

Amy L. Williams

EDUCATION	Massachusetts Institute of Technology , Cambridge, Massachusetts, USA Ph.D., Computer Science February 2010 <ul style="list-style-type: none">• Advisors: Professors David K. Gifford and David E. Housman S.M., Electrical Engineering and Computer Science June 2005 <ul style="list-style-type: none">• Advisor: Professor Michael D. Ernst University of Utah , Salt Lake City, Utah, USA B.S., Computer Science (with Honors) May 2003 <ul style="list-style-type: none">• Advisor: Professor Wilson Hsieh B.S., Mathematics May 2003
AWARDS AND FELLOWSHIPS	Alfred P. Sloan Research Fellowship , 2015 ASHG Trainee Research Semifinalist Award , 2011 Ruth L. Kirschstein National Research Service Award (NRSA) , 2010 NIH post-doctoral fellowship (2010 – 2013) NSF Graduate Research Fellowship , 2003 Pre-doctoral fellowship (2003 – 2006) Dean's List , University of Utah, 1998 – 2002 C.M. Collins Engineering Scholarship , University of Utah, 2002 Early undergraduate admission following high school junior year, 1998
PROFESSIONAL EXPERIENCE	23andMe , Sunnyvale, California, USA <i>Senior Scientist II</i> January 2022 to November 2024 Population Genetics Research & Development Cornell University , Ithaca, New York, USA <i>Associate Professor</i> July 2020 to December 2022 <i>Nancy and Peter Meinig Family Investigator in Life Sciences</i> <i>Assistant Professor</i> August 2014 to June 2020 Columbia University , New York, New York, USA, <i>and</i> Howard Hughes Medical Institute , Chicago, Illinois, USA <i>Work conducted at Columbia University</i> <i>Postdoctoral Research Fellow</i> August 2013 to July 2014 <ul style="list-style-type: none">• Advisor: Professor Molly Przeworski Harvard Medical School , Boston, Massachusetts, USA <i>Postdoctoral Research Fellow</i> October 2009 to July 2013 <ul style="list-style-type: none">• Advisors: Professors David Reich and David Altshuler

Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Graduate Student and Research Assistant January 2004 to October 2009

- Advisors: Professors David K. Gifford and David E. Housman

Teaching Assistant January 2005 to May 2005

- 6.170, Spring Semester 2005: Laboratory in Software Engineering
– Undergraduate course in software engineering principles

University of Utah, Salt Lake City, Utah, USA

Undergraduate Research Assistant May 2002 to May 2003

- Advisors: Professors Wilson Hsieh and Peter Shirley

PREPRINTS

Qiao Y, Jewett EM, McManus KF, Freyman WA, Curran JE, Williams-Blangero S, Blangero J, The 23andMe Research Team, **Williams AL**[#]. *Reconstructing parent genomes using siblings and other relatives. (bioRxiv preprint)*
doi:10.1101/2024.05.10.593578.

Williams CM[#], O'Connell J, Freyman WA, 23andMe Research Team, Gignoux CR, Ramachandran S[#], **Williams AL**[#]. *Phasing millions of samples achieves near perfect accuracy, enabling parent-of-origin classification of variants. (bioRxiv preprint)*
doi:10.1101/2024.05.06.592816.

Avadhanam S, **Williams AL**[#]. *Phase-free local ancestry inference mitigates the impact of switch errors on phase-based methods. (bioRxiv preprint)*
doi:10.1101/2023.12.02.569669.

PUBLICATIONS

[#]Corresponding author *Co-first author Underlined: student/postdoctoral advisee

Massarat AR, Lamkin M, Reeve C, **Williams AL**, D'Antonio M, Gymrek M. *Haptools: a toolkit for admixture and haplotype analysis. Bioinformatics 39(3):btad104*, Mar. 2023.

Avadhanam S, **Williams AL**[#]. *Simultaneous inference of parental admixture proportions and admixture times from unphased local ancestry calls. Am J Hum Genet 109(1):1405-1420*, Aug. 2022.

Smith J^{*}, Qiao Y^{*,#}, **Williams AL**[#]. *Evaluating the utility of identity-by-descent segment numbers for relatedness inference via information theory and classification. G3 Genes|Genomes|Genetics*, Mar. 2022.

Chan AW, Villwock SS, **Williams AL**, Jannink JL. *Sexual dimorphism and the effect of wild introgressions on recombination in Manihot esculenta. G3 Genes|Genomes|Genetics 12(1):jkab372*, Nov. 2021.

