Genetic Analysis of Stress Hormone Levels Affecting Egg Production Capabilities in Chickens

A. Adler
NCSSM Online
North Carolina School of Science and Mathematics
Durham, North Carolina

January 16, 2017

Abstract: Stress hormones in chickens can affect behavioral and physical traits, including egg production in hens. This paper makes use of genomic and statistical tools to determine quantitative trait loci (QTL) in chicken DNA that code for the production of three stress hormones. This was done by creating two models: rst obtaining QTL data, performing multiple genetic mainscans, con dence interval graphs, and effect plots, and later correlating the results with an experiment model of egg production percent for two test groups. It was found that the production of Aldosterone is coded signi cantly on chromosome 5, DHEA on chromosomes 4 and 21, and Corticosterone on chromosomes 3 and 7. It has also been concluded that Corticosterone levels correlate with speci c genotypes at particular loci, and that Corticosterone plays a signi cant role in a chicken's ability to lay eggs. Using the ndings determined by this paper, it is possible to provide geneticists and poultry-owners with valuable information concerning the production of stress hormones and how it can affect egg-laying capabilities.

Key words: Chickens,Gallus Gallus, QTL Analysis, Corticosterone, Aldos-

using genetic and statistical models. These models show the locations where phenotypic traits are coded and can be used for genetic engineering and analysis.

Red Junglefowl Chickens, Gallus Gallus, are considered the ancestors of today's domesticated chickens. Since their domestication around 5000 years ago, these chickens have dispersed and live all over the world[1]. Subspecies of the Red Junglefowl and other chicken breeds are used by humans worldwide for their valuable source of meat and production of edible eggs. Many factors can affect a chicken's ability to lay eggs, including habitat, nutrition, and stress factors. Examples of the stressors chickens face include extreme temperature, lack of food, and predation[2]. In response to these stress factors, chickens produce stress hormones, which are secreted from the adrenal cortex[5]. It is possible to measure these levels of stress hormones, making it feasible to perform genomic studies on these stress hormones in chickens. Using a genetic QTL analysis of chicken DNA data, it is possible to see the specific locations on the chickens' chromosomes that code for production of various hormones.

The term ”QTL” is used throughout this paper to refer to quantitative trait loci, genetic locations on an organism's DNA that have been been statistically proven to have a large influence over a physical trait, or phenotype[7]. In doing this type of research, it is necessary to understand that organisms have genetic information coded on chromosomes, and genes make up proteins, which code for various phenotypes. In order to study these QTL, data must first be created and mapped using computational models. To create the data, a sample size of one type of organism (typically a few hundred) is selected for study. To produce the best results, this sample is often comprised of phenotypically diverse individuals of the same species[7]. Next, phenotypes such as size, and in the case of this paper, hormone levels, are quantitatively measured for each individual. Blood is then extracted from each individual and DNA is separated from other materials, often by the use of a centrifuge. The extracted genetic material is finally run through various computers and DNA sequencers to create the QTL data. Genetic markers are used in this field of study to measure chromosome length (in centimorgans) and to estimate where QTL might be located[7].

The terms ”loci” and ”LOD scores” are also technical terms used frequently in this paper. Loci are the plural form for locus, and are used to describe significant genetic locations that have influence over phenotypes[7]. LOD stands for ”Logarithm of the Odds” and its score can be used to measure genetic linkage of a loci to a phenotype. In genomic research, a LOD score of 3 or greater is generally considered significant in the coding for traits[7].

Figure 1: This diagram shows a pair of chromosomes, and suggests how QTL can be found using genetic markers[3].

Figure 1 shows a diagram of two chromosomes and demonstrates how QTL are located. Pairs of chromosomes are connected by centromeres, the region above the centromere is called the P-arm, and the region below the centromere is called the

Q-arm. QTL are located on either arm and are surrounded by one genetic marker on each side, represented by M1 and M2. These markers facilitate the process of nding signi cant loci and are used in creating the QTL data sets.

