

Progress Report

DBI-0527192: Sequencing the Maize Genome

PI: Richard K. Wilson

Co-PIs: Doreen Ware, Rodney Wing, W. Richard McCombie, Patrick Schnable

Key Personnel: Sandra W. Clifton, Srinivas Aluru, Lincoln Stein, Robert Martienssen

Overview of original project goals:

Project Objectives:

1. Provide the complete sequence and structures of all maize genes and their locations (in linear order) on both the genetic and physical maps of maize.
2. The gene space of B73 maize (gene sequences and adjacent regulatory regions) should be finished to high quality according to currently acceptable standards
3. If applicable, the sizes of gaps between the genes should be estimated and draft sequences of repetitive DNA between genes presented where possible.
4. The sequence will be fully integrated with the genetic and physical maps.
5. Annotation will include gene models, predicted exon/intron structure, incorporation of EST and full-length cDNA data, gene ontology, and relationship with homologs in other organisms, including but not limited to, the other sequenced plant genomes.
6. Annotation will be coordinated with existing maize community and comparative databases with the eventual goal of generating complete curation of the genomic sequences to a standard set by established model organism databases.

Overview of implemented and proposed changes to project goals:

None this year and none anticipated.

Research Activities and Results.

*1. Choosing minimal tiling path of mapped BAC clones and preparing DNA for sequencing.
Responsible PI: R. Wing (AGI)*

Goals: Minimal tiling paths are to be chosen by AGI in consultation with the GSC. The AGI will purify DNA from the selected clones and verify their identity by generating end sequences for each clone. Once confirmed, the DNA will be shipped to the GSC for library construction and sequencing.

Progress: AGI has the following responsibilities for the Maize Genome Sequencing Project: a) Picking a MTP across the maize genome and maintenance of the physical map; b) Clone validation, preparation of sheared BAC DNA and delivery of such DNA to GSC for shotgun library construction; and c) Sequence improvement of 18% of the maize clones selected for finishing. Items 1 and 2 are addressed here. Item 3 is addressed below in the sequence improvement section.

a) Picking a MTP across the maize genome and maintenance of the physical map. As shown in the following flowchart, we have developed and optimized the pipeline from MTP clone selection to BAC library construction.

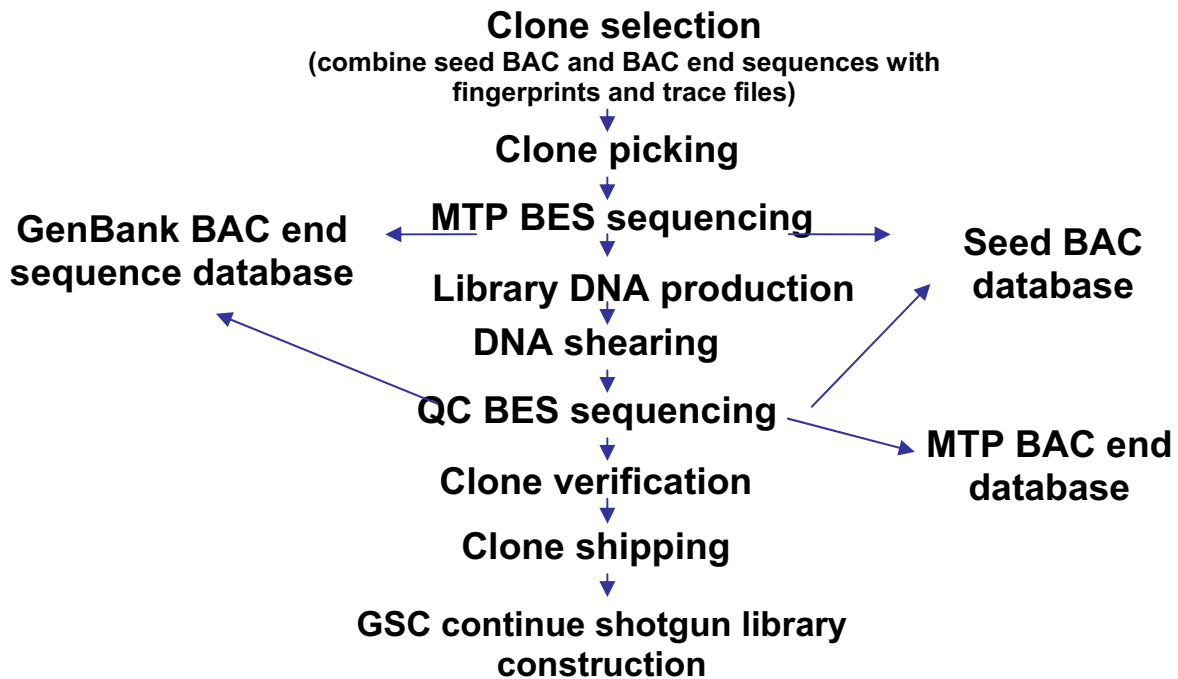


Figure 1: Flowchart for MTP picking and Library Construction

In clone selection, to minimize clone overlap, we employ a seed BAC approach. Seed BACs are selected 800 Kb apart. Then, we perform further walking steps based on seed BAC sequences. In collaboration with CSHL, we have developed a MTP pipeline for this walking process as shown in Figure 2. This pipeline combines all the data from physical map, blastn search of seed BAC against a BES database, and trace display. The pipeline makes the time-consuming selection process of MTP BACs simple and error-proof.

