



HAL
open science

Generalization of the Q(ST) framework in hierarchically structured populations: Impacts of inbreeding and dominance

Philippe Cubry, Ivan Scotti, Sylvie Muratorio, Francois Lefèvre

► **To cite this version:**

Philippe Cubry, Ivan Scotti, Sylvie Muratorio, Francois Lefèvre. Generalization of the Q(ST) framework in hierarchically structured populations: Impacts of inbreeding and dominance. *Molecular Ecology Resources*, 2017, 17 (6), pp.e76-e83. 10.1111/1755-0998.12693 . hal-01576329

HAL Id: hal-01576329

<https://hal.science/hal-01576329>

Submitted on 25 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

3 Philippe Cubry^{1,2}, Ivan Scotti¹, Sylvie Oddou-Muratorio¹, François Lefèvre¹ *

4
5 ¹INRA, UR 629 Ecologie des Forêts Méditerranéennes, URFM, Domaine Saint Paul,
6 AgroParc, 84914 Avignon Cedex9, France

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

8
9 *Keywords: hierarchical Q-statistics, hierarchical F-statistics, quantitative genetics*

10

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

12

13 *Running title: Generalized hierarchical Q-statistics*

14

15 *Abstract: 240 words*

16 *Total: 4570 words*

17 *1 table + 2 figures*

18 **Abstract**

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

36

37

38 Introduction

39 Quantitative measures of adaptive genetic differentiation are needed for the study of local
40 adaptation processes. The Q_{ST} statistic based on a partition of the genetic variance between
41 and within populations is broadly used to quantify populations differentiation at quantitative
42 traits in simple population structures. Its comparison to neutral marker divergence, measured
43 via F_{ST} , provides a useful tool to understand the processes that lead to differentiation (see
44 Leinonen *et al.* 2008 for review). In the case of neutrality, linkage equilibrium, Hardy-
45 Weinberg equilibrium and additive inheritance, the mean value of differentiation at trait level,
46 Q_{ST} , equals the mean value of differentiation at gene level, F_{ST} , and their variances have the
47 same order of magnitude (Whitlock 2008). In other words, the distribution of single locus F_{ST}
48 values provides a reference to which Q_{ST} values may be compared to infer the occurrence of
49 selection leading to divergence (high values of Q_{ST}) or convergence (low values of Q_{ST}).
50 Leinonen *et al.* (2013) highlighted the usefulness of the $Q_{ST} - F_{ST}$ comparison approach for
51 detecting adaptive evolution in various ecological contexts. In particular, this is expected to be
52 a powerful approach to detect recent adaptation because Q_{ST} is partly determined by
53 covariances among QTL, which respond to selection more rapidly than individual allele
54 frequencies (Latta 1998).

55 The $Q_{ST} - F_{ST}$ comparison framework was initially defined assuming strictly additive
56 inheritance of the quantitative trait, Hardy-Weinberg equilibrium, absence of structured
57 linkage disequilibrium (LD), and absence of selection. Violations of these assumptions have a
58 variety of effects. Goudet & Büchi (2006) and Goudet & Martin (2007) showed that
59 dominance generally reduces the mean value of Q_{ST} below the mean value of F_{ST} in the island

60 model, although the exact outcome may depend on demographic history and migration rates
61 among demes (see Whitlock 2008 for review). Whitlock (1999) and López-Fanjul *et al.*
62 (2003) showed that additive-by-additive epistasis tends to have the same effect in general.
63 Thus, with non-additive inheritance, the $Q_{ST} - F_{ST}$ approach is conservative in detecting
64 selection leading to adaptive divergence, but liberal in detecting selection leading to adaptive
65 convergence (Whitlock 2008). Relaxing the Hardy-Weinberg hypothesis, Bonnin *et al.* (1996)
66 extended the definition of Q_{ST} for inbred populations, aiming to keep the property of equal
67 mean values of Q_{ST} and F_{ST} in that case. Regarding the assumption on LD, Whitlock (2008)
68 argued that, in the absence of selection, random LDs for different sets of loci compensate one
69 another and, therefore, have no effect on mean Q_{ST} for neutral traits. Selection induces non-
70 independence among QTL and makes expected Q_{ST} deviate from single-locus F_{ST} . Gavrilets &
71 Hasting (1995) proposed an analytical framework to investigate the impact of QTL covariance
72 on Q_{ST} at each level of the population structure. Kremer & Le Corre (2012) quantified the
73 decoupling between Q_{ST} and F_{ST} under different evolutionary scenarios in a simple (non
74 hierarchical) population structure. Santure & Wang (2009) showed that dominance and
75 inbreeding generally inflate the decoupling between Q_{ST} and F_{ST} .

