAACGTGGATAACTCCGTA
CTGACTACGGCCTCTAGACTTACTTGACCAGATACGCTGTCTTGGCACGTGGATG
GTTTAGAGGAATCACATCCAAGACTGGCTAACGACGAAGCAACTCTTGAGTGAAATTG
TTGTCTCCTGTATTGGGATGCGGGTACTAGATGACTGCAGGGACTCCGACGTTAAGTA
CATTACCCCGTCATAGGCGCCGTTAGGATCACGTTACCGCCATAAGATGGGAGCATGA
CTTCTTCTCCGCTGCGCCACGCCAGTAGTGTATTACTCCTATAACCCCTCTGAGAGTCCG
GAGGCAGGAAATCCGCCACGAATGAGAATGTATTCCCCGACAATCATTATGGGGCGCTC
CTAAGCTTTCCACTCGGTTGAGCCGGCTAGGCCTCTGCCGGAGTTGACGGACT
GCTGCCGACACCCGGGATTGTTAGGGGGTTATTGAGGGCACTCGCAGCCAACCT
GTCGGGACCAGCCGGCTGGTCATCGGGCTTATATAGCGAAATGCCGAGGACCCGGCC
CCACGCTATGGAACGTCTTAGCTCCGGCAGGCAATTAGGACAACGTAAGTATGGCGG
ATATAAACAGAGAAACGGCGAATACCTATTGATCGTACGGTAAATAGCCTCGCG
GAGGCATGTGCCATGCTAGCCTGCGGAGCACTCTAGTTATGCATATGGCCACAGGACA
CTCGTCGCTTCGGATTGCCCTCTATGTGACGGTTTAGGCGACTTATGCTCAGCAC
CGTTAAACCAGACCGACACTAGATCTATAAGGCTGCCACGCCAGACGAGAGCGCACGG
AGATCACCAGCGATCTATGATCGGCGACCATTGTTGAGGACTGGGCGAGAGG
AACTACGGTGCCGCTAACAACCTCTCGGCGTACGTTGAGGACTGGTACTGGGCGAGAGG
TGATCGTACGCTATTGAGGATTGACTGATACCGGAAGACATCTCAGTTGAGTGAAGTGG
TACGACAGAGACCGTGCACCTACCAAAACCTCTAGTCTAAGTTGAGACCAATTGGTAGT
TTGTCCAGAACTCAGATTATCACCAGAGGACGCACGCTCTACCTTATGATCCATTGAT
GTCCCTGAGGCTGCAATACATGCAACCAGGCAGTCAGTCTCGCGTAAGTCCTAGTGCAATG
GGGCTTTTCTGGCCTCGAGAAGAGGAGACGTCAGTCCAGATATCTTAATGTGG
TAATTGGAAGGACTCTGGCCCTCCACCCCTAGGCAGTGTATACTCTTCAAAACGGGC
TATTAGTTATGGGGTCCGAAGATTGAAAAAGGTGAGCGAACCTGGCCACCGGAAAGA
CGGGCTTCAAAGCAACCTGACCACGGTTGCGCGTCCGTATCAAGATCCTCTTAATAAGC
CCCCGTCACTGTTGGTTGAGAGCCCAGGACGGGTTGCCAGATGTGCGACTATATCGC
TTAGTGGCTCTGGGCCGGTGCCTGAGGACTGGCAGTGTACGATCGCACGCCCATGAGAACGATAG
CCCTTGCAACATATTGCGTTTTGTCTTCTGACTTACGAGTACGATCGCACGCCCATGAGAACGATAG
AGCTTCTACCGGTGCGCCTCCGTACGAGTACGATCGCACGCCCATGAGAACGATAG
GTAAACCTGGTGCCTGTGAGCGACAAAAGCTTAAATGGAAATACGGCCATAAGTC
GGTGCAGAACGGCTCGTAGCAATGTTGGTCTGACTATGATCTACATATTACAGGCGGTA
CGTCTGCTCTGGTCAGCCTCTAATGGCTCGTAAGATAGTGCAGCCGCTGGTGTACT
GATGACCTCGGCTCCCCATTGCTACTACGGCGATTCTGGAGAGGCCAGCTGCGTTGCT
AATGTGAGGACAGTGTAGTATTGAAACGATAAGTCCCCACTGGTTGCGCTATCGA

