A Journal of Macroecology

Global Ecology and Biogeography

RESEARCH PAPER

Phenotypic and phylogenetic correlates of geographic range size in Canadian freshwater fishes

Simon Blanchet[1,2]*, Yorick Reyjol[3], Julien April[4,5], Nicholas E. Mandrak[6], Marco A. Rodríguez[7], Louis Bernatchez[4] and Pierre Magnan[7]

[1]*Station d'Ecologie Expérimentale du CNRS à Moulis, U.S.R. 2936, route du CNRS, 09200 Moulis, France,* [2]*CNRS: Evolution et Diversité Biologique, U.M.R 5174, 118 route de Narbonne, F-31062 Toulouse Cedex 4, France,* [3]*ONEMA, The French National Office for Water and Aquatic Ecosystems, DAST, Square Nadar, 94300 Vincennes, France,* [4]*Institut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène Marchand, Université Laval, Québec, Québec, Canada G1V 0A6,* [5]*Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs, 880 Chemin Sainte-Foy, Québec, Québec, Canada G1S 4X4,* [6]*Great Lakes Laboratory for Fisheries and Aquatic Sciences, 867 Lakeshore Road, Burlington, Ontario, Canada 7R 4A6,* [7]*Centre de Recherche sur les interactions basins versants – écosystèmes aquatiques (RIVE), Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières, Québec, Canada G9A 5H7*

*Correspondence: Simon Blanchet, Station d'Ecologie Expérimentale du CNRS à Moulis, U.S.R. 2936, route du CNRS, 09200 Moulis, France.
E-mail: simon.blanchet@ecoex-moulis.cnrs.fr

ABSTRACT

Aim We investigated the relationship between geographic range size (GRS), longitude and latitude (Rapoport's rule) in Canadian freshwater fishes. We tested hypotheses regarding the phenotypic and phylogenetic determinants of GRS to unravel processes driving the spatial patterns of GRS in Canada. Because GRS is negatively correlated with extinction risk, we also aimed at identifying biological proxies that may be used to predict extinction risks.

Location North-America, Canadian Shield.

Methods We built a database combining range area, seven ecological traits, and a molecular phylogeny for native Canadian freshwater fishes. We tested latitudinal and longitudinal patterns in GRS by the mean of Pearson correlations. We combined phylogenetic generalized least squares (PGLS) models and a model selection procedure to tease apart hypotheses (and hence ecological traits) that best explained GRS in our dataset. PGLSs were also used to explore relationships between ecological traits, phylogeny, and species mid-range latitude and longitude. Partial regressions were used to determine direct and indirect relationships driving spatial patterns of GRS in Canadian freshwater fishes.

Results There was a significant and positive correlation between GRS, latitude and longitude. According to PGLSs, three ecological traits, related to the habitat use, migratory and thermal tolerance hypotheses, were significantly correlated with GRS, mid-range longitude and mid-range latitude. Two traits related to locomotion were further related to GRS. There was no phylogenetic effect on GRS (i.e. no phylogenetic conservatism). Partial regressions revealed complex direct and indirect relationships between ecological traits, mid-range latitude, mid-range longitude and GRS.

Main conclusions Our results show that traits related to the ability to use dispersal corridors, as well as traits directly related to mobility, are useful in understanding biodiversity patterns such as Rapoport's rule. However, because of a weak explanatory power, we conclude that using biological proxies of GRS to predict species at risk of extinction would be premature.

Keywords
Biodiversity, biogeography, ecophylogenetics, extinction risk, path analysis, Rapoport's rule.

INTRODUCTION

Species greatly vary in the size of their geographic range, with some being restricted to small areas, whereas others are cosmopolitan (Gaston, 2000, 2009). Understanding the processes driving interspecific variation in geographic range size (GRS) has both fundamental and conservation implications. For instance, because species displaying small GRSs are more prone to extinction (Gaston & Blackburn, 1996a), characterizing the phenotypic attributes that explain the inability of species to
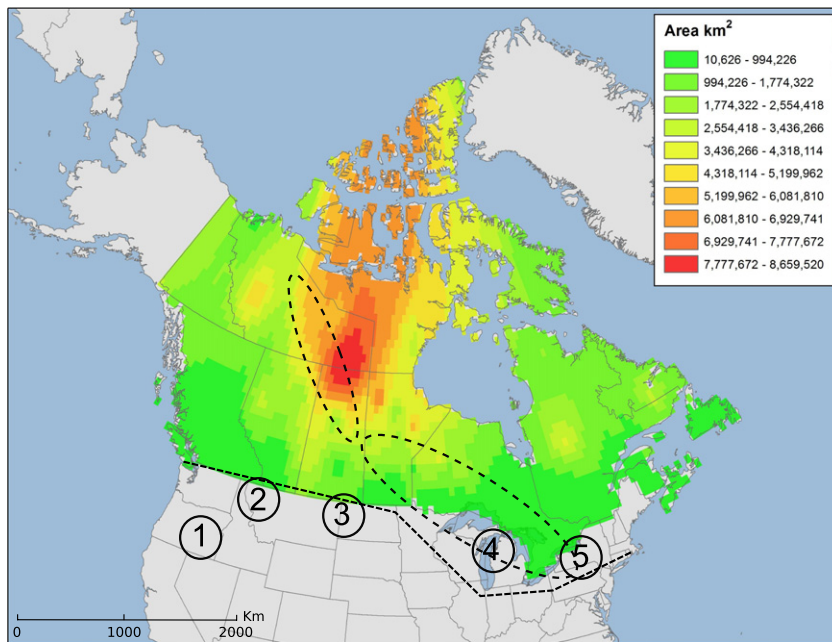
Figure 1 Map presenting the glacial conditions of North America during the Wisconsinan glaciation events (~15,000 yr BP) and the geographic pattern of mean range size of Canadian freshwater fishes. The oblique dotted lines indicate the maximum extent of Wisconsinan ice (the continent was covered by ice above the lines). The circled numbers indicate the main glacial refugia: 1. Columbian; 2. Banff-Jasper; 3. Missourian; 4. Mississippian; and 5. Atlantic. The dotted ovals represent the general location of the pro-glacial lakes that could have been used by fishes to re-colonize Canada. Simplified from Hocutt & Wiley (1986). Colours indicate the mean range size of Canadian freshwater fishes.

colonize wide geographic ranges is essential for predicting species extinction risk (Gaston, 2008a,b; Mace *et al.*, 2010).

From a spatial point of view, interspecific variation in GRS is often characterized by a north–south gradient, whereby species living at higher latitude generally display a larger GRS (Stevens, 1989; Gaston, 2003). Such a general pattern has been observed in several plant and animal species and is referred to as « Rapoport's rule » (Morin & Chuine, 2006; Gaston, 2009). Several hypotheses related to biogeographic, environmental and phenotypic factors have been proposed for understanding processes driving interspecific variation in GRS and its relationship with latitude (e.g. Morin & Chuine, 2006; Mouillot & Gaston, 2007; Calosi *et al.*, 2010; Garcia-Barros & Benito, 2010). For instance, GRS might be « heritable » along a phylogenetic tree, with closely related species sharing similar phenotypic attributes and, hence, a similar GRS (Jablonski, 1987; Hunt *et al.*, 2005; Roy *et al.*, 2009). Although many studies have focused on the relationships between GRS and phenotypic attributes such as body size (Taylor & Gotelli, 1994; Gaston & Blackburn, 1996a; Pyron, 1999; Rosenfield, 2002), no consensus has yet emerged regarding the phenotypic attributes explaining interspecific variation in GRS, and their implications with regard to the origin of Rapoport's rule.

