Combining species distribution models and population genomics underlines the determinants of range limitation in an emerging parasite

Eglantine Mathieu-Bégné, Géraldine Loot, Elise Mazé-Guilmo, Valentin Mullet, Clémence Genthon, Simon Blanchet

▶ **To cite this version:**

HAL Id: hal-03129226
https://hal.inrae.fr/hal-03129226

Submitted on 2 Feb 2021

ECOGRAPHY

Research

Combining species distribution models and population genomics underlines the determinants of range limitation in an emerging parasite

Eglantine Mathieu-Bégné, Géraldine Loot, Elise Mazé-Guilmo, Valentin Mullet, Clémence Genthon and Simon Blanchet

E. Mathieu-Bégné (https://orcid.org/0000-0002-6846-5491) ✉ (eglantine.mb@gmail.com), G. Loot, E. Mazé-Guilmo, V. Mullet and S. Blanchet, Evolution et Diversité Biologique, Univ. Paul Sabatier (UPS), Centre National de la Recherche Scientifique, Ecole Nationale pour la Formation Agronomique (ENFA), Toulouse, France. GL also at: Inst. Univ. de France (IUF), Paris, France. EM-G, VM and SM also at: Station d'Ecologie Théorique et Expérimentale, CNRS, UPS, Moulis, France. – C. Genthon, INRA, Genotoul, Castanet-Tolosan, France.

Understanding the drivers of species distribution is necessary in order to properly predict the future geographical ranges of colonizing species. Yet this task is challenging for species involved in intimate interactions, such as parasites, since their distribution is likely shaped by a complex interplay between environmental-related and host-related factors. Here we developed an original approach combining species distribution models (SDMs) and population genomics to test whether the local environmental conditions or the host genomic background most likely limits the colonization of an emerging freshwater fish ectoparasite, *Tracheliastes polycolpus*. We hypothesized that the absence of *T. polycolpus* in some areas may be due to an unsuitable environment in these areas (the 'environmental suitability hypothesis') and/or to the presence of resistant hosts in these areas (the 'genomic background hypothesis'). Using a SDM set at the French spatial extent, we first found that the environmental conditions of an uninfected area were as suitable for the parasite as those of infected areas. Then, using single nucleotide polymorphisms (SNPs) data at the host genome scale, we demonstrated that there was a strong association between the spatial occurrence of parasites and the host genomic background. In particular, the area in which the parasite was absent sustained a unique host population from a genomic standpoint, and ninety SNPs were significantly associated to the infection status (parasitized versus unparasitized) of individual hosts. We concluded that the spatial distribution of *T. polycolpus* (and its colonization potential) was more likely limited by intrinsic host characteristics associated to parasite resistance, rather than to the local environmental conditions. This study illustrates the usefulness of combining modeling and genomic approaches to reveal the determinants of species distribution and to improve predictions about their future ranges.

Keywords: freshwater, genome wide association study, integrative approaches, invasive species, population genomics, species distribution model

Introduction

Determining the factors limiting species ranges has long fascinated ecologists (Gaston 2009, Wiens et al. 2010). This issue has recently gained particular attention because of the major implications of assessing the future geographical ranges of species facing global change (Guisan and Thuiller 2005, Moor et al. 2015, Narouei-Khandan et al. 2016). This is especially true for emerging parasites that have already been observed to expand (or shift) their ranges due to human activities, which can have direct negative feedback on human health and biodiversity (Kutz et al. 2005, Morgan et al. 2009, Escobar et al. 2017). Nonetheless, parasites are involved in a triple interaction between the parasite, the host and their common environment (Wolinska and King 2009). Hence, both the ecological conditions of the surrounding environment and the factors related to host defense and/or parasite virulence are likely to determine parasite spatial distribution (Grosholz 1994, Bozick and Real 2015).

Studies on the environmental determinants of parasite distribution have mostly focused on the abiotic features of the environment that affect the demographic dynamics of parasites (Lachish et al. 2013, Pérez-Rodríguez et al. 2013, Pickles et al. 2013). For instance, in many parasites, a higher temperature has been shown to be associated with a broader seasonal window for an efficient transmission to hosts or vectors, which facilitates the colonization of new areas (Kutz et al. 2005, Paull et al. 2015, Caminade et al. 2019). Nonetheless, Lafferty (2009) argued that non-climatic environmental factors should also be considered in order to understand and predict the spatial distribution of parasites. In particular, for parasites inhabiting aquatic ecosystems a broad range of environmental factors (e.g. water temperature, precipitations, water discharge, habitat configuration etc.) that can be strongly structured spatially (e.g. along the upstream–downstream gradient in rivers) have been shown to influence parasite distribution (Cardon et al. 2011, Blasco-Costa et al. 2013, Marcogliese 2016). Given the complexity of host–parasite relationships, the most recent modeling approaches actually tend to integrate both abiotic (e.g. climate, but also land cover, or habitat fragmentation) and biotic factors (e.g. hosts distribution, hosts density, biotic interactions) as potential environmental determinants of the spatial distribution of parasites (Giannini et al. 2013, Gehman et al. 2018, Byers et al. 2019).

Alternatively, many studies have aimed to understand the distribution of parasites by focusing on the individual characteristics of hosts (physiology, behavior, immunity) or parasites (virulence factors, infectivity) that underlie infection levels (Grosholz 1994, Morand et al. 1996, Mougeot 2005). The host genomic background (e.g. the diversity of genes involved in the major histocompatibility complex) seems to be a key determinant of the individual infection status (Bernatchez and Landry 2003, Wegner et al. 2003, Aguilar et al. 2016). In geographic areas in which founder effects and/or bottlenecks have led to a heterogeneous spatial distribution of alleles, parasite distribution could be spatially limited because of the presence of specific alleles conferring resistance (Zhang et al. 2015, Charbonnel et al. 2019). Next-Generation Sequencing approaches have recently opened new research avenues by giving access to thousands of both non-coding and coding markers related to host defense at the genome-wide level (Wenzel et al. 2015, Benavides et al. 2016, Zueva et al. 2018).

Although studies that focus on the influence of host and parasite characteristics on the distribution of parasites are usually set at small spatial scales (but see Schwabl et al. 2017, Vajana et al. 2018), they provide crucial complementary information to that gathered at larger spatial scales from approaches linking environmental features to parasite distribution (Cardon et al. 2011, Fourcade et al. 2014, Vajana et al. 2018). The distribution of parasite infecting poïkilotherms hosts has been shown to be driven by multiple environmental features acting either directly on the parasite life-cycle or indirectly by altering host characteristics (Cardon et al. 2011, Blasco-Costa et al. 2015). For instance, lower water flow has been demonstrated to suppress fish immune response which subsequently altered fish parasite distribution (Bakke and Harris 1998, Marcogliese 2001). However, only a very few studies have simultaneously investigated the relative (or joint) influence of environmental determinants (sensu largo) and host individual characteristics (and notably the role of their genomic background) on parasite distribution (Barrett et al. 2013, Fourcade et al. 2014, Schwabl et al. 2017).

