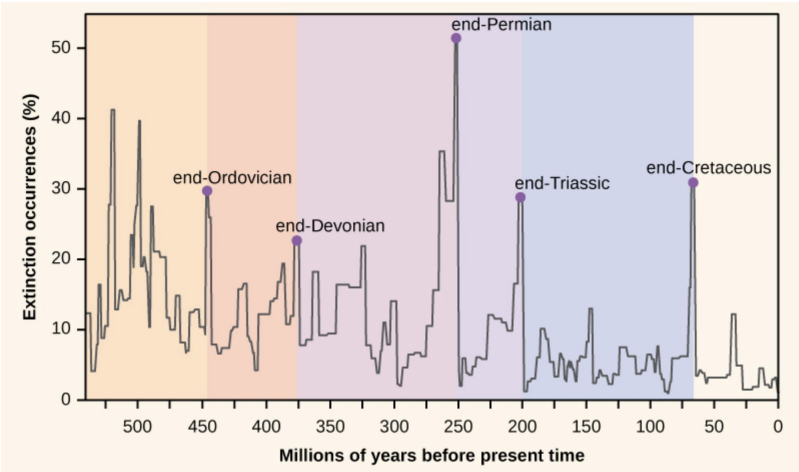
# Using Machine Learning Models and Logistic Regression Analyses to Develop a Comprehensive Understanding of Extinction Risk For Marine Animal Phyla Across the Paleozoic

**Adarsh S. Ambati, Anya Sengupta, Theo Chiang, Dr. Pedro Monarrez, Michael Pimentel-Galvan, Dr. Noel Heim, Dr. Jonathan Payne**

# Basic Info/Background

- Extinction is volatile and relies on many ecological and biological variables
- Nature of extinction risk varies across geologic time; we focus on the Paleozoic Era




542 mya — **Cambrian**
Cambrian Explosion

488 mya — **Ordovician**
Late Ordovician Mass Extinction

444 mya — **Silurian**

416 mya — **Devonian**
The Devonian Mass Extinction

359 mya — **Carboniferous**

299 mya — **Permian**
End Permian Mass Extinction -

252 mya

# Questions

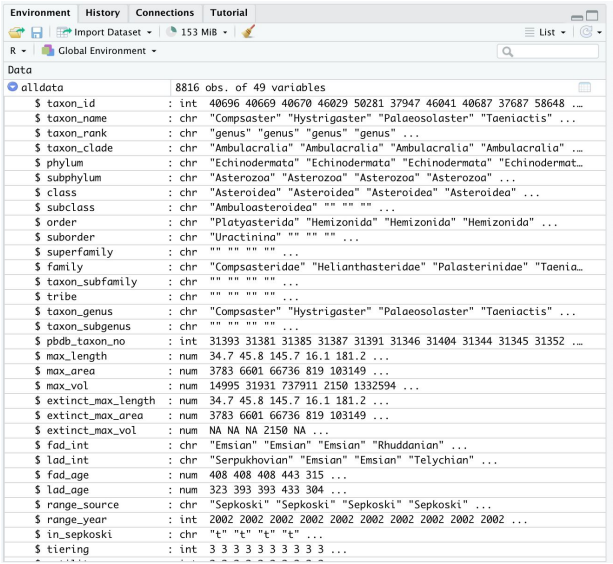
1. Were certain taxonomic groups of organisms preferentially selected for extinction during the Paleozoic era?

2. Were certain types of organisms preferentially selected for extinction during the Paleozoic era?

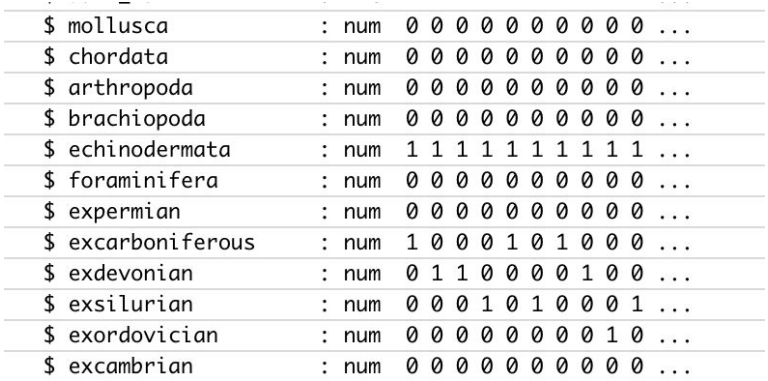
3. Is it possible to predict when a particular genus goes extinct during the Paleozoic using characteristics/descriptors?