Kivisild T, Saag L, Hui R, Biagini SA, Pankratov V, D'Atanasio E, Pagani L, Saag L, Rootsi S, Mägi R, Metspalu E, Valk H, Malve M, Irdt K, Reisberg T, Solnik A, Scheib CL, Seidman DN, **Williams AL**, Tambets K, Metspalu M. *Patterns of genetic connectedness between modern and medieval Estonian genomes reveal the origins of a major ancestry component of the Finnish population. Am J Hum Genet 108(9):1792-1806*, Sep. 2021.

Qiao Y*, [Santerud J*](#), [Basu-Roy S](#), [Hayward C](#), **Williams AL[#]**. *Distinguishing pedigree relationships using multi-way identity by descent sharing and sex-specific genetic maps*. **Am J Hum Genet** **108(1):68-83**, Jan. 2021.

[Hubisz MJ](#), **Williams AL**, [Siepel, A](#). *Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph*. **PLOS Genet** **16(8):e1008895**, Aug 2020.

[Seidman DN](#), [Shenoy SA](#), [Kim M](#), [Babu R](#), [Dyer TD](#), [Lehman DM](#), [Curran JE](#), [Duggirala R](#), [Blangero J](#), **Williams AL[#]**. *Rapid, phase-free detection of long identity-by-descent segments enables effective relationship classification*. **Am J Hum Genet** **106(4):453-466**, Apr. 2020.

[Caballero M](#), [Seidman DN](#), [Qiao Y](#), [Santerud J](#), [Dyer TD](#), [Lehman DM](#), [Curran JE](#), [Duggirala R](#), [Blangero J](#), [Carmi S](#), **Williams AL[#]**. *Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives*. **PLOS Genet** **15(12):e1007979**, Dec 2019.

[Chan AW](#), **Williams AL**, [Jannink JL](#). *A statistical framework for detecting mislabeled and contaminated samples using shallow-depth sequence data*. **BMC Bioinformatics** **19:478**, Dec 2018.

[Ramstetter MD](#), [Dyer TD](#), [Lehman DM](#), [Curran JE](#), [Duggirala R](#), [Blangero J](#), [Mezey JG](#), **Williams AL[#]**. *Inferring identical by descent sharing of sample ancestors promotes high resolution relative detection*. **Am J Hum Genet** **103(1):30-44**, Jul 2018.

• [Video of invited webinar presentation from Broad Institute \[YouTube\]](#)

[Mercader JM](#), [Liao RG](#), [Bell AD](#), [Dymek Z](#), [Estrada K](#), [Tukianinen T](#), ... [18 authors] ... **Williams AL**, ... [54 authors] ... [Altshuler D](#), [Florez JC](#) on behalf of the SIGMA T2D Genetics Consortium. *A loss-of-function splice acceptor variant in IGF2 is protective for type 2 diabetes*. **Diabetes** **66(11):2903-2914**, Nov 2017.

[Ramstetter MD[#]](#), [Dyer TD](#), [Lehman DM](#), [Curran JE](#), [Duggirala R](#), [Blangero J](#), [Mezey JG](#), **Williams AL[#]**. *Benchmarking relatedness inference methods with genome-wide data from thousands of relatives*. **Genetics** **207(1):75-82**, Sep 2017.

[Palmer ND](#), [Goodarzi MO](#), [Langefeld CD](#), ... [22 authors] ... **Williams AL**, ... [21 authors] ... [Watanabe RM](#), [Wagenknecht LE](#). *Genetic Variants Associated with Quantitative Glucose Homeostasis Traits Translate to Type 2 Diabetes in Mexican Americans: The GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium*. **Diabetes** **64(5):1853-66**, May 2015.

Williams AL[#], [Genovese G](#), [Dyer T](#), et al. for the T2D-GENES Consortium. *Non-crossover gene conversions show strong GC bias and unexpected clustering in humans*. **eLife** **2015;4:e04637**, Apr 2015.

The SIGMA Type 2 Diabetes Consortium: [Estrada K](#), [Aukrust I](#), [Bjørkhaug L](#), [Burt NP](#), [Mercader JM](#), [García-Ortiz H](#), [Huerta-Chagoya A](#), [Moreno-Macías H](#), [Walford G](#), [Flannick J](#), **Williams AL[†]**, et al. *Association of a Low-Frequency Variant in HNF1A With Type 2 Diabetes in a Latino Population*. **JAMA** **311(22):2305-14**, Jun 2014.

