Towards unravelling the Rosette agent enigma: Spread and emergence of the co-invasive host-pathogen complex, Pseudorasbora parva-Sphaerothecum destruens

Combe Marine, Cherif Emira, Charrier Amélie, Barbey Bruno, Chague Martine, Georges Carrel, Chasserieau Céline, Foissy Jean-Michel, Gerard Barbara, Gozlan Zachary, et al.

Towards unravelling the Rosette agent enigma: Spread and emergence of the co-invasive host-pathogen complex, *Pseudorasbora parva-Sphaerothecum destruens*

Combe Marine[1*+], Cherif Emira[1*+], Charrier Amélie[2], Barbey Bruno[3], Chague Martine[2], Carrel Georges[4], Chasserieau Céline[5], Foissy Jean-Michel[6], Gerard Barbara[7], Gozlan Zachary[8], Guillouët Jérôme[9], Hérodet Benjamin[10], Laine Manon[11], Masseboeuf Fabrice[12], Mirkovic Ivan[13], Nicolas Delphine[14], Poulet Nicolas[15], Martin Jean-François[16], Gilles André[17] & Gozlan Rodolphe Elie[1+]

[1]ISEM UMR226, Université de Montpellier, CNRS, IRD, EPHE, 34090 Montpellier, France

[2]Laboratoires des Pyrénées et des Landes, 1 rue Marcel David, BP 219, 40004 Mont-de-Marsan, France

[3] Fédération de l'Indre pour la Pêche et la Protection du Milieu Aquatique, FDAAPMA 36, 19 rue des Etats-Unis 36000 Châteauroux, France

[4]INRAE, Centre PACA, UMR RECOVER, F-13182, Aix-en-Provence, France

[5]Fédération de Haute-Savoie pour la Pêche et la Protection du Milieu Aquatique, 2092 route des Diacquenods, Saint Martin Bellevue, 74370 Fillière, France

[6]DIR OFB Paca Corse Domaine du Petit Arbois Pavillon Laënnec – Hall B, Avenue Louis Philibert,13547 Aix-en-Provence, France

[7]Fédération de Loire-Atlantique pour la Pêche et la Protection du Milieu Aquatique, France

[8]Lycée Joffre, 31 rue André Chénier, 34130 Mauguio, France

[9]Fédération Nationale de la Pêche en France et de la Protection du Milieu Aquatique 108-110 rue Saint-Maur, 75011 Paris, France

[10]Fédération de l'Ain pour la Pêche et la Protection du Milieu Aquatique, 638 rue du Revermont, ZAC de la Cambuse, 01440 Viriat, France

[11]Fédération de la Pêche et de Protection du Milieu Aquatique de la Gironde, 10 ZA du Lapin, 33750 Beychac-et-Caillau, France

[12]Fédération des Pyrénées-Atlantiques pour la Pêche et la Protection du Milieu Aquatique, 12 boulevard Hauterive, 64000 Pau, France

[13]Fédération de Seine-Maritime pour la Pêche et la Protection du Milieu Aquatique, 11 cours Clemenceau, 76100 Rouen, France

[14]Tour du Valat, Research Institute for the Conservation of Mediterranean Wetlands, Le Sambuc, 13200 Arles, France

[15]Office Français de la Biodiversité, Allée du Pr Camille Soula, 31400 Toulouse, France

[16]CBGP, Montpellier SupAgro, INRA, CIRAD, IRD, Univ Montpellier, Montpellier, France

[17]UMR 1467 RECOVER, Aix Marseille Univ, INRAE, Centre St Charles, 3 place Victor Hugo, 13331, Marseille, France

[*]equally contributing authors

[+] corresponding authors: marine.combe@ird.fr; emira.cherif@ird.fr; rudy.gozlan@ird.fr

Abstract

The emergence of non-native fungal pathogens is a growing threat to global health, biodiversity, conservation biology, food security and the global economy. Moreover, a thorough understanding of the spread and emergence of pathogens among invasive and native host populations, as well as genetic analysis of the structure of co-invasive host populations, is crucial in terms of conservation biology and management strategies. Here we combined extensive catchment sampling, molecular detection tools and genomic signatures to i) assess the prevalence of the rosette agent *Sphaerothecum destruens* in invasive and native fish populations in contrasting french regions, and ii) characterize the genetic diversity and population structure of its co-invasive and asymptomatic carrier *Pseudorasbora parva*. Although *S. destruens* was not detected in all the fish collected its presence in contrasting freshwater ecosystems suggests that the disease may already be widespread in France. Furthermore, our results show that the detection of *S. destruens* DNA in its asymptomatic carrier *P. parva* is still limited. Finally, we found that *P. parva* populations show a homogeneous genetic and geographical structuring, which raises the possibility of the occurrence of successive introduction events in France from their native and invasive range.

Keywords: Biological invasions, fungal pathogens, aquatic disease emergence, population genomics

Introduction

Species dispersal and habitat colonization are key ecological processes for maintaining species diversity and population persistence (Hubbell, 2001; Lande et al., 2003). However, the dramatic increase in species translocations and biological invasions over the last century due to human activities (Lawler et al., 2006), such as the increase in global trade and transportation networks, has led to a global redistribution of biodiversity (including pathogens) (Ricciardi, 2007) as well as an irrevocable loss of ecosystem functions (Chapin et al., 2000; Al-Shorbaji et al., 2016), the decline in wildlife biodiversity (Gozlan et al., 2006) and the emergence of diseases in livestock and wildlife populations worldwide (Daszak et al., 2000; Woolhouse et al., 2005; Gozlan et al., 2005; Jones et al., 2008; Lymbery et al., 2014).

Emerging fungal pathogens pose a growing threat to global health, ecosystems, food security and the world economy (Fisher et al., 2012). The introduction and spread of non-native fungal pathogens is a serious threat to biodiversity and a major concern for conservation biology. Indeed, the co-introduction of a host and a pathogen can dramatically affect the invaded ecosystem through biodiversity losses, destabilization of foodwebs and the spread of emerging infectious diseases to native host species (Stiers et al., 2011; Gallardo et al., 2016; Crowl et al., 2008; Hatcher et al., 2012). These co-introductions of non-native host-pathogen systems can also create new host-pathogen interactions. For example, if native species are susceptible to the pathogen, the pathogen may become a powerful asset allowing the invading host to win competition for niche space (Price et al., 1986). This is illustrated by the co-introduction in the UK of squirrel pox virus and the invasive grey squirrel (*Sciurus carolinensis*), its asymptomatic carrier, which resulted in high mortality of the native red squirrel (*S. vulgaris*) and accelerated its replacement by the invader (Tompkins et al., 2003). Similarly, the extinction of the native white-clawed crayfish (*Austropotamobius pallipes*) caused by crayfish plague is the result of the co-introduction of the fungus *Aphanomyces astaci* and its healthy carrier crayfish (*Pacifastacus leniusculus*)

105 into Europe (Holdich & Pöckl, 2007). In addition, to understand the processes leading to
106 successful invasion, a thorough understanding of the spread of pathogens and their
107 virulence among invasive and native host populations, as well as genetic analysis of the
108 structure of co-invasive host population, is crucial in terms of conservation biology and
109 management strategies.
110 Among the well-known invasions of host and pathogen, the rise of the rosette agent in
111 Europe remains an epidemiological enigma (Combe & Gozlan, 2018). Indeed, there is still a
112 discrepancy between the magnitude of observed and predicted fish mortalities at specific
113 sites were *S. destruens* was introduced and the low or absent level of monitoring and
114 reporting of *S. destruens* across Europe (Combe & Gozlan, 2018). The topmouth gudgeon
115 *Pseudorasbora parva*, a small freshwater cyprinid fish native to East Asia (East China,
116 Taiwan, Korea and Japan), has been the fastest fish invasion worldwide (Gozlan, 2012;
117 Zhang & Zhao, 2016). In particular, *P. parva* is also the asymptomatic carrier of the rosette
118 agent (*Sphaerothecum destruens*), a generalist parasite, so far phylogenetically located at
119 the fungal-animal boundary, which has proven to be highly virulent to many native European
120 fish species and responsible for major ecological and economic impacts worldwide (Gozlan
121 et al., 2005; Sana et al., 2017; Combe & Gozlan, 2018). Their co-introduction into European
122 countries is the result of accidental translocations from their native range via aquaculture
123 trade of Chinese carps between China and former USSR countries (Gozlan et al., 2010).
124 Then, during the 1960s multiple introductions of *P. parva* took place around the Black Sea,
125 followed by further introductions in the 1980's in Eurasia and North Africa (reviewed in
126 Combe & Gozlan, 2018). Following these initial human-mediated introductions, natural local
127 spread took place across the major European rivers to the Middle-East. While previous
128 studies suggested that European populations of *P. parva* were introduced from two of four
129 distinct native lineages (Sana et al., 2017; Hardouin et al., 2018; Combe & Gozlan, 2018),
130 we recently found that the current genetic clustering of native *P. parva* range was shaped by
131 waves of gene flow from population in southern and northern China and that the invasive
132 genetic diversity was the result of multiple invasion pathways of already admixed native

populations (Brazier et al., 2021). In France, *P. parva* has been observed for the first time in northwestern rivers in 1978-1979 (Allardi & Chancerel, 1988; Poulet et al., 2011), and *S. destruens* has been detected in asymptomatic populations of topmouth gudgeon in Southwestern France (Charrier et al., 2016) and has been suspected of being responsible for two episodes of mortality of brown and rainbow trout in farms and experimental facilities in western France (Boitard et al., 2017). As *P. parva* is now reported in almost all french regions, including the island of Corsica, the potential spread and virulence of the rosette agent in native and invasive fish populations represent a major ecological, health and economic issue for aquatic biodiversity and fish conservation. For example, its co-introduction into a catchment area in southeast Turkey almost led to the total extinction (up to 80%-90% of mortalities) of native fish populations of conservation or economic importance such as sea bass (Ercan et al., 2015; Combe & Gozlan, 2018). However, to date, the extent of distribution of *S. destruens* in any country and the genetic structuring of co-invasive host-pathogen populations have not yet been identified in relation to their native range and to other invasive European countries. This information would be of primary importance to assess whether there is a risk of disease emergence in wild and farmed fish populations at a country level and to better understand the historical connections between the different host-pathogen.

