UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

MEDICC2

Minhyuk Park

3/25/21

Citation

- Petkovic, M., Watkins, T., et al. Whole-genome doubling-aware copy number phylogenies for cancer evolution with MEDICC2 (preprint)
- Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, et al. (2014) Phylogenetic Quantification of Intra-tumour Heterogeneity. PLOS Computational Biology 10(4): e1003535. https://doi.org/10.1371/journal.pcbi.1003535

Background

- Copy Number Aberration (CNA):
  - Gains and losses of segments of the genome
  - Cancer evolutionary process
- Whole Genome Doubling (WGD):
  - The entire genome being duplicated
  - Can lead to tetraploidy (start from diploid and double to tetraploid)
- These come about as part of the evolutionary process of cancer genomes:
  - They contain signal about cancer phylogeny

Minimum Event Distance Problem (MED)

- Given two Copy Number Profiles (CNP), get the number of gains and losses on arbitrary lengths of segments to transform one CNP to another

Problem Statement

- Input:
  - Set of copy number profiles
- Output:
  - Tree topology, internal node copy number profiles, and branch lengths
- Optimizing for:
  - Minimum Event Distance under Whole Genome Doubling

WGD and Cancer Phylogeny Inference

- No methods so far that could infer phylogeny from genomes that include WGD event:
  - Distinguishing between WGD events and multiple gain/loss is hard
  - No evolutionary model exists that identify WGD events in cancer phylogeny
  - MEDICC was limited to arbitrary lengths and not whole genomes
- MEDICC2:
  - Minimum Event Distance for Intra-tumor Copy-number Comparisons
  - Explicitly models clonal/subclonal WGD events

MEDICC2 Overview

- Start with a set of Copy Number Profiles (CNP)
- Get a minimum event distance matrix on all pairs of the input genome
- Neighbor Joining on the distance matrix to get a tree
- Infer ancestral CNPs on the internal nodes
- Assign branch lengths using minimum evolution distance of each parent and child clade

