

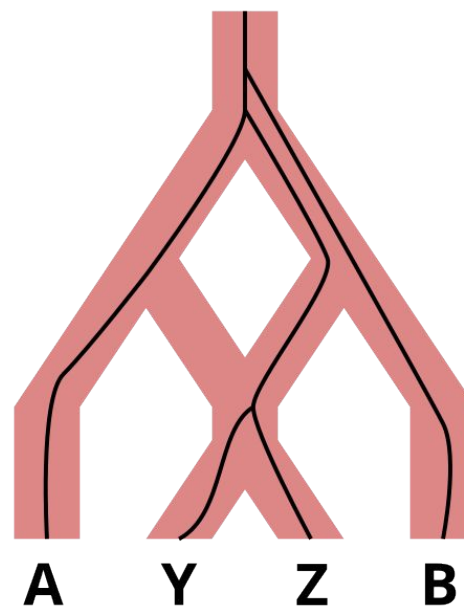
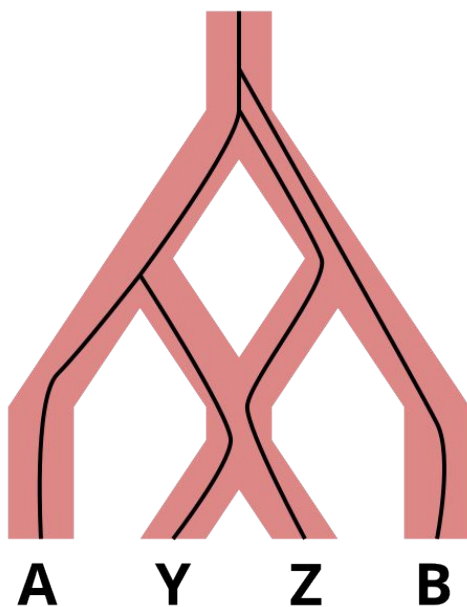
Fast and flexible inference of allopolyploid phylogenies

Huw Ogilvie

Botany 2023, Boise, Idaho

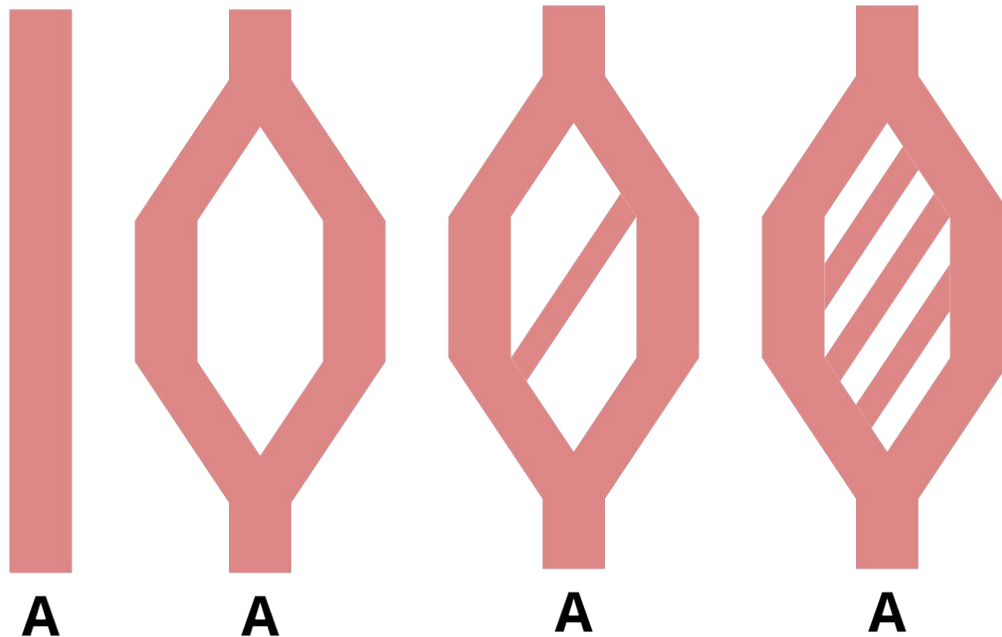
Multispecies network coalescent: multilocus inference

