Nikolai G. Vetr, PhD

Contact Information

Phone: (602) 578-9196
Email: nikgvetr@stanford.edu

LinkedIn: linkedin.com/in/nikolai-vetr
GitHub: github.com/NikVetr/

Skills & Interests

- Probability Modeling
- Multiomic Data Analysis
- Time Series Modeling
- Bayesian Methods
- Monte Carlo Methods
- Causal Inference

- Computer Vision
- Artificial Neural Networks
- Machine Learning
- Optimization Methods
- Science Communication
- Nat. Lang. Processing

- Evolutionary Biology
- Exercise Biology
- Population Genetics
- Data Visualization
- High Perf. Computing
- Biomedical Data Science

Recent Papers

Abell, N., **Vetr, N.***, Montgomery, S., et al. 2024. *A Survey of High Depth Allele-Specific Expression Across Normal Tissues and Ovarian Cancers.* In Prep.

Vetr, N., Gay, N., and Montgomery, S. 2024. *The impact of exercise on gene regulation in association with complex trait genetics.* Nature Communications 15(3346): 1-14. DOI: 10.1038/s41467-024-45966-w.

MoTrPAC Study Group[†]. 2024. *Temporal dynamics of the multi-omic response to endurance exercise training across tissues.* Nature 629(8010): 174-183. DOI: 10.1038/s41586-023-06877-w.

Gates, K., Panicker, A., Biendarra-Tiegs, S., **Vetr, N.**, et al. *Shotgun Immunoproteomics for Identification of Nonhuman Leukocyte Antigens Associated With Cellular Dysfunction in Heart Transplant Rejection.* Transplantation 106(7). DOI: 10.1097/TP.0000000000004012

[*]*dual first authorship*, [†]*Author Group: 2 (of 8)*
(see second page for a description of work)

Leadership

Founder, Applied Bayesian Statistics Research Cluster, *UC-Davis*		*2019 - 2020*
President, Board of Directors, *Wild Animal Initiative*		*2020 - Present*
President, Board of Directors, *Rethink Priorities*		*2023 - Present*

Languages

Programming: R, Stan, BASH, Python, C++, CSS, HTML, JS
Natural: Russian, English, Spanish

Education & Training

Postdoc, Montgomery Lab, Stanford University *Current*
Pathology + Genetics + Biomedical Data Science

PhD, University of California, Davis *2020*
Dissertation: *Exploring and Extending Multivariate Brownian Diffusion Models of Phenotypic Evolution for Bayesian Phylogenetic Inference*
Anthropology + Population Biology + Data Science & Informatics

BA, Vanderbilt University *2013*
Earth & Environmental Sciences + Ecology, Evolution & Organismal Biology
Departmental Honors, *summa cum laude*

Teaching

Associate Instructor, University of California, Davis	*2015 - 2020*
Human Evolution + Primate Evolution + Human Evolutionary Biology	
Carpentries Instructor, Data & Software Carpentries	*2019*
Course Coordinator, Workshop in Applied Phylogenetics	*2019*

Projects Previous **dissertation work** focused on computational methods development in Bayesian phylogenetics. This included developing:

- computational tricks (eg factorizations) to improve scaling of multivariate likelihood calculation
- efficient, tunable proposal distributions and informative (non-I centered) prior distributions for correlation matrices (rank-one updates and downdates w/ Givens rotations to reduce Brownian rate matrix decomposition from $O(n^3)$ to $O(n^2)$)
- extensive simulation experiments to systematically characterize the conditions under which proposed methods succeed and fail
- novel data visualizations (eg *cumulative average resolution* curves) to assist interpretation of the above
- hybrid EM/MCMC approaches to reliably fit otherwise intractable multimodal models (eg phylogenetic diffusion + multivariate ordinal probit)

Current **postdoctoral work** focuses on probability modeling of multiomic dynamics in the biomedical context. Projects here include:

- extensive data analysis and figure generation for preclinical MoTrPAC landscape publication, including integrative proteomic + transcriptomic time series analysis, multiomic & multitissue clustering + network analysis, batch effect adjustment, dimensionality reduction, and uncertainty / error propagation, among other work
- developing novel causal inferential methods, effect standardization workflows, and complex trait enrichment methods, while also leveraging more conventional bioinformatic tools (eg for heritability enrichment)
- performing principled, multilevel modeling of loss-of-heterozygosity and allele-specific expression in a multitissue dataset to identify signatures of ovarian cancer
- developing meta-scientific tools to streamline common computational workflows (eg converting Stan models to interactive R code for predictive simulation, calibration analysis, outlier detection, etc.)

