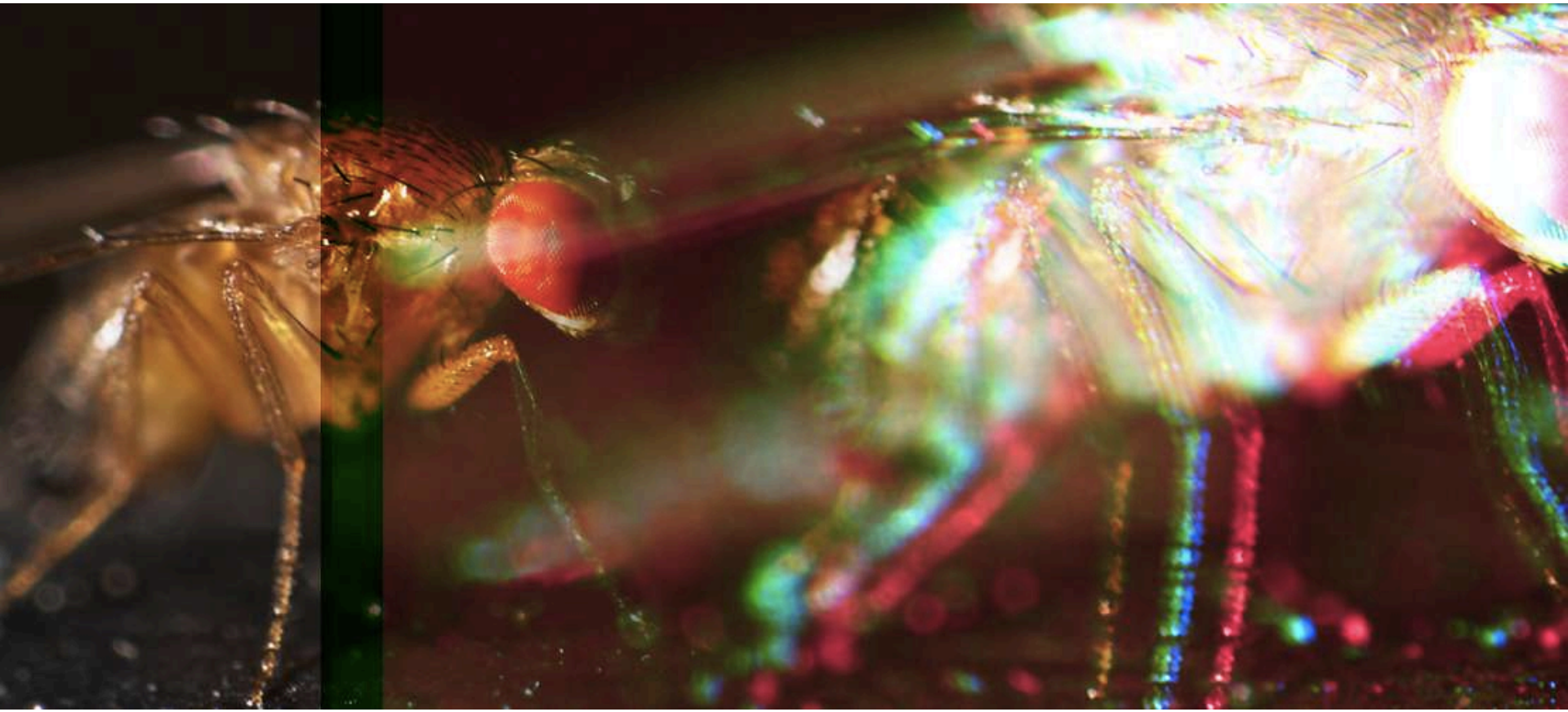
The role of natural transposable element insertions in stress response

Josefa González
Institute of Evolutionary Biology (CSIC-UPF)

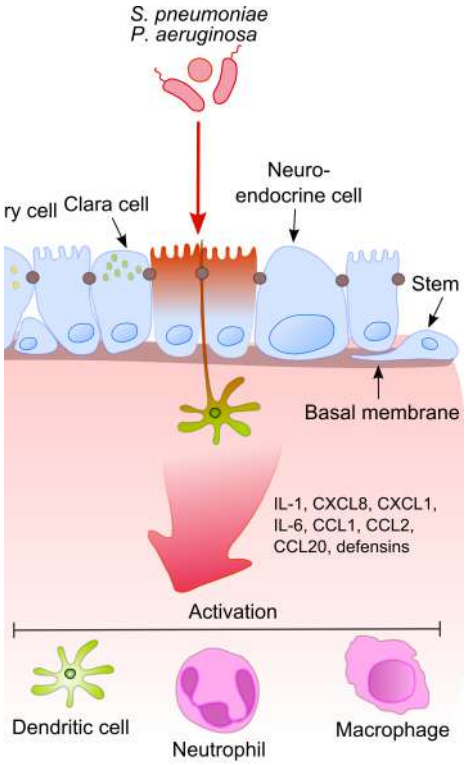


@GonzalezLab_BCN

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

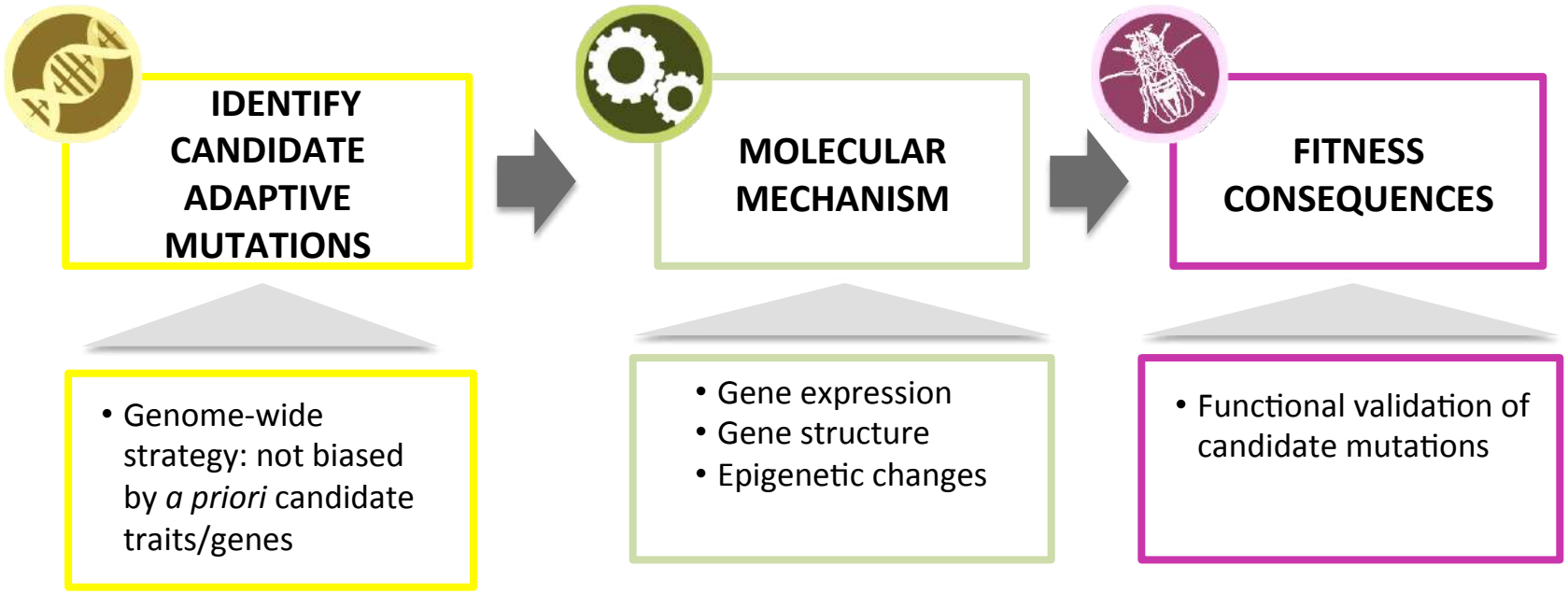
UNDERSTANDING ADAPTATION

Adaptation underlies the ability of species to survive in changing environments, host-pathogen interactions, resistance to pesticides and drugs



