

1 The demographic history of populations experiencing asymmetric
2 gene flow: combining simulated and empirical data.

3 I. PAZ-VINAS,^{*,†,§} E. QUÉMÉRÉ,^{**} L. CHIKHI,^{†,§,#} G. LOOT,^{*,§} and
4 S. BLANCHET^{*,†}

5

6 ^{*} Centre National de la Recherche Scientifique (CNRS); Station d'Écologie Expérimentale du
7 CNRS à Moulis, USR 2936, F-09200 Moulis, France,

8 [†] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier, École
9 Nationale de Formation Agronomique (ENFA); UMR 5174 EDB (Laboratoire Évolution &
10 Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse cedex 4, France,

11 [§] Université de Toulouse, UPS, UMR 5174 (EDB), 118 route de Narbonne, F-31062
12 Toulouse cedex 4, France,

13 ^{**} Institut National de la Recherche Agronomique (INRA), UMR Comportement et Écologie
14 de la Faune Sauvage, INRA, BP 52627 Castanet-Tolosan cedex, F-31326, France,

15 [#] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, P-2780-156 Oeiras, Portugal.

16

17 **Keywords:** source-sink dynamics, demographic change, fish, rivers, ABC

18

19 **Corresponding author:**

20 Ivan PAZ-VINAS

21 Laboratoire Évolution et Diversité Biologique (EDB)

22 UMR 5174 (CNRS - UPS - ENFA)

23 118 route de Narbonne, F-31062 Toulouse cedex 4, France

24 Phone: (+33) 5 61 55 67 35

25 E-mail: ivanpaz23@gmail.com

26

27 **Running title:** Asymmetric gene flow and demographic inferences

28

29 **Abstract**

30 Population structure can significantly affect genetic-based demographic inferences,
31 generating spurious bottleneck-like signals. Previous studies have typically assumed island or
32 stepping-stone models, which are characterized by symmetric gene flow. However, many
33 organisms are characterized by asymmetric gene flow. Here, we combined simulated and
34 empirical data to test if asymmetric gene flow affects the inference of past demographic
35 changes. Through the analysis of simulated genetic data with three methods (i.e.,
36 BOTTLENECK, M-ratio and MSVAR), we demonstrated that asymmetric gene flow biases
37 past demographic changes. Most biases were towards spurious signals of expansion, albeit
38 their strength depended on values of effective population size and migration rate. It is
39 noteworthy that the spurious signals of demographic changes also depended on the statistical
40 approach underlying each of the three methods. To some extent, biases induced by
41 asymmetric gene flow were confirmed in an empirical multi-specific dataset involving four
42 freshwater fish species (*Squalius cephalus*, *Leuciscus burdigalensis*, *Gobio gobio*, *Phoxinus*
43 *phoxinus*). Indeed, all species exhibited signals of bottlenecks across two rivers for two out of
44 the three methods. This suggests that, although potentially biased by asymmetric gene flow,
45 these methods were able to bypass this bias when a bottleneck actually occurred. Our results
46 show that population structure and dispersal patterns have to be considered for proper
47 inference of demographic changes from genetic data.

48 **Introduction**

49 Inferring the demographic history of populations such as changes in effective
50 population size (contractions, expansions) is of prime importance for basic research and
51 conservation issues (Chikhi & Bruford 2005; Leblois *et al.* 2006). Several indirect methods
52 based on the analysis of neutral genetic variation have been developed to that aim (Cornuet &
53 Luikart 1996; Garza & Williamson 2001; Beaumont 1999; Storz & Beaumont 2002). These
54 methods have been largely used to assess the impact of environmental or anthropogenic
55 changes on the demographic history of endangered populations (e.g., Goossens *et al.* 2006;
56 Sousa *et al.* 2008).

57 However, inferring the demographic history of wild populations remains challenging.
58 Indeed, most methods assume that populations can be approximated by simple models such as
59 the Wright-Fisher model (Cornuet & Luikart 1996; Leblois *et al.* 2006). However, wild
60 populations rarely match these assumptions, since most of them are either spatially structured,
61 affected by external gene flow and/or at a non-equilibrium state (Hanski 1998; Broquet *et al.*
62 2010; Chikhi *et al.* 2010). Consequently, any deviations from these simple models may lead
63 to misinterpretations or incorrect inferences (Nielsen & Beaumont 2009; Städler *et al.* 2009;
64 Chikhi *et al.* 2010). Given that the development of inference methods based on complex
65 demographic models poses problems of its own, it is crucial to explore how existing inference
66 methods are robust to deviations from simple models assumptions (Leblois *et al.* 2006;
67 Städler *et al.* 2009; Chikhi *et al.* 2010). Recent programs based on the coalescent framework
68 (Kingman 1982) allow the simulation of genetic data under a wide variety of population
69 models (Hoban *et al.* 2012). Thus, specific simulated genetic datasets can be analyzed to test
70 the potential effects of particular population characteristics on the genetic inference of
71 populations' demographic history. Accordingly, population structure (Nielsen & Beaumont

72 2009; Städler *et al.* 2009; Chikhi *et al.* 2010; Peter *et al.* 2010), sampling scheme (Städler *et*
73 *al.* 2009; Chikhi *et al.* 2010), gene flow reductions (Broquet *et al.* 2010) and isolation-by-
74 distance (Leblois *et al.* 2006) have been identified as generators of false signals of
75 demographic change, with biases towards bottlenecks (e.g., Broquet *et al.* 2010; Chikhi *et al.*
76 2010) and, more rarely, towards expansions (e.g., Leblois *et al.* 2006).

77 A population characteristic that has rarely been considered to date in the context of
78 demographic history inferences is asymmetric gene flow. Differences in habitat quality, social
79 interactions or abiotic constraints (e.g., wind, oceanic currents, river flow or gravity)
80 frequently generate source-sink dynamics and impose asymmetric gene flow on natural
81 populations (Kawecki & Holt 2002). For instance, in riverine freshwater ecosystems,
82 organisms generally experience an inherent downstream-biased gene flow due to the
83 unidirectional water flow of rivers (Hänfling & Weetman 2006; Pollux *et al.* 2009). Such
84 asymmetry in gene flow drastically affects the genetic structure of wild riverine populations,
85 with, for instance, an accumulation of genetic diversity (e.g., number of alleles per locus)
86 downstream (i.e., sink populations, Kawecki & Holt 2002; Hänfling & Weetman 2006).

87 The demography of wild populations is dramatically affected by human pressures, and
88 notably by human-induced habitat fragmentation (Fahrig 2003; Henle *et al.* 2004). Freshwater
89 ecosystems are particularly affected by habitat fragmentation, either through the building of
90 hydroelectric dams or the presence of smaller obstacles like weirs (2 to 3 meters high,
91 Raeymaekers *et al.* 2008; Blanchet *et al.* 2010). In general, habitat fragmentation induces
92 changes in effective population size (N_e) that are theoretically inferable using the methods
93 described above. However, river fragmentation by dams and weirs may strongly affect the
94 movements of fishes, in both upstream and downstream directions. As a result, river
95 fragmentation can alter natural gene flow, either by exacerbating or, on the contrary, by

96 disrupting the natural asymmetric (i.e., downstream-biased) gene flow expected on such
97 ecosystems (Hänfling & Weetman 2006; Raeymaekers *et al.* 2008; but see Horreo *et al.*
98 2011). Although several studies have used coalescent- and frequency-based estimators of N_e
99 in fragmented rivers to infer effects of recent fragmentation (Alò & Turner 2005; Sousa *et al.*
100 2008; Nock *et al.* 2011), none of them have quantified how asymmetric gene flow might
101 affect the inference of past demographic changes that can be drawn from molecular markers
102 in such ecosystems.

103 In this paper, we explored both theoretically and empirically the potential problem
104 posed by asymmetric gene flow to infer temporal changes in N_e . First, we analyzed genetic
105 data simulated under a stationary linear stepping-stone model to test if asymmetric gene flow
106 can generate false signals of demographic changes. This was done using three methods widely
107 used to infer demographic changes: those implemented in the programs BOTTLENECK
108 (Cornuet & Luikart 1996; Piry *et al.* 1999) and MSVAR 1.3 (Beaumont *et al.* 1999; Storz &
109 Beaumont 2002), and the M-ratio method (Garza & Williamson 2001). Second, we used the
110 same three methods to analyze empirical data involving four freshwater fish species (*Squalius*
111 *cephalus*, *Leuciscus burdigalensis*, *Gobio gobio*, *Phoxinus phoxinus*) sampled in two rivers,
112 which differ by their level of anthropogenic fragmentation and asymmetric gene flow.

113

114 **Materials & methods**

115 Simulated data

116 To explore the consequences of asymmetric gene flow on the inference of changes in
117 N_e , we simulated genetic data under 27 different scenarios representing populations
118 experiencing symmetric or asymmetric gene flow but no changes in N_e , and then used this
119 data as input for three methods used to infer changes in N_e .

