1

2

3

4   Article type   : Articles

5

6

7   **Predicting life history parameters for all fishes worldwide**

8

9   Running header: Predicting life history for all fishes

10

11   James T. Thorson[1,*], Stephan B. Munch[2], Jason M. Cope[1], Jin Gao[1,3]

12

13   [1] Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science

14   Center, National Marine Fisheries Service, NOAA, Seattle, WA, USA

15

16   [2] Fish Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries

17   Service, NOAA, 110 Schaffer Road, Santa Cruz, CA, USA;

18

19   [3] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle,

20   WA, USA

21

22   * Corresponding author

23   James.Thorson@noaa.gov

24

25   **Abstract**:

26   Scientists and resource managers need to know life history parameters (e.g., average

27   mortality rate, individual growth rate, maximum length or mass, and timing of maturity) to

28   understand and respond to risks to natural populations and ecosystems. For over one-hundred

29   years, scientists have identified "life history invariants" (LHI) representing pairs of

30 parameters whose ratio is theorized to be constant across species. LHI then promise to allow

31 prediction of many parameters from field-measurements of a few important traits. Using LHI

32 in this way, however, neglects any residual patterns in parameters when making predictions.

33 We therefore apply a multivariate model for eight variables (seven parameters and

34 temperature) in all 32,000 marine fishes, and include taxonomic structure for residuals (with

35 levels for class, order, family, genus, and species). We illustrate that this approach predicts

36 variables probabilistically for taxa with many or few data. We then use this model to resolve

37 three questions regarding life-history parameters in fishes. Specifically we show that: (1) on

38 average there is a 1.24% decrease in the Brody growth coefficient for every 1% increase in

39 maximum size; (2) the ratio of natural mortality rate and growth coefficient is not a LHI but

40 instead varies systematically based on the timing of maturation, where movement along this

41 life-history axis is predictably correlated with species taxonomy; and (3) three variables must

42 be known per species to precisely predict remaining life-history variables. We distribute our

43 predictive model as an R package to allow future life-history predictions for fishes to be

44 conditioned on taxonomy and life-history data for fishes worldwide. This package also

45 contains predictions (and predictive intervals) for mortality, maturity, size, and growth

46 parameters for all described fishes.

47

48 Keywords: Life-history theory; natural mortality; individual growth; meta-analysis;

49 phylogenetic regression; life-history invariant

50 **Introduction**

51 Biological characteristics (e.g., individual growth rates, age at maturation, maximum body

52 length and mass, and adult rates of natural mortality) greatly differ among different animal

53 species. These characteristics (or more precisely, their value averaged among individuals for

54 a species at a given location and time) are sometimes termed "life-history parameters", and

55 these parameters play an important role for understanding and managing populations and

56 ecosystems. In fishes, for example, these life-history parameters are important to determine

57 sustainable harvest rates, estimate harvest rates from changes in average length, and project

58 the likely impact of different management actions (Quinn and Deriso 1999). Unfortunately,

59 some or all of these life-history parameters are unknown or poorly measured for most

60 harvested species. Ideally, these parameters would be estimated at the same time as trends in

61 population abundance and fishery harvest rates, e.g., by using a data-integrated stock

62 assessment model to jointly analyze data from surveys and fisheries (Maunder and Punt

63 2013). However, it is often infeasible (or even impossible) to estimate life-history parameters

64 in a stock assessment model given available data or resources (Magnusson and Hilborn

65 2007).

66   As one example, the average adult rate of natural mortality is perhaps the most difficult-

67 to-estimate life history parameter. Estimating natural mortality using mark-recapture

68 sampling, cohort analysis, multispecies models of predation rates, or data-integrated stock

69 assessment models are each expensive, resource-intensive, and often laden with modeling

70 assumptions. As a consequence, fisheries scientists since the 1950s have sought alternative

71 approaches to informing natural mortality for a given species. One common approach is to

72 infer a likely rate of natural mortality from the value of correlated life-history parameters. A

73 recent review identified 29 different methods for estimating natural mortality from limited

74 data, and concluded that many of these methods fail for at least some species because they

75 were derived based on small samples with limited taxonomic breadth (Kenchington 2014).

76 As an alternative, authors have simultaneously estimated relationships among multiple life-

77 history parameters, but these studies have generally involved small groups of closely-related

78 species rather than fishes in general (Nadon and Ault 2016).

79   Fortunately, life-history variables (including estimates of life-history parameters from

80 field-measurements as well as environmental conditions, e.g., temperature) have been

81 compiled into public databases (e.g., Fishbase; Froese 1990, 2011), that include species

82 within most major fish taxa. In principle, these databases allow us to use phylogenetic

83 information and correlations among life history parameters to obtain better estimates of vital

84 rates for both well-studied and poorly-studied species. Meta-analyses of life history

85 parameters in fishes indicate that multiple parameters can be described using a reduced set of

86 ordinates or groups (Winemiller and Rose 1992). Life history theory (e.g., Roff 1984,

87 Charnov 1993, Jensen 1996) and more recently the metabolic theory of ecology (Brown et al.

88 2004) have provided a conceptual framework that explains observed correlations among

89 traits. In particular, ratios of particular life-history parameters are often theorized to be nearly

90 constant across widely different taxa (Charnov et al. 1991). A special case of life-history

91 relationships occurs when the units are identical for parameters in the numerator and

92 denominator, such that the ratio is dimensionless. Charnov (1993) and others have defined

93 these as "life-history invariants" (LHI), and claimed that these dimensionless numbers in

94 some cases will be conserved across taxa. In practice, life-history theory, LHI, and observed

95 correlations suggest that if we know the adult mass and average environmental temperature

96 of a given species (which are relatively easy to measure in the field), we can use this

97 information to predict other, unknown life-history parameters. Obviously, if local studies are

98   available regarding additional life-history parameters for that given species (e.g., local

99   measurements of age at maturation), then these local results may also improve estimates of

100 other parameters for that species.

101   Many studies have shown that relationships among life-history parameters vary

102 taxonomically, so that relationships are more similar for closely related species than

103 otherwise unrelated species (e.g., Stearns 1983, Crespi and Teo 2002, Thorson et al. 2014b,

104 Nadon and Ault 2016, Dick et al. 2017). Intuitively, this suggests that taxonomic information

105 can also be useful when we predicting life history parameters. One approach to formalize

106 this intuition is to apply a multivariate extension of the standard phylogenetic model for

107 genetic evolution (Felsenstein 1985, Grafen 1989). Using this model, we predict that

108 parameters follow a random evolution over time, where correlations among life-history

109 parameters encode "neutral" (i.e., variable) or "conserved" (constant) relationships. Given

110 information regarding the evolutionary lineage of a given taxonomic tree (i.e., time since

111 divergence for every included species), this evolutionary model can inform correlations

112 among two species by the time that has elapsed since their most recent common ancestor. In

113 the absence of detailed evolutionary information, this tree can instead by approximated using

114 taxonomic information (e.g., by assuming that two species from the same genera are more

115 related than two species in the same family but not the same genera).

116   In this study, we predict life history variables (temperature and life-history parameters)

117 for all >32,000 fishes worldwide while accounting for similarity in the relationships among

118 life-history parameters for fishes that are taxonomically related, explicitly representing

119 residual error including correlations among parameters, and accounting for missing data. We

120 then use results to address the following outstanding questions regarding life-history

121 parameters in fishes: (1) What is the average relationship between the Brody growth

122 coefficient, K, and maximum size, L_∞ (for which different theories have previously been

123 published)? (2) Is the ratio of natural mortality M and the growth coefficient K a LHI, or

124 does it vary systematically (e.g., based on timing of maturity or other life-history

125 parameters)? and (3) How many variables must be known per species to precisely predict

126 remaining life-history variables (i.e., how many parameters are "neutral" vs. "conserved")?

127 We also distribute an R package *FishLife* (https://github.com/James-Thorson/FishLife)

128 containing code to re-run this multivariate life-history model as well as results from fitting it

129 to data from FishBase. This R package can also be used to update predictions for individual

130 stocks based on additional, stock-specific studies, and therefore has use for basic and applied

131   research on fishes worldwide. It also includes an Rshiny graphical user interface, which can

132   be used to visualize predicted life-history parameters for any fish species worldwide.

133   **Methods**

134   **Life history theory and invariants**

135   There are four main difficulties when predicting life-history parameters for all fishes

136   worldwide:

137   1. *Measurement errors*: Field-measurements of life history parameters are imprecise and

138   often correlated (e.g., a negative correlation in estimates of the Brody growth coefficient

139   and asymptotic length). We therefore treat existing records of life-history parameters as

140   imperfect measures of "true" parameters for a given species, where residual errors covary

141   among parameters. Estimating residual covariance is feasible whenever multiple

142   measurements are available for a single species (where we assume a single "true"

143   parameters-value for each species; Thorson et al. (2014a)). This residual covariance is a

144   combination of measurement errors (covariance in field measurements) and biological

145   variation within a given species (e.g., variation in parameters among years or

146   populations). In the following, we assume that field measurements of life-history

147   parameters are unbiased, although future work could address this assumption by

148   estimating average bias relative to a more-trusted data set (see e.g., Thorson et al. 2014a).

149   2. *Incomplete data*: No single field-study will provide estimates of all important life-history

150   parameters – instead, a study will generally report a subset of parameters that are

151   estimated using data from that study's design. To calculate the probability distribution of

152   available data while accounting for covariation in field measurements (as necessary in a

153   likelihood or Bayesian approach to parameter estimation), we must therefore account for

154   "incomplete data", i.e., the values of life-history parameters that are not reported for each

155   individual study. We therefore use standard missing-data methods under the assumption

156   that data are "missing at random" (i.e., that missing values can be imputed from reported

157   values).

