

ion differentiation determined from putative  
adaptive SNP markers in *Eulachon*

*Evelyn Takyi*

May 8, 2017

## Required packages

treemap, readr, devtools, gstudio, pegas, ape, ade4, adegenet, diveRsity, hierfstat, poppr, readxl

## Importing data into R.

```
mon <- read.genepop("EulachonSig.gen")

##
## Converting data from a Genepop .gen file to a genind object...
##
##
## File description: Eulachon significant loci
## Warning in read.genepop("EulachonSig.gen"): Duplicate individual names
## detected. Coercing them to be unique.
##
## ...done.
```

## summary of imported data.

[illegible]

### Names of the 12 populations of samples used in analysis

```
popNames(mon)

## [1] "SS08"      "BELCOL03" "COW02"      "CR12"      "FRAS09"    "KC02"
## [7] "KEM01"     "KEN04"     "KLK02"     "SKE10"     "STIK06"    "TMR01"
```

## convert dataset into genind objects for each population

```
pop_label <- c(rep("SS08",41),rep("BELCOL03",33),
              rep("COW02",37),rep("CR12",22),rep("FRAS09",40),rep("KC02",36),
              rep("KEM01",42),rep("KEN04",71),rep("KLK02",41),rep("SKE10",33),rep("STIK06",66),rep("TMR01",37))

##. Creates a list of genind objects for each population
pops_separat <- seppop(mon,pop=pop_label)
pops_separat$COW02

## /// GENIND OBJECT ///////////
##
## // 37 individuals; 193 loci; 386 alleles; size: 155.5 Kb
##
## // Basic content
## @tab: 37 x 386 matrix of allele counts
## @loc.n.all: number of alleles per locus (range: 2-2)
## @loc.fac: locus factor for the 386 columns of @tab
## @all.names: list of allele names for each locus
## @ploidy: ploidy of each individual (range: 2-2)
## @type: codom
## @call: .local(x = x, i = i, j = j, treatOther = ..1, quiet = ..2, drop = drop)
##
## // Optional content
## @pop: population of each individual (group size range: 37-37)
```

## The mean observed and expected heterozygosity

```
##. Compute the mean observed and expected heterozygosity for each population over all loci
summaryCOW02 <- summary(pops_separat$COW02)
mean(summaryCOW02$Hobs)

## [1] 0.2239624
mean(summaryCOW02$Hexp)

## [1] 0.2246354
```

