

Pierre de Villemereuil¹, Oscar E. Gaggiotti², and Jérôme Goudet³¹*Institut de Systématique, Évolution, Biodiversité (ISYEB), École Pratique des Hautes Études | PSL, MNHN, CNRS, Sorbonne Université, Université des Antilles, Paris, France*²*Scottish Oceans Institute, University of St Andrews, Fife, KY16 8LB, United Kingdom*³*Department of Ecology and Evolution & Swiss Institute of Bioinformatics, University of Lausanne, 1015 Switzerland***Abstract**

1
2
3
4
5
6
7
8
9
10
11

1. Common garden experiments are valuable to study adaptive phenomenon and adaptive potential, in that they allow to study local adaptation without the confounding effect of phenotypic plasticity. The $Q_{ST}-F_{ST}$ comparison framework, comparing genetic differentiation at the phenotypic and molecular level, is the usual way to test and measure whether local adaptation influences phenotypic divergence between populations.

2. Here, we highlight that the assumptions behind the expected equality $Q_{ST} = F_{ST}$ under neutrality correspond to a very simple model of population genetics. While the equality might, on average, be robust to violation of such assumptions, more complex population structure can generate strong evolutionary noise.

12 **Synthesis** We highlight recent methodological developments aimed at overcoming this issue and at providing
13 a more general framework to detect local adaptation, using less restrictive assumptions. We invite
14 empiricists to look into these methods and theorists to continue developing even more general meth-
15 ods.

16 **Keywords:** Local adaptation, common garden, $Q_{ST} - F_{ST}$ comparison, population structure, phenotypic di-
17 vergence, population genetics, quantitative genetics.

Introduction

18
19 There are three main possible responses from organisms subjected to climate change (Parmesan 2006; Aitken,
20 Yeaman, Holliday, Wang, and Curtis-McLane 2008): they can (i) disperse and change their range limits, match-
21 ing the new geographic repartition of their ecological niche, (ii) quickly acclimate to the new climatic condi-
22 tions through (possibly transgenerational) phenotypic plasticity or (iii) in the longer run, evolve genetically
23 to match the new climatic conditions through an adaptive process. Predicting the likeliness of these three out-
24 comes (or any combination thereof) is challenging and requires both reliable models and detailed information

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:

10.1111/1365-2745.13258

This article is protected by copyright. All rights reserved.

25 about the focal species. Probably one of the most difficult tasks is to assess the extent to which phenotypic
26 plasticity or genetic adaptation allow a species to cope with the climatic challenge. This is, in part, because
27 phenotypic plasticity (including epigenetic changes) and adaptive evolution cannot be disentangled from sim-
28 ple observations in the natural environment, and also because they greatly differ in their characteristics to
29 respond to climate change (phenotypic plasticity having a much shorter response time).

30 Fortunately, “common gardens”, experimental settings in which offspring from different populations are
31 raised under the same (sets of) environmental conditions, allow to circumvent this difficulty and to assess the
32 amount of adaptive genetic variation that exists among populations (Kawecki and Ebert 2004; Savolainen, Las-
33 coux, and Merila 2013; de Villemereuil, Gaggiotti, Mouterde, and Till-Bottraud 2016). A particular challenge
34 in the analysis of such experiment is to distinguish between neutral and adaptive genetic variation among
35 populations. Indeed, even in absence of selection, and as a result of the combined effect of drift, mutation
36 and migration, populations are expected to diverge from a phenotypic point of view, just as they do from a
37 genetic one (Lande 1992). As the fixation index (F_{ST}) measures the genetic divergence between populations
38 from a molecular point of view, the phenotypic divergence can be measured through a parameter named
39 Q_{ST} (Spitze 1993). Both can be defined as the ratio of the between-population genetic variance over the total
40 genetic variance (Lande 1992; Spitze 1993):

