Daily Records

Caccone PostDoc

Gus Dunn

January, 2015

# 2015-01-02 (Friday)

## Writing Methods

* [x] Functional Annotations
  + done-ish at 2015-01-02 08:58
* [ ] Linkage
  + **STILL NEED TO DO THIS ANALYSIS…**

## Linkage Analysis

* Still not understanding how you can calculate LD without phased data but it **seems** like many programs claim to…

### PLINK

* [v1.90 user manual: LD section](https://www.cog-genomics.org/plink2/ld)
* [x] create the files needed from the master VCF file (tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf)
  + looks like plink now reads VCF (v1.90): will try this first.
* [x] split data into smaller pieces to parallelize the plink analysis.
  + looks like the --parallel flag will allow plink to take care of this.
* [x] start run(s) on louise.
* [ ] *try running a --blocks PLINK analysis for haplotype blocks to see if its useful?*

**plink commands run and kept:**

plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf --allow-extra-chr \  
 --r gz with-freqs \  
 --out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf\  
 /ld/r\_none\_freqs  
  
plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf --allow-extra-chr \  
 --r gz in-phase with-freqs \  
 --out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf\  
 /ld/r\_none\_phase\_freqs  
  
plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf --allow-extra-chr \  
 --r triangle gz \  
 --out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf\  
 /ld/r\_tri

### Plot PLINK results

* [x] create ipython notebook file
  + [YALE/ddrad58/2015-01-02\_Plot\_PLINK\_results.ipynb](http://nbviewer.ipython.org/github/xguse/ipy_notebooks/blob/master/YALE/ddrad58/2015-01-02_Plot_PLINK_results.ipynb)
  + [2015-01-13]: [YALE/ddrad58/2015-01-05\_Plot\_PLINK\_results.ipynb](http://nbviewer.ipython.org/github/xguse/ipy_notebooks/blob/master/YALE/ddrad58/2015-01-05_Plot_PLINK_results.ipynb)
* [ ] write code to plot

## TODO for Gisella

* [ ] re-read grant bit about bioinformatics and think about how to use Hongyu Zhao.

# 2015-01-03 (Saturday)

## Linkage Analysis

### Plot PLINK results

* [x] **WHAT** should be plotted?
  + [x] what *exactly is* the value telling us?
    - [x] does it already take into account the distance?
      * according to [Wikipedia](http://en.wikipedia.org/wiki/Linkage_disequilibrium#Definition), is simply the correlation coefficient between pairs of loci:
  + It seems like plotting ( divided by distance) **is** warranted where:
    - is location of SNP
    - is location of SNP
* [ ] write code to plot

#### Questions for Andrea

* Some MAFs are zero which causes the LD () to fail. [link](http://localhost:8888/jupiter/notebooks/YALE/ddrad58/2015-01-02_Plot_PLINK_results.ipynb#LD-as-r-for-Scaffold0:)

# 2015-01-05 (Monday)

## Linkage Analysis

### Plot PLINK results

* met with Andrea after showing her what I had and specifically the MAF = 0 for about half the scaffold0 comparisons.
  + **(see meeting notes for more details)**

### Re-Filter original VCF

* the incorrect (--min-allele/--max-allele) filter was used to generate: tsetseFINAL\_14Oct2014\_f2\_53.recode.vcf
* the correct filter is --maf.
* I am doing it myself with MAF = 0.05 (see below).
* **retained only 47.7% sites**
* **I will be re-running my PLINK LD analysis just in case.**

wd238 at compute-1-4 in ~GENOMES/glossina\_fuscipes/annotations/SNPs (py278)   
$ vcftools \  
 --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf \  
 --maf 0.05 \  
 --out tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05 \  
 --recode  
  
VCFtools - v0.1.12b  
(C) Adam Auton and Anthony Marcketta 2009  
  
Parameters as interpreted:  
 --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.vcf  
 --maf 0.05  
 --out tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05  
 --recode  
  
After filtering, kept 53 out of 53 Individuals  
Outputting VCF file...  
After filtering, kept 73297 out of a possible 153650 Sites  
Run Time = 21.00 seconds

### PLINK - rerun

plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf \  
--allow-extra-chr \  
--r gz with-freqs dprime \  
--out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf/ld/r\_none\_freqs\_dprime

## Recover dead positives

### Dissections

* Prepped for dissections and pre-filled the worksheets
* but we are out of the 1.5 ml tubes that I bought for this and I will have to go get some more tomorrow morning.

# 2015-01-06 (Tuesday)

## Linkage Analysis

* emailed Dan about looking over the results.
* Should probably run them by Jeff if he has time too.

