SUPPLEMENTARY MATERIAL 3

		:	10	20	30	40	50	60	
				.					
	Co_1	NTPRRSNLR	KASVAILKQ	WLLDHV:	SNPYPT <mark>DIEKD</mark>	ALAQATDLN	VSQVNNWFINA	RRRILQPL	
	Co_2	RSKQRPNLS	REVVEVLKN	WLFAHS:	SRPYPS <mark>DVE</mark> KT	AMMAETGISI	LLQLNNWFINA	RRRLLQKP	
	Mb_1	NTGGRNNMP	HEVTSRLKE	WFFAHT	SHPYPS <mark>E</mark> QKKR	ELASQCDLTI	LQQINNWFINA	RRRLLNRP	
	Mb_2	SRHCTKRFA:	SSSIDTLKE	WLFAHT	DRPYPTDQDKT	ELMQQTGLD	LMQINNWFINA	RRRLLVKV	
	Aq_TALE	KGKKREKTS	PASQKLLKE	WLFSHSI	RCPYPTEDDKQ	NLCRMTGLSI	LQQLNNWFINA	RRRILPQK	
Maia	Dm_Hth	NQKKRGIFP	KVATNILRA	WLFQHL!	THPYPS <mark>ED</mark> QKK	QLAQDTGLT	ILQVNNWFINA	RRRIVQPM	
weis	NV MEIS	SQKKRGIFP	KAATNIMKA	WLFQHL!	THPYPS <mark>EE</mark> QKR	SLAQETGLT	ILQVNNWFINA	RRRIVQPM	
	Co_PBC-like	SRKRRINLS	REAQQVLND	WFLAHI	EHPYPS <mark>E</mark> SEKE	QLADQTNLT	MRQISTWFANK	KRNRQAQD-	TALE
PBC	Dm Exd	ARRKRRNFS	KQASEILNE	YFYSHL	SNPYPS <mark>EEA</mark> KE	ELARKCGIT	VSQVSNWFGNK	KRIRYKKNI	
	Nv PBX	ARRKRRNFS	KQATEILNE	YFYSHL	SNPYPS <mark>EEA</mark> KE	ELARKCNIS	VAQISNWFGNK	KRIRYKKNI	
T	Dm_Vis	LRKRRGNLP	KSSVKILKR	WLYEHR	YNAYPS <mark>DAE</mark> KF	TLSQEANLT	VLQVCNWFINA	RRRILPEM	
Igif	Nv_Tgif	PKRRRGNLP	KDSVNVLRL	WLWEHR	FNAYPS <mark>EAE</mark> KQ	YLSKAANLS	VLQVCNWFINA	RRRILPDM	
	Dm Ara	LAARRKNAT	RESTATLKA	WLNEHKI	KNPYPT <mark>KGE</mark> KI	MLAIITKMTI	LTQVSTWFANA	RRLKKEN	
Irx	Nv Irx	LAARRKNAT	RETTSTLKA	WLFEHR	KNPYPT <mark>KGE</mark> KI	MLAILTKMTI	LTQVSTWFANA	RRRLKKEN	
	AgIrxa	SAAGSITRR	MRNTAVLVK	WIEDHO	SNPYPT <mark>KAE</mark> KQ	YLAYYSGMN	MTQLSTWFANA	RRIKKIG	
	~								
	Aq Six1/2	GEETSYCFK	EKSRVVLRQ	WYTI	KNAYPSPREKR	QLAEQTGLT	TTQVSNWFKNR	RRORDRAAE	
	Dm Six1	GEETSYCFK	EKSRSVLRD	WYSI	HNPYPSPREKR	DLAEATGLT	TTQVSNWFKNR	RORDRAAE	
Six	Nv Six12	GEETSYCFK	EKSRNILRE	WYSI	HNPYPSPREKR	ELAEGTGLT	TOVSNWFKNR	RORDRAAE	
	Dm Six3	GEOKTHCFK	ERTRSLLRE	WYL	DPYPNPTKKR	ELAKATGLNI	PTOVGNWFKNR	RORDRAAA	
	Nv Six36	GEOKTHCFK	ERTRSLLRE	WYL	DPYPNPTKKR	ELAOATGLT	PTQVGNWFKNR	RORDRAAA	
	Co 8	MRSFRLKKS	SEOVRVLEA	FFRI	- EFPKPRKYOVI	ALCNDTGLL	TEVRNWFRNR	RLKDAKLR	
	Co 7	KKKSKSELD	EKOLRRLNR	AFNI	EDSRPDDEOLA	AIAAKVGLS	EEEVYSWFKAC	RKLTROSN	
	Co 5	KTRRRTTIT	LEOLGMLEE	AFN	RNNLPDCFORT	EVSRTTGLS	RVIRIWFONR	RAKORRSE	
	Co_4	SKRRRTOIS	EOOVRELEM	LFD	VDPWPSAEDKI	ALSRRLELS	OSVOVWFONR	RARAKROD	
	Ny POU4	KKRKRTSIG	AAEKRSLEA	YFA	MNPRPSSDKTA	STAEKLDLS	KNVVRVWFCNC	ROKKKRMK	
	Ny POU1	RRKRRTTIG	LAAKEALEN	HFM	KOTKPSSPEIV	RTADGLELD	EVVRVWFCNR	ROREKRVK	
	Ag Poul	HRKRRTTIG	MSAKERLEO	HF = -0	VOPKPSSSDIT	KVADSLNLD	EVIEVWECNE	ROREKRVR	
POU	Ag PouVI	KRKKRVVYT	PHALSTLNK	YFI.	KEPRPNROTTE	MVAEELDIJ	PEEVRVWFCNK	ROKYKTSN	
	Dm Nub	REKKETSIE	TTTRGALEK	AFI.	NOKPTSEETT	OLADRISME	EVVRVWFCNR	ROKEKRIN	
