Characterizing the impact of transposon dynamics in *Helianthus*

Evan Staton
Department of Botany and Biodiversity Research Centre, UBC
sunflower meeting, 1/13/16

Transposable elements are important sources of variation

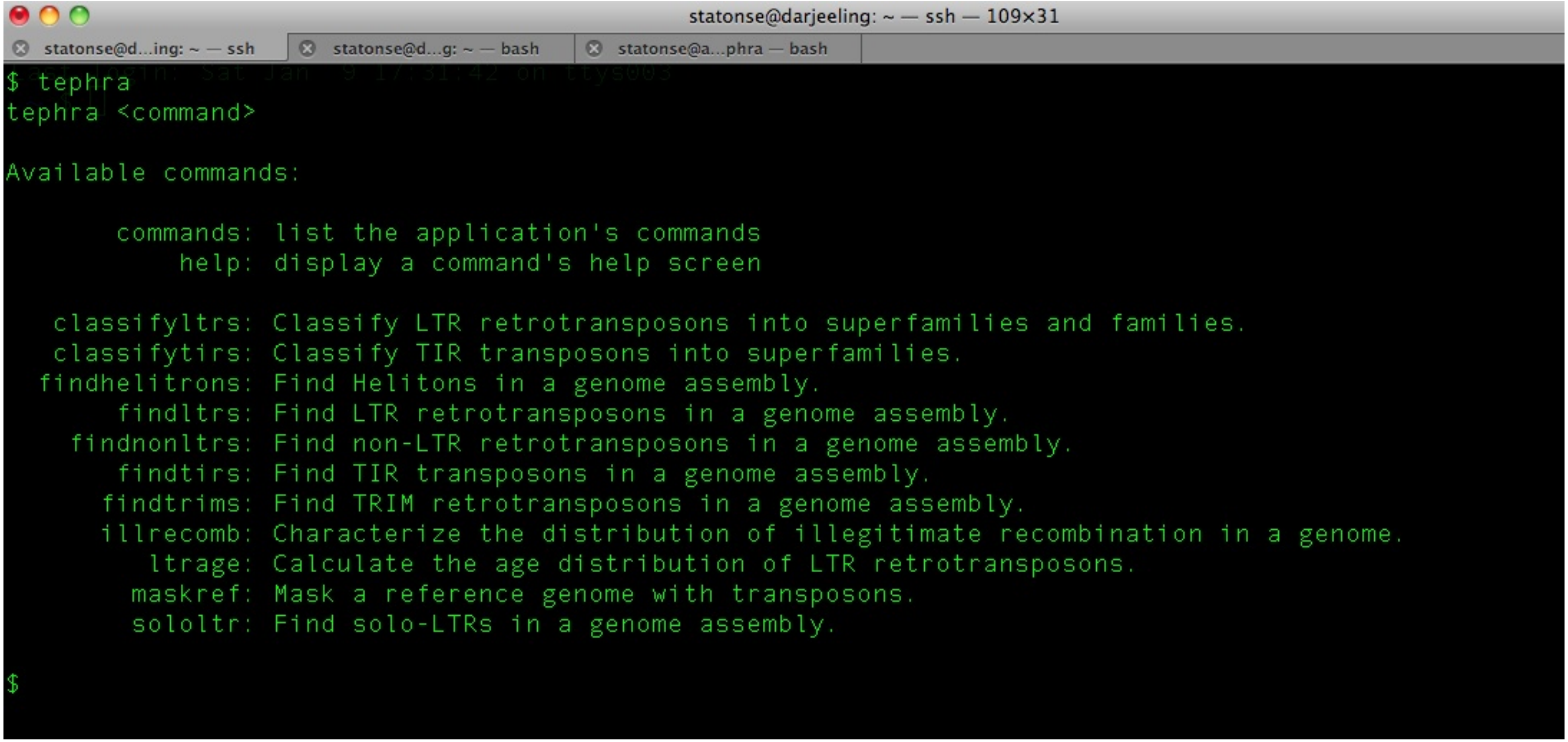
Common sunflower genome description

- Fine-scale structure of the LTR-RTs in the sunflower genome have been thoroughly analyzed
    - Time scales of activity
    - Deletion rates
    - Characteristics of insertion to describe genomic bias
    - Differential gain/loss over time

Common sunflower genome description

- Fine-scale structure of *all TEs* in the sunflower genome
  - Abundance/diversity both within and between species
  - Time scales of activity
  - Deletion rates
  - Differential gain/loss over time