Two potential difficulties can be identified when searching for phenotypic attributes influencing GRS. First, biogeographic history may override relationships between phenotypic attributes and GRS if several independent biogeographic areas are analysed simultaneously, or if several glacial refugia exist within a biogeographic area (Garcia-Barros & Benito, 2010). Indeed, all else being equal, GRS is potentially dependent upon (1) the duration since post-glacial colonization (the greater the duration since post-glacial colonization, the larger potential GRS) and (2) the number of post-glacial dispersal corridors

available from glacial refugia (the greater the number of dispersal corridors from a refugium, the further species will be able to disperse, resulting in larger geographic ranges). It is of prime importance to control for these potential biogeographic effects to disentangle the independent relationships between phenotypic attributes and GRS. Second, the evolutionary dynamics of taxa need to be accounted for since old species are expected to display a larger GRS simply because they have had more time to become widespread in multiple refugia (Roy *et al.*, 2009).

The freshwater fish fauna of Canada is an ideal model to explore relationships between phenotypic attributes of species and their GRS. During the last glaciation event, nearly all of Canada was covered by a permanent ice sheet (Fig. 1), which impeded survival of aquatic organisms, and was colonized by these organisms only within the last 15,000 years (Hocutt & Wiley, 1986; Griffiths, 2010). In addition, the glacial refugia for the Canadian freshwater fish fauna are relatively well established (Fig. 1, Hocutt & Wiley, 1986; Mandrak & Crossman, 1992). When the glaciers melted, a succession of pro-glacial lakes (Fig. 1) and their outlets allowed aquatic organisms to re-colonize this area, hence creating dispersal corridors (Griffiths, 2010). Although the availability of these corridors may vary from one refugium to another, analyzing separately the fauna of each refugium can control for such potential biases. Finally, because the colonization of this area is recent, one can control the problem related to the age of species, as speciation is usually a longer term process (> 15,000 years, Avise *et al.*, 1998). Therefore, the recent colonization history of Canadian watersheds represents a great opportunity to investigate biological attributes favouring colonization at such a large spatial scale, especially for fishes.

The goals of this study are to identify the phenotypic and phylogenetic determinants of GRS in Canadian freshwater

fishes, and to unravel processes driving the relationship between interspecific variation in GRS, longitude and latitude (Rapoport's rule) in this area. To do so, we built a database combining range area, ecological traits, and a complete molecular phylogeny for nearly all native fish species occurring in Canadian watersheds. We tested whether Canadian freshwater fishes displayed a spatial pattern in GRS, and notably a pattern corroborating Rapoport's rule. We then used a model selection procedure to test a series of non-exclusive hypotheses regarding the relationships between phenotypic attributes and range size variation in Canadian freshwater fishes. More specifically, the following hypotheses were tested: (1) the 'life-history traits hypothesis' predicting that certain life-history characteristics (e.g. high fecundity, large body size) should improve the ability of fishes to rapidly colonize and expand their range into novel environment (Glazier & Eckert, 2002); (2) the 'locomotion activity hypothesis' stating that certain morphological characteristics, and notably those related to the locomotion (e.g. body shape), should promote long-distance dispersal and colonization (Gutierrez & Menendez, 1997; Glazier & Eckert, 2002); (3) the 'thermal (or physiological) hypothesis' stating that cold-tolerant species should have larger geographic range because they re-colonized watersheds earlier than other species after the last glaciation (Hocutt & Wiley, 1986); (4) the 'migratory hypothesis' stating that diadromous species should have a larger geographic range because of their ability to use salt water as a pathway for colonizing new areas; (5) the 'habitat use hypothesis' stating that species able to live in lakes (i.e. obligatory or partially lake users) should have a greater geographic range due to their ability to use pro-glacial lakes for colonization (Griffiths, 2010); and (6) the 'phylogenetic hypothesis' stating that closely related species should have similar ranges because the traits underlying dispersal and colonization ability are phylogenetically conserved. Finally, we explored relationships between phenotypic attributes, mid-range latitude and longitude, and we used partial regressions to determine the mechanisms behind the relationship between GRS, longitude and latitude.

MATERIAL AND METHODS

The database

The database contains information on 145 of the 203 native freshwater fish species known for Canada (Nelson *et al.*, 2004); other species were not included because of a lack of functional and/or phylogenetic information. For each species, we documented the GRS (in km^2), the mean latitude of each species GRS (the mid-range latitude), and a series of phenotypic attributes (see below). We focus here only on the range and mid-range latitude and longitude of the Canadian range of the 145 species. Most of these species also occur further south (i.e. in the USA), meaning that the mid-range latitude and GRS were underestimated for most species. However, we were interested in the traits allowing species a range expansion after a glaciated period, and this also allows us to control for the duration since post-glacial

colonization. It is noteworthy that the Canadian GRS was strongly correlated ($r^2 = 0.68$, $P < 0.001$) to the whole range size (i.e. USA + Canadian ranges); our findings could hence be extrapolated to the whole range size of these species. Using ArcMap, the area of the GRS of each species was measured from distribution maps digitized from Scott & Crossman (1973), and the mid-range latitude and longitude were extracted as the centroid of the polygon best representing the Canadian range of each species. Phenotypic attributes were gathered from published dataset and measurements from species drawings in Scott & Crossman (1973). Specifically, two morphological traits (shape and swimming factors; i.e. the ratio of total body length to maximum body width, and the ratio of minimum caudal peduncle width to maximum caudal fin width, respectively) were measured. One physiological trait (thermal preference, semi-categorical variable with five categories: category 1 being species tolerating coldest waters, and category 5 species tolerating warmest waters), migratory status (categorical variable with three categories: non-migratory, amphidromous or potadromous species), habitat use (categorical variable with three categories: living in rivers only, living in lakes only, or living in both rivers and lakes), and eight life-history traits (maximum body length, maximum lifetime, age at maturity, beginning of the spawning period, end of the spawning period, spawning duration, mean fecundity and mean egg size; all continuous variables) were obtained from published studies (Coker *et al.*, 2001; Froese & Pauly, 2006). Collinearity in eight life-history traits was reduced by performing a principal components analysis (PCA). All variables were normalized before the analysis, and the first two axes of the PCA (representing 63% of the total variance) were used to summarize interspecific variation in life-history traits (see Table S1 for details about the PCA). Thermal preference was defined as the preferred summer water temperature measured in the laboratory (see Coker *et al.*, 2001), and was different from the thermal tolerance (i.e. a range of tolerated temperature). We did not use thermal tolerance as data were available for a limited number of fish species only. Overall, this led to seven variables summarizing the phenotypic attributes of species. The two morphological traits were related to the 'locomotion activity hypothesis', the two first axes of the PCA to the 'life-history hypothesis', the physiological trait to the 'thermal hypothesis', the migratory status to the 'migration hypothesis', and the habitat use to the 'habitat use hypothesis' (see Table 1). Collinearity among these seven variables was relatively weak (r < 0.34, Table 1).

