Additional files 3 for Joo et al.

Article title: Common ancestry of heterodimerizing TALE homeobox transcription factors across Metazoa and Archaeplastida.

Authors: Sunjoo Joo, Ming Hsiu Wang, Gary Lui, Jenny Lee, Andrew Barnas, Eunsoo Kim, Sebastian Sudek, Alexandra Z. Worden, and Jae-Hyeok Lee.

The following Supporting Information is available for this article:

Fig. S1. Alignment of TALE homeodomain sequences of the Archae-algal collection.

Fig. S2. Homology domain alignment of KNOX sequences.

Fig. S3. GLX class is defined by PBL-Chloro domain.

Fig. S4. Extensive homology of Picsa_04995 and Klefl_00021_0250 to Chocr_41034 indicates their classification as PBX-Red.

Fig. S5. Intron-retention pattern suggests parallel evolution of KNOX and non-KNOX group classes from common duplicated TALE ancestors.

Fig. S6. Alignment of Class-B TALE proteins in volvocales.

Fig. S7. Alignment of Mam-A TALE proteins in mamiellophyceae.

Fig. S8. Alignment of Mam-B TALE proteins in mamiellophyceae.

Fig. S9. Full-length proteins are necessary for mamiellophyceae non-KNOX TALE proteins to form heterodimers.

Fig. S10. TALE interaction network defined by this study using yeast-two-hybrid assays.

Fig. S11. Identification of MEIS homologs in choanoflagellates.

Fig. S12. Identification of PBX homologs in choanoflagellates.

Fig. S13. Alignment of the MEIS homologs in Amorphea.

Fig. S14. Alignment of the PBX homologs in Amorphea.

Fig. S15. ELK-domain alignment.

Fig. S1. Alignment of TALE homeodomain sequences of the Archaeplastida-algal collection. The 106 sequences were made into an alignment after excluding 20 near identical sequences to reduce redundancy. Animal/amoeba/haptophyte outgroup sequences are included as they share homology with Archaeplastida TALEs outside the homeodomain. The three bars above the sequence numbers show predicted alpha helices. Discarded insertions are noted in red arrowheads.

Fig. S2. Homology domain alignment of KNOX sequences. MEIS class outgroup sequences are included at the bottom. Class label is on the left. KN-A, KN-B, ELK, HOMEOBOX, KN-C1, and KN-C2 domains are labeled on the top. Class groups are labeled by colored bars on the left next to the gene names. Yellow, light green, and green shades in sequences show more than 60%, 80%, or 100% similarity in each column. Gaps between KN-A and KN-B and between KN-B and ELK have been eliminated.
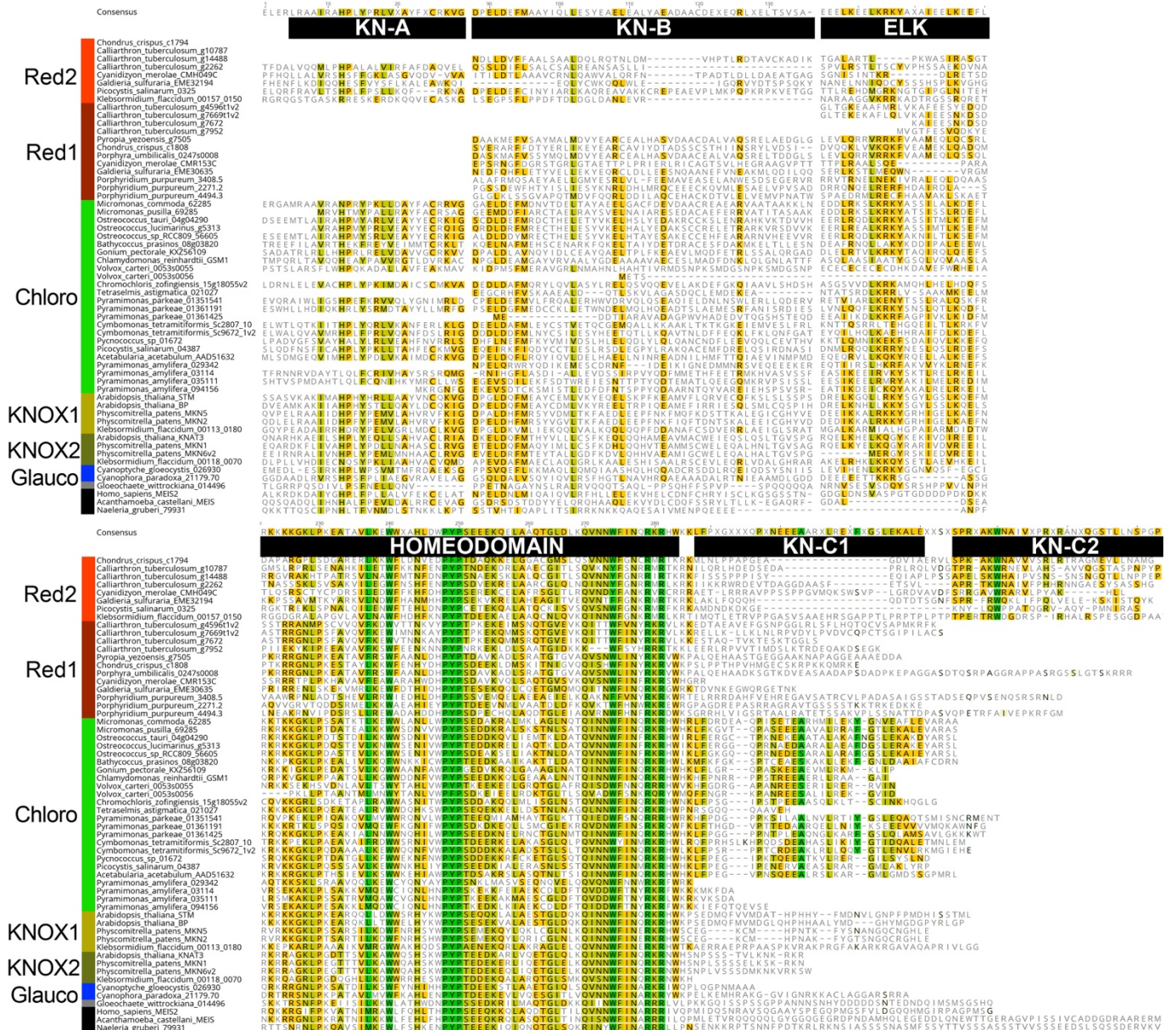
Fig. S3. GLX class is defined by PBL-Chloro domain. (A-C) Three alignments are adjusted with inserting gaps for direct comparison among different PBL-Chloro domains. (A) GLX-Chloro class members. (B) GLX-Basal class members. (C) Three Viridiplantae sequences with strong MEINOX homology domain. PBC-homology is shared among the Chlorophyta non-KNOX sequences.