Here we combined catchment-scale sampling, molecular detection tools and genomic signatures to i) assess the presence and prevalence of *S. destruens* in invasive and native fish populations at a country level, and ii) characterize the genetic diversity and population structure of co-invasive *P. parva* populations to ascertain whether they constitute a heterogeneous genetic pool derived from admixed native populations that could have major implications for the genetic diversity and virulence of the pathogen.

Materials & Methods

Sampling

Fish sampling was conducted in France between March and November in 2017, 2018 and 2019 in collaboration with seven French Department Angling Associations, such as Fédération Départementale de Pêche et de Protection du Milieu Aquatique FDPPMA de l'Ain (Ain), FDPPMA de l'Indre (Indr), FDPPMA de Haute-Savoie (HSAv), FDPPMA de Seine-Maritime (Smar), FDPPMA de Gironde (Gir), FDPPMA des Pyrénées-Atlantiques (PyrA), FDPPMA de Loire-Atlantique (LoirA); the Provence Alpes Côte d'Azur regional direction of the French Biodiversity Agency (OFB) (Cors); and two research institutes:, the Tour du Valat (Brhon) and INRAE (Vauc). For each angling association, specific considerations were followed such as conservation, target species and sensible sites. Fish were collected using nets, traps or even electrofishing. Sampled fish were immediately euthanized with an anesthetic (120 mg/L benzocaïne) and dissected either at the sampling site or at the Laboratory des Pyrénées et des Landes (LPL). Ten freshwater sites, five in northern France and five in southern France, were selected. At each sampling site, 50 individuals of *P. parva* and other native fish species corresponding to a total of 100 individuals were collected and analyzed at the LPL, although for some sites we could not reach this exact number of individuals. For all fish, liver, kidney and spleen were collected and preserved in 70% ethanol to test for the presence of the parasite *S. destruens* and the prevalence of the disease. For each *P. parva*, fin clips were collected, preserved in 70% ethanol and used to assess the genetics of the host population.

DNA extractions

Total DNA was extracted from internal organs (kidney, liver and spleen) previously pooled from all individuals belonging to the same species and sampling site. Two commercial kits, selected on the basis of their column capacity, were used according to the weight of the organs collected and therefore the size of the fish. The NucleoSpin® Tissue kit (Macherey-Nagel) was used for weights up to 25 mg while the NucleoSpin® Soil kit (Macherey-Nagel) was used for higher weights up to 250 mg. Prior to extraction, samples were ground in lysis buffer using a tissue homogenizer (Precellys® 24, Bertin). Extractions were performed

188 according to the manufacturer's instructions for both kits. A negative control was added (kit
189 reagents without sample) for each set of extractions to ensure that the kit reagents were not
190 contaminated.
191
192 **Real-time quantitative PCR**
193 *Primers and probe design.* The primers and probes for the real-time qPCR for the detection
194 of *S. destruens* were designed from the 18S rRNA gene sequences obtained from GenBank.
195 The accession numbers for the *S. destruens* sequences were: AY267344, AY267345,
196 AY267346 and FN996945. Sequences of closely related organisms such as *Dermocystidium*
197 *salmonis* (U21337), *Dermocystidium spp.* (U21336), *Amphibiocystidium ranae* (AY550245),
198 *Rhinosporidium seeberi* (AF118851) and *Ichthyosporea* (KM213082) were aligned with the
199 18S sequences of *S. destruens* in order to find regions specific to *S. destruens*, using Clustal
200 Omega (Sievers et al., 2011). Primers and probe were then designed using Primer3
201 software (Koressaar & Remm, 2007; Untergasser et al., 2012) and synthesized by Eurofins
202 Genomics, with the following sequences: forward primer (SdACH F) 5'-
203 ACCGCCCGTCGCTACTAC-3', reverse primer (SdACH R) 5'-AACTTTTCGGCAGCCTCAC-
204 3' and probe (SdACH P) 5'-(FAM)-TGGCCCTGTACCG-(MGBEQ)-3'.
205 *TaqMan qPCR reaction.* Amplification reactions were performed in a QuantStudio™ 5 Real-
206 Time PCR System (ThermoFisher Scientific). The qPCR reaction mixes (25 µL) consisted of
207 12.5 µL of GoTaq® Probe qPCR Master Mix (Promega), 5 µL of DNA template, and primers
208 and probe at the same final concentration of 200 nM each. Water was added up to 25 µL.
209 Cycling conditions were as follows: 95°C for 2 min, then 40 cycles of 95°C for 15 sec, 60°C
210 for 1 min. A negative control (5 µL of Nuclease-free water instead of DNA template) was
211 added at each qPCR run.
212 *Specificity of the qPCR assay.* The specificity of this new qPCR design was determined
213 using DNA extracts of various fish pathogens, obtained from collections or isolated in the
214 LPL laboratory from naturally infected fish and identified by MALDI-TOF mass spectrometry
215 (Bruker Daltonics) (see Supplementary Table S1). DNA was extracted from pure cultures

using the NucleoSpin® Tissue kit (Macherey-Nagel) following the manufacturer's recommendations. For Ichthyosporea species related to *S. destruens* (*Dermocystidium salmonis*, *Dermocystidium spp.*, *Rhinosporidium seeberi* and *Amphibiocystidium ranae*), synthetic genes (Eurofins Genomics) containing the targeted sequences of 18S rRNA were synthesized and freeze-dried plasmids were resuspended in 1 mL of 10 mM Tris-HCL pH 8 and used in qPCR assays. The DNA concentrations used for this specificity test were 100 ng/µL for all extracts and 10^6 copies for the plasmids.

qPCR assay performance. Our protocols are based on the french PCR standard for animal health NFU47-600. A synthetic gene (Eurofins Genomics) containing the *S. destruens* 18S rRNA sequence targeted by our TaqMan assay was synthesized to determine qPCR assay performance criteria such as efficiency, slope, intercept and linear regression coefficient (r^2). Standard curves were generated from 10-fold serial dilutions in 10 mM Tris-HCl, pH 8 of the plasmid DNA from 10^6 to 1 copy used as PCR template, allowing the determination of the limit of quantification (LOQ) and the identification of the gene copy number. The LOQ is the cut-off point below which three replicates of CT-values tend to disperse one from each other or are indeterminate. The standard curves were performed on three separate days (see Supplementary Fig. S1). The threshold values (CT) were plotted against the corresponding gene copy numbers.

Limit of Detection (*LOD).* The french PCR standard for animal health NFU47-600 states that the LOD should be determined by 6 serial dilutions from the identified LOQ (here LOQ = 100 copies) with 8 replicates of each dilution, repeated three times independently. Considering the 95% confidence interval, the LOD should correspond to the number of plasmid copies giving 23/24 positive results. For this purpose, the synthetic *S. destruens* 18S rRNA gene was diluted by half from the solution corresponding to 100 copies (LOQ) to 3.125 copies to identify the limit of detection (LOD). In total of six dilutions (100, 50, 25, 12.5, 6.25 and 3.125) were tested in a first experiment, which consisted of serial dilutions in 10 mM Tris-HCl, pH 8. In a second experiment three serial dilutions (50, 25 and 12.5) were performed in an uninfected fish DNA solution at 100 ng/µL to further determine the LOD (see

Supplementary Table S2). The fish DNA solution was obtained from the EPC (epithelioma papulosum cyprini ATCC® CRL-2872™) cell line after DNA extraction using the NucleoSpin® Soil kit (Macherey-Nagel).

Disease Prevalence

Each individual sampled was tested by qPCR assays (kidney, liver, spleen being pooled together), thus indicating how many individuals were found infected out of the total number of individuals sampled. The prevalence of the disease was calculated as follows: (number of positive individuals / total number of individuals tested) × 100.

GBS sequencing

A total of 420 DNA samples (410 individuals in total, with 10 individuals replicated on the sequencing lanes) were genotyped for single nucleotide polymorphism (SNP) markers by first digesting the genomic DNA with PstI and then genotyping-by-sequencing (GBS), resulting in an average of 2,225,123 (∓1,108,497) raw sequencing reads per sample (Table 1; see Supplementary Information). Native samples from Brazier et al. (2021) supplemented with individuals from Taiwan were also included in the subsequent analysis to place the french population back into the native range (see Supplementary Table S3). SNP calling was performed using the denovo_map.pl pipeline implemented in Stacks (v2.55) (Catchen et al., 2013). The same combination of parameters used by Brazier et al. (2021) was applied (ustacks; m=10, M=3, maximum of 2 loci per stack and cstacks; n=3). A minimum read depth of 20 was required for each marker. To avoid bias due to large proportions of missing data and maintaining a large sample size, loci with more than 45% missing data were removed from the dataset, resulting in 5028 validated SNP markers with an overall missing data rate of 36%. After trimming, the final dataset contained 409 french individuals from the sampled sites, with an average of 45 individuals per site (Table 1) and 300 individuals from the native range (Table S3).

Genotype phasing and imputation

Haplotype phasing and missing data were inferred for the 708 individuals analyzed using Beagle 5.1 (Browning & Browning, 2007; Browning et al., 2018). Beagle uses the localized haplotype-cluster model and applies an iterative approach to infer the most likely haplotype pair for each individual. At each iteration, phased input data is used to build a localized haplotype-cluster model. Once the model is constructed, the phased haplotypes of each individual are sampled from the induced diploid HMM, conditional on the individual's genotypes. The sampled haplotypes form the input for the next iteration, and so on. In the final iteration, the Viterbi algorithm selects the most likely haplotypes for each individual, conditional on the diploid HMM and the individual's genotype data. For each copy of each individual, missing alleles are randomly imputed according to allele frequencies, and the data for each individual are phased by randomly ordering the genotypes (Browning & Browning, 2007).

