HPC2 and Ubinuclein define a novel family of histone chaperones conserved throughout eukaryotes

S. Balaji*, Lakshminarayan. M. Iyer & L. Aravind*
National Center for Biotechnology Information,
National Library of Medicine,
National Institutes of Health,
Bethesda, MD 20894, USA

Supplementary material

Contents

1.1 Multiple sequence alignment of the alpha helical domain detected in HUN containing proteins of metazoans, basidiomycetes and plants 2

2.1 Multiple sequence alignment of the Hir3-like TPR repeats 4

3.1 Identification of potential roles of HPC2 in distinct contexts like transcription and DNA repair/replication 11

4.1 Identification of largest dense sub-network of protein pairs containing HUN domain protein FLJ25778 constructed based on high mutual information in their phyletic patterns 12

5.1 Analysis of phyletic patterns 13

5.1.1 Procedure to identify orthologous proteins 13

5.1.2 Distribution of correlation coefficients in the phyletic patterns of 70 proteins (Hpc2p and its interacting proteins shown in Fig S3) 15

5.1.3 Distribution of Mutual Information Coefficients (MICs) in phyletic patterns of human cell cycle proteins 15

Table S1: Phyletic occurrence of components of Hir complex, Asf1 and Cac2 16

Table S2: Correlation coefficients in phyletic patterns between Hir/Asf1 complex components 17

Table S3: List of 63 eukaryotic genomes considered for the analysis 18

Supplementary data

1.1 Multiple sequence alignment of the alpha helical domain detected in HUN containing proteins of metazoans, basidiomycetes and plants

```
Predicted Secondary Structure                 ----HHHHHHHHHH    HHHHHHHH-------------HHHHHHHHHHHHHHHH
UM04022.1_Umay_71019877                 486  LPVDPRLQAAFDDL    KAKVDAESF------AVKTKFPPSLKPPLVETAKLAV
CNBI1080_Cneo_134115561                 515  LPPYPAVPQEVARA    LLALRLESLEYDWDHANRSKFPEHLKPSLKRASTIAY  7
LACBIDRAFT_319105_Lbic_170097207        327  HSFHPDLQVGIESI    KQAIAQESW------AQKGKFPPALKPQLAQLALLAI
CC1G_01547_Ccin_169854391               465  NSFHPDLQVAIEEL    KKAIAKESW------EQKGKFPQALKPALSSLALLAI
GSTEN:00015692:G:001_Tnig_47218912      309  GGAPAGLAAAIERL 24 RQAARTSEGE-----SKLKFFTPEINAILLDIELQCR
LOC100023118_Mdom_126340861             456  EGLPIRLEKRIEDL    RIAAKLFDEE-----GRKKFFTQDMNNILLDIELQLQ
D130059P03Rik_Mmus_158749567            436  EGLPVLLEKRIEDL    RVAAKLFDEE-----GRKKFFTQDMNNILLDIELQLQ
KIAA2030_Hsap_156630841                 450  EGLPVLLEKRIEDL    RVAAKLFDEE-----GRKKFFTQDMNNILLDIELQLQ
UBN1_Hsap_6940751                       355  EGLPAPLEKRVKEL    AQAARAAEGE-----SRQKFFTQDINGILLDIEAQTR
UBN1_Ggal_154937332                     350  EGLPAPLEKRIKEL    AQAARAAEGE-----GKQKFFTQDINNIILDIELQTR
GSTEN:00008525:G:001_Tnig_47208786      256  SCLPKPLMKRIEDL    RTASRQFDEE-----GRKKFFTLDMNNILLDIELQVQ
NEMVEDRAFT_v1g202773_Nvec_156393686     332  NYLPTHVESVVMKL    KQAATNAADE-----GKCKFFNKD-------------
AgaP_AGAP002048_Agam_158301289          414  EDIPDQLRRDVDAL    KELARGLSGA----GGKLNLFESKVADLLLRIDDGAR
LOC412133_Amel_110751190                344  DSLPENIVEIVNKL    KACAENNKE------GKAKFFNAAVNALLLSLEKKLR  1
Dper\GL23606_Dper_195144098             439  TSLDADILSTVNCF    KETVKSRDMC-----GKKFTLDNNLSSLLLKVYEALL
TRIADDRAFT_56741_Tadh_196005425         247  KDLPMEIIEMVNQI  1 EIGGSLKDKE--------NFEDSRAQEILNKLASKFR  2
LOC100141652_Tcas_189236704             272  ENLSPDIASVITAI    KRSSELNT-------DPNNIFTSEINILILRLERKCK
yemalpha_Dmel_17136766                  438  TTLDADIVTAVNSF    KEAVKSRDMC-----GKKFNLDVKLSPLLLRVYEAVL
LOC579948_Spur_115772532                361  LELPLQLVDNIQKI    KEAARK-SVE-----GKTKFFNSYVNGLLLKIEHGTR
PHYPADRAFT_171726_Ppat_168059334        349  WPKGTVMERAIQDL 13 GEVQEAPSQP----GAKAKRLPREVKTRLAKVARLAV 26
Os05g0422600_Osat_115463977             351  KTKGTRLERAIRDL 13 VHEAEANCQV-----AVKRRLPQEVKQKLAKVARLSA  1
AT1G21610_Atha_22329715                 330  KSKTSILEKAIREL 13 ENQEADTSSQ-----AVKRRLPRDVKLKLAKVARIAQ  1
consensus/90%                                .s....l...lpph    .......p.......s...bhs..hp..l.ph.....
```

```
Predicted Secondary Structure           ------------HHHHHHHHHH-----HHHHHHHHHHHHHHH-    ---HHHHHHHHHHHHHHHHHHHHHHH--------
UM04022.1_Umay_71019877                 -------MLGEYDDNFFNHLPTIFPYNRFTMMKLVKREFFAQH  4 EDLKQDHLDKLGSMIRDSYPQQLADF--------
CNBI1080_Cneo_134115561                 ------FDHTSEDKGFAICLQSVYPYNVFTLRKLVTKLIYEGY  4 QDCEDEGIRQFKEEIDKDSAMWIEKYETSKRDW-
LACBIDRAFT_319105_Lbic_170097207        -------KLDEYDDHFFNLMPTLFPYNKFTMTKLIKRTVYTDH  4 VERQEILLNELAELARLGFAKAEEEW--------
CC1G_01547_Ccin_169854391               -------RLDEYDDHFFNLMPVLFPYNKFTMTKLIKRTVYNDH  4 QERQDELLKQLEELAKAGFAQAEAEW--------
GSTEN:00015692:G:001_Tnig_47218912      ------EHGGQLRSKVYTHLSSFLPCSRDTVLKRVKRLLLAHT 19 PPGVEGPLQKLKEAVGRAMPEQIARF--------
LOC100023118_Mdom_126340861             ------EQSPVIRSGVYSHLEAFVPCNKDTLVKRLKKLHLNVQ    DDRLREPLQKLKLAVSNVMPEQLFKY--------
D130059P03Rik_Mmus_158749567            ------ELGPVIRSGVYSHLEAFVPCNKETLVKRLKKLHLNVQ    DDRLREPLQKLKLAVSNVMPEQLFKY--------
KIAA2030_Hsap_156630841                 ------ELGPVIRSGVYSHLEAFVPCNKETLVKRLKKLHLNVQ    DDRLREPLQKLKLAVSNVMPEQLFKY--------
UBN1_Hsap_6940751                       ------ELSSQVRSGVYAYLASFLPCSKDALLKRARKLHLYEQ    GGRLKEPLQKLKEAIGRAMPEQMAKY--------
UBN1_Ggal_154937332                     ------ELNSQIRSGVYTHLAAFLPCSKDTLVKRARKLYLYEQ    GGRLKEPLQKLKEAIGRAMPEQMAKY--------
GSTEN:00008525:G:001_Tnig_47208786      ------EQPAEVRSAVYSHLEAFVPCNKEALLKRLKKLSLNIQ    DNRLHTPLLKLKLAVCNVMPEQMARY--------
NEMVEDRAFT_v1g202773_Nvec_156393686     --------------GIYDYLAFYLPCTKDTLLKRAKKLLLTDQ    AGRLREPMAKLKAAVDKVMPEQIKAFEEQVRKHV
AgaP_AGAP002048_Agam_158301289          ------ASGSSTRNAVFRHLEAQLSVSRQSLQLKLKKIRIRKL    EGRTKSLAAKLEDVIADTMPAIVTKY--------
LOC412133_Amel_110751190                ------LPSSSLRLETYGHLARFLPCSKVALQNRAKKLFVQDV    DYKTRDLIKRLKTVIDEIMPSVVCKY--------
Dper\GL23606_Dper_195144098             ------CTDRNERNMVFSHIEYQLQLPKYYILRKGKALRSKEE    KTRTTIALESLRRAVSESMPKAIASY--------
TRIADDRAFT_56741_Tadh_196005425         ------NVCRKMRENVCIYLKHHLPGCKMPVRSVIIKMMLNEV    KDDLNARMNELKEEIDKCMESQQKNYDKM-----
LOC100141652_Tcas_189236704             ------CLGKPSKMKVYEHLATFLKCKTETLTRRAKSLVLDDE    QKKLKNLLSGLKSEINNIMPSLLESY--------
yemalpha_Dmel_17136766                  ------CTDRNERNMVFSHIEYQLQLPKYYMLRKGKQVRAKEE    KRKSTIMLEKLRRAVAVVMPKAVANY--------
LOC579948_Spur_115772532                ------LVNCSKRSAVYSHLACHLPCTKVTLQKRAKKLQLGIQ    DNHLKEPLQRLREGIAQVMPDQEMRF--------
PHYPADRAFT_171726_Ppat_168059334        KQGKVPDELVVRLHGILGHIMRPN------TLKRNLKEMVEHG  3 KQEKEGRLLDIKREVTEMVRNQVTKFK-------
Os05g0422600_Osat_115463977             QGKIPEHELINRLMGIVGHLVHRR------TLKRNMKEMVQSG  3 KQEKAGKLQQVKMEIYEMVKARLATKPK------
AT1G21610_Atha_22329715                 SQGKHSTELINRLMSIVGHLIQLR------SLKRNLKIMIDMG  3 TREKDTRFKQINNEVLDMIKAKVSLMES------
consensus/90%                           ...............hh..l...........pb.bp..h...    ..p.p..h.plp..l...h.....pa........
```

```
Predicted Secondary Structure           HHHHHHHHHH    --------------HHHHHHHHHHHHHHHHHHHH------    HHHHHHHHH---HHHHHHH------
UM04022.1_Umay_71019877                 QAAHKAWVER 33 GEDENIEPTKRWRWSEEMREQVFVLITIENAMSELIN-EKH  7 QYSEINARK---ACYKQMAELFPED
CNBI1080_Cneo_134115561                 EVEVAAWDEK 27 EDIRPPEPQRRFVWTSEMRDIFWQLIENLRVTTELIK-KGA 11 EWTEQGLKS---RLYKKIVDIFPE-
LACBIDRAFT_319105_Lbic_170097207        ERSVVAWDRR 63 KDHGQHPPAKRFRMTETMKNIVWNLVLLSNECCRLEN-EKN  7 QVSEQGLRK---VLYQKIVAAFPE-
CC1G_01547_Ccin_169854391               EKSVAAWDKR 58 GGHAPHPPGKKYRLTEQMKAIIWDLVLLSNECCRLEN-EKN  7 QVSEQGLRK---VLYQKIVAAFPE-
GSTEN:00015692:G:001_Tnig_47218912      RQTCHLHEQV 27 GAKR--GPRKMFRWTEEIRECLSLVLQTKLEEKEKEK--SS    QELEDFLKA---LLENEVKPLWPK-
LOC100023118_Mdom_126340861             QEDCQARSQA 23 PGKRVIGPRKKFHWDDTIRALLCNLVKIKLGCYELET-NKS    QSAEDYLKT---FMETEVKPLWPK-
D130059P03Rik_Mmus_158749567            QEDCQARSQA 23 PGKRVIGPRKKFHWDDTIRTLLCNLVEIKLGCYELEP-NKS    QSAEDYLKS---FMETEVKPLWPK-
KIAA2030_Hsap_156630841                 QEDCQARSQA 23 PGKRVIGPRKKFHWDDTIRTLLCNLVEIKLGCYELEP-NKS    QSAEDYLKS---FMETEVKPLWPK-
UBN1_Hsap_6940751                       QDECQAHTQA 28 GGRRIMGPRKKFQWNDEIRELLCQVVKIKLESQDLERNNKA    QAWEDCVKG---FLDAEVKPLWPK-
UBN1_Ggal_154937332                     QEECQAHTQA 28 GGKRVTGPRKKFQWNDEIRELLCHLVKIKLDGYEYDK-NKA    QSLEDYVKT---FLDGEVKPLWPK-
GSTEN:00008525:G:001_Tnig_47208786      NMDCIAKVAK 18 PGKRVIGPRKKFEWDEKLRSLLCNLVRVKLGCYELEG-KNS    QSVEDYLKA---FMETEVKPLWPK-
NEMVEDRAFT_v1g202773_Nvec_156393686     EEEASASSST 48 EKTKKTVPKRKFIWSDELRELLHDVVTVKVKLFEMSK-TRT    TSAEEYLKE---FFDTEVKQLWPK-
AgaP_AGAP002048_Agam_158301289          ELDCLKVNEL 26 VTQQIRNPKKKYTWNERSRDLLWELYNVRQQLFALVR-PRN    QSEDEVLAE---FLRTKVVPLWPK-
LOC412133_Amel_110751190                VDECQKVAEG 20 DADKSKIPRKRFPWTEEAKKLVSEIASIKSQCYNVLR-PRK    LDMYTYVGS---FMVRQVLPLWPR-
Dper\GL23606_Dper_195144098             EVELHHFAEQ  7 -EVPPKMPRKKFQWTSELRNLLFTVYQARWTSYAVLG-KRK    DSLEDFING---YMKEKVVELWPM-
TRIADDRAFT_56741_Tadh_196005425         TGDNKENSDD 19 -DKK-RLPRKSFQWNETLRQLLEKIVILKEEFTKLSS-NRV    SLSDNWLTELKLFFDKEVKTLWPK-
LOC100141652_Tcas_189236704             EKESQRIIQK 11 -NKSLRLPKRRFQWNEVTKKLLKDILILKKRSLLFEG-KTK    ESLDTQMTT---FLKTEIQVLWPE-
yemalpha_Dmel_17136766                  ETELRTFAEQ  7 -ELPPKMPRKKFQWTSELRHLLYDVYQARWTSYAFLA-KRK    ESLEEFINW---YLKEKVVELWPP-
LOC579948_Spur_115772532                KTECEEVS--    ----------------LKIALLCEVVKVKVKTFEISK-TRT    QSAEEYLKA---FLDTDVKNLWPK-
PHYPADRAFT_171726_Ppat_168059334        GAEAREGSSD  7 SVEKAGTSTEKYKWDHAMEDLLCNLYDQYLEGMD-------    EHKGPQIRK----LYFELAELWPE-
Os05g0422600_Osat_115463977             GAEHKVESID  6 THDDRMALRGKSVMDAVLEDRICDLYDLYVEGMD-------    EDKGPQSRK----LYLEVRN--DQ-
AT1G21610_Atha_22329715                 QAIKPEGATS  5 SVEKPSL--KKFVMDAALEDKLCDLYDIFIDGLD-------    EDQGPQTKK----LYVNLAELWPN-
consensus/90%                           ..p....s..    .........cba.hs...c.bl.plh..b............    p.....hp.....h..pl..haPb.
```

```
Predicted Secondary Structure          -HHHHHHHHHH    HHHHHHHHH---
UM04022.1_Umay_71019877                GWTTSTNISRE    YAQIKKKKDRQE   731\Basidiomycetes
CNBI1080_Cneo_134115561                GYTNNNIISRE 25 VTKLKKRKEEAQ   803|
LACBIDRAFT_319105_Lbic_170097207       GWMSSGQISRD    VSVMKKKYEKEN   601|
CC1G_01547_Ccin_169854391              GWMSSGQISRD    VSAMKKRLEKEQ   734/
GSTEN:00015692:G:001_Tnig_47218912     GWMQSR-----    ------------   561\Metazoans
LOC100023118_Mdom_126340861            GWMQARMLFKE  8 TSAPAKKKVILP   689|
D130059P03Rik_Mmus_158749567           GWMQARMLFKE  8 TSAPAKKKVIPA   669|
KIAA2030_Hsap_156630841                GWMQARMLFKE  8 TSAPAKKKVIPA   683|
UBN1_Hsap_6940751                      GWMQARTLFKE  8 TSILAKKKVMAP   594|
UBN1_Ggal_154937332                    GWMQARTLFKE  8 TSVLAKKKVVAP   588|
GSTEN:00008525:G:001_Tnig_47208786     GWMQARILFKE  8 TGYIAKKKLVPA   484|
NEMVEDRAFT_v1g202773_Nvec_156393686    GWMTSRILFKE  8 TSAPRFKAKKAP   577|
AgaP_AGAP002048_Agam_158301289         GWIRYEDIQKE  5 KALAKGQAAAAA   648|
LOC412133_Amel_110751190               GYMTTHALLKF 22 ISATFSQEKVSG   588|
Dper\GL23606_Dper_195144098            GWMRYDELQRE  5 NSSRRSTEKPK-   651|
TRIADDRAFT_56741_Tadh_196005425        GWMNKRSKIGQ  6 SSSKGGKSKQIS   478|
LOC100141652_Tcas_189236704            GWMSMNALNKI 10 INSSLSSNSPNS   492|
yemalpha_Dmel_17136766                 GWMRLDELQRE  5 NAKLKAKEKPKA   651|
LOC579948_Spur_115772532               DWIQTRNLYKE  8 TYSVGKQRKATP   552/
PHYPADRAFT_171726_Ppat_168059334       GWMDNNGIKAA 14 NKGLKTPLCYPR   609\Plants
Os05g0422600_Osat_115463977            RMKRRRLAAAA 12 AQSLQNMTSTHT   581|
AT1G21610_Atha_22329715                RLMDYRGIKHA 15 AKEMDQTKMKKS   562/
consensus/90%                          sabp.p.h...    ............
```

Figure S1: Multiple sequence alignment of the alpha helical domain detected in HUN containing proteins of metazoans, basidiomycetes and plants. Proteins are denoted by their gene names, species abbreviations and gis separated by underscores. The predicted secondary structure is shown at the top of the alignment where, H: helix and E-strand. Shown below the alignment is the conservation profile at 90% consensus. Consensus abbreviations are as follows: b: big residues (KFILMQRWYE), h: hydrophobic residues (ACFMWYIVL), l: aliphatic residues (IVL), a: aromatic residues (FYWH), and s: small residues (GASVTDNPC). Species abbreviations are as follows: Agam : *Anopheles gambiae*; Amel : *Apis mellifera*; Atha : *Arabidopsis thaliana*; Ccin : *Coprinopsis cinerea*; Cneo : *Cryptococcus neoformans*; Dmel : *Drosophila melanogaster*; Dper : *Drosophila persimilis*; Ggal : *Gallus gallus*; Hsap : *Homo sapiens*; Lbic : *Laccaria bicolor*; Mdom : *Monodelphis domestica*; Mmus *: Mus musculus*; Nvec : *Nematostella vectensis*; Osat : *Oryza sativa*; Ppat : *Physcomitrella patens*; Spur : *Strongylocentrotus purpuratus*; Tadh : *Trichoplax adhaerens*; Tcas : *Tribolium castaneum*; Tnig : *Tetraodon nigroviridis*; Umay : *Ustilago maydis*.

