Generalization of the $Q_{ST}$ framework in hierarchically structured populations: impacts of inbreeding and dominance.

Philippe Cubry$^{1,2}$, Ivan Scotti$^1$, Sylvie Oddou-Muratorio$^1$, François Lefèvre$^1$ *

$^1$INRA, UR 629 Ecologie des Forêts Méditerranéennes, URFM, Domaine Saint Paul, AgroParc, 84914 Avignon Cedex9, France

$^2$IRD, UMR DIADE, 393 Avenue Agropolis, 34394 Montpellier, Cedex 5, France

Keywords: hierarchical Q-statistics, hierarchical F-statistics, quantitative genetics

* Corresponding author. Fax: +33.4.32722902 E-mail: francois.lefevre.2@inra.fr

Running title: Generalized hierarchical $Q$-statistics

Abstract: 240 words

Total: 4570 words

1 table + 2 figures
Abstract

$Q_{ST}$ is a differentiation parameter based on the decomposition of the genetic variance of a trait. In the case of additive inheritance and absence of selection, it is analogous to the genic differentiation measured on individual loci, $F_{ST}$. Thus, $Q_{ST} - F_{ST}$ comparison is used to infer selection: selective divergence when $Q_{ST} > F_{ST}$, or convergence when $Q_{ST} < F_{ST}$. The definition of $Q$-statistics was extended to two-levels hierarchical population structures with Hardy-Weinberg equilibrium. Here, we generalize the $Q$-statistics framework to any hierarchical population structure. First, we developed the analytical definition of hierarchical $Q$-statistics for populations not at Hardy-Weinberg equilibrium. We show that the $Q$-statistics values obtained with the Hardy-Weinberg definition are lower than their corresponding $F$-statistics when $F_{IS} > 0$ (higher when $F_{IS} < 0$). Then, we used an island model simulation approach to investigate the impact of inbreeding and dominance on the $Q_{ST} - F_{ST}$ framework in a hierarchical population structure. We show that, while differentiation at the lower hierarchical level ($Q_{SR}$) is a monotonic function of migration, differentiation at the upper level ($Q_{RT}$) is not. In the case of additive inheritance, we show that inbreeding inflates the variance of $Q_{RT}$, which can increase the frequency of $Q_{RT} > F_{RT}$ cases. We also show that dominance drastically reduces $Q$-statistics below $F$-statistics for any level of the hierarchy. Therefore, high values of $Q$-statistics are good indicators of selection, but low values are not in the case of dominance.
Introduction

Quantitative measures of adaptive genetic differentiation are needed for the study of local adaptation processes. The $Q_{ST}$ statistic based on a partition of the genetic variance between and within populations is broadly used to quantify populations differentiation at quantitative traits in simple population structures. Its comparison to neutral marker divergence, measured via $F_{ST}$, provides a useful tool to understand the processes that lead to differentiation (see Leinonen et al. 2008 for review). In the case of neutrality, linkage equilibrium, Hardy-Weinberg equilibrium and additive inheritance, the mean value of differentiation at trait level, $Q_{ST}$, equals the mean value of differentiation at gene level, $F_{ST}$, and their variances have the same order of magnitude (Whitlock 2008). In other words, the distribution of single locus $F_{ST}$ values provides a reference to which $Q_{ST}$ values may be compared to infer the occurrence of selection leading to divergence (high values of $Q_{ST}$) or convergence (low values of $Q_{ST}$).

Leinonen et al. (2013) highlighted the usefulness of the $Q_{ST} – F_{ST}$ comparison approach for detecting adaptive evolution in various ecological contexts. In particular, this is expected to be a powerful approach to detect recent adaptation because $Q_{ST}$ is partly determined by covariances among QTL, which respond to selection more rapidly than individual allele frequencies (Latta 1998).

