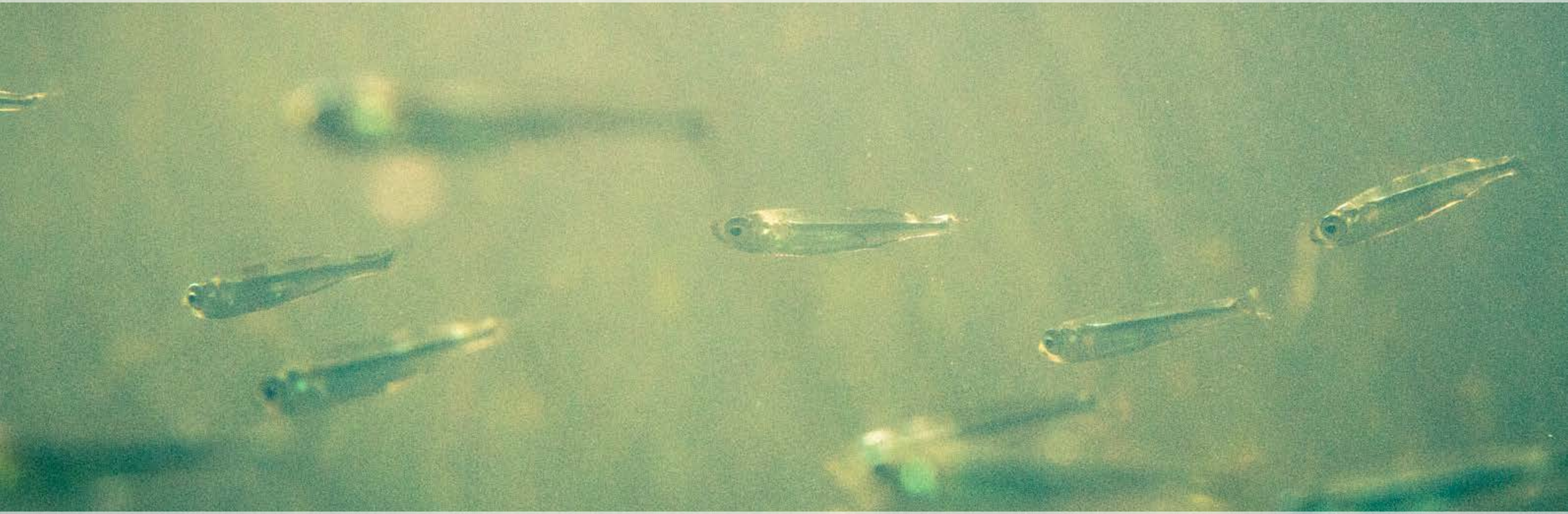
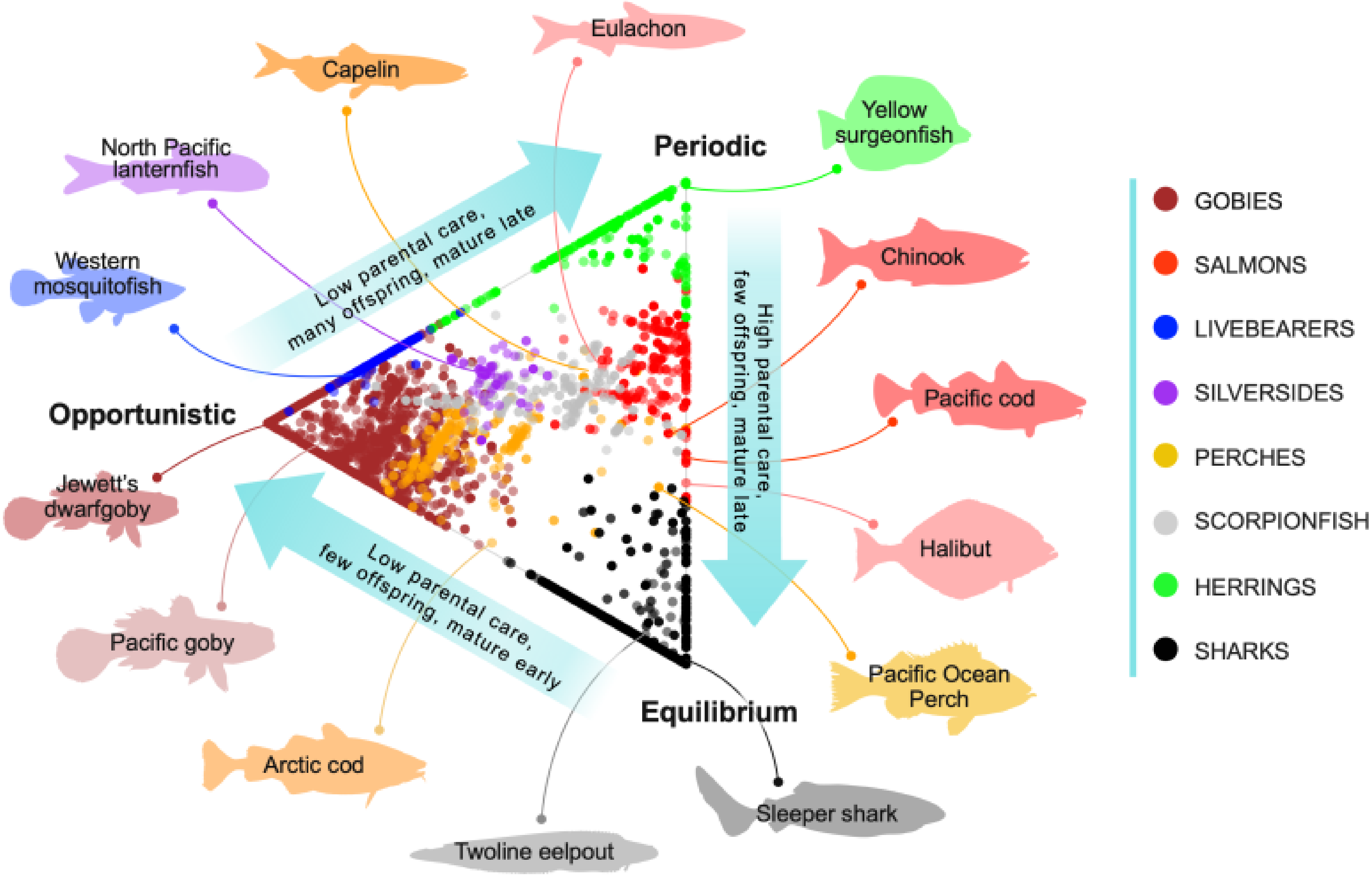
Phylogenetic imputation of reproductive, behavioral, and morphometric traits, and their use in joint species distribution models to understand community assembly



