RESEARCH REPORT
SWINE HEALTH

Title: Implementation of a PRRSV Strain Database (Renewal) - **NPB #06-127**

Investigator: Kay S. Faaberg, PhD (PI now located at NADC in Ames, IA)

Institution: University of Minnesota

Date Submitted: October 30, 2007

Industry Summary:
The PRRSV strain database (http://prrsv.ahc.umn.edu) has been supported by the National Pork Board for three years. The database (PRRSVdb; 7,627 unique sequences) is freely available for web-based queries in order to obtain detailed information about PRRSV sequences (percent similarity to vaccines or other database isolates, ORF5 RFLP analysis, year and state of isolation, phylogenetic relationships, GenBank submission numbers). The project, proposed to fulfill the NPB directive to implement a National PRRSV Sequence Database, has relevance to a number of PRRS Initiative efforts such as surveying potentially new field strains for vaccines, examining sequence conservation among various isolates for improvement of diagnostics, epitope evaluation and assessment of virus spread locally, nationally and internationally. The database is comprised presently of 3 major veterinary diagnostic laboratory sequenced isolates (Minnesota Veterinary Diagnostic Laboratory, South Dakota Animal Disease Research & Diagnostic Laboratory, Manitoba Veterinary Services Branch) and those independently deposited in GenBank. In addition, the appearance of the database has been updated to provide additional flexibility to the PRRSVdb. This new flex-based web interface provides improved phylogeny viewing and user-uploaded sequence analysis, either dependent or independent of the database. Accessing related clinical data will be available through PRRSVdb interaction with other data sources soon. This research is the result of continued interactions of the National Pork Board (Pamela Zaabel), the National Center for Animal Disease (Kay Faaberg) the Minnesota Veterinary Diagnostic Laboratory (MVDL; James Collins and the Molecular Diagnostic Section), South Dakota Animal Disease Research & Diagnostic Laboratory (Jane Christopher-Hennings, Travis Clement), and the Manitoba Veterinary Services Branch (Andre Hamel) with software engineers (Trevor Wennblom, John Crow and Ernest Retzel) located at the Center for Biomedical Research Informatics (CBRI) of the University of Minnesota.

For more information contact:
National Pork Board, P.O. Box 9114, Des Moines, Iowa USA
800-456-7675, **Fax:** 515-223-2646, **E-Mail:** porkboard@porkboard.org, **Web:** http://www.porkboard.org/

Scientific Abstract: The PRRSV database now contains 7,627 unique sequences (>2% nucleotide difference) with accessions from the Minnesota Veterinary Diagnostic Laboratory, the South Dakota Disease Research & Diagnostic Laboratory, the Manitoba Veterinary Services Branch as well as many PRRSV sequences gathered from GenBank. The expansion of the database to include other laboratories has made exploration of possible new field isolates more robust. In this past year, we have also taken the necessary steps to incorporate field surveillance data, made online submission easy for the technicians, and improved the query capability of the database. In so doing, the web-page format was drastically altered, using software tools that have just become available within the past year. Although not yet fully implemented, this new design promises to be more adept at deriving more user-defined information, not just a limited database for sequence viewing and comparison. As the project enters its fourth year, we now can concentrate on providing clinical data, additional nucleotide sequence content and more advanced bioinformatics tools.

Introduction: This report details the improvements made to the initial NPB grant entitled "Implementation of a PRRSV Strain Database" (NPB project #04-118). The inherent merit of the research proposed mirrors the initial community desire to have all PRRSV nucleotide sequences and tools to investigate PRRS disease available in one easy to access web domain. The dedication of the principal researchers, the software engineers and the technicians producing the PRRSV sequences has made the PRRSV Database one of the leaders in web-design and functionality. In the past year, we have successfully expanded the diagnostic laboratories to include a major PRRSV sequencing center, South Dakota Animal Disease Research & Diagnostic Laboratory (SDADRDL). In addition, we opted to transfer the database to a new flex-based interface which shows the MySQL Database on a more user-friendly web page. As a result, the appearance of the database is entirely new and is still in its initial stages, but the supportive data is maintained in the same way behind the scenes. Lastly, to begin to supply limited field surveillance data, two avenues were taken. One was to include requested information on the veterinary diagnostic submission forms and the other was to request related information from the MDVL database. Much of this has been completed and will be described fully below.