# Methodology

- Phyla - Echinodermata, Mollusca, Chordata, Arthropoda, Brachiopoda
- Descriptors - buffering, feeding patterns, motility, oceanic tiering, respiratory organ type, circulatory system type, length, surface area, volume
- Regularized binomial regression models(Step 3) + logistic binomial regression(Step 1/2 )
  - Built Using R
  - Uses Stanford Earth Body Size Dataset (n=8816)



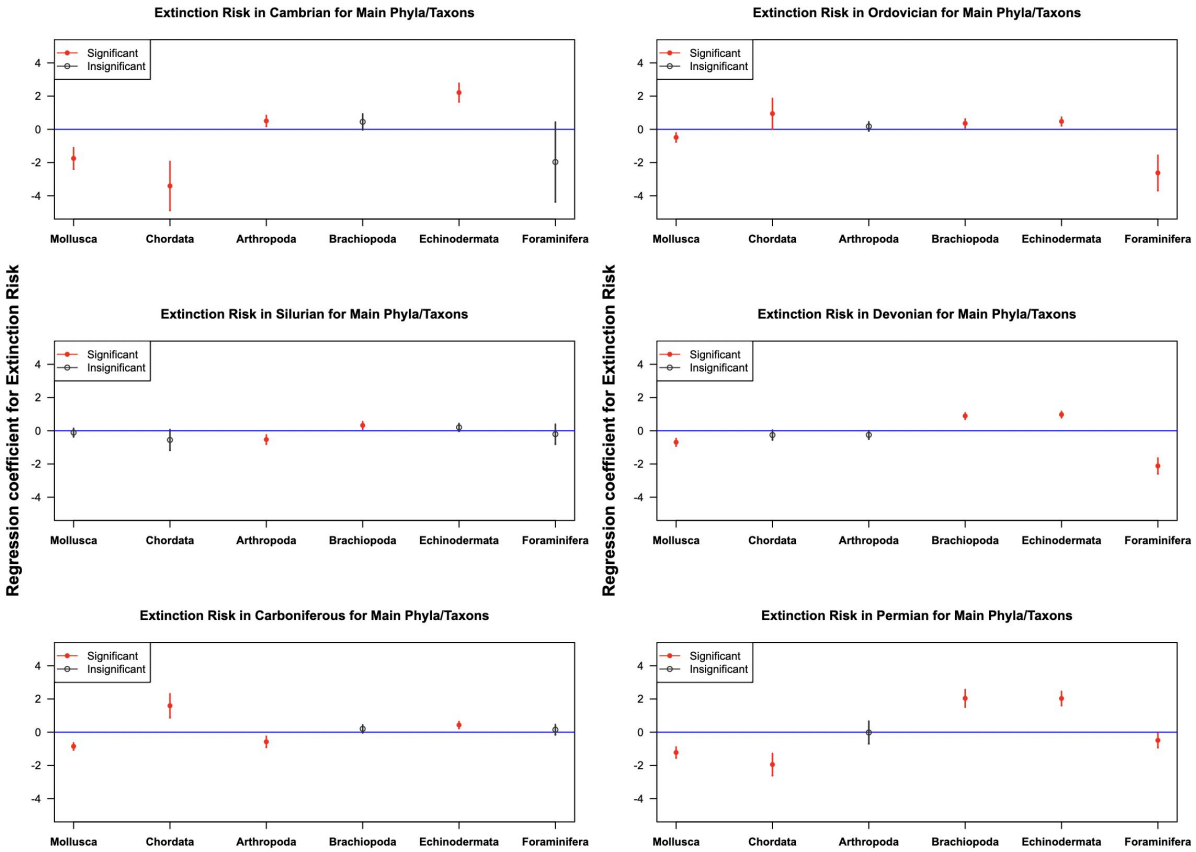
# Methodology Cont.



```
$ mollusca          : num  0 0 0 0 0 0 0 0 0 0 ...
$ chordata          : num  0 0 0 0 0 0 0 0 0 0 ...
$ arthropoda        : num  0 0 0 0 0 0 0 0 0 0 ...
$ brachiopoda       : num  0 0 0 0 0 0 0 0 0 0 ...
$ echinodermata     : num  1 1 1 1 1 1 1 1 1 1 ...
$ foraminifera      : num  0 0 0 0 0 0 0 0 0 0 ...
$ expermian         : num  0 0 0 0 0 0 0 0 0 0 ...
$ excarboniferous   : num  1 0 0 0 1 0 1 0 0 0 ...
$ exdevonian        : num  0 1 1 0 0 0 0 1 0 0 ...
$ exsilurian        : num  0 0 0 1 0 1 0 0 0 1 ...
$ exordovician      : num  0 0 0 0 0 0 0 0 1 0 ...
$ excambrian        : num  0 0 0 0 0 0 0 0 0 0 ...
```

