Disjunction between canola distribution and the genetic structure of its recently described pest, the canola flower midge (Contarinia brassicola)

Erin Campbell[1], Julian Dupuis[2], Jennifer Holowachuk[3], Shane Hladun[3], Meghan Vankosky[3], and Boyd Mori[1]

[1]University of Alberta
[2]University of Kentucky
[3]Agriculture and Agri-Food Canada Saskatoon

June 3, 2020

Abstract

Population genomics is a useful tool in the integrated pest management toolbox for elucidating population dynamics, demography, and histories of invasion. However, next-generation sequencing approaches can be hampered by low DNA input from small organisms, such as insect pests. Here, we use a restriction-site associated DNA sequencing approach combined with whole-genome amplification to assess genomic population structure of a newly described pest of canola, the diminutive canola flower midge, Contarinia brassicola. We find that whole-genome amplification prior to library preparation caused a reduction in the overall number of loci sequenced and an increase in overall sequencing depth but had no discernable impact on genotyping consistency for population genetic analysis. Clustering analyses recovered little geographic structure across the main canola production region, but differentiated several geographically disparate populations at edges of the agricultural zone. Given a lack of alternative hypotheses for this pattern, we suggest these data support alternative hosts for this species and thus our canola-centric view of this midge as a pest has limited our understanding of its biology. These results speak to the need for increased surveying effort across multiple habitats and other potential hosts within Brassicaceae, to elucidate both our ecological and evolutionary knowledge of this species as well as potential management implications.

INTRODUCTION

The canola flower midge (CFM), *Contarinia brassicola* Sinclair, is a newly discovered cecidomyiid fly from the Canadian prairies that forms flower galls on canola, *Brassica napus* L. (Mori *et al.*2019). Canola was initially developed from rapeseed, *Brassica rapa* L. and *B. napus* in the Canadian provinces of Manitoba and Saskatchewan in the 1970s, and has since increased to become one of the largest oilseed crops in the world due to widespread use as livestock feed, biofuel, and cooking oil (Barthet 2016; Canola Council of Canada 2020a). Today, the Canadian Prairies produce and export the largest amount of canola in the world, and the highest levels of Canadian production occur in Saskatchewan (LMC International 2016; Statistics Canada 2019).

CFM is hypothesized to be native to Canada (Mori *et al.* 2019), although knowledge of its biology is limited by the short history of its taxonomic existence. Prior to its description in 2019, the canola midge pests of the Prairie provinces were erroneously thought to be the swede midge, *Contarinia nasturtii* (Kieffer), a morphologically and ecologically similar congener of CFM. Swede midge causes significant crop damage in parts of Europe, Asia, and more recently, as an invasive pest of canola in North America (Hallett *et al.* 2007;

Chen*et al.* 2011). Like swede midge, CFM appears to be multivoltine. Initial adult emergence occurs in June and July, during canola bud formation, with a second generation in August; however, CFM larvae have been observed in the field throughout the summer and into September, suggesting that they may produce more than two generations per year (Chen *et al.* 2011; Andreassen *et al* . 2018; Mori *et al.* 2019; Soroka *et al* . 2019). Larvae are small, up to a few millimeters in length, and they feed within developing canola flower buds. This causes the buds to transform into galls, which then fail to flower or produce seed (Mori *et al.* 2019). Due to their feeding behaviour and ability to produce multiple generations per year, CFM is potentially capable of causing significant impact on Canadian canola crop yields.

While several aspects of CFM ecology have been described (Mori *et al.* 2019; Soroka *et al* . 2019), little is known about CFM population dynamics. Prior genetic investigation of CFM was restricted to specimens sampled primarily from Saskatchewan and use of only a single mitochondrial gene (Mori *et al.* 2019). There have been no assessments of CFM population structure at wider geographic scales, thus limiting effective monitoring and risk assessment across the canola producing region. Population genetics is a powerful tool for integrated pest management, and can inform effective management strategies by elucidating how genetic diversity, population size, and habitat connectivity influence population dynamics (Rollins *et al.* 2006; Tiroesele *et al.* 2014; Pélissié *et al.* 2018; Combs*et al.* 2019). Genetic assessments of population dynamics are particularly important when organisms lack comprehensive historical occurrence records (e.g. Mori *et al* . 2016) or are not easily observed in the field, as is the case with CFM. Next generation sequencing (NGS) approaches, particularly those that use restriction enzymes to digest DNA and ultimately produce large single nucleotide polymorphism (SNP) datasets, have recently become widespread in population genomic studies. These approaches can assess hundreds or thousands of markers across the genome in organisms with no existing genomic resources (Davey & Blaxter 2010; Andrews *et al.* 2016), and often provide a more comprehensive representation of population dynamics compared to one or a few markers (Dussex *et al.* 2016; Vendrami *et al.* 2017; Liu *et al.* 2019).

Small organisms present a challenge to restriction enzyme-based approaches, as these approaches require higher quality and quantity of DNA than traditional gene sequencing. The development of whole genome amplification (WGA) techniques, which amplify genomic DNA prior to NGS library preparation and sequencing, present a possible solution to this problem. However, the application of WGA in NGS datasets is relatively new and there have been few studies to date that have assessed whether WGA is likely to introduce amplification biases that may impact genome coverage and genotyping, particularly in organisms that lack a reference genome (Lovmar *et al.* 2006; El Sharawy *et al.* 2012; Ellegaard *et al.* 2013; Cruaud *et al.* 2018). In the first study to comprehensively investigate WGA for insect population genetics using non-pooled samples, de Medeiros & Farrell (2018) found WGA resulted in sufficient libraries for analysis, albeit with fewer loci. Given the likelihood that WGA techniques will see increased use in SNP-based studies of small organisms, further assessment of amplification and sequencing biases in this context is necessary.

Here we assessed the genomic structure of CFM, and investigated whether the use of WGA prior to NGS introduced differences in locus recovery, SNP genotyping, and estimates of polymorphism that may impact downstream population genomic analyses. We sampled CFM across its known range and compared the population structure recovered with *COI* haplotype data to that of genomic SNPs. This is the first genomic study of CFM, which presents a data-rich foundation for continued study of population dynamics of this pest and highlights several areas for future research to improve risk assessment and monitoring efforts for CFM.

METHODS

CFM surveying, specimen collection, and DNA extraction

A comprehensive survey for CFM was conducted throughout the canola-producing regions of Alberta, Saskatchewan, and Manitoba in 2017 and 2018 (Vankosky *et al. in prep*). Surveyors visited 546 fields, from the northern limit of canola production to the southern limit of CFM range in Alberta and Saskatchewan. In Manitoba, the survey was mostly limited to the agricultural extent in the northwest of the province, with

the exception of a single, additional site in Portage la Prairie. At each site, 100 canola racemes along the edge of each field were examined. All galled flowers found were collected and returned to the laboratory in a refrigerated container. In the laboratory, buds were dissected, and larvae were placed into individual 2 ml tubes and frozen at -80 °C (Vankosky *et al. in prep*). From all survey results, we subsampled sites for genetic analysis by selecting the sites that had the highest CFM densities, defined as any location where more than four larvae were sampled. Our genetic sampling also aimed to maximize the geographic scope across the range of CFM.

