Received: 21 May 2021 | Revised: 4 October 2021 | Accepted: 5 October 2021

DOI: 10.1111/1755-0998.13528

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Chromosome-level genome assembly reveals genomic architecture of northern range expansion in the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae)

Christopher I. Keeling[1,2] | Erin O. Campbell[3] | Philip D. Batista[4] | Victor A. Shegelski[3] | Stephen A. L. Trevoy[3] | Dezene P. W. Huber[4] | Jasmine K. Janes[5,6] | Felix A. H. Sperling[3]

[1]Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, Québec, QC, Canada

[2]Département de biochimie, de microbiologie et de bio-informatique, Université Laval, Québec, QC, Canada

[3]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

[4]Faculty of Environment, University of Northern British Columbia, Prince George, BC, Canada

[5]Biology Department, Vancouver Island University, Nanaimo, BC, Canada

[6]School of Environmental and Rural Studies, University of New England, Armidale, NSW, Australia

Correspondence
Christopher I. Keeling, Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, Québec, QC, Canada.
Email: christopher.keeling@canada.ca

Felix A. H. Sperling, Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada.
Email: felix.sperling@ualberta.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: NET GP 434810-12 and RGPIN-2018-04920; Genomics Research and Development Initiative (GRDI) of the Government of Canada (NRCan)

Abstract

Genome sequencing methods and assembly tools have improved dramatically since the 2013 publication of draft genome assemblies for the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae). We conducted proximity ligation library sequencing and scaffolding to improve contiguity, and then used linkage mapping and recent bioinformatic tools for correction and further improvement. The new assemblies have dramatically improved contiguity and gaps compared to the originals: N50 values increased 26- to 36-fold, and the number of gaps were reduced by half. Ninety per cent of the content of the assemblies is now contained in 12 and 11 scaffolds for the female and male assemblies, respectively. Based on linkage mapping information, the 12 largest scaffolds in both assemblies represent all 11 autosomal chromosomes and the neo-X chromosome. These assemblies now have nearly chromosome-sized scaffolds and will be instrumental for studying genomic architecture, chromosome evolution, population genomics, functional genomics, and adaptation in this and other pest insects. We also identified regions in two chromosomes, including the ancestral-X portion of the neo-X chromosome, with elevated differentiation between northern and southern Canadian populations.

KEYWORDS
chromosome architecture, forest pest management, Hi-C, insect ecology, linkage map, proximity ligation sequencing, sex chromosomes

1 | INTRODUCTION

The availability of high-quality whole genomes facilitates an enhanced understanding of genomic organization, the mechanisms that lead to deviations from "common" architectural patterns, and the potential outcomes of such deviations. For example, a new, highly contiguous genome assembly for the honey bee, *Apis mellifera* L. (Hymenoptera: Apidae), supported the detection of inversions associated with local adaptation in high-altitude populations (Christmas et al., 2019). Comparative genomics has also contributed to a growing recognition of genomic islands of divergence and speciation (Gagnaire et al., 2013; Renaut et al., 2013). Numerous recent studies have highlighted the ecological roles that divergent regions within genomes may have and how they are formed (e.g., Bay & Ruegg, 2017; Ma et al., 2018; Renaut et al., 2013; Tavares et al., 2018).

Neo-sex chromosomes and the degeneration of male sex chromosomes (e.g., Muller's ratchet) present other aspects of speciation genomics that can be explored using high-quality genome assemblies. For example, assessments of neo-sex chromosome formation and prevalence, and the subsequent deterioration of male sex chromosomes, can provide insight into the mechanisms and consequences of their evolution. High-quality insect genomes are increasingly valuable because they present numerous complex sex-determination mechanisms that can affect their stability through recombination, mutation and drift (Blackmon et al., 2016). Recent research has highlighted the interaction between neo-sex chromosomes and reproductive isolation (Bracewell et al., 2017).

The genetics of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae), has been studied for many years (e.g., Kelley & Farrell, 1998; Kelley et al., 2000; Mock et al., 2007; Samarasekera et al., 2012; Sturgeon & Mitton, 1986), but genomic resources have grown considerably in recent years (Cullingham et al., 2018; Keeling et al., 2012; Keeling, Yuen, et al., 2013). Much of this data growth has been in response to ongoing massive population outbreaks in western North America driven by historical forest management practices and climate change, which have resulted in severe forest disturbances and losses (e.g., Raffa et al., 2008; Saab et al., 2014; Sambaraju & Goodsman, 2021). Genetics and genomics were quickly recognized as tools to help model and manage outbreaks by better understanding population dynamics (Goodsman et al., 2016; James et al., 2016); signatures of selection (Batista et al., 2016; Janes et al., 2014); the processes of olfaction (Andersson et al., 2013, 2019; Chiu et al., 2019), pheromone biosynthesis (Aw et al., 2010; Keeling, Chiu, et al., 2013; Keeling et al., 2016; Nadeau et al., 2017), host colonization (Robert et al., 2013) and overwintering (Robert et al., 2016); genome evolution and speciation (Bracewell et al., 2011, 2017; Dowle et al., 2017); and the coevolution between the beetle, its hosts and its symbiotic microorganisms (James et al., 2011).

While recent focus has been on applied pest management outcomes for *D. ponderosae*, there are numerous theoretical benefits as well, and the two often complement each other. For example, the draft mountain pine beetle genome (Keeling, Yuen, et al., 2013) revealed considerable synteny with *Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae), the most closely related genome assembly at the time. That work highlighted the important roles that specific gene families and the neo-XY chromosomal system may have in contributing to the ecological success of the mountain pine beetle (Keeling, Yuen, et al., 2013). That study also noted the lack of linkage map data to further refine our knowledge of the location and formation of the neo-XY system. Traditionally, linkage maps have been used to quantify effect sizes that certain loci have on traits of interest (e.g., quantitative trait loci) (Berlin et al., 2017), which can greatly accelerate breeding programmes and our understanding of evolutionary processes and outcomes. However, linkage maps can also assist in genome assembly of nonmodel organisms (Bartholomé et al., 2014) and provide insight into comparative analyses of genomic architecture (i.e., synteny, collinearity, chromosomal rearrangements, etc.) (Barth et al., 2019; Butler et al., 2017).

The karyotype of *Dendroctonus* spp. (Coleoptera: Curculionidae) varies from five autosome pairs (AA) + neo-XY to 14 AA + Xy_p (Zúñiga et al., 2002), with males being the heterogametic sex in all cases. In *D. ponderosae* and *D. jeffreyi* Hopkins, the karyotype is 11 AA + neo-XY. Neo-sex chromosomes result from the fusion of a sex chromosome with an autosome. Three other *Dendroctonus* spp. have a neo-XY karyotype: *D. adjunctus* Blandford (6 AA + neo-XY), *D. approximatus* Dietz (5 AA + neo-XY) and *D. brevicomis* LeConte (5 AA + neo-XY). It is not known whether the neo-XY chromosomes in these species are orthologous to those in *D. ponderosae* and *D. jefferyi*. In *D. ponderosae*, the neo-X is believed to have originated from an ancestral 12 AA + Xy_p karyotype through the fusion of the ancestral-X chromosome with one copy of the largest autosome (Lanier, 1981). The ancestral-Y (y_p) chromosome was lost, to be replaced by the other copy of the fused autosome, becoming neo-Y (Lanier, 1981; Zúñiga et al., 2002). However, several species in this genus have the ancestral 14 AA + Xy_p karyotype (Zúñiga et al., 2002), suggesting that the large autosome that became part of the neo-sex chromosomes in *D. ponderosae* may itself have been the fusion of three ancestral autosomes. Thus, a large portion of the genome content of *D. ponderosae* is in the sex chromosomes. Apart from understanding the genome architecture, delineating the sex chromosomes and the genes they contain will help us to understand how *D. ponderosae* has become one of the most important pest species in its genus, and a keystone species in western North American pine forests.

The exponential increase in the availability of whole genomes over the past decade has largely been derived from short-read sequences. However, while short reads can be assembled into larger contigs and scaffolds, sequencing gaps and repetitive regions often prevent full contiguous assembly and correct ordering (Fierst, 2015). Many genomes assembled from short reads remain highly fragmented and are best considered draft versions. Long-read sequencing technologies can overcome some issues associated with gaps and repetitive regions (Bleidorn, 2015; Rhoads & Au, 2015). These new technologies originally suffered from higher sequencing error rates but have been steadily improving. Chromosome conformation

capture sequencing, or Hi-C, is a method to study three-dimensional folding of chromosomes that offers an elegant solution to these challenges. Specifically, the Hi-C method detects closely linked sections (~10–100 kb) of DNA via chromatin contacts (van Berkum et al., 2010), producing highly accurate information on chromosome-scale genomic position (Cairns et al., 2016). When the Hi-C approach is coupled with an existing short-read draft genome, scaffolding is dramatically improved by additional anchor points interspersed throughout the short-read assembly (Li et al., 2018).

Here we provide chromosome-level male and female genome assemblies based on new proximity ligation and linkage mapping data to further support research on mountain pine beetle, which continues to exert major ecological and economic effects on pine forests in Canada and the USA. We additionally use whole-genome pool-seq and single nucleotide polymorphism (SNP) data to assess signatures of natural selection and investigate genomic linkage in the context of rapid population, geographical and host range expansions, as well as neo-sex chromosome evolution. Our results broadly characterize the neo-sex chromosomes and indicate that local adaptation in Canadian mountain pine beetle populations appears to be largely mediated by one autosomal region and the ancestral-X portion of the neo-X chromosome.

2 | MATERIALS AND METHODS

2.1 | Genome assembly and annotation

2.1.1 | Insects for genome scaffolding

Female adult beetles were collected in August 2017 from a lodgepole pine forest ~60 km south of Grande Prairie, Alberta, Canada (54.6942°N, 118.6694°W). Specimens from this collection have been vouchered at the E. H. Strickland Entomological Museum at the University of Alberta with UASM nos. 391992–391994.

2.1.2 | Chicago library preparation and sequencing

Dovetail Genomics LLC prepared a Chicago library as described previously (Putnam et al., 2016). Briefly, ~500 ng of high-molecular-weight genomic DNA (gDNA) (mean fragment length = 80 kb) from an individual female beetle (NCBI BioSample: SAMN14918906) was reconstituted into chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with *Dpn*II, the 5′-overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA was purified from associated protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment length and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of the library. The library was sequenced on an Illumina HiSeq X platform to produce 510 million 2 × 150 bp paired-end reads (NCBI SRA: SRR11965410), which provided 735× physical coverage of the genome (1–100-kb pairs).

2.1.3 | Hi-C library preparation and sequencing

Dovetail Genomics LLC prepared a Hi-C library in a manner similar to that described previously (Lieberman-Aiden et al., 2009). Briefly, chromatin from another individual female beetle (NCBI BioSample: SAMN14918916) was fixed in place with formaldehyde in the nucleus and then extracted fixed chromatin was digested with *Dpn*II. The 5′ overhangs were filled in with biotinylated nucleotides and then the free blunt ends were ligated. After ligation, crosslinks were reversed, the DNA was purified from protein, and a library was prepared as above and sequenced on an Illumina HiSeq X platform to produce 514 million 2 × 150-bp paired-end reads (NCBI SRA: SRR11965415), which provided 742× physical coverage of the genome (10–10,000-kb pairs).

2.1.4 | Scaffolding the assembly with HiRise

A schematic for the assembly process is shown in Figure 1. We used the pre-existing male (NCBI BioProject: PRJNA162621) and female (NCBI BioProject: PRJNA179493) genome assemblies (Keeling, Yuen, et al., 2013) as the input *de novo* assemblies for scaffolding, along with their underlying shotgun reads (NCBI SRA: SRR546181 and SRR546185 for male assembly, and SRR546193 for female assembly).

Separately for each sex, the *de novo* assembly, shotgun reads, Chicago library reads and Dovetail Hi-C library reads were used as input data for HIRISE, Dovetail's proprietary software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016). First, shotgun and Chicago library sequences were aligned to the draft input assembly using the SNAP version 1.0beta read mapper (Zaharia et al., 2011). Separations of Chicago read pairs mapped within draft scaffolds were analysed by HIRISE to produce a likelihood model for genomic distance between read pairs; the model was used to identify and break putative misjoins, to score prospective joins and make joins. Second, Hi-C library sequences were aligned and scaffolded following the same method. Third, shotgun sequences were used to close gaps between contigs.