## compute allelic richness for each population

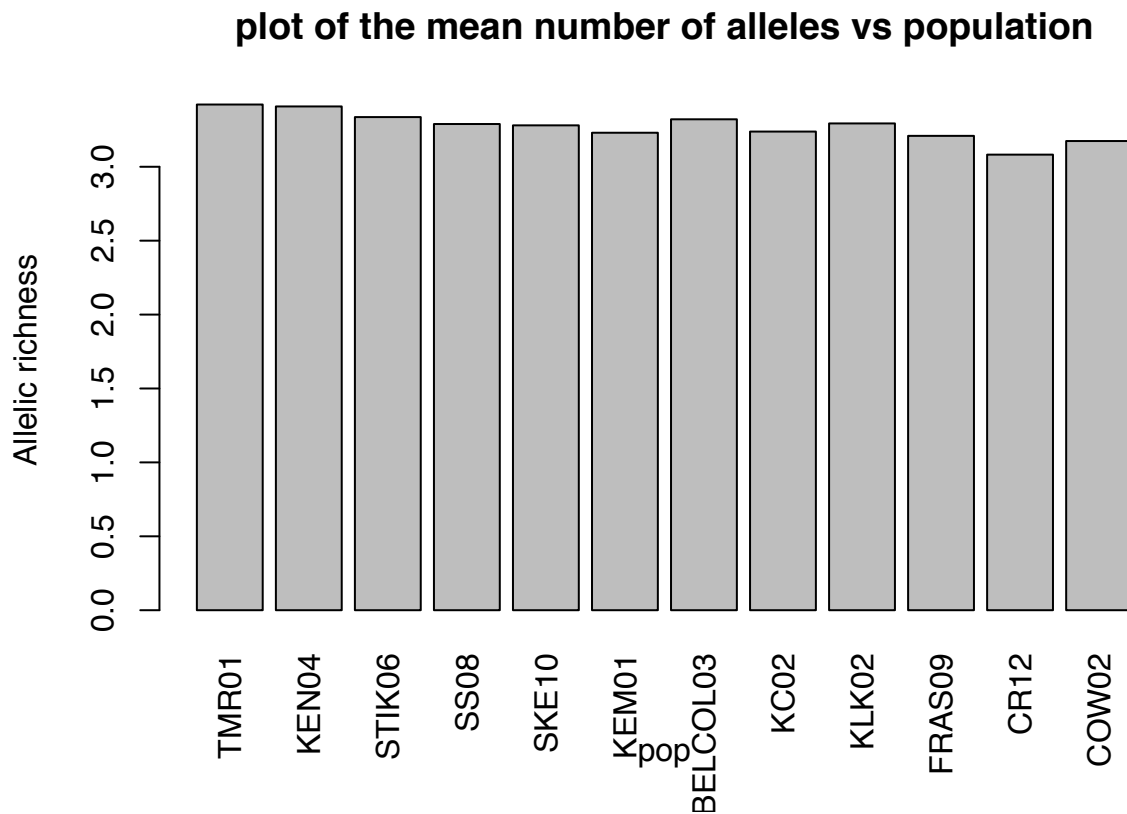
```
ba <- divBasic(infile = "EulachonSig.gen" , outfile = "by", gp = 3, bootstraps = 1000, HWEexact = FALSE, n = 1000)
Allerich <- summary(ba$Allelic_richness)

##. Mean allelic richness and mean heterozygosity for each population
heteroz <- read_excel("~/Downloads/Microsaetlilit/heteroz.xlsx")
heteroz

## # A tibble: 12 × 3
##       pop Heterozygosity Allelicrich
##   <chr>          <dbl>          <dbl>
## 1   TMR01          0.3969          3.421
```

```
## 2      KEN04      0.3921      3.408
## 3      STIK06     0.4058      3.336
## 4      SS08      0.4056      3.289
## 5      SKE10     0.4103      3.280
## 6      KEM01     0.4064      3.230
## 7      BELCOL03  0.4101      3.321
## 8      KC02      0.4132      3.238
## 9      KLK02     0.4078      3.293
## 10     FRAS09    0.4197      3.209
## 11     CR12      0.4251      3.082
## 12     COW02     0.4181      3.174
```

```
#. The mean number of alleles (Ar) were plotted to determine the genetic diversity in each population.
barplot(heteroz$Allelicrich, main="plot of the mean number of alleles vs population",ylab="Allelic richness")
```



## Test for genetic differentiation between pairs of population

```
#. Population structure was determined using the test for genetic differentiation between pairs of populations.
wc(mon)
```

```
## $FST
## [1] 0.05218927
##
## $FIS
## [1] 0.01277384
```

```
GF <- genet.dist(mon, method = "WC84")
GF
```

```
##          SS08      BELCOL03      COW02      CR12      FRAS09
## BELCOL03 0.010964926
## COW02    0.050284725 0.051959708
## CR12     0.076474711 0.084113123 0.031979660
## FRAS09   0.051107361 0.048257454 0.013656907 0.048652160
## KC02     0.020033081 0.009531729 0.049303419 0.089151531 0.040137025
## KEM01    0.003651866 0.011597369 0.053200231 0.081269978 0.048218065
## KEN04    0.080730771 0.078461326 0.083510072 0.105055990 0.095159708
## KLK02    0.012910011 0.007515463 0.049278960 0.080704024 0.043720618
## SKE10    0.012045033 0.006881454 0.051292325 0.084330945 0.048515859
## STIK06   0.011920935 0.008753050 0.047096889 0.081670619 0.052005081
## TMR01    0.095934381 0.091324233 0.096674264 0.123364057 0.103815008
##          KC02      KEM01      KEN04      KLK02      SKE10
## BELCOL03
## COW02
## CR12
## FRAS09
## KC02
## KEM01    0.014783901
## KEN04    0.091851764 0.074487167
## KLK02    0.004208213 0.008749633 0.086386533
## SKE10    0.015758806 0.008320082 0.076885941 0.011709553
## STIK06   0.018175434 0.009495093 0.074535505 0.010840276 0.005994568
## TMR01    0.105964773 0.091485898 0.005083938 0.101830392 0.091415594
##          STIK06
## BELCOL03
## COW02
## CR12
## FRAS09
## KC02
## KEM01
## KEN04
## KLK02
## SKE10
## STIK06
## TMR01    0.088565356
```