Fig. S4. Extensive homology of Picsa_04995 and Klefl_00021_0250 to Chocr_41034 indicates their classification as PBX-Red. MEINOX-homology and PBL-Red domains are indicated by red bars below the alignment.

Fig. S5. Intron-retention pattern suggests parallel evolution of KNOX and non-KNOX group classes from common duplicated TALE ancestors. (A) Intron locations collected from 12 TALE classes are shown with arrows. Half arrows indicate cases where not all the class members share the position. White arrows indicate shared positions in at least two different classes, and black arrows indicate class-specific positions. The numbers above the consensus sequence show 60 amino acid positions; the three-amino-acid extension is denoted as 'abc.' Row color depicts two alternative domain configurations: purple for MEIS/KNOX types, and navy for PBX/GLX types. Class names are colored according to their phylogenetic groups: green for Viridiplantae, red for Rhodophyta, blue for Glaucophyta, and black for outgroups. The numbers following the class names show how many genes provided the intron information. Of the shared positions, purple triangles on the top mark those shared between MEIS/KNOX and PBX/GLX classes, blue triangles mark those shared between GLX and BELL classes, and red triangles mark those shared between KNOX classes. A notable exception is the KNOX-Red1 class, for which three Rhodophyta clades show different intron locations (44/45, 48[2/3] or 53[2/3]), indicating that the 44/45 intron position can indeed be displaced to 48[2/3] or elsewhere, albeit infrequently. The unique 46/47 intron in the PBX-Glauco (Cyapa_20927) would presumably have resulted from a similar displacement in intron position. (B) Distribution of conserved introns among the TALE homeobox classes. Identified TALE classes are mapped on the Arachaeplastida phylogeny. The 44/45 intron is marked by blue outline and the 48[2/3] intron is marked by red outline. Underlines of the class names indicate the presence of a PBC-homology domain. The Archaeplastida phylogeny is modified from figure 1 of Jackson et al. (2015).

(a)

(b)

Fig. S6. Alignment of Class-B TALE proteins in volvocales. Short motifs are conserved among all members in this class over the entire length of the sequence.

Multiple sequence alignment of ChrHD1_exonstr, GonpeHDG1_exonstr, and VcaHDG1_ExonStr with Consensus and Identity tracks.

Positions 1,430–1,500

```
Consensus          XSXXLXXXLXXGXGXGXRTPGEAQGQDPGQQXXAAAXAXXVXXXXXRXXPLLXPRXXSPDIFXXWAXAXXFAAXXAXX
Identity
ChrHD1_exonstr     GSSSLRRQLRQGIGGGG---------DMAAAVAHVRRLSGRHTPLIP--VESPDIFSSWAAEAFAAACAAR
GonpeHDG1_exonstr  PSPELLIRLAGGGTDGRRTPGEAQGQDPGQQAVAAAAAAVVQEVARRLSPLLQPRLSPDIFANWAVADEFAAQAASS
VcaHDG1_ExonStr    SSSQN                                        RQQQLL
```

Positions 1,510–1,580

```
Consensus          XXXXXSGXXXAGGXXXXAGAXAXXXXXXXXXXXXXLXXXXXQFVPLQAXGRXAPAPXXXXXLG
Identity
ChrHD1_exonstr     SGSGGGAAGAGADSAGGTALAAGATAGLPFGAIGSTAVLHSSTAAAATAALGAHTGRVEVAAVGRRASVDARGNSLG
GonpeHDG1_exonstr  NTASDGGGGSSSG--GGATSGAGAEAAEWIRTPQLDSLVADWVRGTPPPSLLGLSDQFVPLQPLATPAPAPPQPAP-
VcaHDG1_ExonStr                                                      Q-SPLQAQGRQQPAP
```

Positions 1,590–1,650

```
Consensus          SLFRSSGGGSGSNGGSGSGGGGGSGGGAGGGAASRPLFRPYTPVSVSGGGGGGGLSVMADMLSTPLLASLDEFXXXXXX
Identity
ChrHD1_exonstr     SLFRSSGGGSGSNGGSGSGGGGGSGGGAGGGAASRPLFRPYTPVSVSGGGGGGGLSVMADMLSTPLLASLDEFLAGTPP
GonpeHDG1_exonstr                                                                          PPAAAT
VcaHDG1_ExonStr
```

Positions 1,660–1,730

```
Consensus          XPXXXXXTXXXXGXGXXXXXEXXXXXXXXXXXXXXGXGXXXXXXXXXAXXXXVAXXXXXPXXXXXEQXXXXATAXX
Identity
ChrHD1_exonstr     PPLQLQADTAGTAGGGSGGSGAVEPHGISPATAGLDAAGHRGGVAAPAYAAQGTAVAAAPAPEPADTDP--LVAATATA
GonpeHDG1_exonstr  SPQDVHPGTVPKRAGAGPASVLELAAAGAIGDRNAGPGSSGALTVATADAGARPVANGLSSLPTSAAAEQYATEATAAT
VcaHDG1_ExonStr
```

Positions 1,740–1,810

```
Consensus          XRGGHGAAXXXXXPXLPPXXXXAXXXAXXXAAXXXXATLXALXXLXXXXXXDPXAAXXLEELXXITXXPXAVXXXXXXXX
Identity
ChrHD1_exonstr     A----AAAPPQLPPWLSATDGATASAGVDSATLAALTALFEHSMDPAAAATALEELSS--LVPSAVAAAAAAVAA
GonpeHDG1_exonstr  GRGGHGAAVTALPLLPPAGGAAAAAADYAASSALATLEALQTLMRTNIDPGAASALLEELQYITADPAAVGGGGGGSGG
VcaHDG1_ExonStr                   P                          LHQQFPSQGHQ
```

Positions 1,820–1,890

```
Consensus          XXXAVVPAAAAGPGPAAGQPETLAXXXXXXVXPXAXXEXXXCXXXXXAXXXPQXXXXXXHHXRXQXXPPHAXXXPX
Identity
ChrHD1_exonstr     AASAVVPAAAAGPGPAAGQPETLAAPMPPAVQAPPAAMEQHLQQRARSQSGAGSTPHQHHYQHHRQAQPPPHAPAPP
GonpeHDG1_exonstr  GSG---------------GGSDELVRNPDDADAEDVEANVAAAGAAARPRAESDVGSTMRQQTPPQALPP
VcaHDG1_ExonStr                         HQQQQHQHQHQQQLPPLPSPQQLPQTPHHSRLQHQQHQQQHQ-
```

Positions 1,900–1,970

```
Consensus          IXXXXXXXLXLXPXAXXXXXXXQXXXQQQXHXXXHXXXXXPXPXXXXXQXQQPQHQQXXXXQQQQQQQQQQHXQPXXXXH
Identity
ChrHD1_exonstr     IHVGTATHPIPPLPTHQPFHHQHHHQHQHDSAHDRPLPPQQQTLQPQQPQHQQQHPPPPPQQQQQQQQPHPHPAYHPH
GonpeHDG1_exonstr  ------LPLHLAPPASEPAGSRQQRQQQQRPHPY------SQLPSHPQLLLIQQQQQQQQQQQQQQQEQEQEH
VcaHDG1_ExonStr    --QHLYPHLNMLPQA------QQQQQHPHLLSQPLPYVPQQHQQRPQ-QQFQTLQQQQQQQQQQQHSQP-
```

