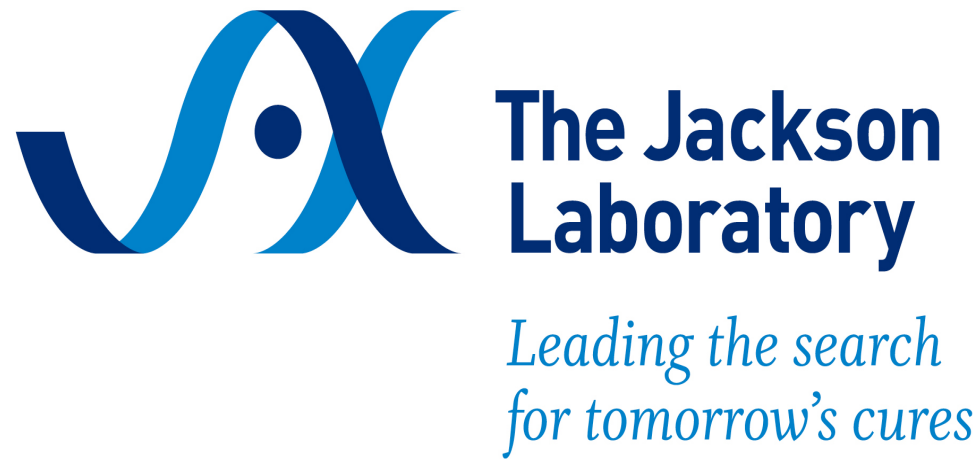


# Genetic diversity and footprints of selection in wild house mice

Raman Akinyanju Lawal\* and Beth Dumont  
The Jackson Laboratory, Bar Harbor, Maine 04609 USA  
\*lawalakinyanju@yahoo.com



