KU LEUVEN

Evolutionary and Quantitative Genetics – I0D53A
Evaluation sheet – Assignment 1
Diversity, structure and selection of three-spined stickleback

Due date of the assignment: 19.11.2023 at 12 h (noon).

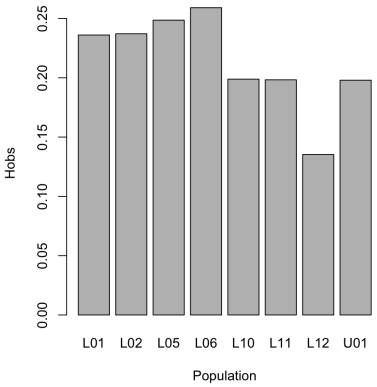
Your last and first name *Plewa Daria*
Your student number 0976669

Please upload your answer in pdf format under Toledo > Assignment on time.

Question 1: In your words, describe me the population genetic structure of the three-spined stickleback populations investigated in Belgium. What environmental factors and processes may have caused this population genetic structure? Support your findings with a figure. [Max. 2 pages]

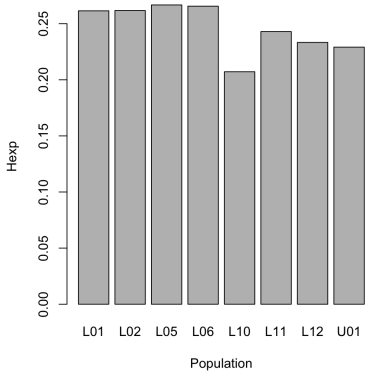
The samples L01, L02, L05, and L06 are coming from the lowland - brackish water, while the L10, L11, L12, and U01 are coming from inland water - freshwater. Expected genetic variation differs slightly from the observed. For the samples coming from the lowland terrains, the genetic variation is more similar to each other than expected. Furthermore, samples from the inland waters also contain more similar genetic variations in comparison to expected variations. The genetic variations from sample L12 is higher than expected one, while the L06 is lower than expected.