2.1 Multiple sequence alignment of the Hir3-like TPR repeats

```
Predicted Secondary Structure              -HHHHHH---HHHHHHHHHHH    HHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHHHH------HHHHHHHHH
SS1G_01112_Sscl_156061996             1  MASFTALNIEPDDQSYDEIDN    TREIQIEDALKLYTAALKLHAQGLPYYPEAAEAYKNLFQSEIFSYPEAVTEFTRL 26
MGG_00105_Mgri_145611812              1  MPSFAAINVEPEENLDEDIED    TRQIQVDEALKLYQNALKLHAQGPKSYDDASKAYDQLFESAIFKLPEAITEYERA 27
NCU04512_Ncra_164429378               1  MPTFQAINVEPEENVDEEIDT    T---------KQIQNALKLHAQGPRFFDEASDAYNALFESEIFRYPESKTEYERT 27
CHGG_04295_Cglo_116195394             1  MAAFSAINVEPEENIDEEVDT    TKDLQVDDALKLFQNALKLHAQGPQSFDQAADAYDALFESEIFKYPESKTDYERF 27
PODANSg7663_Pans_171691398           15  QPAFQAINVEPDEIVDEEIDN    TRDIQVEDALKLFQTALKLHSQGPKFFDDAADAYNALFGSEIFKYPEAKTEYERA 27
SNOG_08212_Pnod_169610231             6  VSSWRPLNVEYEQEAEDEVDD    TQEIQIEEALKLYQTALKYHSEGPESFDKTAAAYRELFASEIFKYNESLSEYQRH 30
FG04030.1_Gzea_46116376               1  MPAFQAINLEPEDNIDEQIDT    TKEIHVDEALKRFQHALKLHAQGPRSREAAEAAYNELFESEIFKYREAKTDYERA 31
hip3_Spom_19113368                    1  MATFTPLNAASEDLDKEKR--    TLEIRIEEAVQIYQNALSAQKQGD--DNAATKYYDELLNVRILR-ELPFTVNSNY  1
BC1G_12898_Bfuc_154295790             1  MATFTALNIEPDDQSYDEIDN    TREIQIEDALKLYTAALKLHAQGLPFYREAAEAYKNLFQSEIFSYPEAVTEFTRL 26
AN3123.2_Anid_67525331                1  MSTWVALNIEPDEAIEEEVDD    TKEIQIEEALKLYQNALKLHSQGPRYYKQAGEAYEALLDSEIFKYPESLSDHKRA 30
KLLA0B03850g_Klac_50303527            1  MSSFAALNLSTNENADESKDH    SRELQVEESSKIFQLALIQLKAKR--FDDAKETFDRLFEIDVIK-PNKWGLYEY-
CAGL0M00286g_Cgla_50293891            1  MSSFSFLNQSPDIEQLEAEEH    TRETQIELCFQLFQDALERLKDDK--FGDANDLFEKLMSQEVIK-PDKWGLYAY-
AGOS_ABL081W_Agos_45185149            1  MSLFTALNSTPHDALLEAEEH    SRELQVEESFRVFQSALFHLKAKR--FDDAEAKFEQLFNIEVLK-PDHWGLYRY-
HIR3_Scer_151945205                   1  MSMFNALNSNIEGEQYEAEEH    SRELQIEQSFNILQDALIDLKNKD--FEKSDSKFQELFQIDVVK-PDRWGMYRN-
Lgig1000008602_Lgig_Lgig1000008602   54  KIRIAALNESSSDGSDSSSSE  2 TKEAKEAEVFGWYNKALVLQRQGN--IFDSESTFLQIIDHPFLDEAKELIENEVE  1
LOC412746_Amel_110758961              1  MIKISALNEESSEESEEEDVP  2 TKEAQ---------------------VYLLETELLDEVQKPEIPD----------  1
Cabin1_Rnor_166157458                 1  MIRIAALNASSTIEDDHEGSF  6 TKEAQEAEAFALYHKALDLQKHDR--FEESAKAYHELLEARLLREAVSS-GDEKE  1
Cabin1_Mmus_70995287                  1  MIRIAALNASSTIEDDHEGSF  6 TKEAQEAEAFALYHKALDLQKHDR--FEESAKAYHELLEARLLREAVSS-GDEKE  1
GSTEN:00021655:G:001_Tnig_47217598    1  -IRIAALNAASTAADESQDPQ  6 TKEAQEAEAFALYHKALDLQKHDK--FEESSKAYHELLKTPLLKEAMPS-DDHRV  1
PHYPADRAFT_139830_Ppat_162675462     42  QFHIAAINDVDTSRDFSEPLA  1 TKEAQESRLTQMYGEALSHLQHGQ--IEKAQSLFQSILQDPISIKAQVR------
Smoe1000001625_Smoe_Smoe1000001625    1  MYAISAINESENSKSRLEPLA  1 TKEAQESHLTKLYQDALSSLQSGD--LANARSLFQAIIQDPLSIKSQRG------
AT4G32820_Atha_15233957               1  MFSIAAINDTESTEKW-EPLA  1 SKEAQEFHLSQTYHDGLLKLQAKD--YDKARELLESILKDPIITNSKVETIA---
OSJNBb0024N19.17_Osat_50872476        1  ---------------------    -------------------------------------------------------
DDBDRAFT_0189500_Ddis_66827817        1  MIRFAAINSSNYSDSEYSEYE 37 SIEAQETYNMKLYSNALKHQQNND--YLQAQQLYQELLQSPMIERINPQSKKSI-
consensus/90%                              ...h.slN...p.........    obp.p.......h..uL...........s...h..lh...h............


Predicted Secondary Structure         --HHHHHHHHHHHHHHHHHHHHHHHHHHH-    HHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHH--HHHHHHHHHHHHH----
SS1G_01112_Sscl_156061996             TPSTLPQILYLAHKNHGQFILDCVKARLR 15 ESQAKVSMDDFGRALVSDESDTELWRRAARIGAMLGSRRISRYCLEAAVEVDDD
MGG_00105_Mgri_145611812              SATSLPQVYYLAYKNHGHFRLDRIRHEAR 16 QTEALGVLSDFEAALDRDPSDAELWRRAARVSAFLKSARVSRYCLEAAIELDDD
NCU04512_Ncra_164429378               IASTLPQTLYLSFKNHGQFIVDRIKHKAR 15 QHDAKKALDEFSAALDYDPSDADLWRKTARIAAFLKSARISRYSLEAAVELDDD
CHGG_04295_Cglo_116195394             IGSSLPQALYLSHKNHGQFVIDRIKHKVR 15 QKDAQRALDDFCVALDRDPSDAELWRKTARVAAFLKSARINRYSLEAAIELEDD
PODANSg7663_Pans_171691398            LSNTLPQAFYLAYKNHGQFIVDRIRRKIR 15 LEDARKALEDFSAALDRDPSDAELWRKTARMAAFLKSSRISRYSLEAAIELDDD
SNOG_08212_Pnod_169610231             APNTLPQILHLSYKNHGQFLLESMQYWVS 10 WNNIRGALNYFSEALDKEDTDLDLWLRAASVSAMLGSKRLTRYCLEAVLDGNDE
FG04030.1_Gzea_46116376               VAASLAQALYLSYKNYGQFFVDKFKDAST 13 HDDGRKVLDSWMTALDQDPSDPELWRKAARFAGAINSNRIKRYCLEAAIELDDD
hip3_Spom_19113368                    KNNALLLLHYLTRKNHGLFLLSKLQSLLS  7 IPVYRDALLDFAIALACDYNDIELWSFVAELAEKLEMPRIQRFALESSFYTGYE
BC1G_12898_Bfuc_154295790             TSSTLPQILYLAYKNHGQFILDCVKARLR 12 DIQAKVSMDDFGRALASDESDTELWRRAARIGAMLGSQRIARYCLEAAVEVDDD
AN3123.2_Anid_67525331                TSSTLQQTIYLSYKNHGQYVLDALRASLQ 16 AESSTTALAQFAEALERDDTDLNLWRQSARLCSALQSYRLTRFCLESVLADDDN
KLLA0B03850g_Klac_50303527            SSPTLDSLRYLAYRNRGILHYQYAKENFK  7 VEHILKTLECLLESLQHGEADNTVTHLLLKIFWSFNSTKLSRTILEYELTRSPE
CAGL0M00286g_Cgla_50293891            SSSIVDNLRYLAYRNRGFFYYKFLMNQYS  8 VDNILKIVENLLESLRHSEADYNVIEVLVQIFESFKSKRLQKLLLENEFMNNDS
AGOS_ABL081W_Agos_45185149            SSPTLDLLRYLAYRNRGVFYYQYVREHAE  7 VDYILKVVENLLQALQHSDGDAAVTGLLLQVFRGFKSKRLQRWILEYELTKQPE
HIR3_Scer_151945205                   SSPTLDNLRYLCYRNRGMYYHLYLENNYE  7 VNCILKAVENLVESIQHSDADFAVTDLLARIFKSFNSVKLERLISEYEFTKQEN
Lgig1000008602_Lgig_Lgig1000008602    AIHPGLALLYSVHKNLAGIALS-------    RQDEKTALQHYLEAVKIDESEVTVWFKIGKIAVNEHNYSLAKLAFEQGLVCNPN
LOC412746_Amel_110758961              KSRPMLSLKYSCFKNIGAIQAI-------    LGNYDEAIENYWEAANLDDSDVTLWYRIGTLALKISNLELACSSFKQGLKCNAN
Cabin1_Rnor_166157458                 LKHPGLILKYSTYKNLAQLAAQ-------    REDLETAMEFYLEAVMLDSTDVNLWYKIGHVALRLIRLPLARHAFEEGLRCNPD
Cabin1_Mmus_70995287                  LKHPGLILKYSTYKNLAQLAAQ-------    REDLETAMEFYLEAVMLDSTDVNLWYKIGHVALRLIRLPLARHAFEEGLRCNPD
GSTEN:00021655:G:001_Tnig_47217598    SKHPGLMLKYSTFKNLATMAAT-------    QDDLETAMEFYLEAVMLDSTDVNMWYKIGQVALRLVRIPLARHAFEEGLNCNPD
PHYPADRAFT_139830_Ppat_162675462      LRLCKSRLRYSALKNLAETFSMQG-----    SGHHVEAVDCFLQAAAIDGKDVVLWNRLGTLSCALGNLNVARRAFEEGLRCSPR
Smoe1000001625_Smoe_Smoe1000001625    KSDAMLHLRFLTYRNLAETYVKQG-----    PEFSAEGLHYYLQAVAIDDKDVVLWNRLGTLACSLGKLNIARKAFEQGLHCSPR
AT4G32820_Atha_15233957               NDNHLHHLRFLALKNLATVFLELG-----    SSHYENALNCYLQAIDLDAKDSVLWNHLGTLSCSMGLLSISRWAFEQGLLCSPN
OSJNBb0024N19.17_Osat_50872476        -----MADRFLTLKNLASVFLQQG-----    SAFYDNALRCYLQAVELDANDSVVWNHLGTLSCSMG------------------
DDBDRAFT_0189500_Ddis_66827817        LSSQMATLKYSTLKNMANINEA-------    IDNKELSLKLSTQAIDLDGRDVSLWYSIGKISMELEKWNISRLALERAIQISPT
consensus/90%                         ........ba.s.+N.u.h..........    ......shp.h..ul..s..D..l....spl...h....l.+..hE..h..p.p


Predicted Secondary Structure         --HHHHHHHHHHHHHHHH-HHHHHHHH-HHHHHH-----------HHHHHHHHH-------------------HHH---HHHHHHHHHH
SS1G_01112_Sscl_156061996             PTIAEVEPQSLEEGFAGE-QLKSLLQL-MSDELSLSHPIMG---P-FKKKTMPQF-------------------LKKHIDPYPFLPNPTS
MGG_00105_Mgri_145611812              PAVGDVEPPSLAEGYAGE-QLKNQLQV-LGDDMALSHPAMG---P-FVKYEVPDF-------------------LKRFLDPLPFLPNPIK
NCU04512_Ncra_164429378               PAVTDFEPPSLAEGFAGE-QLKRQLEV-LDDKMALSHPIMK---S-FREKGLPKF-------------------LERYLDPIPSLPDYAK
CHGG_04295_Cglo_116195394             PAVEEVDPPSLAEGFAGE-QLKQQLQV-LGDDMALTHPIMN---P-FVHREMPPY-------------------LARFMDPMPFLPDPTK
PODANSg7663_Pans_171691398            PAVDEVEPPSLAEGYAGE-DLKDQLQV-LGDEMALGHPIMK---P-FVERGLPPM-------------------LKRYQDPIPFLPNPAK
SNOG_08212_Pnod_169610231             LWENLLRLPGLEEGFAGH-QLRELVER-LEDNVSLMQAPLS---S-MKRKKLSDT-------------------IKKRLNPFSFAPLPAD
FG04030.1_Gzea_46116376               PAVMELDPPSLAEGMAGE-QLKDQLKL-LGDDMALSHPIMA---P-WLKKKMPAL-------------------LKRHLDPIPFLPDPTK
hip3_Spom_19113368                    PFDAERVFTNIDDLNPGK-LISLQNLY--NLLKKLGGIAEP--LP-QLDLSISSLYT-----------------LPSFFPQLPTPSFHRR
BC1G_12898_Bfuc_154295790             PTVAEVEPQSLEEGFAGE-QLKSVLEA-LSDELSLSHPIMG---P-YKKNEMPQF-------------------LKKHIDPYPFLPNATN
AN3123.2_Anid_67525331                RLEVRSEQLGLEETFAEE-GLRKTLHS-VQDRLSVSLVPIK-----KPKKALLKF-------------------FKQQSDPYPYLPSLPD
KLLA0B03850g_Klac_50303527            --------EIRLLSSLSA-MFKDQKRV-FKQYSQLLKKLHVTSLPQNIPTAVEAVL------------------NCIETCDVEEGQIE--
CAGL0M00286g_Cgla_50293891            --------HLLFSGRKRTNILPQLKVL-ISQYINLLKDLHA---SSVLEKKIICNG------------------QAKNTSHLPRTNVKFK
AGOS_ABL081W_Agos_45185149            --------EMLLLGRGRA-LLPDTRGL-VEEYRRVLRDVR----SEPVTEFAARVL------------------TLVKTDSLEVQPLP--
HIR3_Scer_151945205                   --------LSLLLGRHRKFLLNDLTLM-MNNYVELTNKLLV---PNLSDNTIFERY------------------HLEKYKDIKPEPLAFG
Lgig1000008602_Lgig_Lgig1000008602    -HWPCLDRIVTITYAYNN--MWNCLEY-IAKSLELDSSYGK---GLAIRDQIFKEQ---PELKSMSKRL---FSLCDPVVERNKCEKKEAE
LOC412746_Amel_110758961              -HWPCLDNMITALYAVPD--YMNCLLY-ISMALERDPTYVK---GLAFRDKIFKDI---PCLEECYKLYNNDWQLDPPLYTQY--DHLIGD
Cabin1_Rnor_166157458                 -HWPCLDNLITVLYTLSD--YTTCLYF-ICKALEKDCRYSK---GLVLKEKIFEEQ---PCLRKDSLRM---FLKCDMSVHEVSVNAAETQ
Cabin1_Mmus_70995287                  -HWPCLDNLITVLYTLSD--YTTCLYF-ICKALEKDCRYSK---GLVLKEKIFEEQ---PCLRKDSLRM---FLKCDMSIHDVSVSAAETQ
GSTEN:00021655:G:001_Tnig_47217598    -HWPCLDNLITVLYTLSD--YTWCLYF-IAKALERDHCYAK---GLVLKERIFEEQ---PSLRRDTMQM---FMKCHVSIHYVEVEEEERE
PHYPADRAFT_139830_Ppat_162675462      -HWSCMEKLVEILIAVGD--ESACLSV-AKRLLKLSPSHPR---ALQIQHAIEQGTSSVLKIGPRGFDL------LQPEYFGLSFDNKRKL
Smoe1000001625_Smoe_Smoe1000001625    -HWSCMEKLVEVLIAIGD--ETGCISV-VKRLLKFHPSHPR---ALHIKSVIEGKH---HGTGIAGMDG------LEPTHTNLSFQRKRKL
AT4G32820_Atha_15233957               -NWNCMEKLLEVLIAVGD--EVSCLSV-ANLILRHWPSHSR---ALHVKHCIEDTDS--APFAPKGIDK------LEPQHVRLKFLGKRKV
OSJNBb0024N19.17_Osat_50872476        ---NCMEKLLEVLIAICD--EVACLSV-AKLILRSWPSHHR---ALHVKKTIEDAEP--VPFAPRGIDI------LEPKHAKLKFCNKRKS
DDBDRAFT_0189500_Ddis_66827817        -HWLSLEKLIELLYVIGD--DNSCKNIIEQRILQEQPSNLK---AIIIKLLIQRSNQ------SNYNEINEILNNILISNQNSKSYIQSLI
consensus/90%                         -................p..b.....h.h...............s...p..h..............................h........
```

```
Predicted Secondary Structure          ---------------------------HHHHHHH------------------------HHHHHHHHHHHHHH    -EEEE----
SS1G_01112_Sscl_156061996              TMSRSI--------------------DFDEPLQI----LEGRAIINVPER---------SWTSIGHEICGAAQ  5 GVILALPPG 35
MGG_00105_Mgri_145611812               EIAIPK----------------------STKDEE----SLSRLILVAQVA---------DWTNVGLAIIAAFL  7 GVFIEVPEV 41
NCU04512_Ncra_164429378                TIAAPR----------------------PDSDEV----GQPRLVMNIPSY---------SWTELGIALVHFVA  7 AVAIQLPEG 41
CHGG_04295_Cglo_116195394              RLAVKE----------------------LRDTEI----ETARSVLPISIP---------SWSELGAAIVRFVT  7 GVLIQLPDT 40
PODANSg7663_Pans_171691398             SLAVPK----------------------KQTTDV----PPPRHVISAASS---------SWTELGAAMVRFVE  7 AVFLETPDA 40
SNOG_08212_Pnod_169610231              VAHADV----------------------RPMLTQ----PAQRVVLTPTKW---------DWAGLGETILQQFM 14 SINFNIPPD 41
FG04030.1_Gzea_46116376                SLIPPKV-----------SDTQAQTQEDTEMGDA---DNASLTEPKSVS---------SWAELGAKLIN---  5 AEAFDTCDG 40
hip3_Spom_19113368                     SINLCP-------------KI-KKLQLSHDTLHGFL--DLVLYALQINEKKT-------PTRGFPSTLIIH--  8 AESHDLEST 33
BC1G_12898_Bfuc_154295790              TLNRSM--------------------DHDVPLQI----LEGRAIINVPER---------SWTAIGHEICQ---  8 GVILALPQG 39
AN3123.2_Anid_67525331                 NLQDLD--------------------SSRNPLAF----RASTHEIKIDSL---------TWAAIGQAILTFLD 12 SITISLPAD 62
KLLA0B03850g_Klac_50303527             PLLKDI----------------KAMKTEDDEILK----ELDGSNLNLNEL---------SWEELGVSLRS---  5 KVVNILGRK 26
CAGL0M00286g_Cgla_50293891             NILEQL----------------NKLRTEDDELIK----SLDIHDIELDNY---------DWTSLLRSLRE---  5 KSLALFGRN 28
AGOS_ABL081W_Agos_45185149             PVLDKI----------------EQLKDEDDAMMK----QLAEYEVNLEEL---------SWECIAESFRS---  5 KYTNFFSKA 32
HIR3_Scer_151945205                    PILSRI----------------SEMKKQDEEIMK----KLDVFNVTLNEE---------SWDEVAKALKN---  5 KTSSLIGRN 32
Lgig1000008602_Lgig_Lgig1000008602     EYIKEA----------------INIRDKRRELAK----PQPPPVIKLPEKIDKY-----TWKRVGECLVSLIN 76 PKDIEKPGP126
LOC412746_Amel_110758961               KLIAEA----------------KDIAKKWAEACKKEFCPKPLPDLTLRKPIKNY-----TWLDLGECLLDMHR 17 QNSEEISKT 59
Cabin1_Rnor_166157458                  AIVDEA----------------LGLRKKRQALIV----REKEPDLKLVQPIPFF-----TWKCLGESLLAMYN 12 GKRIDLSD- 42
Cabin1_Mmus_70995287                   AIVDEA----------------LGLRKKRQALIV----REKEPDLKLVQPIPFF-----TWKCLGESLLAMYS 12 GKRIDLSD- 42
GSTEN:00021655:G:001_Tnig_47217598     SVLEEA----------------RELRRRRQMLSR----CEPQPDLLLIQPVRSL-----TWKHLGESLLAMYK 12 GRRIDLSE- 47
PHYPADRAFT_139830_Ppat_162675462       ENAYET----------------VQHQKKR---------KVHSIDVHLPEA---------SWFALVNAVTDVLK  1 VSREGLGVA 19
Smoe1000001625_Smoe_Smoe1000001625     DVVPDT-----------------SKRKKR---------KQYSVQLTLQEP---------SWLLLVKATLESLQ  7 LKSEELKEK 21
AT4G32820_Atha_15233957                SDMNQD----------------MDATSKK---------LRKRVQFKLPEA---------SWVALLNILIGIVH  1 -SRETVGIS 20
OSJNBb0024N19.17_Osat_50872476         GDDETH----------------HETVTKK---------SRQNAKLQLTEA---------KWMALLDGILSFLS  4 KSNEDHGAN 34
DDBDRAFT_0189500_Ddis_66827817         KKRDRIQQQQQQQQQQQQQQQQQQQQQQQQQQQQMIIGKQPQLTNELLDKPVEYILENKNWSSLLDLLNSIYN 68 TDTSITTAT131
consensus/90%                          ..........................................h.h...........sW..l...h.....    ......s..
```

```
Predicted Secondary Structure          ----------------    ---------------------------    HHHHH--HHHHH-------------------HHHHHH
SS1G_01112_Sscl_156061996              PTTNNSDERA------  9 KSSDAAL-----PTRKRSQSAAGLRDTP  4 STHKR--SRRIRNRES-TAET-------PIDPTTQYA
MGG_00105_Mgri_145611812               SDTQQTAANTDSTPAT 13 KDSTASL-----GFRKRSQSEAGLADGG  4 VDSKR--PKRARRRDT-AVEE-------VVDPSTIIA
NCU04512_Ncra_164429378                QDTEEAESIVRETKVT 11 KDRTSSL-----PTRKRSQSAAGIPDPG  4 GETKR--SKRTRRRET-APEE-------AMDSATLLA
CHGG_04295_Cglo_116195394              SEEATVEGLTSAPAPE  9 KDPNNSI-----STRKRSQSAAGIPDAL  4 ADVKR--SKRTRRRDT-AVEE-------VMDSATFLA
PODANSg7663_Pans_171691398             SEKPQAAEEESTPMVA  8 TGKERAL-----PSRKRSQSVAGLPDPP  4 AEGKR--SKRTRRRET-GRRE-TAAEE-IVDPARLLA
SNOG_08212_Pnod_169610231              KSAEDDIDKSIVQAQE  9 SSSAIALDAVPAPSRKRSTDSAGLPETA  4 SR-----SKRIRARDT-ITEPSAGGDAVAQEANKQLD
FG04030.1_Gzea_46116376                KDSTPKEQKAGSEGTE  9 QKDSNGA-----SARKRSQSAAGLPDGA  4 AMEKR--SKRVRRRETLQAEE-------TTDPNTLIA
hip3_Spom_19113368                     VPPPSDNLPKPQLLKR  7 ---------------------------    ---RI--SKRSRGRDLRTPSE--------SNLFSLIA
BC1G_12898_Bfuc_154295790              PTTNNSDERT------  9 QLSDAAL-----PTRKRSQSAAGLRDTP  4 STHKR--SRRIRNRENTTVET-------PVDPAAQYA
AN3123.2_Anid_67525331                 RAEKQLIESLEVQSIQ 11 KAEEDDL-----KYPENTSRKRSSGSAI  4 AERLRVKSRRTRLRDS-LAEASTHTDDVVFDQAKYYE
KLLA0B03850g_Klac_50303527             DSSNESPGHSPEHFED  9 NIEVSVE-----REPSAKRQDSTDR---    PAQRS--SKRFKEREKVQEPD-------VEQQVEFHT
CAGL0M00286g_Cgla_50293891             DNSNNENEVTEEK---  5 KVVTTNG-----NNRKRSITEIETK---    HAQRS--SKRFRDRDS-SAAD-------SETPMTFEP
AGOS_ABL081W_Agos_45185149             VQNEREPFAATQAATE  9 ENGGRTE-----SKRDSELPSDSQR---    PAQRS--SKRFKERS--TETN-------EAELLRPHQ
HIR3_Scer_151945205                    ERQEEEQDNESVRADD 10 QTNEEAR-----PNKRTDEHIDSTK---    PLQRS--SKRFKERE--QENS-------KELVMDVHK
Lgig1000008602_Lgig_Lgig1000008602     ISSSQSSFKEDSKVSS 44 QMSTTQE-----SSNTQTTNQDNVDSDC 96 GNKRR--SARVKHTK--DQK--------QEESVNFQE
LOC412746_Amel_110758961               LDTDAVQFEEQSQIKN 32 NISDKDS-----NKSDEKNLKKEDKPQC 37 NSMRR--SARVRSSN--RREA-------ERDDVQLEE
Cabin1_Rnor_166157458                  VTTTSFPLHSPGLLDT  5 DVSGGDK-----SKKGVKRKKTLEESGE    TAKRR--SARVRNTK--CK---------KEEKVDFQG
Cabin1_Mmus_70995287                   VTAASFPLHSPSLLDT  5 DVSGGDK-----SKKGVKRKKTVEESGE    TAKRR--SARVRNTK--CK---------KEEKVDFQG
GSTEN:00021655:G:001_Tnig_47217598     AQPQTSLTQSPVEVTT  7 DVSLADK-----VKKAPKRKRVVEDIGE    TAKRR--SARVRNTK--CK---------KEEKIDFQE
PHYPADRAFT_139830_Ppat_162675462       KDVENASVHQKQQGTG  1 KQEASKL-----PDLMKPSDPTTEWDQP    SQERR--STRLEKLRS-CRRL-------DKEGEGFSH
Smoe1000001625_Smoe_Smoe1000001625     SGQTSAKLTDDLLLNN  1 GCDTVSL-----AEKISPSKLSEVEDSH    AHERR--STRLRSRIK---------------GEEA--
AT4G32820_Atha_15233957                KKKDHCVDSDSSNVSV  1 DCNIERE-----SGGSVKEKEPVFSEEH    PQERR--STRLERLRN--QKP-------EKEGLEFDN
OSJNBb0024N19.17_Osat_50872476         GGNEHDLYHDGESVLS  1 DCRTA-----------VKEKDTNSDREH    PHERR--STRLERLRS--RKS-------GKDEHE-SN
DDBDRAFT_0189500_Ddis_66827817         HSTTTTTTTNSNDKLD 46 LIISVID-----WMLEALNTISREANQF  5 LINKK--IMKIKSIIT--QNT-------TSNNSDYQD
consensus/90%                          ....p..........    ............................    ...+p..S.R.c..p...............p..p...
```

```
Predicted Secondary Structure          --HHHHHHHHH--------------HHHHH-----------    HHHH----------------    --HHHHHHHHHHH--------HHHHHHHHHHHHHH
SS1G_01112_Sscl_156061996              DVFTFVGNMLQKLG-VEDLGTREDLQVALA--VDTATESNE 43 GMSA-----FLEHSGSISQK  8 AEFGGISWFAQRVEK--DWMPLQDVVFEWLRIVLA
MGG_00105_Mgri_145611812               ELFQLIQETTKTLG-INDPDASKKISEILE--SCSSEDRLS 34 GLNA-----FLEHTKPGTSP  7 NETAGLKAFIARING--GWFTLEDCAFRWLSALCH
NCU04512_Ncra_164429378                NLFQMTKNFLENLG-VTDRDTLDRITEILD--SCASDDR-- 16 SLSA-----FLEHSKSPSQR  7 DDVRNIRGFVMRMNS--GWNTIQDVVYEYVKSLAE
CHGG_04295_Cglo_116195394              NLFQTTKNLLENLG-VTDQVTLGRIAEVLD--SCAADDRTG 34 SLSA-----FLEHAKTGSQR  7 DDSRGLKSFVENVNQ--GWNTIHDIAYKYVKALVK
PODANSg7663_Pans_171691398             NLFQTTKNLLENLG-VTDHVTLERIAEVLD--SCASDDRMG 34 SLSA-----FLEHTKSGSQR  7 DETRGVRAFVEKINN--GWYPVQDVVFEFLKIIMP
SNOG_08212_Pnod_169610231              CLYEIVKDIYARLG-VEDVQDPTNLRGLLA--NLPPPPSCS 39 GLNA-----FLGYAKAGTGQ  7 ENQK-LGTFARIVNE--GWLSTKEVAFAWVEFLLS 17
FG04030.1_Gzea_46116376                NLFQMTKSVLEHLG-VEDKSTFAFITELID--SCATENRPG 34 GPSS-----FLEHAKTGSQD  7 NEKQGLRSFARKIGGCRSWMTAEDVAYEWVRAVSQ
hip3_Spom_19113368                     SIFEDYHQLLVNFP---SEGSIPDLDVSTD--SAANGAFSK  5 ----------LAMHSANRLE  3 VPDGYLLQLLSEVNS--LNMVPAELATFFVESMLR  7
BC1G_12898_Bfuc_154295790              DVFRFVGNVLQKIG-VEDLGTREELQVALA--IDKATERND 43 GLSA-----FLEHSNSASQN  8 AEFGGISHFAQKIEK--DWMPLQDVVFEWLRTVLA
AN3123.2_Anid_67525331                 GVFGTVGALLSKLG-VEDLGTFEELRRSVA--SAGERKDSP 49 GLAV-----FLEHSKKTTHK  7 SGEDELFDFIRAVNG--SRLHLHDVVFEWLGCLLR 21
KLLA0B03850g_Klac_50303527             YIINEINQSLHVSG-IEKDISIEELDPNNV--AKRQELCMS 34 QLTM-----LLRSSMFSGDV  8 IEAEDVRRFIRDVND--KKLHFHAVRFLFIKELLK  7
CAGL0M00286g_Cgla_50293891             MFYS-YTKILSLLG-CKAPYDYEAIYPDLM-EVKANDIIYA 39 HFKT-----LLNSKNFQKSD  5 ISQTSLRRFLDKVNR--QQCHFHELRFKLIWFLFD  6
AGOS_ABL081W_Agos_45185149             DFAALFYSCTGILE-VRLEIELSHLNPEAI--PSTYPAYMP 38 QLTS-----LLRSSMFTEEN  8 LPYKEVAEFISTVNS--NRLHFHAVRLLLLQVLLT  7
HIR3_Scer_151945205                    RFFGEFNTLLSYIH-ILPFCDFDTFASKFI-IGSSDKQPEK 39 QLNA-----LLKSNAFDDKE  8 LDSDHIRSFISEVNA--GNLHFHQVRLKLLFKLLG  9
Lgig1000008602_Lgig_Lgig1000008602     LLQKFLPSSLVQND--EEDDEEMRSDETES----------E  1 ETNKKDVV--VKTEDVEYLK  1 TEDKDVREFIKSCI---KNGGIVHLCIKYLITLAH
LOC412746_Amel_110758961               TLRRIFPSTLLPDTVKLTKDDPSKNMDDSM-DTMDLYQLFA  1 QENNSNIIETVKSSESSKSP  1 PDMSQQQNYFGTES---KYTEF--LCT--------
Cabin1_Rnor_166157458                  LLVKFLPSRLRKLD-PEEEDDPFNNYEVQS--EAKLESFSN  1 GPHR------LSFDSATFME    SEKQDVHAFLMENL---TNGGVLELMMRYLKSMGH
Cabin1_Mmus_70995287                   LLVKFLPSRLRKLD-PEEEDDPFNNYEVQA--EAKLESFSN  1 GPHR------LSFDSATFME    SEKQDVHAFLMENL---TNGGLLELMMRYLKSMGH
GSTEN:00021655:G:001_Tnig_47217598     LLYKYLPSRLKKFD-LDNDDESLSNLETQC--DDK-DDLQL  1 HSSS------LPAGPVEYME 23 AEQQDVHSFLLRNV---SNGGILDLMMHYLKAVGQ
PHYPADRAFT_139830_Ppat_162675462       ALKKALEPFVI---------------------------YIL  1 PSTN------VTEESLSAVN    PKSITVLRFLNEVA---SNSGIFHVAQQLLERVVA 15
Smoe1000001625_Smoe_Smoe1000001625     SLTQLLEPFIIDSSRLEVEVDPEASVITHCERVDGSLAYAA  1 ASSS------SPEGSASCVT    VEAREVTQFMQDYA---NNSGLYHVGHAILERVSS 15
AT4G32820_Atha_15233957                DILQYLEKFVLKR-----GFD-----------------RES  1 GSFC------NEES---DPI    SEHAVVSNFVKENL---ENYGAYHMGHLLLEYIAN 17
OSJNBb0024N19.17_Osat_50872476         AITQFLDSFILKRTSIPEKID-----------------CSG  1 GDAS------TPEALTYTPD    REASDVKQFLSKIS---KNCGPLHIGYMLLEEIAQ 15
DDBDRAFT_0189500_Ddis_66827817         YLLLSELNLDKYQQFLFEKKNKHENINNNNNNNNNNNNNNN 12 SDNSNSNNNSVNSNDTTTTT  1 TTTNNKPNIITDGD---YIIEIPTVIIVNEDAYKP  5
consensus/90%                          .h..................p....................    ..........l.........    .....l..ah.p............l...hl..h..
```

```
Predicted Secondary Structure         HHHHH-----HHH----HHHHHH-------------------------------------HHHHH----HHHHHHHHHHH-----------HHHHHH
SS1G_01112_Sscl_156061996             TYLQRQWSAELKV----AVVRMT--------SFLDADIYDR--VRSELELSRSSGEN-STQISELE----EMVETIFELHL-----------DVYSRI 13
MGG_00105_Mgri_145611812              SYTASRWSEERKR----A------------------------YESQNHQDSNPSV--VENQPTME----ALAQMLFELHL-----------DIYERV 13
NCU04512_Ncra_164429378               SYSTEKWSDQMKV----AIVQVI--------SRLDEGIYDRVVYDLEQWQLSDASG--HQPSPDIV----DLVYMLFELHL-----------DVYERI 13
CHGG_04295_Cglo_116195394             TYTEYKWPDQLRT----AVIDVT--------GRLEGSIYDSF--AYELELSPKDVQ--GETISELA----GLAQTLFELHL-----------DVYDHI 13
PODANSg7663_Pans_171691398            TYTGTKWPDHTKT----TLSDVI--------TQFGSSIYERV--TYELDLCLAKGD--HEAHLELV----KLVHMVFELYL-----------DVYERT 13
SNOG_08212_Pnod_169610231             SYMQFKWSDELKR----NLVQII----VNLDDYIHECMLERI-ASLNTRILAEEHQSCTYNFTTFDESQAEIIETLFELHL-----------DIYSLI 13
FG04030.1_Gzea_46116376               SYATAKWSDKMKL----SVVQML--------NKVDSTLYRKI----TEEMELAAGS--AERLADLE----TIVPMLFELHV-----------DIYELI 13
hip3_Spom_19113368                    FYLQQCWGKIFKK----KFTTIC--------ERIESTLHELV----------KASL----QTPEVF----NISQSLFELFL-----------DDYFLA 22
BC1G_12898_Bfuc_154295790             TYLQRQWSAELKV----ATVRMI--------SFLDADIFDRV--QSELELSRSSGEV-STHPSQIE----EITETIFELHL-----------DVYTRI 13
AN3123.2_Anid_67525331                SYVAYQWPTTLKD----VVLQLLSLEDEYIHGKLEEGMQTLEHHILEAQSGTPFRYT-AKHFADLE-----MIQAIYELHL-----------DLYAPM 13
KLLA0B03850g_Klac_50303527            LVTDYFWSPALY-----EVVEFF-------SMAIEVNLYQMI---------------EPSTDESTV---SIAVSIFEIMTNMMG-------NICNDL 18
CAGL0M00286g_Cgla_50293891            LLTSTRWTKQMY-----RDVEWI-------TLSIEKNIFSFI----------------QSDIAKYR----HVALSIFETLINSLV-------ELCEEI 18
AGOS_ABL081W_Agos_45185149            LITDTFWSPILY-----DTIESF-------VLSLESNIYDLV----------------YSQSEKYK----GLGLSFCEVLMNSLG-------GIYTEI 18
HIR3_Scer_151945205                   LIIDYLWESQLL-----KIVLWF-------VFGIESNIFALI----------------NKNKRQCK----YLALSIYELLVNHLG-------NIVEEI 18
Lgig1000008602_Lgig_Lgig1000008602    -QDHKIWQEEVHK----SFLDVF--------KTLRPI-TSFP------DLFSESIP--IDMKYDFS----KIGLVGCELLL-----------DEYVRI 19
LOC412746_Amel_110758961              -KWNQQWSKGILD----IYLQAY--------TFMREH-IPHP------SPFEDIND--DVLKLD-A----EMTLLFGELHT-----------DKWL-- 11
Cabin1_Rnor_166157458                 -KFLLKWPPGLAE----VVLSVY--------HSWRRHSTSLP------NPLLRDCS--NKHIKDMM----LMSLSCMELQL-----------DQWLLT 43
Cabin1_Mmus_70995287                  -KFLLKWPPGLAD----VVLSVY--------HSWRRHSTSLP------NPLLRDCS--NKHIKDMM----LMSLSCMELQL-----------DQWLLT 43
GSTEN:00021655:G:001_Tnig_47217598    -KFLKEWPQGLPA----VVLEVY--------QTWRKYSGGLP------NPLLRDCS--NQHIREML----AMSLACMELQL-----------EQWSHN 42
PHYPADRAFT_139830_Ppat_162675462      EKLTRMWAAERSAACSLFLAEVY--------MDMATSATNDTTA----SKCFHDCN--YNVCRLIE----NATSRGTD-------------------- 29
Smoe1000001625_Smoe_Smoe1000001625    EKYTRHWGNGRSSRCNLFLAELY--------LDAASSSAEPG-------KLLDHCN--YNICRVIE------------------------------- 18
AT4G32820_Atha_15233957               EKLTRHWGRDRKPECSLFLAELY--------HDFDSKRSDIPDA----PSCMVEVT--YHLSKIIE----SVSLDYAI------------------- 26
OSJNBb0024N19.17_Osat_50872476        DKITRGWAQDRSAQCSLFLAELY--------YD-QALCSGSPLA----SSELSNSS--YHLCKVIQ----SVALELPFRTSDGAAKSTNLDLNMESHM 39
DDBDRAFT_0189500_Ddis_66827817        --NLSFYLQNLNLK---KFIKPF--------KLL-PI-IEND------KENENQME--TNHHVEKE----DLSTIKDELLF-----------EIRYKW 11
consensus/90%                         .b....W...b.........p............h.................................h.....-....................
```

```
Predicted Secondary Structure         -HHHH-HHHHHHHHHHHHH-------------EEEEEEH---HHHHHHEE-----HHHHHHHHHHHHHHH    -----EEEE---
SS1G_01112_Sscl_156061996             MTKDRLDRWANFAAFVMRS-----RQHERDDELSLRYLW---ASVFYATVPGESSREHKVFLWTNLRDILN    EAGKPPILL---
MGG_00105_Mgri_145611812              ESRTRLMRWLFLRSQLFDS-----GKAQPSEELTLRFCW---AVVFATTLSDTVSPDHVLDSWRSLREVVE    QAGTAEIFL---
NCU04512_Ncra_164429378               KARIRLGRWWDLATQVSRR-----RYFGTDDALTLRFLW---AAVVFITVTEGVPREHILNCWHSLRDHLA    TTTVPDIVL---
CHGG_04295_Cglo_116195394             EAKCRVDRWADMAGEVIRL-----RTLEPDDPIALRFLW---ASVASTTMAKGTPRDHVLNCWHSLRDFLL    GSAATEVNL---
PODANSg7663_Pans_171691398            EIQLRVDKWMDLTTEVTRQ-----RNPNTDDELSSRFLW---AGVYATTLAKDTPRDHILNCWHSMYDHLL    ESSISEITL---
SNOG_08212_Pnod_169610231             LQSDRLERWSTLAREIVQL-RSACRSNLNMDDLNMRHIW---ASVFQMSVNDDIPPEHVVSAIAELKAILD    AQGDRTIQV---
FG04030.1_Gzea_46116376               DTKYRLGRWLDIASAYLRL-----IDRPANDPLIARFLW---ASVLVSSHAEQPLREHILLMWTSLRDHLA    DEKVPAISL---
hip3_Spom_19113368                    YKKLRCLRWKSLTEQVVEL--QPSCKSSSQRHLIIRNSWARNLLRLTGCSSEAIVENFKQLRCLLQENSD    -----SLEL---
BC1G_12898_Bfuc_154295790             MTKDRLDRWADLSAFVIRS-----RNSEQDDELSLRYLW---ASVFYATVPEDVTREHKVLLWTSLRDILD    KAGKPPISL---
AN3123.2_Anid_67525331                LQQDRLARWSMLARSALTHFIDCCPERVNRERITIRHIW---ASTFHSNMGVDAQREHILLCLEDLKRLFS    CLNEPVLSL---
KLLA0B03850g_Klac_50303527            KLQRKIVRWYEMLLSKPLS-----------TKLKFRFQW---SHFSCLQCTTDVTNSTLITTIEDIENSLK  1 FPDFGTVAY---
CAGL0M00286g_Cgla_50293891            KIEKKITKWNSIIAQVETT----------DRDWTVRYLW---AYYTLIQCTSDIQNCLLSNILEKVHQILM  1 DNSSTVYLY---
AGOS_ABL081W_Agos_45185149            KLEKKIDRWVLLLDQLVF------E-----GKLKFRYLW---SKFCYLQCISDVTDDRIIHSLDYITHELR  1 SSMDIDISY---
HIR3_Scer_151945205                   KVEKRIRSWHTLLEQIAD------EKD---KELYVHFQW---THYCFLQYTCDIVDSRLSETLTSLENTIK  1 SDSSLDIAY---
Lgig1000008602_Lgig_Lgig1000008602    RYLRE-DLWYILFCIGFP------TVIQDSPSYCIRVHW---LRGRFYLLQGDI--EQAVVCFEQVLYYLK  4 KESKLFIKL---
LOC412746_Amel_110758961              LDKLG-TGLP--------------------SEELGVLW---LKANIFLCQGDT--DIVIETLELLLNHMQ  4 QNQNIFLKL---
Cabin1_Rnor_166157458                 RDLFE-DGWL--------------------EFVVRVYW---LKARFLALQGDM--EQALENYDICTEILQ 11 EQRDIVIRL---
Cabin1_Mmus_70995287                  RDLFE-DGWL--------------------EFVVRVYW---LKARFLALQGDM--EQALENYDICTEILQ 11 EQGDIVIRL---
GSTEN:00021655:G:001_Tnig_47217598    KDLFE-NDWL--------------------DVMVRVYW---LKARFLALQGDM--ELALESYDVCTGLLQ  6 EGKHYTISL---
PHYPADRAFT_139830_Ppat_162675462      PLDVR-YDW---------------------SFWARFHW---LSGRLWVHSGQW--EQAHREFERCQSILE  4 TGSCALIFL---
Smoe1000001625_Smoe_Smoe1000001625    SIERD-AGW---------------------IFWARFHW---VYGRYCMLAGNP--KKGFSEFRKCLCLLQ  2 RERYTAIVL---
AT4G32820_Atha_15233957               EVLDY-DTR---------------------SFWARYFW---LSARLSILEDNK--AKALEEYLRCLSLLG  1 EGIGEAPVLIQR
OSJNBb0024N19.17_Osat_50872476        DSSSN-TNC---------------------VFWIRFFW---LSGCLSLSSDCK--EKAYKEFNIALSLMR  4 AKINREFVLL--
DDBDRAFT_0189500_Ddis_66827817        DINMA-TLYFKE------------------------CELLF-DDYNKRNFTTTETT--PTTTTTTTTTTNSNG  4 KNKLDYMII---
consensus/90%                         ....p-..a.......................h.h+..W............p...p.....hp.h...h.    ......h.h...
```

```
Predicted Secondary Structure         -------HHHHHHHHHHHHHHHHHHH----HHHH-------HHHHHH------    HHHHH----HHHHHH-----HHHHHHHHHHHHHHHHHH-----EEEEE
SS1G_01112_Sscl_156061996             TNNAVIAEISTSAADREVSSLTTMEFF----QNLFQPGRSDPLAIIQTLEPVLD 26 TPTIL----RDMWKFLDC-GSTPLRLFLWQRLSEAYASIEYKTKVFSC
MGG_00105_Mgri_145611812              PNNVVMPEVSVKAVDREISKLTTMDFF----LGLFKEDMGDPVAVISSLEPVLN 34 ATQTL----QDLWKFVRG-SSTDLRLFLWKRLGDAYRAIKYNTMQFSC
NCU04512_Ncra_164429378               PNNAVMPEISVAAADREVSKLTTMEFF----LGLFQEEQDNPVSVIDTLEPVLN 35 ASQGL----QDLWNFIEK-SSTDLRLLLWSRLGDAYGKIKYTTKQFSC
CHGG_04295_Cglo_116195394             VNNTTMPEISTAAAEREISKLTTMDFF----LGLFQEDNGNPVAVVESLEPVLN 34 ASQEL----TDLWKFLKG-SNTELRLLLWSRLGEAYGKIQYTTKQFSC
PODANSg7663_Pans_171691398            PNNAVMPEISTAAAEREISKLTTMDFF----LGLFQEDMGDPTSVIDSLEPVLN 34 ASQSL----MDLWKFLKR-SSTELRLLLWSRLGEAYGKISYATKQFSC
SNOG_08212_Pnod_169610231             QNNAIMPEMCIAAVERELVRISMKDFF----LKVFDHDEKDPVTVIESLEPILE 41 QPSPI----VEMRKFLDA-ANVNVRVSLWHRLREAYEAIEYSPKVLSC
FG04030.1_Gzea_46116376               PNNVVMPTISPAAADREISKLTTMDFF----LGLFQDNMDDPVFVIDTLEPVLN 62 ASQSL----KDLWKFLHN-SSTELRLFLWSRLGDAYDAINYSTKKFSC
hip3_Spom_19113368                    LNSQCMADLSVQVVDFELSKLQTVEFFN---TLFMNTKNLDFNAVIKNLEPVLS  6 EDPQA----KFISQFLEK-TSTEFQIHLWYLLYQAYSSAHRPYNSLLC
BC1G_12898_Bfuc_154295790             PNNAVIPEISTSAADREVSSLTTMEFF----QNLFQPDKSDPLAIIQTLEPVLD 26 TPAIL----RDMWKFLDS-GSTSLRLFLWQRLSEAYASIGYKTKVFSC
AN3123.2_Anid_67525331                ANSSIMSELSCEAIDQEISKLNSMDFF----TRIFNPDSEDPVGLIETIEPIVE 13 QSLQR----REMGSFLDR-GDATLRLFLWRRLQDAYRKIDYPPKVVSC
KLLA0B03850g_Klac_50303527            SNYNNISSLSLKTVQSQFSKIKMLRRFTIVD-IENEDEDNTHKDYIEHLYCIL-  5 EGTDF----QSLLSFVNT-SPFLMKLKLWKLLLNYFASRKDQLRFQNC
CAGL0M00286g_Cgla_50293891            PNADFISSFDLSFVDTQLKKLKMIKKLVNVDKHSESDKSEDNNSQFIMLENVL-  5 ENHAVGGDDDDMIDFFMS-APFMLKFKLWELIYMKYVKKNDIEKTTLI
AGOS_ABL081W_Agos_45185149            ANYEHTPRLNMETVQSQLSKIKMMRKFTMVNQMDVDRNDSSNEDQIEALSRVL-  6 DTPED----RSMSEFVNQ-APFLLKIKLWRIVLHHYLAVKDQSNFQIC
HIR3_Scer_151945205                   PNYRHIPALNLNTVQSQKRKIRIIQNITVEDISEDTNSDTHSENHLETLEKVL-  5 PSTNHSNIDEEMVSFIFN-SPFLLKIRLWGVLFSSYVKKSSIQDVQRI
Lgig1000008602_Lgig_Lgig1000008602    QNCHSDQLITPEKVTNRKDNVQRSHSLGQT-QKLFESESYD--KVIDTLLKTFE    ESRNK----VPQSSVAPSERHSQIKL-----LQKALFKDGDMQKCILW
LOC412746_Amel_110758961              PNCKYNSQISIKIVEKKLKSIHRGQKLGEI-QNLYDEKKYA--ELVD-------    -------------------RVQQLSM-----LLDSLWQLQQYEECYVW
Cabin1_Rnor_166157458                 PNLHNDSIVSLEEIDKNLKSLERCQSLEEI-QRLFEAGDYK--AVVQLLRPTLC  4 DRAKH----LEFMTSIPE-RPAQLLL-----LQDSLLRLEEHRQCFEC
Cabin1_Mmus_70995287                  PNLHNDSVVSLEEIDKNLKSLERCQSLEEI-QRLFEAGDYK--AVVHLLRPTLC  4 DRAKH----LEFMTSIPE-RPAQLLL-----LQDSLLRLKDHRQCFEC
GSTEN:00021655:G:001_Tnig_47217598    PNLRVDAAISVEEIEKRLKSLERCQSLEEI-QHLFESADFD--SVVRLLQPTLN  2 SRSKA----LELVSSAPE-RPAQLIL-----LQNALLKLDDFQQCLEC
PHYPADRAFT_139830_Ppat_162675462      PHCKLDKEISLDRVHGKLHELRVENILKHSAAKLLEKGQYE--DLIRLLKPVLL 21 ASVSS----HEL-------KGLEMLI-----TACENIKPPNLVLGLQC
Smoe1000001625_Smoe_Smoe1000001625    PHCKMDKRISLEMAEQKLYEFEVGDLLKNTVAKLDKEEKYL--ELIKVLEPALF 14 -AMQS----QEF-------KALGMLI-----SACQRVKPANFVIQLKC
AT4G32820_Atha_15233957               PHCRRVRELTINRIIHEINLLKIDFLLENNIPEMMEKEFYS--ECVNLLAPLLF 14 EGISS----VEL-------SALEVLI-----KACQKSKPIDVEVYMNC
OSJNBb0024N19.17_Osat_50872476        PHNKLVKLLTADRILREINLVKLESLLWHN-DENINKITHT--EFMELLPPLLL 17 EKVIS----LEL-------GALDVLI-----SACENAKPMNIQVYLDS
DDBDRAFT_0189500_Ddis_66827817        GHCKFDRVISPLILSDKLSKLQDQQEKNSA-GQLFNSKQYD--SVITLLEPLFP    DSLCV----LEFPPL----TNSNVNT-----TTNTSNNNNNNNIGTVF
consensus/90%                         .p......ls...hppphpplp..p.h........p....p....lp.L...h.    ...........h........s..hbh.......p.............h
```

```
Predicted Secondary Structure         HHHHHHHHHHH    -------------HHHHHHHHHHH---HHHHHHHHH-HHHH    HHHHHHHHHHHHHHHHHH    ----------hhhhh
SS1G_01112_Sscl_156061996             SLKCIEILVAE    FRSQDYLNC--DDETRRQKMLVSF---KALDHLLVP-ALTT 15 HVRSTCAALAQLSRILHA 13 SLPQNGAYSAFGSFS
MGG_00105_Mgri_145611812              LLRSIEMIVDD    FERESYLTA--SAETRTSILIHML---RNLDDLLIH-ALHL 14 HLGATAAALAKLSCILHV 13 QNSNVGSSFRK----
NCU04512_Ncra_164429378               FLRSIETIVHD    FERKDYVNT--PQAARKELFMTML---KALDDQIIQ-SLHL 14 HIKTTSAALAKVSCLLHV 13 QPPATG-SLAA----
CHGG_04295_Cglo_116195394             FLRSIETVISD    LESDDYLHS--TDEPRRNQFMSLL---KALDDLIIQ-SLHL 14 HLRSSMSALAKVSCMLHV 12 KPAPASTSDLR----
PODANSg7663_Pans_171691398            FLRSIETIITD    FEHADYLET--PPEPRRALFMTMI---KALDDLIIQ-SLHL 14 HLKSTLSALAKLSCLLHV 12 RPAPANNSTLR----
SNOG_08212_Pnod_169610231             YLRIIETFVTD    FKSTNFQEL--SDVERQVKLLTRF---RVIDDLITK-TLQI 14 HVQSSMASIVQLLHILSA 13 AMPRVEALPAQ-AFV
FG04030.1_Gzea_46116376               CLKSIELIVSD    LEGETYANT--VPESRRLLFMRTL---KSLDELLIT-ALGS 14 HIRSTSAALAKVNCLLHV 12 KPERTGNAASKAAFQ
hip3_Spom_19113368                    AFQSLKIILIR    LCSSSFSIQ--DAGRRQAELLGMLNFSSNLFRIIWQ-KLHE 15 CIRIILIYLRTFAIYVGI  8 IPKPSNPEFDS----
BC1G_12898_Bfuc_154295790             SLKCIEILVAE    FRSEGYLKC--DDETRRHKLLVLF---KALDQLLVP-ALTT 15 HIKSTCAALAQLSRILHA 13 SLPQTGVYSMFGSFN
AN3123.2_Anid_67525331                HLRSIETIIKE    LRSQAYVEE--PGEHRQATLVRWL---KSLDGMLSK-TVTT 15 HVKSSMSAVAFLIRLLHS 13 PGIDLRAALAK-SLE
KLLA0B03850g_Klac_50303527            FFSSIKFLLAR    LSSEEYQIQ--SQLQRQQTLLSTL---THIGEFSKR-FFQV 11 PSRDQLQVLVSIFRLLYP  9 KNASLKSFFKK--AV
CAGL0M00286g_Cgla_50293891            YGIMLKLLVDT    LECKTYRTL--GPKKRQERLLIMM---SCLKDITHN-CIEV 12 WYDDTTSQLVFIFFILFP  8 KLSGTQTFFQK--AT
AGOS_ABL081W_Agos_45185149            YFKVLHVLYKR    LCSKEYSDQ--SQLQRQQTLLSTL---SLMRSFTSL-FIEL 13 QASEYFKLLIDVFILLYP  9 KSATTVSFFKK--AA
HIR3_Scer_151945205                   YFHVLDFMKGA    LTSPVYKES--NPHGRHQMLLTVL---TAIGYLSSQ-LTAI 12 LEDYMFEKLLQTFFFFYT  9 NDVSNKSFFKR--AS
Lgig1000008602_Lgig_Lgig1000008602    GEVCYSEALHY    YRKA-------ATPTQREEWAKTM---VEVLNSLNK-AFEK 12 KIRFSENLIKIIETMMDV  1 PNVTEMPVGTVIPWL
LOC412746_Amel_110758961              AEACLNEAWQN    YLNA-------SDEIEQKKWTTSI---LMALEKLKACTIEV 12 LSRLVQNLVHIVCNQLDV  1 ETTVEMPLETVLPWI
Cabin1_Rnor_166157458                 SDVALNEAVQQ    MLNS-------SDSAAKEEWAATV---TQLLLGMEQ-ALSS 15 LVRLTNNLIQVIDCSMAV  1 EEPKEPYVSSVLPWI
Cabin1_Mmus_70995287                  SDVALNEAVQQ    MLNS-------SESAAKEEWVATV---TQLLLGIEQ-ALSS 15 LVRLTNNLIQVIDCSMAV  1 EEPKEPYVSSVLPWI
GSTEN:00021655:G:001_Tnig_47217598    SELALNEALQQ    LTSAPH-----GTRPTKEEWVCTI---RKLLDGIEV-CFAK 13 LSRLANNLIQLIDLSMST  1 DDPKEPYFFSVLPWI
PHYPADRAFT_139830_Ppat_162675462      RELRLRLLSRA  3 --AEVKIGD--DNDSSFTTLDTEL----------------    ------LLLTLMCHFVQT 16 TQAYSS---------
Smoe1000001625_Smoe_Smoe1000001625    HLKRLEIFCHC 16 SAASGEMDD--DRAKELRTWSKKI---AEEVRLLSR-CAAA 13 LGRLQNALLAVLIHFGSG 12 ENLETS---------
AT4G32820_Atha_15233957               HRRKLQVLLDS  3 GESVVTPKT--SSKNSSESWDHLV---AEEVKAILL-CISQ  4 LDQSGNSRYSDSQNADGI    EEEKKS---------
OSJNBb0024N19.17_Osat_50872476        HRRKMQVLTVA  3 VGSVTTNEG--KKSSDIEFMETMN---RNRLESVVE-AVKD 30 VGDFQSLLLTIMCAAVKM 15 DQLESS---------
DDBDRAFT_0189500_Ddis_66827817        E---FK-----    NKDHIDSDT--KSRLNRIDLLLQS---FLMTSNIRG-VLKM 16 TVNFYITLIDRVFKHDNM    KPKKLDETESTLLSI
consensus/90%                         ....hp.h...    ...........s...pb..h...h.............h..    ..p.....l..h...h..    ...............
```

```
Predicted Secondary Structure         HHHHHHHHHHHHHHHHHHHHHH-------------------HHHHHHHHHHHHH----------------HHHHHHHHHHHHHHH--
SS1G_01112_Sscl_156061996             NFANKLKEMQVRTWALQYAIVKEATAQN--RDLFPTPD----NDLADFLALVHYSLGLRKCCKA------SNKIFLKMMKQEMSRLKHI
MGG_00105_Mgri_145611812              SLLDLLHNFQARTWCLQYCVLKVGIGSN--KELFTKPS----SHLSQLLAAIHQVLGLRKMCKA------SNRIFPKMMRVELLRQTEA
NCU04512_Ncra_164429378               SFLNRLREMQVRTWSLQYTVLKMGINLH--PDIFLTPE----NDLADYLAAVHQVLGLRKYCKS------SNKIFLKMMRVELLKLKNI
CHGG_04295_Cglo_116195394             SFLNKLQEMQVRTWSLQYTMLKVGLNQH--PELFPKPE----NDLADYLAAVHQVLGLRKCCKS------SNKIFLKMMRVELLKQKNI
PODANSg7663_Pans_171691398            SFLNKLQEMQVRTWSLQYTMLKVGIQQH--PDAFPNYE----NDLAEYLAAVHQVLGLRKSCKS------SNKIFLKMMRVELLKQRNI
SNOG_08212_Pnod_169610231             NMSTKLDDMHIRLWMVQYHLLKEALSQN--AENFPTPS----EDLFEFLRHVHHATGVRRFCHQ------SSRHFLRLAKDEMLRLDDV  3
FG04030.1_Gzea_46116376               ALKNKLKEMQIRAWCLQYTMFKAGIN-K--DNCIIGPE----NDLADYLAAIHQVIGIRKFCKA------SNKIFLKVMRVELLKLKNI
hip3_Spom_19113368                    -YAQTVKDSLMSGWCIFYTLFAHLLH----YDFIKADAQ---KLLPQILTAIHSQYSFRGYCSS------SNQSFLELSQTECQRLDA-
BC1G_12898_Bfuc_154295790             NFANKIKEMQVRTWALQYAMVKEATTQN--RDLFPTPD----NDLADFLALVHYSLGLRKCCKV------SNKIFLKMMKQEMSRFKHI
AN3123.2_Anid_67525331                NFKEKMRDMLVRCWVLLYTLLKEAIAQN--QEMFDEPL----EDRVYYLRAVHNSLGLRQMCKR------SRNQFLKLVKSELLALDVE
KLLA0B03850g_Klac_50303527            KSSITIQDIFLTVASLMILCHKFITLEK--ENAKEEFT----GSIIDLISALHMLAGKFEFCDR------FDGNFLKVQEKLLCEFSN-
CAGL0M00286g_Cgla_50293891            KSSKTIKSIFVDILTLLLFHLNIGLKSK--ISSEPELE----NSIIDLVSTTHISISFYNFCDY------ASGNFLQIAENLICQFTN-
AGOS_ABL081W_Agos_45185149            KSSVILKDMFVDISCLLVLYFTGACTQK--EVVNGV------EVATEFIHALHLLIGGFTFCDA------ANGNFLNLAVHFFCSVEG-
HIR3_Scer_151945205                   KSSGKMKDIMIDLATLILYYYDLQAKLR--TPAEQG------IETTELIWSLHTLFGHFHFCDA------SNGKFLDLAEKLLCQFIN-
Lgig1000008602_Lgig_Lgig1000008602    LLYRLIHHEEVKLHSMI-ASQSKDQADVLE-----GAI----STSLMLLSVAHEYLGRHSWCTK------NNGAVLLLFLKVLRKERNK  8
LOC412746_Amel_110758961              LLHYILQYKEDKERAKA-ESSYKNKHNSNY--SESDDEDEGVPASIMILFIAHDFMGRHSWCCY------NEAKLLFFTLNLVIPKLET  7
Cabin1_Rnor_166157458                 ILHRIIWQEEDTFRSLCHQQQLQNPTEEGI--SEMPML----PSSLMLLNTAHEYLGRRSWCCN------SDGALLRFYVHVLQKELAA 10
Cabin1_Mmus_70995287                  ILHRIIWQEEDTFRSLCHQQQLQNPTEEGI--PEMPML----PSSLMLLNTAHEYLGRRSWCCN------SDGALLRFYVQVLQKELAA 10
GSTEN:00021655:G:001_Tnig_47217598    VLYRIIKHEEAAFDCLL-QQHMSVRDDEDE--SDAPIL----PSSLALLNTAHEYLGRRSLCCS------SDGALLKFYVQVLEKELAI 10
PHYPADRAFT_139830_Ppat_162675462      -----CFVDAAIAFCRLQHLQPSVPL----------------EHQVTLLVTLHDVLAEYGLCCAGKDSEGGEGGFLKMSIKHLGALESR 78
Smoe1000001625_Smoe_Smoe1000001625    -----FFVDAAIAFCRVQQLHPALPV----------------EQQVGLLVTVHDILADRGVCAAGASCDEGEGSFLKMAIKHLLSLEVM 49
AT4G32820_Atha_15233957               -----CFLDAAIGFCKLQHLDATIST----------------KYQVELIIRLHDLLAEYGLCCAGKNCAGEEGAFLRFAIKHLLAVDMK 85
OSJNBb0024N19.17_Osat_50872476        -----CLVDAAIAFCKLQHLDPMISI----------------KIQA-----LHDLLAEYGLCCAGRDGEGEEGTFLKFTIKHLMALDVK 49
DDBDRAFT_0189500_Ddis_66827817        FIRILYRYYKKNISYTSLLLLLQIYQIDESKKLKT-------YQTIQLYSMLHSELGKKKQC-L------NEPQFLINYMELLFNVIID 82
consensus/90%                         .....h........sb..............................hl..hH..hu.b.hCp.......sp..hLbh....h..b...
```

```
Predicted Secondary Structure         -HHHHHHHHHHHHHHHHH-----hhhhh----HHHHHHHHHHHHHHHHHHHHHH--HHHHHHHH    HHHHHHHHHHHHHHHHHHHH
SS1G_01112_Sscl_156061996             DKWEDYLGQVLFDLYGIKL-----GVGTFLLEDHGCPIESLDRRTVMNIAEQVIS--LANNMPMK  1 LMKHELKPTIEKMQTAVGQPK
MGG_00105_Mgri_145611812              EDWERDLGQVLYDLHGLKL-----GV---EVEDHGCPPEKLEKRNAIALVEKITI--LADRMSMK  1 LLKSELKTTIEHMSQAIGQTK
NCU04512_Ncra_164429378               ENWEDYLGQVLYDLHGLKL-----GVGIWDVQEHDCPPEKLEKRQALAMVEKISL--LAHRMTMK  1 LLKSDLKNTVESMQQAIGSAK
CHGG_04295_Cglo_116195394             DNWEDYLGQVLYDLHGLKL-----GVGIWEVQDHGCEPENLEKRQALQLVGKVTM--LANRMSMK  1 LLKSDLKTTIERMQQAIGSAK
PODANSg7663_Pans_171691398            ENWEDYLGQVLYDLHGLKL-----GVGIWEVQEHGCEPENLEKRQALQLVEKITV--LANRMSMK  1 LLKSDLKTTIERMQQAIGTTK
SNOG_08212_Pnod_169610231             NSRDTELCQVLHDLYGLKL-----FIEPLECQDFSSTPDILDKKTAFGILPFVMS--QVKKVDLK  1 LPKTELRGTIDKVHAALGRPK
FG04030.1_Gzea_46116376               ENWEDYLGQVLYDLHGLKL-----GAGVWEVQDHACPPEKLEKRQTMQLVERVCI--LANRMPMK  1 LLKSDLKTTIDHMQQTIGQTK
hip3_Spom_19113368                    WENENEILQCVCCRFNLII-----GSEYYVPQSHQSDSVNLTSNDAIKIIPFILG--FAVKRSHG  2 MPRSDQKNALEIICKVIRFPG
BC1G_12898_Bfuc_154295790             DRWEDYLGQVLFDLYGIRL-----GVGTFLLDDHGCPRENLDRRTVMNIAEQVIS--LANNMPMK  1 LMKHELKPTIEKMQTAVGQPK
AN3123.2_Anid_67525331                QDIEADICQILFDIHGVKL-----SLSDHLLSDHGCVPEKLDRSTAIMMIDFVMK--QAKKINIK  1 LSKSELKNTIEKMQQSIGTTK
KLLA0B03850g_Klac_50303527            DSAFTHLKRELWCRYHISI-----GLD-NPEDQHQTHAQPMVKANAIMLSNYFIKFQYQDRNILV  2 NNRSNFKQFFETVMELIGDID
CAGL0M00286g_Cgla_50293891            EEAFTQLKQIMWCKYHYQL-----SGDSSIVSQHKTKAMEMSKMVSLPLGIYLVKLQYQDKHPYI  1 SAKIASKQILDNIVETFGDLS
AGOS_ABL081W_Agos_45185149            SSSFIPLIQILLCKYRLSL-----GGDANSFEDHGAKPQVLEMHNAVRLAKYLIRFEYQNKNPYL  2 TNRSNLKQVIENVIDVIGKIP
HIR3_Scer_151945205                   NDSFLQLKQILWCRYHYAI-----ASDNFSPDLHDTKAVEMEKIHSLPLGTYLIKLQYQNKNPYL  1 SSKTTLKQIMDNIIEKIGDPS
Lgig1000008602_Lgig_Lgig1000008602    EELDVAFEQCIYCLYGHPNK----KSKSRRLHDHNSPPISLTWTRAQAVFEYYKPSTVPEFDSYK  2 TVSTELEGLLRRVFVLVPEE-
LOC412746_Amel_110758961              SKVSKYLEQIFYCLYGHYNKVN--KSRPKHIEEHGVPQMKVTWEGAQLLFDFYKPKQIPEFCSPK  2 TISADTEDLFKRIIKLIPPE-
Cabin1_Rnor_166157458                 EELETALEQCFYCLYSFPSK----KSKARYLEEHSAQQVDLTWEDALFMFEYFKPKTLPEFDSYK  2 TVSADLANLLKRIATIVPRT-
Cabin1_Mmus_70995287                  EELETALEQCFYCLYSFPSK----KSKARYLEEHSAQQVDLIWEDALFMFEYFKPKTLPEFDSYK  2 TVSADLANLLKRIATIVPRT-
GSTEN:00021655:G:001_Tnig_47217598    EELEMAVEQCFFCLYAYPSK----KSKARYLDDHSSPQVELHWNDALFLFRYFKPKSLPEFDSYK  2 TVSAELANLLRRFASIAPLS-
PHYPADRAFT_139830_Ppat_162675462      LGIDYALDQSFFCLYGLNLRGE--ATGPDGLMEHQNTNRGDYKTMEQCAGVLQYILPYAKACSR-    AGLMKLKKVLRAIHKQFSCP-
Smoe1000001625_Smoe_Smoe1000001625    -GLDKALDQCFFCLYGLNLKSDLESGASSDLAFHKTPSGGDFQSKEQCAAVLQYVLPYAKACSK-    TGLGKLRKVLQAVRKQFPQP-
AT4G32820_Atha_15233957               LLINNALDQCFFCLYGLNLRVD--GSYEDELAVHKNTSRGDYQTKEQCVDVFQYILPYAKASSR-    TGLVKLRRVLRAIKKHFSQP-
OSJNBb0024N19.17_Osat_50872476        SSIDSALDQAFFCLYGLKINPD--SCSEDDLAVHKNTSRGDYQTKEQCADVFQYVLPYAKALSFG 20 TGLVKLRRVLRAIRKHFPQP-
DDBDRAFT_0189500_Ddis_66827817        IKVLTDICQCFNCIHSIKL--------SNSLEDHNCSTQYGSNSNSSIRQLKGPGVYMQLFILIK 30 ANKKEIKELIEKIHSRFPD--
consensus/90%                         ......l.Qhhhsba.h................H.s......b...b.h..b.....h.p..s..    ....pb...hc.h.p.hs...
```

```
Predicted Secondary Structure         --HHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHH    H-HH----HHH    HHHHHHHH---hhh--hhhhhhh--
SS1G_01112_Sscl_156061996             --ETPTMLHNSRNISEYLKTPIRPLDLYLAMRGESEIDSISVRTPESPL    ADKGWYYLLGM    LALTKYRS---TKRVSAGSQTDD--
MGG_00105_Mgri_145611812              --STPQMIHNLRNFTEYLKKPIHPLRLYQAWKGTVSIDAVTVTTSEAAL    AEHGWFFLLGM    IAFSKFKQVDLNRRQTPGA-TDD--
NCU04512_Ncra_164429378               --STAQMMHNHRNFNEYLKTPIHPLHLYKALTGNVDLDAVTINTPDSAP    AKHGWFFLLGM    IALTKFKGVDLNRRQTPGA-LDD--
CHGG_04295_Cglo_116195394             --SSPQMAHNLRNYSEYLKAPIHPLHLYQALTGSVDLDAVTINTPESAP    AKQGWFFLLGM    IALTKFKGVDLNRRQTPGA-TDD--
PODANSg7663_Pans_171691398            --SNPQMTHNLRNYTEYLKTPIHPLHLFQALDGSIDLDAVTINTVDSAP    AKHGWYFLLGM    IGLTKFKGVELARRQTPGA-TDD--
SNOG_08212_Pnod_169610231             --QHKDLSFNHKLITAWFKSPIHPASLLNCLKGVGSLSAKHVPAEEAVA    ASMGWYYLMGD    ISLSKFRS---QKRMTQGP-TED--
FG04030.1_Gzea_46116376               --STPQMYHNLRNFTEYLKKPIHPLRLYRALIGDVSVDAVSVNTPETIL    ATHGWFFLLGM    MALTKFKGVELNRRQTPGA-TDD--
hip3_Spom_19113368                    E-NNADVYFNKCAIKEFLERDISPQLTKMLLKSNDILGLREIGSKVVDD    RVRGLYYAQSQ    VLFGYYRS---RLKGSRCI--ND--
BC1G_12898_Bfuc_154295790             --ETPTMLHNSRNITEYLKTPIRPLDFYLAMRGEVEIDFISVHTPESPL    ADKGWYYLLGM    LALTKYRS---TKRTSAGSQTDD--
AN3123.2_Anid_67525331                --AVPPVSYNRRILNAYLKTPINPSELVRAIQGVTDLPFIPVPSQTAVI    ANSGWYFLLGY    AALTKFRS---QKRLNPVP-TTD--
KLLA0B03850g_Klac_50303527            YEANHVLSRNEYFFNQYFTAPITLKTIRDAYAGEFEIEFTSPNDDLQSG    VDGGLFYVSAV  2 LNQYKARK----KSMQARP--SE--
CAGL0M00286g_Cgla_50293891            SQDNYIIMRNLHAFNGYLRSNITSKLFQSCFEGKDKVDLKNPNDEFQKA    MDAGVFYVSGI  2 LNLYKIRK----RSVQARP--SE--
AGOS_ABL081W_Agos_45185149            YAENHILARNVYYFEQYLEAPVTAKIIQEALQGCISIELTKPRDNLQEI    VDLGLYYISGV  2 LNLYKVRK----KTLQARP--SE--
HIR3_Scer_151945205                   TLDNHIISRNSFLLNEYLSRPITADLLKHTFSGATSLYLTSPNDELQQG    MTAGLFYVSSL  2 LGLYKMRK----KSMQARP--SE--
Lgig1000008602_Lgig_Lgig1000008602    --LGGAGKVDT--VVKYIDGDVN------------TPPSF-ETDNLVL-    --KELYYLLAD    ---YYFKN------------KET--
LOC412746_Amel_110758961              --SDPNQVIDE--MMAYIMGERE------------TMPTV-K--KLLPH    SMSTIYYLLAD    ---FYFKN------------GKW--
Cabin1_Rnor_166157458                 --EKPALSMDK--VSAYIEGTSA------------EVPCLPDGADPAPP    VLNELYYLLAD    ---YHFKN------------KEQ--
Cabin1_Mmus_70995287                  --EKPALSMDK--VSAYIEGTSA------------EAPCLPEGADPTPP    VLNELYYLLAD    ---YHFKN------------KEQ--
GSTEN:00021655:G:001_Tnig_47217598    --DTPMLTMED--VAAYIEGMTE------------KVPCLPEGSAPPPP    IIDEIYYLLAD    ---YHFKN------------KEQ--
PHYPADRAFT_139830_Ppat_162675462      --PPDVLSEKS--VDMYLEDPDF------------DEDKLRDMVKSGTS    AELVVNFALHS 11 WPGFVLTK------------EGEEF
Smoe1000001625_Smoe_Smoe1000001625    --PSNILVGNA--VDAFMDDLSF------------DEDKFSEMVINRTN    VDELVSFAFPK 53 SPGFVFTK------------EGEEF
AT4G32820_Atha_15233957               --PDDLLIGNV--IDKFLDDPEL------------CEDKLSYEAGSEGF    LETITKCLIPS 14 WPGFVLTK------------EGEEF
OSJNBb0024N19.17_Osat_50872476        --PYDLLVNNP--LDNFLDGPDS------------CEKILSEIYETNGS    KEAVLNVLFPG 40 YTGFVLKK------------EGGEF
DDBDRAFT_0189500_Ddis_66827817        --PPSRVTKYKSVIDQYMNNEIKELTDELTSGSASTSIPSSPQRLALNL 19 F-TDIYYMYAL  2 -------------------EKDEDC
consensus/90%                         ......h..p...h..ahp................p.............    .-..h.ah.s.    .....hp...............p..

Predicted Secondary Structure         HHHHHHHHHHHHH---HHHHHHHHHHHHHHHH---HHHHHHHH-------h    HHHHHHHHHHHHHHHHHHHHHH
SS1G_01112_Sscl_156061996             LRFAAHFMKLQLQYTVDHWEAWYRLAQCYDLE---LEEEVLWSADKMNNDR  3 VRLQKCSIHCYIMALSTAMRCAD  2
MGG_00105_Mgri_145611812              LRIGQASLRSQIQFTPDRWDAWLRLAECYDYE---LDEAVLWTADKINKDR  3 LKFQRSSIHCYSLALSHSYSMSR  3
NCU04512_Ncra_164429378               LRIAVTFLRLQLQYTPDRWDAWFRLAECFDYE---LDDMVLWTADKMNKDR  3 VKFQRQSIHCYTLALSHSYASTS  8
CHGG_04295_Cglo_116195394             LRIGATFLRLQLQYTPDRWDAWFRLAECFDYE---LDESVLWTADKMNKER  3 VKFQRNAIHSYIMALSHSYAWAS  8
PODANSg7663_Pans_171691398            LRIGATYLRLQLQFTPDRWDAWFRLAECFDYE---LDESVLWSADKMNKDR  3 VKFQRSSIHCYTLALSHSYAWSA  8
SNOG_08212_Pnod_169610231             LSYAQAFFIQDLEYSTDRWETWYRLAQVNDLQ---LEEVVSWTADKINSGS  3 INHQRGAINCYTMAVACAVRE--  2
FG04030.1_Gzea_46116376               LRIGATFLRLQLQFTADRWDAWFRLAECFDYE---LDEAVLWTADKMNKER  3 VKFQRNAIHCYTLALSHSRNLD-  2
hip3_Spom_19113368                    LLVIIKYFLLDLYLNPRRQDSWYTCSSVFSSL---ADEELGWSAEQICLAD  3 NEYRRKAILCNLMALSLPFTQ--
BC1G_12898_Bfuc_154295790             LNFAAHFMKLQLQYTADHWEAWYRLAQCYDLE---LEEEVLWSADKMNGDR  3 VRLQKSSIHCYVMALSTAMRCAD  2
AN3123.2_Anid_67525331                LDEAISWFRQDLEHNTSRWESWYRLAQVWDSK---VEEDITWSADKINNNR  3 VTWQRNAIHCYAMAVATAAKTAE  2
KLLA0B03850g_Klac_50303527            LDAIISTLTTDILYNTKRFESWYLLGKCYSFV---VEDDLTWTSDKLASRD  5 AFAEKKAILCYLMALSLYLSTFK  8
CAGL0M00286g_Cgla_50293891            LDAIITMFKNDILYNTKRFESWYLLGKCYSYI---VEDDLLWTADKISSSE  5 ATTQKRAILCYAMSLTLFFENET  3
AGOS_ABL081W_Agos_45185149            LDSIIETLKADILYKTNRFETWFLLGKCYSYV---VEDDLIWTSDRLVVAS  5 ASIQRKAILCYLMALNLCQLSVD  9
HIR3_Scer_151945205                   LDSIIRMLKNDIIYNTNRFESWILLGKCYSYI---VEDDLIWTSDKITVPE  5 ALTQRKAILCYLMAISIYYSKLD  3
Lgig1000008602_Lgig_Lgig1000008602    -SKAIRFYQCDLCVNPKRLDSWAGTALARMSQ---LEQKLNSMELKLDMPI    HKKAIAALRCFKRAVEINDTTRT  2
LOC412746_Amel_110758961              -SAAIRYYLLDLCFHPTRLSSWTGLAMSSATV---IDTWLNKYTPINEDKL    LAKAKLTQSSYRHAVELDSCNHI  2
Cabin1_Rnor_166157458                 -SKAIKFYMHDICICPNRFDSWAGMALARASR---IQDKLNSNELKSDGPI    WRHATPVLNCFRRALEIDSSNLS  2
Cabin1_Mmus_70995287                  -SKAIKFYMHDICICPNRFDSWAGMALARASR---IQDKLNSNELKSDGPI    WKHATPVLNCFRRALEIDSSNLS  2
GSTEN:00021655:G:001_Tnig_47217598    -SKAIKFYVNDICVCPKRFDSWAGMALARASR---IQEKLNSNELKSDLPI    WKHSQAVLTCFKRALEIDGSNLS  2
PHYPADRAFT_139830_Ppat_162675462      VQTNANYFKYDLCYNPLRFESWLKLAKILDEVKHFFDLLLNDGSKNINVVE    WRKTGILASRVELSRRQSRRSLL  4
Smoe1000001625_Smoe_Smoe1000001625    FEHVSSLFKYDLLYNMLRFESWHKLAGIYDEE---VDLLLNDGSKNVNVLE    WRKNARLTSRVETSRRRSRRCLL  4
AT4G32820_Atha_15233957               EQQNTNLFKYDLLYNPLRFESWEKLGNIYDEA---GSFIRSCCYCYFENVP  4 WRKNSALSQRVETSRRRSRRCLL  4
OSJNBb0024N19.17_Osat_50872476        VQQSANLFKYDLLYNPLRFESWQKLANLYDEE---VDLLLNDGSKHISILD    WRTNTTLIQRVEMGRRHSRRCLL  4
DDBDRAFT_0189500_Ddis_66827817        IDQFRSLLIKDLYCNPNRIDSWQALSKSYSDQ---LKKVTDSDGILIRYLY 46 GGNNDENYQEYELIELKISSLYK 19
consensus/90%                         .p...p.hb.pl.hss.Rh-uW..hu.hbs.....hp..l..s....p...    .........ph..u.........

Predicted Secondary Structure         ---HHHHHHHHHHHHHHHHH-HHHH-------------    HHHHHHHHHHHHHHHH---    ----HHHHHHHHHH----    ---------
SS1G_01112_Sscl_156061996             FGTAEKLSEMYYDFGMRLYA-SSREPFSMEVFWVDEEK 18 KYKLWRYASELFKRSLADR    PDFWMNHFMLGKCQWKMV  9 KPPTNVP--
MGG_00105_Mgri_145611812              DNDSEALYDLYYNFGMRMYA-SSREPFGMEPFKHSEQE 23 DYQVWKYAASLFRKAMARR    PKDWKSAYMYSKCLWKMY  9 STKRTKP--
NCU04512_Ncra_164429378               GGDTEALYDLYHEFGMRMYA-SSREPFAMEPFMHSDQH 21 DYQVWKFAASLFRKAMAGK    PKDWKNPYMIAKCLWKMY  9 KNRHGRP--
CHGG_04295_Cglo_116195394             EDDEEALYDMFHEFGMRMYS-SSRSPFAMEPFQHAEQN 21 DYQVWKFAAHLFRKAMVGK    PKDWKNPYMVAKCLWKMY  9 KNRLRRP--
PODANSg7663_Pans_171691398            EDEKEALNDMYREFGMRMYA-SSREPFAMEPFKHSDQE 21 DYQVWKFAAHLFRKAMEGK    PKEWKNPYMVAKCLWKMY  9 KTRRHRP--
SNOG_08212_Pnod_169610231             AQSQSRVGQMYTEFANRMYA-SSREPFSMEAFQFRENE 22 LYTAWKIANTLYKRAIQEN    PNKWWNYYMLGKCSWKMW 17 STNRVQP--
FG04030.1_Gzea_46116376               THDGDPLHDLYHKFAMRMYA-SSREPFAMEPFQHSDQE 21 KYKVWKFAAKLFRMAMDRK    PDNWKNPYMLSKCYWKMY  9 KDITSKI--
hip3_Spom_19113368                    --DKLFKANVYFDFAMNLYA-SARPPLEMAAFLPSETR 17 VSKVIALAADYFGMSAEL-    SNDWRALYMLGKACRKCG 11 AAALAPT--
BC1G_12898_Bfuc_154295790             FGTAEKLSDMYHDFGMRLYA-SSREPFSMEVFWVGEEK 19 KYKLWKYAAELFKRSLADR    PDFWMNHFMLGKCQWKMI  9 ---PIVP--
AN3123.2_Anid_67525331                PETGALLADLYTDFGIRLYS-SSREPLSMAAFSVADFT 21 VYNVWRLAAFLLRQALVDK    PKNWMTHYMLSKCLWKMF  6 RGSSKRI--
KLLA0B03850g_Klac_50303527            NDNKIVFRNLLESLGKEMLQ-AYVKPMSSMSYTWKLKP 22 ETNILRCILMILAKADTLY  4 ERNWMNPFYISKVHFKTD    RKLFKES--
CAGL0M00286g_Cgla_50293891            VDEQIVIVECLESLANELMI-AVYKPMEALCFDTNLSS 26 DSSIEDTMIKCLTLANRYQ 10 RLNWFNFYSIAIVYSTE-    KLDRKGD--
AGOS_ABL081W_Agos_45185149            MENAFIKREIYEVLALELLN-AYFKPMEGLPFQRDIAP 22 SANIYRATLLVFTNADEM-  8 CKNWLNPHYIASLRFKHE    REDFITD--
HIR3_Scer_151945205                   -DDKKIILEALDDLGSMLIS-GYYNPMNKLCFSWKSSA 24 DFNIEQSIFLCFNRACSLS 10 VLNWSSFYNLAKFFFKTD    GGNNCKL--
Lgig1000008602_Lgig_Lgig1000008602    -IEYGSLAYQLHSHSSRQLKLKDWFSLDQELLQIVNDM    KPEMLQIARNCYWKASEC-  5 EEEWLHHYMMGKIKEKKR    AHPKDYLEH
LOC412746_Amel_110758961              -TEYGSFAYTVHSFCSRLLK-QETHTLSMGRFEILETR    KEEMLEIAYKCFQSCARLF 10 DERWLYQYMLGKIAEKKN    QDPPIYLEH
Cabin1_Rnor_166157458                 -IEYGTMSYALHSFASRQLK-QWRAELPPEVVQQMEDR    RDSMLETARHCFTSAAHC-  5 EEEWLIHYMLGKVAEKQQ    QPPTVYLLH
Cabin1_Mmus_70995287                  -IEYGTMSYALHSFASRQLK-QWRSELPPELVQQMEDR    RDSMLETARHCFTSAARC-  5 EEEWLIHYMLGKVAEKQQ    QPPTVYLLH
GSTEN:00021655:G:001_Tnig_47217598    -IEYGTISYALHSFASRQLK-QWRNELPAEVVKQMVER    RDAMLDTALQCFQGASRC-  5 EEEWLIHYMLGKIAEKRK    KAPVEYLQL
PHYPADRAFT_139830_Ppat_162675462      LAVTPDQKSEVNELLALVYYDSIQNVVPSYDQRRHVPV    HD--ASWIAACQNSFAHF-  6 KPHWTHVFYCGKLVEKLE    QPHSVSLAY
Smoe1000001625_Smoe_Smoe1000001625    LAPQAKHEFQIHELLALVYYETLQNTVPSYDQRSHVQK    KD--ALWTSRCLNAIKHC-  6 RPDWTHLLYLGKLSEKLE    KPSEESFSL
AT4G32820_Atha_15233957               LANSPDQQSEIHELLALVYYDSLQSVVPFYDQRSVLPS    KD--ATWTRFCENSMKHF-  6 RQDWSHAFYMGKLSEKLG    HSYEISLSY
OSJNBb0024N19.17_Osat_50872476        LAKTASDKAQMHEMLALVYYDSLQNVVPFYDQRATLPV    KD--STWETFCRNSMKHF-  6 KAEWLYAFYLGKLCEKLG    HSPAEAFSY
DDBDRAFT_0189500_Ddis_66827817        FRDLALLIFEKHYFIRKSSS-RRYQEKLQSLFRRSLNY    IERAIQLDKQCNLNSPE--    KENWFYPVFKSKLQRKSG    EQPLVYL--
consensus/90%                         ..p.......h.ph..b.h....b..h...........    .....p....hb.......    ..pW...ahhuKh.bK..    .........
```

```
Predicted Secondary Structure         HHHHHHHHHHHH----    -------HHHHHHHHHH----------HHHHHHH---    --HHHHH    HHHHHHHHHHHHHH-
SS1G_01112_Sscl_156061996             TMDMVLSAFVRAIETC  7 QEPILEPHYKLVSLVHKLVARGEME--FKDAAKILQD 13 ESDEDWE    IFILACFRQLRNAD-
MGG_00105_Mgri_145611812              GKELLISALEKAVEVV  9 QDPVIEPHYKILVVLYKLVSRGDMT--AQEAADVLQR 18 DDKEEWE    EYVIKYLCHLRDKD-
NCU04512_Ncra_164429378               TVQAVIKALEKTVEVV  9 QDPILEPHYKILSVINKLVTRGDLP--HQEAADILQR 13 EKDEDWH    EYMLRYMRHLRDKD-
CHGG_04295_Cglo_116195394             TVQAVIDALEKTIEVV  9 QDPILEPHYKIVSVVHKLVMRGDME--IQAAANVLQH 13 VNAEDWE    AYVVRCLKYLRDKD-
PODANSg7663_Pans_171691398            TVKMILRALEKTVEVL  9 QDPILEPHYKIVSVVDKLVRRGDLP--RQEAADLLQR 13 DTPEEWE    AYIIRCLRYLRDKD-
SNOG_08212_Pnod_169610231             TWVSVTDAIINAIDTL  7 SEPILEPHYKLVSIAHKLFQKRAID--HVKGVELLSH 10 ENAEGWE    RYILAVLKSLRSAD-
FG04030.1_Gzea_46116376               GLMTLIDTLKKAVEVA  8 TDPIFEPHYKIVTILHKLVIRGDVP--AKDAAEILSE 16 SALEDWE    EYIIRNLSKLKEKD-
hip3_Spom_19113368                    KS--------------  5 QALLIEPHYALLSNLSKAAIEGSVE--IVQILSYLRR 16 EDVNIYK  4 LFILKALAEMRILD-
BC1G_12898_Bfuc_154295790             TMETVLNTFIRAIETC  7 QEPILEPHYKLASLVHKLVARGEME--FQDAAKLLQD 13 ENEEDWE    IFILACFRQLRNLD-
AN3123.2_Anid_67525331                SLDSVLDSLLDTIDAL  7 SEPIFEPHYKLVSIIHKLVIKEVLT--PAEASKTLVA 10 EDRNSWK    KYILDVLKNLKNAD-
KLLA0B03850g_Klac_50303527            GICLLQDSCRLALLQS  4 SDNILEAHYALVSYCYKCFKDRTFS--LAEAISHLRE  6 LPEEEWQ  8 NLIIRLLKYILSKD-
CAGL0M00286g_Cgla_50293891            MIDNMSKACELANKYT  2 KDSVVEPHVALIHMLYKLYIEGKVT--MEDTLSTLKR  6 FDDDFWI 15 SKLIELLKYVISKD-
AGOS_ABL081W_Agos_45185149            GFDIIIRACNLAMKAS  2 SGSIIEPHYSLVIKCYKSVKHGWLR--PEDALKHLLN  6 QEANFYS 12 VKIITLLRRLLALV-
HIR3_Scer_151945205                   VAKYITQGCQIAYESS  2 KDPIIEPHYLLVNACYKWVKRGVIG--VNEALTLLSK  6 EQEEFWV 15 DKIIRLLRHLLSVD-
Lgig1000008602_Lgig_Lgig1000008602    YRQAAIYLHEDEAVYP 12 AIEALEVFYRLHATILKHLIRDDGT--HDYKLFEEYI293 DSERKIQ 19 ELISSCMKTLNMCLS
LOC412746_Amel_110758961              YAKASELLYHNNAQYP 11 SIEALEVHYRIHASILKYLEQHEGK--PLKKSLGQLL222 ESEKKKE 17 TLIAYCLAGLEQCIL
Cabin1_Rnor_166157458                 YRQAGHYLHEEAARYP 11 AMEALEVYFRLHASILKLLGKPDSG--VSAEVLVSFM208 GAEEQRQ    FLTEQCIASFCLCLS
Cabin1_Mmus_70995287                  YRQAGHYLHEEAARYP 11 AMEALEVYFRLHASILKLLGKPDSG--VGAEVLVSFM208 GAEEQRQ    FLTEQCISSFCLCLS
GSTEN:00021655:G:001_Tnig_47217598    YKKAAHYLHEEAARYP 11 AMEALELHFRLGATILKLLEANESD--LDHELLFSLL245 DRGEQRR    TLVEMCVRALFLCLS
PHYPADRAFT_139830_Ppat_162675462      YKKGIDM---------    NPTAVDPLYRLHASRFKLLCCSRHS--DKNILQVVVG 42 LLNQVWN    TLFDDCAAALQSCFD
Smoe1000001625_Smoe_Smoe1000001625    YKRAIEL---------    NPSAVDSLYRLHRSRLKLLCKGRYH--DMQIIETVAK 31 -IATAWR    ALFKDCISAMQACTD
AT4G32820_Atha_15233957               YKQAMTL---------    NPSAVDPVYRMHASRLKLLNACGKQ--NLEALKVLAS 46 QMEGVWH    MLYNDSLSALGICVE
OSJNBb0024N19.17_Osat_50872476        YNKAVVL---------    NPTAVDPVYRMHASRMKLLYTQGKQ--NLDAIQVVAD 46 LLDKVWH    ILYDDCLCALGTCVE
DDBDRAFT_0189500_Ddis_66827817        ETLNKTIQKLPQISKT    KQNSILPLYRLHADRLKLLLPFETTQIHNDLLTILEK 48 TAMTLVV 14 LLIENSLSALTQIKS
consensus/90%                         ................    ...hl-s.a.l.....Khh............h.....    ...p.b.    .bh...h..h.....

Predicted Secondary Structure         -------HHHHHHHHHHH-    ---HHHHHHHHHHH    HHH------EEE    EEEEE---    -------HHHHHHHHHHHHHHHHHH
SS1G_01112_Sscl_156061996             -KPHWHHRMIARVANILYD  1 AKPDYVQACAARIE    FRESIFTKAMSM    QVWKP---    EVERPGRHFVYMRRYAIFMARLLFLI
MGG_00105_Mgri_145611812              -KSNWQHRMIIRHARILFD 12 GKEVFMGAKAAFSV    LRESMFTKTMVM    NVWKC---    DAERPGRHHVYTEQYVRFMTRLLVAM
NCU04512_Ncra_164429378               -KSNWQHRMIIRHARLLFD 12 KEASKEAASKAFAV    LRENMFTKTMVM    NVWKC---    DAERPGRHHVYTEQYIRYMARLLAVM
CHGG_04295_Cglo_116195394             -KSNWQHRIIMRHARILFD  6 DNESSAAAKAAFNV    LRESMFTKTMVM    NVWKC---    DAERPGRHHVYTEQYIRYMVRVLDVM
PODANSg7663_Pans_171691398            -KSNWQHRIIMRHARVLFP  6 DPANVDAAKAAFAV    LRENMFTKTMVM    NVWKC---    EAERSGRHYVYTWRYIKYVAKILAAL
SNOG_08212_Pnod_169610231             -KSSWHHRIISRSAHIIFD    EGKDSMMAHGAKHE    LTQQMFTKTMAV    QVWKP---    ENERPGRHFVYTTRYTRFFVDLLDLT
FG04030.1_Gzea_46116376               -KANWQHRIIIRHAKILFD  6 SEDRLVNAKAAFGV    LRDSMFTKTMVM    NVWKC---    DAERPGRHHVYTEQYVRLMTKLLVIM
hip3_Spom_19113368                    -KQSWHHRPTYRIAKIMEH    -LGNVQQAKEEMET    LFSYKTSGKSLL    NIWRT---    PNERPGRHFYYGATYSRYLLSLFYKT
BC1G_12898_Bfuc_154295790             -KPHWHHRMIARVANVVYD  1 SNPDYEKACLARSE    FRESIFTKTMSM    QVWKP---    EAERPGRHFVYMRRYALFMARLLFLT
AN3123.2_Anid_67525331                -KANWHHRMAVRSARIIYD    DNKGAIAAAEAKSE    LTQQIFTKTMTI    QVWKP---    EYERPGRHFVYTTRYAYFFVTLLDQL
KLLA0B03850g_Klac_50303527            -KQKWQHRPIYRIAQIKYT    EFEDTDGAMKEMNK    LLALKSVNKNVV    NIWKP---    ENELPGKHFVYAYQYVLFYMDLLNEK
CAGL0M00286g_Cgla_50293891            -KKKWQHEPYFYIAEILFM    EMKDVEKSREWMDN    IISIRSVNKNLV    NIWKP---    DYERPGKHFVFTHNYVMFYLELLKYD
AGOS_ABL081W_Agos_45185149            -KRKWHHRPTYRIARILFD    DFNDVDGAIQEMGV    LMALKSVNKNRV    NIWKP---    EYERPGKHSVYTYQYVMFYLKLLSHK
HIR3_Scer_151945205                   -KKKWQHRPRYRIARILFD    DLGDVNGALEEMDS    LISAKSINKNLV    NIWKP---    DFERPGKHFIYTYQYLVLYLDLLFAI
Lgig1000008602_Lgig_Lgig1000008602    -RFPNHYKSVYRLAYAYTY  2 NHKNLQFARDLLLG 15 FHERKTANFFQ-    GIWKIPVQ    EIDRAGSFASHMNKAICLLLDILKEL
LOC412746_Amel_110758961              -RFSEHYKSFYRLSHFFFN  2 TAKDIIKCRNLLLD 15 FADRKSTNFFN-    GVWHIPNR    EVDRPGSFAFHMSRCVTLLMQVLKET
Cabin1_Rnor_166157458                 -RFPQHYKSLYRLAFLYTY  2 THRNLQWARDVLLG 16 FCERNKTNFFN-    GIWRIPVD    EIDRPGSFAWHMNRSIVLLLKVLAQL
Cabin1_Mmus_70995287                  -RFPQHYKSLYRLAFLYTY  2 THRNLQWARDVLLG 16 FCERNKTNFFN-    GIWRIPVD    EIDRPGSFAWHMNRSIVLLLKVLAQL
GSTEN:00021655:G:001_Tnig_47217598    -RFPQHYKSLYRLAFFYTN  2 THQNLQWARDVLLG 16 FCERNKTNLFNR 73 GIWRIPVD    EIDRPGSFASHMNRSIVLLLEVLYQL
PHYPADRAFT_139830_Ppat_162675462      GELKHFHKARYRLALGLHC  2 EDHDLERAKEELA-    FCFKTHRSLFTI    NMWEIDES 18 MPESSRKFISCVRKYLIFYLTLCEEA
Smoe1000001625_Smoe_Smoe1000001625    GELKHFHKARFCIAQGLYC  2 EEQDWERAKEELA-    FCFRSSRSLFTI    NMWELDGV 21 LPESSRKFITCVRKYLLLYFVLCEKA
AT4G32820_Atha_15233957               GDLKHFHKARYMLAQGLYR  2 GSSDLQRAKEELS-    FCFKSSRSSFTI    NMWEIDGM 21 LPESSRKFITCIRKYLLFYLRLLEET
OSJNBb0024N19.17_Osat_50872476        GELKHFHKARYKLAQGLYR  2 EAGDLERAKEELS-    FCFKSTRSSFTV    NMWEIDGS 21 LSESSRKFITCIRKYMILYLNLLEKN
DDBDRAFT_0189500_Ddis_66827817        -IVSYHHQASYRIVWTIRN  3 KDNFISNAMNEFL-    KFFKPKMSRLSK  1 CIWSLWND    DYVGDGKMDRCFAKYYKLYIQLLEEN
consensus/90%                         .c....a+..hb.A..h..    ...s...A...b..    h......p....    .hWc....    ..-ps.pa.....p.h.hhhplh...

Predicted Secondary Structure         --HHHHHHHHHHHH--------EE-HHH---HHHHHHHHHHHHHHHH----------------------HHHHHHHHHH------------------
SS1G_01112_Sscl_156061996             NDKANMELLAKRVR---KRAVDYYKFSD---VWSECCTIYLRLIRKAA---NIPANVDEV-FRGIS----QEGWDAYSQRIEDSIKD--GSLSHPA---
MGG_00105_Mgri_145611812              DDRVNLEALLRRIR---KKGADFYHFNE---LWHYCVLAYTRLIRQAY---QIPTTEEET-FKNVS----LDEFEIVAERITEWAVKL-KAEEHSA---
NCU04512_Ncra_164429378               NDRANLEAVLRRIR---KKGVDFYHFSE---LWQHGVQIYLKMIRKAF---NIPANDEDP-FKNVA----PEDFEAVAEKIAEWVITP-EAESHAA---
CHGG_04295_Cglo_116195394             KDRVSLEAVLRRIR---KKGADFYHFNE---LWQFGVLAYVKLLRRTF---EVPLTEEDA-FKTLS----QEEFDTVGEKITAWSTTP-AAEDNAA---
PODANSg7663_Pans_171691398            NDRTNLEAVLRRIR---KRGADFYYFNE---LWQYCVQIYLKLIRQTF---NVPSAGEDA-FKTLN----TEEFDVVGEKITEWATTP-AAESHPA---
SNOG_08212_Pnod_169610231             GDKANFEALAKRVR---RKTADFFEHSK---LWQDLCMRYLKMLRKLG---NIPEGQEDSAFKSVN----QEEFNTLALRMEAWCQN--PLTQHPV---
FG04030.1_Gzea_46116376               NDRGNLEQLLRRLR---KKGADFYHFTD---LWQSCCMAYLKLLRQGY---NVSPVADDA-FKALS----SEEFEVVGERIADWAAS--DGAENPL---
hip3_Spom_19113368                    NDKVNFLQFLKRFR---RSSSTIYEHRQ---IWLDIMIKYLEDLRLQH---SVKETQIND-------LP-LVEFKYVYKEISLLDEQK----------
BC1G_12898_Bfuc_154295790             DDKPNMELLAKRVR---KRAVDYFKFGD---VWNECCTIYLRLIRKAA---AIPANVDEV-FRGIT----QEVWEAYSQRLDDSIKD--GSLSHPA---
AN3123.2_Anid_67525331                DDRANLEQLLRRVR---KKQGDFINHTK---LWEDMCLTYARVIRRAA---GISEGHEESVFKPIG----WEEFSSNTARLEELPQL--APESPSL---
KLLA0B03850g_Klac_50303527            HDFMAIGGMVKKLR---RFGSGMINSQD---AINKAVELFVNGAKVKL-SLNEKEHGELL-MQRIP----FPEFVELSEELFNAFKK--GDYQSDV---
CAGL0M00286g_Cgla_50293891            KDAQSIGLLTKKLR---RFSSGMAYIGK---ANDVATSYFIECVTKKY-LLHDKVYTESY-MGNIN----YNDFVKYSQEIVDNSKA--ADIAEDI---
AGOS_ABL081W_Agos_45185149            NDYISLGHAARKMR---RFGSGMVNGSL---ATDKAVEMFINGARAAL-GINEKEHAELL-LPTLN----YQAFNRYSDELIASFNK--ENYSYEV---
HIR3_Scer_151945205                   KDFNTTGLVIKKLR---RFGSGTVNVNE---LLERAINVYTQSAKIKL-QLQDKSYVEQI-LPTLN----YQEFLKISEQLNQVFDQ--GKYPEEI---
Lgig1000008602_Lgig_Lgig1000008602    NDNRMLYNVHVQLLRVPDSGKKFLRDSER-------IYLWKQAYKNTI--DVLMNKCDQL-DKENNE---EKRLSFLMDVYKVWHHGV-HKINK-----
LOC412746_Amel_110758961              NDGRMLMQLSVQLGKIPESDKKYLRDSEREQLSCQALTLCLQSLRTKV--QLMGTSTSDL-AINKNA---DARTQVLLDTYWVYQRAL-KSFQTKEQ--
Cabin1_Rnor_166157458                 RDHSTLLKVSSMLQRTPDQGKKYLRDADRQVLAQRAFILTVKVLEDTL--SELAEGLEHP-GSKACGLS-GARMTTDVSHKASPEDGQ-ESLPHPK---
Cabin1_Mmus_70995287                  RDHSTLLKVSSMLQRTPDQGKKYLRDADRQVLAQRAFILTVKVLEDTL--SELAEGSEHP-GSKACGLS-GARMTTDVSHKASPEDGQ-ESPQHPK---
GSTEN:00021655:G:001_Tnig_47217598    KDHDTLLKVSFMLQRTPDQGKKYLRDGDRQVLAKRAFFLTVKVLEDNLNISALFQVSEQP-HKVPATPSVGEMTTTDVSSKPSSEDGR-TALPK-----
PHYPADRAFT_139830_Ppat_162675462      EDLGTLERAYTCL----KADRKFSLCLED--ISRVALGKYILALARAI--CQTDSEGSTS-QTLLHPLL-EKMFNLYMDFGISWSDSAGLSLAEAGIMN
Smoe1000001625_Smoe_Smoe1000001625    GDYVTLERAYSTL----RIDKRFSICLED--IASAALSSYVHALGATI--TQAESSSSSSS-FSSLQGLL-ERMFNLFMDHGIWWAETS-----------
AT4G32820_Atha_15233957               EDVNTLERAFNSL----RSDKRFSLCVED--LVPVAIGRYVKALVSSM----SRVESAGA-IINPDSQL-EKIFSLFIEQGSIWPDIC--NFPETR---
OSJNBb0024N19.17_Osat_50872476        RDLWTLERAYTYL----RTDKRFALCLGD--IVPVGLGKYLQVLTSAI--RNPEIRRVSG-DASVENLL-EKMFGVFMDHANLWADIS--TIPEVN---
DDBDRAFT_0189500_Ddis_66827817        NDIKNLEIIYFKI----KNDPKFAKEATQ--CYLSCCKVLTNLINLKL-QQQQESQQLFQ-KSISN----TAAGGNDLSNDENPSNS--SALSTIS---
consensus/90%                         pD..sh..h...l....c.s..h....p...h....h....p.h...h.............................pp..................
```

```
Predicted Secondary Structure         --HHHH-------------------------------------------------------------------    ----------------
SS1G_01112_Sscl_156061996             LDALRET-----------------IELKRLN----NGL----AKITPIDDMINDIW--AVLYIQIVRDTPAIDV    SALQQSQID-------
MGG_00105_Mgri_145611812              FDAMKEA-----------------IELKKLN----ANL----TKATPIDDLVSDCY--CLIYSEVGSDLPGPPP    AQILEDRAKQKEA---
NCU04512_Ncra_164429378               LNAMKEA-----------------IELKKLN----ANL----MKAGPIDDLINDCY--SIIHHEMVPEQTASET    PQGGDEHSRAGEGELP
CHGG_04295_Cglo_116195394             LTAMKEA-----------------VEIKKLN----SNL----MKAGPIDDLITDCY--SAIYLHVKPDLLDSDQ    QKLPHHPPGDPE----
PODANSg7663_Pans_171691398            LNAMKEA-----------------VELKKLN----SNL----MKAGPIDDLITDCY--STIYLDVRPELPRPAE    HSTETQPGEDSSQ---
SNOG_08212_Pnod_169610231             LDILRDA-----------------IELKRLN----NGL----MKPLQIDDLIGDTY--ALLYTTIGPSLPPLPS    EQQQQQLVPST-----
FG04030.1_Gzea_46116376               FNCMKEA-----------------IEFKKLN----ANL----MKVAPIDDLLYDCY--SRIYSEIAKTLPGPDP    SKVVEERHHAKEV---
hip3_Spom_19113368                    LSLLHQV-----------------YEIRKLN----NGL----YPTIKVDDFLIDCF--MSLYNEVKSSISPL--    -DANIPNSPSTIT---
BC1G_12898_Bfuc_154295790             LDALRET-----------------IELKKLN----NGL----AKITPIDDMINDIW--AVLYIQVVRDTPAIDI    TTLNQSQID-------
AN3123.2_Anid_67525331                LELLRDA-----------------IELKKLN----NNL----MKVALLEDLIADVY--SRIYEVNLPILL----    ----------------
KLLA0B03850g_Klac_50303527            LDVFLLA-----------------YNLKKGT----NSI--------QFDGVCITIY-FKYFYCPFVENKKQSTP 22 PQSQSQPQLPRSP-S-
CAGL0M00286g_Cgla_50293891            LEVLKIA-----------------YQLKKGT----NSV--------VYDTVCLSLF-FQYIYEPYRKEHPETEP  2 VSDSENKSN-------
AGOS_ABL081W_Agos_45185149            LESLAIS-----------------YQLKKGS----SGI--------AFDGVCLSIY-FKYLYLPWVEEHEAAEA  9 INKNVNPTENETA-S-
HIR3_Scer_151945205                   SSGLKLA-----------------FQLKKGH----SGI--------AFDSVCLGIY-FEYLYFPLARQDQSLTD    VNDENNPAL-------
Lgig1000008602_Lgig_Lgig1000008602    LTELTNK-----------------LFIRTYRLARRDKVDNNIAVIEQASKFCQ--------------------    -QRQNQTKLPQIP---
LOC412746_Amel_110758961              TIQVLAS-----------------LLVDTYK-TYIGNKNLEGNVLEIATKYCNDYNYNRKVLSKFFSSLDKPTE  2 TQPQATSPIPVTT---
Cabin1_Rnor_166157458                 KLPLADGSGPGPEPGGKVGPLHQLPVATDTR-----------DNTEQGGEPKDKER--PPVGPTEPMDTGETAA    RHPDLEPTPRLLP---
Cabin1_Mmus_70995287                  KLPLADGSGPGPEPGGRMGPLNQLPVVTDIR-----------DNTDQGGEQKDKES--PQVGPTEPMDTSEAAG    RHSDLEQTSRLLP---
GSTEN:00021655:G:001_Tnig_47217598    KPRLAEGVSGTGPRDAAQAQLAP-PLLPSSE-----------RTREGLPEKVDPSG--HKAGPEEPMELEAGSW    RRPAGAAELPTAA---
PHYPADRAFT_139830_Ppat_162675462      SPEVAES--------AIYSYTHRYIQLLEAD----KRV----DALEVVNERIRKRF----KNPKLVGAQSGQVC    KHASFAWCRALCT---
Smoe1000001625_Smoe_Smoe1000001625    ESEGQEV--------AVYCYIHRYLHYLEND----CRV----DALELVNERIRKRF----KHQKLSNPHFSQIC    RHATVAWCRALCS---
AT4G32820_Atha_15233957               GPETSES--------SLYRYLHQYIVSLELD----NKV----ETLETINEKIRKRF----KNPKLSNSFSAKVG    RHASLAWCRALII---
OSJNBb0024N19.17_Osat_50872476        SPELSES--------NLYSYIHQYIHLLESD----VRL----DVLEGLNEKIRKRF----KTPKLSNSNFAKIC    KHASLAWCRCILI---
DDBDRAFT_0189500_Ddis_66827817        QTATTTI-----------------STTTIIP--------------EFETILKYVWE--FYLDLSIFPEHRSSIQ    KLVMDSWNFNNQI---
consensus/90%                         ......s...................h.p.p................s.........................    ................

Predicted Secondary Structure         ------------------
SS1G_01112_Sscl_156061996             GEAT------ARASGPMNL  1690
MGG_00105_Mgri_145611812              AAAA------SAEERPTGL  1713
NCU04512_Ncra_164429378               TKPA------AEGEEPKSM  1702
CHGG_04295_Cglo_116195394             SKPD------SSNPKSTSL  1715
PODANSg7663_Pans_171691398            QPPP------SASQDGVGV  1734
SNOG_08212_Pnod_169610231             QPFI------AQHGGALPI  1771
FG04030.1_Gzea_46116376               AAQL------EAAQTETKV  1750
hip3_Spom_19113368                    AKPL------NDEKAINNE  1560
BC1G_12898_Bfuc_154295790             GEAA------ARPSGPMTL  1690
AN3123.2_Anid_67525331                EQVD------EENKEKMKV  1754
KLLA0B03850g_Klac_50303527            PNTA-NNKE-KTAKTPVKN  1625
CAGL0M00286g_Cgla_50293891            ------------------   1593
AGOS_ABL081W_Agos_45185149            PQPA-AEPQ-KSTTKQVSS  1624
HIR3_Scer_151945205                   PSSG-SVIS-KSTPDPTSK  1623
Lgig1000008602_Lgig_Lgig1000008602    HRPA-LTLF-KPQESQSDS  2164
LOC412746_Amel_110758961              SQPQ-ITST-SPQSSQSRK  1820
Cabin1_Rnor_166157458                 GRP--PRDR-GPESRSAEL  1797
Cabin1_Mmus_70995287                  GRP--PRDR-GPESRSAEL  1797
GSTEN:00021655:G:001_Tnig_47217598    GTSAEPLHK-VEGGRTSEL  1929
PHYPADRAFT_139830_Ppat_162675462      ALCS-ITTI-SQ--DTIEA  1637
Smoe1000001625_Smoe_Smoe1000001625    ALAS-VTPL-ERLPDHTPQ  1622
AT4G32820_Atha_15233957               SLAL-ITPL-QPVSSEESQ  1611
OSJNBb0024N19.17_Osat_50872476        KLAS-ITPL-PESMETTDQ  1612
DDBDRAFT_0189500_Ddis_66827817        NTIS-PRKS-SPTFEEVIK  2009
consensus/90%                         ..................
```

