The landscape of fitness effects of putatively functional noncoding mutations in humans

狄晨露, postdoc
Kirk Lohmueller lab
University of California, Los Angeles

浙江大学
生命演化研究中心

Sep.26st, 2024, ZJU

chenludi6@gmail.com
cdi@ucla.edu

The journey: Natural selection and human diseases

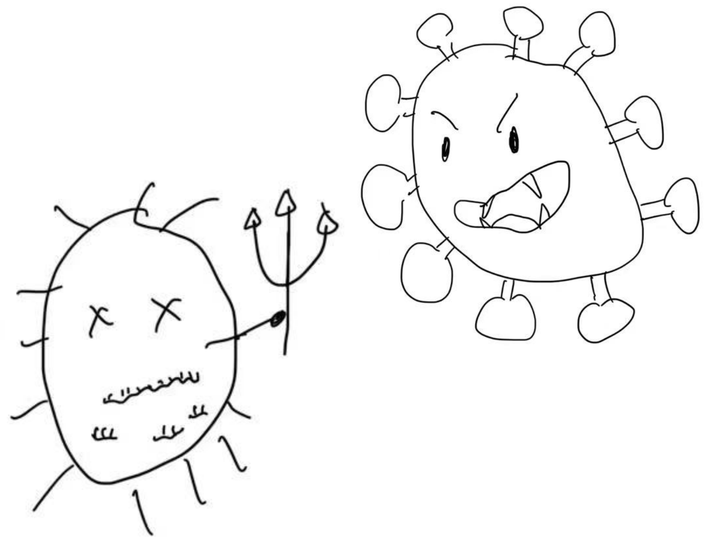
赵烨

李正和

The journey: Natural selection and human diseases

赵烨

李正和

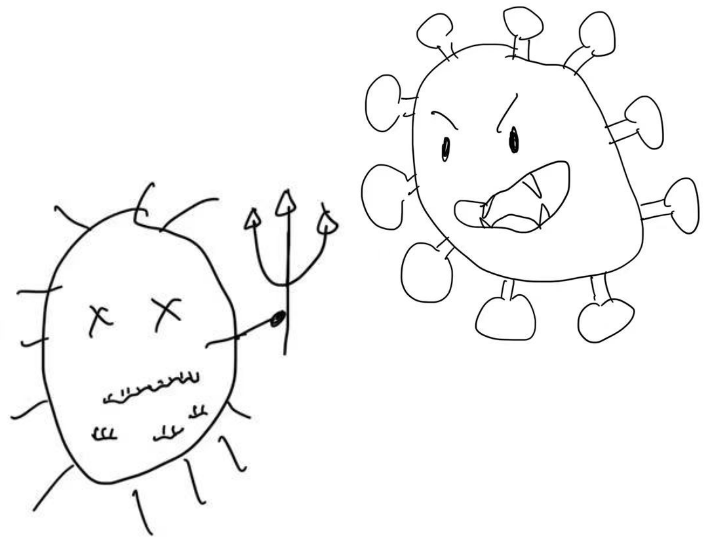


The journey: Natural selection and human diseases

赵烨

李正和

“Nothing in biology makes sense except in the light of Evolution.”



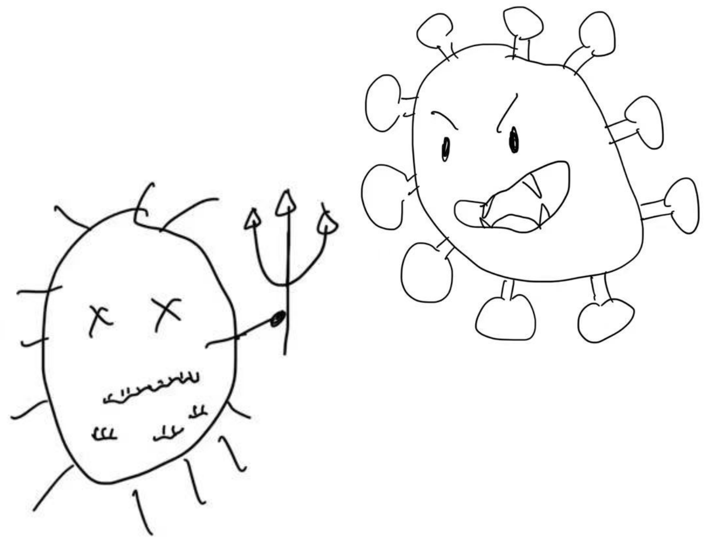
The journey: Natural selection and human diseases

THE UNIVERSITY
OF ARIZONA


赵烨


李正和

"Nothing in biology makes sense except in the light of Evolution."



The journey: Natural selection and human diseases

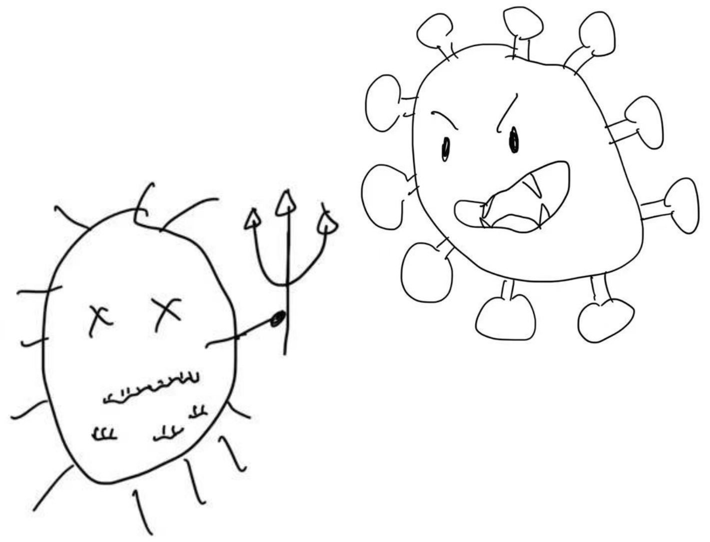
浙江大学
1897
ZHEJIANG UNIVERSITY

THE UNIVERSITY
OF ARIZONA


赵烨


李正和

*“Nothing in biology makes sense
except in the light of Evolution.”*



The journey: Natural selection and human diseases

浙江大学
1897
ZHEJIANG UNIVERSITY

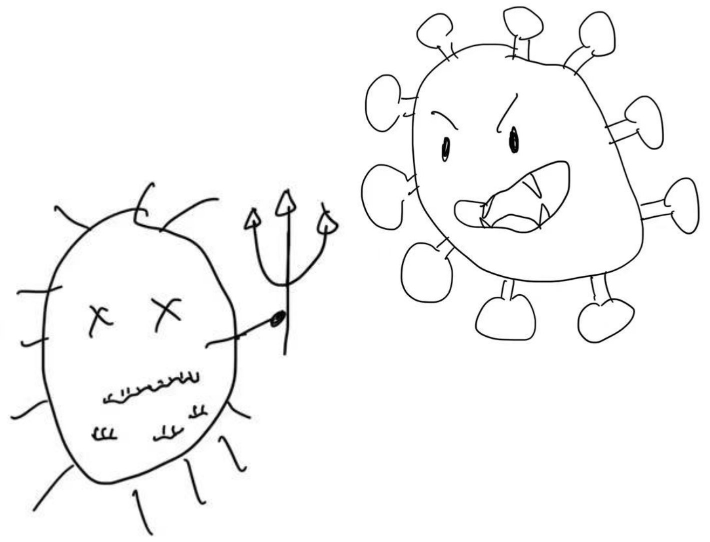
THE UNIVERSITY
OF ARIZONA


David Enard


赵烨


李正和

*“Nothing in biology makes sense
except in the light of Evolution.”*



The journey: Natural selection and human diseases

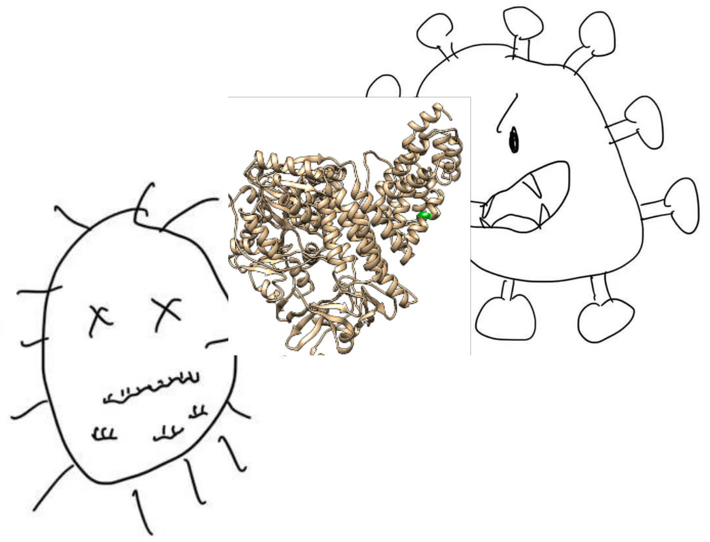
THE UNIVERSITY
OF ARIZONA

David Enard

赵烨

李正和

“Nothing in biology makes sense except in the light of Evolution.”



The journey: Natural selection and human diseases

David Enard

赵烨

李正和

“Nothing in biology makes sense except in the light of Evolution.”

The journey: Natural selection and human diseases

David Enard

Kirk Lohmueller

赵烨

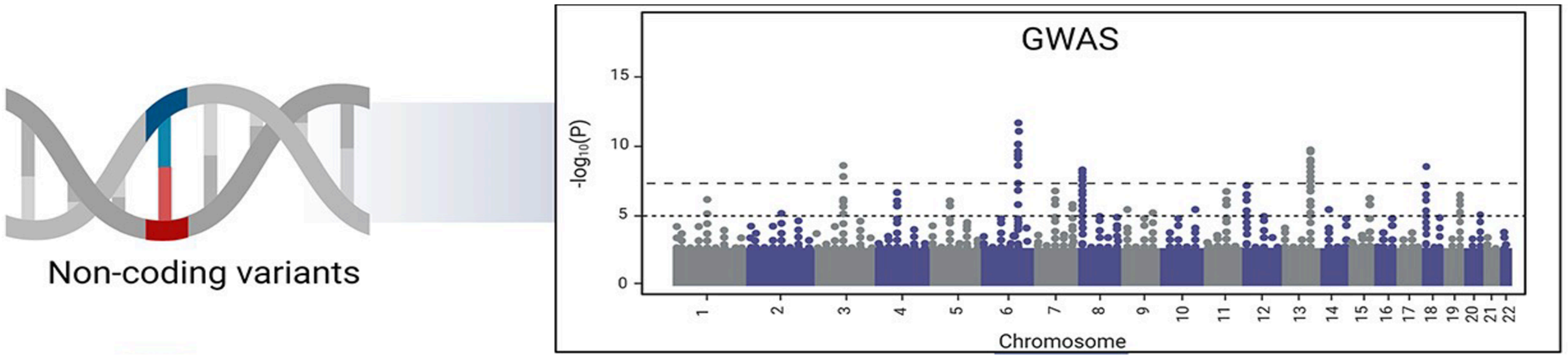
李正和

“Nothing in biology makes sense except in the light of Evolution.”

- What about complex diseases?

Disease associated non-coding variants

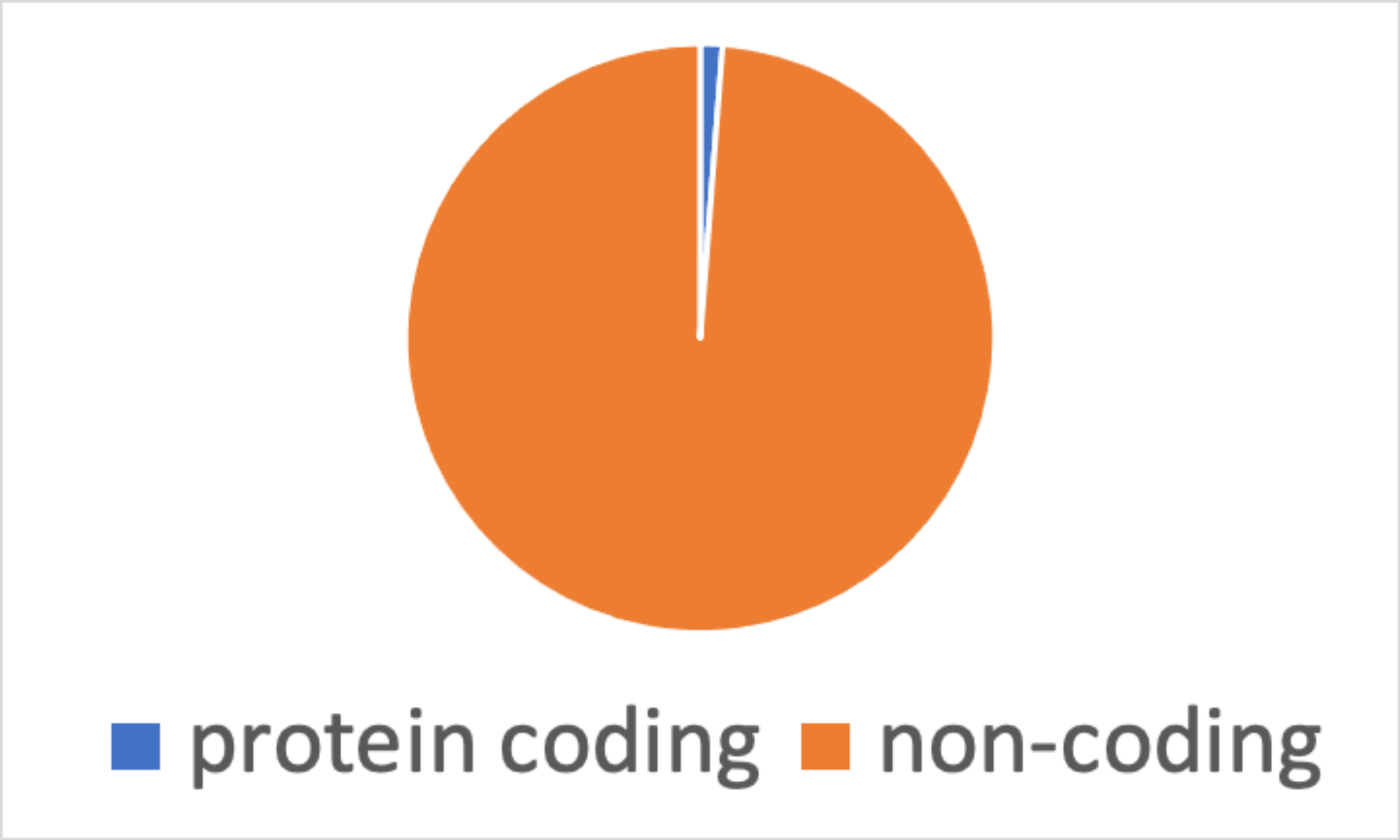
Most variants, including most of the functional ones that contribute to disease risk, are not in protein-coding regions.



Schipper and Posthuma, 2022

Non-coding DNA

Non-coding DNA: DNA does not code for protein



The ENCODE Project Consortium, 2012

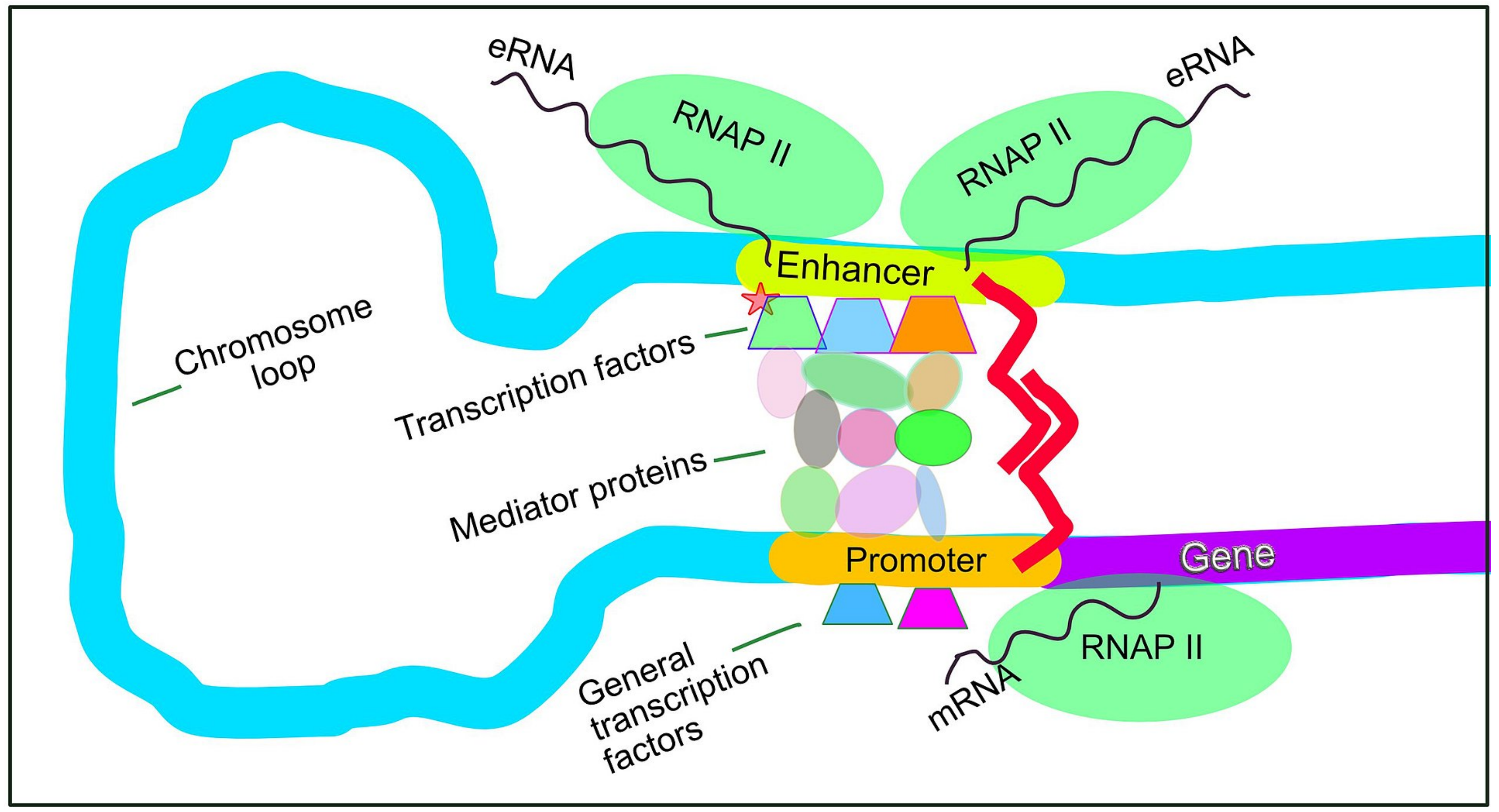
- protein coding: 1.22%
- non-coding: 98.78%

Non-coding DNA

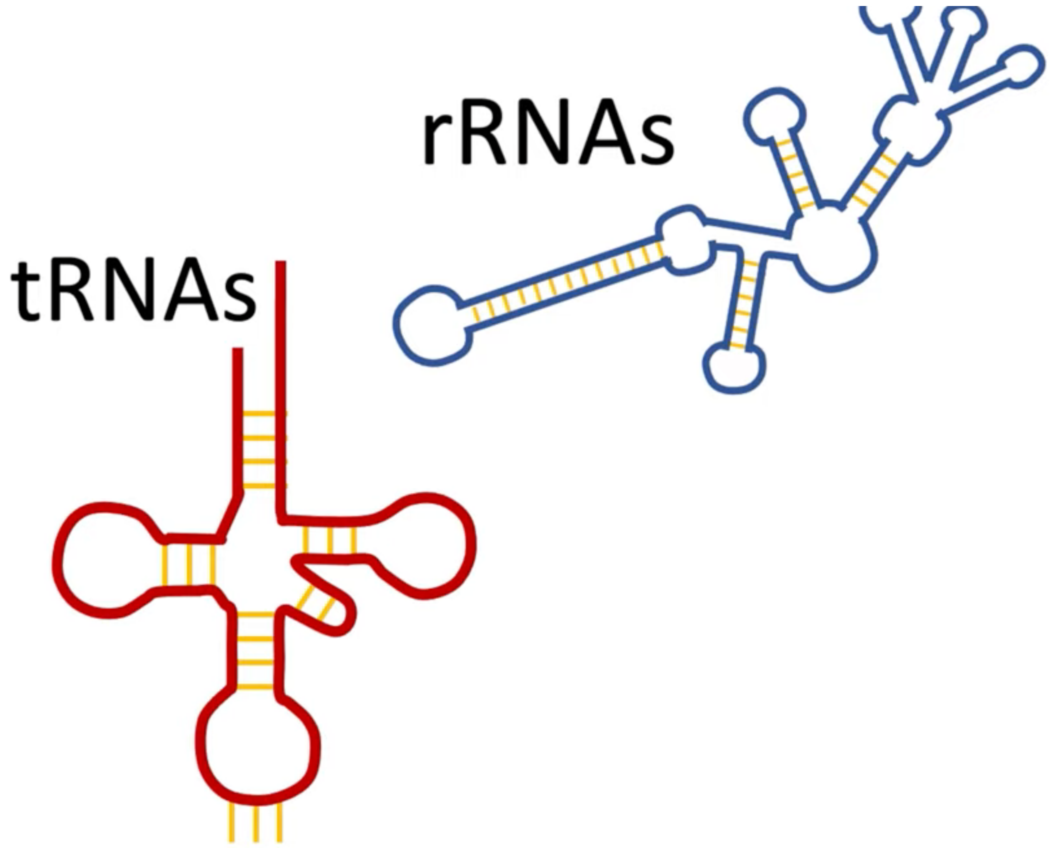
Non-coding DNA: DNA does not code for protein

Gene regulatory sequences

promoters (1% of genome)
enhancers (10% of genome)



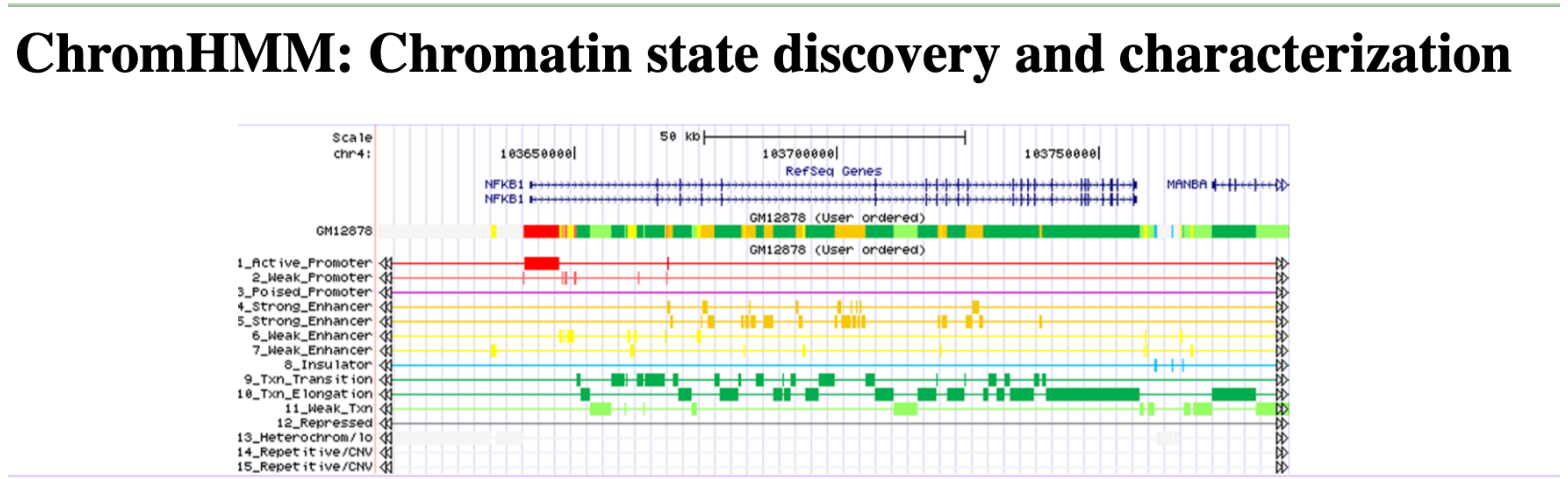
non-coding RNA



Others

introns
telomeres
satellite DNA
…

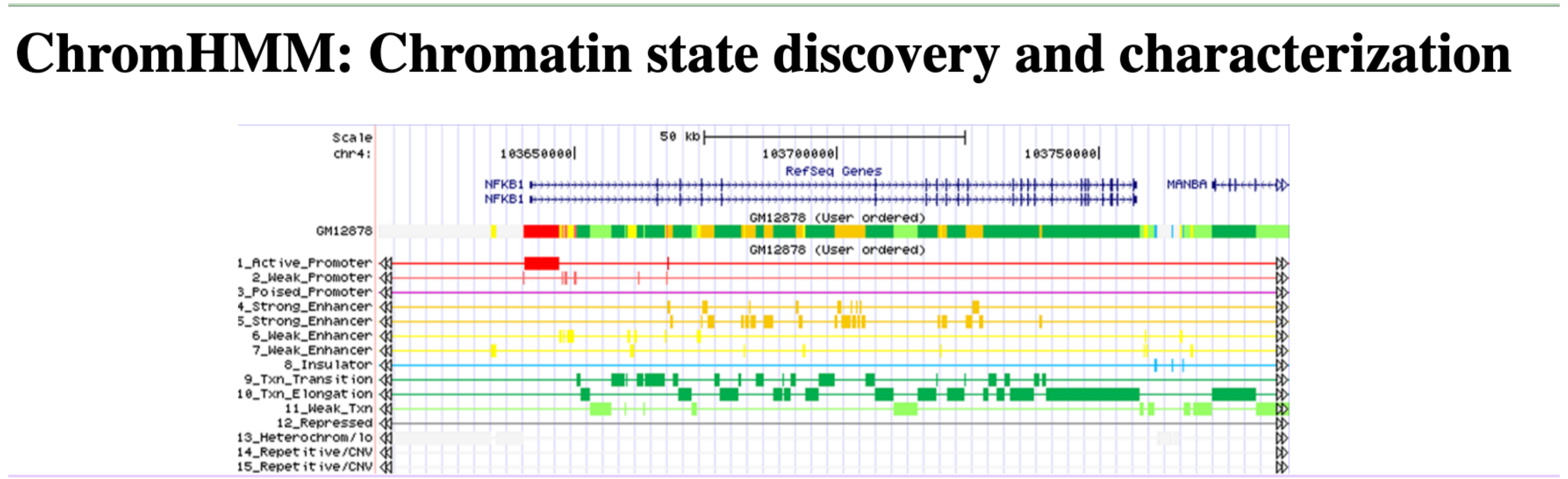
Annotation of non-coding regions


ChromHMM: Chromatin state discovery and characterization

ChromHMM: integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to annotate chromatin states.

Ernst J and Kellis M, 2012; Vu H and Ernst J, 2022

Annotation of non-coding regions



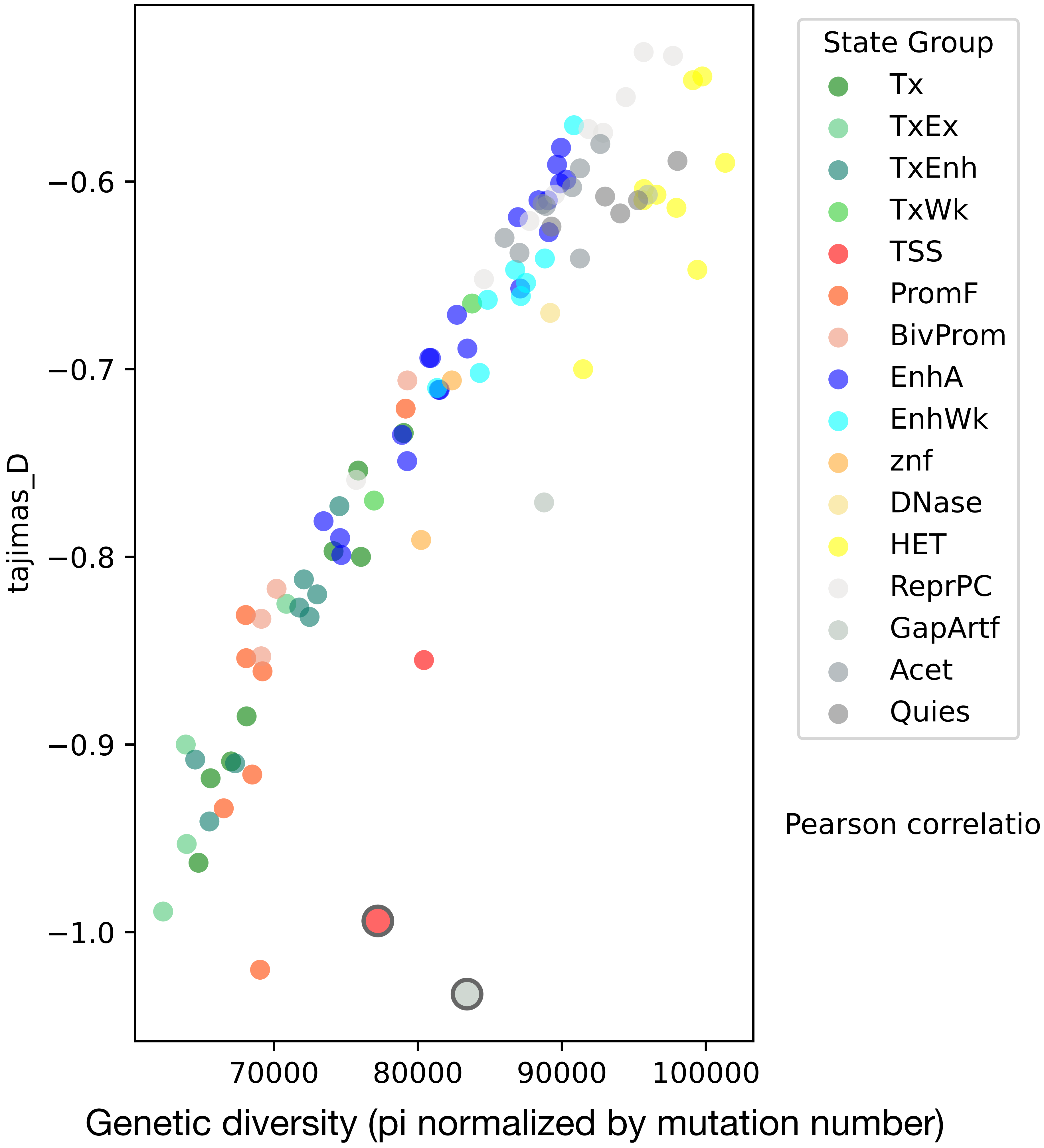
ChromHMM: Chromatin state discovery and characterization

ChromHMM: integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to annotate chromatin states.

Different chromatin states ~ Selection?

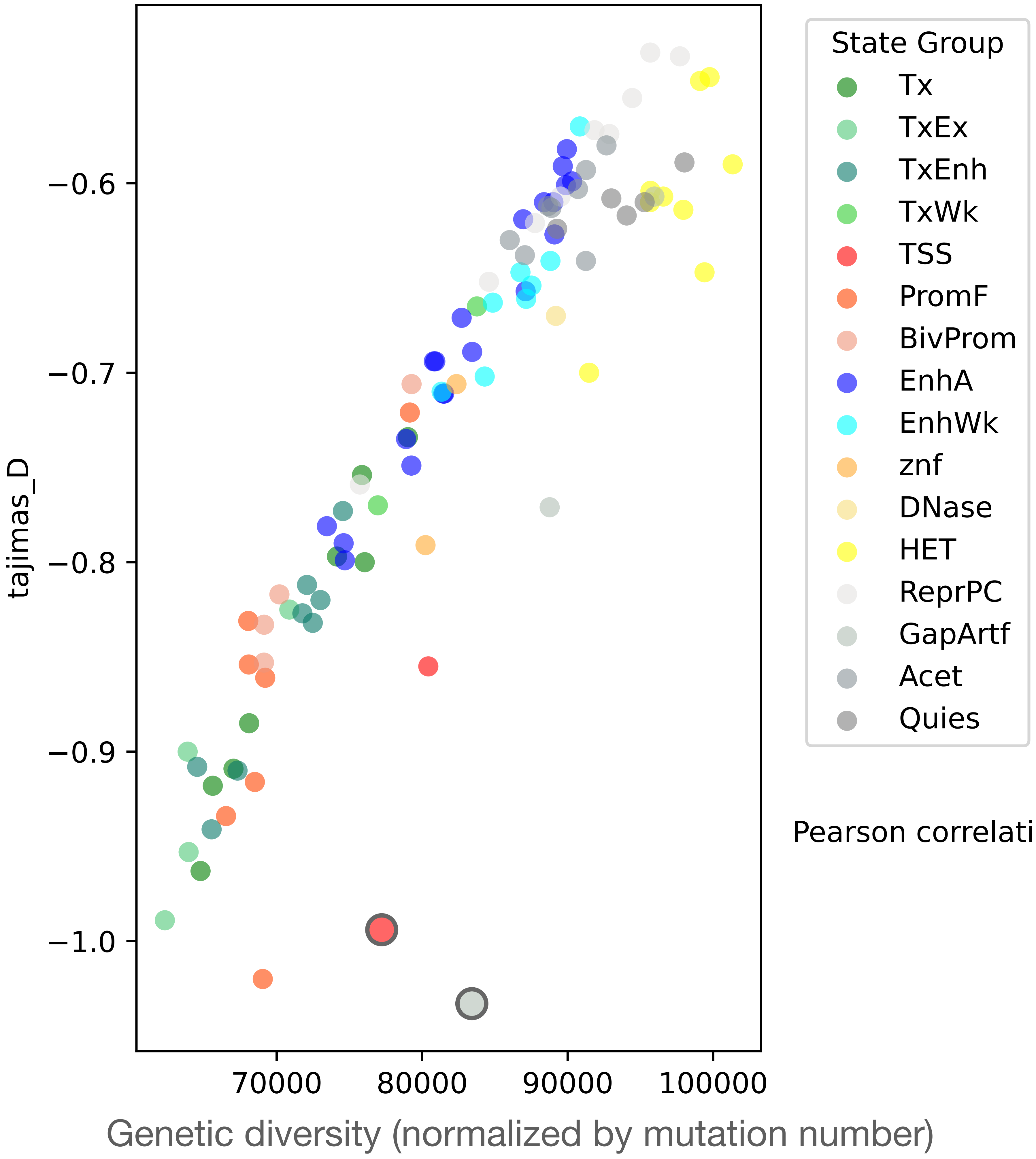
Ernst J and Kellis M, 2012; Vu H and Ernst J, 2022

Patterns of genetic variation vary across noncoding annotations

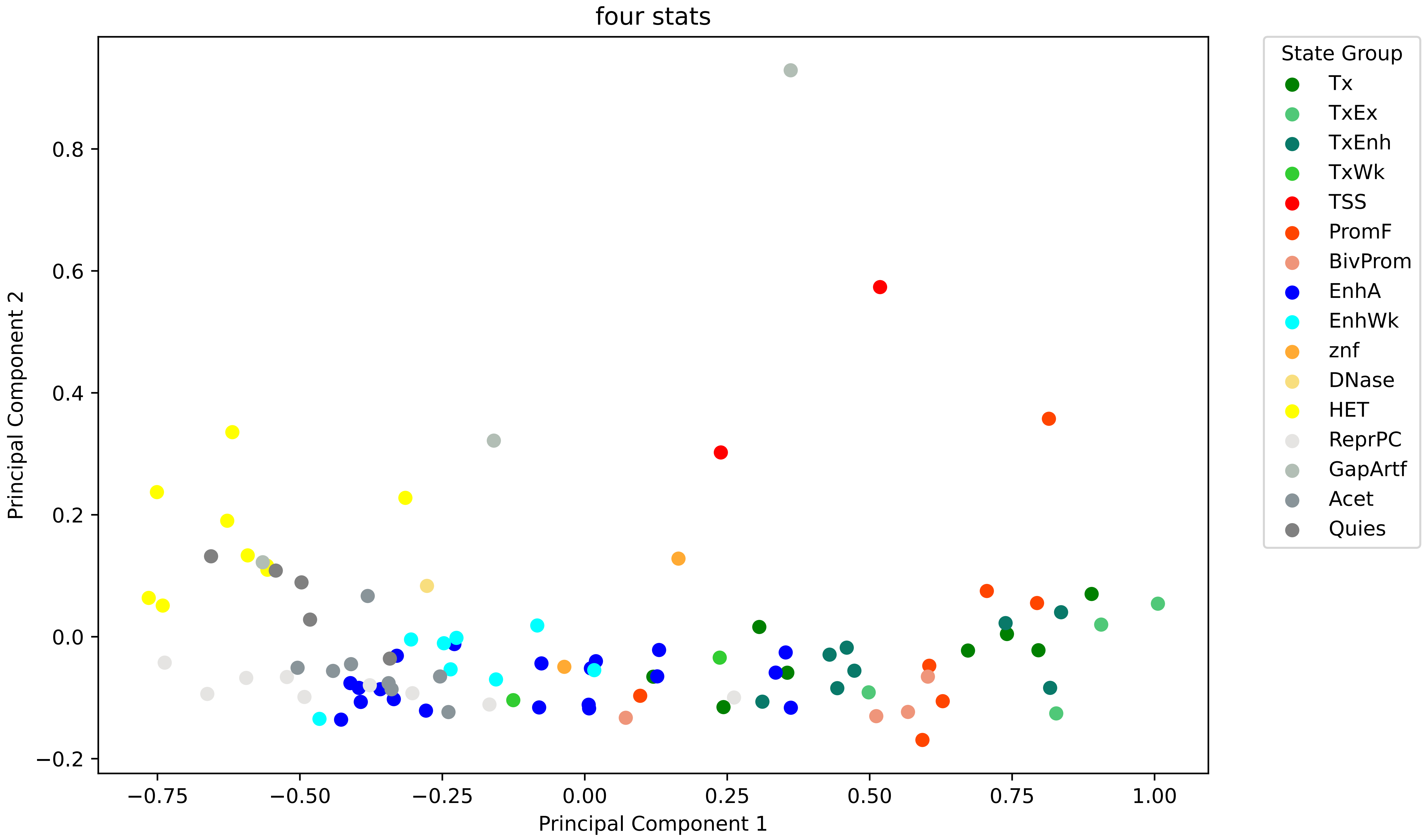


Pearson correlation: 0.90

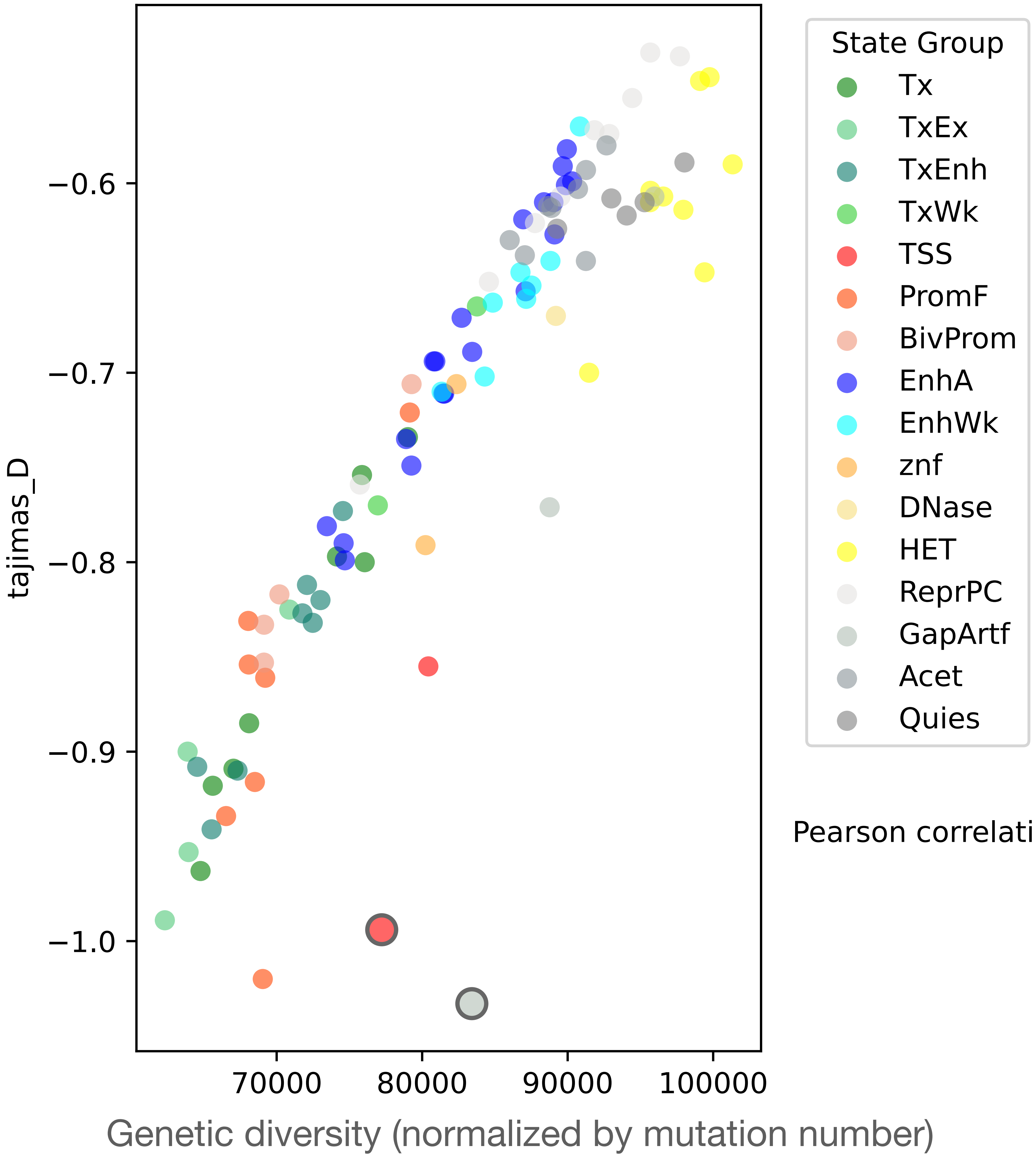
Patterns of genetic variation vary across noncoding annotations



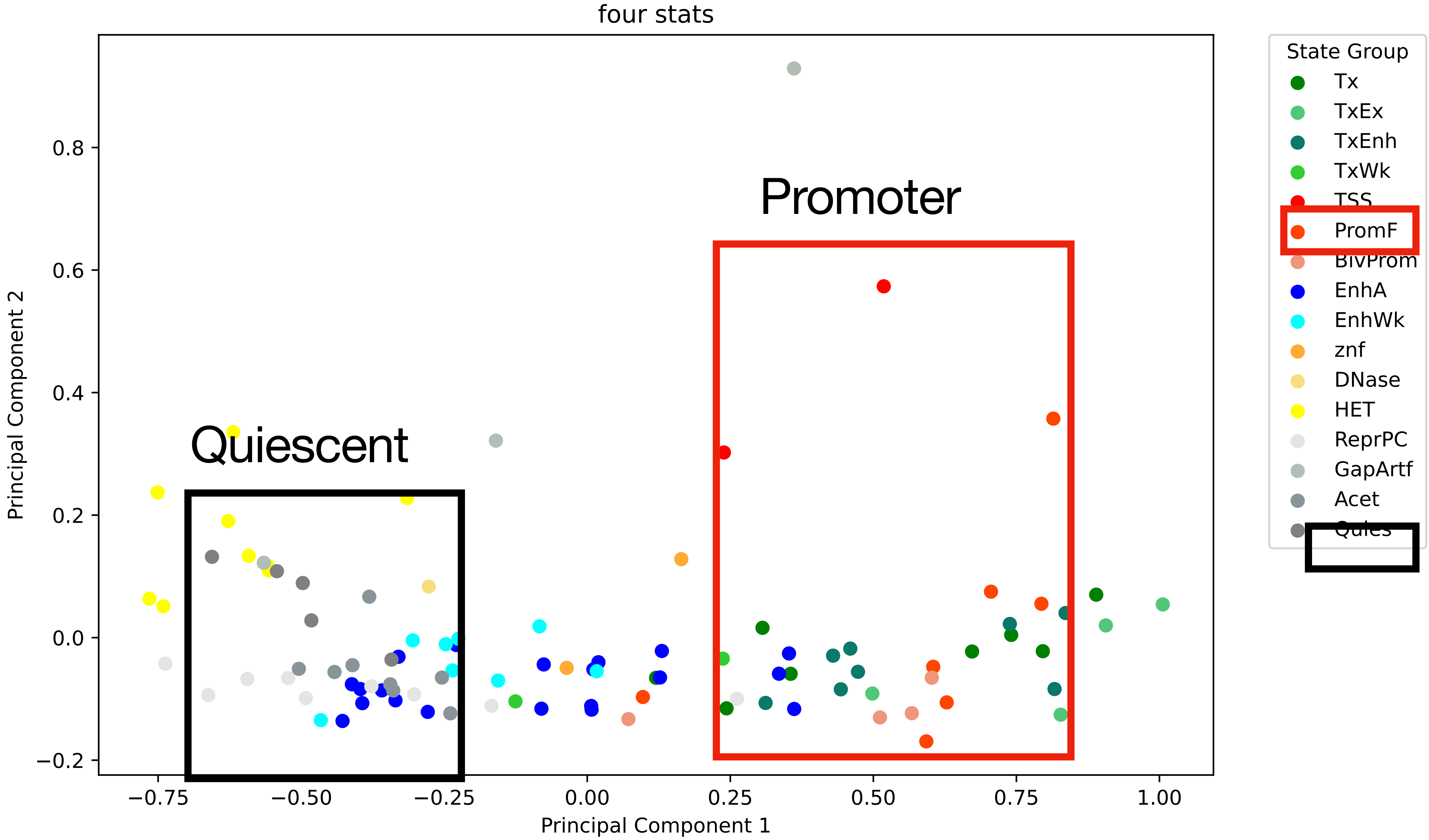
Pearson correlation: 0.90



Patterns of genetic variation vary across noncoding annotations



Pearson correlation: 0.90



What are the fitness effects of non-coding mutations?

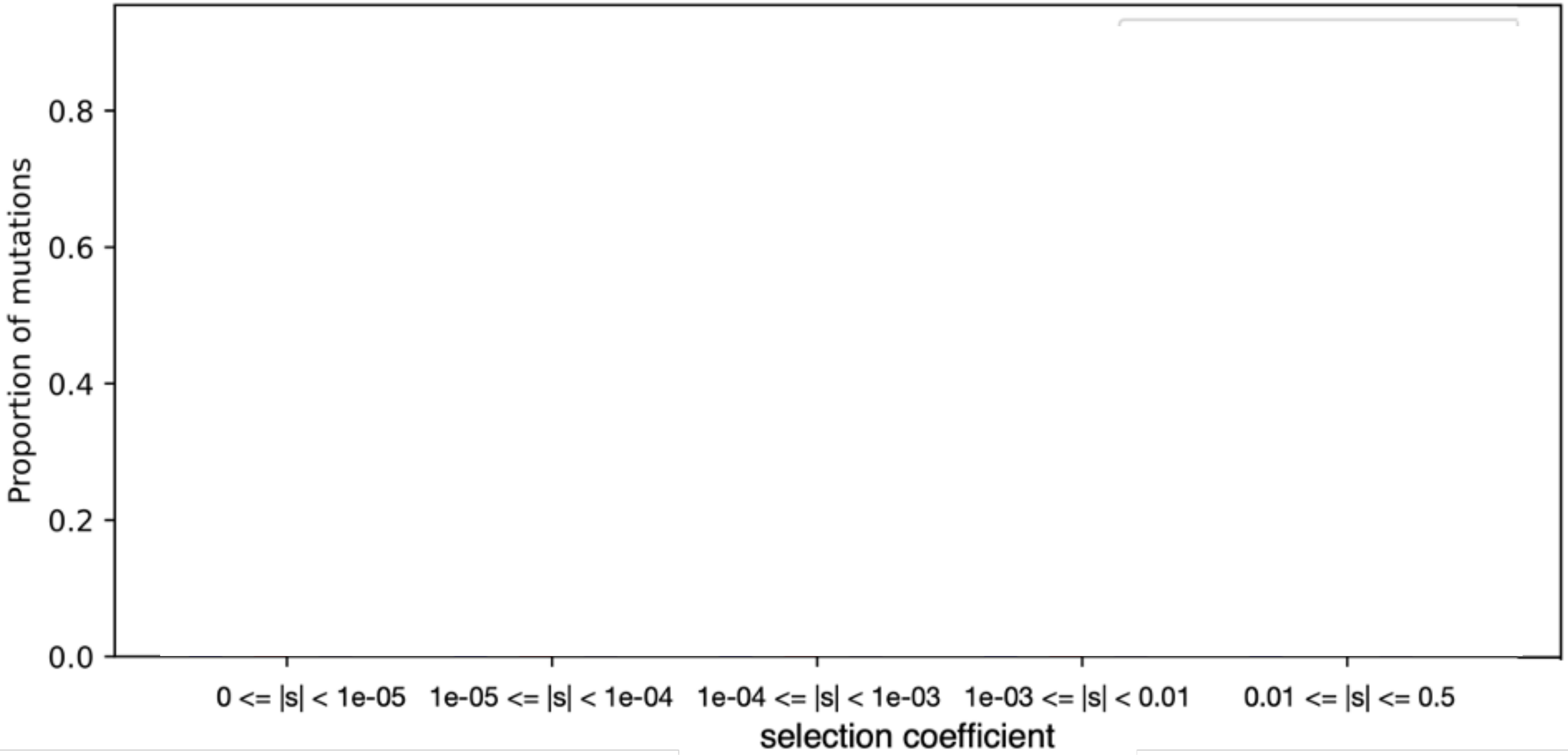
What are the fitness effects of non-coding mutations?

Fitness effects: selection coefficients

What are the fitness effects of non-coding mutations?

