

Supplemental material for:

Biased gene conversion constrains adaptation in *Arabidopsis thaliana*

Tuomas Härmälä^{*,1} and Peter Tiffin^{*}

^{*} Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN

¹ thamala@umn.edu

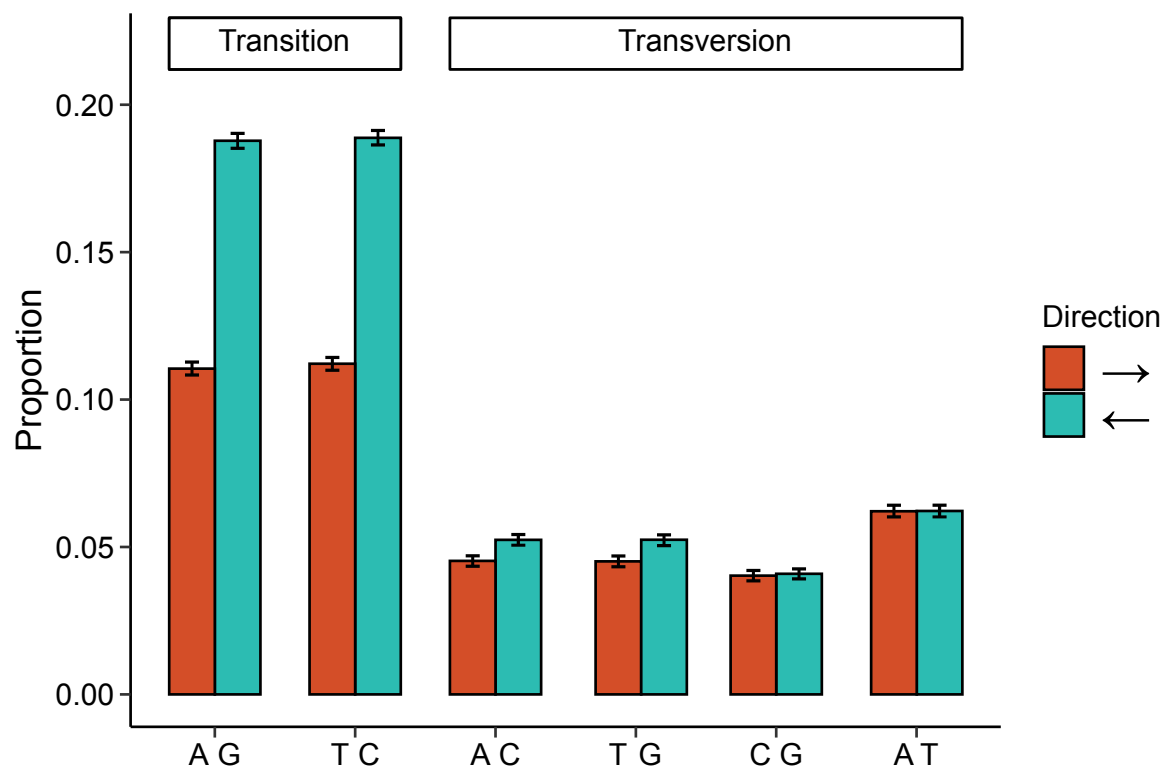


Figure S1. The proportion of different nucleotide changes from ancestral (*A. lyrata*, *C. rubella*, and *A. alpine*) to derived allele in *A. thaliana*. Colors show the direction of change (e.g. for “A G”, red is A → G and teal is A ← G). Error bars show 95% CIs.

