Estimating the respective contributions of human and viral genetic variation to HIV control

István Bartha[1,2], Paul McLaren[1,2], Chanson Brumme[3], Richard Harrigan[3], Amalio Telenti[4] and Jacques Fellay[1,2]

[1]École Polytechnique Fédérale de Lausanne, School of Life Sciences, Lausanne, Switzerland, [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland,

[3]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada, [4]J. Craig Venter Institute, La Jolla, USA

Contact address: istvan.bartha@epfl.ch

SUMMARY

Using paired human and viral genotypes along with data on clinical outcome, we estimate that approximately 33% of variation in setpoint viral load is attributable to host and viral genotypic factors.

We also show that the main predictor is the viral genotype and host determination of viral load is largely dependent on its ability to select for viral variants.

INTRODUCTION

Viral load setpoint is a major correlate of HIV disease progression. Genome-wide association studies have identified common human polymorphisms that together explain no more than 15% of its phenotypic variance. Concurrently, studies of the impact of HIV genetic diversity on viral load have produced highly variable estimates (20-46%). Here we present a joint assessment of the respective contributions of human and viral variation to HIV viral load at setpoint.

RESULTS

We estimated the variance of setpoint viral load attributable to the following predictors:

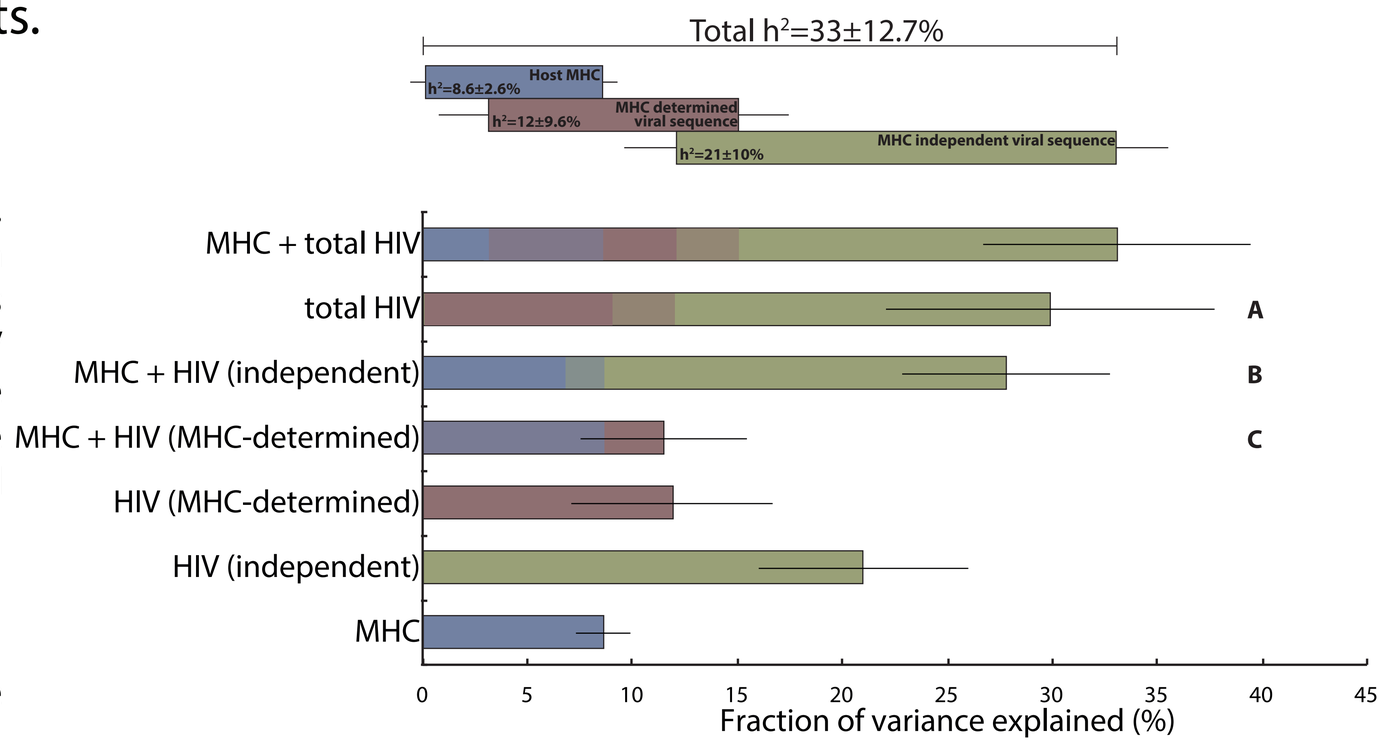
1) human genetic relatedness based on common polymorphisms of the Major Histocompatibility Complex (**MHC**),

2) part of the viral sequence, which is not under host selective pressure, therefore **independent of MHC** (2/3 of the sequence),

3) part of the viral sequence, which is **determined by MHC** (1/3 of the sequence).