To provide such an integrative assessment of the determinants of parasite distribution, we focused on the fish ectoparasite, *Tracheliastes polycolpus* which has rapidly emerged in French watersheds from northeastern Europe (Rey et al. 2015). *Tracheliastes polycolpus* was accidentally introduced in France in the Loire River Basin in the 1920s, most probably by the introduction of its main fish host, *Leuciscus idus*, for aquaculture (Rey et al. 2015). In 15–20 yr *T. polycolpus* has expanded its range from the Loire River Basin to almost all French river basins (Rey et al. 2015). However, repeated surveys by local managers never recorded infection by *T. polycolpus* in the northeastern river basins of France such as the Seine, the Rhin and the Meuse River Basins (Fig. 1). These three river basins thus constitute an intriguing geographically delineated unparasitized area. This is highly surprising because 1) the parasite has been able to rapidly colonize all other river basins in France (which suggests a high dispersal ability), and 2) it is found in sites directly adjacent to these unparasitized river basins (see the inset in Fig. 1) which are connected through man-made channels (that impede host dispersal but not the dispersal of the parasite, Mazé-Guilmo 2016). Examining the environmental determinants and the host individual characteristics that potentially limit the parasite's range expansion is an important, yet challenging, objective that could be fulfilled using integrative approaches.

Here, we aimed to identify the environmental determinants and the individual host characteristics that underlie the spatial distribution of *T. polycolpus*, and that hence explain why its range expansion is currently limited. Specifically, we tested two non-exclusive hypotheses. First, the 'environmental suitability hypothesis' states that environmental features (sensu *largo*) are driving the distribution of *T. polycolpus* in
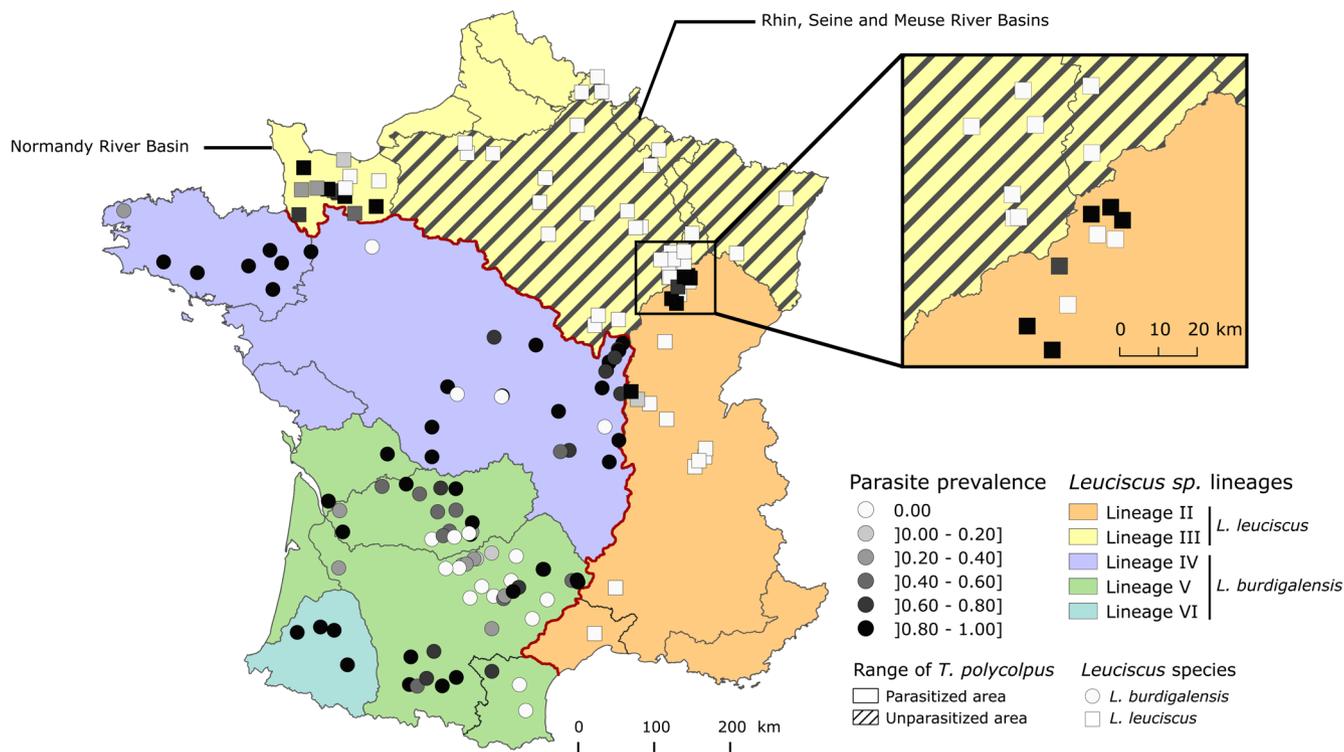
Figure 1. Map showing the distribution of sampled sites used in the SDM approach for the two *Leuciscus* species (dots, *Leuciscus burdigalensis*; squares, *Leuciscus leuciscus*), together with their levels of parasite prevalence. The main *Leuciscus* lineages are also displayed. Grey lines delimitate the main River Basins (and some are indicated). The red line shows range delimitation between the two *Leuciscus* species. The inset is an enlarged representation of the area included in the black square in the main map. Finally, the hatched area refers to the area where hosts are consistently unparasitized. The *L. leuciscus* lineage I is not represented on this map since its range covers central European rivers and is absent in France (Costedoat et al. 2014).

France, and it can be predicted that environmental features of the unparasitized area (Fig. 1) are unsuitable for the parasite life-cycle. Second, the 'genomic background hypothesis' states that the spatial distribution of *T. polycolpus* is driven by individual host characteristics, and it can be predicted that the genomic background of hosts from the unparasitized area is different from that of hosts from parasitized areas and this genomic background likely confers a resistance to *T. polycolpus*. To test these two non-mutually exclusive hypotheses, we combined species distribution models (SDMs) and population genomic approaches. The first approach aimed to statistically test whether the environment was suitable for parasites in the unparasitized area. In contrast, the second approach aimed to test whether the genomic structure of host populations underlined the difference in the host infection status (parasitized versus unparasitized) among geographic areas, and whether some genomic regions were statistically related to the infection status of individual hosts.

Material and methods

Biological models

Tracheliastes polycolpus is a crustacean copepod in which females are ectoparasites of freshwater fish (males are dwarf and free-living, Piasecki 1989). Females anchor to the fins of fish where they are fecundated by a single male and rapidly develop two egg sacs carrying on average one hundred eggs each (Loot et al. 2011). After hatching, free-living copepodids are released into the water; this represents the infective stage of the parasite (Mazé-Guilmo, 2016). Fin damage and inflammation are directly caused by the feeding activities of this parasite, which indirectly affects host fitness (Blanchet et al. 2009b).

In French watersheds, *Leuciscus burdigalensis* and *Leuciscus leuciscus* (rostrum dace and common dace, respectively) are the main host species of *T. polycolpus*. *Leuciscus leuciscus* inhabits rivers from the northern and eastern parts of France and extends its range far east in Eurasia, whereas *L. burdigalensis* is endemic to rivers of southwestern France (Fig. 1). Previous studies that were based on microsatellite and mitochondrial markers demonstrated that these two species are structured into five lineages in France (i.e. three and two lineages for *L. burdigalensis* and *L. leuciscus* respectively, Costedoat et al. 2014), which are geographically delineated by French river basins (Fig. 1). The unparasitized area encompasses three river basins (Rhin, Seine and Meuse, Fig. 1, 2) that are inhabited by a single lineage of *L. leuciscus* (hereafter referred to as 'lineage III-unparasitized', Fig. 1), and this lineage also inhabits the Normandy River Basin where the parasite is present (hereafter referred to as 'lineage III-parasitized', Fig. 1). Despite

strong genetic differentiation, the two *Leuciscus* species have very similar ecological requirements. They both live in running water with moderate water temperature (from 14°C to 24°C in warmer months), and they are gregarious species which usually feed on benthic invertebrates. *Tracheliastes polycolpus* also has several alternative host species including *Parachondrostoma toxostoma*, *Gobio* sp., *Phoxinus phoxinus*, *Rutilus rutilus* and *Squalius cephalus* (Lootvoet et al. 2013). Average prevalence on alternative hosts are lower than that on the main host species, as it generally ranges from 1% to 10% in alternative host species and from 10% to 90% in the main hosts species (Lootvoet et al. 2013).

