An open problem with transposable element abundance

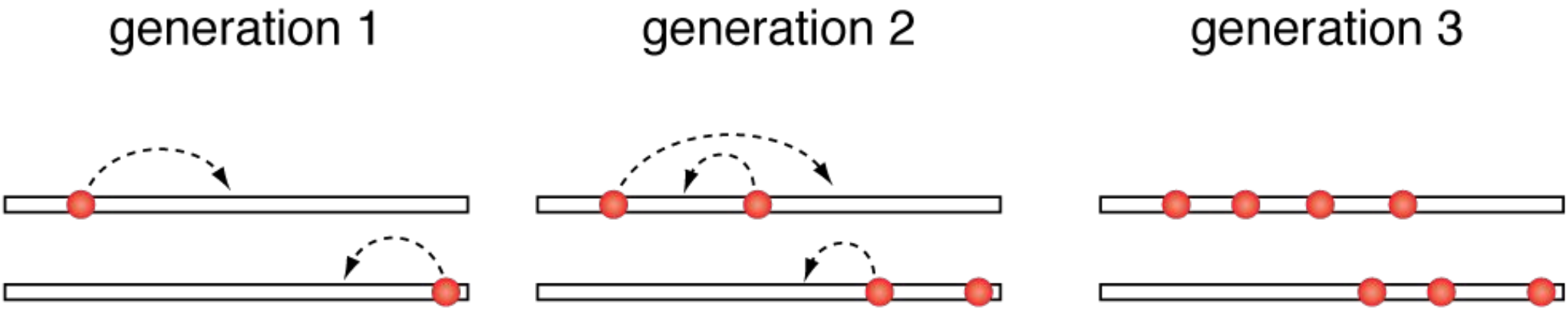
Robert Kofler

Structure of talk

- Some background on TEs
- Insights from a natural TE invasion
- My problem with TE abundance
- Possible solutions

Transposable elements

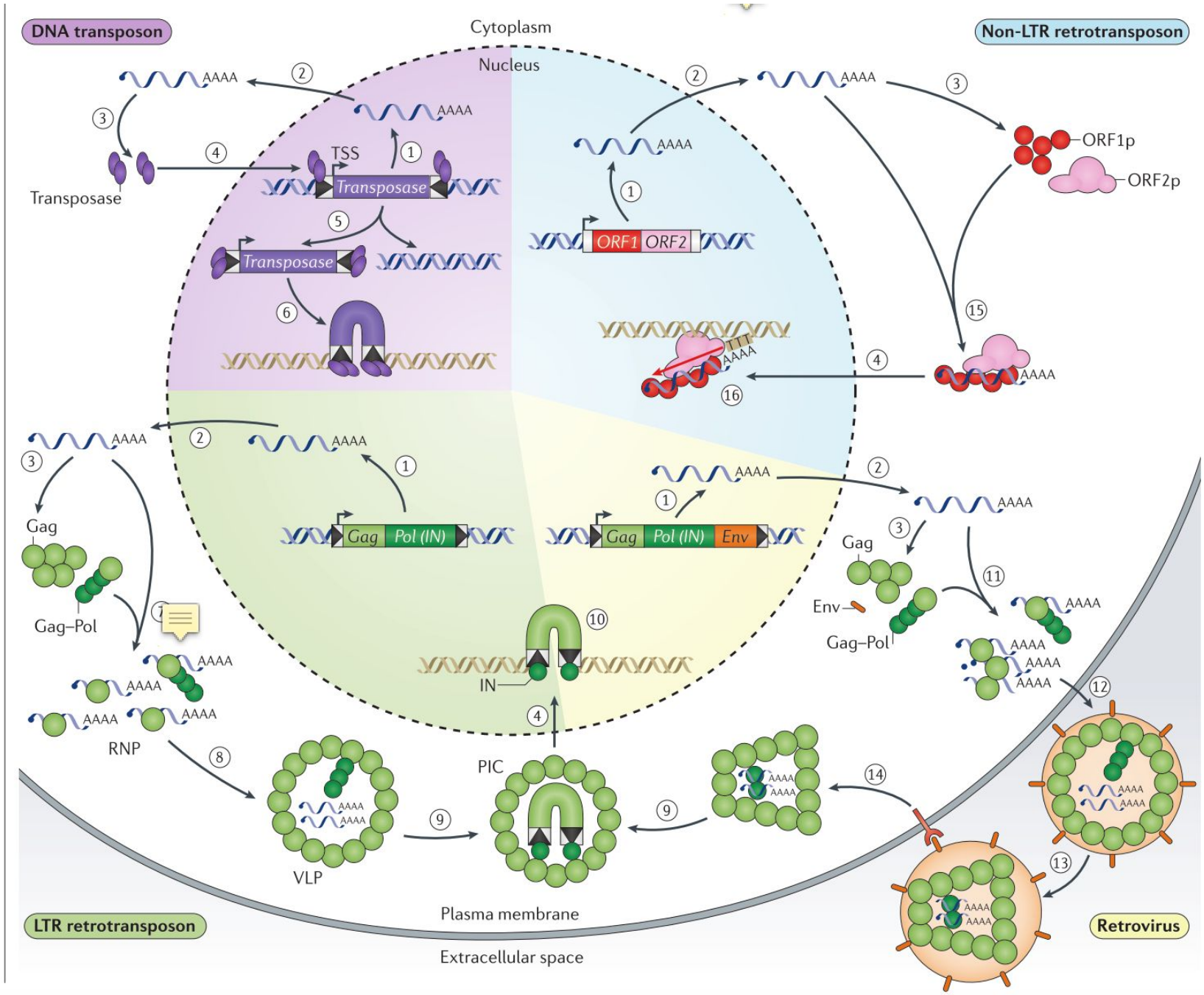
Selfish DNA that proliferates within genomes, even to the detriment of the host



TEs come in different flavours

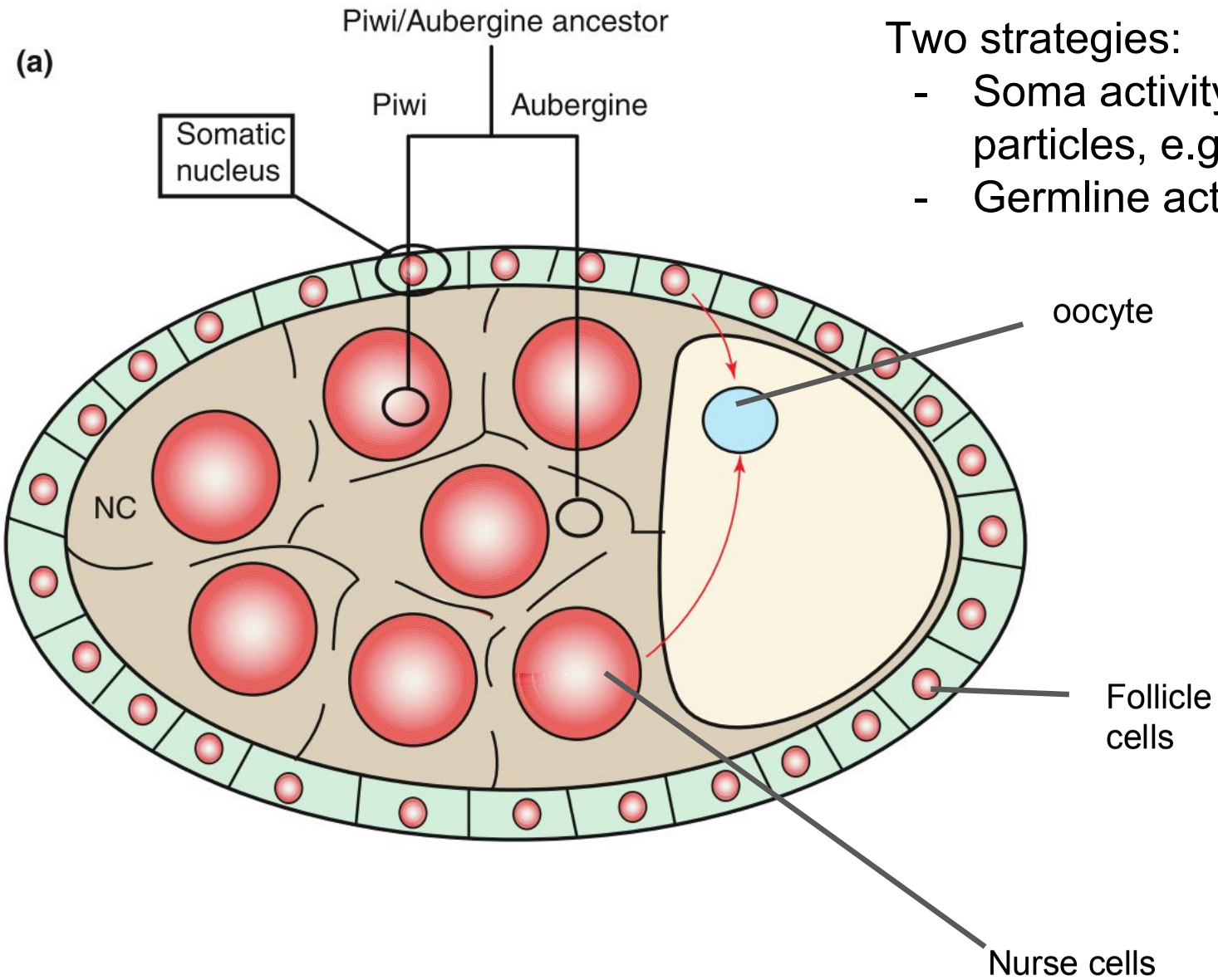
Classification		Structure	TSD	Code	Occurrence
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1–Mariner*	Tase*	TA	DTT	P, M, F, O
	hAT	Tase*	8	DTA	P, M, F, O
	Mutator	Tase*	9–11	DTM	P, M, F, O
	Merlin	Tase*	8–9	DTE	M, O
	Transib	Tase*	5	DTR	M, F
	P	Tase	8	DTP	P, M
	PiggyBac	Tase	TTAA	DTB	M, O
	PIF– Harbinger	Tase* ORF2	3	DTH	P, M, F, O
	CACTA	Tase ORF2	2–3	DTC	P, M, F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P, M, F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M, F, O

Different transposition mechanism



Sultana et al. (2017) NRG

The battleground

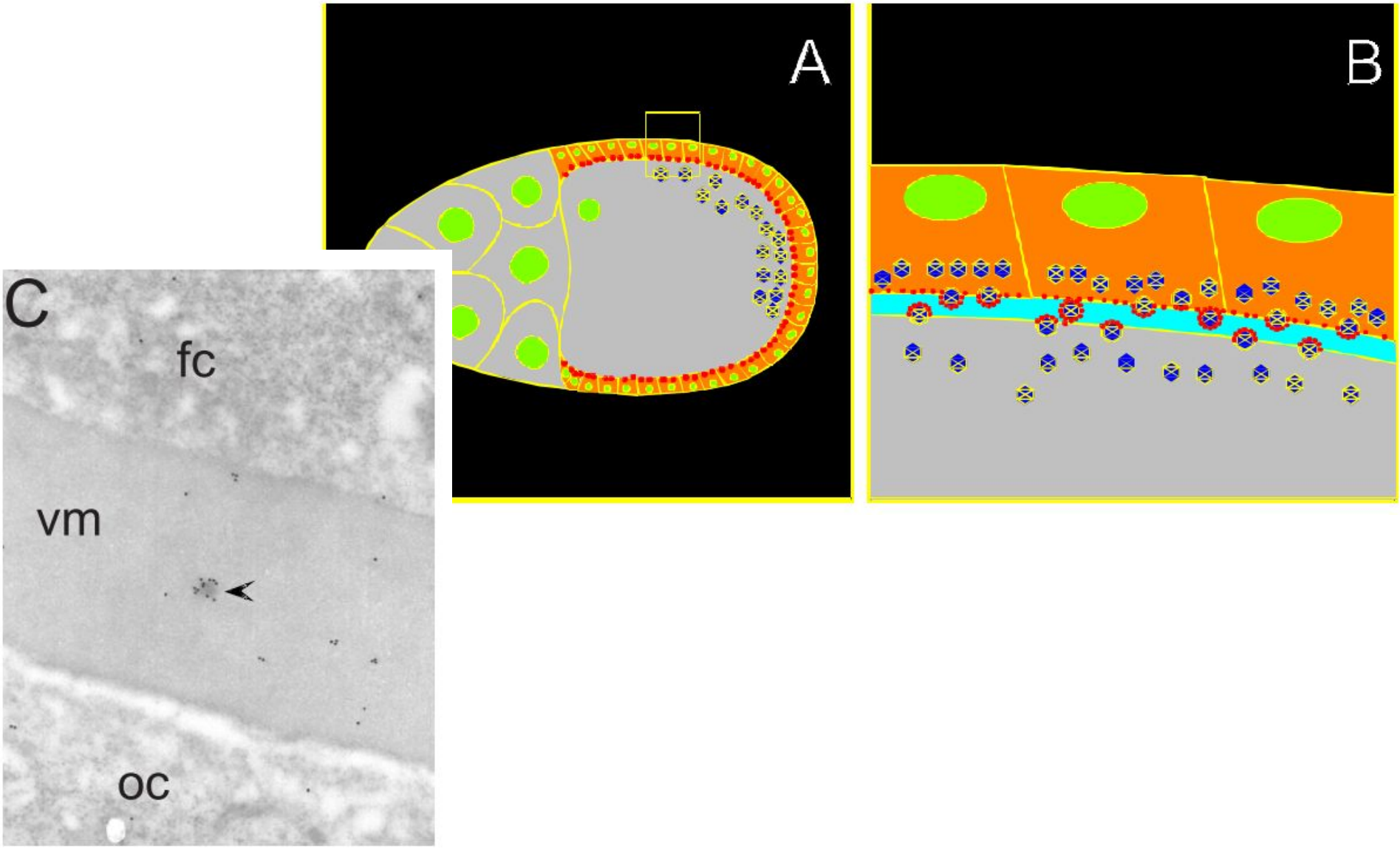


Two strategies:

- Soma activity, virus-like particles, e.g gypsy
- Germline activity

Blumenstiel 2011
Trends in Genetics

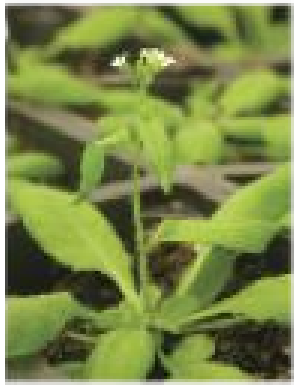
Some virus particles migrate from soma to oocyte



TE composition varies dramatically; a) between species

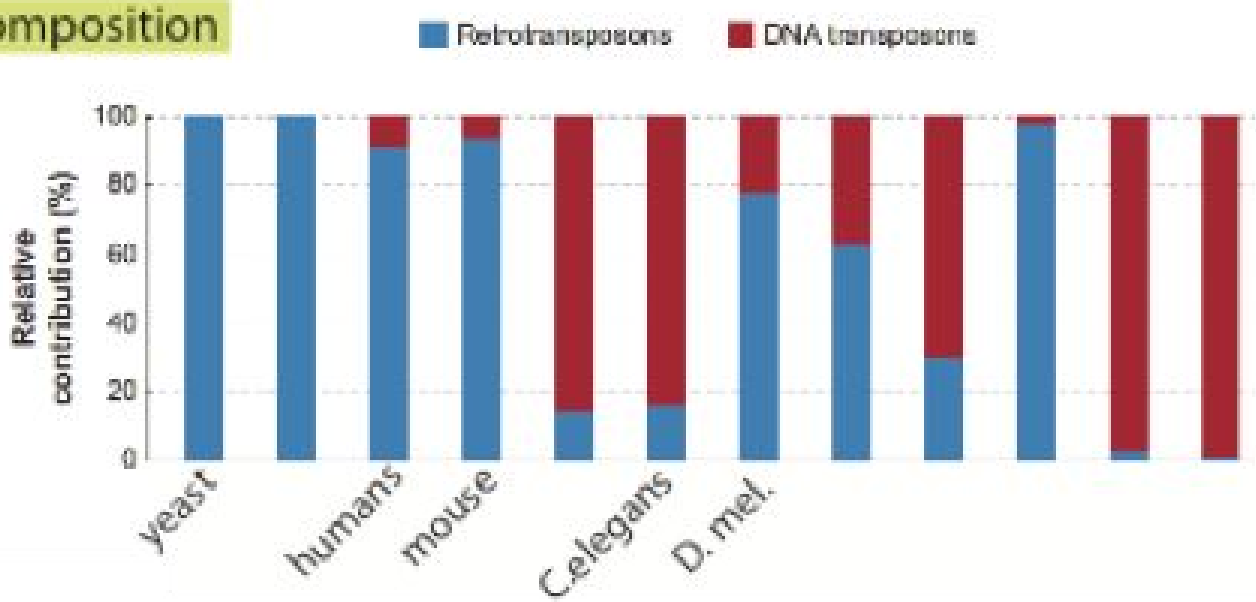
a.) amount

Arabidopsis thaliana : 120Mb



Paris japonica : 149,000Mb

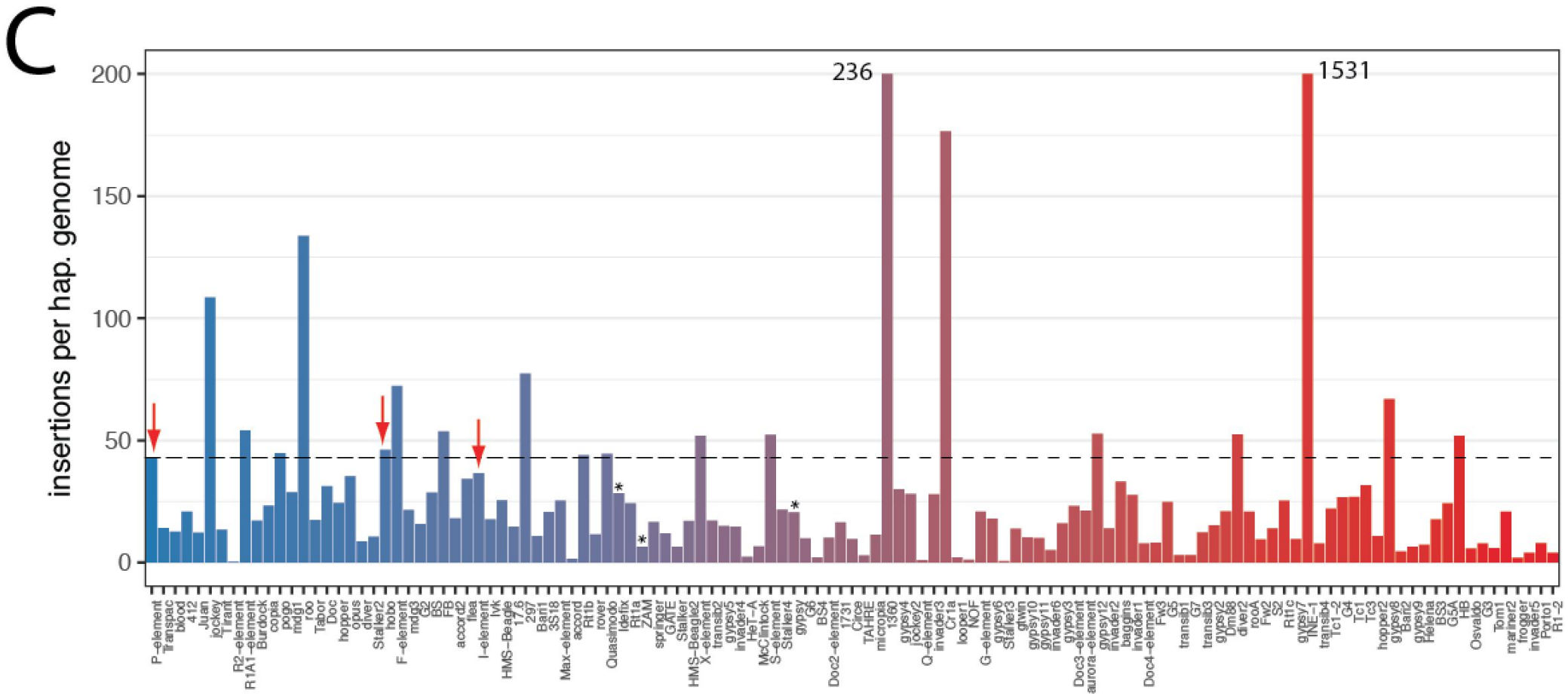
b.) composition



Feschotte 2007

TE composition varies dramatically; a) within species

D. melanogester



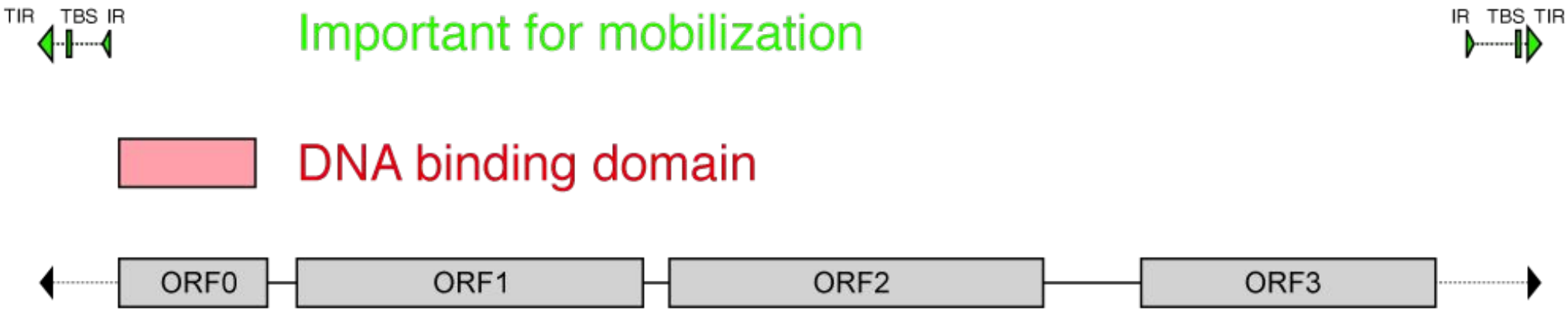
Data from
Kofler 2015 PLoS Genetics

Many open questions

- What are the fitness consequences? DFE: Deleterious, beneficial, neutral?
- How often are TEs invading new species and what is the vector?
- How is the host defending against this onslaught?
- Can the host defence lead to vertical extinction (no active copies survive in a lineage)?
- Is a stable coexistence possible?
- Why is TE abundance so variable between species?

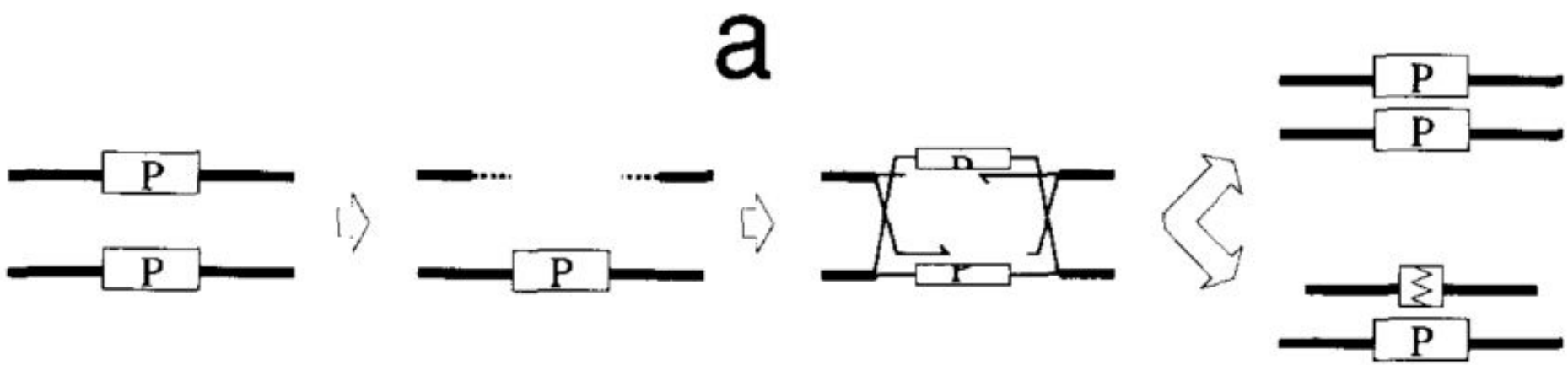
Insights from a natural invasion of the P-element

What is the P-element?
A 2907bp DNA transposon



Lee, Beall and Rio (1998) EMBO Journal
Majumdar and Rio (2015) Microbiology Spectrum

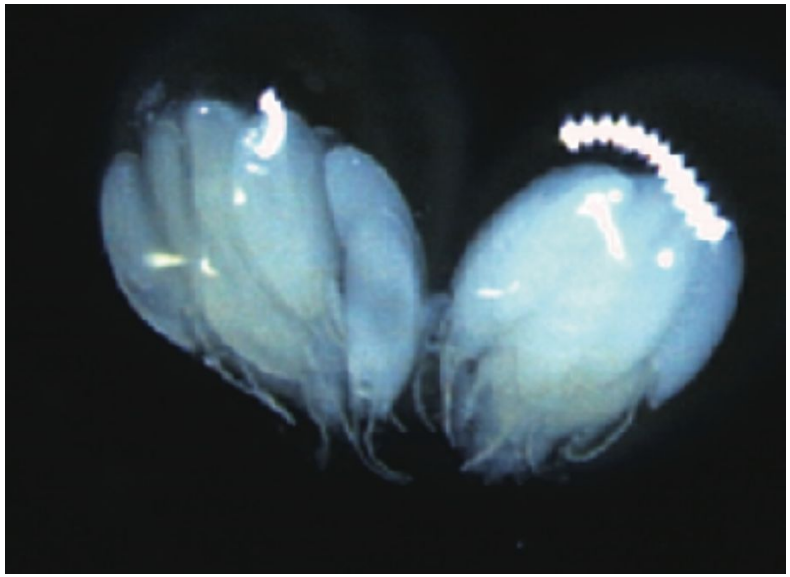
How do cut-and-paste TEs manage to increase in copy numbers?