Female chickens have the ability to produce edible eggs almost daily, all without male fertilization. Many factors contribute to the ef - ciency of egg production, including age, breed,

cross. Along with basic information such as sex, an ID number, and parental grandmothers, this data set contains the phenotypes for brain mass, metatarsus length, and amounts of various stress induced hormones, as well as logarithmically transformed values for some of the phenotypes. Although some of the phenotypes in this data set can't be genetically analyzed, there is a total of 79 phenotypes in this data set.

The first steps in this model included clearing previous data and setting the R directory to a designated desktop folder. After first downloading the R/qtl library, it was loaded into the model[8]. Next, the data was loaded into the R script and information was entered to tell the program that AA represents a homozygous dominant genotype, BB represents a homozygous recessive genotype, AB represents a heterozygous genotype, and a dash (-) represents missing data. The function "jittermap" was then used to move the genetic markers apart slightly so that the results were more reliable. The "summary" and "names" commands printed useful information about the type of cross, number and phenotype names, number of genetic markers, and percentages by genotype. As described in the previous paragraph, these functions describe the data set as having information for 232 individuals, 79 phenotypes, and 739 genetic markers for 29 chromosomes.

The next three commands in the code output graphs that suggest the reliability of the data. Est.rf displays a graph of recombination fractions with LOD scores. Plot.map displays a genetic map that contains all 29 chromosomes, their lengths in centimorgans, and the locations of all 739 markers. Plot.missing shows available data in white and missing data in black.

The next portion of this model renames variables within this data set for later analysis and uses the hist() command to produce histograms for the data of three hormone levels (Aldosterone Log, DHEA Log, and Corticosterone Response).

The final diagnostic used to visualize the reliability of the data was a collection of qq plots that compared theoretical quantities with sample quantities for each of the three chosen phenotypes. In addition to the three list plots, a linear regression line was added to each to more easily see the strength of each correlation.

After performing the initial diagnostic tests, the mainscans could be produced. Before creating mainscans for each of the three hormone phenotypes, a genetic probability map had to be created and genome probability calculations had to be carried out in order to calculate what an ideal scan should look like. For these two tests, the Haldane method and a fixed step width was used[12]. After producing this ideal model, mainscans were created for the traits of Aldosterone, DHEA, and Corticosterone. For each scan, the correct phenotype column was entered, the EM algorithm was used, and the scans were run for 100 permutations[13]. After running each scan, the three mainscans were plotted with confidence thresholds at 67%, 90%, and 95%, as designated by the blue, red, and green lines. After the model finished running and graphing each scan, it also output a text description, showing significant chromosomes, the exact locations, and LOD scores.

The final portion of the R/qtl model analyzed the Corticosterone response in greater detail. First, two confidence interval plots were created for chromosomes 3 and 7 to show more precise genetic locations that code for Corticosterone production. As in the mainscans, a confidence

threshold of 95% was used to obtain accurate results. The green lines near the bottom of the graphs represent length intervals that likely contain signi cant loci that code for Corticosterone response. Two effect plots were created for Corticosterone response, one for each of the loci on chromosomes 3 and 7. These two plots show which genotypes at a speci c loci tend to correlate with a higher, lower, or medium Corticosterone production amount.

After completing and running the R/qtl model, one additional model was constructed in Mathematica to demonstrate how Corticosterone levels in chickens affect egg-laying abilities[9]. This model rst imports the downloaded data set. The data used in this model was found in the paper by Shini[11]. After importing the data, the model deletes the two header rows and transposes the data so that each column is de ned by a variable (age, control egg production percent, and Corticosterone egg production percent). Lastly, this model graphs the two experimental groups as separate lines on the same graph, with age in weeks as the independent variable and egg production percent as the dependent variable.

Results and Discussion

The graphs produced in the rst part of the R/qtl model give basic information about the available genetic data. Shown in gures 2, 3, and 4, these diagnostics include a recombination fraction graph, chromosome map, and missing data graph.

