

Supplementary Figure 8

Contents

```
##load the library
library(motifStack)

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))
}

gpDis <- 2
Fly_DNA <- readDataDoAna("pcmsDatasetFly", groupDistance=gpDis)
unUnique <- Fly_DNA$sig

pcms <- readPCM(file.path("pcmsUni", "14.UniqueCleanCluster"))
dimmer <- sapply(pcms, motifStack::isHomoDimer)
pcms.mono <- lapply(pcms[dimmer], function(.ele){
  pos <- motifStack::getHomoDimerCenter(.ele)
  .ele$mat <- .ele$mat[, 1:as.numeric(pos["pos"])]
  .ele
})
```

```

pcms.mono <- c(pcms[!dimmer], pcms.mono)

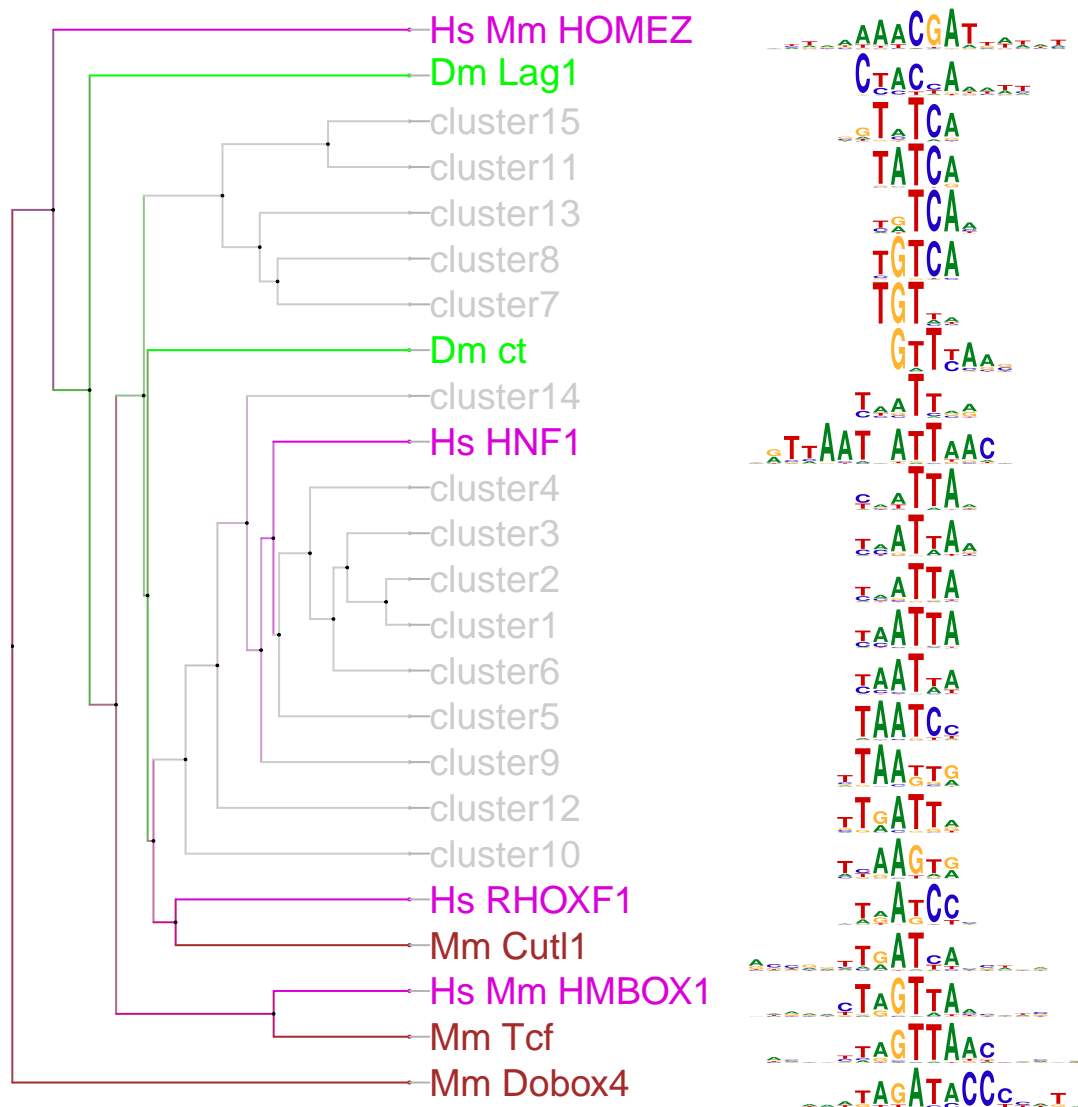
unUnique <- unUnique[!sapply(unUnique, function(.ele) .ele$name) %in% names(pcms)]
names(unUnique) <- paste("Dm_cluster", 1:length(unUnique), sep="")
unUnique <- mapply(function(.ele, n) {.ele@name <- n; .ele}, unUnique, names(unUnique))
pfms <- c(lapply(pcms.mono, pcm2pfm), unUnique)
pfms.cp <- c(lapply(pcms, pcm2pfm), unUnique)
pcms <- lapply(pfms, function(.ele){
  mat <- floor(.ele@mat * 1000)
  new("pcm", mat=mat, name=.ele@name)
})

dir.create("uniMotifTmp")
sta <- lapply(pcms, function(.ele){
  mat <- cbind(rownames(.ele@mat), "|", .ele@mat)
  write.table(mat, file=file.path("uniMotifTmp", paste(.ele@name, "pcm", sep=".")),
    row.names=FALSE, col.names=FALSE, sep="\t", quote=FALSE)
})

uniMotif <- readDataDoAna("uniMotifTmp", groupDistance=NA)
unlink("uniMotifTmp", recursive = TRUE)
attach(uniMotif)
motifs <- pfms.cp[leaves]
motifs=DNAmotifAlignment(motifs, minimalConsensus=3, threshold=.5)
leaveCols <- colorSet[gsub("^((Dm|Mm|Hs)_)ate$", "\\1", leaves)]
leaveCols[grepl("cluster", leaves)] <- "gray80"
leaves <- gsub("Dm_cluster", "cluster", leaves)
motifPiles(phylog=phylog, motifs,
  labels.leaves=leaves, col.leaves=leaveCols, col.tree=leaveCols,
  plotIndex=FALSE, clabel.leaves=1.5)

detach(uniMotif)

```



Supplementary Figure 8: Species-specific motifs in the HD family