Meeting Notes and Things To Do

R. Mark Sharp

2021-08-29

Provided in reverse chronological order

20210827

1. Mark presented progress on countKinshipValues.R, which is still had a minor error, which was corrected after the meeting.
2. Matt had already sent a large a deidentified rhesus pedigree that Mark was able to locate. Once Matt has provided a file with potential parents, Mark can use it to test the code with large files. Changes to allow large pedigrees and large simulations are expected. Currently with small pedigrees simulations of $> 100,000$ iterations are possible.
3. Matt presented ideas regarding updating of the current website. Much of the basic material is still accurate but information regarding installation and running the application is dated.
   - Instructions on creating the pedigrees should be largely correct, however, use of Excel files and LabKey are missing from the instructions.
   - There is a need to add some short instructional videos with the idea that the subjects be very specific and the videos be less than three minutes long.
4. After the meeting, Mark corrected the countKinshipValues functions, added a set of unit tests for the function, and added over 20 more unit tests to eight other functions. A clean build: 0 errors, 0 warnings, 0 notes with updated website pages will be uploaded. However, since this is on the issue8 development branch, it should not affect the current GitHub website.
5. Next function for Mark is to create the function to use the output of **countKinshipValues** to get summary statistics and a function to create histograms. The function to create histograms is of lower priority.

20210806

1. Preparation notes
   a. Mark has not created the function to create the Tukey five number summary.
   b. Mark has created **createSimKinships** that uses **make_SimPed** and **kinship** makes a list object of kinship matrices from simulated pedigrees of possible parents for animals with unknown parents.
   c. Mark has created **cumulateSimKinships**, which makes a list object of kinship matrices from simulated pedigrees of possible parents for animals with unknown parents.
2.

20210514

1. Discussed issue 8 - adding fromCenter to calculate unknown parents kinship value
   a. Matt to work on function to find potential sires and dams for animals lacking parents to be used in the kinship simulation, **kinship.R**.
      1. Only object needed by the function is **ped**.

b. Matt to work on a function to find animals born here but lack parents.
   1. The function will create a list object with three vectors of IDs: animals missing both parents, animals missing the dam, and animals missing the sire.
c. Matt to work on a function to find parents for those identified above
   1. Use ped, list, argument
   2. Argument:
      - Sires with any offspring within plus or minus 1 yr of birth.
      - Dams with any offspring within plus or minus 1 yr of birth but did not have an offspring in focal birth year. This will work for rhesus but needs to be made so that the exclusion period is variable. For example, potential baboon dams would need to be excluded if they had an offspring within 6 months of birth. Thus, the plus or minus period needs to be an optional parameter with a default ot 1 year and the dam exclusion period should default to the gestational length for the species.
d. Matt to work on a function to put them together into a list object **allSimParents** made up of lists where the internal list has the offspring ID (**id**), a vector of representative sires (**sires**), and a vector of representative dams (**dams**). In those cases where the sire or dam is known the vector will contain the known respective sire or dam ID.
e. Mark to create a function (**makeSimPed()**) to use the list object of **allSimParents** to make the simulated pedigree along with the associated unit tests.
f. Mark will create a function to run the simulation. It will take the **allSimParents** and use **makeSimPed()** to create a list of simulated kinship matrices. If memory becomes an issue, summary statistics will have to be accumulated. The simmulated kinship coefficients for any one animal will be normally destributed so this is straight forward.
g. Mark will create a function to take the simulated kinship matrices and generate Tukey's five number summary for each.

20201130

1. Mark met with Amanda to discuss the parent exclusion project and to discuss writing of a paper about nprcgenekeepr.
   a. The possible journals she thought of are listed below.
      1. ILAR
      2. Molecular Ecology
      3. Zoo Biology
      4. Journal of Medical Primatology - probably not good for software
      5. Journal of Genomics - check

20201120

1. Mark and Matt met with the Genetics and Genomics Consortium to present a brief summary of work to date and to discuss potential work going forward.
   1. Provided a table of statistics describing the growth of the software from a set of 8 scripts to a mature package published on CRAN.
   2. Near term plans
      1. Plans to correctly handle false founders made up of animals from within the colony that do not have one or both parents identified.
      2. Offered to help with LabKey integration at each site using LabKey.
      3. Offered to train users and improve tutorials based on user feedback.
      4. Will refactor user interface code to use Shiny modules.
   3. Parent exclusion software
      1. Nearly done with conversion of Java scripts to R
      2. Plan to make standalone Shiny Application

3. May integrate into `nprcgenekeepr`
4. Possible long term plan is to provide predictive modeling of breeding strategies via simulation.
5. Obtained requests from users
   1. ARMS connectivity
   2. Oracle connectivity
   3. Add tracking of individual alleles
      1. The internal code already does this if the animal has geneotypes assigned. However, the user interface is minimal and does not provide output regarding statistical summaries of probabilities for those individual genes being in the current potential breeders.
   4. Group from common colony - assign kinship coefficients (matrix) based on genetic data. Mark does not remember the meaning of this request and needs to find someone who can provide more explanation.

20201117

1. Preparation for Friday's meeting with the Genetics and Genomics Consortium.
2. Review of progress to date
   1. `Software changes`

Simple metrics for the Original Version (20160816) and Current Version 1.0.3 (20200526).

Source	Files	Lines	Code	Blank Lines	Document Lines	Comment Lines
Original Files	8	3621	1927	531	0	1163
Version 1.0.3	293	13776	7775	441	4613	947

Simple metrics for the Original Version (20160816) and Current Version 1.0.3 (20200526) and current version groups of functions.

