Sequencing Impact at the University of Missouri
Prepared by Bob Schnabel, Scott Givan and Diane Oerly with Input from Researchers across Campus

This document was created to convey why it is necessary for the University of Missouri to provide additional computational capacity and disk storage to support the explosion of data that has been and will continue to be generated by enhancements to high throughput sequencing research. A sustainable funding model is also provided.

Executive Summary

It would be an understatement to say that "next-generation" sequencing technology has been revolutionary. Over the last 10 years, sequencing has created a paradigm shift in biological sciences where more and more a component of research involves "just sequence it". This is because the types of data, applications and resulting insights are expanding every year. Further, the volume and speed of data generation are growing exponentially, while the costs to generate these data are decreasing exponentially. The Human Genome Project completed the first draft genome sequence in 2001 at an estimated cost of $3 billion. Next-generation sequencing became mainstream around 2007 and enabled the re-sequencing of a human genome at a cost of approximately $50,000. In late 2015, Illumina announced the availability of their X10 sequencer for use on non-human samples enabling the re-sequencing of a mammalian (human, cow, dog etc.) genome for approximately $1,500 and with an annual throughput of 10,000 genomes per year. The ease, rapidity and cost effectiveness of generating sequence data has created a computational analysis bottleneck. The growth of computational resources on the MU campus has not kept pace with the growth in data generation capability.

In order for Mizzou to maintain a competitive research environment, we need to expand the computational resources available for bioinformatics analysis of large data which include sequence data. It will require an initial investment of $619,000 in early 2016 to build the needed core infrastructure and will require ongoing funding to maintain and expand this infrastructure.

Initial investments (cost share of $231,000) made by Mizzou in 2005 to bring next-generation sequencing to this campus have been returned many-fold. Based on a survey sent to MU researchers in November 2015, a total of 66 grants have been awarded involving sequencing for a total of $87.5M. $7.6M of that is directly attributable to sequence data generation/analysis. In addition, another $7.9M in grant funding has been submitted and remains pending. This research has led to 173 refereed journal articles in top-tier journals producing over 6,000 citations. Additionally, 19 M.S., 62 Ph.D. and 21 postdocs have been trained as a result of these sequence related research projects. Plant and animal researchers at MU have been at the forefront of the next-generation sequencing revolution. However, based on the diversity of grants and papers gathered by the survey, sequence analysis provides a common foundation that ties together many disciplines on campus. As such, investment in computational capacity directed at sequence data analysis will serve the entire campus and provide technological ties between disciplines.

The following is a detailed description of the history of sequencing/bioinformatics, a description of the computation resources required, and a model for sustainability and an analysis of the impacts of next-generation sequencing at Mizzou.

History of Sequencing at MU

The story of high throughput sequencing at MU is the story of how a small investment can produce enormous success. It started with a proposal to the National Science Foundation (NSF). Ten years ago (August, 2005) MU was awarded a Major Research Instrumentation (MRI) award to purchase Illumina instruments (both genotyping and sequencing) totaling $925,500. MU's cost share was about $231,000. The NSF was particularly impressed that the project served BOTH plant sciences and animal sciences and that the shared instrumentation would be housed in MU's DNA Core Facility where it could generate revenue from various other research projects and businesses. The DNA Core Facility staff collaborated with research support computing staff and industry to generate, analyze and distribute the massive quantities of data involved in High Throughput Sequencing (HTS) and genotyping.

This cost share with the NSF led to several significant outcomes including:

- Mark McIntosh was Principal Investigator on a grant from the State of Missouri which contributed $1.3M in salary support that enabled the creation of MU's **Informatics Research Core Facility**. This revenue generating facility provides bioinformatics data management and support for a variety of researchers. As envisioned, the Director, Scott Givan serves as co-investigator on many of the current research grants. The collaborations strengthen research proposals and facilitate high-throughput sequencing-based projects to a broad base of faculty and staff.
- The US Congress selected MU as the location of the National Center for Soybean Biotechnology (NCSB) http://www.soybiotechcenter.org/. Henry Nguyen, of MU's Interdisciplinary Plant Sciences Group who served as PI on the MRI award directs the center. See: http://soybeangenomics.missouri.edu/personnel/nguyenbio.htm MU was selected to host the center based on our record of interdisciplinary research on soybean genetics, genomics, and related sciences. The NCSB is a collaborative program among scientists at the University of Missouri, the USDA-ARS Plant Genetics Unit in Columbia, and the Donald Danforth Plant Science Center in St. Louis, MO to provide innovative molecular approaches that can be applied toward soybean improvement. The NCSB's 40+ researchers have expertise in diverse fields including agronomy, microbiology and plant pathology, chemistry, biochemistry, animal science, food science, molecular biology, engineering, computer science, and agricultural economics. Their many successful collaborations include SOYKB – see: http://engineering.missouri.edu/2015/06/soykb-a-powerful-tool-at-the-junction-of-plant-biology-and-computer-science/. Scientists working in the NCSB successfully compete for funding from a variety of other sources including the Missouri Soybean Merchandising Council (MSMC), federal agencies, and private industry. For information on their research projects, see: http://www.soybiotechcenter.org/research/.
- A historic $5.6 million gift. In Spring of 2004, MU officials announced a historic donation from one of the nation's top private livestock producers. The David W. Gust family, owners of Circle A Ranch in Iberia, MO donated detailed animal performance data on approximately 6,000 animals, including DNA samples. Jerry Taylor, Curators' Professor and Wurdack Chair in Animal Genomics, who was Co-PI on the MRI award- was able to secure this gift which was heralded as the first such gift to a public institution. See: http://cafnr.missouri.edu/alumni-files/gratitude/s2004-gratitude.pdf Animals from this

donation were among the 10,000 animals genotyped on the Illumina genotyping system from the NSF MRI award.

- Mizzou Advantage targeted investment led to the hire of researcher Chris Elsik with the knowledge and skills to support and advance plant AND animal research – http://mizzouadvantage.missouri.edu/people/christine-elsik/. Dr. Elsik and her lab collaborate with scientists from around the world performing a fundamental role in gene prediction, organization of community annotation and the analysis of model organism genomes. Her lab http://genomes.missouri.edu/elsiklab has developed and maintains the Bovine Genome Database, and the Hymenoptera Genome Database (which includes BeeBase, NasoniaBase and the Ant Genomes Portal). Chris is also highly involved in Missouri's EPSCOR project that has received $20 million in funding. Note the Research and Accomplishments information at: http://www.epscormissouri.org/

The initial Illumina sequencing instruments obtained as part of the NSF MRI grants in 2005/2008 have been replaced as new technology/instruments become available. Currently the DNA Core Facility operates two Illumina Hi-Seq and two Illumina Mi-Seq instruments. The DNA Core has seen exponential growth in its services reflective of the demand for sequencing. (**Fig. 1**). This increased demand for sequencing capacity has been due to an increasing number of funded grants that involve sequence generation.

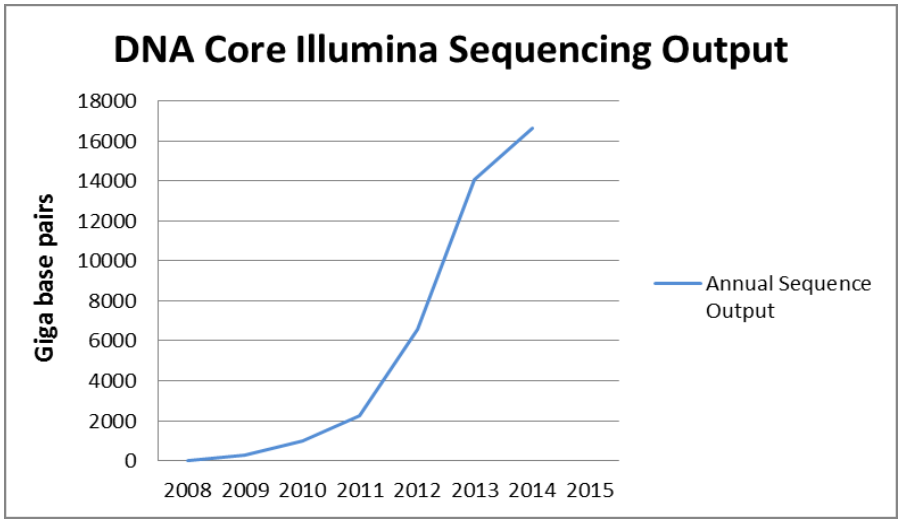

Fig. 1. Exponential growth of sequence data generation at the Mizzou DNA Core facility. http://biotech.missouri.edu/dnacore/initiateproject.html#

Impact of Sequencing Technology at Mizzou

To quantify the impact of the DNA Core Facility sequence data generation, in November, 2015, an e-mail questionnaire was sent to Mizzou researchers involved with sequence generation and analysis. The e-mail contained an Excel template for researchers to list grant information, publications and students trained. Responses to this e-mail survey are summarized below.

A total of 66 grants that involve sequence generation/analysis (**Fig. 2**) have been funded since 2001 (detailed in Appendix A). This represents $87,530,171 in total funding and $7,647,758 directly attributable to sequence data generation/analysis (**Fig. 3 A-B**). As previously noted, the DNA Core Facility received its first DNA sequencer in 2008. This enabled researchers to generate preliminary data with which to write additional grants. The result of this acquisition can be seen in **Fig. 3** where a large number of grant dollars in 2011 primarily coming from two grants to the Animal Genomics group: Integrated Program for Reducing Bovine Respiratory Disease Complex in Beef and Dairy Cattle ($10M) and National Program for the Genetic Improvement of Feed Efficiency in Beef Cattle ($5M). As detailed in Appendix A, and depicted in **Fig. 1**, the total number of grants, total dollars and dollars attributable to sequence generation have continued to increase year over year which is in part responsible for the exponential growth in sequence generation from the DNA Core Facility.

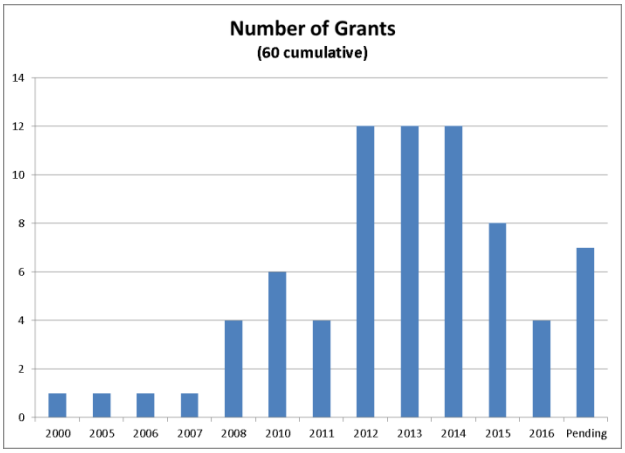

Fig. 2. Number of sequencing related grants funded.

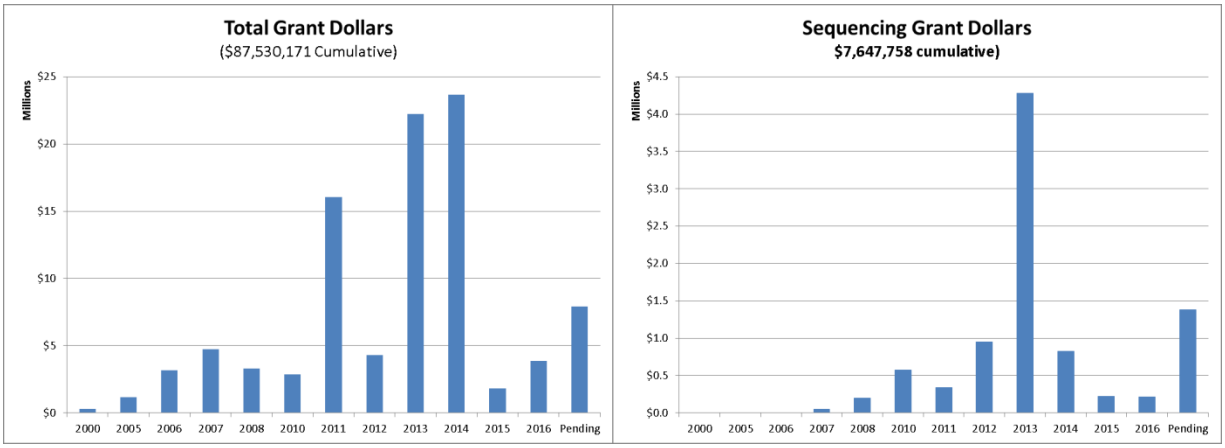

Fig. 3. Total dollars by year (A) and sequencing expenditures (B) for grants received by MU researchers.

The success of MU researchers in competing for grant dollars has also had an impact on the number and quality of peer reviewed papers published. In response to the e-mail survey, researchers provided a list of all publications related to sequencing (**Appendix C**). In order to quantify the impact of these publications, the number of citations per paper and the journal metrics were obtained from Thompson Reuters InCites Journal Citation Reports.

https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action. A total of 173 papers were submitted that have a combined 6716 citations as of 12/1/2015. As can be seen from **Fig. 4**, the increase in the number of papers coincides with the increase in grants and grant dollars obtained and also the amount of sequence data generated. It is common knowledge that there is a "lag" in citations of several years from the time a paper is published. Given the rapid rise in the number of sequencing papers in the last two years (2014-15), one can extrapolate that the citation counts of these papers in the coming years will be significant and positively impact Mizzou's AAU metrics.

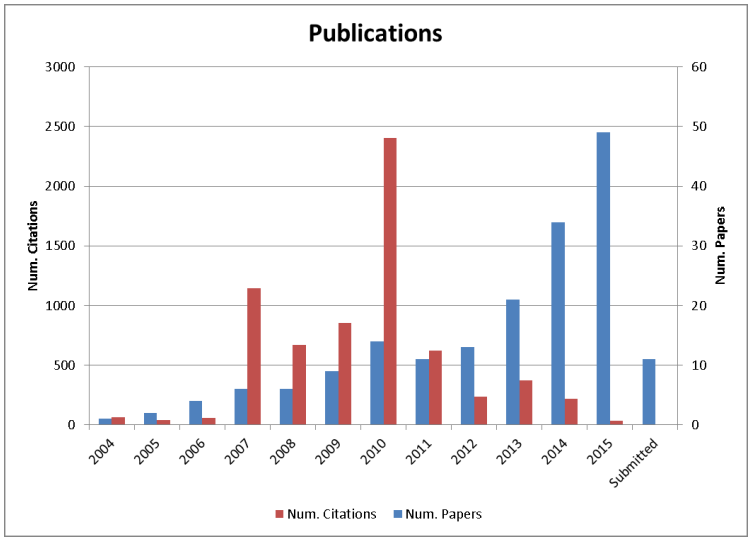

Fig. 4. Total number of sequencing related papers and the numbers of citations.

One paper of particular note is from 2008: Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries. *Nature Methods. 5:247-252.* This was the cover article and currently has 319 citations. This paper is noteworthy because it was the first paper describing the use of next-generation sequencing to directly estimate allele frequencies for the development of a genotyping assay. The assay that was developed was used to genotype over 10,000 cattle samples by the DNA Core Facility using instrumentation acquired from the initial NSF MRI award. The assay itself was marketed by Illumina and has sold over 2 million units worldwide. In 2010, this group received the USDA Secretary's Honor Award - the highest award given by the USDA. All of this was made possible by the initial investment in the Illumina technology at Mizzou.

