

~ Jade Yu Cheng ~

8500 Waterwillow Pl · Orlando, FL 32827, USA

US Mobile: +1-808-206-5059 · Denmark Mobile: +45-9398-1013 · Email: jade.cheng@cs.au.dk · Website: www.jade-cheng.com

Appointments	Current	Data scientist.	Strong Analytics LLC, a DS consulting and ML development firm, USA.	
	2017	Postdoc, Data scientist.	Statistics and Genetics, University of California, Berkeley, USA	
		Postdoc, Data scientist.	Natural History Museum of Denmark, University of Copenhagen, Denmark	
Education	2016	Ph.D.	Computer Science, Bioinformatics; Aarhus University, Denmark	
	2011	M.S.	Information & Computer Science; University of Hawai'i at Mānoa, USA	
		M.S. (dual)	Bioengineering; University of Hawai'i at Mānoa, USA	
	2006	B.S.	Biotechnology; Nankai University, China	
Work Experience	Data scientist · Strong Analytics LLC			
	Current	Design, implement, and deploy machine learning methods and algorithms, construct data science pipelines to power smarter products and automated operations. Topics include text processing, classifiers, fuzzy search, fraud detection, supervised and unsupervised learning, etc.		
	Data scientist, Postdoc · University of California, Berkeley & Natural History Museum of Denmark			
	2015 – 2017	Under the supervision of <i>Prof. Rasmus Nielsen</i> , developed a suite of statistical methods and software to infer population structure, build evolutionary trees, detect selection outliers, etc. Techniques include mixture analysis, model-based classification, quadratic programming, etc.		
	Ph.D. fellow · Aarhus University · Computer Science · Bioinformatics Research Centre			
	2013 – 2016	Under the supervision of <i>Prof. Thomas Mailund</i> , investigated demographic inference with coalescent hidden Markov models, developed numerical optimizations, admixture modeling, etc. Techniques include model inference, hidden Markov model, heuristic-based optimization, etc.		
	Ph.D. fellow · Aarhus University · Centre for Biocultural History			
	2013 – 2014	Contributed as a core member of the Hvorkommerdufra project, a large-scale study of the Danish population's genetic history. Technologies include 23andme API, Plink, Google visualization and maps API, OAuth, JSON, SVG, HTML, CSS, JavaScript, etc.		
	Software developer · Hawai'i Water Resources Research Center			
	2011 – 2012	Created a content management database and a web application that allow authors, editors, and administrators to submit, edit, and revise conference abstracts and manuscripts.		
Software developer, Database administrator · Jab-Ex Corporation (Volunteer)				
2011 – 2014	Designed, developed, and maintained web applications and databases; created a content management system for a professional musician.			
Software developer · Texas Health, Arlington Memorial Hospital, Cardiac Rehabilitation (Volunteer)				
2011 – 2012	Designed and developed an experimental audio system that allows RNs to control and monitor educational audio tracks that patients hear as they exercise.			
Research Experience	M.S. research · University of Hawai'i at Mānoa · Department of Information and Computer Science			
	2009 – 2011	Under the supervision of <i>Prof. Edoardo Biagioni</i> , investigated the OLPC Acoustic tape measure algorithm on the Lego NXT; implementation allowed two peers to measure physical distance using only audio pulses and stock components. Technologies include OLPC, Sugar, MAX232CPE, Lego NXT, NXT-G, LeJOS, Java, Bluetooth, etc.		
	M.S. research · University of Hawai'i at Mānoa · Department of Bioengineering			
	2007 – 2008	Under the supervision of <i>Prof. Marla Berry</i> and <i>Prof. Peter Hoffman</i> , constructed and analyzed expression vectors and target proteins with mutations from selenocysteine to cystin.		
B.S. capstone · Nankai University · Department of Biotechnology · Genetics				
2004 – 2006	Under the supervision of <i>Prof. Ruiyang Chen</i> , studied the systematics and phylogeny of Pinellia; classified species of Pinellia; and determined the rank of a newly found species in this family.			
Teaching Experience	Teaching assistant · Aarhus University		Teaching assistant · University of Hawai'i at Mānoa	
	2016	Q4 Introduction to Data Science	2011	Spring Algorithms
	2015	Q1 Introduction to Bioinformatics (R)	2010	Spring Introduction to Computer Science II (Java)
	2014	Q1 Databases (SQL)	2010	Fall Algorithms
	2013	Q4 Algorithms and Data Structures II	2009	Fall Discrete Mathematics for Computer Science II
	2013	Q3 Algorithms and Data Structures I	2009	Fall Machine-Level and Systems Programming
	2013	Q2 Applied Programming (Python)	2009	Spring Algorithms
	Adjunct professor · Hawai'i Pacific University		2009	Spring Programming Language Theory (Assembly)
	2012	Spring Computer Science II (Java)	2008	Fall Introduction to Computer Science II



~ Jade Yu Cheng ~

8500 Waterwillow Pl · Orlando, FL 32827, USA

US Mobile: +1-808-206-5059 · Denmark Mobile: +45-9398-1013 · Email: jade.cheng@cs.au.dk · Website: www.jade-cheng.com

Selected Publications

Cheng, Jade Yu, Mailund, Thomas, and Nielsen, Rasmus. "Fast admixture analysis and population tree estimation for SNP and NGS data." *Bioinformatics* (2017), p.btx098.