Source	Files	Lines	Code	Blank Lines	Document Lines	Comment Lines
Original Files	8	3621	1927	531	0	1163
Version 1.0.3	293	13776	7775	441	4613	947
1.0.3: functions	163	8272	3007	226	4406	633
1.0.3: ui	11	1782	1506	98	2	176
1.0.3: tests	119	3722	3262	117	205	138

20200423

1. Submission to CRAN was rejected with the following revision requests
   1. considerably reduce the check time of your package to stay below the threshold of 10 minutes. This can for example be achieved by:
      - Omitting the less important and lengthy tests by only running them conditionally if some environment variable is set that you only define on your machine.
      - Using precomputed results for the computational intensive parts in vignettes.
      - Ensure that even after this change the package still includes illustrative examples of the use of the user-facing functions. Please aim at trying to test and also illustrate the use of as much functionality of the package as possible while making considerate use of the computational resources.
   2. Single quotes are used to mark non-English usage, package names, software names, API names and paper titles. Not for names or abbreviations/acronyms. Please omit them where not accurate, e.g. –> Raboin, API,. . .
   3. Please only capitalize names, sentence beginnings and abbreviations/acronyms in the description text of your DESCRIPTION file. e.g. – electronic health records
   4. Missing Rd-tags:
      - addGenotype.Rd: \value

- agePyramidPlot.Rd: \value
- allTrueNoNA.Rd: \value
- assignAlleles.Rd: \value
- calcFE.Rd: \value
- calcFEFG.Rd: \value
- calcFG.Rd: \value
- calcRetention.Rd: \value
- checkGenotypeFile.Rd: \value
- convertSexCodes.Rd: \value
- create_wkbk.Rd: \value
- createExampleFiles.Rd: \arguments,]value
- . . .

5. You have examples for unexported functions which cannot run in this way. Please either add nprcgenekeepr::: to the function calls in the examples, omit these examples or export these functions.
6. Please fix and resubmit, and document what was changed in the submission comments.

2. I will work toward getting these changes in as soon as possible and resubmit.
3. I have not heard from you regarding the user acceptance testing and am curious as whether or not you are going to go forward with that in this period of social distancing.

20200320

1. Notes of things to do prior to submission
   1. Correct spelling package wide - Done 2020319
   2. Update Development_Plans.Rmd
      - Updated and cleaned up a bit on 20200319 in preparation.
      - Sent out a request to Amanda and Matt for input on 20200319. Their input may not arrive before this submission.

20200316

1. Matt will work on colony manager tutorial during last two weeks of March.
2. Mark will provide user acceptance test questions.
3. Mark will provide colony manager tutorial material to Matt.
4. Mark will submit to CRAN during week of March 16, 2020.
5. Amanda will administer the user acceptance testing to collect feedback on using the application. What causes problems, what can and cannot be done, what is missing, what is too complex, etc.

20200203

1. Is the tutorial for CRAN or the Consortium users. need overview of tab structure and order (left to right).
   1. Notes regarding feedback from Amanda on 20200204 I am willing to change "reading" to "uploading", but they do not mean the same thing. When you upload a file, there is no inference of paying attention to the contents. When reading a file, the contents are examined. Reading is the more accurate descriptor in our case.
   I am going to write a separate tutorial explicitly for the consortium members as typical R users would want conventional use of terms and descriptors. For example, Shiny apps are always described as Shiny apps.

20200106

1. Mark will convert nprcmanager to nprcgenekeepr before any other changes.
2. Amanda will provide notes for the Shiny application tutorial with the following goals.
   a. Address the specific needs of consortium members.
   b. Reduce the complexity of the presentation where possible.
   c. Ensure there are no major components left out.
3. Amanda will talk to Dave Lawrence about using his code to guide R development of the *Diversity Report.*
4. Mark will complete submission to CRAN for publication to the consortium.
   a. Mark expects that the package will be accepted on CRAN once examples have been provided for all exported functions. This is not a trivial matter as there are currently 142 exported functions.
   b. Mark will remove exporting of functions that do not need to be exported.
   c. Mark will use code from the unit tests for many of the examples.
5. Mark has completed only the first of the four items from the 20191120 meeting. These will be addressed after the submission to CRAN is successful.

20191218

1. Added ability to export each of the six figures on the **Summary Statistics** page.
2. Need to update shiny application tutorial to reflect the ability to export each of the six figures on the **Summary Statistics** page.

20191130

1. See if https://www.semanticscholar.org/paper/Management-of-genetic-diversity-using-gene-dropping-Khaldari-Javaremi/43dd1975193830b9243c8eef672db134e42e871d/figure/3 is figure that would be beneficial.

20191120

1. Export all graphics
2. Create, error check and export dawn of time pedigree
3. Add heuristic to removed recent progeny as founders.
   a. Add a column - something like "from center" so that "born at site" can be ascertained.
4. Provide a way to clear focal animals added in **Pedigree Browser** tab so that full pedigree is again available. This will also allow a new pedigree to be added via the **Input** tab.

20191115

1. Clear out empty text field (*Filter View*) on __Genetic Value Analysis___ tab. Done 20191115

20191111

1. Send example pedigrees with errors to Amanda.Done 20191111
2. To meet Friday at 4 PM Pacific to go through entire presentation.
3. Code changes since August 16, 2016.
   - Initial package construction was in March of 2017.
   - Start of code additions in August 2017.
   - There are now 736 unit tests covering over 92% of the code.

Table 1: Simple metrics for the Original Version (20160816) and Current Version 1.0.5.9000 (20210421).

Source	Files	Lines	Code	Blank Lines	Document Lines	Comment Lines
Original Files	8	3621	1927	531	0	1163
Version 1.0.5.9000	309	14687	8334	478	4910	965

Table 2: Simple metrics for the Original Version (20160816) and Current Version 1.0.5.9000 (20210421) and current version groups of functions.

Source	Files	Lines	Code	Blank Lines	Document Lines	Comment Lines
Original Files	8	3621	1927	531	0	1163
Version 1.0.5.9000	309	14687	8334	478	4910	965
1.0.5.9000: functions	170	8660	3107	229	4686	638
1.0.5.9000: ui	11	1790	1514	98	2	176
1.0.5.9000: tests	128	4237	3713	151	222	151

- Two tutorials have been developed. One to guide R users in the interactive use of the package functions (public API) and one to guide users of the Shiny application.
- Available via GitHub with MIT license.