The breadth of sequencing related activities is also reflected in the journals that Mizzou researchers have published in (**Appendix B**). The 173 papers were published in 81 journals covering 27 categories/disciplines. Of the 173 papers, 143 (77%) are from journals ranked in the first quartile (Q1) of their discipline with 6% (Q2), 6% (Q3), 4% (Q4) and 5% in journals without a ranking. Sixteen papers were published in nine journals with impact factors >10, again demonstrating the relevance and high impact of sequencing related activities.

Another measure of impact is in the training of graduate students. A total of 16 researchers from 10 departments/divisions indicated that 19 M.S., 62 Ph.D. and 21 postdocs have been trained in

bioinformatics associated with sequence data analysis (**Table 1**). This demonstrates a broad educational impact focused on the common technology of sequence data analysis.

Table 1. Number of students trained in sequence data analysis.

Mizzou PI/Co-I	Dept/Division	Number MS	Number Ph.D.	Number Postdocs
Taylor, JF	Animal Science	4	3	1
Prather, RS	Animal Science	1	2	0
Lyons, LA	Animal Science	1	0	0
Lyons, LA	CVM - VMS	2	0	0
Beerntsen, BT	Veterinary Pathobiology	0	0	1
Xu, D	Computer Science	4	11	3
King, EG	Biological Sciences	1	2	1
Joshi, T	MMI / MUII / CS /IPG	2	3	0
Wall, J	Biochemistry	0	2	2
Spencer, T	Animal Science	0	6	0
Taylor, KH	Pathology and Anatomical Sciences	2	3	0
Stacey, G	Plant Sci/Biochem	1	10	8
Elsik, CG	Animal Science	0	6	0
Johnson, GS	Veterinary Pathobiology	0	3	0
Flint-Garcia, S	USDA-ARS, Div Plant Sci	0	2	2
Pires, JC	Biological Sciences	1	9	3
	Total	19	62	21

Informatics Research Core Facility

As mentioned above, the MU Informatics Research Core Facility (IRCF) works extensively with researchers to manage, analyze and visualize high-throughput data. A significant focus of the IRCF is working with Illumina sequence data, both Hi-Seq and Mi-Seq, generated by the MU DNA Core. One measure of IRCF activities is its broad client base. Since 2011, the IRCF has worked closely with nearly 100 research groups. These research groups come from a number of MU colleges and departments, including Biology, Microbiology, Veterinary Medicine, Agriculture and Natural Resources, Chemistry, Biochemistry and the Medical School. The IRCF also has a number of external clients, including from the University of Texas, Montana State University and Virginia Tech. Another measure of the growth of IRCF is the revenue generated by these activities. As illustrated in **Fig. 5**, since 2011, which is the first year that the IRCF charged for services, revenue has grown to over $100,000 per calendar year.

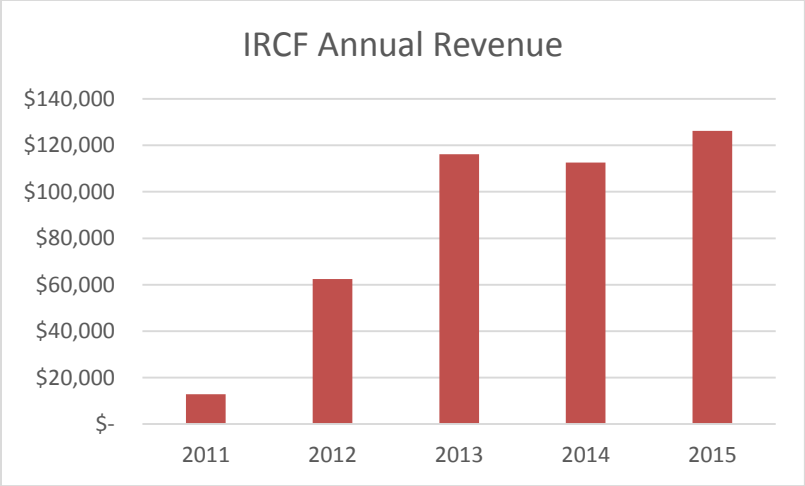

Fig. 5. Annual revenues of the Informatics Research Core Facility demonstrating a rapid increase and sustained demand for services.

Computational Demand

The success of researchers competing for grant dollars has become a double-edged sword. While we have clearly demonstrated the growth of sequencing and its academic impact at Mizzou, the computational resources necessary to continue this growth have been absent. The University maintains a high performance computing (HPC) cluster named **Lewis3** (https://umbc.rnet.missouri.edu/resources/lewis3.html). The original **Lewis** cluster was upgraded in 2015 to include 32 compute nodes with 128 GB of memory each and 8 compute nodes with 256 GB of memory each and each compute node has 24 Intel Xeon cores (960 total CPU cores). While this may seem like a large compute system it is actually quite small compared to our peer research institutions. For example, Texas A&M operates several HPC systems, one of which has 16,900 CPU cores http://sc.tamu.edu/systems/. Likewise, Iowa State University operates several HPC clusters with over 2,500 CPU cores (http://www.hpc.iastate.edu/systems).

HPC systems have traditionally been built based on the characteristics of the type of analyses to be performed. Typically these systems have a large number of compute nodes with a large number of CPU cores but low memory in the range of 12-48 GB/node. For physical sciences (e.g., physics and chemistry), which are generally characterized by extremely large numbers of "small" jobs that individually operate on a small piece of data, the traditional HPC system performs very well. In many cases, an analysis job can be distributed across a large number of nodes/cores.

Bioinformatics analysis, in particular sequence data analysis, has very different compute requirements. Sequence data analysis is characterized by very large input files, even larger intermediary files and most importantly, large memory requirements. These characteristics make sequence analysis on traditional HPC systems either extremely difficult or in most cases impossible. For example, two common analysis pipelines are *de-novo* transcriptome or genome assembly, and error correction. Both often require greater than 256 GB of memory and sometimes as much as 1 TB. The University of Missouri HPC systems generally lack compute nodes capable

of performing these types of analysis. For example, the **Lewis3** cluster only has two compute nodes capable of performing these types of analyses.

The large amount of sequence data generated by Mizzou researchers and the lack of centrally supported high performance computing capable of processing these data have led to the creation of independently acquired and managed systems to handle the sequence data. The Informatics Research Core Facility (IRCF) manages the **BioCluster**, which currently contains 9 servers with a total of 328 processors and 225 TB of storage. The Division of Animal Sciences manages the **MUgenomics** infrastructure which contains 8 servers with a total of 424 CPU cores and 160 TB of storage. While the total CPU count of these systems is smaller than the **Lewis3** system, they are different from **Lewis3** in two very important aspects. The **BioCluster** and **MUgenomics** consist primarily of high memory and high CPU core nodes, most of which have 64 cores and 512 GB memory. These two systems were specifically built to analyze sequence data because the University operated **Lewis** system was incapable of analyzing the sequence data.

The exponential increase in sequence data being generated by an ever increasing number of funded grants demands an investment in a new HPC system designed to meet the computing characteristics of these new data. In order to determine the hardware architecture requirements for a new Informatics HPC we present an analysis of a typical sequence analysis pipeline, variant calling from genomic resequencing.

Sequence Analysis Benchmark

The Broad Institute has created "best practices" guidelines for using the Genome Analysis ToolKit (GATK), which is a widely used analysis pipeline for variant calling https://www.broadinstitute.org/gatk/guide/best-practices.php. The GATK best practices pipeline was implemented on the **MUgenomics** infrastructure. The pipeline, as implemented by the Animal Genomics group, consists of 17 steps. Each of these steps has different hardware requirements in terms of CPU, disk and memory usage. In order to characterize each of these steps, a profiler program was written that captures system statistics at 5 second intervals during the entire pipeline throughput and produces a summary plot.

The server used was **MUgenomics07** with 48 cores (4 x 12 core AMD 6348 2.8 GHz) and 512 GB DDR2 1866 RAM. This server is the Postgres database server for the Animal Genomics group and was chosen because of its fast disk storage system to illustrate the I/O requirements of a sequence data pipeline. The storage on this server consists of 30 - 300 GB Seagate 15K rpm SAS2 drives (ST33000657SS) configured as a 4 TB RAID10 array on an Adaptec 78165 controller.

The data used were genomic sequence data from a domestic dog. This dog had approximately 30X of genome sequence coverage of raw data (2 x 100 bp) from the MU DNA Core's Illumina HiSeq2000 instrument. The gzip compressed input data size was 76.2 Gb. Note that the analysis presented below is part of the Animal Genomics group's standard pipeline and these figures are available for over 400 animals. This animal was chosen as a representative of the total data processed.

Fig. 6 presents the system characteristics of the pipeline. First, note that it took over 21 hours to process the data for this single animal. This pipeline has been highly optimized and tailored to the **MUgenomics** infrastructure to maximize hardware usage and minimize total run time. This is

representative of a typical 30X mammalian genome using this hardware and this pipeline. The Animal Genomics group has over 400 such animals and if restricted to a single server would require almost a year to process the data. The red peaks in **Fig. 6** are characteristic of multithreaded programs that are able to utilize all available CPU cores and thus these steps of the pipeline are CPU-bound which is nearly half (47.3%) of the total required processing time. The blue peaks in **Fig. 6** are characteristic of stages that have high volumes of disk writes where processed data need to be physically written to storage. High disk I/O (>200 MB/s) represents 33.1% of the total processing time. Finally, stages with elevated disk wait (%wa > 1.0) are illustrated in pink. The I/O wait on this server is generally not an issue due to the high throughput of the storage subsystem. Note that the I/O in blue is relative to the right Y-axis and is in MB/s. A significant portion of processing time is spent performing I/O >200 MB/s and a non-trivial amount of time performing I/O at >800 MB/s. This has profound implications for HPC infrastructure where compute nodes use shared data storage and storage bandwidth needs to be a primary design criteria.

The memory subsystem performance is not presented, although cache and dirty bytes are captured in the profiler. This is because this analysis pipeline has been optimized and requires a compute node with 512 GB memory. The large memory is used for cache and significantly speeds up the entire process.

In summary, compute infrastructure for the analysis of sequence data has drastically different resource requirements than provided by traditional HPC environments. These requirements are characterized by compute nodes with high CPU core counts (>40 but preferably 64), high memory (minimum 512 GB) and very high bandwidth to storage (>200 MB/s sustained for long periods).

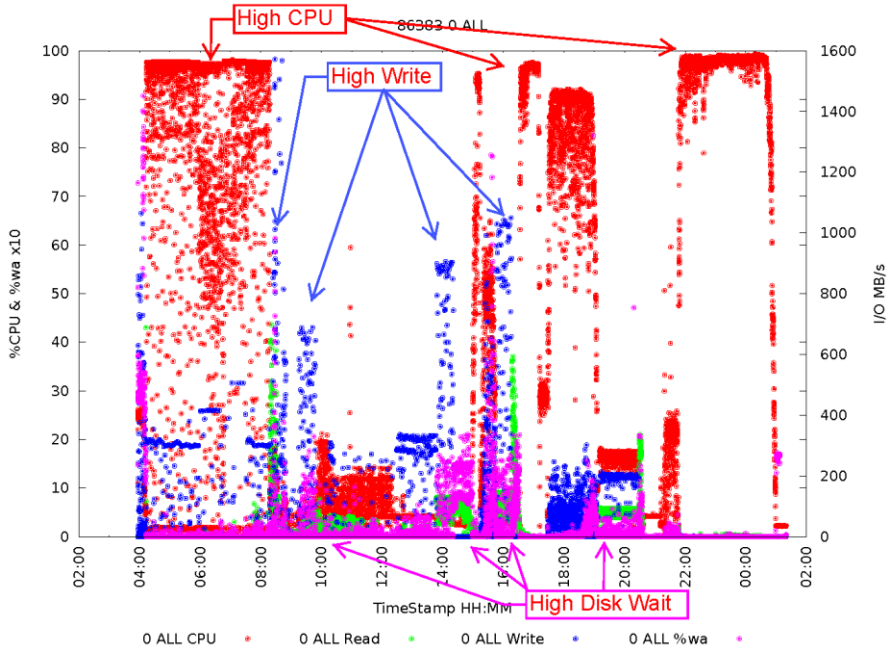

Fig. 6. System profile of the GATK best practices pipeline for variant calling from genomic data.

Key Trends Make Computational Capacity More Critical

The next generation of DNA sequencer will again increase throughputs and lower costs. The implications for Mizzou researchers are significant. A researcher can afford to over-deliver by generating 2-5 times the amount of data that were originally proposed in a sequencing budget. The decrease in cost and increase in data generation make it feasible for researchers to write grants to perform population level analyses. This too will require an increase in the computation required to store, process, and analyze these data. In the past, a lab could process and analyze the data on a small number of machines, for example the **BioCluster** or **MUgenomics** servers. The increased amount of compute power required will be beyond what a single lab can manage and requires an extensive infrastructure to take advantage of multiple nodes to process the data in parallel in order to compute the results in a *competitive time*.

In terms of genomics, a common target is a "1000 Genome Project" for a species. The first Human 1000 genome project was a massive effort http://www.1000genomes.org/participants. Similar projects have been completed for most highly-researched species. Of particular interest at Mizzou are the 1000 Bull Genomes project (Taylor, Schnabel, Decker), 1000 Soybean Genomes project (Nguyen, Xu, Joshi), Maize Genome (Flint-Garcia), 1000 Dog Genomes (Johnson, O'Brien, Schnabel, Taylor) and the cat Nine Lives project (Lyons, Gandolfi). Technological advancements have made it possible for an individual lab to conduct their own 1000 Genomes Project for their favorite species. Therefore, considering active projects at Mizzou, a logical short-term target for building a new bioinformatics cluster is the ability to process 1000 genomes in a reasonable timeframe.

A single ongoing analysis on campus requires the processing of about 1000 genomes and consumes roughly 1 million core-hours of compute time. A single 56 core node would take about two years of compute time to complete the required anlyses. There is only one machine in the campus cluster (**Lewis3**) with sufficient RAM to complete the analysis but it only has 24 cores. A cluster of 20 such **Lewis3** nodes would reduce the processing time down to about 36 days for this single project.

This proposal is to acquire an initial bioinformatics cluster and establish a sustainable funding model which would enable researchers to directly include the cost of the analysis in their grant funding. Research Computing Support Services has already invested over $500K in the General Purpose Research Storage to create a sustainable storage environment that allows researchers to easily include the cost of data storage in their grants. The bioinformatics cluster would be built and tuned specifically for the analyses required to process and analyze sequence data.

Description of the Proposed Hardware Infrastructure

The sequencing projects commonly conducted at Mizzou have the structural characteristics described in **Table 2** and result in about 10 million core hours, and a 20 node cluster could compute this in about one year. The cost of the system is described in **Table 3**.

Table 2. Characteristics of proposed 20 node Bioinformatics cluster.

Analysis Type:	# Cores	RAM (GB)	Scratch (TB)	Jobs /Batch	Batches	Length (Day)	# Labs	Core Hours
Alignment/Variant Calling	64	512	1	1000	1	0.75	5	7,680,000
RNA Assembly- Big	64	1024	1	10	2	14	5	2,150,400
RNA Assembly – Small	64	512	1	100	2	0.5	5	768,000

The cost of such a cluster is estimated as having an initial capital investment of $619,000. Note: that this does not include operational costs such as space, power, cooling, hardware or software support or the expertise to run a High Performance Computing cluster. The division of IT Research Computing Support Services proposes to run the newly acquired cluster as a part of its core mission to provide centralized research computing capability for the campus and the system.

