Pstat: an R Package to Assess Population Differentiation in Phenotypic Traits

by Stéphane Blondeau Da Silva and Anne Da Silva

Abstract The package Pstat for the statistical software R, created by the R Development Core Team, calculates P_{ST} values to assess differentiation among populations from a set of quantitative traits. It also provides confidence intervals and distributions of P_{ST} through the use of the bootstrap method. Variations of P_{ST} in function of the parameter c/h^2 are studied as well. Moreover, the package proposes different useful transformations of the measured phenotypic traits in particular to eliminate any variation resulting from allometric growth (calculation of residuals from linear regressions, Reist standardizations or Aitchison transformation).

Introduction

Understanding the causes governing patterns of morphological variations in the wild represents a fundamental goal of evolutionary biology. In particular, the relative importance of selective and neutral processes behind the observed differentiation remains a crucial question. In this perspective, studies comparing differentiation in quantitative traits and in neutral markers have significantly increased in number over the last ten years (Leinonen et al., 2013). Typically, a set of populations is sampled; the degree of genetic differentiation among populations estimated for a set of molecular markers with the Wright's F_{ST} index (Wright, 1951) is compared to the degree of phenotypic differentiation at a single or a set of quantitative traits, using the Q_{ST} (Spitze, 1993) index to assess the degree of phenotypic divergence. The logic of the F_{ST} and Q_{ST} comparisons relies on the assumption that F_{ST}, obtained by the consideration of neutral markers, reflects the divergence only induced by genetic drift (Reynolds et al., 1983). Hence, F_{ST} provides a null expectation and allows estimating the degree of population differentiation that would be reached without selection (Merilä and Crnokrak, 2001). As a consequence, the comparison between F_{ST} and Q_{ST} leads to three possibilities: (i) $Q_{ST} > F_{ST}$, means that quantitative traits show higher level of differentiation than what would have been expected under the single influence of genetic drift, which implies that natural selection could induce differentiation between populations by favoring different phenotypes (*i.e.* heterogeneous selection), (ii) $Q_{ST} < F_{ST}$, would be explained, as previously, by the influence of the natural selection, but in this case selecting for same optima among populations (*i.e.* homogeneous selection), (iii) $Q_{ST} = F_{ST}$, means that no departure from neutral expectations can be detected and then that the degree of differentiation in the quantitative traits could have been obtained only by genetic drift, even if contribution of natural selection cannot be excluded and estimated.

The study of Spitze (Spitze, 1993) introduced and defined the Q_{ST} quantity as follow (for diploid species and assuming purely additive gene action):

$$Q_{ST} = \frac{\sigma^2_{A_b}}{\sigma^2_{A_b} + 2\sigma^2_{A_w}},$$

where $\sigma^2_{A_b}$ and $\sigma^2_{A_w}$ are the morphological additive genetic variance components between and within populations. In the wild, the estimation of the additive genetic variance components is challenging as breeding design can hardly be conducted or are impossible. Therefore, Q_{ST} is often approximated by P_{ST} (Leinonen et al., 2006), directly calculated from the total phenotypic variance components (*i.e.* with no distinction between relative contributions of genetic and environmental variations):

$$P_{ST} = \frac{c\sigma^2_b}{c\sigma^2_b + 2h^2\sigma^2_w} \quad \text{or} \quad P_{ST} = \frac{\frac{c}{h^2}\sigma^2_b}{\frac{c}{h^2}\sigma^2_b + 2\sigma^2_w}, \tag{1}$$

where σ^2_b and σ^2_w are respectively the phenotypic variances between and within populations. c is an estimate of the proportion of the total variance due to additive genetic effects across populations and h^2, called heritability, is the proportion of phenotypic variance due to additive genetic effects (Brommer, 2011). Clearly, how well P_{ST} approximates Q_{ST}, depends on the parameters c and h^2 (*i.e.* if the values of c and h^2 are known then the phenotypic divergence quantified by P_{ST} would equal Q_{ST}) which implies to remain cautious in the interpretation obtained from P_{ST} (Brommer, 2011).

