Government of Saskatchewan
Ministry of Agriculture

ADF Project Final Report Format

1. Project title and ADF file number.

ADF Project #20110155, Development of a germplasm resource to dissect complex traits in *Brassica napus*

2. Name of the Principal Investigator and contact information.

Isobel Parkin; Agriculture and Agri-Food Canada, Saskatoon RDC, 107 Science Place, Saskatoon, SK, S7N0X2, Canada; isobel.parkin@agr.gc.ca

3. Name of the collaborators and contact information.

Sally Vail; address as above; sally.vail@agr.gc.ca
Steve Robinson; address as above; steve.robinson@agr.gc.ca

4. Abstract/ Summary: *This must include project objectives, results, and conclusions for use in publications and in the Ministry database. Maximum of 300 words in lay language.*

There is a pressing need to develop improved canola varieties due to its importance to the Saskatchewan economy. However, this is hampered by the evolutionary history of canola, since modern varieties are derived from a relatively small subset of the genetic diversity found within available collections of *Brassica napus* germplasm. This is compounded by the narrowing of gene pools available to breeding programs as a result of intensive selection for specific quality traits. The project's main objective was to develop a tool that addressed this concern by developing germplasm (and associated resources), which not only capture the available diversity within the expanse of spring type *B. napus* but also provides a platform for dissecting and manipulating multiple complex traits. The project successfully did this by generating a nested association mapping (NAM) population, the population is composed of 2500 recombinant inbred lines derived from 50 diverse crosses. The founding parental lines were extensively phenotyped in the field and lab, highlighting wide-ranging variability for agronomically important traits. In addition, each of the parental lines and their progeny has been effectively assayed at hundreds of thousands of points across the genome, providing a detailed picture of their genetic composition. The whole population was assayed in the field in 2016 providing preliminary data for association of important traits with the underlying genes controlling their expression. This population can act as an effective link between the extensive genomics resources developed for Brassica species and the downstream breeding programs thus filling a gap often noted by producer groups and stakeholders in the wider industry. As such the population has been used to leverage funding for a number of aligned projects which are studying a range of traits, including protein accumulation and quality, drought and heat stress, root-soil interactions, seed dormancy and veriticillium resistance.

5. Introduction: *Brief project background and rationale.*

Modern canola varieties are derived from a relatively small subset of the genetic diversity found within available collections of *Brassica napus* germplasm (Snowdon and Friedt, 2004) and intensive selection for specific quality traits has further narrowed the genetic diversity within elite breeding lines. The project aimed to exploit diversity from across available collections of *B. napus* and provide a valuable resource for plant breeders to select new germplasm based on both phenotype and genotype.

Molecular research in crop improvement has tended to focus on important traits such as disease resistance that are controlled by a small number of genes and are amenable to linkage analysis approaches using simple segregating populations. However, many highly desirable agricultural traits that are targeted by breeders exhibit a complex inheritance and are often quantitative in nature being controlled by many quantitative trait loci (QTL). To understand such traits, powerful statistical methods have been developed that measure the phenotypic variance observed for the phenotype within a population. Identifying the genomic regions containing these QTL is the first step towards the effective improvement of these traits. This is hampered by the necessity to develop individual mapping populations for each trait of interest and is further compounded by the lack of knowledge as to the penetrance of individual QTL alleles in different genotypic backgrounds and the low frequency of recombination events in such populations that limits the resolution of individual QTL.

More recently, with the development of affordable high-throughput genomics technologies, the genotypes of larger natural populations are being analysed and important phenotypes are being dissected using statistical analyses derived from population genetics in extensive association mapping studies. One of the most advanced strategies being applied to understand complex quantitative traits involve a combination of both linkage and association mapping analyses as exemplified through the generation and use of Nested Association Mapping (NAM) populations (Yu et al, 2008). This powerful approach combines genomics technologies that provide high density genotypic information, the increased precision of association mapping resulting from higher numbers of recombination events, and the enhanced resolving power of linkage analysis. The NAM strategy utilizes recombinant inbred (RI) populations developed from multiple crosses between diverse inbred founder lines and a single reference, or common line. This approach generates multiple related RI lines that are mosaics of their parental lines and have the advantage of possessing homozygous alleles at each locus. The application of genome sequence analysis and molecular markers to the NAM population allows the genome composition of each RI line to be determined. This allows for the efficiencies inherent in the NAM experimental design to be realized. The characterization of the entire genomic DNA sequence of each inbred founder line and the reference line and the identification of individual recombinant DNA chromosomal regions among the RI progeny effectively captures the entire DNA sequence of the 2,500 population at a cost of sequencing only 50 individuals. Each chromosome segment of the RI lines is genetically characterized and by associating these regions with replicated phenotypic data the lines provide a powerful tool for not only identifying QTL but resolving them to small regions of the genome with tagged markers. The precise tagging of these valuable QTL with molecular markers immediately enables these traits to be manipulated using marker assisted breeding programs. Additionally, the high level of genetic precision in which the RI populations map each QTL significantly eases the identification of the underlying gene and its regulatory sequences.