Table 3. Cost characteristics of proposed 20 node Bioinformatics cluster

Item	Cost	Count	Total
Compute Node	$24,000	20	$480,000
10/40 Gigabit Switch	$6,000	1	$6,000
Rack/Power	$3,000	1	$3,000
100TB High Speed Scratch	$120,000	1	$120,000
Miscellaneous	$10,000	1	$10,000
Total			**$619,000**

Sustainability Model

Although some compute infrastructures already exist at Mizzou such as the **BioCluster** and the **MUgenomics** servers, they are not scalable to accommodate the demand of the growing campus community involved in bioinformatics. Computing infrastructure has an inherent lifecycle and must be upgraded and/or replaced. The model we propose involves the University and researchers sharing the costs. Under this model, the University pays for the initial acquisition of a system and researchers "buy-in" by allocating grant money to acquire additional compute nodes, storage and pay for life cycle updates. Recent upgrades of **Lewis3**, the purchase of General Purpose Research Storage and the proposed Bioinformatics cluster investment, position MU to establish a fee structure to make the system sustainable.

Storage

The Research Computing Support Services recently invested over $500K to implement more than one Petabyte (PB) of disc storage for the University's General Purpose Research Storage (GPRS). https://doit.missouri.edu/services/research/general-purpose-research-storage/.

This service offers a sustainable and flexible research data storage environment at reasonable cost. The Division of IT's Research Computing staff manage the data storage and will actively recruit

and support additional investments by researchers and other units on campus. GPRS options available to researchers include:

Individual Storage: Individual researchers are given 10 GB of private storage for private use at no charge. Additional storage is allocated in 256 GB increments at the cost of $10/TB/month.

Project Storage: Storage associated with specific groups of researchers or instruments are allocated on a per project basis. Each project is provided with 10 GB of storage at no charge. Additional storage will be allocated in 256 GB increments at the cost of $10/TB/month.

Special Project Storage: 50 TB of storage have been designated for special cases where the storage will be provided at a reduced cost or at no charge. These projects are expected to be short exploratory projects or special projects evaluated on a case-by-case basis considering the storage needs (size and duration). Special Projects might include promising exploratory research for short periods of time, exemplary research that has little chance of external funding or needs leading to the likely securing of funding, and other special-use cases as determined by MU's CI Council.

Large Storage Investment: Researchers with larger storage needs can invest in half or multiple entire nodes, around 100 TB per node, for five years of service. Investors receive dedicated access to one (half-node) or two (full-node) 10 Gigabit Ethernet (GbE) network storage ports. Nodes will be taken out of service after five years, but as part of a future storage offering, may be placed in a depreciated storage system.

The GPRS investment and cost structure now make it possible to budget for a project's storage costs during the grant writing process. Prior to the 2015 purchase of the GPRS this was not possible at Mizzou because no such infrastructure existed.

Budgeting the Costs in a Project Proposal

This has not traditionally occurred at Mizzou during the grant writing process because these types of cost templates have not been available. This has been for two reasons: 1) the infrastructure did not exist, and 2) there has not been the volume of data on campus with which to build these models.

The initial investment by the University in the Bioinformatics cluster will provide a baseline for estimating the cost structure associated with providing the required computational capacity in a sustainable way similar to the manner in which the General Purpose Research Storage (GPRS) operates. There are several groups on campus (IRCF, Animal Genomics and Soybean) with experience in managing large volumes of sequence data of multiple varieties such as genomic resequencing, RNA-seq, etc. These groups will create a template that new researchers can use to estimate the storage and computational costs associated with their project. For example, the Animal Genomics group has estimated that the cost to store and process a 30X coverage mammalian genome is approximately $264. This includes the cost to store the raw and processed data at $10/TB/mo for five years and the cost of purchasing an additional compute node at $15,000 amortized over 100 animals. Therefore, if someone was writing a new grant application proposing to sequence 100 genomes they would budget $26,400 to enable the sustainability of the Bioinformatics cluster and cover the computational costs associated with the research. The computational costs template can be included in the grant budget justification. Funding agencies, in particular, NSF, NIH and USDA routinely require data management plans and expect applicants to budget and make provisions for long-term data storage.

APPENDIX A

Sequencing related grants funded at MU organized by year of funding.

Species	Mizzou PI/Co-I	Dept /Division	Agency	Funding years	Total budget	Sequencing budget	Title
Soybean	Stacey, G	Plant Sci/Biochem	United Soy Board	2000 - 2005	$300,000		Development of soybean genetics and genomics
Plants	Pires	Biological Science	NSF	2005 - 2011	$1,176,000		Functional genomics of plant polyploids
Plants	Pires	Biological Science	NSF	2006 - 2011	$3,170,304		Toward unraveling the morphological plasticity and genome redundancy of Brassica oleracea
Maize	Evans (Stanford)/Givan	MMI	NSF	2007 - 2013	$4,755,985	$48,000	Functional Genomics of Maize Gametophytes
	Pires	Biological Science	MU Richard Wallace Research Incentive Grant Award	2008 - 2008	$4,000		Improving phylogenetic thinking in biology undergraduates
	Pires	Biological Science	NSF	2008 - 2010	$10,000		COLLABORATIVE RESEARCH: Comparative Investigation of Incipient Sex Chromosome Evolution in the genus Asparagus
	Pires	Biological Science	NSF	2008 - 2011	$275,093		COLLABORATIVE RESEARCH: From Acorus to Zingiber: Assembling the phylogeny of monocots
Soybean	Stacey, G	Plant Sci/Biochem	NSF	2008 - current	$3,000,000	$200,000	Soybean root hairs as a model for single cell systems biology
Human	Taylor K.	Pathology & Anatomical Sciences	Mizzou Advantage	2010 - 2010	$10,000	$10,000	Methylation analysis in leukemia
	Pires	Biological Science	MU Research Board	2010 - 2011	$30,000		Improving DNA barcoding using next-generation sequencing
Human	Taylor K.	Pathology & Anatomical Sciences	NIH	2010 - 2015	$750,000	$130,600	Towards defining the functional methylome in acute lymphoblastic leukemia
Dog	Johnson/O'Brien	Veterinary Medicine	Internal funding	2010 - current	$300,000	$300,000	Mizzou Comparative Canine Sequencing
Dog	O'Brien	Veterinary Medicine	Mizzou Advantage	2010-2010	$70,000	$70,000	Whole genome sequencing of an animal model of cerebral cortical dysplasia: Developing the next generation of genomics for human and animal health
Human trophoblast	Roberts/Ezashi	Animal Sciences	NIH	2010-2015	$1,717,855	$60,000	Pluripotent stem cells as models for normal and diseased trophoblast NIH R01 HD67790

Species	Mizzou PI/Co-I	Dept /Division	Agency	Funding years	Total budget	Sequencing budget	Title
	Pires	Biological Science	NSF	2011 - 2013	$14,951		DISSERTATION RESEARCH: Systematics of the Order Brassicales and Family Brassicaceae, and comparative genomic analyses (nuclear and organellar genomes): a phylogenomic approach
Cow	Taylor	Animal Science	USDA/AFRI	2011 - 2016	$10,000,000	$150,000	Integrated Program for Reducing Bovine Respiratory Disease Complex in Beef and Dairy Cattle
Cow	Taylor	Animal Science	USDA/AFRI	2011 - 2016	$5,000,000	$180,000	National Program for the Genetic Improvement of Feed Efficiency in Beef Cattle.
Metagenomic	Bryda		NIH	2011 - 2016	$1,038,598	$10,000	Rat Resource and Research Center
Mouse	Cummings/Givan		MU Research Board	2012 - 2013	$69,576	$16,992	Characterizing 5-HT and GABA Receptor Expression in 5-HT Deficient Mice
Varroa mite	Elsik	Animal and Plant Sciences	USDA/AFRI	2012 - 2013	$124,066	$124,066	Genome Characterization of the Mite Varroa destructor, the Primary Pest of Honey Bees.
Mosquito	Beerntsen	Veterinary Medicine	Internal funding	2012 - 2013	$18,000	$17,610	Mosquito Innate Immunity: From Immune Recognition Proteins to Signaling Pathways
	Pires	Biological Science	NSF	2012 - 2013	$14,999		Phylogeny and evolution of the Brassica crops and wild relatives (tribe Brassiceae, Brassicaceae): morphological diversity and homoplasy
Bacteria	Wall	Biochemistry	Discovery Project from DOE LBNL-SFA	2012 - 2014	$60,000	$12,000	Development of 3' RNA-seq for Quantifying Transcript End Sites
Mosquito	Beerntsen	Veterinary Medicine	Research Board	2012 - 2014	$62,186	$55,600	Mosquito Immunity: Pathogen Recognition to Signaling
Human	Taylor K.	Pathology &. Anatomical Sciences	Hyundai	2012 - 2015	$250,000	$38,500	Elucidating the role of DNA methylation in pediatric ALL relapse
Hymenopteran Insects	Elsik	Animal and Plant Sciences	USDA/AFRI	2012 - 2015	$392,685	$196,000	Mining Hymenoptera Genomes for Functional Sequences
	Pires	Biological Science	NSF	2012 - 2015	$468,844		Phylogenomics of polyploidy in the Brassicales
Cow	Elsik	Animal and Plant Sciences	USDA/AFRI	2012 - 2015	$660,451	$330,000	The Next Generation Bovine Genome Database
Sheep	Spencer	Animal Science	USDA/AFRI	2012 - 2016	$499,000	$40,000	Physiological Roles of Hydroxysteroid (11-beta) Dehydrogenases and Cortisol in Early Pregnancy
Cow	Spencer	Animal Science	NIH	2012 - 2017	$1,700,000	$120,000	System biology approach to understand endometrial receptivity & pregnancy loss
Cat	Lyons/Gandolfi	Veterinary Medicine & Surgery	Winn Feline Foundation	2013 - 2014	$30,552	$25,000	9 Lives Cat Genome Sequencing Initiative

Species	Mizzou PI/Co-I	Dept /Division	Agency	Funding years	Total budget	Sequencing budget	Title
Dog	O'Brien/Johnson	Veterinary Medicine	Poodle Club of America Foundation	2013 - 2014	$15,000	$12,000	Applying Whole Genome Sequencing to Identify the Mutation Responsible for Polymicrogyria in Standard Poodles.
Cat	Lyons/Gandolfi	Veterinary Medicine & Surgery	Donation	2013 - 2014	$70,000	$70,000	Unrestricted gift
Sheep	Conant	Animal Science	USDA/AFRI	2013 - 2015	$124,000	$80,000	Improving Profitability and Sustainability of Sheep Production
Soybean	Nguyen/Xu/Joshi	Plant Science/CS/MUII/MMI/IPG	United Soybean Board	2013 - 2015	$869,000	$300,000	Large Scale Sequencing of Germplasm to Develop Genomic Resources for Soybean Improvement
Cow	Patterson	Animal Science	USDA/AFRI	2013 - 2016	3,000,000	$504,000	Identification and management of alleles impairing heifer fertility while optimizing genetic gain in Angus cattle
Bacteria	Xu	Computer Science	USDA	2013 - 2016	$450,000	$100,000	Prediction and Control of the Performance of Anaerobic Digestion of Animal Manure through Metagenomics for Renewable Energy
Cow	Elsik	Animal and Plant Sciences	USDA/AFRI	2013 - 2016	$499,876	$250,000	Web-Based Tools for Curation and Display of Locus-Specific Alternate Assemblies
Maize	Buckler PI (Cornell)/Flint-Garcia (Mizzou)	Plant Sciences/USDA	NSF	2013 - 2018	$12,500,000	$2,685,000	Biology of rare alleles in maize and its wild relatives
Cow	Spencer	Animal Science	USDA/AFRI	2013 - 2018	3,000,000	$150,000	Improving Fertliity of Dairy Cattle Using Translational Genomics
Human	Ezashi, Schust/Schulz	Animal Science/ObGyn	NIH	2013 - 2018	1,592,565	$24,000	Pluripotent Stem Cells: Modeling syncytiotrophoblast development and pathogenesis
Cat	Lyons/Gandolfi	Veterinary Medicine & Surgery	Endowment	2013 - current	$80,000	$80,000	Gilbreath McLorn Endowment of Comparative Medicine
	Pires	Biological Science	USDA	2014 - 2015	$13,240		Genomic diversity among morphotypes of Brassica rapa vegetable accessions
Dog	O'Brien/Johnson	Veterinary Medicine	Black Russion Terrier Club of America	2014 - 2015	$12,000	$6,000	Identification of the Mutation Responsible for Juvenile Laryngeal Paralysis & Polyneuropathy in Black Russian Terriers
Cat	Lyons/Gandolfi	Veterinary Medicine & Surgery	National Genographic	2014 - 2015	$70,000	$72,000	Sequencing the genomes of twelve cats
	Pires	Biological Science	MU Research Board	2014 - 2015	$24,922		The evolution of grass photosynthesis
Mouse	Spencer	Animal Science	NIH	2014 - 2016	$415,250	$40,000	Biological Role of Endometrial Glands in Uterine Function
Human	Taylor K.	Pathology & Anatomical Sciences	Gift	2014 - 2016	$28,444	$15,000	Palmer Histiocytosis Research

Species	Mizzou PI/Co-I	Dept /Division	Agency	Funding years	Total budget	Sequencing budget	Title
	Pires	Biological Science	Mizzou Advantage	2014 - 2016	$65,269		The evolution of grass photosynthesis: applications for food and sustainable energy
Mosquito	Franz	Veterinary Pathobiol	NIH	2014 - 2016	$275,000	$20,000	Transgenic resistance of Aedes aegypti to the four serotypes of dengue virus
Brassica (cabbage, kale, cauliflower etc)	Pires	Biological Sciences	NSF	2014 - 2017	2,179,716	$400,000	Polyploidy and plasticity in the crop Brassicas
Bacteria	Wall	Biochemistry	DOE LBNL SFA	2014 - 2017	$530,000	$30,000	Rapid deduction and comparative genomics of survival mechanisms of metal-reducing microbes
Soybean	Walker/Elsik	Animal and Plant Sciences	NSF	2014 - 2019	$20,000,000	$190,000	The Missouri Transect: Climate, Plants, and Community
Cat	Lyons/Gandolfi	Veterinary Medicine & Surgery	MUGiveDirect	2014 - current	$50,000	$50,000	Direct donations for cat sequencing.
Human	Taylor K.	Pathology & Anatomical Sciences	BTCF foundation	2015 - 2016	$59,000	$20,000	Acute lymphoblastic leukemia research
Fruit Fly	Hagen	Animal Science	MU Research Council	2015 - 2016	$9,890	$9,890	Analysis of non-coding RNA in Wolbachia induced cytoplasmic incompatibility
	Pires/Conant	Biological Science/Animal Science	NSF	2015 - 2016	$15,329		DISSERTATION RESEARCH: C4 Photosynthetic Evolution; Sub-types, Diversity, and Function within the Grass Tribe Paniceae
Bacteria	Wall	Biochemistry	Mizzou Advantage	2015 - 2016	$149,954	$15,000	Genetic Requirements for Biofilm Formation of *Desulfovibrio vulgaris* Hildenborough
Cow	Hagen	Animal Science	MU Research Board	2015 - 2016	$48,000	$42,000	Genome-wide DNA methylation in Bovine LOS
Human	Bashkin (UMSL)/Givan	MMI	UM IDIC	2015 - 2016	$100,000	$50,000	Prevention of Cervical Cancer: Footprinting Anti-HPV Drug-DNA Interactions in Live Cells Using Gamma Ray-Generated Hydroxyl Radicals and Next-Generation DNA Sequencing
Sheep	Spencer	Animal Science	USDA/AFRI	2015 - 2019	$452,000	$60,000	Biological Roles of Exosomes/Microvesicles in Conceptus Elongation and Uterine Interactions During Early Pregnancy
Metagenomic	Franklin	Veterinary Pathobiology	NIH	2015 - 2020	$986,454	$25,000	The Mutant Mouse Resource and Research Center (MMRRC) at the University of Missouri
Mouse	Spencer	Animal Science	NIH	2016 - 2018	$422,124	$20,000	Generation of a Transgenic Mouse Model to Study Uterine Gland Function
Bacteria	Taylor/Schnabel	Animal Science	USDA/AFRI	2016 - 2019	$499,993	$18,500	A novel in silico approach to GWAS for enhanced resistance to bacterial mastitis in dairy cattle

Species	Mizzou PI/Co-I	Dept /Division	Agency	Funding years	Total budget	Sequencing budget	Title
Sheep	Spencer	Animal Science	USDA/AFRI	2016 - 2021	1,700,000	$150,000	Exosomes/microvesicles: Novel Mediators of Uterine Receptivity and Conceptus-Maternal Interactions
Human	Taylor K.		NIH	2016 - 2021	1,250,000	$25,000	Functional analyses of differentially methylated regulatory elements in leukemia

APPENDIX B

Summary of peer reviewed journal activity. Shaded journals have an impact factor > 10.

