SUPPLEMENTAL MATERIAL

Hepatitis C virus NS3 sequence diversity and antiviral resistance-associated variant frequency in HCV/HIV coinfection

Table S1. Primer ID primers for cDNA synthesis.

Unique Sequence	Primer ID	Bar-code	HCV-complementary Sequence
ACCTGGCAAGCACGCTCTGGCCTTGAA	**NNNNNNNN**CT	**NNNN**	GAACACCGGGGACCTCATGGTTGTCTC
....	**NNNNNNNN** ..	**NNNN**	............T..............
....	**NNNNNNNN** ..	**NNNN**	...........T..G...........
....	**NNNNNNNN** ..	**NNNN**	A.........................
....	**NNNNNNNN** ..	**NNNN**	.........A................
....	**NNNNNNNN** ..	**NNNN**	...................T.......
....	**NNNNNNNN** ..	**NNNN**	.........................A..
....	**NNNNNNNN** ..	**NNNN**	.....................A.......
....	**NNNNNNNN** ..	**NNNN**	......T....................

Table S2. Most frequent primer ID sequences in two independent sequencing runs.

Subject	Run 1		Run 2	
	Primer ID	**Count**	**Primer ID**	**Count**
M1	CACCGTAT	58	TATCCGGC	22
	TTCGTCGA	40	CTACGACT	22
	CCCGCCCC	39	ACCGAACA	22
	CGATTGCC	38	CCCGCCCC	20
	ACGTGTTC	38	CCCATATA	20
	ATCACATA	37	GAGATTCA	18
	ACCAGTAC	36	GCGCCGCG	17
	CCGCCACC	35	CACCCCGC	17
M2	AACATCCC	37	CGGGTGAA	26
	ATGTGCTA	36	AATGCAAT	26
	AGATGATA	35	TGGCGAAA	22
	TTAACATA	33	GTCGATAT	22
	ACATGCTA	33	ATAACTAT	22
	CATACTCA	32	GTGAGCAC	21
	AACATGAC	32	ACATAGAA	21
C1	CCCCGCCC	19	CCCGCCCC	27
	CTCCCCAC	17	CCCCTCCC	22
	CCGCCCAC	16	CCCCCGCC	21
	CCCCTCCC*	16	CCCCCACC	20
	CCCCCGCC	16	CCCTCCCC	19
	CCCTCCCC	15	CCCCACCC	18
C2	CCTAGTCT	77	AATAACAA	71
	AATACCAA	63	GCCGCGAG	60
	GACACAGG	58	TATTGTTA	55
	GTTGAATC	54	ACCAATAT	54
	CCTACATG	54	TAGCGTAC	52
	TTCATTAA	53	ATCGGAGG	52
	ATTCAAGA	53	ATTGATAA	51

*Primer ID sequences common to both runs are color-coded.

Table S3. Multivariable Linear Regression Models for Pairwise Nucleotide Diversity (π).

N	Variable	β estimate	S.E.[a]	p-value
40	HIV co-infection	-0.00471	0.00191	0.02
	Sequencing depth ($\log_{10}$)	-0.00340	0.00283	0.24
	Serum/plasma HCV RNA ($\log_{10}$ copies/mL)	0.00269	0.00145	0.07

N	Variable	β estimate	S.E.	p-value
38*	African-American race	-0.00319	0.00202	0.12
	Sequencing depth ($\log_{10}$)	-0.00473	0.00301	0.12
	Serum/plasma HCV RNA ($\log_{10}$ copies/mL)	0.00234	0.00156	0.14

N	Variable	β estimate	S.E.	p-value
38[b]	HIV co-infection	-0.00382	0.00229	0.11
	African-American race	-0.00145	0.00223	0.52
	Sequencing depth ($\log_{10}$)	-0.00365	0.00300	0.23
	Serum/plasma HCV RNA ($\log_{10}$ copies/mL)	0.00270	0.00153	0.09

[a]S.E. = standard error.
[b]Excludes subjects with race = Other.

Table S4. Multivariable Linear Regression Models for Tajima's *D*.

N	Variable	β estimate	S.E.[a]	p-value
40	HIV co-infection	-0.11445	0.06787	0.10
	Sequencing depth ($\log_{10}$)	-0.07328	0.10063	0.47
	Serum/plasma HCV RNA ($\log_{10}$ copies/mL)	0.04852	0.05150	0.35

N	Variable	β estimate	S.E.	p-value
38[b]	African-American race	-0.12238	0.06929	0.09
	Sequencing depth ($\log_{10}$)	-0.09367	0.10299	0.37
	Serum/plasma HCV RNA ($\log_{10}$ copies/mL)	0.03580	0.05332	0.51

[a]S.E. = standard error.
[b]Excludes subjects with race = Other.

Table S5. Multivariable Linear Regression Model for Shannon Index.

N	Variable	β estimate	S.E.	p-value
38[b]	African-American race	-0.83755	0.37390	0.03
	HIV co-infection	-0.23619	0.37390	0.53

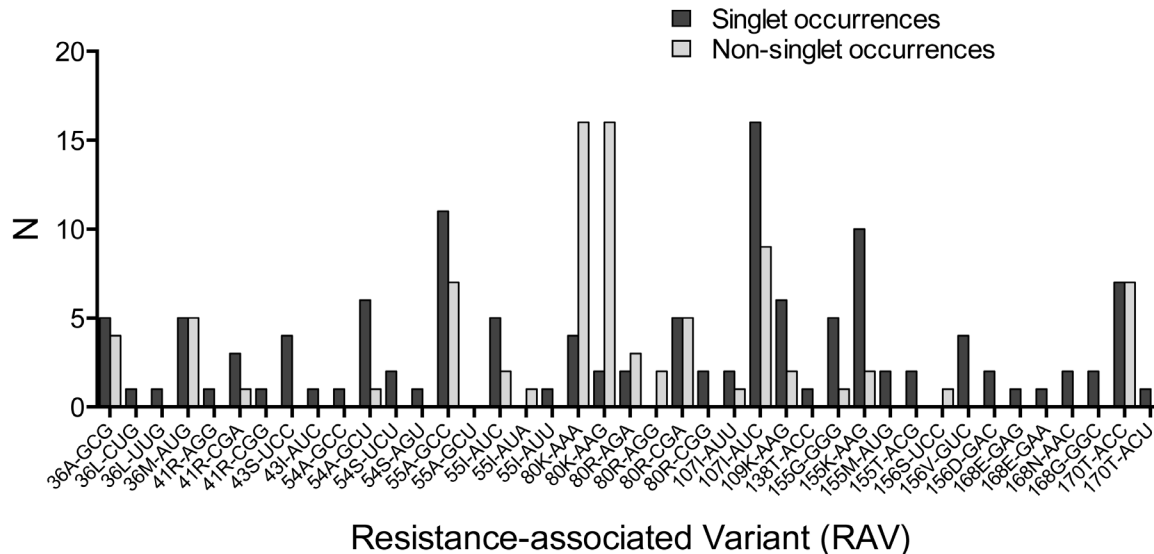[a]S.E. = standard error.
[b]Excludes subjects with race = Other.

Figure S1

Figure S1. Number of subjects with singlet and non-singlet occurrences of individual RAVs. There was substantial association between singlet and non-singlet occurrences of specific RAVs, excluding Q80K (Spearman's rho = 0.606, Monte-Carlo permutation* p<0.001).

*http://www2.sas.com/proceedings/sugi27/p251-27.pdf