120

121 *The population genetics model.* We used the coalescent-based program ms along with the
122 microsat.exe program (Hudson 2002) to simulate genetic data under a strict Stepwise
123 Mutation Model (SMM). Specifically, we approximated a river, by considering a linear
124 stepping-stone population model composed of 10 demes (see Figure 1). All demes had the
125 same effective number of diploid individuals N , which remained constant across generations.
126 Each deme was characterized by three parameters: the scaled mutation rate $\theta = 4N\mu$, where μ
127 represents the neutral mutation rate *per locus*, and two scaled migration rates (M)
128 corresponding to the downstream- and upstream-directed gene flow: $M_{Downstream} = 4Nm$ and
129 $M_{Upstream} = \frac{M_{Downstream}}{a}$, where m is the migration rate and a is a parameter representing the
130 gene flow asymmetry (Figure 1). We used values of $a > 1$ to generate downstream-biased
131 gene flow. Deme 1 and deme 10 in Figure 1 can be considered as the most upstream and
132 downstream demes of the hypothetical river, respectively.

133 *Parameter estimation and exploration.* For all simulations, we assumed a unique neutral
134 mutation rate of $\mu=5.56 \times 10^{-4}$. This value corresponds to the average mutation rate calculated
135 for 49 microsatellite loci in the Cyprinid fish *Cyprinus carpio L.* (Yue *et al.* 2007). For
136 selecting values for all other model parameters (i.e., N , m and a , this combination of
137 parameters will hereafter be referred to ϕ), we first estimated values that best characterizes
138 riverine fish populations by performing ABC-regression analyses (i.e., approximate Bayesian
139 computation, Beaumont *et al.* 2002) based on observed summary statistics compiled for
140 several populations through a literature survey (Table S1). Specifically, we first obtained or
141 computed for sixteen riverine fish populations from fourteen rivers (i) the mean allelic
142 richness *per population* (AR), and (ii) the Pearson's correlation coefficient (r) between the

143 mean AR per sampling location and the distance of each sampling location from the river
144 source. Significant positive correlations between AR and distance from the river source are
145 characteristic of river organisms that experience downstream-biased gene flow asymmetry
146 (Hänfling & Weetman 2006; Blanchet *et al.* 2010). In a second step, we generated a total of
147 1,328,784 different genetic datasets under the population genetics model described above, by
148 drawing values for ϕ from grids, as in Weiss & von Haeseler (1998; see Figure S1). As noted
149 by Beaumont *et al.* (2002), grids of parameters can be seen as uniform priors. For each
150 genetic dataset, fifteen independent microsatellite loci were simulated, and a total of 22
151 diploid individuals were sampled for each deme. As for the literature survey populations, two
152 summary statistics (AR and r) were computed for each simulated dataset.

153 Next, we applied an ABC-regression algorithm (Beaumont *et al.* 2002) to each
154 surveyed population independently, by using the R package "abc" (Csillery *et al.* 2012). For
155 each ABC analysis, we retained 1% of the simulations whose summary statistics were the
156 closest from those calculated for the surveyed population. Imperfect matching between
157 observed and simulated data was corrected by using a local linear regression method
158 (Beaumont *et al.* 2002; Csillery *et al.* 2012). We estimated the median values of ϕ from the
159 corrected posterior distributions of ϕ for each population (see Table S1) and, finally, we
160 averaged these median values over all surveyed populations to obtain a first set of ϕ values:
161 $N=3147$, $m=0.053$ and $a=7.5$ (Table S1). We assumed that this set of ϕ values approximately
162 characterizes riverine fish populations. Then, to explore and generalize the effects of varying
163 N , m and a on the inference of changes in N_e , we explored two additional values per
164 parameter (leading to exploring $N=\{50, 500, 3147\}$, $m=\{0.01, 0.053, 0.1\}$ and $a=\{1, 7.5,$
165 $50\}$), and crossed all parameter values in a full-factorial design so as to generate genetic data
166 under 27 different scenarios. An asymmetry of $a=50$ is probably unrealistic, but the goal here

167 was to explore the effect of asymmetry in extreme conditions so as to explore how it differs
168 from a more realistic scenario (i.e., $a=7.5$). These scenarios were used to generate input
169 genetic data for further demographic history analyses (see § *Demographic history inference*).

170

171 Empirical data

172 *Biological models.* The four fish species considered here are all of the family Cyprinidae,
173 belong to the same trophic level (i.e., they are essentially insectivorous) and differ principally
174 in their maximum body length and dispersal abilities (Bolland *et al.* 2008; De Leeuw &
175 Winter 2008). *Squalius cephalus* (the European chub) and *Leuciscus burdigalensis* (the
176 rostrum dace) are two large-bodied fish (a maximum body length of 600 mm and 400 mm
177 respectively), whereas *Gobio gobio* (the gudgeon) and *Phoxinus phoxinus* (the European
178 minnow) are small-bodied fish (200 mm and 140 mm respectively).

179 *Study area.* Sampling was performed in two rivers that belong to the Adour-Garonne basin
180 drainage (South-western France): the Célé and the Viaur rivers (Figure S2). These rivers
181 present similar abiotic conditions but display differences concerning their level of
182 fragmentation. The Viaur River is highly fragmented with more than 50 small weirs (2-3
183 meters high, constructed within the last 800 years) and two recent hydroelectric dams (30
184 meters high, dating from 60 years ago, see Figure S2). We henceforth refer to this river as the
185 “highly fragmented river”. In the Célé River, ten-fifteen small weirs are found along the river
186 gradient. These were established over the last century and most of them are equipped with
187 fish ladders. The Célé River will be referred to as the “weakly fragmented river”. It is
188 noteworthy that asymmetric gene flow, effective population size and migration rate values
189 have been estimated for all these populations (i.e., a population here refers to a species within

190 a river system) through the ABC-regression algorithms presented above; these eight empirical
191 populations are characterized by a wide range of parameter values (see Table S1).

192 *Sampling design.* During summer 2006, a total of 10 and 11 sites were sampled on the Viaur
193 and Célé rivers respectively (Figure S2). We covered the entire upstream-downstream
194 gradient for both rivers to account for the entire genetic structure of the fish populations. At
195 each site, about 20 individuals *per* species were sampled by electric fishing. Small fragments
196 of pelvic fins were collected and preserved in 70% ethanol for later genetic analyses. *L.*
197 *burdigalensis* and *S. cephalus* were not found in all sampling sites, probably because the
198 habitat (notably temperature) is not favorable for these two species.

199 *Genetic data.* A salt-extraction protocol (Aljanabi & Martinez 1997) was performed to extract
200 genomic DNA from the pelvic fins of fishes. *Phoxinus phoxinus* and *Gobio gobio* were
201 genotyped at eight microsatellite loci, *Squalius cephalus* at ten loci and *Leuciscus*
202 *burdigalensis* at fifteen loci. Loci were amplified using multiplex PCRs and amplified
203 fragments were scored using the software GENEMAPPER® v.4.0 (Applied Biosystems,
204 Foster City, CA, USA). Neither departure from Hardy-Weinberg equilibrium nor null alleles
205 were detected for any of these loci (see Blanchet *et al.* 2010 for further details).

206

207 Demographic history inference

208 We used three approaches to infer past demographic changes through the analysis of
209 genetic data. Two of them are moment-based methods that rely on summary statistics (i.e., the
210 BOTTLENECK method, Cornuet & Luikart 1996; and the M-ratio method, Garza &
211 Williamson 2001) and the third uses a full-likelihood Bayesian approach (i.e., the MSVAR
212 method, Beaumont 1999; Storz & Beaumont 2002). For simulated data, analyses were

213 performed at two different spatial levels: (i) at the deme level, where each deme was analyzed
214 independently (i.e., 10 demes x 27 scenarios = 270 analyses, 22 individuals per analysis) and
215 (ii) at the population level, where all individuals from a same scenario were pooled together in
216 a single analysis (i.e., one analysis per scenario, 220 individuals per analysis). Pooling
217 individuals from multiple sampling locations counters potential biases induced by population
218 structure when looking for demographic changes and improves the characterization of
219 parameters associated to demographic changes at the population level (Chikhi *et al.* 2010).
220 Due to the computational burden inherent to MSVAR, population-level analyses were not
221 performed using this method. For empirical data, analyses were done (i) at the sampling site
222 level (i.e., 74 analyses, ~20-22 individuals *per* analysis) and (ii) at the population level (i.e., 8
223 analyses, between 140 and 220 individuals *per* analysis).

224 *BOTTLENECK method.* We applied the moment-based method of Cornuet & Luikart (1996)
225 as implemented in the BOTTLENECK software (Piry *et al.* 1999). This method compares the
226 expected heterozygosity computed from a sample (H_e) through observed allele frequencies
227 with the expected heterozygosity (H_{eq}) based on the allele frequencies expected at the
228 mutation-drift equilibrium (given the observed number of alleles n_A of the sample). The
229 significance of deviations from mutation-drift equilibrium was tested through Wilcoxon's
230 signed rank tests. For simulated data, we performed analyses assuming the Stepwise Mutation
231 Model (SMM, Piry *et al.* 1999), as it is the mutation model used by ms to simulate the data
232 (Hudson 2002). Additionally, we calculated from the output of BOTTLENECK departures
233 from mutation-drift equilibrium averaged over loci: $\Delta H = H_e - H_{eq}$ (Broquet *et al.* 2010). For
234 empirical data, we performed analyses assuming a Two-Phase mutation Model (TPM), which
235 is more appropriate for empirical microsatellite data (Di Rienzo *et al.* 1994; Piry *et al.* 1999).

236 We parameterized the TPM with 90% single step mutations (Garza & Williamson 2001),
237 assuming a conservative variance among multiple steps of 10.