$$Q_{ST} = \frac{V_B}{V_B + 2V_A}, \quad (1)$$

41 where V_A is the within-population additive genetic variance and V_B is the between population genetic variance.
42 Moreover, in the absence of selection (i.e. for a purely neutral trait), it is expected that $F_{ST} = Q_{ST}$ (Whitlock
43 2008). Thus, hypothesis testing based on a null hypothesis of neutrality, as first suggested by Spitze (1993)
44 has generated a lot of literature focused on the $Q_{ST} - F_{ST}$ comparison (reviewed in Leinonen, O’Hara, Cano,
45 and Merilä 2008; Leinonen, McCairns, O’Hara, and Merilä 2013). In these studies, a phenotypic trait with Q_{ST}
46 significantly larger than the F_{ST} estimated from neutral markers is considered as being under local adaptation,
47 while a Q_{ST} significantly smaller than F_{ST} is taken as a sign of balancing selection (Spitze 1993; Leinonen et
48 al. 2013). This framework of $Q_{ST} - F_{ST}$ comparison has been heavily criticised for being subjected to many
49 issues and limitations. For example, Q_{ST} is notoriously difficult to estimate and its (often large) uncertainty
50 should be carefully accounted for in the comparison (O’Hara and Merilä 2005); it can also be influenced by
51 the effect of dominance and inbreeding (Goudet and Büchi 2006; Goudet and Martin 2006; Santure and Wang
52 2008) and mutation rates might not be comparable between the Quantitative Trait Loci (QTLs) and the neutral
53 markers used to compute F_{ST} (Edelaar, Burraco, and Gomez-Mestre 2011; Edelaar and Björklund 2011). Besides
54 offering criticisms, these studies also provide strategies to overcome some of the limitations, and because such
55 a framework is so crucial to the study of local adaptation, it is still very popular in the literature (Leinonen

56 et al. 2013).

57 Here, we will focus on an issue that has been less extensively discussed and, in our experience, not
58 always considered by empiricists using Q_{ST} : the influence of population structure on the $Q_{ST}-F_{ST}$ comparison.
59 We explain the assumptions behind the “classical” computation of Q_{ST} , show a little example of how popula-
60 tion structure can affect the test for local adaptation and explore some alternatives to perform statistical tests
61 excluding the neutral hypothesis of divergence between populations with less restrictive assumptions.

62 The typical population model

63 Q_{ST} is typically estimated from common garden phenotypic measurements Y , in which each individual i
64 belongs to a distinct natural population of origin p . The average population effects a_p are considered as being
65 of genetic origin and their variance V_B is assimilated to the between-population genetic variance. While Spitze
66 (1993) used an ANOVA to compute the very first empirical Q_{ST} , it is nowadays common (Leinonen et al. 2013)
67 to use a mixed model to compute V_B and the within-population additive genetic variance V_A , using:

$$Y_i = \mu + u_i + a_{p(i)} + e_i, \quad (2)$$

68 where μ is the model intercept, u_i is the individual-level genetic random effect, $a_{p(i)}$ is the population-level
69 genetic random effect (of the population $p(i)$ the individual i belongs to) and e_i is the residual. The variance
70 associated with the a_p effects is V_B and the variance associated with u_i is V_A . The assumptions about the
71 population genetics model lie in how the structure of the random effects are specified. Generally, in order
72 to estimate the a_p effect in the model, one simply uses a “simple” random effect based on the population ID,
73 which results in the assumption that the a_p 's are independent and identically distributed. In a more formal
74 way:

$$a_p \sim \mathcal{N}(0, IV_B), \quad (3)$$

75 where I is the identity matrix and \mathcal{N} is the normal distribution. This identity matrix is akin to assuming
76 populations are structured according to an island model, where all equally-sized populations receive and
77 send the same number of migrants from a common pool (see Box 1A). Note that this model is also the one
78 generally used to derive F_{ST} estimates (Weir and Cockerham 1984, but see Gaggiotti and Foll 2010; Karhunen
79 and Ovaskainen 2012; Weir and Goudet 2017 for other approaches). In practice, this model is quite robust to
80 slight deviations from its assumptions, as shown by decades of experience from population geneticists using
81 F_{ST} (Holsinger and Weir 2009).