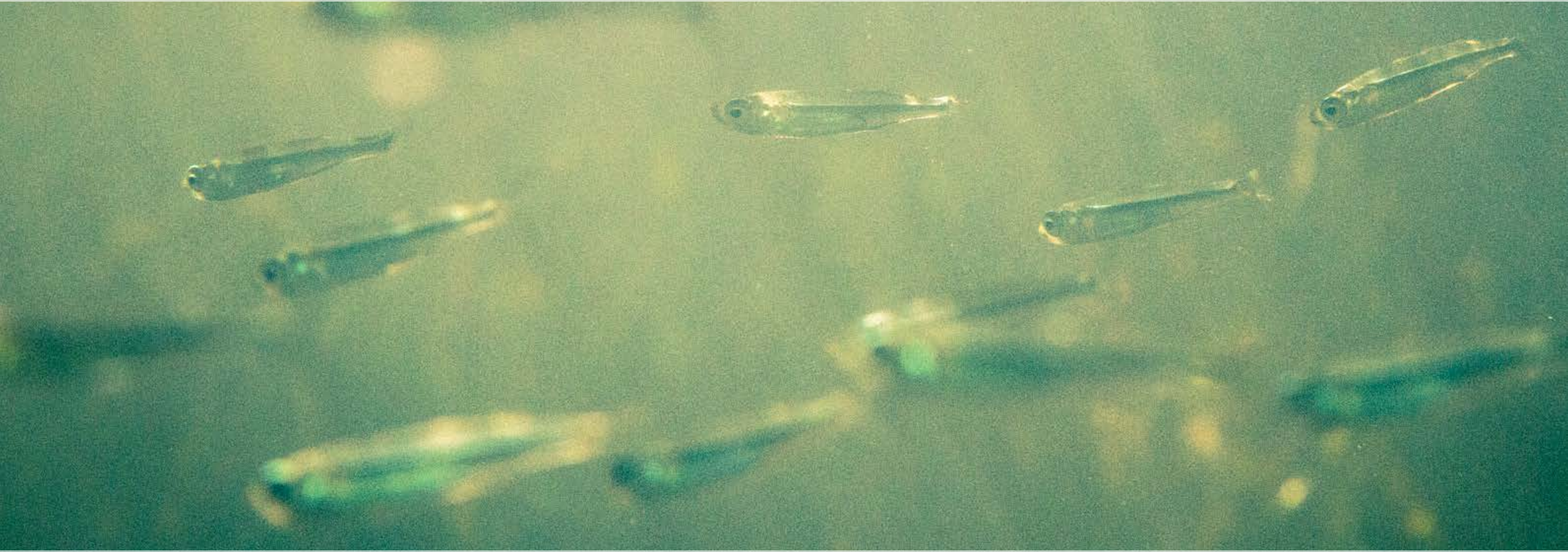
Thorson, Maureaud, Frelat, Mérigot, Friedman,
Palomares, Pinsky, Price, Wainwright
2023, Methods in Ecology and Evolution

© Jonny Armstrong



Eulachon
Capelin
Yellow surgeonfish
North Pacific lanternfish
Periodic
Western mosquitofish
Low parental care, many offspring, mature late
Chinook
High parental care, few offspring, mature late
Opportunistic
Pacific cod
Jewett's dwarfgoby
Halibut
Low parental care, few offspring, mature early
Pacific goby
Pacific Ocean Perch
Equilibrium
Arctic cod
Sleeper shark
Twoline eelpout
GOBIES
SALMONS
LIVEBEARERS
SILVERSIDES
PERCHES
SCORPIONFISH
HERRINGS
SHARKS

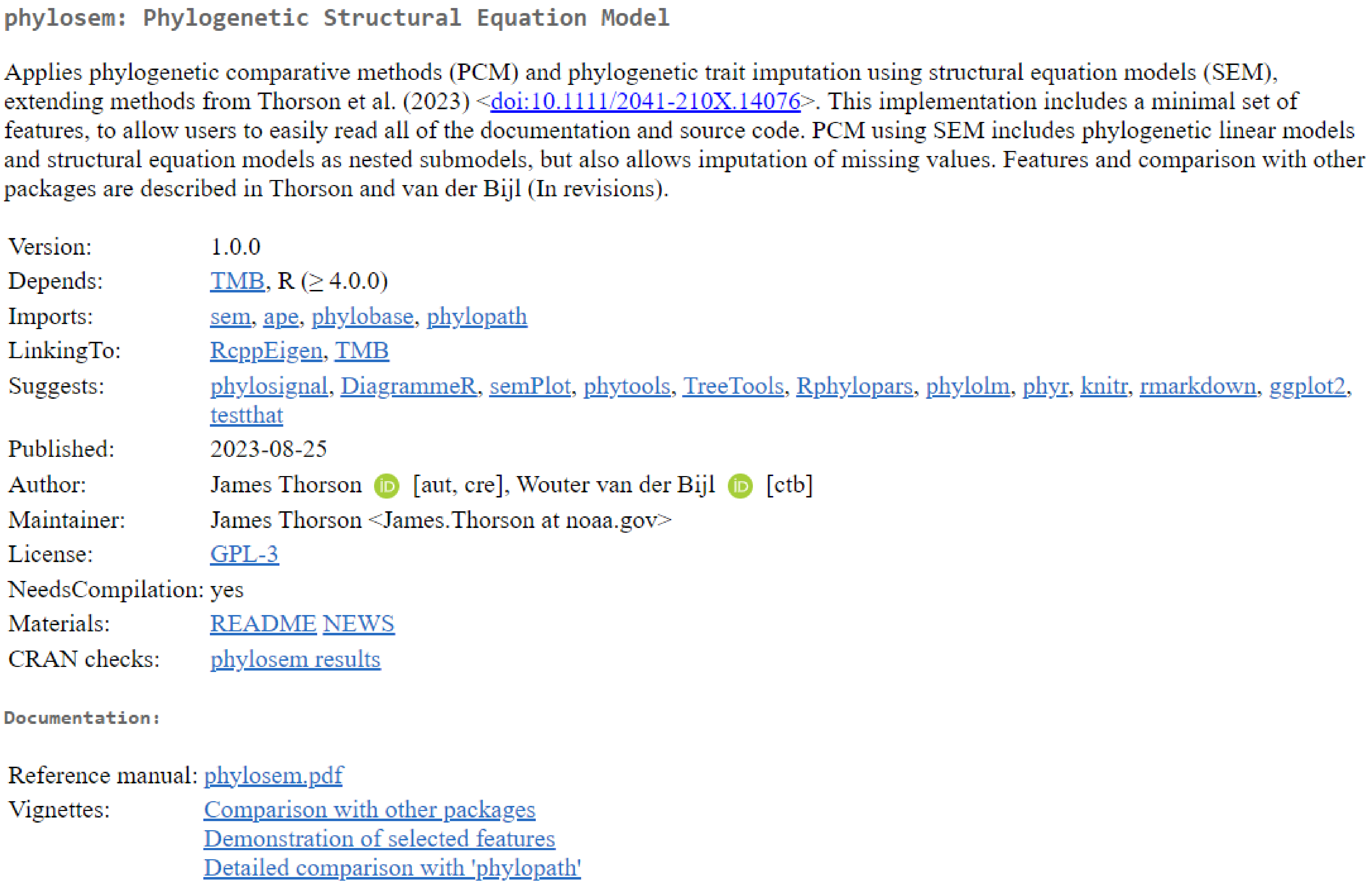
phylosem: a fast and simple R package for phylogenetic inference and trait imputation using structural equation models



