Correction

Correction: Nagy, A., *et al.* Reassessing Domain Architecture Evolution of Metazoan Proteins: Major Impact of Gene Prediction Errors. *Genes* 2011, *2*, 449-501.

Alinda Nagy, György Szláma, Eszter Szarka, Mária Trexler, László Bányai and László Patthy *

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1113 Budapest, Hungary; E-Mails: nagya@enzim.hu (A.N.); szlama@enzim.hu (G.S.); szarka@enzim.hu (E.S.); trexler@enzim.hu (M.T.); banyai@enzim.hu (L.B.)

* Author to whom correspondence should be addressed; E-Mail: patthy@enzim.hu; Tel.: (361) 2-093-537; Fax: (361) 4-665-465.

We found some errors in the published versions of Figure S2, Figure S3 and Figure S8 of our paper [1]. The correct Figures are presented below.

Figure S2. Correction of the sequence of rat DCLK1_RAT by the FixPred protocol. The DA of DCLK1_RAT was found to differ from those of DCLK1_MOUSE and DCLK1_HUMAN: whereas the latter contain two DCX and a Pkinase domain, the rat sequence lacks DCX domains. The sequence DCLK1_RAT_CORRECTED was predicted by the use of alternative gene models and is supported by ESTs FN798821, CF978300 and CB798849. (a) Comparison of the domain architecture of DCLK1_RAT with those of the correct DCLK1_HUMAN, DCLK1_MOUSE and DCLK1_RAT_CORRECTED sequences. (b) Alignment of the sequence of DCLK1_RAT with the correct DCLK_HUMAN, DCLK1_MOUSE and DCLK1_RAT_CORRECTED sequences.
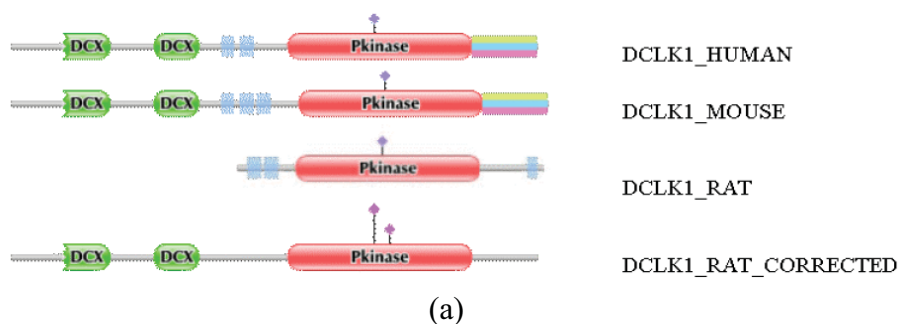

(a)

