# QUANTIFYING THE INDIVIDUAL IMPACT OF ARTIFICIAL BARRIERS IN FRESHWATERS:

# A STANDARDIZED AND ABSOLUTE INDEX OF GENETIC CONNECTIVITY

Jérôme G. Prunier [1] *, Camille Poesy [1], Vincent Dubut [2], Charlotte Veyssière [3], Géraldine Loot [3], Nicolas Poulet [4] & Simon Blanchet [1,3]*

[1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS); Station d'Ecologie Théorique et Expérimentale, UMR 5321, F-09200 Moulis, France

[2] Aix Marseille Univ, CNRS, IRD, Avignon Univ, IMBE, Marseille, France

[3] CNRS, UPS, École Nationale de Formation Agronomique (ENFA) ; UMR 5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse cedex 4, France

[4] Pôle écohydraulique AFB-IMT, allée du Pr Camille Soula, 31400 Toulouse, France

*Corresponding authors:

Jérôme G. Prunier and Simon Blanchet

Station d'Ecologie Théorique et Expérimentale, UMR 5321, F-09200 Moulis, France

2 route du CNRS, 09200 Moulis

Phone: (+33)561040361

E-mail: jerome.prunier@gmail.com, simon.blanchet@sete.cnrs.fr

# Abstract

Fragmentation by artificial barriers is an important threat to freshwater biodiversity. Mitigating the negative aftermaths of fragmentation is of crude importance, and it is now essential for environmental managers to benefit from a precise estimate of the individual impact of weirs and dams on river connectivity. Although the indirect monitoring of connectivity using molecular data constitutes a promising approach, it is still plagued with several constraints preventing a standardized and individual quantification of barrier effects. Indeed, observed levels of genetic differentiation depend on both the age of the obstacle and the effective size of the populations it separates, making difficult comparisons among obstacles. Here, we developed a standardized index of genetic connectivity ($C_{INDEX}$), allowing an absolute and independent assessment of the individual effects of obstacles on connectivity. The $C_{INDEX}$ is the standardized ratio (expressed as a percentage) between the observed genetic differentiation between pairs of populations located on either side of an obstacle and the genetic differentiation expected if this obstacle completely prevented gene flow. The expected genetic differentiation is calculated from simulations taking into account both the age of the barrier and the effective size of the targeted populations. Using both simulated and published empirical datasets, we explored and discussed the validity and the limits of the $C_{INDEX}$. We demonstrated that it allows quantifying genetic effects of fragmentation only a few generations after barrier creation and provides valid comparisons among populations (or species) of different effective populations sizes and obstacles of different ages. The computation of the $C_{INDEX}$ requires a minimum amount of fieldwork and genotypic data, and solves some of the difficulties inherent to the study of artificial fragmentation in rivers and potentially in other ecosystems. This makes the $C_{INDEX}$ a promising and objective tool for managers aiming at restoring connectivity and at evaluating the efficiency of restoration programs.



**Keywords:** Riverscape connectivity, Artificial fragmentation, Genetic differentiation, Simulations, Bio-indicator, Dams, Weirs, Freshwater fish

# Introduction

Heavily impacted by human activities, rivers are at the heart of biodiversity conservation issues (Dudgeon et al., 2006; Reid et al., 2018). Among the various threats to these ecosystems, river fragmentation by artificial barriers is considered as the most widespread and worrying (Couto & Olden, 2018; Nilsson, 2005; Turgeon, Turpin, & Gregory-Eaves, 2019). Weirs and dams, but also pipes and culverts, have long been, and are still, constructed for flow regulation and/or hydropower supply but they often imply a loss of habitat and a reduction in riverscape functional connectivity (that is, species-specific) in freshwater organisms (Birnie-Gauvin, Aarestrup, Riis, Jepsen, & Koed, 2017; Jansson, Nilsson, & Malmqvist, 2007). For fish, artificial fragmentation is known to impact key biological processes such as migration, dispersal and recruitment, and thus viability and productivity of populations and communities (Blanchet, Rey, Etienne, Lek, & Loot, 2010; Poulet, 2007; Turgeon et al., 2019). Given the central role of hydropower as a source of energy, mitigating these negative aftermaths is now of crude importance (Couto & Olden, 2018; Gibson, Wilman, & Laurance, 2017).

Different restoration and mitigation measures may be considered to enhance longitudinal river connectivity, including the removal of obstacles, periodic turbine shutdowns and fishpasses setting (Bednarek, 2001; Poff & Schmidt, 2016; Silva et al., 2018). However, these measures may all result in unintended outcomes (e.g., McLaughlin et al., 2013), or unsatisfactory trade-offs between conservation of biodiversity, preservation of historical and cultural legacy and the maintenance of services provided by obstacles (Gibson et al., 2017; Hand et al., 2018; Roy et al., 2018; Song et al., 2019). Benefiting from a precise estimate of the impact of an obstacle on river connectivity, or from a precise estimate of the gain in connectivity resulting from a restoration action, is therefore essential for environmental managers and for conservation planning (Cooke & Hinch, 2013; Januchowski-Hartley, Diebel, Doran, & McIntyre, 2014; Raeymaekers, Raeymaekers, Koizumi, Geldof, & Volckaert, 2009).

The direct monitoring methods conventionally used in rivers to quantify the functional permeability of an obstacle or the efficiency of a restoration action are video-counting, telemetry and capture-recapture protocols. Although efficient (e.g., Cooke & Hinch, 2013; Hawkins et al., 2018; Junge et al., 2014; Pracheil et al., 2015), these methods are yet associated with technical constraints. In particular,

ecological studies based on video counting or telemetry are often conducted on a limited number of obstacles, whereas robust capture-recapture protocols imply repeated and exhaustive capture sessions, ideally over several years, which involves the mobilization of substantial human and financial resources (Cayuela et al., 2018).

Indirect monitoring based on molecular data constitutes a promising alternative approach, allowing multi-specific studies of dam-induced fragmentation (Selkoe, Scribner, & Galindo, 2015). Among the many analytical procedures developed in recent years to quantify the mobility of organisms on the basis of genetic or genomic data, assignment methods and parentage analyses (Jombart, Devillard, & Balloux, 2010; Pritchard, Stephens, & Donnelly, 2000; Städele & Vigilant, 2016; Wilson & Rannala, 2003) allow the detection of 'real-time' non-effective movements (that is, not necessarily followed by a reproduction event; e.g., Junge et al., 2014; Raeymaekers et al., 2009; Saint-Pé et al., 2018) but they require an extensive sampling of individuals and moderate to high genetic differentiation between populations (Broquet & Petit, 2009; Cayuela et al., 2018).

An alternative method to quantify the permeability of an obstacle from molecular data is simply to measure the level of neutral genetic differentiation between populations located in the immediate upstream and downstream vicinity of an obstacle ("adjacent sampling strategy"), an approach that does not necessarily require large sample sizes or heavy computation: any drop in local functional connectivity due to the creation of a barrier to gene flow is expected to translate into an increase in neutral genetic differentiation (Raeymaekers et al., 2009). However, measures of genetic differentiation may only be considered as correct estimates of barrier effects when comparing obstacles of the *same* age and/or separating populations of *similar* effective size. This is because genetic differentiation primarily stems from genetic drift, that is, from the random fluctuation of allelic frequencies naturally occurring in all populations (Allendorf, 1986). When populations are separated by an obstacle to gene flow, these fluctuations tend to occur independently in each population, leading to a differential distribution of allelic frequencies on either side of the barrier. This process is yet progressive, taking place over several generations (Landguth et al., 2010), and is all the more slow as effective population sizes are large (Broquet & Petit, 2009; Cayuela et al., 2018; Prunier, Dubut, Chikhi, & Blanchet, 2017). As a consequence, it is impossible to attribute the differences in levels of

genetic differentiation observed between obstacles varying in age and/or in the effective size of populations they separate to differences in their actual barrier effects; older obstacles or obstacles separating smaller populations should show higher genetic differentiation than more recent obstacles or obstacles separating larger populations, despite similar barrier effects. Given this drawback, there is an urgent need for the development of a *standardized* and *absolute* index of genetic connectivity that take into account the contribution of both the age of the obstacle and the effective size of populations to observed measures of genetic differentiation. Such an index might allow a quick and robust quantification of individual barrier effects whatever their characteristics, paving the way for informed management prioritization and proper evaluation of restoration measures, along with inter-basins and interspecific comparative studies.

Here, we bridge that gap by developing a user-friendly and standardized index of genetic connectivity, allowing an absolute and independent assessment of the individual effects of obstacles on gene flow. The proposed index ($C_{INDEX}$) is based on the relative comparison of observed measures of genetic differentiation resulting from an adjacent sampling strategy with theoretical machine learning predictions obtained from genetic simulations. Genetic simulations are here used to reflect the expected evolution of allelic frequencies resulting from the interplay between the age of the obstacle and the effective population sizes: the closer the observed measure of genetic differentiation from the one that would be expected in the worst-case scenario (total barrier to gene flow), the lower the index of connectivity. We first present the logic and principles underlying our index. We then use both simulated and published empirical genetic datasets to explore and discuss the validity and the limits of the proposed index. We finally propose several perspectives to use and further improve the index.

# Material and methods

**Principle of the proposed index of genetic connectivity $C_{INDEX}$**

The proposed index of genetic connectivity $C_{INDEX}$ is designed as a standardized estimate of the amount of gene flow that gets through an obstacle separating two adjacent populations. It simply consists in rescaling the observed measure of genetic differentiation $GD_{obs}$ within its theoretical range of variation, taking into account the expected temporal evolution of allelic frequencies resulting from the interplay between the age of the obstacle and the averaged effective sizes of populations. This theoretical range of variation spans from $GD_{min}$ to $GD_{max}$. $GD_{min}$ stands for the theoretical measure of genetic differentiation that would be expected if the obstacle was totally permeable to gene flow (migration rate $m \approx 0.5$). $GD_{min}$ should theoretically equal 0 but the background noise resulting from the concomitant influences of genetic drift, mutations and random sampling may actually lead to positive –yet very low– measures of genetic differentiation. On the other hand, $GD_{max}$ stands for the theoretical measure of genetic differentiation that would be expected under the worst-case scenario, that is, under the hypothesis that the considered obstacle is a total barrier to gene flow ($m = 0$). $GD_{max}$ is expected to increase with time since barrier creation and to decrease with the increase in effective population sizes (Gauffre, Estoup, Bretagnolle, & Cosson, 2008; Landguth et al., 2010). The index of connectivity $C_{INDEX}$ is then computed as follows (see Appendix S1 for details)

$$C_{INDEX} = \left(\frac{\ln(GD_{obs}/GD_{max})}{\ln(GD_{min}/GD_{max})}\right) \times 100 = \left(\frac{ln(GD_{obs}) - ln(GD_{max})}{ln(GD_{min}) - ln(GD_{max})}\right) \times 100 \qquad \text{(eqn. 1)}$$

It thus ranges from 0 % (the observed measure of genetic differentiation is maximum and equals the expected value under the assumption that the considered obstacle acts as a total barrier to gene flow) to 100 % (the observed measure of genetic differentiation is minimum- but not null- and equals the expected value under the assumption that the considered obstacle has no impact on gene flow). $GD_{obs}$ is directly calculated from observed genotypic data collected in populations located at the immediate upstream and downstream vicinity of the obstacle, whereas $GD_{min}$ and $GD_{max}$ are predicted from

theoretical datasets simulated according to three main parameters: the mutation rate $\mu$ of considered genetic markers, and (for $GD_{max}$ only) the effective population size *Ne* of the two considered populations and the age T of the total barrier to gene flow (expressed in number of generations; e.g., Landguth et al., 2010; Lowe & Allendorf, 2010).

