When new human-modified habitats favour the expansion of an amphibian pioneer species: Evolutionary history of the natterjack toad (Bufo calamita) in a coal basin

Leslie Faucher, Laura Hénocq, Cédric Vanappelghem, Stéphanie Rondel, Robin Quevillart, Sophie Gallina, Cécile Godé, Julie Jaquiéry, Jean-François Arnaud

PROF. JEAN-FRANCOIS ARNAUD (Orcid ID : 0000-0001-5009-7144)

When new human-modified habitats favor the expansion of an amphibian pioneer species: evolutionary history of the natterjack toad (*Bufo calamita)* in a coal basin

Leslie Faucher[1], Laura Hénocq[1], Cédric Vanappelghem[1,2], Stéphanie Rondel[3], Robin Quevillart[4], Sophie Gallina[1], Cécile Godé[1], Julie Jaquiéry[5] & Jean-François Arnaud[1*]

[1] Univ. Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France

[2] Conservatoire d'espaces naturels du Nord et du Pas de Calais, Lillers, France

[3] Centre Permanent d'Initiatives pour l'Environnement - Chaîne des Terrils, Loos-en-Gohelle, France

[4] Groupe ornithologique et naturaliste du Nord – Pas-de-Calais, Lille, France

[5] Université de Rennes 1, UMR CNRS 6553 – ECOBIO, 35042 Rennes Cedex, France

* Corresponding author

E-mail: jean-francois.arnaud@univ-lille1.fr

Abstract

Human activities affect microevolutionary dynamics by inducing environmental changes. In particular, land cover conversion and loss of native habitats decrease genetic diversity and jeopardize the adaptive ability of populations. Nonetheless, new anthropogenic habitats can also promote the successful establishment of emblematic pioneer species. We investigated this issue by

examining the population genetic features and evolutionary history of the natterjack toad (*Bufo [Epidalea] calamita*) in northern France, where populations can be found in native coastal habitats and coalfield habitats shaped by European industrial history, along with an additional set of European populations located outside this focal area. We predicted contrasting patterns of genetic structure, with newly settled coalfield populations departing from migration-drift equilibrium. As expected, coalfield populations showed a mosaic of genetically divergent populations with short-range patterns of gene flow, and native coastal populations indicated an equilibrium state with an isolation-by-distance pattern suggestive of post-glacial range expansion. However, coalfield populations exhibited (i) high levels of genetic diversity, (ii) no evidence of local inbreeding or reduced effective population size, and (iii) multiple maternal mitochondrial lineages, a genetic footprint depicting independent colonization events. Furthermore, Approximate Bayesian Computations suggested several evolutionary trajectories from ancient isolation in glacial refugia during the Pleistocene, with biogeographical signatures of recent expansion probably confounded by human-mediated mixing of different lineages. From an evolutionary and conservation perspective, this study highlights the ecological value of industrial areas, provided that ongoing regional gene flow is ensured within the existing lineage boundaries.

Introduction

Anthropogenic land-use changes are by far the most important contemporary cause of native habitat loss and landscape fragmentation (Fischer & Lindenmayer 2007). The loss of habitats and isolation of populations can decrease levels of gene flow among populations and reduce the effective population size (N_e), two key components of population genetic variability (Baguette *et al.* 2013; Ellegren & Galtier 2016). This is a major threat to a wide array of taxa, especially amphibians that are the vertebrate lineage with the highest number of species facing extinction (Beebee & Griffiths 2005; Wake & Vredenburg 2008; Allentoft & O'Brien 2010; Rivera-Ortíz *et al.* 2015). The vulnerability of amphibians facing habitat loss is attributed to numerous factors including relatively limited dispersal capabilities, high breeding site fidelity, a complex life cycle requiring different types of habitats (aquatic habitats for breeding period and/or larvae stages and terrestrial habitats otherwise) often combined with narrow habitat tolerance, and generally small N_e with high demographic fluctuations (Beebee & Griffiths 2005; McCartney-Melstad & Shaffer 2015; Lourenço *et al.* 2017).

Anthropogenic habitats can differ from native habitats with respect to their colonization history and may have different evolutionary effects (Alberti 2015). Urban environments can host trapped and isolated populations. In this context, recent studies investigated the strength of genetic drift induced by reduced functional connectivity across an urban matrix (*e.g.* Munshi-South *et al.* 2013; Beninde *et al.* 2016; Furman *et al.* 2016; Lourenço *et al.* 2017). However, it appears that new anthropogenic habitats can also promote the establishment of pioneer species in highly disturbed areas (Amici *et al.* 2015; Flavenot *et al.* 2015). This is particularly true for post-industrial environments such as former mining areas that may act as biodiversity refugia (Brock *et al.* 2007). Nonetheless, studies dealing with population genetic features in such *a posteriori*-colonized environments are very scarce and, to the best of our knowledge, only focused on a few plant species (Brock *et al.* 2007; Esfeld *et al.* 2008; Prinz *et al.* 2009; Pauwels *et al.* 2012). Our work aims at

bridging this gap by assessing the current genetic patterns and evolutionary history of an amphibian species in natural and reclaimed coal mine environments.

In northern France, three centuries of coal extraction has led to high habitat fragmentation and the formation of spoil heaps, *i.e.* artificial hills constituted by mineral materials extracted with the coal. Spoil heaps (some of which cover 90 ha and exceed 140 m in height) have thin and nutrient-poor soil. Since the cessation of mining activities in the late 1980s, spoil heaps have constituted new continental xeric primary ecosystems suitable for pioneer species, as well as a specialized flora adapted to anthropogenic metal stress (Godin 2002; Lemoine 2012; Pauwels *et al.* 2012). Presented as a negative symbol for a long time, the spoil heaps of northern France are the latest addition to UNESCO's world heritage list and landowners, agencies and volunteer associations all work to protect this cultural and natural heritage and to improve habitats for biodiversity and wildlife monitoring (Lemoine 2012).

In this study, we investigated the population genetic structure of the natterjack toad (*Bufo* [*Epidalea*] *calamita,* Laurenti, 1768) that established in this new human-modified habitat, and examined the related historical process of colonization. The natterjack toad is a protected pioneer anuran species characterized by a wide geographical distribution ranging from central and western Europe and extending northward to Sweden (Rowe *et al.* 2006). This species generally establishes in nutrient-poor environments with open and unshaded light sandy soils, such as coastal dunes, lowland heaths, semi-deserts, and meadows or marshes (Beebee & Denton 1996; Denton *et al.* 1997; Wilkinson & Griffiths 2013). Species occurrence declines in the northern parts of its range where populations are more fragmented and isolated (Allentoft *et al.* 2009). Based on historical records dating back to 1878 and current mapping within the coalfield areas of northern France, the natterjack toad was historically present in inland semi-natural habitats in the direct vicinity of mining areas. Subsequently, this species successfully settled on the slopes and in the surroundings of abandoned spoil heaps that constitute suitable breeding and resting habitats with temporary ponds (Godin 2002). Today, most historical semi-natural populations previously located in the neighborhood of the spoil heaps became extinct or undetectable (see Fig. S1, Godin 2002; Lemoine 2012). Outside the former coalfields, there are numerous populations established along the coastline area in coastal sand dunes and saltmarshes known to be their "native habitats".

We conducted an extensive population genetic structure analysis based on both nuclear and mitochondrial data. We performed an exhaustive sampling of natterjack toad populations located in the anthropogenic habitats of the mining area and the natural coastal areas of northern France. Our main objectives were (i) to determine whether coastal natural and inland coalfield populations differ in terms of patterns of genetic structure and (ii) to assess the colonization and evolutionary history of inland coalfield populations. Owing to founder effects during the establishment of coalfield populations and to high habitat fragmentation impeding gene flow, our expectations were twofold. First, we expected reduced migration and increased drift for mining area populations compared to native coastal populations, with higher levels of genetic differentiation and lower levels of genetic diversity. Next, we hypothesized a common ancestry between coastal natural populations and mining area populations, the latter being thought to be founded by natural colonization events and/or accidental translocations of individuals during former mining activities from previous semi-natural inland habitats hosting the natterjack toad (Godin 2002). Lastly, we discussed the

evolutionary implications and validity of our findings for the long-term management of pioneer species in anthropogenic areas.

Materials and methods

Study area

Northern France is a broad open plain with, in the west, a shoreline composed of cliffs and wide beaches sheltering wildlife protection areas and, in the east, inland areas mainly composed of anthropogenic habitats (Fig. 1A and Fig. S1). The coalfield area of northern France covers 1200 km² from east to west (Fig. 1A). This industrial area illustrates the significant human impact with landscapes shaped over three centuries by coal extraction, which was virtually a mono-industry in this region from the 1700s to the 1900s. Coal activities led to the formation of more than 300 spoil heaps connected by railways along with extensive urban development. Today, the mining area consists of a very fragmented landscape dominated by urban habitats and conventional farming and crossed by several chains of spoil heaps constituting new continental open habitats. Some spoil heaps are still exploited for schist extraction, others have been abandoned or are now managed for leisure or wildlife protection. Natural early successional habitats such as coastal sand dunes are likely to be the primary "native" sites for *B. calamita*. This species has now successfully colonized numerous spoil heaps, potentially from former semi-natural habitats such as marshes or grassland in the neighborhood of coalfield areas (Fig. S1). Our sampling thus focused on the newly colonized mining area and the neighboring coastal area of northern France likely to host native populations. Additional populations located in different areas of the biogeographical range were also collected to calibrate the genetic divergence of coastline and mining area populations of northern France and to shed light on the settlement history of the natterjack toad in this peculiar area (Fig. 1A).

Sampling procedure

Overall, a total of 68 populations (1209 adult individuals) were sampled. We collected DNA from 959 *B. calamita* individuals in 53 different populations in the focal study area in northern France, with a mean (±SD) of 18 ±12 individuals sampled per population: 32 populations (MA-1 to MA-32) were located within the mining area and 21 populations (C-1 to C-21) were located along the coastal area (Fig. 1A, Table 1). To date, no *B. calamita* populations have been recorded between the coastal and coalfield areas, and we consider population sampling to be exhaustive based on reported locations from 20 years ago (Godin 2002, Fig. S1). To assess broad genetic structure patterns and test the origin of northern France populations, 250 additional individuals were collected from 15 populations in other parts of the *B. calamita* geographical distribution with a mean (±SD) of 17 ± 6 individuals per population; 12 of these populations were collected in other regions of France, 2 in Switzerland and 1 in Sweden (see Fig. 1A and Table 1). Individuals were caught by night during the breeding period (from March to June). Each locality was sampled once in 2013 or 2014. A non-invasive sampling method was used (Broquet *et al.* 2006). Plain sterile 15SC Copan (Brescia Italia) swabs were used to take samples of buccal cells for each captured individual and were stored

at -20°C or dried prior to DNA extraction in the laboratory. Individuals were immediately released after buccal swabbing.

