

Phylogenetics to help predict active metabolism

G. GUÉNARD, D. BOISCLAIR, AND P. LEGENDRE

Département de sciences biologiques, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7

Citation: Guénard, G., D. Boisclair, and P. Legendre. 2015. Phylogenetics to help predict active metabolism. *Ecosphere* 6(4):62. <http://dx.doi.org/10.1890/ES14-00479.1>

Abstract. This paper shows how to build predictive models involving phylogenetic information to estimate metabolic traits such as active metabolic costs. Fish swimming cost is often estimated from body mass and swimming speed. The parameters of the relationships between these variables and swimming cost vary among species because each species has its own morphology and physiology. It is now widely recognized that traits are phylogenetically structured. Using new statistical approaches, it is possible to both correct swimming cost models for statistical phylogenetic non-independence and use the inherent phylogenetic signal to improve models. With these models one can extend, to a larger set of species, empirical knowledge about traits that are difficult to obtain; swimming cost is one such trait.

Swimming cost accounts for a large and variable component of a fish energy budget, yet models have only been developed from observations performed on a few species, thereby constraining the scope of bioenergetic models. Here, we propose a method where body mass and swimming speed are used together with phylogeny to predict swimming cost. The resulting model explained a large proportion of the variation (90%) in the forced swimming cost of 16 fish species submitted to forced swimming experiments. We also compared phylogenetically-explicit predictions for forced swimming experiments with experimental results of routine swimming for five species, among which one was not used to build the model. Results confirmed that forced swimming underestimates the cost of unsteady swimming. The phylogenetic modeling could be used to estimate other variables of interest in bioenergetic studies.

Key words: activity; forced swimming; multiple-species modeling; phylogenetic modeling; routine swimming; swimming cost.

Received 2 December 2014; revised 27 January 2015; accepted 29 January 2015; **published** 22 April 2015. Corresponding Editor: D. P. C. Peters.

Copyright: © 2015 Guénard et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. <http://creativecommons.org/licenses/by/3.0/>

† **E-mail:** guillaume.guenard@gmail.com

INTRODUCTION

Active metabolism is an important and variable component of energy budgets (Boisclair and Leggett 1989, Boisclair and Rasmussen 1996, Aubin-Horth et al. 1999). The in situ estimation of fish swimming cost involves models to circumvent the technical difficulty of measuring outcomes of metabolic processes (e.g., oxygen consumption) on individual organisms swimming in the wild. These models are generally log-linear functions involving variables such as body

mass and swimming speed whose implementation requires data obtained from respirometry experiments. These experiments consist in measuring oxygen consumption by depletion inside a sealed, oxygen-tight, vessel called a respirometer (Blazka 1958). To obtain a swimming cost model, suites of such experimental assays have to be performed with specimens having various masses and swimming at different speeds.

When implementing swimming cost models using data from multiple species, observations cannot be considered independent as the species

share different degrees of common ancestry to one another (Felsenstein 1985). Also, the parameters of swimming cost models may vary among species as a consequence of their particular morphological and physiological traits, and the differences among values of the parameters may depend on the phylogenetic distances among the species. Therefore, a swimming cost model estimated using data from one species may not give a good estimate of the swimming cost of another species whereas a general swimming cost model estimated using multiple species data will have biased parameter values if the latter are not estimated while taking the lack of statistical independence among observations into account.

The effect of body mass and swimming speed on swimming cost is fairly simple to understand. A heavier specimen experiences larger hydrodynamic drag than a smaller one as a consequence of its larger cross-section and length. The consequences of swimming speed for a fish are the same as for any object moving in a viscous environment: the organism produces more drag at high than at low speed. Greater drag implies that more work has to be deployed to travel a given distance at a given speed. While experimentally inducing variation in body mass is easily done by selecting individuals of various sizes, the representation of the effect of swimming speed involves many more assumptions regarding fish locomotory behavior. Under the assumption that the conditions in their habitat allow fish to swim steadily, the experiments exploit the fact that many fishes are positively rheotactic (i.e., they will turn to face into an oncoming current) and consist in make them hold a position against a steady flow of water. We will refer to the latter as “forced swimming” experiments. Whether or not (and under what circumstances, if any) steady swimming is a reasonable assumption for modeling real swimming cost has been questioned (Webb 1991). When swimming is unsteady, costs associated with acceleration and turns are substantial and have to be accounted for in experiments. One such approach consists in granting fish with enough freedom of movement to adopt swimming patterns resembling those they would feature under natural conditions (Tang and Boisclair 1993). We will refer to that as “routine swimming” experi-

ments. Response to given experimental conditions designed to induce spontaneous swimming movements may vary among species as a consequence of their particular behavioral traits.

Species traits (e.g., morphological, physiological, behavioral) are the outcome of evolutionary processes and are therefore structured with respects to phylogeny (Felsenstein 1985). Methods now exist to use the among-species phylogenetic patterns of trait variation purposefully, to predict trait values (Phylogenetic Eigenvector Maps [PEM]; Guénard et al. 2013). Knowing the phylogeny and trait values for a set of species, one can quantify the manner whereby a trait evolved using phylogenetic eigenfunctions. The eigenfunctions describe the candidate patterns of among-species trait variation across the phylogeny and are used to model that variation. The resulting model can be used to predict trait values for species where they are unknown. Phylogenetic modeling is especially useful when trait values are difficult, time consuming, or expensive to obtain. When applied with proper care, the PEM method allows researchers and practitioners to maximize the use of information acquired at great expense.