- **Two-step** (first gene trees, second species phylogeny) or **joint**
- **Free** recombination between loci, **no** recombination within loci
- **Deep coalescence** causes gene tree discordance, even absent reticulation
- **Panmixia** within each branch



Searching for the optimal species network

- The number of possible **tree topologies** grows **hyper-exponentially** with the number of taxa
- The number of **network topologies** is **infinite** with **any** taxa
- Existing algorithms to search the space of topologies **mix poorly**
- Existing **likelihood/scoring** functions are **slow**



The Problem of Allopolyploidy

- One type of hybrid speciation
- More common in plants
- **Two** copies are maintained of each chromosome from both parental species
- Multiple pathways exist that lead to allopolyploids

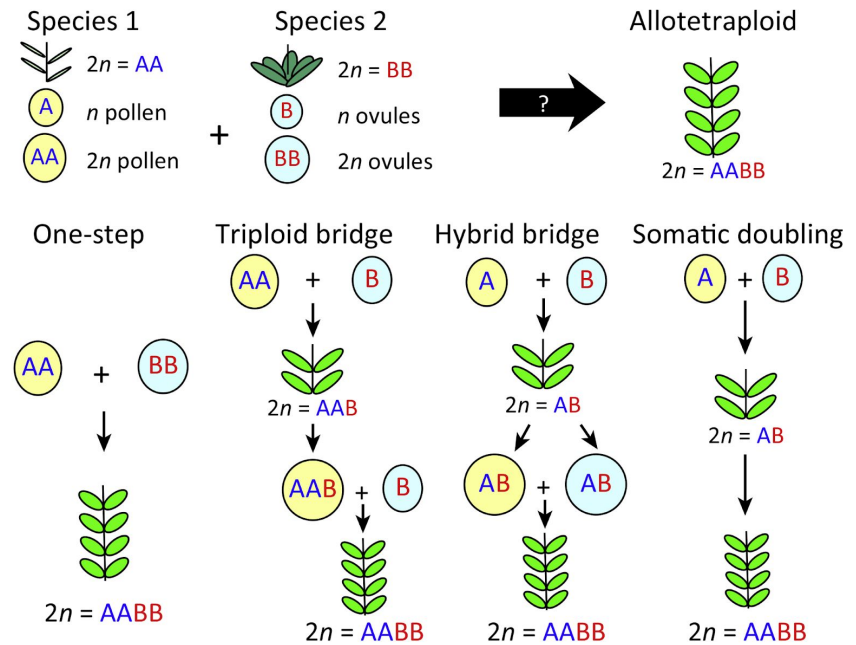


Figure from Mason & Pires (2016),
 “Unreduced gametes: meiotic mishap or
 evolutionary mechanism?” *Trends in Genetics*

Allopolyploidy and Recombination

- In allopolyploids, meiotic recombination may **not** occur between chromosomes derived from different parent species
- Sets of chromosomes from different parent species maintained as **subgenomes**