**Expected measures of genetic differentiation**

We used QuantiNemo2 (Neuenschwander, Michaud, & Goudet, 2019), an individual-based simulator for population genetics, to simulate theoretical datasets that will in turn be used to predict $GD_{min}$ and $GD_{max}$ values. We designed a very simple meta-population composed of two adjacent demes, and we used forward simulations of gene flow between these two demes over 1000 non-overlapping generations. Each deme was initiated with *Ne* individuals and kept at a constant size over generations, with *Ne* taking 93 different values ranging from 30 to 2000 individuals. Genetic polymorphism was based on 15 microsatellite loci and 20 alleles per locus, which corresponds to the number of markers typically used in empirical study focusing on functional connectivity (e.g., Blanchet et al., 2010; Coleman et al., 2018; Storfer, Murphy, Spear, Holderegger, & Waits, 2010). The mutation rate $\mu$, following a stepwise mutation model, was set to $5\times10^{-5}$ or $5\times10^{-4}$, so as to explore the natural variability observed in microsatellite markers (mutation rate ranging from $10^{-6}$ to $10^{-2}$; Li, Korol, Fahima, Beiles, & Nevo, 2002; Schlötterer, 2000; Yue, David, & Orban, 2007). Genotypes were randomly assigned to individuals at the beginning of simulations. The inter-deme migration rate was set to 0.5 for the first 400 generations, a value providing an optimal mixing of populations (panmixia) and mimicking a natural situation without barrier, and then dropped to zero for the last 600 generations, mimicking the creation of a total barrier to gene flow, splitting a “single” population into two subpopulations. With populations being isolated for 600 generations, we made sure our simulations covered a realistic time frame: most artificial barriers in freshwater ecosystems were indeed built between the Middle Ages (12th–15th centuries) and today (Blanchet et al., 2010), which corresponds to a number of generations ranging from 0 to ~ 400 in aquatic organisms such as fish species (assuming a generation time of 2 years for fish species). For each deme size *Ne* and each mutation rate $\mu$, we ran ten simulation replicates, and 30 genotypes were sampled every ten

generations from generation 300 to generation 1000, resulting in a total of 132060 simulated genetic datasets in the *Fstat* format (Goudet, 1995), further converted into the *genepop* format (Rousset, 2008) using R (R Development Core Team, 2014).

For each dataset, we measured the two following pairwise metrics of genetic differentiation: the Hedrick's G''st (Hedrick, 2005; Meirmans & Hedrick, 2011) and the Meirmans' φ'st (Meirmans, 2006), both computed using the R-package *mmod* (Winter, 2012). Other metrics were initially considered, but preliminary analyses indicated that some were dependent on sample size (e.g., the proportion of shared alleles or the Cavalli-Sforza and Edwards' Chord distance; Bowcock et al., 1994; Cavalli-Sforza & Edwards, 1967; see Appendix S2 for details), while others were sensitive to mutation rate and/or did not show enough variability (e.g., the Weir and Cockerham's θst or the Jost's D; Jost, 2008; Weir & Cockerham, 1984; see Appendix S3 for details): they were thus discarded to avoid jeopardizing the validity of the proposed index. We found that the two retained metrics G''st and φ'st were robust to variations in mutation rate and increased quickly after barrier creation, especially in the case of small effective population sizes (Appendix S3), in accordance with theoretical expectations (Lowe & Allendorf, 2010; Meirmans & Hedrick, 2011). All negative G''st and φ'st values were set to 0. In addition to these two measures of genetic differentiation G''st and φ'st, we also computed the averaged expected heterozygosity *He* over the 15 loci in each population. *He* was then averaged over the two populations and further considered as a proxy for effective population sizes. Both theoretical and empirical works indeed indicate that genetic diversity should increase with the increase in effective population sizes (Hague & Routman, 2016; Kimura, 1983; see also Appendix S4). We here focused on mean heterozygosity because, unlike metrics such as allelic richness, heterozygosity values are bounded between 0 and 1, which facilitates comparison between case studies. Moreover, this metric is much more straightforward to calculate for managers than the actual effective population size, since the latter is notoriously difficult to estimate in complex landscapes (Paz-Vinas, Comte, et al., 2013; Wang, 2005). These calculations resulted in a final dataset comprising 132060 lines and the eight following columns: the simulated mutation rate $\mu$, the generation $t$ at which genotypes were recorded, the age T of the barrier computed as $T = (t - 400)$ and expressed in number of generations (the barrier being created at generation 400 in our simulations), the replicate number, the effective

population size *Ne* of each simulated population, the mean expected heterozygosity *He* and the two metrics of genetic differentiation G''st and φ'st.

A first subset of this final dataset was used as a training set in the regression implementation of a Random Forest machine-learning algorithm (Breiman, 2001). The objective was to establish theoretical distributions allowing future predictions of $GD_{max}$ values according to both time and effective population sizes (with *He* as a proxy). These theoretical distributions were designed so as to mimic the temporal inertia in the setting up of genetic differentiation after the creation of a total barrier to gene flow. For each mutation rate $\mu$ and each metric of genetic differentiation *GD* (either G''st or φ'st) computed after the creation of the barrier (i.e., for T > 0), we used the R-package *randomForest* (Liaw & Wiener, 2002) to fit the model $GD \sim T \times He$. We used 200 trees and a sample size of 500, as these values provided very good accuracy (mean squared errors lower than 0.4%). Created *randomForest* R-objects were saved in the form of *.rda* files (the usual file format for saving R objects) and were further used to predict the four possible expected measures of genetic differentiation $GD_{max}$ (two possible metrics of genetic differentiation and two possible mutation rates) between pairs of populations according to both the mean expected heterozygosity *He* (the proxy for effective population sizes) and the number of generations *T* elapsed since barrier creation, using the *predict.randomForest* function.

A second subset of the final dataset was used to predict the four possible measures of genetic differentiation $GD_{min}$ (background signal) that may be expected under the influence of mutations, drift and random sampling between two adjacent populations not separated by any barrier to gene flow. For each of both mutation rates $\mu$ and each of both metrics of genetic differentiation *GD* (either G''st or φ'st) computed before the creation of the barrier (i.e., for T < 0), $GD_{min}$ was simply computed as the mean of simulated *GD* values. These four predicted $GD_{min}$ values were stored in the form of a *.rda* file.

**Computing the index of connectivity $C_{INDEX}$**

Equation 1 allows computing a unique index of connectivity $C_{INDEX}$ for each combination of both a mutation rate $\mu$ ($5\times10^{-5}$ or $5\times10^{-4}$) and a metric of genetic differentiation *GD* (G''st or φ'st). The four

indices are then averaged to get the final connectivity index $C_{INDEX}$ with a 95% confidence interval computed as $1.96 \times SE$, with $SE$ the estimated standard error (i.e., the estimated standard deviation divided by $\sqrt{4}$).

Note that when several genotypic datasets are available for a same obstacle, for instance when several sympatric species are sampled on either side of the obstacle or when several replicates are considered (as is the case of all simulated data in this study), an overall $C_{INDEX}$ can also be estimated using an intercept-only mixed-effect linear model with the various indices $C$ as the response variable and the genotypic dataset as a random effect (Bates, Mächler, Bolker, & Walker, 2015). This procedure allows taking into account the fact that indices of connectivity $C_{INDEX}$ computed from the same dataset are not independent and thus avoids biased estimates for standard error $SE$ (McNeish, 2014). The overall $C_{INDEX}$ is then obtained from the estimated intercept of the model (which simply amounts to calculating the average of indices $C$ across datasets) and the corresponding 95% confidence interval is computed as $1.96 \times SE$, with $SE$ the unbiased standard error as estimated by the mixed-effect model.

The whole procedure was automated within a user-friendly R-function (see Appendix S10). Users are simply expected to provide an empirical genotypic dataset (in the *genepop* format) and a parameter file indicating for each considered obstacle the name of the two adjacent populations (as given in the genotypic dataset) and the number of generations elapsed since barrier creation. This number of generations is to be estimated from the life-history traits of the considered species. Figure 1 provides a flowchart allowing an overall visualization of the process.

**Validation of the connectivity index**

To assess the validity of the proposed $C_{INDEX}$ in response to different levels of obstacle permeability, we again used the program QuantiNemo2 to simulate gene flow between two demes over 1000 non-overlapping generations. Demes were initiated with *Ne* = 50, 100, 250, 500 or 1000 individuals and kept at a constant size over generations. To mimic realistic genetic datasets, each microsatellite locus

was given a unique stepwise mutation rate $\mu$ randomly picked from a log-normal distribution ranging from $5\times10^{-5}$ to $5\times10^{-3}$ with a mean of $5\times10^{-4}$ (see Appendix S5 for details). The inter-deme migration rate was set to 0.5 for the first 400 generations, and then dropped to *m* for the last 600 generations, with *m* ranging from 0 to 0.2 with an increment of 0.01 and from 0.2 to 0.5 with an increment of 0.05, mimicking the creation of a more or less severe barrier to gene flow (total barrier, *m* = 0; no barrier, *m* = 0.5). All other simulation parameters were similar to previous simulations. For each deme size *Ne* and each migration rate *m*, we ran 20 simulation replicates, and 30 genotypes were sampled at generations 405 (age of the barrier T = 5), 410, 415, 420, 425, 450, 500 and 700 (T = 300), resulting in a total of 21600 simulated genetic datasets in the *Fstat* format, further converted into the *genepop* format.