158   3. *Taxonomic similarity:* Many parameters are more similar for taxonomically related

159   species than otherwise unrelated species. Importantly, this similarity is apparent at every

160   taxonomic level, e.g., among taxonomic classes (Charnov et al. 1991), families (Beverton

161   1992), and species (Thorson et al. 2014b, Dick et al. 2017). We therefore specify that

162   parameters vary "hierarchically", e.g., are different for Orders within a Class, Families

163   within an Order, Genera within a Family, and Species within a Genus (as recommended

164   by Hoenig et al. 2016). Measurements are only available for species, so average

165   parameters for higher-level taxa are never directly observed.

166 4. *Unknown functional forms*: There is little agreement about the exact functional-form of

167   the relationships among parameters. For example, recent studies have argued for (Hordyk

168   et al. 2015) and against (Nadon and Ault 2016) a purported dependence of rate-

169   parameters (M/k) and the ratio of length-at-maturity (L_{mat}) and asymptotic maximum

170   length (L_∞). We therefore use a life-history model that can potentially include

171   covariation among multiple parameters, rather than only modelling ratios among pairs of

172   parameters. Published life-history relationships are often multiplicative (i.e., the product

173   of age-at-maturity and natural mortality is approximately constant among species,

174   Beverton and Holt (1959)), so taking logarithms often results in a linear (additive)

175   relationship among parameters, or at least results in relationships that can be closely

176   approximated as being linear.

177 We overcome these four difficulties by specifying a multivariate model for trait evolution

178 along a taxonomic tree, while using replicated samples for each individual species to

179 distinguish trait evolution from residual covariance (e.g., estimation errors when calculating

180 parameters from field data).

Multivariate model for trait-evolution along taxonomic trees

182 Taxonomies are defined such that, for any taxonomic classification, it belongs to exactly one

183 classification for any higher taxonomic level (e.g., every family belongs to exactly one order,

184 etc). Therefore, we can define a set of taxonomic levels (e.g., Class, Order, Family, Genus,

185 and Species), and every taxon (e.g., genus *Sebastes*) is the "child" of exactly one "parent"

186 taxon (e.g., family *Sebastidae*). It is therefore possible to factor the probability distribution

187 for all life-history variables within a taxonomy by specifying a series of conditional

188 probabilities for variables for a given taxon via the variables of "ancestors" (the "parent"

189 taxon, its "parent", etc) (Felsenstein 1985). In the following, we specify a first-order Markov

190 process for trait evolution (i.e., variables for a "child" taxon depend only upon values for its

191 "parent" taxon).

192   We approximate trait evolution via a multivariate random-walk process. However, in the

193 absence of a high-quality chronology representing time-since-divergence for all fishes, we

194 have approximated trait-evolution by estimating covariation for taxonomic Classes, Orders

195 within Classes, families within Orders, genera within families, and species within genera (i.e.,

196 assuming that the time since divergence is approximately equal for all families within any

197 Order, etc.). Specifically, we define the true (unobserved) value $x_{g,j}$ of variable j (of n_J

198 modelled variables) and taxon g (of n_G modelled taxa), where $\mathbf{x}_g$ is the vector of true

199 variables for that taxon. Variables then depend only upon the value of variables for the

200 parent-taxon:

$$\mathbf{x}_g \sim MVN(\mathbf{x}_{p(g)}, \mathbf{\Sigma}_{l(g)}) \tag{1}$$

201 where $p(g)$ is the index of the parent-taxon for child g, and $l(g)$ indicates the taxonomic

202 level of the child-taxon (e.g., $l(g) = 1$ when the child-taxon is a taxonomic Class, and

203 $l(g) = 5$ when the child-taxon is a taxonomic species). The evolutionary-covariance, $\mathbf{\Sigma}_{l(g)}$,

204 may itself vary along the taxonomic tree. For simplicity, we assume that the correlation

205 among variables is identical for all taxonomic levels, but that the variance changes among

206 levels. Specifically, we decompose evolutionary-covariance $\mathbf{\Sigma}_{l(g)}$ into components caused by

207 common factors (sensu "exploratory factor analysis," Lawley 1940) or independent evolution

208 of each variables:

$$\mathbf{\Sigma}_{l(g)} = \lambda_l(\mathbf{L}_\Sigma \mathbf{L}_\Sigma^{\mathrm{T}} + \mathbf{D}) \tag{2}$$

209 where $\mathbf{L}_\Sigma$ is a lower-diagonal n_J by n_Σ matrix ($0 \leq n_\Sigma \leq n_J$) approximating covariance

210 among variables (where $\mathbf{L}_\Sigma$ is the Cholesky decomposition of $\lambda_l^{-1}\mathbf{\Sigma} - \mathbf{D}$ when $n_\Sigma = n_J$), $\mathbf{D}$ is

211 a diagonal matrix with diagonal elements σ_j^2 representing independent component of

212 variance in trait-evolution, and λ_l is the relative covariance explained by taxonomic level l

213 relative to the variance explained by taxonomic Order (i.e., we specify that $\lambda_{l=1} = 1$ to

214 ensure variable identifiability). The assumption that evolutionary-correlation is identical

215 among taxonomic levels is analogous to the assumption in conventional coalescent models

216 (Felsenstein 1973) that the evolutionary covariance does not change over time. Future

217 analyses could explore relaxing this assumption, although it would presumably require a

218 large increase in estimated coefficients (fixed and random effects).

219   We then estimate variables by specifying a distribution for augmented data $\tilde{\mathbf{y}}_i$ for the ith

220 study:

$$\tilde{\mathbf{y}}_i \sim MVN(\mathbf{x}_{g(i)}, \mathbf{V}) \tag{3}$$

221 where $\mathbf{V}$ is the residual covariance, which also includes factor-model ($\mathbf{L}_V$ with rank n_V) and

222 independent variance components (v_j^2). Many studies have missing values (i.e., instances

223 where $y_{i,j} = \mathrm{NA}$), and we assume that these are missing-at-random (i.e., available data are

224 representative of missing data). Given this assumption, missing data can be treated as

225 random with a uniform distribution:

$$\tilde{y}_{ij} = \begin{cases} \varepsilon_{ij} & \text{if } y_{ij} = NA \\ y_{ij} & \text{if } y_{ij} \neq NA \end{cases} \tag{4a}$$

226   where ε_{ij} is the missing value for variable j or observation i, and

$$\varepsilon_{ij} \sim Uniform(lb, ub) \tag{4b}$$

227   where lb and ub are set far below and above the range of data observed, such that further

228   decreases/increases in their values has no impact on model results.

229   We then estimate fixed effects by identifying values that maximize the marginal

230   likelihood function. Fixed effects include evolutionary covariance ($\mathbf{L}_\Sigma$ and σ_j^2), residual

231   covariance ($\mathbf{L}_V$ and v_j^2), and the ratio of evolution covariance among taxonomic levels (λ_l).

232   We also treat as fixed the average value μ_j for each life-history variable j for the common

233   ancestor of the taxonomic tree (Phyllum Chordata). The marginal likelihood is calculated

234   while integrating across the joint probability of data and random effects with respect to

235   random effects. We treat as random the true (latent) value of variables for each taxon ($\mathbf{x}_g$) as

236   well as all missing data (see Eq. 3 and 4b for their distribution). We approximate the

237   marginal likelihood using the Laplace approximation (Skaug and Fournier 2006),

238   implemented using Template Model Builder (Kristensen et al. 2016) within the R statistical

239   environment (R Core Team 2015). TMB also provides the gradients of the marginal

240   likelihood with respect to fixed effects, and we use these gradients in a conventional

241   nonlinear optimizer to identify maximum likelihood estimates (MLE) for fixed effects.

242   Given the MLE for fixed effects, we then calculate standard errors using a generalization of

243   the delta-method and the matrix of 2nd derivatives of the marginal likelihood function with

244   respect to fixed effects.

245   **Database**

246   We compile all available records of eight life-history variables (seven life history parameters

247   and temperature, $n_J = 8$): natural mortality rate (M), Brody growth coefficient (K),

248   asymptotic maximum length (L_∞), asymptotic maximum mass (W_∞), length at maturity

249   (L_{mat}), age at maturity (a_{mat}), maximum age (a_{max}), and average temperature (T). We

250   apply the life-history model to data for all eight variables, although we also report results

251   from a sensitivity analysis where we analyse only the seven life-history variables (i.e.,

252   excluding temperature). Subsets of these variables are often obtained from the same study

253   (e.g., a study providing both k and L_∞ for a single population of a given species), and we

254   record all measurements from a single study as a vector of data

$$y_i = \{\ln(M), \ln(k), \ln(L_\infty), \ln(W_\infty), \ln(L_{mat}), \ln(a_{mat}), \ln(a_{max}), T\} \tag{5}$$

255   for the ith study, while recording missing variables for each study as NA. We use a log-

256   transformation for all variables (except temperature, T) because (1) this transforms all $y_{i,j}$ to

257   have unbounded support (matching the support of the multivariate normal distribution from

258   Eq. 3), and (2) because this results in life-history invariants being expressed as linear

259   constraints that can be approximated using linear equations (e.g., $\ln(M) - \ln(K)$ is

260   approximately constant if M/K is a life-history invariant). All variables were downloaded

261   from FishBase (Froese 1990) using the package *rfishbase* (Boettiger et al. 2015) on Aug. 25,

262   2016, and our database is distributed with R package *FishLife*. In total, we obtained 29,196

263   variables (life-history parameters and temperature) from 9,853 studies ($n_i = 9,853$) for 3,551

264   unique species of the 33,102 fish species that are listed within FishBase. For each species,

265   we also noted the taxonomic Class, Order, and Family for a given Latin binomial (Genus-

266   species) name.

267 **Model exploration**

268   Next, we fitted the evolutionary model for life-history variables to this entire database while

269   varying model complexity in two ways: varying the number of fixed effects used to

270   approximate covariance in trait-evolution (n_Σ) or residual covariation (n_V). We specifically

271   perform a grid-search across all eight combinations for each configuration ($n_\Sigma = \{1,2,...,8\}$

272   and $n_V = \{1,2,...,8\}$), for a total of 64 models. For each model we confirm that the final

273   gradient of the marginal likelihood with respect to each fixed effects was low (<0.00001), and

274   that the hessian of the negative log-likelihood with respect to fixed effects was positive

275   definite. We then record Empirical Bayes predictions of values $\mathbf{x}_g$ for each taxon, and use

276   the generalized delta-method to calculate predictive covariance $\mathrm{Cov}(\mathbf{x}_g)$ for these variables.