Fitness effects: selection coefficients

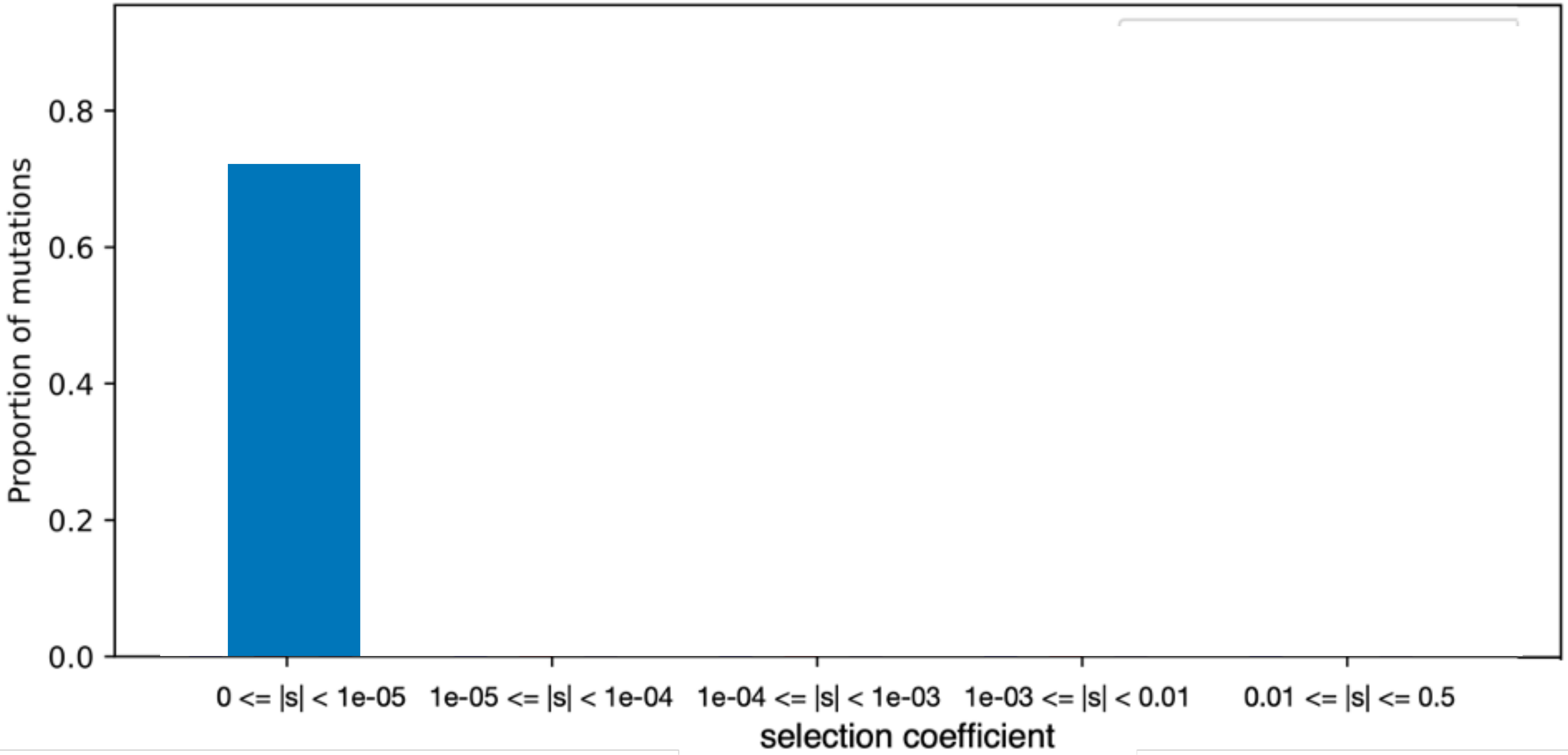
Distribution of Fitness Effects (DFE)



What are the fitness effects of non-coding mutations?

Fitness effects: selection coefficients

Distribution of Fitness Effects (DFE)

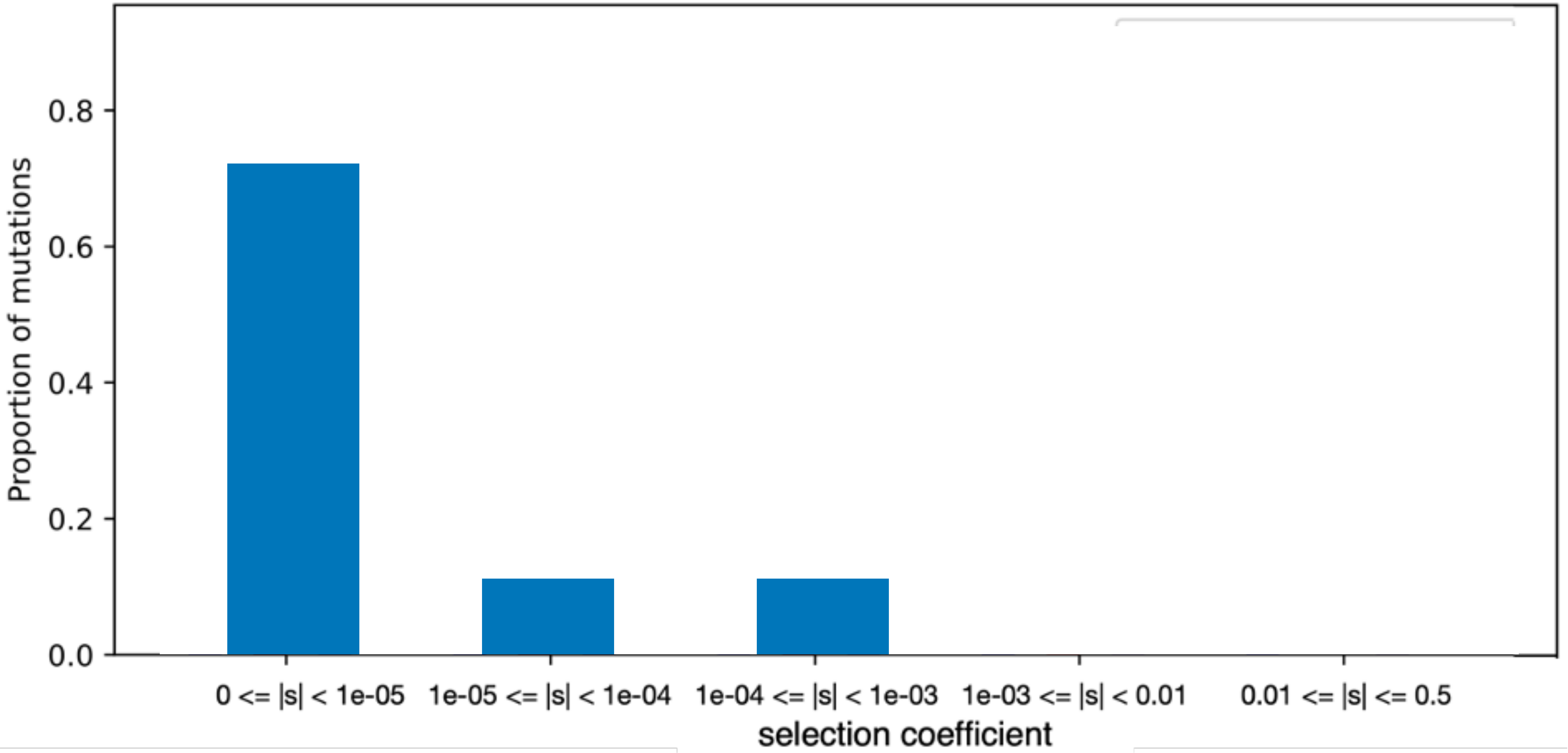


Neutral

What are the fitness effects of non-coding mutations?

Fitness effects: selection coefficients

Distribution of Fitness Effects (DFE)

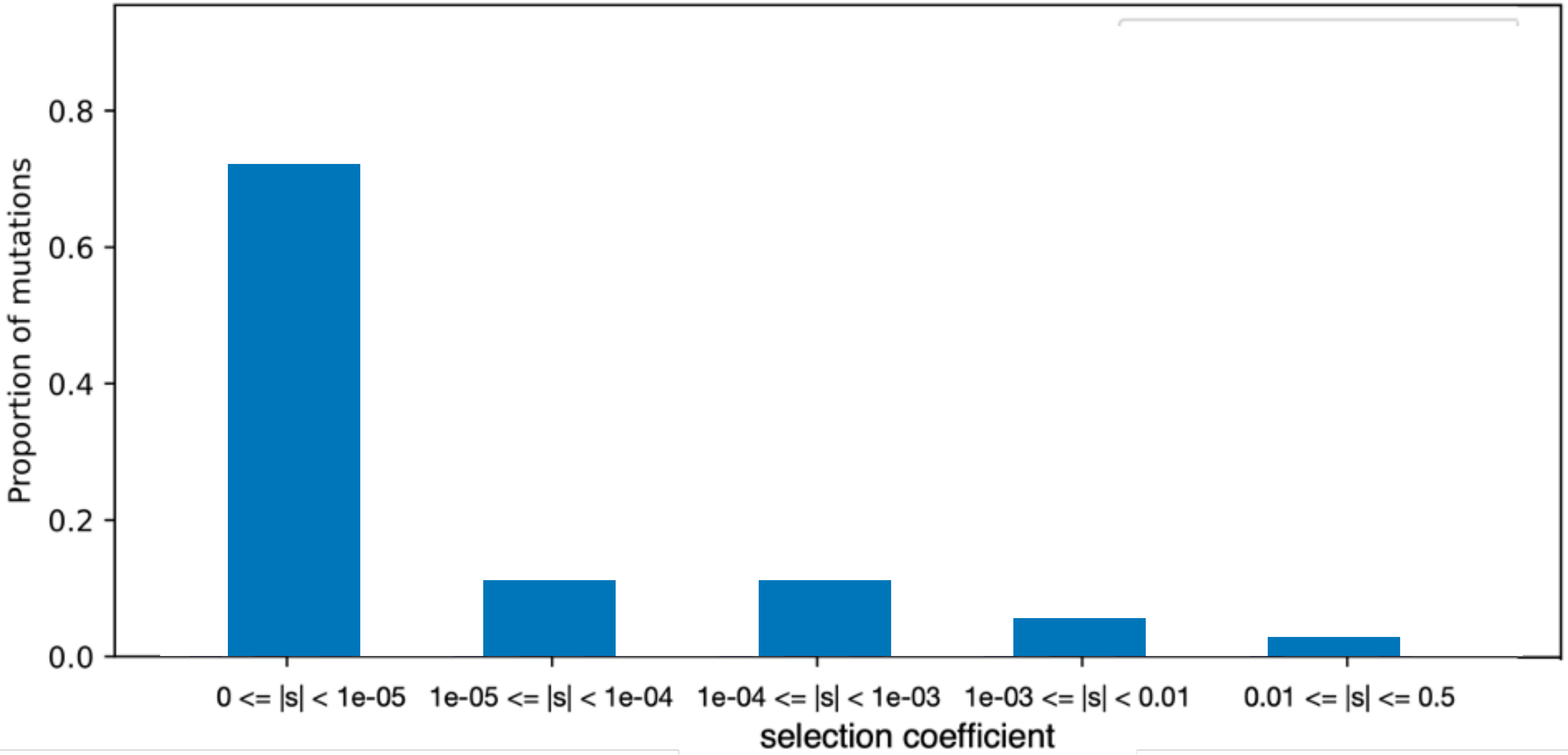


Neutral Weakly deleterious

What are the fitness effects of non-coding mutations?

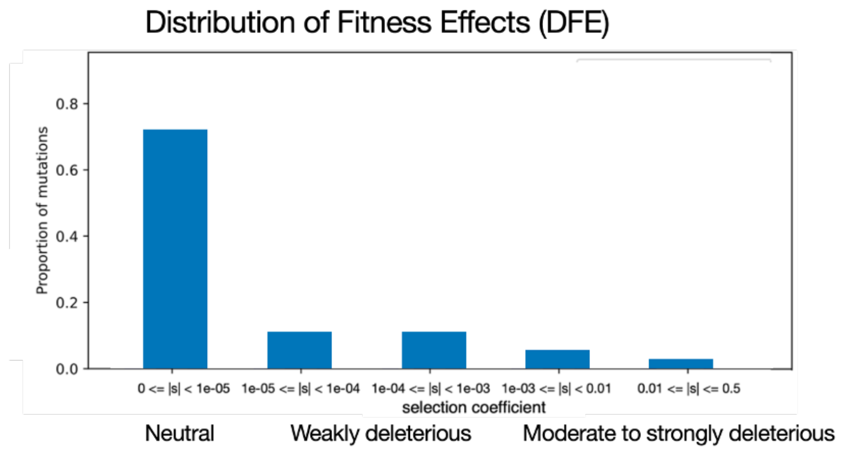
Fitness effects: selection coefficients

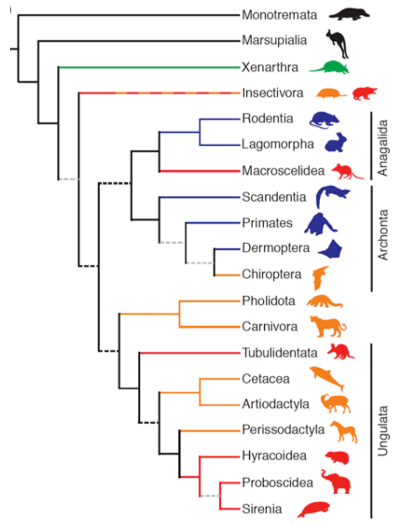
Distribution of Fitness Effects (DFE)



Neutral Weakly deleterious Moderate to strongly deleterious

DFE of non-coding mutations





- Model-based method for inferring Distribution of Fitness effects (DFE)

- DFE of annotated functional non-coding regions in human genome

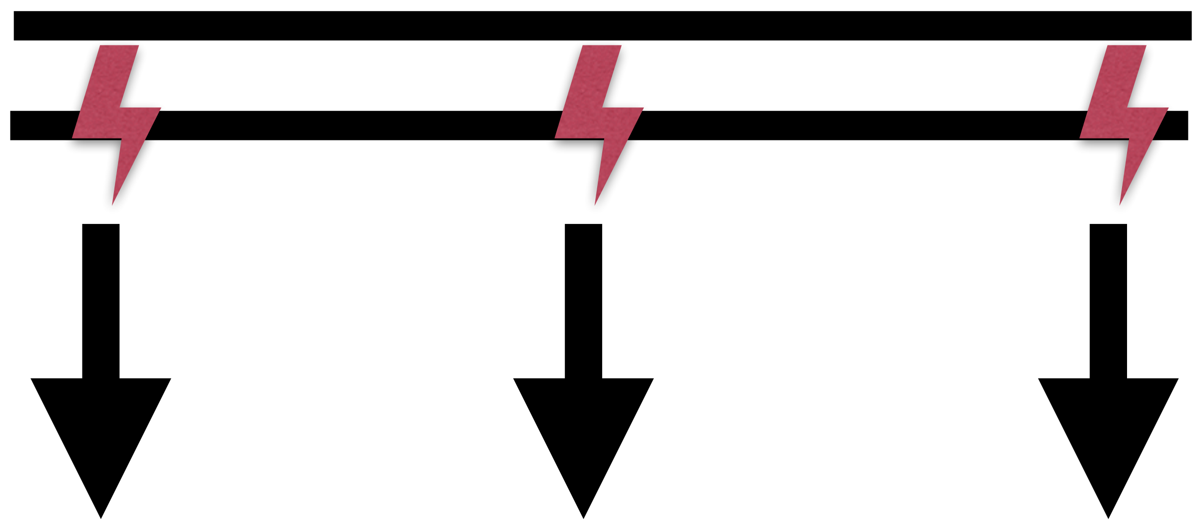
- Negative selection in conserved and non-conserved human genomic region

picture is from https://www.phonandroid.com/tesla-continue-de-decliner-le-cybertruck-a-toutes-les-sauces-avec-un-marteau-de-gym.html
Phylogeny is from Springer et al., 2004

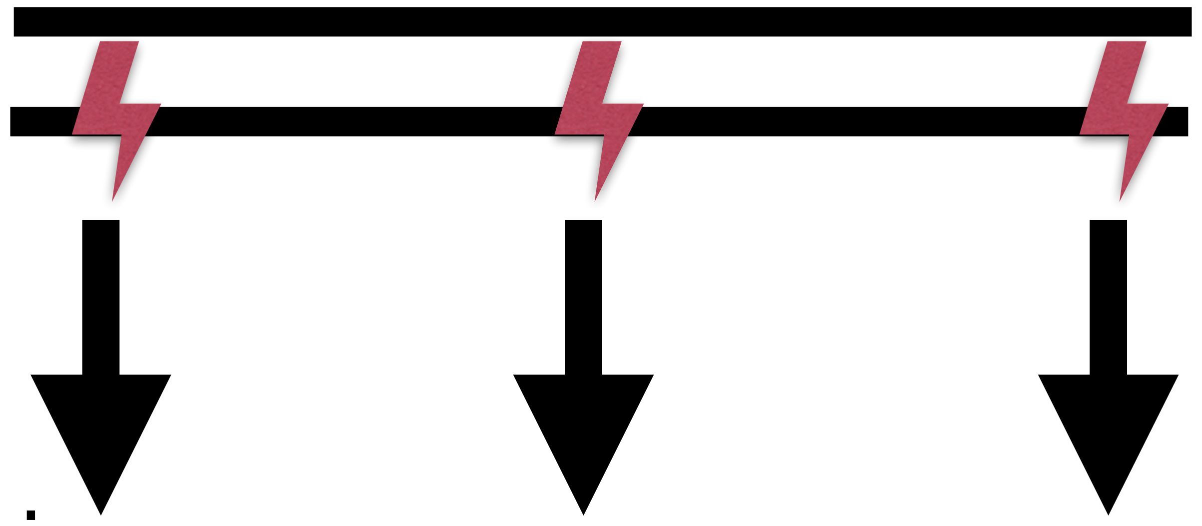
Infer selection coefficient by site frequency spectrum (SFS)

Infer selection coefficient by site frequency spectrum (SFS)

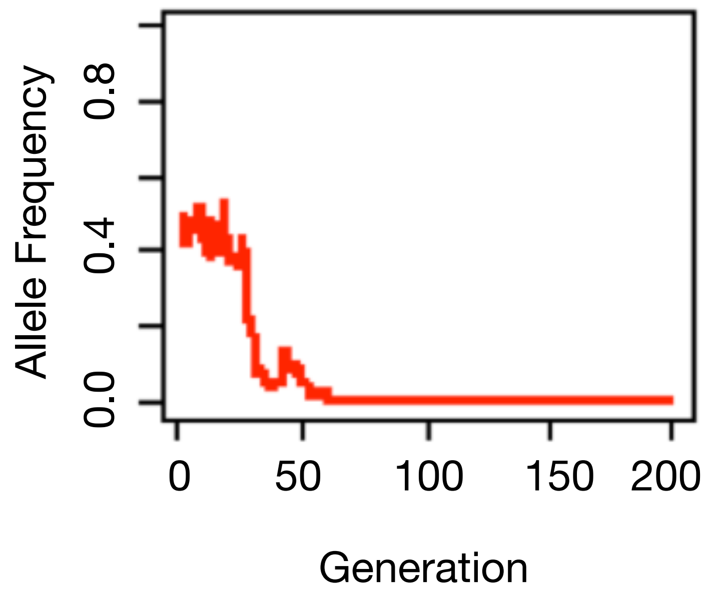
Infer selection coefficient by site frequency spectrum (SFS)



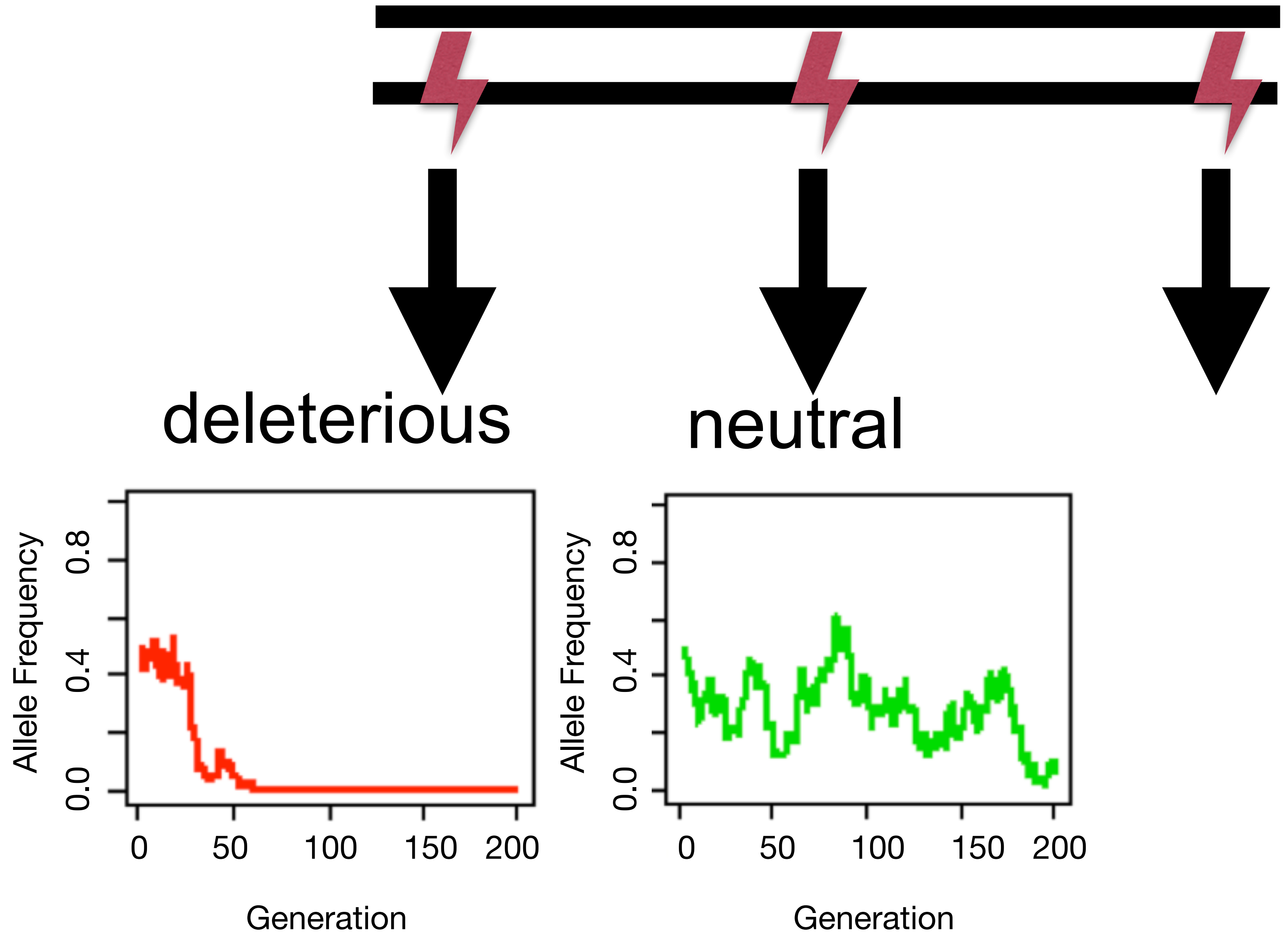
Infer selection coefficient by site frequency spectrum (SFS)



deleterious



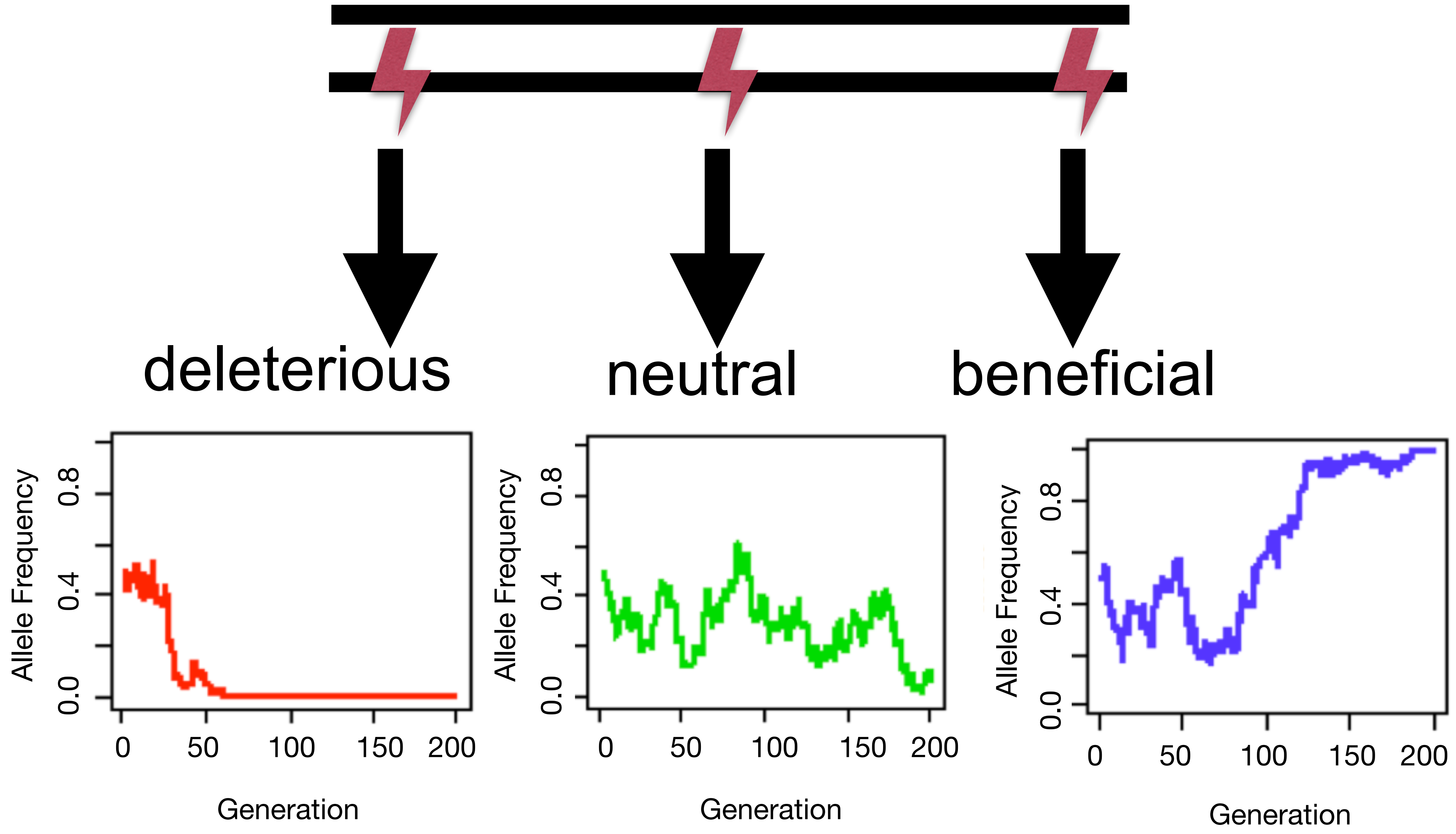
Infer selection coefficient by site frequency spectrum (SFS)



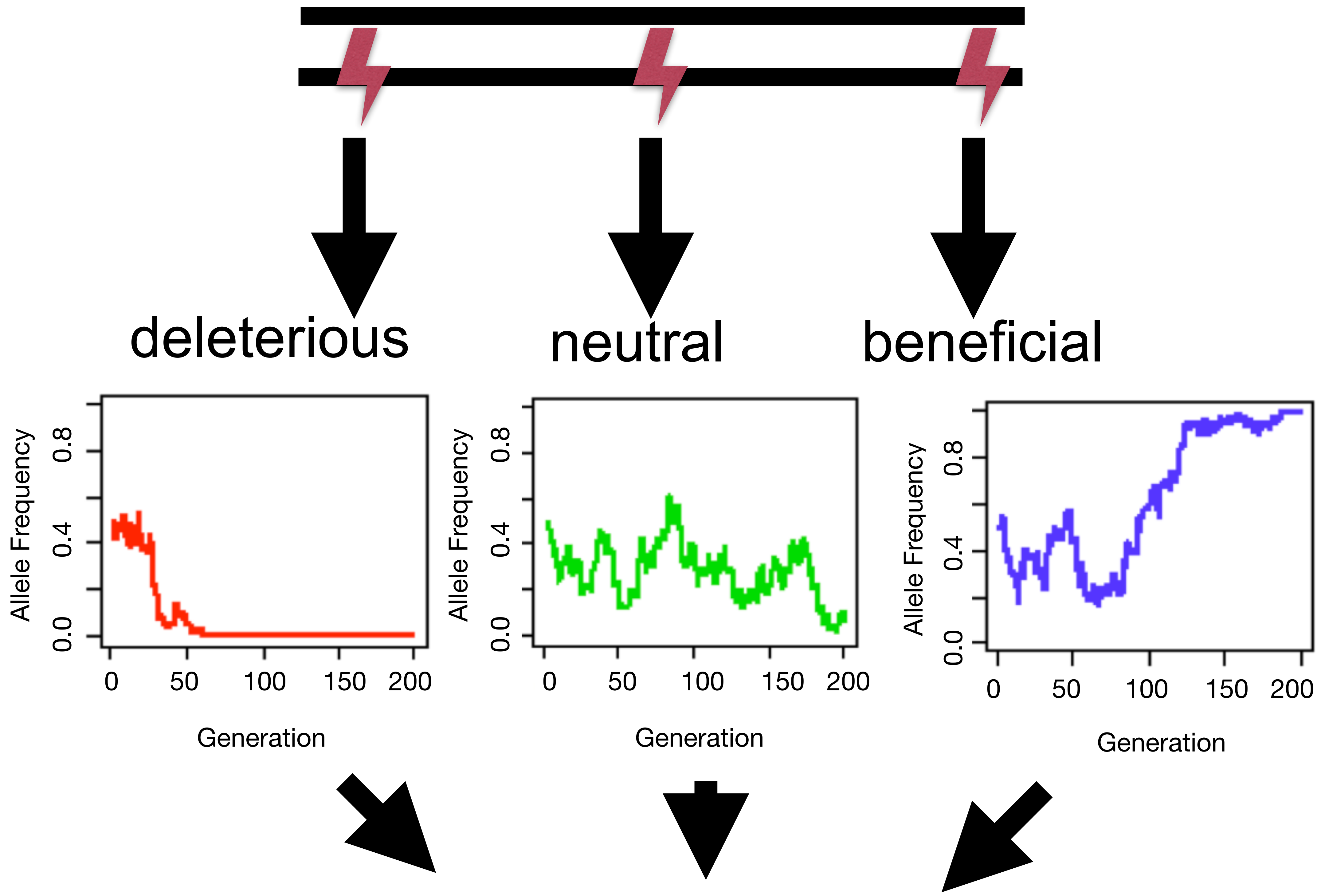
deleterious

neutral

Infer selection coefficient by site frequency spectrum (SFS)



Infer selection coefficient by site frequency spectrum (SFS)

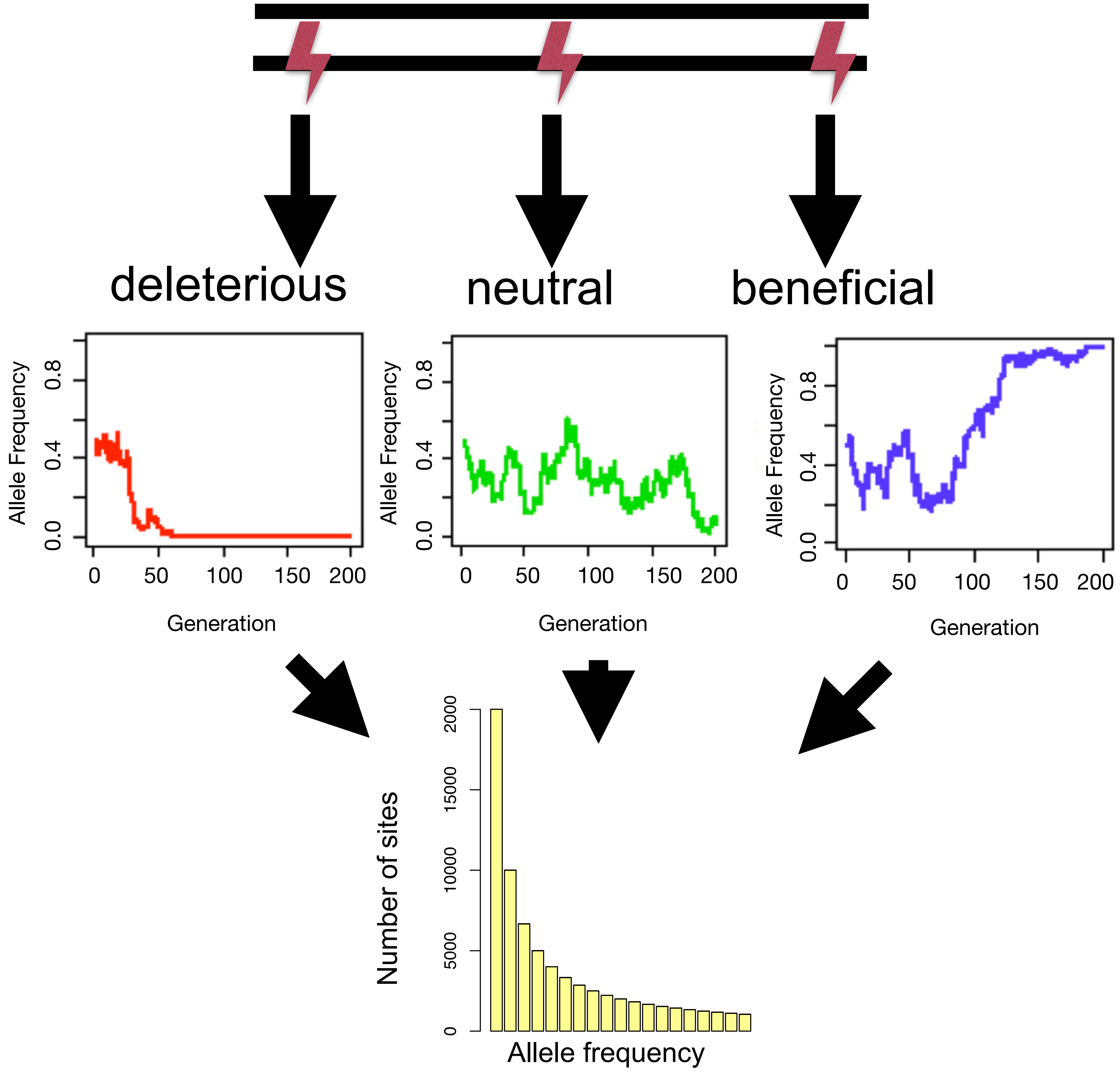


deleterious

neutral

beneficial

Infer selection coefficient by site frequency spectrum (SFS)

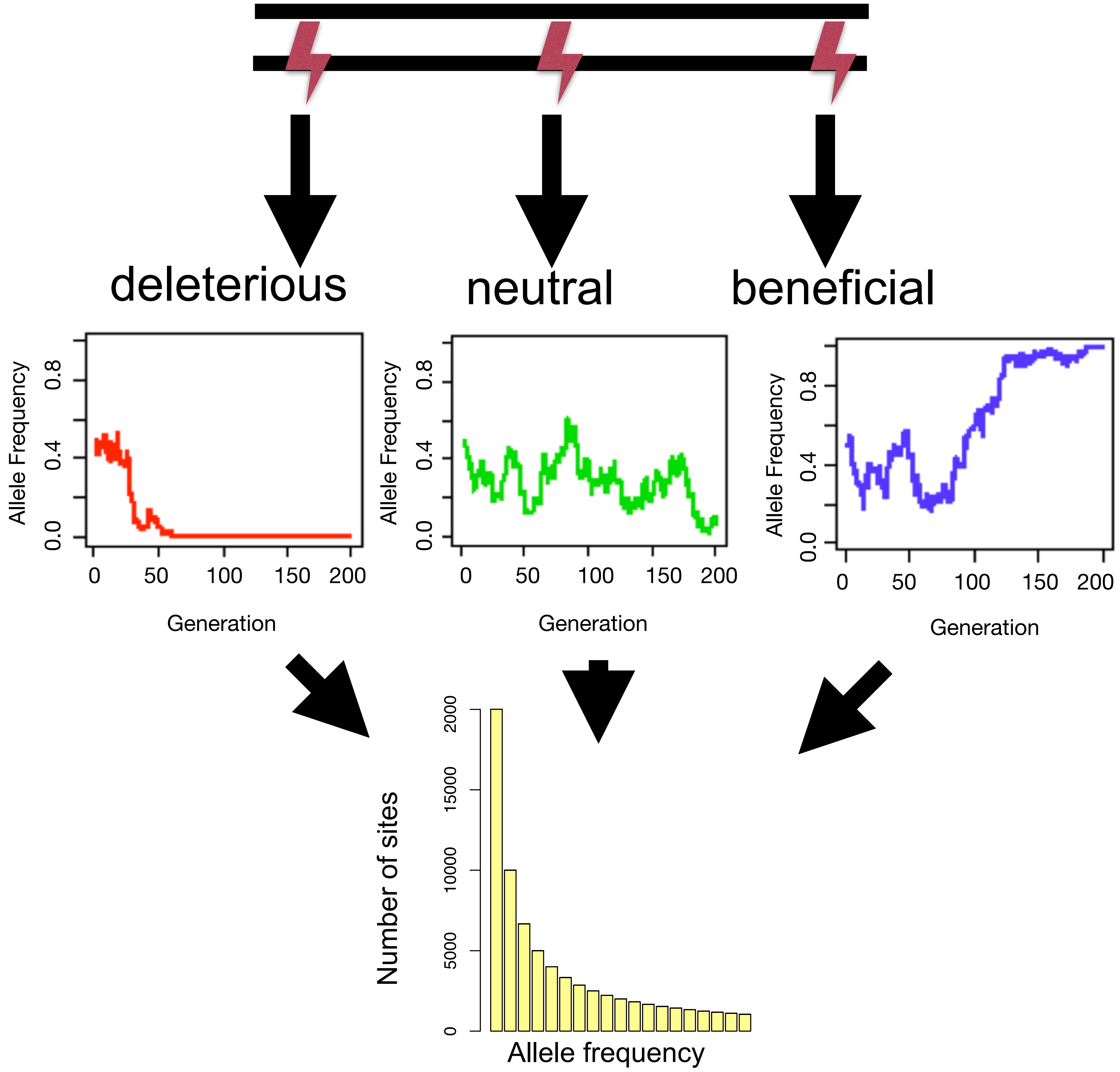


deleterious

neutral

beneficial

Infer selection coefficient by site frequency spectrum (SFS)



deleterious

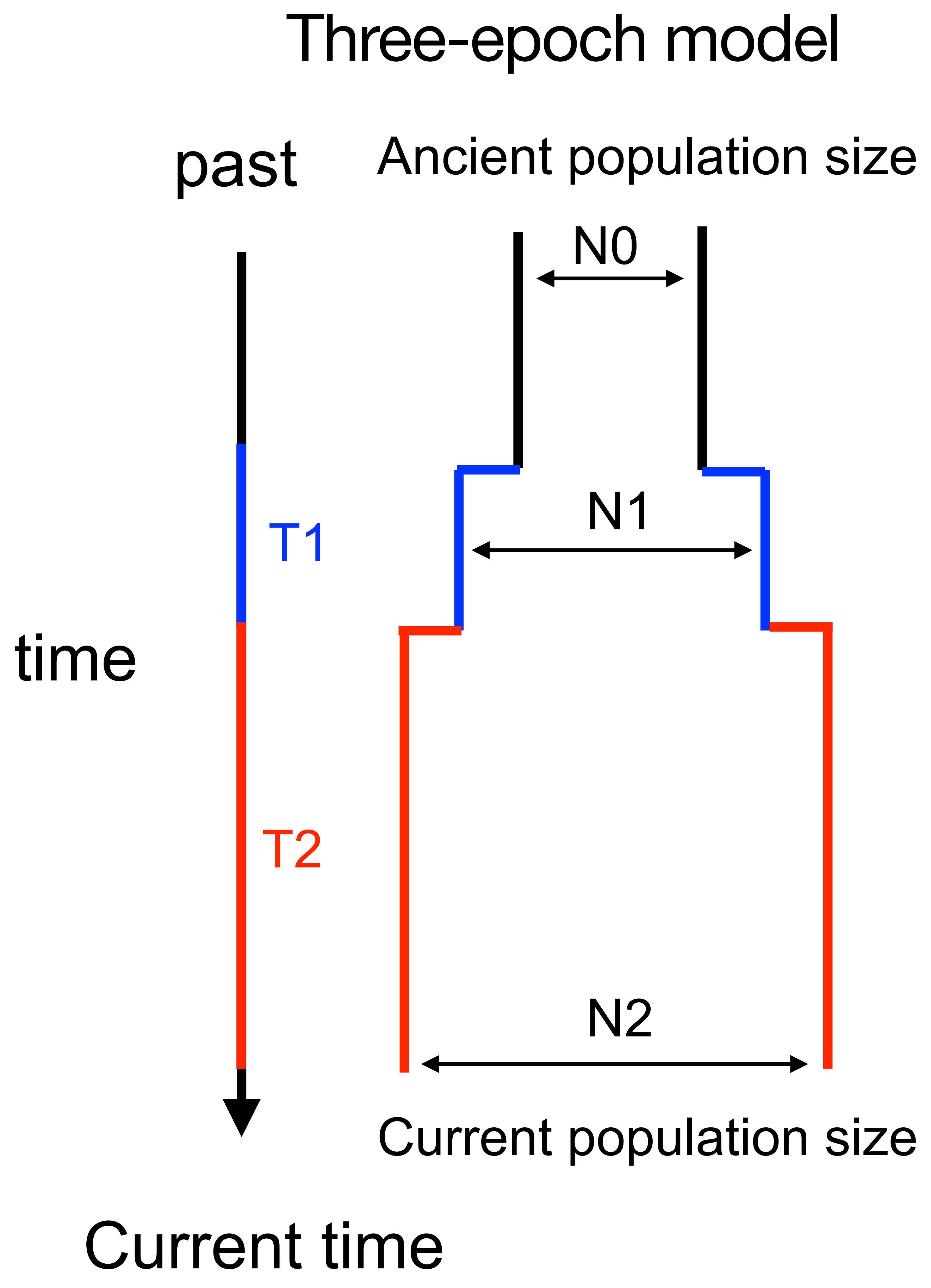
neutral

beneficial

1000 genome project:
Yoruba population (YRI)

Demographic history influences SFS

Demographic history: Changes of population size

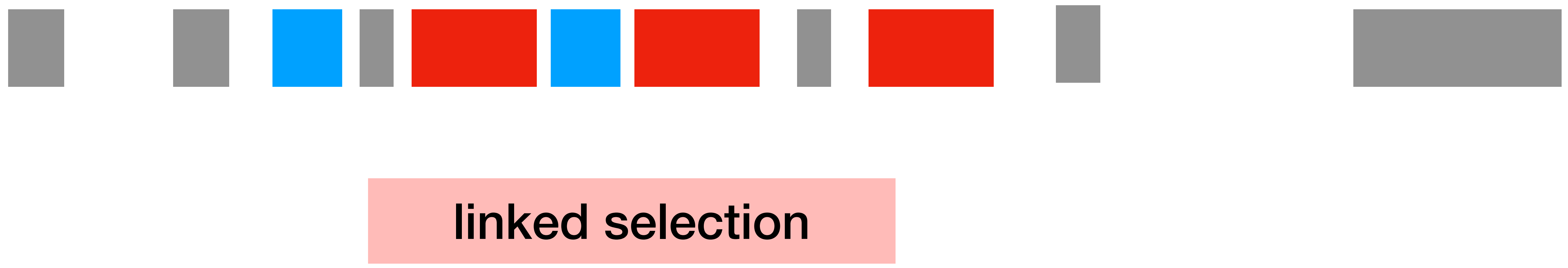


Linked selection influences SFS

◼ quiescent region

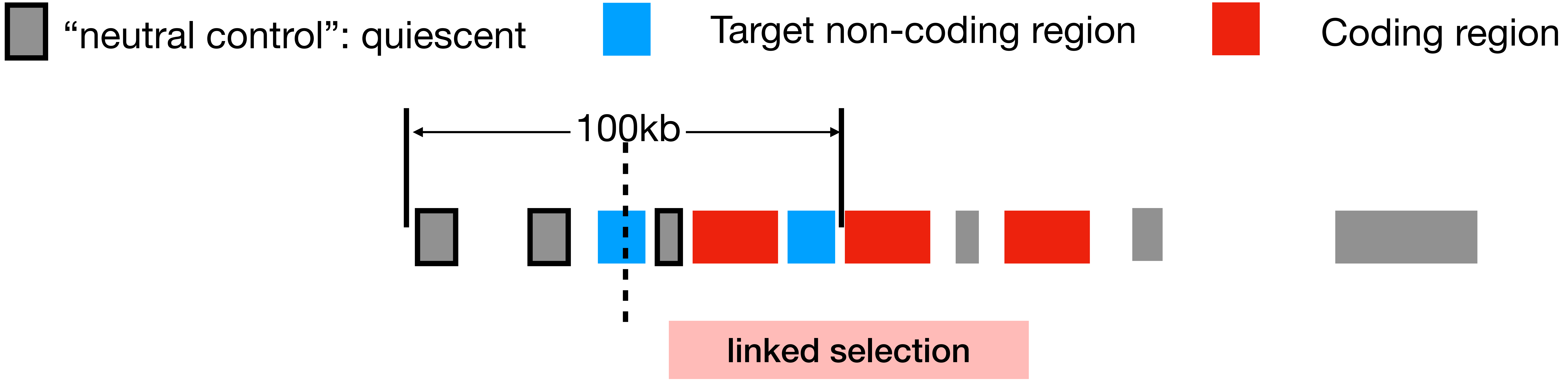
🟦 Target non-coding region

🟥 Coding region



linked selection

Method: Finding “neutral” controls

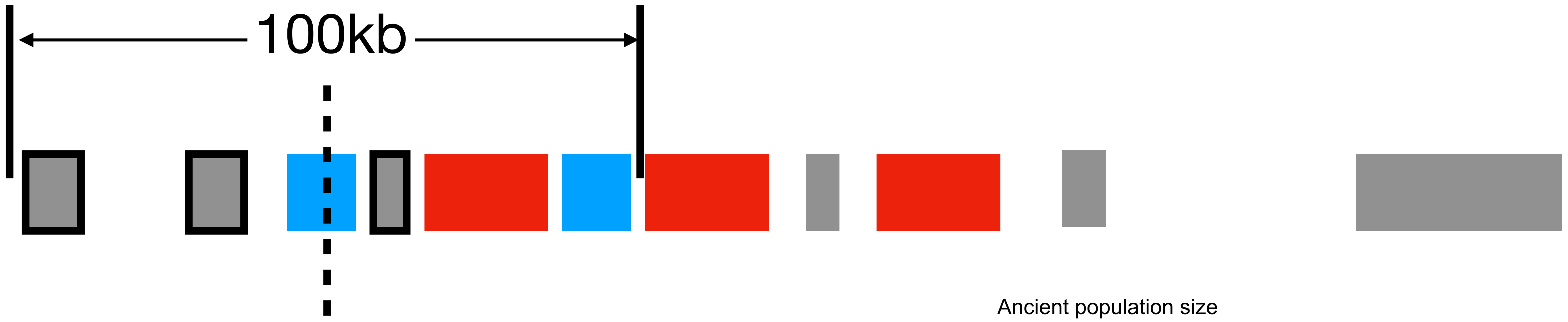


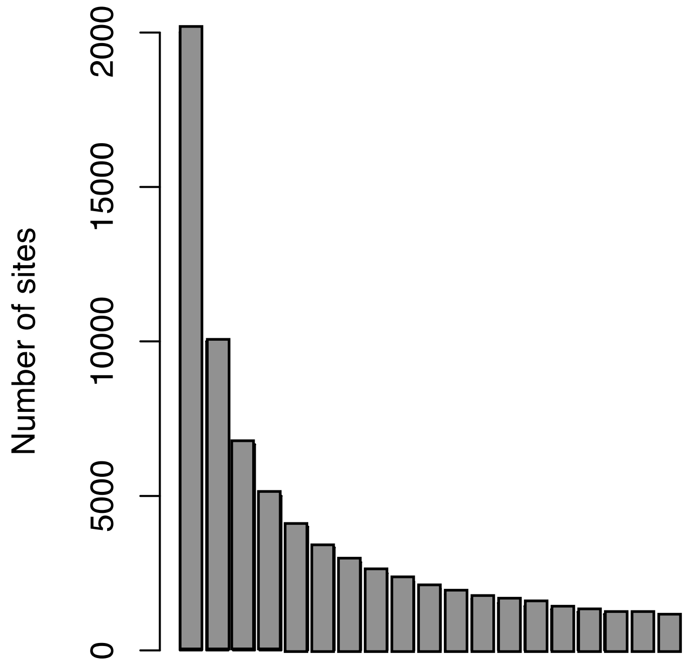
▨ “neutral control”: quiescent 🟦 Target non-coding region 🟥 Coding region

Gutenkunst et al., 2009; Kim et al., 2017

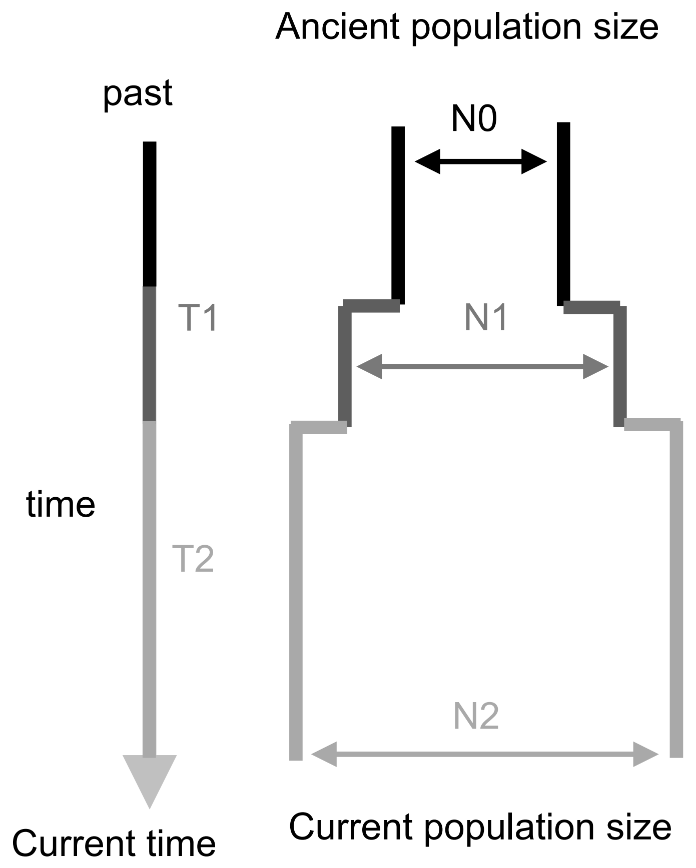
dadi: Infer models of demographic history