A large number of studies have assessed the potential for natural selection to impact morphological evolution by comparing phenotypic divergence with neutral genetic divergence *via* a P_{ST}-F_{ST} approach (*e.g.* Mobley et al. (2011), Lima (2012), He et al. (2013), and Shinn et al. (2015)). While estimations of F_{ST}

values are performed by different R packages as **diveRsity** (Keenan et al., 2013) or **hierfstat** (Goudet, 2005), no R package was created in order to deal with P_{ST} index. In this study, we present the **Pstat** package , which will give users the possibility to easily handle large datasets of quantitative traits and in particular to correct quantitative traits taking into account allometric growth. The package calculates P_{ST} values with their respective bootstrapped confidence intervals offering several options to select individuals, traits or populations that have to be considered. Various plotting tools are also provided for a visual evaluation of P_{ST} values, allowing in particular, users to compare P_{ST} with F_{ST} values. We will henceforth walk through a detailed example to give an overview of the **Pstat** package.

An example to get familiar with the main functions

First of all, the **Pstat** package must be dowloaded and installed from the CRAN repository. After having opened a R session and loaded the required package (`library(Pstat)`), we load the aggregate-level dataset (`data(test)`) into R. This dataset is a data frame containing 200 rows, each of them being an individual of a population of common wetland plants *Juncus effusus* (see Michalski and Durka (2015) and the Dryad Digital Repository (`https://doi.org/10.5061/dryad.bk5hk`)). The data frame contains in its first column the name of the populations (A, B, C, D, and E) to which the individual belongs and the eleven others contain quantitative measures (Table 1).

	Populations	QM1	QM2	QM3	QM4	...
1	A	0.18487253	0.4001979	0.1694021	42	...
2	B	0.24023500	0.4718000	0.2178500	46	...
3	C	0.23499676	0.4686213	0.2060222	25	...
4	B	0.20495223	0.3746026	0.1846816	51	...
5	C	0.20739220	0.4866461	0.2131618	19	...
6	C	0.22545341	0.3770903	0.1882165	28	...
7	C	0.18371681	0.4992361	0.2167194	25	...
...	...	...	...	...	...	...

Table 1: Part of the dataset `test`: the first seven individuals of populations of common wetland plants *Juncus effusus* (see Michalski and Durka (2015)) and the first four columns containing quantitative measures.

The data preparation

Before further running the calculations, the package allows the user to transform its dataset if necessary, in particular to eliminate any variation resulting from allometric growth. Users have the choice between three alternatives:

- determine the residuals of a simple linear regression, one of the quantitative variable being the regressor (Kuhry and Marcus, 1977);

- use an allometric transformation described in Reist (1985);

- use Aitchison's log-ratio transform (Aitchison, 1986).

Among a variety of univariate transformations that aim at separating size and shape variations, Reist showed that adjustments of size using a regression and residuals (the first case) and allometric adjustements to a standard size (the second case) are to be preferred: they allow the complete removal of size variations and have minimal impact on the correlation and covariance structure of the data (Reist, 1985). The third transformation offers the benefit of keeping the same number of variables contrary to the two firsts.

Simple linear adjustments

The first adjustment method provided by **Pstat** package is a simple linear regression. Assuming the existence of linear relationships between the dependant variables and the regressor (one of the eleven quantitative traits), the `Res()` function returns a new data frame (the column containing the quantitative trait used as a regressor being deleted). It contains the residuals of the regression (Kuhry and Marcus, 1977). The four function's arguments are as follows:

- ○ **data**: the studied data frame to be transformed with as many rows as individuals; the first column contains the name of the population to which the individual belongs, the others contain quantitative variables.

- ○ **reg**: the name (or the rank) of the variable chosen as the regressor.

- ○ **Rp**: the names of the populations to be deleted. Default value: $Rp = 0$ (no population removed).

- ○ **Ri**: the line numbers of individuals to be deleted. Default value: $Ri = 0$ (no individual removed).

```
Res(data=test, reg="QM3")
# Res() function running, the explanatory variable being QM3
```

In Table 2 the transformed data frame is presented.

	Populations	QM1	QM2	QM4	...
1	A	0.0339245264	5.621424e-03	6.23817063	...
2	B	0.1001268662	4.497085e-02	8.44522196	...
3	C	0.0922422473	4.966613e-02	-12.11705813	...
4	B	0.0574228940	-3.014565e-02	14.67271191	...
5	C	0.0662351079	6.293798e-02	-18.38127681	...
6	C	0.0787149904	-3.001126e-02	-8.45810788	...
7	C	0.0433557311	7.315960e-02	-12.51293868	...
...	...	...	...	...	...

