Assembling specimen-based trait knowledgebases to test broad-scale drivers of life-history variation

Robert Guralnick
Bryan McLean
(and Daijiang Li, Maggie Hantak)
“Currently, many trait databases focus on a species-level value... This is a logical place to start building a database if many of the questions are focused on comparing central tendencies across species....Having any info is still better than no info, but often we need info on variability across individuals within a species or we want to know how the trait might vary with changes in the environment.”
Digitization of museum specimens has made specimens and specimen data discoverable on the internet.
Standardization doesn’t assure that all the data about specimens is mobilized. In fact, standardization can mean some critical data is hidden in digitized specimen data.

Tamiasciurus hudsonicus

Specific Locality: Goldstream Valley, KK Campbell Residence
Collecting Method: firearm
Collecting Source: wild caught
Event Date: 2012-06-01 to 2012-06-30

Dynamic properties:
age class: young; ear from notch: 23mm; hind foot with claw: 49mm; sex: female; tail length: 121mm; total length: 299mm; weight: 135g

Occurrence Remarks:
Appeared to be first year

Normalized results: Body mass: 135 grams, Total body length: 299mm

How to build the largest knowledge-base of vertebrate body masses ever assembled

Recipe:
1. Take 20 million digitized vertebrate records
2. Decompose the records and look for hidden body size information
3. Extract matching records and harmonize the outputs to same units and format
4. Reassemble into a neat form the extracted data with the original records
5. Serve to the community
6. Make it all reproducible so others can follow the recipe
Our efforts unlocked an enormous number of body mass and length measurements.
Bringing It back to VertNet –
Adapting the portal to make the trait data discoverable
No general relationship between mass and temperature in endothermic species

26 Carnegie Museum of Natural History
CM Birds Collection (2015)

27 American Museum of Natural History
AM Collections (2015)

28 VertNet: a new model for biodiversity data sharing
H Constable, R Guralnick, J Wieczorek, C Spencer, AF Peterson, VertNet Steering Committee (2010)
https://doi.org/10.1371/journal.pbio.1000309
PubMed | Google Scholar

29 University of British Columbia Beaty Biodiversity Museum
Cowan Tetrapod Collection - Birds (2015)

30 University of British Columbia Beaty Biodiversity Museum
Cowan Tetrapod Collection - Mammals (2015)

31 Cornell University Museum of Vertebrates
MUCH MORE THAN JUST MASS AND LENGTH
It's not just body measurements... but life history data

Bryan McLean
What controls rangewide litter size variation in *P. maniculatus* (PEMA)

- *P. maniculatus* has multiple litters per year.
- Shorter breeding season may lead to investment in more offspring since there are fewer opportunities to breed.
But, to quote one of the key researchers to look at broad litter size patterns (Millar, 1989): "It should be remembered that organisms do not respond to latitude, longitude, or elevation per se. Comparisons should involve meaningful measures of the environment."
What controls rangewide litter size variation in *P. maniculatus*

- How do instantaneous variables (temperature of the past few weeks) and more long-term (10 year average of season length) variables affect litter size/embryo count variation?

- Can we preliminarily pick up a signal of change over longer time-scales (decadal changes) with these data?
What data resources exist for examining litter size trends?

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalogNumber</td>
<td>11198</td>
</tr>
<tr>
<td>OtherCatalogNumbers</td>
<td>DZTM: Denver Zoology Tissue Mammal=173; collector number=JRD418</td>
</tr>
<tr>
<td>RecordNumber</td>
<td>JRD418</td>
</tr>
<tr>
<td>RecordedBy</td>
<td>Collector(s): John R. Demboski; Preparator(s): Kayce C. Bell</td>
</tr>
<tr>
<td>IndividualCount</td>
<td>1</td>
</tr>
<tr>
<td>Sex</td>
<td>female</td>
</tr>
<tr>
<td>Preparations</td>
<td>liver (frozen); heart, kidney (frozen); skin, study; skeleton; embryo (ethanol)</td>
</tr>
<tr>
<td>DynamicProperties</td>
<td>sex=female ; total length=153 mm; tail length=66 mm; hind foot with claw=20 mm; ear from notch=21 mm; weight=21.9 g; reproductive data=embryos: 3R, 2L ; crown-rump length=15 mm</td>
</tr>
<tr>
<td>EventDate</td>
<td>2007-07-12</td>
</tr>
<tr>
<td>StartDayOfYear</td>
<td>193</td>
</tr>
<tr>
<td>EndDayOfYear</td>
<td>193</td>
</tr>
<tr>
<td>Year</td>
<td>2007</td>
</tr>
<tr>
<td>Month</td>
<td>07</td>
</tr>
<tr>
<td>Day</td>
<td>12</td>
</tr>
<tr>
<td>VerbatimEventDate</td>
<td>11 July 2007</td>
</tr>
</tbody>
</table>
How many litter size records are there for PEMA in VertNet?
Is reporting of embryo counts changing over time?