2.1.5 | Correction and improvement of HIRISE assemblies

We compared the male and female HIRISE assemblies to each other using NUCMER (part of MUMMER version 3.23_3, Kurtz et al., 2004), and to linkage map information (described below) using ALLMAPS version 0.8.12 (Tang et al., 2015), to identify inconsistencies between the

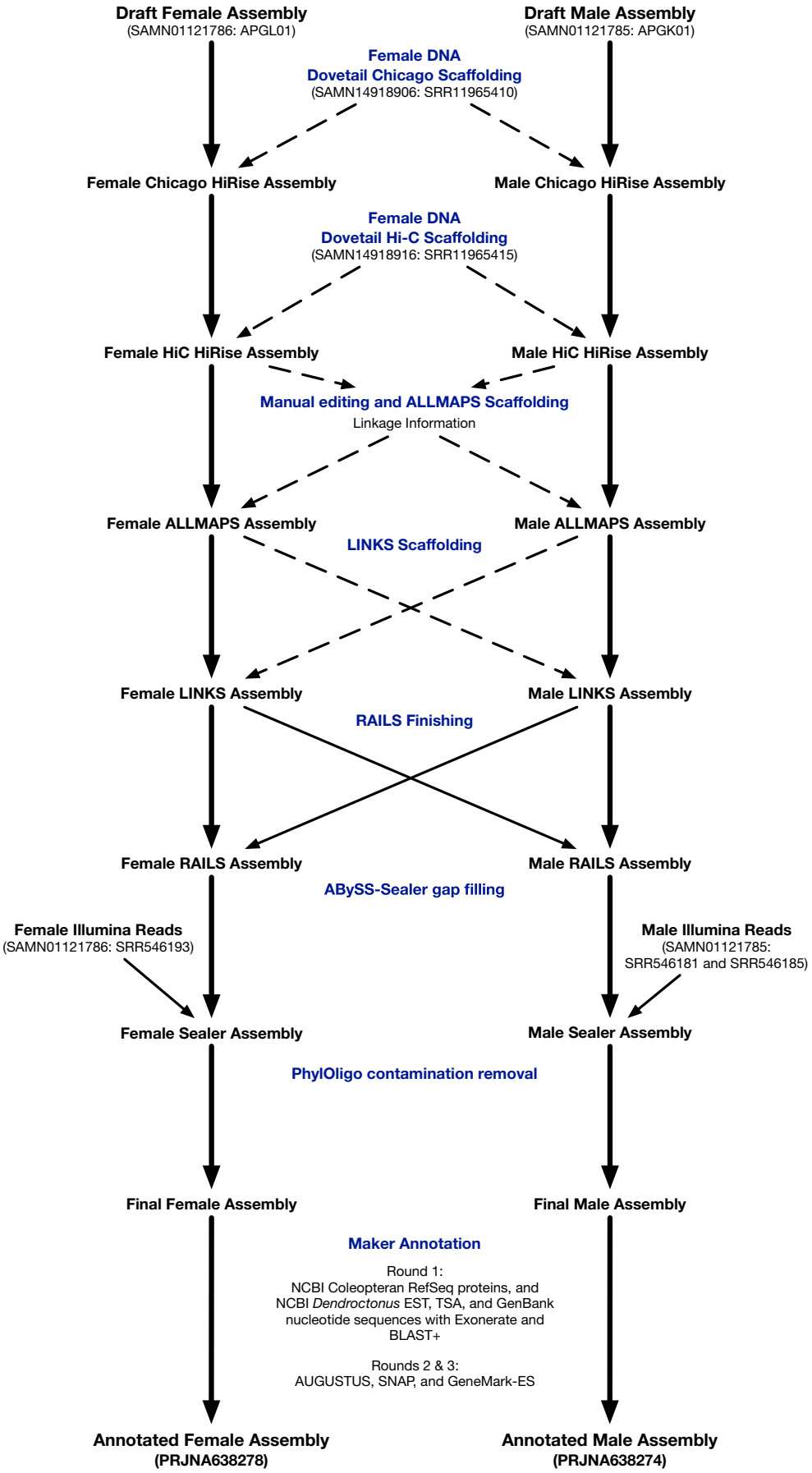
Draft Female Assembly
(SAMN01121786: APGL01)
Draft Male Assembly
(SAMN01121785: APGK01)
Female DNA
Dovetail Chicago Scaffolding
(SAMN14918906: SRR11965410)
Female Chicago HiRise Assembly
Male Chicago HiRise Assembly
Female DNA
Dovetail Hi-C Scaffolding
(SAMN14918916: SRR11965415)
Female HiC HiRise Assembly
Male HiC HiRise Assembly
Manual editing and ALLMAPS Scaffolding
Linkage Information
Female ALLMAPS Assembly
Male ALLMAPS Assembly
LINKS Scaffolding
Female LINKS Assembly
Male LINKS Assembly
RAILS Finishing
Female RAILS Assembly
Male RAILS Assembly
ABySS-Sealer gap filling
Female Illumina Reads
(SAMN01121786: SRR546193)
Male Illumina Reads
(SAMN01121785:
SRR546181 and SRR546185)
Female Sealer Assembly
Male Sealer Assembly
PhylOligo contamination removal
Final Female Assembly
Final Male Assembly
Maker Annotation
Round 1:
NCBI Coleopteran RefSeq proteins, and
NCBI Dendroctonus EST, TSA, and GenBank
nucleotide sequences with Exonerate and
BLAST+
Rounds 2 & 3:
AUGUSTUS, SNAP, and GeneMark-ES
Annotated Female Assembly
(PRJNA638278)
Annotated Male Assembly
(PRJNA638274)

FIGURE 1 Genome assembly pipeline. Each major step en route to the final annotated assemblies is shown in blue text. Sources of the input data (NCBI BioSample, assembly and SRA accessions) are indicated in parentheses. Dashed lines between the two assembly paths indicate sharing of scaffolding information only; solid lines indicate sharing of scaffolding and sequencing information. For annotation of the genomes with MAKER, empirical sequence sources and the *ab initio* gene predictors used are indicated

two assemblies, and between the assemblies and the linkage map information. Aided by NUCMER-based mummerplot comparisons of the relevant scaffolds, we resolved inconsistencies by manually flipping and/or shuffling the Chicago scaffolds that appeared to be misjoined in the final HIRISE step such that both assemblies were mutually consistent and were also consistent with the linkage map information. When there were inconsistencies between the male and female assemblies, we deemed the assembly most consistent with the linkage map information to be correct. At this stage we also shortened to 999 bp those gaps estimated during the original draft assembly to be greater than 1000 bp. We then scaffolded and gap-filled the assemblies with the corresponding assembly from the other sex using LINKS version 1.8.7 (Warren et al., 2015) and RAILS version 1.4.1 (Warren, 2016). We filled the remaining gaps in the scaffolds with ABYSS-SEALER version 2.1.5_1 (Paulino et al., 2015) using shotgun sequence data from the original *de novo* assembly (NCBI SRA: SRR546193 for female assembly, and SRR546181 and SRR546185 for male assembly). We removed contaminant scaffolds identified by PHYLOLIGO version 1.0 (Mallet et al., 2017). We used BUSCO version 4.1.4 (Simao et al., 2015) analysis with the insecta_odb10.2020-09-10 lineage data set (Kriventseva et al., 2019) to assess improvement of gene content on the final genome assemblies, and the input *de novo* assemblies. Assembly consistency plots were generated using JUPITERPLOT version 1.0 (Chu, 2020) with CIRCOS version 0.69-3 (Krzywinski et al., 2009).

2.1.6 | Gene prediction and annotation of genome assemblies

We modelled repetitive sequences with REPEATMODELER version 2.0 (part of DFAM TE TOOLS version 1.0, Dfam Consortium, 2019) and then masked the assemblies with REPEATMASKER version 4.0.7_4 (Smit et al., 2020) using these repetitive sequences and a set of transposable element proteins provided with MAKER. We then completed three rounds of MAKER version 2.31.10_2 (Holt & Yandell, 2011) gene prediction on scaffolds longer than 5 kb. For the first round, we used NCBI Coleopteran RefSeq proteins (downloaded December 11, 2019) for protein homology evidence, and NCBI *D. ponderosae* expressed sequence tags (ESTs) (downloaded December 11, 2019); *D. ponderosae* GenBank FLcDNAs: BT126413-BT128693 and JQ855638-JQ855707; and NCBI *Dendroctonus* spp. TSA accessions: GABX00000000, GAFI00000000, GAFW00000000, GAFX00000000, GDAR00000000, and GGKQ00000000 for EST evidence. Alignments were completed using EXONERATE version 2.4.0 (Slater & Birney, 2005) and BLAST+2.9.0 (Camacho et al., 2009). The results from the first round were used with BUSCO version 3.0.2_2 to train AUGUSTUS version 3.3.3 (Stanke et al., 2006). Round two used AUGUSTUS, SNAP version 2013-11-29_1 (Korf, 2004), and GENEMARK-ES version 2019-05 (Lomsadze et al., 2005) for *ab initio* gene prediction. We used the results from round two to directly train AUGUSTUS for the final round of MAKER along with SNAP and GENEMARK-ES again. We then used INTERPROSCAN version 5.40-77.0 (Jones et al., 2014) and BLAST+version 2.10.0 against the UniProtKB/Swiss-Prot 2020_01 database to functionally annotate the predicted proteins from the gene predictions. BUSCO version 4.1.4 analysis with the insecta_odb10.2020-09-10 lineage data set was completed on the predicted proteins from the first transcript variant of each gene model. The annotated assemblies are available with the following NCBI accessions: BioProjects, PRJNA638274 (male) and PRJNA638278 (female); genome accessions JAFETF000000000 (male) and JAFETG000000000 (female).

2.2 | Linkage mapping

2.2.1 | Controlled crosses

We collected mountain pine beetle larvae (F_0 generation) from infected lodgepole pine trees from Smokey River Lowlands near Grande Prairie (Alberta, Canada; 54.356267°N, 118.318533° W) and BURNCO Quarry near Canmore (Alberta, Canada; 51.067100°N, 115.287283°W) in October 2014 (Figure S1). These sites represent the two genetically distinct subpopulations previously identified within Canada (Batista et al., 2016; Janes et al., 2014; Samarasekera et al., 2012) and are referred to in this study as North and South, respectively. Collected larvae continued development in their natal bolts at 4°C until spring 2015. Next, we placed bolts in opaque emergence boxes at room temperature (~22°C) and mature mountain pine beetles were collected daily for a maximum of 15 days. Freshly emerged mountain pine beetles were sexed using the auditory method described by Rosenberger et al. (2016) to minimize physical damage. We temporarily stored the emerged adults at 4°C until a sufficient number had been collected to begin crosses (maximum 5 days). Further details are provided in Trevoy et al. (2019).

We reared North × South crosses of the F_0 generation in 55-cm bolts from three lodgepole pine, *Pinus contorta* Douglas ex Louden (Pinales: Pinaceae), trees felled in May 2015 at Nojack, Alberta (53.601717°N, 115.587317°W). Females were placed in 1.5-ml microcentrifuge tubes attached to bolts to encourage gallery construction within a specific bolt. Males from the opposite population were placed in the tube 48 hr later. Females or males that failed to enter the bolt within 24 hr were replaced with alternates from the same site of origin. We stored F_0 crosses in a sealed room at ~22°C for 6 weeks, at which point we transferred them to emergence boxes. F_1 progeny were collected over 15 days and physically sexed using the nondestructive auditory method. Mountain pine beetle adults have

been known to establish secondary galleries in different trees (Janes et al., 2016). To ensure that F_1 beetles were assigned to the correct gallery, and therefore the correct cross, each bolt was stripped and visually inspected. Any progeny from a bolt containing multiple galleries was discarded from further analysis.

We crossed F_1 beetles with gallery siblings to produce highly related F_2 individuals. Males tend to guard the entrance of a gallery and pack it with frass (Reid, 1962). Thus, after 7 days, we peeled back the wood around each gallery to a depth of 5 cm and the male was removed for genotyping. In accordance with Amman (1972), females were expected to complete gallery construction within 3 weeks. Thus, we extracted females at a later date in a similar fashion to males. Bolts were placed in emergence boxes at ~22°C until additional F_2 progeny were no longer recovered. All F_2 individuals were stored at −20°C prior to DNA extraction after physical sexing, using the dimorphic seventh tergite (Safranyik & Carroll, 2006). Sexing using the dimorphic seventh tergite was reserved for F_2 individuals because, while more accurate than the auditory method, it typically causes damage to the specimen. The crossing design and family summaries are illustrated in Figure S1. Specimens have been vouchered at the E. H. Strickland Entomological Museum at the University of Alberta with UASM nos. 391988–391991.

2.2.2 | DNA extraction, sequencing and trimming of linkage mapping samples

We extracted DNA from 229 F_1 and F_2 individuals from the North × South crosses using the Qiagen DNeasy Blood and Tissue kit with optional RNase treatment. Samples underwent SNP library preparation following the ddRAD protocol of Peterson et al. (2012) using *Pst*I and *Msp*I restriction enzymes. Illumina 75-bp single-end sequences were produced on a NextSeq500 at the University of Alberta Molecular Biology Services Unit. We demultiplexed the resulting sequence reads using the *process_radtags* module of STACKS version 2.0 (Catchen et al., 2011; Rochette et al., 2019). Remnant Illumina adapters and the *Pst*I cut site at the 5′ end of each read were trimmed using CUTADAPT version 1.10 (Martin, 2011). The trimmed length for each read was 62 bp. We have deposited the demultiplexed, trimmed reads for these samples (NCBI BioSamples SAMN17498181–SAMN17498396) into the NCBI Sequence Read Archive under accession PRJNA694171.

