

Supplemental File for Kitchen et al. *Genomic variants among threatened Acropora corals*

Discriminant Factorial Correspondence Analysis

The R code provided in this document will perform the discriminant analysis on the factorial correspondence scores calculated in GENETIX version 4.05 for the 5 microsatellite markers (Baums et al. 2005) or 8 single nucleotide variant markers (this study). For details on the SNV markers design and targets, see Tables S3-S5.

Load the packages required for the analyses.

```
library(MASS)
library(klaR)
library(dplyr)
library(magrittr)
library(pander)
library(tables)
library(heplots)
library(caret)
library(ggplot2)
library(vegan)
```

Microsatellites

Import the factorial correspondence analysis scores.

```
mst<-read.table("FCA_Msat_output.txt", check.names=FALSE, header=T, row.names=1)
mst
```

##	Pop	Color	F1	F2	F3	F4
## 1000	P	P	-166.76	-578.22	81.50	116.35
## 1545	P	P	-36.17	-615.71	103.11	-243.18
## 1617	P	P	-136.37	-873.41	82.78	950.97
## 3141	P	P	-442.36	-761.24	122.76	1499.93
## 4067	P	P	10.64	-1390.60	-2767.22	-2196.50
## 5551	P	P	157.25	-646.23	231.43	811.11
## 5977	P	P	-77.19	-451.84	-941.87	-277.08
## 8900	P	P	-351.44	-1029.33	-674.23	-200.39
## 13722	P	P	-182.74	-751.31	-184.71	720.78
## 13909	P	P	-35.69	-664.22	-124.26	-410.67
## 1770	C	C	-95.33	1169.51	-7.29	25.95
## 2454	C	C	-403.86	1383.91	-379.28	-709.50
## 2984	C	C	153.48	575.47	-597.66	-376.21
## 8870	C	C	-260.39	954.61	33.44	60.09
## 9412	C	C	-198.07	1122.16	19.08	-237.11
## 9468	C	C	-158.07	952.37	120.85	-346.60
## 13772	C	C	-225.43	1258.48	-103.28	-218.86
## 13829	C	C	-266.37	1231.34	-6.00	-136.33
## 13923	C	C	-473.36	1208.61	-324.08	-31.70
## 1769	F1	F1	-346.02	-507.11	150.11	894.97
## 1834	F1	F1	-128.09	-160.73	-177.20	41.70
## 8939	F1	F1	-238.28	-486.04	685.63	-388.46

```
## 1302 H H 369.55 -222.47 -151.55 -406.53
## 1303 H H 323.75 -140.62 -307.70 -230.14
## 1667 H H -504.38 -382.96 23.41 517.51
## 1835 H H -322.94 325.15 22.32 -177.91
## 1837 H H 92.34 -235.67 312.87 1093.77
## 3070 H H 51.63 -63.79 56.81 504.18
## 3071 H H -272.81 -545.02 269.21 990.38
## 4062 H H -39.46 -433.04 671.42 -88.21
## 5993 H H -314.42 -349.41 409.15 294.00
## 6779 H H -211.09 -131.70 41.08 -58.56
## 6791 H H -357.49 -10.88 -28.72 264.50
## 8867 H H -112.08 1430.93 -244.18 -58.39
## 11274 H H -250.99 -986.90 404.56 -1415.49
## 11293 H H -154.02 -484.90 212.12 155.28
## 11327 H H -416.02 261.26 -522.75 -197.10
## 11328 H H -377.50 261.09 -404.39 220.12
## 11334 H H -38.17 -308.74 658.89 102.65
## 11409 H H -184.61 277.88 574.65 -611.22
## 11414 H H -391.66 67.09 -382.31 277.66
## 13764 H H -394.68 -114.64 -75.91 1257.98
## 13778 H H -493.60 277.56 358.03 499.86
## 13799 H H -14.06 -410.71 -189.49 449.47
## 13807 H H -121.45 -791.46 -578.68 99.64
## 13708 H H -521.45 1486.53 -564.29 104.34
## 14280 H U -324.79 140.59 27.85 436.98
## 14281 H U -303.19 470.75 98.20 -241.57
## 14282 H U 73.16 -125.08 -418.86 -235.21
## 14323 H U -121.45 -791.46 -578.68 99.64
## 14324 H U -274.22 -241.22 1721.37 -528.66
## 14326 H U -138.85 -14.42 947.79 -638.66
## 14327 H U -356.71 425.18 -107.86 270.98
## 14328 H U 503.68 -465.95 1882.85 -1436.88
## 14329 H U -477.73 375.50 -571.56 38.27
## 14330 H U -581.94 635.86 166.37 -252.92
## 14333 H U -362.91 -746.87 -41.81 -307.75
## 14334 H U -558.85 311.24 341.33 584.03
## 14335 H U 2895.78 266.45 -146.95 270.75
## 14338 H U 96.60 -133.44 -257.87 -218.12
## 14331 H U -581.94 635.86 166.37 -252.92
## 14325 H U 503.68 -465.95 1882.85 -1436.88
## 14336 H U 2895.78 266.45 -146.95 270.75
## 14337 H U 2895.78 266.45 -146.95 270.75
## 14339 H U 2895.78 266.45 -146.95 270.75
## 14322 H U -121.45 -791.46 -578.68 99.64
```

