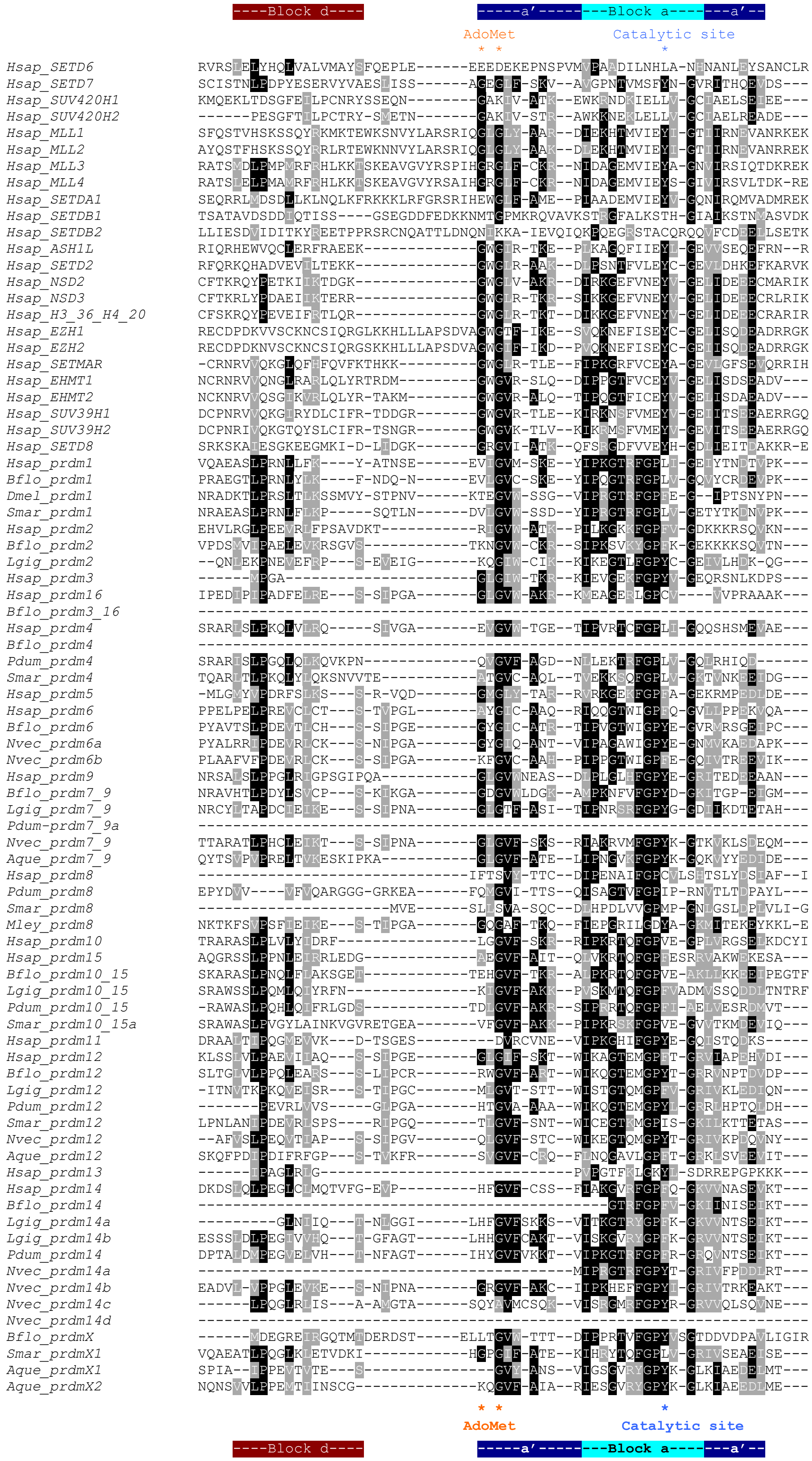
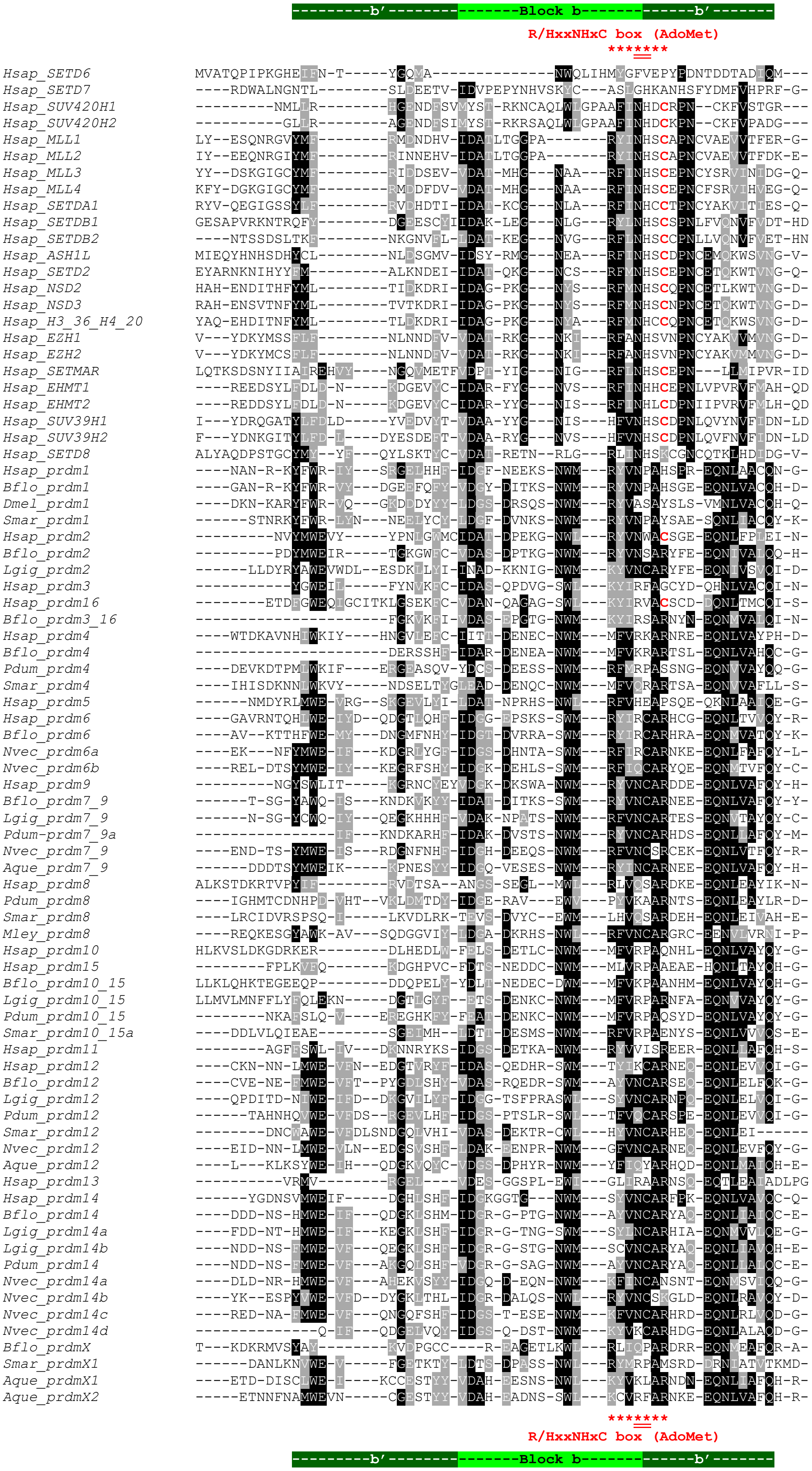
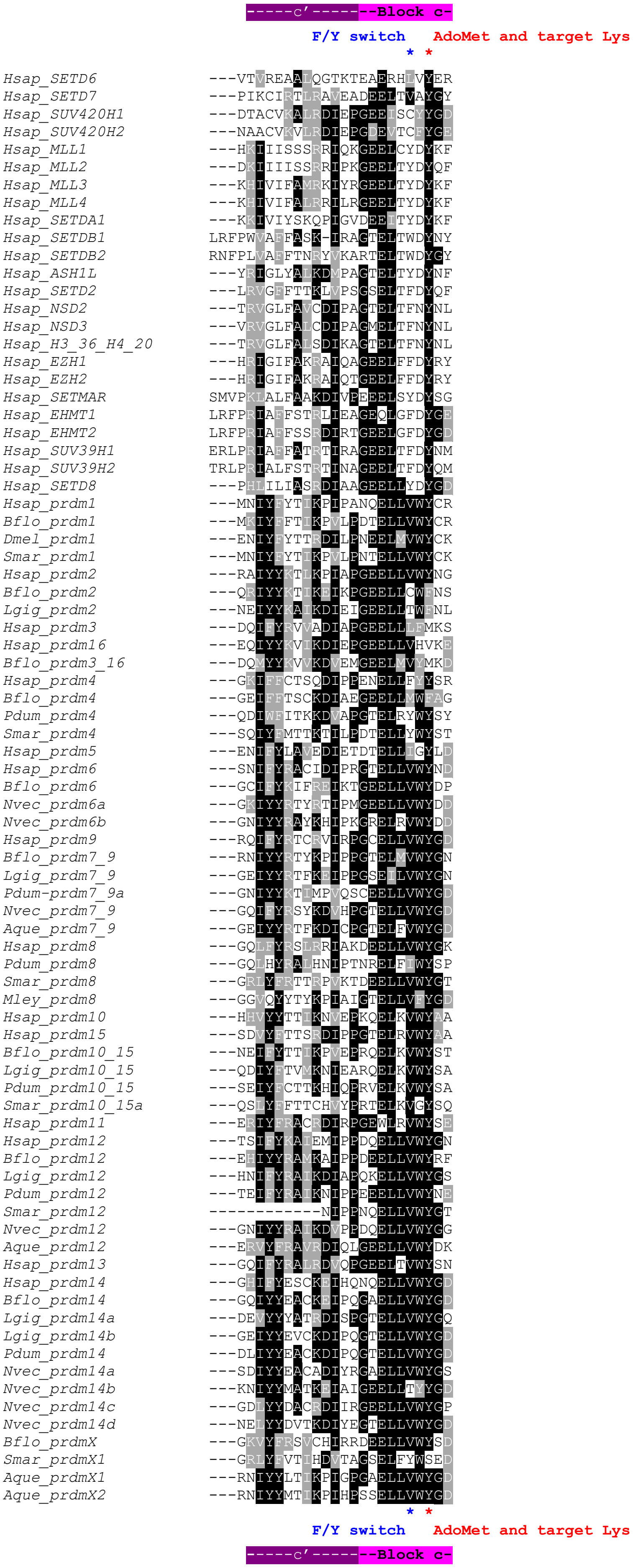


```
                         * *                  *
Hsap_SETD6        RVRSLELYHQLVALVMAYSFQEPLE------EEEDEKEPNSPVMVPAADILNHLA-NHNANLEYSANCLR
Hsap_SETD7        SCISTNLPDPYESERVYVAESLISS------AGEGLF-SKV--AVGPNTVMSFYN-GVRITHQEVDS---
Hsap_SUV420H1     KMQEKLTDSGFEILPCNRYSSEQN--------GAKIV-ATK--EWKRNDKIELLV-GCIAELSEIEE---
Hsap_SUV420H2     ------PESGFTILPCTRY-SMETN-------GAKIV-STR--AWKKNEKLELLV-GCIAELREADE---
Hsap_MLL1         SFQSTVHSKSSQYRKMKTEWKSNVYLARSRIQGLGLY-AAR--DIEKHTMVIEYI-GTIIRNEVANRKEK
Hsap_MLL2         AYQSTFHSKSSQYRRLRTEWKNNVYLARSRIQGLGLY-AAK--DLEKHTMVIEYI-GTIIRNEVANRREK
Hsap_MLL3         RATSMDLPMPMRFRHLKKTSKEAVGVYRSPIHGRGLF-CKR--NIDAGEMVIEYA-GNVIRSIQTDKREK
Hsap_MLL4         RATSLELPMAMRFRHLKKTSKEAVGVYRSAIHGRGLF-CKR--NIDAGEMVIEYS-GIVIRSVLTDK-RE
Hsap_SETDA1       SEQRRLMDSDLLKLNQLKFRKKKLRFGRSRIHEWGLF-AME--PIAADEMVIEYV-GQNIRQMVADMREK
Hsap_SETDB1       TSATAVDSDDIQTISS----GSEGDDFEDKKNMTGPMKRQVAVKSTRGFALKSTH-GIAIKSTNMASVDK
Hsap_SETDB2       LLIESDVIDITKYREETPPRSRCNQATTLDNQNIKKA-IEVQIQKPQEGRSTACQRQQVFCDEELLSETK
Hsap_ASH1L        RIQRHEWVQCLERFRAEEK-------------GWGIR-TKE--PLKAGQFIIEYL-GEVVSEQEFRN--R
Hsap_SETD2        RFQRKQHADVEVILTEKK--------------GWGLR-AAK--DLPSNTFVLEYC-GEVLDHKEFKARVK
Hsap_NSD2         CFTKRQYPETKIIKTDGK--------------GWGLV-AKR--DIRKGEFVNEYV-GELIDEEECMARIK
Hsap_NSD3         CFTKRLYPDAEIIKTERR--------------GWGLR-TKR--SIKKGEFVNEYV-GELIDEEECRLRIK
Hsap_H3_36_H4_20  CFSKRQYPEVEIFRTLQR--------------GWGLR-TKT--DIKKGEFVNEYV-GELIDEEECRARIR
Hsap_EZH1         RECDPDKVVSCKNCSIQRGLKKHLLLAPSDVAGWGTF-IKE--SVQKNEFISEYC-GELISQDEADRRGK
Hsap_EZH2         RECDPDKNVSCKNCSIQRGSKKHLLLAPSDVAGWGIF-IKD--PVQKNEFISEYC-GEIISQDEADRRGK
Hsap_SETMAR       -CRNRVVQKGLQFHFQVFKTHHK---------GWGLR-TLE--FIPKGRFVCEYA-GEVLGFSEVQRRIH
Hsap_EHMT1        NCRNRVVQNGLRARLQLYRTRDM---------GWGVR-SLQ--DIPPGTFVCEYV-GELISDSEADV---
Hsap_EHMT2        NCKNRVVQSGIKVRLQLYR-TAKM--------GWGVR-ALQ--TIPQGTFICEYV-GELISDAEADV---
Hsap_SUV39H1      DCPNRVVQKGIRYDLCIFR-TDDGR-------GWGVR-TLE--KIRKNSFVMEYV-GEIITSEEAERRGQ
Hsap_SUV39H2      DCPNRIVQKGTQYSLCIFR-TSNGR-------GWGVK-TLV--KIKRMSFVMEYV-GEVITSEEAERRGQ
Hsap_SETD8        SRKSKAIESGKEEGMKI-D-LIDGK-------GRGVI-ATK--QFSRGDFVVEYH-GDLIEITDAKKR-E
Hsap_prdm1        VQAEASLPRNLLFK----Y-ATNSE------EVIGVM-SKE--YIPKGTRFGPLI-GEIYTNDTVPK---
Bflo_prdm1        PRAEGTLPRNLYLK----F-NDQ-N------EVLGVC-SKE--YIPQGTRFGPLV-GQVYCRDEVPK---
Dmel_prdm1        NRADKTLPRSLTLKSSMVY-STPNV------KTEGVW-SSG--VIPRGTRFGPFE-G--IPTSNYPN---
Smar_prdm1        NRAEASLPRNLFLKP-----SQTLN------DVLGVW-SSD--YIPRGTRFGPLV-GETYTKDNVPK---
Hsap_prdm2        EHVLRGLPEEVRLFPSAVDKT-----------RIGVW-ATK--PILKGKKFGPFV-GDKKKRSQVKN---
Bflo_prdm2        VPDSMVIPAELEVKRSGVS------------TKNGVW-CKR--SIPKSVKYGPFK-GEKKKKSQVTN---
Lgig_prdm2        --QNLEKPNEVEFRP---S-EVEIG-------KQGIW-CIK--KIKEGTLFGPYC-GEIVLHDK-QG---
Hsap_prdm3        ------MPGA----------------------GLGIW-TKR--KIEVGEKFGPYV-GEQRSNLKDPS---
Hsap_prdm16       IPEDIPIPADFELRE---S-SIPGA-------GLGVW-AKR--KMEAGERLGPCV----VVPRAAAK---
Bflo_prdm3_16     ----------------------------------------------------------------------
Hsap_prdm4        SRARLSLPKQLVLRQ-----SIVGA-------EVGVW-TGE--TIPVRTCFGPLI-GQQSHSMEVAE---
Bflo_prdm4        ----------------------------------------------------------------------
Pdum_prdm4        SRARISLPGQLQLKQVKPN-------------QVGVF-AGD--NLLEKTRFGPLV-GQLRHIQD------
Smar_prdm4        TQARLTLPKQLYLQKSNVVTE-----------ATGVC-AQL--TVEKKSQFGPLV-GKTVNKEEIDG---
Hsap_prdm5        -MLGMYVPDRFSLKS---S-R-VQD-------GMGLY-TAR--RVRKGEKFGPFA-GEKRMPEDLDE---
Hsap_prdm6        PPELPELPREVCLCT---S-TVPGL-------AYGIC-AAQ--RIQQGTWIGPFQ-GVLLPPEKVQA---
Bflo_prdm6        PYAVTSLPDEVTLCH---S-SIPGE-------GYGIC-ATR--TIPVGTWIGPYE-GVRMRSGEIPC---
Nvec_prdm6a       PYALRRIPDEVRLCK---S-NIPGA-------GYGIQ-ANT--VIPAGAWIGPYE-GNMVKAEDAPK---
Nvec_prdm6b       PLAAFVFPDEVRLCK---S-SIPGA-------KFGVC-AAH--PIPPGTWIGPFE-GQIVTREEVIK---
Hsap_prdm9        NRSALSLPPGLRIGPSGIPQA-----------GLGVWNEAS--DLPLGLHFGPYE-GRITEDEEAAN---
Bflo_prdm7_9      NRAVHTLPDYLSVCP---S-KIKGA-------GDGVWLDGK--AMPKNFVFGPYD-GKITGP-EIGM---
Lgig_prdm7_9      NRCYLTAPDCIEIKE---S-SIPNA-------GLGTF-ASI--TIPNRSRFGPYG-GDIIKDTETAH---
Pdum-prdm7_9a     ----------------------------------------------------------------------
Nvec_prdm7_9      TTARATLPHCLEIKT---S-SIPNA-------GLGVF-SKS--RIAKRVMFGPYK-GTKVKLSDEQM---
Aque_prdm7_9      QYTSVPVPRELTVKESKIPKA-----------GLGVF-ATE--LIPNGVKFGPYK-GQKVYYEDIDE---
Hsap_prdm8        -------------------------------IFTSVY-TTC--DIPENAIFGPCVLSHTSLYDSIAF--I
Pdum_prdm8        EPYDVV----VFVQARGGG-GRKEA------FQMGVI-TTS--QISAGTVFGPIP-RNVTLTDPAYL---
Smar_prdm8        ------------------------MVE------SLLSVA-SQC--DLHPDLVVGPMP-GNLGSLDPLVLI-G
Mley_prdm8        NKTKFSVPSFIEIKE---S-TIPGA-------GQGAF-TKQ--FIEPGRILGDYA-GKMITEKEYKKL-E
Hsap_prdm10       TRARASLPLVLYIDRF----------------LGGVF-SKR--RIPKRTQFGPVE-GPLVRGSELKDCYI
Hsap_prdm15       AQGRSSLPPNLEIRRLEDG-------------AEGVF-AIT--QLVKRTQFGPFESRRVAKWEKESA---
Bflo_prdm10_15    SKARASLPNQLFLAKSGET------------TEHGVF-TKR--ALPKRTQFGPVE-AKLLKKEEIPEGTF
Lgig_prdm10_15    SRAWSSLPQMLQIYRFN---------------KIGVF-AKK--PVSKMTQFGPFVADMVSSQDDLTNTRF
Pdum_prdm10_15    -RAWASLPQHLQIFRLGDS------------TDLGVF-AKR--SIPRRTQFGPFI-AELVESRDMVT---
Smar_prdm10_15a   SRAWASLPVGYLAINKVGVRETGEA-------VFGVF-AKK--PIPKRSKFGPVE-GVVTKMDEVIQ---
Hsap_prdm11       DRAALTIPQGMEVVK---D-TSGES---------DVRCVNE--VIPKGHIFGPYE-GQISTQDKS-----
Hsap_prdm12       KLSSLVLPAEVIIAQ---S-SIPGE-------GLGIF-SKT--WIKAGTEMGPFT-GRVIAPEHVDI---
Bflo_prdm12       SLTGLVLPPQLEARS---S-LIPCR-------RWGVF-ART--WIKQGTEMGPYT-GRRVNPTDVDP---
Lgig_prdm12       -ITNVTKPKQVEISR---S-TIPGC-------MLGVT-STT--WISTGTQMGPFV-GRIVKLEDIQN---
Pdum_prdm12       -------PEVRLVVS-----GLPGA-------HTGVA-AAA--WIKQGTEMGPYL-GRRLHPTQLDH---
Smar_prdm12       LPNLANIPDEVRLSPS----RIPGQ-------TLGVF-SNT--WICEGTKMGPIS-GKILKTTETAS---
Nvec_prdm12       --AFVSLPEQVTIAP---S-SIPGV-------QLGVF-STC--WIKEGTQMGPYT-GRIVKPDQVNY---
Aque_prdm12       SKQFPDIPDIFRFGP---S-TVKFR-------SVGVF-CRQ--FLNQGAVLGPFT-GRKLSVEEVIT---
Hsap_prdm13       ------IPAGLRLG-----------------------------PVPGTFKLGKYL-SDRREPGPKKK---
Hsap_prdm14       DKDSLQLPEGLCLMQTVFG-EVP---------HFGVF-CSS--FIAKGVRFGPFQ-GKVVNASEVKT---
Bflo_prdm14       -----------------------------------------------GTRFGPFV-GKIINISEIKT---
Lgig_prdm14a      ---------GLNIIQ---T-NLGGI------LHFGVFSKKS--VITKGTRYGPFK-GKVVNTSEIKT---
Lgig_prdm14b      ESSSLDLPEGIVVHQ---T-GFAGT------LHHGVFCAKT--VISKGVRYGPFK-GRVVNTSEIKT---
Pdum_prdm14       DPTALDMPEGVELVH---T-NFAGT------IHYGVFVKKT--VIPKGTRFGPFR-GRQVNTSEIKT---
Nvec_prdm14a      -------------------------------------------MIPRGTRFGPYT-GRIVFPDDLRT---
Nvec_prdm14b      EADVL-VPPGLEVKE---S-NIPNA-------GRGVF-AKC--IIPKHEFFGPYI-GRIVTRKEAKT---
Nvec_prdm14c      ------LPQGLRLIS---A-AMGTA------SQYAVMCSQK--VISRGMRFGPYR-GRVVQLSQVNE---
Nvec_prdm14d      ----------------------------------------------------------------------
Bflo_prdmX        ------MDEGREIRGQTMTDERDST-----ELLTGVW-TTT--DIPPRTVFGPYVSGTDDVDPAVLIGIR
Smar_prdmX1       VQAEATLPQGLKLETVDKI------------HGPGIF-ATE--KIHRYTQFGPLV-GRIVSEAEISE---
Aque_prdmX1       SPIA--IPPEVTVTE---S---------------GVY-ANS--VIGSGVRYGPYK-GLKIAEDELMT---
Aque_prdmX2       NQNSVVLPPEMTIINSCG--------------KQGVF-AIA--RIESGVRYGPYK-GLKIAEEDLME---
                         * *                  *
```

