

Ohio University

OHIO Open Library

All Forum Articles

College of Arts & Sciences Forum

7-5-2019

Research in Hua Lab Makes Breakthrough on Genomics Studies and Teaching

Ohio University College of Arts & Sciences

Follow this and additional works at: https://ohioopen.library.ohio.edu/cas_forum_all

Recommended Citation

Ohio University College of Arts & Sciences, "Research in Hua Lab Makes Breakthrough on Genomics Studies and Teaching" (2019). *All Forum Articles*. 6730.
https://ohioopen.library.ohio.edu/cas_forum_all/6730

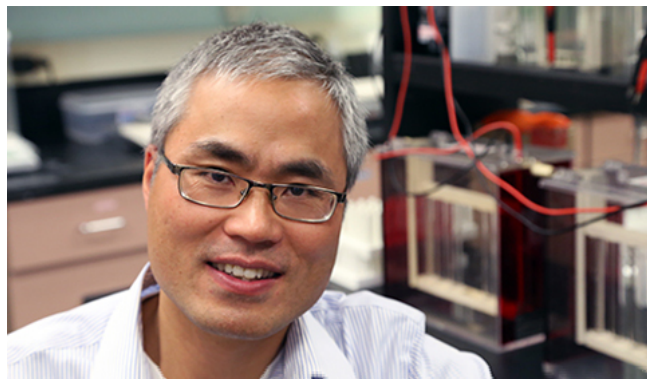
This News Article is brought to you for free and open access by the College of Arts & Sciences Forum at OHIO Open Library. It has been accepted for inclusion in All Forum Articles by an authorized administrator of OHIO Open Library. For more information, please contact deborded@ohio.edu.

Research in Hua Lab Makes Breakthrough on Genomics Studies and Teaching

July 5, 2019

Categories: Research

Tags: environmental and plant biology news, environmental and plant biology research, Faculty Research, Molecular and Cellular Biology news, Molecular and Cellular Biology Research, research, Zhihua Hua



Dr. Zhihua Hua

With the significant improvement of DNA sequencing technologies, more and more genome sequencing data are made available to the public. However, to understand these genome sequences is not easy because they seem to be randomly and repetitively written by four letters, A, T, C, and G, which are called nucleotides, the building blocks of DNA.

How to convert these genome sequencing data into meaningful biological information is one of the most challenging research topics in biology. A recently published study from [Dr. Zhihua Hua's](#) laboratory in the [Environmental & Plant Biology Department](#) and the [Molecular & Cellular Biology](#) program at Ohio University tackles this challenging question using bioinformatics, an interdisciplinary approach combining mathematical modeling, computational programming, and evolutionary biology analysis.

Hua's lab developed an approach using the evolutionary history of a gene superfamily and a tool capable of crunching through data from long-dead siblings as well as cousins in other plants. The results can help uncover previously unknown members of a gene family and help in analyzing the functions of genome sequences.

Starting with the Impossible

Hua developed this interest when he worked with [Dr. Richard D. Vierstra](#) at the University of Wisconsin-Madison. He was challenged by Vierstra to understand the biological functions of 700 unknown genes in *Arabidopsis thaliana*, a mustard weed plant that has been widely used for understanding fundamental biological questions. However, it is not possible for a biologist to characterize the functions of 700 unknown genes. One unknown gene could even take the entire career of a biologist.

This challenge pushed Hua to find solutions outside the box. He developed an idea to let the evolutionary history predict the winners of these 700 unknown genes based on the theory that their activities in the current *Arabidopsis* genome have been selected through millions of years of evolution. Unfortunately, it was not easy to find the evolutionary history of these genes because they were born in the ancestors of *Arabidopsis* at different time periods. Some of them had many siblings that were dead before *Arabidopsis* emerged in the history; some are orphans that are only present in *Arabidopsis*; and some have cousins in many other species.

The key to find these relationships is to discover the siblings and cousins of these unknown genes in many other genomes, which are written by A, T, C, and G. The rapid accumulation of sequenced genomes benefited his approach.

Tracking a Gene Superfamily

In biology, a group of genes that share similar sequence features is called gene superfamily. The 700 unknown genes that Hua was challenged on belong to one particular gene superfamily, called *F-box*, whose protein products recognize abnormal and/or unwanted proteins in all eukaryotic cells to target them for degradation. Many of its known members have been demonstrated to play important roles in a wide range of life processes in plants as well as in humans, including the cell cycle, tumorigenesis, circadian rhythms, etc.

When Hua started to look for siblings and cousins of his 700 unknown genes, he recognized that there were many unreported *F-box* genes that were hidden in many sequenced genomes. How to find these hidden genes became another challenge. After much practice, Hua developed a mathematical algorithm, called Closing-Target-Trimming. The algorithm is to iteratively trim away the members of *F-box* genes that have been reported so that the unreported ones will be exposed in the genome and can then be further carefully annotated.