Perform the linear discriminant analysis (LDA) on the reference sample FCA scores. Then, predict sample membership based on the posterior probabilities.

```
#subset the reference samples
sub=1:46
#LDA on the reference samples
mst.lda <- lda(Pop ~ F1 + F2 +F3 + F4,mst,subset=sub)
mst.lda
```

```
## Call:
```

```
## lda(Pop ~ F1 + F2 + F3 + F4, data = mst, subset = sub)
##
## Prior probabilities of groups:
##      C      F1      H      P
## 0.19565217 0.06521739 0.52173913 0.21739130
##
## Group means:
##      F1      F2      F3      F4
## C -214.1556 1095.16222 -138.24667 -218.9189
## F1 -237.4633 -384.62667 219.51333 182.7367
## H -193.9838 -51.05917 23.52292 149.4913
## P -126.0830 -776.21100 -407.07100 77.1320
##
## Coefficients of linear discriminants:
##      LD1      LD2      LD3
## F1 -0.0005717947 0.0016332345 -0.0029457752
## F2 -0.0022150253 0.0002931081 -0.0004147440
## F3 -0.0004946753 -0.0019202052 0.0003049358
## F4 0.0004886279 0.0002748860 -0.0015108708
##
## Proportion of trace:
##      LD1      LD2      LD3
## 0.9375 0.0601 0.0024

#predict the group membership for the whole data set including test hybrids (n=20)
p<-predict(mst.lda,newdata=mst)
#convert posterior probabilities to %
round(p$posterior, 2)*100

##      C F1 H P
## 1000 0 13 61 26
## 1545 0 12 66 22
## 1617 0 7 25 68
## 3141 0 9 24 67
## 4067 0 0 0 100
## 5551 0 6 54 39
## 5977 0 2 36 62
## 8900 0 4 10 86
## 13722 0 6 27 67
## 13909 0 10 58 32
## 1770 95 0 5 0
## 2454 99 0 1 0
## 2984 58 0 41 1
## 8870 80 0 20 0
## 9412 95 0 5 0
## 9468 89 0 11 0
## 13772 97 0 3 0
## 13829 96 0 4 0
## 13923 92 0 8 0
## 1769 0 13 55 32
## 1834 1 6 81 12
## 8939 0 26 70 4
## 1302 2 3 86 10
## 1303 2 2 84 12
## 1667 0 16 62 22
```

```
## 1835 9 5 84 1
## 1837 0 5 81 14
## 3070 1 4 87 8
## 3071 0 13 56 31
## 4062 0 16 79 4
## 5993 0 18 75 8
## 6779 1 9 84 6
## 6791 1 9 84 6
## 8867 99 0 1 0
## 11274 0 44 46 10
## 11293 0 13 70 16
## 11327 5 5 86 5
## 11328 3 4 87 5
## 11334 0 13 83 4
## 11409 18 6 76 0
## 11414 1 6 83 11
## 13764 0 7 71 22
## 13778 2 9 87 1
## 13799 0 5 61 34
## 13807 0 3 20 77
## 13708 98 0 2 0
## 14280 2 6 88 4
## 14281 23 3 73 0
## 14282 1 3 82 14
## 14323 0 3 20 77
## 14324 1 37 62 0
## 14326 4 14 81 0
## 14327 9 4 86 2
## 14328 4 19 77 0
## 14329 6 4 86 4
## 14330 31 4 64 0
## 14333 0 20 47 33
## 14334 2 9 87 1
## 14335 97 0 3 0
## 14338 1 4 84 10
## 14331 31 4 64 0
## 14325 4 19 77 0
## 14336 97 0 3 0
## 14337 97 0 3 0
## 14339 97 0 3 0
## 14322 0 3 20 77
```