Laboratorial procedures and data quality control

Whole genomic DNA was extracted from swabs using Macherey-Nagel (Düren, Germany) NucleoSpin® 96 trace kits following the standard protocol outlined in the manufacturer's handbook. To assess nuclear DNA polymorphism, all samples were genotyped at 37 nuclear microsatellite loci (Table 1). Of these loci, 15 are described in Rowe *et al.* (1997, 2000) and Rogell *et al.* (2005) and 22 are recently developed loci, described in Faucher *et al.* (2016) (see Table S1). Amplification procedures (PCR), multiplexing and genotyping were carried out following the standard protocols described in Faucher *et al.* (2016). PCR products were analyzed using an ABI Prism 3730xl Analyzer (Applied Biosystems, Foster City, CA) and multilocus genotypes were manually scored using GENEMAPPER 3.7 software. Individuals that failed to amplify or that showed dubious genotypes underwent a second round of PCR. All the 1209 individuals were successfully genotyped at the 37 loci with an overall missing data rate of 1.95%.

Analyses of genetic diversity were applied to 62 populations (out of 68) characterized by a minimum sample size of seven genotyped individuals. The six remaining populations were only used for the Bayesian clustering analyses (see below). Linkage disequilibrium (LD) among pairs of loci and across populations was tested using a log-likelihood ratio statistic described in Raymond & Rousset (1995). Departures from Hardy–Weinberg (HW) equilibrium were tested for hypotheses of either excess or deficit in heterozygotes using a multisample score test, which defines a global test across populations (Rousset & Raymond 1995). In both cases, the Markov Chain method implemented in the software GENEPOP v4.4.3 (Rousset 2008) provided unbiased estimates of probability using the following parameters: 10,000 dememorizations, 1000 batches, and 10,000 iterations per batch. Because multiple statistical tests were performed, we applied a sequential Bonferroni correction to *P*-values for a family-wise error rate of $\alpha = 0.05$ (Rice 1989).

Mitochondrial genetic diversity was also investigated using single nucleotide polymorphism (SNP) with a "kompetitive" allele-specific PCR (KASPar) genotyping assay method and primers defined by K Biosciences (Hoddesdon, UK), as described in Faucher *et al*. (2016). Fifteen SNPs located along a total of 2200 bp from the D-loop and the 16S gene and previously found to be polymorphic were surveyed using amplification procedures and detection of polymorphism as in Faucher *et al.* (2016)*.* An average of 11 individuals per population were genotyped (see Table 1). Overall, the 15 SNPs were successfully amplified for 750 individuals on 855 chosen randomly (1.7% missing SNP data per locus). Individuals with SNP missing data were discarded from further analyses. Finally, each SNP combination was treated as a single haplotype because the mitochondrial genome is inherited as a single linkage unit.

Genetic diversity and contemporary Ne

Each population was described by calculating the total number of alleles observed in nuclear microsatellites (A_T), the average number of alleles per nuclear locus (A_n), the observed heterozygosity (H_O), the gene diversity (H_E), and the number of mitochondrial haplotypes (M_T).

Similarly, allelic richness (A_r) and haplotypic richness (M_r) were estimated after accounting for population variation in sample sizes with a rarefaction method and a hierarchical sampling design implemented in the software package HᴘRᴀRᴇ (Kalinowski 2005). As a measure of individual inbreeding, the internal relatedness (IR) was estimated for each individual using the R package Gᴇɴʜᴇᴛ (Coulon 2009).

Long-term and contemporary N_e shapes current levels of genetic diversity and determines the strength of genetic drift (Ellegren & Galtier 2016). N_e was estimated using the LD method described in Waples (2006) and implemented in NᴇEsᴛɪᴍᴀᴛᴏʀ v2.0 (Do *et al.* 2014). This estimator is fairly robust to violations of the closed-population assumption when migration rates are moderate (Waples & England 2011). Alleles with frequency less than 5% were excluded and 95% confidence intervals were obtained by jackknifing across loci. N_e estimates are very sensitive to reduced sample sizes, even when a large number of loci are employed (Wang 2016). Populations with sample sizes ≤ 15 were therefore excluded from analyses. Overlapping generations may also bias contemporary N_e estimates which should be interpreted with cautions (Waples *et al.* 2014). Nonetheless, LD estimations of N_e allow reliable relative comparisons across populations (*e.g*. Wang 2012; Lourenço *et al.* 2017).

Population genetic affiliation and levels of genetic differentiation

Population genetic affinities was first described with a neighbor-joining (NJ) tree based on pairwise genetic Cavalli-Sforza and Edwards (1967) distances (D_{CE}), using Pᴏᴘᴜʟᴀᴛɪᴏɴs 1.2.32 software (http://bioinformatics.org/~tryphon/populations/). For a visual assessment of the mitochondrial geographical distribution, haplotype distribution was mapped and genealogical relationships among haplotypes were estimated using a median-joining haplotype network drawn with PᴏᴘᴀRᴛ (http://popart.otago.ac.nz).

Next, we used a non-spatially explicit Bayesian clustering analysis, without any *a priori* population affiliation, as implemented in the program SᴛRᴜᴄᴛᴜRᴇ v 2.3.4 (Pritchard *et al.* 2000; Hubisz *et al.* 2009). This analyze was applied on two data sets: (i) a full data set comprising 68 populations; and (ii) a data set comprising 53 populations sampled in the focal area of northern France (Fig. 1A). We used 500,000 burn-in Markov Chain Monte Carlo (MCMC) iterations followed by 5.10^6 MCMC iterations post-burn-in, carried out on 30 replicates for each value of the tested number of clusters (K) ranging from 2 to 53 or 68 (depending on the focal population data set surveyed). The most likely *K* was evaluated using the calculation of the *ad hoc* statistic *ΔK* following Evanno *et al.* (2005). Cʟᴜᴍᴘᴘ v 1.1 (Jakobsson & Rosenberg 2007) and DɪsᴛRᴜᴄᴛ v 1.1 (Rosenberg 2004) were used to identify the most likely clustering solution among replicated runs and for graphical display, respectively. As suggested by Puechmaille (2016), we controlled for the effect of uneven sampling schemes: additional analyses were performed with even sample sizes per geographic cluster previously observed. For this, 22 populations (mean 6 per genetic cluster, totalizing 376 individuals) were randomly chosen. As a complementary analysis free of any genetic assumptions, a spatial principal component analysis (sPCA, Jombart *et al.* 2008) was also performed as described in Fig. S5.

Genetic differentiation was assessed using *F*-statistics according to the Weir & Cockerham (1984) ANOVA procedure on nuclear data across loci and populations using SPAGEDI version 1.4 (Hardy & Vekemans 2002). Mono- and multilocus intra-population fixation indices (F_{IS}), and inter-population fixation indices (F_{ST}) were estimated and tested for significance using 10,000 permutations. Mitochondrial genetic differentiation was further assessed using G_{ST}, an estimator equivalent to F_{ST} (Pons & Petit 1996), also using SPAGEDI and tested for significance with 10,000 individual permutations among populations. To evaluate how observed genetic structure is shaped by contemporary gene flow or is a vestige of past historical events, mean mitochondrial G_{ST} estimates were compared to mean N_{ST} estimates, an index of genetic differentiation accounting for phylogenetic distances between mitochondrial haplotypes (Pons & Petit 1996). Likewise, F_{ST} estimates were also compared to the index of genetic differentiation R_{ST} which is based on variance in allelic size (Slatkin 1995). N_{ST} and R_{ST} estimates were compared to G_{ST} and F_{ST} estimates using 10,000 permutations of genetic distance between haplotypes and of allele size respectively, using SPAGEDI.

Statistical comparison of genetic parameters and spatial genetic structure (SGS) patterns

We tested for significant differences in mean genetic diversity (H_E, H_O, A_r, M_r), N_e estimates, mean *IR* and mean pairwise F_{ST} among the main inferred genetic clusters (see above) in the focal study area in northern France. One-way analysis of variance (ANOVA) was used in cases where the data followed normality and homoscedasticity, while a non-parametric Kruskal-Wallis test was used when assumptions were not met. When statistically significant differences were detected, a *post hoc* Tukey-Kramer HSD's test or a *post hoc* Kruskal multi-comparisons test was performed for ANOVA and Kruskal-Wallis test, respectively. For nuclear genetic diversity, a generalized linear model with loci set as random effect, followed by a multi comparison test on means by Tukey's contrast, was further used to test for significant differences among genetic clusters (R packages, "lme4", Bates *et al.* 2015 and "multcomp", Hothorn *et al.* 2008).

The occurrence of SGS was tested at the individual level by applying spatial autocorrelation analyses as implemented in SPAGEDI. Nason's kinship coefficient F_{ij} (Loiselle *et al.* 1995) was chosen as a pairwise estimator of individual genetic relatedness, as it has robust statistical properties (Vekemans & Hardy 2004). To visualize SGS, F_{ij} values were averaged over a set of 10 distance classes defined to obtain the same number of individual pairs within each distance class. Standard errors of F_{ij} were estimated using a jackknifing procedure over loci. Significance of F_{ij} estimates was calculated by 10,000 permutations of individual locations for the nuclear and the mitochondrial haplotype data set. To quantify and compare the strength of SGS among populations without arbitrarily setting geographical distance intervals, we used the *Sp* statistics (Vekemans & Hardy 2004). Caution is also required when interpreting the extent of positive SGS displayed by the first x-intercept of correlograms (Peakall *et al.* 2003; Favre-Bac *et al.* 2016). To provide an approximation of the true extent of detectable positive SGS, we further performed analyses of cumulative distance classes by increasing distance class sizes spanning the minimum distance between samples to the maximum distance of sampling, as described in Peakall *et al.* (2003). Finally, isolation by distance (IBD) patterns were investigated using polynomial regression and Mantel tests of association as described in Oden

& Sokal (1986) and Smouse *et al.* (1986) to estimate the significance of relationships between pairwise D_{CE} genetic distances and pairwise geographical distances among populations.

Evolutionary history of populations in the focal study area

Based on the genetic clusters depicted in the data set, approximate Bayesian computations (ABC) analyses were used to determine the historical process involved in the settlement of populations. We evaluated whether the observed genetic variability in northern France resulted from (i) an eastern continental origin, (ii) a western coastal origin, or (iii) another unsampled native locality. The objective was also to discriminate between recent human-mediated translocations and colonization from populations historically located in inland habitats in the neighborhood of the mining area. Divergence times were estimated to gain further insights into evolutionary history chronology.