Expected genetic variation of the Spined stickleback



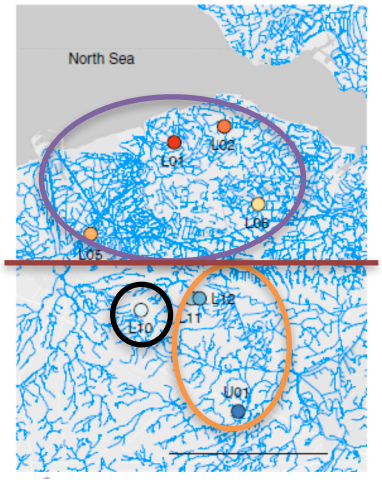
Source: R results

Observed genetic variation of the Spined stickleback



Source: R results

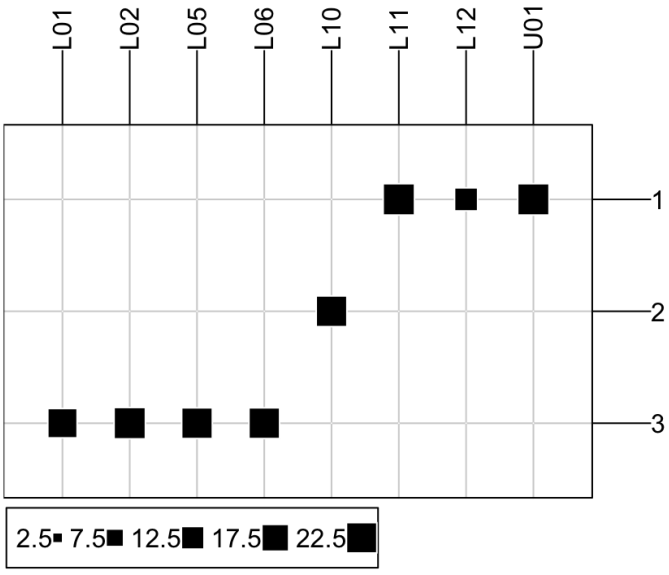
Placement of populations with rivers



Source: The paper

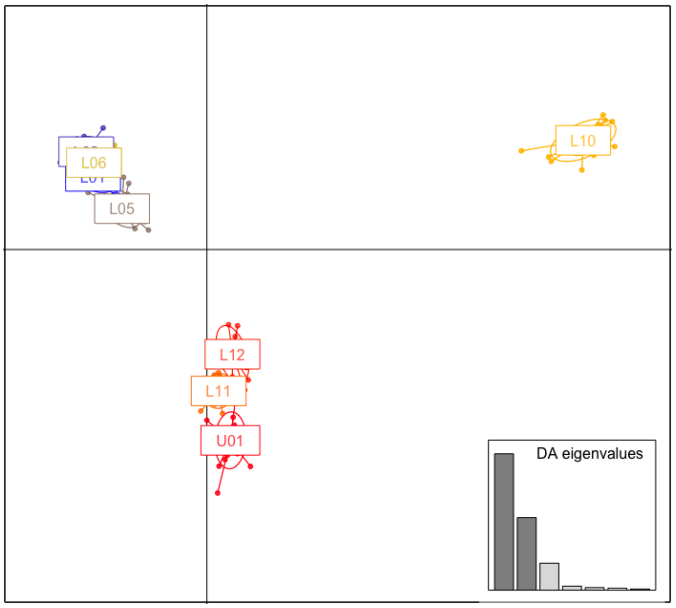
The place of occurrence of those populations causes that difference. Based on the charts we can tell that the closer to the coast, the bigger the migration, genetic flow, and genetic difference would be. Additionally, with a greater distance to the shore, the number of basins and rivers is falling, consequently, the populations up the stream would be more secluded in comparison to those closer to the shore, due to the lesser chance of meeting those populations. Above I also present the chart presenting the rivers in Belgium with the placement of selected populations, with the line representing the lowland and inland populations. Additionally, I clustered the samples regarding the result obtained from PCA analysis presented below.

Clustering results



Source: R results

PCA results

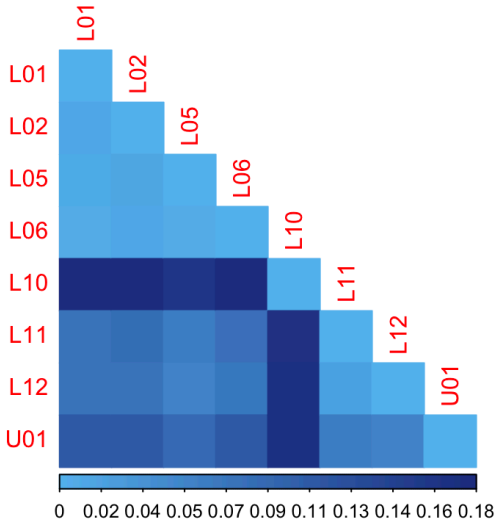


Source: R results

PCA presents additionally the populations that are the most similar to each other (lesser differentiation). The results presented on the PCA chart are consistent with the placement of the populations with the rivers and clustering

Heatmap of Fst presented on the left shows us and confirms the results obtained and presented earlier. From all the samples - samples L10 show the biggest difference in comparison to others. That particular differentiation could be caused by natural selection favoring the ecologically relevant traits, but not being as useful in other areas and populations. Less divergent is U01, the rest of them in comparison to those two are quite similar to each other.

Heatmap of Fst



Source: R results

Resluts of Heatmap

```
> pp
            L01         L02         L05         L06         L10         L11         L12         U01
L01 0.000000000 0.01111834 0.009351406 0.006440617 0.1774012 0.07780712 0.07322023 0.10811995
L02 0.011118344 0.00000000 0.013764927 0.010891155 0.1796818 0.08201762 0.07588466 0.10847427
L05 0.009351406 0.01376493 0.000000000 0.008882026 0.1621167 0.06236764 0.05694312 0.08636127
L06 0.006440617 0.01089115 0.008882026 0.000000000 0.1739582 0.07928417 0.07177411 0.10875132
L10 0.177401220 0.17968178 0.162116657 0.173958247 0.0000000 0.16394808 0.16498014 0.16524019
L11 0.077807116 0.08201762 0.062367637 0.079284174 0.1639481 0.00000000 0.02053868 0.06181800
L12 0.073220231 0.07588466 0.056943124 0.071774114 0.1649801 0.02053868 0.00000000 0.05520445
U01 0.108119953 0.10847427 0.086361268 0.108751320 0.1652402 0.06181800 0.05520445 0.00000000
```

Source: R results

Question 2: What is an AMOVA exactly? Explain with your own words. How do you interpret the hierarchical pattern for the stickleback data? [Max. 1 page]

AMOVA (Analysis of Molecular Variance) is the ANOVA (Analysis of Variance) specified for the nucleotides. ANOVA is a statistical method for the analysis of data with one or several simultaneously influencing factors. That method can explain how factors can influence the results in different groups and with what probability.