Summarize group membership for the four taxon (A. palmata = P, A. cervicornis = C, F1 and hybrid = H)

```
#Table of field calls compared to predicted taxon assignment
```

```
freqtable <- table(Predicted=p$class, Field_Call=mst$Pop)
freqtable
```

```
##           Field_Call
## Predicted C F1 H P
##           C 9 0 6 0
##           F1 0 0 0 0
##           H 0 3 35 4
##           P 0 0 3 6
```

Compile data into a new data frame and write to file. Posterior probabilities estimates were reported in Table

S2.

```
#compile table
mst.dat<-data.frame(p$posterior,
                    ObservedGroup = mst$Pop,
                    PredictedGroup = p$class, p$x)

write.table(mst.dat, file="predictedProbs_DFAC_msat.txt", sep="\t")
```

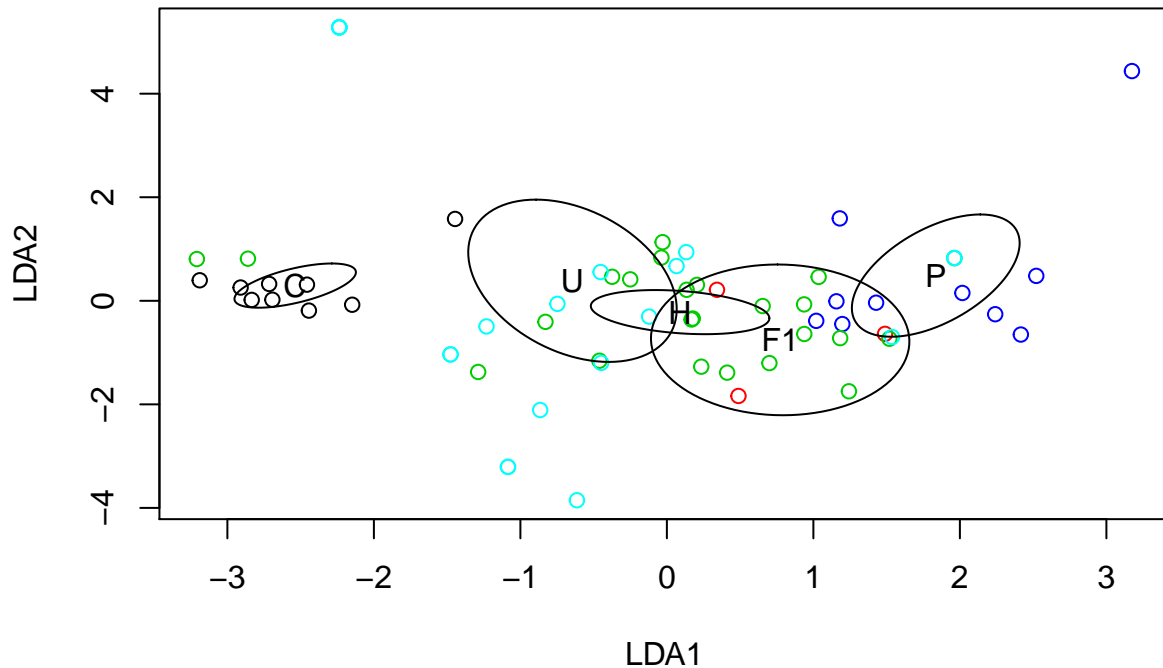
Create Figure 6 panel A.

```
mst$LDA1 <- p$x[,1]
mst$LDA2 <- p$x[,2]
gMeansMsat <- mst %>% group_by(Pop) %>% select(LDA1,LDA2) %>% summarise_each(funs(mean))
gMeansMsat
```

```
## # A tibble: 4 x 3
##   Pop    LDA1    LDA2
##   <fct> <dbl>   <dbl>
## 1 C     -2.54    0.294
## 2 F1    0.773   -0.755
## 3 H     -0.244   0.0566
## 4 P     1.84    0.487
```

```
#calculate ordination ellipses for each group based on LDA1 and LDA2
#vegan covariance ellipse function
veganCovEllipse<-function (cov, center = c(0, 0), scale = 1, npoints = 100)
{
  theta <- (0:npoints) * 2 * pi/npoints
  Circle <- cbind(cos(theta), sin(theta))
  t(center + scale * t(Circle %*% chol(cov)))
}

#ordination for LDA1 and LDA2 grouped by color which equals taxon groups
plot(mst[,7:8],col=mst$Color)
ord<-ordiellipse(mst[,7:8], mst$Color, display = "sites",
                 kind = "se", conf = 0.95, label = T)
```



```
#create new dataframe with ellipses
df_ell <- data.frame()
for(g in levels(mst$Color)){