The $Q_{ST} – F_{ST}$ comparison framework was initially defined assuming strictly additive inheritance of the quantitative trait, Hardy-Weinberg equilibrium, absence of structured linkage disequilibrium (LD), and absence of selection. Violations of these assumptions have a variety of effects. Goudet & Büchi (2006) and Goudet & Martin (2007) showed that dominance generally reduces the mean value of $Q_{ST}$ below the mean value of $F_{ST}$ in the island.
model, although the exact outcome may depend on demographic history and migration rates among demes (see Whitlock 2008 for review). Whitlock (1999) and López-Fanjul et al. (2003) showed that additive-by-additive epistasis tends to have the same effect in general. Thus, with non-additive inheritance, the $Q_{ST} - F_{ST}$ approach is conservative in detecting selection leading to adaptive divergence, but liberal in detecting selection leading to adaptive convergence (Whitlock 2008). Relaxing the Hardy-Weinberg hypothesis, Bonnin et al. (1996) extended the definition of $Q_{ST}$ for inbred populations, aiming to keep the property of equal mean values of $Q_{ST}$ and $F_{ST}$ in that case. Regarding the assumption on LD, Whitlock (2008) argued that, in the absence of selection, random LDs for different sets of loci compensate one another and, therefore, have no effect on mean $Q_{ST}$ for neutral traits. Selection induces non-independence among QTL and makes expected $Q_{ST}$ deviate from single-locus $F_{ST}$. Gavrilets & Hastings (1995) proposed an analytical framework to investigate the impact of QTL covariance on $Q_{ST}$ at each level of the population structure. Kremer & Le Corre (2012) quantified the decoupling between $Q_{ST}$ and $F_{ST}$ under different evolutionary scenarios in a simple (non-hierarchical) population structure. Santure & Wang (2009) showed that dominance and inbreeding generally inflate the decoupling between $Q_{ST}$ and $F_{ST}$.

Besides one-level population structures, hierarchical structures attract increasing interest in ecology (Richardson et al. 2014). Whitlock & Gilbert (2012) stressed the interest of detecting selection across differing levels in spatial hierarchical population structures with varying amounts of gene flow. A hierarchical population structure may also result from temporal patterns of gene flow, e.g. across cohorts within populations. It also occurs in plant or animal breeding programs, e.g. sublining breeding schemes. More generally, understanding evolutionary processes in hierarchical population models helps interpret situations where
undetected within-population substructure may be suspected, e.g. in case of high within-
population environmental heterogeneity leading to microgeographic adaptation (Scotti et al.
2016).

Whitlock & Gilbert (2012) introduced a hierarchical extension of Q-statistics measuring
genetic differentiation at different spatial scales in a hierarchical island model or, more
generally, for each level of any hierarchical population structure. Using the hierarchical
derivations of F-statistics from Slatkin & Voelm (1991) and Yang (1998), it is then possible to
perform Q-statistics – F-statistics comparisons for each hierarchical level of the population
structure. However, these definitions of hierarchical Q-statistics do not account for local
departure from Hardy-Weinberg equilibrium, such as inbreeding, and the impact of
dominance in hierarchical population structure has not yet been investigated.

Here, we extend the framework of $Q_{ST} - F_{ST}$ comparison in hierarchical population structures
with no selection. We first summarize previous analytical developments that led to the
definition of hierarchical Q-statistics in Hardy-Weinberg populations and generalize these
derivations to account for inbreeding. Based on these developments, we analyze the effect of
inbreeding on Q-statistics and the consequences of ignoring it on $Q_{ST} - F_{ST}$ comparisons.

Next, we present the results of simulations in a hierarchical island model and analyze the
impacts of dominance on (i) the Q-statistics and (ii) the relation between Q-statistics and F-
statistics for each level of the hierarchy.

1) Definition of hierarchical Q-statistics for populations at Hardy-Weinberg equilibrium:

a synthesis
. **Simple population structure**

In the case of a single gene, which extends to the case of additive polygenic inheritance with no LD by summing over loci, Wright (1951) showed how overall departure from Hardy-Weinberg equilibrium in a mixture of populations having different allele frequencies affects genetic variance. In a diploid organism, the additive genetic variance in the total population ($\sigma^2_{IT}$) is related to the one that would be achieved in case of Hardy-Weinberg equilibrium ($\sigma^2_{T(0)}$) through the fixation index $F$:

$$\sigma^2_{IT} = (1 + F) \sigma^2_{T(0)} \quad \text{(Eq 1)}$$

Wright also demonstrated that the average within-population variance $\sigma^2_{IS}$ relates to $\sigma^2_{T(0)}$:

$$\sigma^2_{IS} = (1 - F) \sigma^2_{T(0)} \quad \text{(Eq 2)}.$$  

Therefore, due to the additivity of variance components, Wright deduced the variance among populations ($\sigma^2_{ST}$) from the difference between the total variance and the within-population variance:

$$\sigma^2_{ST} = (1 + F) \sigma^2_{T(0)} - (1 - F) \sigma^2_{T(0)} = 2F \sigma^2_{T(0)} \quad \text{(Eq 3)}.$$  

Considering here populations at Hardy-Weinberg equilibrium ($F_{IS} = 0$), the genic differentiation between populations is $F = F_{IT} = F_{ST}$. Lande (1992) resolved these equations for $F_{ST}$:
\[ F_{ST} = \frac{\sigma_{ST}^2}{\sigma_{ST}^2 + 2\sigma_{IS}^2} \]  
(Eq 4).

Spitze (1993) called \(Q_{ST}\) this estimator of genic differentiation based on the variance components. We will further call it \(Q_{ST(0)}\) to indicate that it refers to the particular case where \(F_{IS} = 0\).

### Extension to a hierarchical population structure

The hierarchical extension of \(F_{ST}\) has been described by several authors (Slatkin & Voelmer 1991; Yang 1998). In order to enable \(Q_{ST} - F_{ST}\) comparisons at any level of population structure, Whitlock & Gilbert (2012) proposed a hierarchical extension of \(Q_{ST(0)}\). To do so, they introduced a regional level (R) of structure, above the population level. The corresponding variance component, called \(V_R\) by the authors, is noted here \(\sigma_{RT}^2\) to follow Wright's notation.

Following Wright's partitioning of variances in Eq 3, Whitlock & Gilbert (2012) wrote that

\[ \sigma_{RT}^2 = 2F_{RT}\sigma_{T(0)}^2 \]  
(Eq 5).

Using Eq 2, they also gave the within-region component of variance as

\[ \sigma_{WR}^2 = \left| 1 - F_{RT} \right| \sigma_{T(0)}^2 \]  
(Eq 6).

For each region, between-population (\(\sigma_{SR}\), noted \(V_B\) in Whitlock & Gilbert 2012) and within-population (\(\sigma_{IS}\) or \(V_A\)) components of genetic variance can then be written as:
Expressing $\sigma^2_{SR}$ in terms of $\sigma^2_{IS}$ led to

$$\sigma^2_{SR} = 2 F_{SR} \frac{\sigma^2_{IS}}{1 - F_{SR}}$$  \hspace{1cm} (Eq 9),

Eq 9 provides a quantitative differentiation parameter for the lowest level of the hierarchy, $Q_{SR(0)}$, based on the variance components:

$$Q_{SR(0)} = F_{SR} = \frac{\sigma^2_{SR}}{\sigma^2_{SR} + 2 \sigma^2_{IS}}$$  \hspace{1cm} (Eq 10),

which is the same as Eq 4 for the lowest hierarchical level in the population structure.

For the differentiation at the highest level of the hierarchy, Whitlock & Gilbert (2012) expressed $\sigma^2_{RT}$ in terms of $\sigma^2_{SR}$:

$$\sigma^2_{RT} = 2 F_{RT} \sigma^2_{T(0)} = 2 F_{RT} \frac{\sigma^2_{SR}}{1 - F_{RT} \left(1 - 2 F_{SR}\right)}$$  \hspace{1cm} (Eq 11).

Resolving for $F_{RT}$ and using Eq 10 to replace $F_{SR}$ by its value, they defined $Q_{RT(0)}$ as follows:

$$Q_{RT(0)} = F_{RT} = \frac{\sigma^2_{RT}}{\sigma^2_{RT} + \sigma^2_{SR} + 2 \sigma^2_{IS}}$$  \hspace{1cm} (Eq 12).
Similarly, *Q-statistics* can be extended to any level of hierarchy (demonstration given in supporting information S1). If we note the hierarchical levels 1 to n, from lowest to highest (n=1 in a simple population structure), we get:

$$Q_{n(0)} = F_n = \frac{\sigma_n^2}{\sum_{i=1}^{n} \sigma_i^2 + 2\sigma_{IS}^2}$$

(Eq 13).