Figure S2: Multiple sequence alignment of the Hir3 domain. Proteins are denoted by their gene names, species abbreviations and gis separated by underscores. The predicted secondary structure is shown at the top of the alignment where, H: helix and E-strand. Shown below the alignment is the conservation profile at 90% consensus. Consensus abbreviations are as follows: p: polar residues (QNSTCHKRDE), o: alcohol side chain containing residues (ST), c: charged residues (HRKDE), '-': acidic residues (DE), '+': basic residues (HKR), b: big residues (KFILMQRWYE), h: hydrophobic residues (ACFMWYIVL), l: aliphatic residues (IVL), a: aromatic residues (FYWH), u:tiny (GAS) and s: small residues (GASVTDNPC). Species abbreviations are as follows: Agos *: Ashbya gossypii*; Amel : *Apis mellifera*; Anid : *Aspergillus nidulans*; Atha : *Arabidopsis thaliana*; Bfuc *: Botryotinia fuckeliana*; Cgla : *Candida glabrata*; Cglo : *Chaetomium globosum*; Ddis : *Dictyostelium discoideum*; Gzea : *Gibberella zeae*; Klac : *Kluyveromyces lactis*; Lgig : *Lottia gigantea*; Mgri : *Magnaporthe grisea*; Mmus : *Mus musculus*; Ncra : *Neurospora crassa*; Osat : *Oryza sativa*; Pans : *Podospora anserina*; Pnod : *Phaeosphaeria nodorum*; Ppat : *Physcomitrella patens*; Rnor : *Rattus norvegicus*; Scer : *Saccharomyces cerevisiae*; Smoe : *Selaginella moellendorffii*; Spom *: Schizosaccharomyces pombe*; Sscl : *Sclerotinia sclerotiorum*; Tnig : *Tetraodon nigroviridis*.