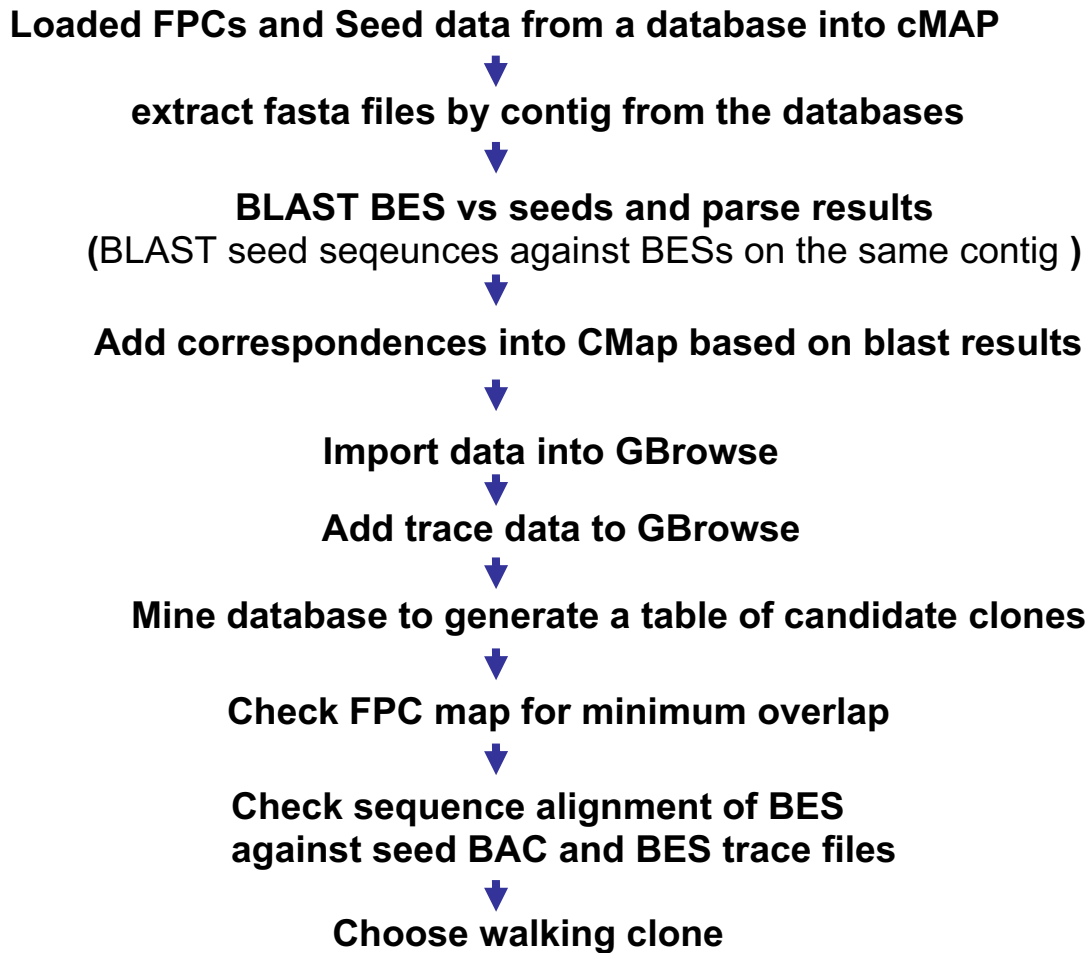


Figure 2: MTP Pipeline Developed by AGI and CSHL

Using these pipelines, we selected 3,400 seed BACs, approximately 800 kb apart along all 10 chromosomes. These seed BACs serve as the basis for further chromosome walking. We also chose 600 BACs to replace partially-sequenced BACs deposited in GenBank to ensure all necessary reagents were available for subsequent sequence improvement. In addition, we selected 200 BACs in a 25-Mb contig as a pilot project that we plan to publish soon that will demonstrate the type and quality of sequence that will be generated for the whole maize genome. Finally, we selected 4,100 walking BACs from sequenced seed BACs. In total, we have selected 8,300 BACs for the sequencing project. Of the selected clones, DNA of 7,800 clones have been prepared and sheared for shotgun library construction.

Our initial estimate of 19,000 MTP clones still stands but we anticipate that number will be reduced to approximately 17,000 MTP clones due to the fact that we selected very large initial seed BACs. We should have a much better estimate of the actual number of MTP BACs by the end of the March 2007.

b) Clone validation, preparation of sheared BAC DNA and delivery of such DNA to GSC for shotgun library construction.

Figure 1 outlines the process of MTP picking and clone processing. To scale up our clone processing pipelines and make them more efficient, AGI and GSC divided the shotgun library production process, with the clone validation, DNA preparation and DNA shearing being performed at AGI. Briefly, once a MTP BAC is picked, the BAC ends are sequenced to determine if the clone is correct. The clone is then prepped and sheared. Our original DNA preparation protocol used a 50ml culture and was very labor intensive and time consuming and was used to prep several hundred seed BACs. Dr. HyeRan Kim developed a new “low volume prep” using 3 ml cultures that was scalable and saved a tremendous amount of time and labor. Sheared DNA derived from this new prep was shipped to GSC and tested in their shotgun library construction group. The tests were very positive so AGI is now using this new prep exclusively to prepare sheared BAC DNA for the project. In addition, the protocol was also transferred to GSC and is now working in their hands and can be used as a backup if necessary.

Once the DNA is sheared and precipitated, we sequence the BAC end one more time to check for any errors in clone handling before the DNA is shipped to GSC.

c) Sequence improvement is just underway at AGI.

Plans: AGI is on target to select and process an additional 960 clones/month until the entire maize genome is covered with BACs. We will continue to pick MTP BACs until we have covered the maize genome that is contained within the physical map. We will begin and implement procedures to fill gaps between contigs and MTP clones where we were unable to obtain overlapping BACs. In addition, AGI will continue to validate, prepare sheared BAC and deliver said DNA to WUGSC for shotgun library construction.

2. BAC sub-clone libraries, production sequencing and assembly, “genespace” boundary determination, automated sequence improvement, and shared finishing of “genespace” in BAC clones. Responsible PI: R. Wilson, GSC

Goals: Provide the complete sequence and structures of all maize genes. The gene space of B73 maize (gene sequences and adjacent regulatory regions) should be of finished quality according to currently acceptable standards. If applicable, the sizes of gaps between the genes should be estimated and draft sequences of repetitive DNA between genes presented where possible.

Progress: The first 4-6 months of the past grant year were spent optimizing procedures and improving computer access and communication among the three Centers (GSC, AGI and CSHL) sharing production and finishing tasks. Bimonthly or monthly conference calls among the three centers involved in the production/finishing procedures have been and are continuing to be used as an effective means of anticipating and dealing with any problems in a timely manner. A smoothly operating protocol is now in place and functioning well. The process is as follows.

A. DNA intake. The data entry process is initiated upon receipt of sheared DNA from AGI. Trays of 96 sheared bacterial artificial chromosomes (BACs) processed by AGI are received by

the GSC and checked into the LIMS database for tracking purposes. They are entered based on position in the physical map and tracked throughout the remaining lab procedures in this system.

B. Library construction. Once entered, the library construction process is initiated. The fragmented pieces of DNA are ligated into sequencing vector and plated to determine the titer of each library. Once a successful library is established, the clone is labeled as “library_done”, in the Oracle database and is moved to the production group for sequence generation.

C. Production process.

i. The plating group produces one large agar plate is for the selection of 2 384-well trays of random subclones. Once the trays are produced, they are entered into the sequencing pipeline, which includes template preparation, and sequence generation on the Applied Biosystems 3730XL sequencing platform. The current production metrics average approximately 90% success rate and 650 bp of phred 20 base quality, or less than one error per 100 bp of raw sequence. These data are then assembled utilizing the assembly algorithm phrap (Green), and this process yields data that range from 4X to 6X statistical coverage across each BAC clone. Upon completion of this process, the projects are classified as “shotgun_done” in the Oracle database.

ii. The first stage of assembled sequence is deposited in the public data repository, National Center for Biotechnology Information (NCBI). The raw sequence trace files also are deposited to the trace archive of NCBI within 24 hours of production.

iii. When projects reach the “shotgun_done status, they proceed to the pre-finishing process where automated directed improvement techniques are employed to improve the overall sequence quality of the projects. Clones are compared to predicted neighbors to ensure a contiguous tiling path, and data is shared between overlapping projects to efficiently utilize all available data. Miss-assemblies are resolved using software tools designed to resolve those regions of the clones. Regions deemed “do finish” are identified based on the goal of the project to improve all gene regions. The process of identifying “do finish” regions has been an ongoing effort in collaboration with CSHL. The initial tagging process utilized a repeat library to identify regions to be finished, but this process was identified early to be less effective than hoped in screening repetitive regions of the genome. Through collaboration with CSHL, a new process for unique sequence identification, described in detail in the CSHL progress report, was developed. This process was developed and thoroughly tested with interaction between the participating centers. Once the clones are sorted and “do finish regions are identified, a software package called autofinish (Gordon) is used to select directed reactions designed to close all gaps and resolve low quality regions in areas deemed necessary for finishing based on the CSHL tagging method. Each project rotates through two cycles of pre-finishing, and then is labeled in the Oracle database as “prefin_done”, and the improved assembly is resubmitted to NCBI.