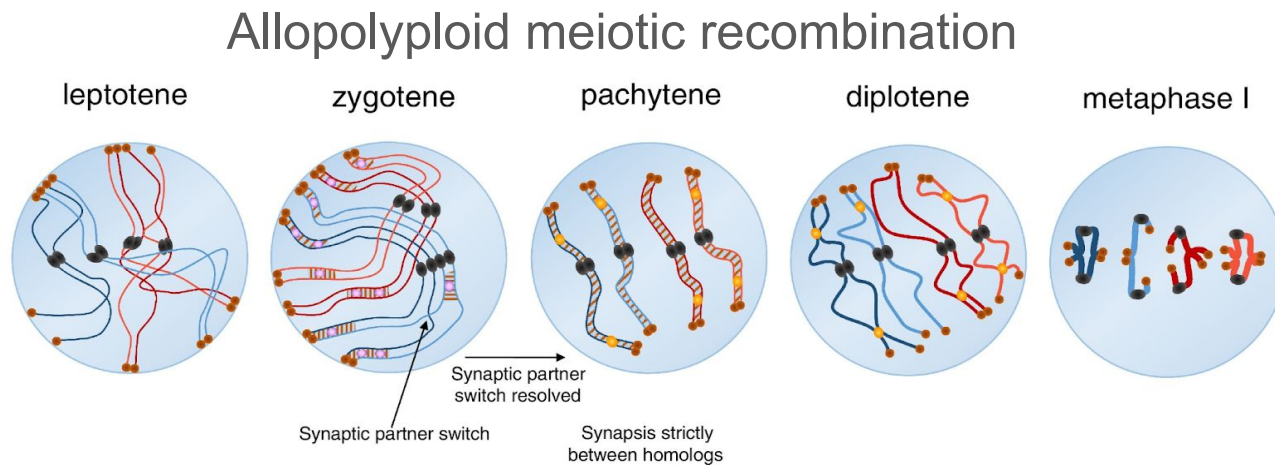
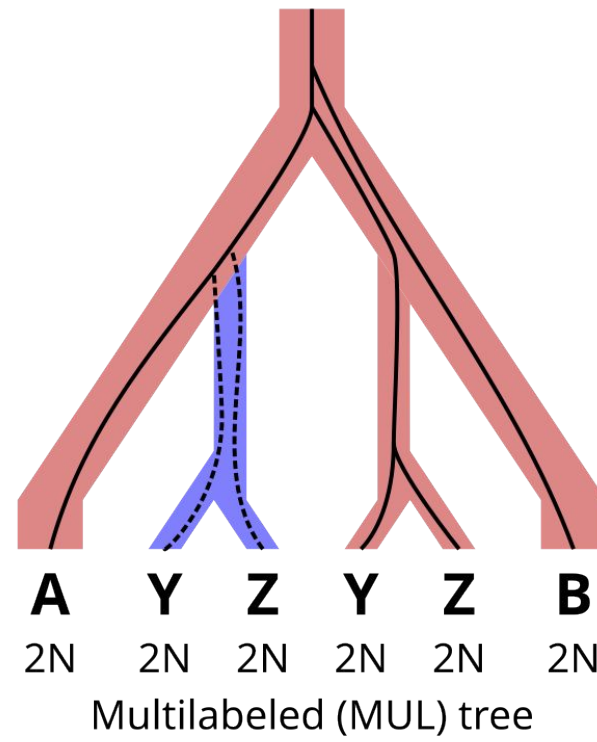
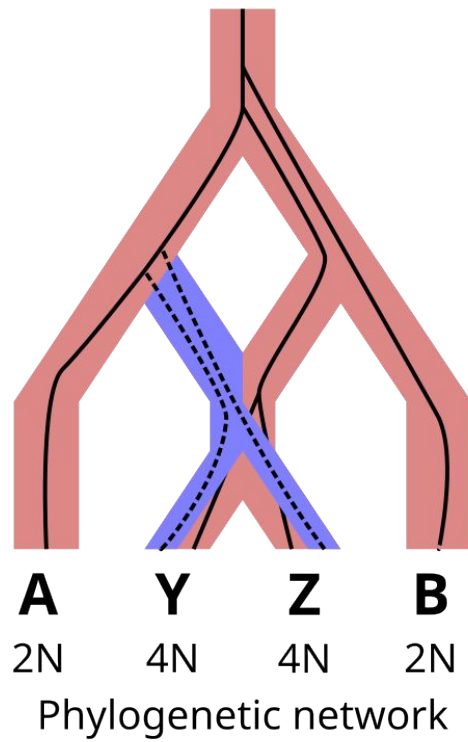


Figure adapted from Lloyd & Bomblies (2015),
“Meiosis in autopolyploid and allopolyploid
Arabidopsis” *Current Opinion in Plant Biology*

We can take advantage of subgenome isolation!