277   To calculate the variance in life-history variables explained on average by knowing

278   taxonomic information for a fish, we calculate the total covariance among variables as:

$$Cov_{total} = \sum_{l=1}^{5} \mathbf{\Sigma}_l + \mathbf{V} \tag{6}$$

279   where the total variance for each individual life-history variable correspond to diagonal

280   elements of Cov_{total}. The variance explained by perfect knowledge of average life-history

281   variables for the highest l^* taxonomic levels then corresponds to:

$$Cov_{explained}(l^*) = \sum_{l=1}^{l^*} \mathbf{\Sigma}_l \tag{7}$$

282 such that the proportion of variance explained by perfect information about the highest l^*

283 taxonomic levels for variable j is $\sum_{l=1}^{l^*} \mathbf{\Sigma}_l (j,j)/[\sum_{l=1}^{5} \mathbf{\Sigma}_l (j,j) + \mathbf{V}(j,j)]$. Similarly, the

284 residual covariance in predictions given perfect information about the first l^* taxonomic

285 levels for a given species $Cov_{pred}(l^*)$ is expected to be:

$$Cov_{pred}(l^*) = Cov_{total} - Cov_{explained}(l^*) - \mathbf{V} = \sum_{l=l^*+1}^{5} \mathbf{\Sigma}_l \qquad (8)$$

286 However, in practice we have imperfect information about average life-history variables for

287 higher taxonomic levels, such that $Cov_{pred}(l^*) > \sum_{l=l^*+1}^{5} \mathbf{\Sigma}_l$, where $Cov_{pred}(l^*) -$

288 $\sum_{l=l^*+1}^{5} \mathbf{\Sigma}_l$ measures our imprecision when estimating average values for higher taxonomic

289 levels.

290   To interpret relationships among different life-history variables, we first predict variables

291 $\mathbf{x}_{pred}$ for a species with an Class, Order, family, genus, and species that is not within the

292 database (where $\mathbf{x}_0 = \mathbf{\mu}$), as well as the predicted covariance $Cov(\mathbf{x}_{pred})$ for these variables

293 (where $Cov(\mathbf{x}_{pred}) = Cov_{total} - \mathbf{V}$). This $Cov(\mathbf{x}_{pred})$ represents the predicted covariation

294 for life-history variables for a fish about which nothing is otherwise known, and it has rank

295 n_J because it is calculated from $\mathbf{\Sigma}_l$ (which itself has rank n_J because $\mathbf{\Sigma}_l$ includes a diagonal

296 component $\mathbf{D}$). We then calculate the eigen-decomposition of this covariance, and examine

297 the relative magnitude of eigenvalues to determine how many dimensions are necessary to

298 explain the vast majority (99%) of trait-covariance. For eigenvalues that explain a substantial

299 proportion of covariance, we then inspect their eigenvectors, where the ratio of two values in

300 a given eigenvector corresponds to the slope of a given axis of covariation. Inferring

301 relationships among variables from their predicted covariation, $Cov(\mathbf{x}_{pred})$, is similar in

302 some ways to major axis regression (MAR), and MAR is a useful for estimating relationships

303 among variables Y and Z whenever (1) an analyst seeks a relationship among Y and Z

304 without specifying either variable as "fixed" or "random", and (2) it is unknown whether

305 future users will be predicting Y from Z or Z from Y (Warton et al. 2006). Unlike

306 conventional MAR, however, we use multiple records of each variable for a given species

307 (several $\mathbf{y}_i$ for the same taxon g) and multiple taxa for the same parent-taxon (several $\mathbf{x}_g$ for

308 the same $\mathbf{x}_{p(g)}$) to separately estimate evolutionary and residual covariance ($\mathbf{\Sigma}_l$ and $\mathbf{V}$), and

309 our ability to separately estimate these two sources of covariance eliminates one major

310 drawback when using conventional MAR (Seim and Sæther 1983, McArdle 1988).

311 **Results**

Descriptive results

312   We first demonstrate our predictions of all eight variables (seven life-history parameters and

313   temperature) for six species, chosen for illustration purposes from the approximately 33,104

314   fish species listed in FishBase (Fig. 1-2). Model selection using the Akaike information

315   criterion (AIC) supports using a model where (1) the evolutionary covariance is the sum of

316   five factors ($n_\Sigma = 5$) and additional, independent variance for each variable, and (2) the

317   residual covariance is the sum of three factors ($n_V = 3$) and additional variance for each

318   factor. We use this model in the following, but confirm that results are essentially unchanged

319   when using other models with similar ΔAIC (Table 1).

320   Model predictions follow a multivariate normal distribution (as implied by the structure

321   of our model), and we visualize this predictive covariance using ellipses to represent a region

322   with a desired coverage (in this case, a 95% predictive ellipse). Predictions at high

323   taxonomic levels (e.g., for all species in class Actinopterygii) have wide confidence ellipses

324   (due to the accumulation of variance when predicting variables based only on high taxonomic

325   levels, see Eq. 7). Similarly, lower taxonomic levels generally have more precise (smaller)

326   predictive ellipses (as expected given Eq. 8). The exception is when predicting variables for

327   a species with no available data. In this case, the prediction is identical to the next-nearest

328   ancestor with available data. For example, any species of genus *Sebastes* without available

329   data has predictions equivalent to the predictive distribution of *Sebastes* (results not shown).

330   Similarly, in the unlikely event that an entire family has no available data, predictions for any

331   species within that family are identical to the predictions for that taxonomic order. In this

332   way, species without available information are guaranteed to have less-precise predictions of

333   life-history variables than species with abundant data.

334   We interpret the relationship among variables using the eigen-decomposition of the

335   predictive covariance for all fishes (Table 2). This covariance among life-history variables is

336   estimated while controlling for residual covariance, e.g., arising from estimation error when

337   calculating life-history parameters from available field data (Table 3). The dominant

338   eigenvector for trait-covariance is associated almost entirely with temperature (loading:

339   -0.995). The value of this eigenvector for natural mortality indicates that a 1 $°C$ increase in

340   temperature is associated with a 5% increase in natural mortality rate on average. This

341   increase in natural mortality is also positively associated with the growth coefficient, and

342   negatively associated with asymptotic weight and maximum age (among other associations).

343   Given its high loading with temperature, we interpret this axis as representing the effect of

345 temperature (or environmental characteristics associated with temperature, e.g., latitude) on

346 metabolic rates.

Question #1: What is the average association between growth rate and maximum size

348 The second axis when analyzing all eight variables explains approximately 80% of remaining

349 variance (after controlling for the impact of temperature), and has strongest association with

350 asymptotic mass (Table 2). Loadings indicate proportional variation in length-at-maturity

351 and asymptotic length, and a greater-than-cubic increase in asymptotic mass with increasing

352 asymptotic length (i.e., 1% increase in asymptotic length is associated with a 3.25% increase

353 in asymptotic mass, or a 3.21% increase for the sensitivity analysis excluding temperature,

354 Table S1). A 10% increase in asymptotic mass is in turn associated with a 2.4% decrease in

355 mortality rate and growth coefficient. Collectively, this axis suggests that body size is the

356 dominant axis of variation for life-history parameters in fishes that are subject to the same

357 temperature, and that increased size is negatively associated with mortality and growth rates.

358 Notably, the allometric-scaling of growth coefficient and asymptotic length ($\frac{-0.258}{0.204} = -1.24$)

359 is between the value of -1.0 predicted by Charnov et al. (2013) and -2.0 predicted by Pauly

360 and Binohlan (1996), and this result is robust to excluding temperature (Table S1, showing

361 $\frac{-0.258}{0.217} = -1.19$).

Question #2: Is the ratio of natural mortality rate and growth coefficient a LHI?

363 The first and second axes of variation explain 97% of total variance, and both have had

364 nearly proportional scaling of mortality (M) and growth coefficient (K). In isolation, these

365 two axes therefore suggest that the ratio M/k is constant among fishes. However, species

366 show large variation in the ratio of natural mortality and growth coefficient (Fig. 3), arising

367 predictably along phylogenetic lineages. In particular, family Sebastidae has relatively low

368 M/k, Salmonidae has relatively high M/k, and Scombridae and Lutjanidae fall somewhere in

369 the middle (3rd panel of Fig. 3).

370   This variation in M/k is generated largely via the 3rd axis of variation (Table 2), which

371 explains 1.6% of total variation. While small, this is nevertheless 60% of the variance

372 remaining after accounting for mass and temperature (or after accounting for mass when

373 excluding temperature, Table. S1). Importantly, this axis is the only component of

374 evolutionary covariance to have substantially different scaling between asymptotic length and

375 length-at-maturity, or between natural mortality rates and growth coefficients. Positive

376 movement along this axis represents a decrease in L_{mat}/L_∞, and we interpret this axis as

377 representing the timing of maturation: optimization of fecundity with respect to age-at-

378   maturity (following Charnov and Berrigan 1991). A decreased L_{mat}/L_∞ is then associated

379   with an increased value of M/k (Fig. 4). Importantly, related taxa (e.g., species within the

380   same taxonomic family) cluster closely together in values of L_{mat}/L_∞ and M/k, suggesting

381   that these parameters evolve gradually within lineages. This implies that species within the

382   same taxon (e.g., species within the same genus) will be highly informative about the value of

383   L_{mat}/L_∞ and M/k. For example, Sebastidae species have relatively high values for $L_{mat}/$

384   L_∞ (where a value of 1 indicates that growth ceases at maturation), and also have relatively

385   low values for M/k compared to other fishes.