## Discriminant Analysis of Principal Components(DAPC)

```
#. using Kmeans and DAPC in adegenet
#. Region membership for each group of sample was determined by DAPC conducted in R with the package A

set.seed(20160308) # Setting a seed for a consistent result
grp <- find.clusters(mon, max.n.clust = 10, n.pca = 20, choose.n.clust = FALSE)

## The "ward" method has been renamed to "ward.D"; note new "ward.D2"

names(grp)

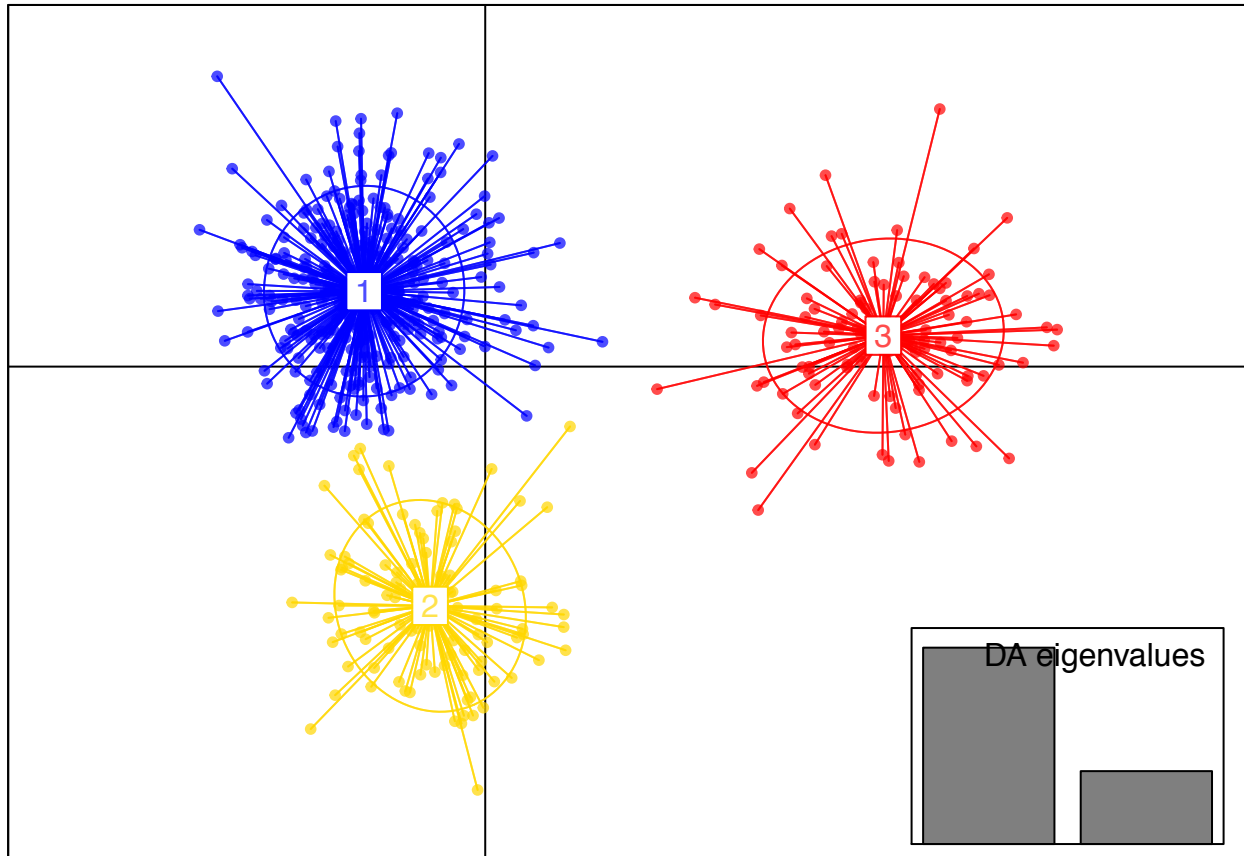
## [1] "Kstat" "stat" "grp" "size"
```