2.2.3 | Data processing and linkage map construction in LEP-MAP3

We used mpileup in SAMTOOLS version 1.9 (Li et al., 2009), and the pileupParser2 and pileup2posterior scripts implemented in LEP-MAP3 (Rastas, 2017), to align trimmed reads to both the male and female Hi-C HIRISE genome assemblies separately. Resulting posterior files were used as input into LEP-MAP3 to produce male- and female-aligned linkage maps. We used identity-by-descent (IBD) scores to verify the assignment of individuals to discrete families by removing individuals with less than 25% IBD to at least half of the individuals within their respective families. The ParentCall2 module imputed missing genotypes in the F_1 parents that were not recovered from the bolts. The Filtering2 module removed markers with high segregation distortion or excessive missing data (data tolerance score of 0.01, following LEP-MAP3 recommendations).

Next, we used the SeparateChromosomes2 module of LEP-MAP3 to separate SNPs into distinct linkage groups representing putative chromosomes. We required retained linkage groups to contain at least 70 SNPs, and set the informativeMask parameter to 23, which excluded markers that were informative only for the fathers (i.e., we retained markers that were either informative for the mothers, or for both mothers and fathers). Including markers informative only for fathers substantially reduced the number of SNPs assigned to linkage groups. We adjusted LOD scores until the number of retained linkage groups closely matched the known number of chromosomes (11 autosomes +2 neo-sex chromosomes) based on mountain pine beetle karyology (Lanier & Wood, 1968). In general, the appropriate LOD score should be similar to the number of chromosomes in the genome (Rastas, 2017).

We then ordered each linkage group in both maps five times using the OrderMarkers2 module of LEP-MAP3 and selected the marker order for each group with the highest likelihood score. We checked each file for incorrect marker ordering by visualizing linkage group graphs with XDOT version 1.1 (Fonseca, 2019). If any of these graphs indicated improper marker ordering, we discarded that replicate, chose the replicate with the next-highest likelihood score, and checked it again. This produced separate SNP recombination distances for male and female specimens in each linkage group, which were used as input for ALLMAPS (Tang et al., 2015). These linkage maps were used to inform joining, ordering and orientation of scaffolds in the male and female Hi-C HIRISE genome assemblies (described above). After these assembly modifications and subsequent steps were completed, we reproduced the male- and female-aligned linkage maps and ALLMAPS figures using the final versions of the male and female genome assemblies and the same parameters described above. We then repeated the marker ordering step in LEP-MAP3, this time outputting a single, sex-averaged distance for each SNP in the linkage groups in order to produce chromosome maps using the LINKAGEMAPVIEW package (Ouellette et al., 2018) in R version 3.6.1 (R Core Team, 2020). We visualized each chromosome as a density map to identify regions with strong genetic linkage, indicated by shorter per-locus centimorgan (cM) distances.

2.3 | Assessment of local adaptation in Canadian mountain pine beetle populations

2.3.1 | Mapping

Paired-end pool-seq reads described previously (Keeling, Yuen, et al., 2013) from unsexed pooled adults were obtained from the

NCBI SRA database (Fairview [FV]—SRR073440; Kananaskis [KA]—SRR086167; Terrace [TR]—SRR073431; Whitecourt [WC]—SRR073441; Cypress Hills [CH]—SRR086168; Valhalla [VA]—SRR086169; Houston [HO]—SRR086170). Paired-end reads were trimmed using BBDUK version 37.25 (Bushnell, 2020) (trimq = 20, minlength = 10) prior to mapping. Two data sets containing northern (HO, TR, WC, FV) and southern (VA, KA, CH) Canadian populations were created and mapped to the final female genome assembly using the BBMAP version 36.92 (Bushnell, 2020) plugin for GENEIOUS version 10.1.3 (Biomatters) with default settings.

2.3.2 | Differentiation and selection

We used POPOOLATION2 (Kofler et al., 2011) to evaluate nucleotide diversity (π), differentiation (F_{ST}) and Tajima's *D* between northern and southern pool-seq data sets. F_{ST} values were calculated using a non-overlapping sliding window approach (min-count −2, min-coverage 4, max-coverage 70, min-covered-fraction 0.2, pool-size 30, window-size 10,000, step-size 10,000). Tajima's *D* and nucleotide diversity values were calculated with NPSTAT version 1 (Ferretti et al., 2013) using a sliding window approach (window_length 10 000).

2.3.3 | Chromosome pathway analysis

In order to compare the gene composition in regions of interest, we performed statistical gene enrichment analysis using the online version of G:PROFILER (version e102_eg49_p15_7a9b4d6) with the g:SCS multiple testing correction method applying a significance threshold of .05 (Raudvere et al., 2019). This tool includes a database of Gene Ontology (GO) (Ashburner et al., 2000; Gene Ontology Consortium, 2021) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) functional annotations based on the draft male mountain pine beetle genome (DendPond_male_1.0). We used BLASTN with the gene models identified in the final female assembly to identify the corresponding gene IDs in this draft genome for analysis.

2.4 | Comparison of linkage groups to SNP cohorts identified in previous work

Trevoy et al. (2019) conducted principal component analyses with the draft female genome assembly to assess linkage disequilibrium in SNPs associated with Canadian population structure. They found that plateaus of high-loading SNPs in linkage disequilibrium (LD) from a number of scaffolds were driving clustering patterns on the first four principal component (PC) axes; plateaus in PCs 1 and 3 were primarily related to geography, PC 2 was sex-linked, and PC 4 was much smaller and not clearly attributed to geography or sex. Because these analyses were performed using the draft assembly, it was not possible at that time to determine whether the high-loading SNPs in each PC axis were linked, or whether they were located on multiple chromosomes and exhibited associations in allele frequency due to selection, drift or some factor other than physical proximity. To determine physical linkage and the chromosomal locations of these SNPs, we assessed the correspondence between the female draft assembly scaffolds containing high-loading SNPs with the highest loadings up to and including the plateaus in each PC shown in Trevoy et al. (2019) and the final female assembly using BLAST+ version 2.10.0 (Camacho et al., 2009). For PCs 1 and 2, we included draft scaffolds containing SNPs that had loadings equal to or greater than 0.05; for PC 3 this cut-off was 0.087, and for PC 4, it was 0.1. We created a custom BLAST database out of the final female assembly, and then used BLASTN to query the draft scaffolds for each PC against the new assembly, specifying a minimum e-value of 10^{-5}. For each PC, hits were sorted first based on e-value and then bitscore, outputting the single best match to the final assembly for each draft assembly scaffold.

2.5 | Assessment of neo-sex chromosome architecture

Identification of the neo-Y and ancestral-X portions of the neo-sex chromosomes may clarify the process of sex chromosome evolution in mountain pine beetle. Previous work has identified probable neo-Y scaffolds (Bracewell et al., 2017; Dowle et al., 2017) and ancestral-X scaffolds (Keeling, Yuen, et al., 2013) in the draft male genome assembly. We assessed the genomic locations of the putative neo-Y scaffolds in the final male genome assembly with BLASTN as described above, using the final male assembly to create a custom database and then blasting the draft scaffolds identified by Dowle et al. (2017) against it. We also used NUCMER to identify the regions of the neo-X chromosome derived from the ancestral-X scaffolds of the draft male assembly predicted by Keeling, Yuen, et al. (2013).

3 | RESULTS

3.1 | Genome assembly and annotation

The original draft genomes used paired-end and mate-pair Illumina library sequencing (Keeling, Yuen, et al., 2013). We made substantial improvements to these assemblies with proximity ligation-based scaffolding with HIRISE; linkage-map/ALLMAPS-informed corrections and scaffolding; further improvements with LINKS, RAILS and ABYSS-SEALER; and PHYLOLIGO-based removal of contaminant scaffolds (Figure 1, Table 1). All of these tools were developed after the original draft assemblies were prepared. A comparison of scaffold sizes between draft and final genome assemblies is shown in Figure S2. The final female and male genome assembly sizes were 223.7 and 224.8 Mb, with N50s/L50s of 16.6 Mb/4 and 16.4 Mb/4, respectively. Gregory et al. (2013) used flow cytometry to estimate a 208-Mb genome size. The non-N portions of the genome assemblies were very similar to this value, 214.0 Mb for the female assembly and 210.5 Mb for the

TABLE 1 Summary statistics for the genome assemblies

	Draft	Final
Female		
Total size (Mb)	261.3	223.7
Non-N length (Mb)	212.8	214.0
Scaffolds		
Number	6520	2136
Longest (Mb)	7.21	63.6
N50 (Mb)/L50	0.466/137	16.6/4
N90 (Mb)/L90	0.0271/943	3.22/12
Average length	40,082	104,747
Number >10 kb	1 696	554
% of genome >10 kb	94.8	97.9
Gene models	12,873	13,393
Transcripts	13,066	16,041
Male		
Total size (Mb)	252.8	224.8
Non-N length (Mb)	201.8	210.5
Scaffolds		
Number	8188	2084
Longest (Mb)	4.26	70.7
N50 (Mb)/L50	0.629/87	16.4/4
N90 (Mb)/L90	0.0150/1422	7.29/11
Average length	30,880	107,864
Number >10 kb	1 928	213
% of genome >10 kb	92.5	97.7
Gene models	13,088	13,601
Transcripts	13,587	16,299

Note: Summary statistics for the starting draft assemblies and the final assemblies. Incremental summary statistics for each step in the assembly process are shown in Table S1.

male assembly. Compared to the draft assemblies, N50 values increased by 26- and 36-fold, and the number of scaffolds decreased by 67% and 75%, respectively. Ninety per cent of each assembly was contained in the largest 12 (female) and 11 (male) scaffolds. Based on linkage mapping information, these 12 largest scaffolds in the female assembly represent the karyotype of this species (11 AA + neo-XX). The male assembly did not contain a large scaffold representing the neo-Y chromosome.

Each step in the assembly process contributed to the improved assemblies, and incremental assembly statistics at each step are shown in Table S1. Chicago HIRISE scaffolding dramatically increased contiguity, reducing the number of scaffolds by 56%–66%. Hi-C HIRISE scaffolding reduced the number of scaffolds by an additional 21%. The linkage map information allowed us to correct misjoins in the HIRISE assemblies and join additional scaffolds. Visualization of the linkage map information with ALLMAPS allowed us to identify several instances where scaffolds from the Chicago HIRISE step were flipped and/or out-of-order with adjacent scaffolds compared to the linkage map information and the assembly from the other sex when they were scaffolded at the Hi-C HIRISE step, even though both assemblies were based upon the same scaffolding information. An example is shown in Figure 2. In total, nine of the 12 largest scaffolds were modified (Figure S3). Hi-C heatmaps for the final assemblies, as well as after the Hi-C HIRISE scaffolding step, are shown in Figure S4.

In one case only, a scaffold from the draft male assembly was flipped and misplaced during the earlier Chicago HIRISE step. Based on linkage map information, ALLMAPS joined three scaffolds to make the neo-X in the female assembly, and four scaffolds to make the neo-X in the male assembly. This made the neo-X scaffold the largest scaffold in both final assemblies, and more than three times larger than the next largest scaffold. The LINKS scaffolding step made only two and six joins, the RAILS step made eight and 11 joins while also filling in 18% and 9% of the existing gaps within scaffolds, and ABYSS-SEALER filled in 38% and 47% of the remaining gaps of the female and male genomes, respectively. We then identified and removed contaminant scaffolds with PHYLOLIGO, encompassing 1.8% and 0.3% of the female and male assemblies, respectively. Most of these contaminant scaffolds from the female and male assemblies matched most similarly to *Serratia* spp. and *Acinetobacter* spp., respectively. Both of these genera in the Gammaproteobacteria have been found in the bark beetle gut bacteriome (Hernández-García et al., 2017). The final assemblies showed good consistency between sexes in both shared synteny and chromosomal arrangement (Figure S5), and also contained 95% of the 1367 Insecta orthologous gene set (insecta_odb10.2020-09-10, Figure S6).

To annotate the genome, we used evidence from coleopteran proteins and *Dendroctonus* spp. transcripts, with *ab initio* methods for gene prediction with three rounds of MAKER3. We identified 13,393 and 13,601 gene models in the female and male genomes, respectively. This represents an ~4% increase from the original draft genome annotations. These gene models contained 91% of the Insecta orthologous gene set (Figure S6), and 74% shared significant homology to proteins in the UniProtKB/Swiss-Prot 2020_01 database. Repetitive elements occupied ~23% and 20% of the female and male genome assemblies, respectively (Table S2). Genome-wide gene content, GC content and repetitive regions are shown in Figure S7.