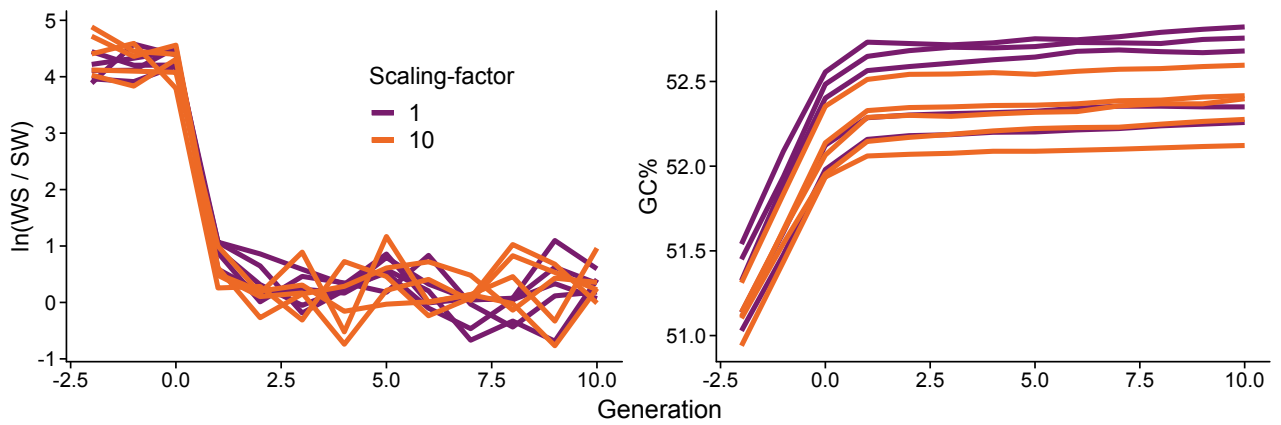


Figure S2. The simulated extent of gBGC with scaled and unscaled parameters. Time in number of generations ($\times 10^4$ for scaled, $\times 10^5$ for unscaled) are shown in horizontal axes. At time zero, population switches to 5% outcrossing. Simulations were run with moderate gene conversion rate (2×10^{-5} scaled, 2×10^{-6} unscaled) and 10% GC over AT repair bias.