Various **personal projects** leverage scientific and statistical computing workflows to achieve diverse aims. These have entailed writing programs that:

- adaptively compress and vectorize raster images using techniques from computer vision, signal processing, and graph theory
- generate custom crossword puzzles that conform to shapes derived from user-supplied pictures
- glue together API calls to create a fully SMS-based, multimodal digital assistant
- infill and text-wrap a passage of text within a user-specified image
- efficiently predict user ratings and provide media recommendations for popular movies
- interface with external software to assist in metalworking, woodworking, pen plotting, and other physical modification of objects
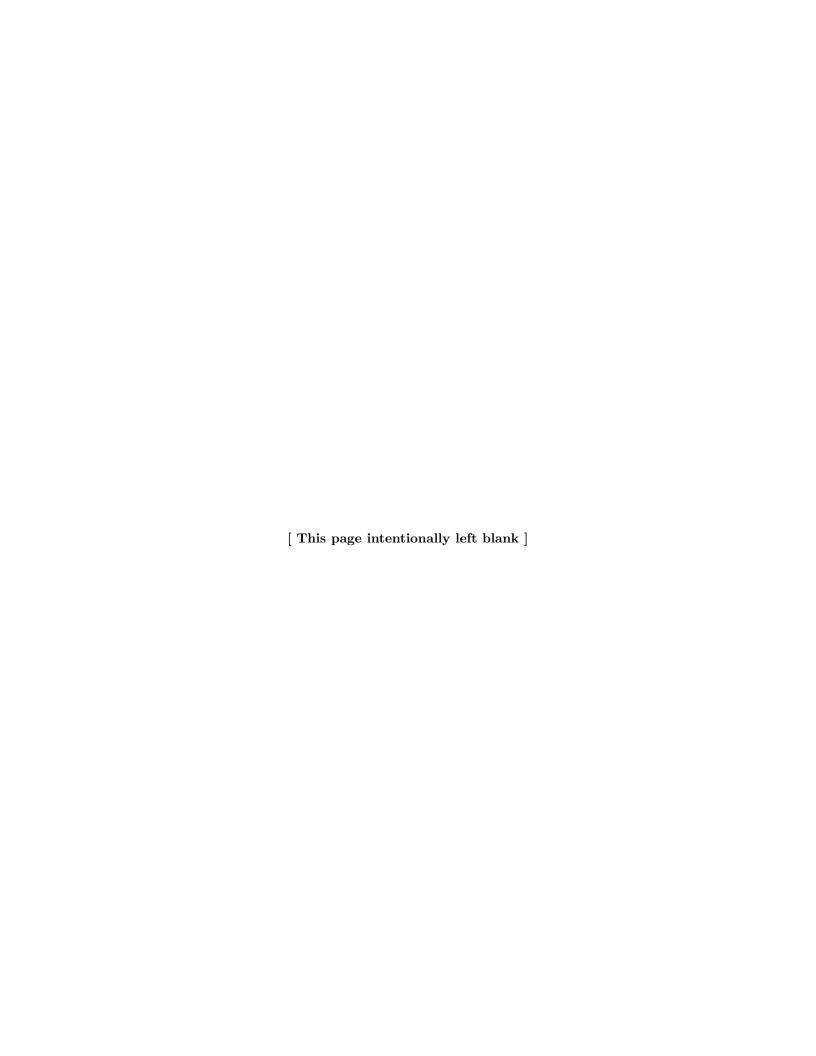
Service **Journal Review**: *Evolution* (2017), *Science Communications* (2018), *Cell Reports* (2021), *Human Genetics and Genomics Advances* (2022)
Grant Review: *WAI Grants* (2021, 2022, 2023)