3.2 | Linkage mapping

3.2.1 | Controlled crosses

From the wild-collected larvae, 74 females and 69 males (F_0 generation) were used to establish 66 North × South crosses, producing the F_1 generation (Figure S1). A total of 92 female and 69 male F_1 beetles were used to establish 66 full-sibling crosses, producing the F_2 generation (Figure S1). Of these full-sibling crosses, 22 (33%) failed to produce offspring and 10 (15%) produced three or fewer offspring. A further 20 (30%) of these F_1 families were discarded due to insufficient brood size (<10 individuals) or evidence

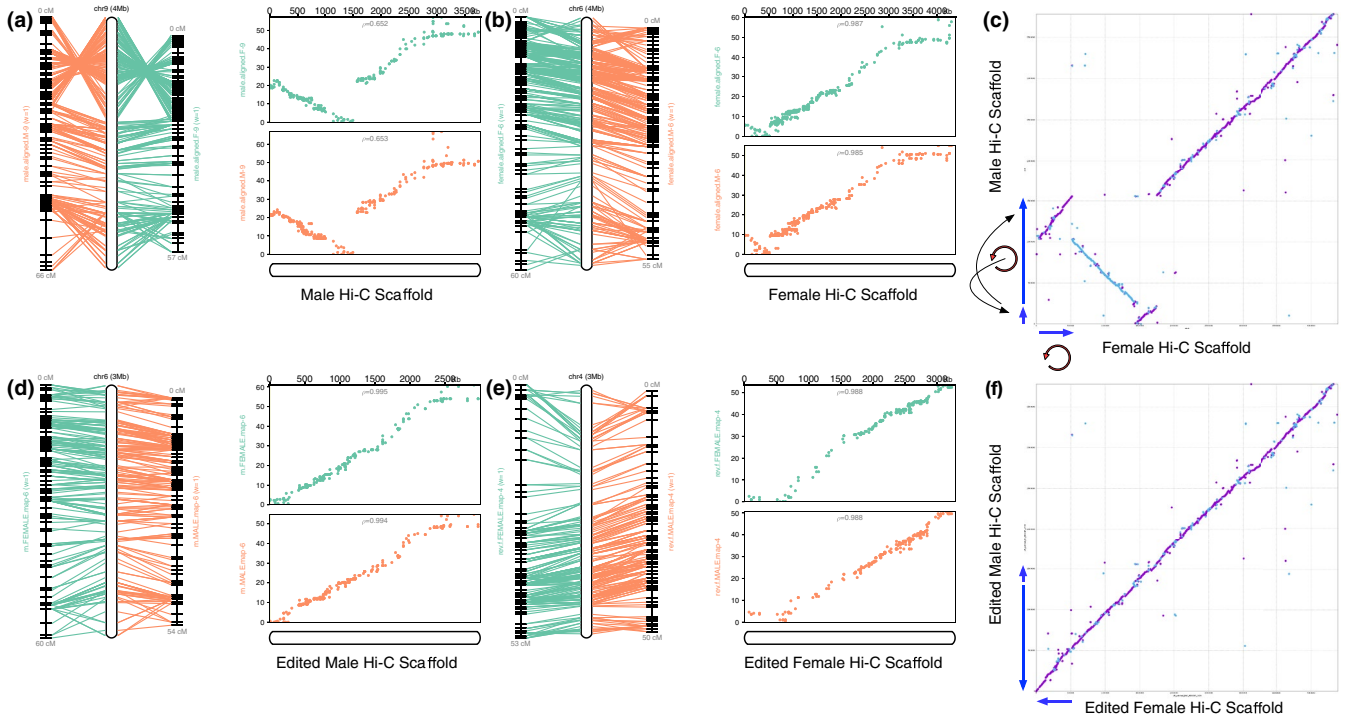

FIGURE 2 Example of manual scaffold editing based on linkage map information. Example of manual repositioning of regions of the scaffolds from the Hi-C step using information from the linkage map and male vs. female dot-plots. (a) ALLMAPS graphs for a male Hi-C scaffold showing position of markers derived from male (orange) and female (green) mapping population samples. Sections with crossing marker positions (left subgraph) and negative slope (right subgraph) indicate regions in the scaffold with orientation inconsistent with the linkage map. Sections in right subgraph that are parallel but not collinear with the trend of the curve indicate regions in the scaffold with ordering inconsistent with the linkage map. (b) ALLMAPS graphs for a female Hi-C scaffold showing position of markers derived from male and female mapping population samples. (c) Dot-plot (mummerplot with data generated by NUCMER) of the male and female Hi-C scaffolds. Blue arrows indicate the position and orientation of the Chicago scaffolds that were inconsistent with linkage map information. Black arrows indicate positional reordering of these scaffolds, and red arrows indicate a flip of the scaffold. (d) ALLMAPS graphs for the same male Hi-C scaffold after manual repositioning showing consistency with the linkage map information. (e) ALLMAPS graphs for the same female Hi-C scaffold after manual repositioning, showing consistency with the linkage map information. (f) Dot-plot of the male and female Hi-C scaffolds after manual repositioning, showing consistency between sexes. Blue arrows demark the position and orientation of the Chicago HIRISE scaffolds after the manual repositioning. For (c) and (f), purple dots indicate forward matches between scaffolds, and cyan dots indicate reverse matches between scaffolds

of gallery mixing (i.e., clear family units could not be identified). Thus, 14 (21%) F_1 broods remained for producing an F_2 generation. A total of 17 F_1 and 212 F_2 individuals were genotyped ($N = 229$), while 10 F_1 specimens (nine males and one female) that were parents of the F_2 broods were not recovered from their bolts (Table S3).

3.2.2 | Genotyping and linkage mapping

Genotyping 229 mountain pine beetles produced 414 million reads and 22,242 SNPs after alignment of samples to the final female genome and 22,087 SNPs after alignment to the final male genome. Thirteen individuals failed IBD testing in LEP-MAP3, including one entire family and three singletons from different families (Table S3). After removing these individuals and imputing genotype information for the 10 missing F_1 specimens, a total of 115 males and 111 females ($N = 226$) in 13 families were used for linkage map construction.

A LOD score of 12 recovered all 11 autosomes and the neo-X chromosome in each linkage map (Figure S3; Figure 3). Increasing the LOD score to values much higher than 12 in the male-aligned linkage maps and/or adjusting the minimum number of SNPs required to form linkage groups failed to recover a putative neo-Y chromosome, but did result in the splitting of other autosomal linkage groups formed from single genomic scaffolds. We were therefore unable to recover a neo-Y linkage group in the male-aligned linkage map. Final female and male linkage maps constructed from the final genome assemblies contained 2795 and 2910 SNPs, respectively. Linkage mapping results are shown in Tables S4–S6.

The recovered linkage groups in the female-aligned linkage map were consistent with those in the male map except for female chromosome 10, which was equivalent to male chromosome 11 and vice versa (Figure 3). Genomic scaffold numbering in this study reflects assembled chromosome size (with chromosome 1 being the largest scaffold) and is therefore not necessarily consistent between the male and female assemblies. In this instance, although male

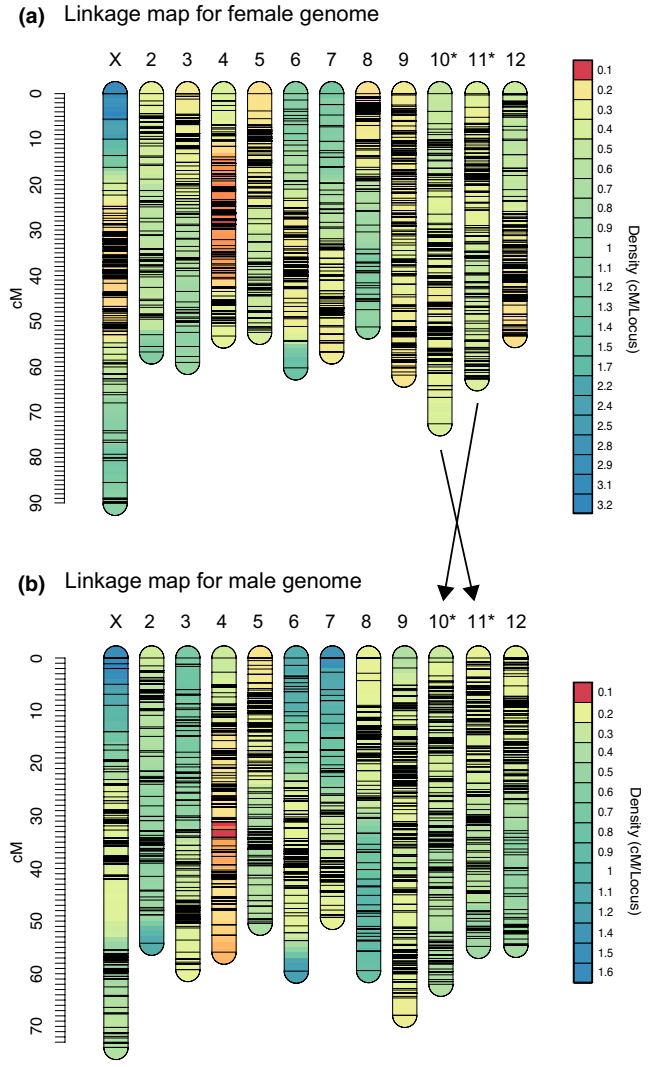

FIGURE 3 Linkage groups representing putative chromosomes for the female-aligned (a) and male-aligned (b) linkage maps. The vertical axis on the left-hand side of the figure indicates the size of each linkage group (cM), and the black horizontal lines in each linkage group depict the positions of SNPs used to construct the maps. The background colour of each linkage group shows recombination density; regions of higher density (exhibiting greater linkage) are indicated by warm colours, and regions with lower density are represented by cooler colours. Note that the female-aligned linkage groups 10 and 11 are labelled in reverse order from the equivalent male-aligned linkage groups, indicated by the arrows. Chromosomes in the genome assemblies are numbered from largest to smallest, and the assembled size of chromosomes 10 and 11 differ between the sexes

chromosome 11 and female chromosome 10 were syntenous, the assembled size in base pairs of the male scaffold was smaller than the female scaffold. The female-aligned linkage groups ranged from ~50 to 90 cM in size, and the equivalent male-aligned linkage map ranged from 50 to 70 cM. This size difference can be attributed to the male and female neo-X linkage groups; the female neo-X was ~20 cM larger than the male copy. Chromosome 4 in both the male- and female-aligned linkage maps contained a region ~30 cM in size with strong genetic linkage (~0.1 cM per locus), although the exact size and chromosomal location of this region differed slightly between the maps (Figure 3).

3.3 | Assessment of local adaptation in northern and southern Canadian mountain pine beetle populations

3.3.1 | Population differentiation and Tajima's *D*

We estimated genome-wide patterns of diversity (π), differentiation (F_{ST}) and Tajima's *D* of the pool-seq data using nonoverlapping 10-kb sliding windows for northern and southern beetle populations. Average pairwise F_{ST} across the female genome was low at $F_{ST} = 0.063$ ($SD = 0.038$), with averages for individual chromosomes ranging from 0.053 ($SD = 0.042$) to 0.092 ($SD = 0.062$) (Figure 4a, Table 2). Average Tajima's *D* was negative for both northern (−0.859) and southern (−0.808) populations (Figure 4b). Two regions exhibited elevated genetic differentiation. The first region was located at the terminal end of the neo-X chromosome between 47 and 64 Mb (Figure 4c). F_{ST} for this region was significantly higher than the rest of the chromosome (average F_{ST} = 0.096 vs. 0.038, unpaired *t* test: $t = 42.497$, $p = 5 \times 10^{-276}$). Average Tajima's *D* in this genomic region was also significantly lower in the northern population compared to the rest of the chromosome (−1.294 vs. −0.913, unpaired *t* test: $t = 20.771$, $p = 5.3 \times 10^{-88}$). This difference in Tajima's *D* was not present in the southern population.