Positions 1,980–2,050

```
Consensus          PXQXPHXSXXXXXXZXXXXDELPDGXXXXXXXXXXXQPHTHHHYPQPHPQHGPXXQHQXXXXXXXXXZXXXQXXXVSXXXX
Identity
ChrHD1_exonstr     PHQHPHQSAAAAAEGQPY------HLPVPAPAGIEQQPHTHHHYPQPHPQHGQPHQHQHYHQHQHHQQYPVAAASQ
GonpeHDG1_exonstr  PRQEP----------------------------------------PHLQHQDAVAMKRERYEASSVS
VcaHDG1_ExonStr    -QDLHLSGSFVRQSVRRDELPDGGSIFSSSFPMGP------------------YQRTASEGVV
```

Positions 2,060–2,130

```
Consensus          XXXPPPPAHHXXHPXGPPPPXXXYGXXXRPPXXGGXXRXXXGXXXXXXXXXXXXXAXXXAGRGDXGAAXXFQQLLADXAGG
Identity
ChrHD1_exonstr     GAAPPPPAHHSPHPQQQPPPYPYPYAHQPPPSGGWARASRGASMTHDASMTHDASMTHA--GASGLVPSFQQLLADDAGG
GonpeHDG1_exonstr            -HPAGPGPHSVGHLQRLLTEGGSFPGDPAAGGQPYYQGAAHPQAHQAAAA---VGG
VcaHDG1_ExonStr      ARMP----AEGPGPPPLYYGNPLRNPPAPASELRTHHG-A--------WDYGAAGA----VGG
```

Positions 2,140–2,210

```
Consensus          XXXXXAGGXGXAXGXXXXXXXQHPXGXXGXXXXQXPXAAAAAAAXXXXXXXXAXXXPXXXXXADXXSAAEAQAXXA
Identity
ChrHD1_exonstr     DDGGTGGGGGPQYGYQHHQHHQHPVTAMATDAAALWAAAAMATAAAAADAAGAGGAAGGGGADVSSAAEAQAKAA
GonpeHDG1_exonstr  AAAAAAAAQGWAFGLSDSSSPLHRAGSVGSMVQVPQSRAAAAAAQRHTAYLAAPPTAPRADAG
VcaHDG1_ExonStr    QTRTVAGGTAQMA--------QQPLGGLGRGMHQLP-PAAAMAVAERQPSGAQQAVLPFSLDAADMA------AMP-
```

Positions 2,220–2,300

```
Consensus          YEWQQYQLQXXXXYXXXXXXXZQXQQQQXXYXXXGQQQQQQXXQQQQOQYGXXXXXXXXXAXXXXXXXXXXXPXQPYB
Identity
ChrHD1_exonstr     YEWQQYQLQLEQYHQQQQYEQYQQQQRQYELYQQQQQQQQQQQQQOQYGQYGNAAATAAATEHPPREGGDHQPPPPP-
GonpeHDG1_exonstr  ------VDLYGTAYIGAQQQQQQQTR----QGQQQQQQHQQQQQOQYGNW----------PSQPVN
VcaHDG1_ExonStr    ----DPNEMYDHSYGSPAQSSGIDGEQYGGRWSTGGVATVSGGDAGGALTYPGRQQPWRD
```

Positions 2,300–2,370

```
Consensus          GYXPXASAXIRQAHQAPPXXXPXXDXYGXXXXDPXSQPPPLESRXXXXAAXAPPPQXRVGXXXXXFAXXASXSXXG
Identity
ChrHD1_exonstr     GYEPYSSA---------PYGDPNLYSIANAQDPSSQPPFLESRPTAAAAASAPPPHAAAGFAAASNSLAAATASGLG
GonpeHDG1_exonstr  EYDPLAAAIRQAHQAPPQSGPAVDMYGRQYDA-------MSVGAADLAPPPQQGRVSYPGMHDTFSTVSTFGS-
VcaHDG1_ExonStr    GYGDHAAS----------PMASRGPDVYGSHFPG---------EPADSAQQ-RVGSENIMQPFAFGGAASTSSGG
```

Positions 2,380–2,440

```
Consensus          IAGNLSGGGATAGAGPGGGGGXSNWTSLQGRVAAGALAYGT-SVAAPSRMINGRGLGFKASXTAAAVAAAPGHGGVG
Identity                                       [Upstream homology]
ChrHD1_exonstr     IAGNLSGGGATAGAGPGGGRPGAMNWTSLQGRVAAGALAYGT-SVAAPSRMINGRGLGFKASXTAAAVAAAPGHGGVG
GonpeHDG1_exonstr                          -GAGGGSSNWTSLQGRVAAGALAYGT-SVAAPSRMINGRGLGFKAS
VcaHDG1_ExonStr                           -GGGGGGGSNWTTLQGRVAAGALAYGT-SVAAPSRMINGRGLGIKSSS
```

Positions 2,450–2,520

```
Consensus          GGGGGGGXGXGXSGGGSGGAGGGAAAGAGVSGGGAGAGAGAGAGVTSSAVTGSGAGGGAGGGGGGGXSGPGPGSKKGRENLPK
Identity                                                                      [Homeobox]
ChrHD1_exonstr     GGGGGGGSGAGGGGGGGGAAAGAGVSGGGAGAGAGAGAGVTSSAVTGSGAGGGAGGGGGGGGSGPGSKKGRENLPK
GonpeHDG1_exonstr  GGGGGG----------------------------------------------------GGGGGANGPGSKKGRENLPK
VcaHDG1_ExonStr    GGGGGGGDGG-----------------------------------------------CGGGGGGGSGPGSKKGRENLPR
```

Positions 2,530–2,600

```
Consensus          XAVXSLKLWVYNHIVHPYPSEDEKEVLCAXTGLDLIQLNNWFINARVRIWKPLITXVFNSXNQPRMAXEACARGDTELL
Identity                     [Homeobox]
ChrHD1_exonstr     NAVTALKQWVYAHIVHPYPSEDEKEVLCAHTGLDLIQLNNWFINARVRIWKPLITQVFNS-NHPRMAAEACARGDGELL
GonpeHDG1_exonstr  TAVNSLKLWVYNHIVHPYPSEDEKEVLCNSTGLDLIQLNNWFINARVRIWKPLITSVFNS-NQPRMAAEACARGDTELL
VcaHDG1_ExonStr    AAVQSLKLWVFNHIVHPYPSEDEKEVLCANTGLDLIQLNNWFINARVRIWKPLITEVFMS-NQPRMAHEACARGDTELL
```