2) **Generalized definitions of Q-statistics when $F_{IS} \neq 0$**

. *Simple population structure*

Wright (1951) generalized Eq 1-3 to relax the assumption of Hardy-Weinberg equilibrium within sub-populations (positive or negative $F_{IS}$). In this case, the *F-statistics* are no longer the same ($F_{IS} \neq 0$ and $F_{IT} \neq F_{ST}$). Accordingly, Wright derived the components of genetic variance, noted $\sigma_{IS(0)}$, $\sigma_{ST(0)}^2$ and $\sigma_{IT(0)}^2$, as follows. Considering that departure from Hardy-Weinberg within populations does not affect the variance between populations, $\sigma_{ST(0)}^2$ remains the same as $\sigma^2_{ST}$: $\sigma^2_{ST(f)} = 2 F_{ST} \sigma^2_{T(0)}$. At the level of the whole population, we now have:

$$\sigma^2_{IT(f)} = (1 + F_{IT}) \sigma^2_{T(0)}$$

Wright then deduced the variance within populations from both previous equations: $\sigma^2_{IS(f)} = (1 + F_{IT} - 2 F_{ST} \sigma^2_{T(0)})$.

Replacing $F_{IT}$ by its expression in terms of $F_{ST}$ and $F_{IS}$ ($F_{IT} = F_{ST} + F_{IS} - F_{ST} F_{IS}$ ) led to

$$\sigma^2_{IS(f)} = (1 + F_{IS}) (1 - F_{ST}) \sigma^2_{T(0)}$$

(Eq 14).
Therefore, using Eq 3 and Eq 14, Bonnin et al. (1996) derived a generalized relation for $Q_{ST(0)}$ in case of inbred populations (the $f$ subscript indicates that any possible value of $F_{IS}$ is now taken into account):

$$Q_{ST(f)} = F_{ST} = \frac{(1 + F_{IS})\sigma_{ST|f}^2}{1 + F_{IS} + 2\sigma_{IS|f}^2}$$ (Eq 15).

For $F_{IS} = 0$, this general equation simplifies to the most commonly used $Q_{ST(0)}$.

**Extension to a hierarchical population structure**

Here, we follow Bonnin et al. (1996) and generalize the hierarchical $Q_{RT(0)}$ and $Q_{ST(0)}$ to the case of populations not at Hardy-Weinberg equilibrium. Following Wright’s demonstration, the variance between regions is not affected by population structure, thus Eq 5 still holds and

$$\sigma_{RT(0)}^2 = 2F_{RT}\sigma_T^2$$.

We now have to introduce two useful parameters: $\sigma_{WR(0)}$, which is the total genetic variance within regions if they were at Hardy-Weinberg equilibrium (i.e. analog to $\sigma_T^2$ but for the within-region additive genetic variance), and $F_{IR}$, which is the $F$-statistic corresponding to individual relative to region level, i.e. the global fixation index within region. $F_{IR}$ relates to $F_{SR}$ and $F_{IS}$ through the relation: $F_{IR} = F_{SR} + F_{IS} - F_{SR}F_{IS}$.

$\sigma_{WR(0)}$ is the one defined in Eq 6, whereas actual $\sigma_{WR|f}^2$ relates to $\sigma_{WR(0)}^2$ following Wright's equations:

$$\sigma_{WR|f}^2 = (1 + F_{IR})\sigma_{WR|0}^2 = (1 + F_{IR})(1 - F_{RT})\sigma_T^2$$ (Eq 16).