  df_ell <- rbind(df_ell, cbind(as.data.frame(with(mst[mst$Color==g,],
    veganCovEllipse(ord[[g]]$cov,ord[[g]]$center,ord[[g]]$scale)))
    ,Color=g))
}

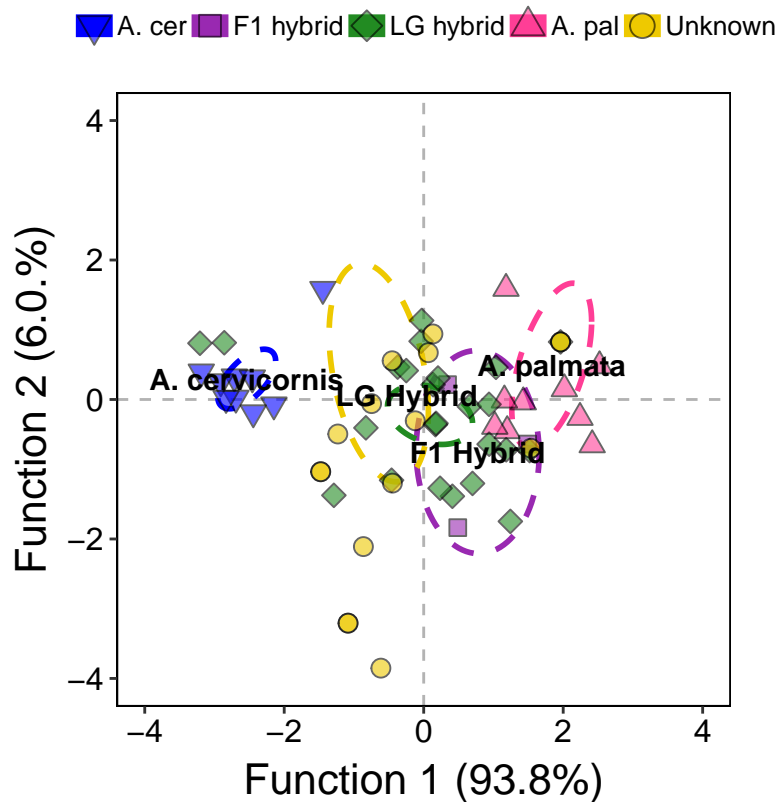
#plot
m<-ggplot(mst, aes(LDA1, LDA2, fill=Color, shape=Color))+
  geom_hline(yintercept=0, col="grey70", linetype=2)+
  geom_vline(xintercept=0, col="grey70", linetype=2)+
  geom_path(data=df_ell, aes(x=LDA1, y=LDA2, colour=Color, group=Color), size=1, linetype=2)+
  geom_point(alpha = 0.6, size=3) +
  xlab("Function 1 (93.8%)")+
  ylab("Function 2 (6.0%)")+
  xlim(-4,4)+
  ylim(-4,4)+
  geom_text(data = gMeansMsat, inherit.aes=FALSE,
    aes(LDA1, LDA2,label = c("A. cervicornis", "F1 Hybrid","LG Hybrid","A. palmata")) ,
    color = "black",
    fontface="bold")+
  scale_fill_manual(values=c("#0000FF","#9929B1","#228B22", "#FF3E96","#EEC900"),
    labels = c("A. cer", "F1 hybrid","LG hybrid","A. pal", "Unknown"))+
  scale_color_manual(values=c("#0000FF","#9929B1","#228B22", "#FF3E96","#EEC900"),
    labels = c("A. cer", "F1 hybrid","LG hybrid","A. pal", "Unknown"))+
```

```

scale_shape_manual(values=c(25,22,23,24,21),
  labels = c("A. cer", "F1 hybrid","LG hybrid","A. pal", "Unknown"))+
guides(color = guide_legend(override.aes = list(size=4)))+
theme(legend.position=c("top"),
  legend.text = element_text(size = 10, color = "black"),
  legend.direction = "horizontal",
  axis.text=element_text(size = 12, color="black"),
  axis.title.x = element_text(vjust = 0, size=16),
  axis.title.y = element_text(vjust = 2, size=16),
  panel.background=element_rect(fill="white"),
  legend.key=element_rect(fill="white"),
  panel.border=element_rect(fill=NA),
  legend.title=element_blank(),
  plot.margin = unit(c(1, 1, 0.5, 0.5), "lines"),
  text = element_text(color = "black")) +
coord_equal()