Positions 2,610–2,680

```
Consensus          XKMNAARNSPTAQLALVAADAXALXCLEQAAQXAAEELXAAAXAAAGGGXXXGXXXXXAXXXXXXGXXXXXXXXGXEGX
Identity
ChrHD1_exonstr     TKMQAARNSPTAQLALVAADASALQNLEQAAQKASEDIAAAAQAAAAGGAKRGSGGTGAKRSKASGALAGAGPAGGEGS
GonpeHDG1_exonstr  GKMNAARNSPTAQLALVAADATALTCLEQAAQAAAEELMAAAKAITAGGG
VcaHDG1_ExonStr    AKMNAARNSPTAQLALVAADAISLSCLEQAAQRKASEVLESARAAAGGSGGKGGISSRAYSMPWRGRRRRGRGRGEEGI
```

Positions 2,690–2,760

```
Consensus          XXGXXXXGXGXXXRRRGCGGGGXXXXXXXXGLGEGDQYQQQYQQHQQQHQHQQHQQQQGADFGAAEFLGDGGAALAEYMA
Identity
ChrHD1_exonstr     AAEEAAGAGPGARSRACGGGGGRSRGTGGGLGEQDQYQQQYQQHQQQHQHQQHQQQQGADFGAAEFLGDGGAALAEYMA
GonpeHDG1_exonstr  -GGSKCGGGSSRRRGCGGGG-----------
VcaHDG1_ExonStr    WRGRRRGRRRRPWRFGGDVVGRGLLSPC
```