3.1 Identification of potential roles of HPC2 in distinct contexts like transcription and DNA repair/replication

We investigated if Hpc2 has multiple roles in the context of cellular processes using the available genetic/protein interactions data on yeast (obtained from the BIOGRID database). For this purpose, we first obtained entire set of genetic/protein interactions of Hpc2 and its sixty nine interacting partners (shown in Fig S3). We then represented this information as an undirected network (Hpc2-part-net), as this representation provides a convenient basis for further analysis. There were 1940 genes/proteins (nodes) and 9505 interactions in Hpc2-part-net. Further, we sought to identify sub-networks in Hpc2-part-net as it might provide insights in to roles of Hpc2. We identified 11890 unique cliques, which are complete sub-networks, in the network using software encoding a heuristic procedure obtained from Yu *et al* {16455753}. However, cliques are a strict definition of sub-network in the context of functional relevance. Hence, we sought to identify dense sub-networks of Hpc2-part-net using the already determined cliques in the network. We defined a dense sub-network as a collection of cliques, determined through a heuristic single linkage clustering algorithm, such that a given clique in the sub-network has at least another clique in the same sub-network which shares 75% or more of its nodes. We also ensured that a clique is uniquely assigned to a sub-network with largest cliques (in terms of number of nodes and edges). This is achieved by employing a standard single linkage clustering method with largest clique in the network as starting point. One of the largest dense sub-networks (see Fig S4 below) in the Hpc2-part-net that contained all components of the Hir /Asf1 complex provided interesting insights on to the distinct roles of Hpc2. This sub-network contained 82 nodes (genes/proteins) and 1206 interactions (genetic/protein interactions) between them.

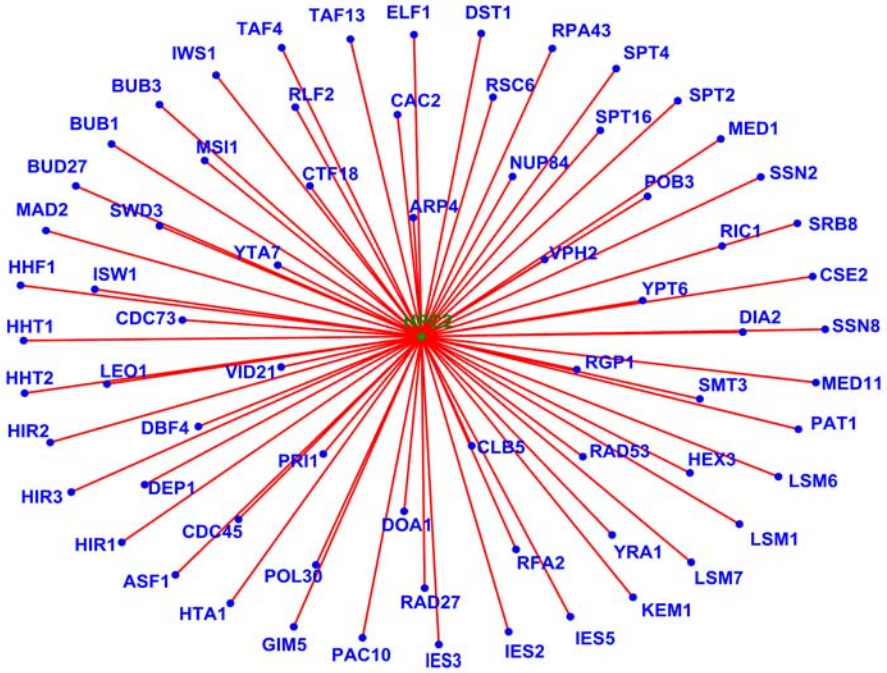

Figure S3: A network representation of set of genes/proteins (69 of them totally) that have genetic/protein interactions as compiled by the BIOGRID database. These interacting partners of the Hpc2 have been approximately grouped according to the functional groups like transcription elongation, chromatin remodeling, mediator complex, RNA processing, replication/repair and kinetochore function.

As shown in Fig S3, Hpc2 has interactions with many genes/proteins that have distinct functions like ubiquitin-pathway (Hex3, Smt3 and Dia2), kinetochore (Mad2, Bud27, Bub1 and Bub3) and RNA processing (like Kem1, Lsm1, Lsm6, Lsm7 and Pat1). We also found that two sets of genes/proteins with few interactions between them which are indicated by red and orange circles in the Fig S4. These genes/proteins are in distinct functional groups namely transcription (marked within red circle) and DNA repair/replication (marked within orange circle). It is interesting to note that Hta1, histone H2A, interacts largely with the members of transcription functional group (marked in red circle).

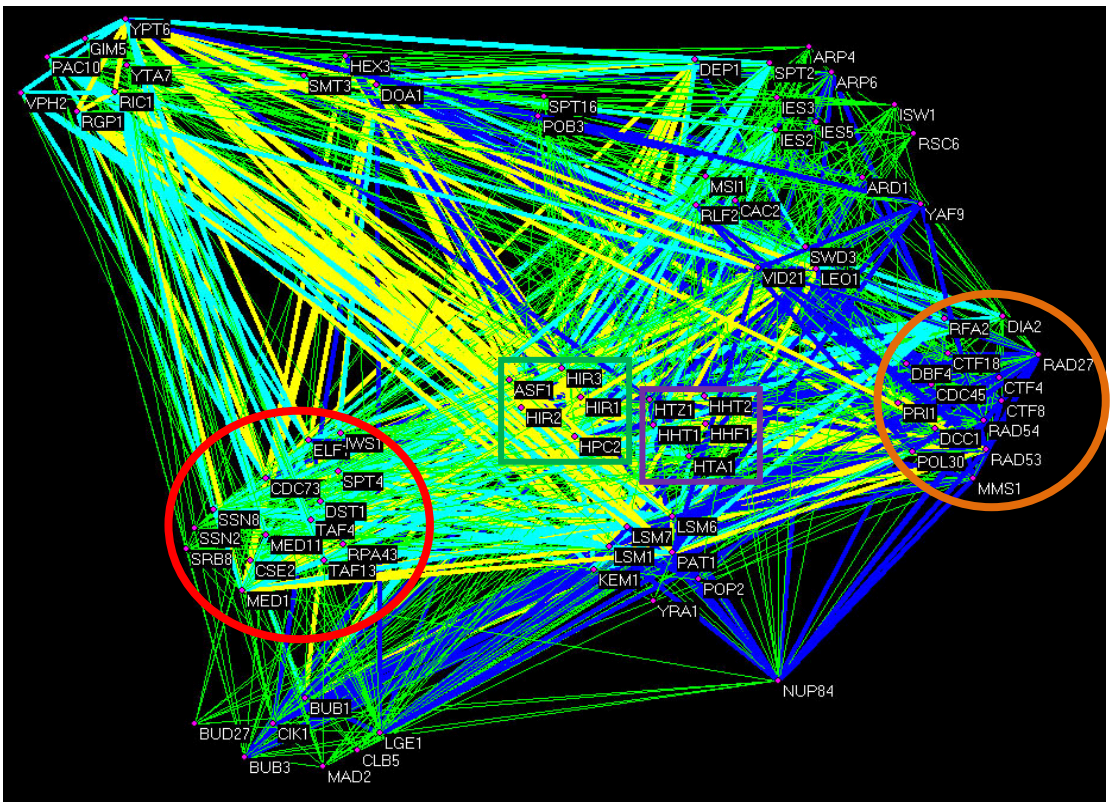

Figure S4: A large dense sub-network in Hpc2-part-net with 82 genes/proteins and 1206 interactions among them. The edges denoted by yellow connects the genes/proteins that are equally part of transcription and repair/replication process, while the cyan and blue edges respectively connects the genes/proteins that have roles specific to transcription (also shown within red circle) and repair/replication (also shown within orange circle). The green edges connects the nodes that belongs to the neither of the above mentioned category.

4.1 Identification of largest dense sub-network of protein pairs containing HUN domain protein FLJ25778 constructed based on high mutual information in their phyletic patterns

A recent study reporting a genome-scale RNAi profiling of cell division in human tissue culture cells (Hela cells) identified 1246 genes that are likely to be involved in the cell cycle {17994010}, including HUN domain protein encoding gene FLJ25778. It would be of particular interest to determine the set of proteins that co-evolve with FLJ25778, which might provide insights in to its cellular roles. For this purpose, we determined the phyletic pattern/profile based on identification of orthologs (see below for details on the methodology) of 1236 protein products of the above-mentioned genes and then evaluated the mutual information coefficient (MIC) for every

possible pairs of proteins in the list. Further, we extracted high MIC pairs of protein (using MIC value of 0.6 or more as threshold). This resulted in the identification of 1208 protein pairs, involving 175 proteins, out of which 14 proteins have high MIC with the HUN domain protein FLJ25778. This information, in its entirety, was represented as undirected network for the purpose of further analysis. In order to better understand the significance of the co-evolution of these proteins with FLJ25778, we seek to find out if these 14 proteins co-evolve even between themselves as this would further strength the findings. For this purpose, we identified all possible cliques in the network (as mentioned in the section 1.1) and we obtained 190 cliques and FLJ25778 was found to occur in two large cliques of 14 proteins. We then collated all the cliques in to dense sub-networks as described in the section 1.1. This resulted in a largest sub-network of 908 edges between 64 proteins, which also included FLJ25778 (see Fig 2B). We then investigated this sub-network in detail and strikingly we found numerous TFs were present in it suggesting potential critical role of FLJ25778 in the transcription regulatory process (see Fig 2B).

5.1 Analysis of phyletic patterns

5.1.1 Procedure to identify orthologous proteins

The procedure, used to identify orthologous proteins in a genome, is described as a flow chart below:

Flowchart describing the procedure to detect orthologs

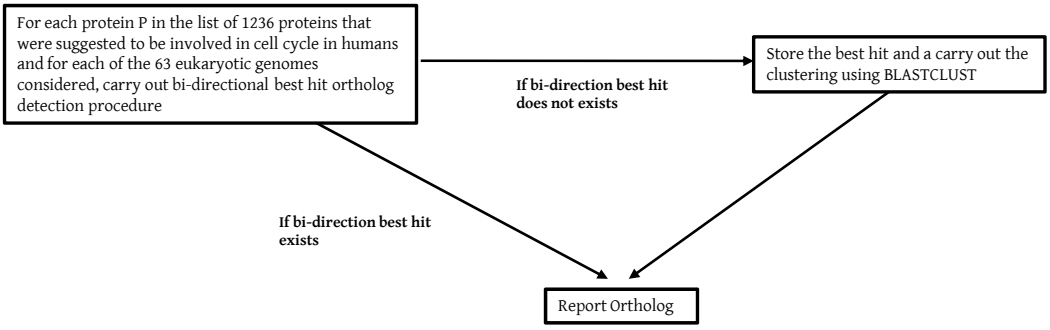

Figure S5: First a bi-directional best hit is performed, if an ortholog is not detected then BLASTCLUST procedure is adopted.

Bi-directional best-hit procedure: For each protein **P** in the list of 1236 human cell cycle proteins (from the list of 1246 cell cycle genes), a BLAST search was performed against the genome of interest (**x**). The best hit, sequence $\mathbf{P^x}$ from genome **x**, was then used as a query and a BLAST search was carried out against the human genome. If the best hit using $\mathbf{P^x}$ as the query happens to be **P** in human, then **P** & $\mathbf{P^x}$ were considered as orthologous proteins. If however $\mathbf{P^x}$ does not pick up **P** from human as its best hit, then a BLASTCLUST {12097341} procedure was adopted.

BLASTclust algorithm: For each of the proteins **P** in human for which the above-mentioned procedure did not pick up orthologous proteins, the best-hit sequence $\mathbf{P^x}$ (using **P** as the query against genome **x**) for each genome were obtained. Thus, for every protein **P**, this procedure gives us a set of proteins which are the best hits from genomes where bi-directional best hit procedure fails. The set of sequences thus obtained along with

the human protein is taken through a BLASTclust procedure using length conservation (L) of 60% and a score density (S) of 60% (Lespinet et al., 2002). Score density (S) is defined as the ratio of number of identical residues in the alignment to the length of the alignment. Documentation is available at: http://bio.ifom-firc.it/docs/blast/readme.bcl.txt and in Lespinet et al {12097341}. This procedure produces clusters of sequences using the single linkage-clustering algorithm. All the sequences belonging to the cluster that also contains the human protein **P** are considered as orthologs (see schematic of the ortholog detection procedure). Analysis of the clusters reveal that the parameters **S**=0.6 and **L**=0.6 performs best with an optimum coverage and lowest false positive rate. See **Fig S6** for a schematic.

Schematic of the procedure to detect orthologous proteins

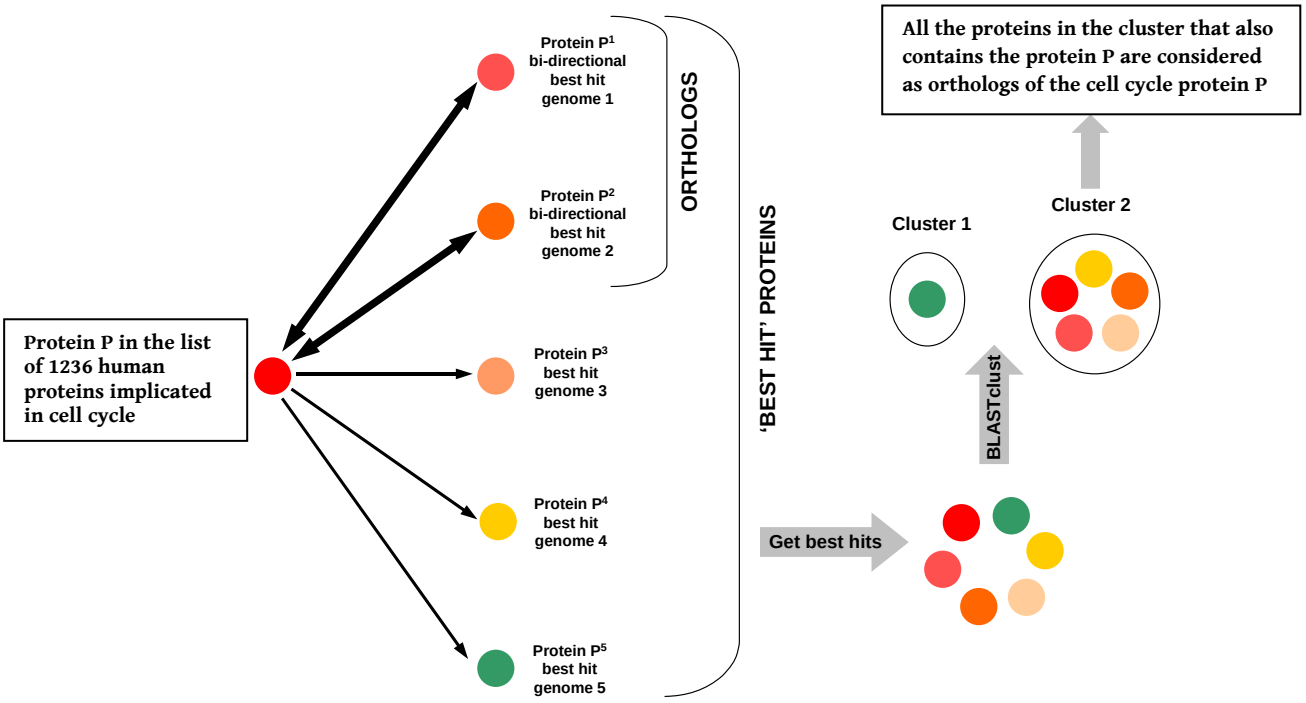

Figure S6: This figure illustrates the steps involved in ortholog detection. Information about orthologs for every genome is stored as an 'orthomap' file and this information was used to construct phyletic patterns for every protein in the list.

5.1.2 Distribution of correlation coefficients in the phyletic patterns of 70 proteins (Hpc2p and its interacting proteins shown in Fig S3)

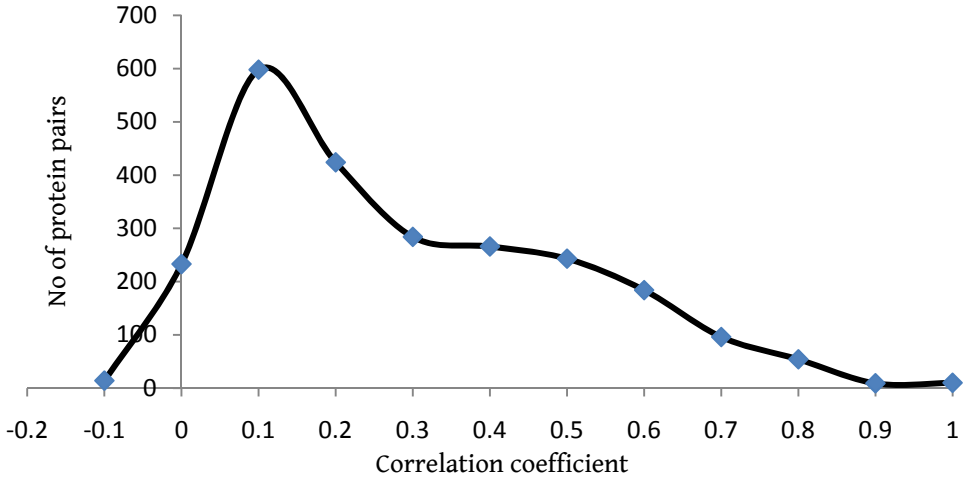

Figure S7: Distribution of correlation coefficient between 1245 pairs of phyletic profiles for 70 proteins (including Hpc2p) corresponding to list of protein/genes that interacts with Hpc2 (Fig S3)and orthology detection was made as described in section 3.1.1 except that only fungal genomes were considered for the analysis. Please see Table S2 for further details.

5.1.3 Distribution of Mutual Information Coefficients (MICs) in phyletic patterns of human cell cycle proteins

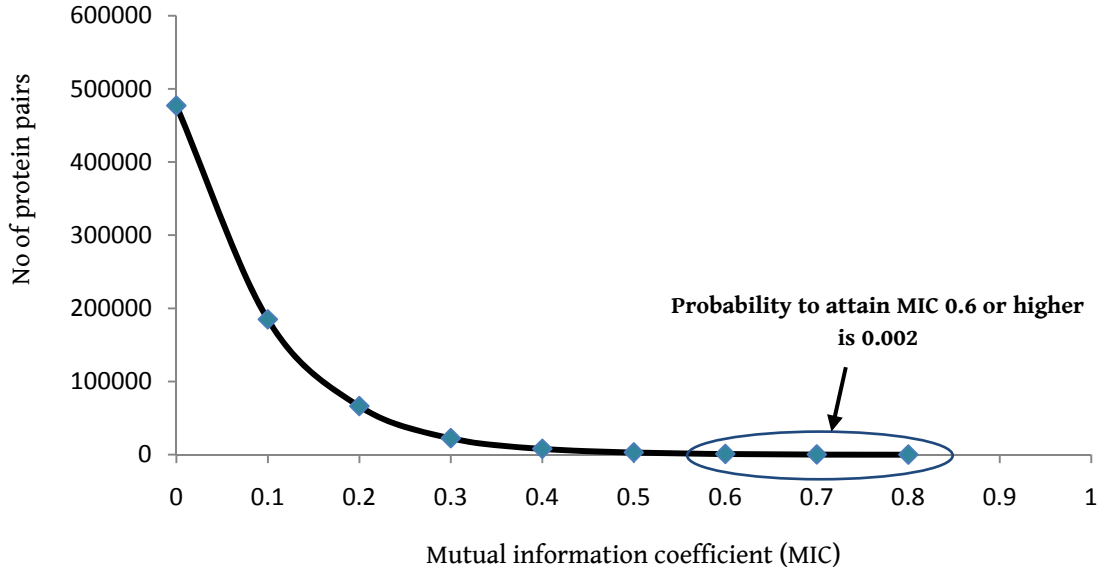

Figure S8: Distribution of MICs between 763230 pairs of phyletic profiles for 1236 proteins corresponding to the list of 1246 human cell cycle genes.

Table S1: Phyletic occurrence of components of Hir complex, Asf1 and Cac2

Protein	**Phyletic occurrence**
Hir1p	Animals, fungi, amoebozoans and plants, the *Guillardia* nucleomorph, chromalveolates apicomplexans, ciliates, oomycetes and diatoms, heteroloboseans, kinetoplastids, parabasalids and diplomonads
Hir2p	Animals, fungi, amoebozoans and plants, the *Guillardia* nucleomorph, chromalveolates apicomplexans, ciliates, oomycetes and diatoms, heteroloboseans, kinetoplastids, parabasalids and diplomonads
Hir3p	Animals, amoebozoans, plants, chlorophyte alga, chromists and heteroloboseans
Hpc2p	*All eukaryotes except Giardia*, kinetoplastids and the *Guillardia* nucleomorph
Asf1p	Animals, fungi, amoebozoans and plants, the *Guillardia* nucleomorph, chromalveolates apicomplexans, ciliates, oomycetes and diatoms, heteroloboseans, kinetoplastids, parabasalids and diplomonads
Cac2p	Animals, fungi, amoebozoans and plants, the *Guillardia* nucleomorph, chromalveolates apicomplexans, ciliates, oomycetes and diatoms, heteroloboseans, kinetoplastids, parabasalids and diplomonads

Table S2: Correlation coefficients in phyletic patterns between Hir/Asf1 complex components

Protein 1	Protein 2	Correlation coefficient
Hpc2p	Hir1p	0.18
Hpc2p	Hir2p	0.67
Hpc2p	Hir3p	0.92
Hpc2p	Asf1p	0.15
Asf1p	Hir1p	0.35
Asf1p	Hir2p	0.15
Asf1p	Hir3p	0.16
Hir1p	Hir2p	-0.02
Hir1p	Hir3p	0.21
Hir2p	Hir3p	0.62

Table S3: List of 63 eukaryotic genomes considered for the analysis

Anopheles_gambiae
Apis_mellifera
Arabidopsis_thaliana
Aspergillus_fumigatus
Aureococcus_anophagefferens_FG
Babesia_bovis
Branchiostoma_floridae_FG
Brugia_malayi_FG
Caenorhabditis_elegans
Candida_glabrata_CBS138
Capitella_spI_FG
Chlamydomonas_reinhardtii_FG
Ciona_intestinalis_FG
Cryptococcus_neoformans_var_JEC21
Cryptosporidium_parvum_NF
Cyanidioschyzon_merolae_FG
Danio_rerio
Daphnia_pulex_FG
Debaryomyces_hansenii_CBS767
Dictyostelium_discoideum_NF
Drosophila_melanogaster
Encephalitozoon_cuniculi
Entamoeba_histolytica_NF
Eremothecium_gossypii
Giardia_lamblia_NF
Gibberella_zeae_NF
Guillardia_theta_NF
Helobdella_robusta_FG
Kluyveromyces_lactis_NRRL_Y-1140
Leishmania_major_NF
Lottia_gigantea_FG
Monosiga_brevicollis_FG
Mus_musculus
Naegleria_gruberi_FG
Nematostella_vectensis_FG
Neurospora_crassa_NF
Ostreococcus_tauri
Pan_troglodytes
Paramecium_tetraurelia_FG
Petromyzon_marinus
Phaeodactylum_tricornutum_FG
Physcomitrella_patens_FG
Phytophthora_ramorum_FG
Phytophthora_sojae_FG
Plasmodium_falciparum
Rattus_norvegicus
Saccharomyces_cerevisiae
Schizosaccharomyces_pombe
Strongylocentrotus_purpuratus
Tetrahymena_thermophila_NF
Tetraodon_nigroviridis_NF
Thalassiosira_pseudonana_FG
Theileria_annulata_NF
Theileria_parva_Muguga_NF
Toxoplasma_gondii_FG
Tribolium_castaneum
Trichomonas_vaginalis_NF
Trichoplax_adhaerens_FG
Trypanosoma_brucei_NF
Trypanosoma_cruzi_CLBrener_NF
Ustilago_maydis_NF
Volvox_carteri_FG
Yarrowia_lipolytica_CLIB99

Supplementary Data

Listed below are the sub-clusters of proteins (each of them is denoted by '#', followed by number of proteins in it), which are sequence hits of Hir1p and Cac2p of yeast for searches against nr database of eukaryotes. These clusters were obtained using BLASTCLUST with sequence identity cut-off of 35% and matching sequence length threshold of 70%, for the largest cluster of 2171 proteins.

The largest cluster of 2171 proteins was obtained by collecting all the PSIBLAST hits of Hir1p (GI:6319463) and Cac2p (GI: 6323534) of yeast and clustering using BLASTCLUST with a sequence identity cut-off of 30% and matching sequence length threshold of 70%.

As mentioned in the main text, all eukaryotic lineages, with the exception of *Giardia and Trichomonas*, showed a clear differentiation of two paralogous groups respectively representing HirA and Cac2 implying that the differentiation of general and temporally specific histone chaperones first emerged prior to the separation of the heterolobosean-kinetoplastid clade from other eukaryotes. HirA and Cac2p clusters (major clusters) are indicated respectively as Hir and Cac2 clusters below.

Trichomonas and Giardia proteins in the first cluster (indicated by #1566) are in a distinct group compared to HirA and Cac2p clusters and they appear to be comparably related to both of them.