iv. At this stage, BAC projects are queued for manual improvement (finishing), which involves the complete resolution of all sequences deemed “do finish” to a standard of less than one error per 10,000 bp. The sequence improvement process effort is divided among AGI, CSHL, and the GSC. All wet lab activities and data storage reside at the GSC with data manipulation efforts being divided among the three sites. All sites conduct improvement

activities on the GSC computer system with remote login methods used at AGI and CSHL. Although the GSC has used remote login efforts for finishing in prior projects, there was significant development required to make this process more efficient on the scale, and complexity of this project. With a wide range of operating systems and remote login locations necessary, compatibility and performance issues were problematic at the beginning of the project. Through collaborations with the participants computer systems staffs and software development groups, most performance and compatibility issues have been overcome. Although significant development was required, the ability to operate on one, well established, efficient system, with one consistent Oracle database LIMS tracking, is now invaluable. Once the manual improvement process begins, regardless of location of the finisher, projects are resubmitted to NCBI at the initiation of the manual improvement efforts with a classification of “active_fin” to indicate that manual efforts are underway. Upon completion of the resolution of all “do finish” regions, the projects are subjected to a rigorous quality control process to ensure that all necessary regions are improved to the appropriate standard. Clones passing this quality control step are deemed “improved”, and submitted to NCBI as “improved”. At this stage, the unique regions of each clone are resolved to high quality sequence, order and orientation is provided where possible for all contigs, and the sequence is sent to the Cold Spring Harbor Laboratory for annotation, further refinement of order, orientation and association of marker information, and incorporation with physical and genetic map information.

Progress: Thus far at the GSC, 4,851 clones have progressed through the library_done stage, 3,296 projects are shotgun_done, 2,630 are prefin_done, and 511 are improved. Although the throughput has not met our original year one goals, the development necessary to achieve final project success has been successful, and the throughput rate has increased significantly, with throughput numbers now on track to meet the project goals. Clone selection has increased to 1,000 clones per month, with library construction, and production sequencing throughput scaling proportionally. Although there is a significant difference between the 511 projects submitted as improved compared to the anticipated 4,750 clones by the end of year 1, there are many indications that this process too, is now ready for significant scale-up and the throughput necessary for attaining the ultimate goals of the project is now in place. With the development of appropriate “do finish” identification, systems performance improvements for offsite finishers, streamlined tracking and communication processes, and clear improvement rules and standards in place, the improvement pipeline is ramping significantly. Each site has increased staff to meet the demand for completion of the project.

Plans: The plans for the remaining years of the project are to employ the systems developed during the first year of the project. The tools and capacity to complete the project are in place, and the throughput of the pipelines will be closely monitored and weakness in the system will continue to be scrutinized and improved to ensure success of the project.

3. Sequence improvement of “genespace” at CSHL. Responsible PI: W. McCombie, CSHL

Progress: The CSHL group working on sequence improvement spent considerable time and effort in year one working with the GSC and AGI to establish the information and systems required for rapid and efficient sequence improvement. This was somewhat more complex than

with other sequencing projects, since we were to improve only the non-repetitive portions of the sequence.

The initial clones that were received were deliberately “over finished” to provide a benchmark to which we could compare our improved products. As a result of this exercise, it was apparent that the software program was tagging more of the sequence as “to be improved” than necessary. We worked extensively with the GSC and the Ware group at CSHL to test various approaches for identification of regions to finish and subsequent results analysis to determine the effectiveness of each approach. These endeavors took most of the first 6 months of the initial grant year. This is described in more detail in the informatics section of this report. In May we began receiving substantial numbers of clones and had a first estimate of how well, and how much of each to improve. We worked at this level until July and then met with the GSC and AGI in St. Louis to improve a software program for tagging sequence regions to improve. The standards were agreed upon, and at that point our sequence improvement rate increased rapidly.

Once the process was stabilized we began training additional finishers. We now have three finishers producing at a rate of about 40-50 BACs per month. Two additional finishers should reach that rate within one-two months. An additional finisher is beginning training, and this additional help should bring us to the required rate.

Our cost for the year is an average of about \$1,390 per clone. However, in looking at our cost per clone per quarter, a dramatic decrease can be seen in the last 2 quarters. The cost in the most recent quarter of the year is \$210. The cost for the most recent month is even lower, at \$176 per clone. We expect that downward trend to continue until our target is reached or exceeded.

Plans: Now that the systems are in place for sequence improvement we will continue increasing our throughput until we are on target to finish our assigned fraction of the genome on schedule. In addition to our already trained finishers, we have two finishers who are increasing their experience and throughput. They should be at maximum capacity within two months.

We will also continue to work with both the GSC team and the Ware group at CSHL to improve the software and tagging. This will include constant monitoring and evaluation of tagging results versus final annotation, as well as enhancing scripts for automating the quality control process. The QC process is still a large time sink, and it is a major target for automation in the coming year in conjunction with the GSC.

4. Sequence improvement of “genespace” at AGI. Responsible PI: R. Wing, AGI

Progress: AGI is responsible for on-line finishing genespace of 18% of pre-finished BACs at GSC. We initiated maize finishing on May, 2006, since then we pre-submitted 171 BACs to GSC out of 435 BACs assigned to AGI for finishing, and 264 BACs are in various progresses in finishing.

Due to the heavy training for project work-bench and other relevant software for on-line finishing, the overall progress was slow in the first 3 months (May-Jul. 2006). However, AGI is

overtaking our original goal. At the same time, AGI hired two additional finishers (now a total 6 finishers) on Oct. 2006 and they are currently in training. From December 2006, we anticipate that all six finishers will be involved in finishing up to 90 maize BACs/month as a mid-term goal.

Plans: We will reach our goal of 130 BACs/month finishing rate within the next 6 months.

5. Bioinformatics

Responsible PI: Doreen Ware, CSHL; Key Personnel: Lincoln Stein, Patrick Schnable, Srinivas Aluru

Goals: The sequence will be fully integrated with the genetic and physical maps. Annotation will include gene models, predicted exon/intron structure, incorporation of EST and full-length cDNA data, gene ontology, and relationship with homologs in other organisms, including but not limited to, the other sequenced plant genomes. Annotation will be coordinated with existing maize community and comparative databases with the eventual goal of generating complete curation of the genomic sequences to a standard set by established model organism databases.

Progress:

A. BAC Tiling Path Clone Selection

Selecting clones in the minimal tiling path began as a largely manual process that soon became a rate-limiting step in the Project. The CSHL and AGI groups worked together to automate this step and eliminate the bottleneck. The process involves comparing the previously sequenced clone (the seed) to BAC-end sequences of clones located nearby within the FPC contig. Clones that overlap the seed BAC, based on the BLAST alignment, are candidates for the next clone to be sequenced. Other criteria, such as overlap on the FPC contig and clone size, are subsequently taken into consideration for the final selection of the next clone.

Increasing throughput was accomplished by automating the data pipeline and setting up visualization methods using existing GMOD tools, CMap and GBrowse (<http://www.gmod.org/node/>). The pipeline uses BLAST to compare the seed BAC sequence with the BAC end sequences, thereby creating correspondences between them. The results are filtered such that BACs that are completely contained within the seed BAC or that are too far away from the seed BAC on the FPC contig are omitted. These results, together with the FPC map, are used to populate the CMap viewer. The BLAST results are also used to populate the GBrowse viewer, along with the trace data such that sequence quality can be simultaneously evaluated by the user. Next, a set of overview web pages are created to direct the scientist to the most useful views for each seed BAC. After the pipeline runs, the CMap view shows the FPC contig and seed BAC along with its correspondences to candidate clones. This allows the scientist to visualize the location of candidate clones within the FPC contig. This view links directly to the BLAST alignment view in Gbrowse. Here, scientists can evaluate the alignment and the trace data together to determine whether the alignment is real as opposed to a spurious one caused by repetitive or low-quality sequences.