Yes, improving....
Where are these records?
What key factors might influence PEMA embryo count?

<table>
<thead>
<tr>
<th>Variable Acronym</th>
<th>Variable</th>
<th>Description</th>
<th>Classification</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>Temperature Seasonality</td>
<td>10-year standard deviation of monthly temperature (degrees Celsius)</td>
<td>long-term</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>PS</td>
<td>Precipitation Seasonality</td>
<td>10-year standard deviation of monthly precipitation (millimeters)</td>
<td>long-term</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>NFFD</td>
<td>Number of Frost-Free Days</td>
<td>10-year average of annual number of frost free days</td>
<td>long-term</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>TB</td>
<td>Mean Breeding Temperature</td>
<td>Average temperature during breeding month and previous month (degrees Celsius)</td>
<td>instantaneous</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>PB</td>
<td>Total Breeding Precipitation</td>
<td>Total precipitation during breeding month and previous month (millimeters)</td>
<td>instantaneous</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>NFFD anomaly</td>
<td>Number of Frost-Free Days Anomaly</td>
<td>Annual deviation from decadal mean number of frost free days</td>
<td>instantaneous</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>FDD anomaly</td>
<td>Freezing Degree Days Anomaly</td>
<td>Annual deviation from decadal mean number of freezing degree days</td>
<td>instantaneous</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>DFFP</td>
<td>Day during Frost-Free Period</td>
<td>Days elapsed after beginning of frost-free period</td>
<td>instantaneous</td>
<td>ClimateNA</td>
</tr>
<tr>
<td>HB</td>
<td>Head-body Length</td>
<td>Total length - tail length (in millimeters)</td>
<td>individual-based</td>
<td>VertNet, NACSM</td>
</tr>
</tbody>
</table>
First Look: Simple Summaries

Overall distribution of embryo counts

There is a signal of latitude just not very strong and latitude is not the predictor of interest

Histogram showing frequency of embryo counts with $M = 4.8$ and $SD = 1.3$

Scatter plots showing embryo count vs. latitude and longitude with $r^2 = 0.04$, $P << 0.01$ and $r^2 < 0.01$, $P = 0.41$ respectively.
• Models with just 10 year average number of frost free days (NFFD) are only marginally worse than models including all climate variables

• Embryo count strong negative relationship with NFFD

• NFFD as a sole predictor explains about 9% of the variance in embryo count

• Including head-body (HB) length marginally improves these models (HB + relationship)
A visual view: Embryo Count Histograms Over NFFD Bins
What about temporal trends?

• Examining temporal trends with incidental data is **fraught with bias issues**

• However, very preliminary analysis using temporal bins of 20 years, and fitting a model with litter size as a response and NFFD, 20 year decadal bins and their interaction as predictors

• Results: We find a significant, negative trend over 20 year bins – which would be expected direction of change

• But.... Hard to believe these results quite yet
Digging into temporal trends?

• The place to start to examine temporal trends is with body mass and length given significant availability of data
• Assembled a large, multisource (>25,000) record adult dataset
• Spatially constrained analysis to 200km grid cells with >100 record and at least 4 decades with >10 records.
• Tested hypotheses of reduced body size due to climate or human pressure
Preliminary Results Using Body Mass

- We assembled MAT (Bergmann’s), MAP, human population density, sex and season collected as predictors of body mass and...

- Treating each cell independently, if we fit this model per cell: \( \text{lm}(\text{mass}_g \sim \text{year} + \text{sex} + \text{season} + \text{MAT} + \text{MAP} + \text{pop}_10\text{km}2_{\log10}) \) 30 of 35 cells have negative year trends, 12 significant negative, and 1 significant positive.

- A linear mixed model with cell as random effect also shows a strong overall size decrease over year or decade BUT there is no relationship with population density or MAT. There is a negative, significant relationship with MAP.

- We still don’t know why we see body size declines. Much more to do...
The Functional Trait Resource for Environmental Studies (FuTRES) project is a collaborative project among four universities (University of Oregon, University of Arizona, University of Florida, and Howard University). The key deliverables of FuTRES are a workflow for assembling functional trait data measured at the specimen level, a database to serve that data, and scientific publications demonstrating the utility of the assembled data.