A molecular-based phylogeny [from a 650 bp fragment of the cytochrome *c* oxidase subunit I (COI) mitochondrial gene] describing the phylogenetic distance between species was developed as follows. DNA sequences obtained from Hubert *et al.* (2008) were aligned using ClustalX (Thompson, 1997). Using the software Mega 5 (Tamura *et al.*, 2011), we determined the best substitution model (GTR+G+I) using a Bayesian information criterion and then computed the maximum likelihood phylogenetic trees to quantify phylogenetic distances among species. The phylogenetic relationship among species was used to test the 'phylogenetic hypothesis'.

Table 1 Correlation matrix of phenotypic attributes selected for the analysis. Each trait has been associated to one of the five hypotheses we sought to test. The sixth hypothesis we tested was related to the molecular phylogeny presented in Fig. 4.

Phenotypic attributes	Life-history traits (axis 1)	Life-history traits (axis 2)	Shape factor	Swimming factor	Water temperature preference[1]	Habitat use[1]	Migratory behaviour[1]
Life-history traits (axis 1)	1	–	–	–	–	–	–
Life-history traits (axis 2)	−0.071	1	–	–	–	–	–
Shape factor	0.172	0.111	1	–	–	–	–
Swimming factor	−0.305	0.209	−0.153	1	–	–	–
Thermal preference*	0.242	0.212	0.056	0.336	1	–	–
Habitat use*	0.029	0.086	0.004	0.117	n.a.	1	–
Migratory behaviour*	0.217	0.134	0.004	0.111	n.a.	n.a.	1

*These three variables are categorical and we present here the percentage of deviance explained by these variables. Deviance was calculated from generalized linear models. n.a., not applicable.

Statistical analyses

We first tested for the existence of the Rapoport's rule in the Canadian freshwater fish fauna by plotting the GRS against the mid-range latitude of each species, and then calculating the significance of the Pearson correlation coefficient for this relationship. This relationship was explored using both untransformed and log-transformed ($\log_e$) GRS (Brown *et al.*, 1996). In addition, we tested for longitudinal patterns in GRS by calculating the significance of the Pearson correlation coefficient between range size and the mid-range longitude. Finally, we represented GRS in two-dimensional geographic space using the digitized range maps of each species that we averaged in grids of 0.7×0.7 degrees (Hawkins & Diniz-Filho, 2006).

We explored the 'phylogenetic hypothesis' by testing for a significant phylogenetic niche conservatism (hereafter 'PNC') in GRS (Cooper *et al.*, 2010). If so, this would provide strong support for the 'phylogenetic hypothesis' because it would indicate that closely related species share similar values of GRS. PNC was tested following the framework proposed by Wiens *et al.* (2010) who recommend fitting Brownian, Ornstein–Uhlenbeck and white noise models, and then comparing the fit of the models. These models of traits evolution were fitted using the Geiger library under the R environment (Harmon *et al.*, 2008), and were compared using the Akaike Information Criteria (AIC).

We tested the five hypotheses related to phenotypic traits through a model comparison procedure (Johnson & Omland, 2004). We used phylogenetic generalized least squares (PGLS) models that include or not phylogeny as an expected covariance matrix since explanatory variables and/or the dependent variables may display phylogenetic conservatism. Three models of traits evolution were considered: white noise, Brownian and Ornstein–Uhlenbeck. In all models, GRS was the dependent variable (log) and the seven variables related to phenotypic attributes were set as fixed effects. To limit the total number of models being analysed and for increasing biological interpretability, we did not include interaction terms. It is noteworthy that preliminary analyses indicated that information gained from including two-term interactions was weak (not shown). We compared the AIC corrected for small sample sizes (AICc)

among all possible models arising from the combination of the seven phenotypic variables. In total, we tested 128 models that included the phylogeny under the Brownian model of evolution, 128 models that included the phylogeny under the Ornstein–Uhlenbeck model of evolution, and 128 models that did not included the phylogeny (the white noise models). Among these 384 models, the model that displayed the lowest AICc value was considered as the 'best model' (Johnson & Omland, 2004). In addition, we calculated the differences in these AICc values between each model and the best model (i.e. ΔAICc). A single best model cannot be assumed if the ΔAICc with other competing models is not greater than 2 U (Johnson & Omland, 2004); therefore, we retained all models with ΔAICc < 2 U as potential candidate models.

To control for the potential effects of the use of different glacial refugia and dispersal corridors, PGLSs were fitted at several spatial extents. PGLSs were first fitted at the Canadian extent by considering all species ($n = 145$) occurring in the region that was covered by the Wisconsinan ice sheet. In the second step, PGLSs were fitted using species only occurring east of the continental divide in the Rocky Mountains ($n = 120$) because the Rocky Mountains split Canadian fishes into two different faunas (Scott & Crossman, 1973; Hocutt & Wiley, 1986), and species living on either side have limited possibilities to expand across this continental divide. The region west of the continental divide was not analysed because sample size (i.e. species richness, $n = 21$) is too low to yield sufficient statistical power. We conducted the same analyses at a finer scale, including species that only used the Mississippian glacial refugium ($n = 57$), the most important one in North America (Fig. 1; Hocutt & Wiley, 1986). Analyses were conducted at this scale control both for differences in the availability of dispersal corridors and for the differential use of glacial refugia. The other known glacial refugia used by the Canadian freshwater fish fauna were not analysed independently because the species richness was too low (Hocutt & Wiley, 1986; Mandrak & Crossman, 1992). The same statistical procedure (PGLSs including or not phylogeny followed by a model selection procedure) was used to test which phenotypic attributes are best correlated with the mid-range latitude and longitude, respectively. For the sake of

clarity, these analyses were performed only at the Canadian extent (i.e. including all species).

To examine direct and indirect relationships among GRS, mid-range latitude, mid-range longitude and the common factors, we used partial regressions to estimate the proportion of explained variance by one factor on a given variable of interests (GRS, mid-range longitude or mid-range latitude), while controlling for the effect of other factors (Cardon *et al.*, 2011). Partial regressions are an extension of partial correlations that allow for consideration of categorical factors. Although partial regression is not the solution to all of the shortcomings of correlative approaches, it does allow identification of spurious correlations. In this analysis, we were specifically interested in testing for, and quantifying, the strength of direct and indirect relationships. Assuming that mid-range latitude, mid-range longitude and GRS all co-vary together, the relationships between common factors and GRS can either be direct or indirect.