For each simulated dataset, we computed the averaged expected heterozygosity *He* and the two pairwise measures of genetic differentiation G''st and φ'st. We then used parameters *T* and *He* to predict the corresponding measures of genetic differentiation $GD_{min}$ and $GD_{max}$ (for both G''st and φ'st) expected under various mutation rates ($5\times10^{-5}$ and $5\times10^{-4}$) using the *predict.randomForest* function (Appendix S10) and the previously created *.rda* files. For each dataset, the four indices of connectivity $C$ were then computed using equation 1. To average datasets across replicates, we finally used intercept-only mixed-effect models (with dataset as a random effect) to get the final mean $C_{INDEX}$ (along with a 95% confidence interval) corresponding to each combination of *N*, *T* and *m*.

**Empirical data**

To assess the behavior of the proposed index of connectivity in real situations, we used two published empirical datasets. The first one is from Gouskov et al. (2016). In this study, authors assessed riverscape fragmentation induced by 37 hydroelectric power stations in the Rhine catchment in Switzerland using data from 2133 European chubs (*Squalius cephalus*) sampled across 47 sites and genotyped at nine microsatellite loci. We selected 23 pairs of populations according to the following criteria: upstream and downstream populations belonged to the same river, were separated by a single dam, were distant from a maximum of 16km (maximum migration distance observed in chub according to Fredrich et al., 2003) and were not separated by any confluence with tributaries. This

selection corresponded to 23 independent dams created between 1893 and 1970, all equipped with fishpasses (Table1; see also Appendix S6 for a map). We considered a generation time of 3 years, as reported in Gouskov et al. (2016) to compute the number of generations elapsed since barrier creation and ran our developed R-function to automatically compute $C_{INDEX}$ values.

The second empirical dataset is from Prunier et al. (2018). In this study, authors assessed the influence of various anthropogenic stressors including riverscape fragmentation induced by weirs on patterns of genetic diversity and differentiation in two freshwater fishes from two distinct rivers in southwestern France. They used data from 1361 Eurasian minnows (*Phoxinus phoxinus*) and 1359 Languedoc gudgeon (*Gobio occitaniae*) sampled across 47 sites (22 in the Célé River and 25 in the Viaur River) and genotyped at 11 and 13 microsatellite loci, respectively. We selected 8 pairs of populations according to the following criteria: upstream and downstream populations belonged to the same river, were separated by a single weir, were distant from a maximum of 1km, were not separated by any confluence with tributaries and were sampled for both species. This selection corresponded to 8 independent weirs (~1 to 3 meters high) created between the 16th and the 20th century (Table 1; see also Appendix S6 for maps). We considered a generation time of 2 years in *P. phoxinus* and 2.5 years in *G. occitaniae* to compute the number of generations elapsed since barrier creation and again ran developed R-function (Appendix S10) to automatically compute $C_{INDEX}$ values for each obstacle, each species and across species.

# Results

## Expected measures of genetic differentiation

The first set of simulations was designed to predict $GD_{min}$ and $GD_{max}$ values, that is, the lower and upper limits of the theoretical range of variation of $GD_{obs}$. Data simulated before the creation of the barrier ($m$ = 0.5; $T$ < 400) were used to predict $GD_{min}$ values whereas data simulated after the creation of the barrier ($m$ = 0; $T$ > 400) were used to predict $GD_{max}$ values. As expected with a migration rate allowing panmixia, $GD_{min}$ values were always very close from 0, ranging from $5.53\times10^{-3}$ to $7.17\times10^{-3}$ for G''st and from $7.46\times10^{-3}$ to $8.51\times10^{-3}$ for φ'st. These values represent the predicted background levels of genetic differentiation, resulting from the sole influences of random processes such as genetic drift, mutations and sampling biases (Figure 2).

$GD_{max}$ values were on the contrary designed to mimic the temporal inertia in the setting up of genetic differentiation after the creation of a total barrier to gene flow. They were predicted from time since barrier creation and averaged expected heterozygosity (a proxy for effective population size) using a Random Forest algorithm, simulated data being used as training sets. With explained variance ranging from 86.8 to 94.2 %, Random Forest models accurately captured variations in measures of genetic differentiation across the parameter space, whatever the considered mutation rate or the considered metric of genetic differentiation (Appendix S7). As expected in absence of gene flow (Figure 2), $GD_{max}$ increased with time since barrier creation and decreased with effective population size (*He*). With predicted $GD_{max}$ values ranging from 0.031 to 0.898 for G''st and from 0.042 to 0.968 for φ'st, both metrics displayed similar distribution patterns across mutation rates, although φ'st systematically showed higher values at low *He*.

## Validation of the connectivity index

The second set of simulations was designed to assess whether the $C_{INDEX}$ correctly reflected the actual level of gene flow between two populations separated by an artificial barrier, beyond the temporal inertia in the setting-up of genetic differentiation. The mean $C_{INDEX}$ values computed over simulated

replicates for each combination of *Ne* (effective population size), *T* (number generations since barrier creation) and *m* (migration rate) showed –as expected- an overall increase with the increase in migration rate, whatever the effective size of populations or the age of the barrier (Figure 3A-H; see Appendix S8 for a visualization against the number *Nm* of effective dispersal events per generation). As expected when population are connected with high migration rates after the barrier setting ($m > 0.2$, a migration rate of 0.5 ensuring panmixia), the 95% confidence intervals for the $C_{INDEX}$ always included values higher than 90% (Figures 3A-H). On the contrary, in absence of gene flow after the barrier setting ($m = 0$), the 95% confidence intervals always included values lower than 10%, except within the first 5 generations after barrier creation (Figure 3A), or within the first 10 generations for very large effective population sizes (Figure 3B, $Ne \geq 1000$). In these cases, the $C_{INDEX}$ was slightly biased upwards, which indicates that we could not totally rule out the noise associated with the measurement of genetic differentiation within the 5-10 first generations after barrier creation (Figure 3I; Appendix S9). Nevertheless, the $C_{INDEX}$ showed valid and consistent values for both lowest and highest migration rates, the two thresholds of 10% (total barrier to gene flow) and 90% (full gene flow) providing robust benchmarks for future interpretation of the index, whatever the age of the obstacle (from generation 15 at least) or the effective size of populations.

For low migration rates ($0 < m \leq 0.05$), the $C_{INDEX}$ showed higher variability, with two noticeable trends. First, all effective population sizes being combined, the $C_{INDEX}$ slightly increased with the increase in time since barrier creation (from generation 15 to generation 300), with a 10 to 20% increase for lowest migration rates (but not for $m = 0$; Figure 3I). Secondly, all generations > 10 being combined, the $C_{INDEX}$ slightly increased with the increase in effective population sizes, with again a 10 to 20% increase for lowest migration rates from $Ne = 50$ to $Ne = 1000$ (Figure 3J). For instance, for $m = 0.05$ after the barrier setting and for the most extreme cases, $C_{INDEX}$ values ranged from 31.1 in smallest populations to 50.1 in largest populations (mean = 41.2) at generation 15 (Figure 3C), and from 55.7 in smallest populations to 86.0 in largest populations (mean = 70.8) at generation 300 (Figure 3H).

**Empirical data**

In the first empirical dataset (Gouskov et al., 2016), monitored dams were created from 1893 to 1970, which corresponds to approximately 13 to 39 generations in *S. cephalus* (Table1). Averaged levels of expected heterozygosity were high and showed little variability (ranging from 0.69 to 0.77), whereas observed measures of genetic differentiation were pretty low, ranging from 0 to 0.032 for φ'st and from 0 to 0.018 for G''st. We found that six dams showed a $C_{INDEX}$ value ranging from 64% to 75%, suggesting from 25 to 36% decrease in genetic connectivity since barrier creation (Figure 4A). The other 17 dams all showed $C_{INDEX}$ values higher or equal to 90% (as indicated by confidence intervals), indicating that populations located on either side of the barrier are properly connected by gene flow. Importantly, $C_{INDEX}$ values were independent from both time since barrier creation (spearman correlation test, $\rho$ = -0.18, p = 0.42) and averaged heterozygosity ($\rho$ = 0.22, p = 0.32) whereas both φ'st and G''st values tended to covary negatively with averaged heterozygosity (p < 0.09) and G''st values to covary positively with time since barrier creation (p < 0.09).

In the second empirical dataset (Prunier et al., 2018), monitored weirs were built between the 16th and the 20th century, that is approximately from 20 to 204 generations in *G. occitaniae* and from 25 to 255 generations in *P. phoxinus*. As previously, averaged levels of expected heterozygosity were high and showed little variability (ranging from 0.58 to 0.72), whereas observed measures of genetic differentiation were pretty low, ranging from 0 to 0.034 for φ'st and from 0 to 0.026 for G''st. The impact of weirs was variable across space and species (Table 1; Figure 4B). In *P. phoxinus*, all weirs were found as highly permeable ($C_{INDEX}$ > 90%) with 7 out of 8 weirs showing a $C_{INDEX}$ of 100%. In *G. occitaniae*, four weirs were found as responsible for a decrease in genetic connectivity since barrier creation ($C_{INDEX}$ < 90%), with $C_{INDEX}$ values ranging from 83.9 (± 1.3%) in the case of barrier CAP in the Viaur River to 49.2 (± 3.8 %) in the case of barrier SCC in the Célé River. When computed across species, none of the weirs were identified as obstacles to overall genetic connectivity ($C_{INDEX}$ ranging from 77.2 to 100 %, with 95% confidence intervals systematically including the 90% threshold; Table 1). As previously, $C_{INDEX}$ values were independent from both time from barrier creation ($\rho$ = 0.21, p = 0.41) and averaged heterozygosity ($\rho$ = 0.18, p = 0.47), but so were φ'st and G''st values (p > 0.29).

# Discussion

Restoring riverscape functional connectivity is of crude importance in terms of biodiversity conservation and it is now often subject to regulatory obligations (e.g. in Europe, the Water Framework Directive 2000/60/EC). However, rivers are subject to many and sometimes contradictory uses (Reid et al., 2018): for practitioners to be able to propose informed trade-offs between restoring riverscape connectivity and maintaining infrastructures and their associated socioeconomic benefits (Hand et al., 2018; Roy et al., 2018; Song et al., 2019), new tools have to be developed, allowing a rapid and reliable quantification of the relative impacts of obstacles to freshwater species movements. Our objective was here develop an operational tool allowing such thorough quantification from a minimum amount of data (Figure 1; see Box 1 for user guidelines).

