

# CIF\_Analysis\_TryTrypDB

David H. Ardell

6/11/2019

This is the CIF Analysis Workflow for Kelly et al. (2019) using the old software for this (logofun and logotax by Eva Freyhult). In this version (v2), I regenerated the alignments.

## Aligning TryTrypDB Data to Preliminary Data and Filtering

Fasta data of Finalized TryTrypDB+Human Gene Data was received from Fateme. There were 3923 total sequences, of which 433 were human and 3490 were from TryTrypDB. Two of the human genes turned out to be selenocysteine (SEC) and were excluded.

I decided to co-align the new and preliminary data, so I could remap earlier work on site-selection and coordinate mapping to the new data. I use Perl one-liners to modify the IDs of data to make it easy to filter out the co-aligned preliminary data later and separate the aligned new data into functional classes. For preliminary data I used the file HomoC\_Cruzi\_red.fas, which had already been site-selected. I had to change two identifiers corresponding to BRENER\_52 and BRENER\_78 so the labeling and filtering strategy would work.

```
faswc final_geneset_Homo_TryTryp.fasta
fasgrep Homo final_geneset_Homo_TryTryp.fasta | faswc
fasgrep -v Homo final_geneset_Homo_TryTryp.fasta | faswc
fastr --degap HomoC_Cruzi_red.fas | fassub -i '$' '-old' > CruzeiHomoOld.fas
faswc CruzeiHomoOld.fas # 1356 old sequences
cat CruzeiHomoOld.fas final_geneset_Homo_TryTryp.fasta | fasgrep -i -- "-old" | faswc # 1356
cat CruzeiHomoOld.fas final_geneset_Homo_TryTryp.fasta > combined.fas
covea -o combined.covea ~/lib/tRNAscan-SE/TRNA2-euk.cm combined.fas
sreformat -u a2m combined.covea | tr '[a-z].' '[A-Z]-' | fasgrep -i -- "-OLD" | faswc # 1356
sreformat -u a2m combined.covea | tr '[a-z].' '[A-Z]-' > combined.covea.fas

# very careful site selection in seaview to match previous site selections

fasgrep -i -- "-OLD" combined.covea.sites.v2.fas | faswc # 1356
fasgrep -vi -- "-OLD" combined.covea.sites.v2.fas > covea.sites.v2.fas
faswc covea.sites.v2.fas # 3923
fasgrep -i "LMAJOR|LTROPICA|LAETHIOPICAL147|LGERBILLILEM452|LTURANICALEM423|LARABICALEM1108" covea.sites.v2.fas > LMAJOR.covea.fas
fasgrep -i "LINFANTUMJPCM5|LDONOVANI" covea.sites.v2.fas > INFANTUM.covea.fas
fasgrep -i "LMEXICANA|LAMAZONENSIS" covea.sites.v2.fas > MEXICANA.covea.fas
fasgrep -i "LPANAMENSIS|LBRAZILIENSIS" covea.sites.v2.fas > VIANNIA.covea.fas
fasgrep -i "LENRIETTII|LSPMARLEM" covea.sites.v2.fas > ENRIETTII.covea.fas
fasgrep -i "LSEYMOURIA|LPYRRHOCORI|CFASCICULATACFCL" covea.sites.v2.fas > LEPTOCRITH.covea.fas
fasgrep -i "TGRAYIANR4|TCRUZI" covea.sites.v2.fas | fasgrep -vi "TCRUZICLBRENER[~EN]" > AMTRYP.covea.fas
fasgrep -i "TBRUCEI|TCONGOLENSE|TEVANSI" covea.sites.v2.fas > AFTRYP.covea.fas
fasgrep -i "HOMO" covea.sites.v2.fas > HOMO.covea.fas
fasgrep -i ALA HOMO.covea.fas > HOMO_A.fas
fasgrep -i CYS HOMO.covea.fas > HOMO_C.fas
fasgrep -i ASP HOMO.covea.fas > HOMO_D.fas
fasgrep -i GLU HOMO.covea.fas > HOMO_E.fas
fasgrep -i PHE HOMO.covea.fas > HOMO_F.fas
```

```

fasgrep -i GLY HOMO.covea.fas > HOMO_G.fas
fasgrep -i HIS HOMO.covea.fas > HOMO_H.fas
fasgrep -i ILE HOMO.covea.fas > HOMO_I.fas
fasgrep -i LYS HOMO.covea.fas > HOMO_K.fas
fasgrep -i LEU HOMO.covea.fas > HOMO_L.fas
fasgrep -i "\WMET" HOMO.covea.fas > HOMO_M.fas
fasgrep -i ASN HOMO.covea.fas > HOMO_N.fas
fasgrep -i PRO HOMO.covea.fas > HOMO_P.fas
fasgrep -i GLN HOMO.covea.fas > HOMO_Q.fas
fasgrep -i ARG HOMO.covea.fas > HOMO_R.fas
fasgrep -i SER HOMO.covea.fas > HOMO_S.fas
fasgrep -i THR HOMO.covea.fas > HOMO_T.fas
fasgrep -i VAL HOMO.covea.fas > HOMO_V.fas
fasgrep -i TRP HOMO.covea.fas > HOMO_W.fas
fasgrep -i IMET HOMO.covea.fas > HOMO_X.fas
fasgrep -i TYR HOMO.covea.fas > HOMO_Y.fas
fasconvert -fo clustalw HOMO_*.fas
perl -e 'foreach my $clade (qw/MAJOR INFANTUM MEXICANA VIANNIA ENRIETTII LEPTOCRITH AMTRYP AFTRYP/){for