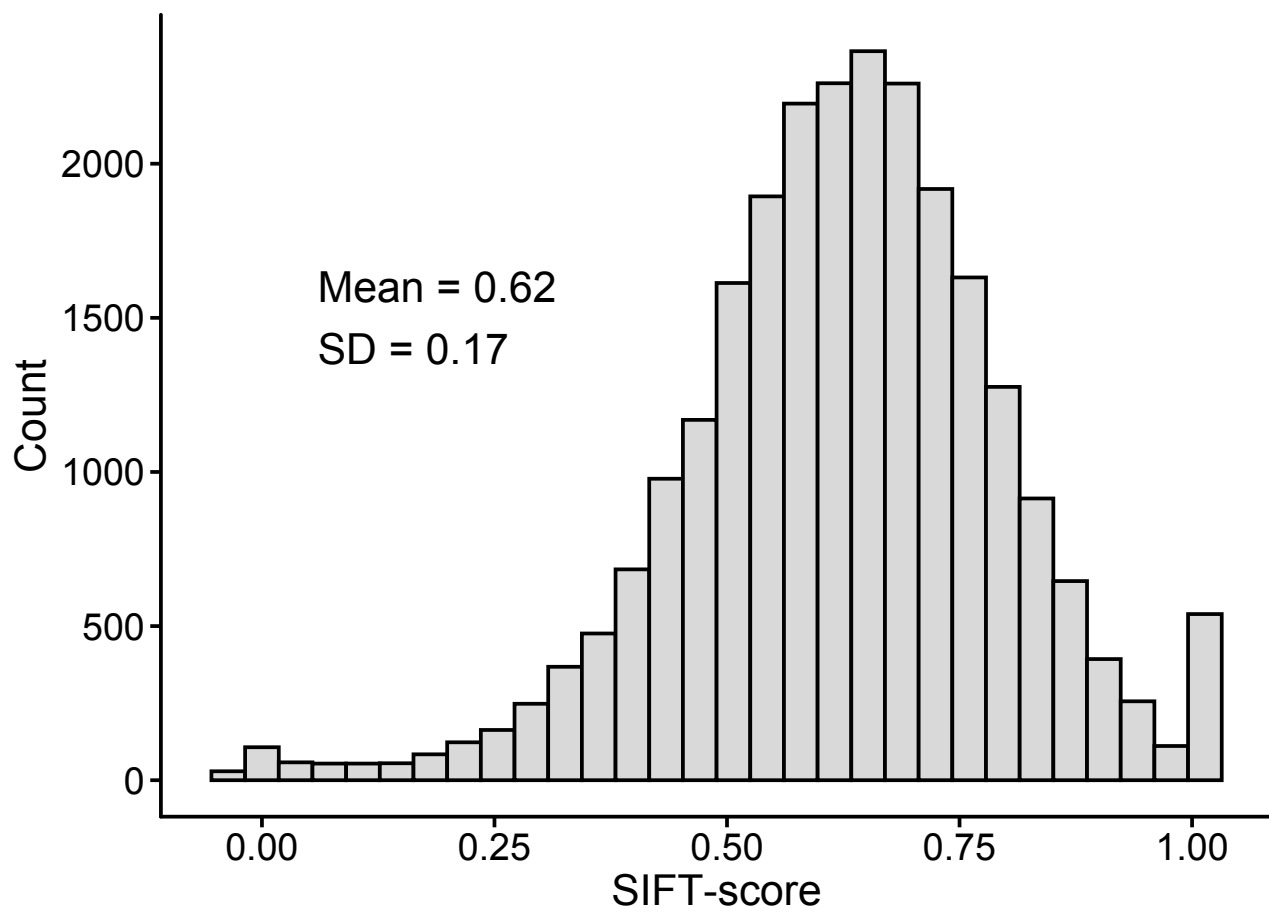


Figure S3. The genome-wide distribution of SIFT-scores.