The development of a NAM population for *B. napus* was a timely endeavor since during the course of the current project a number of advances were made which have further increased our ability to exploit the resource. As anticipated a number of genome sequences became available for *B. napus* (Chalhoub et al, 2014) which provide a reference for all the genotypic data generated during the project. A high-density (~60,000) single nucleotide polymorphism (SNP) array was generated for analyzing *B. napus* (Clarke et al, 2016) that has been used to assay all the RI lines. In addition, collaboration between the Global Institute for Food Security and the University of Saskatchewan has established a Plant Phenotyping and Imaging Research Centre (P2IRC) (https://www.cs.usask.ca/research/phenotyping-centre/) that is developing high-throughput methods for plant phenotyping (both lab and field based) that is using the NAM population as one experimental crop platform (see additional milestone).

6. Methodology: *Include approaches, experimental design, methodology, materials, sites, etc.*

Listed according to the Objectives:

Objective 1: **Identify and phenotypically characterize 50 genetically diverse *B. napus* founder lines**
To identify the founder lines we utilised both genotypic and phenotypic data. A 6000 SNP array was utilised to characterise all available germplasm. Phenotypic data was collated from all available sources to assist with founder lines selection. The founder lines have been phenoyped for multiple traits in both the lab and field, these have been described in the results section.

Objective 2: **Re-sequence the genomes of the 50 founder lines and the elite common line.**
The canola genome sequence of spring type DH12075 was available through the Canadian Canola Genome Sequencing Initiative (CanSeq) and was used as a reference to align sequences of the 50 founder lines and the common elite line. Each of the 51 lines were re-sequenced using high-throughput next generation Illumina sequencing technology based on established lab protocols to generate genomic shot-gun libraries and carry out the sequencing reactions.

Objective 3: **Generate 50 recombinant inbred (RI) *B. napus* populations – SKNAM population**
The NAM population was developed from crosses between each selected founder genotype and the elite common line.

Single seed descent for six generations (effectively 98% of the genotypic variation is fixed), was utilised to develop recombinant inbred lines (RILs). In total the NAM population encompassed greater than 2,500 individuals (at least 50 per cross with 25 each from reciprocal crosses). Genetic markers were employed to confirm the identity of the F1 individuals prior to RI development. The RI lines were generated under contract with an experienced service provider.

Objective 4: **Map the inheritance of genome segments in the NAM population by genotyping the 2,500 RI lines using a high density 60,000 SNP array.**
The 2,500 lines of the NAM population have been genotyped using the 60K SNP Illumina Brassica Infinium array (Clarke et al, 2016). DNA for the lines was extracted using a standard CTAB protocol and array hybridisation was carried out according to the manufacturer's instructions, the arrays were scanned on an Illumina HiScan at AAFC, Saskatoon.

Objective 5: **Multiple seeds of all NAM lines to provide to breeders and researchers**.
Multiplication of the whole population was completed in the greenhouse under standard control conditions. Additional seed is currently being generated in collaboration with the three industry partners.

7. Research accomplishments: *(Describe progress towards meeting objectives. Please use revised objectives if Ministry-approved revisions have been made to original objectives.)*