Table 2: The linear adjusted data frame (with Res()), the explanatory variable being QM3 (only the first seven individuals and the first three corrected quantitative variables are presented in the table).

Reist transformation

In the second adjustment method provided by **Pstat** package, all morphometric measurements can be standardized according to Reist (see Reist (1985)), as was done by Kaeuffer et al. (2012) or He et al. (2013).

Let n be the number of individuals and p the number of quantitative traits. $\exists \, k \in \{1, ..., p\}$, such that the kth trait is the explanatory variable. Let us denote by $(x_i)_{1 \leq i \leq n}$ this variable and for $j \in \{1, ..., p\} \setminus \{k\}$, $(y_{ij})_{1 \leq i \leq n}$ the others. The formula is:

$$\forall i \in \{1, ... n\} \quad \text{and} \quad \forall j \in \{1, ..., p\} \setminus \{k\},$$
$$Y_{ij} = log(y_{ij}) - b_j(log(x_i) - log(\overline{x})),$$

where Y_{ij} is the size adjusted measurement of the jth trait (appart from the explanatory variable) for the ith individual, y_{ij} the original morphometric measurement for the same individual and the same trait, $\overline{x}$ the overall mean of the explanatory variable for all individuals, and x_i its value for the ith individual. $\forall j \in \{1, ..., p\} \setminus \{k\}$, the parameter b_j is estimated for the quantitative trait y_j (*i.e.* $(y_{ij})_{1 \leq i \leq n}$) from the observed data as the slope of the linear regression of $log(y_j)$ on $log(x)$, using all specimens.

The ReistTrans() function returns a corrected data frame (the column containing the quantitative trait used as a regressor being deleted). The function's arguments are the same as those of Res() function: **data**, **reg**, **Rp**, and **Ri**.

```
ReistTrans(test, reg=3)
# ReistTrans() function running, explanatory variable: the third one, namely QM3
```

In Table 3 the transformed data frame is presented.

Aitchison transformation

The third adjustment method provided by **Pstat** package performs a Aitchison log-ratio transformation of measures to account for individual size-effects (Aitchison, 1986). Such a transformation was made, for example, by Peres-Neto and Magnan (2004) or Shinn et al. (2015).

	Populations	QM1	QM2	QM4	...
1	A	-0.7445410	-0.3859875	1.631722	...
2	B	-0.6004703	-0.3462059	1.648348	...
3	C	-0.6167708	-0.3421063	1.388608	...
4	B	-0.6893556	-0.4255755	1.708186	...
5	C	-0.6669355	-0.3300087	1.266323	...
6	C	-0.6456670	-0.4250906	1.446049	...
7	C	-0.7175846	-0.3210021	1.384003	...
...	...	...	...	...	...

Table 3: The Reist adjusted data frame, the explanatory variable being QM3 (only the first seven individuals and the first three corrected quantitative variables are presented in the table).

Let n be the number of individuals and p the number of morphological traits. For $j \in \{1, ..., p\}$, let us denote by $(y_{ij})_{1 \leq i \leq n}$ the quantitative variables. The formula is:

$$\forall i \in \{1, ...n\} \quad \text{and} \quad \forall j \in \{1, ..., p\},$$

$$Y_{ij} = log(y_{ij}) - \frac{1}{p} \sum_{k=1}^{p} log(y_{ik}),$$

where Y_{ij} is the transformed measure of the jth trait for the ith individual and y_{ij} the original value for the ith individual and the jth trait.

The `AitTrans()` function returns a corrected data frame. It has a single argument: **data** (the studied data frame to be transformed with as many rows as individuals, the first column containing the name of the population to which the individual belongs, the others containing quantitative variables).

```
AitTrans(test)
# AitTrans() function running
```

In Table 4 the transformed data frame is presented.

	Populations	QM1	QM2	QM3	QM4	...
1	A	-1.947544	-1.6121417	-1.985498	0.408832854	...
2	B	-1.910214	-1.6170919	-1.952692	0.371908012	...
3	C	-1.834151	-1.5343910	-1.891299	0.192727037	...
4	B	-1.901481	-1.6395625	-1.946710	0.494436831	...
5	C	-1.832709	-1.4622885	-1.820792	0.129251912	...
6	C	-1.801889	-1.5785008	-1.880288	0.292211929	...
7	C	-1.938699	-1.5045418	-1.866950	0.195092172	...
...	...	...	...	...	...	...