238 *M-ratio method.* To detect significant population declines in our datasets, we applied Garza &
239 Williamson's M-ratio test (Garza & Williamson 2001). It is noteworthy that this method
240 (contrary to the two other methods) does not allow the detection of demographic expansions.
241 In bottlenecked populations, the number of alleles on microsatellite loci (n_A) is expected to be
242 reduced more quickly than the range in allele size (r_A). As a result, the ratio $M = n_A / r_A$ will
243 be smaller in bottlenecked populations than in stable populations (Garza & Williamson 2001).
244 Accordingly, we calculated M for both empirical and simulated datasets. Then, we compared
245 M values obtained from our data with 95% critical M values (M_c), calculated from 10,000
246 simulations of stable populations with the Critical_M program (Garza & Williamson 2001).
247 An M value that falls below the M_c value indicates that the population has experienced a
248 significant bottleneck. For simulated scenarios, we assessed M_c values assuming the SMM,
249 and using the θ values previously used to simulate the data. For empirical data, θ was
250 calculated assuming $\mu=5.56 \times 10^{-4}$ and using N_e values reported in Blanchet *et al.* (2010). We
251 assumed a TPM model with a proportion of one-step mutations of 90% and an average size of
252 non-one-step mutations of 3.5 (Garza & Williamson 2001).

253 *MSVAR method.* To detect and quantify changes in N_e , we used a method relying on a
254 hierarchical Bayesian model based on a coalescent framework (as implemented in MSVAR
255 1.3, Beaumont 1999; Storz & Beaumont 2002). This model assumes that a stable, closed
256 population of ancestral size N_1 increased or decreased exponentially to its current size N_0 over
257 a time interval ta (in years). Given lognormal prior distributions and microsatellite data (i.e.,
258 allelic distribution and relative allele sizes), the method infers the model parameters $\Phi =$
259 $\{N_0, N_1, ta, \theta\}$, where $\theta = 4N_0\mu$ and μ is the mutation rate. The posterior probability density of Φ

260 is established through Markov Chain Monte Carlo (MCMC) techniques. Loci are supposed to
261 be independent and to evolve under a strict SMM, but the method is also robust against
262 deviations from strict SMM (Storz & Beaumont 2002; Girod *et al.* 2011). For each MSVAR
263 analysis, we performed four independent runs of 5×10^9 steps, varying the starting values and
264 means for priors and hyperpriors (values in Table S2). Parameters were thinned with an
265 interval of 5×10^4 steps, resulting in output files with 1×10^5 values. To avoid bias induced by
266 the starting values on parameter estimation, the first 10% of the chains was discarded (i.e.,
267 burn-in). We checked the convergence of the chains visually and with the Gelman & Rubin
268 analysis (Gelman & Rubin 1992). We considered that chains converged well when values
269 smaller than 1.1 were obtained (Gelman & Hill 2007).

270 For each independent run of MSVAR, the magnitude of the demographic change was
271 estimated through the calculation of an effect size (i.e., Hedges'd, Hedges & Olkin 1985) and
272 its 95% confidence interval. Hedges'd is a mean standardized difference (i.e., independent of
273 the original scale) between the log of the ancestral population size ($\log(N_I)$) and the log of the
274 current population size ($\log(N_0)$). The standardization of the mean difference is obtained by
275 dividing the mean difference by a pooled standard deviation (formulas in Appendix S1). We
276 combined the four effect sizes of each independent run to calculate a mean effect size (MES)
277 *per* analysis, along with its 95% confidence interval (Rosenberg *et al.* 1997). A MES value
278 whose confidence interval includes zero means that the population did not experience a
279 significant demographic change. Significantly negative values correspond to significant
280 bottlenecks, while significantly positive values are significant population expansions. Pairs of
281 MES were considered as significantly different when their 95% confidence intervals did not
282 overlap. Information about these methods along with an illustrative example is provided in the
283 Appendix S1.

284 For empirical data, we further estimated the beginning of the exponential demographic
285 changes inferred with MSVAR by calculating Bayes' factors (BFs), which measure the
286 weight of evidence of alternative time intervals for ta (i.e., the time of the beginning of the
287 demographic change). BFs were first computed for time periods of 10 years in a sliding
288 window from 0 to 100 years, then for periods of 100 years from 200 to 10,000 years ago. BFs
289 greater than 4 are usually interpreted as positive evidence, while BFs greater than 7 are
290 considered as significant (Storz & Beaumont 2002; Sousa *et al.* 2008). For each species on the
291 highly fragmented river, we also calculated (through the posterior distribution of ta) the
292 probability that the detected demographic changes occurred (i) after dam construction ($p_{(dam)}$,
293 ta between 0-60 years ago), and (ii) after weir construction began ($p_{(weir)}$, ta between 0-800
294 years ago). We considered a generation time of three years for *S. cephalus* and *L.*
295 *burdigalensis*, and of two years for *G. gobio* and *P. phoxinus* (Poncin *et al.* 1987). For the
296 sake of clarity, we present only BFs computed for ta at the population level.

297 *Effects of N , m , a and distance from the source on demographic history inference.* In order to
298 synthesize results obtained from the simulated datasets, we ran Generalized Linear Models
299 (GLMs) to statistically test for each method independently the effects of N , m , a and distance
300 from the putative source (D) on inferences of changes in Ne . In these models, the dependent
301 variables were ΔH , M and MES (calculated at the deme level) for the BOTTLENECK, M-
302 ratio and MSVAR methods respectively. Explanatory variables were N , m , a and D . They
303 were all treated as fixed effects, and we further included all two-term and three-term
304 interactions so as to test the significance of interacting effects between explanatory variables.
305 We assumed Gaussian error terms for all dependent variables and the significance of each
306 fixed effect was assessed using F-ratio tests.

307

308 **Results**

309 Simulated data

310 *BOTTLENECK method.* At the deme level and over all scenarios, 47 datasets (47/270=17.4%)
311 exhibited significant departures from mutation-drift equilibrium. Most of them (32/47=68%)
312 displayed significant heterozygosity deficiencies, which are generally interpreted as signals of
313 demographic expansions. Only 15 demes displayed significant heterozygosity excesses, which
314 are generally interpreted as signals of bottlenecks. At the population level, and over all
315 scenarios, we detected 14 (14/27=51.9%) significant departures from mutation-drift
316 equilibrium, all in the form of heterozygosity deficiencies. Additionally, our GLM-based
317 analysis revealed a significant three-way interaction between N , m and a (Table 1). This
318 analysis indicates that the BOTTLENECK method detected false signals of expansion (i.e.,
319 negative values of ΔH) under moderate (i.e., $a=7.5$) and strong (i.e., $a=50$) gene flow
320 asymmetries, although this pattern was altered by the effective population size at the deme
321 level (Figure 2A-C).

322 *M-ratio method.* At the deme level and over all scenarios, 36.3% of the demes (i.e., 98/270)
323 displayed a significant signal of population decrease. However, at the population level, no
324 significant signals of demographic decline were detected. The GLM-based analysis also
325 highlighted a significant three-way interaction between N , m and a (Table 1). This analysis
326 confirmed that the M-ratio method detected false signals of bottlenecks, but only for
327 symmetric gene flow, and under some specific combinations of N and m (Figure 2D).

328 *MSVAR method.* 41.85% of deme-level datasets (i.e., 113/270) indicated significant signals of
329 demographic change. Among these significant signals, false signals of expansion were more
330 frequent than false signals of bottleneck (69% vs. 31% respectively). According to the GLM

331 analysis, we detected two significant two-term interactions, one implying N and m , and the
332 other implying m and a (Table 1). The first interaction indicated that, irrespective of a , false
333 signals of bottleneck were mainly detected for low values of N and m , whereas false signals of
334 expansion tended to be greater for intermediate values of m (0.053) and large values of N ($>$
335 500, Figure 3A). The second interaction indicates that, irrespective of N , strong signals of
336 false bottlenecks were mainly detected for situations of symmetric gene flow (i.e., $a=1$), but
337 only for low migration rate ($m=0.01$, Figure 3B). In contrast, strong signals of false
338 expansions were detected under several and contrasted combinations of m and a (Figure 3B).
339 Indeed, false signals of expansion were detected under symmetric gene flow and with high
340 migration rate ($m=0.1$), but also under asymmetric gene flow ($a = 7.5$ or 50) and low to
341 medium migration rates ($m = 0.01$ or 0.053 , Figure 3B). We additionally found that, overall,
342 the magnitude of the false demographic expansion increased with the distance from the
343 putative source (Table 1).

344

345 Empirical data

346 *BOTTLENECK method.* At the sampling site level, we detected a significant heterozygosity
347 excess in only one case (i.e., site V8 for *S. cephalus* in the river Viaur, Table S3). In contrast,
348 17 significant heterozygosity deficiencies were detected (Table S3). None of these deviations
349 were significant after Bonferroni corrections. In contrast, at the population level, significant
350 heterozygosity deficiencies were found for all species and in the two rivers (Table 2).

351

352 *M-ratio method.* At the sampling site level, the M-ratio test detected significant bottlenecks at
353 all sites, irrespective of the species and the river (Table S3). At the population level, all

354 populations exhibited significant signals of bottleneck but one (i.e., *G. gobio* in the river Célé;
355 Table 2).