Testing the environmental suitability hypothesis

We aimed to test whether environmental conditions could limit *T. polycolpus* distribution in the host lineage III-unparasitized. We built a SDM at the French spatial extent linking environmental features to parasite prevalence estimated at the sampling site level. Prevalence here refers to the number of parasitized hosts over the total number of hosts sampled in a given site. We used the SDM to test which environmental variables affect the distribution of *T. polycolpus* in France, and then used it to predict parasite prevalence as expected based on the actual environmental features. If the environment suitability hypothesis is verified, we would expect the SDM to predict low parasite prevalence in the

lineage III-unparasitized, which would indicate that the environment is unsuitable for the parasite in this part of France.

Field data
158 sites were sampled in French rivers from 2009 to 2011 (Fig. 1). Sites were sampled using electrofishing by the the Agence Française de la Biodiversité (AFB), the Fédérations Départementales de Pêche et de Protection des Milieux Aquatiques (FDPPMA) or directly by our own team. Sampling was always carried out during summer to avoid seasonal effects on parasite prevalence, and the sampling effort was kept as constant as possible. At each site, the number of dace sampled was recorded, their individual body length (mm) was measured, and their individual infection status (parasitized by *T. polycolpus* or unparasitized) was determined. Data on infection status was used to calculate parasite prevalence at each site (number of parasitized dace divided by the total number of dace sampled). A total of 1935 dace were sampled (12 dace per site in average), among which 908 *L. burdigalensis* (468 were parasitized) and 1027 *L. leuciscus* (257 were parasitized). In most sites, a piece of pelvic fin of each dace (5 mm²) was sampled and preserved in 95% alcohol.

Environmental data
We focused on thirteen abiotic and biotic variables that may explain the spatial distribution of *T. polycolpus*. We focused on biotic variables measured at the individual host level (but

Figure 2. Map showing the distribution of sampled sites used in the population genomic approach. Grey lines delimitate the main river basins and their names are indicated in white rectangles for relevant ones. The hatched area refers to area where hosts are consistently unparasitized.

averaged at the site level, see hereafter), and both abiotic and biotic variables measured at the site level (see Table 1 for a detailed list).

At the individual level, we considered host body length (mm), expected heterozygosity (He) computed from a set of microsatellite data (Mazé-Guilmo 2016) and identity of the main host species of *T. polycolpus* on which we recorded infections (i.e. either *L. burdigalensis* or *L. leuciscus*, Table 1). Each of these variables was averaged at the site level before being included in the SDM.

At the site level, we first extracted biotic variables related to the fish community composition using the AFB database (Poulet et al. 2011, Blanchet et al. 2014). This database contains fish occurrence data from yearly fish sampling in many sites evenly distributed across France. For each sampling site we drew a list of fish species occurrence over the period 2009–2011, from which we summed the total number of fish species ('fish species richness') and the total number of host species for *T. polycolpus* ('number of host species', Table 1) defined according to Lootvoet et al. (2013).

We then extracted abiotic variables belonging to four main categories namely climatic variables, hydrographic variables, landscape variables and habitat fragmentation variables

(Table 1). Water temperature was not available across all sampling sites and was thus modeled based on air temperature variables. More specifically, we predicted the mean annual water temperature based on a combination of the mean annual air temperature, and the mean monthly air temperatures for January and July extracted for each site from the SAFRAN database (LeMoine 2002) using a random forest model (see Chevalier et al. 2018 for more details). Hydrographic variables were extracted from the 'Réseau Hydrographique Théorique' database (RHT, Pella et al. 2012) at each site: altitude (m), river slope (%), riverine distance of each site from its river source (km), area of the upstream watershed (km²), Strahler index (i.e. the hierarchical level in the hydrographic network of each river), river width (m), river depth (m), average size of bed sediment (mm), minimum average monthly river flow ($m^3 s^{-1}$) and average natural river flow ($m^3 s^{-1}$). Landscape variables were extracted from the Corine land cover 2012 database (<www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>) as percentages of main land cover types within a 10 km radius around each site using the first level of land cover description (agricultural land, artificialized lands, forest and semi-natural lands and wetlands grouped with bodies of water). We then synthetized hydrographic

Table 1. Summary of the variables used in the species distribution model developed to identify the underlying determinants of *T. polycolpus* distribution in France. For each variable the organization level at which the variable has been measured is indicated along with the name of the variable and the associated biological hypotheses.

Level	Explicative variables	Hypotheses
Individual level	Expected heterozygosity	Global level of individual heterozygosity has been shown to impact parasite resistance in many host species, including dace (Blanchet et al. 2009a, Cardon et al. 2011)
	Dace body length	Larger hosts are expected to harbor more parasites because they are older (and less resistant) and because their bodies have a larger surface area (Cardon et al. 2011)
	Species identity	We expect that the two main hosts species (i.e. *L. leuciscus* and *L. burdigalensis*) on which we measured prevalence may differ in their intrinsic ability to resist the parasite
Site level (biotic)	Number of hosts species	*T. polycolpus* is a generalist species (i.e. it can be found at lower prevalence on other cyprinid species than *Leuciscus* sp., Lootvoet et al. 2013) and may be affected by the dilution effect: the higher the host species richness, the lowest the prevalence (Johnson et al. 2008)
	Fish species richness	'Fish species richness' refers to the same hypothesis than above (the dilution effect) except that it focuses on all species of the fish community since it can include potential host species that are fully resistant (Roche et al. 2012)
Site level (abiotic)	Climatic variables	
	Water temperature (modeled from air temperatures)	Temperature affects several aspects of the parasite life-cycle (Cardon et al. 2011, Mazé-Guilmo et al. 2016) and may affect host resistance
	Hydrographic variables	
	Hydrology (PC1)	We expect that the weakly mobile infective larvae may be favored in waters with low velocities/discharges
	Topology (PC2)	Different topologic characteristics (slope, altitude) could define habitats that are more or less suitable to the parasite: for instance, sites at higher altitude may be less suitable
	Landscape variables	
	Landscape 1 and Landscape 2 (PC1 and PC2)	Landscape composition around the site describe a particular level of anthropogenic pressures that can affect the success of non-native species
	Fragmentation variables	
	Dam density	Dams affect water velocity/discharge and may influence the survival of the infective larva
	Weirs density	In the same way, weirs affect water velocity/discharge and may also influence the survival of the infective larvae
	Weirs position	Weirs position relates to the same hypothesis than above

and landscape variables using principal component analyses (one PCA per category) in order to avoid both collinearity and over-parametrization in subsequent linear models (Supplementary material Appendix 1 Fig. A1, Table A1). For the hydrographic variables, we conserved the first two axes (58% and 15% of the variance respectively, Supplementary material Appendix 1 Fig. A1a, Table A1) that we named 'Hydrology' and 'Topology' respectively. For landscape variables we conserved the first two axes (50% and 28% of the variance respectively, Supplementary material Appendix 1 Fig. A1b, Table A1) that we arbitrarily named 'Landscape 1' and 'Landscape 2'. Finally, we extracted fragmentation variables from the 'Référentiel des Obstacles à l'Ecoulement' database (ROE, <www.onema.fr/REFERENTIEL-DES-OBSTACLES-A-L>) to quantify habitat fragmentation by weirs (number of obstacles being < 3 m height) or dams (number of obstacles being > 3 m height) within a 10 km radius around each sampling site. We input whether a weir was present directly upstream, downstream or both upstream and downstream for each sampling site (within a 2 km radius) to generate a more direct estimate of fragmentation at the sampling site level (categorical variable: upstream, downstream, both upstream and downstream or no weir at all). All abiotic characteristics were extracted using a Geographic Information System (QGIS Development Team 2008) and PCAs were run using the R package 'ade4' (Dray et al. 2007).