Usually it is recommended to use column_spec before collapse_rows, especially in LaTeX, to get a des

20191014

1. **Progress Since Last Meeting**

    a. Code changes
       - Added filter to pedigree IDs available for breeding group formation so that only animals at the institution (exit == NA) and animals with recorded birth dates (birth != NA) are potential breeding group members.
    b. Documentation Updates
       - Shiny application tutorial first draft is nearly complete. Breeding group formation remains.
    c. CRAN submission preparation
       - Have begun using RHUB tools to prepare for the CRAN submission.
         – library(rhub)
         – library(usethis)
         – cran_prep <- check_for_cran()
         – cran_prep$cran_summary()
         – usethis::use_cran_comments()

2. **Meeting Notes**

    a. Mark will fly up on 19th leave 21st
    b. Need to ensure access to WiFi and video.
    c. Finish Shiny tutorial
    d. Recheck all files to ensure animals are deidentified via obfuscations of IDs, birth dates, and exit dates.
    e. Plan to have a CRAN submission prior to November 20, 2019, meeting.
    f. Want to develop a better name. This should be done prior to CRAN submission.
    g. Matt is to test LabKey connection.
    h. Amanda and Mark will see if Wayne can move forward on Mark's access to PRIMe
    i. Mark will work on **Genetic Production**, **Genetic Value Analysis**, and **Founders** as described in the meeting notes for 20190916.

20190916

1. **Production Calculation**

   Amanda corrected and clarified how to calculate *Production* using the

   a. The Production Status is calculated on September 09, 2019.
   b. Births = count of all animals in group born since January 1, 2017 through December 31, 2018, that lived at least 30 days. Animals born after December 31, 2018, are not counted.
   c. Dams = count of all females in group that have a birth date on or prior to September 09, 2016.
   d. Production = Births / Dams
   e. Production Status (color)
      1. Shelter and pens
         1. Red – "< 0.51"
         2. Yellow – ">= 0.51 & < 0.54"
         3. Green – ">= 0.54"
      2. Corrals
         1. Red – "< 0.61"
         2. Yellow – ">= 0.61 & < 0.65"
         3. Green – ">= 0.65"

2. **Genetic Production**

   Percent of females ">= 3.5 at the start of the 2 calendar year period that have not produced offspring in the past 2 calendar years. Filter out animals over ?? (start with 18) years old."

3. **Genetic Value Analysis** Have Mark look at algorithm and develop simple options for handling progeny that are missing parental information.

   One option is to not calculate genetic value of animals < minParentAge.

4. **Founders** We need to get a better understanding of the issues surrounding the identification of founders and how founders affect the genetic value analysis. As a first step Mark will develop code to list the founders and the count of male and female founders. See how they affect the genetic value analysis.

   In the very short table below **x123** has a founder as a dam while **x122** does not have a founder for a sire because of the respective values of *dam_from_center* and *sire_from_center*.

id	sire	sire_from_center	dam	dam_from_center	birth
x123	s123	Y		N	1988/04/21
x122		Y	d123	Y	1989/08/18

5. **Matt Schultz and R Setup for nprcmanager**

   Mark is to contact Matt to arrange a time to work with him to make sure he get nprcmanger set up as a repository on his computer.

   Mark will also send the interactive tutorial to him and Amanda. Done 20190916.

20190826

1. Genetic Diversity Graphic
   a. Labels Done 20190908
      - Breeding Group
      - High-Low - Value
      - Indian Origin - Origin
      - Fecundity - Production

- Kinship With Male - Inbreeding
- Genotype Phenotype - Flags
  b. Labels of groups on left
  c. Labels of genetic diversity labels top and 45 degree angle
  d. Genetic Diversity - Genetic Diversity Report
2. Send list of errors that can be detected. Done 20190826
3. Send rhesus MHC data pedigrees Done 20190826
4.

20190810

This was an *ad hoc* meeting to get clarification on heat map development for genetic diversity reporting. This meeting was a response to questions in a July 18, 2019, email from Mark in response to an email from Amanda with the subject of "genetic diversity reporting questions".

As a result of this meeting, the following columns are defined as follows

1. GENETIC DIVERSITY REPORTING
   1. What are the proportions of high and low genetic value breeding-age adults in the group?
      1. RED > 0.5 LOW VALUE
      2. 0.5 >= YELLOW >= 0.30 LOW VALUE
      3. GREEN =< 0.3 LOW VALUE
   2. Are all members of the breeding group Indian-origin rhesus macaques?
      1. RED – >= 1 HYBRID OR CHINESE ANIMALS
      2. YELLOW – >= 1 BORDERLINE HYBRID ANCESTRY & 0 HYBRID OR CHINESE ANIMALS
      3. GREEN – 0 HYBRID OR CHINESE ANIMALS & 0 BORDERLINE HYBRID ANCESTRY = 10% <= 15% chinese: Proportion of chinese ancestry is borderline Yellow; Red > 15%; This information is found under animal history under genetics main tab – genetic ancestry.
   3. Among all females in group age >= 3 years what is percentage with a kinship coefficient <= 0.0156 with at least 1 male >= 5 years old in the group.
      1. Denomonator = count of females >= 3 years of age
      2. Numerator = count of females with kinship coefficient <= 0.0156 with at least 1 male >= 5 years old in the group.
      3. Thresholds
         1. Red – < 0.6
         2. Yellow – >= 0.6 & <= 0.9
         3. Green – > 0.9
   4. Fecundity
      1. Denominator = count of females >= 3 years of age
      2. Numerator = count of births that live > 30 days
      3. Thresholds
         1. Shelter and pens
            1. Red – < 0.6
            2. Yellow – >= 0.6 & <= 0.63
            3. Green – > 0.63
         2. Corrals
            1. Red – < 0.5
            2. Yellow – >= 0.5 & <= 0.53
            3. Green – > 0.53
   5. Flagged for genotype of phenotype
      1. Thresholds
         1. Red – >= 3 group members flagged
         2. Yellow – < 3 and >= 1 group members flagged

3. Green – 0 group members flagged

20190715

1. To prepare for the meeting
    1. Work on tutorial to go through the Goldilocks path.
    2. Work on tutorial to look at all possible data input errors
    3. Submit new version to GitHub and Travis-ci
2. We discussed heat map type issues with regard to realtime examination of kinship relationships within breeding groups
    1. Amanda is to send Mark a pedigree with breeding group information and an updated version of the questions in an Excel workbook
    2. Mark is to use those data and the questions to develop some visualizations for Amanda's review.
3. The interactive tutorial was discussed a bit.
    1. Priscilla Williams, a current graduate student, who Mark has worked with for several years is working through the tutorial.
    2. This discussion brought up the topic of making use of the Shiny application as simple as possible.
4. Mark is going to put together the Shiny application tutorial next

20190603

1. Joint Working Group Meeting week before Thanksgiving. We will meet with the Breeding Colony Managers Group. November 19-21, 2019.

