

Group A formation proteins from randomForest analysis

>epi_00394

MSFNEIIFKNFKQNVTHYAIYLLSLVISVVLVYFSFVTLKYAHHL
HGNQSFPAIKEGSQVGSYFLFIIIVVFLLYANFLFLKRRGRELSLLQIIGLTKKDIMK
MIMLEQLMTFMMTTIVGIIILGIFGSKILLMIVLRLLLGINVSVSIIFNHAILETLLLI
AVSVYLVIVFQSYVYLLKRSIKELASDVNKKEFSHTRTTLGEVVLGLLGIIMIVSGYIM
STKLVDNVETIIQPFAILFLTVIGSYFFRSTVSLIFKAIQRLRNGTVSVTDVMFTSP
IIYRVKKNAFSLTVMTVVSVAITVSVLCAALSRSTLTNEVLLSSPHDVTLKNQKQANE
LAFKLNRRNIEHYNYKEVVYAKVYKDLHLFSEGVYRPKEITVTSVKYIPNVSTKKGQT
DIIIPRGLKDVVKTDKKRTAYVGSKKFRLKVSRLRGINKVYFMSDVDRGRPTLILND
EDYQKIREHIKEKNIVSQYGFDLKNKNDLPELEKLVSSINEDIETRSEAASEISSLTG
ILLFVTSFLGITFLIAAGCIIYIKQIDETEDELNYTILRKLGFTHQDMSKGLKLVKI
FNFGLPLIIALLHAYFASLAYMHLNMVNTNQIPIFIVMAVYTATYAVFAIIAYNHSKRT
IKHSI

>epi_00430

MKNLFLSSVAVILSICLVLLLFLATNQNALAKVHKTICTFTKVK
TVSHNDIFPNLNMNHVAEASEKNSDHQKNVKNKTNQSVRTQASYDFNPQECKVIALNY
AENHLSKKYNIERTFNESLSHSVYVFSSENKLNHQTQVKNKSGTVHASNINNH

>epi_00538

MKDWTINNHSKRQLIIQSNHDDISIIIEPLRNGSFKILAEFNALN
TDYIQVYDHTLVVSIHEKEISVFDRSSQSI

>epi_00542

MKLISKSIYLSFIFFIIFSLMLFTNQSFANDNKRQKLPKPGDIIIVT
KGPVLNGFFGHCSLAIDHDTVLQIEGPGDKPMTEDFESYRDRYGKGNKNEWIKVYRCSH
PNAGTKAAQWAKKHYENSSEYLVTFNLKSETFTYCTKIIYQAYKYGVDKNSVSDHGL
YIISPYALTDNFTDEYKLLKLVKTY

>epi_01643

MKNQDKIQNVVKKLLSSLGVNIKKTKSRLDIINTLPTSNOVSHLK

>epi_01831

MQQQKIKRFLAHFVLDHVTIEYLKYGHQLTIEQLKLLIFILHFT
ENHKEDLSLNMIIFYKNYQKNQLLKSITHLYEFNWISKKRHPYDQRRLVITLTQNQCS
KITQLIEELEHFLEVKSTLINEINHSTLLSYLKYKCHSQFRVIEQSCTSQHLTLEELYL
LGLLIIISDNKTTFKSIKVHALKGIAMGPIIKTLQSKGYLIKRSRDERYIVLTLRK
EKVNVIQSEIEECYNKLEQGIQHV

>epi_01832

MRENYŠIRVFNNLAINTEKALSVKVFYCLKGKCEVTINVQKHIL
EKDDIAFVVMNDTYSLLSNTETMCCIIDIPIHRYLSQKDTQFLISGTKIDESSRDRIK
YWILKILELHCLSYNDVSEIQRLIQFLLELSYLKPKLVDNKFYLSEDIHQYLVDH
HDSKINKHELAEAVNLSNQALTSMFQKTPFQTFNQYLNQLRLKFCFLIDILTTHKPIEE
IAIDHGFFHYSRFIQLFKNTYGYTPKLIIRDYIATSIFKNTAAEIDLDRHFLMNIHEL
QDLDSKIIISKYIKMSDKGKKYRSYDIYIEDNRSTILDQEQIVHIKRNLSLSQKSMRY
VIELNYTSMIENKELCRYEMLKILRFCSGLNLIPTFKIITDRHDTFTSKEKMAKLT
QMLFIMLREFNQLEIEFIVEDMMLKQVVQLKKMISSYFEYKLNRYIKNEKIGNVNYQ
NLEKQVTQIFIPIDQLHLYIKEMSFVKVILETSYLTDLKSEVVEYPWLQHIHTLIKMC
GLVRGVLIQPSFDYATTYTYHSNLKPHHILSYVIRMFNQLRGTIVYKDDAIIMTKYKY
EYQAIAIFLNRIKENVNHQQLIFSDIQSLNHAIEYIEEMPLLNMENNELDSHIPYP
HQYWLAKLKSSESQQAIVDLPKMSIAHLTFLCS