Statistical analyses

We developed a SDM to explain *T. polycolpus* prevalence at the site level by using a model accounting for spatial dependence among sites as we detected a strong signal of spatial autocorrelation for *T. polycolpus* prevalence (Moran test, p-value < 0.001). We used the R package 'spaMM' (Rousset and Ferdy 2014) and its function 'corrHLfit' to specify geographic coordinates of each sampling site into continuous random terms.

We first built a full model fitted on the dataset covering the current range of *T. polycolpus* (i.e. parasitized areas, Fig. 1, n = 128 sites) to test the significance of each variable on parasite prevalence. It included all explicative variables (Table 1) as well as the quadratic term for body length because preliminary visual inspections suggested a non-linear relationship between prevalence and body length. A binomial error terms distribution with a logit link function was fitted to the prevalence data. Then, likelihood ratio tests based on PQL/L fits were performed by contrasting the full model to a model excluding the explicative variable of interest, so as to generate a χ^2 statistic and its associated p-value for each explicative variable.

Secondly, we parametrized the model on a restricted dataset including 75% of the sampling sites covering the current range of *T. polycolpus* (i.e. parasitized areas, Fig. 1). The parameterized model allowed predicting *T. polycolpus* prevalence both on the remaining 25% of the dataset covering the current range of *T. polycolpus* and in sites out of the range limit of *T. polycolpus* (Lineage III-unparasitized, Fig. 1). This procedure was reiterated 100 times to account for the uncertainty arising from the

selection of the calibration dataset. The predictive power of the model was assessed at the host lineage level and at the site level within each host lineage using the mean absolute error (MAE) computed using the R package 'caret' (Kuhn 2012). We finally compared predicted prevalence between lineages using contrast tests using the R package 'lsmeans' and a binomial generalized linear mixed model (Lenth 2016).

Testing the role of the genetic background

We aimed to test whether genomic background of hosts could explain the individual infection status of *T. polycolpus*, and hence its range limitation in the lineage III-unparasitized (Fig. 1). In this section, models were set at the individual level, with the response variable being whether or not a given individual host is parasitized. We investigated whether some genomic structure underlined the absence of *T. polycolpus* in individuals from the lineage III-unparasitized. Secondly, we investigated genomic variation associated with *T. polycolpus* occurrence at the individual level.

Data collection and sequencing

DNA was extracted from the piece of pelvic fin as described in Aljanabi and Martinez (1997), and collected on 96 individual hosts belonging to the species *L. leuciscus* so as to reflect the variability of parasite prevalence and the different lineages (lineage II, lineage III-parasitized and lineage III-unparasitized). Overall, 28 individuals from the lineage II were sampled at four sampling sites belonging to the Rhone River Basin (Fig. 2) and we selected half of them with parasites and the other half without. We then sampled 28 individuals from the lineage III-parasitized at five sampling sites belonging to the Normandy River Basin (Fig. 2), and we selected half of them with parasites and the other half without. Finally, 40 individuals from to the lineage III-unparasitized were sampled at five sampling sites localized in the Rhin, Seine and Meuse Rivers Basins (Fig. 2).

We used a paired-end restriction site-associated DNA sequencing approach (RAD-seq, Baird et al. 2008) to sequence a reduced representation of the host genome into reads of 145 bp. Briefly, DNA quality was tested using a spectrophotometer (Nanodrop ND8000), quantified using a fluorimeter and standardized to 200 ng. DNA was sheared through sonication, digested using the *sbf1* enzyme for one hour at 37°C and then inactivated at 65°C. Adapters were ligated and samples were individually barcoded. Sequences were amplified through ten PCR cycles using the kit KAPA HiFi HotStart ReadyMix. Libraries were paired-end sequenced on one lane of an Illumina HISeq3000 containing 96 independent banks (one per individual) and using the chemistry V4 from kits TruSeq (2 × 125 pb). Libraries were finally quality checked using BioAnalyzer. Library preparation and sequencing were performed at the GeT-PlaGe core facility (Toulouse, France).

Data filtering and catalogue construction

We relied on the 'Stacks' pipeline (ver. 2.3.4) to both filter the raw dataset and identify SNPs genotyped at the individual

level (Catchen et al. 2013). First, we removed reads with unidentified bases and/or with low quality scores (below a phred33 quality score of 20) using 'process_radtags'. PCR duplicates were also removed using 'clone_filter'. To build our catalogue of loci, alleles and ultimately SNPs from the filtered reads, we ran the 'denovo_map.pl' using as parameters value: $m = 3$, $M = 2$, n = 4 (i.e. set of parameters maximizing the number of SNPs having high likelihood according to preliminary tests). Finally, 'populations' was used to select loci with a minimum allele frequency above 3%, with a likelihood over −20, with a minimum coverage depth greater than 3, occurring in the three populations and in at least 40% of the individuals for each population. We also kept only one SNP per locus to limit linkage disequilibrium between SNPs. When applied to the whole database (i.e. 95 individuals because one individual was discarded due to its poor sequencing quality), this parameter set resulted in a catalog of 14 255 SNPs. A final filtering excluded SNPs with more than 30% of missing values from further analyzes using the library 'poppr' in R (Kamvar et al. 2014). This resulted in a final dataset of 2543 SNPs genotyped from 95 individuals.

Population genetic structure

A discriminant analysis of principal component (DAPC) was conducted using 'adegenet' (Jombart 2008) in R to first identify a potential genetic structure among populations using the 'find.cluster' function. All principal components were kept, and the best number of clusters was selected using the Bayesian information criterion (BIC). A second DAPC was run using 30 principal components (Supplementary material Appendix 1 Fig. A2) and the k number of clusters selected previously. The $k − 1$ discriminant functions (i.e. individuals' scores extracted along the $k − 1$ discriminant axis) were then extracted. We finally constructed a generalized linear model (GLM) with a quasibinomial error term linking parasite prevalence at the individual level to the $k − 1$ discriminant functions of the DAPC to test their significance. In this model, we also controlled for the host body length and the individual level of genomic diversity (measured as expected heterozygosity, He, calculated from allelic frequencies over all 2543 SNPs using the 'adegenet' R package). These later variables were included as continuous explicative variables (together with their respective quadratic terms) since they have been shown to be (sometimes non-linearly) related to individual parasite prevalence in this host–parasite system (Blanchet et al. 2009a).

Genome wide association study

We used a genome wide association study (GWAS) approach to test whether prevalence at the individual level was significantly related to particular SNPs. We relied on GLMs with a quasibinomial error term linking parasite prevalence at the individual level (parasitized or unparasitized, response variable) to the genotype (coded as a categorical factor: homozygote for allele a or b and heterozygote ab) of each SNP. One model per SNP marker (i.e. 2543) was run in which we controlled for population structure using the $k − 1$ discriminant

functions described previously. We also included the host body length and He as continuous co-variables with their respective quadratic terms. Type I ANOVAs were performed to extract the p-value associated to each SNP marker. p-values were then corrected for multiple comparisons using a false discovery rate approach (FDR) and using the 'fdrtool' R package (Strimmer 2008).

The R software (ver. 3.5.3, <www.r-project.org>) was used to run statistical analyses when no other software is mentioned.