Figure S2. *Cont.*

```
                        1                                                   50
dclk1_rat_corrected     MSFGRDMELE HFDERDKAQR YSRGSRVNGL PSPTHSAHCS FYRTRTLQTL
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      MSFGRDMELE HFDERDKAQR YSRGSRVNGL PSPTHSAHCS FYRTRTLQTL
       dclk1_human      MSFGRDMELE HFDERDKAQR YSRGSRVNGL PSPTHSAHCS FYRTRTLQTL


                        51                                                  100
dclk1_rat_corrected     SSEKKAKKVR FYRNGDRYFK GIVYAISPDR FRSFEALLAD LTRTLSDNVN
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      SSEKKAKKVR FYRNGDRYFK GIVYAISPDR FRSFEALLAD LTRTLSDNVN
       dclk1_human      SSEKKAKKVR FYRNGDRYFK GIVYAISPDR FRSFEALLAD LTRTLSDNVN


                        101                                                 150
dclk1_rat_corrected     LPQGVRTIYT IDGLKKISSL DQLVEGESYV CGSIEPFKKL EYTKNVNPNW
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      LPQGVRTIYT IDGLKKISSL DQLVEGESYV CGSIEPFKKL EYTKNVNPNW
       dclk1_human      LPQGVRTIYT IDGLKKISSL DQLVEGESYV CGSIEPFKKL EYTKNVNPNW


                        151                                                 200
dclk1_rat_corrected     SVNVKTTSAS RAVSSLATAK GGPSEVRENK DFIRPKLVTI IRSGVKPRKA
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      SVNVKTTSAS RAVSSLATAK GGPSEVRENK DFIRPKLVTI IRSGVKPRKA
       dclk1_human      SVNVKTTSAS RAVSSLATAK GSPSEVRENK DFIRPKLVTI IRSGVKPRKA


                        201                                                 250
dclk1_rat_corrected     VRILLNKKTA HSFEQVLTDI TDAIKLDSGV VKRLYTLDGK QVMCLQDFFG
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      VRILLNKKTA HSFEQVLTDI TDAIKLDSGV VKRLYTLDGK QVMCLQDFFG
       dclk1_human      VRILLNKKTA HSFEQVLTDI TDAIKLDSGV VKRLYTLDGK QVMCLQDFFG


                        251                                                 300
dclk1_rat_corrected     DDDIFIACGP EKFRYQDVLL LDESECRVVK STSYTKIASA SRRGTTKSPG
        dclk1_rat       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
       dclk1_mouse      DDDIFIACGP EKFRYQDDFL LDESECRVVK STSYTKIASA SRRGTTKSPG
       dclk1_human      DDDIFIACGP EKFRYQDDFL LDESECRVVK STSYTKIASS SRRSTTKSPG


                        301                                                 350
dclk1_rat_corrected     PSRRSKSPAS TSSVNGTPGS QLSTPRSGKS PSPSPTSPGS LRKQRISQHG
        dclk1_rat       ~~~~~~~~MLE LIEVNGTPGS QLSTPRSGKS PSPSPTSPGS LRKQRISQHG
       dclk1_mouse      PSRRSKSPAS TSSVNGTPGS QLSTPRSGKS PSPSPTSPGS LRKQRISQHG
       dclk1_human      PSRRSKSPAS TSSVNGTPGS QLSTPRSGKS PSPSPTSPGS LRKQRSSQHG


                        351                                                 400
dclk1_rat_corrected     GSSTSLSSTK VCSSMDENDG PGE....... .........E ESDEGFQIPA
        dclk1_rat       GSSTSLSSTK VCSSMDENDG PGE....... .........E ESDEGFQIPA
       dclk1_mouse      GSSTSLSSTK VCSSMDENDG PGEGDELGRR HSLQRGWRRE ESEEGFQIPA
       dclk1_human      GSSTSLASTK VCSSMDENDG PGE....... .........E VSEEGFQIPA


                        401                                                 450
dclk1_rat_corrected     TITERYKVGR TIGDGNFAVV KECIERSTAR EYALKIIKKS KCRGKEHMIQ
        dclk1_rat       TITERYKVGR TIGDGNFAVV KECIERSTAR EYALKIIKKS KCRGKEHMIQ
       dclk1_mouse      TITERYKVGR TIGDGNFAVV KECIERSTAR EYALKIIKKS KCRGKEHMIQ
       dclk1_human      TITERYKVGR TIGDGNFAVV KECVERSTAR EYALKIIKKS KCRGKEHMIQ


                        451                                                 500
dclk1_rat_corrected     NEVSILRRVK HPNIVLLIEE MDVPTELYLV MELVKGGDLF DAITSTSKYT
        dclk1_rat       NEVSILRRVK HPNIVLLIEE MDVPTELYLV MELVKGGDLF DAITSTSKYT
       dclk1_mouse      NEVSILRRVK HPNIVLLIEE MDVPTELYLV MELVKGGDLF DAITSTSKYT
       dclk1_human      NEVSILRRVK HPNIVLLIEE MDVPTELYLV MELVKGGDLF DAITSTNKYT


                        501                                                 550
dclk1_rat_corrected     ERDASGMLYN LASAIKYLHS LNIVHRDIKP ENLLVYEHQD GSKSLKLGDF
        dclk1_rat       ERDASGMLYN LASAIKYLHS LNIVHRDIKP ENLLVYEHQD GSKSLKLGDF
       dclk1_mouse      ERDASGMLYN LASAIKYLHS LNIVHRDIKP ENLLVYEHQD GSKSLKLGDF
       dclk1_human      ERDASGMLYN LASAIKYLHS LNIVHRDIKP ENLLVYEHQD GSKSLKLGDF
```

(b)

Figure S2. *Cont*.

```
                        551                                                 600
   dclk1_rat_corrected  GLATIVDGPL YTVCGTPTYV APEIIAETGY GLKVDIWAAG VITYILLCGF
             dclk1_rat  GLATIVDGPL YTVCGTPTYV APEIIAETGY GLKVDIWAAG VITYILLCGF
           dclk1_mouse  GLATIVDGPL YTVCGTPTYV APEIIAETGY GLKVDIWAAG VITYILLCGF
           dclk1_human  GLATIVDGPL YTVCGTPTYV APEIIAETGY GLKVDIWAAG VITYILLCGF

                        601                                                 650
   dclk1_rat_corrected  PPFRGSGDDQ EVLFDQILMG QVDFPSPYWD NVSDSAKELI NMMLLVNVDQ
             dclk1_rat  PPFRGSGDDQ EVLFDQILMG QVDFPSPYWD NVSDSAKELI NMMLLVNVDQ
           dclk1_mouse  PPFRGSGDDQ EVLFDQILMG QVDFPSPYWD NVSDSAKELI NMMLLVNVDQ
           dclk1_human  PPFRGSGDDQ EVLFDQILMG QVDFPSPYWD NVSDSAKELI TMMLLVDVDQ

                        651                                                 700
   dclk1_rat_corrected  RFSAVQVLEH PWVNDDGLPE NEHQLSVAGK IKKHFNTGPK PSSTAAGVSV
             dclk1_rat  RFSAVQVLEH PWVNDDGLPE NEHQLSVAGK IKKHFNTGPK PSSTAAGVSV
           dclk1_mouse  RFSAVQVLEH PWVNDDGLPE NEHQLSVAGK IKKHFNTGPK PSSTAAGVSV
           dclk1_human  RFSAVQVLEH PWVNDDGLPE NEHQLSVAGK IKKHFNTGPK PNSTAAGVSV

                        701                                                 750
   dclk1_rat_corrected  IATTALDKER QVFRRRRNQD VRGRYKAQPA PPELNSESED YSPSSSETVR
             dclk1_rat  IATTALDKER QVFRRRRNQD VRGRYKAQPA PPELNSESED YSPSSSETVR
           dclk1_mouse  IATTALDKER QVFRRRRNQD VRSRYKAQPA PPELNSESED YSPSSSETVR
           dclk1_human  IATTALDKER QVFRRRRNQD VRSRYKAQPA PPELNSESED YSPSSSETVR

                        751
   dclk1_rat_corrected  SPNSPF
             dclk1_rat  SPNSPF
           dclk1_mouse  SPNSPF
           dclk1_human  SPNSPF
```