[†]Performed sample selection for exome sequencing and contributed to association analyses

The SIGMA Type 2 Diabetes Consortium: Williams AL[‡], Jacobs SBR, Moreno-Macías H, et al. *Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico.* **Nature** **506:97-101**, Feb 2014.

‡Led the genetics analysis and wrote the genetics portions of paper.

The 1000 Genomes Project Consortium[§]. *An integrated map of genetic variation from 1,092 human genomes.* **Nature** **491:56-65**, Nov 2012.

§Developed HAPMIX extension and applied it to the Latino populations.

Williams AL[#], Patterson N, Glessner J, Hakonarson H, Reich D. *Phasing Many Thousands of Genotyped Samples.* **Am J Hum Genet**, **91:238-51**, Aug 2012.

Fejerman L, Chen GK, Eng C, Huntsman S, Hu D, **Williams A**, Pasaniuc B, John EM, Via M, Gignoux C, Ingles S, Monroe KR, Kolonel LN, Torres-Mejía G, Pérez-Stable EJ, Burchard EG, Henderson BE, Haiman CA, Ziv E. *Admixture mapping identifies a locus on 6q25 associated with breast cancer risk in US Latinas.* **Hum Mol Genet**, **21:1907-17**, Apr 2012.

Williams AL[#], Housman DE, Rinard MC, Gifford DK. *Rapid Haplotype Inference for Nuclear Families.* **Genome Biol**, **11:R108**, Oct 2010.

Williams A, Thies W, Ernst MD. *Static Deadlock Detection for Java Libraries.* **In Proceedings of the 2005 European Conference on Object-Oriented Programming (ECOOP)**. Glasgow, Scotland, UK. Jul 2005.

Williams A, Barrus S, Morley K, Shirley P. *An efficient and robust ray-box intersection algorithm.* **Journal of Graphics Tools**, **10:49-54**, Jun 2005.

GRANT
SUPPORT

NIH/NIGMS R35GM133805 (Role: **PI**)

Scalable methods for the characterization and analysis of families in large genomic datasets.

Award period:

9/1/2019 – 7/31/2024

Total amount:

\$1,606,482

Alfred P. Sloan Research Fellowship, Alfred P. Sloan Foundation (Role: **PI**)

Award period:

9/15/2015 – 9/14/2017

Total amount:

\$50,000

NIH/NIGMS R01GM102192 (Role: **Subaward PI**; PI: Adam Siepel)

Computational methods for human genomic data integration: demography, selection, and functional potential.

Award period

9/1/2014 – 8/31/2017

Total amount:

\$148,441

CONFERENCE
TALKS

“Reconstructing parent genotypes at genotyping array accuracy using siblings and other relatives”

Society for Molecular Biology and Evolution

July 2024

“Estimating relationships by combining DNA from multiple siblings”

RootsTech

March 2022

“Exploiting multi-way identity-by-descent to detect pedigree relationships, cluster relatives, and infer long-range phase” (**Invited Speaker**)

Society for Molecular Biology and Evolution July 2021

“Reconstructing parent DNA and evaluating relatives at HAPI-DNA.org”
 RootsTech February 2021

“Reconstructing Grandma’s Genome”
 RootsTech February 2020

“Using DNA from many samples to distinguish pedigree relationships of close relatives”
 Family History Technology Workshop February 2020

“Inferring identical by descent sharing of sample ancestors promotes high resolution relatedness detection”
 Probabilistic Modeling in Genomics (PROBGEN) September 2017

“Inferring the genomes of mothers and fathers using genotype data from a set of siblings”
 Family History Technology Workshop February 2017

“Pedigree reconstruction in the era of many thousands of samples”
 Probabilistic Modeling in Genomics (PROBGEN) September 2016

“Inferring local ancestry by jointly analyzing admixed samples” (Invited Speaker)
 Probabilistic Modeling in Genomics (PROBGEN) October 2015

“Inferring local ancestry by jointly analyzing admixed samples” (Invited Speaker)
 Models and Inference in Population Genetics Workshop September 2015

“Fine-scale properties of non-crossover gene conversion in humans.” (Invited Speaker)
 New York Area Population Genomics Workshop January 2015

“The fine-scale landscape of meiotic non-crossover gene conversion.”
 64th Annual Meeting of The American Society of Human Genetics October 2014

“Non-crossover gene conversions show strong allelic bias and unexpected clustering in humans”
 Biological Sequence Analysis and Probabilistic Models Workshop July 2014