Results

Testing the environmental suitability hypothesis

In the parasitized area, *Tracheliastes polycolpus* prevalence was significantly related to host body length, host species identity, temperature, hydrology, the first synthetic variable of landscape composition and the position of weirs around the sampling site (Table 2). These relationships suggested that *T. polycolpus* prevalence was higher in *L. burdigalensis* than in *L. leuciscus* and that *T. polycolpus* prevalence was higher for hosts with an intermediate body length than for hosts with a smaller or larger body length (Supplementary material Appendix 1 Table A2). Furthermore, *T. polycolpus* prevalence was higher in cold sites localized in small rivers and preferentially associated with agricultural lands rather than warm sites localized in large rivers and associated with forest or semi-natural lands (Supplementary material Appendix 1 Table A2, Fig. A1). Finally, fragmentation by weirs had a significant impact on *T. polycolpus* prevalence, with higher prevalence expected in sites in which a weir was present downstream (compared to other positions of weirs or no weir at all, Table 2 and Supplementary material Appendix 1 Table A2).

Computation of model MAE indicated an overall good predictive power of the model, at the lineage level

Table 2. Likelihood ratio tests [χ^2, degrees of freedom (df) and p-values] testing the significance of each of the explicative variables included in the SDM that was developed to identify the underlying determinants of *T. polycolpus* distribution in France. Significant variables are indicated in bold.

Variable	χ^2_LR	df	p_value
Expected heterozygosity	0.031	1	0.860
Dace body length	**11.380**	**1**	**0.001**
(Dace body length)^2	**5.563**	**1**	**0.018**
Species identity	**13.330**	**1**	**< 0.001**
Number of hosts species	0.539	1	0.463
Fish species richness	1.031	1	0.310
Water temperature	**5.554**	**1**	**0.018**
Hydrology	**4.591**	**1**	**0.032**
Topology	1.621	1	0.203
Landscape 1	**9.619**	**1**	**0.002**
Landscape 2	1.820	1	0.177
Dams density	0.474	1	0.491
Weirs density	0.413	1	0.520
Weirs position	**8.849**	**3**	**0.031**

(MAE = 0.151, Fig. 3) and at the site level within most host lineages (ranging from 0.258 in lineage II to 0.367 in lineage VI). In the five host lineages localized in the current range of *T. polycolpus*, prevalences predicted from our model were very close to observed prevalences (except in the lineage VI, Fig. 3). However, in the unparasitized area of the lineage III (i.e. lineage III-unparasitized, Fig. 3), the predicted parasite prevalence was high (0.36) and similar to the prevalence values observed and predicted in areas and lineages in which *T. polycolpus* is currently present (i.e. *L. burdigalensis*, lineage II, lineage III-parasitized, Fig. 3). Specifically, in lineage III-unparasitized, the predicted parasite prevalence was not significantly different neither from the prevalence predicted in the lineage III-parasitized (Contrast test, p = 0.220, Supplementary material Appendix 1 Fig. A3) nor from the prevalence predicted in the lineage II (Contrast test, p = 0.102, Supplementary material Appendix 1 Fig. A3). This indicated that the environment was not unsuitable for *T. polycolpus* in the lineage III-unparasitized.

Testing the role of the genetic background

The clustering analysis based on 2543 SNPs revealed that the best clustering was achieved for *k* = 3 (Supplementary material

Appendix 1 Fig. A4). The two first axes of the DAPC performed for *k* = 3 and based on 30 PCA components summarized 67.3% of the total variance (Fig. 4). The first axis discriminated the two dace lineages (i.e. lineage II and lineage III), as well as individuals of the lineage III belonging to the parasitized area from individuals of the same lineage but belonging to the unparasitized area (Fig. 4 and Supplementary material Appendix 1 Fig. A5). Interestingly, the second axis discriminated individuals localized in the parasitized areas from individuals localized in the unparasitized area (lineages II and lineage III-parasitized versus lineage III-unparasitized, Fig. 4). This suggests that the genomic background of dace localized in the unparasitized area is different from the genomic background of dace belonging to the lineage II, but also from dace belonging to the lineage III but localized in the parasitized area. Furthermore, we found that both the first and second discriminant functions had a significant effect on individual infection status (F = 9.750; df = 88,1; p = 0.002 and F = 77.830; df = 88,1; p < 0.001 respectively).

Finally, when controlling for population structure, the GWAS revealed 90 SNPs significantly associated to *T. polycolpus* prevalence measured at the individual level for *L. leuciscus* (at a FDR threshold of 5%, Supplementary material Appendix 1 Table A3).
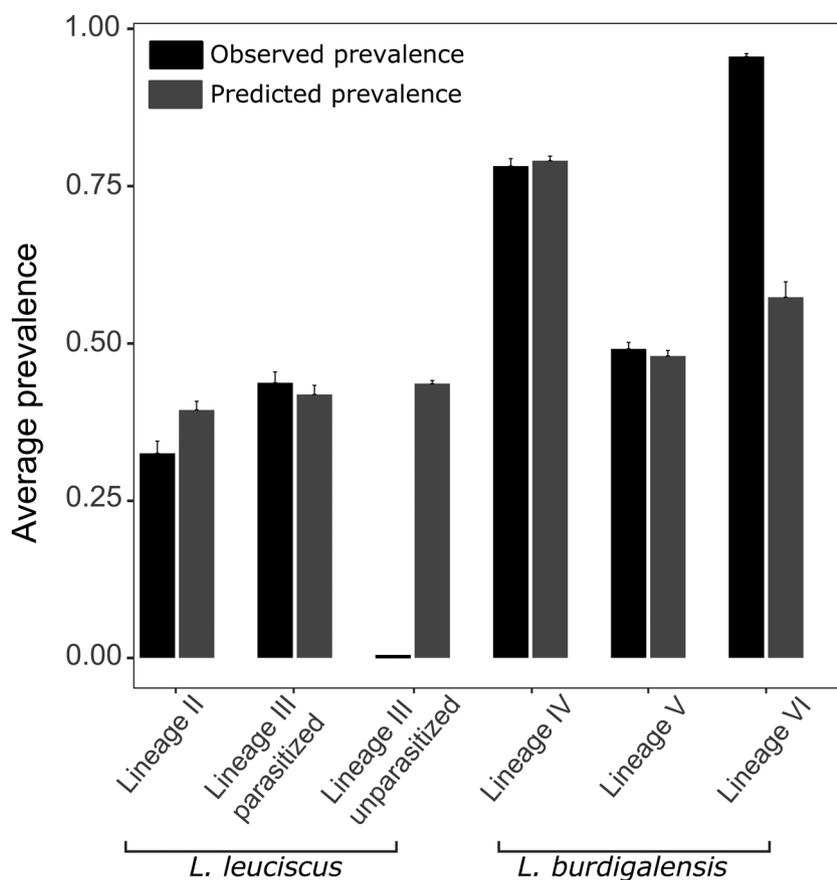
Figure 3. Barplot showing the average prevalence for each dace lineage. Black bars correspond to data measured in the wild. Grey bars correspond to values predicted by the SDM model (averaged over 100 iterations). Error-bars are standard deviations from the mean.
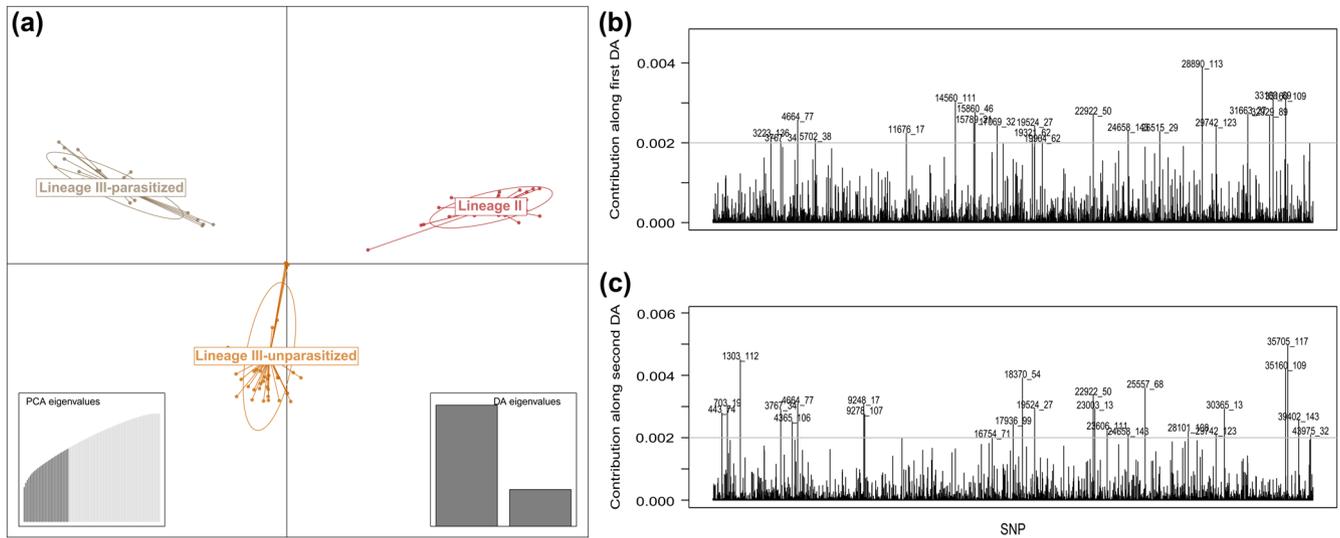
Figure 4. (a) Scatter plot from a DAPC based on genomic data (2543 SNP markers) showing population differentiation between the three genomic clusters previously inferred. Parasitized and unparasitized refer to the area (Fig. 1) not individual fish condition. The 30 first axes of a PCA were retained and two axes were retained for the DAPC as eigen-values plots show. (b–c) Contribution of each SNP to the multivariate DAPC analysis. These plots show the extent to which each allele is driving genomic population differentiation along the first (b) and (c) second axis of the DAPC. Each vertical line corresponds to the contribution of a unique SNP. SNP whose value is above the grey horizontal line are SNP contributing to more than 2‰ to the genomic differentiation between clusters. The SNPs are identified as follows: loci number_position of the SNP.