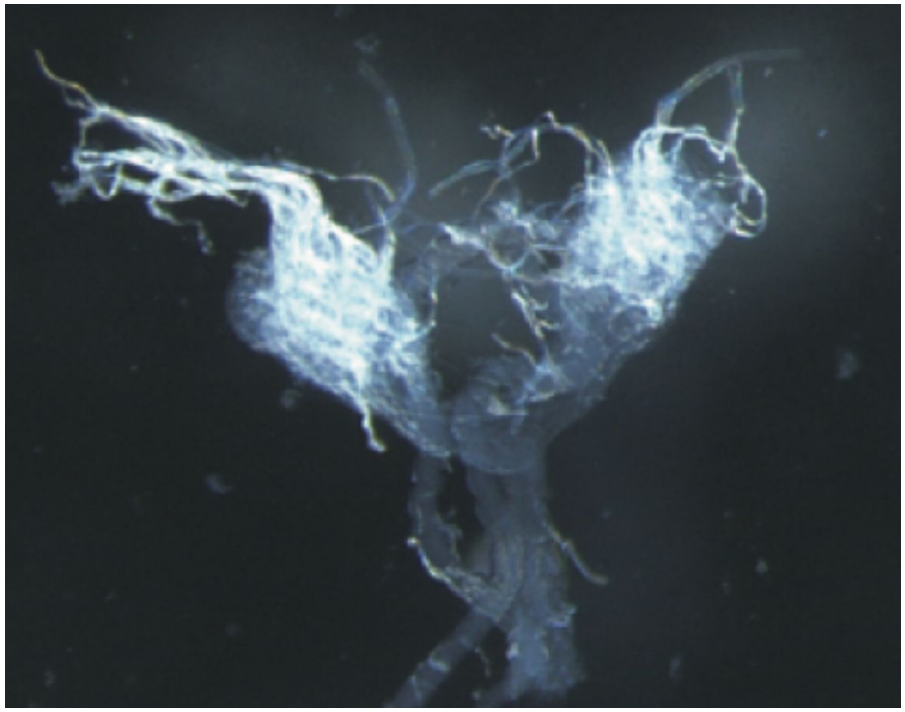


Engels 1990 Cell

How was the P-element discovered?

By doing this to a fly ovary

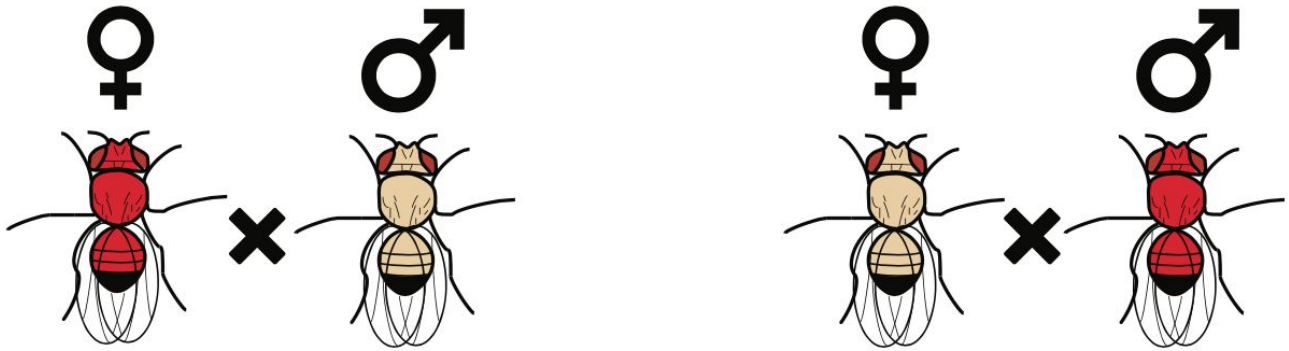




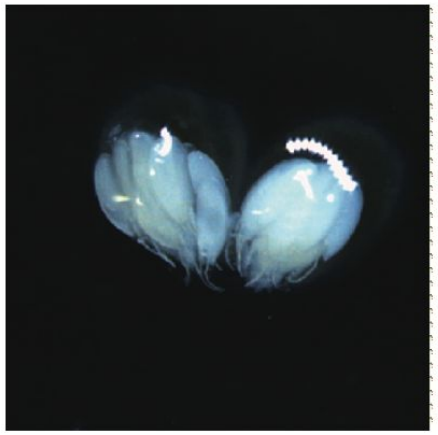
Hill, Schlötterer and Betancourt (2006) PLoS Genetics

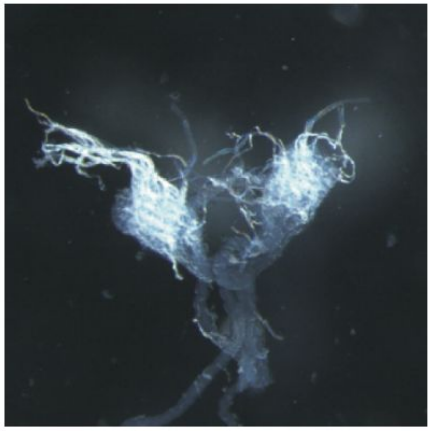
Hybrid dysgenesis

P



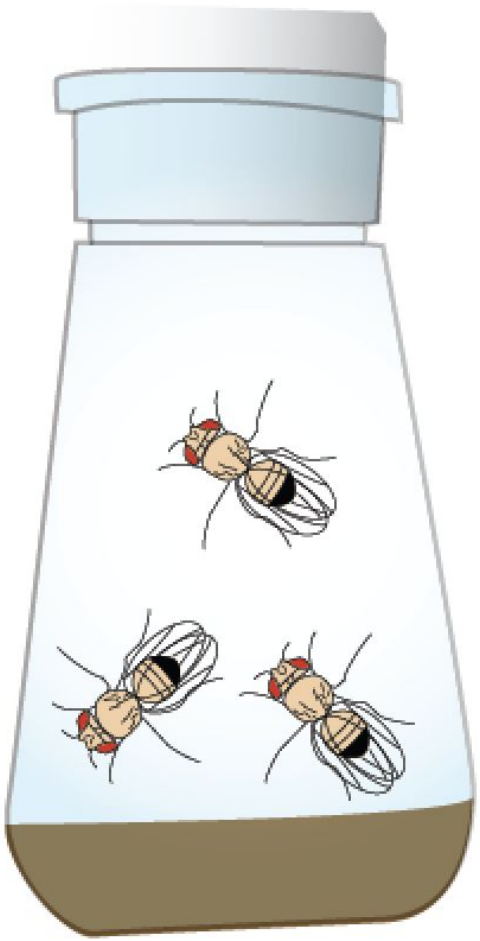
F1

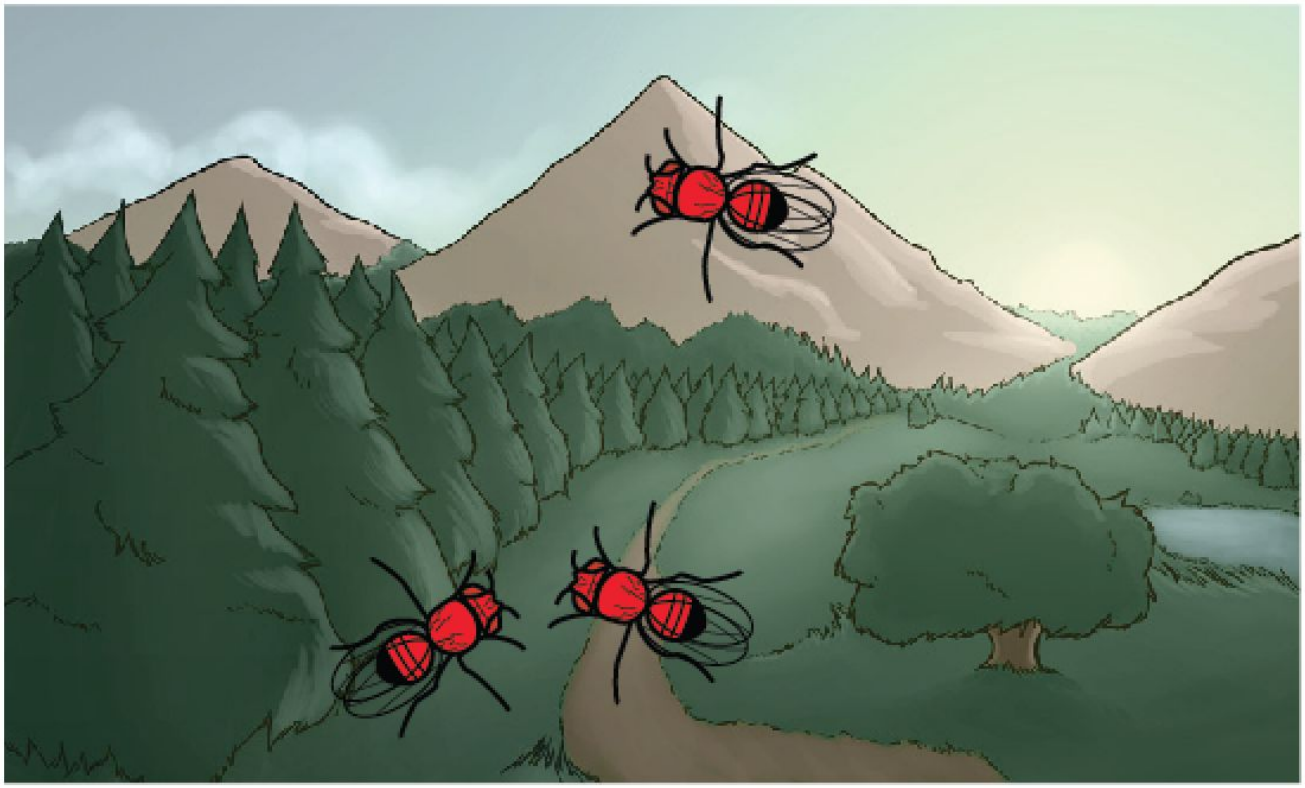




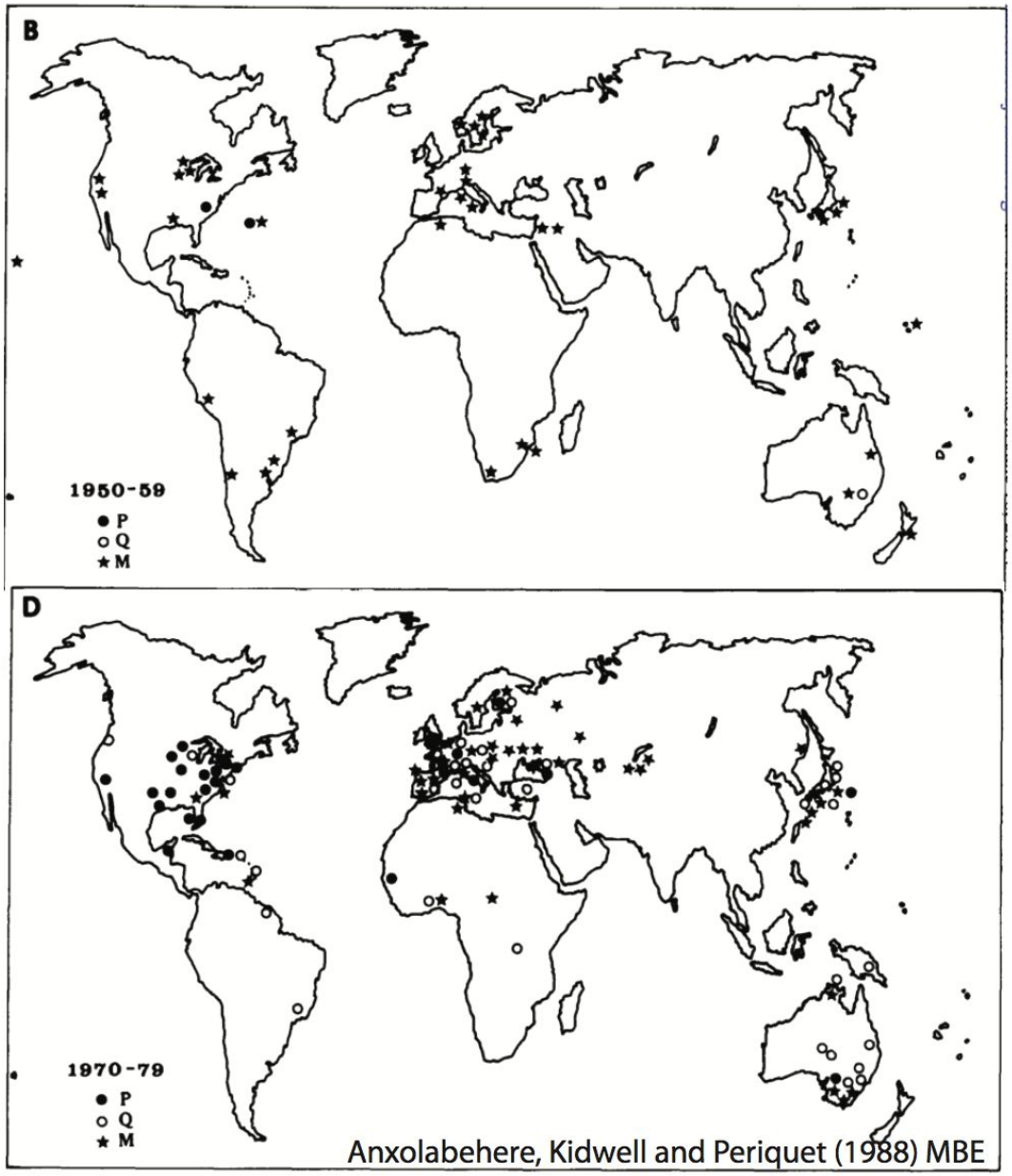
Hill, Schlötterer and Betancourt (2016) PLoS Genetics
Kidwell, Kidwell and Sved (1977) Genetics

Why are natural strains infected but lab strains not?





Rapid invasion in natural *D. melanogaster* populations



Anxolabehere, Kidwell and Periquet (1988) MBE

Never found in related Drosophila species

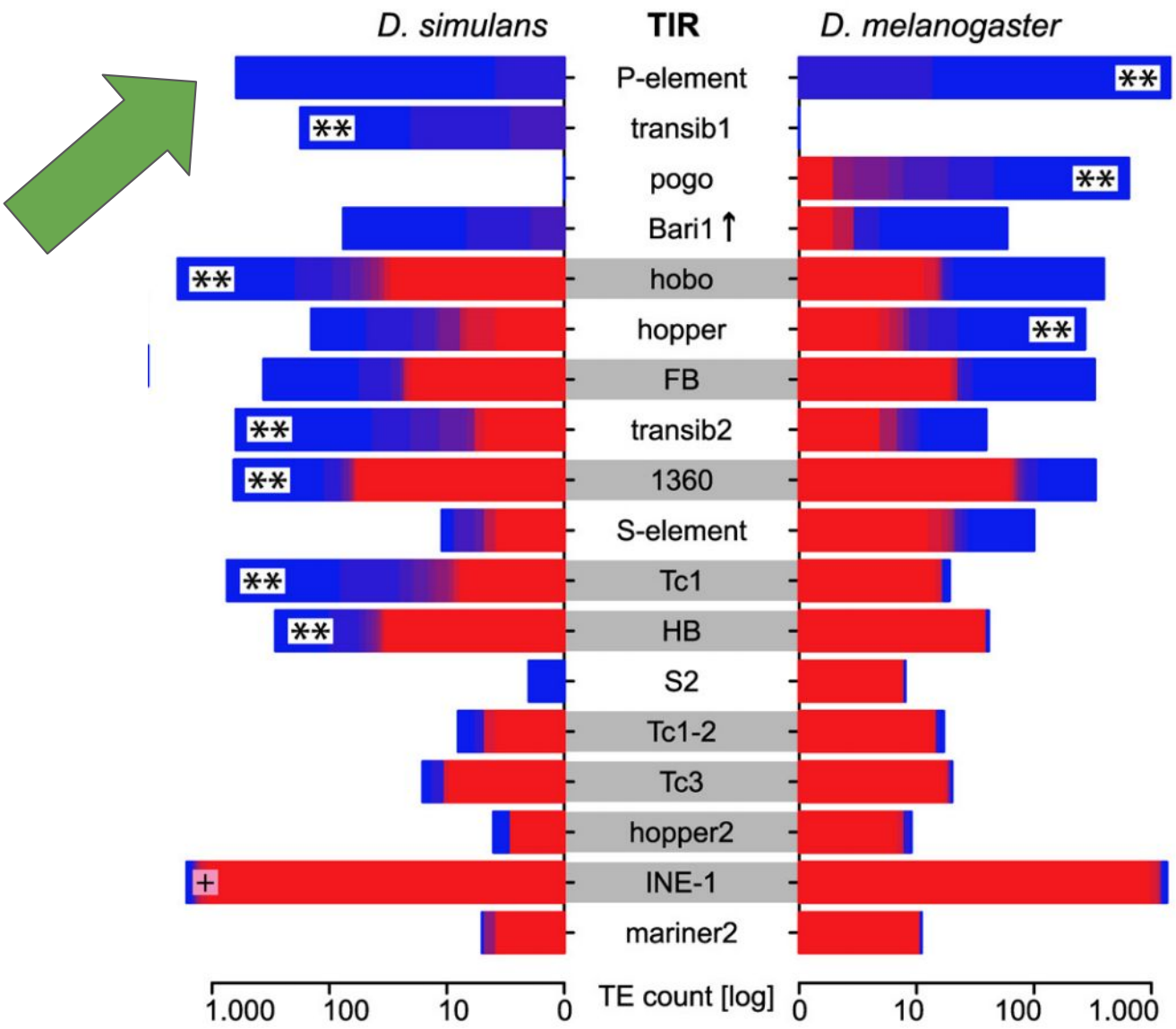
Table 1 The distributi

Species	pπ25.1
D. melanogaster	+
D. simulans	–
D. mauritiana	–
D. yakuba	–
D. teissieri	–
D. pseudoobscura	–

Brookfield 1986 Nature

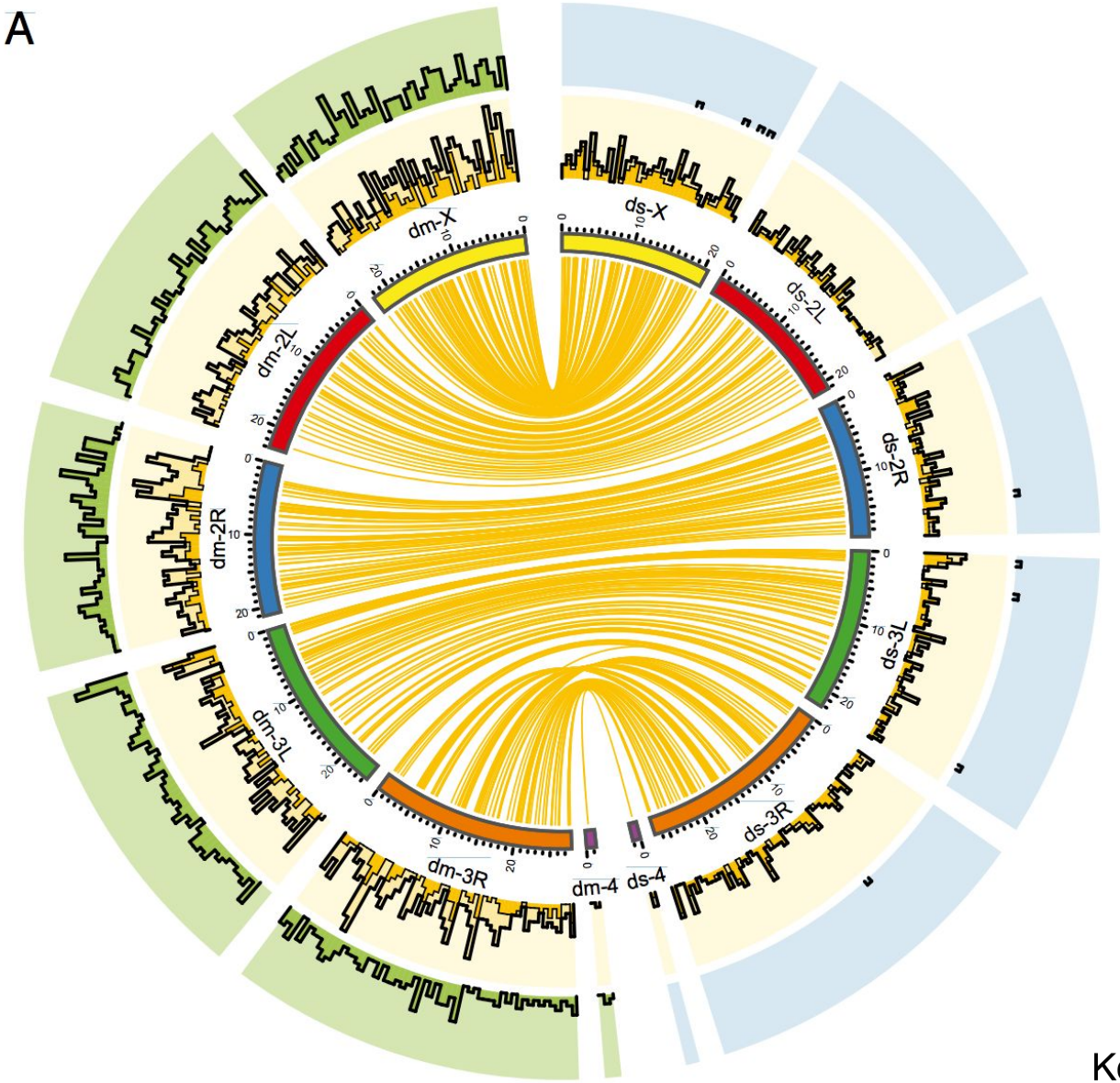
until..
we sampled *D. simulans* from South Africa in 2012

Comparing the TE abundance between *Dmel* and *Dsim*: a signal where none should be found