Scenarios were evaluated in a two-step analysis. A first set of scenarios (set A) aimed to analyze the origin of each target genetic unit in northern France by using the nuclear genetic units depicted outside the focal study area in northern France as putative sources (Fig. 2A). Based on the results yielded by set A, a second set of scenarios (set B) aimed to depict the relationship among the different target genetic units and to estimate their divergence times taking into account all the observed genetic units (Fig. 2B). Overall, three demographic scenarios were defined for each analysis step following each of the three hypotheses on population origins. ABC analyses were computed using the software DIYABC v.2 (Cornuet *et al.* 2014).

For all scenarios, an identical set of microsatellite loci was simulated. Due to computational limitations and the infinite number of possible scenarios when numerous populations are considered, inferences were based on finite set of geographically delimited groups. These genetically homogenous groups were defined using Bayesian clustering analysis (see above). In each group of the focal study area, populations with the smallest nuclear pairwise genetic differentiation index (F_{ST} < 0.075) were selected (N=95, N=132 and N=65 for coastal, western coalfield and eastern coalfield populations, respectively) to represent the whole group for comparisons with data simulated under the different scenarios, as suggested in Lombaert *et al.* (2014). For the genetic units outside the focal study area, all individuals belonging to each external genetic clusters were included to guarantee sufficient sample size (N= 94 in western lineage and N= 103 in eastern lineage). The origin(s) of target genetic units in the focal study area was then tested based on adjacent groups and by simplifying the scenarios as much as possible to avoid poor estimation of parameters (Bertorelle *et al.* 2010). Assuming post-glacial recolonization hypothesis with leading-edge expansions along the Atlantic coastline (Rowe *et al.* 2006), scenarios were set with divergence time in generations (t_1 to t_4), putative ancestral effective population size (N_a), current effective population size estimates (N_1 to N_5) and effective sizes of founder populations (N_{f1}, N_{f2}, N_{f3}, N_{f4}) (see Table S2). Analyses were performed with historical, demographic and mutational parameter values drawn from the prior distributions described in Table S2 (*e.g.*, Cornille *et al.* 2013; Dussex *et al.* 2014; Lombaert *et al.* 2014). Detailed steps of each analysis are described in Table S2.

Results

Genetic diversity and contemporary Ne

Among the 37 nuclear microsatellite loci, 2 were monomorphic (Buca3 and BC38, see Table S1). Exact tests demonstrated no LD among loci. Two loci, Bcalµ9 and Bc22, were excluded due to technical artifacts (*i.e.,* null alleles and/or allele drop-out). The remaining loci perfectly matched expectations of HW proportions, with A_r ranging from 1.59 to 5.01 (Table S1). Levels of intra-population genetic diversity were very variable with the Swedish population being characterized by the lowest level of nuclear and mitochondrial diversity (Table 1). Based on the 35 nuclear microsatellite loci used in subsequent analyses, multilocus A_T ranged from 52 alleles to 157 alleles per population. Sixteen out 62 populations showed an overall significant deviation from HW equilibrium, which was due to only few significant single-locus effects, involving either excess or deficit in heterozygotes (Table 1). Effective population sizes ranged from 7 to 1506, but were mostly comprised between 10 and 60, with a mean of 93 and a median of 45 (Table 1).

In terms of mitochondrial genetic diversity, 21 haplotypes were found with three major haplotypes accounting for 58% of the mitochondrial diversity, and other haplotypes differing from one of these major haplotypes by one to four mutation steps. Haplotypes were closely related because only 16% of the haplotype pairwise comparisons indicated haplotypic divergence of more than five mutation steps (Fig. 3A). Over the whole dataset, populations exhibited from one to five different haplotypes (Table 1 and Fig. 3B). Only one major haplotype (Hap 11) was shared by the focal study populations in northern France and the other populations (Fig. 3B). Interestingly, of the 12 haplotypes which were exclusively observed in the focal study area, 8 haplotypes occurred only in coalfield populations (Fig. 3B).

Population genetic affiliation and levels of genetic differentiation

We found significant genetic differentiation over all sampled populations, with a mean multilocus F_{ST} of 0.177 ±0.008 ($P < 0.001$) for the nuclear data and a mean G_{ST} of 0.487 ($P < 0.001$) for the mitochondrial data. 98.6% of nuclear pairwise F_{ST} and 89% of mitochondrial pairwise G_{ST} were significant, ranging from 0.00 to 0.582 and from 0.00 to 1.000, respectively. Additionally, a significant phylogeographical signal was detected for both nuclear and mitochondrial data when using a random balanced subset of 22 populations over the whole sampled area to avoid bias due to uneven sampling (see Table S3).

Clear genetic partitioning was detected over the whole data set of sampled populations and within the focal study area in northern France. Bayesian clustering, sPCA and NJ tree results concurred, indicating the occurrence of five distinct genetic clusters. Bayesian clustering analysis suggested two levels of structure in the whole data set because the K vs ΔK distribution was multimodal, with a hierarchical structure at $K = 2$ and at $K = 5$ (Fig. S2A). When $K = 2$, the inferred genetic clusters consisted of one group composed of individuals from the focal study area (northern France) and one group of individuals from other regions. This result arose due to uneven sampling efforts. Additional runs involving randomly chosen, more balanced data sets showed a clear partitioning, splitting the subset into three genetic clusters according to geographical location (Fig.

S3). The second modal $K = 5$ obtained in the analysis of the whole data set showed genetic clusters in clear concordance with geographical location (Fig. 1B). Very few admixed individuals occurred within populations: more than 88% of individuals from western France (WF-1 to WF-5) and eastern France (E-1 to E-6) were assigned to their respective cluster with at least 90% membership probability. In the focal study area (northern France), individuals were split among three genetically distinct clusters according to their spatial distribution: coastline populations, western and eastern coalfield populations with admixed populations in the western coalfield cluster (Fig. 1B). Genetic affinities among populations depicted by tree topology (Fig. S4) and sPCA (Figure S5A) unambiguously confirmed this pattern. Furthermore, mitochondrial data mirrored spatial affiliation patterns depicted by nuclear data: each of the five clusters identified were characterized by a specific distribution of haplotypes (Fig. 3B).

When focusing only on the focal study area in northern France, Bayesian analyses confirmed the genetic divergence among coastline, western coalfield and eastern coalfield populations and yielded identical patterns of clustering, with two major modes at $K = 2$ (coastal *versus* coalfield populations) and $K = 3$ (coastal, western and eastern coalfield populations) (Fig. S1B). The sPCA synthetic map of the first three global scores distinguished identical population affiliation along with a further genetic break between southern and northern littoral populations (Fig. S5B). An overall significant genetic differentiation still occurred with a mean multilocus F_{ST} estimate of 0.110 ±0.009 ($P< 0.001$) and $G_{ST} = 0.403$ ($P< 0.001$). No phylogeographical signal was detected within the focal study area in northern France: R_{ST} (0.109; $P< 0.001$) and N_{ST} (0.413; $P< 0.001$) did not significantly differ from the F_{ST} and G_{ST} (Table S3).

Statistical comparison of genetic parameters and SGS patterns

Levels of genetic diversity and inbreeding were compared among the three groups defined within the focal study area in northern France. No significant difference occurred regarding F_{IS} or A_r (Fig. 4 A, B). Similarly, no significant difference was observed in terms of single-sample N_e estimates (Fig. 4C, Kruskal Wallis; $\chi^2 = 6.2472$; $P > 0.05$). Nonetheless, as showed by H_e, H_o and M_r estimates, coastal populations appeared to be significantly less genetically diverse compared with eastern coalfield populations or both eastern and western coalfield populations (ANOVA, $\chi^2 = 27.664$, df=2, $P < 0.001$, ANOVA, $\chi^2 = 24.305$, df=2, $P < 0.001$, Kruskal Wallis; $\chi^2 = 11.441$; $P < 0.01$ respectively, Fig. 4 D, E and F). Coastal populations also showed a significantly higher -individual inbreeding level (ANOVA; $F = 9.184$; $P < 0.001$) compared with coalfield populations (Fig. 4 G). Trends for lower mean pairwise nuclear F_{ST} but higher mean mitochondrial pairwise G_{ST} were observed in coastal populations (ANOVA; $F = 3.624$; $P < 0.05$ and Kruskal Wallis; $\chi^2 = 16.637$; $P < 0.001$, Fig. 4 H and I).

SGS depicted by the correlograms in Figure 5 showed a consistent pattern of decreasing genetic relatedness with increasing geographical distance. Strength of SGS was 5 to 60-fold higher for mitochondrial DNA data set compared with the nuclear DNA data set (Fig. 5, A, B, C, D, E, and F). There was a considerable difference in the extent of SGS between coastal populations and inland populations. Nuclear kinship coefficient dropped to zero at 40 km until reaching significant negative value around 80 km in the case of coastal populations whereas this drop occurred at much shorter spatial scale (5 km) in the two coalfield clusters (Fig. 5, A, B and C). Similar patterns hold for

mitochondrial data, especially for coalfield populations for which a more compact spatial-autocorrelation pattern was found (Fig. 5, D, E and F). Analyses of cumulative distance classes unambiguously confirmed this contrasting extent of positive SGS (Fig. S6). A significant IBD both along the coastline and within the mining area was observed (r_z = 0.514 along the coastline, r_z = 0.693 in western coalfield and r_z=0.516 in eastern coalfield; all at P< 0.001, Fig. 5 G, H and I). Linear relationships clearly hold only at very short geographical distances in the coalfield area (< 5 km) whereas IBD extends over larger geographical distances along the coastline.

Evolutionary history of populations in the focal study area

Pre-evaluation of prior-parameter combinations and results of model checking can be found in Table S5 and Fig. S7. For the first sets of scenarios aiming to evaluate the putative origins of the genetic units observed in northern France, the relative posterior probabilities calculated for each scenario provided the strongest statistical support for scenario A1, suggesting a common western origin for the three northern lineages (P (L) = 0.976, 95% CI [0.967-0.985], P (L) = 0.948, 95% CI [0.936 – 0.960], P (L) = 0.949, 95% CI [0.908-0.990] for the coastal, the western and the eastern coalfield areas, respectively) (see Table S4 for detailed results). The low type II error rate calculated for these most likely scenarios suggest that these probability estimations are robust, with a low probability of choosing a false scenario.

The relative posterior probabilities calculated for each scenario of simulation set B, that aimed at deciphering the relationships among the different genetic units of Northern France, provided the strongest statistical support for scenario B1 (P(L) = 0.913, 95% CI [0.892-0.934]), which also showed the lowest type II error rate (Fig. 2 B). This scenario suggests that the eastern cluster diverged 300,000 years ago (Table S4 and S5), followed by the western cluster, during the last glacial maximum (138,000 years ago). Finally, the coastal genetic cluster divergence was dated to 52,000 years ago and the within-mining area divergence shaping the eastern and western genetic differentiation was estimated to 14,000 years ago.

