

QIHUI ZHU

684 Farmington Ave, West Hartford, CT 06119

Lab: 860-837-2439 Cell: (706)-201-8288 Email: qihui.zhu@jax.org

AREAS OF EXPERTISE

✓ NGS Data Analysis	✓ Cytogenetics	✓ Microarray
✓ Genomics	✓ Bioinformatics	✓ SNP/SV/InDel
✓ MySQL	✓ UNIX platform	✓ PERL/R Programming

RESEARCH EXPERIENCE

JACKSON LABORATORY FOR GENOMIC MEDICINE

2013 - present

Associate Research Scientist

Development and Maintenance of Clinical CGH/SNP array analysis platform

- ▶ Took charge of the Agilent comparative genomic hybridization (CGH) and single nucleotide polymorphisms (SNP) array platforms, which have been widely used in Brigham and Woman's Hospital Clinical Cytogenetics Laboratory, Dana Farber/Harvard Cancer Center Cytogenetics Core and Jackson Laboratory for Genomic Medicine to generate the reports for the cytogeneticists and facilitate clinical diagnostics.
- ▶ Designed and developed new analysis software for Affymetrix Cytoscan[®] HD array to detect chromosomal imbalances and visualize with easy-to-use graphical interface. This platform currently being tested provides the broadest coverage and highest performance for detecting human chromosomal aberrations that associate with diseases.
- ▶ Created clinical copy number variation (CNV) aberration database to facilitate clinical research and updated monthly.

Design of customized high-resolution genome-wide microarray

- ▶ Designed 8X60K customized microarray to identify variants associated with congenital diaphragmatic hernia (CDH).
- ▶ Applied this microarray to more than 1500 patients and controls in the CDH cohort.

- ▶ Identified structural variants that are significantly associated with CDH risk.

BRIGHAM and WOMEN'S HOSPITAL, HARVARD MEDICAL SCHOOL

2011 - 2013

Postdoctoral Researcher

Analysis of Deep Sequencing Data in the Multiple Primate Genomes

- ▶ Developed a pipeline to detect SNPs and Insertion-deletion polymorphisms (INDELs) by combining multiple algorithms (GATK, Samtools, and Pindel, etc) in 20 deep sequenced human and nonhuman primate samples. The calling and filtering parameters have been adjusted to provide the best results for this dataset. The sensitivity and specificity of identified SNPs are 99.87% and 95.12%.
- ▶ Discovered hotspots of structural variation that are evolutionarily relevant and threw new lights upon genetic susceptibility to auto-immune and congenital diseases.
- ▶ Scanned 20,000+ homologous genes in the primate lineages and detected different selective pressures on the different categories of disease genes to help find new targets for drug discovery and understand how heritable diseases arise.

Investigation of Balancing Selection on Non-coding Regions with Regulatory Function in Human Genome

- ▶ Developed a new approach to identify functional non-coding regions associated with human phenotypic variation and disease at the genome level. Provided the first experimental data to strongly support the notion that ancient variation can indeed affect regulatory function in modern humans. Complemented and expanded the recent comparative genomic analyses of archaic human genomes.
- ▶ Identified a ~36 kb locus (NE1 locus) at the upstream of *APOBEC3* locus that has an ancient substructure, predating Human-Neandertal sequence divergence and is therefore likely to be under balancing selection.
- ▶ Characterized this locus at the population level by 1,092 samples from 15 populations in the 1,000 Genomes Project.
- ▶ Investigated structural variants within and around the *APOBEC3* locus and their functional roles against retroviral infections, such as HIV.

UNIVERSITY OF GEORGIA

2007-2011

Postdoctoral Research Associate

Identification of Rare Point Mutations in Large Populations by Next-Gen Sequencing (NGS)

- ▶ Conducted pioneering research on plant breeding by using NGS technology.
- ▶ Constructed a 6X genomic library of orphan crop *Eragrostis tef* to identify ‘green revolution’ dwarfing genes by PCR-based library screening.
- ▶ Planted and pooled 21,210 EMS mutagenized lines based on sample size calculation. Sequenced quantified PCR amplicons of candidate genes by 454 technology (over 400,000 reads) with 1,058 pairs of tagged primers.
- ▶ Developed a pipeline in PERL after evaluating multiple existing alignment and SNP calling approaches to align each read to two homoelogenous reference sequences to identify SNPs and filter candidates based on various criteria due to strong PCR/sequencing bias and PCR/sequencing errors.
- ▶ Identified candidate mutations that associated with dwarfing stature and validated 7 independent mutations by Sanger sequencing. Facilitated to increase grain yield.

