Development of Genetic Resources for the Identification of Disease Resistance Genes in *Cicer arietinum*

By

PARKER WHEELER

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

PLANT PATHOLOGY

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Professor Douglas Cook, Chair

Professor Gitta Coaker

Assistant Professor Tiffany Lowe-Power

Committee in Charge

2023

Abstract

Chickpeas (*Cicer arietinum*) act as an important source of nutrition in many developing countries. In recent years, the productivity of the chickpea crop has been negatively impacted by *Fusarium oxysporum* f. sp. *ciceri.* To address this problem, efforts have been made to find resistance genes present in chickpea germplasm to combat this disease. This project is focused on improving genomic resources in chickpea with the goal of identifying a resistance gene that was predicted previously in a quantitative trait locus (QTL) study. Three cultivated accessions were sequenced and analyzed for structural and genetic differences and an assay was designed for fine mapping the QTL region conferring disease resistance. The sequencing resulted in improvement of the reference genome and expansion of chickpea genomic resources. The assay design will be applied in further research in an attempt to identify the causal gene for Fusarium wilt resistance which can be applied to breeding efforts to deploy more resistant cultivars.

Introduction/Background

Chickpeas (*Cicer arietinum*) are a vital food source for many developing countries. Their seeds contain a high level of vitamins and minerals and are densely packed with protein, making them a naturally nutrient rich food (Wallace et al. 2016). Ethiopia is one of the world's top producers of chickpea, and the largest producer in Africa, generating 457,320 metric tons of crop in 2020 (FAO). Over 90% of farmed chickpeas are consumed in the same country in which they are grown (Kassie et al, 2009). This means that chickpeas are critical in the health and economy in Ethiopia, providing a dense nutrient source as well as a source of income for Ethiopian growers.

While chickpeas are an advantageous crop for fulfilling nutritional needs, they are severely lacking in cultivars that can maintain high productivity when planted at scale. Most yield reductions are due to biotic and abiotic stresses, and there are insufficient disease resistant and stress tolerant cultivars available to combat these problems (Millan et al, 2006). Additionally, within the cultivated chickpea germplasm, there is a very low level of diversity, making trait development and molecular mapping a challenge (Sunkad et al. 2019, Abbo et al. 2003).

The chickpea crop in Ethiopia is threatened by a variety of plant diseases, one of the most devastating being Fusarium wilt, caused by *Fusarium oxysporum* f. sp. *ciceris* (Foc). Susceptible plants are infected by Foc penetrating the root tissue primarily through the root meristem or root hairs, without the need for wounding or specialized penetration structures (Jiménez-Fernández et al, 2013). Once inside the root, the fungus proceeds to colonize the xylem tissue, leading to systemic infection (Jiménez-Fernández et al, 2013). Foc is a soilborne pathogen and is known to persist in the soil in the form of chlamydospores for up to six years, even after planting of

susceptible crops has ceased (Singh et al. 2007). When a plot is infected with Fusarium wilt, the yield losses can be immense, typically 10-40%, but up to 100% yield loss in severe cases (Garg et al. 2018). The persistence of chlamydospores means that each subsequent year, the losses due to Fusarium wilt will increase as susceptible plant hosts are growing in an increasingly infested environment. Since it is both soilborne and forms long-lasting survival structures, *F. oxysporum* is extremely difficult to remove from fields once it is present. Fungicides can be applied via fumigation, but this is hazardous to the environment and extremely costly, so it is not a feasible option for the growers in Ethiopia. The difficulty of treatment makes resistant cultivars the preferred option, and a crucial tool in overcoming the threat of Fusarium wilt disease in developing countries (Singh et al. 2007).

Eight different pathogenic races of Foc have been identified to date, which include 0, 1A, 1B/C, 2, 3, 4, 5, and 6 (Jendoubi et al. 2017). Races 0, 2, 3, and 4 have been reported as being present in Ethiopia and nearby countries (Shehabu et al. 2008). It is important that resistance to these races be well understood when developing wilt resistant cultivars for use in Ethiopia. Resistance has been previously described in several cultivars, but ongoing development of new resistance strategies will proactively remedy any resistance-breaking events. Of the four cultivars present in Ethiopia, race 0 causes yellowing, while races 2, 3, and 4 cause wilting symptoms (Kelly et al. 1994). To date, no cultivar possesses robust resistance to all pathogenic races of Foc.

Previously, a novel quantitative trait locus (QTL) has been identified on chromosome 4 which segregates as a single recessive locus and confers resistance to a broad range of Ethiopian Foc strains. This QTL was discovered using a genotyping-by-sequencing method (RADseq) (Davey

and Blaxter, 2010) in a cross between a resistant parent and susceptible parent, Dhera and JG 62, respectively. RADseq identified 836 segregating SNP markers, which was sufficient to identify a physically broad major QTL responsible for Fusarium wilt resistance, spanning over 15 Mbp. While this QTL confirms the presence of a major effect resistance gene, a more precise location of the responsible gene is necessary for marker-assisted breeding. The current project focused primarily on resistant Dhera and susceptible JG 62, but also developed genomic sequence information for WR 315, another known resistant accession.

Towards this end, this research aimed to develop molecular genetic, genomic, and germplasm tools to fine map the gene underlying the recessive resistance to Foc. Genomic data was used to identify high density single nucleotide polymorphisms (SNPs), SNPs were converted to PCR-based assays for allele discrimination, and new recombinant populations are being developed to screen individual plants for informative recombination events on chromosome four. These tools and populations will provide the basis to narrow the causal haplotype and nominate candidate genes responsible for the observed resistance. Importantly, the resulting linked molecular markers will provide tools for breeding programs which can apply the information to introgress the resistance allele into other chickpea varieties.

Results and Discussion

I. Sequencing Genomes

Due to the low genetic diversity between cultivated chickpea accessions, it was necessary to sequence each parent genome to identify polymorphisms between them. PacBio HiFi sequencing was employed due to the extremely long reads that are produced, which can span regions with

many repeats, allowing for greater confidence in assembling and comparing repeat-rich regions among the sequenced accessions. While sequencing the genomes does not guarantee the identification of useful SNPs, it does give a more complete view of each parent genome. Additionally, long reads are useful for marker development and improvement of the quality of the chickpea reference genome which used HiSeq2000, a short-read technology (Varshney et al, 2013).

Assessment of raw sequencing reads based on FastQC results indicated that the data was of high quality based on GC content, sequence length, and Phred scores (Table 1). Surprisingly, GenomeScope, a k-mer analysis software, predicted that the Dhera genotype has a genome size of 413 Mb, and JG 62 has a genome size of 382 Mb. This estimate is about 300 Mb smaller than the reported chickpea genome size. This is likely due to an overestimation in coverage by the GenomeScope software which has been reported in other cases (Pflug et al, 2020). The coverage reported by GenomeScope was nearly twice what was expected for each genome, ultimately resulting in an underestimate of genome size. By contrast, assembly with Hifiasm 0.0.16.0 (Cheng et al. 2021) yielded genomes of the expected size for the three sequenced accessions, from ~710 to ~730 Mpb, and gapless contig N50s between 12.8 and 45.4 Mbp.

High contiguity assemblies such as those obtained here were impossible even as recently as 2022 and can be attributed to improvements both in sequencing platforms (PacBio Revio) and assembly tools. In the JG 62 genome, for example, the entire genome is in 415 gapless pieces and fully half of the genome is in 6 ungapped contigs. Indeed, several chromosomes are present as single ungapped molecules and the largest contig is 120.6 Mbp. By contrast, when the CDC

Frontier genome was reported in 2013 (Varshney et al) the reference genome contained 62,619 contigs with an N50 of 23.5 kb, which represent roughly a 1,000-fold improvement in assembly quality. Moreover, compared to the 2013 genome, which could confidently assign 345 Mb to chromosomes with many misplaced scaffolds, the data reported here provides unambiguous assignment of >700 Mpb to chromosomes with minimal or no placement error.

The completeness of these three genomes relative to the reference genome allows for a better understanding of repeat-rich regions, structural variation, and polymorphisms between cultivated accessions. Moreover, gene discovery will be both more accurate (essentially no gaps in the gene space) and complete, and for the first time, centromere structure can be compared among accessions.

