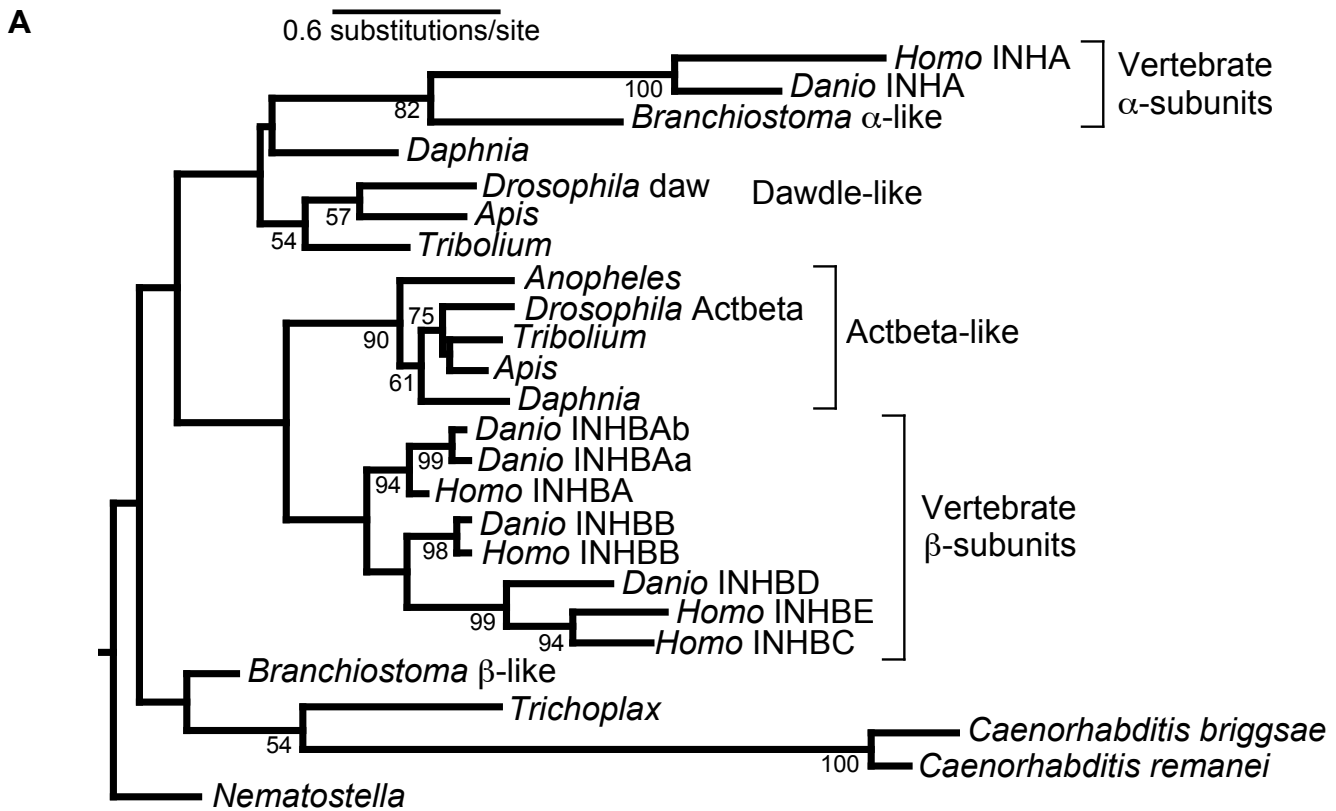


Figure S2. Large-scale estimate of inhibin/activin phylogeny. A) Maximum likelihood (ML) estimate of phylogeny for genes retrieved from annotated animal genomes by BLAST (Altschul et al., 2009) searches. The alignment included only the TGF β domain of genes that have human inhibin/activin genes as their top hit (they all satisfied the bidirectional best hit criterion). This analysis was used the WAG (Whelan and Goldman, 2001) model with the a proportion of sites assumed to be invariant and the remaining sites at different rates drawn from a Γ -distribution (with a shape parameter estimated by ML). Support for clades reflects the percentage of 100 bootstrap replicates. **B)** Alignment used for part A of this figure, with the 3 amino acid insertion that unites the lancelet (*Branchiostoma*) α -subunit-like protein with the vertebrate (in this alignment, human and zebrafish) inhibin α -subunit highlighted using yellow. **C)** Alignment for the lancelet α -subunit-like protein with the lancelet β protein mature domain sequence. **D)** Alignment for the mature domain sequences between the human inhibin α -subunit, chicken inhibin α -subunit and human inhibin β A subunit. The red box that used in the **C)** and **D)** highlighted the regions that we focused on.



