Kochia\_QTL\_mapping\_workshop

## Extracting genotype information from a vcf file

Here, we take a vcf file that has been filtered, and extract information from the genotype field.

#install.packages('vcfR')  
library(vcfR)

##   
## \*\*\*\*\* \*\*\* vcfR \*\*\* \*\*\*\*\*  
## This is vcfR 1.12.0   
## browseVignettes('vcfR') # Documentation  
## citation('vcfR') # Citation  
## \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*

#read in an actual vcf file and convert/save as a vcfR object  
vcf <- read.vcfR( "Kochia\_F2\_GBS.vcf", verbose = FALSE)  
  
#extract the genotypes from the file  
head(vcf[,1:5])

## [1] "\*\*\*\*\* Object of class 'vcfR' \*\*\*\*\*"  
## [1] "\*\*\*\*\* Meta section \*\*\*\*\*"  
## [1] "##fileformat=VCFv4.1"   
## [2] "##source=\"GATK haplotype Caller\""   
## [3] "##some more comment lines about how the variant calling was done"  
## [1]   
## [1] "\*\*\*\*\* Fixed section \*\*\*\*\*"  
## CHROM POS ID REF ALT QUAL FILTER  
## [1,] "chr1" "0" NA "A" "T" "2441" NA   
## [2,] "chr1" "1000000" NA "A" "T" "3992" NA   
## [3,] "chr1" "2000000" NA "A" "T" "2103" NA   
## [4,] "chr1" "3000000" NA "A" "T" "825" NA   
## [5,] "chr1" "4000000" NA "A" "T" "1888" NA   
## [6,] "chr1" "5000000" NA "A" "T" "2553" NA   
## [1]   
## [1] "\*\*\*\*\* Genotype section \*\*\*\*\*"  
## FORMAT 1 2   
## [1,] "GT:AD:DP:GQ:PL" "0/1:10,11:21:0,0:0" "0/0:25,0:25:0,0:0"  
## [2,] "GT:AD:DP:GQ:PL" "0/1:9,9:18:0,0:0" "0/0:16,0:16:0,0:0"  
## [3,] "GT:AD:DP:GQ:PL" "1/1:0,20:20:0,0:0" "0/0:19,0:19:0,0:0"  
## [4,] "GT:AD:DP:GQ:PL" "1/1:0,16:16:0,0:0" "0/0:18,0:18:0,0:0"  
## [5,] "GT:AD:DP:GQ:PL" "1/1:0,28:28:0,0:0" "0/0:19,0:19:0,0:0"  
## [6,] "GT:AD:DP:GQ:PL" "1/1:0,17:17:0,0:0" "0/0:23,0:23:0,0:0"  
## 3 4   
## [1,] "0/1:13,9:22:0,0:0" "0/1:10,8:18:0,0:0"   
## [2,] "0/1:16,14:30:0,0:0" "0/1:13,12:25:0,0:0"  
## [3,] "0/1:15,11:26:0,0:0" "0/1:7,9:16:0,0:0"   
## [4,] "1/1:0,9:9:0,0:0" "0/1:13,11:24:0,0:0"  
## [5,] "1/1:0,14:14:0,0:0" "0/1:15,16:31:0,0:0"  
## [6,] "1/1:0,22:22:0,0:0" "0/1:7,11:18:0,0:0"   
## [1]   
## [1] "Unique GT formats:"  
## [1] "GT:AD:DP:GQ:PL"  
## [1]

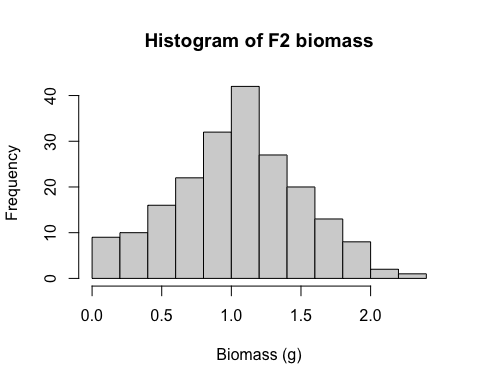
geno <- extract.gt(vcf)  
geno\_df <- as.data.frame(geno, sep = ',')  
head(geno\_df[,1:5])

## 1 2 3 4 5  
## chr1\_0 0/1 0/0 0/1 0/1 0/1  
## chr1\_1000000 0/1 0/0 0/1 0/1 0/1  
## chr1\_2000000 1/1 0/0 0/1 0/1 0/1  
## chr1\_3000000 1/1 0/0 1/1 0/1 0/1  
## chr1\_4000000 1/1 0/0 1/1 0/1 0/1  
## chr1\_5000000 1/1 0/0 1/1 0/1 0/1

write.csv(geno\_df, file = "Kochia\_genotypes.csv")  
  
#Assign alleles based on parental genotypes by calling an external python script  
#Arguments IN ORDER: /path/to/python\_script.py /path/to/geno\_df.csv Name\_of\_R\_parent Name\_of\_S\_parent /path/to/alleles\_out.csv  
system('python assign\_alleles.py ./Kochia\_genotypes.csv R\_parent S\_parent ./Kochia\_alleles.csv', wait = FALSE)

## Loading information into a “cross” object using the qtl2 package

library(qtl2)  
#load in the data by calling the control file  
cross <- read\_cross2("kochia\_qtl\_control\_file.yaml")  
  
#plot a histogram of biomass  
hist(cross$pheno, xlab = "Biomass (g)", main = "Histogram of F2 biomass")



##Prep the data for a QTL scan and conduct the scan

#calculate genetic probabilities at each marker. Default error rate is .0001  
pr <- calc\_genoprob(cross=cross, map=cross$gmap, error\_prob = .002)  
  
#look at the names of the entries in the pr data structure  
names(pr)

## [1] "chr1" "chr6"

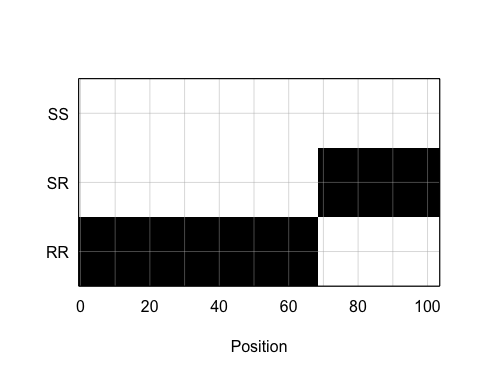
#look at the dimension names for the chromosome 6 array  
dimnames(pr$'chr6')