Selected Grants & Awards

NIH T15	*2021*
Excellence in Data Science Community Training and Outreach	*2019, 2020*
Outstanding Graduate Student Teaching Award Nominee	*2016, 2019, 2020*
1st Place Picnic Day Exhibit Award in "Secrets of Nature" Category	*2017*
NSF Graduate Research Fellowship	*2015*

[**This page intentionally left blank**]

Nikolai Vetr
Stanford University, California
nikgvetr@stanford.edu — github.com/NikVetr/ — linkedin.com/in/nikolai-vetr
(last update: June 6, 2024)

Education and Training

Postdoc, Montgomery Lab, Stanford University — *Current*
Pathology + Biomedical Data Science

PhD, University of California, Davis — *2020*
Dissertation: *Exploring and Extending Multivariate Brownian Diffusion Models of Phenotypic Evolution for Bayesian Phylogenetic Inference*
Committee: Tim Weaver (Chair), Shara Bailey, Peter Wainwright
GPA: 4.0/4.0 — Anthropology + Population Biology + Data Science & Informatics

BA, Vanderbilt University — *2013*
GPA: 3.9/4.0 — Earth & Environmental Sciences + Ecology, Evolution & Organismal Biology
Departmental Honors, *summa cum laude*

Additional Education

Discussion Groups — *2013 - 2020*
I've served as an active participant in groups on Computational Molecular Evolution, Paleoanthropology, Quantitative Genetics, and Machine Learning. I've been the founder and lead in groups on Python, Applied Linear Algebra, Bayesian Data Analysis, and Deep Learning.

Workshops — *2013 - 2020*
I've attended or led multiple workshops ranging in duration from 2h to 9d on Universal Design, Science Communication, Bayesian Phylogenetics, Database Management, Natural Language Processing, Carpentries Instructor Training, and other topics.

Languages

	Programming	Natural
Fluent:		Russian, English
Proficient:	R, RevBayes, Stan, BASH	Spanish
Beginner:	Python, C++, CSS, HTML, JS	

Service

Journal Review

Evolution (2017), *Science Communications* (2018), *Cell Reports* (2021), *Human Genetics and Genomics Advances* (2022)

Grant Review

WAI Grants (2021, 2022, 2023)

Leadership

Founder & Lead, Applied Bayesian Statistics Research Cluster *2019 - 2020*
I founded, coordinated, and supervised an interdisciplinary research cluster of 70+ scientists and statisticians across numerous career stages (PhD Student, Postdoc, PI, Industry Researcher), financially and spiritually sponsored by the UC-Davis *Data Science & Informatics* Unit

Director & President, Board of Directors, Wild Animal Initiative *2020-Present*
I serve on the Board of Directors at the *Wild Animal Initiative*, a non-profit dedicated to researching the lives of wild animals. In this role, I help to oversee the organization's activities and policies as described in their mission statement and bylaws, approve financial decisions, decide on board meeting agendas, and engage in related activities. I've also served as chair of both the DEI and Science committees

Director & President, Board of Directors, Rethink Priorities *2023-Present*
I serve on the Board of Directors at *Rethink Priorities*, a non-profit focused on cause prioritization and research in animal welfare, artificial intelligence, climate change, and global health and development. Responsibilities are similar to above. Over several years I have also extensively consulted on or performed statistical model fitting, data visualization, and related activities

Projects

Postdoctoral Work

My postdoctoral work in the Montgomery Lab focuses on exploring and developing statistical methods to better understand the molecular mechanisms by which genetic and environmental variation structure phenotypic variation, especially insofar as that understanding might have translational applications in contexts such as drug discovery and personalized medicine. Current and recent projects include:

- Exploring inferential performance of **sparse feature selection** methods under conditions of strong input interdependence [LINK]
- Integrating GWAS and eQTL mapping results with transcriptomic, proteomic, and ATAC-seq data to help **identify the molecular transducers of physical activity** through the MoTrPAC Consortium [LINK]
- Streamlining the Bayesian workflow by creating text-processing utilities to **deterministically translate Stan code to R** (or other interpreted languages), allowing for dynamic exploration of model structure and easier automated acquisition of posterior predictive probabilities or distributions (for simulation experiments, or to assess observation-specific goodness of fit) [LINK]
- Designing novel probability models to infer differential **allele-specific expression patterns linked to ovarian cancer** throughout the course of tumor development using data from mmPCR-seq [LINK]
- Developing **integrative methods to link multiomic response with trait expression** that appropriately propagate upstream inferential uncertainty [LINK]
- Fitting tractable **high-D interaction models involving multiple filtration steps** for false-negative aware tests of intersectionality [LINK]
- Propagating inferential uncertainty to **mitigate 0-biasing effects of differential power** on estimates of interdependence [LINK]
- Hierachically modeling the **temporal dynamics of whole-trancriptome expression data** following environmental perturbation using linear combinations of logistic functions to characterize temporal behavior in expected rates of gamma-poisson distributed read counts [LINK]
- Exploring causal inferential implications of **aging clocks** [LINK, example]

- Developing easy-to-use tools for **frequentist testing** of novel generative immune response models
- Developing and implementing a **dynamic request intake framework** atop Google Workspace with Google Apps Script to assist administrative staff with coordinating disparate requests with scaling group size
- Describing **novel informative correlation matrix distributions** for use in multivariate, multilevel diffusion models [LINK]
- Understanding the proteomic signatures of **allogeneic renal transplant rejection** [LINK]
- Exploring sources of linkage disequilibrium unrelated to crossing over (eg assortative mating) in a simulation framework [LINK]
- Collaborating on the description and implementation of **Bayesian differential expression models** [LINK],
- Describing theoretical expectations for **rare-variant effects on outlier detection** [LINK, example]

Dissertation Work

My dissertation work primarily centered on Bayesian computational phylogenetics methods development, specifically extending and exploring multivariate Brownian (mvBM) diffusion models to novel character evolutionary applications and contexts [LINK]. Components of this this work included:

- Through simulation experiments, **exploring the informativeness and reliability of continuous character alignments** at retrieval of simulating tree topology, with subsequent application to a craniometric dataset
- The **effects of model misspecification** in the above, both with respect to the evolutionary process and with respect to accommodating missing data and measurement error
- Developing and implementing novel optimizations to **reduce the time complexity** of likelihood calculation under the phylogenetic mvBM model from $O(n^3)$ to $O(n^2)$, modularizing particular operations to **avoid redundant computation** during Met-Hastings, and inventing substantially **more efficient, tunable proposal distributions** over correlation matrices of arbitrary dimensionality while inherently respecting PSD constraints. In total, **reducing the runtime required** to fit this popular evolutionary model under typical empirical conditions from months to hours
- **Comparing model-based phylogenetic methods to more common heuristic (maximum parsimony, distance) methods**, both in simulation and empirically (using ape and old-world monkey datasets), and contrasting their performance to methods that use molecular (e.g. nucleotide sequence) data
- **Joint inference of ancestral character states** in a monkey dataset, with additional theory development on imputing unobserved tip characters in a phylogenetically sensible way
- Layering a **high-dimensional probit model atop mvBM to accommodate discrete traits** and comparing it with both more traditional and novel CTMC (*Continuous Time Markov Chain*) models of discrete trait substitution, with application to human dental variation

In addition to the above dissertation contents, parallel work involving related skills included:

- **Comparison of morphological relaxed clock models** in a species-rich fish dataset whose tree height straddles the K-Pg boundary [LINK]
- **Approximating truncated biogeographic diffusions using data augmentation** on phylogenies to sample migration histories while respecting (land, water) barriers to explore latitudinal / climactic drivers of morphological differentiation in modern humans [LINK]