Genomic DNA was extracted from whole specimens sampled at 16 localities (Table S1) using a QIAamp DNA Micro Kit (QIAGEN). The final DNA concentration of each sample (either with or without WGA, see below) was standardized to 20 ng/µl for library preparation following the two-enzyme genotyping-by-sequencing (GBS) method of Poland *et al.*(2012).

Whole genome amplification, library preparation, and sequencing

Given the small body size of CFM and the relatively high amount of input-DNA required for GBS (200 ng per sample), consistently isolating enough DNA from each specimen was challenging. Recently developed WGA methods, such as the REPLI-g WGA Mini Kit (QIAGEN) hold promise for NGS studies of small organisms. The REPLI-g mini kit uses multiple displacement amplification to amplify genomic DNA (Cheung & Nelson 1996), and typical usage can produce an average product length of 10 kb. These kits advertise uniform DNA amplification however some studies have suggested that they can introduce amplification biases, impacting genome coverage, and they have also been reported to co-amplify contaminant DNA (Ellegaard *et al.* 2013; de Medeiros & Farrell 2018). Although a handful of studies have used such WGA kits for NGS of small organisms (Blair *et al.* 2015; Onyango et al. 2015; Cruaud *et al.*2018; de Medeiros & Farrell 2018), only two studies have assessed the impact of amplification biases in non-pooled samples of individuals using restriction enzyme-based SNP genotyping methods, a suite of techniques that includes GBS. Blair *et al.* (2015) tested the effect of WGA on locus recovery and genotyping using relatively high levels of input DNA (100 ng), per manufacturer's specifications, and reported essentially no difference in locus recovery or genotyping between treatments. A similar study using variable quantities of input DNA (as low as 6 ng) found that genome coverage appeared to be impacted by sample-specific differences in the amount of DNA used for WGA (de Medeiros & Farrell, 2018).

To test the effect of WGA on GBS approaches of small insect samples, we created GBS libraries with and without WGA for 24 of the CFM samples collected in 2017 (n = 48 libraries). Given preliminary results of these 48 libraries, the remaining 96 CFM samples collected in 2017 and 2018 underwent WGA prior to library preparation. GBS library preparation largely followed Poland *et al.* (2012), and any modifications to this protocol are detailed in Erlandson *et al.* (2019). Paired-end sequencing was conducted in two runs using an Illumina HiSeq 2500: the 24 individuals used to assess the effect of WGA on GBS library preparation were pooled and sequenced separately from the remaining 96 individuals. A 439 basepair region of the mitochondrial *COI* gene was also amplified for each specimen and sequenced on an ABI 3730xl Sanger sequencer following Mori *et al.* (2019). All sequencing (GBS and *COI*) was conducted at the National Research Council of Canada Laboratory (Saskatoon, Saskatchewan, Canada).

Data processing and Stacks parameter testing

GBS sequence data was demultiplexed on the Cedar cluster hosted by Compute Canada using the *process_radtags* module in Stacks v. 2.3 (Rochette *et al.* 2019). Parameter testing following Paris*et al.* (2017) was conducted on the 24 individuals sequenced with and without the REPLI-g treatment (herein referred to as the "WGA test dataset") using the *denovo_map.pl* script to determine the optimal values of the M and n parameters during subsequent*de novo* locus construction and SNP calling. The Mparameter controls the number of mismatches allowed between stacks in the same individual, which represent unique alleles, and the $n$parameter controls the number of mismatches in stacks across individuals as they are merged into loci (Catchen *et al.* 2011; Rochette*et al.* 2019). We tested values between 1 and 9 for both parameters. Lower values of M and n permit fewer mismatches between stacks and, barring exceptionally high levels of natural

polymorphism, should be more optimal in regional studies such as this one, where few geographic barriers exist between populations (Paris*et al.* 2017).

Following the recommendations in Paris *et al.* (2017), we additionally set the m parameter to 3, which controls the minimum allele depth, and used the *r80* principle, a stringent approach to data filtering that retains only loci that are present in 80% of the dataset. When genomic data are assembled *de novo,* there is risk of constructing loci from contaminant DNA, and some studies have reported that WGA can increase the representation of such contaminants in raw sequence reads (Ellegaard *et al.* 2013; de Medeiros & Farrell 2018). However, contaminant DNA, if present, is typically unequally distributed among samples, so using the *r80* parameter should reduce this risk (Paris *et al.* 2017); de Medeiros & Farrell (2018) found that a similar, stringent filtering approach was effective at removing such contaminants from their dataset. We assessed the number of recovered loci, polymorphic loci, and SNPs across each value of M and n independently for the WGA and non-WGA sequences in the WGA test dataset to assess any differences in the data that might be attributed to this treatment prior to GBS library preparation.

For CFM population genomic analyses, we processed all the WGA sequences from both sequencing runs together (n = 120, herein referred to as the "population genetic dataset"), specified a minimum minor allele frequency of 3%, limited the number of SNPs output per locus to one using the *–write_random_snp* option in the *populations*module of Stacks to reduce genomic linkage, and removed any individuals with more than 50% missing data. *COI* sequences for the same specimens were aligned and quality checked following Mori *et al.*(2019).

Population genetic analyses

We conducted hierarchical clustering analyses of SNPs for the 16 sampled localities in the population genetic dataset using principal components analysis (PCA) and the program Structure 2.3.4 (Pritchard *et al.*2000). Principal Component Analyses (PCAs) were conducted using*glPca* in adegenet (Jombart 2008) implemented in R 3.6.1 (R Core Team 2019), and plotted with ggplot2 (Wickham 2009). Structure was set to use the admixture model and correlated allele frequencies and was run with and without using sampling locations as a prior (*locprior*vs. *nolocprior*). We tested K = 1-20 with 20 independent replicates per value of K . Each value of K ran for 400,000 MCMC reps with a burn-in period of 200,000 and we averaged runs using CLUMPAK v1.1 (Kopelman *et al.* 2015). Following the recommendations of Janes *et al.* (2017), we considered multiple metrics when determining the optimal value of K , including comparison to the PCA, LnPr(X $|K$) (Pritchard*et al.* 2000), ΔK (Evanno *et al.* 2005), and the statistics proposed by Puechmaille (2016). We calculated the latter with StructureSelector (Li & Liu 2017) using a population map corresponding to collection localities, and a threshold for cluster placement set to 0.5.