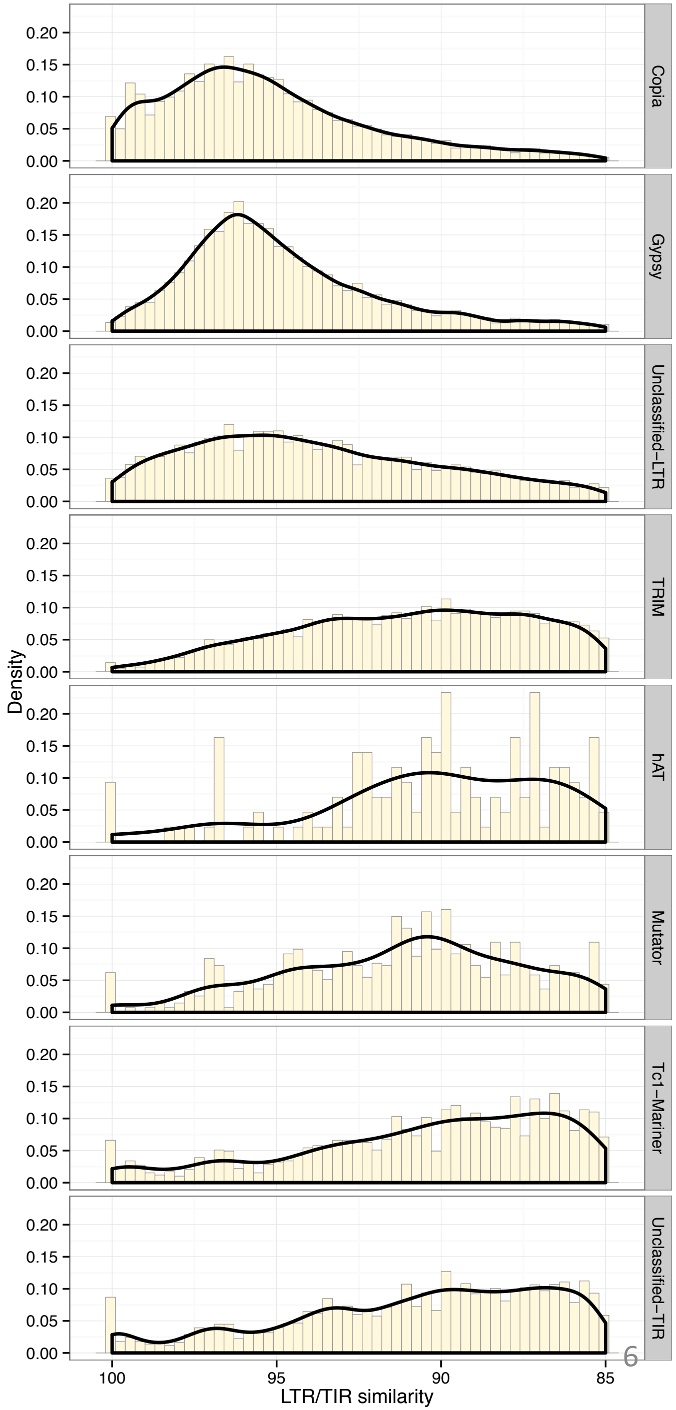
TE annotation with Tephra



```
$ tephra
tephra <command>

Available commands:

       commands: list the application's commands
           help: display a command's help screen

   classifyltrs: Classify LTR retrotransposons into superfamilies and families.
   classifytirs: Classify TIR transposons into superfamilies.
  findhelitrons: Find Helitons in a genome assembly.
       findltrs: Find LTR retrotransposons in a genome assembly.
    findnonltrs: Find non-LTR retrotransposons in a genome assembly.
       findtirs: Find TIR transposons in a genome assembly.
      findtrims: Find TRIM retrotransposons in a genome assembly.
      illrecomb: Characterize the distribution of illegitimate recombination in a genome.
        ltrage: Calculate the age distribution of LTR retrotransposons.
        maskref: Mask a reference genome with transposons.
        sololtr: Find solo-LTRs in a genome assembly.

$
```

Tephra – github.com/sestaton/tephra

The time scales of sunflower transposon activity



Tephra – github.com/sestaton/tephra

Intraspecific variation in *Helianthus annuus*

Compare 288 sunflower breeding lines representing USDA, INRA, Oil and non-Oil and “core 12” (Mandel et al. 2013)

1) *What is the total level of TE variation between lines?*

Intraspecific variation in *Helianthus annuus*

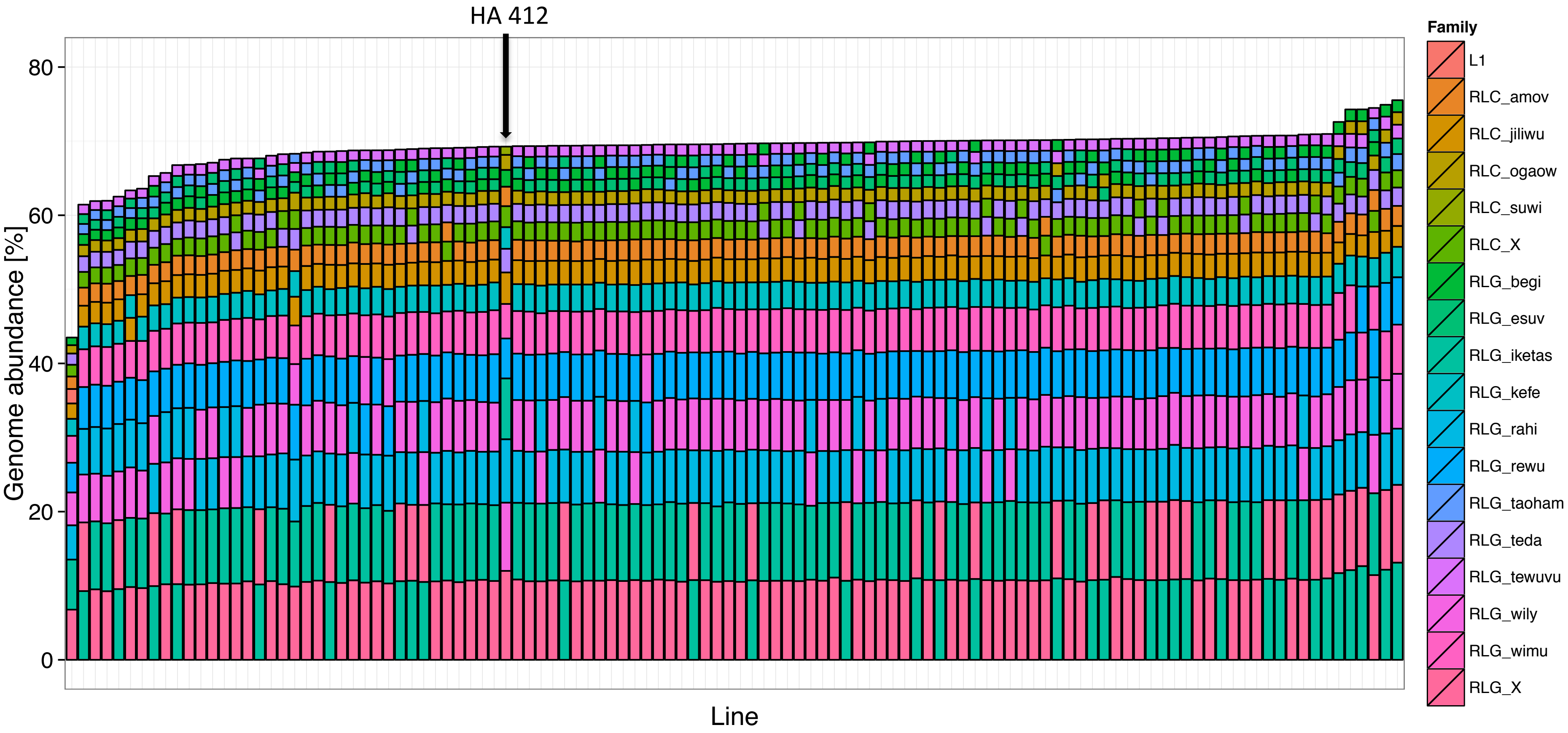
Compare 288 sunflower breeding lines representing USDA, INRA, Oil and non-Oil and “core 12” (Mandel et al. 2013)

