

An age-of-allele test of neutrality for transposable element insertions

Justin Blumenstiel (1)

Fiona He (2)

Casey M. Bergman (2)

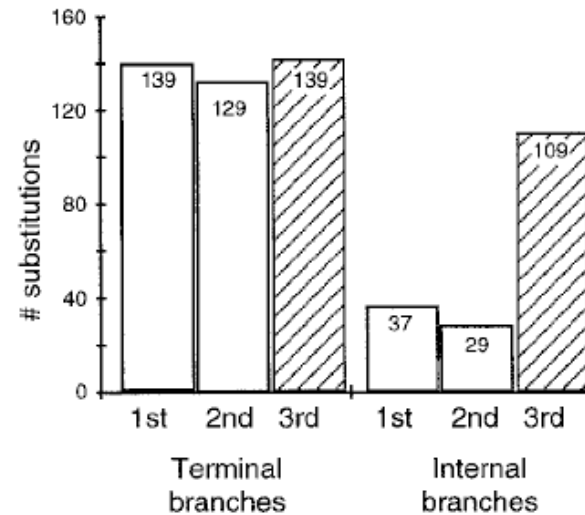
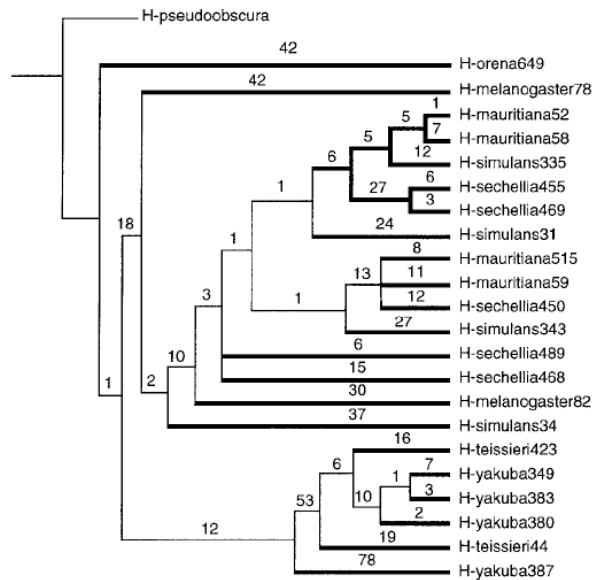
(1) Dept. of Ecology & Evolutionary Biology, University of Kansas