(b)

Figure S3. Evidence that SYWM_CAEEL is mispredicted. The Swiss-Prot SYWM_CAEEL sequence arose by *in silico* fusion of the gene encoding the worm ortholog of PEX10 proteins and the worm ortholog of SYWM proteins. Note that no EST supports the existence of the fusion protein and that separate translation of these genes is supported by EST sequences BJ806113 of *Caenorhabditis elegans* and EST DR782673 of *Caenorhabditis remanei*. (a) Alignment of the mispredicted fusion sequence SYWM_CAEEL with its corrected constituents, PEX10_CAEEL and SYWM_CAEEL_CORRECTED; (b). Alignment of the FixPred predicted sequence of worm PEX10_CAEEL with orthologous PEX10 sequences; (c) Alignment of the FixPred corrected sequence SYWM_CAEEL_CORRECTED with orthologous SYWM sequences.

```
                        1                                                   50
           pex10_caeel  MNTYVAEIGE IVRSQRRDEE YIEDITERLS RVSKELLGQR TWIRWFPYLK
            sywm_caeel  MNTYVAEIGE IVRSQRRDEE YIEDITERLS RVSKELLGQR TWIRWFPYLK
  sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

                        51                                                  100
           pex10_caeel  SIASTLYYTS TVVLGNQTLG EEYVHLFESN GLERTVPSIP SRISFVLLHS
            sywm_caeel  SIASTLYYTS TVVLGNQTLG EEYVHLFESN GLERTVPSIP SRISFVLLHS
  sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

                        101                                                 150
           pex10_caeel  AFPLISNYLI QKAESTLTHP STESFLGIPI RKNQKARQSF LDVFFWLRTK
            sywm_caeel  AFPLISNYLI QKAESTLTHP STESFLGIPI RKNQKARQSF LDVFFWLRTK
  sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

(a)

Figure S3. *Cont.*

```
              151                                                          200
pex10_caeel   LFPQLQRAHI ALFYITGAYY SIARRFTGIR FLSASAHSDI PALKVYRFLG
 sywm_caeel   LFPQLQRAHI ALFYITGAYY SIARRFTGIR FLSASAHSDI PALKVYRFLG
sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              201                                                          250
pex10_caeel   YITLIQLAVS IGISLYSFLE QEKFNNKLKK EKKENNGGSD RNLDENSLFH
 sywm_caeel   YITLIQLAVS IGISLYSFLE QEKFNNKLKK EKKENNGGSD RNLDENSLFH
sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              251                                                          300
pex10_caeel   PTFQCSICLE NKNPSALFCG HLFCWTCIQE HAVAATSSAS TSSARCPQCR
 sywm_caeel   PTFQCSICLE NKNPSALFCG HLFCWTCIQE HAVAATSSAS TSSARCPQCR
sywm_caeel_corrected  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~M IFSGKFTSHL

              301                                                          350
pex10_caeel   LEFQPRDVTP LLNL***~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   LEFQPRDPNN LRLLSTSTHP TIYFTGIQPT GIPHLGNFFG SIEPWTELQN
sywm_caeel_corrected  LNYGFK.PNN LRLLSTSTHP TIYFTGIQPT GIPHLGNFFG SIEPWTELQN

              351                                                          400
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   SVDKNILMML SVVDQHAISL GPLPANELRQ NTHQMTASLI ACGVDPNRTL
sywm_caeel_corrected  SVDKNILMML SVVDQHAISL GPLPANELRQ NTHQMTASLI ACGVDPNRTL

              401                                                          450
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   LFRQSDVPQI AQISWILGSL QTTSKLARLP QYKEKKERFK KGDIPVGLLT
sywm_caeel_corrected  LFRQSDVPQI AQISWILGSL QTTSKLARLP QYKEKKERFK KGDIPVGLLT

