Genetic diversity in invasive populations of Argentine stem weevil allows evolution of resistance to biocontrol

Thomas W.R. Harrop[1], Marissa F. Le Lec[1], Ruy Juaregui[2], Shannon Taylor[1], Sarah Inwood[1], John Skelly[1], Siva Ganesh (**sp?**)[2], Rachael Ashby[3], Jeanne Jacobs[3], Stephen Goldson[4], Peter K. Dearden[1]

Goldson’s dissection ppl?

[1] University of Otago
[2] AgResearch Palmerston North
[3] AgResearch Lincoln?
[4] BPRC

Abstract

The abstract should outline the purpose of the paper and the main results, conclusions and recommendations, using clear, factual, numbered statements

- context and need for the work
- approach and methods used
- main results (2-3 points)

Synthesis and applications

- wider implications and relevance to management or policy

Keywords

Naughty weevils, Invasive species, Molecular evolution, ???

Introduction

New Zealand pastures are highly modified landscapes that suffer from severe pest impacts [**???**]. The susceptibility of pasture to pests may be due to low plant and animal diversity, resulting in low biotic resistance to invasive species [**???**]. The Argentine stem weevil (ASW), *Listronotus bonariensis* Kuschel, is a particularly destructive invasive pest of pasture, reaching densities of 700 adults per m2 and causing economic impacts up to NZ$200M per annum [**???**,**??**,**??**]. Conventional, chemical control of ASW is not possible because **xyz** (**ref**). To complement endophyte-based plant resisance [**???**,**??**], the solitary wasp *Microctonus hyperodae* Loan (Hymenoptera: Braconidae) was released for biological control of *L. bonariensis* in 1992. Within three years of its release, parasitism of ASW by *M. hyperodae* reached 90% [**???**], reducing or eliminating damage by ASW [*e.g.* **???**,**??**,**??**]. Once established, failure of biological control is **rare/unprecedented?**, because **xyz** (**ref**).

Although ASW was initially well managed by this system, biological control of ASW by *M. hyperodae* failed after c.14 generations [**???**,**??**,1]. This failure may be the result of adaptation in weevil populations resulting from selection pressure by the parasitoid [**???**,1]. Because ASW reproduces sexually, ASW populations may have greater capacity to evolve than populations of *M. hyperodae*, which reproduces parthenogenetically. Empirical modelling of the ASW–*M. hyperodae* interaction indicated that resistance is inevitable when hosts have more genetic variation than their predator [**???**], *e.g.* because of genetic recombination. Although this work establishes a theoretical pathway for resistance to occur, examples of evolution of resistance to classical biological control have not been reported (**are we sure?**).

Measuring the variation in populations of the host and the parasitoid is required to explain this case of evolution of resistance to biocontrol. We address this by genotyping-by-sequencing of a geographical survey of Argentine stem weevil populations in New Zealand. Our experiments reveal a high proportion of unstructured variation across 12 populations from the North and Sourth islands of New Zealand, consistent with high heterozygosity and gene flow between populations. We found a geographical cline associated with regions of the genome, indicating a level of local adaptation within populations, but no evidence of adaptation at this resolution in parasitised weevils compared to parasitoid-free weevils. Our results show that the amount of genetic variation in NZ populations of ASW is far higher than detected by traditional molecular markers [**???**,**??**],, suggesting that ASW populations evolved resistance via weak selection acting on variants of minor effect that existed before the introduction of *M. hyperodae*. (**Last sentence OTT?**)

Materials and methods

Weevil sampling

From Goldson & co:

- weevil collection details for geographic survey
- collection and processing/dissection details for parasitised *vs.* unparasitised expt

The map was plotted with the ggmap package for ggplot2 [2].

Reduced-representation genome sequencing and processing

From AgResearch:

- details on DNA extraction, GBS pipeline and sequencing

We used a strict processing pipeline to prepare the raw GBS reads for locus assembly. Samples were demultiplexed with zero allowed barcode mismatches to 91–93 b reads, depending on barcode length. Reads were trimmed by searching for adaptors with a minimum match of 11 b. Reads shorter than 80 b after trimming were discarded. All remaining reads were truncated to 80 b to account for unmatched adaptor sequence < 11 b that may have been present at the end of reads. To remove overamplified samples, we calculated the GC content for each library and discarded samples with median read GC > 45%. We followed the recommended steps for optimising parameters [3,4] before assembling loci *de novo* using Stacks [5]. The code we used to process the raw reads, optimise parameters and assemble loci is hosted at github.com/TomHarrop/stacks-asw and github.com/MarissaLL/asw-para-matched.