76 Besides one-level population structures, hierarchical structures attract increasing interest in
77 ecology (Richardson *et al.* 2014). Whitlock & Gilbert (2012) stressed the interest of detecting
78 selection across differing levels in spatial hierarchical population structures with varying
79 amounts of gene flow. A hierarchical population structure may also result from temporal
80 patterns of gene flow, e.g. across cohorts within populations. It also occurs in plant or animal
81 breeding programs, e.g. sublining breeding schemes. More generally, understanding
82 evolutionary processes in hierarchical population models helps interpret situations where

83 undetected within-population substructure may be suspected, e.g. in case of high within-
84 population environmental heterogeneity leading to microgeographic adaptation (Scotti *et al.*
85 2016).

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

94 Here, we extend the framework of $Q_{ST} - F_{ST}$ comparison in hierarchical population structures
95 with no selection. We first summarize previous analytical developments that led to the
96 definition of hierarchical *Q-statistics* in Hardy-Weinberg populations and generalize these
97 derivations to account for inbreeding. Based on these developments, we analyze the effect of
98 inbreeding on *Q-statistics* and the consequences of ignoring it on $Q_{ST} - F_{ST}$ comparisons.
99 Next, we present the results of simulations in a hierarchical island model and analyze the
100 impacts of dominance on (i) the *Q-statistics* and (ii) the relation between *Q-statistics* and *F-*
101 *statistics* for each level of the hierarchy.

102

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

104 **a synthesis**

105 . Simple population structure

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

112
$$\sigma_{IT}^2 = (1 + F) \sigma_{T(0)}^2 \quad (\text{Eq 1})$$

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

114
$$\sigma_{IS}^2 = (1 - F) \sigma_{T(0)}^2 \quad (\text{Eq 2}).$$

115 Therefore, due to the additivity of variance components, Wright deduced the variance among
116 populations (σ_{ST}^2) from the difference between the total variance and the within-population
117 variance:

118
$$\sigma_{ST}^2 = (1 + F) \sigma_{T(0)}^2 - (1 - F) \sigma_{T(0)}^2 = 2F \sigma_{T(0)}^2 \quad (\text{Eq 3}).$$

119 Considering here populations at Hardy-Weinberg equilibrium ($F_{IS} = 0$), the genic
120 differentiation between populations is $F = F_{IT} = F_{ST}$. Lande (1992) resolved these equations
121 for F_{ST} :

$$F_{ST} = \frac{\sigma_{ST}^2}{\sigma_{ST}^2 + 2\sigma_{IS}^2} \quad (\text{Eq 4}).$$

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 & Voelm 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 σ_{RT}^2 to follow Wright's notation.

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

$$\sigma_{RT}^2 = 2 F_{RT} \sigma_{T(0)}^2 \quad (\text{Eq 5}).$$

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

$$\sigma_{WR}^2 = (1 - F_{RT}) \sigma_{T(0)}^2 \quad (\text{Eq 6}).$$

For each region, between-population (σ_{SR}^2 , noted V_B in Whitlock & Gilbert 2012) and within-population (σ_{IS}^2 or V_A) components of genetic variance can then be written as:

139 $\sigma_{SR}^2 = 2 F_{SR} \sigma_{WR}^2 = 2 F_{SR} (1 - F_{RT}) \sigma_{T(0)}^2$ (Eq 7),

140 $\sigma_{IS}^2 = (1 - F_{SR}) \sigma_{WR}^2 = (1 - F_{SR}) (1 - F_{RT}) \sigma_{T(0)}^2$ (Eq 8).

