

# Michael Alonge | Computational Biologist

(831) 201-2788 • malonge11@gmail.com • <https://github.com/malonge> • michaelalonge.com

## CAREER SUMMARY

---

I am interested in how eukaryotic genome structure influences function. As a Research Associate at Driscoll's and a Ph.D. student at JHU, I have used bioinformatics tools and developed new methods to study genome structure. I developed the widely used RagTag scaffolder which I used to scaffold a highly complete wheat genome assembly. I have worked closely with geneticists to discover natural structural variation in tomato causing important domestication and improvement phenotypes. I recently aided with the complete and near-complete assembly of a human and Arabidopsis thaliana genome, respectively.

## EDUCATION

---

### Ph.D. Candidate, Computer Science

Johns Hopkins University, Baltimore, MD

Sep 2017 – Nov 2021 (expected)

M.S.E. awarded August 2019

*Advised by Michael Schatz*

### B.S., Biomolecular engineering

UC Santa Cruz, Santa Cruz, CA

Sep 2009 – May 2014

## SKILLS

---

- Genome assembly
- Comparative genomics
- Pan-genomics
- Variant discovery
- Bioinformatics
- Scientific computing
  - Python/R
  - Numpy/Scipy/Pandas
  - Grid computing
- Adobe Illustrator
- git/GitHub
- some basic molecular biology lab experience

## PUBLICATIONS

---

**Google Scholar** - [bit.ly/33e8ICL](https://bit.ly/33e8ICL)

Mc Cartney, Shafin and Alonge, et al. "Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies." *bioRxiv* (2021).

<https://doi.org/10.1101/2021.07.02.450803> \*

Naish, **Alonge** and Wlodzimierz, et al. "The genetic and epigenetic landscape of the Arabidopsis centromeres." *bioRxiv* (2021). <https://doi.org/10.1101/2021.05.30.446350> \*

**Alonge** and Wang, et al. "Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato." *Cell* (2020).  
<https://doi.org/10.1016/j.cell.2020.05.021> \*

**Alonge** and Shumate, et al. "Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies." *Genetics* (2020).  
<https://doi.org/10.1534/genetics.120.303501> \*

**Alonge, Michael**, et al. "RaGOO: fast and accurate reference-guided scaffolding of draft genomes." *Genome biology* 20.1 (2019): 1-17. <https://doi.org/10.1186/s13059-019-1829-6> \*

**Alonge, Michael**, and Michael C. Schatz. "A master regulator of regeneration." *Science* 363.6432 (2019): 1152-1153. <https://doi.org/10.1126/science.aaw6258> \*

Nurk, Sergey, et al. "The complete sequence of a human genome." *bioRxiv* (2021).  
<https://doi.org/10.1101/2021.05.26.445798>

Pereira, Lara, et al. "Natural genetic diversity in tomato flavor genes." *Frontiers in Plant Science* 12 (2021): 914. <https://doi.org/10.3389/fpls.2021.642828>

Chen, Li-Yu, et al. "The bracteatus pineapple genome and domestication of clonally propagated crops." *Nature Genetics* (2019): 1-10. <https://doi.org/10.1038/s41588-019-0506-8>

Wenger, Aaron M., et al. "Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome." *Nature Biotechnology* (2019): 1-8.  
<https://doi.org/10.1038/s41587-019-0217-9>

Soyk, Sebastian, et al. "Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato." *Nature Plants* (2019): 1.  
<https://doi.org/10.1038/s41477-019-0422-z>

Buti, Matteo, et al. "The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry)." *GigaScience* 7.4 (2017): 1-14.  
<https://doi.org/10.1093/gigascience/giy010>

\* Primary or co-primary authorship

## RESEARCH

---

### **The Tomato Structural Variant Landscape**

By sequencing 100 tomato varieties with long-reads, we established the first large-scale database of natural structural variants in any crop. Genome-wide, I found that breeding introgressions from wild material substantially altered the genome structure of modern varieties. At specific loci, I used bioinformatics techniques to resolve large gene duplications underlying traits.