Objective 1. ***Incorporate SDSU sequences into the PRRSV relational database***
Objective 2. ***Further refine phylogeny analysis***
Objective 3. ***Expand the available field surveillance data by accessing limited information from other database sources***

Materials & Methods: The PRRSV database, after the second year of funding, consisted of over 5000 unique PRRSV sequences (http://prrsv.ahc.umn.edu). In the past year, the MVDL private database expanded to 5,126 individual PRRSV sequences, only some of which were novel to the database. The South Dakota Animal Disease Research & Diagnostic Laboratory Database has added 3,274 sequences, only 964 of which were unique. Imported GenBank sequences total 433. A total of 7,627 unique sequences (>2% nucleotide difference) are part of the PRRSV database, with 429 yet to be reviewed. The duplicate sequences were noted and electronically "filed" with the isolation date and place.

The software tools imported and implemented include The Basic Local Alignment Search Tool for database specific use (BLAST; http://www.ncbi.nlm.nih.gov/BLAST/) ClustalW (http://www.ebi.ac.uk/clustalw/), Phylip (http://evolution.gs.washington.edu/phylip.html). Interproscan (http://www.ebi.ac.uk/InterProScan/), Protein Family Alignment Annotation Tool (PFAAT; http://pfaat.sourceforge.net/), Jalview (http://www.jalview.org/), and MFold (http://frontend.bioinfo.rpi.edu/applications/mfold/; M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406-15, (2003)). Native software developed by CBRI include new software code to result in less errors and better software problem identification, ORF detection software and RFLP analysis. The web framework is Ruby on Rails™, but now has been integrated with Flex, an additional framework that helps you build dynamic, interactive rich Internet applications. The PRRSV database itself is still based on MySQL software. Note that the rich blend of software from several domains are all open source and were blended to work together by the expert software programmers Trevor Wennblom and John Crow at CBRI.

Results:

Objectives completed this year but funded in prior year (NPB #05-162). There were some improvements to the PRRSV database that we accomplished this past year. They include:

a. Addition of nucleotide similarity/difference matrices to the output of sequence comparison. Shown in Figure 1, under the comparison tab of the new flex-based web-page (described below), the user can now access the percentage identity between any sequences compared.