Objectives	Progress
1) Identify and phenotypically characterize 50 genetically diverse *B. napus* founder lines	Identification of founder lines - **COMPLETE**: Almost 400 spring type *B. napus* lines were genotyped with over 5000 SNP markers. These lines include germplasm from all world germplasm collections and germplasm developed through the breeding program at AAFC Saskatoon. This program has utilised crosses between winter and spring types to generate high vigour germplasm with spring like characteristics, developed germplasm with novel seed quality profiles and introgressed diverse genetic variation through the use of inter-specific crosses. In addition, a number of resynthesised lines have been utilised in population development that has generated germplasm with novel characteristics including improved water use efficiency. The SNP data and all available phenotypic data was collated and used to inform the final selection of the 50 founder lines. Early consultation with representatives from interested seed companies assisted with ensuring selection of germplasm that had characteristics of value for inclusion in downstream breeding efforts. The common line ACS N33 (N99-508) was selected from the AAFC breeding program (Quantum/LG3260), which is a homozygous elite line that is pre-adapted to Prairie conditions. The choice of this line will ensure that the NAM population can be readily assessed under Canadian field conditions, increasing the efficacy of the developed population. Phenotypic characterisation of founder lines - **COMPLETE**: Although the phenotyping of the founder lines is effectively complete, we continue to have requests to assess variation for additional traits, these new efforts will lead to further projects in the coming years. Field based studies: The founder lines and common line were phenotyped for four field seasons under Saskatchewan conditions. In 2015 and 2016, the lines were grown in additional locations in Saskatchewan (Outlook, Scott) and Alberta

	(Beaverlodge). At Outlook the lines were grown under irrigated and non-irrigated conditions to look for variation in water use efficiency. A number of parameters have been assessed in the field and on field grown seed, including plant emergence, biomass accumulation prior to flowering (NDVI), days to flower, days to maturity, lodging index, height at maturity, general agronomic notes, NIR analyses (seed oil content, protein content and fibre), fatty acid and glucosinolate profiles. We can provide a detailed breakdown of each of these traits on request. A demonstration field trial plot of the founder lines was established in 2015 and was visited as part of the field tour during the International Rapeseed Congress which was held in Saskatoon and further additional tours that year for colleagues and collaborators (**Figure 1**). Example of lab based studies: In 2015, a substantial greenhouse assessment of the lines was undertaken, with over 30 individual parameters being measured for each of the lines (example figure from analyses of these data is shown in **Figure 2**). In 2016, the founder lines were assessed by a collaborator in Europe and found to house substantial variation for resistance to Verticillium wilt (**Figure 3**).
2) Re-sequence the genomes of the 50 founder lines and the elite common line.	**COMPLETE.** Illumina genomic DNA sequencing libraries were generated for each of the founder lines and the common parent. The libraries were sequenced to a depth of approximately 10 genome equivalents. The reads were cleaned and trimmed prior to alignment to the reference spring type genome sequence (DH12075) with a 74-79% mapping efficiency (only unique matches considered). These data will be used to impute the sequence of each RI line after SNP genotyping; however, in the first instance we utilised the data to assess the exact genotypic composition of each founder line. Sequence variant discovery was carried out using the UnifiedGenotyper in GATK, ~17 million raw sequence variations were identified across all lines. After filtering 4,300,272 SNP loci were called across the genome with fairly equivalent numbers of SNPs in both the A and C genomes. These data have characterised the inheritance of each region of the genome and indicate which lines share which regions of the genome. In addition these data identified a number of chromosomal rearrangements in each of the *B. napus* lines (**Figure 4**). Such rearrangements could prove to be highly significant since they have been found to be correlated with major adaptive changes in *B. napus*, for example the low glucosinolate phenotype in canola is the result of one such change.
3) Generate 50 recombinant inbred (RI) *B. napus* populations – SKNAM population	**COMPLETE.** Initial crosses were made at AAFC and the resultant F1 progeny were confirmed through marker analyses to represent the correct combination of parental lines. F2 seed was provided to Haplotech (Winnipeg, MB) who carried out single seed descent on 25 lines from reciprocal F1s for each of the 50 crosses. F6 seed was generated for 2800 lines, three populations had less than 50 lines in total due to either infertility or extended generation times. However, many populations had greater than 50 lines available.
4) Map the inheritance of genome segments in the NAM population by genotyping the 2,500 RI lines using a high density 60,000 SNP array.	**COMPLETE.** DNA was generated from all 2800 available RI lines and each of the lines has been hybridised to the Brassica 60K Illumina Infinium array. This work was only completed in February of 2017, so analyses of the resultant data is still progressing and it is anticipated that the full maps of all the lines will be completed by the end of the summer in 2017.