	Dm vvl	KRKKRTSIE	VSVKGALEO	HFH	KOPKPSAOEIT	SLADSLOLE	EVVRVWFCNR	ROKEKRMT	
	Ny Tsi	PTRVRTVLNI	EKOLHTLET	CYN	ANPRPDAMMKE	OLVENTGLS	PRVIRVWFONK	RCKDKKKN	
	Dm Lmx	PKRPRTTLN	TOORRAFKA	SFE	VSPKPCRKVRE	NLAKDTGLSI		RAKVKKTO	
	Ny Lmx	PKRPRTILT	SOORKVEKS	AFE	ISSKPCRKVRE	ELSRETGLS	VRVVOVWFONC	RAKVOKSV	
	Ny I.TM	SKRARTEIS	NDOLAFLKV	AVA	SPETTLEDRE	RIAKETGLD	MRVVOVWFONR	RAKOKRLS	
LIM	Ag Lin11	KGKTRTSIN	PKOLTVLOA	TYE	KEPRPSRAMRE	DLAAOTGLT	KVTOVWFONR	RSKDKKDG	
	Ag Lim3	OKRPRTTIS	KOLDLIKT	AYC	VSPKPSRHVRO	ELSOKTGLO	MRVVOVWFONK	RAKDKRTK	
	Dme Lim1	SKRPRTTIK	AKOLEVIKT	AFN	TPKPTRHIRE	OLAKETGLP	MRVTOVWFONK	RSKERRMK	
	Dm Lim3	NKRPRTTTT	AKOLETIKT	AYNI	SPKPARHVRE	OLSODIGLD	MRVVOVWFONR	RAKEKRLK	
	Ag 050d	PKRTRTAYS	NSOLDOLEL	TFA	TTHYPDVFTRE	DISBRIGTR	DRIOVWFONR	RARFRKOE	non-
	Ag 050b	TKKKRMTYT	KOOKDALES	YFY	DSYPDTOARE	NMSEALGITI	PEKVOVWFONR	RAKCRKRE	TALE
	Ag 050c	PKKTRTOFS	PKOLVYLEE	CFLI	KNRFPSAKERE	STAFFLDLT		RAKHRRKS	
	Ag 050a	RKRNRTVYS	TDOLKOLED	SFK	NPYPDKLMRD	NLAFELDMS	KKVNVWFONR	RVKLKKHS	
	Co paired-like	ASTNKRYFL	SDOLELLES	FYD	ONKEPKPSDEE	ALAAKMDESI	RARTOOWFRNR	RAKERRTO	
Dred	Ny dmbx	SOTTRTRFT	PYOTOVINE	TFS	SSAYIDASTCG	OLARLIGIS	SRSTOTWFKNK	RYKLRTOA	
Pra-	Dm otp	OKRHRTRFT	PAOLNELER	CFSI	XTHYPDIFMRE	ETAMRIGIT	SRVOVWFONR	RAKWKKRK	
like	Ny otp	OKRHRTRFT	PAOLNELER	CFA	RTHYPDVFMRE	ELAARIGIT	SRVOVWFONR	RAKWKKRK	
	Dm repo	KKKTRTTTT	AYOLEELER	AFE	RAPYPDVFARE	ELATKINIS	SRVOVWFONR	RAKWRKHE	
	Dm rx	HRRNRTTFT	TYOLHELER	AFE	KSHYPDVYSRE	ELAMKUNLP	VRVOVWFONR	RAKWRROE	
		LORNRTSFT	NDOIDSLEK	EFE	RTHYPDVFARE	RIACKICLP	CARTOVWESNE	RAKWRREE	
	Ny PAXC	LERNETTET	PDOLEMLEK	EFE	KSHYPDVATRE	ELANKIDMS	ARVOVWESNE	RAKWRRHO	
	Ny By	ORBNRTTFT	KOOLOELEK	VFE	KHYPDIALRE	ELAAKINISI	CARTOVWFONE	RAKWRKLO	
	Ny Emxa	PKRIRTAFT	PTOLLHLEN		KNHYIVGTERK	OLASYLNIS	TOIKVWFOND	RTKWKROO	
	Ny Hox1	KHRKRMAV	RIOUTEU	EF	FTRYI.TKFPPT	EMARMIDIA	SBUAKIMEUND	RWKMKKDM	
	Dm AbdA	RRRCROTVT	REOTIETER	EF	NHYLTRERKI	ETAHALCI.T	SBUIKIMEUND	RMKI.KKFT	
	Dm Antr	RKBCBOTVT	RVOTIFIER	EF	ENRYL TRACKI	ETAHALCIT	SBUIKIMEUND	BWKMKKEN	
ANTP	Dm NK	KKK7buara	CBUIERIER		KKAIGVGLDU	EMARITMUT	CLOAKIMEOND TOTATATATATATATATATATATATATATATATATATAT	RTKWKKOD	
	Dm BarH1	UBKYBLYEL	DHOLOTIER	SFFI	ROKYLGVOFPO	ELAHKI DI GI		RUKWKDOT	
	Ag NK234	KRRPRCLEG	HTOLALIED	RYA	CORAT DO A ORKO		CAUNK I MEUND	RAKNKBUU	
	Ag Hex	RKAARMRFS	OFOTOTIED	RF01	CHYLLPADRK	TTANSTOM	SBOAKLMEONK	RAOCKRSP	
	Co LAG1	GISTKPTKP	NEDNDVIEV NEDNDVIEV	EYKI	KKKRMTPERAA	DITKKLDKT	PEYIMTWFHPE	RNADKPSK	
LAG1	Ny LAG1	GVKDKK-IN	FTPNPFCFK	VY0	TINKPPSDRIL	GLSKOIGWT	REVERWERHR	RMOSKPSL	
LAGI	Dm LAG1	GIRSSRPKK	AANVPILEK	TYA	KSTRLDKKKLV	PLSKOTDMS	EREIERWWRLB	RAODKPST	