## Abstract

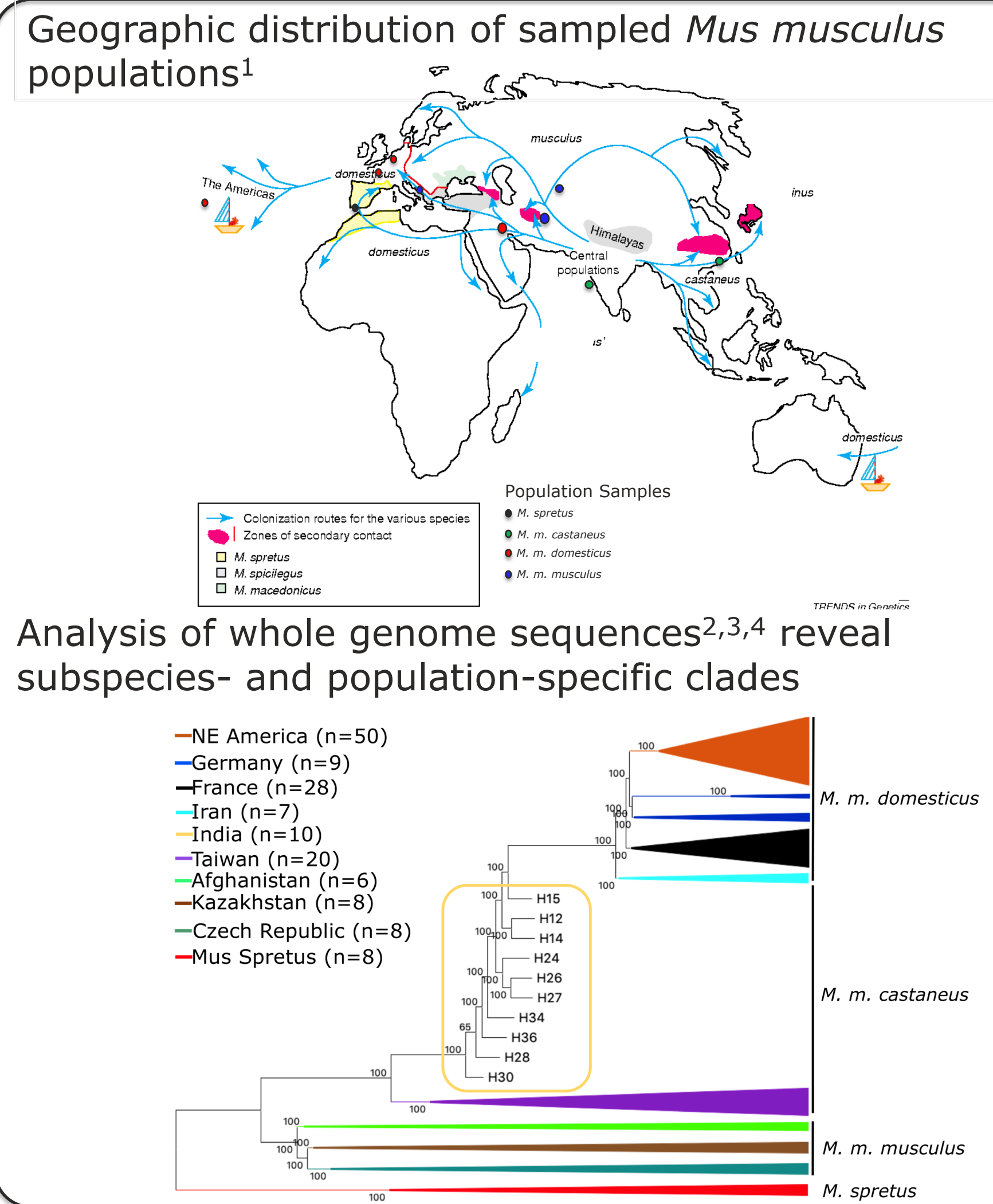
A major goal of evolutionary genetics is to decipher the mechanisms of adaptation to diverse environments. Here, we used 154 whole-genome sequences from wild-caught house mice to perform genome-wide scans for selection. These samples span three core house mice subspecies (*Mus musculus domesticus*, *M. m. castaneus*, *M. m. musculus*) and the outgroup taxon, *M. spretus*, and include organisms inhabiting diverse ecological environments. We employ multiple established population genetic methods and conservative statistical thresholds to spotlight 71 loci in *M. m. domesticus*, 52 in *M. m. castaneus*, 48 in *M. m. musculus*, and 50 in *M. spretus* that are evolving non-neutrally. Included among these loci are well-established targets of positive selection among mammals, including olfactory receptors, genes involved in reproduction, and *Epas1*, which has been previously implicated in physiological adaptation to hypoxic environments. Several loci are also evolving via different evolutionary selection regime in different subspecies, including *Pkd2*, *Hbb-bh2*, *Prl2c3*, *Cntnap2*, *Lrrc25*, and *Susd6*. Our analyses also underscore a key role for balancing selection in the maintenance of genetic diversity at several genes, including *Cwc22*, *Zswim2* and *Fam171b*. Taken together, our findings comprise a catalog of putative signals of positive and balancing selection in a powerful biomedical model system poised for facile discovery of the genetic and underlying physiological mechanisms of adaptation in the wild.

## Rationale



- The house mouse, *Mus musculus*, is the premiere mammalian model system for biomedical research
- In the wild, house mice inhabit remarkably diverse environments, including deserts, high altitude regions, and tropical climates
- We **hypothesize** that the genomes of wild house mice exhibit strong signals of adaptation to these variable environments

## Data



## Approach

Positive selection leaves characteristic footprints in patterns of DNA diversity and divergence. To identify genes evolving via positive selection in wild mouse populations, we computed multiple **diversity** and **divergence** statistics in 20kb windows (10kb step) across wild mouse genomes.