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

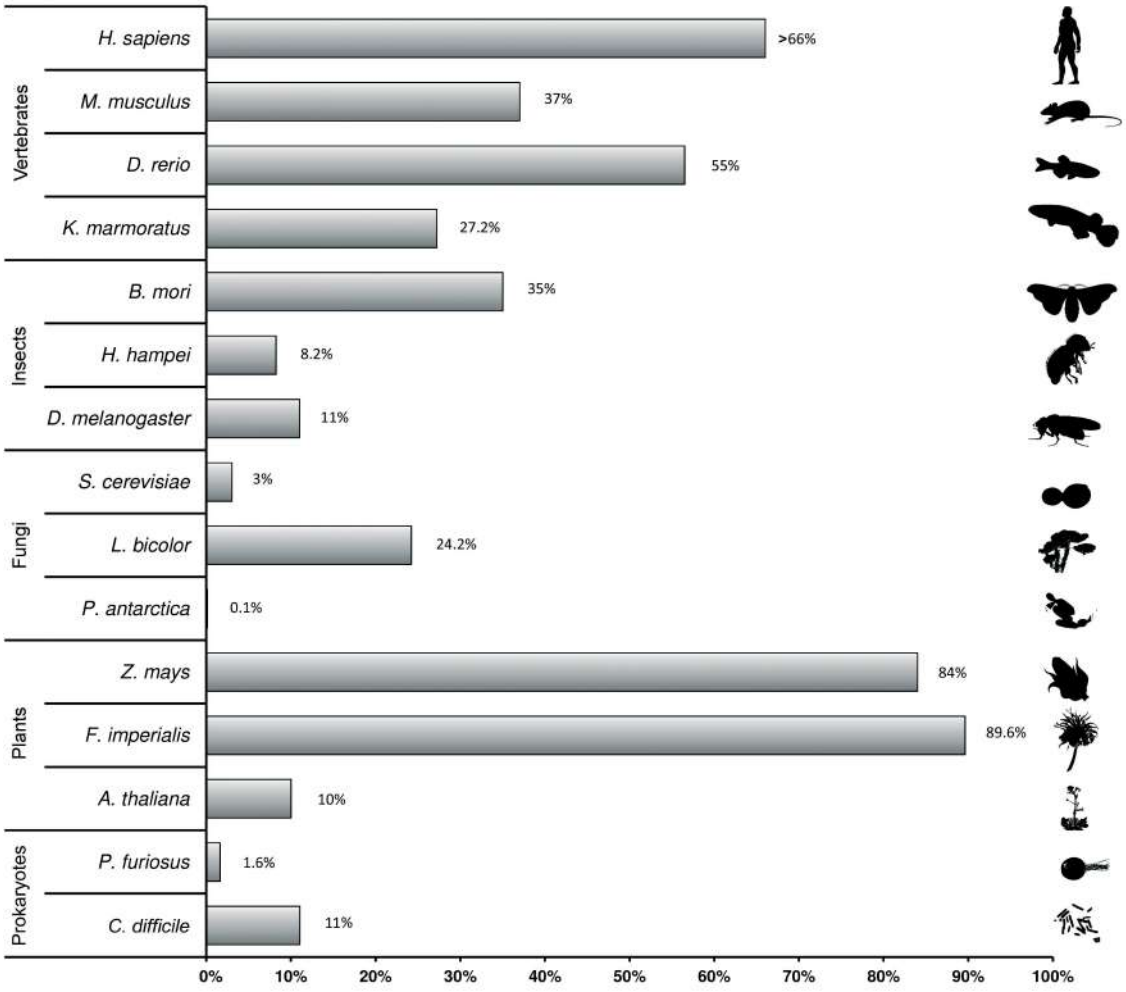
UNDERSTANDING ADAPTATION



Barret and Hoekstra 2008; Storz and Wheat 2010

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

TEs represent a sizable proportion of all genomes



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

TEs are diverse

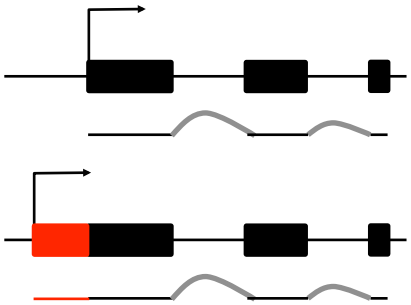
Class I: retrotransposons

Class II: DNA transposons

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

TEs are a potent source of regulatory mutations

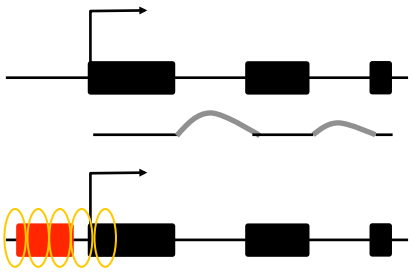
Alternative transcription start sites



Batut et al 2013 Genome Res

Merenciano et al 2016 PLoS Gen

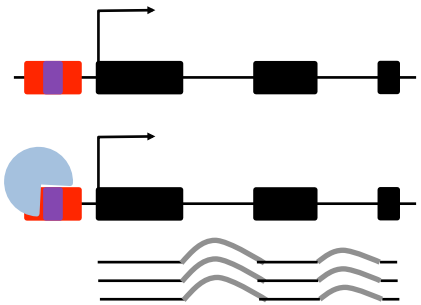
Epigenetic marks



Lee and Karpen 2017 eLIFE

Guio et al 2018 Sci Reports

Addition of TFBSs



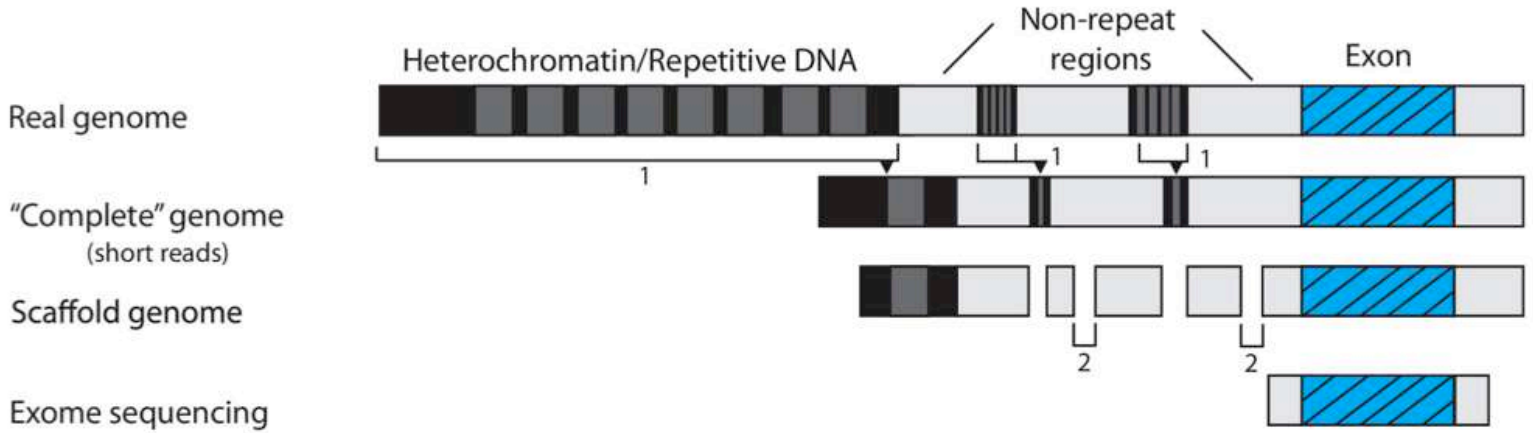
Sundaram et al 2014 Genome Res

Chuong et al 2016 Science

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

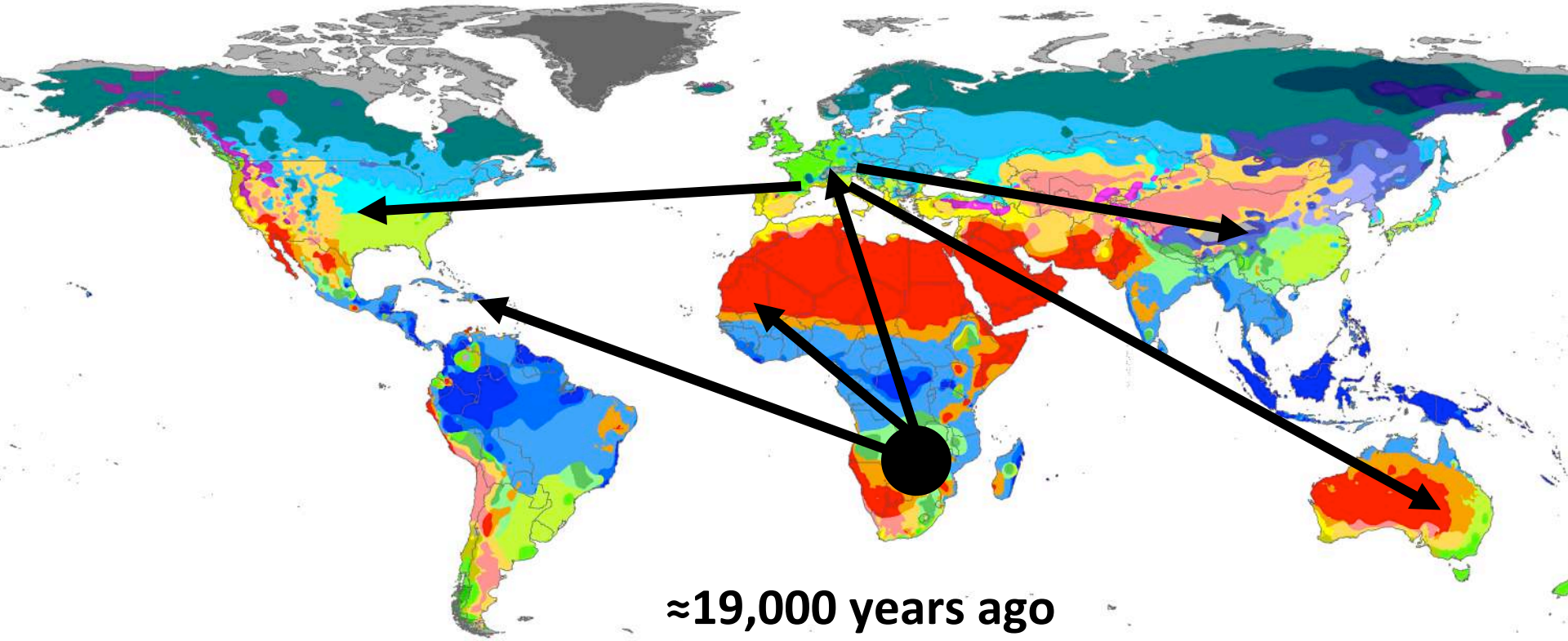
Ignored because difficult to annotate

- Most current genomes



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

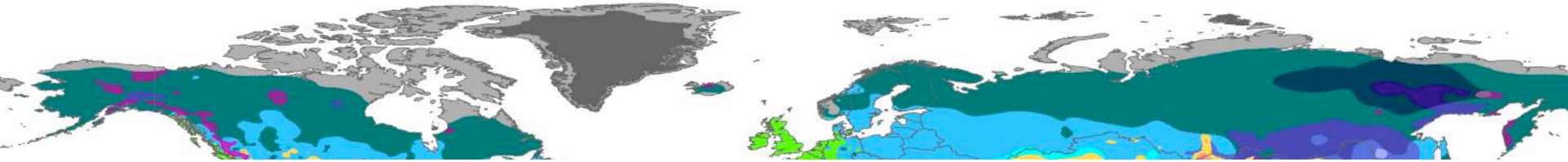
Why *Drosophila melanogaster*?



Le and Stephan 2006, Thornton and Andolfatto 2006, Arguello et al 2019

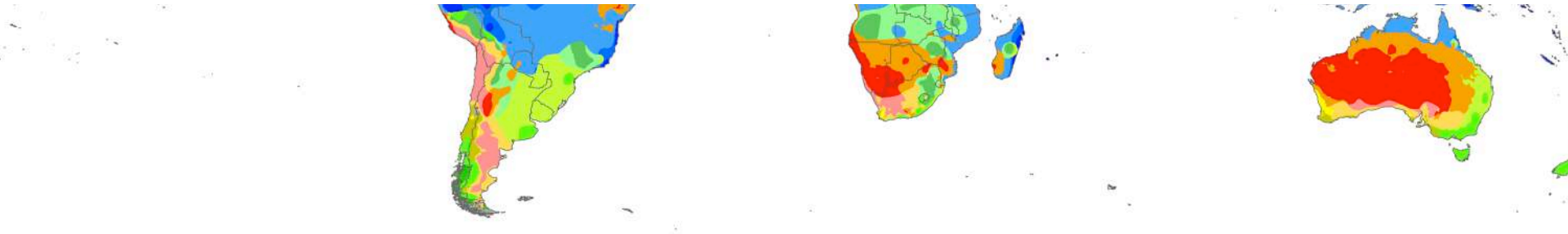
GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Why *Drosophila melanogaster*?



Recent adaptations must be common in this species

- **new pathogens**
- **new diet**
- **range of temperatures**
- **toxic environments**



Le and Stephan 2006, Thornton and Andolfatto 2006, Arguello et al 2019

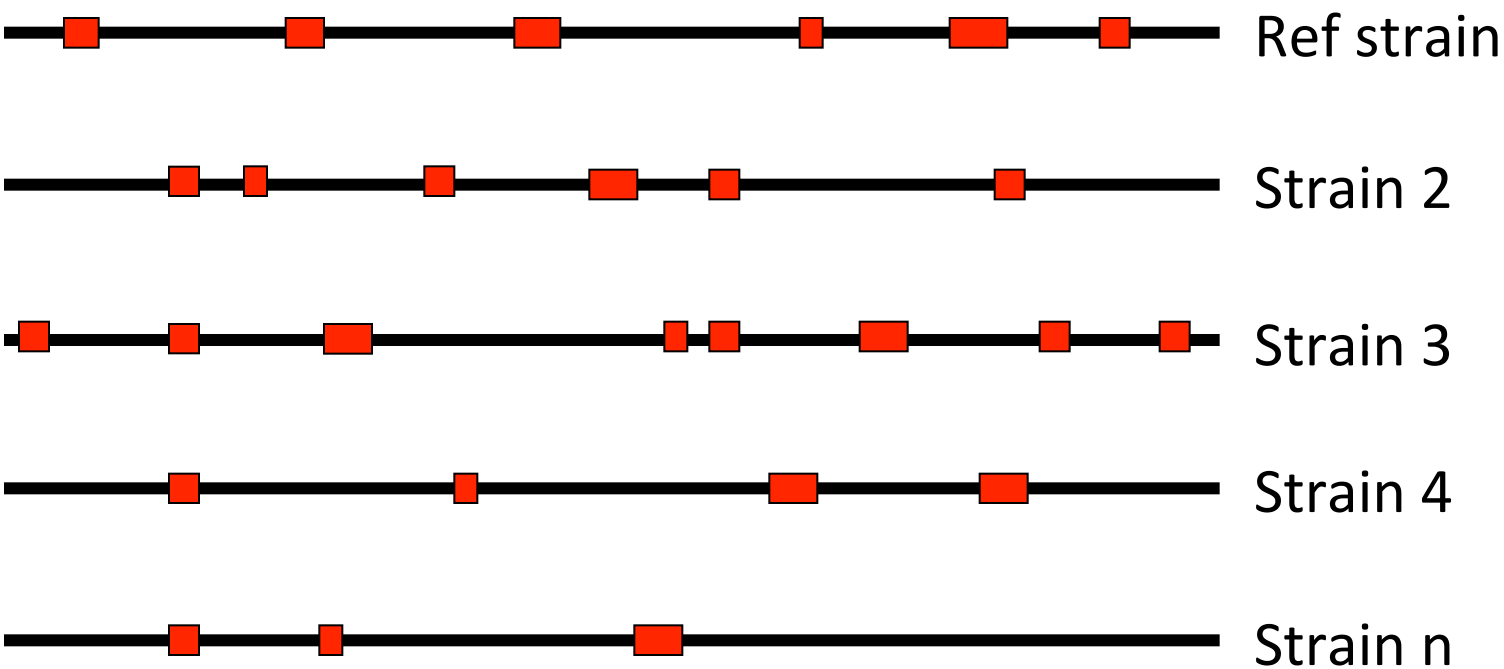
GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Transposable elements in Drosophila

Ref strain

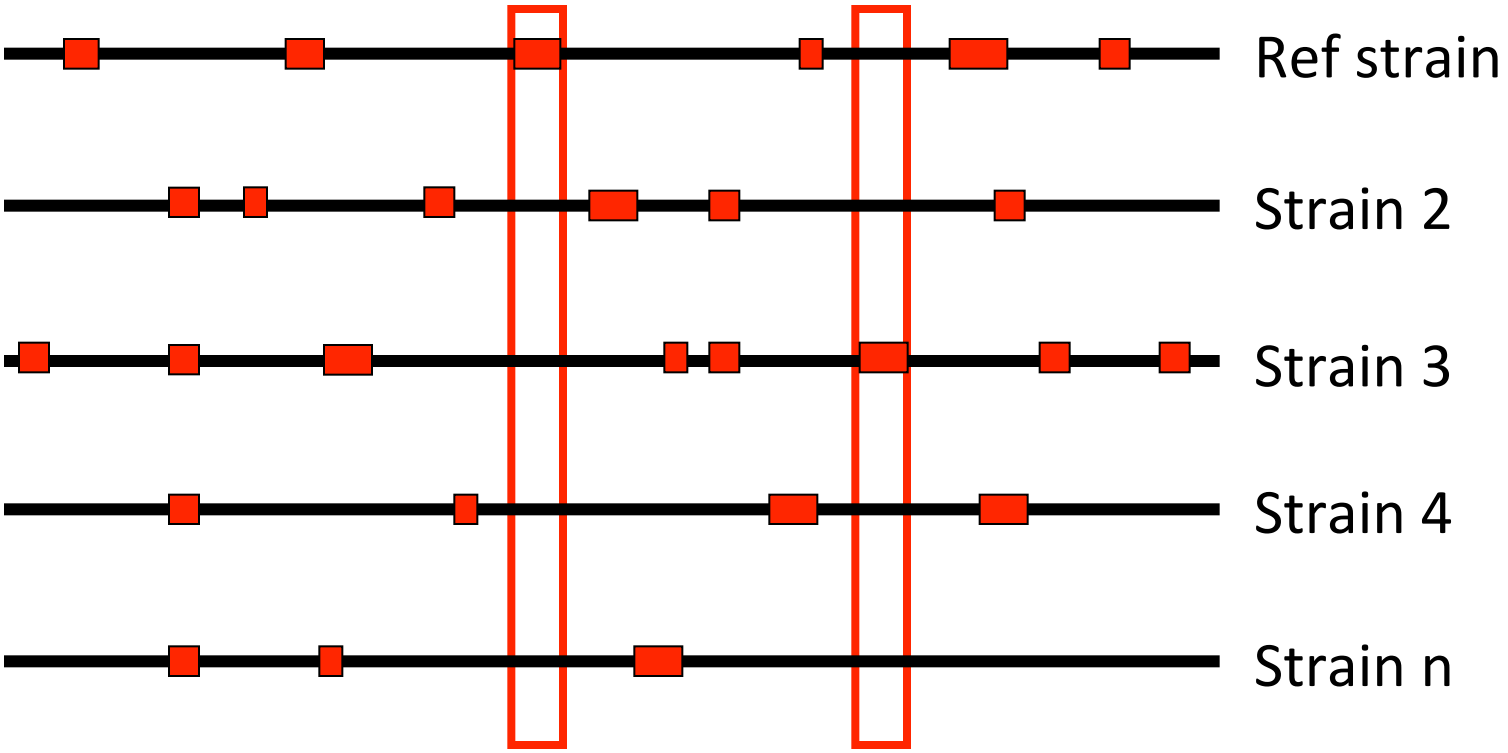
GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Transposable elements in Drosophila