James Thorson and Wouter van der Bijl

In press, Journal of Evolutionary Biology

© Jonny Armstrong



phylosem: Phylogenetic Structural Equation Model

Applies phylogenetic comparative methods (PCM) and phylogenetic trait imputation using structural equation models (SEM), extending methods from Thorson et al. (2023) <doi:10.1111/2041-210X.14076>. This implementation includes a minimal set of features, to allow users to easily read all of the documentation and source code. PCM using SEM includes phylogenetic linear models and structural equation models as nested submodels, but also allows imputation of missing values. Features and comparison with other packages are described in Thorson and van der Bijl (In revisions).

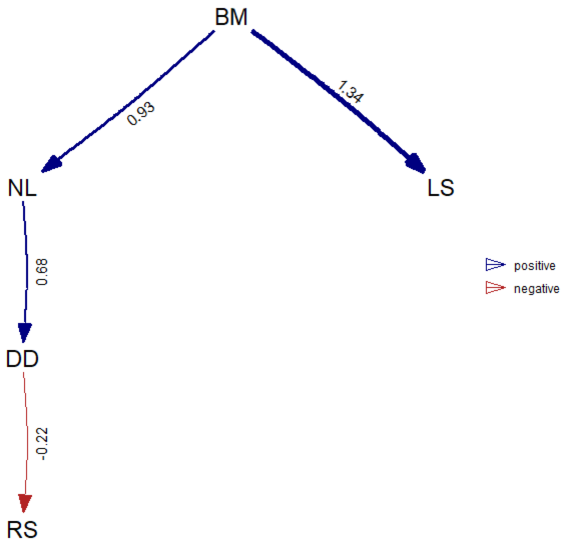
Version:	1.0.0
Depends:	TMB, R (≥ 4.0.0)
Imports:	sem, ape, phylobase, phylopath
LinkingTo:	RcppEigen, TMB
Suggests:	phylosignal, DiagrammeR, semPlot, phytools, TreeTools, Rphylopars, phylolm, phyr, knitr, rmarkdown, ggplot2, testthat
Published:	2023-08-25
Author:	James Thorson [aut, cre], Wouter van der Bijl [ctb]
Maintainer:	James Thorson <James.Thorson at noaa.gov>
License:	GPL-3
NeedsCompilation:	yes
Materials:	README NEWS
CRAN checks:	phylosem results

Documentation:

Reference manual:	phylosem.pdf
Vignettes:	Comparison with other packages Demonstration of selected features Detailed comparison with 'phylopath'

User-friendly R-package *phylosem*:

- Input format:
  - *ape*: specify phylogeny / taxonomy trees
  - *sem:* specify structural equation models
- Output format:
  - *phylopath:* model selection and averaging
  - *phylolm:* coefficient estimates
  - *phylo4d*: plot traits and trees
  - *sem:* plot path coefficients

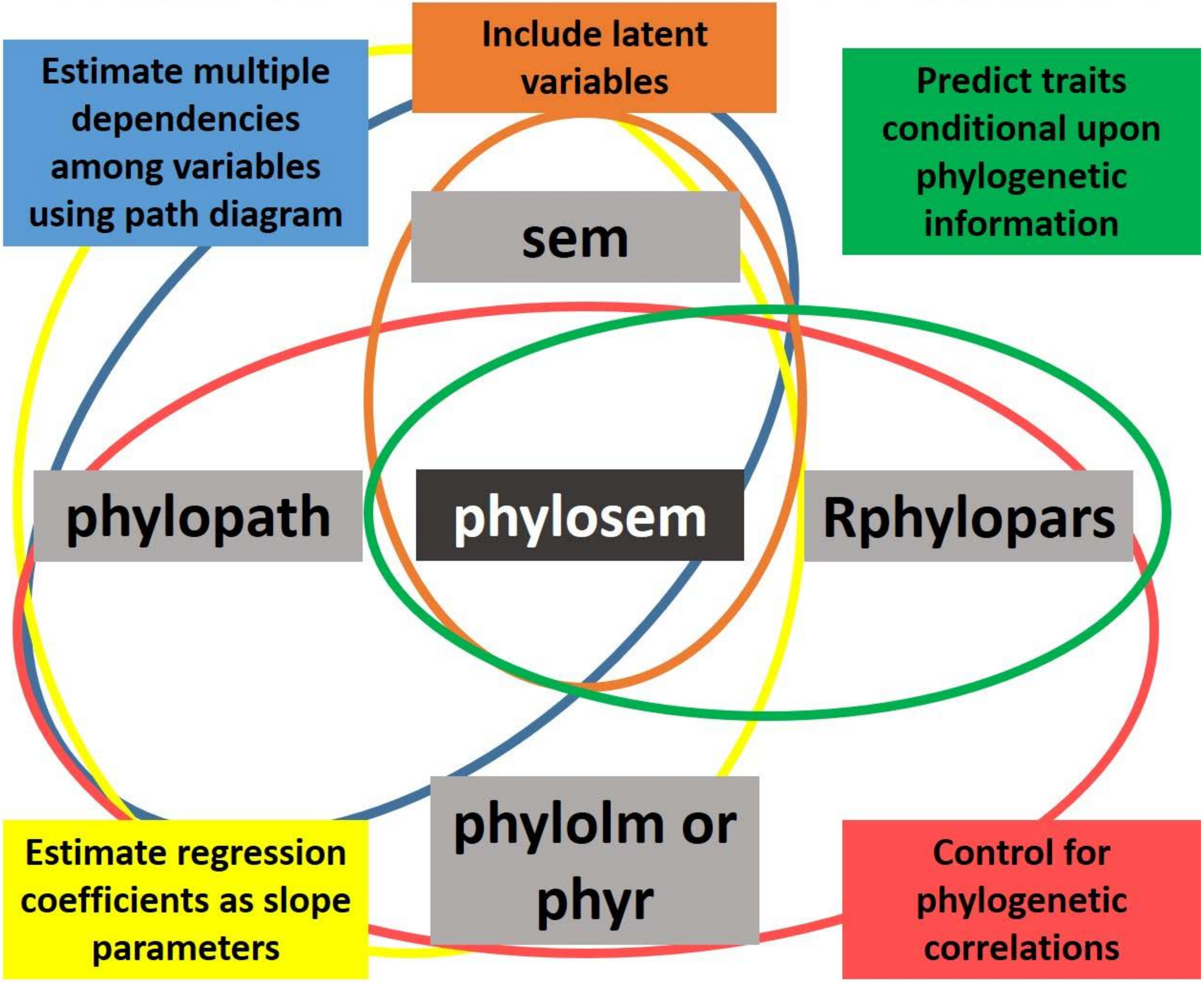


```
# Load data set
library(phylopath)

# Run phylosem
model = "
  DD -> RS, p1
  BM -> LS, p2
  BM -> NL, p3
  NL -> DD, p4
"
psem = phylosem( sem = model,
         data = rhino,
         tree = rhino_tree )

plot( as(psem,"fitted_DAG") )
```