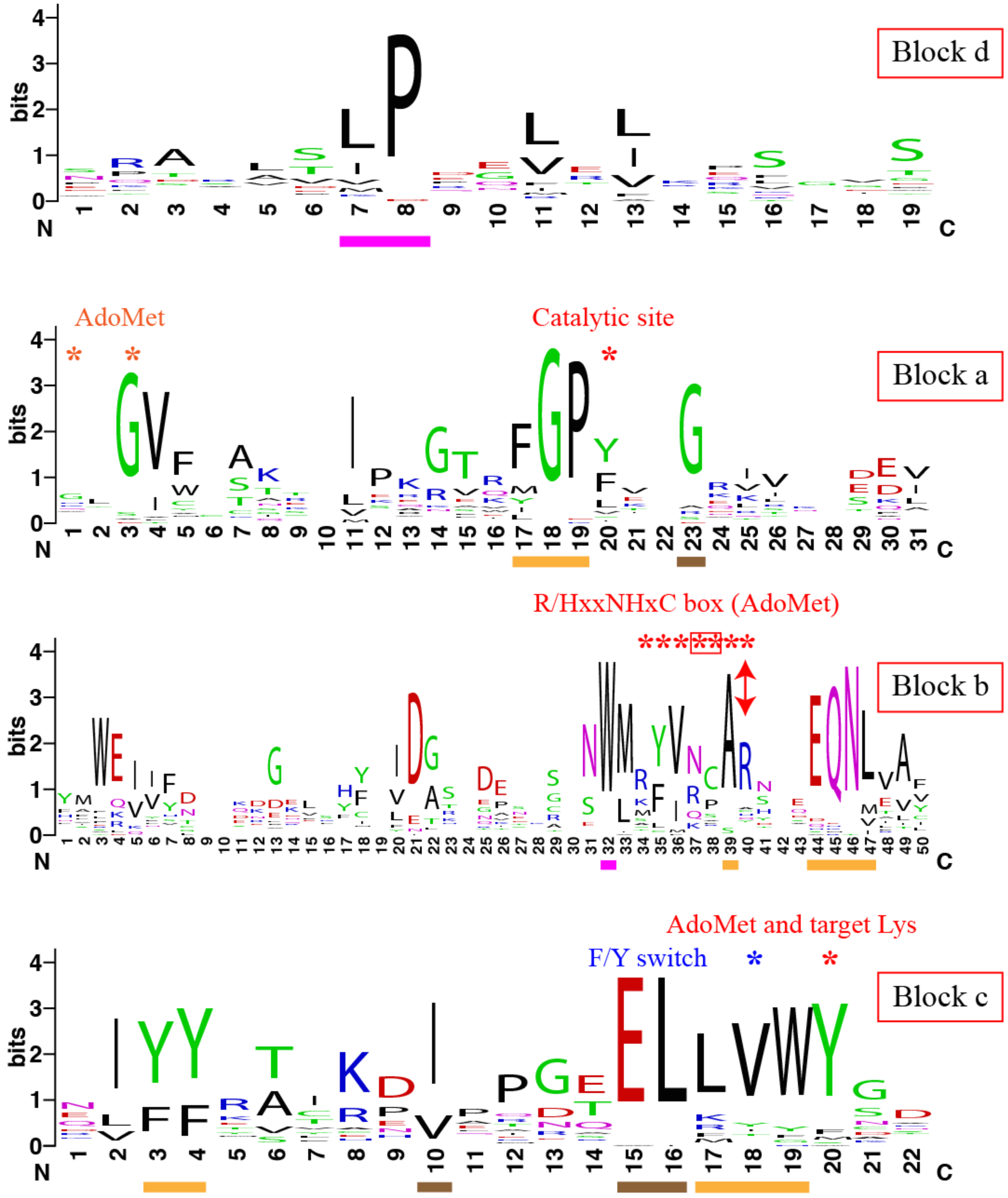


```
                                      ---------b′--------------Block b--------------b′-------
                                                               R/HxxNHxC box (AdoMet)
                                                                        *******

Hsap_SETD6              MVATQPIPKGHEIFN-T-----YGQMA--------------NWQLIHMYGFVEPYPDNTDDTADIQM---
Hsap_SETD7              ----RDWALNGNTL--------SLDEETV-IDVPEPYNHVSKYC---ASLGHKANHSFYDMFVHPRF-G-
Hsap_SUV420H1           ---------NMLLR--------HGENDFSVMYST-RKNCAQLWLGPAAFINHDCRPN--CKFVSTGR---
Hsap_SUV420H2           ----------GLLR--------AGENDFSIMYST-RKRSAQLWLGPAAFINHDCKPN--CKFVPADG---
Hsap_MLL1               LY--ESQNRGVYMF--------RMDNDHV-IDATLTGGPA-------RYINHSCAPNCVAEVVTFER-G-
Hsap_MLL2               IY--EEQNRGIYMF--------RINNEHV-IDATLTGGPA-------RYINHSCAPNCVAEVVTFDK-E-
Hsap_MLL3               YY--DSKGIGCYMF--------RIDDSEV-VDAT-MHG---NAA---RFINHSCEPNCYSRVINIDG-Q-
Hsap_MLL4               KFY-DGKGIGCYMF--------RMDDFDV-VDAT-MHG---NAA---RFINHSCEPNCFSRVIHVEG-Q-
Hsap_SETDA1             RYV-QEGIGSSYLF--------RVDHDTI-IDAT-KCG---NLA---RFINHCCTPNCYAKVITIES-Q-
Hsap_SETDB1             GESAPVRKNTRQFY--------DGEESCYIIDAK-LEG---NLG---RYLNHSCSPNLFVQNVFVDT-HD
Hsap_SETDB2             ----NTSSDSLTKF--------NKGNVFL-LDAT-KEG---NVG---RFLNHSCCPNLLVQNVFVET-HN
Hsap_ASH1L              MIEQYHNHSDHYCL--------NLDSGMV-IDSY-RMG---NEA---RFINHSCDPNCEMQKWSVNG-V-
Hsap_SETD2              EYARNKNIHYYFM---------ALKNDEI-IDAT-QKG---NCS---RFMNHSCEPNCETQKWTVNG-Q-
Hsap_NSD2               HAH-ENDITHFYML--------TIDKDRI-IDAG-PKG---NYS---RFMNHSCQPNCETLKWTVNG-D-
Hsap_NSD3               RAH-ENSVTNFYML--------TVTKDRI-IDAG-PKG---NYS---RFMNHSCNPNCETQKWTVNG-D-
Hsap_H3_36_H4_20        YAQ-EHDITNFYML--------TLDKDRI-IDAG-PKG---NYA---RFMNHCCQPNCETQKWSVNG-D-
Hsap_EZH1               V---YDKYMSSFLF--------NLNNDFV-VDAT-RKG---NKI---RFANHSVNPNCYAKVVMVNG-D-
Hsap_EZH2               V---YDKYMCSFLF--------NLNNDFV-VDAT-RKG---NKI---RFANHSVNPNCYAKVMMVNG-D-
Hsap_SETMAR             LQTKSDSNYIIAIREHVY----NGQVMETFVDPT-YIG---NIG---RFLNHSCEPN---LLMIPVR-ID
Hsap_EHMT1              ----REEDSYLFDLD-N-----KDGEVYC-IDAR-FYG---NVS---RFINHHCEPNLVPVRVFMAH-QD
Hsap_EHMT2              ----REDDSYLFDLD-N-----KDGEVYC-IDAR-YYG---NIS---RFINHLCDPNIIPVRVFMLH-QD
Hsap_SUV39H1            I---YDRQGATYLFDLD-----YVEDVYT-VDAA-YYG---NIS---HFVNHSCDPNLQVYNVFIDN-LD
Hsap_SUV39H2            F---YDNKGITYLFD-L----DYESDEFT-VDAA-RYG---NVS---HFVNHSCDPNLQVFNVFIDN-LD
Hsap_SETD8              ALYAQDPSTGCYMY--YF---QYLSKTYC-VDAT-RETN--RLG---RLINHSKCGNCQTKLHDIDG-V-
Hsap_prdm1              ----NAN-R-KYFWR-IY---SRGELHHF-IDGF-NEEKS-NWM---RYVNPAHSPR-EQNLAACQN-G-
Bflo_prdm1              ----GAN-R-KYFWR-VY---DGEEFQFY-VDGY-DITKS-NWM---RYVNPAHSGE-EQNLVACQH-D-
Dmel_prdm1              ----DKN-KARYFWR-VQ---GKDDDYYY-LDGS-DRSQS-NWM---RYVASAYSLS-VMNLVACQH-Q-
Smar_prdm1              ------STNRKYFWR-LYN---NEELYCY-LDGF-DVNKS-NWM---RYVNPAYSAE-SQNLIACQY-K-
Hsap_prdm2              ---------NVYMWEVY-----YPNLGWMCIDAT-DPEKG-NWL---RYVNWACSGE-EQNLFPLEI-N-
Bflo_prdm2              ---------PDYMWEIR-----TGKGWFC-VDAS-DPTKG-NWM---RYVNSARYFE-EQNIVALQQ-H-
Lgig_prdm2              ------LLDYRYAWEVWDL--ESDKLLYI-INAD-KKNIG-NWM---KYVNCARYFE-EQNIVSVQI-D-
Hsap_prdm3              -----------YGWEIL-----FYNVKFC-IDAS-QPDVG-SWL---KYIRFAGCYD-QHNLVACQI-N-
Hsap_prdm16             --------ETDFGWEQIGCITKLGSEKFC-VDAN-QAGAG-SWL---KYIRVACSCD-DQNLTMCQI-S-
Bflo_prdm3_16           ----------------------FGKVKFI-VDAS-EPGTG-NWM---KYIRSARNYN-EQNMVALQI-N-
Hsap_prdm4              ----WTDKAVNHIWKIY----HNGVLEFC-IITT-DENEC-NWM---MFVRKARNRE-EQNLVAYPH-D-
Bflo_prdm4              ----------------------DERSSHF-IDAR-DENEA-NWM---MFVKRARTSL-EQNLVAHQC-G-
Pdum_prdm4              ----DEVKDTPMLWKIF----ERGEASQV-YDCS-DEESS-NWM---RFMRPASSNG-EQNVVAYQQ-G-
Smar_prdm4              ----IHISDKNNLWKVY-----NDSELTYGLEAD-DENQC-NWM---MFVQRARTSA-EQNVVAFLL-S-
Hsap_prdm5              ------NMDYRLMWE-VRG--SKGEVLYI-LDAT-NPRHS-NWL---RFVHEAPSQE-QKNLAAIQE-G-
Hsap_prdm6              ----GAVRNTQHLWE-IYD--QDGTLQHF-IDGG-EPSKS-SWM---RYIRCARHCG-EQNLTVVQY-R-
Bflo_prdm6              ----AV--KTTHFWE-MY---DNGMFNHY-IDGT-DVRRA-SWM---RYIRCARHRA-EQNMVATQY-K-
Nvec_prdm6a             ----EK---NFYMWE-IF---KDGRLYGF-IDGS-DHNTA-SWM---RFIRCARNKE-EQNLFAFQY-L-
Nvec_prdm6b             ----REL-DTSYMWE-IY---KEGRFSHY-IDGK-DEHLS-SWM---RFIQCARYQE-EQNMTVFQY-C-
Hsap_prdm9              ---------NGYSWLIT-----KGRNCYEYVDGK-DKSWA-NWM---RYVNCARDDE-EQNLVAFQY-H-
Bflo_prdm7_9            ------T-SG-YAWQ-IS---KNDKVKYY-IDAT-DITKS-SWM---RYVNCARNEE-EQNLVAFQY-Y-
Lgig_prdm7_9            ------N-SG-YCWQ-IY---QEGKHHHF-VDAK-NPATS-NWM---RFVNCARTES-EQNVTAYQY-C-
Pdum-prdm7_9a           ----------------IF---KNDKARHF-IDAK-DVSTS-NWM---RYVNCARHDS-EQNLLAFQY-M-
Nvec_prdm7_9            ----END-TS-YMWE-IS---RDGNFNHF-IDGH-DEEQS-NWM---RFVNCSRCEK-EQNLVTFQY-R-
Aque_prdm7_9            ------DDDTSYMWEIK-----KPNESYY-IDGQ-VESES-NWM---RYINCARNEE-EQNLVAFQY-H-
Hsap_prdm8              ALKSTDKRTVPYIF--------RVDTSA--ANGS-SEGL--MWL---RLVQSARDKE-EQNLEAYIK-N-
Pdum_prdm8              ----IGHMTCDNHPD-VHT--VKLDMTDY-IDGE-RAV---EWV---PYVKAARNTS-EQNLEAYLR-D-
Smar_prdm8              ----LRCIDVRSPSQ-I-----LKVDLRK-TEVS-DVYC--EWM---LHVQSARDEH-EQNLEIVAH-E-
Mley_prdm8              ----REQKESGYAWK-AV---SQDGGVIY-LDGA-DKRHS-NWL---RFVNCARGRC-EENVLVRNI-P-
Hsap_prdm10             HLKVSLDKGDRKER--------DLHEDLW-FELS-DETLC-NWM---MFVRPAQNHL-EQNLVAYQY-G-
Hsap_prdm15             --------FPLKVFQ-------KDGHPVC-FDTS-NEDDC-NWM---MLVRPAAEAE-HQNLTAYQH-G-
Bflo_prdm10_15          LLKLQHKTEGEEQP--------DDQPELY-YDLT-NEDEC-DWM---MFVKPAANHM-EQNLVAYQH-G-
Lgig_prdm10_15          LLMVLMNFFLYFQLEKN-----DGTLGYF--ETS-DENKC-NWM---MFVRPARNFA-EQNVVAYQY-G-
Pdum_prdm10_15          --------NKAFSLQ-V----EREGHKFY-FEAT-DENKC-NWM---MFVRPAQSYD-EQNLVAYQY-G-
Smar_prdm10_15a         ----DDLVLQIEAE--------SGEIMH--LDTT-DESMS-NWM---RFVRPAENYS-EQNLVVVQS-E-
Hsap_prdm11             --------AGFFSWL-IV---DKNNRYKS-IDGS-DETKA-NWM---RYVVISREER-EQNLLAFQH-S-
Hsap_prdm12             ----CKN-NN-LMWE-VFN--EDGTVRYF-IDAS-QEDHR-SWM---TYIKCARNEQ-EQNLEVVQI-G-
Bflo_prdm12             ----CVE-NE-FMWE-VFT--PYGDLSHY-VDAS-RQEDR-SWM---AYVNCARSEQ-EQNLELFQK-G-
Lgig_prdm12             ----QPDITD-NIWE-IFD--DKGVILYF-IDGG-TSFPRASWL---SYVNCARNPQ-EQNLELVQI-G-
Pdum_prdm12             ------TAHNHQVWE-VFDS--RGEVLHF-IDGS-PTSLR-SWL---TFVQCARSPE-EQNLEVVQI-G-
Smar_prdm12             --------DNCWAWE-VFDLSNDGQLVHI-VDAS-DEKTR-CWL---HYVNCARHEQ-EQNLEI------
Nvec_prdm12             ----EID-NN-LMWE-VLN--EDGSVSHF-LDAK-EENPR-NWM---GFVNCARNEQ-EQNLEVFQY-G-
Aque_prdm12             ----L---KLKSYWE-IH---QDGKVQYC-VDGS-DPHYR-NWM---YFIQYARHQD-EQNLMAIQH-E-
Hsap_prdm13             ----------VRMV--------RGEL----VDES-GGSPL-EWI---GLIRAARNSQ-EQTLEAIADLPG
Hsap_prdm14             ------YGDNSVMWEIF-----DGHLSHF-IDGKGGTG---NWM---SYVNCARFPK-EQNLVAVQC-Q-
Bflo_prdm14             ----DDD-NS-HMWE-IF---QDGKLSHM-IDGR-G-PTG-NWM---AYVNCARYAQ-EQNLIAIQC-E-
Lgig_prdm14a            ----FDD-NT-HMWE-IF---KEGKLSHF-IDGR-G-TNG-SWM---SYLNCARHIA-EQNMVVLQE-G-
Lgig_prdm14b            ----NDD-NS-FMWE-VF---QEGKLSHF-IDGR-G-STG-NWM---SCVNCARYAQ-EQNLIAVQH-E-
Pdum_prdm14             ----NDD-NS-FMWE-VF---AKGQLSHF-VDGR-G-SAG-NWM---AYVNCARYAQ-EQNLLALQC-E-
Nvec_prdm14a            ----DLD-NR-HMWE-VF---AHEKVSYY-IDGQ-D-EQN-NWM---KFINCANSNT-EQNMSVIQQ-G-
Nvec_prdm14b            ----YK--ESPYVWE-VFD--DYGKLTHL-IDGR-DALQS-NWL---RYVNCSKGLD-EQNLRAVQY-D-
Nvec_prdm14c            ----RED-NA-FMWE-VF---QNGQFSHF-IDGS-T-ESE-NWM---KFVNCARHRD-EQNLRLVQD-G-
Nvec_prdm14d            ---------------Q-IF---QDGELVQY-IDGS-Q-DST-NWM---KYVKCARHDG-EQNLALAQD-G-
Bflo_prdmX              T---KDKRMVSYAY--------KVDPGCC----R-EAGETLKWL---RLIQPARDRR-EQNMEAFQR-A-
Smar_prdmX1             ------DANLKNVWE-V-----FGETKTY-LDTS-DPASS-NWL---RYMRPAMSRD-DRNIATVTKMD-
Aque_prdmX1             ----ETD-DISCLWE-I----KCCESTYY-VDAH-EESNS-NWL---KYVKLARNDN-EQNLIAFQH-R-
Aque_prdmX2             -----ETNNFNAMWEVN-----CGESTYY-VDAH-EADNS-SWL---KCVRFARNKE-EQNLVAFQH-R-

                                                                        *******
                                                               R/HxxNHxC box (AdoMet)
                                      ---------b′--------------Block b--------------b′-------
```