All sequenced genomes were aligned to each other and the reference genome 96029, another cultivated accession (Figures 1-6). When aligned to each other, the JG 62 genome is highly colinear to the reference genome 96029, another cultivated accession (Figure 1). WR 315 also aligns quite closely to 96029, although it shows some areas of potential duplications (Figure 2). The WR 315 genome has not yet been improved to a chromosome level assembly, which may be responsible for some added noise in the dot plots with this genome. In addition, JG 62 and WR 315 align quite closely to each other (Figure 3). Dhera, however, shows some divergence from the 96029 reference genome, JG 62, and WR 315. When Dhera is aligned to WR 315, the dot plot shows a lack of colinearlity between the genomes, particularly in chromosomes one, four, seven, and two (Figure 4). When Dhera is aligned to either 96029 or JG 62, the dot plot shows a Dhera-specific translocation between chromosomes 1 and 6 (Figures 5 and 6). Moreover,

repetitive sequences on chromosome 4, apparently at the centromere, lack simple co-linearity between the genomes as shown in the comparison of Dhera with JG 62. (Figure 6). The divergence of repeats on chromosome four is consistent with the possibility that Dhera has retained a centromeric region from a wild progenitor during prior breeding efforts. Centromeric regions are repeat-rich and highly plastic (Das et al, 2017), and diverge quickly during speciation (Henikoff et al, 2001). Even when centromeric regions are highly similar they are typified by low recombination rates (Choo 1998). Foreign chromosomes, which typically lack extended centromere sequence similarity compared to native chromosomes, cause further reductions to recombination, may impose linkage drag on linked genic regions, and can be the source of genetic distortion and meiotic non-disjunction (Henikoff et al. 2001). Deokar et al. (2014) reported strong segregation distortion on chickpea chromosome 4, with chromosome four having almost no recombination in the region of interest to this study. Interestingly, numerous studies have reported high rates of polymorphism on chromosome 4 relative to other chromosomes. Nevertheless, there is evidence for recombination occurring between repeats in centromeric regions of plant cells (Fernandes et al. 2019), meaning that recombinants may be found if F2 populations produced in this project are of sufficient size.

A SNP density map between Dhera and JG 62 further supports that there is high dissimilarity between the two genomes over a span of chromosome four, most notably between approximately 14 Mb to 40 Mb, which includes the same region identified previously by the QTL study and the assembly dot plots (Figure 7). Chromosome four also appears to have the highest SNP density of any chromosome between Dhera and JG 62 (Figure 7).

II. Candidate Gene Evaluation

The Legume Information System's CicerMine was used to search for candidate disease resistance genes present on chromosome four (https://mines.legumeinfo.org/cicermine/begin.do). This search yielded fifteen different candidate genes, which may have been limited by the annotation data as well as the terms used in the gene descriptions (Table 2, Figure 8). The identified genes shared many Interpro tags such as "disease resistance protein", "NB-ARC", "P-loop containing nucleoside triphosphate hydrolase", and "leucine rich repeat", and GO terms including "defense response", "ADP binding", and nucleoside-triphosphate activity".

Annotations were further assigned to discovered genes/proteins based on homology, either using the CicerMine database or by searching the National Center for Biotechnology Information using BLAST. Four genes were annotated as CC-NBS-LRR proteins, including Ca4g086900 and three tandemly-arrayed genes (Ca4g342800, Ca4g343100 and Ca4g343300), with the latter genes having greatest similarity to the RPP13-like protein 2 in *Arabidopsis thaliana*. Three additional genes, including two in tandem, were annotated as TIR-NBS-LRRs proteins (Ca4g179100, Ca4g019900, and Ca4g019800), while two genes correspond to LRR-type proteins (Ca4g142700 and Ca4g178800) and could be PAMP receptors. Five additional genes (Ca4g137800, Ca4g137900, Ca4g192200, Ca4g243300 and Ca4g243400) encode dirigent-like proteins, which while not involved in plant pathogen signaling, they are implicated in lignan formation during plant defense response.

Of the 16 candidate genes that were identified, only three are located within the QTL region in the CDC Frontier reference genome (Figure 8). A local BLAST was performed to align each of

the candidate genes from the reference genome to the Dhera and JG 62 chromosome level assemblies. Based on the number and distribution of BLAST hits, the identified candidate genes are similar in sequence (>99.9% identity), copy number and location between resistant Dhera and susceptible JG 62 genomes. Although potentially functional differences in protein coding regions could be investigated using tools such as SIFT (Ng and Henikoff, 2003), it seems likely that other genes contribute to the recessive trait (see below). The QTL region is large (nearly 20 Mb) and contains almost 1,500 previously predicted genes which may increase as a result of the production of higher quality genomes. In any case, sequence prospecting will be more efficient once the causal region is further defined based on fine mapping efforts.

Since the previously identified QTL is recessive, a literature search was done to see if these candidate genes have been reported as recessive in other contexts. There has been one report of a dirigent protein in maize, which confers salt tolerance as a recessive gene (Wang et al. 2022). Similarly, there was one report of a gene responsible for *Ralstonia solanacearum* resistance in *Arabidopsis thaliana* which was characterized as a recessive *RRS1-R* gene, described as being a subtype of the class of TIR-NBS-LRRs (Deslandes et al. 2002). However, NBS-LRRs are typically reported as dominant genes, and dirigent proteins are involved in lignification, but generally not considered to be disease resistant genes. Thus, a search for more commonly reported recessive disease resistance genes was conducted.

Typically, recessive disease resistant loci are not comprised of a standard resistance gene (R gene), but rather a mutation in a susceptibility gene (van Esse et al. 2020). The mutation of this susceptibility gene results in the loss-of-function of a process on which the pathogen was

dependent on for successful infection of the plant. There are several well-known occurrences of these recessive, loss of susceptibility genes in action. Most notably the *Mildew resistance locus O* (*Mlo*) genes, which confer powdery mildew loss-of-function recessive resistance across a broad array of plant species (Kusch and Panstruga, 2017). While it may be possible that one of the identified classical disease resistance genes is responsible for Fusarium wilt resistance, it appears to be unlikely based on prior publications involving recessive resistance genes.

III. F2 populations

Genetic crosses were made between Dhera (R) and JG 62 (S). Sixteen F1 seed were germinated, and the plants grown to produce approximately 50 F2 seeds per plant, which will result in around 800 total seeds. These F2 seeds are recombinants of the F1 parents and are being used in conjunction the molecular markers to identify new informative recombinants within the Ca4 interval associated with Fusarium wilt resistance.

IV. Development of a SNP dataset and genotyping assays.

Single nucleotide polymorphisms (SNP) for genetic mapping were identified by aligning the Dhera genome sequence against JG 62. In total, 726,065 SNP were identified. As shown in Figure 7, SNP density varied discretely across the genome, including regions of non-alignment (white spaces in the plot) that are apparent deletions in the Dhera genome (see chromosomes 3, 4 and 6). Highest SNP density was on chromosome 4 between 15 and 40 Mbp, including a central non-aligned region that is the putative centromere implicated in the dot plot alignment (Figure 6).

SNPs were prioritized for molecular marker development based on local polymorphism rates. High ranking candidates were polymorphic at the focal SNP, but monomorphic between JG 62 and Dhera in the flanking 100 bp region. This filtering reduced the total number of candidate SNP to 6,630. All SNPs were submitted for primer design by IDT. Polymorphic SNP along with 100 bp of flanking DNA sequence were used for development of oligonucleotide PCR primers. SNP for genotyping were further filtered by selecting oligonucleotide pairs with balanced GC content, absent homopolymers and low predicted primer-dimer formation. In total, 52 primer pairs and fluorescent reagents were ordered to use as allele-specific oligonucleotide PCR markers (PCR Allele Competitive Extension assays, or PACE markers), nine of which are in the region of interest on chromosome 4 (Figure 9). Moving forward, the PACE assays designed in this project will be utilized in assessment of the F2 populations produced by this work. The results of this assay will determine recombination events which, if found, can be used for fine mapping the QTL region, and eventually narrowing down the causal gene. This information can then be deployed in breeding programs to develop new cultivars that are resistant to Fusarium wilt.

