

---

## BIOGRAPHICAL SKETCH: C. Titus Brown

NAME: C. Titus Brown



---

eRA COMMONS USER NAME: BROWNCT

---

POSITION TITLE: Associate Professor, UC Davis

---

EDUCATION/TRAINING:

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Reed College, Portland, OR	BA	6/1997	Mathematics
California Institute of Technology, Pasadena, CA	PhD	6/2007	Developmental Biology and Genomics
California Institute of Technology, Pasadena, CA	-	5/2008	Developmental Biology and Genomics

### A. Personal Statement

I am focused on advancing biology through better, more flexible, and more advanced data analysis – what I call “data-intensive biology”. My primary interest is in refining hypotheses using high-throughput sequencing data, as well as integrating multiple data types, and I have largely worked in non-model systems and environmental microbiology. I have worked extensively at the interface of computation and microbiology, contributing to projects ranging from regulatory analysis of single microbial genomes, to simple and complex metagenome analysis (Goffredi et al, 2014). As a professor, my lab has contributed significantly to advances in data analysis and software development for metagenome and transcriptome assembly (Pell et al., 2012; Brown et al., 2012), successfully analyzing several extremely large soil metagenomes (Howe et al., 2014). At UC Davis, I will be running many workshops and training sessions for the purposes of advancing collaborations as well as better understanding the technical challenges that stand in the way of advancing the entire field; the great demand for computational training in these areas places me in an excellent position to identify and target these blocking challenges to drive my own research.

### B. Positions and Honors

2008-2014	Assistant Professor, Computer Science and Engineering / Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI.
2014-2019	Moore Foundation Data-Driven Discovery Investigator.
2010-2013-	Director, two-week intensive workshop in Analyzing Next-Generation Sequencing Data. Scientific Advisory Board member, iPlant Collaborative, 2013-present.
2015	Faculty, MBL Microbial Diversity course.
2012-2013	TA and Lecturer, MBL STAMPS course.
2006-2008	TA and Lecturer, MBL Embryology course.

## C. Contribution to Science

Google Scholar profile: <http://scholar.google.com/citations?user=O4rYanMAAAAJ&hl=en>

### 1. Scalable data structures and algorithms for metagenome and transcriptome assembly. (2008-present)

As a professor, my lab tackled the problem of scaling *de novo* metagenome and transcriptome assembly methods to the extremely large data sets emerging from Illumina sequencers. At the time, one of the biggest blocks to making robust use of non-model transcriptomes and metagenomes was an inability to search or assemble these data sets. We have developed novel data structures ([Pell et al., 2012](#)), implemented advanced data structures in a fast and reusable open library ([Zhang et al., 2014](#); the khmer software), built novel algorithms ([Brown et al., 2012](#)), and applied these novel data structures and algorithms to previously unresolvable data sets such as immense and diverse soil metagenomes ([Howe et al., 2014](#)). Our software is widely used (100-10,000 users), and has helped drive the field of metagenome assembly towards considerably more powerful techniques. It is now included in the Illumina TruSeq Long Read technology pipeline.

### 2. Robust and open scientific software development. (2000-present)

For 15 years, I have championed a central role for good, open, and sustainable software development practices in science. As software becomes ever more important for data analysis and modeling in science, robust, reusable, and remixable software can accelerate the practice of both computational and experimental science. As a participant in the Software Carpentry training program, as a blogger, and as a reviewer of bioinformatics software papers, I have helped develop and drive better software practice. Our paper on “Best Practices for Scientific Computing” ([Wilson et al., 2014](#)) was the most downloaded paper at PLOS Biology last year, and has received over 70 citations in under a year. I routinely engage with other software developers online and in workshops to help advance the methods, process, and culture of scientific computing in biology.

### 3. Training and teaching in data-intensive biology. (2008-present)

I believe that one of the biggest challenges facing biology today is the lack of basic computational expertise amongst early career researchers in biology. To help address this problem, I've developed over 10 different peer-instruction workshops and short courses, and taught over 40 workshops in the last 6 years. This includes my flagship two week NIH-funded summer course, “Analyzing Next-Generation Sequencing Data” (2010-present; [bioinformatics.msu.edu/ngs-summer-course-2015](http://bioinformatics.msu.edu/ngs-summer-course-2015)), as well as several short courses on sequence analysis, data management, and over a dozen Software Carpentry-like short courses. All of my course materials and presentations are posted online and available under an open Creative Commons license (e.g. see [angus.readthedocs.org/en/2014](http://angus.readthedocs.org/en/2014)), and receive well over 100,000 “hits” a year. At my new position at UC Davis, I have been specifically tasked with significantly expanding my training program in data intensive biology ([ivory.idyll.org/blog/2014-davis-and-training.html](http://ivory.idyll.org/blog/2014-davis-and-training.html)), and all of my research will be integrated with my training moving forward.

### 4. Comparative sequence analysis and Gene Regulatory Networks (1999-2008).

As a graduate student, I developed a platform for comparative sequence analysis that helped identify dozens of *cis*-regulatory elements in the sea urchin ([Brown et al., 2005](#)). This software played a significant role in developing the sea urchin endomesoderm network, one of the largest and most detailed extant developmental gene networks ([Davidson et al., 2002](#)). My intellectual contribution was to work at the intersection of biological needs and computational possibility to identify a robust approach and implement it in an easy-to-use piece of software that let dozens of biologists apply it without my help or support.

## 5. The Avida platform for studying evolution. (1993-1997)

As an undergraduate, I worked with Chris Adami and Charles Ofria to develop the Avida evolutionary platform for bottom-up evolutionary studies ([Adami and Brown, 1994](#)). This platform is now one of the most widely used platforms in computational studies of evolution, and is one of the foundations of the BEACON NSF STC for studying “evolution in action.” My intellectual contribution was in connecting the scientific needs to effective and robust computation.

## D. Research Support

1. Moore Data Driven Discovery Investigator Award (2014-2019). This five-year career award is aimed at the development of new techniques and approaches for harnessing large biological data sets. It explicitly supports the development of robust, general software for sharing both raw data and analyses, as well as field-specific software development. I am the sole PI.
2. NIH BIGDATA R01, “Low-memory Streaming Prefilter for Biological Sequencing Data” (2013-2016). This supports our software and algorithm development for sequence analysis, with the goal of enabling very scalable analysis of large sequencing data sets. I am the sole PI.
3. NIH NHGRI R25, “Analyzing Next-Generation Sequencing Data” (2013-2016). This workshop grant supports a two-week summer workshop on training in next-generation sequencing data analysis, as well as the development of associated materials. I am the sole PI.
4. NSF OCI Supplement to NSF BEACON STC, “Materials and Workshops for Cyberinfrastructure Education in Biology” (2013-2016). This supplement to the BEACON STC supports collaboration between several NSF BIO centers (including iPlant and C-MORE) to develop workshops for data-intensive biology. The Data Carpentry workshop series ([data-carpentry.org](http://data-carpentry.org)) emerged partly from this funding. I am the sole PI.
5. USDA NIFA, “Easy to use Web interfaces for sequence analysis” (2010-2014). This four year grant supported our core software development work on improving non-model sequence analysis, and led to the development of the khmer software package. I was the sole PI.