356

357 *MSVAR* method. At the sampling site level, most sampling sites displayed significant
358 bottlenecks (i.e., all MES values were significantly negative), a pattern that holds true for all
359 species and rivers (Figure 4). There were no clear spatial patterns along the upstream-
360 downstream gradient (i.e., demographic changes did not tend to be larger either downstream
361 or upstream, Figure 4). However, there were striking site-to-site MES discrepancies. For
362 instance, for *P. phoxinus*, we found no significant demographic changes in downstream sites
363 for both the Célé and Viaur rivers (i.e., the MES 95% CI included 0), while other sites were
364 characterized by signals of bottlenecks of diverse magnitudes (Figure 4D).

365 Concerning population level analyses, we found significant bottlenecks for all species
366 and rivers (Figure 5). These analyses indicated that the magnitude of the bottleneck tended to
367 be stronger for the two largest species (*S. cephalus* and more particularly *L. burdigalensis*)
368 than for the two smallest species (*G. gobio* and *P. phoxinus*; Figure 5). Furthermore, the
369 magnitude of the bottleneck was significantly stronger in the highly fragmented river for *L.*
370 *burdigalensis* and *G. gobio* (Figure 5).

371 Regarding the dating of the detected bottlenecks, we estimated that they most probably
372 occurred more than 800 years ago (Figure 6) and thus before dam or weir construction.
373 Accordingly, the probabilities that these bottlenecks occurred after dam or weir construction
374 on the highly fragmented river were very low for all species ($p_{(dam)} < 0.007$, $p_{(weir)} < 0.052$). Only
375 *P. phoxinus* showed a non negligible $p_{(weir)}$ of 0.238. Over all species, the population declines
376 tended to be more ancient in the highly fragmented river than in the weakly fragmented river,

377 except for *L. burdigalensis* (Figure 6). At the intra-river level, *ta* estimations were also
378 congruent for all species but *L. burdigalensis*. This species revealed the most ancient *ta* values
379 on the weakly fragmented river (Figure 6A) whereas it showed one of the most recent
380 bottlenecks on the highly fragmented river (Figure 6B).

381

382 **Discussion**

383 As expected, our simulated data showed that asymmetric gene flow can bias the
384 genetically-based inference of past demographic changes. We notably demonstrated that
385 asymmetric gene flow can -under certain conditions of migration rate and effective population
386 size- generate false signals of population expansion. Interestingly, this tendency was detected
387 in our empirical data, but only for one of the three inference methods we used. In contrast, the
388 other two methods revealed strong signals of bottleneck for the four fish species and across
389 the two rivers sampled, which are characterized by different levels of asymmetric gene flow
390 (see Table S1).

391

392 *Effects of gene flow asymmetry on demographic history inferences*

393 In most cases of significant –although spurious- demographic changes, our simulations
394 showed that asymmetric gene flow generates false signals of demographic expansion.
395 However, this pattern was sensitive to other population parameters, namely the migration rate
396 and the effective population size. We indeed detected strong interactive effects of these
397 population parameters on signals of false demographic changes. These interactive effects are
398 yet difficult to biologically interpret, and make difficult to withdraw general predictions about
399 the effect of asymmetric gene flow on estimates of historical demographic changes in natural

400 systems. Our results hence demonstrate the importance of simultaneously considering
401 multiple parameters such as the effective population size and the migration rate when testing
402 the robustness of analytical methods through simulations.

403 The effect of asymmetric gene flow on demographic change inferences was also
404 dependent on the method we used. Indeed, contrary to the MSVAR and the BOTTLENECK
405 methods, the M-ratio method was not affected by asymmetric gene flow, as we found no clear
406 evidence that downstream-biased asymmetric gene flow led to false signals of bottleneck.
407 However, under conditions of symmetric gene flow, the M-ratio method tended to detect false
408 signals of bottleneck, especially under low to moderate migration rates. As demonstrated
409 previously for the MSVAR method (Chikhi *et al.* 2010), this may be due to the confounding
410 effects of population structure, and of the sampling scheme on the representativeness of
411 genetic diversity.

412 We further observed correlations between distance from the upstream deme and the
413 magnitude of the demographic expansion (only for the MSVAR method). These differences
414 between upstream and downstream demes are probably the result of a source-sink like
415 dynamic, whereby downstream demes act as sinks and receive an excess of alleles through
416 downstream-directed migration (Kawecki & Holt 2002; Morrissey & de Kerckhove 2009).
417 Such source-sink dynamics generally lead to a gradual increase of allelic richness along the
418 upstream-downstream gradient in rivers (Hänfling & Weetman 2006; Blanchet *et al.* 2010),
419 and may therefore produce signals similar to those generated by demographic expansions.
420 This may be because the number and frequencies of alleles actually observed in downstream
421 sites are different than what expected under a demographically stable model. Finally, we
422 found that the symmetric gene flow scenario led to patterns of false bottlenecks (only for low

423 migration rate), as expected from previous simulations in n-island and two-dimensional
424 stepping-stone models (Städler *et al.* 2009; Chikhi *et al.* 2010).

425

426 *Effect of asymmetric gene flow on fish population demographic histories*

427 We detected significant population bottlenecks for all species in the two rivers when
428 we analyzed the empirical data. Because two out of the three methods (MSVAR and M-ratio
429 methods) were concordant in highlighting significant bottlenecks, we could reasonably
430 assume that these populations had actually experienced demographic declines. However,
431 significant signals of expansions were identified for all species and rivers at the population
432 level using the BOTTLENECK method. This result is consistent with that obtained for the
433 simulated data (see above), suggesting that, in wild populations, this method may be subjected
434 to the type of bias induced by asymmetric gene flow. Overall, this would suggest that, despite
435 asymmetric gene flow may theoretically affect the inference of demographic changes (our
436 simulations), some inference methods may be powerful enough to bypass this type of bias
437 when a population has actually experienced a bottleneck.

438 We tested such a hypothesis by running an additional analysis in which we simulated a
439 scenario where the population was subjected to (i) a bottleneck of magnitude and timing
440 similar to that estimated for the empirical data, and (ii) post-bottleneck ϕ values equal to the
441 mean values estimated from the literature survey (i.e., $N=3147$, $a=7.5$, $m=0.053$). We found
442 that MSVAR detected a significant bottleneck (results not shown), which suggests that at least
443 under some conditions, MSVAR can bypass the bias induced by asymmetry. It is noteworthy
444 that we also detected a significant bottleneck using the M-ratio test, whereas BOTTLENECK
445 detected a significant heterozygosity deficiency (i.e., a population expansion signal).

446 Regarding our empirical data, we note however that some sampling sites did not display
447 significant demographic changes. For instance, the absence of significant bottlenecks for *P.*
448 *phoxinus* in downstream sites suggests that asymmetric gene flow was probably strong
449 enough in these sites to counterbalance the effect of ancient bottlenecks. This means that
450 more simulations varying both asymmetric gene flow and the characteristics (i.e., magnitude,
451 date and type) of demographic changes are required to refine the conditions under which
452 MSVAR adequately detects population size changes.

453 To summarize, our study suggests that the BOTTLENECK method may be less suited
454 than the MSVAR and M-ratio methods to infer demographic changes in wild populations
455 experiencing asymmetric gene flow. This conclusion is apparently solid, since our empirical
456 dataset includes fish populations covering a wide range of values regarding their levels of
457 asymmetric gene flow (i.e., $1.893 < a < 9.135$), migration rate (i.e., $0.042 < m < 0.078$) and
458 effective population size (i.e., $546.488 < N < 8,088.188$; see Table S1). But, given the fact we
459 do not know the actual demographic history of these populations, we should remain cautious.
460 An important lesson from this is perhaps that each methods looks at the genetic data from a
461 slightly different angle, and uses different aspects of genetic diversity measures, which may in
462 the end mean that the methods could be used jointly once we better understand their joint
463 properties.

464 From a biological point of view, we surprisingly found that the dating of the
465 bottlenecks experienced by these populations was similar for three of the four species. For all
466 species, we found that the corresponding demographic declines were ancient and pre-dated
467 the construction of the weirs and dams. For the highly fragmented river, the most likely
468 inferred dates for the beginning of the bottlenecks range from 2,000 to 8,000 years ago, which
469 contrasts with the first known mill weirs in this river (~800 years ago). Such dating suggests

470 that these bottlenecks occurred after the last glacial period (i.e., Würm glacial period, $ta <$
471 10,000 years), more precisely between the Atlantic and the middle Subatlantic chronozones of
472 the Holocene (Mangerud *et al.* 1974). These important bottlenecks might have been generated
473 by different events, such as post-glacial colonization (Hänfling *et al.* 2002; Swatdipong *et al.*
474 2010), environmental stochastic events or random catastrophes (Hedrick & Miller 1992;
475 Lande 1993). The dating obtained with the MSVAR method might only be loosely related to
476 any particular event. Improving our knowledge in the paleoenvironmental history of the
477 studied region would certainly help in understanding the potential causes of such strong
478 population declines. Moreover, in the case of a series of expansions and contractions (which
479 are likely to have happened in many natural systems), it is unclear which event would be
480 “identified” by MSVAR (Quéméré *et al.* 2012; Salmons *et al.* 2012). Simulation of multiple
481 events may thus be necessary for improving our interpretation of MSVAR outputs.

