**Zhongyun (Julie) Huang**

**Email:zhongyunbell@gmail.com  Phone: (617)-513-8852  LinkedIn:** [**linkedin.com/in/zhongyunhuang**](https://www.linkedin.com/in/zhongyunhuang) **GitHub: github.com/Zhongyun-Huang**

**SKILLS**

**Data Science:** Linux/Unix, Python (pandas), R, Perl, bash/shell, awk, SQL, machine learning, linear mixed-effect model, cloud computing, algorithm design, database, mathematical statistics, RESTful API, Approximate Bayesian computation, Flask

**Bioinformatics:** NGS, GBS, sequence assembly and alignment tools (BWA, Bowtie, BLAST, CLC workbench), SNP calling, polymorphism analysis tools (VCFtools, BCFtools), RNA-seq, BaseSpace, Galaxy, Tuxedo Suites, Agilent SureSelect, target capture bait design, phylogenetic analysis (PHYLIP, RAxML), population genetics tools (STRUCTURE, PCA, Tajima’s D, pi, theta), Gene Ontology analysis, selection scans (BayeScan, Lositan)

**EXPERIENCE**

**Insight Data Science Fellow,** Boston, MA Jan 2017 - present

* Built CRCmicrobe, a tool to predict patient’s risk of colorectal cancer based on gut microbiome sequencing results
* Trained and validated various machine learning methods, including logistics regression, SVC, SVM, random forest and nearest neighbors classification
* Developed an interactive front end using Flask

**Research & Development Summer Intern, Dow AgroSciences,** Indianapolis, IN 05/2016-08/2016

* Classified and predicted pedigree information based on genomic variation data utilizing and unsupervised machine learning approach
* Developed algorithm and constructed pipeline for pedigree quality improvement in collaboration with genomis and breeding research team
* Reduced research cost by improving breeding efficiency with NGS data manipulation
* Filed an internal invention disclosure with project product

**Graduate Researcher, University of Massachusetts Amherst,** MA 09/2011-12/2017

* Implemented model-based clustering method for inferring population structure using NGS data
* Revealed the genetic origins of a study system by phylogeny construction and Principal Component Analysis
* Scanned for signature of selection by FST analysis to explore genes enhancing plant physiology performance
* Surveyed candidate genes underlying important agricultural traits by performing Agilent SureSelect sequence capture technique, developed in-house pipelines in python and perl to analyze polymorphisms in SureSelect results, and discovered multiple events of trait acquisition through functional genetic polymorphisms
* Validated predicted candidate genes by molecular biology methods
* Constructed RNA-seq library to detect differential gene expression patterns between weedy rice and their crop ancestors
* Analyzed RNA-seq data with Tuxedo method
* Drew statistical inferences by analyzing multivariate physiology data with linear mixed-effect model
* De novo assembled the Illumine sequences of 15 fungus genomes, and called SNPs and DIPs by CLC workbench to account for the pathogenicity of the fungal pathogen

**EDUCATION**

**Ph.D in Population Genomics**, University of Massachusetts09/2011-05/2017(Expected)

**BS. In Biology**, Fudan University, Shanghai, China 09/2007-06/2011