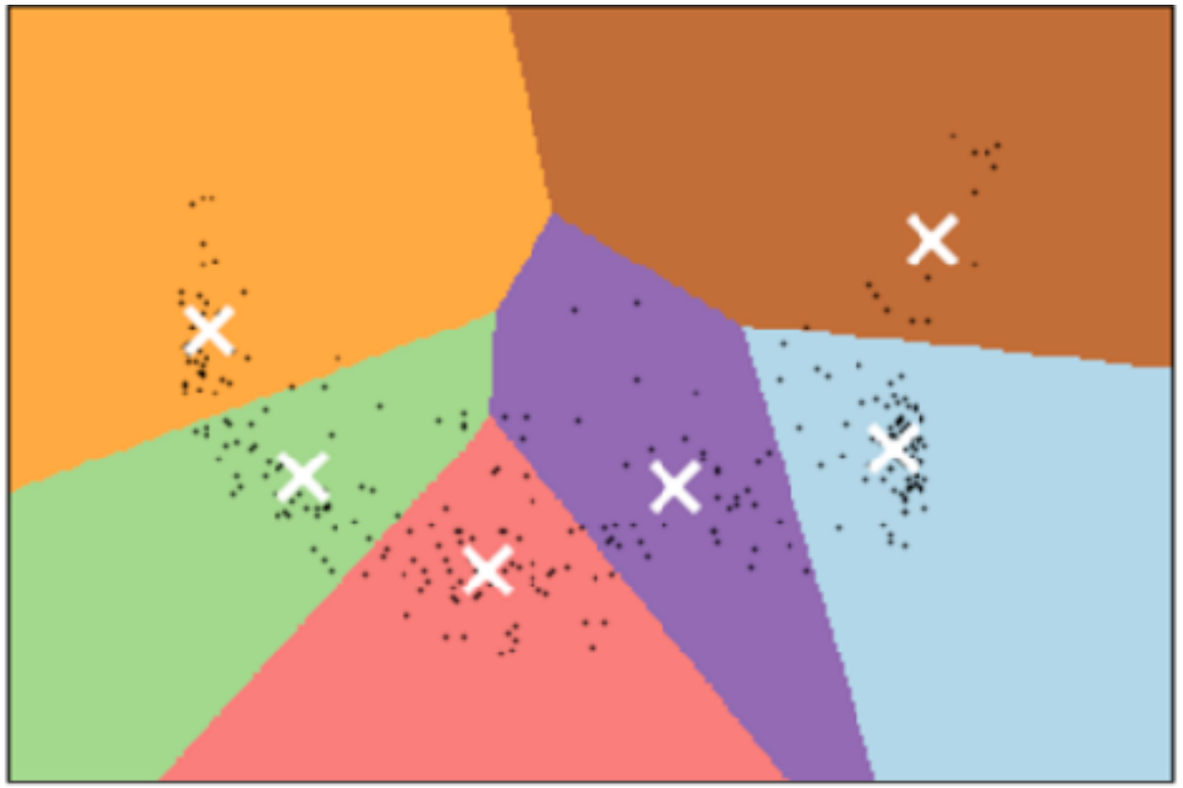
MATT JOHNSON, CHICAGO BOTANIC GARDEN

mjohnson@chicagobotanic.org

@mossmatters



MarkerMiner
Locus development made easy

OUTLINE

Locus selection: two strategies

Examples at different phylogenetic scales

PAFTOL probe design from 1KP

Hands On With MarkerMiner

PROBE DESIGN: TWO STRATEGIES

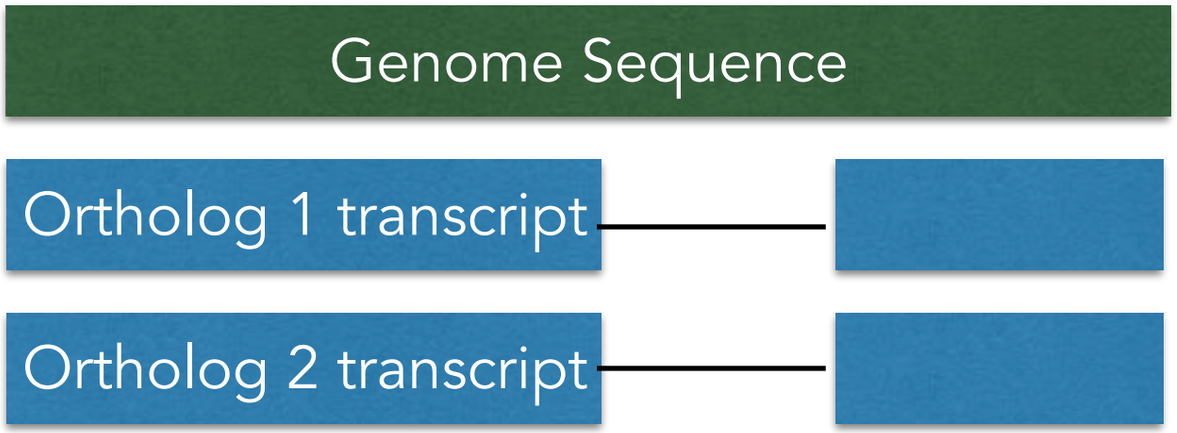
Genome Skimming

- 10x - 20x genome coverage
- Intron location information
- May not be suitable for large genomes
- Organellar genome assembly

Transcriptome Sequencing

- 20-30 million RNASeq reads
- Reduced genome representation: exons only
- Tissue and time dependent
- Can multiplex several species

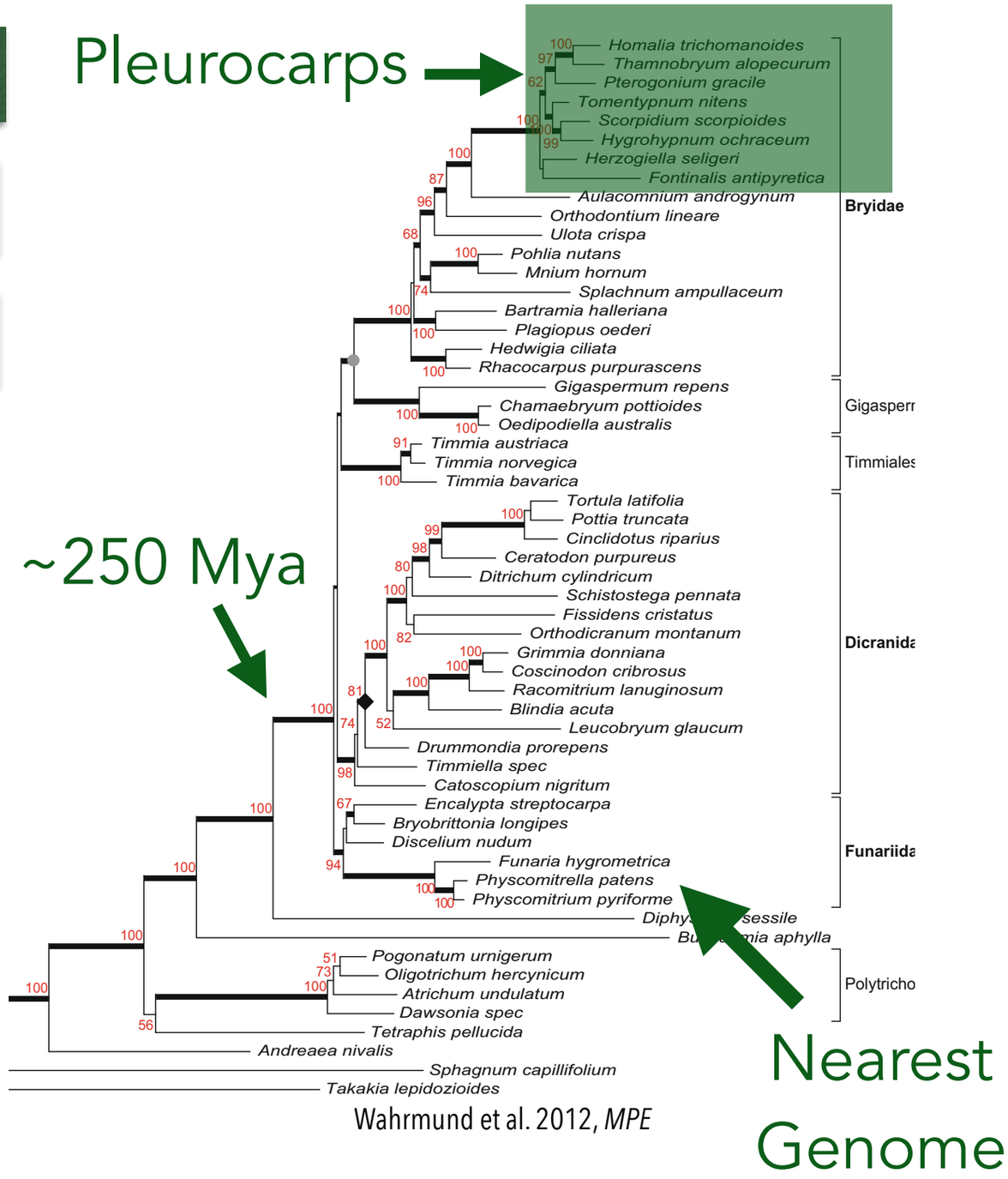
TARGET DESIGN EXAMPLES: MOSS TREE OF LIFE



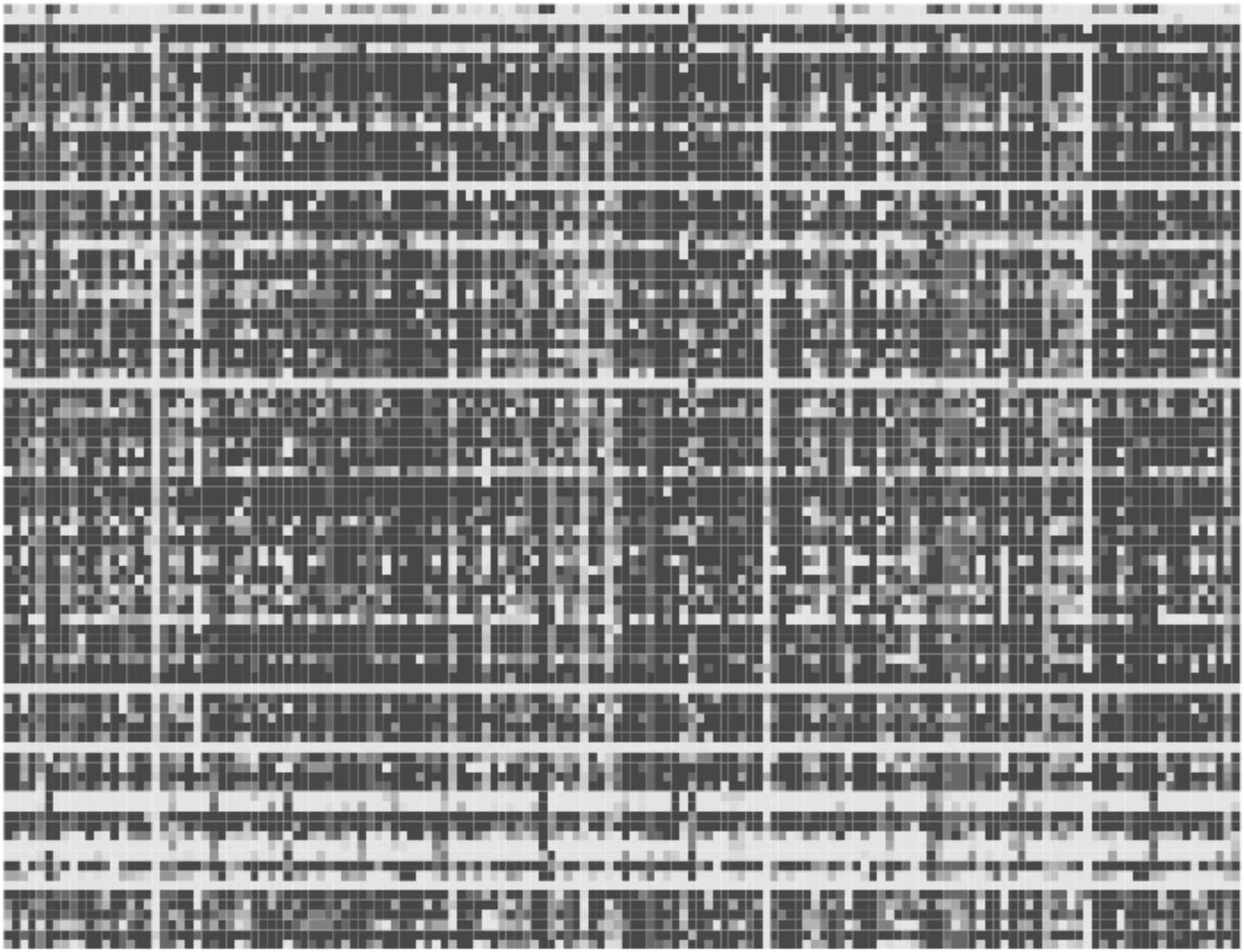
Align moss transcripts (1KP) to *Physcomitrella* genome

