

Evolutionary drivers of transposable elements accumulation in swallowtail butterflies and passerine birds

Alba Marino^(1,2), Pelle Scholten⁽¹⁾, Eliette Reboud⁽¹⁾, Bérénice Lafon⁽¹⁾, Marie-Ka Tilak⁽¹⁾, Fabien Condamine⁽¹⁾, Annabelle Haudry⁽³⁾, Anna-Sophie Fiston-Lavier^(1,4), Benoit Nabholz⁽¹⁾

(1) University of Montpellier, ISEM, UMR 5554, CNRS, IRD, Montpellier, France; (2) University of Pavia, Department of Biology and Biotechnology, Pavia, Italy; (3) University Claude Bernard Lyon 1, LBBE, UMR 5558, CNRS, VAS, Villeurbanne, France; (4) Institut Universitaire de France, Paris, France

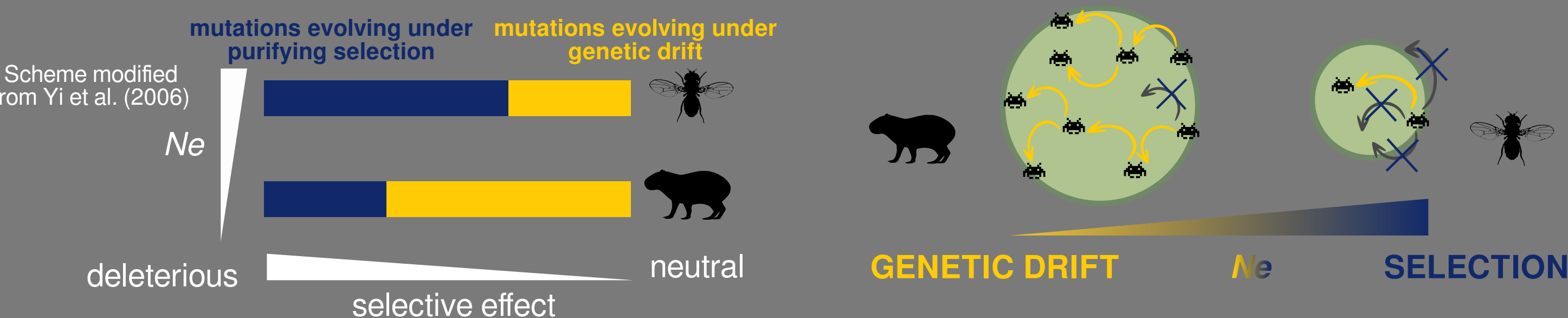
✉ alba.marino@unipv.it

Background

Animals display a dramatic variation in their **genome size** and **transposable element (TE) content**, but the evolutionary causes behind such different patterns remain debated

The Drift-Barrier hypothesis

- The effective size of a population (N_e) determines how efficiently slightly deleterious DNA insertions are eliminated
- Species impacted by **high genetic drift levels (low N_e)** are expected to accumulate more TEs and evolve larger genomes



- Although originally detected across the tree of life (Lynch and Conery, 2003), this pattern was not found across animal genomes

Marino et al. *eLife* 2025

Questions & Approach

- Are genome size and TE content variation explained by different genetic drift levels at **shorter evolutionary scales**?
- Is reduced N_e reflected in increased accumulation of recent TEs?