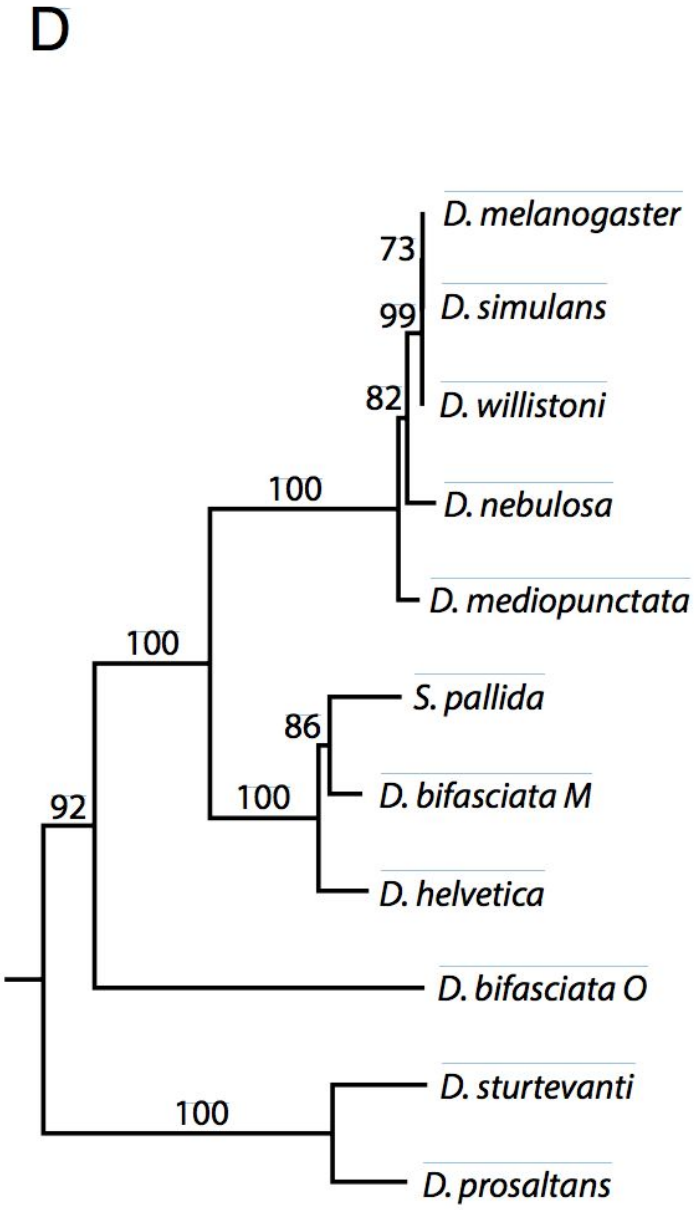
Kofler et al. 2015 PLoS Genetics

Genomic distribution in two populations of *Dmel* and *Dsim*



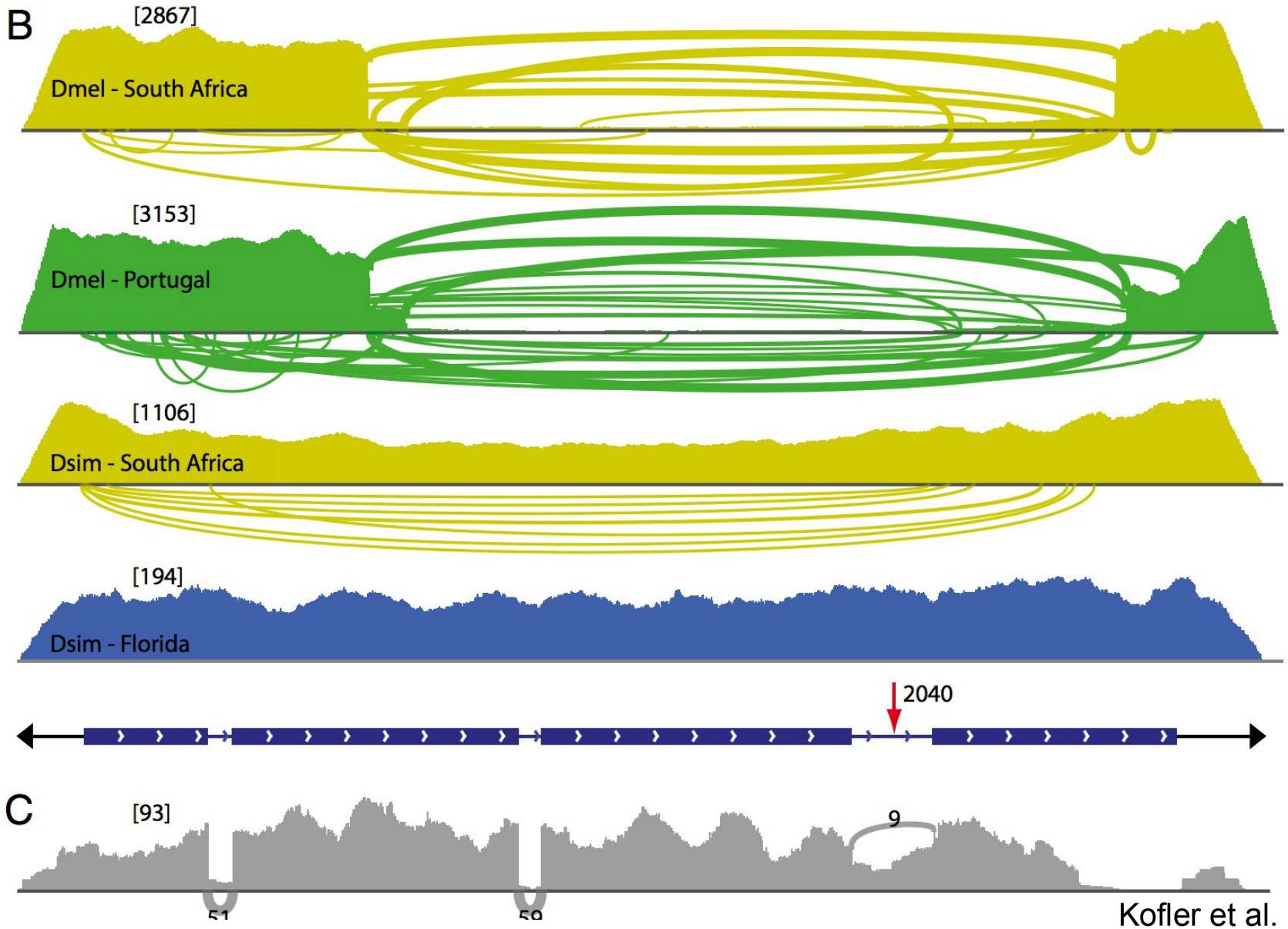
Kofler et al. 2015 PNAS

Origin: probably HT from *D.melanogaster*



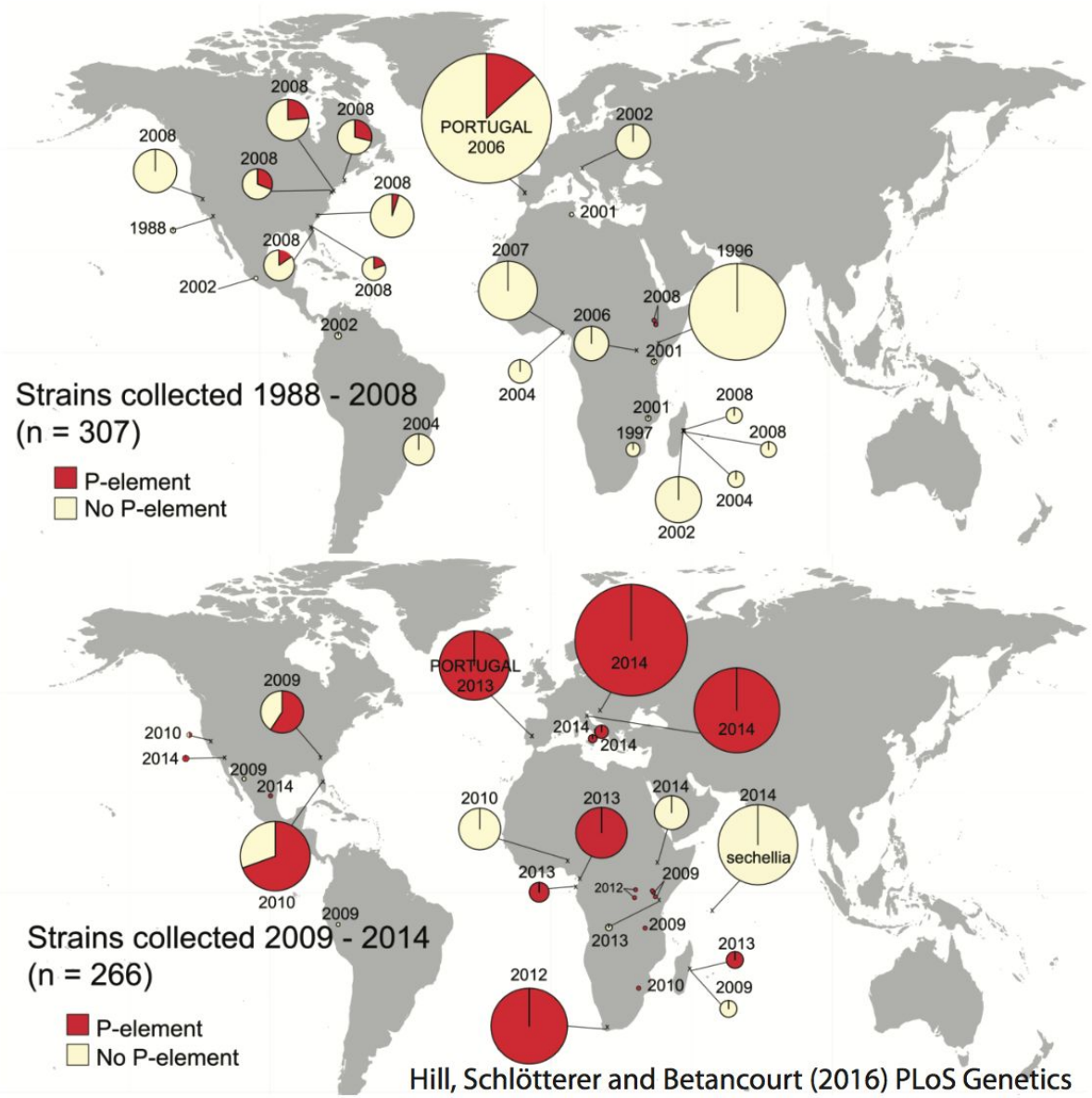
Kofler et al. 2015 PNAS

Dsim P-element is mostly full length and has a single SNP



Kofler et al. 2015 PNAS

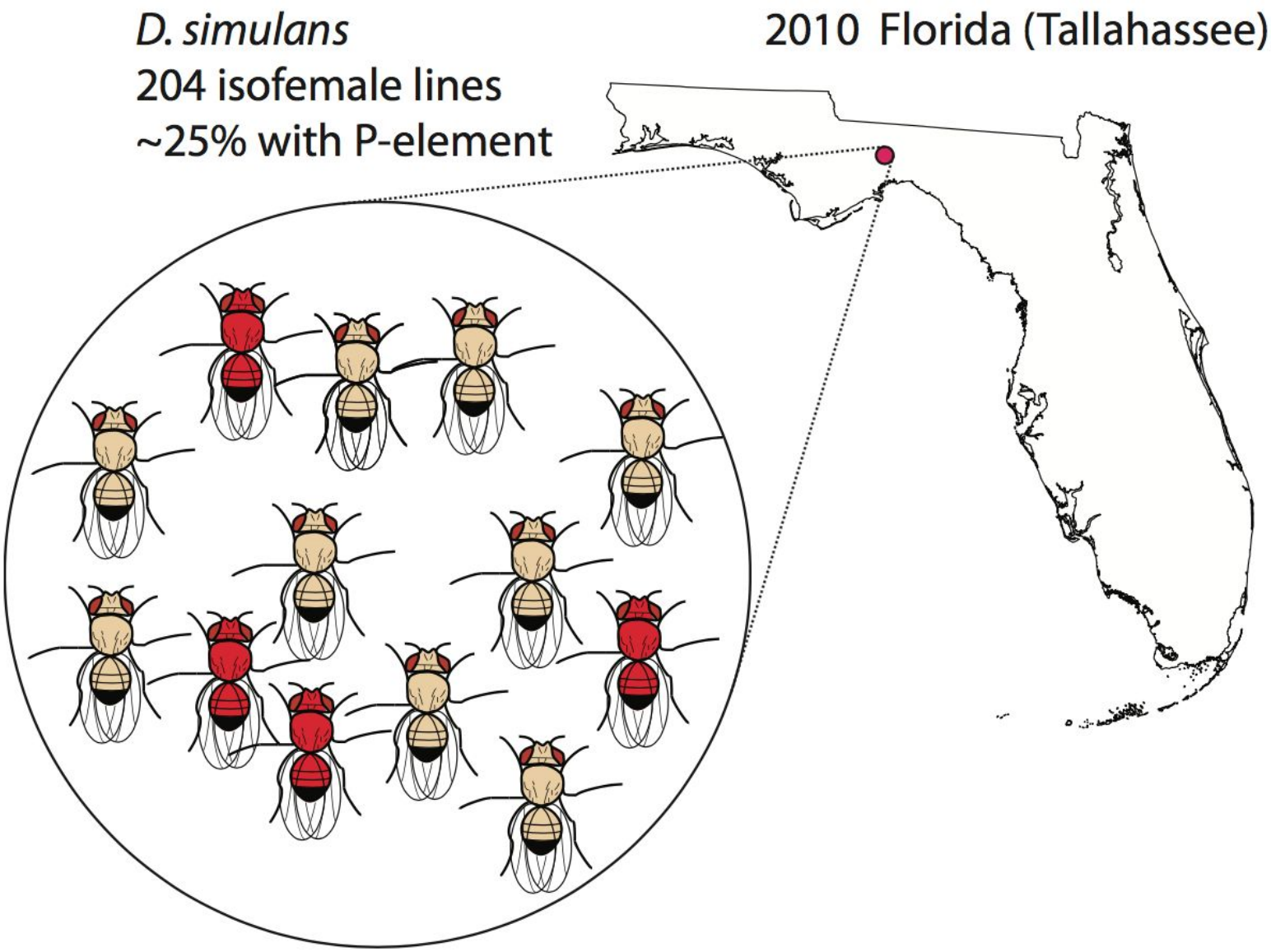
Rapid spread in worldwide populations



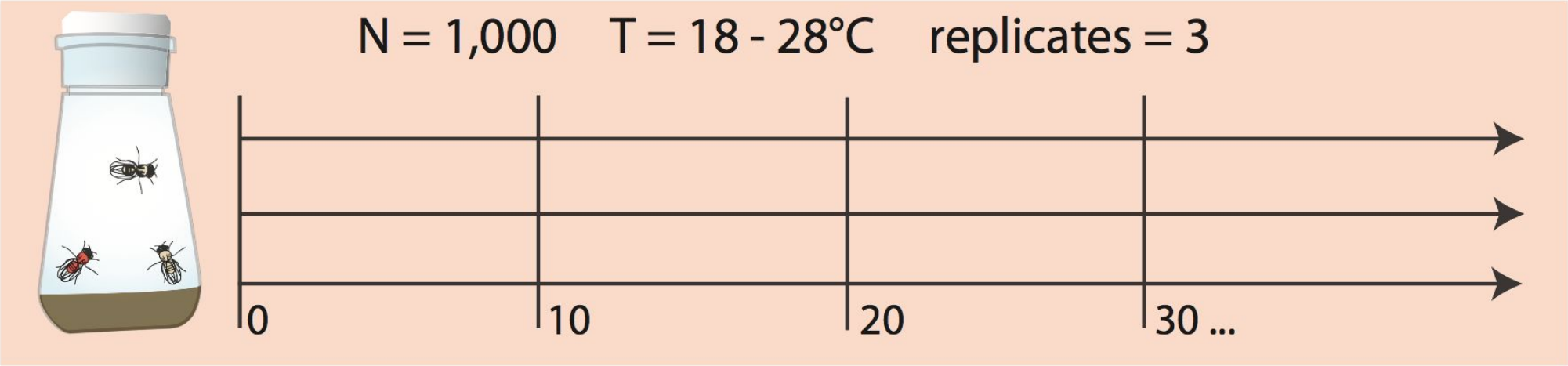
Hill, Schlötterer and Betancourt (2016) PLoS Genetics

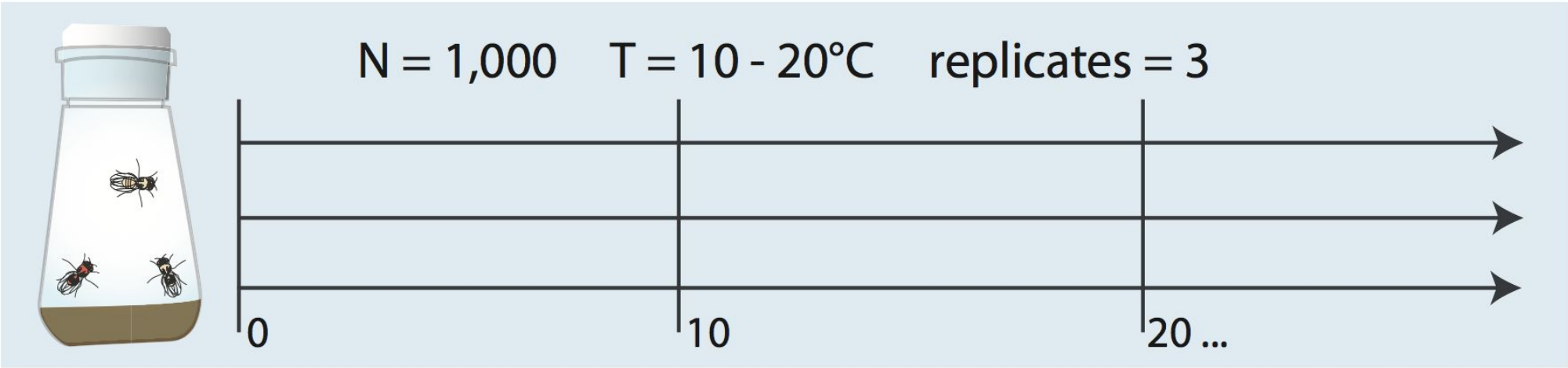
An unique opportunity: monitoring a natural TE invasion

A natural population at an early stage of a P-element invasion



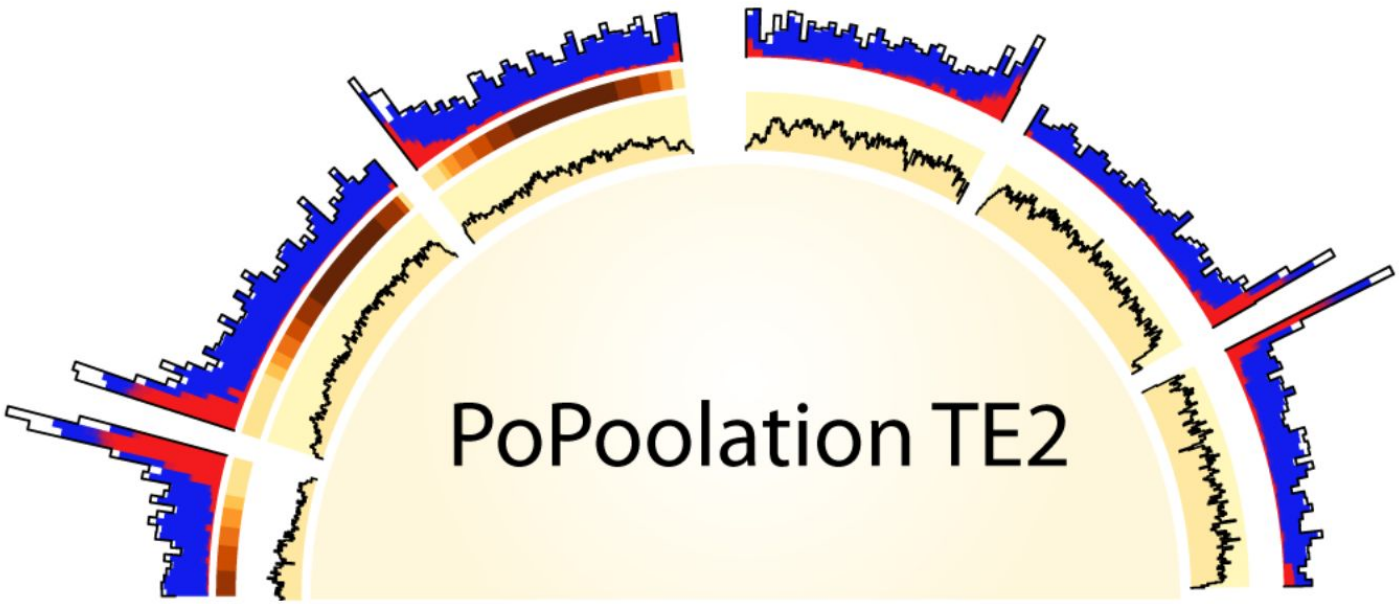
Experimental populations





Kofler et. al; submitted

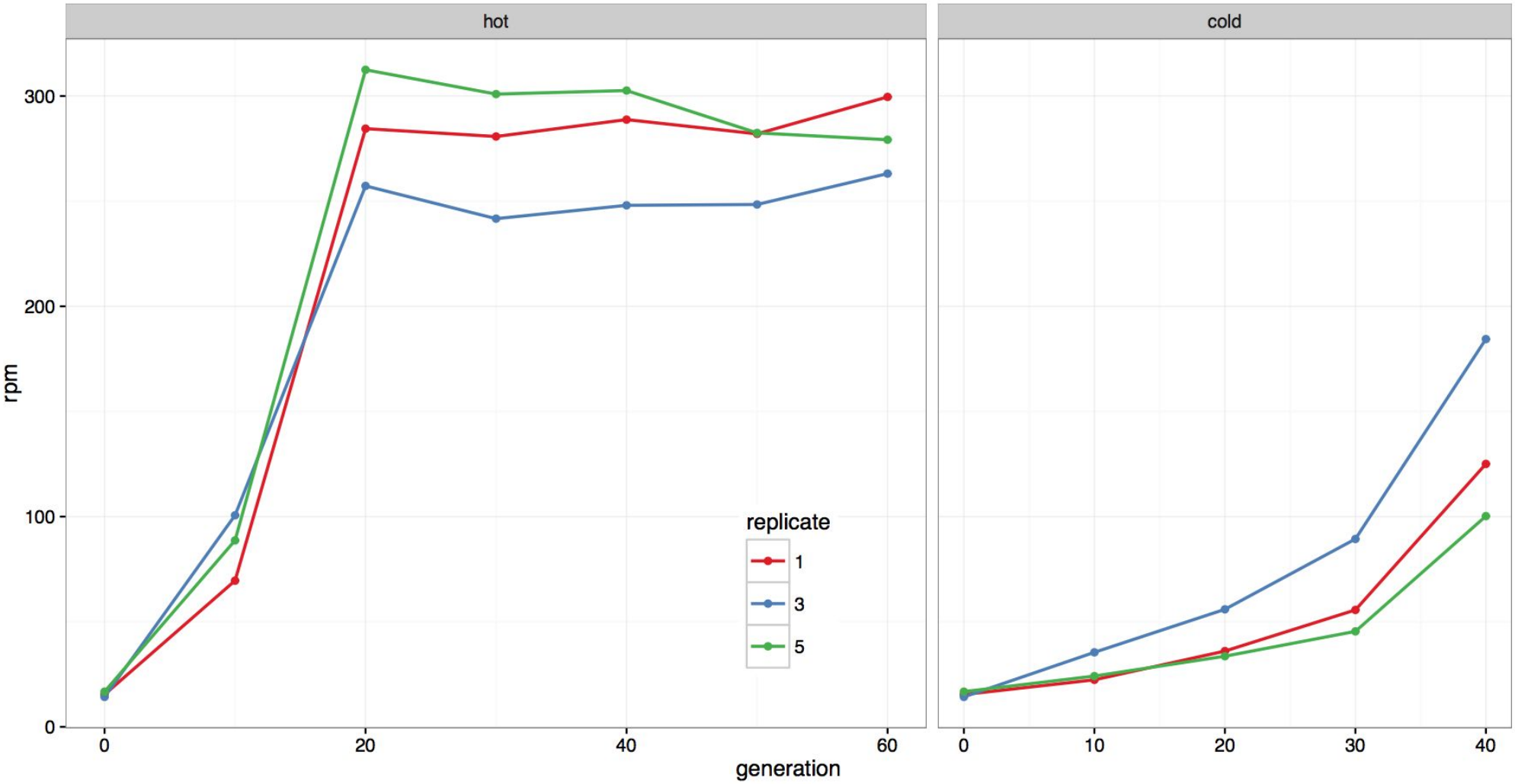
Estimate TE abundance in pooled samples



- enables a fast comparison of TE abundance between multiple pooled samples (implemented in Java and accepts bam-file)
- avoids biases caused by differences in insert size and read numbers

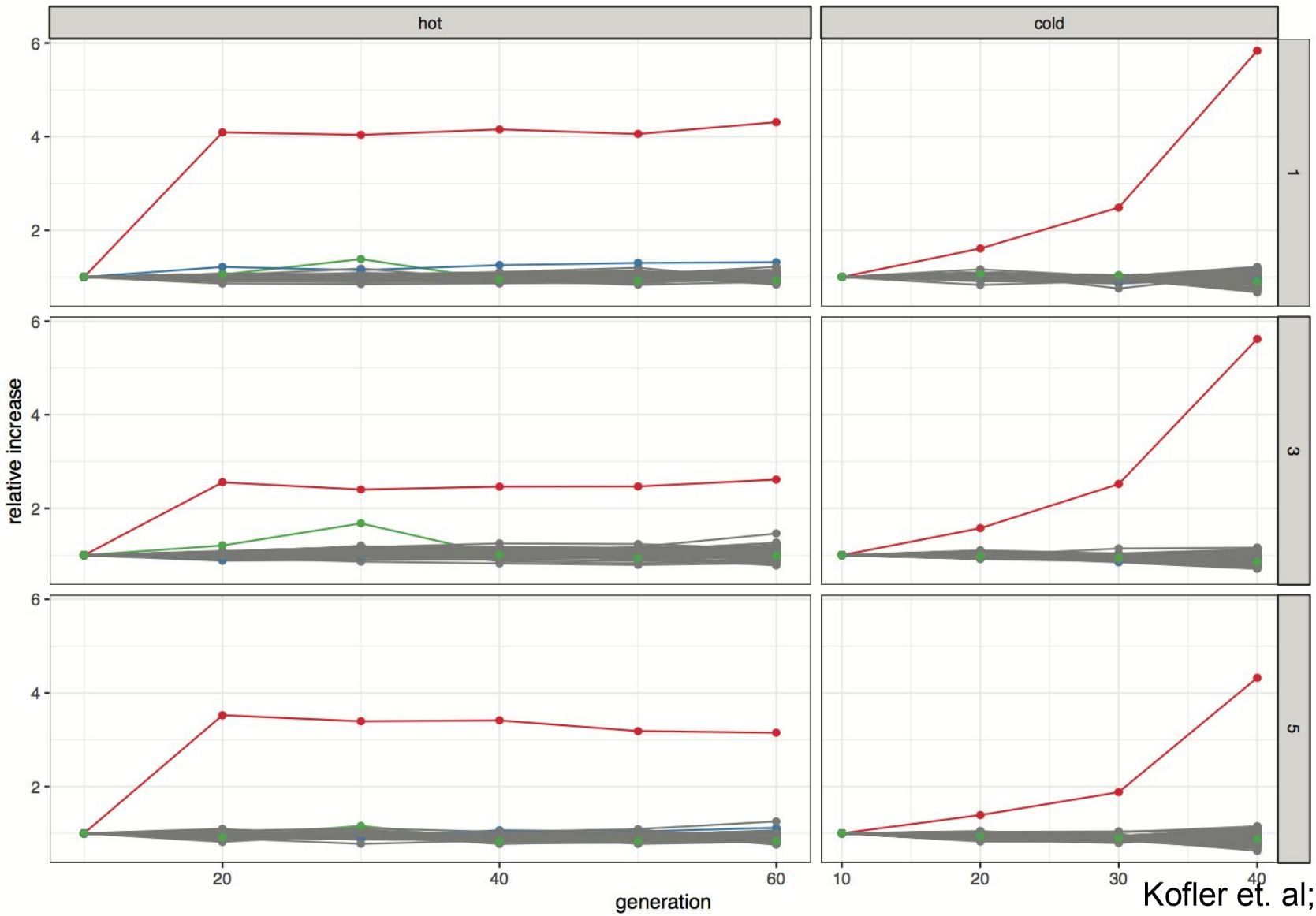
Kofler, Gómez-Sánchez and Schlötterer (2016) MBE