Discussion

Intraspecific genetic variation is a key component of biodiversity: it determines ecological and evolutionary outcomes such as the potential of a given species to respond to environmental changes associated with human activities (DiBattista 2008; Alberti 2015; Mimura *et al.* 2017). Recent studies investigated the impact of anthropogenic land conversion, like urbanization or quarrying activities, on genetic diversity (*e.g.* Flavenot *et al.* 2015; Beninde *et al.* 2016; Lourenço *et al.* 2017). However, the effect of post-industrial areas such as coalfields on patterns of genetic structure remains to be investigated. To the best of our knowledge, among the very few population genetics studies performed in such human-modified environments (*e.g.* Brock *et al.* 2007; Esfeld *et al.* 2008), this work is the first describing in detail current genetic patterns for a vertebrate species. We found that coalfield areas shelter distinct genetic lineages of the natterjack toad, and that these populations are also characterized by different patterns of genetic structure compared to

populations found in natural habitats. Our findings and their evolutionary and conservation implications are discussed below.

Level and spatial arrangement of genetic variability in anthropogenic and natural areas

Populations established in urbanized fragmented and/or human-modified colonized habitats may display lower levels of genetic diversity, higher inbreeding and lower contemporary N_e, especially for amphibian species (McCartney-Melstad & Shaffer 2015; Rivera-Ortíz *et al.* 2015). For instance, Munshi-South *et al.* (2013) found a significant decline in genetic diversity in urban populations of the Dusky salamander in New York City. In contrast, Furman *et al.* (2016) found no evidence of genetic erosion and increased genetic differentiation in natural and artificially constructed wetlands in a wood frog, despite extensive urbanization. In the same way, isolated urban populations of the Fire salamander did not exhibit lower levels of genetic diversity and showed similar N_e compared to rural populations (Lourenço *et al.* 2017).

In our study, natural coastal populations of natterjack toad met a regional gene flow/genetic drift equilibrium with a continuous IBD pattern over increasing spatial scales. Conversely, a positive SGS was observed at much shorter spatial scales in inland coalfields areas. This highlights a different regional dynamics, with genetic drift being much more influential than gene flow beyond a spatial scale of 5 km. Such a departure from gene flow/drift equilibrium is found in a wide array of taxa (*e.g.* Hutchison & Templeton 1999; Austin *et al.* 2004; Hänfling & Weetman 2006; Phillipsen *et al.* 2015). Besides, our results are supported by higher levels of nuclear and mitochondrial genetic differentiation among coalfield populations compared with the coastal populations over a similar geographical area. Overall, this pattern supports a mosaic of genetically distinct populations with genetic breaks occurring over a few kilometers probably due to the unfavorable habitat (highly urbanized or intensive farming) surrounding spoil heaps in coalfield areas. Such pronounced SGS is a common feature in reptilian and amphibian populations where natural or human-induced dispersal barriers are prevailing factors in disrupting gene flow (Vos *et al.* 2001; Arioli *et al.* 2010; Munshi-South *et al.* 2013; Sotiropoulos *et al.* 2013; Beninde *et al.* 2016).

However, despite coalfield populations being isolated due to habitat fragmentation, we observed no decrease in the levels of gene diversity in natterjack toad. In contrast, coalfield populations showed higher levels of genetic diversity and lower inbreeding levels compared with natural coastal populations. Moreover, we did not detect any differences in terms of contemporary N_e between natural coastal and coalfield populations, a result also found in Lourenço *et al*. (2017) when comparing Fire salamanders in rural and urbanized environments. Our N_e estimates ranged from 7 to 600 and were comparable to effective population sizes observed in other pond-breeding amphibians located outside urbanized environments, like Yosemite toads or Fire salamanders (Wang *et al.* 2011; Wang 2012; Lourenço *et al.* 2017). Our N_e estimates were also higher than those estimated by Beebee (2009) in a range of natural British and European continental *B*. *calamita* populations. Inland coalfield estimated values of N_e were even in the upper range of N_e for amphibian populations whose values are usually expected to be under 100 and frequently closer to 10 (reviewed in Phillipsen *et al.* 2011). Altogether, our results support the conservation value of post-industrial areas because they host populations with substantial levels of genetic diversity and consistent contemporary N_e. Our results are very similar to the few studies conducted on similar

environments: plant species living in former mining areas, like Orchidaceae, Compositaea, Chenopodiacea or Brassicaceae, also exhibited a high genetic variation that did not differ from natural populations (Brock *et al.* 2007; Esfeld *et al.* 2008; Prinz *et al.* 2009; Pauwels *et al.* 2012).

Whereas lower levels of genetic diversity and higher inbreeding levels in natural coastline populations can be attributed to postglacial leading-edge colonization that involves strong genetic drift (Rowe *et al.* 2006; Slatkin & Excoffier 2012), the high levels of genetic diversity within the coalfield area may be a result of independent colonization events from surrounding populations. Indeed, colonizers from multiple sources are known to enhance intra-population genetic diversity (Whitlock & McCauley 1990). Among the few available comparable studies, this process of colonization has been suggested in *Suaeda maritima*, a halophyte plant species that repeatedly colonized former potash mining dumps in Germany (Prinz *et al.* 2009, see also Esfeld *et al.* 2008). In our case study, *B. calamita* is known to be a pioneer species with high dispersal capabilities (Miaud *et al.* 2000; Sinsch *et al.* 2012). Thus, the most parsimonious hypothesis is that toads successfully dispersed from natural populations initially established in former semi-natural habitats (humid meadows or pastures) located in the vicinity of the coalfields (Godin 2002). Additionally, accidental human-mediated introductions of individuals from unknown sources, during mining activities or currently by individual translocations, cannot be ruled out, as observed in the European tree frog (Andersen *et al.* 2004). The exceptional levels of mitochondrial diversity found in the two coalfield entities reinforce the likelihood of multiple waves of immigration events from diverse sources. Together, the occurrence of distinct maternal lineages and the absence of equilibrium between gene flow and drift suggest an establishment of colonizers from previous semi-natural areas surrounding the coalfields, potentially facilitated by trade operations related to mining activities.

Population evolutionary history

Local interactions between the micro-evolutionary processes of gene flow and genetic drift do not exclude biogeographic historical processes as key factors shaping the current genetic structure (*e.g.*, Austin *et al.* 2004; Dussex *et al.* 2014). Indeed, it is now well established that Pleistocene climate cycling had a major role in driving the present patterns of genetic structure for several taxa in Europe (Hewitt 1999; Stewart *et al.* 2010). The divergence time estimates in each most likely ABC scenario were consistent and our results suggest a common history but distinct subsequent re-colonization pathways of each evolutionary unit located in the focal area (northern France). The three northern genetic units and the eastern and the western genetic units located south of the focal study area diverged during the Pleistocene from around -300,000 to -14 000 years ago. Different glacial refugia during climatic cooling drove population genetic divergence following a south-to-north pathway: firstly an eastern genetic unit split, secondly a western genetic unit split, thirdly a northern coastal genetic unit split and finally a recent split of the eastern and western coalfield genetic units in the last 14 000 years. Several studies in a wide array of taxa have already raised the hypothesis of multiple European glacial refugia during the Pleistocene (Hewitt 1999; Cornille *et al.* 2013; Havrdovà *et al.* 2015; Ursenbacher *et al.* 2015). Rowe *et al.* (2006) suggest that local but relatively recent and short-lived glacial refugia, located in northern France and central-eastern Europe from 20,000 to 9000 years ago, explain the lineages observed throughout the natterjack toad geographical range distribution. These long-term isolation events in different refugia

may explain the observed divergence between northern genetic units and the western genetic unit. Our salient result was the common ancestry of the three genetic units depicted in the focal study area, dating back around 50,000 years ago. Former inland semi-natural populations that subsequently established in the neighboring coalfield areas have therefore a common history with natural coastal populations. The timescale is however too short to get more precise inferences using ABC analyses. It should also be noted that the occurrence of multiple mtDNA lineages in this post-glacial expansion area may be suggestive of an impact of recent human activities that mixed different post-glacial lineages, which may inflate time divergence estimates (Lombaert *et al.* 2014). Future studies involving assignment tests and a comprehensive sampling across the species' range are required to gain further insights into the evolutionary history of observed lineages.

Evolutionary and conservation implications

Ignoring the nature of intraspecific genetic variation in management decisions can lead to negative consequences for biodiversity conservation (Allentoft & O'Brien 2010; McCartney-Melstad & Shaffer 2015; Mimura *et al.* 2017). We depicted clear genetic boundaries in contrasting habitats, which can be used for defining conservation and management units. Our results (large levels of genetic diversity and N_e) suggest that the erosive effect of genetic drift is not currently a major threat for natterjack populations in coalfield areas. Based on the few comparable studies involving plant populations established in post-mining areas, Prinz *et al.* (2009) argued that this may be the rule rather than an exception. Hence, former mining areas can act as refugia areas and offset natural habitat loss for pioneer species living in early successional habitats (Esfeld *et al.* 2008). Altogether, these results highlight the value of former mining areas for the long-term conservation of animal and plant pioneer species, as found for other non-urbanized areas such as quarries (*e.g.* Flavenot *et al.* 2015). Highly genetically diverse natterjack populations established in the mining areas have probably benefited from accidental human-mediated transport, lower predation risk and/or reduced competition with the common toad (Beebee & Denton 1996). For natural coastal populations, additional investigations are needed to determine whether population reinforcement is required. Nevertheless, transfers across such ecologically well-differentiated habitats (coastal *vs* coalfield habitats) should be discouraged to avoid possible outbreeding depression. Indeed, local adaptation to coalfield environment conditions might already have been established (*e.g.* Pauwels *et al.* 2012 in Brassicaceae). The challenge is now to replace the past colonizer and migrant flow by ongoing effective regional gene flow because we clearly demonstrated non-equilibrium conditions with gene flow occurring only at short spatial scales (less than 10 km). The protection of spoil heaps and surrounding open habitats is essential for the conservation of viable populations. However, the whole landscape configuration remains a key factor for the long-term management of connecting pathways among populations living in such exceptional areas inherited from European industrial history.

Data Accessibility

- Nuclear mitochondrial DNA sequences: GenBank accessions KX237573–KX237594.

- Mitochondrial DNA sequences used to define SNP markers: GenBank accessions KX237595–KX237630.

- Coordinates of sample locations, nuclear and mitochondrial genetic data: DRYAD: doi:10.5061/dryad.b89k7.

Author Contributions

JFA, CV, SR and JJ conceived and designed the study. LF, SR, RQ, JJ, LH and JFA conducted the sampling. LF, LH and CG genotyped the samples. LF, CG and JFA designed the new nuclear microsatellite loci and mitochondrial SNPs. LF, JFA, SG and JJ analyzed the data. LF and JFA wrote the manuscript. All authors provided editorial comments.

