Supplementary Information

to

Epistasis and pleiotropy affect the modularity of the genotype-phenotype map of cross-resistance in HIV-1

by

Robert Polster, Christos J. Petropoulos, Sebastian Bonhoeffer, and Frédéric Guillaume

Pleiotropy and epistatic pleiotropy

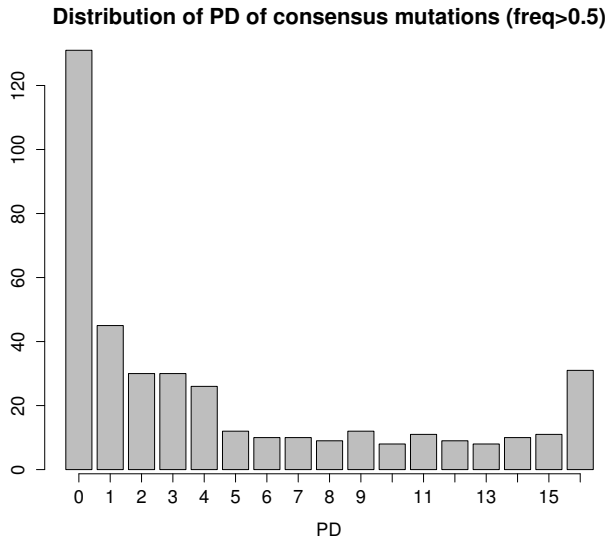

Figure 1: Distribution of PD of the 403 most frequent mutations with frequency > 0.5. They correspond to the amino acid consensus sequence of protease and reverse transcriptase.

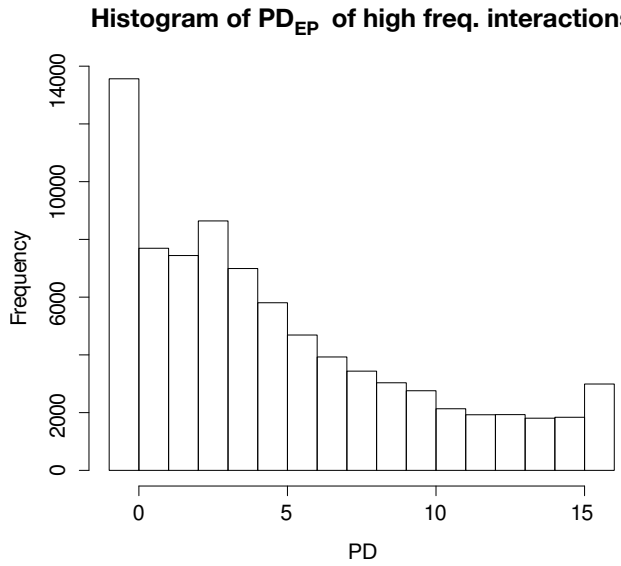

Figure 2: Distribution of pleiotropy among high-frequency mutation interactions (frequency > 0.5)

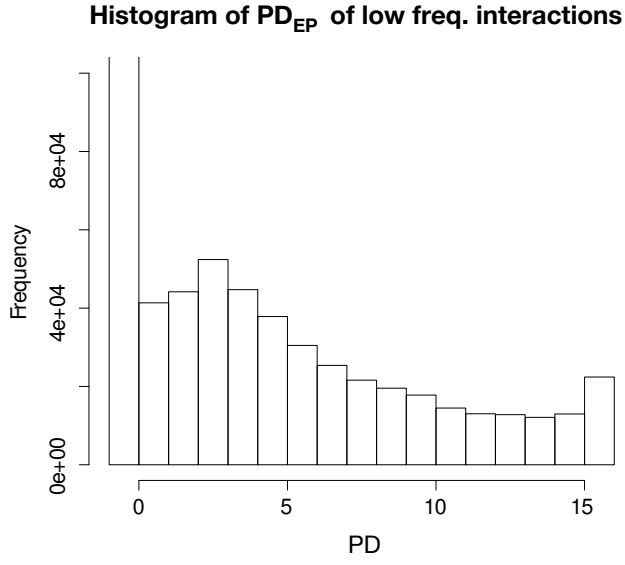

Figure 3: Distribution of pleiotropy among low-frequency mutation interactions (frequency < 0.5).