### PLINK --make-bed

wd238 at compute-1-4 in ~GENOMES/glossina\_fuscipes/annotations/SNPs (py278)   
$ plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf \  
> --allow-extra-chr \  
> --maf 0.05 \  
> --make-bed \  
> --out tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink  
  
PLINK v1.90b2o 64-bit (25 Nov 2014) https://www.cog-genomics.org/plink2  
(C) 2005-2014 Shaun Purcell, Christopher Chang GNU General Public License v3  
Logging to tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink.log.  
48251 MB RAM detected; reserving 24125 MB for main workspace.  
--vcf: 73k variants complete.  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink-temporary.bed  
+  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink-temporary.bim  
+  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink-temporary.fam  
written.  
73297 variants loaded from .bim file.  
53 people (0 males, 0 females, 53 ambiguous) loaded from .fam.  
Ambiguous sex IDs written to  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink.nosex .  
Using 1 thread (no multithreaded calculations invoked).  
Before main variant filters, 53 founders and 0 nonfounders present.  
Calculating allele frequencies... done.  
Total genotyping rate is 0.965098.  
0 variants removed due to MAF threshold(s) (--maf/--max-maf).  
73297 variants and 53 people pass filters and QC.  
Note: No phenotypes present.  
--make-bed to  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink.bed +  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink.bim +  
tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink.fam ...  
done.

### PLINK --blocks

* running with --blocks option to look at estimated haplotype blocks

wd238 at compute-1-4 in ~GENOMES/glossina\_fuscipes/annotations/SNPs (py278)   
$ plink --vcf tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf \  
> --allow-extra-chr \  
> --blocks no-pheno-req no-small-max-span \  
> --out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf\  
 /ld/blocks\_nophenoreq\_nosmallmaxspan  
  
PLINK v1.90b2o 64-bit (25 Nov 2014) https://www.cog-genomics.org/plink2  
(C) 2005-2014 Shaun Purcell, Christopher Chang GNU General Public License v3  
Logging to plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf\  
 /ld/blocks\_nophenoreq\_nosmallmaxspan.log.  
48251 MB RAM detected; reserving 24125 MB for main workspace.  
--vcf: 73k variants complete.  
  
...  
  
73297 variants loaded from .bim file.  
53 people (0 males, 0 females, 53 ambiguous) loaded from .fam.  
Ambiguous sex IDs written to  
plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf\   
 /ld/blocks\_nophenoreq\_nosmallmaxspan.nosex  
.  
Using 1 thread (no multithreaded calculations invoked).  
Before main variant filters, 53 founders and 0 nonfounders present.  
Calculating allele frequencies... done.  
Total genotyping rate is 0.965098.  
73297 variants and 53 people pass filters and QC.  
Note: No phenotypes present.  
--blocks: 8040 haploblocks written to  
plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf\  
 /ld/blocks\_nophenoreq\_nosmallmaxspan.blocks  
.  
Extra block details written to  
plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf\  
 /ld/blocks\_nophenoreq\_nosmallmaxspan.blocks.det  
.  
Longest span: 199.985kb.

### Plot PLINK results

* cleaned up a few things
* added residual plots following the regplots

## Recover dead positives

* need to meet with Kirsten (emailed her to schedule a time)
  + subject: “*Short meeting to talk about the dead positives screen?*”

### Dissections

* getting more:
  + tubes
  + PBS
  + Pens
* need to get receipt(s) from Kirsten regarding the dissection dish order
  + emailed with subject: “*did you ever send me the receipt(s) for the stuff you ordered for the dissections over the internet?*”

## Bonizzoni *et al*: Insecticide Resistance

* I am SUPER late on this!

# 2015-01-07 (Wednesday)

## Bonizzoni *et al*: Insecticide Resistance

**Status:** COMPLETE

* finished reviewing the main text
* emailed it to her
* will not be going over the legends or figs

## Meeting with Serap Aksoy

**Time:** 10:00 AM to 11:30AM

### Discussed

* how to log Iowa tsetse samples
* student to do much f the logging after a spreadsheet is devised
* location of the other RNA midguts
  + she said she thought they didnt get any but then thought she remembered that Brian tried extracting RNA from at least a few infected midguts with no success.
  + \*\*She said she needs the carcasses of the infected flies too which I did not remember (**I need to bring this up with Gisella bc this is a major reduction of our expected infected material…**).
    - *this doesn’t really make sense anyway since i dont think we preserved the bodies for RNA.*
* having me send an ad or two for a postdoc position for her lab to my friends

### Action items

**Status:** IN PROGRESS

* [x] create simple excel sheet to track Iowa samples
  + [x] email sheet to Brian and Serap
    - subject: *Spreadsheet to record Iowa sample materials*
  + **NOTE:** sheet is a google sheet named [Iowa\_tsetse\_material\_inventory](https://docs.google.com/spreadsheets/d/1SeoKnRQ0djB1xjyVGQy-NikNFTzQ-wL-hXCwUZAoNqw/edit?usp=sharing)
* [x] locate extra RNA midguts in our freezers
  + [x] email Aksoy, Brian, Michelle to schedule pickup
    - subject: *Many more midguts and heads for RNA*
* [ ] send feelers and ads to friends about postdoc position in her lab
* [ ] contact Gisella about Serap wanting the carcasses…