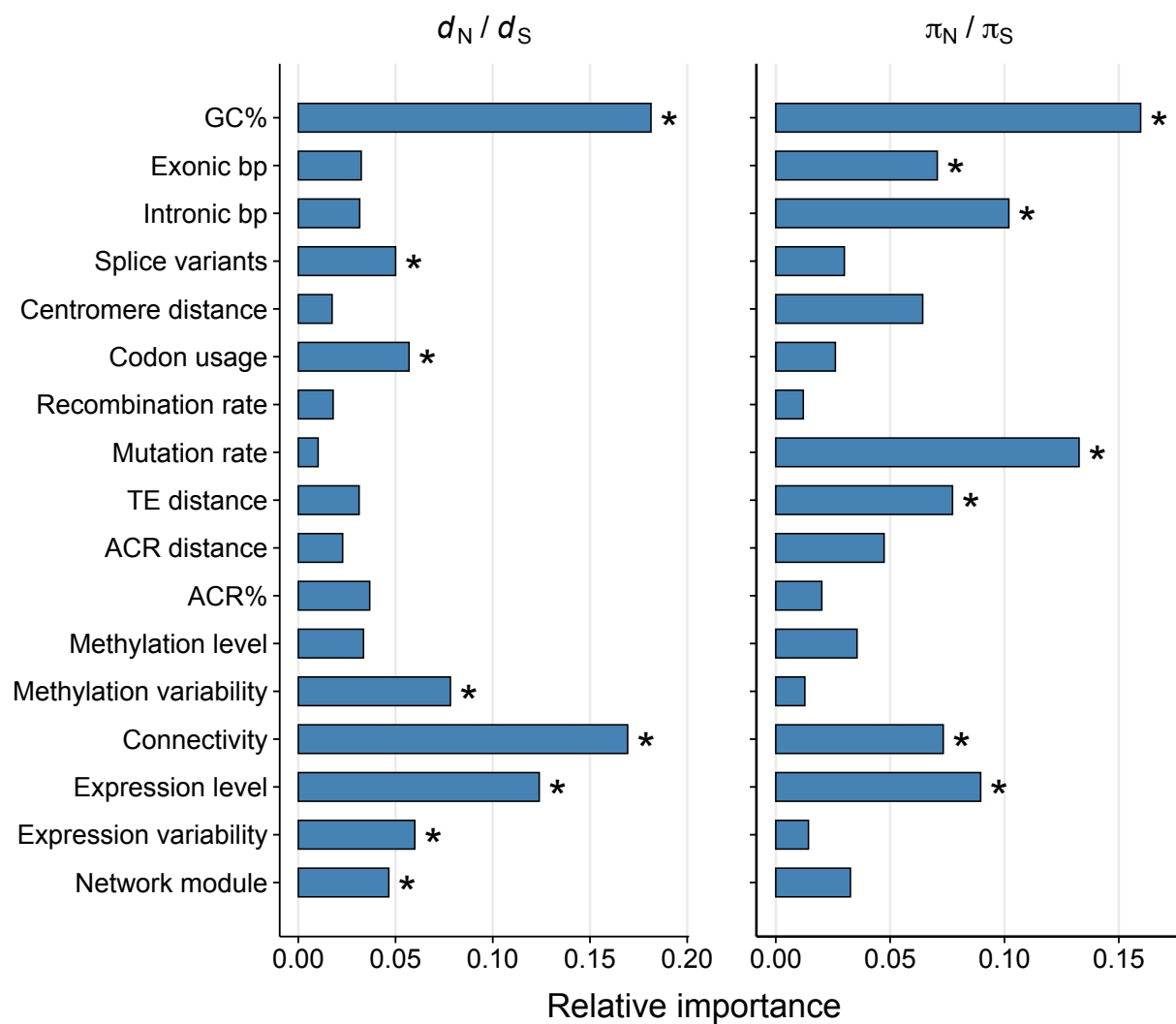


Figure S4. Variable importance from Extra-Trees models for d_N/d_S and π_N/π_S .

* $P < 0.05$ (Bonferroni corrected).