Journal Name	Num. Papers	Impact Factor	Category	Rank/total	Quartile
Acta Veterinaria Scandinavica	1	1.377	Veterinary Sciences	36/133	Q2
American Journal of Physiology-Regulatory, Integrative and Comparative Physiology	1	3.106	Physiology	25/83	Q2
Animal Genetics	1	2.207	Agriculture, Dairy & Animal Science	3/57	Q1
Applied and Environmental Microbiology	2	3.668	Biotechnology & Applied Microbiology	34/163	Q1
Archives of Virology	1	2.390	Virology	18/33	Q3
Artificial Intelligence in Medicine	1	2.019	Computer Science, Artificial Intelligence	40/123	Q2
Bioinformatics	2	4.981	Biochemical Research Methods	8/79	Q1
Biology of Reproduction	4	3.318	Reproductive Biology	5/30	Q1
BMC Bioinformatics	4	2.576	Biochemical Research Methods	34/79	Q2
BMC Evolutionary Biology	1	3.368	Evolutionary Biology	17/46	Q2
BMC Genomics	18	3.986	Biotechnology & Applied Microbiology	26/163	Q1
BMC Plant Biology	2	3.813	Plant Sciences	22/204	Q1
BMC Systems Biology	1	2.435	Mathmatical & Computational Biology	13/57	Q1
BMC Veterinary Research	1	1.777	Veterinary Sciences	21/133	Q1
Cancer Research	1	9.329	Oncology	11/211	Q1
Cellular Reprogramming	2	1.788	Cell & Tissue Engineering	17/21	Q4
Cold Spring Harbor Symposia on Quantitative Biology	1	0.856	Biochemistry & Molecular Biology	224/261	Q4
Crop Science	1	1.575	Agronomy	23/81	Q2
DNA Research	1	5.477	Genetics & Heredity	23/167	Q1
Epigenetics	3	4.780	Biochemistry & Molecular Biology	56/290	Q1
FASEB JOURNAL	1	5.043	Biochemistry & Molecular Biology	50/290	Q1
Frontiers in Microbiology	1	3.989	Microbiology	27/119	Q1
Frontiers in Plant Science	2	3.948	Plant Sciences	19/204	Q1
Functional Plant Biology	1	3.145	Plant Sciences	34/204	Q1
Genes and Development	1	10.798	Cell Biology	16/184	Q1
Genetics	4	5.963	Genetics & Heredity	21/167	Q1
Genome	1	1.424	Biotechnology & Applied Microbiology	114/163	Q3
Genome Biology	3	10.381	Biochemistry & Molecular Biology	5/163	Q1
Genome Biology and Evolution	1	4.229	Evolutionary Biology	9/46	Q1
Genome Research	5	14.630	Biochemistry & Molecular Biology	4/290	Q1
Human Genomics	1	2.146	Genetics & Heredity	105/167	Q3

Journal Name	Num. Papers	Impact Factor	Category	Rank/total	Quartile
International Journal of Data Mining and Bioinformatics	1	0.495	Mathmatical & Computational Biology	56/57	Q4
International Journal of General Systems	1	1.637	Computer Science, Theory & Methods	18/102	Q1
Investigative Ophthalmology and Visual Science	1	3.404	Opthalmology	7/57	Q1
Journal of Animal Science	1	2.108	Agriculture, Dairy & Animal Science	5/57	Q1
Journal of Bioinformatics and Computational Biology	2	0.783	Mathmatical & Computational Biology	48/57	Q4
Journal of Experimental Botany	3	5.526	Plant Sciences	12/204	Q1
JOURNAL OF PROTEOME RESEARCH	1	4.245	Biochemical Research Methods	14/79	Q1
Journal of Reproduction and Development	1	1.515	Agriculture, Dairy & Animal Science	13/57	Q1
Journal of Veterinary Internal Medicine	4	1.879	Veterinary Sciences	17/133	Q1
Molecular Breeding	1	2.246	Agronomy	14/81	Q1
Molecular Genetics and Genomics	1	2.728	Biochemistry & Molecular Biology	144/290	Q2
Molecular Genetics and Metabolism	2	2.625	Biochemistry & Molecular Biology	149/290	Q3
Molecular Plant Pathology	1	4.724	Plant Sciences	16/204	Q1
MOLECULAR PLANT-MICROBE INTERACTIONS	1	3.944	Plant Sciences	20/204	Q1
Molecular Psychiatry	1	14.496	Biochemistry & Molecular Biology	5/290	Q1
Molecular Reproduction and Development	2	2.527	Biochemistry & Molecular Biology	159/290	Q3
Nature	2	41.456	Multidisciplinary Sciences	1/57	Q1
Nature Communications	1	11.470	Multidisciplinary Sciences	3/57	Q1
Nature Genetics	1	29.352	Genetics & Heredity	2/167	Q1
Nature Methods	1	32.072	Biochemical Research Methods	1/79	Q1
Neurobiology of Disease	2	5.078	Neurosciences	39/252	Q1
New Phytol	1	7.672	Plant Sciences	6/204	Q1
Nucleic Acids Research	8	9.112	Biochemistry & Molecular Biology	20/290	Q1
PHYSIOLOGICAL ENTOMOLOGY	1	1.416	Entomology	31/92	Q2
Physiological Genomics	3	2.374	Cell Biology	128/184	Q3
Plant Biotechnology Journal	3	5.752	Biotechnology & Applied Microbiology	15/163	Q1
Plant Cell	3	9.338	Plant Sciences	4/204	Q1
Plant Genetic Resources-Characterization and Utilization	1	0.580	Plant Sciences	165/204	Q4
Plant Journal	2	5.972	Plant Sciences	10/204	Q1
Plant Physiology	4	6.841	Plant Sciences	8/204	Q1
Plant Science	1	3.607	Plant Sciences	27/204	Q1
PLoS Biology	1	9.343	Biochemistry & Molecular Biology	17/290	Q1
PLoS Genetics	4	7.528	Genetics & Heredity	14/167	Q1

Journal Name	Num. Papers	Impact Factor	Category	Rank/total	Quartile
PLoS Neglected Tropical Diseases	1	4.446	Parasitology	5/36	Q1
PLOS ONE	19	3.234	Multidisciplinary Sciences	9/57	Q1
Proceedings of the National Academy of Sciences USA	5	9.674	Multidisciplinary Sciences	4/57	Q1
Proceedings of the Royal Society B: Biological Sciences	1	5.051	Biology	8/85	Q1
RNA-A PUBLICATION OF THE RNA SOCIETY	1	4.936	Biochemistry & Molecular Biology	53/290	Q1
Science	1	33.611	Multidisciplinary Sciences	2/57	Q1
Science of the Total Environment	1	4.099	Environmental Sciences	18/223	Q1
World Mycotoxin Journal	1	2.157	Mycology	13/24	Q3
Asian J. Plant Pathology	1				
Current Opinion in Insect Science	1				
G3: Genes|Genomes|Genetics	2				
Genome Announcements	1				
International Journal of Bioinformatics Research and Applications	1				
International Society for Microbial Ecology Journal	1				
Journal of Computer and Science Technology	1				
The Journal of Physiology	1				
Theoretic Applied Genetics	1				

APPENDIX C

Sequencing Research Related Publications

Submitted (10)

Valliyodan B, Qiu D, Patil G, Zeng P, Huang J, Dai L, Chen C, Zeng L, **Joshi T,** Song L, Vuong T, Musket T, **Xu D**, Shannon JG, Shifeng C, Liu X, Nguyen HT. Landscape of genomic diversity and trait discovery in soybean. 2015. *Submitted*.

Liu Y, Khan SM, Wang J, Chen S, Rynge M, Wang J, Maldonado dos Santos JV, Valliyodan B, Merchant N, **Nguyen HT**, **Xu D**, **Joshi T**. PGen: Large-Scale Pegasus Workflow for Genomic Variation Analysis in SoyKB. Bioinformatics Application Notes. 2015 *Submitted*.

Oswaldo Valdés-López, Josef Batek, Nicolas Gomez'Hernandez, Ning Zhang, **Trupti Joshi**, **Dong Xu**, Kim K, Hixson, Karl K, Weitz, Joshua T. Aldrisch, Ljiljana Paša-Tolic and **Gary Stacey**. Soybean root hairs grown under heat stress show global changes in their transcriptional and proteomic profile. Frontiers in Plant Science. 2015 *Submitted*.

Lyons, L.A., Creighton, E.K., Alhaddad, H., Beale, H.C., Grahn, R.A., Rah, H., Maggs, D.J., Helps, C.R., and **Gandolfi, B**. (2015). Whole Genome Sequencing Identifies an AIPL1 Variant in Persian Cats as a New Model for Leber's Congenital Amaurosis. *Submitted*.

Gandolfi, B., Alamri, S., Darby, W.G., Adhikari, B., Lattimer, J.C., Malik, R., Wade, C.M., **Lyons, L.A**., Cheng, J., Bateman, J.F., McIntyre, P., Lamande, S.R., and Haase, B. (2015). A dominant TRPV4 variant implicated in osteochondrodysplasia of Scottish fold cats. *Submitted*.

Bickhart DM, L Xu, JL Hutchison, JB Cole, DJ Null, SG Schroeder, J Song, JF Garcia, TS Sonstegard, CP Van Tassell, **RD Schnabel**, **JF Taylor**, HA Lewin, and GE Liu. 2015. The tandem distribution pattern of cattle CNV facilitates its population genetic analyses. DNA Res. *Submitted*.

Taylor JF, LK Whitacre, JL Hoff, PC Tizioto, **JW Kim**, **JE Decker** and **RD Schnabel**. 2015. Lessons from cattle genome and transcriptome sequencing. Genet Sel Evol *Submitted*.

Forde N, Maillo V, O'Gaora P, Simintiras CA, Sturmey RG, Ealy AD, **Spencer TE**, Gutierrez-Adan A, Rizos D, Lonergan P. Sexually dimorphic gene expression in male and female bovine conceptuses at the initiation of implantation. Proc Natl Acad Sci USA; *Submitted*.

Oswaldo Valdés-López, Josef Batek, Nicolas Gomez'Hernandez, Cuong T. Nguyen, Mariel Carolina Isidra-Arellano, Ning Zhang, Trupti Joshi, Dong Xu, Kim K, Hixson, Karl K, Weitz, Joshua T. Aldrisch, Ljiljana Paša-Tolic and Gary Stacey (2015) Soybean root hairs grown under heat stress show global changes in their transcriptional and proteomic profile. Frontiers Plant Sci. *Submitted*.

Emma L Ivansson, Kate Megquier, Sergey V Kozyrev, Eva Murén, Izabella Baranowska Körberg, Ross Swofford, Michele Koltookian, Noriko Tonomura, Rong Zeng, Ana L Kolicheski, Liz Hansen, Martin L Katz, Gayle C Johnson, Gary S Johnson, Joan R Coates, Kerstin Lindblad-Toh. Variants within SP110 modify risk of canine degenerative myelopathy: a model for amyotrophic lateral sclerosis. Nature Communications, *Submitted.*

2016 (1)

Liu, Z, J Cook, S Melia-Hancock, K Guill, **C Bottoms**, A Garcia, O Ott, RJ Nelson, J Recker, P Balint-Kurti, S Larsson, N Lepak, E Buckler, L Trimble, W Tracy, MD McMullen, and **SA Flint-Garcia**. Expanding maize genetic resources with pre-domestication alleles: maize-teosinte introgression populations. (2016) *The Plant Genome*. doi: 10.3835/plantgenome2015.07.0053

2015 (51)

Whitacre LK, PC Tizioto, **JW Kim**, TS Sonstegard, LJ Alexander, JF Medrano, **RD Schnabel**, **JF Taylor** and **JE Decker**. 2015. What's in your next-generation sequence data? An exploration of unmapped DNA and RNA sequence reads from the bovine reference individual. *BMC Genomics* 16: 1114.

Almamun, M., Levinson, B.T., van Swaay, A.C., Johnson, N.T., McKay, S.D., Arthur, G.L., **Davis, J.W.** and **Taylor, K.H.**, 2015. Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia. *Epigen*. 10(9):882-90. PMID:26308964. citations 0.

Bickhart DM, Hutchison JL, Xu L, **Schnabel RD**, **Taylor JF**, Reecy JM, Schroeder S, Van Tassell CP, Sonstegard TS, Liu GE. 2015. RAPTR-SV: a hybrid method for the detection of structural variants. *Bioinformatics 2015, Feb. 1-7*. citations 1.

Brechenmacher, L., T.H.N. Nguyen, N. Zhang, T-H Jun, D. Xu, M.A. Rouf Mian, G. Stacey. (2015) Identification of soybean proteins and genes differentially regulated in near isogenic lines differing in resistance to aphid infestation *J. Proteome Res*. (in press). citations 0.

Brooks KE, Burns GW, **Spencer TE**. Peroxisome proliferator activator receptor gamma (PPARG) regulates conceptus elongation in sheep. *Biol Reprod* 2015; 92(2):42. citations 1.

Brooks KE, **Spencer TE**. Biological roles of interferon tau (IFNT) and type I IFN receptors in elongation of the ovine conceptus. *Biol Reprod* 2015; 92(2):47. citations 1.

Chen Z, Hagen DE, Elsik CG, Ji T, Morris CJ, Moon LE, Rivera RM. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *Proc Natl Acad Sci U S A*. 2015 Apr 14;112(15):4618-23. citations 0.

Chenglin Chai, Yongqin Wang, **Trupti Joshi**, Babu Valliyodan, Silvas Prince, Lydia Michel, **Dong Xu** and **Henry T Nguyen**. Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance. *BMC Genomics*, 16:596 doi:10.1186/s12864-015-1743-6, 2015. citations 0.