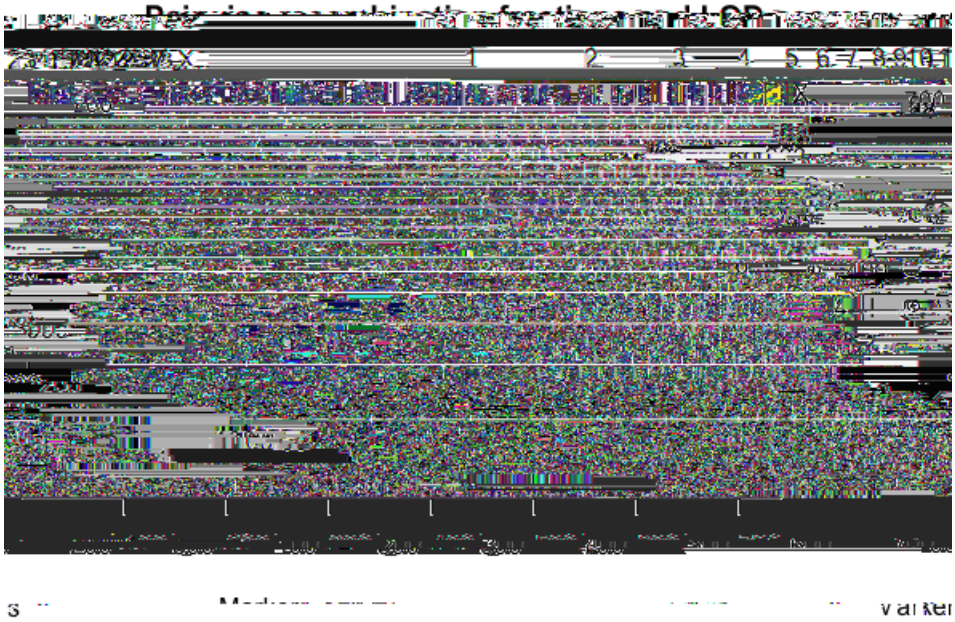

Figure 2: This graph displays the pairwise recombination fractions with LOD scores for each marker on all 29 chromosomes.