5) Multiple seeds of all NAM lines to provide to breeders and researchers	**COMPLETE.** Seed of all available RI lines was multiplied in the winter of 2015-16. There was sufficient seed to provide a small sample to each of the industry partners in the project and also to complete a preliminary field assessment in 2016, see below. Further seed increases will be required due to the demands on the population - the industry partners and AAFC have agreed to share this burden in the first instance. However, this is unlikely to be sufficient to meet the demands on the population.
6) Additional Work – Field assessment of NAM population in 2016	In the growing season of 2016, 2300 of the NAM RI Lines (RIL) were grown and assessed in a field nursery near Saskatoon, SK. The trial was set up in a Type II Modified Augmented Design (MAD) where the primary or reoccurring check was the Reference Line (RL), NAM-0. When analyzed, the values from the plots of the primary check will be used to adjust means for spatial variability across the field for the adjacent un-replicated NAM RIL plots. Fourteen NAM Founder Lines (FL) were selected as secondary checks were grown with ten replicated plots across the trial to generate error values for testing differences between un-replicated means of lines. The trial consisted of about 3200 plots that were a single row each, 3m long and spaced 60cm apart. Out of the 50 RIL sub-populations (ie. populations generated from crosses with each of the FL), 40 had at least 45 up to 69 RILs grown in the field in 2016. The remaining 10 sub-populations were slightly less represented with as low as 15 RILs due to either infertility and/or lateness in maturity which prevented sufficient seed for field assessment in 2016. In-field data collected during the growing season on the plots included stand/emergence counts, NDVI (at 4-6 leaf stage and again pre-bolting), days to flowering, plant height, canopy depth, lodging ratings, days to maturity and pre-harvest pod shattering. In addition, the trial was imaged four times throughout the growing season using a UAV equipped with remote sensing equipment with the eventual goal of extracting vegetative indices from the images. This was done in collaboration with the P2IRC (**Table 1**). The unexpected snowfall incurred the first week of October in 2016 unfortunately jeopardized obtaining estimates of seed yield from the plots as only about 15% of the plots had been harvested with a combine prior to this event. The remainder of the plots were harvested by hand and threshed prior to submitting seed for quality analyses. Currently, the seed is being analyzed for contents of oil, protein, fiber fractions, glucosinolate and fatty acid profiles. Upon completion of seed quality analyses, a MAD data analysis pipeline will be employed to obtain adjusted means for further analysis (http://probes.pw.usda.gov/bioinformatics_tools/MADPipeline/index.html; Australian Journal of Crop Science, 7:1789-1800). **Preliminary Results:** When compared to the diverse NAM FLs and the highly adapted RL, the NAM RILs showed extreme values for flowering, maturity and height variables; however, most of the RILs were less extreme in range compared to the FLs as shown by the concentration of values within the first and third quartiles (**Figure 5**). Furthermore, median and mean values of the combined RIL population were between combined FL and RL values for flowering, maturity and height. Interestingly, the median and mean values for the RILs were substantially lower than the FLs and slightly lower than the plots of the RL. Although the range of values and means for lodging indicated some lines lodged much more than commercial checks, the lowest values

	within the RILs were comparable to those of the relatively un-lodged commercial checks. Results on the raw values demonstrate the utility of the NAM RIL population for further analysis and use as breeding lines. In order to further study the genetics of complex agronomic traits, moderation in adaptation traits, such as flowering and maturity, is critical for field-based assessment. Furthermore, the results to date bode well for accessing diversity to recapture novel traits in adapted breeding lines.

add additional lines as required

8. Discussion: *Provide discussion necessary to the full understanding of the results. Where applicable, results should be discussed in the context of existing knowledge and relevant literature. Detail any major concerns or project setbacks.*

The project goals were achieved with the development of a structured population of *B. napus* that can be used for the study of complex traits in both the field and lab environment. We are still in the process of analysing the massive amount of genotypic data that we have created but the use of the SNP array has simplified the processing of this data, which should ensure rapid completion of this task. This will allow us to begin to identify genomic regions controlling traits of interest.

There were no major setbacks in the development of the population, we have a small number of sub-populations that could be expanded (those with lower fertility and/or late maturing); however, the value of these for studying field-based traits is questionable. We have some concerns over the continued maintenance and distribution of the lines, since this is a labour and space intensive, time consuming and a relatively expensive commitment. In the short term we have negotiated a seed increase of all the lines with our industry partners and we will encourage the inclusion of funds for population maintenance in any project that requests the lines. Our only other concern is how we store the data generated and provide adequate access to the community, we are in discussions to develop a database with a web enabled front end to allow such access. Although we currently have limited resources for this work, we hope to leverage support from the P2IRC project, which is intending to use the NAM population in its work to apply high throughput imaging to crop plants.

