# Genetic variation associated with a geographical cline in New Zealand populations of Argentine Stem Weevil

Rachael Ashby3  
Peter K. Dearden1  
Stephen Goldson4  
Thomas W.R. Harrop1  
Sarah Inwood1  
Jeanne Jacobs3  
Ruy Juaregui2  
Marissa F. Le Lec1  
Shannon Taylor1

Alphabetical for now! Sample collectors, goldson’s dissection crew?

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

**weevils are naughty**

## Introduction

## Materials and methods

### Collections *etc*. for GBS

Weevils were collected from blah

### GBS sequencing and processing

DNA was extracted blah

### Genome assembly

An **Illumina TruSeq PCR-free?** library was generated from DNA extracted from a single, male Argentine stem weevil (**moar deets**). A total of **X GB** of paired-end 100 b and paired-end 150 b reads were generated from the TruSeq PCR-free library. After removing common sequencing contaminants and trimming adaptor sequences using BBMap (**ref**), a short-read-only genome was assembled with meraculous [**ref**].

WGA of single indiv

ONT stuff

Assembly tricks

Analysis (BUSCO, RepeatModeller)

### Genome-based analyses *e.g.* *F*ST

Catalog was mapped with bwa mem etc.

### 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**. The code we used to assemble the ASW genome is hosted at [github.com/TomHarrop/asw-flye-withpool](https://github.com/TomHarrop/asw-flye-withpool). We used snakemake (Köster & Rahmann, 2012) to arrange analysis steps into workflows and monitor dependencies, and Singularity (Kurtzer, Sochat, & Bauer, 2017)make to capture the computing environment. The final results and all intermediate steps 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](https://github.com/TomHarrop/asw-gbs-genome-paper).X

## Results

### Variation in NZ populations of Argentine stem weevil

To measure the variation in NZ populations of ASW, we collected samples from X Y Z locations. We found lots of sweet variation.

### The Argentine stem weevil genome

To map the variation in NZ population of ASW to regions of the genome, we constructed a draft assembly. We initially attempted to assemble a genome from a single individual using short-read sequencing. This resulted in a fragmented assembly with poor BUSCO scores (**genome\_table**). Because of the high heterozygosity in the single-individual short-read library (**SI**), we 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 **X GB** of reads with an *N*50 length of **Y KB**. The low read *N*50 length is caused by branching of the genomic DNA during WGA by Φ29 DNA polymerase [**ref**]. Assembling the single-individual, long-read-only genome resulted in improved contiguity and BUSCO scores (**genome\_table**). We detected an extreme level of repeats in the single-individual, long-read-only genome (**genome\_table**). To produce a second ONT dataset with longer reads, we extracted HMW DNA from a pool of **x** individuals. Sequencing this DNA on the MinION sequencer produced **X GB** of reads with an *N*50 length of **Y KB**. We constructed a draft genome using the pooled, long-read dataset for contig construction, and the single-individual, long-read dataset for assembly polishing. This resulted in a more contiguous assembly, but a large number of redundant contigs (**genome\_table**), presumably because of the high rate of heterozygosity in the pooled, long-read dataset. After using the short-read, single-individual sequencing data with the purge\_haplotigs pipeline to remove redundant contigs (**ref**), we produced a final, draft genome of **X GB** (**genome\_table**). **Something about the repetitiveness**. We used this draft genome for subsequent analysis.

### Variation associates with a North-South cline

etc. etc.

## Discussion

## Authors’ contributions

## Acknowledgements

## Data availability

## References

Köster, J., & Rahmann, S. (2012). Snakemakea scalable bioinformatics workflow engine. *Bioinformatics*, *28*(19), 2520–2522. doi: [10.1093/bioinformatics/bts480](https://doi.org/10.1093/bioinformatics/bts480)X

Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. *PLOS ONE*, *12*(5), e0177459. doi: [10.1371/journal.pone.0177459](https://doi.org/10.1371/journal.pone.0177459)X