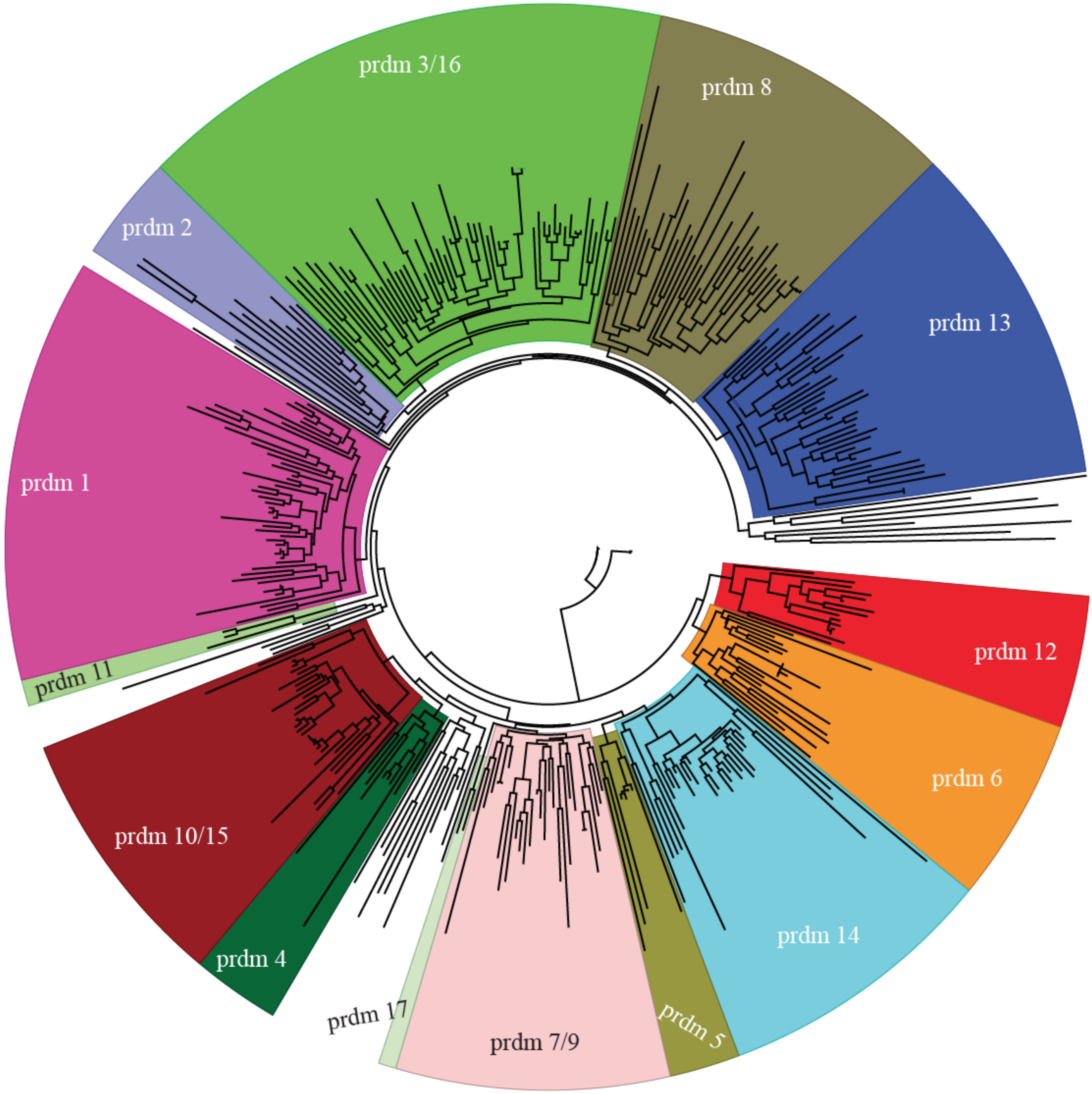


```
                              -----c’-------Block c-
                                     F/Y switch   AdoMet and target Lys
                                               * *
Hsap_SETD6                ---VTVREAALQGTKTEAERHLVYER
Hsap_SETD7                ---PIKCIRTLRAVEADEELTVAYGY
Hsap_SUV420H1             ---DTACVKALRDIEPGEEISCYYGD
Hsap_SUV420H2             ---NAACVKVLRDIEPGDEVTCFYGE
Hsap_MLL1                 ---HKIIISSSRRIQKGEELCYDYKF
Hsap_MLL2                 ---DKIIISSRRIPKGEELTYDYQF
Hsap_MLL3                 ---KHIVIFAMRKIYRGEELTYDYKF
Hsap_MLL4                 ---KHIVIFALRRILRGEELTYDYKF
Hsap_SETDA1               ---KKIVIYSKQPIGVDEEITYDYKF
Hsap_SETDB1               LRFPWVAFFASK-IRAGTELTWDYNY
Hsap_SETDB2               RNFPLVAFFTNRYVKARTELTWDYGY
Hsap_ASH1L                ---YRIGLYALKDMPAGTELTYDYNF
Hsap_SETD2                ---LRVGFFTTKLVPSGSELTFDYQF
Hsap_NSD2                 ---TRVGLFAVCDIPAGTELTFNYNL
Hsap_NSD3                 ---VRVGLFALCDIPAGMELTFNYNL
Hsap_H3_36_H4_20          ---TRVGLFALSDIKAGTELTFNYNL
Hsap_EZH1                 ---HRIGIFAKRAIQAGEELFFDYRY
Hsap_EZH2                 ---HRIGIFAKRAIQTGEELFFDYRY
Hsap_SETMAR               SMVPKLALFAAKDIVPEEELSYDYSG
Hsap_EHMT1                LRFPRIAFFSTRLIEAGEQLGFDYGE
Hsap_EHMT2                LRFPRIAFFSSRDIRTGEELGFDYGD
Hsap_SUV39H1              ERLPRIAFFATRTIRAGEELTFDYNM
Hsap_SUV39H2              TRLPRIALFSTRTINAGEELTFDYQM
Hsap_SETD8                ---PHLILIASRDIAAGEELLYDYGD
Hsap_prdm1                ---MNIYFYTIKPIPANQELLVWYCR
Bflo_prdm1                ---MKIYFFTIKPVLPDTELLVWYCR
Dmel_prdm1                ---ENIYFYTTRDILPNEELMVWYCK
Smar_prdm1                ---MNIYFYTIKPVLPNTELLVWYCK
Hsap_prdm2                ---RAIYYKTLKPIAPGEELLVWYNG
Bflo_prdm2                ---QRIYYKTIKEIKPGEELLCWFNS
Lgig_prdm2                ---NEIYYKAIKDIEIGEELLTWFNL
Hsap_prdm3                ---DQIFYRVVADIAPGEELLLFMKS
Hsap_prdm16               ---EQIYYKVIKDIEPGEELLVHVKE
Bflo_prdm3_16             ---DQMYYKVVKDVEMGEELMVYMKD
Hsap_prdm4                ---GKIFFCTSQDIPPENELLFYYSR
Bflo_prdm4                ---GEIFFTSCKDIAEGEELLMWFAG
Pdum_prdm4                ---QDIWFITKKDVAPGTELRYWYSY
Smar_prdm4                ---SQIYFMTTKTILPDTELLYWYST
Hsap_prdm5                ---ENIFYLAVEDIETDTELLIGYLD
Hsap_prdm6                ---SNIFYRACIDIPRGTELLVWYND
Bflo_prdm6                ---GCIFYKIFREIKTGEELLVWYDP
Nvec_prdm6a               ---GKIYYRTYRTIPMGEELLVWYDD
Nvec_prdm6b               ---GNIYYRAYKHIPKGRELRVWYDD
Hsap_prdm9                ---RQIFYRTCRVIRPGCELLVWYGD
Bflo_prdm7_9              ---RNIYYRTYKPIPPGTELMVWYGN
Lgig_prdm7_9              ---GEIYYRTFKEIPPGSEILVWYGN
Pdum-prdm7_9a             ---GNIYYKTIMPVQSCEELLVWYGD
Nvec_prdm7_9              ---GQIFYRSYKDVHPGTELLVWYGD
Aque_prdm7_9              ---GEIYYRTFKDICPGTELFVWYGD
Hsap_prdm8                ---GQLFYRSLRRIAKDEELLVWYGK
Pdum_prdm8                ---GQLHYRALHNIPTNRELFIWYSP
Smar_prdm8                ---GRLYFRTTRPVKTDEELLVWYGT
Mley_prdm8                ---GGVQYYTYKPIAIGTELLVFYGD
Hsap_prdm10               ---HHVYYTTIKNVEPKQELKVWYAA
Hsap_prdm15               ---SDVYFTTSRDIPPGTELRVWYAA
Bflo_prdm10_15            ---NEIFYTTIKPVEPRQELKVWYST
Lgig_prdm10_15            ---QDIYFTVMKNIEARQELKVWYSA
Pdum_prdm10_15            ---SEIYFCTTKHIQPRVELKVWYSA
Smar_prdm10_15a           ---QSLYFFTTCHVYPRTELKVGYSQ
Hsap_prdm11               ---ERIYFRACRDIRPGEWLRVWYSE
Hsap_prdm12               ---TSIFYKAIEMIPPDQELLVWYGN
Bflo_prdm12               ---EHIYYRAMKAIPPDEELLVWYRF
Lgig_prdm12               ---HNIFYRAIKDIAPQKELLVWYGS
Pdum_prdm12               ---TEIFYRAIKNIPPEEELLVWYNE
Smar_prdm12               ------------NIPPNQELLVWYGT
Nvec_prdm12               ---GNIYYRAIKDVPPDQELLVWYGG
Aque_prdm12               ---ERVYFRAVRDIQLGEELLVWYDK
Hsap_prdm13               ---GQIFYRALRDVQPGEELTVWYSN
Hsap_prdm14               ---GHIFYESCKEIHQNQELLVWYGD
Bflo_prdm14               ---GQIYYEACKEIPQGAELLVWYGD
Lgig_prdm14a              ---DEVYYYATRDISPGTELLVWYGQ
Lgig_prdm14b              ---GEIYYEVCKDIPQGTELLVWYGD
Pdum_prdm14               ---DLIYYEACKDIPQGTELLVWYGD
Nvec_prdm14a              ---SDIYYEACADIYRGAELLVWYGS
Nvec_prdm14b              ---KNIYYMATKEIAIGEELLTYYGD
Nvec_prdm14c              ---GDLYYDACRDIIRGEELLVWYGP
Nvec_prdm14d              ---NELYYDVTKDIYEGTELLVWYGD
Bflo_prdmX                ---GKVYFRSVCHIRRDEELLVWYSD
Smar_prdmX1               ---GRLYFVTIHDVTAGSELFYWSED
Aque_prdmX1               ---RNIYYLTIKPIGPGAELLVWYGD
Aque_prdmX2               ---RNIYYMTIKPIHPSSELLVWYGD
                                               * *
                                     F/Y switch   AdoMet and target Lys
                              -----c’-------Block c-
```