## Meeting with Andrea

### Discussed

* problem re-running the figure generation R script
* couldn’t open the PNG writer bc no X11 on the cluster and ssh -Y wasn’t working even though it did last time…
* I added some code to the R script to specify `png(type=“cairo”)
* waiting to hear the outcome
  + *program ran but output was not what was expected: many more graphs than last time*
  + I expect user error

## Recover dead positives

* Meeting with Kirstin tomorrow at 1 or 2 PM

# 2015-01-08 (Thursday)

## Admin stuff

### Serap’s postdoc advertisement

* posted to facebook

### Lab meetings

* email from Jeff:
  + subject: *lab meetings*
  + body:
* Gus,
* Any lab meetings set up? We should grab Aris and Yiota who might give a joint one. Then also the new Anthropology guy, Eduardo Fernandez-Duque.
* Jeff
* doodle poll:
  + [link](http://doodle.com/hc4r8gdi6wnse425) sent to pc\_labs
  + not certain if Maggie is on that list yet
  + emailed Carol for her email in case not
    - She **WAS** on the list when I sent the link
    - Carol replied with current list: recorded below

### Current pc\_labs list

#separator-CACCONE#  
adalgisa.caccone@yale.edu  
carol.mariani@yale.edu  
danielle.edwards@yale.edu  
nphavill@fs.fed.us  
jrichardson@providence.edu  
giovanna.carpi@yale.edu  
katharine.walter@yale.edu  
gus.dunn@yale.edu  
  
#separator-POWELL#  
jeffrey.powell@yale.edu  
kirstin.dion@yale.edu  
andrea.gloria-soria@yale.edu  
b.evans@yale.edu  
joshua.richardson@yale.edu  
  
#separator-TEMP-ROTATION-UNDERGRAD#  
christian.hernandez@yale.edu  
elaine.guevara@yale.edu  
andres.valdivieso@yale.edu  
mkcorley@gmail.com  
alexis.halyard@yale.edu  
pkotsakiozi@hotmail.com  
aristeidis.parmakelis@yale.edu

## Recover dead positives

### Meeting with Kirstin

* see: [Kirsten-2015-01-08.md](file:///home/gus/Dropbox/repos/git/markdown-docs/notes/meetings/Kirsten-2015-01-08/Kirsten-2015-01-08.md)
* sent above for confirmation or amendment to Kirstin
  + She approves

# 2015-01-09 (Friday)

## Sarah Licensing Exam

* I was taking care of the kids all morning

## New project brainstorming

## PLINK: Fst

* defining population ID file for:
  + tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf
  + tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf.popdef

plink --bfile tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.plink \  
--allow-extra-chr \  
--within tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf.popdef \  
--fst \  
--out plink\_out/tsetseFINAL\_14Oct2014\_f2\_53.recode.renamed\_scaffolds.maf0\_05.vcf/fst/out

* This keeps giving me errors:

Warning: No samples named in --within file remain in the current analysis.  
Using 1 thread (no multithreaded calculations invoked).  
Before main variant filters, 53 founders and 0 nonfounders present.  
Calculating allele frequencies... done.  
Total genotyping rate is 0.965098.  
73297 variants and 53 people pass filters and QC.  
Note: No phenotypes present.  
Error: --fst requires at least two nonempty clusters.

# 2015-01-10 (Saturday)

## PLINK: Fst

* Still getting errors
* looks like its bc there is no sex info attached to the samples
* looking for other was to do this stuff: found EggLib-py

## Install EggLib

* installation needs bio++
* see tomorrow

## Install Bio++ (bpp)

* install script ([bpp-setup.sh](http://biopp.univ-montp2.fr/Download/bpp-setup.sh)) obtained from [bio++ website](http://biopp.univ-montp2.fr/wiki/index.php/Installation).
  + altered install script to fit system (louise) and renamed [install\_bpp\_2.2.0.sh](/home/gus/remote_mounts/louise/scripts/installs/install_bpp_2.2.0.sh).
* see tomorrow

# 2015-01-11 (Sunday)

## Install EggLib

* finished bio++install
* installing other things (actually may just module-ize versions already installed on louise:
  + [x] gsl (already installed)
  + [x] clustalw (linked to louise/~MAIN\_APPS/clustalw/clustalw-2.0.12-linux-i686-libcppstatic)
  + [x] muscle
  + [x] paml
  + [x] phyml
  + [x] primer3
  + [x] phylip

## Install Bio++ (bpp)

* had to ammend the install script to include the bpp install location in $PATH so it can use itself to install/build parts of itself
* as far as I can tell the only things that fail to install now are the GUI-based stuff that needs Qt. As I am using this on th cluster I dont need/want these so I am going to proceed as if this succeeded.