Cheng, Jade Yu, and Mailund, Thomas. "Ancestral population genomics using coalescence hidden Markov models and heuristic optimisation algorithms." *Computational biology and chemistry* 57 (2015): 80-92.

Ilardo, Melissa A., Ida Moltke, Thorfinn S. Korneliussen, **Jade Cheng**, et al. "Physiological and Genetic Adaptations to Diving in Sea Nomads." *Cell* 173, no. 3 (2018): 569-580.

Alves, Joel M, Carneiro Miguel, **Cheng, Jade Yu**, ... Jiggins, Francis M. "Parallel adaptation of rabbit populations to myxoma virus." 2018 Submitted

Athanasiadis, Georgios, **Jade Y. Cheng**, Bjarni J. Vilhjálmsson, Frank G. Jørgensen, Thomas D. Als, Stephanie Le Hellard, Thomas Espeseth et al. "Nationwide genomic study in Denmark reveals remarkable population homogeneity." *Genetics* (2016): genetics-116.

Malaspinas, Anna, Michael C. Westaway, Craig Muller, Vitor C. Sousa, Oscar Lao, Isabel Alves, Anders Bergström Georgios Athanasiadis, **Jade Y Cheng**, et al. "A genomic history of Aboriginal Australia." *Nature* (2016).

Lan, Tianying, **Jade Y. Cheng**, Aakrosh Ratan, ... Thomas Mailund, and Charlotte Lindqvist. "Genome-wide evidence for a hybrid origin of modern polar bears." *bioRxiv* (2016): 047498.

Athanasiadis, Georgios, Frank G. Jørgensen, **Jade Y. Cheng**, Peter C. Kjærgaard, Mikkel H. Schierup, and Thomas Mailund. "Spitting for science: Danish high school students commit to a large-scale self-reported genetic study." *PLoS One* 11, no. 8 (2016): e0161822.

Cheng, Jade Yu. "Learning with Admixture: Modeling, Optimization, and Applications in Population Genetics." *forkningsdatabasen Denmark* (2016).

Cheng, Jade Yu. "Demographic Inference with Coalescent Hidden Markov Model and More." *cs.au.dk* (2014).

Selected Software

nemtree : interactive visualization for phylogenetic trees www.jade-cheng.com/trees

nemepos : interactive visualization for partially ordered sets www.jade-cheng.com/graphs/nemepos

'Ohana: a suite of software to infer population structure, admixture, and selection www.jade-cheng.com/ohana

AdmxCoalHMM: coalescence hidden Markov modeling and optimizations www.github.com/jade-cheng/jocx

Dissemination

2017 Invited, School of Mathematics and Statistics, Victoria University of Wellington, New Zealand

2017 Society for Molecular Biology & Evolution, Austin, USA

2015 Probabilistic Modeling in Genomics, Cold Spring Harbor Laboratory New York, USA

2015 Visiting, Center for Theoretical Evolutionary Genomics, UC Berkeley, USA

2015 Asia Pacific Bioinformatics, HsinChu Taiwan

2014 Society for Bioinformatics in Northern Europe, Oslo, Norway

2014 Big Data, MADALGO, Aarhus, Denmark

Technologies

Languages Python, Java, C++11, JavaScript, TypeScript, PHP, SQL, C, C#, LaTeX, R, x86 Assembly, BAT, BASH, make, LISP, Mathematica, NXT-G, HTML, CSS, XML, SVG, JSON

Libraries Pandas, Spark, Spark SQL, Spark ML, UDF, pySpark, Flask, D3js, numpy, scipy, matplotlib, nltk, scikit-learn, pyVCF, ape, JRE, JDBC, Swing, BLAS, LAPACK, jQuery, jQuery UI, phpMyAdmin, GD, 23andme API, Google maps and visualization API, OAuth 2.0, Java Applets, Highcharts JS, LeJOS, XStream, POSIX, BSD Sockets, Pthreads, Mono, .NET 4.0

Database PostgreSQL, Orator ORM, DataGrip, MySQL, MySQL Workbench, IBM DB2, MS SQL, UCSC Genome Browser, Server and Report Builder

ML topics Regression, Classification, Clusterization, Decision tree, SVM, Naive Bayes, kNN, K-means, PCA, Dimensionality reduction, Gradient boosting, Outlier detection, Supervised and unsupervised learning, NLP, OCR, Numerical optimization, Hidden Markov model, etc.

Other REST API, Docker, Amazon AWS, S3, EC2, Parquet, HPC cluster, SLURM, Git, Linux, GNOME, PyCharm, Eclipse, Inkscape, Google Docs, Atom, Lint4J, Shell, Galaxy, Nachos, VirtualBox, UNIX Networking, Wireshark, Visual Studio 2005-2008, PIC16F873A, MPLAB, Click Modular Router, NetBeans, OLPC, Lego NXT

Personal

From: Chengdu, China Languages: Mandarin Chinese and English Citizenship: USA

References

Rasmus Nielsen Statistics & Genetics, UC Berkeley & U of Copenhagen rasmus_nielsen@berkeley.edu

Thomas Mailund CS & Bioinformatics, Aarhus University mailund@birc.au.dk

Edoardo Biagioni Information & CS, University of Hawai'i at Mānoa esb@ics.hawaii.edu

Ralph Freese Math, University of Hawai'i at Mānoa ralph@hawaii.edu