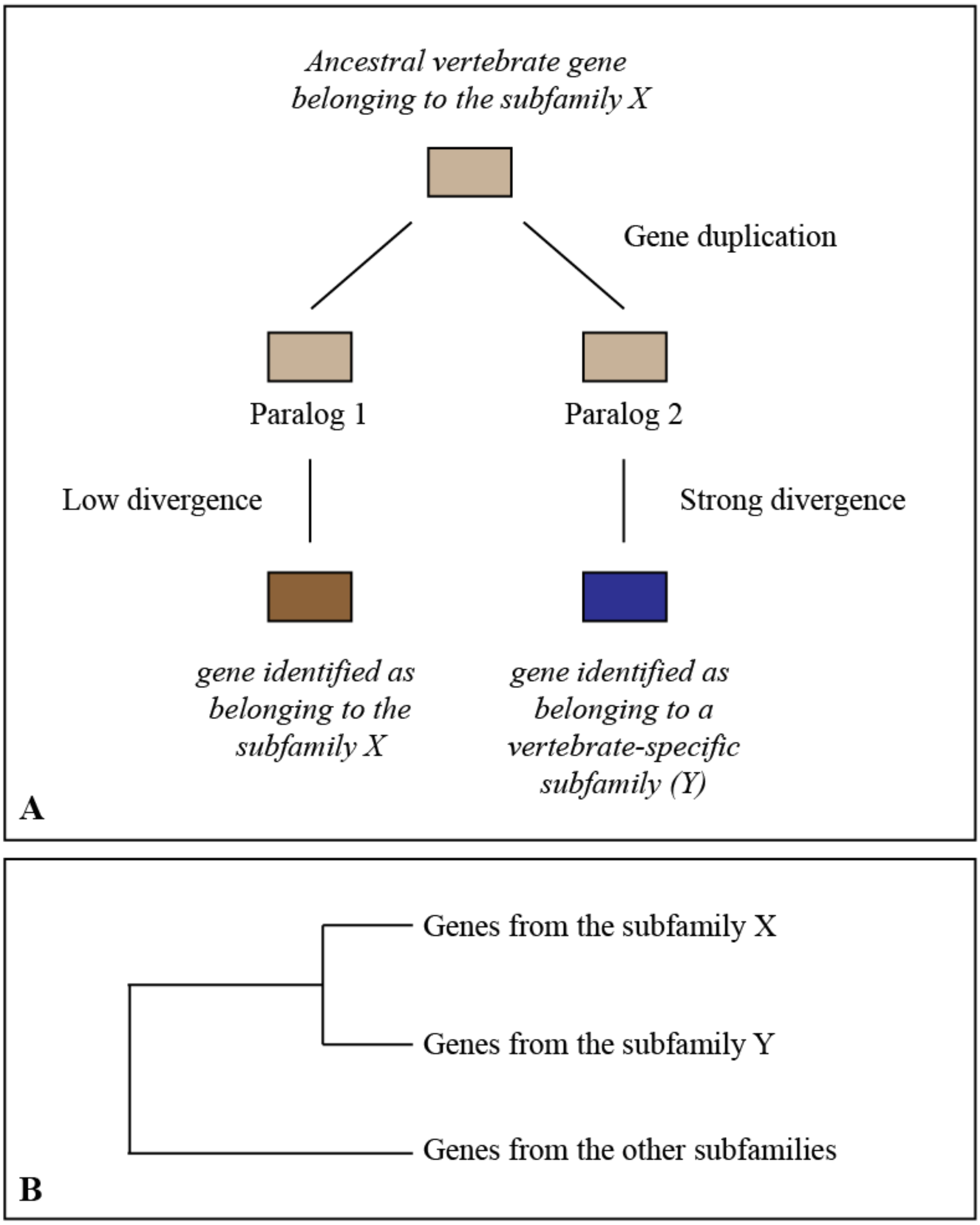
Supplementary Figure S1: Multiple alignment of the SET domain of Prdm and non-Prdm proteins. Four conserved blocks of sequence are indicated – blocks *a*, *b* and *c* are indicated as in Dillon et al. (2005) and Hohenauer and Moore (2012). Conservation extends around these blocks, defining the *a'*, *b'* and *c'* blocks. Important functional residues and motifs discussed in Supplementary Text 1 are indicated. Underlined asterisks in block *b* indicate the position of the N-H motif discussed in Supplementary Text 1.