▨ “neutral control”: quiescent 🟦 Target non-coding region 🟥 Coding region





Site Frequency Spectrum (SFS)
of control sites



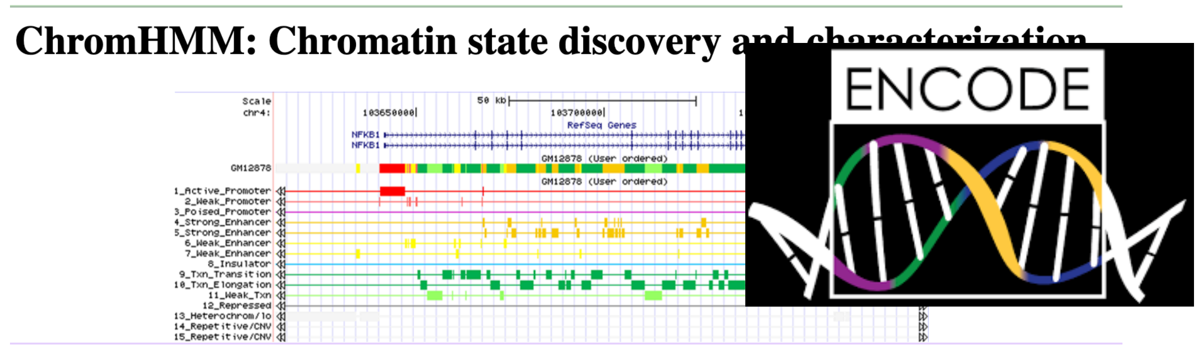
Demographic model

Gutenkunst et al., 2009; Kim et al., 2017

Workflow

Step1: Annotate genomic regions

🟦 Target non-coding region

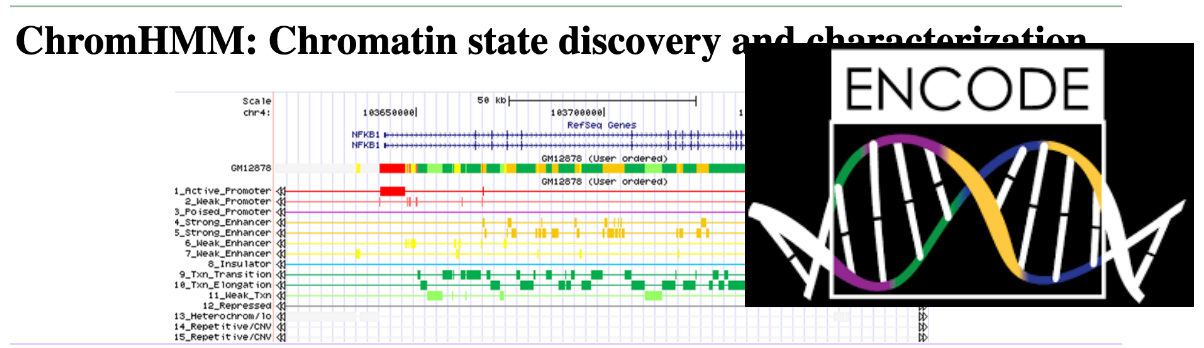


ChromHMM: Chromatin state discovery and characterization

Workflow

Step1: Annotate genomic regions

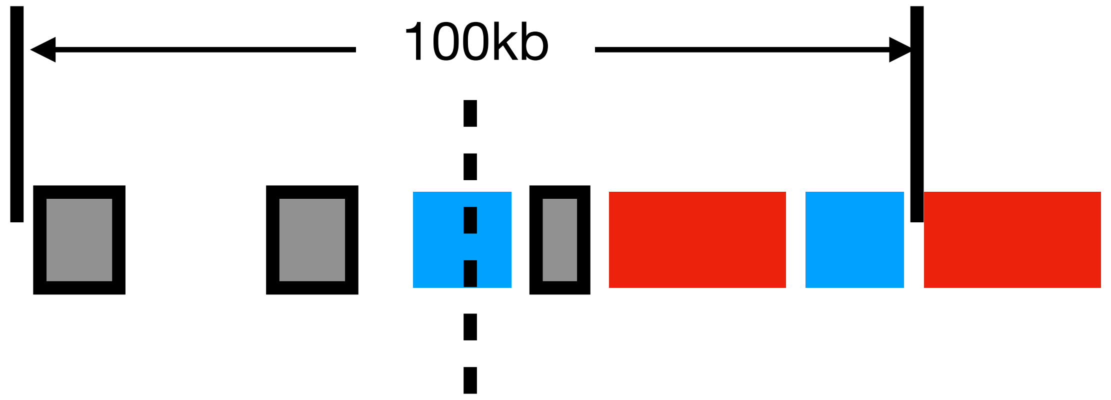
🟦 Target non-coding region



ChromHMM: Chromatin state discovery and characterization

Step2: Find quiescent control

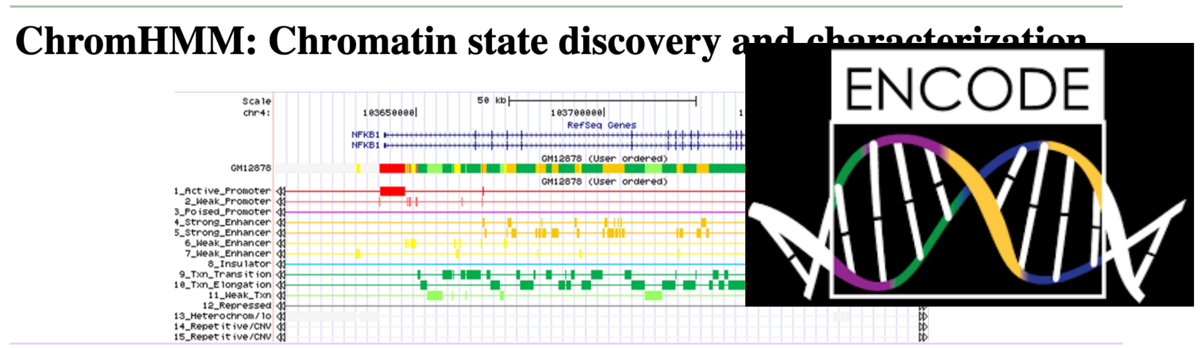
▨ “neutral control”: quiescent



Workflow

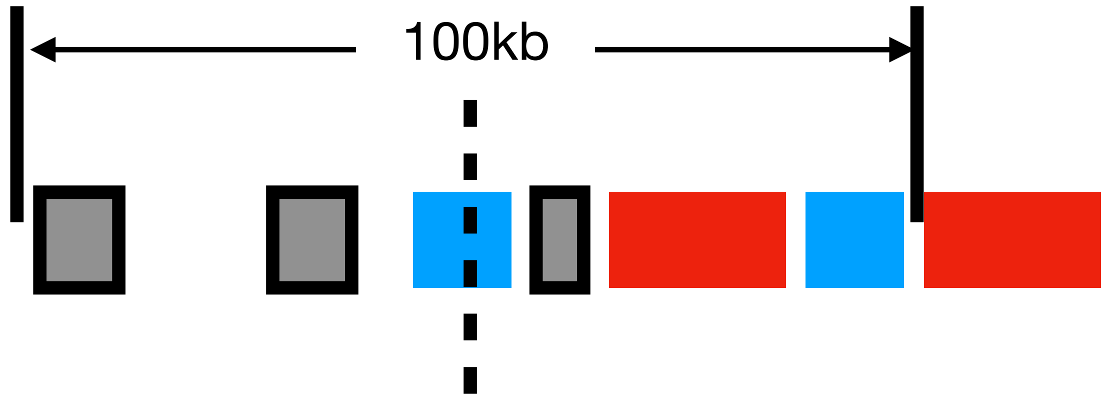
Step1: Annotate genomic regions

🟦 Target non-coding region

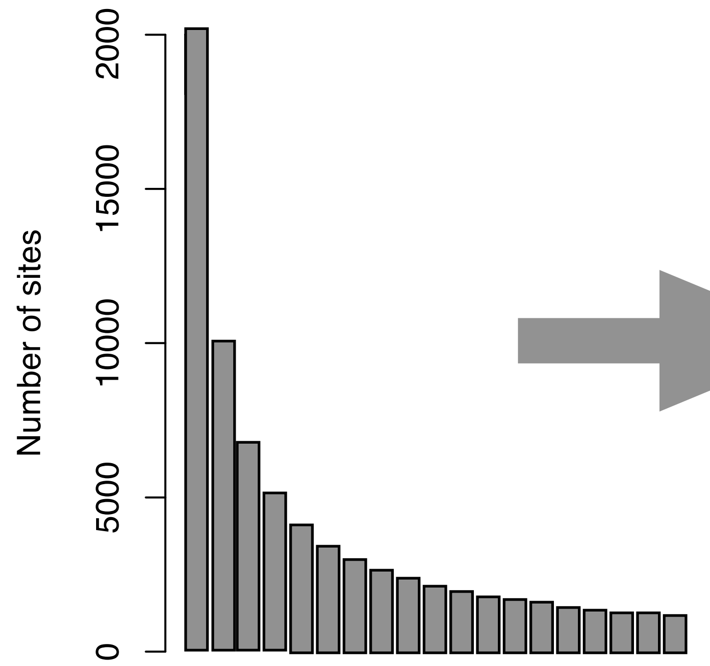

ChromHMM: Chromatin state discovery and characterization

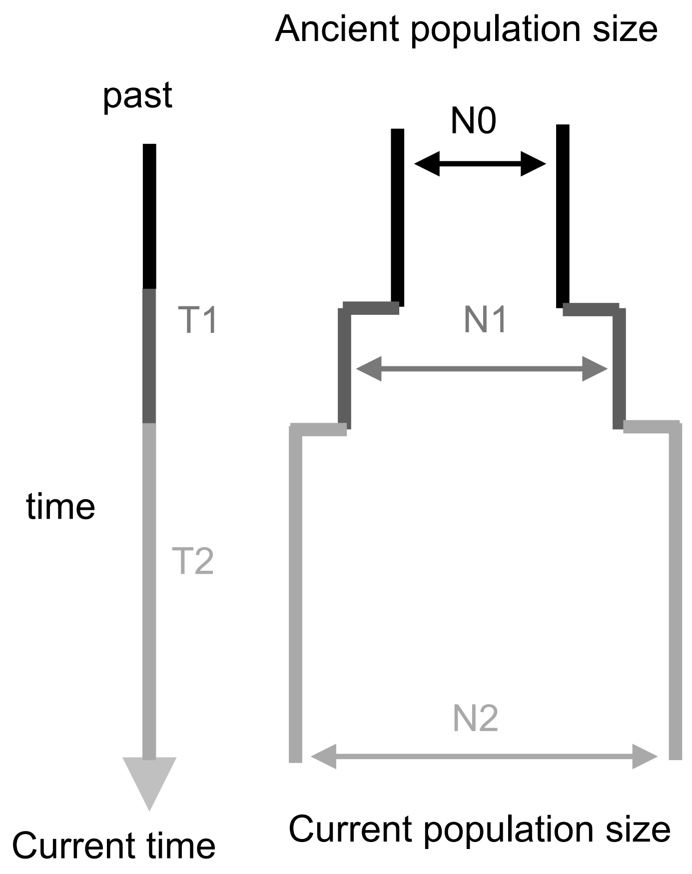
Step2: Find quiescent control

▨ “neutral control”: quiescent



Step3: Infer demographic model

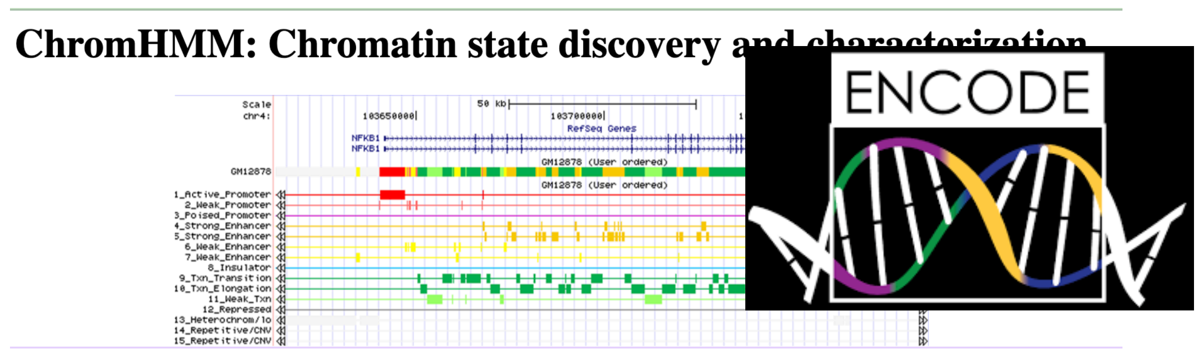




Workflow

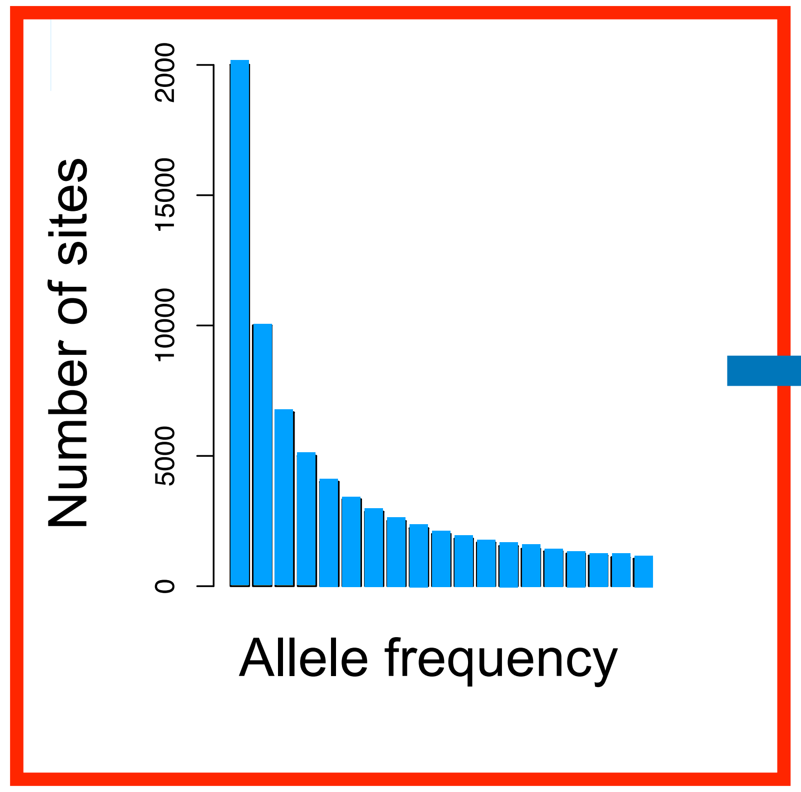
Step1: Annotate genomic regions

🟦 Target non-coding region



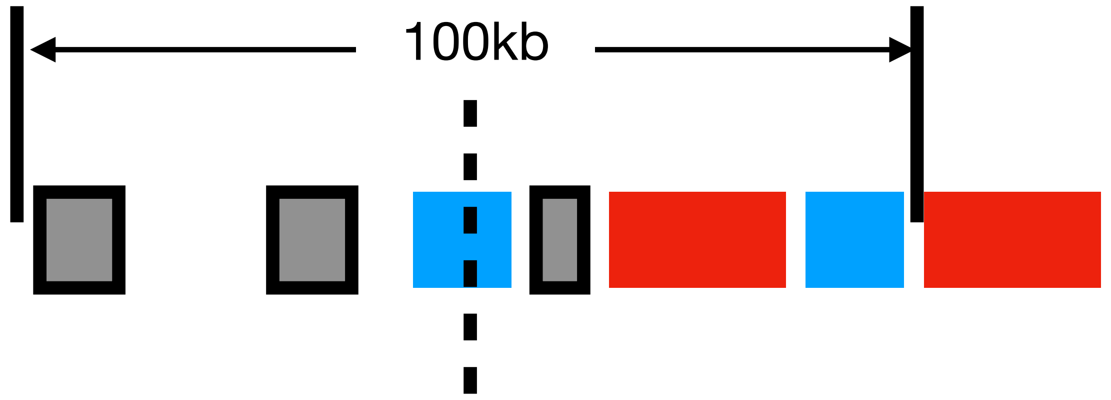
ChromHMM: Chromatin state discovery and characterization

Step4: Infer Distribution of Fitness Effecs (DFE)

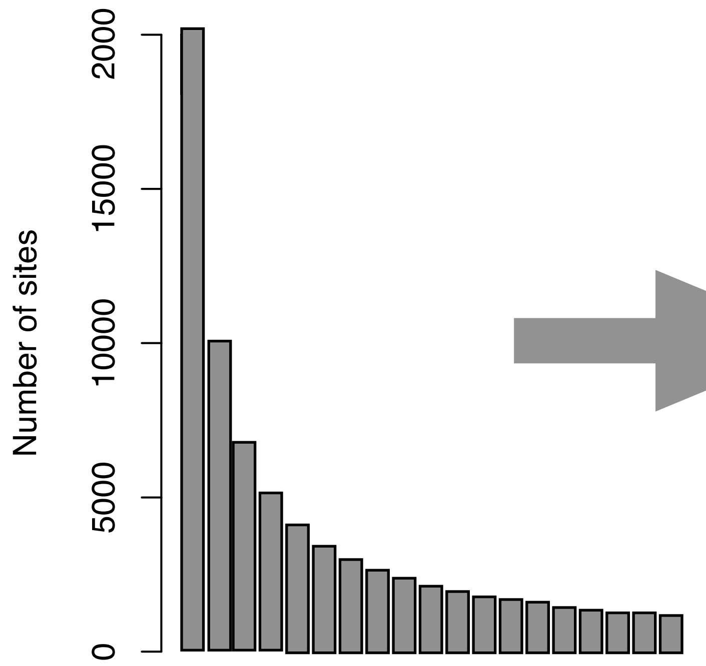


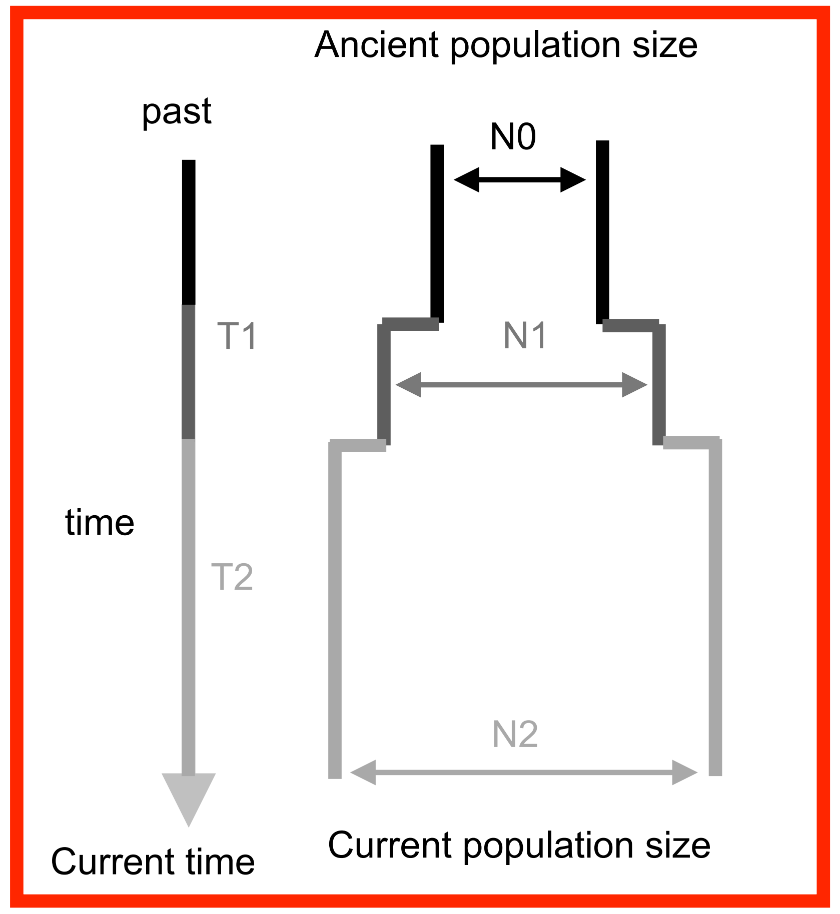
Step2: Find quiescent control

▨ “neutral control”: quiescent

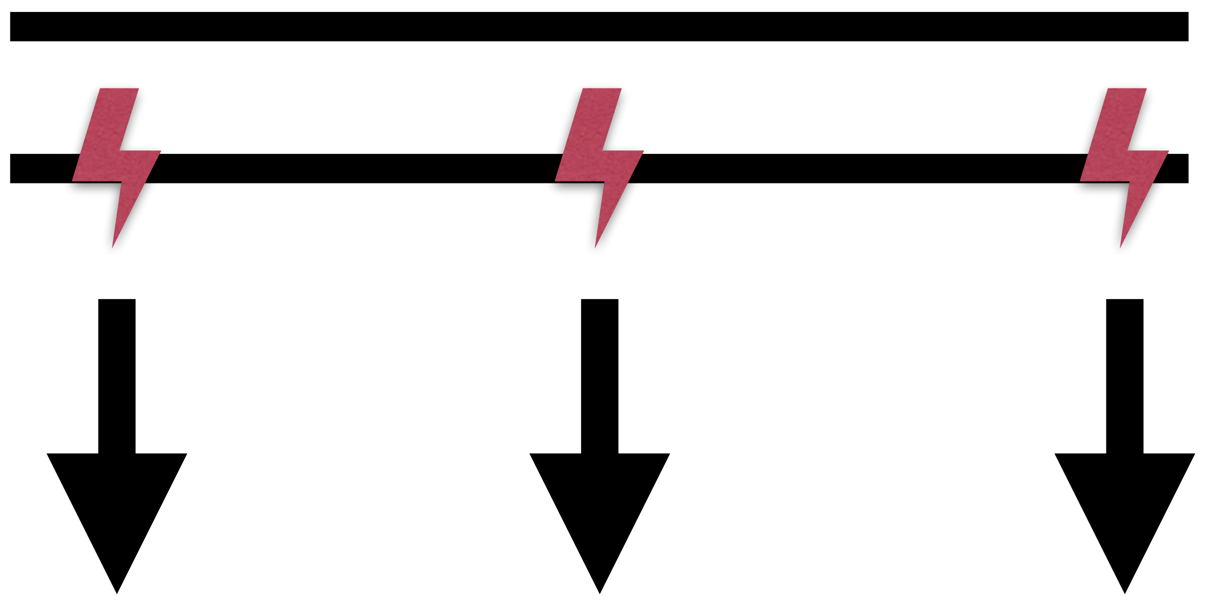


Step3: Infer demographic model





Fit-dadi: models of demographic history and natural selection

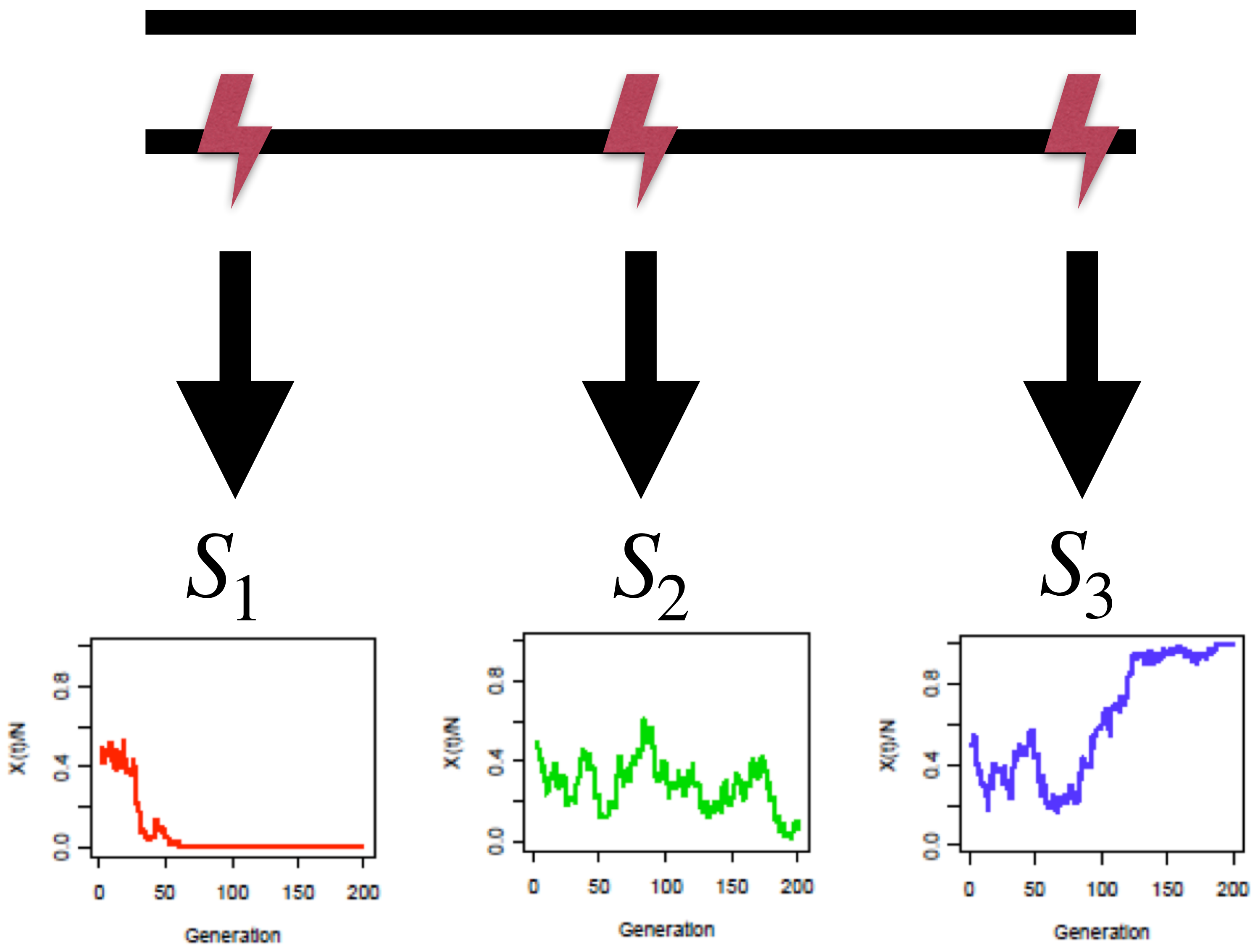


Sequence with L bases

mutations~Poisson(λ=2$N\mu L$)

Gutenkunst et al., 2009; Kim et al., 2017

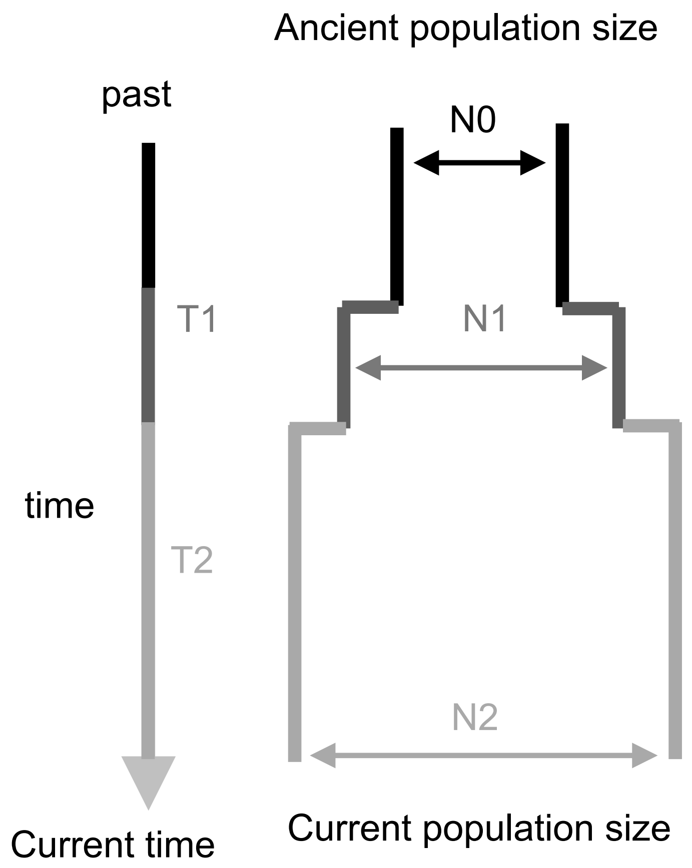
Fit-dadi: models of demographic history and natural selection



Sequence with L bases

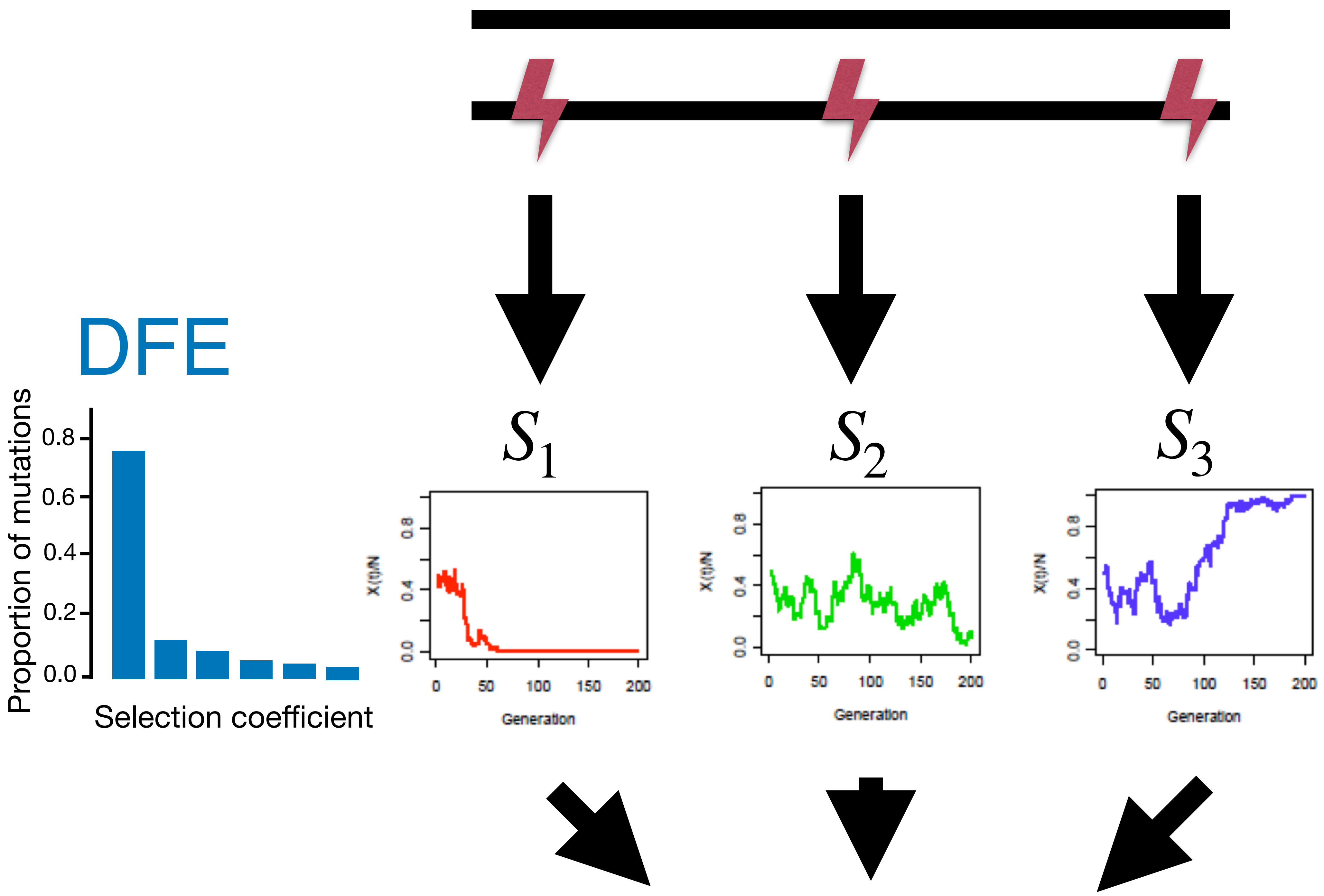
mutations~Poisson($\lambda=2N\mu L$)

Frequency of each mutation follows the: Wright-Fisher and inferred demographic model



Gutenkunst et al., 2009; Kim et al., 2017

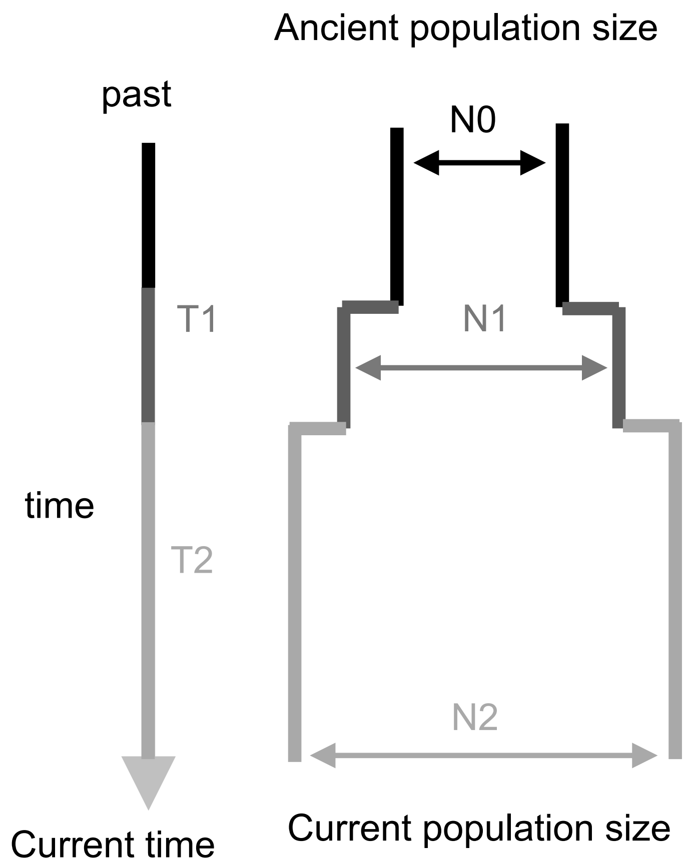
Fit-dadi: models of demographic history and natural selection



Sequence with *L* bases

mutations~Poisson($\lambda=2N\mu L$)

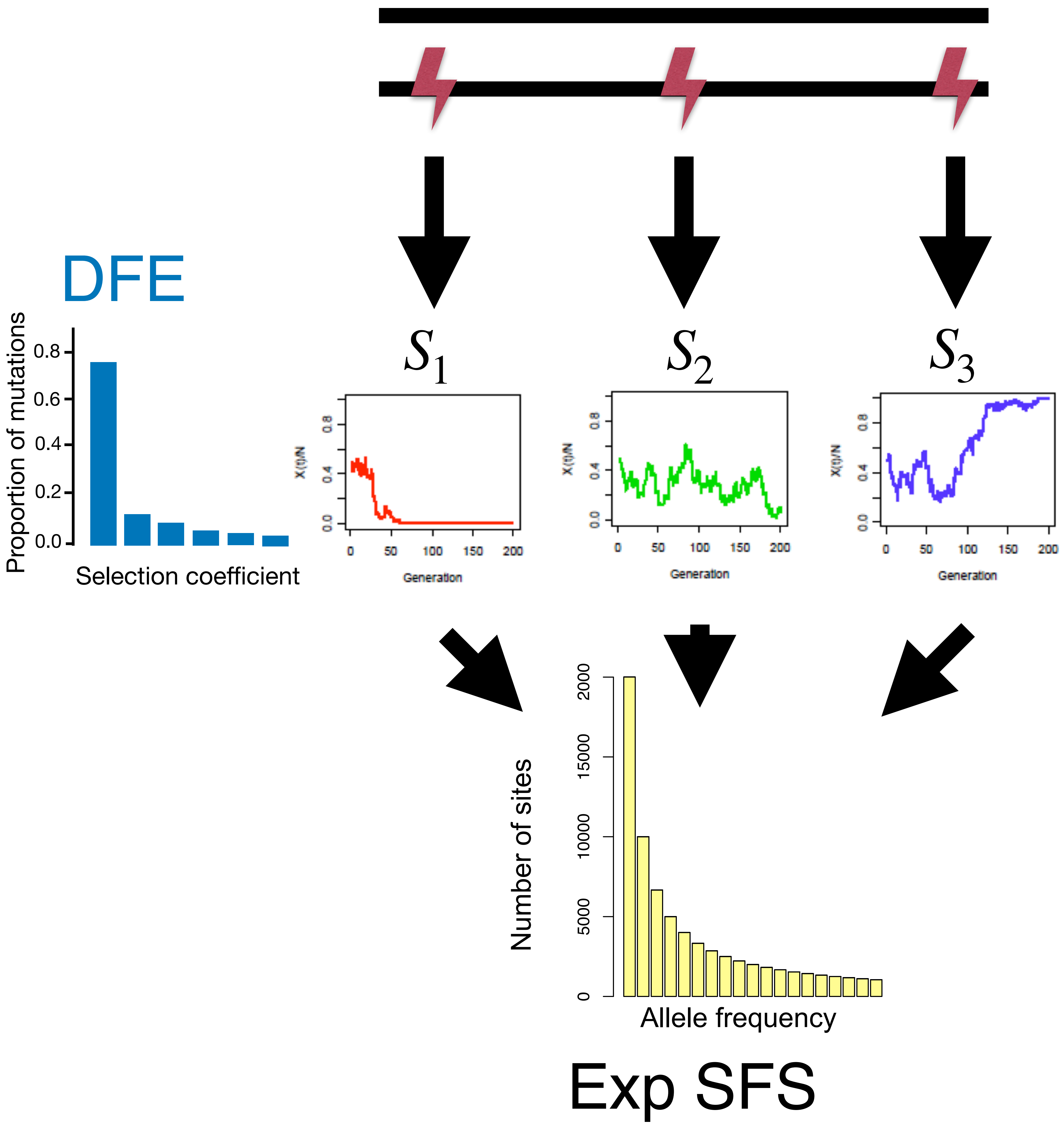
Frequency of each mutation follows the: Wright-Fisher and inferred demographic model



Parameter: Distribution of Fitness effects

Gutenkunst et al., 2009; Kim et al., 2017

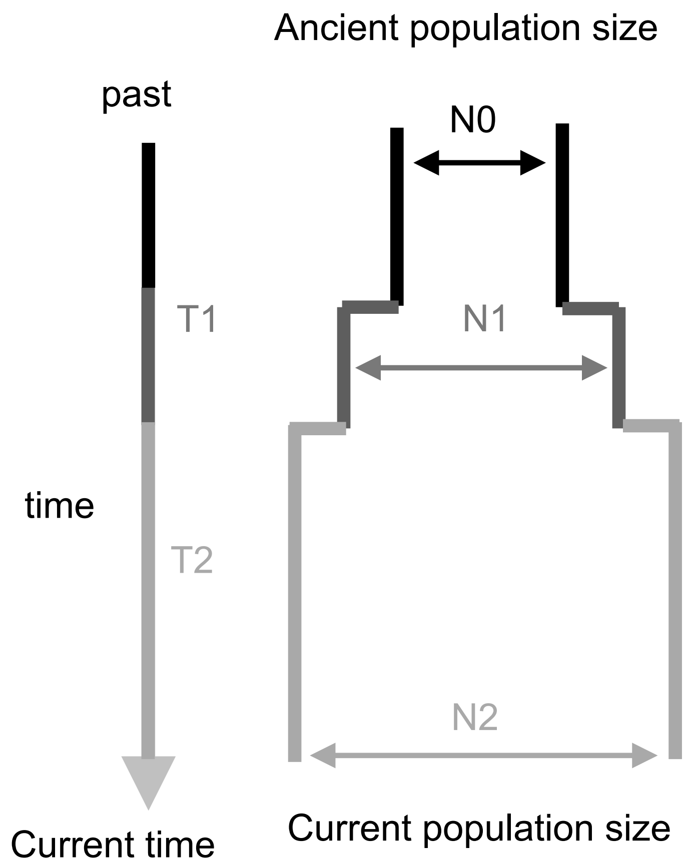
Fit-dadi: models of demographic history and natural selection



Sequence with *L* bases

mutations~Poisson($\lambda=2N\mu L$)

Frequency of each mutation follows the: Wright-Fisher and inferred demographic model

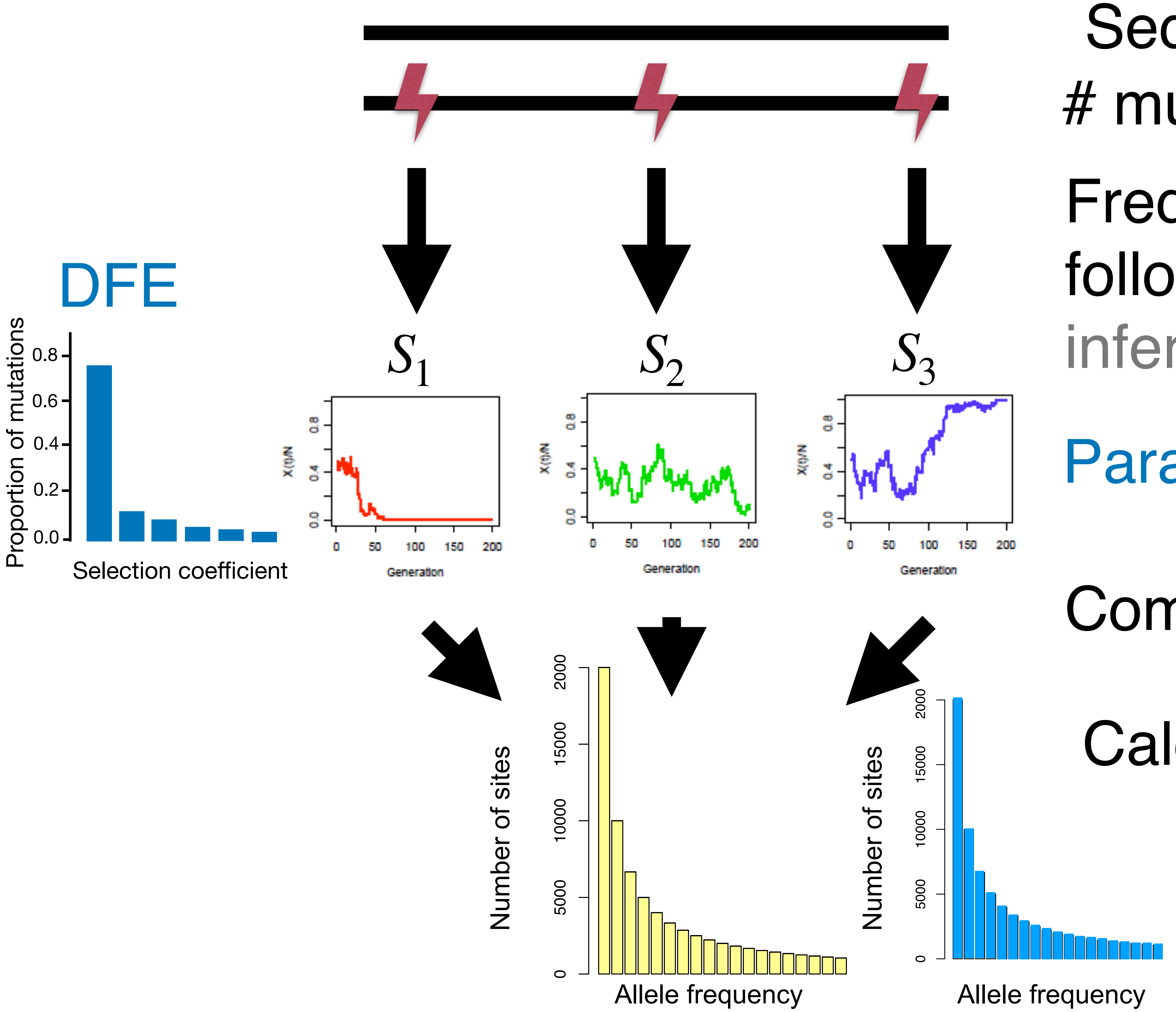


Parameter: Distribution of Fitness effects

Combine mutations to make SFS

Gutenkunst et al., 2009; Kim et al., 2017

dadi: models of demographic history and natural selection



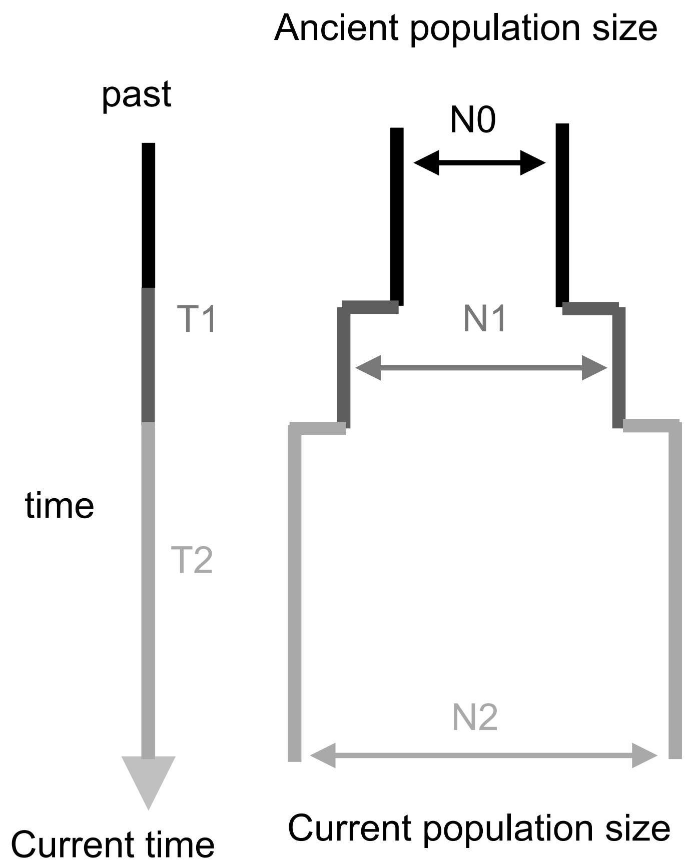
Exp SFS

Obs SFS from non-coding region

Sequence with *L* bases

\# mutations~Poisson($\lambda=2N\mu L$)

Frequency of each mutation follows the: Wright-Fisher and inferred demographic model