1566
145476531 hypothetical protein GSPATT00027923001 [Paramecium tetraurelia strain d4-2]
145545345 hypothetical protein GSPATT00003825001 [Paramecium tetraurelia strain d4-2]
145508654 hypothetical protein GSPATT00038832001 [Paramecium tetraurelia strain d4-2]
145492580 hypothetical protein GSPATT00034367001 [Paramecium tetraurelia strain d4-2]
145545815 hypothetical protein GSPATT00023922001 [Paramecium tetraurelia strain d4-2]
145482657 hypothetical protein GSPATT00030591001 [Paramecium tetraurelia strain d4-2]
145533739 hypothetical protein GSPATT00002617001 [Paramecium tetraurelia strain d4-2]
145544112 hypothetical protein GSPATT00023520001 [Paramecium tetraurelia strain d4-2]
145504572 hypothetical protein GSPATT00038639001 [Paramecium tetraurelia strain d4-2]
145528438 hypothetical protein GSPATT00039495001 [Paramecium tetraurelia strain d4-2]
145524131 hypothetical protein GSPATT00039395001 [Paramecium tetraurelia strain d4-2]
145551881 hypothetical protein GSPATT00026567001 [Paramecium tetraurelia strain d4-2]
149238730 conserved hypothetical protein [Lodderomyces elongisporus NRRL YB-4239]
134117680 hypothetical protein CNBL0890 [Cryptococcus neoformans var. neoformans B-3501A]
145517799 hypothetical protein GSPATT00039209001 [Paramecium tetraurelia strain d4-2]
156847202 hypothetical protein Kpol_1048p59 [Vanderwaltozyma polyspora DSM 70294]
169851989 hypothetical protein CC1G_08510 [Coprinopsis cinerea okayama7#130]
145551917 hypothetical protein GSPATT00026585001 [Paramecium tetraurelia strain d4-2]
156843356 hypothetical protein Kpol_1024p42 [Vanderwaltozyma polyspora DSM 70294]
50286567 unnamed protein product [Candida glabrata]
50309993 unnamed protein product [Kluyveromyces lactis]
156051964 hypothetical protein SS1G_07389 [Sclerotinia sclerotiorum 1980]
45190361 AEL246Cp [Ashbya gossypii ATCC 10895]
58270264 transcription initiation factor tfiid 90 kda subunit (tafii-90) [Cryptococcus neoformans var. neoformans JEC21]
148710078 TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor [Mus musculus]
194205706 PREDICTED: TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa [Equus caballus]
50294333 hypothetical protein CAGL0M05335g [Candida glabrata CBS138]
145499207 hypothetical protein GSPATT00036975001 [Paramecium tetraurelia strain d4-2]
62898962 Transcription initiation factor TFIID subunit 5 variant [Homo sapiens]
145550147 hypothetical protein GSPATT00025698001 [Paramecium tetraurelia strain d4-2]
47220310 unnamed protein product [Tetraodon nigroviridis]
1732075 TBP-associated factor [Homo sapiens]
26354080 unnamed protein product [Mus musculus]
47117222 TAF5_MOUSE Transcription initiation factor TFIID subunit 5 (Transcription initiation factor TFIID 100 kDa subunit) (TAF(II)100) (TAFII-100) (TAFII100)
1932938 TFIID subunit TAFII100 [Homo sapiens]
109090468 PREDICTED: similar to TBP-associated factor 5 [Macaca mulatta]
114632660 PREDICTED: TBP-associated factor 5 [Pan troglodytes]
127798463 TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa [Homo sapiens]
194041961 PREDICTED: TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa [Sus scrofa]
1491718 hTAFII100 [Homo sapiens]
145517103 hypothetical protein GSPATT00012568001 [Paramecium tetraurelia strain d4-2]

6319675 Subunit (90 kDa) of TFIID and SAGA complexes, involved in RNA polymerase II transcription initiation and in chromatin modification; Taf5p [Saccharomyces cerevisiae]
68477027 likely TFIID and SAGA complex component Taf5p [Candida albicans SC5314]
151946586 TafII90 [Saccharomyces cerevisiae YJM789]
157822291 TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor [Rattus norvegicus]
145523572 hypothetical protein GSPATT00039381001 [Paramecium tetraurelia strain d4-2]
154309684 hypothetical protein BC1G_07312 [Botryotinia fuckeliana B05.10]
71895077 TBP-associated factor 5 [Gallus gallus]
145476893 hypothetical protein GSPATT00028105001 [Paramecium tetraurelia strain d4-2]
182765471 TBP-associated factor 5 [Xenopus laevis]
150866779 TFIID and SAGA subunit [Pichia stipitis CBS 6054]
145478957 hypothetical protein GSPATT00029139001 [Paramecium tetraurelia strain d4-2]
50424771 hypothetical protein DEHA0F15235g [Debaryomyces hansenii CBS767]
195539684 Unknown (protein for MGC:186040) [Xenopus tropicalis]
116198367 conserved hypothetical protein [Chaetomium globosum CBS 148.51]
169843862 hypothetical protein CC1G_10527 [Coprinopsis cinerea okayama7#130]
171692415 unnamed protein product [Podospora anserina]
195032704 GH11222 [Drosophila grimshawi]
195437614 GK24403 [Drosophila willistoni]
164426529 hypothetical protein NCU04136 [Neurospora crassa OR74A]
45768550 Taf5 protein [Danio rerio]
70990416 transcription initiation factor TFIID subunit [Aspergillus fumigatus Af293]
119497155 transcription initiation factor TFIID subunit, putative [Neosartorya fischeri NRRL 181]
119570031 TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa, isoform CRA_b [Homo sapiens]
145229293 hypothetical protein An01g05300 [Aspergillus niger]
145606002 hypothetical protein MGG_06742 [Magnaporthe grisea 70-15]
121703171 transcription initiation factor TFIID subunit, putative [Aspergillus clavatus NRRL 1]
156847448 hypothetical protein Kpol_1028p23 [Vanderwaltozyma polyspora DSM 70294]
50546811 hypothetical protein [Yarrowia lipolytica]
115387245 conserved hypothetical protein [Aspergillus terreus NIH2624]
169766708 hypothetical protein [Aspergillus oryzae RIB40]
145484466 hypothetical protein GSPATT00031074001 [Paramecium tetraurelia strain d4-2]
145548369 hypothetical protein GSPATT00025202001 [Paramecium tetraurelia strain d4-2]
71018393 hypothetical protein UM03280.1 [Ustilago maydis 521]
16944416 related to TRANSCRIPTION INITIATION FACTOR TFIID 100 KDA SUBUNIT [Neurospora crassa]
119189439 hypothetical protein CIMG_04767 [Coccidioides immitis RS]
189194491 transcription initiation factor TFIID subunit 5 [Pyrenophora tritici-repentis Pt-1C-BFP]
167533195 predicted protein [Monosiga brevicollis MX1]
195118272 GI21646 [Drosophila mojavensis]
186703819 General transcriptional corepressor TUP1 [Zygosaccharomyces rouxii]
154280194 conserved hypothetical protein [Ajellomyces capsulatus NAm1]
173067 TUP1 protein
6319926 General repressor of transcription, forms complex with Cyc8p, involved in the establishment of repressive chromatin structure through interactions with histones H3 and H4, appears to enhance expression of some genes; Tup1p [Saccharomyces cerevisiae]
151943895 deoxythymidine monophosphate uptake protein [Saccharomyces cerevisiae YJM789]
190406502 glucose repression regulatory protein TUP1 [Saccharomyces cerevisiae RM11-1a]
195388320 GJ17774 [Drosophila virilis]
195123045 GI20797 [Drosophila mojavensis]
195382211 GJ20531 [Drosophila virilis]
195025710 GH20712 [Drosophila grimshawi]
300084 transcription factor TFIID dTAFII80 subunit [Drosophila melanogaster, embryo, Peptide, 704 aa]
17136870 TBP-associated factor 5 CG7704-PA [Drosophila melanogaster]
194758044 GF11078 [Drosophila ananassae]
194884145 GG22709 [Drosophila erecta]
195333257 GM20485 [Drosophila sechellia]
195427012 GK20967 [Drosophila willistoni]
195483603 GE13065 [Drosophila yakuba]
195582298 GD25948 [Drosophila simulans]
70984084 transcriptional repressor TupA/RocA [Aspergillus fumigatus Af293]
159122348 transcriptional repressor TupA/RocA, putative [Aspergillus fumigatus A1163]
17137500 ebi CG4063-PA [Drosophila melanogaster]
194853455 GG24716 [Drosophila erecta]
195350093 GM16739 [Drosophila sechellia]
195470256 GE16791 [Drosophila yakuba]
195575539 GD23022 [Drosophila simulans]
198459431 GA20529 [Drosophila pseudoobscura pseudoobscura]
146423322 hypothetical protein PGUG_00968 [Pichia guilliermondii ATCC 6260]
190345057 hypothetical protein PGUG_00968 [Pichia guilliermondii ATCC 6260]
194766525 GF20651 [Drosophila ananassae]

114581405 PREDICTED: hypothetical protein [Pan troglodytes]
109458577 PREDICTED: similar to CG3436-PA, isoform A [Rattus norvegicus]
125985353 GA17928 [Drosophila pseudoobscura pseudoobscura]
195147240 GL18875 [Drosophila persimilis]
170586452 F-box-like/WD-repeat protein ebi [Brugia malayi]
195172728 GL20028 [Drosophila persimilis]
148683736 notchless homolog 1 (Drosophila) [Mus musculus]
2245632 transcriptional repressor TUP1 [Kluyveromyces lactis]
50311047 TUP1_KLULA [Kluyveromyces lactis]
145478263 hypothetical protein GSPATT00028791001 [Paramecium tetraurelia strain d4-2]
67516021 hypothetical protein AN0292.2 [Aspergillus nidulans FGSC A4]
115768161 PREDICTED: similar to TAF5 protein [Strongylocentrotus purpuratus]
145524157 hypothetical protein GSPATT00015386001 [Paramecium tetraurelia strain d4-2]
157339727 unnamed protein product [Vitis vinifera]
145546219 hypothetical protein GSPATT00024124001 [Paramecium tetraurelia strain d4-2]
168034550 TFIID and SAGA complex subunit [Physcomitrella patens subsp. patens]
145511403 hypothetical protein GSPATT00010466001 [Paramecium tetraurelia strain d4-2]
193652571 PREDICTED: similar to AGAP011678-PA [Acyrthosiphon pisum]
79520288 TAF5 (TBP-ASSOCIATED FACTOR 5); nucleotide binding [Arabidopsis thaliana]
145547002 hypothetical protein GSPATT00024517001 [Paramecium tetraurelia strain d4-2]
169625031 hypothetical protein SNOG_15782 [Phaeosphaeria nodorum SN15]
74004640 PREDICTED: similar to WD repeat, SAM and U-box domain containing 1 [Canis familiaris]
115469236 Os06g0649500 [Oryza sativa (japonica cultivar-group)]
164656491 hypothetical protein MGL_3408 [Malassezia globosa CBS 7966]
91083813 PREDICTED: similar to wd-repeat protein [Tribolium castaneum]
156550101 PREDICTED: similar to wd-repeat protein [Nasonia vitripennis]
170035924 transcription initiation factor TFIID subunit 5 [Culex quinquefasciatus]
157115015 wd-repeat protein [Aedes aegypti]
158301063 AGAP011678-PA [Anopheles gambiae str. PEST]
50545019 YlTUP1 [Yarrowia lipolytica]
145510074 hypothetical protein GSPATT00038925001 [Paramecium tetraurelia strain d4-2]
19075402 transcription factor TFIID complex subunit 5 Taf72 [Schizosaccharomyces pombe]
50285811 unnamed protein product [Candida glabrata]
19113046 TATA-binding protein associated factor (PMID 11279037) [Schizosaccharomyces pombe]
72393257 hypothetical protein, conserved [Trypanosoma brucei TREU927]
74007099 PREDICTED: similar to transducin (beta)-like 1 X-linked [Canis familiaris]
45200863 AGL234Wp [Ashbya gossypii ATCC 10895]
149244152 transcriptional repressor TUP1 [Lodderomyces elongisporus NRRL YB-4239]
149634556 PREDICTED: similar to TBP-associated factor [Ornithorhynchus anatinus]
171686988 unnamed protein product [Podospora anserina]
11066216 AF197225_1 TUPA [Emericella nidulans]
46111239 hypothetical protein FG02501.1 [Gibberella zeae PH-1]
47679343 Tup1p [Exophiala dermatitidis]
116191811 hypothetical protein CHGG_05623 [Chaetomium globosum CBS 148.51]
145351604 predicted protein [Ostreococcus lucimarinus CCE9901]
170592499 WD40 associated region in TFIID subunit family protein [Brugia malayi]
19113785 transcriptional corepressor Tup11 [Schizosaccharomyces pombe 972h-]
149256553 PREDICTED: hypothetical protein [Mus musculus]
15240710 WD-40 repeat family protein [Arabidopsis thaliana]
34395322 putative WD-40 repeat protein family [Oryza sativa Japonica Group]
39953425 hypothetical protein MGG_08829 [Magnaporthe grisea 70-15]
50409865 hypothetical protein DEHA0A13838g [Debaryomyces hansenii CBS767]
170099900 chromatin associated protein [Laccaria bicolor S238N-H82]
199429183 DEHA2A13420p [Debaryomyces hansenii]
110761148 PREDICTED: similar to TBP-associated factor 5 CG7704-PA [Apis mellifera]
1698504 rco-1 gene product
85109447 hypothetical protein NCU06205 [Neurospora crassa OR74A]
145495202 hypothetical protein GSPATT00035326001 [Paramecium tetraurelia strain d4-2]
39577682 Tup1p [Pichia angusta]
41054115 TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor [Danio rerio]
164660164 hypothetical protein MGL_1388 [Malassezia globosa CBS 7966]
47214090 unnamed protein product [Tetraodon nigroviridis]
66806843 WD40 repeat-containing protein [Dictyostelium discoideum AX4]
169612435 hypothetical protein SNOG_09340 [Phaeosphaeria nodorum SN15]
194680022 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF-associated factor 65 beta) (PAF65-beta) [Bos taurus]
156048464 hypothetical protein SS1G_08963 [Sclerotinia sclerotiorum 1980]
156360693 predicted protein [Nematostella vectensis]
154297425 hypothetical protein BC1G_12309 [Botryotinia fuckeliana B05.10]

125543891 hypothetical protein OsI_011263 [Oryza sativa (indica cultivar-group)]
145532892 hypothetical protein GSPATT00039708001 [Paramecium tetraurelia strain d4-2]
169777547 hypothetical protein [Aspergillus oryzae RIB40]
189190562 transcriptional repressor rco-1 [Pyrenophora tritici-repentis Pt-1C-BFP]
156050805 hypothetical protein SS1G_07990 [Sclerotinia sclerotiorum 1980]
7657439 PCAF associated factor 65 beta isoform a [Homo sapiens]
19527184 TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor [Mus musculus]
73952622 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF associated factor 65 beta) (PAF65-beta) [Canis familiaris]
109020106 PREDICTED: PCAF associated factor 65 beta isoform 2 [Macaca mulatta]
118088239 PREDICTED: similar to PCAF associated factor 65 beta [Gallus gallus]
121703536 wd-repeat protein [Aspergillus clavatus NRRL 1]
126307247 PREDICTED: similar to PCAF associated factor 65 beta [Monodelphis domestica]
149758059 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF-associated factor 65 beta) (PAF65-beta) [Equus caballus]
169642580 Taf5l protein [Rattus norvegicus]
194042627 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF-associated factor 65 beta) (PAF65-beta) [Sus scrofa]
55742466 TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa [Xenopus tropicalis]
119467836 wd-repeat protein [Neosartorya fischeri NRRL 181]
148224790 hypothetical protein LOC734892 [Xenopus laevis]
169776901 hypothetical protein [Aspergillus oryzae RIB40]
171695104 unnamed protein product [Podospora anserina]
147907010 TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa [Xenopus laevis]
19113822 transcriptional corepressor Tup12 (PMID 10567571) [Schizosaccharomyces pombe]
115398498 transcriptional repressor rco-1 [Aspergillus terreus NIH2624]
126313963 PREDICTED: similar to Notchless homolog 1 (Drosophila) [Monodelphis domestica]
119196467 transcriptional repressor [Coccidioides immitis RS]
188037290 TupA [Emericella nidulans]
29465691 TupA protein [Penicillium marneffei]
145250079 hypothetical protein An15g00140 [Aspergillus niger]
195442412 GK18046 [Drosophila willistoni]
145532355 hypothetical protein GSPATT00018731001 [Paramecium tetraurelia strain d4-2]
156056266 hypothetical protein SS1G_05485 [Sclerotinia sclerotiorum 1980]
47227921 unnamed protein product [Tetraodon nigroviridis]
66814316 hypothetical protein DDBDRAFT_0206475 [Dictyostelium discoideum AX4]
66812260 transcriptional repressor TUP1 [Dictyostelium discoideum AX4]
5032159 transducin beta-like 1X [Homo sapiens]
145243266 hypothetical protein An11g01650 [Aspergillus niger]
158255360 unnamed protein product [Homo sapiens]
194227657 PREDICTED: similar to transducin (beta)-like 1 X-linked [Equus caballus]
114687677 PREDICTED: transducin beta-like 1X isoform 2 [Pan troglodytes]
145603105 hypothetical protein MGG_04473 [Magnaporthe grisea 70-15]
198422939 PREDICTED: similar to transducin (beta)-like 1 X-linked receptor 1 isoform 1 [Ciona intestinalis]
67968479 unnamed protein product [Macaca fascicularis]
146423584 hypothetical protein PGUG_01096 [Pichia guilliermondii ATCC 6260]
157348457 unnamed protein product [Vitis vinifera]
190345162 hypothetical protein PGUG_01096 [Pichia guilliermondii ATCC 6260]
154276878 transcriptional repressor TUP1 [Ajellomyces capsulatus NAm1]
145535916 hypothetical protein GSPATT00039749001 [Paramecium tetraurelia strain d4-2]
58267636 general transcriptional repressor [Cryptococcus neoformans var. neoformans JEC21]
33146605 putative TATA box binding protein-associated factor [Oryza sativa Japonica Group]
196001987 hypothetical protein TRIADDRAFT_37330 [Trichoplax adhaerens]
195168990 GL13421 [Drosophila persimilis]
114626281 PREDICTED: PRP4 pre-mRNA processing factor 4 homolog [Pan troglodytes]
148697254 transducin (beta)-like 1 X-linked, isoform CRA_a [Mus musculus]
195012996 GH15383 [Drosophila grimshawi]
125979507 GA19511 [Drosophila pseudoobscura pseudoobscura]
145506354 hypothetical protein GSPATT00038668001 [Paramecium tetraurelia strain d4-2]
19074975 TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 72/90-100 kDa [Encephalitozoon cuniculi GB-M1]
115471661 Os07g0405100 [Oryza sativa (japonica cultivar-group)]
15227373 EMB2776/LIS (LACHESIS); nucleotide binding [Arabidopsis thaliana]
30794128 putative small nuclear ribonucleoprotein Prp4p [Arabidopsis thaliana]
21355245 CG6322 CG6322-PA [Drosophila melanogaster]
195328406 GM24327 [Drosophila sechellia]
195591096 GD12400 [Drosophila simulans]
109461730 PREDICTED: similar to CG3436-PA, isoform A [Rattus norvegicus]
194871924 GG15805 [Drosophila erecta]
195494811 GE22143 [Drosophila yakuba]

194750731 GF23908 [Drosophila ananassae]
156084696 WD-repeat protein, putative [Babesia bovis]
171687265 unnamed protein product [Podospora anserina]
47227128 unnamed protein product [Tetraodon nigroviridis]
118345960 hypothetical protein TTHERM_00700970 [Tetrahymena thermophila SB210]
47226994 unnamed protein product [Tetraodon nigroviridis]
71410152 notchless homolog, putative [Trypanosoma cruzi strain CL Brener]
70991757 WD repeat protein [Aspergillus fumigatus Af293]
109039252 PREDICTED: similar to WD repeat domain 51A [Macaca mulatta]
119195649 hypothetical protein CIMG_02199 [Coccidioides immitis RS]
195376195 GJ13130 [Drosophila virilis]
62088796 transducin beta-like 1X variant [Homo sapiens]
145551644 hypothetical protein GSPATT00026449001 [Paramecium tetraurelia strain d4-2]
195126092 GI12988 [Drosophila mojavensis]
121719733 pre-mRNA splicing factor, putative [Aspergillus clavatus NRRL 1]
149637996 PREDICTED: similar to WD repeat domain 51B [Ornithorhynchus anatinus]
62089180 PRPF4 protein variant [Homo sapiens]
119469599 WD repeat protein [Neosartorya fischeri NRRL 181]
159123728 pre-mRNA splicing factor, putative [Aspergillus fumigatus A1163]
169769711 hypothetical protein [Aspergillus oryzae RIB40]
119484210 pre-mRNA splicing factor, putative [Neosartorya fischeri NRRL 181]
195999252 hypothetical protein TRIADDRAFT_53592 [Trichoplax adhaerens]
67540670 hypothetical protein AN6505.2 [Aspergillus nidulans FGSC A4]
67594391 notchless [Cryptosporidium hominis TU502]
126644100 notchless [Cryptosporidium parvum Iowa II]
149237194 hypothetical protein LELG_04446 [Lodderomyces elongisporus NRRL YB-4239]
114691919 PREDICTED: transducin beta-like 1Y [Pan troglodytes]
121699266 WD repeat protein [Aspergillus clavatus NRRL 1]
159477777 hypothetical protein CHLREDRAFT_119935 [Chlamydomonas reinhardtii]
115491083 conserved hypothetical protein [Aspergillus terreus NIH2624]
145613920 hypothetical protein MGG_01259 [Magnaporthe grisea 70-15]
164427194 hypothetical protein NCU03244 [Neurospora crassa OR74A]
149639953 PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
170049573 will die slowly [Culex quinquefasciatus]
115396860 hypothetical protein ATEG_04891 [Aspergillus terreus NIH2624]
119920654 PREDICTED: similar to transducin (beta)-like 1 X-linked [Bos taurus]
26326543 unnamed protein product [Mus musculus]
145508622 hypothetical protein GSPATT00038818001 [Paramecium tetraurelia strain d4-2]
149638310 PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
39948719 hypothetical protein MGG_08708 [Magnaporthe grisea 70-15]
118099168 PREDICTED: hypothetical protein [Gallus gallus]
158259841 unnamed protein product [Homo sapiens]
164427237 hypothetical protein NCU11251 [Neurospora crassa OR74A]
48116262 PREDICTED: similar to CG6322-PA [Apis mellifera]
62859665 transducin (beta)-like 1 X-linked receptor 1 [Xenopus (Silurana) tropicalis]
115453063 Os03g0355200 [Oryza sativa (japonica cultivar-group)]
126297571 PREDICTED: similar to PRP4 pre-mRNA processing factor 4 homolog (yeast) isoform 2 [Monodelphis domestica]
118084170 PREDICTED: hypothetical protein [Gallus gallus]
126297568 PREDICTED: similar to PRP4 pre-mRNA processing factor 4 homolog (yeast) isoform 1 [Monodelphis domestica]
126331160 PREDICTED: hypothetical protein [Monodelphis domestica]
168044694 predicted protein [Physcomitrella patens subsp. patens]
2708305 U4/U6 small nuclear ribonucleoprotein hPrp4 [Homo sapiens]
15150805 transducin beta-like 1Y [Homo sapiens]
24431950 PRP4 pre-mRNA processing factor 4 homolog [Homo sapiens]
67526937 hypothetical protein AN3926.2 [Aspergillus nidulans FGSC A4]
146088241 notchless homolog [Leishmania infantum JPCM5]
147907316 hypothetical protein LOC779369 [Xenopus laevis]
150866427 hypothetical protein PICST_63551 [Pichia stipitis CBS 6054]
154338545 notchless homolog, putative [Leishmania braziliensis MHOM/BR/75/M2904]
157870387 notchless homolog, putative [Leishmania major]
189053699 unnamed protein product [Homo sapiens]
194225655 PREDICTED: similar to U4/U6 small nuclear ribonucleoprotein Prp4 (U4/U6 snRNP 60 kDa protein) (WD splicing factor Prp4) (hPrp4) (PRP4 homolog) [Equus caballus]
2653736 U4/U6 snRNP 60 kDa protein [Homo sapiens]
26354947 unnamed protein product [Mus musculus]
48146327 PRPF4 [Homo sapiens]
55925589 PRP4 pre-mRNA processing factor 4 homolog [Mus musculus]
67522022 hypothetical protein AN1468.2 [Aspergillus nidulans FGSC A4]
73971978 PREDICTED: similar to PRP4 pre-mRNA processing factor 4 homolog [Canis familiaris]

77735613 PRP4 pre-mRNA processing factor 4 homolog [Bos taurus]
109110534 PREDICTED: PRP4 pre-mRNA processing factor 4 homolog [Macaca mulatta]
109129878 PREDICTED: transducin beta-like 1X isoform 1 [Macaca mulatta]
157820821 PRP4 pre-mRNA processing factor 4 homolog [Rattus norvegicus]
197098526 PRP4 pre-mRNA processing factor 4 homolog [Pongo abelii]
198422937 PREDICTED: similar to transducin (beta)-like 1 X-linked receptor 1 isoform 2 [Ciona intestinalis]
2853277 WD splicing factor Hprp4p [Homo sapiens]
126139399 WD-repeat protein required for cell viability [Pichia stipitis CBS 6054]
187607519 hypothetical protein LOC100145441 [Xenopus tropicalis]
148223275 nuclear receptor co-repressor complex subunit TBLR1 [Xenopus laevis]
156542767 PREDICTED: similar to wd-repeat protein isoform 4 [Nasonia vitripennis]
46107170 hypothetical protein FG00468.1 [Gibberella zeae PH-1]
109129874 PREDICTED: transducin beta-like 1X isoform 2 [Macaca mulatta]
71022699 hypothetical protein UM05432.1 [Ustilago maydis 521]
85115850 hypothetical protein NCU00794 [Neurospora crassa OR74A]
116181204 hypothetical protein CHGG_01230 [Chaetomium globosum CBS 148.51]
119910305 PREDICTED: similar to WD repeat-containing protein 88 (PQQ repeat and WD repeat-containing protein) [Bos taurus]
121703357 WD repeat protein [Aspergillus clavatus NRRL 1]
145351828 predicted protein [Ostreococcus lucimarinus CCE9901]
10383804 WD-repeat protein involved in ribosome biogenesis; may interact with ribosomes; required for maturation and efficient intra-nuclear transport or pre-60S ribosomal subunits, localizes to the nucleolus; Rsa4p [Saccharomyces cerevisiae]
46122641 hypothetical protein FG05698.1 [Gibberella zeae PH-1]
50305243 unnamed protein product [Kluyveromyces lactis]
70990200 ribosome biogenesis protein Rsa4 [Aspergillus fumigatus Af293]
119497387 WD repeat protein [Neosartorya fischeri NRRL 181]
126343427 PREDICTED: similar to nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 [Monodelphis domestica]
156538050 PREDICTED: hypothetical protein [Nasonia vitripennis]
167534993 predicted protein [Monosiga brevicollis MX1]
169766866 hypothetical protein [Aspergillus oryzae RIB40]
169849193 hypothetical protein CC1G_00847 [Coprinopsis cinerea okayama7#130]
170111430 predicted protein [Laccaria bicolor S238N-H82]
186703915 WD repeat-containing protein YCR072C [Zygosaccharomyces rouxii]
189203063 U4/U6 small nuclear ribonucleoprotein Prp4 [Pyrenophora tritici-repentis Pt-1C-BFP]
10434648 unnamed protein product [Homo sapiens]
12006104 AF268193_1 IRA1 [Homo sapiens]
12006108 AF268195_1 IRA1 [Mus musculus]
19913371 transducin (beta)-like 1 X-linked receptor 1 [Homo sapiens]
26331128 unnamed protein product [Mus musculus]
31543001 IRA1 protein [Mus musculus]
50543284 hypothetical protein [Yarrowia lipolytica]
68473930 transcriptional repressor of filamentous growth [Candida albicans SC5314]
74003665 PREDICTED: similar to nuclear receptor co-repressor/HDAC3 complex subunit isoform 2 [Canis familiaris]
115398810 WD-repeat protein 5 [Aspergillus terreus NIH2624]
119883224 PREDICTED: similar to transducin (beta)-like 1X-linked receptor 1 (predicted) [Bos taurus]
157822061 transducin (beta)-like 1X-linked receptor 1 [Rattus norvegicus]
45190866 AER263Cp [Ashbya gossypii ATCC 10895]
145535406 hypothetical protein GSPATT00039746001 [Paramecium tetraurelia strain d4-2]
148724926 novel protein similar to vertebrate transducin (beta)-like 1X-linked receptor 1 (TBL1XR1, zgc:110312) [Danio rerio]
156390600 predicted protein [Nematostella vectensis]
156542769 PREDICTED: similar to wd-repeat protein isoform 3 [Nasonia vitripennis]
157136853 wd-repeat protein [Aedes aegypti]
48097023 PREDICTED: similar to IRA1 protein isoform 1 [Apis mellifera]
125556286 hypothetical protein OsI_023124 [Oryza sativa (indica cultivar-group)]
125598048 hypothetical protein OsJ_021311 [Oryza sativa (japonica cultivar-group)]
145506384 hypothetical protein GSPATT00038682001 [Paramecium tetraurelia strain d4-2]
156847482 hypothetical protein Kpol_1028p41 [Vanderwaltozyma polyspora DSM 70294]
169772829 hypothetical protein [Aspergillus oryzae RIB40]
198422965 PREDICTED: similar to PRP4 pre-mRNA processing factor 4 homolog [Ciona intestinalis]
58266782 RNA processing-related protein [Cryptococcus neoformans var. neoformans JEC21]
156390324 predicted protein [Nematostella vectensis]
171695704 unnamed protein product [Podospora anserina]
189234886 PREDICTED: similar to wd-repeat protein [Tribolium castaneum]
47215488 unnamed protein product [Tetraodon nigroviridis]
67971632 unnamed protein product [Macaca fascicularis]
131888158 transducin (beta)-like 1X-linked [Danio rerio]
148726563 novel protein (zgc:63617) [Danio rerio]
156542765 PREDICTED: similar to wd-repeat protein isoform 1 [Nasonia vitripennis]
195996461 hypothetical protein TRIADDRAFT_18846 [Trichoplax adhaerens]
20302751 AF391290_3 unknown [Branchiostoma floridae]

156057299 hypothetical protein SS1G_04380 [Sclerotinia sclerotiorum 1980]
157124855 wd-repeat protein [Aedes aegypti]
41054303 PRP4 pre-mRNA processing factor 4 homolog [Danio rerio]
45360769 PRP4 pre-mRNA processing factor 4 homolog [Xenopus tropicalis]
193666990 PREDICTED: similar to wd-repeat protein [Acyrthosiphon pisum]
47216991 unnamed protein product [Tetraodon nigroviridis]
109043586 PREDICTED: similar to nuclear receptor co-repressor/HDAC3 complex subunit [Macaca mulatta]
170052593 U4/U6 small nuclear ribonucleoprotein Prp4 [Culex quinquefasciatus]
119193472 hypothetical protein CIMG_01113 [Coccidioides immitis RS]
158285020 AGAP002144-PA [Anopheles gambiae str. PEST]
158297568 AGAP007739-PA [Anopheles gambiae str. PEST]
19075884 notchless-like protein [Schizosaccharomyces pombe]
72110055 PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]
72001544 ZC302.2a [Caenorhabditis elegans]
109113986 PREDICTED: Notchless gene homolog isoform 1 [Macaca mulatta]
193592097 PREDICTED: similar to AGAP007739-PA [Acyrthosiphon pisum]
50288717 unnamed protein product [Candida glabrata]
156542771 PREDICTED: similar to wd-repeat protein isoform 2 [Nasonia vitripennis]
169602957 hypothetical protein SNOG_04483 [Phaeosphaeria nodorum SN15]
125577707 hypothetical protein OsJ_033138 [Oryza sativa (japonica cultivar-group)]
156383749 predicted protein [Nematostella vectensis]
169642556 LOC398608 protein [Xenopus laevis]
125547300 hypothetical protein OsI_014355 [Oryza sativa (indica cultivar-group)]
170590149 U4/U6 small nuclear ribonucleoprotein hPrp4 [Brugia malayi]
145353453 predicted protein [Ostreococcus lucimarinus CCE9901]
193627197 PREDICTED: similar to IRA1 protein [Acyrthosiphon pisum]
118387616 WD-repeat protein HUSSY-07, putative [Tetrahymena thermophila SB210]
115695227 PREDICTED: similar to nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 [Strongylocentrotus purpuratus]
195437532 GK24428 [Drosophila willistoni]
116181536 hypothetical protein CHGG_01396 [Chaetomium globosum CBS 148.51]
123470660 transducin, putative [Trichomonas vaginalis G3]
168010281 predicted protein [Physcomitrella patens subsp. patens]
114668129 PREDICTED: Notchless gene homolog isoform 1 [Pan troglodytes]
145253246 hypothetical protein An16g08380 [Aspergillus niger]
189235571 PREDICTED: similar to wd-repeat protein [Tribolium castaneum]
37682095 TUWD12 [Danio rerio]
41056233 TUWD12 [Danio rerio]
123232887 WD repeat domain 51B [Danio rerio]
145512722 hypothetical protein GSPATT00039041001 [Paramecium tetraurelia strain d4-2]
194758597 GF14879 [Drosophila ananassae]
45549090 Notchless CG2863-PA [Drosophila melanogaster]
119112932 AGAP002221-PA [Anopheles gambiae str. PEST]
194853592 GG24639 [Drosophila erecta]
195350141 GM16656 [Drosophila sechellia]
195388370 GJ19693 [Drosophila virilis]
56753746 SJCHGC04552 protein [Schistosoma japonicum]
62087802 Notchless gene homolog variant [Homo sapiens]
118150558 hypothetical protein LOC777721 [Danio rerio]
119600573 notchless homolog 1 (Drosophila), isoform CRA_c [Homo sapiens]
189011602 notchless homolog 1 [Rattus norvegicus]
198433744 PREDICTED: similar to notchless homolog 1 [Ciona intestinalis]
62460466 Notchless gene homolog [Bos taurus]
21703860 Notchless gene homolog [Mus musculus]
34922640 NLE1_HUMAN Notchless protein homolog 1
56205885 notchless homolog 1 (Drosophila) [Mus musculus]
62234461 Notchless gene homolog isoform a [Homo sapiens]
74189962 unnamed protein product [Mus musculus]
74221147 unnamed protein product [Mus musculus]
109113988 PREDICTED: Notchless gene homolog isoform 2 [Macaca mulatta]
119600572 notchless homolog 1 (Drosophila), isoform CRA_b [Homo sapiens]
150951674 predicted protein [Pichia stipitis CBS 6054]
170594191 Guanine nucleotide-binding protein beta subunit 2 [Brugia malayi]
194215266 PREDICTED: similar to WD repeat-containing protein 88 (PQQ repeat and WD repeat-containing protein) [Equus caballus]
195032822 GH10505 [Drosophila grimshawi]
195118224 GI18023 [Drosophila mojavensis]
197102232 notchless homolog 1 [Pongo abelii]
149724681 PREDICTED: similar to Notchless protein homolog 1 [Equus caballus]
158508590 WD repeat, SAM and U-box domain containing 1 [Gallus gallus]
196010591 hypothetical protein TRIADDRAFT_28728 [Trichoplax adhaerens]

73948532 PREDICTED: similar to CG17437-PA [Canis familiaris]
119600571 notchless homolog 1 (Drosophila), isoform CRA_a [Homo sapiens]
195575583 GD22949 [Drosophila simulans]
40557601 notchless-like protein [Solanum chacoense]
118100271 PREDICTED: hypothetical protein [Gallus gallus]
145512725 hypothetical protein GSPATT00039042001 [Paramecium tetraurelia strain d4-2]
170579174 WD-repeat protein HUSSY-07 [Brugia malayi]
195147194 GL19252 [Drosophila persimilis]
195470300 GE16018 [Drosophila yakuba]
4127781 Notchless protein [Drosophila melanogaster]
115480854 Os10g0104500 [Oryza sativa (japonica cultivar-group)]
115486121 Os11g0594200 [Oryza sativa (japonica cultivar-group)]
118398681 hypothetical protein TTHERM_00760520 [Tetrahymena thermophila SB210]
126339618 PREDICTED: hypothetical protein [Monodelphis domestica]
30048137 Wdsub1 protein [Mus musculus]
145491297 hypothetical protein GSPATT00006103001 [Paramecium tetraurelia strain d4-2]
145522492 hypothetical protein GSPATT00014623001 [Paramecium tetraurelia strain d4-2]
26665869 WD repeat domain 51B [Homo sapiens]
114646150 PREDICTED: WD repeat domain 51B isoform 3 [Pan troglodytes]
149642967 WD repeat domain 51B [Bos taurus]
193657087 PREDICTED: similar to notchless homolog 1 [Acyrthosiphon pisum]
123425974 WD repeat protein, putative [Trichomonas vaginalis G3]
149067084 rCG49027 [Rattus norvegicus]
157350695 unnamed protein product [Vitis vinifera]
157753920 Hypothetical protein CBG07064 [Caenorhabditis briggsae AF16]
21752385 unnamed protein product [Homo sapiens]
27882062 Nle-pending-prov protein [Xenopus laevis]
47059149 WD repeat domain 51B [Mus musculus]
62079091 WD repeat, SAM and U-box domain containing 1 [Rattus norvegicus]
66472582 notchless homolog 1 [Danio rerio]
109099816 PREDICTED: similar to WD repeat, SAM and U-box domain containing 1 [Macaca mulatta]
118405102 notchless homolog 1 [Xenopus tropicalis]
126253825 WSDU1_CHICK WD repeat, SAM and U-box domain-containing protein 1
126326215 PREDICTED: similar to WD repeat, sterile alpha motif and U-box domain containing 1, isoform 1 [Monodelphis domestica]
145356501 predicted protein [Ostreococcus lucimarinus CCE9901]
148232415 Notchless gene homolog [Xenopus laevis]
170047140 WD repeat protein [Culex quinquefasciatus]
189571681 WD repeat, sterile alpha motif and U-box domain containing 1 [Homo sapiens]
149639594 PREDICTED: similar to WD repeat, sterile alpha motif and U-box domain containing 1 [Ornithorhynchus anatinus]
149730639 PREDICTED: WD repeat, sterile alpha motif and U-box domain containing 1 [Equus caballus]
156552151 PREDICTED: similar to notchless homolog 1 (Drosophila) [Nasonia vitripennis]
168009638 predicted protein [Physcomitrella patens subsp. patens]
58037281 WD repeat, SAM and U-box domain containing 1 [Mus musculus]
126326217 PREDICTED: similar to WD repeat, sterile alpha motif and U-box domain containing 1, isoform 2 [Monodelphis domestica]
126338290 PREDICTED: hypothetical protein [Monodelphis domestica]
198432947 PREDICTED: similar to predicted protein [Ciona intestinalis]
15237273 WD-40 repeat family protein / notchless protein, putative [Arabidopsis thaliana]
17536759 W07E6.2 [Caenorhabditis elegans]
71897347 transducin (beta)-like 1 X-linked receptor 1 [Gallus gallus]
74758690 WDR88_HUMAN WD repeat-containing protein 88 (PQQ repeat and WD repeat-containing protein)
102470001 PQQ repeat and WD repeat domain containing [Homo sapiens]
114578944 PREDICTED: hypothetical protein [Pan troglodytes]
114676524 PREDICTED: hypothetical protein [Pan troglodytes]
194902576 GG17291 [Drosophila erecta]
71028554 hypothetical protein TP04_0285 [Theileria parva strain Muguga]
91094533 PREDICTED: similar to notchless homolog 1 [Tribolium castaneum]
125587509 hypothetical protein OsJ_011656 [Oryza sativa (japonica cultivar-group)]
156374161 predicted protein [Nematostella vectensis]
57525363 WD repeat domain 51B [Gallus gallus]
73966983 PREDICTED: similar to Notchless homolog 1 (WD-repeat protein HUSSY-07) [Canis familiaris]
148230284 hypothetical protein LOC734519 [Xenopus laevis]
121700168 nuclear migration protein NudF [Aspergillus clavatus NRRL 1]
147841127 hypothetical protein [Vitis vinifera]
194906265 GG12011 [Drosophila erecta]
195503263 GE10447 [Drosophila yakuba]
123975934 transducin, putative [Trichomonas vaginalis G3]
194226643 PREDICTED: similar to WD repeat domain 51B [Equus caballus]
46121319 hypothetical protein FG05038.1 [Gibberella zeae PH-1]
156355398 predicted protein [Nematostella vectensis]

67478341 WD domain containing protein [Entamoeba histolytica HM-1:IMSS]
167388307 hypothetical protein, conserved [Entamoeba dispar SAW760]
119496197 WD repeat protein [Neosartorya fischeri NRRL 181]
145519343 hypothetical protein GSPATT00013588001 [Paramecium tetraurelia strain d4-2]
145528177 hypothetical protein GSPATT00016878001 [Paramecium tetraurelia strain d4-2]
194381934 unnamed protein product [Homo sapiens]
70994968 WD repeat protein [Aspergillus fumigatus Af293]
84996467 hypothetical protein, conserved [Theileria annulata]
119194625 hypothetical protein CIMG_01687 [Coccidioides immitis RS]
50754099 PREDICTED: similar to WD repeat domain 51A [Gallus gallus]
45190337 AEL269Cp [Ashbya gossypii ATCC 10895]
71021149 hypothetical protein UM04658.1 [Ustilago maydis 521]
109892806 NUDF_ASPOR Nuclear distribution protein nudF homolog
154420400 WD repeat protein, putative [Trichomonas vaginalis G3]
71018161 hypothetical protein UM03164.1 [Ustilago maydis 521]
121702197 WD repeat protein [Aspergillus clavatus NRRL 1]
195445078 GK11178 [Drosophila willistoni]
50554175 hypothetical protein [Yarrowia lipolytica]
114668131 PREDICTED: Notchless gene homolog isoform 2 [Pan troglodytes]
145207304 WDR88 protein [Homo sapiens]
164656851 hypothetical protein MGL_3096 [Malassezia globosa CBS 7966]
119631816 WD repeat, sterile alpha motif and U-box domain containing 1, isoform CRA_c [Homo sapiens]
197098256 WD repeat domain 51B [Pongo abelii]
46106236 hypothetical protein FG00414.1 [Gibberella zeae PH-1]
148878210 Wdr51b protein [Mus musculus]
198449874 GA20446 [Drosophila pseudoobscura pseudoobscura]
47224493 unnamed protein product [Tetraodon nigroviridis]
194765270 GF22879 [Drosophila ananassae]
18542932 Putative notchless protein homolog [Oryza sativa (japonica cultivar-group)]
161078943 CG7568 CG7568-PB [Drosophila melanogaster]
71024559 hypothetical protein UM06362.1 [Ustilago maydis 521]
115491733 nuclear migration protein nudF [Aspergillus terreus NIH2624]
195036342 GH18693 [Drosophila grimshawi]
195341223 GM12236 [Drosophila sechellia]
195574771 GD17720 [Drosophila simulans]
195109540 GI24459 [Drosophila mojavensis]
67540054 NUDF_EMENI Nuclear migration protein nudF [Aspergillus nidulans FGSC A4]
194385880 unnamed protein product [Homo sapiens]
145522917 hypothetical protein GSPATT00039354001 [Paramecium tetraurelia strain d4-2]
62860034 WD repeat domain 51B [Xenopus (Silurana) tropicalis]
78707608 Notchless, putative, expressed [Oryza sativa (japonica cultivar-group)]
133778015 WDR51A protein [Homo sapiens]
198450475 GA17766 [Drosophila pseudoobscura pseudoobscura]
194664076 PREDICTED: similar to WD repeat domain 5B [Bos taurus]
119887667 PREDICTED: similar to WD repeat, SAM and U-box domain-containing protein 1 [Bos taurus]
157104339 wd-repeat protein [Aedes aegypti]
169625214 hypothetical protein SNOG_15874 [Phaeosphaeria nodorum SN15]
169768132 hypothetical protein [Aspergillus oryzae RIB40]
50545651 hypothetical protein [Yarrowia lipolytica]
159484634 predicted protein [Chlamydomonas reinhardtii]
169773515 hypothetical protein [Aspergillus oryzae RIB40]
67540430 hypothetical protein AN6385.2 [Aspergillus nidulans FGSC A4]
123508237 WD repeat protein, putative [Trichomonas vaginalis G3]
134111008 hypothetical protein CNBD3750 [Cryptococcus neoformans var. neoformans B-3501A]
38322722 DKFZP434C245-like protein [Takifugu rubripes]
189195740 cytosolic iron-sulfur protein assembly protein 1 [Pyrenophora tritici-repentis Pt-1C-BFP]
195394423 GJ10548 [Drosophila virilis]
156396456 predicted protein [Nematostella vectensis]
169621913 hypothetical protein SNOG_14169 [Phaeosphaeria nodorum SN15]
123397965 WD repeat protein, putative [Trichomonas vaginalis G3]
123447904 WD repeat protein, putative [Trichomonas vaginalis G3]
109113697 PREDICTED: similar to WD repeat and SOCS box-containing 1 isoform 1 isoform 3 [Macaca mulatta]
114668577 PREDICTED: hypothetical protein isoform 3 [Pan troglodytes]
149640263 PREDICTED: similar to nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 [Ornithorhynchus anatinus]
169859606 hypothetical protein CC1G_07088 [Coprinopsis cinerea okayama7#130]
71408172 hypothetical protein Tc00.1047053503905.40 [Trypanosoma cruzi strain CL Brener]
74136113 WSB1 protein [Takifugu rubripes]
71013096 hypothetical protein UM02408.1 [Ustilago maydis 521]
71403276 hypothetical protein Tc00.1047053507705.20 [Trypanosoma cruzi strain CL Brener]

