Development of Molecular Markers from the Transcriptome of an Ant-Mimicking Spider for Population Genetic Analyses

Anna Schell

Faculty Mentors: Jennifer Apple and Josephine Reinhardt

Department of Biology, SUNY Geneseo, Geneseo NY

Introduction

The ant-mimicking spider, *Myrmarachne formicaria*, is a species of jumping spider that is originally native to Eurasia, but it has been sighted in North America starting in 2001 (Bradley et al. 2006). This spider is distinctive because of the way it imitates ants by moving its front two legs around like an ant's antennae. As it is not native to North America, it must have been introduced to the country, but there is not currently enough information to determine whether there was only one introduction event or several different events. By analyzing the ant-mimicking spider's genetic material for variation, it is possible to determine if there was a single introduction event from one source population, or if it arrived here through multiple dispersal events (Cristescu 2015).

In previous work, fragments of mitochondrial DNA from 27 spiders collected from three states were sequenced and analyzed for variation (Mills et al. 2021). These sequences were found to be almost identical between individuals, with only two polymorphisms occurring in a few samples. This suggested very low genetic variation within the North American *M.formicaria* population. However, these regions of mitochondrial DNA may not be the most variable within species, so it may not be ideal to use them for genetic markers to assess variation.

Fig 1. (left) A map showing where *M. formicaria* were collected for a study of mitochondrial DNA variation.
Fig 2. (right) A female *M. formicaria* near her silken shelter on a leaf.

Microsatellites instead might be a more effective molecular marker to assess variation within a species. These markers consist of a short portion of DNA, usually only a few base pairs in length, that is repeated many times in succession. An efficient and cost-effective way of doing this is to examine the transcriptome. In this project, I am searching the transcriptome for these repetitive sequences, some of which may be useful as genetic markers.

Results

The entire transcriptome contains 6954 transcripts, and 2255 of those transcripts contained three letter repeats. Out of 70 transcripts chosen with three letter repeats, 30 have had possible primers identified. Ten had significant spider protein similarities. Many of the transcripts did not translate well to a protein, and some of the ones that did translate did not match consistently with other spiders.

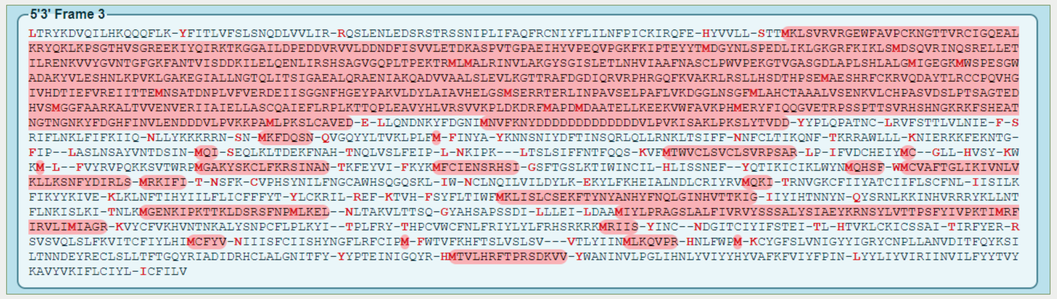
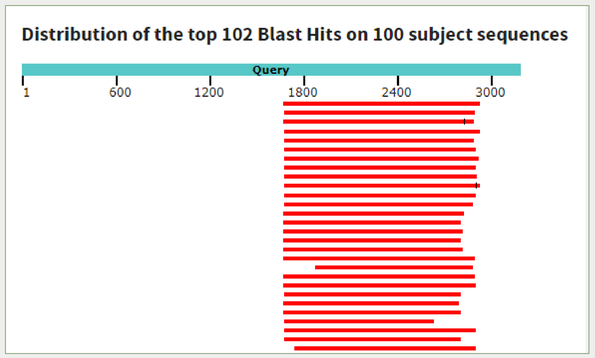


Fig 4. A translated frame of one of the spider DNA transcripts. One highlighted region is long and uninterrupted, which suggests that this region of DNA may code for a protein.



Some of the translated transcripts matched with specific known proteins from various spiders. These proteins include histidine ammonia-lyase, lipoma-preferred partner homolog, cytochrome P450 4V2, and phospholipase A2 group XV. Some transcripts only matched with uncharacterized or hypothetical proteins found in spiders, so it is possible that those regions of DNA did not code for a normal protein. Common spider matches included the common house spider (*Parasteatoda tepidariorum*), a dwarf spider (*Oedothorax gibbosus*), African social velvet spider (*Stegodyphus mimosarum*) and others in the *Stegodyphus* genus, and the wasp spider (*Argiope bruennichi*).

Fig 5. (left) A graphical summary of the BLAST results which indicates that most similar sequences matched to the DNA at about 1700-2800 base pairs.