SNP pairwise F_{ST} was calculated in R using StAMPP (Pembleton *et al.* 2013) with 1,000 bootstrap permutations and a Benjamini-Hochberg p-value correction. Expected and observed heterozygosity (H_e and H_o , respectively) were calculated in dartR (Gruber & Georges, 2019). Isolation by distance (IBD) analysis using Euclidean distance and a Mantel test with 10,000 permutations was conducted using the R packages sna (Butts 2019), geosphere (Hijmans 2019), and adegenet (Jombart 2008). Due to potentially different biological scenarios impacting the correlation between genetic and geographic distance (e.g. a single genetic cline versus two or more distinct clines, Meirmans 2012; Teske*et al.* 2018; Maitra *et al.* 2019), the densities between points were visualized with a kernel density estimation function using the package MASS (Venables & Ripley 2002).

PopART (Leigh & Bryant 2015) was used to construct a minimum spanning network of *COI* haplotypes.

GIS mapping

To assess whether population genetic structure corresponded to landscape or ecological factors, we used QGIS (QGIS Development Team 2019) to overlay Canadian canola spatial density and soil zone data (open.canada.ca) on maps depicting the survey locations and average genetic clustering output by Structure for the CFM population dataset. The canola overlay depicts crop inventory values reported by AAFC

(averaged between 2009-2018) as rasters that indicate the level of expected canola spatial density at each geographic location; regions of green indicate high canola density and regions of pale yellow represent low density. Yearly canola inventory maps were not available, so we were unable to consider any impact of temporal changes in regional canola inventory on CFM population structure. The soil zone overlay depicts the approximate agricultural extent of the Canadian Prairies, and was used to define the northern boundary of the CFM survey (see survey methods above).

RESULTS

CFM surveys

CFM surveys in 2017 and 2018 recovered larvae at 135 of the 547 sites sampled in the northern prairie regions of Manitoba, Saskatchewan, and Alberta (excluding the Peace River Region) (Fig. 1), albeit in low numbers (<4) at most sites. Areas with positive larval records broadly corresponded to the black, dark grey, and, to a lesser extent, dark brown soil zones where canola production is the highest (Fig. 1A & B, Canola Council of Canada 2020b). These regions are bordered to the north by parkland or boreal forest, and to the south by drier regions where other Brassicaceae crops, such as mustard, are produced in higher quantities than canola (Diverse Field Crops Cluster 2020).

Sequence data characteristics and de novo *locus construction*

a. WGA test dataset

Samples treated with WGA prior to GBS library preparation had higher numbers of raw sequence reads relative to the non-WGA samples, however this read abundance was not evenly distributed across individuals (Table S1). The WGA test dataset (24 individuals sequenced with and without WGA = 48 sub-libraries) produced a total of 445.5 million raw sequence reads; 57.2 million reads were attributed to the non-WGA treated sequences and the remaining 388.4 million to the WGA treated sequences (Appendix 1: Table A1). After quality filtering, the number of retained reads dropped to 8.7 million and 80.4 million, respectively. Approximately 68% of the total sequencing reads were discarded during quality filtering due to adapter contamination, while only 2.2% of the total reads were discarded due to low quality. Across samples, 8 of the 24 samples represented approximately 80% of the WGA raw sequence reads (min: 21.6 million, max: 70.3 million, mean: 38.9 million, Appendix 1: Table A2). The remaining 16 samples contained markedly fewer raw sequencing reads (min: 2.7 million, max: 8.7 million, mean: 4.8 million). While the non-WGA samples had a more even distribution of raw reads across samples, the same proportion of samples (8 of 24) still contained the majority (55%) of the non-WGA raw reads (min: 2.9 million, max: 6 million, mean: 3.9 million, Appendix 1: Table A2), and 5 of these highly-sequenced individuals were the same between treatments.

Next, we assessed the number of invariant loci, polymorphic loci, and SNPs for each tested value of M and n using the 48 libraries in the WGA test dataset (24 with WGA and 24 without). Following Paris *et al.* (2017), we chose parameter values for$M$ and n that optimized both the number of polymorphic loci and SNPs, and for both the WGA and non-WGA treatments these values were maximized at M $2n$ 2. In the resulting dataset, we observed large differences in the number of polymorphic loci, SNPs, and overall read depth between the two treatments. The non-WGA samples had more than twice the number of loci and SNPs than the samples treated with WGA, and the mean depth of coverage in these sequences was approximately 30% that of the WGA samples (Appendix 1: Table A3). However, the mean number of SNPs per locus between treatments (non-WGA = 2.4, WGA = 2.1, Table A3) and values of observed heterozygosity (non-WGA = 0.15, WGA = 0.13, Table A3) were similar. Additionally, pairwiseF_{ST} calculations between the WGA and non-WGA sequences for each population were zero (Appendix 1: Table A4), and a PCA of this dataset clustered libraries by sample, not WGA treatment. (Appendix 1: Fig. A1).

b. Second sequencing run and population genetic dataset

The second sequencing run (96 individuals treated with WGA prior to sequencing) produced a total of 354.9 million sequence reads, which was reduced to 69.3 million after quality filtering; here, 70.7% of sequence

reads were removed during quality filtering due to adapter contamination, and 1.1% were discarded due to low quality (Appendix 1: Table A1). Both the 24 WGA libraries from the WGA test dataset and these 96 libraries were used to create the population genetic dataset, however 14 individuals containing more than 50% missing data were additionally removed; after filtering, this dataset contained 106 individuals and 1,702 SNPs (Appendix 1: Table A3) and was used for all subsequent SNP analyses.

SNP population genomic analyses

Results of PCA and Structure were concordant and supported hierarchical population structure within this dataset. In the PCA, the first and second principal components (PCs) of the 16 localities recovered two highly divergent populations from the eastern edge of the sampled region in Manitoba: Swan River and Portage la Prairie (Fig. 2A). Two Albertan localities on the western edge of our sampling region, Athabasca and Sangudo, were less distinct but the combined effect of PC 1 and PC 2 clustered them apart from the remaining 12 central localities. These western and eastern sampling edges broadly coincide with the boundaries of canola production in the Canadian Prairies, excluding the Peace River Region of Alberta, a geographically disparate region in the Boreal Plains northwest of the rest of the prairies (western-most cluster of survey points in Fig. 1); we did not recover any CFM larvae from this region in our 2017 or 2018 surveys. Hierarchical PCA omitting the divergent Manitoba localities (i.e. "14 localities") separated the two aforementioned western Alberta localities along PC 1 and PC 2 (Fig. 2B). Further hierarchical PCA omitting the divergent Manitoba and Alberta localities (i.e. "12 localities") recovered little additional substructure, although three localities, Fairy Glen, Preeceville, and Dauphin, had some individuals that appeared to be genetically distinct along PC 1 and PC 2 and others that clustered with the remaining central localities (Fig. 2C).