Genome assembly

To produce the short read dataset, an Illumina TruSeq PCR-free 350bp insert library was generated from DNA extracted from a single, male Argentine stem weevil collected from endophyte-free hybrid ryegrass (*Lolium perenne* × *Lolium multiflorum*) at Lincoln, New Zealand. Library preparation and sequencing were performed by Macrogen Inc. (Seoul, Republic of Korea). A total of 158 GB of 100 b and 150 b paired-end reads were generated from the TruSeq PCR-free library. After removing common sequencing contaminants and trimming adaptor sequences using BBTools [6], the short-read-only genome was assembled with meraculous [7–9]. Reproducible code for assembling the short-read dataset and assessing the assemblies is hosted at github.com/tomharrop/asw-nopcr.

To produce long reads from a single individual, we produced high molecular weight DNA from a single, male ASW collected from Ruakura, New Zealand, using a modified QIAGEN Genomic-tip 20/G extraction protocol [10]. We amplified the DNA using Φ29 multiple displacement amplification (QIAGEN REPLI-g Midi Kit) and debranched the amplified DNA using T7 Endonuclease I (New England Biolabs) according to the Oxford Nanopore Technologies Premium whole genome amplification protocol version WGA_kit9_v1. Amplified DNA was sequenced on 6 R9.4.1 flowcells using a MinION Mk1B sequencer (Oxford Nanopore Technologies). We also extracted high molecular weight DNA from three pools, each of 20 unsexed individuals collected from Ruakura, New Zealand. We sequenced this pooled DNA on 5 R9.4.1 flowcells, following the Genomic DNA by Ligation protocol (SQK-LSK109; Oxford Nanopore Technologies). We removed adaptor sequences from the long reads with Porechop 0.2.4 (github.com/rrwick/Porechop) and assembled with Flye 2.6 [11]. Reproducible code for assembling and assessing the long-read ASW genomes is hosted at github.com/TomHarrop/asw-flye-withpool.

All genome assemblies were assessed by size and contiguity statistics and BUSCO analysis [12]. Redundant contigs were removed from the combined, long read assembly with Purge Haplotigs 0b9afdf [13] using a low, mid and high cutoff of 60, 120 and 190, respectively. We used the Dfam TE Tools Container v1.1

(github.com/Dfam-consortium/TETools) with RepeatModeler 2.0.1 [14] and RepeatMasker 4.1.0 [15] to estimate the repeat content of the long read genomes.

Genome-based analyses, F_{ST}, etc. etc.

- Catalog mapping *e.g.* `bwa mem`

Reproducibility and data availability

Raw sequence data for the ASW genome are hosted at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under accession **TBA**. We used `snakemake` [16] to arrange analysis steps into workflows and monitor dependencies, and `Singularity` [17] to capture the computing environment. Using the code repositories listed in each methods section, the final results can be reproduced from the raw data with a single command using `snakemake` and `Singularity`. The source for this manuscript is hosted at github.com/TomHarrop/asw-gbs-genome-paper.

Results

The Argentine stem weevil genome is repetitive and polymorphic

To construct a reference for genotyping populations of Argentine stem weevils, we produced a draft assembly of the ASW genome. We initially attempted assembly from a single individual using PCR-free, short read sequencing. This resulted in a fragmented assembly with low BUSCO scores (Table 1). *k*-mer analysis on the raw short reads suggested 2.1 polymorphisms per 100 bp and a genomic repeat content of at least 28% in the individual we sequenced (**Supporting Information**). We then attempted to produce a long-read genome assembly using whole-genome amplification (WGA) of high molecular weight (HMW) DNA from a single individual, followed by sequencing on the Oxford Nanopore Technologies (ONT) MinION sequencer. We produced 29.8 GB of quality-filtered reads with an N_{50} length of 9.0 KB. The raw read N_{50} length was reduced by debranching of the amplified DNA by T7 Endonuclease I, which is necessary following multiple displacement amplification (see methods). Assembling the single individual, long read genome resulted in improved contiguity and BUSCO scores (Table 1). Consistent with the raw short read data, we detected an **extreme level (how much?)** of repeats in the single individual, long read genome (Table 1). To improve assembly across long repeats, we produced a second ONT dataset with longer reads from HMW DNA from two pools of 20 individuals each. Sequencing these samples on the MinION sequencer produced a total of 12.0 GB of quality-filtered reads with an N_{50} length of 19.5 KB. Assembling the longer reads generated from the pooled sample alone resulted in a more contiguous genome, but with lower BUSCO scores (Table 1). We constructed a combined, long-read genome using the pooled, long-read dataset for contig construction, and the single-individual, long-read dataset for assembly polishing. This improved the BUSCO scores, but produced a large number of redundant contigs (Table 1), presumably because of the high rate of heterozygosity in the pooled, long-read dataset. Finally, we used the PCR-free, short read sequencing data from a single individual with the Purge Haplotigs pipeline to remove redundant contigs from the combined long read assembly [13]. This resulted in a final draft assembly of 1.1 GB with an N_{50} length of 122.3 kb and a BUSCO completeness of 83.9%.