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

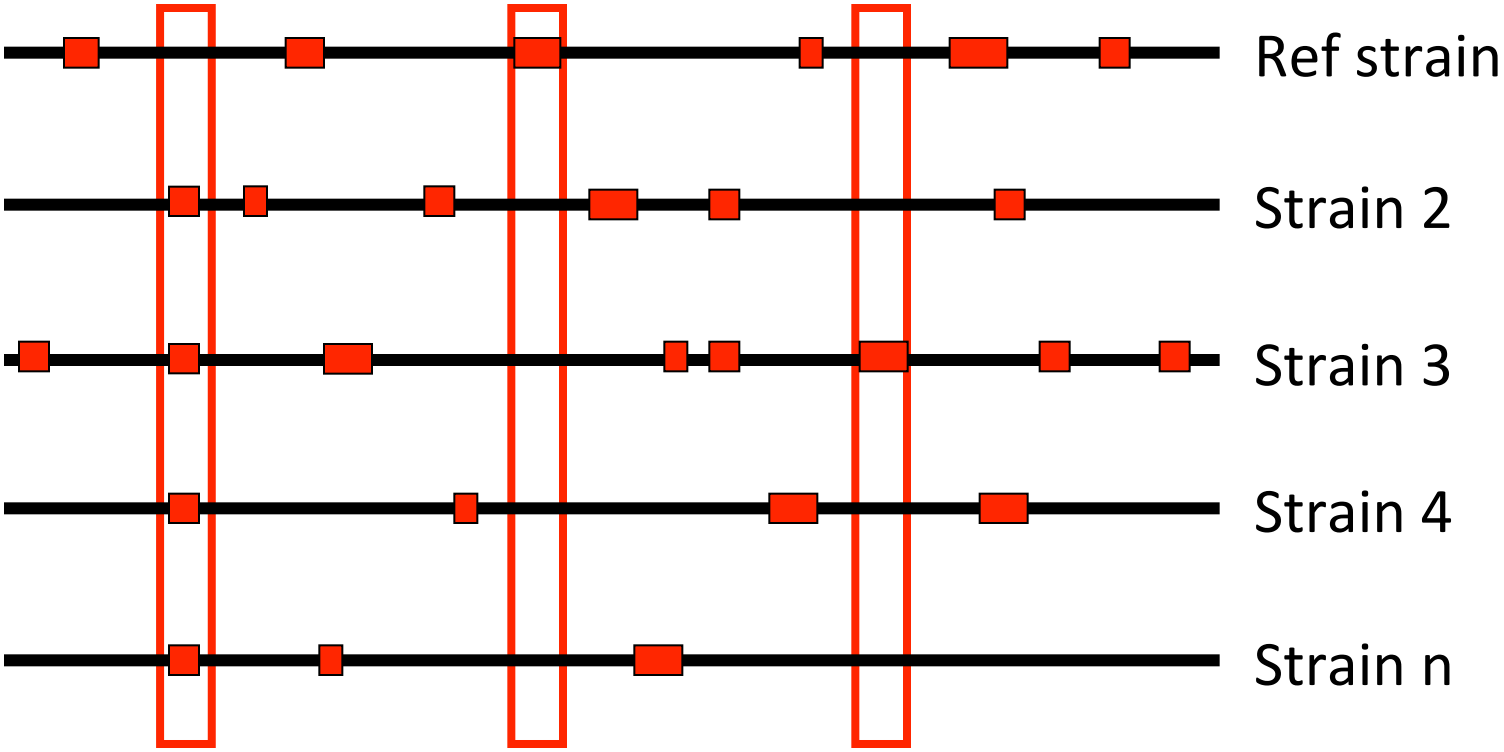
Transposable elements in Drosophila



Strong selection against TE insertions due to ectopic recombination (Petrov et al 2011 Mol Biol Evol)

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Transposable elements in Drosophila

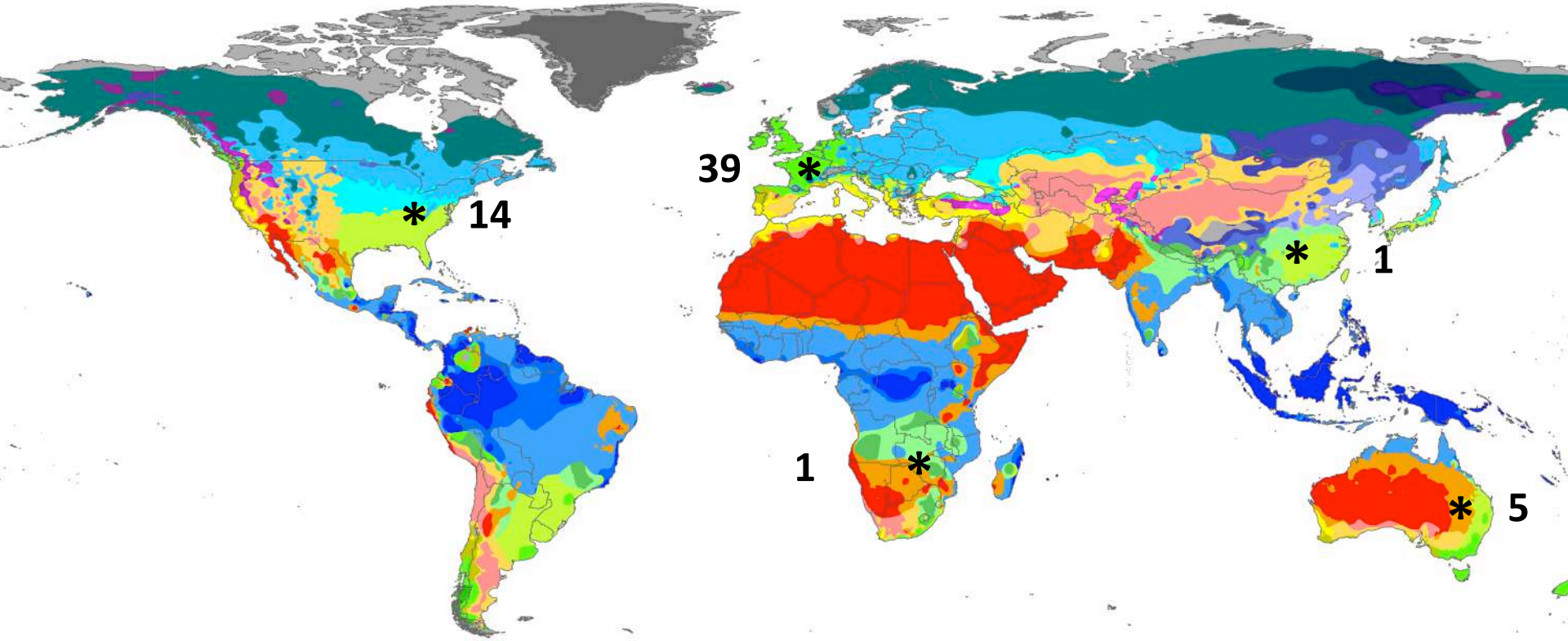


Adaptive effect overcomes the initial disadvantage?

Strong selection against TE insertions due to ectopic recombination (Petrov et al 2011 Mol Biol Evol)

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Identify candidate adaptive mutations



91 samples from 60 natural populations

Rech et al 2019 PLoS Gen @rechgab

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

DROSEU
EUROPEAN DROSOPHILA
POPULATION GENOMICS CONSORTIUM

Thomas Flatt
Uni Fribourg

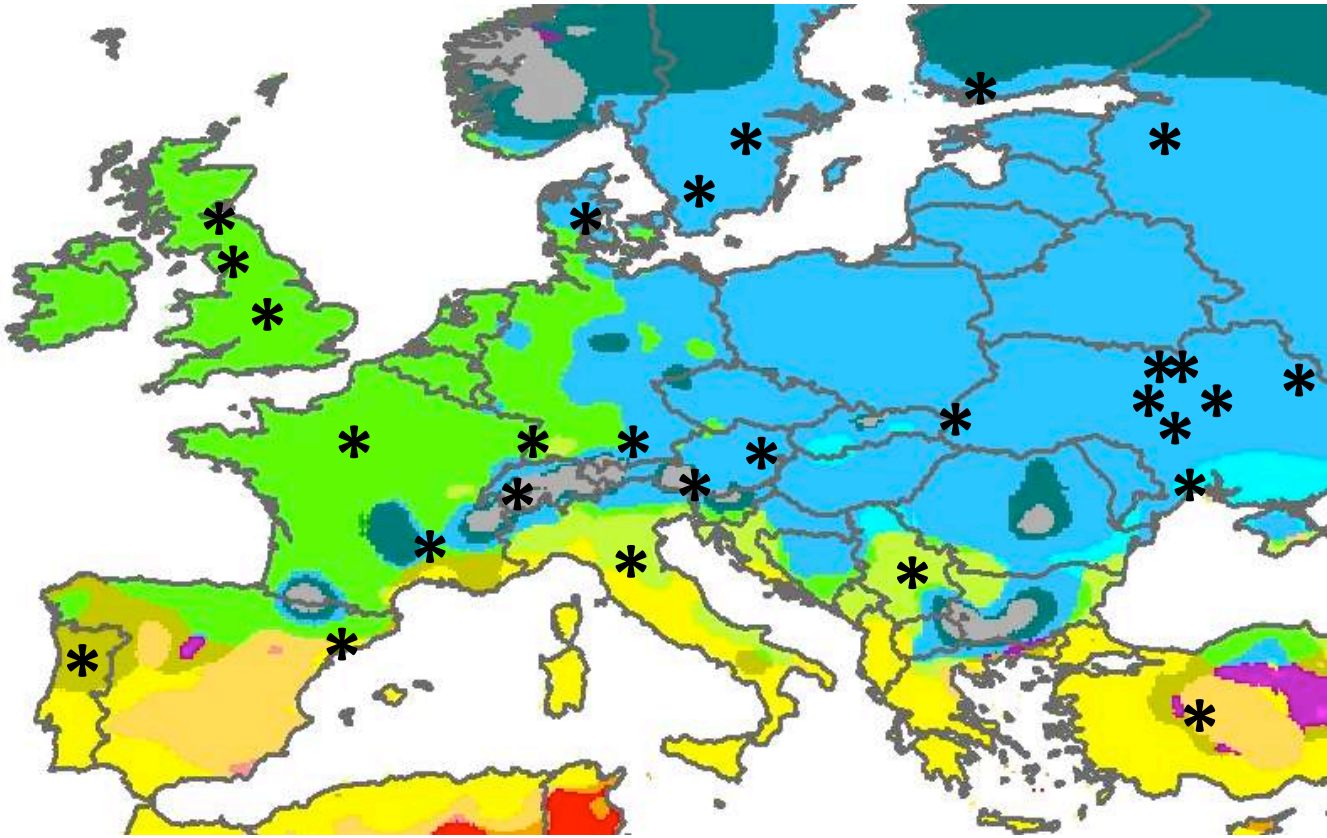
Martin Kapun
Uni Zurich

J. González
IBE (CSIC-UPF)

61 labs: 24 European countries + Morocco, Israel, US, Colombia, Ecuador, Australia

Pooled-seq data:
2014: 32 populations
2015: 39 populations
2016: 38 populations

Kapun et al 2018, bioRxiv



www.droseu.net

eseb
European Society for Evolutionary Biology

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

MELANOGASTER
CATCH THE FLY!