# 2015-01-12 (Monday)

## Install EggLib

* finished installing external helper programs:
  + clustalw
  + phylip
* made a modules file for the whole group: egglib\_helpers

## Install Bio++ (bpp)

* forgot to write a modules file for this.
* doing it now

# 2015-01-13 (Tuesday)

## Recover dead positives

* meeting
* see meeting notes [gisella\_kirsten\_2015-01-14.md](file:///home/gus/Dropbox/repos/git/markdown-docs/notes/meetings/gisella_kirsten_2015-01-14/gisella_kirsten_2015-01-14.md)

## Maps stuff

* Need to update our Northern Uganda map with the latest Village location data
* Spartan.utils.maps.gps
  + coding the functions to take all trap GPS coords for a village and return one GPS coord set for each that represents the central tendency for simplified plotting

## TsetseSampleDB

* adding names/village codes to the list of village-to-code map ([village\_id\_map.csv](file:///home/gus/Dropbox/repos/git/TsetseCheckout/TsetseCheckout/data/village_id_map.csv))

## Helping Aris

* getting mpirun mrbayes to run on grace
* took ~ 1 hour.

# 2015-01-15 (Thursday)

## Science Fair

* 8 to 12:30

## Meeting about Kenya Tsetse

Members:

* Gisella
* Serap
* Michelle
* Brian
* Gus

Notes (bad bc they were taken with my phone):

* Gpd samples
* DNA samples
* What column data will be needed for this Gpd Kenya collection
* Re-circulate collection template excel and protocol etc

# 2015-01-16 (Friday)

## Collection Spreadsheet review

* point is to make sure we can use this for the Kenya “simpler”
* after speaking with Gisella, I am going to add a few of the “simpler” column heading to the normal collection spreadsheet and write a bunch of notes explaining that not everything needs to be filled in for everything.
* collection spreadsheet: [Example\_collection\_template\_kenya.xls](file:///home/gus/Dropbox/uganda%20data/collection_sheet_templates/Example_collection_template_kenya.xls)
  + **status:** finished
* summary spreadsheet: [Example\_summary\_template\_kenya.xls](file:///home/gus/Dropbox/uganda%20data/collection_sheet_templates/Example_summary_template_kenya.xls)
  + **status:** in progress

### Email explanation

**Subject:**

**Body:**

## Updating maps: current trap locations

* created new spreadsheet:[collection\_meta\_data/meta\_data.ods](file:///home/gus/Dropbox/uganda%20data/collection_meta_data/meta_data.ods) to store current state of stuff like the village-to-symbol map, etc.
* [ ] collecting trap GPS data to file: [TsetseCheckout/data/village\_id\_map.csv](file://CREATE_ME)
* [ ] collecting all village-to-symbol maps that I have to [meta\_data.ods](file:///home/gus/Dropbox/repos/git/TsetseCheckout/TsetseCheckout/data/village_id_map.csv)

### spartan dev: GPS stuff

* pycharm and ipython

## Phone for Dan

I emailed Dan the following:

**Subject:** Phone call for you

**Body:** Just fielded a call for you from Karan(Karen?) Peart from the Yale Public affairs and communications office.

She would like you to call her back at your earliest convenience (432-1326).

Gus

# 2015-01-18 (Sunday)

## Sarah is sick

* short day: 10am to 1:45pm

## Updating maps: current trap locations

* ipython: [2015-01-16\_convert\_fall2014\_trap\_gps\_village\_names.ipynb](file:///home/gus/Dropbox/common/ipy_notebooks/YALE/maps_stuff/2015-01-16_convert_fall2014_trap_gps_village_names.ipynb)

### spartan dev: GPS stuff

* working on teaching GPSCoordTree how to get mean coordinates

# 2015-01-19 (Monday)

## Andrea: quick chat

* wants to re-run the ddRAD pipeline since I(we) found some issues with at least one of the command line runs’ options.
* I agree
* [ ] #todo: I am installing [Stacks 1.24](http://creskolab.uoregon.edu/stacks/) for her on louise and trying to set up the web-based analysis part
* should not change TOO much about the results and will end up being MUCH more replicatable due to the use of publicly accessible data from vectorbase.
* we can continue to work with the current data until the new set is done and just adjust the work to accommodate the new stuff at the end.

## Updating maps: current trap locations

### spartan dev: GPS stuff

* working on teaching GPSCoordTree how to get mean coordinates

## Install Stacks

* [installation guide](http://creskolab.uoregon.edu/stacks/manual/#install)

## Collection Spreadsheet review

* adding explanation text to the [fly\_collection\_basic.md](file:///home/gus/Dropbox/repos/git/markdown-docs/protocols/fly_collection_basic/fly_collection_basic.md) document.
* **STATUS:**
  + having issues getting validation and drop-down lists to carry over into “empty” rows
  + plan to fix it by copying a template row into like 1000 rows
  + still need to execute the above tomorrow bc computer is acting a fool and I have to go home to sick Sarah and Liam.