2. Work on loop code: have list of animals involved in loops with counts, number of loops, number of animals in loops. Not high priority.

3. DCM wants to report on problems in current breeding groups prior occurrence.

    - Amanda has developed an initial set of questions.
        – Are all Indian origin?
        – Has alpha male been in group > 3 years.
        – Are kinship coefficients of >=0.0156 between male and females >= 2.5 years old (settable) within a breeding group.
        – Are kinship coefficients of >=0.0156 between males >= 2.5 years old (settable) within a breeding group.
        – Are offspring equally distributed among female breeders.
        – Are any members flagged to be genotyped. This likely does not reside in demographics.

4. Does Amanda know what the *2014-10-16_ResPed_v1.1.txt*, *BreedingGroups1_4MendozaTest.csv*, *Jmac_studbook_20180711.csv*, *Jmac_studbook_20180711.txt*, and *MendozaC1C2newharemstest.csv* files represent? Can they be use for instruction? We are not going to use them. Done 20190622 Are some to be used with the Example_Pedigree file? We are not going to use them. Done 20190622

5. Decide which pedigree files to leave in package. Done 20190622

    - All pedigree files except an obfuscated baboon (qcPed) and an obfuscated rhesus (examplePedigree) were moved out of the package.
    - The qcPed and examplePedigree are in the package as data, which can be obtained by the user with
        – **qcPed <- nprcmanager::qcPed**
        – **examplePedigree <- nprcmanager::examplePedigree**
    - Three versions of the qcPed pedigree were included as actual files for the getPedigree() unit tests.

6. Figure out how to best provide example pedigree files. See point immediately above. Done 20190622

7. Items brought forward from previous meetings to prevent forgetting them.
   1. From 20190311
      - Amanda sent Mark the 2015-02-14_Genetic_metrics_white paper_Final.docx paper for him to review and collect ideas for a renewed ORIP Reporting tab. Amanda suggested that we need to propose collection of these metrics from all primate centers. Genetic and Genomics Working Group website: https://nprcresearch.org/primate/genetics-genomics/genetics-genomics-working-group.php
   2. From 20190408
      - Consider adding the ability to ignore founders if the user selects that option.
      - Go through all primary data structure names and definitions to update and correct.
      - Edit all function descriptions to update and correct

20190429

1. Updated Amanda's R, R packages, and installed version 0.5.10 (20190428) of nprcmanager.
2. Demonstrated that new version of nprcmanager displayed suspicious parent table in ErrorTab. Done 20190428
3. Decided on new wording of "One or both parents are below the minimum parental age. Check both parent and offspring birth dates." to place before the Suspicious Parents table. Done 20190428
4. Removed row label column in Suspicious Parents table. Done 20190428
5. Plan to add ability to read in pedigrees in Excel format to Input tab. Done 20190519
6. Still need to complete items listed in 20190408 meeting notes. (See below.)
7. Will provide a "Goldilocks" tutorial of Shiny application first.

20190408

1. Test for error when pedigree is not available when forming groups
2. Make sure harems are formed correctly when males are in seed animals
3. List of things found prior to the meeting to do before next meeting:
   - Unit test for fillGroupMembersWithSexRatio() Done 20190328
   - Go through all primary data structure names and definitions to update and correct.
   - Edit all function descriptions to update and correct
   - Correct nprcmanager.R file. This includes the following and more. Check, correct, and create where needed function lists including lists of all functions, pedigree file testing functions, genetic value functions, plotting functions, breeding group formation functions, and gene dropping functions (if possible). Done 20190515
   - Consider making a summary function for the gvReport structure. Done 20190602
   - Consider adding the ability to ignore founders if the user selects that option.

Ballou & Lacy describe genome uniqueness as "the proportion of simulations in which an individual receives the only copy of a founder allele." We have interpreted this as meaning that genome uniqueness should only be calculated for living, non-founder animals. Alleles possessed by living founders are not considered when calculating genome uniqueness.

We have a differing view on this, since a living founder can still contribute to the population. The function below calculates genome uniqueness for all living animals and considers all alleles. It does not ignore living founders and their alleles.

Our results for genome uniqueness will, therefore differ slightly from those returned by Pedscope. Pedscope calculates genome uniqueness only for non-founders and ignores the contribution of any founders in the population. This will cause Pedscope's genome uniqueness estimates to possibly be slightly higher for non-founders than what this function will calculate.

20190311

1. Remove explanatory text above sex ratio radio buttons. Done 20190311
2. Ignore kinship between females at or above the minimum parent age (Yes/No). Done 20190311.
3. Invoice for Jan 12 - Mar 11, 2019 sent 20190311.
4. Renamed (Pyramid Plot to Age Pyramid Plot) and moved to the right of Pedigree Browser. Done 20190311
5. Amanda sent Mark the 2015-02-14_Genetic_metrics_white paper_Final.docx paper for him to review and collect ideas for a renewed ORIP Reporting tab. Amanda suggested that we need to propose collection of these metrics from all primate centers. Genetic and Genomics Working Group website: https://nprcresearch.org/primate/genetics-genomics/genetics-genomics-working-group.php
6. Still have two features to add and one graphic element to correct.
   - Allow user to enter the number of groups of seed animals to be entered. See *Seed Groups* from last month's notes. Done 20190406
   - Add **minParentAge** to check box for *Use minimum parent age for age of animals to be grouped with the mother*. Done 20190406
7. To meet again on 20190325 if Mark makes sufficient progress otherwise we will meet on 20190408 at 4PM PAC.