82 Theoretically, the equality $F_{ST} = Q_{ST}$ under neutrality should be valid, *on average*, for population structures

Box 1: Comparing island and stepping-stone models

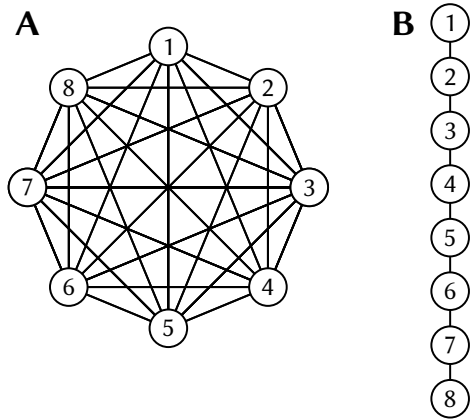
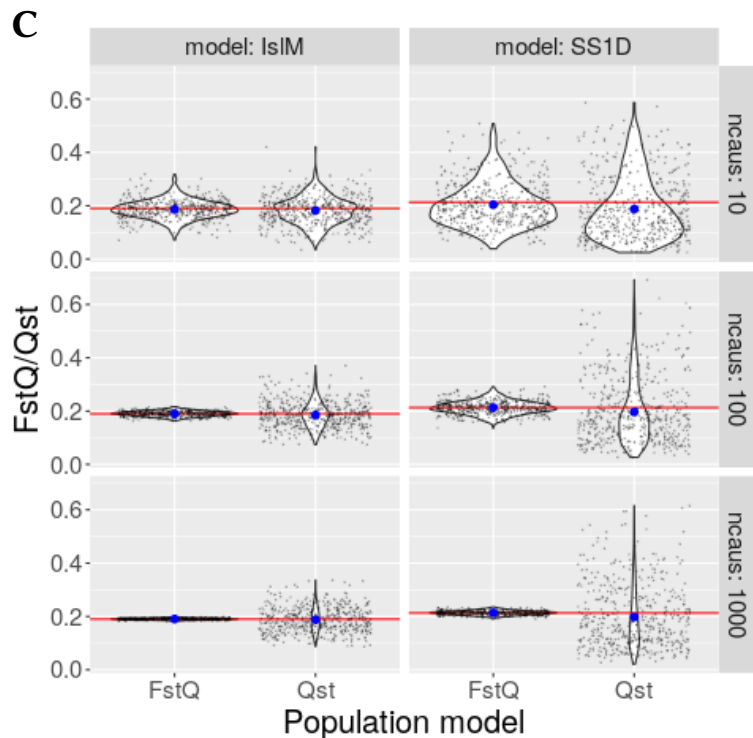


Figure C shows the results. For each scenario (IsIM: Island model, SS1D: Stepping-Stone), 500 neutral, purely additive and genetically determined traits are simulated with either 10, 100 or 1000 (top to bottom row) causal loci and effect sizes drawn from a normal distribution. The red horizontal line on each panel is the genomic F_{ST} , the violin plots labelled F_{ST}^Q correspond to F_{ST} estimated from causal loci, and the violin plots labelled Q_{ST} show the distribution of Q_{ST} for each scenario. Dots are the observations. The blue points show the mean value of the observations. For both models, F_{ST}^Q is essentially unbiased, with more variation when the number of loci encoding the trait is small, as expected. For Q_{ST} , while more variable than F_{ST}^Q , the estimation for the island model are centered around their expected value, with an empirical 95% percentile interval [0.1, 0.27]. For the one-dimensional stepping stone model, despite the mean being close to its expected value, a majority of Q_{ST} estimates are smaller than the expected value, with a mode around 0.1 and the range of variation is extremely wide, from 0 to 0.7, and an empirical 95% percentile interval [0.05, 0.44]. Such a variation due to very large evolutionary stochasticity for Q_{ST} (note that statistical sampling errors have been minimized here, since V_A is the phenotypic variance and V_B is estimated from 20 populations, so using the statistical framework in Equation 2 would most likely aggravate the issue) makes carrying a test for the null hypothesis $Q_{ST} = F_{ST}$ prone to a high level of type I errors, if not accounting for such effect of population structure. Indeed, for the island model, 5% of the simulated Q_{ST} are significant at the 5% nominal level when using the modified Lewontin-Krakauer test proposed by Whitlock and Guillaume (2009), while 30% are significant for the one dimensional stepping-stone model.