The new streamlined process has dramatically reduced the time required for scientists to choose next clone to push into sequencing, dropping from 20 minutes to a mere 5 minutes.

B. Annotation

The goal of annotation is to identify and characterize biologically relevant features using both comparative and predictive approaches. The first year of this endeavor has focused on the development of protocols that will form the basis of the annotation pipeline. This includes testing of gene prediction and alignment tools, parameter optimization, and selection and acquisition of datasets for comparative analyses. For method evaluation, we have relied upon a set of 100 sequenced BAC clones that were selected at random to be representative of the maize genome (NSF #0211851). We have also used the annotation of these BACs provided by the Munich Information Center for Protein Sequences (MIPS) as a reference (Haberer et al. 2005. *Plant Physiology* 139:1612-1624). Testing has also been performed on an expanded set of 574 additional BAC sequences, including a sample of 135 clones sequenced by GSC. While some analysis protocols can be directly adopted from the Gramene Project (www.gramene.org) with little modification, new approaches are required to address characteristics that are specific to the maize genome.

Repetitive sequences, composed predominantly of long terminal repeat (LTR)-retrotransposons, make up ~66% of the genome. These complicate the detection of low-copy number genes and cause spurious alignments. It was therefore a high priority to define best practices for detecting and masking repeat sequences. We have evaluated repeat libraries from both Iowa State University (<http://magi.plantgenomics.iastate.edu/repeatdb.html>) and TIGR (http://maize.tigr.org/repeat_db.shtml) using RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*. 1996-2004; <http://www.repeatmasker.org>). Our findings, and those that are published (Emrich et al. 2004. *Bioinformatics* 20:140-147), showed that these libraries contain a small number of “contaminating” sequences that are neither transposons nor repetitive in the maize genome. Such sequences were also included in the MIPS REdat database (<http://mips.gsf.de/proj/plant/webapp/recat/>), which was compiled from these and other sources. The inadvertent inclusion of non-repetitive sequences in these libraries, which caused false-positive masking of several families of low-copy coding sequences, can be attributed to their adjacency to characterized transposable elements. With regard to sensitivity, the tested libraries were comprehensive in detecting both type I and II transposable elements within the BAC sequence test set. Nevertheless, we sought a method of repeat detection that did not require manual curation of a repeat database and could handle potentially novel repeats as they are encountered over the course of the Project. We therefore implemented mathematically defined repeats (MDR) (Yu et al. 2002. *Science* 296:79-92), using the 0.45X shotgun sequence (DOE Joint Genome Institute) as a uniform sampling of the maize genome. The Vmatch software package (Stefan Kurtz: www.vmatch.de) was used to generate a persistent suffix array index of constituent 20-mers within the shotgun sequence. In our pipeline, BAC sequences are subsequently used to query the suffix array, resulting in a semi-quantitative report of repeat levels over the length of the sequence. Our analysis using test BACs showed that masking by MDR gave similar sensitivity to that using the publicly-available repeat libraries, and corrected the problem of false-positive masking.

While it is necessary to exclude repetitive sequences to perform certain computational analyses, it is also important to regard transposons as biological entities that have played a substantial role in shaping the evolution of the maize genome. Users of the genome browser will

benefit not only from viewing repetitiveness in a quantitative sense, but also in understanding qualitatively the types of transposons present in their region of interest. Repeat classification is also essential for understanding how proliferation of different transposon classes contributed to the expansion of the maize genome. To classify repeat sequences, we are utilizing the MIPS Repeat Element Catalog (MIPS-REcat, <http://mips.gsf.de/proj/plant/webapp/recat/RecatTreeFrameset.jsp>). This database, when used in conjunction with the MIPS-REdat sequence library, provides systematic hierarchical classification of transposons and other repeat elements (Haberer et al. 2005. *Plant Physiology* 139:1612-1624). Evaluation of the MIPS classification tree on the BAC test set showed effective classification of type I (retroelement) and type II (DNA transposon) mobile elements, and effective subclassification of retroelements into the Ty1/copia and Ty3/gypsy families. To provide more refined classification and functional annotation of the LTR retroelements we are collaborating with Jeff Bennetzen (University of Georgia) and Phillip SanMiguel (Purdue University). Retroelements are frequently fragmented in the maize genome due to transposition of new elements into old, a phenomenon known as nested transposition that presumably affects their biological activity. Our collaborators provided a database of prototypical full-length retroelement sequences on which we were able to define the flanking LTR sequences. These reference sequences make it possible to display the “intactness” of retroelements, from which users can infer the history of transposition events and how this history might have affected present-day function.

Of course, most users will be interested in the gene space between the repetitive regions. We are using both *ab initio* gene prediction and evidence-based gene-build approaches to define protein-coding genes. Previous research has shown that FGENESH trained on monocot sequences (<http://www.softberry.com/berry.phtml>) provides the most accurate gene predictions in maize, among five *ab initio* programs tested (Yao et al. 2005. *Plant Mol. Biol.* 57:445-460). In our hands we found that FGENESH gives better predictions on non-masked sequence than on masked, the latter causing false-joins of exons from unrelated genes across repetitive regions. On non-masked sequences approximately 76% of resulting gene models lie within the repetitive fraction of the genome, the vast majority of which encode retroelement polyproteins. These gene models can be easily segregated from the others for purposes of display and downstream analysis. Approximately 18% of gene predictions showed homology to previously annotated protein sequences (mostly from rice), while 6% showed no homology. The latter may be attributed to the identification of novel genes, false-positive gene identification, or mis-prediction of genes (for example on the wrong strand or reading frame). Two flaws of FGENESH are failure to predict genes within genic regions (estimated at 18% by Yao et al. 2005), and prediction of truncated gene models. Prediction accuracy may be improved by training FGENESH on a larger dataset of maize genes. Softberry has expressed their willingness to retrain FGENESH and we have set aside budget to carry this forward. We are also working with Brad Barbazuk (Donald Danforth Plant Science Center) to adopt TWINSCAN software, which is being trained to annotate the maize genome (NSF #0501758; <http://maize.danforthcenter.org/abinitio.htm>). TWINSCAN integrates probability models like those of other *ab initio* programs with information from the alignments between two genomes. Early tests show promise at improving predictions compared to FGENESH (Barbazuk, personal communication). Ultimately, the most accurate gene models will come from evidence-based gene-builds. We are implementing the Ensembl gene-build procedure (Curwen et al. 2004.

Genome Research 14:942-950) which has been used on numerous genome annotation projects (<http://www.ensembl.org/index.html>). This method is automated and entirely based on experimental evidence. In principal, evidence can include protein, cDNA, and EST sequences derived from maize (targeted build) as well as from closely related species (similarity build), such as rice. In practice the Ensembl pipeline will be most effective with availability of the ~30,000 full-length maize cDNA sequences (NSF award # [0501857](#); PIs: Y. Yu, C. Soderlund, V. Walbot) over the next few years. As an intermediate step we have adapted the BLAT alignment and post-processing pipeline from the Gramene Project to display alignments of both genomic and expressed sequences from a wide range of grass species (see http://www.gramene.org/documentation/Alignment_docs/to_Maize/index.html). These alignments reveal conserved regions that generally correspond to protein-coding regions.