Table 4: The Aitchison adjusted data frame (only the first seven individuals and the first four corrected quantitative variables are presented in the table).

Phenotypic differentiation evaluation and confidence intervals

Pst values

We are interested in determining the phenotypic differentiation across the five populations for each of the eleven quantitative traits of the example dataset. The function `Pst()` enables the user to determinate the P_{ST} values of each trait (Leinonen et al., 2006). It can also return the confidence intervals associated with the P_{ST} values. These confidence intervals are constructed following bootstrapping approach which relies on random sampling with replacement (Efron and Tibshirani, 1993). Indeed, single variable data frames are built from the studied data frame by sampling individuals from it, replacement allowing individuals to be selected more than once or even not selected. It should be noted that all these built data frames have the same individuals number as the original one. P_{ST} values of each of these data frames are calculated to determine the required confidence interval.

`Pst()` function's arguments are as follows:

- **data**: the studied data frame with as many rows as individuals; the first column contains the name of the population to which the individual belongs, the others contain quantitative variables.

- **ci**: if $ci = 1$ the confidence intervals are added to P_{ST} values. Default value: $ci = 0$.

- **csh**: the $\frac{c}{h^2}$ value. Default value: $\frac{c}{h^2} = 1$.

- **va**: a vector containing the selected variables names or numbers (*i.e.* those of the quantitative measures considered). If $va = 0$ all the variables are selected. Default value: $va = 0$.

- **boot**: the number of data frames generated to determine the confidence interval with the bootstrap method. Default value: $boot = 1000$.

- **Pw**: the names of the two populations considered to obtain pairwise P_{ST}. Default value: $Pw = 0$ (no pairwise analysis).

- **Rp**: the names of the populations to be deleted. Default value: $Rp = 0$ (no population removed).

- **Ri**: the line numbers of individuals to be deleted. Default value: $Ri = 0$ (no individual removed).

- **pe**: the confidence level of the calculated interval. Default value: $pe = 0.95$.

Let us apply the `Pst()` function to `test` dataset:

```
Pst(test, csh=0.2, Pw=c("C","D"))
# Only populations C and D are considered, we obtain pairwise Pst values

# Here is the output:
[1] "Populations sizes are:"
 C  D
76 32
   Quant_Varia Pst_Values
 1         QM1  0.1749659
 2         QM2  0.7460913
...        ...        ...
 4        QM10  0.9800028
```

Here is a second example:

```
Pst(test, va=c(2,"QM7"), ci=1, boot=10000, Ri=c(5,117:121), pe=0.99)
# Pst values of the two selected variables (the second one and QM7) and their 99%
# confidence intervals are here returned.

# The output is:
[1] "Populations sizes are:"
 A  B  C  D  E
12 76 72 30  4
  Quant_Varia Pst_Values 99 %_LowBoundCI 99 %_UpBoundCI
1         QM2  0.8561307       0.7826177      0.9198395
2         QM7  0.8851413       0.7722856      0.9376501
```

As can be seen in both examples, `Pst()` function returns a data frame.

Distribution of Pst

The multiple P_{ST} stemming from the bootstrap method enable the user to obtain a distribution histogram of P_{ST}. The function `BootPst()` returns the different values obtained, *via* the bootstrapping approach, of a single quantitative trait and/or constructs the repartition histogram of P_{ST} (of the considered trait). There is no need to focus on the arguments this fonction shares with the function `Pst()` (**data**, **csh**, **boot**, **Pw**, **Rp**, **Ri**, and **pe**), let us rather look at its specific arguments:

- **opt**: if $opt = 0$ all the boot values of P_{ST} are returned, if $opt = "ci"$ these ordered values and the confidence interval are returned, and if $opt = "hist"$ these ordered values and the distribution histogram of P_{ST} are returned. Default value: $opt = 0$.

- **va**: the name (or number) of the quantitative measure considered.

- **bars**: the maximum number of bars the histogram may have: indeed, on the x-axis, interval $[0, 1]$ is divided into *bars* parts (there may exist unfilled bars). Default value: $bars = 20$.