386   **Question #3: How many parameters must be known to predict remaining parameters?**

387   Inspecting the eigen-decomposition of predictive variance for fishes shows that three axes

388   explain 99% of total variance in all eight variables (Table 2). Given results from this eigen-

389   decomposition, we infer that we could eliminate 99% of total variance when predicting the

390   remaining life-history variables if abundant sampling data allowed us to have near-perfect

391   measurements of temperature, maximum size, and either natural mortality or individual

392   growth rates for a given species. As sensitivity analysis, we also re-run the model while

393   excluding temperature (Appendix S1 Table S1). In this case, the first two axes explain 95%

394   of total variation, and the eigenvalues and eigenvectors for these axes are strongly similar to

395   the 2nd and 3rd axes of the model involving temperature.

396   Finally, we visualize the proportion of variance explained for each variable using this

397   model (Fig. 5) when decomposing total covariance into covariance explained by Class,

398   Order, family, genus and species, as well as residual covariance (where residual covariance is

399   shown in Table 3). This decomposition shows that residual covariance explains

400   approximately 10% of total variation on average across variables. Residual variation is

401   greatest (approximately 20%) for growth coefficients (K) and mortality (M) rates, where

402   estimates of both are expected to be most variable given typical field-measurements, and is

403   lowest (approximately 5%) for maximum length and length-at-maturity. Evolutionary-

404   covariance is greatest for different family within a given order, and is relatively similar for

405   other taxonomic levels.

406   **Discussion**

407   In this study, we have applied a multivariate model for the evolution of life-history

408   parameters to a global database of fish variables including individual growth, size, mortality,

409   maturity, and average temperature. We have then used this model to resolve three ongoing

410   questions regarding fish life history, as discussed in detail below. We also distribute our

411 results as an R package (with a graphical user interface) that can be used to predict life-

412 history parameters for any of the nearly 33,000 described fish species. These predictions are

413 informed using data in FishBase as well as taxonomic similarity to other fish species with

414 many or few data.

415   Predictions from our model indicate that variation among all eight variables can be

416 described using three main dimensions, representing temperature ($°C$), size (maximum body

417 mass, W_∞), and the timing of maturation (covariation among $\frac{M}{k}$ and $\frac{L_{mat}}{L_\infty}$). The first

418 dimension conforms to the metabolic theory of ecology (Brown et al. 2004), which predicts

419 that temperature controls metabolic rates underlying individual growth and activity levels

420 among taxa worldwide. Similarly, the second dimension conforms to the size-spectrum

421 theory of marine communities (Andersen et al. 2009), which predicts that growth and

422 mortality are defined by asymptotic mass. However, we also find that a third dimension (the

423 biological trade-offs between growth, maturity, and longevity) explains an important

424 component of variation (60% of variance after controlling for temperature and maximum

425 size). By identifying a third dimension of covariation, our results contrast with previous

426 empirical studies of fish life history, which predict growth, mortality, and maturity from a

427 species' asymptotic mass (e.g., Eq. 3 and 5c of Charnov et al. 2013). Interestingly, our three

428 dimensions correspond to the three predictors for natural mortality estimated by Pauly

429 (1980), although this latter study did not include maturity as either predictor or response

430 variable.

431   We conclude that three dimensions are sufficient to capture 99% of covariation among

432 our eight variables, and a natural question is whether this number of dimensions is consistent

433 with existing life-history theory. We therefore compile a list of previously published

434 relationships among these eight life-history variables for fishes that are based entirely upon

435 theoretical optimization of fitness rather than empirical observations (Appendix S2). This

436 table shows that reported relationships can often be derived from one-another. For example,

437 Jensen (1996) attributes three LHI to Beverton and Holt (Eq. TB1-4, TB1-5, and TB1-6 in

438 Appendix S2). However, the third LHI (TB1-6) can be derived from the others given that

439 growth follows a von Bertalanffy growth curve (TB1-1). We therefore identify a minimal set

440 of life history relationships that can be used to derive or approximate all other reported

441 relationships (Table 4). We refer to this as a "minimal and sufficient" set of relationships,

442 because it involves the minimal number of equations (i.e., eliminates equations that can be

443 derived from others) but still is sufficient to contain (or approximate) the other reported

444   relationships. These six relationships represent von Bertalanffy growth (T4-1); allometric

445   scaling of weight-at-length (T4-2), maximizing fecundity with respect to age at maturity (T4-

446   3) and the Brody growth coefficient (T4-4), the definition of maximum age (T4-5), and the

447   metabolic scaling with temperature (T4-6). We also hypothesize which variables in these six

448   equations are likely to be conserved or varying among fish species. By examining the

449   number of free parameters that can be used to calculate all eight variables in this study (see

450   Table 4 caption), we predict that all eight life-history variables can be well-described via 3-4

451   dimensions. This prediction is generally consistent with our empirical conclusion that all

452   variables can be well predicted given perfect knowledge of three variables for each species.

453   Across all fishes, we also conclude that asymptotic length $\ln(L_\infty)$ increases 24% faster

454   than any decrease in growth coefficient $\ln(K)$ among species, i.e., $\ln(K) \propto -1.24 \ln(L_\infty)$ as

455   indicated by the 2nd axis of variation. This estimate arises after using replicated sampling for

456   each individual species to estimate the residual covariance for these two parameters arising

457   from estimation errors: as expected, this residual covariance is negative between $\ln(L_\infty)$ and

458    $\ln(K)$ (see Table 3). Following Charnov and Berrigan (1991), we interpret this covariance

459   between $\ln(L_\infty)$ and $\ln(K)$ as species optimizing their fecundity with respect to growth rates,

460   with resulting impacts on asymptotic length and mass. Our predicted coefficient (using a

461   variant of major axis regression) is closer to the relationship $\ln(k) \propto -\ln(L_\infty)$ predicted by

462   Charnov et al. (2013) than the relationship $\ln(k) \propto -2\ln(L_\infty)$ predicted by Pauly and

463   Binohlan (1996). The product of growth coefficient and asymptotic mass is the anabolic rate

464   for juveniles (Mangel 2006, Charnov et al. 2013), so our results suggests that large-bodied

465   species on average have a higher juvenile anabolic rate (i.e., are able to access greater

466   densities of available food as juveniles).

467   Lastly, we conclude that the ratio of natural mortality rate and the Brody growth

468   coefficient ($\frac{M}{K}$) is not itself conserved across species (i.e., not a "life history invariant" as

469   claimed by Jensen (1996 Eq. 8)). Instead species with slow growth relative to mortality

470   (relatively high $\frac{M}{K}$) also have lower maturation length relative to maximum length (relatively

471   low $\frac{L_{mat}}{L_\infty}$), such that the ratio of age-at-maturity (a_{mat}) and maximum age (a_{max}) is

472   approximately conserved among species. This result explains large differences in $\frac{M}{K}$ among

473   taxonomic groups, e.g., where *Sebastes* species have very low $\frac{M}{K}$ and high $\frac{L_{mat}}{L_\infty}$ relative to

474   other taxa (Thorson et al. 2014b). This association between $\frac{M}{K}$ and $\frac{L_{mat}}{L_\infty}$ is not a new

475   discovery (Holt 1958, Beverton 1992, Hordyk et al. 2015). In fact, it is easy to justify

476   theoretically given the assumption that lifetime reproductive output per recruit is proportional

477   to survival-to-maturity and mass at maturity (we follow the presentation from Mangel (2006

478   Chap. 2)):

$$F(a_{mat}) = \exp\left(-M a_{mat}\right) \times \alpha L_{mat}^{\beta}$$

479   where α is the average mass per standardized volume and β represents ontogenic changes in

480   body shape. Optimizing fitness $F(a_{mat})$ with respect to a_{mat} yields:

$$a_{mat} = \frac{1}{k}\log\left(\frac{M + \beta K}{M}\right)$$

481   where substituting L_{mat} for a_{mat} and re-arranging then yields:

$$\frac{L_{mat}}{L_{\infty}} = \left(\frac{\beta K}{M + \beta K}\right)$$

482   This derivation therefore suggests that $\frac{L_{mat}}{L_{\infty}}$ and $\frac{M}{k}$ should have a nonlinear association.

483   Despite this theoretical evidence, however, authors have periodically suggested that $\frac{M}{k}$ is

484   constant among species (e.g., Jensen 1996). Our study also estimates taxonomic variation

485   around this relationship, e.g., where salmonids have higher $\frac{L_{mat}}{L_{\infty}}$ than predicted from their $\frac{M}{k}$

486   (i.e., are closer to determinate growth than otherwise predicted), whereas Scombrids are the

487   opposite (less determinate growth than otherwise predicted). We hope that this combination

488   of theoretical and empirical evidence will put to rest the notion that $\frac{M}{k}$ is a life-history

489   invariant, and will encourage future research to explain residual patterns around life-history

490   relationships among taxa.

491   Finally, we note several shortcomings of our current study, which could potentially be

492   explored in future research. Most importantly, previous evolutionary studies have often used

493   rates of molecular evolution rather than taxonomic information to account for correlations

494   among species (Felsenstein 1973, 1985). Using the "coalescent model" for molecular

495   evolution, an analyst can model an evolutionary tree as being shaped by standard population-

496   genetic forces. For this reason, authors have argued that the coalescent model provides a

497   coherent statistical basis for analyzing species lineages. However, the coalescent model may

498   be difficult to apply in practice because it involves several unknown parameters including

499   effective population size, recombination rate, and selection history (i.e., selective and random

500   mutation rates). In particular, a large evolutionary tree (like we analyses here) may include

501   branches with large differences in evolutionary rates, and these differences may lead to

502 difficulties in reconstructing the species tree, or in accurately approximating correlations

503 among species based only on time since divergence (Takahashi et al. 2001, Rosenberg and

504 Nordborg 2002). Therefore, we have instead used a taxonomic tree to approximate the

505 evolutionary time since divergence for all pairs of species. Our current approach is similar to

506 previous fisheries meta-analyses that have used nested random effects to explain variation in

507 species parameters at different taxonomic levels (Foss-Grant et al. 2016), and we recommend

508 future research conducting a detailed comparison of the two methods. We also note that

509 future research could explore different evolutionary-covariance occurring at different

510 taxonomic levels, where different covariance could be interpreted as representing different

511 evolutionary constraints operating at short (species-level) or long (family- or class-level)

512 evolutionary scales. This exploration would not be possible using the conventional

513 coalescent model, so we see this potential for future research as one advantage to our current

514 approach.