Discussion

We identified factors limiting the spatial distribution of an emerging parasite in its invasive range. Combining a species distribution model and population genomics approaches we suggest that host characteristics, such as the genomic background of individuals living in the unparasitized area, are more likely than environmental features to explain the current spatial distribution of *Tracheliastes polycolpus*, and hence why it has not invaded some areas.

We first highlighted a complex role of environmental factors involving multifactorial drivers of *T. polycolpus* spatial distribution. Climate is one of the most studied drivers in parasite distribution (Kutz et al. 2005, Barrett et al. 2012, Caminade et al. 2019). We consistently found a significant effect of water temperature. In particular, cold sites were associated with high prevalence, which matches the biology of this parasite species and also the host habitat preferences (Mazé-Guilmo et al. 2016, Franke et al. 2019). However, we also found significant effects of various environmental drivers such as host related factors (host body length and host species), the river topography, the surrounding landscape composition and the anthropogenic fragmentation. Overall, these later findings corroborate previous findings on this species. For instance, as in Blanchet et al. (2009a), we found that parasite attachment is favored in hosts with intermediate body length, probably because of a trade-off between the surface available to the parasite to anchor and age-related ability of hosts to resist/tolerate the parasite (Cardon et al. 2011). Similarly, we found that sites localized in small rivers with low stream velocity (negative effect of the 'Hydrology'

variable, Supplementary material Appendix 1 Fig. A1) and with a weir downstream were associated to high prevalence. These effects were expected as the free-living infective stage of this parasite requires low water velocity for development (Loot et al. 2004). More generally, such a multifactorial environmental determinism of *T. polycolpus* distribution strongly supports the growing view that integrative distribution models accounting for multiple environmental drivers, including biological interactions, are particularly powerful for predicting the spatial distribution of parasites (Lafferty 2009, Giannini et al. 2013, Marcogliese 2016).

Based on predictions from this species distribution model, we further showed that environmental suitability was unlikely to explain the range limitation of *T. polycolpus*. Prevalence levels predicted from environmental conditions characterizing the area in which the parasite is currently absent were similar to those characterizing areas in which the parasite is recurrently observed. This suggests that these conditions are suitable for *T. polycolpus*. One could argue that in unparasitized French river basins we may have missed some key environmental variables better predicting the occurrence of the parasite. Nonetheless, our species distribution model showed very good predictive performances for all other host lineages highlighting a very good predictive power. It is noteworthy that predicted parasite prevalences were underestimated for the host lineage VI. Interestingly, the lineage VI is also the most phylogenetically differentiated lineage of *L. burdigalensis* and is actually considered as a separate endemic species by some authors (*L. bearnensis*, Kottelat and Freyhof 2007 but see Costedoat et al. 2014). Discrepancy between predicted and observed prevalence in this phylogenetically divergent

lineage as well as in the host lineage III-unparasitized suggest that other factors than environmental suitability, notably host genomic background, could determine patterns of *T. polycolpus* prevalence, and consequently its range limitation.

Using a population genomic approach based on 2543 SNPs, we showed that host genomic background is associated with *T. polycolpus* distribution, both at the individual and population levels. We first revealed a marked host population structure in *L. leuciscus* that spatially matches the distribution of *T. polycolpus*. We showed that the lineage III proposed by Costedoat et al. (2014) is actually separated into a parasitized 'sub-lineage' localized in the Normandy River Basins (Fig. 2) and unparasitized 'sub-lineage' in the Rhin-Seine-Meuse River Basins. Previous studies on the genetic structure of the *Leuciscus* species complex were based on microsatellite, allozyme and mitochondrial markers, and covered a more extended spatial extent but with fewer sampled individuals per location (Costedoat et al. 2006, 2014). Here, the finer genetic structure that we detected probably results from the higher resolution of the SNP panel we used, as well as from the more intensive sampling of the *Leuciscus* species within lineages. For technical reasons, we focused on a subset of *T. polycolpus* host distribution in France (Fig. 1, 2). We can thus speculate that within-lineage genomic structure of host populations could be a more general mechanism underlying patterns of *T. polycolpus* prevalence across France (e.g. within the lineage II, where southern sites sampled also display null prevalence, Fig. 1). We further found that 90 SNPs were associated with individual infection status, even when controlling for population structure. Both the genetic uniqueness of host populations localized in the unparasitized area and the association of SNPs with individual infection status suggest that *T. polycolpus* range limitation is due to local host resistance, hence providing a strong support to the 'host genomic background hypothesis'.