The proposed index of genetic connectivity $C_{INDEX}$ can be easily and automatically computed from a simple set of upstream and downstream genotypes collected once and in the direct vicinity of a putative barrier, provided the number of generations elapsed since barrier creation is known. Based on two complementary metrics of genetic differentiation (G''st and φ'st) preliminary chosen so as to limit any possible bias, the $C_{INDEX}$ simply scales the observed level of genetic differentiation ($GD_{obs}$) with respect to a theoretical range of variation spanning from the background noise expected in the absence of any barrier to gene flow ($GD_{min} \sim 0$, $C_{INDEX}$ = 100%) to the maximal level of differentiation expected if the obstacle was a total barrier to gene flow ($GD_{max}$, $C_{INDEX}$ = 0%), the latter taking into account both the time since barrier creation and the effective population size. Using numerous simulations to explore the interplay between time since barrier creation, mutation rate and averaged expected heterozygosity (a proxy for effective population size), we were able to obtain $GD_{max}$ values for a large range of biologically realistic parameters (Figure 2). As expected, $GD_{max}$ values progressively increased with time since barrier creation and decreased with averaged expected heterozygosity. Mutation rate also influenced $GD_{max}$ patterns: as expected, higher mutation rates accelerate genetic differentiation through time when population sizes are small to medium.

The $C_{INDEX}$ showed constant patterns of increase with the increase in migration rates (from $m = 0$ to $m = 0.2$), whatever the number of generations since barrier creation and the effective population size (Figure 3). For lowest migration rates ($m \leq 0.05$), we yet found that it could underestimate barrier effects in the first 5 to 10 generations after the creation of the obstacle. As a conservative strategy, we suggest that the $C_{INDEX}$ should not be used to assess the permeability of obstacles separating populations for less than 10 generations. It is yet noteworthy that the $C_{INDEX}$ can be applied to any type of organisms and thus that species with low generation time (such as some invertebrate species) may be considered as good candidates to investigate the impact of recently built barriers. For lowest migration rates ($m \leq 0.05$), we also found that $C_{INDEX}$ values slightly increased with both time since barrier creation (from generations 15 to 300) and effective population sizes (from $Ne = 50$ to $Ne = 1000$; Figure 3I-J). These trends have to be kept in mind when comparing intermediate $C_{INDEX}$ values, ranging from ~15 to ~70% (see Box 1 for an illustration).

Nevertheless, the $C_{INDEX}$ provides a promising individual quantification of both the short- and long-term genetic effects of dam-induced fragmentation, allowing robust comparisons among species or populations with different population sizes, and obstacles of different ages (from generation 15 at least) and types. When applied to empirical data, the $C_{INDEX}$ allowed identifying several obstacles partially limiting gene flow in at least two freshwater fish species (Figure 4): six out of the 23 dams monitored by Gouskov et al. (2016) and four out of the eight weirs monitored by Prunier et al. (2018) showed a $C_{INDEX}$ lower than 90%, in chubs and gudgeons, respectively. In each dataset, computed $C_{INDEX}$ values were systematically independent from both time from barrier creation and averaged heterozygosity, indicating that the $C_{INDEX}$ properly takes into account the differential evolution of allelic frequencies on either side of the barrier. Interestingly, the two most recent weirs in Prunier et al. (2018; the SCC weir on the Célé River and the CIR weir on the Viaur River, both built in the 1960's, i.e. ~20 gudgeon generations ago; Table 1) showed contrasted results in gudgeons: the SCC weir was identified as the most impactful obstacle ($C_{INDEX} = 49.2 \pm 3.8\%$), whereas the CIR weir was found as highly permeable to gene flow ($C_{INDEX} = 86.7 \pm 15.1\%$). These contrasted results suggest that distinct typological features (height, slope, presence of a secondary channel, etc.) may differently affect fish

mobility (Baudoin et al., 2014). Furthermore, none of the monitored weirs were identified as barriers to gene flow in minnows, in accordance with personal field observations and previous findings on the same two rivers (Blanchet et al., 2010). Although understanding how obstacle typological features and fish traits might interact and shape riverscape patterns of functional connectivity was beyond the scope of this study, these results suggest that future comparative studies based on the proposed $C_{INDEX}$ might provide thorough insights as to the determinants of dam-induced fragmentation in various freshwater organisms (Richardson, Brady, Wang, & Spear, 2016), including fish but also other taxa such as macro-invertebrates that display very contrasting traits related to dispersal (e.g., Alp, Keller, Westram, & Robinson, 2012).

Despite its strong operational potential, the $C_{INDEX}$ does not come without some limitations. First of all, it is important to remember that this index is a measure of genetic connectivity, not demographic connectivity (Lowe & Allendorf, 2010), and thus cannot directly provide any counting of the actual number of crossing events. If immigrants do not reproduce, the actual crossing of dozens of individuals, although suggesting high permeability, might not translate into high migration rates, resulting in low $C_{INDEX}$ values (Figure 3). Furthermore, a migration rate has to be interpreted in regard of effective population size: a migration rate of 0.05 actually corresponds to 2.5 effective dispersal events per generation in populations of size 50, but to 50 effective dispersal events in populations of size 1000 (Appendix S8). Despite higher permeability in the latter case, the contribution of these 50 immigrants to local recruitment might still be limited when compared to the contribution of residents (Lowe & Allendorf, 2010). Although this is more of an inherent characteristic of the index than a real limitation, it is important to keep this specificity in mind when interpreting it.

Secondly, the computation of the $C_{INDEX}$ relies on the assumption that, beyond the background signal of genetic differentiation that is expected under the sole influences of genetic drift, mutations and random sampling ($GD_{min}$), the observed measures of genetic differentiation $GD_{obs}$ only stem from dam-induced fragmentation. This assumption is only valid when sampling adjacent populations, located at the immediate upstream and downstream vicinity of the considered obstacle (Figure 1). It

thus implies the exclusion of migratory fish species, though at the heart of great conservation issues (e.g., Junge et al., 2014; Klütsch et al., 2019): complex life-cycles such as anadromy ("river-sea-river" migrations), catadromy ("sea-river-sea" migrations) or potamodromy ("river-lake-river" migrations) indeed preclude the delineation of upstream and downstream populations and do not allow proper estimates for the $C_{INDEX}$. In non-migratory fish species, this assumption also limits the use of the $C_{INDEX}$ in large-scale studies, in which the distance between the upstream and the downstream sampling sites lies beyond the dispersal capacities of the studied species. It certainly leaves room for manoeuvre, as illustrated with the empirical dataset from Gouskov et al. (2016): we could for instance select pairs of populations located up to 16km apart, but this was only possible because of the high mobility of chubs, and performed in an illustrative purpose. In low-mobility species for instance, a non-adjacent sampling might bias the $C_{INDEX}$ downwards and hence overestimate the measure of fragmentation, as observed measures of genetic differentiation would result from dam-induced fragmentation but also from other processes such as Isolation-by-Distance (e.g. Coleman et al., 2018). We thus strongly encourage practitioners to consider an adjacent sampling design as often as possible, although we readily acknowledge that this may not always be an easy task given safety and accessibility considerations. Furthermore, fish may not always be found in the direct vicinity of obstacles. For instance, the conversion of a river into a reservoir after the creation of a dam often leads to major habitat modification and shifts in species composition (Bednarek, 2001), which can force adapting the sampling design. A solution might be to capture the resultant background signal of genetic differentiation by simulating $GD_{min}$ values under various scenarios of isolation (Isolation-by-Distance, Isolation-by-Resistance, etc.; McRae, 2006) in a way similar to the simulation of $GD_{max}$ values in this study (Figure 2). We yet believe that the variety, the complexity and the specificity of such scenarios would preclude the computation of standardized $C_{INDEX}$ scores, comparable among obstacles, species and studies. Although it might in some instances be considered a technical constraint, we argue that only a strict adjacent sampling design can warrant unbiased and reliable $C_{INDEX}$ estimates.

Finally, the proposed $C_{INDEX}$ does not take into account the possible asymmetric gene flow created by barriers, as fish might struggle or even fail to ascent an obstacle (sometimes despite the presence of

dedicated fishpasses; Silva et al., 2018) whereas dam discharge might on the contrary further increase or even force downstream movements (Pracheil et al., 2015). Although quantifying the asymmetric permeability of obstacles appears of crude importance for informed conservation measures, the proposed $C_{INDEX}$ currently relies on the use of classical pairwise measures of genetic differentiation that assume symmetric gene flow. Future developments will be required to allow the $C_{INDEX}$ to provide unbiased and distinct standardized scores for both upstream and downstream barrier effects. In the meanwhile, it may be interesting to also assess the validity of the $C_{INDEX}$ in quantifying the effects of terrestrial obstacles, since asymmetric gene flow is not necessarily as pronounced as in river systems: provided that populations are sampled in the direct vicinity of the obstacle, the $C_{INDEX}$ might as well provide a standardized quantification of road-induced fragmentation.

## Conclusion

We here laid the groundwork for an operational tool dedicated to the individual and standardized quantification of the impact of artificial barriers on riverscape functional connectivity from measures of genetic differentiation. The proposed index of genetic connectivity $C_{INDEX}$ is designed to take into account the temporal inertia in the evolution of allelic frequencies resulting from the interplay between the age of the obstacle and the effective sizes of populations. Provided only adjacent populations are sampled, the $C_{INDEX}$ allows a rapid and thorough ranking of obstacles only a few generations after their creation. The $C_{INDEX}$ in its current form still suffers from some limitations, and it should be seen as the preliminary version of a future powerful bio-indicator of habitat fragmentation, rather than as an end-product. We call conservation and population geneticists to pursue the development of such an index, as we –as scientists– need to help managers resolve complex and urging social problems. Nonetheless, the $C_{INDEX}$ is robust, only requires a minimum amount of fieldwork and genotypic data and already solves several difficulties inherent to the study of dam-induced fragmentation in river systems, making it a promising tool for the restoration of riverscape connectivity. The $C_{INDEX}$ may allow practitioners to objectively identify obstacles that do not present any substantial conservation issue (from a

connectivity perspective) and help them target their efforts and resources towards the most impactful ones. Similarly, it may allow tracking the expected temporal decrease in genetic differentiation after obstacle removal or fishpass setting (Landguth et al., 2010; Schwartz, Luikart, & Waples, 2007) and thus help evaluate the success of local mitigations and restoration measures in response to regulatory obligations (Silva et al., 2018). Finally, it might as well provide a standardized quantification of road-induced fragmentation, a critical issue in terrestrial ecology.

## Acknowledgements

We warmly thank all the colleagues and students who helped with field sampling. We are also grateful to Dr. A. Gouskov and C. Vorburger for details about their data. This work has been financially supported by grants to SB from the Agence Française pour la Biodiversité and from the Région Occitanie (CONAQUAT).

## Supporting Information

Appendix S1: Details on the $C_{INDEX}$ formula.

Appendix S2: Effect of sample size on various measures of genetic differentiation.

Appendix S3: Effect of both mutation rate and time since barrier creation on various measures of genetic differentiation.

Appendix S4: Relationship between mean expected heterozygosity and mean effective population size.

Appendix S5: Distribution of mutation rates used in the second set of simulations.

Appendix S6: Localization of obstacles in the two empirical datasets.

Appendix S7: Relationship between observed and predicted measures of genetic differentiation.