482

483 *Conclusion*

484 Recent years have shown that several factors can play significant roles in producing
485 non-equilibrium patterns, such as isolation by distance (Leblois *et al.* 2006), population
486 structure (Städler *et al.* 2009; Chikhi *et al.* 2010, Peter *et al.* 2010), rapid decreases of gene
487 flow (i.e., fragmentation, Broquet *et al.* 2010), spatial expansions (Edmonds *et al.* 2002), or
488 departures from the assumed mutation model (Chikhi *et al.* 2010). However, the
489 consequences of asymmetrical gene flow have been neglected. Our simulations confirm our
490 expectation that asymmetric gene flow may generate biases when inferring demographic
491 changes from genetic data. However, the direction and magnitude of such biases depended
492 upon other population characteristics such as migration rate and effective population size.

493 This study demonstrates the complexity of inferring demographic changes from genetic data
494 in wild populations, and the importance of integrating multiple parameters in simulations
495 aiming at testing the robustness of inference methods in population genetics (e.g., Heller *et al.*
496 *in press*).

497 In spite of these potential biases, our multi-specific empirical data suggests that, if
498 used with care and conjointly, most inference methods appear suitable to infer demographic
499 changes in populations experiencing asymmetric gene flow. Indeed, our empirical data
500 suggest that asymmetric gene flow was unlikely to have caused the bottlenecks observed in
501 the eight wild fish populations. We also found that if a major bottleneck was responsible of
502 the patterns observed, it was unlikely to have been caused by recent anthropogenic
503 fragmentation. However, we cannot claim that we have identified unambiguously the factors
504 generating the strong bottlenecks observed in all fish species, even if they dated around the
505 same period.

506 The last twenty years have seen major improvements in population genetics inference,
507 in particular with the development of full-likelihood methods. Our results and those from
508 previous studies clearly demonstrate that population structure and dispersal patterns have to
509 be considered for properly inferring the demographic history of wild populations (Chikhi *et al.*
510 *2010*; Girod *et al.* *2011*). An important step for future studies will be to quantify the ability
511 of emerging methods (such as those based on approximate Bayesian computations) to
512 efficiently disentangle signals of demographic changes from false signals arising from
513 population structure (see Peter *et al.* *2010* for instance).

514

515 **Acknowledgements**

516 We thank Éric Petit, Thomas Broquet, Vincent Dubut, Jérôme Chave, Camille Pagès,
517 Guillaume Evanno, Raphael Leblois and three anonymous reviewers for their constructive
518 and stimulating comments. Olivier Rey, Gaël Grenouillet, Loïc Tudesque, Muriel Gevrey,
519 Laetitia Buisson, Sébastien Brosse, Leslie Faggiano and Fabien Leprieur are thanked for their
520 help in the field. We also thank the CALMIP group, in particular Boris Dintrans and Nicolas
521 Renon. This work was performed using HPC (High Performance Computing) resources from
522 CALMIP (allocation 2010-P1003). We are grateful to Radika Michniewicz for correcting and
523 editing the English. The authors also thank the "Agence de l'Eau Adour-Garonne" for
524 financial support and the "Génopole Toulouse" for help with genotyping. IP is financially
525 supported by a MESR ("Ministère de l'Enseignement Supérieur et de la Recherche") PhD
526 scholarship. This work has been done in two research units (EDB & EcoEx CNRS Moulis)
527 that are part of the "Laboratoire d'Excellence (LABEX) entitled TULIP (ANR -10-LABX-
528 41).

529

530 **References**

- 531 Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic
532 DNA for PCR-based techniques. *Nucleic Acids Research* **25**, 4692-4693.
- 533 Alò D, Turner TF (2005) Effects of habitat fragmentation on effective population size in the
534 endangered Rio Grande silvery minnow. *Conservation Biology* **19**, 1138-1148.
- 535 Beaumont MA (1999) Detecting population expansion and decline using microsatellites.
536 *Genetics* **153**, 2013-2029.

- 537 Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in
538 population genetics. *Genetics* **162**, 2025-2035.
- 539 Blanchet S, Rey O, Etienne R, Lek S, Loot G (2010) Species-specific responses to landscape
540 fragmentation: implications for management strategies. *Evolutionary Applications* **3**,
541 291-304.
- 542 Bolland JD, Cowx IG, Lucas MC (2008) Movements and habitat use of wild and stocked
543 juvenile chub, *Leuciscus cephalus* (L.), in a small lowland river. *Fisheries*
544 *Management and Ecology* **15**, 401-407.
- 545 Broquet T, Angelone S, Jaquiere J, *et al.* (2010) Genetic bottlenecks driven by population
546 disconnection. *Conservation Biology* **24**, 1596-1605.
- 547 Chikhi L, Bruford MW (2005) Mammalian population genetics and genomics. *Mammalian*
548 *Genomics* (ed. Ruvinsky A, Marshall Graves J). CABI Publishers, UK. **Chapter 21**,
549 539-583.
- 550 Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of
551 population structure, genetic diversity and the sampling scheme on the detection and
552 quantification of population size changes. *Genetics* **186**, 983-U347.
- 553 Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting
554 recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.
- 555 Csillery K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian
556 computation (ABC). *Methods in Ecology and Evolution* **3**, 475-479.

- 557 De Leeuw JJ, Winter HV (2008) Migration of rheophilic fish in the large lowland rivers
558 Meuse and Rhine, the Netherlands. *Fisheries Management and Ecology* **15**, 409-415.
- 559 Di Rienzo A, Peterson AC, Garza JC, *et al.* (1994) Mutational processes of simple-sequence
560 repeat loci in human populations. *Proceedings of the National Academy of Sciences of
561 the United States of America* **91**, 3166-3170.
- 562 Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an
563 expanding population. *Proceedings of the National Academy of Sciences of the United
564 States of America* **101**, 975-979.
- 565 Fahrig L (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology
566 Evolution and Systematics* **34**, 487-515.
- 567 Garza JC, Williamson EG (2001) Detection of reduction in population size using data from
568 microsatellite loci. *Molecular Ecology* **10**, 305-318.
- 569 Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*
570 Cambridge University Press, New York.
- 571 Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences.
572 *Statistical Science* **7**, 457-472.
- 573 Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion
574 from microsatellite data: a simulation-based evaluation of the msvar method. *Genetics*
575 **188**, 165-U287.
- 576 Goossens B, Chikhi L, Ancrenaz M, *et al.* (2006) Genetic signature of anthropogenic
577 population collapse in orang-utans. *PLoS Biology* **4**, 285-291.

- 578 Hänfling B, Hellemans B, Volckaert FAM, Carvalho GR (2002) Late glacial history of the
579 cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular*
580 *Ecology* **11**, 1717-1729.
- 581 Hänfling B, Weetman D (2006) Concordant genetic estimators of migration reveal
582 anthropogenically enhanced source-sink population structure in the River Sculpin,
583 *Cottus gobio*. *Genetics* **173**, 1487-1501.
- 584 Hanski I (1998) Metapopulation dynamics. *Nature* **396**, 41-49.
- 585 Hedges LV, Olkin I (1985) *Statistical methods for meta-analysis* Academic Press, New York.
- 586 Hedrick PW, Miller PS (1992) Conservation genetics - Techniques and fundamentals.
587 *Ecological Applications* **2**, 30-46.
- 588 Heller R, Chikhi L, Sigiesmund HR (*in press*) The confounding effect of population structure
589 on Bayesian skyline plot inferences of demographic history. *PLoS ONE*.
- 590 Henle K, Lindenmayer DB, Margules CR, Saunders DA, Wissel C (2004) Species survival in
591 fragmented landscapes: where are we now? *Biodiversity and Conservation* **13**, 1-8.
- 592 Hoban S, Bertorelle G, Gaggiotti OE (2012) Computer simulations: tools for population and
593 evolutionary genetics. *Nature Reviews Genetics* **13**, 110-122.
- 594 Horreo JL, Martinez JL, Ayllon F, *et al.* (2011) Impact of habitat fragmentation on the
595 genetics of populations in dendritic landscapes. *Freshwater Biology* **56**, 2567-2579.
- 596 Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic
597 variation. *Bioinformatics* **18**, 337-338.

- 598 Kawecki TJ, Holt RD (2002) Evolutionary consequences of asymmetric dispersal rates.
599 *American Naturalist* **160**, 333-347.
- 600 Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability*
601 **19**, 27-43.
- 602 Lande R (1993) Risks of population extinction from demographic and environmental
603 stochasticity and random catastrophes. *American Naturalist* **142**, 911-927.
- 604 Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over
605 large recombining regions in a subdivided population with a complex history.
606 *Bioinformatics* **20**, 2485-2487.
- 607 Leblois R, Estoup A, Streiff R (2006) Genetics of recent habitat contraction and reduction in
608 population size: does isolation by distance matter? *Molecular Ecology* **15**, 3601-3615.
- 609 Mangerud J, Andersen ST, Berglund BE, Donner JJ (1974) Quaternary stratigraphy of
610 Norden, a proposal for terminology and classification. *Boreas* **3**, 109-126.
- 611 Morrissey MB, de Kerckhove DT (2009) The maintenance of genetic variation due to
612 asymmetric gene flow in dendritic metapopulations. *American Naturalist* **174**, 875-
613 889.
- 614 Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular*
615 *Ecology* **18**, 1034-1047.
- 616 Nock CJ, Overden JR, Butler GL, *et al.* (2011) Population structure, effective population size
617 and adverse effects of stocking in the endangered Australian eastern freshwater cod
618 *Maccullochella ikei*. *Journal of Fish Biology* **78**, 303-321.

