

Shiya Song

COMPUTATIONAL BIOLOGIST · GENOMICS DATA SCIENTIST · BIOINFORMATICS PH.D

☎ 734-545-0480 | ✉ song.shiya@gmail.com | 🌐 songsy | in Shiya Song | 🐦 @ShiyaSong

Summary

I'm a passionate and experienced senior genomics data scientist specialized in algorithmic development for DNA related product. I develop statistical algorithms to interpret millions of genotypes and pedigrees and provide new genomic discoveries for consumers. I'm skilled in population & statistical genetics, next generation sequencing analysis, statistical inference, machine learning and deep learning and have successfully applied my skills in a cross-functional industry team to help deliver consumer value.

Education

University of Michigan

Ann Arbor, MI

PH.D IN BIOINFORMATICS & MASTER IN STATISTICS

2012 - 2016

- Dissertation title: Demographic and population separation history inference based on whole genome sequences
- Completion of deeplearning.ai **Deep Learning Specialization**: Completion certificate

Skills

Languages/Computing

Python, R, C++, Bash, Perl; AWS, SLURM/HPC, Hadoop/EMR, Spark; Scikit-learn, Keras; Snakemake

Statistical analysis

statistical inference, hidden Markov models, machine learning, deep learning, network analysis

NGS analysis

read alignment, variant/CNV calling, denovo assembly, statistical phasing, sequence alignment, RNA-seq analysis

Experience

AncestryDNA

San Francisco, CA

GENOMICS DATA SCIENTIST (SENIOR STAFF)

Aug. 2016 - PRESENT

- Leading a team of 5 scientists on ancestry inference algorithm and reference panel development
- Develop novel methods on pedigree reconstruction and controlling fine-scale population structure in GWAS studies
- Built and launched a random forests model to assign confidence on ethnicity estimates to improve consumer experience; designed a Convolutional Neural Network based method to infer ethnicity
- Improved array design and performed marker validation
- Provide ongoing scientific support on DNA origin story, DNA circle and Thrulines product. Wrote two white papers, co-authored one journal publication and file 4 non-provisional patents as lead inventor.

University of Michigan

Ann Arbor, MI

GRADUATE STUDENT RESEARCH ASSISTANT

2012 - 2016

- Developed a computational pipeline to analyze and reconstruct global haplotypes with high accuracy using next generation sequencing data from pools of fosmid clones as part of 1000 Genomes Project Phase III
- Inferred population split times and migration rates using reconstructed haplotypes from different populations by combining Pairwise Sequential Markovian Coalescent model with Approximate Bayesian Computation
- Analyzed over 100 contemporary village/breed dogs and ancient dog whole genome sequences using population genetics methods (PCA, ADMIXTURE, f3/f4, G-PhoCS) to understand the evolutionary and demographic history of dogs since the primary wolf divergence.

Publications

Link to Google Scholar

- Laura R. Botigué*, **Shiya Song***, Amelie Scheu*, Shyamalika Gopalan, Amanda L. Pendleton, Matthew Oetjens, Angela Taravella, Timo Seregély, Andrea Zeeb-Lanz, Rose-Marie Arbogast, Dean Bobo, Kevin Daly, Martina Unterländer, Joachim Burger, Jeffrey M. Kidd, Krishna R. Veeramah. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature Communications*, 8. (*authors contributed equally)
- Eunjung Han, Peter Carbonetto, Ross E. Curtis, Yong Wang, Julie M. Granka, Jake Byrnes, Keith Noto, Amir R. Kermany, Natalie M. Myres, Mathew J. Barber, Kristin Rand, **Shiya Song**, Theodore Roman, Erin Battat, Kenneth G. Chahine, Catherine A. Ball. (2017). Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nature Communications*, 8, 14238.
- **Shiya Song**, Elzbieta Sliwerskas, Sarah Emery, Jeffrey M. Kidd. Modeling human population separation history using physically phased genomes. *Genetics* (2016): genetics-116
- 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68-74. (Contribute Supplemental 6.3 Section)
- Kimberly F. McManus*, Joanna L. Kelley*, **Shiya Song***, Krishna Veeramah, August E. Woerner, Laurie S. Stevison, Oliver A. Ryder, Great Ape Genome Diversity Consortium, Jeffrey M. Kidd, Jeff Wall, Carlos D. Bustamante, and Michael Hammer. (2015). Inference of Gorilla Demographic and Selective History from Whole-Genome Sequence Data. *Molecular biology and evolution*, 32(3), 600-612. (*authors contributed equally)