Following Eq 14 and Eq 6, $\sigma_{IS|f}^2$ can be written in terms of $F_{IS}$ and $F_{SR}$:
\[ \sigma_{IS,f}^2 = \left(1 + F_{IS}\right) \sigma_{W,f}^2 = \left(1 + F_{SR}\right) \sigma_{T,0}^2 \quad (Eq \ 17). \]

\[ \sigma_{SR,0}^2 \text{ can be written as: } \sigma_{SR,f}^2 = 2F_{SR}\sigma_{W,0}^2 = 2F_{SR}\left(1 - F_{RT}\right)\sigma_{T,0}^2 \quad (Eq \ 18). \]

Expressing \( \sigma_{SR,0}^2 \) in terms of \( \sigma_{IS,0}^2 \), we get:

\[ \sigma_{SR,f}^2 = 2F_{SR}\left(1 + F_{IS}\right)\sigma_{IS,f}^2 \quad (Eq \ 19). \]

Resolving this equation for \( F_{SR} \) gives the generalized definition of \( Q_{SR,f} \) when populations are not panmictic:

\[ Q_{SR,f} = F_{SR} = \frac{\left(1 + F_{IS}\right)\sigma_{SR,f}^2}{\left(1 + F_{IS}\right)\sigma_{SR,f}^2 + 2\sigma_{IS,f}^2} \quad (Eq \ 20), \] which is equivalent to Eq 15.

\( Q_{RT,f} \) is calculated accordingly, replacing \( F_{SR} \) by its value in Eq 11 for \( \sigma_{RT}^2 \):

\[ \sigma_{RT,f}^2 = 2F_{RT}\frac{\sigma_{SR,f}^2}{2\left(1 - F_{RT}\right)\left(1 + F_{IS}\right)\sigma_{SR,f}^2 + 2\sigma_{IS,f}^2} \quad (Eq \ 21). \]

Resolving this for \( F_{RT} \) gives the generalized definition of \( Q_{RT,f} \):

\[ Q_{RT,f} = F_{RT} = \frac{\left(1 + F_{IS}\right)\sigma_{RT,f}^2}{\left(1 + F_{IS}\right)\sigma_{RT,f}^2 + \left(1 + F_{IS}\right)\sigma_{SR,f}^2 + 2\sigma_{IS,f}^2} \quad (Eq \ 22). \]

This formula extends to any level of hierarchical structure as follows (demonstration given in supporting information S1):
Table 1 summarizes the generalized formulae of the $Q$-statistics that are unbiased analogs of $F_{ST}$ when $F_{IS} \neq 0$, for any number of levels in population structure.

In empirical studies, the specific definitions of $Q$-statistics that assume Hardy-Weinberg equilibrium ($Q_{ST(0)}$) are more often used than the generalized definitions ($Q_{ST(f)}$). In cases where $F_{IS} \neq 0$ for whatever reason, $Q_{ST(0)}$ is a biased analog of $F_{ST}$, which may affect the $Q_{ST} - F_{ST}$ comparison approach. To quantify this bias, replacing $\sigma_{IS}^2$ and $\sigma_{ST}^2$ by their values (Eq 3 and Eq 14, respectively) in Eq 15 leads to:

$$Q_{ST(0)} = \frac{F_{ST}}{1 + F_{IS} \left[ 1 - F_{ST} \right]}$$

which is also valid for $Q_{SR(0)}$.

Similarly, it results from Eq 22 that:

$$Q_{RT(0)} = \frac{F_{RT}}{1 + F_{IS} \left[ 1 - F_{SR} \left( 1 - F_{RT} \right) \right]}$$

Since $1 - F_{ST}$, $1 - F_{SR}$ and $1 - F_{RT}$ are all positive, these relations show that neglecting departure from Hardy-Weinberg equilibrium in the computation of $Q$-statistics provide lower values than the related $F$-statistics in case of inbreeding ($F_{IS} > 0$) or higher values in case of excess of heterozygotes ($F_{IS} < 0$). The difference is larger for intermediate values of differentiation. In a hierarchical population structure, strong differentiation at lower levels (e.g high $F_{SR}$ value) reduces the bias at upper levels ($Q_{RT(0)} - F_{RT}$).
3) Impact of dominance on the \( Q_{ST} - F_{ST} \) comparison in hierarchical populations

We used the quantiNEMO simulation platform (Neuenschwander et al. 2008) to simulate a hierarchical island model made of five regions and four populations within each region, i.e. 20 populations in total, and 100 individuals per population. We used three hierarchical structures where the ratio of within- to between-region migration was set to 6, 15 and 30, respectively.