```

m



Single Nucleotide Variants

Repeat the steps above, but this time using the SNV FCA scores.

```

#import data
snp<-read.table("FCA_snps_output.txt", check.names=FALSE, header=T, row.names=1)
#view FCA scores

```

snp

##	Pop	Color	F1	F2	F3	F4
## 1000	P	P	-1000.22	-84.50	13.47	-42.83
## 1545	P	P	-1000.22	-84.50	13.47	-42.83
## 1617	P	P	-1000.22	-84.50	13.47	-42.83
## 3141	P	P	-1000.22	-84.50	13.47	-42.83
## 4067	P	P	-1000.22	-84.50	13.47	-42.83
## 5551	P	P	-1000.22	-84.50	13.47	-42.83
## 5977	P	P	-1000.22	-84.50	13.47	-42.83
## 8900	P	P	-1000.22	-84.50	13.47	-42.83
## 13722	P	P	-1000.22	-84.50	13.47	-42.83
## 13909	P	P	-1000.22	-84.50	13.47	-42.83
## 1770	C	C	995.97	48.86	-0.92	-7.13
## 2454	C	C	995.97	48.86	-0.92	-7.13
## 2984	C	C	995.97	48.86	-0.92	-7.13
## 8870	C	C	995.97	48.86	-0.92	-7.13
## 9412	C	C	995.97	48.86	-0.92	-7.13
## 9468	C	C	995.97	48.86	-0.92	-7.13
## 13772	C	C	995.97	48.86	-0.92	-7.13
## 13829	C	C	995.97	48.86	-0.92	-7.13
## 13923	C	C	995.97	48.86	-0.92	-7.13
## 1769	F1	F1	-2.12	-17.82	6.28	-24.98
## 1834	F1	F1	-2.12	-17.82	6.28	-24.98
## 8939	F1	F1	-2.12	-17.82	6.28	-24.98
## 1302	H	H	-252.93	-48.84	302.39	139.40
## 1303	H	H	-245.21	-227.55	-70.62	280.06
## 1667	H	H	-134.66	85.24	-76.83	-26.07
## 1835	H	H	-110.83	-301.95	-111.78	84.73
## 1837	H	H	-2.12	-17.82	6.28	-24.98
## 3070	H	H	-418.48	591.87	164.86	53.03
## 3071	H	H	-530.70	305.06	-206.25	166.69
## 4062	H	H	-882.14	-32.61	65.26	228.54
## 5993	H	H	-648.77	253.17	-258.04	-104.68
## 6779	H	H	-2.12	-17.82	6.28	-24.98
## 6791	H	H	-110.83	-301.95	-111.78	84.73
## 8867	H	H	770.85	-340.69	135.97	71.62
## 11274	H	H	-2.12	-17.82	6.28	-24.98
## 11293	H	H	-548.32	641.07	46.43	40.08
## 11327	H	H	995.97	48.86	-0.92	-7.13
## 11328	H	H	995.97	48.86	-0.92	-7.13
## 11334	H	H	-2.12	-17.82	6.28	-24.98
## 11409	H	H	995.97	48.86	-0.92	-7.13
## 11414	H	H	995.97	48.86	-0.92	-7.13
## 13764	H	H	887.27	-235.28	-118.98	102.58
## 13778	H	H	-2.12	-17.82	6.28	-24.98
## 13799	H	H	-2.12	-17.82	6.28	-24.98
## 13807	H	H	-657.02	356.94	-71.63	149.78
## 13708	H	H	-2.12	-17.82	6.28	-24.98
## 14280	H	U	-2.12	-17.82	6.28	-24.98
## 14281	H	U	-2.12	-17.82	6.28	-24.98
## 14282	H	U	-2.12	-17.82	6.28	-24.98
## 14323	H	U	-2.12	-17.82	6.28	-24.98
## 14324	H	U	-2.12	-17.82	6.28	-24.98


```
## 14326 H U -2.12 -17.82 6.28 -24.98
## 14327 H U -2.12 -17.82 6.28 -24.98
## 14328 H U -2.12 -17.82 6.28 -24.98
## 14329 H U -2.12 -17.82 6.28 -24.98
## 14330 H U -2.12 -17.82 6.28 -24.98
## 14333 H U -2.12 -17.82 6.28 -24.98
## 14334 H U -2.12 -17.82 6.28 -24.98
## 14335 H U -2.12 -17.82 6.28 -24.98
## 14338 H U -2.12 -17.82 6.28 -24.98
## 14331 H U -2.12 -17.82 6.28 -24.98
## 14325 H U -2.12 -17.82 6.28 -24.98
## 14336 H U -2.12 -17.82 6.28 -24.98
## 14337 H U -2.12 -17.82 6.28 -24.98
## 14339 H U -2.12 -17.82 6.28 -24.98
## 14322 H U -2.12 -17.82 6.28 -24.98
```