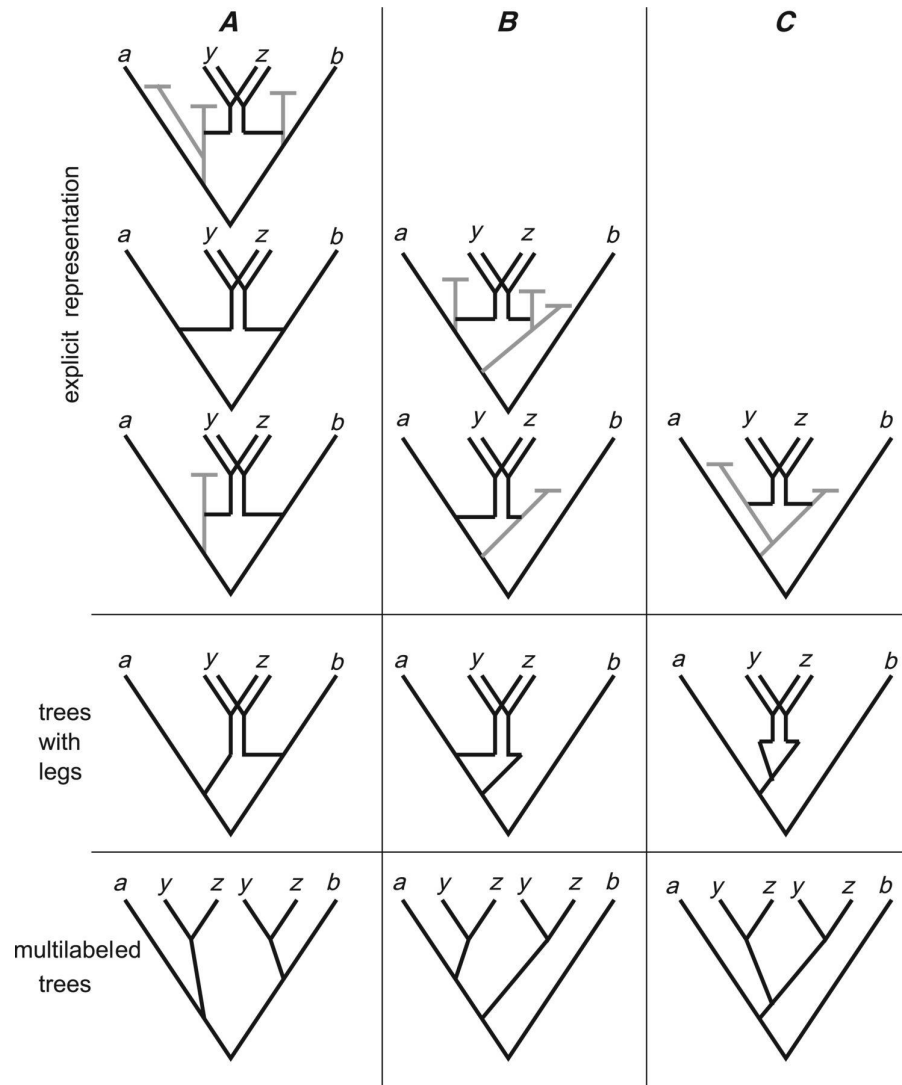
- Only a **finite** number of valid networks (when ploidy is **known**)
- Transform to MUL tree and use **faster** tree **likelihood/scoring** functions



AlloppNET (2013)

- Joint inference (Bayesian) of gene trees and network
- Uses *BEAST to compute likelihood of MUL tree
- Exactly two diploid species and one tetraploid clade required
- Accounts for **some** kinds of incomplete sampling
- Mau, Newton & Larget (1999) move to alter tetraploid clade
- New move to alter “legs”