Table 1. Assembly statistics for the final draft genome and intermediate assemblies. n.d.: not determined.

	Short read	Single individual, long read	Pooled, long read	Combined, long read	Final draft
Assembly length (Gb)	1.3	1.2	1.2	1.7	1.1
N_{50}	53046	4523	2958	5281	2681
N_{50} length (kb)	7.1	74.4	112.6	86.4	122.3
Complete single-copy BUSCOs (%)	32.7	72.2	71.0	69.2	78.8
Complete multiple-copy BUSCOs (%)	17.2	7.5	5.9	17.4	5.1
Repeat content (%)	n.d.	x	x	x	~67.8

Variation in NZ populations of Argentine stem weevil

To measure genetic variation in invasive New Zealand populations of ASW, we collected individuals from 7 sites in the North Island and 5 sites in the South Island of New Zealand (Figure 1A). We genotyped individuals with a modified genotyping-by-sequencing (GBS) protocol [18]. After strict filtering of the raw GBS data, we mapped reads from each individual against our draft genome and used gstacks to assemble loci [5]. For analysis, we removed loci with more than two alleles, minor allele frequency less than 0.05, or missing genotypes in more than 20% of individuals. We also removed individuals missing genotypes at more than 20% of loci. Our final dataset comprised 7–15 individuals per location (total 116), genotyped at 20,445 loci containing 4,363 biallelic sites. The mean observed heterozygosity for these sites ranged from 0.18–0.21 across populations (Figure 1B), and pairwise F_{ST} values between populations ranged from 0.024–0.051 (Figure 1C). Principal components analysis (PCA) of the genotypes revealed overlapping populations of ASW, with 9.4% of total variance explained by the first two components (Figure 1D). These results suggest that NZ populations of ASW are highly heterozygous, but the variation is not highly structured between populations. This is consistent with a large effective population size and high gene flow between populations.