References

Alberti M (2015) Eco-evolutionary dynamics in an urbanizing planet. *Trends in Ecology and Evolution*, **30**, 114–126.

Allentoft ME, O'Brien J (2010) Global amphibian declines, loss of genetic diversity and fitness: a review. *Diversity*, **2**, 47–71.

Allentoft ME, Siegismund HR, Briggs L, Andersen LW (2009) Microsatellite analysis of the natterjack toad (*Bufo calamita*) in Denmark: populations are islands in a fragmented landscape. *Conservation Genetics*, **10**, 15–28.

Amici V, Landi S, Frascaroli F *et al.* (2015) Anthropogenic drivers of plant diversity: perspective on land use change in a dynamic cultural landscape. *Biodiversity and Conservation*, **24**, 3185–3199.

Andersen LW, Fog K, Damgaard C (2004) Habitat fragmentation causes bottlenecks and inbreeding in the European tree frog (*Hyla arborea*). *Proceedings of the Royal Society of London, Series B*,

271, 1293–1302.

Arioli M, Jakob C, Reyer H-U (2010) Genetic diversity in water frog hybrids (*Pelophylax esculentus*) varies with population structure and geographic location. *Molecular Ecology*, **19**, 1814–1828.

Austin JD, Lougheed SC, Boag PT (2004) Controlling for the effects of history and nonequilibrium conditions in gene flow estimates in northern bullfrog (*Rana catesbeiana*) populations. *Genetics*, **168**, 1491–1506.

Baguette M, Blanchet S, Legrand D, Stevens VM, Turlure C (2013) Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, **88**, 310–326.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Beebee TJC (2009) A comparison of single-sample effective size estimators using empirical toad (*Bufo calamita*) population data: genetic compensation and population size-genetic diversity correlations. *Molecular Ecology*, **18**, 4790–4797.

Beebee TJC, Denton JS (1996) *The natterjack toad conservation handbook (Species recovery program)*. English Nature, Northminster House, Peterborough PE1 1UA.

Beebee TJC, Griffiths RA (2005) The amphibian decline crisis: a watershed for conservation biology? *Biological Conservation*, **125**, 271–285.

Beninde J, Feldmeier S, Werner M *et al.* (2016) Cityscape genetics: structural vs. functional connectivity of an urban lizard population. *Molecular Ecology*, **25**, 4984–5000.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Brock J, Aboling S, Stelzer R, Esch E, Papenbrock J (2007) Genetic variation among different populations of Aster tripolium grown on naturally and anthropogenic salt-contaminated habitats: implications for conservation strategies. *Journal of Plant Research*, **120**, 99–112.

Broquet T, Berset-Braendli L, Emaresi G, Fumagalli L (2006) Buccal swabs allow efficient and reliable microsatellite genotyping in amphibians. *Conservation Genetics*, **8**, 509–511.

Cavalli-Sforza L., Edwards AW. (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Cornille A, Giraud T, Bellard C *et al.* (2013) Postglacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, **22**, 2249–2263.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0 : a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

Coulon A (2009) GENHET: an easy-to-use R function to estimate individual heterozygosity. *Molecular*

Ecology Resources, **10**, 167–169.

Denton JS, Hitchings SP, Beebee TJC, Gent A (1997) A recovery program for the natterjack toad (*Bufo calamita*) in Britain. *Conservation Biology*, **11**, 1329–1338.

DiBattista JD (2008) Patterns of genetic variation in anthropogenically impacted populations. *Conservation Genetics*, **9**, 141–156.

Do C, Waples RS, Peel D *et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (*Ne*) from genetic data. *Molecular Ecology Resources*, **14**, 209–214.

Dussex N, Wegmann D, Robertson C (2014) Postglacial expansion and not human influence best explains the population structure in the endangered kea (*Nestor notabilis*). *Molecular Ecology*, **23**, 2193–2209.

Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nature Reviews Genetics*, **17**, 422–433.

Esfeld K, Hensen I, Wesche K *et al.* (2008) Molecular data indicate multiple independent colonizations of former lignite mining areas in Eastern Germany by *Epipactis palustris* (Orchidaceae). *Biodiversity and Conservation*, **17**, 2441–2453.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Faucher L, Godé C, Arnaud J-F (2016) Development of nuclear microsatellite loci and mitochondrial single nucleotide polymorphisms for the natterjack toad, *Bufo* (*Epidalea*) *calamita* (Bufonidae), using next generation sequencing and Competitive Allele Specific PCR (KASPar). *Journal of Heredity*, **107**, 660–665.

Favre-Bac L, Mony C, Ernoult A, Burel F, Arnaud J-F (2016) Ditch network sustains functional connectivity and influences patterns of gene flow in an intensive agricultural landscape. *Heredity*, **116**, 200–212.

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, **16**, 265–280.

Flavenot T, Fellous S, Abdelkrim J, Baguette M, Coulon A (2015) Impact of quarrying on genetic diversity: an approach across landscapes and over time. *Conservation Genetics*, **16**, 181–194.

Furman BLS, Scheffers BR, Taylor M, Davis C, Paszkowski CA (2016) Limited genetic structure in a wood frog (*Lithobates sylvaticus*) population in an urban landscape inhabiting natural and constructed wetlands. *Conservation Genetics*, **17**, 19–30.

Godin J (2002) Degré de rareté, évolution de la distribution et particularités de l'herpétofaune de la région Nord - Pas de Calais. *Bulletin de la Société Herpétologique de France*, **104**, 16–35.

Hänfling B, Weetman D (2006) Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics*, **173**,

1487–1501.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Havrdovà A, Douda J, Krak K *et al.* (2015) Higher genetic diversity in recolonized areas than in refugia of *Alnus glutinosa* triggered by continent-wide lineage admixture. *Molecular Ecology*, **24**, 4759–4777.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biometrical Journal*, **50**, 346–363.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kalinowski ST (2005) HP-RARE 1.0 : a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Lemoine G (2012) Interêt, gestion, protection et valorisation des terrils miniers dans le Nord–Pas-de-Calais. *Annales de la Société Géologique du Nord*, **19**, 1–8.

Loiselle BA, Sork VL, John N, Graham C (1995) Spatial genetic structure of a tropical understory shurb, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Lombaert E, Guillemaud T, Lundgren J *et al.* (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Molecular Ecology*, **23**, 5979–5997.

Lourenço A, Álvarez D, Wang IJ, Velo-Antón G (2017) Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Molecular Ecology*, **26**, 1498–1514.

McCartney-Melstad E, Shaffer HB (2015) Amphibian molecular ecology and how it has informed conservation. *Molecular Ecology*, **24**, 5084–5109.

Miaud C, Sanuy D, Avrillier J-N, Sanuy D, Miaud C (2000) Terrestrial movements of the natterjack

toad *Bufo calamita* (Amphibia, Anura) in a semi-arid, agricultural landscape. *Amphibia-Reptilia*, **21**, 357–369.

Mimura M, Yahara T, Faith DP *et al.* (2017) Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, **10**, 121–139.

Munshi-South J, Zak Y, Pehek E (2013) Conservation genetics of extremely isolated urban populations of the northern dusky salamander (*Desmognathus fuscus*) in New York City. *PeerJ*, **1**, e64.

Oden NL, Sokal RR (1986) Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Systematic Zoology*, **35**, 608–617.

Pauwels M, Vekemans X, Godé C *et al.* (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytologist*, **193**, 916–928.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes. *Evolution*, **57**, 1182–1195.

Phillipsen IC, Funk WC, Hoffman EA, Monsen KJ, Blouin MS (2011) Comparative analyses of effective population size within and among species: ranid frogs as a case study. *Evolution*, **65**, 2927–2945.

Phillipsen IC, Kirk EH, Bogan MT, Mims MC, Julian D (2015) Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects moderate dispersal. *Molecular Ecology*, **24**, 54–69.

Pons O, Petit R (1996) Measuring and testing genetic differentiation with ordered vs. unordered alleles. *Genetics*, **144**, 1237–1245.

Prinz K, Weising K, Hensen I (2009) Genetic structure of coastal and inland populations of the annual halophyte *Suaeda maritima* (L.) *dumort*. in Central Europe, inferred from amplified fragment length polymorphism markers. *Plant Biology*, **11**, 812–820.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puechmaille SJ (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, **16**, 608–627.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Rivera-Ortíz FA, Aguilar R, Arizmendi MDC, Quesada M, Oyama K (2015) Habitat fragmentation and genetic variability of tetrapod populations. *Animal Conservation*, **18**, 249–258.

Rogell B, Gyllenstrand N, Höglund J (2005) Six polymorphic microsatellite loci in the Natterjack toad, *Bufo calamita*. *Molecular Ecology Notes*, **5**, 639–640.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rousset F, Raymond M (1995) Testing heterozygote excess. *Genetics*, **140**, 1413–1419.

Rowe G, Beebee TJC, Burke T (1997) PCR primers for polymorphic microsatellite loci in the anuran amphibian *Bufo calamita*. *Molecular Ecology*, **6**, 401–402.

Rowe G, Beebee TJC, Burke T (2000) A further four polymorphic microsatellite loci in the natterjack toad *Bufo calamita*. *Conservation Genetics*, **1**, 371–372.

Rowe G, Harris DJ, Beebee TJC (2006) Lusitania revisited: a phylogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range. *Molecular Phylogenetics and Evolution*, **39**, 335–346.

Sinsch U, Oromi N, Miaud C, Denton JS, Sanuy D (2012) Connectivity of local amphibian populations: modelling the migratory capacity of radio-tracked natterjack toads. *Animal Conservation*, **15**, 388–396.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, **191**, 171–181.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Sotiropoulos K, Eleftherakos K, Tsaparis D *et al.* (2013) Fine scale spatial genetic structure of two syntopic newts across a network of ponds: implications for conservation. *Conservation Genetics*, **14**, 385–400.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London, Series B*, **277**, 661–671.

Ursenbacher S, Guillon M, Cubizolle H *et al.* (2015) Postglacial recolonization in a cold climate specialist in western Europe: patterns of genetic diversity in the adder (*Vipera berus*) support the central – marginal hypothesis. *Molecular Ecology*, **24**, 3639–3651.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Vos CC, Antonisse-De Jong AG, Goedhart PW, Smulders MJ (2001) Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity*, **86**,

598–608.

Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11466–11473.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.

Wang J (2016) A comparison of single-sample estimators of effective population sizes from genetic marker data. *Molecular Ecology*, **25**, 4692–4711.

Wang IJ, Johnson JR, Johnson BB, Shaffer HB (2011) Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, *Ambystoma californiense*. *Conservation Genetics*, **12**, 911–920.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167–184.

Waples RS, Antao T, Luikart G (2014) Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, **197**, 769–780.

Waples RS, England PR (2011) Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, **189**, 633–644.