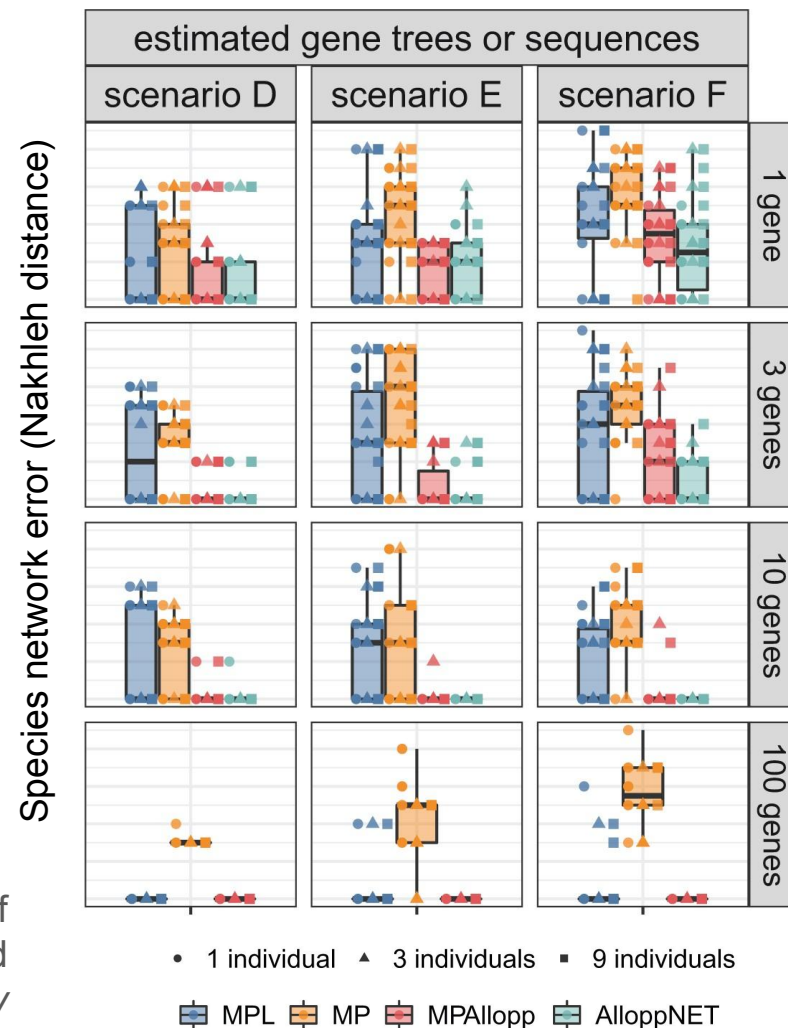
Jones, Sagitov & Oxelman (2013), "Statistical Inference of Allopolyploid Species Networks in the Presence of Incomplete Lineage Sorting" *Systematic Biology*



MPAllopp (2022)

- Two-step inference (gene trees as input)
- **100x faster** than AlloppNET
- Scales to larger datasets
- Uses the maximum parsimony tree scoring function from Maddison (1997)
- **Minimize Deep Coalescence**
- Searches the space of networks using an SPR-like move
- **Fixed polyploid clade membership** (iterate over all possible combinations)
- Does **not** support “**bubbles**” (cycle with diameter = 2 edges)

Yan et al. (2022) "Maximum Parsimony Inference of Phylogenetic Networks in the Presence of Polyploid Complexes" *Systematic Biology*