Positions 2,770–2,833

```
Consensus          TYGHLAGDFAAAMASFAAAGAGVMDPHAAAAAAAVAAAAGAGGAXXXGGGGGGCXGXSXTDDXXXXY
Identity
ChrHD1_exonstr     TYGHLAGDFAAAMASFAAAGAGVMDPHAAAAAAAVAAAAGAGGAPGEGGGGGGAYGDSGTDDPLGYY
GonpeHDG1_exonstr  ------GGAGAGGGGGGGGSGESV---SMEGY
VcaHDG1_ExonStr
```
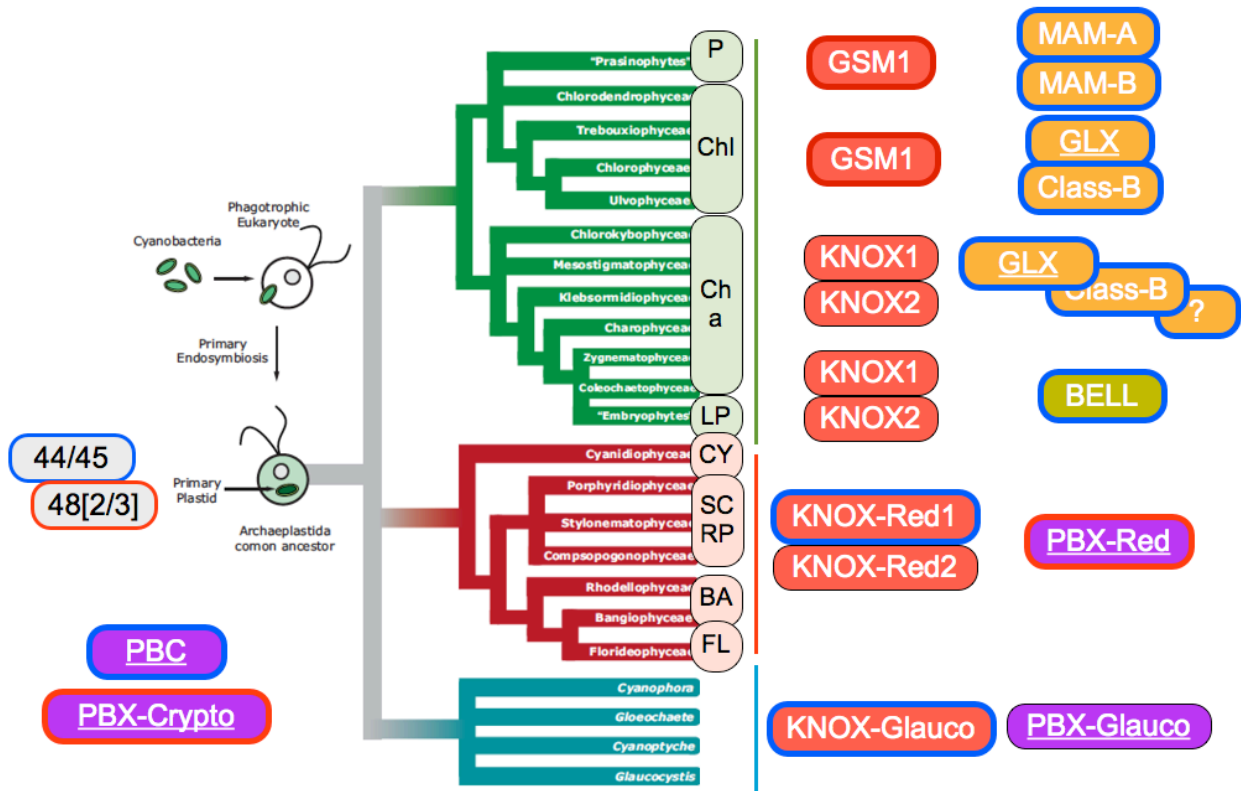
Fig. S7. Alignment of Mam-A TALE proteins in mamiellophyceae. Short motifs (Box1-4) are conserved among all members in this class over the entire length of the sequence. Red reverse triangle at 548-549 shows the truncation position of Micco_Mam-A-tr used in Yeast-two-hybrid analysis.
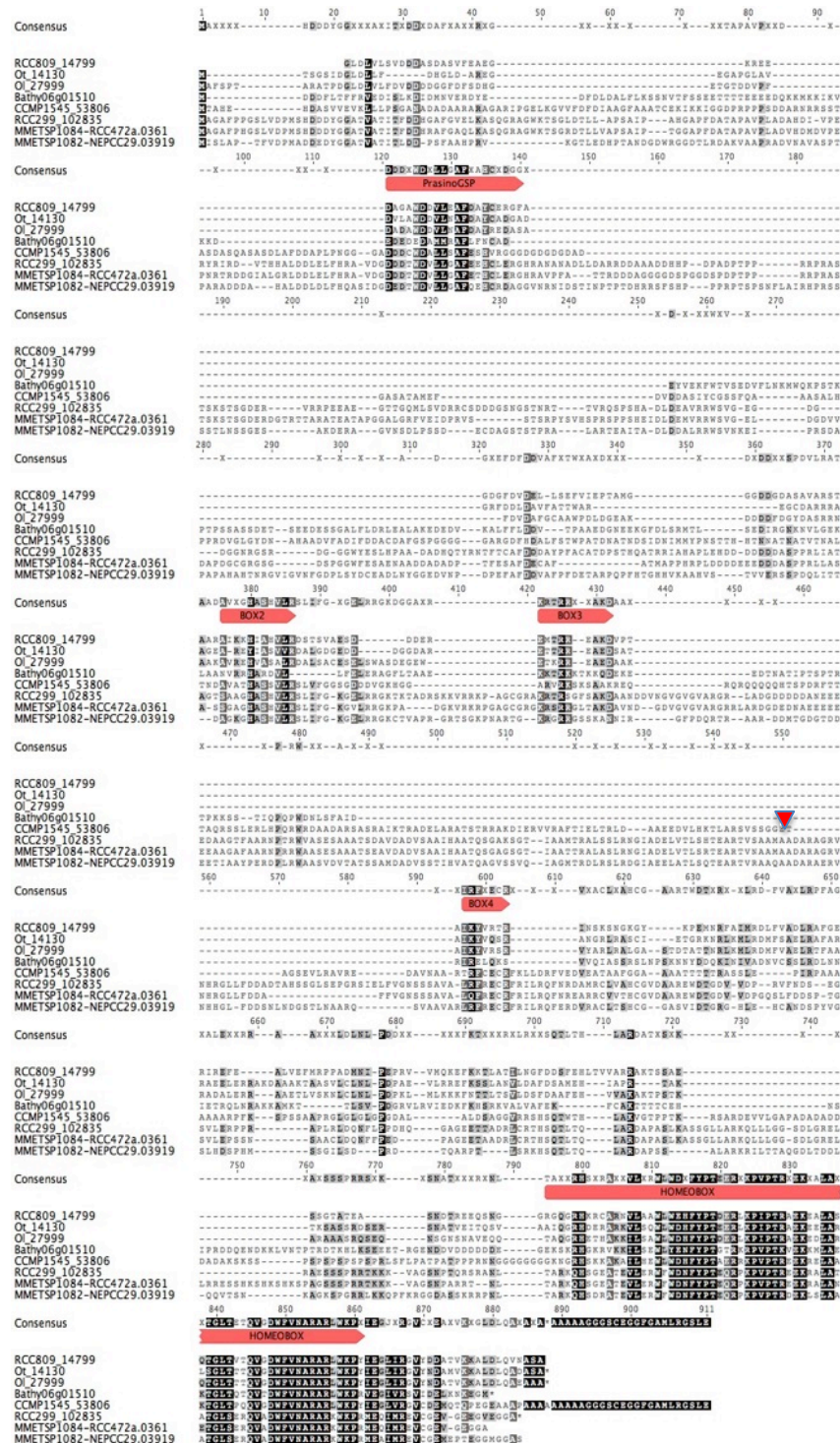
Fig. S8. Alignment of Mam-B TALE proteins in mamiellophyceae. A conserved motif is found between 180-197 amino acids in the alignment. Red reverse triangle at 100-101 shows the truncation position of Ostta_Mam-B-tr used in Yeast-two-hybrid analysis. Homology is restricted to a single homology-A domain ouside the homeodomain.
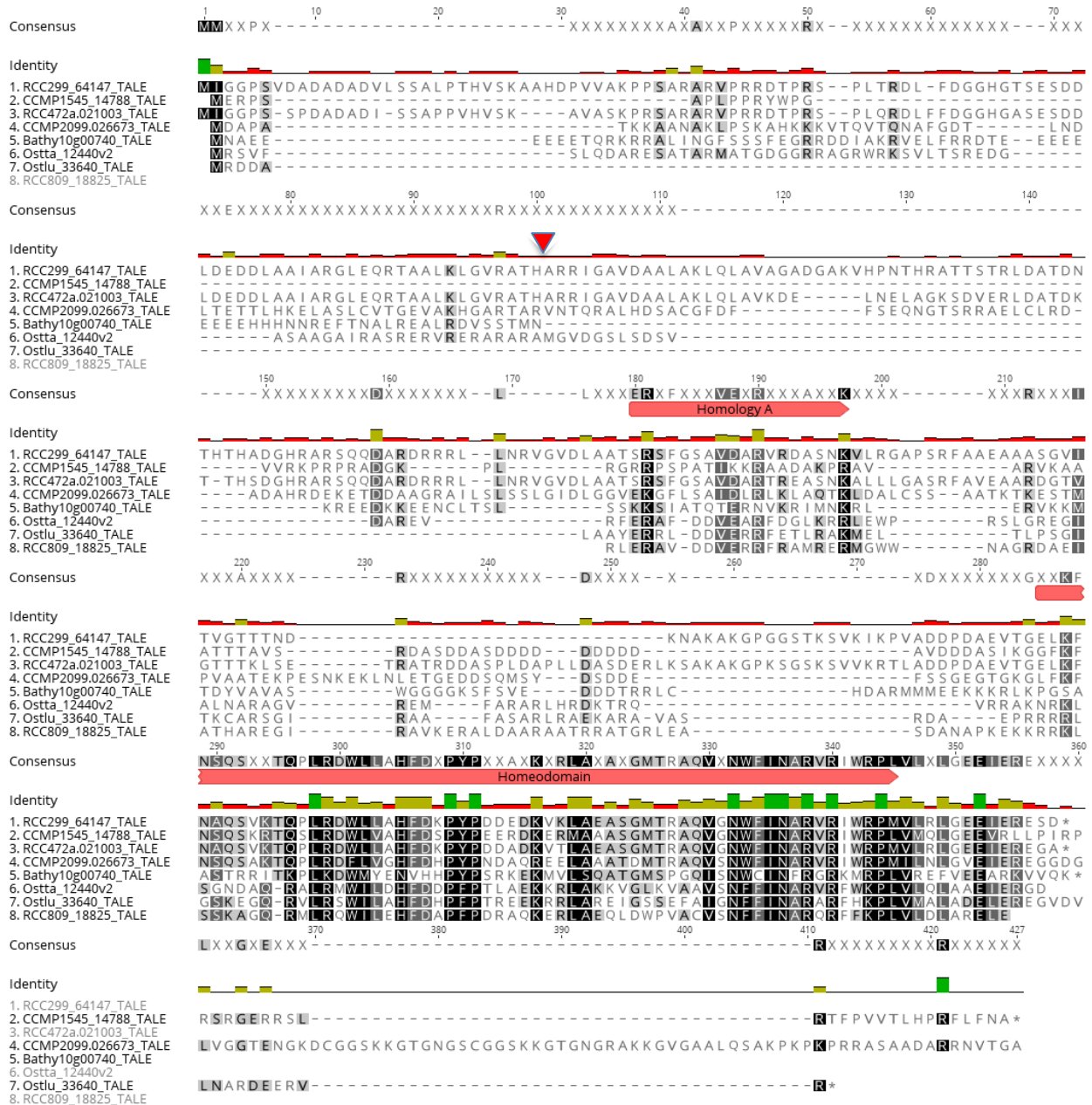
Fig. S9. Full-length proteins are necessary for mamiellophyceae non-KNOX TALE proteins to form heterodimers. Left and Right: Yeast-two-hybrid assays on Ade-/His-/Leu-/Trp- medium. The construct information for the prey conjugated with the GAL4 DNA-binding domain and for the bait conjugated with the GAL4 transcriptional activation domain is given in the table below.
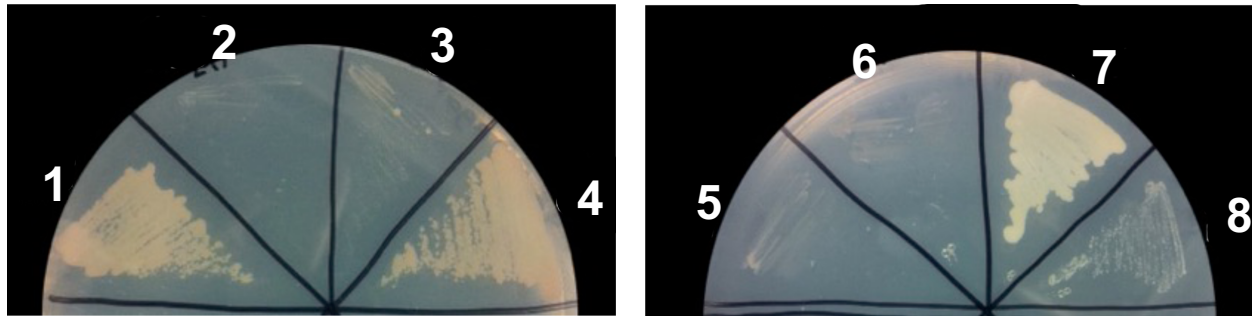
Sector	DBD	AD	Sector	DBD	AD
1	Ostta_KNOX	Ostta_Mam-A	5	Micco_Mam-B	T-Ag
2	Ostta_KNOX	Ostta_Mam-B-tr	6	Micco_Mam-A-tr	Micco_KNOX
3	Ostta_Mam-A	Ostta_Mam-B-tr	7	Micco_Mam-B	Micco_KNOX
4	Ostta_Mam-A	Ostta_KNOX	8	Micco_Mam-B	Micco_Mam-A-tr