grp\$grp

```
## 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018
## 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1
## 019 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 037 038 039 040 041 042 043 044 045 046 047 048 049 050 051 052 053 054
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 055 056 057 058 059 060 061 062 063 064 065 066 067 068 069 070 071 072
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 073 074 075 076 077 078 079 080 081 082 083 084 085 086 087 088 089 090
## 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 091 092 093 094 095 096 097 098 099 100 101 102 103 104 105 106 107 108
## 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
## 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
## 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
## 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
## 2 2 2 2 2 2 1 2 2 2 2 1 1 1 1 1 1 1
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198
## 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
## 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3
## 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
## 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288
## 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3
## 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306
## 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324
## 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1
## 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
## 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
## 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
## 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
## 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468
## 1 1 1 1 1 1 1 1 1 1 1 1 3 3 3 3 3 3
```

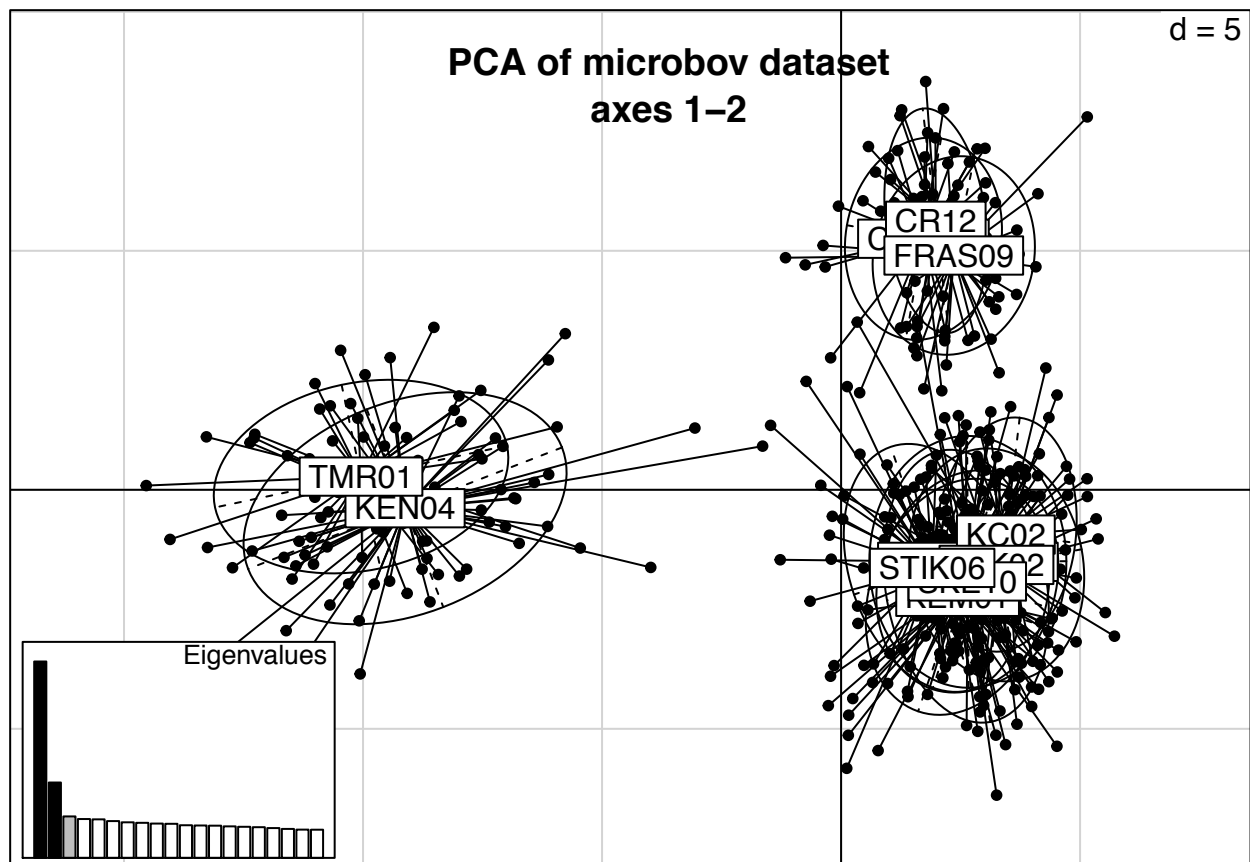
```
## 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486
##    3    3    3    3    3    3    3    3    3    3    3    3    3    3    3    3    3    3
## 487 488 489 490 491 492 493 494
##    3    3    3    3    3    3    3    3
## Levels: 1 2 3
```

```
dapc1 <- dapc(mon, grp$grp, n.pca = 20, n.da = 6)
scatter(dapc1) # plot of the group
```



