

# Supporting Information

Heffer et al. 10.1073/pnas.1010746107

## *Bombyx mori* Ftz (Bm-Ftz)

MSSAATTNNRVSMDWWNNYNTSDPNSKQSKVNYKQYPTMANNYYQNYYPHNGCYGNW  
LSANCSSEKLCGTPTEGIVKSEPNWHGYSANLQTGSNGMDAIKKWQEMMNQYTQQQYNNY  
YNQTNMLCANTSVEKEVTDVRSMNSPFGCSISENYGSPQSFTSDFKSISPLGDDSPH  
LTTPKTTSQNPYFFACDNSTREISQVQVYENNNWEKNNETTSGKECNLSQFHGGF  
QKVDDQTSIKKDAVGAPSPNEVAQSSMGAEPQCDTMTRVQAVGDNADYTEENKMAAP  
EVQAYYPWMKGVSGDTKKEGSKRTRQTYTFQTLLEKEFHFNKYLRRRIEVSHALG  
LTERQIKIWFQNRRMKAKKDGLTTSPEPEFALDIGATKLGNNVSEYIDPRQQIGLPEY  
PNYHAGSVQPNIGHIPENMNSHCMMPYGGMLPKM

## *Apis mellifera* Ftz (Am-Ftz)

MTDSTFAAEMDYFWLVEATEHDDNDIYDLNLEDICPOTSQUESTSSGSGSGSGSGSHSA  
MNSVQSVTVQANQGMEMAQNQMTQSGA1GMEMAQNQMTQNQMTQGQMMGMEEMAQNQMT  
QSQMMEMAQCQMAQGQMMGMEEMAQNQMTQSGQMMGAQNQAMGQSESVESSSL  
DKPMEMEAQCGQPLCQESPQSPSTSSSVSSNDGSLVNVER  
QRQVNVNQRVPVTQSIINYSGYVQNQVYPIRQPLILNNSCMQOATATTFQPQPSCHANL  
NGNLNVRKQNGIRSNLNRNEQQQQQQQLMRQQQLQQQQLQQQQLRLHHQQQLHQQMLL  
QQQQQLQQQQQLQQQQQLHQQMLLQQQQQLHQQQLLQQQQQHFQQQQQLHQQMLLQQQ  
QQQQHQQLQQOEYQQQQQRAEEQRSSLQQQQPOLYLRQQENQMQMLQOPHTSRQQIQSQ  
TOPOTLAQSPQIQSOSQPQTOQQQPQOSQOMNYSWIKPSTSKEGETSNNNNGQKRTRQTY  
TRYQTYLEKEFHFNRYLTRNRLIEIARALHLTERQVKLWFQNRRMKAKKFNKDAENG  
DKGTGRPRRNQSFTRNVTATATRATVITATVATATATVTSINGAASTAAHYQTATAM  
PSCMQQQPPQQHQ  
NQPPL

## *Callosobruchus maculatus* Ftz (Cm-Ftz)

MSASQFGSCSEYYNQVGYNFYDQARMPGYYGNNGYQYHNAYPYGGYTEKRDAFAEAYGD  
VKEEPSSACRFDAHANQGYSNPVCEPDDSIIRRSPVINQYQPTGYGSLATSLSPPRAANE  
DDSTTGSSAAEKTGKMEDESSA  
SKPGEKITYDTELKTHSPADYEVHASNMS  
LDCDELSSCGKEKTTSEAEDALTAAQNNYE  
YPWMRSNDHTAKCNKTRQTYTRYQTYLE  
LEKEFHFNKYLTRRRRIEIAHTLCLTERQIKIWFQNRRMKAKKFDKLAVPAQOVDPTI  
QDVNMNQHLYPAMSPATPASYYSCGSSEGLPETSTLDANRNSFDANRNSFDAEVAPLT  
ALKNIPEGPPLSPPO