The second region of elevated genetic differentiation was on chromosome 4 between 2.7 and 6.7 Mb (average $F_{ST} = 0.193$ vs. 0.061, unpaired *t* test: $t = 64.535$, $p = 5.1 \times 10^{-239}$) (Figure 4d). This second region exhibited a significant difference in Tajima's *D* between northern and southern populations (Tajima's $D = -0.615$ vs. 0.109, unpaired *t* test: $t = 27.756$, $p = 7.5 \times 10^{-118}$) and was located within the region of strong genetic linkage in female-aligned linkage group 4 (Figure 3a). In the linkage map, this region was roughly 23–30 cM in size and contained 147 of the 515 SNPs (~29%) in this linkage group. Upstream of this second region (start of chromosome to 2.7 Mb) we detected significantly lower levels of F_{ST} compared to the rest of the chromosome located downstream of this region (6.7 Mb to end of chromosome) (average $F_{ST} = 0.045$ vs. 0.065, unpaired *t* test: $t = 19.043$, $p = 2.1 \times 10^{-68}$). Tajima's *D* in this upstream region also showed a significant difference between northern and southern populations (Tajima's $D = 0.076$ vs. −0.155, unpaired *t* test: $t = 5.841$, $p = 8.76 \times 10^{-9}$). No differences between populations in Tajima's *D* were apparent in the downstream region.

We characterized the gene composition of these two chromosomes using GO annotation and KEGG pathway enrichment analysis. Overall, the neo-X chromosome contained genes significantly enriched in the Cellular Component GO terms: cell periphery

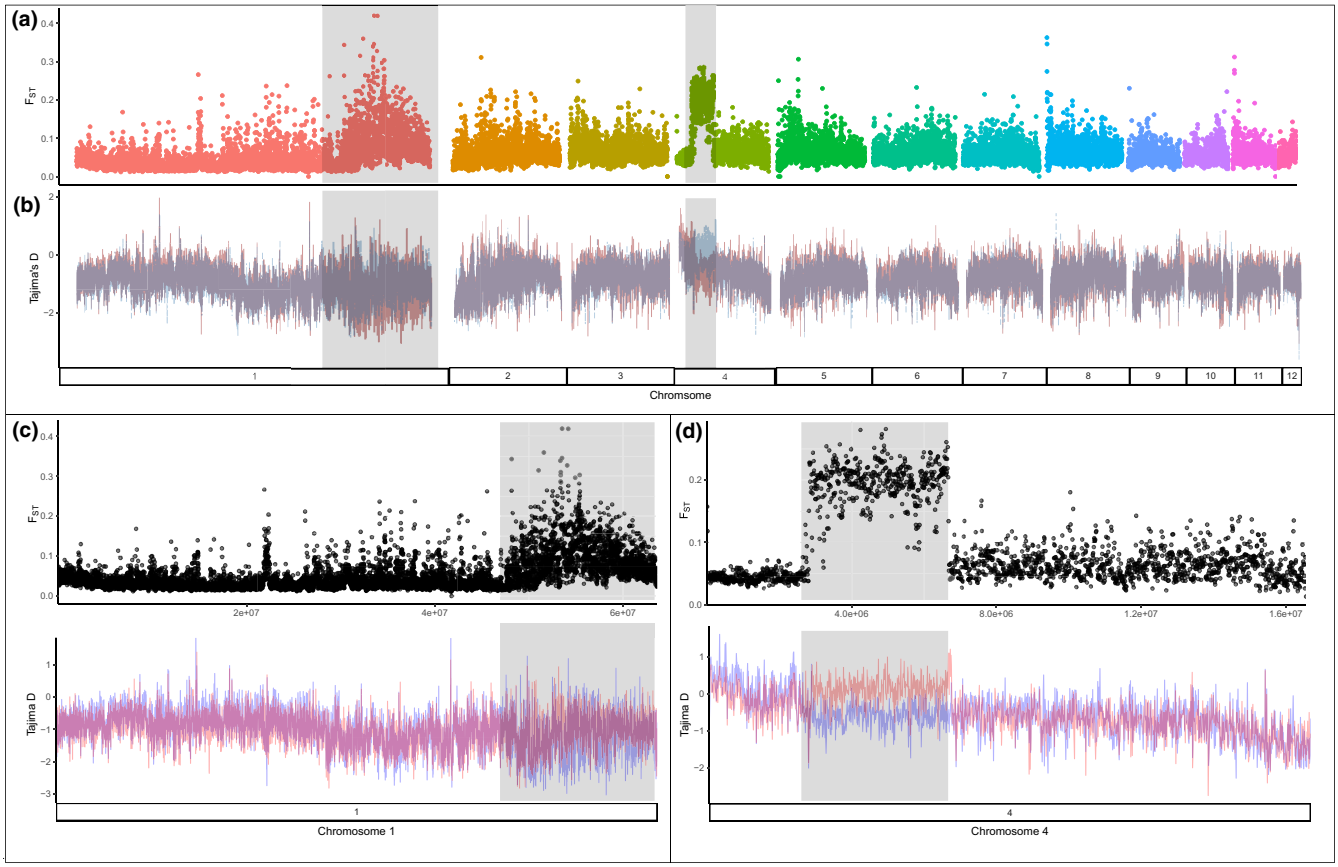

FIGURE 4 Patterns of genome-wide differentiation and Tajima's *D* of the female mountain pine beetle genome. (a) Manhattan plot for F_{ST} was calculated in 10-kb nonoverlapping sliding windows with each point representing a single estimate of F_{ST} ordered at the relative genomic position and coloured according to chromosome number. (b) Line graph of Tajima's *D* calculated in 10-kb nonoverlapping sliding windows for northern (red) and southern (blue) pool-seq populations of mountain pine beetle. (c) Manhattan plot for F_{ST} and line graph of Tajima's *D* for neo-X. (d) Manhattan plot for F_{ST} and line graph of Tajima's *D* for chromosome 4. Regions of elevated F_{ST} and Tajima's *D* are highlighted in grey

(GO:0071944), plasma membrane (GO:0005886) and SWI/SNF complex (GO:0016514) (File S1). The 18-Mb region at the terminal end of the neo-X chromosome, containing 880 genes, was significantly enriched in genes with the Molecular Function GO term: catalytic activity, acting on a protein (GO:0140096); Biological Process GO terms: transport (GO:0006810) and establishment of localization (GO:0051234); and the Cellular Component GO terms: SWI/SNF complex (GO:0016514) and gap junction (GO:0005921). No KEGG pathway was significantly enriched in this region, or the neo-X chromosome overall.

On chromosome 4, neither the region of elevated differentiation between 2.7 and 6.7 Mb, nor the region downstream had any significantly enriched GO terms. However, the region upstream contained genes significantly enriched in the Molecular Function GO terms: alcohol-forming fatty acyl-CoA reductase activity (GO:0102965), fatty-acyl-CoA reductase (alcohol-forming) activity (GO:0080019), oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (GO:0016620), and oxidoreductase activity, acting on the aldehyde or oxo group of donors (GO:0016903). No KEGG pathways were significantly enriched for chromosome 4.

3.4 | Comparison of linkage groups to SNP cohorts identified in previous work

Janes et al. (2014) identified 208 of 1536 SNPs that appeared as unique outliers over 27 sampling sites in British Columbia and Alberta, Canada. We found that 99% of the 1536 SNPs could be found in the female assembly, and 98% were found in the largest 11 scaffolds (Figure S7). No SNPs were found on scaffold 12. Ninety-eight per cent of the 208 outlier SNPs were found in the largest 11 scaffolds and did not appear clustered. Scaffolds 1 and 4 had proportionally fewer outlier SNPs compared to the other scaffolds.

Using the PC loading plateaus described above and in Trevoy et al. (2019), we identified and extracted scaffolds from the draft female assembly (Keeling, Yuen, et al., 2013) that had the highest loadings for each of the first four axes (Table S7). The 48 draft scaffolds identified from PC 1, which Trevoy et al. (2019) found to be related to geography, contained 101 SNPs and had top BLAST hits to 16 scaffolds in the final female assembly. Chromosome 4 contained 66 of these SNPs. PC 2 contained 214 SNPs on 60 draft scaffolds. With only one exception, all of these draft scaffolds had top BLAST

TABLE 2 Canadian mountain pine beetle average nucleotide diversity (π), Tajima's *D* and F_{ST} estimates

Partition	Diversity (π) North (%)	Diversity (π) South (%)	Tajima's *D* North	Tajima's *D* South	F_{ST} (North vs. South)
Ancestral-autosomal region of neo-X	1.267	1.265	−0.965	−1.050	0.038
Ancestral-X region of neo-X	0.431	0.569	−1.214	−0.913	0.096
Autosomes	1.016	1.194	−0.790	−0.813	0.067
Chromosome 2	0.741	0.859	−0.880	−0.950	0.063
Chromosome 3	0.996	1.195	−0.793	−0.805	0.068
Chromosome 4	1.301	1.587	−0.599	−0.496	0.092
Chromosome 5	0.986	1.197	−0.824	−0.838	0.068
Chromosome 6	0.926	1.117	−0.791	−0.895	0.064
Chromosome 7	0.949	1.096	−0.738	−0.784	0.062
Chromosome 8	0.838	1.021	−0.804	−0.789	0.068
Chromosome 9	1.077	1.203	−0.846	−0.916	0.055
Chromosome 10	1.191	1.431	−0.867	−0.879	0.059
Chromosome 11	1.256	1.547	−0.840	−0.873	0.062
Chromosome 12	1.723	2.079	−0.749	−0.868	0.058

Note: We used POPOOLATION2 (Kofler et al., 2011) and the final female genome assembly to evaluate nucleotide diversity (π), differentiation (F_{ST}) and Tajima's *D* between northern and southern pool-seq data sets. Nucleotide diversity graphs with a nonoverlapping sliding window across the 12 largest scaffolds are shown in Figure S9.

hits to the neo-X chromosome in the female assembly. This is consistent with the sex-based clustering along PC 2 found in Trevoy et al. (2019). PC 3 contained 50 SNPs on 11 draft scaffolds. These draft scaffolds aligned to chromosome 4, and nine of them were also found in PC 1. Of the 68 unique SNPs from PCs 1 and 3 that corresponded to female chromosome 4, 13 were located within the region of elevated F_{ST} and Tajima's *D* on that chromosome, as described above (Table S8). Finally, two draft scaffolds containing 25 SNPs from PC 4 matched to chromosome 9 in the final female assembly.

3.5 | Assessment of neo-sex chromosome architecture

We previously identified six large scaffolds in the draft male assembly with low nucleotide variation that we predicted constituted the ancestral-X portion of the neo-X chromosome (Keeling, Yuen, et al., 2013). In the new assemblies, these scaffolds were found at the terminal end of the neo-X scaffold in the same orientation and almost the same order as previously predicted, starting at ~58.3 and 51.1 Mb for the male and female neo-X scaffolds, respectively (Figure S8).

As the neo-Y did not assemble to the chromosome level, to determine the position of the ancestral X, we used BLASTN to determine where 800 putative neo-Y scaffolds in the draft assembly found by Dowle et al. (2017) were located in the new male assembly. Of these draft scaffolds, 454 matched to the male neo-X chromosome, and the remaining 346 draft scaffolds corresponded to 202 other

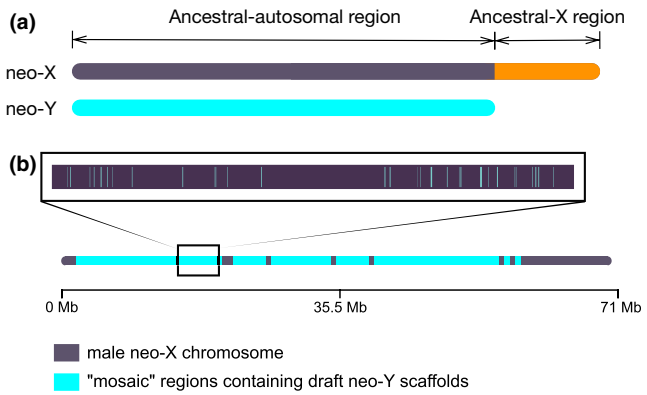

FIGURE 5 A schematic of the male neo-X chromosome. (a) Schematic of the neo-sex chromosomes. The neo-X chromosome originates from the fusion of an ancestral autosome (shown in dark grey) and the ancestral-X chromosome (shown in orange). The neo-Y chromosome (shown in cyan) originates from the other copy of the ancestral autosome found in the neo-X. (b) Schematic of the male neo-X chromosome. Regions of cyan indicate where neo-Y scaffolds identified from the draft male assembly by Dowle et al. (2017) mapped on to the male neo-X chromosome in the new genome assembly. The inset image at the top of the figure illustrates that these scaffolds are short and not contiguous along the chromosome, but rather interspersed among neo-X regions. These regions indicate enduring similarity of the neo-sex chromosomes in the regions originating from the ancestral autosome. The dark grey region on the far-right portion of the chromosome indicates a large region where putative neo-Y scaffolds failed to map, and probably represents the ancestral-X portion of the chromosome

scaffolds in the final genome (Table S9). The male neo-X chromosome spans 70,670,744 bp in length, and the 454 draft scaffolds corresponding to the neo-X chromosome were located between base pairs 1,440,820 and 59,293,732 (Table S10). These draft scaffolds were not contiguous but were instead interspersed along this region of the neo-X chromosome (Figure 5). The remaining portion of the final neo-X chromosome downstream of base pair 59,293,732 lacked any putative neo-Y scaffolds and also broadly overlapped with the region of elevated F_{ST} and Tajima's *D* on the final female neo-X chromosome, described above. All these data consistently support the same location of the ancestral-X at the terminal end of the neo-X chromosome.