## [[1]]  
## [1] "1" "2" "3" "4" "5" "6"   
## [7] "7" "8" "9" "10" "11" "12"   
## [13] "13" "14" "15" "16" "17" "18"   
## [19] "19" "20" "21" "22" "23" "24"   
## [25] "25" "26" "27" "28" "29" "30"   
## [31] "31" "32" "33" "34" "35" "36"   
## [37] "37" "38" "39" "40" "41" "42"   
## [43] "43" "44" "45" "46" "47" "48"   
## [49] "49" "50" "51" "52" "53" "54"   
## [55] "55" "56" "57" "58" "59" "60"   
## [61] "61" "62" "63" "64" "65" "66"   
## [67] "67" "68" "69" "70" "71" "72"   
## [73] "73" "74" "75" "76" "77" "78"   
## [79] "79" "80" "81" "82" "83" "84"   
## [85] "85" "86" "87" "88" "89" "90"   
## [91] "91" "92" "93" "94" "95" "96"   
## [97] "97" "98" "99" "100" "101" "102"   
## [103] "103" "104" "105" "106" "107" "108"   
## [109] "109" "110" "111" "112" "113" "114"   
## [115] "115" "116" "117" "118" "119" "120"   
## [121] "121" "122" "123" "124" "125" "126"   
## [127] "127" "128" "129" "130" "131" "132"   
## [133] "133" "134" "135" "136" "137" "138"   
## [139] "139" "140" "141" "142" "143" "144"   
## [145] "145" "146" "147" "148" "149" "150"   
## [151] "151" "152" "153" "154" "155" "156"   
## [157] "157" "158" "159" "160" "161" "162"   
## [163] "163" "164" "165" "166" "167" "168"   
## [169] "169" "170" "171" "172" "173" "174"   
## [175] "175" "176" "177" "178" "179" "180"   
## [181] "181" "182" "183" "184" "185" "186"   
## [187] "187" "188" "189" "190" "191" "192"   
## [193] "193" "194" "195" "196" "197" "198"   
## [199] "199" "200" "R\_parent" "S\_parent"  
##   
## [[2]]  
## [1] "SS" "SR" "RR"  
##   
## [[3]]  
## [1] "chr6\_0" "chr6\_1000000" "chr6\_2000000" "chr6\_3000000"   
## [5] "chr6\_4000000" "chr6\_5000000" "chr6\_6000000" "chr6\_7000000"   
## [9] "chr6\_8000000" "chr6\_9000000" "chr6\_10000000" "chr6\_11000000"   
## [13] "chr6\_12000000" "chr6\_13000000" "chr6\_14000000" "chr6\_15000000"   
## [17] "chr6\_16000000" "chr6\_17000000" "chr6\_18000000" "chr6\_19000000"   
## [21] "chr6\_20000000" "chr6\_21000000" "chr6\_22000000" "chr6\_23000000"   
## [25] "chr6\_24000000" "chr6\_25000000" "chr6\_26000000" "chr6\_27000000"   
## [29] "chr6\_28000000" "chr6\_29000000" "chr6\_30000000" "chr6\_31000000"   
## [33] "chr6\_32000000" "chr6\_33000000" "chr6\_34000000" "chr6\_35000000"   
## [37] "chr6\_36000000" "chr6\_37000000" "chr6\_38000000" "chr6\_39000000"   
## [41] "chr6\_40000000" "chr6\_41000000" "chr6\_42000000" "chr6\_43000000"   
## [45] "chr6\_44000000" "chr6\_45000000" "chr6\_46000000" "chr6\_47000000"   
## [49] "chr6\_48000000" "chr6\_49000000" "chr6\_50000000" "chr6\_51000000"   
## [53] "chr6\_52000000" "chr6\_53000000" "chr6\_54000000" "chr6\_55000000"   
## [57] "chr6\_56000000" "chr6\_57000000" "chr6\_58000000" "chr6\_59000000"   
## [61] "chr6\_60000000" "chr6\_61000000" "chr6\_62000000" "chr6\_63000000"   
## [65] "chr6\_64000000" "chr6\_65000000" "chr6\_66000000" "chr6\_67000000"   
## [69] "chr6\_68000000" "chr6\_69000000" "chr6\_70000000" "chr6\_71000000"   
## [73] "chr6\_72000000" "chr6\_73000000" "chr6\_74000000" "chr6\_75000000"   
## [77] "chr6\_76000000" "chr6\_77000000" "chr6\_78000000" "chr6\_79000000"   
## [81] "chr6\_80000000" "chr6\_81000000" "chr6\_82000000" "chr6\_83000000"   
## [85] "chr6\_84000000" "chr6\_85000000" "chr6\_86000000" "chr6\_87000000"   
## [89] "chr6\_88000000" "chr6\_89000000" "chr6\_90000000" "chr6\_91000000"   
## [93] "chr6\_92000000" "chr6\_93000000" "chr6\_94000000" "chr6\_95000000"   
## [97] "chr6\_96000000" "chr6\_97000000" "chr6\_98000000" "chr6\_99000000"   
## [101] "chr6\_100000000" "chr6\_101000000" "chr6\_102000000" "chr6\_103000000"

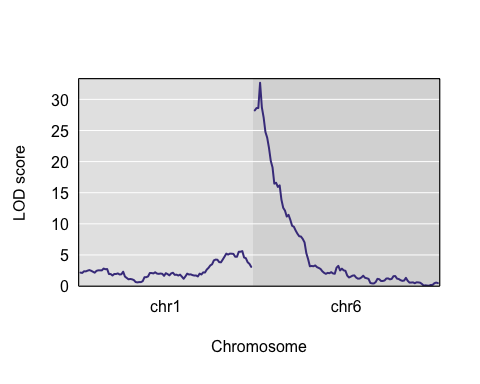
#view the first three rows of genotype probabilities for a genotyped marker on chromosome 6  
(pr$`chr6`)[1:3,,"chr6\_52000000"]

## SS SR RR  
## 1 1.007912e-11 2.007896e-07 9.999998e-01  
## 2 2.007492e-07 9.999996e-01 2.007492e-07  
## 3 2.007492e-07 9.999996e-01 2.007492e-07

#plot the probabilities of individual 1 genotypes across chromosome 6  
plot\_genoprob(pr, map=cross$gmap, ind = 1, chr = "chr6")

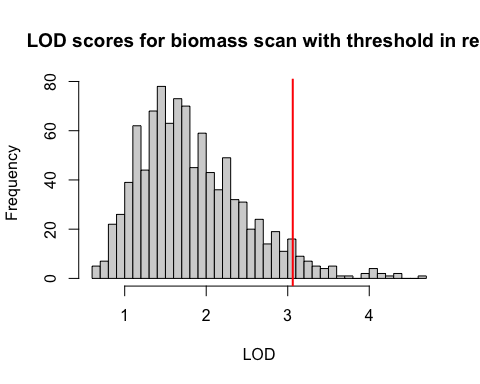


#haley-knott regression plots genotype probability on a spectrum (x-axis) and phenotype on the y-axis. A linear regression (including any covariates you provide) will show if there is an effect of genotype on phenotype.  
#perform genome scan using haley-knott regression on the data previously generated  
out <- scan1(genoprobs=pr, pheno =cross$pheno)  
  
#plot the LOD scores for the liver LOD scores  
plot\_scan1(out, map = cross$gmap, lodcolumn = "biomass")



## Using permutation to determine a significance threshold for QTL peaks

#permutation mixes the phenotypes up so there should be no relationship between genotype and phenotype  
#the 95th percentile calculated LOD score from these permutation is considered the 95% confidence threshold for statistical significance   
operm <- scan1perm(genoprobs = pr, pheno = cross$pheno, n\_perm = 1000)  
  