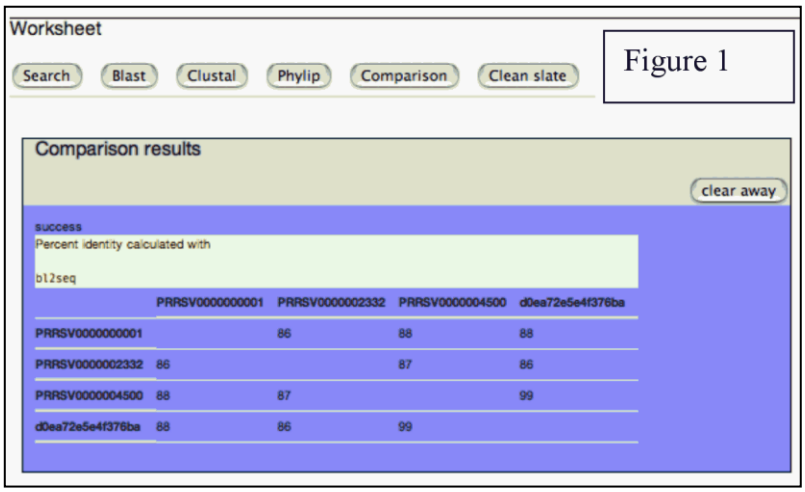


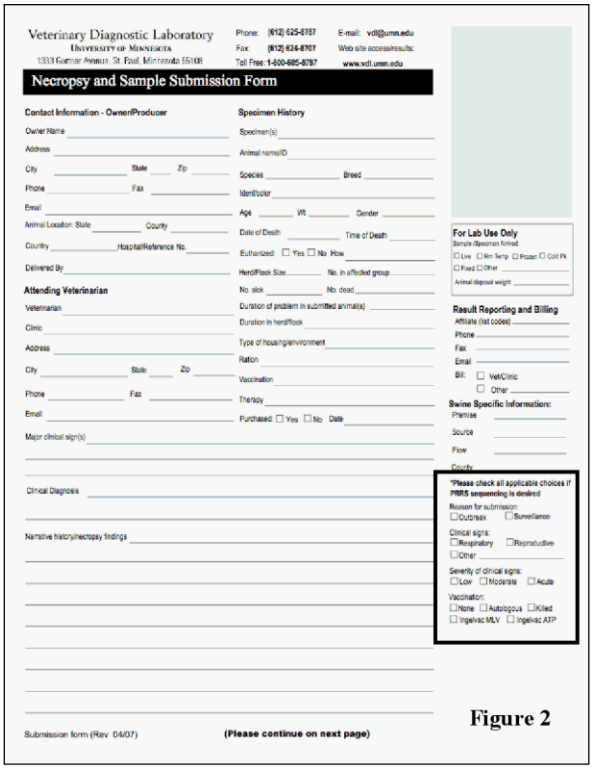
Figure 1

b. Importation of all other GenBank sequences. >1600 sequences were added with both GenBank Accession numbers as well as a PRRSVdb specific identifier. Thus, the database user can compare an input sequence to GenBank accessions made worldwide without switching to the actual National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/)

c. With the help of MVDL personnel, we redesigned the manual veterinary submission form to account for the information requested by the NPB (Figure 2):
   i. Date of isolation
   ii. Place of isolation
   iii. Reason for submission (outbreak or surveillance)
   iv. Respiratory or reproductive disease or both
   v. Severity of the clinical signs (low, moderate, acute)
   vi. Prior Exposure (naïve, acclimation, serum inoculation, vaccination)

d. Addition of columns to the MySQL database for the importation of the clinical information described above.

Therefore, we have completed, from the PRRSVdb end) all the necessary steps for clinical PRRS disease surveillance data associated with a particular nucleotide sequence, both of which reside in the private MVDL database, to now be accessible in the PRRSVdb.


Veterinary Diagnostic Laboratory
University of Minnesota
1333 Gortner Avenue, St. Paul, Minnesota 55108
Phone: (612) 625-8707
Fax: (612) 624-8707
Toll Free: 1-800-605-8787
E-mail: vdl@umn.edu
Web site access/results: www.vdl.umn.edu

Necropsy and Sample Submission Form

Contact Information - Owner/Producer
Owner Name
Address
City State Zip
Phone Fax
Email
Animal Location: State County
Country Hospital/Reference No.
Delivered By

Attending Veterinarian
Veterinarian
Clinic
Address
City State Zip
Phone Fax
Email
Major clinical sign(s)

Specimen History
Specimen(s)
Animal name/ID
Species Breed
Identification
Age Wt Gender
Date of Death Time of Death
Euthanized: Yes No How
Herd/Flock Size No. in affected group
No. sick No. dead
Duration of problem in submitted animal(s)
Duration in herd/flock
Type of housing/environment
Ration
Vaccination
Therapy
Purchased: Yes No Date

Clinical Diagnosis
Narrative history/necropsy findings

For Lab Use Only
Sample/Specimen Arrived
Live Rm Temp Frozen Cold Pk
Fixed Other
Animal disposal weight

Result Reporting and Billing
Affiliate (list codes)
Phone
Fax
Email
Bill: Vet/Clinic
Other

Swine Specific Information:
Premise
Source
Flow
County

*Please check all applicable choices if PRRS sequencing is desired
Reason for submission:
Outbreak Surveillance
Clinical signs:
Respiratory Reproductive
Other
Severity of clinical signs:
Low Moderate Acute
Vaccination:
None Autologous Killed
Ingelvac MLV Ingelvac ATP

Submission form (Rev. 04/07)
(Please continue on next page)

Figure 2

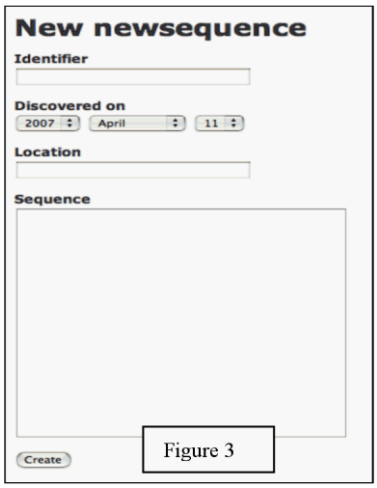


Figure 3

e. New online nucleotide sequence submission form (Figure 3). This new online format alleviates the need for PI Faaberg nucleotide sequence review before submission to the database. Thus, both the MVDL and the SDADRDL can immediately deposited their sequences online with all information needed close at hand. This web-page form is accessible only to certain IP addresses; it is not available to investigators from outside approved diagnostic laboratories.