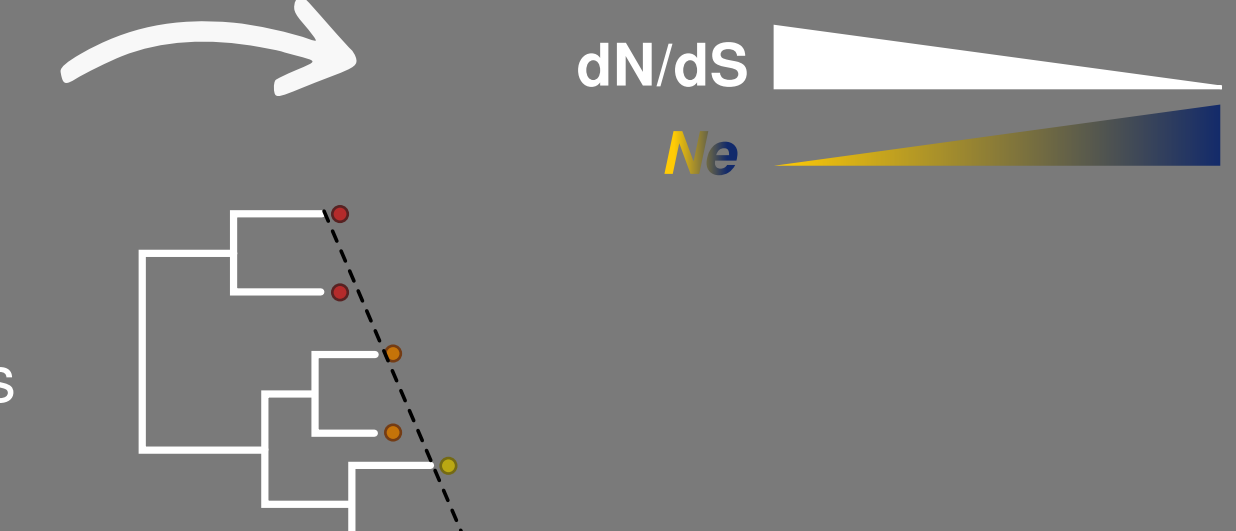
- We leverage two taxa characterized by **contrasted genome sizes and contrasted N_e**



- N_e is deduced from genome-wide dN/dS

more non-synonymous substitutions accumulate at small N_e
synonymous substitutions tend to accumulate neutrally

Relationship between N_e and its proxy:



- Species relatedness is accounted for in correlations (phylogenetic generalized least squares, PGLS)

Results

Papilionidae

- PGLS of dN/dS as predictor of genome size and TE content

Genome size/TE content - dN/dS	Slope	Adjusted-R ²	p-value
Genome size	0.954	-0.029	0.745
All TEs	5.084	0.001	0.315
LINE	1.971	-0.030	0.788
SINE	-30.181	0.012	0.280
LTR	1.559	-0.032	0.889
DNA	7.258	-0.001	0.334
RC	12.325	-0.018	0.517

The remarkable genome size and TE content variation are not associated to dN/dS variation (N_e proxy)

- PGLS of dN/dS as predictor of TE gain rate on terminal branches and full-length (FL) TE copy number

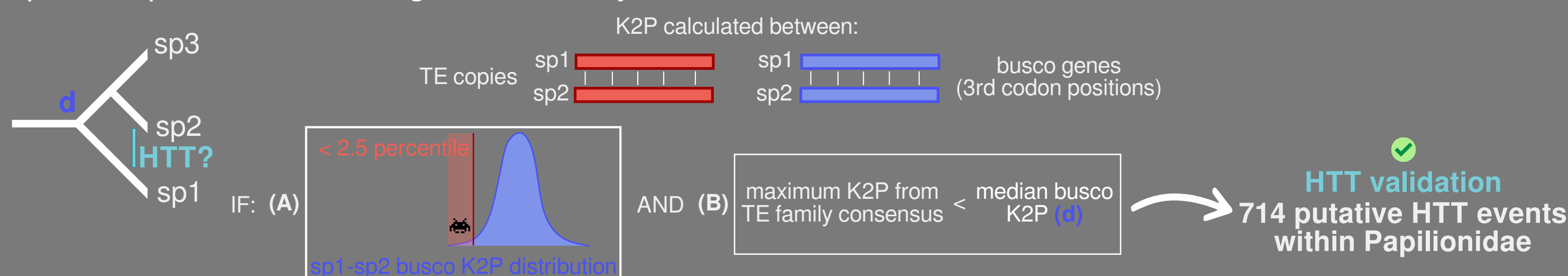
TE gain rate - dN/dS	Slope	Adjusted-R ²	p-value
All TEs	11.492	0.056	0.149
LINE	8.858	0.010	0.284
SINE	51.490	0.009	0.248
LTR	13.951	-0.008	0.370
DNA	13.270	0.076	0.114
RC	8.017	-0.030	0.54

FL TE copies - dN/dS	Slope	Adjusted-R ²	p-value
All TEs	-3.680	0.044	0.127
LINE	-15.426	-0.008	0.400
SINE	-9.019	-0.047	0.762
LTR	-4.861	-0.030	0.778
DNA	-0.023	0.007	0.276
RC	3.199	-0.031	0.860