Chengli Xu, Xiangwu Ju, Dandan Song, Fengming Huang, Depei Tang, Zhen Zou, Chao Zhang, **Trupti Joshi**, Lijuan Jia, Weihai Xu, Kai-Feng Xu, Qian Wang, Yanlei Xiong, Zhenmin Guo, Xiangmei Chen, Fumin Huang, Jiantao Xu, Ying Zhong, Yi Zhu, Yi Peng, Li Wang, Xinyu Zhang, Rui Jiang, Dangsheng Li, Tao Jiang, **Dong Xu**, Chengyu Jiang. An association analysis between psychophysical characteristics and genome-wide gene expression changes in human adaptation to the extreme climate at the Antarctic Dome Argus. *Molecular Psychiatry*, 20:536–544; doi:10.1038/mp.2014.72, 2015. citations 0.

Daiyong Denga, Ning Zhangb, **Dong Xu**, Mary Reeda, Fengjing Liua, Guolu Zheng. Polymorphism of the glucosyltransferase gene (ycjM) in Escherichia coli and its use for tracking human fecal pollution in water. Science of the Total Environment. 537:260–267, 2015. citations 0.

Edger, P. P., H. M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glackner, A. E. Platts, D. G. Heckel, J. P. Der, E. K. Wafula, M. Tang, J. A. Hofberger, A. Smithson, J. C. Hall, M. Blanchette, T. E. Bureau, S. I. Wright, M. E. Schranz, M. S. Barker, **G. C. Conant**, N. Wahlberg, H. Vogel, J. C. Pires and C. W. Wheat. (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, U.S.A.*, **112**: 8362-8366. citations 3.

Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, Hagen DE. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Res*. 2015 Nov 17. pii: gkv1208.

Elsik CG, Unni DR, Diesh CM, Tayal A, Emery ML, Nguyen HN, Hagen DE. Bovine Genome Database: new tools for gleaning function from the Bos taurus genome. *Nucleic Acids Res*. 2015 Oct 19. pii: gkv1077.

Ericsson A.C., Davis J.W., Spollen W., Bivens N., **Givan S.A**., Hagan, C., **McIntosh M.**, Franklin C.L. (2015). Effects of Vendor and Genetic Background on the Composition of the Fecal Microbiota of Inbred Mice. *PLOS ONE*, DOI: 10.1371/journal.pone.0116704. citations 6. citations 7.

Franz, AWE, Balaraman, V, Fraser, MJ. 2015. Disruption of dengue virus transmission by mosquitoes. *Current Opinion in Insect Science* 6, 1-9. citations 0.

Gandolfi, B., Grahn, R.A., Creighton, E.K., Williams, D.C., Dickinson, P.J., Sturges, B.K., Guo, L.T., Shelton, G.D., Leegwater, P.A., Longeri, M., Malik, R., and **Lyons, L.A.** (2015). COLQ variant associated with Devon Rex and Sphynx feline hereditary myopathy. *Anim Genet*. In press citations 0.

Garima Kushwaha, Mikhail Dozmorov, Jing Qiu, Huidong Shi, **Dong Xu**. Hypomethylation Coordinates Antagonistically with Hypermethylation in Cancer Development: A Case Study of Leukemia. *Human Genomics*. In press. citations 0.

Gilliam DH, Kolicheski A, **Johnson GS**, Mhlanga-Mutangadura T, **Taylor JF**, **Schnabel RD**, **Katz ML**. 2015. Golden Retriever dogs with neuronal ceroid lipofuscinosis have a two-base-pair deletion and frameshift in CLN5. *Molecular Genetics and Metabolism* Jun-Jul;115(2-3):101-9. citations 0.

Givan S.A., Zhou M-Y, Bromert K., Bivens N., Chapman L.F. (2015). Genome sequences of Pseudoalteromonas strains ATCC BAA-314, ATCC 70018, and ATCC 70019. *Genome Announcements* 3(3):e00390-15. doi:10.1128/genomeA.00390-15 citations 0.

Gunvant Patil, Babu Valliyodan, Rupesh Deshmukh, Silvas Prince, Bjorn Nicander, Mingzhe Zhao, Humira Sonah, Li Song, Li Lin, Juhi Chaudhary, Yang Liu, Trupti Joshi, **Dong Xu** and **Henry T. Nguyen**. Soybean (Glycine max) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics*, 16:520, 2015. citations 0.

Hossain MS, **Joshi T**, **Stacey G**. System approaches to study root hairs as a single cell plant model: current status and future perspectives. *Front Plant Sci*. 2015 May 19;6:363. doi: 10.3389/fpls.2015.00363. eCollection 2015. citations 0.

Jinrong Wan, Tri Vuong, Yongqing Jiao, **Trupti Joshi**, Hongxin Zhang, **Dong Xu** and **Henry T Nguyen**. Whole-genome gene expression profiling revealed genes and pathways potentially involved in regulating interactions of soybean with cyst nematode (Heterodera glycines Ichinohe). *BMC Genomics*, 16:148, doi:10.1186/s12864-015-1316-8, 2015. citations 2.

Juexin Wang, Jason Vander Heiden, Namita T Gupta, Mate Nagy, **Trupti Joshi**, Babu Valliyodan, Haiying Shi, Yanchun Liang, **Henry T. Nguyen**, Jing Zhang and **Dong Xu**. A Bayesian Model for Detection of High-order Interactions among Genetic Variants in Genome-wide Association Studies. *BMC Genomics*. In press. citations 0.

Kiwamu Tanaka, Sung-Hwan Cho, Hyeyoung Lee, An Q. Pham, Josef M. Batek, Shiqi Cui, Jing Qiu, Saad M. Khan, **Trupti Joshi**, Zhanyuan J. Zhang, **Dong Xu** and **Gary Stacey**. Effect of lipo-chitooligosaccharide on early growth of C4 grass seedlings. *J. Exp. Bot.* doi: 10.1093/jxb/erv260, 2015. citations 0.

Marsden CD, Vecchyo DO, **Johnson GS**, **O'Brien DP**, **Taylor JF**, Ramirez O, Vilà C, Marques-Bonet T, **Schnabel RD**, Wayne RK and Lohmueller KE. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Accepted Proceedings of the National Academy of Sciences USA*. citations 0.

Mhlanga-Mutangadura T, **Johnson GS**, **Schnabel RD**, **Taylor JF**, **Johnson GC**, **Katz ML**, Shelton GL, Lever T, Giuliano E, Granger N, Shomper J and **O'Brien DP**. A homozygous RAB3GAP1:c.743delC mutation is associated with juvenile onset laryngeal paralysis and polyneuropathy with ocular abnormalities and neuronal vacuolation in Black Russian Terrier dogs. *Accepted Neurobiology of Disease*. citations 0.

Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., **King E.G.**, Long, A.D., and Macdonald, S.J. In Press. Identifying loci contributing to resistence to xenobiotics in Drosophila. *PLoS Genetics*. citations 0.

Noyes HA, Daly D, Goodhead I, Kaye S, Kemp SJ, Kenny J, Saccheri I, **Schnabel RD**, **Taylor JF**, and Hall N. A simple procedure for directly obtaining haplotype sequences of diploid genomes. *BMC Genomics 16:642*. citations 0.

Prince SJ, Li S, Qiu D, Maldonado dos santos JV, Chai C, **Joshi T**, Patil G, Valliyodan B, Vuong T, Murphy M, Krampis K, Tucker DM, Biyashev R, Dorrance AE, Maroof MAS, **Xu D**, Shannon G, **Nguyen HT**. Genetic variants in root architecture-related genes in a Glycine soja accession, a potential resource to improve cultivated soybean. *BMC Genomics*. 2015 Feb 25;16:132. doi: 10.1186/s12864-015-1334-6. citations 3.

R. Michael Roberts, Shinichiro Yabe, Ying Yang, and **Toshihiko Ezashi** (2015) A Human Stem Cell Model for Creating Placental Syncytiotrophoblast, the Major Cellular Barrier that Limits Fetal Exposure to Xenobiotics. In: Stem Cells in Toxicology and Medicine (Saura S Sahu, Editor) John Wiley & Sons Ltd., Chichester, UK. In Press citations 0.

Rohdin C, Gilliam D*, O'Leary C, **O'Brien DP**, **Coates, J**, **Johnson G**, Jäderlund, K: A KCNJ10 mutation previously identified in the Russell group of terriers also occurs in Smooth-Haired Fox Terriers with hereditary ataxia and in related breeds. 2015. *Acta Veterinaria Scandinavica*. 2015, 57:26. citations 0.

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJ, Hasselmann M, Lozier JD, Robertson HM, Smagghe G, Stolle E, Van Vaerenbergh M, Waterhouse RM, Bornberg-Bauer E,

Klasberg S, Bennett AK, Câmara F, Guigó R, Hoff K, Mariotti M, Munoz-Torres M, Murphy T, Santesmasses D, Amdam GV, Beckers M, Beye M, Biewer M, Bitondi MM, Blaxter ML, Bourke AF, Brown MJ, Buechel SD, Cameron R, Cappelle K, Carolan JC, Christiaens O, Ciborowski KL, Clarke DF, Colgan TJ, Collins DH, Cridge AG, Dalmay T, Dreier S, du Plessis L, Duncan E, Erler S, Evans J, Falcon T, Flores K, Freitas FC, Fuchikawa T, Gempe T, Hartfelder K, Hauser F, Helbing S, Humann FC, Irvine F, Jermiin LS, Johnson CE, Johnson RM, Jones AK, Kadowaki T, Kidner JH, Koch V, Köhler A, Kraus FB, Lattorff HM, Leask M, Lockett GA, Mallon EB, Antonio DS, Marxer M, Meeus I, Moritz RF, Nair A, Näpflin K, Nissen I, Niu J, Nunes FM, Oakeshott JG, Osborne A, Otte M, Pinheiro DG, Rossié N, Rueppell O, Santos CG, Schmid-Hempel R, Schmitt BD, Schulte C, Simões ZL, Soares MP, Swevers L, Winnebeck EC, Wolschin F, Yu N, Zdobnov EM, Aqrawi PK, Blankenburg KP, Coyle M, Francisco L, Hernandez AG, Holder M, Hudson ME, Jackson L, Jayaseelan J, Joshi V, Kovar C, Lee SL, Mata R, Mathew T, Newsham IF, Ngo R, Okwuonu G, Pham C, Pu LL, Saada N, Santibanez J, Simmons D, Thornton R, Venkat A, Walden KK, Wu YQ, Debyser G, Devreese B, Asher C, Blommaert J, Chipman AD, Chittka L, Fouks B, Liu J, O'Neill MP, Sumner S, Puiu D, Qu J, Salzberg SL, Scherer SE, Muzny DM, Richards S, Robinson GE, Gibbs RA, Schmid-Hempel P, Worley KC. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* 2015 Apr 24;16:76. citations 10.

Saritas-Yildirim B, Childers CP, Elsik CG, Silva EM. Identification of REST targets in the Xenopus tropicalis genome. *BMC Genomics*. 2015 May 14;16:380. citations 0.

Sha Lu, Xiaoyan Yin, William Spollen, Ning Zhang, **Dong Xu**; Kristin Bilyeu, James Schoelz, and Zhanyuan Zhang. Analysis of the siRNA-mediated gene silencing process targeting three homologous genes controlling soybean seed oil quality. *PLoS One*. DOI: 10.1371/journal.pone.0129010, 2015. citations 0.

Silvas J Prince, **Trupti Joshi**, Raymond N Mutava, Naeem Syed, Joao V Maldonado dos Santos, Gunvant Patil, Li Song, JiaoJiao Wang, Li Lin, Wei Chen, J. Grover Shannon, Babu Valliyodan, **Dong Xu**, **Henry T Nguyen**. Comparative analysis of drought-responsive transcriptome in soybean lines contrasting for canopy wilting. *Plant Science*. 240:65–78, 2015. citations 0.

Silvas J Prince, Li Song, Dan Qiu, Joao V Maldonado dos Santos, Chenglin Chai, **Trupti Joshi**, Gunvant Patil, Babu Valliyodan, Tri D Vuong, Mackensie Murphy, Konstantinos Krampis, Dominic M Tucker, Ruslan Biyashev, Anne E Dorrance, MA Saghai Maroof, **Dong Xu**, J Grover Shannon and **Henry T Nguyen**. Genetic variants in root architecture-related genes in a Glycine soja accession, a potential resource to improve cultivated soybean. *BMC Genomics*, 16:132, doi:10.1186/s12864-015-1334-6, 2015. citations 3.

Syed NH, Prince SJ, Mutava RN, Patil G, Li S, Chen W, Babu V, **Joshi T**, Khan S and **Nguyen HT**. Core clock, SUB1, and ABAR genes mediate flooding and drought responses via alternative splicing in soybean. *Journal of Experimental Botany*. 2015. doi:10.1093/jxb/erv407. citations 0.

Tanaka K, Cho SH, Lee H, Pham AQ, Batek JM, Cui S, Qiu J, Khan SM, **Joshi T**, Zhang ZJ, **Xu D**, **Stacey G**. Effect of lipo-chitooligosaccharide on early growth of C4 grass seedlings. *J Exp Bot*. 2015 Jun 6. pii: erv260. [Epub ahead of print] citations 0.

Taxis TM, Wolff S, Zhang C, Dai J, **Schnabel RD**, **Taylor JF**, **Kerley MS**, **Pires JC**, Gregg SJ, **Lamberson WR** and **Conant GC**. The players may change but the game remains: Network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Accepted Nucleic Acid Research*. citations 0.

Tizioto PC, **Kim JW**, Seabury CM, **Schnabel RD**, Gershwin LJ, Van Eenennaam AL, Toaff-Rosenstein R, Neibergs HL, the Bovine Respiratory Disease Complex Coordinated Agricultural Project Research Team, and **Taylor JF**. Immunological Response to Single Pathogen Challenge with Agents of the Bovine Respiratory Disease Complex: An RNA-sequence analysis of the bronchial lymph node transcriptome. *Accepted PLoS One*. citations 0.

Tizioto PC, Coutinho LL, **Taylor JF**, **Decker JE**, **Schnabel RD**, Rosa KO, Oliveira PSN, Souza MM, Mourão GB, Tullio RR, Chaves AS, Lanna DPD, Zerlotini A, Mudadu MA and Regitano LCA. 2015. Global liver gene expression differences in Nelore steers with divergent residual feed intake phenotypes. *BMC Genomics 16:242*. citations 0.

Wan J, Vuong T, Jiao Y, **Joshi T**, Zhang H, Cui S, Qiu J, **Xu D**, and **Nguyen HT**. Whole-genome gene expression profiling revealed genes and pathways potentially involved in regulating the interactions of soybean with cyst

nematode (Heterodera glycines Ichinohe). *BMC Genomics* 2015, 16:148. doi:10.1186/s12864-015-1316-8. citations 2.

Whitworth, K.M., J. Mao, K. Lee, W.G. Spollen, M.S. Samuel, **E.M. Walters**, L.D. Spate, **R.S. Prather**. 2015. Transcriptome analysis of pig in vivo, in vitro fertilized and nuclear transfer blastocyst stage embryos treated with histone deacetylase inhibitors post-fusion and activation reveals changes in the lysosomal pathway. *Cellular Reprogramming* 17:243-258. citations 0.

Yongqing Jiao, Tri D. Vuong, Yang Liu, Zenglu Li, Jim Noe, Robert Robbins, **Dong Xu**, J. Grover Shannon and **Henry Nguyen**. Identification and characterization of quantitative trait loci underlying resistance to southern root-knot and reniform nematodes in soybean accession PI 567516C. *Molecular Breeding*, 2015-35:131, DOI: 10.1007/s11032-015-0330-5, 2015. citations 0.