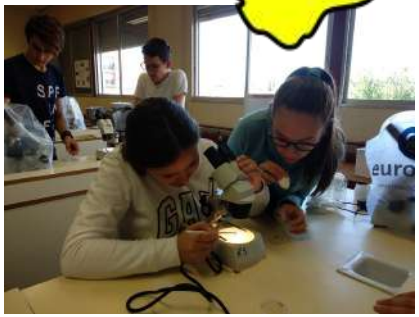
#MelanogasterCTF
http:// melanogaster.eu

CEIP Ramón y Cajal,
Pina de Ebro, Zahara Alonso

IES Benjamín Jarnes,
Osera de Ebro, Joan Torró

IES Eladio Cabañero, Tomelloso
Antonio Buendía y Ma. José Gómez

IES Antonio José Cavanilles
Alicante, Pepa Juan

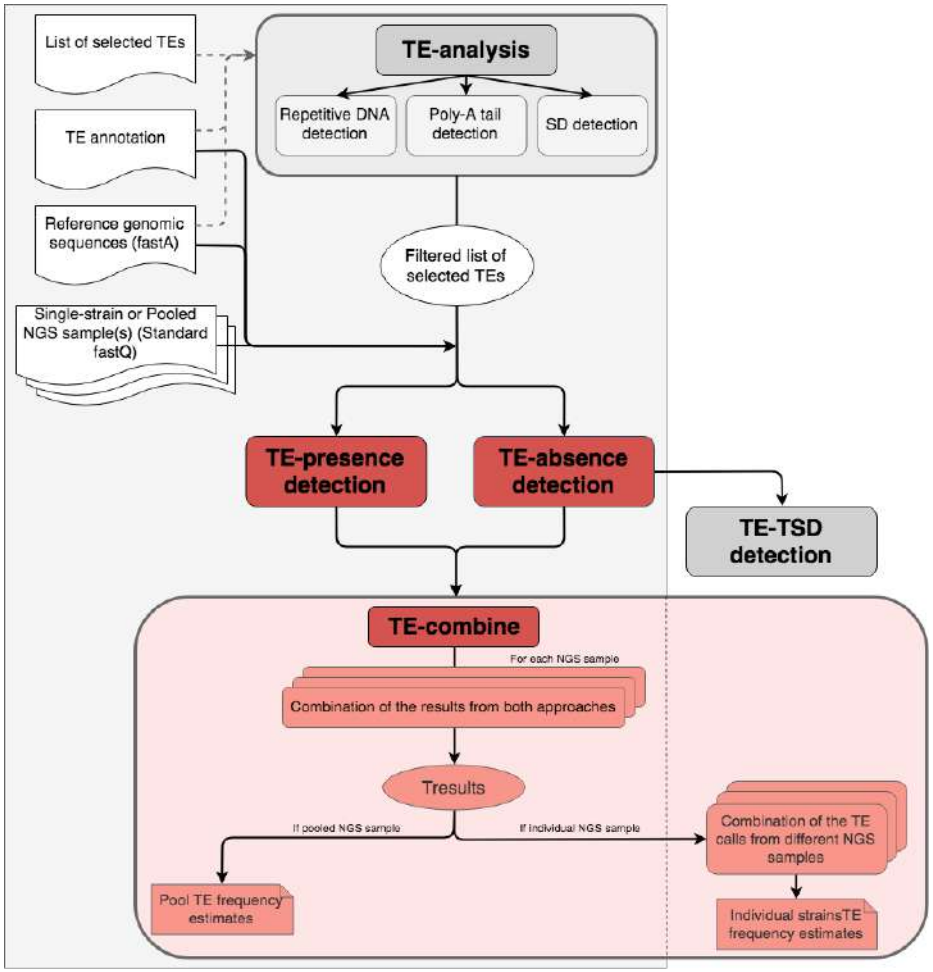


IES Alonso de Ercilla,
Ocaña
Belén Roldán

IES José de Mora, Baza
Silvana Castillo y David Gallardo

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
FECYT FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA
erc European Research Council

T-lex3: accurate assessment of TE population frequencies



Maria Bogaerts
PhD student

Individual and pooled NGS datasets
Detects TE presence and absence
Identifies miss-annotated TEs
Drosophila and humans reference insertions

New mapper: BWA-MEM
New filtering steps: increased accuracy
Available in Github: improved manual, test datasets
https://github.com/GonzalezLab/T-lex3

Bogaerts-Márquez et al (submitted)

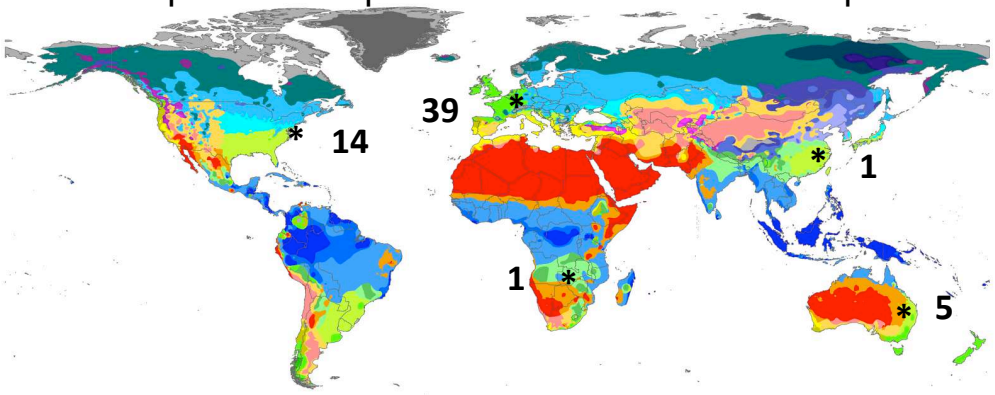
isem
Institut des Sciences de l'Evolution-Montpellier

GCCGCACTC CAA AC AGC CAA GAAACC González•Lab
Evolutionary and Functional Genomics

Stress, behaviour and development are shaped by TEs

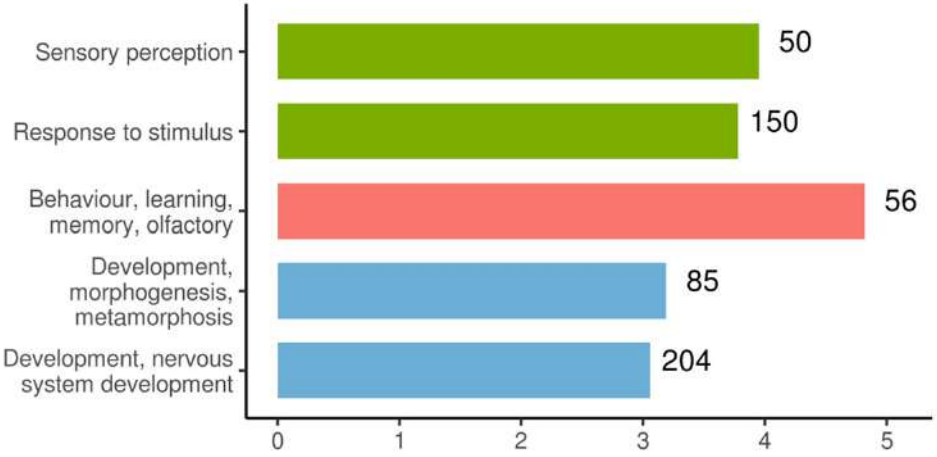
Gabriel Rech
Postdoc

Population frequencies estimates in 91 samples



- Age
- Length

- Selective sweeps
  - iHS
  - nSL
  - H12
- F_{ST}



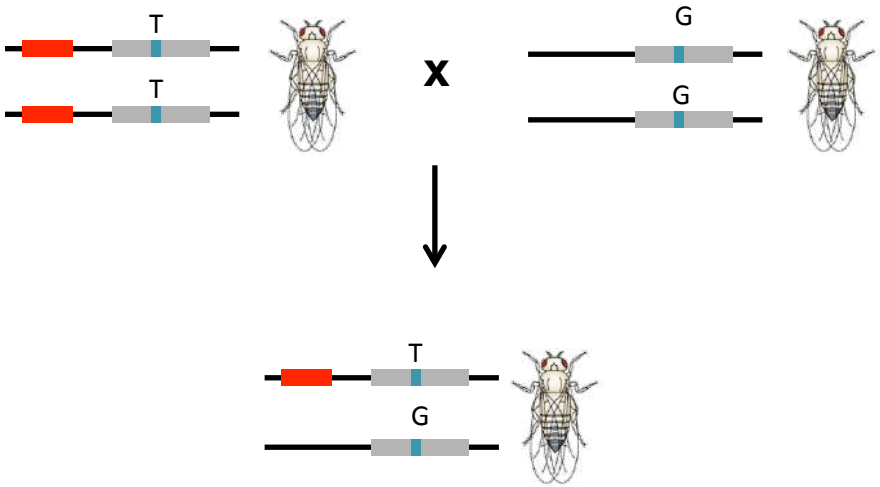
Rech et al 2019 PLoS Gen @rechgab

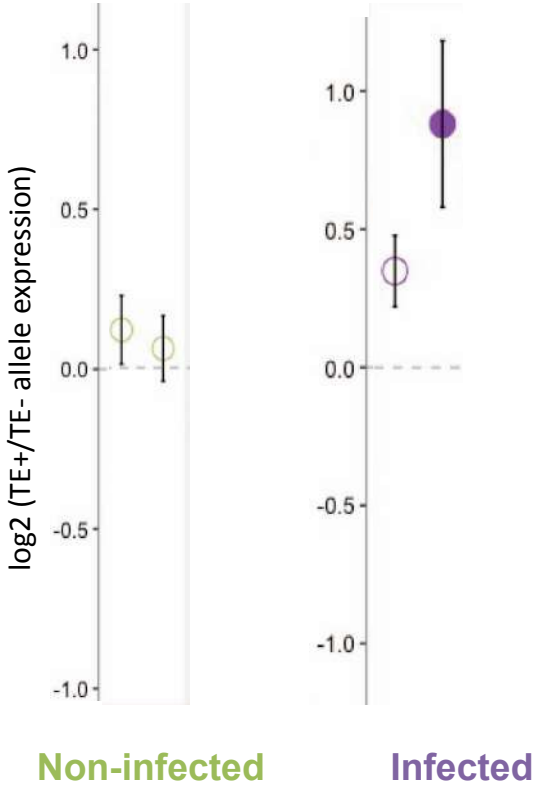
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

FBti0019386 allele is up-regulated in infected conditions

Anna Ullastres PhD

Allele-specific expression analyses

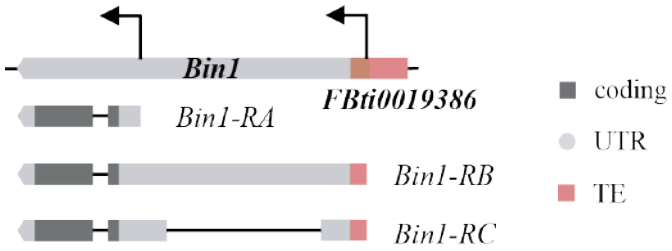




Ullastres et al 2019 bioRxiv

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

FBti0019386 adds a TSS

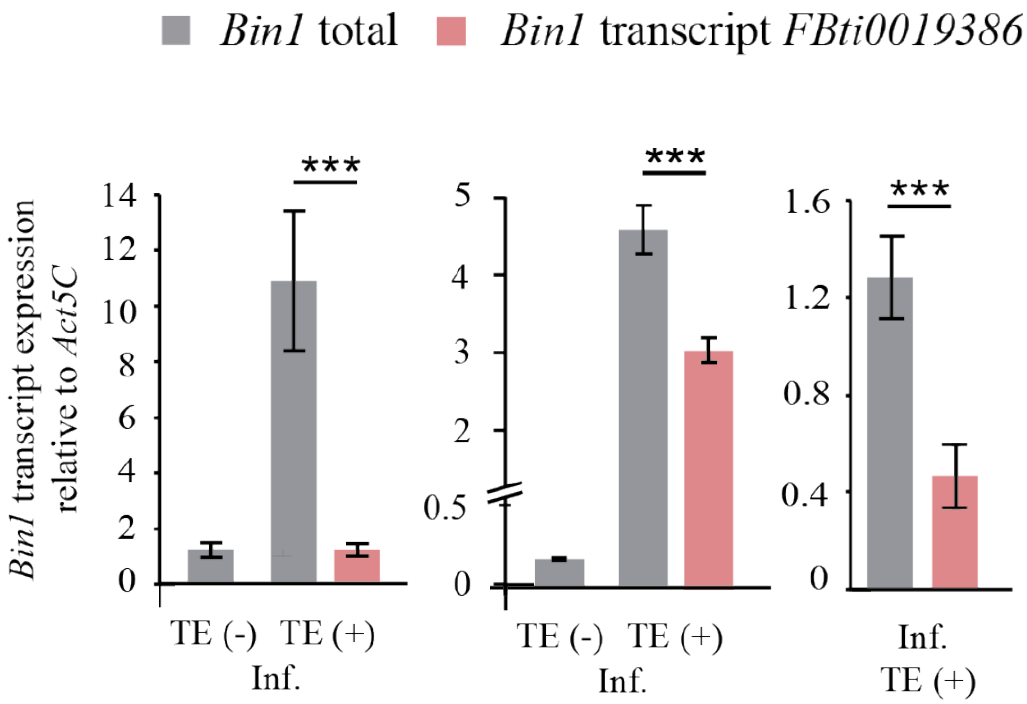


	FBti0019386 (-)	*FBti0019386* (+)
N-Inf.	*Bin1-RA*	*Bin1-RA*
Inf.	*Bin1-RA*	*Bin1-RA* *Bin1-RC* *Bin1-RD* *Bin1-RE*