#this outputs a list of the maximum LOD score achieved from each permutation  
#plot these values  
{hist(operm[,'biomass'], breaks = 50, xlab = "LOD", main = "LOD scores for biomass scan with threshold in red")  
abline(v = summary(operm)[,'biomass'], col = 'red', lwd = 2)}



#operm has calculated the 95th percentile (5% of random LOD are above this line)  
#return thresholds of different alphas  
summary(operm, alpha= c(.02, .05))

## LOD thresholds (1000 permutations)  
## biomass  
## 0.02 3.47  
## 0.05 3.06

#to find significant peaks, we use the findpeaks function with the scan output, thresholds, and marker map  
#save the permutation data so we can call it later  
thr <- summary(operm)  
  
#find the peaks with 95% confidence (thr) and find the genomic interval that is 95% likely to contain a causative locus (based on a bayesian credible interval)  
find\_peaks(scan1\_output = out, map = cross$gmap, threshold = thr, prob = 0.95, expand2markers = FALSE)

## lodindex lodcolumn chr pos lod ci\_lo ci\_hi  
## 1 1 biomass chr1 91 5.617069 81 93  
## 2 1 biomass chr6 3 32.661039 3 3

#find multiple peaks on a chromosome with 'peakdrop' to tell how far the LOD needs to drop between peaks  
find\_peaks(scan1\_output = out, map = cross$gmap, threshold = thr, peakdrop = 1.8, prob = 0.95, expand2markers = FALSE)

## lodindex lodcolumn chr pos lod ci\_lo ci\_hi  
## 1 1 biomass chr1 91 5.617069 81 93  
## 2 1 biomass chr6 3 32.661039 3 3

## Extracting predicted genes in your QTL intervals

#this bash command will filter only predicted genes in your chr1 interval and extract the predicted function of the protein product (all from the gff annotation file)  
#in this interval we find many proteins, but the Glutathione S-transferase-like protein is an interesting candidate for sure  
grep gene Kochia\_annotation\_formatted.gff3 | awk '$1=="chr1"' | awk '$4>81000000' | awk '$4<93000000' > ./chr1\_qtl\_genes.txt  
  
grep gene Kochia\_annotation\_formatted.gff3 | awk '$1=="chr1"' | awk '$4>81000000' | awk '$4<93000000' | cut -d$'\t' -f 9 | rev | cut -d '=' -f 1 | rev | sort | uniq > ./chr1\_qtl\_gene\_names.txt  
  