Weir BS, Cockerham CC (1984) Estimating *F*-Statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock M, McCauley D (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Wilkinson JW, Griffiths RA (2013) Amphibian declines and conservation in Britain. In: *Amphibian biology. Status of conservation and decline of amphibians: Eastern hemisphere. Western Europe* (eds Heatwole H, Wilkinson JW), p. 108. Pelagic Publishing, Exeter, UK.

Table 1: Estimates of nuclear and mitochondrial genetic variation in 68 populations of *Bufo [Epidalea] calamita.* Sample size (N), total number of alleles (A_T), the mean number of alleles (A_n), the observed heterozygosity (H_O), expected heterozygosity (H_E), the nuclear allelic richness (A_r) and the mean intra-population fixation index (F_{IS}) according to Weir & Cockerham (1984). Significance of F_{IS} was tested using a *U*-test with 10,000 iterations for each population over all loci and corrected using a Bonferroni adjustment technique (Rice 1989). Also reported are the number of genotyped individuals for mitochondrial SNP polymorphism (N_mito), the number of different mitochondrial haplotypes (M_T) and the haplotype richness (M_r), and the one-sample effective population size (N_e) estimates using linkage disequilibrium method with jackknife 95% CIs. ABC Analyses: specific populations used in ABC analyses. –: Not applicable (insufficient sampling size).

* *P* < 0.05; ** *P* < 0.01; *** *P* < 0.001

ID	Area	Latitude (WGS84)	Longitude (WGS84)	N	A_T	A_n	H_O	H_E	A_r	F_{IS}	N_mito	M_T	M_r	N_e	95% CIs	ABC analyses
C-01	North Coastline	2.5398	51.0821	9	118	3.371	0.471	0.523	3.19	0.104 *	9	3	2.96	–	–	X
C-02	North Coastline	2.5040	51.0728	17	129	3.686	0.525	0.504	3.12	-0.043	15	2	2	53	30.7 - 149	
C-03	North Coastline	2.5044	51.0686	8	107	3.057	0.468	0.467	2.97	-0.001	8	2	2	–	–	
C-04	North Coastline	2.4501	50.9760	2	63	–	–	–	3.29	–	2	2	–	–	–	
C-05	North Coastline	2.2778	51.0199	29	133	3.800	0.473	0.521	2.90	0.094 ***	10	1	1	46.9	33.3 - 73.8	
C-06	North Coastline	2.1699	50.9883	13	134	3.829	0.51	0.504	3.3	-0.013	13	2	2	–	–	X

1

ID	Location														
C-07	North Coastline	2.1744	50.9877	10	120	3.429	0.534	0.501	3.16	-0.071 *	10	2	2	-	−	
C-08	North Coastline	1.9540	50.9830	9	122	3.486	0.509	0.507	3.07	-0.005	9	2	2	−	−	
C-09	North Coastline	1.7849	50.9519	10	117	3.343	0.489	0.497	3.14	0.017	10	2	2	−	−	X
C-10	North Coastline	1.6803	50.8945	4	83	−	−	−	3.24	−	4	1	-	−	−	
C-11	North Coastline	1.6292	50.7968	12	119	3.400	0.486	0.514	3.12	0.046	11	1	1	−	−	X
C-12	North Coastline	1.5788	50.6490	11	111	3.171	0.423	0.443	2.91	0.047	10	3	2.92	−	−	
C-13	North Coastline	1.5876	50.5991	18	129	3.686	0.465	0.457	3.03	-0.02	13	2	2	12.5	9.8 - 16.2	
C-14	North Coastline	1.5987	50.5984	10	125	3.571	0.483	0.486	3.27	0.008	7	1	1	-	−	X
C-15	North Coastline	1.5785	50.5968	8	121	3.457	0.443	0.471	3.33	0.065	7	1	1	-	−	X
C-16	North Coastline	1.6237	50.5829	21	109	3.114	0.45	0.412	2.64	-0.093 **	14	1	1	8.6	6.5 - 11.5	
C-17	North Coastline	1.6244	50.5823	16	102	2.914	0.409	0.387	2.50	-0.057	12	1	1	28	15.9 - 73.4	
C-18	North Coastline	1.5832	50.4869	20	145	4.143	0.499	0.485	3.30	-0.029	12	3	2.67	198.7	73.8 - ∞	X

C-19	North Coastline	1.5758	50.456	13	138	3.943	0.479	0.486	3.35	0.015	12	3	2.84	–	–	X
C-20	North Coastline	1.5889	50.4398	11	124	3.543	0.463	0.455	3.32	-0.021	11	3	2.88	–	–	
C-21	North Coastline	1.5758	50.4350	22	157	4.486	0.516	0.498	3.15	-0.037	14	3	2.94	316.1	81.3 - ∞	X
MA-01	Northern Coalfield	2.3279	50.5772	22	104	2.971	0.439	0.455	2.71	0.035	15	2	1.93	13.1	10 - 17.5	
MA-02	Northern Coalfield	2.3291	50.5788	15	102	2.914	0.493	0.474	3.25	-0.039	10	4	3.84	6.9	4.9 - 9.4	
MA-03	Northern Coalfield	2.5198	50.4786	27	127	3.629	0.544	0.528	2.93	-0.031 *	25	4	3.66	20.2	16.2 - 25.8	
MA-04	Northern Coalfield	2.5218	50.4858	2	67	–	–	–	3.23	–	2	2	–	–	–	
MA-05	Northern Coalfield	2.5675	50.4591	26	107	3.057	0.47	0.455	3.11	-0.032	15	2	2	22.3	16.2 - 32.9	
MA-06	Northern Coalfield	2.7380	50.5018	11	130	3.714	0.506	0.516	2.66	0.019	15	2	2	–	–	X
MA-07	Northern Coalfield	2.7400	50.4615	34	140	4.000	0.518	0.534	4.00	0.034	11	4	3.87	36.7	28.9 - 48.4	X
MA-08	Northern Coalfield	2.7469	50.4576	25	140	4.000	0.533	0.533	4.00	-0.002	22	5	4.25	18.2	15.1-22.4	X
MA-09	Northern Coalfield	2.7534	50.2618	10	122	3.486	0.513	0.512	3.14	-0.002	16	3	3	–	–	
MA-10	Northern Coalfield	2.7574	50.4856	10	126	3.600	0.531	0.521	3.6	-0.021	10	5	4.84	–	–	X

0

MA-11	Northern Coalfield	2.7656	50.4461	33	129	3.686	0.486	0.485	3.69	0.000	14	2	1.76	18.3	15.1-22.3	
MA-12	Northern Coalfield	2.7797	50.4461	36	125	3.571	0.507	0.490	3.57	-0.035	14	2	1.76	57.9	40.7-92.9	
MA-13	Northern Coalfield	2.7967	50.4038	30	130	3.714	0.523	0.501	3.71	-0.044	15	4	3.87	50.9	36.5-79.3	
MA-14	Northern Coalfield	2.8022	50.492	52	134	3.829	0.504	0.500	3.39	-0.008	11	3	2.87	96.1	64.1-172.4	X
MA-15	Northern Coalfield	2.8110	50.4625	32	111	3.171	0.464	0.444	3.11	-0.048	15	5	4.38	49.4	33.5-83.7	
MA-16	Northern Coalfield	2.9823	50.4583	6	95	–	–	–	2.94	–	6	2	–	–	–	
MA-17	Northern Coalfield	3.2041	50.3168	50	135	3.857	0.494	0.484	3.21	-0.026	15	4	3.72	29	24.5-34.7	
MA-18	Northern Coalfield	3.2332	50.3449	18	136	3.886	0.544	0.552	3.06	0.015	15	3	2.93	50.2	33.1-94.3	
MA-19	Northern Coalfield	3.2371	50.3447	13	128	3.657	0.497	0.533	3.38	0.069*	11	4	3.76	–	–	X
MA-20	Northern Coalfield	3.2381	50.3453	21	142	4.057	0.554	0.553	3.35	-0.002	14	3	2.75	16.6	13.2-21.4	
MA-21	Northern Coalfield	3.2409	50.3827	17	127	3.629	0.506	0.508	3.15	-0.031	15	3	2.72	75.8	39.1-474.4	X
MA-22	Northern Coalfield	3.2457	50.3751	38	142	4.057	0.545	0.529	3.27	0.003	13	5	4.53	51.2	38.6-72.3	

MA-23	Northern Coalfield	3.2899	50.3425	40	118	3.371	0.561	0.504	2.90	-0.116 ***	10	2	2	39.3	29.3-56	
MA-24	Northern Coalfield	3.3054	50.3581	7	109	3.114	0.567	0.513	3.31	-0.117 **	7	2	2	–	–	X
MA-25	Northern Coalfield	3.3059	50.3575	8	119	3.400	0.529	0.532	3.34	0.008	7	2	2	–	–	X
MA-26	Northern Coalfield	3.3060	50.3584	13	130	3.714	0.549	0.554	3.34	0.008	13	4	3.79	–	–	X
MA-27	Northern Coalfield	3.3073	50.3585	7	117	3.343	0.527	0.539	3.24	0.026	7	3	3	–	–	
MA-28	Northern Coalfield	3.5190	50.4135	9	105	3.000	0.517	0.501	2.85	-0.034	3	2	–	–	–	
MA-29	Northern Coalfield	3.5193	50.413	16	111	3.171	0.498	0.497	2.86	-0.001	10	3	3	54.7	27.8-333.5	
MA-30	Northern Coalfield	3.5208	50.4133	14	108	3.086	0.516	0.500	3.11	-0.034	11	3	2.99	–	–	
MA-31	Northern Coalfield	3.5670	50.4044	10	115	3.286	0.503	0.501	3.01	-0.005	10	3	2.92	–	–	
MA-32	Northern Coalfield	3.5706	50.4076	34	133	3.800	0.519	0.517	2.74	-0.003	14	4	3.71	83.7	52.3-179.3	
NEF-01	Northern France	3.9334	49.5534	13	128	3.657	0.527	0.514	3.36	-0.027	11	4	3.98	-	–	
E-01	Eastern France	5.6668	47.5337	16	144	4.114	0.501	0.557	3.00	0.093 ***	14	2	2	68.9	34-764.6	X
E-02	Eastern France	5.0678	47.4317	23	146	4.171	0.471	0.572	3.4	0.153 ***	13	2	2	22.9	17.7-31.1	X