Select genes expressed in at least two pleurocarpous mosses

Design probes from multiple sequences



TARGET DESIGN EXAMPLES: MOSS TREE OF LIFE

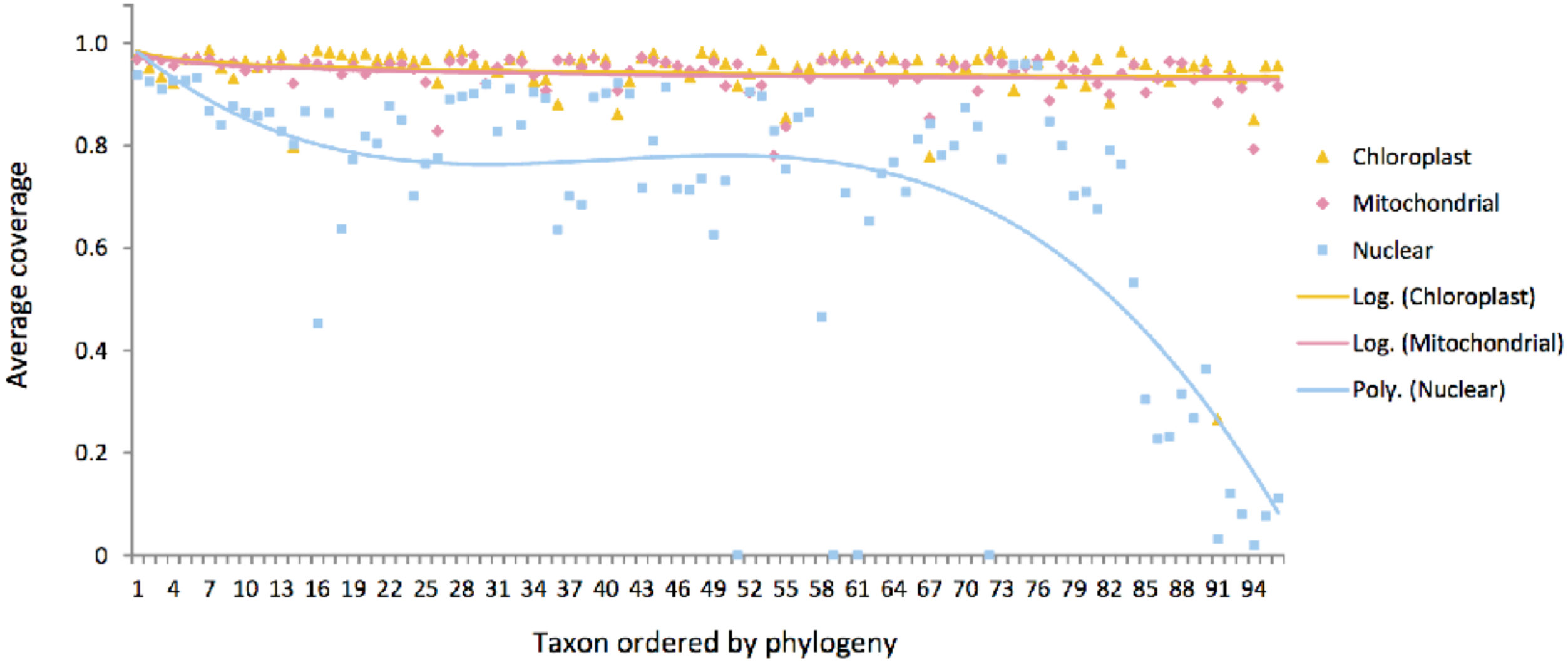


96 Samples

150 Genes

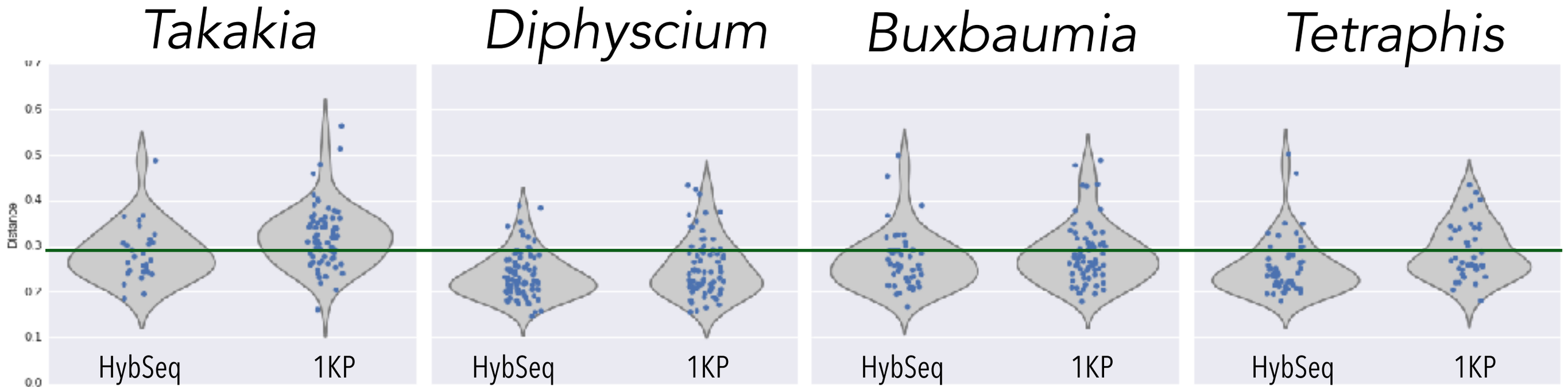
Liu, Johnson, et al., in prep

TARGET DESIGN EXAMPLES: MOSS TREE OF LIFE



Liu, Johnson, et al., in prep

TARGET DESIGN EXAMPLES: MOSS TREE OF LIFE



Comparing divergence between probe sequences and:

Sequences recovered by HybPiper (left)
Transcripts from 1KP (right)

In mosses far diverged from *Physcomitrella*

Liu, Johnson, et al., in prep

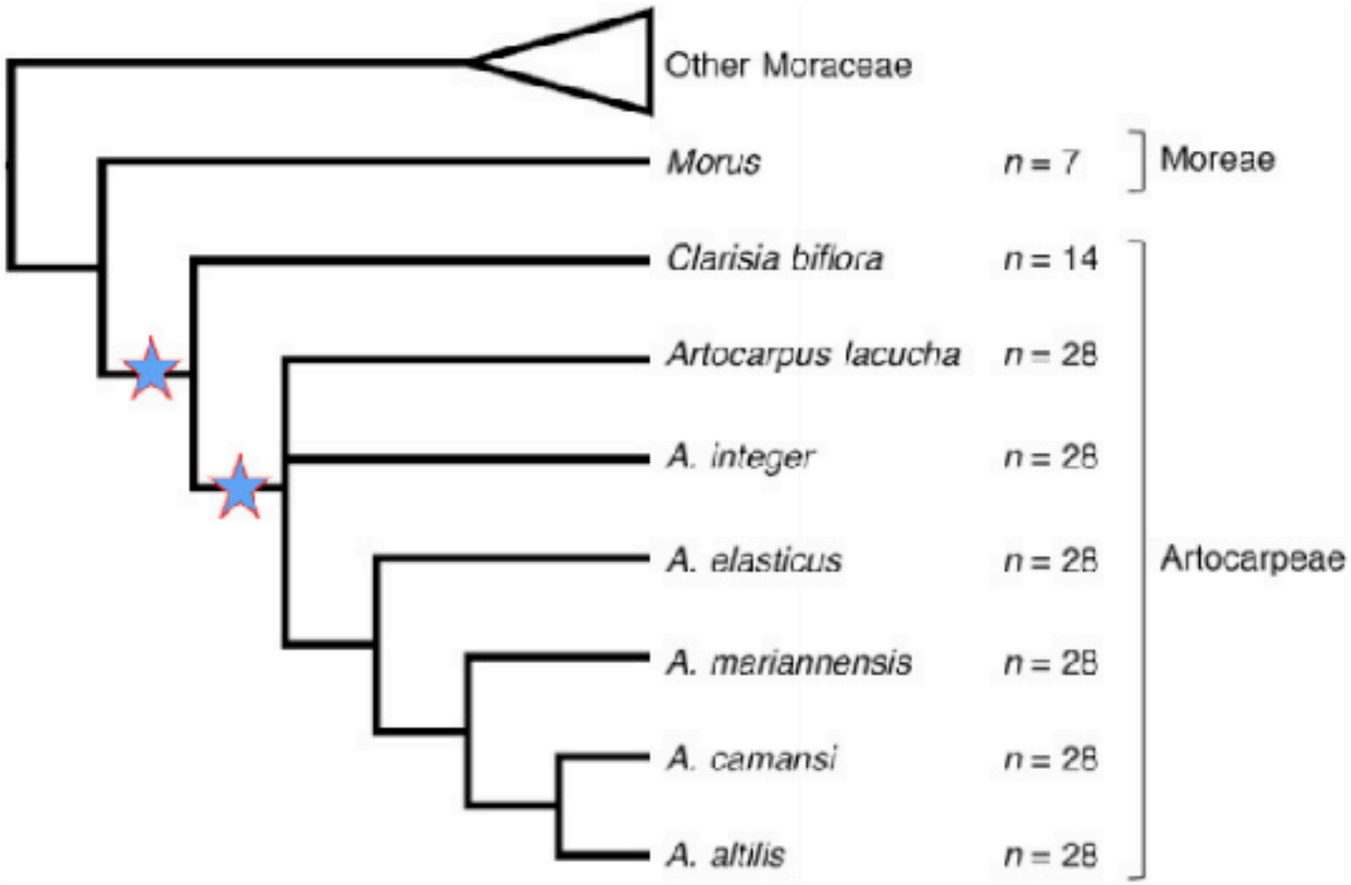
THE IMPORTANCE OF MULTIPLE ORTHOLOGS PER GENE

More accurate alignment

Provides redundancy (free extra tiling)