1) *What is the total level of TE variation between lines?*

2) *Is there significant variation in TE family abundance between lines?*

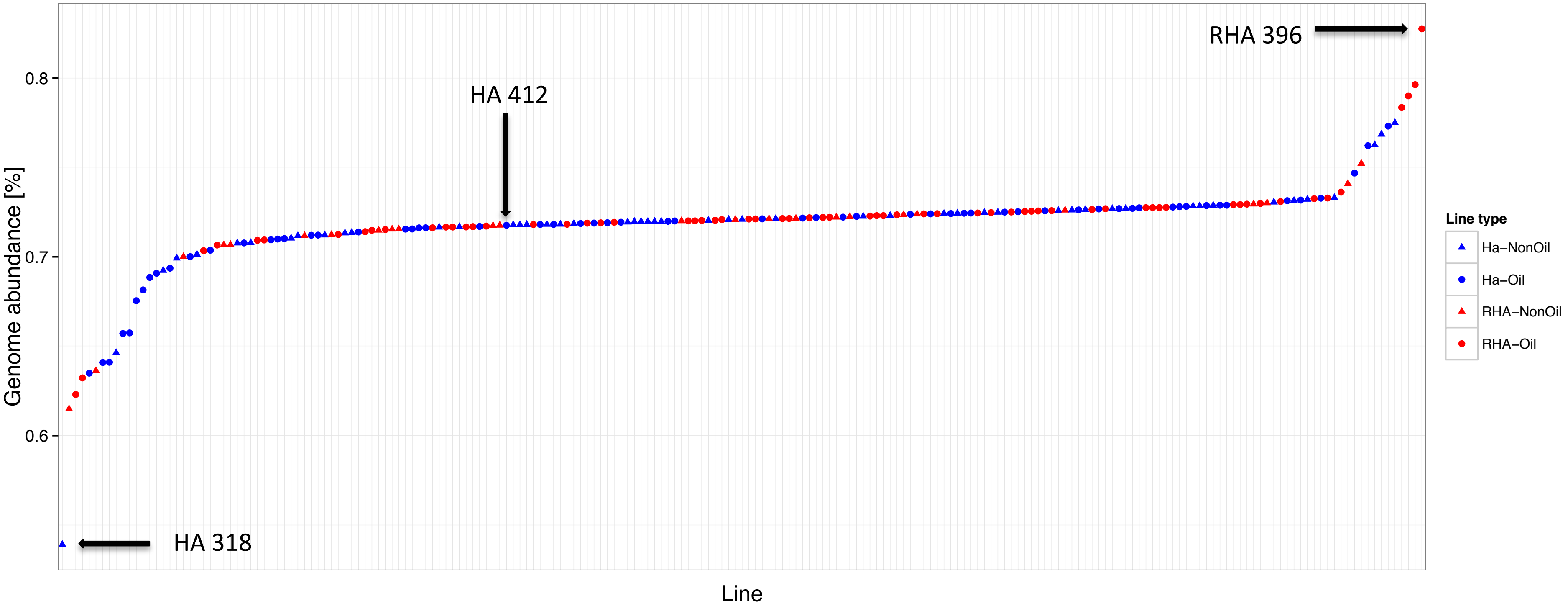
3) *Is there any difference between Oil and non-Oil lines or heterotic groups?*