In each hierarchical structure, our sets of parameter values were mostly similar to Goudet & Büchi (2006). We used four global migration rates: \( m=0.002, 0.01, 0.05 \) and 0.2. Selfing rate was 0 or 0.8, leading to \( F_{IS} \) values around 0.65 at the end of the simulations. The genetic architecture of the trait consisted in 100 QTL, 10 alleles per QTL with a k-allele mutation model parameter set to 0.001, and the variance of allelic effects was set to unity. The genotypic value of an individual was computed as:

\[
G = \sum_{i \in \text{loci}} a_i + a_{i'} + k_{ii'} |a_i - a_{i'}|
\]

where \( a_i \) and \( a_{i'} \) are the allelic effects and \( k_{ii'} \) is the dominance parameter. Inheritance was either purely additive (\( k_{ii'}=0 \) for all allelic pairs) or with dominance (\( k_{ii'} \) drawn from a normal distribution of mean 0 and variance 1, thus allowing over-dominance and under-dominance). Default parameters were used for the initial genotype frequencies, i.e. initial populations were maximally polymorphic in respect to allele frequencies. The hierarchical population structure was allowed to evolve 1000 generations. Each set of varying parameters (ratio of within- to between-region migration, global migration rate, selfing rate and dominance) was replicated 10 times, i.e. 480 simulations were run in total. Setting files are provided as supporting information S2.

At the end of each simulation, inbreeding \( (F_{IS}) \) and within-population additive variance \( (\sigma^2_{IS0}) \)
values were directly taken from quantiNEMO outputs. We used the matrix of genotypes of the last generation to estimate the hierarchical $F$-statistics ($F_{RT}$, $F_{SR}$) using hierfstat package in R version 3.3.1. We used the matrix of genotypic values to estimate the hierarchical components of genetic variance ($\sigma^2_{RT(f)}$, $\sigma^2_{SR(f)}$) from a random linear model and maximum likelihood estimator using lme4 package in R.

These simulations illustrate an interesting feature of differentiation in hierarchical population structures. We present here the results obtained with a ratio of within- to between-region migration set to 15 (more details are given in supporting information S3; similar analyses with the other values of this ratio, 6 and 30, are provided in supporting information S4). By definition, the differentiation at regional level, $Q_{RT}$, not only depends on the variance among regions ($\sigma^2_{RT(0)}$) but also on the variance between populations within regions ($\sigma^2_{SR(0)}$, see Eq. 22 above). Migration jointly influences both levels of the hierarchy and its effect on $\sigma^2_{RT(0)}$ may be counterbalanced by its effect on $\sigma^2_{SR(0)}$. Therefore, unlike $Q_{SR}$ (or $Q_{ST}$ in a simple population structure), $Q_{RT}$ is not a monotonic function of migration. This clearly appears in Figure 1 in the purely additive case: when migration is very low (here when $m=0.002$), global and within-region differentiation ($Q_{ST(global}$ and $Q_{SR}$, respectively) are maximized but differentiation at regional level ($Q_{RT}$) is reduced. In our simulations, differentiation at regional level was maximum when migration rate was $m=0.01$.

In the case of pure additivity, inbreeding marginally increases the mean values of hierarchical $Q$-statistics and increases more significantly the variance of $Q_{RT}$ (Figure 1). Dominance drastically reduces $Q$-statistics, globally and for each hierarchical level (Figure 1).
In simulations under pure additivity, the relation between $Q$-statistics and $F$-statistics followed the expectation that mean $Q_{SR}$ equals mean $F_{SR}$ as in a simple population structure (Figure 2). The relation was slightly different at regional level: in the absence of inbreeding, mean $Q_{RT}$ was slightly lower that mean $F_{RT}$. This discrepancy reflects the presence of covariance terms between alleles that would deserve further investigation. With inbreeding, the variance of $Q_{RT}$ increased, thus increasing the number of occurrences where $Q_{RT} > F_{RT}$ (Figure 2). With dominance, $Q$-statistics were always lower than $F$-statistics at the global level, which is consistent with the results obtained in a simple population structure by Goudet and Büchi (2006), and this remained true for each hierarchical level irrespective of the presence or absence of inbreeding (Figure 2).