- High-dimensional multilevel generalized linear mixed models (GLMMs; Gamma-Poisson, in this case) in Stan to **heart transplant patient immune response** data both before and after an acute rejection episode [LINK]
- A hierarchical univariate Ornstein-Uhlenbeck model in Stan to the 'evolution' of **nitrogen concentrations in manure ponds** across California to predict values after an arbitrary amount of time has passed in an arbitrary pond [LINK]

Personal Projects

I've written many dozens of one-off scripts devoted to data visualization, text mining, web scraping, animation, replicating published analyses or algorithms from scratch, generating artwork, solving everyday difficulties, and exploring off-the-cuff ideas proposed by myself or colleagues. Sometimes these require the flexibility of implementing everything from scratch, sometimes I'm able to make do with glueing a bunch of API calls together.

Recent examples here include but are not limited to:

- Exploring the space of valid crossword puzzles via MCMC to **generate crosswords that adhere to specific shapes determined by a user-specified images** (and hints + answers) [LINK, example]
- Implementing on AWS a **multimodal, SMS-based chat assistant** for family not quite able to operate browser- or app-based interfaces, but very familiar with sending text messages [LINK]
- As above, but instead of a chatGPT clone, hooking the API calls to a voice interface to create an admittedly high latency, **LLM-powered voice assistant** (not unlike Amazon Alexa or Google Home) to run on family computers.
- **Generating board (or other) meeting minutes from video recordings**, interleaving image-to-text output at the appropriate timestams extracted using temporal variance in multivariate picture heterogeneity, also incorporating speaker diarization methods to better identify who motioned this or seconded that [LINK]
- Leveraging wavelets and other transforms to **compress and vectorize raster images** [LINK, example], with initial attempts to expand into higher dimensional spaces (eg time or 3D objects) [LINK, example]
- Automated **infilling and text-wrapping** of arbitrary images with arbitrary text [LINK, example]
- Automated **processing and overlay of in/outpainted images** [LINK, example]
- **Generating seating charts** under various (attitudinal, geographic) optimization criteria for an ice-breaker activity, including in the multi-round iterated case where participants move [LINK, example]
- **Generating walking directions** using graph operations and API calls that prioritize other weighted criteria besides time to destination (eg regularly spaced amenities, density of PoI, beauty as proxied by Flickr geolocated metadata, etc.) [WiP]
- **Generating laser cutter input files** for parallel tag / fob creation (to tag deboxed boardgames, provide custom swag, create 3D engravings by discretizing depth maps, etc.) [WiP, LINK]
- Making other **small animations** leveraging basic geometric and linear algebraical operations [LINK, example]
- Fitting multilevel GLMMs in Stan to **consumer dietary and attitudinal response datasets** (zero-inflated Poisson and ordered logit, respectively) measuring the effects of exposure to a relevant advertisement urging dietary change [LINK]
- Implementing an efficient, conditional multivariate ordinal probit model for **personalized movie rating predictions**, exploiting basic properties of Schur complements [LINK]
- Writing a **"falling rain" parody generator** for the Matrix 4 release [LINK, example]

I've also consulted individuals & organizations in a minor role (e.g. over a $\sim$ 1h meeting) on many additional projects throughout my postdoc and graduate school.

Field & Labwork

Archaeological / Paleontological Excavation   *2013*
During the Summer 2013 field season I helped in excavation efforts ongoing at La Ferrassie, a Neandertal archaeological and paleontological site in Savignac-de-Miremont, France

Water Quality Analysis   *2011*
I performed water quality analyses (of salinity, reactive nitrogen concentration, etc.) of a large set of streams distributed across the North Island of New Zealand

Stable Isotope Ecology   *2012 - 2013*
I examined how environmental and ecological conditions were recorded in enamel stable isotope ratios (carbon and oxygen) and dental microwear textures for several marsupial taxa throughout Australasia

Teaching

Associate Instructor*, University of California, Davis   *2015 - 2020*
As instructor-of-record, I've taught three quarters of an upper-division paleoanthropology course, two quarters of an upper-division evolutionary primatology course, and one quarter of an introductory course on human evolution. In these roles, I created or modified all the lab and lecture materials, designed and graded assignments and tests, mentored individual students, delivered lectures, and either supervised teaching assistants or administered and oversaw lab activities