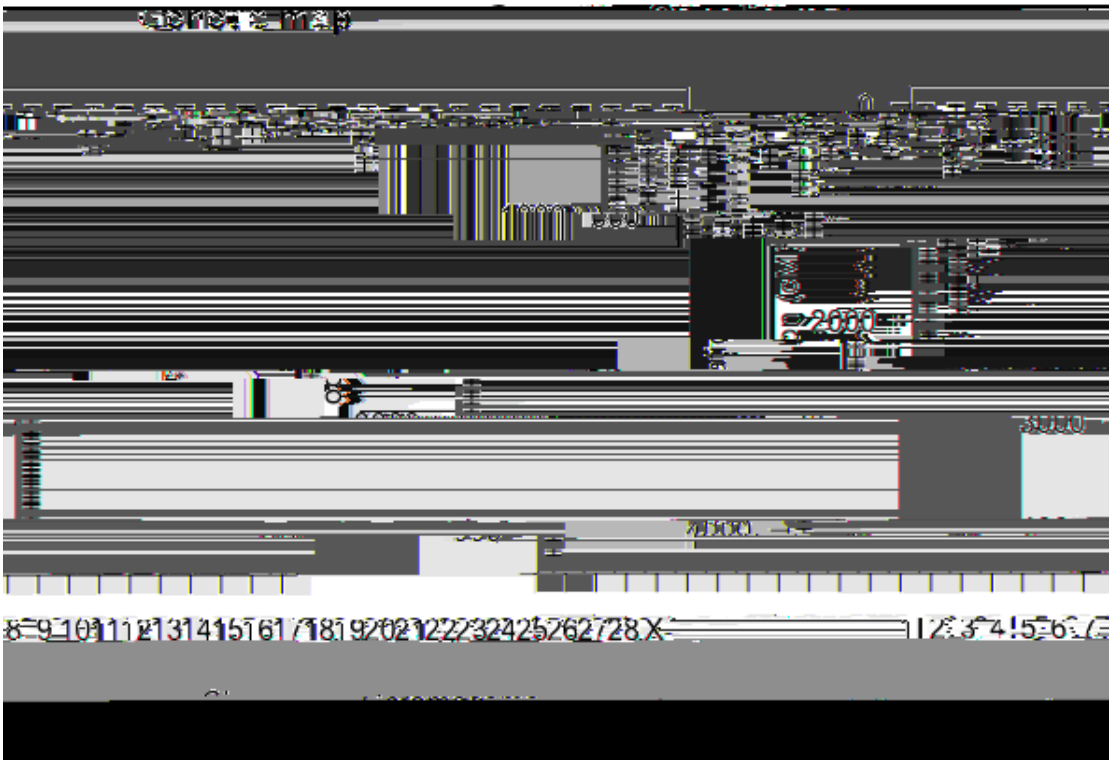

Figure 3: This graph displays the locations of all 739 genetic markers on the chromosomes. Each horizontal line represents a genetic marker and each vertical line represents the length of each chromosome, measured in centimorgans. Chromosome 1 is clearly the longest chromosome, and also seems to have the most genetic markers in this map.

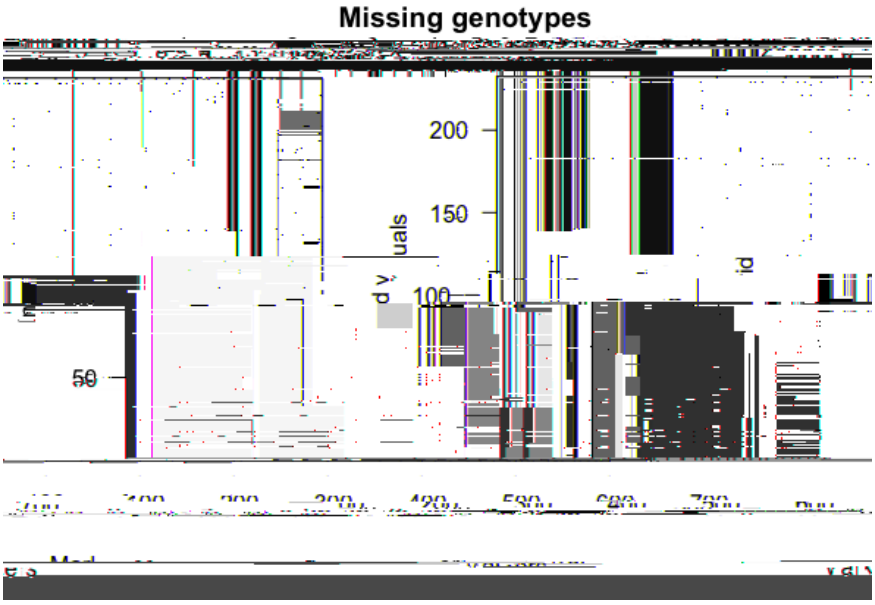

Figure 4: This plot shows available genetic data in white and missing data in black. Because there

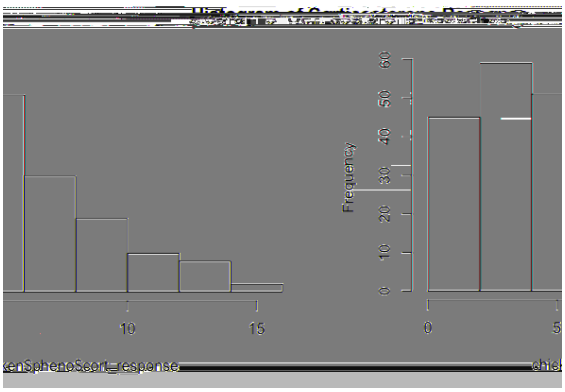

Figure 7: This graph shows a histogram of the Corticosterone values. Although shifted slightly to the left, the values are in a bell-shaped curve, meaning the results will be moderately reliable.

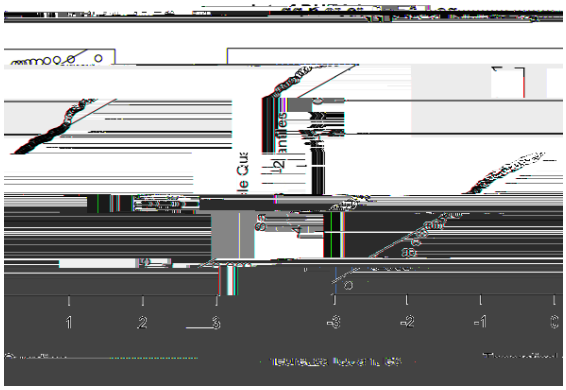

Figure 9: This is a qq plot of theoretical and sample DHEA quantities. The linear regression is almost 45 degrees and the data points have slight variation, which demonstrates near-perfect correlation and strong results.