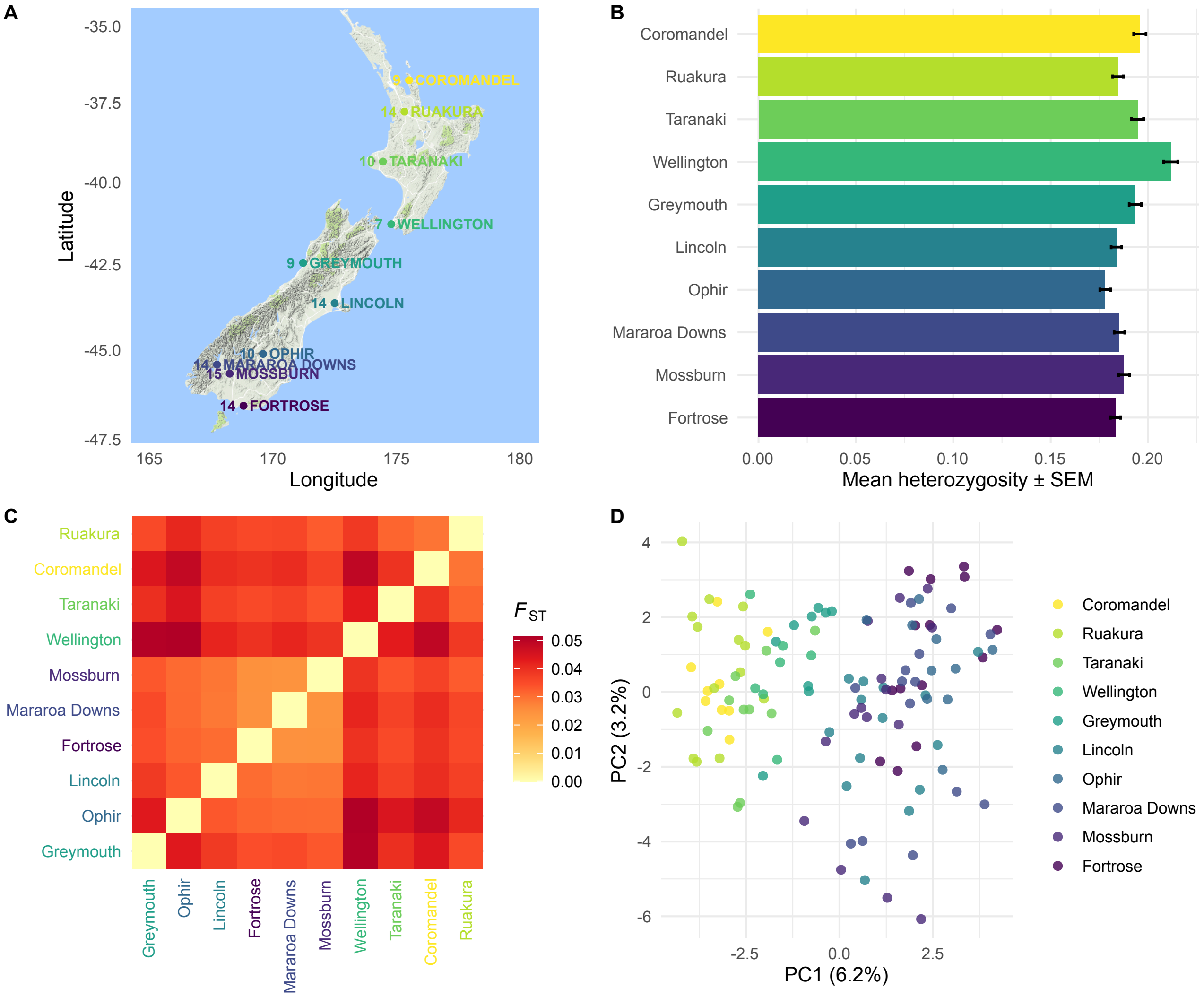


Figure 1. Genetic diversity in NZ populations of Argentine stem weevil. **A** Weevil sampling locations. We collected Argentine stem weevils from 4 locations in the North Island and 6 locations in the South Island of New Zealand. The number of weevils genotyped from each location is shown on the map. Greymouth is on the South Island but on the North side of the Alpine divide. **B** Mean observed heterozygosity (H_O) across 4,363 variant sites for each population. **C** Pairwise F_{ST} values between populations. **D** Pricipal components analysis (PCA) of 116 individuals genotyped at 4,363 biallelic sites. The first two principal components (PC1 and PC2) are shown. The populations overlap on PC1 and PC2, but weevils sampled from higher latitudes tend to have lower scores on PC1. PC1 and PC2 together explain 9.4% of variance in the dataset, indicating a high level of unstructured genetic variation in weevil populations.

Genetic variation is not associated with parasitism by a biocontrol agent

To detect variation associated with parasitism by *Microctonus hyperodae* (*i.e.* selection exerted by the biocontrol agent), we genotyped weevils that had also been tested for the presence of a parasitoid larva. These weevils were collected from **Lincoln, New Zealand?** and **Ruakura, New Zealand?**, because of the decline in parasitism rate recorded at these locations [1]. After filtering and assembly, we genotyped **X** parasitised weevils and **Y** weevils without a detected parasitoid at the same 20,445 loci used for the geographical samples, which contained 4,579 biallelic sites in this second dataset. **We did not detect SNPs that were associated with the presence of a parasitoid larva, although we were able to detect SNPs that were associated with the location the weevil was collected. (Figure to show this).** This suggests that the developing resistance of the weevil to biocontrol [1] is not related to within-population genetic variation that allows some weevils to avoid parasitism or its effects.

Genetic variation between NZ weevils associates with a geographical cline

Although geographic location explains a small proportion of the genetic variance between ASW individuals, parasitism rates vary at different sites in NZ (**citation?**) We addressed this by testing whether differences between populations were related to selection in defined regions of the genome. We used discriminant analysis of principal components to find genetic variability associated with differences between populations (DAPC; **???**). The major linear discriminant, which explains 97.7% of between-population variations, separates populations from North and South of the Alpine divide (Figure 2A), although admixture was evident in all populations (Figure 2B). **Regions of the genome associated with this.**

New Zealand population of Argentine stem weevils is large and diverse, with multiple introductions

[No] evidence for reduced diversity since/on introduction of invasive populations to NZ. We can't do historical demographics / N_e because of the distance between GBS loci.