cat ./chr1\_qtl\_gene\_names.txt

## 1,4-alpha-glucan-branching enzyme 1, chloroplastic/amyloplastic (EC 2.4.1.18) (Starch branching enzyme I)  
## 1,4-alpha-glucan-branching enzyme 2, chloroplastic/amyloplastic (EC 2.4.1.18) (Q-enzyme) (Starch-branching enzyme IIB)  
## 3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase (3-beta-HSD) [Includes: 3-beta-hydroxy-Delta(5)-steroid dehydrogenase (EC 1.1.1.145) (3-beta-hydroxy-5-ene steroid dehydrogenase) (Progesterone reductase); Steroid Delta-isomerase (EC 5.3.3.1) (Delta-5-3-ketosteroid isomerase)]  
## 3-hydroxy-3-methylglutaryl coenzyme A reductase 2-B (HMG-CoA reductase 2) (Hydroxymethylglutaryl-CoA reductase) (PgHMGR2) (EC 1.1.1.34)  
## 40S ribosomal protein S12  
## 5'-3' exoribonuclease 4 (EC 3.1.13.-) (Protein ACC INSENSITIVE 1) (Protein ETHYLENE INSENSITIVE 5) (Protein EXORIBONUCLEASE 4)  
## 50S ribosomal protein 6, chloroplastic (CL25) (Plastid-specific 50S ribosomal protein 6) (PSRP-6)  
## 50S ribosomal protein L10e  
## 65-kDa microtubule-associated protein 3 (AtMAP65-3) (Protein PLEIADE)  
## 65-kDa microtubule-associated protein 6 (AtMAP65-6)  
## A-kinase anchor protein 17A (AKAP-17A) (721P) (B-lymphocyte antigen) (Protein XE7) (Protein kinase A-anchoring protein 17A) (PRKA17A) (Splicing factor, arginine/serine-rich 17A)  
## AP2-like ethylene-responsive transcription factor SNZ (Protein SCHNARCHZAPFEN)  
## AT-hook motif nuclear-localized protein 18  
## AT-hook motif nuclear-localized protein 28  
## AT-hook motif nuclear-localized protein 6  
## ATP-dependent DNA helicase CHL1 (EC 3.6.4.12) (Chromosome loss protein 1)  
## ATP-dependent DNA/RNA helicase DHX36 (EC 3.6.4.12) (EC 3.6.4.13) (DEAD/H box polypeptide 36) (DEAH-box protein 36) (G4-resolvase-1) (G4R1) (MLE-like protein 1) (RNA helicase associated with AU-rich element protein)  
## ATP-dependent zinc metalloprotease FtsH 3 (EC 3.4.24.-)  
## Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)  
## Acetylpolyamine amidohydrolase (APAH) (EC 3.5.1.-) (Acetylcadaverine deacetylase) (Acetylpolyamine deacetylase) (Acetylputrescine deacetylase) (EC 3.5.1.62)  
## Acyl-CoA-binding domain-containing protein 2 (Acyl-CoA binding protein 2)  
## Adenine phosphoribosyltransferase 1, chloroplastic (APRT 1) (AtAPT1) (EC 2.4.2.7)  
## Adenosine deaminase-like protein (EC 3.5.4.-) (Adenosine deaminase-like protein isoform 1) (N6-mAMP deaminase) (HsMAPDA) (N6-methyl-AMP aminohydrolase)  
## Albumin-2 (24 kDa albumin) (LS-24) (PA2)  
## Alkaline phosphatase D (APaseD) (EC 3.1.3.1)  
## Ankyrin repeat and SAM domain-containing protein 1A (Odin)  
## Auxin response factor 5 (Auxin-responsive protein IAA24) (Transcription factor MONOPTEROS)  
## Auxin-induced protein 15A  
## Auxin-induced protein 6B  
## Basic leucine zipper 4 (AtbZIP4)  
## Beta-glucosidase 10 (Os4bglu10) (EC 3.2.1.21)  
## Bifunctional protein FolD 4, chloroplastic (Protein EMBRYO DEFECTIVE 3127) (Tetrahydrofolate dehydrogenase/cyclohydrolase 4) [Includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9)]  
## Bromodomain-containing protein 1 (Bromodomain and PHD finger-containing protein 2)  
## COP9 signalosome complex subunit 7 (CSN complex subunit 7) (Protein FUSCA 5)  
## Calcineurin B-like protein 1  
## Calcium-dependent protein kinase 21 (EC 2.7.11.1)  
## Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type (EC 7.2.2.10) (Calcium pump)  
## Catalase isozyme 1 (EC 1.11.1.6)  
## Catalase-2 (EC 1.11.1.6)  
## Cell division control protein 2 homolog D (EC 2.7.11.22) (EC 2.7.11.23)  
## Cellulose synthase-like protein G3 (AtCslG3) (EC 2.4.1.-)  
## Chaperone protein DnaJ  
## Chaperone protein dnaJ 72 (AtDjC72) (AtJ72)  
## Chlorophyll a-b binding protein 8, chloroplastic (LHCI type III CAB-8)  
## Chlorophyll a-b binding protein CP24 10A, chloroplastic (CAB-10A) (LHCP)  
## Citrate synthase, mitochondrial (EC 2.3.3.16)  
## Coiled-coil domain-containing protein 115  
## Copia protein (Gag-int-pol protein) [Cleaved into: Copia VLP protein; Copia protease (EC 3.4.23.-)]  
## Cyclin-D3-2 (G1/S-specific cyclin-D3-2) (CycD3;2)  
## Cyclin-dependent kinase 1 (CDK1) (EC 2.7.11.22) (EC 2.7.11.23) (Cell division control protein 2 homolog) (Cell division protein kinase 1) (p34 protein kinase)  
## Cysteine-rich protein 2 (CRP-2)  
## Cysteine-rich receptor-like protein kinase 25 (Cysteine-rich RLK25) (EC 2.7.11.-)  
## Cytochrome P450 704B1 (Long-chain fatty acid omega-hydroxylase) (EC 1.14.14.80)  
## Cytochrome P450 714A1 (EC 1.14.-.-) (EUI-like P450 A1)  
## Cytochrome P450 81Q32 (CrCYP81Q32) (EC 1.14.-.-)  
## Cytokinin riboside 5'-monophosphate phosphoribohydrolase (EC 3.2.2.n1) (Protein LONELY GUY homolog) (LOG homolog)  
## DNA repair protein XRCC3 (X-ray repair cross-complementing protein 3)  
## DNA repair protein XRCC3 homolog (X-ray repair cross-complementing protein 3 homolog) (AtXRCC3)  
## DNA-directed RNA polymerase I subunit RPA2 (EC 2.7.7.6) (DNA-directed RNA polymerase I polypeptide 2) (RNA polymerase I subunit 2)  
## DNA-directed RNA polymerase III subunit 2 (DNA-directed RNA polymerase III subunit RPC2) (DNA polymerase I subunit C2) (EC 2.7.7.6) (Nuclear RNA polymerase C2)  
## Dipeptidyl peptidase family member 6 (EC 3.4.14.-)  
## Disease resistance protein RGA2 (Blight resistance protein RPI) (RGA2-blb)  
## Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1A (Ribophorin IA) (RPN-IA) (Ribophorin-1A)  