4 | DISCUSSION

4.1 | Genome assembly

The draft mountain pine beetle genomes were assembled a decade ago from paired-end and mate-pair Illumina sequences during the early development of long-read sequencing and proximity ligation sequencing technologies. Since then, analytical tools have improved and new assembly tools have been developed to incorporate long-read and proximity ligation sequencing technologies (Amarasinghe et al., 2020). An upsurge of chromosome-level assemblies is currently being published using these new approaches and tools, including for many insects. We used proximity ligation sequencing and the improved analytical tools, along with linkage map information, to enhance the draft mountain pine beetle genome assemblies. Most of the assembly content was present in chromosome-sized scaffolds. Linkage map information provided critical complementary information that highlighted inconsistencies in the proximity ligation-based assemblies. Although we saw no evidence of draft scaffolds being incorrectly associated with each other, we found several instances of incorrect local ordering and orientation (Figure S3). These inconsistencies encompass (i) assembly artefacts, (ii) any conflicts between the two females used for Chicago and HiC HIRISE scaffolding, and (iii) any conflicts between male and female draft assemblies. In almost all cases, the inconsistencies resulted from orientation and ordering errors of specific Chicago assembly scaffolds during the HiC HIRISE step. This suggests that assembly artefacts were the major contributor to the observed inconsistencies between the male and female HiC assemblies. In the neo-X chromosome, the final Hi-C interaction heatmaps show a possible remaining inverted region that was not supported by the linkage map information (Figure S4). This typifies the recurring challenges of this approach. However, long-read sequencing and bioinformatic tools to reduce these errors continue to develop and improve (e.g., Nakabayashi & Morishita, 2020) and should resolve these issues in the future.

At present, genomes have been published for only two other bark beetle species (Curculionidae: Scolytinae): the coffee borer beetle, *Hypothenemus hampei* (Ferrari) (Vega et al., 2015), and the European spruce bark beetle, *Ips typographus* L. (Powell et al., 2020). Although the latter was completed with long-read PacBio sequences, our use of the complementary approaches of proximity ligation sequencing and linkage map data for scaffolding existing draft genomes resulted in more contiguous assemblies. However, gene content was similarly complete between assemblies (95.6% for *I. typographus* vs. 95.2% and 94.8%, for the female and male mountain pine beetle assemblies, respectively, based upon BUSCO version 4.1.4 with insecta_odb10.2020-09-10). These percentages suggest that the draft assemblies used for our scaffolding were missing some gene content. Future long-read sequencing data, such as PacBio or Nanopore, could improve this gene content and potentially resolve the neo-XY chromosomes fully.

4.2 | Genomic architecture reflects geographical divergence and adaptation

Divergence between northern and southern Canadian populations of mountain pine beetle has been shown in several population genetic studies (Batista et al., 2016; Janes et al., 2014; Samarasekera et al., 2012). Using available genome-wide pool-seq data (Keeling, Yuen, et al., 2013) and our newly assembled genomes, we evaluated this genetic differentiation at the chromosome level. The mountain pine beetle genome displayed two notable regions of elevated genetic divergence. The first divergent region is on the terminal end of the neo-X chromosome (Figures 4c and 5a). This 18-Mb region represents the ancestral-X portion of the neo-X chromosome. Genes in this region were significantly enriched in GO terms associated with transport, chromatin remodelling, gap junctions and catalytic activity (File S1). This finding is consistent with other work that has identified a similar pattern of increased divergence and/or population differentiation on sex chromosomes, particularly the X chromosome (summarized in Presgraves, 2018). Also of note, the region of increased genetic differentiation in the mountain pine beetle occurred on the ancestral-X portion of the neo-X chromosome and not the neo portion, a former autosome. Beaudry et al. (2020) recovered a similar pattern in plant neo-sex chromosomes. They suggested that a reduction in gene flow on the neo-sex chromosomes probably coincided with the X-autosomal fusion event, and that higher genetic differentiation in this region may be the result of local adaptation, or the unmasking and/or overall higher rate of recessive or deleterious alleles on the neo-X chromosome relative to autosomes.

Previous work by Bracewell et al. (2017) on mountain pine beetle populations in the northern USA contrasts with our findings on genetic divergence in Canadian populations, as they did not identify differences in genetic divergence between the ancestral-autosomal and ancestral-X regions of the neo-X chromosome. However, that study reported higher genetic diversity in the ancestral-autosomal region. This differs from our findings where the ancestral-X region had the lowest levels of nucleotide diversity (Table 2). We also found the ancestral-autosomal region of neo-X had relatively similar levels

of diversity to those of the other autosomes, but also showed the lowest levels of differentiation. In each chromosome and subregion, we saw a higher level of genetic diversity compared to Bracewell et al. (2017), which may be explained by the recent population expansion in Canadian populations (Table 2; Figure S9).

The second region with substantially increased divergence involves 4 Mb near the beginning of chromosome 4 (Figure 4d). This region shows a north–south difference in Tajima's *D*, with the southern population having positive Tajima's *D* values compared to the negative-shifted values in the northern population that are typical of most of the rest of chromosome 4 and other chromosomes. Positive Tajima's *D* values in the southern population suggest that genes within this region are undergoing balancing selection, while negative Tajima's *D* values in the north highlight an excess of rare alleles in this region. It remains unknown if these polymorphisms play a role in population expansion or local adaptation of mountain pine beetles in northern British Columbia and Alberta. It is possible that this pattern of differentiation indicates a chromosomal inversion between the northern and southern populations, with both inversion variants being present in the southern population. This could alter the selection occurring on genes within this genomic region and could also alter patterns of gene expression near the inversion breakpoints (Durmaz et al., 2021). However, our linkage mapping results did not indicate patterns of recombination on chromosome 4 that were consistent with an inversion, and so the differentiated F_{ST} and Tajima's *D* in this region may instead reflect the recent expansion and establishment of mountain pine beetle populations north and east of their historical Canadian range. Given the size of this differentiated region and the biological importance of the genes involved in protein synthesis and gene regulation found to be enriched in this part of the chromosome, it presents an interesting target for investigating differences in local adaptation between populations of the mountain pine beetle.

4.3 | Neo-sex chromosome characterization and its applications to studies of chromosomal evolution

Alignment of the draft neo-Y scaffolds to the neo-X revealed the approximate boundary between the ancestral-autosomal and ancestral-X portions of the neo-X chromosome. Our data suggest that the ancestral-X region may be slightly larger than 10 Mb (Figure 5), or ~14% of the total length of the neo-X chromosome.

Using linkage mapping and genome assemblies, we were also able to confirm that the markers in each of the first four principal components (PCs) identified in Trevoy et al. (2019) were physically linked to three chromosomes, suggesting that LD in these SNPs is due to physical linkage rather than drift or similar selective pressures on several independently segregating markers. Results from the Tajima's *D* and F_{ST} analyses further suggest that some of these linked markers were located in chromosomal regions found to be undergoing genetic differentiation within Canadian populations. Trevoy et al. (2019) also showed that high PC 2 loadings distinguished SNPs in beetles that were almost entirely heterozygous in males and homozygous in females. They hypothesized that this reflects sex-specific nucleotide changes in paralogous neo-X linked vs. neo-Y linked genes. If true, the male-specific heterozygosity found by Trevoy et al. (2019) actually resulted from erroneous alignment of neo-Y reads to the ancestral-autosomal region of the neo-X. Support for this conclusion comes from neo-Y scaffold locations on the draft genome, as identified by Dowle et al. (2017), mapping to our neo-X chromosome (Figure 5). Furthermore, if any of the draft genome scaffolds containing high-loading SNPs on the PC 2 axis identified by Trevoy et al. (2019) truly represented the male copy of the neo-X, we would expect some of them to align to the ancestral-X region as well, but this was not the case. Rather, we identify a mosaic-like pattern across neo-Y scaffolds that is consistent with both the recent origin of mountain pine beetle neo-sex chromosomes and the gradual accumulation of chromosome-specific mutations.

Karyotype varies across *Dendroctonus* and includes both typical Xy_p sex chromosomes (as found for the majority of beetles) and neo-sex chromosomes (Zúñiga et al., 2002), as in mountain pine beetle. We have generated the first chromosome-level assembly in this genus, and others are forthcoming (Casola et al., 2020; Keeling et al., 2020). The mountain pine beetle neo-sex chromosomes represent ~30% of its genome content. Genome assemblies from the other *Dendroctonus* spp. should identify which ancestral autosomes became part of the neo-sex chromosomes in mountain pine beetle, and the consequences of such large sex chromosomes for mountain pine beetle physiology and ecology.

This improved mountain pine beetle genome may provide another window for studying chromosome evolution relating to neo-XY development and Y chromosome degeneration. Currently, much of our understanding of neo-XY systems comes from studies of *Drosophila* spp. (Wei & Bachtrog, 2019), while mammalian and *Drosophila* systems provide insight into Y degeneration (e.g., Muller's ratchet) (Charlesworth & Charlesworth, 2000). However, neither mammals nor *Drosophila* are ideal systems for these studies because divergence times among species can be considerable—their Y chromosomes tend to be quite old, exhibiting substantial degeneration (Charlesworth & Charlesworth, 2000). Robertsonian translocations between autosomes and ancestral sex chromosomes appear to be a common mechanism for neo-XY development (Ferretti et al., 2020; Wei & Bachtrog, 2019). Such neo-chromosomes are believed to shift to achiasmatic meiosis shortly after they arise and spread through a population, preventing any further recombination of the neo-Y chromosome (Blackmon & Demuth, 2014; Wei & Bachtrog, 2019). The resulting lack of recombination has several consequences for the neo-Y chromosome. Specifically, deleterious mutations can accumulate and the effective population size of the neo-Y chromosome is greatly reduced, in part because males will have differential success in mating, but also because the neo-Y chromosome is usually smaller and contains fewer genes than its neo-X counterpart (Charlesworth & Charlesworth, 2000). These processes represent the precursor to

degeneration of Y chromosomes. Muller's ratchet is often used to further explain such degeneration. The ratchet refers to the process of individuals in a population with the lowest levels of mutation on the Y chromosome being lost until a new equilibrium is reached in the population with a degenerate Y, at which point the ratchet will move ahead another notch and the individuals with the next-lowest levels of mutation will be lost (Gordo et al., 2002). Thus, evolutionarily more recent neo-Y systems, like that of mountain pine beetle, are particularly valuable.

Several other features of the mountain pine beetle genome, coupled with the improved genomic resource presented here, make it an interesting chromosomal evolution study system. First, compared to traditional chiasmatic or achiasmatic XY systems, the ancestral Xy_p karyotypes of many *Dendroctonus* spp. may more readily result in neo-XY systems (Blackmon & Demuth, 2014; Dutrillaux & Dutrillaux, 2017). Second, Y turnover, due to degeneration followed by neo-Y gains, is common in many beetle groups (Blackmon & Demuth, 2014), and is seen in approximately one third of *Dendroctonus* species (Zúñiga et al., 2002). Third, neo-Y haplotype groups corresponding to geography and low hybrid viability have been identified in the mountain pine beetle (Dowle et al., 2017). Thus, further study of the mountain pine beetle genome may allow us to answer a variety of questions relating to Muller's ratchet-related concepts, and implications for neo-Y chromosome turnover, in phylogeographical and cyclical population irruption contexts.

In conclusion, we completed proximity ligation-based scaffolding of the draft genome of *Dendroctonus ponderosae*, supported by linkage mapping, to generate chromosome-level assemblies. These assemblies support genome-level investigations of many biological processes for this keystone species and silvicultural pest and will serve as a valuable resource for functional and evolutionary studies of other *Dendroctonus* species, other Scolytinae and Coleoptera more generally.