To illustrate the effect of population structure on the $F_{ST} - Q_{ST}$ contrast, we compare the results of simulations of the island model (A), where all populations exchange the same number of migrants with all others, and the one dimensional stepping stone model (B), where each population exchanges migrants with its nearest neighbours only. The data were generated with the ms program (Hudson 2002), assuming a genome of 20 chromosomes each 100'000 base pair long. Migration was adjusted to obtain an overall F_{ST} around 0.2 for the two population structures. In both sets of simulations, the number of populations is 20, and 500 individuals per population are sampled.



83 other than the island model (Whitlock 1999). We show in Box 1 this equality to hold for the island model and,
84 *on average* indeed, for the one-dimensional stepping-stone model. However, for this last model, a majority of
85 traits gave Q_{ST} values (much) lower than F_{ST} , and some gave Q_{ST} values much higher than F_{ST} (Box 1C). We
86 argue that the test $Q_{ST} = F_{ST}$ is not reliable in general because of the much larger evolutionary stochasticity
87 associated with Q_{ST} , which inflates type I errors. Hence implementation of the test should take into account
88 such evolutionary variance. This is particularly important when the population structure deviates from a
89 simple island model because the increase in evolutionary stochasticity is much more pronounced for Q_{ST}
90 than for F_{ST} .

91 **Toward a solution: more general strategies**

92 It is difficult to come up with a universal model of neutral evolution, that can account for any sort of population
93 structure. However, there have been some methodological developments that aim at testing local adaptation
94 in a common garden setting with more general assumptions about population structure. Here, we wish to
95 highlight two of them.

96 Ovaskainen, Karhunen, Zheng, Arias, and Merilä (2011) method is not based on a direct $Q_{ST} - F_{ST}$ compar-
97 ison. Instead, it is based on a theoretically motivated neutral model of phenotypic divergence that allows for
98 differences in migration and drift among populations, as well as preferential migration between populations.
99 To achieve this, the identity matrix I is replaced by a between-population relatedness matrix (here noted B)
100 in Equation 3:

$$a_p \sim \mathcal{N}(0, BV_B), \quad (4)$$

101 To estimate this matrix from neutral marker data the same authors propose an extension of the F-model (Gag-
102 giotti and Foll 2010) that also allows the simultaneous estimation of the other parameters associated with
103 Equation 2. Because B is a matrix, and not just a single number as is F_{ST} , this framework offers both a more
104 accurate description of the population structure and more power to detect deviation from neutrality. It does so
105 by alleviating the issues affecting the direct comparison of Q_{ST} with F_{ST} and provides a statistical test (coined
106 “S-test”) measuring the deviation of the population means themselves from the neutral expectation. A more
107 recent implementation of the method in R (driftsel, Karhunen, Merilä, Leinonen, Cano, and Ovaskainen 2013)
108 first estimates the matrix using neutral markers and an admixture F-model (Karhunen and Ovaskainen 2012),
109 and then incorporates quantitative trait data to estimate all remaining parameters and further refine the esti-
110 mate of the matrix. Finally, a new alternative method to estimate the B matrix (which can then be used by
111 driftsel) is provided by the unified approach to characterise population structure and individual relatedness
112 and inbreeding recently put forward by Weir and Goudet (2017). Using this model of neutral evolution ap-

113 plicable to both molecular and quantitative trait data also allows for the further addition of environmental
114 information, and as a result, perform more powerful statistical tests using habitat information (Karhunen,
115 Ovaskainen, Herczeg, and Merilä 2014) or single environmental values (de Villemereuil, Mouterde, Gaggiotti,
116 and Till-Bottraud 2018), while accounting for the (potentially confounding) effect of genetic drift and popula-
117 tion structure.