515   Secondly, we have not explicitly included many variables that might be expected to

516 control opportunities for growth or reproduction in fishes. For example, species parameters

517 like maximum size are associated with depth and latitude (Cailliet et al. 2001, Gertseva et al.

518 2010). These patterns may be driven to some extent by temperature (Pauly 2010), and our

519 model therefore indirectly accounts for depth and latitude via our inclusion of temperature

520 measurements. However, other factors controlling opportunities for growth include predation

521 and food availability (Winemiller 1989). We have not included any variable to directly

522 measure differences in food availability or predation, and identifying such a variable (e.g.,

523 using regional variation in primary production) could perhaps improve life-history

524 predictions in fishes. Likewise, human-induced changes in life history parameters could alter

525 the relationship among traits (Conover and Munch 2002), so including a variable representing

526 exploitation history may similarly improve predictions in some fishes. While a large portion

527 of variability is captured in the three main dimensions presented here, further precision and

528 ecological insight could be gaining by including additional, ecologically relevant variables in

529 the life-history analysis.

530   Thirdly, we have used data from FishBase without attempting to discriminate between

531 high and low-quality studies (e.g., between estimates of natural mortality arising from high-

532 quality tag-recapture vs. lower-quality catch curves for lightly impacted populations).

533 Previous work suggests that FishBase records are on average an unbiased measure of high

534 quality for individual species (Thorson et al. 2014a). We have also assimilated records that

535 include a large proportion of missing values (i.e., 63% of all values are missing) and this high

536 proportion of missing records may impact our ability to precisely estimate the number of

537 dimensions for representing trait-covariation. To account for this high proportion of missing

538 data, we used a "rank-reduction" technique to estimate covariation among variables (Zuur et

539 al. 2003, Warton et al. 2015), and this allowed us to select a parsimonious number of

540 parameters for both evolutionary and residual covariation. However, it is difficult to assess

541 the potential impact of missing data on estimating the dimensionality of covariation, and we

542 recommend future statistical exploration regarding this topic.

543   We distribute an R package *FishLife* containing code to replicate or update our analysis,

544 and packaging results when fitted to FishBase data. This R package includes predictions for

545 all marine fishes, where species without any available data are predicted using information

546 from all related taxa. The package also includes a function to update predictions of life-

547 history parameters for a given species based on user-supplied data. In this way, it can be

548 used to generate species-specific predictions of life-history parameter that combine both

549 global and local data. Many previous studies have developed relationships for predicting

550 these same variables, e.g., where life-history data are combined with assumed functional

551 relationships among variables to estimate a small set of model parameters, and these

552 relationships and model parameters are then used to predict life-history variables for

553 unobserved species (Pauly and Binohlan 1996, Froese and Binohlan 2000, Hordyk et al.

554 2015, Nadon and Ault 2016). By contrast, our approach uses both model parameters (fixed

555 effects) as well as patterns in residuals among taxa (random effects) when predicting life-

556 history variables for any species, and therefore bridges between analyses of all fish species

557 (Pauly and Binohlan 1996, Charnov et al. 2013) and analyses restricted to specific taxa

558 (Thorson et al. 2014b, Nadon and Ault 2016). We believe that our mixed-effect approach

559 (and R package *FishLife*) has several benefits relative to previous regression-approaches

560 including: (1) it predicts life-history variables probabilistically for species while using

561 information about related taxa to inform predictions; (2) it predicts uncertainty for species

562 based on the quantity of data that are available, so that species with many field-measurements

563 of life-history variables have predictions that are more precise than species with only a few

564 field measurements; (3) it uses the full set of life-history variables available for a given

565 species to predict all unknown variables, e.g., rather than calculating natural mortality from a

566 single variable (i.e., maximum age) and ignoring all others (e.g., individual growth rates).

567 **Acknowledgements**

568 We thank R. Froese for the tremendous effort in compiling records for all fishes globally into

569 FishBase, and D. Pauly for decades of research that has highlighted the utility of FishBase of

570 fish life-history analysis. We also thank P. Neubauer for preliminary discussions regarding

571 phylogenetic regression, and O. Hamel, J. Hastie, M. Mangel, and one anonymous reviewer

572 for comments on an earlier draft.

573 **Bibliography**

574 Andersen, K. H., K. D. Farnsworth, M. Pedersen, H. Gislason, and J. E. Beyer. 2009. How

575   community ecology links natural mortality, growth, and production of fish

576   populations. ICES Journal of Marine Science: Journal du Conseil 66:1978–1984.

577 Beverton, R., and S. Holt. 1959. A review of the lifespans and mortality rates of fish in

578   nature, and their relation to growth and other physiological characteristics. Pages

579   142–177 *in* G. E. W. Wolstenholme and M. O'Conner, editors. Ciba Foundation

580   Symposium-The Lifespan of Animals (Colloquia on Ageing). J. and A. Churchil Ltd.,

581   London.

582 Beverton, R. J. H. 1992. Patterns of reproductive strategy parameters in some marine teleost

583   fishes. Journal of Fish Biology 41:137–160.

584 Boettiger, C., S. Chamberlain, D. T. Lang, and P. Wainwright. 2015. rfishbase: R Interface to

585   "FishBase."

586 Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a

587   metabolic theory of ecology. Ecology 85:1771–1789.

588 Cailliet, G. M., A. H. Andrews, E. J. Burton, D. L. Watters, D. E. Kline, and L. A. Ferry-

589   Graham. 2001. Age determination and validation studies of marine fishes: do deep-

590   dwellers live longer? Experimental Gerontology 36:739–764.

591 Charnov, E. L. 1993. Life history invariants: some explorations of symmetry in evolutionary

592   ecology. Oxford University Press, Oxford, UK.

593 Charnov, E. L., D. Berrigan, and R. J. H. Bevertron. 1991. Dimensionless numbers and the

594   assembly rules for life histories [and discussion]. Philosophical Transactions of the

595   Royal Society of London. Series B: Biological Sciences 332:41–48.

596 Charnov, E. L., H. Gislason, and J. G. Pope. 2013. Evolutionary assembly rules for fish life

597   histories. Fish and Fisheries 14:213–224.

598 Conover, D. O., and S. B. Munch. 2002. Sustaining fisheries yields over evolutionary time

599   scales. Science 297:94–96.

600 Crespi, B. J., and R. Teo. 2002. Comparative phylogenetic analysis of the evolution of

601   semelparity and life history in salmonid fishes. Evolution 56:1008–1020.

602 Dick, E. J., S. Beyer, M. Mangel, and S. Ralston. 2017. A meta-analysis of fecundity in

603   rockfishes (genus Sebastes). Fisheries Research 187:73–85.

604 Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous

605   characters. American Journal of Human Genetics 25:471.

606 Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist:1–15.

607 Foss-Grant, A. P., E. F. Zipkin, J. T. Thorson, O. P. Jensen, and W. F. Fagan. 2016.

608   Hierarchical analysis of taxonomic variation in intraspecific competition across fish

609   species. Ecology 97:1724–1734.

610 Froese, R. 1990. FishBase: An information system to support fisheries and aquaculture

611   research. ICLARM Fishbyte 8:21–24.

612 Froese, R. 2011. The science in FishBase. Pages 47–54 *in* V. Christensen and J. Maclean,

613   editors. Ecosystem approaches to fisheries: a global perspective. First edition.

614   Cambridge University Press, Cambridge, UK.

615 Froese, R., and C. Binohlan. 2000. Empirical relationships to estimate asymptotic length,

616   length at first maturity and length at maximum yield per recruit in fishes, with a

617   simple method to evaluate length frequency data. Journal of Fish Biology 56:758–

618   773.

619 Gertseva, V. V., J. M. Cope, and S. E. Matson. 2010. Growth variability in the splitnose

620   rockfish Sebastes diploproa of the northeast Pacific Ocean: pattern revisited. Marine

621   Ecology Progress Series 413:125–136.

622 Grafen, A. 1989. The Phylogenetic Regression. Philosophical Transactions of the Royal

623   Society of London. Series B, Biological Sciences 326:119–157.

624 Hoenig, J. M., A. Y.-H. Then, E. A. Babcock, N. G. Hall, D. A. Hewitt, and S. A. Hesp.

625   2016. The logic of comparative life history studies for estimating key parameters,

626   with a focus on natural mortality rate. ICES Journal of Marine Science: Journal du

627   Conseil 73:2453–2467.

628 Holt, S. J. 1958. The evaluation of fisheries resources by the dynamic analysis of stocks, and

629   notes on the time factors involved. ICNAF Special Publication 1:77–95.

630 Hordyk, A., K. Ono, K. Sainsbury, N. Loneragan, and J. Prince. 2015. Some explorations of

631   the life history ratios to describe length composition, spawning-per-recruit, and the

632   spawning potential ratio. ICES Journal of Marine Science: Journal du Conseil

633   72:204–216.

634   Jensen, A. L. 1996. Beverton and Holt life history invariants result from optimal trade-off of

635   reproduction and survival. Canadian Journal of Fisheries and Aquatic Sciences

636   53:820–822.

637   Kenchington, T. J. 2014. Natural mortality estimators for information-limited fisheries. Fish

638   and Fisheries 15:533–562.

639   Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. 2016. TMB: Automatic

640   Differentiation and Laplace Approximation. Journal of Statistical Software 70:1–21.

641   Lawley, D. N. 1940. VI.—The Estimation of Factor Loadings by the Method of Maximum

642   Likelihood. Proceedings of the Royal Society of Edinburgh 60:64–82.

643   Magnusson, A., and R. Hilborn. 2007. What makes fisheries data informative? Fish and

644   Fisheries 8:337–358.

645   Mangel, M. 2006. The theoretical biologist's toolbox: quantitative methods for ecology and

646   evolutionary biology. Cambridge University Press, Cambridge, UK.