ACKNOWLEDGEMENTS

This research was supported by a Genomics Research and Development Initiative (GRDI) grant from the Government of Canada (to C.I.K.), a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2018-04920, to F.A.H.S.), and an NSERC Strategic Partnership Grants for Networks (NET GP 434810-12, to D.P.W.H. and F.A.H.S.) to the TRIA Network, with contributions from Alberta Agriculture and Forestry, fRI Research, Manitoba Conservation and Water Stewardship, Natural Resources Canada—Canadian Forest Service, Northwest Territories Environment and Natural Resources, Ontario Ministry of Natural Resources and Forestry, Saskatchewan Ministry of Environment, West Fraser Timber Co. Ltd and Weyerhaeuser Canada Ltd. This research was enabled in part by computing support provided by WestGrid (www.westgrid.ca) and Compute Canada Calcul Canada (www.computecanada.ca).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Designed research: C.I.K., E.O.C., P.D.B., V.A.S., S.A.L.T., D.P.W.H., J.K.J., F.A.H.S. Performed research: C.I.K., E.O.C., P.D.B., V.A.S., S.A.L.T. Contributed new reagents or analytical tools: S.A.L.T. Analysed data: C.I.K., E.O.C., P.D.B., V.A.S., S.A.L.T., D.P.W.H., J.K.J., F.A.H.S. Wrote the paper: C.I.K., E.O.C., P.D.B., J.K.J. All authors have read and approved the final version of this manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Supporting Information of this article or are openly available in NCBI at https://www.ncbi.nlm.nih.gov with the following accession numbers. Genome assemblies: BioProjects PRJNA638274 (male) and PRJNA638278 (female); Genome accessions JAFETF000000000 (male) and JAFETG000000000 (female). Chicago sequencing: BioProject PRJNA638289, BioSample SAMN14918906, SRA SRR11965410. Hi-C sequencing: BioProject PRJNA638296, BioSample SAMN14918916, SRA SRR11965415. ddRAD sequencing: BioProject PRJNA694171, BioSamples SAMN17498181-SAMN17498396.

ORCID

Christopher I. Keeling https://orcid.org/0000-0001-7885-8558
Erin O. Campbell https://orcid.org/0000-0001-8546-0636
Philip D. Batista https://orcid.org/0000-0002-9598-2793
Victor A. Shegelski https://orcid.org/0000-0003-2438-9447
Stephen A. L. Trevoy https://orcid.org/0000-0003-1217-406X
Dezene P. W. Huber https://orcid.org/0000-0002-6495-1759
Jasmine K. Janes https://orcid.org/0000-0002-4511-2087
Felix A. H. Sperling https://orcid.org/0000-0001-5148-4226

REFERENCES

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 30. https://doi.org/10.1186/s13059-020-1935-5

Amman, G. D. (1972). Mountain pine beetle brood production in relation to thickness of lodgepole pine phloem. *Journal of Economic Entomology*, *65*, 138–140. https://doi.org/10.1093/jee/65.1.138

Andersson, M. N., Grosse-Wilde, E., Keeling, C. I., Bengtsson, J. M., Yuen, M. M. S., Li, M., Hillbur, Y., Bohlmann, J., Hansson, B. S., & Schlyter, F. (2013). Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, *Ips typographus* and *Dendroctonus ponderosae* (Coleoptera: Curculionidae: Scolytinae). *BMC Genomics*, *14*, 198.–https://doi.org/10.1186/1471-2164-14-198

Andersson, M. N., Keeling, C. I., & Mitchell, R. F. (2019). Genomic content of chemosensory genes correlates with host range in wood-boring beetles (*Dendroctonus ponderosae*, *Agrilus planipennis*, and *Anoplophora glabripennis*). *BMC Genomics*, *20*(1), 690. https://doi.org/10.1186/s12864-019-6054-x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *The Gene Ontology Consortium. Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Aw, T., Schlauch, K., Keeling, C. I., Young, S., Bearfield, J. C., Blomquist, G. J., & Tittiger, C. (2010). Functional genomics of mountain pine beetle (*Dendroctonus ponderosae*) midguts and fat bodies. *BMC Genomics*, *11*(1), 215. https://doi.org/10.1186/1471-2164-11-215

Barth, J. M. I., Villegas-Ríos, D., Freitas, C., Moland, E., Star, B., André, C., Knutsen, H., Bradbury, I., Dierking, J., Petereit, C., Righton, D., Metcalfe, J., Jakobsen, K. S., Olsen, E. M., & Jentoft, S. (2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. *Molecular Ecology*, *28*(6), 1394–1411. https://doi.org/10.1111/mec.15010

Bartholomé, J., Mandrou, E., Mabiala, A., Jenkins, J., Nabihoudine, I., Klopp, C., Schmutz, J., Plomion, C., & Gion, J.-M. (2014). High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist*, *206*(4), 1283–1296. https://doi.org/10.1111/nph.13150

Batista, P. D., Janes, J. K., Boone, C. K., Murray, B. W., & Sperling, F. A. H. (2016). Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). *Ecology and Evolution*, *6*(17), 6292–6300. https://doi.org/10.1002/ece3.2367

Bay, R. A., & Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 20162414. https://doi.org/10.1098/rspb.2016.2414

Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2020). Ancestral and neo-sex chromosomes contribute to population divergence in a dioecious plant. *Evolution*, *74*(2), 256–269. https://doi.org/10.1111/evo.13892

Berlin, S., Hallingbck, H. R., Beyer, F., Nordh, N.-E., Weih, M., & Rnnberg-Wstljung, A.-C. (2017). Genetics of phenotypic plasticity and biomass traits in hybrid willows across contrasting environments and years. *Annals of Botany*, *120*(1), 87–100. https://doi.org/10.1093/aob/mcx029

Blackmon, H., & Demuth, J. P. (2014). Estimating tempo and mode of Y chromosome turnover: Explaining Y chromosome loss with the fragile Y hypothesis. *Genetics*, *197*(2), 561–572. https://doi.org/10.1534/genetics.114.164269

Blackmon, H., Ross, L., & Bachtrog, D. (2016). Sex determination, sex chromosomes, and karyotype evolution in insects. *Journal of Heredity*, *108*(1), 78–93. https://doi.org/10.1093/jhered/esw047

Bleidorn, C. (2015). Third generation sequencing: Technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, *14*(1), 1–8. https://doi.org/10.1080/14772000.2015.1099575

Bracewell, R. R., Bentz, B. J., Sullivan, B. T., & Good, J. M. (2017). Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. *Nature Communications*, *8*, 1–14. https://doi.org/10.1038/s41467-017-01761-4

Bracewell, R. R., Pfrender, M. E., Mock, K. E., & Bentz, B. J. (2011). Cryptic postzygotic isolation in an eruptive species of bark beetle (*Dendroctonus ponderosae*). *Evolution*, *65*(4), 961–975. https://doi.org/10.1111/j.1558-5646.2010.01201.x

Bushnell, B. (2020). *BBMap short read aligner*. https://sourceforge.net/projects/bbmap/

Butler, J. B., Vaillancourt, R. E., Potts, B. M., Lee, D. J., King, G. J., Baten, A., Shepherd, M., & Freeman, J. S. (2017). Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. *BMC Genomics*, *18*, 397. https://doi.org/10.1186/s12864-017-3782-7

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Vrnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.-M., Osborne, C., Fraser, P., & Spivakov, M. (2016). CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biology*, *17*(1), 1–17. https://doi.org/10.1186/s13059-016-0992-2

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Casola, C., Landa, S., Hjelman, C., Jonika, M., Sullivan, B., Kyre, B., Rieske-Kinney, L. K., & Blackmon, H. (2020). *Using comparative genomics to identify species-specific targets for RNAi in Dendroctonus bark beetles*. 2020 Entomology Virtual Annual Meeting of the Entomological Society of America.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*, 171–182. https://doi.org/10.1534/g3.111.000240/-/DC1

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *355*, 1563–1572. https://doi.org/10.1098/rstb.2000.0717

Chiu, C. C., Keeling, C. I., & Bohlmann, J. (2019). Cytochromes P450 preferentially expressed in antennae of the mountain pine beetle. *Journal of Chemical Ecology*, *45*(2), 178–186. https://doi.org/10.1007/s10886-018-0999-0

Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., & Webster, M. T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*, *28*(6), 1358–1374. https://doi.org/10.1111/mec.14944

Chu, J. (2020). *JupiterPlot*. https://github.com/JustinChu/JupiterPlot

Cullingham, C. I., Janes, J. K., Hamelin, R. C., James, P. M. A., Murray, B. W., & Sperling, F. A. H. (2018). The contribution of genetics and genomics to understanding the ecology of the mountain pine beetle system. *Canadian Journal of Forest Research*, *49*(7), 721–730. https://doi.org/10.1139/cjfr-2018-0303

Dfam Consortium (2019). *Dfam TE Tools Container*. https://github.com/Dfam-consortium/TETools

Dowle, E. J., Bracewell, R. R., Pfrender, M. E., Mock, K. E., Bentz, B. J., & Ragland, G. J. (2017). Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology*, *26*(21), 6071–6084. https://doi.org/10.1111/mec.14342

Durmaz, E., Kerdaffrec, E., Katsianis, G., Kapun, M., & Flatt, T. (2021). *How selection acts on chromosomal inversions*. *eLS*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0028745

Dutrillaux, A. M., & Dutrillaux, B. (2017). Evolution of the sex chromosomes in beetles. I. The loss of the Y chromosome. *Cytogenetic and Genome Research*, *152*(2), 97–104. https://doi.org/10.1159/000478075

Ferretti, A., Milani, D., Palacios-Gimenez, O. M., Ruiz-Ruano, F. J., & Cabral-de-Mello, D. C. (2020). High dynamism for neo-sex chromosomes: Satellite DNAs reveal complex evolution in a grasshopper. *Heredity*, *125*(3), 124–137. https://doi.org/10.1038/s41437-020-0327-7

Ferretti, L., Ramos-Onsins, S. E., & Perez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular Ecology*, *22*(22), 5561–5576. https://doi.org/10.1111/mec.12522

Fierst, J. L. (2015). Using linkage maps to correct and scaffold *de novo* genome assemblies: Methods, challenges, and computational tools. *Frontiers in Genetics*, *6*, 1–8. https://doi.org/10.3389/fgene.2015.00220

Fonseca, J. (2019). *xdot: Interactive viewer for Graphviz dot files*. Version 1.1. https://pypi.org/project/xdot/

Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*(9), 2483–2497. https://doi.org/10.1111/evo.12075

Gene Ontology Consortium (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Goodsman, D. W., Koch, D., Whitehouse, C., Evenden, M. L., Cooke, B. J., & Lewis, M. A. (2016). Aggregation and a strong Allee effect in a cooperative outbreak insect. *Ecological Applications*, *26*(8), 2623–2636. https://doi.org/10.1002/eap.1404

Gordo, I., Navarro, A., & Charlesworth, B. (2002). Muller's ratchet and the pattern of variation at a neutral locus. *Genetics*, *161*(2), 835–848. https://doi.org/10.1093/genetics/161.2.835

Gregory, T. R., Nathwani, P., Bonnett, T. R., & Huber, D. P. (2013). Sizing up arthropod genomes: An evaluation of the impact of environmental variation on genome size estimates by flow cytometry and the use of qPCR as a method of estimation. *Genome*, *56*(9), 505–510. https://doi.org/10.1139/gen-2013-0044

Hernández-García, J. A., Briones-Roblero, C. I., Rivera-Orduña, F. N., & Zúñiga, G. (2017). Revealing the gut bacteriome of *Dendroctonus* bark beetles (Curculionidae: Scolytinae): Diversity, core members and co-evolutionary patterns. *Scientific Reports*, *7*(1), 13864. https://doi.org/10.1038/s41598-017-14031-6

Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*, 491. https://doi.org/10.1186/1471-2105-12-491

James, P. M., Coltman, D. W., Murray, B. W., Hamelin, R. C., & Sperling, F. A. H. (2011). Spatial genetic structure of a symbiotic beetle-fungal system: Toward multi-taxa integrated landscape genetics. *PLoS One*, *6*(10), e25359. https://doi.org/10.1371/journal.pone.0025359

James, P. M. A., Janes, J. K., Roe, A. D., & Cooke, B. J. (2016). Modeling landscape-level spatial variation in sex ratio skew in the mountain pine beetle (Coleoptera: Curculionidae). *Environmental Entomology*, *45*(4), 790–801. https://doi.org/10.1093/ee/nvw048

Janes, J. K., Li, Y., Keeling, C. I., Yuen, M. M. S., Boone, C. K., Cooke, J. E. K., Bohlmann, J., Huber, D. P. W., Murray, B. W., Coltman, D. W., & Sperling, F. A. H. (2014). How the mountain pine beetle (*Dendroctonus ponderosae*) breached the Canadian Rocky Mountains. *Molecular Biology and Evolution*, *31*(7), 1803–1815. https://doi.org/10.1093/molbev/msu135

Janes, J. K., Roe, A. D., Rice, A. V., Gorrell, J. C., Coltman, D. W., Langor, D. W., & Sperling, F. A. H. (2016). Polygamy and an absence of fine-scale structure in *Dendroctonus ponderosae* (Hopk.) (Coleoptera: Curcilionidae) confirmed using molecular markers. *Heredity*, *116*, 68–74. https://doi.org/10.1038/hdy.2015.71

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Keeling, C. I., Campbell, E. O., Willsey, T., Strong, W., Huber, D. P. W., Sperling, F., & Bleiker, K. (2020). *Genome sequencing in bark beetles*. 2020 Entomology Virtual Annual Meeting of the Entomological Society of America.