141 Expressing σ_{SR}^2 in terms of σ_{IS}^2 led to

142 $\sigma_{SR}^2 = 2 F_{SR} \frac{\sigma_{IS}^2}{(1 - F_{SR})}$ (Eq 9),

143 Eq 9 provides a quantitative differentiation parameter for the lowest level of the hierarchy,

144 $Q_{SR(0)}$, based on the variance components:

145 $Q_{SR(0)} = F_{SR} = \frac{\sigma_{SR}^2}{\sigma_{SR}^2 + 2 \sigma_{IS}^2}$ (Eq 10),

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

147 For the differentiation at the highest level of the hierarchy, Whitlock & Gilbert (2012)

148 expressed σ_{RT}^2 in terms of σ_{SR}^2 :

149 $\sigma_{RT}^2 = 2 F_{RT} \sigma_{T(0)}^2 = 2 F_{RT} \frac{\sigma_{SR}^2}{(1 - F_{RT}) 2 F_{SR}}$ (Eq 11).

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

151 $Q_{RT(0)} = F_{RT} = \frac{\sigma_{RT}^2}{\sigma_{RT}^2 + \sigma_{SR}^2 + 2 \sigma_{IS}^2}$ (Eq 12).

152 Similarly, *Q-statistics* can be extended to any level of hierarchy (demonstration given in
 153 supporting information S1). If we note the hierarchical levels 1 to n, from lowest to highest
 154 (n=1 in a simple population structure), we get:

$$Q_{n(0)} = F_n = \frac{\sigma_n^2}{\left[\sum_{i=1}^n \sigma_i^2 \right] + 2\sigma_{IS}^2} \quad (\text{Eq 13}).$$

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

157 . Simple population structure

158 Wright (1951) generalized Eq 1-3 to relax the assumption of Hardy-Weinberg equilibrium
 159 within sub-populations (positive or negative F_{IS}). In this case, the *F-statistics* are no longer the
 160 same ($F_{IS} \neq 0$ and $F_{IT} \neq F_{ST}$). Accordingly, Wright derived the components of genetic variance,
 161 noted $\sigma_{IS(f)}^2$, $\sigma_{ST(f)}^2$ and $\sigma_{IT(f)}^2$, as follows. Considering that departure from Hardy-Weinberg
 162 within populations does not affect the variance between populations, $\sigma_{ST(f)}^2$ remains the same

163 as σ_{ST}^2 : $\sigma_{ST(f)}^2 = 2 F_{ST} \sigma_{T(0)}^2$. At the level of the whole population, we now have:

164 $\sigma_{IT(f)}^2 = (1 + F_{IT}) \sigma_{T(0)}^2$. Wright then deduced the variance within populations from both

165 previous equations: $\sigma_{IS(f)}^2 = (1 + F_{IT} - 2 F_{ST}) \sigma_{T(0)}^2$.

166 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

167 $\sigma_{IS(f)}^2 = (1 + F_{IS})(1 - F_{ST}) \sigma_{T(0)}^2$ (Eq 14).

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

$$172 \quad Q_{ST(f)} = F_{ST} = \frac{(1 + F_{IS})\sigma_{ST(f)}^2}{(1 + F_{IS})\sigma_{ST(f)}^2 + 2\sigma_{IS(f)}^2} \quad (\text{Eq 15}).$$

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

174 . Extension to a hierarchical population structure

175 Here, we follow Bonnin *et al.* (1996) and generalize the hierarchical $Q_{RT(0)}$ and $Q_{ST(0)}$ to the
 176 case of populations not at Hardy-Weinberg equilibrium. Following Wright's demonstration,
 177 the variance between regions is not affected by population structure, thus Eq 5 still holds and
 178 $\sigma_{RT(f)}^2 = \sigma_{RT(0)}^2 = 2F_{RT}\sigma_{T(0)}^2$. We now have to introduce two useful parameters: $\sigma_{WR(0)}^2$, which is
 179 the total genetic variance within regions if they were at Hardy-Weinberg equilibrium (i.e.
 180 analog to $\sigma_{T(0)}^2$ but for the within-region additive genetic variance), and F_{IR} , which is the F -
 181 *statistic* corresponding to individual relative to region level, i.e. the global fixation index
 182 within region. F_{IR} relates to F_{SR} and F_{IS} through the relation: $F_{IR} = F_{SR} + F_{IS} - F_{SR}F_{IS}$.