# 2015-01-20 (Tuesday)

## Sarah still sick

* stayed home to help with particularly crazy morning

## Meeting with Gisella and Andrea: ddRAD paper

* Met at 10:30 AM
* summarized in [gisella\_andrea\_2015-01-20.md](file:///home/gus/Dropbox/repos/git/markdown-docs/notes/meetings/gisella_andrea_2015-01-20/gisella_andrea_2015-01-20.md).

## ddRAD todos

* [in progress] read Mark’s tryp paper for the LD stuff
* [¤] email Mark to have a short sit-down to go over my results and ask about his work

## Install Stacks

* [installation guide](http://creskolab.uoregon.edu/stacks/manual/#install)

### Prerequisites

**Visualization:**

* [ ] DB2 Pear Module: <http://pear.php.net/package/MDB2/>
* [ ] MDB2 MySQL driver: <http://pear.php.net/package/MDB2_Driver_mysql/>
* [ ] PHP
* [ ] MySQL
* [ ] Perl DBI module installed with the MySQL driver [CPAN/dist/DBD-mysql/](http://search.cpan.org/dist/DBD-mysql/)

**Spreadsheet export:**

* [ ] Perl module: [Spreadsheet::WriteExcel](http://search.cpan.org/~jmcnamara/Spreadsheet-WriteExcel-2.37/)

**Performance improvement:**

* [X] samtools for reading BAM files (already installed)
* [ ] Google’s SparseHash class to lower memory usage <http://code.google.com/p/sparsehash/>

### Stacks

**INSTALLATION PROBLEMS NOTES:**

* Having issues getting the samtools includes and libs configured for the ./configure command.
* plan to build on jupiter using the ARCH ABS and copy the include/lib directories to louise

**download location:**

* status: *in progress*
* downloaded [stacks-1.24.tar.gz](http://creskolab.uoregon.edu/stacks/source/stacks-1.24.tar.gz) to gus@louise/src.

**install script:**

* status: *in progress*
* [gus@louise/scripts/installs/install\_XXXX](file:///home/gus/remote_mounts/louise/scripts/installs/install_XXXXXXX)

**module file:**

* status: *in progress*
* [gus@louise/.local/environment-modules/Modules/3.2.10/my\_modulefiles/XXXXX/XXXXX](/home/gus/remote_mounts/louise/.local/environment-modules/Modules/3.2.10/my_modulefiles/XXXXX/XXXXX)

**software root:**

* status: *in progress*
* [gus@louise/home/gus/remote\_mounts/louise/.local/easybuild/software/XXXXX/XXXXX](/home/gus/remote_mounts/louise/.local/easybuild/software/XXXXX/XXXXX)

## Install SparseHash

**download location:**

* status: *complete*
* downloaded [sparsehash-2.0.2.tar.gz](https://sparsehash.googlecode.com/files/sparsehash-2.0.2.tar.gz) to gus@louise/src.

**install script:**

* status: *written and run*
* [gus@louise/scripts/installs/install\_sparsehash\_2.0.2.sh](file:///home/gus/remote_mounts/louise/scripts/installs/install_sparsehash_2.0.2.sh)

**module file:**

* status: *complete but not tested*
* [gus@louise/.local/environment-modules/Modules/3.2.10/my\_modulefiles/sparsehash/2.0.2](/home/gus/remote_mounts/louise/.local/environment-modules/Modules/3.2.10/my_modulefiles/sparsehash/2.0.2)

**software root:**

* status: *installed*
* [gus@louise/home/gus/remote\_mounts/louise/.local/easybuild/software/sparsehash/2.0.2](/home/gus/remote_mounts/louise/.local/easybuild/software/sparsehash/2.0.2)

# 2015-01-21 (Wednesday)

## Family still VERY sick

* stayed home with Liam while Sarah took Clementine and herself to the doctor
* got to work at 12:30
* had to go home so Sarah could sleep bc Clem was not letting her
* went home at 1:30
* came back at 3:30
* home again at 6:00
* work again at 8:30

## Manual install of Samtools/htslib

### htslib (built with Arch ABS on jupiter)

ABANDONING THIS METHOD FOR NOW. TOO MANY PROBLEMS WITH INTEGRATING CERTAIN INCLUDE AND LIB DIRECTRIES WHEN COMPILING ACCROSS DEPENDENCIES. TRYING EASYBUILD AND ITS *TOOLCHAIN* PARADIGM FOR NOW.