RESULTS

The distribution of GRS based on untransformed data is highly skewed to the right, indicating that most species have a relatively small geographic range (mean $\pm$ SD = 1.57×10^6 km^2 $\pm$ 2.22×10^6, skewness = 1.93, kurtosis = 5.96, Fig. 2a). Based on the log-transformed data, the distribution is more normal and slightly skewed to the left (mean $\pm$ SD = 13.35 $\pm$ 2.08, skewness = -1.33, kurtosis = 6.65, Fig. 2b). We detected a significant pattern whereby the GRS of species increases linearly and positively with the mid-range latitude ($r = 0.696$, d.f. = 143, t = 11.60, $P < 0.001$, Fig. 3a). When the GRS is log transformed, the relationship with the mid-range latitude becomes curvilinear [generalized linear model (GLM) including a quadratic term, r^2 = 0.604], that is, the GRS increases until 60° latitude and then slightly decreases (Fig. 3b). Additionally, we detected a slight, but significant, longitudinal pattern in GRS, since GRS tended to be smaller for fishes living in the most eastern part of the Canada ($r = -0.227$, d.f. = 113, t = -2.48, $P = 0.014$, not shown). However, this negative correlation was probably a geographical artefact due to the smaller continental landmass availability in eastern Canada (Fig. 1); arguably, this correlation became positive when a GLM including both mid-range longitude and mid-range latitude was computed (not shown). When plotted in two dimensions, spatial patterns confirmed that mean range sizes tended to be larger at higher latitude, but that range sizes were maximal in size in the centre of the Canada and slightly decreased in the upper north limit of the country (Fig. 1).

We did not detect any significant phylogenetic conservatism for the GRS of Canadian freshwater fishes: the white noise model of evolution had the lowest AIC value (AIC = 626.95) compared with the Brownian and Ornstein–Uhlenbeck models of evolution (AICs = 627.91 and 630.95, respectively); therefore, closely related species do not share a similar GRS (Fig. 4).

At the Canadian extent, the best PGLS model retained for explaining the GRS includes five variables related to four hypotheses: migratory status related to the 'migratory' hypoth-
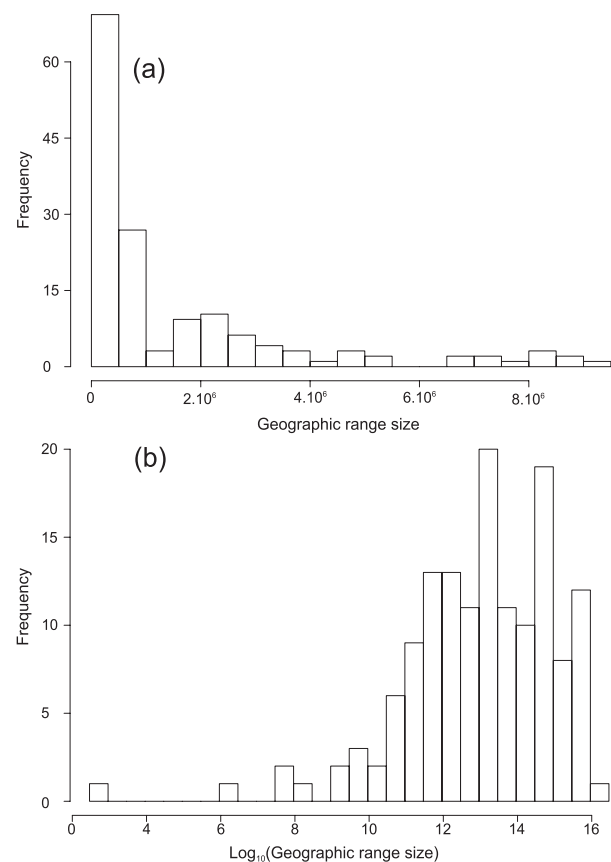
Figure 2 Frequency distribution of (a) geographic range size and (b) log-transformed geographic range size of Canadian freshwater fishes.

esis; habitat use related to the 'habitat use' hypothesis; thermal preference related to the 'thermal' hypothesis; and shape factor and swimming factors related to the 'locomotion activity' hypothesis (Table 2). Three other models had a ΔAICc < 2, indicating that they can also be considered as good candidate models (Table 2). These additional models all included migratory status, habitat use, thermal preference and the shape factor (Table 2). Each also included variables related to the 'locomotion activity' or the 'life-history' hypotheses (see Table 2 for details). All these models included the phylogeny as an expected covariance matrix, and were best fitted using the Ornstein–Uhlenbeck model of evolution. For the sake of clarity, only the best model will be interpreted. This model explained 24.02% of the observed variance in GRS. According to this model, potadromous species with cold thermal preferences have larger geographic ranges (Fig. 5a, 5b) and riverine species have smaller geographic ranges than lacustrine species (Fig. 5c). Species with high values of shape and swimming factors had higher GRS (coefficient slopes: a = 0.375 $\pm$ 0.027 and a = 3.258 $\pm$ 0.662 for the shape and swimming factors, respectively).

We found similar results when PGLSs were run using species occurring only east of the continental divide, or using species from the Mississippian refugium. In each case, the best model
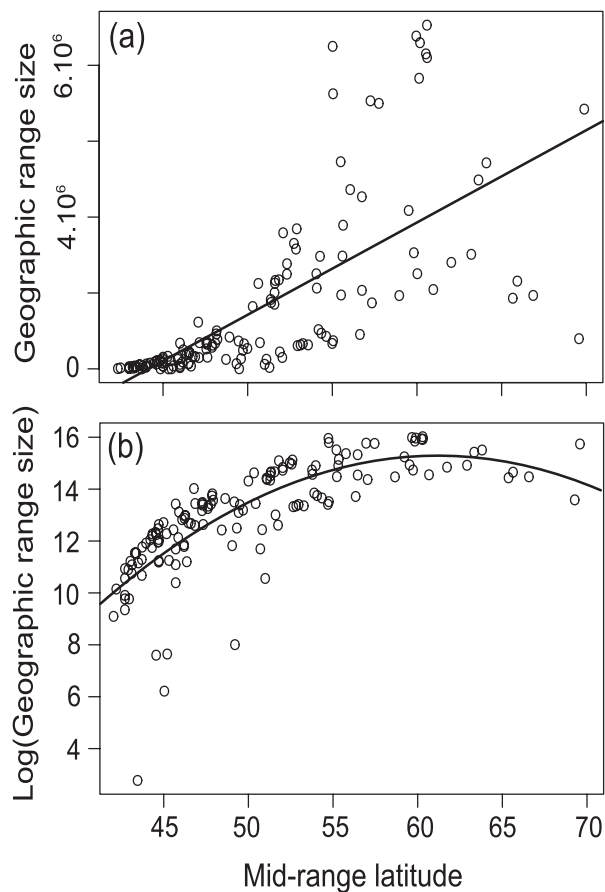
Figure 3 Relationship between mid-range latitude and geographic range size of Canadian freshwater fishes. The relationship is shown for (a) raw and (b) log-transformed geographic range size data.

included migratory status, habitat use, thermal preference, and shape and swimming factors (Tables S2, S3), and the models were best fit using the Ornstein–Uhlenbeck model of evolution (Tables S2, S3). We identified other models having a ΔAICc < 2, and these models included one of the life-history trait axes (the same than for the whole dataset, Tables S2, S3). Overall, these reduced-scale analyses strengthened our findings that the 'thermal', 'habitat use', 'migratory' and 'locomotion ability' hypotheses were the best for explaining the GRS of the Canadian fish fauna.