118 Sometimes, the population structure is so complex that even defining populations can be very difficult and,
119 in the end, a fairly subjective process (Waples and Gaggiotti 2006). Martins, Caye, Luu, Blum, and François
120 (2016) have shown that, in such cases, an equivalent to F_{ST} can be derived even in absence of delimited
121 populations, using either the genomic proportion of an individual assigned to a given cluster (the Q matrix of
122 the Structure program) or the individual scores along the different axes obtained from a Principal Component
123 Analysis of the genotypes. This later approach was used by Josephs, Berg, Ross-Ibarra, and Coop (2019) to
124 define an equivalent to Q_{ST} (coined Q_X) in absence of explicitly defined populations. However, to circumvent
125 the need for a direct comparison to F_{ST} , Josephs et al. (2019) suggest testing for a phenotypic excess of variance
126 (F -test) along some of the Principal Component axes retained for the analysis. The difficulty in the analysis lies
127 in identifying the set of first axes defined to be “among” populations (the other lower axes being considered
128 “within”). Josephs et al. (2019) offer various ways of defining this limit between among and within populations,
129 ranging from using an arbitrary threshold to using the Tracy-Widom test.

130 Conclusion

131 Since it was first proposed by Spitze (1993), the $Q_{ST} - F_{ST}$ comparison framework has been an invaluable tool
132 to investigate the prevalence and characteristics of local adaptation. Nonetheless, we encourage empiricists to
133 consider the alternative strategies discussed here to better account for various effects of population structure.
134 As these new strategies themselves have their own limitations, we also urge theorists to continue developing
135 new methods to study the phenotypic impact of local adaptation in common garden, while accounting for
136 population structure and evolution stochasticity as accurately as possible.

137 Authors contribution

138 PdV led the writing of the manuscript with critical contributions from OEG and JG. JG performed the simula-
139 tion analysis, with analysis from all authors. All authors agreed to the publication.