647   Maunder, M. N., and A. E. Punt. 2013. A review of integrated analysis in fisheries stock

648   assessment. Fisheries Research 142:61–74.

649   McArdle, B. H. 1988. The structural relationship: regression in biology. Canadian Journal of

650   Zoology 66:2329–2339.

651   Nadon, M. O., and J. S. Ault. 2016. A stepwise stochastic simulation approach to estimate

652   life history parameters for data-poor fisheries. Canadian Journal of Fisheries and

653   Aquatic Sciences 73:1–11.

654   Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and

655   mean environmental temperature in 175 fish stocks. Journal du Conseil international

656   pour l'Exploration de la Mer 39:175–192.

657   Pauly, D. 2010. Gasping fish and panting squids: oxygen, temperature and the growth of

658   water-breathing animals. International Ecology Institute, Oldendorf/Luhe, Germany.

659   Pauly, D., and C. Binohlan. 1996. FishBase and AUXIM as tools for comparing life-patterns,

660   growth and natural mortality of fish: applications to snappers and groupers. Page 449

661   et al., Biology, Fisheries and Culture of Tropical Groupers and Snappers. ICLARM

662   Conf. Proc.

663   Quinn, T. J., and R. B. Deriso. 1999. Quantitative Fish Dynamics. First edition. Oxford

664   University Press, Oxford, UK.

665   R Core Team. 2015. R: A Language and Environment for Statistical Computing. R

666   Foundation for Statistical Computing, Vienna, Austria.

667   Roff, D. A. 1984. The evolution of life history parameters in teleosts. Canadian Journal of
668   Fisheries and Aquatic Sciences 41:989–1000.

669   Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the
670   analysis of genetic polymorphisms. Nature Reviews Genetics 3:380–390.

671   Seim, E., and B.-E. Sæther. 1983. On rethinking allometry: which regression model to use?
672   Journal of Theoretical Biology 104:161–168.

673   Skaug, H., and D. Fournier. 2006. Automatic approximation of the marginal likelihood in
674   non-Gaussian hierarchical models. Computational Statistics & Data Analysis 51:699–
675   709.

676   Stearns, S. C. 1983. The Influence of Size and Phylogeny on Patterns of Covariation among
677   Life-History Traits in the Mammals. Oikos 41:173–187.

678   Takahashi, K., Y. Terai, M. Nishida, and N. Okada. 2001. Phylogenetic Relationships and
679   Ancient Incomplete Lineage Sorting Among Cichlid Fishes in Lake Tanganyika as
680   Revealed by Analysis of the Insertion of Retroposons. Molecular Biology and
681   Evolution 18:2057–2066.

682   Thorson, J. T., J. M. Cope, and W. S. Patrick. 2014a. Assessing the quality of life history
683   information in publicly available databases. Ecological Applications 24:217–226.

684   Thorson, J. T., I. G. Taylor, I. Stewart, and A. E. Punt. 2014b. Rigorous meta-analysis of life
685   history correlations by simultaneously analyzing multiple population dynamics
686   models. Ecological Applications 24:315–326.

687   Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and
688   F. K. C. Hui. 2015. So Many Variables: Joint Modeling in Community Ecology.
689   Trends in Ecology & Evolution 30:766–779.

690   Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby. 2006. Bivariate line-fitting
691   methods for allometry. Biological Reviews 81:259–291.

692   Winemiller, K. O. 1989. Patterns of variation in life history among South American fishes in
693   seasonal environments. Oecologia 81:225–241.

694   Winemiller, K. O., and K. A. Rose. 1992. Patterns of life-history diversification in North
695   American fishes: implications for population regulation. Can. J. Fish. Aquat. Sci.
696   49:2196–2218.

697   Zuur, A. F., I. D. Tuck, and N. Bailey. 2003. Dynamic factor analysis to estimate common
698   trends in fisheries time series. Canadian Journal of Fisheries and Aquatic Sciences
699   60:542–552.

700

701

702

703 Supporting Information

704 Additional supporting information may be found in the online version of this article at

705 http://onlinelibrary.wiley.com/doi/10.1002/eap.xxxx/suppinfo

706

707 Data Availability

708 Data available from the Github Repository: https://doi.org/10.5281/zenodo.826921

709

710

711 Table 1 – Model selection results for the multivariate model of fish life-history variables.

Number of factors for covariance among taxa	Number of factors for observation covariance	ΔAIC
5	3	0.0
5	4	0.4
4	4	3.8
6	3	6.0
6	4	6.4
5	5	7.1
4	5	9.3

712

713

714 Table 2 – Summary of predictive covariance among all eight life history variables. Predictive

715 covariance is decomposed into eight orthogonal components, where the first component

716 explains the maximum variance, the second explains the maximum variance after accounting

717 for the 1st component, etc. The first three components explain 99% of total variance, and we

718 list the predicted variance for each (1st row), proportion of total variance (2nd row),

719 cumulative proportion of variance (3rd row), and association of that component with different

720 life-history variables (which we term "loadings"). The ratio of two loadings represents the

721 average relationship between two variables along the axis of variation represented by that

722 eigenvector.

	Eigenvectors		
	#1	#2	#3
Variance decomposition			
Eigenvalue	52.348	10.09	1.051
Proportion of variance	0.816	0.157	0.016
Cumulative proportion of variance	0.816	0.973	0.990
Loadings			
$\ln(L_\infty)$	-0.019	0.258	0.120
$\ln(K)$	0.043	-0.204	0.360
$\ln(W_\infty)$	-0.039	0.838	0.393
$\ln(a_{max})$	-0.039	0.200	-0.523
$\ln(a_{mat})$	-0.042	0.207	-0.481
$\ln(M)$	0.050	-0.205	0.437
$\ln(L_{mat})$	-0.022	0.241	0.048
Temperature	0.995	0.079	-0.060

723

724   Table 3 – Measurement covariance (lower-triangle, including diagonal) and resulting correlation (upper triangle, indicated by bold-font) for

725   replicated measurements of each life-history variable for a given species (e.g., the 2nd row and 1st column shows the negative measurement

726   covariance between $\ln(K)$ and $\ln(L_\infty)$).

	$\ln(L_\infty)$	$\ln(K)$	$\ln(W_\infty)$	$\ln(a_{max})$	$\ln(a_{mat})$	$\ln(M)$	$\ln(L_{mat})$	Temperature
$\ln(L_\infty)$	0.063	**-0.585**	**0.964**	**0.205**	**0.545**	**-0.299**	**0.598**	**-0.035**
$\ln(K)$	-0.082	0.311	**-0.562**	**-0.354**	**-0.399**	**0.584**	**-0.168**	**0.087**
$\ln(W_\infty)$	0.189	-0.244	0.609	**0.185**	**0.514**	**-0.280**	**0.566**	**-0.029**
$\ln(a_{max})$	0.018	-0.07	0.051	0.127	**0.448**	**-0.564**	**0.234**	**-0.154**
$\ln(a_{mat})$	0.043	-0.069	0.127	0.050	0.100	**-0.389**	**0.663**	**-0.140**
$\ln(M)$	-0.037	0.158	-0.106	-0.097	-0.060	0.236	**0.041**	**0.161**
$\ln(L_{mat})$	0.034	-0.021	0.099	0.019	0.047	0.005	0.050	**-0.081**
Temperature	-0.025	0.137	-0.064	-0.154	-0.124	0.220	-0.051	7.912

727

728

729

730   Table 4 – Minimal (i.e., non-redundant) and sufficient (i.e., not missing any relations) set of

731   theoretical life-history relationships among the eight life history variables

732     $\{M, K, a_{mat}, L_{mat}, L_\infty, W_\infty, a_{max}, T\}$ analyzed in this study. These relationships involve six

733   exogenous parameters, $\{\delta, \alpha, \beta, E, A, C\}$, and if these parameters are constant for all fishes,

734   we expect that all eight variables are constrained within a subspace with 2 degrees of freedom

735   (= 8 variables − 6 relations). We hypothesize that: α (average mass per standardized

736   volume) does not vary systematically among fishes; β (ontogenic changes in body shape)

737   will vary systematically among fishes; δ (allometric scaling of pre-mature survival to length-

738   at-maturity) does not vary systematically; A (anabolic rate) does not vary systematically; E

739   (activation energy from metabolic theory) does not vary systematically; and C (average

740   mortality rate when $T \gg E$) may or may not vary systematically. We therefore hypothesize

741   that the value for all eight variables for each fish can be described via 3-4 coordinates,

742     $\{R, T, \beta\}$ or $\{R, T, \beta, C\}$, depending upon whether C is constant or varies among fishes.

Eq.	Relationship	Theoretical justification	Notes
T4-1	$L_{mat} = L_\infty \exp(-K a_{mat})$	Individual energy gain/loss	-
T4-2	$\log(W_\infty) = \log(\alpha) + \beta \log(L_\infty)$	Spatial scaling of individual body size	α is average tissue mass per volume; β is ontogenic changes in body shape
T4-3	$\delta \dfrac{K}{M} = \dfrac{R}{1-R}$ Where $R = \dfrac{L_{mat}}{L_\infty}$	Maximizing individual fecundity w.r.t. a_{mat}	Eq. T1-1 and T1-3 imply: $$\frac{L_{mat}}{L_\infty} = \frac{\delta}{\delta + M/K}$$ and: $$a_{mat} = \frac{1}{K}\log\left(\frac{\delta K + M}{M}\right)$$
T4-4	$L_\infty = A \times K^{-h}$ Where $h = \dfrac{R-1}{R}\log(1-R)$	Maximizing individual fecundity w.r.t. K	A is anabolic rate
T4-5	$a_{max} = \dfrac{\ln(0.01)}{M}$	True by definition of a_{max} as the age to which 1% of post-larval individuals survive	-

| T4-6 | $M = C \times \exp\left(-\dfrac{E}{T}\right)$ | Metabolic scaling of biological rates and temperature | E is activation energy (~0.6-0.7 eV) divided by Boltzmann's constant (8.617×10⁻⁵ eV/°K) C is proportionality constant between M and metabolic energy |

E is activation energy (~0.6-0.7 eV) divided by Boltzmann's constant (8.617×10^{-5} eV/°K)

C is proportionality constant between M and metabolic energy

743

744

745 Fig. 1 – Predictive distribution for eight life-history variables of three species of genus

746 Thunnus (*T. alalunga*, *T. thynnus*, and *T. albacares*), as well as the predictive distribution for

747 genus Thunnus and its ancestral taxa (family Scombridae, order Perciformes, and class

748 Actinopterygii). Panels show the 95% predictive distribution for all life-history variables in

749 our database: individual growth (x-axis) and natural mortality rate (y-axis; top-left right);

750 asymptotic maximum weight (x-axis) and asymptotic maximum length (y-axis; top-right

751 panel); maximum age (x-axis) and age at maturity (y-axis; bottom-left panel); and length at

752 maturity (x-axis) and average temperature for the species' spatial distribution (y-axis;

753 bottom-right panel).