              451                                                          500
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   YPLLQAADVL TFKATTVPVG EDQSQHLNLL GGLAYAFNKT YETEIFPIPK
sywm_caeel_corrected  YPLLQAADVL TFKATTVPVG EDQSQHLNLL GGLAYAFNKT YETEIFPIPK

              501                                                          550
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   QLTRESHARI RSLREPEKKM SKSSGGPRSR IEITDSRSTI IEKCQKAQSD
sywm_caeel_corrected  QLTRESHARI RSLREPEKKM SKSSGGPRSR IEITDSRSTI IEKCQKAQSD

              551                                                          600
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   NAGKVTYDKE NRLAVSNLLD LYSAVTKTQT SEIDFSNWTT LDLKMNLAEA
sywm_caeel_corrected  NAGKVTYDKE NRLAVSNLLD LYSAVTKTQT SEIDFSNWTT LDLKMNLAEA

              601                                                          650
pex10_caeel   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 sywm_caeel   VDKRLAPIRQ KFEELQNTGE VDKVLTENGE KAREIAEKNL EEIRRTIGFL
sywm_caeel_corrected  VDKRLAPIRQ KFEELQNTGE VDKVLTENGE KAREIAEKNL EEIRRTIGFL
```

(a)

SYWM_CAEEL

PEX10_HUMAN

PEX10_CAEEL

Figure S3. *Cont.*
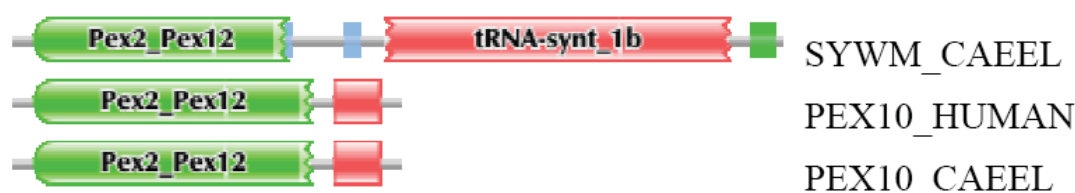
```
           1                                                        50
pex10_human  ~~MAPAAASP PEVIRAAQKD EYYRGGLRSA AGGALHSLAG ARKWLEWR..
pex10_macfa  ~~MAPAAASP PEVIRAAQKD EYYRGGLRSA AGGALHSLAG ARKWLEWR..
pex10_caeel  ~~MNTYVAEI GEIVRSQRRD EEYIEDITER LSRVSKELLG QRTWIRWF..
pex10_pican  MFKLLSFANA PAIVRANQKD SYFESRLHNQ LLDVVKAIKG SHFVHKYP..
pex10_schpo  ~MHLSAHIDP LQIILCTEID EACIQFIKSQ IEGIARA.CG PRMQANFEGV

           51                                                       100
pex10_human  ..KEVELLSD VAYFGLTTLA GYQTLGEEYV SIIQVDPSRI H...VPSSLR
pex10_macfa  ..KEVELLSD VAYFGLTTLA GYQTLGEEYV SIVRVDPSQT R...VPSWLR
pex10_caeel  ..PYLKSIAS TLYYTSTVVL GNQTLGEEYV HLFESNGLER T...VPSIPS
pex10_pican  ..EELRTLAT ALYLCLTTLV GSKTLGEEYV DLVYVSRDGR K...IPKFAS
pex10_schpo  LIPYVDVLGK FLY..RACCL RYATMGEEAA RIVLAKQDRS KGLVLATTGE

           101                                                      150
pex10_human  RGVLVTLHAV LP....YLLD KALLPLEQEL QADPDSGRPL QGSLGPGGRG
pex10_macfa  RGVLVTLHAV LP....YLLD KVLLPLEQEL QADPDSGRPS QGSLVPGGRG
pex10_caeel  RISFVLLHSA FPLISNYLIQ KA....ESTL T......HPS TESFL.....
pex10_pican  RFGFVVAYVL FPYAVRQLLQ K.LKAQQSRL .......... .......AQL
pex10_schpo  RMTSLIFSLV IDLVGVH.VN KLLKQASYSS SFKLPFG... LRNLLPEAVI

           151                                                      200
pex10_human  CSGARRWMRH HTATLTEQQR RALLRAVFVL RQGL.ACL.Q RLHVAWFYIH
pex10_macfa  CSGVRRWVRR HTATLTEQQR RALLRAAFVL RQGL.ACL.Q QLHVAWFYIH
pex10_caeel  ..GIP..IRK N.....QKAR QSFLDVFFWL RTKLFPQL.Q RAHIALFYIT
pex10_pican  VSGV...... .......... .SYMNVMDLL .......... NLHLALFYFT
pex10_schpo  SK......EK HLVYILNSFK PILLKLVSII R...FLCLTM KGHCA.....