## Dynamin-1 (EC 3.6.5.5) (B-dynamin) (D100) (Dynamin, brain)  
## Dynamin-like 120 kDa protein, mitochondrial (EC 3.6.5.5) (Optic atrophy protein 1 homolog) [Cleaved into: Dynamin-like 120 kDa protein, form S1]  
## Dynamin-related protein 1E (Dynamin-like protein 4) (Dynamin-like protein DLP2) (Dynamin-like protein E)  
## E3 ubiquitin-protein ligase RDUF1 (EC 2.3.2.27) (RING and DUF1117 domain-containing protein 1) (AtRDUF1) (RING-type E3 ubiquitin transferase RDUF1)  
## E3 ubiquitin-protein ligase SDIR1 (EC 2.3.2.27) (Protein SALT- AND DROUGHT-INDUCED RING FINGER 1) (RING-type E3 ubiquitin transferase SDIR1)  
## E3 ubiquitin-protein ligase UPL3 (Ubiquitin-protein ligase 3) (EC 2.3.2.26) (HECT ubiquitin-protein ligase 3) (HECT-type E3 ubiquitin transferase UPL3) (Protein KAKTUS)  
## EPIDERMAL PATTERNING FACTOR-like protein 2 (EPF-like protein 2) [Cleaved into: MEPFL2]  
## ETO1-like protein 1 (Ethylene overproducer 1-like protein 1)  
## Endo-1,4-beta-xylanase 1 (Xylanase 1) (EC 3.2.1.8) (1,4-beta-D-xylan xylanohydrolase 1)  
## Endo-1,4-beta-xylanase 2 (Xylanase 2) (EC 3.2.1.8) (1,4-beta-D-xylan xylanohydrolase 2)  
## Endoglucanase 24 (EC 3.2.1.4) (Endo-1,4-beta glucanase 24)  
## Ent-kaurene oxidase 2 (OsKO2) (EC 1.14.14.86) (Cytochrome P450 701A6) (Ent-kaurene oxidase-like 2) (OsKOL2) (OsKOS3) (Protein DWARF 35)  
## Equilibrative nucleotide transporter 1 (AtENT1) (Nucleoside transporter ENT1)  
## Ergosterol biosynthetic protein 28 homolog  
## Ethylene receptor (EC 2.7.13.3)  
## F-box/kelch-repeat protein At5g42350  
## Filament-like plant protein 4 (AtFPP4)  
## Flavonoid 3',5'-hydroxylase 1 (F3'5'H) (EC 1.14.14.81) (CYPLXXVA1) (Cytochrome P450 75A1)  
## Flavonol synthase/flavanone 3-hydroxylase (FLS) (EC 1.14.11.9) (EC 1.14.20.6) (Fragment)  
## G1/S-specific cyclin-D2  
## G1/S-specific cyclin-E  
## GTP-binding protein Rit2 (EC 3.6.5.2) (Ras-like protein expressed in neurons) (Ras-like without CAAX protein 2)  
## Gamete and mating-type specific protein A (EC 3.4.22.-)  
## Geraniol 8-hydroxylase (EC 1.14.14.83) (Cytochrome P450 76B10) (Geraniol 10-hydroxylase) (SmG10H)  
## Germin-like protein 3-1  
## Germin-like protein 9-1  
## Germin-like protein 9-3  
## Gibberellin 2-beta-dioxygenase 7 (EC 1.14.11.13) (GA 2-oxidase 7) (Gibberellin 2-beta-hydroxylase 7) (Gibberellin 2-oxidase 7)  
## Gibberellin 20-oxidase-like protein (EC 1.14.11.-)  
## Gibberellin-regulated protein 11 (GAST1 protein homolog 11)  
## Gibberellin-regulated protein 7 (GAST1 protein homolog 7)  
## Glutamate receptor ionotropic, NMDA 3B (GluN3B) (N-methyl-D-aspartate receptor subunit NR3B) (NMDAR3B) (NR3B) (NMDA receptor 4) (Nr4)  
## Glutathione S-transferase-like protein FUS3 (EC 2.5.1.-) (Fusarin biosynthesis protein 3)  
## Glyceraldehyde-3-phosphate dehydrogenase B, chloroplastic (EC 1.2.1.13) (NADP-dependent glyceraldehydephosphate dehydrogenase subunit B)  
## Glycerol-3-phosphate acyltransferase, chloroplastic (GPAT) (EC 2.3.1.15)  
## Growth-regulating factor 9 (AtGRF9) (Transcription activator GRF9)  
## Guanidinoacetate N-methyltransferase (EC 2.1.1.2)  
## HMG1/2-like protein  
## HMG1/2-like protein (Protein SB11)  
## HORMA domain-containing protein 1  
## Heptahelical transmembrane protein 1 (PAQR family protein HHP1)  
## Heterogeneous nuclear ribonucleoprotein A0 (hnRNP A0)  
## High mobility group B protein 14 (Nucleosome/chromatin assembly factor group D 14)  
## High mobility group B protein 3 (High mobility group protein B 2) (AtHMGbeta2) (HMG beta 2) (Nucleosome/chromatin assembly factor group D 03) (Nucleosome/chromatin assembly factor group D 3)  
## Homeobox protein knotted-1-like 2 (Fragment)  
## IQ domain-containing protein IQM3 (IQ motif-containing protein 3)  
## Inositol-tetrakisphosphate 1-kinase 6 (EC 2.7.1.134) (Inositol 1,3,4-trisphosphate 5/6-kinase 6) (Inositol-triphosphate 5/6-kinase 6) (Ins(1,3,4)P(3) 5/6-kinase 6) (OsITP5/6K-6) (OsITPK6) (EC 2.7.1.159)  
## Interferon-related developmental regulator 2 (Protein SKMC15)  
## Kelch-like protein 3  
## Kinesin-like protein KIN-10A (Phragmoplast-associated kinesin-related protein 2) (AtPAKRP2)  
## Kinesin-like protein NACK1 (NPK1-activating kinesin-1)  
## Kunitz trypsin inhibitor 5 (AtKTI5) (Kunitz trypsin inhibitor 2) (AtKTI2)  
## Lariat debranching enzyme (EC 3.1.-.-)  
## Lecithin-cholesterol acyltransferase-like 4 (EC 2.3.1.-)  
## Leucine-rich repeat receptor-like protein kinase PEPR2 (EC 2.7.11.1) (Elicitor peptide 1 receptor 2) (PEP1 receptor 2)  
## Lissencephaly-1 homolog (Pronuclear migration abnormal protein 1)  
## MLO-like protein 5 (AtMlo5)  
## MTOR-associated protein MEAK7 (MEAK7) (TBC/LysM-associated domain-containing protein 1) (TLD domain-containing protein 1)  
## Manganese transporter pdt1  
## Mannose/glucose-specific lectin (Fragment)  
## Maturase K (Intron maturase)  
## Midasin (AtMDN1) (Dynein-related AAA-ATPase MDN1) (MIDAS-containing protein 1) (Protein DWARF AND SHORT ROOT 1)  
## Mitogen-activated protein kinase kinase kinase ANP1 (EC 2.7.11.25) (Arabidopsis NPK1-related kinase 1)  
## Mitotic checkpoint serine/threonine-protein kinase BUB1 (hBUB1) (EC 2.7.11.1) (BUB1A)  
## Multiple C2 and transmembrane domain-containing protein  
## N-acetylglucosamine kinase 1 (GlcNAc kinase 1) (EC 2.7.1.59) (Hexokinase 1) (EC 2.7.1.1)  
## NADH dehydrogenase [ubiquinone] iron-sulfur protein 1, mitochondrial (EC 7.1.1.2) (Protein EMBRYO DEFECTIVE 1467)  
## NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial (Complex I-18 kDa) (CI-18 kDa) (Complex I-AQDQ) (CI-AQDQ) (NADH-ubiquinone oxidoreductase 18 kDa subunit)  
## NADH-dependent flavin oxidoreductase iccE (EC 1.-.-.-) (Ilicicolin H biosynthesis cluster protein E)  
## NADH-dependent flavin oxidoreductase nadA (EC 1.-.-.-) (Aflatoxin biosynthesis protein nadA)  
## NADPH dehydrogenase (EC 1.6.99.1)  
## NADPH dehydrogenase 3 (EC 1.6.99.1) (Old yellow enzyme 3)  
## NADPH-dependent aldehyde reductase ARI1 (EC 1.1.1.-)  
## NPC intracellular cholesterol transporter 1 (Niemann-Pick C1 protein)  
## Nuclear/nucleolar GTPase 2  
## Nucleotide-sugar uncharacterized transporter 2  
## Oxalate oxidase GF-2.8 (EC 1.2.3.4) (Germin GF-2.8)  
## Pantothenate kinase 2 (AtPANK2) (Pantothenic acid kinase 2) [Includes: Pantothenate kinase (EC 2.7.1.33); 4'-phosphopantetheine phosphatase (EC 3.1.3.