Annotation of protein sequences will utilize existing pipelines established by the Gramene and Ensembl projects. This includes identification of functional domains using InterPro searches and assignment of gene ontologies (GO) (see http://www.gramene.org/documentation/protein_help.html). Comparative genomics analysis, including the identification of conserved syntenic regions, orthologues, paralogues, and protein clusters will utilize the Compara multi-species analysis and database schema (<http://www.ensembl.org/info/software/compara/index.html>). For nucleotide-level whole genome alignments we are currently investigating the WABA (Kent, WJ & Zahler, A.M. 2000. Genome Res. 10:1115-25) and BLASTZ (Schwartz, S et al. 2003. Genome Res. 13:103-7; Kent, WJ, et al. 2003. PNAS 100:11484-11489) alignment tools.

Although CSHL is largely responsible for annotation, the ultimate delivery of meaningful data to the user requires teamwork at all levels of the Project. The staffs of CSHL and GSC have worked closely to develop a standardized GenBank submission record that will be accurately parsed as primary annotation for the sequenced clone. The Maize Project differs from other clone-by-clone sequencing projects in that most clones will not be sequenced beyond the Phase I level. Thus most clones will be represented as multiple contigs. Information on how these contigs are oriented and associated into scaffolds will be essential to users of the genome browser. We have therefore established methods to encode this information within the GenBank record submitted by GSC, so that it can be conveyed by the CSHL annotation team to the user.

C. The Maize Genome Browser

Equally important to generating meaningful annotation is its delivery to the user community. The first release of the MaizeSequence Browser (www.maizesequence.org) went live on 28 September 2006. The browser and database infrastructure are powered by Ensembl (<http://www.ensembl.org/index.html>), which has proven itself highly robust and flexible in the service of genome projects. The interface provides convenient entry points to the genome, both by searching and browsing, and displays salient features, including predicted genes, markers, repeats, and expressed and conserved regions. A high priority was to make searching and browsing easy for all members of the maize community, whether geneticist, breeder, or molecular biologist. Thus entry is currently possible by sequence accession, physical position, genetic position, and by conserved synteny with rice. The BLAST search engine is to be added in December 2006.

The genome browser is designed to operate at three levels of genomic resolution. In MapView, the highest, most global perspective, the genome is displayed at the fingerprint contig level (based on the July 2005 release of the AGI agarose FPC map). The maize core bin map is displayed alongside the physical map and the density of markers, clones, and accessioned BACs provide an overview of genome organization as well as progress of the Project. CytoView provides detailed visualization of the AGI physical map. BAC clones, hybridized markers (including many classes of overgoes), and simple sequence repeats are displayed here. Sequenced clones in GenBank are color-coded depending on their source and level of annotation. This view also serves as a launching point for ContigView, which displays the genome at the nucleotide level of resolution. ContigView is provided for clones that have reached HTGS_IMPROVED status and for clones publicly available outside of the current project. It displays primary sequence annotations such as *ab initio* gene prediction (FgenesH), repeats, and alignment to EST, cDNA, and GSS sequences. In the future ContigView will also serve as a launching point to view secondary annotations such as comparative maps (CMap), orthologue/parologue/syntenly analysis (Compara), and protein sequence characterization (Interpro/GO).

D. CSHL Maize Team Tools/Infrastructure

Research data and software development documents are centralized using a Project wiki. Issues, whether conceptual (new ideas for approaching annotation) or discrete (simple bug fixes), are tracked and ticketed among developers using Mantis. The Browser code and ancillary software such as the `cshl_tagger` (tags BACs with mathematical repeats), are maintained with subversion.

E. ISU Efforts

The ISU team has been developing a scaffolding approach that makes use of retrotransposons. Preliminary experiments conducted on maize BACs have been promising. Using 257 "random" BACs finished by the GSC downloaded from GenBank it was possible using this approach to conservatively obtain 1.5 additional scaffolds per BAC.

Plans: In the second year of the project, we are planning to introduce various enhancements by extending the set of tools available to the community to explore the progressing maize sequence (software improvements) as well as computationally improving existing data (data improvements). One software feature is an automated notification system that allows end users to "subscribe" to specific regions of the maize genome. The system, leveraged by the annotation pipeline, will notify users when a region of interest has an updated sequence or marker alignment. Another goal is to use Ensembl's Distributed Annotation System (DAS) infrastructure to provide alignments of procured data sets (such as mRNAs and full-length cDNAs). A third feature is the visual integration of the larger-scale FPC view with the more targeted, sequence-based BAC view, that will provide a uniform browsing context.

The annotation pipeline will also leverage genome-wide data improvements by integrating the maize physical map with the genetic map, as well as, subsequently, the optical map. FPC markers with known sequence will be aligned to updated BAC sequences. The alignments, or lack thereof, will serve to validate and verify markers hybridized during FPC map construction. The browser will, in turn, allow users to compare the original FPC hybridizations

with the updated annotations using graphical tracks. The maize optical map, currently under development at the University of Wisconsin, is already being integrated into the BAC selection process. A visual track displaying optical bands on a given BAC will further corroborate contig order-and-orientation.

At ISU during the coming year the utility of using the transposon scaffolding approach for sequencing complex plant genomes at low coverage will be explored.

6. Outreach

Responsible PIs/Key Personnel: R. Wilson (maize research community, GSC education program), W. McCombie (maize research community, CSHL public outreach program), all co-PIs (maize research community)

Goals: Development of a maize genome public education website at the CSHL DNA Learning Center primarily aimed at the general public, but will also include materials for non-science college students, as well as high school students. PI Schnable, along with other key personnel, will deliver public lectures to better educate the lay public about plant genetics and genomics.

NIH-funded outreach program at GSC is to be leveraged by including a section on maize/plant genetics/genomics.

Progress: In the past year outreach activities to the maize community by the Ware group included soliciting maize researchers for preliminary requirements for the maize genome sequence site, coordination with the maize community database MaizeGDB, establishing appropriate contacts with existing and future maize research initiatives and attending annual meetings. In order to establish browser requirements we solicited maize researchers at CSHL, Missouri, Iowa and the Plant Gene Expression center for feedback on the existing Gramene browser and specific maize requirements. This was done via phone calls, in person meetings, and email exchanges. To enhance the existing working relationship with MaizeGDB, Dr. Lawrence and her group spent a day at CSHL reviewing the browser, discussing mechanism for linking between the project sites, data exchange, and establishing a working model for feedback between the groups. To make the broader maize community aware of on going efforts, Drs. Lawrence and Ware coauthored “MGSC: Gramene and MaizeGDB cooperate to provide access to sequences and related data” published as part of the Maize Genetic Cooperation Newsletter volume 80, describing efforts to coordinate on the delivery of the maize sequence to the community. In order to build upon existing resources we have been working closely with several community members for data integration. These include annotation of maize retrotransposon elements with Drs. Phillip San Miguel and Jeff Benentzen, the maize optical map with Dr. Schwartz group and gene predictions with Dr. Brad Barbuzak.

The outreach program at CSHL is largely focused on developing a website with 3D graphics for high school students and the general public. We have been waiting for enough data

to support this activity. Now that the first large, “super contig” is nearing completion, we will begin building the website and incorporating this data.