When PGLSs were applied to the mid-range latitude (at the Canadian extent), the best model included four variables. Three of them (thermal preference, habitat use, migration status) were common to the best model predicting the GRS (Table 2). The last one was the second axis of the PCA for life-history traits and was related to the 'life-history' hypothesis. This PCA axis summarized the beginning and end of the spawning period, as well as the mean egg size (Table S1). This model was best fitted using the Ornstein–Uhlenbeck model of evolution. This model explained 52.53% of the total variance observed in mid-range latitude.

When PGLSs were applied to the mid-range longitude (at the Canadian extent), the best model included was very similar to the one selected for the mid-range latitude (Table 2). It includes thermal preference, habitat use, migration status and the second PCA axis for life-history traits. However, this model was best fitted using the Brownian model of evolution. This model explained 37.26% of the total variance observed in mid-range longitude.

Partial regressions relating GRS, mid-range latitude, mid-range longitude and common explanatory factors (see Fig. 6) demonstrated that the significant relationship detected between thermal preference and GRS was no longer significant once the effects of other variables (notably mid-range latitude) were controlled (Fig. 6). Therefore, the relationship between thermal preference and GRS was spurious. Because the relationship between thermal preference and mid-range latitude was still significant in the partial regression (Fig. 6), the spurious relationship between thermal preference and GRS was probably due to thermal preference that correlates directly with mid-range latitude, which co-varies with GRS (i.e. an indirect relationship between thermal tolerance and GRS). This analysis further reveals that the relationships between habitat use and GRS, and between migratory status and GRS, were both direct and indirect (Fig. 6). Shape and swimming factors were directly related to GRS (Fig. 6). Finally, this analysis reveals that mid-range latitude was related only to thermal preference and migratory status (other relationships were spurious), whereas mid-range longitude was related only to habitat use and the second axis of the life-history traits PCA (Fig. 6).

DISCUSSION

We show that the Canadian freshwater fish fauna is characterized by a clear spatial pattern of GRS – species living at higher latitude display a greater GRS (i.e. Rapoport's rule, Griffiths, 2010). There was also a longitudinal pattern, whereby species tended to have greater GRS at higher longitude, although this pattern was weak and became apparent only once the latitudinal pattern was controlled for. By combining phylogenetic regressions and model selection procedures, we further highlight the ecological traits underlying this pattern. Habitat use, migratory status, thermal preference and swimming ability traits affect either directly or indirectly the propensity of a species to occupy a large GRS.

Rapoport's rule has been described in many organisms (Gaston, 2003, 2008b, 2009; Morin & Chuine, 2006). In freshwater fishes, this pattern was identified for several faunas including freshwater fishes in North America (Griffiths, 2010). Although we used a more restricted dataset than Griffiths (2010), our results confirm those previous results, demonstrating that Rapoport's rule holds true even for a geographic area that has been colonized only recently (i.e. < 15,000 yr BP, Hocutt & Wiley, 1986). It is noteworthy that when GRS was log transformed, the correlation between GRS and mid-range latitude pointed to a curvilinear relationship in which GRS reaches a maximum for species living at latitude around 60° N, and then
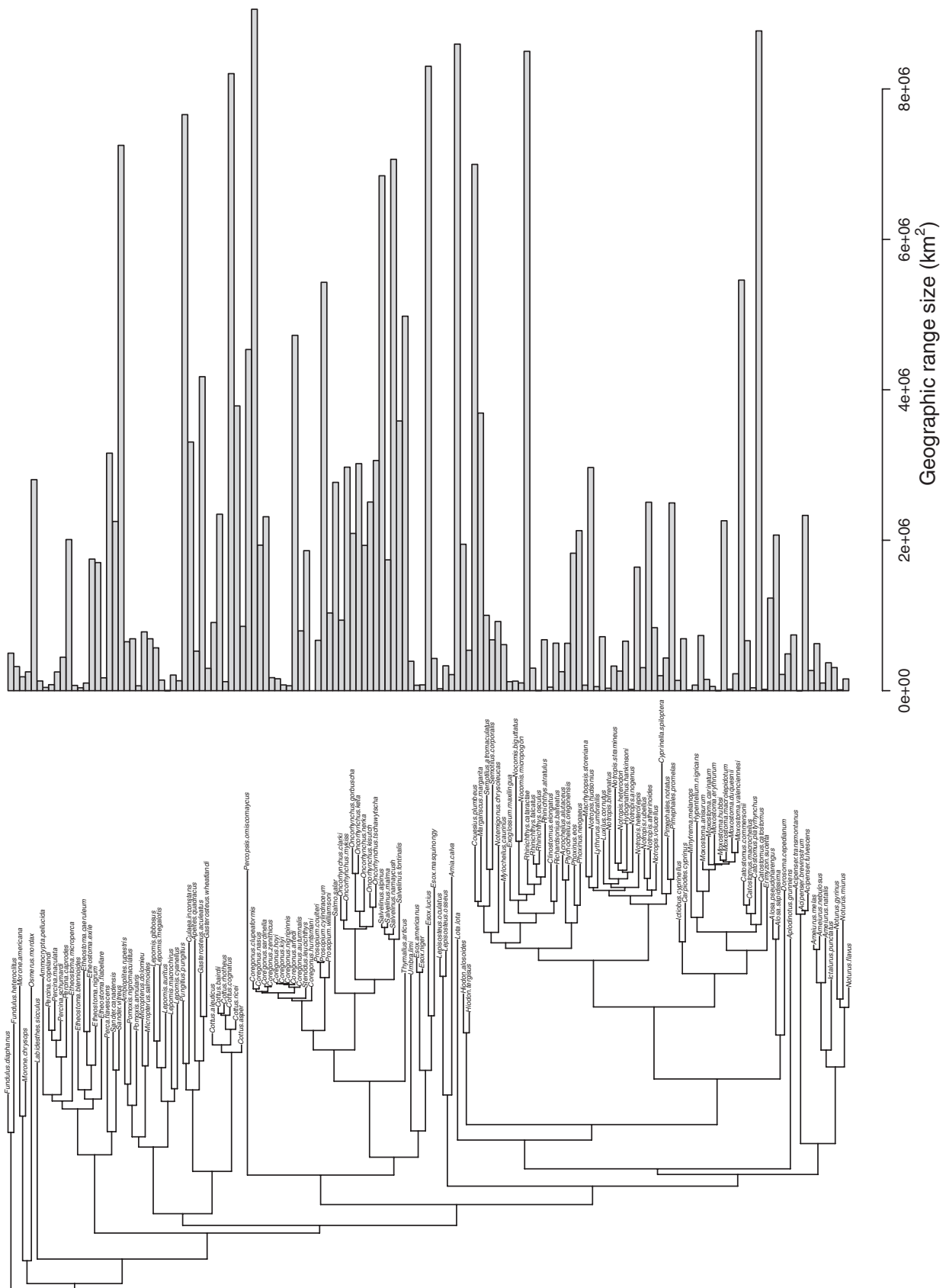
Figure 4 Relationships between the molecular phylogeny and the geographic range size. We detected no significant phylogenetic conservatism for geographic range size.