AMOVA table

```
> stickamova
$call
ade4::amova(samples = xtab, distances = xdist, structures = xstruct)

$results
                            Df    Sum Sq   Mean Sq
Between group                1  3975.506 3975.5062
Between samples Within group   6  6108.425 1018.0709
Within samples               114 42532.004  373.0878
Total                        121 52615.936  434.8424

$componentsofcovariance
                                        Sigma          %
Variations  Between group                48.25634  10.365256
Variations  Between samples Within group  44.21450   9.497086
Variations  Within samples               373.08776  80.137658
Total variations                        465.55860 100.000000

$statphi
                        Phi
Phi-samples-total 0.1986234
Phi-samples-group 0.1059532
Phi-group-total   0.1036526
```

Source: R

Monte Carlo permutation

```
> sticksignif
class: krandtest lightkrandtest
Monte-Carlo tests
Call: randtest.amova(xtest = stickamova, nrepet = 999)

Number of tests:   3

Adjustment method for multiple comparisons:   none
Permutation number:   999
                        Test       Obs    Std.Obs   Alter Pvalue
1  Variations within samples 373.08776 -50.203944    less  0.001
2 Variations between samples  44.21450  28.760887 greater  0.001
3   Variations between group  48.25634   4.366969 greater  0.001
```

Source: R

	Samples	Samples within groups	Groups
Statistical significance	Yes	Yes	Yes
Variance %	80.137658	9.497086	10.365256
Mean	373.0878	1018.0709	3975.5062
Phi	0.1986234	0.1059532	0.1036526

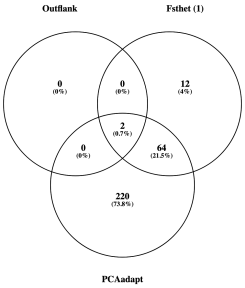
Variances within, and between samples are statistically significant, and the variance between groups is also significant. The variance between samples within groups is quite similar (9.5 and 10.36%), while the variance between all samples shows a great percentage difference (80%). Additionally, phi represents the level of differentiation statistics. The greater the result if it is, the bigger difference the results show. I contained all the results in the table above. We can see from the obtained results that the biggest difference represents the result from all samples, samples within groups and groups are more alike to each other.

Question 3: In your own words, describe me the outlier SNP pattern in the three-spined stickleback data and what has caused it (consider both the evolutionary drivers and genomic processes (for example, how are outlier SNPs distributed in the genome?).

In the analysis of outlier SNP were used 3 methods: FstHet, OutFLANK, and Pcadapt. Each of the methods uses different algorithms that detect the SNPs. To obtain the best results it is still encouraged to use all of those methods. FstHet was able to detect 78 significant SNPs (using betahat), OutFLANK 2, and PCAadapt 287 (using Benjamini-Hochberg procedure). 2 SNPs finally were detected by all 3 methods, and 64 by PCAadapt and Fsthet simultaneously. Algorithms used in OutFLANK could be the most strict in comparison to the other 2 methods, Fsthet having less strict algorithms and PCAadapt the least strict ones. Based on the results we obtained 2 SPNs: scaffold_99_316918 and scaffold_119_26420 that have been detected as outliers within all methods.

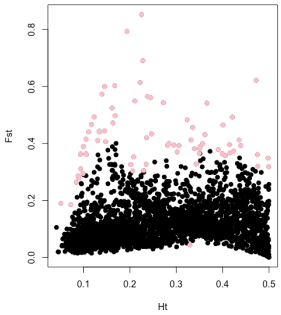
Particular SNPs can be favorable in certain ecotypes and populations. Those SNPs could emerge from different processes that occur in populations such as mutations, gene flow, natural selection, bottleneck, or genetic drift. The Manhattan plots below represent 2 main regions in which the most of obtained SNPs are located. Those 2 regions can be associated with traits that increase the chance of survival and reproduction. Additionally, those particular SPNs could be associated with each other causing linkage disequilibrium in inheritance.

Venn diaghram

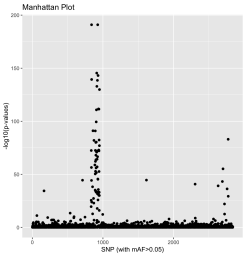


Source: Venny

Fsthet - betahat



PCAadapt



Source: R

OutFLANK