20190225

1. Mark to add additional unit tests for new seed group formation code.
2. Items to add to instructions:
   a. Discuss what to do about overlapping group formation. Instruct the user to run the Make Groups command again to get a new set of groups. The new set will overlap with the prior results, but will be mutually exclusive within a run.
3. Mark to send an invoice at end of February. Amanda will send the amount of funding available. Done 20190311
4. Seed Groups
   - Have user enter number of seed groups to form and then present that number of windows.
   - Change label to "Optional: Seed Groups with Specific Animals". Done 20190311
   - Have it in its own gray box
5. Pull downs to right in one column: Done 20190309
   a. Number of groups desired
   b. Animals will be grouped with mother that are below age (years): [] Use minimum parent age. <– default – Have not figured out how to have **minParentAge** from Input tab to show up in the BreedingGroupFormation tab. Have dropdown disappear if checkbox above is selected.
   c. Animals with kinship above . . . .
   d. Dropdown with
      - Ignore females at or above the minimum parent age
      - Do not ignore females at or above the minimum parent age
   e. Modify kinship dropdown to add common relationship nomenclature in parentheses.
   f. Number of simulations
6. Left column input: Done 20190309
   a. Make Groups in color and larger
   b. Enter the group to view
   c. Export Current Group
   d. Export Current Group Kinship Matrix
7. Example emails
   a. Example: Whether or not one of 10 females would work with a particular male.
   b. Amanda is to collect other similar emails to use as ideas for development of vignettes.
8. Next meeting March 11, 2019, at 4PM PAC.

20190114

1. Amanda provided a new logo to use, which represents all of the National Primate Research Centers. This is necessary because the ONPRC parent organization does not allow them to have a logo. Mark will incorporate that logo into the application and add some comments regarding development supported by ONPRC and SNPRC. Added 20190119.

2. Discussed harem formation and formation of groups with a set sex ratio. Done 20190103.

3. Discussed what to do about overlapping group formation. Decided to instruct the user to run the Make Groups command again to get a new set of groups. The new set will overlap with the prior results, but will be mutually exclusive within a run. Added to things to do for 20190225 meeting 20190224

4. Discussed layout of breeding group formation tab by using a mockup tool. Decided to have six workflows selectable in a radio button configuration. The user interface will redraw based on the radio button selected to reflect the user interface elements needed within the selected workflow. Done 20190224

5. Need to test group formation with a candidate animal not in the pedigree or not in the genetic analysis. Found that this causes a application crash. Done 20190224

6. Need to test group formation with a candidate animal not in the pedigree or not in the genetic analysis. Found that groups are not formed, but an error is not displayed. The user must remove the false Id to get group formation to work. Done 20190224.

7. Make an error reporting function that informs the user what has occurred when an animal not in the pedigree or not in the genetic analysis is entered as a seed animal or a candidate animal.

8. Change "Animals will be ignored below this age" to "Animals will be grouped with the mother below age". Done 20190224.

9. Change to Animals with kinship above this value will be excluded. Done 20190224.

10. Change to Include kinship in display of groups. Done 20190224.

11. Disable breeding group formation tab until genetic value analysis has occurred. Have disabled tab display text explaining that genetic value analysis must be performed first.

12. Have 6 boxes for seed animals. User can provide seed animals for up to six groups. Done 20190224

13. The next meeting will be 20190211 at 4 PM Pacific time.

20181210

1. Mark reported that he was having trouble implementing the user interface elements needed for items 4 and 5 of the 20181105 meeting notes. Specifically the need to have dynamically generated text input boxes for the number of groups to be formed is the issue. He has not figured out either the desired interface behavior or the technical methods needed for this feature. This is going to delay getting the tutorials written, which he had hoped to complete in first draft form during December.

After some discussion, all agreed that the features described in items 4 and 5 are sufficiently important to delay the preparation of the tutorials and subsequent technical paper a few weeks. Thus, Mark will work on getting these implemented in December with plans on working on the tutorials in January.

Ability to form harems. Done 20181230.

Ability to form breeding groups with specified sex ratios. Done 20180103.

1. Amanda and Mark had fairly extensive discussions regarding how group selection is to be done. The current software forms groups so that there is no overlap of animals among any of the groups. The plan is for Mark to implement a group selection procedure that will make two types of groups.

One group type is what is currently done by the software (no overlap of animals among any of the groups formed.) These groups will be made up of all of the animals that can be used in groups based on kinship and sex ratio criteria such that none of the groups have any individuals in common.

The second type of group is a collection of sub-groups where each sub-group is a group of the first type. Thus, there is no overlap of animals within any one sub-group of groups and there is potential overlay among the various subgroups. This is complected and hard to follow so it is illustrated with the list below where each letter represent a specific animal. Overlapping Groups 1-4 have within them sets of animal that have no animals in common with other unique sets within the group.

For example, animal **H** appears in set 1 of each overlapping group, but it does not appear more than once in any on group. Also, animal **E** appears in group 1, 2, and 4 and not in group 3.

- Overlapping Group 1 - Unique Set 1: I,B,J,L,H - Unique Set 2: D,G,U,S,Q - Unique Set 3: E,P,F,A,C
- Overlapping Group 2 - Unique Set 1: M,V,H,Z,K - Unique Set 2: O,T,W,D,X - Unique Set 3: C,J,E,F,B
- Overlapping Group 3 - Unique Set 1: H,I,F,X,K - Unique Set 2: Q,Y,A,D,S - Unique Set 3: T,B,C,G,M
- Overlapping Group 4 - Unique Set 1: H,V,I,T,Z - Unique Set 2: E,W,M,N,O - Unique Set 3: P,Q,J,G,U

1. Amanda and Mark are to develop a better descriptor for the button that causes the pedigree information to be read in tested and sent to the pedigree browser function. Mark has changed it to *Read and Check Pedigree* for now. Done 20181212.
2. Mark is to reorganize the **Input** tab to place the description of the minimum parent age in hovertext and make the check pedigree button more evident. Done 20181212.
3. The next meeting will be 20190114 at 4 PM Pacific time.