-)]  
## Pentatricopeptide repeat-containing protein At1g63080, mitochondrial  
## Pentatricopeptide repeat-containing protein At2g39230, mitochondrial (Protein LATERAL ORGAN JUNCTION)  
## Pentatricopeptide repeat-containing protein At4g14850 (Protein LOVASTATIN INSENSITIVE 1)  
## Pentatricopeptide repeat-containing protein At4g20770  
## Pentatricopeptide repeat-containing protein At5g11310, mitochondrial  
## Pentatricopeptide repeat-containing protein At5g16640, mitochondrial  
## Pentatricopeptide repeat-containing protein At5g57250, mitochondrial  
## Peptidyl-prolyl cis-trans isomerase FKBP10 (PPIase FKBP10) (EC 5.2.1.8) (FK506-binding protein 10) (FKBP-10) (Rotamase)  
## Peptidyl-prolyl cis-trans isomerase FKBP16-4, chloroplastic (PPIase FKBP16-4) (EC 5.2.1.8) (FK506-binding protein 16-4) (AtFKBP16-4) (Immunophilin FKBP16-4) (Rotamase)  
## Peptidyl-prolyl cis-trans isomerase FKBP17-2, chloroplastic (PPIase FKBP17-2) (EC 5.2.1.8) (FK506-binding protein 17-2) (AtFKBP17-2) (Immunophilin FKBP17-2) (Rotamase)  
## Peroxisome biogenesis protein 7 (Peroxin-7) (AtPEX7) (Peroxisomal targeting signal type 2 receptor) (Pex7p)  
## Phenylalanine--tRNA ligase beta subunit, cytoplasmic (EC 6.1.1.20) (Phenylalanyl-tRNA synthetase beta subunit) (PheRS)  
## Phospholipase D delta (AtPLDdelta) (PLD delta) (EC 3.1.4.4)  
## Phytosulfokines 3 (AtPSK3) [Cleaved into: Phytosulfokine-alpha (PSK-alpha) (Phytosulfokine-a); Phytosulfokine-beta (PSK-beta) (Phytosulfokine-b)]  
## Poly(rC)-binding protein 1 (Alpha-CP1) (Heterogeneous nuclear ribonucleoprotein E1) (hnRNP E1)  
## Potassium-transporting ATPase ATP-binding subunit (EC 7.2.2.6) (ATP phosphohydrolase [potassium-transporting] B chain) (Potassium-binding and translocating subunit B) (Potassium-translocating ATPase B chain)  
## Pre-mRNA-splicing factor CWC22  
## Probable 26S proteasome regulatory subunit p27  
## Probable GABA transporter 2  
## Probable GTP-binding protein OBGM, mitochondrial  
## Probable branched-chain-amino-acid aminotransferase (BCAT) (EC 2.6.1.42)  
## Probable disease resistance protein At5g63020 (pNd11)  
## Probable glucuronosyltransferase Os04g0650300 (EC 2.4.-.-) (OsGT43B)  
## Probable glutathione S-transferase GSTU1 (EC 2.5.1.18)  
## Probable inactive serine/threonine-protein kinase bub1  
## Probable leucine-rich repeat receptor-like protein kinase IMK3 (EC 2.7.11.1) (Protein INFLORESCENCE MERISTEM RECEPTOR-LIKE KINASE 3) (Protein MERISTEMATIC RECEPTOR-LIKE KINASE)  
## Probable methyltransferase At1g29790 (EC 2.1.1.-)  
## Probable methyltransferase PMT22 (EC 2.1.1.-)  
## Probable pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH8 (EC 3.6.4.13) (DEAH RNA helicase homolog PRP2)  
## Probable proline transporter 2  
## Probable protein kinase UbiB (EC 2.7.-.-) (Ubiquinone biosynthesis protein UbiB)  
## Probable protein phosphatase 2C 5 (OsPP2C05) (EC 3.1.3.16)  
## Probable protein phosphatase 2C T23F11.1 (PP2C) (EC 3.1.3.16)  
## Probable serine/threonine-protein kinase CST (EC 2.7.11.1) (Protein CAST AWAY)  
## Probable xyloglucan glycosyltransferase 12 (EC 2.4.1.-) (Cellulose synthase-like protein C12) (AtCslC12)  
## Protein ABC transporter 1, mitochondrial (ABC1At) (AtABC1) (EC 2.7.-.-)  
## Protein ACTIVITY OF BC1 COMPLEX KINASE 3, chloroplastic (ABC1-LIKE KINASE 3) (EC 2.7.-.-) (EC 2.7.11.1) (Protein REPRESSOR OF BDR1)  
## Protein DETOXIFICATION 26 (AtDTX26) (Multidrug and toxic compound extrusion protein 26) (MATE protein 26)  
## Protein DOG1-like 4  
## Protein ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1 (EC 2.3.1.-)  
## Protein ETHYLENE-INSENSITIVE 2 (OsEIN2) (Protein MAO HUZI 7)  
## Protein FAR-RED IMPAIRED RESPONSE 1  
## Protein FAR1-RELATED SEQUENCE 1  
## Protein FAR1-RELATED SEQUENCE 4  
## Protein FAR1-RELATED SEQUENCE 6  
## Protein IWS1 homolog (IWS1-like protein)  
## Protein MAIN-LIKE 2  
## Protein MAINTENANCE OF MERISTEMS  
## Protein MANNAN SYNTHESIS-RELATED 2 (AtMSR2) (EC 2.4.1.-) (O-fucosyltransferase 12) (O-FucT-12) (O-fucosyltransferase family protein)  
## Protein MOS2  
## Protein NLP7 (AtNLP7) (NIN-like protein 7) (Nodule inception protein-like protein 7)  
## Protein NO VEIN (Protein EMBRYO DEFECTIVE 2597)  
## Protein O-mannosyl-transferase TMTC1 (EC 2.4.1.109) (Transmembrane and TPR repeat-containing protein 1)  
## Protein PHLOEM PROTEIN 2-LIKE A10 (AtPP2-A10)  
## Protein SOB FIVE-LIKE 5 (AtSOFL5)  
## Protein canopy-1 (Protein D121)  
## Protein of unknown function  
## Protein p34  
## Protein patched homolog 2 (PTC2)  
## Purple acid phosphatase (EC 3.1.3.2) (Zinc(II) purple acid phosphatase)  
## Putative COX1/OXI3 intron 2 protein  
## Putative DUF21 domain-containing protein At1g03270 (CBS domain-containing protein CBSDUF4)  
## Putative NADPH dehydrogenase C5H10.10 (EC 1.6.99.1) (Old yellow enzyme homolog 2)  
## Putative NADPH-dependent methylglyoxal reductase GRP2 (EC 1.1.1.283) (Cytoplasmic antigenic protein 2)  
## Putative disease resistance protein RGA3 (Blight resistance protein B149) (RGA1-blb)  
## Putative dynamin-related protein 4A  
## Putative lipase ROG1 (EC 3.1.-.-) (Revertant of glycogen synthase kinase mutation protein 1)  
## Putative pentatricopeptide repeat-containing protein At3g23330  
## Putative receptor-like protein kinase At1g80870 (EC 2.7.11.1)  
## Putative replication factor C small subunit L499 (RFC small subunit L499) (Clamp loader small subunit L499)  
## Putative ribonuclease H protein At1g65750 (EC 3.1.26.4)  
## Putative serine protease K12H4.7 (EC 3.4.-.-)  
## Putative transcription factor bHLH086 (Basic helix-loop-helix protein 86) (AtbHLH86) (bHLH 86) (Protein ROOT HAIR DEFECTIVE 6-LIKE 1) (AtRSL1) (Protein RHD SIX-LIKE 1) (Transcription factor EN 113) (bHLH transcription factor bHLH086)  
## Pyrethroid hydrolase (EC 3.1.1.88)  
## Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha (PFP) (6-phosphofructokinase, pyrophosphate dependent) (PPi-PFK) (Pyrophosphate-dependent 6-phosphofructose-1-kinase)  