```

## Function Logos

```

logofun --states ACGT -f 1 -b 1 -o HOMO.v3 HOMO_*.aln
logofun --states ACGT -f 1 -b 1 -o MAJOR.v3 MAJOR_*.aln
logofun --states ACGT -f 1 -b 1 -o INFANTUM.v3 INFANTUM_*.aln
logofun --states ACGT -f 1 -b 1 -o MEXICANA.v3 MEXICANA_*.aln
logofun --states ACGT -f 1 -b 1 -o VIANNIA.v3 VIANNIA_*.aln
logofun --states ACGT -f 1 -b 1 -o AMTRYP.v3 AMTRYP_*.aln
logofun --states ACGT -f 1 -b 1 -o AFTRYP.v3 AFTRYP_*.aln
logofun --states ACGT -f 1 -b 1 -o LEPTOCRITH.v3 LEPTOCRITH_*.aln
logofun --states ACGT -f 1 -b 1 -o ENRIETTII.v3 ENRIETTII_*.aln

```

## ID Logos

```

logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_MAJOR_ID 'HOMO_*.aln' 'MAJOR_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_INFANTUM_ID 'HOMO_*.aln' 'INFANTUM_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_MEXICANA_ID 'HOMO_*.aln' 'MEXICANA_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_VIANNIA_ID 'HOMO_*.aln' 'VIANNIA_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_ENRIETTII_ID 'HOMO_*.aln' 'ENRIETTII_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_LEPTOCRITH_ID 'HOMO_*.aln' 'LEPTOCRITH_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_AFTRYP_ID 'HOMO_*.aln' 'AFTRYP_*.aln'
logotax -b 3 -l infodiff -p other --states 'ACGT' -o HOMO_AMTRYP_ID 'HOMO_*.aln' 'AMTRYP_*.aln'