(2) Faculty of Life Sciences, University of Manchester

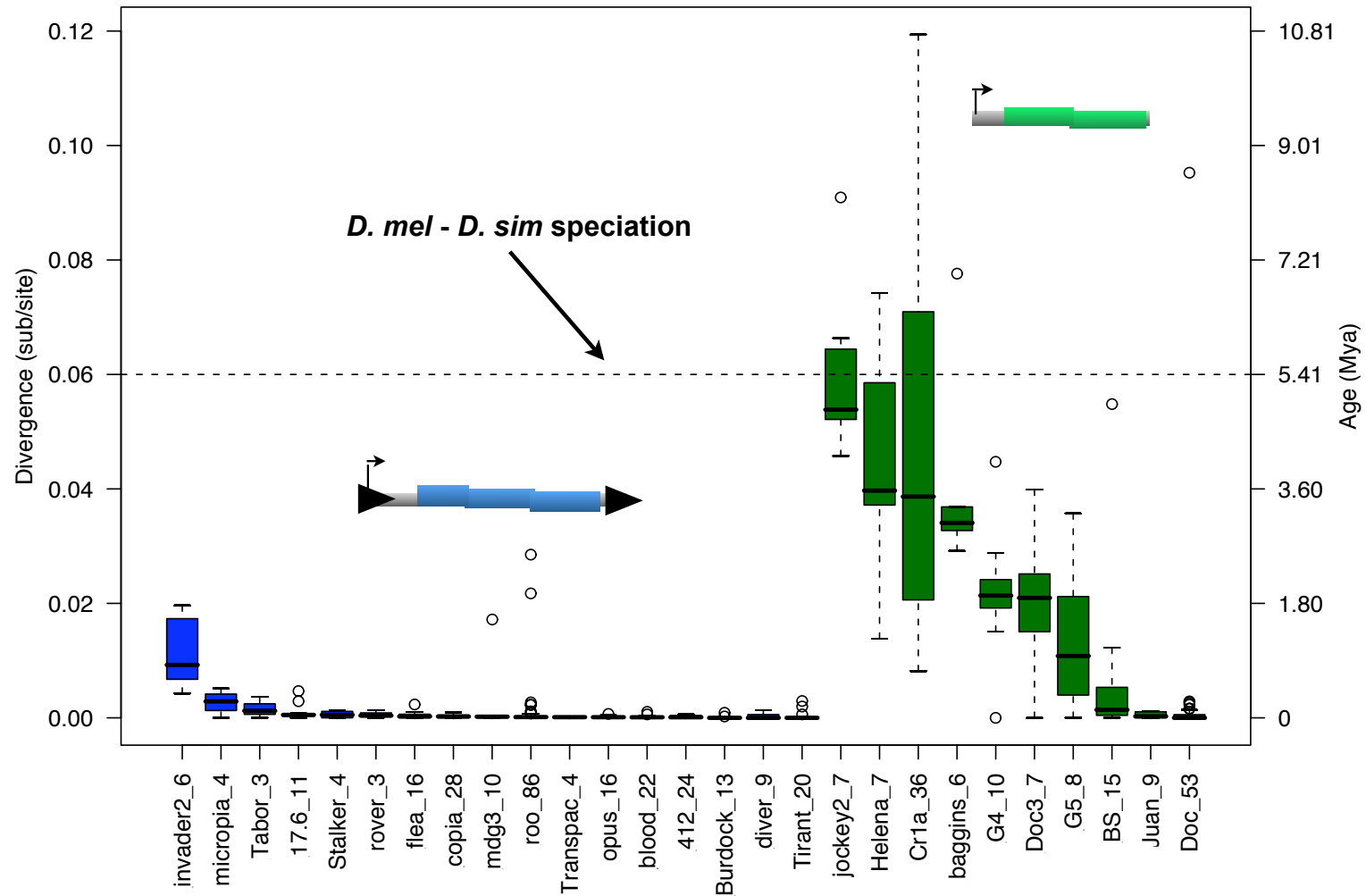
A brief introduction to models of TE evolution

- **Selfish DNA sequences, intra-genomic parasites**
- **Transposition rates \gg excision rates**
- **Equilibrium maintained by transposition-selection balance**
- **Mechanism of negative selection is debated**
- **TE insertions at low frequency because of negative selection**
- **Recent genomic evidence casts doubt on the assumption of equilibrium and inference of negative selection**