Appendix S8: $C_{INDEX}$ responses to the increase in the number of effective dispersal events (*Nm*).

Appendix S9: Rational for the observed bias in $C_{INDEX}$ values in the very first generations after barrier creation.

Appendix S10: Walkthrough for the computation of the $C_{INDEX}$

Appendix S11: https://doi.org/10.6084/m9.figshare.9698879.v1 (R-objects for the computation of the $C_{INDEX}$)

# REFERENCES


Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, *5*(2), 181-190. (WOS:A1986C753400010).

Alp, M., Keller, I., Westram, A. M., & Robinson, C. T. (2012). How river structure and biological traits influence gene flow : A population genetic study of two stream invertebrates with differing dispersal abilities: Biological traits and gene flow in stream invertebrates. *Freshwater Biology*, *57*(5), 969-981. doi: 10.1111/j.1365-2427.2012.02758.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). doi: 10.18637/jss.v067.i01

Baudoin, J.-M., Burgun, V., Chanseau, M., Larinier, M., Ovidio, M., Sremski, W., … Voegtle, B. (2014). *The ICE protocol for ecological continuity. Assessing the passage of obstacles by fish. Concepts, design and application*. Paris: Onema.

Bednarek, A. T. (2001). Undamming Rivers : A Review of the Ecological Impacts of Dam Removal. *Environmental Management*, *27*(6), 803-814. doi: 10.1007/s002670010189

Birnie-Gauvin, K., Aarestrup, K., Riis, T. M. O., Jepsen, N., & Koed, A. (2017). Shining a light on the loss of rheophilic fish habitat in lowland rivers as a forgotten consequence of barriers, and its implications for management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(6), 1345-1349. doi: 10.1002/aqc.2795

Blanchet, S., Rey, O., Etienne, R., Lek, S., & Loot, G. (2010). Species-specific responses to landscape fragmentation : Implications for management strategies. *Evolutionary Applications*, *3*(3), 291-304. doi: 10.1111/j.1752-4571.2009.00110.x

Bowcock, A. M., Ruizlinares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavallisforza, L. L. (1994). High-resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, *368*(6470), 455-457. (ISI:A1994ND12000063).

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. (WOS:000170489900001).

Broquet, T., & Petit, E. J. (2009). Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 193-216. doi: 10.1146/annurev.ecolsys.110308.120324

Cavalli-Sforza, L. L., & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *American journal of human genetics*, *19*, 233.

Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J.-S., Clobert, J., Besnard, A., & Bernatchez, L. (2018). Demographic and genetic approaches to study dispersal in wild animal populations : A methodological review. *Molecular Ecology*, *27*(20), 3976-4010. doi: 10.1111/mec.14848

Coleman, R. A., Gauffre, B., Pavlova, A., Beheregaray, L. B., Kearns, J., Lyon, J., … Sunnucks, P. (2018). Artificial barriers prevent genetic recovery of small isolated populations of a low-mobility freshwater fish. *Heredity*. doi: 10.1038/s41437-017-0008-3

Cooke, S. J., & Hinch, S. G. (2013). Improving the reliability of fishway attraction and passage efficiency estimates to inform fishway engineering, science, and practice. *Ecological Engineering*, *58*, 123-132. doi: 10.1016/j.ecoleng.2013.06.005

Couto, T. B., & Olden, J. D. (2018). Global proliferation of small hydropower plants—Science and policy. *Frontiers in Ecology and the Environment*. doi: 10.1002/fee.1746

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., … Sullivan, C. A. (2006). Freshwater biodiversity : Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), 163-182. doi: 10.1017/S1464793105006950

Fredrich, F., Ohmann, S., Curio, B., & Kirschbaum, F. (2003). Spawning migrations of the chub in the River Spree, Germany. *Journal of Fish Biology*, *63*(3), 710-723. doi: 10.1046/j.1095-8649.2003.00184.x

Gauffre, B., Estoup, A., Bretagnolle, V., & Cosson, J. F. (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, *17*(21), 4619-4629. doi: 10.1111/j.1365-294X.2008.03950.x

Gibson, L., Wilman, E. N., & Laurance, W. F. (2017). How Green is 'Green' Energy? *Trends in Ecology & Evolution*, *32*(12), 922-935. doi: 10.1016/j.tree.2017.09.007

Goudet, J. (1995). FSTAT (Version 1.2) : A Computer Program to Calculate F-Statistics. *Journal of Heredity*, *86*(6), 485-486. doi: 10.1093/oxfordjournals.jhered.a111627

Gouskov, A., Reyes, M., Wirthner-Bitterlin, L., & Vorburger, C. (2016). Fish population genetic structure shaped by hydroelectric power plants in the upper Rhine catchment. *Evolutionary Applications*, *9*(2), 394-408. doi: 10.1111/eva.12339

Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, *116*(1), 92-98. doi: 10.1038/hdy.2015.76

Hand, B. K., Flint, C. G., Frissell, C. A., Muhlfeld, C. C., Devlin, S. P., Kennedy, B. P., … Stanford, J. A. (2018). A social-ecological perspective for riverscape management in the Columbia River Basin. *Frontiers in Ecology and the Environment*, *16*(S1), S23-S33. doi: 10.1002/fee.1752

Hawkins, P. R., Hortle, K. G., Phommanivong, S., & Singsua, Y. (2018). Underwater video monitoring of fish passage in the Mekong River at Sadam Channel, Khone Falls, Laos. *River Research and Applications*, *34*(3), 232-243. doi: 10.1002/rra.3239

Hedrick, P. W. (2005). A Standardized Genetic Differentiation Measure. *Evolution*, *59*(8), 1633-1638. doi: 10.1111/j.0014-3820.2005.tb01814.x

Jansson, R., Nilsson, C., & Malmqvist, B. (2007). Restoring freshwater ecosystems in riverine landscapes : The roles of connectivity and recovery processes. *Freshwater Biology*, *52*(4), 589-596. doi: 10.1111/j.1365-2427.2007.01737.x

Januchowski-Hartley, S. R., Diebel, M., Doran, P. J., & McIntyre, P. B. (2014). Predicting road culvert passability for migratory fishes. *Diversity and Distributions*, *20*(12), 1414-1424. doi: 10.1111/ddi.12248

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components : A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. doi: 10.1186/1471-2156-11-94

Jost, L. (2008). $G_{ST}$ and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015-4026. doi: 10.1111/j.1365-294X.2008.03887.x

Junge, C., Museth, J., Hindar, K., Kraabøl, M., & Vøllestad, L. A. (2014). Assessing the consequences of habitat fragmentation for two migratory salmonid fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *24*(3), 297-311. doi: 10.1002/aqc.2391

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Klütsch, C. F. C., Maduna, S. N., Polikarpova, N., Forfang, K., Aspholm, P. E., Nyman, T., … Hagen, S. B. (2019). Genetic changes caused by restocking and hydroelectric dams in demographically bottlenecked brown trout in a transnational subarctic riverine system. *Ecology and Evolution*, ece3.5191. doi: 10.1002/ece3.5191

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKELVEY, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179-4191. doi: 10.1111/j.1365-294X.2010.04808.x

Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites : Genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, *11*(12), 2453-2465. doi: 10.1046/j.1365-294X.2002.01643.x

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18-22.

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, *19*(15), 3038-3051. doi: 10.1111/j.1365-294X.2010.04688.x

McLaughlin, R. L., Smyth, E. R. B., Castro-Santos, T., Jones, M. L., Koops, M. A., Pratt, T. C., & Vélez-Espino, L.-A. (2013). Unintended consequences and trade-offs of fish passage. *Fish and Fisheries*, *14*(4), 580-604. doi: 10.1111/faf.12003

McNeish, D. M. (2014). *Analyzing Clustered Data with OLS Regression : The Effect of a Hierarchical Data Structure*. *40*, 6.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561.

Meirmans, P. G. (2006). Using the Amova Framework to Estimate a Standardized Genetic Differentiation Measure. *Evolution*, *60*(11), 2399-2402. doi: 10.1111/j.0014-3820.2006.tb01874.x

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure : FST and related measures. *Molecular Ecology Resources*, *11*(1), 5-18. doi: 10.1111/j.1755-0998.2010.02927.x

Neuenschwander, S., Michaud, F., & Goudet, J. (2019). QuantiNemo 2 : A Swiss knife to simulate complex demographic and genetic scenarios, forward and backward in time. *Bioinformatics*, *35*(5), 886-888. doi: 10.1093/bioinformatics/bty737

Nilsson, C. (2005). Fragmentation and Flow Regulation of the World's Large River Systems. *Science*, *308*(5720), 405-408. doi: 10.1126/science.1107887

Paz-Vinas, I., Comte, L., Chevalier, M., Dubut, V., Veyssiere, C., Grenouillet, G., … Blanchet, S. (2013). Combining genetic and demographic data for prioritizing conservation actions : Insights from a threatened fish species. *Ecology and Evolution*, *3*(8), 2696-2710. doi: 10.1002/ece3.645

Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G., & Blanchet, S. (2013). The demographic history of populations experiencing asymmetric gene flow : Combining simulated and empirical data. *Molecular Ecology*, *22*(12), 3279-3291. doi: 10.1111/mec.12321

Poff, N. L., & Schmidt, J. C. (2016). How dams can go with the flow. *Science (New York, N.Y.)*, *353*(6304), 1099-1100. doi: 10.1126/science.aah4926

Poulet, N. (2007). Impact of weirs on fish communities in a piedmont stream. *River Research and Applications*, *23*(9), 1038-1047. doi: 10.1002/rra.1040

Pracheil, B. M., Mestl, G. E., & Pegg, M. A. (2015). Movement through Dams Facilitates Population Connectivity in a Large River. *River Research and Applications*, *31*(5), 517-525. doi: 10.1002/rra.2751

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959. (WOS:000087475100039).

Prunier, J. G., Dubut, V., Chikhi, L., & Blanchet, S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. *Methods in Ecology and Evolution*, *8*(12), 1866-1877. doi: 10.1111/2041-210X.12820

Prunier, J. G., Dubut, V., Loot, G., Tudesque, L., & Blanchet, S. (2018). The relative contribution of river network structure and anthropogenic stressors to spatial patterns of genetic diversity in two freshwater fishes : A multiple-stressors approach. *Freshwater Biology*, *63*(1), 6-21. doi: 10.1111/fwb.13034

R Development Core Team. (2014). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. Consulté à l'adresse http://www.R-project.org

Raeymaekers, J. A. M., Raeymaekers, D., Koizumi, I., Geldof, S., & Volckaert, F. A. M. (2009). Guidelines for restoring connectivity around water mills : A population genetic approach to the management of riverine fish. *Journal of Applied Ecology*, *46*(3), 562-571. doi: 10.1111/j.1365-2664.2009.01652.x

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., … Cooke, S. J. (2018). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*. doi: 10.1111/brv.12480

Richardson, J. L., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular ecology*, *25*(4), 849–863.