We have also developed a “standard presentation” to describe the project to other scientists at meetings. Co-PI McCombie will be presenting this at the “Crop Genomics, Trait Analysis and Breeding” conference at the Wellcome Trust Genome Center in Hinxton on November 8. We will also be presenting this to the community at additional meetings to describe the project and the types of data being produced.

Dr. Wilson and co-PIs have made several presentations to the plant biology community about the maize genome sequencing project:

1. Plant Genomics European Meeting, Venice, Italy, October 2006.
2. National Academy of Science Workshop on Agricultural Biotechnology for the Global Public Good, Chennai, India, October 2006.
3. Plant Genomics in China VII, Harbin, China, August 2006.
4. China Agricultural University, China, August 2006.
5. Monsanto, St. Louis. May 2006
6. Biology of Genomes, Cold Spring Harbor, May 2006
7. Maize Genetics, Asilomar, March 2006
8. Advances in Genome Biology, February 2006
9. Plant and Animal Genomes, San Diego, January 2006

Plans:

CSHL: We will begin building the maize website at the DNA Learning Center, using the data that is just becoming available in a quantity that will allow for this. We have also developed a “standard presentation” to describe the project to other scientists at meetings. Co-PI McCombie will be presenting this at the “Crop Genomics, Trait Analysis and Breeding” conference at the Wellcome Trust Genome Center in Hinxton on November 8. We will also be presenting this to the community at additional meetings to describe the project and the types of data being produced.

GSC: We will be making presentations to the community at the Plant and Animal meeting in San Diego, CA in January, 2007 and at the Maize Genetics Meeting in ST Charles, IL in March 2007. We will be working with maize.gdb to post periodic announcements of progress to the community and will be submitting a brief paper to the Annual Maize Cooperative Newsletter, again regarding sequencing progress.

MGPAC Report

The Maize Genome Project Advisory Committee (MGPAC) reviewed the project at the Washington University Genome Sequencing Center on Oct. 3, 2006. The group heard presentations about the components being conducted at the WUGSC, the Arizona Genome Institute, Cold Spring Harbor Laboratory, the Joint Genome Institute, and Iowa State University.

Overall, the major components of the project are moving along. The overall progress is a little behind schedule, but not by a significant enough margin to warrant concern. This is the early stages of the project and it is not unusual for a large project to ramp up more slowly than expected, especially when coordination among multiple sites is involved. Moreover, the sequencing capacity of the WUGSC is sufficiently in excess over what is called for so that it should be a relatively straightforward process to make up for lost time. This is all predicated on the provision of enough BAC clones to take advantage of the sequencing capacity, and at this stage it appeared that activities at the AGI were directed appropriately at achieving this. Similarly, the development of the informatics contribution at the CSHL was also developing appropriately. Thus the committee felt that the overall progress in the project was good.

One area that had not matured well was the subproject to purify chromosome 10 by flow sorting and use this material for investigating the use of whole genome shotgun methods in sequencing a complex genome such as maize. At this stage highly purified chromosome 10 had not been achieved and the yield of the best material was lower than desired. The MGPAC encouraged Dr. Rokhsar to consider alternate plans for redirecting funds for this part of the project. Overall, the committee supported a plan for investigating whole genome shotgun assembly using the reads produced from the BAC clones as a partial redirection of the project. When the entire BAC tiling path is sequenced to 6x, this will represent an approximation to a whole genome shotgun dataset, limited by the clone coverage of the genome, the higher read coverage in regions of BAC overlap, among other issues. Nevertheless this would not incur additional sequencing costs and could be explored with a set of BACs that showed sufficient continuity even if it did not represent the whole tiling path to get some idea of how assembly software would need to be tuned.

An alternative proposal, "Plan B", was proposed by Dr. Rokhsar. This involved doing low coverage shotgun sequencing of a second maize inbred line, Mo17, for SNP discovery. The MGPAC was supportive about the general idea of sample sequencing additional maize genotypes to enhance SNP discovery. However, continued shotgun sequencing of Mo17 was not strongly supported without first considering the full range of ideas for use of the funds. The committee discussed other higher priority activities such as sample sequencing across diversity lines in maize, using the emerging JGI sorghum sequence to help anchor the B73 sequence, and other possible ways to leverage ongoing and new sequence information. The MGPAC strongly advocates for a Plan B that emanates from the maize community through direct communication with the Maize Genetics Executive Committee (MGEC, via Sarah Hake, chair). Any redirection of funds should be consistent with community direction and integrated with comparative maize sequence efforts that are newly funded. Dr. Rokhsar was encouraged to prepare a proposal that is guided by the community gold standard, and that contains more budgetary and technical details. A follow-up review by the MGPAC of an alternate Plan B proposal is perhaps warranted.

The committee also recognized ambiguity in the relationship between the JGI component and the main Washington University coordinated project. While the use of the Washington University data for testing whole genome assembly approaches seemed straightforward to coordinate, the relation of the proposed sample sequencing project at JGI to the main project was not clear. The MGPAC hopes that the development of Plan B will include feedback from the Washington University center as to how to integrate these two activities.

Finally, there was discussion about the need for regular communication about project progress to the maize community at regular intervals. A possible forum for this would be through announcements in maize.gdb or as a submitted brief paper to the Annual Maize Cooperative Newsletter, in addition to regular attendance and presentations at the Annual Maize Genetics Conference in March. Communication will also improve once the maizesequence.org website is disseminated publicly to the community using the maize listserv and by announcements on [maizegdb](http://maizegdb.org). The PIs are encouraged to invest time in this form of communication in the near future.

Submitted by George Weinstock, Chair

Investigator: Richard K. Wilson

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: Large Scale Genome Sequencing

Source of Support: US NIH/NHGRI (Wilson)

Total Award Amount: \$ 136,400,000 Total Award Period Covered: 11/01/03 - 11/30/06

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 7.20 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: Comparative Genomics of the *Enterobacteriaceae*

Source of Support: US NIH/NIAID

Total Award Amount: \$ 2,020,143 Total Award Period Covered: 09/01/02 – 08/31/07

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 0.12 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: A genomic approach to Parasites from the Phylum Nematode

Source of Support: US NIH/NIAID

Total Award Amount: \$ 2,722,459 Total Award Period Covered: 03/01/04 - 02/28/08

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 0.24 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: Genomics of Acute Myelogenous Leukemia (Proj. 1.)