Methods

I. DNA Extraction, Sequencing, Assembly

Plant nuclei were prepared for each sample (Dhera, JG 62, and WR 315) according to the PacBio procedure for isolating nuclei from plant tissue using LN2 disruption (PacBio). This protocol was amended at several steps. The amount of beta-mercaptoethanol in the nuclei isolation buffer was increased from 0.25% to 0.8% to address issues of DNA quality. Additionally, two additional repetitions of the nuclei isolation buffer wash step were added (for five repetitions total) to improve sample purity.

High molecular weight DNA was extracted from prepared plant nuclei using a Nanobind plant nuclei kit (PacBio). The PacBio kit was used to try and improve sample quality for the HiFi sequencing. The samples were tested for purity with Nanodrop and for concentration with Qubit prior to library preparation. The resulting high molecular weight DNA was sent for sequencing on the PacBio Sequel II system for the Dhera and WR 315 genotypes and the Revio system for the JG 62 genotype at the UC Davis Genome Center. The Genome Center prepared SMRTbell libraries for each sample. Each genotype was assigned one SMRT cell for collecting HiFi read data.

After sequencing, quality control tests were run on the data. FastQC was used to give an overall snapshot of sequence quality, GC content, and possible contamination (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Jellyfish, a k-mer counting program, was used to create a histogram of k-mers in the sequencing data (Marcais and Kingsford 2011). This data was interpreted in GenomeScope, which predicted genome size and heterozygosity for each sample (Vurture et al 2017).

After quality control, each genome was assembled using hifiasm, an assembly program designed for resolving haploid genomes produced by HiFi sequencing (Cheng et al 2021). These assemblies were subsequently aligned to the reference genome, CDC Frontier, using minimap2 (Li 2018). The HiFi-specific preset “asm20” was applied during minimap2 alignment to CDC Frontier for both genotypes.

The Dhera and JG 62 genome assemblies were improved using Bionano scaffolding data from previous optical mapping. Each .paf file produced by the assemblies was converted to an .agp file for manipulation. The .agp file was manipulated to construct larger scaffolds from the contigs produced by hifiasm using agptools (https://warrenlab.github.io/agptools/) to keep or reject each contig. Based on dot plots created for each chromosome, contigs were either joined to create larger scaffolds or removed from the alignment.

After high quality chromosome-level assemblies were produced for both JG 62 and Dhera, the two genomes were aligned using minimap2 and plotted against each other and the reference using DotPlotly.

Once the cross-genome alignments were completed, sequences were compared using paftools and SNPs were cataloged. SNPs were prioritized for conversion to allele-specific oligonucleotide markers (PCR Allele Competitive Extension assays, 3CR Bioscience) based the absence of other SNP in the surrounding 100 bp region, and on considerations for PCR such as GC content, homopolymeric runs, and likelihood of primer-dimer formation.

II. Candidate Gene Evaluation

The reference chickpea genome, variety "CDC Frontier", was sequenced and assembled prior to initiation of this project (Varshney et al. 2013). This genome was annotated using a combination of evidence from homology to *Medicago truncatula*, transcript prediction using Stringtie and *ab inito* modeling using Maker. Functional predictions were developed using standard tools, including the InterPro and GeneOntology databases and vocabularies. These annotations are stored in a "data mine" on the Legume Information System (LIS) website

(https://mines.legumeinfo.org/cicermine/begin.do). Using this database, the previously identified QTL region on chromosome four was mined for candidate disease resistance genes using a simple search terms strategy, including “disease resistance”, “NB-ARC”, “NLR”, and defense-associated cues.

III. F_2 Populations, SNP selection, Testing for Recombinants

Chickpea genotypes Dhera and JG 62 were artificially hybridized using the protocol described by Kalve and Tadege (2017). Chickpea seeds were nicked using nail clippers and planted in three-gallon pots filled with potting soil (UC mix). Plants were grown for 5 weeks until first flowering, at which point crosses between genotypes began. JG 62 (purple flowers) was used as the female parent and Dhera (white flowers) as the male parent, since female parents with purple flowers are reportedly more receptive to pollen (Kalve and Tadege 2017). JG 62 flowers were carefully emasculated using forceps when the petals were the same length as the sepals. Dhera flowers were chosen for pollen when the flowers were half-open, while the pollen was orange and sticky. Collected pollen was manually transferred from the Dhera flower to the stigma of the JG 62 flower immediately following emasculation. Crosses were marked with dated tags to monitor for success. Each successful cross produced one to two F1 seed(s), which were harvested at maturity when the seed pod and peduncle had become completely yellowed and dry.

SNP markers were identified using paftools which produced a vcf file including all SNPs and indels. SNP markers that were produced from the genome sequence data were validated against parent genotypes Dhera and JG 62 to ensure they were polymorphic. The SNP assay was

designed in collaboration with Integrated DNA Technologies (IDT) to produce a PACE (PCR Allele Competitive Extension) SNP genotyping assay.

Conclusion and Further Directions

In summary, the results of this project further validated that the QTL region previously identified in Dhera is unique in comparison to other cultivated accessions, suggesting that it could be responsible for conferring Fusarium wilt resistance. This was shown both in the alignments between genomes and the further assessment of SNP density between Dhera and JG 62. Additionally, the candidate gene search produced several potential responsible genes that fall in the region of interest, and the improved annotation of these genomes could reveal more candidate genes present in this region in the future.

The markers designed in this project will be tested for allele discrimination using parental DNA. Informative SNP markers will be used both to validate the F1 as successful crosses and then to assess recombination within the candidate QTL region in the F2 seed.

If recombination is found, the F3 seed will be planted and genotyped using the same assay to determine regions of heterozygosity or homozygosity around the recombination event. Progeny with informative genotypes will then be allowed to set seed, which will be later collected and sent to Ethiopia. Once in Ethiopia, the F4 seed will be planted in sick plots to test for phenotypic traits indicative of Fusarium wilt resistance. Based on the phenotypic data and the fine mapping data, the QTL region will be narrowed down, enabling more focused gene searches to eventually identify the causal resistance gene and to nominate molecular markers for introgression breeding.

	Dhera Genome	JG 62 Genome	WR 315 Genome
FastQC Data (raw sequence)			
Total sequences	2,239,319	2,616,041	547,724
Sequence length	155-60491	221-63237	130-50216
% GC	32	32	32
Average length (estimated)	15000	18000	16000
Average Phred score (estimated)	92	39	92
QUAST Data (assembly)			
Number of contigs	577	415	606
Largest contig	47,523,781	120,552,8890	32,060,181
Total length	722,108,579	729,313,849	709,917,745
N50	19,381,490	45,318,072	12,759,852
L50	13	6	19

Table 1: Genome sequence quality statistics. FastQC is from raw sequence data, QUAST is from primary assemblies.