B

```

Homo INHA      R -ARR-----STPLMSWPWSPSALRLLQRPPEEPAAHANCHRVALNISFQELGWERWIVPPSFIHYCHGGC-----GLHIP-----
Homo INHBA    R RRRR---GLECD-----GKV-NICCKQFFVDFKIDGNDWIIAPSGYHANYCEGEC-----PSHIAGTSGSLSLPH
Homo INHBB    R IRKR---GLECD-----GRT-NLCCRQFFIDFRLIGNDWIIAPTGYGYNYCEGSC-----PAYLAGVPGSASSPH
Homo INHBC    Q IHRR---GIDCQ-----GGS-RMCCRQEFFVDFREI GHWNDWIIQPEGYAMNFCIGQC-----PLHLAGMPTIAASPH
Homo INHBE    R ARRR---TPTCE-----PAT-PLCCRRDHYVDFELGWRDWWLQPEGYQLNYCSGGC-----PPHLAGSPGIAASPH
Danio INHA     R -SRR-----A-PKIPWSPDAIENKRPASQG-----TDCRRQIEISFEDLGNDWIIVHPKSFYFYCHGNC-----SSAER-----
Danio INHBAa  R RRRR---GLECD-----GKA-RVCCRQFFVDFKIDGNDWIIAPSGYHANYCEGDC-----ASNVASITGNSLSLPH
Danio INHBAb  R RSKR---GLECD-----GKI-RVCCRQFFVDFKIDGNDWIIAPSGYHANYCEGDC-----PSHVASITGNSLSLPH
Danio INHBB   R IRKR---GLECD-----GNGGLCCRQFFIDFRLIGNDWIIAPAGYGYNYCEGSC-----PAYMAGVPGSASSPH
Danio INHBD   A LSKR---SLRCG-----DDV-TVCCCKDFYIKFRDIQWDDWIIAPEGYHMNYCMGQC-----PQHLSGSPGIAASPH
Branchiostoma α-like R -HKRD--VSNDCNDEDEDEKDFISSRNETS-----PSP-RSCCRSLVSVFKIDGWDGWMEPAEFAHYCLGQC-----ASYNLSP-----H
Branchiostoma β-like DGRSNECCREKFYVDFKIDAWDDWII SPKGYANFCTGSC-----QGTILPRY-----HH
Anopheles    R LRRR---ALDCT-----GAPNEQCKQKQFYVDFKALKDDWII RPHGYANYCKGSC-----HL-ADRFSSEY
Apis daw-like N RPKR---NSNCL-----PEM-KECCRDLEYINFDIGNDWII LHPNGYHAYFCRGC-----SS--AASLTSGSLY
Apis Actbeta-like R VRRR---AIECS-----GAIKGCCQKQFYVDFKALGDDWIIAPQGYANYCRGDC-----AAG-HRTPDTFLNLY
Tribolium daw-like R RPKR---ATKCS-----AGR-NECCRESFYVDFADIGWDDWII SPAGYDAYFCRGC-----VT--ASALTLSATQH
Tribolium Actbeta-like R VRRR---ALDCS-----EDSGNCCQKQFYVDFKALGDDWIIAPQGYANYCRGDC-----GL--HRTPDVTVYH
Drosophila daw SRQR---SINCS-----SGM-TECCREHLYISFRDIGWSDWII LKPEGYANYFCRGC-----SS--VASVTQAASHH
Drosophila Actbeta R VRRR---AVDCG-----GALNGCCKESEFYVDFKALGDDWIIAPRGYANYCRGDC-----TGS-FRTPDTFTQFH
Daphnia daw-like SRQKRQSVTFRCG-----ATV-TQCCRESLVSFVKDVGWDDWII VAPSGYHAYFCRGC-----RT--MTAPASSANTH
Daphnia Actbeta-like R -RRT---PLDCS-----PTT-SACCKQFFVDFRDLGDDWIIAPSGYANYCRGDCGGGAGGAGGAR---A-PEFSLSYH
C. remanei   KSDCI-----GCCIVPFYVNFTEIGWDDWII LKPEGYANYCSGGC-----SN--ELDENYQMMK
C. briggsae  KSNCD-----GCCITPFYVNFTEIGWDDWII LYPFGWNYCSGGC-----NE--KSDNYEIMK
Trichoplax   VNACR-----KQS-THCCVEEYINFEIGWSDWII SPQGTANYCRGTC-----FGMDSFGY-----R
Nematostella K QRKR---AIHCS-----PQM-RECCQKQFFVDFEEMGDDWII LVPKGFNANYCTGSC-----YGHILPVY-----HH

Homo INHA     - - - - -PNLSLFPVPGAPPTPAQPYSLPGAQPCCAALPGTMRPLHVRTTSDGGYSFKYETVPNLLTQHCACI
Homo INHBA    S TVINHYRMRG-----HSPFANKLSCCVPTKLRPMSMLY---YDQNI IKKD I QNMIVEECGCS
Homo INHBB    T AVVNQYMRG-----LNP-GTVNSCCIPTKLRMSMLY---FDDEYNIVKRDVPMNIVEECGCA
Homo INHBC    T AVLNLLKANT-----AAGTGGGSCCVPTARRPLSLLY---YDRDSNIVKTDIPDMVVEACGCS
Homo INHBE    S AVFSLKANN-----PWP--ASTSCCVPTARRPLSLLY---LDHNGNIVKTDVPMVVEACGCS
Danio INHA     - - - - -ITPILGITQCCAPVPEMSKSLRFTTSDGGYSFKYETLPI N IPEECNCI
Danio INHBAa  S TVISHYRIRG-----YSPFNISCCVPTLRAMSMLY---YNEEQI IKKD I QNMIVEECGCS
Danio INHBAb  S TVINHYRMRG-----YSPFNISCCVPTLRAMSMLY---YNEEQI IKKD I QNMIVEECGCS
Danio INHBB   T AVVNQYMRG-----MSP-GVNSCCIPTKLRMSMLY---FDDEYNIVKRDVPMNIVEECGCA
Danio INHBD   A SVFSQKANG-----IN--TAVSCCVPIQRPLSMVY---FNSQHTIVKTDVPMNIVEECGCT
Branchiostoma α-like A A IMDQV-----VSPVYEMKCCVVPVPGQMDLLI RYSFDGGKTISTQHIPNI IVNTCGCL
Branchiostoma β-like T SVLQRVALSQKD-----RETRKLTPTCCPTPKMSALSMLY---FDNDQVIFNKNLPMNIVEECGCS
Anopheles    H VYIDQYRRQGGAGNRGLGKMHKAGGGGGGAGGAG-----LAGTHQCCAPVKYSPMSLIF---YGPDR I I KQDLAKMIVEECGP
Apis daw-like T HVIEEYRKMD-----RLAGMPCCAPLFPSPMSLIF---YGPSN I I KRDLPKMVVEECGP
Apis Actbeta-like N NVIRKLNKNGMT-----IHRKNEIVPCSPQLSPQLLY---VDSNNTITQKTLPMNIVEECGCM
Tribolium daw-like T HVIEEYRKTH-----HLSGMPCCAPLFPSSMSLIF---YGPDMT I I KRDLPKMVVEECGP
Tribolium Actbeta-like N SILQKVMYQGNKN-----RGRSPEITPCCAATQFQPLQVY---MDNNTMTKVLNIVEECGCV
Drosophila daw S SIKMLLSTSGA-----NKSLEIVPCCAKYQSSQLVY---MDSNTATVTKTLPNIVEECGCR
Drosophila Actbeta A HFIEEYRK-----MGLMNGHRPCAPIKFSSMSLIF---YDGD-GI I KRDLPKMVVEECGP
Daphnia daw-like A SLQKLVARERVP-----IARPHLAPCCAPTRLSPLVIFY---SDENNVIKQRTLPNIVEECGCSL
Daphnia Actbeta S HVIDQ000000-----MFANYLQCCCAPTKFSSMSLIF---FGPDM I I KRDLPKMVVEECG
C. remanei   L SSRDQ-----SMIPEPCAPNYYSIDMIV---ALTFQD I RKTTHGMRALSCST
C. briggsae  L LVTDQ-----SLIPKPCAPNYYSIDIII---ALSSVD I RTRAHGMRALSCS
Trichoplax   L KILQKL-----SEEDNSLIPCCSPGKMSLTHIY---QDSNGYLSHWI LPMNIVEECQCF
Nematostella T EVIQKVAL-----RKT-RELSPCCAPTRKMPDLSLLY---YDKDENLQENVSNMIVEECGCS