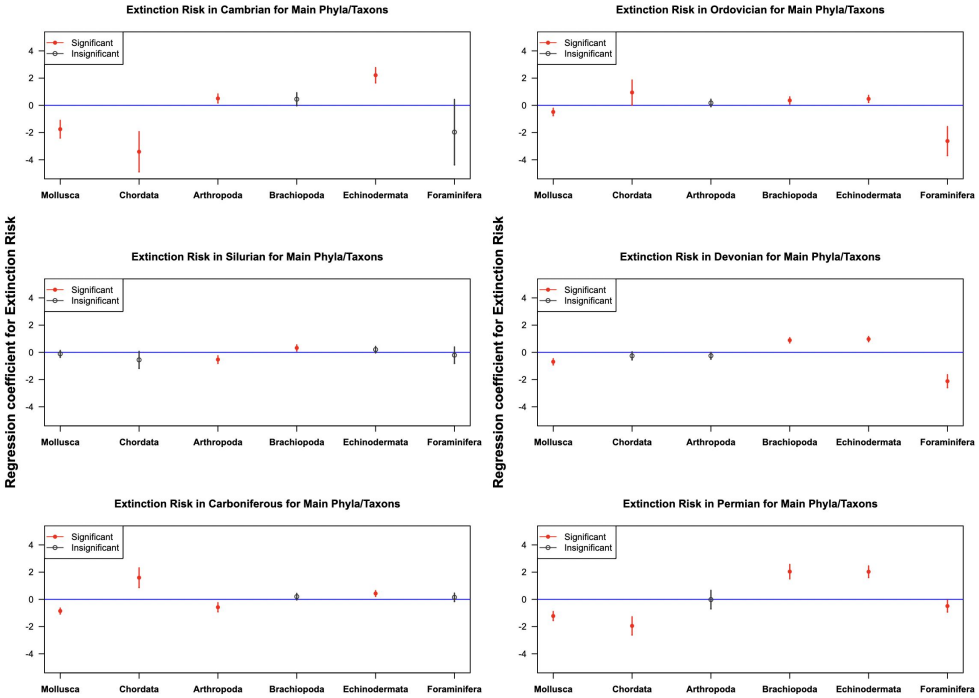
- Step 1: Binomial Logistic Regression Analysis on Phyla/Class/Order during each stage of Paleozoic identifying likeliness of extinction on each class
- Step 2: Binomial Regression Analysis on Phyla/Class/Order during each stage of Paleozoic identifying likeliness of extinction on each genus descriptor (ie predatory feeding, facultative motility, benthic tiering)
- Step 3: Developing simple machine learning model (using regularization) to predict whether a taxonomic group/specimen goes extinct in a specific period

# Step 1:

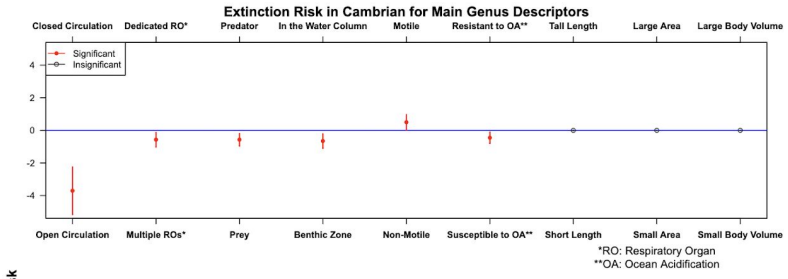

Extinction Risk in Cambrian for Main Phyla/Taxons
Extinction Risk in Ordovician for Main Phyla/Taxons
Extinction Risk in Silurian for Main Phyla/Taxons
Extinction Risk in Devonian for Main Phyla/Taxons
Extinction Risk in Carboniferous for Main Phyla/Taxons
Extinction Risk in Permian for Main Phyla/Taxons
Regression coefficient for Extinction Risk
Regression coefficient for Extinction Risk
Significant
Insignificant
4
2
0
-2
-4
Mollusca
Chordata
Arthropoda
Brachiopoda
Echinodermata
Foraminifera

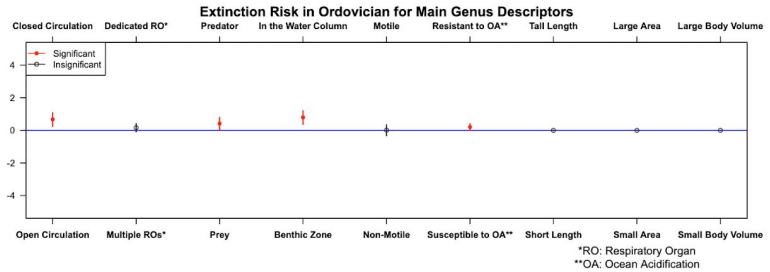
# Step 1: Key Findings

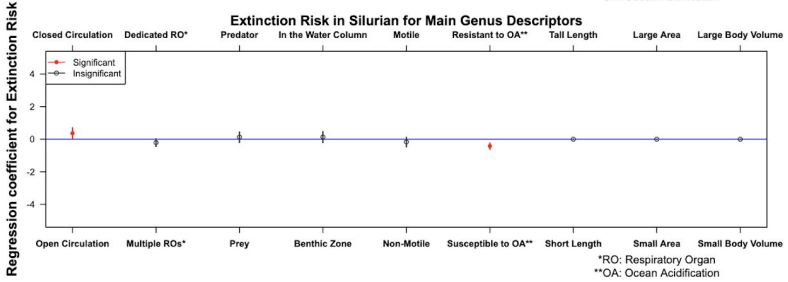
- **23** of **36** data points were significant
  - **12** data points had a significant greater extinction risk
  - **11** data points were significantly selected for survival.
- Patterns
  - Mollusca — consistently selected for survival
  - Brachiopoda— (major extinction events in Devonian and Permian, reflected in graphs)
  - Echinodermata — consistently selected for extinction
- **Extinction Risk was not uniformly felt across all major phyla in the Paleozoic**