Expands phylogenetic breadth

TARGET DESIGN EXAMPLES: ARTOCARPUS GENOME SKIM



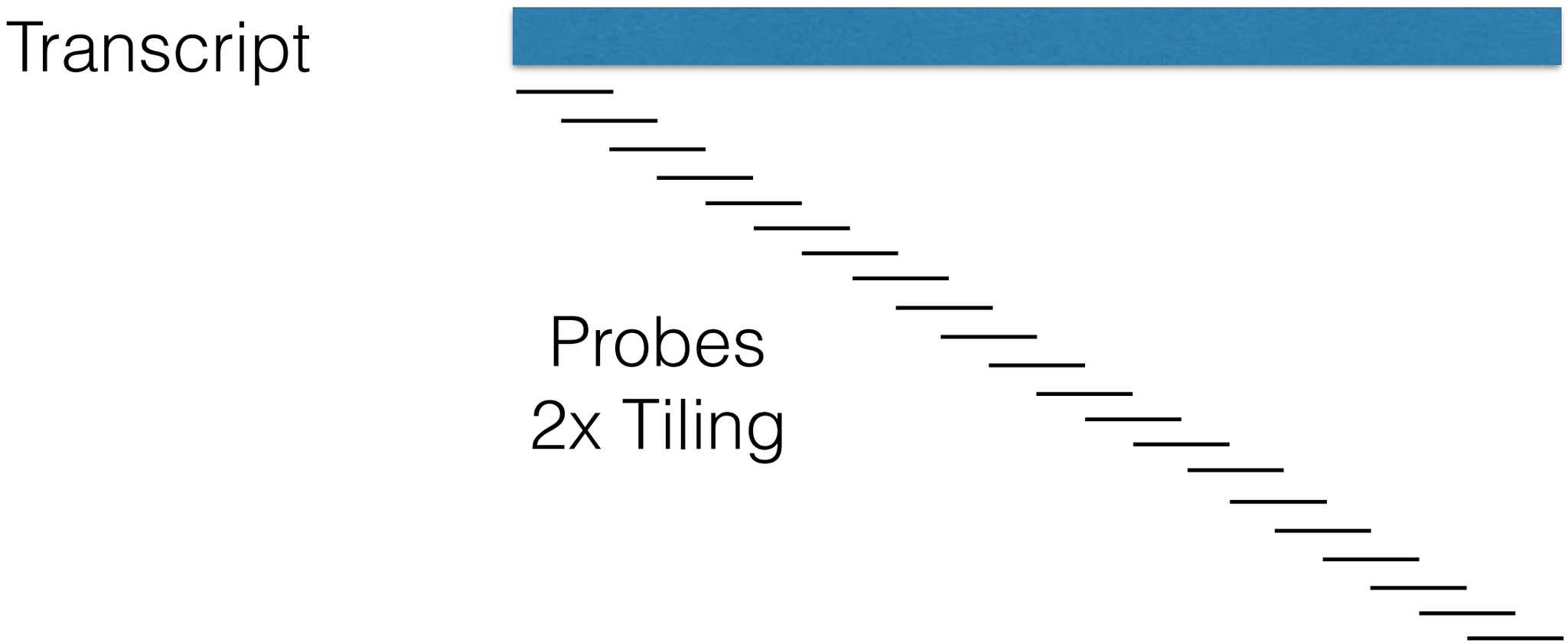
17x Whole-Genome Sequencing (*Artocarpus camansi*)

Determined orthology with *Morus*

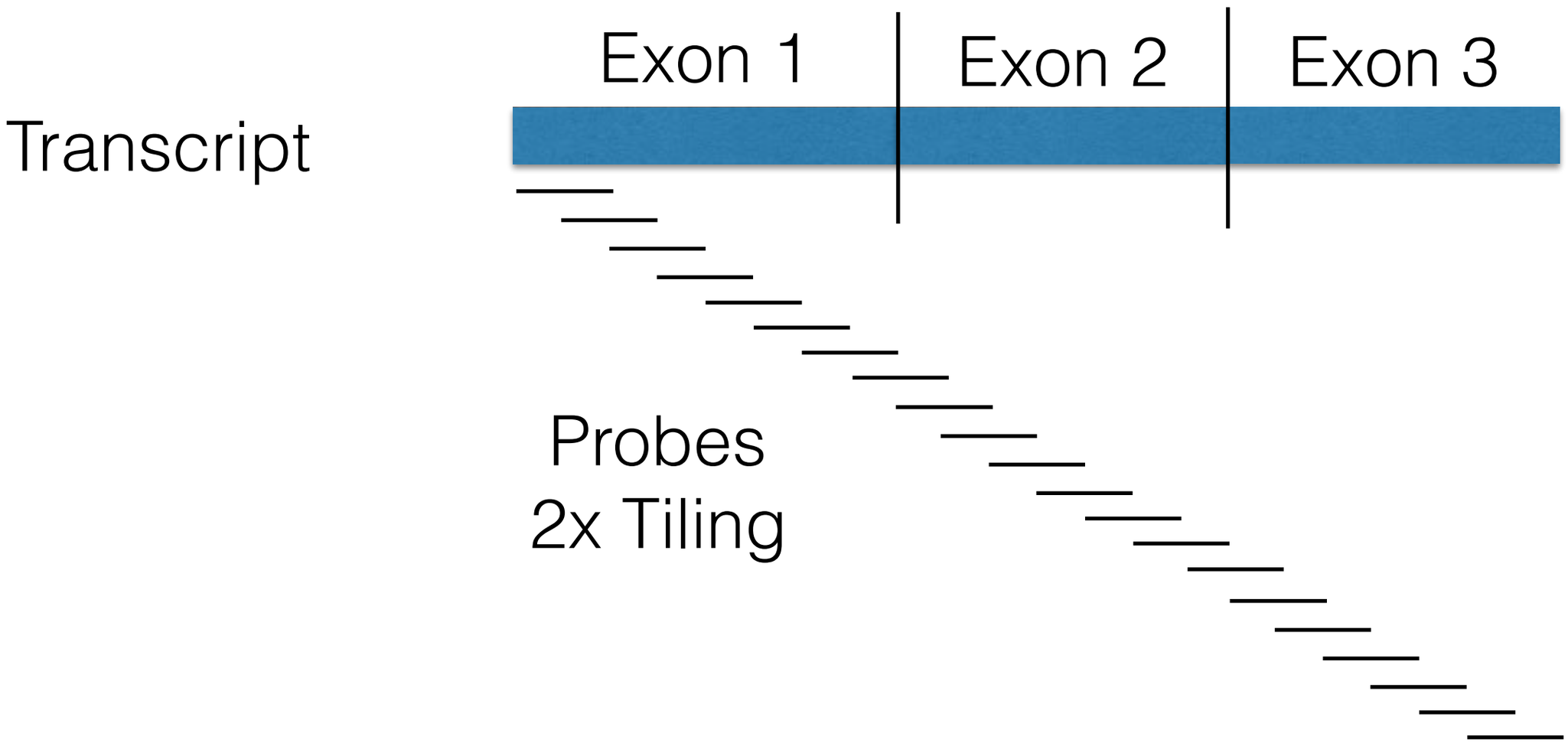
333 "phylogeny" genes, plus MADS-Box and volatiles