Intraspecific variation in *Helianthus annuus*

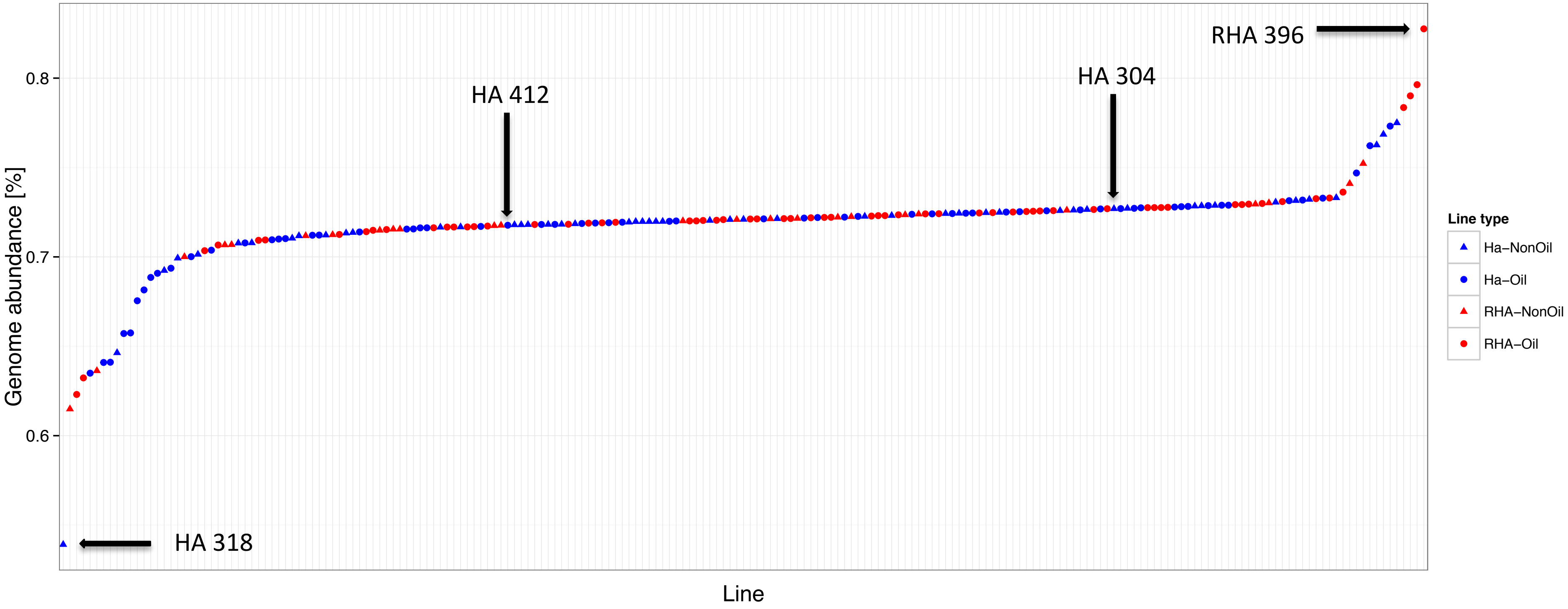


HA lines

Intraspecific variation in *Helianthus annuus*



Intraspecific variation in *Helianthus annuus*



Intraspecific variation in *Helianthus annuus*

Mann-Whitney-Wilcoxon test for sign. family-level variation between Oil and non-Oil lines

Measured the total variation between lines by source

Significant variation between 64 Oil/non-Oil lines

Lines vary by ~30% in total repeat abundance

Min ~48.1% - HA 318
Max ~83.9% - RHA 396

Conclusions

Compare 288 sunflower breeding lines representing USDA, INRA, Oil and non-Oil and “core 12” (Mandel et al. 2013)

1) *What is the total level of TE variation between lines?*

Conservatively, 30% total variation.

Conclusions

Compare 288 sunflower breeding lines representing USDA, INRA, Oil and non-Oil and "core 12" (Mandel et al. 2013)

2) *Is there significant variation in TE family abundance between lines?*

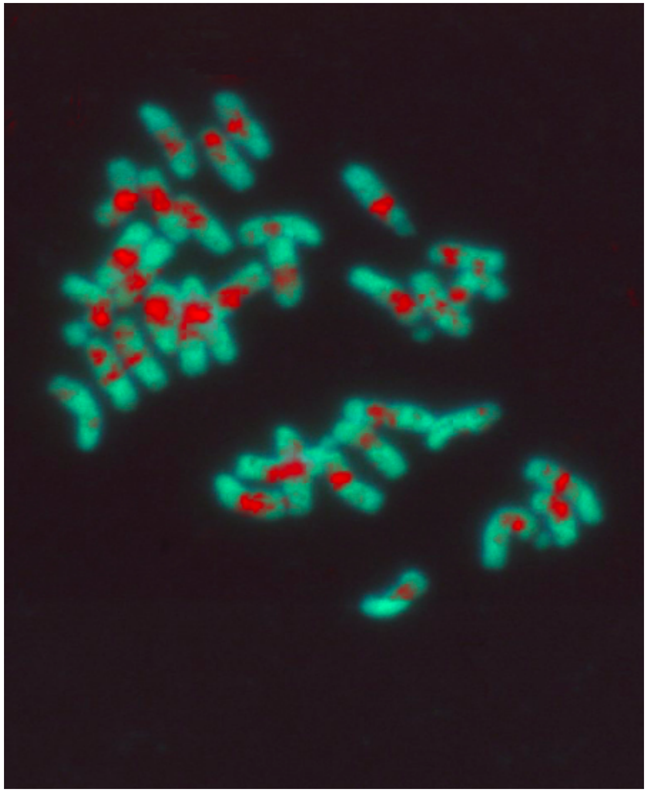
Significant variation exists for DNA TEs and retrotransposons.

3) *Is there any difference between Oil and non-Oil lines, or heterotic groups?*

Approx. 25% lines differ significantly.

Characterizing centromeres in sunflower

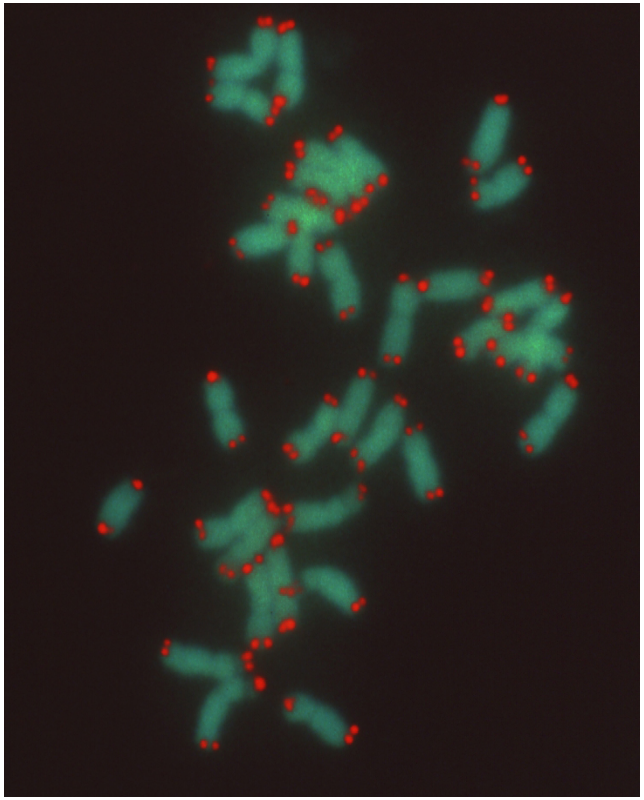
- Centromere cores are composed of satellite repeats and LTR-RTs



