SURF Proposal

Heidi Heejin Yang
heidi.h.yang@berkeley.edu

Research Statement

Identifying the genomic mechanisms that drive divergence of rapidly evolving species, remains a major unknown in evolutionary biology. This project seeks to elucidate the potential role of transposable elements (TEs) in adaptive radiation and the associated evolution of diverse phenotypes (3). Transposable elements (TEs), DNA sequences that can change position within a genome, have been suggested as playing a role in initiating divergence. I propose to use an adaptive radiation of *Tetragnatha* spiders, which display different stages of adaptive radiation across different Hawaiian islands, to test the expectation that TE activity is higher in the youngest branches of the radiation and radiating Hawaiian species have higher TE activity than a non-radiating mainland species. To test this I will: 1) provide a valuable chromosome-level reference genome of the Hawaiian *Tetragnatha brevignatha* that will allow comparison of positional information relative to an existing genome of a mainland species, *T. versicolor*, and 2) utilize existing transcriptome data from across the radiation together with the reference genome to compare TE expression in radiating and non-radiating *Tetragnatha* species.

Background Topic and Rationale for Research

Adaptive radiations are rapid bursts of diversification within a lineage that allow the component species to adapt and fill many different ecological roles (1). While the ecological and evolutionary aspects have been extensively studied, little is known about the genomic mechanisms that produce such high genetic and phenotypic diversity (3). One potential genomic component are transposable elements (TEs), DNA sequences that can change its position within a genome. TEs can produce a wide variety of mutations, but their potential role in eliciting reproductive isolation and/or phenotypic change associated with adaptive radiations remains unclear (4). McClintock (1984) first proposed that TE activity may increase in response to "challenges to the genome" (8). Since adaptive radiations frequently occur when a species colonizes a new area, presenting new biotic or abiotic stresses (2), these changes may trigger the deregulation of the genome and activate TEs (1, 3).

Comparative studies have shown a higher level of TE activity in radiating lineages than in related non-radiating lineages, including the African cichlid fishes and *Anolis* lizards (3). TE activity within these groups has also been associated with the rise of novel traits; in African cichlids a TE insertion in the cis-regulatory region of a pigmentation gene led to the evolution of egg-spots (7), and accumulation of TEs in *Hox* gene clusters, genes responsible for morphological differences between taxa, has been observed in *Anolis* lizards (6). Given the evidence of increased TE activity in adaptive radiations and their potential to produce novel traits, much needed is a study that examines how TEs are involved in reproductive isolation

versus niche differentiation; and for this, we need a system that displays populations and species in the earliest stages of adaptive radiation.

The Hawaiian Islands offer the opportunity to examine adaptive radiation across a vast time spectrum, and the radiation of Hawaiian *Tetragnatha* spiders is an ideal study system because it has followed this chronosequence (9, 10). Previous work has highlighted the initial stages of adaptive radiation in Hawaiian *Tetragnatha* spiders (11); additionally, work through the California Conservation Genomics Project will produce a reference genome of the California species, *Tetragnatha versicolor*. Using the temporal aspect of diversification and information of their mainland sister taxa, this project will compare TE activity between older radiating, younger radiating, and non-radiating mainland *Tetragnatha* species to test if TE activity increases in response to novel ecological conditions and early on within a radiation.

Research Plan: Methodologies and Timeline

I hypothesize that TE activity occurs early on in a lineage due to the deregulation of previously silenced regions in response to new abiotic or biotic stresses. Thus, considering the green "spiny leg" clade within the radiating *Tetragnatha* lineage, the activity of TEs will be higher in the species on the youngest islands of Maui and Hawai'i compared to those on the older islands of Kaua'i and O'ahu, but both will have higher numbers than the mainland taxa (Figure 1). To test this I plan to use a two-pronged approach: (1) I will analyze transcriptome data from mainland and Hawaiian *Tetragnatha* species from a previous SURF-funded study by undergraduate student Cory Berger (12) to compare transcript levels of TEs of *T. kauainesis, T. tantalus, T. brevignatha* (Maui), *T. brevignatha* (Hawai'i), and *T. versicolor*; (2) I will construct a high-quality chromosome-level reference genome of *T. brevignatha*, building on ongoing work in the lab involving the chromosome-level genome of *T. versicolor*. The genome will be used to accurately annotate where TEs may exist, and I will map the existing transcriptome data onto it to estimate the expression of TE loci for each species.