>epi_01980

MDIKQŠSEKQGRPHHLSDSRTVLKRNFIPIPTYILLQSIPIIIV
VFGSLGITAMITQQAPPQWLYHCSLSLSFVIAQGLILIIFYKIHQSVINDVMKQQWII
AKNKIIKIVIVAVVYLLLLLIIRVIGTSLPNHLSYHLTQSEQRTLGLFKSPYVLLVTF
ISMVFLRPMVEQIIYRYLIIHELKGVWNRQFVIGLSIVIETIVHVYDMASIFEIFPYI
VIASAATILYIKSRDNLIVAYIFQVILQCILFIEILCKYTNF

>epi_02055

MSFYEFMQNFLGDDTPLGELVDWINQDINFPKDVTSQKEILTYF
RNHPCPENIPVTSIKRALAVFNQFTKV

>epi_02098

MKKISLIATTVLTGLLLFPSVNDTTTHAAEVTSHDAQAVAIQAM
KNSGGNPDQLQNFKKVKDKGDYFTIDINNKSAGVGTYKVKYKNGVVLYKSGNYGEYSQL
NTRQCYVAQDIATTSSQVKRHSTQQTQAVDSTRELNSYVGVQVQSSDLPQQSSSDMLP
NTGMKDKNFNTNGIISLFLLLTAGFITLYHQPLRKMS

>epi_02108

MSFYHFIQNFVGDDETPLGKLATCINQDEDLPMEEETAHNILNYF
NQLNYFDDDCIEAVKRSLSLYEQSKVAL

>epi_02134

MTFKILWKMFKQNFVNQRHIIVPPFISVISILFGIEYILLSLTTN
IYFNEHHPKELKISAIIGIVFMTMLLFIFLIYANHFVMNHRKKEFALNMALGMEKKHLR
LIIILIELFIQFIISAILSIVGGYLFGEFFMLFNKIVNTHQPQLSDYFPDVLMSKITL
TMLLSLMIILFVINNFKISFKNSLQLLIKNKSKEKSRVRLIIFLILGLIFIGIGYY
LAIKPNTAIGSLGIIFFAILSTLIGTYLLFVSLGSIIVLEMLQKLDHYYYKPNHFFFAIA
GLKSRVKSSAIGLATISFMCTFLIVTLSMTVSTYRNMDFRFEFAFKNDYAGYYIGDFH
KDSKIQRKIENLKKDIRQEVPTGQFKIYARGMVGAEIQGGLKHKKLRQTVSSGLFNF
GNKQKFNSFISYINKSDYNKNNKIKLDDDEIAISTSVSLFKMKTLNIFGKTYRVKY
IESTNIDNLLYADGITLIVNQQQLMDRIVNEYRNHNENLIITPNQVQTAVEFNVLKE
KDKLNHRIKKIGVQHDIEFQVKKQNLMLWKQVNSSLIFVGSVSVLVLIGIFLMMYYK
QVSEGHEDRDAYITMKQLGLDEILIKKTINKQVIWVFLIPVIVAIHTLAAFRIIYSV
LGIVGQYDLGLYATSYVGVIVVFIIFYSMMYWITSRIYYTMINDKH

>epi_02135

MLLEVKNVKKVYGRGMNTTMALNQMNLEIDENEFVAIMGESGSG
KSTLLNLIATFDRTTEGLIKLDGLPLNQLKNKDIARFRREMMGFVFQDFNVLNTMSNK
DNILMPLVLANERPIMQKRLMEISEQLGIEDLLEKYPSEISGGQKQRIAIARALIA
PKLLLADEPTGALDSKTSKNLMCLFRKINQKHQITILMVTHSNIDASYANRVIIFIKDGR
LYHEIYRGEESQTDYQKRIADSLAILNGVGD

>epi_02136

MKFAYIQSIRNEISIIILIIILLFFALIFVVFSLPFDAYVLAISII
LLLMCVRWIKYLSFKKNEHLKDKVAYLEHELAHVKNQQIEYRNDVESYFLTWHQIK
TPITASQLLLERNEENVNRVQRQIVHIDNYTSLALSYLKLLNEESDMITKVTVDLL
IRPLILKYRIQFIEQKTIHYEKSEDIILTDAQWASIMIEQLLNNALKYAKGKDIWID
FDVANQTLQIKDNGIGISKADIPKIFDKGYSGFNGRLNEQSTGIGLFIQVHIANHLNI
QVTVQSELNHGTVFFIHFTEK