Swimming cost measurements are expensive to obtain and fish form a very diverse group that has evolved over 530 million years. Hence, multiple species phylogenetic modeling appears to be a pragmatic alternative to building separate swimming cost models for every species of interest. The method also helps solve the problem for rare or endangered species, on which experimentation would likely be forbidden, or species that are costly to obtain (e.g., abyssal). It allows one to build synthetic models involving multiple species while implicitly correcting for the parameter estimation bias associated with using data from multiple species that have not evolved independently of one another.

The purpose of the present study is to illustrate how swimming cost information from published studies, combined with that obtained from DNA sequences and phylogenetic estimation methods, can be used to build multi-species, phylogenetically-explicit and unbiased swimming cost models.

METHODS

Data sources

The data set used to develop the multi-species swimming cost model was collected by Boisclair and Tang (1993) from 24 published studies. In addition to forced and routine swimming experiments, the original data set includes a third type called “directed swimming” experiments, in which fish were trained to follow a target area moving above the respirometer (Muir et al. 1965, Wohlschlag and Cameron 1967, Muir and Niimi 1972). For the sake of simplicity, and since the directed swimming experiments were performed on only two species (*Kuhlia sandvicensis*: Aholehole; Muir et al. 1965, Muir and Niimi 1972 and *Lagodon rhomboides*: Pinfish; Wohlschlag and Cameron 1967), we decided to focus our analysis on the species with forced swimming cost data (i.e., the type of experiment with the most abundant data), excluding data from directed swimming experiments, and to use the routine swimming data as an illustrative application of phylogenetic swimming cost models. For the latter application, we also added data from two studies on the brook trout (*Salvelinus fontinalis*; Tang and Boisclair 1995, Tang et al. 2000). Each entry of the data set synthesizes the result of a swimming cost experiment and features fish body mass (g fresh mass), swimming speed (cm s^{-1}), the type of experiment (forced or routine swimming) and the net cost of swimming in terms of oxygen consumption ($\text{mgO}_2 \text{ h}^{-1}$). The latter was obtained by subtracting standard metabolic rates and osmoregulation costs from gross rates of oxygen consumption measured in respirometers (see Boisclair and Tang 1993 for details).

Phylogenetic analysis

We used DNA sequences downloaded from the U.S. National Center for Biotechnology Information's GenBank website to estimate the fish phylogeny (<http://www.ncbi.nlm.nih.gov/genbank/>; database: Nucleotide). Sequences encompass the mitochondrial genomes and the nuclear 28S, 18S, and 5.8S rRNA transcripts and internal transcribed spacers (ITS 1 and ITS 2). We focused on complete sequences as much as possible and, otherwise, took the longest partial sequences available. When data on our species

were not available, we borrowed sequences from related species in the same genus or family, unless other species in the same genus or family were present in our data set. Multiple DNA sequences alignments were performed separately for each gene using program Muscle version 3.8.31 (Edgar 2004; <http://www.drive5.com/muscle>). We estimated phylogenies using a maximum likelihood method (Felsenstein 1981, Felsenstein and Churchill 1996; program: fdnaml from package EMBOSS version 6.3.1: Rice et al. 2000) on the super-alignment obtained by concatenating sequences for all genes. Species were ordered in decreasing order of their number of aligned base pairs to help the tree construction algorithm by including species with sparse DNA data into trees after the species for which more complete DNA data was available.

Modeling approach

The modeling approach used was similar to that described in Guénard et al. (2011). It involved the computation of a multiple regression model in which swimming cost descriptors were used together with phylogenetic eigenvectors. Phylogenetic eigenvectors are linearly independent variables obtained from the structure of the phylogenetic tree (topology and branch lengths; Diniz-Filho et al. 1998, Desdevisse et al. 2003, Guénard et al. 2013); they allow one to represent phylogenetic patterns of variation, for instance in multiple regression. We preferred to use phylogenetic eigenvectors rather than other alternatives (e.g., Martins and Hansen 1997, Garland and Ives 2000, Bruggeman et al. 2009) because of the relative simplicity of that approach and its ability to assess interaction between the phylogeny and swimming cost descriptors. Swimming cost descriptors include body mass (M) and swimming speed (V). The original phylogenetic tree had a single tip per species and therefore had to be expanded because multiple experiments had been performed for each species. We obtained an expanded tree having a tip for each experimental result by binding the root of a star tree having branch lengths of 0 and as many tips as the number of experimental results to each tip of the species tree. This approach assumes that phylogenetic patterns of variation at the intra-specific level (i.e., among the different experimental specimens

of a same species) are negligible.

The swimming cost model was described by the following linear equation:

$$\begin{aligned} \log_{10} C_{i,sp,M,V} = & b_0 + b_M \log_{10} M_i + b_V \log_{10} V_i \\ & + \sum_j b_{ij} u_{sp,j} + \sum_k b_{M \times u_k} u_{sp,k} M_i \\ & + \sum_l b_{V \times u_l} u_{sp,l} V_i \\ & + \varepsilon_i \end{aligned} \quad (1)$$

where $C_{i,type,M,V}$ is the swimming cost of specimen i with mass M (g wet mass) belonging to species sp , forced to swim at speed V (cm s^{-1}). Coefficients b_{\dots} are the regression parameters associated with the different components of the model (b_0 is the intercept of the model; \times in subscripts denote parameters associated with interactions), $u_{sp,\dots}$ are the loadings of species sp on eigenvectors, j are the eigenvectors involved in modeling the main effect of the phylogeny, k are the eigenvectors involved in modeling the interaction between body mass and phylogeny, l are the eigenvectors involved in modeling the interaction between swimming speed and phylogeny, and ε_i is a normal deviate with mean 0 and variance σ_ε .