           201                                                      250
pex10_human  GVFYHLAKRL TGITYLRVRS LPGEDLRARV SYRLLGVISL LHLVLS..MG
pex10_macfa  GVFYHLAKRL TGITYLRVRS LPGEDLRARV SYRLLGVVSL LHLVLS..VG
pex10_caeel  GAYYSIARRF TGIRFLSASA ..HSDIPALK VYRFLGYITL IQLAVS..IG
pex10_pican  GKYYQFAKRL FGLRYAFGYR VDKNQQRARG NYELLGLLII FQTVFKNVAN
pex10_schpo  ....TVSQLL LGLKYISLDE INPEEKK... ..KVLTLLLL L........G

           251                                                      300
pex10_human  LQ.LYGFRQR Q....RARKE WRLHRGLSHR RASLEERAVS RNP..L..CT
pex10_macfa  LR.LYGFRQR Q....RARKE WRLHRGLSHR RGSLEERAVS RNP..L..CT
pex10_caeel  IS.LYSFLEQ EKFNNKLKKE KKENNGGSDR ..NLDENSLF .HP..TFQCS
pex10_pican  LRKLWGATKT VQDSGDL..I YRFRDQTSDV IDLADPKVLP YLPEASRTCM
pex10_schpo  SRLIASILQH SNSYFDQHTI ....SSITDE RDLEDKNKLP FIPEGNRKCS

           301                                                      350
pex10_human  LCLEERRHPT ATPCGHLFCW ECI....... TAWCS.SKAE CPLCREKFPP
pex10_macfa  LCLEERRHPT ATPCGHLFCW ECI....... TAWCS.SKAE CPLCREKFPP
pex10_caeel  ICL.ENKNPS ALFCGHLFCW TCIQEHAVAA TSSASTSSAR CPQCRLEFQP
pex10_pican  LCLSPMKDPS CGECGHVFCW KCVLD..... ..WVK.ERQE CPLCRAKMRE
pex10_schpo  LCMEFIHCPA ATECGHIFCW SCI....... NGWTS.KKSE CPLCRAFSSP

           351      362
pex10_human  QKLIYLRHYR ~~
pex10_macfa  QKLIYLRHYR ~~
pex10_caeel  RDVTPLLNL* **
pex10_pican  SQLLPLR~~~ ~~
pex10_schpo  SKIILLR~~~ ~~
```

(b)

Figure S3. *Cont.*


```
                        1                                                    50
       sywm_human       ~~~~~~~MAL HSMRKARERW SFIRALHKGS AAAPALQKDS KKRVFSGIQP
       sywm_mouse       ~~~~~~~MAL FSVRKARECW RFIRALHKGP AATLAPQKES GERVFSGIQP
       sywm_yeast       MSNKQAVLKL ISKR.....W ..ISTVQRAD FKLNSEALHS NATVFSMIQP
sywm_caeel_corrected    ~~~~~~~~MI FSGKFTSHLL NY..GFKPNN LRLLSTSTHP .TIYFTGIQP

                        51                                                   100
       sywm_human       TGILHLGNYL GAIESWV... RLQDEYDSVL YSIVDLHSIT VPQ.DPAVLR
       sywm_mouse       TGILHLGNYL GAIESWV... NLQEEYDTVI YSIVDLHSIT VPQ.DPTVLQ
       sywm_yeast       TGCFHLGNYL GATRVWTDLC ELKQPGQELI FGVADLHAIT VPKPDGEMFR
sywm_caeel_corrected    TGIPHLGNFF GSIEPWTELQ NSVDKNILMM LSVVDQHAIS LGPLPANELR

                        101                                                  150
       sywm_human       QSILDMTAVL LACGINPEKS ILFQQSQVSE HTQLSWILSC MVRLPRLQHL
       sywm_mouse       QSILDMTAVL LACGINPEKS ILFQQSKVSE HTQLSWILTC MVRLPRLQHL
       sywm_yeast       KFRHEAVASI LAVGVDPEKA SVIYQSAIPQ HSELHWLLST LASMGLLNRM
sywm_caeel_corrected    QNTHQMTASL IACGVDPNRT LLFRQSDVPQ IAQISWILGS LQTTSKLARL

                        151                                                  200
       sywm_human       HQWKAKT.TK QKHDG..... .......TVG LLTYPVLQAA DILLYKSTHV
       sywm_mouse       HQWKAKA.AN RKHDG..... .......TVG LLTYPVLQAA DILCYKSTHV
       sywm_yeast       TQWKSKSNIK QSTNGDYLVN DSDVGKVRLG LFSYPVLQAA DILLYKSTHV
sywm_caeel_corrected    PQYKEKKERF KKGD...... ......IPVG LLTYPLLQAA DVLTFKATTV

                        201                                                  250
       sywm_human       PVGEDQVQHM ELVQDLAQGF NKKYGE.FFP VPESIL.TSM KKVKSLRDPS
       sywm_mouse       PVGEDQVQHM ELVQDLARSF NQKYGE.FFP LPKSIL.TSM KKVKSLRDPS
       sywm_yeast       PVGDDQSQHL ELTRHLAEKF NKMYKKNFFP KPVTML.AQT KKVLSLSTPE
sywm_caeel_corrected    PVGEDQSQHL NLLGGLAYAF NKTYETEIFP IPKQLTRESH ARIRSLREPE