Fast invasion in hot - slow at cold; Plateauing of invasion in hot



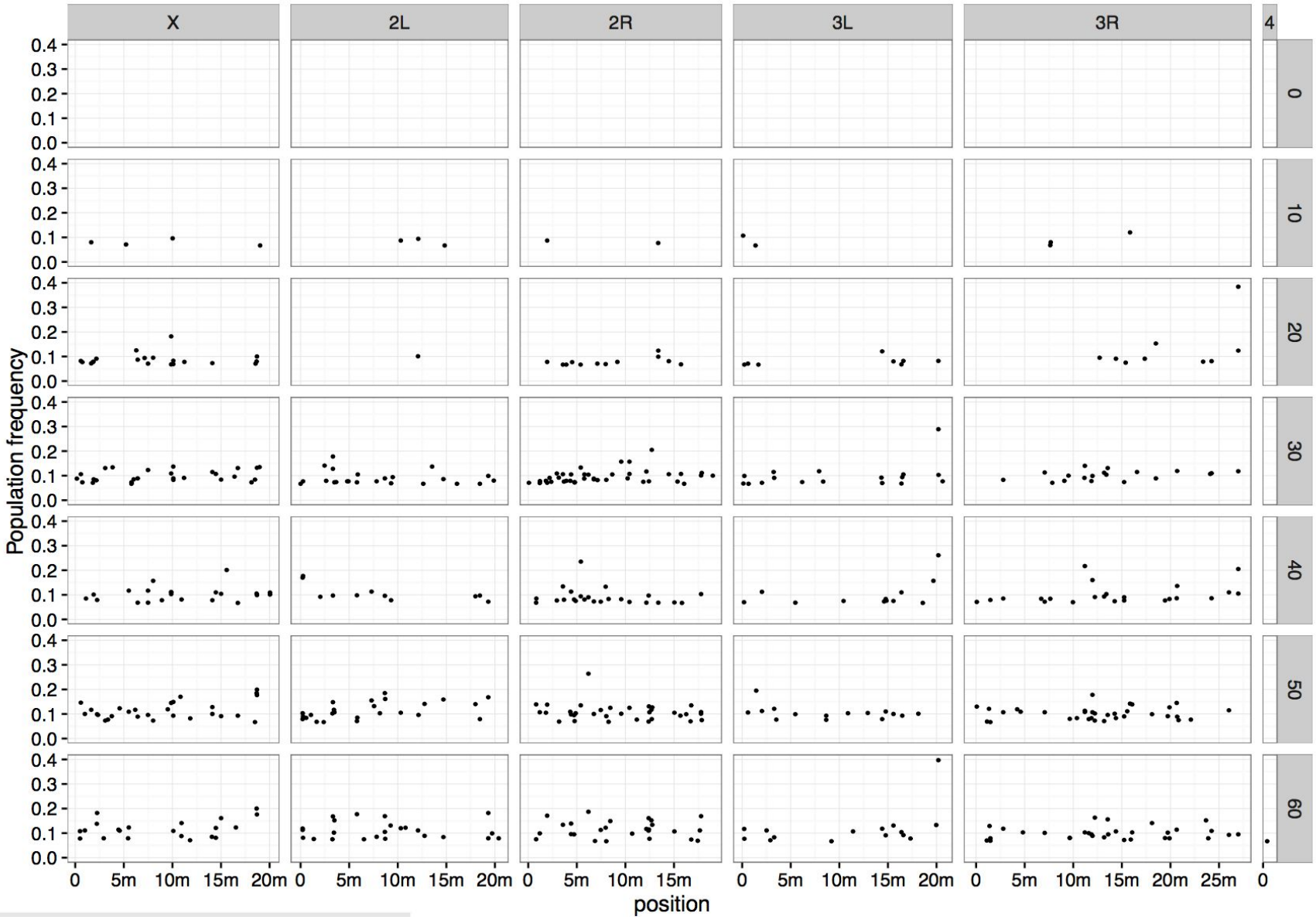
Kofler et. al; submitted

Only the P-element rises in frequency; all other families remain at a constant level

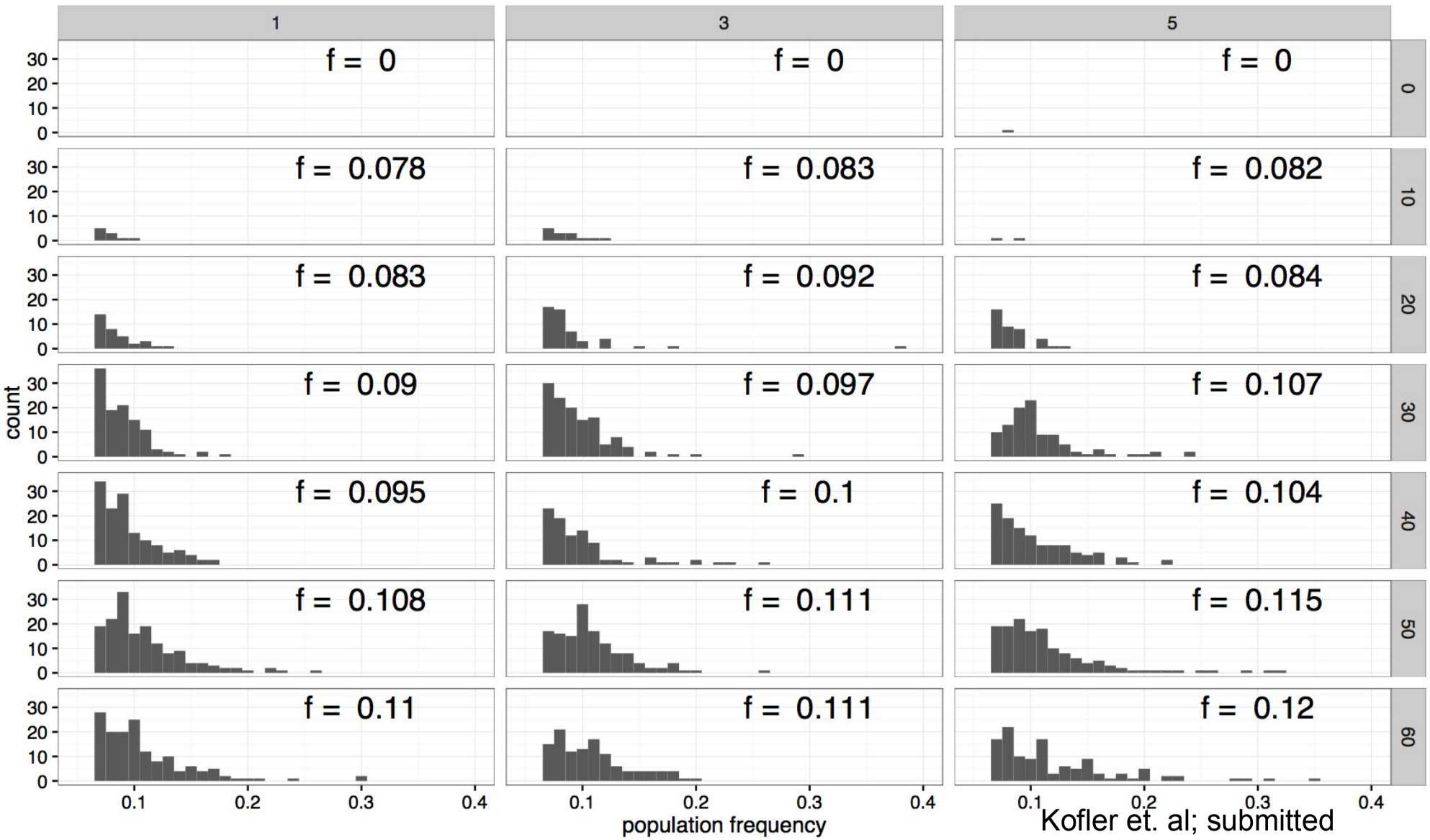


Kofler et. al; submitted

Low frequency insertions emerge

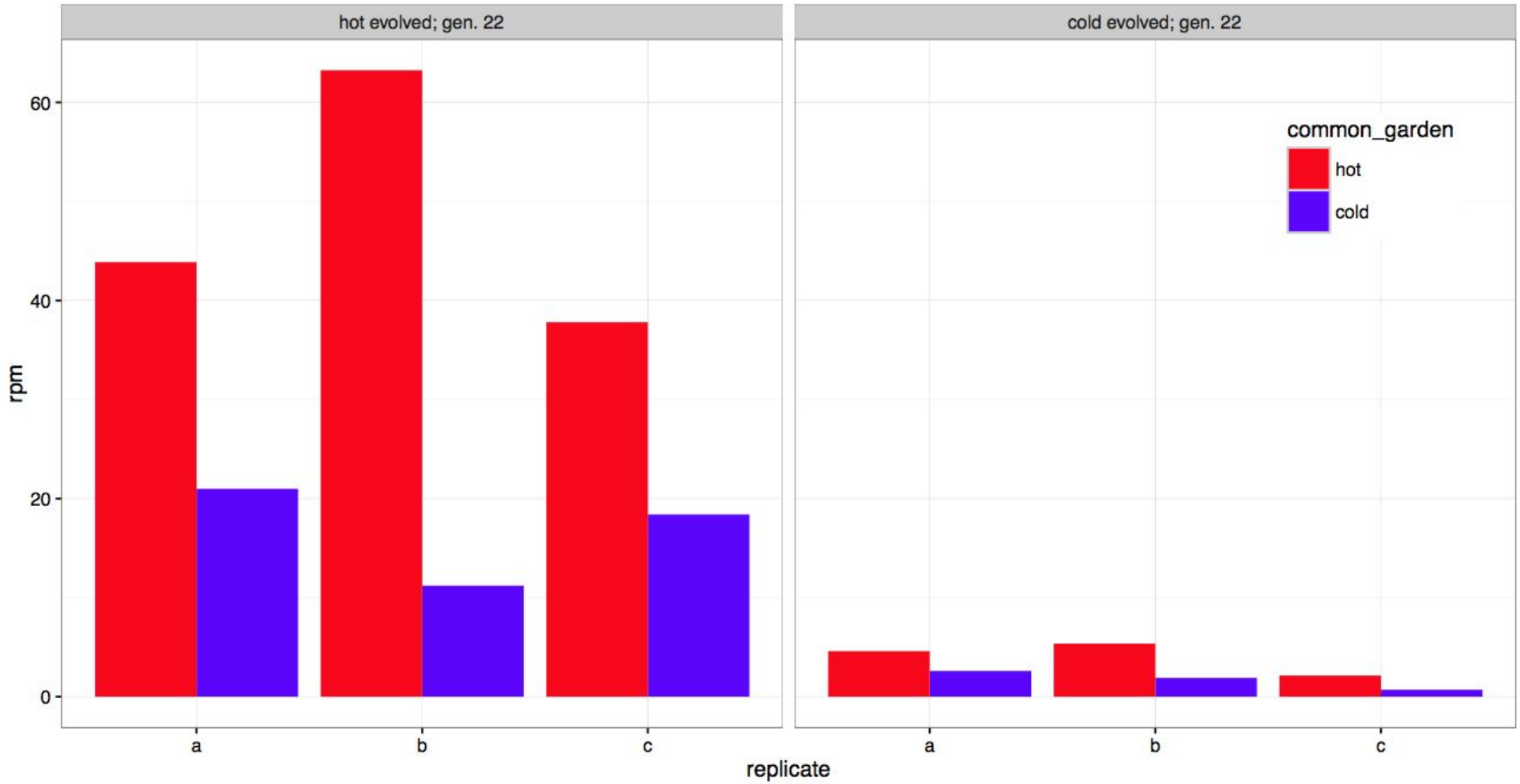


Average population frequency increases



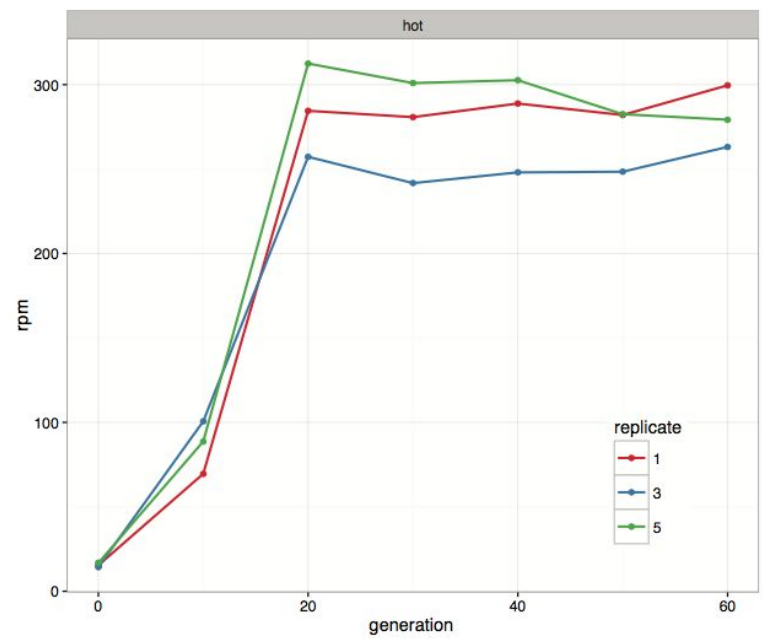
Kofler et. al; submitted

P-element expression is lower at cold temp.



Dynamics of hot invasion

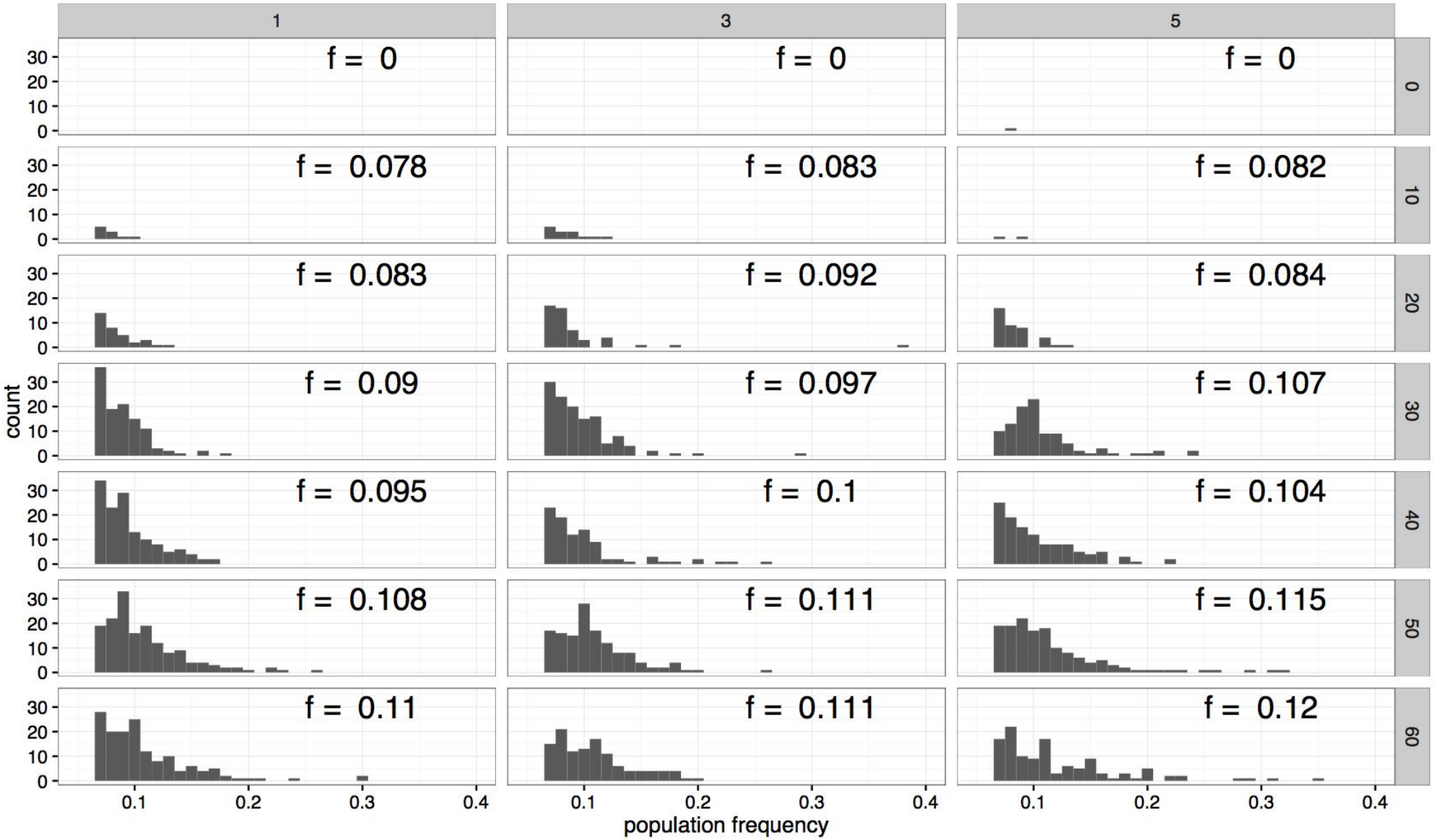
1. high activity until generation 20
2. activity declines at the plateau
3. some insertions rise in frequency due to drift



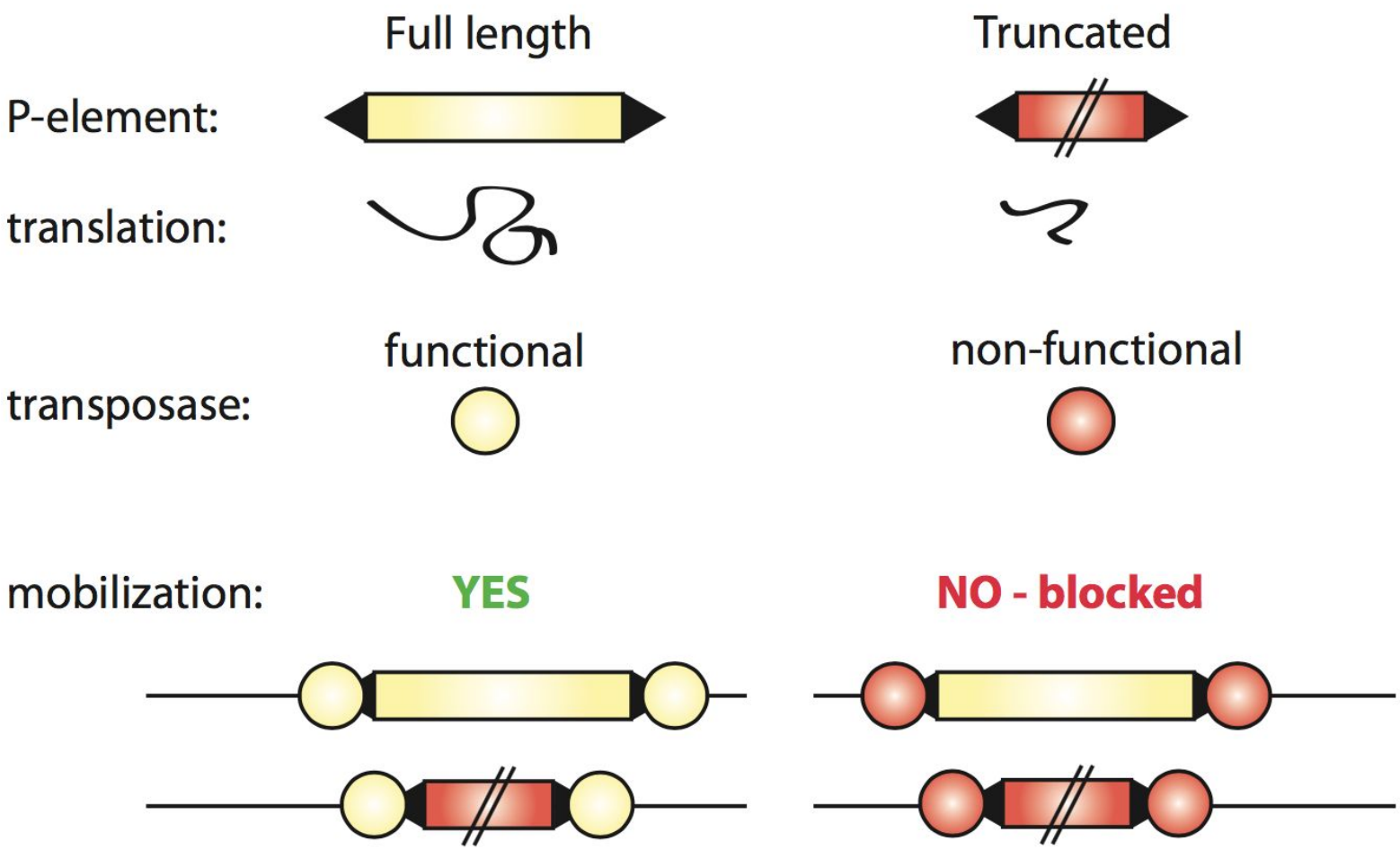
Possible causes of the decreasing activity

1. all possible insertion sites occupied (P-element has a strong insertion bias)
2. truncated copies may downregulate activity
3. piRNAs may downregulate activity

1) No fixation of insertions suggests that the majority of possible insertion sites is not occupied