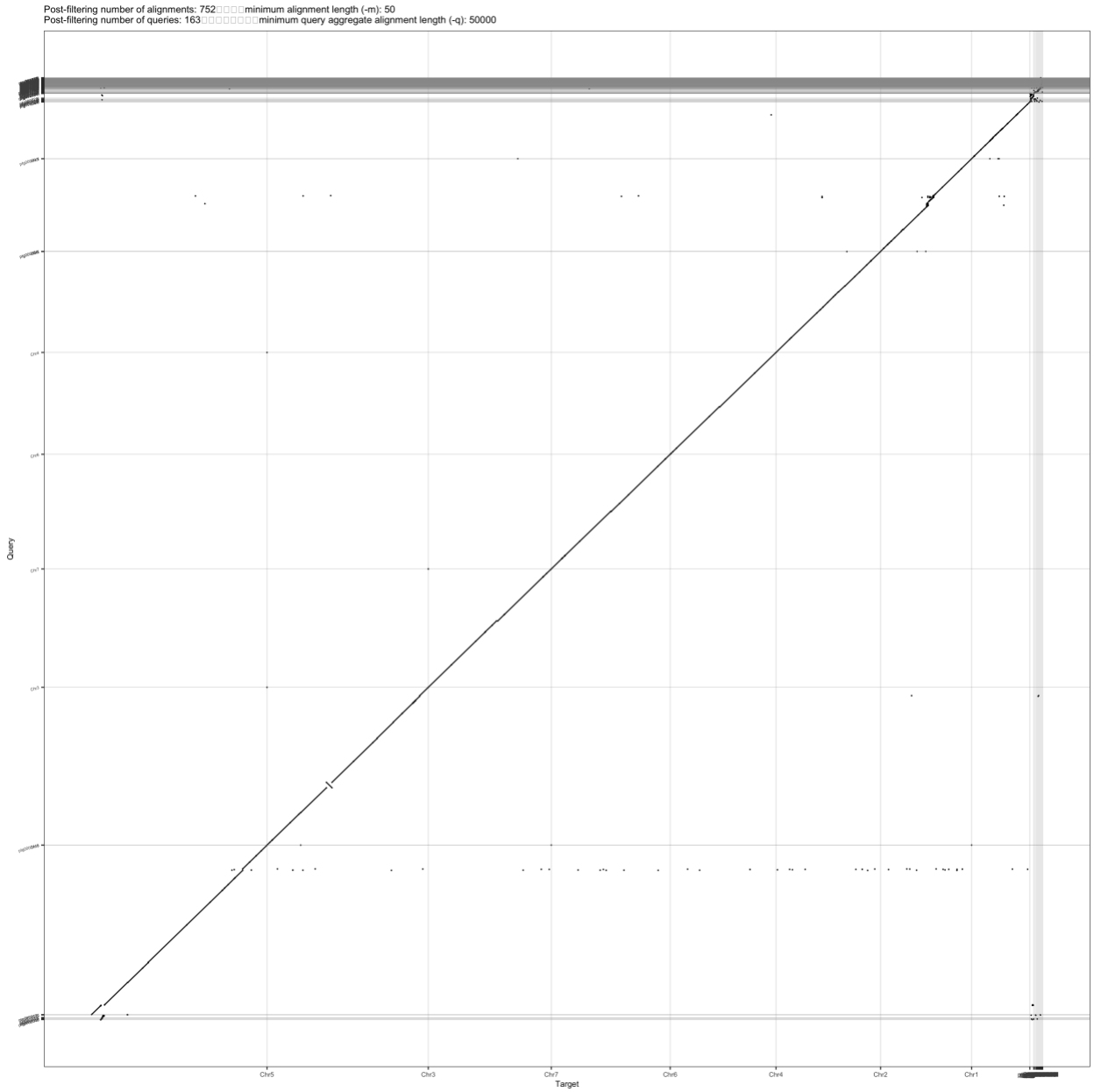

Figure 1: Dot plot, 96029 (x axis) and JG 62 chromosome level assembly (y axis)

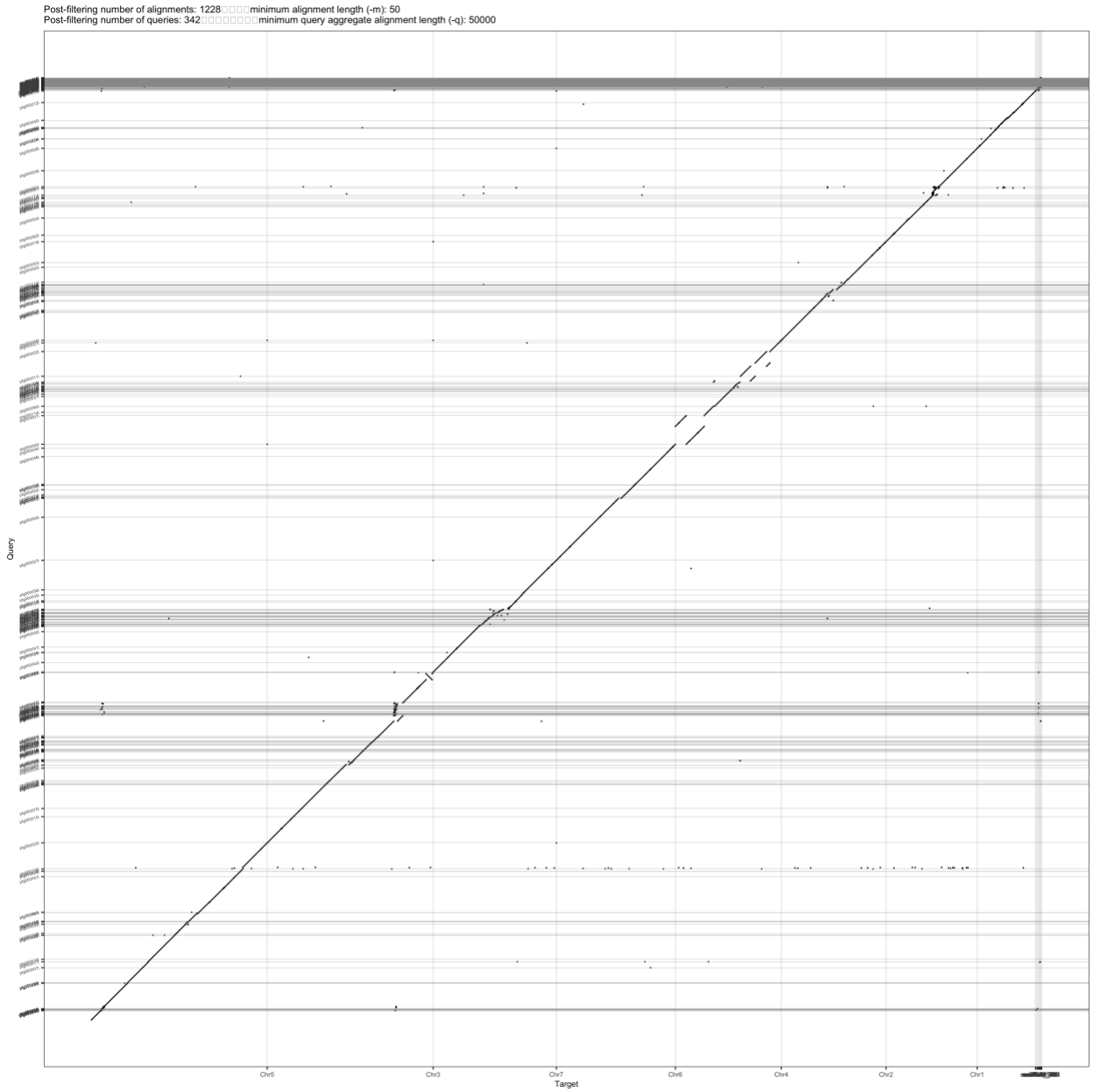

Figure 2: Dot plot, 96029 (x axis) and WR 315 primary assembly (y axis)

