

Shiya Song

COMPUTATIONAL BIOLOGIST · GENOMICS DATA SCIENTIST · BIOINFORMATICS PH.D

☎ 734-545-0480 | ✉ song.shiya@gmail.com | 🏠 songsy.github.io/ | 🌐 [songsy](https://songsy.com) | in [Shiya Song](#) | 🐦 [@ShiyaSong](#)

Summary

I'm a passionate and experienced computational biologist specialized in biomarker analysis for clinical trial development, biomarker and clinical data standardization, harmonization, ML/AI algorithm development for prediction of various clinical endpoints and interactive app development for analysis and visualization. I'm skilled in bioinformatics, statistical analysis, immuno-oncology, machine learning and deep learning and have successfully applied my skills in a cross-functional industry team.

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; SQL; AWS, SLURM/HPC, Hadoop/EMR, Spark; Scikit-learn, Keras; Snakemake
Statistical analysis	statistical inference, linear mixed models, hidden Markov models, machine learning, deep learning, network analysis
NGS analysis	read alignment, variant/CNV calling, denovo assembly, statistical phasing, RNA-seq analysis, dbSNP

Experience

Amgen Inc.

South San Francisco, CA

SENIOR SCIENTIST, DEPT OF CLINICAL BIOMARKER AND DIAGNOSTIC IN PRECISION MEDICINE

Apr. 2020 - PRESENT

- Lead a team of 3 scientists (serve as people manager and individual contributor) and 2 functional service providers to build biomarker platform which includes 1) building automatic data ingestion pipeline to standardize and harmonize biomarker and clinical data, 2) building automatic analysis pipeline to find associated analytes regards to various clinical endpoints and perform predictive analytics/modeling 3) building an interactive web application for analysis and visualization
- Perform biomarker analysis in support of immuno-oncology clinical programs, especially BiTE pipelines.

AncestryDNA

San Francisco, CA

GENOMICS DATA SCIENTIST (SENIOR STAFF)

Aug. 2016 - Apr. 2020

- Project lead on ancestry inference algorithm (hidden Markov models) and reference panel development
- Develop novel statistical methods on pedigree reconstruction (composite likelihood)
- 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
- Provide scientific support on DNA origin story, DNA circle and Thrulines product. Contributed three white papers, co-authored two journal publications, presented at local conferences and filed 6 non-provisional patents as lead inventor (list found in Google Scholar).

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

- Wang, Yong, **Shiya Song**, Joshua G. Schraiber, Alisa Sedghifar, Jake K. Byrnes, David A. Turissini, Eurie L. Hong, Catherine A. Ball, and Keith Noto. "Ancestry inference using reference labeled clusters of haplotypes." *BMC bioinformatics* 22, no. 1 (2021): 1-14.
- Hateley, Shannon, Angelica Lopez-Izquierdo, Chuanchau J. Jou, Scott Cho, Joshua G. Schraiber, **Shiya Song**, Colin T. Maguire et al. "The history and geographic distribution of a KCNQ1 atrial fibrillation risk allele." *Nature communications* 12, no. 1 (2021): 1-10.
- Yu, He, **Shiya Song**, Jiazi Liu, Sheng Li, Lu Zhang, Dajun Wang, and Shu-Jin Luo. Effects of the Qinghai-Tibet Railway on the Landscape Genetics of the Endangered Przewalski's Gazelle (*Procapra przewalskii*). *Scientific reports* 7, no. 1 (2017): 17983.
- 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)