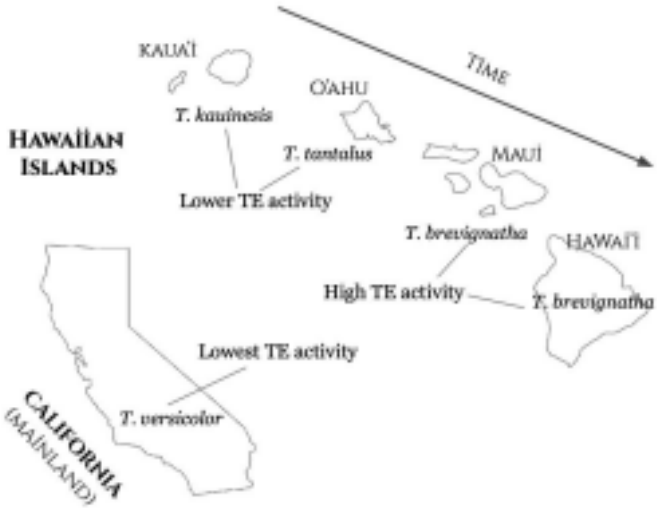

Figure 1: Proposed levels of TE activity in *Tetragnatha* species by geographic area

Using the available transcriptomes of *Tetragnatha* species, I will quantify preliminary TE levels from the transcriptomes using SalmonTE, a program that performs pseudomapping of TEs using RNAseq data. This will provide TE expression data across the archipelago that will offer a functional gene analysis of TEs, which I will compare to the full chromosome-level genome

analysis.

Spider genomes are difficult to sequence because of repetitive sequences and low GC content, so a variety of sequencing and preparation methods are needed. I will be preparing libraries for three complementary sequencing methodologies to produce a chromosome-level genome. I will obtain live specimens of the Maui species, *T. brevignatha*, through a colleague from EvoLab at UC Berkeley. To extract the DNA I will remove the legs and grind the tissue in liquid nitrogen to keep the DNA intact. Library prep for Illumina® and Hi-C will be performed in the EvoLab, and PacBio® library prep and all sequencing will occur at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

After library prep, I will perform the baseline assembly, using the high coverage long-read PacBio® sequencing. Since long reads tend to have high error rates, the medium coverage Illumina® HiSeq sequencing will help resolve issues in the baseline assembly. To get chromosome-level data, Hi-C sequencing will be needed to scaffold sequences in a chromosomal conformation. I will scaffold the contigs, or sequence fragments, using the HiRise software platform to build the final draft genome, and use the *T. brevignatha* transcriptome data to annotate the genome, and additionally use RepBase, a TE library, to annotate possible TEs.

Finally, using the genome assembly annotated with TEs, I will map transcriptomes of the different *Tetragnatha* species to determine and compare the levels of TE expressions using Telescope, a software program that provides locus-specific estimates of expression. This project will culminate in a final honors thesis and manuscript for publication as the first study to compare transcriptome and genome analyses of TEs across time in the context of adaptive radiations.