Gardner et al., APPS, 2016

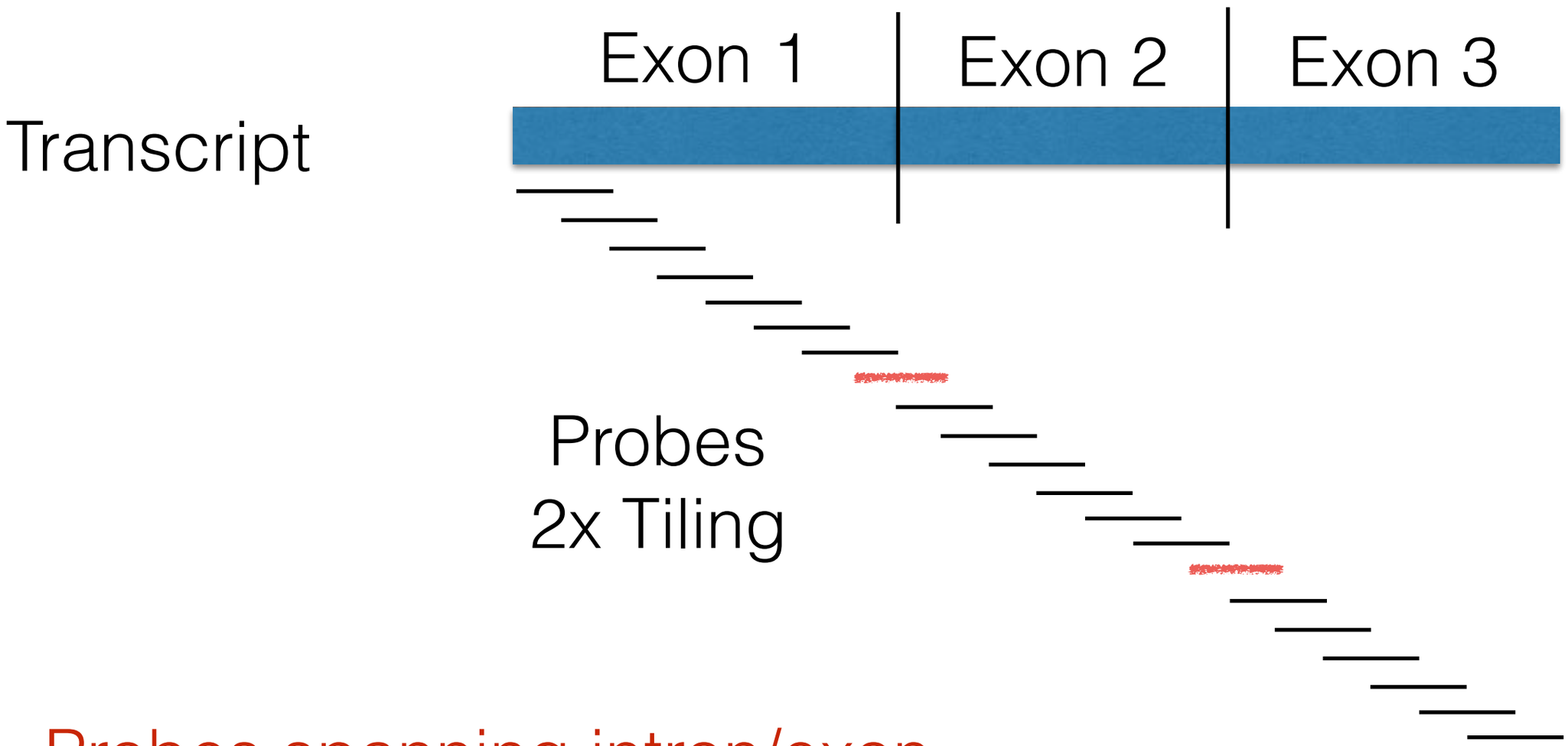
TARGET DESIGN EXAMPLES: USING GENOME SEQUENCE



TARGET DESIGN EXAMPLES: USING GENOME SEQUENCE

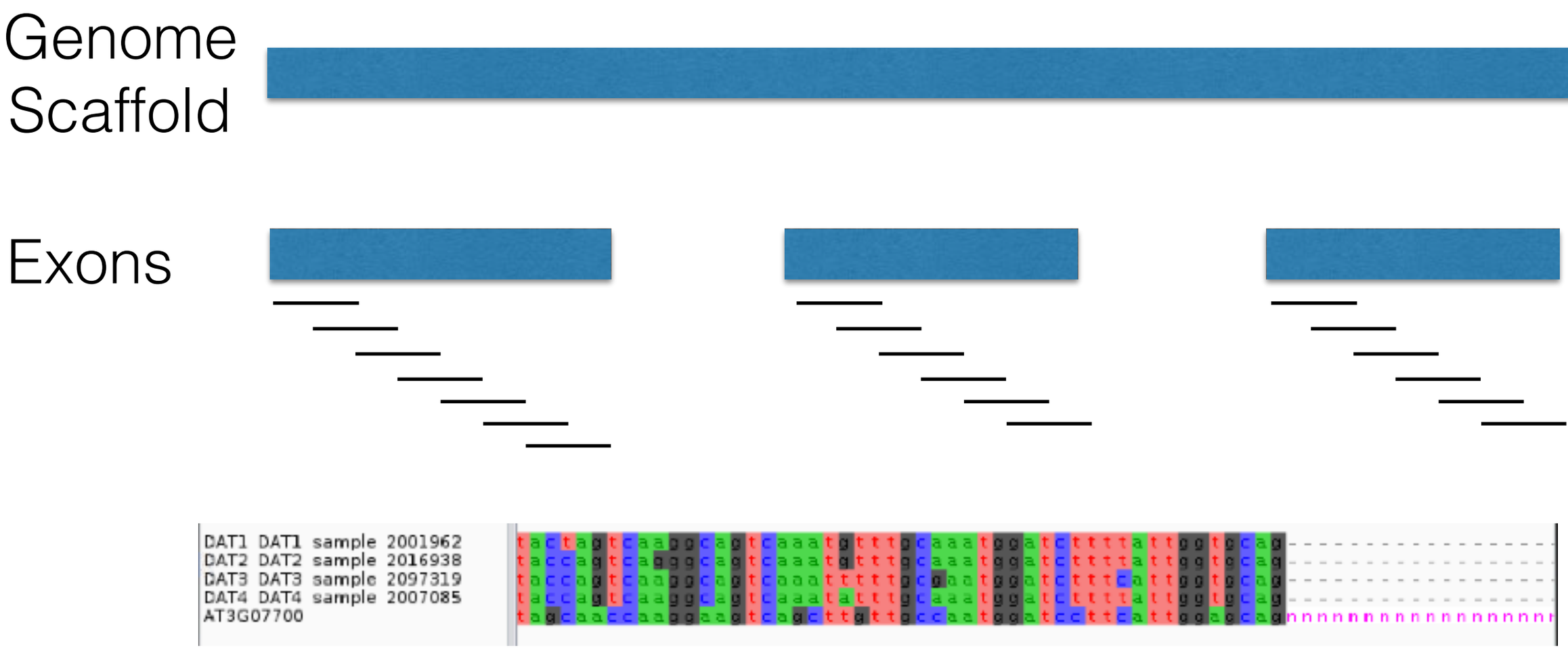


TARGET DESIGN EXAMPLES: USING GENOME SEQUENCE



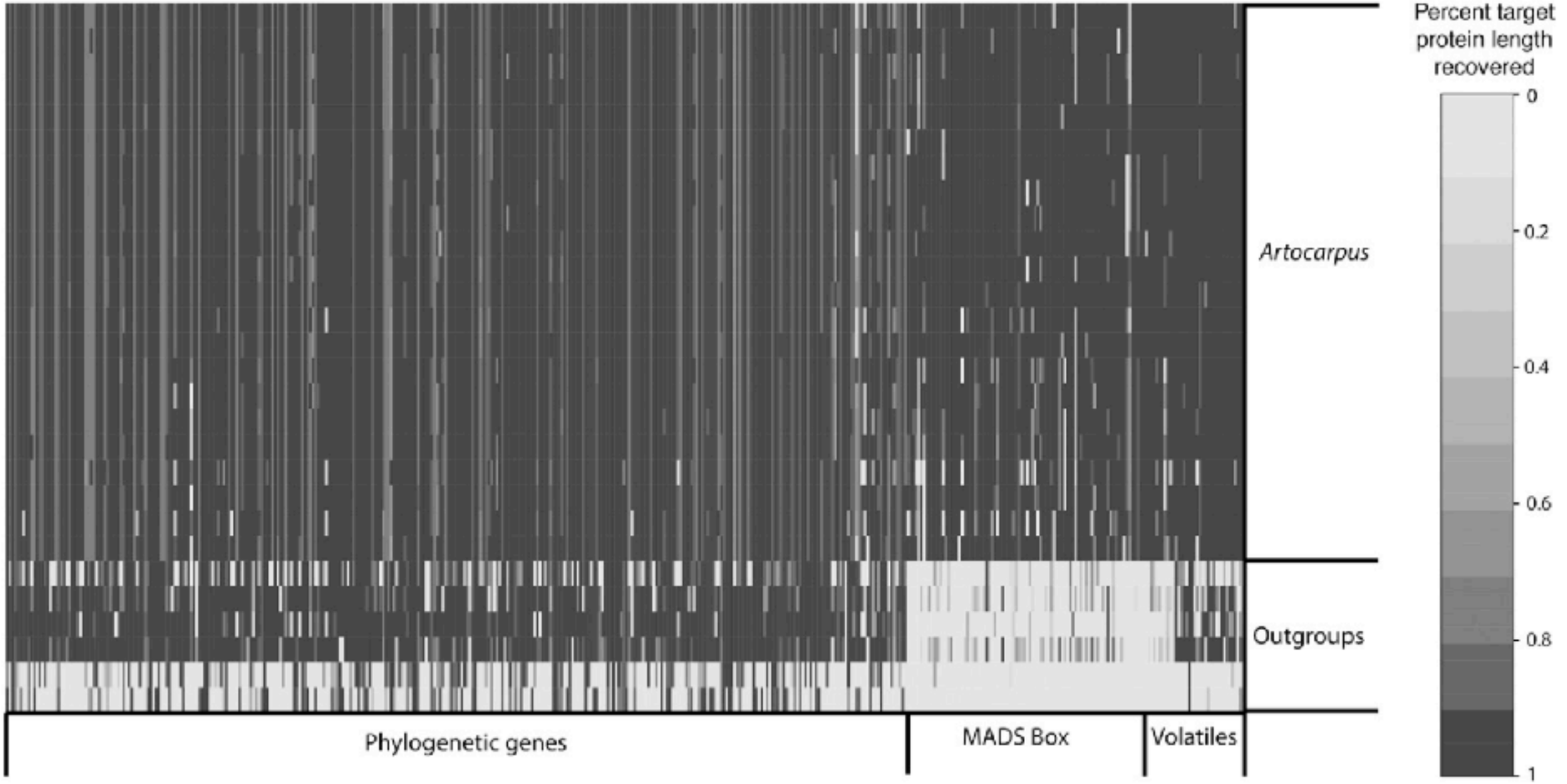
Probes spanning intron/exon boundaries may not work

TARGET DESIGN EXAMPLES: USING GENOME SEQUENCE



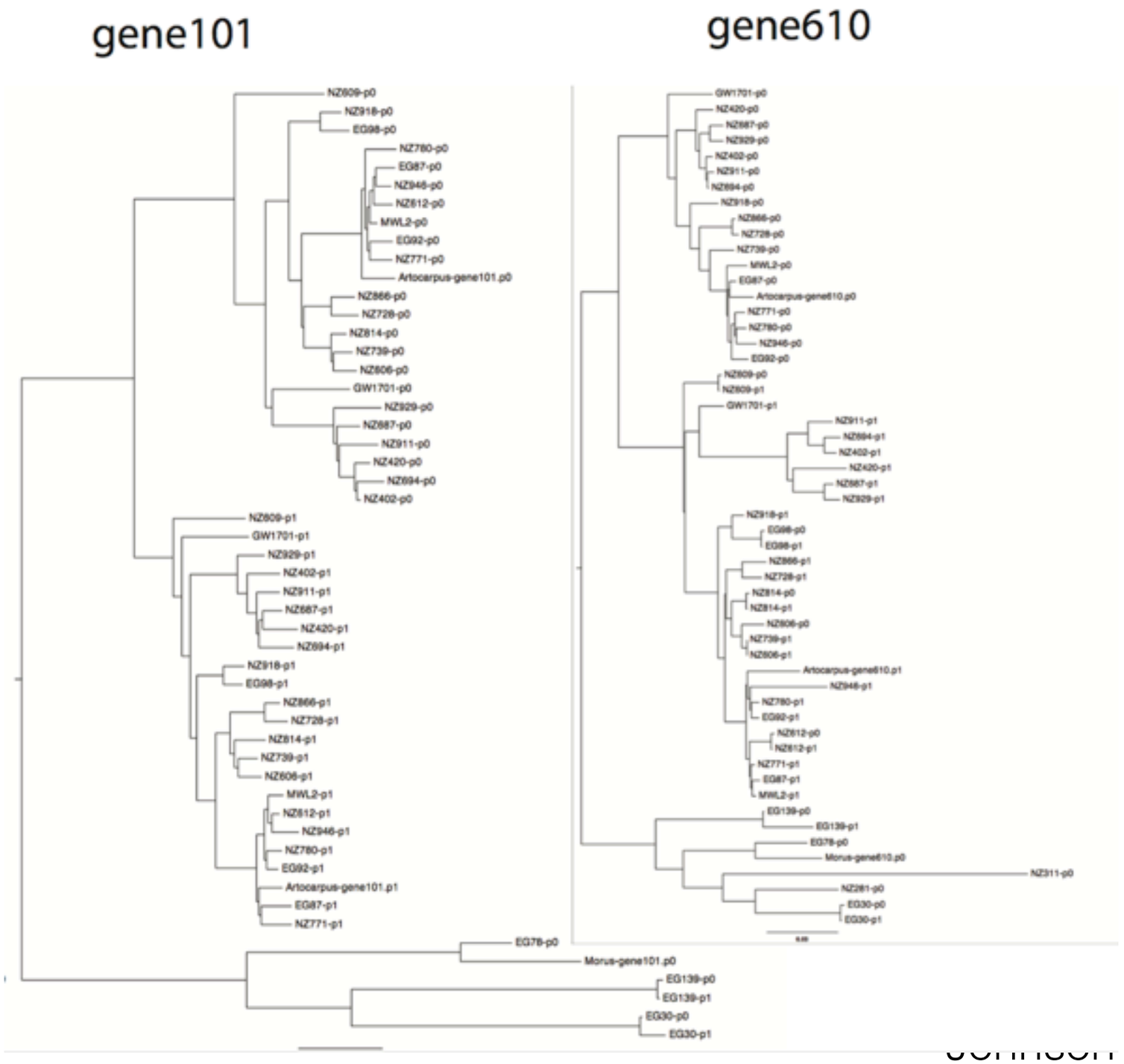
With genome, probes extend to intron/exon boundaries
Remainder of probe can be filled with T