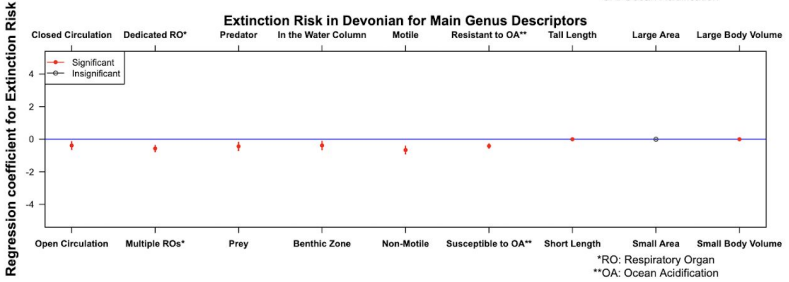


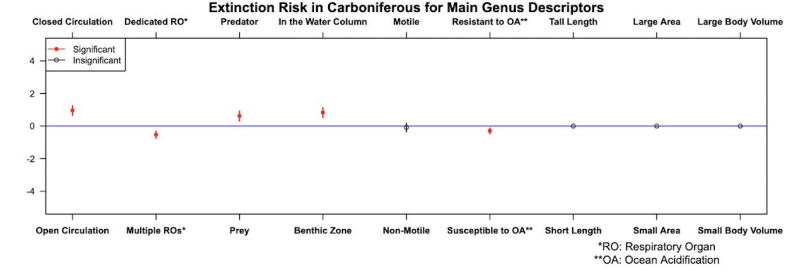
# Step 2:

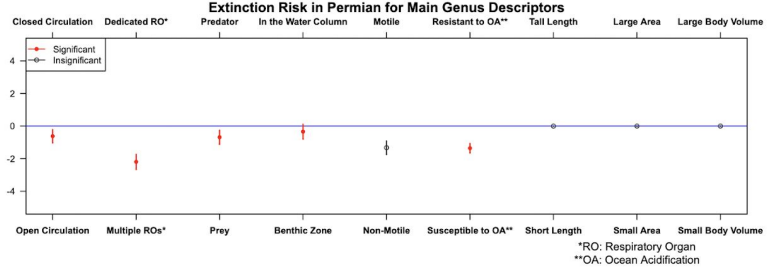

Extinction Risk in Cambrian for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification


Extinction Risk in Ordovician for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification


Regression coefficient for Extinction Risk
Extinction Risk in Silurian for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification

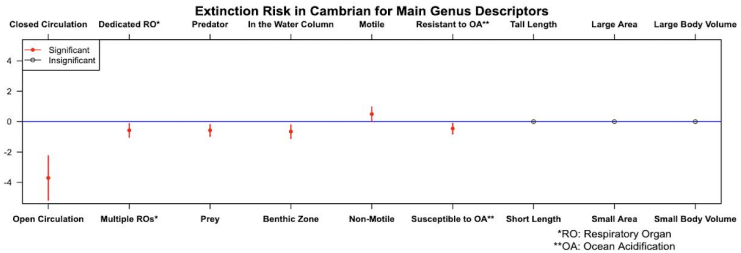

Regression coefficient for Extinction Risk
Extinction Risk in Devonian for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification

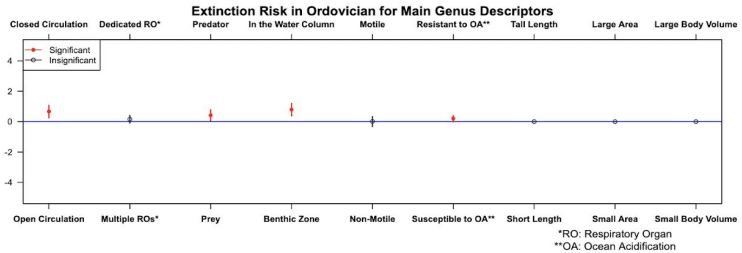

Extinction Risk in Carboniferous for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification

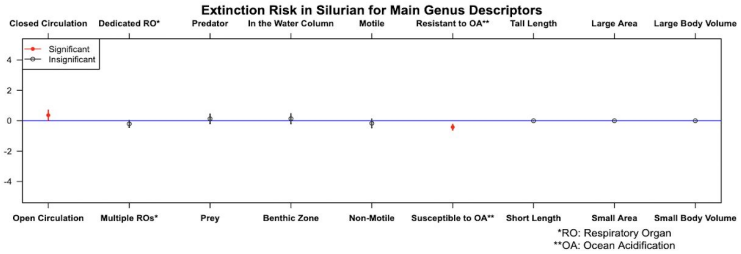

Extinction Risk in Permian for Main Genus Descriptors
Closed Circulation
Dedicated RO*
Predator
In the Water Column
Motile
Resistant to OA**
Tall Length
Large Area
Large Body Volume
Significant
Insignificant
Open Circulation
Multiple ROs*
Prey
Benthic Zone
Non-Motile
Susceptible to OA**
Short Length
Small Area
Small Body Volume
*RO: Respiratory Organ
**OA: Ocean Acidification

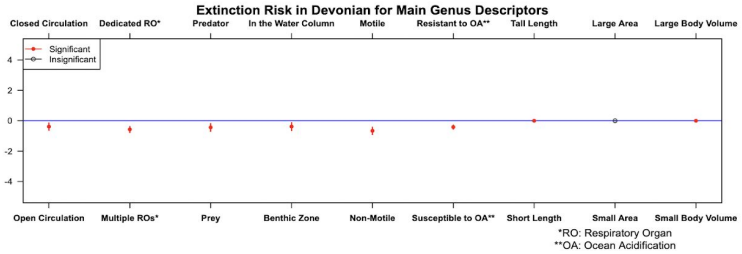
# Step 2: Key Findings

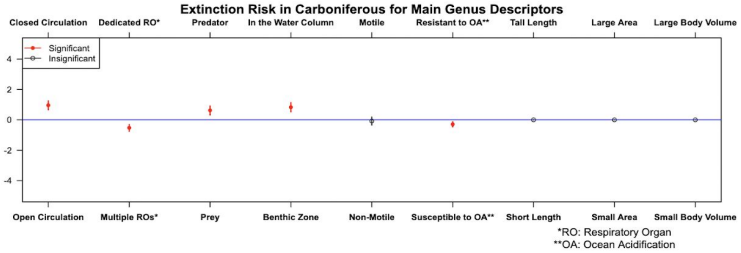
- **30** of **54** data points were significant
- Patterns
  - **Body Size**
  - Circulatory Systems (very volatile)
  - Respiratory Organ Type

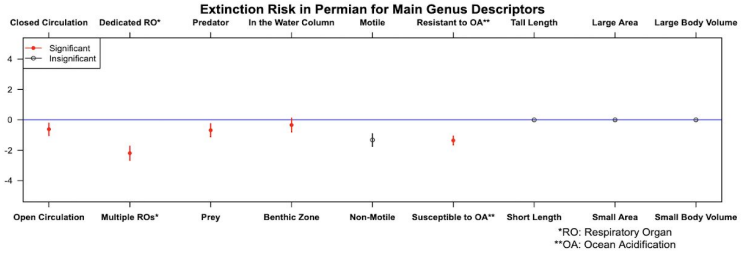


# Step 3: Machine Learning Model

- 6 binomial regression models with regularization for each period
  - Regularization is helps in two major ways:
    - Reducing the variance of the model so as to not cause it to become overfit to the training data
    - Helps determine the features that causes the model to increase in variance and removes or shrinks their contribution to the model. (Essentially, features that do not predict have their coefficients reduced)
- Features:
  - Circulatory systems
  - Type of Respiratory Organ
  - Feeding (Predator or Prey)
  - Tiering (Water Column or Benthic)
  - Motility (Freely Moving or Non-motile)
  - Ability to withstand Ocean Acidification
  - Maximum length
  - Maximum Area
  - Maximum Volume
- **Predicting Outcome:** whether or not a genus went extinct in a particular period

# Step 3: Machine Learning Model

Three types of regression models (lasso, elastic net, and ridge regression) were tested:

- **Lasso:** Focused on Feature Elimination (penalizes by removing)
- **Ridge Regression:** Focused on Feature Coefficient Reduction (penalizes by shrinking)
- **Elastic Net:** middle of both
- Nomenclature: Lasso : $\alpha = 1$; Ridge : $\alpha = 0$; Elastic net: $0 < \alpha < 1$
- The best model identified by running the program for $\alpha$ values between 0-1

# Step 3: Machine Learning Model

| circ | respOrgan | feeding | tiering | motility | oceanacidificationwithstand | max_length | max_area | calc_max_vol | excambrian |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 0 | 34.700000 | 3782.7603 | 1.499532e+04 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 181.200000 | 103149.2939 | 1.332594e+06 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 18.232000 | 1044.2837 | 2.613302e+03 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 55.450000 | 9659.4628 | 5.353129e+04 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 64.057376 | 12891.0460 | 7.920197e+04 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 37.300395 | 4370.9591 | 1.824551e+04 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 37.269746 | 4363.7789 | 1.820484e+04 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 40.702486 | 5204.6525 | 2.312418e+04 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 91.940000 | 26555.7700 | 2.112418e+05 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 41.897815 | 5514.8362 | 2.501452e+04 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 43.493046 | 5942.7783 | 2.768527e+04 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 10.141566 | 323.1171 | 5.316888e+02 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 45.810729 | 6593.0184 | 3.187570e+04 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 | 46.668914 | 6842.3497 | 3.352294e+04 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 27.050000 | 2298.7112 | 3.482660e+04 | 0 |

| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Cambrian | 0 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.1 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.2 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.3 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.4 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.5 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.6 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.7 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.8 | 0.01 | 0.92 | 0.93 | 0.8 |
| Cambrian | 0.9 | 0.01 | 0.08 | 0.5 | 0.79 |
| Cambrian | 1 | 0.01 | 0.91 | 0.5 | 0 |

| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Ordovician | 0 | 0.012589254 | 0.83 | 0.67 | 0.73 |
| Ordovician | 0.1 | 0.01 | 0.83 | 0.68 | 0.73 |
| Ordovician | 0.2 | 0.012589254 | 0.83 | 0.67 | 0.73 |
| Ordovician | 0.3 | 0.01 | 0.83 | 0.67 | 0.73 |
| Ordovician | 0.4 | 0.01 | 0.83 | 0.67 | 0.73 |
| Ordovician | 0.5 | 0.01 | 0.83 | 0.66 | 0.73 |
| Ordovician | 0.6 | 0.01 | 0.82 | 0.66 | 0.73 |
| Ordovician | 0.7 | 0.01 | 0.82 | 0.66 | 0.73 |
| Ordovician | 0.8 | 0.01 | 0.82 | 0.66 | 0.74 |
| Ordovician | 0.9 | 0.01 | 0.82 | 0.61 | 0.74 |
| Ordovician | 1 | 0.01 | 0.82 | 0.61 | 0.75 |

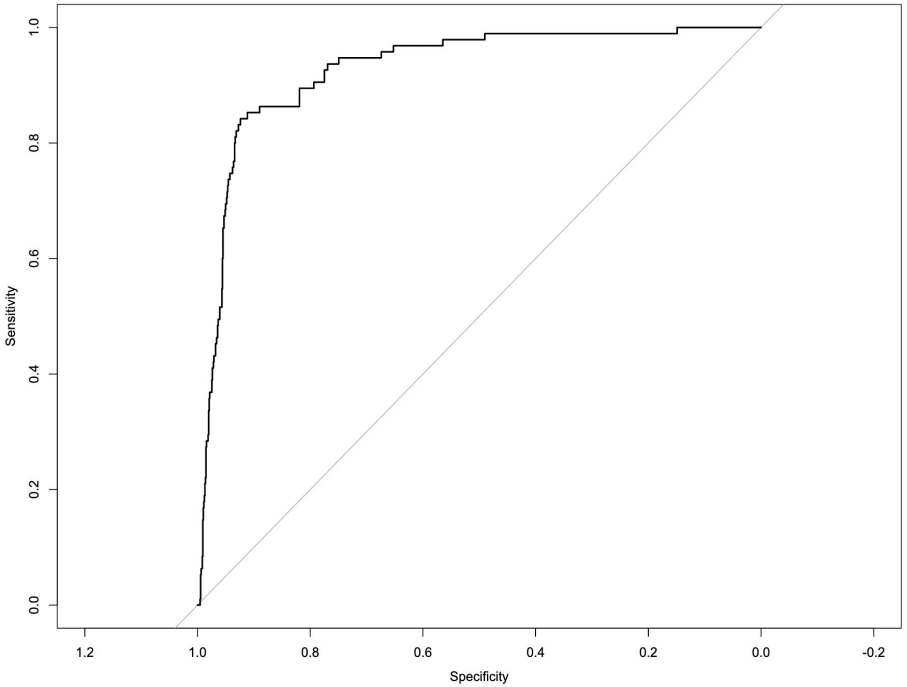
| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Silurian | 0 | 0.005416508 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.1 | 0.001219581 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.2 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.3 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.4 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.5 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.6 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.7 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.8 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 0.9 | 0.000578704 | 0.91 | 0.6 | 0.89 |
| Silurian | 1 | 0.000578704 | 0.91 | 0.6 | 0.89 |