Rousset, F. (2008). GENEPOP ´007 : A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103-106. (WOS:000253827100016).

Roy, S. G., Uchida, E., de Souza, S. P., Blachly, B., Fox, E., Gardner, K., … Hart, D. (2018). A multiscale approach to balance trade-offs among dam infrastructure, river restoration, and cost. *Proceedings of the National Academy of Sciences*, *115*(47), 12069-12074. doi: 10.1073/pnas.1807437115

Saint-Pé, K., Blanchet, S., Tissot, L., Poulet, N., Plasseraud, O., Loot, G., … Prunier, J. G. (2018). Genetic admixture between captive-bred and wild individuals affects patterns of dispersal in a brown

trout (*Salmo trutta*) population. *Conservation Genetics*, *19*(5), 1269-1279. doi: 10.1007/s10592-018-1095-2

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, *109*(6), 365-371. doi: 10.1007/s004120000089

Schwartz, M., Luikart, G., & Waples, R. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25-33. doi: 10.1016/j.tree.2006.08.009

Selkoe, K. A., Scribner, K. T., & Galindo, H. M. (2015). Waterscape Genetics—Applications of Landscape Genetics to Rivers, Lakes, and Seas. In N. Balkenhol, S. A. Cushman, A. T. Storfer, & L. P. Waits (Éd.), *Landscape Genetics* (p. 220-246). doi: 10.1002/9781118525258.ch13

Silva, A. T., Lucas, M. C., Castro-Santos, T., Katopodis, C., Baumgartner, L. J., Thiem, J. D., … Cooke, S. J. (2018). The future of fish passage science, engineering, and practice. *Fish and Fisheries*, *19*(2), 340-362. doi: 10.1111/faf.12258

Song, C., Omalley, A., Roy, S. G., Barber, B. L., Zydlewski, J., & Mo, W. (2019). Managing dams for energy and fish tradeoffs : What does a win-win solution take? *Science of The Total Environment*, *669*, 833-843. doi: 10.1016/j.scitotenv.2019.03.042

Städele, V., & Vigilant, L. (2016). Strategies for determining kinship in wild populations using genetic data. *Ecology and Evolution*, *6*(17), 6107-6120. doi: 10.1002/ece3.2346

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics : Where are we now? *Molecular Ecology*, *19*(17), 3496-3514. doi: 10.1111/j.1365-294X.2010.04691.x

Turgeon, K., Turpin, C., & Gregory-Eaves, I. (2019). Dams have varying impacts on fish communities across latitudes : A quantitative synthesis. *Ecology Letters*, ele.13283. doi: 10.1111/ele.13283

Wang, J. L. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *360*(1459), 1395-1409. (WOS:000231317800005).

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, *38*(6), 1358-1370. (WOS:A1984TY40400017).

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, *163*(3), 1177-1191. (WOS:000182046900029).

Winter, D. J. (2012). MMOD : An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, *12*(6), 1158-1160. doi: 10.1111/j.1755-0998.2012.03174.x

Yue, G. H., David, L., & Orban, L. (2007). Mutation rate and pattern of microsatellites in common carp (Cyprinus carpio L.). *Genetica*, *129*(3), 329-331. doi: 10.1007/s10709-006-0003-8

# Box

**Box 1 : Guidelines for the use and the interpretation of the $C_{INDEX}$.**

The $C_{INDEX}$ allows an individual and standardized quantification of the impact of artificial barriers on riverscape functional connectivity from snapshot measures of genetic differentiation. Here, we provide a guideline for practitioners:

- **Species**: any freshwater species whose local effective population sizes are lower than 1000 can be considered.
- **Obstacle:** any obstacle whose age corresponds to a minimum of 10-15 generations and a maximum of 600 generations for the studied species can be considered.
- **Sampling:** populations are sampled in the immediate upstream and downstream vicinity of the obstacle, with a minimum of 20-30 individuals per populations.
- **Genetic data**: Individual genotypes are based on a set of polymorphic microsatellite markers.
- **Computation:** The $C_{INDEX}$ is computed in R thanks to a user-friendly script provided in Appendix S11 (see also Appendix S10 for a walkthrough).
- **Interpretation for $C_{INDEX}$ < 10%:** A $C_{INDEX}$ value lower than 10% (or whose 95% confidence interval includes the 10% threshold) indicates no gene flow between populations (total barrier effect), whatever the age of the obstacle of the effective size of populations.
- **Interpretation for $C_{INDEX}$ > 90%:** A $C_{INDEX}$ value higher than 90% (or whose 95% confidence interval includes the 90% threshold) indicates full genetic connectivity (no barrier effect), whatever the age of the obstacle of the effective size of populations.
- **Interpretation for intermediate $C_{INDEX}$ values:** Intermediate $C_{INDEX}$ values can be used to rank obstacles according to their barrier effect. However, for $C_{INDEX}$ values ranging from ~15 to ~70%, the $C_{INDEX}$ tends to slightly increase with both the increase in time since barrier creation and the increase in effective population sizes (Figure 3I-J).
- Obstacles with $C_{INDEX}$ values that do not differ by more than 15 to 20% but that are characterized by very different ages and / or population sizes (as indicated by expected heterozygosity) should be

considered as possibly having comparable barrier effects, except of course if the ranking of obstacles based on $C_{INDEX}$ values goes against these trends. Consider for instance an obstacle A of age 20 (in generations) and an obstacle B of age 300. If $C_{INDEX}$(A) = 20% and $C_{INDEX}$(B) = 40%, both obstacles should be considered as possibly having the same impact on gene flow. On the contrary, if $C_{INDEX}$(A) = 40% and $C_{INDEX}$(B) = 20%, obstacle B can be confidently considered as more impactful than obstacle A.

# Table

Table 1. Main characteristics and results for the obstacles selected from empirical datasets (Original publication: (1) Gouskov et al., 2016; (2) Prunier et al., 2018). For each obstacle, the table indicates the name of the river, the date of creation, the distance between upstream and downstream sampled populations, the considered species (*Sc*: *Squalius cephalus*; *Go* : *Gobio occitaniae*; *Pp*: *Phoxinus phoxinus*), the number of generations elapsed since barrier creation, the mean expected heterozygosity (*He*), and the computed $C_{INDEX}$ along with its 95% confidence inverval. In bold, obstacles that were found as significant barriers to gene flow in the considered species.

| River | Obstacle | Creation date | Upstream-Downstream distance (km) | Species | Number of elapsed generations | He | $C_{INDEX}$ | 95%CI | Original publication |
|---|---|---|---|---|---|---|---|---|---|
| Rhine | Barr05 | 1966 | 12.02 | *Sc* | 14.67 | 0.76 | 85.50 | 16.53 | (1) |
| Rhine | Barr06 | 1914 | 8.28 | *Sc* | 32.00 | 0.75 | 100 | 0 | (1) |
| Rhine | Barr07 | 1933 | 7.19 | *Sc* | 25.67 | 0.75 | 100 | 0 | (1) |
| Rhine | Barr08 | 1941 | 12.86 | *Sc* | 23.00 | 0.73 | 88.51 | 2.05 | (1) |
| Rhine | Barr09 | 1920 | 6.38 | *Sc* | 30.00 | 0.73 | 90.02 | 4.31 | (1) |
| Rhine | Barr10 | 1956 | 12.42 | *Sc* | 18.00 | 0.69 | 100 | 0 | (1) |
| **Rhine** | **Barr11** | **1964** | **4.79** | ***Sc*** | 15.33 | 0.69 | 74.98 | 13.82 | (1) |
| **Aar** | **Barr13** | **1902** | **1.91** | ***Sc*** | 36.00 | 0.76 | 64.21 | 8.21 | (1) |
| **Aar** | **Barr14** | **1953** | **10.75** | ***Sc*** | 19.00 | 0.76 | 63.59 | 5.58 | (1) |
| Aar | Barr15 | 1945 | 7.50 | *Sc* | 21.67 | 0.75 | 86.49 | 15.39 | (1) |
| Aar | Barr16 | 1929 | 5.61 | *Sc* | 27.00 | 0.77 | 100 | 0 | (1) |
| Aar | Barr17 | 1893 | 3.14 | *Sc* | 39.00 | 0.77 | 100 | 0 | (1) |
| Aar | Barr18 | 1917 | 15.55 | *Sc* | 31.00 | 0.76 | 96.60 | 4.09 | (1) |
| **Aar** | **Barr19** | **1896** | 1.93 | *Sc* | 38.00 | 0.76 | 74.49 | 4.06 | (1) |
| **Aar** | **Barr20** | **1896** | 6.82 | *Sc* | 38.00 | 0.75 | 69.41 | 3.90 | (1) |
| Aar | Barr21 | 1970 | 6.00 | *Sc* | 13.33 | 0.76 | 100 | 0 | (1) |
| Aar | Barr22 | 1970 | 9.51 | *Sc* | 13.33 | 0.77 | 100 | 0 | (1) |
| Aar | Barr23 | 1939 | 9.22 | *Sc* | 23.67 | 0.76 | 100 | 0 | (1) |
| Aar | Barr24 | 1900 | 11.67 | *Sc* | 36.67 | 0.76 | 100 | 0 | (1) |
| Aar | Barr25 | 1968 | 5.81 | *Sc* | 14.00 | 0.76 | 100 | 0 | (1) |
| Aar | Barr26 | 1963 | 4.84 | *Sc* | 15.67 | 0.75 | 100 | 0 | (1) |
| **Limmat** | **Barr32** | **1933** | **16.08** | ***Sc*** | **25.67** | **0.73** | **71.30** | **2.10** | **(1)** |
| Limmat | Barr33 | 1933 | 3.24 | *Sc* | 25.67 | 0.72 | 100 | 0 | (1) |
| **Célé** | **CLA** | **1500** | **0.18** | ***Go*** | **204** | **0.60** | **82.03** | **5.28** | **(2)** |
| Célé | SCA | 1500 | 0.09 | *Go* | 204 | 0.63 | 92.73 | 3.88 | (2) |
| **Célé** | **SCC** | **1960** | **0.2** | ***Go*** | **20** | **0.64** | **49.23** | **3.78** | **(2)** |
| Viaur | SEG | 1600 | 0.11 | *Go* | 164 | 0.58 | 100 | 0 | (2) |
| Viaur | CAM | 1600 | 0.49 | *Go* | 164 | 0.62 | 100 | 0 | (2) |
| **Viaur** | **CAP** | **1700** | **0.55** | ***Go*** | **124** | **0.61** | **83.85** | **1.33** | **(2)** |
| **Viaur** | **SJU** | **1800** | **1.07** | ***Go*** | **64** | **0.62** | **81.98** | **3.98** | **(2)** |
| Viaur | CIR | 1960 | 0.97 | *Go* | 20 | 0.62 | 86.72 | 15.10 | (2) |
| Célé | CLA | 1500 | 0.18 | *Pp* | 255 | 0.54 | 90.29 | 5.14 | (2) |
| Célé | SCA | 1500 | 0.09 | *Pp* | 255 | 0.57 | 100 | 0 | (2) |
| Célé | SCC | 1960 | 0.2 | *Pp* | 25 | 0.58 | 100 | 0 | (2) |
| Viaur | SEG | 1600 | 0.11 | *Pp* | 205 | 0.63 | 100 | 0 | (2) |
| Viaur | CAM | 1600 | 0.49 | *Pp* | 205 | 0.61 | 100 | 0 | (2) |
| Viaur | CAP | 1700 | 0.55 | *Pp* | 155 | 0.67 | 100 | 0 | (2) |
| Viaur | SJU | 1800 | 1.07 | *Pp* | 105 | 0.70 | 100 | 0 | (2) |
| Viaur | CIR | 1960 | 0.97 | *Pp* | 25 | 0.70 | 100 | 0 | (2) |
| Célé | CLA | 1500 | 0.18 | *Go-Pp* | / | / | 86.16 | 8.10 | (2) |
| Célé | SCA | 1500 | 0.09 | *Go-Pp* | / | / | 96.37 | 7.12 | (2) |
| Célé | SCC | 1960 | 0.2 | *Go-Pp* | / | / | 74.61 | 49.76 | (2) |
| Viaur | SEG | 1600 | 0.11 | *Go-Pp* | / | / | 100 | 0 | (2) |
| Viaur | CAM | 1600 | 0.49 | *Go-Pp* | / | / | 100 | 0 | (2) |
| Viaur | CAP | 1700 | 0.55 | *Go-Pp* | / | / | 91.93 | 15.83 | (2) |
| Viaur | SJU | 1800 | 1.07 | *Go-Pp* | / | / | 90.99 | 17.66 | (2) |
| Viaur | CIR | 1960 | 0.97 | *Go-Pp* | / | / | 93.36 | 13.02 | (2) |