The same qualitative patterns as in Figure 1 and Figure 2 were obtained with the other ratios of within- to between-region migration (see supporting information S4). The only noticeable quantitative difference was the non-linear effect of the migration rate on $Q_{RT}$, which co-varied with the ratio of within- to between-region migration, less pronounced with the lowest ratio.

Discussion

In this note, we proposed generalized definitions of $Q$-statistics for hierarchical population structures that keep the property of analogy to hierarchical $F$-statistics even when the lowest level of the population structure is not at Hardy-Weinberg equilibrium, in case of neutrality. This property is essential to infer selection from the $Q_{ST} - F_{ST}$ comparison, and these generalized definitions should be preferred to the simpler ones that assume Hardy-Weinberg equilibrium. In random mating populations, one generation of panmixia is theoretically enough to restore equilibrium after disturbance and, therefore, we expect little departure from
Hardy-Weinberg equilibrium in practice. However, inbred mating systems or unrevealed population substructure may generate positive \( F_{IS} \), while partial asexuality may generate negative \( F_{IS} \) (Stoeckel & Masson 2014). In this study, we quantified the discrepancy between \( Q\)-statistics and \( F\)-statistics induced by the use of \( Q\)-statistics definitions that do not account for an existing sub-structure in a population system, i.e. non-null \( F_{IS} \) values. This leads to under-estimate \( Q\)-statistics when \( F_{IS} \) is positive, and vice-versa. Therefore, when positive \( F_{IS} \) is neglected in the computation of \( Q\)-statistics, the inference of adaptive divergence when \( Q\)-statistics are significantly higher than \( F\)-statistics is conservative but, reversely, inference of adaptive convergence cannot be concluded when \( Q\)-statistics are lower than \( F\)-statistics. The reverse effect occurs with negative \( F_{IS} \) values. Interestingly, in a hierarchical population structure, high differentiation at the lowest hierarchical level reduces the bias due to neglecting \( F_{IS} \) on the upper level \( Q\)-statistics.

The analytical framework leading to the definition of generalized \( Q\)-statistics relies on the assumptions of additivity and independence among neutral loci. Here, in a hierarchical island model, we show that dominance drastically reduces hierarchical \( Q\)-statistics \( (Q_{RT}, Q_{SR}) \) below their related \( F\)-statistics \( (F_{RT}, F_{SR}) \), as initially shown by Goudet & Büchi (2006) and Goudet & Martin (2007) in a simple island model. Therefore, dominance makes the \( Q_{ST} - F_{ST} \) comparison a conservative approach when inferring selective divergence from high \( Q_{ST} \) values but no conclusion can be drawn from low \( Q_{ST} \) values. For empirical studies, it should be noticed that, even in the absence of selection, historical and demographic processes can generate random statistical dependencies among loci that are captured in the \( Q\)-statistics but not in the \( F\)-statistics. These covariances contribute to the high variance of \( Q\)-statistics. In our simulated hierarchical island model, discrepancies between \( Q\)-statistics and \( F\)-statistics were
more likely at the upper level of the hierarchy ($Q_{RT} - F_{RT}$), and the variance of $Q_{RT}$ was particularly sensitive to inbreeding in the case of pure additivity. Therefore, the power to detect selection from the $Q_{ST} - F_{ST}$ approach is lower between regions than within regions.

When there is dominance, the mean and the variance of $Q$-statistics are both reduced. To account for these neutral sources of covariance among loci, $Q$-statistics should rather be compared to measures of neutral differentiation patterns that account for these disequilibria, such as the $CF_{ST}$ proposed by Kremer et al. (1997). Statistical developments (Martin et al, 2008) and the recent availability of genome-wide resources have provided some clues for the analysis of polygenic adaptation (Berg & Coop 2014; Yeaman 2015; Stephan 2016).