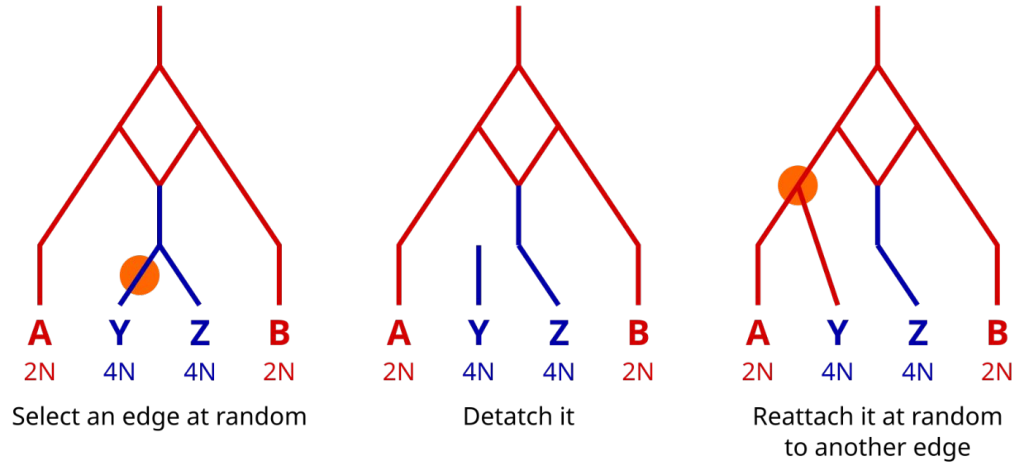
MP-SUGAR (now)

- Reimplemented in **PhyNetPy**
 - A brand new, user-friendly, multi-method framework for inference of trees and networks
 - <https://phylogenomics.rice.edu/html/phyNETpy.html>
- 100x faster than MPAllopp
- **10,000x faster** than AlloppNET
- New move that **searches entire network space**
- **No requirement** to sample diploid relatives

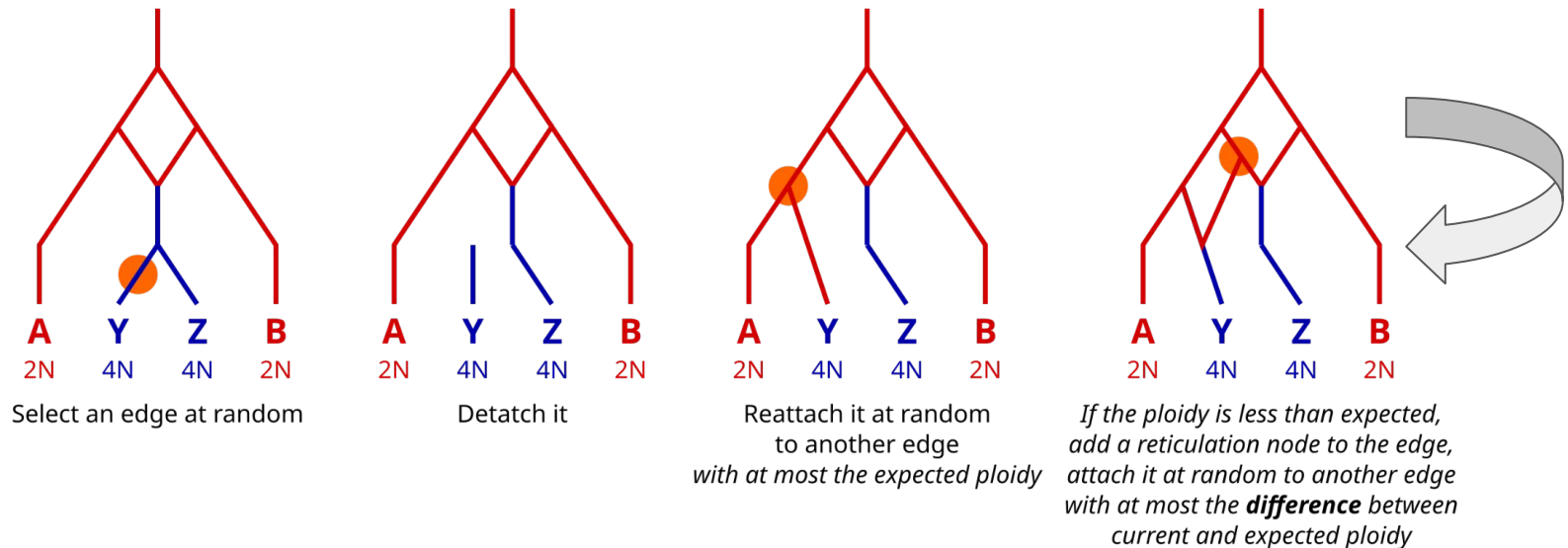


“Switch parentage” move

Subtree prune-and-regraft (SPR)