**ABS build:**

cd /home/gus/remote\_mounts/louise/src/ABS/  
tar -xf htslib.tar.gz  
cd htslib  
makepkg

**Install script:**

* [gus@louise/scripts/installs/install\_htslib\_1.1.sh](file:///home/gus/remote_mounts/louise/scripts/installs/install_htslib_1.1.sh)

## EasyBuild installs

### EasyBuild 1.16.1

**Install script:**

* [gus@louise/scripts/installs/install\_easybuild\_1.16.1.sh](file:///home/gus/remote_mounts/louise/scripts/installs/install_easybuild_1.16.1.sh)

### Install samtools-1.1

wd238 at compute-21-15 in ~ (py278)   
$ md load EasyBuild/1.16.1  
  
wd238 at compute-21-15 in ~ (py278)   
$ eb SAMtools-1.1-goolf-1.4.10.eb --try-toolchain=goolf,1.4.10-no-OFED --robot  
  
...

* this will be installing the whole toolchain and all samtools dependencies so it was executed in a screen.
* I am going home to sleep while this works [2015-01-21 21:33].
* emailing Andrea first to let her know its not gonna be ready when I told her.

**STATUS: [2015-01-22 07:52]**

* Build seems to have **SUCCEEDED**

# 2015-01-22 (Thursday)

## Collection documentation files

* fixed/kludged the collection template to keep cell-notes and verification by including “dummy” entries up to 500.
  + collection template: [Example\_collection\_template\_kenya.xlsx](file:///home/gus/Dropbox/uganda%20data/collection_sheet_templates/Example_collection_template_kenya.xlsx)
  + summary template: [Example\_summary\_template\_kenya.xlsx](file:///home/gus/Dropbox/uganda%20data/collection_sheet_templates/Example_summary_template_kenya.xlsx)
* amended and committed v0.2.1 of [fly\_collection\_basic.md](file:///home/gus/Dropbox/repos/git/markdown-docs/protocols/fly_collection_basic/fly_collection_basic.md) to the repo with custom message:
* “protocols/fly\_collection\_basic/fly\_collection\_basic.md commited at version: v0.2.1”

## EasyBuild installs

### Install GSL-1.16

wd238 at compute-21-15 in ~ (py278)   
$ eb GSL-1.16-goolf-1.4.10.eb --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:**

== COMPLETED: Installation ended successfully  
== Results of the build can be found in the log file \  
 /home2/wd238/.local/easybuild/software/GSL/1.16-goolf-1.4.10-no-OFED/\  
 easybuild/easybuild-GSL-1.16-20150122.075231.log  
== Build succeeded for 1 out of 1

### Install sparsehash-2.0.2

wd238 at compute-21-15 in ~ (py278)  
eb google-sparsehash-2.0.2-goolf-1.4.10.eb --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:**

== COMPLETED: Installation ended successfully  
== Results of the build can be found in the log file \  
 /home2/wd238/.local/easybuild/software/google-sparsehash/ \  
 2.0.2-goolf-1.4.10-no-OFED/easybuild/ \  
 easybuild-google-sparsehash-2.0.2-20150122.080547.log  
== Build succeeded for 1 out of 1

### Install Stacks-1.03

**NOTE:** this is not for use per se but to generate the config and module files so that I can modify them for the latest version of Stacks and install *THAT* version.

wd238 at compute-21-15 in ~ (py278)   
$ eb Stacks-1.03-goolf-1.4.10.eb --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:**

== COMPLETED: Installation ended successfully  
== Results of the build can be found in the log file \  
 /home2/wd238/.local/easybuild/software/Stacks/ \  
 1.03-goolf-1.4.10-no-OFED/easybuild/ \  
 easybuild-Stacks-1.03-20150122.081337.log  
== Build succeeded for 1 out of 1

### Install Stacks-1.24

**easyconfig file:**

* altered the one generated when building Stacks-1.03
  + [gus@louise/.local/easybuild/ebfiles\_repo/Stacks/Stacks-1.03-goolf-1.4.10-no-OFED.eb](file:///home/gus/remote_mounts/louise/.local/easybuild/ebfiles_repo/Stacks/Stacks-1.03-goolf-1.4.10-no-OFED.eb)
* [gus@louise/scripts/installs/easybuild/Stacks-1.24-goolf-1.4.10-no-OFED.eb](file:///home/gus/remote_mounts/louise/scripts/installs/easybuild/Stacks-1.24-goolf-1.4.10-no-OFED.eb)

#### Attempt 01

wd238 at compute-21-15 in ~ (py278)  
eb Stacks-1.24-goolf-1.4.10-no-OFED.eb --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:** FAILED

* couldn’t find Stacks-1.24-goolf-1.4.10-no-OFED.eb
* basically expected.

#### Attempt 02

wd238 at compute-21-15 in ~ (py278)  
eb $HOME/.local/easybuild/ebfiles\_repo/Stacks/Stacks-1.24-goolf-1.4.10-no-OFED.eb \  
 --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:** FAILED

* error log: [gus@louise/scripts/installs/easybuild/failure\_logs/easybuild-Stacks-1.24-20150122.090021.pPmjV.log](file:///home/gus/remote_mounts/louise/scripts/installs/easybuild/failure_logs/easybuild-Stacks-1.24-20150122.090021.pPmjV.log)
* looks like it cant find sparsehash for the linking
* will add samtools and sparsehash to the easyconfig file as dependencies and or build dependencies.