Table 2 Model selection for the geographic range size, the mid-range latitude and the mid-range longitude of Canadian freshwater fishes. Overall, for each three traits, 384 models were compared using AICc (128 models including phylogeny under a Brownian model of evolution, 128 models including phylogeny under a Ornstein–Uhlenbeck model of evolution, 128 models excluding phylogeny). The table only shows the models with ΔAIC < 2 (i.e. the best fit models). For each each of the three traits, the models with the lowest AIC are highlighted in bold. 'X' indicates the variables that were included in the model(s). For the geographic range size and the mid-range latitude, all selected models were best fitted using the Ornstein–Uhlenbeck model of trait evolution as an expected covariance matrix. For the mid-range longitude, all selected models were best fitted using the Brownian model of trait evolution as an expected covariance matrix.

	Geographic range size				Mid-range latitude		Mid-range longitude	
	Model 1	Model 2	Model 3	Model 4	**Model 1**	Model 2	**Model 1**	Model 2
Traits								
Life-history 1 ('life-history' hypothesis)	–	–	–	X	–	X	–	X
Life-history 2 ('life-history' hypothesis)	–	X	X	X	**X**	X	**X**	X
Shape factor ('locomotion activity' hypothesis)	**X**	X	X	X	–	–	–	–
Swimming factor ('locomotion activity' hypothesis)	**X**	–	X	–	–	–	–	X
Thermal preference ('thermal' hypothesis)	**X**	X	X	X	**X**	X	**X**	X
Migration status ('migratory' hypothesis)	**X**	X	X	X	**X**	X	**X**	X
Habitat use ('habitat use' hypothesis)	**X**	X	X	X	**X**	X	**X**	X
AICc	**575.31**	575.87	578.42	577.15	**-296.13**	-294.96	**1135.82**	1137.65
Weight	**0.3**	0.22	0.17	0.12	**0.303**	0.168	**0.47**	0.187

AIC, Akaike information criteria.
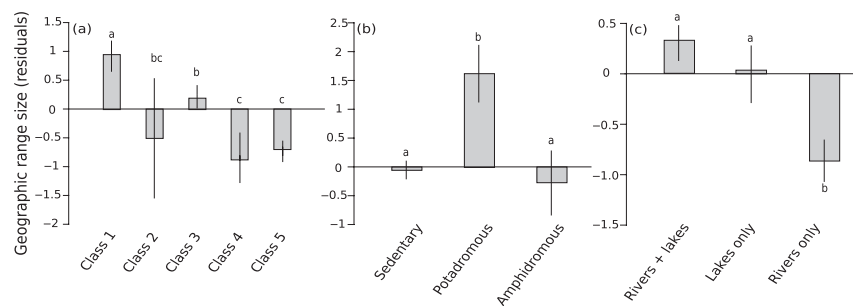
Figure 5 Relationships between geographic range size and (a) water temperature preference (class 1 is cold-tolerant species, class 5 is warm-tolerant species, and classes 2–4 are intermediate species), (b) migratory behaviour, and (c) habitat use ('Rivers + lakes' are species using both types of habitat; 'Rivers only' are species living only in rivers; and 'Lakes only' are species of Canadian freshwater fishes living only in lakes. The same letter above the bars indicates categories that did not differ significantly (contrast tests, $P > 0.05$) one from each other.

slightly decreases at higher latitudes. This pattern was obvious when mean range sizes were plotted in two dimensions (Fig. 1), and may reflect an inherent limitation of the method we used to define the latitude of each species; with the mid-point method, as mid-range latitude values approach the border, there will be range size constraints, which have consequences on range size patterns (i.e. the mid-domain effect, Colwell & Hurtt, 1994). Finally, we highlight a weak longitudinal pattern illustrating that more easterly species tended to have greater GRS; this may correspond to species that use the Mississippian and Atlantic refugia (see Fig. 1) from which they colonized.

Our phylogenetic regressions highlight six ecological characteristics of species that were significantly correlated with GRS, mid-range longitude and/or mid-range latitude. Importantly, three of them (habitat use, migratory status, thermal preference) were common to the three geographic components, and were

related to the 'habitat use', 'migratory' and 'thermal' hypotheses. These findings were congruent over all spatial extents we considered here. This indicates that our findings were robust to the biogeographic and evolutionary histories of species. For instance, the relationships between phenotypic attributes and GRS may be confounded if two faunas that occupied two different glacial refugia during the last glaciation are analysed simultaneously (Garcia-Barros & Benito, 2010). Because we accounted for such potential biases, we were able to focus on independent relationships between phenotypic attributes, GRS and mid-range latitude.

Another limitation of our understanding of processes underlying large-scale biodiversity patterns is that most inferences at such scales are made from correlations. Using partial regressions, we showed that migratory status and thermal preference were directly correlated with mid-range latitude; life-history
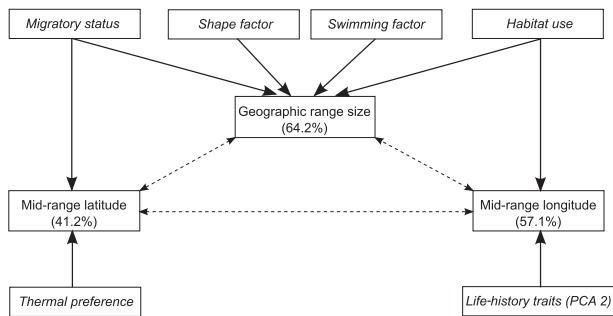
Figure 6 Direct and indirect relationships between geographic range size, mid-range latitude, mid-range longitude, thermal preference, migratory status, habitat use, shape and swimming factors, and life-history traits of Canadian freshwater fish species. Assuming a covariation between mid-range latitude, mid-range longitude and geographic range size (the double-headed and dotted arrow, all significant at $\alpha = 0.05$), we performed partial regressions analysis to quantify the significance of each factor on mid-range latitude, mid-range longitude and geographic range size, respectively. For the sake of clarity, only the significant paths (symbolized as arrows) are indicated in the figure (F-tests, all P-values < 0.05). All other paths (e.g. thermal preference to geographic range size) were not significant after controlling for other variables and are not shown. Numbers under brackets indicate the total proportion of deviance explained when all factors were accounted for.

traits and habitat use were directly correlated with mid-range longitude; and migratory status, habitat use, and swimming and shape factors were directly correlated with GRS. All other relationships (and notably the correlation between thermal tolerance and GRS) were actually spurious or indirect. This means that inferring processes from the apparent correlation between thermal preference and GRS would have been misleading, as this correlation is actually not direct. Furthermore, our results showed that different phenotypic attributes affect the three components of the GRS and Rapoport's rule, and we propose biological scenarios that may explain relationships highlighted in Fig. 6.