In Structure analyses, the use of sampling location as a prior (*locprior*) did not produce substantial differences in cluster assignments when compared to the analyses that did not incorporate this information (*nolocprior*), thus we focus only on the latter here. We found variable support for an optimal value of K : LnPr(X $|K$) displayed only a gradual plateau starting at $K = 5$ to 7, ΔK values were generally low (maximum $\Delta K = 21.8$) but supported $K = 2$, 5, 7, and 9, and the Puechmaille statistics supported $K = 5$, 6, and 7 (Fig. S1). Visualization of bar charts for all values of K indicated hierarchical structure that matched the results of the PCA: $K = 2$ and 3 separated the two eastern-most Manitoba localities and $K = 4$ separated the two western-most Alberta localities. At $K = 5$ and 6, some individuals from two Saskatchewan localities (Fairy Glen and Preeceville) formed a distinct cluster, as was observed in the PCA (Fig. 2). Beyond $K = 6$ there was little meaningful structure and additional clusters were generally represented by low Q -ratios (all bar charts presented in Fig. S1). Additionally, independent hierarchical Structure analyses of the large central cluster (12 localities) supported the same divisions as the $K = 6$ results (Fig. 2D, Fig. S1), further supporting $K = 6$ as the optimal value of K . Finally, two specimens sampled in Sangudo and Athabasca clustered with the central population rather than with their collection locality and likely represent migrants (Fig. 2D).

IBD analysis using Euclidean distance and pairwiseF_{ST} /(1-F_{ST}) values for all 16 localities was highly significant ($r^2 = 0.33$, p-value = 0.004, Fig. 3A), and remained significant after removing the eastern Portage la Prairie and Swan River localities ($r^2 = 0.26$, p-value = 0.03, Fig. 3B). However, pairwise point densities indicated "islands" of data points rather than a single cline tracking the regression line as would be expected if genetic divergence increased linearly with geographic distance. After additionally removing the Sangudo and Athabasca localities, IBD analysis of the remaining 12 central localities was not significant ($r^2 = 0.04$, p-value = 0.38, Fig. 3C), suggesting that the four divergent localities were the primary drivers of the aforementioned relationships.

Values of expected and observed heterozygosity were moderate and generally similar within each population, except for Swan River and Portage la Prairie, which both had slight heterozygote excess ($H_o = 0.24$, $H_e = 0.16$ in both populations, Table 1), and North Battleford, which had lower observed values of heterozygosity ($H_O = 0.14$,$H_E = 0.21$). We note however that the North Battleford population had far higher levels of missing data than the other populations (average missing data of North Battleford population = 45%;

average missing data across remaining populations = 9%). PairwiseF_{ST} values ranged from 0 - 0.39 (Table 1), and were lower between central populations (0 - 0.17), and higher in comparisons including at least one of the four divergent populations (Swan River, Portage la Prairie, Sangudo, and Athabasca) recovered in the PCA and Structure analyses (0.13 - 0.39).

COI haplotype mapping and summary statistics

Due to missing nucleotide (nt) sequence at the 5' and/or 3' ends in 20 specimens (min. missing = 7 nt, max. missing = 80 nt, Table S1), we created a masked dataset using the modal sequence of those missing regions for each collection locality to ensure haplotype mapping was not biased by missing data. Two specimens additionally failed to sequence and were omitted from the *COI* dataset (final n = 104). The minimum spanning haplotype network depicted a single large haplogroup and nine additional smaller haplogroups (Fig. 1C). Central populations (indicated by light blue and pink colours) had the greatest amount of haplotype diversity, however overall haplotype variation was low (number of segregating sites = 16, number of parsimony-informative sites = 13), and there was no clear spatial relationship to haplotype variation; except for the Swan River and Portage la Prairie populations, each population had sequences in more than one haplogroup. The Swan River and Portage la Prairie haplotypes were identical and clustered in the large haplogroup with several specimens collected from central populations as well as the western Sangudo population. The Athabasca population was moderately distinct, and clustered mostly in a smaller haplogroup along with a few other specimens from central populations.

DISCUSSION

Here, we present the first genomic analysis of population structure for CFM. We used WGA to generate GBS libraries from these small insects, and although we found some impact of WGA on the resulting raw sequence data, there was no appreciable impact on filtered datasets and subsequent population genomic analyses. Overall, the GBS dataset recovered very little population structure across the majority of the sampled area, although much more so than the comparable mitochondrial dataset, and the only strongly differentiated populations were geographically disparate and located at the edges of the canola production region. Given a lack of alternative explanations for this pattern, we use these data to expound on the hypothesis that CFM is a native species that has unrecognized hosts both within and outside of the main agricultural zone, which is where research on this newly described species has focused thus far.

Whole genome amplification and GBS sequencing performance

The observed differences in sequencing coverage between treatments in the WGA test dataset may be attributed to multiple factors. Five of the eight most highly sequenced samples were the same between the WGA and non-WGA treatments, so those specimens may have had higher initial molecular weight DNA compared to the other 16 individuals, which could result in more sequence tags being cut and amplified (Andrews *et al.* 2016). However, this does not sufficiently explain the overall greater number of sequence reads attributed to the WGA samples. It is possible that differences in quantification accuracy between the WGA and non-WGA treatments prior to library preparation resulted in greater amplification and sequencing output of the WGA samples, however we are unable to follow-up on this possibility.

Perhaps most significantly, we observed a high level of adapter contamination in both sequencing runs, regardless of WGA treatment. This is generally the result of input DNA fragments being shorter than the 150 bp sequencing length, thus leading to adapter sequence integration into the 3' ends of the sequencing reads and subsequent sequencing of these regions (Illumina 2020). Bioanalyzer results for the WGA test and population genetic datasets confirmed that a high proportion of short insert fragment lengths were present in our final libraries (shorter than 150 bp excluding sequencing adapters, results not shown). Nonoptimal size selection via our restriction enzyme choice or substantial shearing of an intermediate library product may produce short fragments, but due to a lack of size quantification in intermediate steps we were unable to verify the stage at which this occurred. Despite this, after processing the retained sequence reads using Stacks, we were successful in assembling a moderate number of loci with sufficient read depth for population genomic analyses (Appendix 1: Table A3). The impact of this adapter contamination would have affected

both treatments in the WGA test dataset equivalently, as both groups of samples were pooled together after primer ligation; thus, while a greater number of useable sequencing reads would have likely increased the overall number and depth of retained loci, this contamination does not appear to have compromised the study, analytically.

Our results also indicate a tradeoff between sequencing coverage and read depth when using WGA prior to GBS (Appendix 1: Table A3). This is concordant with the findings of de Medeiros & Farrell (2018), who found that samples with less input DNA were more prone to reduced genome coverage after sequencing. Our results differ from those of Blair *et al.* (2015) and Cruaud *et al.* (2018), who both found negligible differences in genome coverage and sequencing depth when comparing WGA and non-WGA samples. However, we note that Blair *et al.* (2015) used much higher quantities of input DNA for WGA than our study system permitted, and Cruaud *et al.* (2018) pooled individuals so they were unable to make the same individual comparisons presented here, and in de Medeiros & Farrell (2018).