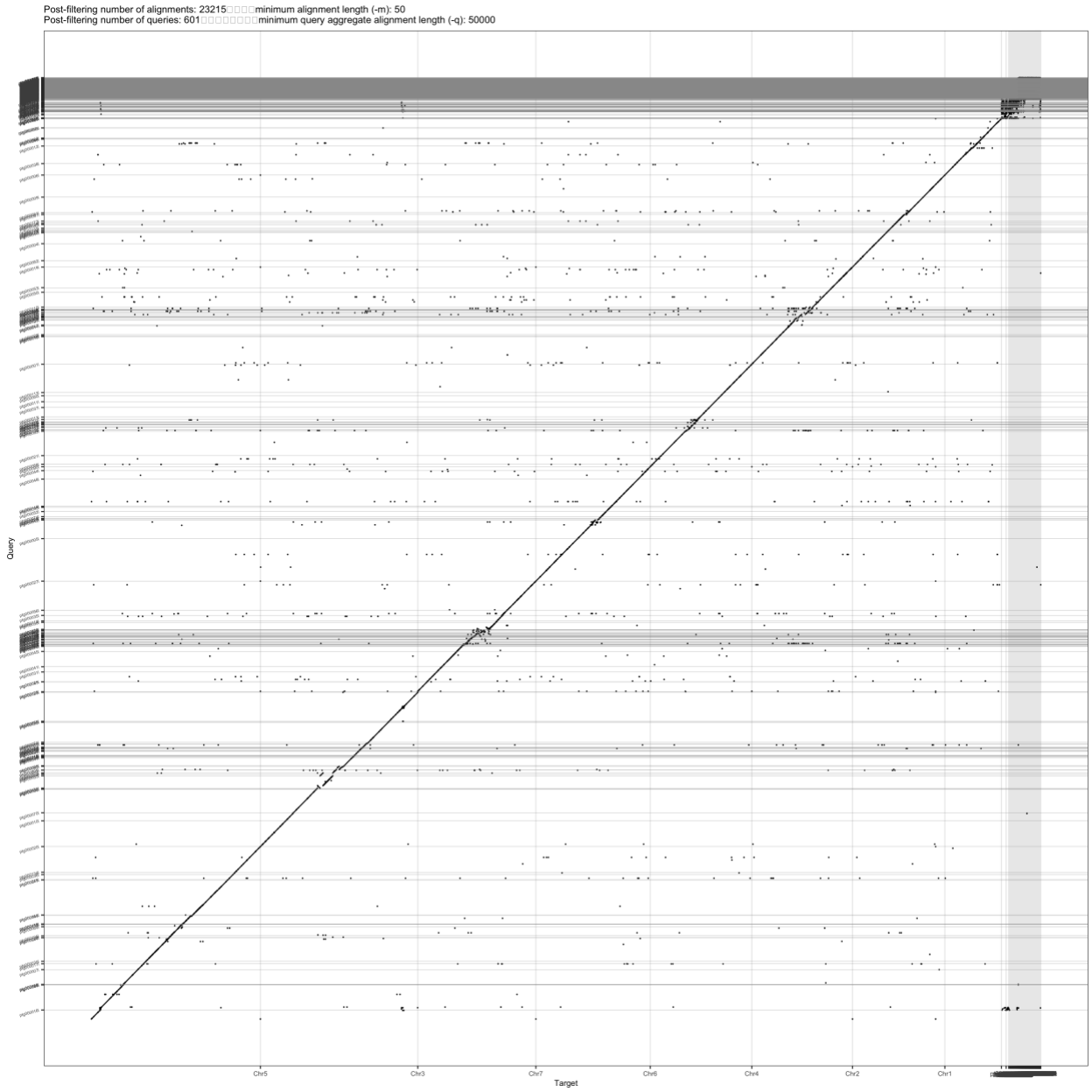

Figure 3: Dot plot, JG 62 chromosome level assembly(x axis) and WR 315 primary assembly (y axis)

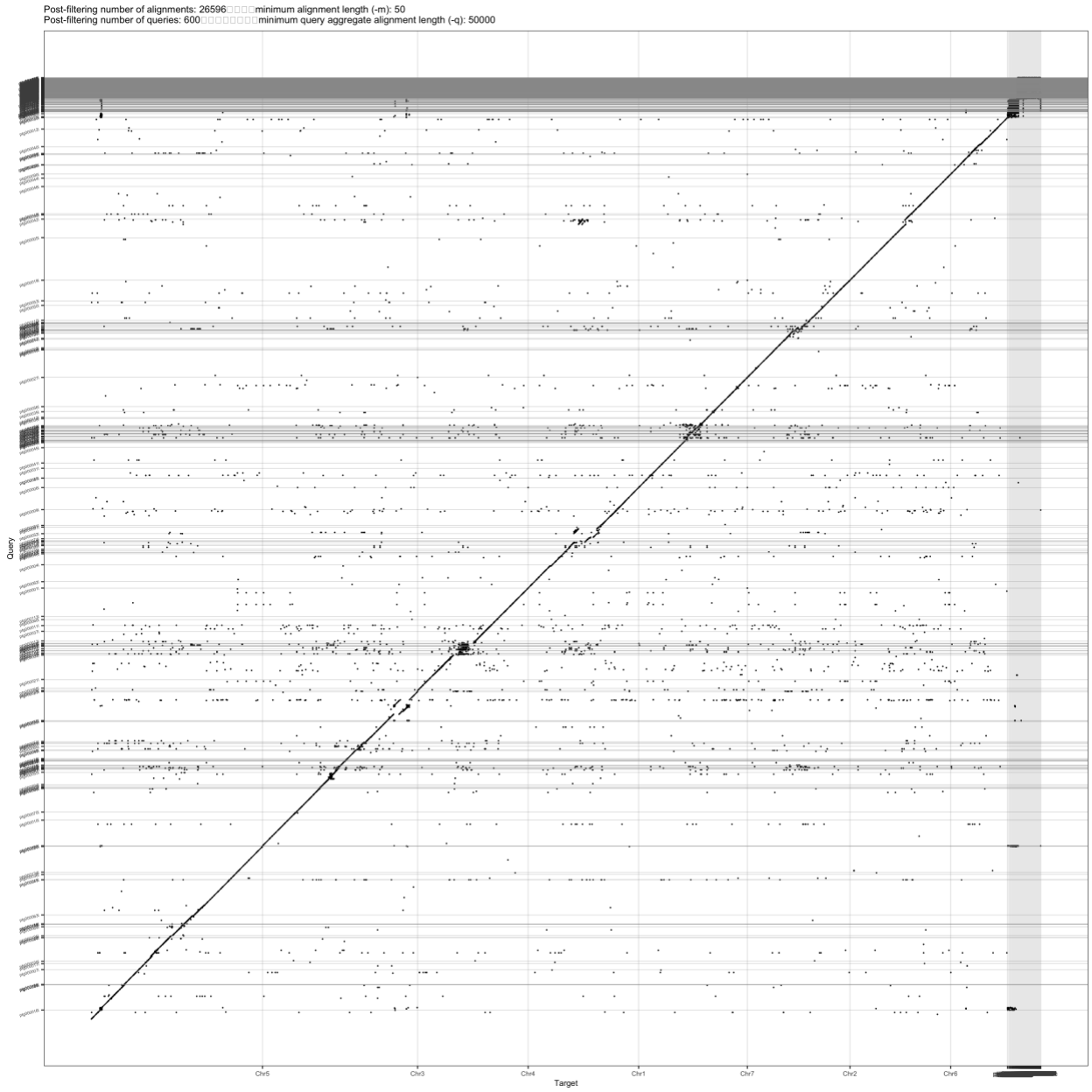

Figure 4: Dot plot, Dhera chromosome level assembly (x axis) and WR 315 primary assembly (y axis)