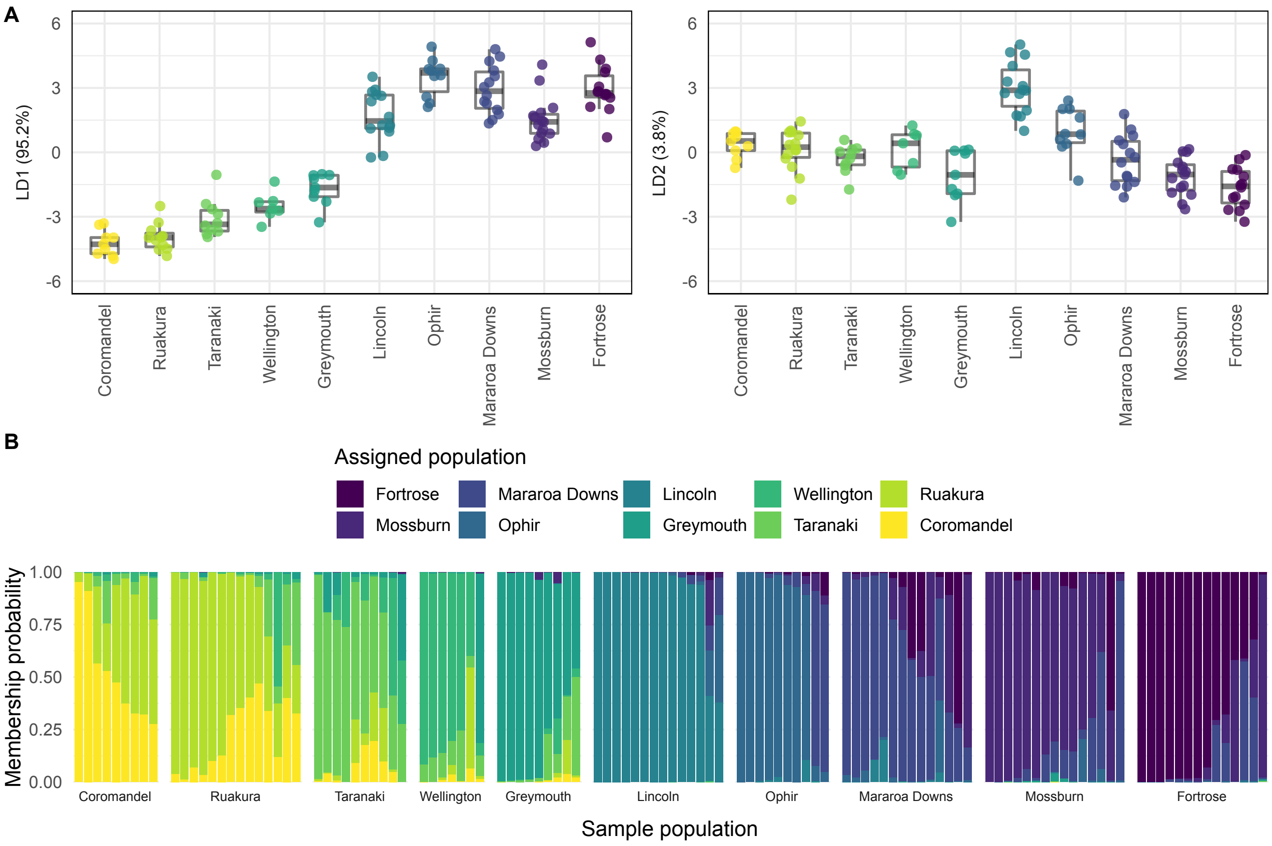
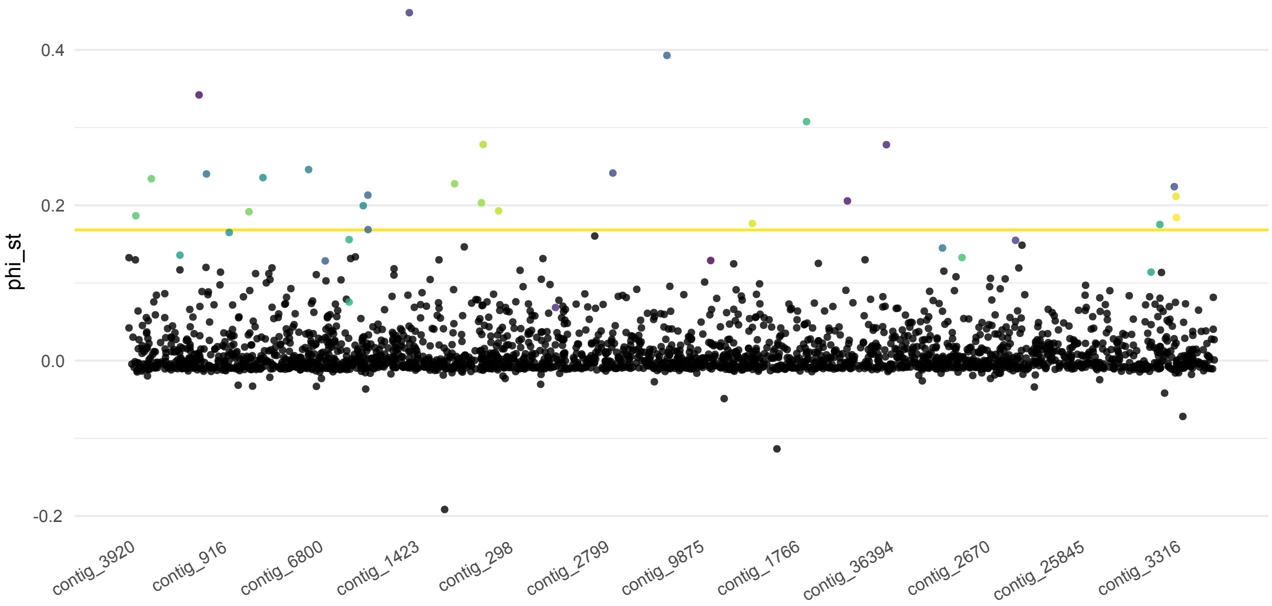
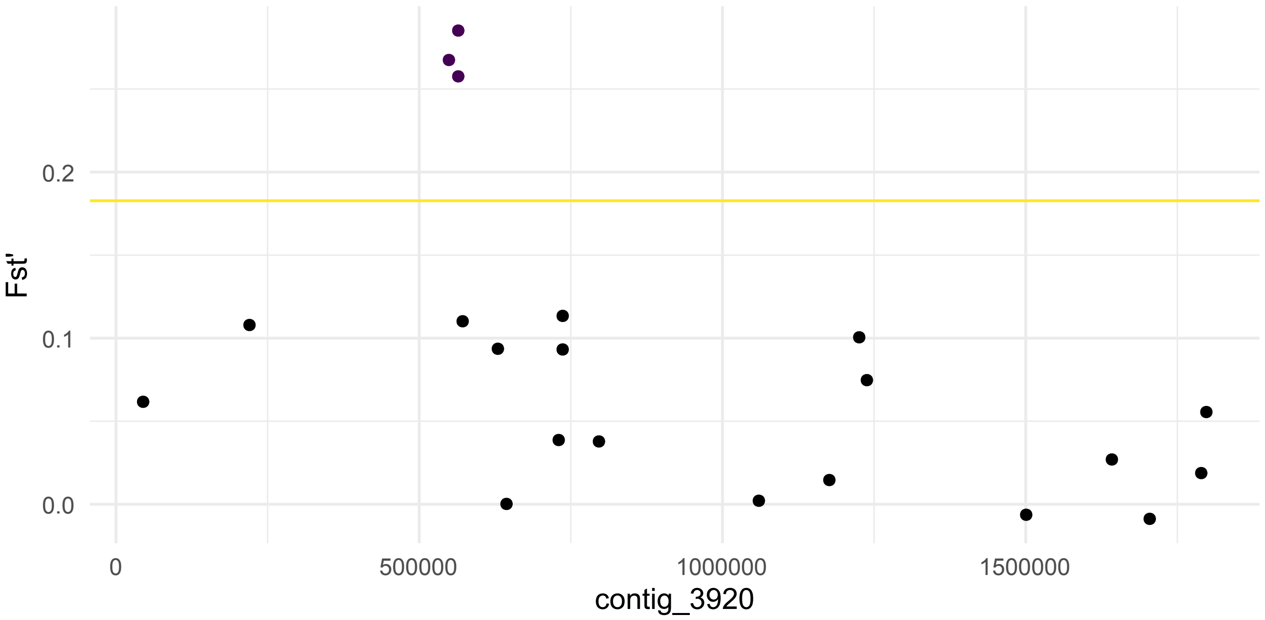