In a hierarchical population structure, we showed a specific effect of interaction between migration and genetic drift on the differentiation among “regions” (or any other kind of upper-level of the hierarchy, not necessarily due to spatial structure) that has interesting consequences for the conservation of the total genetic diversity. The migration rate has a non-monotonic effect on the differentiation among “regions”. Very low migration rates induce high differentiation among “populations” within-“regions”, which is also related to high gene diversity at “regional” level (data not shown). In this case, “populations” become quasi-independent samples submitted to genetic drift and each “region”, having several of these independent “populations”, captures a good level of gene diversity, thus reducing the divergence among “regions”. In other words, in a hierarchical population structure, the impact of genetic drift on gene diversity at “regional” level can be reduced not only by high migration rates but also by very low migration rates, and this effect is increased when the ratio of within- to between-“region” migration is high.
Acknowledgement

This work was supported by Agence Nationale de la Recherche (ANR) under the FLAG project ANR-12-ADAP-0007-01. We are grateful to Antoine Kremer, Thomas Boivin, Catherine Bastien, Leopoldo Sanchez, Vincent Segura and anonymous reviewers for their helpful comments.

References


Goudet J, Martin G (2007) Under neutrality, $Q_{ST} \leq F_{ST}$ when there is dominance in an island model. *Genetics*, **176**, 1371-1374.


Data accessibility

Detailed quantiNEMO parameter sets and simulation outputs are provided as supporting information.

Author contributions

PC, IS, SOM and FL designed the research and wrote the paper. PC and FL made the analytical developments and the simulations.

Tables and Figures

Table 1: Definition of generalized \textit{Q-statistics} taking into account inbreeding in hierarchically structured populations. See the main text for the definitions of the parameters.

Figure 1: \textit{Q-statistics} in a hierarchical island model with five regions and four populations per region. The ratio of within- to between-region migration was set to 15 and four migration rates were used, with or without inbreeding, with or without dominance. Differentiation parameters are computed globally ($Q_{\text{ST,global}}$, among the 20 populations), among regions ($Q_{\text{RT}}$) and among populations within regions ($Q_{\text{SR}}$).

Figure 2: Relation between \textit{Q-statistics} and \textit{F-statistics} in the same hierarchical island model as Figure 1. Symbols reflect the migration rates (o: $m=0.002$; $\Delta$: $m=0.01$; +: $m=0.05$; x: $m=0.2$).

Supporting information

Supporting information S1: Demonstration of the generalized equations Eq 13 and Eq 23

Supporting information S2: Settings file used to elaborate 480 neutral evolutionary scenarios in a two-levels hierarchical population structure with the quantiNEMO simulation platform.

Supporting information S3: Simulation outputs, diversity and differentiation parameters for each simulation run with a ratio of within- to between-region set to 15.

Supporting S4: Additional results equivalent to Figure 1 and Figure 2 obtained with two other ratios of within- to between-region migration, respectively set to 6 and 30.
Table 1

<table>
<thead>
<tr>
<th>Population structure</th>
<th>Definition of generalized hierarchical Q-statistics</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-level structure</td>
<td>( Q_{ST/f} = \frac{1 + F_{IS}}{1 + F_{IS}} \frac{\sigma^2_{ST/f}}{\sigma^2_{ST/f} + 2 \sigma^2_{IS/f}} ) (Eq 15)</td>
<td>(1)</td>
</tr>
<tr>
<td>2-levels hierarchy</td>
<td>( Q_{SR/f} = \frac{1 + F_{IS}}{1 + F_{IS}} \frac{\sigma^2_{SR/f}}{\sigma^2_{SR/f} + 2 \sigma^2_{IS/f}} ) (Eq 20)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>( Q_{RT/f} = \frac{1 + F_{IS}}{1 + F_{IS}} \frac{\sigma^2_{RT/f}}{\sigma^2_{RT/f} + 1 + F_{IS} \sigma^2_{SR/f} + 2 \sigma^2_{IS/f}} ) (Eq 22)</td>
<td>(2)</td>
</tr>
<tr>
<td>n-levels hierarchy</td>
<td>( Q_{n(f)} = F_n = \frac{1 + F_{IS}}{\sum_{i=1}^{n} \sigma^2_i + 2 \sigma^2_{IS}} ) (Eq 23)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

References: (1) Bonnin et al, 1996; (2) this paper (see text for other references that funded these derivations)