183 $\sigma_{WR(0)}^2$ is the one defined in Eq 6, whereas actual $\sigma_{WR(f)}^2$ relates to $\sigma_{WR(0)}^2$ following Wright's
 184 equations:

$$185 \quad \sigma_{WR(f)}^2 = (1 + F_{IR})\sigma_{WR(0)}^2 = (1 + F_{IR})(1 - F_{RT})\sigma_{T(0)}^2 \quad (\text{Eq 16}).$$

186 Following Eq 14 and Eq 6, $\sigma_{IS(f)}^2$ can be written in terms of F_{IS} and F_{SR} :

187 $\sigma_{IS(f)}^2 = (1 + F_{IS})(1 - F_{SR})\sigma_{WR(0)}^2 = (1 + F_{IS})(1 - F_{SR})(1 - F_{RT})\sigma_{T(0)}^2$ (Eq 17).

188 $\sigma_{SR(f)}^2$ can be written as: $\sigma_{SR(f)}^2 = 2F_{SR}\sigma_{WR(0)}^2 = 2F_{SR}(1 - F_{RT})\sigma_{T(0)}^2$ (Eq 18).

189 Expressing $\sigma_{SR(f)}^2$ in terms of $\sigma_{IS(f)}^2$, we get: $\sigma_{SR(f)}^2 = 2F_{SR} \frac{\sigma_{IS(f)}^2}{(1 + F_{IS})(1 - F_{SR})}$ (Eq 19). Resolving

190 this equation for F_{SR} gives the generalized definition of $Q_{SR(f)}$ when populations are not
191 panmictic:

192 $Q_{SR(f)} = F_{SR} = \frac{(1 + F_{IS})\sigma_{SR(f)}^2}{(1 + F_{IS})\sigma_{SR(f)}^2 + 2\sigma_{IS(f)}^2}$ (Eq 20), which is equivalent to Eq 15.

193 $Q_{RT(f)}$ is calculated accordingly, replacing F_{SR} by its value in Eq 11 for σ_{RT}^2 :

194 $\sigma_{RT(f)}^2 = 2F_{RT} \frac{\sigma_{SR(f)}^2}{2(1 - F_{RT}) \frac{(1 + F_{IS})\sigma_{SR(f)}^2}{(1 + F_{IS})\sigma_{SR(f)}^2 + 2\sigma_{IS(f)}^2}}$ (Eq 21).

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

196 $Q_{RT(f)} = F_{RT} = \frac{(1 + F_{IS})\sigma_{RT(f)}^2}{(1 + F_{IS})\sigma_{RT(f)}^2 + (1 + F_{IS})\sigma_{SR(f)}^2 + 2\sigma_{IS(f)}^2}$ (Eq 22).

197 This formula extends to any level of hierarchical structure as follows (demonstration given in
198 supporting information S1):

$$Q_{n(f)} = F_n = \frac{(1 + F_{IS})\sigma_n^2}{\left[\sum_{i=1}^n (1 + F_{IS})\sigma_i^2 \right] + 2\sigma_{IS}^2} \quad (\text{Eq 23}).$$

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 σ_{IS}^2 and σ_{ST}^2 by their values (Eq 3 and

Eq 14, respectively) in Eq 15 leads to: $Q_{ST(0)} = \frac{F_{ST}}{1 + F_{IS}(1 - F_{ST})}$, which is also valid for $Q_{SR(0)}$.

Similarly, it results from Eq 22 that: $Q_{RT(0)} = \frac{F_{RT}}{1 + F_{IS}(1 - F_{SR})(1 - F_{RT})}$. 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:
$$G = \sum_{loci} a_i + a_{i'} + k_{ii'} |a_i - a_{i'}|$$
 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 ($\sigma^2_{IS(f)}$)

236 values were directly taken from quantiNEMO outputs. We used the matrix of genotypes of the
237 last generation to estimate the hierarchical *F-statistics* (F_{RT} , F_{SR}) using *hierfstat* package in R
238 version 3.3.1. We used the matrix of genotypic values to estimate the hierarchical components
239 of genetic variance ($\sigma^2_{RT(f)}$, $\sigma^2_{SR(f)}$) from a random linear model and maximum likelihood
240 estimator using *lme4* package in R.

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

254 In the case of pure additivity, inbreeding marginally increases the mean values of hierarchical
255 *Q-statistics* and increases more significantly the variance of Q_{RT} (Figure 1). Dominance
256 drastically reduces *Q-statistics*, globally and for each hierarchical level (Figure 1).