- Plant centromeres are a hotspot for unequal homologous recombination (coldspot for recombination)
- Centromeres are highly unstable at the DNA level
- Core repeat in *H. annuus* is a 105 bp subunit

Characterizing telomeres in sunflower

- Telomeres are composed primarily of satellite repeats



- Plants appear to have a conserved 5’-(TTTAGGG)-3’ telomeric repeat

Future directions – major outstanding questions

I. How have DNA transposons contributed to gene function and copy number?

II. How have TEs contributed to phenotypic evolution and ecological diversity?

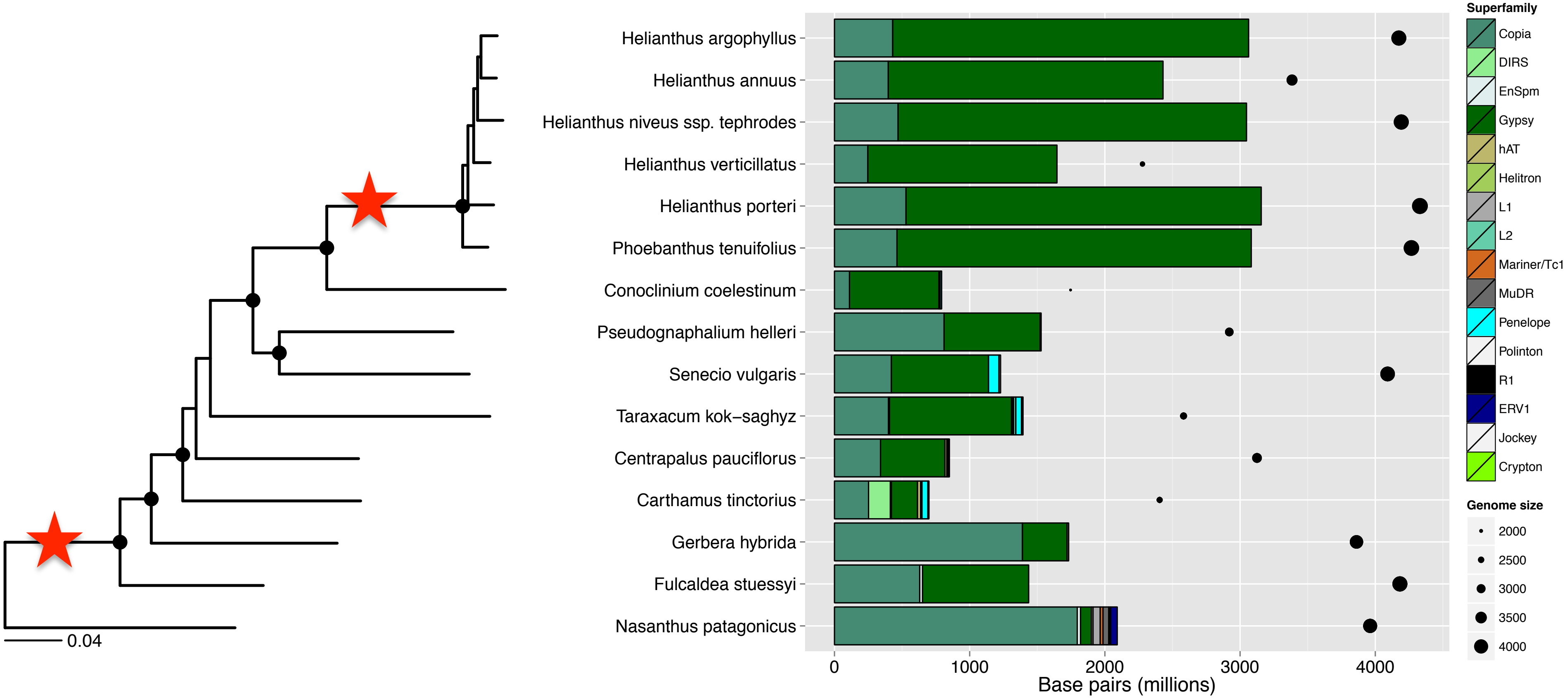
III. Has rapid evolution of centromeres played a role in speciation of the Asteraceae?