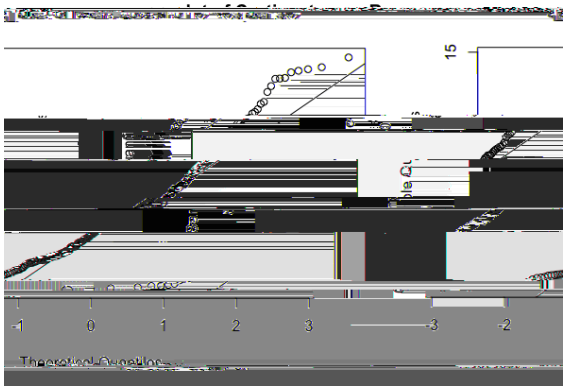

Figure 10: The nal qq plot is a graph of theoretical and sample Corticosterone values. This graph shows a strong correlation, suggesting reliable results.

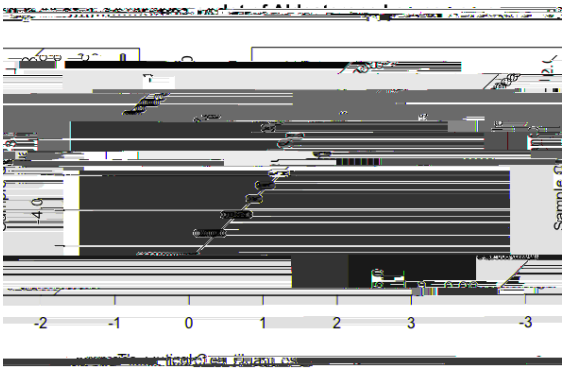

Figure 8: This is a qq plot of theoretical and sample Aldosterone quantities. The linear regression is almost a 45 degree diagonal line and the data points are close to the regression, suggesting a near-perfect correlation and strong results.

The next portion of the R/qtl model produced three phenotypic mainscans, for Aldosterone level, DHEA level, and Corticosterone level. Figures 11, 12, and 13 show these mainscans with text output embedded. The x-axis displays the locations of each genetic marker, and the y-axis displays LOD scores. High peaks represent high LOD scores, indicating signif-

icant loci that likely contribute to the coding of each trait. Loci with LOD scores greater than 3 can be considered signi cant and loci with LOD scores greater than the 95% con dence threshold were identi ed in this paper. Con dence thresholds were included for 67%, 90% and 95% con dence, shown by the blue, red, and green lines in each scan.

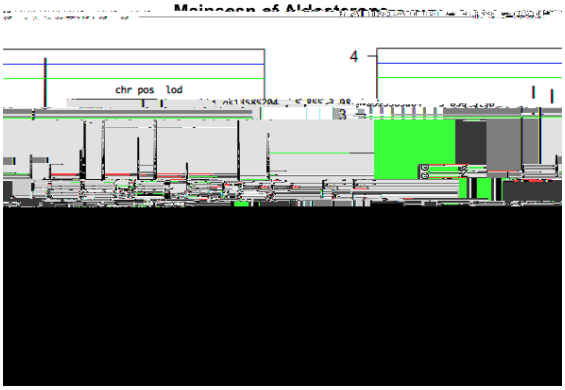

Figure 11: This graph shows an Aldosteron Mainscan. Three high peaks appear to be abo a LOD score of 3 and only one appears to b above the 95% threshold. It is easy to visualiz this signi cant locus as being on chromosome 5, and this is veri ed with the text output, which shows a signi cant locus on chromosome 5, at a location of 855 centimortans, and with a LOD score of 3.98.

Figure 12: This graph shows a DHEA Mainscan. Multiple high peaks appear to be above a LOD score of 3 and two appears to be above the 95% green threshold line. After visual analysis it seems one very high peak is located near chromosome 21 and one lower peak near chromosome 4. These results are con rmed with the text summary, which suggests two signi cant loci: one on chromosome 4, at 1174 centimorgans, with a LOD score of 4.75 and one on chromosome 21, near position 0, with a LOD score of 6.91.