140 Data Availability

141 We provide the code for replicating the analysis in Supplementary Information.

142 References

- 143 Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., and Curtis-McLane, S. (2008). Adaptation, migration or extirpation:
144 climate change outcomes for tree populations. *Evolutionary Applications* 1.1, 95–111.
- 145 de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., and Till-Bottraud, I. (2016). Common garden experiments in the
146 genomic era: new perspectives and opportunities. *Heredity* 116.3, 249–254.
- 147 de Villemereuil, P., Mouterde, M., Gaggiotti, O. E., and Till-Bottraud, I. (2018). Patterns of phenotypic plasticity and local
148 adaptation in the wide elevation range of the alpine plant *Arabis alpina*. *Journal of Ecology* 106.5, 1952–1971.
- 149 Edelaar, P. and Björklund, M. (2011). If FST does not measure neutral genetic differentiation, then comparing it with QST
150 is misleading. Or is it? *Molecular Ecology* 20.9, 1805–1812.
- 151 Edelaar, P., Burraco, P., and Gomez-Mestre, I. (2011). Comparisons between QST and FST—how wrong have we been?
152 *Molecular Ecology* 20.23, 4830–4839.
- 153 Gaggiotti, O. E. and Foll, M. (2010). Quantifying population structure using the F-model. *Molecular Ecology Resources*
154 10.5, 821–830.
- 155 Goudet, J. and Martin, G. (2006). Under neutrality, $QST \leq FST$ when there is dominance in an island model. *Genetics*
156 176.2, 1371–1374.
- 157 Goudet, J. and Büchi, L. (2006). The effects of dominance, regular inbreeding and sampling design on QST, an estimator
158 of population differentiation for quantitative traits. *Genetics* 172.2, 1337–1347.
- 159 Holsinger, K. E. and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and inter-
160 preting F ST. *Nature Reviews Genetics* 10.9, 639–650.
- 161 Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18.2,
162 337–338.
- 163 Josephs, E. B., Berg, J. J., Ross-Ibarra, J., and Coop, G. (2019). Detecting adaptive differentiation in structured populations
164 with genomic data and common gardens. *Genetics* 211.3, 989–1004.
- 165 Karhunen, M., Merilä, J., Leinonen, T., Cano, J. M., and Ovaskainen, O. (2013). DRIFTSEL: an R package for detecting
166 signals of natural selection in quantitative traits. *Molecular Ecology Resources* 13.4, 746–754.
- 167 Karhunen, M., Ovaskainen, O., Herczeg, G., and Merilä, J. (2014). Bringing habitat information into statistical tests of
168 local adaptation in quantitative traits: a case study of nine-spined sticklebacks. *Evolution* 68.2, 559–568.
- 169 Karhunen, M. and Ovaskainen, O. (2012). Estimating population-level coancestry coefficients by an admixture F model.
170 *Genetics* 192.2, 609–617.
- 171 Kawecki, T. J. and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters* 7.12, 1225–1241.
- 172 Lande, R. (1992). Neutral theory of quantitative genetic variance in an island model with local extinction and colonization.
173 *Evolution* 46.2, 381–389.

- 174 Leinonen, T., O'Hara, R. B., Cano, J., and Merilä, J. (2008). Comparative studies of quantitative trait and neutral marker
175 divergence: a meta-analysis. *Journal of Evolutionary Biology* 21.1, 1–17.
- 176 Leinonen, T., McCairns, R. J. S., O'Hara, R. B., and Merilä, J. (2013). QST–FST comparisons: evolutionary and ecological
177 insights from genomic heterogeneity. *Nature Reviews Genetics* 14.3, 179–190.
- 178 Martins, H., Caye, K., Luu, K., Blum, M. G. B., and François, O. (2016). Identifying outlier loci in admixed and in continuous
179 populations using ancestral population differentiation statistics. *Molecular Ecology* 25.20, 5029–5042.
- 180 O'Hara, R. B. and Merilä, J. (2005). Bias and precision in QST estimates: problems and some solutions. *Genetics* 171.3,
181 1331–1339.
- 182 Ovaskainen, O., Karhunen, M., Zheng, C., Arias, J. M. C., and Merilä, J. (2011). A new method to uncover signatures of
183 divergent and stabilizing selection in quantitative traits. *Genetics* 189.2, 621–632.
- 184 Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution,
185 and Systematics* 37, 637–669.
- 186 Santure, A. W. and Wang, J. (2008). The joint effects of selection and dominance on the QST - FST contrast. *Genetics*
187 181.1, 259–276.
- 188 Savolainen, O., Lascoux, M., and Merila, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics* 14.11,
189 807–820.
- 190 Spitze, K. (1993). Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* 135.2,
191 367–374.
- 192 Waples, R. S. and Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for
193 identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15.6, 1419–1439.
- 194 Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38.6,
195 1358–1370.
- 196 Weir, B. S. and Goudet, J. (2017). A unified characterization of population structure and relatedness. *Genetics* 206.4, 2085–
197 2103.
- 198 Whitlock, M. C. (1999). Neutral additive genetic variance in a metapopulation. *Genetics Research* 74.3, 215–221.
- 199 — (2008). Evolutionary inference from Q_{ST} . *Molecular Ecology* 17.8, 1885–1896.
- 200 Whitlock, M. C. and Guillaume, F. (2009). Testing for spatially divergent selection: comparing Q_{ST} to F_{ST} . *Genetics* 183.3,
201 1055–1063.