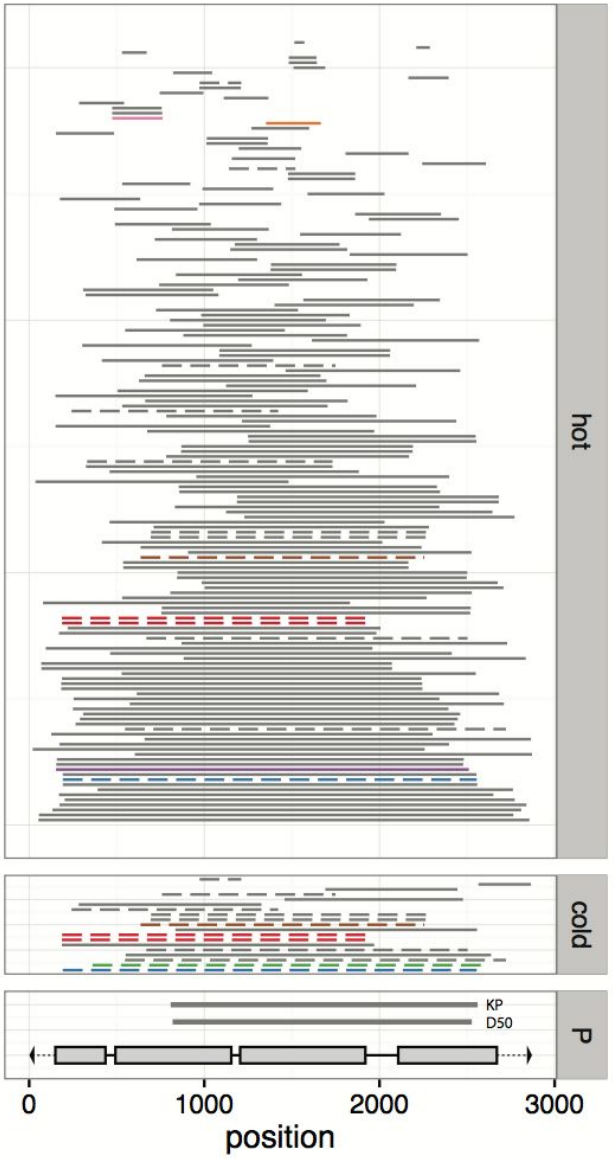


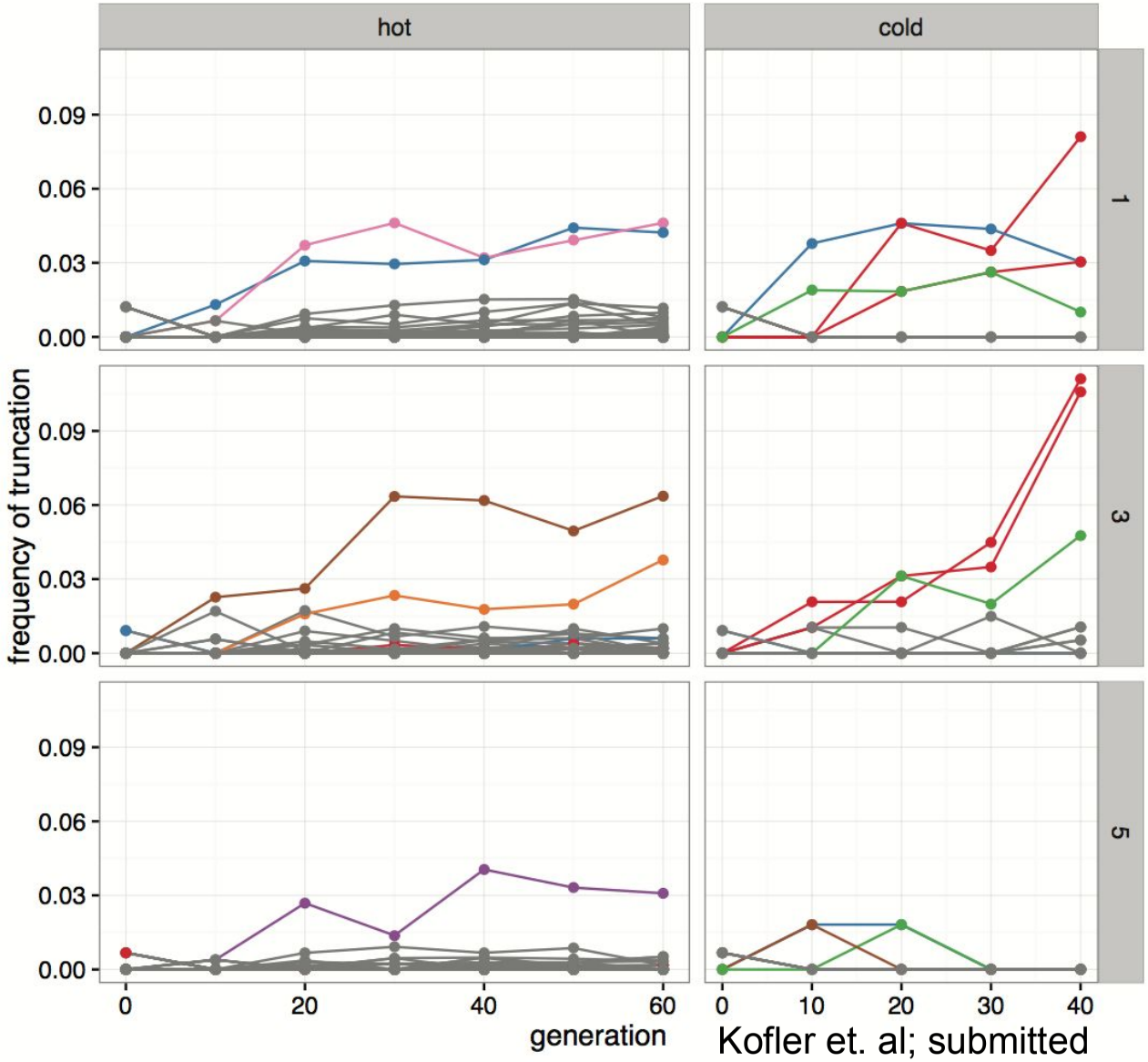
2) Truncated P-elements may downregulate activity by blocking binding sites



Black et al. (1987) EMBO Journal
Lee, Beall and Rio (1998) EMBO Journal

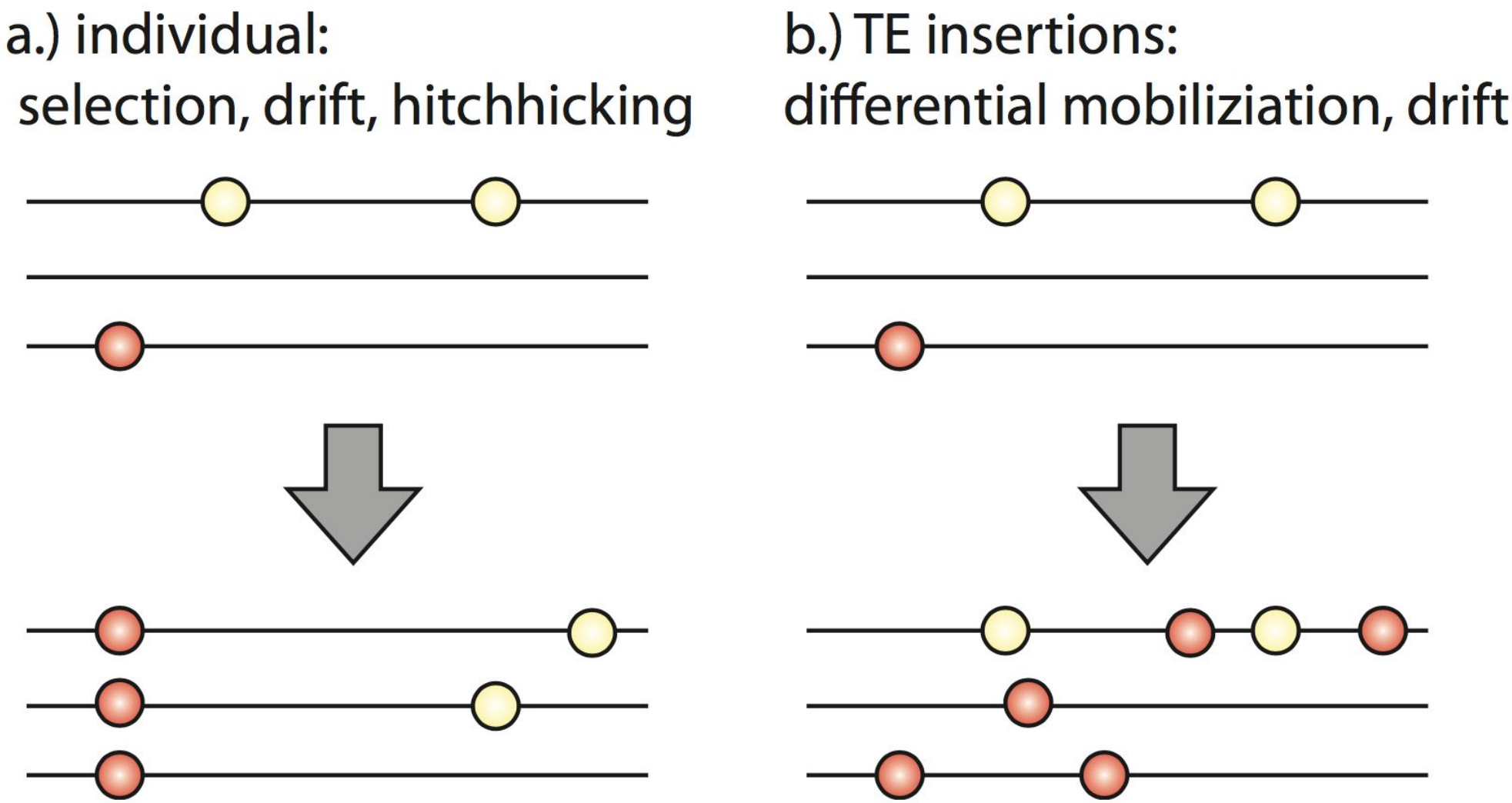
Truncated copies rapidly emerge - some rise in frequency





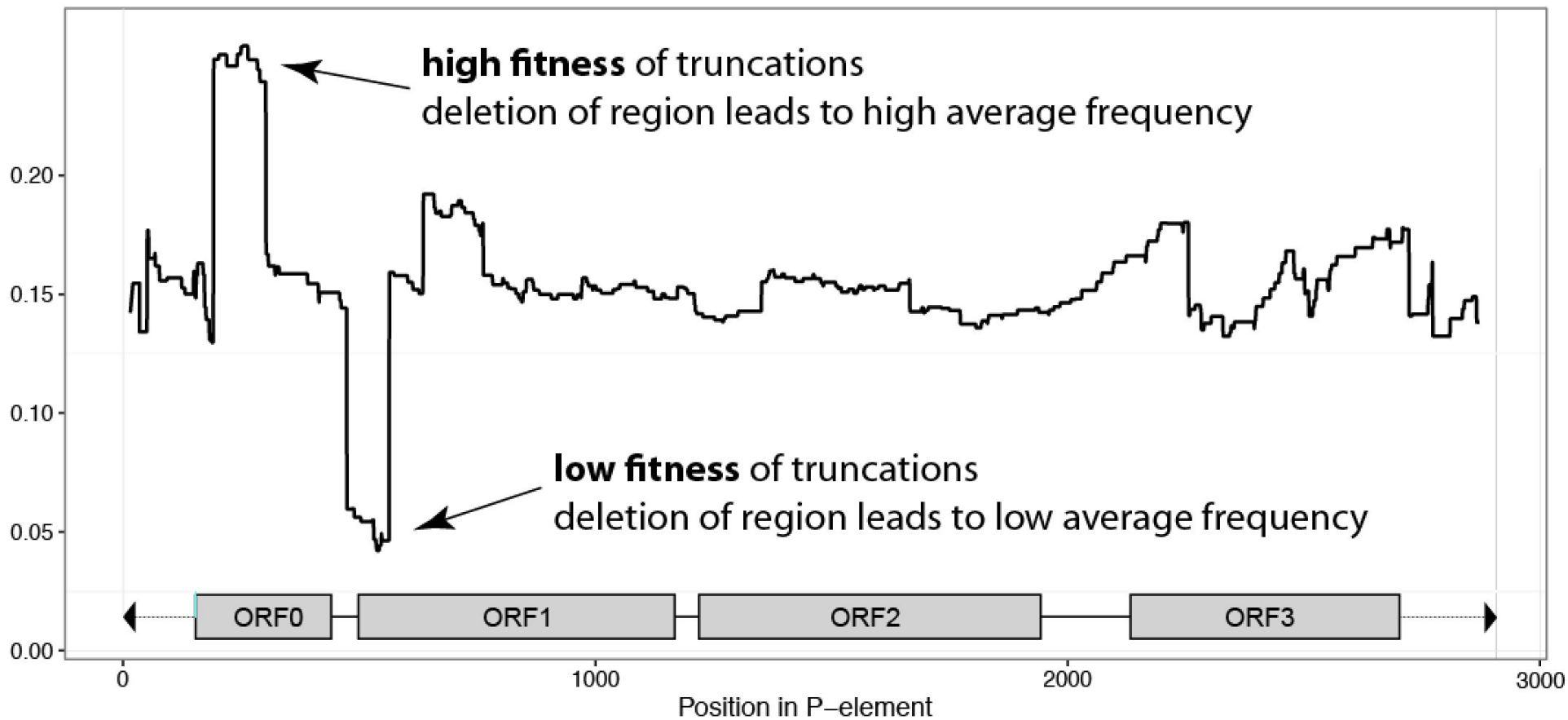
Kofler et. al; submitted

TEs may be selected at two levels



Insights from truncation trajectories

We used a "fitness landscape", in which the frequency of a truncation reflects its relative fitness. For every position in the P-element we averaged the frequency of all truncations covering the site.