Let us apply the `BootPst()` function to `test` dataset:

```
BootPst(test, opt="ci", va="Body_length", Rp=c("A","E"))
# The ordered values and the associated 95% confidence interval are obtained thanks to
# the bootstrap method (only three populations are considered).

# The output is:
[1] "The studied quantitative variable is:"
[1] "Body_length"
[1] "Populations sizes are:"
 B  C  D
76 76 32
[1] "95 % confidence interval determined by 1000 bootstrap values:"
[1] 0.8757057 0.9585423
    [1] 0.7938426 0.8338286 0.8510682 0.8512374 0.8545911 0.8551115 0.8552097
    [8] 0.8637057 0.8641575 0.8644145 0.8659723 0.8671139 0.8671265 0.8676122
   [15] 0.8686147 0.8702277 0.8708352 0.8711419 0.8718030 0.8721783 0.8734932
   ...
  [995] 0.9621794 0.9625852 0.9634700 0.9644283 0.9650500 0.9689611
```
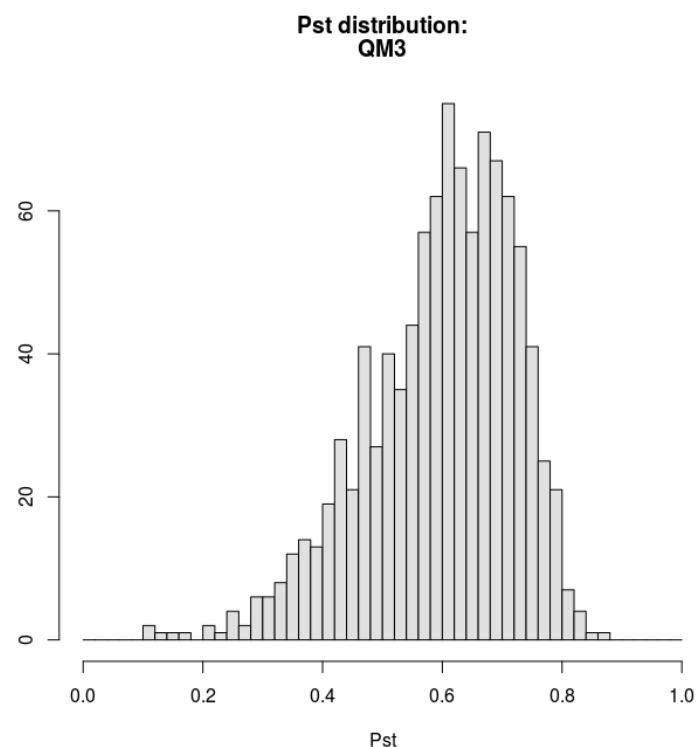
Here is a second example:

```
BootPst(test, opt="hist", va=3, bars=50)
# The ordered values and the distribution histogram are obtained thanks to the
# bootstrap method (see Figure 1).
```


Pst distribution: QM3

Figure 1: P_{ST} distribution histogram of QM3.

```
# The output is:
[1] "The studied quantitative variable is:"
[1] "QM3"
[1] "Populations sizes are:"
 A  B  C  D  E
12 76 76 32  4
[1] "1000 bootstrap values and Pst distribution:"
    [1] 0.1062747 0.1076470 0.1269888 0.1593121 0.1775196 0.2050347 0.2111617
    [8] 0.2327508 0.2401064 0.2487401 0.2588179 0.2589942 0.2623706 0.2722956
   [15] 0.2827915 0.2860497 0.2935858 0.2947525 0.2954878 0.2995198 0.3003267
```

```
...
[995] 0.8211326 0.8253874 0.8293417 0.8318546 0.8420100 0.8635299
```

We note that `BootPst()` function returns a vector containing P_{ST} bootstrap values.

## Variations of Pst values and visual comparison with Wright's Fst index

Plots illustrating how a comparison between neutral differentiation (F_{ST}, see Wright (1951)) and the values of P_{ST} (defined by Formula (1)) depends on the $\frac{c}{h^2}$ ratio, have been exploited by Brommer (2011) and Lima (2012) in their articles. The **Pstat** package provides the plotting tools to perform such an analysis with the function `TracePst`. It shares with the above function the following arguments: **data**, **boot**, **Pw**, **Rp**, **Ri**, and **pe**. The new arguments are:

- **va**: a vector containing the selected variables names or numbers (*i.e.* those of the quantitative measures considered). If $va = 0$ all the variables are selected. Default value: $va = 0$.
- **ci**: if $ci = 1$ the confidence interval of P_{ST} is plotted. Default value: $ci = 1$.
- **Fst**: the value of Wright's F_{ST}, if avalaible. Default value: $F_{ST} = -1$ (in this latter case the value of F_{ST} is unavailable).
- **xm**: the maximum on x-axis. Default value: $xm = 2$.
- **pts**: the number of points used to plot the curves. Default value: $pts = 30$.