Figure S18. An illustrative homeobox alignment, showing the *Capsaspora* and *Monosiga* homeobox domains in contrast to other metazoan families. Taxa used includes Nv (*Nematostella vectensis*), Aq (*Amphimedon queenslandica*), Co (*Capsaspora owczarzaki*), Dm (*Drosophila melanogaster*) and Mb (*Monosiga brevicollis*).



Figure S19. Maximum likelihood tree from the homeobox domain of TALE HDs. The tree is rooted using *Arabidopsis thaliana* sequences as outgroup. Statistical support was obtained by RAxML with 100-bootstrap replicates (bootstrap value, BV) and Bayesian Posterior Probabilities (BPP). Both values are shown on key branches. A black dot indicates BV > 90% and BPP > 0.95. Taxa used adapted from Larroux et al. Nv (*Nematostella vectensis*), Aq (*Amphimedon queenslandica*), Mb (*Monosiga brevicollis*), Co (*Capsaspora owczarzaki*), Dm (*Drosophila melanogaster*), Mm (*Mus musculus*), Sc (*Saccharomyces cerevisiae*), Um (*Ustilago maydis*), At (*Arabidopsis thaliana*), Nc (*Neurospora crassa*) and Y1 (*Yarrowia lypolitica*).



Figure S20. Maximum likelihood tree of non-TALE homeobox domains including all *Capsaspora* non-TALE homologs. The tree is rooted using the midpoint-rooted tree option. Statistical support was obtained by RAxML with 100-bootstrap replicates (BV) and Bayesian Posterior Probabilities (BPP). Both values are shown on key branches. Mb (*Monosiga brevicollis*), Co (*Capsaspora owczarzaki*). Metazoan branches depicted in red and fungal branches in green.



Figure S21. A) An illustrative homeobox alignment, showing the Capsaspora6 with some other archetypical Prd-like genes from different species. In grey the aminoacids that define Prd-like class (Galliot et al. 1999). Black triangle shows the intron position. **B)** Maximum likelihood tree from the homeobox domain of non-TALE HDs from ANTP, Prd-Like, POU and LIM-HD classes. The tree is rooted using the midpoint-rooted tree option. Statistical support was obtained by RAxML with 100-bootstrap replicates (BV) and Bayesian Posterior Probabilities (BPP). Both values are shown on key branches. A black dot indicates BV > 90% and BPP > 0.95. Taxa used Nv (*Nematostella vectensis*), Aq (*Amphimedon queenslandica*), Co (*Capsaspora owczarzaki*) and Dm (*Drosophila melanogaster*).



Figure S22. Maximum likelihood tree from the TLC domain. The tree is rooted using the midpoint-rooted tree option. Statistical support was obtained by RAxML with 100-bootstrap replicates (BV) and Bayesian Posterior Probabilities (BPP). Both values are shown on key branches. A black dot indicates BV > 90% and BPP > 0.95. Taxa used Nv (*Nematostella vectensis*), Aq (*Amphimedon queenslandica*), Mb (*Monosiga brevicollis*), Co (*Capsaspora owczarzaki*), Hs (*Homo sapiens*), Sc (*Saccharomyces cerevisiae*), Um (*Ustilago maydis*) and Dd (*Dictyostelium discoideum*). PFAM domain architecture displayed in the different classes.



Figure S23. Maximum likelihood tree of CP2 domain containing genes. The tree is rooted using fungi. Statistical support was obtained by RAxML with 100-bootstrap replicates (BV) and Bayesian Posterior Probabilities (BPP). Both values are shown on key branches. A black dot indicates BV > 90% and BPP > 0.95.