



Photos OK.
Download with
QR code.

A Method to Map Structural Variation



Luca Comai ¹, Kirk Amundson ¹, Benny Ordóñez ¹, Xin Zhao ¹, Guilherme T. Braz ², Jiming Jiang ², and Isabelle M. Henry ¹

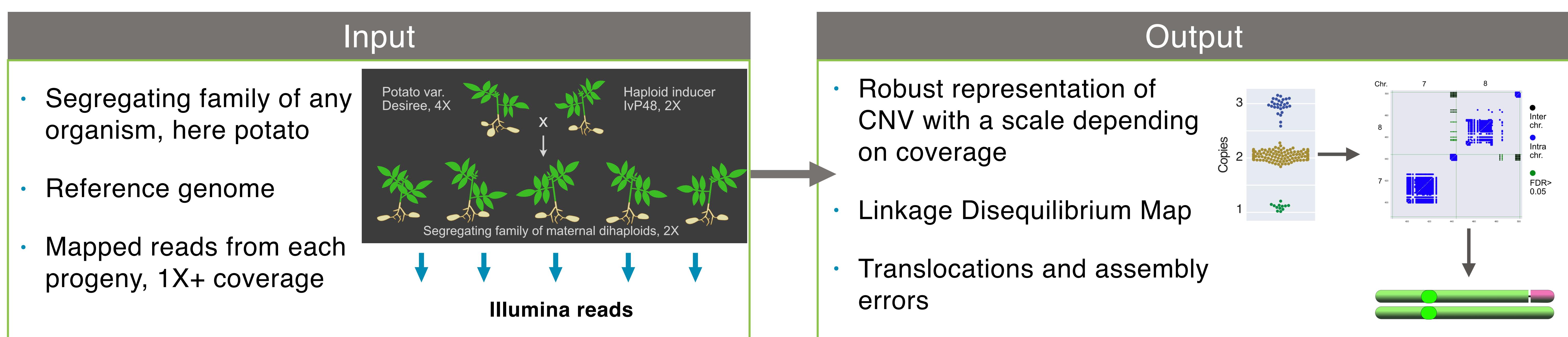
¹ Department of Plant Biology and Genome Center, University of California Davis, Davis, CA 95616