#LDA on the reference samples

```
snp.lda <- lda(Pop ~ F1 + F2 +F3 + F4,snp,subset=sub)
snp.lda
```

Call:

```
## lda(Pop ~ F1 + F2 + F3 + F4, data = snp, subset = sub)
```

##

Prior probabilities of groups:

```
## C F1 H P
```

```
## 0.19565217 0.06521739 0.52173913 0.21739130
```

##

Group means:

```
## F1 F2 F3 F4
```

```
## C 995.97000 48.86000 -0.92 -7.13000
```

```
## F1 -2.12000 -17.82000 6.28 -24.98000
```

```
## H 45.30292 33.96583 -11.28 44.46292
```

```
## P -1000.22000 -84.50000 13.47 -42.83000
```

##

Coefficients of linear discriminants:

```
## LD1 LD2
```

```
## F1 -0.002757999 -0.0005285885
```

```
## F2 -0.003743833 0.0011171809
```

```
## F3 0.002797797 -0.0027536483
```

```
## F4 -0.009370000 0.0125459867
```

##

Proportion of trace:

```
## LD1 LD2
```

```
## 0.9498 0.0502
```

#predict the group membership for the whole data set including test hybrids (n=20)

```
p<-predict(snp.lda,newdata=snp)
```

#convert posterior probabilities to %

```
round(p$posterior, 2)*100
```

```
## C F1 H P
```

```
## 1000 0 0 0 100
```

```
## 1545 0 0 0 100
```

```
## 1617 0 0 0 100
```

```
## 3141 0 0 0 100
```

```

## 4067  0  0  0 100
## 5551  0  0  0 100
## 5977  0  0  0 100
## 8900  0  0  0 100
## 13722 0  0  0 100
## 13909 0  0  0 100
## 1770  88  0 12  0
## 2454  88  0 12  0
## 2984  88  0 12  0
## 8870  88  0 12  0
## 9412  88  0 12  0
## 9468  88  0 12  0
## 13772 88  0 12  0
## 13829 88  0 12  0
## 13923 88  0 12  0
## 1769  0 25 74  0
## 1834  0 25 74  0
## 8939  0 25 74  0
## 1302  0  9 90  0
## 1303  0  0 100  0
## 1667  1 15 84  0
## 1835  0  8 92  0
## 1837  0 25 74  0
## 3070  2  2 96  0
## 3071  1  0 99  0
## 4062  0  2 98  0
## 5993  0 31 61  8
## 6779  0 25 74  0
## 6791  0  8 92  0
## 8867 11  6 83  0
## 11274 0 25 74  0
## 11293 2  1 97  0
## 11327 88  0 12  0
## 11328 88  0 12  0
## 11334 0 25 74  0
## 11409 88  0 12  0
## 11414 88  0 12  0
## 13764 59  0 41  0
## 13778 0 25 74  0
## 13799 0 25 74  0
## 13807 0  0 99  0
## 13708 0 25 74  0
## 14280 0 25 74  0
## 14281 0 25 74  0
## 14282 0 25 74  0
## 14323 0 25 74  0
## 14324 0 25 74  0
## 14326 0 25 74  0
## 14327 0 25 74  0
## 14328 0 25 74  0
## 14329 0 25 74  0
## 14330 0 25 74  0
## 14333 0 25 74  0
## 14334 0 25 74  0

```

```
## 14335 0 25 74 0
## 14338 0 25 74 0
## 14331 0 25 74 0
## 14325 0 25 74 0
## 14336 0 25 74 0
## 14337 0 25 74 0
## 14339 0 25 74 0
## 14322 0 25 74 0
```

```
#Table of field calls compared to predicted taxon assignment
freqtable <- table(Predicted=p$class, Field_Call=snp$Pop)
freqtable
```

```
##           Field_Call
## Predicted C F1 H P
##           C    9 0 5 0
##           F1   0 0 0 0
##           H    0 3 39 0
##           P    0 0 0 10
```

```
#compile table
snp.dat<-data.frame(p$posterior,
                    ObservedGroup = snp$Pop,
                    PredictedGroup = p$class, p$x)

write.table(snp.dat, file="predictedProbs_DFAC_SNPs.txt", sep="\t")

snp$LDA1 <- p$x[,1]
snp$LDA2 <- p$x[,2]
gMeanssnp <- snp %>% group_by(Pop) %>% select(LDA1,LDA2) %>% summarise_each(funs(mean))
```

```
## Adding missing grouping variables: `Pop`
## `summarise_each()` is deprecated.
## Use `summarise_all()`, `summarise_at()` or `summarise_if()` instead.
## To map `funs` over all variables, use `summarise_all()`
```

```
gMeanssnp
```

```
## # A tibble: 4 x 3
##   Pop      LDA1      LDA2
##   <fct>   <dbl>   <dbl>
## 1 C      -2.72   -0.711
## 2 F1      0.465   -0.501
## 3 H      -0.0936  0.0180
## 4 P       3.65    -0.292
```