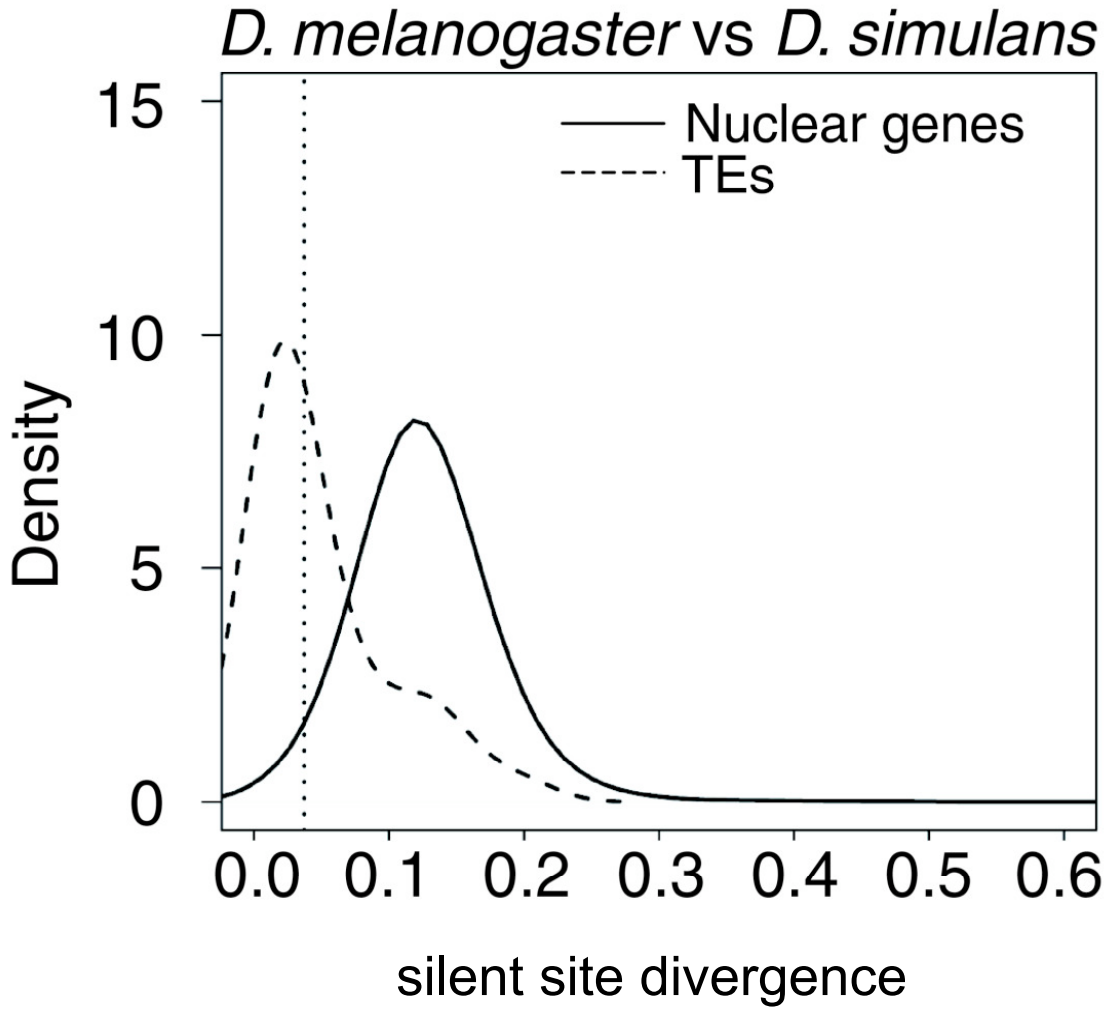
Estimating the age of 'pseudogene-like' retrotransposon insertion alleles



Retrotransposon demographics in *D. melanogaster*



Horizontal transfer of *D. melanogaster* TE families



Project aims

- **Develop a non-equilibrium model of neutral TE evolution that relaxes the assumption of a constant TE insertion rate.**
- **Obtain allele frequency data for a large sample of TEs in ancestral and derived populations of *D. melanogaster*.**
- **Test whether observed TE allele frequencies are consistent with ages of TE insertion estimated from genomic data to infer forces controlling TE evolution.**