Fig. S10. TALE interaction network defined by this study using yeast-two-hybrid assays.
(A) Summary diagram for the TALE interaction network. (B) Yeast-two-hybrid assays for the cross-species interaction of TALE proteins. Only one of the possible reciprocal combinations of GAL4 domain conjugations is provided for simplicity. Large X indicates no yeast in the sector. -LTHA: Leu-/Trp-/His-/Ade- medium; -LT: Leu-/Trp- medium.

Fig. S11. Identification of MEIS homologs in choanoflagellates.

Fig. S12. Identification of PBX homologs in choanoflagellates.

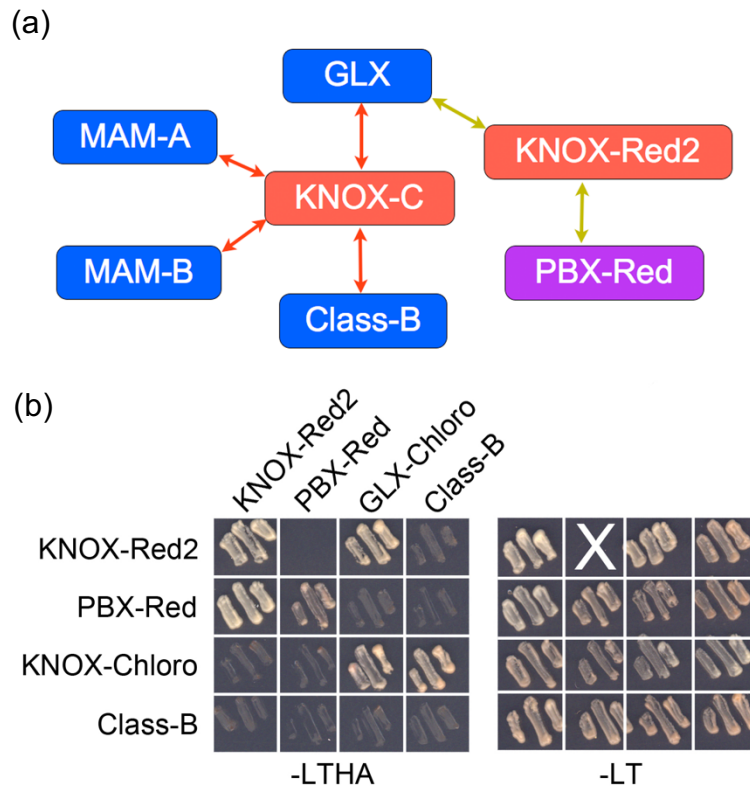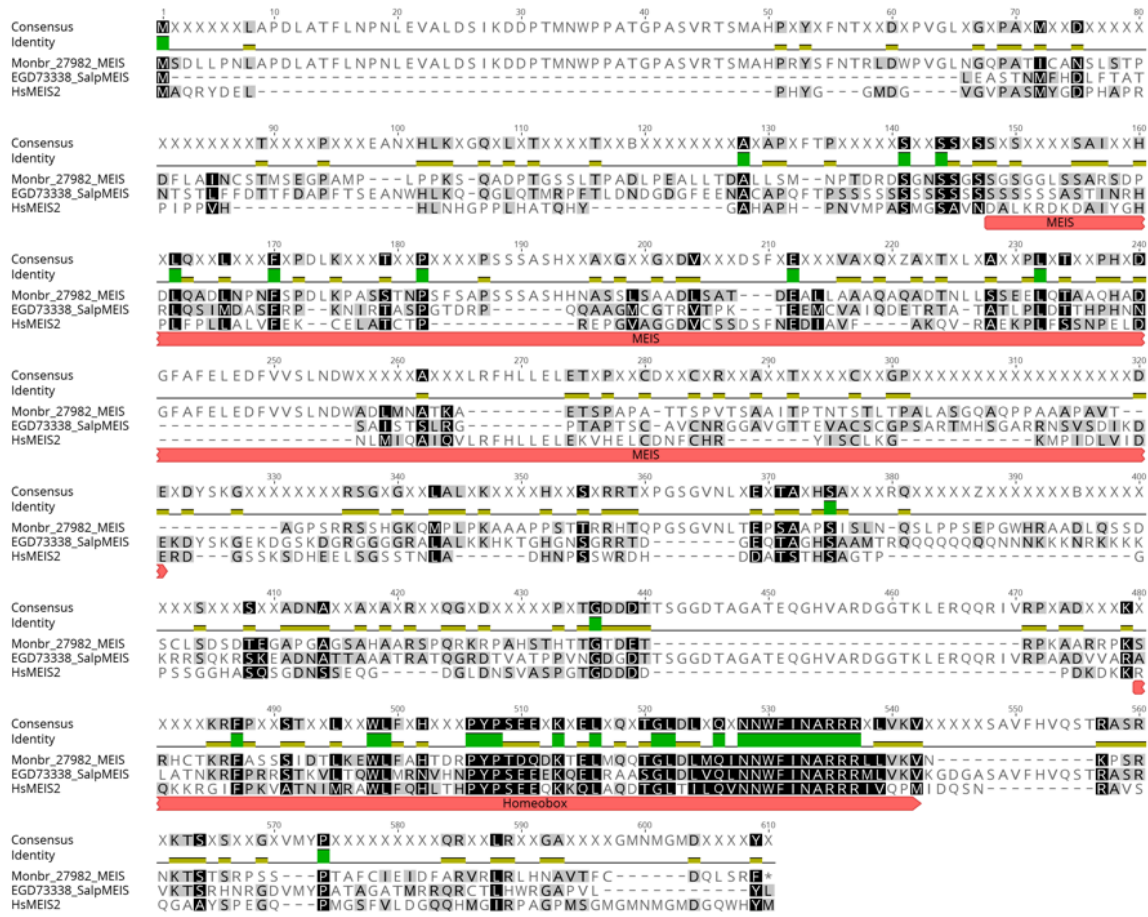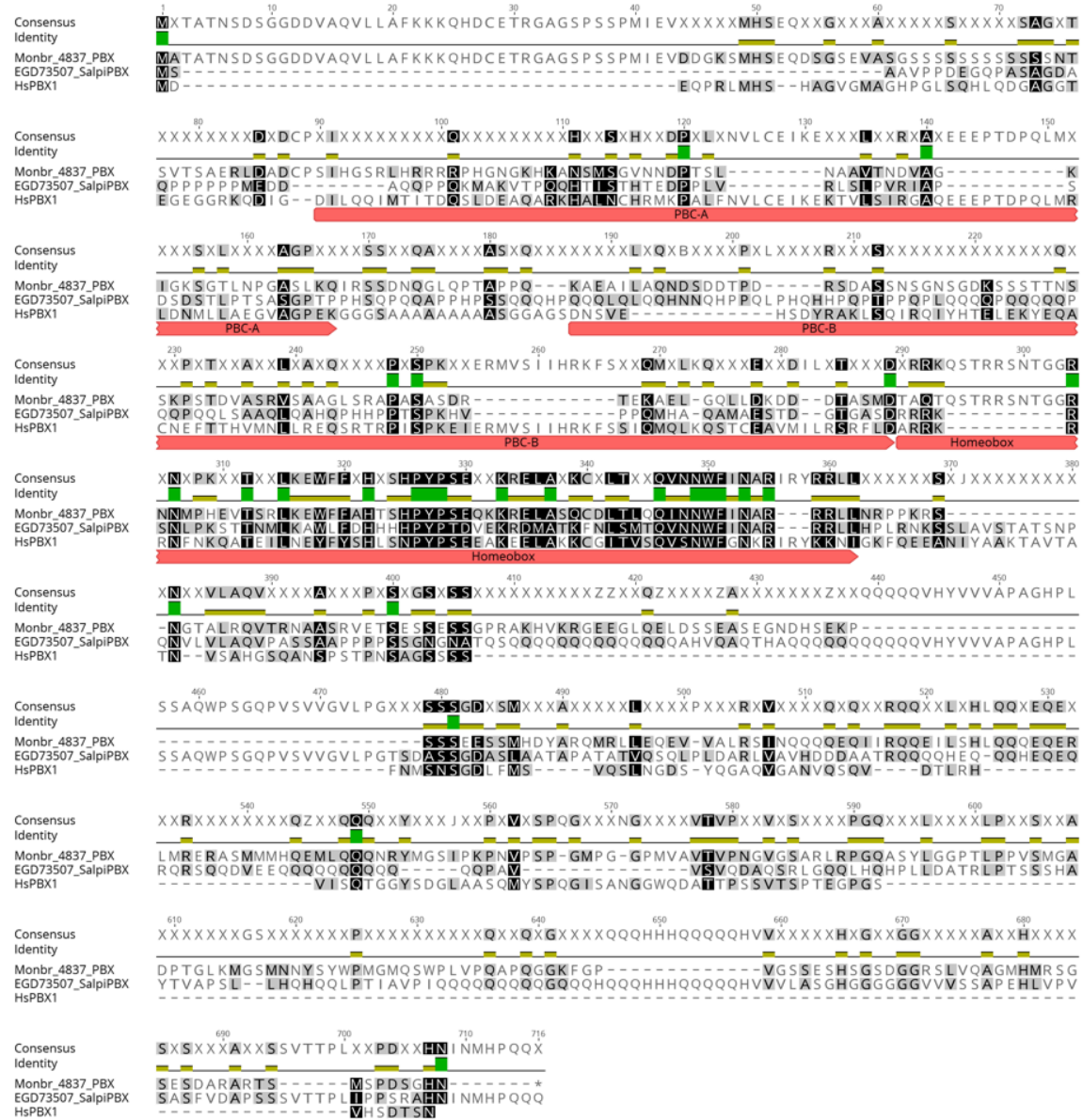**Fig. S13**. Alignment of the MEIS homologs in Amorphea.

```
                              1         10        20        30          40          50        60        70
1. Capowc_EFW44798_TALE  DTESDGTPSKRARLDNGWSATSAG------NNNNNNSSNNNNNNN--NNPLPPPPPGMSAYPPPPHAMQ
2. Thetra_KNC56447_TALE  NGSAGGPGNEQAPPQQVP-----------LSPTHTSSPTSASSNAPLSALNMVSGEALAS---------
3. Monbr_27982_MEIS      SSGSGGLSSARSDPDLQADLNPNFSPDLKPASSTNPSFSAPSSSASH--HNASSLSAADLSA------TDE
4. Salros_EGD73338_MEIS  DNDGDGFEENACAPQFTP------------SSSSSSSSSSSSSSSS--SSASTINRHRLQS------IMD

                              80        90        100       110       120       130       140
1. Capowc_EFW44798_TALE  APPQPMQMAYASN-YSGHPYYMPSPYPDQSMPYNYSYGMPVSQPVPPVHHTHAEYKPTHASANSVAAAATA
2. Thetra_KNC56447_TALE  ------------------------------------------------------TTSETKVTTTA
3. Monbr_27982_MEIS      ALLAAAQAQADTNLLSSEELQTAAQHADGFAFELEDFVVSLNDWADLMNATKAETSPAPATTSPVTSAAIT
4. Salros_EGD73338_MEIS  ASFRPKNIRTASP-GTDRPQQAAGMCGTRVTPKTEEMCVAIQD-----ETRTATATLPLDTTHPHNNSAIS

                              150       160       170       180       190       200       210
1. Capowc_EFW44798_TALE  APTVVPAAPTSDG--AHKTKPLPTDPPHTNKPGSSNES-----ERVIEKPASSE-----------STQPRD