To avoid overfitting and ensure that the resulting model be as general as possible, we regularized the multiple regression model using ridge regression (see Legendre and Legendre [2012] for a description of the method). We estimated the ridge regression penalty factor that gave the best out-of-the-sample predictions, which we assessed using jackknife cross-validation, removing one species at a time, each time estimating the model with the remaining species and making predictions for the removed species using the resulting model. That procedure was repeated for all species in the data set. Comparison between the observed and predicted forced swimming costs was performed using the cross-validation R-squared calculated as follows:

$$R_{cv}^2 = 1 - \frac{(n-1) \sum_{i=1}^n (y_i - y_{pred,i})^2}{n \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

where y_i and $y_{pred,i}$ are the observed and predicted values of swimming cost, respectively, for individual i and \bar{y} is the mean observed swimming cost. The R_{cv}^2 takes the value 1 for a perfect prediction and the value 0 when the mean square prediction error equals the mean square deviation from the mean (i.e., the observed

variance), in which case the model is no better than to simply take the mean. It is therefore interpreted in the same way as an adjusted coefficient of determination (R_{adj}^2) in regression analysis. The R_{cv}^2 has no negative boundary, as there is no theoretical limit to how badly a model may predict reality. The penalty factor was estimated as that maximizing R_{cv}^2 by gradient descent using the bound constrained (to allow only for positive penalties) L-BFGS-B algorithm (Byrd et al. 1995). In addition to the prediction coefficient presented above, we assessed the accuracy of the model for individual predictions i by computing the deviation factor, dev_i , obtained as follows:

$$dev_i = \begin{cases} \frac{y_{pred,i}}{y_i} - 1 & y_{pred,i} > y_i \\ 0 & y_{pred,i} = y_i \\ 1 - \frac{y_i}{y_{pred,i}} & y_{pred,i} < y_i \end{cases} \quad (3)$$

That factor has no measurement units and indicates the number of times models overestimate ($dev_i > 0$) or underestimate ($dev_i < 0$) the observed swimming cost. Values close to 0 indicate accurate predictions.

All calculations other than sequence alignment and tree estimation were performed using the R language for statistical computing (R Development Core Team 2014) and R packages ape (Paradis et al. 2004), MASS (Venables and Ripley 2002), and MPSEM (Guénard et al. 2013).

RESULTS

Data mining

The data set we obtained comprises 976 experiments on 17 fish species reported in 26 studies (Table 1). These experiments were performed on specimen having mass ranging from 1.1 to 1432 g and swimming at speeds ranging from 0.314 to 143 cm s^{-1} ; their swimming costs ranged from 2.7 $\mu\text{gO}_2 \text{ h}^{-1}$ to 923.7 $\text{mgO}_2 \text{ h}^{-1}$. Data were sparse for most species. The widest range of body mass and swimming speeds was observed for the Sockeye salmon (*Oncorhynchus nerka*; body mass: 3.38–1432 g, swimming speed: 0.314–143 cm s^{-1}) while the second largest such range was obtained for the Rainbow trout (*Oncorhynchus mykiss*; body mass: 27.0–471 g, swimming speed: 10.2–103 cm s^{-1}), both pertain-

Table 1. Studies from which swimming cost data were obtained.

Species (common name)	Reference	Type	<i>n</i>
<i>Carassius auratus</i> (Goldfish)	Smit (1965)	Routine	54
		Forced	8
	Kutty (1968)	Forced	17
	Smit et al. (1971)	Forced	64
<i>Centropomus undecimalis</i> (Snook)	Pérez-Pinzón and Lutz (1991)	Forced	15
<i>Kuhlia sandvicensis</i> (Aholehole)	Muir et al. (1965)	Routine	11
	Muir and Niimi (1972)	Forced	54
<i>Limanda limanda</i> (Common dab)	Duthie (1982)	Forced	45
<i>Liza macrolepis</i> (Mullet)	Kutty (1969)	Routine	3
		Forced	8
<i>Melanogrammus aeglefinus</i> (Haddock)	Tyler (1969)	Forced	25
<i>Micropterus salmoides</i> (Largemouth bass)	Beamish (1970)	Forced	32
<i>Microstomus kitt</i> (Lemon sole)	Duthie (1982)	Forced	31
<i>Oncorhynchus mykiss</i> (Rainbow trout)	Rao (1968)	Forced	75
	Kutty (1968)	Forced	11
	Webb (1971)	Forced	31
	Weatherley et al. (1982)	Routine	8
		Forced	26
	Wood and Randall (1973)	Forced	21
	Brett (1964)	Forced	108
<i>Oncorhynchus nerka</i> (Sockeye salmon)	Brett (1965)	Forced	21
	Brett and Glass (1973)	Forced	44
	Febry and Lutz (1987)	Forced	12
	Farmer and Beamish (1969)	Forced	76
<i>Oreochromis mossambicus</i> (Mozambique tilapia)	Duthie (1982)	Forced	48
<i>Oreochromis niloticus</i> (Nile tilapia)	Priede and Holliday (1980)	Forced	33
<i>Platichthys flesus</i> (European flounder)	Beamish (1990a)	Forced	12
<i>Pleuronectes platessa</i> (European plaice)	Tang and Boisclair (1995)	Routine	21
<i>Salvelinus namaycush</i> (Lake trout)	Tang et al. (2000)	Routine	24
<i>Salvelinus fontinalis</i> (Brook trout)	Tarby (1981)	Forced	10
	Beamish (1990a)	Forced	6
	Beamish (1990b)	Forced	22

ing to the same genus. The other species featured narrower ranges of swimming speeds, generally spanning factors of 10 or less, and even much narrower ranges of body masses: three species were only represented by a single value of body mass whereas most other species were represented by 2 or 3 values spanning factors of 2–5. Among these data, we found 855 values, representing 16 species, from forced swimming experiments (the Brook trout, *Salvelinus fontinalis*, had no forced swimming cost value), that we used to build the phylogenetic forced swimming cost model. Comparisons between phylogenetically-explicit forced swimming predictions and routine swimming observations were performed using the remaining 121 values (representing five species) from routine swimming experiments.