Genetic diversity and population structure

A hierarchical approach was adopted to study and place back the french *P. parva* diversity within the native range. First, the genotypes from France were analyzed to test whether the french group is a continuous homogeneous population or a structured population. Secondly, the analyses were performed at the level of the area of origin (France vs. native range).

Expected heterozygosity (He), observed heterozygosity (Ho), diversity (Pi), private alleles (Pa), variant sites (Vs), inbreeding coefficient (Fis) and population differentiation indices Fst and **Φ**st on haplotype data were estimated by kernel-smoothed calculations with 1000 bootstrap resamples using the populations program implemented in Stacks (v2.55) (Catchen et al., 2013). In addition to allele occurrence frequencies, pairwise **Φ**st analysis takes advantage of the evolutionary distance between alleles and provides insight into the patterns of relationships between populations (Holsinger & Weir, 2009).

The genetic distance between individuals based on dissimilarity was estimated by calculating the number of allelic differences between two samples using the bitwise.dist

function in the popprv2.9.1 package (Kamvar et al., 2014). To visualize the population structure, a minimum spanning network (MSN) clustering multilocus genotypes (MLG) by the previously estimated genetic distances was constructed using the popr.msn function (popprv2.9.1 package).

The relationship between individuals was inferred by an Identity-By-State (IBS) analysis through a genome-wide average IBS pairwise distances using the function snpgdsIBS from the SNPRelate package and followed by a hierarchical cluster analysis using the snpgdsHCluster function (SNPRelate Package). A multidimensional scaling analysis (MSA) based on IBS pairwise distances was also performed using the cmdscale function of the R stats package (v3.6.2) (R Core Team, 2020).

Principal component analysis (PCA) and discriminant analysis of principal components (DAPC) were used to estimate clustering first between the french groups, then between the french population and the native groups. The find.clusters algorithm was run with 10 K and ten repetitions for each value of K. The optim.a.score function was used to identify the optimal number of principal components to support the DAPC. The Bayesian information criterion was used to select the most likely number of genetic clusters. All these analyses are implemented in the adegenet package (v 2.1.3) (Jombart & Ahmed, 2011). The R-dependent analyses were run in R version 4.0.5 (R Core Team, 2020).

The fastSTRUCTURE program (Raj et al., 2014) was used to estimate the most likely number of genetic clusters K among the french and native groups. Structure_threader (Pina-Martins et al., 2017) program was used to parallelize and automate fastSTRUCTURE runs. Both simple prior and logistic prior models were used for ancestry estimates, the latter being recommended for identifying subtle structure (Raj et al., 2014). K values ranging from 1 to 10 were analyzed. The best value was evaluated in two ways: using the chooseK algorithm implemented in fastSTRUCTURE and the Puechmaille statistics adapted to uneven sampling (Puechmaille, 2016), implemented in the STRUCTURE SELECTOR web interface (Li & Liu, 2018). Then, CLUMPAK (Kopelman et al., 2015) implemented in STRUCTURE SELECTOR was used with the default options to summarize the results and represent

graphically the sample assignment probability for each K.

Results

Validation of the qPCR assay for *S. destruens* in fish tissues

The first step in validating the specificity of this new qPCR assay was to verify that amplification reactions for *S. destruens* DNA could not induce cross-amplification reactions of other fish pathogens. As this generalist parasite is now well known to infect a wide variety of fish species (Combe & Gozlan, 2018), we targeted bacterial strains infecting salmonids (Harrell et al., 1986; Hedrick et al., 1989; Arkush et al., 1998; Boitard et al., 2017) such as *Aeromonas salmonicida* subsp. *salmonicida*, *Renibacterium salmoninarum*, *Tetracapsuloides bryosalmonae*, *Flavobacterium psychrophilum* and *Yersinia ruckeri,* but also bacteria infecting a wider spectrum of fish species (*Photobacterium damselae* subsp. *damselae* and *Photobacterium damselae* subsp. *piscicida*) as well as closely related species (18S rRNA synthetic gene from *Amphibiocystidium ranae*, *Rhinosporidium seeberi*, *Dermocystidium* spp., *Dermocystidium salmonis* and other *Ichthyosporea* spp.) and showing 95% sequence identity with the 18S rRNA targeted gene from *S. destruens* when subjected to NCBI BLAST analysis. All DNA extracts tested and the negative control resulted in no amplification, while DNA from the North American isolate RA-1 (ATCC® 50643™) of *S. destruens* produced efficient qPCR amplification (C_T = 20.63) (Table S1).

Next, qPCR performance (efficiency, slope, intercept and linear regression correlation) was assessed using 10-fold serial dilutions of a plasmid solution harboring the *S. destruens* 18S rRNA gene sequence ranging from 10^6 copies to 1 copy. Three separate qPCR runs were performed to validate the reproducibility of the results. Each standard curve was obtained by plotting the threshold values (C_T) against the corresponding copy numbers. Here the linear range of this plot was validated over the range of 10^2 to 10^6 copies. The limit of quantification (LOQ), defined by the smallest amount of analyte that can be quantified, corresponds here to 10^2 copies (mean C_T value = 33.6) (Fig. S1). The PCR amplification efficiencies obtained

were 102.96, 99.13 and 98.31% and were calculated using the following formula: $E = (10^{(-1/slope)})-1$, using the slope values for each standard curve. For each assay, the intercepts are also shown (b) with the linear regression correlation (R^2), which are very close to 1 in all cases. All these criteria demonstrate a good performance of qPCR and validate the use of this new tool for the detection of *S. destruens* DNA in fish tissues.

By definition, the limit of detection (LOD) is the smallest amount of analyte in a sample that can be detected with a 95% confidence interval. To determine this LOD, it is necessary to test different concentrations of plasmids with at least 8 technical replicates per dilution and to repeat the experiment on three separate days. Thus, each dilution tested gave 24 results. Given the 95% confidence interval, LOD should correspond to the plasmid copy numbers giving 23/24 positive results. LOD was estimated to be between 25 (24/24 positive results) and 12.5 (20/24 positive results) copies of the target 18S rRNA (Table S2). In order to resolve the exact gene copy number that corresponds to the LOD and to more accurately assess the exact qPCR conditions, another set of assays was carried out by replacing the Tris buffer with uninfected fish DNA extract (Table S2). In this second assay, the 25 copies dilution gave 23/24 positive results and is therefore considered to represent the LOD of *S. destruens* qPCR system. These parameters were taken into account when analyzing the samples by differentiating between *S. destruens* positive samples that are detected and quantifiable (C_T < LOQ (C_T 33.6) or detected but not quantifiable (C_T > LOD (25 DNA copies) but C_T > LOQ (C_T 33.6)).

Detection of *S. destruens* in native and invasive fish samples in France

While fish sampling was spread out between 2017-2019, all angling associations sampled during the same period of the year, between March and November. Whilst May-September represents the optimal period for proliferation, transmission and thus detection of *S. destruens* (Ercan et al., 2015), we acknowledge that sampling in March, April, October and November may have impacted the efficiency of *S. destruens* detection and thus lead to an underestimation of the prevalence of this parasite. Only the angling associations of Ain,

Cors, Gir, Indr, HSAv, Smar sent samples strictly following the protocol detailed above. In some cases, either fewer *P. parva* individuals or fewer native fish species were sampled and therefore analyzed (Table 2). *S. destruens* DNA of *S. destruens* was detected in fish samples collected in Ain, Indr, Cors, Brhon and Gir, indicating thus a wide spread of this parasite across freshwater rivers in France (Fig. 1). We found at least 6 different fish species infected with *S. destruens*; its co-invasive asymptomatic carrier *P. parva* (prevalence between 2-4%) as previously reported (Charrier et al., 2016); 3 native fish species found infected with this parasite for the first time such as *Alburnus alburnus* (bleak, 9% prevalence), *Rhodeus amarus* (European bitterling , 20% prevalence) and *Rutilus rutilus* (roach, 4% prevalence); 2 Genus for which specific species were not determined such as *Gobio sp.* (3 species of gudgeons are reported in France, 10% prevalence) and *Phoxinus sp.* (5 species of minnows are reported in France, 2% prevalence).

Genetic diversity and structure of *P. parva* populations at country level

Analysis of the genetic diversity of the nine groups sampled from the northern (Smar), central (Indr), central-eastern (Ain, HSAv), south-eastern (Brhon, Vauc, Cors) and south-western (PyrA, Gir) regions of France showed low overall diversity within the groups (Table 1). The estimated percentages of polymorphic loci ranged from 8.59% to 13.4%, with the Smar group and Indr group having the lowest and the highest rate, respectively (Table 1). Low values of observed heterozygosity (Ho) and nucleotide diversity (Pi) were also found in the analyzed groups.

The lowest values of Ho and Pi were observed in the Cors group (0.009 and 0.0116, respectively), while the Vauc group had the highest values (Ho: 0.0166 and Pi: 0.0151) (Table 1). To test whether the french population is structured, the genetic differentiation and relationship between the nine french groups were assessed using four complementary analyses. First, the pairwise Fst between groups showed little genetic differentiation between the french groups with values ranging from 0.0064 to 0.0199 and the highest value being observed between the PyrA and Gir groups (Fig. 2A). Secondly, the genetic distance

between individuals and the MSN showed that the samples do not cluster by groups, without geographical differentiation (Fig. 2B). Finally, this lack of genetic and geographical structure was also observed by hierarchical clustering based on the pairwise IBS and by PCA analysis (Fig. 2C and 2D). Our results showed unambiguously and surprisingly that individuals sampled across the country cluster in a unique homogeneous deme rather than in regionally associated demes.

