Transposable Element Annotation Meeting (TEAM)
April 18-24, 2014
McGill Bellairs Research Institute
Folkestone, St. James, Barbados

TE Annotation Benchmark Proposal

1. Name

Manually Curated Regions of *Arabidopsis lyrata*

2. Authors

Douglas Hoen, Glenn Hickey, Thomas Bureau, Mathieu Blanchette
McGill University, School of Computer Science & Department of Biology
Contact (DH): douglas.hoen@mcgill.ca

3. Description

A detailed manual curation of several short regions of the *A. lyrata* genome based on multiple types of evidence including homology, repetitiveness, structure, functional genomics (e.g., siRNA), and comparative genomics (e.g., orthology).

4. Specification

	Description	Comments
Type	Real	20 x 100-kbp
Primary Uses	Sensitivity, specificity	Initially created to assess performance of TE HMM
Additional Uses	--	--
Taxa	Arabidopsis lyrata (eudicot)	--
Source	D. Hoen	--
Documentation	TBD	--
Version	0.1	--
Other	Status	Under construction

5. Details

We randomly selected twenty 100-kbp regions of the *A. lyrata* genome (JGI Build 1) and manually annotated all TEs with any evidentiary support based on multiple lines of evidence (see below). For each TE, we qualitatively assessed the certainty of: its existence, the position of its termini, and its superfamily. Evidence considered:

- Homology: RepeatMasker with several libraries (Jurka et al. 2005; Buisine et al. 2008; Hollister et al. 2011; de la Chaux et al. 2012; Haudry et al. 2013); conserved domains (custom)
- Repetitiveness: RepeatModeller, 15-mer depth, self-LASTZ (custom)
- Structural: LTR_FINDER, sliding-window LASTZ (custom), poly-A (custom)
- Functional: siRNA, mRNA
- Comparative: LASTZ orthology chains to eight Brassicaceae spp. (Haudry et al. 2013)
- Other: tandem repeats (TRF), gaps (N's), gene models (FGenesH)

6. References

Buisine N, Quesneville H, Colot V. 2008. Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets. *Genomics* **91**(5): 467-475.

de la Chaux N, Tsuchimatsu T, Shimizu KK, Wagner A. 2012. The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. *Mobile DNA* **3**(1): 2.

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet* **45**(8): 891-898.

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci USA* **108**(6): 2322-2327.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467.