“A genome-wide estimate of the meiotic gene conversion rate in humans.”
 Biology of Genomes May 2013

“Identification of a novel genome-wide significant association with type 2 diabetes risk in Mexican and Mexican Americans.”
 62nd Annual Meeting of The American Society of Human Genetics November 2012

“Phasing of Many Thousands of Genotyped Samples.”
 61nd Annual Meeting of The American Society of Human Genetics October 2011
Semifinalist for Trainee Research Award

“Static Deadlock Detection for Java Libraries.”
 European Conference on Object-Oriented Programming (ECOOP) July 2005

INVITED TALKS	<i>“Reconstructing parental genomes and cross-chromosome phasing using data from millions of people”</i>	
	– UC Berkeley	November 2024
	– Brigham Young University	November 2024
	<i>“Reconstructing parental genomes and near perfect phasing using data from millions of people”</i>	
	– Stanford University	May 2024
	– Women in Data Science (Stanford University)	May 2024
	– University of Georgia	March 2024
	<i>“Inferring haplotypes with millions of samples and estimating parent admixture proportions and times”</i>	
	– Brown University	January 2023
	<i>“Reconstructing parent genotypes using data from siblings”</i>	
	– New York Genome Center	April 2022
	<i>“New methods to classify pedigree relationships and reconstruct parent DNA”</i>	
	– University of Southern California	February 2022
	<i>“Exploiting multi-way identity-by-descent to detect pedigree relationships, cluster relatives, and infer long-range phase”</i>	
	– University of Utah	October 2021
	<i>“New methods for classifying relatives in large datasets”</i>	
	– 23andMe	July 2020
	<i>“Distinguishing pedigree relationships using multi-way identity by descent segments detected in phase-free genotypes”</i>	
– Mount Sinai	November 2019	
<i>“Reconstructing ancestor genomes using genetic data from close relatives”</i>		
– FamilySearch	April 2019	
– Utah Valley University	April 2019	
<i>“Inferring identical by descent sharing of sample ancestors promotes high resolution relative detection”</i>		
– Broad Institute of Harvard and MIT (Web-based talk)	May 2018	
<i>“Leveraging and inferring properties of ungenotyped ancestors enables high resolution relatedness detection”</i>		
– Brigham Young University	April 2018	
– 23andMe	April 2018	
– Ithaca College	February 2018	
– Brown University	November 2017	
<i>“Relatedness inference in modern samples: Leveraging multi-way signals to improve resolution”</i>		
– Weill Cornell Medicine	March 2017	

- “Towards large-scale disease and ancestry analyses in diverse samples”
 – Population Architecture using Genomics and Epidemiology (PAGE) Steering Committee Meeting November 2015
- “Haplotype inference of large datasets and applications to gene conversion and disease studies”
 – Brigham Young University March 2014
 – Cornell University February 2014
 – New York University February 2014
 – University of Chicago February 2014
- “Meiotic gene conversion in humans: rate, sex ratio, and GC bias”
 – University of Chicago June 2013
- “Analyzing gene conversion event rate and structure”
 – Texas Biomedical Research Institute April 2013
- “Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico”
 – Icahn School of Medicine at Mount Sinai November 2013
 – Texas Biomedical Research Institute April 2013
 – Harvard Medical School March 2013
 – Broad Institute Annual Retreat November 2012
- “Phasing of Many Thousands of Genotyped Samples”
 – Harvard Medical School March 2013
 – Columbia University February 2013
 – Broad Institute of Harvard and MIT February 2013

OUTREACH

Informing the broader non-scientific community about my research is a passion I spend both time and a small amount of personal money to pursue. I am especially interested in reaching the lay genetic genealogy community, an avid group of family history practitioners. To that end, I have presented at the RootsTech family history conference three times, with recorded videos from some pandemic years available on YouTube. I also built and continue to maintain the website <https://hapi-dna.org/>, which provides user-friendly web-based versions of several tools my lab developed. (Notably I built the web versions of the tools and did not ask trainees to do so.) The website also includes helpful information such as the rates at which various relatives share identity-by-descent (IBD) segments. It is linked to from the International Society of Genetic Genealogy (ISOGG) website and receives an average of roughly 50 views per day.