9. Conclusions and Recommendations: *Highlight significant conclusions based on the previous sections, with emphasis on the project objectives specified above. Provide recommendations for the application and adoption of the project.*

- The developed NAM population is an extremely valuable resource for studying multiple traits relevant to sustainable canola production.
- The NAM population has already been used as a foundation to apply for a number of successful research projects (drought and heat tolerance, seed dormancy, hybrid vigour) and additional project submissions are pending.
- It is recommended that a searchable database is developed that will allow access to the generated data since this would be a valuable mechanism for further disseminating information about this resource.

10. Success stories/ practical implications for producers or industry: *Identify new innovations and /or technologies developed through this project; and elaborate on how they might impact the producers /industry.*

The NAM population has been the basis of successful applications to ADF, GFII and the AAFC Canadian Crop Genomics Initiative. Each of these projects has targeted the identification of genomic regions underlying important traits for continued canola development. These projects are in the process of defining germplasm that contain valuable alleles for improving a number of traits and at the same time identifying molecular markers that can be used to rapidly introgress these regions into current commercial lines. It would be anticipated that improvements determined by such projects would be realised in the next 5-10 years in the form of improved commercial varieties.

In addition, although hard to evaluate the project was funded by three industry partners who have been selecting material from the germplasm throughout, this material will be much more advanced and likely to have a much earlier impact at the production level.

11. Patents/ IP generated/ commercialized products: *List any products developed from this research.*

The germplasm - NAM founders and NAM population.

12. List technology transfer activities: *Include presentations to conferences, producer groups or articles published in science journals or other magazines.*

The NAM founder lines have been transferred under MTA to collaborators to initiate project work such as the Verticillium wilt screening (U of Goettingen, Germany) and protein analyses (U of Manitoba).

The NAM population was described in presentations given by Dr. Isobel Parkin (examples below).

The NAM population was presented at the 14th International Rapeseed Congress (July 5-7th, 2015) held in Saskatoon in the form of a poster (see below) and also during the field tours. The founder lines of the NAM population were grown in demonstration plots and Dr. Sally Vail described the project objectives and goals to national and international visitors (**Figure 1**).

The NAM population was presented at the AAFC booth at the Crop Production Show in Saskatoon, January 11-17th, 2016.

Presentations:
Parkin, I (2014) Development of a germplasm resource to dissect complex traits in *Brassica napus.* Genomic Selection Workshop, Eckernförde, Germany, Nov 18-19th.

Parkin, I (2015) Impact of Genomics on Brassica Genetics and Breeding- a sequence level view of U's triangle. International Rapeseed Congress, Saskatoon, July 7th.

Parkin, I (2016) Application of Genomics technologies to Canola crop improvement, Canola Council Discovery Forum, Winnipeg, October 27th..

Poster:
A.S.K. Shunmugam, R. Soolanayakanahally, L. Yang, K. Horner, E. Higgins, M. Lewis, **I. Parkin, S. Vail, S. Robinson** Characterizing growth physiology, flowering phenology, yield components and seed quality attributes in founder lines of a spring *Brassica napus* Nested Association Mapping population. 14th International Rapeseed Congress, July 5-7th, 2015, Saskatoon.

13. List any industry contributions or support received.

The three provincial Prairie Canola Producer organisations provided funding to match the contributions of ADF.

Three industry partners each contributed $100K (total $300K) during the course of the project. The industry partners are also contributing to the seed increase of the whole population.

14. Is there a need to conduct follow up research? *Detail any further research, development and/or communication needs arising from this project.*

To fully exploit the population and fulfill current demand it would be highly desired to have a commercial multiplication of the lines.

Additional research funds were applied for and received by Dr. Raju Soolanayakanahally from SK ADF to assess the NAM founder lines for variability in drought and heat stress. This could lead to multiple avenues of future research.

Additional phenotype data needs to be acquired for the entire population to allow robust analyses of agronomic traits, the preliminary data generated in 2016 will potentially assist in narrowing down populations that could be identified to focus analyses of particular traits; however, data from the whole population would be extremely powerful for determining genetic factors controlling yield component traits. There is much discussion about exploiting the population to study nutrient use efficiency in the crop, which would be a valuable area for further investigation.

As mentioned above funds to develop an integrated, expandable and easily accessed data portal for the NAM lines and the associated data is a priority.

A number of individual projects targeting particular traits are being discussed, for example the variation observed for Verticillium wilt suggests that the NAM project could be an excellent platform for studying this potential new threat for Saskatchewan canola production. There is also interest in screening the population for resistance to swede midge, another new problem that is beginning to emerge for canola production.