### **Genome assembly scaffolding and improvement algorithms**

I developed RaGOO/RagTag, an open-source collection of software tools for scaffolding and improving modern genome assemblies. RagTag performs homology-based assembly correction, scaffolding, and patching/gap-filling. It also can merge scaffoldings from multiple genomic maps. <https://github.com/malonge/RagTag>

### **High-quality reference genome assembly**

I have contributed to many human and plant reference genome assembly projects. As a member of the T2T consortium, I used the latest sequencing technologies to polish and patch the first-ever complete human reference genome. After scaffolding a highly repetitive wheat genome assembly, I analyzed pervasive gene duplications missing from the reference. I have led multiple plant genome assembly projects with a focus on the Solanum. I have recently assembled a near-complete *Arabidopsis thaliana* genome including all five centromeres.

## TEACHING

---

**Teaching Assistant, Johns Hopkins, Computational Genomics: Sequences** **Fall 2019**

I created grading rubrics and assisted CAs in grading assignments and tests. I also answered questions on an online question board and held weekly office hours. Additionally, I provided guidance for research project groups and helped grade final project oral presentations.

*Course taught by Ben Langmead, Ph.D.*

## PROFESSIONAL EXPERIENCE

---

**Research Associate I and II: Computational Biology**  
(I) Nov 2014 – Mar 2017 • (II) Mar 2017 – Aug 2017

**Driscoll's, Watsonville CA**

I led computational biology projects with the ultimate goal of assisting breeders and molecular biologists. I especially focused on establishing reference genome assemblies to facilitate downstream genomics analyses. I also assisted with wet-lab protocols and plant phenotyping.  
*Supervised by Judson Ward, Ph.D.*

## JOURNAL/CONFERENCE REVIEWS

---

• Nature Genetics      • Genome Biology      • ISMB      • Nature Communications

## CONFERENCES AND CONSORTIA

---

### Talks

**The Boyce Thompson Institute Nanopore Symposium 2019**, Ithaca, NY  
*Using ONT sequencing to resolve complex structural variants in plants*

**Galaxy Works “Structural Variant Detection using ONT Data” Webinar 2021**, Virtual  
*SVs, Introgressions, and the Necessity of Crop Pan-Genomics*

### Posters

**Intelligent Systems for Molecular Biology 2018**, Chicago, IL  
*RaGOO: Fast reference-guided ordering and orienting of genome assembly contigs*

**Biological Data Sciences 2018**, Cold Spring Harbor, NY  
*RaGOO: Fast reference-guided ordering and orienting of genome assembly contigs*

**Plant & Animal Genome 2019**, San Diego, CA  
*RaGOO: Fast reference-guided ordering and orienting of genome assembly contigs*

**Genome Informatics 2019**, Cold Spring Harbor, NY  
*Long-Read Sequencing Reveals the Structural Variant Landscape in Tomato*

**Solanaceae 2019**, Jerusalem, Israel (“Best poster” awardee)  
*Long-Read Sequencing Reveals the Structural Variant Landscape of Tomato*

**Plant & Animal Genome 2020**, San Diego, CA  
*Major impacts of widespread structural variation on gene expression and crop improvement in tomato*

**Solanaceae 2020, Virtual**

*Towards ultra-accurate and complete chromosome-scale Solanaceae reference genomes*

**REFERENCES**

---

**Zachary Lippman, Ph.D.**

*Professor and HHMI Investigator, CSHL*  
lippman@cshl.edu

**Michael Schatz, Ph.D.**

*Bloomberg Distinguished Professor of  
Computer Science and Biology, JHU*  
mschatz@jhu.edu

**Judson Ward, Ph.D.**

*Founder and CTO, Ohalo Genetics*  
jud@ohalogenetics.com

**Steven Salzberg, Ph.D.**

*Professor of Biomedical Engineering,  
Computer Science, and Biostatistics and  
Director of the Center for Computational  
Biology, JHU*  
*McKusick-Nathans Institute of Genetic  
Medicine, JHU School of Medicine*  
salzberg@jhu.edu