Teaching Assistant*, University of California, Davis & Vanderbilt University   *2012 - 2020*
I have taught a total of 14 discussion sections for courses on human evolution and paleoanthropology, as well as the lab for an introductory cell biology course at Vanderbilt University. As a TA, I occasionally gave guest lectures on various subjects (mostly statistical phylogenetics)

Outreach Lecturer, University of California, Davis   *2013 - 2020*
I've given multiple yearly talks on human evolution to elementary and middle school students on location and visiting UC-Davis, as well as during campus-wide events (e.g. *Picnic Day*). I've also assisted in miscellaneous workshops targetted at adults, e.g. on NLP during UC-D Data Science Health Day

Course Coordinator, Workshop in Applied Phylogenetics   *2019*
I served as a course coordinator for world-renowned, widely-attended, week-long workshop in applied computational Bayesian phylogenetics held in May 2019 at the Bodega Marine Laboratory

Carpentries Instructor   *2019*
I have completed Instructor Training and Checkout for The Carpentries organization and am fully certified to teach courses in the Data and Software Carpentries

**student evaluations available upon request*

Research Output

Papers

[1] Abell, N., **Vetr, N.***, Montgomery, S., et al. 2024. *A Survey of High Depth Allele-Specific Expression Across Normal Tissues and Ovarian Cancers.* In Prep.

[2] **Vetr, N.**, Gay, N., and Montgomery, S. 2024. *The impact of exercise on gene regulation in association*

with complex trait genetics. Nature Communications 15(3346): 1-14. DOI: 10.1038/s41467-024-45966-w. [LINK]

[3] **MoTrPAC Study Group**[†]. 2024. *Temporal dynamics of the multi-omic response to endurance exercise training across tissues.* Nature 629(8010): 174-183. DOI: 10.1038/s41586-023-06877-w. [LINK]

[4] Carpenter, B., Chen, S., and **Vetr, N**. 2022. *Bayesian Models of Gene Expression and RNA Sequencing.* [LINK]

[5] Gates, K., Panicker, A., Biendarra-Tiegs, S., **Vetr, N.**, Higuita, M., Nelson, T., Pereira, N., and Griffiths, L. 2021. *Shotgun Immunoproteomics for Identification of Nonhuman Leukocyte Antigens Associated With Cellular Dysfunction in Heart Transplant Rejection.* Transplantation 106(7). DOI: 10.1097/TP.0000 000000004012 [LINK]

[6] **Vetr, N.**, Bailey, S., Moore, B., and Weaver, T. 2020. *Human Population History from Discrete Dental Traits Under an Approximate Multivariate Ordinal Probit.* [LINK]

[7] **Vetr, N.** and Weaver, T. 2020. *Primate Phylogenetics with Landmark Data: A Model-Based, Multivariate Brownian Approach.* [LINK]

[8] **Vetr, N.**, May, M., Moore, B., and Weaver, T. 2019. *Bayesian Phylogenetic Inference Under Multivariate Brownian Motion.* [LINK]

[9] **Vetr, N.**, DeSantis, L., Yann, L., et al. 2013. *Is Rapoport's Rule a Recent Phenomenon? A Deep Time Perspective on Potential Causal Mechanisms.* Biology Letters 9(5): 1-5. DOI: 10.1098/rsbl.2013.0398 [LINK]

*dual first authorship
[†]Author Group: 2 (of 8)

Conference Presentations

[1] **Vetr, N.G.**, Abell, N.S., Tsang, E., Cheng, N., Linder, S., Smith, K.S., Uzunovi, J., Lamaze, F., Ardlie, K., Awadalla, P., Li, J.B., and S.B. Montgomery. 2024. *High Depth Allele-Specific Expression Across Healthy Tissues and Ovarian Cancer.* Abstract submitted to the American Society of Human Genetics annual meeting, Denver CO.