Reported differences in sequencing depth between treatments did not appear to impact *de novo* locus construction and SNP calling in the WGA test dataset, which was consistent with other studies (Blair *et al.* 2015; Cruaud *et al.* 2018; de Medeiros & Farrell 2018). Pairwise F_{ST} comparisons, observed heterozygosity, and PCA indicated little difference in genotyping between treatments when they were filtered together (Appendix 1: Table A3, A4; Fig. A1). Additionally, similarity in the average number of SNPs per locus between treatments (Table A3) suggests that the relatively greater number of SNPs found in the non-WGA treatment can be attributed to greater genomic coverage, rather than an inflation in the level of measured polymorphism. Our results suggest that, despite the potential for unequal amplification of genomic DNA by WGA, this approach is not likely to produce significant biases that impact downstream *de novo* SNP calling, provided that read depth is sufficient. Therefore, we suggest that the benefits of WGA (namely, facilitating the use of single specimens of small species for NGS) in studies that seek to randomly sample markers across the genome outweigh the potential shortcoming of reduced genome coverage.

Population structure of CFM in the Canadian Prairies

We found little overall geographic structuring related to either canola production or soil zone in the 16 populations included in this study, although nuclear SNP analyses recovered substantially more population structure than *COI* haplotype analysis (Fig. 1). Both PCA and Structure analyses using SNPs recovered only four markedly divergent populations (Swan River and Portage la Prairie in Manitoba, and Athabasca and Sangudo in Alberta), located near the edges of canola production in those regions (Fig. 1, Fig. 2). While this may be indicative of an edge effect (sensu Cook 1961), other populations were also sampled near the edges of canola production but were not genetically distinct. The Portage la Prairie population is a possible exception to this, as these individuals were sampled from a research farm (Canada-Manitoba Crop Diversification Centre) located in the city of Portage la Prairie, and as a result may have reduced opportunities for gene flow with other CFM populations located on more rural farmland.

The remaining 12 central populations formed a large genetic cluster spanning eastern Alberta, Saskatchewan, and western Manitoba. Within this central cluster, Structure analysis indicated two distinct sources of genetic ancestry that were not clearly related to sampling geography (Fig. 1, Fig. 2D), and which was particularly pronounced in the Fairy Glen and Preeceville populations. Pairwise $F_{ST}$was also low between these central populations (Table 1), and IBD analysis was non-significant (Fig. 3C) suggesting few geographic barriers to gene flow. This is consistent with the homogenous landscape throughout much of the Canadian Prairies, and the high level of contemporary canola inventory in the sampling region of this study (Fig. 1A).

COI haplotype diversity was relatively low overall, and the four divergent populations in the SNP-based analyses were not distinct for *COI* . These results are consistent with contemporary, widespread gene flow facilitated by large-scale canola production in the Canadian Prairies. It is possible that differences in recovered population structure between SNPs and *COI* are due to temporal differences in habitat connectivity resulting from year-over-year changes in canola inventory, and/or changes in effective population sizes of CFM due to regional and temporal differences in insecticide use. The *COI*gene represents only a single haploid

marker, and if our sampling coincided with a period of greater effective population size and connectivity, we may expect to have less population structure in one or a few markers compared to thousands of diploid nuclear SNPs (Dussex*et al.* 2016; Liu *et al.* 2019). Data for historical year-over-year canola inventory production numbers or insecticide spray records are unavailable over this broad geographic range, so we cannot test this hypothesis.

Canola myopia

This study provides a much-needed foundation for understanding the population genetics and demography of CFM. However, we still know little about the historical ecology and evolution of this species, or whether CFM is likely to be a significant risk to canola production in North America. Notably, the hypothesis that CFM is native, based on its widespread distribution as well as its mitochondrial DNA and parasitoid diversity (Mori *et al.* 2019), remains speculative. The disjunct distributions of highly differentiated population genomic units in canola producing regions may provide additional evidence for this speculation, and lines of reasoning for future research. Our surveying and sampling was limited to the canola production regions across the Prairie provinces. Given the short history of widespread canola production in Canada (ca. 40 years), if CFM is native then it must have some native (and/or naturalized) hosts within or outside of this geographic region. Alternative host associations have yet to be thoroughly evaluated for this species, although CFM larvae and galls were found on mustard (*Brassica juncea* va. Centennial Brown) grown in a small plot on an AAFC research farm in Melfort, Saskatchewan (BAM, unpublished). This locality is outside of the typical mustard growing region of southwestern Saskatchewan and inside the primary distribution of CFM. If alternative hosts do exist for this species, our canola-centric sampling may have anthropogenically biased our assessments of population structure in two ways: 1) these geographically disparate, differentiated populations at the edge of the canola production region may represent bleed-over genetic structure from an alternative and more geographically widespread host range, and 2) the lack of strong differentiation across the majority of our sampling region may reflect a relatively recent bottleneck onto the anthropogenic host. Thus, if CFM is native, as hypothesized, our current assessment of population structure may suffer from the presence of unsampled "ghost populations" (sensu Beerli 2004) present on alternative hosts outside of the canola production region. Failing to sample ghost populations can decrease confidence in population assignments of sampled individuals (Beerli 2004; Slatkin 2004), and may explain the substructure recovered in our clustering analyses (Fig. 1, Fig. 2), and the ambiguous support for an optimal value of K in Structure analyses (Fig. S1).

This is largely supposition at this point; however, given the lack of alternative hypotheses to explain the disjunct pattern of highly differentiated populations at the edges of the canola production region, we believe it deserves follow up research. Most importantly, it will be vital to increase survey efforts to other Brassicaceae both within and outside the canola production region. Pheromone monitoring tools (e.g. Mori *et al.* 2016) would greatly facilitate this survey effort. Expanded surveying to also include more diverse potential habitats will provide important information about the range and host preferences of this species, and facilitate comparisons of regional or host-associated population densities that may inform CFM risk assessments and monitoring. Temporal sampling throughout the growing season will also help to clarify the number of generations that CFM can produce each year, and elucidate the ecological and population dynamics of early versus late generations. Finally, if our hypothesis of alternative hosts is substantiated, CFM may provide a unique model system for studying the consequences of a contemporary host shift onto a major commercial crop species, thus informing both the evolution of insect-plant relationships as well as impacts on pest management (Chen 2016; Bernal & Medina 2018; Bernal *et al.* 2019).

DATA ACCESSIBILITY Raw NGS sequencing files and mitochondrial DNA sequences are in the process of being deposited on NCBI SRA and GenBank, respectively.

AUTHOR CONTRIBUTIONS

BAM, MAV, JH, and SH contributed to study design, field collection and DNA sequencing. EOC, JRD, and BAM conducted analyses and wrote the manuscript. All authors provided feedback on manuscript drafts.