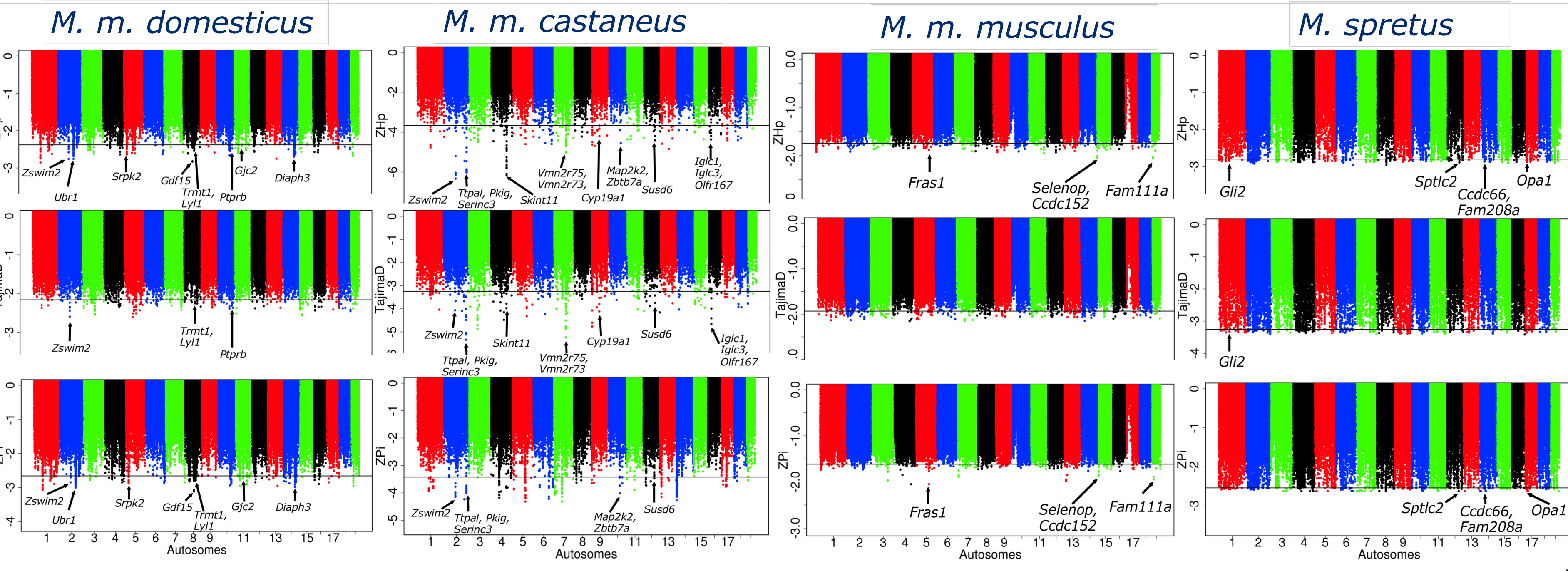
	Statistic	Expectation Under Positive Selection
Diversity	Tajima's D	Negative
	H <sub>p</sub>	Low
Divergence	Pi	Low
	F <sub>ST</sub>	High
	D <sub>XY</sub>	High

We focus on genes residing in windows in the extreme 0.1% tail of the distribution of each test statistic.