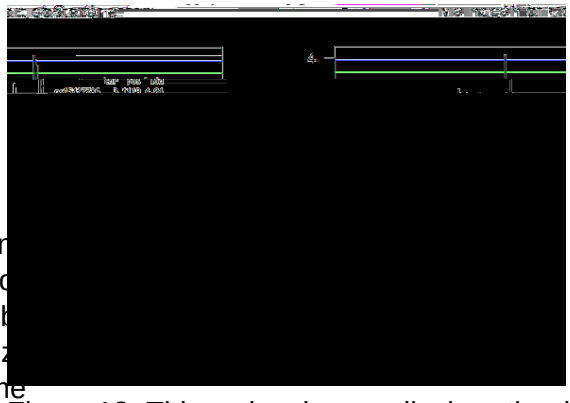

Figure 13: This nal mainscan displays the data for Corticosterone. Multiple high peaks appear to be above a LOD score of 3 and two appear to be above the 95% green threshold line. It is easy to visualize these signi cant loci as being on chromosomes 3 and 7. These results are also veri ed with the text output, which shows signi cant loci on chromosome 3, at 1119 centimorgans, with a LOD score of 4.04, and chromosome 7, at 342 centimorgans with a LOD score of 4.05. Because the two LOD scores for these two loci are so similar, it is likely that they both code almost equally for Corticosterone production.

After identifying the two signi cant loci for Corticosterone production, two con dence in-

terval scans were created. The point of using this type of diagnostic is to localize the individual signi cant loci of chromosome 3, as seen in gure 14. The second is of loci on chromosome 7 and can be seen in gure 15. The green line near the bottom of each graph represents a range that likely contains the most signi cant loci on each chromosome.

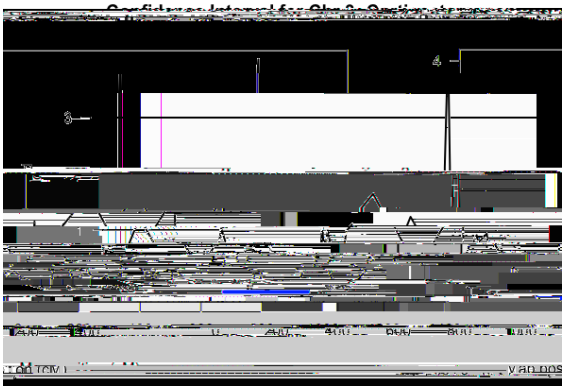

Figure 14: This con dence interval plot shows a signi cant interval of about 840 to 1120 centimorgans on chromosome 3, as suggested by the green interval line.

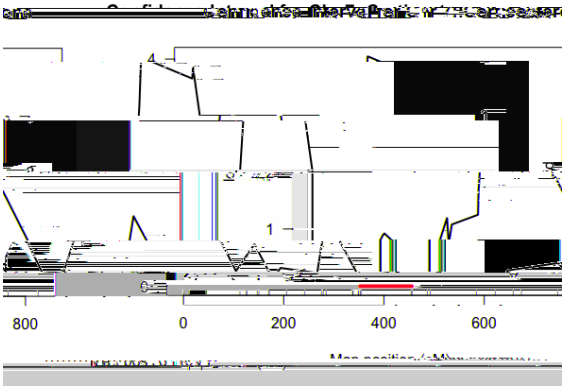

Figure 15: This con dence interval plot shows a signi cant interval of about 355 to 460 centimorgans on chromosome 7, as suggested by the green interval line.

The nal part of the R/atl model produced two

production differently. A computational chemistry model could be used to analyze hormone structure differences or dissociation constants, and could provide insight as to why different hormones cause different effects.

Acknowledgements

The author would like to thank Mr. Robert Gotwals for teaching the Introduction to Computational Science Course and providing the necessary knowledge to carry out this project. The author would also like to thank Dr. Dirk-Jan de Koning for his help with nding research data. Lastly, thanks is extended to the North Carolina School of Science and Mathematics for providing access to this class and numerous academic tools.