TARGET DESIGN EXAMPLES: ARTOCARPUS GENOME SKIM



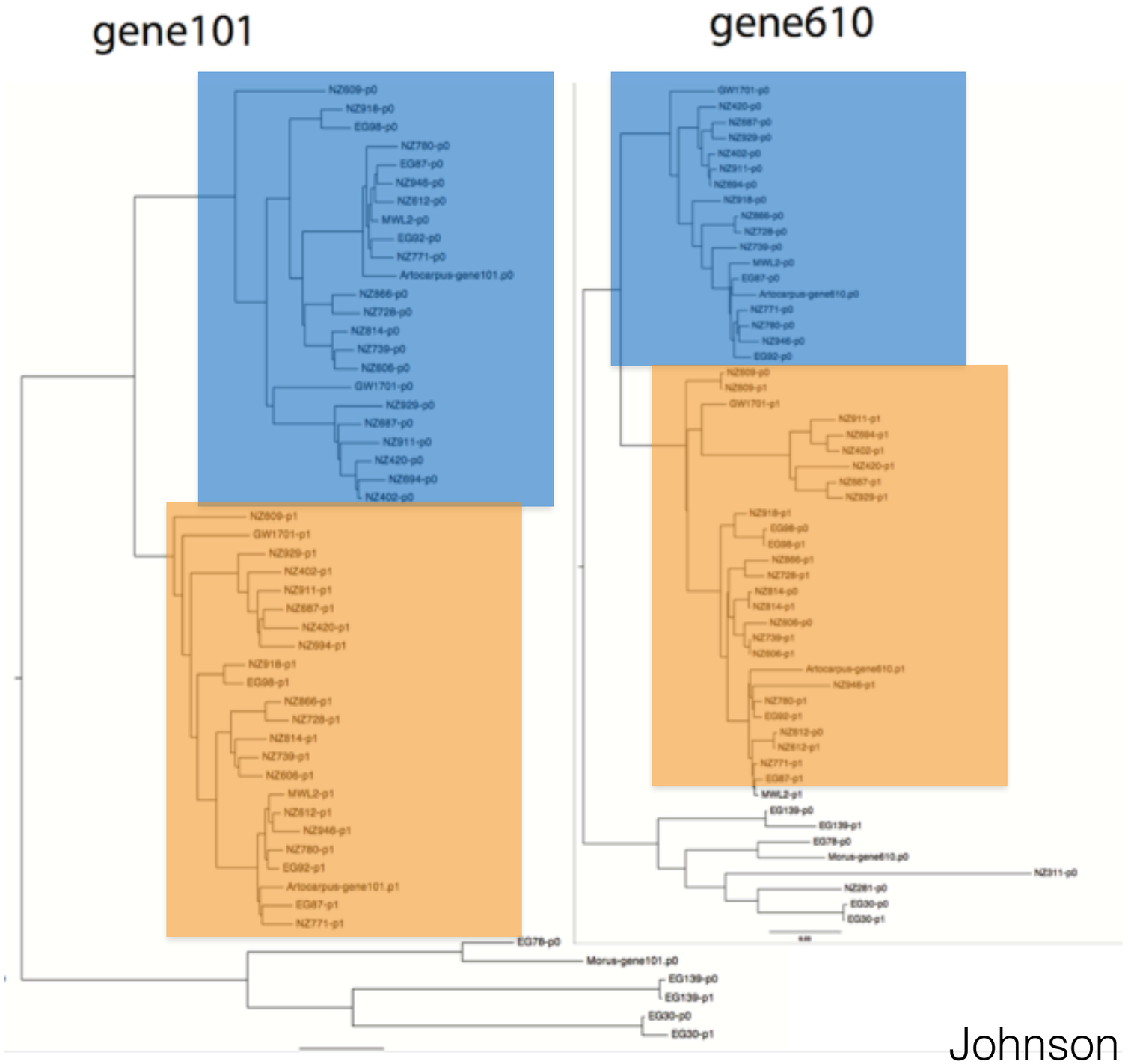
Johnson et al., APPS, 2016

TARGET DESIGN EXAMPLES: ARTOCARPUS GENOME SKIM



Johnson et al., APPS, 2016

TARGET DESIGN EXAMPLES: ARTOCARPUS GENOME SKIM



Johnson et al., APPS, 2016

PAFTOL PROBE DESIGN

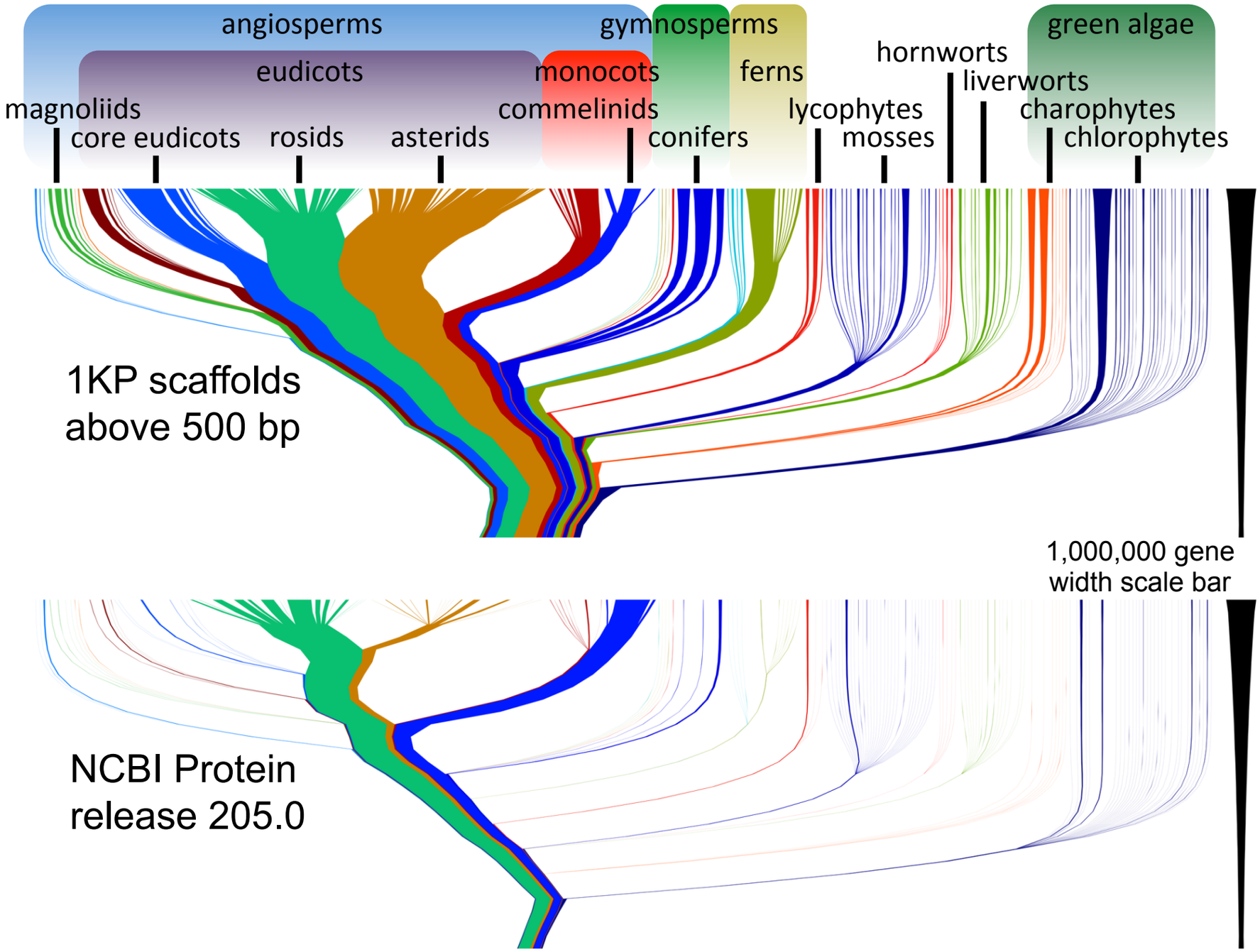
Build a genus level phylogeny of flowering plants

One Kit to Rule Them All?

Which genes?
Which orthologs?

Data: 400 genes from 1000 Plant Transcriptome Project (1KP)