We found three genes whose sequences were available for all 17 species: cytochrome oxidase subunit I (9 complete, 8 partial), cytochrome b (13 complete, 4 partial), and DNA coding for the 16S ribosomal RNA (11 complete, 6 incomplete). Complete ribosomal genomes were available for

nine species whereas the sequence coding for the small subunit (12S) ribosomal RNA was available for 15 species (9 complete, 6 partial). Sequences for nuclear ribosomal RNA and associated intergenic spacers were sparse: 18S RNA was available for eight species while the sequence for intergenic spacer 2 (that between 5.8S and 28S rRNA) was only available for a single species and therefore could not be used to estimate a phylogeny. Species shared a minimum of 135 and a maximum of 4457 base pairs with one another (median value: 849 base pairs).

Phylogenetic analysis

We rooted the estimated phylogenetic tree on infra-class Teleostei (Fig. 1). The bi-partition occurs between the ostariophysians, for which the Goldfish (*Carassius auratus*, family: Cyprinidae) is the sole representative, and super-orders Protacanthopterygii and Acanthopterygii. The separation of the latter super-orders immediately follows, with order Protacanthopterygii represented by the four salmonids. Among the

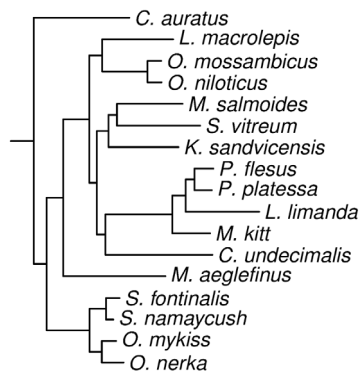


Fig. 1. The phylogenetic tree estimated by maximum likelihood using a selection of mitochondrial and nuclear ribosomal DNA sequences obtained from the NCBI GenBank.

salmonids, the tree shows the expected separation between genera *Salvelinus* (Lake trout, *S. namaycush*, Brook trout, *S. fontinalis*) and *Oncorhynchus* (Rainbow trout, *O. mykiss*, and Sockeye salmon, *O. nerka*). The first bi-partition in the acanthopterygian subtree separates order Gadiformes (Haddock, *Melanogrammus aeglefinus*) from the remaining species (orders Perciformes and Pleuronectiformes). Then follows a second separation between a first group composed of families Sparidae (represented by the Mullet, *Liza macrolepis*) and Cichlidae (represented by genus *Oreochromis*: the Nile tilapia, *O. niloticus*, and Mozambique tilapia, *O. mossambicus*); these two families belong to order Perciformes, leaving a second group composed of the remaining species. As expected from taxonomy, the groups split according to the families (the Mullet splitting from the two Tilapias). The remaining species split into a first group composed of the representatives of families Centrarchidae (Largemouth bass, *Micropterus salmoides*), Percidae (Walleye, *Sander vitreum*), and Kuhlidae (Aholehole, *Kuhlia sandvicensis*) and a second group composed of the Snook (*Centropomus undecimalis*, family Centropomidae) and members of order Pleuronectiformes. In the first group, the Aholehole splits from a group composed of the Largemouth bass and the Walleye. In the second group, the Snook splits from the Pleuronectiformes. The Pleuronectiformes subtree separates the Lemon sole (*Microstomus kitt*) from the remaining species and then the Common dab

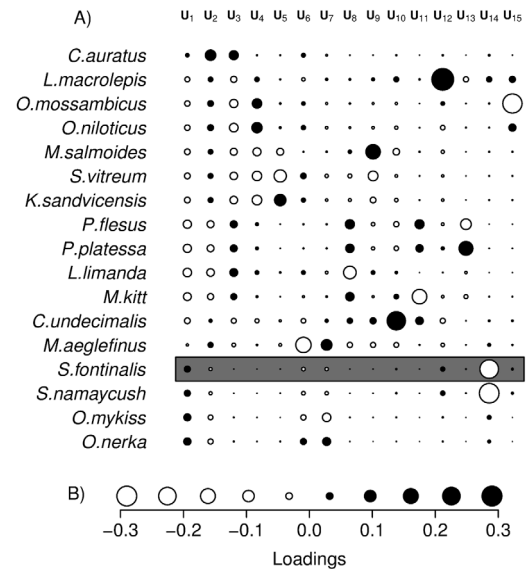


Fig. 2. The phylogenetic eigenvectors that were selected to model swimming cost: (A) Species loadings on the eigenfunctions showing the patterns of phylogenetic variation that they represent, (B) legend associating species loadings to the circle sizes (absolute values proportional to the surface area of bubbles) and colors (black and white marker: negative and positive species loadings, respectively). Prediction scores for the Brook trout (*Salvelinus fontinalis*), which is used exclusively to make predictions, are highlighted in gray.

(*Limanda limanda*) from the European flounder (*Platichthys flesus*) and the European plaice (*Pleuronectes platessa*). While the estimated tree agrees with taxonomy most of the time, it suggests that order Perciformes (and, possibly, suborder Percoidei) is paraphyletic since order Pleuronectiformes is embedded into it.