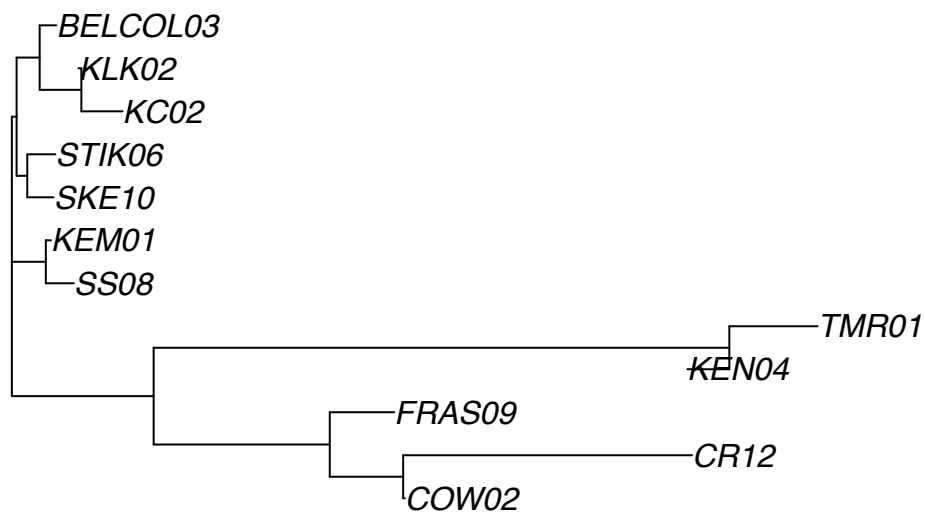
structure of the various groups with population sample names

```
#This plane is to show the structure of the various groups of population
s.class(pca1$li, pop(mon))
title("PCA of microbov dataset\naxes 1-2")
add.scatter.eig(pca1$eig[1:20], 3,1,2)
```



## Neighbourjoining Tree

```
#A neighbour-joining tree plotted to visualize the distances for each of these loci sets based on the di
tree <- nj(GF)
plot.phylo(tree)
```



## Reference

### publication

Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015) Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources* 15(6): 1421–1434. <http://dx.doi.org/10.1111/1755-0998.12400>

### Data obtained from Dryad

Candy JR, Campbell NR, Beacham TD, Grinnell MH, Narum SR, Larson WA (2015) Data from: Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.1797v>