```

## KLD Logos



```
# THESE ARE THE COORDINATES FOR THE TRNA STRUCTURE BACKBONE IN THE FIGURES
```

```
line.x <- c(6.875,6.500,6.125,5.750,5.375,5.000,4.625,4.625,5.000,5.000,2.375
,2.750,2.375,2.750,2.500,2.875,3.250,2.875,2.500,2.125,1.750,1.375,1.000
,0.625,0.250,0.625,1.000,1.500,1.125,1.500,1.125,1.500,1.125,1.500,1.125
,1.500,1.125,0.625,1.000,1.375,1.750,2.125,2.500,2.875,2.375,2.750,2.375
,2.750,2.375,4.250,4.250,3.875,4.250,3.875,4.250,4.250,3.875,3.500,3.125
,2.750,2.250,1.875,1.500,1.125,1.500,1.875,2.250,2.750,3.125,3.500,3.875
,4.250,4.625,5.000,5.375,5.750,6.125,6.500,6.875,7.250,7.625,8.000,8.375);
```

```
line.y <- c(8.875,8.500,8.875,8.500,8.875,8.500,8.875,7.375,7.000
,3.500,3.500,3.875,4.250,4.625,5.125,5.500,5.875,6.250,6.625
,7.000,7.375,7.000,6.625,6.250,5.875,5.500,5.125,4.625,4.250
,3.875,3.500,3.125,2.750,2.375,2.000,1.625,1.250,0.750,0.375
,0.000,-0.375,0.000,0.375,0.750,1.250,1.625,2.000,2.375,2.750
,2.750,3.875,4.250,4.625,5.000,5.375,8.500,8.875,8.500,8.875
,8.500,8.000,8.375,8.750,9.125,9.500,9.875,10.250,9.750,10.125
,9.750,10.125,9.750,10.125,9.750,10.125,9.750,10.125,9.750,10.125
,9.750,10.125,9.750,10.125);
```

```
## THESE ARE FOR THE SPRINZL COORD LABELS
```

```
coord.labels <- c("1","5","10","14A","18","20C","25","30","35","40","45","50","55","60","65","70");
coord.labels.x <- c(6.875,5.375,2.375,2.875,1.750,0.625,1.125,1.500,1.750,2.750,3.875,3.875,1.875,2.250
coord.labels.y <- c(8.875,8.875,3.500,5.500,7.375,5.500,3.500,1.625,-0.375,1.625,4.250,8.875,8.375,10.250);
```

```
xbump <- 0.35;
ybump <- 0.4;
```

```
up <- c(5,13,14,15,16,17);
up.coord.labels <- coord.labels[up];
up.coord.labels.x <- coord.labels.x[up];
up.coord.labels.y <- coord.labels.y[up] + ybump;
```

```
dn <- c(1,2,9,11,12);
dn.coord.labels <- coord.labels[dn];
dn.coord.labels.x <- coord.labels.x[dn];
dn.coord.labels.y <- coord.labels.y[dn] - ybump;
```

```
lt <- c(3,4,7);
lt.coord.labels <- coord.labels[lt];
lt.coord.labels.x <- coord.labels.x[lt] - xbump;
lt.coord.labels.y <- coord.labels.y[lt];
```

```
rt <- c(6,8,10);
rt.coord.labels <- coord.labels[rt];
rt.coord.labels.x <- coord.labels.x[rt] + xbump;
rt.coord.labels.y <- coord.labels.y[rt];
```

```
all.bubble <- function(df,name="bubble",alpha=0.5,fact=0.5,area=TRUE,legend=FALSE) {
  gains <- (df$gainbits * df$gainfht);
```

```

convs <- (df$convbites * df$convfht);
map2rgb <- function (c) { rgb(t(col2rgb(c))/255,alpha=alpha);}
colormap <- function (g,c) {
  y <- rep(0,length(g));
  y[g < 0.48 & c < 0.44] <- rgb(t(col2rgb("white"))/255,alpha=alpha);
  y[g >= 0.48 & g < 0.95 & c < 0.44] <- rgb(t(col2rgb("darkred"))/255,alpha=alpha);
  y[g >= 0.95 & c < 0.44] <- rgb(t(col2rgb("red"))/255,alpha=alpha);
  y[g < 0.48 & c >= 0.44 & c < 0.70] <- rgb(t(col2rgb("darkblue"))/255,alpha=alpha);
  y[g >= 0.48 & g < 0.95 & c >= 0.44 & c < 0.70] <- rgb(t(col2rgb("darkmagenta"))/255,alpha=alpha);
  y[g >= 0.95 & c >= 0.44 & c < 0.70] <- map2rgb("deeppink");
  y[g < 0.48 & c >= 0.70] <- map2rgb("blue");
  y[g >= 0.48 & g < 0.95 & c >= 0.70] <- map2rgb("blueviolet");
  y[g >= 0.95 & c >= 0.70] <- map2rgb("magenta");
  y;
}
colors <- colormap(gains,convs);

widthmap <- function (g,c) {
  y <- rep(1,length(g));
  y[g < 0.48 & c < 0.44] <- 1;
  y[g >= 0.48 | c >= 0.44] <- 2;
  y[g >= 0.95 | c >= 0.70] <- 3;
  y;
}
widths <- widthmap(gains,convs);

for (class in levels(df$aa)) {
  filenm <- paste(name,"_",class,".pdf",sep="");
  pdf(file=filenm,version="1.4");
  bubbleplot(
    df$x[df$aa == class],
    df$y[df$aa == class],
    (df$fbits[df$aa == class] * df$fht[df$aa == class]),
    fact=fact, #0.265165 = sqrt(2*0.375^2)/2
    area=area,
    fg = rgb(t(col2rgb("black"))/255),
    bg = colors[df$aa == class],
    box=FALSE,
    axes=FALSE,
    lwd=2
  );
  lines(line.x,line.y);
  text(labels=up.coord.labels,x=up.coord.labels.x,y=up.coord.labels.y);
  text(labels=dn.coord.labels,x=dn.coord.labels.x,y=dn.coord.labels.y);
  text(labels=lt.coord.labels,x=lt.coord.labels.x,y=lt.coord.labels.y);
  text(labels=rt.coord.labels,x=rt.coord.labels.x,y=rt.coord.labels.y);
  if (legend) {
    legend.x <- rep(df$x[df$aa == "X" & df$state == "A" &
      (df$sprinzl == "68" | df$sprinzl == "70" |
        df$sprinzl == "72" )] + c(0.2,0.3,0.4),3);
    legend.y <- c(rep(df$y[df$aa == "X" & df$state == "A" & df$sprinzl == "31"],3),
      rep(df$y[df$aa == "X" & df$state == "A" & df$sprinzl == "27"],3),
      rep(df$y[df$aa == "X" & df$state == "A" & df$sprinzl == "23"],3));
  }
}