ACKNOWLEDGEMENTS

We would like to thank members of Agriculture and Agri-Food Canada, Canola Council of Canada, and Alberta Agriculture and Forestry, in particular Scott Meers and Shelley Barkley, for their help in collecting larval samples. This study contains information licensed under version 2.0 of the Open Government License - Canada. We also wish to recognize computational support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). Research funding was provided by the Alberta Canola Producers Commission (Alberta Canola) and the Saskatchewan Canola Development Commission (SaskCanola) as part of the Canola Agronomic Research Program (CARP Grant 2017.12) to MAV and BAM, and the University of Alberta to BAM.

REFERENCES

Agriculture and Agri-Food Canada. 2011. The Case for Canadian Mustard. Available from: http://www.agr.gc.ca/resources/prod/Internet-Internet/MISB-DGSIM/ATS-SEA/PDF/4512-eng.pdf. [Accessed May 11 2020).

Andreassen, L., Soroka, J., Grenkow, L., Olfert, O., and Hallett, R. 2018. Midge (Diptera: Cecidomyiidae) injury to Brassicaceae in field trials in northeastern Saskatchewan, Canada. *The Canadian Entomologist* , **150** , 637-651.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. 2016. Harnessing the power of RADseq for ecological and evoluationary genetics. *Nature Reviews Genetics* , **17** , 81-92.

Barthet, J. J. 2016. Canola: Overview. Pp. 237-241. IN Wrigley, C. W., Corke, H., Seetharaman, K., and Faubion, J. (Eds.) *Encyclopedia of Food Grains* (2nd edition), Academic Press.

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations.*Molecular Ecology* , **13** , 827-836.

Bernal, J. S., Dávila-Flores, A. M., Medina, R. F., Chen, Y. H., Harrison, K. E., and Berrier, K. A. 2019. Did maize domestication and early spread mediate the population genetics of corn leafhopper?*Insect Science* , **26** , 569-586.

Bernal, J. S. and Medina, R. F. 2018. Agriculture sows pests: how crop domestication, host shifts, and agricultural intensification can create insect pests from herbivores. *Current Opinion in Insect Science* ,**26** , 76-81.

Blair, C., Campbell, C. R., and Yoder, A. D. 2015. Assessing the utlity of whole genome amplified DNA for next-generation molecular ecology.*Molecular Ecology Resources* , **15** , 1079-1090.

Butts, C. T. 2019. sna: Tools for social network analysis. R package version 2.5. https://CRAN.R-project.org/package=sna

Canola Council of Canada. 2020a. What is Canola? Available from: https://www.canolacouncil.org/oil-and-meal/what-is-canola/ [Accessed April 22, 2020].

Canola Council of Canada. 2020b. Effects of Soil Characteristics. Available from: https://www.canolacouncil.org/canola-encyclopedia/field-characteristics/effects-of-soil-characteristics/ [Accessed on May 11 2020].

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. *G3* , **1** , 171-182.

Chen, Y. H. 2016. Crop domestication, global human-mediated migration, and the unresolved role of geography in pest control. *Elementa: Science of the Anthropocene* , **4** , 000106.

Chen, M., Shelton, A. M., Hallett, R. H., Hoepting, C. A., Kikkert, J. R., and Wang, P. 2011. Swede midge (Diptera: Cecidomyiidae), ten years of invasion of crucifer crops in North America. *Journal of Economic*

Entomology , **104** , 709-716.

Cheung, V. G. and Nelson, S. F. 1996. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc. Natl. Acad. Sci. USA* , **93** , 14676-14679.

Combs, M., Byers, K., Himsworth, C., and Munshi-South, J. 2019. Harnessing population genetics for pest management: theory and application for urban rats. *Human-Wildlife Interactions* ,**13** , 250-263.

Cook, L. M. (1961) The edge effect in population genetics. The American Naturalist 95:295-307

Cruaud, A., Groussier, G., Genson, G., Sauné, L, Polaszek, A., and Rasplus, J-Y. 2018. Pushing the limits of whole genome amplification: successful sequencing of RADseq library from a single microhymenopteran (Chalcidoidea, *Trichogramma*). *PeerJ* , **6** , e5640.

Davey, J. W., and Blaxter, M. L. 2010. RADseq: next generation population genetics. *Briefings in Functional Genomics* ,**9** , 416-423.

de Medeiros, B. A. S., and Farrell, B. D. 2018. Whole genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci. *PeerJ* , **6** , e5089.

Diverse Field Crops Cluster. 2020. Mustard. Available from: https://www.dfcc.ca/mustard [Accessed on May 11 2020].

Dussex, N., Chuah, A., and Waters, J. M. 2016. Genome-wide SNPs reveal fine-scale differentiation among wingless stonefly. *Evolution* ,**70** , 38-47.

Ellegaard, K. M., Klasson, L, and Andersson, S. G. E. 2013. Testing the reproducibility of multiple displacement amplification on genomes of clonal endosymbiont populations. *PLoS ONE* , **8** , e82319.

El Sharawy, A., Warner, J., Olson, J., Forster, M., Schilhabel, M. B., Link, D. R., Rose-John, S., Schreiber, S., Rosenstiel, P., Brayer, J., and Franke, A., 2012. Accurate variant detection across non-amplified and whole genome amplified DNA using targeted next generation sequencing. *BMC Genomics* , **13** , 500.

Erlandson, M. A., Mori, B. A., Coutu, C., Holowachuk, J., Olfert, O. O., Gariepy, T. D., and Hegedus, D. D. 2019. Examining population structure of a bertha armyworm, *Mamestra configurata* (Lepidoptera: Noctuidae), outbreak in western North America: implications for gene flow and dispersal. *PLoS ONE* , **14** , e0218993.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology, 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Gruber, B., and Georges, A. 2019. dartR: Importing and analyzing SNP and silicodart data generated by genome-wide restriction fragment analysis. R package version 1.1.11. https://CRAN.R-project.org/package=dartR

Hallett, R. H., Goodfellow, S. A., and Heal, J. D. 2007. Monitoring and detection of the swede midge (Diptera: Cecidomyiidae). *The Canadian Entomologist* , **139** , 700-712.

Hijmans, R. J. 2019. geosphere: Spherical trigonometry. R package version 1.5-10. https://CRAN.R-porject.org/package=geosphere

Illumina. 2020. Adapter trimming: Why are adapter sequences trimmed from only the 3' ends of reads? Available from: https://support.illumina.com/bulletins/2016/04/adapter-trimming-why-are-adapter-sequences-trimmed-from-only-the–ends-of-reads.html [accessed April 22 2020].

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. Molecular Ecology, 26, 3594–3602. https://doi.org/10.1111/mec.14187

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* , **24** , 1403-1405.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources, 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Leigh, J. W., and Bryant, D. 2015. PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution* ,**6** , 1110-1116.