Ullastres et al 2019 bioRxiv

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

FBti0019386 allele expression is background dependent



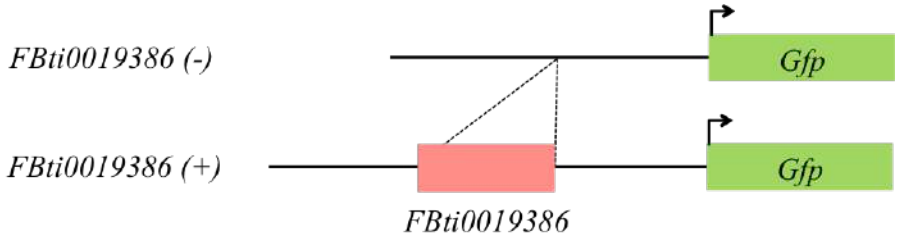
Ullastres et al 2019 bioRxiv

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

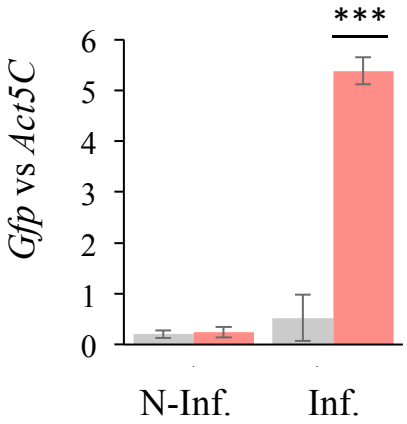
FBti0019386 drives the expression of a reporter gene

Vivien
Horváth
PhD

Enhancer assay design



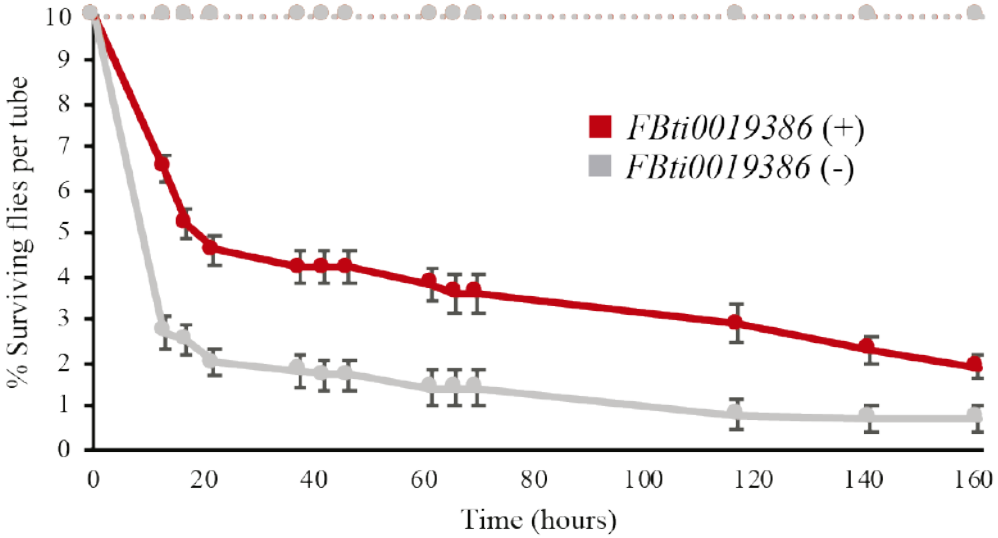
qRT-PCR *Gfp*



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

FBti0019386 is associated with increased infection tolerance

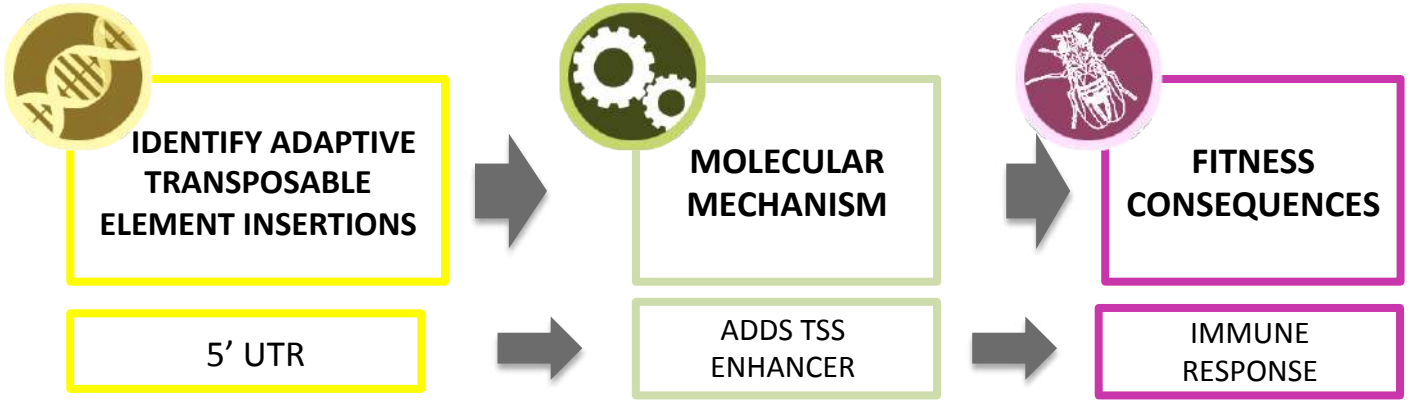
Miriam Merenciano PhD



Ullastres et al 2019 bioRxiv

González•Lab
Evolutionary and Functional Genomics

TEs are an important source of adaptive mutations



Ullastres et al.
In prep

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

TEs are an important source of adaptive mutations

IDENTIFY ADAPTIVE TRANSPOSABLE ELEMENT INSERTIONS	MOLECULAR MECHANISM	FITNESS CONSEQUENCES	
5' UTR	ADDS TSS ENHANCER	IMMUNE RESPONSE	Ullastres et al. **In prep**
5' UTR	AFFECTS EXPRESSION	IMMUNE RESPONSE	Ullastres et al. **In prep**
3' UTR	AFFECTS POLY-A TRANSCRIPT LENGTH	XENOBIOTIC RESPONSE	Mateo et al. **PLoS Genetics**. 2014
INTERGENIC	ADDS TFBS AND EPIGENETIC MARKS	OXIDATIVE STRESS RESPONSE	Guio et al. **Mol Ecol**. 2014 Guio et al **Sci Rep** 2018
INTRONIC	ADDS TFBS	HEAVY METAL RESPONSE	Le et al. **Sci Rep** 2017
5' UTR	ADDS A TSS	COLD STRESS RESPONSE	Merenciano et al. **PLoS Genetics** 2016
INTRONIC	AFFECTS EXPRESSION	COLD STRESS RESPONSE	Falqués et al. **In prep**
INTRONIC	AFFECTS EXPRESSION LEVEL	FASTER DEVELOPMENTAL TIME	Ullastres et al. **Mol Biol Evol**. 2015

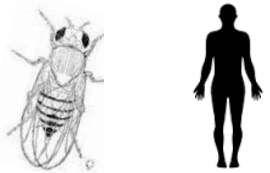
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

What is the impact of TEs in the regulation of stress-response genes?

Stress-response pathways are highly conserved across organisms

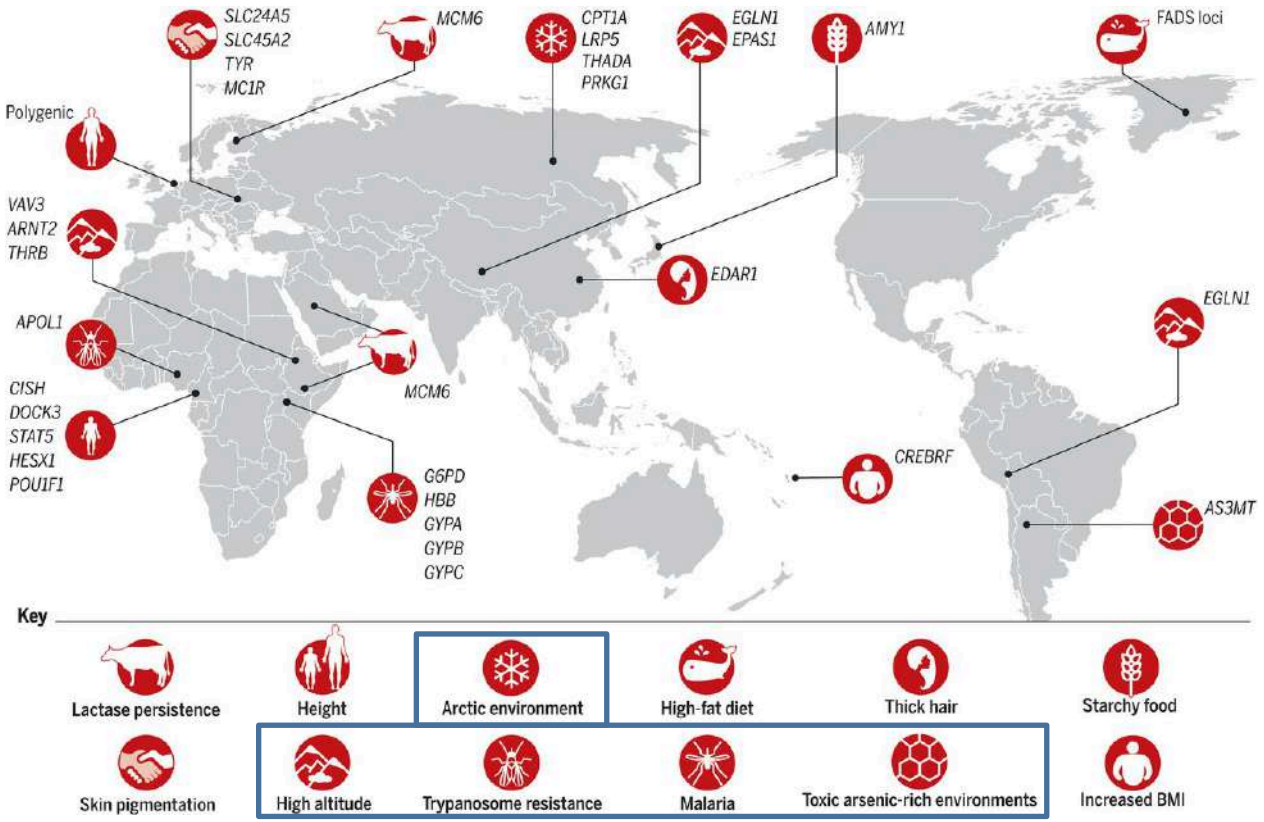
Oxidative stress	Cnc/Nrf2
Xenobiotic stress	Cnc/Nrf2
Hypoxia stress	HIF1
Heavy-metal stress	MTF1
Heat shock stress	HSF
Immune stress	caudal, Deaf1, nub, XBP1, dorsal