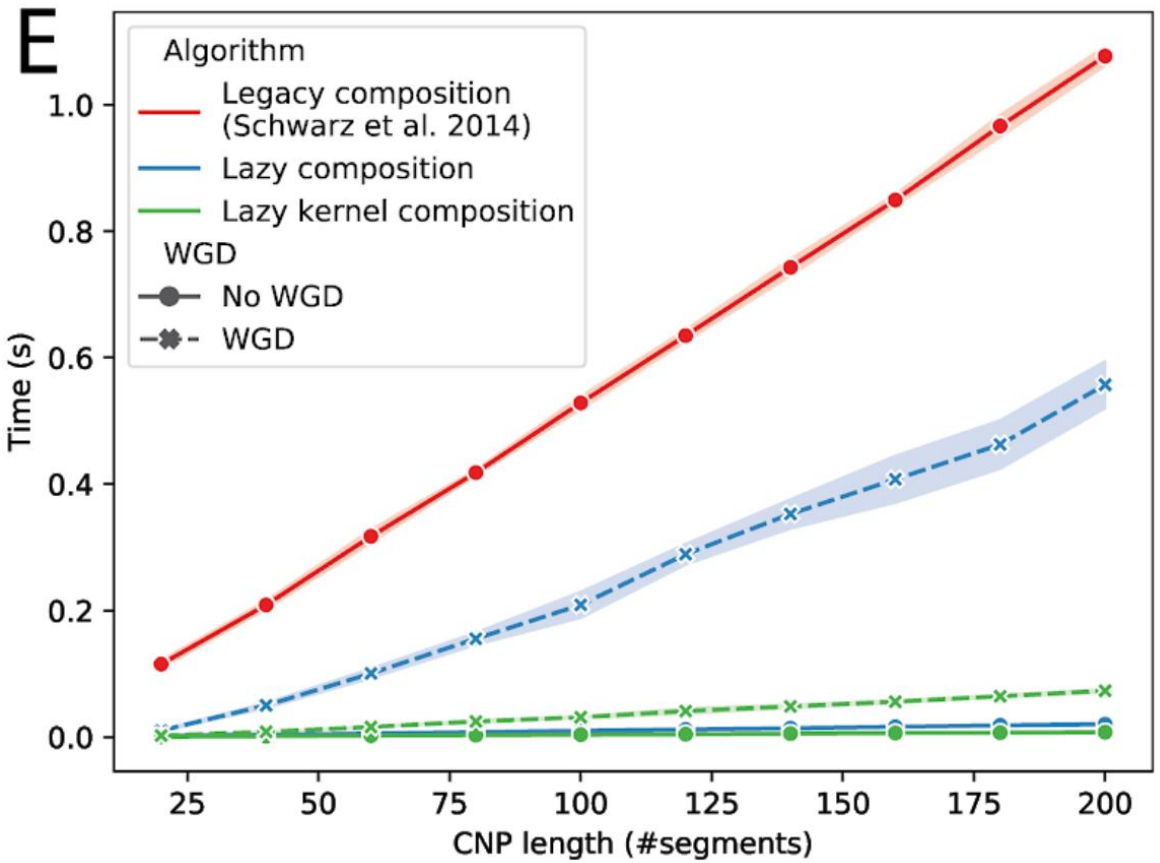
Finite-State Transducer

- Finite State Machine that has:
  - A read tape
  - A write tape
- To solve mimimum event distance:
  - Input and output alphabet are copy numbers
  - Internal states and their transitions represent all possible transformations from input to output

More on Solving MED-WGD

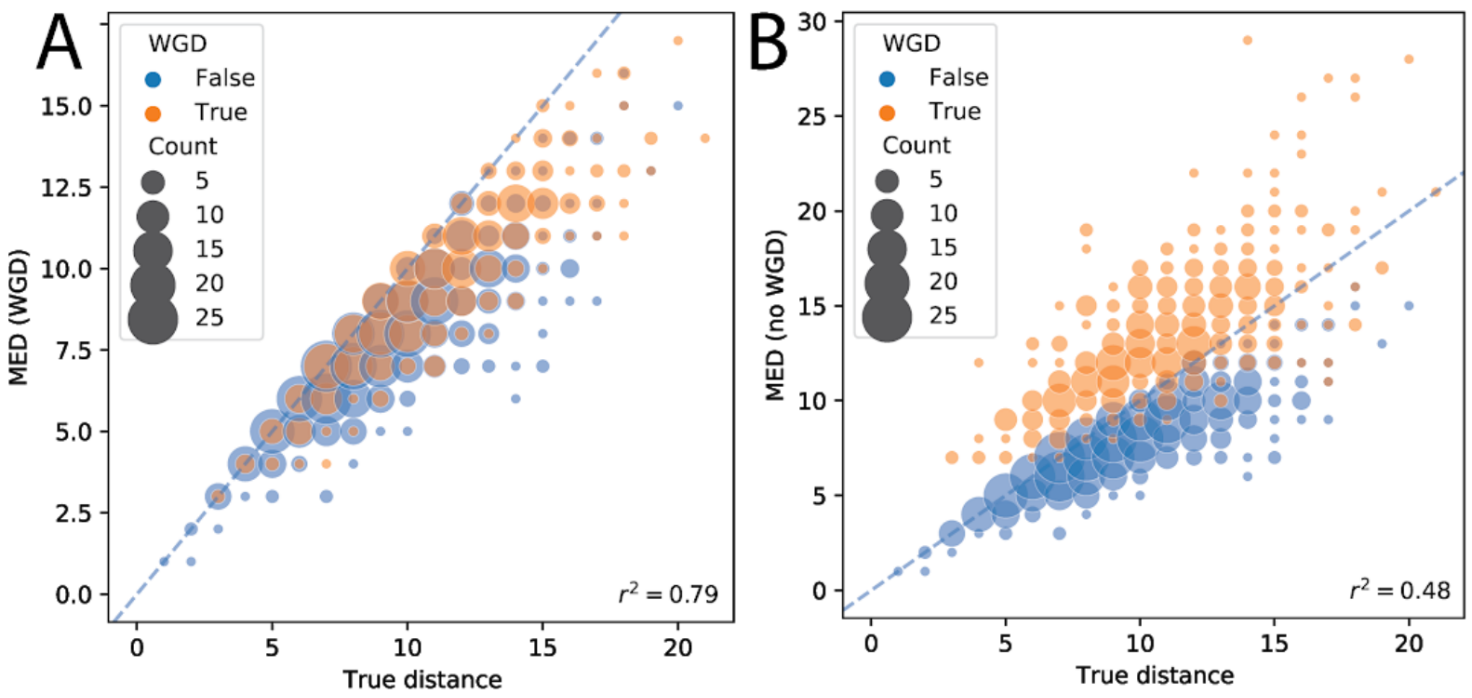
- Using a finite-state transducer with input and outputs being the copy numbers:
  - Let's call this T[x,y] where x is the input sequence and y is the output sequence
  - Composition of gets us sequential events
- $T_{MED-WGD} = T_{LOH} \circ T_{WGD} \circ T_L \circ T_G$
  - We know that all the loss of heterozygosity events can be considered first
  - WGD should be considered right away to reduce non-determinism
  - Then all the losses
  - Then all the gains
  - Getting the shortest distance through $T_{MED-WGD}[x, y]$ should give us the minimum event distance accounting for whole genome doubling from x to y