Relationship of french deme *P. parva* and native range

Φst values ranged from 0.009 to 0.146 (Fig. 3A; See Supplementary Table S4).The pairwise **Φ**st values between the french deme and each native deme were relatively high and ranged from 0.040 to 0.146. Japan and Taiwan showed the greatest differentiation with the french deme with **Φ**st equal to 0.146 and 0.097, respectively. On the other hand, the lowest differentiation values (French vs. native) were observed with the North China (0.040) and Northeast China (0.047) demes. The differentiation between the french deme and South and Southeast China demes were equivalent (0.058) (Fig. 3A; See Supplementary Table S4). The IBS pairwise distances used to perform a multidimensional scaling analysis showed a clear compact cluster with predominantly french individuals also containing several individuals from each native deme except Japanese and South China demes (Fig. 3B).

We sought to determine the genetic structure and relationship between the french deme and the native range. To select the most plausible number of genetic clusters K, several criteria were evaluated. The values of the Bayesian Information Criterion for the K-means clustering suggested a K=8-9 (see Supplementary Information). The chooseK algorithm based on marginal likelihood and Puechmaille's statistics suggested a K=9 (see Supplementary Fig. S2, S3) for fastSTRUCTURE analysis. For all values of K, the results converged to aggregate french samples to one independent cluster (Fig. 3C). The results of all the analyses show a french cluster different from the native clusters (Fig. 3). Admixtures and lower genetic differentiation between the french deme and several native demes exist however (Fig. 3C; Fig. S3).

Discussion

This is the first study assessing the distribution and the prevalence of the rosette agent *S. destruens* among invasive and native fish species in different regions in France. We found that *S. destruens* is widely distributed throughout France, in regions such as Ain (central-eastern France), Indre (central France), Gironde (south-western France), Bouches-du-Rhône and Corsica island (south-eastern France) and exhibiting a range of climates going from temperate, oceanic and Mediterranean, respectively, but also in contrasting freshwater habitats such as lotic and lentic ones. Moreover, our broad environmental screening of *S. destruens* distribution in France shows new native fish species infected with this parasite. The disease prevalence ranging between 2% and 20% is in the range of those previously found in wild fish in other countries (Combe & Gozlan, 2018), such as in the Netherlands where a prevalence of *S. destruens* of about 25% has been found in the native *G. aculeatus* (Spikmans et al., 2020), and other countries (Combe & Gozlan, 2018). However, the rosette agent was not detected in all host populations of *P. parva* despite a homogeneous genetic background between french populations and a similar invasion history. These findings lead us to raise two hypothesis: 1) a loss of pathogen has occurred due to its absence in the founding *P. parva* population, either because a small number of pathogen-free founders (i.e. *P. parva* hosts for which the pathogen was not able to survive/replicate in their native range, highly unlikely) has invaded the new location or 2) because although present in *P. parva* founder hosts *S. destruens* was unable to survive translocation (i.e. actual absence, unlikely) or 3) the presence of the rosette agent in each *P. parva* individual (i.e. asymptomatic carrier) is of such low prevalence and potentially associated with a non-homogeneous distribution in fish organs that our detection protocol could not detect its presence at a site (likely). Based on its co-introduction with *P. parva* into Europe, the wide distribution of *P. parva* in France and previous reports of *S. destruens*'s presence and impact on freshwater fish biodiversity (reviewed in Combe & Gozlan, 2018) the first scenario seems highly

466 unlikely. Interestingly, we found *P. parva* hosts infected with the rosette agent in Ain (central-
467 eastern, France), Bouches-du-Rhône and Corsica island (south-eastern France) and Indre
468 (central France), regions that are far away from each other, notably for Corsica island that is
469 separated from the continent by the Mediterranean Sea and where high virulence levels of
470 *S. destruens* were reported (personal communication). Surprisingly, while in some regions
471 neither the host *P. parva* nor native fish species were infected with the rosette agent, in
472 Gironde *P. parva* individuals seem parasite-free although native species such as *R. rutilus*
473 and *Gobio sp.* were *S. destruens* positive. Taken together, our findings along with the results
474 from Spikmans et al. (2020) showing the strong correlation between *S. destruens*
475 occurrence and *P. parva* presence *via* water body analyses, tend to indicate that *S.*
476 *destruens* is widely distributed across freshwater ecosystems in France. Furthermore,
477 despite its presence and likely introduction by *P. parva* hosts, detecting *S. destruens* in
478 asymptomatic carriers who typically have a low copy number of the infective agent remains
479 challenging. Moreover, it is worth noting that prior to the study of Ercan et al. (2015) showing
480 large fish population collapses in Turkey, no one had noticed the significant increase in fish
481 mortality rates across Europe following *P. parva* and *S. destruens* co-invasion, despite
482 laboratory evidence. This lack of disease detection can also be attributed to 1) the chronic
483 nature of the disease leading to slow and periodic mortality rates in the wild, 2) the difficulties
484 in monitoring fish mortality in freshwater systems and 3) the non-inclusion of this parasite in
485 the existing list of noticeable diseases in Europe, including France (Combe & Gozlan, 2018).
486 Here our results clearly highlight the risk of disease emergence in wild fish in France and
487 beyond, and potentially in fish of economic importance. Thus, it is necessary for
488 policymakers and stakeholders to react quickly to control its spread from already
489 contaminated reservoirs to still healthy ecosystems.

490 It has been hypothesized that different *S. destruens* strains (i.e. isolates) could have been
491 co-introduced in Europe with distinct *P. parva* lineages and sequencing of the ITS-1 genetic
492 marker of *S. destruens* strains has shown that European strains (UK and Turkey) cluster

493 with respect to North-American strains but do not intermingle with each other and are closely
494 related to Chinese strains (Ercan et al., 2015; Sana et al., 2017; Hardouin et al., 2018). This
495 was consistent with the invasion history of *P. parva* populations in Europe (Combe & Gozlan,
496 2018; Brazier et al., 2021), thus raising the possibility that host-parasite populations
497 eventually invaded northern and southern France from at least two sources of native
498 population, originating from north and south of the Yangtze River, respectively. Such
499 scenarios of an ancient host-parasite coevolution and co-introduction could have direct
500 implications on disease risk, e.g. increased transmission and virulence of the pathogen to
501 naive fish populations. For example, all identified strains of *S. destruens* have been shown to
502 be highly virulent to local fish populations in England, Turkey and the USA, although they are
503 asymptomatic in *P. parva* populations (Combe & Gozlan, 2018). Due to the absence of *S.*
504 *destruens* strains in France, we performed here an analysis of the genetic diversity of french
505 *P. parva* populations to check their potential heterogeneity and admixed origin of
506 introduction, as recently found for some other European populations (Brazier et al., 2021).
507 Surprisingly, our molecular screening and genetic assignment of the nine sampled groups
508 revealed a lack of genetic and geographic structure within and between *P. parva*
509 populations, with the French populations representing a single homogeneous gene pool.
510 Furthermore, no strong links are shared between french and native populations, with the
511 entire French range constituting an independent genetic pool. Therefore, our results suggest
512 that the French (continental and insular) population of *P. parva* could not have originated
513 from a direct introduction from the native range, but is likely to have originated from a
514 successful invasive population through an invasive bridgehead effect. Indeed, bridgehead
515 effects could stimulate the invasion process and could be due to either the evolution of
516 higher invasiveness in the primary introduced population (e.g. acquisition of new traits
517 increasing establishment and spread success), or to an increased abundance in the
518 bridgehead region, leading to a higher probability of movement to new non-native habitats
519 compared to native populations (Lombaert et al., 2010; Bertelsmeier & Keller, 2018).
520 However, empirical evidence to confirm these assumptions is still lacking. Nevertheless,

521 evidence of successful invasive populations becoming the source of new introductions in
522 distant territories is increasingly available for several invasive species such as the red
523 swamp crayfish (*Procambarus clarkii*), grapevine downy mildew (*Plasmopara viticola*) and
524 common ragweed (*Ambrosia artemisiifolia*) (van Boheemen et al., 2017; Oficialdegui et al.,
525 2019; Fontaine et al., 2021). Recently, Brazier et al. (2021) found a similar homogeneous
526 and specific pattern for *P. parva* populations in Italy, raising the question of the genetic
527 structure of other invaded Mediterranean regions such as Spain or Morocco.

528 Although our results require further analysis by adding the range of invasions, they provide
529 insight into the complex origin of the French *P. parva* populations, and allow us to infer the
530 origin of *S. destruens* strains within a country, as observed for France. Overall low but
531 equivalent genetic diversity was observed within each French group (Table 1). Nevertheless,
532 there was a clear reduction in genetic diversity within the French deme compared to each
533 native deme, highlighting that the invasion of *P. parva* in France was accompanied by a
534 substantial loss of genetic diversity (Fig. 3D). Therefore, if the co-introduction of *P. parva/S.*
535 *destruens* is considered as a host-pathogen complex, we can assume that French
536 populations of the rosette agent should also show low genetic variation. However, further
537 genetic analysis will be required to clearly assess the potential pathway of introduction of *P.*
538 *parva* populations into France from primary European invasive demes, the resulting genetic
539 diversity of the co-invasive pathogen and the potential high virulence of *S. destruens* strains
540 in France.

541

542 **Conclusion**

543 Although showing relatively low levels of prevalence, *S. destruens* is widespread among
544 invasive and native fish populations from contrasting regions of France, suggesting that
545 there is a high risk of disease emergence. We also found that co-introduced *P. parva*
546 populations represent a homogeneous gene pool between these regions, indicating a

potential comparable homogeneity of the rosette agent population. Overall, our results provide a better understanding of pathogen emergence resulting from secondary dispersal of a successful invasive population in a host-pathogen co-introduction context, paving the path for the transition from invasion biology to invasion epidemiology to assess risks of disease emergence associated with biological invasions.