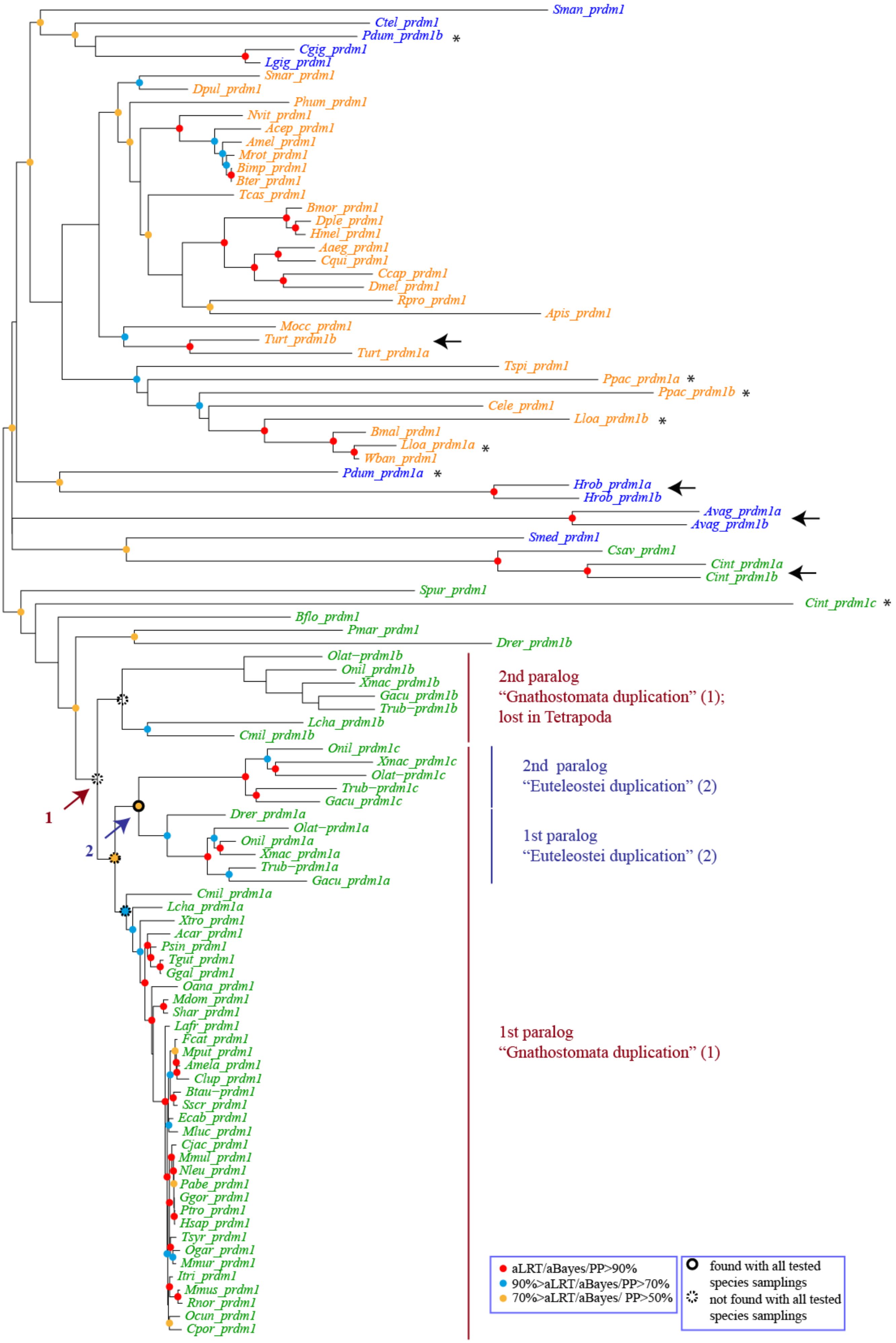
Supplementary Figure S2: Conserved regions in the SET domain of Prdm proteins. Sequence logos have been generated (with WebLogo) from multiple alignments of the four conserved blocks defined in Fig. S1. Only Prdm sequences were included in these multiple alignments. The red double arrow indicates a position where a cysteine is found in most non-Prdm SET domains in block *b*. The two asterisks in a red box indicate the position of the N-H motif in block *b* discussed in Supplementary Text 1. The colored lines (pink, orange and brown) under the number of the residues correspond to the conserved residues and motifs described in Supplementary Text 1: brown for residues strongly conserved in most Prdm and non-Prdm proteins, pink for residues strongly conserved in most Prdm and a few non-Prdm proteins, and orange for residues well conserved in Prdm but not in non-Prdm proteins.



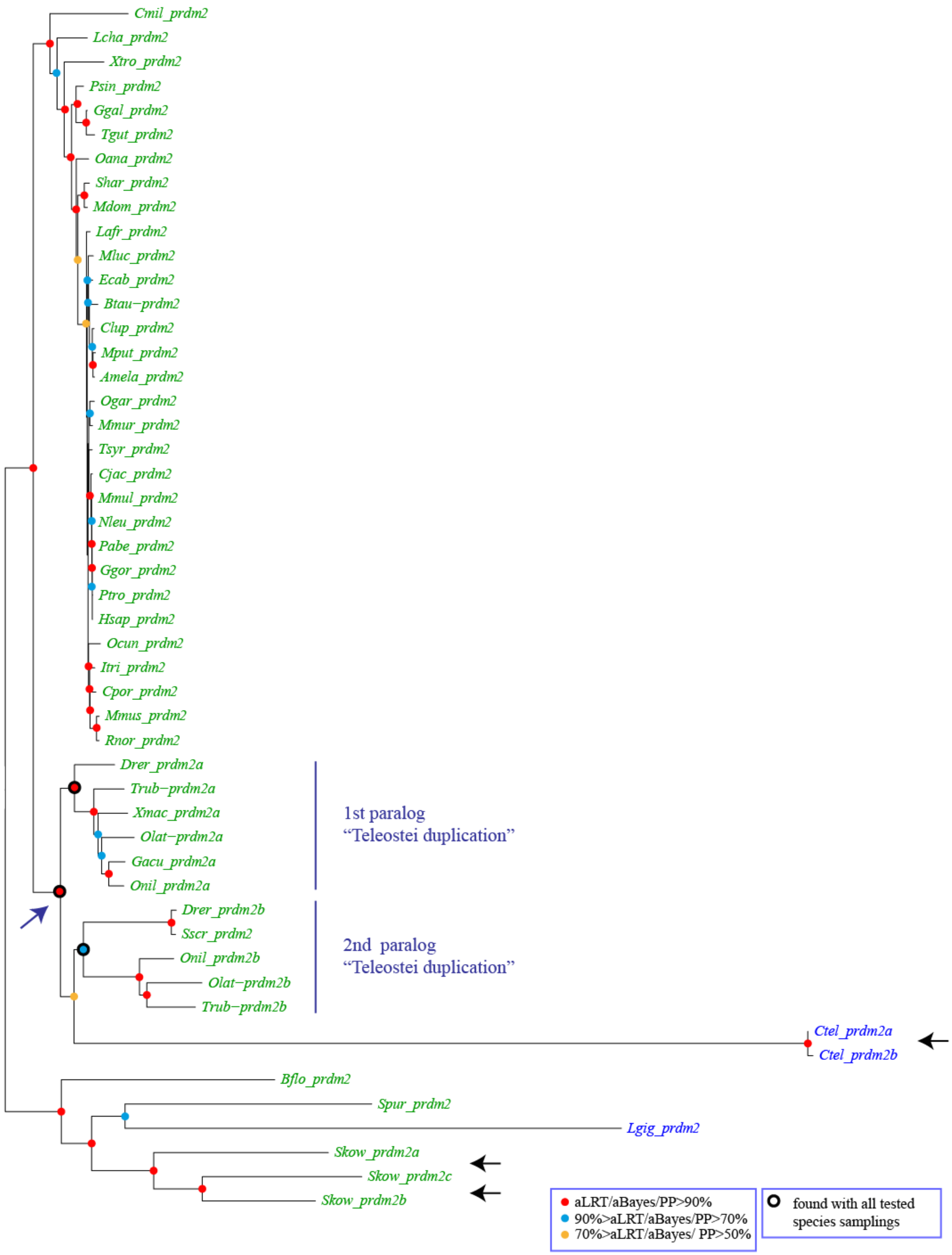
Supplementary Figure S3: Phylogenetic analysis of *Prdm* genes. An unrooted Maximum-likelihood (ML) tree is shown. See Table S7 for the statistical support of the different monophyletic groups. Colored monophyletic groups correspond to the Prdm subfamilies listed in Tables 1 and S1. Uncolored parts of the tree correspond to 'orphan' genes.



Supplementary Figure S4: Origin of the vertebrate specific subfamilies. (A) A model for the origin of the vertebrate specific subfamilies through gene duplication. (B) Expected phylogenetic tree in the case of the model described in (A).