15. Acknowledgements. *Include actions taken to acknowledge support by the Ministry of Agriculture and the Canada-Saskatchewan Growing Forward 2 bilateral agreement.*

The support provided by the Ministry of Agriculture and the other funders was gratefully acknowledged in all presentations made describing the project, in addition they will be acknowledged in multiple publications that are currently being drafted.

16. Appendices: *Include any additional materials supporting the previous sections, e.g. detailed data tables, maps, graphs, specifications, literature cited*

Figure 1: Dr. Sally Vail describing the NAM population project during field tours organised in 2015.

Figure 2: Scatter plot matrix of traits estimated in greenhouse 2015 (a) leaf characteristics (CCI-chlorophyll content index, F_v/F_m-fluorescence) (b) gas exchange (A-photosynthetic rate, g_s-stomatal conductance, E-transpiration rate, C_i/C_a-ratio of intercellular to ambient CO_2 concentration) and (c) flowering phenology (DTB-days to bolting, DTF-days to flowering). Linear regression fit and correlation coefficient of traits are presented with diagonal histograms representing phenotypic variance for each trait.

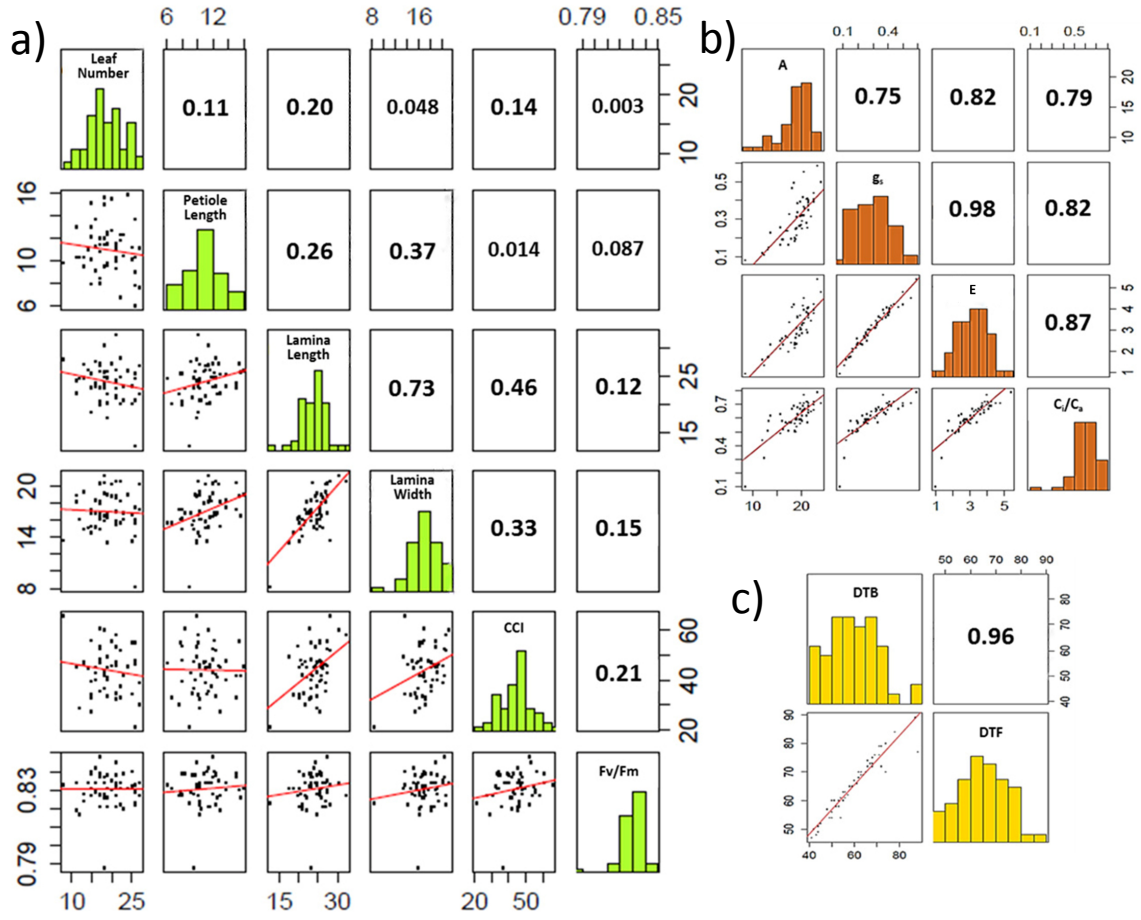


Figure 3: Screening of NAM founders (AUDPC - area under the defense response curve) for resistance to Verticillium wilt, Falcon and Express are the susceptible and resistant checks, respectively.