Acknowledgments

This work was funded by the ROSETTA project (AFB-IRD) and E.C received a post-doctoral fellowship from IRD. We acknowledge Sylvain Santoni for his precious experience and advice in GBS data analysis, as well as Mélanie Lesne, Laurie Lamothe, Emilie Merle, Carine Bellet and Mélanie Régo for their implication in DNA extractions and molecular biology at the LPL laboratory.

Conflict of interest

The authors declare no conflict of interest.

Credit author statement

Combe Marine: Conceptualization, Data curation, Original draft preparation

Cherif Emira: Data curation, Software, Formal analysis, Original draft preparation

Charrier Amélie: Methodology, Investigation, Data validation

Barbey Bruno, Chague Martine, Carrel Georges, Chasserieau Céline, Foissy Jean-Michel, Gerard Barbara, Guillouët Jérôme, Hérodet Benjamin, Laine Manon, Masseboeuf Fabrice, Mirkovic Ivan, Nicolas Delphine, Poulet Nicolas: Methodology, Investigation

Gozlan Zachary: Software and mapping

Martin Jean-François, Gilles André: Data curation, Software, Formal analysis

573 **Gozlan Rodolphe Elie:** Conceptualization, Supervision, Funding acquisition, Original draft
574 preparation
575 **ALL:** Commenting on the manuscript
576
577 **References**
578 Allardi, J., & Chancerel, F. (1988). Note Ichtyologique-Sur la présence en France de
579 Pseudorasbora parva (Schlegel, 1842). *Bulletin Français de la Pêche et de la Pisciculture*,
580 (308), 35-37.
581
582 Al-Shorbaji, F., Roche, B., Britton, R., Andreou, D., & Gozlan, R. (2017). Influence of
583 predation
584 on community resilience to disease. *Journal of Animal Ecology*, *86*(5): 1147–1158.
585 doi.org/10.1038/emi.2016.46.
586
587 Arkush, K.D., Frasca, S., & Hedrick, R.P. (1998). Pathology associated with the rosette
588 agent,a systemic protist infecting salmonid fishes. *Journal of Aquatic Animal Health*, *10*(1):
589 1–11. doi.org/10.1577/1548-8667(1998)010<0001:PAWTRA>2.0.CO;2.
590
591 Bertelsmeier, C. & Keller, L. (2018). Bridgehead effects and role of adaptive evolution in
592 invasive populations. *Trends in Ecology & Evolution*, 33:7.
593 doi.org/10.1016/j.tree.2018.04.014.
594
595 Brazier, T., CHERIF, E., Martin, J. F., Gilles, A., Blanchet, S., Zhao, Y., ... & Gozlan, R.
596 (2021). A tale of an invader: Reconstructing the genomic history of invasive topmouth
597 gudgeon (Pseudorasbora parva) populations. *Authorea Preprints*.
598 doi.org/10.22541/au.161417292.23392023/v1.
599
600 Boitard, P.M., Charrier, A., Labrut, S., & Jamin, M. (2017). First detection of *Sphaerothecum*

destruens in salmonids in France. *Bulletin of the European Association of Fish Pathologists*, 37: 198–204.

Browning, S.R., & Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097. doi.org/10.1086/521987.

Browning, B.L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, *103*(3), 338–348. doi.org/10.1016/j.ajhg.2018.07.015.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, *22*(11), 3124–3140. doi.org/10.1111/mec.12354.

Chapin, F.S.C., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., et al. (2000). Consequences of changing biodiversity, Nature, 405: 234–42. doi.org/10.1038/35012241.

Charrier, A., Peudpiece, M., Lesne, M. & Daniel, P. (2016). First report of the intracellular fish parasite *Sphaerothecum destruens* associated with the invasive topmouth gudgeon (*Pseudorasbora parva*) in France. *Knowledge & Management of Aquatic Ecosystems*, 417:44. doi.org/10.1051/kmae/2016031.

Combe, M. & Gozlan, R.E. (2018). The rise of the rosette agent in Europe: an epidemiological enigma. *Transbound & Emerging Diseases*, 65: 1474–1481. doi.wiley.com/10.1111/tbed.13001

627 Crowl, T.A., Crist, T.O., Parmenter, R.R., Belovsky, G., & Lugo, A.E. (2008). The spread of
628 invasive species and infectious disease as drivers of ecosystem change. *Frontiers in*
629 *Ecology and the Environment*, *6*(5), 238–246. doi.org/10.1890/070151
630
631 Daszak, P., Cunningham, A.A., & Hyatt, A.D. (2000). Emerging infectious diseases of wildlife
632 - threats to biodiversity and human health. *Science*, 287(5452), 443–449.
633 doi.org/10.1126/science.287.5452.443.
634
635 Ercan, D., Andreou, D., Sana, S., Öntas, C., Baba, E., Top, N. et al. (2015). Evidence of
636 threat to European economy and biodiversity following the introduction of an alien pathogen
637 on the fungal-animal boundary. *Emerging Microbes and Infection*, 4:9.
638 doi.org/10.1038/emi.2015.52.
639
640 Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. et al.
641 (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature,* 484: 186–
642 194. doi.org/10.1038/nature10947
643
644 Fontaine, M. C., Labbé, F., Dussert, Y., Delière, L., Richart-Cervera, S., Giraud, T., &
645 Delmotte, F. (2021). Europe as a bridgehead in the worldwide invasion history of grapevine
646 downy mildew, Plasmopara viticola. *Current Biology*, 31 (10), 2155–2166.
647 doi.org/10.1016/j.cub.2021.03.00
648
649 Gallardo, B., Clavero, M., Sánchez, M. I., & Vilà, M. (2016). Global ecological impacts of
650 invasive species in aquatic ecosystems. *Global change biology*, *22*(1), 151–163.
651 doi.org/10.1111/gcb.13004
652
653 Gozlan, R.E. (2012). *Pseudorasbora parva* temminck and schlegel (topmouth gudgeon). A
654 Handbook of Global Freshwater Invasive Species Abingdon: Earthscan.

655 Gozlan, R., St-Hilaire, S., Feist, S. *et al.* Disease threat to European fish. *Nature* 435, 1046
656 (2005). https://doi.org/10.1038/4351046a
657
658 Gozlan, R.E., Peeler, E.J., Longshaw, M., St-hilaire, S., & Feist, S.W. (2006). Effect of
659 microbial pathogens on the diversity of aquatic populations, notably in Europe. *Microbes &*
660 *Infection*, 8: 1358–1364. doi.org/10.1016/j.micinf.2005.12.010.
661
662 Gozlan, R.E., Andreou, D., Asaeda, T., Beyer, K., Bouhadad, R., Burnard, D., et al (2010).
663 Pan-continental invasion of *Pseudorasbora parva*: towards a better understanding of
664 freshwater fish invasions. *Fish and Fisheries*, *11*(4), 315–340. doi.org/10.1111/j.1467-
665 2979.2010.00361.x.
666
667 Hardouin, E.A., Andreou, D., Zhao, Y., Chevret, P., Fletcher, D.H., Britton, J.R., et al. (2018).
668 Reconciling the biogeography of an invader through recent and historic genetic patterns: The
669 case of topmouth gudgeon *Pseudorasbora parva*. *Biological Invasions*, 20: 2157–2171.
670 doi.org/10.1007/s10530-018-1693–4.
671
672 Harrell, L. W., Elston, R. A., Scott, T. M., & Wilkinson, M. T. (1986). A significant new
673 systemic disease of net-pen reared chinook salmon (Oncorhynchus tshawytscha) brood
674 stock. *Aquaculture*, *55*(4), 249-262.
675
676 Hatcher, M.J., Dick, J.T., & Dunn, A. M. (2012). Disease emergence and invasions.
677 *Functional Ecology*, 26(6), 1275–1287. doi.org/10.1111/j.1365-2435.2012.02031.x.
678
679 Holdich, D.M., & Pöckl, M. (2007). Invasive crustaceans in European inland waters. In
680 Biological invaders in inland waters: Profiles, distribution, and threats (pp. 29-75). Springer,
681 Dordrecht. Holsinger & Weir, 2009.

Hubbell, S.P. (2001). The unified neutral theory of biodiversity and biogeography. In Monographs in Population Biology. Princeton Univ Press. doi:10.4249/scholarpedia.8822

Hedrick, R.P., Friedman, C.S. & Modin, J. (1989). Systemic infection in Atlantic salmon Salmo salar with a Dermocystidium-like species. Diseases of Aquatic Organisms 7, 171–177.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi.org/10.1093/bioinformatics/btr521.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., et al. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990–993. doi.org/10.1038/nature06536.

Kamvar, Z.N., Tabima, J.F., & Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. doi.org/10.7717/peerj.281.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*, *15*(5), 1179-1191. doi.org/10.1111/1755-0998.12387

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289–1291. doi.org/10.1093/bioinformatics/btm091.

709

710 Lande, R., Engen, S. & Saether, B-E. (2003). Stochastic population dynamics in ecology and
711 conservation. Oxford University Press. doi:10.1093/acprof:oso/9780198525257.001.0001.
712

713 Lawler, J.J., Aukema, J.E., Grant, J.B., Halpern, B.S., Kaveira, P., Nelson, C.R. et al. (2006).
714 Conservation science: a 20-year report card. *Frontiers in Ecology and the Environment*, 4(9):
715 473–480. doi.org/10.1890/1540-9295(2006)4[473:CSAYRC]2.0.CO;2.

716 Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize
717 the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1),
718 176–177. doi.org/10.1111/1755-0998.12719.

