

INGDIF - SIMPLE MEASURE BASED ON ANALYSIS OF GENETIC DIFFERENTIATION

Naris Pojskić^{1*}

¹Laboratory for Bioinformatics and Biostatistics, University of Sarajevo - Institute for Genetic Engineering and Biotechnology, Bosnia and Herzegovina

*Correspondence

E-mail: naris.pojskic@ingeb.unsa.ba

Received

July, 2017

Accepted

November, 2017

Published

December, 2017

Copyright: ©2017 Genetics & Applications, The Official Publication of the Institute for Genetic Engineering and Biotechnology, University of Sarajevo

Notes and Comments

Population differentiation based on genetic diversity was subject of many previous scientific studies. Consequently, various methods were suggested. The most widely used method was fixation index F_{ST} , as a part of F_{IS} , F_{IT} and F_{ST} parameters which were proposed by Wright (1943, 1951, 1965). The main objective is to hierarchically estimate genetic variation in populations. Nei (1973, 1987) suggested G_{ST} as more appropriate methods, with θ (Cockerham 1969, 1973; Weir et Cockerham 1984), and Φ_{ST} (Excoffier et al. 1992) introduced later on as more adequate methods for molecular markers. Wright's F_{ST} has range between 0 and 1 where 0 indicates absence of differentiation, while 1 shows absolute divergence with no shared alleles. This method helps to quantify and compare level of genetic differentiation among populations. Since, in practice, when multialleles loci are applied, F_{ST} value of 1 is almost never observed for fixation indices (Wright 1978; Hedrick 1999; Jost 2008). This fact reduces application of fixation indices when highly polymorphic markers (e.g microsatellites) are used (Hedrick 1999). However, certain literature suggests that Nei's G_{ST} and Wier and Cockerham's θ are

flawed in the sense that 1 does not represent maximal differentiation. Arguing about practical applicability of standard genetic differentiation methods, Jost (2008) suggested allelic diversity (Δ) to be base for measuring the genetic differentiation D_{ST} as indicator of divergence (D). Jost considers that this approach corrects sampling bias, does not suffer the flaws of F-statistics and, being related to diversity, is more adequate. Nilsryman and Olofleimar (2009) concluded in their study that D_{ST} suffers the same problems as other measures, and that G_{ST} is still more appropriate method.

In order to achieve improved scale, standardization of fixation indices was proposed (Hedrick, 2005). The approach is to estimate G'_{ST} as scaling of the observed G_{ST} to its maximum attainable value. Similar to that, standardization of AMOVA (Excoffier et al. 1992) was suggested by Meirmans (2006). If we apply standardization of fixation indices, in case of absence of shared alleles, value of genetic differentiation will reach 1 regardless of genetic variation within population. In case of pairwise F_{ST} , there is a matrix of genetic differentiation between each pair of populations not showing values across loci.

We suggest simple measure based on the analyses of genetic differentiation. The purpose of this measurement is to estimate deviation of each locus F_{ST} (G_{ST} , θ) from overall value. This index of genetic differentiation (I) is represented as $(F_{STk} - F_{ST})/F_{ST}$, where F_{STk} is value for individual locus and F_{ST} for all loci. Index for each locus shows intensity and direction of deviation from overall value. Statistical probability is estimated by permutation (recommended 1000-10000). Thus, it is possible to make assessment of the relevance of impact of

```
#####
#          INGDIF - Index of Genetic Differentiation          #
#          Naris Pojskic (naris.pojskic@ingeb.unsa.ba)        #
#          University of Sarajevo                             #
#          Institute for Genetic Engineering and Biotechnology #
#          Laboratory for Bioinformatics and Biostatistics     #
#          Bioinfo Research Group - www.bioinfo.ba            #
#####

## This function calculates deviation Fst of loci from overall average across loci or haplotype diversity
## The input file should be named input.csv
## Results will be saved as output.csv

mydata <- read.csv("input.csv")
mydata
a <- mydata[,c(1)]
a
loc <- as.character(a)
b <- mydata[,c(2)]
b
av <- round(mean(b),3)
i <- round((b-av)/av,3)
s <- 1
n <- ncol(t(i))
xbar <- 0
z <- (xbar-i)/(s/sqrt(n))
z
pval <- 2*pnorm(-abs(z))
pv <- round(pval,4)
r <- sum(abs(i))
rgd <- round((abs(i)/r)*100,2)
rgd
tabela <- t(rbind(loc, b, t(i), pv, t(rgd)))
colnames(tabela) <- c("Loc", "G", "I", "p-value", "R%")
tabela
tloc <- c(loc)
plabe <- paste(tloc, rgd) # add percents to labels
barplot(i, horiz=TRUE, main="Index of Genetic Differentiation (I)", xlab="Index", names.arg = loc)
dev.new()
pie(rgd, labels=plabe, main="Pie Chart of Relative index (R%)")
```

Figure 1. Presentation of INDIF R script code

genetic differentiation at individual loci to the overall genetic differentiation. Additionally, relative genetic differentiation (R) is proportion of F_{ST} deviation of every locus within total deviation (all observed loci from overall F_{ST} value).

