I studied the greenbug aphid (Schizaphis graminum), which has been recognized as a pest of grain crops and grasses for over 150 years and was first reported in North America in 1882 (University of Florida, Featured Creatures). In warm climates, most insects are female and they reproduce via parthenogenesis (Kansas State University, Sorghum Insects). Greenbugs are agriculturally important because they can feed on over 70 different species of plants. The basis for this project was to attempt to improve the greenbug aphid genome assembly using in order to better understanding its biology and why it have different host plants unlike other aphids. The question was, can we use sequencing data generated from 10X Chromium to improve scaffolding of the existing assembly? The results show that we did in fact improve the genome scaffolding using BWA and ARCS. These results are important because improving the genome allows us compare the greenbug genome to other aphids and measure gene expression (mRNA levels) as it feeds different plants to understand why it has such a broad host range.

**Method and Experimental Design**

We leveraged several unix-based command programs to integrate 10X Chromium and transcriptome data into the greenbug assembly.

1) Initially learned a Linux operating system and how to use the command line.
2) We mapped 150 x 150 nt PE Illumina HiSeq X-Ten data from 10X libraries to the existing assembly using a short read aligner called BWA (Li and Durbin, 2009).
3) Then we used ARCS scaffolding tool to identify PE reads that could be used to join scaffolds (Yeo et al, 2017).
4) Additionally, we mapped mRNA transcripts from the de novo transcriptome assembly to the newly assembled scaffold assembly using the blat program (Kent, 2002).
5) Finally the SCUBAT program was used to look for identify transcripts that spanned more than one scaffold in order to assemble them.

**Conclusions**

Overall, the refinement and improvement of the greenbug aphid genome is significant in understanding the biology of these insects. This information can now be used to compare the greenbug genome to other aphid species to identify factors that enable it to live on so many more species of grasses than other aphids can.

**Future Directions**

Although 10X Chromium data significantly improved scaffolding, we could also integrate long-read sequencing data to fill some of the gaps in this assembly using a program called PBelly. After improving the genome assembly to the best of our abilities with currently available library technologies and assembly algorithms, it can then be used to understand more about the insect itself.

The next step in research would be a combination observing how greenbug aphids interact with their different plant species and also by comparing gene expression levels (mRNA) of the aphids as they feed on different hosts using an approach called RNA-Seq with the genome as a reference. We could also comparing the genome to the genomes of other aphids to identify characteristics that may allow greenbug to feed on a wider variety of hosts.