Villanueva-Cañas, Horvath et al 2019 Nuc Acids Res

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

Adaptation in humans is also related to several stress-responses



Fan et al 2016 Science

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

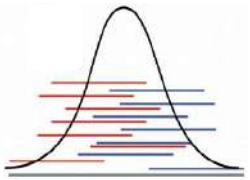
PREDICTIONS

TFBMs

PWMs

TFFMs

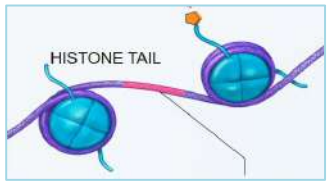
TFBSs



ADDITIONAL EVIDENCE

Enhancer features

Open chromatin/ histone marks

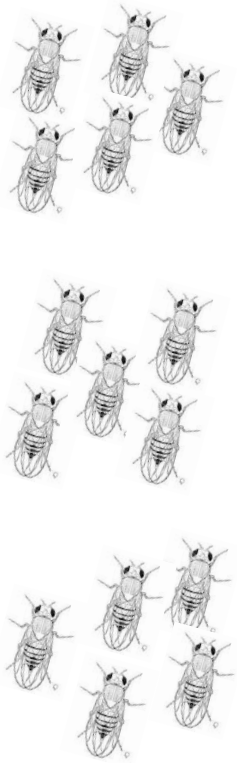


CBP co-binding

Genomic location

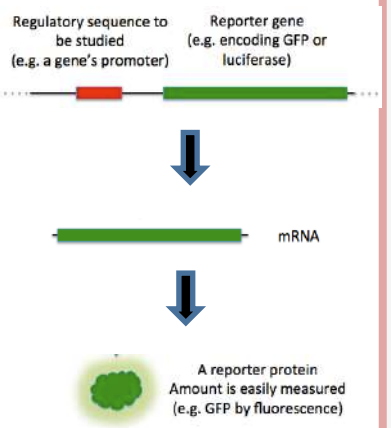
Regulatory region/ stress-related genes

Population frequencies



VALIDATION

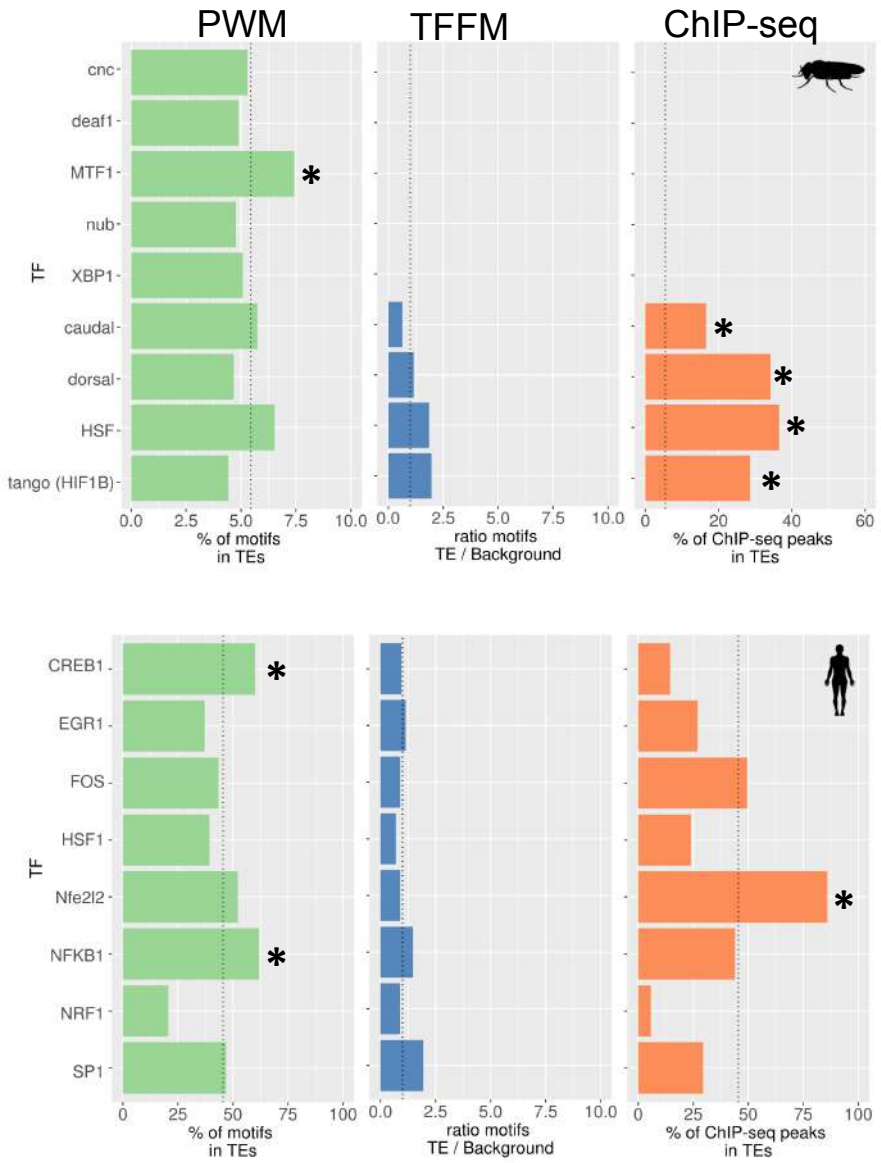
In vivo enhancer assays



Villanueva-Cañas, Horvath et al 2019 Nuc Acids Res

GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

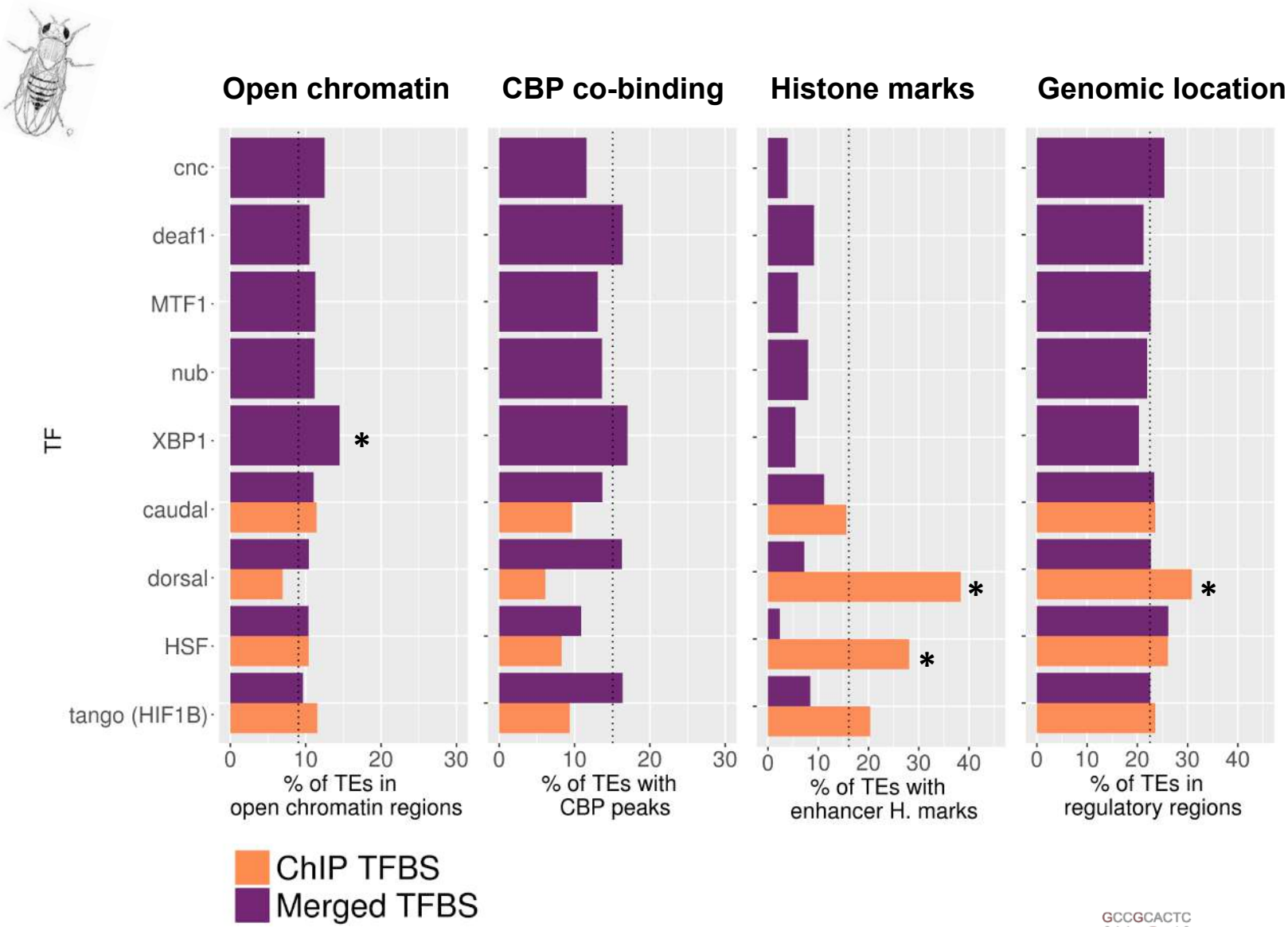
TEs are a considerable source of stress-response TFBS



- Proportion of TFBS in TEs is TF-specific
  - 17% to 37% in Drosophila
  - 6% to 86% in humans

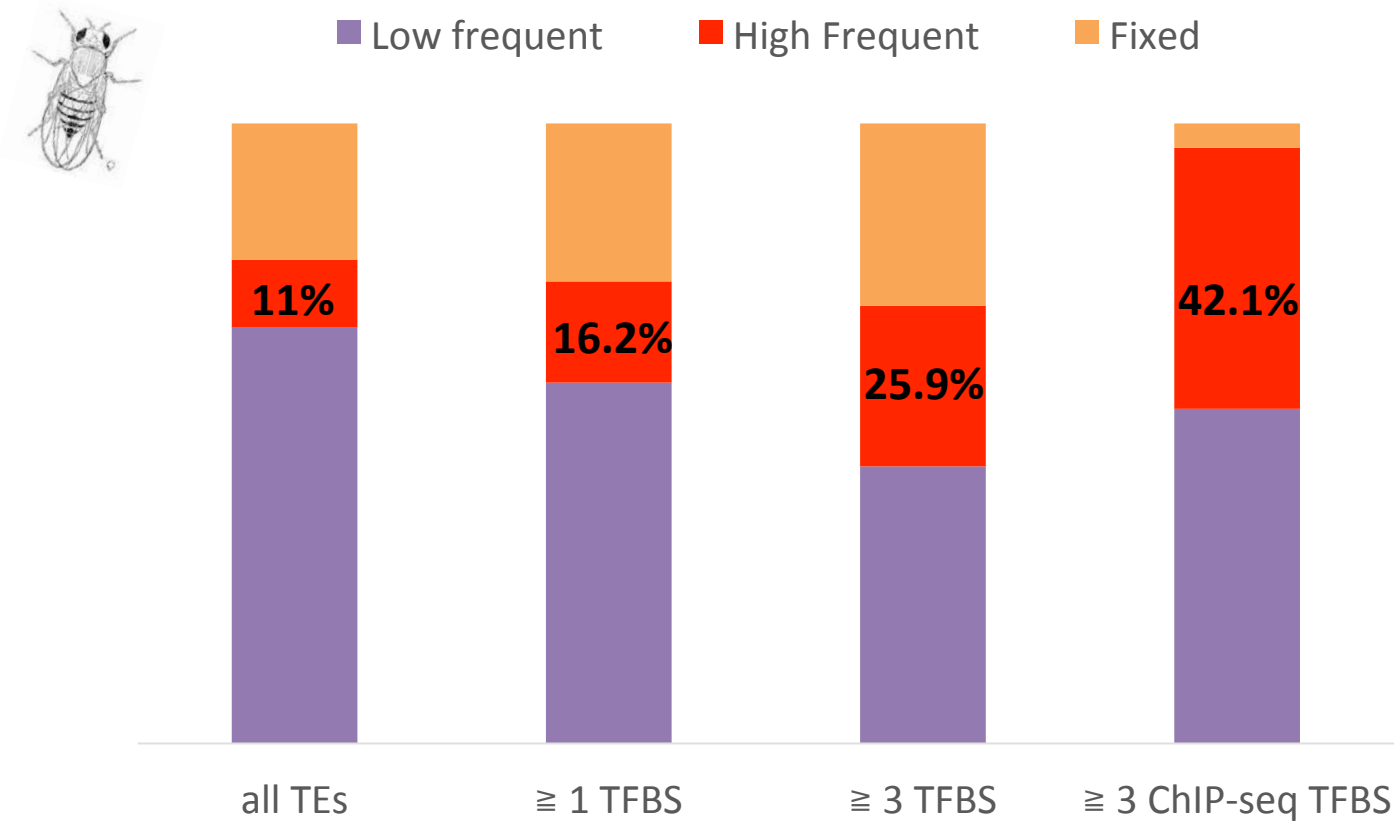
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

TEs containing TFBS are not globally enriched for enhancer features



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

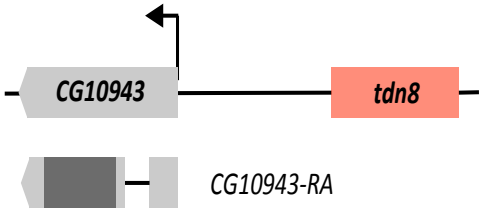
TEs with ≥ 3 TFBS are present at high population frequencies



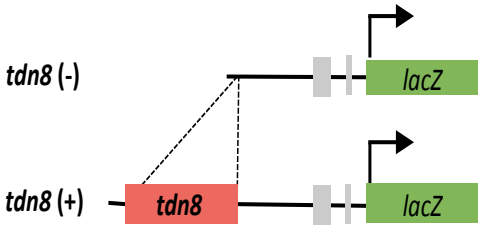
GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Functional validation of candidate TEs