2. Thetra_KNC56447_TALE  -PGAGSADANPGGRPIGSSMLVTTETETTNSTNNHSQLHGDAEEHVGDMAPVNESYSVAG-----SENGGG
3. Monbr_27982_MEIS      -PTNTSTLTPALAS--GQAQPPAAAPAVTAGPSRRSSH----------------------------
4. Salros_EGD73338_MEIS  TSLRGPTAPTSCAVCNRGGAVGTTEVACSCGPSARTMHSGARRNSVSDIKDEKD-YSKGEKDGSKDGRGGG

                              220       230       240       250       260       270       280
1. Capowc_EFW44798_TALE  NVQHPPPQSGGQQPSQAHHQNG-----SEAPVTPNTLPHASTFLDSPRGAGKSPREDGPNPDP----SNG
2. Thetra_KNC56447_TALE  DRARAAKRSKLSSSAAVTRQT-------AEVDSGSDSDYV--EGKKQDGGNGKRKRSSASTKRG-------
3. Monbr_27982_MEIS      GKQMPLPKAAAPPSTTRRHTQPGSGVNLTEPSAAPSISLN-QSLPPSEPGWHRAADLQSSDSCLSDSDTEG
4. Salros_EGD73338_MEIS  GRALALKKHKTGHGNSGRRTD------GEQTAGHSAAMTRQQQQQQQNNNKKKNRKKKKKRRSQKRSKE

                              290       300       310       320       330       340       350
1. Capowc_EFW44798_TALE  AKMEQLADPTQPVQLGDPMYSHFSVKQAKAEPKSSSEFHFTWSKDGQHTVLTRGRSKQRPN-LSREVVEVL
2. Thetra_KNC56447_TALE  ---------------------------GGAAKSR----------------RWGGNRASSRLFPPQAKAVF
3. Monbr_27982_MEIS      APGAGSAHAARSPQRKRPAHSTHTTGTDETRPKAA---------------RRPKSRHCTKRFASSSIDTL
4. Salros_EGD73338_MEIS  ADNATTAAATRATQGRDTVATPPVNGDGDTTSGGDTAGATEQGHVARDGGTKLERQQRTNKRFPRRSTKVL

                              360       370       380       390       400       410       420
1. Capowc_EFW44798_TALE  KNWLFAHSSRPYPSDVEKTAMMAETGIISLLQLNNWFINARRRLLQKPD
2. Thetra_KNC56447_TALE  ETWLKDHLDEPYPEEEDVLRMAHDADVKPDQVLQWFINARRRHPLLVNRNFKRCRRRYSTISDRASTQRRK
3. Monbr_27982_MEIS      KEWLFAHTDRPYPTDQDKTELMQQTGLDLMQINNWFINARRRLLVKVN------------KPSRNKTSTS
4. Salros_EGD73338_MEIS  TQWLMRNVHNPYPSEEEKQELRAASGLDLVQLNNWFINARRRMLVKVKGDGASAVFHVQSTRASRVKTSRH

                              430       440       450       458
1. Capowc_EFW44798_TALE
2. Thetra_KNC56447_TALE  SQRRTRSPAEPATAPMMTGTGGQPVAASFLFG
3. Monbr_27982_MEIS      RPSS---PTAFCIEIDFARVRLRLHNAVTFCD
4. Salros_EGD73338_MEIS  NRGDVMYPATAGATMRRQRCTLHWRGAPVLY
```