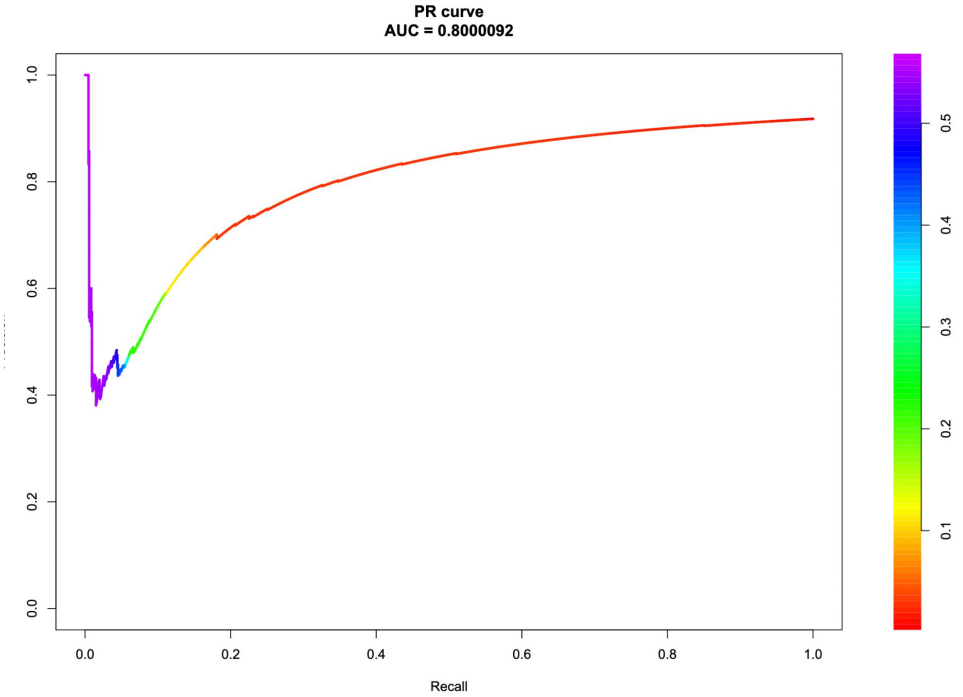
| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Devonian | 0 | 0.025118864 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.1 | 0.015848932 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.2 | 0.019952623 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.3 | 0.006309573 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.4 | 0.003981072 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.5 | 0.003981072 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.6 | 0.003162278 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.7 | 0.001584893 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.8 | 0.002511886 | 0.79 | 0.64 | 0.75 |
| Devonian | 0.9 | 0.001995262 | 0.79 | 0.64 | 0.75 |
| Devonian | 1 | 0.003162278 | 0.79 | 0.64 | 0.75 |

| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Carboniferous | 0 | 0.003541732 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.1 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.2 | 0.011383466 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.3 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.4 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.5 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.6 | 0.003541732 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.7 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.8 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 0.9 | 0.002915452 | 0.83 | 0.6 | 0.77 |
| Carboniferous | 1 | 0.002915452 | 0.83 | 0.6 | 0.77 |

| Period | Alpha Value | Lambda Value | Accuracy | AUROC | AUPR |
|---|---|---|---|---|---|
| Permian | 0 | 0.002915452 | 0.84 | 0.62 | 0.79 |
| Permian | 0.1 | 0.002915452 | 0.84 | 0.62 | 0.79 |
| Permian | 0.2 | 0.002915452 | 0.84 | 0.63 | 0.78 |
| Permian | 0.3 | 0.002915452 | 0.84 | 0.62 | 0.79 |
| Permian | 0.4 | 0.002915452 | 0.84 | 0.56 | 0.82 |
| Permian | 0.5 | 0.002915452 | 0.84 | 0.56 | 0.82 |
| Permian | 0.6 | 0.002915452 | 0.84 | 0.56 | 0.82 |
| Permian | 0.7 | 0.002915452 | 0.84 | 0.56 | 0.82 |
| Permian | 0.8 | 0.002915452 | 0.84 | 0.45 | 0.83 |
| Permian | 0.9 | 0.002915452 | 0.84 | 0.46 | 0.83 |
| Permian | 1 | 0.002915452 | 0.84 | 0.46 | 0.83 |

# Step 3: Machine Learning Model Results




Area under ROC Curve of **0.93** (ability to distinguish extinction and survival) and Area under PR Curve of **0.80** (fewer prediction errors)

# Step 3: Machine Learning Model Results

| | Cambrian Model Features | Score |
|---|---|---|
| 1 | Circulation System | 3.3422079 |
| 2 | Motility | 2.27727063 |
| 3 | Ocean Acidification Resistance | 1.12284708 |
| 4 | Feeding Patterns | 0.42942449 |
| 5 | Respiratory Organ System | 0.37019192 |
| 6 | Tiering | 0.02890546 |
| 7 | Maximum Length | 0.00479994 |
| 8 | Maximum Area | 4.23E-07 |
| 9 | Calculated Maximum Volume | 2.39E-09 |

| | Ordovician Model Features | Score |
|---|---|---|
| 1 | Motility | 1.06918333 |
| 2 | Ocean Acidification Resistance | 0.85323525 |
| 3 | Tiering | 0.33049509 |
| 4 | Feeding Patterns | 0.32654485 |
| 5 | Respiratory Organ System | 0.15505047 |
| 6 | Circulation System | 0.02294367 |
| 7 | Maximum Length | 0.00124485 |
| 8 | Maximum Area | 4.10E-06 |
| 9 | Calculated Maximum Volume | 1.08E-08 |