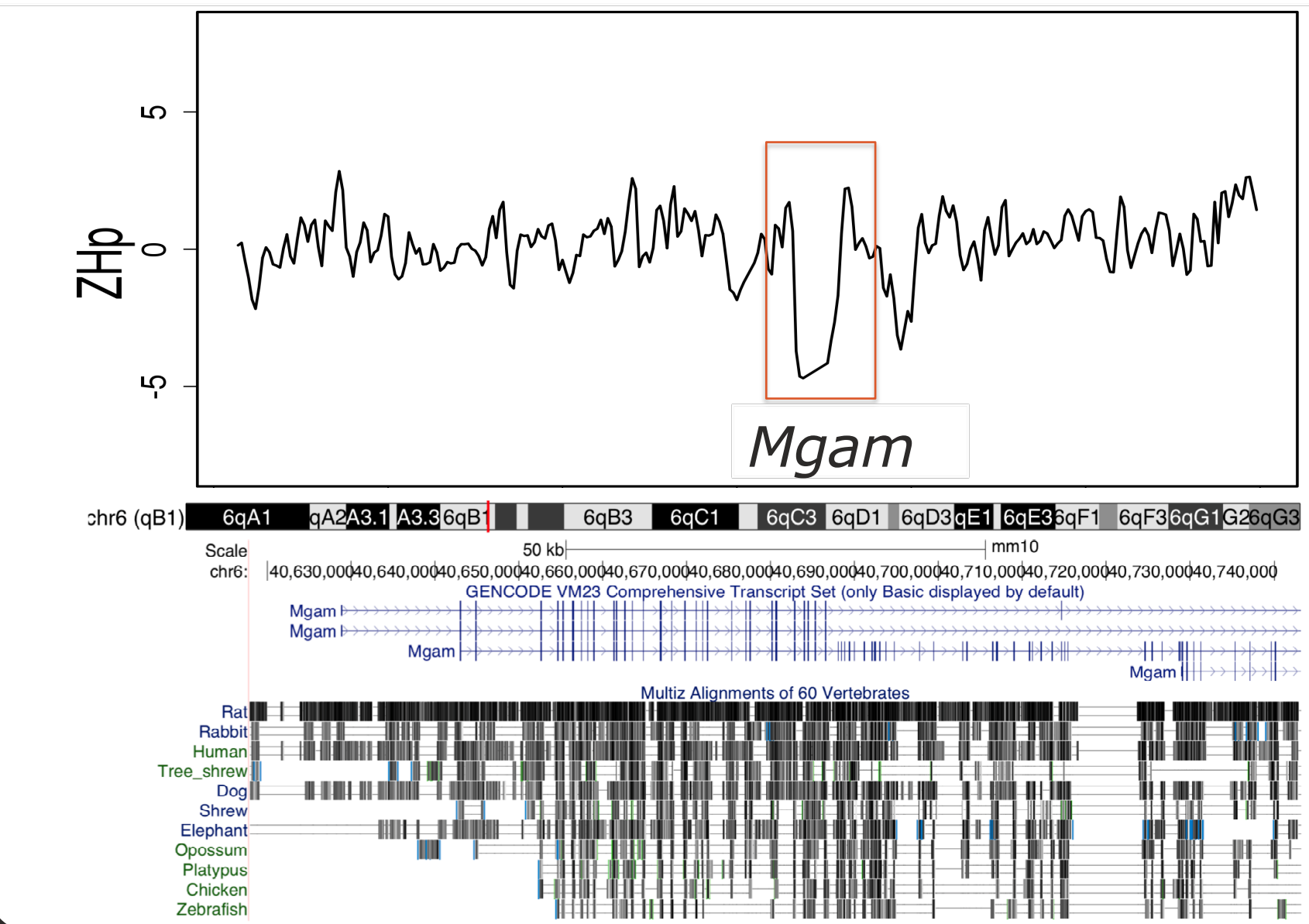
**A unique aspect of our study is the ability to simultaneously identify targets of positive selection on multiple evolutionary scales: within populations in each subspecies, within each subspecies, and between subspecies.**