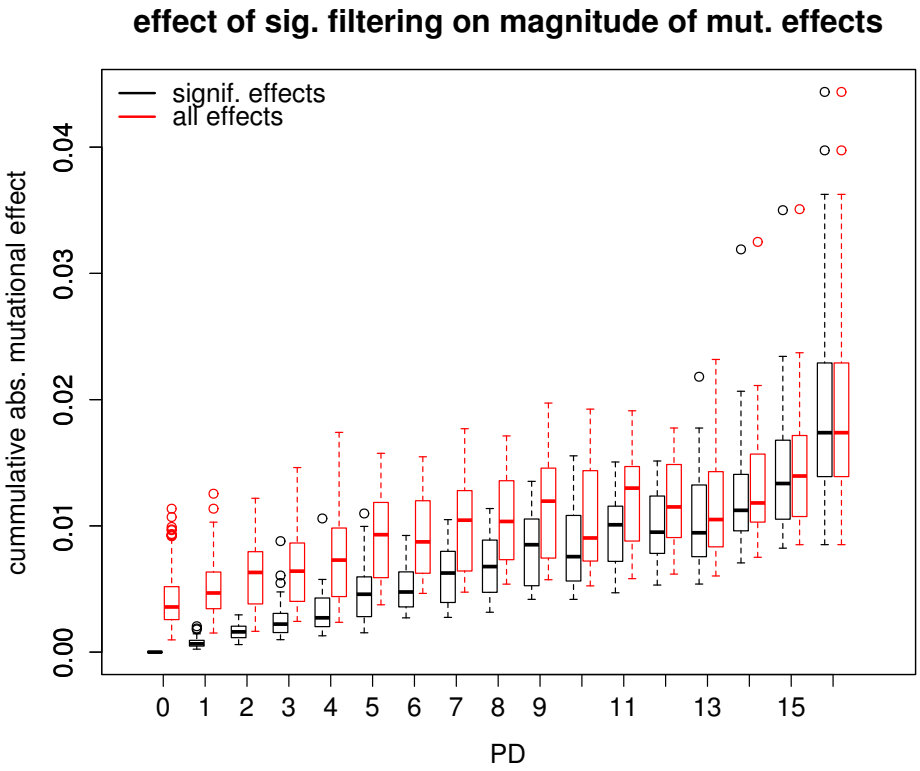

Figure 4: Distributions of the sum of absolute mutational effect sizes in all environments, counting only significant effects (black), or all 16 measured effects (red), for each PD class.

PCA stability

The stability of the three first principal components of the PCA on the ME, MEEP, and SEQ **G**-matrices is tested by downsampling the sequence space and by bootstrapping it. In each case, 10,000 replicates were performed. We used the vector correlation between the original PCs and the replicate PCs as our statistics, reported below as the average absolute value of the 10,000 replicates. The sample sizes used for downsampling are reported in the first column. Sampling is without replacement. Bootstrapping is performed by sampling with replacement within the whole sequence space.

Table 1

	ME			**MEEP**			**SEQ**		
sample size	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
10	0.9968	0.9024	0.5234	0.8791	0.6377	0.4760	0.9137	0.6381	0.4382
100	0.9998	0.9939	0.7512	0.9924	0.9638	0.8061	0.9944	0.9189	0.7678
1000	0.9999	0.9994	0.9373	0.9993	0.9967	0.9788	0.9994	0.9923	0.9783
5000	0.9999	0.9998	0.9913	0.9999	0.9994	0.9962	0.9999	0.9985	0.9963
10000	0.9999	0.9999	0.9960	0.9999	0.9997	0.9983	0.9999	0.9993	0.9983
25000	0.9999	0.9999	0.9988	0.9999	0.9999	0.9995	0.9999	0.9998	0.9995
bootstrap	0.9999	0.9999	0.9994	0.9999	0.9999	0.9997	0.9999	0.9998	0.9997
C.I. 2.5%	0.9999	0.9999	0.9969	0.9999	0.9998	0.9988	0.9999	0.9996	0.9991
C.I. 97.5%	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Hierarchical clustering of viral fitness among drugs