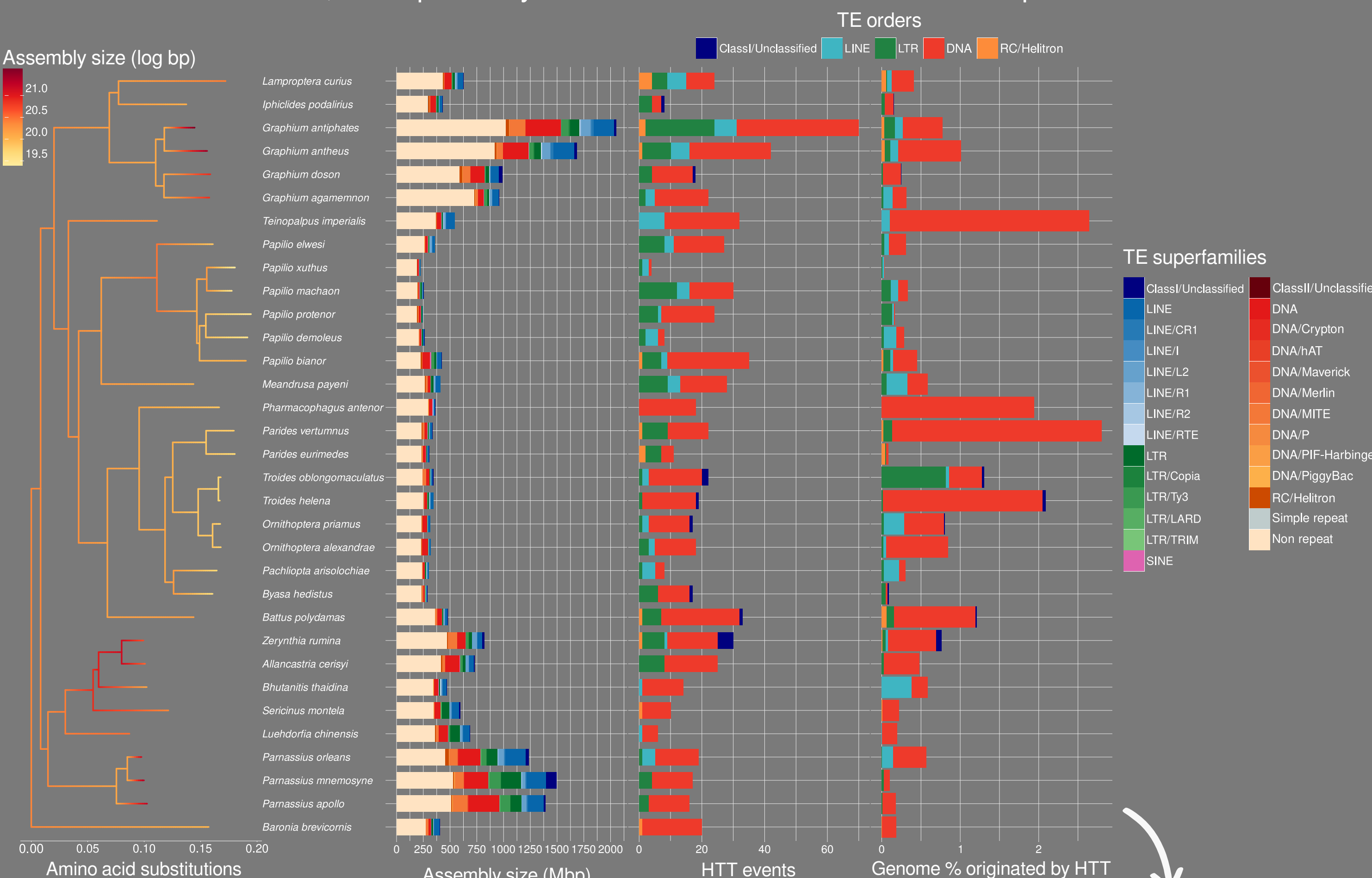
TE accumulation rate and the number of FL elements are not predicted by genetic drift levels either

Lepidoptera are known to be a hotspot clade for horizontal transfer of TEs (HTT) (Reiss et al. 2019): are larger genomes and TE amounts related to the frequent exchange of TE families across species?