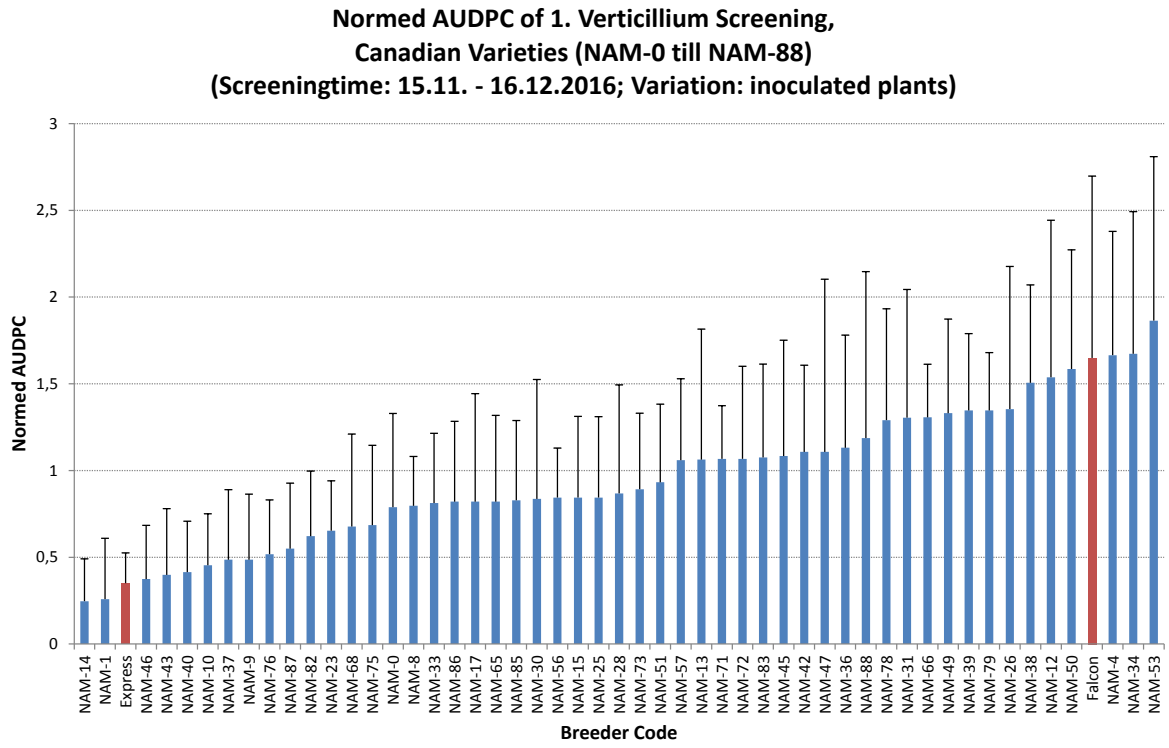


Figure 4: Genome rearrangements in the *B. napus* lines that result from homoeologous recombination events between the A and C genome could be identified through mapping of the sequence data to the reference genome. In the upper panel the duplication of almost half a chromosome is shown by the increase in read depth, while the corresponding loss of the homoeologous region is shown in the panel below by the reduced read depth.