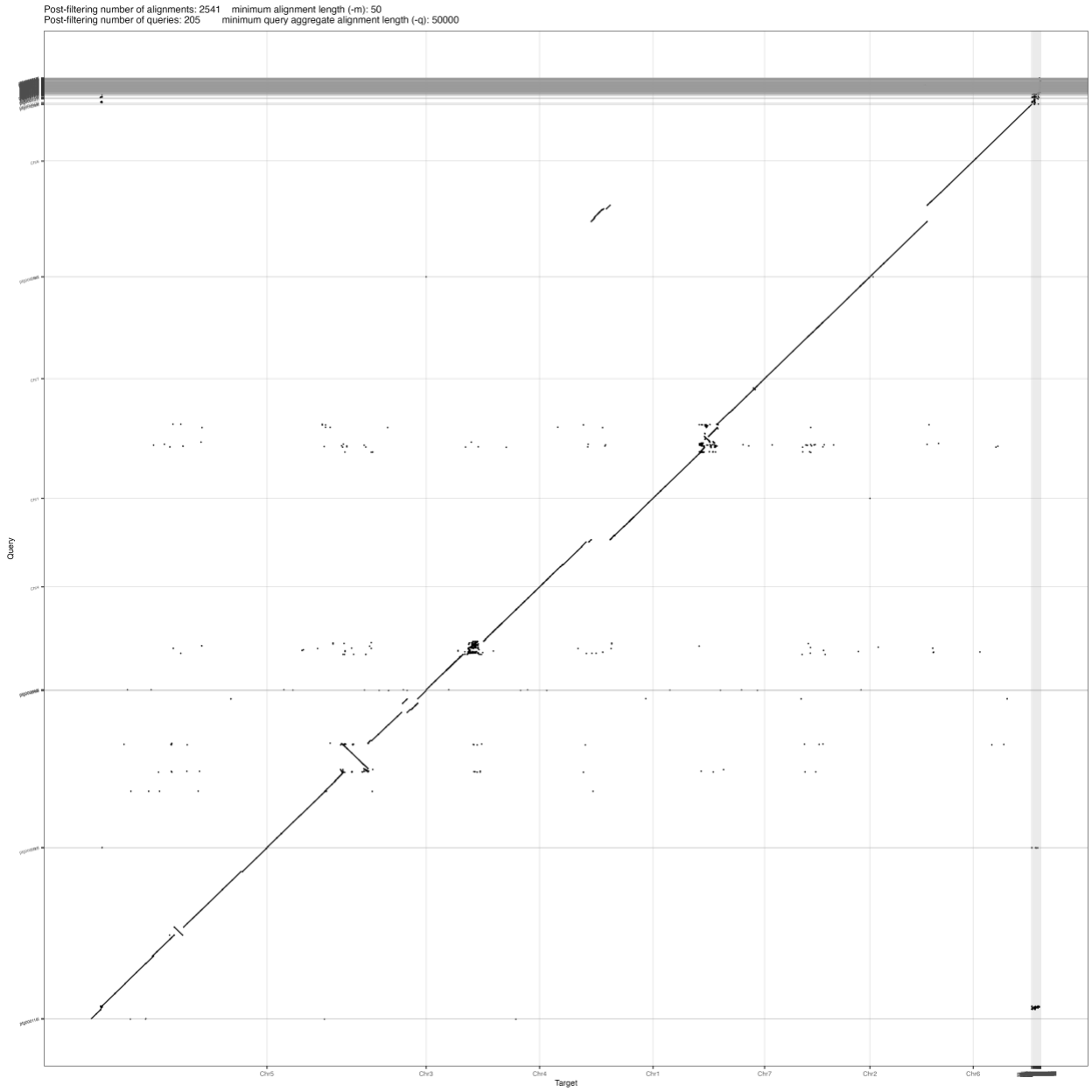

Figure 5: Dot plot, 96029 (x axis) and Dhera chromosome level assembly (y axis)

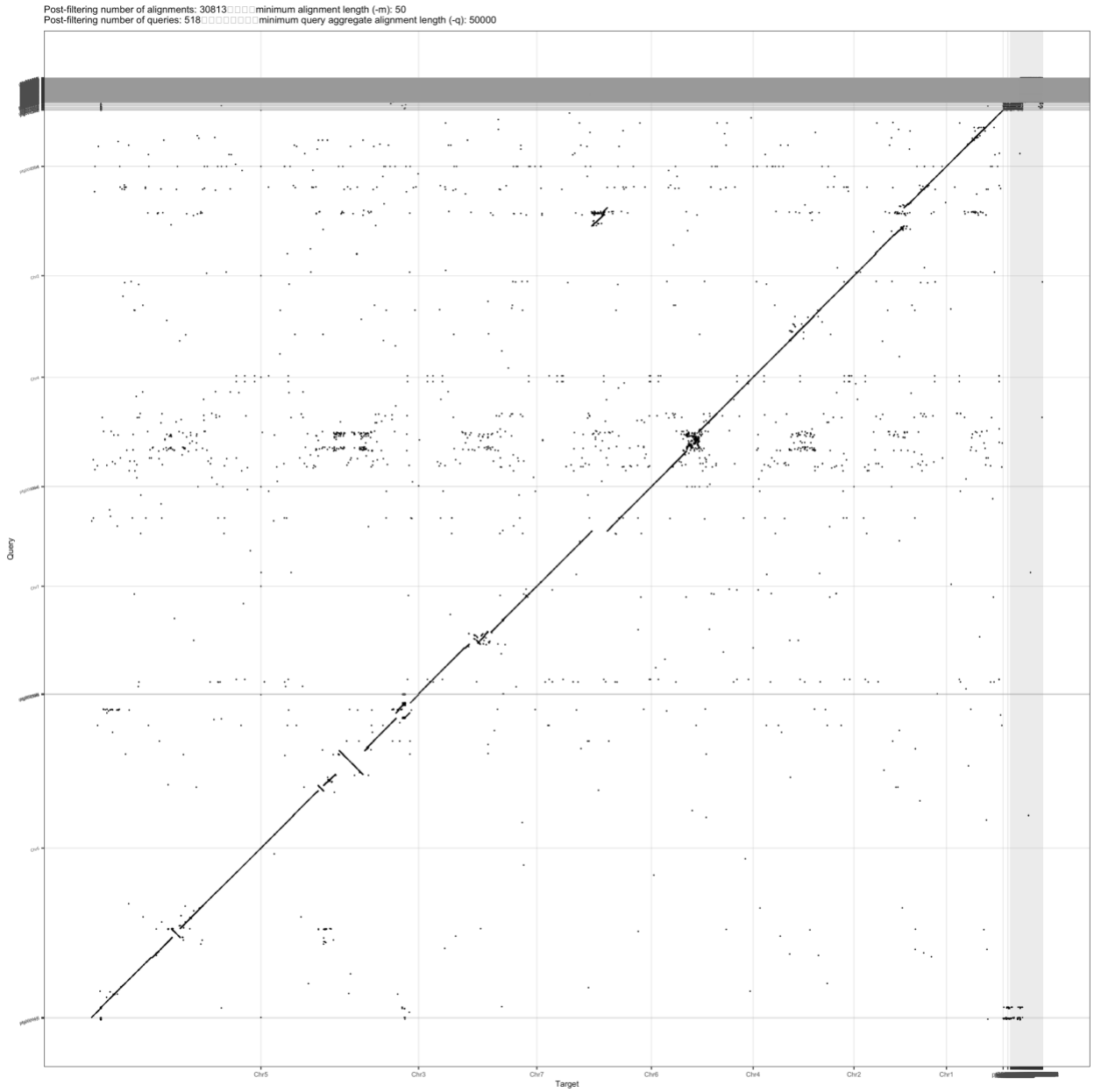

Figure 6: Dot plot, JG 62 chromosome level assembly (x axis) and Dhera chromosome level assembly (y axis)