                        251                                                  300
       sywm_human       AKMSKSDPDK LATVRITDSP EEIVQKFRKA VTDFTSE.VT YDPAGRAGVS
       sywm_mouse       SKMSKSDPDK LATVRITDSP EEIVQKFRKA VTDFTSE.VT YEPDSRAGVS
       sywm_yeast       KKMSKSDPNH DSVIFLNDEP KAIQKKIRKA LTDSISDRFY YDPVERPGVS
sywm_caeel_corrected    KKMSKSSGGP RSRIEITDSR STIIEKCQKA QSD.NAGKVT YDKENRLAVS

                        301                                                  350
       sywm_human       NIVAVHAAVT GLSVEEVVRR SAGMNTAR.Y KLAVADAVIE KFAPIKREIE
       sywm_mouse       NMVAIHAAVS GLSVEEVVRS SAGLDTAR.Y KLLVADAVIE KFAPIRKEIE
       sywm_yeast       NLINIVSGIQ RKSIEDVVED VSRFNNYRDF KDYVSEVIIE ELKGPRTEFE
sywm_caeel_corrected    NLLDLYSAVT KTQTSEI..D FSNWTTL.DL KMNLAEAVDK RLAPIRQKFE

                        351                                       388
       sywm_human       KLKLDKDHLE KVLQIGSAKA KELAYTVCQE VKKLVGFL
       sywm_mouse       KLKMDKDHLR KVLLVGSAKA KELASPVFEE VKKLVGIL
       sywm_yeast       KYINEPTYLH SVVESGMRKA REKAAKNLAD IHKIMGF~
sywm_caeel_corrected    ELQNTGE.VD KVLTENGEKA REIAEKNLEE IRRTIGFL
```

(c)

Figure S8. Correction of the sequence of the XP_416936 protein of *Gallus gallus* with the FixPred protocol. The DA of the GNOMON predicted protein XP_416936 was found to differ from those of GAS6_MOUSE, GAS6_RAT, GAS6_HUMAN: whereas the latter contain a signal peptide, a Gla, three EGF_CA, a Laminin_G_1 and a Laminin_G_2 domain, XP_416936 lacks the N-terminal signal peptide and Gla domain. The sequence XP_416936_CORRECTED was predicted by the use of ESTs CD217792, BM439645 and BU115578. (a) Comparison of the DAs of XP_416936, XP_416936_CORRECTED with those of GAS6_MOUSE, GAS6_RAT and GAS6_HUMAN. Note that some of the four EGF_CA domains of GAS6 proteins are detected with E-values >0.0001 and are not represented in the DA images generated by Pfam. (b) Alignment of the sequences of XP_416936, XP_416936_CORRECTED with those of GAS6_MOUSE, GAS6_RAT and GAS6_HUMAN.

(a)
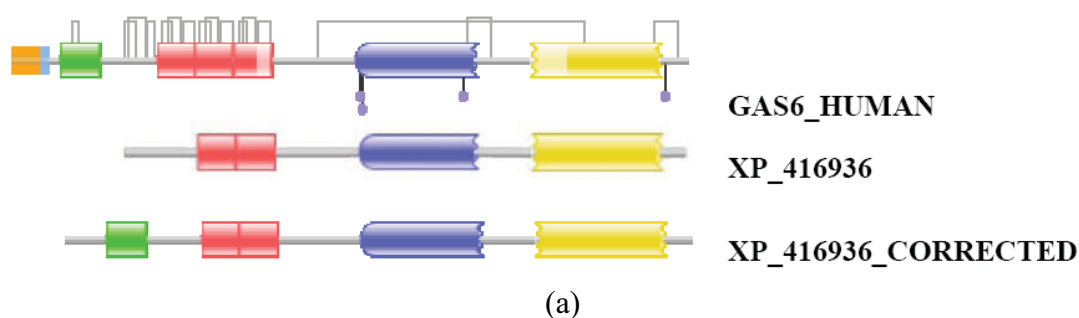
```
                    1                                                    50
         xp_416936  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
xp_416936_corrected  ~~~~~~~~~M PQPLGAALLL ALLAADCSQA VLLRAPEAAQ FLRQRQRRAY
        gas6_mouse  MPP..PPGPA A.ALGTALLL LLLASESSHT VLLRAREAAQ FLRPRQRRAY
          gas6_rat  MPP..PPGPT A.ALGTALLL LLLASESSHT VLLRAREAAQ FLRPRQRRAY
        gas6_human  MAPSLSPGPA ALRRAPQLLL LLLAAECALA ALLPAREATQ FLRPRQRRAF