									6							
E-03	Eastern France	5.4833	46.7500	20	94	2.686	0.345	0.373	3.57	0.078 **	12	1	1	62.7	26.2- ∞	X
E-04	Switzerland	6.3632	46.5355	17	109	3.114	0.436	0.490	1.92	0.111 ***	7	1	1	42.3	23.5-134	X
E-05	Switzerland	6.4196	46.5025	6	94	–	–	–	3.40	–	4	2	–	–	–	X
E-06	Lorraine, France	6.1050	48.5509	10	108	3.086	0.501	0.490	3.45	-0.027	9	2	2	–	–	
SEF-01	Alpes, France	5.8811	44.2881	21	152	4.343	0.496	0.576	3.17	0.096 ***	10	1	1	1506.1	88.1- ∞	X
SWF-01	Lot, France	1.7075	44.8536	20	137	3.914	0.527	0.562	2.32	0.059 **	11	2	2	6.9	5.5-8.6	
WF-01	Western, France	-3.0028	48.8768	20	97	2.771	0.348	0.355	2.70	0.004	8	1	1	9	6.5-12.6	X
WF-02	Western, France	-3.0438	48.7871	21	75	2.143	0.31	0.327	2.65	0.031	7	1	1	40.5	16- ∞	X
WF-03	Western, France	-1.1667	48.2167	25	139	3.971	0.495	0.528	2.92	0.052 *	13	2	2	95.9	46.2-1805.3	X
WF-04	Western, France	-1.9920	48.016	22	112	3.200	0.449	0.444	2.43	-0.012	15	2	2	54.6	30.4-172.9	X
WF-05	Western, France	-1.9500	47.8833	6	93	–	–	–	2.46	–	2	1	–	–	–	X
N-01	Sweden	13.4687	55.7106	10	52	1.486	0.144	0.157	3.32	0.089 *	10	1	1	–	–	

Figure 1: (A) Location of the 68 populations of *Bufo [Epidalea] calamita* sampled in Western Europe. Inset map focuses on fine-scale geographical distribution of 53 populations located in the focal study area in northern France, with circle sizes being proportional to the number of sampled individuals. Shaded area indicates the location of the coalfield area in northern France. **(B)** Results of Bayesian clustering assignment probability of individuals (barplot on the left) and of populations (maps on the right) giving their membership in K= 5 genetic clusters. Each individual is represented by a thin line partitioned into K colored segments displaying the individual estimated membership fractions in K clusters. Each population is represented by a diagram partitioned into K colors displaying the mean population estimated membership fractions in K clusters.

Figure 2: Graphical representation of competing scenarios of the settlement of *Bufo [Epidalea] calamita* in the focal study area in northern France, tested by approximate Bayesian computation. N_a, N_1 to N_5, N_{f1} to N_{f4} refer to effective population sizes of putative ancestral, standing populations and past transitory populations, respectively, and, t_1 to t_4 refer to divergence times (prior settings of model parameters are given in Table S2). The first set of scenarios **(A)** aims to determine the origins of each northern lineage by evaluating the relationships of each northern target unit with respect to lineages located outside northern France (eastern and western France lineages). The second set of scenarios **(B)** aims to determine the northern France lineage relationships taking into account all the lineages observed. All hypotheses assumed a shared ancestor and ancient divergence of western and eastern France lineages from the last climate cooling (see Results on genetic affiliations).

Figure 3: (A) Median joining network depicting the relationships among 21 haplotypes (H1 to H21) observed in 68 sampled populations of *Bufo [Epidalea] calamita*. Mitochondrial diversity is based on a combination of 15 mitochondrial DNA SNPs coded as a single haplotype. Each black trait represents one nucleotide difference. **(B)** Maps showing the geographical distribution of mitochondrial diversity for 68 sampled populations of *B. calamita.* Inset map refers to the fine-scaled geographical distribution of populations sampled in the focal study area in northern France. Colors refer to the haplotype identities given in the haplotype network.
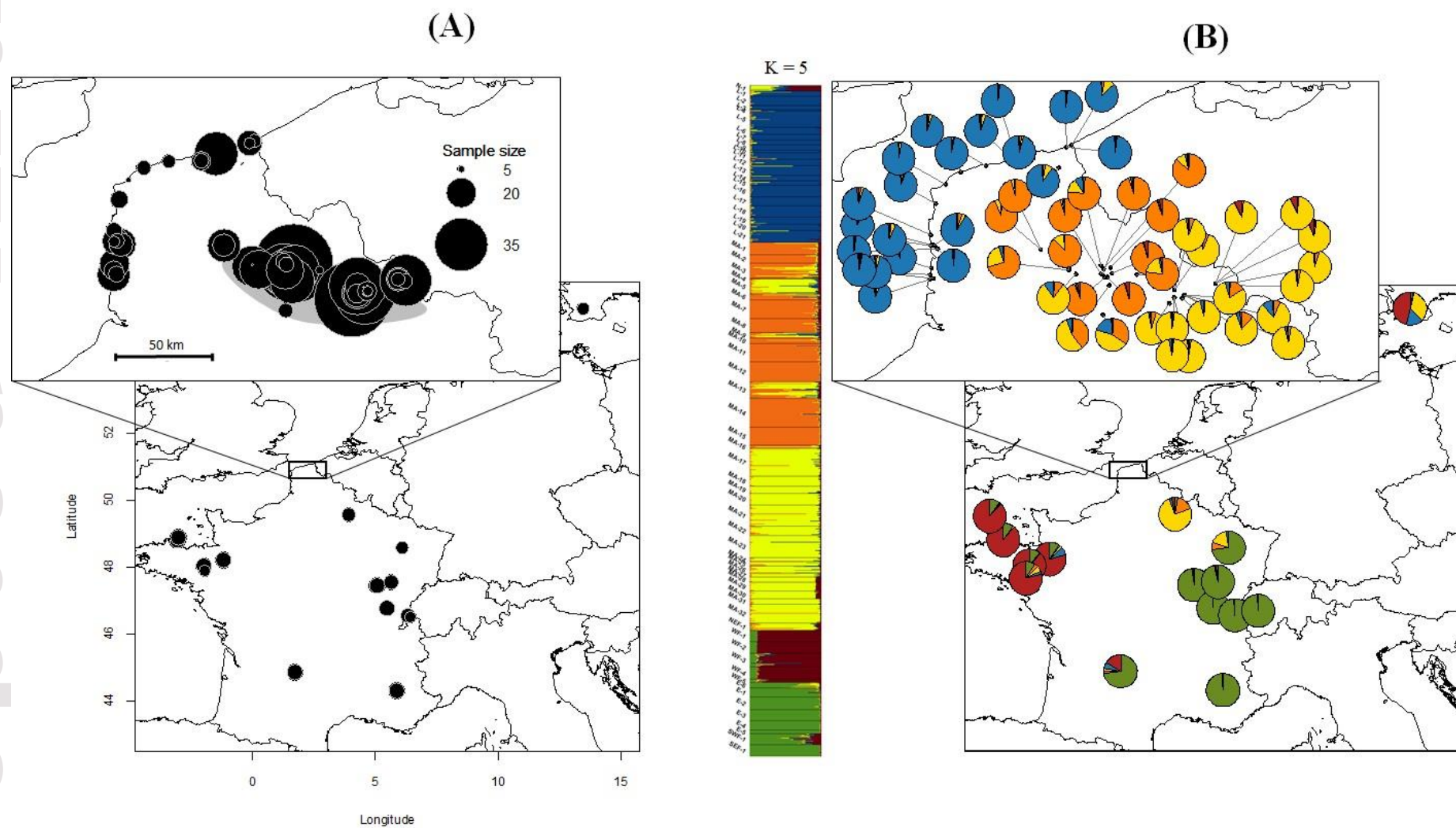
Figure 4: Variation in **(A)** mean multilocus intra-population fixation index (F_{IS}), **(B)** nuclear allelic richness (A_r), **(C)** effective population size estimates on nuclear data, **(D)** gene diversity (H_e) and **(E)** observed heterozygosity (H_o) among the three genetic clusters depicted within the focal study area in northern France: coastal population cluster, western coalfield population cluster and eastern coalfield population cluster. Also indicated are **(F)** mitochondrial haplotypic richness, **(G)** internal relatedness *IR*, **(H)** mean pairwise F_{ST} estimates and **(I)** mean mitochondrial pairwise G_{ST}. Significant differences tested by a Tukey-Kramer HSD test or Kruskal Wallis multi-comparison test are indicated by letters [a], [b] and [c].

Figure 5: Spatial genetic structure in *Bufo [Epidalea] calamita* populations located within the focal study area in northern France and defined as belonging to three distinct genetic clusters: coastline populations, western coalfield populations and eastern coalfield populations. Average pairwise kinship coefficients (F_{ij}, Loiselle *et al.* 1995) between individuals are plotted against geographical distance for nuclear microsatellites (**A**, **B**, **C**) and mitochondrial haplotypes (**D**, **E**, **F**) for the three clusters of populations. Dashed lines indicate the upper and lower 95% CI limits of no significant spatial genetic structure (● $P < 0.05$, ● $P > 0.05$). Also indicated for each cluster are the *Sp* statistics and their statistical significance, allowing comparisons of the strength of spatial genetic structure.

Scatterplots of chord genetic distance (D_{CE}; Cavalli-Sforza & Edwards 1967) against pairs of Euclidean geographical distance for the three genetic clusters (**G**, **H**, **I**). Polynomial regression lines are plotted for spatial scales indicated in color, orange: < 5 km, green:] 5 km, 20 km], blue:] 20 km, 50 km [and purple: >50 km. The number of pairwise comparisons is comprised between 18 and 78. Mantel coefficients r_z computed for each spatial scale and regression adjusted R-squared (R^2) are also indicated. $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Figure 1

Figure 2

(A) Set A : Origin of populations in the focal study area (northern France)

➤ A1 Western origin (coast)

N_a

N_f

N_4 N_x N_5

Western France   Eastern France

Coastline
Western coalfield
Eastern coalfield

➤ A2 Eastern origin (inland)

N_a

N_f

N_4 N_x N_5

Western France   Eastern France

Coastline
Western coalfield
Eastern coalfield

Early post-glacial recolonisation — t_4

Colonisation of northern area — t_3

Current

➤ A3 Older divergence (other origin ?)