An age-of-allele model for TE insertions

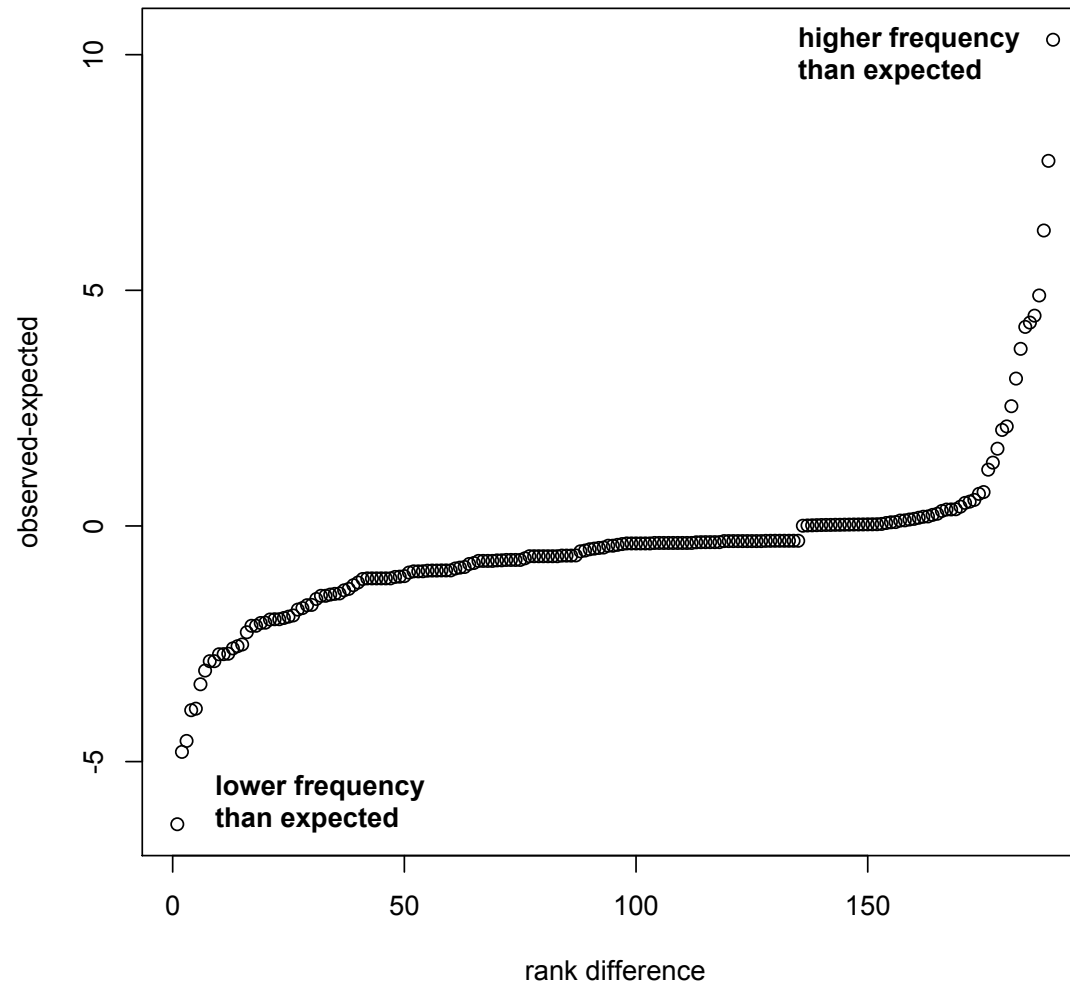
- Question: what is the probability that an allele of age t is present in i copies in a sample of n chromosomes?
- Calculate probability of i descendants from a single ancestor given j ancestors (Feller 1957)
- Calculate probability of j ancestors at time t under standard neutral model (Tavare 1984)
- Calculate probability of insertion at time t given s substitutions in a fragment of length l under Poisson process (Bayes 1763)

$$P(i | s_l, n) = \sum_{j=1}^n P(i | j, n) \left(\int_0^{\infty} P(j | t, n) P(t | s_l) dt \right)$$