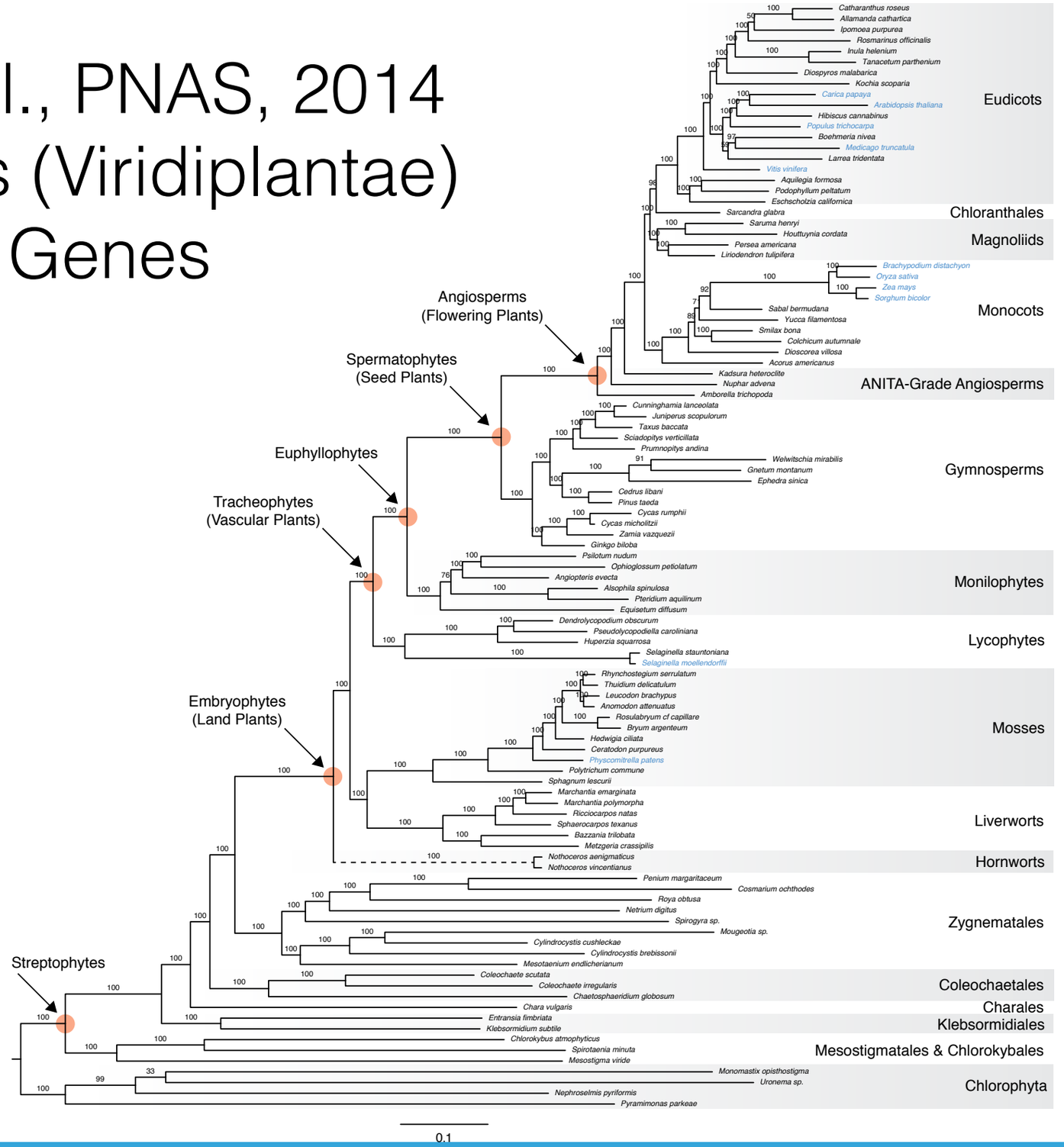
PAFTOL HYBSEQ PROJECT: COVERING ALL ANGIOSPERMS



onekp.com

PAFTOL HYBSEQ PROJECT: COVERING ALL ANGIOSPERMS

Wickett et al., PNAS, 2014
107 Species (Viridiplantae)
852 Genes



PAFTOL HYBSEQ PROJECT: COVERING ALL ANGIOSPERMS

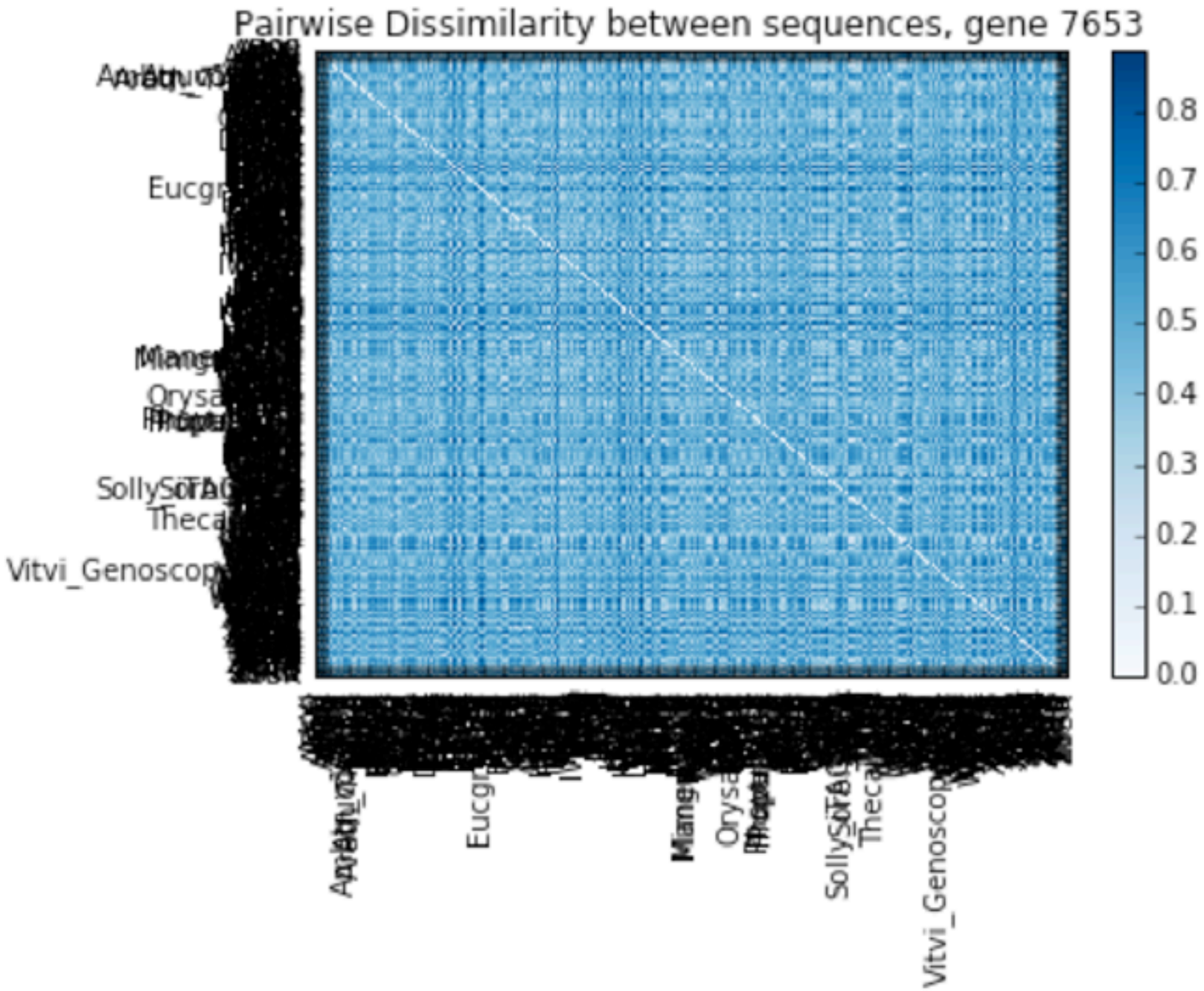
Data: 400 loci used for phylogenetics

Orthology already determined (using genomes)

Pre-screened for low-copy in Viridiplantae

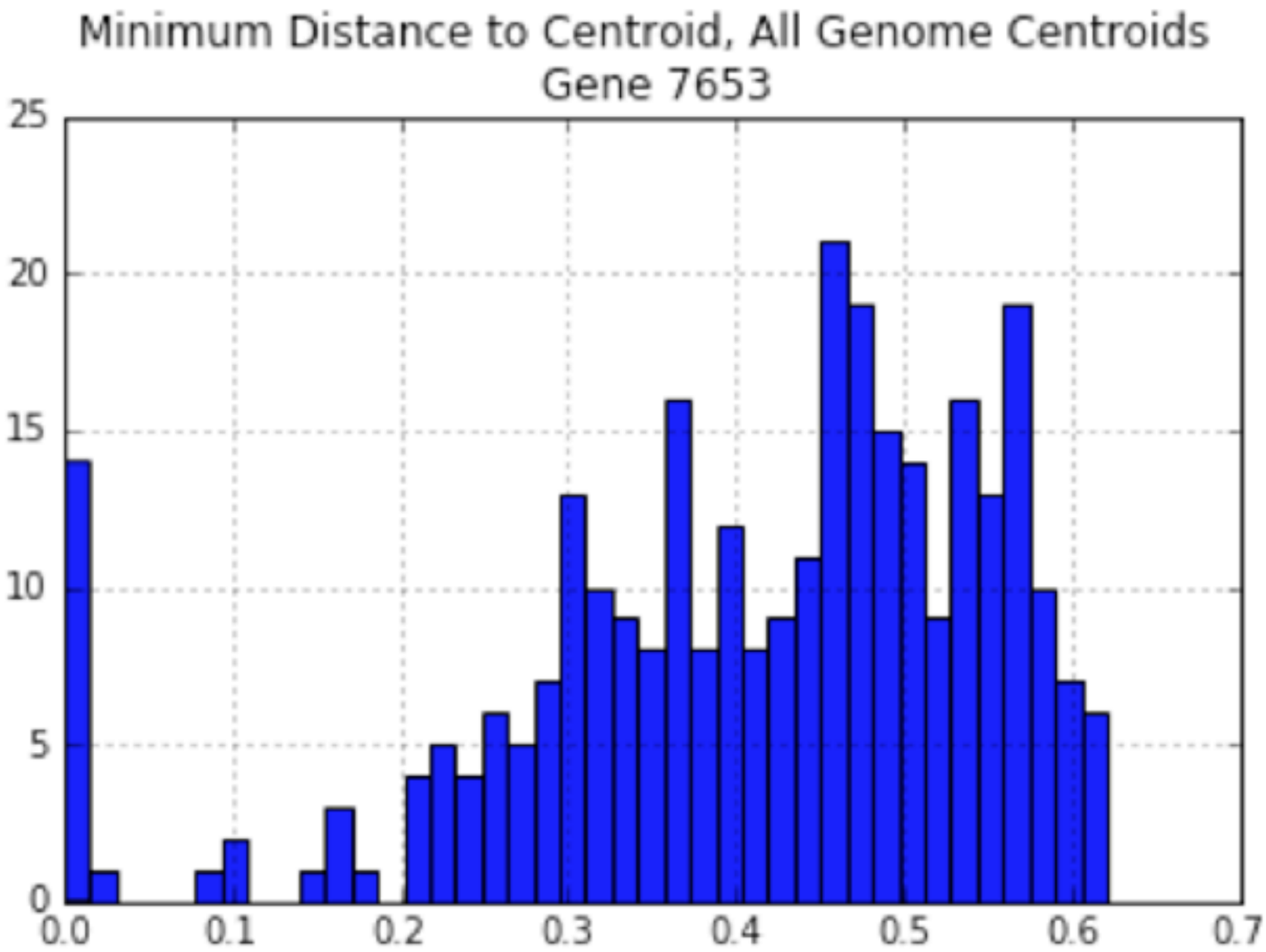
Too many sequences to start (600+ Angiosperms)

SELECTING SEQUENCES: MACHINE LEARNING



SELECTING SEQUENCES: MACHINE LEARNING

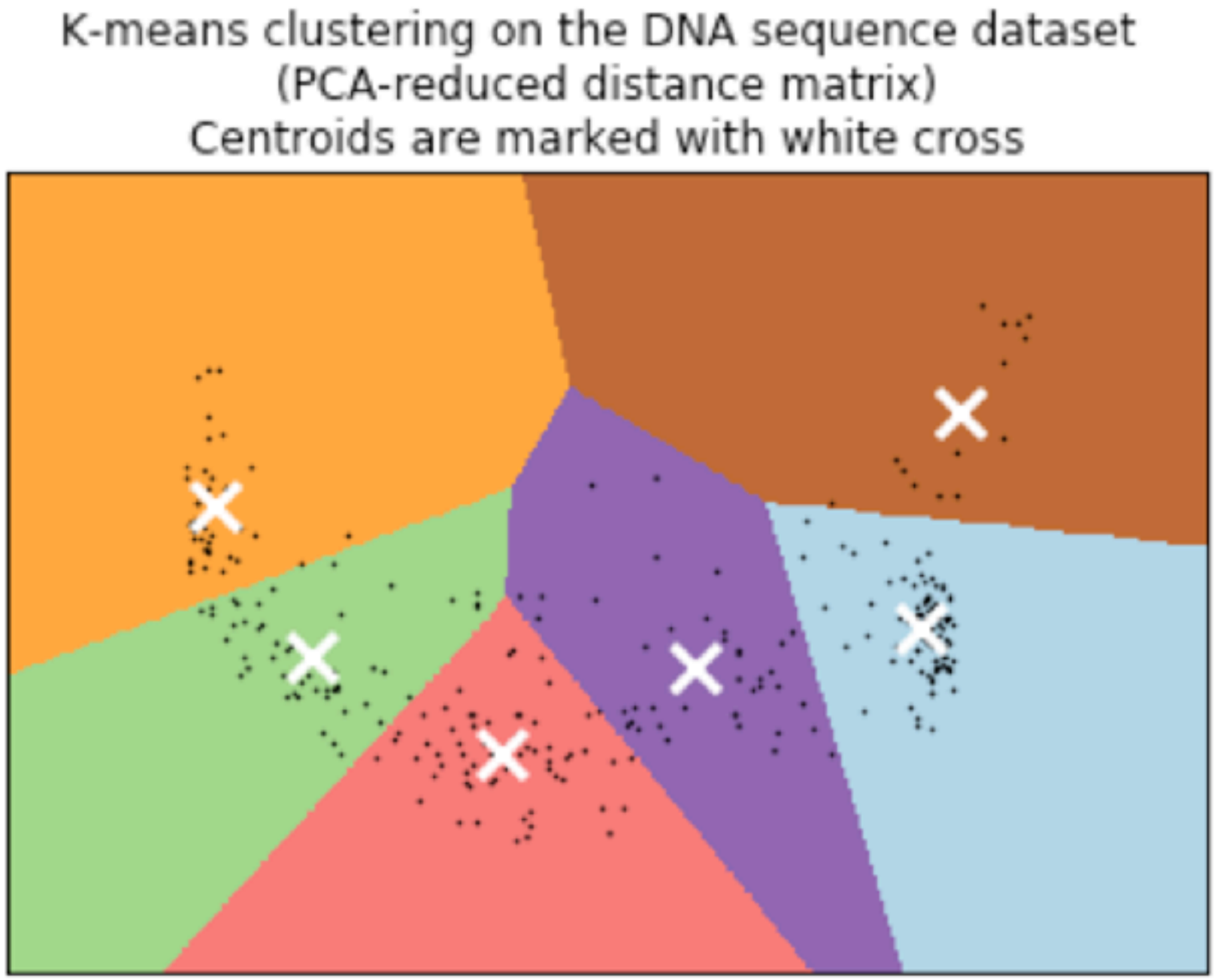
What if we only select from genomes?