N_a

N_f

N_4 N_5 N_x

Western France   Eastern France

Coastline
Western coalfield
Eastern coalfield

Early post-glacial recolonisation — t_4

Colonisation of northern area — t_3

Current

(B) Set B: Relationship between focal study populations (northern France)

➤ B1

N_a

N_{f3}

N_{f2}

N_{f1}

N_3 N_2 N_1 N_4 N_5

Eastern coalfield   Western coalfield   Coastline   Western France   Eastern France

➤ B2

N_a

N_{f3}

N_{f2}

N_{f1}

N_1 N_3 N_2 N_4 N_5

Coastline   Eastern coalfield   Western coalfield   Western France   Eastern France

Early post-glacial recolonisation — t_4

t_3

Colonisation of northern area — t_2

Current — t_1

➤ B3

N_a

N_{f3}

N_{f2}

N_{f1}

N_2 N_1 N_3 N_4 N_5

Western coalfield   Coastline   Eastern coalfield   Western France   Eastern France

Early post-glacial recolonisation — t_4

t_3

Colonisation of northern area — t_2

Current — t_1

Figure 3

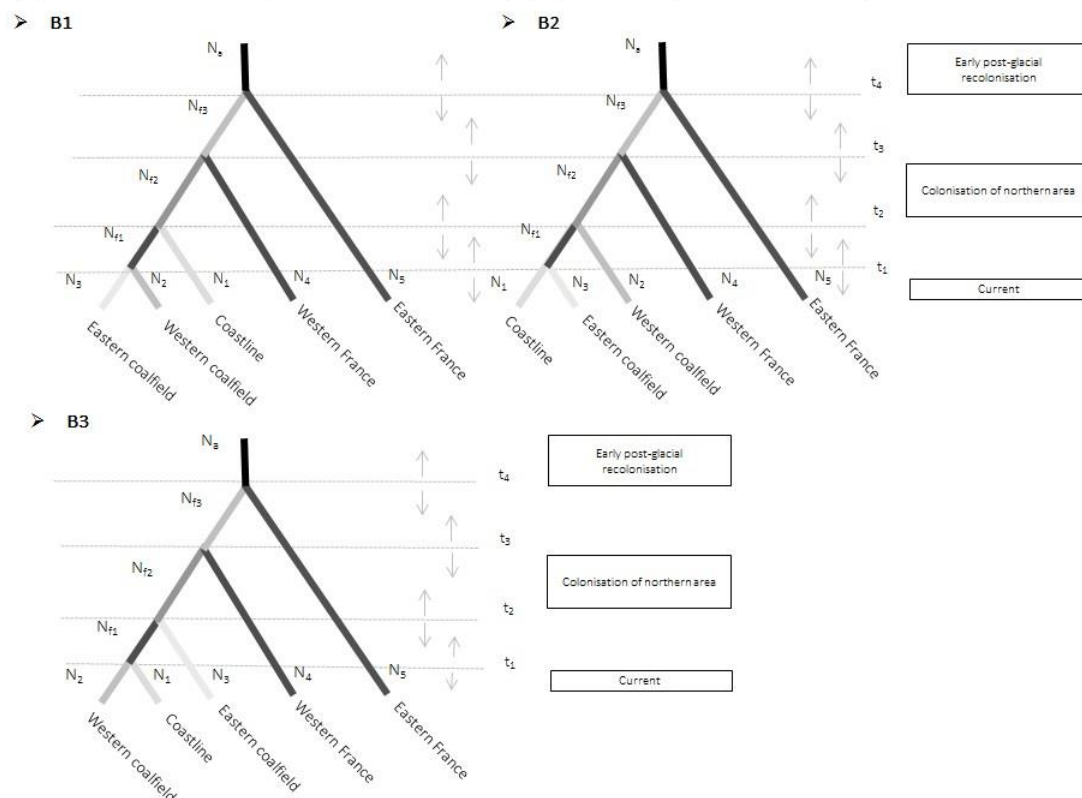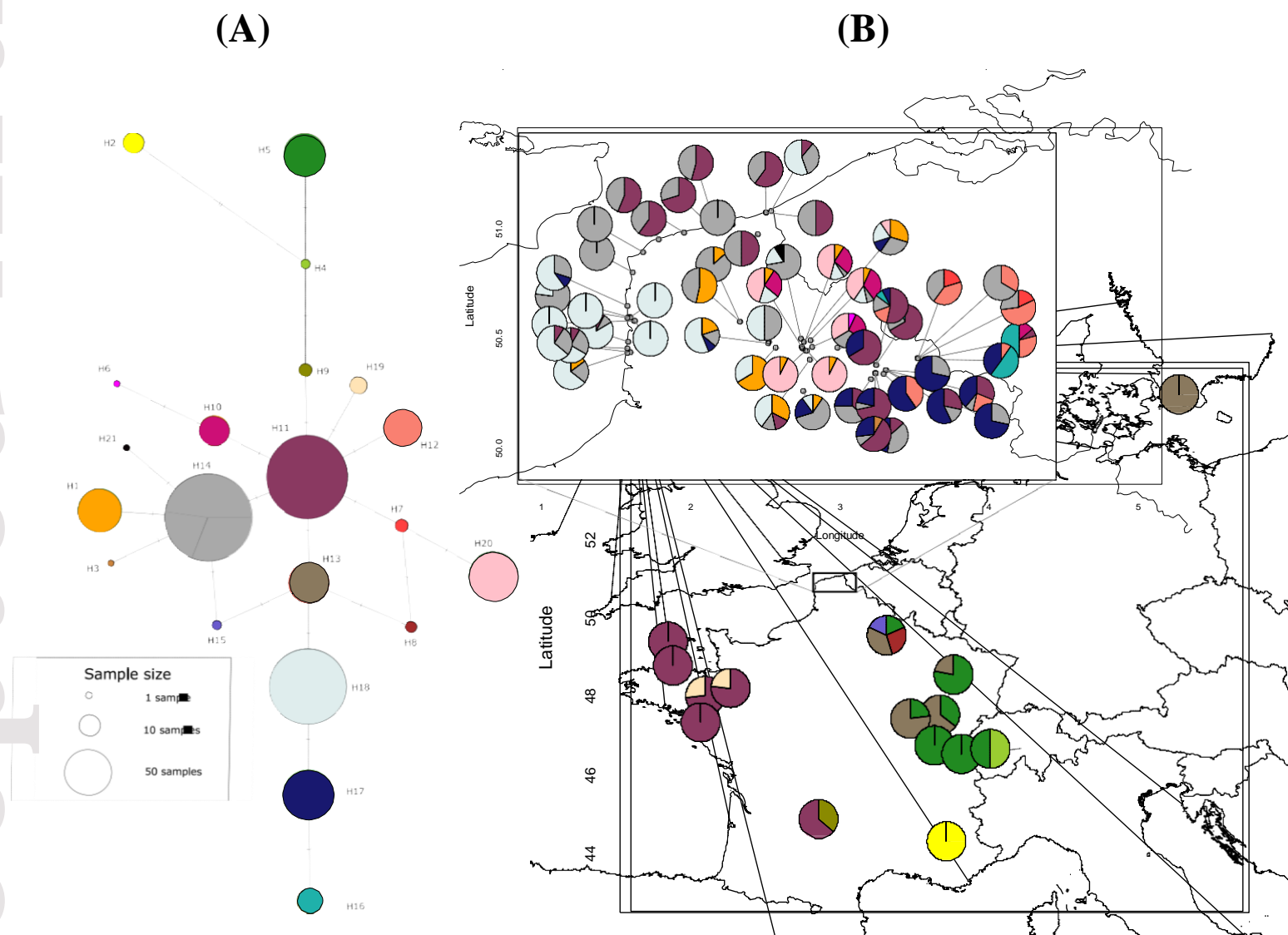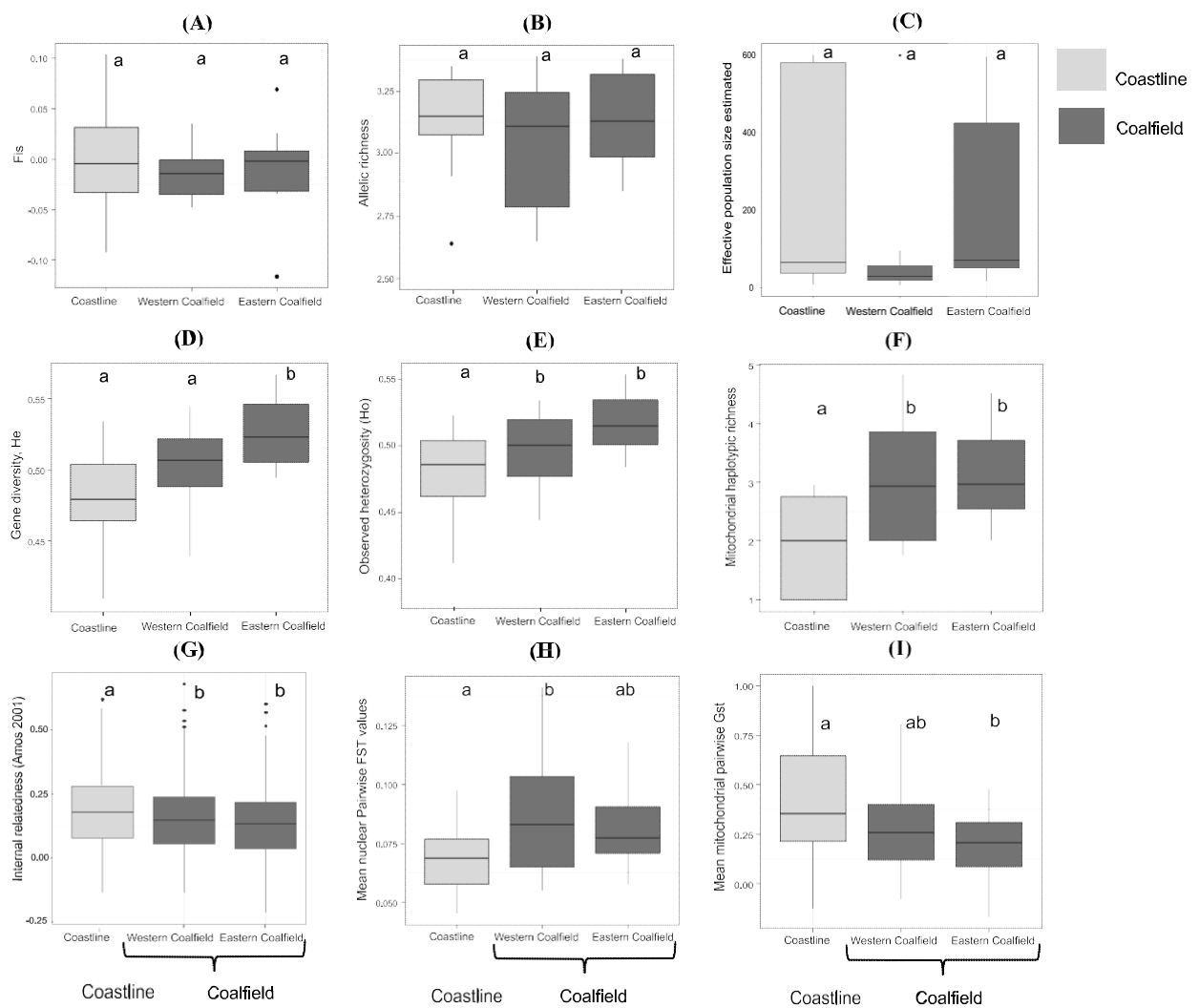**(A)**   **(B)**

Figure 4

Figure 5 :