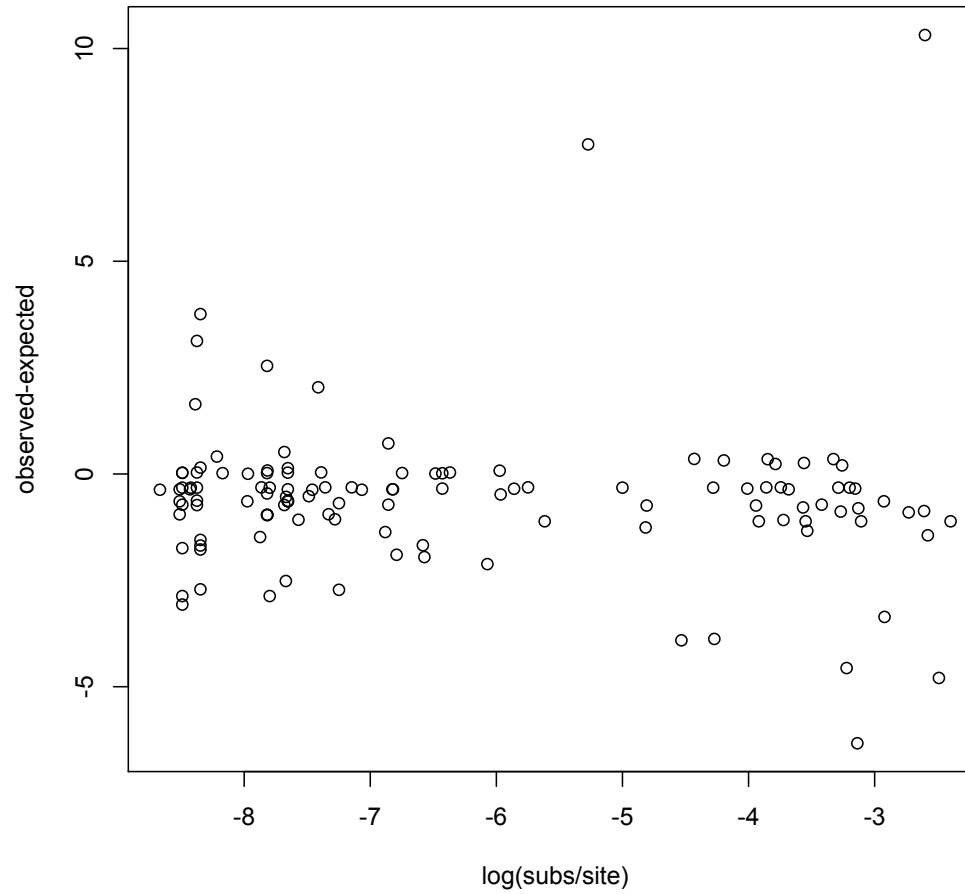
Allele frequency data for TE insertions

- 190 loci (90 LTR and 100 non-LTR)
- 2 PCR per loci per strain (TE+flank / L+R flanking regions)
- 12 strains from 2 populations - Zimbabwe (from Stephan Lab) & North Carolina (from Mackay Lab)
- Insertion in genomic sequence is included as 13th allele to account for ascertainment bias
- Individual strain allele frequency data consistent with pooled strain allele frequency data from Gonzalez *et al.* (2008)

Fit of expected allele frequency under neutral model to observed frequency in North Carolina



Expected allele frequency fits observed allele frequency over a wide range of ages



Preliminary observations

- Majority of TE insertions in North Carolina are at or close to expected frequency given age since insertion under neutrality
- Some loci deviate strongly from predicted frequency and may reflect loci under positive and negative selection
- Model accurately predicts observed allele frequency over wide range of insertion ages
- Model parameterized with current estimate of African population size leads to poor predictions but yields better fit with ancestral population size

Ongoing and Future Work

- Use model to generate maximum likelihood estimate of N_e under assumption that insertion alleles are neutral
- Analysis of the fit of model to data according to:
 - TE class
 - TE family
 - X vs. autosome
 - recombination
- Resolution of best summary statistic(s) to assess global fit of the model to the data

Acknowledgements

Justin Blumenstiel
Fiona He

Nicolas Svetec
Wolfgang Stephan