719 Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010).
720 Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PloS*
721 *one*, *5*(3), e9743. doi.org/10.1371/journal.pone.0009743
722

723 Lymbery, A.J., Morine, M., Kanani, H.G., Beatty, S., Morgan, D.L. (2014). Co-invaders: The
724 effect of alien parasites on native hosts. *International Journal of Parasitology: Parasites and*
725 *Wildlife*, 3: 171–177. doi.org/10.1016/j.ijppaw.2014.04.002.
726

727 Oficialdegui, F.J., Clavero, M., Sánchez, M.I., Green, A.J., Boyero, L., Michot, T.C., et al.
728 (2019). Unravelling the global invasion routes of a worldwide invader, the red swamp
729 crayfish (*Procambarus clarkii*). *Freshwater Biology*, *64*(8), 1382-1400.
730 doi.org/10.1111/fwb.13312
731

732 Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An
733 improved method for automation and parallelization of programs structure, fastStructure and
734 MavericK on multicore CPU systems. *Molecular Ecology Resources*, *17*(6), e268-e274.

735 doi:10.1111/1755-0998.12702

736

737 Poulet, N., Beaulaton, L., & Dembski, S. (2011). Time trends in fish populations in
738 metropolitan France: insights from national monitoring data. *Journal of Fish Biology*, *79*(6),
739 1436-1452. doi.org/10.1111/j.1095-8649.2011.03084.x

740

741 Price, P. W., Westoby, M., Rice, B., Atsatt, P. R., Fritz, R. S., Thompson, J. N., & Mobley, K.
742 (1986). Parasite mediation in ecological interactions. *Annual review of ecology and*
743 *systematics*, *17*(1), 487-505.

744

745 Puechmaille, S.J. (2016). The program structure does not reliably recover the correct
746 population structure when sampling is uneven: subsampling and new estimators alleviate the
747 problem. *Molecular Ecology Resources*, *16*(3), 608–627. doi.org/10.1111/1755-0998.12512.

748

749 Raj, A., Stephens, M., & Pritchard, J.K. (2014). fastSTRUCTURE: variational inference of
750 population structure in large SNP data sets. *Genetics*, *197*(2), 573–589.
751 doi.org/10.1534/genetics.114.164350

752

753 Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global
754 change? *Conservation Biology: the journal of the society for conservation biology*, 21(2):
755 329–336. doi.org/10.1111/j.1523-1739.2006.00615.x.

756

757 Sana, S., Hardouin, E.A., Gozlan, R.E., Ercan, D., Tarkan, A.S., Zhang, T., et al. (2017).
758 Origin and invasion of the emerging infectious pathogen *Sphaerothecum destruens*.
759 *Emerging Microbes & Infections*, 6(8): e76. doi.org/10.1038/emi.2017.64.

760

761 Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., et al. (2011). Fast,
762 scalable generation of high-quality protein multiple sequence alignments using Clustal

763 Omega. *Molecular Systems Biology*, 7:539. doi.org/10.1038/msb.2011.75.
764
765 Spikmans, F., Lemmers, P., op den Camp, H. J., van Haren, E., Kappen, F., Blaakmeer, A.,
766 et al. (2020). Impact of the invasive alien topmouth gudgeon (*Pseudorasbora parva*) and its
767 associated parasite *Sphaerothecum destruens* on native fish species. *Biological Invasions*,
768 22(2), 587-601. doi.org/10.1007/s10530-019-02114-6.
769
770 Stiers, I., Crohain, N., Josens, G., & Triest, L. (2011). Impact of three aquatic invasive
771 species on native plants and macroinvertebrates in temperate ponds. *Biological Invasions*,
772 *13*(12), 2715–2726. doi.org/10.1007/s10530-011-9942-9.
773
774 Tompkins, D.M., White, A.R., & Boots, M. (2003). Ecological replacement of native red
775 squirrels by invasive greys driven by disease. *Ecology Letters*, 6(3), 189–196.
776 doi.org/10.1046/j.1461-0248.2003.00417.x.
777
778 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen,
779 S. G. (2012). Primer3—new capabilities and interfaces. Nucleic acids research, 40(15),
780 e115–e115. doi.org/10.1093/nar/gks596.
781
782 van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., &
783 Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion
784 characterize the introduction history of Ambrosia artemisiifolia in Europe and Australia.
785 *Molecular Ecology*, *26*(20), 5421–5434. doi.org/10.1111/mec.14293.
786
787 Woolhouse, M.E.J., & Gowtage-Sequeria, S. (2005). Host range and emerging and
788 Reemerging pathogens. *Emerging Infectious Diseases*, 11(12), 1842–1847.

789 doi.org/10.3201/eid1112.050997.

790

791 Zhang, C., & Zhao, Y. (2016). Species diversity and distribution of inland fishes in China.
792 Beijing: Science Press.

793

794 **Figures caption**

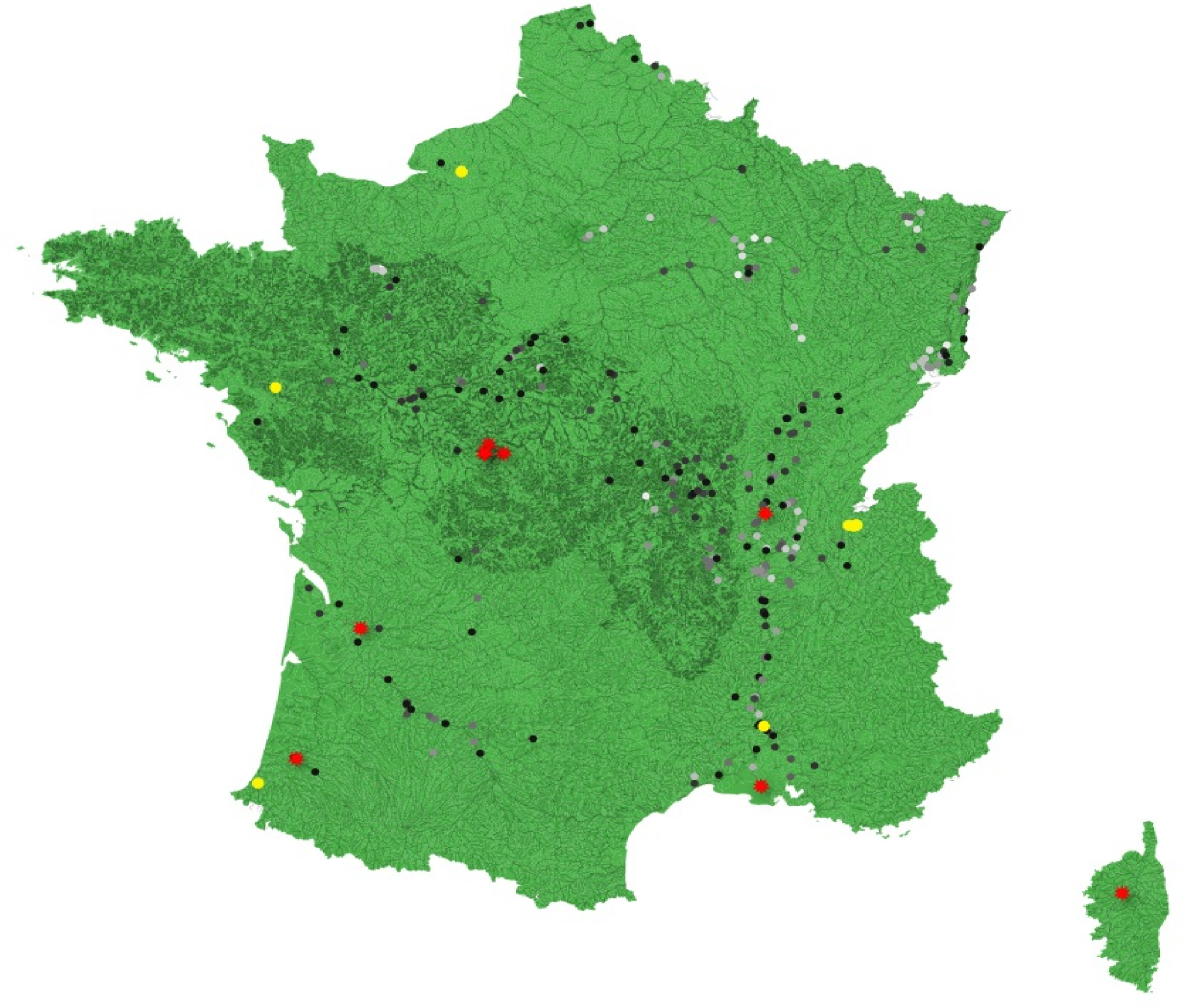
795 **Fig. 1.** Distribution of *Pseudorasbora parva* and the fish pathogen *Sphaerothecum destruens*
796 in France. The gradient from light grey dots to black dots indicates the distribution of *P.*
797 *parva* populations in 1995, 2000, 2005 and 2007 respectively; Yellow dots are sites where *S.*
798 *destruens* was tested and not detected. Red dots are sites where *S. destruens* was tested
799 and found in fish species. Rivers systems are shown in light grey.