- 619 Peter BM, Wegmann D, Excoffier L (2010) Distinguishing between population bottleneck
620 and population subdivision by a Bayesian model choice procedure. *Molecular Ecology*
621 **19**, 4648-4660.
- 622 Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting
623 recent reductions in the effective population size using allele frequency data. *Journal*
624 *of Heredity* **90**, 502-503.
- 625 Pollux BJA, Luteijn A, Van Groenendael JM, Ouborg NJ (2009) Gene flow and genetic
626 structure of the aquatic macrophyte *Sparganium emersum* in a linear unidirectional
627 river. *Freshwater Biology* **54**, 64-76.
- 628 Poncin P, Melard C, Philippart JC (1987) Use of temperature and photoperiod in the control
629 of the reproduction of 3 European cyprinids, *Barbus barbus* (L), *Leuciscus cephalus*
630 (L) and *Tinca tinca* (L), reared in captivity – Preliminary Results. *Bulletin Français*
631 *De La Pêche Et De La Pisciculture*, 1-12.
- 632 Quéméré E, Amelot X, Pierson J, Crouau-Roy B, Chikhi L (2012). Genetic data suggest a
633 natural pre-human origin of open habitats in northern Madagascar and question the
634 deforestation narrative in this region. *Proceedings of the National Academy of*
635 *Sciences of the United States of America* **109**, 13028-33.
- 636 Raeymaekers JAM, Maes GE, Geldof S, *et al.* (2008) Modelling genetic connectivity in
637 sticklebacks as a guideline for river restoration. *Evolutionary Applications* **1**, 475-488.
- 638 Rosenberg MS, Adams DC, Gurevitch J (1997) Metawin: Statistical software for meta-
639 analysis with resampling tests. *Sunderland, MA, US: Sinauer Associates.* **iv**,, 65.

- 640 Salmons J, Salamolard M, Fouillot D, *et al.* (2012) Signature of a pre-human population
641 decline in the critically endangered reunion island endemic forest bird *Coracina*
642 *newtoni*. *PLoS ONE* **7(8)**, e43524.
- 643 Sousa V, Penha F, Collares-Pereira MJ, Chikhi L, Coelho MM (2008) Genetic structure and
644 signature of population decrease in the critically endangered freshwater cyprinid
645 *Chondrostoma lusitanicum*. *Conservation Genetics* **9**, 791-805.
- 646 Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The impact of sampling
647 schemes on the site frequency spectrum in nonequilibrium subdivided populations.
648 *Genetics* **182**, 205-216.
- 649 Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and
650 contraction: An empirical analysis of microsatellite DNA variation using a
651 hierarchical Bayesian model. *Evolution* **56**, 154-166.
- 652 Swatdipong A, Primmer CR, Vasemagi A (2010) Historical and recent genetic bottlenecks in
653 European grayling, *Thymallus thymallus*. *Conservation Genetics* **11**, 279-292.
- 654 Weiss G, von Haeseler A (1998) Inference of population history using a likelihood approach.
655 *Genetics* **149**, 1539-1546.
- 656 Yue GH, David L, Orban L (2007) Mutation rate and pattern of microsatellites in common
657 carp (*Cyprinus carpio L.*). *Genetica* **129**, 329-331.

658

659 **Data accessibility**

660 R scripts for analyzing MSVAR outputs and for simulate genetic data with ms, empirical
661 microsatellite datasets and simulated microsatellite datasets are available at Dryad Digital
662 Repository doi:10.5061/dryad.5sc31.

663 **Authors' contributions**

664 IP, SB, GL, EQ and LC wrote the paper. SB, IP and GL designed the study and
665 managed the project. IP, SB, EQ and LC implemented the methods and analyzed the results.
666 All authors read and approved this version of the manuscript.

667

For Review Only

668 **TABLE 1: Results for the Generalized Linear Models used to synthesize results**
 669 **obtained from the analyses of simulated datasets with (i) BOTTLENECK (associated**
 670 **dependent variable = ΔH), (ii) the M-ratio method (i.e., M), and (iii) MSVAR (i.e., MES).**
 671 **"NS" indicates p-values > 0.05; * indicates p-values < 0.05; ** indicates p-values < 0.01;**
 672 ***** indicates p-values < 0.001. Significant effects indicates that explanatory variables**
 673 **significantly affect one of the three dependent variables, each being related to one of the**
 674 **three methods used to infer demographic changes. Significant single terms are not**
 675 **interpreted when they are involved in significant interaction terms.**

676

Explanatory variables	Dependent variables		
	ΔH	M	MES
Distance from the source (D)	NS	NS	***
Effective population size (N)	***	***	***
Migration rate (m)	***	***	NS
Asymmetry coefficient (a)	***	***	*
$D*N$	NS	**	NS
$D*m$	NS	NS	NS
$D*a$	NS	NS	NS
$N*m$	***	***	**
$m*a$	**	**	***
$N*a$	***	***	NS
$D*N*m$	NS	NS	NS
$D*m*a$	NS	NS	NS
$D*N*a$	NS	NS	NS
$N*m*a$	**	***	NS

677

678 **TABLE 2: Results for the Wilcoxon's sign rank tests computed by BOTTLENECK for**
 679 **the empirical data and for the M-ratio test. For the two methods, analyses were**
 680 **conducted at the population level assuming a TPM mutation model.**

681

Species	River	Status	Wilcoxon excess	Wilcoxon deficiency	M (sd)
<i>S. cephalus</i>	Viaur	highly fragmented	0.997 ^{NS}	0.005**	0.571 (0.217)*
<i>L. burdigalensis</i>	Viaur	highly fragmented	0.999 ^{NS}	0.002**	0.563 (0.197)*
<i>G. gobio</i>	Viaur	highly fragmented	0.996 ^{NS}	0.006**	0.6931 (0.233)*
<i>P. phoxinus</i>	Viaur	highly fragmented	0.980 ^{NS}	0.027*	0.748 (0.165)*
<i>S. cephalus</i>	Célé	weakly fragmented	0.999 ^{NS}	0.001**	0.5839 (0.146)*
<i>L. burdigalensis</i>	Célé	weakly fragmented	0.999 ^{NS}	<0.001**	0.664 (0.203)*
<i>G. gobio</i>	Célé	weakly fragmented	0.980 ^{NS}	0.027*	0.788 (0.201)
<i>P. phoxinus</i>	Célé	weakly fragmented	1.000 ^{NS}	0.008*	0.739 (0.171)*

682 For the BOTTLENECK analyses: * indicates a significant deviation from mutation-drift equilibrium (p-value \leq
 683 0.05); ** indicates a significant deviation from mutation-drift equilibrium after sequential Bonferroni corrections
 684 for all populations, and ^{NS} means that there is not a significant deviation from mutation-drift equilibrium (p-value
 685 > 0.05). Significant H_e excesses are evidences of recent population decreases. Significant H_e deficiencies can be
 686 interpreted as evidences of recent demographic expansion. For the M-ratio test: * indicates a significant M value
 687 (i.e., $M \leq M_c$), which is interpreted as a significant signal of population decrease, and ^{NS} means that the test is not
 688 significant (i.e., $M > M_c$).

689

690 **Figure legends**

691 **FIGURE 1.** Diagram representing the linear stepping-stone model with asymmetric
692 gene flow. Black circles are demes. $M_{\text{Downstream}}$ characterizes downstream-directed
693 gene flow, while M_{Upstream} indicates upstream-directed gene flow. Here, deme one is
694 considered as the most upstream deme of a hypothetical river.

695 **FIGURE 2.** Barplots representing values of ΔH (A, B and C) and M (D, E and F) in
696 function of three interacting parameters (as revealed by the GLM-approach: N , m
697 and a). Vertical lines correspond to the standard error. * means that the population
698 has experienced a significant bottleneck (i.e., $M < M_c$).

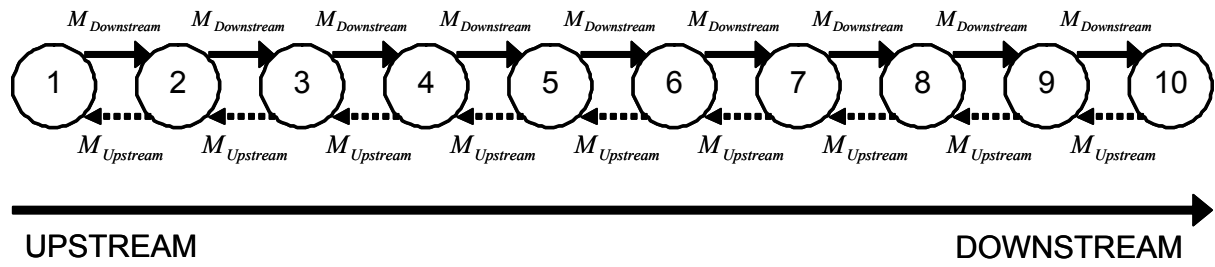
699 **FIGURE 3.** Barplots representing values of mean effect sizes (MES) in function of
700 two different two-term interactions (as revealed by the GLM-approach): (A)
701 interaction between the parameters N and m , and (B) interaction between m and a .
702 Vertical lines correspond to the standard error.