Estimating the MHC attributable heritability of HIV viral load yielded an estimate of h^2=8.6% (SD=2.6%). The MHC independent viral sequence explained 21% (SD=10%) of the phenotypic variation. The MHC-determined viral sequence showed an estimated h^2=11.9% (SD=9.6%).
The combination of the three predictors explained 33.1% (SD=12.7%) of the variance of setpoint viral load. The majority (at least 29.9%) of that 33.1% heritability is attributable to the viral sequence, as the MHC contributes little to the combination of the two viral predictors (**Table** and **Figure**).

The three predictors are not independent because the explained variance of the combined predictors is less than the sum of the explained variance of the individual predictors.



A) The overlap between contributions of MHC-determined and MHC-independent viral variation is likely attributable to non-random correlations between viral variants (i.e. linkage disequilibrium in viral sequence).

B) The overlap between contributions of host MHC and MHC-independent viral variation suggests that we didn't have the statistical power to identify all viral variants that are under host pressure.

C) The fact that the host does not add additional explanatory power to the MHC-determined viral sequence excludes that MHC influences viral load independent of the viral genotype.

CONCLUSION

This is the first estimate of the combined and respective contributions of the host and the viral genomes to the observed variability of HIV viral load. We showed that both the pathogen and host genomes have detectable impacts on the clinical outcome of infection, which are however not independent. These results suggest that the main predictor of clinical outcome is the viral genotype and that host determination of viral load is largely dependent on its ability to select viral variants.

Models	MHC		HIV (MHC independent)		HIV (MHC-determined)				
	h^2	$SD(\%)$	h^2	$SD(\%)$	h^2	$SD(\%)$	total h^2	total $SD(\%)$	n
MHC	8.6	2.6					8.6	2.6	1488
HIV (MHC independent)			21.0	10.0			21.0	10.0	1045
HIV (MHC-determined)					11.9	9.6	11.9	9.6	890
MHC + HIV (MHC independent)	4.7	1.9	23.1	9.8			27.8	9.9	1029
MHC + HIV (MHC-determined)	5.8	2.3			5.7	7.6	11.5	7.9	874
total HIV			26.9	13.2	3	8.5	29.9	15.7	886
MHC + total HIV	3.1	1.3	30.0	12.0	0	4.0	33.1	12.7	870

METHODS

Human genotype data across the MHC region, full-length consensus HIV sequences and setpoint viral load results were available for 1029 treatment naïve individuals of European ancestry, infected with HIV-1 subtype B.

Heritability estimation was carried out with GCTA[1] using kernel matrices representing human and viral relatedness.

The human genetic relatedness matrix was derived from 27 common polymorphisms of the MHC shown to strongly influence viral load selected by LASSO.

The relatedness matrix of the viral samples was computed from phylogenetic trees as follows. Bootstrapped phylogenetic trees were inferred by RAxML, and rooted to the group M ancestral sequence. The phylogenetic covariance matrix was derived from the phylogenetic trees by taking the branch length of the shared ancestry (the distance from the root to the most recent common ancestor) [2]. This assumes that the trait is evolving in a neutral manner on the tree, i.e. the phenotypic relatedness is proportional to the amount of shared ancestry. The covariance matrix was then converted to a correlation matrix in order to have ones in the diagonal.

Regions of the viral sequence under host selective pressure were identified by logistic regression as described in [3], with a liberal p-value threshold of $1x10^{-6}$.

Clinical site was included as fixed effect covariate in the analyses.

1. J. Yang et al., Nat. Genet. 42, 565–9 (2010). 2. E. Housworth, E. P. Martins, M. Lynch, Am. Nat. 163, 84–96 (2004). 3. I. Bartha et al., Elife. 2, e01123 (2013).