Green Bug Aphid Genome

Connie McDowell[1] and Dr. Erin Scully[2,3]
[1]Department of Computer Engineering, College of Engineering, Kansas State University
[2]USDA-ARS-CGAHR, Stored Product Insect and Engineering Research Unit, Manhattan, KS
[3]Department of Entomology, College of Agriculture, Kansas State University

Undergraduate Research Experience in Entomology

K-State 2025 K-STATE Research and Extension

Abstract

I studied the greenbug aphid (*Shizaphis 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.

Purpose

Despite its broad host range, no genomic resources are available for greenbug, which would allow us to identify genetic factors that allow it to consume so many different types of plants. The purpose of this work is to try and improve the existing greenbug genome assembly.

Questions, Hypotheses, and Predictions

Question: Can we use short read sequencing data generated from 10X Chromium libraries and transcriptome data to improve the genetic scaffolding on the greenbug aphid?

Hypothesis: 10X data and the transcriptome will improve the previous genome.

Study System

The greenbug "wheat" aphid (*Schizaphis graminum*) is a major pest of sorghum and other bioenergy grasses in the same family, along with wheat. This particular aphid can attack over 70 different grass species, which is troublesome because they can also vector diseases, such as barley yellow dwarf virus, and salivary proteins injected into the plant during feeding break down chlorophyll and plant cell walls. Overall, the reduction in chlorophyll results in chlorosis, or yellowing of the aerial tissues, which significantly reduces the photosynthetic potential of the plant. Combined with nutrient loss from aphid feeding, heavily infested plants can appear stunted in growth, which ultimately results in significantly reduced grain and biomass yields. Aphids also produce honeydew which can allow mold to grow on the plants and interferes with harvesting equipment. The aphids which are of the order hemiptera are parthenogenic meaning they can reproduce without a male but can sometimes switch to a sexual stage if needed and although typically wingless, can produce wings when stressed.

http://entomology.k-state.edu/extension/insect-information/crop-pests/sorghum/greenbug.html

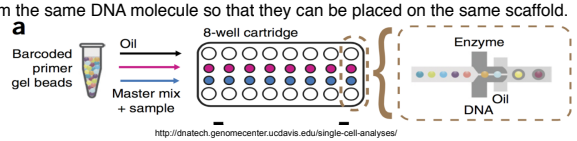
http://txppipm.blogspot.com/2014/01/new-invasive-aphid-on-sorghum-in-texas.html

Methods and Experimental Design

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

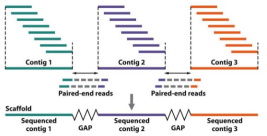
1)I initially learned a Linux operating system and how to use the command line.
2)We mapped 150 x 150 nt PE Illumina HiSeq X-Ten data from 10X libraries to the existing assembly using a short read aligner called BWA (Li and Durbin, 2009).
3)Then we used ARCS scaffolding tool to identify PE reads that could be used to join scaffolds (Yeo et al, 2017).
4)Additionally, we mapped mRNA transcripts to from a *de novo* transcriptome assembly to the newly scaffolded 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.

A. 10X Library prep: 1 nanogram of high molecular weight (HMW) DNA recovered from a single insect is partitioned into microfluidic chambers that each hold 1 molecule of DNA. All pieces of DNA from the same molecule are tagged with the same barcode. This information is exploited by the assembler to identify fragments from the same DNA molecule so that they can be placed on the same scaffold.

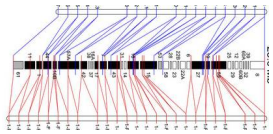


http://dnatech.genomecenter.ucdavis.edu/single-cell-analyses/

B. Because the sequencing libraries were derived from HMW DNA, it has the potential to create high quality genomic scaffolds using long-range paired end linkages.



https://www.slideshare.net/AdrinBezOrtega/longap-an-integrated-genome-assembly-platform-for-ion-torrent-data



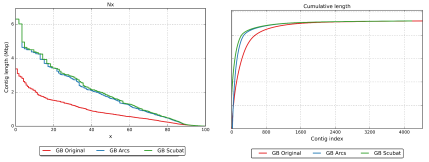
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137746

Results

	Existing Greenbug Assembly
Number Contigs	30,056
Number Scaffolds	4,419
N50 Contig	26,585
N50 Scaffold	703,343
Max Contig	200,512
Max Scaffold	3,238,898
Number of Gaps	25,637
Total Assembly Length (with Ns)	334 Mb
BUSCO	S:90.8%, D:2.8%
	F:2.1%, M:4.3%

Arcs Scaffolded Greenbug Assembly	SCUBAT Scaffold Greenbug Assembly
29,712	29,677
4,075	4,040
27,717	27,200
1,577,655	1,684,793
200,512	200,512
5,935,080	5,935,080
25,637	25,637
334 Mb	334 Mb
S:91.1%, D:2.6%	S:91.3%, D:2.5%
F:2.1%, M:4.2%	F:2.0%, M:4.2%

ARCs scaffolding increased the N50 scaffold and max scaffold lengths by ~50% and improved BUSCO recovery.
In addition, a higher percentage of the genome was placed in fewer scaffolds after ARCs. Few transcripts spanning scaffolds were identified using SCUBAT indicating that few gene models were broken across scaffolds.

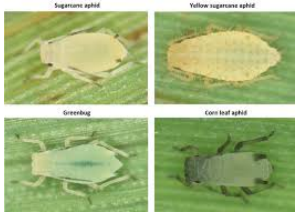


Conclusions

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 PBJelly. 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.



https://www.morningagclips.com/sugarcane-aphid-occurrence-in-aug/

References

Barlow, V. M. and Godfrey, L.D. UC Pest Management Guidelines "Small Grains". 2007. *Univeristy of California Agriculture and Natural Resources* 3466.

Kent, W.J. 2002. BLAT-The BLAST-Like Alignment Tool. Genome Research 12, 656-664.

Li, H.,and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760.

Michaud, J.P. 2014. Sorghum Insects. *Kansas State University*.

Nagata, Russell T. and Nuessly, Gregg S. 2005. Featured Creatures "greenbug". *University of Florida EENY-353*.

SCUBAT (Scaffolding Contigs Using BLAT And Transcripts): https://github.com/elswob/SCUBAT

Yeo, S., Coombe, L., Chu, J., Warren, R.L., and Birol, I. 2017. ARCS: Assembly Roundup by Chromium Scaffolding. bioRxiv: doi: http://dx.doi.org/10.1101/100750

Acknowledgements

We thank Dr. Scott Geib, USDA-ARS-Pacific Basin Agricultural Research Center Hilo, HI and Drs. Scott Sattler, Nathan Palmer, and Gautam Sarath, USDA-ARS-Wheat, Sorghum, and Forage Research Unit Lincoln, NE. Funding for this project was provided by USDA-NIFA Grant ID 2011-67009-30026entitled "The impacts of lignin modifications on fungal pathogen and insect interactions in sorghum for cellulosic and thermal bioenergy" awarded to SS and USDA-NIFA Grant ID 2011-67009-30096 entitled "Mitigating insect herbivory of warm-season bioenergy grasses – getting ahead of the curve" awarded to GS.