>epi_02137

MKIFIVEDDLVIAESLANELSKWNYEVHVVDNFEKIMEDFRRVE
PQLVLLDINLPTLNGFHWCEMRKVSHVPIMFISSRDNDMDQILAIQMGDDFIEKPF
NLSLTVAKIQALLRRTYDLSIARDEIAVKGCKLIVDEAKLSKDNEHVQLSLELQILK
LLFQONENKYVSRALIEKWESENFIDDNTLAVNMTRLRKLLSIGVDDLIETKKNVG
YRV

>sap_00201

MTKLEVINPATNEVLERLDYATHEQINHQIKQAHQAFQNWKKVDAHERSAKLAQWAQLID
DHQDELARLITLEGKPLAEAKGEVAYANSYVKWYAEAEAKRVYGRTPANSPSKKIVIDK
FPVGVGAIPTWNFPAAMITRKMAPALAAGCTIICKPAVKTPLTTIKLVLAHQAGFPKD
AISYIIASGKDAGDIFTNHSLISKVFTFTGSTAVGKSLIESSAQVKNVTMELGGLAPLIV
HNDADIEAAVDGTIASKFRNAGQTCICANRIYVHEDIAEAYNEKLEKVVHALSVGDGLKE
EVKIGPLIDNQAVEKVLTHIKDAQEKGGQLSRSIEDIQALGGNFKLPVITNANLDMKAM
HEETFGPVAPVMTYSDLDEVINIANDTEFGLASYFFTNDYRTGLKLFNELEYGVIGWNDG
GPSAAHAPFGGLKESGYGREGGVEGIEPYLETKYLSIQE

>sap_00203

MKNHLHRWVIEALSFMALGLFGSLIIGLILETIGNQTLIPQIDLSFLSEIGKVAMSVTGA
AIGVAIAYGLGAKPLVIFSCVIVGTLYTTFGGGAVGAYIATLIVVELSRFYASKTKIDI
IVTPLLTIIVIGGAIKLFGPALNEFMGTGLGKVIIVATEQQPLIMGILVSVLFLGLALTAPI
SSAALALMLDLSGLAAGAATIGCAAQMVGFVNGYKDNWGSIGLSIGIGTSMQVQVNIK
HPMILVPTIASAVVAPVMTTLFPMTNNAAGAGMGTSGLIGQIMTINTMGGSLQTWLLII
IFHIIIPAVISFALYKYFFKKQWIMPGDQLIHVAN

>sap_00398

MKVFLKLGWFFKQKTSYLIGLGTLLLIALIEIVPPQIIGRTIDEMTSNRLTPRLLAIYL

FILVITAILTYVLRVWRLSIFGTSQKLGKILRTYLYKKYTEMSAIFQNRRTGDLMAHA
TNDIRAVQNAAGAGILMIADSLITGGTVIITMATTVSWKLTLIAMIPLPFMVLLTSIYGS
LLSKGFKKAQAAFSKLNDKTQESVAGIKVTKTFGYEPSDQADFKHLSDDVVAKNLKVKSI
DALFDPTITLVIIGMSYFLSIAFGAQMVFHNDISLGQLITFNTYLGMLVWPLLALGLFFNI
VQRAKASYERIEEIGELPNDIDTSYVIDERPOGDIRFNINQFYFPGNEDQGIYDIHFTIK
EGSTVGIVGRTGSGKSALIRLLREFDTQHAQDIEFGGHPIRDYNVESLRAQFGYVPOEH
FLFSTTIRNNIAFSNETIDDEKIFEASKLSHIHDDIMQFSKDYQTVVGERGVSLSGGQKQ
RVSIARALLVNPEILILDDSLSAVDAQTEEAILENLERLRSGKTNIITAHRMSAVKHADL
IIVMNEGRIIERGNHATLMSKKGWYYDYQAAQALQEQLSRNLDLSLTCKGDGEND