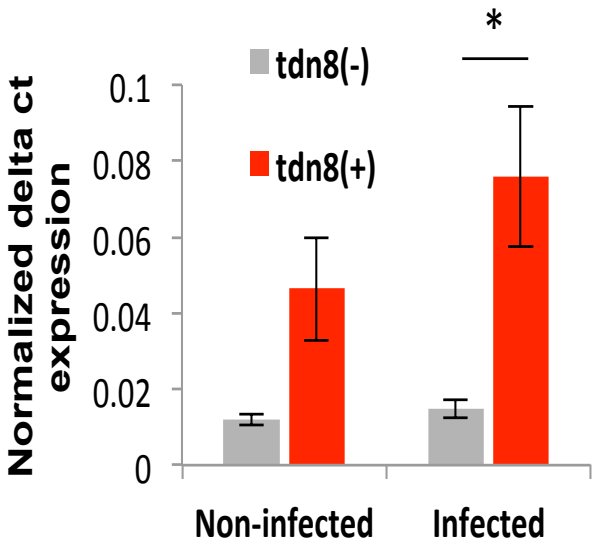
A. *CG10943* transcript



B. Enhancer assay design



C. qRT-PCR *lacZ* transgenic flies gut



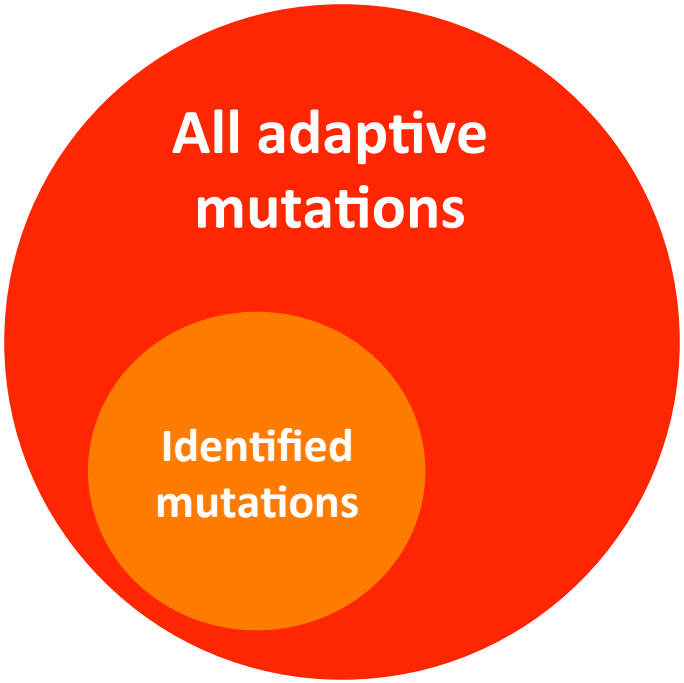
GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

Functional validation of candidate TEs

We validated six out of 12 TEs tested

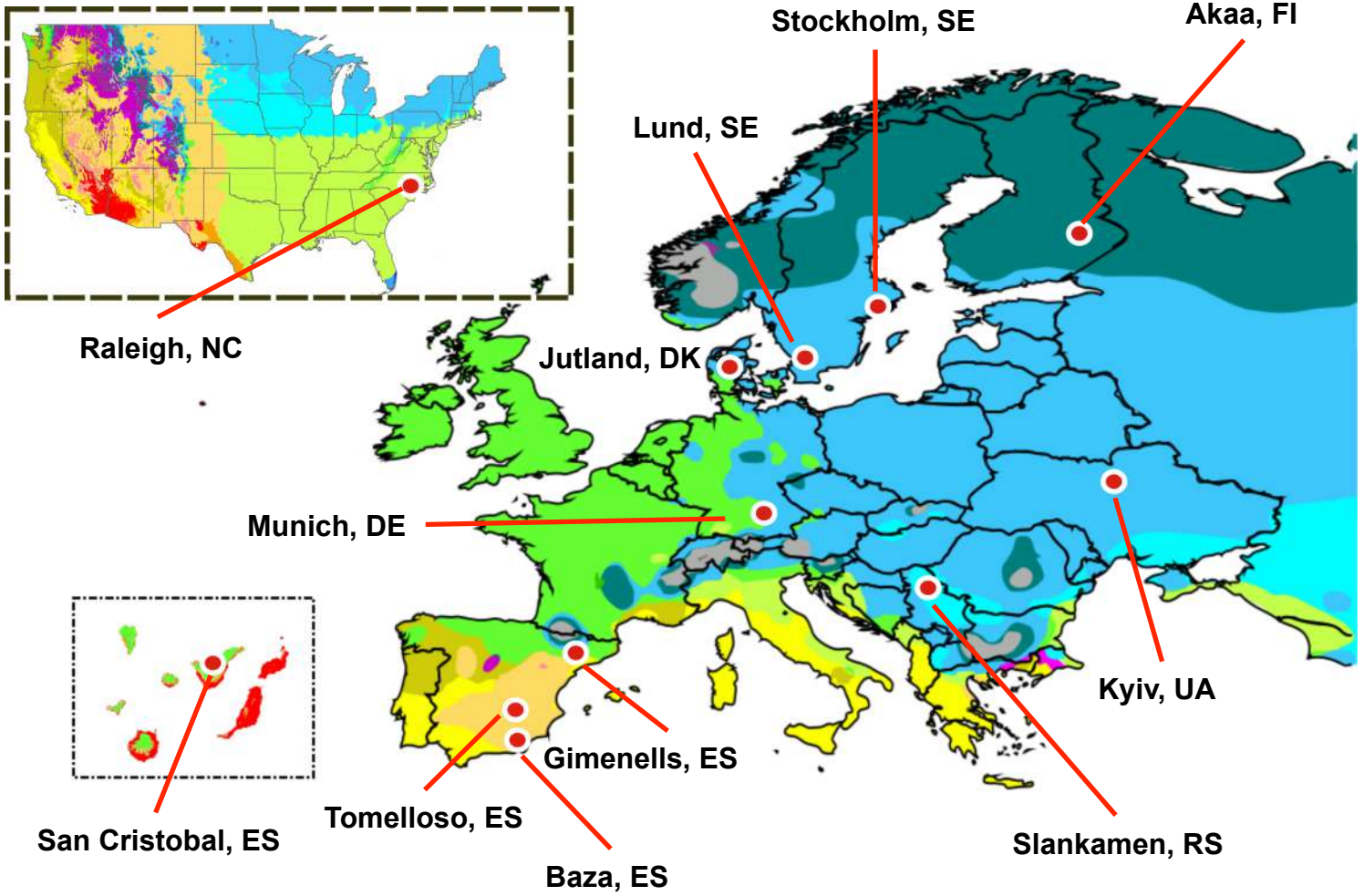
- All six TEs affected expression of the reporter gene under stress conditions, suggesting that these TEs are indeed involved in stress gene regulation
- Five of the six TEs add TFBS that were already present in the promoter region of the nearby gene
- All six TEs were either LTR or LINE from different families: *roo*, *jockey*, *invader4*, *Rt1b*, *BS*, and *transpac* while most of the non-validated TEs were TIRs
- Non-validated TEs:
  Some might required genomic context that is not present in the vector
  May serve as a landing pad for TFs to scan DNA
  Buffer of extra binding sites to trap TF

Villanueva-Cañas, Horvath et al 2019 Nuc Acids Res



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

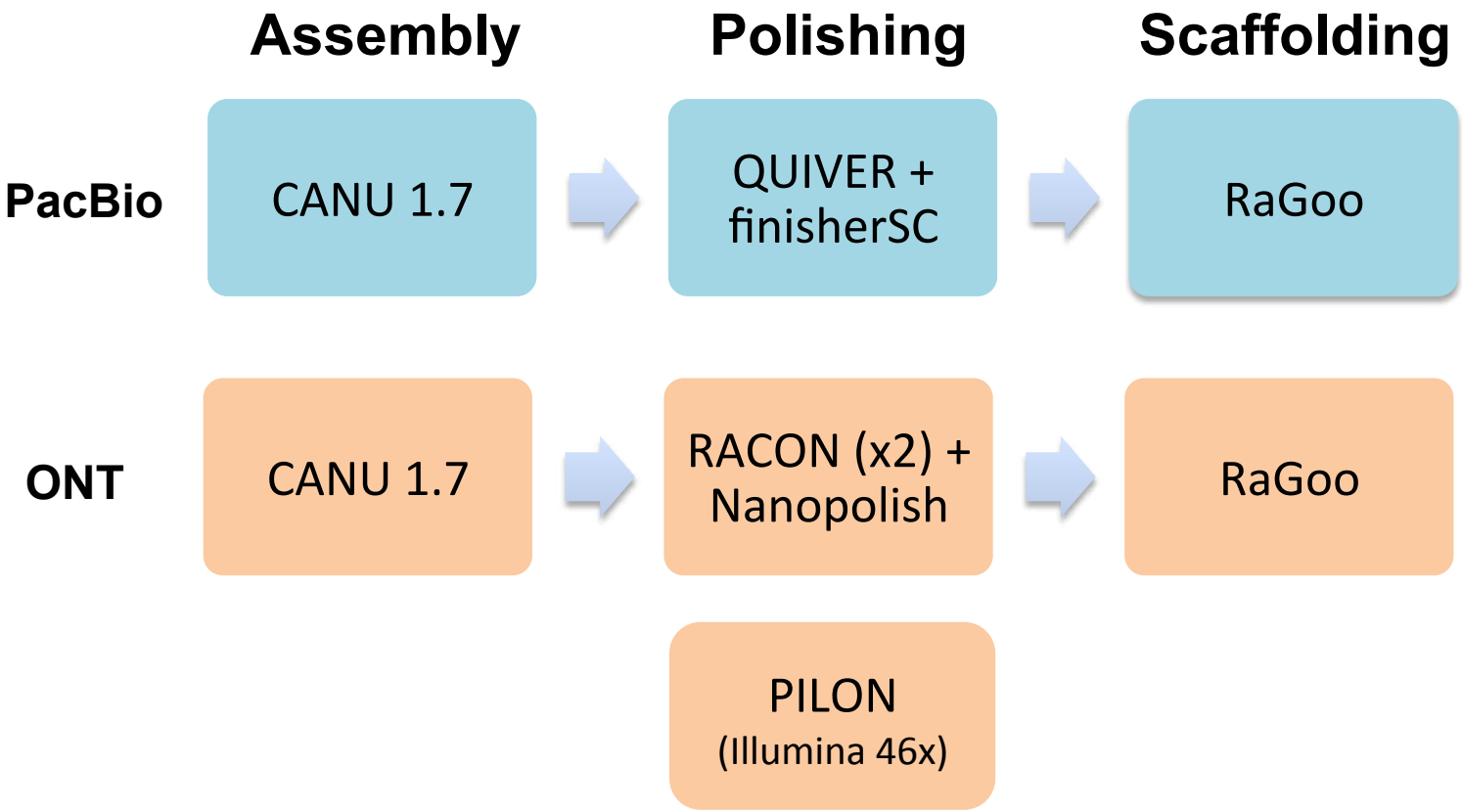
13 new *Dmel* reference genomes



@rechgab

González•Lab Evolutionary and Functional Genomics

13 new *Dmel* reference genomes



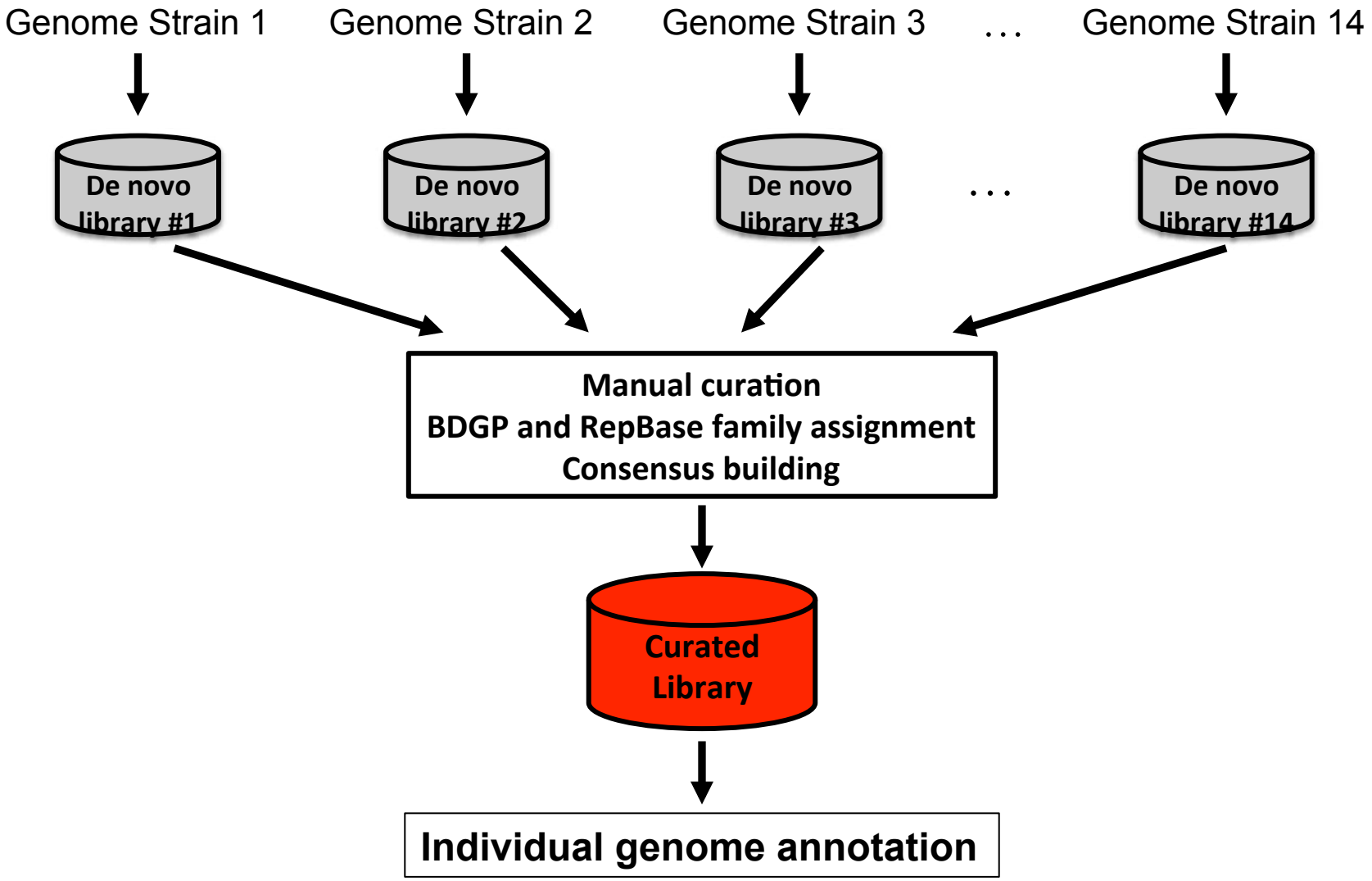
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