References

[1] Gautier, Zoe. ”Gallus gallus (red junglefowl).” Animal Diversity Web. N.p., n.d. Web. 23 Dec. 2016.

https://naldc.nal.usda.gov/download/34971/PDF.

[12] Chen, Zehua. Statistical methods for QTL mapping. Boca Raton: CRC Press,(Boca)no0s-

Data Tables and Code

Figure 5: Egg Production Affected by Corticosterone Datatable

R/qtl Code

```
# R script for analyzing QTL data in Chickens
# Alexander Adler
# Chicken Stress Hormone QTL dataset
# December 27, 2016

# clean things up
rm(list=ls())
# set working directory
setwd("/Users/aadler/Desktop/RFolder")


# load the QTL library
# NOTE!  I first had to INSTALL the library using: install.packages("qtl")
# Now I can use the package qtl
library(qtl)
```

```
# Load the data
chicken <- read.cross("csv", file="Chicken.csv",
genotypes = c("AA","AB","BB"), na.strings = "-", alleles = c("A","B"))
# Use jittermap to move markers apart slightly so my results are better
jittermap(chicken)
#  A summary of the cross gives me some basic data
summary(chicken)
# the names function tells me what phenotypes are in this dayaset
names(chicken$pheno)
# take a look at my data, make sure it's pretty clean
chicken <- est.rf(chicken)
plot.rf(chicken)
# It's nice to see my genetic map -- all of the horizontal lines are genetic
# markers that have been inserted
plot.map(chicken)
#  It's often the case that I have missing data -- plot.missing shows me
# where it is
plot.missing(chicken)

# renames variables for later use
ALD <- chicken$pheno$aldosterone_log2
DHEA <- chicken$pheno$DHEA_log
CORT <- chicken$pheno$cort_response

# histogram of Aldosterone Log phenotype
hist(chicken$pheno$aldosterone_log2, main = "Histogram of Aldosterone Log")
# histogram of DHEA Log phenotype
hist(chicken$pheno$DHEA_log, main = "Histogram of DHEA Log")
# histogram of Corticosterone phenotype
hist(chicken$pheno$cort_response, main = "Histogram of Corticosterone
Response")


# Another diagnostic (qq plots with linear regression lines)....if my data is
# relatively clean, I should get nice 45 degree diagonal lines
qqnorm(ALD, main = "qq plot of Aldosterone Log")
qqline(ALD, main = "qq plot of Aldosterone Log")
qqnorm(DHEA, main = "qq plot of DHEA Log")
qqline(DHEA, main = "qq plot of DHEA Log")
```

```
qqnorm(CORT, main = "qq plot of Corticosterone Response")
qqline(CORT, main = "qq plot of Corticosterone Response")


# Now I'm going to generate a mainscan.  First, I calculate what the scan
# should look like, so I'm going to calculate a genetic probability map.
chicken <- calc.genoprob(chicken, step = 2.0, off.end = 0.0, error.prob =
1.0e-4, map.function = "haldane", stepwidth = "fixed" )

# Run a simulated geno probability calculation
chicken <- sim.geno(chicken, step = 2.0, n.draws=32, error.prob = 1.0e-4,
map.function = "haldane", stepwidth = "fixed" )

# Perform the mainscan for the CORT Response QTL
# I'm going to run this Cort Response scan for 100 "permulations"
chicken.scanCORT <- scanone(chicken, pheno.col = 26, model = "normal",
method = "em")
chicken.scanCORT.perm <- scanone(chicken, pheno.col = 26, model = "normal",
method = "em", n.perm = 100)
# plot the CORT response mainscan
plot(chicken.scanCORT, main = "Mainscan of Corticosterone")
# I'm putting threshold lines at 67% confidence, 90% confidence, and 95%
# confidence.
thresh <- summary(chicken.scanCORT.perm, alpha = c(0.33, 0.10, 0.05))