Phylogenetic model

We estimated 15 phylogenetic eigenvectors (named u_1 – u_{15}) from the sub-tree of the 16 species for which forced swimming cost data were available; then we estimated the prediction scores for the Rainbow trout, *O. mykiss*, using its phylogenetic location with respect to the other species (Fig. 2). The phylogenetic eigenvectors represent particular patterns of phylogenetic variation that were selected to intervene in the model. It is noteworthy that the eigenvectors describe not only the phylogenetic patterns

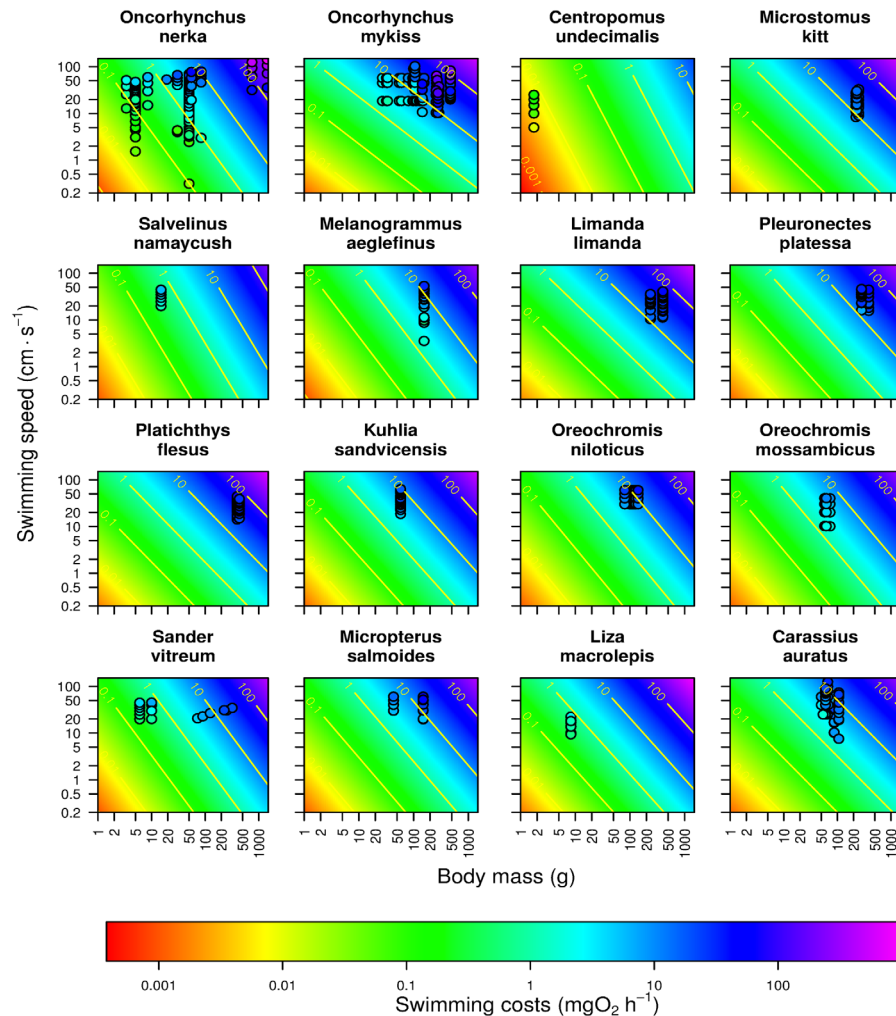


Fig. 3. Swimming cost as a function of swimming speed and body mass: the observed values correspond to the background color of the markers whereas the values fitted by the phylogenetic model for the range of observed swimming speeds and body masses correspond to the background color of the plotting area.

associated to the topology but also the branch length structure of the tree.

The estimated ridge regression penalty factor was 38.66, with an R_{cv}^2 of 0.81 obtained from the set of cross-validation trials that maximized predictive power. The R_{adj}^2 (adjusted for the number of parameters in the model) of the phylogenetic forced swimming cost model was 0.90 whereas the R_{adj}^2 for the partial model, with only the (\log_{10}) body mass and the (\log_{10}) swimming speed, was 0.85 (Figs. 3, 4A).

The most extreme model deviations was an overestimation ($dev = 24.4$) for an experiment with a 62 g Goldfish swimming at 25 cm s^{-1} and

that was reported by Smit et al. (1971) to have had an oxygen consumption rate of 0.193 mg h^{-1} , with the model prediction being 4.90 mg h^{-1} . The second most extreme value was also an overestimation ($dev = 21.8$) of swimming cost for an experiment with a 50 g Sockeye salmon swimming at 0.314 cm s^{-1} , which was reported by Brett (1964) to have consumed oxygen at a rate of $2.7 \text{ } \mu\text{g h}^{-1}$ (the lowest value in the whole data set) while the prediction was $61.5 \text{ } \mu\text{g h}^{-1}$. All other predictions (853 over 855 = 99.7%) were within a deviation factor of ± 10 , and about three quarters (73.5%) within a deviation factor of ± 1 . On a specific basis, the median absolute deviation