Objective 1. ***Incorporate SDSU sequences into the PRRSV relational database***.

a. *Train student employee at SDSU on method of submission* – At the beginning of the funding cycle, an SDSU student (Travis Clement) was rapidly trained on the procedure utilized to submit sequences to the PRRSVdb.

b. *Import SDSU PRRSV ORF5 sequences into PRRSVdb* – A login at UMN-CCGB was created for submission of SDSU sequences. 1000 PRRSV sequence ID's (ID's 6000-6999) were forwarded to SDSU. An additional number of PRRSV sequence ID's were forwarded later in the year (7000-9274). SDSU assigned an ID to a nucleotide sequence, periodically updated the related spreadsheet, and then forwarded the spreadsheet to the curator of the PRRSVdb. To date, SDSU has submitted 3274 sequences.

b. *Setup SDSU VDL with redesigned veterinary diagnostic submission form* – The NPB Swine Health Committee established criteria in 2006 that they wanted added to the PRRSVdb, namely 1) Reason for submission (outbreak or surveillance), 2) Respiratory or reproductive disease or both, 3) Severity of the clinical signs (low, moderate, acute), 4) Prior Exposure- (naïve, acclimation, serum inoculation, vaccination) and 5) Zip code (prefer premise ID). This list was sent to SDSU. SDSU now routinely puts out the questions as described (minus item 5) to all of their e-mail clients (sequencing requests and results are transmitted via e-mail) and only one client answers the questions routinely (6 total have been returned). They report that it is difficult to get them to submit the answers to the database questions. To get the SDSU submission forms revised will take some time. Jane Christopher-Hennings did submit the PRRSVdb questions to her Director and Associate Director, but nothing has been added at this time. She will continue to work on this aim in the following year.