754 Fig. 2 – Predictive distribution for eight life-history variables of three species of genus

755 *Sebastes* (*S. alutus*, *S. pinniger*, and *S. crameri*), see Fig. 2 caption for details

756 Fig. 3 – Marginal predictive distribution for ratio of natural mortality rate and the Brody

757 growth coefficient (derived from the predictive covariance for $\ln(M)$ and $\ln(K)$) for each

758 taxon in the life history database, highlighting results for four taxonomic families

759 (Sebastidae, Salmonidae, Scombridae, Lutjanidae) that are frequently fished.

760 Fig. 4 – Predictive distribution (showing the bivariate 95% predictive interval) for M/K (x-

761 axis) vs. L_{mat}/L_{max} (y-axis) for each taxon in the life history database (see Fig. 5 caption for

762 details), where the bottom panel shows the predicted relationship from Holt (1958), see Table

763 1, Eq. T1.3.

764 Fig. 5 – Proportion of variance explained by Class, Order, family, genus, species, and

765 residual covariance calculated using Eq. 6-7 (Class is darkest grey; residual is lightest grey;

766 and other levels are ordered according to the preceding list). Variables differ in the

767 magnitude of residual covariance, but the proportion explained by taxonomic levels is

768   otherwise proportional among variables. "Comb." refers to the sum of variance for all
769   variables, and thus represents variance-explained on average among variables.
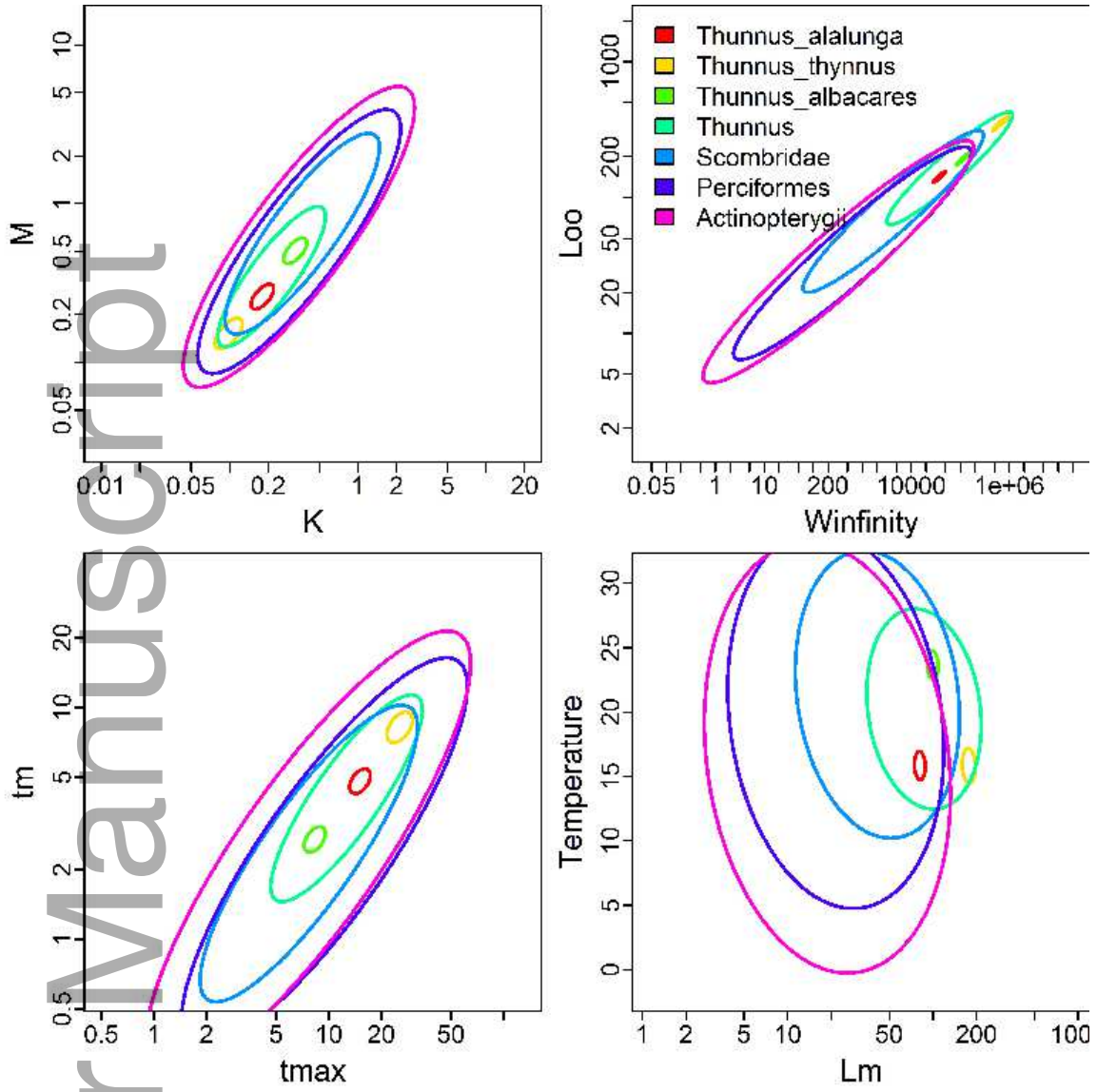
eap_1606_f1.jpg
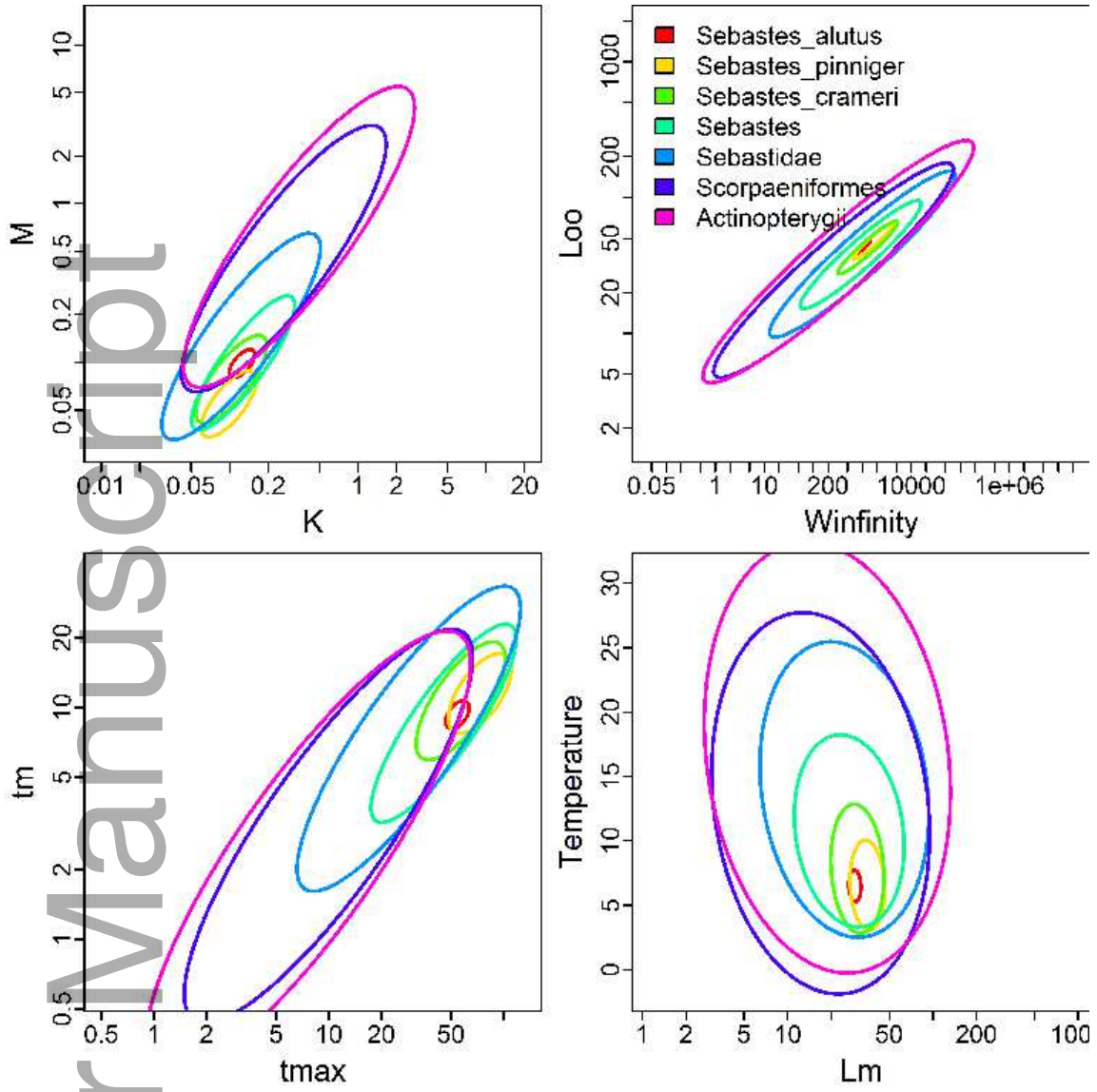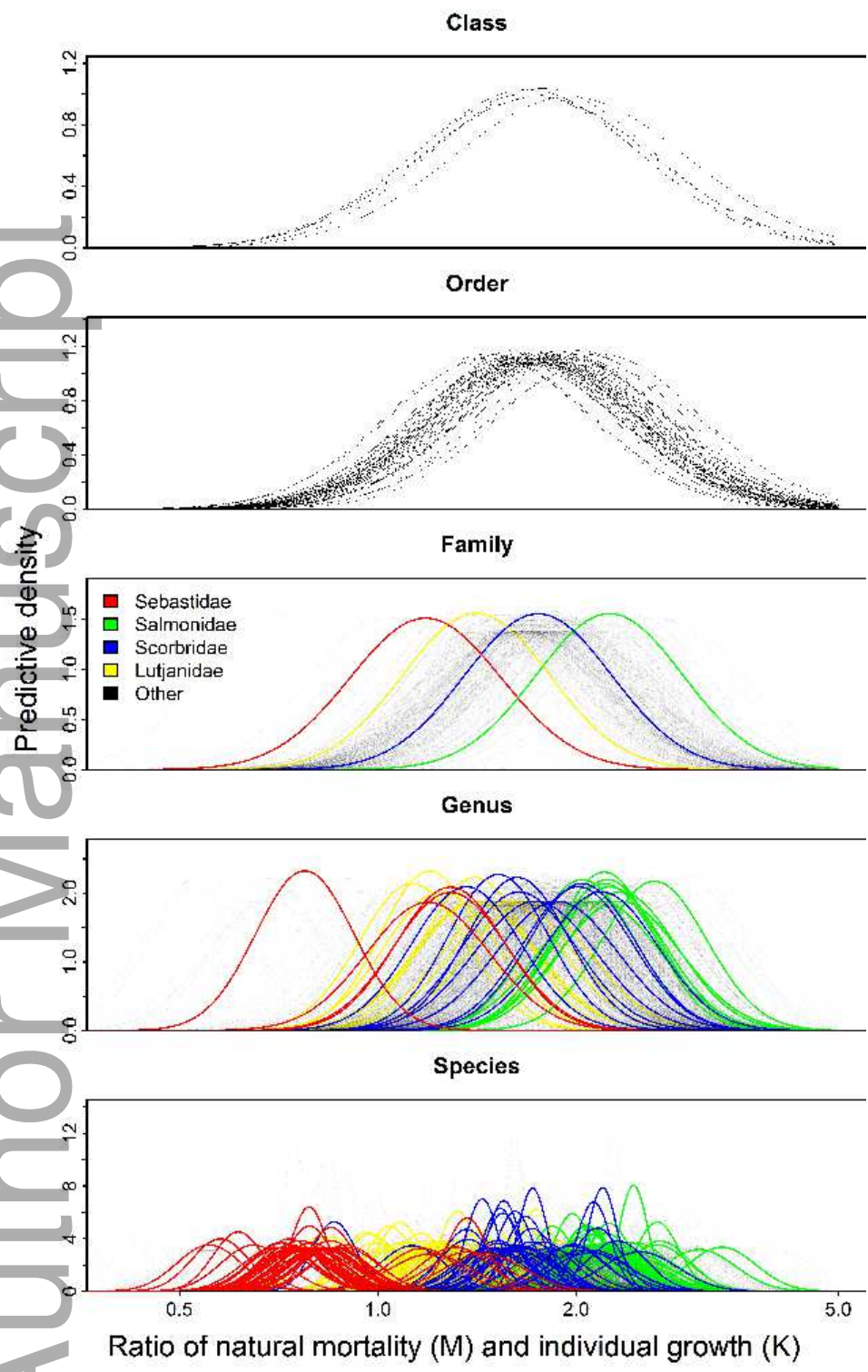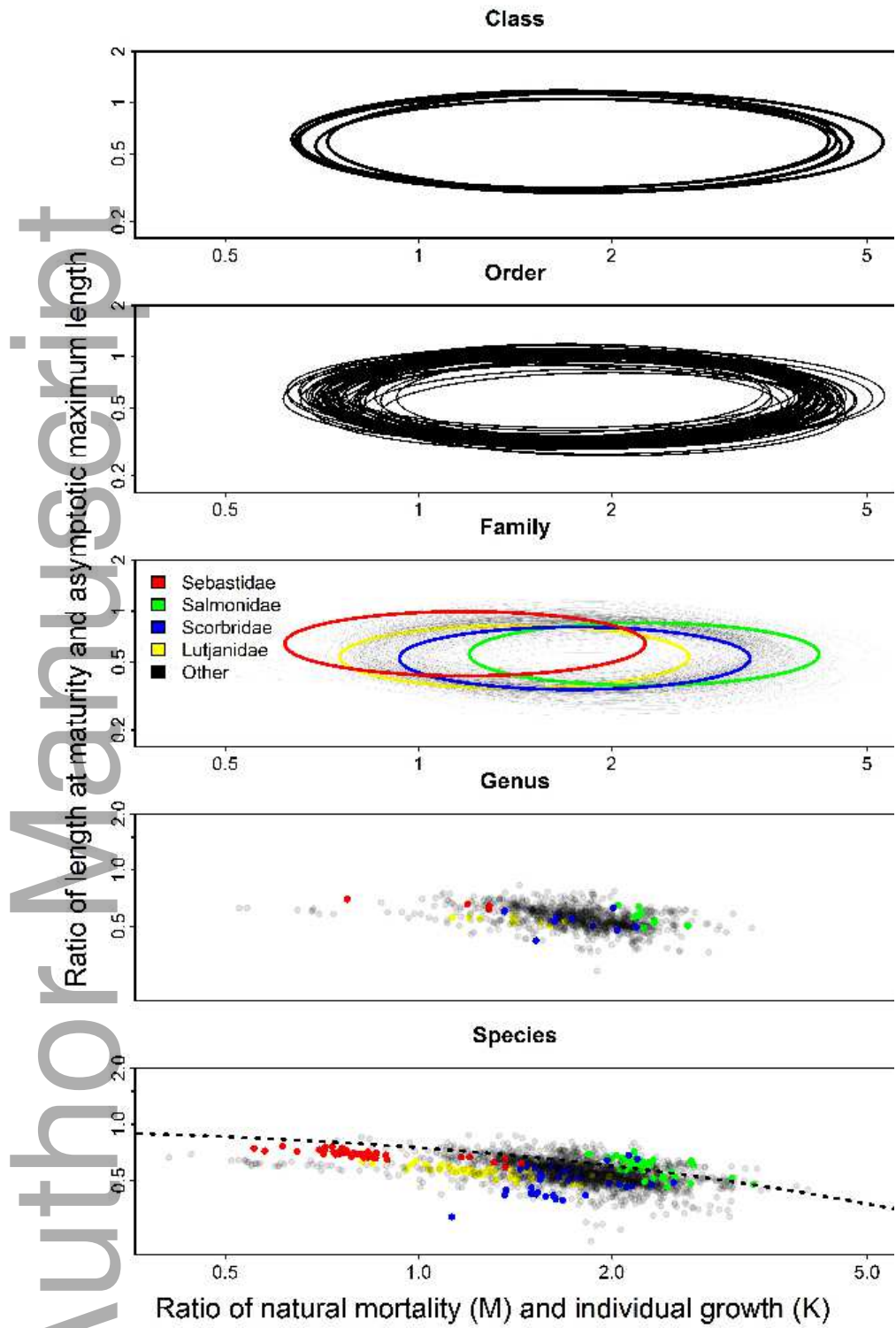
eap_1606_f2.jpg