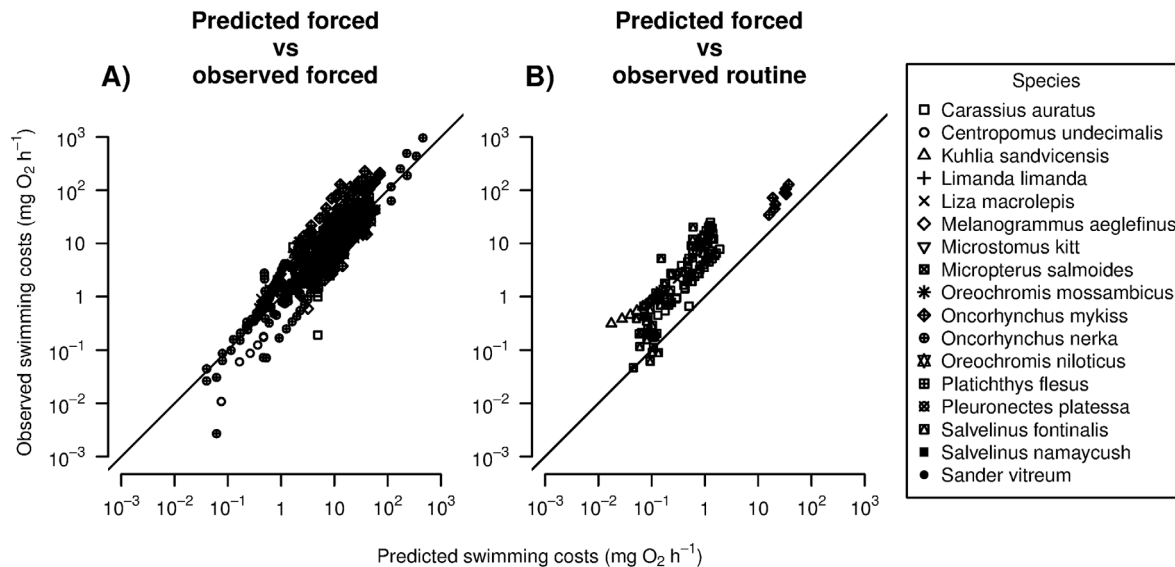


Fig. 4. Observed and predicted swimming costs. (A) Forced swimming cost predicted by the phylogenetic forced swimming cost models following leave-one-species-out cross-validation trials, (B) Routine swimming costs predicted using the phylogenetic forced swimming cost model. The diagonal line is the 1:1 line on which correct predictions would fall.

factor ranged from 0.25 (Largemouth bass) to 1.91 (Snook) with a mean of 0.67.

Comparison to routine swimming

Comparison between the specific estimates of forced swimming cost and routine swimming cost indicates that most values were found at a relatively constant distance above the 1:1 log-log line and that forced swimming cost underestimated routine swimming (Fig. 4B) cost by a deviation factor of -6.4 on average (*K. sandvicensis*: -9.5 , *C. auratus*: -7.2 , *L. macrolepis*: -7.0 , *S. fontinalis*: -5.6 , and *O. mykiss*: -1.9). That finding agrees with that of Boisclair and Tang (1993) whose results suggest that forced swimming costs underestimate routine swimming costs by factors ranging from 6.4 to 14.0.

DISCUSSION

In the present study, we showed that phylogenetic modeling can be used to enhance our ability to predict elements of metabolism. We illustrated that possibility using fish swimming cost but adaptations of the same approach may be used to predict other relevant quantities for other types of organisms. For instance, paramete-

ters of models predicting standard or basal metabolic rate, urinary and fecal losses, maximum consumption rate, or swimming capacity, may also vary among species and may possibly feature phylogenetic structures that may be estimated from phylogeny. The number of species used here (17: 16 to build the forced swimming cost model and 1 used exclusively for swimming cost comparisons) is modest in comparison to the number of extant teleost species (≈ 27000), showing that phylogenetic modeling is achievable with a fairly small number of species.

It is worth mentioning that the quality of the available data set, in terms of the coverage of body mass values and swimming speeds for the different species, was somewhat remote from the ideal. The ideal data set to reliably discriminate the actual effect of body mass, swimming speed, and phylogeny would have consisted in a set of standard body masses repeated for a standard set of swimming speeds, with evenly replicated body mass–swimming speed combinations, repeated for all species. Such a stringent sampling scenario would have insured the absence of collinearity among descriptors, more especially between the experimental conditions (body

masses and swimming speeds) and phylogeny. In the present study, we found that body mass was strongly related to phylogeny ($R^2_{adj} = 0.67$, $F_{15,839} = 115.5$, $p < 0.0001$), as was, though to a lesser extent, swimming speed ($R^2_{adj} = 0.17$, $F_{15,839} = 12.63$, $p < 0.0001$); body mass and swimming speed were only slightly correlated ($R^2_{adj} = 0.025$, $F_{1,839} = 23.00$, $p < 0.0001$). Although body mass at a particular life stage is a trait that we expect to be related to phylogeny, in experimental settings this factor could be artificially controlled (i.e., by selecting fish of particular sizes) to reduce the collinearity between the phylogenetic eigenvectors and body mass. In the present study, we used ridge regression to mitigate that issue as much as numerically possible, yet it was not possible to do so entirely given the sparse coverage of different body masses (Fig. 3). Because of the irregular manner in which body mass and swimming speed are spread among the species, it is likely that their joint apparent contribution in the model ($R^2_{adj} = 0.85$) is exaggerated. A greater fraction of that variation could well have been associated to phylogenetic eigenvectors, had the sampling been more regularly distributed. We regard that situation as being a likely limitation of exploiting meta-analytical data sets stemming from the fact that the original purpose of the data was indeed different from that of the meta-analysis. By opposition, the original studies were focusing on a particular range of sizes and swimming speeds of a single (or two) species.

Phylogenetic bioenergetic models could be implemented with relatively little information insofar as that information is correctly targeted at representing the broadest possible range of conditions (e.g., body mass, swimming speed, temperature) and spanning broad phylogenies, including, for instance, other Osteichthyes such as species of sturgeons in the case of fish models. For example, one could define four different masses, e.g., 0.01, 0.1, 0.5 and 1 times the mean adult size; and four different speeds, e.g., 0.1, 0.5, 1 and 2 body length s^{-1} . Repeating each treatment on different specimens, say three, and performing experiments with, say, 20 species would amount to a sample size of 960, which is similar to the one we had in this study. Such a data set would, however, form a better design for model construction: the orthogonality of the

controlled variables, body mass and swimming speed, in the model would insure that their contributions to the model would not be artificially distorted by sampling-related issues but, instead, be the result of the species' own features (Guénard et al. 2014). Similar rationales would apply to phylogenetic modeling of standard metabolic rate, swimming capacity, or maximum consumption rate, with the standard set of swimming speeds being replaced by a standard set of temperatures.