Proposed Timeline

Prior to SURF program: Review protocols; receive specimens and library prep kits Week 1: DNA Extraction; Tutorials for SalmonTE programs for transcriptomic analysis Week 2: Illumina Library Prep; Set-up server account and download transcriptome genomes Week 3: Hi-C Library Prep; Run preliminary analyses for *T. brevignatha* transcriptomes Week 4: PacBio Library Prep at QB3; Bring Illumina and Hi-C to QB3 for Sequencing; SalmonTE analysis of *T. brevignatha* transcriptome
Week 5: SalmonTE analysis of *T. tantalus, T. kauainesis*
Weeks 6: Analyze data of young and old radiating species using SalmonTE and R Week 7: Begin running assembly program on PacBio® reads; prepare figures and presentation Week 8: Review software programs for genome assembly; finalize and practice presentation Into The Semester: Resolve Assembly Errors, Scaffold Assemblies using DoveTail HiRise, annotate genome using *T. brevignatha* transcriptome, use Telescope for TE mapping and expression level estimation

Qualifications

Having taken an interdisciplinary approach to my studies, majoring in Molecular Environmental Biology and Geography and minoring in Data Science, I feel that I have the knowledge, skills, and ability to learn new information quickly to take on this interdisciplinary project. Last semester I took Practical Genomics, where I learned how the process of Illumina sequencing works, various applications used to analyze genomic and transcriptomic data, and different algorithms used to assemble genomes. Genomics is a heavily tech-oriented field, but having taken fundamental courses in computer and data science, I have strong programming skills and am capable of analyzing large amounts of sequencing data.

I will be collaborating with members of the EvoLab on this project, who have worked with the QB3 Genomics Facility and have performed the necessary protocols for DNA sequencing. This project will also complement and utilize the genome of *T. versicolor* assembled through an initiative to create genomic data for multiple species in California, the California Conservation Genomics Project, along with the available *Tetragnatha* transcriptomes. Additionally, I have been working in the lab with my faculty advisor, Dr. Rosemary Gillespie and PhD candidate, Natalie Graham since January 2019. In their lab I have used a data science approach to understand the biotic interactions among Hawai‘i arthropods, including Tetragnatha spiders and their prey. Furthermore, for a class project in Practical Genomics I analyzed novel data from high-throughput sequencing to elucidate the patterns of phylogenetic diversity across Arthropoda in Hawai‘i.

Citations

1. Schluter, D., 2000. *The ecology of adaptive radiation*. OUP Oxford.
2. Losos, J.B., 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism: American Society of Naturalists EO Wilson Award address. *The American Naturalist*, *175*(6), pp.623-639.
3. Berner, D. and Salzburger, W., 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), pp.491-499.
4. Chénais, B., Caruso, A., Hiard, S. and Casse, N., 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*, *509*(1), pp.7-15.
5. Oliver, K.R. and Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. *Bioessays*, *31*(7), pp.703-714.
6. Feiner, N., 2016. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1840), P.20161555.
7. Santos, M.E., Braasch, I., Boileau, N., Meyer, B.S., Sauteur, L., Böhne, A., Belting, H.G., Affolter, M. and Salzburger, W., 2014. The evolution of cichlid fish egg-spots is linked

with a cis-regulatory change. *Nature communications*, *5*(1), pp.1-11.
8. McClintock, B., 1984. The significance of responses of the genome to challenge. 9. Wagner, W.L. and Funk, V.A., 1995. *Hawaiian biogeography*. Smithsonian Institution Press. 10. Gillespie, R.G., 2002. HAWAIIAN SPIDERS OF THE GENUS TETRAGNATHA IV NEW, SMALL SPECIES IN THE SPINY LEG CLADE. *The Journal of Arachnology*, *30*(1), pp.159-172.
11. Cotoras, D.D., Bi, K., Brewer, M.S., Lindberg, D.R., Prost, S. and Gillespie, R.G., 2018. Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. BMC evolutionary biology, 18(1), pp.1-13. 12. Berger, Cory Ailin; Michael S Brewer; Nobuaki Kono; Hiroyuki Nakamura; Kazuharu Arakawa; Susan R Kennedy; Hannah M Wood; Seira A Adams; Rosemary G Gillespie. 2021. Shifts in morphology, gene expression, and selection underlie web loss in Hawaiian Tetragnatha spiders. BMC Evolutionary Biology, In press