13 new *Dmel* reference genomes

Technology	Genome	Coverage	Genome size	N50	BUSCO
PacBio	MUN	117X	142,0	7.8Mb	98.5%
	TOM	86X	**150,4**	3.2Mb	98.1%
ONT	FIN	107X	142,7	**18.9Mb**	**98.6%**
	LUN	96X	138,1	2.0Mb	98.4%
	B22	90X	142,4	3.0Mb	97.3%
	COR	82X	**138,1**	7.8Mb	98.0%
	JUT	81X	138,4	4.0Mb	98.4%
	SLA	68X	143,7	0.8Mb	97.3%
	GIM	65X	142,3	6.8Mb	**98.6%**
	R177	61X	141,9	14.5Mb	97.4%
	TEN	60X	140,5	1.1Mb	97.4%
	KIE	51X	143,8	3.8Mb	98.4%
	R375	47X	141,2	13.5Mb	96.1%

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

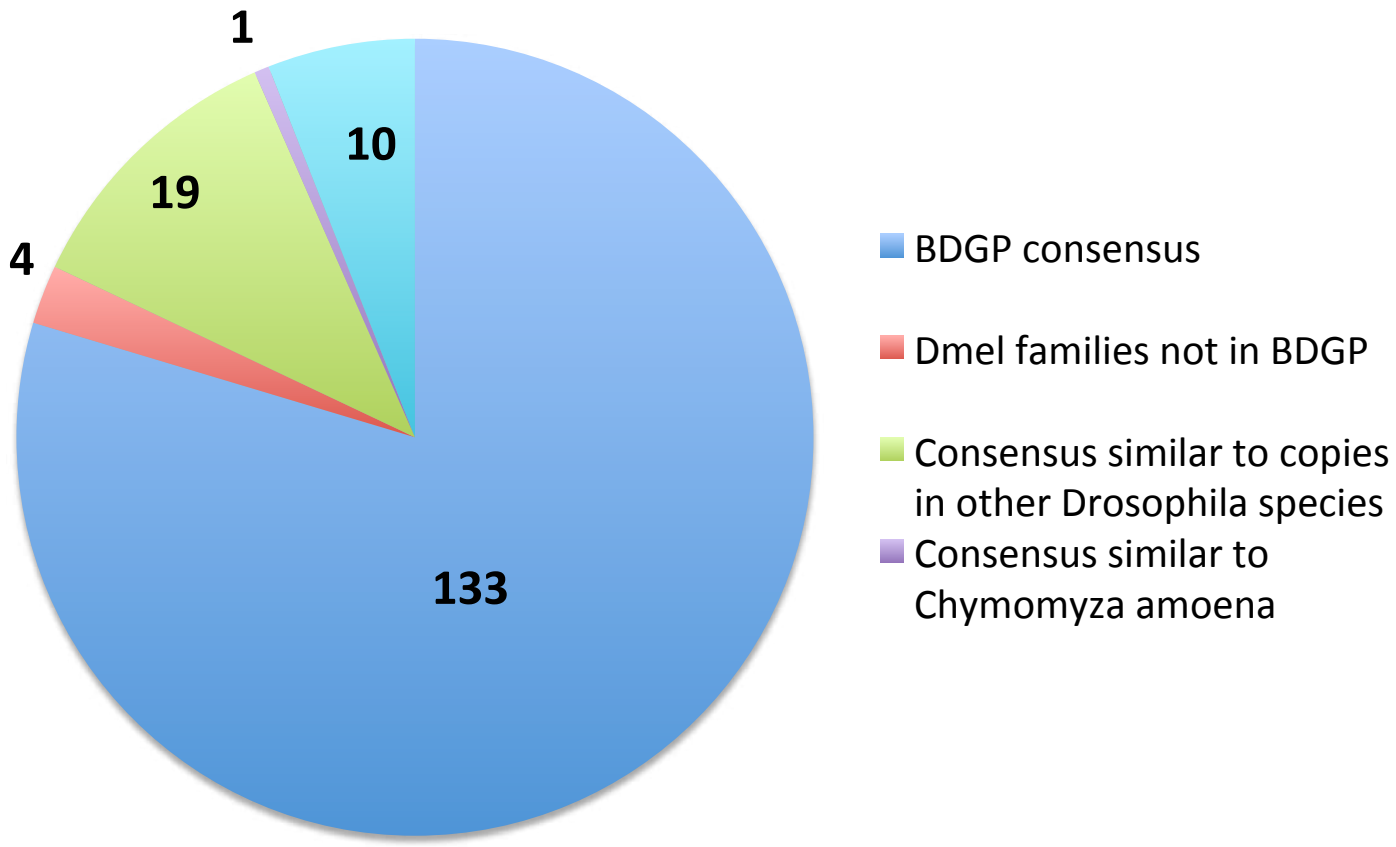
De novo annotation of TEs with REPET



Flutre et al 2011 PLoS One, Quesneville et al 2005 PLoS Comp Bio

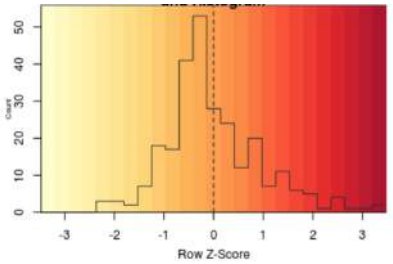
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

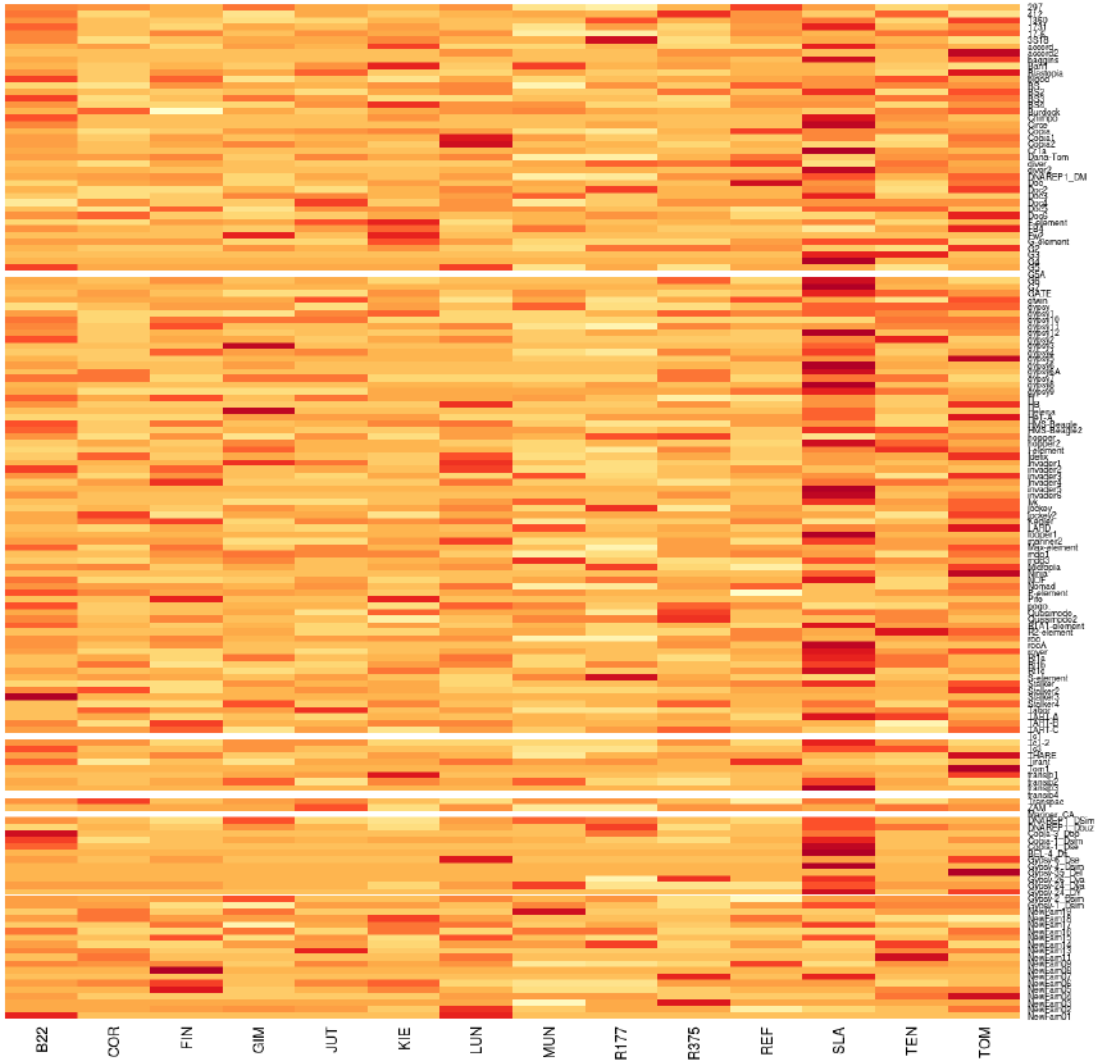
TE annotation in the new 13 reference genomes



GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

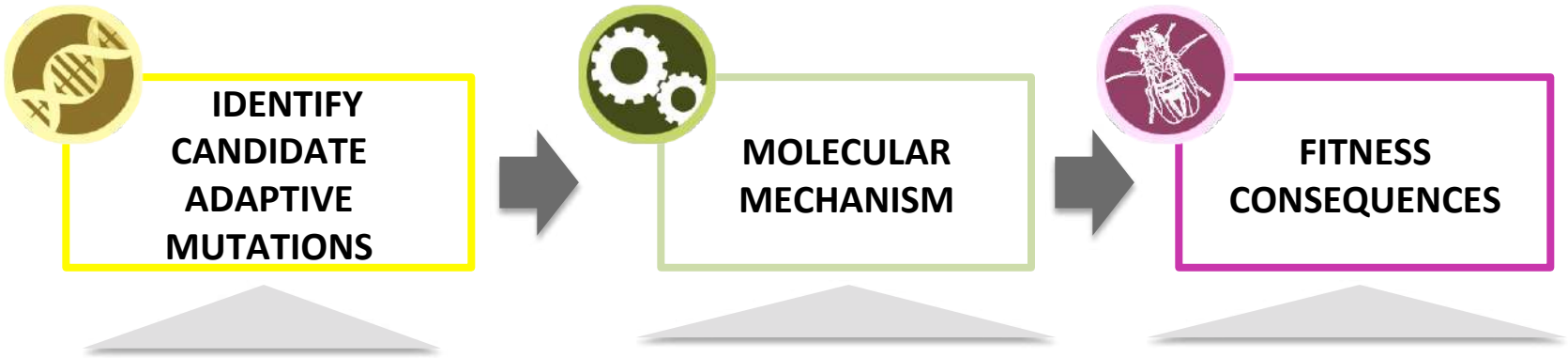
Genomes differ in family abundance and total copy number





GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics

SUMMARY



- **300 polymorphic TEs likely to be enriched for adaptive insertions**

- **Several case studies show the diversity of molecular mechanisms**
- **TEs adds stress-related TFBS (Nucleic Acids Res, 2019)**

- **Stress, development and behavior**
- TEs affect immune-response genes (bioRxiv)

Barret and Hoekstra 2008; Storz and Wheat 2010

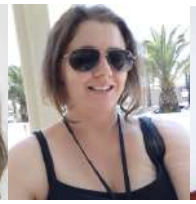
GCCGCACTC CAA AC AGC González•Lab CAA GAAACC Evolutionary and Functional Genomics

Acknowledgments

Gabriel Rech
Postdoc

Sara Guirao-Rico
Postdoc

Judit Salces
Postdoc



Llew Green
Postdoc

Carlos Vargas
Postdoc

Laura Aguilera
Lab tech

Miriam Merenciano
PhD student

Maria Bogaerts
PhD student

Vivien Horvath
PhD student

Hadi Quesneville
URGI, INRA, Versailles FRANCE

Roberto Torres
Fruitfly, adaptive science outreach

FUNDING

European Research Council
Established by the European Commission

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca

eseb
European Society for Evolutionary Biology

CSIC
Consejo Superior de Investigaciones Científicas

cost
European Cooperation in Science and Technology

RESEARCH NETWORKS

AdaptNet
GENOMICS OF ADAPTATION NETWORK

DROSEU
EUROPEAN DROSOPHILA POPULATION GENOMICS CONSORTIUM

EuroScitizen
Building on scientific literacy in Evolution

josefa.gonzalez@ibe.upf-csic.es
www.gonzalezlab.eu

GCCGCACTC
CAA AC
AGC González•Lab
CAA
GAAACC Evolutionary and Functional Genomics