The comparison of forced and routine swimming cost, using this time an analysis taking phylogeny into account, further confirmed the conclusion of Boisclair and Tang (1993) that forced swimming cost models underestimate fish routine swimming cost by a large extent. Routine swimming cost involves spontaneous accelerations and turns that are standard features of the locomotory behavior of most fish species (e.g., Webb 1991). Most of the information available on fish swimming costs comes from forced swimming cost experiments. Nevertheless, it is important for users of swimming cost models to experimentally confirm that the context of their study enables their target fishes to swim steadily (i.e., at a constant speed, without accelerations and turns). A failure of the experiment to meet that assumption would lead users astray, by a large margin, about the actual size of swimming costs.

Provided that sufficient effort is deployed to collect the necessary data, we regard phylogenetic modeling as a keystone method that can allow biologists to apply bioenergetic modeling in practice to species other than the well-studied ones and in contexts where multiple species are often involved. With phylogenetic models, the great efforts that have to be deployed to acquire information on metabolic rates, which traditionally involves performing numerous experiments, can be optimized, since variation in metabolic rates among the species of a group can be explicitly represented by phylogenetic eigenvectors used in an explanatory manner. For instance, a study improving our knowledge of fish swimming costs by adding two more species to the existing 16 species forced swimming cost model may be regarded as a more relevant addition to knowledge than a study providing two single-species models because the former

would give us the ability to predict swimming costs for more than two species. We hope phylogenetic modeling will help scientists unfold the full potential of bioenergetic modeling in practice by increasing the “species scope” of the models predicting the different compartments of the energy budget.

ACKNOWLEDGMENTS

We are thankful to the many people who helped us during the elaboration of the present study by providing useful comments and suggestions. G. Guénard was supported by a Strategic Network Grant to D. Boisclair, P. Legendre, and 11 other scientists (NETGP/370899-2008), and an Individual Discovery grant from the Natural Sciences and Engineering Council of Canada (NSERC) to P. Legendre. That paper is an original contribution of NSERC Hydronet.

LITERATURE CITED

- Aubin-Horth, N., J. Gingras, and D. Boisclair. 1999. Comparison of activity rates of 1+ yellow perch (*Perca flavescens*) from population of contrasting growth rates using underwater video observations. *Canadian Journal of Fisheries and Aquatic Sciences* 56:1122–1132.
- Beamish, F. W. H. 1970. Oxygen consumption of largemouth bass, *Micropterus salmoides*, in relation to swimming speed and temperature. *Canadian Journal of Zoology* 48:1221–1228.
- Beamish, F. W. H. 1990a. Metabolism and photoperiod in juvenile lake charr and walleye. *Environmental Biology of Fishes* 29:201–207.
- Beamish, F. W. H. 1990b. Swimming metabolism and temperature in juvenile walleye, *Stizostedion vitreum vitreum*. *Environmental Biology of Fishes* 27:309–314.
- Blazka, P. 1958. The anaerobic metabolism in fish. *Physiological Zoology* 31:117–128.
- Boisclair, D., and W. C. Leggett. 1989. The importance of activity in bioenergetic models applied to actively foraging fishes. *Canadian Journal of Fisheries and Aquatic Sciences* 46:1859–1867.
- Boisclair, D., and J. B. Rasmussen. 1996. Empirical analysis of the influence of environmental variables associated with lake eutrophication on perch growth, consumption, and activity rate. *Annales Zoologici Fennici* 33:507–515.
- Boisclair, D., and M. Tang. 1993. Empirical analysis of the influence of swimming pattern on the net energetic cost of swimming in fishes. *Journal of Fish Biology* 42:169–183.
- Brett, J. R. 1964. The respiratory metabolism and swimming performances of young sockeye salmon. *Journal of the Fisheries Research Board of Canada* 21:1183–1226.
- Brett, J. R. 1965. The relation of size to rate of oxygen consumption and sustained swimming speed of sockeye salmon (*Oncorhynchus nerka*). *Journal of the Fisheries Research Board of Canada* 22:1491–1501.
- Brett, J. R., and N. R. Glass. 1973. Metabolic rates and critical swimming speeds of sockeye salmon (*Oncorhynchus nerka*) in relation to size and temperature. *Journal of the Fisheries Research Board of Canada* 30:379–387.
- Bruggeman, J., J. Heringa, and B. W. Brandt. 2009. Phylopars: estimation of missing parameter values using phylogeny. *Nucleic Acid Research* 37(supplement 2):W179–W184.
- Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16:1190–1208.
- Desdève, Y., P. Legendre, L. Azouzi, and S. Morand. 2003. Quantifying phylogenetically structured environmental variation. *Evolution* 57:2647–2652.
- Diniz-Filho, J. A. F., C. E. R. de Sant’Ana, and L. M. Bini. 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution* 52(5):1247–1262.
- Duthie, G. G. 1982. The respiratory metabolism of temperature-adapted flatfish at rest and during swimming activity and the use of anaerobic metabolism at moderate swimming speeds. *Journal of Experimental Biology* 97:359–373.
- Edgar, R. 2004. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research* 32:1792–1797.
- Farmer, G. J., and F. W. H. Beamish. 1969. Oxygen consumption of tilapia nilotica in relation to swimming speed and salinity. *Journal of the Fisheries Research Board of Canada* 26:2807–2821.
- Febry, R., and P. L. Lutz. 1987. Energy partitioning in fish: The activity related cost of osmoregulation in a euryhaline cichlid. *Journal of Experimental Biology* 128:63–85.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Ecology* 17:368–376.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Felsenstein, J., and G. A. Churchill. 1996. A hidden Markov model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution* 13:93–104.
- Garland, T. J., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist* 155:346–364.
- Guénard, G., P. Legendre, and P. Peres-Neto. 2013. Phylogenetic eigenvector maps (PEM): a frame-