20181105

1. Do not report as an error the wrong sex for animals added into the pedigree and appear as both a sire and dam without an ego record. These need to be reported as an error because they are both a sire and a dam. Done 20181208
2. Make a combined logo for Oregon and SNPRC. Have ONPRC on top using blue and green. Check on the University of Oregon website for a color palette. Have ONPRC in the lighter color similar to the SNPRC color scheme. Have the macaque and oval in the same blue color. Clean up the resolution of the Oregon logo as it is currently fuzzy. Some research found the following link OHSU Color Palette.

- Alert #C34D36
- Body text #555555
- Button text #805B16
- Button pre-fade #FFD769
- Button post-fade #FFC529
- "Currently accepting patients" marker #67B445
- Desktop navigation hover (darker blue) #093561
- Desktop primary navigation #0E4D8F
- Footer background #0E4D8F
- Footer link text #B9B5B5
- Form error handling #C34D36
- Glossary Definition background #FDFDE2
- Glossary Definition border #F5F26B
- Link #0072FF
- Link hover state #0E4D8F
- Navigation active (current page) link #A4CAFA
- Navigation background darker #E5E4E4
- Navigation background lighter #F3F3F3
- Promo background #F3F3F3
- Subsite name in header type #0E4D8F Done 20181208

3. Check the outliers found in the first box and whisker plot when the file "Example_Pedigree.csv" is analyzed. The outliers are exactly as described: Data beyond the end of the whiskers are called "outlying" points and are plotted individually. The jittered points are overlaid on top of the boxplot and its outliers. Done 20181112

4. Add the ability to select a sex ratio for group formation. Ratios are to be female to male starting with 1:1 and progressing by 0.5 on the female side, holding the male side at 1. Go up to a ratio of 10:1. The progression will look something like 1:1, 1.5:1, 2:1, 2.5:1 ... 10:1.

5. Add the ability to use harems with 1 male and any number of females. One way to do this would be to put any number of males in with the females in the list of candidates and then specify that single male harems are to be formed with a specified number of groups formed. This will allow the select of the best males to be selected from the excess male candidates.

6. Beth is to contact Mark at least a week prior to the next meeting to set up a time to diagnose why her installation of the software is not working. Rcpp package was not being replaced during installation and update of other packages. This prevented nprcmanager from being installed. Beth manually removed the old version of the Rcpp package, installed a newer version and then installed nprcmanager and rmsutilityr. Mark will do some research to see if others have reported similar problems. Working with Beth prompted the development of a function that reports the version of the application on the input tab so users can easily find out which version they are using. Done 20181109.

7. Until we work on LabKey connectivity again, we have decided to postpone Item 2 from two meetings ago: Give user the option to save a skeleton configuration file to their home directory if they do not have a configuration file.

8. We reaffirmed the plan for Mark to finish the tutorials (one for an interactive use of the software and one for using the Shiny application) and to get a draft of a technical paper written by the end of the year.

9. Not discussed during the meeting was the need to develop additional unit tests to cover all of the new functions created to handle the PEDSYS and military formatted dates (YYYYMMDD), which look like an integer. Most of those unit tests have been made, but additional ones are needed to provide full or near full coverage. Done 20181112

10. Miscellaneous

- Found that breeding groups being formed included unknown animals that had been added as placeholders for unknown parents. Those were removed from consideration. Done 20181119
- Worked with Terry Hawkins on 20181119 and 20181120 to make sure the LabKey code was working. He had an error in his base URL specification so that initial efforts failed and the error message was misleading. I have trapped the error with a tryCatch function and am sending a message to the log file. This needs to be tested.

11. The next meeting is scheduled for 20181126. Due to conflicts in Mark's schedule, we rescheduled for 20181210.

20181022

1. Allow the use of Ego for ID. Done 20181022
2. Remove the display of the change columns. Done 20181103
3. Remove old error tabs each time a file browsing occurs. Done 20181103
4. List columns that should be there id, sire, dam, sex, birth when columns are listed as missing. Done 20181103
5. Improve the reporting of multiple errors of the same type. Perhaps "The first 5 records have the following errors: . In total there were records with the same type of error. Please check and correct the pedigree file." Changes are improved over the suggestions above, but similar. Done 20181103
6. Correct "Genetic Uniqueness Score" to "Genome Uniqueness Score". Done 20181022
7. Item 2 from last meeting list: Give user the option to save a skeleton configuration file to their home directory if they do not have a configuration file.
8. During the next meeting, we will work with Beth on clearly defining how breeding group formation

tab should work. In the mean time, Mark will examine the current behavior and code so that he understands exactly what is being done.

9. The near term goal is to produce a technical paper within the next six months that describes the software. Mark will get this started by completing two tutorial documents. The first will be a tutorial on how to use the major functions from the R console in an interactive mode. The second will be a tutorial on how to use the Shiny application. The first tutorial should be completed some time in December.

10. Amanda and Mark discussed the goal of enhancing the scope of software to include the ability to do phenotype – genotype association studies. Amanda has some ideas regarding direction. Mark needs some guidance as to what techniques to investigate.

Below are some relevant articles found from a simple web search with the Google search engine using the search elements: "genetic association studies in pedigrees" * https://www.stats.ox.ac.uk/~mcvean/gwa4.pdf * https://dx.doi.org/10.1186%2F1753-6561-8-S1-S26 * https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1459209/ * https://www.ncbi.nlm.nih.gov/pubmed/12687644 11. Not mentioned during the meeting but taken from other correspondence. Mark will Contact Daniel Nicolalde (fdnicolalde@primate.wisc.edu; (608) 890-4592; INFORMATICS AND DATA SERVICES) to see if he will assist in getting the LabKey interface working with the Wisconsin National Primate Research Center system. 12. The next meeting is scheduled for November 5, 2018 at 4 PM Pacific Time.

