

**Sarah B. Kingan, Ph.D.**  
Staff Scientist, Bioinformatics Applications, PacBio

 @drsarahdoom  
<https://github.com/skingan>  
<https://www.linkedin.com/in/sarahkingan/>

## Summary

- Organismal biologist with 10 years of experience in bioinformatics focusing on development of high quality genomics resources for pure and applied research.
- *de novo* genome assembly expert in plant, animal, and human markets
- Leader of diverse internal teams investigating *de novo* genome assembly using HiFi (CCS) data and novel library preparation protocols for low DNA input samples
- Leader and contributor of bioinformatics analysis in collaborations with key opinion leaders resulting in top-tier publications
- Represent PacBio at scientific conferences, communicating best practices and promoting our technologies

## Skills

I have strong analytical and leadership skills with experience leading internal technical teams, collaborating with external researchers in industry, academia and government, developing new computational methods, testing new PacBio software before release, and supporting users through technical documentation and marketing collateral.

- Programming: Python [proficient], Perl [expert], R [expert], Snakemake [proficient], bash [expert]
- Software development tools: Jira, Git, Confluence
- Cluster computing (HPC, Slurm, SGE), Cloud computing (AWS)
- Bioinformatics software: genome assembly, read alignment, comparative genomics, genome annotation

## Project Highlights

### **De novo Genome Assembly Application Team**

- Role: Team Lead
- Goal: develop best practices and technical recommendation for customers for *de novo* genome assembly with HiFi reads.
- Directed team including members from marketing, software, applications, and tech support to design and execute analysis of HiFi assembly methods on diverse human and non-human samples.
- Deliverables:
  - Best Practices Guide for *de novo* Assembly
  - Speaker for PacBio webinar on HiFi genome assembly July 2019

### **Low DNA Input Protocol for PacBio**

- Role: Application team leader and bioinformatics scientist
- Goal: reduce DNA input requirements for *de novo* genome assembly with PacBio; expand PacBio into insect, clinical, and DTC consumer human genomics market
- Assisted in development and testing of data processing tools for amplified libraries (PCR duplication detection, adapter trimming) and *de novo* genome assembly (FALCON)
- Developed quality metrics for measuring success of the product, generated technical training documents for customers
- Protocol documentation and support: <https://www.pacb.com/blog/now-available-low-dna-input-workflow-for-smrt-sequencing/>
- Peer Reviewed Publication: Kingan et al. (2019) A High-Quality *De novo* Genome Assembly from a Single Mosquito Using PacBio Sequencing. Genes 10(1), 62. <https://www.mdpi.com/2073-4425/10/1/62>
- Invited Talks: i5K webinar (March 6, 2019), Chan Zuckerberg Biohub (March 29, 2019), Plant and Animal Genomes Conference (Bioinformatics session, Phase Genomics Workshop, January 2019)

### **Software Development and Support: de novo Genome Assembly and Phasing**

- Co-developer of FALCON-Phase, a novel approach to extend phasing of PacBio diploid genome assemblies using HiC data.

- Conceived of method, recruited collaborator from partner company (Phase Genomics), senior author on preprint, worked with engineers to translate proof-of-concept software into production code available in bioconda.
- Additional work on FALCON and FALCON-Unzip feature development, documentation, testing, and advanced customer support and training.
- Publication: Senior author of Kronenberg et al. (2019) Extended haplotype phasing of de novo genome assemblies with FALCON-Phase. bioRxiv. doi: <https://doi.org/10.1101/327064>
- pb-assembly support documentation: <https://github.com/PacificBiosciences/pb-assembly>

### **High-Quality Reference Genome Projects**

- **In collaboration with academic and industry scientists, I have led or contributed to many genome assembly projects**
- **Technical Lead**
  - **Spotted Lanternfly**, first Sequel II genome assembly (Kingan *et al.* 2019)
  - **Single mosquito assembly**, low DNA input protocol proof of concept sample (Kingan *et al.* 2019)
  - Jamaican Lion ***Cannabis***, performed assembly and consulted on downstream work (McKernan *et al.* 2019)
  - ***Aedes aegypti* Mosquito**, co-first author on large collaboration (Matthews *et al.* 2018)
- **Contributor**
  - **Triobinning** method, ran human genome assemblies (Koren *et al.* 2018)
  - Endangered **Hawaiian Crow**, developed haplotype diversity analysis (Sutton *et al.* 2018)
  - **Waterbuffalo**, ran genome assembly, consulted on downstream work (Low *et al.* 2019)
  - **Wheat**, hexaploid genome, ran genome assembly (Zimin *et al.* 2017)
  - **Hummingbird and zebrafish**, ran genome assemblies (Korlach *et al.* 2017)
  - **Hops**, ran genome assembly, consulted on downstream work (<http://hopbase2.cgrb.oregonstate.edu/>, in review)
  - **Barberry** (Bartaula *et al.* 2019)
  - **Brahman – Angus Cattle Genomes** (Low et al. 2019)
  - **Shea Tree** (in prep)

### **Education and Employment History**

#### **Pacific Biosciences, Menlo Park, CA**

*Staff Scientist, Bioinformatics Applications*

Development, testing and application of new methods for genome assembly with internal and external (academic and industry) scientists. Application of variant detection and IsoSeq methods to novel samples.

May 2016 - present

#### **Lummei Analytics, San Mateo, CA**

*Co-Founder*

Design and automation of NGS workflows for academic collaboration on evolutionary studies of model species.