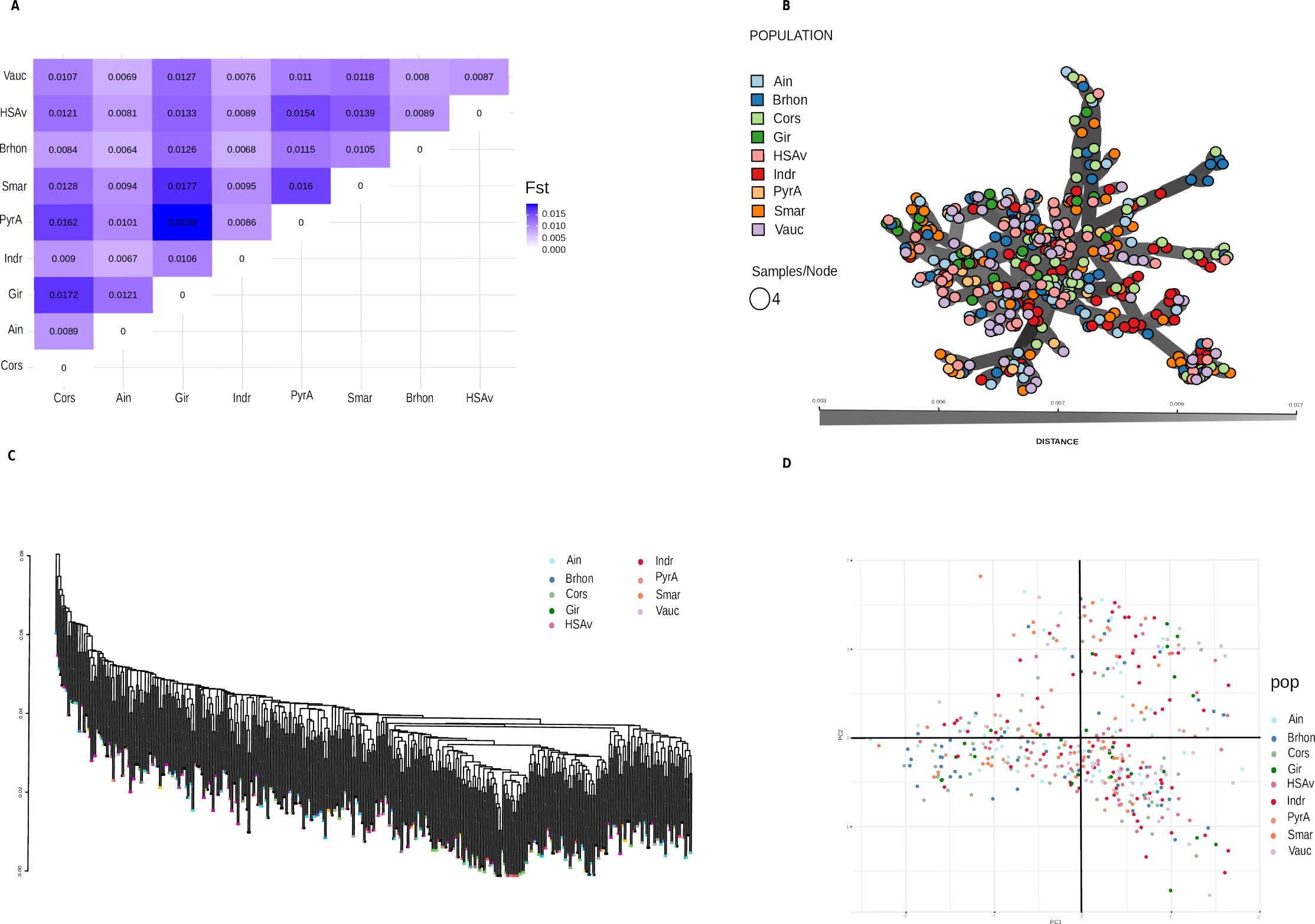
800

801 **Fig. 2.** Genetic diversity and structure of the French *P. parva* groups. (A) Pairwise Fst. (B)
802 Minimum spanning network (MSN), each node is a genotype and the edges represent the
803 genetic distance, nodes with identical genetic distance are connected, lighter the edges, the
804 longer the distance. (C) Hierarchical clustering. (D) Principal component analysis.

805

806 **Fig. 3.** Genetic clustering and relationship of the french deme of *P. parva* with the native
807 range. (A) Pairwise **Φ**st. (B) Multidimensional scaling plot of the french deme and native
808 range. based on the identity by state (IBS) distance (C) Discriminant analysis of principal
809 components of the french and native range populations, axes represent the first two linear
810 discriminants (LD). (D) Nucleotide diversity in the french deme and in each native range.




A
Vauc 0.0107 0.0069 0.0127 0.0076 0.011 0.0118 0.008 0.0087
HSAv 0.0121 0.0081 0.0133 0.0089 0.0154 0.0139 0.0089 0
Brhon 0.0084 0.0064 0.0126 0.0068 0.0115 0.0105 0
Smar 0.0128 0.0094 0.0177 0.0095 0.016 0
PyrA 0.0162 0.0101 0.0199 0.0086 0
Indr 0.009 0.0067 0.0106 0
Gir 0.0172 0.0121 0
Ain 0.0089 0
Cors 0
Cors Ain Gir Indr PyrA Smar Brhon HSAv
Fst
0.015
0.010
0.005
0.000
B
POPULATION
Ain
Brhon
Cors
Gir
HSAv
Indr
PyrA
Smar
Vauc
Samples/Node
4
0.003 0.006 0.007 0.009 0.017
DISTANCE
C
Ain
Brhon
Cors
Gir
HSAv
Indr
PyrA
Smar
Vauc
0.08 0.06 0.04 0.02 0.00
D
PC2
PC1
pop
Ain
Brhon
Cors
Gir
HSAv
Indr
PyrA
Smar
Vauc

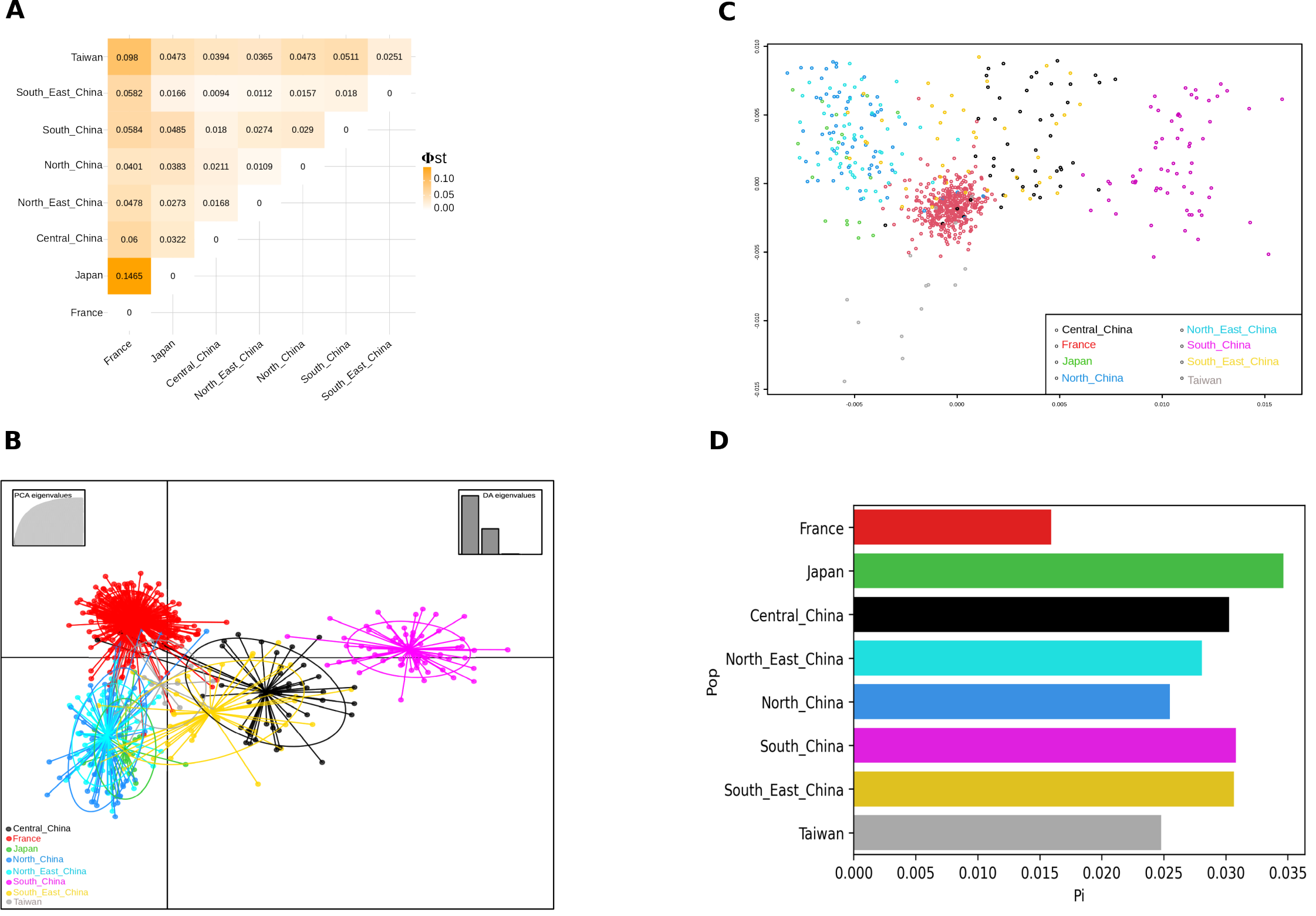

A
Taiwan 0.098 0.0473 0.0394 0.0365 0.0473 0.0511 0.0251
South_East_China 0.0582 0.0166 0.0094 0.0112 0.0157 0.018 0
South_China 0.0584 0.0485 0.018 0.0274 0.029 0
North_China 0.0401 0.0383 0.0211 0.0109 0
North_East_China 0.0478 0.0273 0.0168 0
Central_China 0.06 0.0322 0
Japan 0.1465 0
France 0
Φst
0.10
0.05
0.00
France
Japan
Central_China
North_East_China
North_China
South_China
South_East_China
C
Central_China
France
Japan
North_China
North_East_China
South_China
South_East_China
Taiwan
B
PCA eigenvalues
DA eigenvalues
Central_China
France
Japan
North_China
North_East_China
South_China
South_East_China
Taiwan
D
France
Japan
Central_China
North_East_China
North_China
South_China
South_East_China
Taiwan
Pop
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
Pi

Table 1. French population properties with geographic origin and diversity indices.