[2] Katz, D.H., Jin, C.A., ..., G., **Vetr, N.G.**, ..., Snyder, M., Montgomery, S., Ashley, E.A., Wheeler, M.T., for the MoTrPAC Study Group. 2024. *The Multi-omic, Multi-tissue Response to Acute Endurance and Resistance Exercise: Results from the Molecular Transducers of Physical Activity Consortium.* Abstract submitted to the American Heart Association Scientific Sessions, Chicago, IL. (to be presented by D. H. Katz, authors omitted for brevity)

[3] **Vetr, N.**, Gay, N., and S. Montgomery. 2022. *The impact of exercise on gene regulation in association with complex trait genetics.* Poster session presented at the American Society of Human Genetics' annual meeting, Los Angeles, CA.

[4] **Vetr, N.**, Gay, N., and S. Montgomery. 2022. *Integrating Exercise-Induced Gene Expression and Human Phenotypic Variation.* Poster session presented at NIH-NLM T15 Training Conference, Buffalo, NY.

[5] Ungar, R., Li, T., Ersaro, N., **Vetr, N.G.**, Battle, A., and S. Montgomery. 2022. *Functional impact of rare variants and sex across the X chromosome and autosomes.* Poster session presented at the American Society of Human Genetics' annual meeting, Los Angeles, CA. (Presented by R. Ungar)

[6] Panicker, A.J., Gates, K.V., Biendarra-Tiegs, S.M., **Vetr, N.G.**, Lopera Higuita, M., Nelson, T.J., Pereira, N.L., and L.G. Griffiths. 2021. *Identification of Non-Human Leukocyte Antigens Associated with Cellular Dysfunction in Heart Transplant Rejection.* Poster session presented at the American Heart Association's annual Scientific Sessions, Nov. 13-15, Boston, MA, and virtual. (Presented by A.J. Panicker)

[7] **Vetr, N**., May, M. Moore, B., and T. Weaver. 2018. *Adapting multivariate Brownian diffusion models to Bayesian inference of human population history and phylogeny.* Oral session presented at the 87th Annual Meeting of the American Association of Physical Anthropologists, Austin, TX.

[8] **Lashinsky, N**., DeSantis, L. Yann, S. Donohue, and R. Haupt. 2013. *Is Rapoport's Rule a Recent Phenomenon? A Deep Time Perspective on Potential Causal Mechanisms.* Oral session presented at the 73rd Annual Meeting of the Society of Vertebrate Paleontology, Raleigh, NC.

[9] Desantis, L., **Vetr, N***., Romer, J., Greshko, M., and Loffredo, L. 2013. *University Students' Acceptance of Climate Change and Evolution: Are Skeptics just Anti- Science?* Poster presented despite second authorship at the 73rd Annual Meeting of the Society of Vertebrate Paleontology, Raleigh, NC. (Presented by N. Vetr)

[10] **Vetr, N***. and DeSantis, L. 2012. *Assessing how the extant lesser forest-wallaby Dorcopsulus vanheurni and Lumholtz's tree kangaroo Dendrolagus lumholtzi record their environment: establishing a baseline for studies of extinct marsupials.* Poster presented at the 5th Annual Meeting of the Southeastern Association of Vertebrate Paleontology, Asheville, NC.

*previously published under the name Lashinsky, N.

Grants & Awards

NIH T15	*2021*
Excellence in Data Science Community Training and Outreach	*2019, 2020*
Outstanding Graduate Student Teaching Award Nominee	*2016, 2019, 2020*
Conference Travel Award	*2012, 2018*
1st Place Picnic Day Exhibit Award in "Secrets of Nature" Category	*2017*
NSF Graduate Research Fellowship	*2015*
Summer Research Grant	*2014, 2015*
Graduate Scholars Fellowship	*2013*
Departmental Honors in Earth and Environmental Sciences	*2013*
Eugene H. Vaughan Undergraduate Research Assistantship in Geology	*2012*
Geology Travel Grant	*2012*
Vanderbilt Undergraduate Summer Research Grant	*2012*
Ross Family Scholarship	*2012*
National Merit Scholarship	*2009*