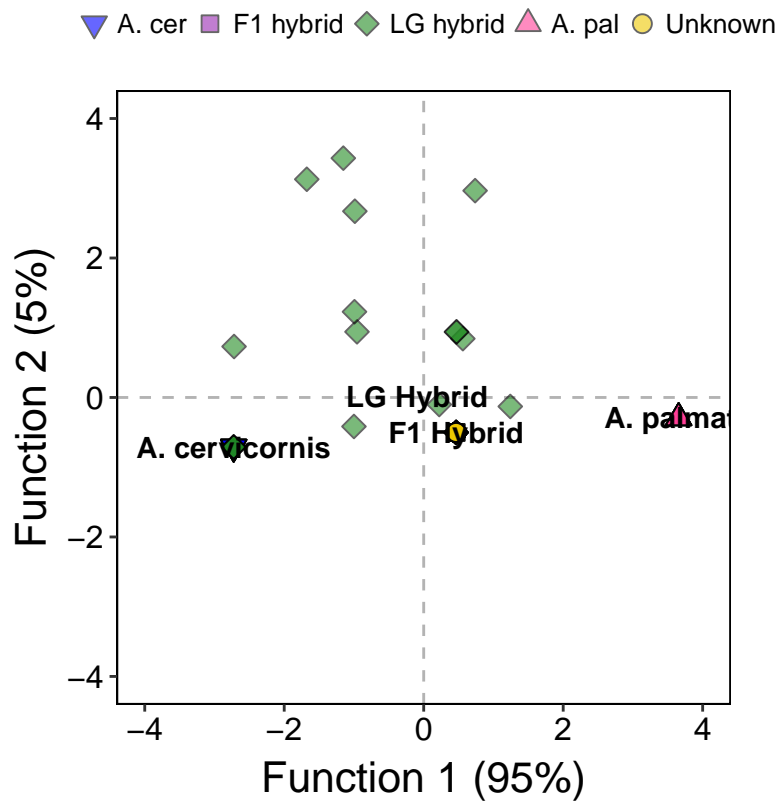
```
s<-ggplot(snp, aes(LDA1, LDA2, fill=Color, shape=Color))+
  geom_hline(yintercept=0, col="grey70", linetype=2)+
  geom_vline(xintercept=0, col="grey70", linetype=2)+
  geom_point(alpha = 0.6, size=3) +
  xlab("Function 1 (95%)" )+
  ylab("Function 2 (5%)" )+
  xlim(-4,4)+
  ylim(-4,4)+
  geom_text(data = gMeanssnp, inherit.aes=FALSE,
            aes(LDA1, LDA2,label = c("A. cervicornis", "F1 Hybrid","LG Hybrid","A. palmata"))) ,
```

```

    color = "black",
    fontface="bold")+
scale_fill_manual(values=c("#0000FF", "#9929B1", "#228B22", "#FF3E96", "#EEC900"),
    labels = c("A. cer", "F1 hybrid", "LG hybrid", "A. pal", "Unknown"))+
scale_shape_manual(values=c(25,22,23,24,21),
    labels = c("A. cer", "F1 hybrid", "LG hybrid", "A. pal", "Unknown"))+
guides(color = guide_legend(override.aes = list(size=4)))+