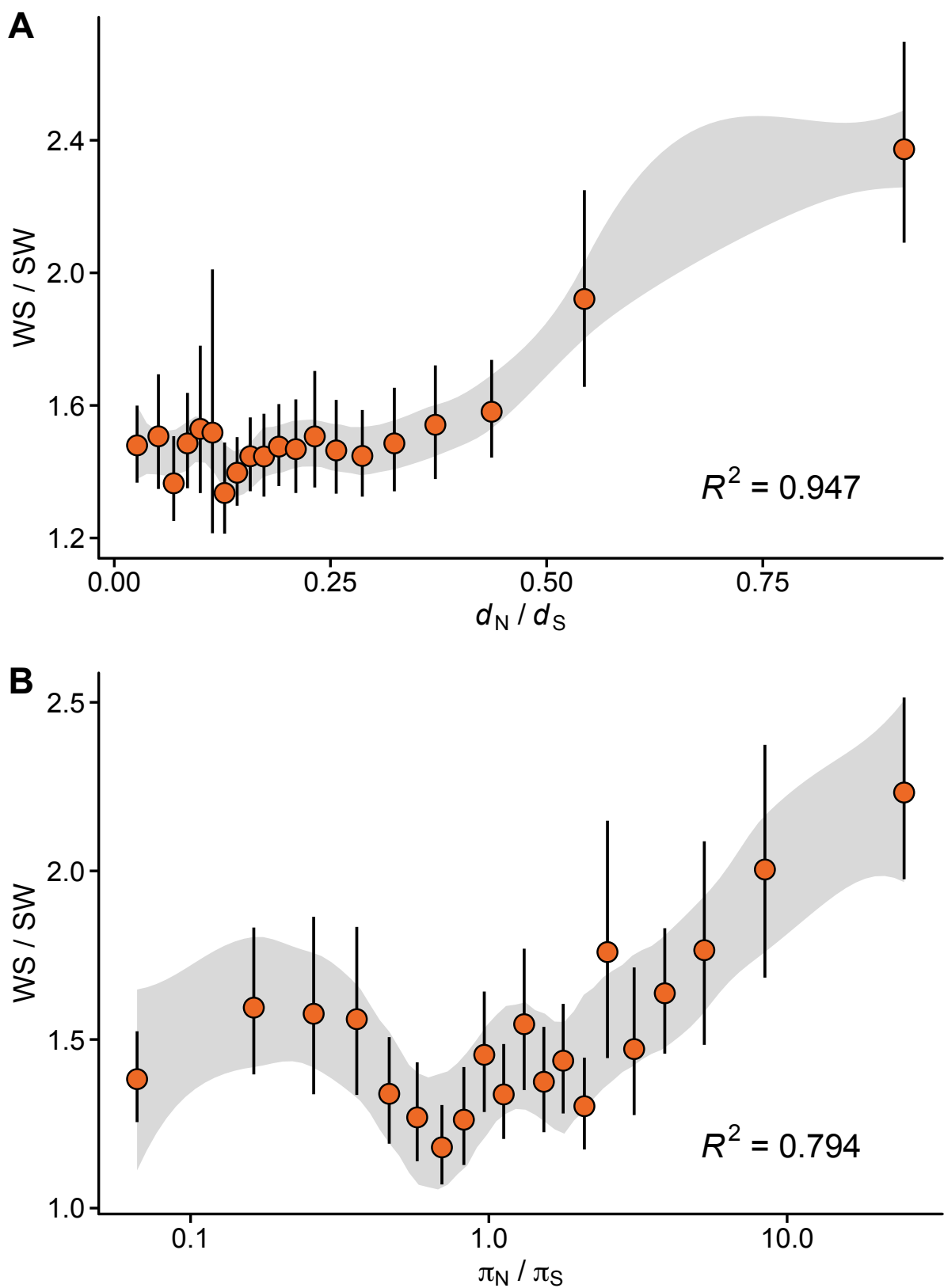


Figure S5. Relationship between WS/SW , a measure of gBGC, and d_N/d_S (A) and π_N/π_S (B). Data were split into 20 bins of equal size. Figure shows means (circles) and 95% CIs (error bars) estimated for each bin. Also shown are 95% CI and R^2 for a loess-model (shaded area) fit on the binned data.