PROFESSIONAL SERVICE

- Ad hoc reviewer, National Institutes of Health (MRAA) Study Section 2024
 Reviewer, National Science Foundation (CISE) Panel 2019
 Ad hoc reviewer, National Institutes of Health (BMRD) Study Section 2018
 Reviewer, Intelligent Systems for Molecular Biology (ISMB)/ECCB Conference 2018
 Reviewer, National Heart Lung and Blood Institute (NHLBI) Special Emphasis Panel 2016
 Reviewer, National Human Genome Research Institute (NHGRI) Special Emphasis Panel 2015
 Reviewer, American Society of Human Genetics 65th Annual Meeting 2015
 Reviewer, Intelligent Systems for Molecular Biology (ISMB)/ECCB Conference 2015
 Ad hoc reviewer for manuscripts submitted to the following journals: 2010-Present
 ◊Guest editor

- Nature
- Nature Genetics
- eLife[◇]
- Proceedings of the National Academy of Sciences
- Nature Communications
- Genome Research
- American Journal of Human Genetics
- Molecular Biology and Evolution
- PLOS Genetics[◇]
- PLOS Computational Biology
- Bioinformatics
- Genetics
- European Journal of Human Genetics
- Annals of Human Genetics
- Diabetologia
- PLOS ONE
- Molecular Ecology Resources

UNIVERSITY SERVICE	The following service was performed at Cornell University:	
	- Director of Graduate Studies, Computational Biology	2020-2021
	- Faculty Search Committee, Computational Biology	2019-2021
	- Admissions Committee, Computational Biology and/or Computational Biology and Medicine Graduate Fields	2014-2021
	- Lead PI overseeing Computational Biology Compute Cluster	2018-2021
	- Advisor to up to 13 undergraduate students majoring in Biometry & Statistics and/or Biological Sciences	2015-Present
	- Advisor to up to 10 freshman Biological Sciences majors	2017-2019
	- Faculty Search Committee, Computational Biology	2018-2019
	- Senator, CALS Faculty Senate	2017-2018
	- Program Committee, Computational Biology and Medicine Graduate Field Tri-Institutional program of Cornell University, Weill Cornell Medical College, and Sloan-Kettering Institute	2016-2017
	- Faculty Search Committee, Biological Statistics & Computational Biology	2015-2016
	- Consulted Life Science Advisory Committee concerning computational biology on campus	Oct 2015
	- Faculty Search Committee, Molecular Biology and Genetics Dept.	2014-2015

TEACHING Computational Genetics and Genomics: BTRY 4840/6840, cross-listed as CS 4775 (Fall semester, every year from 2015 to 2020), Cornell University. Up to 57 undergraduate and graduate students. Introduction to computational algorithms for analyzing genetic and genomic data. Assignments include bi-weekly statistical and Python-based computational problem sets and a final project implementing an algorithm from the literature or a new method typically based on the student's research.

Statistical and Computational Genetics: BTRY 7200 (Spring semester, every year from 2016 to 2021), Cornell University. 10 students, primarily graduate. Weekly readings and discussions of recent literature on computational genetics or population genetics.

MENTORING **Doctoral students**
 Siddharth Avadhanam, Computational Biology, Cornell University, 2018-2023.
 Jens Sannerud, Genetics, Genomics, & Development, Cornell University, 2017-2023.
 Daniel Seidman, Computational Biology & Medicine, Cornell University, 2017-2022.

Ying Qiao, Computational Biology, Cornell University, 2017-2021.
Melissa Hubisz, Computational Biology, Cornell University, co-chair, 2014-2019.
Monica Ramstetter, Computational Biology, Cornell University, co-chair, 2015-2017.

Doctoral committees

Cole M. Williams, Computational Biology, Brown University, 2021-Present.
Travis Wrightsman, Plant Breeding, Cornell University, 2019-2024.
Runxi Shen, Computational Biology, Cornell University, 2018-2022.
Jonathan Hughes, Ecology & Evolutionary Biology, Cornell University, 2017-2022.
Afra Shafquat, Computational Biology, Cornell University, 2017-2021.
Konnor La, Computational Biology & Medicine, Cornell University, 2020.
Ariel Chan, Plant Breeding, Cornell University, 2016-2019.

Master's students

Rajesh Bollapragada, M.Eng., Computer Science, 2019.
Ramya Babu, M.S., Computer Science, 2017-2018.
Joshua Cohen, M.Eng., Computer Science, 2016.

Undergraduate researchers

Yilei Huang, B.S., Biometry & Statistics, 2020-2021.
Michael Merrill, B.S., Information Science, 2017.
Arjun Biddanda, B.S., Computer Science, 2015.

Updated November 29, 2024