Grp	Region	N	Sites	%Polymorphic_loci	Pa	Ho (SD)	Pi	Fis
Cors	Corse	48	5028	8.8106	31	0.0096(0.0547)	0.0116	0.01696
Ain	Ain	50	5028	12.0326	55	0.0134(0.0734)	0.0131	0.01047
Gir	Gironde	26	5028	8.7509	34	0.0156(0.0821)	0.0150	0.00765
Indr	Indre	69	5028	13.4049	86	0.0162(0.0831)	0.0151	0.01079
PyrA	Pyrénées-Atlantiques	20	5028	6.4240	14	0.0134(0.0801)	0.0132	0.00652
Smar	Seine-Maritime	49	5028	8.5918	54	0.0129(0.0763)	0.0127	0.01095
Brhon	Bouches-du-Rhone	50	5028	11.1177	37	0.0116(0.0613)	0.0127	0.01527
HSAv	Haute-Savoie	50	5028	9.8050	33	0.0145(0.0791)	0.0138	0.00868
Vauc	Vaucluse	47	5028	12.6690	41	0.0166(0.0848)	0.0151	0.00469

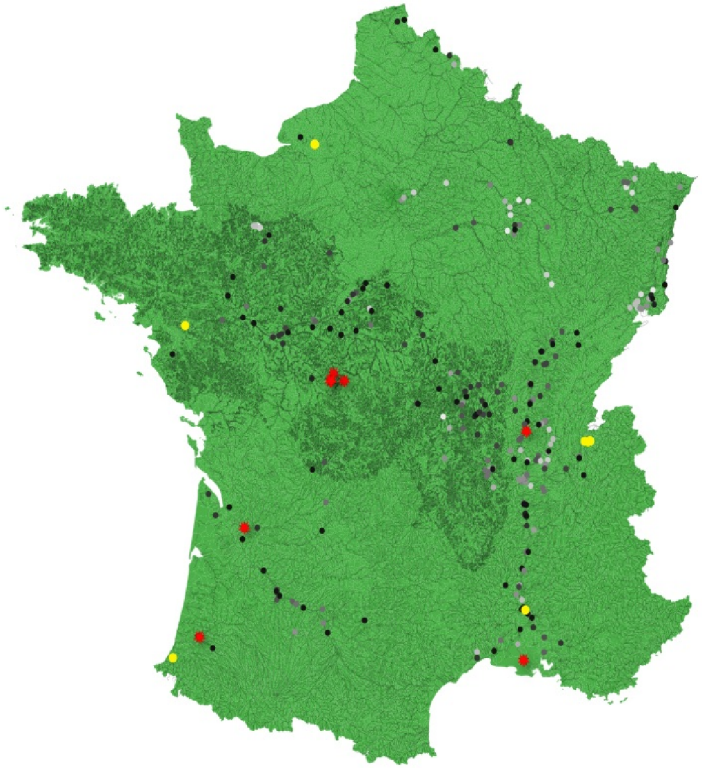
Note: Grp IDs will be used as tags in the manuscript.

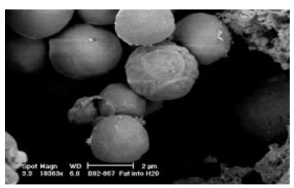
Abbreviations: Grp, population; N, number of genotyped individuals; Sites, number of fixed and variant sites; %Polymorphic, percentage of polymorphic loci; PA, number of private alleles within each population; Ho, observed heterozygosity; SD, Standard Deviation; Pi, nucleotide diversity.

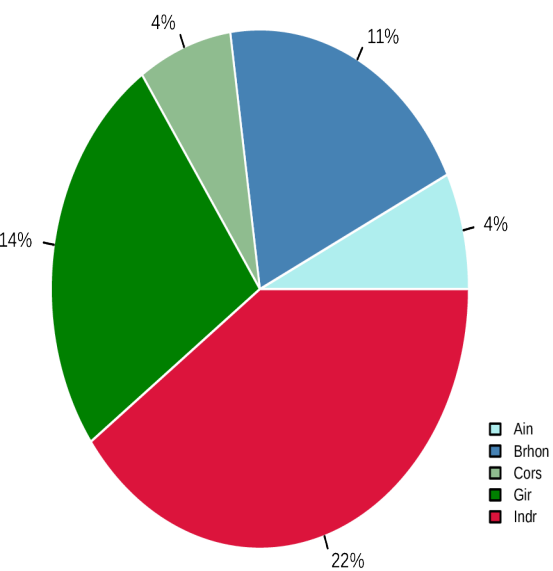
Table 2. Detection of *S. destruens* DNA in invasive and native fish populations in France. Region: angling association (abbreviation); Date: sampling date; N: number of individuals; qPCR: number of positive individuals; Ct: Ct-value; DV: disease prevalence (%); NA: missing data. Whilst three species of gudgeon and 5 species of minnows are reported in France, *Phoxinus sp.* and *Gobio sp.* were not identified at the species level here.

Region	Date	Genus/Species	N	qPCR	Ct	DV (%)
Ain (Ain)	06/09/17	*Pseudorasbora parva*	50	2	33.7 31.1	4
		Phoxinus sp.	50	0	0	0
		Perca fluviatilis	18	0	0	0
		Rutilus rutilus	13	0	0	0
		Chondrostoma nasus	19	0	0	0
Corse (Cors)	19/09/18	*Pseudorasbora parva*	50	1	33.6	2
		Phoxinus sp.	50	1	29.2	2
		Salmo trutta	38	0	0	0
Bouches-du-Rhône (Brhon)	01/10/18	*Pseudorasbora parva*	50	1	36.7	2
		Lepomis gibbosus	36	0	0	0
		Abramis brama	22	0	0	0
		Rutilus rutilus	2	0	0	0
		Scardinius erythrophthalmus	1	0	0	0
	26/06/19	*Pseudorasbora parva*	50	0	0	0
		Lepomis gibbosus	29	0	0	0
		Alburnus alburnus	11	1	36.1	9
		Abramis brama	1	0	0	0
		Cyprinus carpio	1	0	0	0
Gironde (Gir)	15/11/18	*Pseudorasbora parva*	26	0	0	0
		Rutilus rutilus	50	2	33.9 36.5	4
		Gobio sp.	10	1	37.8	10
		Squalius cephalus	11	NA	NA	NA
		Lepomis gibbosus	11	NA	NA	NA
		Phoxinus sp.	11	NA	NA	NA
		Anguilla anguilla	6	NA	NA	NA
Indre (Indr)	26/09/17	*Pseudorasbora parva*	50	1	36.64	2
		Rutilus rutilus	50	0	0	0
		Abramis brama	10	0	0	0
		Carassius carassius	10	0	0	0
		Rhodeus amarus	10	2	36.1 36.2	20
		Ameiurus melas	10	0	0	0
		Perca fluviatilis	10	0	0	0

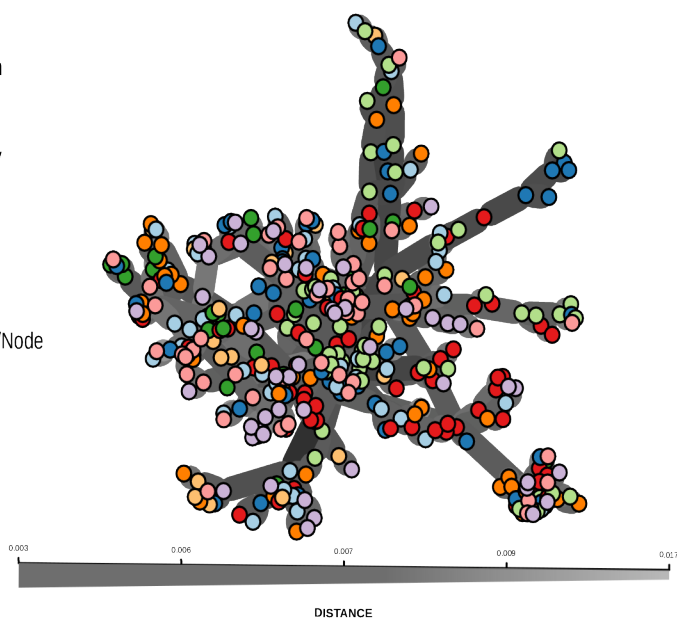
Loire Atlantique (LoirA)	15/03/18	*Pseudorasbora parva*	9	0	0	0
Pyrénées Atlantiques (PyrA)	20/10/17	*Pseudorasbora parva*	20	0	0	0
Haute Savoie (HSAv)	24/10/17	*Pseudorasbora parva*	50	0	0	0
	09/10/17	*Salmo trutta*	20	0	0	0
	09/10/17	*Squalius cephalus*	20	0	0	0
	09/10/17	*Chondrostoma nasus*	5	0	0	0
	09/11/17	*Cyprinus carpio*	4	0	0	0
	09/11/17	*Barbus barbus*	20	0	0	0
Seine Maritime (Smar)	05/09/17	*Pseudorasbora parva*	49	0	0	0
		Lepomis gibbosus	80	0	0	0
		Scardinius erythrophthalmus	7	0	0	0
		Abramis brama	2	0	0	0
		Carassius carassius	7	0	0	0
		Rutilus rutilus	2	0	0	0
Vaucluse (Vauc)	17/07/18	*Rutilus rutilus*	27	0	0	0
		Blicca bjoerkna	18	0	0	0
		Squalius cephalus	50	0	0	0
	15/05/19	*Pseudorasbora parva*	47	0	0	0
		Chondrostoma nasus	8	0	0	0
		Rutilus rutilus	13	0	0	0
		Gobio sp.	22	0	0	0
		Rhodeus amarus	20	0	0	0
		Blicca bjoerkna	8	0	0	0
		Scardinius erythrophthalmus	3	0	0	0
		Alburnus alburnus	30	0	0	0
		Chondrostoma toxostoma	2	0	0	0






4%
11%
4%
14%
22%
Ain
Brhon
Cors
Gir
Indr


POPULATION
Ain
Brhon
Cors
Gir
HSAv
Indr
PyrA
Smar
Vauc
Samples/Node
4


0.003
0.006
0.007
0.009
DISTANCE