The emergence of resistance against *T. polycolpus* in northeastern river basins is an intriguing question and two main hypotheses can be formulated: one related to macro-evolutionary processes and the other one related to recent and rapid adaptation of hosts to this invasive parasite. The 'macro-evolutionary' hypothesis assumes that resistance to *T. polycolpus* emerged during the differentiation of *Leuciscus* lineages over time. Costedoat et al. (2006) dated the differentiation of the lineage III back to the Pleistocene (~500 000 yr ago), with the colonization of river basins of northern France from eastern Europe. The differentiation between the lineage III-parasitized and lineage III-unparasitized (and hence the emergence of resistance) may have consequently occurred during late Pleistocene, probably after the former Channel River had retracted, hence causing a loss of connectivity between Normandy River Basins and Rhin-Seine-Meuse River Basins (which occurred ~10–12 000 yr ago, Antoine et al. 2007, Hugueny et al. 2011, Dias et al. 2014). Alternatively, the recent and rapid adaptation scenario assumes rapid adaptive changes in hosts along with the introduction of *T. polycolpus* in France during the 1920s (Rey et al. 2015). Some examples of rapid evolution of resistance to disease in wild host populations have been documented (Duffy and Sivars-Becker 2007, Bonneaud et al. 2011, Epstein et al. 2016), but these examples remain rare. In the *L. leuciscus*–*T. polycolpus* system, adaptive evolution toward resistance cannot be excluded since we found significant SNP-infection status associations, but whether emergence of resistance is due to drift or to adaptive evolution, and whether this occurred far in the past or very recently are questions that would need to be fully tested.

Overall, our results demonstrated the importance of combining approaches for predicting the distribution range of co-evolving and interacting species. Species distribution modelling approaches focusing on parasite species have, so far, mainly focused on environmental factors (Ostfeld et al. 2005, Caminade et al. 2019). However, our results suggest that predictions of future parasite distribution (such as northern range expansion as often predicted in the context of climate change, Altizer et al. 2013, Carter 2018), when only based on environmental drivers should be carefully interpreted. Here, we indeed provided an empirical example in which the presence of a parasite would have been expected in an area where the parasite actually never achieved to infect local hosts. This suggests that ecological features alone are not sufficient to explain the range and potential spread of emerging parasites (Pérez-Rodríguez et al. 2013, Anderson 2017). Furthermore, the genomic-based approach allowed demonstrating that the host genomic background was likely a critical factor for explaining *T. polycolpus* distribution both at the population and the individual levels. Population genomic approaches allow accounting for evolutionary processes leading to parasite resistance such as local adaptation, which, from our results, appears to be a crucial process to take into account in order to improve predictions regarding the potential spread of emerging pathogens. To conclude, our study illustrates the usefulness of integrative approaches to reveal the determinants of emerging pathogens at large spatial scales as well as to predict future spread.

Data availability statement

Data available from Figshare Digital Repository: <https://doi.org/10.6084/m9.figshare.13142528> (Mathieu-Bégné et al. 2020).

References

Aguilar, J. R. et al. 2016. MHC-I provides both quantitative resistance and susceptibility to blood parasites in blue tits in the wild. – J. Avian Biol. 47: 669–677.

Aljanabi, S. M. and Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. – Nucleic Acids Res. 25: 4692–4693.

Altizer, S. et al. 2013. Climate change and infectious diseases: from evidence to a predictive framework. – Science 341: 514–519.

Anderson, R. P. 2017. When and how should biotic interactions be considered in models of species niches and distributions? – J. Biogeogr. 44: 8–17.

Antoine, P. et al. 2007. Pleistocene fluvial terraces from northern France (Seine, Yonne, Somme): synthesis, and new results from interglacial deposits. – Quater. Sci. Rev. 26: 2701–2723.

Baird, N. A. et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. – PLoS One 3: e3376.

Bakke, T. A. and Harris, P. D. 1998. Diseases and parasites in wild Atlantic salmon (*Salmo salar*) populations. – Can. J. Fish. Aquat. Sci. 55: 247–266.

Barrett, T. M. et al. 2012. Host–parasite distributions under changing climate: *Tsuga heterophylla* and *Arceuthobium tsugense* in Alaska. – Can. J. For. Res. 42: 642–656.

Barrett, M. A. et al. 2013. Climate change, predictive modeling and lemur health: assessing impacts of changing climate on health and conservation in Madagascar. – Biol. Conserv. 157: 409–422.

Benavides, M. V. et al. 2016. Genomic regions associated with sheep resistance to gastrointestinal nematodes. – Trends Parasitol. 32: 470–480.

Bernatchez, L. and Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? MHC studies in nonmodel vertebrates. – J. Evol. Biol. 16: 363–377.

Blanchet, S. et al. 2009a. Evidence of parasite-mediated disruptive selection on genetic diversity in a wild fish population. – Mol. Ecol. 18: 1112–1123.

Blanchet, S. et al. 2009b. Why do parasitized hosts look different? Resolving the 'chicken-egg' dilemma. – Oecologia 160: 37–47.

Blanchet, S. et al. 2014. Regional vs local drivers of phylogenetic and species diversity in stream fish communities. – Freshwater Biol. 59: 450–462.

Blasco-Costa, I. et al. 2013. Upstream–downstream gradient in infection levels by fish parasites: a common river pattern? – Parasitology 140: 266–274.

Blasco-Costa, I. et al. 2015. Biogeography of parasitism in freshwater fish: spatial patterns in hot spots of infection. – Ecography 38: 301–310.

Bonneaud, C. et al. 2011. Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. – Proc. Natl Acad. Sci. USA 108: 7866–7871.

Bozick, B. A. and Real, L. A. 2015. Integrating parasites and pathogens into the study of geographic range limits. – Quart. Rev. Biol. 90: 361–380.

Byers, J. E. et al. 2019. What factors explain the geographical range of mammalian parasites? – Proc. R. Soc. B 286: 20190673.

Caminade, C. et al. 2019. Impact of recent and future climate change on vector-borne diseases: climate change and vector-borne diseases. – Ann. N. Y. Acad. Sci. 1436: 157–173.

Cardon, M. et al. 2011. Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal analysis: causalities in host–parasite interactions. – J. Anim. Ecol. 80: 657–667.

Carter, N. 2018. Climate change, disease range shifts and the future of the Africa lion. – Conserv. Biol. 32: 1207–1210.

Catchen, J. et al. 2013. Stacks: an analysis tool set for population genomics. – Mol. Ecol. 22: 3124–3140.

Charbonnel, N. et al. 2019. Differential immune gene expression associated with contemporary range expansion of two invasive rodents in Senegal. – bioRxiv ver5 peer-reviewed and recommended by PCI Ecology.

Chevalier, M. et al. 2018. Interactions between species attributes explain population dynamics in stream fishes under changing climate. – Ecosphere 9: e02061.

Costedoat, C. et al. 2006. Heterogeneous colonization pattern of European Cyprinids, as highlighted by the dace complex (Teleostei: Cyprinidae). – Mol. Phylogenet. Evol. 41: 127–148.

Costedoat, C. et al. 2014. Révision de la taxonomie ichtyologique en métropole: validation moléculaire du guide de détermination et investigation des aires de répartition. – Rap. Tech., ONEMA, France.

Dias, M. S. et al. 2014. Global imprint of historical connectivity on freshwater fish biodiversity. – Ecol. Lett. 17: 1130–1140.

Dray, S. et al. 2007. The ade4 package: implementing the duality diagram for ecologists. – J. Stat. Softw. 22: 1–20.

Duffy, M. A. and Sivars-Becker, L. 2007. Rapid evolution and ecological host–parasite dynamics: rapid evolution and disease dynamics. – Ecol. Lett. 10: 44–53.

Epstein, B. et al. 2016. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. – Nat. Comm. 7: 12684.

Escobar, L. E. et al. 2017. Inferring the ecological niche of toxoplasma gondii and *Bartonella* spp. in wild felids. – Front. Vet. Sci. 4: 172.