Sman_prdm1
Ctel_prdm1
Pdum_prdm1b *
Cgig_prdm1
Lgig_prdm1
Smar_prdm1
Dpul_prdm1
Phum_prdm1
Nvit_prdm1
Acep_prdm1
Amel_prdm1
Mrot_prdm1
Bimp_prdm1
Bter_prdm1
Tcas_prdm1
Bmor_prdm1
Dple_prdm1
Hmel_prdm1
Aaeg_prdm1
Cqui_prdm1
Ccap_prdm1
Dmel_prdm1
Rpro_prdm1
Apis_prdm1
Mocc_prdm1
Turt_prdm1b
Turt_prdm1a
Tspi_prdm1
Ppac_prdm1a *
Ppac_prdm1b *
Cele_prdm1
Lloa_prdm1b *
Bmal_prdm1
Lloa_prdm1a *
Wban_prdm1
Pdum_prdm1a *
Hrob_prdm1a
Hrob_prdm1b
Avag_prdm1a
Avag_prdm1b
Smed_prdm1
Csav_prdm1
Cint_prdm1a
Cint_prdm1b
Spur_prdm1
Cint_prdm1c *
Bflo_prdm1
Pmar_prdm1
Drer_prdm1b
Olat−prdm1b
Onil_prdm1b
Xmac_prdm1b
Gacu_prdm1b
Trub−prdm1b
Lcha_prdm1b
Cmil_prdm1b
2nd paralog
“Gnathostomata duplication” (1);
lost in Tetrapoda
Onil_prdm1c
Xmac_prdm1c
Olat−prdm1c
Trub−prdm1c
Gacu_prdm1c
2nd paralog
“Euteleostei duplication” (2)
Drer_prdm1a
Olat−prdm1a
Onil_prdm1a
Xmac_prdm1a
Trub−prdm1a
Gacu_prdm1a
1st paralog
“Euteleostei duplication” (2)
1
2
Cmil_prdm1a
Lcha_prdm1a
Xtro_prdm1
Acar_prdm1
Psin_prdm1
Tgut_prdm1
Ggal_prdm1
Oana_prdm1
Mdom_prdm1
Shar_prdm1
Lafr_prdm1
Fcat_prdm1
Mput_prdm1
Amela_prdm1
Clup_prdm1
Btau−prdm1
Sscr_prdm1
Ecab_prdm1
Mluc_prdm1
Cjac_prdm1
Mmul_prdm1
Nleu_prdm1
Pabe_prdm1
Ggor_prdm1
Ptro_prdm1
Hsap_prdm1
Tsyr_prdm1
Ogar_prdm1
Mmur_prdm1
Itri_prdm1
Mmus_prdm1
Rnor_prdm1
Ocun_prdm1
Cpor_prdm1
1st paralog
“Gnathostomata duplication” (1)
● aLRT/aBayes/PP>90%
● 90%>aLRT/aBayes/PP>70%
● 70%>aLRT/aBayes/ PP>50%
○ found with all tested species samplings
○ not found with all tested species samplings

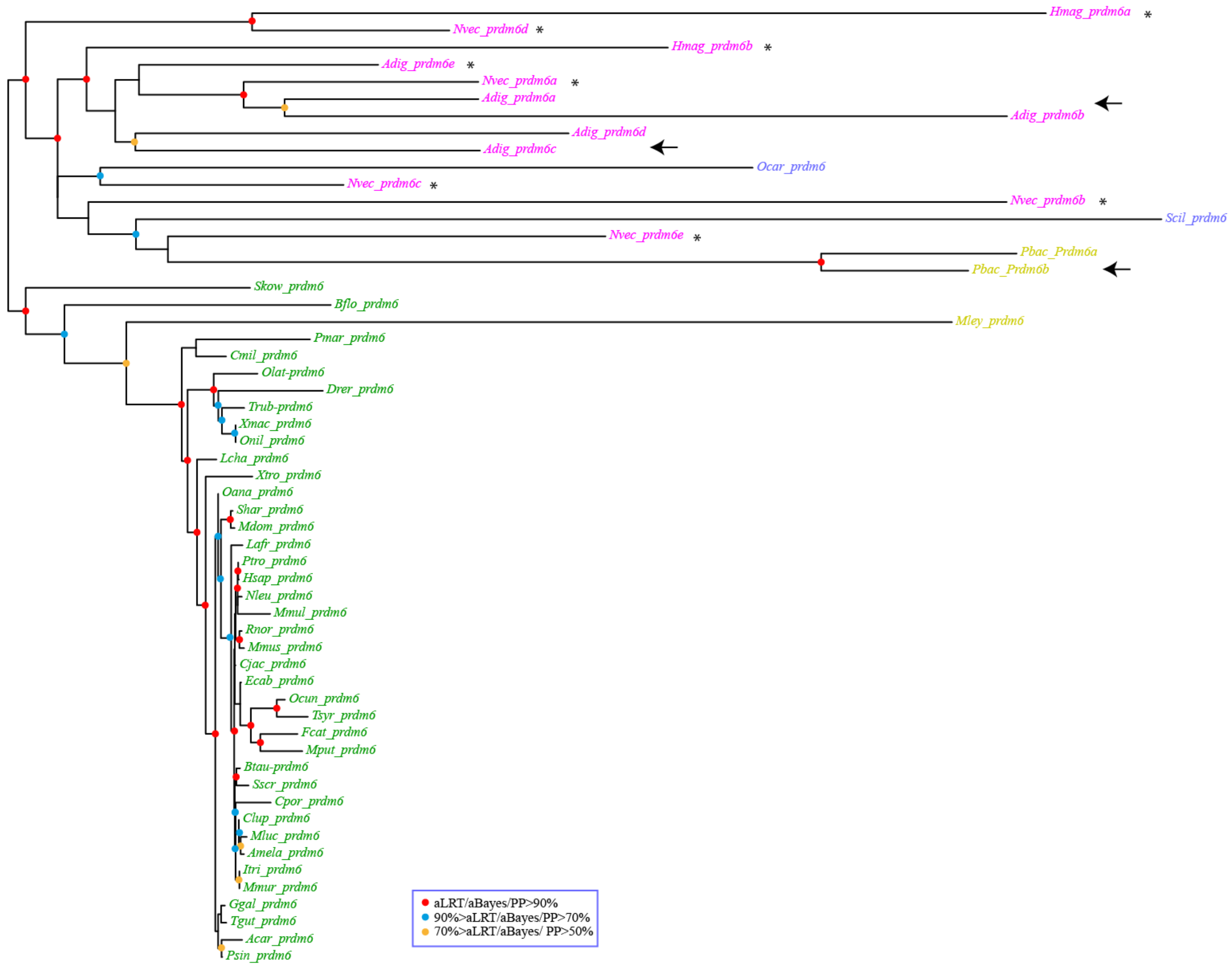
Supplementary Figure S5: Phylogenetic analysis of the Prdm1 subfamily. A Maximum-likelihood (ML) tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by Bayesian inference (BI). Statistical supports (aLRT and aBayes values for ML; posterior probabilities, PP, for BI) are indicated on the nodes by color circles (color code is indicated in the figure). Nodes without color circles are not statistically supported and/or not found by both ML and BI methods. Species names are abbreviated using the first letter of the genus name followed by the three first letters of the species name. All abbreviations can be found in Table 1. Black arrows indicate pairs of paralogs that are closely related in the phylogenetic tree, whereas asterisks denotes paralogs that are not closely associated in the phylogenetic tree. Duplications that likely occurred in Gnathostomata and Euteleostei are indicated (red and purple arrows respectively). Monophyletic groups that allowed to position these duplications are also highlighted. Some but not all these monophyletic groups are also found in ML trees constructed with several different species samplings (sampling 1 : all sequences minus sequences that appear the most divergent in the multiple alignment and phylogenetic tree, *i.e.*, *Cint_prdm1C*, *Sman_prdm1*, *Smed_prdm1*, *Avag_prdm1a*, *Avag_prdm1b*, *Hrob_prdm1a*, *Hrob_prdm1b*, and *Pdum_prdm1a* ; sampling 2 : only deuterostome genes ; sampling 3 : only chordate genes ; sampling 4 : only chordate genes minus *Pmar_prdm1*, *Drer_prdm1b*, and *Cint_prdm1c*).


Cmil_prdm2
Lcha_prdm2
Xtro_prdm2
Psin_prdm2
Ggal_prdm2
Tgut_prdm2
Oana_prdm2
Shar_prdm2
Mdom_prdm2
Lafr_prdm2
Mluc_prdm2
Ecab_prdm2
Btau−prdm2
Clup_prdm2
Mput_prdm2
Amela_prdm2
Ogar_prdm2
Mmur_prdm2
Tsyr_prdm2
Cjac_prdm2
Mmul_prdm2
Nleu_prdm2
Pabe_prdm2
Ggor_prdm2
Ptro_prdm2
Hsap_prdm2
Ocun_prdm2
Itri_prdm2
Cpor_prdm2
Mmus_prdm2
Rnor_prdm2
Drer_prdm2a
Trub−prdm2a
Xmac_prdm2a
Olat−prdm2a
Gacu_prdm2a
Onil_prdm2a
1st paralog
“Teleostei duplication”
Drer_prdm2b
Sscr_prdm2
Onil_prdm2b
Olat−prdm2b
Trub−prdm2b
2nd paralog
“Teleostei duplication”
Ctel_prdm2a
Ctel_prdm2b
Bflo_prdm2
Spur_prdm2
Lgig_prdm2
Skow_prdm2a
Skow_prdm2c
Skow_prdm2b
● aLRT/aBayes/PP>90%
● 90%>aLRT/aBayes/PP>70%
● 70%>aLRT/aBayes/ PP>50%
○ found with all tested species samplings

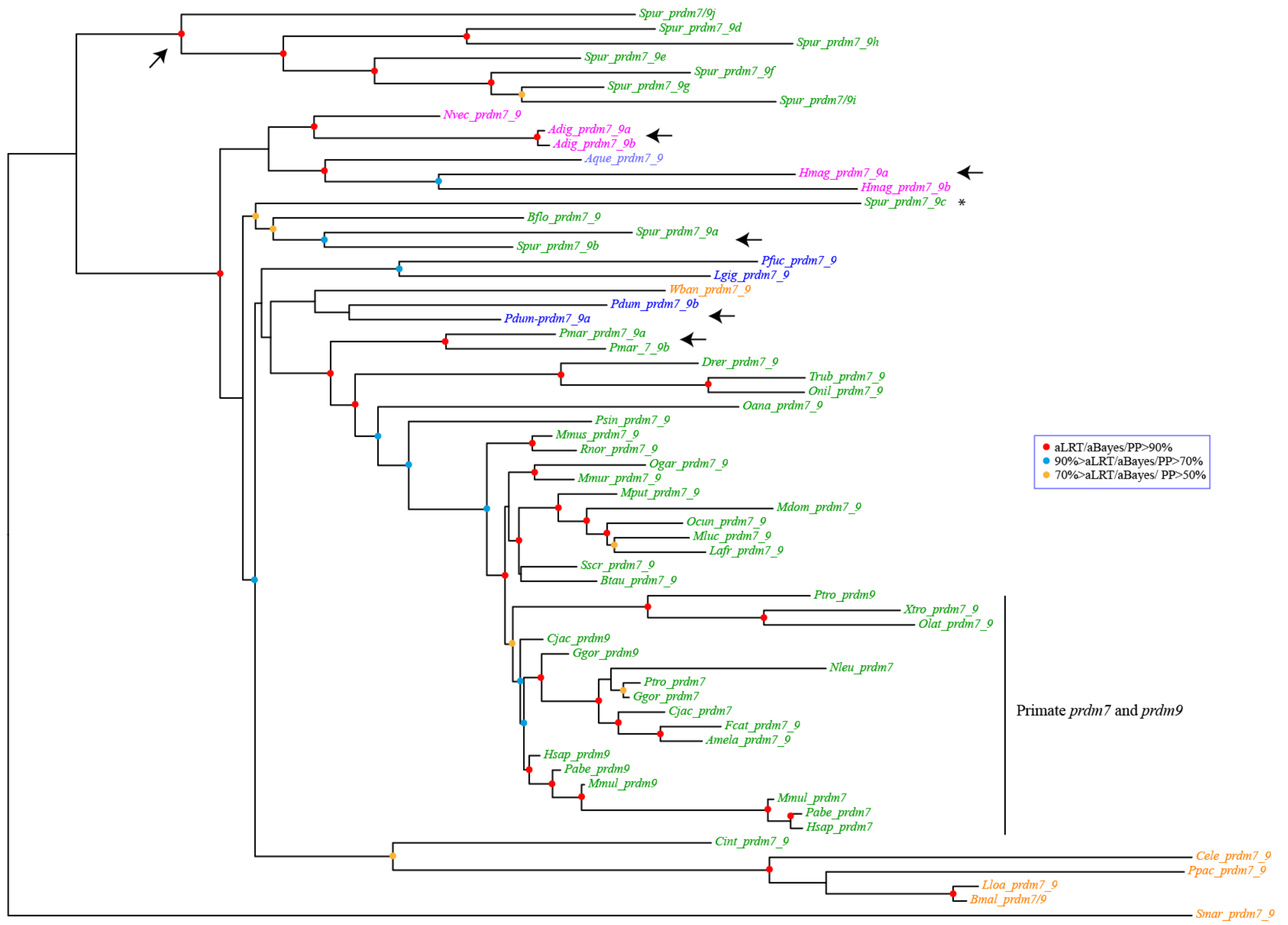
Supplementary Figure S6: Phylogenetic analysis of the Prdm2 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5. A duplications that likely took place in Teleostei is indicated (purple arrow). Monophyletic groups that allowed to position this duplication are also highlighted. These monophyletic groups are also found in ML trees constructed with several different species samplings (sampling 1: only deuterostome genes; sampling 2: only chordate genes; sampling 3: only vertebrate genes).

Supplementary Figure S7: Phylogenetic analysis of the Prdm4 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.