20180917

1. Change Pedigree browser button label to say "Trim pedigree based on focal animals". Done 20180917
2. Give user the option to save a skeleton configuration file to their home directory if they do not have a configuration file.
3. Incorporate the error detection functions into the shiny application so that if a defective pedigree file is selected by the user a new tab is created with a description of the errors detected.

a. The user should be moved to the error report tab. Done 20181020
b. The error report tab should be placed immediately after the input tab. Changed to immediately before the input tab. Done 20181020

4. Beth is to try the current (0.3.31) or later version of nprcmanager with the pedigree file that has been failing.

a. If that file still causes an error, Beth is to send it to Mark. Received file from Beth 20181018.
b. Mark will change the code to better handle the issues observed, update the package, and have Beth test again. All code changes made for all known file errors completed 20181020.

20180820

Meeting was canceled.

Topics I wanted to bring up.

1. Check on ability of Beth and Amanda to install package.
2. Demonstrate use of **qcStudbook** in interactive mode.
3. See if I should follow up with users who have had trouble loading a file.
4. Demonstration to genetics and genomics group
5. Breeding group formation. Workflow and goals
6. Genetic value analyses. Workflow and goals

20180711 Meeting Notes

Summary Statistics tab

1. Use default box and whisker plot with red dots for outliers.

a. Changes made. Done 2018-07-14
b. Hover text to box and whisker plots added. Done 2018-07-14

Pedigree Browser tab

1. Change Update breeding colony button to " " Waiting on Amanda. Changed to "Update Focal Animals" based on instructions from Amanda. Done 2018-08-04
2. Add function to "Update Breeding Colony" button to browse for a file; add animal IDs to the window where they could be pasted. Started 2018-07-29 getting file not updating window. Done 2018-08-15
3. Change first bullet point to reflect change in function of the "Update Focal Animal" button. Done 2018-08-15
4. Move first bullet point to far right. Done 2018-08-15
5. Remove second bullet point. However, provide this information when the file has any type of error. This is complete for one error at a time. Currently the **qcStudbook** function does not continue scanning the file when an error is found. Done 2018-07-17
6. Change "Display UIDs for partial parentage" to "Display UIDs". Consider coming up with a better name than "UID". Add hovertext to explain what this button does. Proposed explanatory text: "Unknown IDs are created by the application for all animals with only one parent. They begin with a capital **U**." Done 2018-08-15
7. Change "Trim pedigree based on specified population" to "Trim Pedigree". Add hovertext to explain what this button does. See what "Trim Pedigree" button does. Does it remove animals? Its intent is to reduce the number of animals being examined. It does reduce the number by removing all individuals not related to the focal animals. Does it change genetic analysis. It does not change the genetic analysis for the animals left. Obviously for animals removed, it does. The name of this button will likely follow from what the new name for "Update Breeding Colony" button. Done 2018-08-16
8. Add a better description for the Search field. Done 2018-08-16

Error Handling

1. Develop a full set of error messages appropriate for all types of input files.

a. Added improved error detection and reporting for missing required columns - Done 2018-07-17
b. Added the ability to call qcStudbook interactively with a flag (**reportErrors**) set to TRUE that causes qcStudbook to report back a list with all errors detected within the input file. Not all errors can be detected in a single pass because some errors preclude checking for others. For example, if the birth date column has an invalid format, the parental age cannot be checked. If no errors are found with **reportErrors** set to TRUE and NULL value is returned. **reportErrors** defaults to **FALSE**.
c. This emphasizes the need for a good set of corrupt input files for testing and unit test development.

2. Amanda and Beth will provide me with some files with errors.

a. Amanda provided a file with two error types: bad birth date type - integer instead of character representation of a date and animal appearing as a sire and a dam. Done 20180720
b. Both errors were caught, but the first error was only seen after correcting the first. It would be better to report all errors and set a flag that told the program to stop at the end of parsing the entire pedigree.
c. Three files were added to the example files: *Example_Pedigree.csv*, *Pedigree_File_Example_134M_dam_removed.csv*, and *Pedigree_File_Example_CSV.csv* Done 20180721
d. These files need to be documented.

3. Refactored some code within **qcStudbook** to a separate function **unknown2NA**. Done 2018-07-18
4. Added bad date string format detection in **convertDates**, which is now **convertDate**. Also added unit tests for **convertDate**. Done 2018-07-21
5. Added ability to handle dates with NA to **convertDates** Done 2018-07-22
6. Added unit test coverage of **updateProgress** calls in *test_reportGV.R* Done 2018-07-22
7. Corrected a bug and added corollary unit test for **rankSubjects**. Done 2018-07-29
8. Added test for missing **age** column in *orderReport* function. This also allows correct ordering when the age column is missing. Done 2018-07-29
9. Added unit tests for *findLoops* function. Done 2018-07-30
10. Added check for class on objects in *convertDates* function, because it was failing when objects were already dates. It now passes those objects through untouched. Done 2018-08-04
11. Added some debug code for updating focal animals (still named breeding colony update in code) Done 2018-08-15

Multiple tabs

1. Come up with a different word for "Breeders". What do you think about *Managed Animals*

a. Amanda suggested *Focal Animals* – this is being adopted.

Miscellaneous items

1. Added automatic generation of a web site

a. 2018-07-17 first draft
b. Need to first move content that has been prepared to the correct location. Done 2018-08-16

2. Changed read.csv to read.table in code to emphasize we are able to read multiple file types.

a. 2018-07-17

3. Amended documentation for the **addGenotype** function to indicate that it is assuming the *genotype* object was opened by **checkGenotypeFile**. 2018-07-17
4. Added use of convenience functions **get_and_or_list** and **is_valid_data_str** from the *github/rmsharp/rmsharp/rmsutilityr* repository. 2018-07-21
5. Added documentation for data elements **finalRpt** and **rpt**, which are both created by the *reportGV* function. 2018-07-29
6. Added brief tutorial on how to use *findLoops* function. 2018-08-04

Questions that came up after the meeting

1. What is the meaning of the following:

a. Text under **Export** button on *Pedigree Browser* tab. "A population must be defined before proceeding to the Genetic Value Analysis''.
b. Check box above **Export** button on *Pedigree Browser* tab. "Trim pedigree based on specified population''. Its intent is to reduce the number of animals being examined. It does reduce the number by removing all individuals not related to the focal animals. Does it change genetic analysis. It does not change the genetic analysis for the animals left. Obviously for animals removed, it does. Done 2018-08-16