Analysis of Population Structure and Evolutionary Impact of Disease Resistance Genes (*R* Genes)

- ▶ Built an *R* gene database by degenerate primer amplification and EST database mining of a biofuel species switchgrass (*Panicum virgatum*). Screened fosmid library and sequenced 12 fosmids containing target *R* genes.
- ▶ Analyzed *R* gene nucleotide polymorphisms and population structure from seven representative populations. Provided a useful basis for germplasm management. Estimated allelic diversity and provided a valuable criterion for sampling strategy.
- ▶ Scored common rust and blast diseases from 63 populations and conducted association study between *R* gene haplotypes and disease responses.

Transposable Elements (TEs) Analysis in the maize and *selaginella* Genome

- ▶ Developed a pipeline to integrate a few existing computational methods of annotating TEs. Annotated all types of TEs in the lycophyte genome *Selaginella* by structure, mathematical and homology based methods.
- ▶ Compared TEs in *Selaginella* with other lineages and discovered *Selaginella* contains a wealth of TEs despite its small genome size. Discovered recent activity and estimated their insertion times. Performed comparison of the assembled whole genome scaffolds and the unassembled shotgun reads in maize and *selaginella*. Provided informative evidence of assembly bias and evaluated assembly quality.

Genome-wide Comparative Analysis of Transcription Factors (TFs)

- ▶ Collected all TFs from two subspecies of Asian cultivated rice, *Oryza sativa* ssp. *japonica* and *O. sativa* ssp. *indica*.
- ▶ Identified homologous TF pairs and removed false positives.
- ▶ Estimated the rate of evolution for putative homologies and explored natural selections on some TFs.
- ▶ Validated the candidate TFs with some differentiations in *japonica* rice and *indica* rice by biological experiments at population level.

Construction of the Database of Poplar Transcription Factors (DPTF: <http://dptf.cbi.pku.edu.cn>)

- ▶ Predicted TFs in the whole genome of *Populus trichocarpa* by computational approaches and manual curation.
- ▶ Annotated TFs extensively, including similarity searches against major databases (Uniprot, RefSeq, EMBL, TRANSFAC etc.) and data mining from EST and microarray data. Provided comprehensive information for the putative TFs.
- ▶ Designed the database in MySQL, built a web browser of DPTF by PHP.

INSTITUTE OF BOTANY, CHINESE ACADEMY OF SCIENCES, BEIJING, CHINA 2002-2005
Research Assistant

Analysis of Multi-locus Nucleotide Variations

- ▶ Collected all the important lineages and their wild ancestors of cultivated rice.
- ▶ Detected a dramatic reduction of diversity in cultivated rice relative to its wild progenitors.
- ▶ Simulated bottlenecks by coalescence under various lengths and population sizes to better understand the domestication process, and discovered a severe domestication bottleneck.
- ▶ Estimated linkage disequilibrium (LD). Greatly facilitated LD mapping.

Phylogeography Study of Natural Populations

- ▶ Explored a new approach to utilize the fast-evolving transposable elements based on the allele genealogical analysis on two transposons in intron region.
- ▶ Demonstrated that the analysis of DNA sequence variations in transposable elements was a powerful tool for phylogeography study at population level.

Phylogeny Reconstruction

- ▶ Sequenced intron regions of four nuclear genes in all of A-genome species of *Oryza* genus.
- ▶ Resolved the fast evolved A-genome phylogenetic relationship by intron sequences

combined with transposable element insertion/deletion events analysis.