703 **FIGURE 4.** Sampling site level mean effect sizes (MES) calculated for all species
704 and rivers. Black squares characterize the weakly fragmented river (C  l  ) sites,
705 while white squares represent highly fragmented river's sites (Viaux). Dashed lines
706 represent the non-significant relationships between MES values and the distance
707 from the source at each site determined by GLMs. Grey vertical lines represent
708 MES' 95% confidence intervals (CIs). MES whose CIs include zero means that no
709 significant demographic changes have been detected. Negative values correspond to
710 significant bottlenecks. Intra-river and intra-specific MES can be easily compared
711 by seeing if their respective CIs overlap. Two MES are considered significantly
712 different when their CIs did not overlap.

713 **FIGURE 5.** Mean effect sizes (MES) for all species and rivers calculated at the
714 population level. Grey vertical lines represent MES' 95% confidence intervals
715 (CIs). Two MES are considered significantly different if their CIs did not overlap.
716 Here, we symbolized only the significance of intra-specific comparisons (i.e.,
717 comparison between MES of the highly fragmented vs. the weakly fragmented river
718 for a single species). NS indicates no significant intra-specific difference between
719 weakly fragmented vs. highly fragmented river and *** means significant
720 difference.

721 **FIGURE 6.** Bayes' factors (BFs) for the time of the beginning of the demographic
722 changes (ta) calculated for the four species for the weakly fragmented river (A) and
723 the highly fragmented river (B). Results correspond to the population level
724 analyses. BFs greater than 4 are considered as "positive evidences", while BFs
725 greater than 7 are considered as significant. Dashed vertical lines correspond to the
726 construction of dams ($ta = 60$ years) and to the beginning of weir construction ($ta =$
727 800 years).