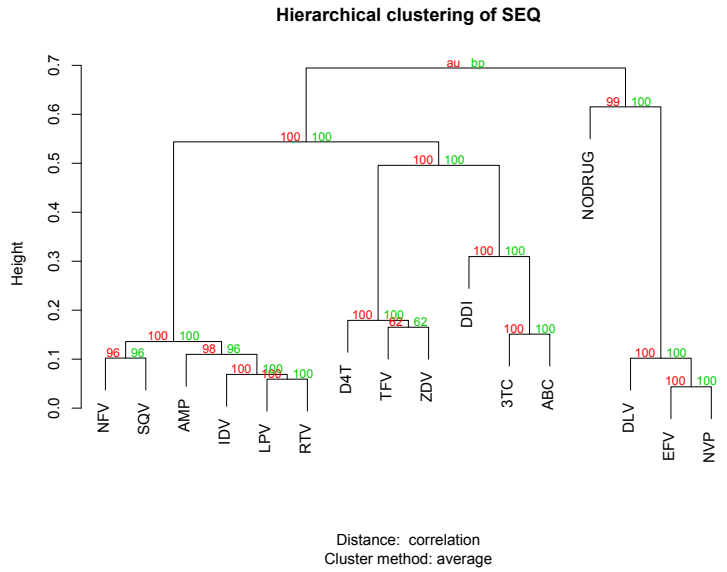

Figure 5: Hierarchical clustering of HIV-1 sequence fitness from the raw data (SEQ matrix). Bootstrap significance support is indicated for approximately unbiased p-values (au, red) and bootstrap probability p-values (bp, green) obtained from the `pvclust` package in R.

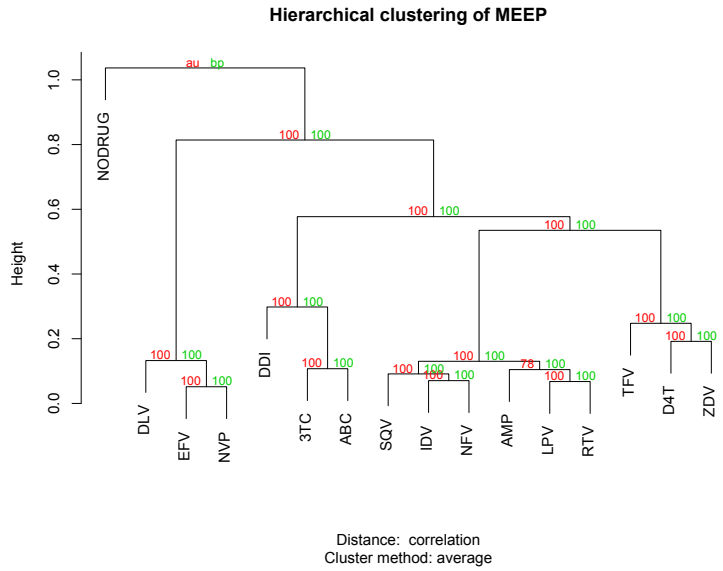

Figure 6: Hierarchical clustering of HIV-1 sequence fitness reconstructed from the MEEP model.

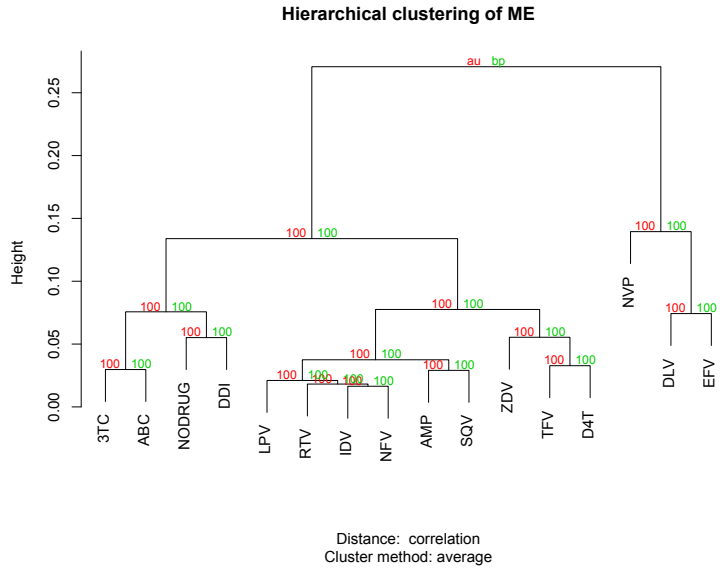

Figure 7: Hierarchical clustering of HIV-1 sequence fitness reconstructed from the main effects of the MEEP model.