EDUCATION

Ph.D., Genetics, 2005

Institute of Botany, the Chinese Academy of Sciences, Beijing, China

M.S., Genetics and Biotechnology, 2002

School of Life Sciences, Nanjing Forestry University, China

B.S., Biology, 1999

College of Forest Resources and Environment, Nanjing Forestry University, China

PUBLICATIONS

1. Gokcumen O, Tischler V, Tica J, **Zhu Q**, Iskow R, Lee E, Fritz M, Langdon A, Stutz A, Pavlidis P, Benes V, Mills R, Park PJ, Lee C, Korbel J. 2013. Primate genome architecture influences structural variation mechanisms and functional consequences. *Proceedings of the National Academy of Sciences USA*, 110(39): 15764-15769.
2. **Zhu Q**, Bennetzen JL, Smith SM. 2013. Isolation and diversity analysis of resistance gene homologues from switchgrass. *G3: Genes/Genomes/Genetics*, 6(3): 1031-1042.
3. Gokcumen O*, **Zhu Q***, Mulder LC, Iskow RC, Austermann C, Scharer CD, Raj T, Boss JM, Sunyaev S, Price A. 2013. Balancing selection on a regulatory region exhibiting ancient variation that predates human–Neandertal divergence. *PLoS Genetics*, 9(4): e1003404. (* co-first authors)
4. **Zhu Q**, Smith SM, Ayele M, Yang L, Jogi A, Chaluvadi SR, Bennetzen JL. 2012. High throughput Discovery of mutations in tef semi-dwarfing genes by next generation sequencing analysis. *Genetics*, 192(3): 819-829.
5. Iskow RC, Gokcumen O, Abyzov A, Malukiewicz J, **Zhu Q**, Sukumar AT, Pai AA, Mills RE, Habegger L, Cusanovich DA. 2012. Regulatory element copy number differences shape primate expression profiles. *Proceedings of the National Academy of Sciences USA*, 109(31): 12656-12661.
6. Gokcumen O, Babb PL, Iskow RC, **Zhu Q**, Shi X, Mills RE, Ionita-Laza I, Vallender EJ, Clark AG, Johnson WE. 2011. Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. *Genome biology*,

- 12(5): R52.
7. Banks JA, Nishiyama T, Hasebe M, ... **Zhu Q** ... Grigoriev IV. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science*, 332(6032): 960-963.
 8. He K, Guo A-Y, Gao G, **Zhu Q**, Liu X-C, Zhang H, Chen X, Gu X, Luo J. 2010. Computational identification of plant transcription factors and the construction of the PlantTFDB database. *Methods in Molecular Biology*, 674: 351-368.
 9. Zhang LB, **Zhu Q**, Wu ZQ, Ross-Ibarra J, Gaut BS, Ge S, Sang T. 2009. Selection on grain shattering genes and rates of rice domestication. *New Phytologist*, 184(3): 708-720.
 10. Schnable PS, Ware D, Fulton RS, ... **Zhu Q** ... Wilson RK. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326(5956): 1112-1115.
 11. Peng Z-Y, Zhang H, Liu T, Dzikiewicz KM, Li S, Wang X, Hu G, Zhu Z, Wei X, **Zhu Q**, Sun Z, Ge S, Ma L, Li L, Deng XW. 2009. Characterization of the genome expression trends in the heading-stage panicle of six rice lineages. *Genomics*, 93(2): 169-178.
 12. Guo A-Y, **Zhu Q**, Gu X, Ge S, Yang J, Luo J. 2008. Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene*, 418(1): 1-8.
 13. Guo A-Y, Chen X, Gao G, Zhang H, **Zhu Q**, Liu X-C, Zhong Y-F, Gu X, He K, Luo J. 2008. PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Research*, 36D: D966-969.
 14. **Zhu Q**, Guo A-Y, Gao G, Zhong Y-F, Xu M, Huang M, Luo J. 2007. DPTF: a database of poplar transcription factors. *Bioinformatics*, 23(10): 1307-1308.
 15. **Zhu Q**, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Molecular Biology and Evolution*, 24(3): 875-888.
 16. Shi X, Wang X, Li Z, **Zhu Q**, Yang J, Ge S, Luo J. 2007. Evidence that natural selection is the primary cause of the guanine-cytosine content variation in rice genes. *Journal of Integrative Plant Biology*, 49(9): 1393-1399.
 17. Guo A-Y, **Zhu Q**, Chen X, Luo J-C. 2007. GSDS: a gene structure display server. *Hereditas (Beijing)* (formerly *Yi Chuan*), 29(8): 1023-1026. (Simplified Chinese with English abstract)
 18. Wang X, Shi X, Li Z, **Zhu Q**, Kong L, Tang W, Ge S, Luo J. 2006. Statistical inference of chromosomal homology based on gene colinearity and applications to *Arabidopsis* and rice. *BMC bioinformatics*, 7(1): 447.
 19. Shi X, Wang X, Li Z, **Zhu Q**, Tang W, Ge S, Luo J. 2006. Nucleotide substitution pattern in rice paralogues: Implication for negative correlation between the synonymous substitution rate and codon usage bias. *Gene*, 376(2): 199-206.

