Supporting information Script S1 and Table S1

Transcriptomic differences between euryhaline and stenohaline malaria vector sibling species in response to salinity stress

Hilary A. Uyhelji, Changde Cheng, Nora J. Besansky

Script S1. The R script used to compute RIF scores.

```
#R SCRIPT FOR CALCULATING RIF-1
#Generated by Changde Cheng and Hilary A Uyhelji based on the methods of Hudson et al.
 2012 (http://www.biomedcentral.com/1471-2164/13/356) and references therein
#2014.05.06

#Read expression data into rd, here as TMM-normalized counts per million

rd = read.csv("cpmTMMnorm_expression.csv", header=TRUE)
dim(rd)

#Place the data for 18 h Anopheles merus and 18 h An. coluzzii into the file m.
m = rd[,2:19]

#Make the gene names (AGAP IDs) the row names
rownames(m)=rd[,1]

#Make data frame that corresponds to the matrix
datTraits = data.frame( sps = rep(c(rep("maf",3),rep("aga",3)),3), water = rep(c(rep("fresh",1),
 rep("salty",2)),6));
datTraits

#Because spearman correlations cannot be computed properly if the standard deviation is zero,
 remove genes unless their cpmTMM expression is >0 in at least 10 libraries
keep = rowSums(m>0)>=10
table(keep)
dim(m)
m = m[keep,]
dim(m)


#Convert m into a matrix
m = as.matrix(m)

#Check the m matrix
head(m)

#Make datasets for just the species An. coluzzii, ie aga, and just An. merus, i.e. maf.
m.aga = m[ , datTraits$sps == 'aga']
m.maf = m[ , datTraits$sps == 'maf']
dim(m.aga)
colnames(m.aga)
dim(m.maf)
colnames(m.maf)
```

```
#Compute spearman correlations, but first transpose the m matrix so that you do a gene-by-gene
  correlation and not condition-by-condition correlation.
aga.cor = cor(t(m.aga),method='spearman')
sum(is.na(aga.cor))
maf.cor = cor(t(m.maf),method='spearman')
sum(is.na(maf.cor))


#Now that you have pairwise correlations for all genes, done independently for each species,
  take the difference in correlation
dcor = maf.cor - aga.cor

#Check dcor output
dim(dcor)
sum(is.na(dcor))

#Read in a list of gene names, ie AGAP IDs, for all significantly differently expressed genes
  (DEGs). Here, doing this for genes that were DE in either or both species, in any saltwater vs.
  0% (freshwater) contrast, where saltwater is either 10% or 20% salinity.
my.de = read.csv('DEG_10vs0_or_20vs0_EitherSp.csv',header=F)
dim(my.de)

#Compute the average abundance (expression level) of each DEG
my.abundance = apply(m[ rownames(m) %in% my.de$V1,],1,mean)

#To check my.abundance
head(my.abundance)
dim(my.abundance)

#Calculate differential expression
my.diff = m[,datTraits$sps == 'maf'] - m[,datTraits$sps == 'aga']

dim(my.diff)
colnames(my.diff)

my.diff = my.diff[ rownames(m) %in% my.de$V1,]
dim(my.diff)
my.diff = apply(my.diff,1,mean)
dim(my.diff)
head(my.diff)

#To check: this can be omitted if desired by using the hashtag # in front of the line
write.csv(my.diff, 'my.diff.csv')
write.csv(my.abundance, 'my.abundance.csv')

#Here computing PIF, as abundance-weighted differential expression.
```

```
myPIF = my.abundance * my.diff

#To check myPIF
write.csv(myPIF, 'myPIF.csv')
dim(myPIF)
head(myPIF)

#Now compute RIF as the product of PIF times the square of the dcor matrix, where dcor is the
  difference between species' gene-by-gene correlation matrices.
myRIF = myPIF %*% dcor[rownames(m) %in% my.de$V1,]^2
nde = nrow(my.de)
myRIF = myRIF/nde
dim(myRIF)

#Read in a list of gene names for all transcription factors in the PEST genome. Alternatively,
  read in a list of genes in QTL peaks.
myTF = read.csv('TranscriptionFactors_AgamP4.2_liberal_list.csv')

#Determine RIF values for these transcription factors, and scale it. Scaling converts raw RIF
  scores into Z-scores.
myRIF.tf = myRIF[,rownames(m) %in% myTF$Gene.stable.ID]
myRIF.tf.scaled = scale(myRIF.tf)

write.csv(myRIF.tf, 'RIF_UNscaled_0to20.csv')
write.csv(myRIF.tf.scaled,'RIF_scaled_0to20.csv')
```

Table S1. Number of significantly differentially expressed (DE) and unchanged genes (FDR<0.05, $|\log_2$ fold change$|>2$) from the 2012 experiment. All tests are of the contrast of expression in SW vs. FW. For *An. coluzzii* SW was 20% the salinity of seawater, and for *An. merus* salinity was 50%.

	Total DE	**No change**	**Up in SW**	**Down in SW**
An. coluzzii 18 h	19 (0.2%)	11006 (99.8%)	13 (0.1%)	6 (0.1%)
An. merus 18 h	351 (3.2%)	10674 (96.8%)	185 (1.7%)	166 (1.5%)
An. coluzzii 66 h	14 (0.1%)	11011 (99.9%)	11 (0.1%)	3 (0.0%)
An. merus 66 h	423 (3.8%)	10602 (96.2%)	234 (2.1%)	189 (1.7%)