theme(legend.position=c("top"),
    legend.text = element_text(size = 10, color = "black"),
    legend.direction = "horizontal",
    axis.text=element_text(size = 12, color="black"),
    axis.title.x = element_text(vjust = 0, size=16),
    axis.title.y = element_text(vjust = 2, size=16),
    panel.background=element_rect(fill="white"),
    legend.key=element_rect(fill='white'),
    panel.border=element_rect(fill=NA),
    legend.title=element_blank(),
    plot.margin = unit(c(1, 1, 0.5, 0.5), "lines"),
    text = element_text(color = "black")) +
coord_equal()

```

s

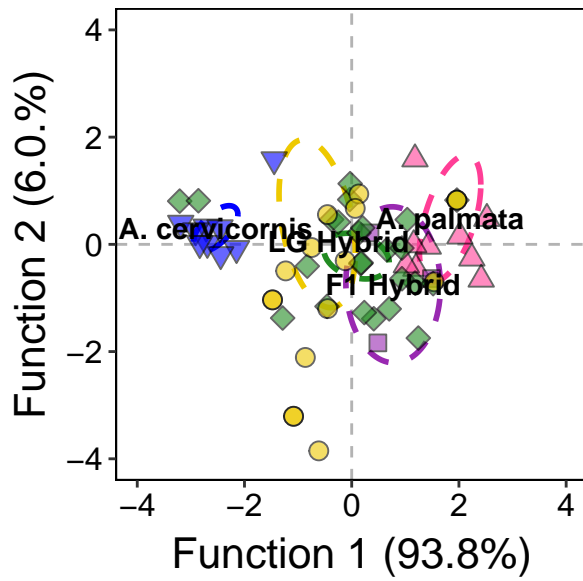


Final Figure 6

```
#putting it all together
library(ggpubr)
ggarrange(m,s,labels = c("A", "B"),ncol=2)
```

A

▼ A. cer
 ■ F1 hybrid
 ◆ LG hybrid
 ▲ A. pal
 ● Un



B

▼ A. cer
 ■ F1 hybrid
 ◆ LG hybrid
 ▲ A. pal
 ● Un