# Figures

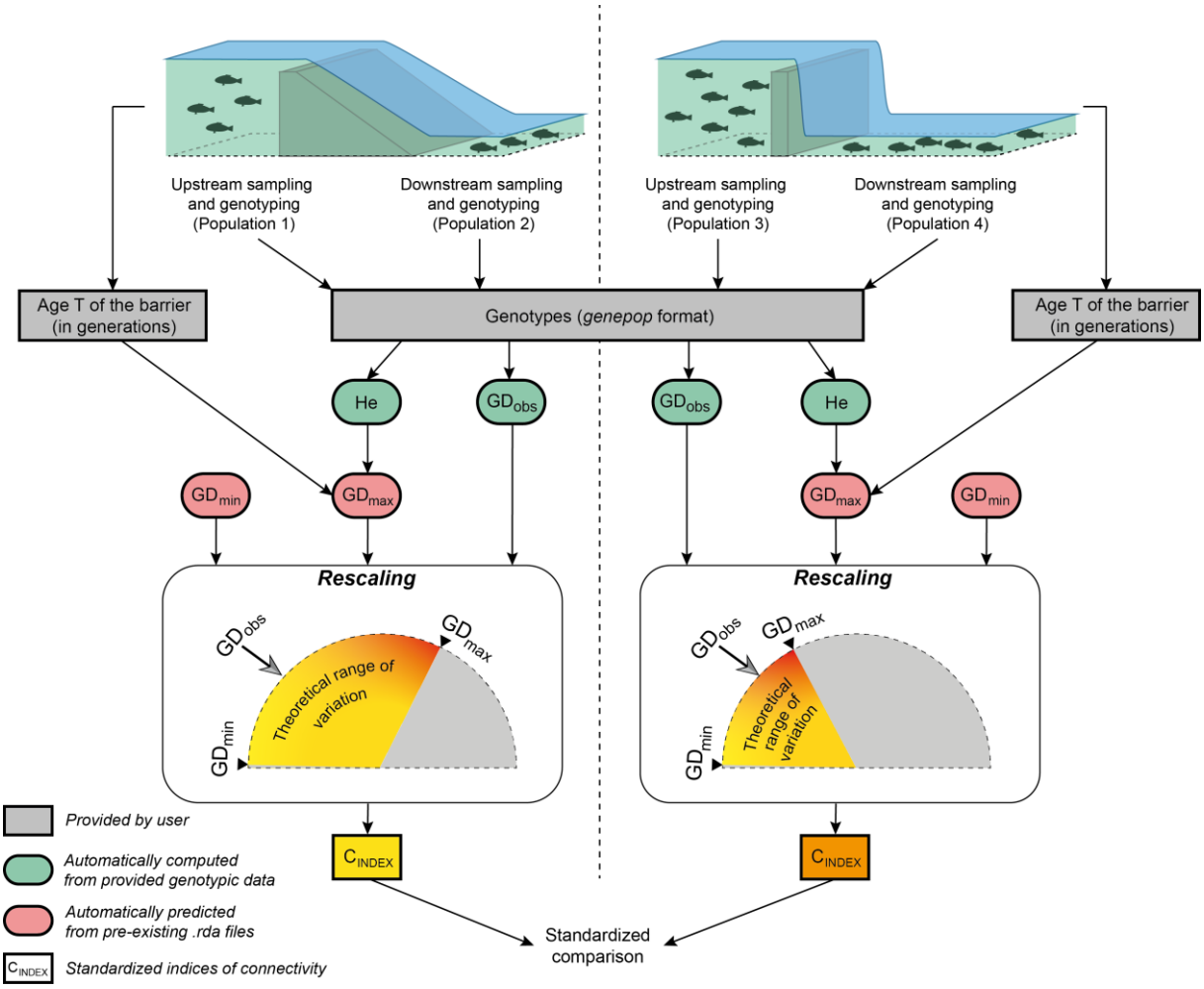


Figure 1. Flowchart illustrating the major steps in calculating the index of genetic connectivity for two independent obstacles. This flowchart refers to a user-friendly script provided in Appendix S11 (see also Appendix S10 for a walkthrough). After the sampling of populations located at the immediate upstream and downstream vicinity of each obstacle, users only have to provide a file of genotypes in the *genepop* format and a file of parameters indicating, for each obstacle, the names of the sampled populations and the number $T$ of generations elapsed since the creation of the obstacle. Observed measures of genetic differentiation $GD_{obs}$ and mean expected heterozygosity $He$ are automatically computed from provided genotypic data. $GD_{min}$ and $GD_{max}$ values, both delimiting the theoretical range of variation of $GD_{obs}$, are automatically predicted from pre-existing .rda files, $GD_{max}$ values depending on both $He$ and $T$. The computation of the index basically amounts to rescaling $GD_{obs}$

within its theoretical range (see main text for details), thus allowing standardized comparisons of the permeability of various obstacles, whatever their age, the considered species or the effective size of sampled populations.

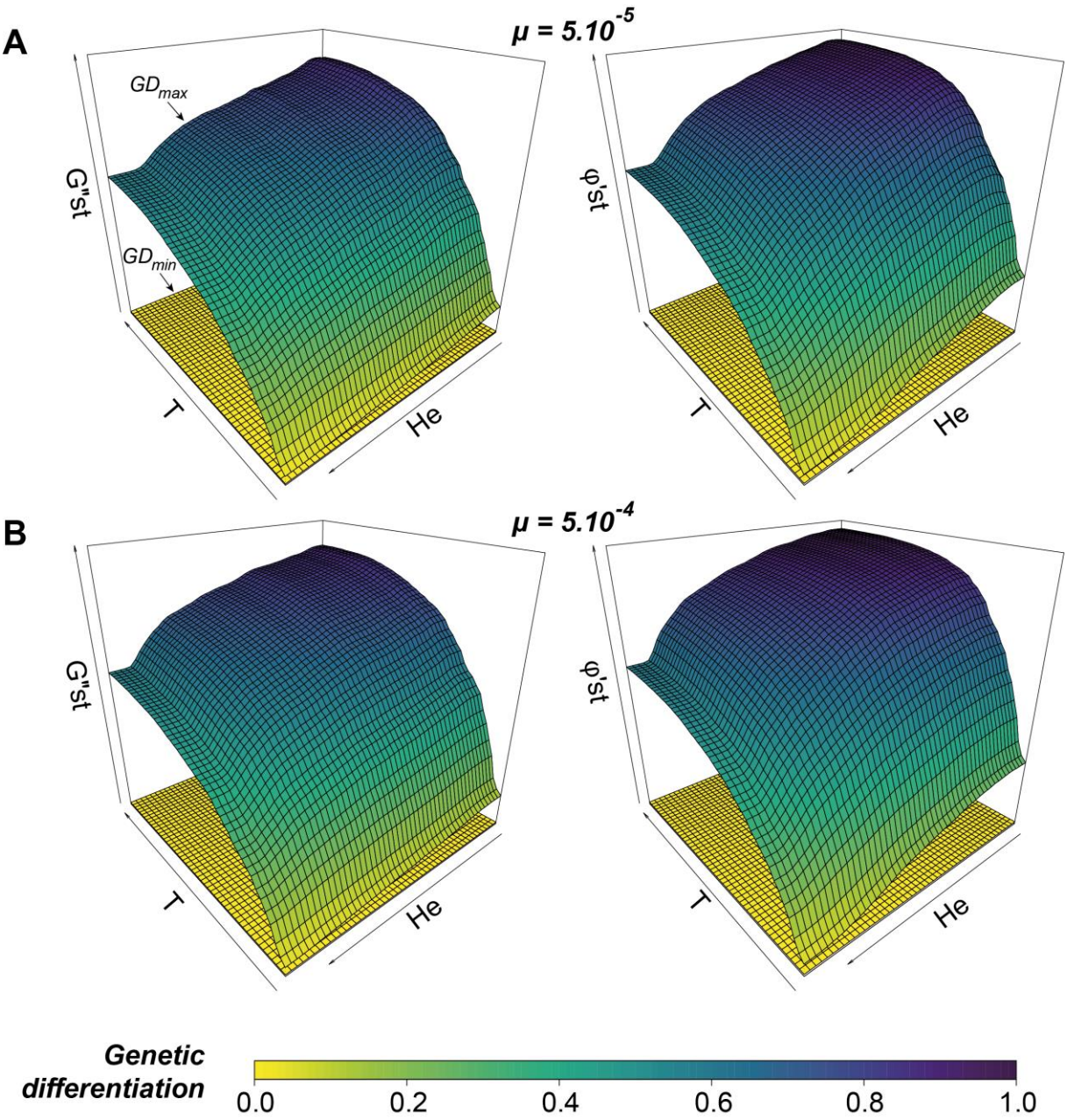


Figure 2. For each mutation rate (panels A and B) and each metric of genetic differentiation (G''st on the left and φ'st on the right), predicted $GD_{max}$ variations across the parameter space defined by the time $T$ elapsed since total barrier creation (from 0 to 600 generations) and the averaged expected heterozygosity (*He*, a proxy for effective population size, ranging from 0 to 0.93) for pairs of adjacent populations. $GD_{min}$ values are represented at the bottom of each graph. $GD_{min}$ and $GD_{max}$ surfaces together delimit the theoretical range of variation for any observed measure of genetic differentiation $GD_{obs}$.