728

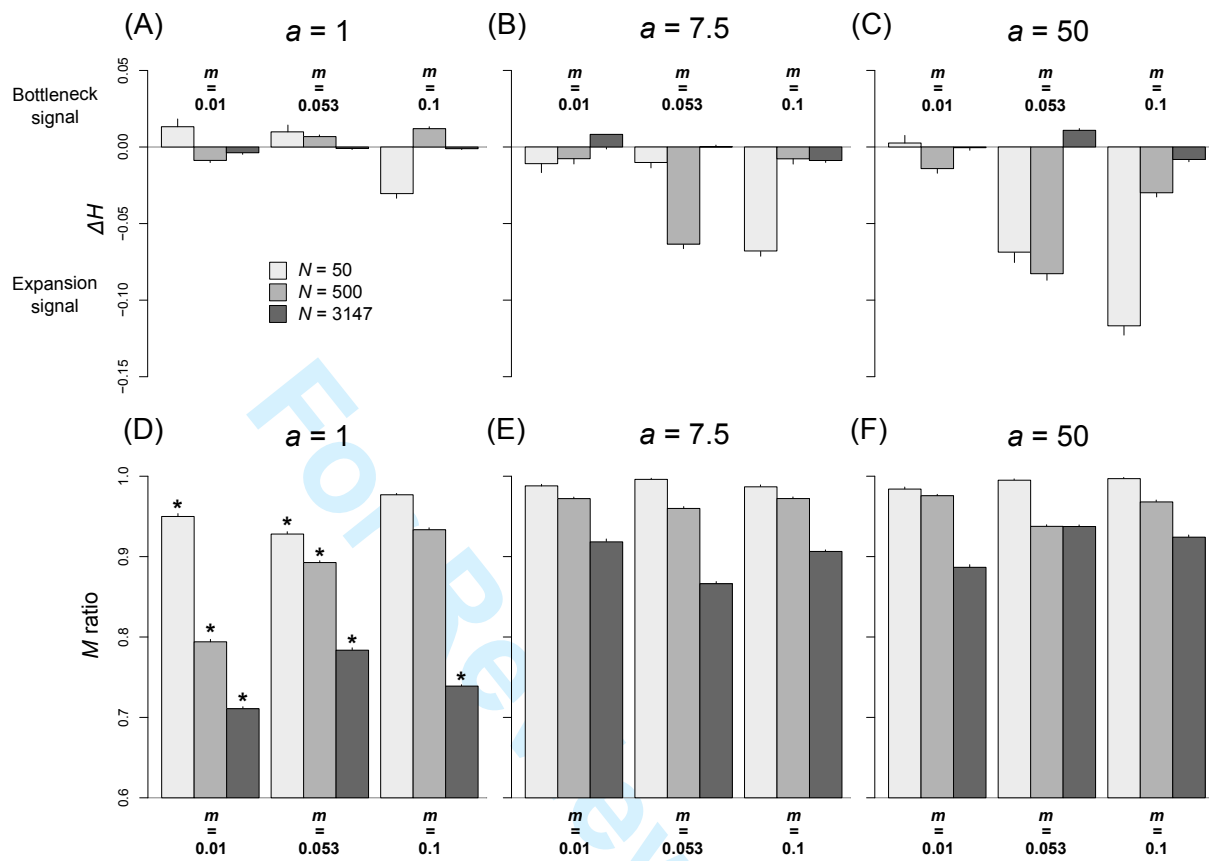
729 **FIGURE 1**

730

731

732

For Review Only

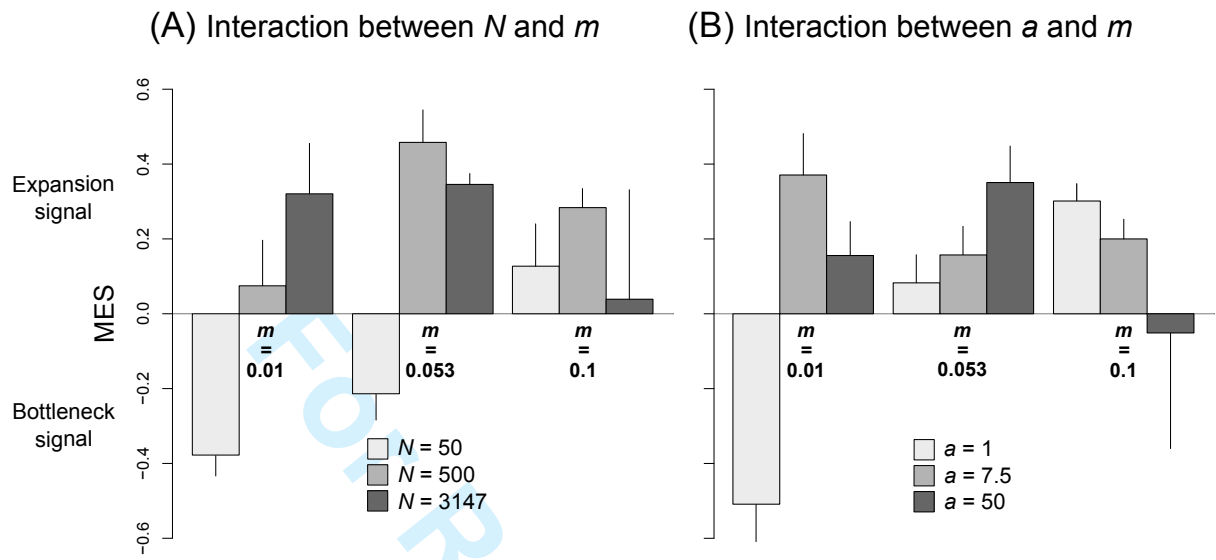
733 **FIGURE 2**

734

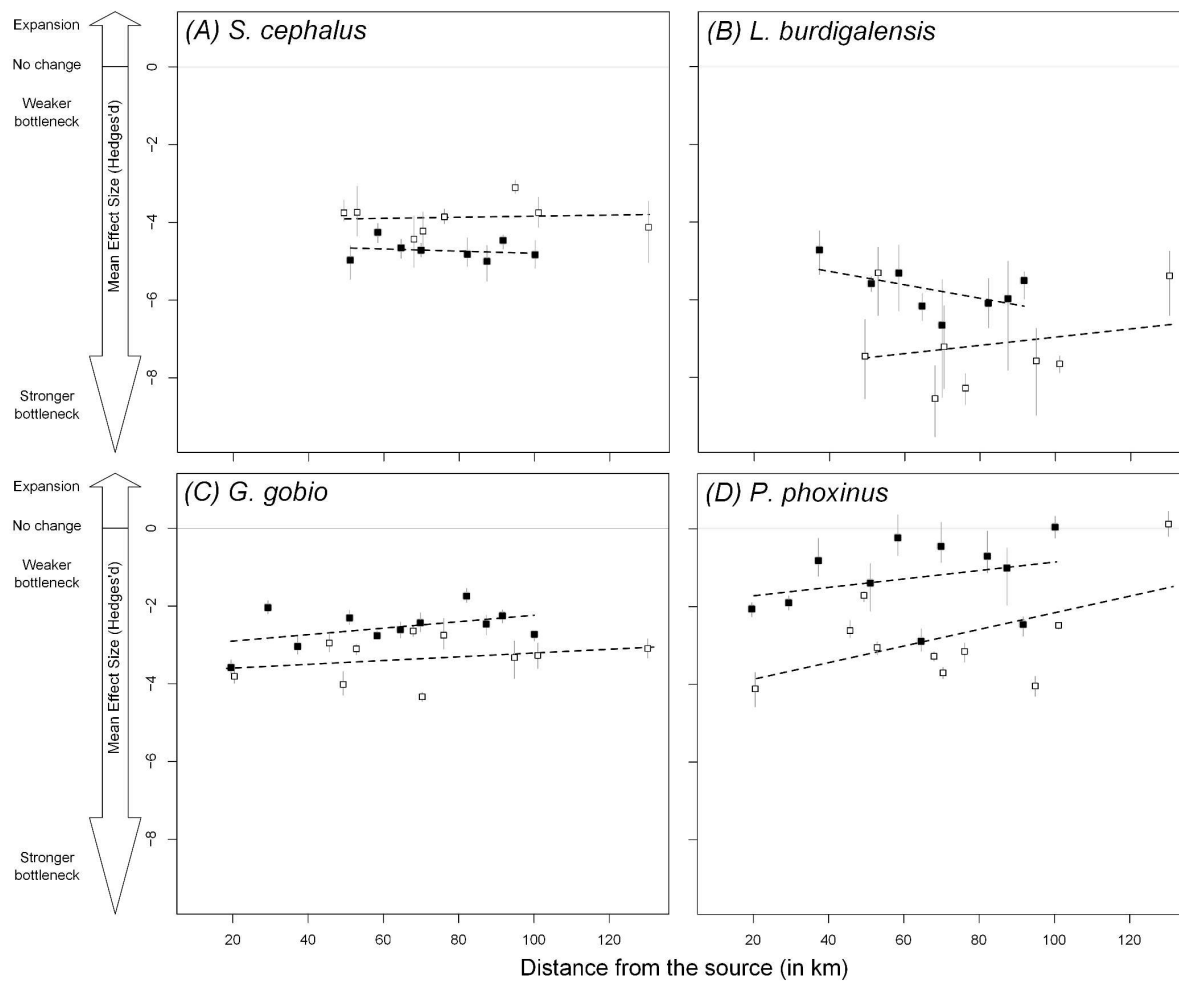
735

736 **FIGURE 3**

737

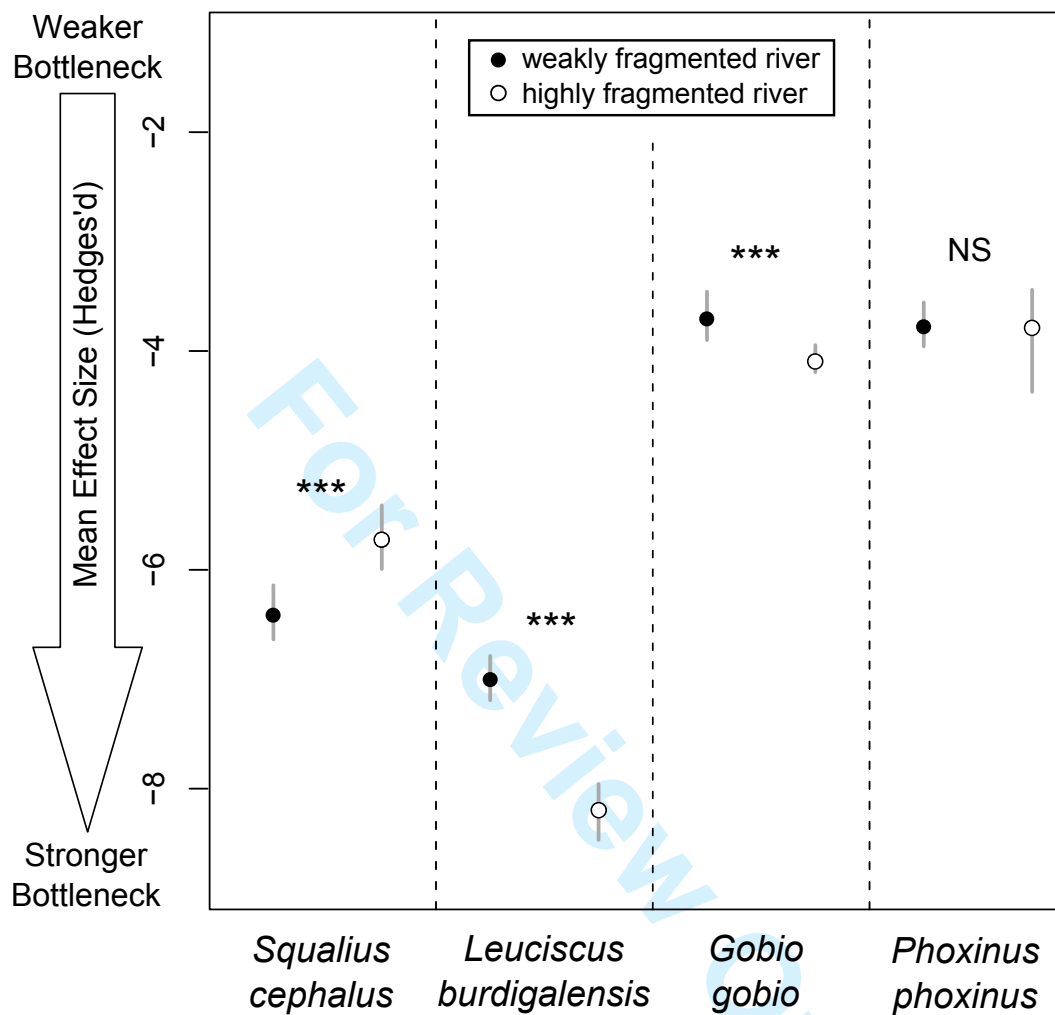


739 **FIGURE 4**



740

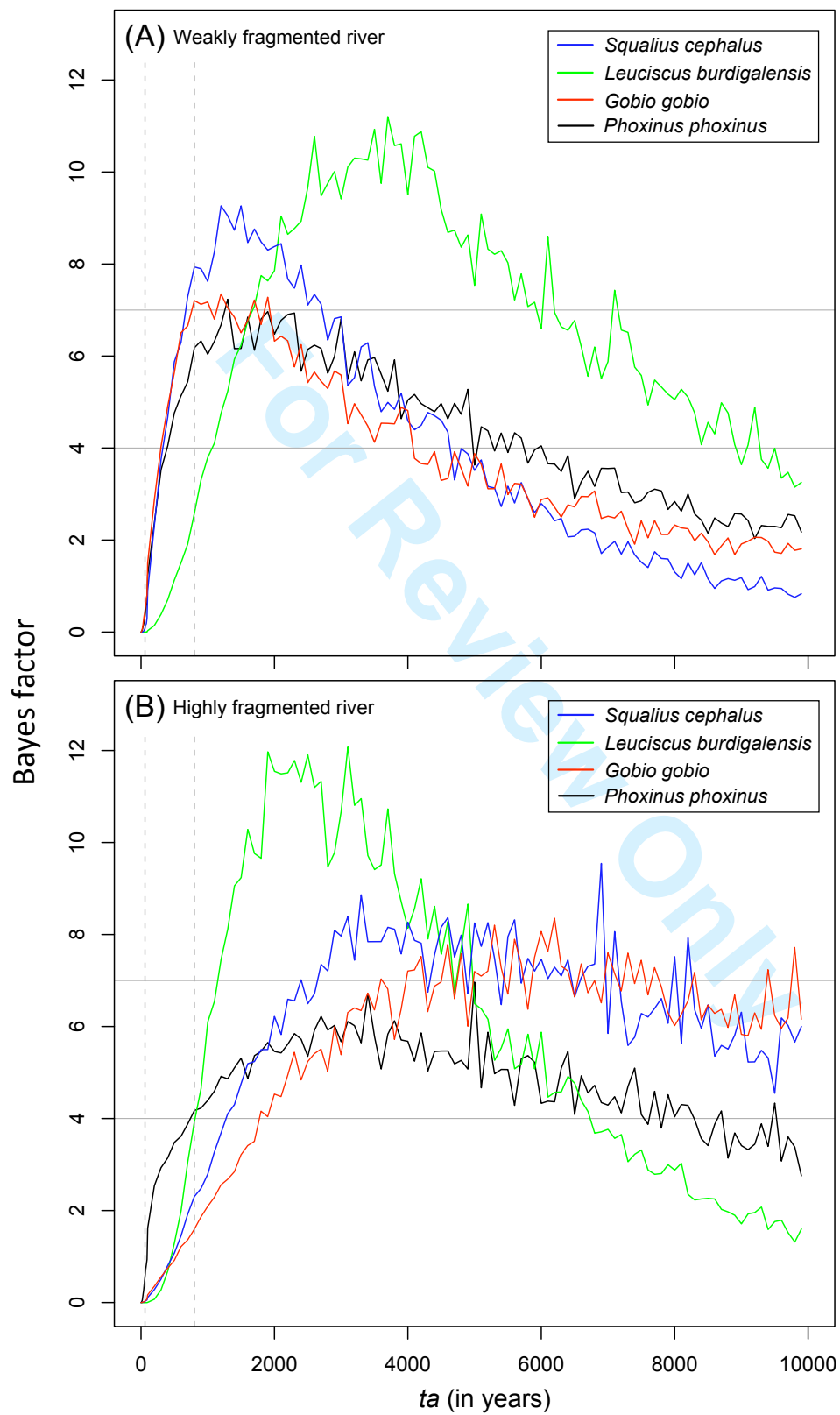
only

741 **FIGURE 5**

742

743

744

745 **FIGURE 6**

746