#### Attempt 03

* added the below to the easyconfig file:
  + builddependencies = [('SAMtools', '1.1'), ('google-sparsehash', '2.0.2')]

wd238 at compute-21-15 in ~ (py278)  
eb $HOME/.local/easybuild/ebfiles\_repo/Stacks/Stacks-1.24-goolf-1.4.10-no-OFED.eb \  
 --try-toolchain=goolf,1.4.10-no-OFED --robot  
  
== temporary log file in case of crash /tmp/easybuild-0T7khn/easybuild-O8eCl9.log  
ERROR: EasyBuild crashed with an error \  
 (at easybuild/software/EasyBuild/1.16.1/lib/python2.7/site-packages/\  
 easybuild\_framework-1.16.1-py2.7.egg/easybuild/tools/robot.py:232 in \  
 resolve\_dependencies): Irresolvable dependencies encountered: \  
 SAMtools/1.1-goolf-1.4.10-no-OFED, google-sparsehash/2.0.2-goolf-1.4.10-no-OFED

**STATUS:** FAILED

* error log: [gus@louise/scripts/installs/easybuild/failure\_logs/easybuild-O8eCl9.log](file:///home/gus/remote_mounts/louise/scripts/installs/easybuild/failure_logs/easybuild-O8eCl9.log)

#### IRC session with author/devs

* one problem was that I dont need to keep using --try-toolchain=goolf,1.4.10-no-OFED since the local easyconfig (Stacks-1.24-goolf-1.4.10-no-OFED.eb) being passed to eb already defines the toolchain.
* the build still fails however

### Install zlib-1.2.8

**NOTE:**

* this is because things seem to need it when building stacks-1.24?
* doesn’t seem like this was the case?

wd238 at compute-21-15 in ~ (py278)   
$ eb zlib-1.2.8-goolf-1.4.10.eb --try-toolchain=goolf,1.4.10-no-OFED --robot

**STATUS:** SUCCESSFUL

== COMPLETED: Installation ended successfully  
== Results of the build can be found in the log file \  
 /home2/wd238/.local/easybuild/software/zlib/\  
 1.2.8-goolf-1.4.10-no-OFED/easybuild/\  
 easybuild-zlib-1.2.8-20150122.115448.log  
== Build succeeded for 1 out of 1

# 2015-01-23 (Friday)

## Meeting with Alexis

* 10:00 to 12:20
* talked about overall project and helped her come up with stuff to talk about for 2 minutes next Tuesday.
* Gisella was supposed to be here but double booked the time so will meet with Alexis individually.
* Gave Alexis my email and asked her to email me so that i would get hers
* so far [17:00] has not emailed me.

## EasyBuild installs

* forked and cloned [easybuild-easyconfigs](https://github.com/xguse/easybuild-easyconfigs) git repo

## Doc appointment

* left desk around 14:00 and got back around 15:20

# 2015-01-24 (Saturday)

## Sarah sprained/broke? her ankle

* Sarah slipped while trying to shovel snow (?! WHY ?!)
* had to go back home soon after arrival

# 2015-01-25 (Sunday)

## EasyBuild installs

* adding new forked git-repo of easyconfigs to easybuild through environment variables in my [gus@louise/.zshrc](file:///home/gus/remote_mounts/louise/.zshrc).
* make check failing was caused by the easyconfig file setting runtest='check'.
  + I removed this
* install still fails but seems to try to repeat itself and fails the SECOND TIME?
  + ...  
    == building and installing Stacks/1.24-goolf-1.4.10-no-OFED...  
    == fetching files...  
    == creating build dir, resetting environment...  
    == unpacking...  
    == patching...  
    == preparing...  
    == configuring...  
    == building...  
    == testing...  
    == installing...  
    == creating build dir, resetting environment...  
    == unpacking...  
    == patching...  
    == preparing...  
    == configuring...  
    == building...  
    == FAILED: Installation ended unsuccessfully...  
    ...
* checking the logs seems to suggest that the BAM include files are still not working
  + testing by removing the reqs from the configure script
    - NO sparsehash and NO bam: **SUCCEEDS**
    - YES sparsehash and NO bam: **SUCCEEDS**
    - NO sparsehash and YES bam: **FAILS**
* **TO TRY TOMORROW:**
  + ‘clone’ environment from one of the “test\_reports” in [failure\_logs](file:///home/gus/remote_mounts/louise/scripts/installs/easybuild/failure_logs/) and try to run ./configure; make; make install; manually.