```

C

```

N-terminal Wrist region
Lancelet α-like R -HKRDVSNDCNDEDEDEKDFISSRNETSPPSR-SCCRRLSVSVFKIDGWDGWMEPAEFAHYCLGQASYNLSPSHPAAIMDQV-----VSPVYEMKCC [94]
Lancelet β      ??????D-----GRSN-ECCREKFYVDFKDIANDDWIISP KGYANFCTGSCOGTILPRYHHTSVLQRVALSQKDRTRKLTCC [85]

```

D

```

N-terminal region
Human mature α STPLMSWPWSPSALRLLQRPPEEPAAHANCHRVALNISFQ~ELGWERWIV 49
Chicken mature α SA~~~~VPWSPAALSLLQRPSEDVAAHTNCRRASLNISFE~ELGWDNWIV 45
Human mature βA GLECDGKV~~NICCKQFFVSVFK~DIGWNDWII 30

Wrist region
Human mature α YPPSFIFHYCHGGCGLHI*PPNLSLFPVPGAPPT~~~~PAQPYSLPGAQP 94
Chicken mature α HPSSFVVFHYCHGNCAEGHG~~~~~LSHRLGVQL 73
Human mature βA APSGYHANYCEGECPSHI*AGTSGSSLSLFSHSTVINQYRLRGHNPFANLKS 79

Human mature α CCAALPGTMRPLHVRTTSDGGYSFKYETVPNLLTQHCA*CI 134
Chicken mature α CCAALPGTMRSLRVRTTSDGGYSFKYETVPNILAQDCTCV 112
Human mature βA CCV~PTKLRPMSMLY~DDGQNI IKKD I QNMIVEECGCS 116

```

References:

- Altschul, S.F., Gertz, E.M., Agarwala, R., Schäffer, A.A., and Yu, Y.K. (2009). PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res* 37, 815-824.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.
- Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution* 61, 2743-2749.
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18, 691-699.