- HTT between two species were validated if (A) the Kimura 2P divergence (K2P) between TE families is lower than expected according to the comparison of vertically inherited orthologous genes, and (B) no copies older than the K2P estimated at the species split are found in a given TE family

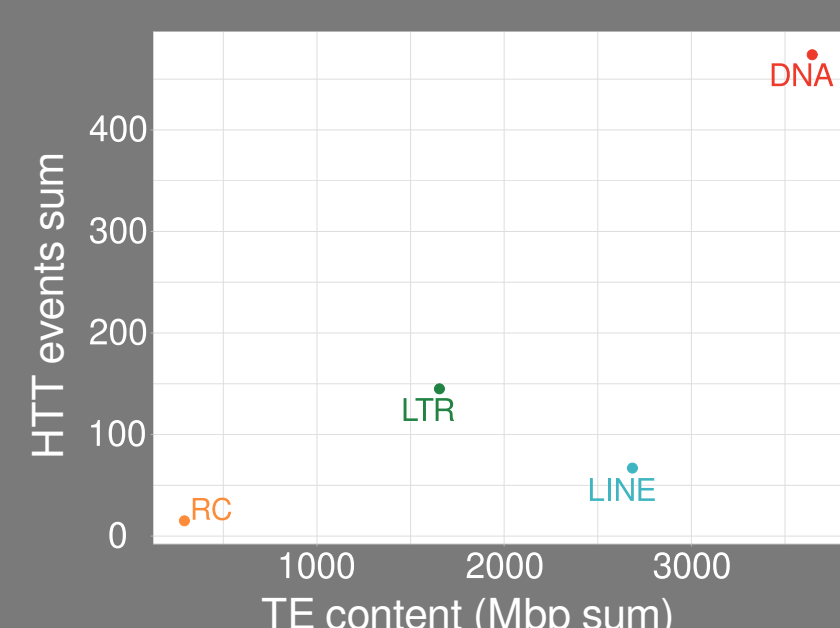


- Genome size variation, TE superfamily annotations and evaluation of HTT impact



Up to almost 3% of genome is made up by laterally transferred, mostly DNA elements

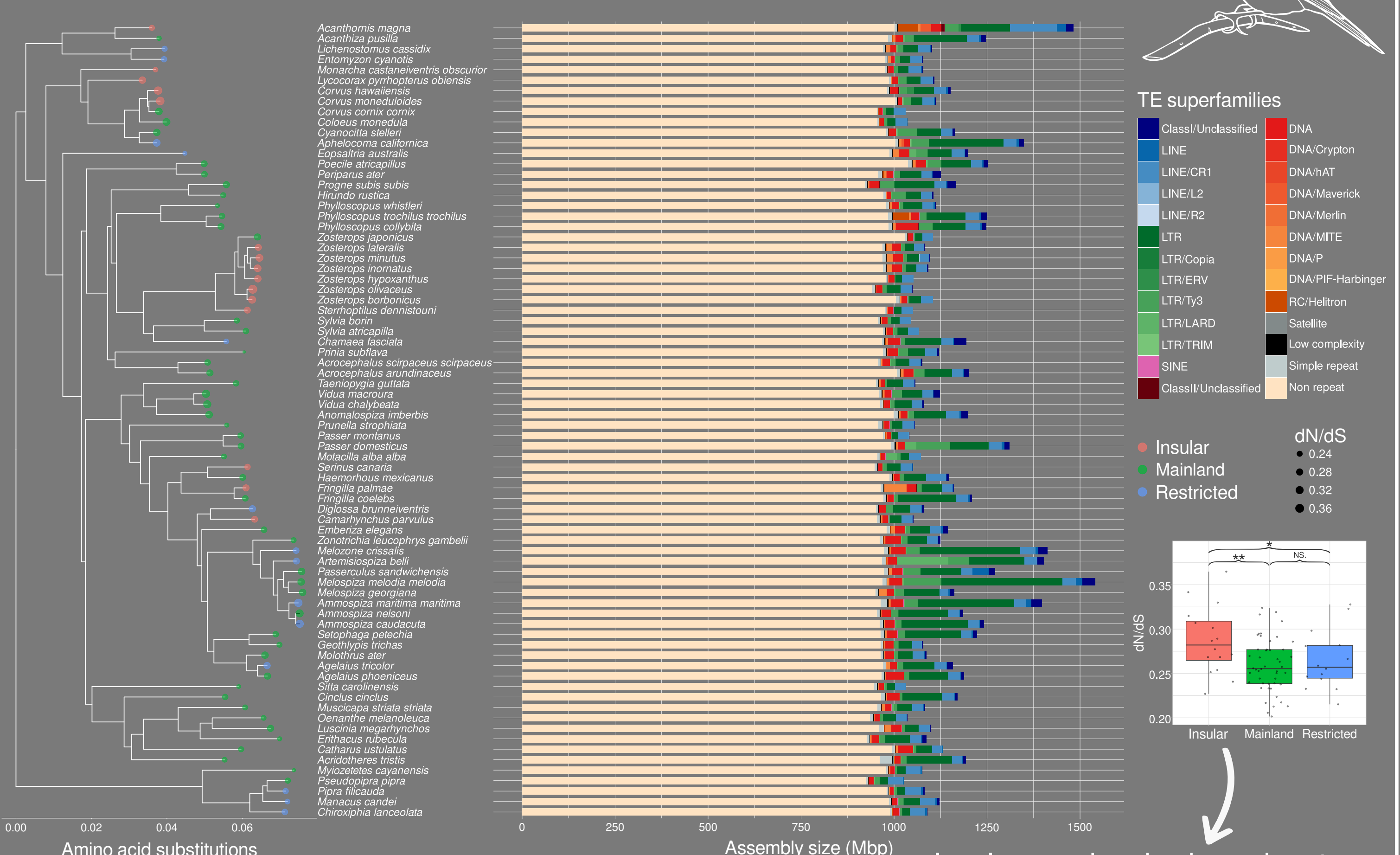
Species with larger genomes and TE contents undergo more HTT events and present larger portions occupied by laterally transferred TE families