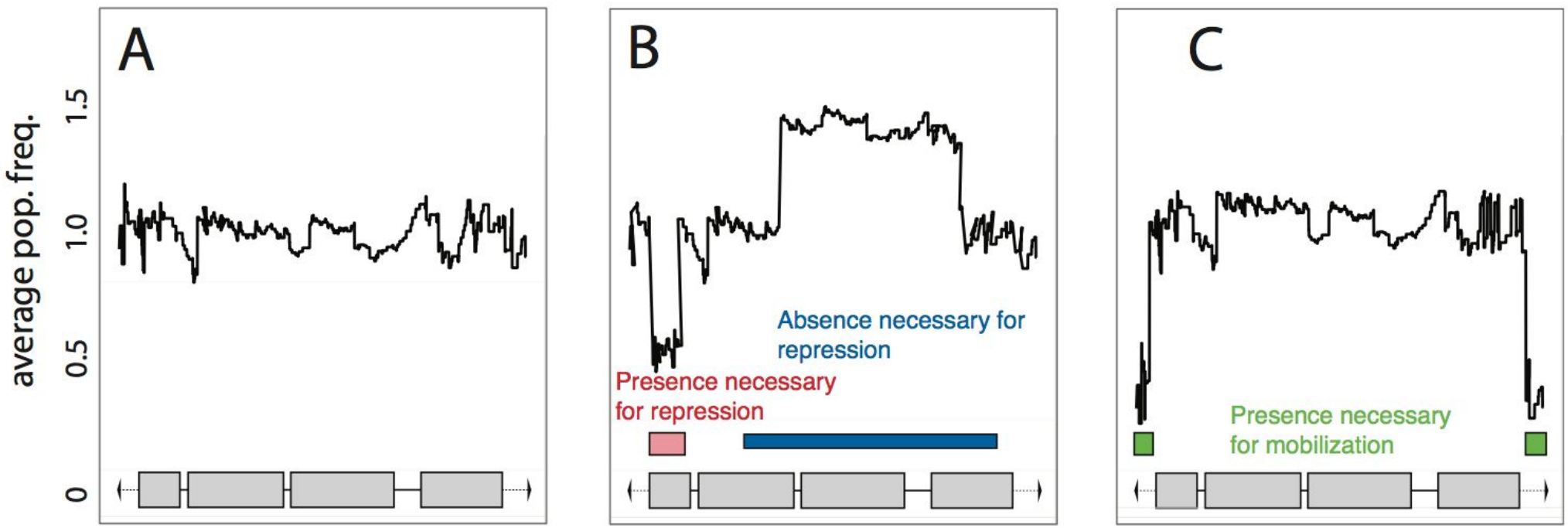


Fitness landscape of truncations allows to test hypothesis

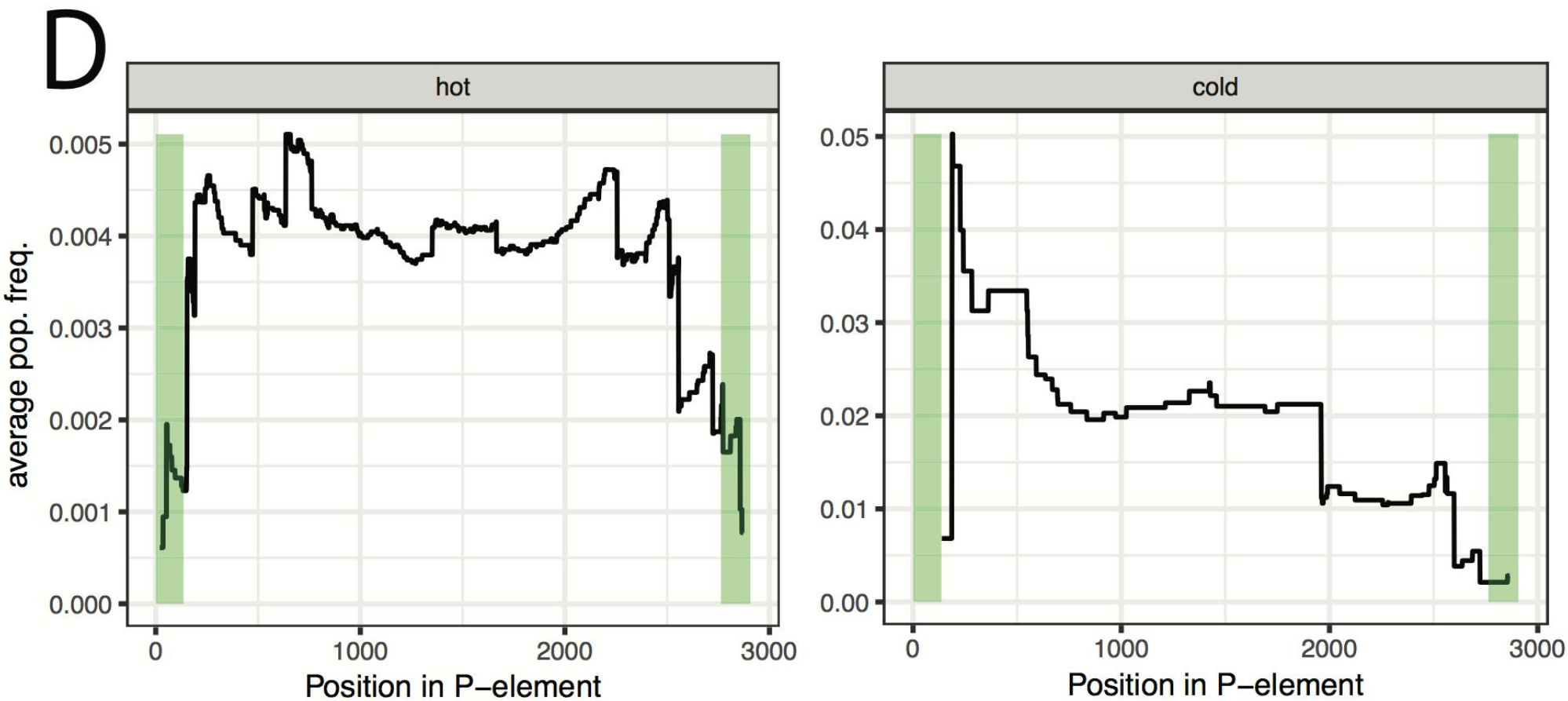
drift,
hitchhiking

pos. selection of truncations suppressing P-element activity

differential mobilization

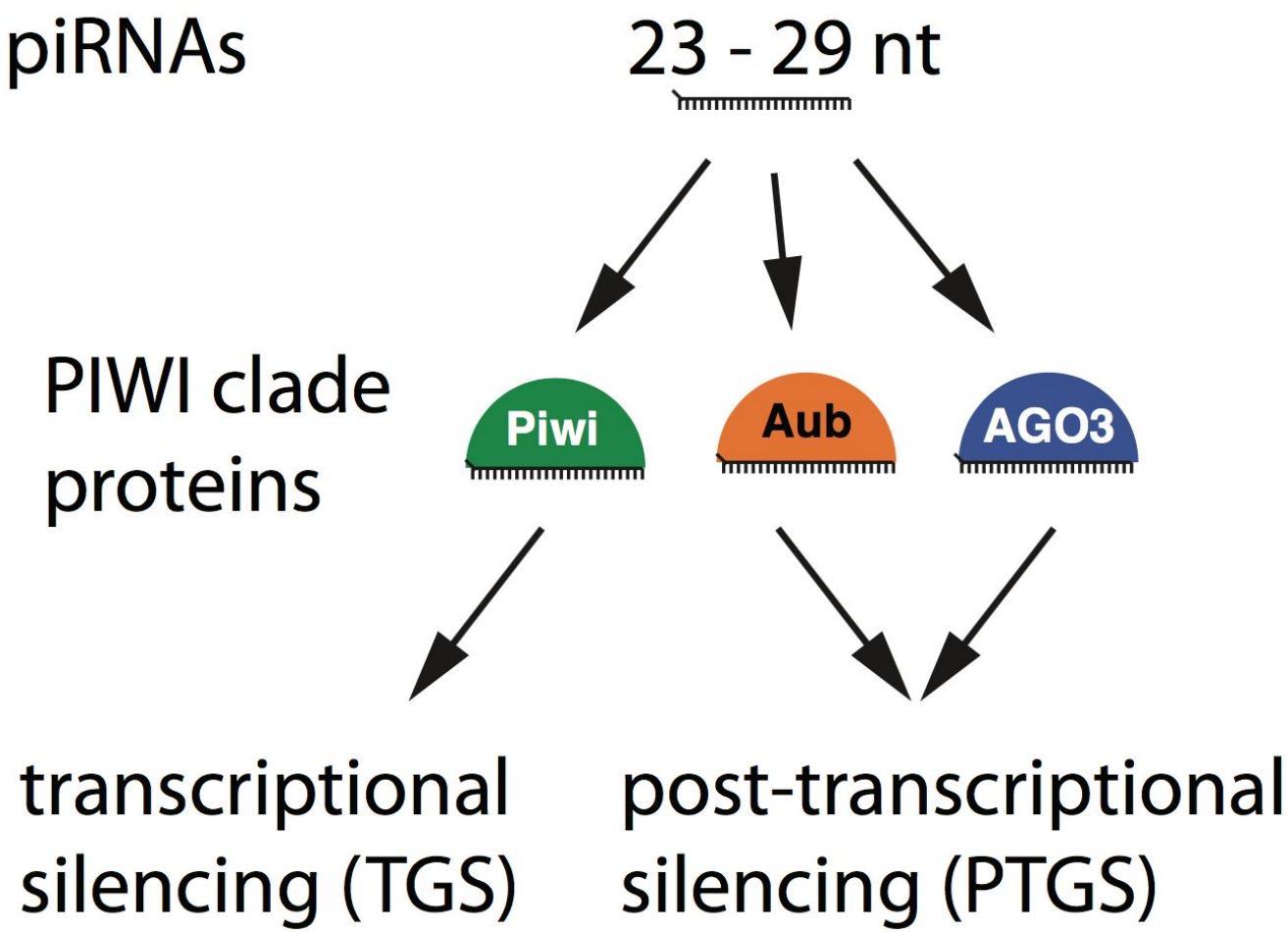


Differential mobilization governs trajectories of truncations



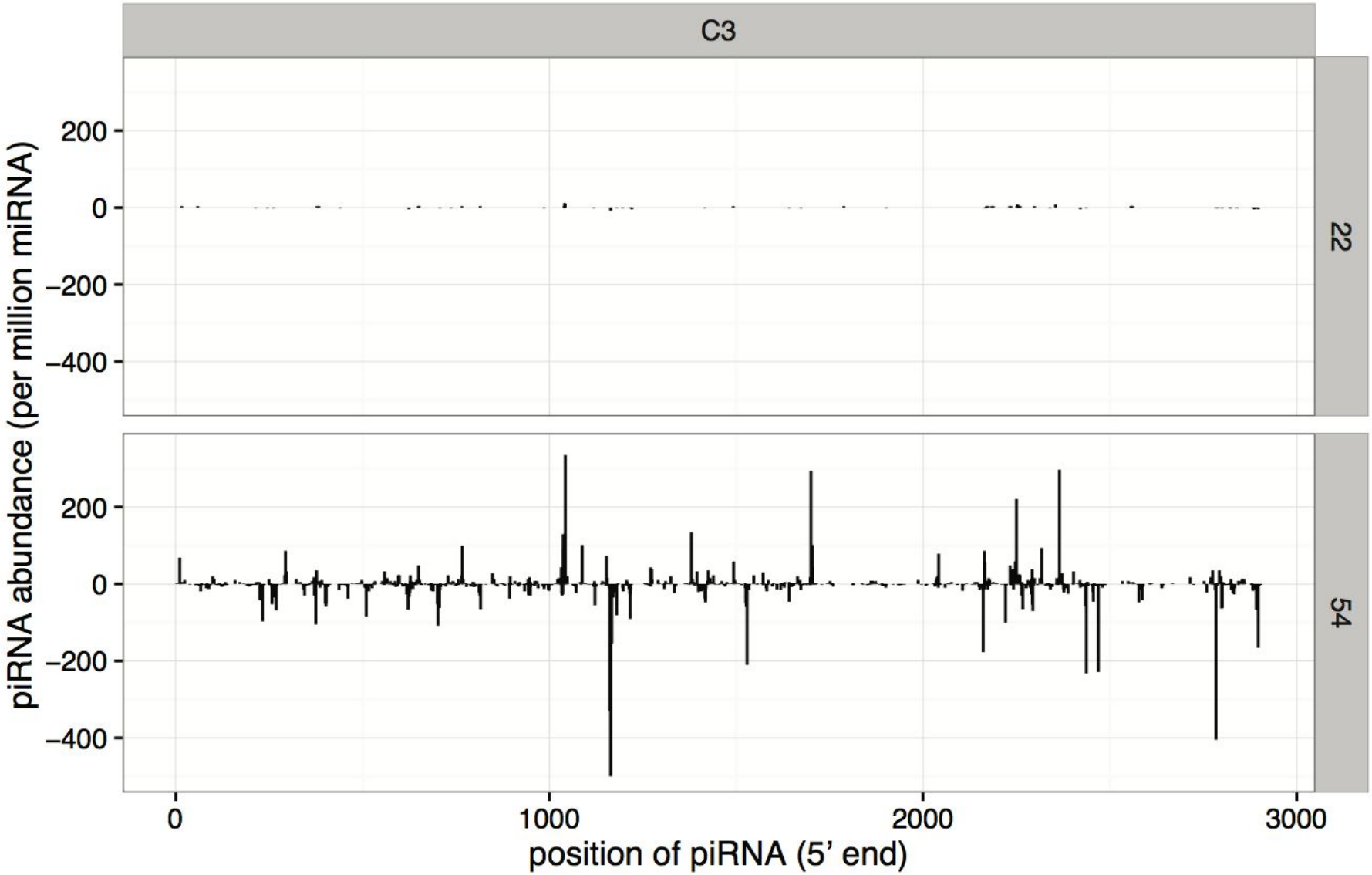
Kofler et. al; submitted

3.) piRNAs suppress TE activity



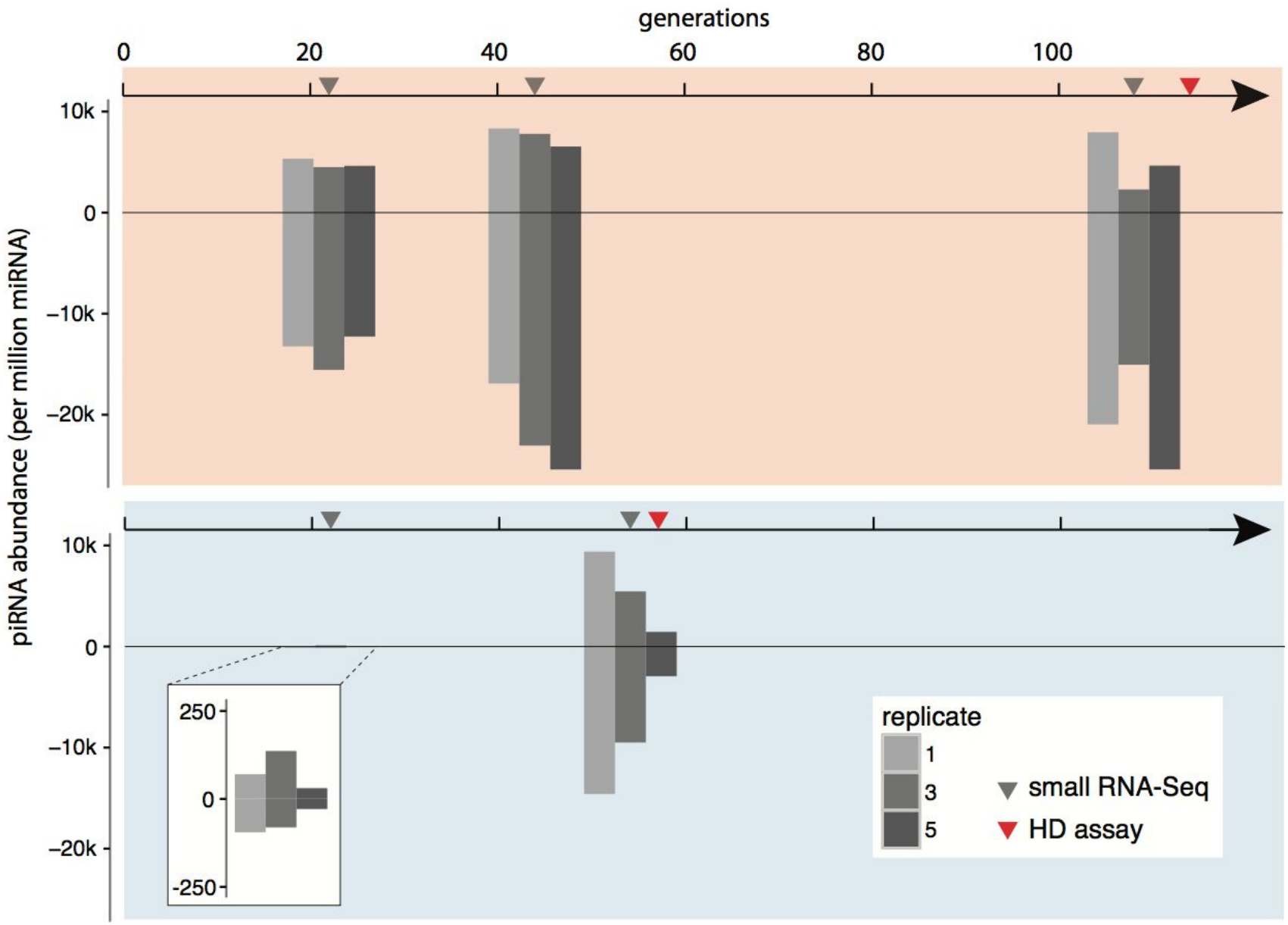
Brennecke et al. (2007) Cell
Gunarwadane et al. (2007) Science

piRNAs emerge in the course of the invasion

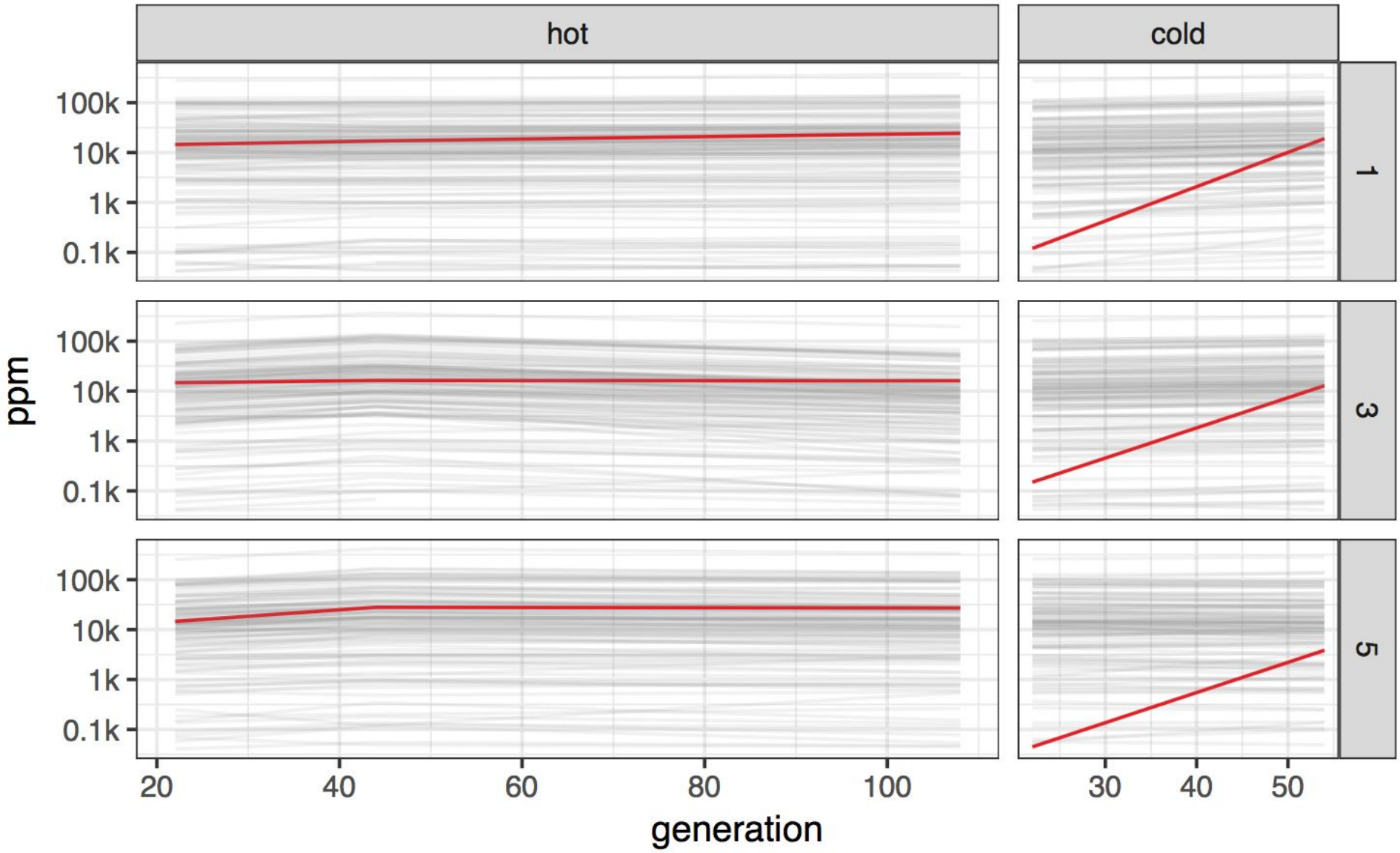


Kofler et. al; submitted

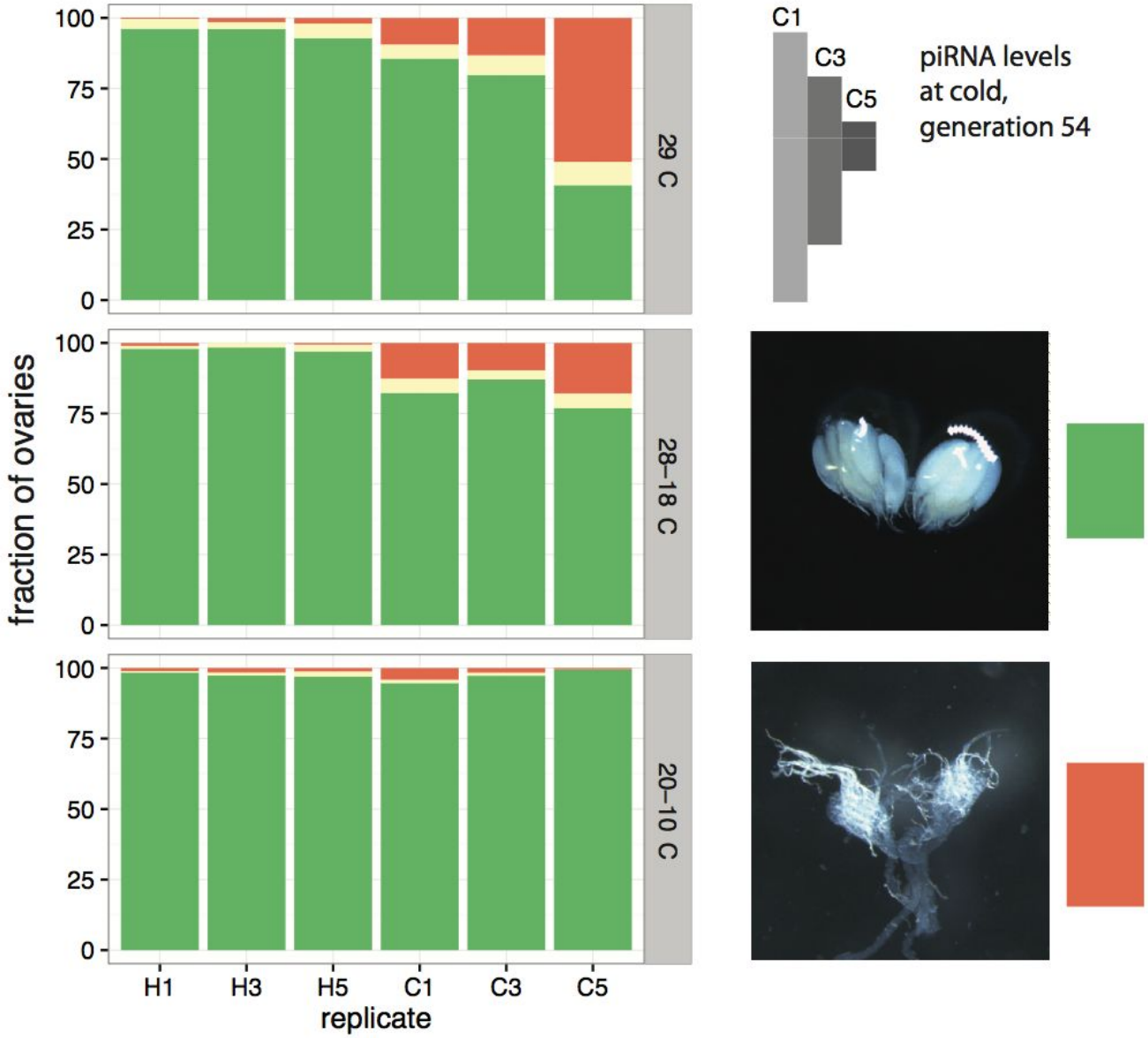
Fast emergence of piRNAs at hot - slow at cold



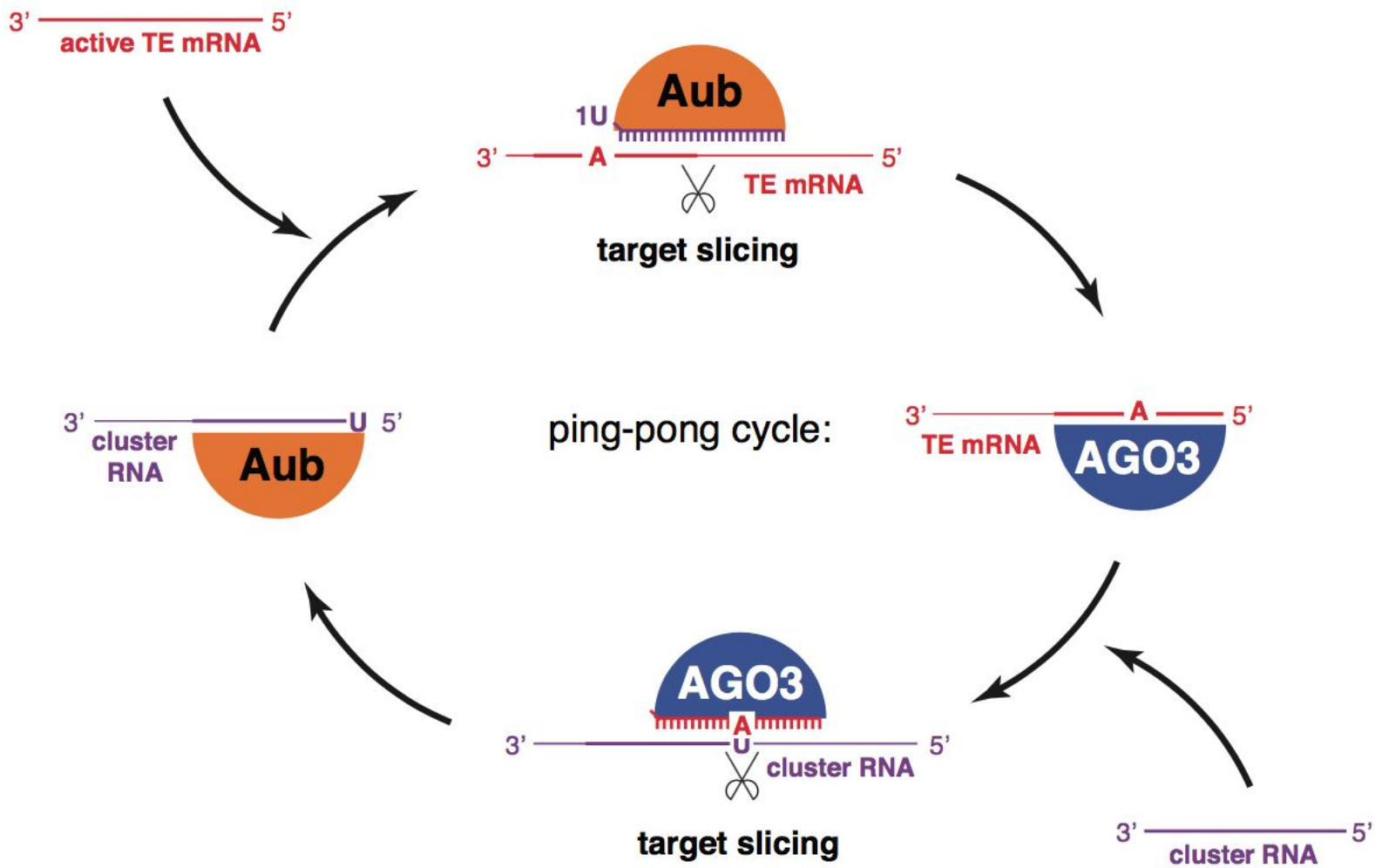
Only P-element piRNA levels increase at cold