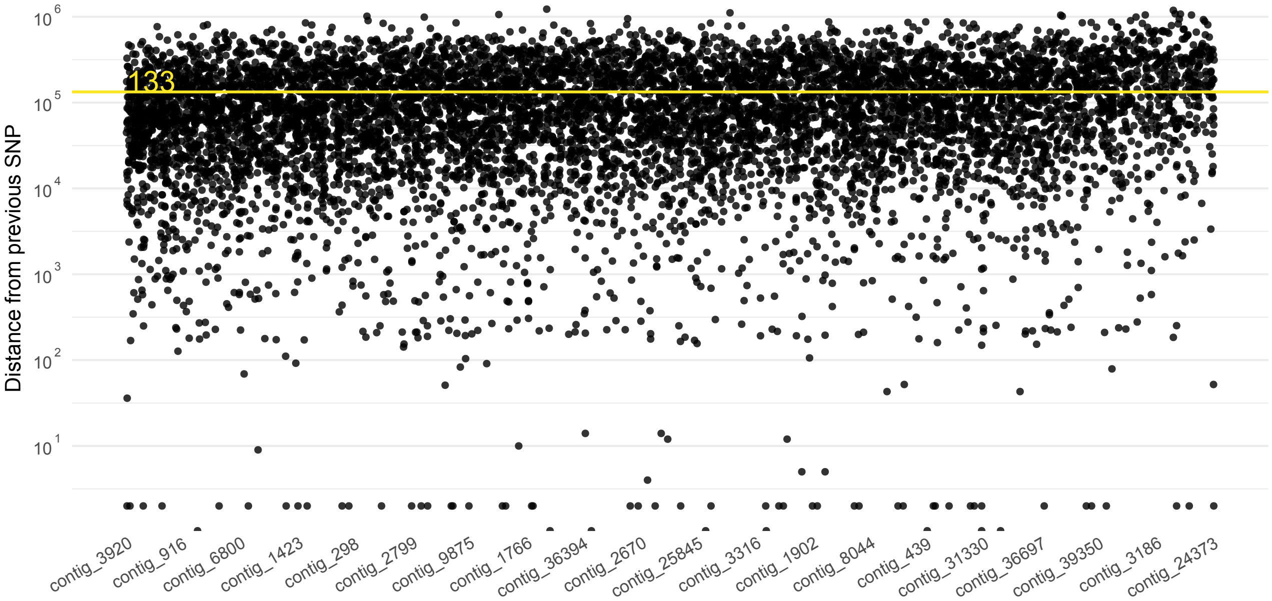
Figure 2. **A** Discriminant analysis of principle components (DAPC) of 116 individuals genotyped at 4,363 biallelic sites. Linear discriminant 1 (LD1) explains 97.7% of between-group variability. Individuals from North of the Alpine divide (see Figure 1) have negative coordinates on LD1, whilst individuals from South of the Alpine divide have positive coordinates. **B** Membership probabilities for each individual. All populations contain individuals with high posterior probabilities of membership to other populations, consistent with admixture (**?**).