Li, Y. L., & Liu, J. X. (2017). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources, 18(1), 176–177. https://doi.org/10.1111/1755-0998.12719

Liu, Y., Chen, L., Duan, X-Z., Zhao, D-S., Sun, J-T., and Hong, X-Y. 2019. Genome-wide single nucleotide polymorphisms are robust in resolving fine-scale population genetic structure of the small brown planthopper, *Laodelphax striatellus* (Fallén) (Hemiptera: Delphacidae). *Journal of Economic Entomology* , **112** , 2362-2368.

LMC International. 2016. The economic impact of canola on the Canadian economy. Report for the Canola Council of Canada, Winnipeg, Canada, December 2016. LMC International, Oxford, United Kingdom. Available from: https://www.canolacouncil.org/markets-stats/industry-overview/economic-impact-of-the-canola-industry/ [accessed April 22 2020].

Lovmar, L., and Syvänen, A-C. 2006. Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Human Mutation* , **27** , 603-614.

Maitra, A., Cunha-Machado, A. S., de Souza Leandro, A., Medeiros da Costa, F., and Scarpassa, V. M. 2019. Exploring deeper genetic structures: *Aedes aegypti* in Brazil. *Acta Tropica* ,**195** , 68-77.

Meirmans, P. G. 2012. The trouble with isolation by distance. *Molecular Ecology* , **21** , 2839–2846.

Mori, B. A., Andreassen, L., Heal, J. D., Dupuis, J. R., Soroka, J. J., and Sinclair, B. J. 2019. A new species of *Contarinia* Rondani (Diptera: Cecidomyiidae) that induces flower galls on canola (*Brassicaceae*) in the Canadian prairies. *The Canadian Entomologist* , **151** , 131-148.

Mori, B. A., Davis, C. S., and Evenden, M.L. 2016. Genetic diversity and population structure identify the potential source of the invasive red clover casebearer moth, *Coleophora deauratella* , in North America.*Biological Invasions* , **18** , 3595 -3609.

Onyango, M. G., Beebe, N. W., Gopurenko, D., Bellis, G., Nicholas, A., Ogugo, M., Djikeng, A., Kemp, S., Walker, P. J., and Duchemin, J-B. 2015. Assessment of population genetic structure in the arbovirus vector midge, *Culicoides brevitarsis* (Diptera: Ceratopogonidae), using multi-locus DNA microsatellites. *Veterinary Research* ,**46** , 108.

Paris, J. P., Stevens, J. R., and Catchen, J. M. 2017. Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution* ,**8** , 1360-1373.

Pélissié, B., Crossley, M. S., Cohen, Z. P., and Schoville, S. D. 2018. Rapid evolution in insect pests: the importance of space and time in population genomics studies. *Current Opinion in Insect Science* ,**28** , 8-16.

Pembleton, L. W., Cogan, N. O. I., and Forster, J. W. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* ,**13** , 946-952.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* ,**7** , e32253.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* ,**155** , 945-959.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. Molecular Ecology Resources, 16, 608–627. https://doi.org/10.1111/1755-0998.12512

QGIS development team. 2019. QGIS geographic information system. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

REPLI-g Mini Kit product details (2020) https://www.qiagen.com/ca/products/next-generation-sequencing/single-cell-low-input/repli-g/repli-g-mini-kit/#orderinginformation

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* , **28** , 4737-4754.

Rollins, L. A., Woolnough, A. P., and Sherwin, W. B. 2006. Population genetic tools for pest management: a review. *Wildlife Research* ,**33** , 251-261.

Slatkin, M. 2004. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* , **14** , 67-73.

Soroka, J., Grenkow, L, Mori, B., and Andreassen, L. 2019. Injury by *Contarinia* (Diptera: Cecidomyiidae) to early-seeded and late-seeded canola in northeastern Saskatchewan, Canada, and assessment of seed treatments for midge control. *The Canadian Entomologist* ,**151** , 219-235.

Statistics Canada. 2019. Production of principal field crops, November 2019. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/191206/dq191206b-eng.htm. [Accessed April 22 2020].

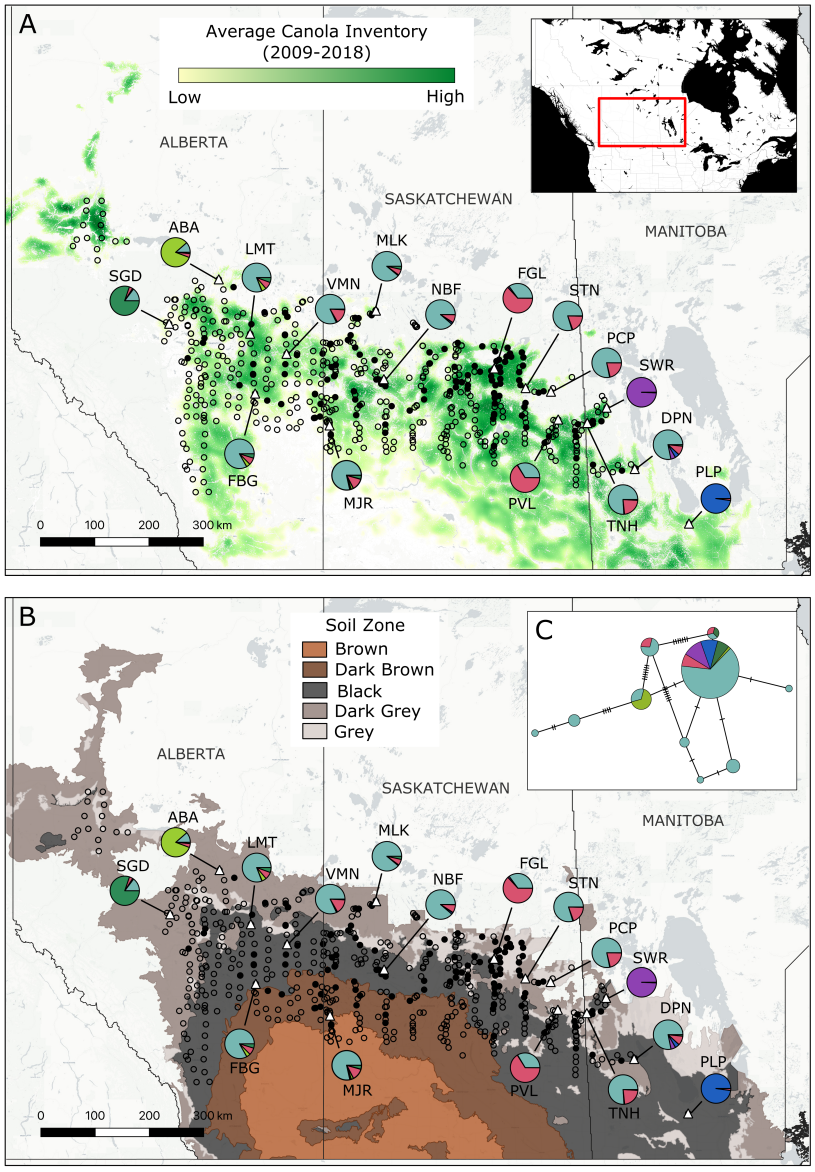
Teske, P. R., Golla, T. R., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C. D., von der Heyden, S., Chiazzari, B., van Vuuren, B. J., and Beheregaray, L. B. 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports* , **8** , 8448.