## Results

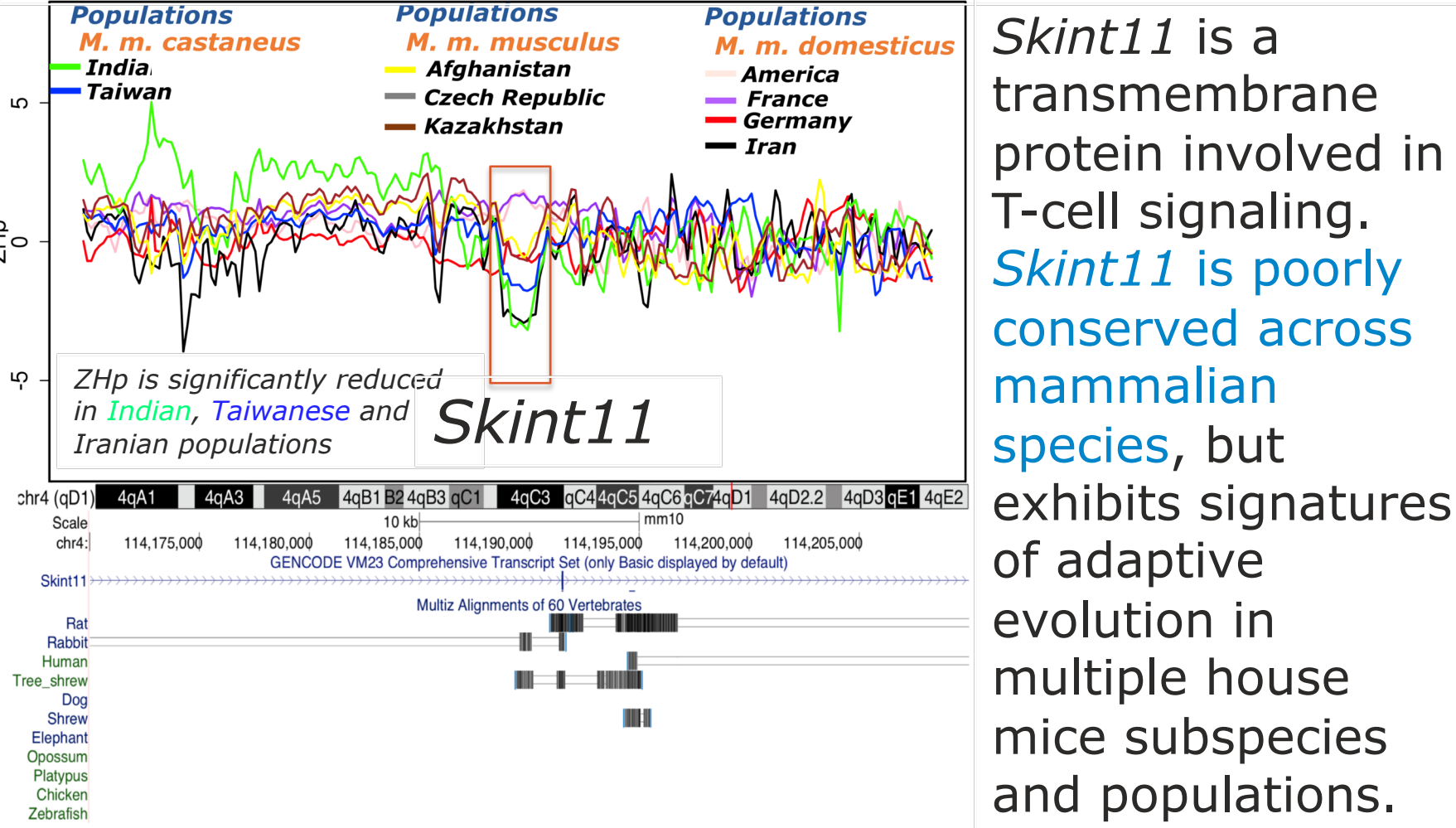
**Widespread signals of positive selection in wild mouse genomes.** Normalized values for three diversity statistics are shown for aggregate samples within a subspecies. Horizontal lines denote empirical 0.1% cutoffs. Candidate genes within outlier regions are annotated. Due to unequal numbers of males and females across populations, the X and Y chromosomes were excluded from analyses.



**Population specific signal of recent positive selection** at *Mgam* in the Iranian population of *M. m. domesticus*. *Mgam* is a glucoamylase involved in starch digestion and is well-conserved across mammals. **We speculate that mutations in this gene** may have facilitated wild mouse adaptation as commensals with human societies subsiding on grain-rich diets.



## Subspecies and population-specific signals of positive selection



*Skint11* is a transmembrane protein involved in T-cell signaling. *Skint11* is poorly conserved across mammalian species, but exhibits signatures of adaptive evolution in multiple house mice subspecies and populations.

Other notable examples of loci with population- or subspecies-specific signals of selection:

- Epas1*** – a recurrent target of adaptation to high altitudes in mammals
- Kiss1*** – a metastasis suppressor gene that is also implicated in gonadotrophin signaling
- Cry1*** – a transcriptional regulator of the circadian clock
- Opa1*** – a dynamin-like GTPase that localizes to the inner mitochondrial membrane