Features available in alternative software



Estimate multiple dependencies among variables using path diagram

Include latent variables

Predict traits conditional upon phylogenetic information

sem

phylopath

phylosem

Rphylopars

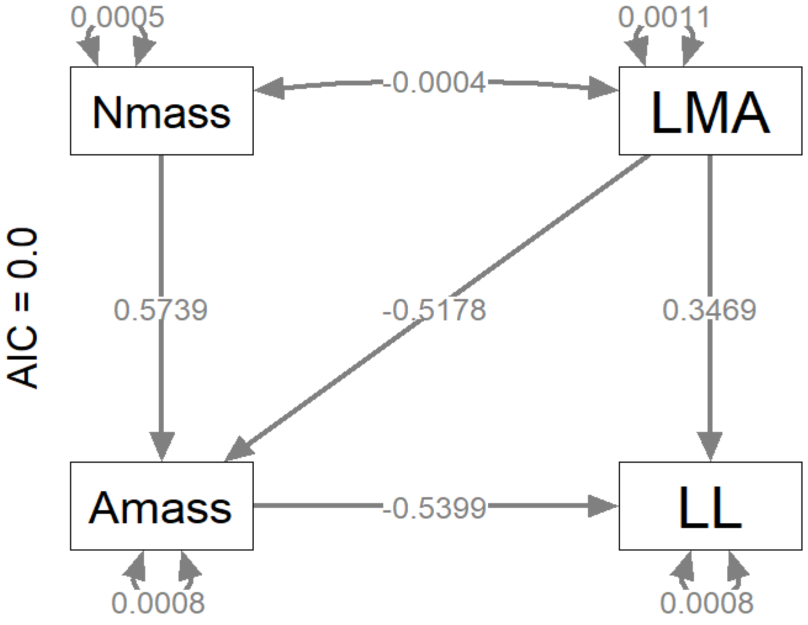
phylolm or phyr

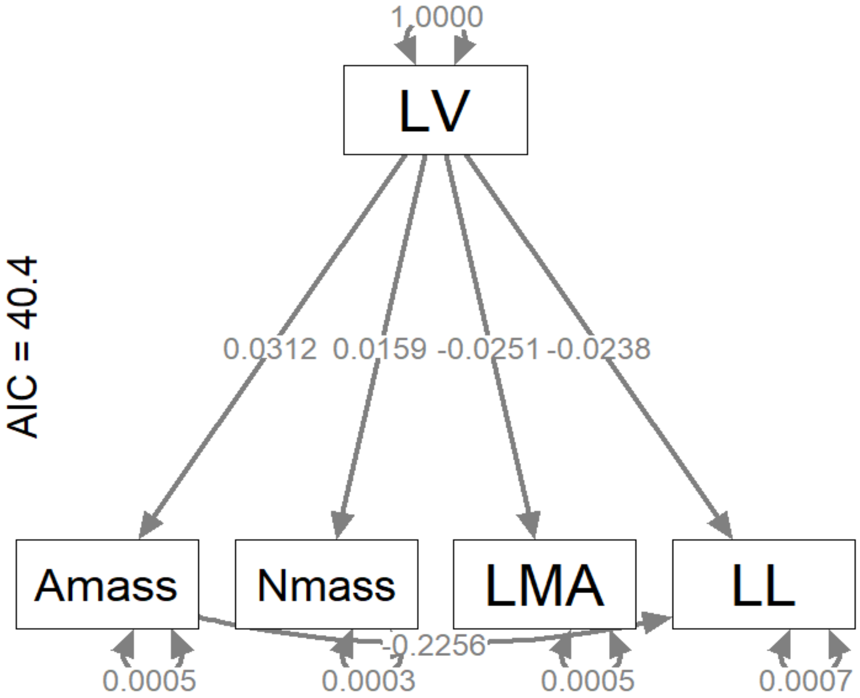
Estimate regression coefficients as slope parameters

Control for phylogenetic correlations

Resolve life-history tradeoffs:

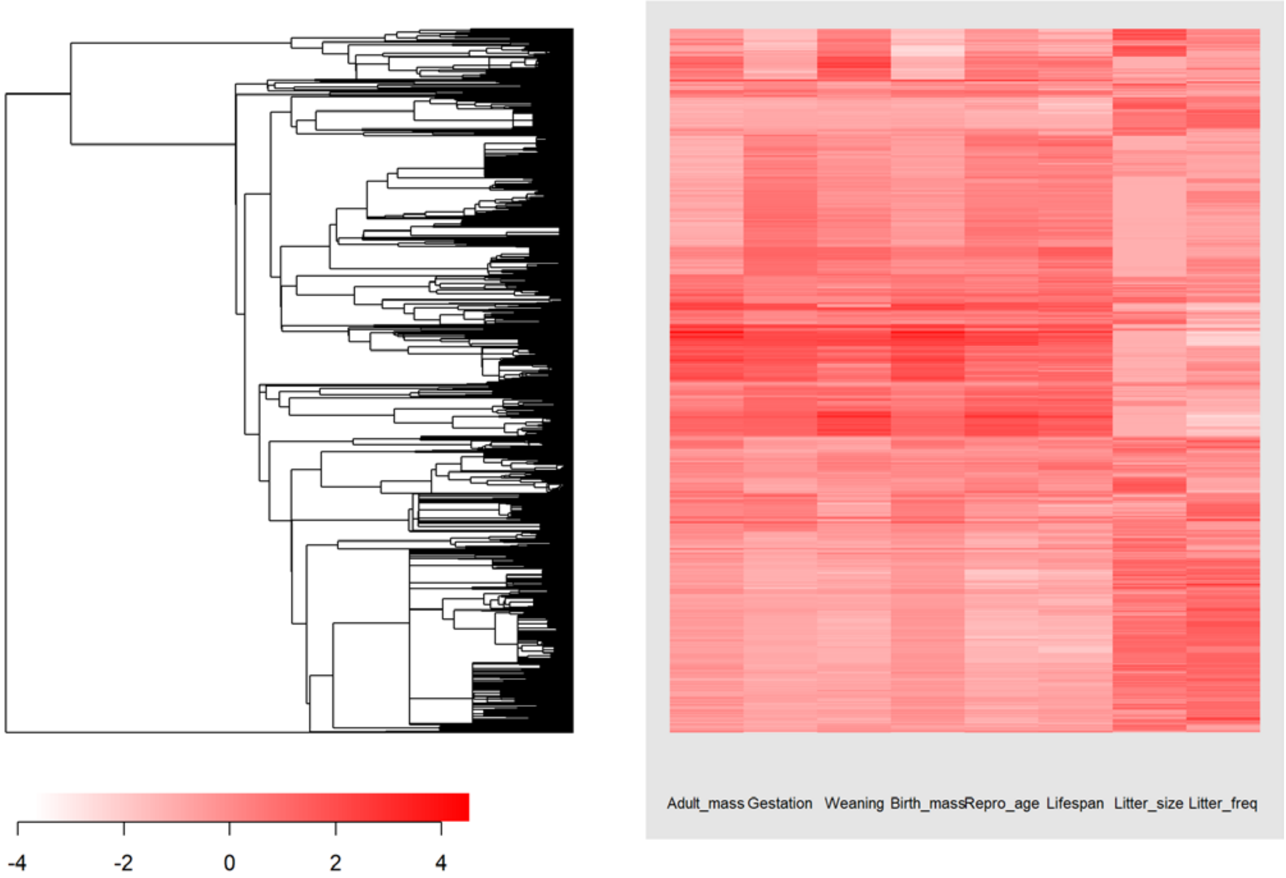
- Leaf energy budget
- two relationships hypothesized by Shipley et al. (2006) among four measured leaf traits
  - Amass: photosynthetic rate
  - Nmass: leaf nitrogen content
  - LMA: leaf mass per area
  - LL: leaf lifespan
- Estimates parsimony and imputes values jointly



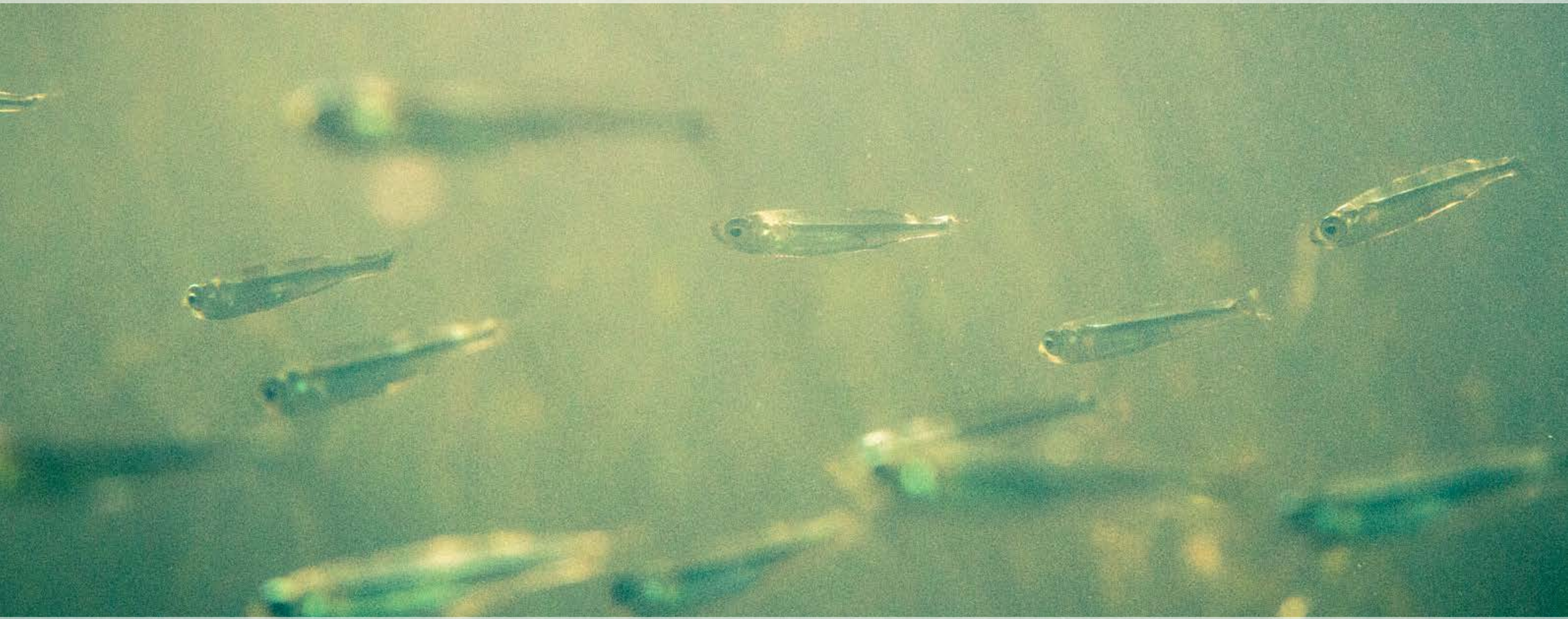


User-friendly R-package *phylosem*:

- Applied to PanTHERIA for 8 traits and 3734 species
- Takes 120 sec
  - Approx. 1000 times faster than joint-Gibbs method from Hassler et al. (2022)



Trees for fishes: The neglected role for phylogenetic comparative methods in fisheries science



Thorson

Revisions in review, Fish and Fisheries

PCM in fisheries

Comparative life-history methods in fisheries:

- Widely used to infer:
  - Biological reference point proxies (e.g., F_msy / M)
  - Natural mortality rate (e.g., Hoenig method)
  - Fecundity parameters (e.g., EJ Dick rockfish model)

- Could be used more for:
  - Ecosystem parameters (e.g., Q/B from Palomares and Pauly 1998)
  - Energy density (borrowed from similar species)

PCM in fisheries

Approaches:

1. Linear model

$$\mathrm{lm}(\ \log(M) \sim 1 + \log(Tmax)\ )$$

2. "Nested taxonomic" linear mixed model

$$\mathrm{lmer}(\ \log(M) \sim 1 + (1|Family) + (1|Genus) + \log(Tmax)\ )$$

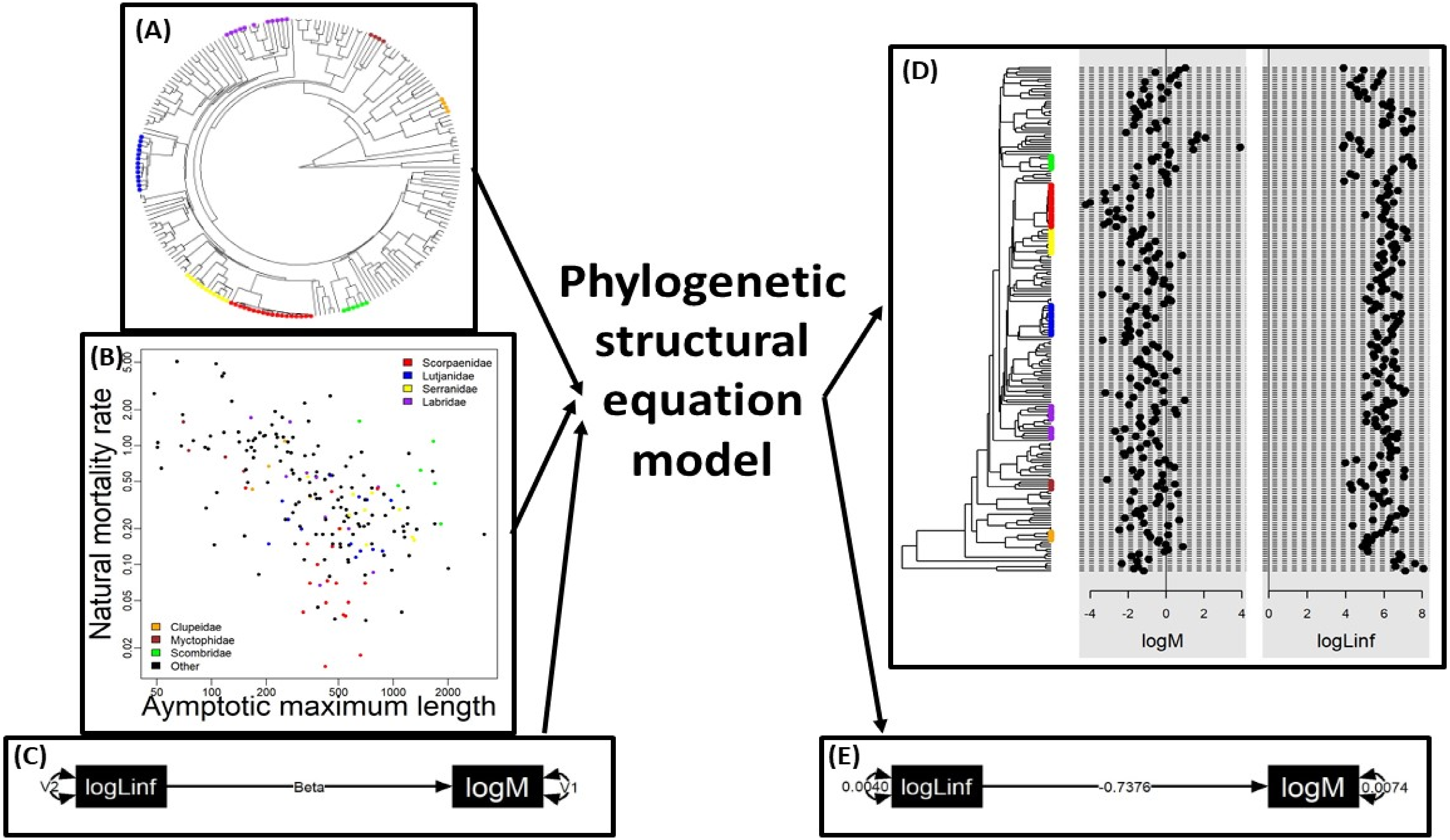
3. Phylogenetic linear model

$$\mathrm{plm}(\log(M) \sim 1 + \log(Tmax)\,, \mathrm{phy} = \mathrm{tree})$$

PCM in fisheries

Approaches:

4. Phylogenetic structural equation model

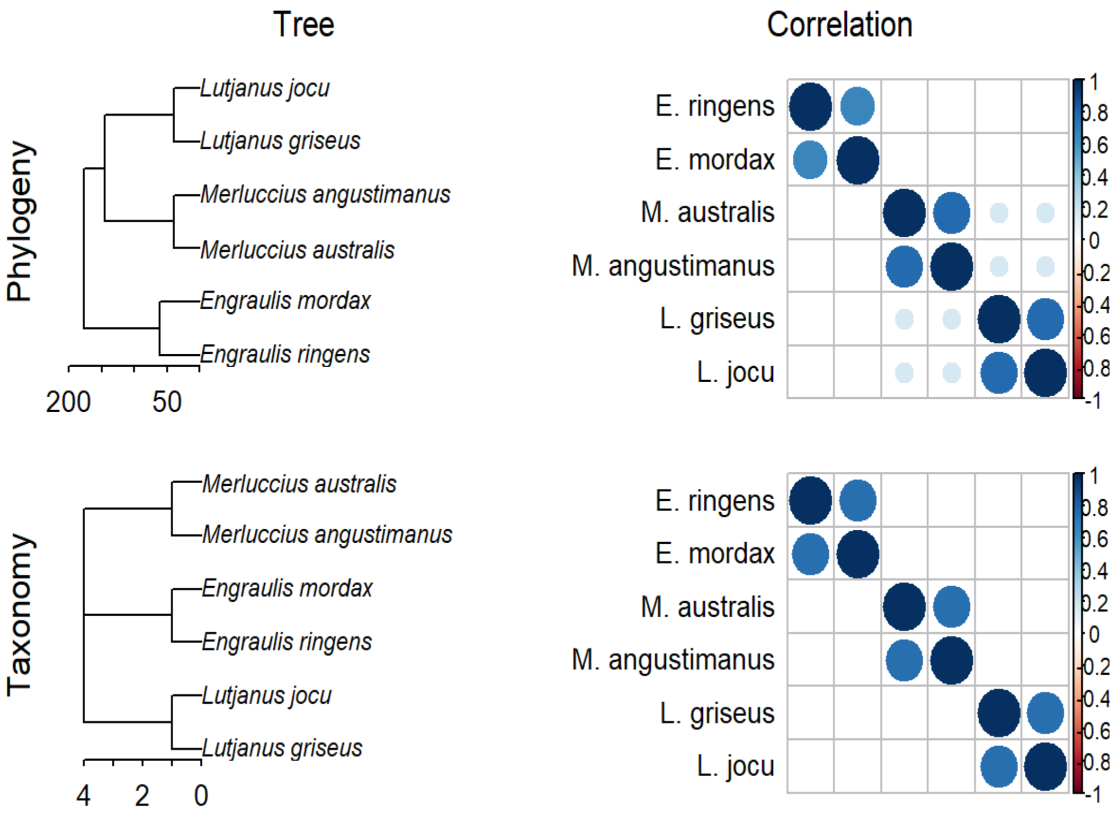


Phylogenetic structural equation model

PCM in fisheries

Advantages of PSEM

- Includes taxonomic mixed model as submodel
  - E.g., convert taxonomy to a tree (see on right)
- Estimates missing values and relationships jointly
- Avoids separate models depending on data availability



Natural mortality prediction

Gold-standard database:

- Then et al. (2015) contains 200 “quality” estimates of natural mortality rate
- Conventional analysis:
  - With longevity:

$$\log(M) \sim 1 + \log(Tmax)$$

  - Without longevity

$$\log(M) \sim 1 + \log(K) + \log(Linf)$$

- Phylogenetic SEM:

$$\log(K) \rightarrow \log(t_{max})$$
$$\log\left(L_{inf}\right) \rightarrow \log(t_{max})$$
$$\log(t_{max}) \rightarrow \log(M)$$

  - i.e., $\log(M)$ is independent of growth conditional on longevity

Natural mortality prediction

Alternative models

- Model #1: Exclude natural mortality
- Model #2: Exclude natural mortality and longevity

Jackknife experiment

1. Exclude data and fit model
2. Predict excluded value and compare with known value
3. Repeat for each datum

Natural mortality prediction

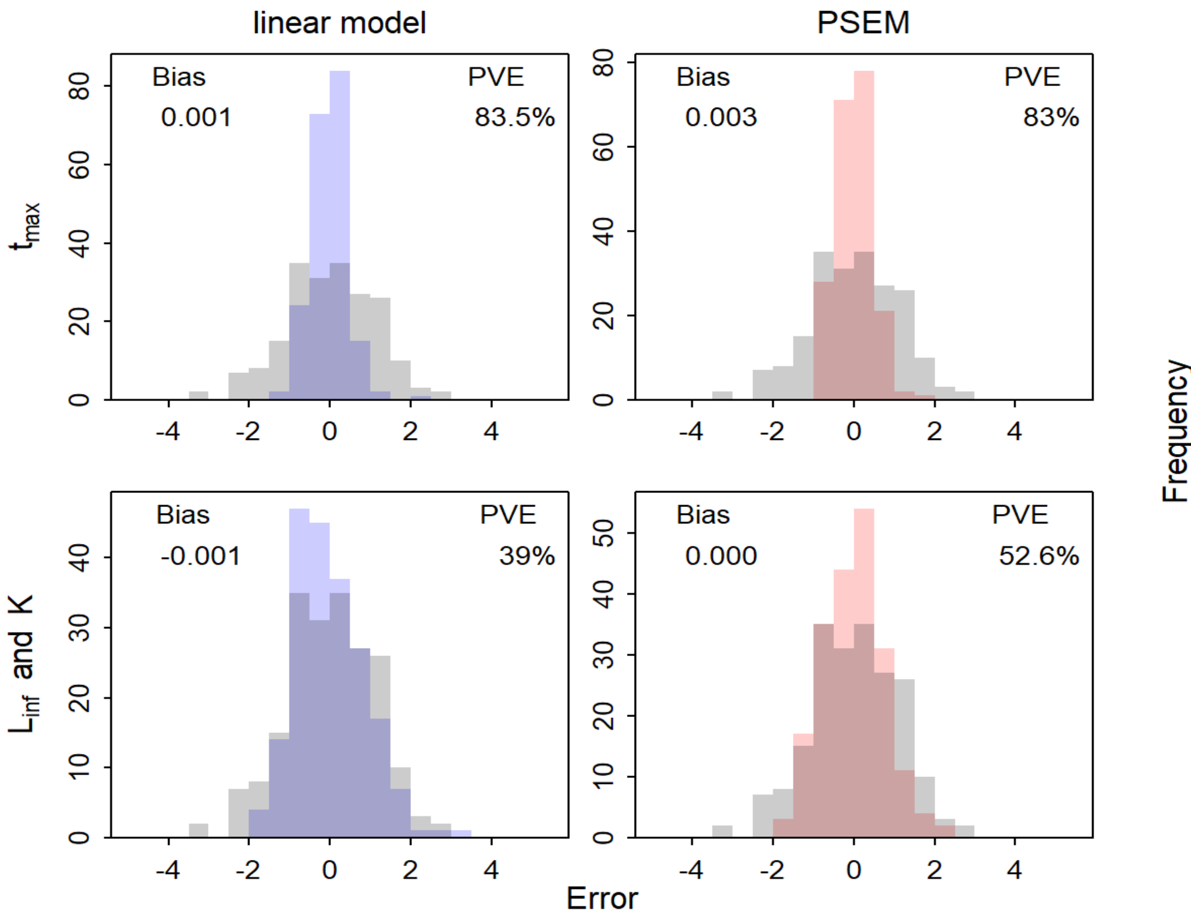
Using longevity:

- … results in high PVE
- … has little phylogenetic signal

Using growth parameters:

- … results in lower PVE
- … shows large improvement from using PSEM

PVE: Percent variance explained
(0% is bad; 100% is good)



Grey: Null model (prediction = mean of data)
Blue: Linear model
Red: Phylogenetic structural equation model

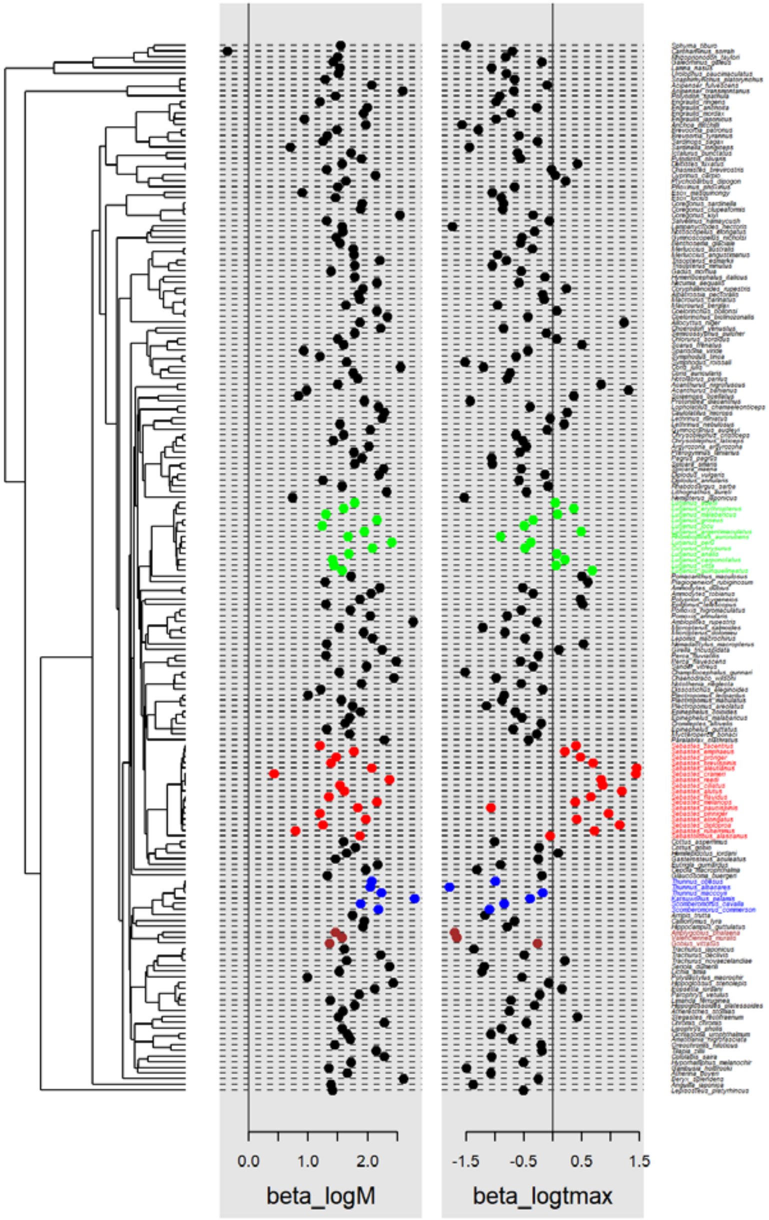
Natural mortality prediction

Beta_logtmax

- Intercept when predicting Tmax from Linf and K
- High for rockfishes

Beta_logM

- Intercept when predicting logM from logTmax
- Not evolutionarily conserved



Natural mortality prediction

Application to AFSC flatfishes (provided by Meaghan Bryan) and Pacific cod (provided by Steve Barbeaux)

Scientific.name	Common name	Current M	Amax (average of top 5 speciments)	K (estimated or assumed)	Linf (estimated or assumed)	Sex	PSEM using longevity	PSEM using growth
Reinhardtius hippoglossoides	Greenland turbot	0.112	48.8	0.109604	891.318	Female	0.119912	0.194795
Reinhardtius hippoglossoides	Greenland turbot	0.112	48.8	0.178908	707.48	Male	0.119897	0.264549
Atheresthes evermanni	Kamchatka flounder	0.11	53	0.098	796	Female	0.097394	0.218989
Atheresthes evermanni	Kamchatka flounder	0.11	53	0.149	607.3	Male	0.097483	0.29321
Lepidopsetta polyxystra	northern rock sole central	0.2	31.8	0.225	503.84	Female	0.172265	0.325518
Lepidopsetta polyxystra	northern rock sole west	0.2	31.8	0.227	404.014	Female	0.172279	0.358415
Lepidopsetta polyxystra	northern rock sole central	0.232	31.8	0.285	412.687	Male	0.172288	0.39155
Lepidopsetta polyxystra	northern rock sole west	0.254	31.8	0.342	333.564	Male	0.172322	0.462853
Lepidopsetta bilineata	southern rock sole central	0.2	28	0.158	515.768	Female	0.195973	0.277147
Lepidopsetta bilineata	southern rock sole west	0.2	28	0.201	460.14	Female	0.195982	0.322206
Lepidopsetta bilineata	southern rock sole central	0.253	28	0.181	414.74	Male	0.195986	0.321794
Lepidopsetta bilineata	southern rock sole west	0.271	28	0.267	371.75	Male	0.196011	0.397816
Gadus macrocephalus	Pacific cod	0.34	14	0.11	1129.58	Both	0.386623	0.201187
Gadus macrocephalus	Pacific cod	0.435	14	0.109	1132.17	Both	0.386626	0.200212

Acknowledgements

Collaborators on related work

- Steve Munch
- Olaf Jensen
- Jason Cope
- Wesley Patrick
- Jin Gao
- Daniel Pauly
- Rainer Froese

Co-authors

- Maureaud, Frelat, Mérigot, Friedman, Palomares, Pinsky, Price, Wainwright
- Wouter van der Bijl

Explorations

- Steve Barbeaux
- Meaghan Bryan

Questions/Requests for Plan Team

Endorsement:

- Any concerns or additional research needed before using phylosem for imputing life-history parameters in council process?

Prioritization:

- Is it worth the effort to explore estimators for M across a larger set of AFSC stocks?
- Any suggests for parameters that are worth imputing (e.g., thermal minima/maxima, age at maturity, etc)?