| | Silurian Model Features | Score |
|---|---|---|
| 1 | Motility | 1.32779174 |
| 2 | Circulation System | 1.06206068 |
| 3 | Respiratory Organ System | 0.70851463 |
| 4 | Feeding Patterns | 0.6192809 |
| 5 | Ocean Acidification Resistance | 0.59558913 |
| 6 | Tiering | 0.59509338 |
| 7 | Maximum Length | 0.00258455 |
| 8 | Maximum Area | 3.22E-06 |
| 9 | Calculated Maximum Volume | 3.32E-08 |

| | Devonian Model Features | Score |
|---|---|---|
| 1 | Ocean Acidification Resistance | 0.70835892 |
| 2 | Respiratory Organ System | 0.45876009 |
| 3 | Motility | 0.22309167 |
| 4 | Circulation System | 0.06226351 |
| 5 | Maximum Length | 0.00555858 |
| 6 | Maximum Area | 2.00E-05 |
| 7 | Calculated Maximum Volume | 4.32E-08 |
| 8 | Feeding Patterns | 0 |
| 9 | Tiering | 0 |

| | Carboniferous Model Features | Score |
|---|---|---|
| 1 | Circulation System | 2.32544688 |
| 2 | Ocean Acidification Resistance | 0.73146419 |
| 3 | Respiratory Organ System | 0.65025381 |
| 4 | Maximum Length | 0.00189793 |
| 5 | Feeding Patterns | 0 |
| 6 | Tiering | 0 |
| 7 | Motility | 0 |
| 8 | Maximum Area | 0 |
| 9 | Calculated Maximum Volume | 0 |

| | Permian Model Features | Score |
|---|---|---|
| 1 | Circulation System | 2.18390017 |
| 2 | Respiratory Organ System | 1.04694225 |
| 3 | Ocean Acidification Resistance | 0.93551679 |
| 4 | Feeding Patterns | 0.86094481 |
| 5 | Tiering | 0.74849748 |
| 6 | Motility | 0.70993468 |
| 7 | Maximum Length | 0.000241 |
| 8 | Calculated Maximum Volume | 2.56E-08 |
| 9 | Maximum Area | 0 |

# Major Takeaways

- Extinction Risk is not uniform across both geologic history or across taxonomic groups
- Certain traits can act as indicators for higher extinction risk; however, these too vary across geologic history
- These traits can even be used to create relatively accurate predictive models approximating what period a genus went extinct
- **One major goal of our project was to act as a comprehensive foundation for future research:** each one of the traits or phyla from stages one and two can be further studied.

# Future Research

- Completing analysis across the rest of geologic history
  - Identifying patterns of how extinction risk changes for each phyla and each trait across every period in Earth's history
    - Ie. How anoxic conditions affect extinction risk for each phlya/trait
    - Ie. How mass extinctions affect extinction risk for each phlya/trait

- Testing out Decision Tree or Random Forest Regression Models to predict exact first and last appearance in geologic history for each genus

- Look into Building Neural Nets for this type of prediction

# Acknowledgements

Dr. Monarrez

Dr. Saltzman

Dr. Heim

Dr. Payne

Michael Pimentel

# Citations


- Bush, A.M., Bambach, R.K., and Daley, G. 2007.Changes in theoretical ecospace utilization in marine fossil assemblages between the mid-Paleozoic and late Cenozoic. Paleobiology 33, 76–97.
- Heim, N. A., Bakshi, S., Buu, L., Chen, S., Heh, S., Jain, A., Noll, C., Patkar, A., Rizk, N., Sundararajan, S., Villante, I., Knope, M. L., Payne, J. L. 2020. Respiratory medium and circulatory anatomy constrain size evolution in marine macrofauna. Paleobiology, 1-16. doi:10.1017/pab.2020.16
- Novack-Gottshall, P. M. (2008). Ecosystem-wide body-size trends in Cambrian–Devonian marine invertebrate lineages. *Paleobiology, 34*(2), 210-228. doi:10.1666/0094-8373(2008)034[0210:ebticm]2.0.co;2
- N. A. Heim, Knope M. L., Schaal. E. K., Wang S. C., Payne, J. L. 2015. Cope's rule in the evolution of marine animals. Science 347, 867-870.
- Payne, J. L., Bush, A. M., Chang, E. T., Heim, N. A., Knope, M. L. and Pruss, S, B. 2016 Extinction intensity, selectivity and their combined macroevolutionary influence in the fossil record. Biol. Lett.122016020220160202
- Bush, A. M, Pruss, S. B. 2013. Theoretical ecospace for ecosystem paleobiology: energy, nutrients, biominerals, and macroevolution. In Ecosystem paleobiology and geobiology (eds AM Bush, S Pruss, JL Payne), pp. 1 – 20. New Haven, CT: The Paleontological Society.
- Knoll, A. H., Bambach, R. K., Payne, J. L., Pruss, S., Fischer, W.W. 2007 Paleophysiology and end-Permian mass extinction. Earth Planet. Sci. Lett. 256, 295–313. (doi:10.1016/j.epsl.2007.02.018)
- Knope, M., Heim, N., Frishkoff, L. and Payne J. L.. Limited role of functional differentiation in early diversification of animals. Nat Commun 6, 6455 (2015).