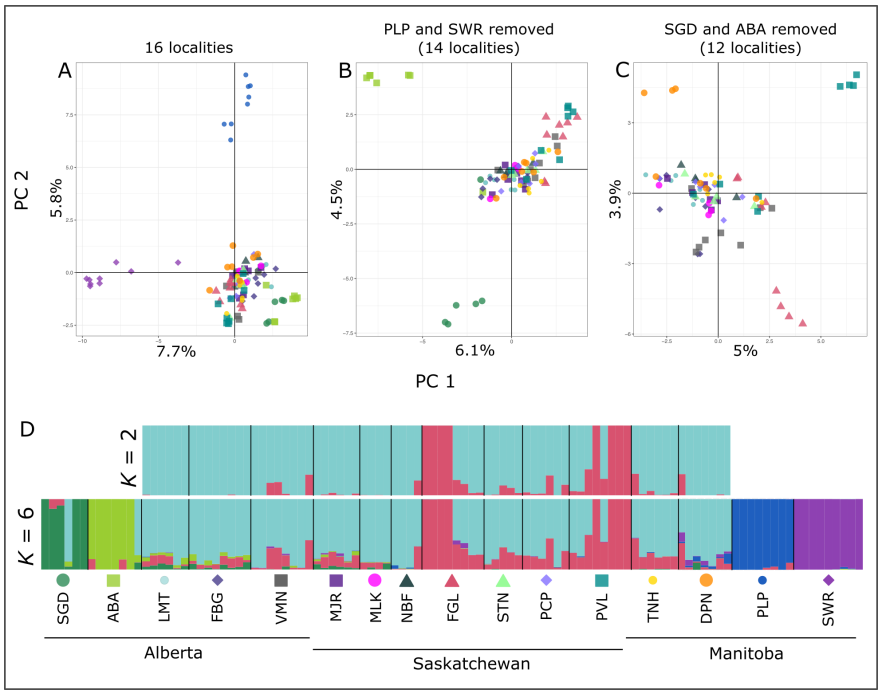
Tiroesele, B., Skoda, S. R., Hunt, T. E., Lee, D. J., Molina-Ochoa, J., and Foster, J. E. 2014. Population structure, genetic variability, and gene flow of the bean leaf beetles, *Cerotoma trifurcata,* in the Midwestern United States. *Journal of Insect Science* , **14** , 62.

Venables, W. N. and Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Vendrami, D. L. J., Telesca, L, Weigand, H., Weiss, M., Fawcett, K., Lehman, K., Clark, M. S., Leese, F., McMinn, C., Moore, H., and Hoffman, J. I. 2017.RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society Open Science* , **4** , 160548.

Wickham, H. 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, United States of America. https://ggplot2.org





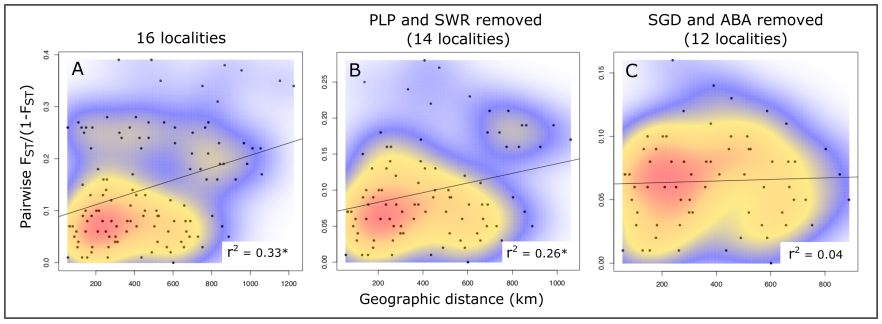


Table 1: Population-level summary statistics for the 16 localities in the CFM population dataset. Lower diagonal indicates pairwise population F_{ST} values. F_{ST} calculations with p-values < 0.05 after Benjamini-Hochberg correction are bolded. N = sample size; H_o = observed heterozygosity; H_e = expected heterozygosity. Locality abbreviations follow Fig. 1.

	PAIRWISE POPULATION F_{ST}																HETEROZYGOSITY	
LOCALITY	SGD	ABA	LMT	FBG	VMN	MJR	MLK	NBF	FGL	STN	PCP	PVL	TNH	SWR	DPN	PLP	N	H_O	H_E
SGD	--																6	0.24	0.17
ABA	**0.25**	--															7	0.22	0.17
LMT	**0.13**	**0.15**	--														6	0.21	0.21
FBG	**0.13**	**0.14**	**0.01**	--													8	0.22	0.23
VMN	**0.16**	**0.18**	**0.05**	**0.04**	--												8	0.23	0.22
MJR	**0.13**	**0.17**	0.01	**0.02**	**0.03**	--											6	0.23	0.20
MLK	**0.22**	**0.24**	**0.05**	**0.04**	**0.06**	**0.03**	--										4	0.16	0.15
NBF	**0.27**	**0.28**	**0.05**	**0.08**	**0.10**	**0.08**	**0.10**	--									4	0.14	0.21
FGL	**0.21**	**0.23**	**0.09**	**0.09**	**0.09**	**0.08**	**0.12**	**0.16**	--								8	0.20	0.20
STN	**0.16**	**0.18**	0.00	**0.02**	**0.04**	0.01	**0.06**	**0.07**	**0.07**	--							5	0.22	0.21
PCP	**0.16**	**0.18**	**0.03**	**0.04**	**0.06**	**0.02**	**0.07**	**0.08**	**0.09**	**0.01**	--						6	0.21	0.20
PVL	**0.19**	**0.21**	**0.08**	**0.08**	**0.08**	**0.07**	**0.13**	**0.14**	**0.09**	**0.06**	**0.07**	--					8	0.21	0.20
TNH	**0.16**	**0.19**	**0.03**	**0.05**	**0.06**	**0.03**	**0.10**	**0.10**	**0.08**	0.01	**0.03**	**0.07**	--				6	0.21	0.21
SWR	**0.37**	**0.38**	**0.27**	**0.26**	**0.25**	**0.26**	**0.35**	**0.39**	**0.28**	**0.26**	**0.26**	**0.27**	**0.26**	--			9	0.24	0.16
DPN	**0.17**	**0.19**	**0.05**	**0.07**	**0.09**	**0.05**	**0.11**	**0.12**	**0.11**	**0.04**	**0.06**	**0.10**	**0.06**	**0.26**	--		7	0.21	0.21
PLP	**0.34**	**0.35**	**0.22**	**0.22**	**0.23**	**0.23**	**0.31**	**0.34**	**0.27**	**0.24**	**0.24**	**0.26**	**0.25**	**0.39**	**0.22**		8	0.24	0.16