² Department of Plant Biology, Michigan State University, East Lansing, MI 48824

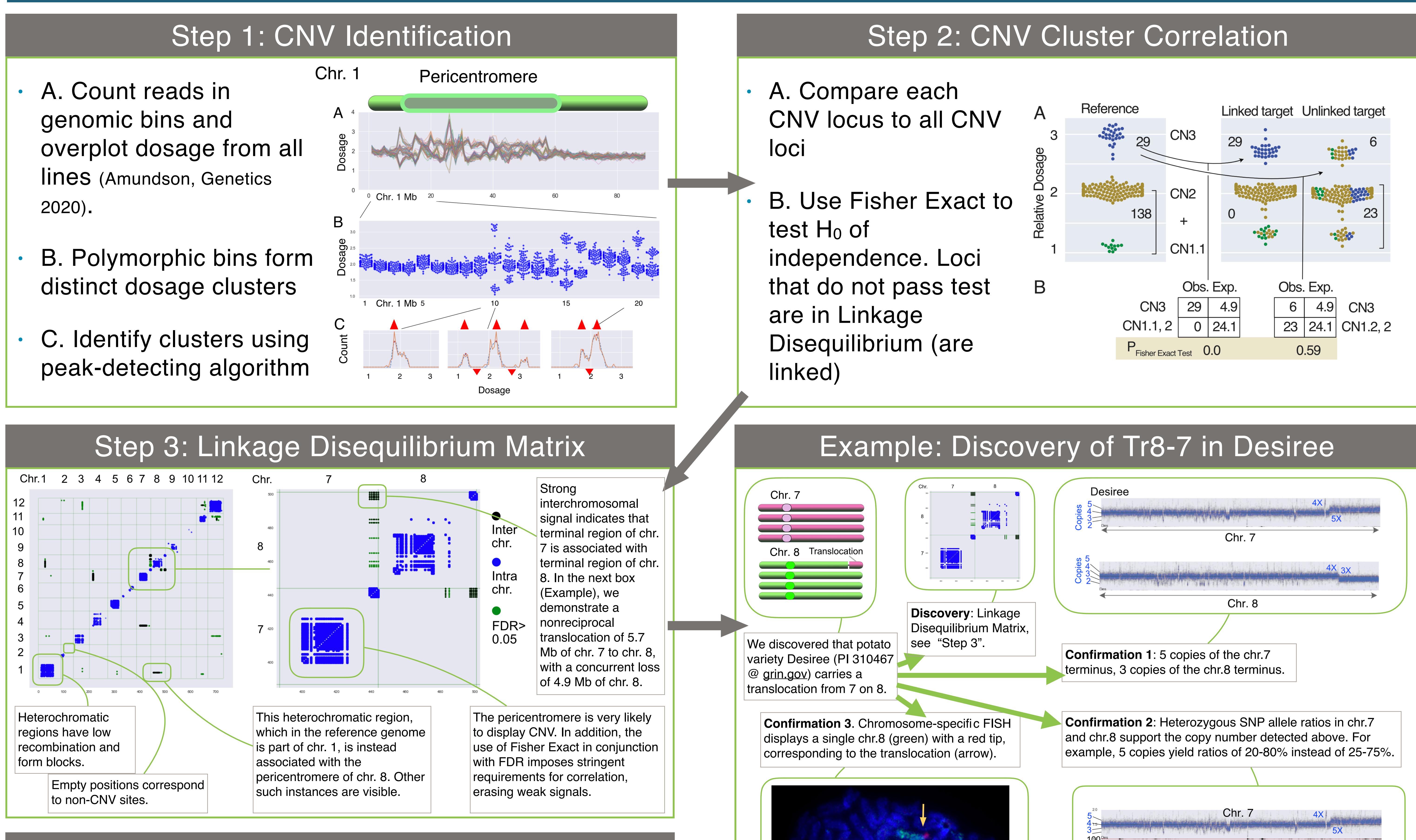
Abstract

Genomic Structural variation (SV) is common and has a profound effect on the phenotype of individuals. Chromosomal translocations are an important type of SV, but also a common spurious effect of defective genome assemblies. In theory, they can be identified by detecting unexpected linkage, by sequencing unexpected junctions, and by fluorescent in situ hybridization (FISH). In practical terms, however, identifying and characterizing a translocation is not simple, particularly in the absence of prior evidence pointing to its location. We study structural genomic variation and developed a method to identify translocations that is based on low-pass Illumina sequencing of related segregating individuals. By comparing sequence coverage in individuals such as siblings, we can cluster genomic dosage states at SV loci. We then detect unexpected linkage through covariance analysis of these copy number variable loci. We demonstrate this method using a population of dihaploid individuals produced by haploid induction crosses in cultivated, autotetraploid potato (*Solanum tuberosum*). In this background, we document frequent nonreciprocal chromosomal translocations as well as possible assembly errors. The deleterious effect of unbalanced translocations in potato is likely buffered by polyploidy. Furthermore, purging of deleterious SV by meiosis is hindered by clonal propagation.

Overview



Method



The Nitty-Gritty

Steps

0. Map reads to reference genome using DESeq to yield SAM files.
- 1a. Derive read counts per bin using BIN-by-SAM. Mean bin count should be >100. Bin size depends on coverage: large scale CNV can be detected using 1X coverage and 1 Mb bin. Divide bin count by population or parent mean count to standardize dosage, setting mean value corresponding to ploidy.
- 1b. Detect dosage clusters using the Python `peakutils` pkg in Jupyter Notebook*. For every bin with variable copy number, each individual is assigned a copy state.
2. Test independence of copy states using Fisher Exact in *ad hoc* Python Jupyter Notebook. Correct for multiple testing error using 5% False Discovery Rate according to Benjamini and Hochberg.
3. Draw graphs using Pandas-Matplotlib-Seaborn. Identify candidate translocations by visual inspection of matrix or by filtering output tables for interchromosomal LD.

*A similar pipeline is available in R. The Python Notebooks and this posters are available (QR code)

Conclusions

We developed a method to map CNV in potato. It has proven effective by identifying a chr7-8 translocation in Desirée and a chr1-4 translocation in var. Alca Tarma (not shown). We have also detected sex-specific CNV in a diploid persimmon family. The method relies on easily obtained inputs, and can work with low coverage data. It can be implemented in 1-2 hours on a personal computer, although very small bins (<100,000 kb) may require a server. Its use should facilitate the identification of translocations in different species.

Funding:
NSF-PGRP
IOS-1444612