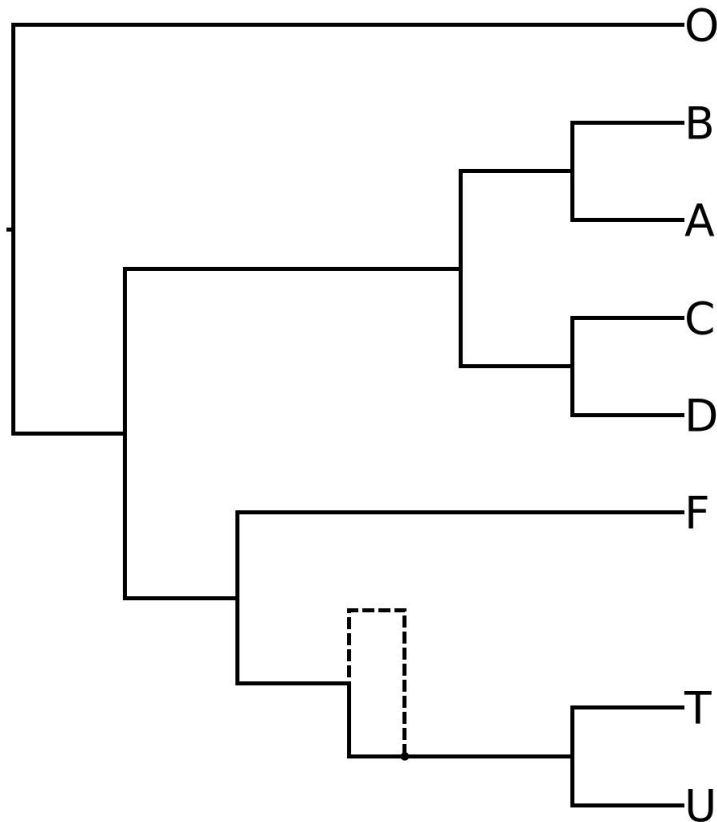


Switch parentage

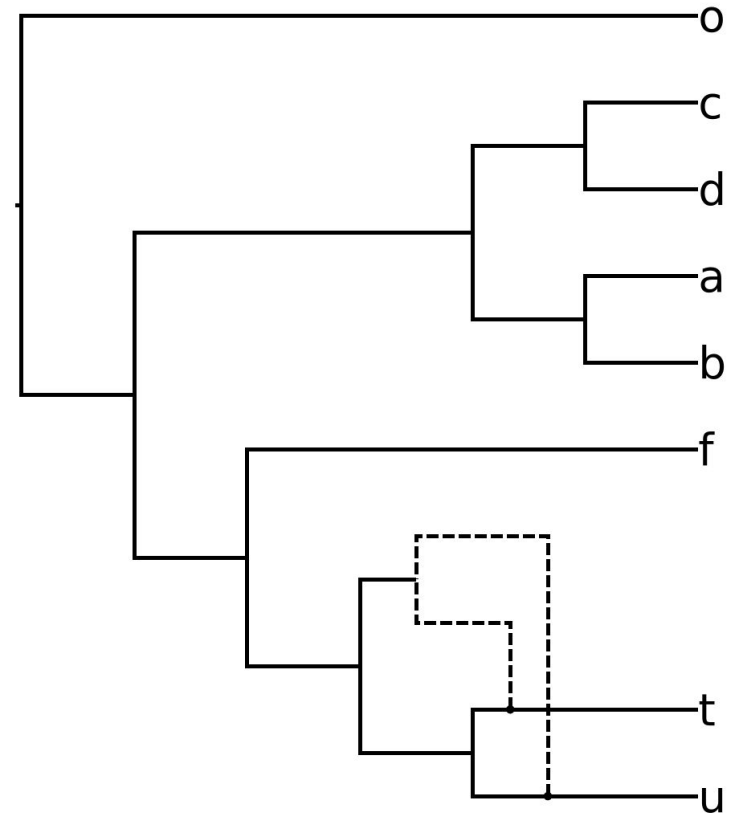


Scenario with neither parent lineage sampled

MP-SUGAR (correct)