Regions of low variation (high F_{ST}) in the stem weevil genome. Needs to be redone with Bayescan.



Regions of low variation (high F_{ST}) on the longest contig. Doesn't work with large distances between loci.



Distance between successive SNPs.

Discussion

Short read assembly failed for this genome because of the extreme repeat content. The final draft assembly had a repeat content of **67.8%** (Table 1), with a maximum repeat size of 17.7 kb and a repeat N_{50} length of 485 bp. The non-repetitive regions had an N_{50} length of 1066 bp. The heterozygosity in weevil populations and lack of an inbred, laboratory strain made pooling individuals for sequencing undesirable. Our assembly strategy of contig construction with the longest reads, followed by assembly polishing with long reads from a single individual, and then redundant contig removal with PCR-free short reads from another single individual allowed us to improve the contiguity and completeness of the stem weevil genome (Table 1). Our final genome is draft quality and we expect gaps in the assembly at larger repeat regions that were not sufficiently covered by long reads.

We were unable to estimate historical demographics because the GBS markers were too sparse in the genome to detect runs of homozygosity. Whole-genome resequencing, which is now widely available at low cost and high throughput, would enable these analyses.

Authors' contributions

Acknowledgements

Data availability

References

1. Tomasetto, F.; Tylianakis, J.M.; Reale, M.; Wratten, S.; Goldson, S.L. Intensified agriculture favors evolved resistance to biological control. *Proceedings of the National Academy of Sciences* **2017**, 201618416. doi: 10.1073/pnas.1618416114.

2. Kahle, D.; Wickham, H. Ggmap: Spatial Visualization with ggplot2. *The R Journal* **2013**, *5*, 144. doi: 10.32614/RJ-2013-014.

3. Paris, J.R.; Stevens, J.R.; Catchen, J.M. Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution* **2017**, *8*, 1360–1373. doi: 10.1111/2041-210X.12775.

4. Rochette, N.C.; Catchen, J.M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* **2017**, *12*, 2640. doi: 10.1038/nprot.2017.123.

5. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An analysis tool set for population genomics. *Molecular Ecology* **2013**, *22*, 3124–3140. doi: 10.1111/mec.12354.

6. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*; Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), 2014;Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

7. Chapman, J.A.; Ho, I.; Sunkara, S.; Luo, S.; Schroth, G.P.; Rokhsar, D.S. Meraculous: De Novo Genome Assembly with Short Paired-End Reads. *PLoS ONE* **2011**, *6*. doi: 10.1371/journal.pone.0023501.

8. Chapman, J.A.; Ho, I.Y.; Goltsman, E.; Rokhsar, D.S. Meraculous2: Fast accurate short-read assembly of large polymorphic genomes. *arXiv:1608.01031 [cs, q-bio]* **2016**. Retrieved from http://arxiv.org/abs/1608.01031.

9. Goltsman, E.; Ho, I.; Rokhsar, D. Meraculous-2D: Haplotype-sensitive Assembly of Highly Heterozygous genomes. *arXiv:1703.09852 [q-bio]* **2017**. Retrieved from http://arxiv.org/abs/1703.09852.