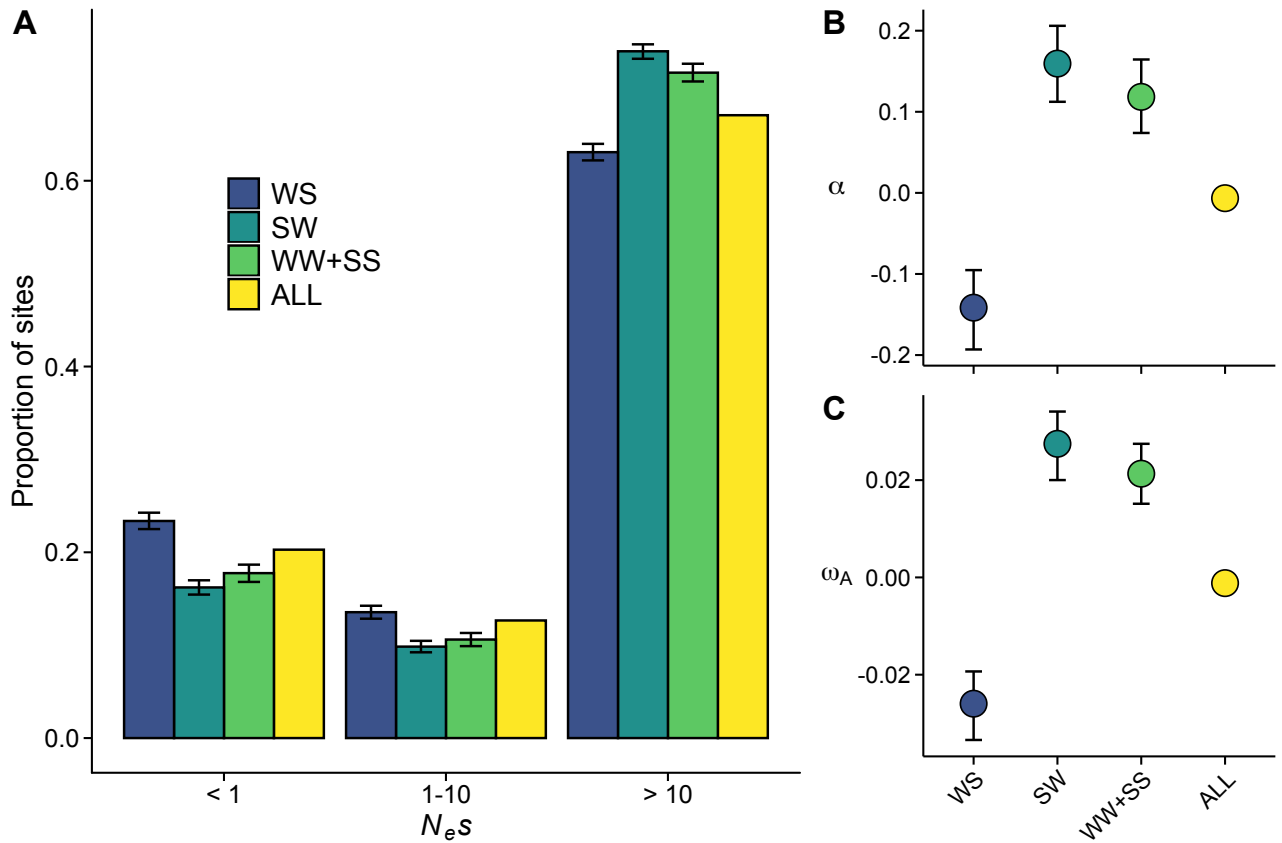


Figure S6. Apparent strength of negative and positive selection at different DAF-sites after removing hypermutable mGC sites. **A:** The distribution of fitness effects (DFE). Nonsynonymous sites were divided into three bins based on the strength of purifying selection ($N_e s$): nearly neutral, intermediate, and highly deleterious, respectively. **B:** The proportion of sites fixed by positive selection (α). **C:** The rate of adaptive substitutions relative to the neutral mutation rate (ω_A). For all three figures, error bars show 95% CIs.

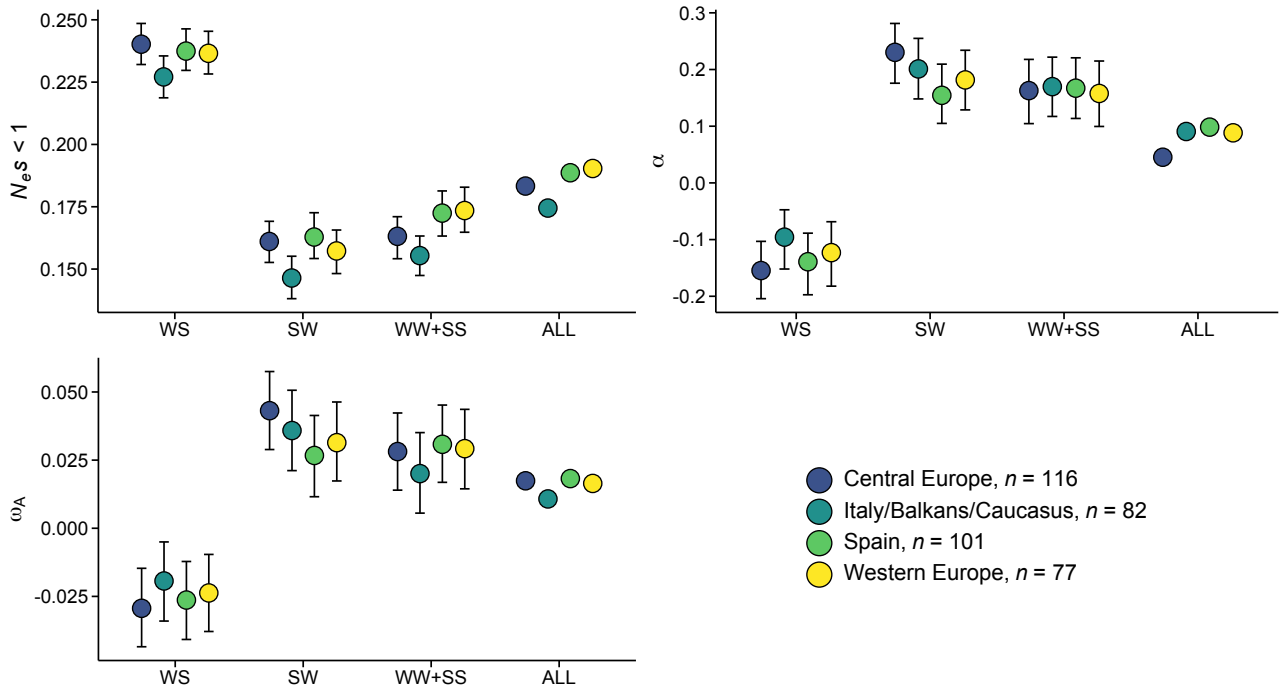


Figure S7. The proportion of nearly neutral mutations ($N_e s < 1$), the proportion of sites fixed by positive selection (α), and the rate of adaptive substitutions relative to the neutral mutation rate (ω_A) for four largest admixture groups. Error bars show 95% CIs.