```

```

legend.z <- rep(2.2,9);
legend.c <- colormap(c(0,0,0,0.5,0.5,0.5,1,1,1),c(0,0.5,1,0,0.5,1,0,0.5,1));
bubbleplot(legend.x, legend.y, legend.z,fact=fact,area=area,bg=legend.c,
  add=TRUE,
  box=FALSE,axes=FALSE,lwd=2);

}
prime.x <- df$x[df$aa == "X" & df$state == "A" &
  (df$sprinzi == "1" | df$sprinzi == "75")];
prime.y <- df$y[df$aa == "X" & df$state == "A" &
  (df$sprinzi == "1" | df$sprinzi == "75")];
text(labels=c("5'", "3'"),x=prime.x,y=prime.y,adj=c(-0.9,0));
dev.off();
}
}

all.bubble(MAJOR,name="MAJOR");

```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```
all.bubble(INFANTUM,name="INFANTUM");
```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```
all.bubble(MEXICANA,name="MEXICANA");
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
## Radius scale factor is 0.5
```

```
all.bubble(VIANNIA,name="VIANNIA");
```

```
## Radius scale factor is 0.5
```



```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

all.bubble(ENRIETTII,name="ENRIETTII");

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```
all.bubble(LEPTOCRITH,name="LEPTOCRITH");
```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```
all.bubble(AMTRYP,name="AMTRYP");
```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

```
## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

all.bubble(AFTRYP,name="AFTRYP")

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5
```

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5

## Radius scale factor is 0.5