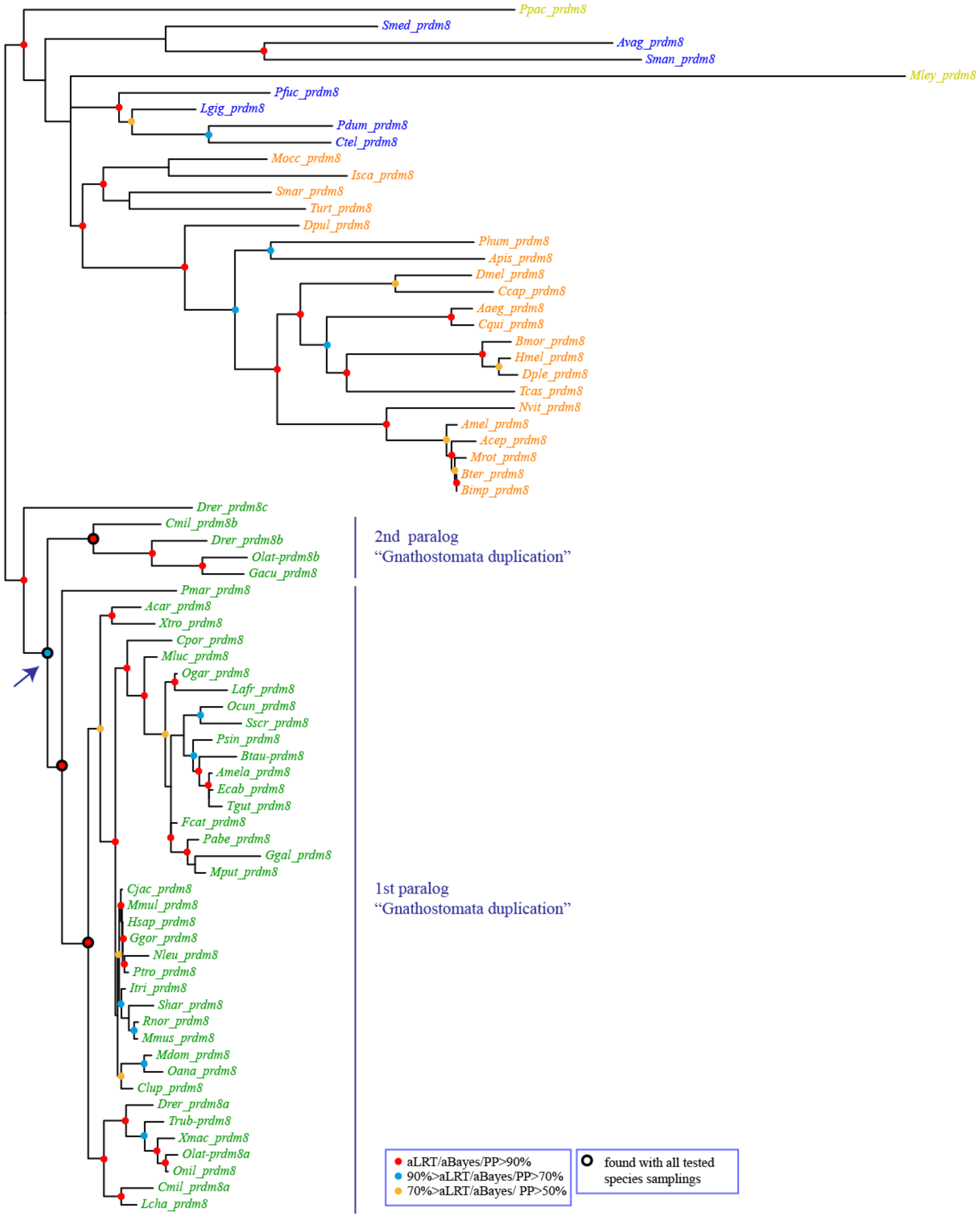
Supplementary Figure S8: Phylogenetic analysis of the Prdm5 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.



Supplementary Figure S9: Phylogenetic analysis of the Prdm6 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.



Supplementary Figure S10: Phylogenetic analysis of the Prdm7/9 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5. To test the robustness of the tree, we constructed ML trees with several different species samplings (sampling 1: only deuterostome genes; sampling 2: only chordate genes; sampling 3: only vertebrate genes; sampling 4: only mammalian genes).

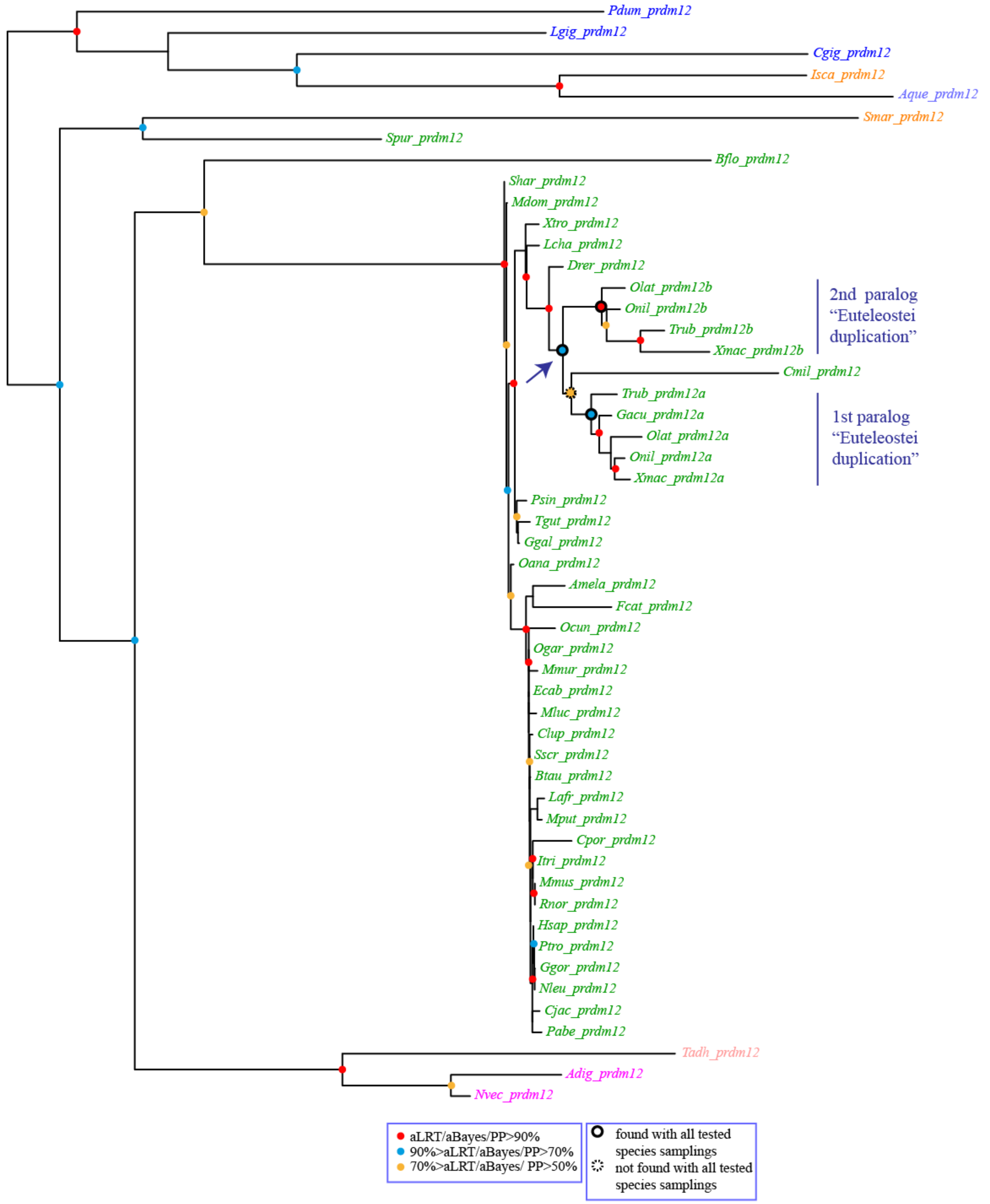


Supplementary Figure S11: Phylogenetic analysis of the Prdm8 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in

Supplementary Fig. S5. A duplications that likely took place in Gnathostomata is indicated (purple arrow). Monophyletic groups that allowed to position this duplication are also highlighted. These monophyletic groups are also found in ML trees constructed with several different species samplings (sampling 1: all sequences minus *Ppac_prdm8*, *Smed_prdm8*, *Avag_prdm8*, *Sman_prdm8*, and *Mley_prdm8;* sampling 2: only vertebrate genes; sampling 3: only vertebrate genes minus *Drer_prdm8c*).

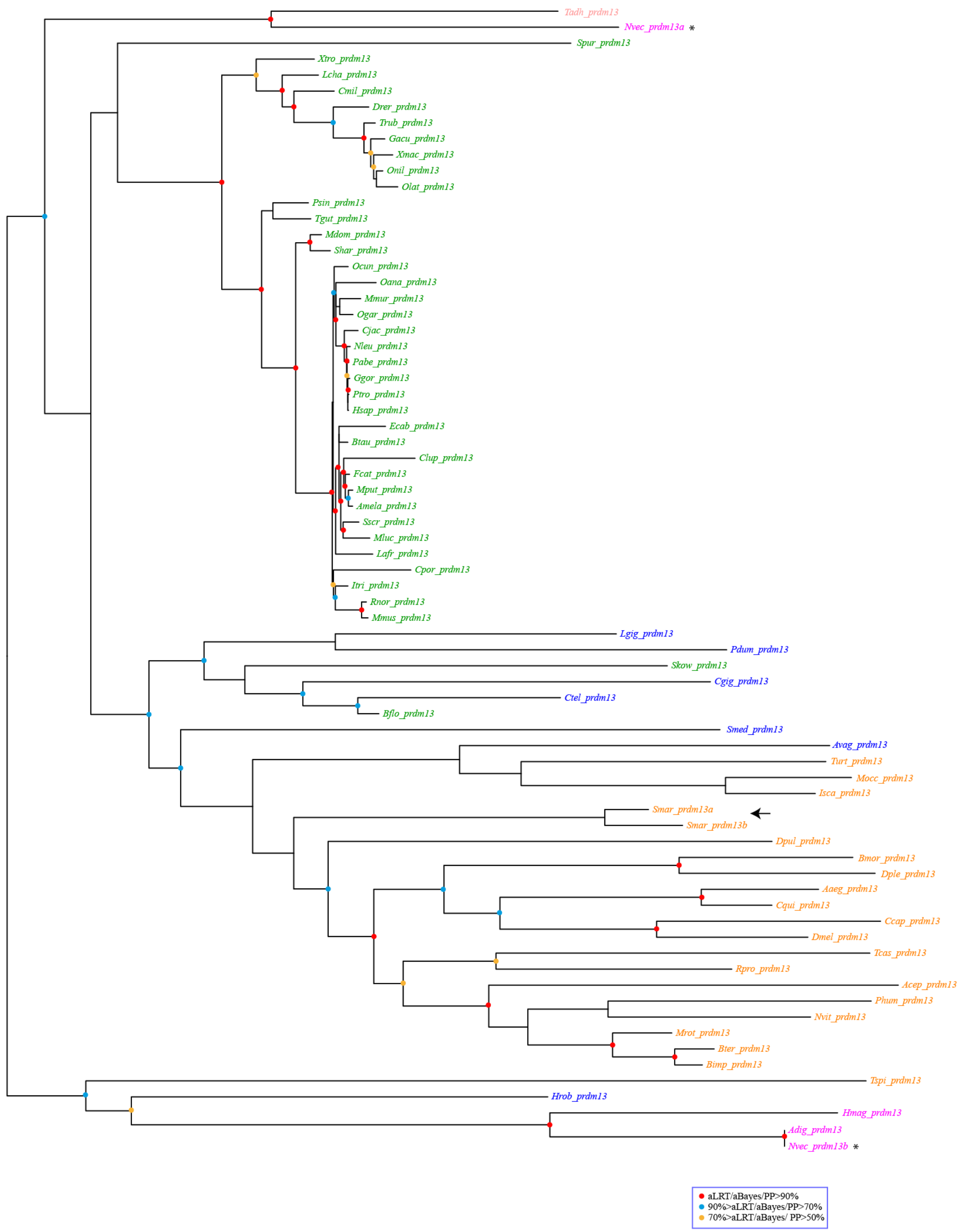

Lcha_prdm11
Oana_prdm11
Mdom_prdm11
Shar_prdm11
Tgut_prdm11
Ggal_prdm11
Acar_prdm11
Psin_prdm11
Btau-prdm11
Sscr_prdm11
Lafr_prdm11
Mmur_prdm11
Ogar_prdm11
Cjac_prdm11
Mmul_prdm11
Nleu_prdm11
Hsap_prdm11
Ptro_prdm11
Ggor_prdm11
Tsyr_prdm11
Rnor_prdm11
Mmus_prdm11
Cpor_prdm11
Ocun_prdm11
Itri_prdm11
Mluc_prdm11
Clup_prdm11
Mput_prdm11
Amela_prdm11
Fcat_prdm11
Ecab_prdm11
Drer_prdm11
● aLRT/aBayes/PP>90%
● 90%>aLRT/aBayes/PP>70%
● 70%>aLRT/aBayes/ PP>50%

Supplementary Figure S12: Phylogenetic analysis of the Prdm11 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.