Keeling, C. I., Chiu, C. C., Aw, T., Li, M., Henderson, H., Tittiger, C., Weng, H. B., Blomquist, G. J., & Bohlmann, J. (2013). Frontalin pheromone biosynthesis in the mountain pine beetle, *Dendroctonus ponderosae*, and the role of isoprenyl diphosphate synthases. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(47), 18838–18843. https://doi.org/10.1073/pnas.1316498110

Keeling, C. I., Henderson, H., Li, M., Yuen, M., Clark, E. L., Fraser, J. D., Huber, D. P. W., Liao, N. Y., Docking, T. R., Birol, I., Chan, S. K., Taylor, G. A., Palmquist, D., Jones, S. J. M., & Bohlmann, J. (2012). Transcriptome and full-length cDNA resources for the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major insect pest of pine forests. *Insect Biochemistry and Molecular Biology*, *42*(8), 525–536. https://doi.org/10.1016/j.ibmb.2012.03.010

Keeling, C. I., Li, M., Dullat, H. K., Henderson, H., Yuen, M. M. S., & Bohlmann, J. (2016). Quantitative metabolome, proteome and transcriptome analysis of midgut and fat body tissues in the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, and insights into pheromone biosynthesis. *Insect Biochemistry and Molecular Biology*, *70*, 170–183. https://doi.org/10.1016/j.ibmb.2016.01.002

Keeling, C. I., Yuen, M. M. S., Liao, N. Y., Docking, T. R., Chan, S. K., Taylor, G. A., Palmquist, D. L., Jackman, S. D., Nguyen, A., Li, M., Henderson, H., Janes, J. K., Zhao, Y., Pandoh, P., Moore, R., Sperling, F. A. H., Huber, D. P. W., Birol, I., Jones, S. J. M., & Bohlmann, J. (2013). Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biology*, *14*, R27. https://doi.org/10.1186/gb-2013-14-3-r27

Kelley, S. T., & Farrell, B. D. (1998). Is specialization a dead-end?: The phylogeny of host use in *Dendroctonus* bark beetles. *Evolution*, *52*, 1731–1743. https://doi.org/10.1111/j.1558-5646.1998.tb02253.x

Kelley, S. T., Farrell, B. D., & Mitton, J. B. (2000). Effects of specialization on genetic differentiation in sister species of bark beetles. *Heredity*, *84*(Pt 2), 218–227. https://doi.org/10.1046/j.1365-2540.2000.00662.x

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Korf, I. (2004). Gene finding in novel genomes [Research Support, Non--U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811. https://doi.org/10.1093/nar/gky1053

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics [Research Support, Non-U.S. Gov't]. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lanier, G. N. (1981). Cytotaxonomy of dendroctonus. In M. W. Stock (Ed.), *Application of genetics and cytology in insect systematics and evolution, Proceedings of the 1980 Annual Meeting of the Entomological Society of America* (pp. 33–66). Forest, Wildlife and Range Experimental Station, University of Idaho.

Lanier, G. N., & Wood, D. L. (1968). Controlled mating, karyology, morphology, and sex-ratio in the *Dendroctonus ponderosae* complex. *Annals of the Entomological Society of America*, *61*(2), 517–526. https://doi.org/10.1093/aesa/61.2.517

Li, C., Lin, F., An, D., Wang, W., & Huang, R. (2018). Genome sequencing and assembly by long reads in plants. *Genes*, *9*(1), 6–14. https://doi.org/10.3390/genes9010006

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., &

Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506. https://doi.org/10.1093/nar/gki937

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R. J., Lascoux, M., Nevo, E., & Liu, J. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), E236–E243. https://doi.org/10.1073/pnas.1713288114

Mallet, L., Bitard-Feildel, T., Cerutti, F., & Chiapello, H. (2017). PhylOligo: A package to identify contaminant or untargeted organism sequences in genome assemblies. *Bioinformatics*, *33*(20), 3283–3285. https://doi.org/10.1093/bioinformatics/btx396

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

Mock, K. E., Bentz, B. J., O'neill, E. M., Chong, J. P., Orwin, J., & Pfrender, M. E. (2007). Landscape-scale genetic variation in a forest outbreak species, the mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology*, *16*(3), 553–568. https://doi.org/10.1111/j.1365-294X.2006.03158.x

Nadeau, J. A., Petereit, J., Tillett, R. L., Jung, K., Fotoohi, M., MacLean, M., Young, S., Schlauch, K., Blomquist, G. J., & Tittiger, C. (2017). Comparative transcriptomics of mountain pine beetle pheromone-biosynthetic tissues and functional analysis of CYP6DE3. *BMC Genomics*, *18*(1), 311. https://doi.org/10.1186/s12864-017-3696-4

Nakabayashi, R., & Morishita, S. (2020). HiC-Hiker: A probabilistic model to determine contig orientation in chromosome-length scaffolds with Hi-C. *Bioinformatics*, *36*(13), 3966–3974. https://doi.org/10.1093/bioinformatics/btaa288

Ouellette, L. A., Reid, R. W., Blanchard, S. G., & Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics*, *34*(2), 306–307. https://doi.org/10.1093/bioinformatics/btx576

Paulino, D., Warren, R. L., Vandervalk, B. P., Raymond, A., Jackman, S. D., & Birol, I. (2015). Sealer: A scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics*, *16*, 230. https://doi.org/10.1186/s12859-015-0663-4

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Powell, D., Groβe-Wilde, E., Krokene, P., Roy, A., Chakraborty, A., Löfstedt, C., Vogel, H., Andersson, M. N., & Schlyter, F. (2020). A highly contiguous genome assembly of a major forest pest, the Eurasian spruce bark beetle *Ips typographus*. *bioRxiv*. https://doi.org/10.1101/2020.11.28.401976

Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. *Molecular Ecology*, *27*(19), 3822–3830. https://doi.org/10.1111/mec.14777

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Research*, *26*(3), 342–350. https://doi.org/10.1101/gr.193474.115

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Raffa, K. F., Aukema, B. H., Bentz, B. J., Carroll, A. L., Hicke, J. A., Turner, M. G., & Romme, W. H. (2008). Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. *BioScience*, *58*(6), 501–517. https://doi.org/10.1641/B580607

Rastas, P. (2017). Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*, *33*(23), 3726–3732. https://doi.org/10.1093/bioinformatics/btx494

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Reid, R. W. (1962). Biology of the mountain pine beetle, *Dendroctonus monticolae* Hopkins, in east Kootenay region of British Columbia II. Behaviour in the host, fecundity, and internal changes in the female. *Canadian Entomologist*, *94*, 605–613. https://doi.org/10.4039/Ent94605-6

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., & Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827. https://doi.org/10.1038/ncomms2833

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robert, J. A., Bonnett, T., Pitt, C., Spooner, L. J., Fraser, J., Yuen, M. M. S., Keeling, C. I., Bohlmann, J., & Huber, D. P. W. (2016). Gene expression analysis of overwintering mountain pine beetle larvae suggests multiple systems involved in overwintering stress, cold hardiness, and preparation for spring development. *PeerJ*, *4*, e2109. https://doi.org/10.7717/peerj.2109

Robert, J. A., Pitt, C., Bonnett, T. R., Yuen, M. M. S., Keeling, C. I., Bohlmann, J., & Huber, D. P. W. (2013). Disentangling detoxification: Gene expression analysis of feeding mountain pine beetle illuminates molecular-level host chemical defense detoxification mechanisms. *PLoS One*, *8*(11), e77777. https://doi.org/10.1371/journal.pone.0077777

Rochette, N. C., Rivera-Colon, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rosenberger, D. W., Venette, R. C., & Aukema, B. H. (2016). Sexing live mountain pine beetles *Dendroctonus ponderosae*: Refinement of a behavioral method for *Dendroctonus* spp. *Entomologia Experimentalis et Applicata*, *160*(2), 195–199. https://doi.org/10.1111/eea.12463

Saab, V. A., Latif, Q. S., Rowland, M. M., Johnson, T. N., Chalfoun, A. D., Buskirk, S. W., Heyward, J. E., & Dresser, M. A. (2014). Ecological consequences of mountain pine beetle outbreaks for wildlife in western North American forests. *Forest Science*, *60*(3), 539–559. https://doi.org/10.5849/forsci.13-022

Safranyik, L., & Carroll, A. L. (2006). The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In L. Safranyik, & B. Wilson (Eds.), *The mountain pine beetle-A synthesis of biology, management, and impacts on lodgepole pine* (pp. 3–66). Natural Resources Canada, Canadian Forest Service. https://cfs.nrcan.gc.ca/publications?id=26116

Samarasekera, G. D. N. G., Bartell, N. V., Lindgren, B. S., Cooke, J. E., Davis, C. S., James, P. M., Coltman, D. W., Mock, K. E., & Murray, B. W. (2012). Spatial genetic structure of the mountain pine beetle (*Dendroctonus ponderosae*) outbreak in western Canada: Historical patterns and contemporary dispersal. *Molecular Ecology*, *21*(12), 2931–2948. https://doi.org/10.1111/j.1365-294X.2012.05587.x

Sambaraju, K. R., & Goodsman, D. W. (2021). Mountain pine beetle: An example of a climate-driven eruptive insect impacting conifer forest

ecosystems. *CAB Reviews*, *16*(18), 1–18. https://doi.org/10.1079/PAVSNNR202116018

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Smit, A. F. A., Hubley, R., & Green, P. (2020). *RepeatMasker*. http://www.repeatmasker.org/

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, *34*, W435–W439. https://doi.org/10.1093/nar/gkl200

Sturgeon, K. B., & Mitton, J. B. (1986). Allozyme and morphological differentiation of mountain pine beetles *Dendroctonus ponderosae* Hopkins (Coleoptera: Scolytidae) associated with host tree. *Evolution*, *40*(2), 290–302. https://doi.org/10.1111/j.1558-5646.1986.tb00471.x

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., & Lu, J. (2015). ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biology*, *16*, 3. https://doi.org/10.1186/s13059-014-0573-1

Tavares, H., Whibley, A., Field, D. L., Bradley, D., Couchman, M., Copsey, L., Elleouet, J., Burrus, M., Andalo, C., Li, M., Li, Q., Xue, Y., Rebocho, A. B., Barton, N. H., & Coen, E. (2018). Selection and gene flow shape genomic islands that control floral guides. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 11006–11011. https://doi.org/10.1073/pnas.1801832115

Trevoy, S. A. L., Janes, J. K., Muirhead, K., & Sperling, F. A. H. (2019). Repurposing population genetics data to discern genomic architecture: A case study of linkage cohort detection in mountain pine beetle (*Dendroctonus ponderosae*). *Ecology and Evolution*, *9*(3), 1147–1159. https://doi.org/10.1002/ece3.4803

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J., & Lander, E. S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments*, *39*, 1–7. https://doi.org/10.3791/1869

Vega, F. E., Brown, S. M., Chen, H., Shen, E., Nair, M. B., Ceja-Navarro, J. A., Brodie, E. L., Infante, F., Dowd, P. F., & Pain, A. (2015). Draft genome of the most devastating insect pest of coffee worldwide: The coffee berry borer, *Hypothenemus hampei*. *Scientific Reports*, *5*, 12525. https://doi.org/10.1038/srep12525

Warren, R. L. (2016). RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. *The Journal of Open Source Software*, *1*(7), 116. https://doi.org/10.21105/joss.00116

Warren, R. L., Yang, C., Vandervalk, B. P., Behsaz, B., Lagman, A., Jones, S. J., & Birol, I. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience*, *4*, 35. https://doi.org/10.1186/s13742-015-0076-3

Wei, K. H., & Bachtrog, D. (2019). Ancestral male recombination in *Drosophila albomicans* produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. *PLoS Genetics*, *15*(11), e1008502. https://doi.org/10.1371/journal.pgen.1008502

Zaharia, M., Bolosky, W. J., Curtis, K., Fox, A., Patterson, D., Shenker, S., Stoica, I., Karp, R. M., & Sittler, T. (2011). Faster and more accurate sequence alignment with SNAP. *arXiv*, 1111.5572. https://arxiv.org/abs/1111.5572

Zúñiga, G., Cisneros, R., Hayes, J. L., & Macias-Samano, J. (2002). Karyology, geographic distribution, and origin of the genus *Dendroctonus* Erichson (Coleoptera: Scolytidae). *Annals of the Entomological Society of America*, *95*(3), 267–275. https://doi.org/10.1603/0013-8746(2002)095[0267:KGDAOO]2.0.CO;2

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Keeling, C. I., Campbell, E. O., Batista, P. D., Shegelski, V. A., Trevoy, S. A. L., Huber, D. P. W., Janes, J. K., & Sperling, F. A. H. (2021). Chromosome-level genome assembly reveals genomic architecture of northern range expansion in the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae). *Molecular Ecology Resources*, 00, 1–19. https://doi.org/10.1111/1755-0998.13528