Let us apply the `TracePst()` function to `test` dataset:

```
# Aitchison adjustment method:
trans_test=AitTrans(test)

# Plots illustrating how comparisons between Fst and Pst depends on c/h^2:
TracePst(trans_test, Fst=0.3, xm=3)
[1] "Populations sizes are:"
 A  B  C  D  E
12 76 76 32  4
```
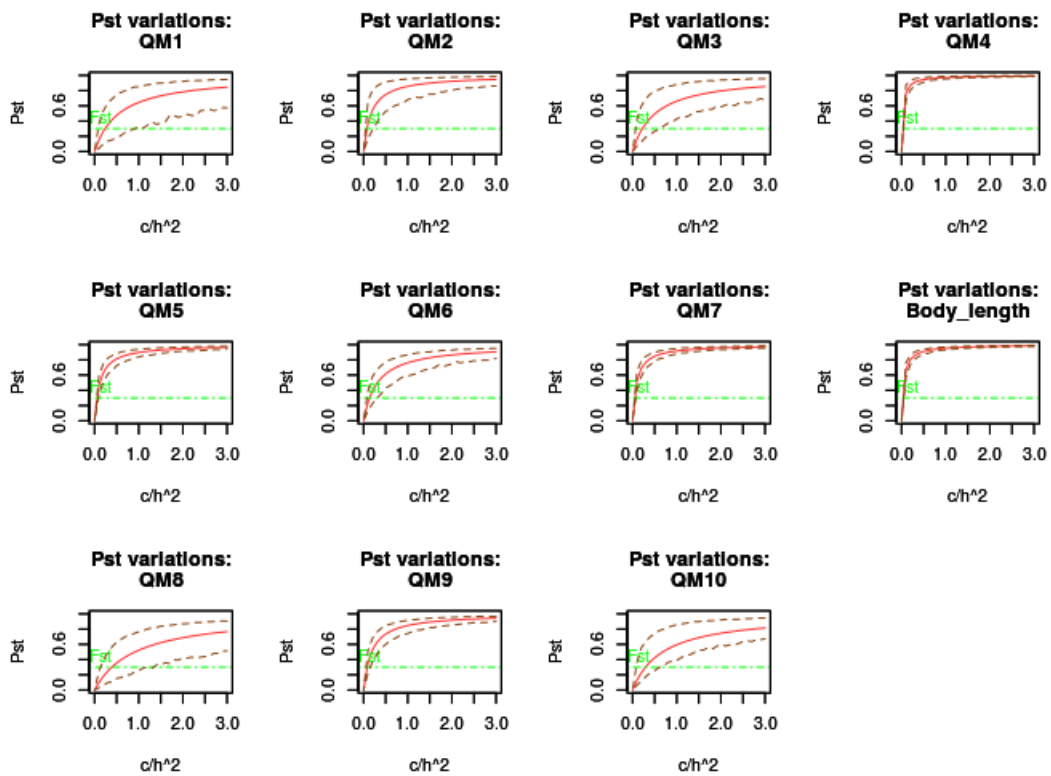

Figure 2: Plots illustrating how comparisons between F_{ST} and P_{ST} depends on $\frac{c}{h^2}$

The horizontal dotted green line (on Figure 2) marks the value of F_{ST}, whereas the P_{ST} values and their lower and upper 95% confidence intervals are plotted in red.

Conclusion

The use of P_{ST}-F_{ST} comparisons has increased rapidly in the last few years, showing the interest of this approach for assessing essential questions in the field of evolutionary and ecological genetics. The **Pstat** package logically completes existing packages allowing to estimate F-statistics and provides a relevant tool improving access to such approaches.