10. Harrop, T. HMW DNA extraction for insects. **2018**. doi: 10.17504/protocols.io.pnwdmfe.

11. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **2019**, 1. doi: 10.1038/s41587-019-0072-8.

12. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. doi: 10.1093/bioinformatics/btv351.

13. Roach, M.J.; Schmidt, S.A.; Borneman, A.R. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **2018**, *19*, 460. doi: 10.1186/s12859-018-2485-7.

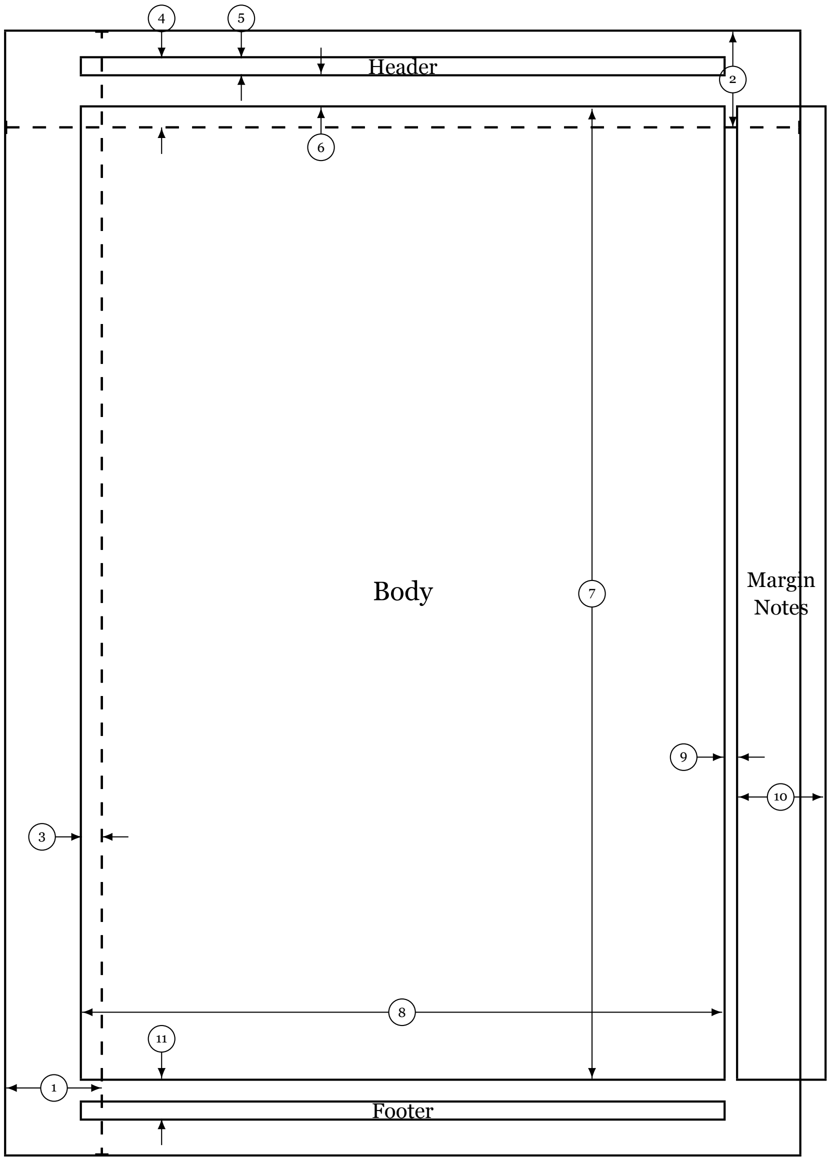
14. Smit, A.F.A.; Hubley, R. RepeatModeler Open-1.0 2015.

15. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2015.

16. Köster, J.; Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520–2522. doi: 10.1093/bioinformatics/bts480.

17. Kurtzer, G.M.; Sochat, V.; Bauer, M.W. Singularity: Scientific containers for mobility of compute. *PLOS ONE* **2017**, *12*, e0177459. doi: 10.1371/journal.pone.0177459.

18. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE* **2011**, *6*, e19379. doi: 10.1371/journal.pone.0019379.



1	one inch + \hoffset	2	one inch + \voffset
3	\oddsidemargin = -15pt	4	\topmargin = -52pt
5	\headheight = 12pt	6	\headsep = 25pt
7	\textheight = 731pt	8	\textwidth = 483pt
9	\marginparsep = 11pt	10	\marginparwidth = 65pt
11	\footskip = 30pt		\marginparpush = 5pt (not shown)
	\hoffset = 0pt		\voffset = 0pt
	\paperwidth = 597pt		\paperheight = 845pt