Supplementary Figure S13: Phylogenetic analysis of the Prdm12 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in

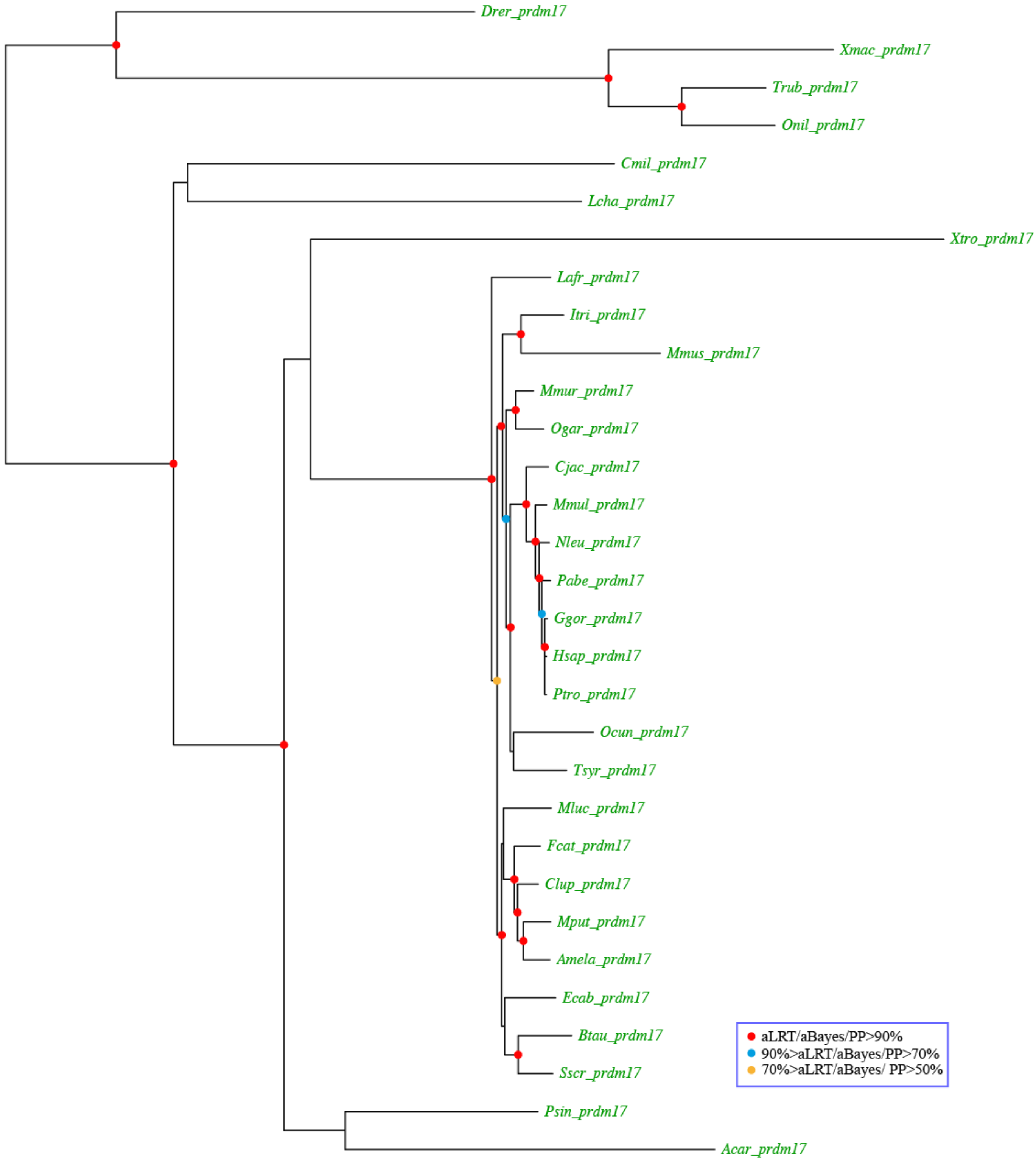
Supplementary Fig. S5. A duplications that likely occurred in Euteleostei is indicated (purple arrow). Monophyletic groups that allowed to position this duplication are also highlighted. An odd feature of the phylogenetic tree is the grouping of the single gene from *Callorhinchus* with the *Prdm12a* genes. In phylogenetic trees constructed with four different partial samplings of species (sampling 1: only deuterostome genes; sampling 2: only chordate genes; sampling 3: only vertebrate genes; sampling 4: only vertebrate genes minus *Cmil_prdm12*), we found the same monophyletic groups for the teleost genes than in the tree constructed with the full dataset. The only difference is that the *Cmil_prdm12* gene is found as outgroup to the teleost genes in the trees constructed with species samplings 1 to 3.



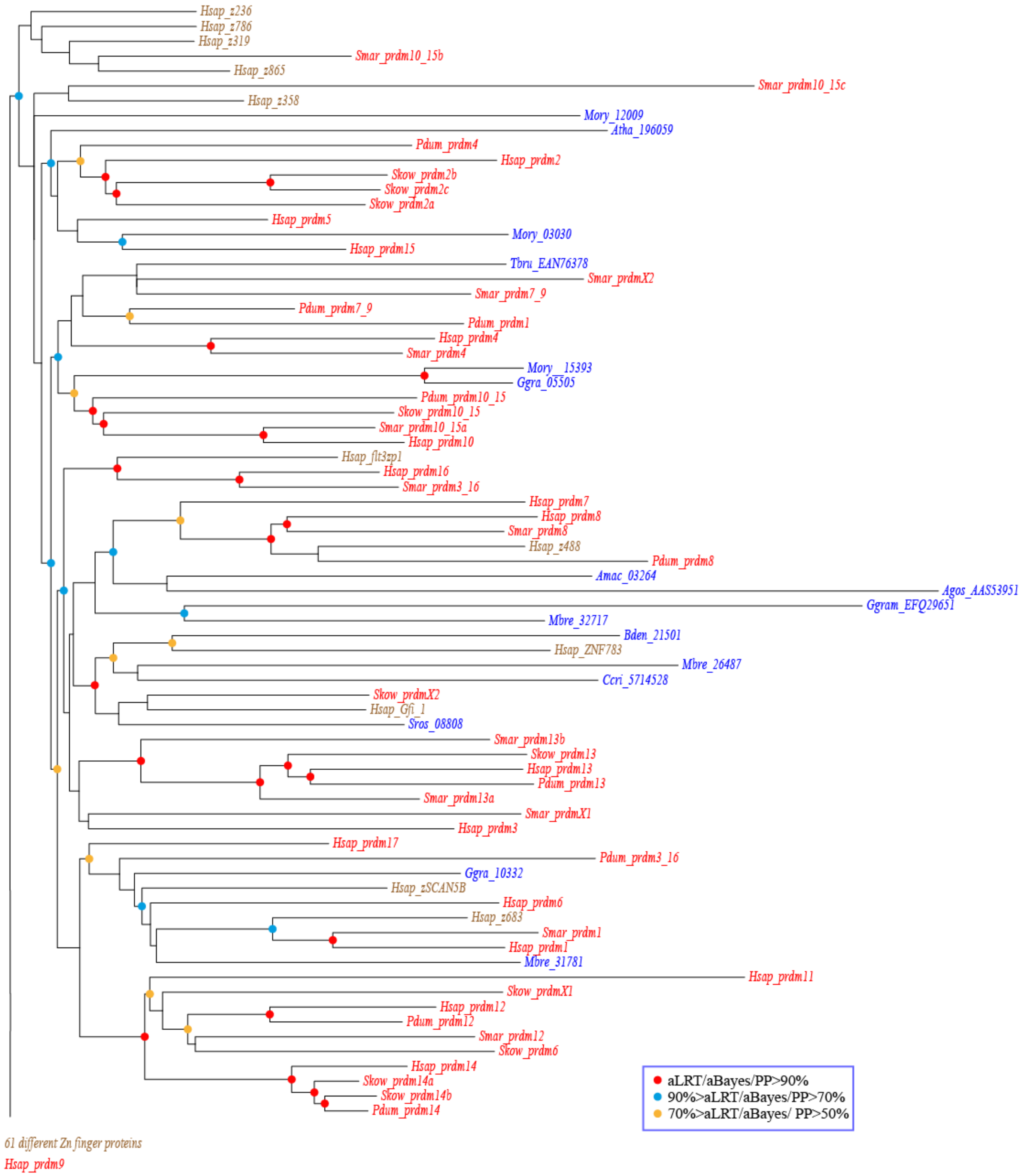
Supplementary Figure S14: Phylogenetic analysis of the Prdm13 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.


Adig_prdm14
Nvec_prdm14d *
Nvec_prdm14a *
Nvec_prdm14c *
Nvec_prdm14b *
Cgig_prdm14c
Cgig_prdm14d
Cgig_prdm14e
Skow_prdm14a *
Bflo_prdm14
Skow_prdm14b *
Spur_prdm14
Cmil_prdm14b *
Cmil_prdm14a *
Lcha_prdm14
Xtro_prdm14
Itri_prdm14
Lafr_prdm14
Mmus_prdm14
Rnor_prdm14
Ecab_prdm14
Fcat_prdm14
Clup_prdm14
Mput_prdm14
Amela_prdm14
Mluc_prdm14
Btau_prdm14
Sscr_prdm14
Tsyr_prdm14
Mmul_prdm14
Pabe_prdm14
Nleu_prdm14
Ggor_prdm14
Hsap_prdm14
Ptro_prdm14
Cjac_prdm14
Cpor_prdm14
Mmur_prdm14
Ogar_prdm14
Ocun_prdm14
Shar_prdm14
Mdom_prdm14
Tgut_prdm14
Ggal_prdm14
Psin_prdm14
Oana_prdm14
Smed_prdm14
Acar_prdm14
Drer_prdm14
Trub_prdm14
Olat_prdm14
Gacu_prdm14
Onil_prdm14
Xmac_prdm14
Ctel_prdm14b *
Pdum_prdm14
Ctel_prdm14a *
Cgig_prdm14a
Lgig_prdm14a
Hrob_prdm14
Lgig_prdm14b
Cgig_prdm14b
● aLRT/aBayes/PP>90%
● 90%>aLRT/aBayes/PP>70%
● 70%>aLRT/aBayes/ PP>50%

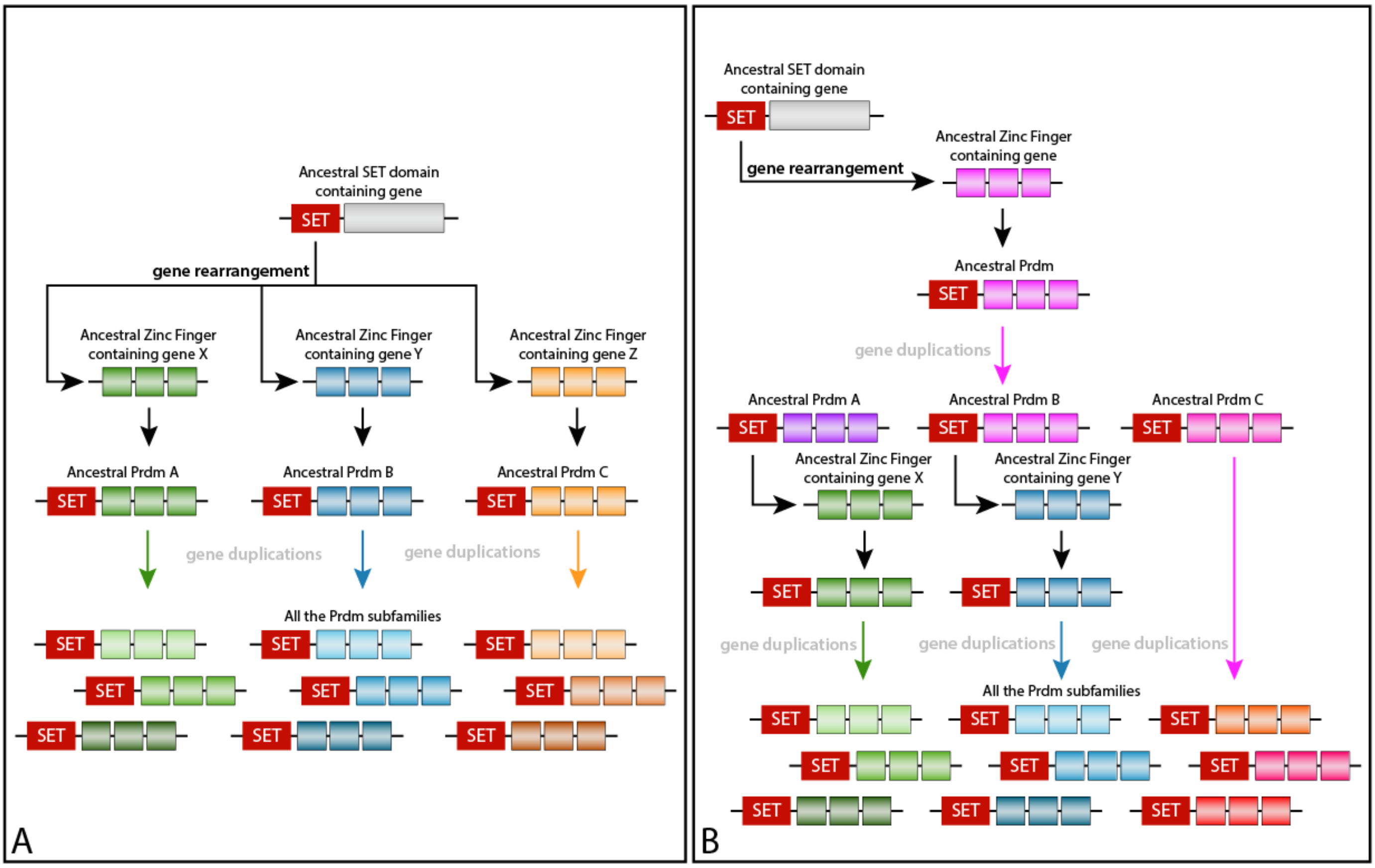
Supplementary Figure S15: Phylogenetic analysis of the Prdm14 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5. In addition, red arrows point to two groups of mollusk genes and correspond to a gene duplication that has occurred in the mollusk lineage.



Supplementary Figure S16: Phylogenetic analysis of the Prdm17 subfamily. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5.



Supplementary Figure S17: Phylogenetic analysis of the Zn finger and Prdm proteins. A ML tree is shown. Midpoint rooting has been used. A similar tree topology has been obtained by BI. Legend is as in Supplementary Fig. S5. Additional species abbreviations can be found in Table S17.



Supplementary Figure S18: Two alternative hypotheses for the origin of the *Prdm* genes in animals. (A) The *Prdm* genes originate through the rearrangements between a single ancestral SET domain containing gene and several different genes encoding Zn fingers. The ancestral *Prdm* genes subsequently diversified through gene duplications. (B) The *Prdm* genes originate through the rearrangements between a single ancestral SET domain containing gene and a single gene encoding Zn fingers. Diversification subsequently occurred through gene duplications and gene rearrangements with different genes encoding Zn fingers.