AAAGTGAACCTCATAACACATGCTGTCCCACGCACATGGATGATTGGACAAATTGATT
CGAGTCTGATCAACCTTCACACAGATCTAGAACATCGAAAGCAGTGATCTCCGGGTGCGA
AATAAAAATACTAGGTAACTAGAGGGCTGCGACGTTCAAACGTTGGTCCGTCTGAACC
GCCATCCAGGATCACGTCGCCCTGAAAAAAAGATATCAGGAACACTCTCCTCAGCAGT
CTGGTCTATGGAAACTACAGGACTAACCTCCTGGCAACCAGGGGCTGGGAATCTGTCA
CATGAGTCAAGGTATTGCTCGATAATCTACTCCAGGCATCTAACTTTCCCAGTCGCT
TAAGCGGGCCTGCCCTTCTGCCTGTCGATCCATAGGACTCGTGCACAGCGCAGGGCTT
AGTCGAGGAGAAAATATCCGGGCCAAAGACAACCAGCATCTCGCTTGCCTAACCC
CCCTACATGCTGTTAGAGAACATCAGTGGAAACCCGGTGCCAGGCATGGAATGTCCTT
AACTCTGGCAGGAAATTAAAGGGAACGTATATACAACGCAAAGAAGCTGGAAAATTGGC
GAGGGAAATCCTGTTCTGTCTATCCAAGAACATGGCATGAGGTGGCAACCGTCGTGCTAG
CGTACAGGGTGCACTTGTAACCATTGGGACACCGGACACTCGCTGTTTCGAAATTAC
CCTTAAGCGCGGGTATTGAACCAGGCTTATGCCAAGATCGTAGCAAGCAGACTCAA
CAAGATATATTTGCCCGCCTAACAGACGAAACTAGTTGGAGGTTATGGAGCATACTATC
ACGTGGCGGCCACTGGTGAGTTACTACACCCCCAGGGCAACGTTGATGCTCCTAAAAAA
ACTCTGGCTGGACGCAAGCGTAACACCCGTGTCACCTCATAATGTTGCAATTCAAGG
GCTTGATCTACACTGGATTGCCATTCTCTCAAAGTATTATGCAGGACGGCGTGCCTC
CATGTAACCTGTCATAACTTACCTGAGACTAGTTGGAAGTGTGGCTAGATCTTGCTCA
CGCATCTAGTCGGTCCACGTTGGTTTAAGATCCAATGATCTCAAAACGCTGCAAGA
TTCTCAACCTGTTACTAACGCGCTGGGTCTACTCCAGCGGATTTTATCTAAAGAC
GATGAGAGGGAGTATTGTCAGACCACATAGCTTCAATGTCCTGATCGGAAGGATCGTTG
GCGCCCGACCCTCAGACTCTGTAGTGAATTGTCCTGAGCCATTGCATGCGAGATCG
GTAGATTGATAGGGGATACAGAATATCCCTGGATGCAATAGACGGACAGCTTGGTATCCT
AAGCGTAGTCGCGCGTCCGAA

Input

Output

genes.mit.edu/cgi-bin/genscanw_py.cgi

GENSCAN Output

View gene model output: [PSI](#) | [PDF](#)

GENSCAN 1.8 Date run: 1-Dec-12A Time: 18:26:33

Sequence /tmp/12_01_14-18:26:33.fasta : 18000 bp : 51.09% C+G : Isochore 3 (51 - 57 C+G%)

Parameter matrix: HumanIso.smt

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr...

1.81 Intr + 2315 2637 123 0 0 61 -3 156 0.321 5.89
1.82 Intr + 4342 4575 234 0 0 5 22 179 0.238 5.42
1.83 Intr + 6670 6793 124 0 1 65 46 88 0.532 2.16
1.84 Intr + 7093 7228 136 2 1 63 111 68 0.367 5.93
1.85 Intr + 7357 7516 108 1 0 64 78 117 0.406 7.48
1.86 Intr + 8243 8314 72 2 0 78 65 69 0.632 1.82

Suboptimal exons with probability > 1.000

Exnum Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr...

NO EXONS FOUND AT GIVEN PROBABILITY CUTOFF

Predicted peptide sequence(s):