Yan, Z., M.S. Hossain, S. Arikit, O. Valdes-Lopez, J. Zhai, J. Wang, M. Libault, T., Ji, L.-J. Qiu, B. Meyers, and G. Stacey. 2015. Identification of microRNAs and their mRNA targets during soybean nodule development: Functional analysis of the role of miR393-3p in soybean nodulation. *New Phytol*. 207 (3):748–759. citations 1.

Yan, Zhe; Hossain, Md Shakhawat; Valdés-López, Oswaldo; Hoang, Nhung T; Zhai, Jixian; Wang, Jun; Libault, Marc; Brechenmacher, Laurent; Findley, Seth; Joshi, Trupti; Qiu, Lijuan; Sherrier, D. Janine; Ji, Tieming; Meyers, Blake; Xu, Dong; Stacey, Gary (2015) Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. *Plant Biotech. Journal*. citations 1.

Yongqing Jiao, Tri D. Vuong, Yan Liu, Clinton Meinhardt, Yang Liu, **Trupti Joshi**, Perry B. Cregan, **Dong Xu**, J. Grover Shannon, **Henry T. Nguyen**. Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theor Appl Genet*. 128:15-23, 2015. citations 0.

CP Zeferino, KD Wells, ASAMT Moura, RA Murarolli, GE Rottinghaus, DR Ledoux. 2015. Gene expression in the kidneys of broilers fed ochratoxin A for different time periods. World Mycotoxin Journal, 1-12. citations 0.

Zhe Yan, Md Shakhawat Hossain, Oswaldo Valdes-Lopez, Nhung T. Hoang, Jixian Zhai, Jun Wang, Marc Libault1, Laurent Brechenmacher, Seth Findley, **Trupti Joshi**, Lijuan Qiu, D. Janine Sherrier, Tieming Ji, Blake C. Meyers, **Dong Xu** and **Gary Stacey**. Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. *Plant Biotechnol*. J., doi: 10.1111/pbi.12387, 1-10, 2015. citations 1.

Zhou A, Hillesland KL, He Z, Schackwitz W, Tu Q, Zane GM, Ma Q, Qu Y, Stahl DA, **Wall JD**, Hazen TC, Fields MW, Arkin AP, and Zhou J. 2015. Rapid selective sweep of pre-existing polymorphisms and slow fixation of new mutations in experimental evolution of *Desulfovibrio vulgaris*. *International Society for Microbial Ecology Journal 9(11):2360-72.* citations 0.

Yan Z, Hossain MS, Valdés-López O, Hoang NT, Zhai J, Wang J, Libault M, Brechenmacher L, Findley S, **Joshi T**, Qiu L, Sherrier DJ, Ji T, Meyers BC**, Xu D**, **Stacey G**. Identification and functional characterization of soybean root hair microRNAs expressed in response to Bradyrhizobium japonicum infection. 2015. *Plant Biotechnology Journal* (2015), pp. 1–10. doi: 10.1111/pbi.12387. citations 1.

2014 (36)

Agler C, Nielsen DM, Urkasemisin G, Singleton A, Tonamura N, Sigurdsson S, Tang R, Linder, K, Arepalli S, Hernandez D, LIndblad-Toh, K, van dee Leemput J, Matsinger-Reif A, **O'Brien DP**, Bell J. Harris T, Steinberg S, Olby NJ: Canine Hereditary Ataxia in Old English Sheepdogs and Gordon Setters is Associated with a Defect in the Autophagy Gene encoding RAB24. *PLOS Genetics* Feb 2014 6;10(2):e1003991 citations 2.

Almamun, M., Levinson, B.T., Gater, S.L., **Schnabel, R.D**., Arthur, G.L., Davis, J.W., and **Taylor, K.H.**, 2014. Genome-wide DNA methylation analysis in precursor B-cell. *Epigen*. 9(12):1588-95 PMID:25484143 citations 1.

Burke, M.K., **King, E.G**, Shahrestani, P, Rose, M.R., and Long, A.D. 2014. Genome-wide association study of extreme longevity in Drosophila melanogaster. *Genome Biology and Evolution*. 6(1): 1-11. citations 5.

Burns G, Brooks K, Wildung M, Navakanitworakul R, Christenson L, **Spencer TE**. Extracellular vesicles in luminal fluid of the ovine uterus. *PLoS One* 2014; Mar 10;9(3):e90913. citations 9.

Chettoor A.M., **Givan S.A.**, Cole R.A., Coker C.T., Unger-Wallace E., Vejlupkova Z., Vollbrecht E., Fowler J.E., Evans M.M. (2014). Discovery of novel transcripts and gametophytic functions via RNA-Seq analysis of maize gametophytic transcriptomes. *Genome Biology* 2014, 15:414 doi:10.1186/s13059-014-0414-2. citations 8

Creason, A.L., Vandeputte, O.M, Savory, E.A, Davis II, E.W., Putnam, M.L., Hu, E.a, Swader-Hines, D., Mol, A., Baucher, M., Prinsen, E., Zdanowska, M., **Givan, S.A**., El Jaziri, M., Loper, J.E., Mahmud, T., Chang, J.H. (2014). Analysis of genome sequences from plant pathogenic Rhodococcus reveals genetic novelties in virulence loci. *PLOS One*, DOI: 10.1371/journal.pone.0101996. citations 2.

Ellison, M., **G. C. Conant**, R. Cockrum, K. Austin, M. Becchi, H. Truong, W. Lamberson, and K. Cammack. (2014) Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Research*, **21**: 115-125. citations 2.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Deng J, Devreese B, Elhaik E, Evans JD, Foster LJ, Graur D, Guigo R; HGSC production teams, Hoff KJ, Holder ME, Hudson ME, Hunt GJ, Jiang H, Joshi V, Khetani RS, Kosarev P, Kovar CL, Ma J, Maleszka R, Moritz RF, Munoz-Torres MC, Murphy TD, Muzny DM, Newsham IF, Reese JT, Robertson HM, Robinson GE, Rueppell O, Solovyev V, Stanke M, Stolle E, Tsuruda JM, Vaerenbergh MV, Waterhouse RM, Weaver DB, Whitfield CW, Wu Y, Zdobnov EM, Zhang L, Zhu D, Gibbs RA; Honey Bee Genome Sequencing Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*. 2014 Jan 30;15:86. citations 49

Filant J, Lydon JP, **Spencer TE**. Integrated chromatin immunoprecipitation sequencing and microarray analysis identifies FOXA2 target genes in the glands of the mouse uterus. *FASEB J* 2014; 28:230-243. citations 4.

Franz, AWE, Sanchez-Vargas, I, Raban, RR, Black IV, WC, James, AA, Olson, KE. 2014. Fitness impact and stability of a transgene conferring resistance to dengue-2 virus following introgression into a genetically-diverse *Aedes aegypti* strain. *PLoS Neglected Tropical Diseases* 8(5): e2833. citations 9.

Guo J, **O'Brien DP**, Mhlanga-Mutangadura T, Olby NJ, **Taylor JF**, **Schnabel RD**, **Katz ML** and **Johnson GS**. 2015. A rare homozygous truncating MFSD8 mutation in the whole genome sequence of a Chinese Crested dog with neuronal ceroid lipofuscinosis. *BMC Veterinary Research 10:960*. citations 0.

Guo J, **Johnson GS**, Brown HA, Provencher ML, da Costa R, Mhlanga-Mutangadura T, **Taylor JF**, **Schnabel RD**, **O'Brien DP** and **Katz ML**. 2014. A CLN8 nonsense mutation in the whole genome sequence of a mixed breed dog with neuronal ceroid lipofuscinosis and Australian Shepherd ancestry. *Molecular Genetics and Metabolism 112(4):302-309*. citations 3.

Gilliam D, **O'Brien D**, **Coates J**, **Johnson GS**; **Johnson G**, Mhlanga-Mutangadura T, Hansen L, **Taylor JF** and **Schnabel RD**. 2014. A homozygous KCNJ10 mutation in jack russell terriers and related breeds with spinocerebellar ataxia with myokymia, seizures, or both. *Journal of Veterinary Internal Medicine 28(3):871-7*. citations 8.

Halley YA, SE Dowd, **JE Decker**, PM Seabury, E Bhattarai, D Rollins, B Ruzicka, IR Tizard, MJ Peterson, **JF Taylor** and CM Seabury. 2014. A draft de novo genome assembly for the Northern Bobwhite (Colinus virginianus) reveals evidence for a rapid decline in effective population size beginning in the late Pleistocene. *PLoS One* 9:e90240. citations 5.

Heidi M Appel, Howard Fescemyer, Juergen Ehlting, David Weston, Erin Rehrig, **Trupti Joshi**, **Dong Xu**, Joerg Bohlmann, **Jack Schultz**. Transcriptional responses of Arabidopsis thaliana to chewing and sucking insect herbivores. *Frontiers in Plant Science*. doi: 10.3389/fpls.2014.00565, 2014. citations 5.

Hillesland KL, Lim S, Flowers JJ, Turkarslan S, Pinel N, Zane GM, Elliott N, Qin Y, Wu L, Baliga NS, Zhou J, **Wall JD**, and Stahl DA. 2014. Erosion of functional independence early in the evolution of a microbial mutualism. *Proceedings of the National Academy of Sciences USA* 111(41):14811-7. citations 3.

Joshi T, Fitzpatrick MR, Chen S, Liu Y, Zhang H, Endacott RZ, Gaudiello EC, **Stacey G**, **Nguyen HT**, **Xu D**. Soybean Knowledge Base (SoyKB): A web resource for integration of soybean translational genomics and molecular breeding. *Nucleic Acids Research*. 2014. Jan;42(Database issue):D1245-52. doi: 10.1093/nar/gkt905. citations 13.

Keller KL, Rapp-Giles BJ, Semkiw ES, Porat I, Brown SD, and **Wall JD**. 2014. New model for electron flow for sulfate reduction in Desulfovibrio alaskensis G20. *Applied and Environmental Microbiology* 80(3):855-868. citations 8.

King, E.G., Kislukhin, G, Walters, K.N., and Long, A.D. 2014. Using Drosophila melanogaster to identify chemotherapy toxicity genes. *Genetics*. 198(1): 31 – 43. citations 1.

King, E.G., Sanderson, B.J., McNeil, C.L., Long, A.D., and Macdonald, S.J. 2014. Genetic Dissection of the Drosophila melanogaster Female Head Transcriptome Reveals Widespread Allelic Heterogeneity. *PLoS Genetics*. 10(5): e1004322. citations 5.

Korte HL, Fels SR, Christensen GA, Price MN, Kuehl JV, Zane GM, Deutschbauer AM, Arkin AP, and **Wall JD**. 2014. Genetic basis for nitrate resistance in Desulfovibrio strains. *Frontiers in Microbiology* 5:153 doi: 10.3389/fmicb.2014.00153. citations 7.

Langewisch T, Zhang H, Vincent R, **Joshi T**, **Xu D**, Bilyeu K. Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. *PLoS One*. 2014 Apr 11;9(4):e94150. doi: 10.1371/journal.pone.0094150. eCollection 2014. citations 4.

Li L., Eichten S.R., Shimizu R., Petsch K., Yeh C.T., Wu W., Chettoor A.M., **Givan S.A.**, Cole R.A., Fowler J.E., Evans M.M., Scanlon M.J., Yu J., Schnable P.S., Timmermans M.C., Springer N.M., Muehlbauer G.J. (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology*, doi:10.1186/gb-2014-15-2-r40. citations 21.

Maillo V, O'Gaora P, Forde N, Besenfelder U, Havlicek V, Burns GW, **Spencer TE**, Gutierrez-Adan A, Lonergan P, Rizos D. Oviduct-embryo interactions in cattle: two-way traffic or a one-way street? *Biol Reprod* 2014; 92(6):144. citations 0.

Marriage, T.N., **King, E.G**., Long, A.D., and Macdonald, S.J. 2014. Fine-mapping Nicotine Resistance Loci in Drosophila Using a Multiparent Advanced Generation Intercross Population. *Genetics*. 198(1): 45 – 57. citations 4.

O'Brien DP, Leeb T: DNA Testing in Neurologic Disease. Invited review, *Journal of Veterinary Internal Medicine* 2014;28(4):1186–1198. citations 1.

Siwaret Arikit, Rui Xia, Atul Kakrana, Kun Huang, Jixian Zhai, Zhe Yan , Oswaldo Valdés-López, Theresa A. Musket, Henry T. Nguyen, Gary Stacey, Blake C. Meyers. (2014) An atlas of soybean small RNAs demonstrates regulation by phased siRNAs of hundreds of coding genes. *Plant Cell* 26: 4584-4601. citations .

Michael D Roberts, Ryan G Toedebusch, Kevin D Wells, Joseph M Company, Jacob D Brown, Clayton L Cruthirds, Alexander J Heese, Conan Zhu, George E Rottinghaus, Thomas E Childs, Frank W Booth. Nucleus accumbens neuronal maturation differences in young rats bred for low versus high voluntary running behaviour. 2014. The Journal of Physiology 592 (10), 2119-2135. citations 4.

Spate, L.D., A. Brown, B.K. Redel, K.M. Whitworth, C.N. Murphy, **R.S. Prather**. 2014. Dickkopf-related protein 1 inhibits the WNT signaling pathway and improves pig oocyte maturation. *PLoS ONE*. doi: 10.1271/journal.pone.0095114. PMID: 24739947. citations 0.

Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, Gowin J, Gronenberg W, Hermansen RA, Hu H, Hunt BG, Huylmans AK, Khalil SM, Mitchell RD, Munoz-Torres MC, Mustard JA, Pan H, Reese JT, Scharf ME, Sun F, Vogel H, Xiao J, Yang W, Yang Z, Yang Z, Zhou J, Zhu J, Brent CS, Elsik CG, Goodisman MA, Liberles DA, Roe RM, Vargo EL, Vilcinskas A, Wang J, Bornberg-Bauer E, Korb J, Zhang G, Liebig J. Molecular traces of alternative social organization in a termite genome. *Nat Commun*. 2014 May 20;5:3636. citations 24.

Ryan G Toedebusch, Michael D Roberts, **Kevin D Wells**, Kayla M Kanosky, Jaume Padilla, Nathan T Jenkins, James W Perfield, Jamal A Ibdah, **Frank W Booth**, **R Scott Rector**. Unique transcriptomic signature of omental adipose tissue in Ossabaw swine: a model of childhood obesity. 2014. Physiological genomics 46 (10), 362-375. citations 1.

Trupti Joshi, Michael R. Fitzpatrick, Shiyuan Chen, Yang Liu, Hongxin Zhang, Ryan Z. Endacott, Eric C. Gaudiello, **Gary Stacey**, **Henry T. Nguyen**, **Dong Xu**. Soybean Knowledge Base (SoyKB): A web resource for integration

of soybean translational genomics and molecular breeding. *Nucleic Acids Research*. 42(1):D1245-D1252, 2014. citations 13.

Truyen N Quach, Hanh TM Nguyen, Babu Valliyodan, **Trupti Joshi**, **Dong Xu**, **Henry T Nguyen**. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. *Molecular Genetics and Genomics*. doi:10.1007/s00438-014-0978-2, 2014. citations 0.