Low piRNA levels at cold consistent with incomplete suppression of the P-element

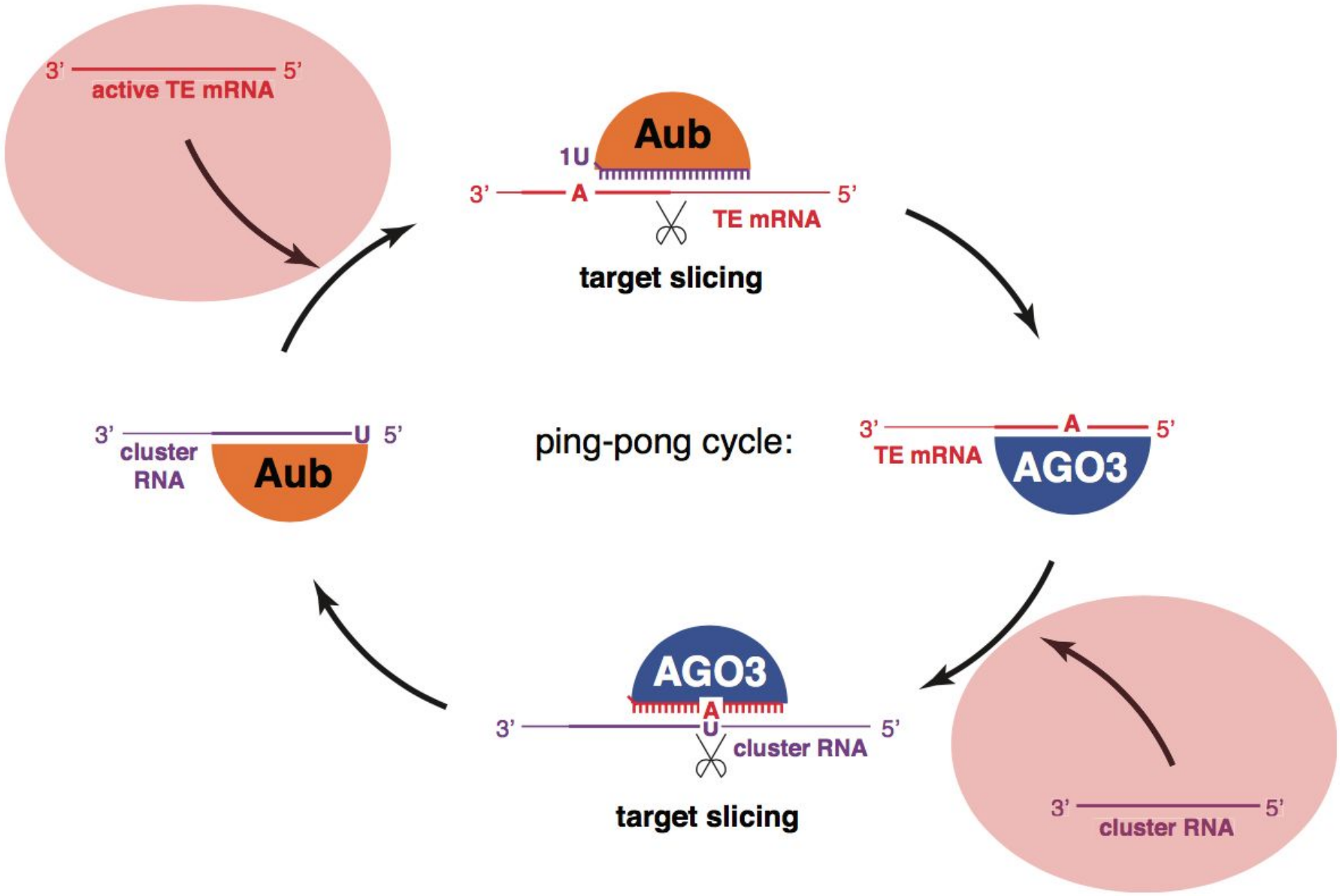


The ping-pong cycle amplifies TE silencing

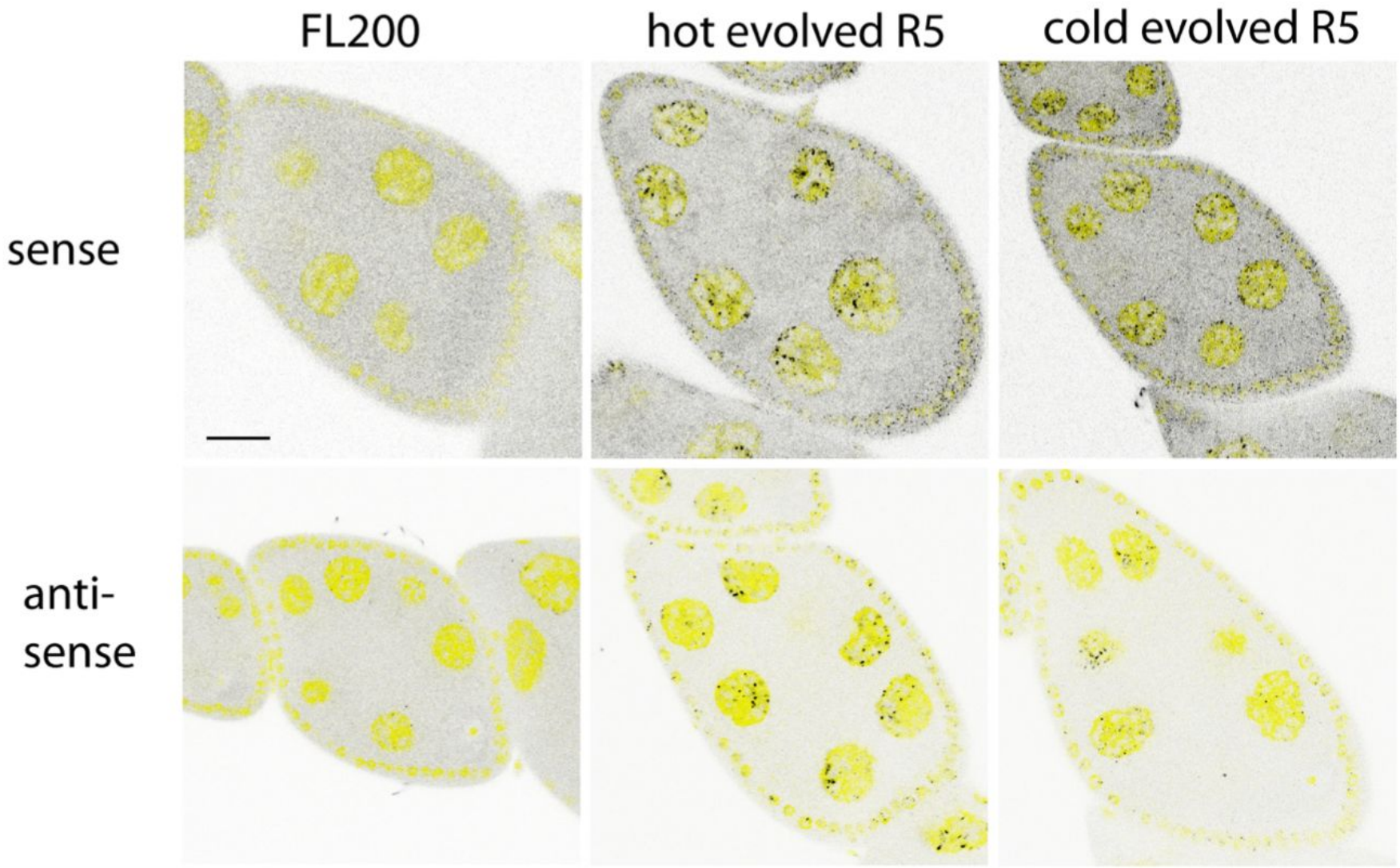


Brennecke et al. (2007) Cell
Gunarwadane et al. (2007) Science

Ping-pong requires sense and antisense mRNA

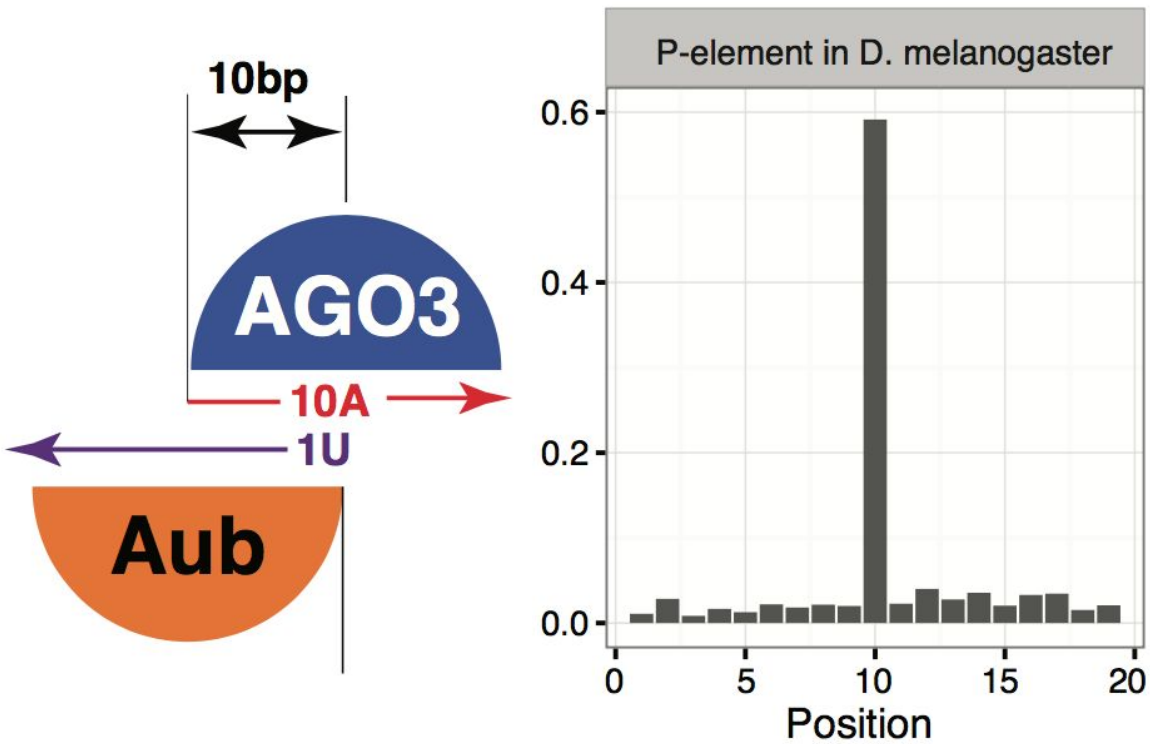


Single molecule RNA-FISH shows sense and antisense transcripts in the germline

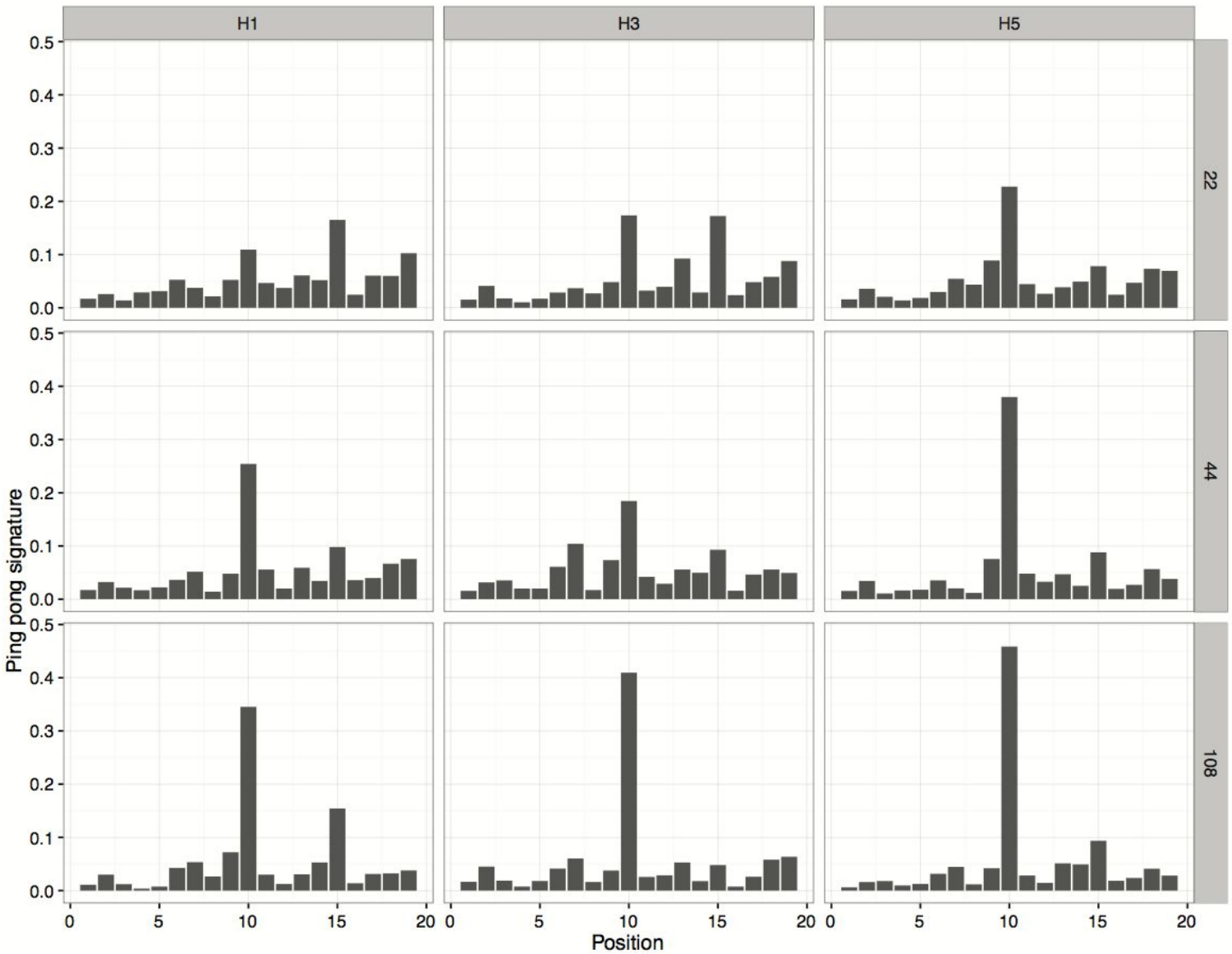


Ping-pong signature shows extent of silencing amplification

The ping-pong signal is a probability distribution, showing for 20bp within the 5′ end of a randomly picked sense piRNAs the probability that an antisense piRNA exists having the given distance. A strong peak at position **10** suggests efficient piRNA amplification.



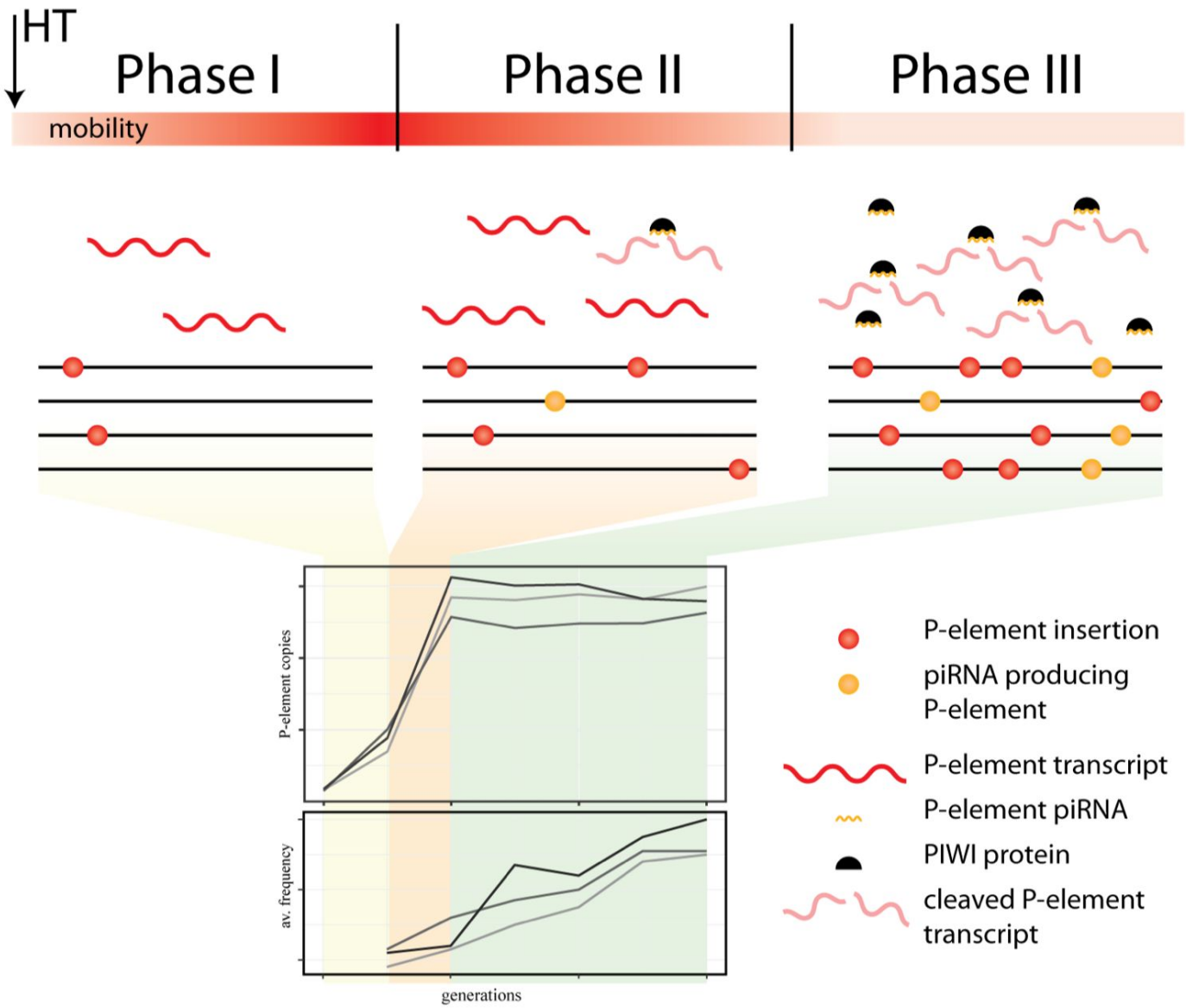
Amplification of silencing increases during the invasion



Summary: Why is activity decreasing

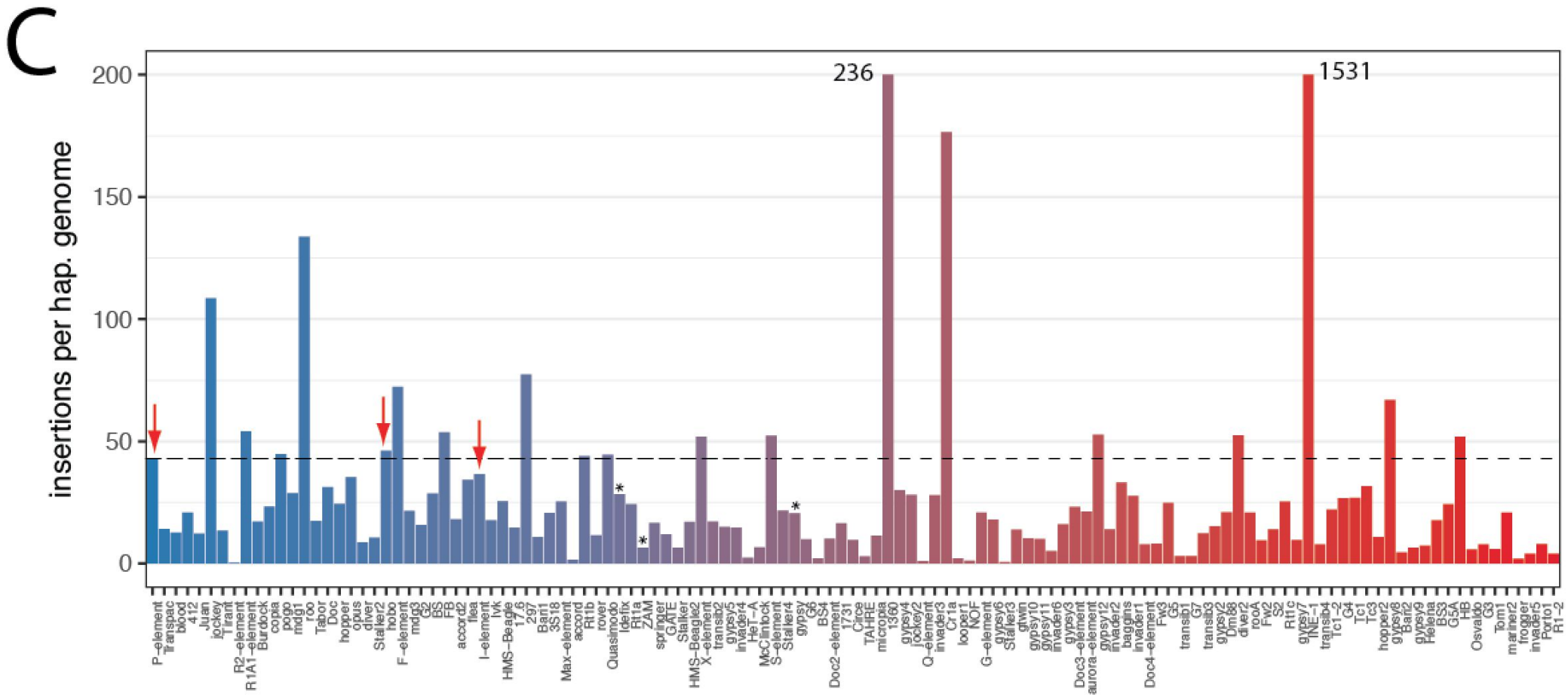
- all possible insertion sites occupied ✗
- truncated copies downregulate activity ?
- piRNAs ✓

Shotgun silencing



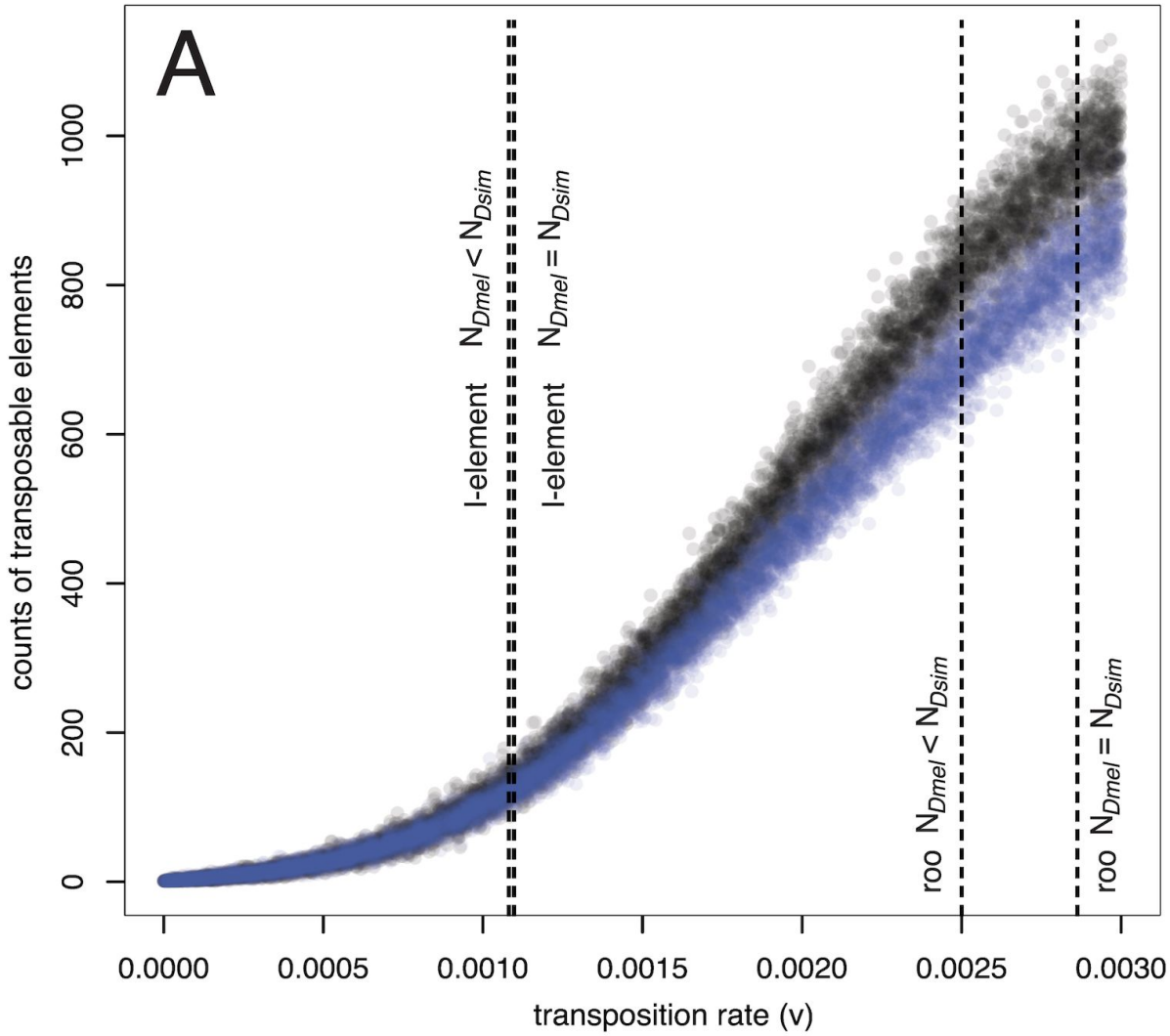
A problem with our understanding of TE abundance

We know that TE abundance varies within species



Data from
Kofler 2015 PLoS Genetics

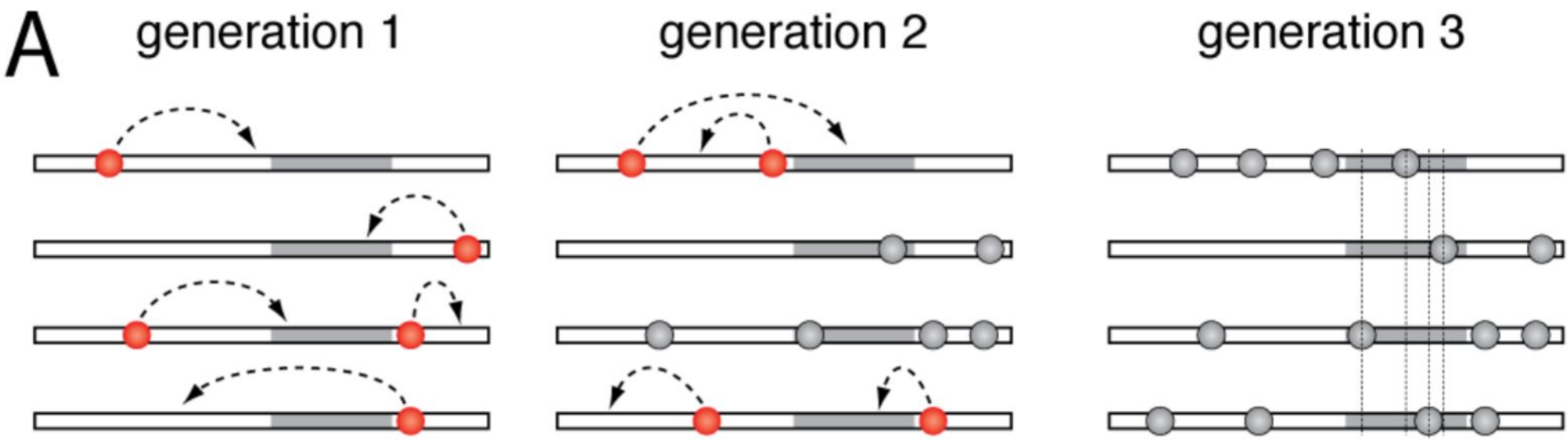
Previous explanation - differential equilibrium with transposition selection balance



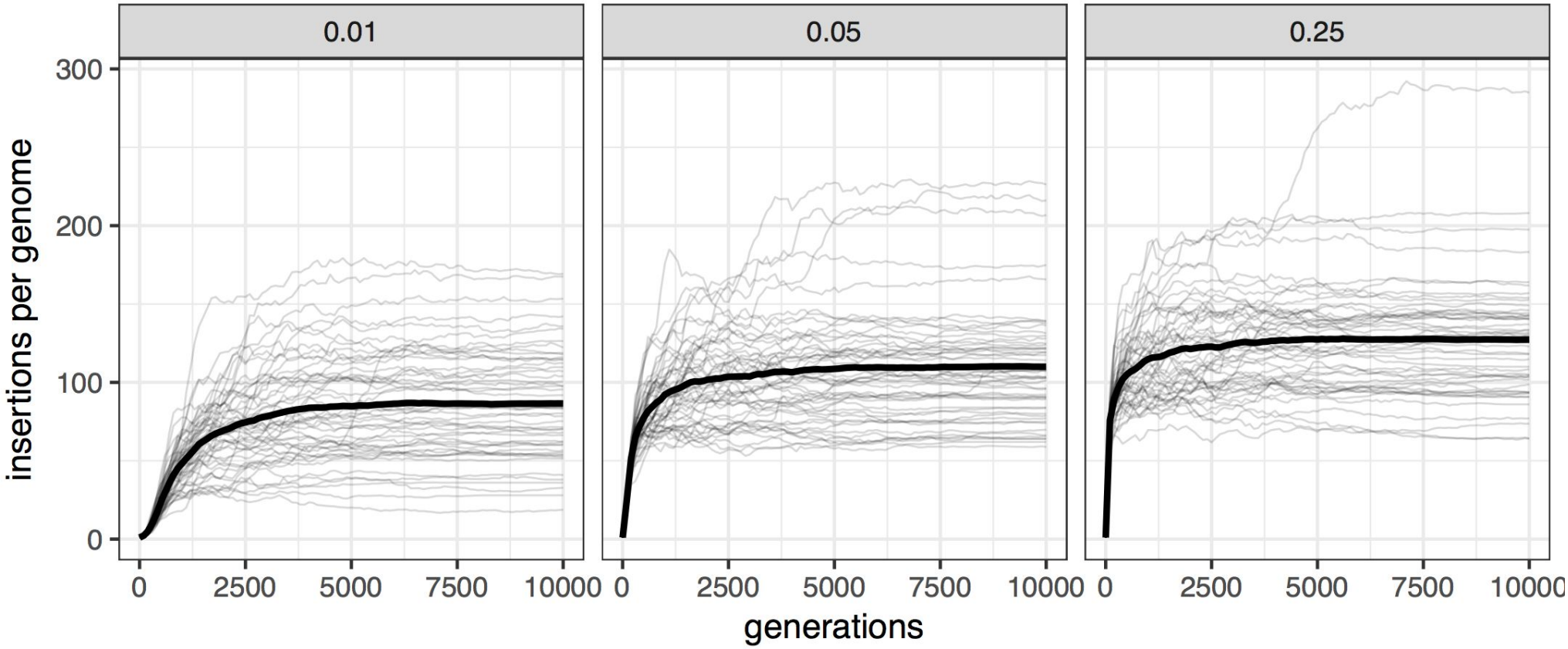
Key:
Transposition rate differences among families

Kofler 2015 PLoS Genetics

The novel “trap model”



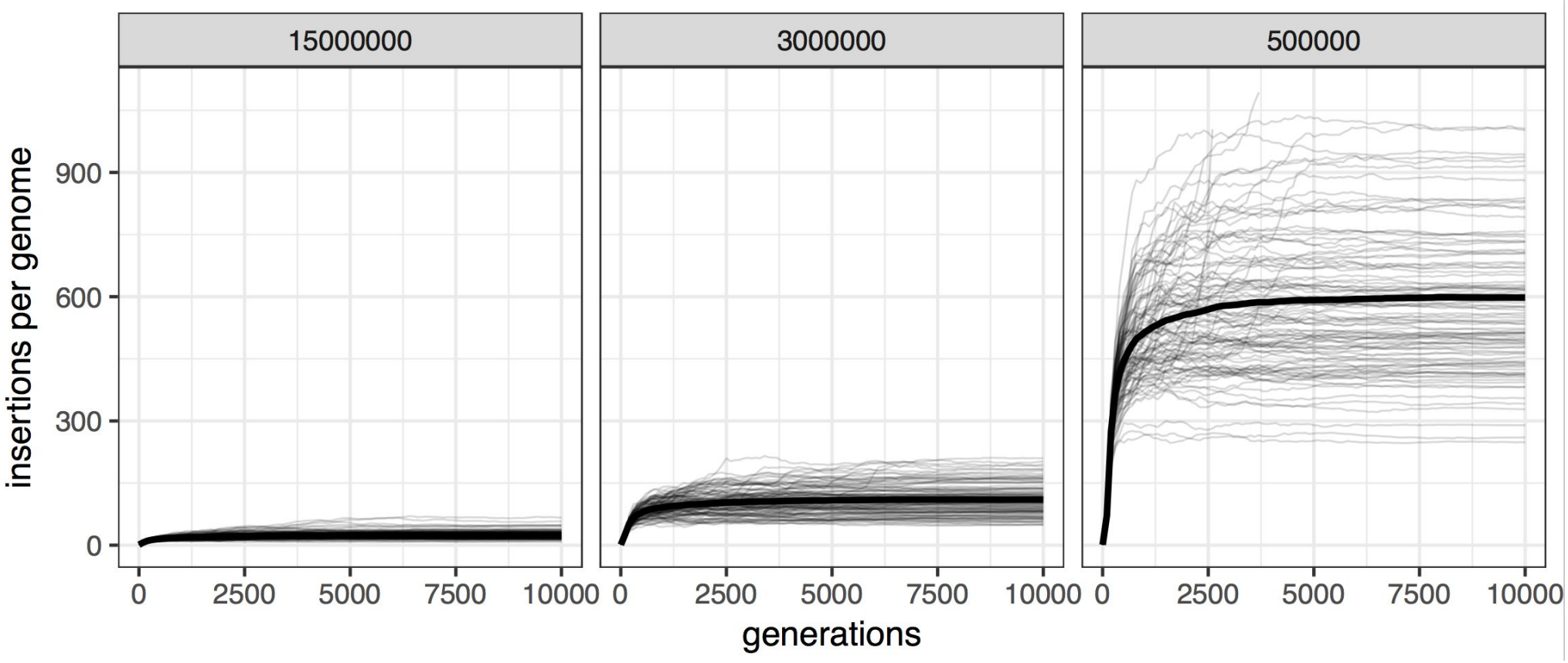
Trap model predicts equal TE abundance irrespective of the transposition rate