Fig. S14. Alignment of the PBX homologs in Amorphea.

```
                              1        10         20         30         40         50         60
1. Capowc_EFW47516_TALE   TPTNRVGFASHDVTAAQLDVCLAQERAACAASETPSLAEIRGLFAFLDMQLLKVAHR--L
2. Spharc_KNC86410_TALE     SEQARQYDELHRQIH--QLLELLADGPERNARD-SEAK--------QMLDRIHNLIN
3. Salros_EGD73507_PBX    PQPLQQQQPQQQQQPQQPQ--QLSAAQLQAHQPHHPP-TSPK--------HVPPQMHAQ-A
4. Monbr_4837_PBX         SSSTTNSSKPSTDVASRVSAAGLSRAPASASDRTEK--AELG--------QLLDKDDD--T
5. HsPBX1                 HTELEKYEQACNEFTTHVM--NLLREQSRT-RPISP--KEIE--------RMVSIIHRK-F

                                   70         80         90        100        110        120
1. Capowc_EFW47516_TALE   SSLEYQLKRKVYTKVIFLKSTFLDSRKRRINLSREAQQVLNDWFLAHIEHPYPSESEKEQL
2. Spharc_KNC86410_TALE   DSIDEALQKH--------GTGASTGAKRRRQLPPAAHSILKKWLYENQDCPYPSDEQKQQL
3. Salros_EGD73507_PBX    MAESTD-----------GTGASDRRRKRSNLPKSTTNMLKAWLFDHHHHPYPTDVEKRDM
4. Monbr_4837_PBX         ASMDTAQT----------QSTRRSNTGGRNNMPHEVTSRLKEWFFAHTSHPYPSEQKKREL
5. HsPBX1                 SSIQMLKQSTCEAVMILRSRFLDARRKRRNFNKQATEILNEYFYSHLSNPYPSEEAKEEL

                               130        140        150  152
1. Capowc_EFW47516_TALE   ADQTNLTMRQISTWFANKRNRQAQD
2. Spharc_KNC86410_TALE   AEECQLEIQQIRNWFANTRNRKRHLF
3. Salros_EGD73507_PBX    ATKFNLSMTQVNNWFINARRRLLHPLRNKS
4. Monbr_4837_PBX         ASQCDLTLQQINNWFINARRRLLNRPPKRS
5. HsPBX1                 AKKCGITVSQVSNWFGNKRIRYKKNI
```

Fig. S15. ELK-domain alignment.

Consensus
Identity

1. CCMP720_035111_KNOX-R1_site_1
2. Ostlu_5313_KNOX_site_1
3. Ostta_04g04290_KNOX_site_1
4. RCC809_56605_KNOX_site_1
5. AT4G08150_BP_site_1
6. CCMP720_094156_KNOX-R1_site_1
7. CCMP720_03114_KNOX-R1_site_1
8. AT1G62360_STM_site_1
9. Picsa_04387_KNOX_site_1
10. RCC299_62285_KNOX_site_1
11. Cyagl_026930_KNOX_site_1
12. CCMP726_01361425_KNOX_site_1
13. CCMP720_029342_KNOX-R1_site_1
14. CCMP1545_69285_KNOX_site_1
15. CCMP726_01361191_KNOX_site_1
16. Bathy08g03820_KNOX_site_1
17. Cyapa_20927.63_PBC
18. Cyapa_20927.64_PBC
19. Glowi_03052_PBC