However, very few LINEs are exchanged between species, in spite of being the second most abundant order across Papilionidae

Passeriformes

- Genome size variation, TE superfamily annotations and dN/dS values at terminal branches



Insular species do show signatures of increased genetic drift

- PGLS of dN/dS as predictor of genome size, TE content and TE accumulation rate on terminal branches

Genome size/TE content - dN/dS	Slope	Adjusted-R ²	p-value
Genome size	0.385	0	0.317
All TEs *	3.976	0.044	0.039
LINE	1.242	0.004	0.254
SINE	-3.540	-0.007	0.443
LTR *	6.032	0.045	0.037
DNA	3.687	0.031	0.071
RC	4.210	-0.007	0.503

Higher genetic drift (large dN/dS) is associated to larger TE loads, but not to larger genomes

The rate of TE accumulation on terminal branches is also positively correlated to drift levels

- PGLS of dN/dS as predictor of copy number and accumulation rate of full-length (FL) TEs

FL TE copies - dN/dS	Slope	Adjusted-R ²	p-value
All TEs **	2.226	0.125	0.001
LINE	2.771	0.025	0.092
SINE	-3.093	-0.016	0.809
LTR *	2.122	0.059	0.020
DNA	1.829	0.009	0.204
RC *	6.347	0.057	0.022

The accumulation of FL, recently inserted elements is even more strongly positively correlated with dN/dS

Conclusions

- In passerine birds, an increased accumulation of TEs is associated with higher drift levels, notably when only FL insertions are considered: this suggests that small N_e promote the fixation of young, putatively active and more deleterious TEs
- Such drift-barrier effect is not observed anymore for genome size, likely as a result of the very quick erosion of avian TEs
- No evidence of N_e affecting genome size or TE content variation is found in Papilionidae. HTT is pervasive and more frequent in larger genomes, but does not provide a global explanation for the proliferation of all TE orders

Take home messages

Stronger genetic drift can drive TE accumulation over short evolutionary timescales, but this pattern is not consistent among taxa. The drift-barrier hypothesis fails to account for genome size variation as a whole.

References

Lynch M., and Conery J.S. (2003) "The origins of genome complexity." *Science* 302.5649: 1401-1404.
 Yi S.V. (2006) "Non-adaptive evolution of genome complexity." *Bioessays* 28.10: 979-982.
 Reiss D., et al. (2019) "Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot." *PLoS genetics* 15.2: e1007965.