Source of Support: US NIH/NCI (T. Ley)

Total Award Amount: \$ 2,250,754 Total Award Period Covered: 09/19/03 - 08/31/07

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 1.20 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: Comparative Microbial Genome Analysis of the Human-Bacteroides symbiosis

Source of Support: US NSF EF 0333284 (Gordon)

Total Award Amount: \$ 1,993,732 Total Award Period Covered: 11/01/03 – 10/31/06

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 0.24 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support

Project/Proposal Title: Sequencing the Maize Genome

Source of Support: US NSF

Total Award Amount: \$29,450,000 Total Award Period Covered: 11/1/2005-10/31/2008

Location of Project: Washington University School of Medicine

Person-Months Per Year Committed to the Project. Cal: 2.40 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support
Project/Proposal Title: Genomics of Acute Myelogenous Leukemia (Proj. 1.)
Source of Support: US NIH/NCI (T. Ley)
Total Award Amount: \$ 2,784,056 Total Award Period Covered: 09/01/07 - 08/31/12
Location of Project: Washington University School of Medicine
Person-Months Per Year Committed to the Project. Cal: 1.20 Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support
Project/Proposal Title: Center for Large Scale Genome Sequencing and Analysis
Source of Support: US NIH/NHGRI (Wilson)
Total Award Amount: \$ 156,000,000 Total Award Period Covered: 11/01/06 - 11/30/10
Location of Project: Washington University School of Medicine
Person-Months Per Year Committed to the Project. Cal: **7.20** Acad: 0.00 Sumr: 0.00

Support: Current Pending Submission Planned in Near Future *Transfer of Support
Project/Proposal Title: Metagenomic studies of the gut microbiomes of obese and lean twins
Source of Support: US NIH (Gordon)
Total Award Amount: \$ 1,000,000 Total Award Period Covered: 7/01/07-6/30/12
Location of Project: Washington University School of Medicine
Person-Months Per Year Committed to the Project. Cal: 0.12 Acad: 0.00 Sumr: 0.00

Current and Pending Support
See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to			
Investigator: W. RICHARD McCOMBIE	Other agencies (including NSF) to which this proposal		
Support:	<input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future
			<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: Systematic Determination of the Rice Gene Set Source of Support: USDA-204-35604-14228 Total Award Amount: \$ 193,354 Total Award Period Covered: 02/1/04-01/31/07 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 3 Acad: Sumr:			
Support:	<input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future
			<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: CSHL Cancer Center Support Grant/Nucleic Acid Chemistry Shared Resource Section Source of Support: NIH-5P30 CA45508-19 Total Award Amount: \$ 9,837 Total Award Period Covered: 08/1/05-07/31/10 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 0.6 Acad: Sumr:			
Support:	<input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future
			<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: VCA-Finishing the Rice Genome Source of Support: NSF/D.B.I - 0321683 Total Award Amount: \$ 352,851 Total Award Period Covered: 9/1/03 – 8/31/07 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 1.8 Acad: Sumr:			
Support:	<input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future
			<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: VCA: Virtual Center for Plant Evolutionary Genomics Source of Support: NSF-DBI-0421604 Total Award Amount: \$98,761 Total Award Period Covered: 10/1/04-09/30/09 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 1.8 Acad: Sumr:			
Support:	<input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future
			<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: YIA: Genomics of Rice, Sorghum and Maize Source of Support: NSF – DBI-0333074 Total Award Amount: \$ 14,000 Total Award Period Covered 09/1/03-08/31/08 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 0 Acad: Sumr:			
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			

Current and Pending Support

See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to			
Investigator: W. RICHARD McCOMBIE	Other agencies (including NSF) to which this proposal		
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: TRPGR: Characterization of Rice Genomes and Transcriptomes using Novel Sequencing Technologies Source of Support: NSF. - 0608405 Total Award Amount: \$301,352 Total Award Period Covered 09/1/06 – 08/31/09 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 1.2 Acad: Sumr:			
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: Improving Cereal Genome Resources for Cereal Crop Improvement Source of Support: USDA. 58-1907-4-444 Total Award Amount: \$ 34,374 Total Award Period Covered: 09/15/04-09/14/09 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 0.24 Acad: Sumr:			
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: Sequencing the Maize Genome Source of Support: NSF. – DBI 0527192 Total Award Amount: \$302,385 Total Award Period Covered: 11/15/05 – 11/14/08 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 2.6 Acad: Sumr:			
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: Arabidopsis 2010: A Comprehensive Resource for Analysis of Arabidopsis Gene Function Source of Support NSF. Total Award Amount: \$374,560 Total Award Period Covered: 09/1/06 – 08/31/09 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: 1.2 Acad: Sumr:			
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: High Throughput Sequencer (Equipment) Source of Support: NIH Total Award Amount: \$500,000 Total Award Period Covered: 04/1/07 – 03/31/08 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Cal: Acad: Sumr:			
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			

Current and Pending Support

See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to			
Investigator: W. RICHARD McCOMBIE	Other agencies (including NSF) to which this proposal		
Support: <input type="checkbox"/> Current	<input checked="" type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: Methods for Genome Wide Detection of Variation in Transcription Start sites			
Source of Support: NIH			
Total Award Amount: \$150,000		Total Award Period Covered: 03/1/07 – 02/28/09	
Location of Project: Cold Spring Harbor Laboratory			
Person-Months Per Year Committed to the			
	Cal: 0.6	Acad:	Sumr:
Support: <input type="checkbox"/> Current	<input checked="" type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: TRPGR: Construction of comprehensive sequence-indexed transposon resources for maize			
Source of Support: NSF			
Total Award Amount: \$7,159,676		Total Award Period Covered: 07/01/07 – 06/30/11	
Location of Project: University of Florida			
Person-Months Per Year Committed to the			
	Cal: 1.20	Acad:	Sumr:
Support: <input type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project:			
Person-Months Per Year Committed to the			
	Cal:	Acad:	Sumr:
Support: <input type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project:			
Person-Months Per Year Committed to the			
	Cal:	Acad:	Sumr:
Support: <input type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project: Cold Spring Harbor Laboratory			
Person-Months Per Year Committed to the			
	Cal:	Acad:	Sumr:
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.			
Investigator: Doreen Ware	Other agencies (including NSF) to which this proposal has been/will be submitted.		
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Sequencing The Maize Genome DBI-0527192 Source of Support: Washington Univ/NSF Total Award Amount: \$436,025 Total Award Period Covered: 11/15/06 – 10/31/08 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Project. Cal: 1.8 Acad: Sumr:			
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: ORZA Map Alignment Project DBI-0321678 Source of Support: Univ. Arizona/Stein Total Award Amount: \$0 Total Award Period Covered: 10/01/03-09/30/07 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Project. Cal: 0.6 Acad: Sumr:			
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Molecular basis of local adaptation Source of Support: USC Total Award Amount: \$41,186 Total Award Period Covered: 04/01/07-03/31/12 Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Project. Cal: .036 Acad: Sumr:			
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: High density scoreable markers for maize trait dissection Source of Support: NSF/Cornell Univerisy Total Award Amount: \$62,349 Total Award Period Covered: 10/01/06-09/30/08 Location of Project: Cold Spring Harbor laboratory Person-Months Per Year Committed to the Project. Cal: .036 Acad: Sumr:			
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: SORGHUM BIOMASS/FEEDSTOCK GENOMICS RESEARCH FOR BIOENERGY (ROONEY) Source of Support: DOE/TAMU Total Award Amount: \$ 22,641 Total Award Period Covered: Location of Project: Cold Spring Harbor Laboratory Person-Months Per Year Committed to the Project. Cal: .036 Acad: Sumr:			
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			

NSF Current and Pending Support

Rod A. Wing

CURRENT

Project/Proposal Title: Oryza Map Alignment Project
PI: Rod Wing; CoPI's- Scott Jackson, Lincoln Stein
Source of Support: NSF
Total Award Amount: \$9,743,546 (70% spent by 9/30/05)
Total Award Period Covered: 10/01/03-09/30/07
Location of Project: University of Arizona, Purdue, Cold Spring Harbor Laboratory
Person-Months Per Year Committed to the Project : ACAD: 2.40 months

Project/Proposal Title: Finishing the Rice Genome
PI: W.R. Mc Combie; CoPI's- Yeisoo Yu, Carol Soderlund, Rod Wing
Source of Support: NSF
Total Award Amount: \$4,202,799
Total Award Period Covered: 10/01/03-09/30/06
Location of Project: Cold Spring Harbor Laboratory, University of Arizona
Person-Months Per Year Committed to the Project: ACAD: 1.80 months