Description	Scientific Name	Common Name	Max Score	Query Cover	E value	Per. Ident
phospholipase A2 group XV-like [Uloborus diversus]	Uloborus diversus	NA	624	39%	0.0	69.30%
phospholipase A2 group XV [Parasteatoda tepidariorum]	Parasteatoda tepidariorum	common house spider	608	38%	0.0	69.78%
Phospholipase A2 group XV like protein [Argiope bruennichi]	Argiope bruennichi	NA	607	38%	0.0	69.38%
Group XV phospholipase A2 [Stegodyphus mimosarum]	Stegodyphus mimosarum	NA	602	39%	0.0	66.19%
Group XV phospholipase A2 [Araneus ventricosus]	Araneus ventricosus	NA	600	38%	0.0	67.65%
phospholipase A2 group XV [Trichonephila inaurata madagascariensis]	Trichonephila inaurata madagascariensis	NA	588	38%	0.0	66.50%
phospholipase A2 group XV-like [Stegodyphus dumicola]	Stegodyphus dumicola	NA	588	39%	0.0	64.90%
phospholipase A2 group XV [Trichonephila clavipes]	Trichonephila clavipes	NA	588	38%	0.0	66.50%
phospholipase A2 group XV [Nephila pilipes]	Nephila pilipes	giant wood spider	587	38%	0.0	66.18%

Using BLAST, it was also possible to find out what part of the spider sequence the database was matching known proteins against, as well as how well they matched. Known protein matching ranged from 10% to 60% depending on the transcript.

Fig 6. BLAST results showing which species and proteins matched well with *M.*formicaria's DNA transcript.

Methods

The transcriptome was assembled using RnaSPAdes, a de novo transcriptome assembler available through the Galaxy bioinformatics platform. The software Gramene (Tello-Ruiz et al. 2022) was used to identify all transcripts containing repeats.

Using the software program R (R Core Team, 2023) and RStudio (RStudio Team, 2023), the output from Gramene was filtered to only include sequences that have three letter repeats in them, as these would be less likely to disrupt an open reading frame. This set of transcripts was then sorted by the number of repeats, as larger microsatellites are more likely to show polymorphism.

Starting with sequences that contain the most repeats, the transcripts were entered into Primer3 to search for potential primers that can bind to the DNA surrounding the repeated sequences. These primers can be used later to amplify the DNA using PCR.

```
121 AAAAAATCTAACTAGCGTTCTTTAAAACTTGACAAAAGCGCCAGTCCATATTCTCTTCAA
                              >>>>>>>>>>>>>>>>>>>>

181 TAGCCACCATCTTTTATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT
                    ****************************************

241 AATAATAATAATAATAATAATAATAATAATAATAAAAGAAAAATCTAAAATATCGAGTAA
    **********************************

301 TCACAAGTCTAGAACGTATGCTTACACTTTTTTGTATTTTCAGCCCTAAACATAATCTAT
    <<<<<<<<<<<<<<<<<<<<<<<<
```

Fig 3. A portion of a *M.formicaria* DNA transcript, with the repeated sequences signified by asterisks and the pair of primers signified by arrows.

After identifying promising transcripts of DNA, the sequences were translated to see if they contained a protein-coding sequence. In addition, the sequences were run in BLAST, a program which compares a segment of DNA to sequences contained in a database. This was done to confirm the identity of the DNA and ensure that it belonged to spiders, and not to microbial contaminants.

Next Steps

Many of the spider species that showed similarities with *M. formicaria* sequences were not very closely related to jumping spiders. This could be due to the fact that there has been relatively little work done involving spider genomics, including jumping spiders. However, it is promising that the transcripts seem to be related to other known spider proteins.

Matching primers have been ordered for several of the most promising transcripts, and they will be used in PCR to amplify the microsatellite. The products of the PCR will be run on an agarose gel to see if the amplification was successful. If so, products will be sent to an external lab for fragment analysis using an automated sequencer. These markers will then be further tested to see if they show enough variability to be useful tools for population genetic analyses, both in the spider's native and introduced ranges.

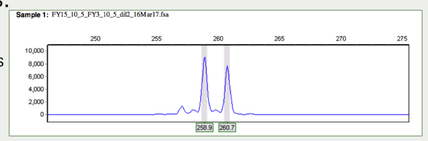


Fig 7. Fragment analysis output showing a heterozygote with a 259 and a 261 allele.

References

Bradley, R. A., Cutler, B., and Hodge, M. (2006). The first records of *Myrmarachne formicaria* (Araneae, Salticidae) in the Americas. The Journal of Arachnology, 34(2), 483-484.

Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular Ecology*, 24, 2212–2225.

Mills, C., Fleishman, D., and Apple, J.L. (2021, June 24-July 1). Genetic diversity in a nonnative ant-mimicking spider in its invaded range [Poster presentation]. American Arachnological Society Annual Meeting (virtual).

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

R Studio Team (2023). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Tello-Ruiz, M. K., Jaiswal, P., & Ware, D. (2022). Gramene: a resource for comparative analysis of plants genomes and pathways. In *Plant Bioinformatics: Methods and Protocols* (pp. 101-131). New York, NY: Springer US.

Acknowledgments

Cassidy Mills performed the RNA isolation from *M. formicaria* making the assembly of its transcriptome possible. Cassidy Mills and Daniel Fleischman carried out the initial mitochondrial DNA work. We are also grateful to the Geneseo Student Association and the Geneseo Foundation for research funding through the Undergraduate TRAC Grant Program.

GENESEO
DEPARTMENT OF
Biology