Two main phenotypic attributes affecting mid-range latitude are the migratory status and thermal tolerance. Specifically, mid-range latitude tended to be higher for migratory species (either anadromous or potadromous) that prefer cold temperature. These findings are consistent with the hypothesis that pro-glacial lakes, that is, the main corridors for fishes following glacial periods (Fig. 1, Hocutt & Wiley, 1986), were cold environments that would have favoured the early re-colonization of cold-tolerant species such as salmonids. Contrary to previous studies, we considered thermal preference rather than thermal tolerance per se. Hypotheses behind thermal tolerance and thermal preference differ: the former allows testing hypotheses regarding the ability of species to cope with environmental heterogeneity, whereas the latter relates to earlier colonization (Stevens, 1989; Addo-Bediako et al., 2000; this study). Future studies should focus on how thermal tolerance per se affects GRS in fishes. Under such a scenario, highly mobile species that

lacked cold-water tolerance may have had to wait thousands of years before extending their distribution range northwards, at which point many ephemeral dispersal corridors no longer existed. Similarly, migratory species (notably anadromous species) may have used alternative dispersal corridors, such as marine ecosystems (Griffiths, 2006, 2010; Reyjol et al., 2007), to disperse northwards. Regarding the mid-range longitude, we found independent relationships with habitat use and reproductive timing (second PCA axis). Species able to live at least partially in lakes tended to have greater mid-range longitude, which may be explained by pro-glacial lakes being important dispersal corridors after the last glaciation. Accordingly, these lakes were mainly southeast–northwest oriented (Fig. 1, Hocutt & Wiley, 1986) and may have been used by species that used the Mississippian and Atlantic refugia to re-colonize western parts of the continent. The second axis of the PCA was mainly related to the timing of the reproduction period (Table S1), and it is likely that these species were able to produce propagules during the cold conditions that followed the glaciated period. Finally, regarding GRS, we identified independent relationships with migratory status, habitat use, and swimming and shape factors. Potadromous species living at least partially in lakes had greater GRSs. As for the mid-range longitude, these results agree with the predictions that primarily riverine and non-migratory fish species were probably limited in their early post-glaciation dispersal because of their inability to use pro-glacial lakes as dispersal corridors (Griffiths, 2006, 2010). Swimming and shape factors were related to the 'locomotion' activity, and we found that species having higher swimming and shape factors displayed greater GRS. It is hence likely that those morphological characteristics may allow long-distance dispersal and colonization (Gutierrez & Menendez, 1997; Glazier & Eckert, 2002).

The 'phylogenetic' hypothesis was not, or at most, weakly supported since we failed to detect significant phylogenetic conservatism for GRS. Contrary to previous studies corroborating the 'phylogenetic' hypothesis (Jablonski, 1987; Hunt et al., 2005), we analysed present-day species at relatively small phylogenetic (i.e. within a vertebrate group) and temporal (< 15,000 years) scales. Webb & Gaston (2005) stated that a phylogenetic signal in GRS would indirectly validate a role for biology (i.e. the ecological and life-history characteristics of species) in the determination of range size. We provide evidence that, even in the absence of phylogenetic signal, certain species traits can actually determine, at least partially, GRS of fishes. However, we failed to find significant effects of body size, a trait traditionally believed to influence GRS (Taylor & Gotelli, 1994; Gaston & Blackburn, 1996a; Pyron, 1999; Rosenfield, 2002). This trait may be acting at smaller spatial scales in relation to feeding behaviour or predator avoidance, that is, to biotic drivers of spatial organization (Reyjol et al., 2008). Alternatively, it is possible that this trait was negligible in our analyses because we considered a large set of life-history and morphological traits that better reflects the complexity of traits-geographic range relationships. This result highlights that merging macroecological analyses and functional ecology may constitute a promising way to understand spatial patterns and processes of large-scale biodiversity patterns.

To summarize, we identified biological traits that affect, either directly or indirectly, the three components of spatial patterns in GRS (GRS, mid-range longitude, mid-range latitude) of a recently colonizing fauna, Canadian freshwater fishes. Our results provide support for a significant role of migratory behaviour, habitat and thermal preferences of fish species as these traits were related to the three components examined. The importance of thermal tolerance in explaining GRS, mid-range longitude and mid-range latitude is consistent with results found for other animal groups (Addo-Bediako *et al.*, 2000; Calosi *et al.*, 2010). However, our results show that both traits related to the ability to use dispersal corridors and traits related to mobility per se may be important for predicting and understanding the Rapoport's rule, and more generally, spatial patterns in GRS in animals.

Our trait-based models explained twice as much variance for mid-range latitude and mid-range longitude (~53 and ~37%, respectively) than for GRS (~24%). We can speculate that mid-range latitude and longitude may better reflect the south–north and east–west patterns of colonization than GRS, which may better reflect range expansion. If so, this would suggest that functional traits (at least those analysed here) would be more important in explaining the post-glacial latitudinal and longitudinal shifts in the geographic distribution of species than range expansions. However, a large amount of variance remains unexplained in our models, particularly for GRS. This indicates that simple biological traits may not be comprehensive predictors of the GRS of species, even when phylogenetic relatedness is explicitly considered.

The conclusion that simple biological traits are poor predictors of GRS has important conservation implications. GRS has often been considered as an important characteristic for predicting the extinction risk of species (Gaston & Blackburn 1996b; Purvis *et al.*, 2000; Lee & Jetz, 2011). Species with small GRS are expected to be more sensitive to environmental changes and, hence, more prone to extinction (Purvis *et al.*, 2000; Lee & Jetz, 2011). A main quest in conservation biology is to identify biological traits correlated with GRS, to be used as biological proxies of extinction risk in predictive models. Our results indicate that it would be premature to use simple biological traits to predict GRS of – and hence extinction risk for – freshwater fishes.

ACKNOWLEDGEMENTS

REFERENCES

Addo-Bediako, A., Chown, S.L. & Gaston, K.J. (2000) Thermal tolerance, climatic variability and latitude. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 739–745.

Avise, J.C., Walker, D. & Johns, G.C. (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 1707–1712.

Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range size: size, shape, boundaries and internal structure. *Annual Review of Ecology, Evolution and Systematics*, **27**, 597–623.

Calosi, P., Bilton, D.T., Spicer, J.I., Votier, S.C. & Atfield, A. (2010) What determines a species' geographical range? Thermal biology and latitudinal range size relationships in European diving beetles (Coleoptera: Dytiscidae). *Journal of Animal Ecology*, **79**, 194–204.

Cardon, M., Loot, G., Grenouillet, G. & Blanchet, S. (2011) Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal analysis. *Journal of Animal Ecology*, **80**, 657–667.

Coker, G.A., Portt, C.B. & Minns, C.K. (2001) Morphological and ecological characteristics of Canadian freshwater fishes. Canadian Manuscript Report of Fisheries and Aquatic Sciences 2554.

Colwell, R.K. & Hurtt, G.C. (1994) Nonbiological gradients in species richness and a spurious Rapoport effect. *The American Naturalist*, **144**, 570–595.

Cooper, N., Jetz, W. & Freckleton, R.P. (2010) Phylogenetic comparative approaches for studying niche conservatism. *Journal of Evolutionary Biology*, **23**, 2529–2539.

Froese, R. & Pauly, D. (2006) FishBase. World Wide Web electronic publication, version (12/2006). Available from http://www.fishbase.org/home.htm (accessed December 2006).

Garcia-Barros, E. & Benito, H.R. (2010) The relationship between geographic range size and life history traits: is biogeographic history uncovered? A test using the Iberian butterflies. *Ecography*, **33**, 392–401.

Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.

Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.

Gaston, K.J. (2008a) Biodiversity and extinction: the importance of being common. *Progress in Physical Geography*, **32**, 73–79.

Gaston, K.J. (2008b) Species-range size distributions: products of speciation, extinction and transformation. *Philosophical Transactions: Biological Sciences*, **353**, 219–230.

Gaston, K.J. (2009) Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1395–1406.

Gaston, K.J. & Blackburn, T.M. (1996a) Global scale macroecology: interactions between population size, geographic range size and body size in the Anseriformes. *Journal of Animal Ecology*, **65**, 701–711.

Gaston, K.J. & Blackburn, T.M. (1996b) Rarity and body size: importance of generality. *Conservation Biology*, **10**, 1295–1298.

Glazier, D. & Eckert, S. (2002) Competitive ability, body size and geographical range size in small mammals. *Journal of Biogeography*, **29**, 81–92.

Griffiths, D. (2006) Patterns and process in the ecological biogeography of European freshwater fish. *Journal of Animal Ecology*, **75**, 734–751.

Griffiths, D. (2010) Pattern and process in the distribution of North American freshwater fish. *Biological Journal of the Linnean Society*, **100**, 46–61.

Gutierrez, D. & Menendez, R. (1997) Patterns in the distribution, abundance and body size of carabid beetles (Coleoptera?: Caraboidea) in relation to dispersal ability. *Journal of Biogeography*, **24**, 903–914.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**, 129–131.

Hawkins, B.A. & Diniz-Filho, J.A.F. (2006) Beyond Rapoport's rule: evaluating range size patterns of New World birds in a two-dimensional framework. *Global Ecology and Biogeography*, **15**, 461–469.

Hocutt, C. & Wiley, E.O. (1986) *The zoogeography of North American freshwater fishes*. Wiley, New York.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J. & Bernatchez, L. (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE*, **3**, e2490.

Hunt, G., Roy, K. & Jablonski, D. (2005) Species-level heritability reaffirmed: a Comment on 'On the heritability of geographic range sizes'. *The American Naturalist*, **166**, 129–135.

Jablonski, D. (1987) Heritability at the species level: analysis of geographic ranges of Cretaceous mollusks. *Science*, **238**, 360–363.

Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**, 101–108.

Lee, T.M. & Jetz, W. (2011) Unraveling the structure of species extinction risk for predictive conservation science. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1329–1338.

Mace, G.M., Collen, B., Fuller, R.A. & Boakes, E.H. (2010) Population and geographic range dynamics: implications for conservation planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3743–3751.

Mandrak, N.E. & Crossman, E.J. (1992) Postglacial dispersal of freshwater fishes into Ontario. *Canadian Journal of Zoology*, **70**, 2247–2259.

Morin, X. & Chuine, I. (2006) Niche breadth, competitive strength and range size of tree species: a trade-off based framework to understand species distribution. *Ecology Letters*, **9**, 185–195.

Mouillot, D. & Gaston, K.J. (2007) Geographical range size heritability: what do neutral models with different modes of speciation predict? *Global Ecology and Biogeography*, **16**, 367–380.

Nelson, J.S. Crossman, E.J., Espinosa-Pérez, H., Findley, L.T., Gilbert, C.R., Lea, R.N. & Williams, J.D. (2004) *Common and scientific names of fishes from the United States, Canada, and Mexico* 6th edn. American Fisheries Society, Bethesda, MD. Special Publication 29.

Purvis, A., Gittleman, J., Cowlishaw, G. & Mace, G. (2000) Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1947–1952.

Pyron, M. (1999) Relationships between geographical range size, body size, local abundance, and habitat breadth in North American suckers and sunfishes. *Journal of Biogeography*, **26**, 549–558.

Reyjol, Y., Hugueny, B., Pont, D., Bianco, P.G., Beier, U., Caiola, N., Casals, F., Cowx, I., Economou, A., Ferreira, T., Haidvogl, G., Noble, R., de Sostoa, A., Vigneron, T. & Virbickas, T. (2007) Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, **16**, 65–75.

Reyjol, Y., Rodriguez, M.A., Dubuc, N., Magnan, P. & Fortin, R. (2008) Among- and within-tributary responses of riverine fish assemblages to habitat features. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 1379–1392.

Rosenfield, J.A. (2002) Pattern and process in the geographical ranges of freshwater fishes. *Global Ecology and Biogeography*, **11**, 323–332.

Roy, K., Hunt, G., Jablonski, D., Krug, A.Z. & Valentine, J.W. (2009) A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1485–1493.

Scott, W.B. & Crossman, E.J. (1973) *Freshwater fishes of Canada*. Fisheries Research Board of Canada, Ottawa, Bulletin 184.

Stevens, G. (1989) The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, **133**, 240–256.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Taylor, C.M. & Gotelli, N.J. (1994) The macroecology of *Cyprinella*: correlates of phylogeny, body size, and geographical range. *The American Naturalist*, **144**, 549–569.

Thompson, J. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Webb, T.J. & Gaston, K.J. (2005) Heritability of geographic range sizes revisited: a reply to Hunt et al. *The American Naturalist*, **166**, 136–143.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Davies, T.J., Grytnes, J.-A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M. & Stephens, P.R. (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, **13**, 1310–1324.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Table S1 Loadings of a principal components analysis (PCA) used to summarize interspecific variation in eight life-history traits. Major loadings are shown in bold.

Table S2 Outputs of the model selection procedure for the geographic range size of Canadian freshwater fishes east of the continental divide. Overall, 384 models were compared using AICc (128 models including phylogeny under a Brownian model of evolution, 128 models including phylogeny under a Ornstein–Uhlenbeck model of evolution, 128 models excluding phylogeny). The table only shows the five models with ΔAIC < 2 (i.e. the best fit models). A cross indicates the variables that were included in the model. These five selected models were best fitted using the Ornstein–Uhlenbeck model of trait evolution as an expected covariance matrix.

Table S3 Outputs of the model selection procedure for the geographic range size of Canadian freshwater fishes that used the Mississippian refugium during the last ice age. Overall, 384 models were compared using AICc (128 models including phylogeny under a Brownian model of evolution, 128 models including phylogeny under a Ornstein–Uhlenbeck model of evolution, 128 models excluding phylogeny). The table only shows the four models with ΔAIC < 2 (i.e. the best fit models). A cross indicates the variables that were included in the model. These four selected models were best fitted using the Ornstein–Uhlenbeck model of trait evolution as an expected covariance matrix.

BIOSKETCH

Simon Blanchet is a researcher at the Station d'Ecologie Expérimentale du CNRS à Moulis. He conducts researches on the causes and consequences of fish diversity from local to global scales, and on the ecology and evolution of host–parasite interactions. His research interests include the effects of global changes on freshwater ecosystems.

Author contributions: S.B., Y.R., J.A., N.E.M., M.A.R., L.B. and P.M. conceived the ideas; S.B., Y.R. and J.A. analysed the data; S.B., Y.R., J.A., N.E.M., M.A.R., L.B. and P.M. wrote the paper.

Editor: José Alexandre Diniz-Filho