- work to model and predict species traits. *Methods in Ecology and Evolution* 4:1120–1131.
- Guénard, G., P. C. von der Ohe, D. de Zwart, P. Legendre, and S. Lek. 2011. Using phylogenetic information to predict species tolerances to toxic chemicals. *Ecological Applications* 21:3178–3190.
- Guénard, G., P. C. von der Ohe, S. C. Walker, S. Lek, and P. Legendre. 2014. Using phylogenetic information and chemical properties to predict species tolerances to pesticides. *Proceedings of the Royal Society B* 281:20133239.
- Kutty, M. N. 1968. Respiratory quotients in goldfish and rainbow trout. *Journal of the Fisheries Research Board of Canada* 25:1689–1725.
- Kutty, M. N. 1969. Oxygen consumption of mullet *Liza macroleptis* with special reference to swimming velocity. *Marine Biology* 4:239–242.
- Legendre, P., and L. Legendre. 2012. Numerical ecology. Third English edition. Elsevier Science, Amsterdam, The Netherlands.
- Martins, E., and T. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149:646–667.
- Muir, B. S., G. J. Nelson, and K. W. Bridges. 1965. A method for measuring swimming speed in oxygen consumption studies on aholehole *Kuhlia sandvicensis*. *Transactions of the American Fisheries Society* 94:378–382.
- Muir, B. S., and A. J. Niimi. 1972. Oxygen consumption of the euryhaline fish aholehole (*Kuhlia sandvicensis*) with reference to salinity, swimming, and food consumption. *Journal of the Fisheries Research Board of Canada* 29:67–77.
- Paradis, E., J. Claude, and K. Strimmer. 2004. Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pérez-Pinzón, M. A., and P. L. Lutz. 1991. Activity related cost of osmoregulation in the juvenile snook (*Centropomus undecimalis*). *Bulletin of Marine Science* 48:58–66.
- Priede, I. G., and F. G. T. Holliday. 1980. The use of a new tilting tunnel respirometer to investigate some aspects of metabolism and swimming activity of the plaice (*Pleuronectes platessa* L.). *Journal of Experimental Biology* 85:295–309.
- Rao, G. M. M. 1968. Oxygen consumption of rainbow trout (*Salmo gairdneri*) in relation to activity and salinity. *Canadian Journal of Zoology* 46:781–785.
- R Development Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rice, P., I. Longden, and A. Bleasby. 2000. Emboss: The european molecular biology open software suite. *Trends in Genetics* 16:276–277.
- Smit, H. 1965. Some experiments on oxygen consumption of goldfish (*Carassius auratus* L.) in relation to swimming speed. *Canadian Journal of Zoology* 43:623–633.
- Smit, H., J. M. Amelink-Koutstaal, J. Vijverberg, and J. C. von Vaupel-Klein. 1971. Oxygen consumption and efficiency of swimming goldfish. *Comparative Biochemistry and Physiology A* 39:1–28.
- Tang, M., and D. Boisclair. 1993. The influence of the size of enclosures on the swimming characteristics of juvenile brook trout. *Canadian Journal of Fisheries and Aquatic Sciences* 50:1786–1793.
- Tang, M., and D. Boisclair. 1995. Relationship between respiration rate of juvenile brook trout (*Salvelinus fontinalis*), water temperature and swimming characteristics. *Canadian Journal of Fisheries and Aquatic Sciences* 52:2138–2145.
- Tang, M., D. Boisclair, C. Ménard, and J. Downing. 2000. Influence of weight, swimming characteristics, and water temperature on the cost of swimming in brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences* 57:1482–1488.
- Tarby, M. J. 1981. Metabolic expenditure of waleye (*Stizostedion vitreum vitreum*) as determined by rate of oxygen consumption. *Canadian Journal of Zoology* 59:882–889.
- Tyler, P. 1969. Relationship between oxygen consumption and swimming speed in the haddock, *Melanogrammus aeglefinus*. *Nature* 221:274–275.
- Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Fourth edition. Springer-Verlag, New York, New York, USA.
- Weatherley, A. H., S. C. Rogers, D. G. Pincock, and J. R. Patch. 1982. Oxygen consumption of active rainbow trout, *Salmo gairdneri* Richardson, derived from electromyograms obtained by radiotelemetry. *Journal of Fish Biology* 20:479–489.
- Webb, P. W. 1971. The swimming energetics of trout: II. oxygen consumption and swimming efficiency. *Journal of Experimental Biology* 55:521–540.
- Webb, P. W. 1991. Composition and mechanics of routine swimming of rainbow trout, *Oncorhynchus mykiss*. *Canadian Journal of Fisheries and Aquatic Sciences* 48:583–590.
- Wohlschlag, D. E., and J. N. Cameron. 1967. Assessment of a low level stress on the respiratory metabolism of the pinfish (*Lagodon rhomboides*). *Contributions in Marine Science* 12:160–171.
- Wood, C. M., and D. J. Randall. 1973. The influence of swimming activity on water balance in the rainbow trout (*Salmo gairdneri*). *Journal of Comparative Physiology* 82:257–276.