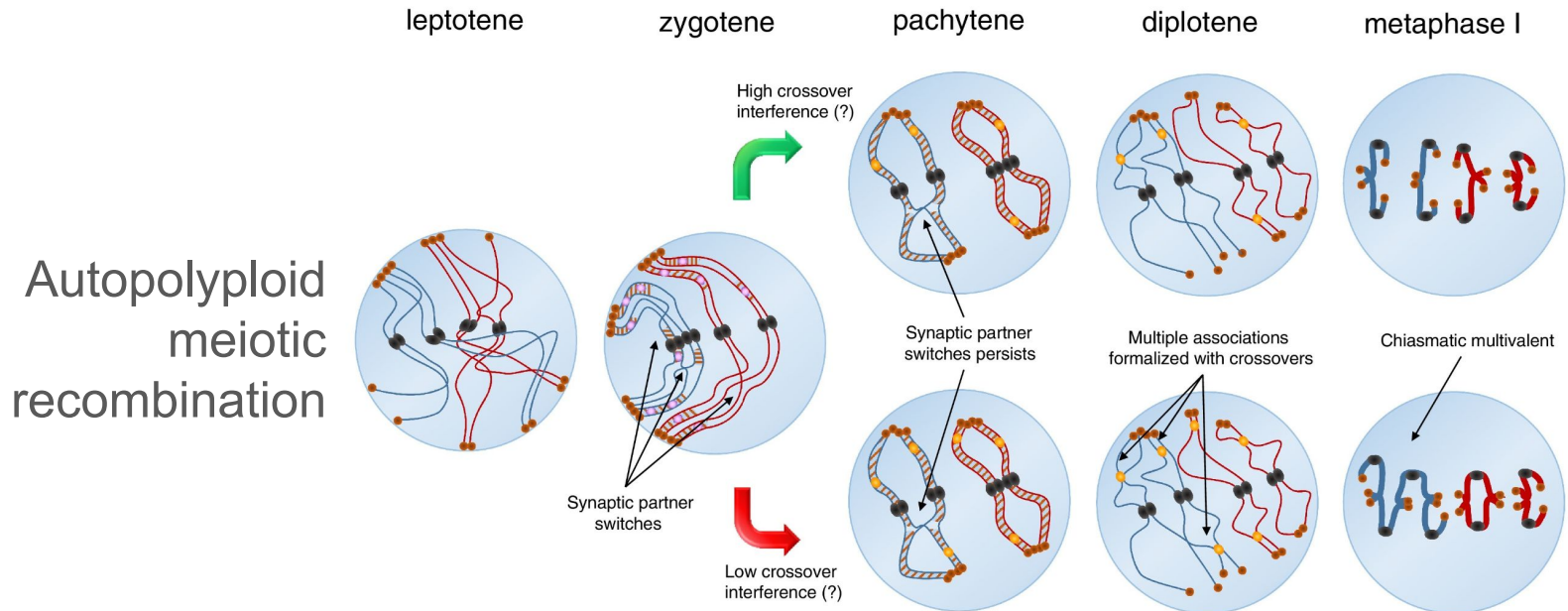
MPAllopp (incorrect)



Future directions

- Simulation study with more complex scenarios
 - Ploidy greater than 4N
 - No diploid relatives
- Add support for autopolyploidy
- Continue to squash bugs

Figure adapted from Lloyd & Bomblies (2015),
 “Meiosis in autopolyploid and allopolyploid
Arabidopsis” *Current Opinion in Plant Biology*



Acknowledgements



Mark Kessler

Software Engineer



Luay Nakhleh

Dean of Engineering & co-PI



Zhi Yan

PhD student

Department of Computer Science, Rice University, Houston Texas