

Supplementary Figure 7. Species-specific motifs in the HD family

```
getMatAlignOut <- function(pcmppath, outpath="output",
                           groupDistance=NA, trim=0.2){
  pcms <- readPCM(pcmppath)
  pfms<-lapply(pcms,pcm2pfm)
  matalign_path <- "./app/matalign-v4a"
  neighbor_path <- "./app/neighbor.app/Contents/MacOS/neighbor"
  system(paste("perl MatAlign2tree.pl --in . --pcmppath", pcmppath,
               "--out", outpath,
               "--matalign", matalign_path,
               "--neighbor", neighbor_path,
               "--tree", "UPGMA"))
  newickstrUPGMA <-
    readLines(con=file.path(outpath, "NJ.matalign.distMX.nwk"))
  phylog <- newick2phylog(newickstrUPGMA, FALSE)
  leaves <- names(phylog$leaves)
  motifs <- pfms[leaves]
  if(!is.na(groupDistance)){
    motifSig <-
      motifSignature(motifs, phylog,
                    groupDistance=groupDistance,
                    min.freq=1, trim=trim)
    sig <- signatures(motifSig)
    gpCol <- sigColor(motifSig)
  }else{
    motifSig <- NA
    sig <- NA
    gpCol <- NA
  }

  return(list(phylog=phylog, sig=sig, gpCol=gpCol,
             motifs=DNAmotifAlignment(motifs),
             leaves=leaves,
             unaligned.pfms=motifs))
}

pcmppath <- dir("pcmsUni", include.dirs=TRUE)[7]
matAlignOut <- lapply(file.path("pcmsUni", pcmppath),
                      getMatAlignOut, groupDistance=4)
description <- c("Species-specific motifs")
colorSet <- c("Dm"="#00FC00",
              "Mm"="brown", "Ms"="#F69156",
              "Hs"="#D900D9")
```

Species-specific motifs

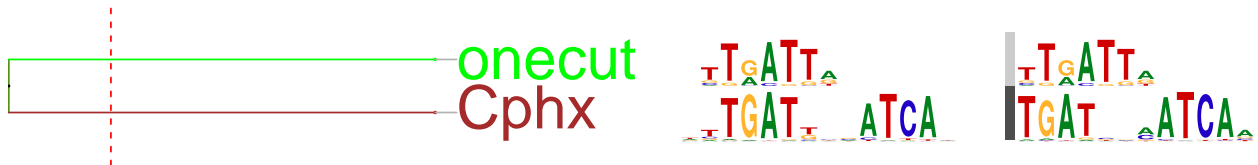
```
leaveNames <- gsub('^(Dm|Mm|Ms|Hs)_', '', .ele$leaves)

species <- colorSet[gsub("^(Dm|Mm|Ms|Hs).*$", "\\1", .ele$leaves)]
```

```

motifFiles(phylog=.ele$phylog, .ele$motifs,
            .ele$sig,
            col.tree=species, col.leaves=species,
            col.pfms2=.ele$gpCol,
            col.pfms2.width=.01, labels.leaves=leaveNames,
            plotIndex=c(FALSE,TRUE), IndexCex=1.5,
            groupDistance=4, clabel.leaves=3)

```



Supplementary Figure 7a: Species-specific motifs

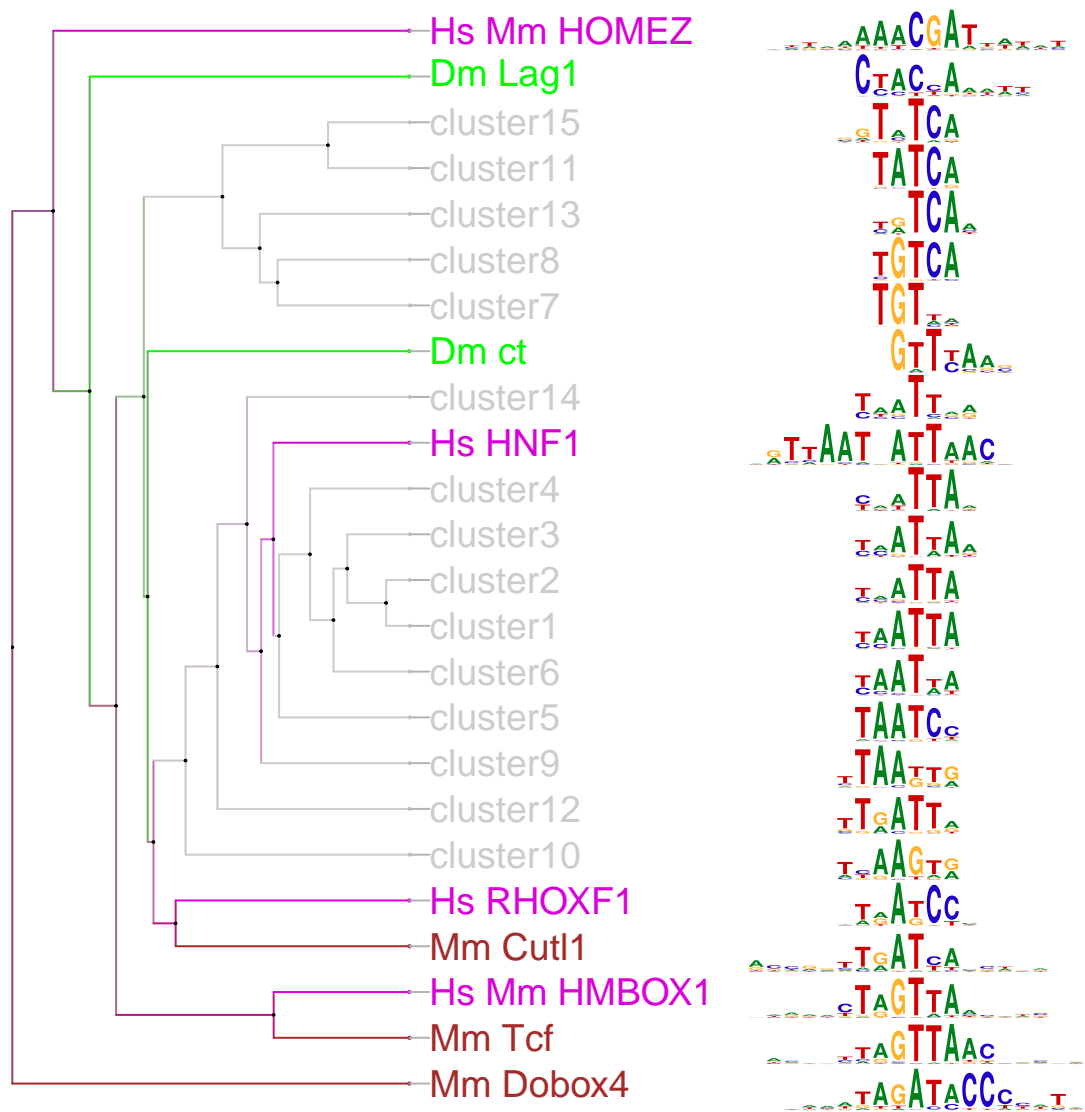
```

colorSet <- c("Dm"="#00FC00", "b1h"="#00FC00", "sw"="#008080", "bml"="darkgreen",
              "Mm"="brown", "MmDREAM"="blue", "Ms"="#F69156",
              "Hs"="#D900D9")

## function to read example data
readDataDoAna <- function(pcmPath, outputPath="output", groupDistance=2.5, trim=0.2){
  pcms <- readPCM(pcmPath)
  pfms<-lapply(pcms,pcm2pfm)
  matalign_path <- "./app/matalign-v4a"
  neighbor_path <- "./app/neighbor.app/Contents/MacOS/neighbor"
  system(paste("perl MatAlign2tree.pl --in . --pcmPath", pcmPath, "--out", outputPath,
               "--matalign", matalign_path, "--neighbor", neighbor_path, "--tree","UPGMA"))
  newickstrUPGMA <- readLines(con=file.path(outputPath, "NJ.matalign.distMX.nwk"))
  phylog <- newick2phylog(newickstrUPGMA, FALSE)
  phylog <- reorderUPGMAtree(phylog, pfms)
  leaves <- names(phylog$leaves)
  motifs <- pfms[leaves]
  if(!is.na(groupDistance)){
    motifSig <- motifSignature(motifs, phylog, groupDistance=groupDistance,
                              min.freq=1, trim=trim)
    sig <- signatures(motifSig)
    gpCol <- sigColor(motifSig)
  }else{
    motifSig <- NA
    sig <- NA
    gpCol <- NA
  }

  return(list(phylog=phylog, sig=sig, gpCol=gpCol,
              motifs=DNAmotifAlignment(motifs, minimalConsensus=3),
              leaves=leaves, unaligned.pfms=motifs))
}

```



Supplementary Figure 7b: Species-specific motifs in the HD family plotted together with common motif clusters