Variation between species

I. How do the patterns in *H. annuus* compare with other Asteraceae?

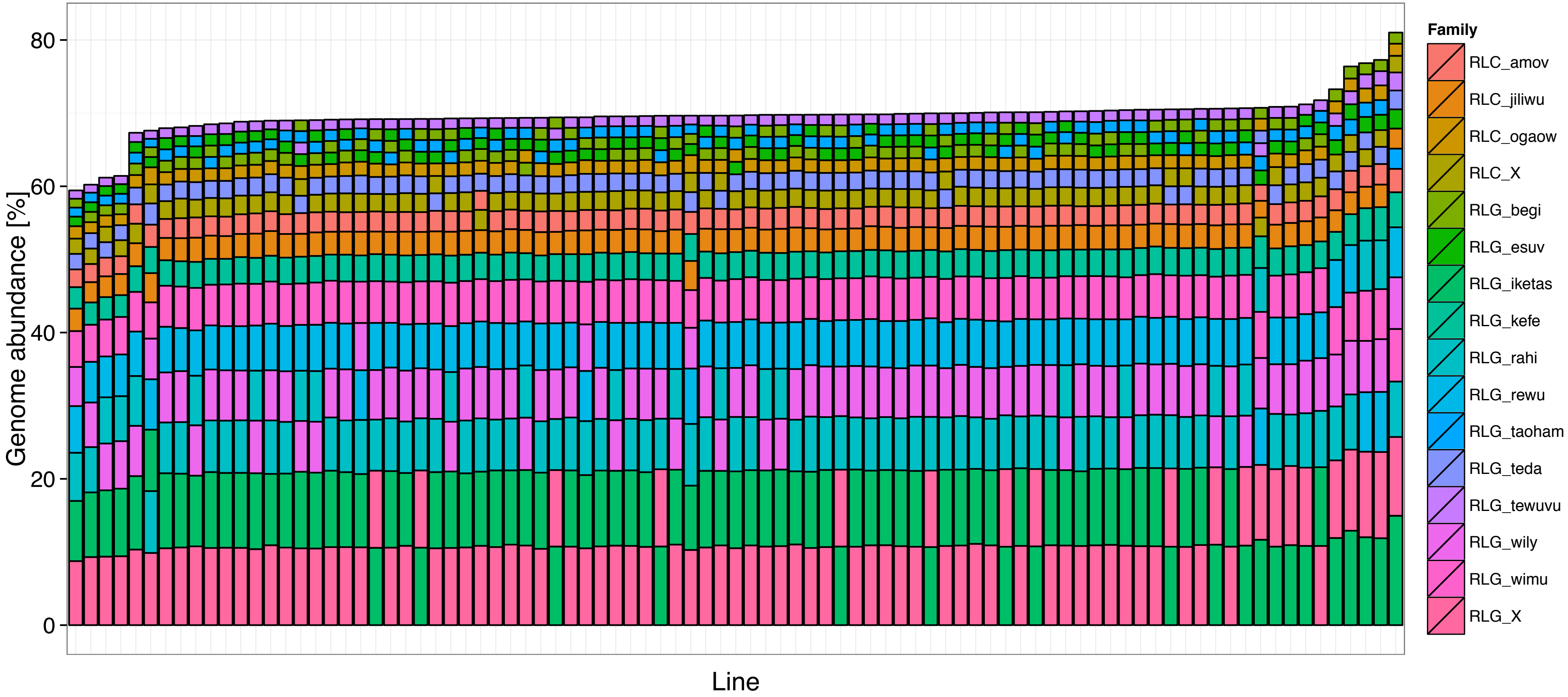
II. In comparison with all TEs, is there a unique history of LTR-RTs in *H. annuus*?

Phylogenetic transitions coincide with genome turnover



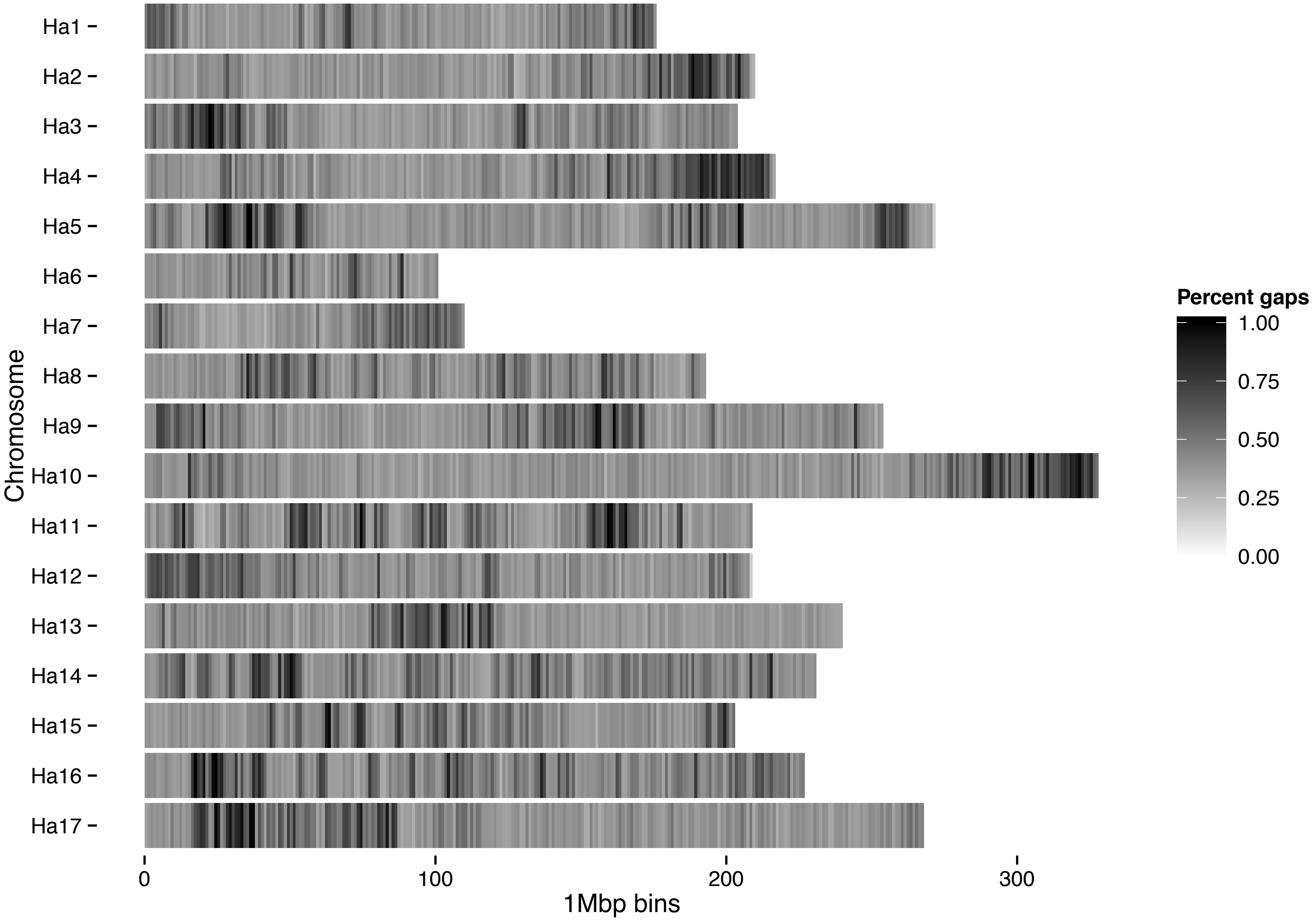
Staton and Burke 2015, *Bioinformatics* (Transposome software)
Staton and Burke 2015, *BMC Genomics*