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

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

271 **Discussion**

272 In this note, we proposed generalized definitions of *Q-statistics* for hierarchical population
273 structures that keep the property of analogy to hierarchical *F-statistics* even when the lowest
274 level of the population structure is not at Hardy-Weinberg equilibrium, in case of neutrality.
275 This property is essential to infer selection from the $Q_{ST} - F_{ST}$ comparison, and these
276 generalized definitions should be preferred to the simpler ones that assume Hardy-Weinberg
277 equilibrium. In random mating populations, one generation of panmixia is theoretically
278 enough to restore equilibrium after disturbance and, therefore, we expect little departure from

279 Hardy-Weinberg equilibrium in practice. However, inbred mating systems or unrevealed
280 population substructure may generate positive F_{IS} , while partial asexuality may generate
281 negative F_{IS} (Stoeckel & Masson 2014). In this study, we quantified the discrepancy between
282 Q -statistics and F -statistics induced by the use of Q -statistics definitions that do not account
283 for an existing sub-structure in a population system, i.e. non-null F_{IS} values. This leads to
284 under-estimate Q -statistics when F_{IS} is positive, and vice-versa. Therefore, when positive F_{IS}
285 is neglected in the computation of Q -statistics, the inference of adaptive divergence when Q -
286 statistics are significantly higher than F -statistics is conservative but, reversely, inference of
287 adaptive convergence cannot be concluded when Q -statistics are lower than F -statistics. The
288 reverse effect occurs with negative F_{IS} values. Interestingly, in a hierarchical population
289 structure, high differentiation at the lowest hierarchical level reduces the bias due to
290 neglecting F_{IS} on the upper level Q -statistics.

291 The analytical framework leading to the definition of generalized Q -statistics relies on the
292 assumptions of additivity and independence among neutral loci. Here, in a hierarchical island
293 model, we show that dominance drastically reduces hierarchical Q -statistics (Q_{RT} , Q_{SR}) below
294 their related F -statistics (F_{RT} , F_{SR}), as initially shown by Goudet & Büchi (2006) and Goudet
295 & Martin (2007) in a simple island model. Therefore, dominance makes the $Q_{ST} - F_{ST}$
296 comparison a conservative approach when inferring selective divergence from high Q_{ST} values
297 but no conclusion can be drawn from low Q_{ST} values. For empirical studies, it should be
298 noticed that, even in the absence of selection, historical and demographic processes can
299 generate random statistical dependencies among loci that are captured in the Q -statistics but
300 not in the F -statistics. These covariances contribute to the high variance of Q -statistics. In our
301 simulated hierarchical island model, discrepancies between Q -statistics and F -statistics were

302 more likely at the upper level of the hierarchy ($Q_{RT} - F_{RT}$), and the variance of Q_{RT} was
303 particularly sensitive to inbreeding in the case of pure additivity. Therefore, the power to
304 detect selection from the $Q_{ST} - F_{ST}$ approach is lower between regions than within regions.
305 When there is dominance, the mean and the variance of Q -statistics are both reduced. To
306 account for these neutral sources of covariance among loci, Q -statistics should rather be
307 compared to measures of neutral differentiation patterns that account for these disequilibria,
308 such as the CF_{ST} proposed by Kremer *et al.* (1997). Statistical developments (Martin *et al.*,
309 2008) and the recent availability of genome-wide resources have provided some clues for the
310 analysis of polygenic adaptation (Berg & Coop 2014; Yeaman 2015; Stephan 2016).

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

324 **Acknowledgement**

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

329 **References**

Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. *PLoS Genet*, **10**, e1004412.

Bonnin I, Prosperi JM, Olivieri I (1996) Genetic markers and quantitative genetic variation in *Medicago truncatula* (*Leguminosae*): a comparative analysis of population structure. *Genetics*, **143**, 1795–1805.

Gavrilets S, Hasting A (1995) Dynamics of polygenic variability under stabilizing selection, recombination, and drift. *Genetical Research*, **65**, 63–74.

Goudet J, Büchi L (2006) The effects of dominance, regular inbreeding and sampling design on Q_{ST} , an estimator of population differentiation for quantitative traits. *Genetics*, **172**, 1337–1347.

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

Kremer A, Zanetto A, Ducouso A (1997) Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics*, **145**, 1229–1241.

Kremer A, Le Corre V (2012) Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity*, **108**, 375–385.