20180611 Meeting Notes

Input tab

1. Move Column content into description column separate with colon - Done
2. Change Column Name to Allowable Name - Done
3. Names have alphanumeric plus "_","-", and" ". - Done
4. Remove " Other characters have not been tested." - Done
5. Change "Any animals listed as a Sire or Dam that do not have their own row or line entry as an Ego will be added." to A new row entry will be added for any Sire or Dam that do not already have their own row as an Ego. - Done
6. Change "Parents will be checked to ensure their own Ego entry is the correct sex." to "Animals will be checked to ensure that their sex is consistent throughout the file."" - Done
7. Use Allele_1 Allele_2 - Done

Summary Statistics tab

1. Remove background on histograms - Done
2. Add box and whisker plot to right of summary statistics histograms - Done
3. Use Glossary for definition of terms

20180501 Meeting Notes

1. The second bullet point under the Input File Handling section of the Input tab says "Please be aware that animals with no parents will be treated as founders in these calculations, i.e., sources of new genetic variation in the colony."

I am assuming that this is fine and stated only to keep users aware that if they have parental information it has to be entered.

2. The third bullet point under the Input File Handling section of the Input tab says "Designation of two alleles is required as currently there is no accommodation for partial genetic information for an individual."

Is this will be a problem that needs to be addressed?

3. Reminder for Mark only: The third bullet point under the Input File Handling section of the Input tab says "If the Age column is provided, the program will use the user-specified age."

Make sure this is true. I may also want to check for internal inconsistencies.

4. In the Input File Handling section of the Input tab the last bullet point immediately above the first table says "Genotype data may be supplied within the pedigree file or in a separate genotype file. Only two additional columns (first and second) are required when the genotypes are provided within the pedigree file."

This should be explained more fully and the program behavior may need to be changed. Currently the program assumes the columns first and second are there and are correctly formed. Currently there is no check to ensure that first_name and second_name data elements are consistent with first and second data elements.

5. In the Input File Handling section of the Input tab and under the first table of column names, there is a section describing what is done by the qcStudbook function to correct data and create new columns.

Should we provide a summary of changes made to the input data such as changes in column names, sex identifiers, etc., removal of duplicated rows and columns that are added?

Need name for genotype (first_name, second_name) Need new name for breeders only file type. Reword "Animals without birth dates are not considered." When minimum parent age is included, animals without birth dates are not rejected.

Change update breeding colony (pedigree browser) by adding explanatory text.

Miscellaneous Accomplishments

These should be stored elsewhere

20180429   Test coverage is 86.32%.

20171004   Unit test coverage is 84.79%.

20170921   Unit test coverage is about 15%.

20170919   We have successfully installed and used *nprcmanager* on Microsoft (MS) Windows 7, MS Windows 10, and MacOS 10.12.6 running R 3.4.1.

20170917   I discovered why the plot of genetic uniqueness scores was not as expected. It was from an analysis where 0 was the threshold setting, which I think we should consider removing as an option. I have set the default to 3, which is what Amanda Vinson's paper indicated that ONPRC typically used.

I got the logging system integrated into the package. I am using the package futile.logger (funny name). Note the check box at the bottom of the side panel on the two attached images. When the Debug on check box is checked (it is not checked by default), the application writes to a file in the users home directory named *nprcmanager.log.* Currently I am only logging events occurring the the *server.R* file, because that is where I tend to have most of my problems show up. I have attached an example file in which turned on logging at the debug level and then read in a couple of pedigree files.

I added code coverage reports to the automated build system running on Travis-CI.org. Currently we only have 5.68% of the code being tested with the unit tests I created thus far. I do not know the code well enough to know what percentage we will have when I feel comfortable, but I know most of the functions have no tests at all.

20170916   Can now assign known genotypes to individuals and to incorporate that information into the gene dropping routine. This has been done at the function level in the *geneDrop* function. The pedigree submitted via the UI can optionally contain genetic information.

20170915   Added a genetic uniqueness plot to the Summary Statistics tab.

20170911   I have connected the Travis CI (continuous integration) tool to my github.com/rmsharp/nprcmanager so that when I check in code it automatically tries to build it from scratch using packages from CRAN. Travis CI is now automatically building nprcmanager without errors or warnings. This puts us considerably closer to being able to offer the package via CRAN if we want.

20170323 Meeting notes

I met with Jack Kent, Deborah Newman, and Charles Peterson about how to use genetic data in a gene dropping simulation.

From: R. Mark Sharp msharp@TxBiomed.org

Subject: Re: Genetic management of colonies

Date: March 23, 2017 at 5:01:07 PM CDT

To: Jack Kent jkent@txbiomed.org

Cc: Deborah Newman dnewman@txbiomedgenetics.org, Charles Peterson charlesp@txbiomed.org

Jack, Debbie, and Charles,

Thank you for meeting with me. I appreciate your help in clarifying the various stages of the issues associated with providing genetic management guidance via simulation with partially known MHC data.

1. I will look more closely into what will be needed to add genetic information and gene frequency information into the gene dropping routines in the kinship2 package. The initial implementation will use only a single locus, which should be a sufficient for MHC data

   A. First step will be to ensure we get expected proportions of genes in children of individuals when we provide genetic data to all founders using a gene frequency based algorithm.

   B. Second step will be to ensure we get expected proportions of genes in children when all parental generations have fully known genotypes.

   C. Third step will be to ensure we get expected proportions of genes in children when parental generations have partially known genotypes and all other genotypes are uninformative.

   D. Fourth step will be to ensure we get expected proportions of genes in children when parental generations have partially known genotypes and remaining parental genes are determined by gene frequency.

2. We will not worry about getting rid of genes in the pedigrees since colony managers need only not breed animals that carry unwanted alleles.

3. I will produce a small number of pedigree drawings using the data Debbie has provided to get feedback on preferences and additional requirements.

4. I will not deal with paternity and maternity issues within the pedigrees at this time.

Thanks again.