abline(h=thresh[1], col = "blue")
abline(h=thresh[2], col = "red")
abline(h=thresh[3], col = "green")
# I'd like to see a text-based output of my CORT scan
summary(chicken.scanCORT, perm=chicken.scanCORT.perm, lodcolumn = 1,
alpha = 0.05)

# Perform the mainscan for the Aldosterone  QTL
# I'm going to run this ALD scan for 100 "permulations"
chicken.scanALD <- scanone(chicken, pheno.col = 78, model = "normal",
method = "em")
chicken.scanALD.perm <- scanone(chicken, pheno.col = 78, model = "normal",
method = "em", n.perm = 100)
# plot the ALD mainscan
plot(chicken.scanALD, main = "Mainscan of Aldosterone")
# I'm putting threshold lines at 67% confidence, 90% confidence, and 95%
# confidence.
```

```
thresh <- summary(chicken.scanALD.perm, alpha = c(0.37, 0.10, 0.05))
abline(h=thresh[1], col = "blue")
abline(h=thresh[2], col = "red")
abline(h=thresh[3], col = "green")
# I'd like to see a text-based output of my ALD scan
summary(chicken.scanALD, perm=chicken.scanALD.perm, lodcolumn = 1,
alpha = 0.05)


# Perform the mainscan for the DHEA QTL
# I'm going to run this DHEA scan for 100 "permulations"
chicken.scanDHEA <- scanone(chicken, pheno.col = 79, model = "normal",
method = "em")
chicken.scanDHEA.perm <- scanone(chicken, pheno.col = 79, model = "normal",
method = "em", n.perm = 100)
# plot the DHEA mainscan
plot(chicken.scanDHEA, main = "Mainscan of DHEA")
# I'm putting threshold lines at 67% confidence, 90% confidence, and 95%
confidence.
thresh <- summary(chicken.scanDHEA.perm, alpha = c(0.37, 0.10, 0.05))
abline(h=thresh[1], col = "blue")
abline(h=thresh[2], col = "red")
abline(h=thresh[3], col = "green")
# I'd like to see a text-based output of my DHEA scan
summary(chicken.scanDHEA, perm=chicken.scanDHEA.perm, lodcolumn = 1,
alpha = 0.05)



# Confidence Interval Plots for CORT
# First CI plot for CORT
CIchr3 <- bayesint(chicken.scanCORT, chr=3, prob=0.95)
plot(chicken.scanCORT, chr=3, lodcolumn = 1, main = "Confidence Interval for
Chr 3: Corticosterone")
lines(x=CIchr3[c(1,3), 2], y=c(0,0), type = "l", col = "green", lwd=4)
CIchr3[c(1,3),2]
# second CI plot for CORT
CIchr7 <- bayesint(chicken.scanCORT, chr=7, prob=0.95)
plot(chicken.scanCORT, chr=7, lodcolumn = 1, main = "Confidence Interval for
Chr 7: Corticosterone")
```

```
lines(x=CIchr7[c(1,3), 2], y=c(0,0), type = "l", col = "green", lwd=4)
CIchr7[c(1,3),2]


# do an effect plot for CORT
par(mfrow=c(1,2))
firstCORTeffect <- find.marker(chicken, chr = 3, pos = 1119)
effectplot(chicken, pheno.col = 26, mname1 = firstCORTeffect, main =
"Effect Plot for Cort: Chr 3")
secondCORTeffect <- find.marker(chicken, chr = 7, pos = 352)
effectplot(chicken, pheno.col = 26, mname1 = secondCORTeffect, main =
"Effect Plot for Cort: Chr 7")

# All done
detach(cross)
#EOF
```

Mathematica Code

```
eggData = Import["/Users/aadler/Desktop/EggData.csv"];
Grid[eggData, Frame -> All]

eggDataNoLabels = Delete[eggData, 1];
eggDataNoLabels2 = Delete[eggDataNoLabels, 1];
{age, eggProduction, eggProduction2} = Transpose[eggDataNoLabels2];

control = Transpose[{age, eggProduction}];
cortTreated = Transpose[{age, eggProduction2}];
dataplot =
 ListLinePlot[{control, cortTreated}, Mesh -> Full,
  AxesLabel -> {Age in weeks, Egg Production Percent}, PlotLabels ->
 {Control, Cortisol Treated},
  PlotLabel -> "Egg Production Percent by Age"]
```