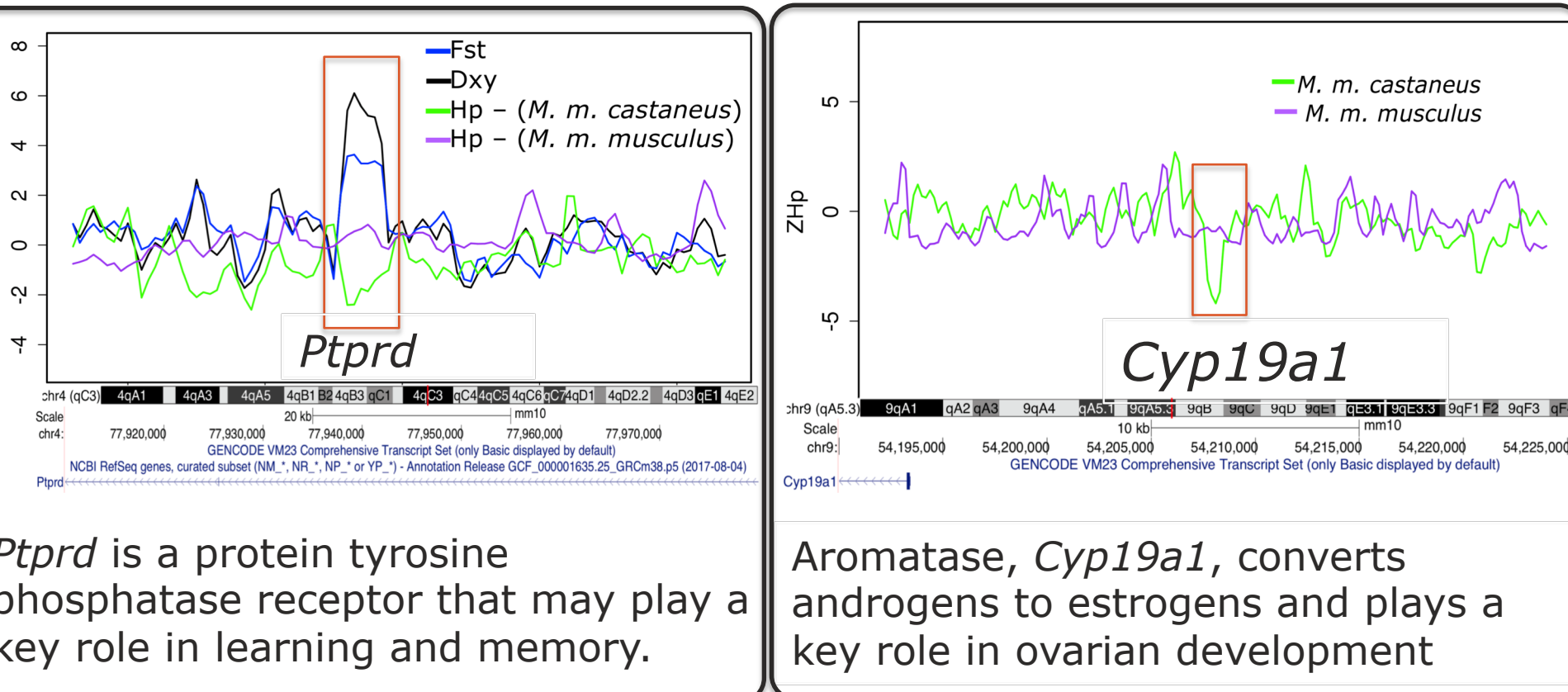
## Funding

This project is supported by Startup funds and NIH grants R00 GM110332 and R35 GM133415 to BLD.