Table 1. An example of results of analysis of the index of genetic differentiation within INGDIF

Loc	G	I	p-value	R%
WGA69	0.111	0.682	0.031	11.63
WGA4	0.043	-0.348	0.2711	5.93
WGA1	0.091	0.379	0.2307	6.46
WGA89	0.089	0.348	0.2711	5.93
WGA9	0.156	1.364	0.0000	23.26
WGA118	0.002	-0.97	0.0022	16.54
WGA202	0.077	0.167	0.5974	2.85
WGA276	0.034	-0.485	0.1251	8.27
WGA376	0.051	-0.227	0.4729	3.87
WGA349	0.007	-0.894	0.0047	15.25

Loc - locus annotation; G - value of the F_{ST} ; I - the value of index with p-value; R% - relative deviation of individual loci against overall F_{ST}

It is expressed as $R = (I/\Sigma I_n) * 100$ where I is absolute value of individual locus while ΣI_n is the sum of deviation of all observed loci from overall F_{ST} value. Results have percentage type of values. The abovementioned calculations can be made using INGDIF R script (Figure 1.) at <http://lbb.ingeb.unsa.ba/INGDIF.html>.

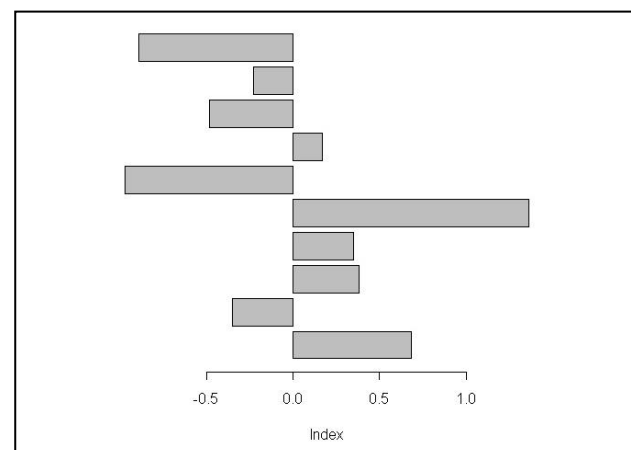


Figure 2. View of index of genetic differentiation

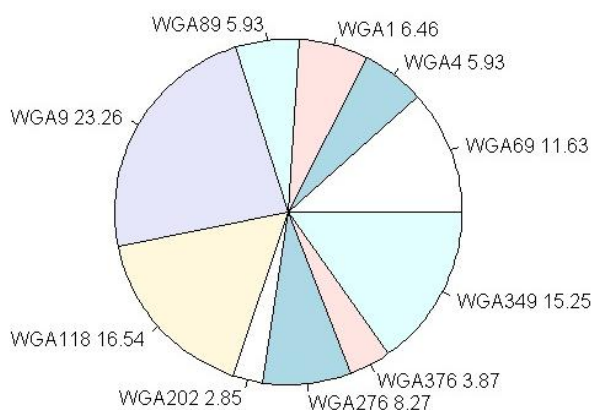


Figure 3. View of relative genetic differentiation

The result of the estimation of the genetic differentiation index contains the locus annotation, the value of the observed genetic differentiation, the value of index with p value, as well as relative deviation of individual loci against overall genetic differentiation (Table 1.)

The evaluation of the genetic differentiation index enables more clearly view of contribution of individual loci to overall genetic differentiation among populations in terms of the intensity and the level of reduction or increase (Figure 2,3).

References

- Cockerham CC (1969) Variance of gene frequencies. *Evolution*, 23:72-84.
- Cockerham CC (1973) Analyses of gene frequencies. *Genetics*, 74:679-700.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131:479-491.
- Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution*, 53:313-318.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, 59:1633-1638.
- Jost L (2008) GST and its relatives do not measure differentiation. *Mol Ecol*, 17:4015-4026.
- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, 60:2399-2402.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, 70:3321-3323.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ryman N, Leimar O (2009) GST is still a useful measure of genetic differentiation — a comment on Jost's D. *Molecular Ecology*, 18:2084-2087.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38:1358-1370.

Wright S (1943) Isolation by distance. *Genetics*, 28:114-138.

Wright S (1951) The genetical structure of populations. *Annu Eugen*, 15:323-354.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19:395-420.

Wright S (1978) *Evolution and the Genetics of Populations, Vol. 4: Variability Within and Among Natural Populations*. The University of Chicago Press, Chicago.