eap_1606_f3.jpg
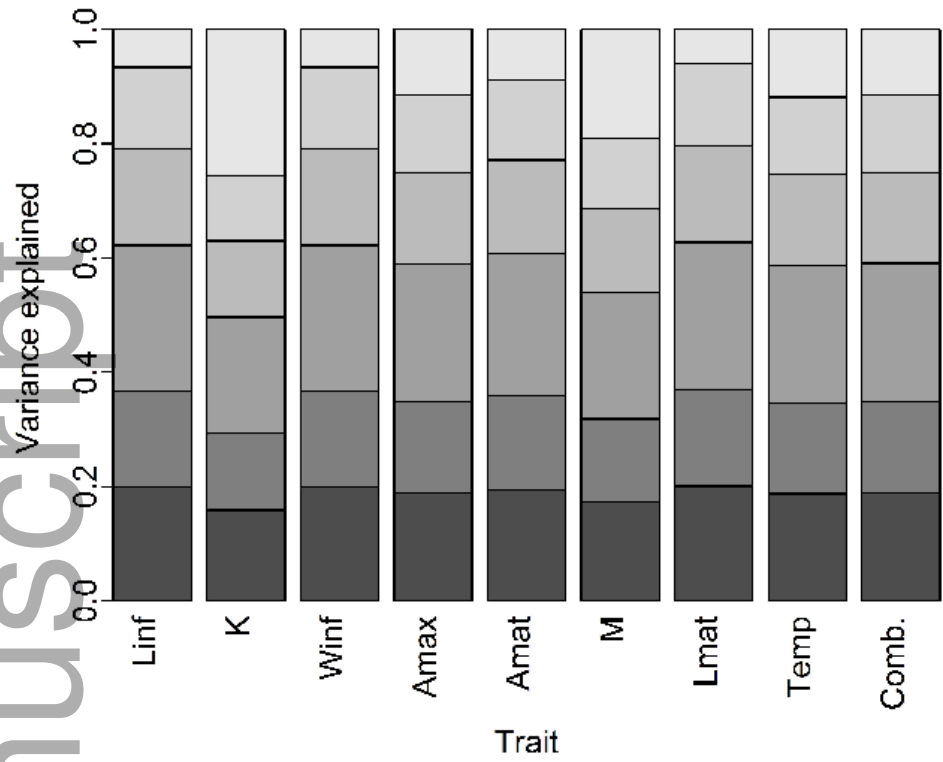
eap_1606_f4.jpg

eap_1606_f5.jpg