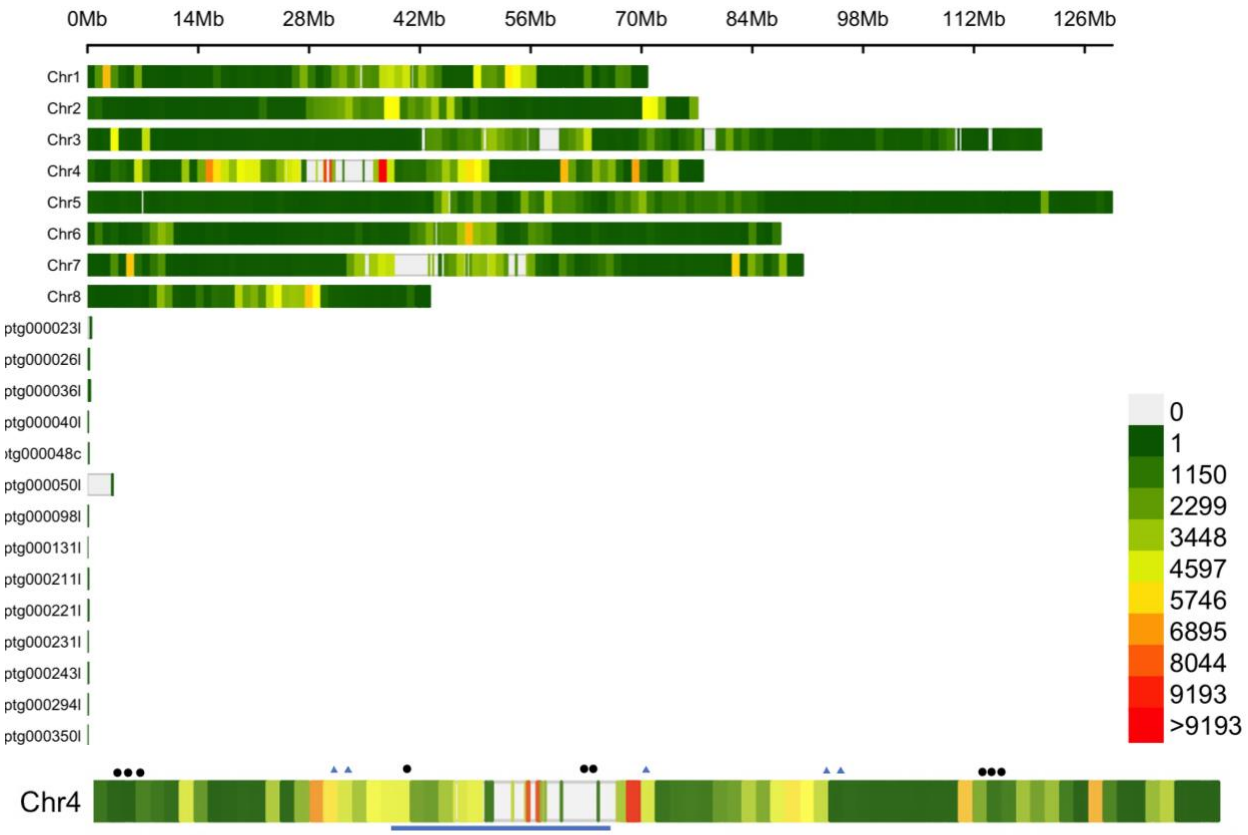

Figure 7: SNP density map, Dhera to JG 62. White regions (labeled "0" in the key) show regions lacking sufficient alignment data. Chromosome 4 diagram at the bottom of the figure shows in more detail where the QTL region is (underlined), the NBS-LRR type candidate proteins (black circles), and dirigent-like candidate proteins (blue triangles).

Gene Name	Description
Ca4g019800	disease resistance protein (TIR-NBS-LRR class), putative; IPR000767 (Disease resistance protein), IPR008808 (Powdery mildew resistance protein, RPW8 domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g019900	disease resistance protein (TIR-NBS-LRR class), putative; IPR000767 (Disease resistance protein), IPR001611 (Leucine-rich repeat), IPR008808 (Powdery mildew resistance protein, RPW8 domain), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0005515 (protein binding), GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g037400	disease resistance protein (NBS-LRR) [Glycine max]; IPR000767 (Disease resistance protein), IPR025875 (Leucine rich repeat 4), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g086900	ATP binding; GTP binding; nucleotide binding; nucleoside-triphosphatases; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0000166 (nucleotide binding), GO:0006952 (defense response), GO:0017111 (nucleoside-triphosphatase activity), GO:0043531 (ADP binding)
Ca4g137800	Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)
Ca4g137900	Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)
Ca4g142700	LRR and NB-ARC domain disease resistance protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g178800	disease resistance family protein / LRR family protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g179100	Disease resistance protein (TIR-NBS-LRR class) family; IPR000767 (Disease resistance protein), IPR001611 (Leucine-rich repeat), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0005515 (protein binding), GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g192200	Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)
Ca4g243300	Disease resistance-responsive (dirigent-like protein) family protein; IPR004265 (Plant disease resistance response protein)
Ca4g243400	Disease resistance-responsive (dirigent-like protein) family protein
Ca4g342800	disease resistance protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g343100	disease resistance protein; IPR000767 (Disease resistance protein), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0006952 (defense response), GO:0043531 (ADP binding)
Ca4g343300	disease resistance protein [Glycine max]; IPR002182 (NB-ARC), IPR027417 (P-loop containing nucleoside triphosphate hydrolase); GO:0043531 (ADP binding)

Table 2: Candidate genes and descriptions

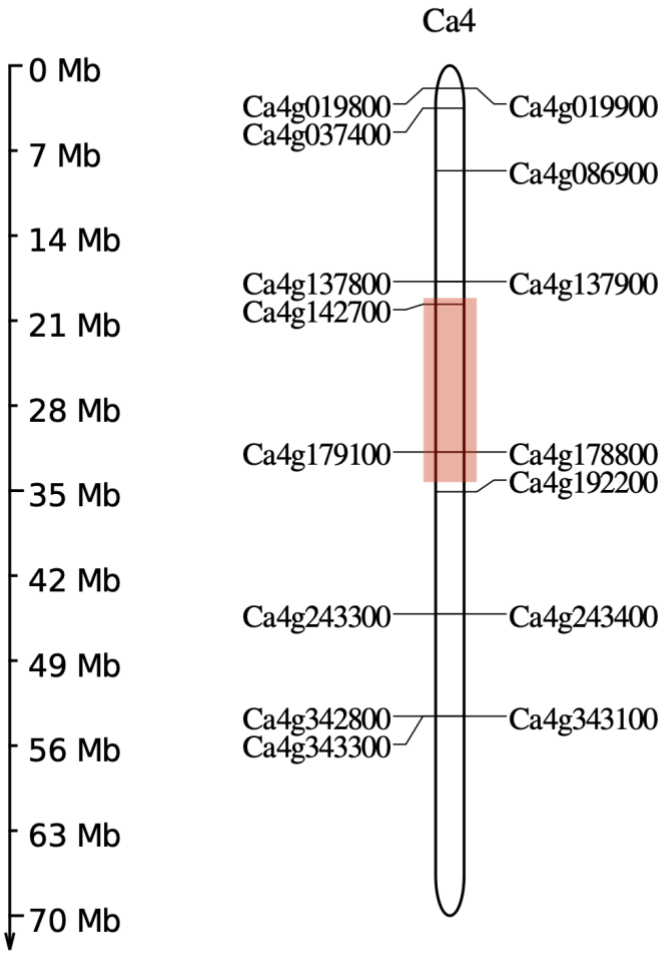

Figure 8: Candidate gene map based on CDC Frontier coordinates, QTL region is highlighted