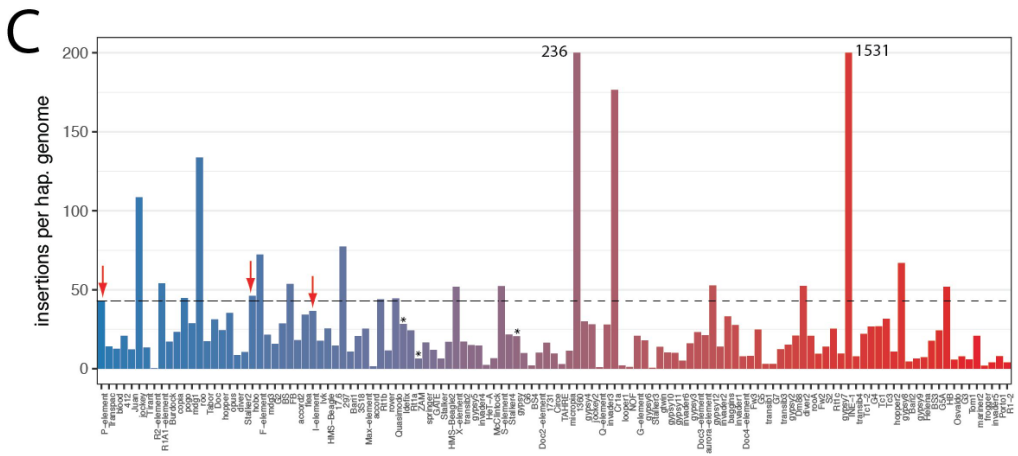
1.) Same plateauing level irrespective of the transposition rate.
2.) Plateauing level maintained over many generations

- 68 million possible insertion sites
- 3 million piRNA cluster sites
- Recombination rate of *D. melanogaster*
- Population size of 1000
- Neutrality of TE insertions
- Start insertions: 1000 randomly distributed over the population

Solely varying piRNA cluster sizes make a difference; But that's identical within species



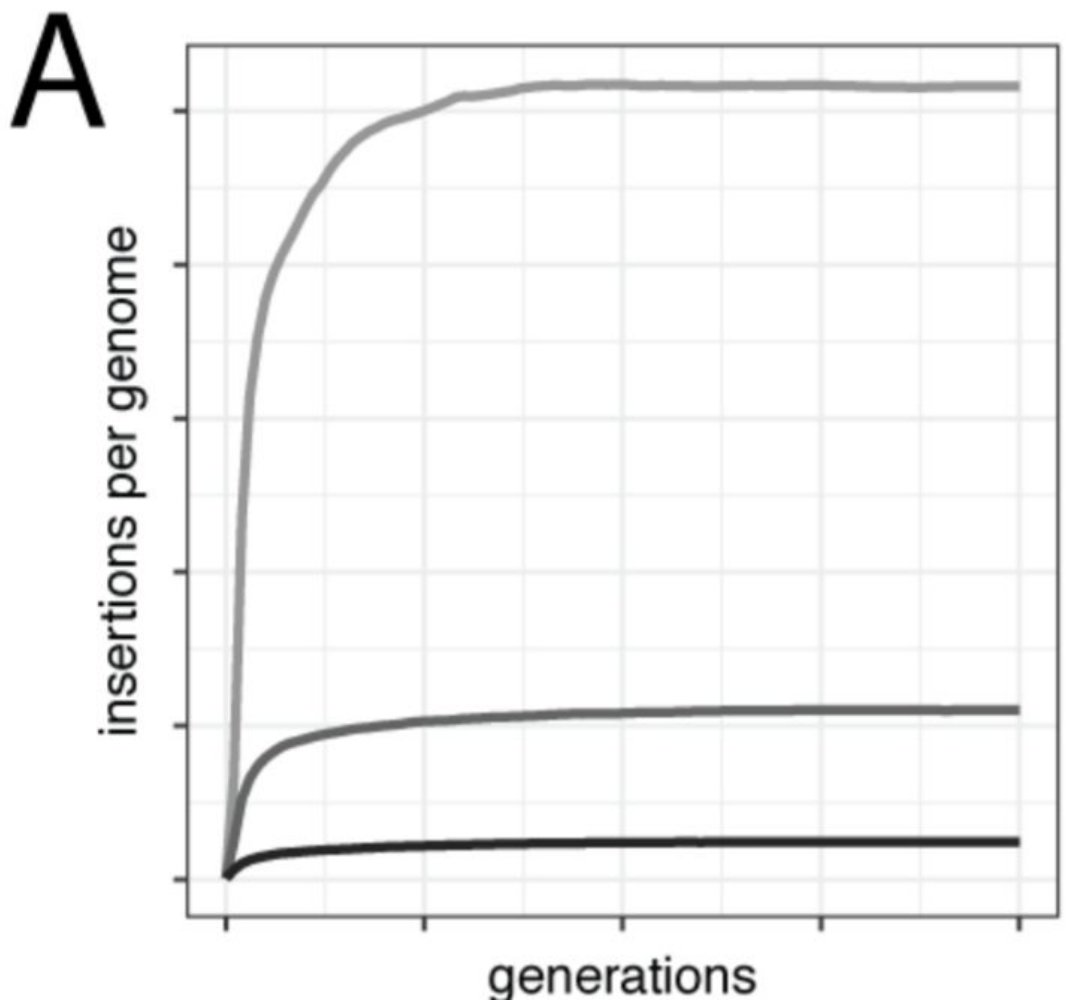
Major Question:
Why is the abundance of TE families so different within species?



Possible explanations

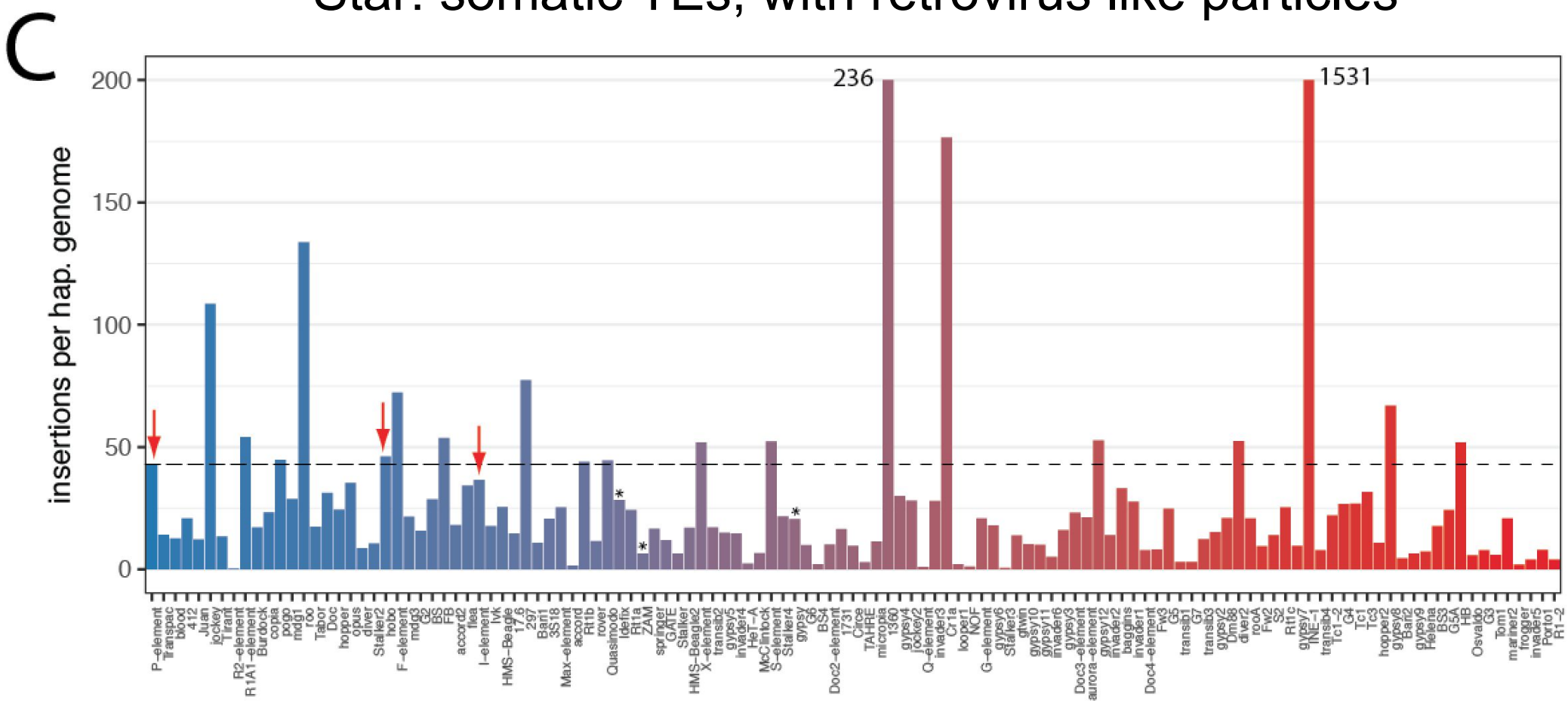
Incomplete quantitative understanding of trap model

Some families may require more than one piRNA cluster insertion.



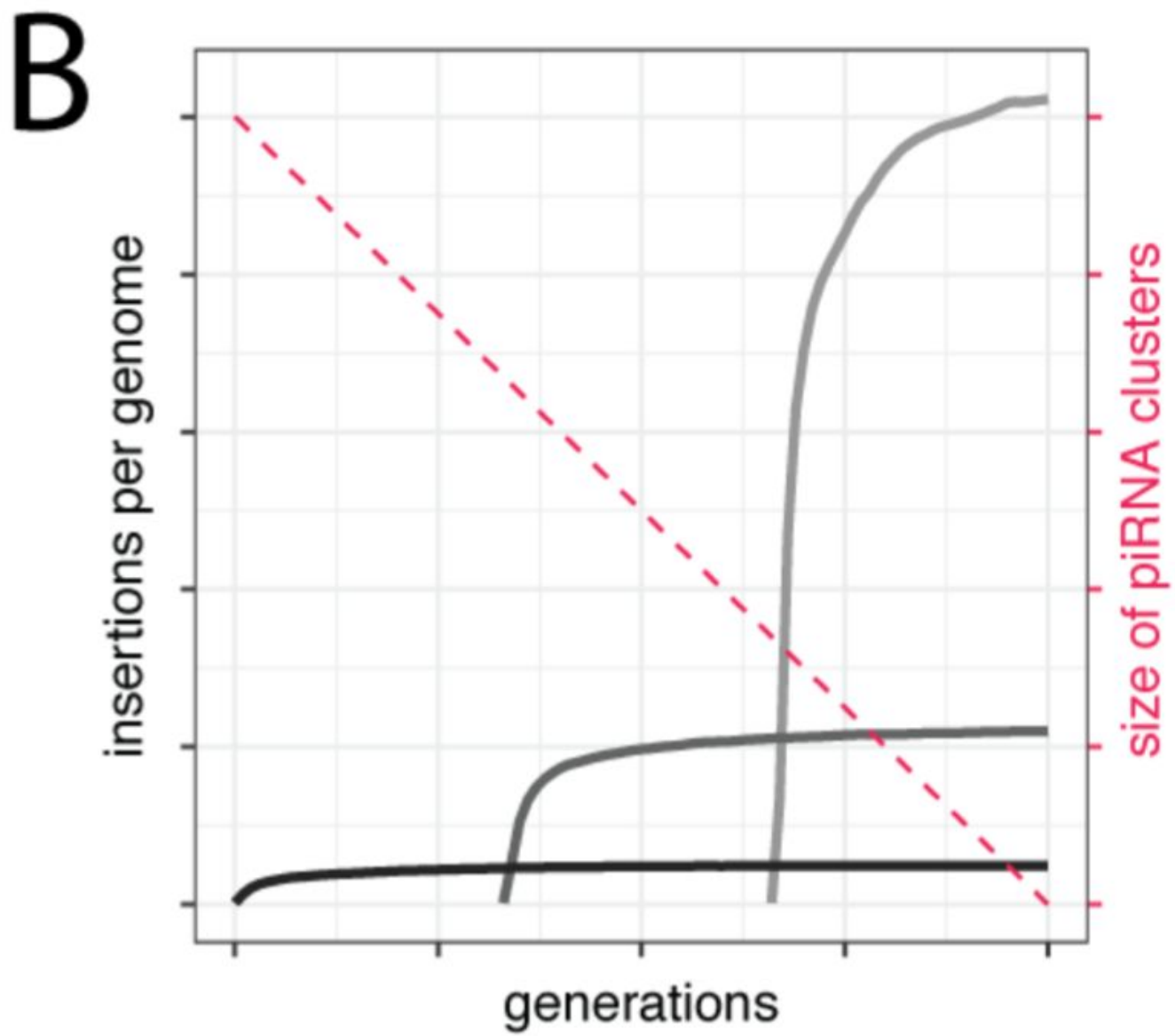
Two differentially silenced groups of TEs (soma vs germline) suggests incomplete understanding)

Star: somatic TEs, with retrovirus like particles

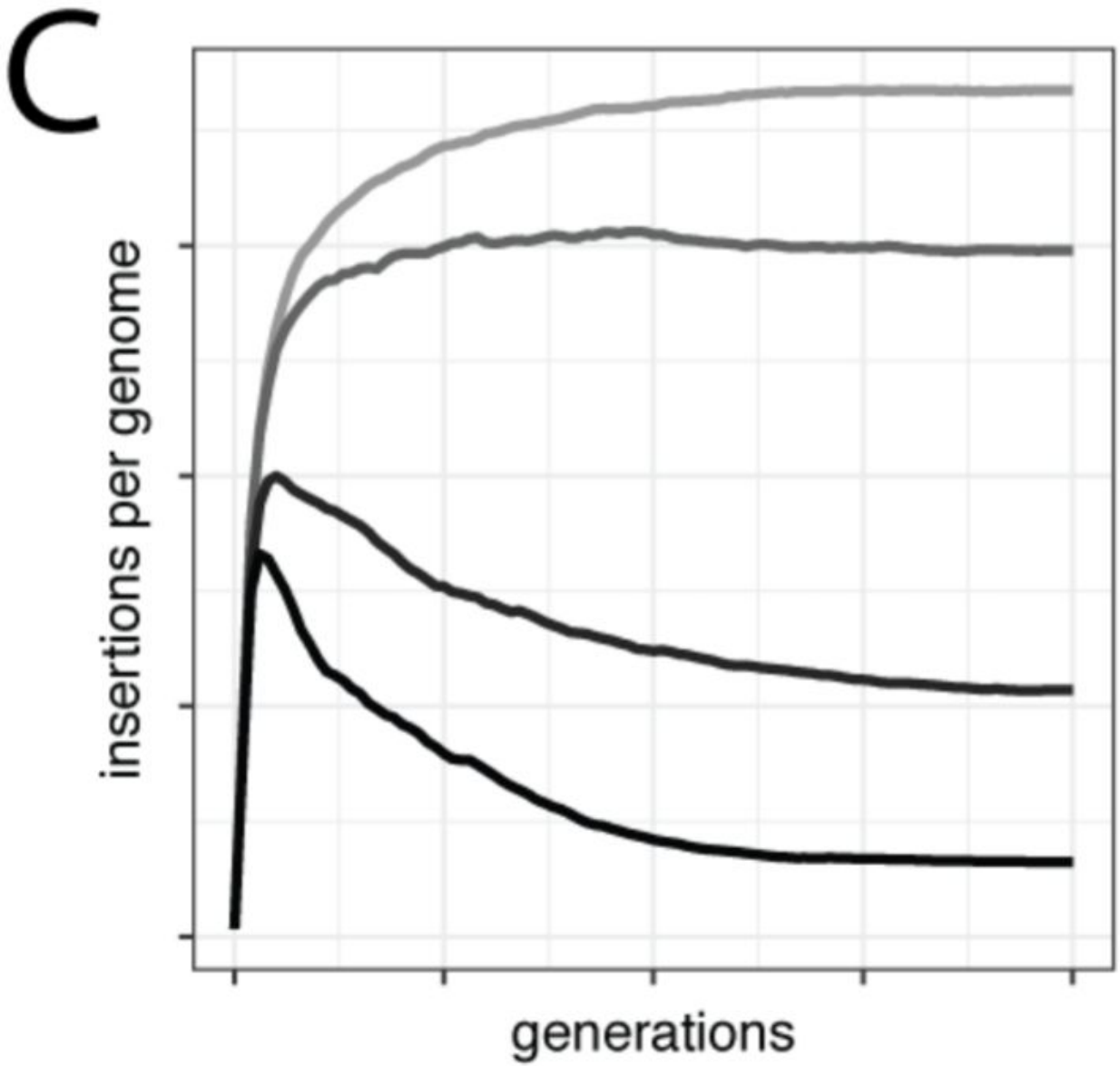


Data from
Kofler 2015 PLoS Genetics

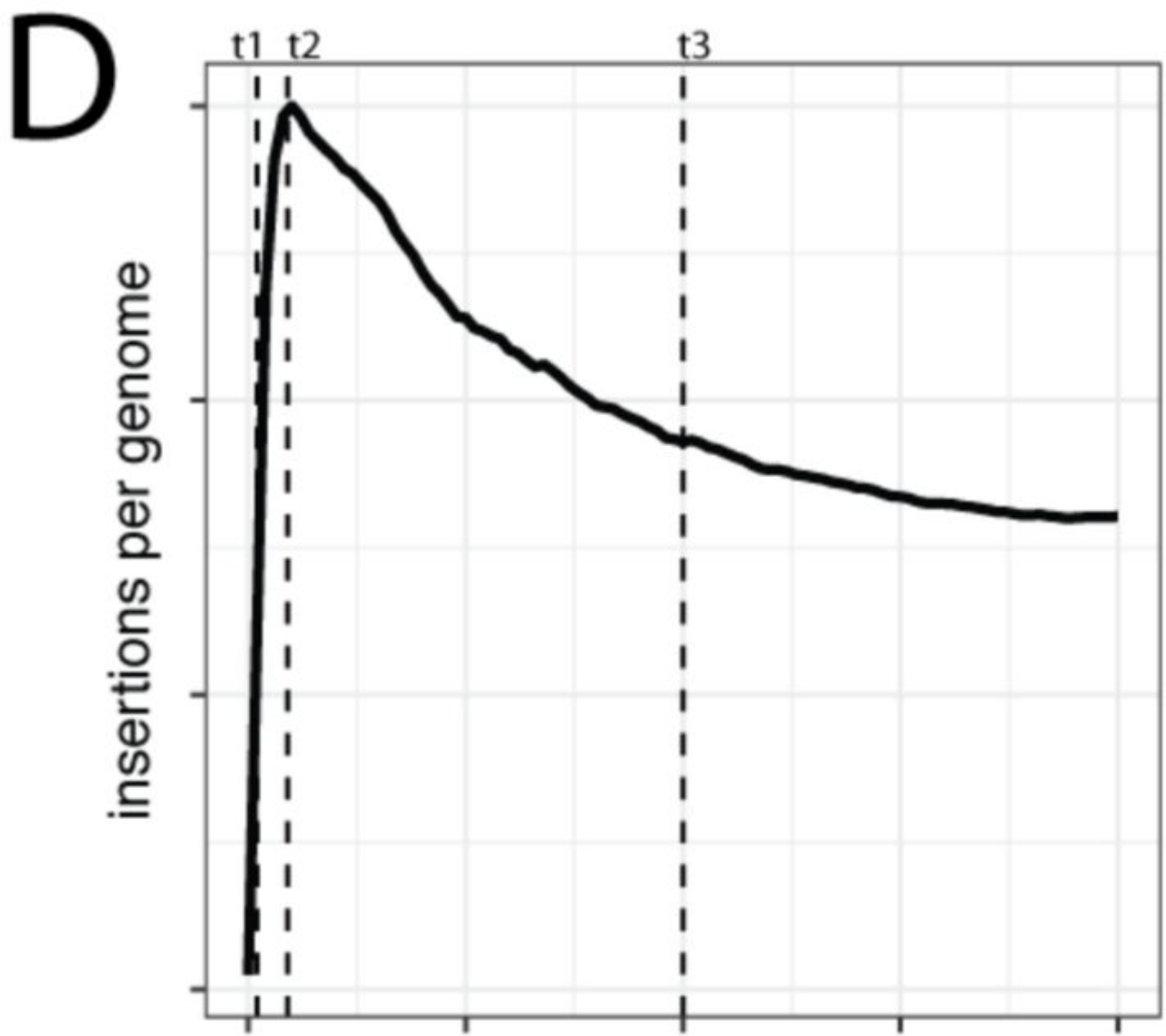
Rapidly evolving piRNA cluster size



Differential neg. selection among families

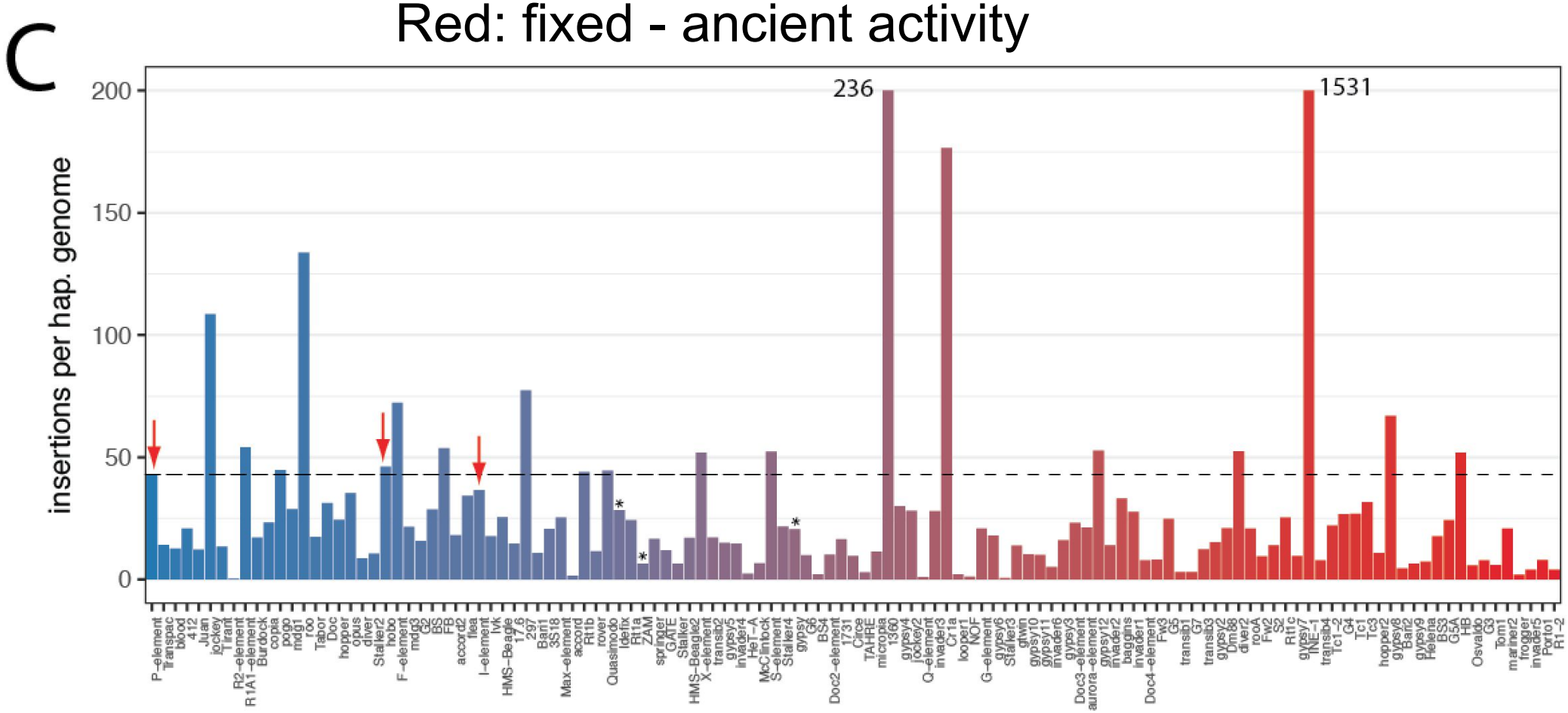


Different stage at the life cycle



TEs certainly are at different stages of the life cycle

Blue: segregating - recent activity
Red: fixed - ancient activity



Data from
Kofler 2015 PLoS Genetics

Which factors dominate?
Are they all equally important?
How can we test this?

Acknowledgments

Christian Schlötterer

Kirsten-Andre Senti

Viola Nolte

Ray Tobler

Daniel Gomez Sanchez

Filip Wierzbicky