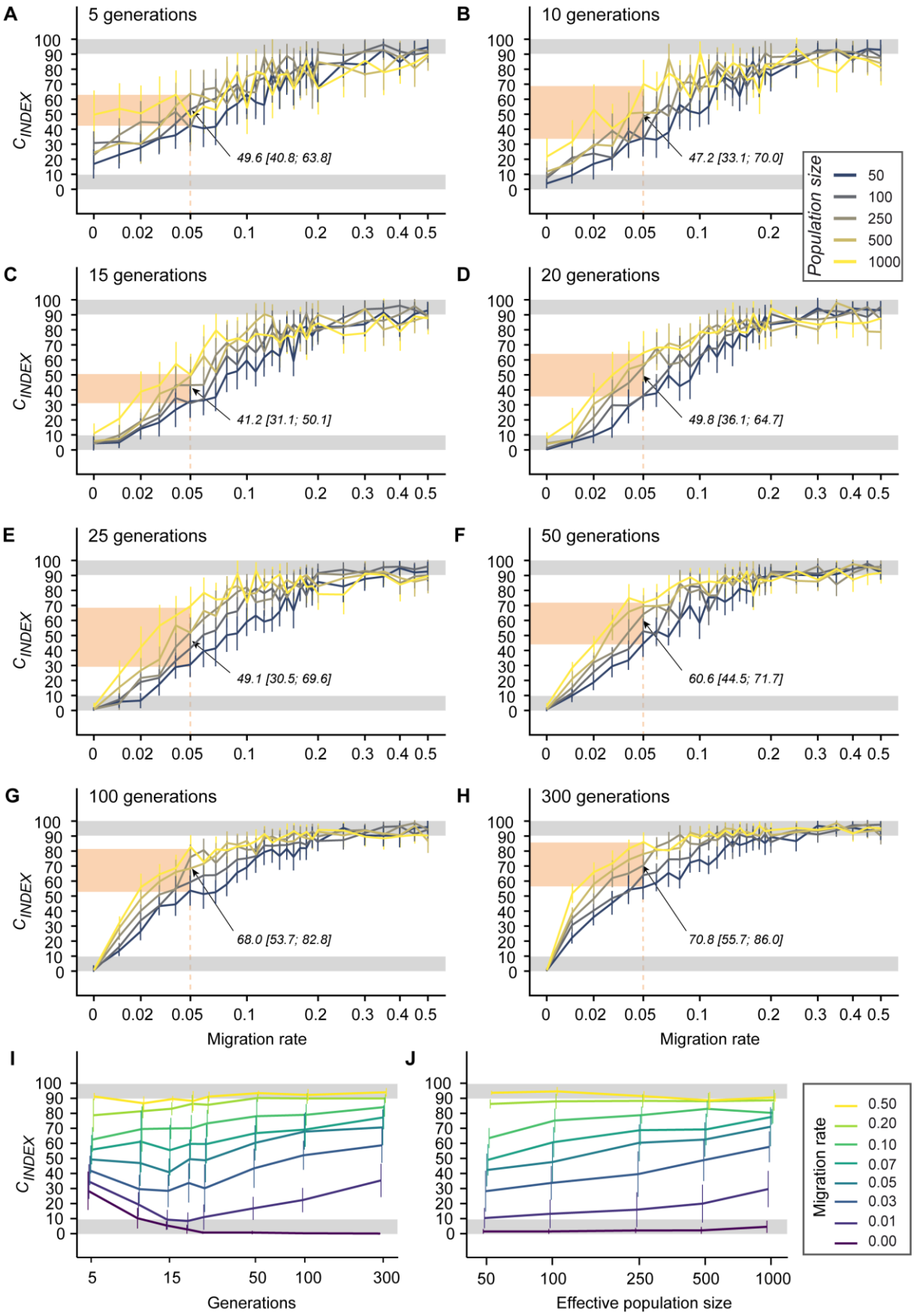
A
5 generations
49.6 [40.8; 63.8]
B
10 generations
47.2 [33.1; 70.0]
C
15 generations
41.2 [31.1; 50.1]
D
20 generations
49.8 [36.1; 64.7]
E
25 generations
49.1 [30.5; 69.6]
F
50 generations
60.6 [44.5; 71.7]
G
100 generations
68.0 [53.7; 82.8]
H
300 generations
70.8 [55.7; 86.0]
C INDEX
Migration rate
Population size
50
100
250
500
1000
I
J
Generations
Effective population size
Migration rate
0.50
0.20
0.10
0.07
0.05
0.03
0.01
0.00

Figure 3. Panels A-H: $C_{INDEX}$ responses to the increase in migration rate (*m*, on a logarithmic scale) for five different population sizes *Ne* (colored lines) and from 5 to 300 generations after barrier creation (panels A to H). All $C_{INDEX}$ values were averaged over 20 simulated replicates and plotted with 95% confidence intervals. In each panel, the shaded pink surface and the black arrow indicate the range of variation of $C_{INDEX}$ values for a migration rate of 0.05 (mean value across effective population sizes, minimum and maximum values into brackets). Panel I-J: $C_{INDEX}$ responses to the increase in time since barrier creation (panel I) and to the increase in effective population size (panel J) for eight different migration rates *m* (colored lines). The mean $C_{INDEX}$ values computed over simulated replicates were here averaged over effective population sizes (panel I) or over generations (excluding generations $\leq$ 10; panel J) and plotted with standard deviations. In all panels, shaded grey areas represent the ranges of variations in which the monitored obstacle can be considered as acting as a total barrier to gene flow ($C_{INDEX} < 10\%$) or, on the contrary, as allowing full genetic connectivity ($C_{INDEX} > 90\%$).

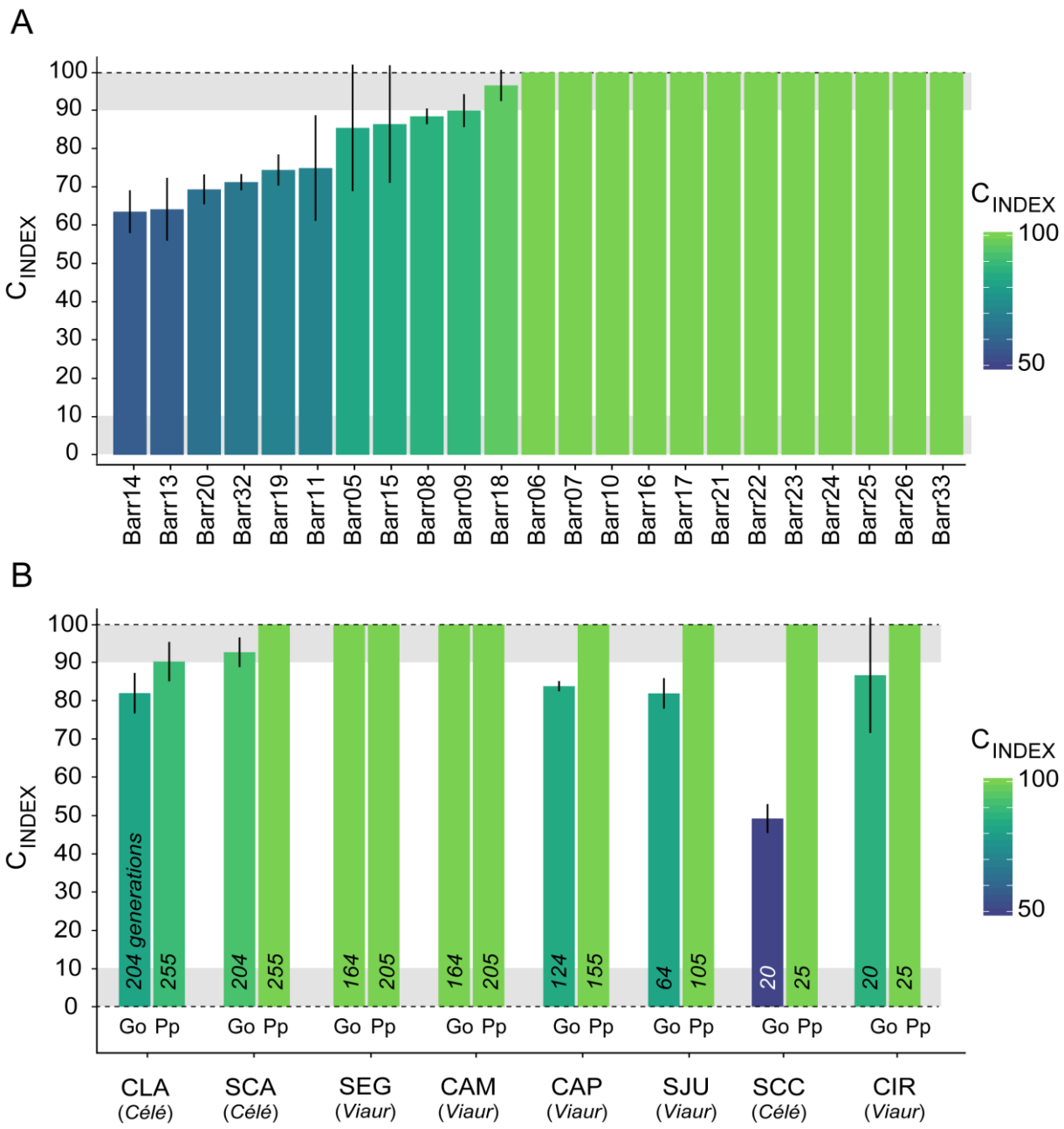


Figure 4. $C_{INDEX}$ values as computed from empirical datasets. Panel A: Results for the 23 dams selected from Gouskov et al. (2016), ranked according to their inferred impact on chubs' genetic connectivity (from the most impactful on the left to the less problematic ones on the right). Panel B: For each monitored species (Go: *Gobio occitaniae*; Pp: *Phoxinus phoxinus*), results for the 8 weirs selected from Prunier et al. (2018), ranked according to their date of creation (indicated by the number of generations elapsed since barrier creation in each species).