## References

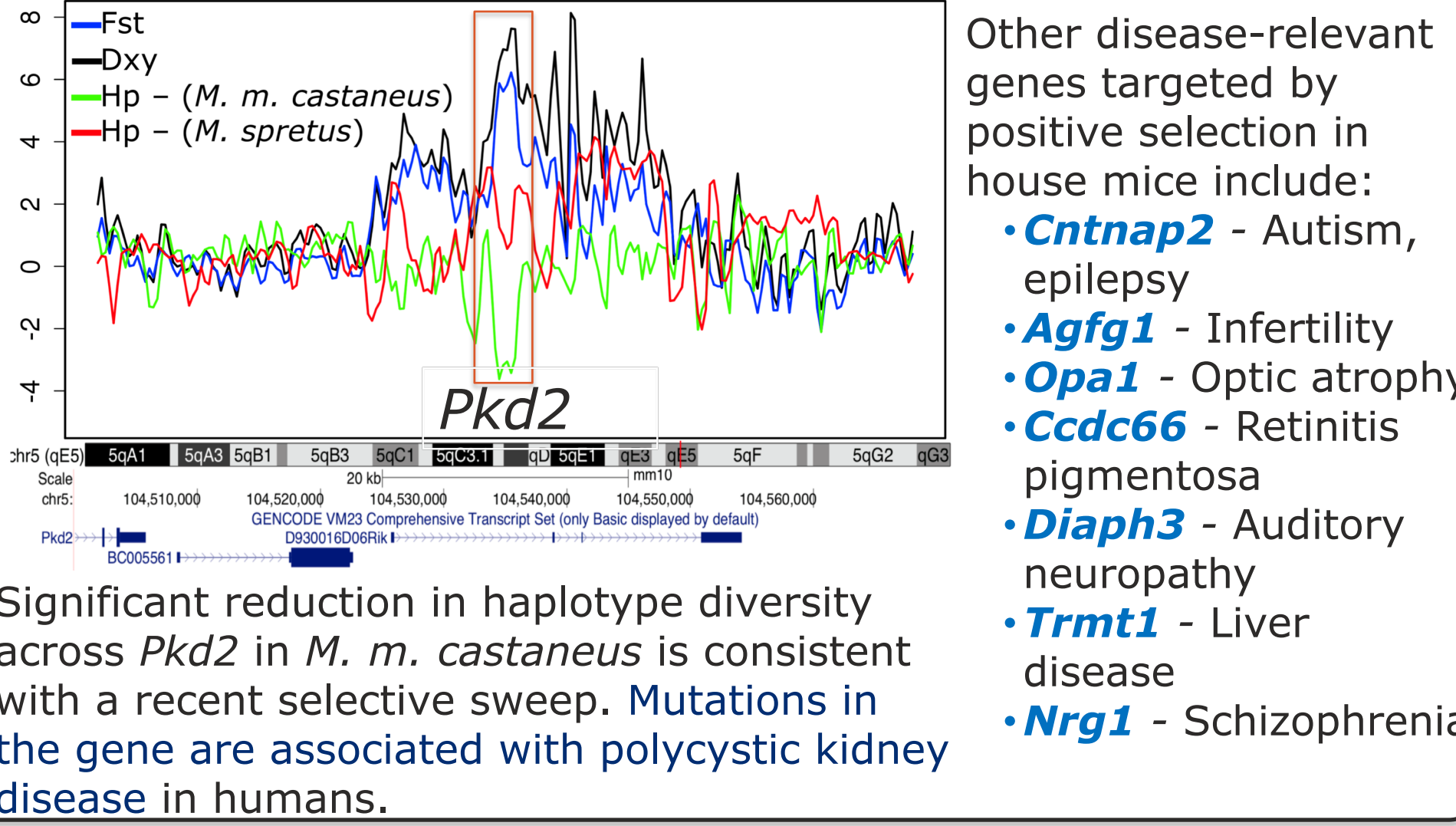
- Jean LG and François RB, 2003. Trends Genet. 19:24-31
- Harr et al. 2016. Sci Data. Article no.: 160075
- Phifer-Rixey et al. 2018. PLoS Genet. Vol. 2
- Davies 2015. (Thesis). Oxford University, UK

## Genes evolving via adaptive evolution between subspecies include loci with key functions in neuronal signaling and spermatogenesis