146182822 hypothetical protein TTHERM_00762930 [Tetrahymena thermophila SB210]
164660198 hypothetical protein MGL_1405 [Malassezia globosa CBS 7966]
72392833 hypothetical protein, conserved [Trypanosoma brucei TREU927]
145489191 hypothetical protein GSPATT00033053001 [Paramecium tetraurelia strain d4-2]
145608398 hypothetical protein MGG_03150 [Magnaporthe grisea 70-15]
41054505 WD repeat and SOCS box-containing 1 [Danio rerio]
118095051 PREDICTED: hypothetical protein [Gallus gallus]
145476749 hypothetical protein GSPATT00028032001 [Paramecium tetraurelia strain d4-2]
145521356 hypothetical protein GSPATT00014103001 [Paramecium tetraurelia strain d4-2]
45361625 WD repeat and SOCS box-containing 1 [Xenopus tropicalis]
5410332 WSB-1 [Homo sapiens]
6563198 AF112205_1 WSB-1 protein [Homo sapiens]
18677720 WD repeat and SOCS box-containing 1 isoform 1 [Homo sapiens]
22760676 unnamed protein product [Homo sapiens]
31419799 WD repeat and SOCS box-containing 1 [Mus musculus]
45382387 WD repeat and SOCS box-containing 1 [Gallus gallus]
62896711 WD SOCS-box protein 1 isoform 1 variant [Homo sapiens]
73966919 PREDICTED: similar to WD SOCS-box protein 1 isoform 1 isoform 2 [Canis familiaris]
74204384 unnamed protein product [Mus musculus]
74220242 unnamed protein product [Mus musculus]
83816974 WD repeat and SOCS box-containing 1 isoform 1 [Mus musculus]
109113695 PREDICTED: similar to WD repeat and SOCS box-containing 1 isoform 1 isoform 6 [Macaca mulatta]
110349999 WD repeat and SOCS box-containing 1 isoform 1 [Rattus norvegicus]
114647205 PREDICTED: WD SOCS-box protein 2 isoform 3 [Pan troglodytes]
114668573 PREDICTED: hypothetical protein isoform 4 [Pan troglodytes]
119911554 PREDICTED: similar to WD repeat and SOCS box-containing 1 isoform 10 [Bos taurus]
149724074 PREDICTED: WD repeat and SOCS box-containing 1 [Equus caballus]
194380610 unnamed protein product [Homo sapiens]
156034665 hypothetical protein SS1G_13267 [Sclerotinia sclerotiorum 1980]
156554779 PREDICTED: hypothetical protein [Nasonia vitripennis]
198454171 GA26333 [Drosophila pseudoobscura pseudoobscura]
66805987 putative dynein regulator [Dictyostelium discoideum AX4]
71410872 hypothetical protein Tc00.1047053509563.30 [Trypanosoma cruzi strain CL Brener]
71652564 hypothetical protein Tc00.1047053503897.40 [Trypanosoma cruzi strain CL Brener]
72392903 hypothetical protein, conserved [Trypanosoma brucei TREU927]
145529465 hypothetical protein GSPATT00039533001 [Paramecium tetraurelia strain d4-2]
146079233 hypothetical protein, conserved [Leishmania infantum JPCM5]
154333191 hypothetical protein LbrM10_V2.0880 [Leishmania braziliensis MHOM/BR/75/M2904]
157865409 hypothetical protein, conserved [Leishmania major]
47221500 unnamed protein product [Tetraodon nigroviridis]
113195628 WD repeat-containing protein 69 [Danio rerio]
145496786 hypothetical protein GSPATT00036117001 [Paramecium tetraurelia strain d4-2]
149720611 PREDICTED: similar to WD SOCS-box protein 2 [Equus caballus]
145527034 hypothetical protein GSPATT00039485001 [Paramecium tetraurelia strain d4-2]
149632946 PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
47085759 WD repeat domain 51A [Danio rerio]
66472584 hypothetical protein LOC553608 [Danio rerio]
110759973 PREDICTED: similar to CG7568-PA [Apis mellifera]
119618523 WD repeat and SOCS box-containing 2, isoform CRA_a [Homo sapiens]
21758953 unnamed protein product [Homo sapiens]
32189425 WD repeat domain 69 [Homo sapiens]
45361545 WD repeat-containing protein 69 [Xenopus tropicalis]
68163493 hypothetical protein LOC363267 [Rattus norvegicus]
73967256 PREDICTED: similar to platelet-activating factor acetylhydrolase beta subunit isoform 5 [Canis familiaris]
73994108 PREDICTED: similar to CG7568-PA [Canis familiaris]
75077029 WDR69_MACFA WD repeat-containing protein 69
114583737 PREDICTED: hypothetical protein [Pan troglodytes]
115495385 WD repeat domain 69 [Bos taurus]
116284318 Si:dkey-223n17.5 protein [Danio rerio]
125598075 hypothetical protein OsJ_021338 [Oryza sativa (japonica cultivar-group)]
147904210 hypothetical protein LOC734280 [Xenopus laevis]
152031728 WDR69_DANRE WD repeat-containing protein 69
156366072 predicted protein [Nematostella vectensis]
195996275 hypothetical protein TRIADDRAFT_49686 [Trichoplax adhaerens]
198428088 PREDICTED: similar to WD repeat-containing protein 69 [Ciona intestinalis]
159465765 predicted protein [Chlamydomonas reinhardtii]
169852680 hypothetical protein CC1G_01084 [Coprinopsis cinerea okayama7#130]
195440620 GK10412 [Drosophila willistoni]
167536841 predicted protein [Monosiga brevicollis MX1]

195152906 GL21559 [Drosophila persimilis]
195173067 GL15696 [Drosophila persimilis]
195176099 GL25364 [Drosophila persimilis]
195176101 GL25365 [Drosophila persimilis]
198465358 GA10144 [Drosophila pseudoobscura pseudoobscura]
26327737 unnamed protein product [Mus musculus]
114668133 PREDICTED: Notchless gene homolog isoform 3 [Pan troglodytes]
145239027 hypothetical protein An08g00420 [Aspergillus niger]
168012126 predicted protein [Physcomitrella patens subsp. patens]
17137196 Lissencephaly-1 CG8440-PA, isoform A [Drosophila melanogaster]
114668575 PREDICTED: WD repeat and SOCS box-containing 1 isoform 2 [Pan troglodytes]
125545296 hypothetical protein OsI_012668 [Oryza sativa (indica cultivar-group)]
125807773 GA21079 [Drosophila pseudoobscura pseudoobscura]
149063503 WD repeat and SOCS box-containing 2, isoform CRA_a [Rattus norvegicus]
157110225 platelet-activating factor acetylhydrolase isoform 1b alpha subunit [Aedes aegypti]
189240395 PREDICTED: similar to Lissencephaly-1 CG8440-PA [Tribolium castaneum]
193645819 PREDICTED: similar to GA21079-PA [Acyrthosiphon pisum]
194755010 GF11864 [Drosophila ananassae]
194882617 GG20577 [Drosophila erecta]
195056834 GH22777 [Drosophila grimshawi]
195122266 GI20573 [Drosophila mojavensis]
195380804 GJ21422 [Drosophila virilis]
198431299 PREDICTED: similar to Platelet-activating factor acetylhydrolase IB subunit alpha (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-1) [Ciona intestinalis]
4557741 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit [Homo sapiens]
7305363 platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit [Mus musculus]
17056921 Miller-Dieker lissencephaly protein [Homo sapiens]
17056923 Miller-Dieker lissencephaly protein [Homo sapiens]
27807199 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit [Bos taurus]
41152229 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit a [Danio rerio]
41152231 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit b [Danio rerio]
45383504 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa [Gallus gallus]
45433584 platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit [Xenopus (Silurana) tropicalis]
47212444 unnamed protein product [Tetraodon nigroviridis]
47523580 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit [Sus scrofa]
67463777 C Chain C, Paf-Ah Holoenzyme: Lis1ALFA2
74184724 unnamed protein product [Mus musculus]
74190682 unnamed protein product [Mus musculus]
74195430 unnamed protein product [Mus musculus]
74204167 unnamed protein product [Mus musculus]
75914613 Lis1a [Danio rerio]
77539766 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit [Pan troglodytes]
90081346 unnamed protein product [Macaca fascicularis]
109112726 PREDICTED: similar to platelet-activating factor acetylhydrolase beta subunit [Macaca mulatta]
148228177 LIS1 [Xenopus laevis]
156544931 PREDICTED: similar to platelet-activating factor acetylhydrolase isoform 1b alpha subunit [Nasonia vitripennis]
167533957 predicted protein [Monosiga brevicollis MX1]
197099568 platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa [Pongo abelii]
449693 1919424A Miller-Dieker lissencephaly gene
38322769 DKFZP434C245-like protein [Tetraodon nigroviridis]
148222751 WD SOCS-box protein 2 [Xenopus laevis]
156386518 predicted protein [Nematostella vectensis]
187607147 hypothetical protein LOC100145242 [Xenopus tropicalis]
194385094 unnamed protein product [Homo sapiens]
195436222 GK22131 [Drosophila willistoni]
196010635 hypothetical protein TRIADDRAFT_50647 [Trichoplax adhaerens]
152926156 transducin family protein [Carica papaya]
30698820 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
31377770 WD repeat domain 51A [Homo sapiens]
84370141 WD repeat domain 51A [Bos taurus]
91207986 WD51A_HUMAN WD repeat-containing protein 51A
115903823 PREDICTED: similar to platelet-activating factor acetylhydrolase 45K chain - bovine [Strongylocentrotus purpuratus]
194042982 PREDICTED: similar to WD SOCS-box protein 2 [Sus scrofa]
66563906 PREDICTED: similar to TUWD12 [Apis mellifera]
73985357 PREDICTED: similar to WD repeat domain 51A isoform 1 [Canis familiaris]
73994651 PREDICTED: similar to WD repeat and SOCS box containing protein 2 (WSB-2) (CS box-containing WD protein) isoform 1 [Canis familiaris]
114647207 PREDICTED: WD SOCS-box protein 2 isoform 1 [Pan troglodytes]
115496374 WD SOCS-box protein 2 [Bos taurus]

126314249 PREDICTED: similar to platelet-activating factor acetylhydrolase 45K chain - bovine [Monodelphis domestica]
151556994 WD repeat and SOCS box-containing 2 [Bos taurus]
58037201 WD repeat domain 51A [Mus musculus]
91207987 WD51A_MOUSE WD repeat-containing protein 51A
195377710 GJ13548 [Drosophila virilis]
8886487 AF163324_1 WSB-2 [Homo sapiens]
8923881 WD SOCS-box protein 2 [Homo sapiens]
17554220 LIS-1 (human lissencephaly gene) related family member (lis-1) [Caenorhabditis elegans]
20532278 WSB2_MOUSE WD repeat and SOCS box-containing protein 2 (WSB-2) (SOCS box-containing WD protein SWiP-2)
31982686 WSB-2 protein [Mus musculus]
56090230 WD repeat and SOCS box-containing 2 [Rattus norvegicus]
61098210 guanine nucleotide binding protein (G protein), beta polypeptide 1 [Gallus gallus]
114647209 PREDICTED: WD SOCS-box protein 2 isoform 2 [Pan troglodytes]
149063504 WD repeat and SOCS box-containing 2, isoform CRA_b [Rattus norvegicus]
157776510 hypothetical protein CBG21225 [Caenorhabditis briggsae AF16]
195018998 GH14795 [Drosophila grimshawi]
145503147 hypothetical protein GSPATT00007594001 [Paramecium tetraurelia strain d4-2]
145523600 hypothetical protein GSPATT00015130001 [Paramecium tetraurelia strain d4-2]
194750027 GF10412 [Drosophila ananassae]
167518311 predicted protein [Monosiga brevicollis MX1]
195127848 GI11850 [Drosophila mojavensis]
170030259 platelet-activating factor acetylhydrolase 1b alpha subunit [Culex quinquefasciatus]
66825073 hypothetical protein DDBDRAFT_0190190 [Dictyostelium discoideum AX4]
194377876 unnamed protein product [Homo sapiens]
47229875 unnamed protein product [Tetraodon nigroviridis]
148226456 hypothetical protein LOC379573 [Xenopus laevis]
189196917 cell cycle control protein cwf17 [Pyrenophora tritici-repentis Pt-1C-BFP]
74000171 PREDICTED: similar to Guanine nucleotide-binding protein beta subunit 5 (Transducin beta chain 5) (Gbeta5) isoform 3 [Canis familiaris]
109081165 PREDICTED: similar to guanine nucleotide-binding protein, beta-5 subunit isoform b isoform 1 [Macaca mulatta]
116060356 FOG: RCC1 domain (ISS) [Ostreococcus tauri]
118793080 AGAP011849-PA [Anopheles gambiae str. PEST]
194701794 unknown [Zea mays]
6754018 guanine nucleotide-binding protein, beta-5 subunit isoform 1 [Mus musculus]
17568701 K04G11.4 [Caenorhabditis elegans]
20336270 guanine nucleotide-binding protein, beta-5 subunit isoform b [Homo sapiens]
41152187 guanine nucleotide binding protein (G protein), beta 5 [Danio rerio]
74000173 PREDICTED: similar to guanine nucleotide-binding protein, beta-5 subunit isoform b isoform 1 [Canis familiaris]
74418608 G protein beta 5 subunit [Uta stansburiana]
109048741 PREDICTED: WD SOCS-box protein 2 [Macaca mulatta]
146078992 WD-repeat containing protein [Leishmania infantum JPCM5]
148524804 G-protein beta subunit 5L [Cyprinus carpio]
149019162 guanine nucleotide binding protein, beta 5, isoform CRA_b [Rattus norvegicus]
157865263 WD-repeat containing protein; hypothetical protein, conserved [Leishmania major]
50426611 hypothetical protein DEHA0G08789g [Debaryomyces hansenii CBS767]
115454653 Os03g0685600 [Oryza sativa (japonica cultivar-group)]
9955107 A Chain A, Crystal Structure Of The C-Terminal Wd40 Domain Of Tup1
159483209 centriole proteome protein [Chlamydomonas reinhardtii]
115812851 PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]
2104937 truncated form platelet-activating factor acetylhydrolase 45kD subunit [Mus musculus]
24663767 CG10191 CG10191-PA [Drosophila melanogaster]
147771302 hypothetical protein [Vitis vinifera]
157357657 unnamed protein product [Vitis vinifera]
169775423 hypothetical protein [Aspergillus oryzae RIB40]
195494095 GE20071 [Drosophila yakuba]
149415092 PREDICTED: similar to PRP4 pre-mRNA processing factor 4 homolog (yeast), partial [Ornithorhynchus anatinus]
157335595 unnamed protein product [Vitis vinifera]
194870404 GG13777 [Drosophila erecta]
195158821 GL13584 [Drosophila persimilis]
157109775 wd-repeat protein [Aedes aegypti]
170031167 wd-repeat protein [Culex quinquefasciatus]
195327372 GM24599 [Drosophila sechellia]
195590058 GD12669 [Drosophila simulans]
57529443 WD repeat domain 57 (U5 snRNP specific) [Gallus gallus]
73985355 PREDICTED: similar to WD repeat domain 51A isoform 2 [Canis familiaris]
123426657 WD repeat protein, putative [Trichomonas vaginalis G3]
157353777 unnamed protein product [Vitis vinifera]
170577867 platelet-activating factor acetylhydrolase IB alpha subunit, putative [Brugia malayi]
146417975 hypothetical protein PGUG_02683 [Pichia guilliermondii ATCC 6260]

190346487 hypothetical protein PGUG_02683 [Pichia guilliermondii ATCC 6260]
193654883 PREDICTED: similar to AGAP007739-PA, partial [Acyrthosiphon pisum]
157134516 wd-repeat protein [Aedes aegypti]
170060657 wd-repeat protein [Culex quinquefasciatus]
170071614 wd-repeat protein [Culex quinquefasciatus]
196015990 hypothetical protein TRIADDRAFT_51125 [Trichoplax adhaerens]
19112019 WD repeat protein Swd3 [Schizosaccharomyces pombe]
50428732 putative WD G-beta repeat protein [Oryza sativa (japonica cultivar-group)]
119591287 WD repeat domain 69, isoform CRA_b [Homo sapiens]
145248483 hypothetical protein An13g03550 [Aspergillus niger]
50421955 hypothetical protein DEHA0E05775g [Debaryomyces hansenii CBS767]
52345880 MGC89488 protein [Xenopus tropicalis]
199432541 DEHA2E05016p [Debaryomyces hansenii]
47222534 unnamed protein product [Tetraodon nigroviridis]
158300631 AGAP012029-PA [Anopheles gambiae str. PEST]
3023832 GBB1_TOBAC Guanine nucleotide-binding protein subunit beta-1
3023839 GBB2_TOBAC Guanine nucleotide-binding protein subunit beta-2
3023841 GBB_NICPL Guanine nucleotide-binding protein subunit beta (NpGPB1)
3023842 GBB_SOLTU Guanine nucleotide-binding protein subunit beta
10048265 AF249743_1 heterotrimeric GTP-binding protein subunit beta 1 [Nicotiana tabacum]
56755769 SJCHGC06010 protein [Schistosoma japonicum]
193599074 PREDICTED: similar to guanine nucleotide-binding protein subunit beta 5 [Acyrthosiphon pisum]
17552164 Temporarily Assigned Gene name family member (tag-125) [Caenorhabditis elegans]
70986635 vegetative incompatibility WD repeat protein [Aspergillus fumigatus Af293]
167519935 predicted protein [Monosiga brevicollis MX1]
198424921 PREDICTED: similar to guanine nucleotide-binding protein, beta-5 subunit [Ciona intestinalis]
3023859 GBB3_TOBAC Guanine nucleotide-binding protein subunit beta
148689181 WD repeat domain 51A, isoform CRA_a [Mus musculus]
58269196 U5 snRNP-specific 40 kDa protein [Cryptococcus neoformans var. neoformans JEC21]
134114449 hypothetical protein CNBG4530 [Cryptococcus neoformans var. neoformans B-3501A]
157345227 unnamed protein product [Vitis vinifera]
156402335 predicted protein [Nematostella vectensis]
195121738 GI19111 [Drosophila mojavensis]
195383766 GJ22241 [Drosophila virilis]
7496291 T15181 hypothetical protein C18E3.5 - Caenorhabditis elegans
4902500 G protein, beta subunit [Geodia cydonium]
46111533 hypothetical protein FG02648.1 [Gibberella zeae PH-1]
118377979 hypothetical protein TTHERM_00787270 [Tetrahymena thermophila SB210]
169776724 hypothetical protein [Aspergillus oryzae RIB40]
169864946 hypothetical protein CC1G_10643 [Coprinopsis cinerea okayama7#130]
198415836 PREDICTED: hypothetical protein [Ciona intestinalis]
7239511 00400. [Arabidopsis thaliana]
12832304 unnamed protein product [Mus musculus]
156087667 WD domain, G-beta repeat containing protein [Babesia bovis]
159475928 transcriptional repression protein [Chlamydomonas reinhardtii]
167517451 predicted protein [Monosiga brevicollis MX1]
116789145 unknown [Picea sitchensis]
145529271 hypothetical protein GSPATT00017384001 [Paramecium tetraurelia strain d4-2]
145534540 hypothetical protein GSPATT00019334001 [Paramecium tetraurelia strain d4-2]
149238431 hypothetical protein LELG_04124 [Lodderomyces elongisporus NRRL YB-4239]
157775051 hypothetical protein CBG09206 [Caenorhabditis briggsae AF16]
167535706 predicted protein [Monosiga brevicollis MX1]
187032248 Hypothetical protein CBG08899 [Caenorhabditis briggsae]
56755415 SJCHGC06272 protein [Schistosoma japonicum]
119597846 guanine nucleotide binding protein (G protein), beta 5, isoform CRA_e [Homo sapiens]
145245727 hypothetical protein An12g00370 [Aspergillus niger]
150865758 WD domain protein [Pichia stipitis CBS 6054]
157817821 WD repeat domain 51A [Rattus norvegicus]
198418235 PREDICTED: similar to Transcription initiation factor TFIID subunit 5 (Transcription initiation factor TFIID 100 kDa subunit) (TAF(II)100) (TAFII-100) (TAFII100), partial [Ciona intestinalis]
126272320 PREDICTED: hypothetical protein [Monodelphis domestica]
145345045 predicted protein [Ostreococcus lucimarinus CCE9901]
193606315 PREDICTED: similar to AGAP005359-PA [Acyrthosiphon pisum]
194389928 unnamed protein product [Homo sapiens]
195181472 GL15634 [Drosophila persimilis]
164657574 hypothetical protein MGL_2899 [Malassezia globosa CBS 7966]
6714707 hypothetical protein [Homo sapiens]
115750265 PREDICTED: hypothetical protein, partial [Strongylocentrotus purpuratus]
194755767 GF11669 [Drosophila ananassae]

17864654 will die slowly CG17437-PA [Drosophila melanogaster]
126330274 PREDICTED: similar to WD repeat domain 57 (U5 snRNP specific) [Monodelphis domestica]
150865806 hypothetical protein PICST_60492 [Pichia stipitis CBS 6054]
194768847 GF22217 [Drosophila ananassae]
194913287 GG12938 [Drosophila erecta]
195477589 GE16263 [Drosophila yakuba]
154279498 conserved hypothetical protein [Ajellomyces capsulatus NAm1]
169845179 hypothetical protein CC1G_00488 [Coprinopsis cinerea okayama7#130]
183231001 transcription initiation factor TFIID subunit 5, putative [Entamoeba histolytica HM-1:IMSS]
198425806 PREDICTED: similar to WD repeat domain 5 [Ciona intestinalis]
70982352 U5 snRNP complex subunit [Aspergillus fumigatus Af293]
76157588 SJCHGC09556 protein [Schistosoma japonicum]
83406042 WDR51A protein [Homo sapiens]
119177669 hypothetical protein CIMG_07750 [Coccidioides immitis RS]
119489225 U5 snRNP complex subunit, putative [Neosartorya fischeri NRRL 181]
121709131 U5 snRNP complex subunit, putative [Aspergillus clavatus NRRL 1]
168034305 predicted protein [Physcomitrella patens subsp. patens]
169782078 hypothetical protein [Aspergillus oryzae RIB40]
24640523 Gbeta5 CG10763-PA [Drosophila melanogaster]
73950485 PREDICTED: similar to WD-repeat protein 57 [Canis familiaris]
109098168 PREDICTED: similar to WD repeat domain 51B [Macaca mulatta]
114052052 WD repeat domain 57 (U5 snRNP specific) [Bos taurus]
115298670 U5 snRNP-specific protein (Prp8-binding) [Mus musculus]
148694366 guanine nucleotide binding protein, beta 5, isoform CRA_d [Mus musculus]
149694963 PREDICTED: similar to WD repeat domain 57 (U5 snRNP specific) [Equus caballus]
194763703 GF20976 [Drosophila ananassae]
195029537 GH22022 [Drosophila grimshawi]
195132861 GI21774 [Drosophila mojavensis]
195397113 GJ16947 [Drosophila virilis]
195432210 GK19997 [Drosophila willistoni]
197381061 hypothetical protein LOC313056 [Rattus norvegicus]
3820594 U5 snRNP-specific 40 kDa protein [Homo sapiens]
73994655 PREDICTED: similar to WD repeat and SOCS box containing protein 2 (WSB-2) (CS box-containing WD protein) isoform 3 [Canis familiaris]
90085014 unnamed protein product [Macaca fascicularis]
114555221 PREDICTED: WD repeat domain 57 (U5 snRNP specific) isoform 1 [Pan troglodytes]
115298668 WD repeat domain 57 (U5 snRNP specific) [Homo sapiens]
116054964 putative WD40 protein Ciao1 (ISS) [Ostreococcus tauri]
154311126 hypothetical protein BC1G_06681 [Botryotinia fuckeliana B05.10]
189054029 unnamed protein product [Homo sapiens]
195049015 GH24861 [Drosophila grimshawi]
195049507 GH24921 [Drosophila grimshawi]
197097664 WD repeat domain 57 (U5 snRNP specific) [Pongo abelii]
12846941 unnamed protein product [Mus musculus]
17488592 AC090119_2 unknown protein [Takifugu rubripes]
17507905 G Protein, Beta subunit family member (gpb-2) [Caenorhabditis elegans]
125981535 GA10552 [Drosophila pseudoobscura pseudoobscura]
125983742 GA14510 [Drosophila pseudoobscura pseudoobscura]
126324702 PREDICTED: similar to SOCS box-containing WD protein SWiP-2 [Monodelphis domestica]
154305580 hypothetical protein BC1G_08559 [Botryotinia fuckeliana B05.10]
156065917 hypothetical protein SS1G_00969 [Sclerotinia sclerotiorum 1980]
164661589 hypothetical protein MGL_1185 [Malassezia globosa CBS 7966]
171677452 unnamed protein product [Podospora anserina]
195134797 GI14411 [Drosophila mojavensis]
195393724 GJ18759 [Drosophila virilis]
195432074 GK19916 [Drosophila willistoni]
42733993 similar to Arabidopsis thaliana (Mouse-ear cress). Putative splicing factor [Dictyostelium discoideum]
48096118 PREDICTED: similar to Lissencephaly-1 CG8440-PA, isoform A [Apis mellifera]
66824199 WD40 repeat-containing protein [Dictyostelium discoideum AX4]
71021279 hypothetical protein UM04723.1 [Ustilago maydis 521]
72108812 PREDICTED: similar to G protein beta 5 subunit [Strongylocentrotus purpuratus]
85095314 hypothetical protein NCU05797 [Neurospora crassa OR74A]
156361180 predicted protein [Nematostella vectensis]
166235371 B Chain B, The Multifunctional Nature Of Gbeta5RGS9 REVEALED FROM ITS Crystal Structure
14029844 AF369757_1 G-protein beta 5 [Ambystoma tigrinum]
20302740 AF391288_4 unknown [Branchiostoma floridae]
51315951 GBB5_TAMST Guanine nucleotide-binding protein subunit beta-5 (Transducin beta chain 5) (Gbeta5)
74193479 unnamed protein product [Mus musculus]
90086137 unnamed protein product [Macaca fascicularis]

91087273 PREDICTED: similar to guanine nucleotide-binding protein beta 5 (g protein beta5) [Tribolium castaneum]
115471407 Os07g0252000 [Oryza sativa (japonica cultivar-group)]
118601168 guanine nucleotide binding protein (G protein), beta 5 [Xenopus tropicalis]
149642745 guanine nucleotide-binding protein, beta-5 subunit [Bos taurus]
156374196 predicted protein [Nematostella vectensis]
158261683 unnamed protein product [Homo sapiens]
169784297 hypothetical protein [Aspergillus oryzae RIB40]
170032381 WD repeat protein 57 [Culex quinquefasciatus]
189054539 unnamed protein product [Homo sapiens]
194034710 PREDICTED: similar to G protein beta subunit 5 short [Sus scrofa]
195335374 GM21820 [Drosophila sechellia]
195584363 GD11312 [Drosophila simulans]
197101527 guanine nucleotide binding protein (G protein), beta 5 [Pongo abelii]
4138841 G-protein beta subunit GPB1 [Filobasidiella neoformans]
18401018 EMB1345 (EMBRYO DEFECTIVE 1345); nucleotide binding [Arabidopsis thaliana]
39945422 hypothetical protein MGG_04693 [Magnaporthe grisea 70-15]
56754257 SJCHGC05198 protein [Schistosoma japonicum]
157125093 guanine nucleotide-binding protein beta 5 (g protein beta5) [Aedes aegypti]
159472284 predicted protein [Chlamydomonas reinhardtii]
170032642 guanine nucleotide-binding protein subunit beta 5 [Culex quinquefasciatus]
66519588 PREDICTED: similar to CG3436-PA, isoform A [Apis mellifera]
73957859 PREDICTED: similar to guanine nucleotide-binding protein, beta 2 isoform 7 [Canis familiaris]
73980251 PREDICTED: similar to WD-repeat protein 39 (Ciao 1 protein) isoform 2 [Canis familiaris]
157116667 wd-repeat protein [Aedes aegypti]
167539822 guanine nucleotide-binding protein subunit beta, putative [Entamoeba dispar SAW760]
183231684 guanine nucleotide-binding protein subunit beta, putative [Entamoeba histolytica HM-1:IMSS]
198423802 PREDICTED: similar to WD repeat domain 57 (U5 snRNP specific) [Ciona intestinalis]
118780497 AGAP009506-PA [Anopheles gambiae str. PEST]
156398837 predicted protein [Nematostella vectensis]
159479754 splicing factor, component of the U4/U6-U5 snRNP complex [Chlamydomonas reinhardtii]
170113765 mycorrhiza-induced NACHT/WD40-repeat domain protein, NWD [Laccaria bicolor S238N-H82]
50313334 guanine nucleotide binding protein beta subunit 2 [Lentinula edodes]
66544556 PREDICTED: similar to G5 CG10763-PA [Apis mellifera]
71004368 hypothetical protein UM00703.1 [Ustilago maydis 521]
72014208 PREDICTED: similar to Prp8bp-pending protein [Strongylocentrotus purpuratus]
145507954 hypothetical protein GSPATT00038800001 [Paramecium tetraurelia strain d4-2]
145529834 hypothetical protein GSPATT00017647001 [Paramecium tetraurelia strain d4-2]
156549788 PREDICTED: similar to guanine nucleotide-binding protein beta 5 (g protein beta5) [Nasonia vitripennis]
157167899 wd-repeat protein [Aedes aegypti]
170067686 will die slowly [Culex quinquefasciatus]
170581114 Biotin/lipoate A/B protein ligase family protein [Brugia malayi]
195394854 GJ10433 [Drosophila virilis]
116201043 conserved hypothetical protein [Chaetomium globosum CBS 148.51]
154285234 predicted protein [Ajellomyces capsulatus NAm1]
161701430 G-protein beta subunit [Rhizoctonia solani]
189314268 G-protein beta subunit [Rhizoctonia solani]
20129125 CG3436 CG3436-PA, isoform A [Drosophila melanogaster]
41053421 WD repeat domain 57 (U5 snRNP specific) [Danio rerio]
66818199 G protein b-subunit [Dictyostelium discoideum AX4]
91083243 PREDICTED: similar to guanine nucleotide-binding protein beta 2 (g protein beta2) [Tribolium castaneum]
91093477 PREDICTED: similar to AGAP009506-PA [Tribolium castaneum]
109391151 lissencephaly protein 1-like [Chlamydomonas reinhardtii]
125985431 GA17451 [Drosophila pseudoobscura pseudoobscura]
148693879 WD repeat domain 61, isoform CRA_a [Mus musculus]
156843146 hypothetical protein Kpol_526p37 [Vanderwaltozyma polyspora DSM 70294]
157117947 guanine nucleotide-binding protein beta 2 (g protein beta2) [Aedes aegypti]
158284471 AGAP012731-PA [Anopheles gambiae str. PEST]
158301959 AGAP001506-PA [Anopheles gambiae str. PEST]
170054462 guanine nucleotide-binding protein beta 2 [Culex quinquefasciatus]
194766421 GF20700 [Drosophila ananassae]
194853260 GG24701 [Drosophila erecta]
195035271 GH11537 [Drosophila grimshawi]
195114434 GI17028 [Drosophila mojavensis]
195350019 GM16719 [Drosophila sechellia]
195386418 GJ24581 [Drosophila virilis]
195433100 GK23911 [Drosophila willistoni]
195470182 GE16626 [Drosophila yakuba]
7160787 guanine nucleotide-binding protein beta subunit [Calliphora vicina]
47221560 unnamed protein product [Tetraodon nigroviridis]

66514532 PREDICTED: similar to G protein -subunit 76C CG8770-PA [Apis mellifera]
91076958 PREDICTED: similar to AGAP007626-PA [Tribolium castaneum]
112982984 will die slowly [Bombyx mori]
116180242 hypothetical protein CHGG_00749 [Chaetomium globosum CBS 148.51]
156547315 PREDICTED: similar to guanine nucleotide-binding protein beta 2 (g protein beta2) [Nasonia vitripennis]
195431226 GK19416 [Drosophila willistoni]
24654584 CG10931 CG10931-PA [Drosophila melanogaster]
146417863 hypothetical protein PGUG_02628 [Pichia guilliermondii ATCC 6260]
156545416 PREDICTED: similar to ENSANGP00000010898 [Nasonia vitripennis]
169781398 hypothetical protein [Aspergillus oryzae RIB40]
190346440 hypothetical protein PGUG_02628 [Pichia guilliermondii ATCC 6260]
195153389 GL17219 [Drosophila persimilis]
19068020 G-protein beta subunit 1 [Phytophthora infestans]
157346216 unnamed protein product [Vitis vinifera]
164658608 hypothetical protein MGL_2225 [Malassezia globosa CBS 7966]
169850105 hypothetical protein CC1G_08354 [Coprinopsis cinerea okayama7#130]
194693802 unknown [Zea mays]
194696962 unknown [Zea mays]
195166006 GL27199 [Drosophila persimilis]
197320656 G protein beta subunit 1 [Phytophthora sojae]
15224356 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
56607084 gpb-2 [Caenorhabditis briggsae]
58266936 hypothetical protein [Cryptococcus neoformans var. neoformans JEC21]
91077142 PREDICTED: similar to will die slowly [Tribolium castaneum]
115620239 PREDICTED: similar to heterotrimeric guanine nucleotide-binding protein beta subunit isoform 2 [Strongylocentrotus purpuratus]
118385633 hypothetical protein TTHERM_01113100 [Tetrahymena thermophila SB210]
134110748 hypothetical protein CNBD2480 [Cryptococcus neoformans var. neoformans B-3501A]
148683063 guanine nucleotide binding protein, beta 1, isoform CRA_a [Mus musculus]
194880857 GG21819 [Drosophila erecta]
195487743 GE11895 [Drosophila yakuba]
119631821 WD repeat, sterile alpha motif and U-box domain containing 1, isoform CRA_g [Homo sapiens]
146169607 hypothetical protein TTHERM_00194680 [Tetrahymena thermophila SB210]
149567171 PREDICTED: similar to WD repeat domain 57 (U5 snRNP specific), partial [Ornithorhynchus anatinus]
198460650 GA10650 [Drosophila pseudoobscura pseudoobscura]
121014 GBB_LOLFO Guanine nucleotide-binding protein subunit beta
1730218 GBB_LYMST Guanine nucleotide-binding protein subunit beta
37223059 visual G-protein beta subunit [Loligo pealei]
46391574 G protein beta subunit [Pinctada fucata]
66523907 PREDICTED: similar to Guanine nucleotide-binding protein subunit beta 1 isoform 1 [Apis mellifera]
115620241 PREDICTED: similar to heterotrimeric guanine nucleotide-binding protein beta subunit isoform 1 [Strongylocentrotus purpuratus]
125530898 hypothetical protein OsI_031422 [Oryza sativa (indica cultivar-group)]
158288821 AGAP000444-PA [Anopheles gambiae str. PEST]
196007722 G protein beta 1 subunit [Trichoplax adhaerens]
6643 G-protein [Caenorhabditis elegans]
193436 G-protein beta subunit
306785 G protein beta subunit
984551 G protein beta 2 subunit [Mus musculus]
984553 G protein beta 1 subunit [Rattus norvegicus]
984555 G protein beta 2 subunit [Rattus norvegicus]
1942174 B Chain B, Heterotrimeric Complex Of A Gt-AlphaGI-Alpha Chimera And The Gt-Beta-Gamma Subunits
3023838 GBB1_XENLA Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 (Transducin beta chain 1) (XGbeta1)
3913720 GBB1_HOMAM Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 (Transducin beta chain 1)
4099537 G protein beta1 subunit [Rattus norvegicus]
4504053 guanine nucleotide-binding protein, beta-3 subunit [Homo sapiens]
11055998 guanine nucleotide-binding protein, beta-4 subunit [Homo sapiens]
11127727 AF277161_1 G-protein B1 subunit [Ambystoma tigrinum]
11127729 AF277162_1 G-protein B3 subunit [Ambystoma tigrinum]
11177902 guanine nucleotide-binding protein, beta-3 subunit [Rattus norvegicus]
12652731 Guanine nucleotide binding protein (G protein), beta polypeptide 3 [Homo sapiens]
13937391 guanine nucleotide-binding protein, beta-2 subunit [Mus musculus]
17534675 G Protein, Beta subunit family member (gpb-1) [Caenorhabditis elegans]
17933598 G protein beta-subunit 13F CG10545-PA, isoform A [Drosophila melanogaster]
19113627 splicing factor Spf38 [Schizosaccharomyces pombe]
20257506 AF501886_1 guanine nucleotide binding protein beta 4 [Mus musculus]
20502976 guanine nucleotide-binding protein, beta-3 subunit [Mus musculus]
31214985 AGAP005911-PA [Anopheles gambiae str. PEST]
31542899 guanine nucleotide-binding protein, beta-4 subunit [Mus musculus]

47086811 guanine nucleotide binding protein (G protein), beta polypeptide 1 [Danio rerio]
47087315 guanine nucleotide binding protein (G protein), beta polypeptide 1, like [Danio rerio]
47229791 unnamed protein product [Tetraodon nigroviridis]
51315872 GBB1_CRIGR Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 (Transducin beta chain 1)
51872141 G protein beta 1 subunit [Litopenaeus vannamei]
53130298 hypothetical protein [Gallus gallus]
55742083 guanine nucleotide binding protein (G protein), beta polypeptide 2 [Xenopus tropicalis]
55742672 guanine nucleotide-binding protein, beta-3 subunit [Canis lupus familiaris]
56757698 SJCHGC05537 protein [Schistosoma japonicum]
57110000 PREDICTED: similar to guanine nucleotide-binding protein, beta-4 subunit [Canis familiaris]
58331839 guanine nucleotide binding protein, beta 3 [Xenopus tropicalis]
61806564 guanine nucleotide binding protein (G protein), beta polypeptide 2 [Danio rerio]
62078551 guanine nucleotide binding protein (G protein), beta 4 [Rattus norvegicus]
72004626 PREDICTED: similar to WD40 protein Ciao1 [Strongylocentrotus purpuratus]
73996099 PREDICTED: transducin beta-3-subunit [Canis familiaris]
74186513 unnamed protein product [Mus musculus]
74199309 unnamed protein product [Mus musculus]
74212166 unnamed protein product [Mus musculus]
74217716 unnamed protein product [Mus musculus]
90079179 unnamed protein product [Macaca fascicularis]
91085467 PREDICTED: similar to guanine nucleotide-binding protein subunit beta-like [Tribolium castaneum]
114052388 heterotrimeric guanine nucleotide-binding protein beta subunit [Bombyx mori]
115969879 PREDICTED: hypothetical protein, partial [Strongylocentrotus purpuratus]
119888032 guanine nucleotide-binding protein subunit beta-like [Microplitis mediator]
119891872 PREDICTED: similar to G-protein beta-subunit [Bos taurus]
126306623 PREDICTED: similar to Guanine nucleotide binding protein (G protein), beta polypeptide 1 [Monodelphis domestica]
126309287 PREDICTED: similar to cell adhesion molecule NCAM [Monodelphis domestica]
126339999 PREDICTED: similar to transducin beta-3-subunit [Monodelphis domestica]
145477231 hypothetical protein GSPATT00028275001 [Paramecium tetraurelia strain d4-2]
148222671 guanine nucleotide binding protein (G protein), beta polypeptide 1 [Xenopus laevis]
148224158 beta 1 subunit of heterotrimeric GTP-binding protein [Xenopus laevis]
148236273 guanine nucleotide binding protein (G protein), beta polypeptide 3 [Xenopus laevis]
148524806 G-protein beta subunit 1 [Cyprinus carpio]
149642747 guanine nucleotide-binding protein, beta-4 subunit [Bos taurus]
149712433 PREDICTED: similar to G-protein beta-subunit [Equus caballus]
149731126 PREDICTED: guanine nucleotide binding protein (G protein), beta polypeptide 4 [Equus caballus]
149758330 PREDICTED: guanine nucleotide binding protein (G protein), beta polypeptide 1 [Equus caballus]
156383572 predicted protein [Nematostella vectensis]
157138353 guanine nucleotide-binding protein beta 3 (g protein beta3) [Aedes aegypti]
157138355 guanine nucleotide-binding protein beta 3 (g protein beta3) [Aedes aegypti]
157169473 guanine nucleotide-binding protein beta 3 (g protein beta3) [Aedes aegypti]
157278441 guanine nucleotide-binding protein beta subunit [Oryzias latipes]
158254574 unnamed protein product [Homo sapiens]
158294989 AGAP005913-PA [Anopheles gambiae str. PEST]
170042642 guanine nucleotide-binding protein subunit beta 1 [Culex quinquefasciatus]
170042644 guanine nucleotide-binding protein subunit beta 1 [Culex quinquefasciatus]
170042647 guanine nucleotide-binding protein subunit beta 1 [Culex quinquefasciatus]
170062956 guanine nucleotide-binding protein subunit beta 1 [Culex quinquefasciatus]
182636902 G-protein beta subunit [Meloidogyne javanica]
189069163 unnamed protein product [Homo sapiens]
193596402 PREDICTED: similar to heterotrimeric guanine nucleotide-binding protein beta subunit [Acyrthosiphon pisum]
193669209 PREDICTED: similar to AGAP009506-PA [Acyrthosiphon pisum]
194763048 GF20505 [Drosophila ananassae]
195044502 GH12882 [Drosophila grimshawi]
195135794 GI14221 [Drosophila mojavensis]
195457086 GK15234 [Drosophila willistoni]
197100735 guanine nucleotide binding protein (G protein), beta polypeptide 1 [Pongo abelii]
198461644 GA23154 [Drosophila pseudoobscura pseudoobscura]
12832206 unnamed protein product [Mus musculus]
31542399 WD40 protein Ciao1 [Mus musculus]
50421219 hypothetical protein DEHA0D16841g [Debaryomyces hansenii CBS767]
56912225 cytosolic iron-sulfur protein assembly 1 homolog [Rattus norvegicus]
62898075 WD40 protein Ciao1 variant [Homo sapiens]
73666448 G protein beta 1 subunit [Uta stansburiana]
73980249 PREDICTED: similar to WD-repeat protein 39 (Ciao 1 protein) isoform 1 [Canis familiaris]
82697393 WD repeat domain 39 [Bos taurus]
109103855 PREDICTED: WD repeat domain 39 [Macaca mulatta]
126303559 PREDICTED: hypothetical protein [Monodelphis domestica]
149727322 PREDICTED: similar to Protein CIAO1 (WD repeat-containing protein 39) [Equus caballus]