# 2015-01-26 (Monday)

## Carl Zimmer Writing Workshop

* 10:00 to 12:00
* Notes made in notebook to be transfered here when I have time (blizzard approaching)

# 2015-01-27 (Tuesday)

## SNOW-pocalypse

* was told to stay home by Yale
* came in by mistake but left soon after realizing

# 2015-01-28 (Wednesday)

## EasyBuild installs

### Stacks

#### Stacks no BAM

* testing to make sure it works
* abandoning this bc we decided that we dont need to run this step over right now

#### Stacks yes BAM

* still not building correctly

## tmux

* starting point [easyconfig](https://github.com/fgeorgatos/easybuild.experimental/blob/539bd104d158c9f41b45d60115f6bf1b7155e11e/contrib/pkgsrc/20141219/t/tmux-1.9a-goolf-1.4.10.eb)
* abandoning this for now
* simply not crucial

## ddRAD stuff

* [X] email Mark about Tryp LD analysis in his paper
  + [X] follow up with him on this “Re: Short meeting to chat about genomic scale LD analysis?”
* [X] read Tryp paper for same
* [X] install latest Stacks version on louise and make runable by Andrea
* [X] return hapFLK script to original code and copy altered version to new name
  + [X] let Andrea know (*acknowledged*)
* [ip] Generate descriptive statistics and figures of the LD results as a whole rather than by contig where possible

# 2015-01-29 (Thursday)

## ddRAD stuff

* [ip] Generate descriptive statistics and figures of the LD results as a whole rather than by contig where possible
* [2015-01-28\_Plot\_PLINK\_results\_cumulative.ipynb](file:///home/gus/Dropbox/common/ipy_notebooks/YALE/ddrad58/2015-01-28_Plot_PLINK_results_cumulative.ipynb)

## Robert’s stuff

* [-ip-] Pick out 26 flies (13 M, 13 F) from each area we want to look at for Robert’s work in March
  + sent first set to Kirstin and Alexis
    - 13 M/13 F from Oyam/Kole trip in 2014-07 (see Table )
* [ ] make sure we have an updated map with all the villages from *Spring and Summer 2014*
* [X] Meet with Gisella to pick out which locations Robert’s data will come from while looking at the updated map.
  + [ ] she told me to pick the areas and give her a table/report on which and why. The issues to consider are:
    - wide representation of population areas in the North
    - allows temporal comparisons as well

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Collection Date | Species | Sex | Teneral | Village | Fly |
| 2014-07-15 | *G. f. fuscipes* | F | NT | OD | 017 |
| 2014-07-15 | *G. f. fuscipes* | F | NT | OCA | 031 |
| 2014-07-15 | *G. f. fuscipes* | F | NT | OCA | 039 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | AKA | 045 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | AKA | 052 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | AKA | 056 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | AKA | 062 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | AKA | 068 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | OCA | 092 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | OCA | 100 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | ACA | 120 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | OD | 137 |
| 2014-07-16 | *G. f. fuscipes* | F | NT | OD | 149 |
| 2014-07-15 | *G. f. fuscipes* | M | NT | OD | 020 |
| 2014-07-15 | *G. f. fuscipes* | M | NT | OCA | 022 |
| 2014-07-15 | *G. f. fuscipes* | M | NT | OCA | 025 |
| 2014-07-15 | *G. f. fuscipes* | M | NT | OCA | 026 |
| 2014-07-15 | *G. f. fuscipes* | M | NT | OCA | 035 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | AKA | 049 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | AKA | 063 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | OCA | 095 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | ACA | 116 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | ACA | 125 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | ACA | 129 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | OCA | 146 |
| 2014-07-16 | *G. f. fuscipes* | M | NT | OD | 155 |

Samples given to Alexis for DNA extraction from a single leg.

## Rob H Jobs

### Iowa samples

* [X] catalog the boxes that we got here at **ESC**.
* [X] email Gisella about getting Rob over to **EPH** to start on their samples.

### *G. f. fuscipes* samples 2014

* [X] organize boxes of *carcass* and *midgut* samples from 2014-03 to 2014-08 in Gisella freezer by month.
* [-ip-] standardize the collection spreadsheets to prepare for automated import to TsetseSampleDB.

## Updating maps: current trap locations

### spartan dev: GPS stuff

* [-ip-] teaching GPSCoordTree how to get mean coordinates
  + [X] make GPSCoord hashable
  + [-ip-] fix GPSCoordTree.\_add\_levels(): getting an empty list somewhere o something that is causing a None to be returned.
* still not fixed but trying a new tactic
  + converting GPSCoordTree to use autovivification trick with an extra key at each node that holds links to all gps\_objs found below it.
  + see [2014-12-26\_functional\_annotation\_table\_generator.ipynb](http://nbviewer.ipython.org/github/xguse/ipy_notebooks/blob/master/YALE/ddrad58/2014-12-26_functional_annotation_table_generator.ipynb) for example of autovivfication method.