Nov 2015 – Dec 2018

#### **University of Rochester, Rochester, NY**

*Postdoctoral Research Associate*

Genomics of *Drosophila* and non-model arthropods. NGS workflows for genome assembly, genome annotation, gene expression, comparative population genomics analysis.

June 2009 – May 2016

#### **Harvard University, Cambridge, MA**

Ph.D., Organismic and Evolutionary Biology, Hartl Lab

Dissertation: Population genomics of gene drive in *Drosophila*

September 2004 – June 2009

### **Selected Talks and Webinars**

#### **Variant calling and *de novo* genome assembly with PacBio HiFi reads**

PacBio webinar, 17 July 2019

<https://www.pacb.com/videos/webinar-variant-calling-and-de-novo-genome-assembly-with-pacbio-hifi-reads/>

#### ***de novo* Genome Assembly with PacBio**

Invited Talk at Chan Zuckerberg Biohub, San Francisco, 29 March 2019

#### **High quality PacBio genomes from single insects: implications for vector research**

i5K Webinar, 6 March 2019, <https://youtu.be/2xs9MaofTzE>

### **FALCON-Phase integrates PacBio and HiC data for de novo assembly, scaffolding and phasing of diploid genomes**

Invited Talk, Bioinformatics Workshop, Plant and Animal Genomes XXVII, San Diego 15 January, 2019

### **PacBio long reads and HiC: Recipes for generating high quality de novo genome assemblies for plants and animals**

Invited Talk, Phase Genomics Workshop, Plant and Animal Genomes XXVII, San Diego, 14 January 2019

### **Challenges in de novo assembly of complex plants**

Invited Talk, CannMed, UCLA, Los Angeles 23 October 2018

### **Understanding, Curating, and Analyzing your Diploid Genome Assembly**

PacBio Asia Pacific Webinar, 31 October 2017

Workshop Talk at PacBio West Coast User Group Meeting, Stanford University, Palo Alto CA, 6 September, 2017

Workshop Talk at PacBio East Coast User Group Meeting, UMD, Baltimore, 27 June, 2017

### **Selected Publications**

McKernan KJ, Helbert Y, Kane LT, Ebling H, Zhang L, Liu B, Eaton Z, McLaughlin S, **Kingan S**, Baybayan P, Concepcion G, Jordan M, Riva A, Barbazuk W, Harkins T. (2020) Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. *bioRxiv* <https://doi.org/10.1101/2020.01.03.894428>

**Kingan SB**, Urban J, Lambert CC, Baybayan P, Childers AK, Coates BS, Scheffler B, Hackett K, Korlach J, Geib SM. (2019) A High-Quality Genome Assembly from a Single, Field-collected Spotted Lanternfly (*Lycorma delicatula*) using the PacBio Sequel II System. *Gigascience* 8:10, [giz122](https://doi.org/10.1093/gigascience/giz122). <https://academic.oup.com/gigascience/article/8/10/giz122/5586996>

Kronenberg ZK, Hall RJ, Hiendleder S, Smith TPL, Sullivan ST, Williams JL, **Kingan SB**. (2018) FALCON-Phase: Integrating PacBio and Hi-C data for phased diploid genomes. *bioRxiv*. <https://doi.org/10.1101/327064>.

**Kingan SB**, Heaton H, Cudini J, Lambert CC, Baybayan P, Galvin DB, Durban R, Korlach J, Lawniczak MKN. (2019) A High-Quality De Novo Genome Assembly from a Single Mosquito using PacBio Sequencing. *Genes* 10: 62. <https://doi.org/10.3390/genes10010062>

Low WY, Tearle R, Bickhart DM, Rosen BD, **Kingan SB**, Swale T, Thibaud-Nissen F, Murphy TD, Young R, Lefevre L, Hume DA, Collins A, Ajmone-Marsan P, Smith TPL, Williams JL. (2019) Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Com* 10:260. <https://doi.org/10.1038/s41467-018-08260-0>

McKernan K, Helbert Y, Kane LT, Ebling H, Zhang L, Liu B, Eaton Z, Sun L, Dimalanta ET, **Kingan SB**, Baybayan B, Press M, Barbazuk W, Harkins T. (2019) Cryptocurrencies and Zero Mode Wave guides: An unclouded path to a more contiguous *Cannabis sativa* genome assembly. Preprint: <https://osf.io/n98gp/>

Matthews BL, Dudchenko O, **Kingan SB**, et al. (2019) Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* 563:501. <https://doi.org/10.1038/s41586-018-0692-z>

Sutton JT, Helmkampf M, Steiner CC, Bellinger MR, Korlach J, Hall R, Baybayan P, Muehling J, Gu J, **Kingan S**, Masuda BM, Ryder OA. (2018) A High-Quality, Long-Read De Novo Genome Assembly to Aid Conservation of Hawaii's Last Remaining Crow Species. *Genes* 9:E393. <https://doi.org/10.3390/genes9080393>

Koren S, Rhie A, Walenz BP, Diltz AT, Bickhart DM, **Kingan SB**, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. (2018) De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotech.* 36:1174. doi: <https://doi.org/10.1038/nbt.4277>.

Zimin AV, Puiu D, Hall R, **Kingan S**, Clavijo BJ, Salzberg SL (2017) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* 6, <https://doi.org/10.1093/gigascience/gix097>

Korlach J, Gedman G, **Kingan SB**, Chin C-S, Howard JT, Audet J-N, Cantin L, Jarvis ED. (2017). *De novo* PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience* 6, <https://doi.org/10.1093/gigascience/gix085>