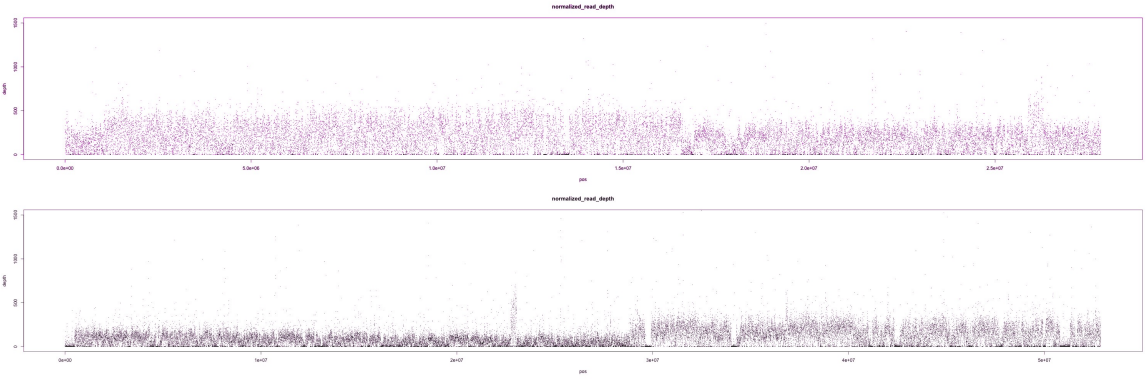
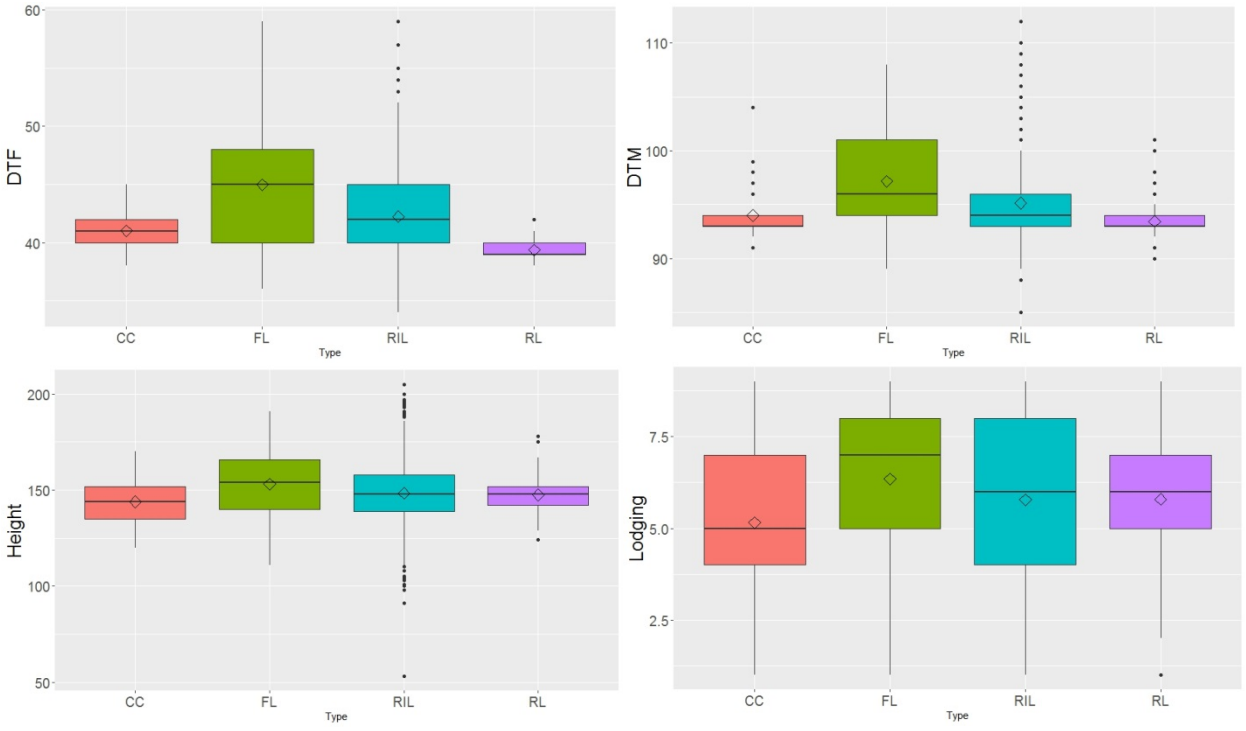


Table 1: Aerial imaging on the Spring *Brassica napus* NAM RIL nursery in 2016*:

Date of Flight	Camera used for Imaging	Height of Flight	Resolution
13/06/2016	RGB	20m	5.5 mm
24/06/2016	Multispectral	20m	13.6 mm
09/07/2016	Multispectral	20m	13.6 mm
27/07/2016	Multispectral	20m	13.6 mm

***Compliments of Sudhakar Duddu and Steve Shirtliffe**

Figure 5: Box-plot demonstrating range in raw values for plots for Days to Flower (DTF), Days to Maturity (DTM), measured Height and visual Lodging scores for Commercial Checks (CC), Founder Lines (FL), NAM Recombinant Inbred Lines (RIL) and the Reference Line (RL).



Notes:
Boxes indicates the range from the 1st to 3rd quartile of values for the populations, whiskers indicate the approximate 95% confidence interval, median indicated by mid-box horizontal line and ◊ indicates the mean. Outlying values are indicated with '●'.
Height was measured in centimeters and Lodging Scores of 1=upright where 9=completely lodged.