## RIO-type serine/threonine-protein kinase Rio1 (EC 2.7.11.1) (EC 3.6.3.-)  
## Receptor expression-enhancing protein 4  
## Receptor-like protein kinase 5 (EC 2.7.10.1) (EC 2.7.11.1) (Protein HAESA)  
## Replication factor A 51 kDa subunit (RP-A p51) (Replication factor-A protein 1) (RF-A protein 1) (Single-stranded DNA-binding protein P51 subunit)  
## Replication protein A 70 kDa DNA-binding subunit C (OsRPA70c) (Replication factor A protein 1C) (Replication protein A 1C)  
## Retinoblastoma-related protein 2 (OsRBR2)  
## Ribosome biogenesis protein TSR1 (20S rRNA accumulation protein 1)  
## Ribosome biogenesis protein bms1  
## SNF1-related protein kinase regulatory subunit gamma-1 (AKIN subunit gamma-1) (AKING1) (AKINgamma1) (CBS domain-containing protein CBSCBS1)  
## SUPPRESSOR OF ABI3-5 (REQUIRED FOR SNC4-1D protein 1) (Splicing factor SUA)  
## SWI/SNF complex subunit SWI3C (AtSWI3C) (Transcription regulatory protein SWI3C)  
## Serine carboxypeptidase-like 50 (EC 3.4.16.-)  
## Serine/threonine-protein kinase BRI1-like 1 (EC 2.7.11.1) (BRASSINOSTEROID INSENSITIVE 1-like protein 1)  
## Serine/threonine-protein kinase Nek5 (EC 2.7.11.1) (NimA-related protein kinase 5) (AtNEK6) (AtNek5)  
## Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B (PP6-ARS-B) (Serine/threonine-protein phosphatase 6 regulatory subunit ARS-B) (Ankyrin repeat domain-containing protein 44)  
## Serine/threonine-protein phosphatase 7 long form homolog (EC 3.1.3.16) (Protein MAIN-LIKE 3)  
## Serpin B4 (Leupin) (Peptidase inhibitor 11) (PI-11) (Squamous cell carcinoma antigen 2) (SCCA-2)  
## Short-chain dehydrogenase/reductase family 42E member 1 (EC 1.1.1.-)  
## Sodium-coupled neutral amino acid transporter 2 (Amino acid transporter A2) (Solute carrier family 38 member 2) (System A amino acid transporter 2) (System A transporter 1) (System N amino acid transporter 2)  
## Solute carrier family 45 member 3 (Prostate cancer-associated protein 6) (Prostein)  
## Spermidine hydroxycinnamoyl transferase (EC 2.3.1.-) (BAHD-like hydroxycinnamoyl transferase)  
## Squalene monooxygenase SE1 (EC 1.14.14.17) (Squalene epoxidase 1) (PgSQE1) (SE) (SE1) (gse)  
## Squamosa promoter-binding-like protein 7  
## Sterol regulatory element-binding protein cleavage-activating protein (SCAP) (SREBP cleavage-activating protein)  
## Subtilisin-like protease SBT1.3 (EC 3.4.21.-) (Subtilase subfamily 1 member 3) (AtSBT1.3)  
## Sugar transport protein 4 (Hexose transporter 4)  
## Sulfate adenylyltransferase (EC 2.7.7.4) (ATP-sulfurylase) (Sulfate adenylate transferase) (SAT)  
## Syntaxin-61 (AtSYP61) (Osmotic stress-sensitive mutant 1)  
## TPR repeat-containing protein ZIP4 (Protein ZIP4 homolog) (OsZIP4)  
## Testis-expressed protein 10 homolog  
## Tetratricopeptide repeat protein 1 (TPR repeat protein 1)  
## Tetratricopeptide repeat protein 7A (TPR repeat protein 7A)  
## Transcription factor LRL2 (Basic helix-loop-helix protein 69) (AtbHLH69) (bHLH 69) (Protein DEFECTIVE REGION OF POLLEN 2) (Protein LJRHL1- LIKE 2) (AtLRL2) (Transcription factor EN 94) (Transcription factor bHLH69) (bHLH transcription factor bHLH069)  
## Transcription factor MYB52 (Myb-related protein 52) (AtMYB52) (Protein ABA-HYPERSENSITIVE 1)  
## Transcription factor TGA1 (DNA-binding protein TGA1a-like protein) (bZIP transcription factor 47) (AtbZIP47)  
## Transport and Golgi organization protein 2  
## Transposon TX1 uncharacterized 149 kDa protein (ORF 2)  
## Transposon Ty3-I Gag-Pol polyprotein (Gag3-Pol3) (Transposon Ty3-2 TYA-TYB polyprotein) [Cleaved into: Capsid protein (CA) (p24); Spacer peptide p3; Nucleocapsid protein p11 (NC); Ty3 protease (PR) (EC 3.4.23.-) (p16); Spacer peptide J; Reverse transcriptase/ribonuclease H (RT) (RT-RH) (EC 2.7.7.49) (EC 2.7.7.7) (EC 3.1.26.4) (p55); Integrase p52 (IN); Integrase p49 (IN)]  
## Tryptophan decarboxylase 2 (EC 4.1.1.-)  
## U-box domain-containing protein 11 (EC 2.3.2.27) (Plant U-box protein 11) (RING-type E3 ubiquitin transferase PUB11)  
## U-box domain-containing protein 17 (EC 2.3.2.27) (Plant U-box protein 17) (RING-type E3 ubiquitin transferase PUB17)  
## U-box domain-containing protein 26 (EC 2.3.2.27) (Plant U-box protein 26) (RING-type E3 ubiquitin transferase PUB26)  
## U-box domain-containing protein 8 (EC 2.3.2.27) (Plant U-box protein 8) (RING-type E3 ubiquitin transferase PUB8)  
## UBA domain-containing protein 7  
## UDP-glucose 4-epimerase GEPI42 (EC 5.1.3.2) (Galactowaldenase) (UDP-galactose 4-epimerase)  
## UDP-glycosyltransferase 79B10 (EC 2.4.1.-)  
## UDP-glycosyltransferase 79B6 (EC 2.4.1.-)  
## Ubiquitin-conjugating enzyme E2 C (EC 2.3.2.23) ((E3-independent) E2 ubiquitin-conjugating enzyme C) (EC 2.3.2.24) (E2 ubiquitin-conjugating enzyme C) (Ubiquitin carrier protein C) (Ubiquitin-protein ligase C)  
## Ubiquitin-conjugating enzyme E2-18 kDa (EC 2.3.2.23) (E2 ubiquitin-conjugating enzyme 7) (Ubiquitin carrier protein) (Ubiquitin-protein ligase)  
## Uncharacterized RNA-binding protein C4F6.14  
## Uncharacterized WD repeat-containing protein alr3466  
## Uncharacterized acetyltransferase At3g50280 (EC 2.3.1.-)  
## Uncharacterized calcium-binding protein C800.10c  
## Uncharacterized mitochondrial protein AtMg00310 (ORF154)  
## Uncharacterized protein KIAA0930 homolog  
## Uncharacterized protein RP120  
## Uncharacterized protein in hydrogenase 1 5'region (Fragment)  
## Uncharacterized protein sll0103  
## Uric acid transporter UacT  
## Uroporphyrinogen decarboxylase 1, chloroplastic (UPD1) (URO-D1) (EC 4.1.1.37)  
## V-type proton ATPase subunit a, vacuolar isoform (V-ATPase a 1 subunit) (V-ATPase 95 kDa subunit) (Vacuolar pH protein 1) (Vacuolar proton pump a subunit) (Vacuolar proton translocating ATPase subunit a 1)  
## Vacuolar import and degradation protein 27  
## Vacuolar protein sorting-associated protein 2 homolog 2 (AtVPS2-2) (Charged multivesicular body protein 2 homolog 2) (ESCRT-III complex subunit VPS2 homolog 2)  
## Very-long-chain aldehyde decarbonylase CER1 (EC 4.1.99.5) (Protein ECERIFERUM 1)  
## Voltage-gated ClC-type chloride channel ClcB  
## WD repeat-containing protein PCN (Protein POPCORN)  
## WEB family protein At1g75720  
## Wall-associated receptor kinase-like 15 (EC 2.7.11.-)  
## Werner syndrome ATP-dependent helicase homolog (EC 3.6.4.12) (Exonuclease WRN) (EC 3.1.-.-)  
## Xylulose kinase (Xylulokinase) (EC 2.7.1.17)  
## Xylulose kinase 2 (Atxk-2) (Xylulokinase 2) (EC 2.7.1.17) (1-Deoxy-D-Xylulokinase) (DXK)  
## Zinc finger BED domain-containing protein 4  
## Zinc finger BED domain-containing protein DAYSLEEPER (Transposase-like protein DAYSLEEPER)  
## Zinc finger BED domain-containing protein RICESLEEPER 1 (Transposase-like protein RICESLEEPER 1)  
## Zinc finger protein ZAT4  
## bZIP transcription factor 60 (AtbZIP60)