Project/Proposal Title: Comparative Evolutionary Genomics of Cotton
PI: J. Wendel; CoPIs- A. Gingle, A. Paterson, R. Wing
Source of Support: NSF
Total Award Amount: \$1,047,034
Total Award Period Covered: 10/01/02-09/30/07
Location of Project: Iowa State, University of Georgia, University of Arizona
Person-Months Per Year Committed to the Project : ACAD: 0.48 months

Project/Proposal Title: VCA: SoyMap, an integrated map of soybean for resolution and dissection of multiple genome duplication events
PI: Scott Jackson, Senior Investigator - Rod Wing
Source of Support: NSF
Total Award Amount: \$1,373,592
Total Award Period Covered: 05/01/05-04/31/08
Location of Project: Purdue, University of Arizona
Person-Months Per Year Committed to the Project : ACAD: 0.60 months

Project/Proposal Title: Acquisition of Instrumentation for Omics Research at the University of Arizona
PI: David Gang, Co PI- David Galbraith, Rod Wing
Source of Support: NSF (MRI)
Total Award Amount: \$192,350
Total Award Period Covered: 12/01/05-11/30/08
Location of Project: University of Arizona
Person-Months Per Year Committed to the Project : ACAD : 0.24 months

**NSF Current and Pending Support
Rod A. Wing**

CURRENT

Project/Proposal Title: Sequenced insertion lines for rice functional genomics
PI: V. Sundaresan CoPI- Rod Wing
Source of Support: USDA
Total Award Amount: \$750,000 (UA \$76,800)
Total Award Period Covered: 11/01/05-10/31/08
Location of Project: UC Davis
Person-Months Per Year Committed to the Project: ACAD: 1%

Project/Proposal Title: Sequencing the Maize Genome
PI: Rick Wilson, CoPI's: W.R. McCombie, Rod Wing, Pat Schnable, Doreen Ware
Source of Support: NSF
Total Award Amount: \$30,000,000 (UA \$2,201,983)
Total Award Period Covered: 09/01/05 – 08/31/08
Location of Project: Washington University, Cold Spring Harbor Laboratory, University of Arizona
Iowa State University
Person-Months Per Year Committed to the Project: ACAD : 0.75 months

Project/Proposal Title: Evolutionary Genomics of a Rice Centromere
PI: Jiming Jiang, CoPI's- Rod Wing, HyeRan Kim, Scott Jackson
Source of Support: NSF
Total Award Amount: (UA \$1,458,309)
Total Award Period Covered: 8/01/2006-07/31/2011
Location of Project: University of Wisconsin-Madison
Person-Months Per Year Committed to the Project : SUMMER: 0.02 months

Project/Proposal Title: MO: Kartchner Caverns: Habitat Scale Community Structure and Function in Carbonate Caves
PI: Raina Maier CoPI's- Leland Pierson III, Barry Pryor, Rod A Wing
Source of Support: NSF
Total Award Amount: \$1,600,000
Total Award Period Covered: 07/01/2006-06/30/2011
Location of Project: University of Arizona
Person-Months Per Year Committed to the Project : ACAD: 0.01 months

Project/Proposal Title: Highly Reduced Genomes of Coresident Bacterial Symbionts of Xylem-Feeding Insects: Ecological and Evolutionary Implications
PI: Nancy Moran, CoPI- Rod A. Wing
Source of Support: NSF
Total Award Amount: \$ 440,793
Total Award Period Covered: 09/01/06-08/31/09
Location of Project: University of Arizona
Person-Months Per Year Committed to the Project : ACAD: 1.0 month

NSF Current and Pending Support
Rod A. Wing

CURRENT

Project/Proposal Title: Sequencing of chromosome 3 short arms from the AA, BB, CC, BBCC genomes of wild relatives of rice for comparative functional and evolutionary genomics

PI: Rod Wing, CoPI's Scott Jackson, Steven Rounsley, Lincoln Stein

Source of Support: NSF

Total Award Amount: \$2,735,151

Total Award Period Covered: 10/01/06-8/31/08

Location of Project: University of Arizona – BIO5, Purdue University, Cold Spring Harbor Laboratory

Person-Months Per Year Committed to the Project : SUMMER: 0.5 month

Project/Proposal Title: Genome Evolution in Diploid and Polyploid cotton

PI: Jonathan Wendel, CoPI's Rod Wing, Andrew Paterson, Adah Leshem-Ackerman

Source of Support: NSF

Total Award Amount: (UA \$534,176)

Total Award Period Covered: 01/01/07-12/31/09

Location of Project: Iowa State University, University of Arizona, University of Georgia

Person-Months Per Year Committed to the Project : SUMMER: 0.5 month

PENDING

Project/Proposal Title: GEPR, alignment of physical maps of rice and two inbreds of Zea mays

PI: Joachim Messing, CoPI- Rod A. Wing

Source of Support: NSF

Total Award Amount: (UA \$2,245,094)

Total Award Period Covered: 07/01/07-6/30/2010

Location of Project: Rutgers, University of Arizona-BIO5

Person-Months Per Year Committed to the Project : ACAD: 1.2 months

Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.			
Investigator: Patrick S. Schnable	Other agencies (including NSF) to which this proposal has been/will be None		
Support: <input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: VCA-High-Density Genetic Map of Maize Transcripts			
Source of Support: NSF-PG (DBI-0321711)			
Total Award Amount: No Cost Extension		Total Award Period Covered: 10/1/03-9/30/06 (extension through 9/30/07)	
Location of Project: Iowa State University			
Person-Months Per Year Committed to the Project.		Cal: 2.7	Acad: Sumr:
Support: <input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: VCA-A Rice Oligo Chip			
Source of Support: NSF-PG (DBI-0313887)			
Total Award Amount: No Cost Extension		Total Award Period Covered: 10/1/03-9/30/06 (extension through 9/30/07)	
Location of Project: Iowa State University			
Person-Months Per Year Committed to the Project.		Cal: 0.5	Acad: Sumr:
Support: <input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: Essential Nature of Fatty Acid Elongation in Plant Development			
Source of Support: NSF (DBI-0344852)			
Total Award Amount: \$122,397		Total Award Period Covered: 4/15/04-3/31/07	
Location of Project: Iowa State University			
Person-Months Per Year Committed to the Project.		Cal: 0.5	Acad: Sumr:
Support: <input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: Recombination Mechanisms in Maize			
Source of Support: USDA-NRI (05-00962)			
Total Award Amount: \$240,000		Total Award Period Covered: 7/1/05-6/30/07	
Location of Project: Iowa State University			
Person-Months Per Year Committed to the Project.		Cal: 0.5	Acad: Sumr:
Support: <input checked="" type="checkbox"/> Current	<input type="checkbox"/> Pending	<input type="checkbox"/> Submission Planned in Near Future	<input type="checkbox"/> *Transfer of Support
Project/Proposal Title: Acetyl-CoA: The Precursor for High-Energy Phytochemicals			
Source of Support: DOE (DE-FG02-01ER15170)			
Total Award Amount: \$96,198		Total Award Period Covered: 7/15/04-7/14/07	
Location of Project: Iowa State University			
Person-Months Per Year Committed to the Project.		Cal: 0.5	Acad: Sumr:
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			