Lande R (1992) Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution*, **46**, 381–389.

Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist*, **151**, 283–292.

Leinonen T, O’Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1–17.

Leinonen T, McCairns RJS, O’Hara RB, Merilä J (2013) Q_{ST} – F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Review Genetics*, **14**, 179–190.

Comment citer ce document :

Cubry, P., Scotti, I., Oddou-Muratorio, S., Lefèvre, F. (2017). Generalization of the QST framework in hierarchically structured populations: Impacts of inbreeding and dominance. *Molecular Ecology Resources*, 17 (6), e76 - e83. , DOI : 10.1111/1755-0998.12693

- López-Fanjul C, Fernández A, Toro MA (2003) The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genetics*, **164**, 1627-1633.
- Martin G, Chapuis E, Goudet J (2008) Multivariate Q_{ST} – F_{ST} comparisons: a neutrality test for the evolution of the G matrix in structured populations. *Genetics*, **180**, 2135-2149.
- Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNEMO: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552-1553.
- Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution*, **29**, 165-176.
- Santure AW, Wang J (2009). The joint effects of selection and dominance on the Q_{ST} - F_{ST} contrast. *Genetics*, **181**, 259-276.
- Scotti I, Gonzalez-Martinez SC, Budde KB, Lalagüe H (2016). Fifty years of genetic studies: what to make of the large amounts of variation found within populations? *Annals of Forest Science*, **73**, 69-75.
- Slatkin M, Voelm L (1991) F_{ST} in a hierarchical island model. *Genetics*, **127**, 627–629.
- Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, **135**, 367–374.
- Stephan W (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, **25**, 79–88.
- Stoeckel S, Masson JP (2014) The exact distributions of F_{IS} under partial asexuality in small finite populations with mutation. *PLoS ONE*, **9**, e85228.
- Whitlock MC (1999) Neutral additive variance in a metapopulation. *Genetical Research*, **74**, 215-221.
- Whitlock MC (2008) Evolutionary inference from Q_{ST} . *Molecular Ecology*, **17**, 1885–1896.
- Whitlock MC, Gilbert KJ (2012) Q_{ST} in a hierarchically structured population. *Molecular Ecology Resources*, **12**, 481–483.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.
- Yang RC (1998) Estimating hierarchical F -Statistics. *Evolution*, **52**, 950–956.
- Yeaman S (2015) Local adaptation by alleles of small effect. *The American Naturalist*, **186**(Suppl), S74-S89.

330 **Data accessibility**

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

333 **Author contributions**

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

336 **Tables and Figures**

337 Table 1: Definition of generalized *Q-statistics* taking into account inbreeding in
338 hierarchically structured populations. See the main text for the definitions of the parameters.

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

344 Figure 2: Relation between *Q-statistics* and *F-statistics* in the same hierarchical island model
345 as Figure 1. Symbols reflect the migration rates (o: $m=0.002$; Δ : $m=0.01$; +: $m=0.05$; x:
346 $m=0.2$).

347 **Supporting information**

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

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

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

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

355 Table 1

356

Population structure	Definition of generalized hierarchical <i>Q</i> -statistics	Ref
1-level structure	$Q_{ST(f)} = \frac{(1 + F_{IS})\sigma_{ST(f)}^2}{(1 + F_{IS})\sigma_{ST(f)}^2 + 2\sigma_{IS(f)}^2} \quad (\text{Eq 15})$	(1)
2-levels hierarchy	$Q_{SR(f)} = \frac{(1 + F_{IS})\sigma_{SR(f)}^2}{(1 + F_{IS})\sigma_{SR(f)}^2 + 2\sigma_{IS(f)}^2} \quad (\text{Eq 20})$ $Q_{RT(f)} = \frac{(1 + F_{IS})\sigma_{RT(f)}^2}{(1 + F_{IS})\sigma_{RT(f)}^2 + (1 + F_{IS})\sigma_{SR(f)}^2 + 2\sigma_{IS(f)}^2} \quad (\text{Eq 22})$	(2)
n-levels hierarchy	$Q_{n(f)} = F_n = \frac{(1 + F_{IS})\sigma_n^2}{\left[\sum_{i=1}^n (1 + F_{IS})\sigma_i^2 \right] + 2\sigma_{IS}^2} \quad (\text{Eq 23})$	(2)

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

Version postprint