Fourcade, Y. et al. 2014. Continental-scale patterns of pathogen prevalence: a case study on the corncrake. – Evol. Appl. 7: 1043–1055.

Franke, F. et al. 2019. Consequences of divergent temperature optima in a host–parasite system. – Oikos 128: 869–880.

Gaston, K. J. 2009. Geographic range limits: achieving synthesis. – Proc. R. Soc. B 276: 1395–1406.

Gehman, A.-L. M. et al. 2018. Host and parasite thermal ecology jointly determine the effect of climate warming on epidemic dynamics. – Proc. Natl Acad. Sci. USA 115: 744–749.

Giannini, T. C. et al. 2013. Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants. – Ecography 36: 649–656.

Grosholz, E. D. 1994. The effects of host genotype and spatial distribution on trematode parasitism in a bivalve population. – Evolution 48: 1514–1524.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.

Hugueny, B. et al. 2011. Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach: extinction rates within freshwater fish communities. – Global Ecol. Biogeogr. 20: 449–463.

Johnson, P. T. J. et al. 2008. Diversity and disease: community structure drives parasite transmission and host fitness. – Ecol. Lett. 11: 1017–1026.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. – Bioinformatics 24: 1403–1405.

Kamvar, Z. N. et al. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal and/or sexual reproduction. – PeerJ 2: e281.

Kottelat, M. and Freyhof, J. 2007. Handbook of European freshwater fishes. – Publications Kottelat, Cornol, Switzerland and Berlin, Germany.

Kuhn, M. 2012. The caret package. – R Foundation for Statistical Computing, Vienna, Austria, <https://cran.r-project.org/package=caret>.

Kutz, S. J. et al. 2005. Global warming is changing the dynamics of Arctic host–parasite systems. – Proc. R. Soc. B 272: 2571–2576.

Lachish, S. et al. 2013. Spatial determinants of infection risk in a multi-species avian malaria system. – Ecography 36: 587–598.

Lafferty, K. D. 2009. The ecology of climate change and infectious diseases. – Ecology 90: 888–900.

Le Moine, P. 2002. Description de l'analyse des champs de surface sur la France par le système SAFRAN. – Rap. Tech., Météo-France CNRM/GMME/MC2, France.

Lenth, R. V. 2016. Least-squares means: the R package lsmeans. – J. Stat. Softw. 69: 1–33.

Loot, G. et al. 2004. The effects of the ectoparasite *Tracheliastes polycolpus* (Copepoda: Lernaeopodidae) on the fins of rostrum dace (*Leuciscus leuciscus burdigalensis*). – Par. Res. 94: 16–23.

Loot, G. et al. 2011. Determinants of life-history traits in a fish ectoparasite: a hierarchical analysis. – Parasitology 138: 848–857.

Lootvoet, A. et al. 2013. Patterns and processes of alternative host use in a generalist parasite: insights from a natural host–parasite interaction. – Funct. Ecol. 27: 1403–1414.

Marcogliese, D. J. 2001. Implications of climate change for parasitism of animals in the aquatic environment. – Can. J. Zool. 79: 1331–1352.

Marcogliese, D. J. 2016. The distribution and abundance of parasites in aquatic ecosystems in a changing climate: more than just temperature. – Integr. Comp. Biol. 56: 611–619.

Mathieu-Bégné, E. et al. 2020. Data from: Combining species distribution models and population genomics underlines the determinants of range limitation in an emerging parasite. – Figshare Digital Repository, <https://doi.org/10.6084/m9.figshare.13142528>.

Mazé-Guilmo, E. 2016. Etude du potentiel de colonisation des parasites: une approche intégrative. – Thèse de Doctorat de l'Université Paul Sabatier, France.

Mazé-Guilmo, E. et al. 2016. Local adaptation drives thermal tolerance among parasite populations: a common garden experiment. – Proc. R. Soc. B 283: 20160587.

Moor, H. et al. 2015. Predicting climate change effects on wetland ecosystem services using species distribution modeling and plant functional traits. – Ambio 44: 113–126.

Morand, S. et al. 1996. Parasite–host coevolution and geographic patterns of parasite infectivity and host susceptibility. – Proc. R. Soc. B 263: 119–128.

Morgan, E. R. et al. 2009. Canine pulmonary angiostrongylosis: the influence of climate on parasite distribution. – Parasitol. Int. 58: 406–410.

Mougeot, F. 2005. Elevated spring testosterone increases parasite intensity in male red grouse. – Behav. Ecol. 17: 117–125.

Narouei-Khandan, H. A. et al. 2016. Global climate suitability of citrus huanglongbing and its vector, the Asian citrus psyllid, using two correlative species distribution modeling approaches, with emphasis on the USA. – Eur. J. Plant Pathol. 144: 655–670.

Ostfeld, R. et al. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. – Trends Ecol. Evol. 20: 328–336.

Paull, S. H. et al. 2015. How temperature shifts affect parasite production: testing the roles of thermal stress and acclimation. – Funct. Ecol. 29: 941–950.

Pella, H. et al. 2012. Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. – Géomorphologie 18: 317–336.

Pérez-Rodríguez, A. et al. 2013. Finding the appropriate variables to model the distribution of vector-borne parasites with different environmental preferences: climate is not enough. – Global Change Biol. 19: 3245–3253.

Piasecki, W. 1989. Life cycle of *Tracheliastes maculatus* Kollar, 1835 (Copepoda, Siphonostomatoida, Lernaeopodidae). – Wiadomości Parazytologiczne.

Pickles, R. S. A. et al. 2013. Predicting shifts in parasite distribution with climate change: a multitrophic level approach. – Global Change Biol. 19: 2645–2654.

Poulet, N. et al. 2011. Time trends in fish populations in metropolitan France: insights from national monitoring data. – J. Fish Biol. 79: 1436–1452.

QGIS Development Team. 2008. QGIS Geographic Information System. – Open Source Geospatial Foundation Project, <http://qgis.org>.

Rey, O. et al. 2015. Elucidating the spatio-temporal dynamics of an emerging wildlife pathogen using approximate Bayesian computation. – Mol. Ecol. 24: 5348–5363.

Roche, B. et al. 2012. Linking community and disease ecology: the impact of biodiversity on pathogen transmission. – Phil. Trans. R. Soc. B 367: 2807–2813.

Rousset, F. and Ferdy, J.-B. 2014. Testing environmental and genetic effects in the presence of spatial autocorrelation. – Ecography 37: 781–790.

Schwabl, P. et al. 2017. Prediction and prevention of parasitic diseases using a landscape genomics framework. – Trends Parasitol. 33: 264–275.

Strimmer, K. 2008. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. – Bioinformatics 24: 1461–1462.

Vajana, E. et al. 2018. Combining landscape genomics and ecological modelling to investigate local adaptation of indigenous ugandan cattle to East Coast fever. – Front. Genet. 9: 385.

Wegner, K. M. et al. 2003. Parasite selection for immunogenetic optimality. – Science 301: 1343–1343.

Wenzel, M. A. et al. 2015. Genome-wide association and genome partitioning reveal novel genomic regions underlying variation in gastrointestinal nematode burden in a wild bird. – Mol. Ecol. 24: 4175–4192.

Wiens, J. J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology: niche conservatism, ecology and conservation. – Ecol. Lett. 13: 1310–1324.

Wolinska, J. and King, K. C. 2009. Environment can alter selection in host–parasite interactions. – Trends Parasitol. 25: 236–244.

Zhang, L. et al. 2015. Major histocompatibility complex alleles associated with parasite susceptibility in wild giant pandas. – Heredity 114: 85–93.

Zueva, K. J. et al. 2018. Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). – Mar. Genom. 39: 26–38.