Too far for HybSeq!

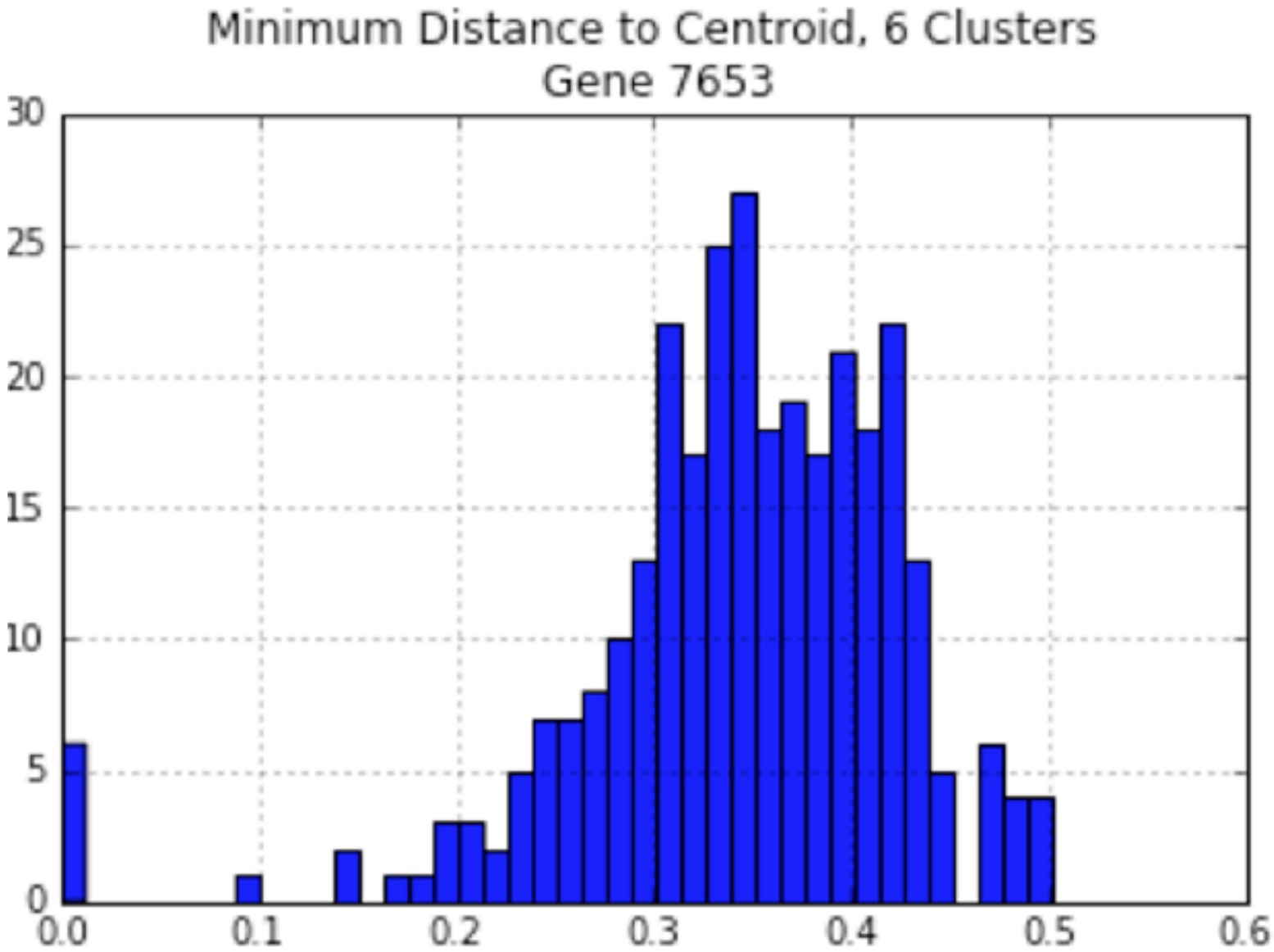
SELECTING SEQUENCES: MACHINE LEARNING

Reducing the dimensionality with K-means clustering



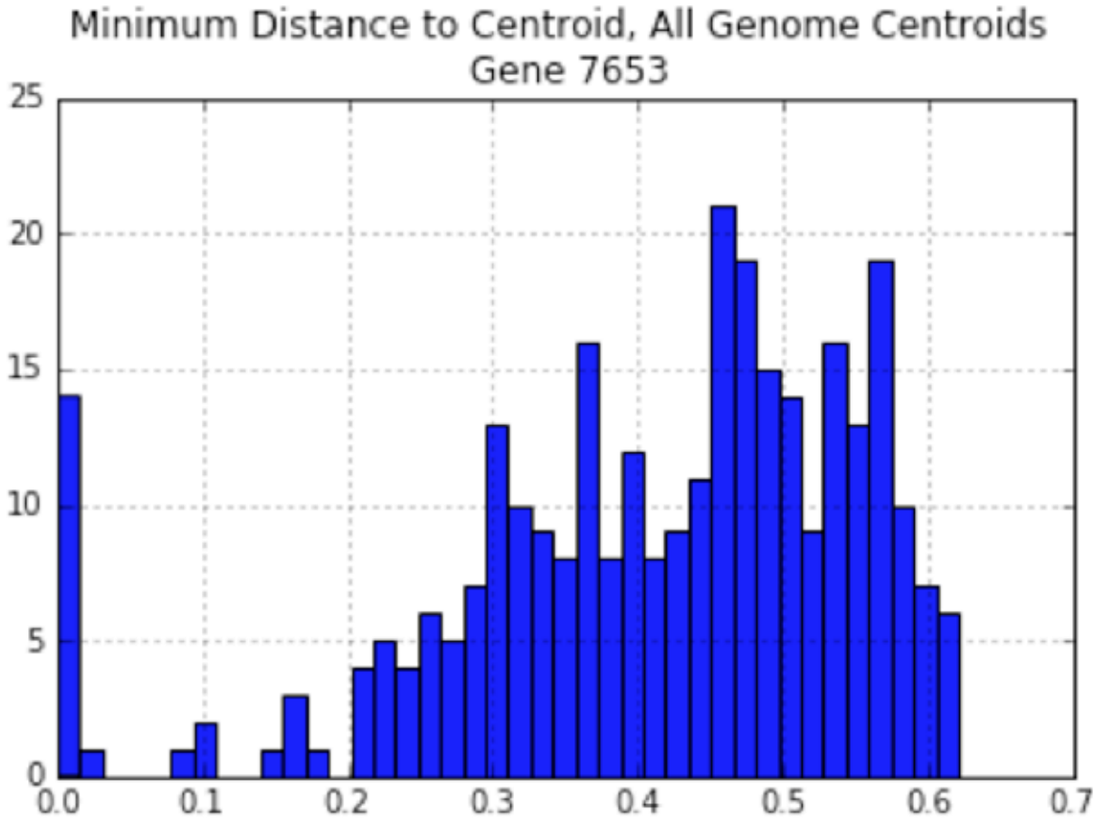
SELECTING SEQUENCES: MACHINE LEARNING

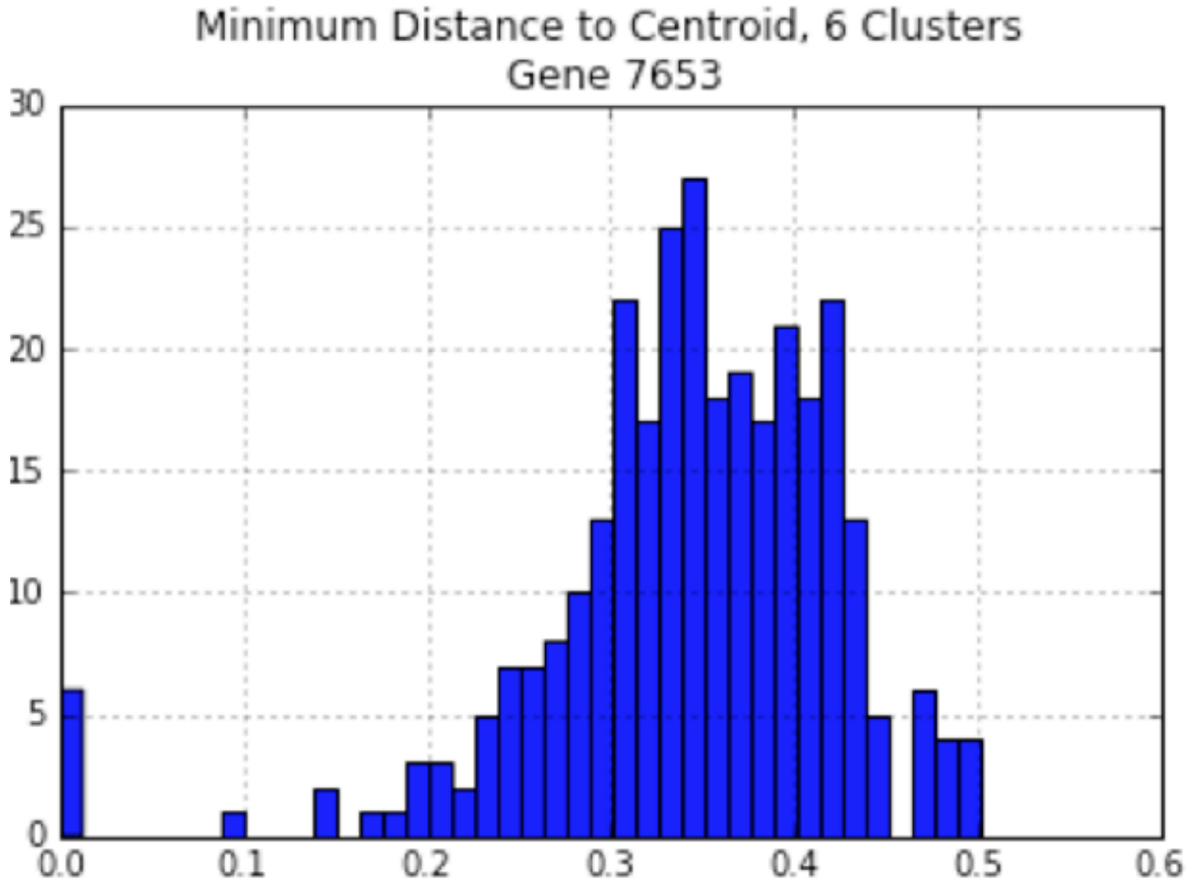
Reducing the dimensionality with K-means clustering