Parameter: Distribution of Fitness effects

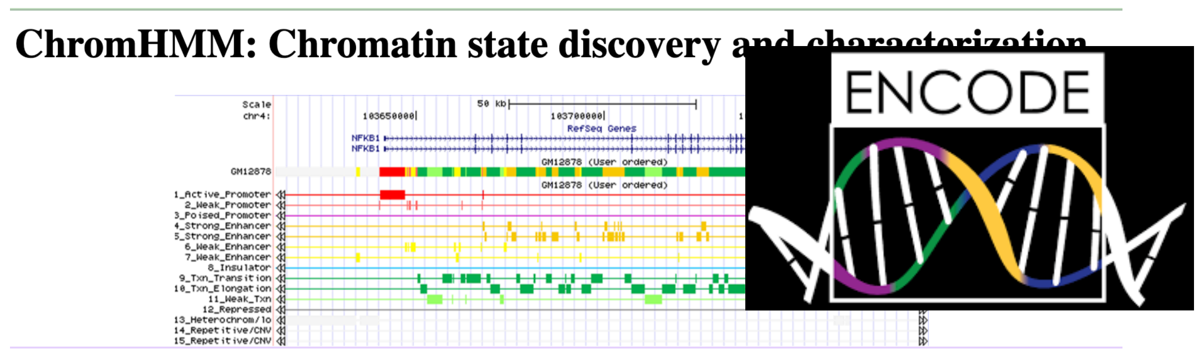
Combine mutations to make SFS

Calculate likelihood of models

Workflow

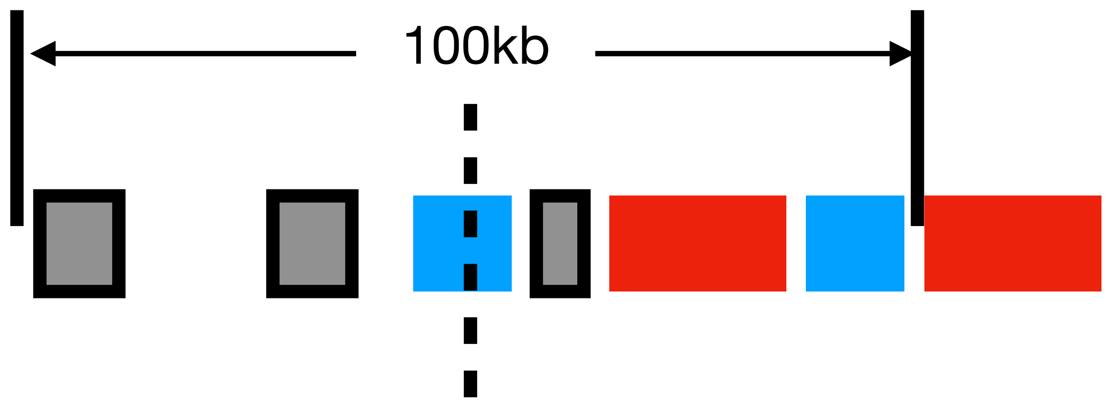
Step1: Annotate genomic regions

🟦 Target non-coding region

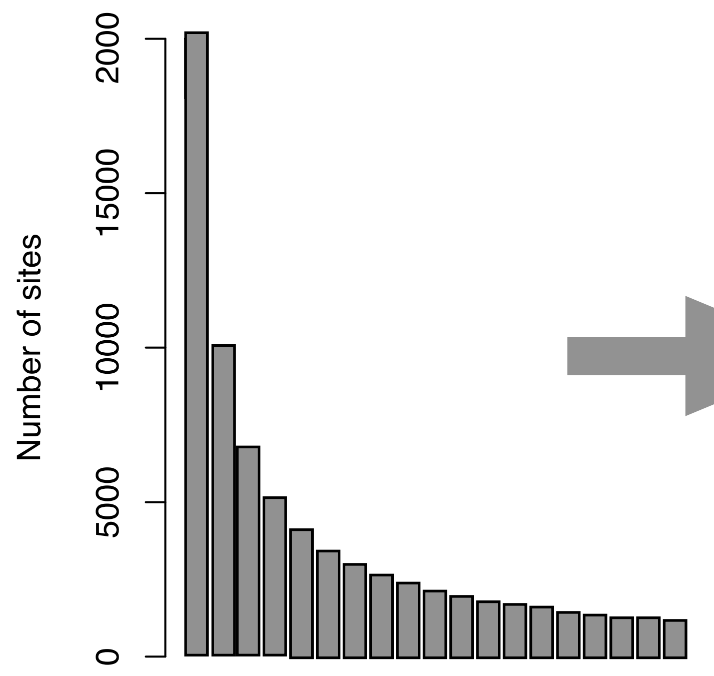


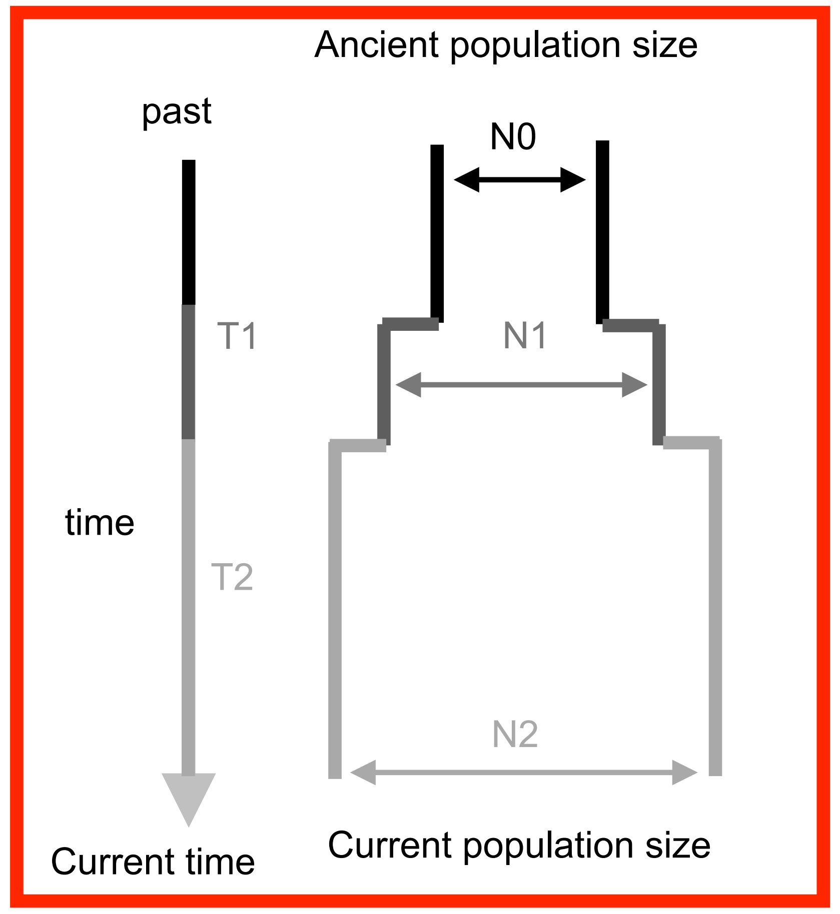
Step2: Find quiescent control

▨ "neutral control": quiescent

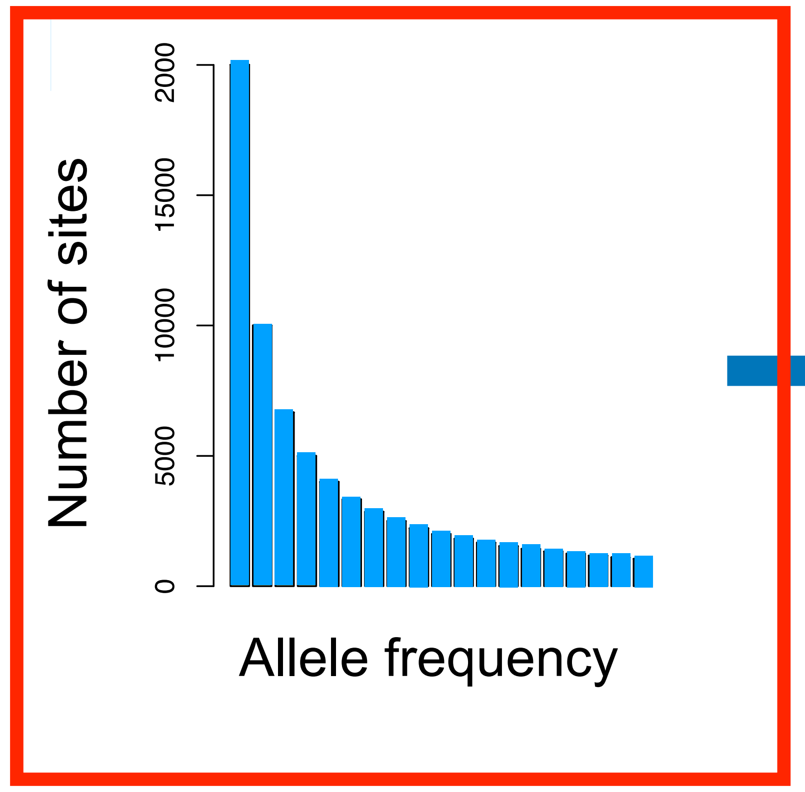


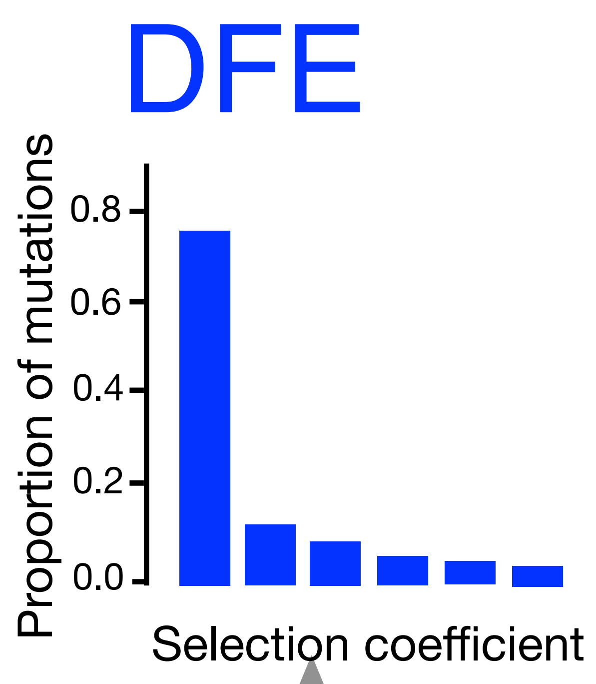
Step3: Infer demographic model



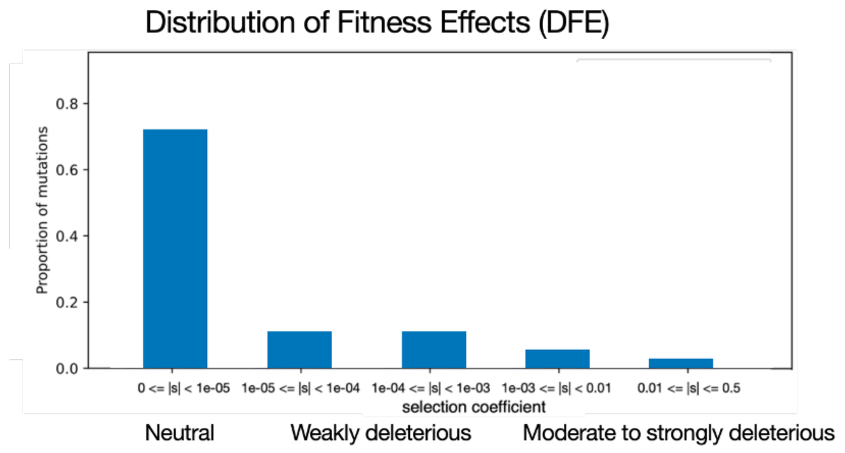


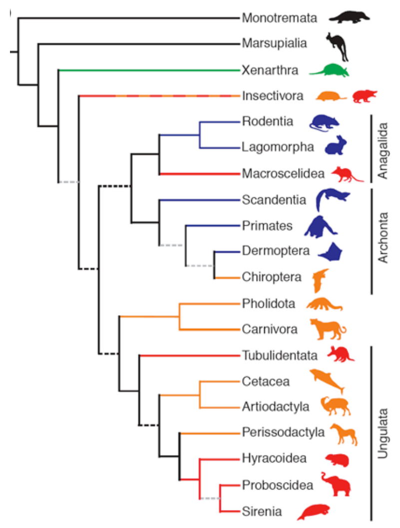
Step4: Infer Distribution of Fitness Effecs (DFE)





DFE of non-coding mutations

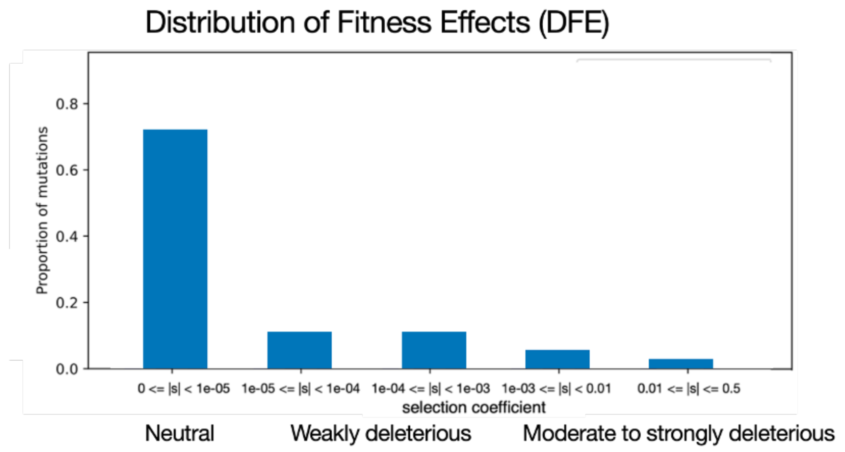


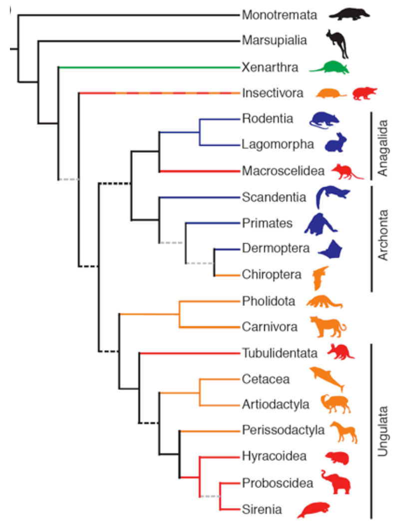


- Model-based method for inferring Distribution of Fitness effects (DFE)
- DFE of annotated functional non-coding regions in human genome
- Negative selection in conserved and non-conserved human genomic region

picture is from https://www.phonandroid.com/tesla-continue-de-decliner-le-cybertruck-a-toutes-les-sauces-avec-un-marteau-de-gym.html
Phylogeny is from Springer et al., 2004

DFE of non-coding mutations

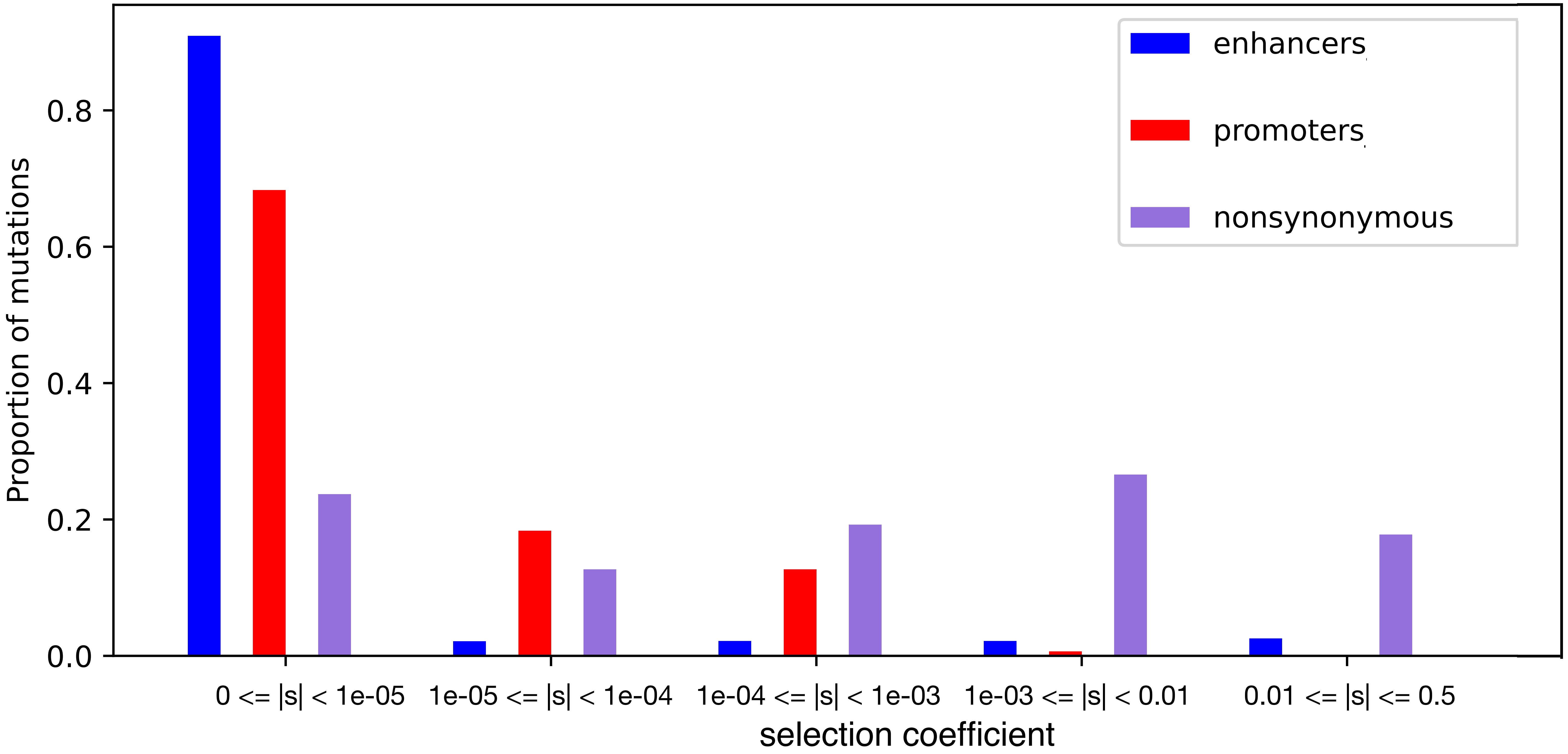




- Model-based method for inferring Distribution of Fitness effects (DFE)
- DFE of annotated functional non-coding regions in human genome
- Negative selection in conserved and non-conserved human genomic region

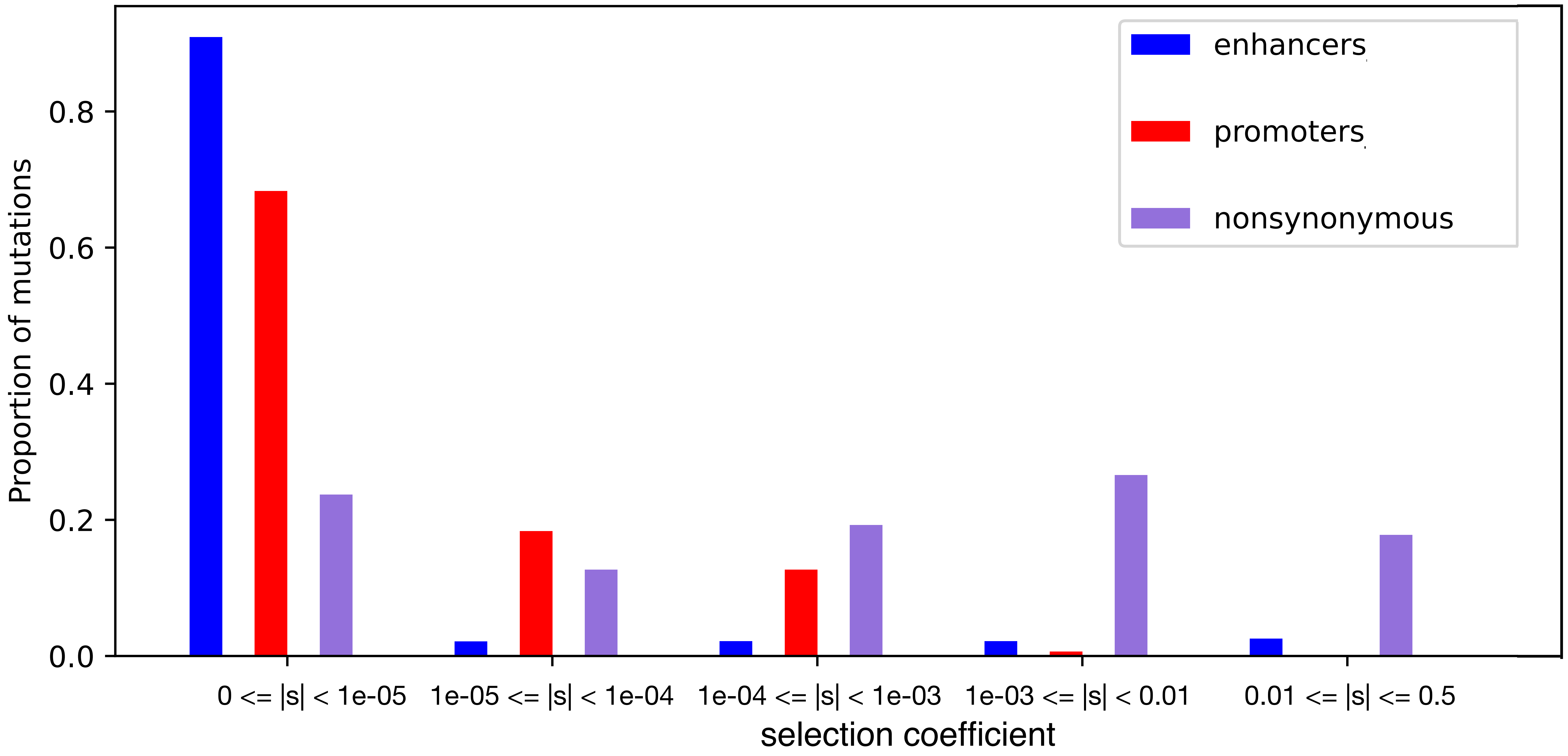
picture is from https://www.phonandroid.com/tesla-continue-de-decliner-le-cybertruck-a-toutes-les-sauces-avec-un-marteau-de-gym.html
Phylogeny is from Springer et al., 2004

DFE of enhancers and promoters



DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

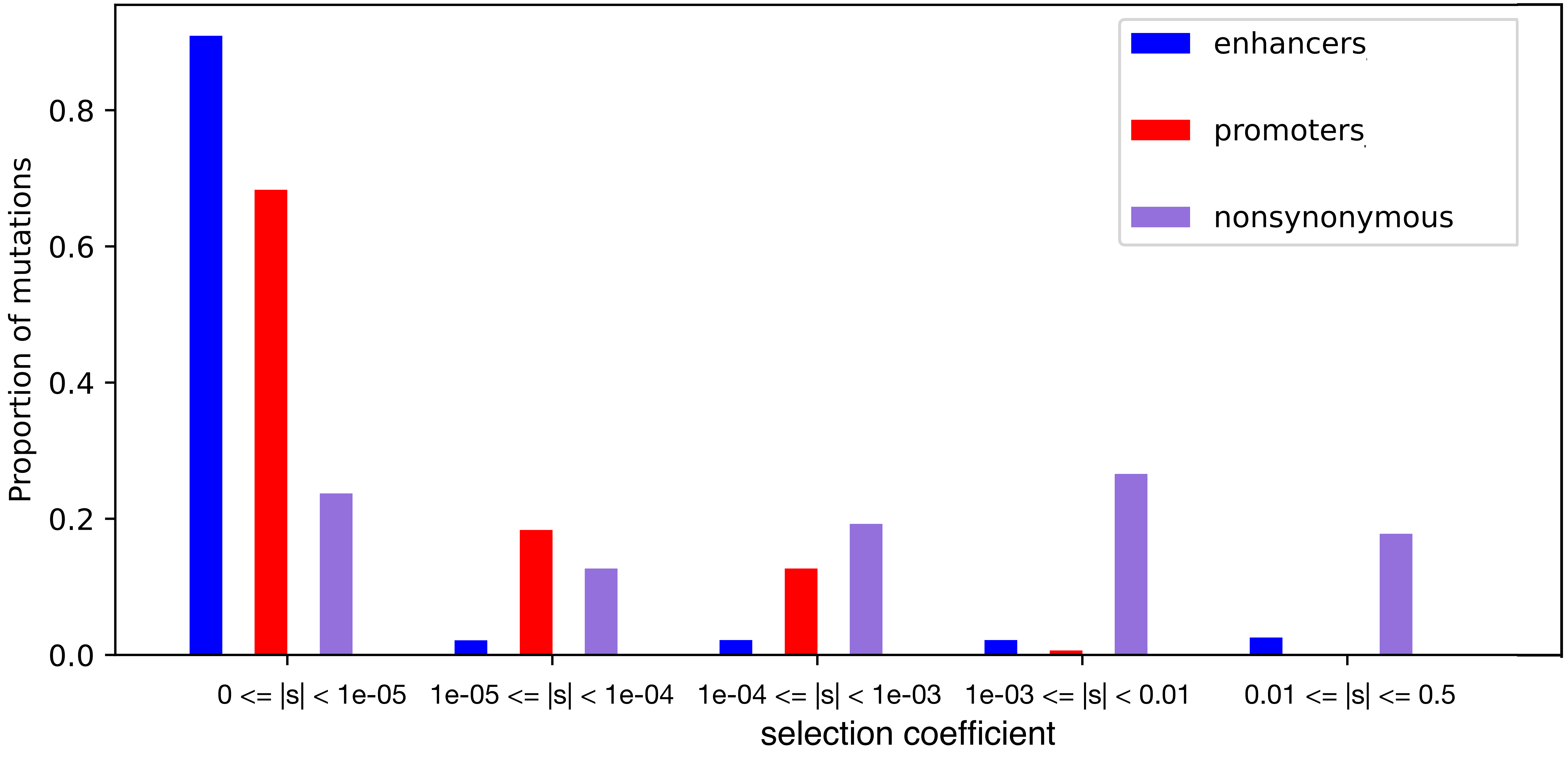
DFE of enhancers and promoters



Neutral

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of enhancers and promoters

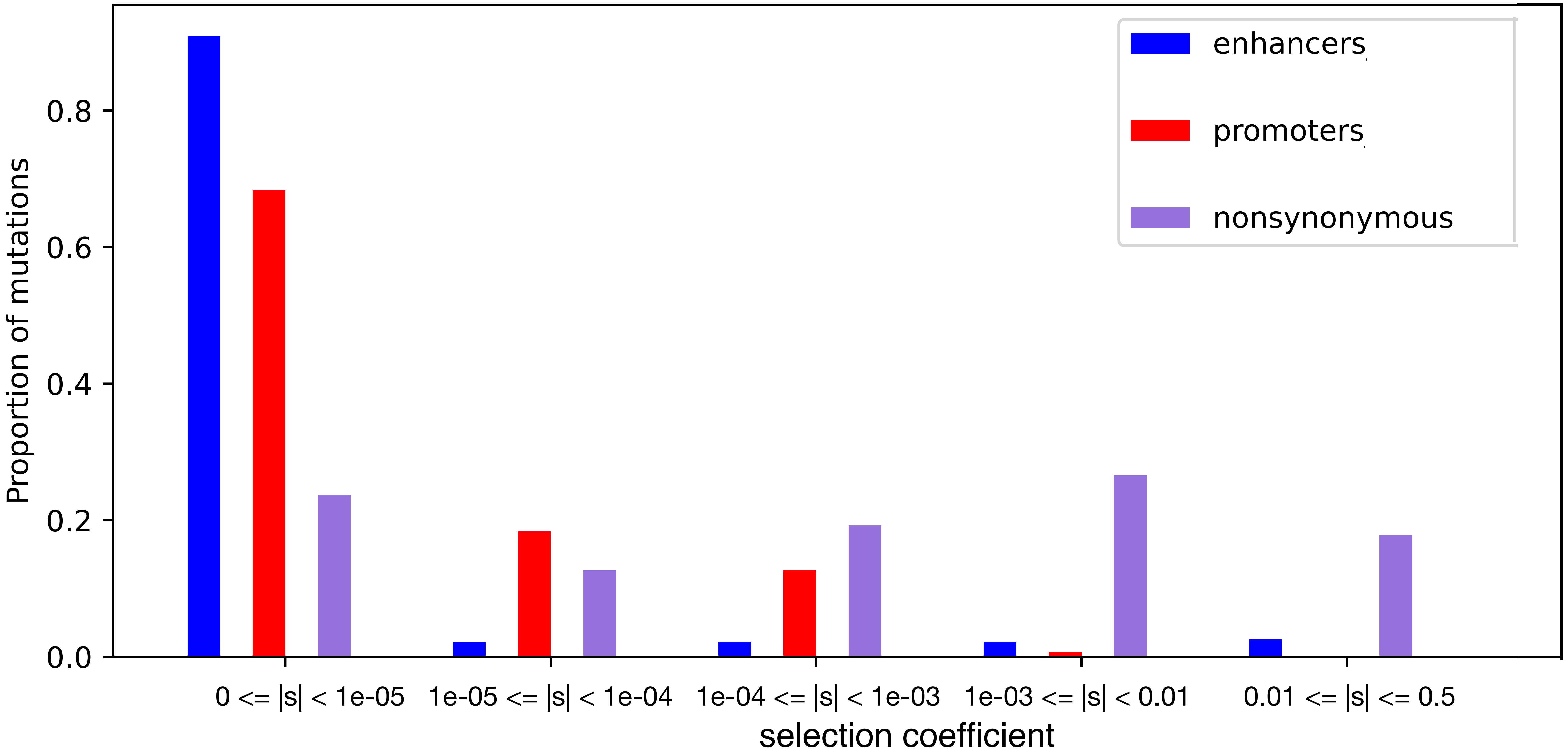


Neutral

enhancers > promoters

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of enhancers and promoters



Neutral

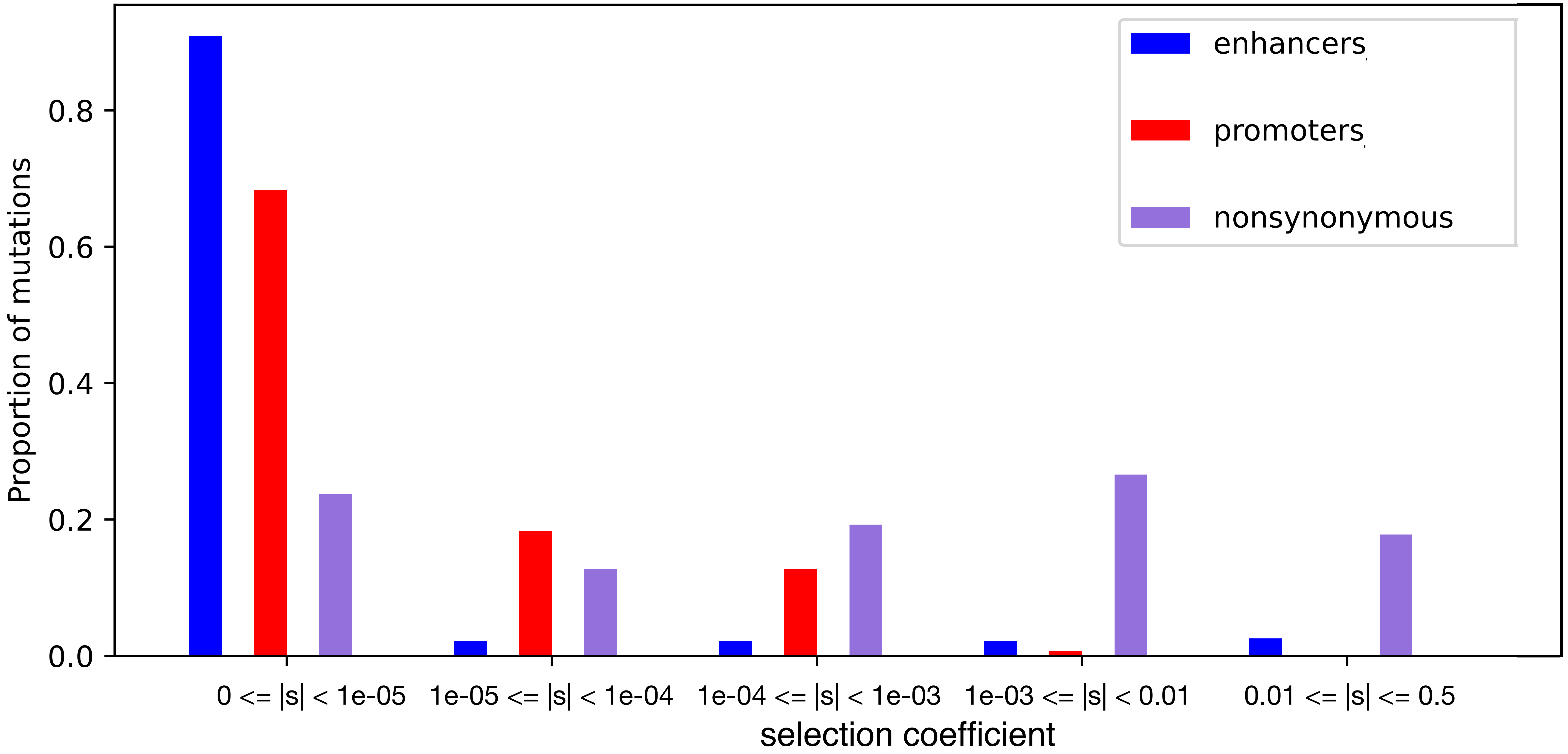
enhancers > promoters

Weakly deleterious

promoters > enhancers

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of enhancers and promoters



Neutral

enhancers > promoters

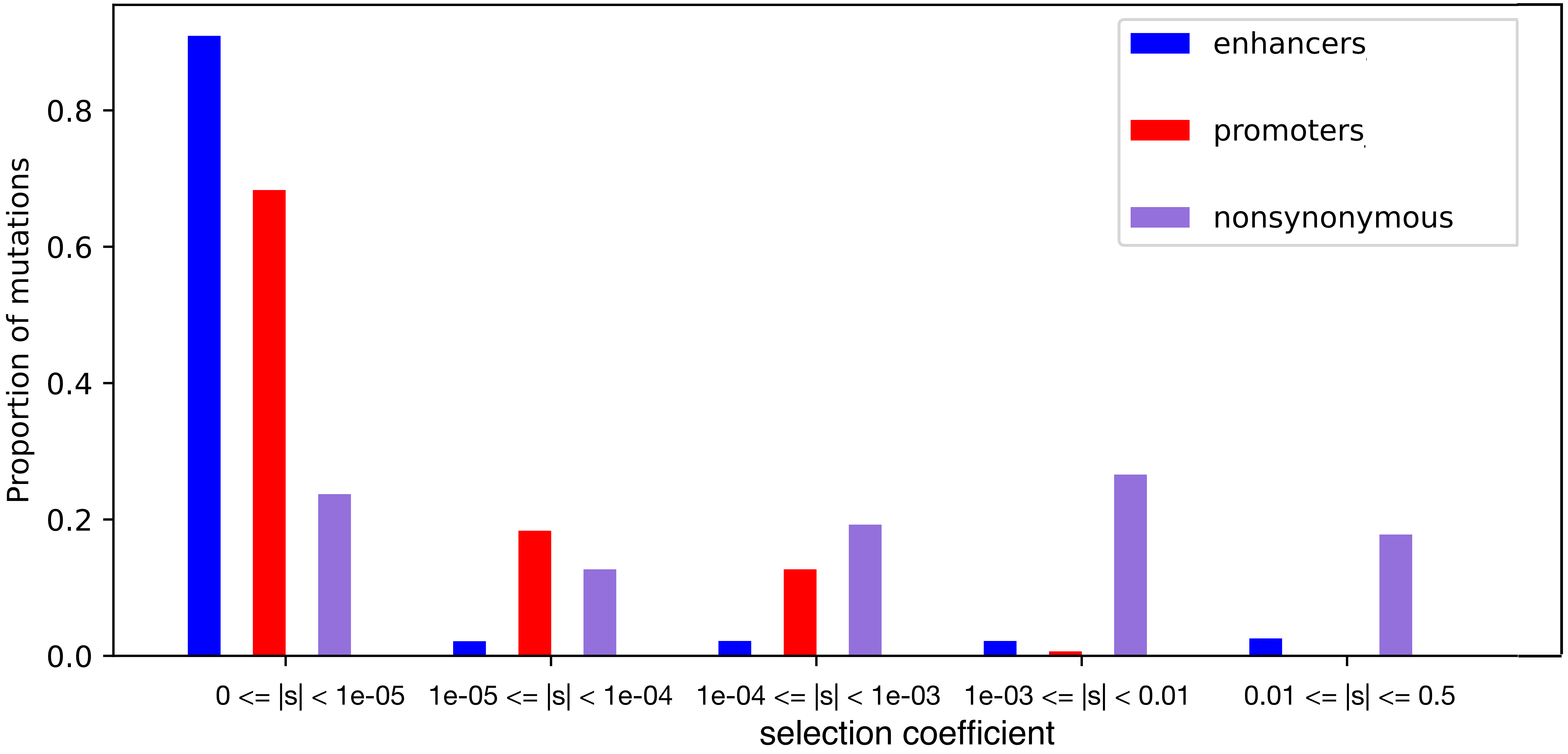
Weakly deleterious

promoters > enhancers

Moderate to strongly deleterious

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of enhancers and promoters



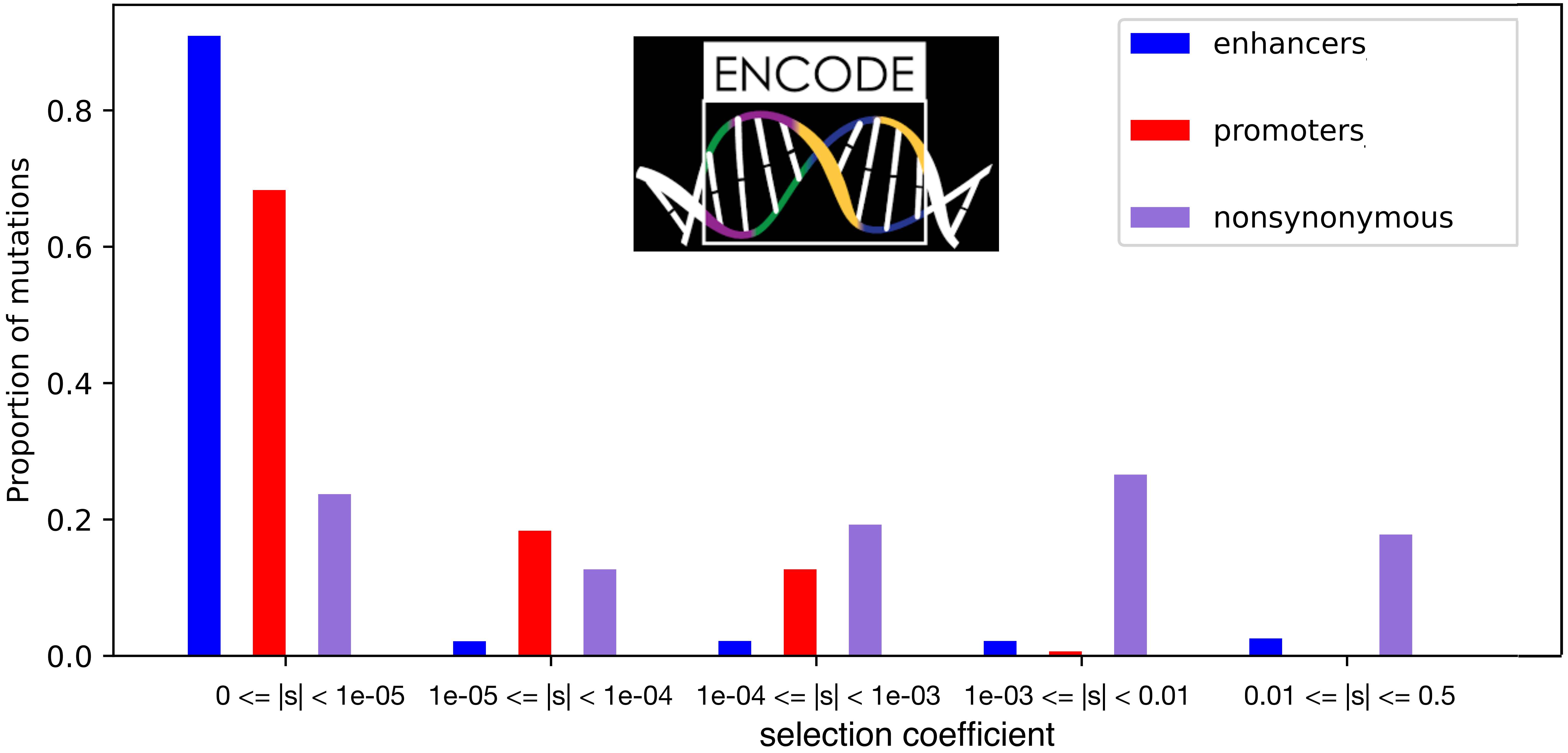
Neutral
enhancers > promoters

Weakly deleterious
promoters > enhancers

Moderate to strongly deleterious
small proportion

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

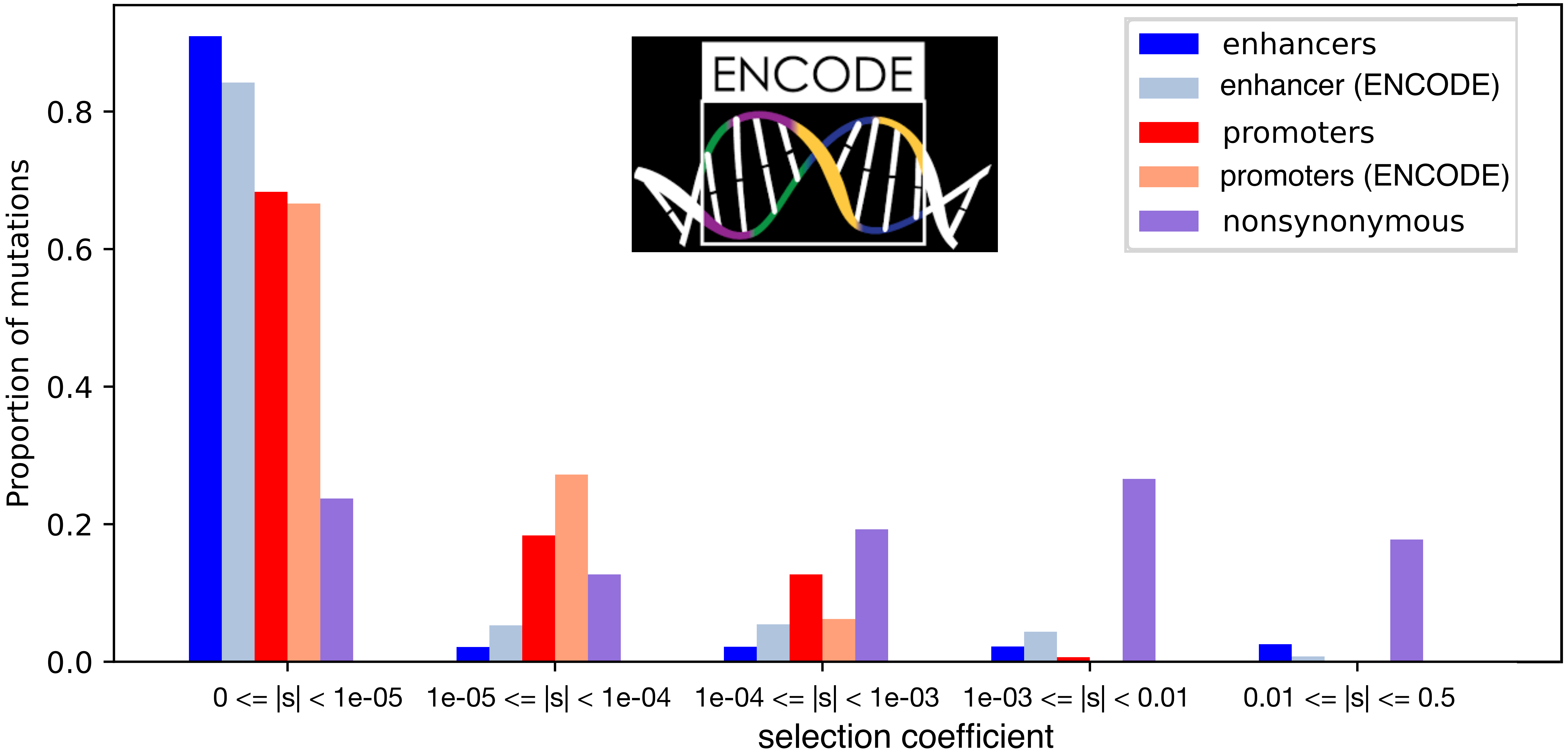
DFE of enhancers and promoters



Neutral	Weakly deleterious	Moderate to strongly deleterious
enhancers > promoters	promoters > enhancers	small proportion

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

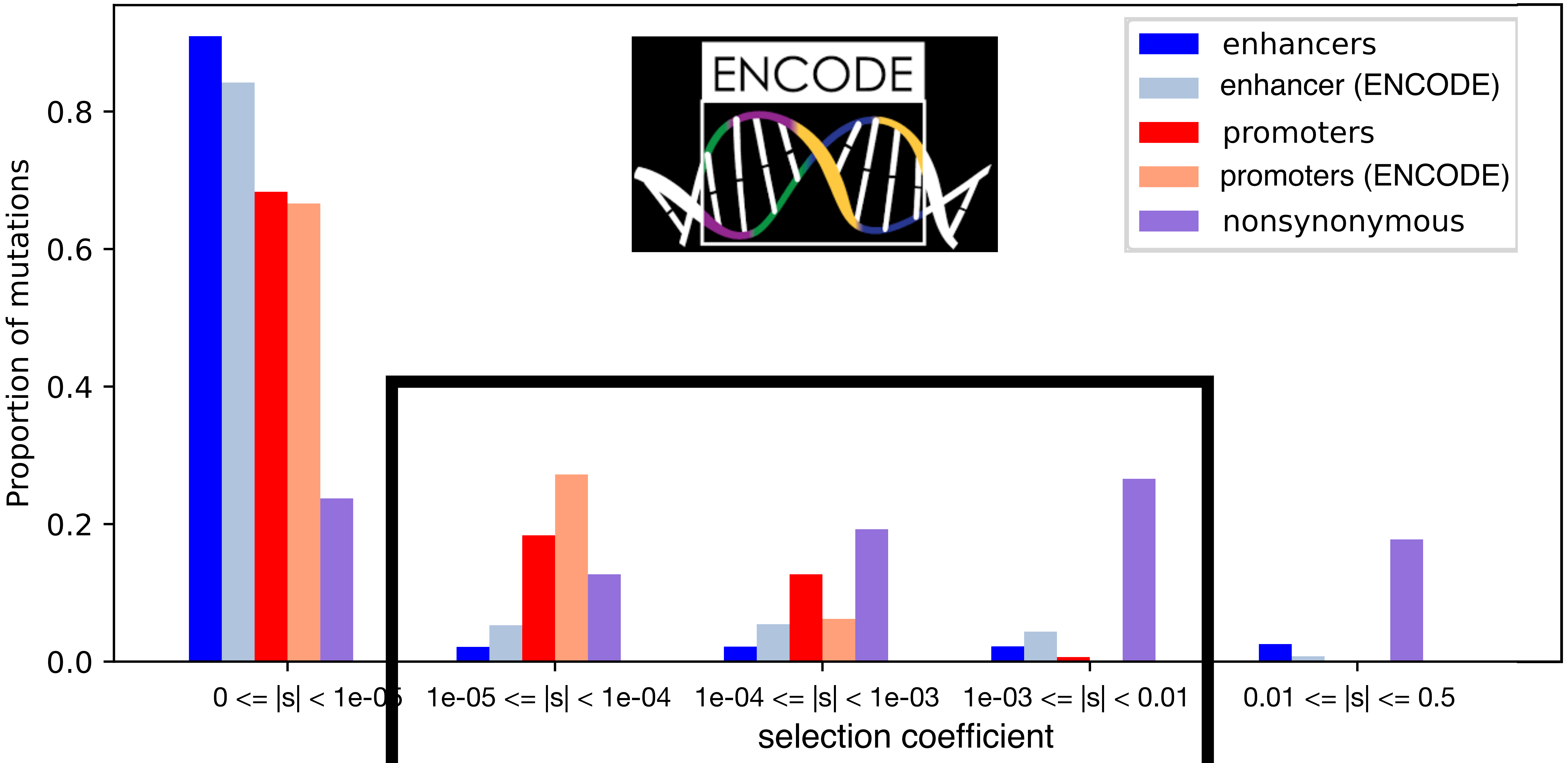
DFE of enhancers and promoters



Neutral	Weakly deleterious	Moderate to strongly deleterious
enhancers > promoters	promoters > enhancers	small proportion

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of enhancers and promoters



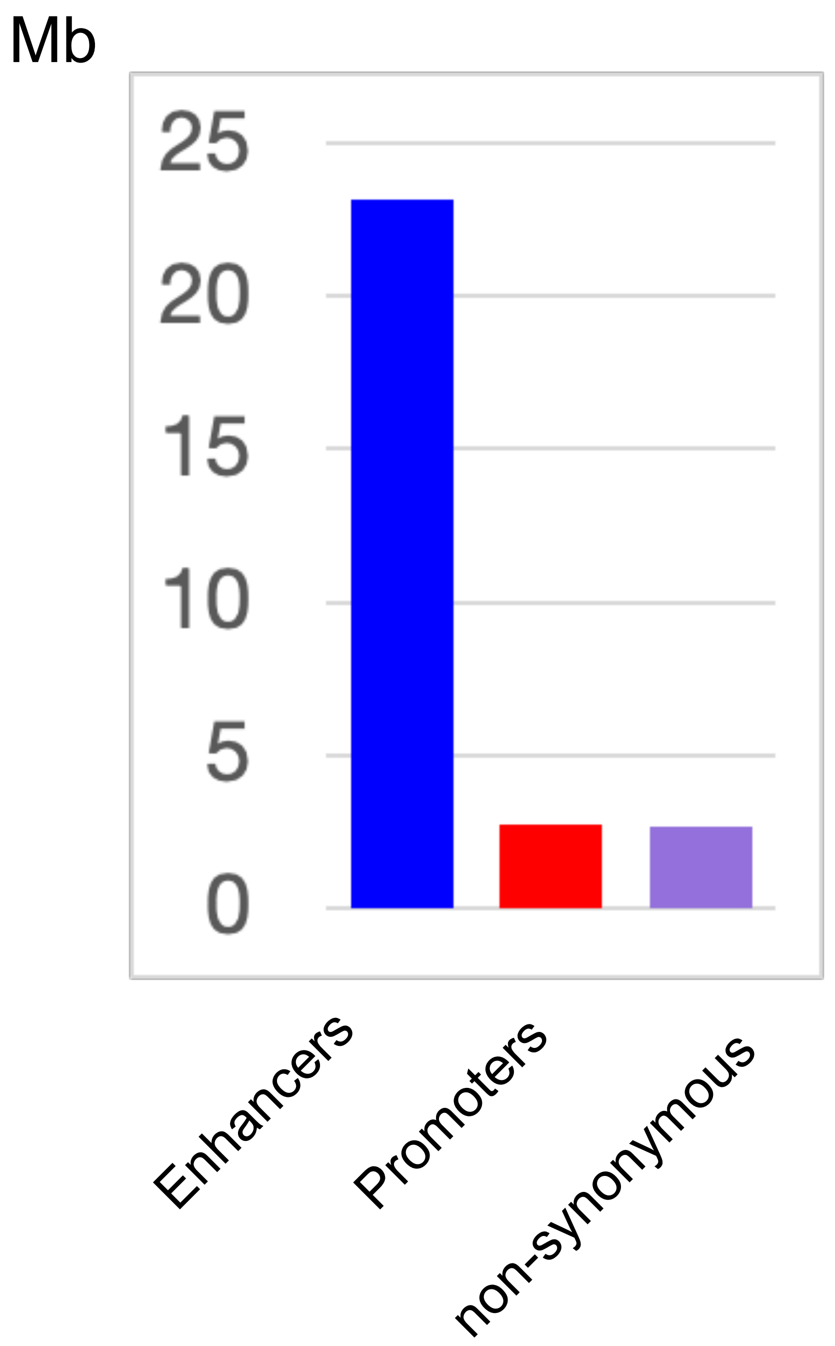
Neutral Weakly deleterious Moderate to strongly deleterious

enhancers > promoters promoters > enhancers small proportion

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

Large number of deleterious mutations

number of sites



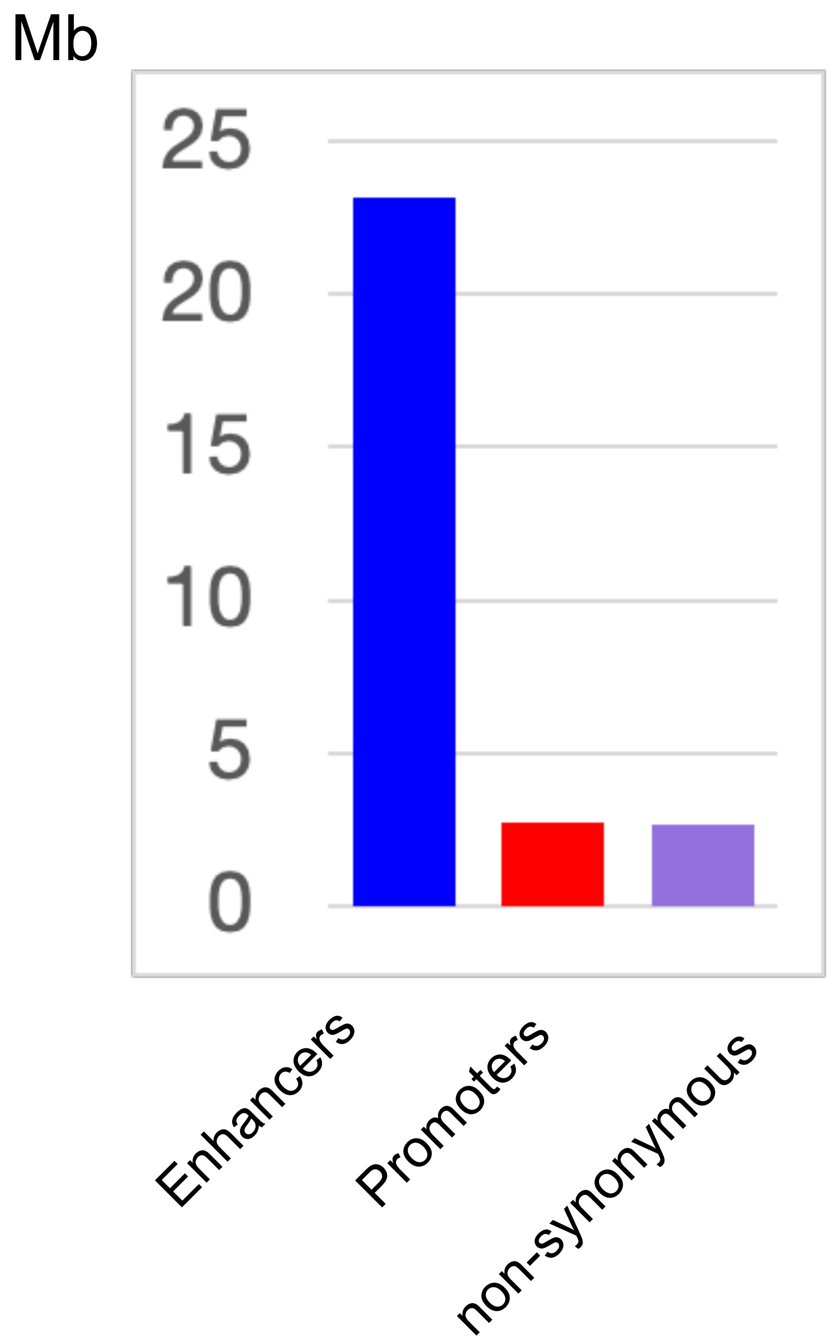
Large number of deleterious mutations

of mutations

$= L \times \mu$

L ↓ number of sites

μ ↓ mutation rate



Large number of deleterious mutations

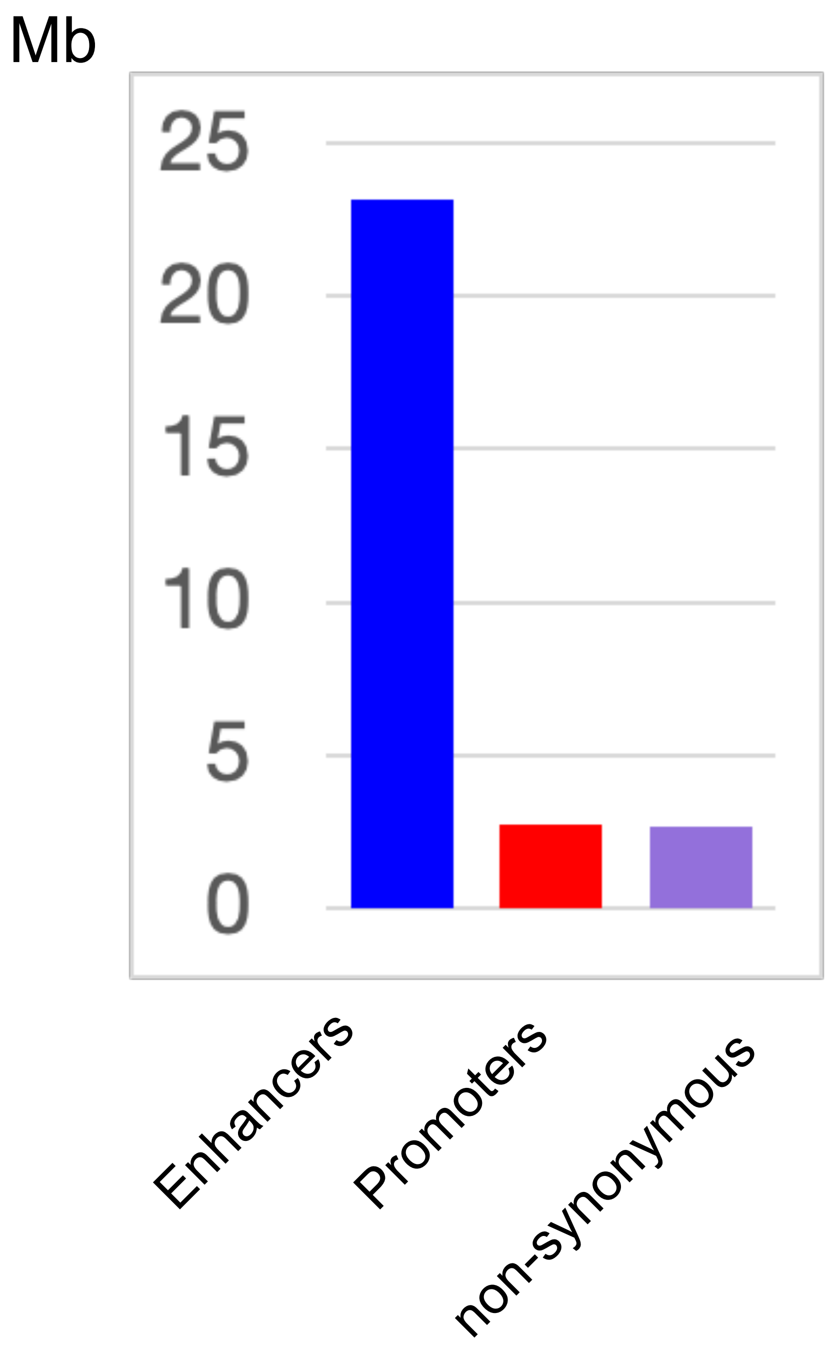
of mutations with given S

$$= L \times \mu \times P(s)$$

L ↓ number of sites

μ ↓ mutation rate

$P(s)$ ↓ probability of mutations with selection coefficient *s*



Large number of deleterious mutations

of mutations with given S

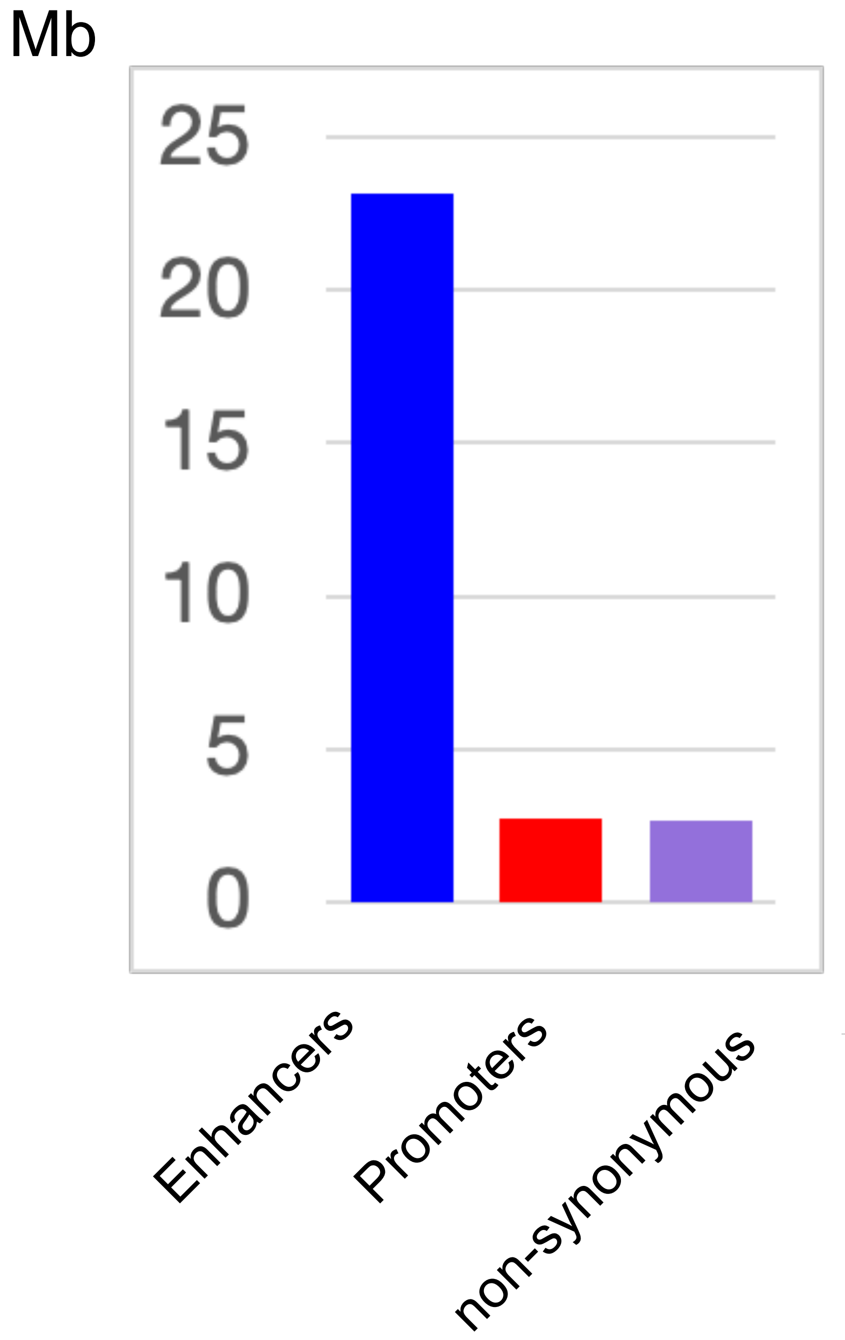
= L × μ × $P(s)$

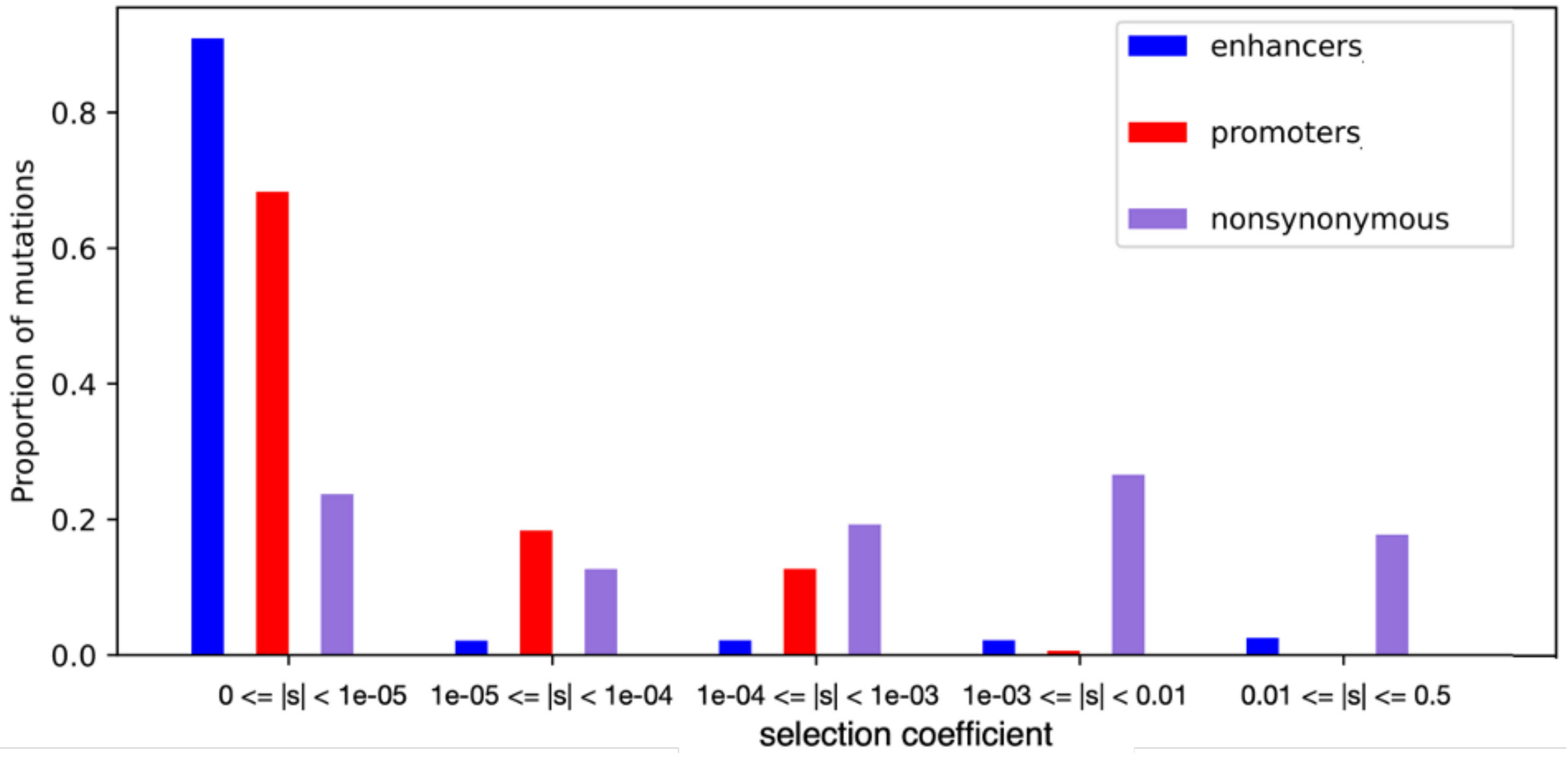
L ↓ number of sites

μ ↓ mutation rate

$P(s)$ ↓ probability of mutations with selection coefficient s

Distribution of Fitness Effects (DFE)





Large number of deleterious mutations

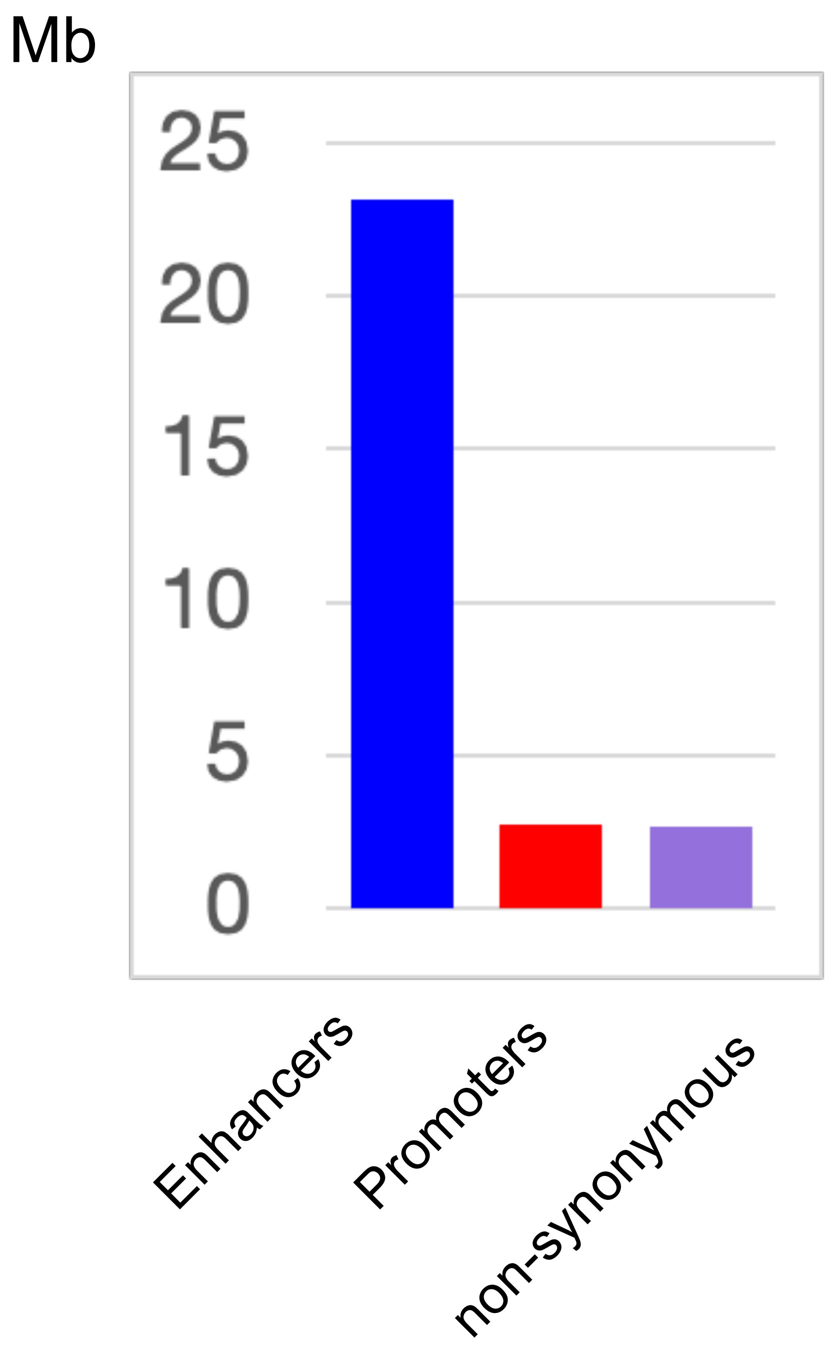
of mutations with given S

= L × μ × $P(s)$

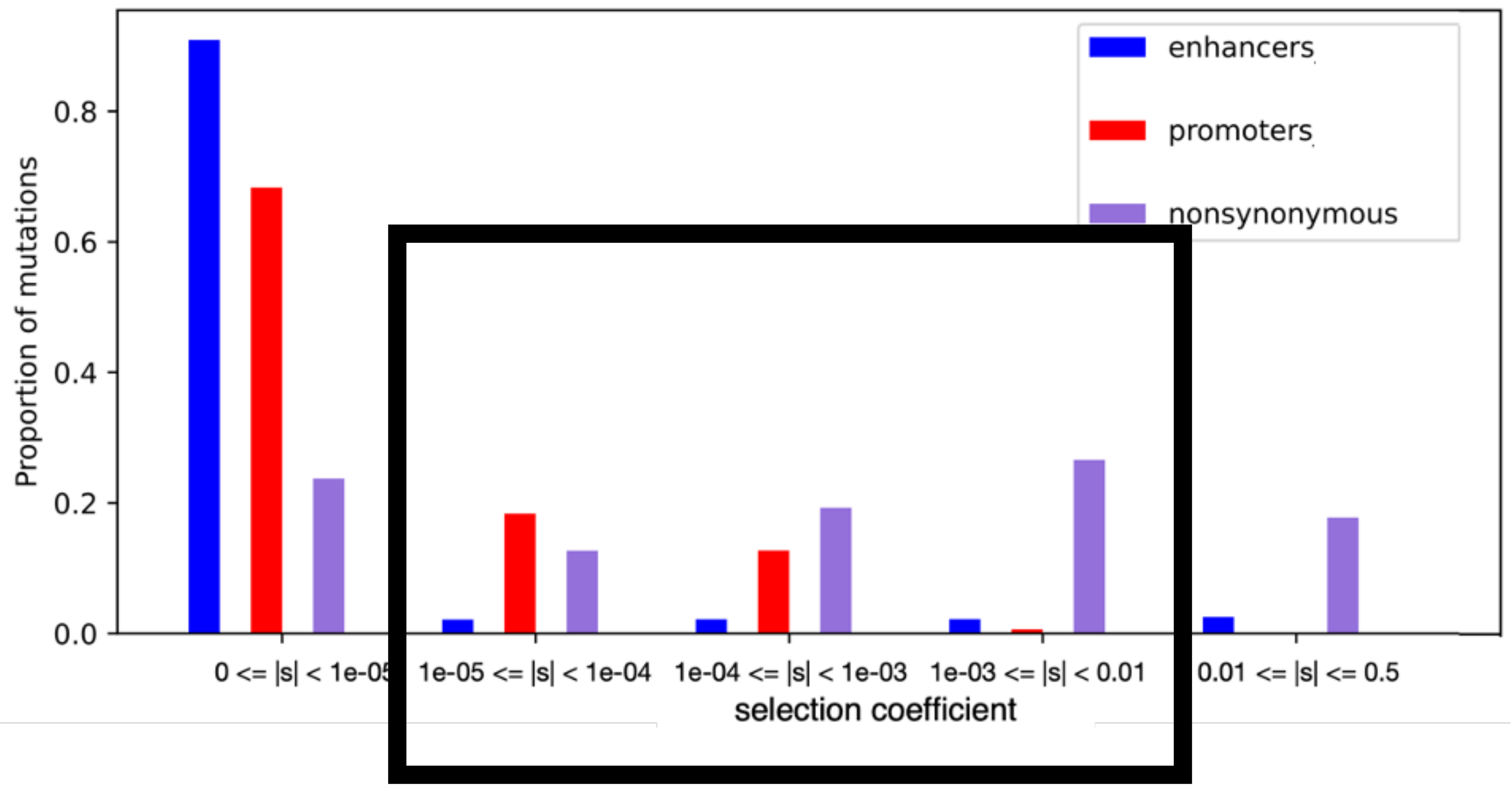
L ↓ number of sites

μ ↓ mutation rate

$P(s)$ ↓ probability of mutations with selection coefficient *s*



Distribution of Fitness Effects (DFE)



Weakly to moderately deleterious

Large number of deleterious mutations

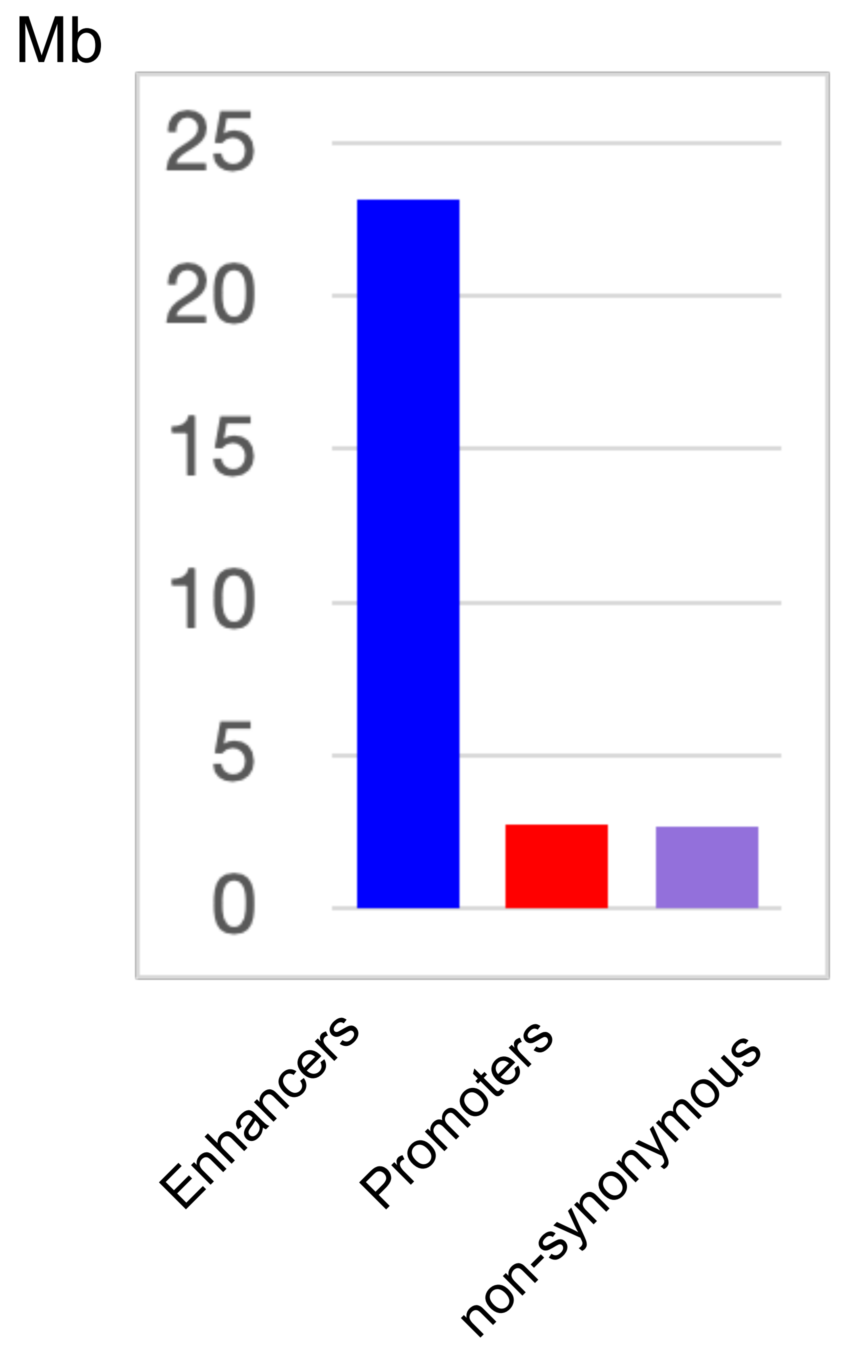
of mutations with given S

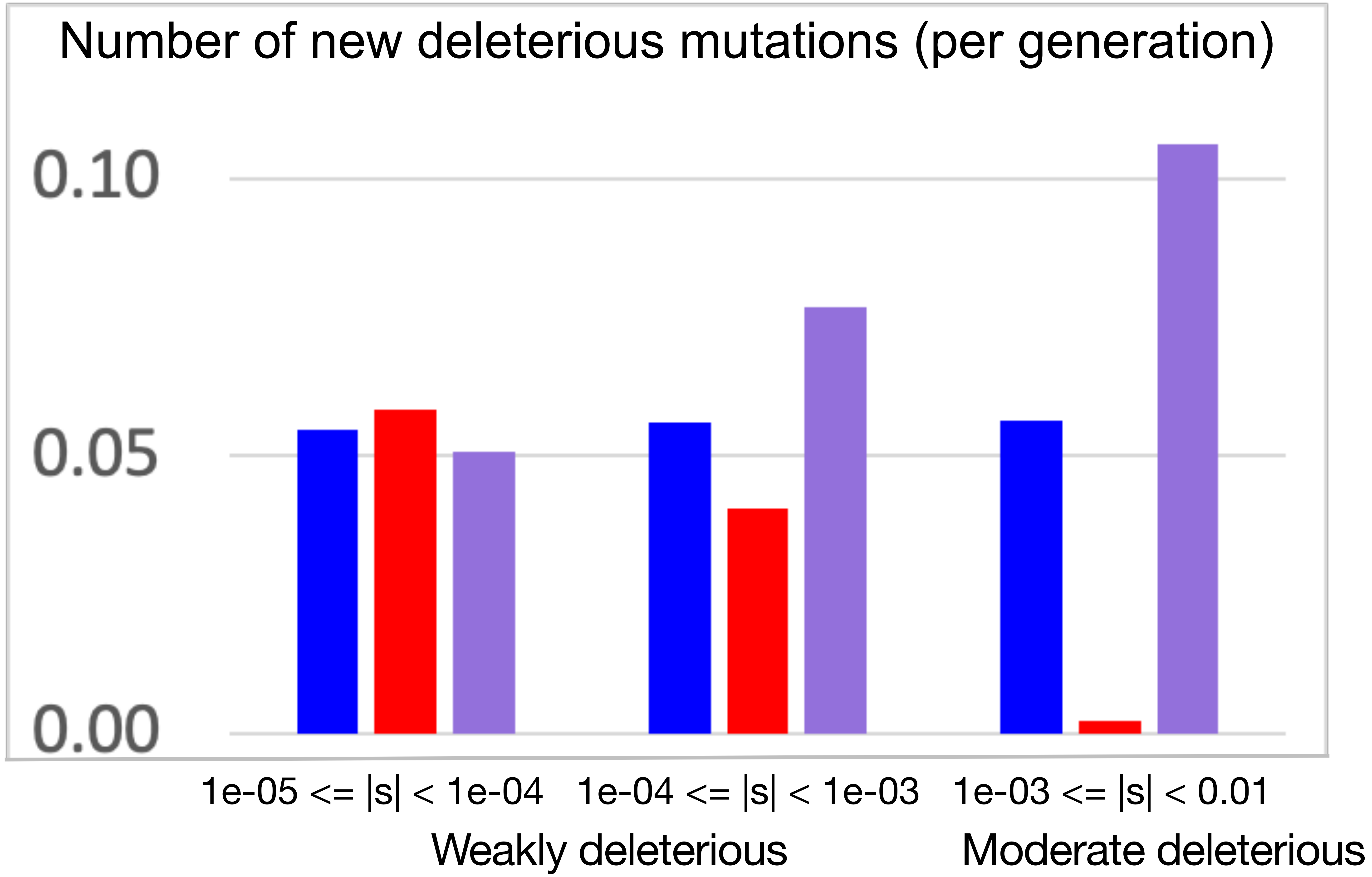
$= L \times \mu \times P(s)$

mutation rate

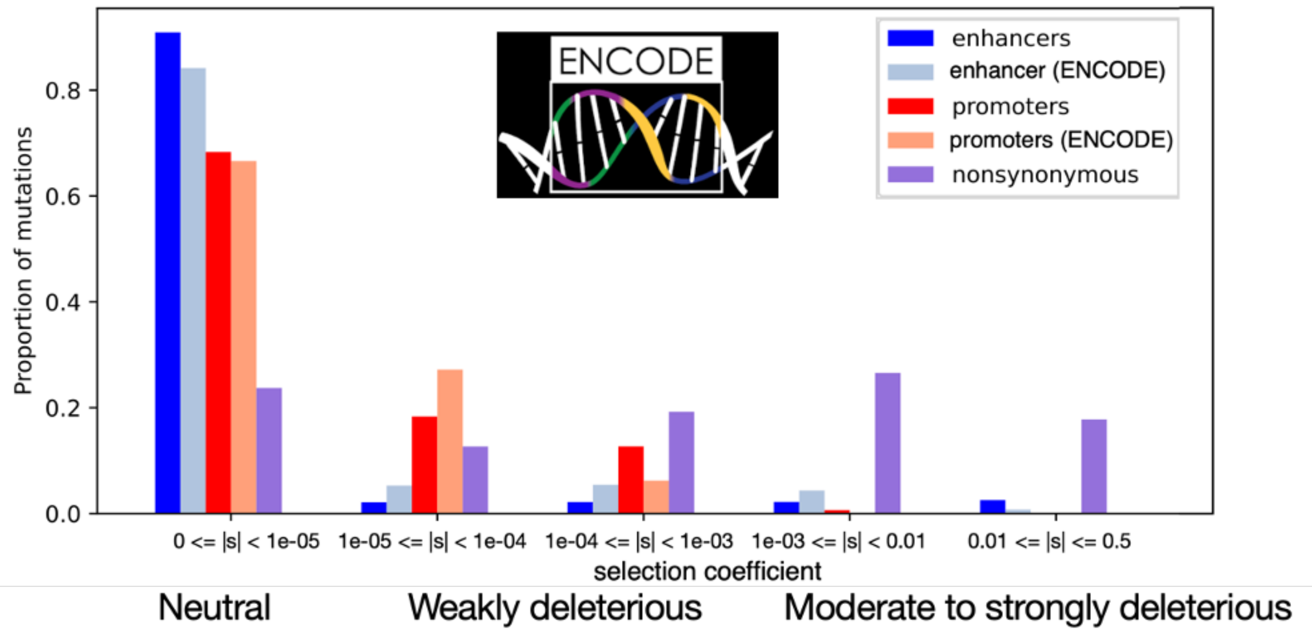
number of sites

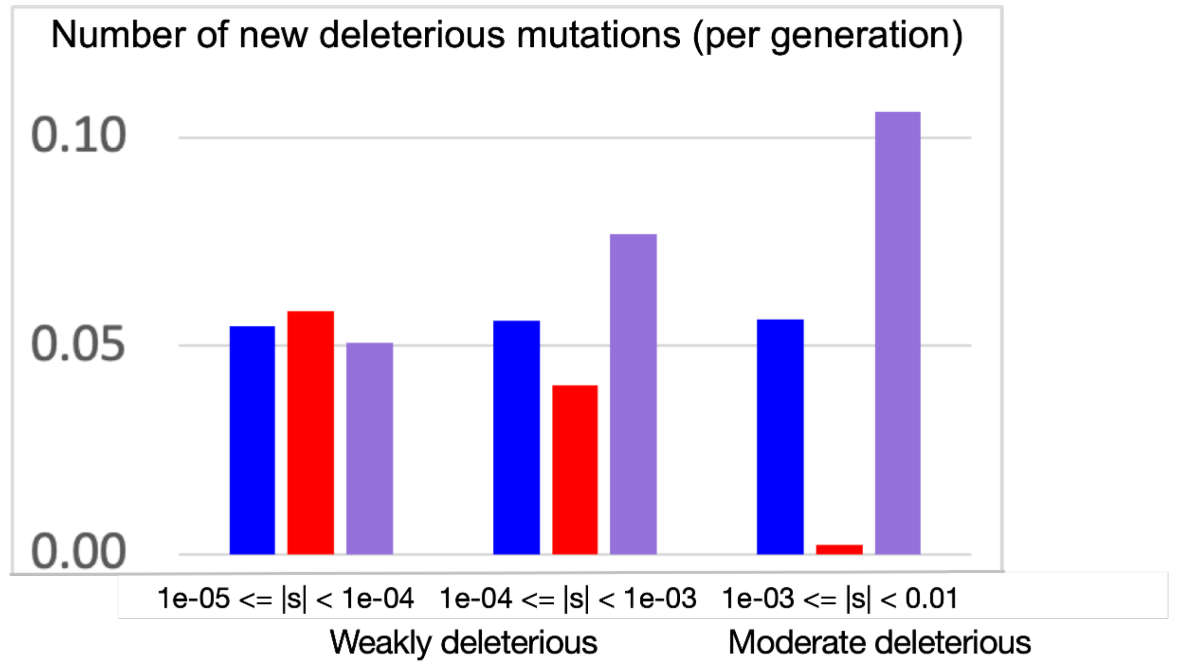
proba





Conclusion

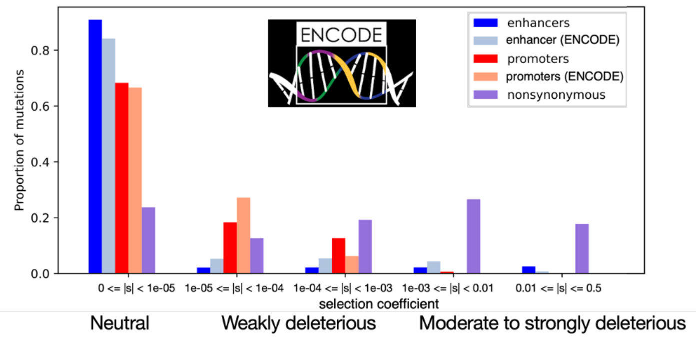




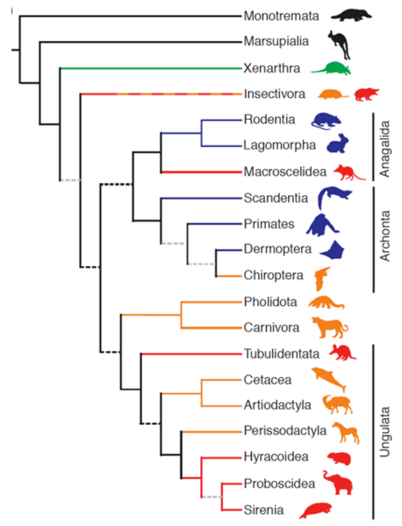
- Negative selection in annotated functional non-coding regions: fitness effects If mutations at enhancers, promoters and other annotated states
- Large number of weak to moderately deleterious mutations in enhancers and promoters that is comparable to non-synonymous sites

DFE of non-coding mutations

- Model-based method for inferring Distribution of Fitness effects (DFE)



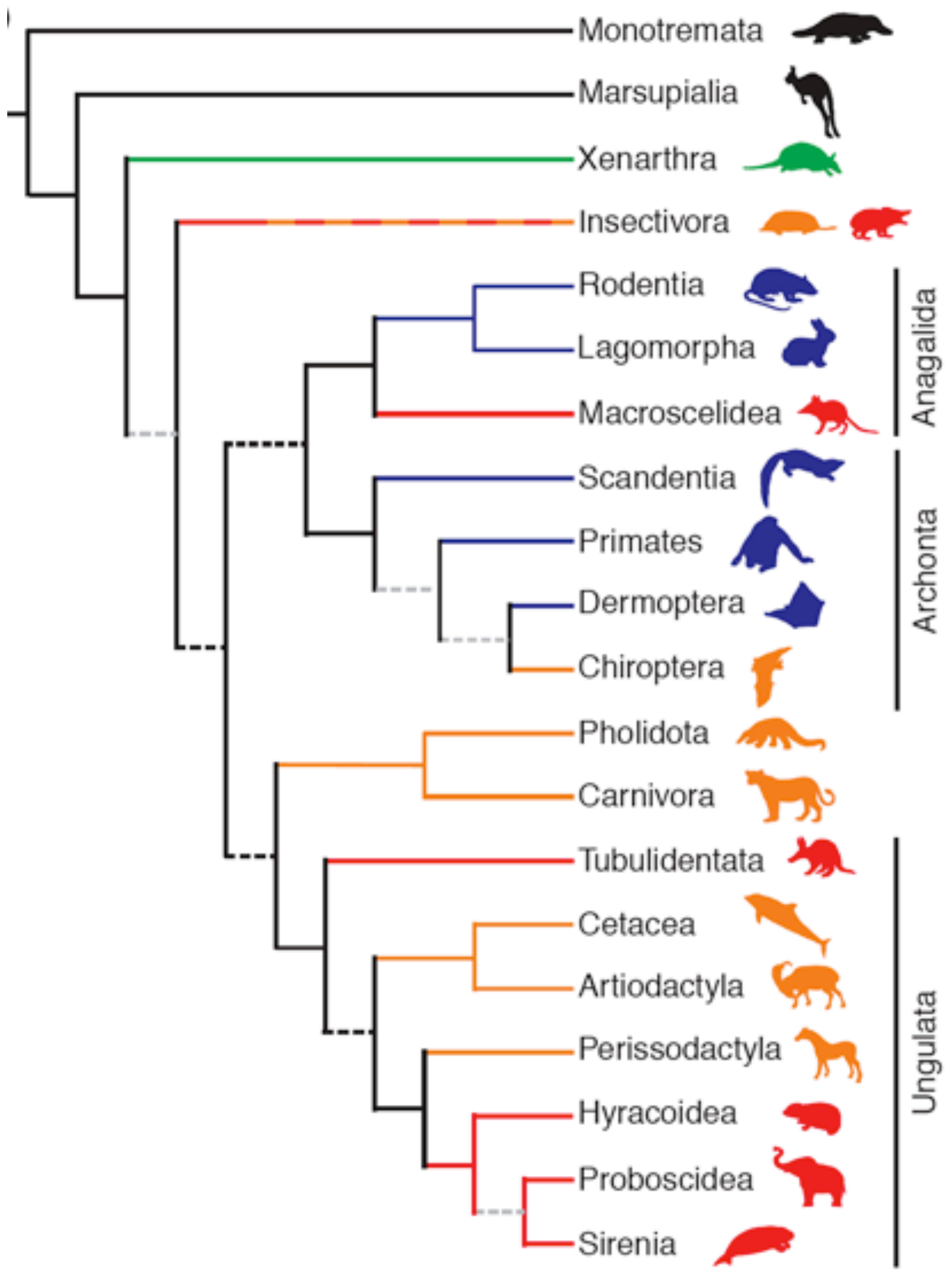
- DFE of annotated functional non-coding regions in human genome



- Negative selection in conserved and non-conserved human genomic region

picture is from https://www.phonandroid.com/tesla-continue-de-decliner-le-cybertruck-a-toutes-les-sauces-avec-un-marteau-de-gym.html
Phylogeny is from Springer et al., 2004

Mutations at conserved sites are more deleterious

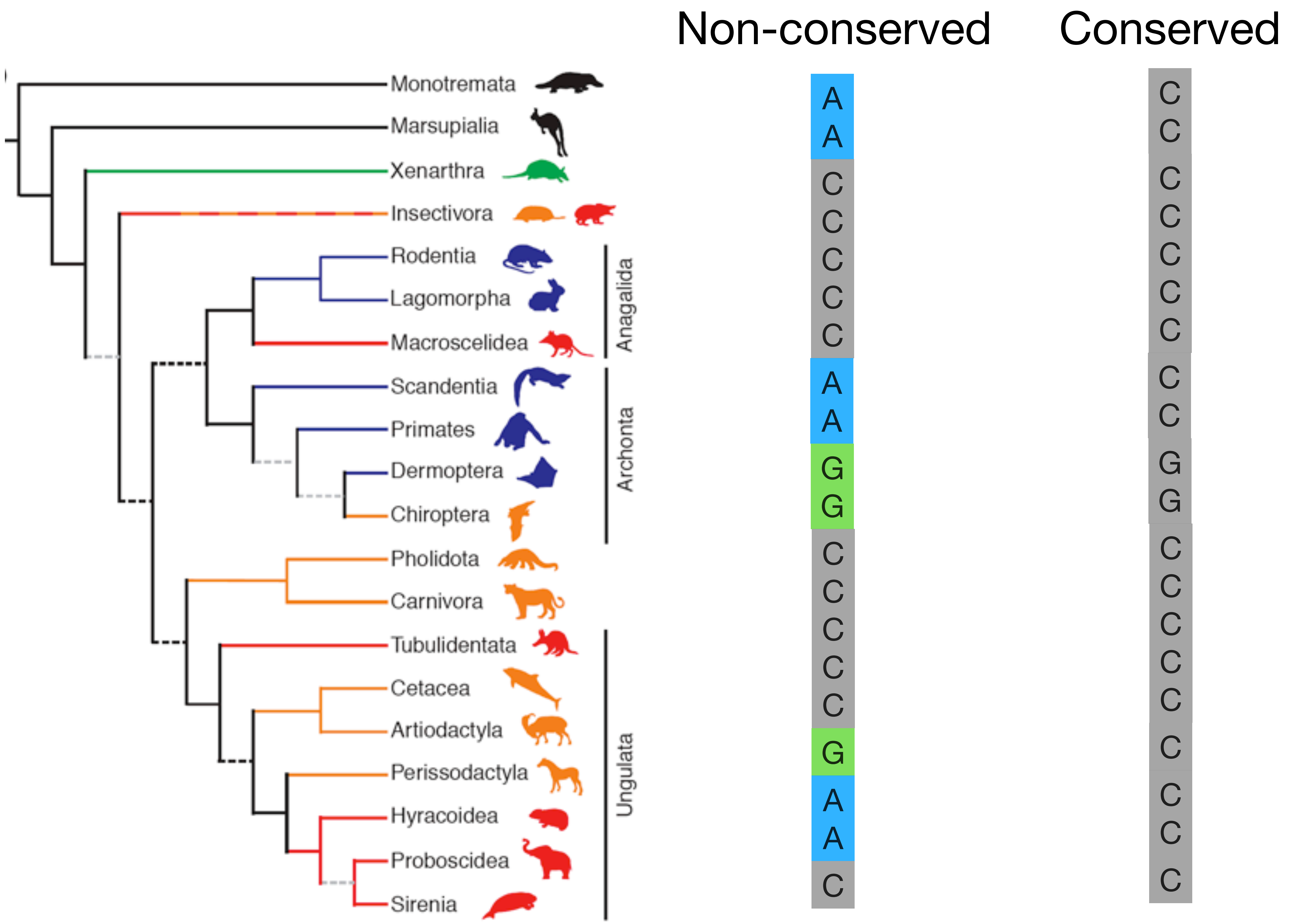


Non-conserved	Conserved
A	C
A	C
C	C
C	C
C	C
C	C
C	C
A	C
A	C
G	C
G	C
C	C
C	C
C	C
C	C
C	C
G	C
A	C
A	C
C	C

Phylogeny is from Springer et al., 2004

Mutations at conserved sites are more deleterious

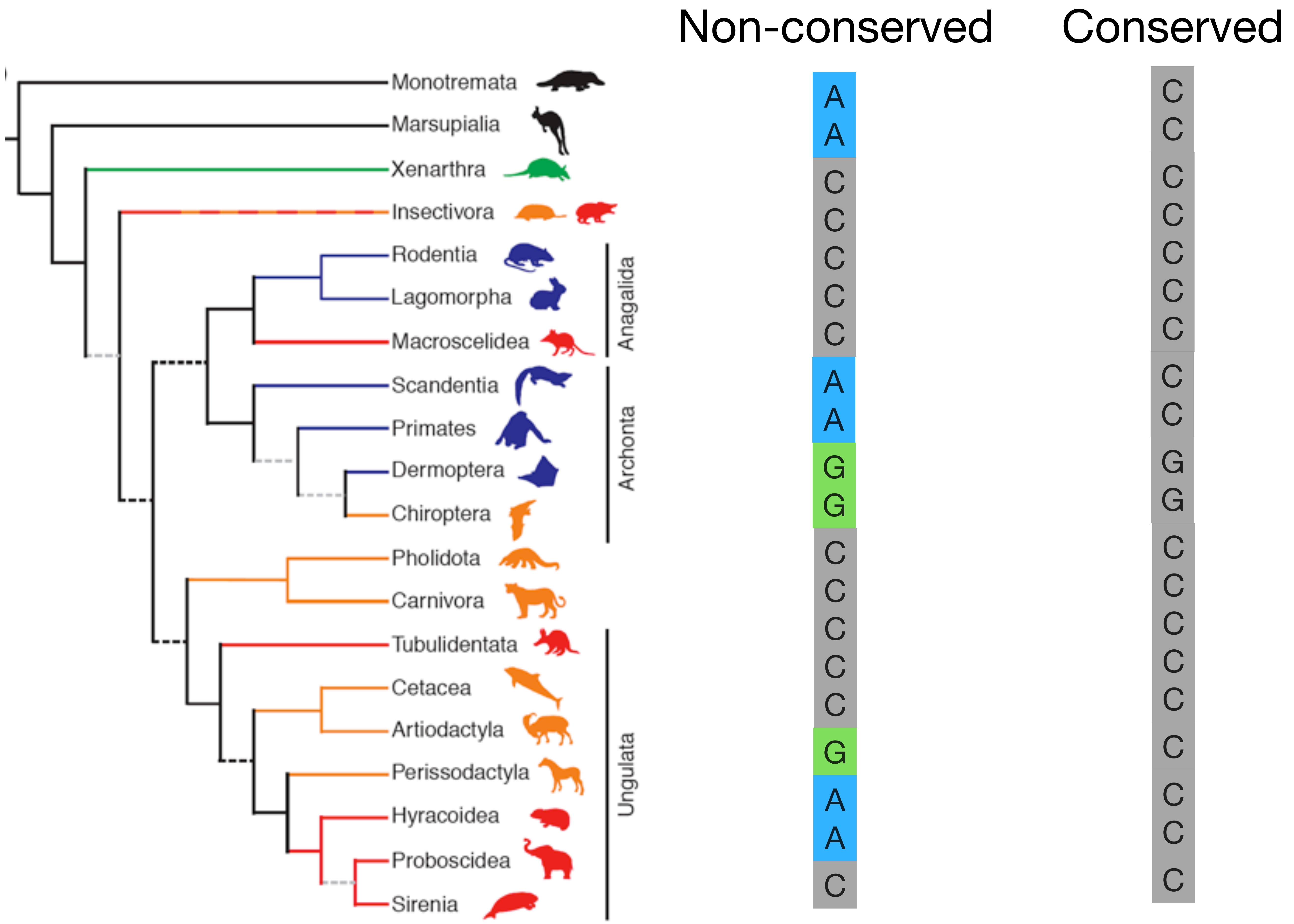
PhastCons scores: PHAST (PHylogenetic Analysis with Space/Time models)



Phylogeny is from Springer et al., 2004

Mutations at conserved sites are more deleterious

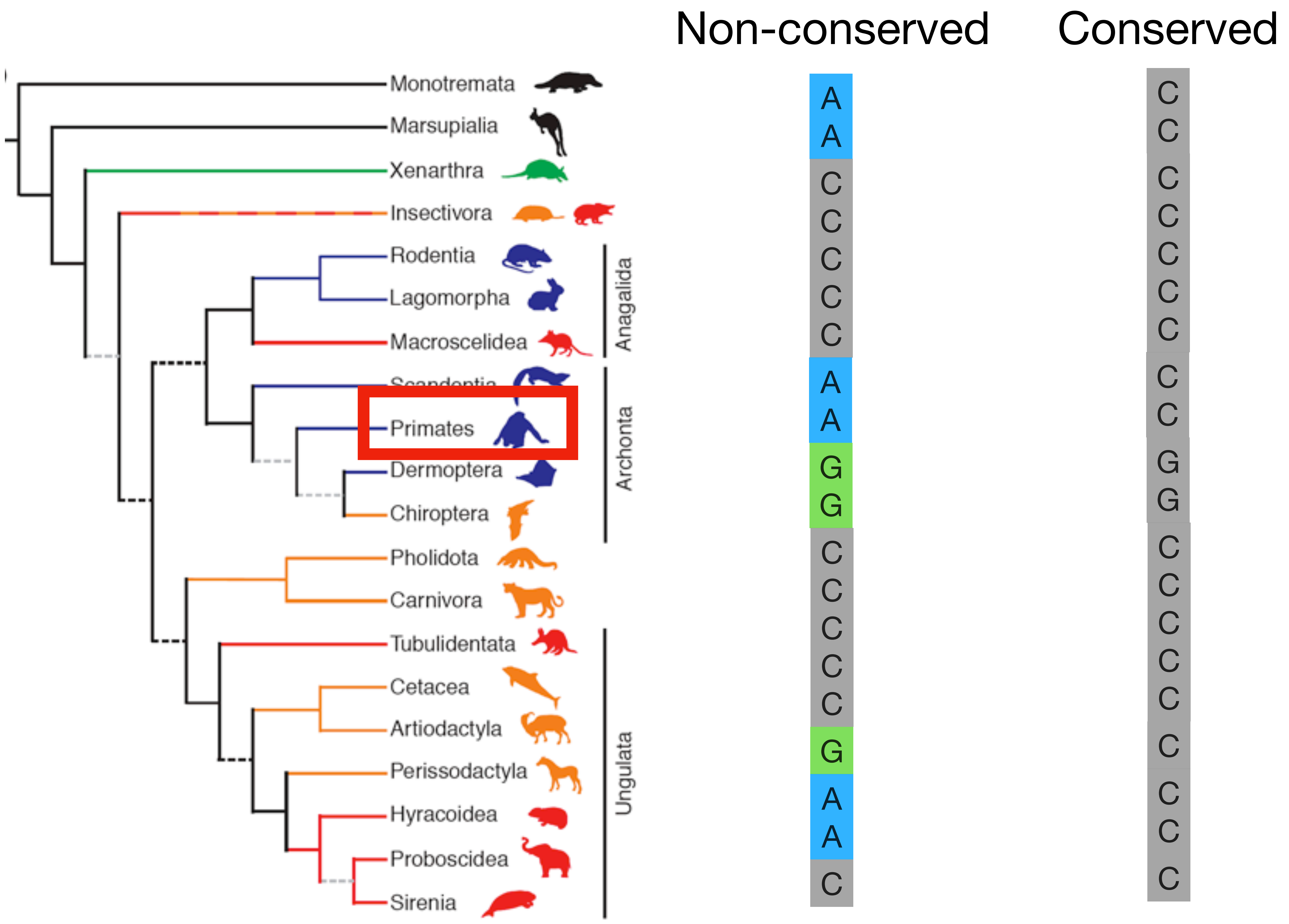
PhastCons scores: PHAST (PHylogenetic Analysis with Space/Time models)
Prob(being conserved) ∈ [0,1]. 1 is for the most conserved sites



Phylogeny is from Springer et al., 2004

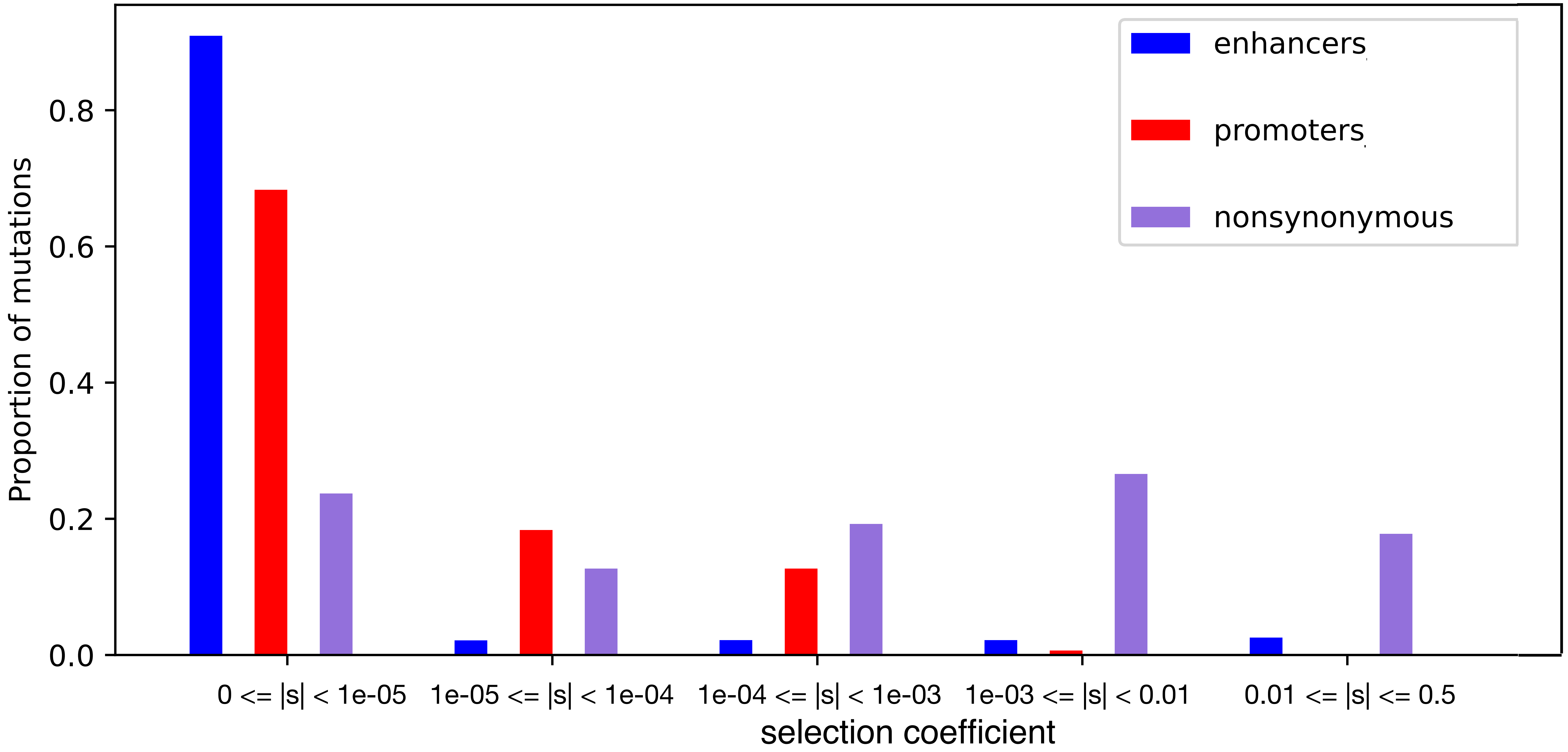
Mutations at conserved sites are more deleterious

PhastCons scores: PHAST (PHylogenetic Analysis with Space/Time models)
Prob(being conserved) ∈ [0,1]. 1 is for the most conserved sites



Phylogeny is from Springer et al., 2004

DFE of enhancers and promoters



Neutral

enhancers > promoters

Weakly deleterious

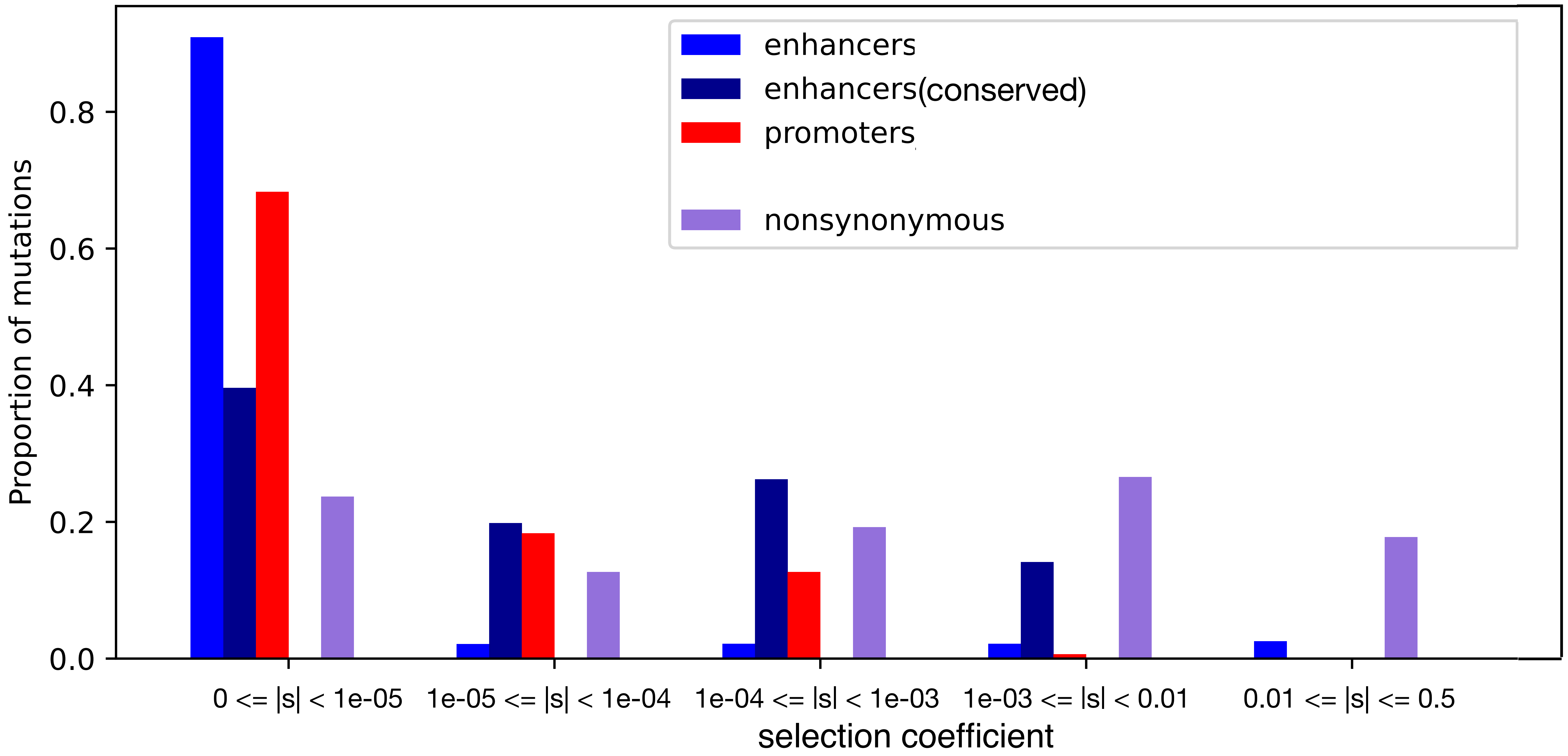
promoters > enhancers

Moderate to strongly deleterious

small proportion, enhancers > promoters

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

Mutations at conserved sites are more delerious

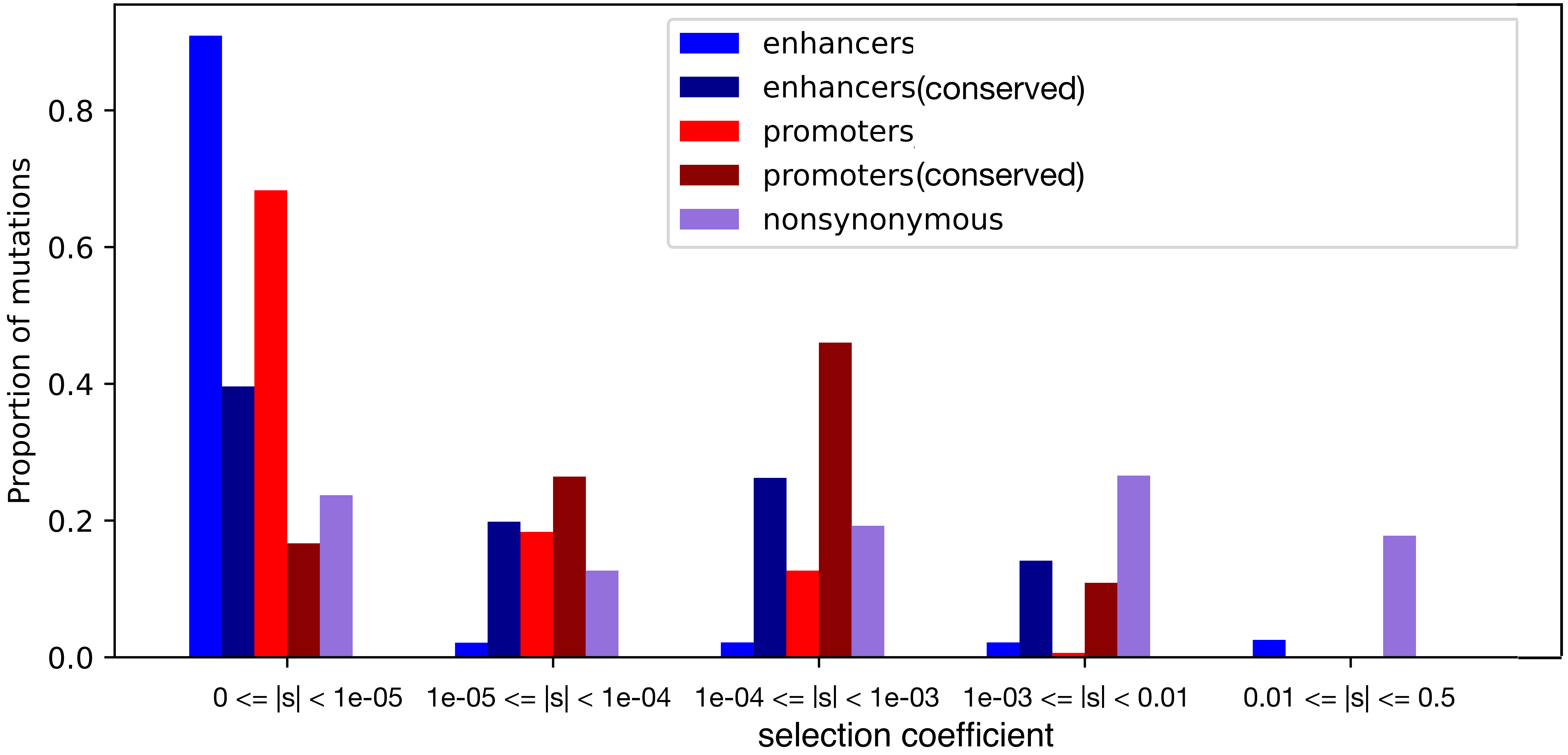


Neutral Weakly deleterious Moderate to strongly deleterious

Negative selection: conserved > non-conserved.

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

Mutations at conserved sites are more delerious



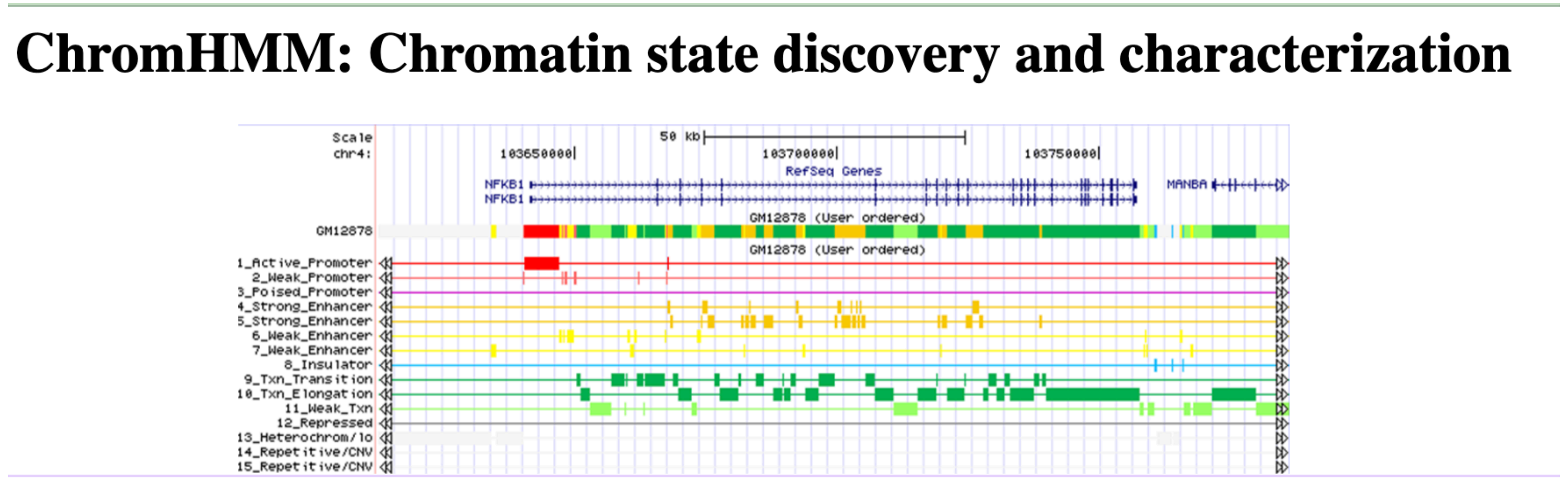
Neutral Weakly deleterious Moderate to strongly deleterious

Negative selection: conserved > non-conserved.

DFE of non-synonymous sites from Kim et al., 2017, 1KG, EUR

DFE of mutations in regions of differing phylogenetic constraint

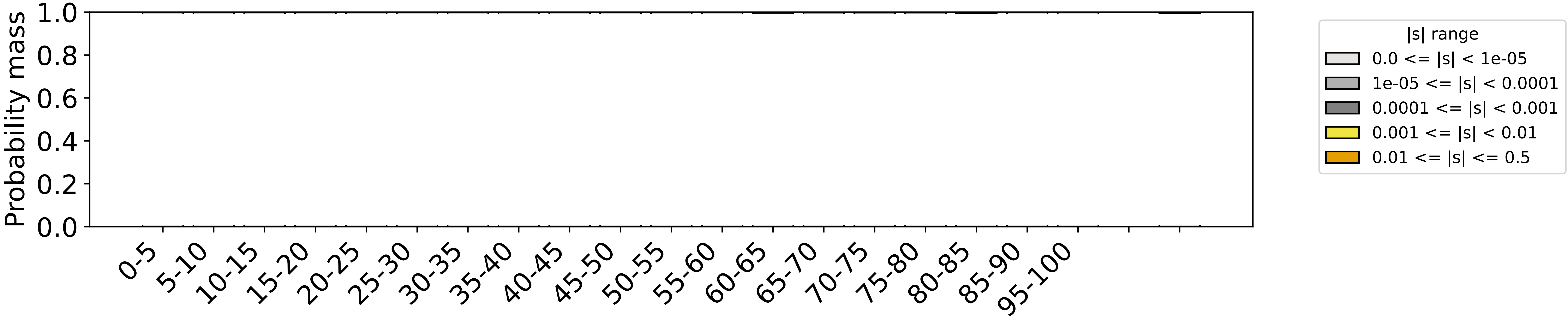
DFE of putative functional non-coding regions
(~50% of the genome)



ChromHMM: Chromatin state discovery and characterization

DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



Neutral

Weakly D

Moderately D

Strongly D

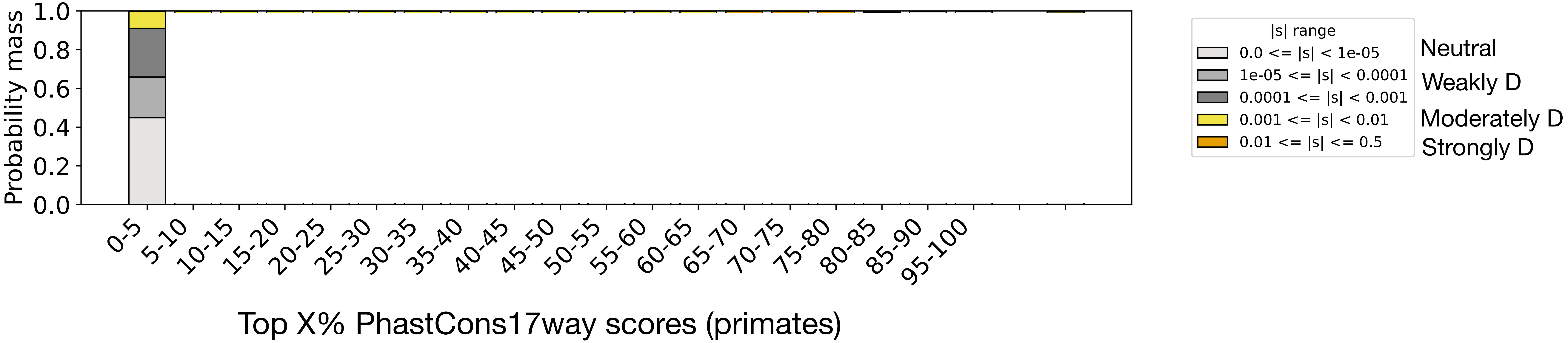
Top X% PhastCons17way scores (primates)

Most conserved

less conserved

DFE of mutations in regions of differing phylogenetic constraint

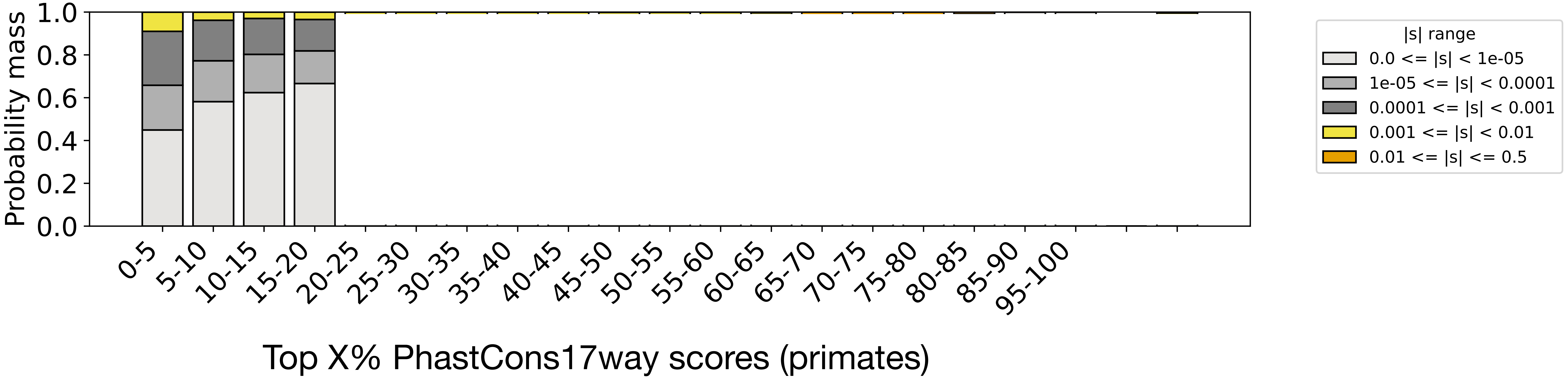
DFE of putative functional non-coding regions



Most conserved less conserved

DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



Neutral

Weakly D

Moderately D

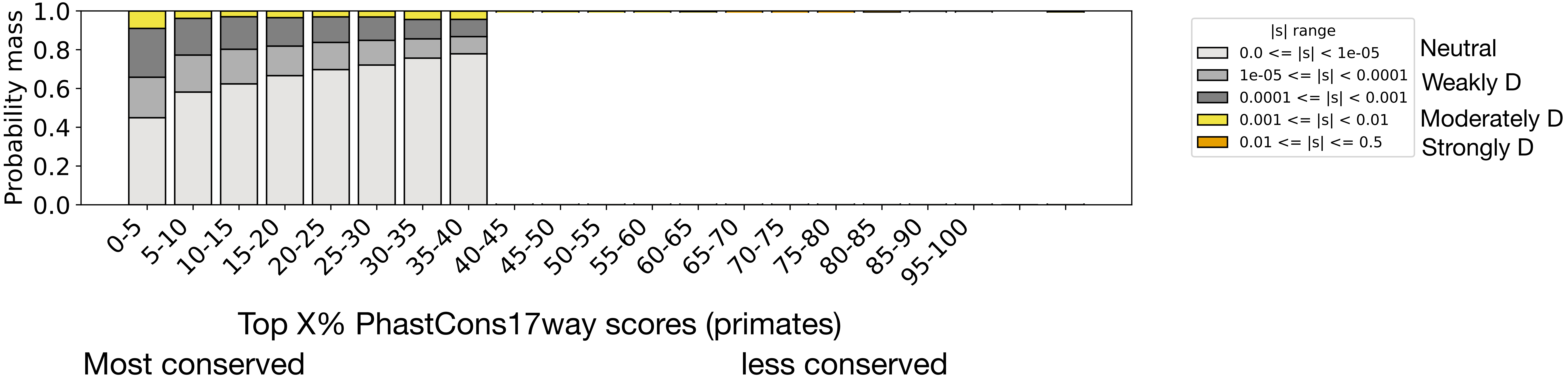
Strongly D

Most conserved

less conserved

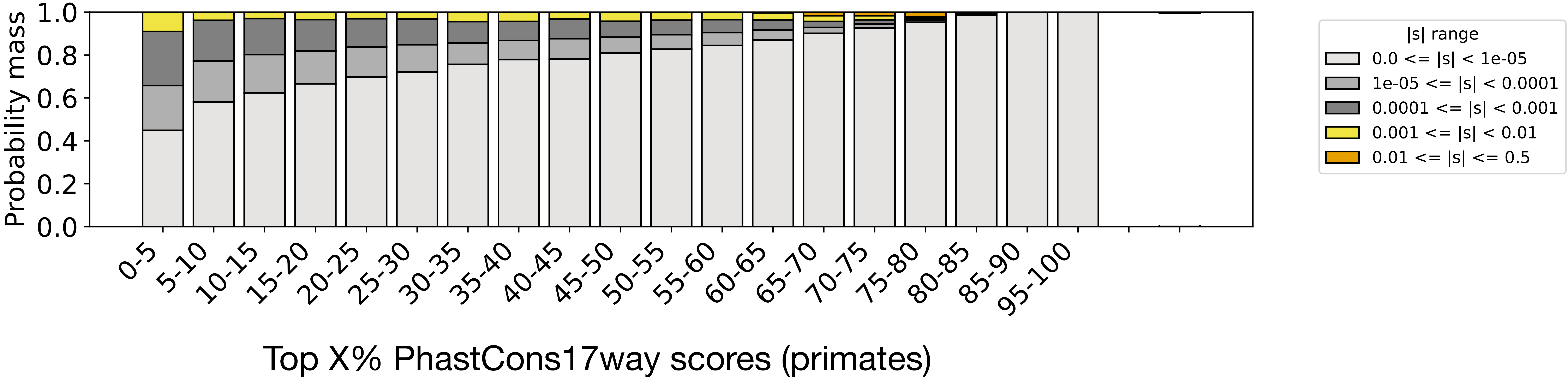
DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



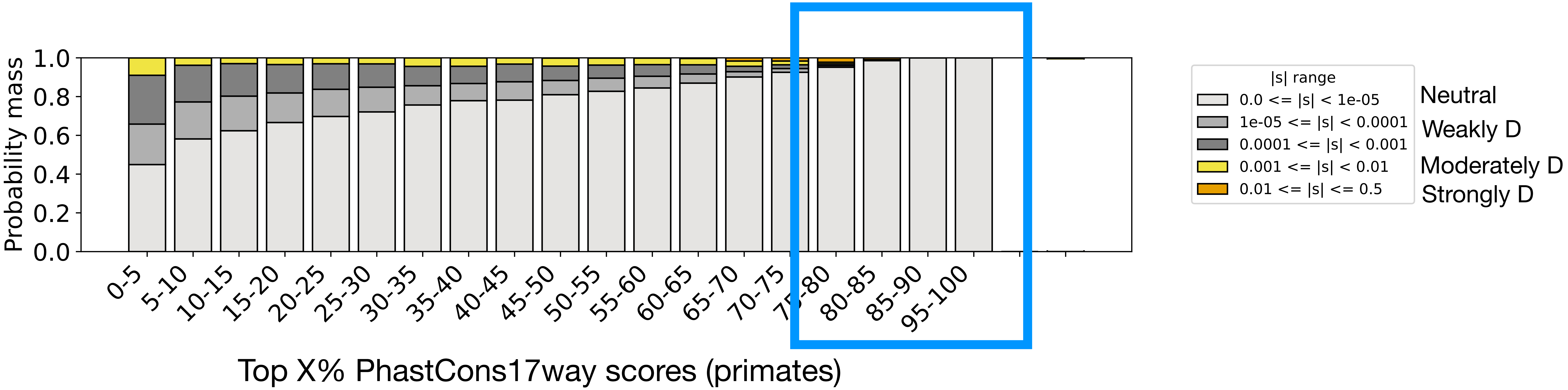
Neutral
Weakly D
Moderately D
Strongly D

Most conserved

less conserved

DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



Neutral
Weakly D
Moderately D
Strongly D

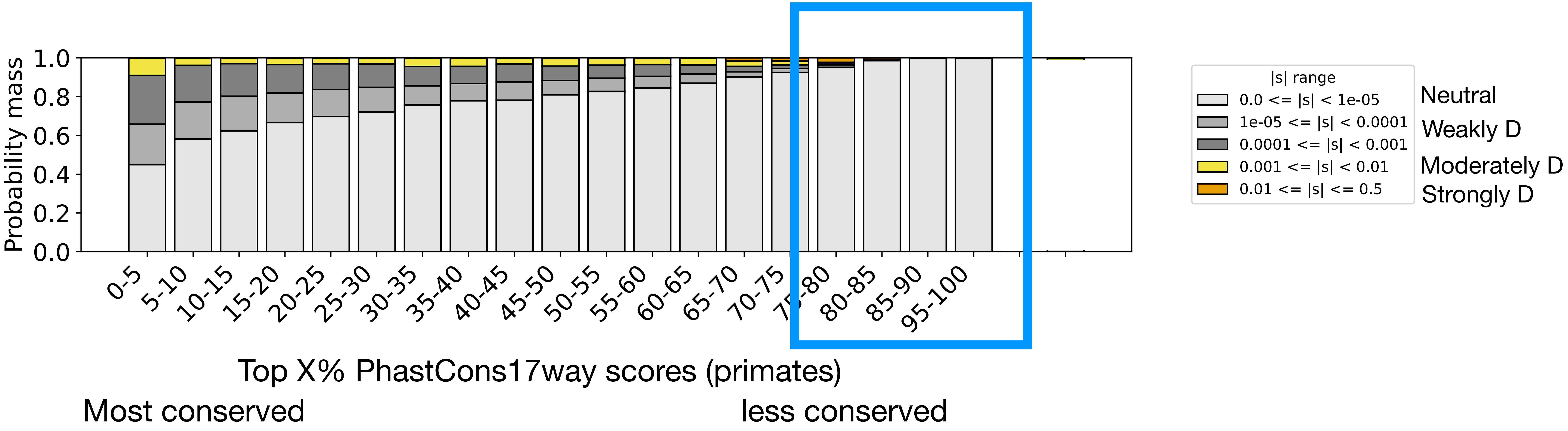
Most conserved

less conserved

DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions

Models for DFE

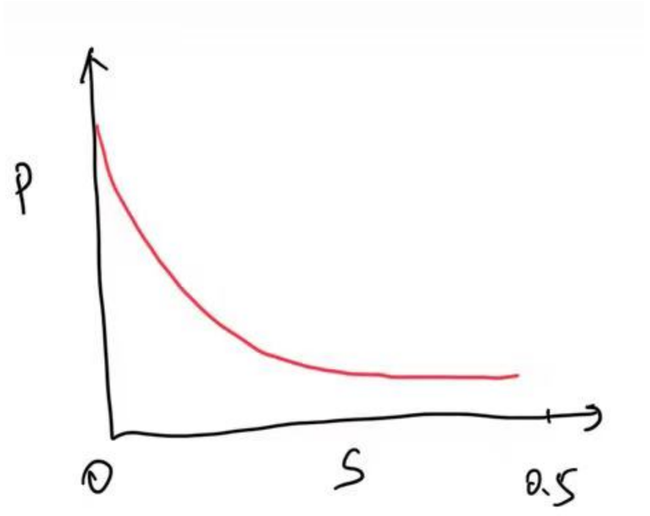


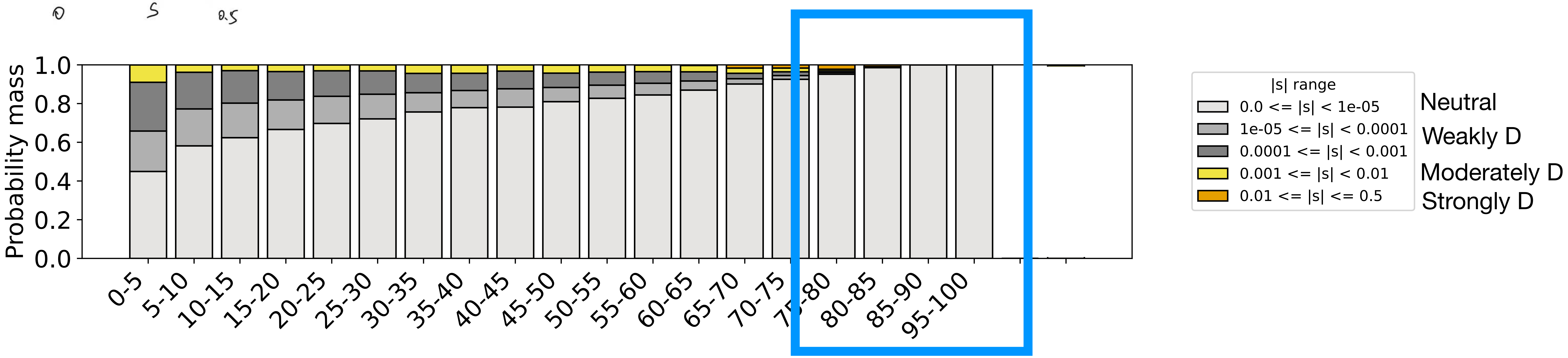
DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions

Models for DFE

Gamma distribution





Top X% PhastCons17way scores (primates)

Most conserved

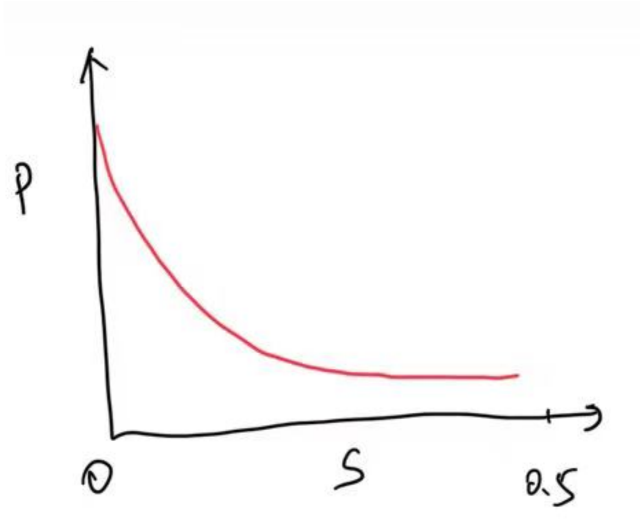
less conserved

DFE of mutations in regions of differing phylogenetic constraint

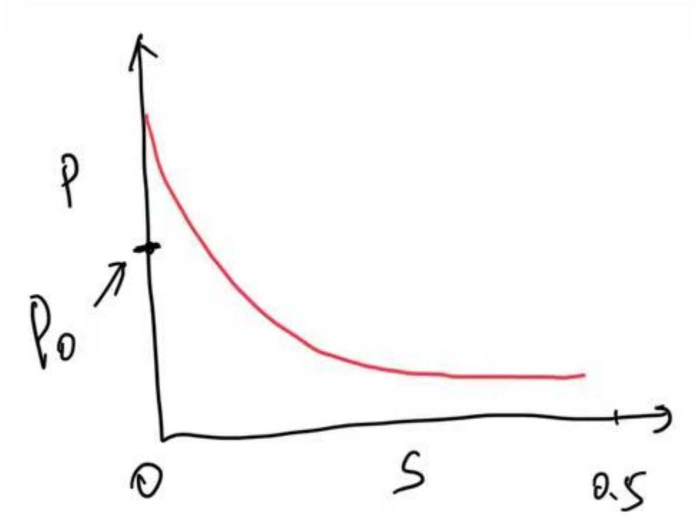
DFE of putative functional non-coding regions

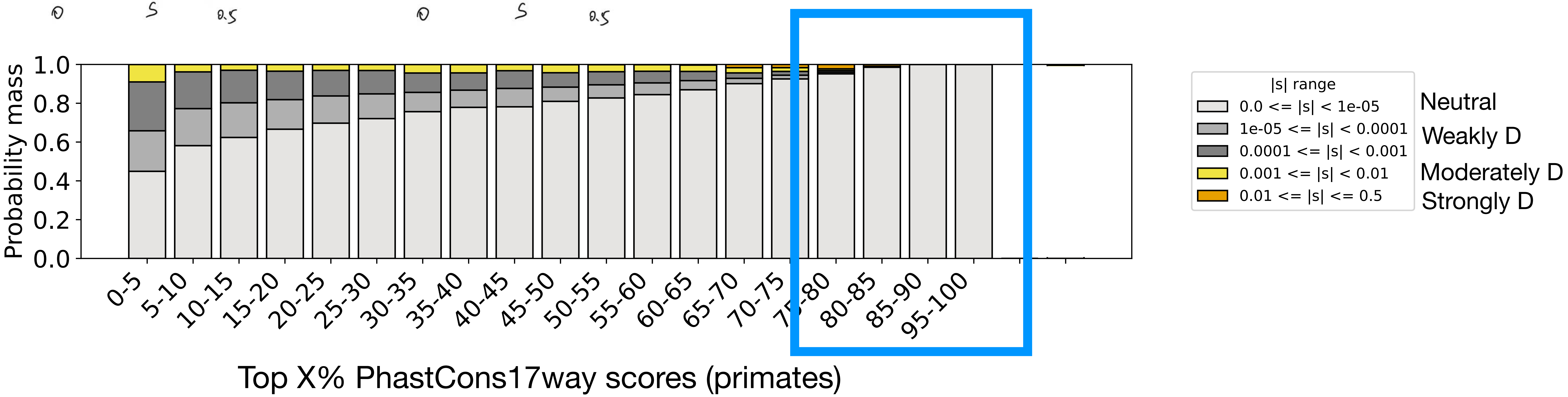
Models for DFE

Gamma distribution



Neu+Gamma distribution





Most conserved

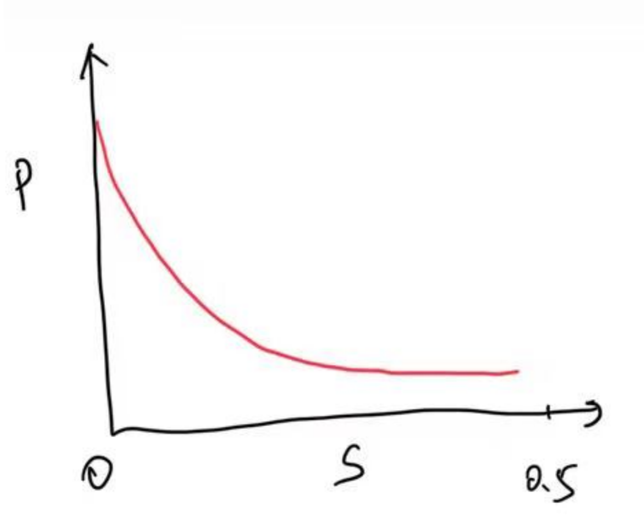
less conserved

DFE of mutations in regions of differing phylogenetic constraint

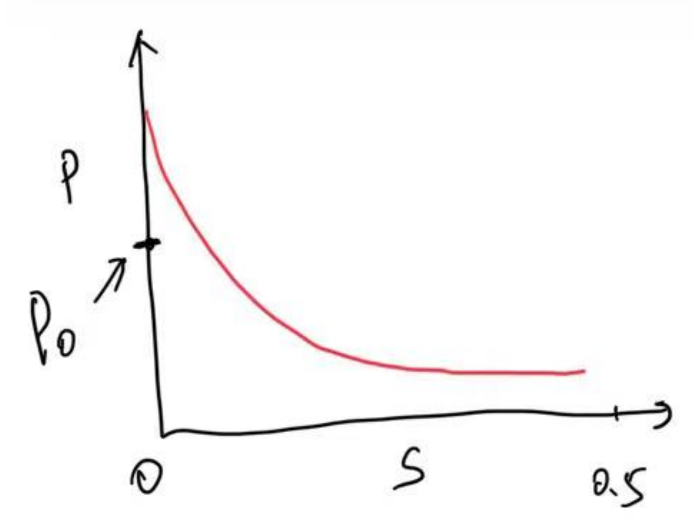
DFE of putative functional non-coding regions

Models for DFE

Gamma distribution

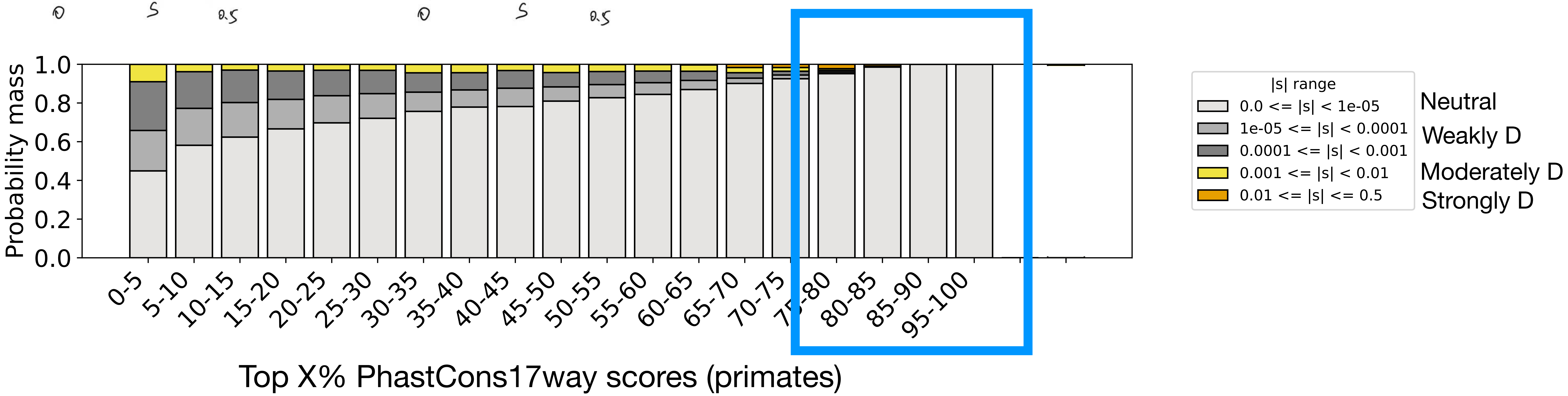


Neu+Gamma distribution



Neutral model

P(s=0)=1

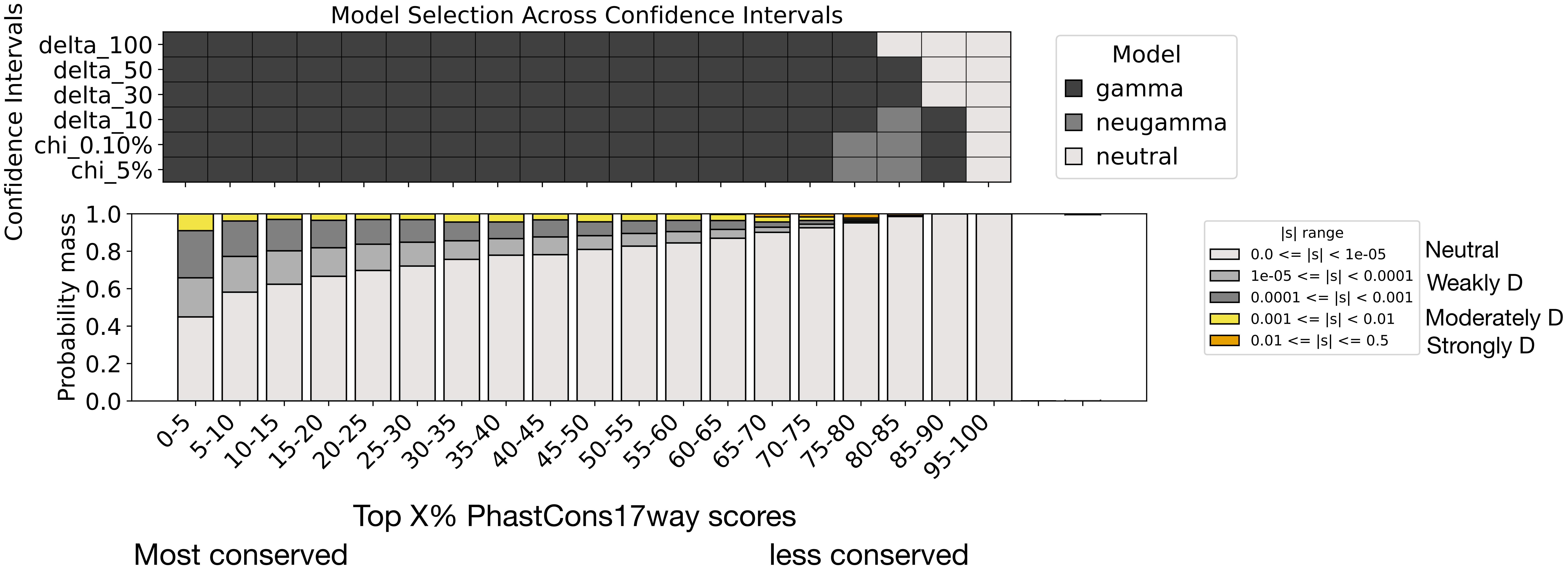


Most conserved

less conserved

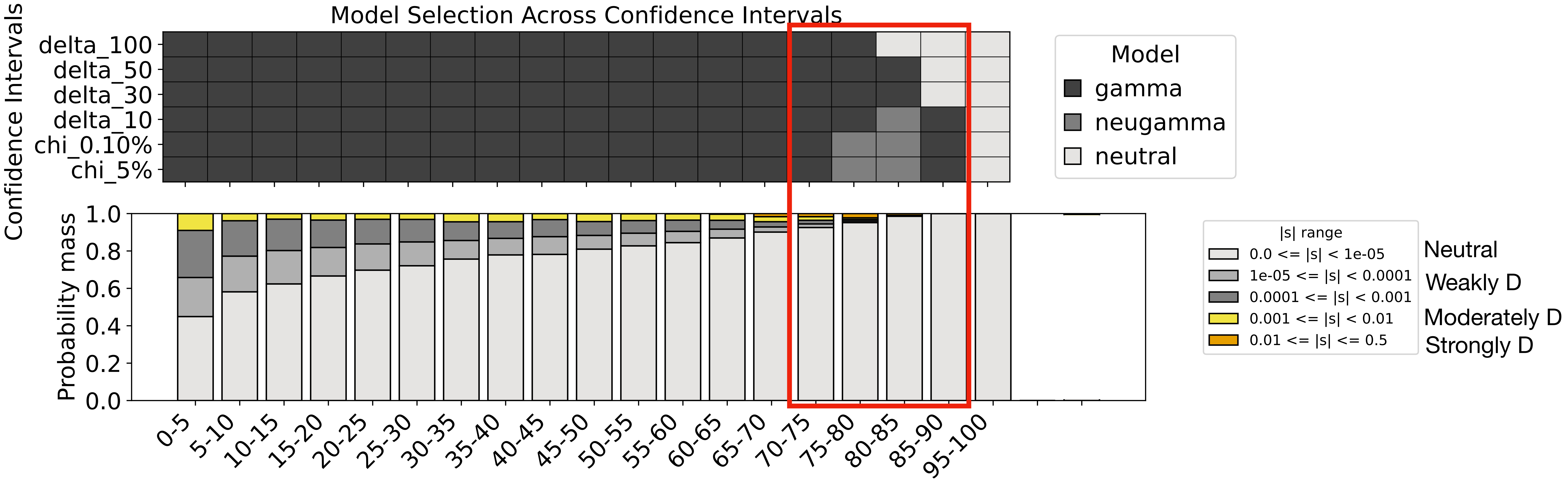
DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



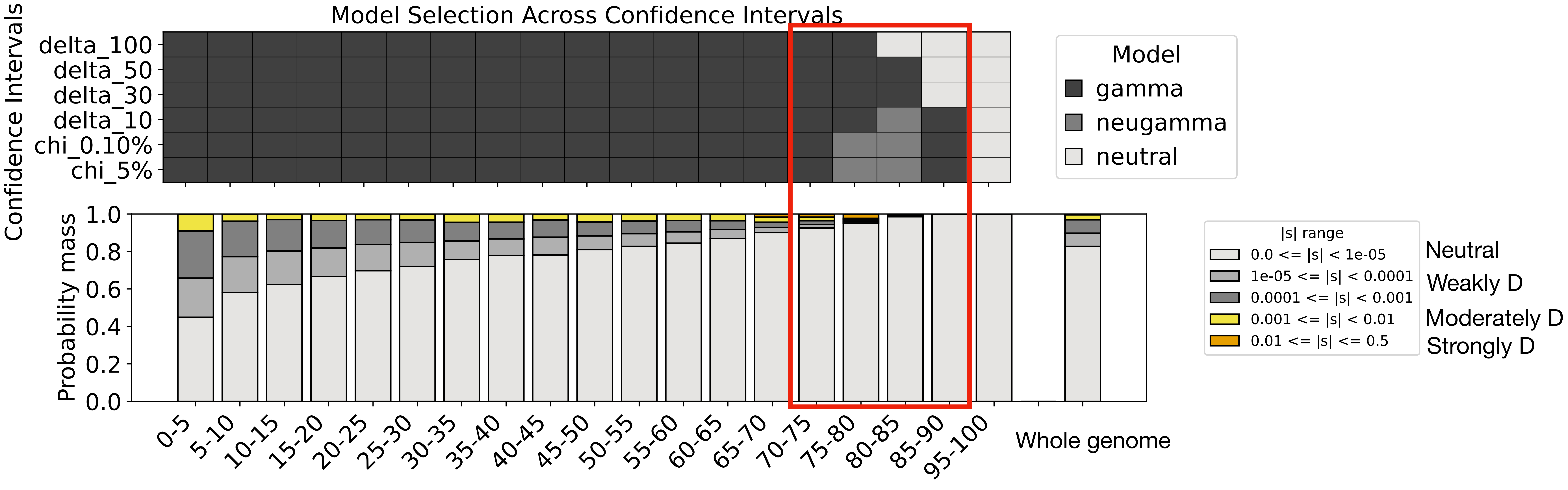
Top X% PhastCons17way scores

Most conserved

less conserved

DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions

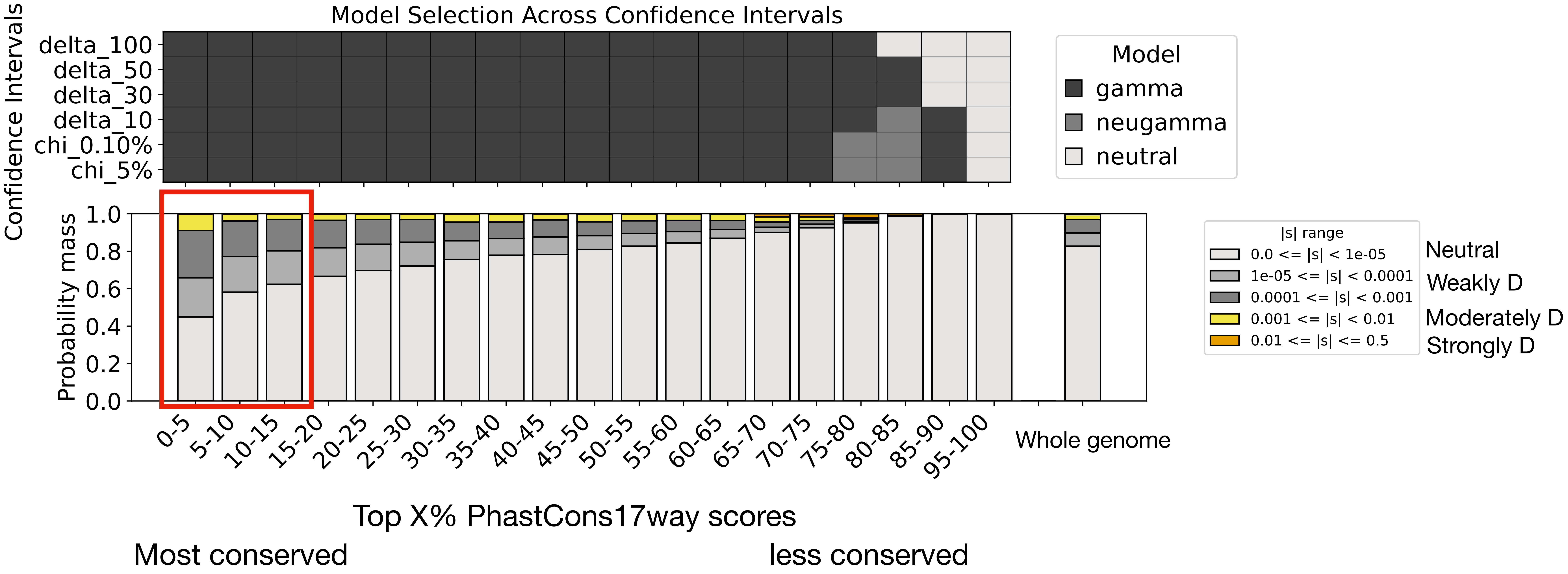


Top X% PhastCons17way scores

Most conserved less conserved

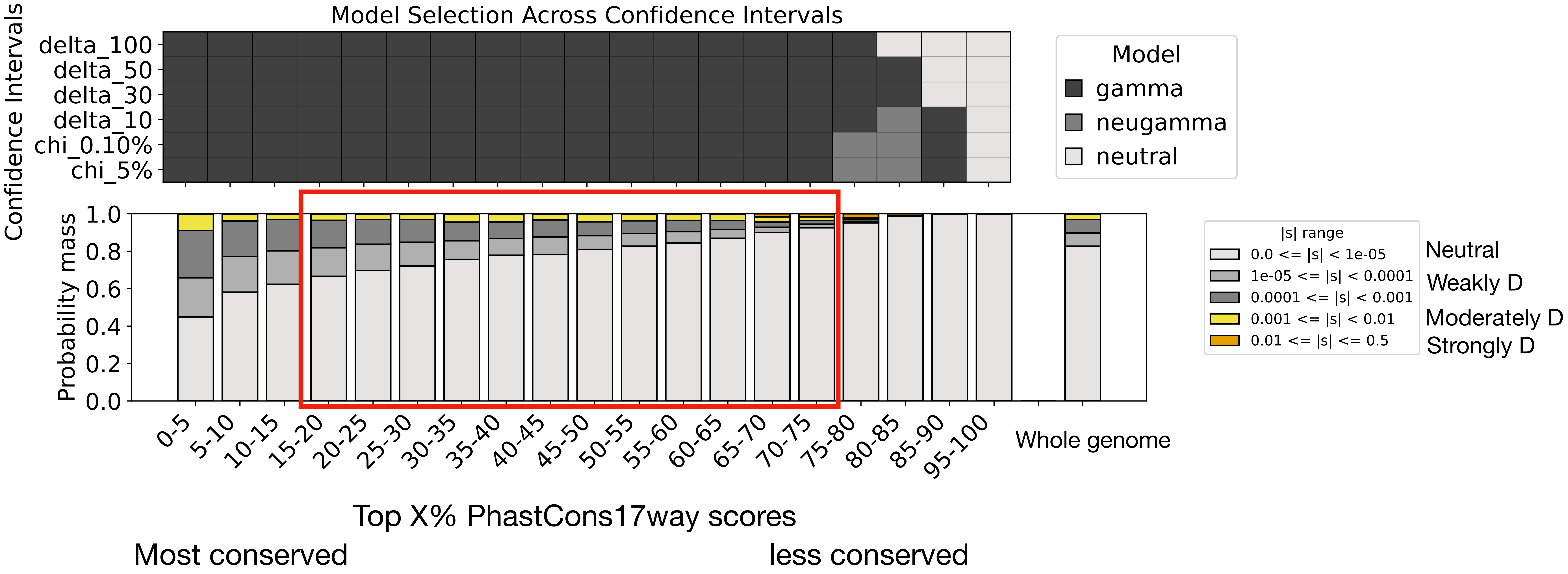
DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



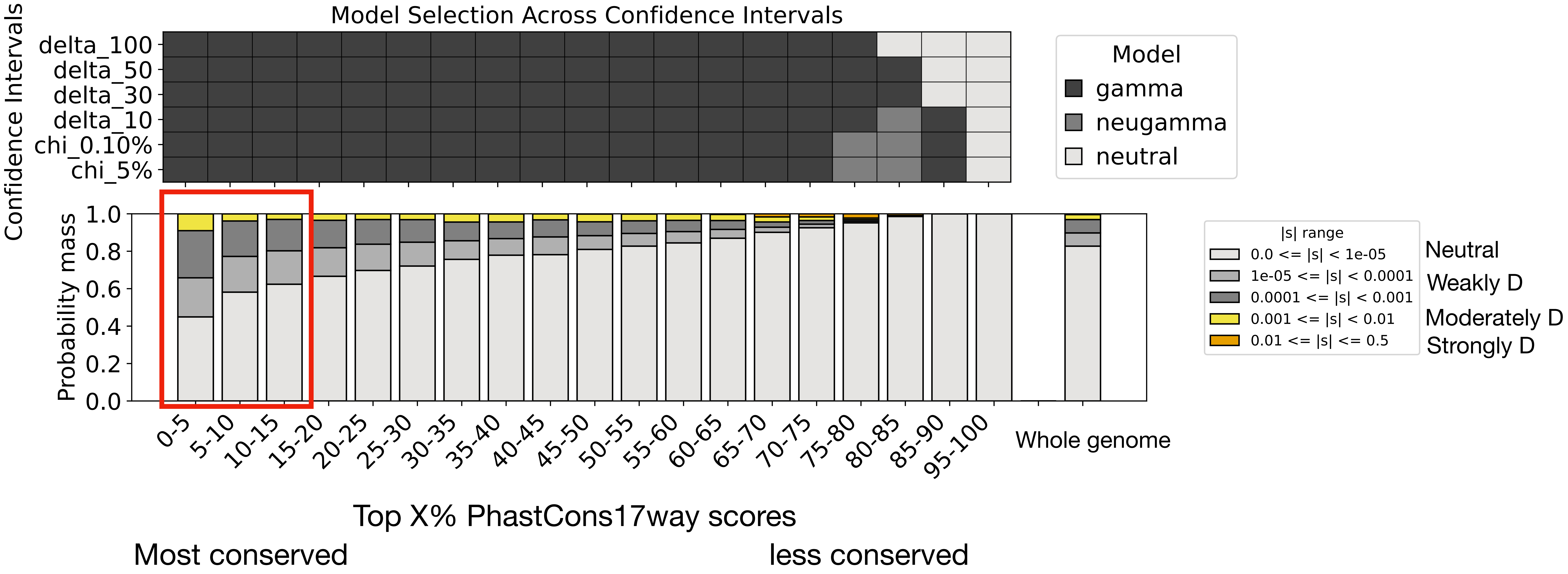
DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions

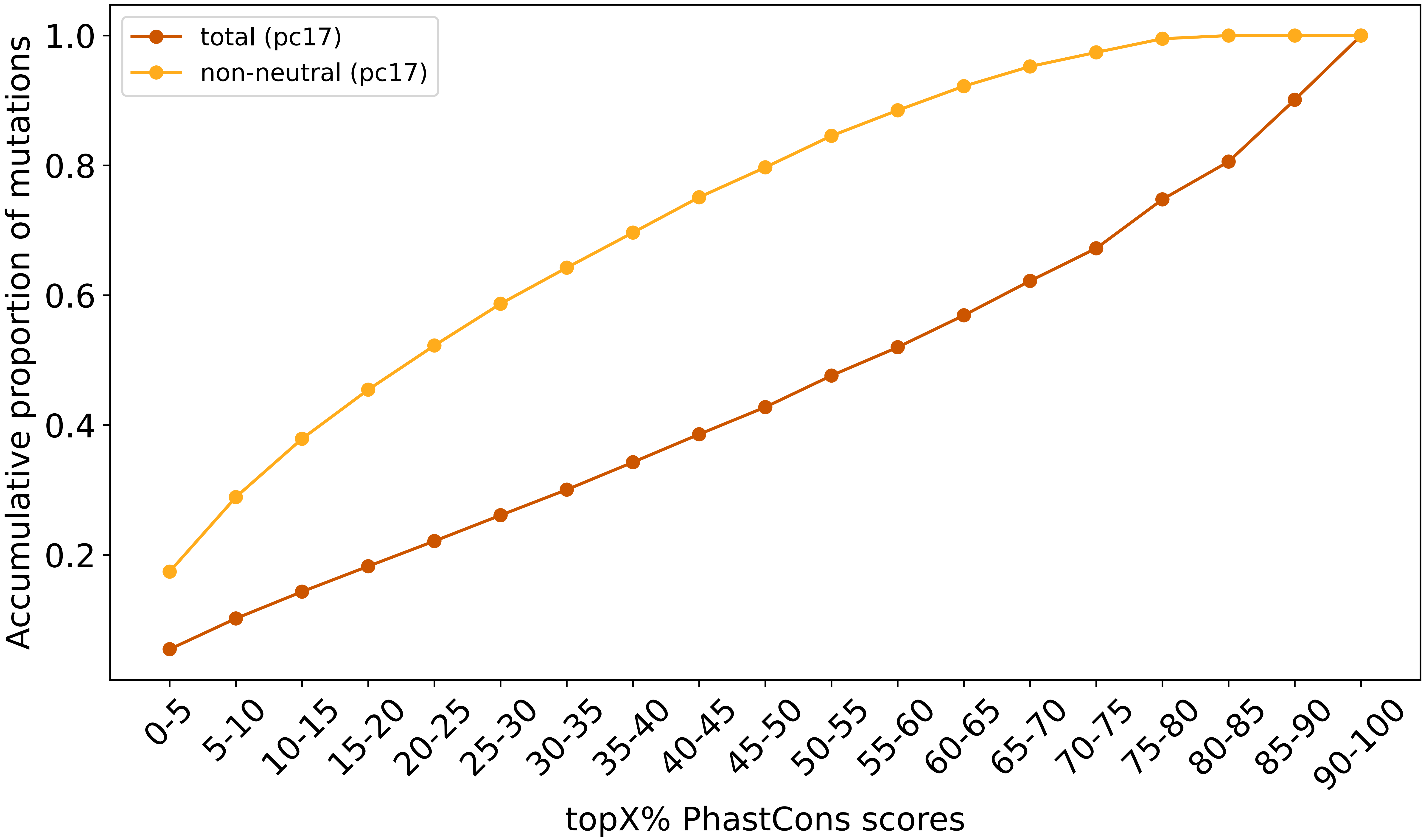


DFE of mutations in regions of differing phylogenetic constraint

DFE of putative functional non-coding regions



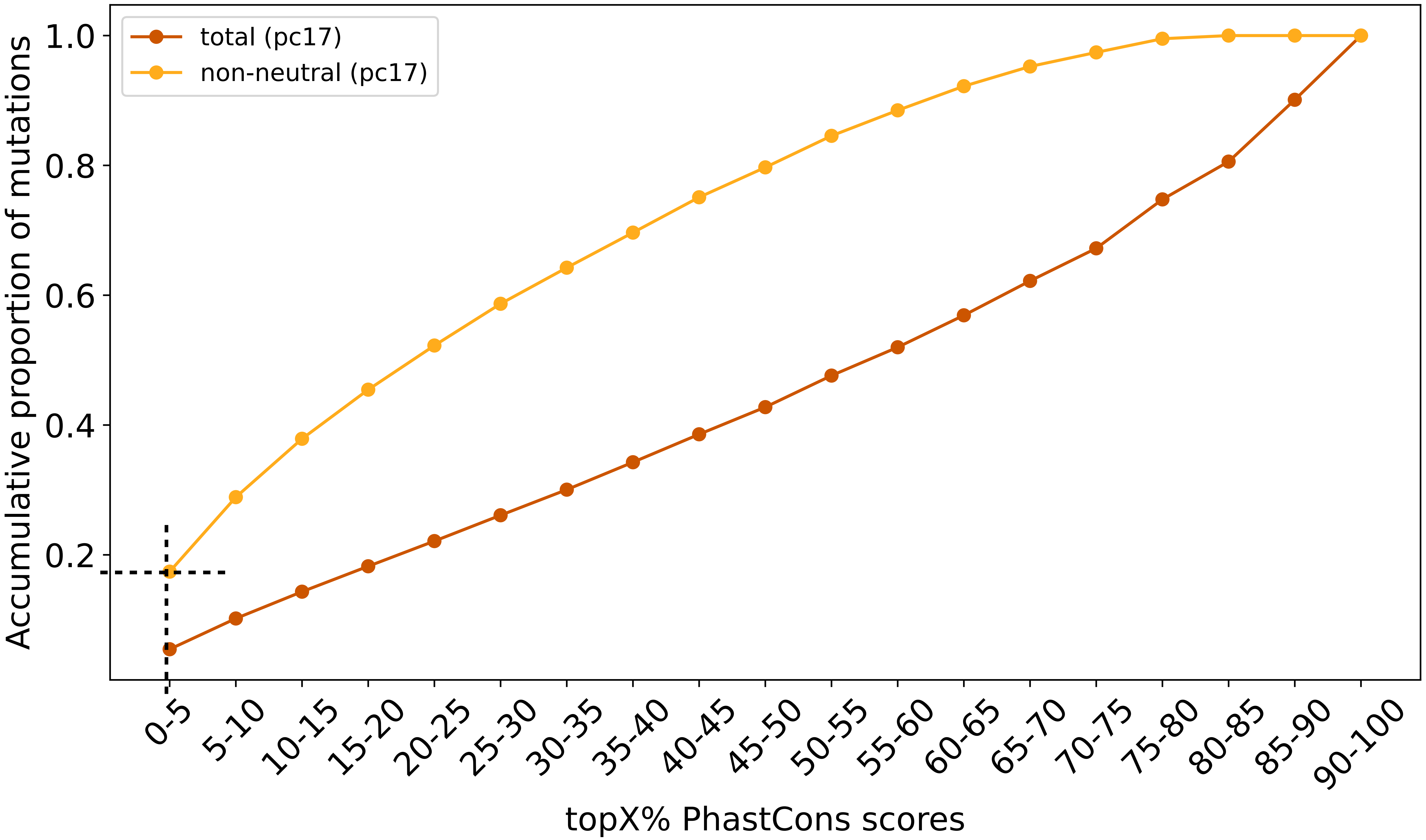
Phylogenetic constraint has limited power to identify deleterious mutations



Most conserved

less conserved

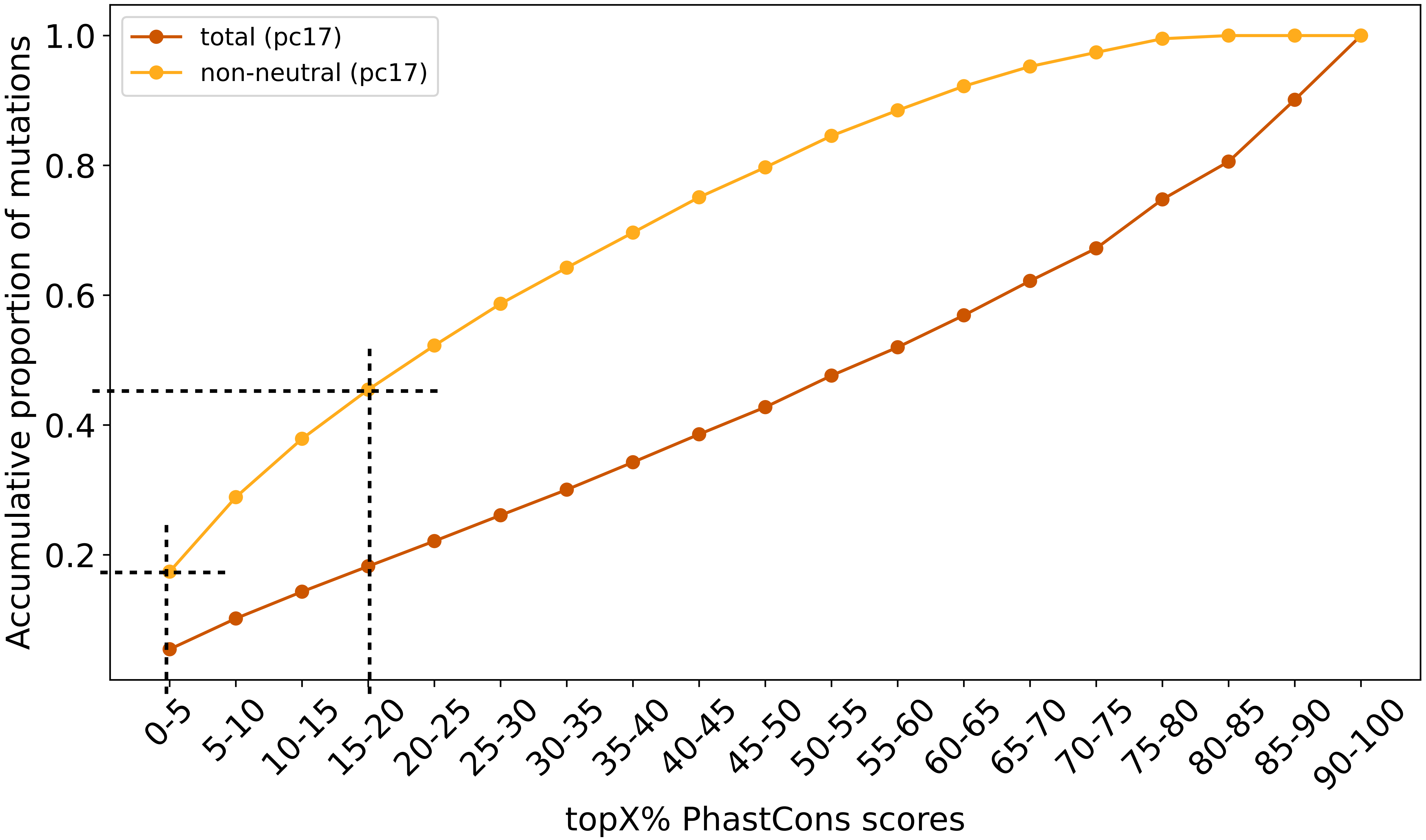
Phylogenetic constraint has limited power to identify deleterious mutations



Most conserved

less conserved

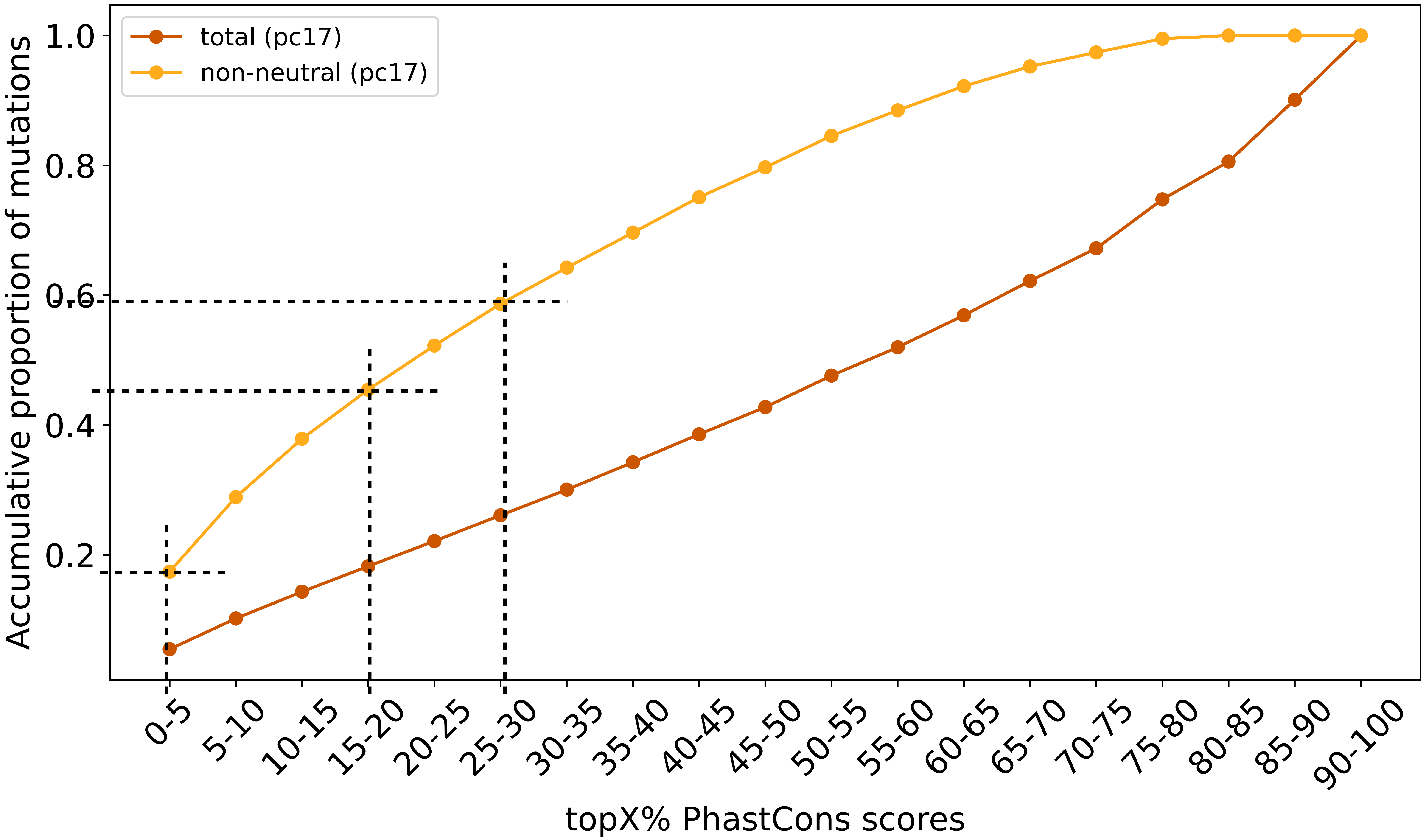
Phylogenetic constraint has limited power to identify deleterious mutations



Most conserved

less conserved

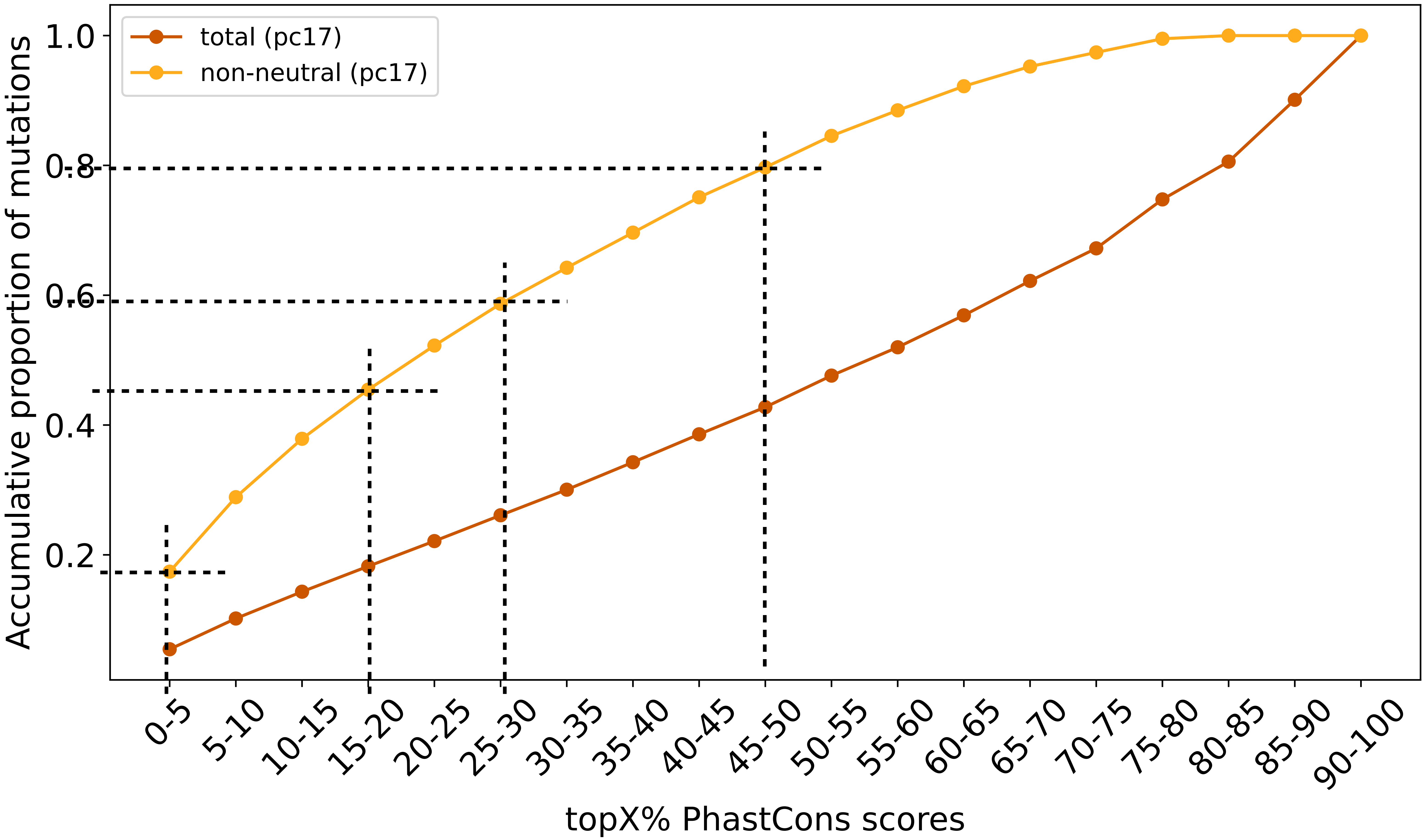
Phylogenetic constraint has limited power to identify deleterious mutations



Most conserved

less conserved

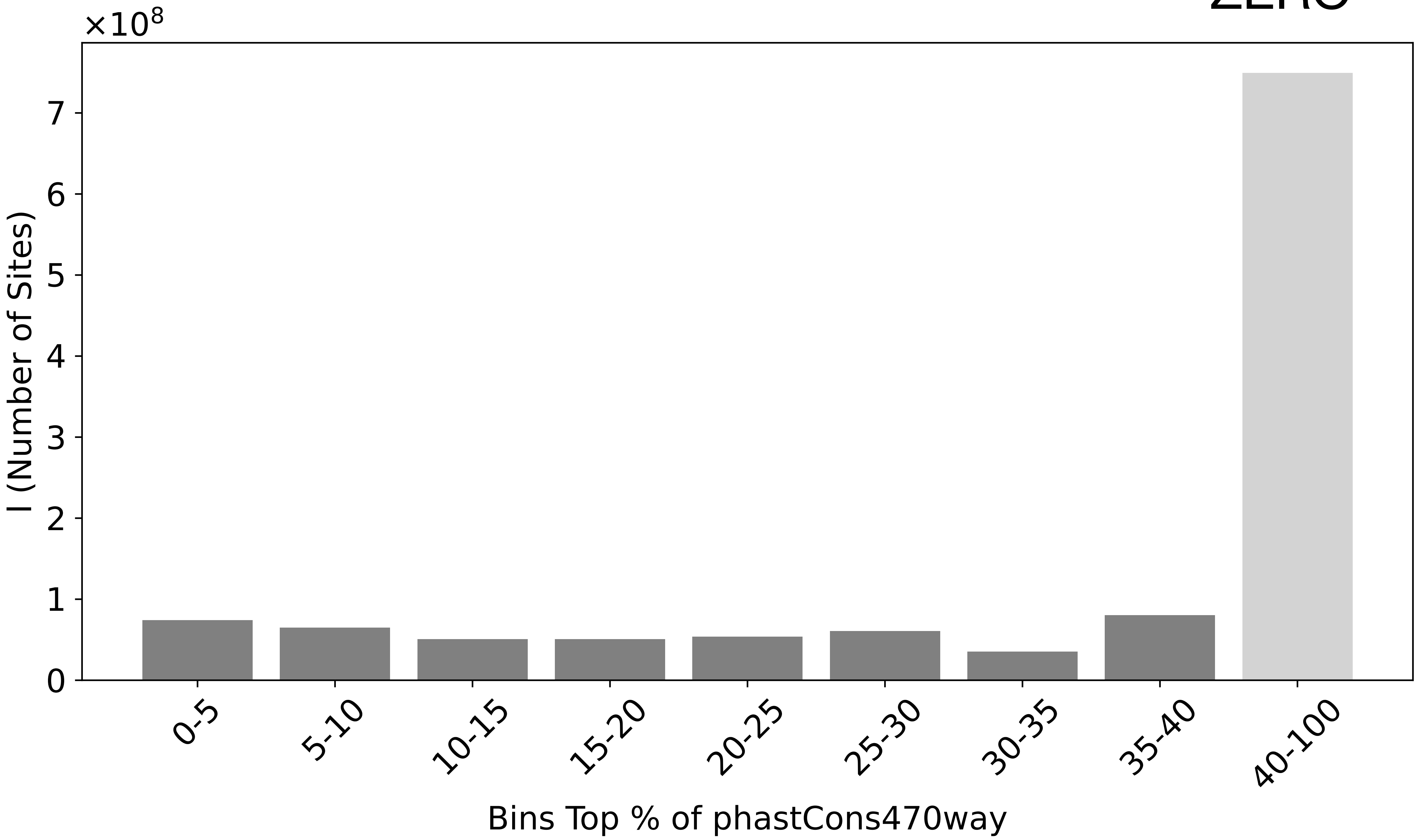
Phylogenetic constraint has limited power to identify deleterious mutations



Most conserved

less conserved

Constraint in mammals

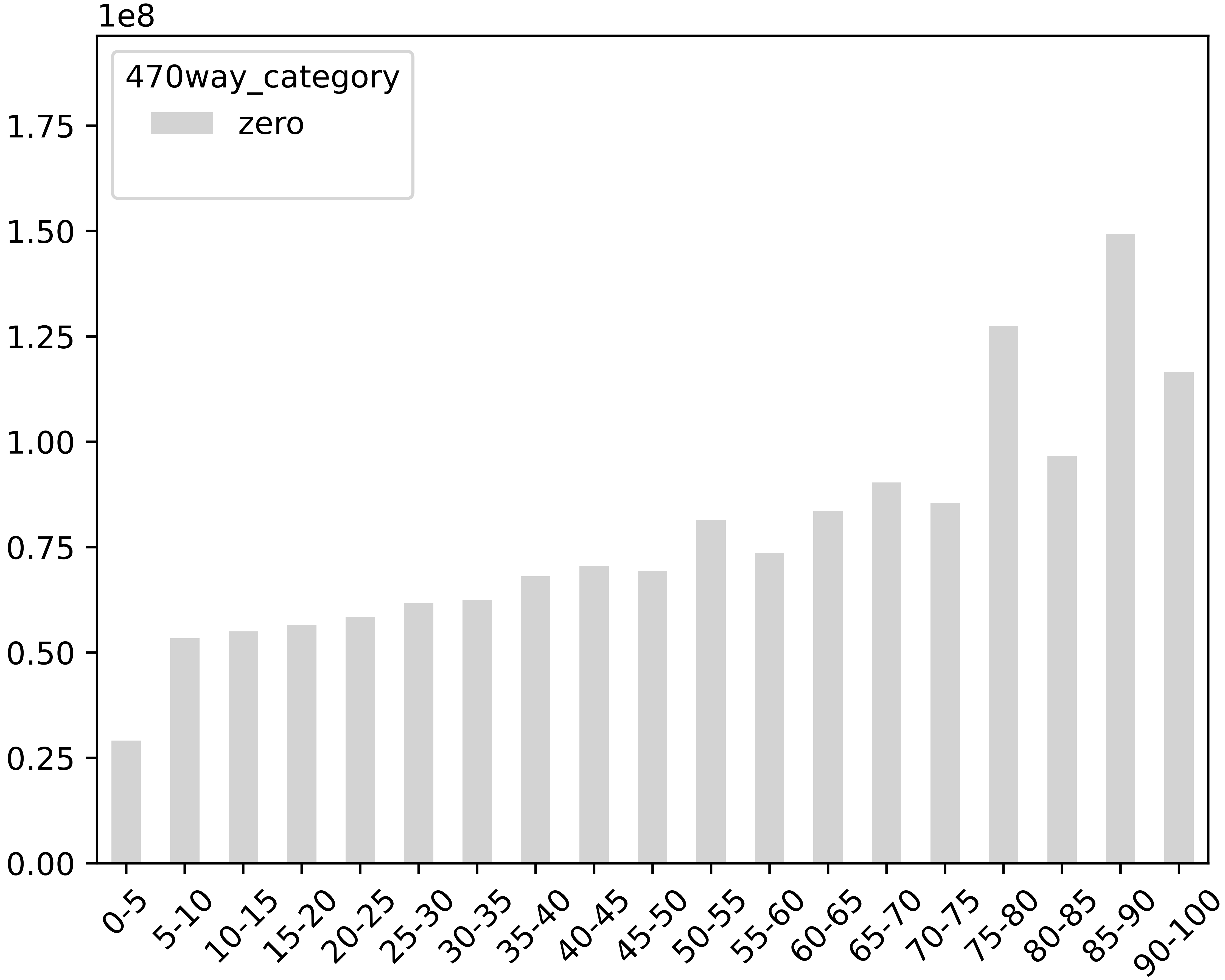


ZERO

Most conserved

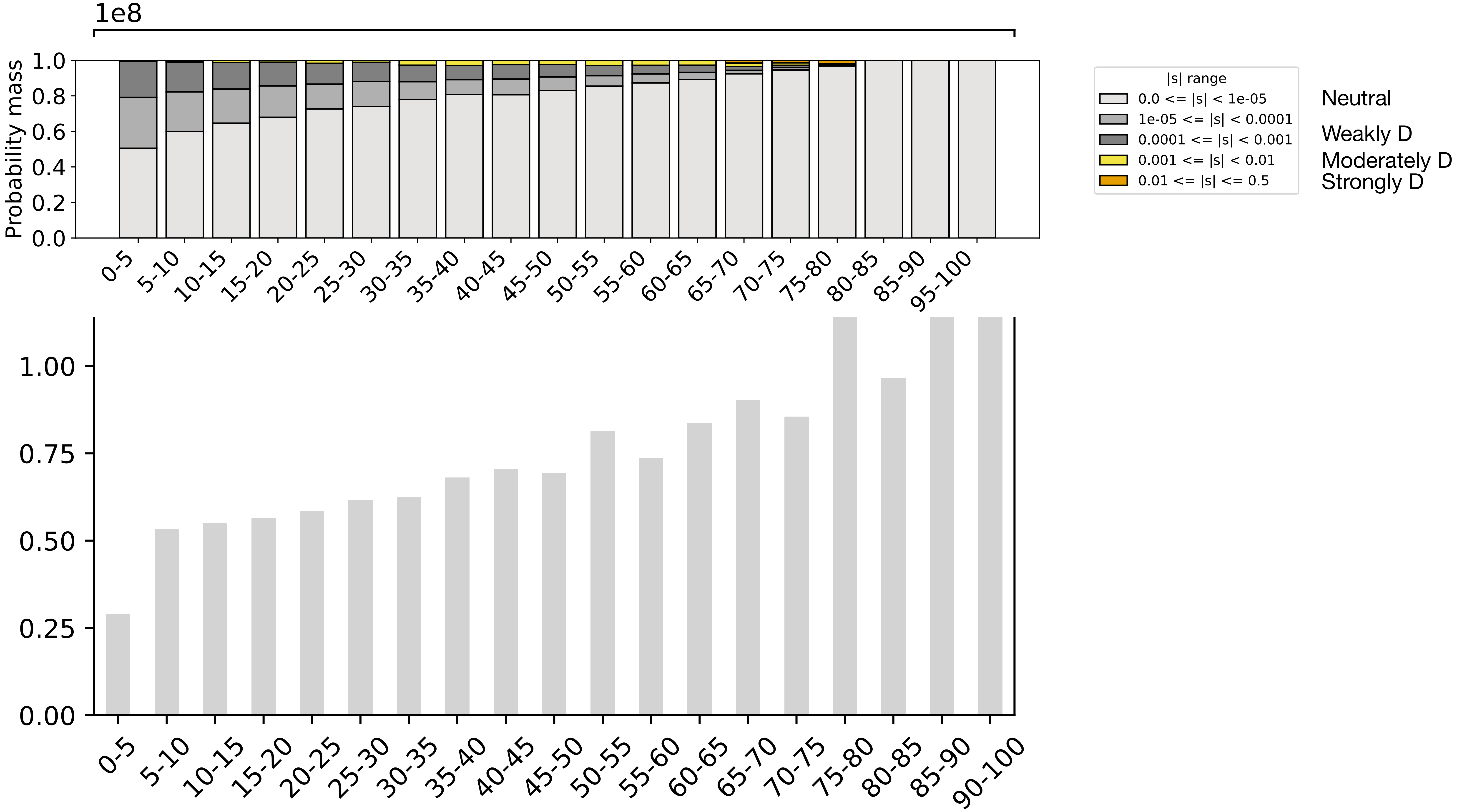
less conserved

Some non-conserved sites in mammals are conserved in primates



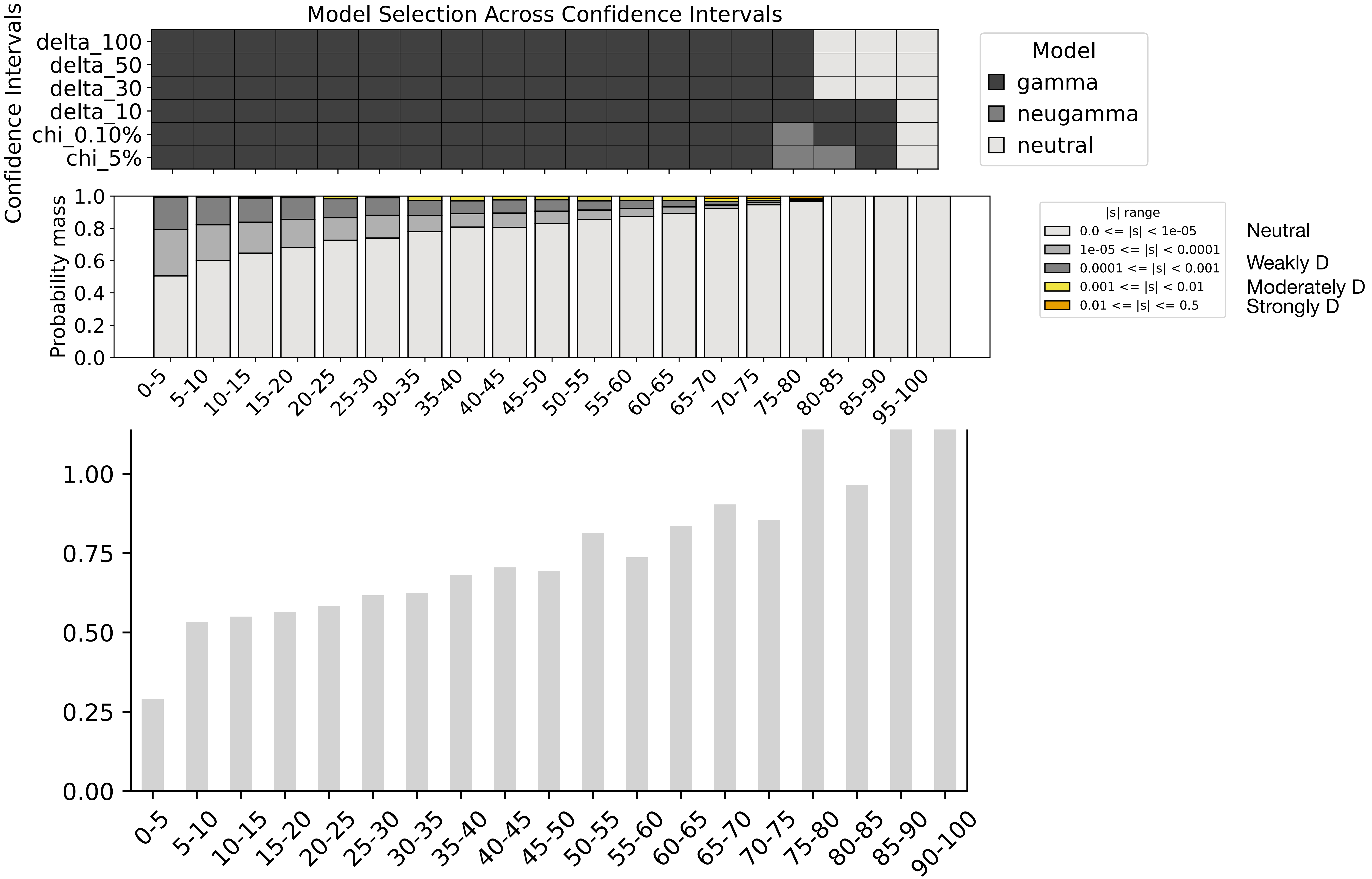
PastCons 470way score=0, Top X% PhastCons17way scores

Some non-conserved sites in mammals are conserved in primates



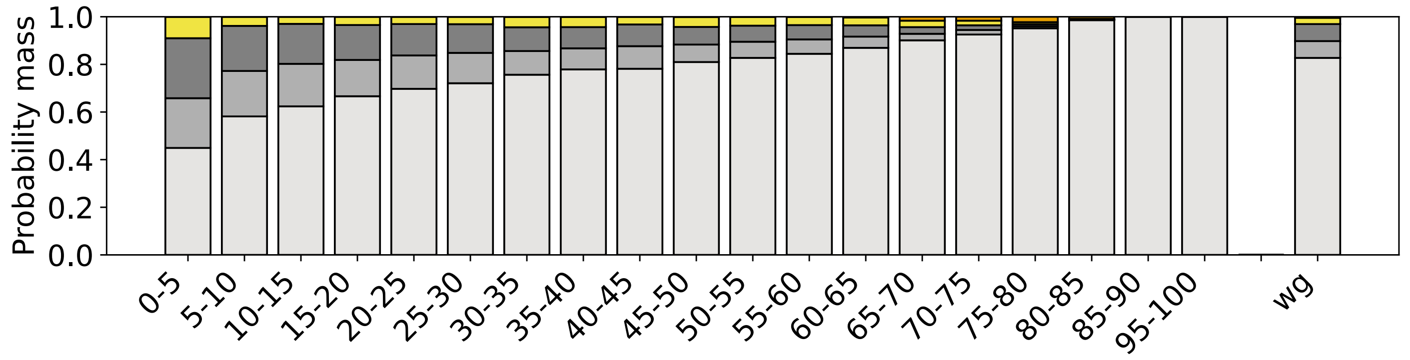
PastCons 470way score=0, Top X% PhastCons17way scores

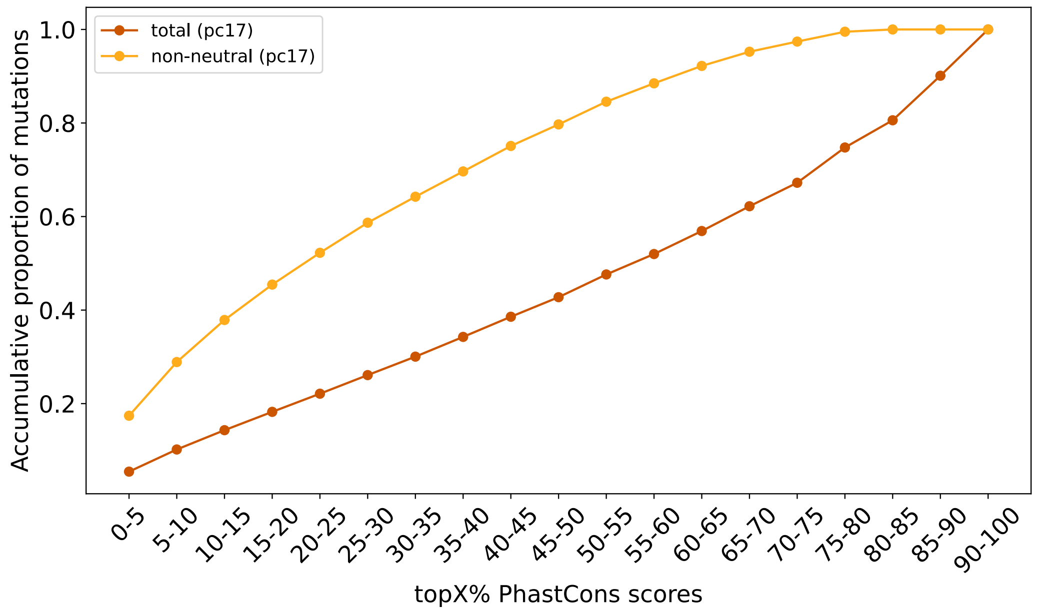
Some non-conserved sites in mammals are conserved in primates



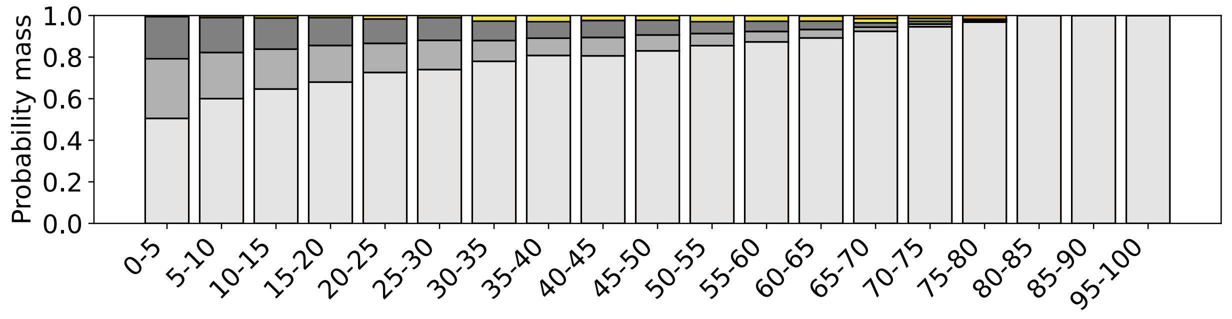
PastCons 470way score=0, Top X% PhastCons17way scores

Conclusion





non-conserved among mammals



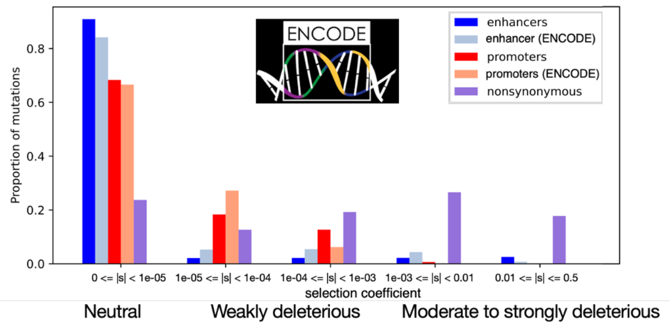
- Conserved sites —— Stronger negative selection

- Highly conserved sites only include limited deleterious mutations

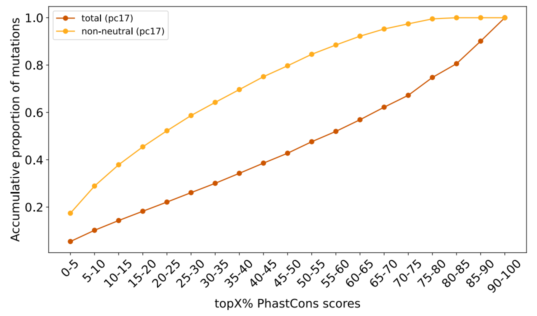
- Non-conserved in mammal but conserved in primates & negative selection in humans: A turnover of selection across evolutionary time.

Take home

- Model-based method for inferring Distribution of Fitness effects (DFE)



- DFE of functional non-coding regions



- Limited negative selection can be captured by constraint.

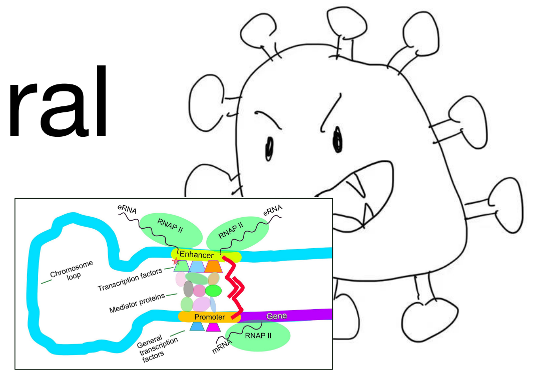
picture is from https://www.phonandroid.com/tesla-continue-de-decliner-le-cybertruck-a-toutes-les-sauces-avec-un-marteau-de-gym.html
Phylogeny is from Springer et al., 2004

Current and future work:

Kirk Lohmueller

Sriram Sankararaman

- How do regulatory elements contribute to defending against viral infections?



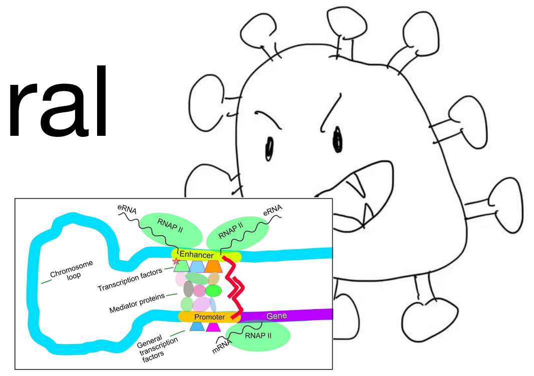
Current and future work:

Ucla

Kirk Lohmueller

Sriram Sankararaman

- How do regulatory elements contribute to defending against viral infections?

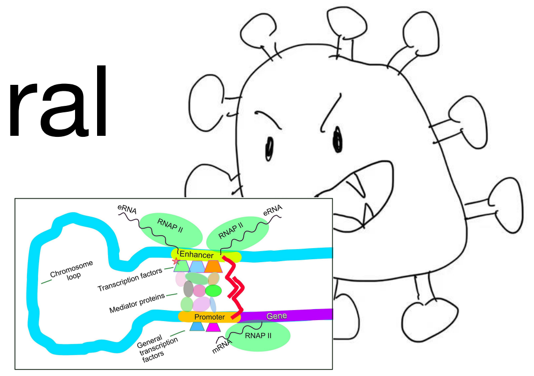


- How do proteins and REs evolve together?

Current and future work:

Kirk Lohmueller Sriram Sankararaman

- How do regulatory elements contribute to defending against viral infections?

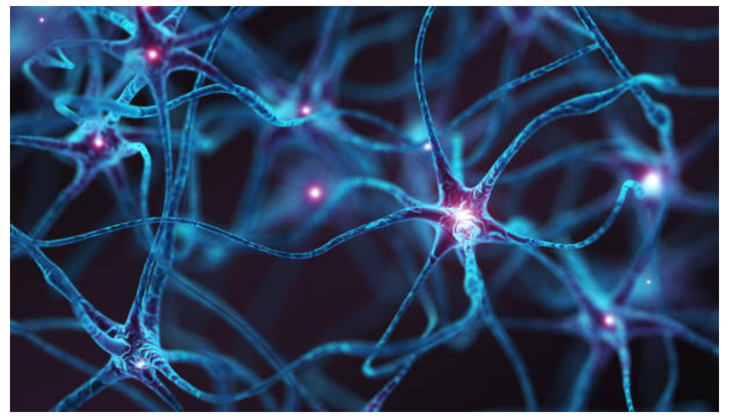


- How do proteins and REs evolve together?

Cal State

Northridge

Eduardo Guerra Amorim



Daniela Soto Jonathan Flint

- Natural selection in regulatory elements involved in different neural pathways

Gustavo Barosso Aaron Ragsdale

- Inference of background selection

Acknowledgement

Lohmueller Lab!

Kirk Lohmueller
Diana Aguilar
Sneha Chakraborty
Chris Kyriazis
Aina Martinez Zurita
Joh Mah
Swetha Ramesh

Thank yoU!

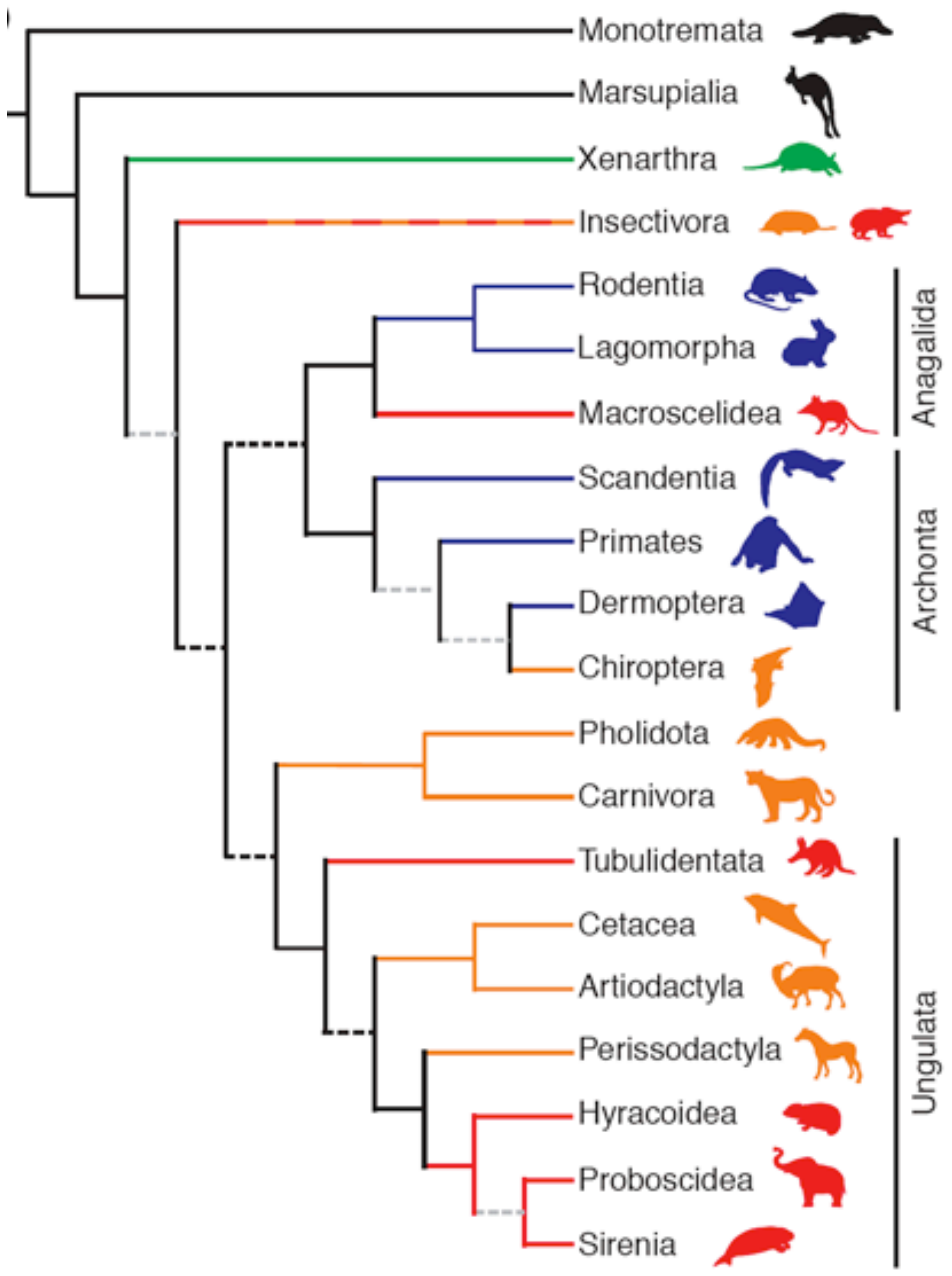
Q & A

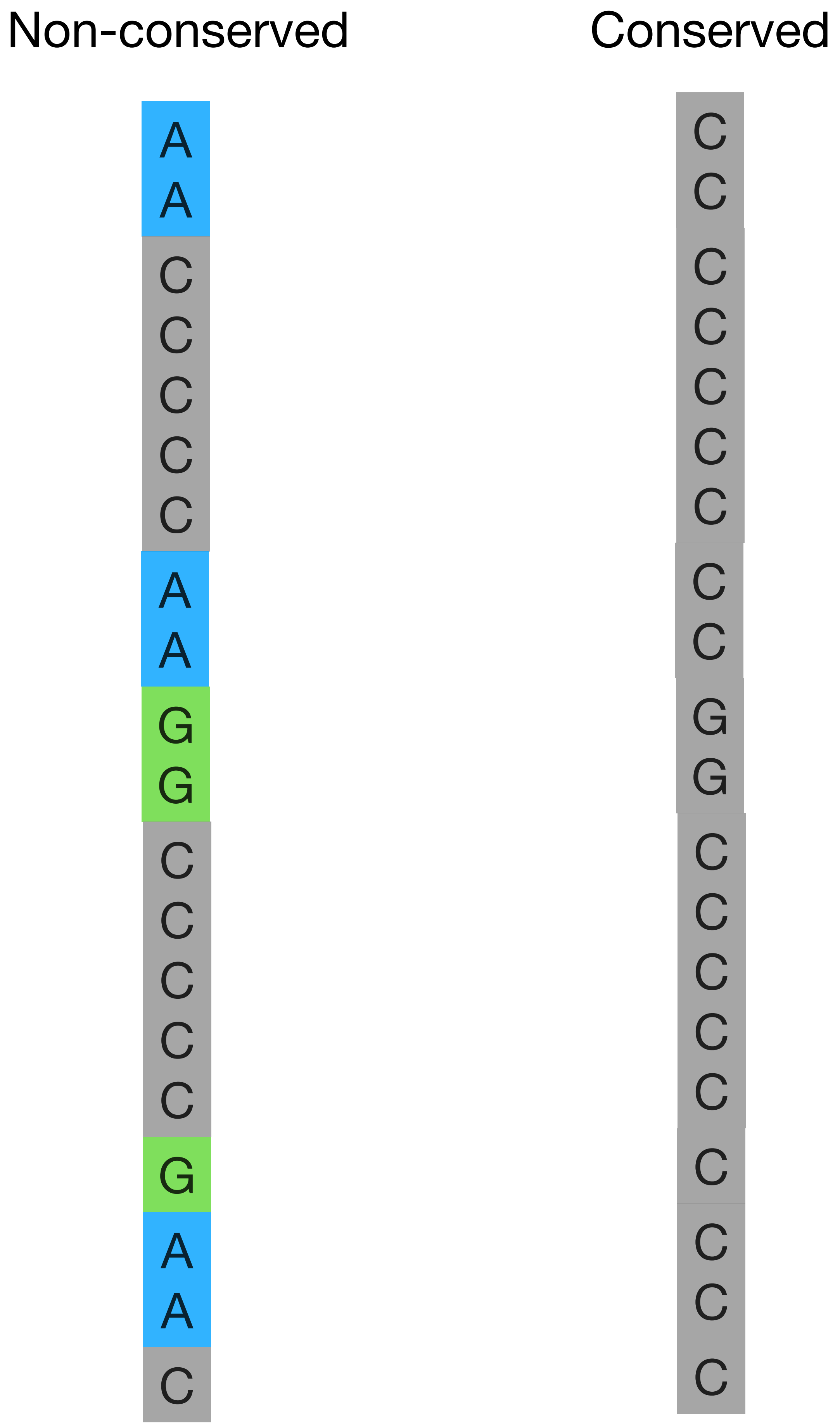
cdi@ucla.edu

X: DiChenlu

Mutations at conserved sites are more deleterious

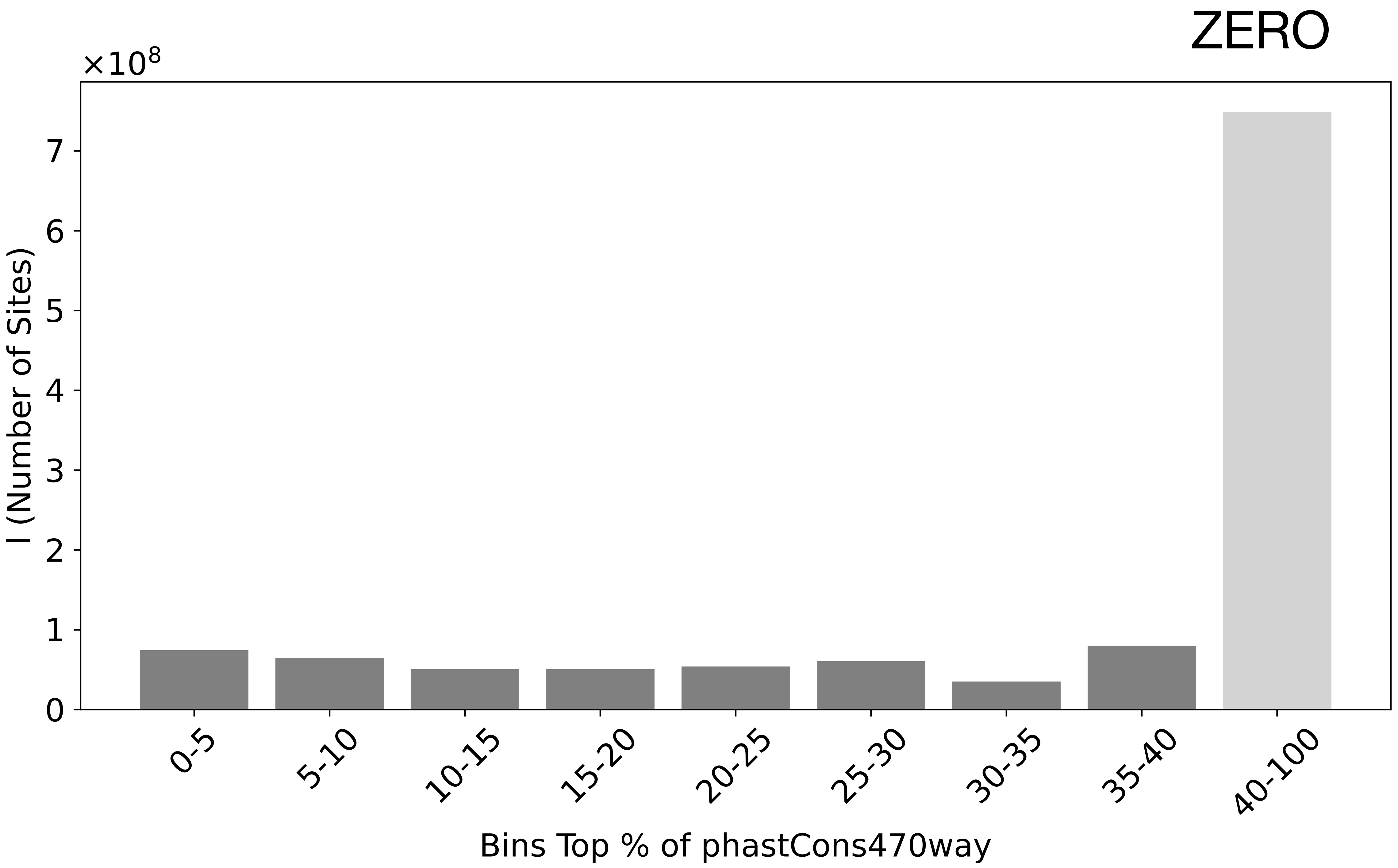
Go deeper: Mammal phylogeny

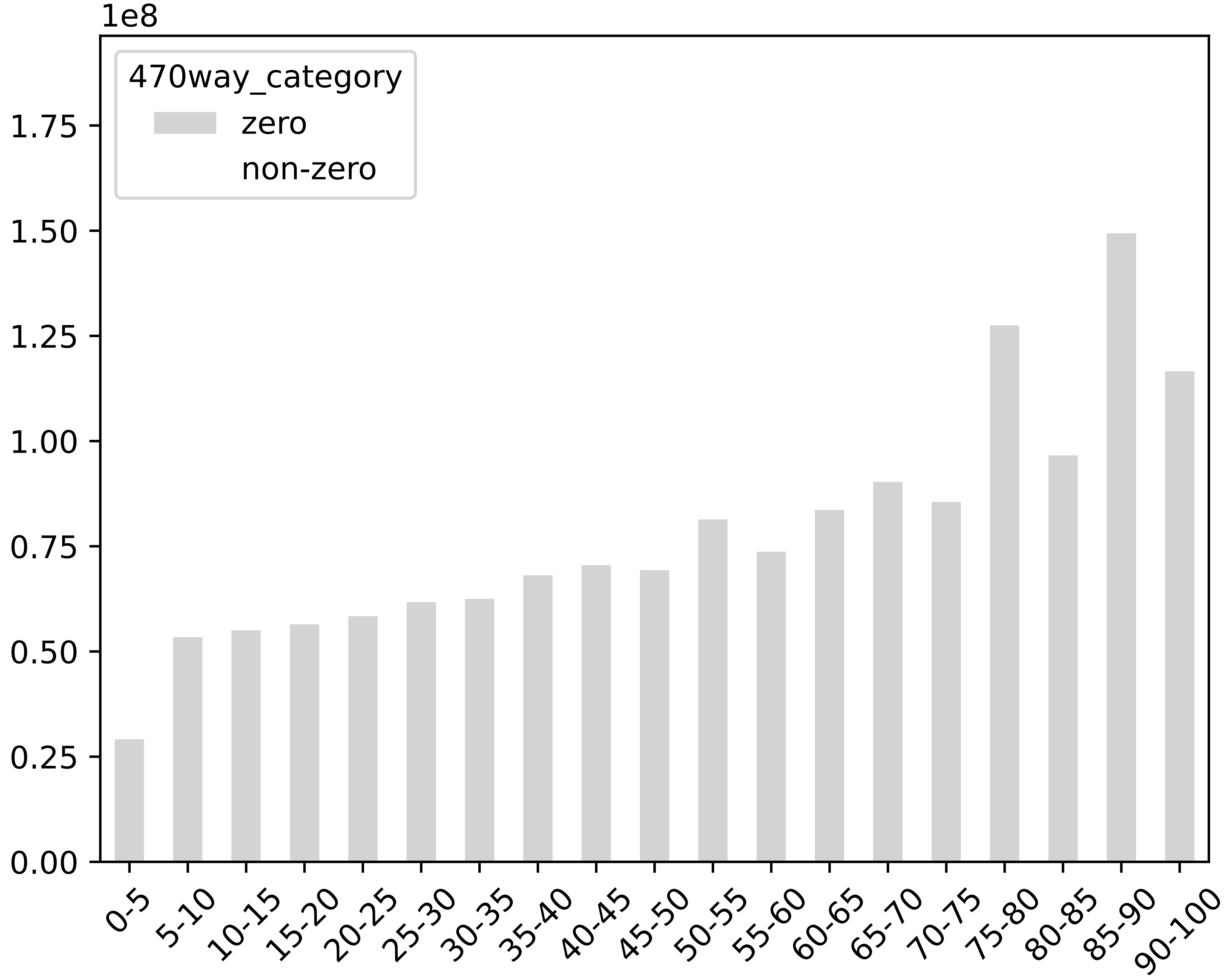




Non-conserved	Conserved
A	C
A	C
C	C
C	C
C	C
C	C
C	C
A	C
A	C
G	G
G	G
C	C
C	C
C	C
C	C
C	C
G	C
A	C
A	C
C	C

Constraint in mammals

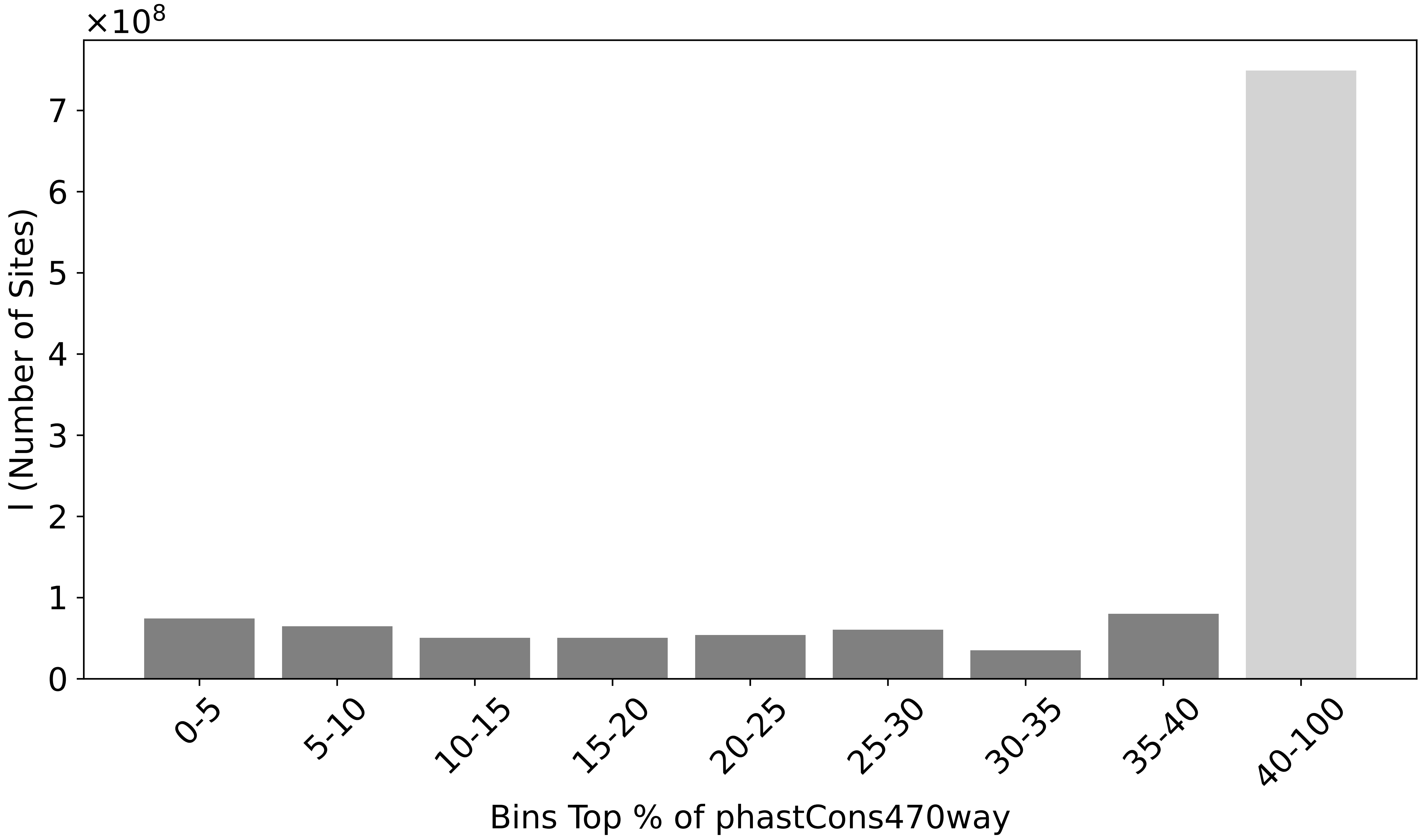


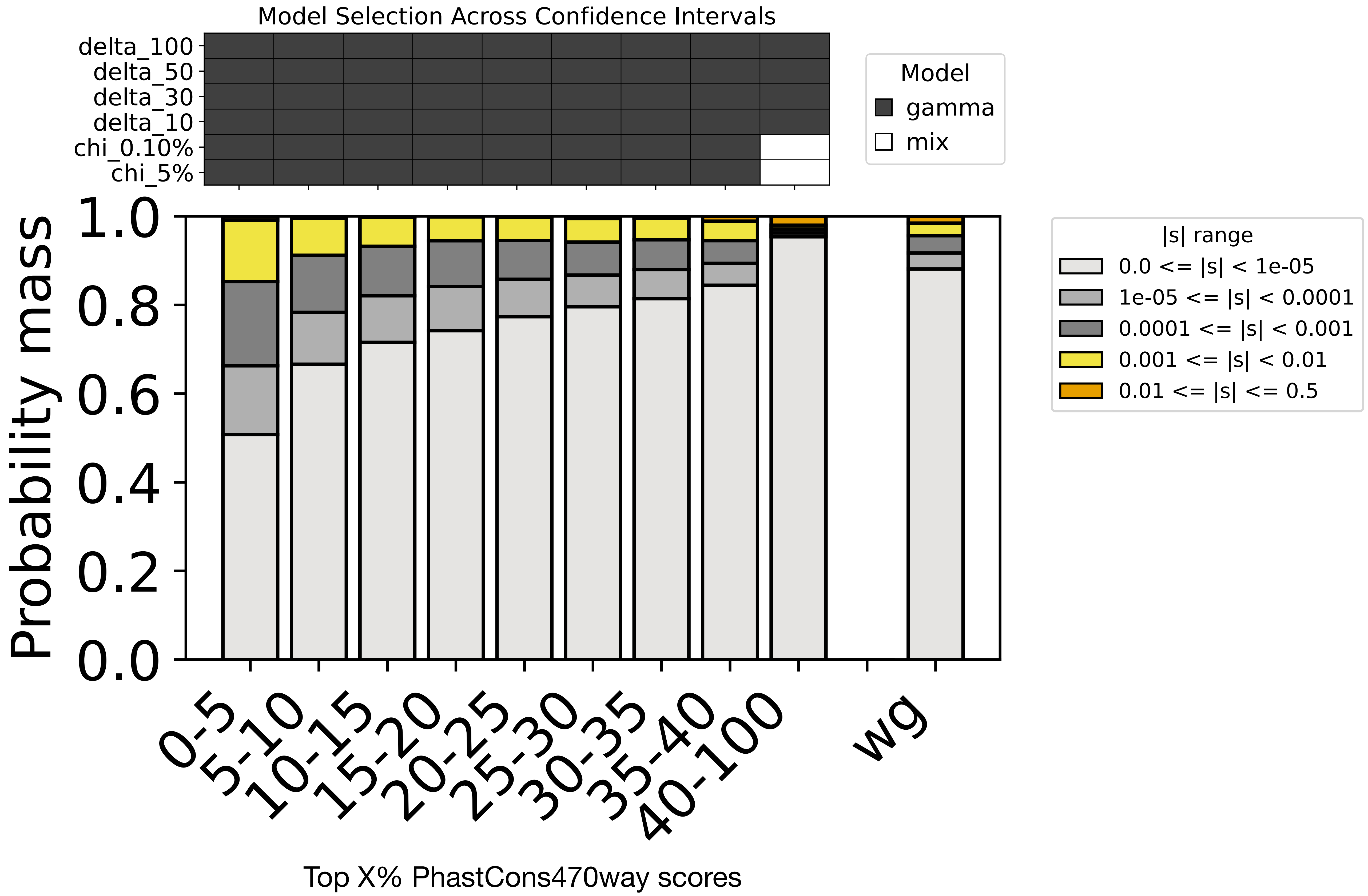


PastCons 470way score=0, Top X% PhastCons17way scores

Constraint in mammals

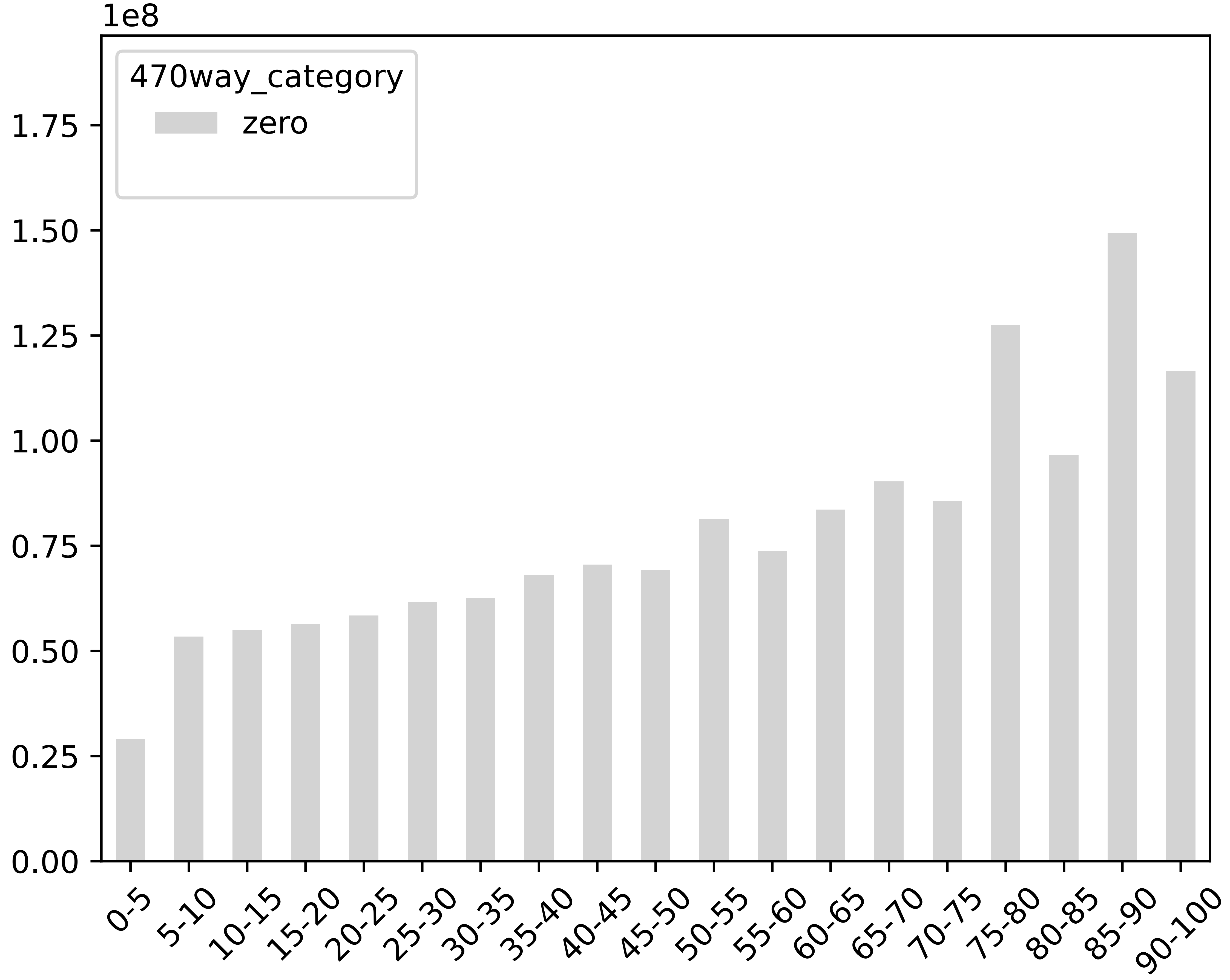
ZERO





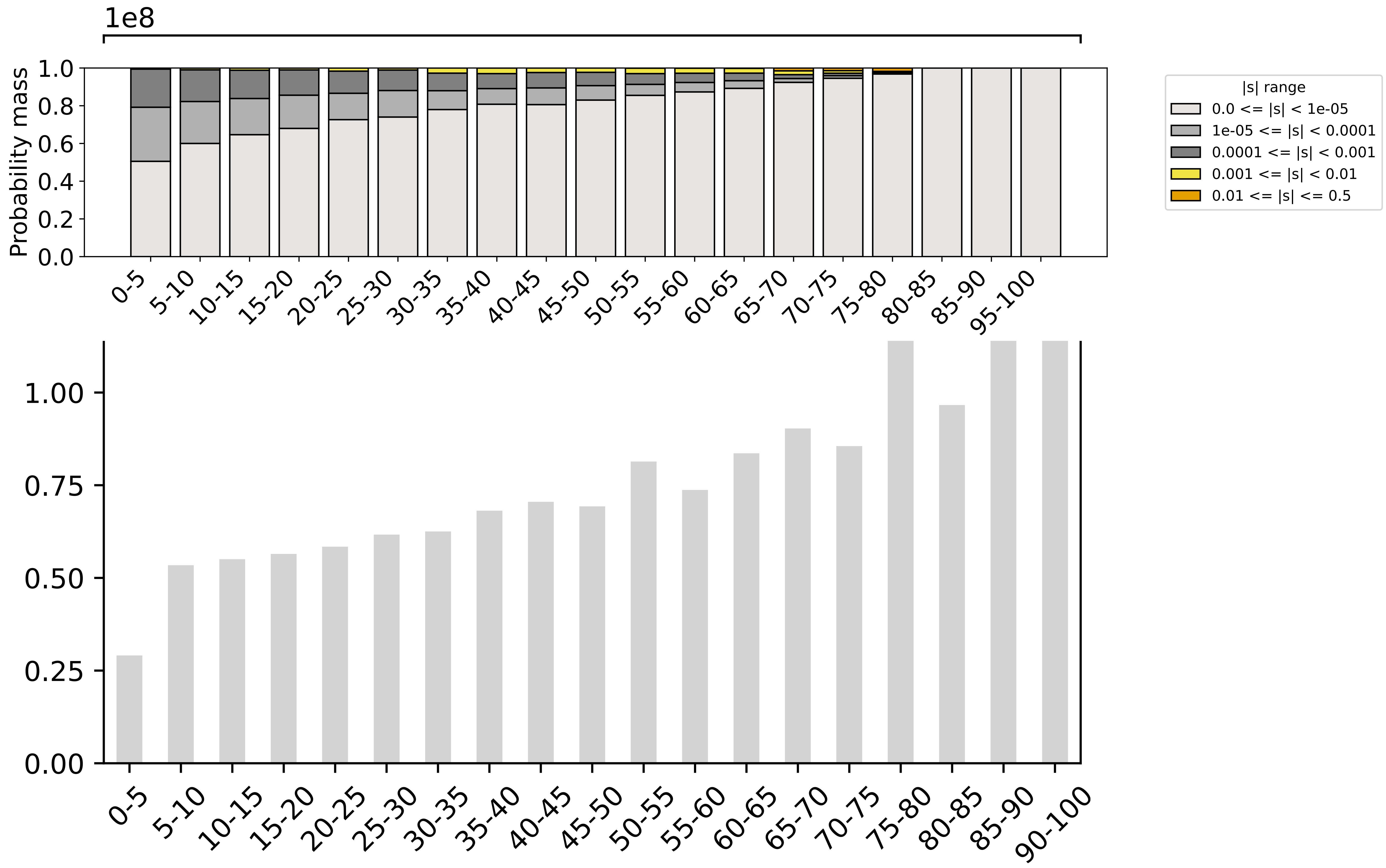
Top X% PhastCons470way scores

Constraint in mammals



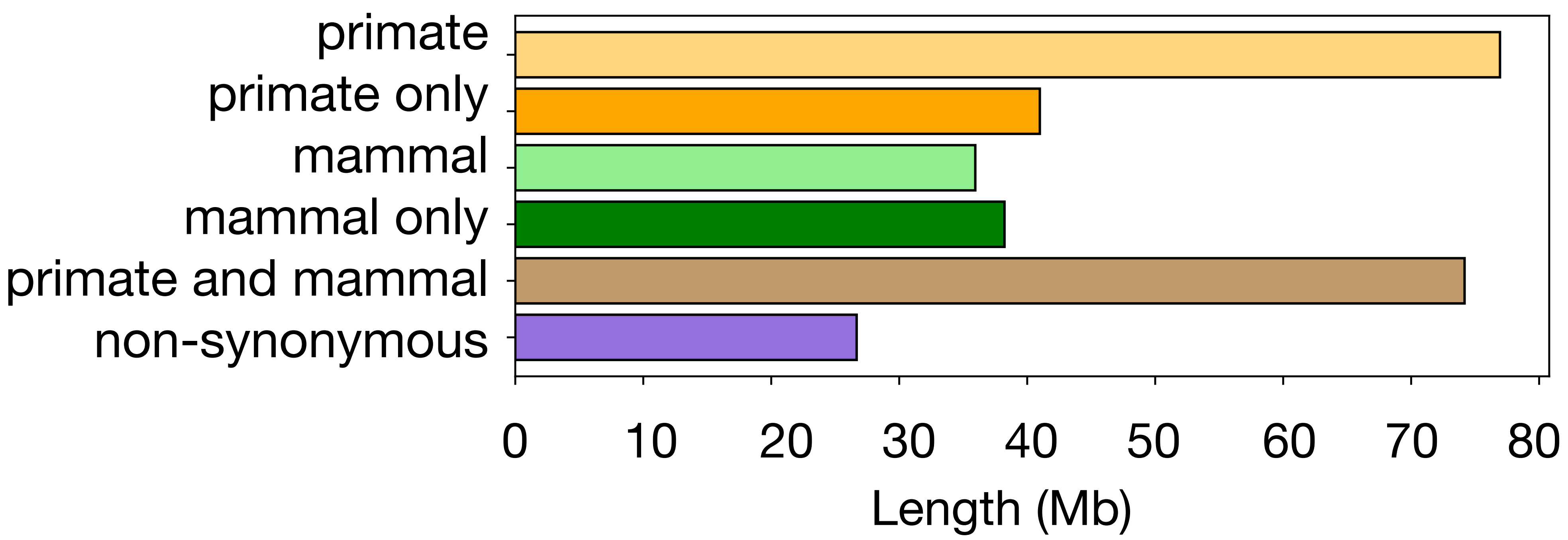
PastCons 470way score=0, Top X% PhastCons17way scores

Constraint in mammals

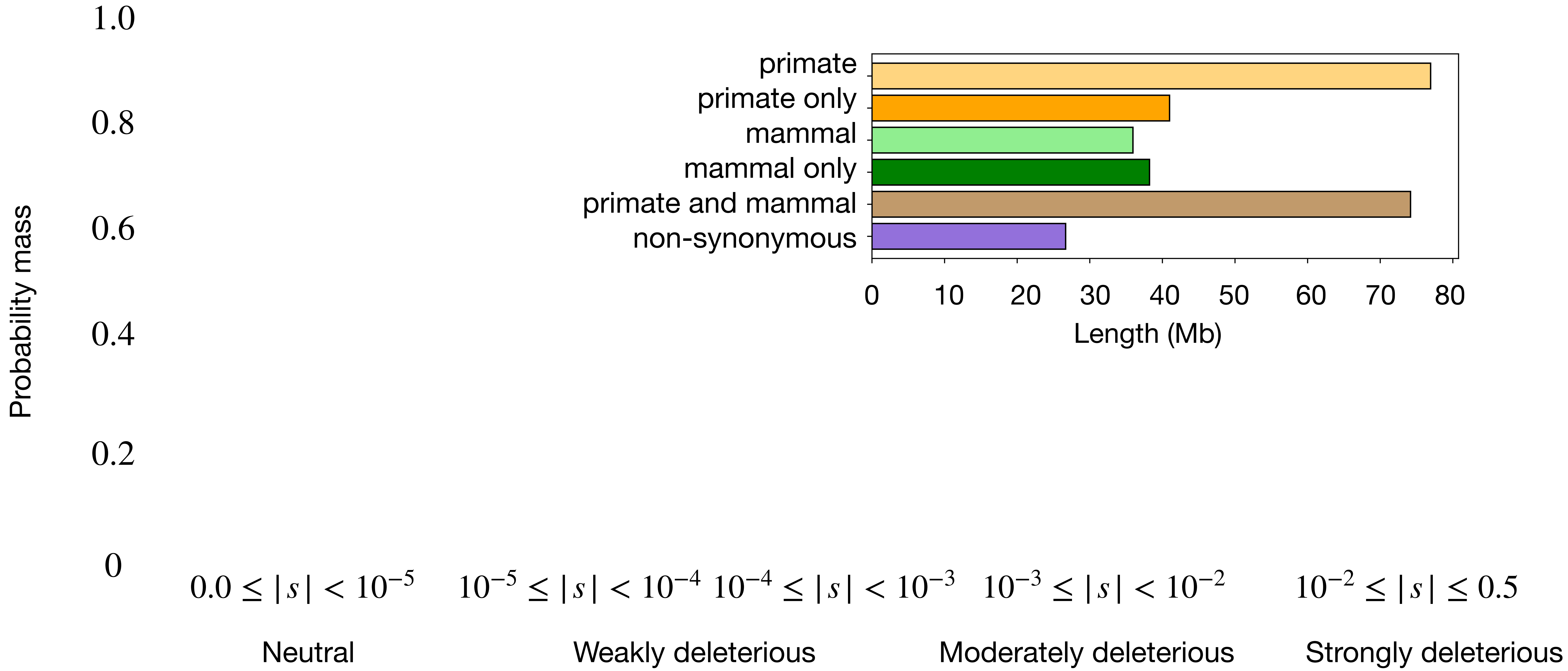


PastCons 470way score=0, Top X% PhastCons17way scores

Evidence for changing selection pressures over recent time



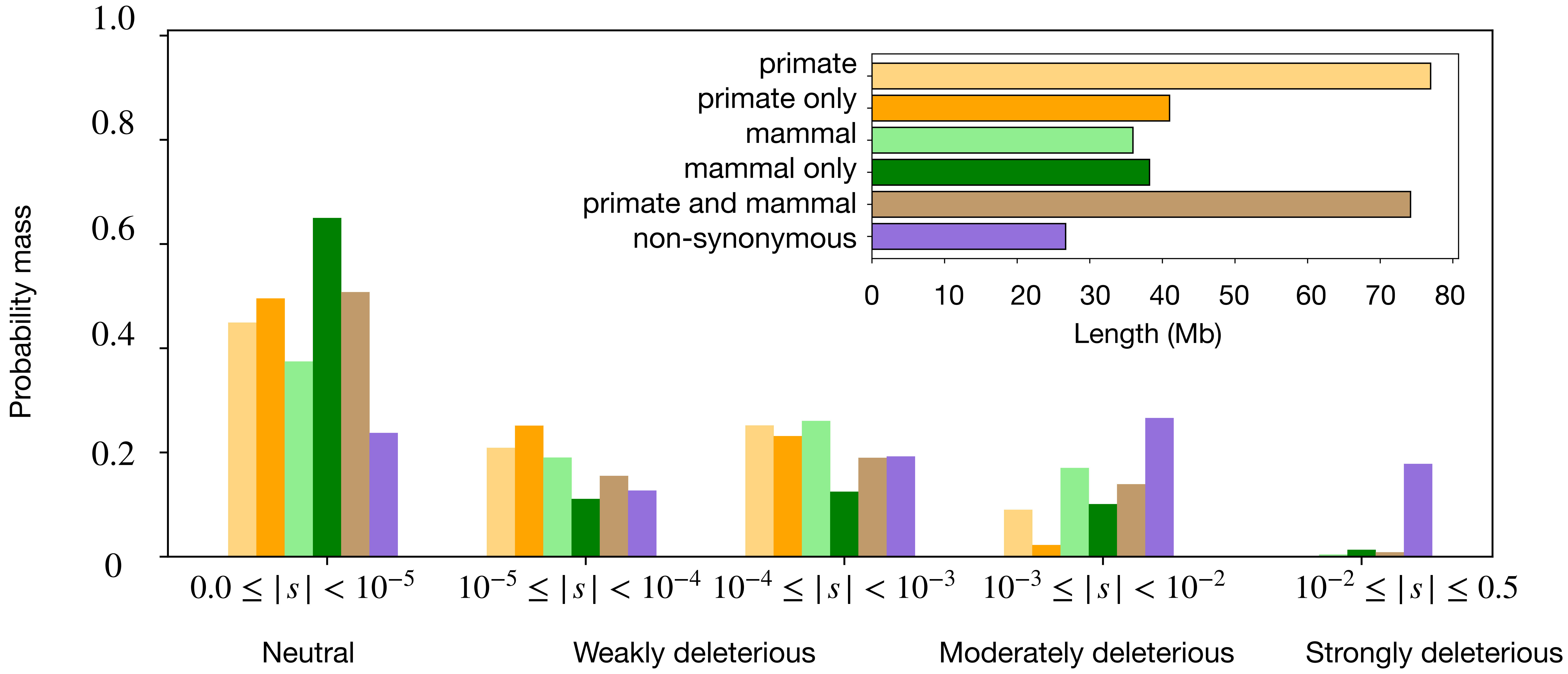
Evidence for changing selection pressures over recent time



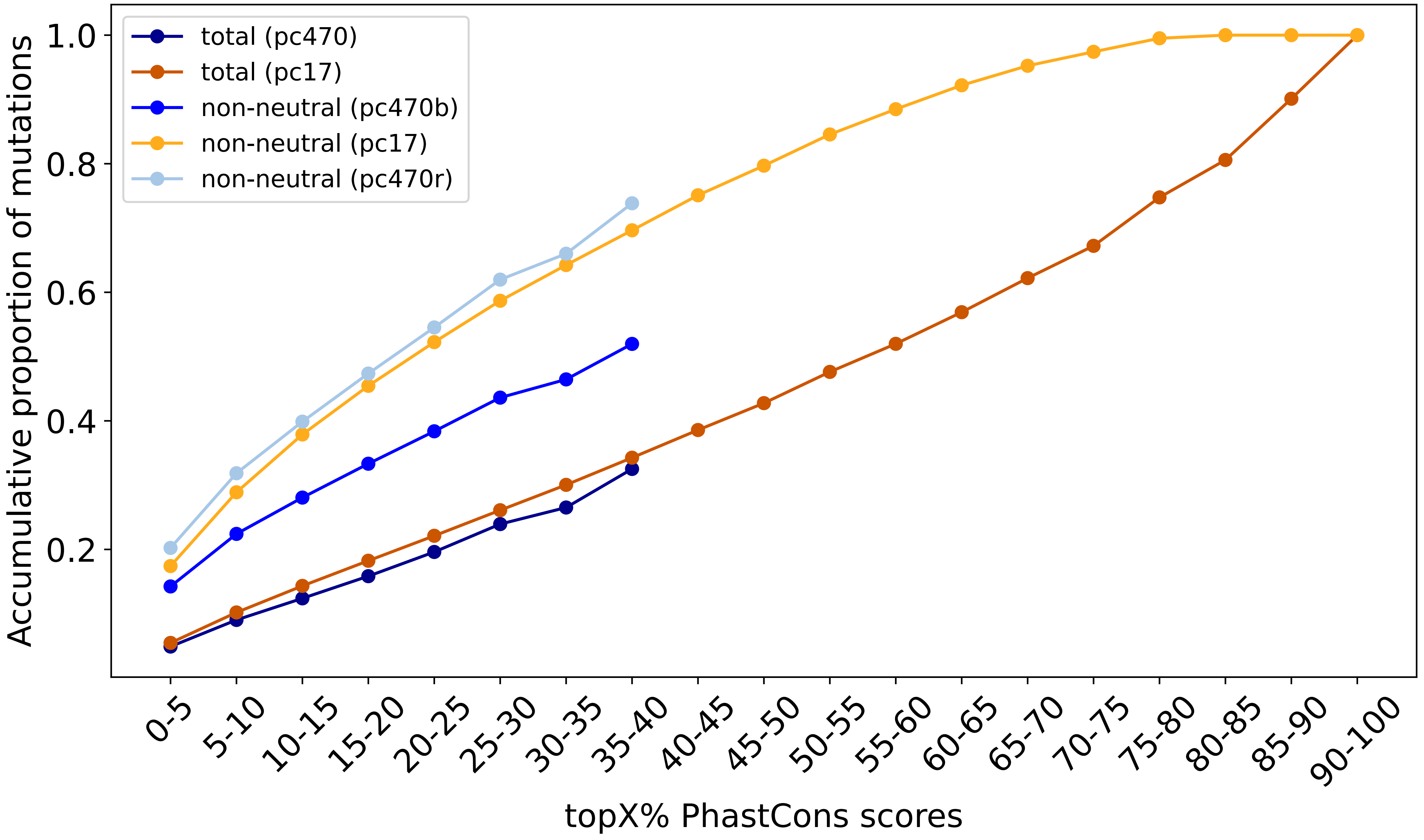
DFEs of the top 5% conserved sites

Evidence for changing selection pressures over recent time

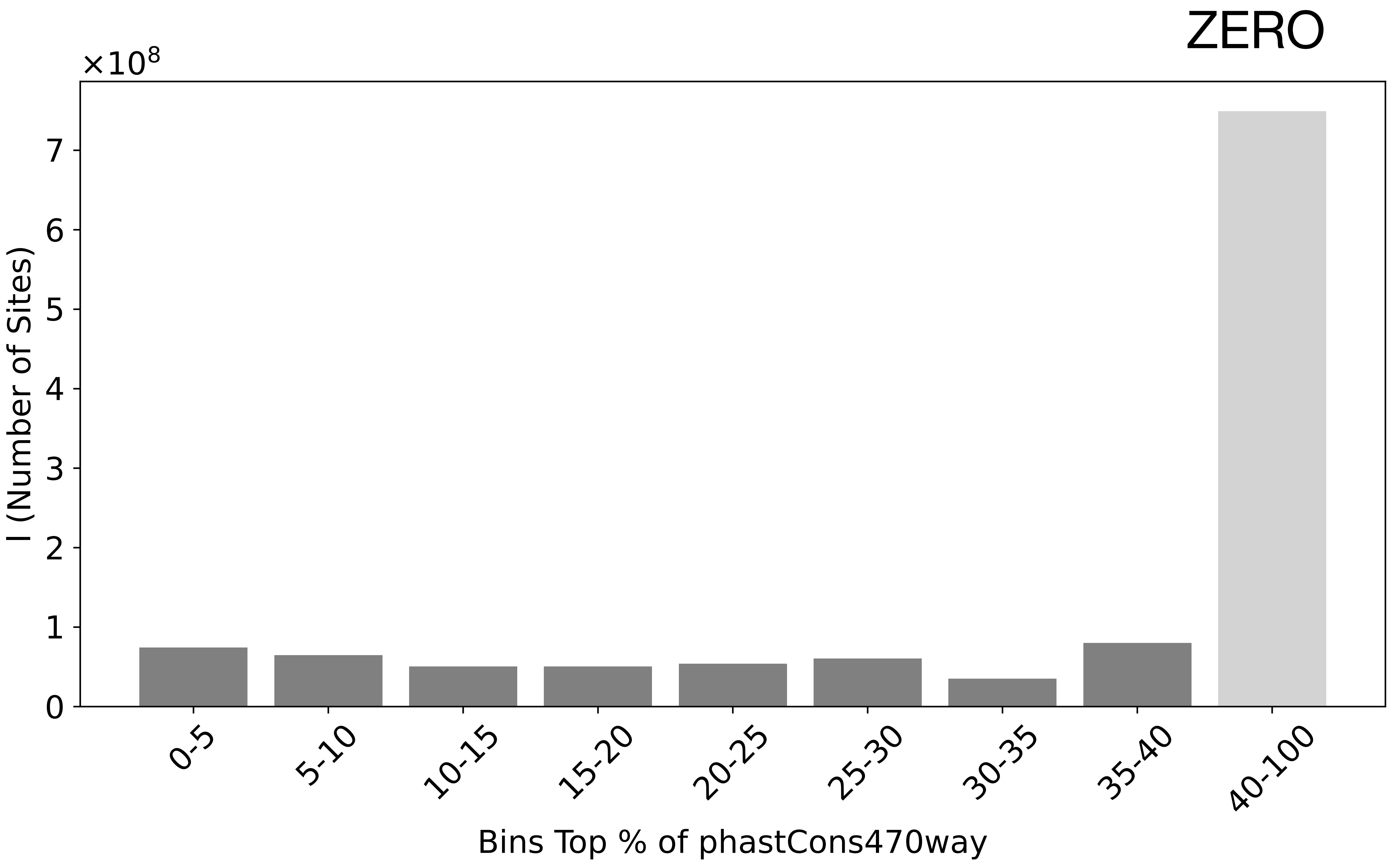
DFEs of the top 5% conserved sites



Phylogenetic constraint has limited power to identify deleterious mutations



Constraint in mammals



PastCons 470way scores