Intraspecific variation in *Helianthus annuus*

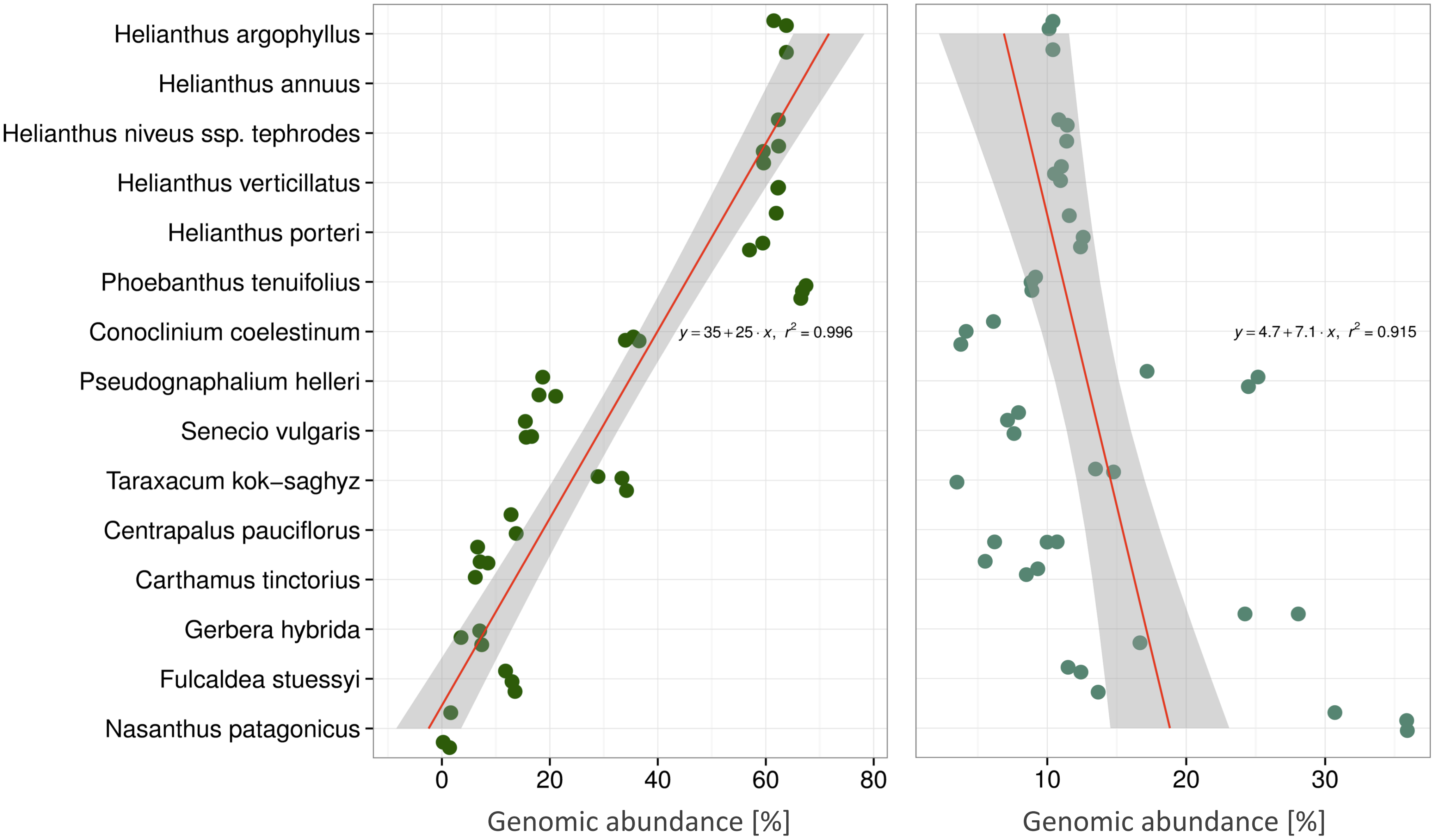


RHA lines

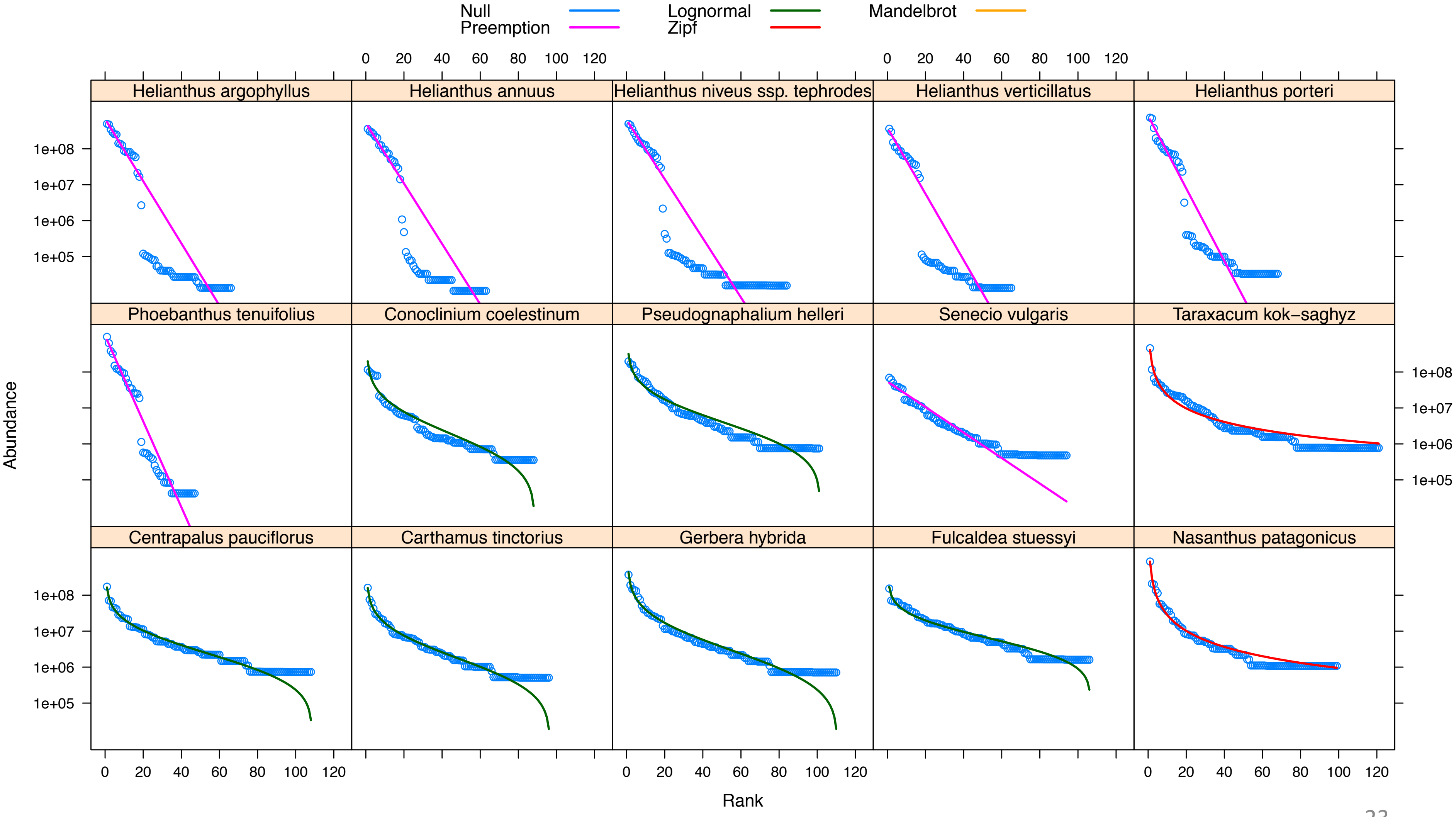
Caveats: ‘gaps’ in our data



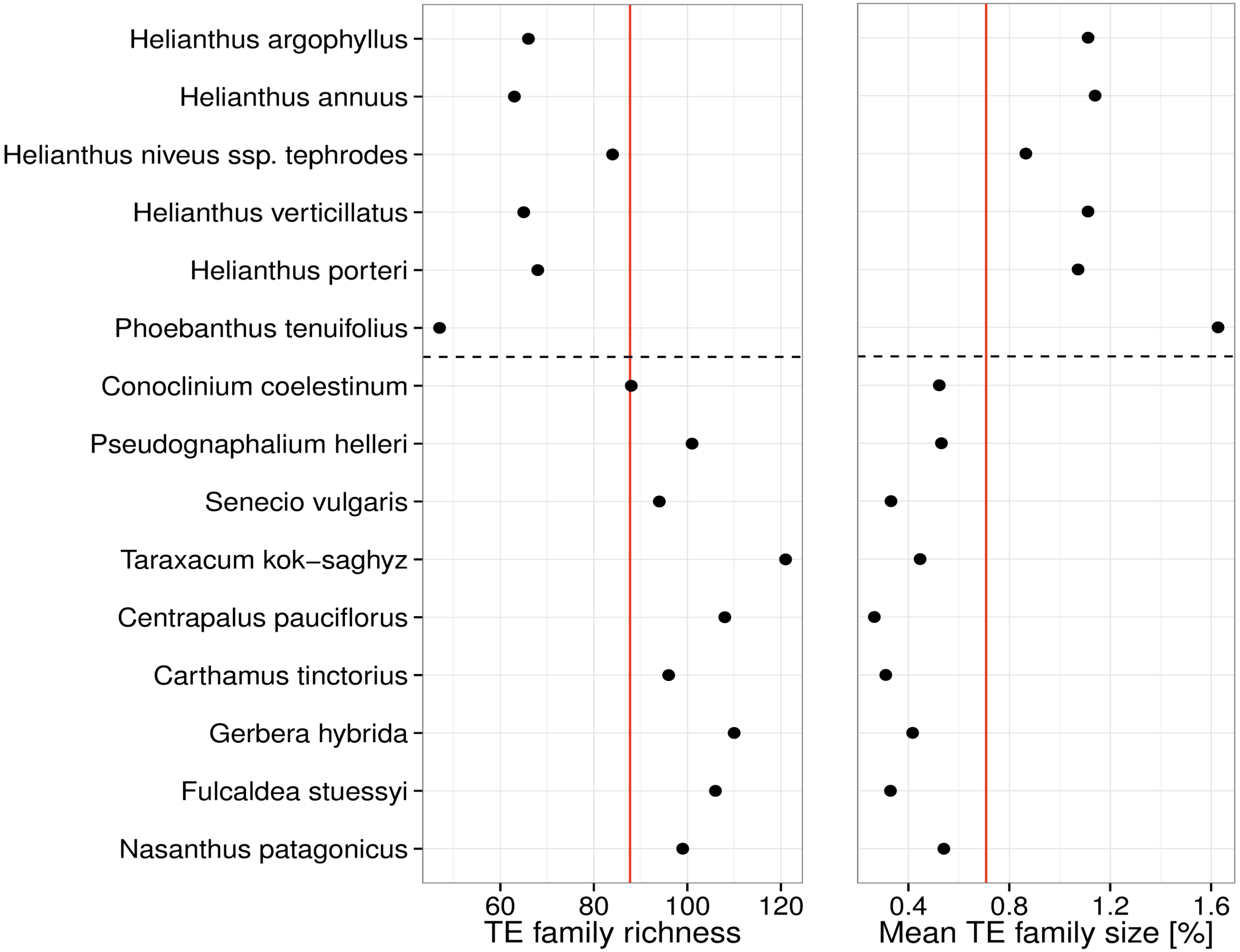
Non-random patterns of change in TE superfamily abundance



RAD plots for each community of TE families



Community level changes in TE properties in the Heliantheae



TE richness is negatively correlated with genome size