20. Gao G, Zhong Y, Guo A, **Zhu Q**, Tang W, Zheng W, Gu X, Wei L, Luo J. 2006. DRTF: a database of rice transcription factors. *Bioinformatics*, 22(10): 1286-1287.
21. **Zhu Q**, Ge S. 2005. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist*, 167(1): 249-265.
22. Ge S, Guo Y-L, **Zhu Q**. 2005. Molecular phylogeny and divergence of the rice tribe Oryzeae, with special reference to the origin of the genus *Oryza*. Rice is life: scientific perspectives for the 21st century. In Toriyama K, Heong K L, Hardy B (eds). Los Banos, Philippines: International Rice Research Institute Publications, p40-44.
23. Yin T, **Zhu Q**, Huang M, Wang M. 2004. History and progress of the genomics studies in the model system of perennial plant species. *Journal of Systematics and Evolution* (formerly *Acta Phytotaxonomica Sinica*), 42(5): 464-479. (Simplified Chinese with English abstract)
24. **Zhu Q**, Pan H, Zhuge Q, Yin T, Zou H, Huang M. 2002. Analysis of genetic structure of natural populations of *Castanopsis fargesii* by RAPDs. *Journal of Integrative Plant Biology* (formerly *Acta Botanica Sinica*), 44(11): 1321-1326.

C O N F E R E N C E S

- **Zhu Q**, et al. Primate genome architecture linked with formation mechanisms and functional consequences of structural variation. **The 14th International Meeting on Human Genome Variation and Complex Genome Analysis (HGV2013)**. Seoul, South Korea. Sep 30-Oct 2, 2013 (Invited talk)
- **Zhu Q**, et al. Balancing selection on a regulatory region exhibiting ancient variation that predates Human–Neandertal divergence. **The Biology of Genomes Meeting**. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. May 7-11, 2013 (Poster)
- **Zhu Q**, et al. Balancing selection on a regulatory region exhibiting ancient variation that predates Human–Neandertal divergence. **2013 Harvard Pathology Retreat**. Boston, MA. May 3, 2013 (Poster)
- **Zhu Q**, et al. Natural selection on a regulatory region with ancient variation predating Human-Neandertal divergence. **The 2012 American Society of Human Genetics meeting**. San Francisco, CA. Nov 6-10, 2012 (Poster)
- **Zhu Q**, et al. Balancing selection on a non-coding region with regulatory function that predates Human-Neandertal divergence. **BWH Pathology Annual Research Celebration**. Boston, MA. Apr 27, 2012 (Poster)
- **Zhu Q**, et al. Improvement of orphan crop *Eragrostis tef* by Next-Gen sequencing. **2011 Plant and Animal Genome Conference**. San Diego, CA. Jan 15-17, 2011 (Invited talk)
- **Zhu Q**, et al. Isolation and diversity analysis of resistance gene. **The 4th BioEnergy Science**

- Center Retreat.** Asheville, NC. Jun 20-23, 2010 (Poster)
- **Zhu Q**, et al. Diversity in resistance gene homologues from switchgrass. **The 3rd BioEnergy Science Center Retreat.** Asheville, NC. Jun 22-24, 2009 (Poster)
 - **Zhu Q**, et al. Population structure of resistance gene homologues (RGHs) from switchgrass germplasm. **The 51st Maize Genetics Conference.** St. Charles, IL. Mar 12-15, 2009 (Poster)
 - **Zhu Q**, et al. Isolation and characterization of resistance gene homologues (RGHs) from switchgrass germplasm. **The 2nd BioEnergy Science Center Retreat.** Chattanooga, TN, Dec 1-3, 2008 (Poster)
 - **Zhu Q**, et al. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives. **Plant Genome Horizons - Vistas and Visions Conference.** Royal Botanic Gardens, Kew. London, UK. Apr 16-17, 2007 (Poster)
 - **Zhu Q** and Ge S. Molecular phylogeny of the A-genome species of *Oryza*, with a reference to origin of the Asian cultivated rice. **The 2nd Annual Symposium on Life Science.** Chinese Academy of Sciences. Beijing, China. Nov 27, 2005 (Invited talk)

P R O F E S S I O N A L A C T I V I T I E S

- ▶ Primary mentor for CURE summer student from Harvard College 2013
- ▶ Reviewer: The Plant Cell. Plant Cell Report. BMC research notes. Heredity. Frontiers in Genomic Assay Technology. The Harvard undergraduate research journal.
- ▶ Member:

ASHG (The American Society of Human Genetics)	2012-2013
CCMC (Cancer Cytogenomics Microarray Consortium)	2012-2013
Harvard School of Public Health PQG (Program in Quantitative Genomics)	2011-2013

A C A D E M I C A W A R D S

- ▶ 2013 Human Genome Variation Conference Scholarship
- ▶ 2013 Travel award from the 14th International Meeting on Human Genome Variation and Complex Genome Analysis
- ▶ 2006-2007 First Grade Research Fellowship of China
- ▶ 2005 President Excellent Award for Graduate Students of Chinese Academy of Sciences (Awarded to top 1% of graduating class)
- ▶ 2005 Di-Ao Fellowship Award of Chinese Academy of Sciences (Awarded to top 1% of graduating class)