#this will do the same for the chr6 interval  
#notice the Cytochrome P450 protein in the interval. We will have to validate to be sure it is responsible for resistance.  
grep gene Kochia\_annotation\_formatted.gff3 | awk '$1=="chr6"' | awk '$4>2500000' | awk '$4<3500000' > ./chr6\_qtl\_genes.txt  
  
grep gene Kochia\_annotation\_formatted.gff3 | awk '$1=="chr6"' | awk '$4>2500000' | awk '$4<3500000' | cut -d$'\t' -f 9 | rev | cut -d '=' -f 1 | rev | sort | uniq > ./chr6\_qtl\_gene\_names.txt  
  
cat ./chr6\_qtl\_gene\_names.txt

## ADP-ribosylation factor-like protein 5A  
## ADP-ribosylation factor-like protein 8d (AtARL8d) (ADP-ribosylation factor-like A1B) (AtARLA1B)  
## AP2-like ethylene-responsive transcription factor At2g41710  
## APO protein 1, chloroplastic (Accumulation of photosystem I protein 1) (Protein ACCUMULATION OF PHOTOSYSTEM ONE 1)  
## ATP synthase subunit d, mitochondrial (ATPase subunit d)  
## Aspartyl protease 37 (EC 3.4.23.-)  
## BTB/POZ domain-containing protein SR1IP1 (Protein ATSR1-INTERACTION PROTEIN 1)  
## Casein kinase II subunit alpha-1 (CK II) (EC 2.7.11.1) (Casein kinase alpha 1) (AtCKA1)  
## Cathepsin B-like cysteine proteinase 4 (EC 3.4.22.-) (Cysteine protease-related 4)  
## Cell wall integrity protein scw1 (Strong cell wall protein 1)  
## Chaperone protein DnaJ  
## Chromatin modification-related protein EAF3  
## Cyclin-D5-1 (G1/S-specific cyclin-D5-1) (CycD5;1)  
## Cysteine-rich receptor-like protein kinase 26 (Cysteine-rich RLK26) (EC 2.7.11.-)  
## Cytochrome P450 81Q32 (CrCYP81Q32) (EC 1.14.-.-)  
## DNA damage-binding protein 2 (DDB p48 subunit) (DDBb) (Damage-specific DNA-binding protein 2) (UV-damaged DNA-binding protein 2) (UV-DDB 2)  
## Ethylene-responsive transcription factor ERF021  
## GH3 domain-containing protein  
## Gibberellin-regulated protein 8 (GAST1 protein homolog 8)  
## L10-interacting MYB domain-containing protein  
## LRR receptor-like serine/threonine-protein kinase GSO1 (EC 2.7.11.1) (Protein GASSHO 1) (Protein SCHENGEN 3)  
## MYB-like transcription factor ODO1 (Protein ODORANT1) (PhODO1)  
## Mitochondrial import receptor subunit TOM70 (Mitochondrial precursor proteins import receptor) (Translocase of outer membrane 70 kDa subunit) (Translocase of outer mitochondrial membrane protein 70)  
## Naringenin,2-oxoglutarate 3-dioxygenase (EC 1.14.11.9) (FHT) (Flavanone-3-hydroxylase) (F3H)  
## POC1 centriolar protein homolog (Proteome of centrioles 1)  
## Pentatricopeptide repeat-containing protein At3g03580  
## Peroxidase 2 (EC 1.11.1.7) (Fragment)  
## Pimeloyl-[acyl-carrier protein] methyl ester esterase (EC 3.1.1.85) (Biotin synthesis protein BioH) (Carboxylesterase BioH)  
## Probable WRKY transcription factor 33 (WRKY DNA-binding protein 33)  
## Probable alpha-L-arabinofuranosidase C (ABF C) (Arabinosidase C) (EC 3.2.1.55)  
## Probable glucuronosyltransferase Os10g0205300 (EC 2.4.-.-)  
## Probable prolyl 4-hydroxylase 9 (AtP4H9) (EC 1.14.11.2)  
## Prolyl 4-hydroxylase subunit alpha-1 (4-PH alpha-1) (EC 1.14.11.2) (Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-1)  
## Protein ALP1-like (EC 3.1.-.-)  
## Protein ETHYLENE-INSENSITIVE 2 (OsEIN2) (Protein MAO HUZI 7)  
## Protein FRIGIDA  
## Protein NONRESPONDING TO OXYLIPINS 2, mitochondrial  
## Protein TRI1  
## Protein of unknown function  
## Putative U-box domain-containing protein 46 (EC 2.3.2.27) (Plant U-box protein 46) (RING-type E3 ubiquitin transferase PUB46)  
## Putative UPF0481 protein At3g02645  
## Putative invertase inhibitor (Pollen allergen Pla a 1) (allergen Pla a 1.0101)  
## Putative nuclease HARBI1 (EC 3.1.-.-) (Harbinger transposase-derived nuclease)  
## Putative oxidoreductase C6G10.06 (EC 1.-.-.-)  
## Putative tyrosine-protein phosphatase auxilin (EC 3.1.3.48) (DnaJ homolog subfamily C member 6)  
## RNA exonuclease 1 (EC 3.1.-.-) (RNase H(70))  
## RNA polymerase II C-terminal domain phosphatase-like 2 (FCP-like 2) (EC 3.1.3.16) (Carboxyl-terminal phosphatase-like 2) (AtCPL2) (CTD phosphatase-like 2)  
## Ribonuclease III domain-containing protein RNC1, chloroplastic (Chloroplast ribonuclease III domain protein)  
## Small COPII coat GTPase SAR1 (EC 3.6.5.-)  
## Subtilisin-like protease SBT1.8 (EC 3.4.21.-) (Subtilase subfamily 1 member 8) (AtSBT1.8)  
## Tetratricopeptide repeat protein 1 (TPR repeat protein 1)  
## Transposon Ty3-I Gag-Pol polyprotein (Gag3-Pol3) (Transposon Ty3-2 TYA-TYB polyprotein) [Cleaved into: Capsid protein (CA) (p24); Spacer peptide p3; Nucleocapsid protein p11 (NC); Ty3 protease (PR) (EC 3.4.23.-) (p16); Spacer peptide J; Reverse transcriptase/ribonuclease H (RT) (RT-RH) (EC 2.7.7.49) (EC 2.7.7.7) (EC 3.1.26.4) (p55); Integrase p52 (IN); Integrase p49 (IN)]  
## UBX domain-containing protein 6 (UBX domain-containing protein 1)  
## Ubiquitin-conjugating enzyme E2 30 (EC 2.3.2.23) (E2 ubiquitin-conjugating enzyme 30) (Ubiquitin carrier protein 30)  
## Uncharacterized protein At2g23090  
## Uncharacterized protein MJ1426  
## Uncharacterized protein ycf36  
## Vacuolar protein sorting-associated protein 13A (Chorea-acanthocytosis protein homolog) (Chorein) (Fragment)  
## Vesicle transport protein USE1 (USE1-like protein)  
## Zinc finger MYM-type protein 1  
## Zinc finger protein 4  
## tRNA-dihydrouridine(47) synthase [NAD(P)(+)]-like (EC 1.3.1.-) (tRNA-dihydrouridine synthase 3-like)