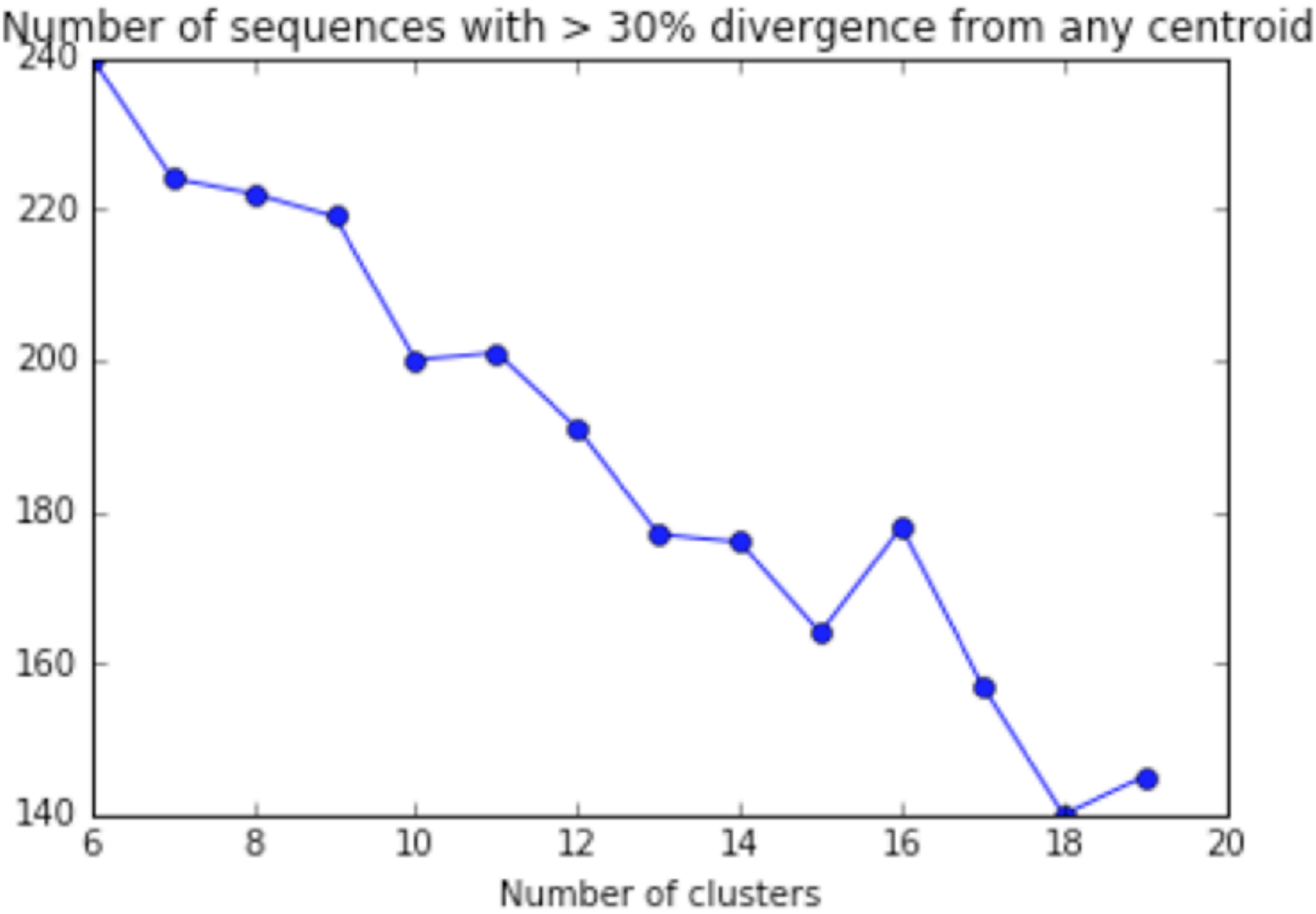
SELECTING SEQUENCES: MACHINE LEARNING

Reducing the dimensionality with K-means clustering



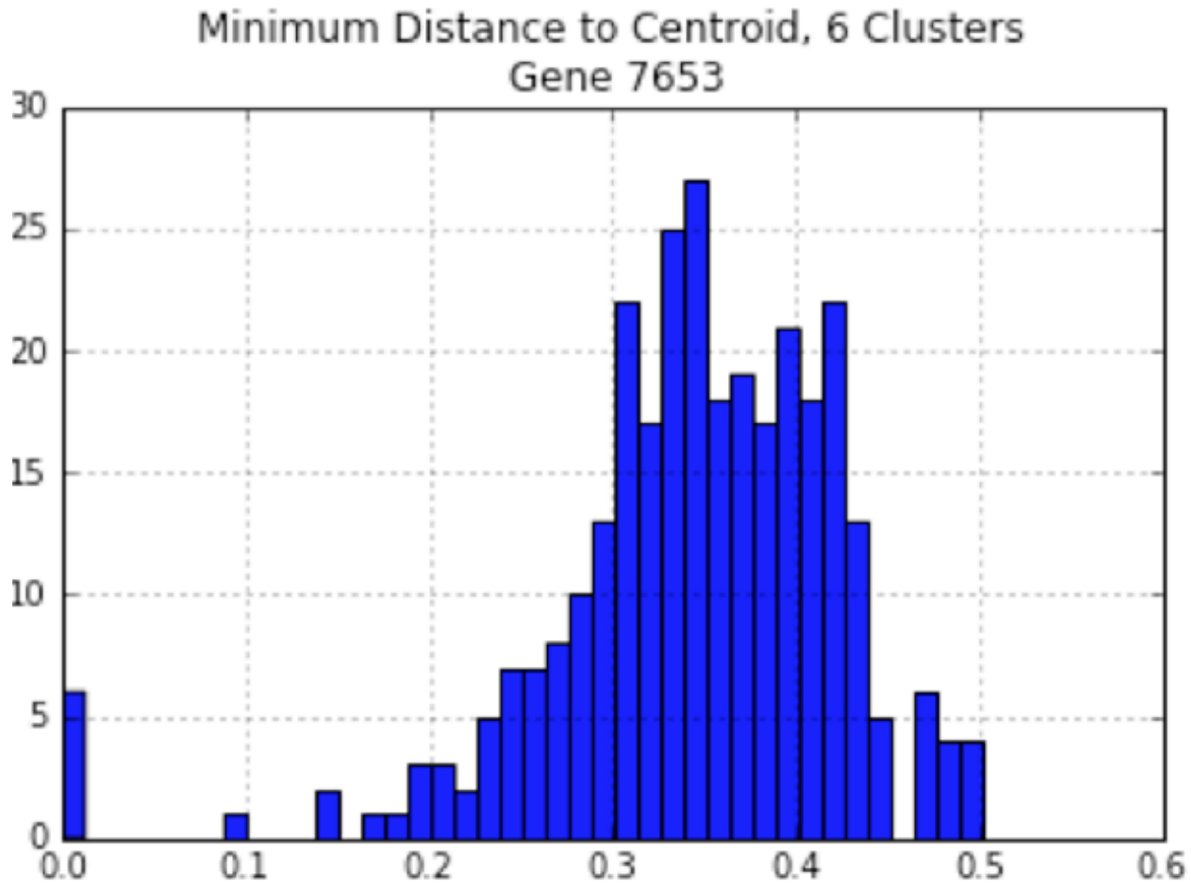


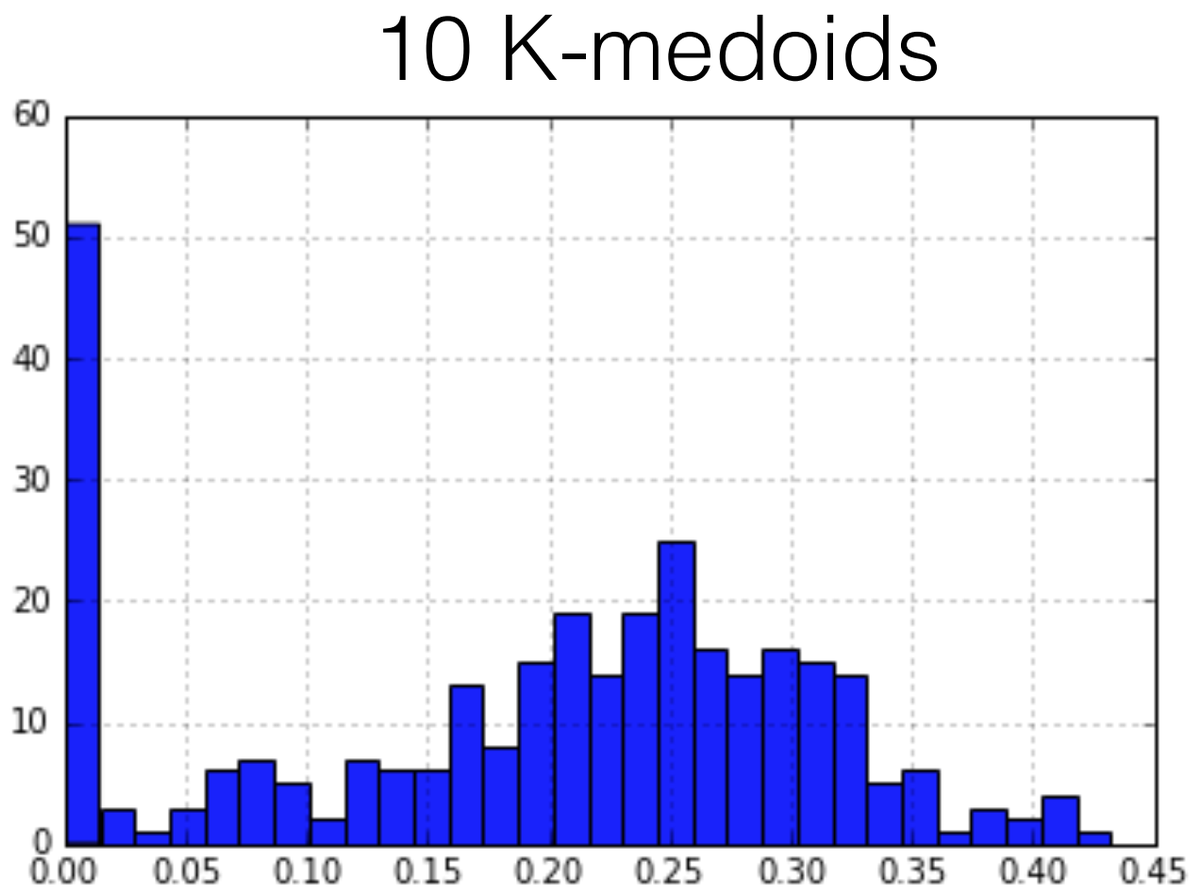
SELECTING SEQUENCES: MACHINE LEARNING



603 total angiosperm sequences

SELECTING SEQUENCES: MACHINE LEARNING





603 total angiosperm sequences

PAFTOL ANGIOSPERM PROBE DESIGN

Using k-medoids method, 354 loci selected

Between 6 and 15 medoids represent > 95% of all 1KP angiosperms

80,000 probes, designed by MycroArray (Michigan, USA)

Pilot project: Sequencing 288 Angiosperms

Stay tuned...

HANDS ON EXERCISE

MarkerMiner

Locus development made easy

Chamala et al., APPS, 2015

Align transcriptomes to existing genomic resources

Select single-copy loci

Generate alignments ready to submit for probe design

Command line or web interface

IMPORTANT WEBSITES

atmo.cyverse.org

github.com/mossmatters/KewHybSeqWorkshop

mossmatters.com