## *Dermestes maculatus* Ftz (Dmac-Ftz)

MSASTGNYDYWSQHPPTYQQYRSNIPLLSSSERPLTNVSPASLNYYNEIDNYRSVTN  
LNGFNPYGYMNEGLLKTVNFKLRSTVNDYGISADIISNNPEPINTHSENNYIQNFT  
SPNFQVHPNSGGLNDATNISPBMSTDITSPKKEIEDDSPA  
TKPHIRKPYDFYE  
TNKPDYQNFYSHVNEFACNKNIKTTPFAVIPQDEINSENISNTNSVTPTNNI  
YPWMKANAEGATNHGGKRTQTYTRYQTYLEKEFHFNKYLTRRRRIEIAHALCLSERQIKIW

FQNRRMKAKKDNKFTIQLQEFTEDINMNQNQLIANSPCANNALYMSNVSPQESTSTGGQEPN  
ALNEGIVEALTQFRNISGPPCIS

## *Forficula auricularia* Ftz (Fa-Ftz)

MKNSGKFILIAYLMAIS  
TPQRKTTSGSKRSRQTSRYQTYLEKEFHFNKYLTRRR  
IEIANAQLETERQIWIWFQNRRMKAKKTRSTEADMNLAAITALTESAFLPKANLAEPMF  
PLTSPMVNTTEPVPPDSDPQDLGVLGATIGPGKAPLVTQQLNSTLQPTQEDPAKLPSON  
RVHRDWPIMLRRGR

## *Thermobia domestica* Ftz (Td-Ftz)

MSSAAYFSNGSGVTTNCWGSTNGGLSSHEQNPNYQPFYFTHPASSTSKYGLVSSTYSE  
HHLPVLTGTPSHHPFPVPRYSTSPSVAATATNTPHNPNTLGRKPDSEPTTTTTESSP  
PITSTTPVSVATTANNNNNLQPDSDFFSSARTNDHSPPSVSQLFMDSGRDHLANGCKV  
SSFCAPNNIVGPDSLMLVQQGFDVTRPLDCLQQPFVGKGPANY  
TPWMKSYTDTHGHPK  
RTRQTYTRFQTYLEKEFHFNKYLTRRRRIEIAHSLGLSERQIKIWFQNRRMKAKKEIK  
MQPQFVNSGNTEDDILEKGMMATPPDAQVFQDKDVIMQNQIHIPFAGIKPENLQFPTIKE  
EFQHQTDISCISSDT

## *Pedetontus saltator* Ftz (Ps-Ftz)

MKSQAVGRHIY  
1PQMSVYSSPGPKRTRQTYTRVQTYLEKEFHFNRYLTRRRRIEIAH  
VLAGTERQIKIWFQNRRMKAKKESKLQEVREHEYVGQDSTVSETPVSATPSTESIKILE  
HSIPPIKVEAGMLINS

## *Folsomia candida* Ftz (Fc-Ftz)

MVTSPESNSSISPLRDVKSEKNMSPDGKEKDVGSTRIE  
YPPWLMRGSYGLKNTTSPRSP  
SSEDNISPSSSKRTRQTYTRCQTYLEKEFHFNKYLTRRRRDLAKMLTLSERQIKIW  
FQNRRMKAKKEVKGHVVASDLVQRHGNTNSENSCYGECTSS

## *Artemia salina* Ftz (As-Ftz)

MNPYFLPSQFPQSPFFGTQNTDVNNNDGSKKFFQACFQPRQINWACDFKSDYDAQKDLTHN  
TYVERSENPQHCMRSGYYPNTFVQFSTPGFVPTQHQMNSNNSIAPLQGITIPMPGOKRT  
RQTYTYQTYLELEKEFELYNRYLTRVRMDISSLQQLTERQIKIWFQNRRMKAKKENKNE  
TNFRSSQSCDASDEMVTSSMTQ

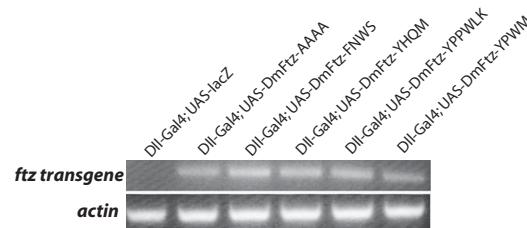
**Fig. S1.** Ftz protein sequences from diverse arthropods. Predicted Ftz protein sequences from *Bombyx mori*, *Apis mellifera*, *Callosobruchus maculatus*, *Dermestes maculatus*, *Forficula auricularia*, *Thermobia domestica*, *Pedetontus saltator*, *Folsomia candida*, *Artemia salina*, as indicated. Genomic and cDNA sequences were isolated by RACE and modified, gene-specific AFLP (Materials and Methods). Motifs of interest are highlighted (LXXLL, green; YPWM, blue; homeodomain, gray). For all sequences isolated or annotated from genomic DNA, introns contained consensus splice donor (GT) and acceptor (AG) sites, and a polypyrimidine stretch at the 3' end. The linker region between YPWM and homeodomain varied between 5 and 14 residues (Table S1).

**Antennapedia**  
*D.mel* -LYPWMR--RKRGQTYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-  
*A.mel* -LYPWMR--RKRGQTYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-  
*T.cas* -LYPWMR--RKRGQTYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-  
*A.pis* -LYPWMR--RKRGQTYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-  
*D.pul* -LYPWMR--RKRGQTYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN-

**Sex combs reduced**  
*D.mel* -IYPWMK--TKRQRTSYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-  
*A.mel* -IY<sub>1</sub>WMK--TKRQRTSYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-  
*T.cas* -IYPWMK--TKRQRTSYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-  
*A.pis* -IYPWMK--TKRQRTSYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH-  
*D.pul* -IYPWMK--TKRQRTSYTRYQTELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKHG-

**Ultrabithorax**  
*D.mel* -FYPWMA--RRRGQTYTRYQTELEKEFHTNHYLTRRRRIEAMAHALCLTERQIKIWFQNRRMKLKKEI-  
*A.mel* -FYPWMA--RRRGQTYTRYQTELEKEFHTNHYLTRRRRIEAMAHALCLTERQIKIWFQNRRMKLKKEI-  
*T.cas* -FYPWMA--RRRGQTYTRYQTELEKEFHTNHYLTRRRRIEAMAHALCLTERQIKIWFQNRRMKLKKEI-  
*A.pis* -FYPWMA--RRRGQTYTRYQTELEKEFHTNHYLTRRRRIEAMAHALCLTERQIKIWFQNRRMKLKKEI-  
*D.pul* --FYPWMA--RRRGQTYTRYQTELEKEFHTNHYLTRRRRIEAMAHALCLTERQIKIWFQNRRMKLKKEI-

**Fig. S2.** Conservation of YPWM motif and homeodomain in other Hox proteins. The residues flanking the YPWM motif and homeodomain sequences of Antennapedia, Sex combs reduced, and Ultrabithorax are highly conserved. Sequences were obtained from the genomes of *Drosophila melanogaster* (*D.mel*), *Apis mellifera* (*A.mel*), *Tribolium castaneum* (*T.cas*), *Acyrtosiphon pisum* (*A.pis*), and *Daphnia pulex* (*D.pul*). Only nonsynonymous amino acid substitutions are highlighted.



**Fig. S3.** Expression levels of “degen-YPWM” transgenes are similar. A *DII*-Gal4 driver was used to express degen-YPWM transgenes in developing imaginal discs during *Drosophila* development. L1 larvae were collected, and cDNA was made from 1  $\mu$ g of RNA. RT-PCR confirmed a *ftz* transgene product was detected in all degen-YPWM lines, but was absent in the control line (*lacZ*). Actin levels in all samples were similar. For each transgene, only one antennal-to-leg phenotype was seen; there was no variation.

**Table S1. The *ftz* genes generally contain small introns and encode short linker regions between the YPWM motif and homeodomain**

Species	Intron size (bp)	# Residues in linker
<i>D. mel</i>	150	11
<i>A. gam</i>	59	11
<i>A. aeg</i>	60	?
<i>B. mor</i>	89	10
<i>A. mel</i>	1657	13
<i>N. vit</i>	157	14
<i>T. cas</i>	50	9
<i>D. mac</i>	60	9
<i>A. pis</i>	224	?
<i>S. gre</i>	>679*	11
<i>F. aur</i>	335	5
<i>P. sal</i>	141	8
<i>S. car</i>	150	10
<i>D. pul</i>	?	8

\*Complete intron sequence not published.