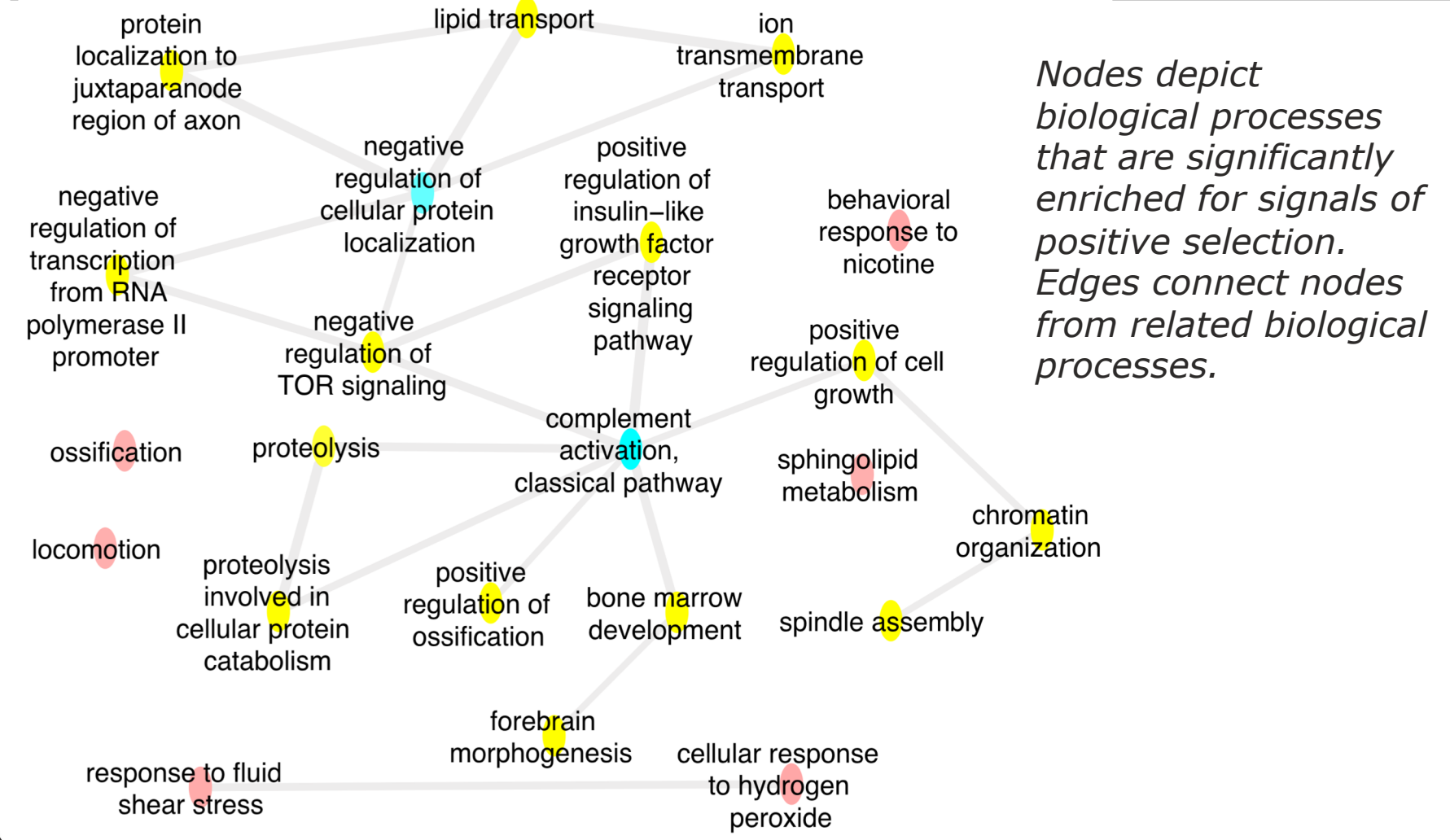
**Positive selection may drive the emergence of behavioral and reproductive barriers between subspecies.**

## Many disease-relevant genes display signals of adaptive evolution in wild house mice



- Other disease-relevant genes targeted by positive selection in house mice include:
- Cntnap2*** - Autism, epilepsy
  - Aggf1*** - Infertility
  - Opa1*** - Optic atrophy
  - Ccdc66*** - Retinitis pigmentosa
  - Diaph3*** - Auditory neuropathy
  - Trmt1*** - Liver disease
  - Nrg1*** - Schizophrenia

## Genes evolving via positive selection in wild mouse populations are implicated in diverse biological processes and pathways



## Conclusions & Future Directions

- We have developed a comprehensive catalog of targets of recent adaptive evolution in wild house mice
- Identified targets play roles in diverse biological processes, including processes relevant to speciation and human disease.
- The targets of positive selection within subspecies are largely distinct from genes evolving via adaptive evolution between subspecies.
- To understand whether observed signals of selection derive from new mutations or standing genetic variation, we are currently sequencing a population of *M. m. bactrianus* mice from the presumed *M. musculus* ancestral zone.