Valdes-Lopez O, Khan SM, Schmitz RJ, Cui S, Qiu J, Zhu M, Cheng JJ, **Joshi T**, **Xu D**, Diers B, Ecker JR, **Stacey G**. Genotypic variation of gene expression during the soybean innate immunity response. *Plant Genetic Resources: Characterization and Utilization* (2014) 12(S1); S27–S30. doi:10.1017/S1479262114000197. citations 0.

William Yang, Kenji Yoshigoe, Xiang Qin, Jun S Liu, Jack Y Yang, Andrzej Niemierko, Youping Deng, Yunlong Liu, A Keith Dunker, Zhongxue Chen, Liangjiang Wang, **Dong Xu**, Hamid R Arabnia, Weida Tong and Mary Qu Yang. Identification of genes and pathways involved in kidney renal clear cell carcinoma. *BMC Bioinformatics*, 15(Suppl 17):S2, doi:10.1186/1471-2105-15-S17-S2, 2014. citations 0.

Zhao, M., J.J. Whyte, G.M. Hopkins, M.D. Kirk, **R.S. Prather**. 2014. Methylated DNA immunoprecipitation and high-throughput sequencing (MeDIP-seq) by using low amounts of genomic DNA. *Cellular Reprogramming* 16:175-184. doi: 10.1089/cell.2014.0002. Epub 2014 Apr 28. PMID: 24773292 citations 1.

2013 (22)

Chapple RH, **Wells KD**, **Givan SA**, **Kim JW**, McKay SD, **Schnabel RD**, **Taylor JF**. A characterization of the rat developmental liver transcriptome. 2013. *Physiological Genomics 45:301-311*. citations 7.

Ditzler M.A., Lange M.J., Debojit B., Bottoms C.A., Virkler K.F., Sawyer A.W., Whatley A.S., Spollen W., **GivanS.A.**, Burke D.H. (2013). High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Research*, doi:10.1093. citations 14

Fels SR, Zane GM, Blake SM, and **Wall JD**. 2013. Rapid transposon liquid enrichment sequencing (TnLE-seq) for gene fitness evaluation in underdeveloped bacterial systems. *Applied and Environmental Microbiology* 79:7510-7517. citations 5.

Hao Zhang, Zhi Li, Yanpu Li, Yuanning Liu, Junxin Liu, Xin Li, Tingjie Shen, Yunna Duan, Minggang Hu, **Dong Xu**. A computational method for predicting regulation of human microRNAs on the influenza virus genome. *BMC Systems Biology*. 7(Suppl 2):S3 doi:10.1186/1752-0509-7-S2-S3, 2013. citations 2.

Isom, S.C., J.R. Stevens, R. Li, W.G. Spollen, L. Cox, L.D. Spate, **R.S. Prather**. 2013. Transcriptional profiling of peri-attachment porcine embryos by RNA-Seq. *Physiological Genomics* 45:577-589. doi: 10.1152/physiolgenomics.00094.2012. Epub 2013 May 21. PMID: 23695885. citations 6.

Joshi T, Valliyodan B, Wu JH, Lee SH, **Xu D**, **Nguyen HT**. Genomic differences between cultivated soybean, G. max and its wild relative G. soja. *BMC Genomics*. 2013, 14(Suppl 1):S5. citations 4.

Kislukhin, G., **King, E.G**., Walters, K.N., Macdonald, S.J., and Long, A.D. 2013. The genetic architecture of methotrexate toxicity is similar in Drosophila melanogaster and humans. *G3: Genes|Genomes|Genetics* 3(8):1301-1310. citations 4.

Minten MA, Bilby TR, Bruno RGS, Allen CC, Madsen CA, Wang Z, Sawyer JE, Bauersachs S, Neibergs HL, Geary TW, **Spencer TE**. Effects of fertility on gene expression and function of the bovine endometrium. *PLoS One* 2013; 8(8): e69444. citations 9.

Morrill, B., L. Cox, A. Ward, S. Heywood, **R.S. Prather**, S.C. Isom. 2013. Targeted DNA methylation analysis by high throughput sequencing in porcine peri-implantation embryos. *Journal of Reproduction and Development* June 24;59(3):314-320. Epub 2013 Feb 22. PMID: 23428632. citations 2.

Parks JM, Johs A, Podar M, Bridou R, Hurt RA Jr, Smith SD, Tomanicek SJ, Qian Y, Brown SD, Brandt CC, Palumbo AV, Smith JC, **Wall JD**, Elias DA, and Liang L. 2013. The genetic basis for bacterial mercury methylation. *Science* 339(6125):1332-5. citations 95.

Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL, Armbruster PA. RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, Aedes albopictus. *J Exp Biol*. 2013 Nov 1;216(Pt 21):4082-90. doi: 10.1242/jeb.089508. citations 7.

Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL, Armbruster PA. Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, Aedes albopictus. *Proceedings of the Royal Society B: Biological Sciences*. 2013 Mar 20;280(1759):20130143. citations 18.

Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG, Armbruster PA. Transcriptome sequencing as a platform to elucidate molecular components of the diapause response in the Asian tiger mosquito, Aedes albopictus. *Physiol Entomol*. 2013 Jun;38(2):173-181. citations 6.

Anne Roulin, Paul Auer, Marc Libault, Jessica Schlueter, Andrew Farmer, Greg May, Gary Stacey, Rebecca W. Doerge and Scott A. Jackson (2013) The fate of duplicated genes in a polyploidy plant genome. *Plant Journal* 73: 143–153. citations 58.

Michael D Roberts, Jacob D Brown, Lauren P Oberle, Alexander J Heese, Ryan G Toedebusch, **Kevin D Wells**, Clayton L Cruthirds, John A Knouse, J Andries Ferreira, Thomas E Childs, Marybeth Brown, **Frank W Booth**. Phenotypic and molecular differences between rats selectively bred to voluntarily run high vs. low nightly distances. 2013. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 304 (22), 1024-1035. citations 11.

Schmitz RJ, He Y, Valdés-López O, Khan SM, **Joshi T**, Urich MA, Nery JR, Diers B, **Xu D**, **Stacey G**, Ecker JR. Ecker. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res*. 2013 Oct;23(10):1663-74. doi: 10.1101/gr.152538.112. citations 48.

Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L, Reese JT, Hunt BG, Graur D, Elhaik E, Kriventseva EV, Wen J, Parker BJ, Cash E, Privman E, Childers CP, Muñoz-Torres MC, Boomsma JJ, Bornberg-Bauer E, Currie CR, Elsik CG, Suen G, Goodisman MA, Keller L, Liebig J, Rawls A, Reinberg D, Smith CD, Smith CR, Tsutsui N, Wurm Y, Zdobnov EM, Berger SL, Gadau J. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res*. 2013 Aug;23(8):1235-47. doi: 10.1101/gr.155408.113. citations 55.

Sohel MH, Hoelker M, Nofersti S, Salilew-Wondim D, Tholen E, Looft C, Rings F, Uddin J, **Spencer TE**, Schellander K, Tesfaye D. Exosomal and non-exosomal transport of circulatory microRNAs in follicular fluid: Implications for bovine oocyte developmental competence. *PLoS One* 2013; 8(11):e78505. citations 23.

Wang, Z., Cao, R., **Taylor, K.H**., Briley, A., Caldwell, C.W., and Cheng, J., 2013. The properties of genome conformation and spatial gene interaction and regulation networks of normal and malignant human cell types. *PLoS ONE* 8(3):e58793. citations 7.

Yuanning Liu, Yaping Chang, Chao Zhang, Qingkai Wei, Jingbo Chen, Huiling Chen, **Dong Xu**. Influence of mRNA features on siRNA interference efficacy. *Journal of Bioinformatics and Computational Biology*. 11(3):1341004. doi: 10.1142/S0219720013410047, 2013. citations 2.

Yuanning Liu, Yaping Chang, **Dong Xu**, Zhi Li, Hao Zhang, Jie Li, Mingyao Tian. Optimized Design of siRNA Based on Multi-featured Comparison and Analysis of H1N1 Virus. *International Journal of Data Mining and Bioinformatics*. 7(4):345-57, 2013. citations 0.

Zhu, Mingzhu, Dahmen, Jeremy, Stacey, Gary, Cheng, Jianlin (2013) Predicting gene regulatory networks of soybean nodulation from RNA-seq transcriptome data. *BMC Bioinformatics* 14:278. citations 3.

2012 (13)

Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, **Schnabel RD**, Ventura M, **Taylor JF**, Garcia JF, Van Tassell CP, Sonstegard TS, Eichler EE and Liu GE. Individualized cattle copy number and segmental duplication maps using next generation sequencing. *Genome Research 22:778-790*. citations 59.

Chao Zhang, Shun-Fu Xu, and **Dong Xu**. Risk Assessment of Gastric Cancer Caused by Helicobacter pylori Using CagA Sequence Markers. *PLoS ONE*. 7(5): e36844. doi:10.1371/journal.pone.0036844, 2012. citations 6.

Givan S.A., Bottoms C.A., Spollen W.G. Computational Analysis of RNA-seq (2012). *Methods in Molecular Biology, RNA Abundance Analysis*, V883:201-219. citations 3

Jungmin Ha, Brian Abernathy, William Nelson, David Grant, Xiaolei Wu, Henry Nguyen, Gary Stacey, Rod Wing, Randy C. Shoemaker, and Scott A. Jackson (2012) Integration of the draft sequence and physical map as a

framework for genomic research in soybean (Glycine max (L.) Merr.) and wild soybean (Glycine soja Sieb. and Zucc). *Genes, Genomes and Genetics* 2: 321-329. citations 2.

Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, Wang Z, Libault M, Brechenmacher L, Valliyodan B, Wu X, Cheng J, **Stacey G**, **Nguyen HT**, **Xu D**. Soybean Knowledge Base (SoyKB): A Web Resource for Soybean Translational Genomics. *BMC Genomics* 2012, 13(Suppl 1):S15. citations 25.

King, E.G., Macdonald, S.J., and Long, A.D. 2012. Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics* 191(3): 935 – 949. (Cover article: http://www.genetics.org/content/191/3.cover-expansion) citations 33.

King, E.G., Merkes, C.M., McNeil, C.L., Hoofer, S.R., Sen, S., Broman, K.W., Long, A.D., and Macdonald, S.J. 2012. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Research* 22:1558 – 1566. citations 39.

Pei, L., Choi, J.H., Liu, J., Lee, E.J., McCarthy, B., Wilson, J.M., Speir, E., Awan, F., Arthur, G.L., Schnabel, J.L., **Taylor, K.H.**, Wang, X., **Xu, D**., Ding, H.F., Munn, D.H., Caldwell, C.W., and Shi, H., 2012. Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics* 1; 7(6):567-78. citations 33.

Redel, B.K., A.N. Brown, L.D. Spate, K.M. Whitworth, **J.A. Green**, **R.S. Prather**. 2012. Glycolysis in preimplantation development is partially controlled by the Warburg Effect. *Mol. Reprod. and Dev*.79:262-271. PMID: 22213464 citations 25.

Roeseler DA, Sachdev S, Buckley DM, **Joshi T**, Wu DK, **Xu D**, Hannink M, Waters ST. Waters. Elongation Factor 1 alpha1 and genes associated with Usher syndromes are down-stream targets of GBX2. *Plos One*. 2012; 7(11). citations 4.

Rowley E.R., Fox S.E., Bryant D.W., Sullivan C.M., **Givan S.A.**, Mehlenbacher S.A., Mockler T.C. (2012). Assembly and Characterization of the European Hazelnut 'Jefferson' Transcriptome. *Crop Science*, 52(6):2679-2686. citations 6.

Xue-Cheng Zhang, Maira C. M. Freire, Mi Ha Le, Glen Hartman, Greg Upchurch, Kerry Pedley, and Gary Stacey (2012) Genetic diversity and origins of Phakopsora pachyrhizi (soybean rust) isolates in the United States. *Asian J. Plant Pathology* 6: 52-65.

Mingzhu Zhu, Xin Deng, Trupti Joshi, Dong Xu, Gary Stacey, Jianlin Cheng (2012) Reconstructing differentially co-expressed gene modules and regulatory networks of soybean cells. *BMC Genomics* 13:437. citations 3.

2011 (12)

Farias FH, Zeng R, Wininger FA, **Taylor JF**, **Schnabel RD**, *McKay SD*, Sanders DN, Lohi H, Seppälä EH, Wade CM, Lindblad-Toh K, **O'Brien DP**, **Katz ML** and **Johnson GS.** 2011. A truncating mutation in ATP13A2 is responsible for adult-onset neuronal ceroid lipofuscinosis in Tibetan terriers. *Neurobiology of Disease 42:468-474*. citations 46.

Fisher CA, EK Bhattarai, JB Osterstock, SE Dowd, PM Seabury, M Vikram, RH Whitlock, Y Schukken, **RD Schnabel**, **JF Taylor**, JE Womack and CM Seabury. 2011. Evolution of the bovine TLR gene family and member associations with Mycobacterium avium subspecies paratuberculosis infection. *PLoS One* 6:e27744. citations 12.

Gyan Prakash Srivastava, Mamatha Hanumappa, Garima Kushwaha, **Henry T. Nguyen** and **Dong Xu**. PRIMEGENS-v2: Homolog-specific PCR primer design for profiling splice variants. *Nucleic Acid Research*. 39(10):e69, 2011. citations 4.

Guan Ning Lin, Chao Zhang, **Dong Xu**. Polytomy identification in microbial phylogenetic reconstruction. *BMC Genomics*. 5(Suppl 3):S2 doi:10.1186/1752-0509-5-S3-S2, 2011. citations 4.

Kimbrel J.A., **Givan S.A.**, Temple T.N., Johnson K.B. and Chang, J.H. (2011). Genome sequencing and comparative analysis of the carrot bacterial blight pathogen, Xanthomonas hortorum pv. carotae M081, for insights into pathogenicity and applications in molecular diagnostics. *Molecular Plant Pathology*, 12: no. doi: 10.1111/j.1364-3703.2010.00694.x citations 9.

Lee EJ, Pei L, Srivastava G, **Joshi T**, Kushwaha G, Choi JH, Robertson KD, Wang X, Colbourne JK, Zhang L, Schroth GP, **Xu D**, Zhang K, Shi H. Targeted bisulfite sequencing by solution hybrid selection and massively parallel sequencing. *Nucleic Acids Research* 2011;39(19). citations 24.

SM Rustemeyer, WR Lamberson, DR Ledoux, K Wells, KJ Austin, KM Cammack. Effects of dietary aflatoxin on the hepatic expression of apoptosis genes in growing barrows. 2011. Journal of Animal Science 89 (4), 916-925. citations 16.

Seabury CM, EK Bhattarai, **JF Taylor**, GG Viswanathan, SM Cooper, DS Davis, SE Dowd, ML Lockwood and PM Seabury. 2011. Genome-wide polymorphism and comparative analyses in the white-tailed deer (Odocoileus virginianus): A model for conservation genomics. *PLoS One*. 6:e15811. citations 14.

Shulaev V., et al. (2011). The genome of woodland strawberry (Fragaria vesca). *Nature Genetics*, doi:10.1038/ng.740. citations 319.

OswaldoValdés-López, Sandra Thibivilliers, Jing Qiu, Wayne Wenzhong Xu, Tran H.N. Nguyen, Marc Libault, Brandon H. Le, Robert Goldberg, Curt Hill, Glen Hartman, Brian Diers, and Gary Stacey (2011) Identification of quantitative trait loci controlling gene expression during the innate immunity response of soybean. *Plant Physiol*. 157: 1975-1986. citations 12.