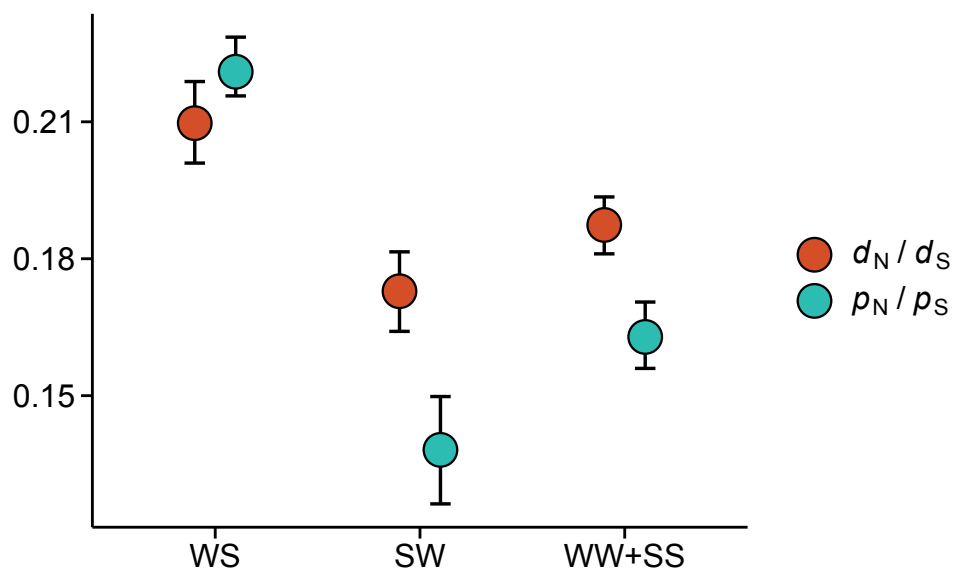


Figure S8. d_N/d_S and p_N/p_S at different DAF-sites for *A. thaliana*. Error bars show 95% CIs.

Table S1. Spearman's rank correlation between genomic features and measures of selective constraint.

Feature	d_N/d_S		π_N/π_S	
	Pairwise	Partial ¹	Pairwise	Partial ¹
GC%	-0.21*	-0.15*	-0.18*	-0.15*
Exonic bp	~0	~0	~0	~0
Intronic bp	-0.03*	~0	-0.04*	~0.04*
Splice variants	0.07*	0.07*	0.03	0.03
Centromere distance	0.02	0.02	0.07*	0.05*
Codon usage	0.07*	0.03*	0.07*	0.04*
Recombination rate	-0.02	-0.02	-0.07*	-0.05*
Mutation rate	~0	~0	-0.03*	~0
TE distance	0.01	~0	0.02	-0.01
ACR distance	0.11*	~0	0.10*	~0
ACR%	-0.12*	~0	-0.10*	0.01
Methylation level	-0.02*	~0	0.01	~0
Methylation variability	0.09*	~0	~0	~0
Connectivity	-0.18*	-0.04*	-0.10*	~0
Expression level	-0.31*	-0.18*	-0.23*	-0.15*
Expression variability	0.11	0.06*	0.08	0.02

¹Partial correlation after controlling for all other features.

* $P < 0.05$ (Bonferroni corrected).

Table S2. Relative importance and Spearman's rank correlation between genomic features and GC%.

Feature	Relative importance	Spearman's ρ	
		Pairwise	Partial ¹
Exonic bp	0.04*	0.13*	0.03*
Intronic bp	0.11*	-0.24*	-0.08*
Splice variants	0.05*	-0.18*	-0.10*
Centromere distance	0.04*	0.12*	0.03*
Codon usage	0.12*	-0.15*	-0.07*
Recombination rate	0.02*	-0.12*	-0.06*
Mutation rate	0.01*	-0.04*	-0.02
TE distance	~0	0.07*	0.03*
ACR distance	~0	-0.04*	0.06*
ACR%	0.06*	0.11*	0.06*
Methylation level	0.13*	-0.27*	-0.03*
Methylation variability	0.21*	0.26*	0.04*
Connectivity	0.03*	0.11*	0.02
Expression level	0.14*	0.20*	0.22*
Expression variability	0.04*	0.26*	0.03*

¹Partial correlation after controlling for all other features.* $P < 0.05$ (Bonferroni corrected).