>sap_00399

MTENANLTAKDQGSALIRLFKYTLPYKWIIVLAFITLILSTIASMMTPYMKIFIDDYLT
PRHFPKETMVWLIVIFISIQLIGAITLYFSQYLFQYLAFKVIQQLRIDAFNKLKGLGMRY
FDKVPGGSIVSRLTNDTETIVDMIVGVFSTFIMAFFMMISSYIMMFVLDVKKLALIALIFL
PIIMIILASYRKYSAFLFSKSRQRLSDLNKLAESIEGMKIIQAFNQERRLNKEFNKIND
EHYQYMLKTVKLDLSLLLRPAISSISIFAVVMILGYFGVISFTTGITAGVVFVAFVQYMERF
FEPINQVSNLNLILQQALVSASRVFALINDDTYEPQQEANNNDNAIETGEIEFDNVFSYD
GETDVLKNIISLTAKPGEMIALVGHGTGSGKSSIINLFMRFYEFNRGDIKIDGNSIKKIPKT
ELKEKIGLVLQDAFMFYGTIASNIKLYHPSMTFFEQVKAAAEFVHANHFIEKLPNQYQHKV
IEKGSASFSSGERQLIAFARTIATNPKILILDEATANIDSETEEQIQQSLNKMRKGRITLA
IAHRLSTIQDADQIFVLNKGEIVERGTHAQLIAQKGIYHNMYLLQNG

>sap_00643

MSVLVLGGAGYIGSHAVDQLISRGYDVVVVDNLGTGHRAAVNKDARFYEGRDIRDKPFLDQ
VFTTETIEGVFHFCAYSLVGESVEKPLAYFNNNVNGLQVLLVEMYDHQVKHIFSSAAV
YGEPTVPITEADSKTPTSPYGESKLMMEKMMHWSHNAYGVNYAALRYFNVAGAKEEGTI
GEDHRPETHLVPIVLQVALGQREALTIFGDDYDTEGSCIRDYLHVVDLIDAHILAYQHL
QNGGESGAFNLGSSQGYSVFEILEAARKVTNKP IEAKVGPRRAGDPSKLVASSDKAQSIL
GWKPKHDNIHDIIDSAWRWHNQHPNGYQEDTEV

>sap_00760

MNISNPKIYEQVADIILENIKSEDYQVGDKLPISIQKLSKQFGVSTASVREALNALRTIGV
IEMKQGYGTFIKQIEPTFFFEFGDKFNLSLVQIKELLELREIVESATVANAAARYRNKADLEK
LATALSVMGEAVIDGTSGEKADLEFHLAIAHAAQNSLLVLELLNNISELMQNSMEETRKIF
IYDKQKTMTKLLEEHEMIYHAILEQDDKTAVKAMQSHLLEVKQTIILANFKET

>sap_01003

MKTKHLSLKIIVIALVLGIAIGSIFNVFSNTAFVENVDKYLEFNVIGQIFLNLI FMLVVPV
FVSIVLVGVVGVGDPKLLGGIGIKTITFFLTTTATAIAIIIGITLALVLQPGAGKSELLNSDD
VSSYQQTLEKEDSSQDSAAKQTFDQTLINLFPKNPLQAMTDENMLQIISFAIFIGVGIIM
VGSKAQMVHKFLEQTNDVLMYIVTMMMSVFAPIGTFGLVAHAFTGAGFGAIQQLGMYFFI
VLLGLGIHFFVYGAAVKFMKSSPLKFFKDFIPAITLGFSSASSSTATLPVSLECTKKMG
VRREIASFVQPLGATINMDGTAIMQGVATIFIAQISGVTLVSQIITVVVIAVVASIGTA
GVPGVGLIMLAMVLNAVGLDPAAGIILGIDRLLDMTRTSGNITGDAACALVISNAEERK
LAKERTLNE

>sap_01607

MKYKTVFDIIGPTMVGPPSSHTAGAVRIGLVARDLNFQPKQADIYLYGSFMETYKGHGT
DVALVGGLLYDTHDDRIQTSLETAEEAGMKVNFIEMAEERSHPNTAIINMRDGDKEISV
EGVSIIGGGKIEVVAINGFNIAISGNYPALLVFKHDTFGTIGRVANILGDSSINVGSMQVS
RKEKGDQALMTCELDDAVNDETIIEKIKNVDGVVTVSLMGDA

>sap_02162

MKYKTVFDIIGPTMVGPPSSHTAGAVRIGLVARDLNFQPKQADIYLYGSFMETYKGHGT
DVALVGGLLYDTHDDRIQTSLETAEEAGMKVNFIEMAEERSHPNTAIINMRDGDKEISV
EGVSIIGGGKIEVVAINGFNIAISGNYPALLVFKHDTFGTIGRVANILGDSSINVGSMQVS
RKEKGDQALMTCELDDAVNDETIIEKIKNVDGVVTVSLMGDA