                    51                                                   100
         xp_416936  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~MTV DVVGVEYFYP KYLACIHKYG
xp_416936_corrected  QIFEETKQGH LERECVEEHC SKEEAREVFE NDPETEYFYP KYLACIHKYG
        gas6_mouse  QVFEEAKQGH LERECVEEVC SKEEAREVFE NDPETEYFYP RYQECMRKYG
          gas6_rat  QVFEEAKQGH LERECVEEVC SKEEAREVFE NDPETDYFYP RYQECMRKYG
        gas6_human  QVFEEAKQGH LERECVEELC SREEAREVFE NDPETDYFYP RYLDCINKYG

                    101                                                  150
         xp_416936  SPYTRSPDFL TCVHNLPNQC SPDPCYKEGT VRCEDLKGDF YCECKRGWQG
xp_416936_corrected  SPYTRSPDFL TCVHNLPNQC SPDPCYKEGT VRCEDLKGDF YCECKRGWQG
        gas6_mouse  RPEEKNPDFA KCVQNLPDQC TPNPCDKKGT HICQDLMGNF FCVCTDGWGG
          gas6_rat  RPEDKNPNFA TCVKNLPDQC TPNPCDKKGT QLCQDLMGNF FCLCKDGWGG
        gas6_human  SPYTKNSGFA TCVQNLPDQC TPNPCDRKGT QACQDLMGNF FCLCKAGWGG

                    151                                                  200
         xp_416936  KTCDKDIDEC KVQHGGCNQI CLNKLGSYRC SCYSGYAL.K DSKTCEDIDE
xp_416936_corrected  KTCDKDIDEC KVQHGGCNQI CLNKLGSYRC SCYSGYAL.K DSKTCEDIDE
        gas6_mouse  RLCDKDVNEC VQKNGGCSQV CHNKPGSFQC ACHSGFSLAS DGQTCQDIDE
          gas6_rat  RLCDKDVNEC SQKNGGCSQV CHNKPGSFQC ACHSGFSLQS DNKSCQDIDE
        gas6_human  RLCDKDVNEC SQENGGCLQI CHNKPGSFHC SCHSGFELSS DGRTCQDIDE
```

(b)

Figure S8. *Cont.*

```
                     201                                                 250
           xp_416936 CAASADICGE ARCKNLISSY ECVCDAGYRY DEQRKTCDDI NECEERLCEQ
 xp_416936_corrected CAASADICGE ARCKNLISSY ECVCDAGYRY DEQRKTCDDI NECEERLCEQ
           gas6_mouse CTDS.DTCGD ARCKNLPGSY SCLCDEGYTY SSKEKTCQDV DECQQDRCEQ
             gas6_rat CTADS.DTCGD ARCKNLPGSY SCLCDKGYTY SSKEKTCQDV DECQQDRCEQ
           gas6_human CADS.EACGE ARCKNLPGSY SCLCDEGFAY SSQEKACRDV DECLQGRCEQ


                     251                                                 300
           xp_416936 MCVNSPGSYT CHCDGRGGVK LSQDMNTCE. .......... ..........
 xp_416936_corrected MCVNSPGSYT CHCDGRGGVK LSQDMNTCE. .......... ..........
           gas6_mouse TCVNSPGSYT CHCDGRGGLK LSPDMDTCE. .......... ..........
             gas6_rat TCVNSPGSYT CHCNGRGGLK LSPDMDTCE. .......... ..........
           gas6_human VCVNSPGSYT CHCDGRGGLK LSQDMDTCEL EAGWPCPRHR RDGSPAARPG


                     301                                                 350
           xp_416936 .......... .......... ..NIIPCVPF AVAKSVKSLY LGRMFSGTPV
 xp_416936_corrected .......... .......... ..NIIPCVPF AVAKSVKSLY LGRMFSGTPV
           gas6_mouse .......... .......... ..DILPCVPF SMAKSVKSLY LGRMFSGTPV
             gas6_rat .......... .......... ..DILPCVPF SMAKSVKSLY LGRMFSGTPV
           gas6_human RGAQGSRSEG HIPDRRGPRP WQDILPCVPF SVAKSVKSLY LGRMFSGTPV


                     351                                                 400
           xp_416936 IRLRFKRKQL TRLVAEFDFR TFDPEGILFF AGGHQDSTWV VLALRKGRLE
 xp_416936_corrected IRLRFKRKQL TRLVAEFDFR TFDPEGILFF AGGHQDSTWV VLALRKGRLE
           gas6_mouse IRLRFKRLQP TRLLAEFDFR TFDPEGVLFF AGGRSDSTWI VLGLRAGRLE
             gas6_rat IRLRFKRLQP TRLLAEFDFR TFDPEGVLFF AGGRSDSTWI VLGLRAGRLE
           gas6_human IRLRFKRLQP TRLVAEFDFR TFDPEGILLF AGGHQDSTWI VLALRAGRLE