170062958 conserved hypothetical protein [Culex quinquefasciatus]
189054959 unnamed protein product [Homo sapiens]
195109502 GI23133 [Drosophila mojavensis]
199431778 DEHA2D15466p [Debaryomyces hansenii]
50726890 guanine nucleotide binding protein (G protein), beta polypeptide 3 [Danio rerio]
114550604 PREDICTED: hypothetical protein [Pan troglodytes]
170062337 CIAO1 [Culex quinquefasciatus]
50294620 hypothetical protein CAGL0M08646g [Candida glabrata CBS138]
71896295 wdr57 protein [Xenopus tropicalis]
71993616 Y18D10A.9 [Caenorhabditis elegans]
79323100 EMB1345 (EMBRYO DEFECTIVE 1345); nucleotide binding [Arabidopsis thaliana]
116059921 putative protein (ISS) [Ostreococcus tauri]
118083304 PREDICTED: similar to G-protein beta-subunit [Gallus gallus]
147899844 hypothetical protein LOC379514 [Xenopus laevis]
148236149 U5 snRNP-specific 40 kDa protein (hPrp8-binding) [Xenopus laevis]
157132872 wd-repeat protein [Aedes aegypti]
168037038 predicted protein [Physcomitrella patens subsp. patens]
193634160 PREDICTED: similar to wd-repeat protein [Acyrthosiphon pisum]
194225819 PREDICTED: similar to LOC787677 protein [Equus caballus]
198438361 PREDICTED: similar to Protein CIAO1 (WD repeat-containing protein 39) [Ciona intestinalis]
109157928 A Chain A, Structure Of Wdr5
109468132 PREDICTED: similar to will die slowly CG17437-PA [Rattus norvegicus]
115465309 Os05g0552300 [Oryza sativa (japonica cultivar-group)]
158285326 AGAP007626-PA [Anopheles gambiae str. PEST]
159475571 hypothetical protein CHLREDRAFT_149069 [Chlamydomonas reinhardtii]
195054196 GH22558 [Drosophila grimshawi]
195999006 conserved hypothetical protein [Trichoplax adhaerens]
13173175 AF326340_1 unknown protein i8 [Aedes aegypti]
19335616 unknown protein [Aedes aegypti]
19335618 unknown protein [Aedes aegypti]
19922278 Ciao1 CG12797-PA [Drosophila melanogaster]
66807159 WD40 repeat-containing protein [Dictyostelium discoideum AX4]
123470809 guanine nucleotide-binding protein beta subunit, putative [Trichomonas vaginalis G3]
123475041 guanine nucleotide-binding protein beta subunit, putative [Trichomonas vaginalis G3]
125539816 hypothetical protein OsI_007444 [Oryza sativa (indica cultivar-group)]
125807283 GA11817 [Drosophila pseudoobscura pseudoobscura]
157109057 hypothetical protein AaeL_AAEL005208 [Aedes aegypti]
194754134 GF12085 [Drosophila ananassae]
194882955 GG22391 [Drosophila erecta]
195334483 GM20174 [Drosophila sechellia]
195436246 GK22124 [Drosophila willistoni]
195486114 GE12280 [Drosophila yakuba]
195583506 GD25651 [Drosophila simulans]
45198717 AFR199Cp [Ashbya gossypii ATCC 10895]
57525219 WD repeat domain 5 [Gallus gallus]
58332678 WD repeat domain 5 [Xenopus tropicalis]
73967734 PREDICTED: similar to WD repeat domain 5 [Canis familiaris]
74184311 unnamed protein product [Mus musculus]
115439261 Os01g0686800 [Oryza sativa (japonica cultivar-group)]
126336934 PREDICTED: hypothetical protein [Monodelphis domestica]
147902750 WD repeat domain 5 [Xenopus laevis]
149410283 PREDICTED: similar to Chain A, Structure Of Wdr5 [Ornithorhynchus anatinus]
156051526 hypothetical protein SS1G_07170 [Sclerotinia sclerotiorum 1980]
156545541 PREDICTED: similar to wd-repeat protein [Nasonia vitripennis]
157428074 WD repeat domain 5 [Bos taurus]
169403953 WD repeat domain 5 [Danio rerio]
194697160 unknown [Zea mays]
195449505 GK22499 [Drosophila willistoni]
196004508 hypothetical protein TRIADDRAFT_55834 [Trichoplax adhaerens]
15235470 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
50309847 unnamed protein product [Kluyveromyces lactis]
62859923 hypothetical protein LOC549641 [Xenopus (Silurana) tropicalis]
66825749 hypothetical protein DDBDRAFT_0190506 [Dictyostelium discoideum AX4]
115385356 conserved hypothetical protein [Aspergillus terreus NIH2624]
123445505 guanine nucleotide-binding protein beta subunit, putative [Trichomonas vaginalis G3]
193807928 U5 snrnp-specific protein, putative [Plasmodium knowlesi strain H]
881422 G beta like protein
50549007 hypothetical protein [Yarrowia lipolytica]
71026364 hypothetical protein TP03_0734 [Theileria parva strain Muguga]

91992949 GNB1 protein [Homo sapiens]
110749303 PREDICTED: similar to CG12797-PA [Apis mellifera]
164659688 hypothetical protein MGL_1967 [Malassezia globosa CBS 7966]
194742241 GF17851 [Drosophila ananassae]
195330209 GM23851 [Drosophila sechellia]
195499660 GE26002 [Drosophila yakuba]
195572099 GD18659 [Drosophila simulans]
198418997 PREDICTED: similar to WD repeat domain 38 [Ciona intestinalis]
17505659 C18E3.5 [Caenorhabditis elegans]
17551498 F08G12.2 [Caenorhabditis elegans]
21356075 CG3909 CG3909-PA [Drosophila melanogaster]
71028560 guanine nucleotide-binding protein [Theileria parva strain Muguga]
84996473 guanine-nucleotide binding protein, putative [Theileria annulata]
157768410 Hypothetical protein CBG17375 [Caenorhabditis briggsae AF16]
168032626 predicted protein [Physcomitrella patens subsp. patens]
195056705 GH22988 [Drosophila grimshawi]
195122304 GI18956 [Drosophila mojavensis]
195381325 GJ21562 [Drosophila virilis]
6320473 Essential protein involved in assembly of cytosolic and nuclear iron-sulfur proteins; Cia1p [Saccharomyces cerevisiae]
7023854 unnamed protein product [Homo sapiens]
23199987 WD repeat domain 5B [Homo sapiens]
28279952 Ciao1 protein [Danio rerio]
50233904 cytosolic iron-sulfur protein assembly 1 homolog [Danio rerio]
109033390 PREDICTED: similar to WD repeat domain 5B [Macaca mulatta]
114588833 PREDICTED: WD repeat domain 5B [Pan troglodytes]
145357095 predicted protein [Ostreococcus lucimarinus CCE9901]
145499449 hypothetical protein GSPATT00037097001 [Paramecium tetraurelia strain d4-2]
149731568 PREDICTED: similar to WD repeat domain 5B [Equus caballus]
151942244 cytosolic iron-sulfur protein assembly protein [Saccharomyces cerevisiae YJM789]
157356023 unnamed protein product [Vitis vinifera]
159476554 U5 snRNP-specific protein [Chlamydomonas reinhardtii]
197098102 WD repeat domain 5B [Pongo abelii]
66820452 hypothetical protein DDB_0185122 [Dictyostelium discoideum AX4]
109659824 putative G-protein beta subunit [Dugesia japonica]
112491198 A Chain A, Wdr5delta23
114051355 WD repeat domain 61 [Bombyx mori]
1749825 G protein beta-subunit-like protein [Nicotiana plumbaginifolia]
21312318 WD repeat domain 5B [Mus musculus]
45185777 ACR091Wp [Ashbya gossypii ATCC 10895]
67078490 WD repeat domain 5B [Rattus norvegicus]
71028286 U5 small nuclear ribonucleoprotein [Theileria parva strain Muguga]
118481885 unknown [Populus trichocarpa]
126322724 PREDICTED: similar to Chain A, Structure Of Wdr5 [Monodelphis domestica]
153862517 putative domain containing guanine nucleotide-binding protein [Theileria sp. China 2]
62088146 guanine nucleotide-binding protein, beta-3 subunit variant [Homo sapiens]
82621178 ArcA2 protein-like [Solanum tuberosum]
91091184 PREDICTED: similar to wd-repeat protein [Tribolium castaneum]
156393484 predicted protein [Nematostella vectensis]
195063051 GH25104 [Drosophila grimshawi]
4589834 LeArcA1 protein [Solanum lycopersicum]
4589836 LeArcA2 protein [Solanum lycopersicum]
15221916 guanine nucleotide-binding family protein / activated protein kinase C receptor, putative / RACK, putative [Arabidopsis thaliana]
15229589 guanine nucleotide-binding family protein / activated protein kinase C receptor (RACK1) [Arabidopsis thaliana]
21593440 guanine nucleotide-binding protein, putative [Arabidopsis thaliana]
124512326 u5 snrnp-specific protein, putative [Plasmodium falciparum 3D7]
145537415 hypothetical protein GSPATT00039789001 [Paramecium tetraurelia strain d4-2]
169781068 hypothetical protein [Aspergillus oryzae RIB40]
170030890 WD repeat protein 61 [Culex quinquefasciatus]
3023847 GBLP_MEDSA Guanine nucleotide-binding protein subunit beta-like protein
3023858 GBLP_SOYBN Guanine nucleotide-binding protein subunit beta-like protein
17541222 RACK1 (mammalian Receptor of Activated C Kinase) homolog family member (rack-1) [Caenorhabditis elegans]
47213175 unnamed protein product [Tetraodon nigroviridis]
157756917 Hypothetical protein CBG03402 [Caenorhabditis briggsae AF16]
195999260 histone H3 recognition and presentation By the Wdr5 module Of the Mll1 complex [Trichoplax adhaerens]
115455059 Os03g0725400 [Oryza sativa (japonica cultivar-group)]
156097803 U5 snRNP-specific 40 kDa protein, putative [Plasmodium vivax SaI-1]
156549230 PREDICTED: similar to ENSANGP00000025955 [Nasonia vitripennis]
169785713 hypothetical protein [Aspergillus oryzae RIB40]

195149700 GL10827 [Drosophila persimilis]
195355457 GM22591 [Drosophila sechellia]
199426943 YALI0E20031p [Yarrowia lipolytica]
60729641 JC7987 receptor for activated C kinase, RACK protein - Plasmodium falciparum
70952179 u5 snrnp-specific protein [Plasmodium chabaudi chabaudi]
70954333 guanine nucleotide-binding protein [Plasmodium chabaudi chabaudi]
83314491 hypothetical protein PY02499 [Plasmodium yoelii yoelii str. 17XNL]
124512250 guanine nucleotide-binding protein, putative [Plasmodium falciparum 3D7]
37720868 TUP1-like protein [Pichia angusta]
67904196 hypothetical protein AN9085.2 [Aspergillus nidulans FGSC A4]
68074629 u5 snrnp-specific protein [Plasmodium berghei strain ANKA]
114050869 WD repeat domain 61 [Bombyx mori]
156084690 receptor for activated C kinase, RACK protein [Babesia bovis]
156097727 activated protein kinase C receptor, putative [Plasmodium vivax SaI-1]
193807887 guanine nucleotide-binding protein, putative [Plasmodium knowlesi strain H]
15233721 VIP3 (VERNALIZATION INDEPENDENCE 3); nucleotide binding [Arabidopsis thaliana]
47117687 receptor for activated C kinase 1 [Toxoplasma gondii]
67482507 guanine nucleotide-binding protein subunit beta 2-like 1, putative [Entamoeba histolytica HM-1:IMSS]
67483660 guanine nucleotide-binding protein subunit beta 2-like 1, putative [Entamoeba histolytica HM-1:IMSS]
145353601 Receptor of activated protein kinase C 1A, component of 40S small ribosomal subunit [Ostreococcus lucimarinus CCE9901]
146092531 activated protein kinase c receptor (LACK); p36 LACK protein; guanine nucleotide-binding protein beta subunit-like protein [Leishmania infantum]
147765784 hypothetical protein [Vitis vinifera]
156552750 PREDICTED: similar to wd-repeat protein [Nasonia vitripennis]
167380821 receptor for activated protein kinase C, putative [Entamoeba dispar SAW760]
167387319 receptor for activated protein kinase C, putative [Entamoeba dispar SAW760]
67467723 guanine nucleotide-binding protein subunit beta 2-like 1, putative [Entamoeba histolytica HM-1:IMSS]
157356025 unnamed protein product [Vitis vinifera]
167391605 receptor for activated protein kinase C, putative [Entamoeba dispar SAW760]
13991907 AF368031_1 receptor for activated protein kinase C RACK1 [Heliothis virescens]
28201165 receptor for activated protein kinase C homolog [Mamestra brassicae]
29841234 similar to GenBank Accession Number U30261 G protein beta subunit-like [Schistosoma japonicum]
53148463 receptor for activated protein kinase C homolog [Plutella xylostella]
71159504 activated C kinase 1 receptor [Choristoneura fumiferana]
71159506 activated C kinase 1 receptor [Choristoneura fumiferana]
115345341 receptor for activated protein kinase C RACK 1 isoform 1 [Bombyx mori]
121543663 putative activated C kinase 1 receptor [Maconellicoccus hirsutus]
161015757 activated C kinase 1 receptor [Spodoptera exigua]
2290597 RACK1 [Drosophila melanogaster]
17137396 Receptor of activated protein kinase C 1 CG7111-PA [Drosophila melanogaster]
67083997 guanine nucleotide-binding protein [Ixodes scapularis]
74025270 guanine nucleotide-binding protein subunit beta - like protein [Trypanosoma brucei TREU927]
91089633 PREDICTED: similar to receptor for activated protein kinase C-like [Tribolium castaneum]
119631819 WD repeat, sterile alpha motif and U-box domain containing 1, isoform CRA_e [Homo sapiens]
149287130 G protein beta subunit-like protein [Ornithodoros parkeri]
156837142 hypothetical protein Kpol_297p7 [Vanderwaltozyma polyspora DSM 70294]
159479982 receptor of activated protein kinase C 1 [Chlamydomonas reinhardtii]
189192008 WD repeat containing protein 5 [Pyrenophora tritici-repentis Pt-1C-BFP]
193599204 PREDICTED: similar to will die slowly CG17437-PA [Acyrthosiphon pisum]
194862950 GG23503 [Drosophila erecta]
195034142 GH11378 [Drosophila grimshawi]
195118463 GI18083 [Drosophila mojavensis]
195385922 GJ16744 [Drosophila virilis]
195471411 Rack1 [Drosophila yakuba]
3859574 activated protein kinase C receptor homolog [Trypanosoma vivax]
15229187 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
112491015 A Chain A, Crystal Structure Of Wdr5HISTONE H3 COMPLEX
114319093 receptor for activated protein kinase C-like [Blattella germanica]
125985911 GA20111 [Drosophila pseudoobscura pseudoobscura]
156548258 PREDICTED: similar to receptor for activated protein kinase C-like [Nasonia vitripennis]
193664759 PREDICTED: similar to will die slowly CG17437-PA [Acyrthosiphon pisum]
193698857 PREDICTED: similar to will die slowly CG17437-PA [Acyrthosiphon pisum]
195015990 GH15081 [Drosophila grimshawi]
195094713 GH10054 [Drosophila grimshawi]
195128835 GI13728 [Drosophila mojavensis]
195379714 GJ14074 [Drosophila virilis]
3023852 GBLP_NEUCR Guanine nucleotide-binding protein subunit beta-like protein (Cross-pathway control WD-repeat protein cpc-2)
3132790 CACK protein [Crithidia fasciculata]
39945474 hypothetical protein MGG_04719 [Magnaporthe grisea 70-15]

46136709 GBLP_NEUCR Guanine nucleotide-binding protein beta subunit-like protein (Cross-pathway control WD-repeat protein cpc-2) [Gibberella zeae PH-1]
50425399 hypothetical protein DEHA0F23265g [Debaryomyces hansenii CBS767]
52000514 G-protein beta like WD repeat protein [Fusarium oxysporum]
67527799 GBLP_NEUCR Guanine nucleotide-binding protein beta subunit-like protein (Cross-pathway control WD-repeat protein cpc-2) [Aspergillus nidulans FGSC A4]
70993334 G-protein comlpex beta subunit CpcB [Aspergillus fumigatus Af293]
85095363 hypothetical protein NCU05810 [Neurospora crassa OR74A]
88766385 G-protein beta subunit [Metarhizium anisopliae]
116201077 guanine nucleotide-binding protein beta subunit-like protein [Chaetomium globosum CBS 148.51]
116793583 unknown [Picea sitchensis]
119499942 G-protein comlpex beta subunit CpcB [Neosartorya fischeri NRRL 181]
121708550 G-protein comlpex beta subunit CpcB [Aspergillus clavatus NRRL 1]
145229997 hypothetical protein An01g08850 [Aspergillus niger]
151935569 putative WD40 protein [Xanthoria elegans]
154301805 guanine nucleotide-binding protein beta subunit [Botryotinia fuckeliana B05.10]
155966234 receptor for activated protein kinase C-like protein [Lepeophtheirus salmonis]
168004433 predicted protein [Physcomitrella patens subsp. patens]
168054151 predicted protein [Physcomitrella patens subsp. patens]
168056562 predicted protein [Physcomitrella patens subsp. patens]
169764553 hypothetical protein [Aspergillus oryzae RIB40]
171677424 unnamed protein product [Podospora anserina]
189191366 WD repeat containing protein 51A [Pyrenophora tritici-repentis Pt-1C-BFP]
47216142 unnamed protein product [Tetraodon nigroviridis]
110758702 PREDICTED: similar to WD repeat domain 61 isoform a [Apis mellifera]
112490205 A Chain A, Wdr5 And Histone H3 Lysine 4 Dimethyl Complex At 2.1 Angstrom
112490208 A Chain A, Wdr5 And Unmodified Histone H3 Complex At 2.25 Angstrom
114626678 PREDICTED: similar to OTTHUMP00000022116 [Pan troglodytes]
125553233 hypothetical protein OsI_020175 [Oryza sativa (indica cultivar-group)]
147783779 hypothetical protein [Vitis vinifera]
156084858 u5 snRNP-specific 40 kDa protein, putative [Babesia bovis]
157347948 unnamed protein product [Vitis vinifera]
167524539 predicted protein [Monosiga brevicollis MX1]
196000012 hypothetical protein TRIADDRAFT_21018 [Trichoplax adhaerens]
13752360 AF320333_1 G protein beta subunit-like protein Rkp1 [Schizosaccharomyces pombe]
19114682 RACK1 homologue Cpc2 [Schizosaccharomyces pombe]
58331911 hypothetical protein LOC496482 [Xenopus tropicalis]
113865883 WD repeat domain 38 [Homo sapiens]
116782369 unknown [Picea sitchensis]
118341407 WD repeat domain 38 [Homo sapiens]
168024906 predicted protein [Physcomitrella patens subsp. patens]
169847107 hypothetical protein CC1G_01901 [Coprinopsis cinerea okayama7#130]
171679235 unnamed protein product [Podospora anserina]
6624971 transducin beta like 1 [Mus musculus]
6650997 AF124742_1 guanine nucleotide-binding protein; RACKI [Euprymna scolopes]
67615505 guanine nucleotide-binding protein [Cryptosporidium hominis TU502]
76156138 SJCHGC05548 protein [Schistosoma japonicum]
112491200 A Chain A, Wdr5 In Complex With Unmodified H3k4 Peptide
160331299 gblp [Hemiselmis andersenii]
196000054 hypothetical protein TRIADDRAFT_20668 [Trichoplax adhaerens]
3023855 GBLP_LEIMA Guanine nucleotide-binding protein subunit beta-like protein (Antigen LACK)
11139411 AF125254_1 activated protein kinase C receptor LACK [Leishmania panamensis]
13625467 LACK protective antigen [Leishmania donovani]
13991856 AF363975_1 p36 LACK protein [Leishmania major]
13991858 AF363976_1 p36 LACK protein [Leishmania mexicana]
13991860 AF363977_1 p36 LACK protein [Leishmania amazonensis]
58392406 AGAP010173-PA [Anopheles gambiae str. PEST]
146092534 activated protein kinase c receptor (LACK); p36 LACK protein; guanine nucleotide-binding protein beta subunit-like protein [Leishmania infantum]
149262197 PREDICTED: similar to platelet-activating factor acetylhydrolase 45K chain - bovine [Mus musculus]
154340729 activated protein kinase c receptor (LACK) [Leishmania braziliensis MHOM/BR/75/M2904]
154431140 activated C kinase protein [Leishmania donovani chagasi]
157872022 activated protein kinase c receptor (LACK); guanine nucleotide-binding protein beta subunit-like protein [Leishmania major]
187033013 Hypothetical protein CBG07981 [Caenorhabditis briggsae]
194033585 PREDICTED: similar to LOC787677 protein [Sus scrofa]
116667226 A Chain A, Structural Basis For The Specific Recognition Of Methylated Histone H3 Lysine 4 By The Wd-40 Protein Wdr5
145481499 hypothetical protein GSPATT00030007001 [Paramecium tetraurelia strain d4-2]
157168005 receptor for activated protein kinase c (rack1) [Aedes aegypti]
3746838 38kDa splicing factor; SPF 38 [Homo sapiens]

15451392 hypothetical protein [Macaca fascicularis]
84996217 U5 snRNP-specific subunit, putative [Theileria annulata]
118376560 Will die slowly protein, putative [Tetrahymena thermophila SB210]
125986748 GA18537 [Drosophila pseudoobscura pseudoobscura]
168018581 WD40 repeat protein, COMPASS complex protein [Physcomitrella patens subsp. patens]
168039898 WD40 repeat protein, COMPASS complex protein [Physcomitrella patens subsp. patens]
168042307 WD40 repeat protein, COMPASS complex protein [Physcomitrella patens subsp. patens]
195160072 GL16277 [Drosophila persimilis]
145522997 hypothetical protein GSPATT00014847001 [Paramecium tetraurelia strain d4-2]
156368809 predicted protein [Nematostella vectensis]
195035807 GH11684 [Drosophila grimshawi]
195343116 GM18660 [Drosophila sechellia]
195401172 GJ16167 [Drosophila virilis]
189237943 PREDICTED: similar to wd-repeat protein [Tribolium castaneum]
195049891 GH13463 [Drosophila grimshawi]
195063055 GH25102 [Drosophila grimshawi]
195079162 GH11784 [Drosophila grimshawi]
195079166 GH11782 [Drosophila grimshawi]
12849020 unnamed protein product [Mus musculus]
13376840 WD repeat domain 61 [Homo sapiens]
41053609 hypothetical protein LOC393827 [Danio rerio]
45360929 WD repeat domain 61 [Xenopus tropicalis]
48146783 REC14 [Homo sapiens]
73951322 PREDICTED: similar to WD repeat domain 61 isoform 1 [Canis familiaris]
74150336 unnamed protein product [Mus musculus]
84000167 WD repeat domain 61 [Bos taurus]
91088247 PREDICTED: similar to WD repeat domain 51B [Tribolium castaneum]
94536819 WD repeat domain 61 [Gallus gallus]
147900923 MGC81859 protein [Xenopus laevis]
158300125 AGAP010318-PA [Anopheles gambiae str. PEST]
194039487 PREDICTED: similar to WD repeat domain 61 [Sus scrofa]
197127190 putative recombination protein REC14 [Taeniopygia guttata]
198425920 PREDICTED: similar to WD repeat domain 61 [Ciona intestinalis]
109098932 PREDICTED: similar to WD repeat and SOCS box-containing protein 2 (WSB-2) (CS box-containing WD protein) [Macaca mulatta]
58037455 WD repeat domain 38 [Mus musculus]
158291205 AGAP012525-PA [Anopheles gambiae str. PEST]
194706570 unknown [Zea mays]
116180278 hypothetical protein CHGG_00767 [Chaetomium globosum CBS 148.51]
148693880 WD repeat domain 61, isoform CRA_b [Mus musculus]
156399668 predicted protein [Nematostella vectensis]
194388882 unnamed protein product [Homo sapiens]
25091532 WSL17_SCHMA WD repeat-containing protein SL1-17
115446591 Os02g0543400 [Oryza sativa (japonica cultivar-group)]
159108753 WD-40 repeat protein [Giardia lamblia ATCC 50803]
167523641 predicted protein [Monosiga brevicollis MX1]
168065866 predicted protein [Physcomitrella patens subsp. patens]
169766972 hypothetical protein [Aspergillus oryzae RIB40]
66529744 PREDICTED: similar to mitogen-activated protein kinase organizer 1 isoform 1 [Apis mellifera]
159482890 predicted protein [Chlamydomonas reinhardtii]
18424859 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
73957855 PREDICTED: similar to guanine nucleotide-binding protein, beta 2 isoform 5 [Canis familiaris]
157339972 unnamed protein product [Vitis vinifera]
195996583 hypothetical protein TRIADDRAFT_52348 [Trichoplax adhaerens]
195999118 hypothetical protein TRIADDRAFT_37206 [Trichoplax adhaerens]
193705826 PREDICTED: similar to predicted protein [Acyrthosiphon pisum]
47222782 unnamed protein product [Tetraodon nigroviridis]
115471097 Os07g0205200 [Oryza sativa (japonica cultivar-group)]
159479442 hypothetical protein CHLREDRAFT_120222 [Chlamydomonas reinhardtii]
156408371 predicted protein [Nematostella vectensis]
82541632 hypothetical protein PY00516 [Plasmodium yoelii yoelii str. 17XNL]
159476014 hypothetical protein CHLREDRAFT_119580 [Chlamydomonas reinhardtii]
2982616 arcA 3 [Nicotiana tabacum]
159465351 notchless-like protein [Chlamydomonas reinhardtii]
349828 Miller-Dieker lissencephaly protein
115620243 PREDICTED: similar to heterotrimeric guanine nucleotide-binding protein beta subunit isoform 3 [Strongylocentrotus purpuratus]
115389746 hypothetical protein ATEG_03200 [Aspergillus terreus NIH2624]
50978464 hypothetical protein [Homo sapiens]

125543892 hypothetical protein OsI_011264 [Oryza sativa (indica cultivar-group)]
17481223 G protein beta subunit [Halocynthia roretzi]
195997775 hypothetical protein TRIADDRAFT_4112 [Trichoplax adhaerens]
159478088 hypothetical protein CHLREDRAFT_72099 [Chlamydomonas reinhardtii]
26345856 unnamed protein product [Mus musculus]
55925393 hypothetical protein LOC492811 [Danio rerio]
115767186 PREDICTED: similar to WD repeat domain 51B [Strongylocentrotus purpuratus]
74272665 G protein beta subunit-like polypeptide [Chlamydomonas incerta]
169844320 hypothetical protein CC1G_03675 [Coprinopsis cinerea okayama7#130]
2654167 activated protein kinase C receptor homolog LACK [Leishmania donovani]
15208165 hypothetical protein [Macaca fascicularis]
164659298 hypothetical protein MGL_1772 [Malassezia globosa CBS 7966]
20143912 WD repeat and SOCS box-containing 1 isoform 2 [Homo sapiens]
73966905 PREDICTED: similar to WD SOCS-box protein 1 isoform 2 isoform 4 [Canis familiaris]
156394489 predicted protein [Nematostella vectensis]
194206419 PREDICTED: similar to WD repeat domain 61 [Equus caballus]
148694885 RIKEN cDNA 1700123D08, isoform CRA_a [Mus musculus]
114658338 PREDICTED: hypothetical protein [Pan troglodytes]
73994653 PREDICTED: similar to WD repeat and SOCS box containing protein 2 (WSB-2) (CS box-containing WD protein) isoform 2 [Canis familiaris]
145528930 hypothetical protein GSPATT00039521001 [Paramecium tetraurelia strain d4-2]
194679284 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF-associated factor 65 beta) (PAF65-beta), partial [Bos taurus]
167521958 predicted protein [Monosiga brevicollis MX1]
126297983 PREDICTED: similar to Chain A, Structure Of Wdr5 [Monodelphis domestica]
77745452 G protein beta subunit-like protein [Solanum tuberosum]
47209012 unnamed protein product [Tetraodon nigroviridis]
110750071 PREDICTED: similar to TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 5L (PCAF-associated factor 65 beta) (PAF65-beta) [Apis mellifera]
189538279 PREDICTED: hypothetical protein [Danio rerio]
109510715 PREDICTED: similar to TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor [Rattus norvegicus]
133754986 CG1796 [Drosophila orena]
169785695 hypothetical protein [Aspergillus oryzae RIB40]
90085383 unnamed protein product [Macaca fascicularis]
170595503 Guanine nucleotide-binding protein beta subunit 1 [Brugia malayi]
197126385 Tango4 [Drosophila simulans]
76155166 SJCHGC00746 protein [Schistosoma japonicum]
156373139 predicted protein [Nematostella vectensis]
159468029 hypothetical protein CHLREDRAFT_115915 [Chlamydomonas reinhardtii]
195093257 GH25232 [Drosophila grimshawi]
149047775 WD repeat, SAM and U-box domain containing 1, isoform CRA_a [Rattus norvegicus]
156408666 predicted protein [Nematostella vectensis]
145522318 hypothetical protein GSPATT00014536001 [Paramecium tetraurelia strain d4-2]
159150808 transport and golgi organization 4 [Drosophila simulans]
159150810 transport and golgi organization 4 [Drosophila melanogaster]
159150812 transport and golgi organization 4 [Drosophila melanogaster]
159150816 transport and golgi organization 4 [Drosophila melanogaster]
118098854 PREDICTED: similar to SOCS box-containing WD protein SWiP-2 [Gallus gallus]
196017887 hypothetical protein TRIADDRAFT_51237 [Trichoplax adhaerens]
50553520 hypothetical protein [Yarrowia lipolytica]
195179508 GL26972 [Drosophila persimilis]
171689846 unnamed protein product [Podospora anserina]
194380666 unnamed protein product [Homo sapiens]
193883168 hypothetical protein CG1796 [Drosophila melanogaster]
169606868 hypothetical protein SNOG_06484 [Phaeosphaeria nodorum SN15]
62955321 hypothetical protein LOC550366 [Danio rerio]
156345519 hypothetical protein NEMVEDRAFT_v1g145097 [Nematostella vectensis]
163111 G protein beta subunit
145351018 predicted protein [Ostreococcus lucimarinus CCE9901]
190896128 U5 snRNP-specific protein-like factor [Populus tremula]
190896130 U5 snRNP-specific protein-like factor [Populus tremula]
190896134 U5 snRNP-specific protein-like factor [Populus tremula]
190896144 U5 snRNP-specific protein-like factor [Populus tremula]
190896148 U5 snRNP-specific protein-like factor [Populus tremula]
190896150 U5 snRNP-specific protein-like factor [Populus tremula]
190896158 U5 snRNP-specific protein-like factor [Populus tremula]
190896164 U5 snRNP-specific protein-like factor [Populus tremula]
190896182 U5 snRNP-specific protein-like factor [Populus tremula]
149631401 PREDICTED: hypothetical protein, partial [Ornithorhynchus anatinus]

196006980 hypothetical protein TRIADDRAFT_26455 [Trichoplax adhaerens]
169866155 hypothetical protein CC1G_13542 [Coprinopsis cinerea okayama7#130]
195560933 GD13813 [Drosophila simulans]
169785655 hypothetical protein [Aspergillus oryzae RIB40]
93141136 chloroplast heterotrimeric G-protein complex protein [Solanum tuberosum]
183230496 WD domain containing protein [Entamoeba histolytica HM-1:IMSS]
17391209 Tbl1xr1 protein [Mus musculus]
126321880 PREDICTED: similar to WD repeat domain 5, [Monodelphis domestica]
115928964 PREDICTED: hypothetical protein, partial [Strongylocentrotus purpuratus]
74203936 unnamed protein product [Mus musculus]
114627453 PREDICTED: hypothetical protein [Pan troglodytes]
159489584 predicted protein [Chlamydomonas reinhardtii]
124358707 putative WD repeat protein [Cryptomeria japonica]
124358709 putative WD repeat protein [Cryptomeria japonica]
124358711 putative WD repeat protein [Thujopsis dolabrata]
124358713 putative WD repeat protein [Chamaecyparis formosensis]
124358715 putative WD repeat protein [Chamaecyparis pisifera]
124358719 putative WD repeat protein [Chamaecyparis obtusa]
124358721 putative WD repeat protein [Chamaecyparis obtusa]
10764839 AF306867_1 unknown [Ochlerotatus triseriatus]
93141126 chloroplast heterotrimeric G-protein complex protein [Solanum lycopersicum]
93141132 chloroplast heterotrimeric G-protein complex protein [Nicotiana tomentosiformis]
93141134 chloroplast heterotrimeric G-protein complex protein [Capsicum annuum]
93141124 chloroplast heterotrimeric G-protein complex protein [Physalis sp. TA1367]
93141128 chloroplast heterotrimeric G-protein complex protein [Solanum melongena]
93141138 chloroplast heterotrimeric G-protein complex protein [Coffea canephora]
119607779 WD repeat domain 31, isoform CRA_d [Homo sapiens]
93141156 chloroplast heterotrimeric G-protein complex protein [Solanum lycopersicoides]
93141130 chloroplast heterotrimeric G-protein complex protein [Petunia axillaris subsp. parodii]
93141148 chloroplast heterotrimeric G-protein complex protein [Solanum chilense]
93141154 chloroplast heterotrimeric G-protein complex protein [Solanum pennellii]

93
109122915 PREDICTED: similar to Fzr1 protein [Macaca mulatta]
50287713 unnamed protein product [Candida glabrata]
199433044 DEHA2E20966p [Debaryomyces hansenii]
169869638 hypothetical protein CC1G_12115 [Coprinopsis cinerea okayama7#130]
150865463 Anaphase promoting complex, Cdc20, Cdh1, and Ama1 subunits [Pichia stipitis CBS 6054]
118401574 hypothetical protein TTHERM_00441940 [Tetrahymena thermophila SB210]
189201816 hypothetical protein PTRG_06911 [Pyrenophora tritici-repentis Pt-1C-BFP]
50311095 unnamed protein product [Kluyveromyces lactis]
72393359 cell division cycle protein, putative [Trypanosoma brucei TREU927]
158288148 AGAP009338-PA [Anopheles gambiae str. PEST]
29569816 Cdc20 [Branchiostoma floridae]
156355083 predicted protein [Nematostella vectensis]
58270876 cell division control protein [Cryptococcus neoformans var. neoformans JEC21]
115910737 PREDICTED: similar to Cdc20 [Strongylocentrotus purpuratus]
145510144 hypothetical protein GSPATT00009880001 [Paramecium tetraurelia strain d4-2]
145549458 hypothetical protein GSPATT00003880001 [Paramecium tetraurelia strain d4-2]
145553433 hypothetical protein GSPATT00027347001 [Paramecium tetraurelia strain d4-2]
91078224 PREDICTED: similar to WD repeat-containing protein slp1 [Tribolium castaneum]
170590470 fizzy-related protein [Brugia malayi]
110766269 PREDICTED: similar to fizzy CG4274-PA [Apis mellifera]
3298595 fizzy1 [Xenopus laevis]
45360545 cell division cycle 20 homolog [Xenopus tropicalis]
57529659 cell division cycle 20 [Gallus gallus]
148232481 fizzy1 [Xenopus laevis]
148232740 cell division cycle 20 homolog [Xenopus laevis]
6330805 KIAA1242 protein [Homo sapiens]
156362470 predicted protein [Nematostella vectensis]
198419385 PREDICTED: similar to R33374_1 [Ciona intestinalis]
468032 p55CDC
468034 p55CDC
8885513 Cdc20 [Mus musculus]
12653679 CDC20 protein [Homo sapiens]
25282463 cell division cycle 20 homolog [Rattus norvegicus]
26344966 unnamed protein product [Mus musculus]
73978048 PREDICTED: similar to Cell division cycle protein 20 homolog (p55CDC) [Canis familiaris]
114555938 PREDICTED: cell division cycle 20 [Pan troglodytes]

114555940 PREDICTED: cell division cycle 20 [Pan troglodytes]
118402582 cell division cycle 20 [Homo sapiens]
126305835 PREDICTED: similar to cell division cycle 20 [Monodelphis domestica]
126723421 cell division cycle 20 homolog [Bos taurus]
165377264 cell division cycle 20 homolog [Mus musculus]
178056456 cell division cycle 20 homolog [Sus scrofa]
6463679 Fzr1 [Homo sapiens]
37537753 FZR_HUMAN Fizzy-related protein homolog (Fzr) (Cdh1/Hct1 homolog) (hCDH1) (CDC20-like protein 1)
47085811 cell division cycle 20 homolog [Danio rerio]
148725403 cell division cycle 20 homolog [Danio rerio]
41053776 fizzy/cell division cycle 20 related 1 [Danio rerio]
9789959 Fzr1 protein [Mus musculus]
59709489 fizzy/cell division cycle 20 related 1 [Sus scrofa]
73987195 PREDICTED: similar to Fizzy-related protein homolog (Fzr) (Cdh1/Hct1 homolog) isoform 1 [Canis familiaris]
74208570 unnamed protein product [Mus musculus]
83035083 fizzy/cell division cycle 20 related 1 [Bos taurus]
126323390 PREDICTED: similar to FZR1 protein [Monodelphis domestica]
147898443 frizzy-related protein 1 [Xenopus laevis]
149034416 fizzy/cell division cycle 20 related 1 (Drosophila) (predicted), isoform CRA_a [Rattus norvegicus]
149760005 PREDICTED: similar to Fizzy/cell division cycle 20 related 1 (Drosophila) [Equus caballus]
187608337 fizzy/cell division cycle 20 related 1 [Xenopus (Silurana) tropicalis]
28277337 LOC398044 protein [Xenopus laevis]
156395597 predicted protein [Nematostella vectensis]
156537047 PREDICTED: similar to GA15568-PA [Nasonia vitripennis]
189515588 PREDICTED: similar to frizzy-related protein 1 [Danio rerio]
19114073 sleepy homolog Slp1 [Schizosaccharomyces pombe]
66562652 PREDICTED: similar to retina aberrant in pattern CG3000-PA, isoform A isoform 1 [Apis mellifera]
91077232 PREDICTED: similar to retina aberrant in pattern CG3000-PA isoform 1 [Tribolium castaneum]
73987199 PREDICTED: similar to Fizzy-related protein homolog (Fzr) (Cdh1/Hct1 homolog) isoform 3 [Canis familiaris]
158301189 AGAP002114-PA [Anopheles gambiae str. PEST]
2326419 fizzy-related protein [Drosophila melanogaster]
24639727 retina aberrant in pattern CG3000-PA, isoform A [Drosophila melanogaster]
125982807 GA15568 [Drosophila pseudoobscura pseudoobscura]
194764131 GF20850 [Drosophila ananassae]
194888369 GG18523 [Drosophila erecta]
195059791 GH17898 [Drosophila grimshawi]
195130229 GI15421 [Drosophila mojavensis]
195397509 GJ16380 [Drosophila virilis]
195447890 GK25147 [Drosophila willistoni]
195477058 GE16840 [Drosophila yakuba]
195565101 GD16281 [Drosophila simulans]
157106216 cell division cycle 20 (cdc20) (fizzy) [Aedes aegypti]
194383632 unnamed protein product [Homo sapiens]
145525559 hypothetical protein GSPATT00016040001 [Paramecium tetraurelia strain d4-2]
195163934 GL14550 [Drosophila persimilis]
126342272 PREDICTED: similar to R33374_1 [Monodelphis domestica]
47228337 unnamed protein product [Tetraodon nigroviridis]
109003441 PREDICTED: cell division cycle 20 [Macaca mulatta]
167527017 predicted protein [Monosiga brevicollis MX1]
19922858 fizzy-related 2 CG16783-PA [Drosophila melanogaster]
87240604 Cdc20/Fizzy; WD40-like [Medicago truncatula]
164663253 hypothetical protein MGL_0523 [Malassezia globosa CBS 7966]
145546753 hypothetical protein GSPATT00024391001 [Paramecium tetraurelia strain d4-2]
170056401 WD repeat-containing protein srw1 [Culex quinquefasciatus]
13549094 AF348674_1 p55CDC [Sus scrofa]
149493365 PREDICTED: similar to fizzy-related protein, partial [Ornithorhynchus anatinus]