This algorithm was first published in his postdoctoral work in 2011 and was recognized as a top 10 percent most-cited work of the journal, *PLoS ONE*, in its 10th anniversary in 2017.

Building the Tool to Find New Superfamily Members

How to apply the Closing-Target-Trimming algorithm to find unreported superfamily members was also challenging. In order to read through all the nucleotides in a genome, the best way is to use a computer. This requires a good skill in computational programming.

With very little prior programming experience, Hua urged himself to master one very efficient text-mining program, called Perl, and he developed a Perl package to deploy the Closing-Target-Trimming algorithm in finding unreported superfamily members in any genomes. His algorithm and program turned out to be very effective in genomic studies of gene superfamilies.

At his own lab at Ohio University, he and his group members have applied this approach to expand the research into other gene superfamilies. To date, the Hua lab has published six articles in this field in several prestigious journals, including *Cell*, *Plant Journal*, *Peer J*, *International Journal of Molecular Sciences*, and *PLoS ONE*.

To develop vigorous research and teaching programs, Hua has been striving to transmit his research skills to students both in his lab and in his classes and also to the science community. In the past five years as a faculty member at Ohio University, he developed an upper-level course, P BIO 4280/5280 Laboratory in Genomics Techniques (also cross-listed as MCB 5280), in teaching genomics techniques. He published a textbook for this course in early 2019 to introduce programming and genome evolutionary analysis skills to biology students across the campus.

To further benefit biologists with little programming experience in studying the genomics of gene superfamilies, Hua and his undergraduate student, **Matthew Early**, recently published two programming packages, **CTT** and **CTTdocker**, that employ the Closing-Target-Trimming algorithm for discovering hidden superfamily members in any sequenced genomes. This study is remarkable in that it demonstrated that while the current genome annotations have on average 15 percent of the genes undiscovered, they can be re-discovered by these programs. The study has also demonstrated the effectiveness of the programs. For example, the CTT program was able to go through 40.9 billion nucleotides and find new genes within four days. This work is equivalent to read through 15 different volumes of the American Heritage Dictionary of the English Language to find sentences that share a similar pattern within one hour.

This work has been published recently on PLoS ONE in an article titled "[Closing target trimming and CTTdocker programs for discovering hidden superfamily loci in genomes.](#)"

These research projects involved two graduate students, [Paymon Doordian](#) and [Peifeng Yu](#); two PACE undergraduate students, **William Vu** (Chemistry & Biochemistry major) and **Early** (Computer Science major); and one visiting scholar, Dr. Zhenyu Gao, from Dr. Hua's lab. The research was funded by a number of sponsors, including Ohio University Startup fund, PACE funds, OURC grant, Baker grant, and National Science Foundation-CAREER Award to Hua.

- See "[Hua's NSF CAREER Grant Will Involve Big Data, Many Students.](#)"

References

- Hua, Z.*, and Early, M.J.(u) (2019). [Closing Target Trimming and CTTdocker programs for discovering hidden superfamily loci in genomes](#). *PLoS ONE* 14, e0209468.
- Hua, Z.*, and Yu, P.(g) (2019). [Diversifying evolution of the ubiquitin-26s proteasome system in Brassicaceae and Poaceae](#). *Int J Mol Sci*, 20, 3226.
- Hua, Z.*, and Gao, Z.(v) (2019). [Adaptive and degenerative evolution of the S-Phase Kinase-Associated Protein 1-Like family in Arabidopsis thaliana](#). *PeerJ* 7, e6740.
- Marshall, R.S., Hua, Z., Mali, S., McLoughlin, F., and Vierstra, R.D. (2019). [ATG8-Binding UIM proteins define a new class of autophagy adaptors and receptors](#). *Cell* 177, 766-781.
- Hua, Z.* (2019). [Data Oriented Genomics Techniques](#) (Textbook for PBIO 4/5280, MCB5280). Figshare.
- Hua, Z.* (2019) [CTTdocker Package](#).
- Hua, Z.*, and Early, M.J.(u) (2018). [CTT Perl Package](#).
- Zhang, G., and Hua, Z.* (2018). [Genome comparison implies the role of Wsm2 in membrane trafficking and protein degradation](#). *PeerJ* 6, e4678.
- Hua, Z.*, Doroodian, P.(g), and Vu, W.(u) (2018). [Contrasting duplication patterns reflect functional diversities of ubiquitin and ubiquitin-like protein modifiers in plants](#). *Plant J* 95, 296-311.
- Hua, Z., Zou, C., Shiu, S.H., and Vierstra, R.D. (2011). [Phylogenetic comparison of F-Box \(FBX\) gene superfamily within the plant kingdom reveals divergent evolutionary histories indicative of genomic drift](#). *PLoS ONE* 6, e16219.