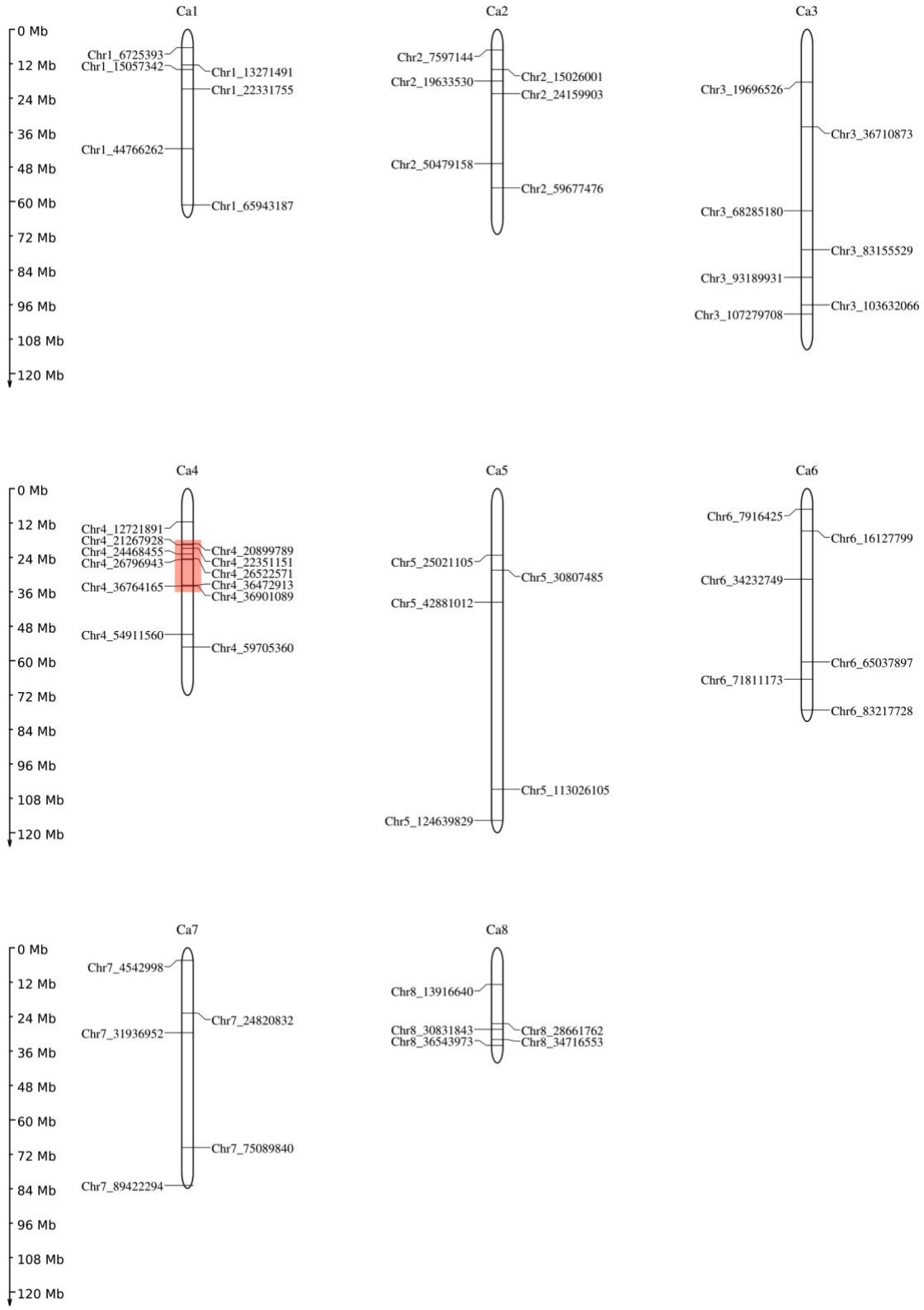

Figure 9: Selected SNP markers mapped to each chromosome, the QTL region is highlighted

Works Cited

Abbo, S., Berger, J., Turner, Neil C. Viewpoint: Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Functional Plant Biology* **30**, 1081-1087 (2003).

Cheng, H., Concepcion, G.T., Feng, X. *et al.* Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021). https://doi.org/10.1038/s41592-020-01056-5

Choo, K. H. Andy. Why Is the Centromere So Cold? *Genome Research* **8.2**, 81-82 (1998). https://doi.org/10.1101/gr.8.2.81

Das, A., Smoak, E.M., Linares-Saldana, R. *et al*. Centromere inheritance through the germline. *Chromosoma* **126**, 595–604 (2017). https://doi.org/10.1007/s00412-017-0640-y

Davey, J.W., Blaxter, M.L. RADSeq: next-generation population genetics. *Brief Funct Genomic* **9**, 416-423 (2010).

Deokar, A.A., Ramsay, L., Sharpe, A.G. *et al.* Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* **15**, 708 (2014). https://doi.org/10.1186/1471-2164-15-708

Deslandes, L., Olivier, J., Theulières, F. *et al.* Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *PNAS.* **99.** 2404-2409 (2002).

Fernandes, Joiselle B., Wlodzimierz, P., Henderson, Ian R. Meiotic recombination within plant centromeres. *Current Opinion in Plant Biology*. **48**, 26-35 (2019). https://doi.org/10.1016/j.pbi.2019.02.008

Garg, T., Mallikarjuna, B.P., Thudi, M. *et al.* Identification of QTLs for resistance to Fusarium wilt and Ascochyta blight in a recombinant inbred population of chickpea (*Cicer arietinum* L.). *Euphytica* **214**, 45 (2018). https://doi.org/10.1007/s10681-018-2125-3

Heng Li, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, Volume 34, Issue 18, September 2018, Pages 3094–3100, https://doi.org/10.1093/bioinformatics/bty191

Henikoff, S., Kami, A., Malik, H.S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293, 1098-1102 (2001).

Jiménez-Fernández, D., Landa, B., Kang, S. Quantitative and Microscopic Assessment of Compatible and Incompatible Interactions between Chickpea Cultivars and *Fusarium oxysporum* f. sp. *ciceris* Races. *PLOS ONE* 8(4): e61360 (2013). https://doi.org/10.1371/journal.pone.0061360

Kalve, S., Tadege, M. A comprehensive technique for artificial hybridization in Chickpea (*Cicer arietinum*). *Plant Methods* **13**, 52 (2017). https://doi.org/10.1186/s13007-017-0202-6

Kassie, Menale, et al. "Current situation and future outlooks of the chickpea sub-sector in Ethiopia." *ICRISAT and EIAR (http://www. icrisat. org/tropicallegumesII/pdfs/Current_Situation. pdf)* (2009).

Kazuhide Rikiishi, Manabu Sugimoto, Masahiko Maekawa, Transcriptomic analysis of developing seeds in a wheat (*Triticum aestivum* L.) mutant RSD32 with reduced seed dormancy, Breeding Science, 2021, Volume 71, Issue 2, Pages 155-166, https://doi.org/10.1270/jsbbs.20016, https://www.jstage.jst.go.jp/article/jsbbs/71/2/71_20016/_article/-char/en

Kelly, A., et al. "Use of genetic fingerprinting and random amplified polymorphic DNA to characterize pathotypes of *Fusarium oxysporum* f. sp. *ciceris* infecting chickpea." *Phytopathology* 84.11 (1994): 1293-1298.

Kusch, S., Panstruga, R. *mlo*-Based Resistance: An Apparently Universal "Weapon" to Defeat Powdery Mildew Disease, *Molecular Plant-Microbe Interactions,* **30**: 179-189 (2017). https://doi.org.10.1094/MPMI-12-16-0255-CR

Marçais, G., Kingsford, C., A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers, *Bioinformatics*, Volume 27, Issue 6, March 2011, Pages 764–770, https://doi.org/10.1093/bioinformatics/btr011

Mekonnen, Solomon A. Genetic Diversity of Ethiopian *Capsicum* spp and Molecular Mapping of the *Up* Gene in Pepper (*Capsicum spp.*). 2020. Seoul National University, PhD dissertation. http://publication.eiar.gov.et:8080/xmlui/bitstream/handle/123456789/3481/FinalThesis_afterDefense_SolomonBK.pdf?sequence=1&isAllowed=y

Millan, T., Clarke, H.J., Siddique, K.H.M. *et al.* Chickpea molecular breeding: New tools and concepts. *Euphytica* **147**, 81–103 (2006). https://doi.org/10.1007/s10681-006-4261-4

Ng, P., Henikoff, S., SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812-3814 (2003).

Pflug J.M., Holmes, V.R., Burrus, C., *et al*. Measuring Genome Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in Beetles (*Coleoptera*). G3 (Bethesda) 10(9):3047-3060 (2020). DOI: 10.1534/g3.120.401028

Singh G, Chen W, Rubiales D, Moore K, Sharma YR, Gan Y. 2007. Diseases and their management. In: Yadav SS, Redden R, Chen W, Sharma B, editors. Chickpea breeding and management. Wallingford: CAB International, p. 497–519.

Shehabu, M., Seid Ahmed & Parshotam K. Sakhuja (2008) Pathogenic variability in Ethiopian isolates of *Fusarium oxysporum* f. sp. *ciceris* and reaction of chickpea improved varieties to the isolates, International Journal of Pest Management, 54:2, 143-149, DOI: 10.1080/09670870701780153

Sunkad, G., Deepa, H., Shruthi, T.H. *et al.* Chickpea wilt: status, diagnostics and management. *Indian Phytopathology* **72**, 619–627 (2019). https://doi.org/10.1007/s42360-019-00154-5S

van Esse, H. P., Reuber, T. L., and van der Does, D. (2020), Genetic modification to improve disease resistance in crops. *New Phytol*, **225**: 70-86. https://doi.org/10.1111.nph/15967

Varshney, R., Song, C., Saxena, R. *et al.* Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol* **31**, 240–246 (2013). https://doi.org/10.1038/nbt.2491

Wang, Y., Cao, Y., Liang, X. *et al.* A dirigent family protein confers variation of Casparian strip thickness and salt tolerance in maize. *Nat Commun* **13**, 2222 (2022). https://doi.org/10.1038/s41467-022-29809-0