70
114625877 PREDICTED: coronin, actin binding protein, 2A [Pan troglodytes]
158285517 AGAP007525-PA [Anopheles gambiae str. PEST]
45190720 AER114Wp [Ashbya gossypii ATCC 10895]
170058296 coronin [Culex quinquefasciatus]
66507360 PREDICTED: similar to coronin, actin binding protein, 2B [Apis mellifera]
156550141 PREDICTED: similar to ENSANGP00000029333 [Nasonia vitripennis]
122890249 coronin, actin binding protein 2A [Mus musculus]
195029501 GH19864 [Drosophila grimshawi]
73971866 PREDICTED: similar to coronin, actin binding protein, 2A [Canis familiaris]
195119438 GI19814 [Drosophila mojavensis]
195149175 GL10967 [Drosophila persimilis]

195401595 GJ18509 [Drosophila virilis]
195435441 GK15589 [Drosophila willistoni]
24586098 coro CG9446-PA, isoform A [Drosophila melanogaster]
194863990 GG10792 [Drosophila erecta]
195356947 GM11075 [Drosophila sechellia]
195474215 GE24410 [Drosophila yakuba]
126335085 PREDICTED: similar to Coronin, actin binding protein, 2A [Monodelphis domestica]
1654313 WD protein IR10 [Homo sapiens]
16554583 coronin, actin binding protein, 2A [Homo sapiens]
16554585 coronin, actin binding protein, 2A [Homo sapiens]
84000407 coronin, actin binding protein, 2A [Bos taurus]
109110840 PREDICTED: similar to coronin, actin binding protein, 2A [Macaca mulatta]
158259037 unnamed protein product [Homo sapiens]
194225528 PREDICTED: coronin, actin binding protein, 2A [Equus caballus]
58865772 coronin, actin binding protein 2A [Rattus norvegicus]
74178398 unnamed protein product [Mus musculus]
149252483 PREDICTED: similar to coronin, actin binding protein 2A [Mus musculus]
197100107 coronin, actin binding protein, 1B [Pongo abelii]
14149734 coronin, actin binding protein, 1B [Homo sapiens]
55636459 PREDICTED: coronin, actin binding protein, 1B isoform 4 [Pan troglodytes]
73982819 PREDICTED: similar to coronin, actin binding protein, 1B isoform 1 [Canis familiaris]
194218536 PREDICTED: coronin, actin binding protein, 1B [Equus caballus]
119595032 coronin, actin binding protein, 1B, isoform CRA_c [Homo sapiens]
164420723 coronin, actin binding protein, 1B [Bos taurus]
189238186 PREDICTED: similar to GA21791-PA [Tribolium castaneum]
130487941 coronin, actin binding protein, 1B [Oryctolagus cuniculus]
6753494 coronin, actin binding protein 1B [Mus musculus]
9506507 coronin, actin-binding protein, 1B [Rattus norvegicus]
149061951 coronin, actin-binding protein, 1B, isoform CRA_a [Rattus norvegicus]
148230198 coronin homolog [Xenopus laevis]
73982825 PREDICTED: similar to coronin, actin binding protein, 1B isoform 4 [Canis familiaris]
196011976 hypothetical protein TRIADDRAFT_36434 [Trichoplax adhaerens]
74197281 unnamed protein product [Mus musculus]
73982823 PREDICTED: similar to coronin, actin binding protein, 1B isoform 3 [Canis familiaris]
58864882 coronin, actin binding protein 6 [Mus musculus]
148747298 coronin, actin binding protein 6 [Rattus norvegicus]
157118635 coronin [Aedes aegypti]
41281697 coronin, actin binding protein 6 isoform A [Mus musculus]
41281703 coronin, actin binding protein 6 isoform B [Mus musculus]
149053464 coronin, actin binding protein 6, isoform CRA_b [Rattus norvegicus]
149053465 coronin, actin binding protein 6, isoform CRA_c [Rattus norvegicus]
195145703 GL23179 [Drosophila persimilis]
148670416 coronin, actin binding protein 2A, isoform CRA_a [Mus musculus]
149045845 rCG54891, isoform CRA_a [Rattus norvegicus]
41055464 coronin, actin binding protein, 1A [Danio rerio]
148701072 coronin, actin binding protein 1B, isoform CRA_b [Mus musculus]
149061952 coronin, actin-binding protein, 1B, isoform CRA_b [Rattus norvegicus]
41055325 hypothetical protein LOC393367 [Danio rerio]
149641758 PREDICTED: similar to coronin 6/clipin E type C isoform 2 [Ornithorhynchus anatinus]
73982821 PREDICTED: similar to coronin, actin binding protein, 1B isoform 2 [Canis familiaris]
41281700 coronin, actin binding protein 6 isoform C [Mus musculus]
149053463 coronin, actin binding protein 6, isoform CRA_a [Rattus norvegicus]
156379196 predicted protein [Nematostella vectensis]
148670417 coronin, actin binding protein 2A, isoform CRA_b [Mus musculus]
26336148 unnamed protein product [Mus musculus]
61555058 coronin, actin binding protein, 1B [Bos taurus]
38181570 Coro1b protein [Rattus norvegicus]
194381300 unnamed protein product [Homo sapiens]
62088738 coronin, actin binding protein, 1B variant [Homo sapiens]

63 **Cac2 Cluster**
157138123 chromatin assembly factor i P60 subunit [Aedes aegypti]
194757958 GF11107 [Drosophila ananassae]
195026398 GH20632 [Drosophila grimshawi]
195425417 GK10714 [Drosophila willistoni]
194857846 GG25206 [Drosophila erecta]
195153457 GL17201 [Drosophila persimilis]
198460559 GA11885 [Drosophila pseudoobscura pseudoobscura]
13569829 AF367178_1 chromatin assembly factor-1 p105 subunit [Drosophila melanogaster]

24652458 Caf1-105 CG12892-PA [Drosophila melanogaster]
195383666 GJ20141 [Drosophila virilis]
195475358 GE21486 [Drosophila yakuba]
158285483 AGAP007544-PA [Anopheles gambiae str. PEST]
195120846 GI20194 [Drosophila mojavensis]
195333147 GM20520 [Drosophila sechellia]
170051150 chromatin assembly factor 1 subunit B [Culex quinquefasciatus]
74001435 PREDICTED: similar to Chromatin assembly factor 1 subunit B (CAF-1 subunit B) (Chromatin assembly factor I p60 subunit) (CAF-I 60 kDa subunit) (CAF-Ip60) (M-phase phosphoprotein 7) [Canis familiaris]
115951537 PREDICTED: similar to chromatin assembly factor-I p60 subunit [Strongylocentrotus purpuratus]
156836756 hypothetical protein Kpol_361p3 [Vanderwaltozyma polyspora DSM 70294]
91077390 PREDICTED: similar to AGAP007544-PA [Tribolium castaneum]
11096297 AF308859_1 FAS2 [Glycine max]
148229121 chromatin assembly factor 1 subunit B [Bos taurus]
149637398 PREDICTED: similar to chromatin assembly factor-I p60 subunit [Ornithorhynchus anatinus]
126325351 PREDICTED: similar to chromatin assembly factor-I p60 subunit [Monodelphis domestica]
21312470 chromatin assembly factor 1 subunit B [Mus musculus]
67078530 chromatin assembly factor 1, subunit B (p60) [Rattus norvegicus]
118403006 chromatin assembly factor 1 [Xenopus laevis]
148237970 chromatin assembly factor-1 p60 subunit [Xenopus laevis]
170095769 chromatin assembly complex 1 subunit B/CAC2 [Laccaria bicolor S238N-H82]
56201446 chromatin assembly factor-1p60 [Gallus gallus]
56799384 chromatin assembly factor 1, subunit B (p60) [Gallus gallus]
148671809 chromatin assembly factor 1, subunit B (p60), isoform CRA_a [Mus musculus]
4885105 chromatin assembly factor 1 subunit B [Homo sapiens]
114684070 PREDICTED: chromatin assembly factor 1 subunit B [Pan troglodytes]
189053718 unnamed protein product [Homo sapiens]
194226233 PREDICTED: similar to Chromatin assembly factor 1 subunit B (CAF-1 subunit B) (Chromatin assembly factor I p60 subunit) (CAF-I 60 kDa subunit) (CAF-Ip60) (M-phase phosphoprotein 7) [Equus caballus]
50291113 unnamed protein product [Candida glabrata]
156550977 PREDICTED: similar to chromatin assembly factor i P60 subunit [Nasonia vitripennis]
170587927 Chromatin assembly factor 1 subunit B [Brugia malayi]
125559885 hypothetical protein OsI_026565 [Oryza sativa (indica cultivar-group)]
45190883 AER280Cp [Ashbya gossypii ATCC 10895]
194695526 unknown [Zea mays]
115474415 Os08g0108200 [Oryza sativa (japonica cultivar-group)]
150865774 hypothetical protein PICST_89595 [Pichia stipitis CBS 6054]
50308895 unnamed protein product [Kluyveromyces lactis]
198426117 PREDICTED: similar to chromatin assembly factor-1p60 [Ciona intestinalis]
25143535 Y71G12B.1a [Caenorhabditis elegans]
25143538 Y71G12B.1b [Caenorhabditis elegans]
42573790 FAS2 (FASCIATA 2); nucleotide binding [Arabidopsis thaliana]
157745187 Hypothetical protein CBG22204 [Caenorhabditis briggsae AF16]
168051044 chromatin assembly complex 1 subunit B/CAC2 [Physcomitrella patens subsp. patens]
157423647 Unknown (protein for IMAGE:8920590) [Xenopus tropicalis]
6323534 Component of the chromatin assembly complex (with Rlf2p and Msi1p) that assembles newly synthesized histones onto recently replicated DNA, required for building functional kinetochores, conserved from yeast to humans; Cac2p [Saccharomyces cerevisiae]
151946061 chromatin assembly factor-I (CAF-I) p60 subunit [Saccharomyces cerevisiae YJM789]
68489498 hypothetical protein CaO19.6670 [Candida albicans SC5314]
193690490 PREDICTED: similar to AGAP007544-PA [Acyrthosiphon pisum]
66507612 PREDICTED: similar to Caf1-105 CG12892-PA [Apis mellifera]
157353903 unnamed protein product [Vitis vinifera]
145342055 predicted protein [Ostreococcus lucimarinus CCE9901]
116061747 WD40 repeat-containing protein (ISS) [Ostreococcus tauri]
42573792 FAS2 (FASCIATA 2); nucleotide binding [Arabidopsis thaliana]
156377221 predicted protein [Nematostella vectensis]
196006409 hypothetical protein TRIADDRAFT_26010 [Trichoplax adhaerens]
18424846 FAS2 (FASCIATA 2); nucleotide binding [Arabidopsis thaliana]

58 **HirA cluster**
109093321 PREDICTED: similar to HIR (histone cell cycle regulation defective, S. cerevisiae) homolog A [Macaca mulatta]
119924678 PREDICTED: similar to HIR histone cell cycle regulation defective homolog A [Bos taurus]
2623856 HIRA homolog [Drosophila melanogaster]
198423921 PREDICTED: similar to predicted protein [Ciona intestinalis]
194896907 GG19653 [Drosophila erecta]
195480395 GE15728 [Drosophila yakuba]
195163770 GL14721 [Drosophila persimilis]
2879829 HIRA [Drosophila melanogaster]

24640390 Hira CG12153-PA [Drosophila melanogaster]
54650948 AT04626p [Drosophila melanogaster]
195353507 GM17533 [Drosophila sechellia]
195447048 GK25579 [Drosophila willistoni]
74095993 Tuple1/HirA [Takifugu rubripes]
126324878 PREDICTED: similar to HIRA [Monodelphis domestica]
22773842 HIRA [Gallus gallus]
146345432 HIRA_CHICK Protein HIRA (TUP1-like enhancer of split protein 1) (cHIRA)
198469914 GA11439 [Drosophila pseudoobscura pseudoobscura]
45385815 HIR histone cell cycle regulation defective homolog A [Gallus gallus]
927419 HIRA [Homo sapiens]
21536485 HIR histone cell cycle regulation defective homolog A [Homo sapiens]
73995866 PREDICTED: similar to HIR (histone cell cycle regulation defective, S. cerevisiae) homolog A isoform 1 [Canis familiaris]
114685075 PREDICTED: HIR (histone cell cycle regulation defective, S. cerevisiae) homolog A isoform 4 [Pan troglodytes]
1359985 HIRA protein [Mus musculus]
1771288 HIRA [Mus musculus]
52426778 histone cell cycle regulation defective homolog A [Mus musculus]
62658286 PREDICTED: similar to histone cell cycle regulation defective homolog A isoform 1 [Rattus norvegicus]
20387079 HIRA protein [Xenopus laevis]
147900031 HIRA protein [Xenopus laevis]
84620806 HIRA [Carassius auratus gibelio]
84620808 HIRA [Carassius auratus]
125817880 PREDICTED: similar to HIR histone cell cycle regulation defective homolog A [Danio rerio]
194763162 GF21157 [Drosophila ananassae]
156374281 predicted protein [Nematostella vectensis]
195398975 GJ15893 [Drosophila virilis]
195041723 GH12580 [Drosophila grimshawi]
195133300 GI16203 [Drosophila mojavensis]
158260641 unnamed protein product [Homo sapiens]
194228528 PREDICTED: HIR histone cell cycle regulation defective homolog A (S. cerevisiae) [Equus caballus]
195028660 GH21785 [Drosophila grimshawi]
170032337 histone transcription regulator [Culex quinquefasciatus]
47227669 unnamed protein product [Tetraodon nigroviridis]
193690647 PREDICTED: similar to AGAP009488-PA [Acyrthosiphon pisum]
156543858 PREDICTED: similar to ENSANGP00000010454 [Nasonia vitripennis]
110778110 PREDICTED: similar to HIRA protein homolog (dHIRA) [Apis mellifera]
158288329 AGAP009488-PA [Anopheles gambiae str. PEST]
73995870 PREDICTED: similar to HIRA protein (TUP1 like enhancer of split protein 1) isoform 3 [Canis familiaris]
114685079 PREDICTED: HIR (histone cell cycle regulation defective, S. cerevisiae) homolog A isoform 1 [Pan troglodytes]
194389634 unnamed protein product [Homo sapiens]
434983 TUP1 like enhancer of SPLIT gene 1 [Homo sapiens]
157113602 histone transcription regulator [Aedes aegypti]
194043448 PREDICTED: HIR histone cell cycle regulation defective homolog A (S. cerevisiae) [Sus scrofa]
149019815 similar to histone cell cycle regulation defective homolog A isoform 1 [Rattus norvegicus]
434993 TUP1-like enhancer of split gene 1 [Mus musculus]
148665122 histone cell cycle regulation defective homolog A (S. cerevisiae), isoform CRA_a [Mus musculus]
3746658 Hira isoform [Drosophila melanogaster]
31127112 Histone cell cycle regulation defective homolog A (S. cerevisiae) [Mus musculus]
17862164 LD11036p [Drosophila melanogaster]
196014858 hypothetical protein TRIADDRAFT_32472 [Trichoplax adhaerens]

41
119182907 hypothetical protein CIMG_06449 [Coccidioides immitis RS]
169776159 hypothetical protein [Aspergillus oryzae RIB40]
115437504 conserved hypothetical protein [Aspergillus terreus NIH2624]
62089360 WD repeat domain 26 variant [Homo sapiens]
73961497 PREDICTED: similar to WD-repeat protein 26 [Canis familiaris]
116198051 hypothetical protein CHGG_07181 [Chaetomium globosum CBS 148.51]
197245766 Unknown (protein for IMAGE:6638325) [Xenopus laevis]
119590126 WD repeat domain 26, isoform CRA_a [Homo sapiens]
145233591 hypothetical protein An02g10400 [Aspergillus niger]
134047968 WDR26_MOUSE WD repeat-containing protein 26
118087837 PREDICTED: similar to WD repeat domain 26 [Gallus gallus]
119499063 WD domain protein [Neosartorya fischeri NRRL 181]
121719107 WD domain protein [Aspergillus clavatus NRRL 1]
134035357 WDR26_XENTR WD repeat-containing protein 26
159125975 catabolite degradation protein, putative [Aspergillus fumigatus A1163]
134024531 LOC733878 protein [Xenopus tropicalis]
122936386 LOC100037096 protein [Xenopus laevis]

157358331 unnamed protein product [Vitis vinifera]
68396095 PREDICTED: hypothetical protein LOC564794 [Danio rerio]
171694049 unnamed protein product [Podospora anserina]
168035605 predicted protein [Physcomitrella patens subsp. patens]
168037590 predicted protein [Physcomitrella patens subsp. patens]
125539056 hypothetical protein OsI_006684 [Oryza sativa (indica cultivar-group)]
125581740 hypothetical protein OsJ_006154 [Oryza sativa (japonica cultivar-group)]
157350479 unnamed protein product [Vitis vinifera]
50547865 hypothetical protein [Yarrowia lipolytica]
148681181 WD repeat domain 26 [Mus musculus]
116004337 WD repeat domain 26 [Bos taurus]
126306893 PREDICTED: similar to WD repeat domain 26, [Monodelphis domestica]
157818723 WD repeat domain 26 [Rattus norvegicus]
115445651 Os02g0294600 [Oryza sativa (japonica cultivar-group)]
19114336 heterotrimeric G protein beta subunit Gnr1 [Schizosaccharomyces pombe 972h-]
149246535 PREDICTED: similar to myocardial ischemic preconditioning upregulated protein 2 isoform 1 [Mus musculus]
189529474 PREDICTED: similar to WD repeat-containing protein 26 [Danio rerio]
145602624 hypothetical protein MGG_09518 [Magnaporthe grisea 70-15]
194387698 unnamed protein product [Homo sapiens]
47205197 unnamed protein product [Tetraodon nigroviridis]
47205447 unnamed protein product [Tetraodon nigroviridis]
70985192 catabolite degradation protein [Aspergillus fumigatus Af293]
154290922 hypothetical protein BC1G_15393 [Botryotinia fuckeliana B05.10]
47208646 unnamed protein product [Tetraodon nigroviridis]

36 **Hir1 Cluster**
116202341 hypothetical protein CHGG_09055 [Chaetomium globosum CBS 148.51]
39974527 hypothetical protein MGG_00590 [Magnaporthe grisea 70-15]
121702031 histone transcription regulator Hir1, putative [Aspergillus clavatus NRRL 1]
119173681 hypothetical protein CIMG_10272 [Coccidioides immitis RS]
145239265 hypothetical protein An08g01600 [Aspergillus niger]
146324907 HIR1_YARLI Protein HIR1
169763824 hypothetical protein [Aspergillus oryzae RIB40]
46121933 hypothetical protein FG05344.1 [Gibberella zeae PH-1]
121925435 HIR1_PHANO Protein HIR1
70995162 histone transcription regulator Hir1 [Aspergillus fumigatus Af293]
154287230 conserved hypothetical protein [Ajellomyces capsulatus NAm1]
119496027 histone transcription regulator Hir1, putative [Neosartorya fischeri NRRL 181]
189189004 hypothetical protein PTRG_00508 [Pyrenophora tritici-repentis Pt-1C-BFP]
164426709 hypothetical protein NCU04035 [Neurospora crassa OR74A]
39979226 related to histone transcription regulator [Neurospora crassa]
171682236 unnamed protein product [Podospora anserina]
154323876 hypothetical protein BC1G_00337 [Botryotinia fuckeliana B05.10]
115391707 conserved hypothetical protein [Aspergillus terreus NIH2624]
50406340 hypothetical protein DEHA0A06996g [Debaryomyces hansenii CBS767]
67521658 hypothetical protein AN1286.2 [Aspergillus nidulans FGSC A4]
199428981 DEHA2A06908p [Debaryomyces hansenii]
150863699 protein involved in cell-cycle regulation of histone transcription [Pichia stipitis CBS 6054]
149248578 conserved hypothetical protein [Lodderomyces elongisporus NRRL YB-4239]
19113367 hira protein Hip1 [Schizosaccharomyces pombe]
50302561 unnamed protein product [Kluyveromyces lactis]
156056661 hypothetical protein SS1G_04061 [Sclerotinia sclerotiorum 1980]
68472703 hypothetical protein CaO19.9647 [Candida albicans SC5314]
6319463 Non-essential transcriptional corepressor involved in the cell cycle-regulated transcription of histone H2A, H2B, H3 and H4 genes; contributes to nucleosome formation, heterochromatic gene silencing, and formation of functional kinetochores; Hir1p [Saccharomyces cerevisiae]
50286763 unnamed protein product [Candida glabrata]
151946387 histone regulation protein [Saccharomyces cerevisiae YJM789]
190408834 HIR1 [Saccharomyces cerevisiae RM11-1a]
45201264 AGR168Wp [Ashbya gossypii ATCC 10895]
156846878 hypothetical protein Kpol_1032p61 [Vanderwaltozyma polyspora DSM 70294]
50553596 hypothetical protein [Yarrowia lipolytica]
146420911 hypothetical protein PGUG_02079 [Pichia guilliermondii ATCC 6260]
190345996 hypothetical protein PGUG_02079 [Pichia guilliermondii ATCC 6260]

33
149238792 WD-repeat protein pop3 [Lodderomyces elongisporus NRRL YB-4239]
164657400 hypothetical protein MGL_2812 [Malassezia globosa CBS 7966]
156043661 conserved hypothetical protein [Sclerotinia sclerotiorum 1980]

58261992 hypothetical protein CNM01130 [Cryptococcus neoformans var. neoformans JEC21]
154316869 conserved hypothetical protein [Botryotinia fuckeliana B05.10]
189196668 WD repeat containing protein pop3 [Pyrenophora tritici-repentis Pt-1C-BFP]
157110635 vegetatible incompatibility protein HET-E-1, putative [Aedes aegypti]
169604496 hypothetical protein SNOG_05260 [Phaeosphaeria nodorum SN15]
199434100 DEHA2G10824p [Debaryomyces hansenii]
169862261 hypothetical protein CC1G_06967 [Coprinopsis cinerea okayama7#130]
190348268 hypothetical protein PGUG_04790 [Pichia guilliermondii ATCC 6260]
145347992 predicted protein [Ostreococcus lucimarinus CCE9901]
118779379 AGAP001036-PA [Anopheles gambiae str. PEST]
164426421 WD-repeat protein pop3 [Neurospora crassa OR74A]
168037964 predicted protein [Physcomitrella patens subsp. patens]
116786838 unknown [Picea sitchensis]
116789501 unknown [Picea sitchensis]
125981461 GA15597 [Drosophila pseudoobscura pseudoobscura]
170054243 vegetatible incompatibility protein HET-E-1 [Culex quinquefasciatus]
195131599 GI15823 [Drosophila mojavensis]
195394005 GJ18676 [Drosophila virilis]
19112474 WD repeat protein Pop3 [Schizosaccharomyces pombe]
195046794 GH24634 [Drosophila grimshawi]
18858099 CG3004 CG3004-PA [Drosophila melanogaster]
167535194 predicted protein [Monosiga brevicollis MX1]
194890443 GG18968 [Drosophila erecta]
195350520 GM11356 [Drosophila sechellia]
195481724 GE15442 [Drosophila yakuba]
15227141 transducin family protein / WD-40 repeat family protein [Arabidopsis thaliana]
194766945 GF22378 [Drosophila ananassae]
195446886 GK25539 [Drosophila willistoni]
167381683 WD repeat-containing protein pop3, putative [Entamoeba dispar SAW760]
66801173 WD40 repeat-containing protein [Dictyostelium discoideum AX4]

30
109039317 PREDICTED: RNA, U3 small nucleolar interacting protein 2 [Macaca mulatta]
193204866 T02H6.1a [Caenorhabditis elegans]
115454195 Os03g0625900 [Oryza sativa (japonica cultivar-group)]
194704588 unknown [Zea mays]
125587176 hypothetical protein OsJ_011323 [Oryza sativa (japonica cultivar-group)]
27881844 Rnu3ip2 protein [Danio rerio]
194754140 GF12083 [Drosophila ananassae]
195154384 GL17526 [Drosophila persimilis]
195486122 GE12278 [Drosophila yakuba]
62471723 CG33505 CG33505-PA [Drosophila melanogaster]
114587142 PREDICTED: RNA, U3 small nucleolar interacting protein 2 isoform 2 [Pan troglodytes]
195334489 GM20172 [Drosophila sechellia]
195056690 GH22986 [Drosophila grimshawi]
195583512 GD25649 [Drosophila simulans]
157817630 RRP9, small subunit (SSU) processome component, homolog [Rattus norvegicus]
195436252 GK22122 [Drosophila willistoni]
163915599 Unknown (protein for IMAGE:5572436) [Xenopus laevis]
4759276 RNA, U3 small nucleolar interacting protein 2 [Homo sapiens]
21704248 U3 snoRNP-associated protein [Mus musculus]
72063080 PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]
114587140 PREDICTED: RNA, U3 small nucleolar interacting protein 2 isoform 1 [Pan troglodytes]
169646808 RNA, U3 small nucleolar interacting protein 2 [Danio rerio]
125544932 hypothetical protein OsI_012304 [Oryza sativa (indica cultivar-group)]
118096677 PREDICTED: similar to U3 snoRNP associated 55 kDa protein [Gallus gallus]
148689203 RNA, U3 small nucleolar interacting protein 2, isoform CRA_a [Mus musculus]
149018649 RNA, U3 small nucleolar interacting protein 2 (predicted), isoform CRA_b [Rattus norvegicus]
168013375 predicted protein [Physcomitrella patens subsp. patens]
198426332 PREDICTED: similar to U3 small nucleolar RNA-interacting protein 2 (U3 small nucleolar ribonucleoprotein-associated 55 kDa protein) (U3 snoRNP-associated 55 kDa protein) (U3-55K) (RRP9 homolog) [Ciona intestinalis]
196006357 hypothetical protein TRIADDRAFT_26083 [Trichoplax adhaerens]
168048765 predicted protein [Physcomitrella patens subsp. patens]

24 **Cac2 Cluster**
58259203 chromatin assembly complex protein [Cryptococcus neoformans var. neoformans JEC21]
189202282 chromatin assembly factor 1 protein [Pyrenophora tritici-repentis Pt-1C-BFP]
154321710 hypothetical protein BC1G_01002 [Botryotinia fuckeliana B05.10]
154272067 conserved hypothetical protein [Ajellomyces capsulatus NAm1]

119480017 chromatin assembly factor 1 subunit B, putative [Neosartorya fischeri NRRL 181]
119178430 hypothetical protein CIMG_08051 [Coccidioides immitis RS]
121710328 chromatin assembly factor 1 subunit B, putative [Aspergillus clavatus NRRL 1]
145238600 hypothetical protein An07g08730 [Aspergillus niger]
171695366 unnamed protein product [Podospora anserina]
116182316 hypothetical protein CHGG_01786 [Chaetomium globosum CBS 148.51]
169767950 hypothetical protein [Aspergillus oryzae RIB40]
156045954 hypothetical protein SS1G_09253 [Sclerotinia sclerotiorum 1980]
46107568 hypothetical protein FG00667.1 [Gibberella zeae PH-1]
50546881 hypothetical protein [Yarrowia lipolytica]
67526655 hypothetical protein AN3785.2 [Aspergillus nidulans FGSC A4]
164424885 hypothetical protein NCU08357 [Neurospora crassa OR74A]
70989371 chromatin assembly factor 1 subunit B [Aspergillus fumigatus Af293]
159128946 chromatin assembly factor 1 subunit B, putative [Aspergillus fumigatus A1163]
145615874 hypothetical protein MGG_03737 [Magnaporthe grisea 70-15]
50425165 hypothetical protein DEHA0F20328g [Debaryomyces hansenii CBS767]
164661673 hypothetical protein MGL_1227 [Malassezia globosa CBS 7966]
19115362 WD repeat protein Cac2 [Schizosaccharomyces pombe]
115390669 conserved hypothetical protein [Aspergillus terreus NIH2624]
169595082 hypothetical protein SNOG_00274 [Phaeosphaeria nodorum SN15]

20
118383986 hypothetical protein TTHERM_00684450 [Tetrahymena thermophila SB210]
17559120 Temporarily Assigned Gene name family member (tag-135) [Caenorhabditis elegans]
157763132 Hypothetical protein CBG09714 [Caenorhabditis briggsae AF16]
170584821 Pre-mRNA splicing protein prp5 [Brugia malayi]
18859793 Transport and Golgi organization 4 CG1796-PA [Drosophila melanogaster]
195356101 GM13251 [Drosophila sechellia]
195566355 GD17060 [Drosophila simulans]
194768072 GF19378 [Drosophila ananassae]
194896028 GG17704 [Drosophila erecta]
195469864 GE16491 [Drosophila yakuba]
195132835 GI21768 [Drosophila mojavensis]
157126630 striatin, putative [Aedes aegypti]
195048288 GH24167 [Drosophila grimshawi]
125981355 GA14743 [Drosophila pseudoobscura pseudoobscura]
195448366 GK10085 [Drosophila willistoni]
195403107 GJ18540 [Drosophila virilis]
193662061 PREDICTED: similar to pleiotropic regulator 1 [Acyrthosiphon pisum]
66529875 PREDICTED: similar to CG1796-PA [Apis mellifera]
158294509 AGAP005631-PA [Anopheles gambiae str. PEST]
159480808 predicted protein [Chlamydomonas reinhardtii]

18
149412478 PREDICTED: similar to F-box and WD-40 domain protein 11 [Ornithorhynchus anatinus]
126291455 PREDICTED: similar to F-box and WD-40 domain protein 11 [Monodelphis domestica]
23956270 F-box and WD-40 domain protein 11 [Mus musculus]
73953434 PREDICTED: similar to F-box and WD-40 domain protein 11 isoform 2 [Canis familiaris]
114603476 PREDICTED: similar to F-box and WD-40 domain protein 11 [Pan troglodytes]
26006203 mKIAA0696 protein [Mus musculus]
3327206 KIAA0696 protein [Homo sapiens]
56207156 F-box and WD-40 domain protein 11 [Mus musculus]
66542058 PREDICTED: similar to supernumerary limbs CG3412-PA [Apis mellifera]
158301421 AGAP001944-PA [Anopheles gambiae str. PEST]
26334549 unnamed protein product [Mus musculus]
48928046 F-box and WD repeat domain containing 11 isoform B [Homo sapiens]
74147468 unnamed protein product [Mus musculus]
86129432 F-box and WD repeat domain containing 11 [Gallus gallus]
194219605 PREDICTED: F-box and WD repeat domain containing 11 [Equus caballus]
41152004 F-box and WD-40 domain protein 11b [Danio rerio]
48928048 F-box and WD repeat domain containing 11 isoform A [Homo sapiens]
13445757 AF339101_1 beta-transducin repeat-containing protein [Heterodera glycines]

15
12963745 spermatid WD-repeat protein [Mus musculus]
59938780 WD repeat domain 31 isoform 1 [Homo sapiens]
109110524 PREDICTED: WD repeat domain 31 isoform 3 [Macaca mulatta]
164663893 WD repeat domain 31 [Rattus norvegicus]
12856409 unnamed protein product [Mus musculus]

21687048 WD repeat domain 31 isoform 3 [Homo sapiens]
73971980 PREDICTED: similar to WD-repeat protein 31 [Canis familiaris]
114626236 PREDICTED: WD repeat domain 31 isoform 2 [Pan troglodytes]
194669728 PREDICTED: similar to WD repeat domain 31 [Bos taurus]
149738634 PREDICTED: similar to WD repeat domain 31 [Equus caballus]
118099166 PREDICTED: similar to WDR31 protein [Gallus gallus]
50925847 WD repeat domain 31 [Rattus norvegicus]
114626238 PREDICTED: similar to WDR31 protein isoform 1 [Pan troglodytes]
148699196 WD repeat domain 31, isoform CRA_b [Mus musculus]
149059621 WD repeat domain 31, isoform CRA_b [Rattus norvegicus]

7 **HirA cluster**
186510672 HIRA (ARABIDOPSIS HOMOLOG OF HISTONE CHAPERONE HIRA); nucleotide binding / protein binding [Arabidopsis thaliana]
75335621 HIRA_ARATH Protein HIRA (Histone regulator protein)
157328608 unnamed protein product [Vitis vinifera]
168019698 histone transcription regulator HIRA [Physcomitrella patens subsp. patens]
115480741 Os09g0567700 [Oryza sativa (japonica cultivar-group)]
168046781 histone transcription regulator HIRA [Physcomitrella patens subsp. patens]
162463112 HIRA [Zea mays]

7
193669465 PREDICTED: similar to AGAP006113-PA [Acyrthosiphon pisum]
195382503 GJ20455 [Drosophila virilis]
194754309 GF12876 [Drosophila ananassae]
195149383 GL11179 [Drosophila persimilis]
198456160 GA15470 [Drosophila pseudoobscura pseudoobscura]
58388279 AGAP006113-PA [Anopheles gambiae str. PEST]
156538931 PREDICTED: similar to conserved hypothetical protein, partial [Nasonia vitripennis]

6
146417979 hypothetical protein PGUG_02685 [Pichia guilliermondii ATCC 6260]
190346489 hypothetical protein PGUG_02685 [Pichia guilliermondii ATCC 6260]
199434022 DEHA2G08184p [Debaryomyces hansenii]
68470454 hypothetical protein CaO19.11771 [Candida albicans SC5314]
50426615 hypothetical protein DEHA0G08833g [Debaryomyces hansenii CBS767]
150865807 Histone transcription regulator HIRA, WD repeat superfamily [Pichia stipitis CBS 6054]

6
72389570 hypothetical protein, conserved [Trypanosoma brucei TREU927]
71410693 hypothetical protein Tc00.1047053509985.20 [Trypanosoma cruzi strain CL Brener]
71424492 hypothetical protein [Trypanosoma cruzi strain CL Brener]
154343664 hypothetical protein LbrM33_V2.0060 [Leishmania braziliensis MHOM/BR/75/M2904]
146097407 hypothetical protein [Leishmania infantum]
157874625 hypothetical protein, conserved [Leishmania major]

5
48122105 PREDICTED: similar to CG3957-PA [Apis mellifera]
45550397 CG10459 CG10459-PA [Drosophila melanogaster]
195332895 GM21146 [Drosophila sechellia]
195581962 GD10677 [Drosophila simulans]
194881113 GG21900 [Drosophila erecta]

4
189195658 WD repeat containing protein 61 [Pyrenophora tritici-repentis Pt-1C-BFP]
169604274 hypothetical protein SNOG_05148 [Phaeosphaeria nodorum SN15]
145239011 hypothetical protein An08g00340 [Aspergillus niger]
67521860 hypothetical protein AN1387.2 [Aspergillus nidulans FGSC A4]

3
169864819 hypothetical protein CC1G_11339 [Coprinopsis cinerea okayama7#130]
58270988 transcription corepressor [Cryptococcus neoformans var. neoformans JEC21]
58270986 transcription corepressor [Cryptococcus neoformans var. neoformans JEC21]

3
145483475 hypothetical protein GSPATT00005111001 [Paramecium tetraurelia strain d4-2]
145552958 hypothetical protein GSPATT00027105001 [Paramecium tetraurelia strain d4-2]
118394837 hypothetical protein TTHERM_01308010 [Tetrahymena thermophila SB210]

3
171693507 unnamed protein product [Podospora anserina]
154299947 hypothetical protein BC1G_10864 [Botryotinia fuckeliana B05.10]
156053568 hypothetical protein SS1G_05631 [Sclerotinia sclerotiorum 1980]

3
118374853 hypothetical protein TTHERM_00219420 [Tetrahymena thermophila SB210]
145543280 hypothetical protein GSPATT00023102001 [Paramecium tetraurelia strain d4-2]
145550431 hypothetical protein GSPATT00025841001 [Paramecium tetraurelia strain d4-2]

3
67471866 WD domain containing protein [Entamoeba histolytica HM-1:IMSS]
167393569 WD repeat-containing protein, putative [Entamoeba dispar SAW760]
167394403 guanine nucleotide-binding protein subunit beta-1, putative [Entamoeba dispar SAW760]

3
58332006 SEH1-like (S. cerevisiae [Xenopus tropicalis]
148223623 hypothetical protein LOC734650 [Xenopus laevis]
170109879 predicted protein [Laccaria bicolor S238N-H82]

3
145611116 hypothetical protein MGG_11532 [Magnaporthe grisea 70-15]
121706478 nuclear pore complex subunit (SEC13), putative [Aspergillus clavatus NRRL 1]
116202287 conserved hypothetical protein [Chaetomium globosum CBS 148.51]

3
71400433 hypothetical protein Tc00.1047053511897.28 [Trypanosoma cruzi strain CL Brener]
68477267 hypothetical protein CaO19.11260 [Candida albicans SC5314]
149053642 notchless homolog 1 (Drosophila) (predicted) [Rattus norvegicus]
2
20810487 WDSUB1 protein [Homo sapiens]
119631815 WD repeat, sterile alpha motif and U-box domain containing 1, isoform CRA_b [Homo sapiens]

2
168031665 predicted protein [Physcomitrella patens subsp. patens]
168001685 predicted protein [Physcomitrella patens subsp. patens]

2
145497575 hypothetical protein GSPATT00006685001 [Paramecium tetraurelia strain d4-2]
145527384 hypothetical protein GSPATT00002224001 [Paramecium tetraurelia strain d4-2]

2
145520643 hypothetical protein GSPATT00001821001 [Paramecium tetraurelia strain d4-2]
145529948 hypothetical protein GSPATT00017704001 [Paramecium tetraurelia strain d4-2]

2
149023215 WD repeat domain 39, isoform CRA_d [Rattus norvegicus]
33390985 WD40 protein Ciao1-like protein [Crassostrea gigas]

1
66807047 transcription initiation factor TFIID subunit [Dictyostelium discoideum AX4]
169862669 hypothetical protein CC1G_09511 [Coprinopsis cinerea okayama7#130]
145349146 predicted protein [Ostreococcus lucimarinus CCE9901]
164655237 hypothetical protein MGL_4084 [Malassezia globosa CBS 7966]
170591626 HIRA protein. [Brugia malayi]
71021701 hypothetical protein UM04934.1 [Ustilago maydis 521]
164662569 hypothetical protein MGL_0181 [Malassezia globosa CBS 7966]
156548996 PREDICTED: similar to MGC130867 protein [Nasonia vitripennis]
167538599 predicted protein [Monosiga brevicollis MX1]
85001415 wd40-repeat protein, putative [Theileria annulata]
159482424 hypothetical protein CHLREDRAFT_132099 [Chlamydomonas reinhardtii]
170587909 F-box domain containing protein [Brugia malayi]
123454327 WD repeat protein, putative [Trichomonas vaginalis G3]
158294415 AGAP005577-PA [Anopheles gambiae str. PEST]
3123027 CORO7_RAT Coronin-7 (Crn7) (70 kDa WD repeat tumor rejection antigen)
126649177 WD-40 repeat protein family / small nuclear ribonucleoprotein Prp4p-related [Cryptosporidium parvum Iowa II]
167522375 predicted protein [Monosiga brevicollis MX1]
58266512 chromatin binding protein [Cryptococcus neoformans var. neoformans JEC21]
123424969 hypothetical protein TVAG_286080 [Trichomonas vaginalis G3]

157358738 unnamed protein product [Vitis vinifera]
145518786 hypothetical protein GSPATT00013334001 [Paramecium tetraurelia strain d4-2]
115482406 Os10g0465000 [Oryza sativa (japonica cultivar-group)]
76154626 SJCHGC06036 protein [Schistosoma japonicum]
189242180 PREDICTED: similar to histone cell cycle regulation defective homolog A (S. cerevisiae), partial [Tribolium castaneum]
116058285 Beta-transducin family (WD-40 repeat) protein (ISS) [Ostreococcus tauri]
168063273 predicted protein [Physcomitrella patens subsp. patens]
47087648 chromatin assembly factor 1, subunit B [Danio rerio]
114593752 PREDICTED: similar to Chain A, Structure Of Wdr5 [Pan troglodytes]
149555875 PREDICTED: hypothetical protein [Ornithorhynchus anatinus]
195063046 GH25103 [Drosophila grimshawi]
56752743 SJCHGC09299 protein [Schistosoma japonicum]
76154308 SJCHGC06035 protein [Schistosoma japonicum]
159489194 predicted protein [Chlamydomonas reinhardtii]
68069283 guanine nucleotide-binding protein [Plasmodium berghei strain ANKA]
70724647 G-protein beta subunit [Cryptococcus gattii]