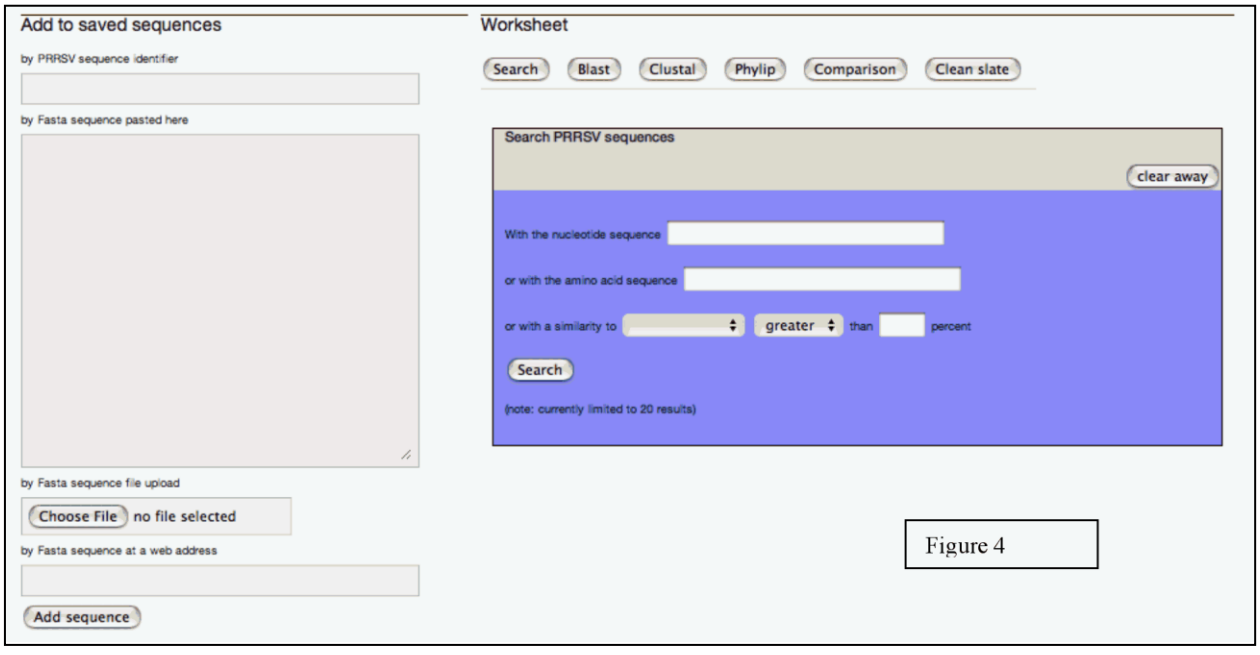


Figure 4

Objective 2. ***Further refine phylogeny analysis.***

A new user interface (shows the MySQL Database on a user friendly web page) became available - called a Flex-based interface. Trevor Wennblom, who has built all aspects of the database, made a decision to migrate the PRRSV Database into this format, making addition of new columns of information and different ways of searching the materials much more facile. A stable version of this new webpage before analysis begins is shown in Figure 4.

a. *Provide uploading tool for nucleotide sequences to be aligned*. In this new view, on the left hand side are a number of options to display sequences you want to work with (by PRRSV database identifier, by cutting and pasting in a user sequence, by upload a specific fasta file, or by accessing a sequence located at another IP address). The user can upload as many sequences as they wish, just clicking on add sequence after the user has filled in one of the boxes. The main advantage to this screen is that you are no longer limited to looking only at one user sequence. This is illustrated in Figure 5. An individually user may add as many PRRSV sequences as desired.

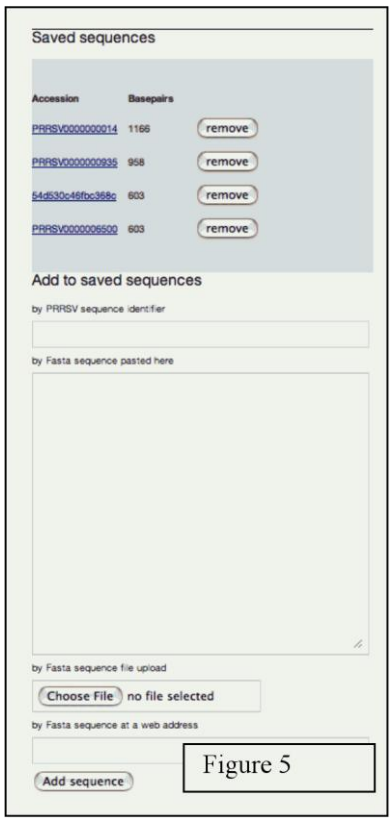


Figure 5

b. *Redesign phylogenetic analysis to allow for PRRSV database independent use* – The ability of the user to submit personal sequences to the database finally allows for PRRSVdb independent phylogeny study. Once all sequences are uploaded on the ride side of the web-page (Figure 6, bottom panel), the user would click on the clustal button. The resulting alignment appears on the right side in the top box (Figure 6, top panel). I have shown the results of a user sequence compared with PRRSVdb sequences, but database independent use is completed in the same manner.

c. Notice that the Jalview and PFAAT (alignment editors with extra capabilities) have not been implemented yet. These two external programs, as well as other standard features (such as RFLP, place and date of isolation, etc. will become available to the new PRRSVdb over the next few months.

Objective 3. ***Expand the available field surveillance data by accessing limited information from other database sources***