Runtime Comparison on Simulated Data



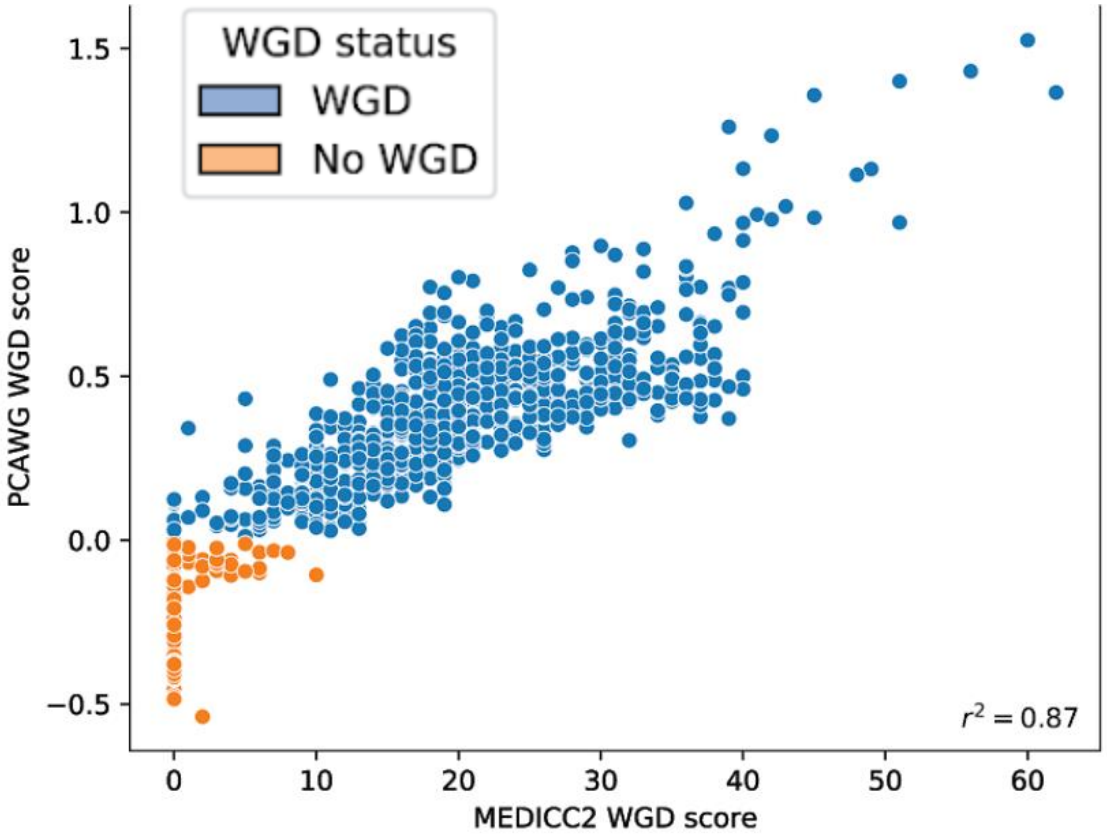
- The authors implemented a lazy computation of shortest distance through the finite-state transducers
- Notice how much faster the lazy version is compared to the original legacy version

Accuracy Comparison on Simulated Data



- This is on simulated data with WGD
- Left takes into account WGD while the right does not
- Not taking into account WGD makes MEDICC2 overestimate the true distances

Classification Comparison on Biological Data I



- PCAWG is the published golden standard
- MEDICC2 is not a classifier:
  - Take the difference of scores between the MED while ignoring WGD and MED while accounting for WGD
  - If this difference is high, then there must have been WGD

Classification Comparison on Biological Data II

	MEDICC	
PCAWG	**no WGD**	**WGD**
no WGD	1937	23
WGD	8	810

- PCAWG is the published golden standard
- MEDICC2 is not a classifier:
  - Take the difference of scores between the MED while ignoring WGD and MED while accounting for WGD
  - If this difference is high, then there must have been WGD

Grainger College of Engineering

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN