

# Supporting Information for [[main paper title]]

Aaron P. Ragsdale\*

February 28, 2021

## Supporting Information

Yet to include: - Tables with data presented in figures 5 and 6

### The diffusion equation, etc

$$\frac{\partial \psi}{\partial t} = \frac{1}{2} \sum_{1 \leq i, j \leq 3} \frac{\partial^2}{\partial x_i \partial x_j} \left[ \frac{x_i (\delta_{i=j} - x_j) \psi}{N(t)} \right] - \frac{\rho}{2} \left( \frac{\partial}{\partial x} \right)$$

### Moment equation and evolution operators

Drift

Recombination

Mutation

General selection

Moment closure via jackknife approximation

### Data analysis

Grouping Thousand Genomes populations based on clustering

Citation for Alex's work: <https://www.abstractsonline.com/pp8/#!/9070/presentation/2384>

---

\*aaronpeaceragsdale@gmail.com

## Supplementary Tables

Table S1: General selection model for diploids and dominance models.

Diploid genotype	General model	Simple dominance	Gene-based dominance
$AB / AB$	$1 + s_{AB/AB}$	$1 + 2s_A + 2s_B$	$1 + 2s$
$AB / Ab$	$1 + s_{AB/Ab}$	$1 + 2s_A + 2s_B h_B$	$1 + 2s$
$AB / aB$	$1 + s_{AB/aB}$	$1 + 2s_A h_A + 2s_B$	$1 + 2s$
$AB / ab$	$1 + s_{AB/ab}$	$1 + 2s_A h_A + 2s_B h_B$	$1 + 2sh$
$Ab / Ab$	$1 + s_{Ab/Ab}$	$1 + 2s_A$	$1 + 2s$
$Ab / aB$	$1 + s_{Ab/aB}$	$1 + 2s_A h_A + 2s_B h_B$	$1 + 2s$
$Ab / ab$	$1 + s_{Ab/ab}$	$1 + 2s_A h_A$	$1 + 2sh$
$aB / aB$	$1 + s_{aB/aB}$	$1 + 2s_B$	$1 + 2s$
$aB / ab$	$1 + s_{aB/ab}$	$1 + 2s_B h_B$	$1 + 2sh$
$ab / ab$	1	1	1

Table S2: Haploid epistasis model.

Haplotype	Fitness
$AB$	$(1 + s_A + s_B)(1 + \epsilon)$
$Ab$	$1 + s_A$
$aB$	$1 + s_B$
$ab$	1

Table S3: Thousand Genomes Project population descriptions for populations used in this study.

Code	Description	Region
ESN	Esan in Nigeria	Africa
GWD	Gambian in Western Divisions in the Gambia	Africa
LWK	Luhya in Webuye, Kenya	Africa
MSL	Mende in Sierra Leone	Africa
YRI	Yoruba in Ibadan, Nigeria	Africa
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	Europe
GBR	British in England and Scotland	Europe
FIN	Finnish in Finland	Europe
IBS	Iberian Population in Spain	Europe
TSI	Toscani in Italia	Europe
CDX	Chinese Dai in Xishuangbanna, China	East Asia
CHB	Han Chinese in Beijing, China	East Asia
CHS	Southern Han Chinese	East Asia
JPT	Japanese in Tokyo, Japan	East Asia
KHV	Kinh in Ho Chi Minh City, Vietnam	East Asia

Table S4: Tamija's  $D$  for classes of coding mutations, both within annotated domains and outside of domains.

Population	Mutation type	Region	Tajima's $D$
ESN	Synonymous	All	-0.882
		In domain	-0.854
		Not in domain	-0.921
	Missense	All	-1.414
		In domain	-1.535
		Not in domain	-1.293
	Loss of function	All	-1.483
		In domain	-2.156
		Not in domain	-1.282
GWD	Synonymous	All	-1.011
		In domain	-0.981
		Not in domain	-1.052
	Missense	All	-1.566
		In domain	-1.678
		Not in domain	-1.452
	Loss of function	All	-1.697
		In domain	-2.328
		Not in domain	-1.501
LWK	Synonymous	All	-1.109
		In domain	-1.088
		Not in domain	-1.139
	Missense	All	-1.589
		In domain	-1.700
		Not in domain	-1.477
	Loss of function	All	-1.666
		In domain	-2.278
		Not in domain	-1.477
MSL	Synonymous	All	-0.983
		In domain	-0.959
		Not in domain	-1.017
	Missense	All	-1.501
		In domain	-1.603
		Not in domain	-1.400
	Loss of function	All	-1.559
		In domain	-2.303
		Not in domain	-1.332
YRI	Synonymous	All	-0.928
		In domain	-0.898
		Not in domain	-0.971
	Missense	All	-1.467
		In domain	-1.586
		Not in domain	-1.348
	Loss of function	All	-1.624
		In domain	-2.237
		Not in domain	-1.424
CEU	Synonymous	All	-0.417
		In domain	-0.392

Table S4: Tamija's  $D$  for classes of coding mutations, both within annotated domains and outside of domains. (*continued*)

Population	Mutation type	Region	Tajima's $D$
FIN	Missense	Not in domain	-0.452
		All	-1.248
		In domain	-1.404
	Loss of function	Not in domain	-1.082
		All	-1.501
		In domain	-2.196
	Synonymous	Not in domain	-1.280
		All	-0.058
		In domain	-0.047
	Missense	Not in domain	-0.075
		All	-0.883
		In domain	-1.048
	Loss of function	Not in domain	-0.710
		All	-1.200
		In domain	-2.034
	Synonymous	Not in domain	-0.906
		All	-0.319
		In domain	-0.300
GBR	Missense	Not in domain	-0.345
		All	-1.120
		In domain	-1.276
	Loss of function	Not in domain	-0.954
		All	-1.313
		In domain	-2.178
	Synonymous	Not in domain	-0.997
		All	-0.689
		In domain	-0.664
	Missense	Not in domain	-0.724
		All	-1.424
		In domain	-1.560
IBS	Loss of function	Not in domain	-1.279
		All	-1.636
		In domain	-2.349
	Synonymous	Not in domain	-1.378
		All	-0.650
		In domain	-0.625
	Missense	Not in domain	-0.685
		All	-1.422
		In domain	-1.568
	Loss of function	Not in domain	-1.266
		All	-1.655
		In domain	-2.349
TSI	Synonymous	Not in domain	-1.397
		All	-0.374
		In domain	-0.366
	Missense	Not in domain	-0.385
		All	-1.179
		In domain	-0.366
CDX	Synonymous	Not in domain	-0.385
		All	-1.179
		In domain	-0.366
	Missense	Not in domain	-0.385
		All	-1.179
		In domain	-0.366

Table S4: Tamija's  $D$  for classes of coding mutations, both within annotated domains and outside of domains. (*continued*)

Population	Mutation type	Region	Tajima's D
CHB	Loss of function	In domain	-1.323
		Not in domain	-1.026
		All	-1.360
		In domain	-2.194
		Not in domain	-1.062
	Synonymous	All	-0.598
		In domain	-0.593
		Not in domain	-0.606
	Missense	All	-1.389
		In domain	-1.528
		Not in domain	-1.239
	Loss of function	All	-1.586
In domain		-2.344	
Not in domain		-1.298	
CHS	Synonymous	All	-0.544
		In domain	-0.545
		Not in domain	-0.544
	Missense	All	-1.334
		In domain	-1.499
		Not in domain	-1.150
	Loss of function	All	-1.559
		In domain	-2.290
		Not in domain	-1.292
JPT	Synonymous	All	-0.371
		In domain	-0.368
		Not in domain	-0.376
	Missense	All	-1.194
		In domain	-1.355
		Not in domain	-1.019
	Loss of function	All	-1.410
		In domain	-2.272
		Not in domain	-1.086
KHV	Synonymous	All	-0.576
		In domain	-0.562
		Not in domain	-0.596
	Missense	All	-1.346
		In domain	-1.473
		Not in domain	-1.210
	Loss of function	All	-1.535
		In domain	-2.294
		Not in domain	-1.269

## Supplementary Figures

- 1) Accuracy of jackknife for varying sample size
- 2) Repeat of figures 2, 3, and 4 but for average  $D$  instead of  $\sigma_d^2$
- 3)

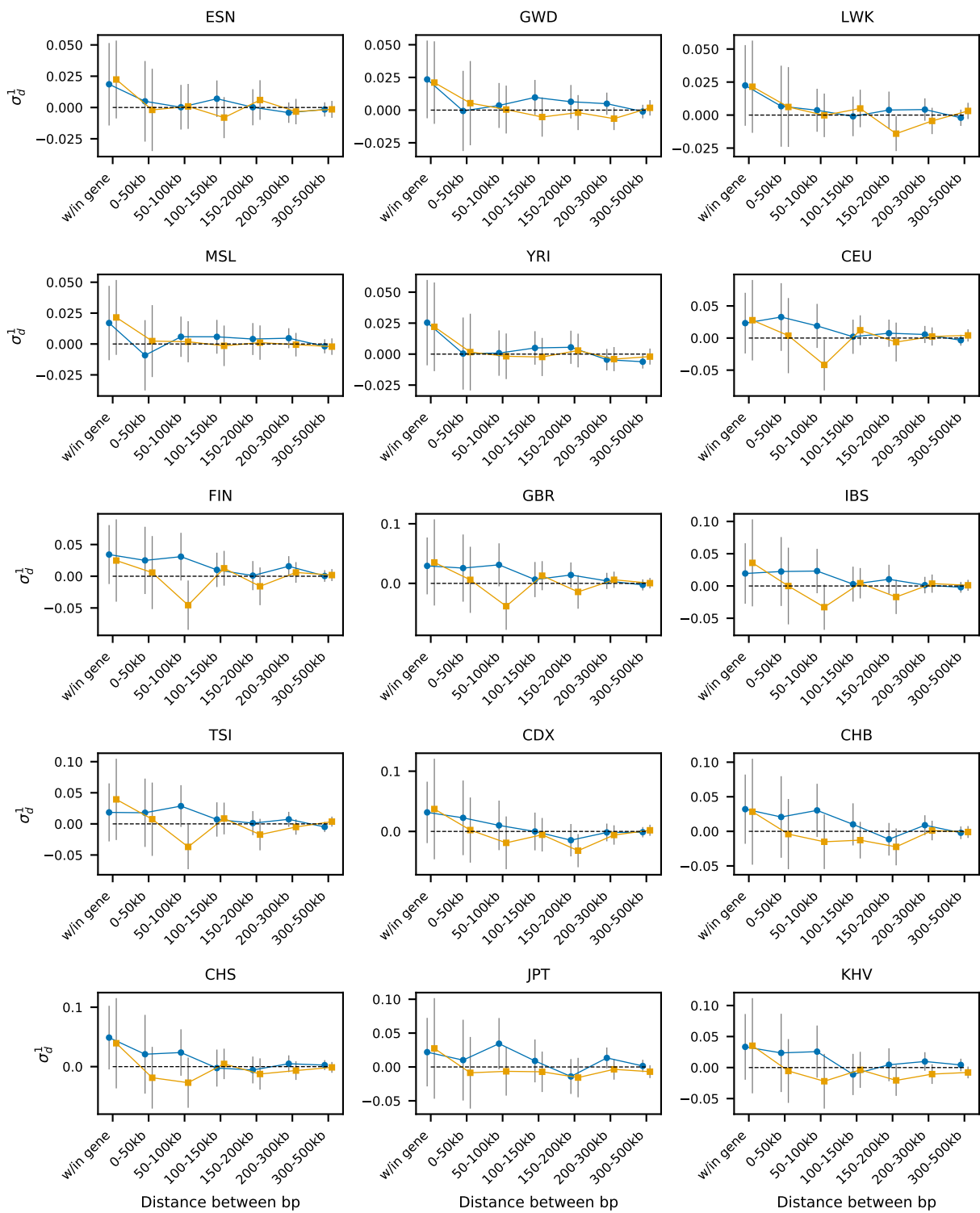


Figure S1: **LD within and between protein-coding genes.** This is the caption.

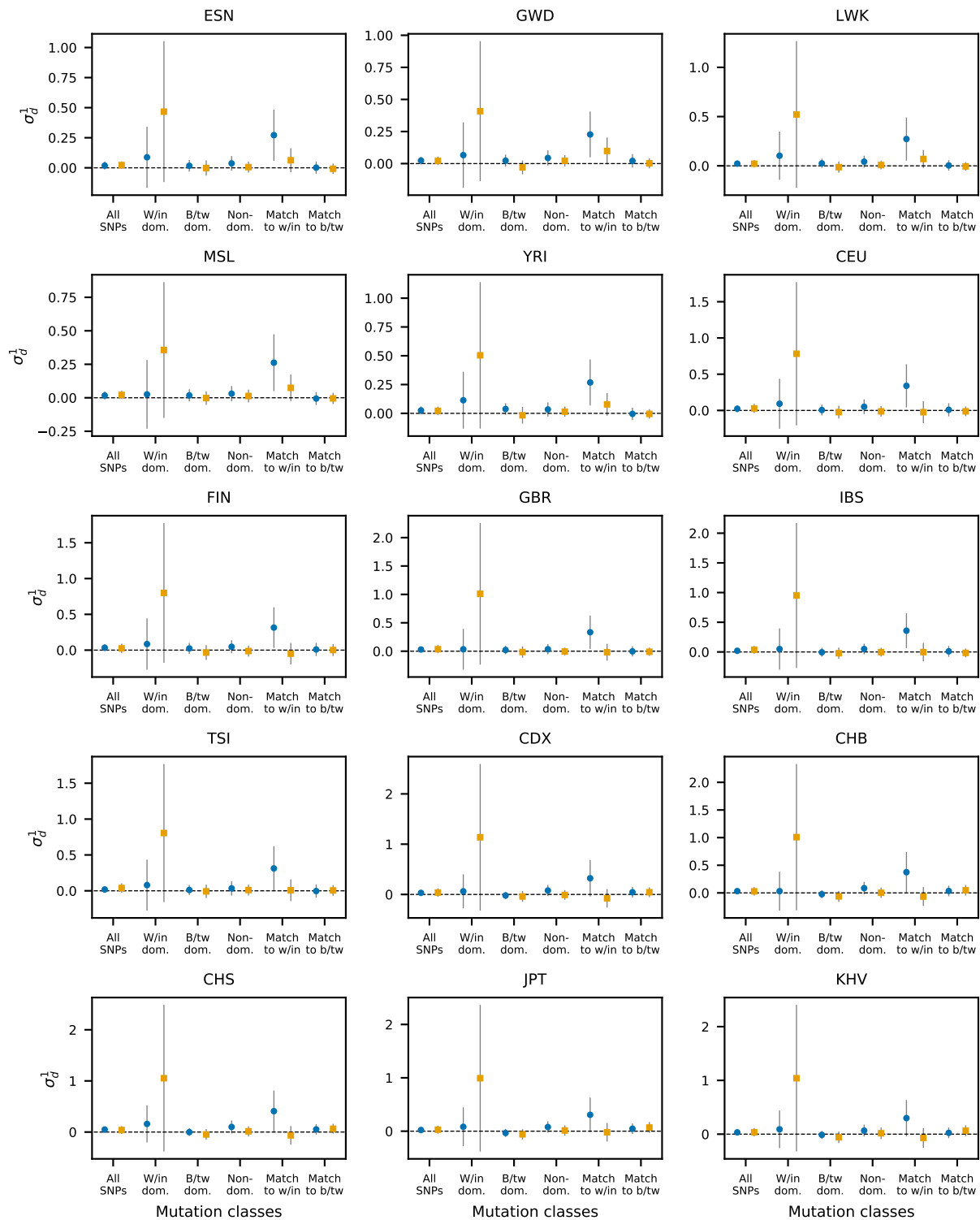


Figure S2: LD within and between coding domains and pairs outside domains at matched distances. This is the caption.



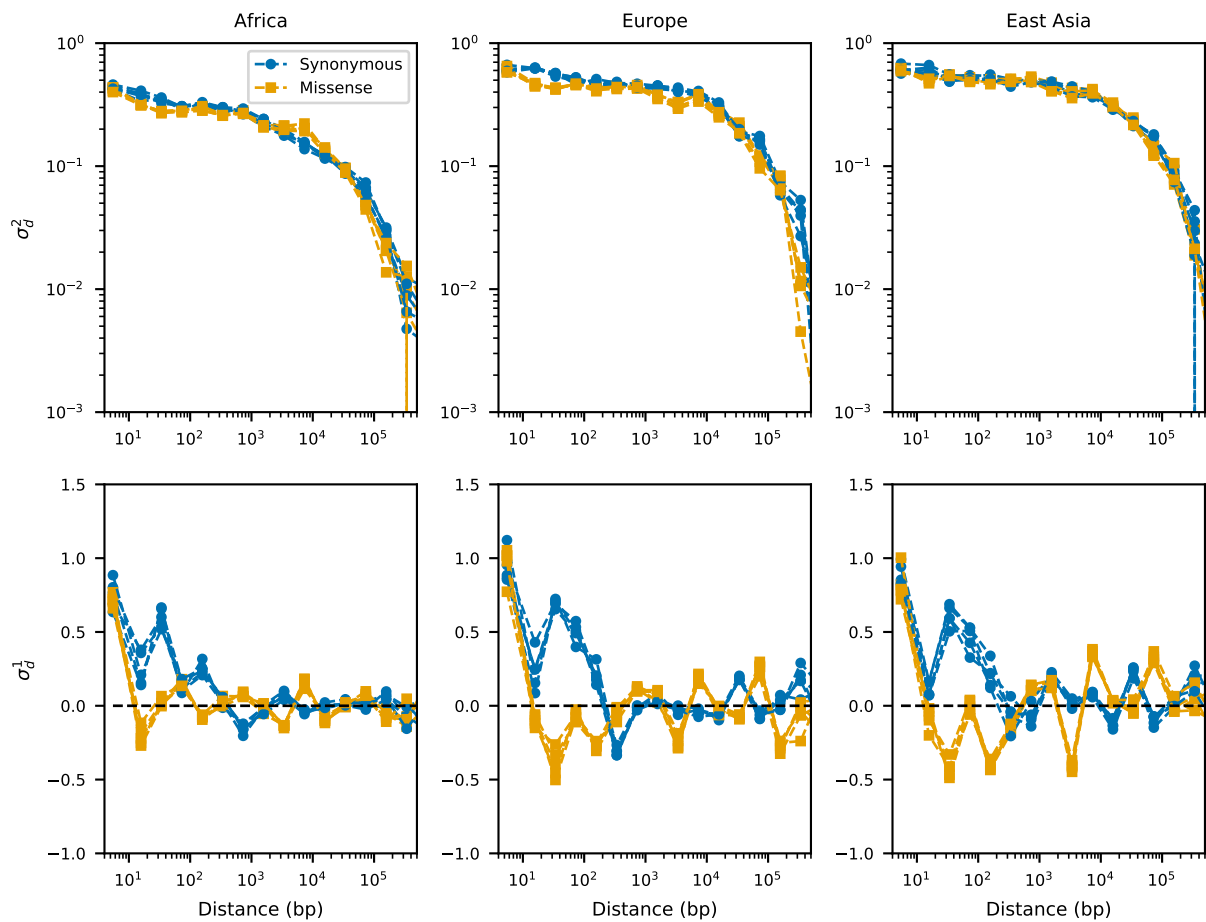


Figure S3: LD decay for synonymous and missense mutations for pairs of mutations that fall outside of domains.

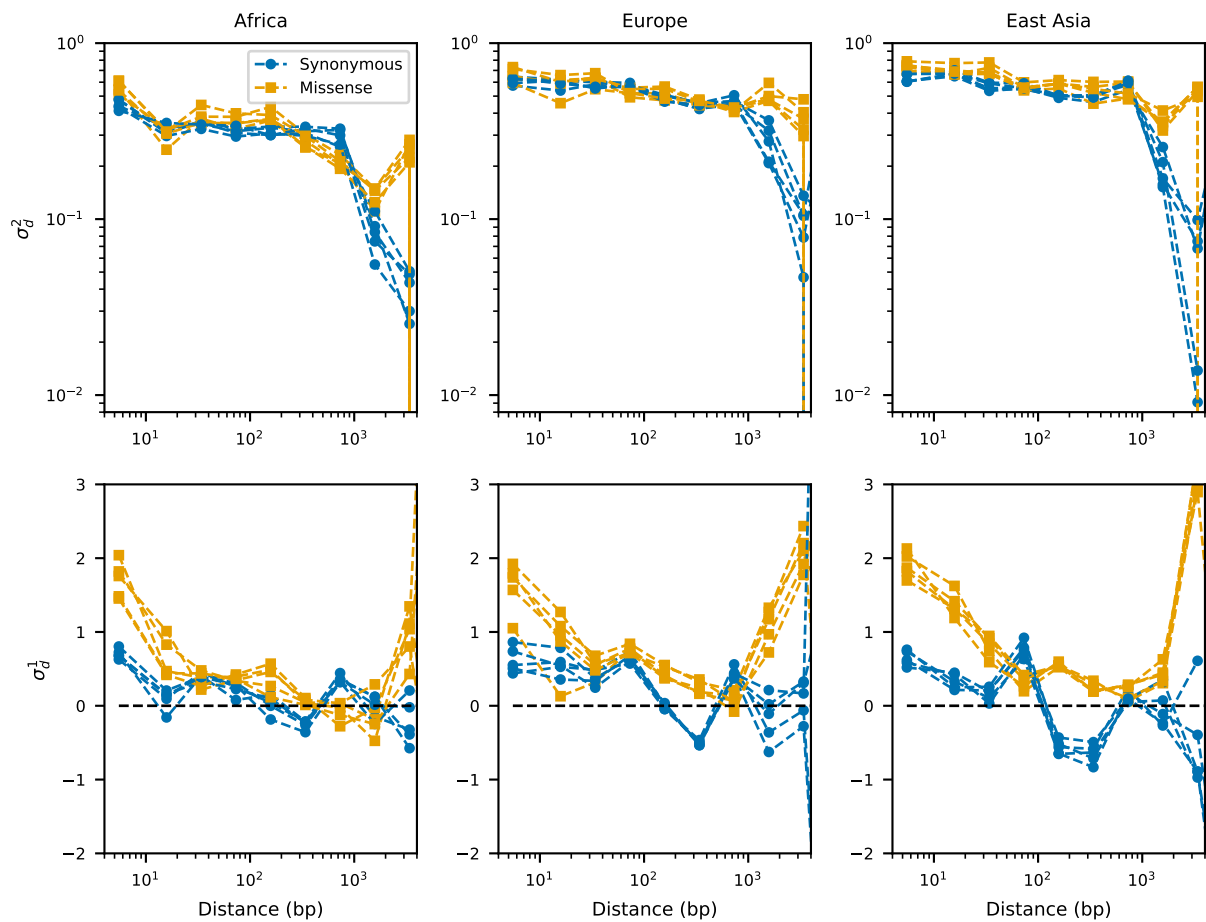


Figure S4: LD decay for synonymous and missense mutations for pairs of mutations that fall inside the same domains.

## Supporting References