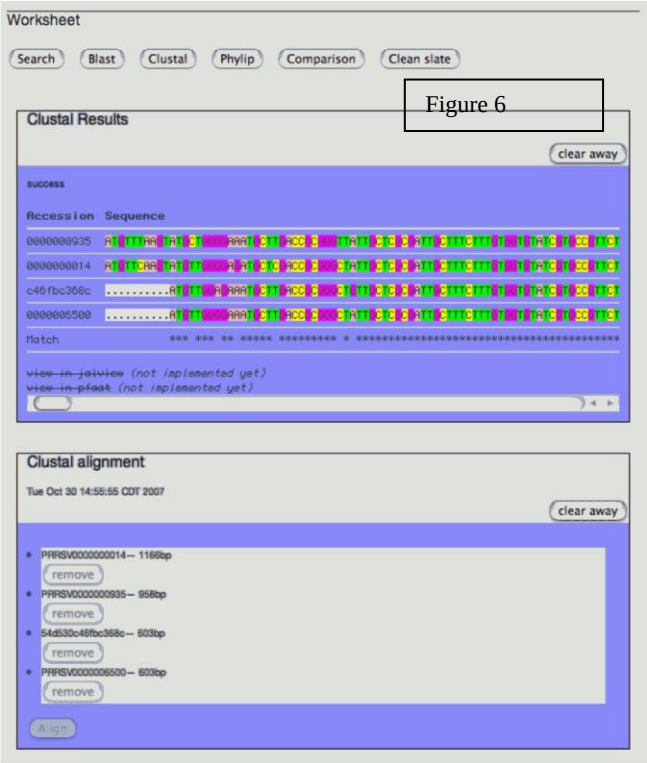


Figure 6

a. *Generate diagnostic case number format used in MVDL* – We have developed a software tool to automatically change all cases originally sent to the database manager (e.g., D07-1111) to the format used by the MVDL database (D07-001111). Thus, we can quickly convert all MVDL ORF5 submissions to this new format.

b. *Migrate relevant field information from MVDL database to PRRSVdb* - We have catalogued the items from the MVDL Database to be added to the PRRSV database and the specific information outlined by the Swine Health Committee forwarded to me in July of 2006. These have been added as columns to the supporting MySQL database but not yet visible on the associated web pages, as we continue to migrate to the new format.

i. Presently, the MVDL electronic sample submission form requests the following relevant criteria:

1. Abortion
2. Central Nervous System
3. Respiratory
4. Septicemia
5. Sudden Unexpected Death
6. Other

ii. The NPB Swine Heath Committee has found that the criteria listed above in 1c are appropriate.

b. In order to capture all relevant information, we added data submission fields for each of these criteria to the PRRSV MySQL database, described above. The zip code, often a swine farm specific identifier, will not be included at this time due to confidentiality issues. These columns will be shown on the web page as the Flex-based interface is brought on line. John Crow, a coinvestigator with CBRI, has spoken to the MVDL database manager (Mary Thurn), and we are assured that data integration with the PRRSVdb will go forward shortly.

Discussion: This funded project has been ongoing since 2004. We originally sought to be a one-stop shop for easy viewing of PRRSV sequences and the information that can be derived from them, such as RFLP, date of sequence emergence, place of origin, N-glycosylation, etc. However, it became clear that additional tools and field surveillance data were needed to augment this original intent. This last has proven to be quite difficult, as veterinary diagnostic laboratories do not require such information in order to access their services and are reluctant to revise their submission forms that already are extremely crowded. However, the MVDL has revised their manual submission form and the data (if supplied by the submitting veterinarian/producer) is now recorded electronically. It is anticipated that the MVDL will start sending us the required information within the next few months.

The development of the database and associated web-pages may appear to be lagging at times, but the underlying software and hardware have undergone extensive revision as new ideas are brought forth and new tools become available. In addition, we acknowledge that there are additional tools to migrate from the original database format to the new Flex-based web-page format. However, the difficult task of creating the new appearance and enabling more advanced queries has been accomplished.

The development of the PRRSVdb and the modifications to the database design as well as the web-page design are progressive, both in terms of database content but in the tools that are used in ongoing revision. It is difficult to assess how this database is perceived within the PRRS Disease community, but to our knowledge it is one of the few publicly available databases that strives to put information on PRRSV in an easy to navigate, easy to understand format. In the next year, we have been funded to bring in additional sequences from Iowa State University and Hong Kong, and alongside of continued database expansion we will continue to strive for the enhancement of database content and query strategies.
One abstract describing the PRRSVdb has been published:

Faaberg, K. S., and Wennblom, T. 2005. The PRRSV nucleotide sequence database. 2005 International PRRS Symposium, No. 5, Saint Louis, MO, 2005.

One manuscript is in preparation:
Wennblom, T., et al. The PRRSV nucleotide sequence database.

Contact Information:

Kay S. Faaberg, PhD
Research Microbiologist USDA-ARS B-14
Virus and Prion Diseases of Livestock
National Animal Disease Center
2300 Dayton Avenue
Ames, IA 50010
Phone: (515) 663-7259
Fax: (515) 663-7458
Email: kay.faaberg@ars.usda.gov