                     401                                                 450
           xp_416936 LQLKYSGIGR VTSSGPLINH GMWQTISVEE LERSLVVKVN RDAVMRIAVS
 xp_416936_corrected LQLKYSGIGR VTSSGPLINH GMWQTISVEE LERSLVVKVN RDAVMRIAVS
           gas6_mouse LQLRYNGVGR ITSSGPTINH GMWQTISVEE LERNLVIKVN KDAVMKIAVA
             gas6_rat LQLRYNGVGR ITSSGPTINH GMWQTISVEE LDRNLVIKVN KDAVMKIAVA
           gas6_human LQLRYNGVGR VTSSGPVINH GMWQTISVEE LARNLVIKVN RDAVMKIAVA


                     451                                                 500
           xp_416936 GDLFTLDKGL YQLNLTVGGI PFKTKDLIVP INPRLDGCLR AWNWLNGEDS
 xp_416936_corrected GDLFTLDKGL YQLNLTVGGI PFKTKDLIVP INPRLDGCLR AWNWLNGEDS
           gas6_mouse GELFQLERGL YHLNLTVGGI PFKESELVQP INPRLDGCMR SWNWLNGEDS
             gas6_rat GGLFQLERGL YHLNLTVGGI PFKESDLVQP INPRLDGCMR SWNWLNGEDS
           gas6_human GDLFQPERGL YHLNLTVGGI PFHEKDLVQP INPRLDGCMR SWNWLNGEDT


                     501                                                 550
           xp_416936 TIQETIKMNE RMQCFAVAGR GSFYPGRGFA IFNLTYMQPS SRNETKTSWK
 xp_416936_corrected TIQETIKMNE RMQCFAVAGR GSFYPGRGFA IFNLTYMQPS SRNETKTSWK
           gas6_mouse AIQETVKANT KMQCFSVTER GSFFPGNGFA TYRLNYTRTS LDVGTETTWE
             gas6_rat AIQETVKANT KMQCFSVTER GSFFPGNGFA FYSLNYTRTS LDVGTETTWE
           gas6_human TIQETVKVNT RMQCFSVTER GSFYPGSGFA FYSLDYMRTP LDVGTESTWE


                     551                                                 600
           xp_416936 IEVNAVIQPA TDTGVMFALV TEDAS.VPLS LSLVDYHSTK KLKQQFVILA
 xp_416936_corrected IEVNAVIQPA TDTGVMFALV TEDAS.VPLS LSLVDYHSTK KLKQQFVILA
           gas6_mouse VKVVARIRPA TDTGVLLAL. VGDDDVVPIS VALVDYHSTK KLKKQLVVLA
             gas6_rat VEVVARIRPA TDTGVLMAL. VGDKDVVLLS VALVDYHSTK KLKKQLVVLA
           gas6_human VEVVAHIRPA ADTGVLFALW APDLRAVPLS VALVDYHSTK KLKKQLVVLA


                     601                                                 650
           xp_416936 VEDTVVSRLA LNLCDKKEHS VDILLKKDQL SLRVDGMEGE RELSTSELED
 xp_416936_corrected VEDTVVSRLA LNLCDKKEHS VDILLKKDQL SLRVDGMEGE RELSTSELED
           gas6_mouse VEDVALALME IKVCDSQEHT VTVSLREGEA TLEVDGTKGQ SEVSTAQLQE
             gas6_rat VENVALALME IKVCDSQEHT VTVSLRDGEA TLEVDGTKGQ SEVSTAQLQE
           gas6_human VEHTALALME IKVCDGQEHV VTVSLRDGEA TLEVDGTRGQ SEVSAAQLQE
```

(b)

Figure S8. *Cont.*

```
            651                                                   700
   xp_416936  SLSILESSLQ SPVKTYVGGL PDVNVTATPV TAFYHGCMTV KLRSKALDLD
xp_416936_corrected  SLSILESSLQ SPVKTYVGGL PDVNVTATPV TAFYHGCMTV KLRSKALDLD
   gas6_mouse  RLDTLKTHLQ GSVHTYVGGL PEVSVISAPV TAFYRGCMTL EVNGKILDLD
     gas6_rat  RLDLLKTRLQ GSVLTFVGGL PDVQVTSTPV TAFYRGCMTL EVNGKTLDLD
   gas6_human  RLAVLERHLR SPVLTFAGGL PDVPVTSAPV TAFYRGCMTL EVNRRLLDLD

            701              722
   xp_416936  EALYKHSDIT SHSCPPVEAG P~
xp_416936_corrected  EALYKHSDIT SHSCPPVEAG P*
   gas6_mouse  TASYKHSDIT SHSCPPVEHA TP
     gas6_rat  TASYKHSDIT SHSCPPVEHV TA
   gas6_human  EAAYKHSDIT AHSCPPVEPA AA
```

(b)

We apologize for any inconvenience caused to the readers.

References

1. Nagy, A.; Szláma, G.; Szarka, E.; Trexler, M.; Bányai, L.; Patthy, L. Reassessing Domain Architecture Evolution of Metazoan Proteins: Major Impact of Gene Prediction Errors. *Genes* **2011**, *2*, 449–501.