Bibliography

J. Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London, New York, 1986. 416 pp. [p2, 3]

J. E. Brommer. Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. *Journal of Evolutionary Biology*, 24:1160–1168, 2011. URL https://doi.org/10.1111/j.1420-9101.2011.02268.x. [p1, 7]

B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993. [p4]

J. Goudet. HIERSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Resources*, 5(1):184–186, 2005. URL https://doi.org/10.1111/j.1471-8286.2004.00828.x. [p2]

Y. He, R. Li, J. Wang, S. Blanchet, and S. Lek. Morphological variation among wild populations of chinese rare minnow (Gobiocypris rarus): Deciphering the role of evolutionary processes. *Zoological Science*, 30:475–483, 2013. URL https://doi.org/10.2108/zsj.30.475. [p1, 3]

R. Kaeuffer, C. L. Peichel, D. I. Bolnick, and A. P. Hendry. Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, 66(2):402–418, 2012. URL https://doi.org/10.1111/j.1558-5646.2011.01440.x. [p3]

K. Keenan, P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8):782–788, 2013. URL https://doi.org/10.1111/2041-210X.12067. [p2]

B. Kuhry and L. F. Marcus. Bivariate linear models in biometry. *Systematic Zoology*, 26(2):201–209, 1977. URL https://doi.org/10.2307/2412842. [p2]

T. Leinonen, J. M. Cano, H. Makinen, and J. Merilä. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*, 19:1803–1812, 2006. URL https://doi.org/10.1111/j.1420-9101.2006.01182.x. [p1, 4]

T. Leinonen, R. J. McCairns, R. B. O'Hara, and J. Merilä. Qst–Fst comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, 14:179–190, 2013. URL https://doi.org/10.1038/nrg3395. [p1]

M. R. Lima. Genetic and morphometric divergence of an invasive bird: The introduced house sparrow (Passer domesticus) in Brazil. *PloS One*, 7(12), 2012. URL https://doi.org/10.1371/journal.pone.0053332. [p1, 7]

J. Merilä and P. Crnokrak. Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, 14:892–903, 2001. URL https://doi.org/10.1046/j.1420-9101.2001.00348.x. [p1]

S. G. Michalski and W. Durka. Separation in flowering time contributes to the maintenance of sympatric cryptic plant lineages. *Ecology and Evolution*, 5(11):2172–2184, 2015. URL https://doi.org/10.1002/ece3.1481. [p2]

K. B. Mobley, D. Lussetti, F. Johansson, G. Englund, and F. Bokma. Morphological and genetic divergence in swedish postglacial stickleback (Pungitius pungitius) populations. *BMC Evolutionary Biology*, 2011. URL https://doi.org/10.1186/1471-2148-11-287. [p1]

P. R. Peres-Neto and P. Magnan. The influence of swimming demand on phenotypic plasticity and morphological integration: a comparison of two polymorphic charr species. *Oecologia*, 140:36–45, 2004. URL https://doi.org/10.1007/s00442-004-1562-y. [p3]

J. D. Reist. An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. *Canadian Journal Zoology*, 63:1429–1439, 1985. URL https://doi.org/10.1139/z85-213. [p2, 3]

J. Reynolds, B. S. Weir, and C. C. Cockerham. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, 105:767–779, 1983. [p1]

C. Shinn, S. Blanchet, G. Loot, S. Lek, and G. Grenouillet. Phenotypic variation as an indicator of pesticide stress in gudgeon: Accounting for confounding factors in the wild. *Science of the Total Environment*, 538:733–752, 2015. URL https://doi.org/10.1016/j.scitotenv.2015.08.081. [p1, 3]

K. Spitze. Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. *Genetics*, 135:367–374, 1993. [p1]

S. Wright. The genetical structure of populations. *Annals of Human Genetics*, 15:323–354, 1951. URL https://doi.org/10.1111/j.1469-1809.1949.tb02451.x. [p1, 7]

Stéphane Blondeau Da Silva, corresponding author
XLIM-Mathis, UMR n°7252 CNRS-Université de Limoges
123, avenue Albert Thomas 87060 Limoges
France
blondeaudasilva@xlim.fr

Anne Da Silva
INRA, UMR n°1061 Génétique Moléculaire Animale
Université de Limoges, UMR n°1061 Génétique Moléculaire Animale
123, avenue Albert Thomas 87060 Limoges
France
anne.blondeau@unilim.fr