Zeng R*, Farias F*, **Johnson G**, *McKay SD*, **Schnabel RD**, **Decker JE**, **Taylor JF**, Mann C, **Katz M**, **Johnson G**, **Coates Joan** and **O'Brien D**. 2010. A truncated retrotransposon disrupts the GRM1 coding sequence in Coton de Tulear dogs with Bandera's neonatal ataxia. *Journal of Veterinary Internal Medicine. 25:267-272*. citations 15.

Jixian Zhai, Dong-Hoon Jeong, Emanuele De Paoli, Sunhee Park, Benjamin D. Rosen, Yupeng Li, Alvaro J. González, Zhe Yan, Scott A. Jackson, Gary Stacey, Doug R. Cook, Pamela J. Green, D. Janine Sherrier, and Blake C. Meyers (2011) microRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes and Development* 25: 2540-2553. citations 161.

2010 (14)

Bauer, B.K., S.C. Isom, L.D. Spate, K.M. Whitworth, W.G. Spollen, S.M. Blake, **G.K. Springer**, C.N. Murphy, **R.S. Prather**. 2010. Transcriptional profiling by deep sequencing identifies differences in mRNA transcript abundance in in vivo vs. in vitro cultured porcine blastocyst stage embryos. *Biology of Reproduction* 83:791-798. PMID: 20668257 citations 27.

Choi, J.H., Li, Y., Guo, J., Pei, L., Rauch, T.A., Kramer, R.S., Macmil, S.L., Wiley, G.B., Bennett, L.B., Schnabel, J.L., **Taylor, K.H**., Kim, S., Xu, D., Sreekumar, A., Pfeifer, G.P., Roe, B.A., **Caldwell, C.W**., Bhalla, K.N., and Shi, H., 2010. Genome-wide DNA methylation maps in follicular lymphoma cells determined by methylation-enriched bisulfite sequencing. *PLoS ONE* 5, e13020. citations 22.

Fahlgren N., Jogdeo S., Kasschau K.D., Sullivan C.M., Chapman E.J., Laubinger S., Smith L.M., Dasenko M., **Givan S.A.**, Weigel D., Carrington J.C. (2010). MicroRNA Gene Evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* 2010 Apr;22(4):1074-1089. citations 115.

Farias FHG, **Johnson GS**, **Taylor JF**, Giuliano E, **Katz ML**, Sanders DN, **Schnabel RD**, *McKay SD*, Khan S, Gharahkhani P, O'Leary CA, Pettitt L, Forman OP, Boursnell M, McLaughlin B, Ahonen S, Lohi H, Hernandez-Merino E, Gould DJ, Sargan DR and Mellersh C. 2010. An ADAMTS17 Splice Donor Site Mutation in Dogs with Primary Lens Luxation. *Investigative Ophthalmology and Visual Science 51:4716-4721*. citations 33.

Isom, S.C., W.G. Spollen, S.M. Blake, B.K. Bauer, **G.K. Springer**, **R.S. Prather**. 2010. Transcriptional profiling of day 12 porcine embryonic disc and trophectoderm samples using ultra-deep sequencing technologies. *Molecular Reproduction and Development* 77:812-819. PMID: 20722012 citations 12.

Joshi T, Yan Z, Libault M, Jeong DH, Park S, Green PJ, Sherrier DJ, Farmer A, May G, Meyers BC, **Xu D**, **Stacey G**. Prediction of novel miRNAs and associated target genes in Glycine max. *BMC Bioinformatics*, 11(Suppl 1):S14, 2010. citations 57.

Kimbrel J.A., **Givan S.A.**, Halgren A.B., Creason A.L., Mills D.I., Banowetz G.M., Armstrong D.J., Chang J.H. (2010). An improved, high-quality draft genome sequence of the Germination-Arrest Factor-producing Pseudomonas fluorescens WH6. *BMC Genomics* 2010, 11:522. citations 28.

Libault M, Farmer A, **Joshi T**, Takahashi K, Langley RJ, Franklin LD, He J, **Xu D**, May G, **Stacey G**. An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *Plant J.* 2010 Jul 1;63(1):86-99. citations 152.

Packer R, Patterson E, **Taylor JF**, **Coates J**, **Schnabel R** and **O'Brien D**. 2010. Characterization and mode of inheritance of a paroxysmal dyskinesia in Chinook dogs. *Journal of Veterinary Internal Medicine 24:1305-1313.* citations 6.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T,Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, **Shu S**, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, **Joshi T**, Libault M, Sethuraman A, Zhang XC, Shinozaki K, **Nguyen HT**, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, **Stacey G**, Shoemaker RC, Jackson SA. Genome Sequence of the Palaeopolyploid Soybean. *Nature*. 463:178-83, 2010. citations 1180.

Seabury CM, PM Seabury, **JE Decker**, **RD Schnabel**, **JF Taylor** and JE Womack. 2010. Diversity and evolution of 11 innate immune genes in Bos taurus taurus and Bos taurus indicus cattle. *Proc. Natl. Acad. Sci. USA*. 107:151-6. citations 30.

Vogel J.P., Bevan M., et al. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463, 763-768. citations 675.

Wu X, Ren C, **Joshi T**, Vuong T, Xu D, **Nguyen HT**. SNP discovery by high-throughput sequencing in soybean. *BMC Genomics* 2010, 11:469. citations 47.

Dongyu Zhao, Yan Wang, Di Luo, Xiaohu Shi, Liupu Wang, **Dong Xu**, Jun Yu, Yanchun Liang. PMirP: A pre-microRNA prediction method based on structure-sequence hybrid features. *Artificial Intelligence in Medicine*. 49(2):127-132. 2010. citations 18.

2009 (9)

Thibivilliers S, Joshi T, Campbell KB, Scheffler B, Boerma R, Xu D, Cooper B, Nguyen HT, Stacey G (2009) EST sequencing of Phaseolus vulgaris genes responsive to Uromyces appendiculatus infection also identifies soybean genes responding to Phakopsora pachyrhizi infection. *BMC Plant Biology* 9:46.

Navdeep Gill, Seth Findley, Jason G. Walling, Christian Hans, Jianxin Ma, Jeff Doyle, Gary Stacey, and Scott Jackson (2009) Molecular and chromosomal evidence for allopolyploidy in soybean Glycine max (L.) Merr. *Plant Physiology* 151: 1167-1174. citations 68.

Marc Libault, Trupti Joshi, Kaori Takahashi, Andrea Hurley-Sommer, Kari Puricelli, Sean Blake, Dong Xu, Henry Nguyen, and Gary Stacey (2009) Large scale analysis of soybean regulatory gene expression identifies a Myb gene involved in nodule development. *Plant Physiol*. 151: 1207-1220. citations 30.

Libault, Marc, Andrew Farmer, Laurent Brechenmacher, William L. Franck, Jenny Drnevich, Raymond J. Langley, Damla D. Bilgin, Osman Radman, Steven J. Clough, Gregory May and Gary Stacey (2009) Complete transcriptome of the soybean root hair cell, a single cell model, and its alteration in response to Bradyrhizobium japonicum infection. *Plant Physiol*. 152: 541-552. citations 80.

Guan Ning Lin, Zhipeng Cai, Guohui Lin, Sounak Chakraborty, and **Dong Xu**, ComPhy: Prokaryotic Composite Distance Phylogenies Inferred from Whole-Genome Gene Sets, *BMC Bioinformatics*, 10(Suppl 1):S5doi:10.1186/1471-2105-10-S1-S5, 2009. citations 10.

Fahlgren N., Kasschau K.D., Chapman E.J., Sullivan C.M., Cumbie J.S., Montgomery T., Gilbert S.D., Dasenko M., **Givan S.A.** and Carrington J.C. (2009). Computational and Analytical Framework for Small RNA Profiling by High-Throughput Sequencing. *RNA* 15(5):992-1002. citations 71.

Filichkin S.A., Priest H.D., **Givan S.A.**, Shen R., Bryant D.W., Fox S.E., Wong W-K, Mockler T.C. (2009). Deep transcriptome sequencing reveals a complex network of alternative splicing in plants. *Genome Research* 10.1101/gr.093302.109. citations 343.

Ramos AM, Crooijmans RPMA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, **Schnabel RD**, Van Tassell CP, **Taylor JF**, Wiedmann RJ, Schook LB and Groenen MAM. 2009. Design of a

high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE 4:e6524*. citations 232.

Sandra Thibivilliers, **Trupti Joshi**, Kimberly B. Campbell, Brian Scheffler, **Dong Xu**, Bret Cooper, **Henry T. Nguyen** and **Gary Stacey**. Generation of Phaseolus vulgaris ESTs and investigation of their regulation upon Uromyces appendiculatus infection. *BMC Plant Biology* 9:46 (27 Apr 2009). citations 20.

2008 (6)

Laurent Brechenmacher, Moon-Young Kim, Jijun Zou, Marisol Benitez, Min Li, Crystal B. McAlvin, Trupti Joshi, Bernarda Calla, Mei Phing Lee, Reena Philip, Marc Libault, Lila O. Vodkin, Dong Xu, Suk-Ha Lee, Steven J. Clough, Gary Stacey. (2008) Transcription profiling of soybean supernodulation by Bradyrhizobium japonicum. *Mol. Plant-Microbe Int.* 21: 631-645. citations 13.

Gyan P. Srivastava, Juyuan Guo, Huidong Shi, and **Dong Xu**. PRIMEGENS-v2: Genome-wide primer design for analyzing DNA methylation patterns of CpG islands. *Bioinformatics*, 4(17):1837-42, 2008. citations 13.

Michael T.P., Mockler T.C., Breton G., McEntee C., Byer A., Trout J.D., Hazen S.P., Shen R., Priest H.D., Sullivan C.M., **Givan S.A.**, Yanovsky M., Hong F., Kay S.A., Chory J. (2008). Network discovery pipeline elucidates conserved time of day specific cis-regulatory modules. *PLoS Genetics* 4(2): e14. doi:10.1371/journal.pgen.0040014 citations 200.

Mockler T.C., Michael T.P., Priest H.D., Shen R., Sullivan C.M., **Givan S.A.**, McEntee C., Kay S.A., Chory J. (2008). DIURNAL: The Diurnal Project: Diurnal and Circadian Expression Profiling, Model-based Pattern Matching, and Promoter Analysis. *Cold Spring Harbor Symposia on Quantitative Biology* 2007 72:1, 353-363. citations 105.

Randy Shoemaker, David Grant, Terry Olson, Wesley C. Warren, Rod Wing, Perry Cregan, Bindu Joseph, Montona Futrell-Griggs, Will Nelson, Jon Davito, Jason Walker, John Wallis, Colin Kremitski, Debbie Scheer, Sandy Clifton, Tina Graves, Henry Nguyuen, Xiaolei Wu, Mingcheng Luo, Jan Dvorak, Steve Cannon, Jeff Thomkins, Jeremy Schmutz, Gary Stacey and Scott Jackson (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. *Genome* 51: 294-302. citations 30.

Van Tassell CP, Smith TPL, Matukumalli LK, **Taylor JF**, **Schnabel RD**, Lawley CT, Haudenchild CD, Moore SS, Warren WC and Sonstegard TS. 2008. Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries. *Nature Methods. 5:247-252 cover article*. citations 319.

2007 (6)

Fahlgren N., Howell M.D., Kasschau K.D., Chapman E.J., Sullivan C.M., Cumbie J.S., **Givan S.A.**, Law T.F., Grant S.R., Dangl J.L., Carrington J.C. (2007). High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. *PLoS ONE*. 2, e219. citations 513.

Howell M.D., Fahlgren N., Chapman E.J., Cumbie J.S., Sullivan C.M., **Givan S.A**., Kasschau K.D., Carrington J.C. (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. *The Plant Cell* 19, 926-942. citations 170.

James O. Allen, Christiane M. Fauron, Patrick Minx, Leah Westgate, Swetha Oddiraju, Guan Ning Lin, Louis Meyer, Hui Sun, Kyung Kim, Chunyan Wang, Feiyu Du, **Dong Xu**, Michael Gibson, Jill Cifrese, Sandra W. Clifton, and Kathleen J. Newton. Comparisons among Two Fertile and Three Male-Sterile Mitochondrial Genomes of Maize. *Genetics*. 177:1173-1192, 2007. citations 80.

Kasschau K.D., Fahlgren N., Chapman E.J., Sullivan C.M., Cumbie J.S., **Givan S.A.**, Carrington J.C. (2007). Genome-Wide Profiling and Analysis of Arabidopsis siRNAs. *PLoS Biology* 5, e57. citations 176.

Rajesh Kumar, Jing Qiu, **Trupti Joshi**, Babu Valliyodan, **Dong Xu**, **Henry T. Nguyen**. Single Feature Polymorphism Discovery in Rice. *PLoS ONE* 2(3): e284. doi:10.1371/journal.pone.0000284, 2007 citations 45.

Taylor, K.H., Kramer, **R.S., Davis**, J.W., Guo, J., Duff, D.J., **Xu, D**., **Caldwell, C.W**., and Shi, H., 2007. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res*. 67, 8511-8518. **[Article featured on cover].** citations 160.

2006 (4)

Andrew Nunberg, Joseph A. Bedell, Mohammad A. Budiman, Robert W. Citek, Sandra W. Clifton, Lucinda Fulton, Deana Pape, Zheng Cai, **Trupti Joshi**, **Henry Nguyen**, **Dong Xu** and **Gary Stacey**. Survey sequencing of soybean elucidates the genome structure and composition. *Functional Plant Biology*. 33(8):765–773, 2006. citations 6.

Xiufeng Wan, Guohui Lin and **Dong Xu**. Rnall: An Efficient Algorithm for Predicting RNA Local Secondary Structural Landscape in Genomes. *Journal of Bioinformatics and Computational Biology*. 4:1015-1031, 2006. citations 10.

Xiu-Feng Wan, Jizhong Zhou, and **Dong Xu**. CodonO: a new informatics method measuring synonymous codon usage bias. *International Journal of General Systems*. 35:109-125, 2006. citations 5.

Xiaomeng Wu, Xiu-Feng Wan, Gang Wu, **Dong Xu**, and Guohui Lin. Phylogenetic Analysis Using Complete Signature Information of Whole Genomes and Clustered Neighbor-Joining Method. *International Journal of Bioinformatics Research and Applications*. 2:219-248, 2006. citations 35.

2005 (2)

Xiu-Feng Wan and **Dong Xu**. Intrinsic terminator prediction and its application in Synechococcus sp. WH8102. *Journal of Computer and Science Technology*. 20: 465-482, 2005. citations 7.

Xiu-Feng Wan, Tao Ren, Kai-Jian Luo, Ming Liao, Gui-Hong Zhang, Jin-Ding Chen, Wei-Sheng Cao, Ye Li, Ning-Yi Jin, **Dong Xu** and Chao-An Xin. Genetic characterization of H5N1 Avian Influenza Viruses isolated in Southern China during the 2003-04 avian influenza breakouts. *Archives of Virology*. 150(6):1257-1266, 2005. citations 33.

2004 (1)

Xiu-Feng Wan, **Dong Xu**, Andris Kleinhofs, and Jizhong Zhou. Quantitative relationship between codon usage bias and GC composition across the unicellular genomes. *BMC Evolutionary Biology*, **4**:19, 2004 (http://www.biomedcentral.com/1471-2148/4/19/). citations 63.